Università degli Studi di Ferrara

DOCTORAL COURSE IN PHYSICS

CYCLE XXXIV

COORDINATOR Prof. Eleonora LUPPI

DEEP LEARNING IN GALAXY CLUSTERS

Scientific/Disciplinary Sector (SDS) FIS/05

Candidate

Dr. Giuseppe ANGORA

Supervisor

Prof. Piero ROSATI

Co-supervisor

Dr. Massimo Meneghetti (INAF-BO)

YEARS 2018/2021

Università degli Studi di Ferrara

Sezioni

Dottorati di ricerca

Il tuo indirizzo e-mail

giuseppe.angora@unife.it

Oggetto:

Dichiarazione di conformità della tesi di Dottorato

Io sottoscritto Dott. (Cognome e Nome)

Angora Giuseppe

Nato a:

Torre del Greco

Provincia:

Napoli

Il giorno:

14/10/1989

Avendo frequentato il Dottorato di Ricerca in:

Fisica

Ciclo di Dottorato

34

Titolo della tesi:

Deep Learning in Galaxy Clusters

Titolo della tesi (traduzione):

- Deep Learning in Galaxy Clusters: Deep Learning in Ammassi di Galassie -

Tutore: Prof. (Cognome e Nome)

Rosati Piero

Settore Scientifico Disciplinare (S.S.D.)

FIS/05

Parole chiave della tesi (max 10):

galaxy clusters, strong gravitational lensing, data analysis, machine learning, deep learning, galaxy redshift

Consapevole, dichiara

CONSAPEVOLE: (1) del fatto che in caso di dichiarazioni mendaci, oltre alle sanzioni previste dal codice penale e dalle Leggi speciali per l'ipotesi di falsità in atti ed uso di atti falsi, decade fin dall'inizio e senza necessità di alcuna formalità dai benefici conseguenti al provvedimento emanato sulla base di tali dichiarazioni; (2) dell'obbligo per l'Università di provvedere al deposito di legge delle tesi di dottorato al fine di assicurarne la conservazione e la consultabilità da parte di terzi; (3) della procedura adottata dall'Università di Ferrara ove si richiede che la tesi sia consegnata dal dottorando in 2 copie, di cui una in formato cartaceo e una in formato pdf non modificabile su idonei supporti (CD-ROM, DVD) secondo le istruzioni pubblicate sul sito : http://www.unife.it/studenti/dottorato alla voce ESAME FINALE – disposizioni e modulistica; (4) del fatto che l'Università, sulla base dei dati forniti, archivierà e renderà consultabile in rete il testo completo della tesi di dottorato di cui alla presente dichiarazione attraverso l'Archivio istituzionale ad accesso aperto "EPRINTS.unife.it" oltre che attraverso i Cataloghi delle Biblioteche Nazionali Centrali di Roma e Firenze. DICHIARO SOTTO LA MIA RESPONSABILITA': (1) che la copia della tesi depositata presso l'Università di Ferrara in formato cartaceo è del tutto identica a quella

presentata in formato elettronico (CD-ROM, DVD), a quelle da inviare ai Commissari di esame finale e alla copia che produrrà in seduta d'esame finale. Di conseguenza va esclusa qualsiasi responsabilità dell'Ateneo stesso per quanto riguarda eventuali errori, imprecisioni o omissioni nei contenuti della tesi; (2) di prendere atto che la tesi in formato cartaceo è l'unica alla quale farà riferimento l'Università per rilasciare, a mia richiesta, la dichiarazione di conformità di eventuali copie. PER ACCETTAZIONE DI QUANTO SOPRA RIPORTATO

Dichiarazione per embargo

6 mesi

Richiesta motivata embargo

1. Tesi in corso di pubblicazione

Liberatoria consultazione dati Eprints

Consapevole del fatto che attraverso l'Archivio istituzionale ad accesso aperto "EPRINTS.unife.it" saranno comunque accessibili i metadati relativi alla tesi (titolo, autore, abstract, ecc.)

Firma del dottorando

Ferrara, li 03/03/2022 Firma del Dottorando Giuseppe Angora

Firma del Tutore

Visto: Il Tutore Si approva Firma del Tutore Piero Rosati

Abstract

Galaxy clusters play an important role in modern cosmology and astrophysics. They act as cosmic laboratories where we can study galaxy formation and evolution, and improve our understanding of the nature of Dark Matter using dynamical and gravitational lensing methods. As powerful gravitational lenses, clusters can be used as natural cosmic telescopes thus extending our detection limit of faint sources and revealing the most distant galaxies. Over the last decades, strong gravitational lensing has been widely used to study the mass distribution on both galaxy and cluster scales. In this context, dedicated surveys with Hubble Space Telescope (HST) and ground-based extensive spectroscopic campaigns have provided data with extraordinary quality, particularly over the last decade. The richness of these data sets, however, cannot be compared with the impressive data volume that upcoming surveys (such as ESA Euclid satellite, Vera Rubin Observatory or James Webb Space Telescope) will generate over the next years. The volume and the complexity of these new datasets can be efficiently dealt using machine learning and deep learning methods, which enable the exploration of hidden correlations within a multi-dimensional parameter space, a discipline which has had a phenomenal development in recent years in many different fields, becoming the dominant methodology over standard methods. In this thesis, we take advantage of this multidisciplinary tool to enable a number of scientific investigations of cluster internal structure and background source population. Specifically, we addressed three complementary issues which exploit imaging and spectroscopic observations of a sample of a dozen galaxy clusters.

As a first application, we implemented deep learning architectures to select galaxy cluster members in galaxy clusters, in the redshift range $0.2 - 0.6$, which is a critical first step for a variety of cluster studies, such as galaxy evolution in dense environments, cluster mass estimates, strong lensing models. By using HST multi-band images alone, convolution neural networks (CNNs) were used to disentangle member galaxies from background and foreground sources, once they were suitably trained with a large sample of spectroscopically confirmed sources in several cluster fields (from VLT VIMOS and MUSE observations), thus avoiding the complicated and time consuming photometric measurement process. We performed several experiments by studying the model dependence on member redshift, magnitude and colours, and characterised limits and capabilities of several CNN models. We compared the deep network performance with other conventional photometric-based approaches, finding that CNNs are able to classify members with a superior purity-completeness rate ($\gtrsim 90\%$), and showing stable results across the parameter space. In view of the currently available and upcoming large area surveys (e.g. Rubin LSST observatory), we also repeated these experiments using multi-band images of a subsample of galaxy clusters from the Subaru 8 m telescope, finding acceptable performances, opening interesting prospects for the next generation surveys.

As a second step, we focused on the identification of galaxy-galaxy strong lenses (GGSL) in galaxy clusters, which can be used to study the internal mass distribution of clusters, traced by the sub-halo population around cluster member galaxies, and can later be compared with cosmological simulations. In this work, we opted for a methodology that combines the need to simulate a large number of GGSL to train deep neural networks, while maintaining the imaging

complexity of real observations. By exploiting high-precision cluster lens models available for 8 clusters (with redshift in $0.2 - 0.6$), we used the estimated deflection angle maps to simulate thousands of realistic strong-lenses in real HST images by ray tracing background sources on the lens plane. Thus, GGSLs are reproduced by taking into account both the sub-halo and the cluster scale mass distributions which affect the morphology, brightness and frequency of galaxy-scale lensing events. We found that deep networks trained on this realistic dataset are able to detect a large fraction of real strong-lensing events, with a limited number of false negative events. We characterised the model performance by exploring their dependence on several parameters for the sources and the galaxy lenses. Moreover, we processed hundreds of cluster members, which are either spectroscopically confirmed or selected with the aforementioned CNN methodology, to test deep model generalisation capabilities and to search for galaxy-galaxy strong-lens candidates.

Finally, we implemented a 3D spectroscopy cross-correlation tool on the MUSE integral field spectrograph data to measure redshifts in an automated and computationally efficient fashion, a crucial ingredient to enable the cluster investigations described above. In fact, the mining of spectroscopic information allows us to build datasets used to train neural networks, to confirm cluster galaxy membership, to measure the redshift of the lens and the source in lensing events. Optimised to be executed on graphic processors, this tool is able to process an entire MUSE dataset in a few tens of seconds, by cross-correlating the 90 000 spectra included in the data cube with a sample of spectral templates. We validated the tool by comparing the redshift measurements of ~ 270 galaxies, with $z \in (0, 7)$, in the cluster MACS J0416.1-2403 with those obtained with conventional techniques relying on heavy user interactions. Even though the tool is still under development, particularly when dealing with low signal-to-noise spectra, our preliminary results appear rather promising and will soon be applied routinely on MUSE data. In addition, such a 3D cross-correlation provides reliable velocity maps of lensed galaxies with a vast improvement over pixel-by-pixel manual measurements by conventional methods.

The methodologies developed in this thesis can be extended beyond the HST imaging data with a relatively modest effort and promise to have important applications with the upcoming next generation facilities. These include for example deep observations of galaxy clusters with the James Webb Space Telescope, and generally wide area surveys from space and ground (Euclid, Rubin LSST) which will discover up to 10^5 galaxy clusters and groups, and will require automated machine learning based methodologies to fully exploit their astrophysical and cosmological content.

Contents

Acknowledgments

First and foremost, I would like to express my deepest gratitude to Piero Rosati that guided me during the PhD years. His ideas, experience and teachings were fundamental for the development of this work, for my growth as a researcher and for my future career.
I owe a special thanks to Amata Mercurio, Massimo Brescia and Massimo Meneghetti for the critical discussions and the crucial advice throughout the PhD years, this work would not have been possible without their support and nurturing. This thesis is the result of a rewarding collaboration among exceptional researchers: Claudio Grillo, Eros Vanzella and Pietro Bergamini. I would like to thank them for their indispensable suggestions.
Last, but not least, I thank Daniela and my family, which support my decisions and my outlandish ideas. Their encouragement and affection have been indispensable for all my accomplishments.

List of publications

Title: The search for galaxy cluster members with deep learning of panchromatic HST imaging and extensive spectroscopy.
Authors: Angora G., Rosati P., Brescia M., Mercurio A., Grillo C., Caminha G., Meneghetti M., Nonino M., Vanzella E., Bergamini P., Biviano A. and Lombardi M.
Status and Journal: Astronomy & Astrophysics, Volume 643, id.A177, 25 pp. Angora et al. 2020

Title: A new high-precision strong lensing model of the galaxy cluster MACS J0416.1-2403. Robust characterization of the cluster mass distribution from VLT/MUSE deep observations.
Authors: Bergamini P., Rosati P. ; Vanzella E., Caminha G. B., Grillo C., Mercurio A., Meneghetti M., Angora G., Calura F., Nonino M., Tozzi P.
Status and Journal: Astronomy & Astrophysics, Volume 645, id.A140, 15 pp. Bergamini et al. 2021

Title: A strong lensing model of the galaxy cluster PSZ1 G311.65-18.48.
Authors: Pignataro G. V., Bergamini P., Meneghetti M., Vanzella E., Calura F., Grillo C., Rosati P., Angora G., Brammer G., Caminha, G. B., Mercurio A., Nonino M., Tozzi P.
Status and Journal: Astronomy & Astrophysics, Volume 655, id.A81, 16 pp. Pignataro et al. 2021

Title: CLASH-VLT: Abell S1063. Cluster assembly history and spectroscopic catalogue.
Authors: Mercurio A., Rosati P., Biviano A., Annunziatella M., Girardi M., Sartoris B., Nonino M., Brescia M., Riccio G., Grillo C., Balestra I., Caminha G. B., De Lucia G., Gobat R., Seitz S., Tozzi P., Scodeggio M., Vanzella E., Angora G., Bergamini P., Borgani S., Demarco R., Meneghetti M., Strazzullo V., Tortorelli L., Umetsu K., Fritz A., Gruen D., Kelson D., Lombardi M., Maier C., Postman M., Rodighiero G. Ziegler B.
Status and Journal: Astronomy & Astrophysics, Volume 656, id.A147, 24 pp. Mercurio et al. 2021

Title: A novel approach to the classification of terrestrial drainage networks based on deep learning and preliminary results on solar system bodies.
Authors: Donadio C., Brescia M., Riccardo A., Angora G., Delli Veneri M., Riccio G.
Status and Journal: Scientific Reports, Volume 11, article id. 5875. Donadio et al. 2021

Title: Deep learning searching for strong lenses in galaxy cluster with HST imaging.
Authors: Angora G., Rosati P., Brescia M., Mercurio A., Meneghetti M., Grillo C., Caminha G., Nonino M., Vanzella E., Bergamini P., Biviano A.
Status and Journal: Close to submission to Astronomy & Astrophysics.

List of Figures

List of Tables

Chapter 1

Introduction

Galaxy clusters are the largest gravitationally bound systems in the Universe. They play a key role in the ongoing understanding of some of the biggest issues in astrophysics and cosmology: the nature of dark matter and dark energy. They are composed of hundreds of galaxies (called "members" or "cluster members") moving at thousands of kilometers per second, trapped within the cluster potential well. Due to their masses and sizes ($\sim 10^{15} M_{\odot}$, ~ 1 Mpc), they serve as cosmic laboratories where studying a wide range of physical and astrophysical phenomena: the formation and evolution of cosmic structures, the galaxy population developments, the chemical evolution of the Universe, plasma physics on the largest scales and relativistic particle acceleration. Furthermore, galaxy clusters, by acting as powerful gravitational lenses, extend the detection limits of faint background sources by revealing high-z galaxies that would not be observable otherwise; their estimated mass profiles provide important constrains on the dark matter distribution and allow to test cosmological models. Actually, the first observational evidence of dark matter has been found by Zwicky (1937) which compared the dynamical mass of clusters measured with the virial theorem, with the luminous mass estimated by considering the galaxy mass-to-luminosity ratio. Now we know that the dark matter is the principal mass component in galaxy clusters (about $80-90\%$), the remaining baryonic part is mainly composed of hot X-ray emitting plasma, while star contribution to the cluster mass is just $1-2\%$.

The first galaxy cluster catalogue was published by Abell (1958), it contained $\sim$ 1700 clusters identified by searching galaxy overdensity within a specified solid angle (combined with further selection criteria). Such approach is still in use today exploiting optical or near-infrared imaging (e.g., Rykoff et al. 2014; Rettura et al. 2014), even if other techniques have been developed: galaxy clusters can be detected by identifying their X-ray emitting gas content (e.g., Pacaud et al. 2016), by using the distortion in cosmic microwave background due to the thermal Sunyaev–Zel'dovich effect (Sunyaev & Zeldovich 1972) at millimeter wavelengths (e.g., Bleem et al. 2015; Planck Collaboration et al. 2016), by searching for significant concentration of Lyman$-\alpha$ emitters (Jiang et al. 2018), or by constructing weak lensing convergence maps, where galaxy clusters appear as peaks (Gavazzi & Soucail 2007; Shan et al. 2012).

According to the cold dark matter scenario, supported by several independent studies (Peebles 1982; Bond et al. 1982; Blumenthal et al. 1982, 1984), galaxies and galaxy clusters were built up through a hierarchical clustering process dynamically dominated by dark matter in an expanding Universe, a model known as ΛCDM, initially proposed by Press & Schechter (1974). Assuming such hierarchical structure formation paradigm of the universe, cosmological numerical simulations clearly describe formation and evolution of galaxy clusters across time (Borgani & Kravtsov 2011; Chiang et al. 2013), as shown in Fig. 1.1: a cluster begins with its collapse on the peaks of the dark matter density field and hierarchically grows through accretion and mergers of small halos.

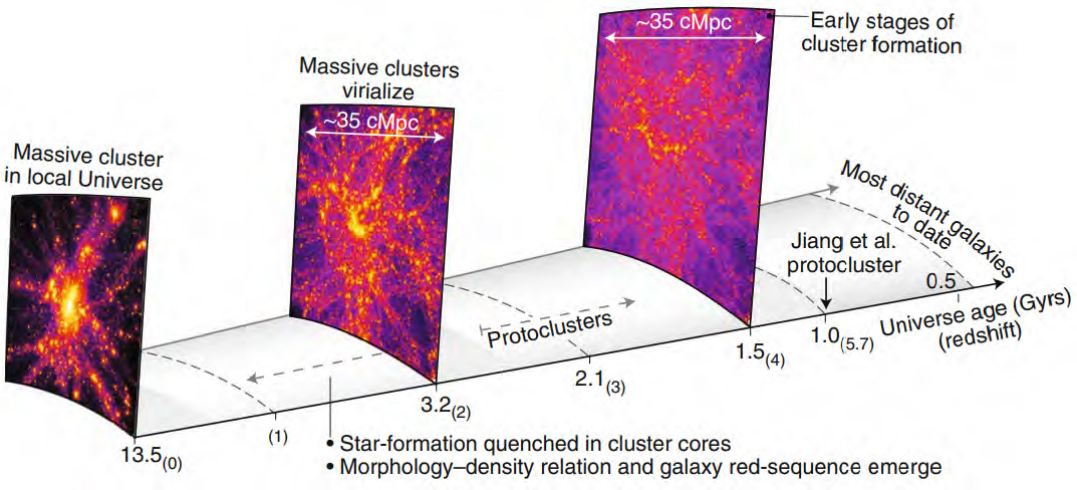


Figure 1.1: A schematic flow of the cosmic evolution of a massive galaxy cluster, starting from the early formation stage ($\sim$ 1 billion years after the Big Bang, $z \sim 4$), until the assembling of a massive bound structure ($M \sim 10^{15} M_{\odot}$, after 10 Gyrs). The axis labels the cosmic time, in terms of age of Universe (Gyrs) and redshfit (between brackets). The three images, describing the dark matter density field, are extracted from cosmological hydrodynamical simulations (Borgani & Kravtsov 2011) at three different redshfits ($z = 4, 2, 0$). Image from Rosati (2018).

In the last decades, strong gravitational lensing has turned out as a powerful tool to measure the total mass distribution in the core of galaxy clusters and so, once the baryonic counterpart has been independently mapped, to study the dark matter distribution. By combining mass tracing achieved through weak leaning (Umetsu et al. 1999; Umetsu & Broadhurst 2008; Umetsu et al. 2011; Umetsu 2013; Umetsu et al. 2014, 2018; Schneider et al. 2000; Hoekstra et al. 2013; Gruen et al. 2013), galaxy dynamics (Stock et al. 2015; Biviano et al. 2013; Balestra et al. 2016) or X-ray analysis (Gómez et al. 2012) with strong lensing techniques, it is possible to reconstruct cluster mass density profile over a large range of radial scales, from kpc to Mpc (Newman et al. 2009, 2011; Merten et al. 2015; Caminha et al. 2017a,b, 2019; Bergamini et al. 2019, 2021b). Furthermore, by characterising cluster substructures, ΛCDM paradigms can be tested by comparing the reconstructed mass profile with N-body or hydrodynamical simulations (Diemand & Moore 2011; Genel et al. 2014; Meneghetti et al. 2020). Finally, due to the flux magnification of lensed sources, galaxy clusters act as cosmic telescopes, allowing to reveal high-z faint (lensed) galaxies (Zheng et al. 2012; Vanzella et al. 2017b, 2020, 2021).

In recent years, dedicated Hubble Space Telescope observations, e.g. Cluster Lensing And Supernova survey with Hubble (CLASH, Postman et al. 2012a), Hubble Frontier Field survey (HFF, Lotz et al. 2017) and Reionization Lensing Cluster Survey (RELICS Coe et al. 2019), have provided supreme-quality multi-band imaging of massive galaxy clusters. Besides space-born surveys, ground-based wide-field imaging of SUBURU/Suprime-Cam (Miyazaki et al. 2002) can be added to extend the analysis to the outer regions of clusters. These observations can be exploited together with an extensive spectroscopic coverage, part of CLASH-VLT VIMOS programme (Rosati et al. 2014), combined with archival observations carried out with the integral field spectrograph MUSE on the VLT (Bacon et al. 2014). Nevertheless, this data richness is just the tip of a deep iceberg: upcoming space and ground based surveys, e.g. the ESA Euclid satellite (Laureijs et al. 2011), the Vera Rubin Observatory (Ivezić et al. 2019) or the James Webb Space Telescope (Gardner et al. 2006), are set to routinely produce tens of terabytes of data of unprecedented quality and complexity on a daily basis: an unprecedented “data-revolution" (also called “data-tsunami" or, with a little of irony, “datageddon") which has been compared with

other human social drastic changes, like the bronze age transition (Hey 2012). Indeed, these volumes of data can be dealt only through a novel framework, delegating most of the work to automatic tools and by exploiting all advances in high-performance computing, machine learning, data science and computer visualisation (Brescia et al. 2018). This paradigm transition has led Astronomy into a *new age*, involving astronomers and data-scientists to develop a new generation of instruments, looking for new methodologies and acquiring new skills in order to handle many different astronomical problems. The ensemble of techniques, methods and approaches to explore these intensive volume of data is known as Data Mining, whose applications within the context of Astronomy led to a new discipline which combines Astronomy, computer science and statistics: the *Astroinformatics* (Djorgovski et al. 2006; Borne et al. 2009). Within this framework, Machine Learning (ML) and Deep Learning (DL) paradigms embed the intrinsic data-driven learning capability to explore huge amounts of multi-dimensional data by searching for hidden correlations within the data parameter space.

Such multidisciplinary paradigms have guided the present thesis, whose main scientific goals are: the development of a new method to select cluster members and to identify galaxy-galaxy strong-lenses in galaxy clusters, using only image-based classification algorithms, thereby avoiding the time-consuming photometry extraction, particularly challenging in galaxy clusters. These classification tasks have been addressed by implementing and testing a plethora of deep learning architectures. Given the restrict number of known strong-lenses in galaxy clusters, we opted for an approach that combines the need to simulate large sets of training images with the complexity of high-resolution HST imaging of clusters. Realistic strong-lensing events were obtained utilising high-precision lens models developed by our research group for a sample of eight clusters from the CLASH and HFF surveys.

Furthermore, in this work, we also developed a fast, automatic, GPU-optimised cross-correlation tool, able to process the whole MUSE data cube ($\sim 90\,000$ spectra) in a few tens of seconds, a critical feature, which will become essential in the next future, with the upcoming data-intensive surveys. Even if this is not a pure deep learning method, it represents a method with which extract (i.e. to *mine*) information from astronomical data, and, more important, it represents a crucial ingredient in order to build spectroscipically confirmed dataset with which neural networks can be trained and tested. Indeed, it can complement the observations by providing cluster galaxies membership, measuring the lens and source redshift in lensing events, extracting kinematical information. The tool has been used to estimate galaxy velocity maps and measure galaxy redshift, with $z \in (0, 7)$. Although the tool is under implementation, our first results indicate that the redshift for the 70% of the processed spectra can be correctly recovered and velocity maps of background lensed spiral galaxies can be accurately reconstructed (RMS $\sim 3\,\mathrm{km\,s^{-1}}$) in a fully automated fashion.

The thesis is structured as follows. Chapter 2 introduces the physical and mathematical concepts necessary for the development of this work. It is split into two parts: (*i*) a summary of the physical proprieties of galaxies and the models used to describe them; (*ii*) an illustration of the gravitational lensing formalism, by underlining the mathematical instruments exploited to build the strong-lensing simulations. Chapter 3 presents an overview of the machine learning and deep learning techniques, by also emphasising their usage in the astronomical context, together with a systematic description of the implemented architectures. Chapter 4 outlines the imaging and spectroscopic data used in this work, by summarising the surveys which provided these data. In Chapter 5, we show the networks capabilities to identify cluster members using images, by varying the experiment configurations in order to analyse model dependence on member redshift, train and test sample sizes, source magnitude and colours. We also compare our results with other catalogue-based approaches and describe the process to identify new members by complementing the spectroscopic sets. In Chapter 6, we illustrate the methodology with

which we simulated thousands of galaxy-galaxy strong-lens examples handling the deflection angle maps provided by cluster lens models, we discuss the networks classification results by considering false negative and false positive distributions, establishing networks limits. We also test network capabilities to identify real strong-lenses by processing an ensemble of known events and hundreds of spectroscopic and candidate members. Chapter 7 illustrates the GPU-based cross-correlation tool, with a computing time benchmark, a validation phase involving simulated galaxy spectra. We measure galaxy velocity maps and automatically estimate the redshifts for an ensemble of galaxies (in a wide z range $0-7$). Finally, we draw our conclusion in Chapter 8.

Throughout the thesis, we adopt a flat ΛCDM cosmology model with Ω_M=0.3, Ω_Λ= 0.7, and $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. All of the astronomical images are oriented with north to the top and east to the left. Unless otherwise specified, magnitudes are in the AB system.

Chapter 2

Galaxy populations and Gravitational Lensing

In this chapter, I outline the physical and mathematical concepts necessary for the development of this work. It is structured into two parts: (*i*) a description of the physical properties of galaxies and their population in galaxy clusters (Sec. 2.1); (*ii*) a summary of the gravitational lensing formalism, focused on the key ingredient used for the strong-lensing simulation (Sec. 2.2).

2.1 Galaxy description

In this section, we describe the main properties of galaxies and the galaxy population in clusters. Sec. 2.1.3 is particularly important, as we illustrate the galaxy surface brightness by describing the equations used to simulate galaxy-galaxy strong lenses by injecting a source behind a lens cluster galaxy (see Chap. 6). In Sec. 2.1.4, we list the main galaxy mass density profiles used in the cluster lens models and for the source-lens plane mapping when simulating strong-lenses (see Sec. 6.2). Similar mass density profiles were used to simulate galaxy velocity maps thus validating the cross-correlation tool (see Sec. 7.3.1).

2.1.1 Galaxy morphological classification

Galaxies occur in different shapes and sizes: some have smooth profiles with elliptical isophotes, others have spiral arms with an elliptical-like bulge, and still others have irregular or peculiar morphologies. Before the advent of modern instruments, these objects appeared as blurred diffuse sources and were listed in astronomical catalogues (e.g. *Messier* catalogue and *New General Catalogue*) as *nebulae* (Schneider 2006). Their existence was established only in the 1920s, a decade opened with the *Great Debate* between Harlow Shapley and Heber Curtis on the nature of the nebulae. Such question was solved in 1925 by Edwin Hubble, who exploited the discovery of Cepheids in the Andromeda "nebula" to derive a distance of $\sim$ 300 kpc and, even if this measurement is $\sim$ 3 times smaller than the distance estimated today, it provided a clear evidence of the extra-galactic nature of Andromeda and marked the beginning of extra-galactic astronomy (Mo et al. 2010).

Based on morphological features, Edwin Hubble set out galaxies in a sequence (Hubble 1926, 1936), that, with later additions and modifications, is still used today (see Fig. 2.1). Hubble recognised three main galaxy types: *ellipticals*, *lenticulars* and *spirals*, classifying as *irregulars* galaxies that would not fit in any of the other classes. Based on Hubble sequence, galaxies were also classified as "early-type" and "late-type", describing the galaxy life cycle progression from ellipticals to spirals, and, although this hypothesis has now been discarded, the terms are still

in use, even if they lost their original meaning. Nowadays galaxies are classified according to other properties (e.g., spectral parameters as emission or absorption lines, spectral distribution, integrated colour, fraction of gas content, level of star formation), which correlate with Hubble's morphological classification. Here we restrict the galaxy description to the main classes identified by Hubble (Longair 2008; Mo et al. 2010):

- Elliptical galaxies (E): these galaxies show no structural features, they are characterised by a smooth, elliptical surface brightness and have red photometric colour. Their absolute magnitude ranges from the brightest known galaxies, having $M_B \sim -24$, to the fainter dwarf ellipticals (dE), with $M_B \sim -18$. Ellipticals are divided into eight sub-classes, corresponding to their ellipticities $\epsilon = 1 - b/a$, where a and b are the semi-major and semi-minor isophotal axis; so, the common notation En indicates the isophote shape, with $n = \mathrm{int}(10 \cdot \epsilon)$, i.e. the closest integer to the value $10 \cdot \epsilon$. Given the lack of hot, young stars, elliptical spectra are characterised by the so-called 4000Å break (see Fig. 2.2): most of the light emerges at wavelengths > 4000Å with the typical K star absorption lines; while, in the bluer part, the spectrum shows strong H and K absorption lines of calcium.

- Spiral galaxies (S-SB): they consist of a disk with spiral arm structure (often with a bar) surrounding a central bulge. They are divided into two sub-classes: normal spiral (S) and barred spiral (SB). Their arms are characterised by clumps of bright O and B stars, HII regions, molecular clouds and dust absorption. Spirals are further classified as Sa, Sb, Sc, according to three criteria: (*i*) the openness of the winding arms, (*ii*) the resolution degree of the arms and (*iii*) the size of the spheroidal components relative to the disk. In contrast to ellipticals, they also emit at blue and ultraviolet wavelengths (see Fig. 2.2), due to the presence of young, hot stars, whose light heat and ionise interstellar medium (ISM), resulting into strong emission lines.

- Lenticulars (S0-SB0): they appear as transitional stage between ellipticals and spirals. Lenticular light disribution can be decomposed into a central bulge and an extensive disk without spiral arms or HII regions. Depending on whether or not they show a bar, these galaxies are divided into S0 or SB0.

- Irregulars (Irr): they are galaxies with only weak (Irr I) or no (Irr II) regular structures. They have neither a dominating bulge nor a rotationally symmetric disk. They appear as patchy systems with few HII rgions. They are arranged at the end of the spiral sequence (in the Sd galaxies). Many of these irregulars are similar to the Milky Way galaxy satellites, the Magellanic Clouds, which became the prototypes these Sm (or SBm) galaxies, in which the spiral structure is reduced to a single stubby arm. Irregulars without spiral structure are called Im galaxies.

Besides the morphological classification, galaxies can be arranged as a function of their spectral distribution (e.g., Fig. 2.2). In particular, the galaxy colour distribution (e.g. F606−F814, or g−r) is bimodal, with a narrow red peak and and broader blue distribution (Bell et al. 2004). Such bimodality is even more evident by arranging galaxies on a colour-magnitude diagram (as example of this bimodality see the colour panel and colour-magnitude diagram in Fig. 5.2), where it is possible to identify: (*i*) a nearly horizontal sequence, which is called "red-sequence", composed by non-star-forming or passive early-type, gas and dust poor galaxies; and (*ii*) a "blue cloud" consisting of star-forming, gas and dust rich, late-type galaxies (Schneider 2006; Mo et al. 2010). Clearly, such separation is not sharp: based on spectral analysis, Dressler & Gunn (1983) found evidence of an intermediate galaxy class, the so-called E+A galaxies, whose spectra are characterised by strong Balmer absorption lines, due to a conspicuous A-star population, but no

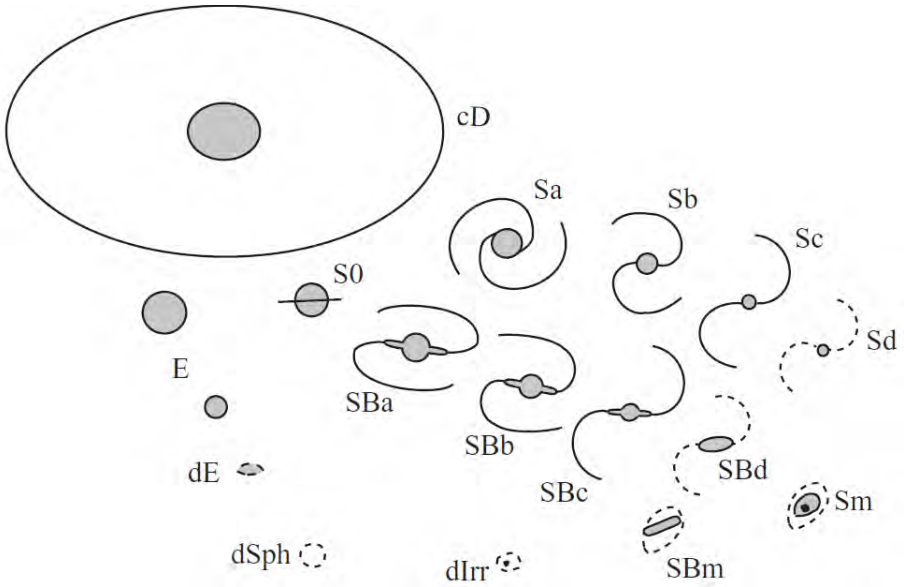


Figure 2.1: Morphological galaxy classification: a modified form of Hubble's diagram. Image from Sparke & Gallagher (2007).

significant OII or Hα emission lines. This suggests a stellar formation process which quenched in the last 1-2 Gyrs (Mo et al. 2010). Even if this transition is not completely understood, galaxy-galaxy mergers seem to trigger the starburst mechanism which leads to the E+A stage (Yang et al. 2004; Bekki et al. 2005).

2.1.2 Galaxy population in clusters

As early as the 1930s, it was realised that the galaxy morphological mixture depends on the environment, so, galaxy clusters, characterised by dense environments, host a large fraction of early-type galaxies (e.g., the analysis of "clusters of nebulae" carried out by Hubble & Humason 1931). Evidences of a morphology-density correlation has been found by Dressler (1980, 1984), which showed that the fraction of spiral galaxies decreases from $\sim 60\%$ in the lowest density regions to less than $\sim 10\%$ in the highest density regions, while elliptical and lenticular fraction reveals the opposite trend: the fraction of E + S0 galaxies is $\sim 80\%$ in regular clusters compared to $\sim 30\%$ in the field. More recently, it has been found that galaxies in denser environments are more massive, redder, more concentrated, less gas-rich and with lower specific star-formation rates (e.g., Kauffmann et al. 2004; Weinmann et al. 2006; Renzini 2006). This dichotomy suggests that galaxies undergo morphological transformation in dense environments: as a galaxy falls into the gravitational well, its gas component experiences a ram pressure (Gunn & Gott 1972) exerted by the intra-cluster medium (ICM), which strips the galactic gas, and, since stars and dark matter are not affected, the gas is left behind while the galaxy fall in the cluster (Sparke & Gallagher 2007; Treu et al. 2003); this mechanism, together with the thermal evaporation of the galactic inter-stellar medium due to the hotter ICM (Cowie & Songaila 1977), the turbulent and viscous stripping of the ISM (Nulsen 1982), results into a quenching of star formation. However, such in-falling galaxy interacts also with the cluster gravitational potential: tidal compression of ICM could even led to an increase of the star formation rate, while tidal truncation of the galaxy outskirt can change the galaxy structure or contribute to star formation rate decreasing (Treu et al. 2003).

The center of almost all of galaxy clusters is dominated by the presence of the Brightest Cluster Galaxy (BCG). These massive and luminous galaxies are known as cD galaxies (i.e.

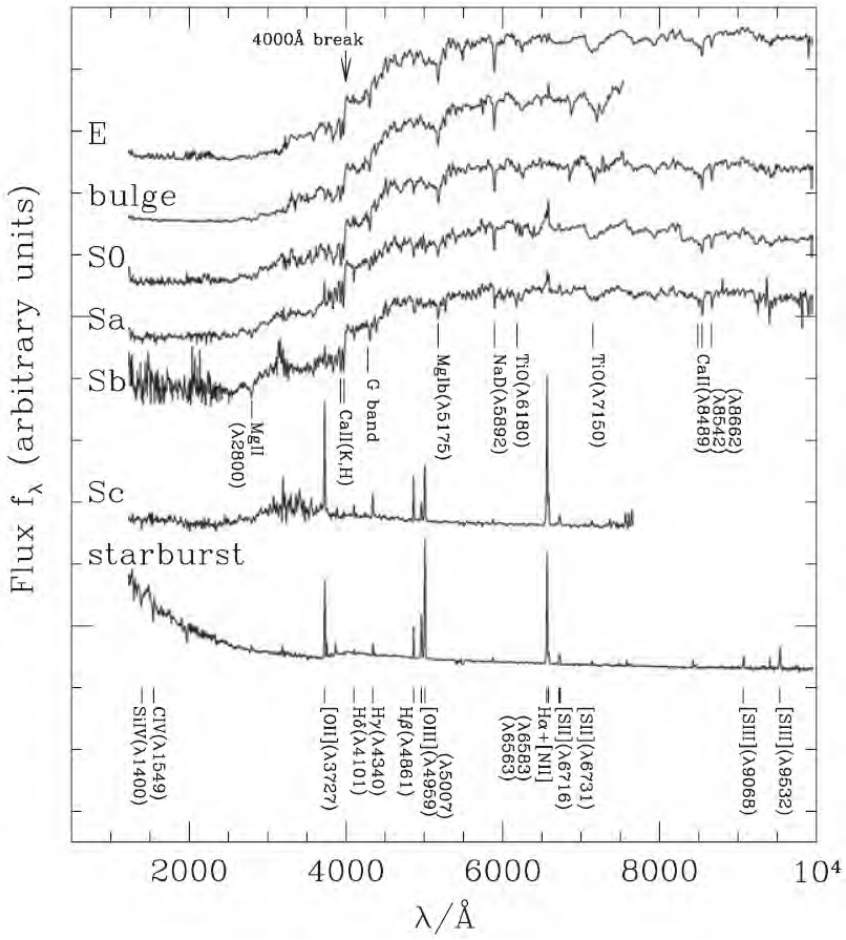


Figure 2.2: Spectra of different types of galaxies from the ultraviolet to the near-infrared. From ellipticals to late-type galaxies, the blue continuum and emission lines became stronger. The panel also shows the main absorption and emission lines. Image from Mo et al. (2010).

central diffuse). Even if they seem giant ellipticals, BSGs have several features which distinguish them from early-type galaxies (Kormendy 1982). They are the most massive known galaxies ($\gtrsim 10^{12} M_\odot$) and their light covers up to $\sim 30\%$ of the entire visible light of a rich cluster (Mo et al. 2010), but their most distinctive characteristic is the the extraordinary extended and diffuse stellar outer envelope (Longair 2008). cD galaxies are found only in regions where the galaxy density exceeds $1h^{-3}\mathrm{Mpc}^{-3}$, compared with an average galaxy density of $10^{-2}h^{-3}\mathrm{Mpc}^{-3}$ (Dressler 1984).

Based on the studied carried out by Oemler (1974), galaxy clusters are distinguished into three main types according to their galaxy content: (*i*) *cD clusters*, with a dominant central cD galaxy and an high fraction of E and S0 galaxies; (*ii*) *spiral-rich clusters* where spirals are about the 50% of the galaxy population, not too dissimilar from that of the field; and (*iii*) *spiral-poor clusters* with no dominant cD galaxy and an higher fraction of S0. Thus, regular, cD clusters are systems which had time to achieve the dynamical equilibrium, whereas for the other systems this relaxation process is still ongoing.

2.1.3 Galaxy surface brightness

A convenient and wide-used way of describing the light distribution of elliptical and spiral galaxies is the formulation proposed by Sérsic (1963, 1968), as a generalisation of de Vaucouleurs $R^{1/4}$ law (de Vaucouleurs 1948), based on which the one-dimensional surface brightness is a function of the isophotal semi-major axis length R, also known as $R^{1/n}$ profile (Mo et al. 2010;

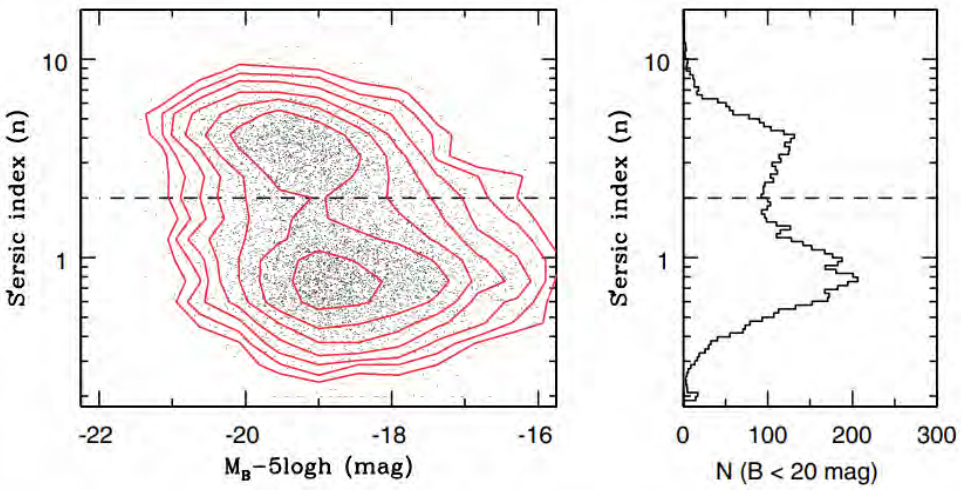


Figure 2.3: *Left panel*: The observed logarithm of the Sérsic index (n) as a function of the absolute magnitude B (together with 0.2−dex contours). *Right panel*: the Sérsic index logarithmic distribution. The dotted line indicates a separability threshold between bulge and disc dominated galaxies. Image from Driver et al. (2006).

Longair 2008):

$$I(R) = I_e \exp\left\{ -\beta_n \left[\left(\frac{R}{R_e} \right)^{1/n} - 1 \right] \right\} \tag{2.1}$$

where R_e is the effective radius enclosing half of the total light emitted, I_e is the surface brightness within R_e, n is the so-called *Sérsic index* whose value determines the degree of concentration of the profile and β_n is a normalisation constant which ensure that the total light sums to the total luminosity emitted by the galaxy.

This formalism can be used to distinguish between disc and bulge dominated galaxies. It can been shown that a galaxy sample split into two populations (e.g., the work conducted by Driver et al. 2006 on the Millennium Galaxy Catalogue, Liske et al. 2003): one centred on $n = 4$, corresponding to ellipticals, and the other centred on $n = 1$, corresponding to disc-dominated galaxies; the distinction occurs at $n \sim 2$ (see Fig. 2.3).

According to Eq. 2.1, the luminosity within a given radius R is given by integrating:

$$L(R) = \int_0^R 2\pi R' I(R') dR' = 2\pi \frac{n \exp(\beta_n)}{\beta_n^{2n}} I_e R_e^2 \, \gamma\left(2n, \beta_n \left(\frac{R}{R_e} \right)^{1/n} \right) \tag{2.2}$$

where γ is the incomplete gamma function. Thus, the total luminosity predicted by the profile is:

$$L_{\rm tot} = \lim_{R \to \infty} L(R) = 2\pi \frac{n \exp(\beta_n)}{\beta_n^{2n}} I_e R_e^2 \, \Gamma(2n) \tag{2.3}$$

where Γ is the gamma function. These two last equations can be used to obtain a relation between b_n and n, indeed, by definition of effective radius, it results $L(R_e) = L_{\rm tot}/2$, by replacing R_e in Eq. 2.2 and comparing this equation with Eq. 2.3, it is possible to write:

$$\Gamma(2n) - 2\gamma(2n, \beta_n) = 0 \tag{2.4}$$

which is a non-linear equation that can only be solved numerically. Analytical expressions which approximate the value of β_n have been found by expanding gamma functions (e.g., Prugniel & Simien 1997; Ciotti & Bertin 1999). In this work, since we are interested to simulate galaxy with $n \in [1, 2]$, we exploited the following expression (Capaccioli 1989):

$$\beta_n \sim 1.9992n - 0.3271 \quad \text{for} \quad n \in (0.5, 10) \tag{2.5}$$

which well approximates β_n for Sérsic index $n \in [1, 5)$ with errors $< 10^{-2}$ (e.g., see Fig. 2 in Graham 2001).

In order to generalise Eq. 2.1, i.e. to include the ellipticity of the isophotes (parametrised with the axis ratio, q), their orientation (according to the position angle, φ), for an arbitrary source position $y_s = (y_{s_1}, y_{s_2})$, given a rigid grid of points $y = (y_1, y_2)$ in which evaluate the surface brightness $I(R)$, the semi-major axis length R is computed with:

$$R_1 = (y_1 - y_{s_1}) \cos\varphi + (y_2 - y_{s_2}) \sin\varphi \tag{2.6}$$

$$R_2 = -(y_1 - y_{s_1}) \sin\varphi + (y_2 - y_{s_2}) \cos\varphi \tag{2.7}$$

$$R = \sqrt{(R_1/q)^2 + R_2^2} \tag{2.8}$$

2.1.4 Galaxy mass density profiles

In this section we introduce three widely used models which describe the mass density distribution of galaxies:

- Singular Isothermal Sphere (SIS), a one-parameter model whose mass density profile $\rho(r)$ is expressed as (Schneider 2006):

$$\rho_{\rm SIS}(r) = \frac{\sigma_0^2}{2\pi G r^2} \tag{2.9}$$

  where the Line of Sight (LoS) velocity dispersion σ_0 is the sole free parameter. This distribution suffers of two issues which make the model unphysical: it has a diverging density as $r \to 0$ and an infinite total mass ($m(r) \propto r$) for $r \to \infty$.

- Pseudo Isothermal Elliptical Mass Distribution (PIEMID, Kassiola & Kovner 1993): this model introduces the core radius r_c as additional parameter. The mass density profile, in its circular form, is given by:

$$\rho_{\rm PIEMD}(r) = \frac{\sigma_0^2}{2\pi G} \frac{1}{r^2 + r_c^2} \tag{2.10}$$

  So, this profile is characterised by a finite central density $\rho_{\rm PIEMD}(r = 0) = \frac{\sigma_0^2}{2\pi G r_c^2}$, with a flat core since $d\rho_{\rm PIEMD}(r)/dr \,\big|_{r=0} = 0$; while for $r \gg r_c$ the PIEMD model behaves like a SIS profile, with an unbound total mass.

- dual Pseudo Isothermal Ellipctical mass distribution (dPIE, Limousin et al. 2005; Elíasdóttir et al. 2007), which introduces an extra parameter: the truncation radius r_{cut}. The dPIE mass density profile is expressed as:

$$\rho_{\rm dPIE}(r) = \frac{\rho_0}{\left(1 + \frac{r^2}{r_{core}^2}\right)\left(1 + \frac{r^2}{r_{cut}^2}\right)} \quad \text{with} \quad r_{cut} > r_{core} \quad \text{and} \quad \rho_0 = \frac{\sigma_0^2}{2\pi G} \frac{r_{cut} + r_{core}}{r_{core}^2 r_{cut}} \tag{2.11}$$

  The dPIE profile has a flat core with central density ρ_0. For $r \in (r_{core}, r_{cut})$ the profile behaves like a SIS model ($\rho(r) \propto r^{-2}$), while for $r \gg r_{cut}$ it decreases as $\rho(r) \propto r^{-4}$.

2.2 Strong Gravitational Lensing

The idea that gravity could bend the light ray is dated back to the eighteenth century, formulated by Isaac Newton according to his corpuscular theory of light (Newton 1704). However, accurate predictions of this deflection were made only after the publication of Albert Einstein's theory of general relativity (Einstein 1916), with Arthur Eddington's expedition to Principe Island, who exploited a solar eclipse to measure the deflection angle for dozens of stars in the Hyades open cluster. Results were later published by Dyson et al. (1920), causing, inter alia, a remarkable prominence in the press of the time. Actually, Eddington's expedition provided the first evidence in support of Einstein's theory.

According to the General Relativity, the bending of the light ray is caused by the curvature of the space-time in a region due to the presence of a massive object. This light ray deflection caused by a mass distributions along the LoS (Line of sight) is called gravitational lensing, whereas the masses, determining the space-time deformation, are named gravitational lenses.

In this chapter I summarise the gravitational lensing fundamentals: Sect. 2.2.1 is dedicated to the lensing refraction index and the deflection angle, the lens equation is derived in Sect. 2.2.2, while the lensing potential proprieties are described in Sect. 2.2.3, the first and second order approximation of the lens equation are shown in Sect. 2.2.4, finally, in Sect. 2.3 the cluster lens models are briefly described.

2.2.1 Lensing refraction index and deflection angle

The light ray path can be described by Fermat's principle, which states that the photons travel along trajectories of stationary optical length respect to variation of the path, i.e.the light path is the one that is covered in the least time, it follows the so-called geodesic; formally:

$$\delta \int_A^B n(\mathbf{S}(l))dl = 0 \tag{2.12}$$

where n is the index of refraction, $\mathbf{S}(l)$ is the light path, A and B are the starting and ending point. To find the index of refraction, we assume the "weak-field" first-order approximation, i.e. the gravitation potential ϕ is much smaller that the squared speed of light, c^2, $\phi/c^2 \ll 1$, which is a valid assumption for most of the lensing phenomena.

In general, the space-time infinitesimal interval (a.k.a line element), ds^2, i.e. the infinitesimal distance between neighboring points, can be written as:

$$ds^2 = g_{\mu\nu} dx^\mu x^\nu \tag{2.13}$$

where $g_{\mu\nu}$ are the components of the metric tensor g, while dx^μ and dx^ν are the infinitesimal displacements along the four space-time directions. Within the weak-field approximation, the metric $g_{\mu\nu}$ can be written as (Carroll et al. 2004):

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \quad \text{with} \quad |h_{\mu\nu}| \ll 1 \tag{2.14}$$

where $\eta_{\mu\nu}$ is the Minkowski metric (diag$[-1, 1, 1, 1]$) and $h_{\mu\nu}$ is a slight perturbation acting on a flat space-time. By also assuming a static gravitational field (i.e. characterised by a variation much more slowly than the speed of light), the perturbation element takes the diagonal form $h_{\mu\nu} = \text{diag}[\frac{-2\phi}{c^2}, \frac{-2\phi}{c^2}, \frac{-2\phi}{c^2}, \frac{-2\phi}{c^2}]$. By replacing this term in Eq. 2.14, the line element in Eq. 2.13 can be written as:

$$ds^2 = -\left(1 + \frac{2\phi}{c^2}\right)c^2 dt^2 + \left(1 - \frac{2\phi}{c^2}\right)d\mathbf{x}^2 \tag{2.15}$$

Solving the perturbed geodesic for a photon (i.e. with a null proper time, $ds^2 = 0$), immediately allows to derive the speed of light c', as measured by an external observer (Meneghetti 2019):

$$c' = \frac{|d\mathbf{x}|}{dt} = c\sqrt{\frac{1+\frac{2\phi}{c^2}}{1-\frac{2\phi}{c^2}}} \simeq c\left(1+\frac{2\phi}{c^2}\right) \tag{2.16}$$

where we have exploited the weak-field approximation $\frac{2\phi}{c^2} \ll 1$. Finally, from this equation, it is immediate to derive the refraction index:

$$n = \frac{c'}{c} = \frac{1}{\left(1+\frac{2\phi}{c^2}\right)} \simeq 1 - \frac{2\phi}{c^2} \tag{2.17}$$

Since $\phi < 0$, it always results in $n \geq 1$ and, thus, the light speed is lower than in vacuum.

Eq. 2.12 is a standard variational problem, which leads to the well known Euler equations. By introducing a curve parameter λ, such that $\frac{d\mathbf{S}}{d\lambda} \equiv \mathbf{e}$, where $\mathbf{e}$ is the unit tangent vector to the light path, it is possible to demonstrate (see, for example, Meneghetti 2019 or Congdon & Keeton 2018) that all the solution to the variational problem satisfy the Euler equation:

$$\frac{d}{d\lambda}(n\mathbf{e}) - \nabla n = 0 \tag{2.18}$$

by manipulating this latter equation:

$$\frac{d\mathbf{e}}{d\lambda} = \frac{1}{n}\Big[\nabla n - \mathbf{e}(\nabla n \cdot \mathbf{e})\Big] = \frac{1}{n}\nabla_\perp \ln n \sim -\frac{2}{c^2}\nabla_\perp \phi \tag{2.19}$$

where we have replaced the refraction index expression (Eq. 2.17) and approximated $\ln n \sim \frac{2\phi}{c^2}$. Thus, the deflection angle of the light ray, due to the gravitational potential, is given by the integral of $-\dot{\mathbf{e}} = -d\mathbf{e}/d\lambda$ along the light path:

$$\hat{\boldsymbol{\alpha}} = \frac{2}{c^2}\int_{\lambda_A}^{\lambda_B} \nabla_\perp \phi d\lambda \tag{2.20}$$

Since $\nabla_\perp \phi$ points away from the center, $\hat{\boldsymbol{\alpha}}$ has the same direction. By also considering a light path coincident with the z-axis and by assuming small deviation of photon trajectory, we can replace in Eq. 2.20 the integration variable λ with z. We also impose that the lens is localised in $z = 0$. Moreover, since most of the light deflection occurs within a region whose dimension is typically much smaller than the distances between observer and lens and between lens and source, the lens can be considered *thin*. With this *thin screen approximation* (Narayan & Bartelmann 1996) the integration limits in Eq. 2.20 can let to infinity, and the deflection angle can be expressed as:

$$\hat{\boldsymbol{\alpha}} = \frac{2}{c^2}\int_{-\infty}^{+\infty} \nabla_\perp \phi dz \tag{2.21}$$

The typical gravitation lensing system is shown in Fig. 2.4, where the lens, located at redshift z_L (i.e. at an angular diameter distance D_L), deflects the light coming from a background source placed at redshift z_S (i.e. at an angular diameter distance D_S). The optical axis is the line perpendicular to both source and lens plane. The distance of the source and the lens from the optical axis is $\boldsymbol{\eta} = \boldsymbol{\beta} D_S$ and $\boldsymbol{\xi} = \boldsymbol{\theta} D_L$, respectively.

Within the thin screen approximation, the lens mass distribution can be projected along the line of sight, characterised by a surface mass density:

$$\Sigma(\boldsymbol{\xi}) = \int \rho(\boldsymbol{\xi}, z) dz \tag{2.22}$$

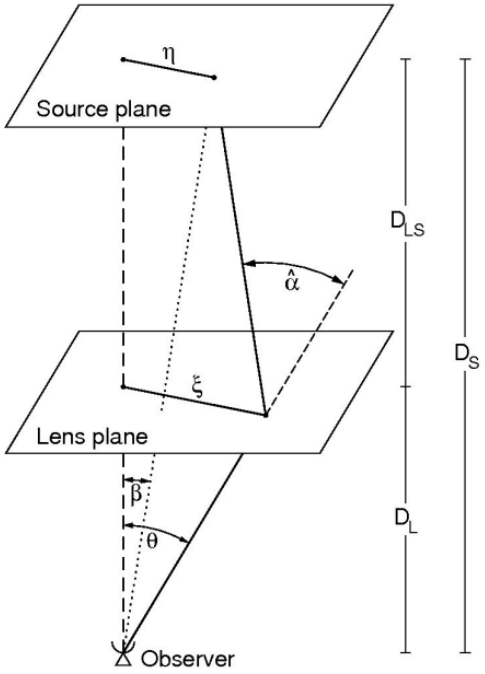


Figure 2.4: Sketch of a typical gravitational lensing system. D_S and D_L are the angular diameter distances, respectively, of the lens and the source from the observer, while D_{LS} is the distance between the lens and the source plane. η and ξ are the 2D distances, respect to the optical axis, on the source and lens plane, respectively. Image from Bartelmann & Schneider (2000).

and the deflection angle can be expressed as the sum of the light ray bending due to all the mass element in the lens plane (Narayan & Bartelmann 1996):

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \frac{4G}{c^2} \int \frac{(\boldsymbol{\xi} - \boldsymbol{\xi}')\Sigma(\boldsymbol{\xi}')}{|\boldsymbol{\xi} - \boldsymbol{\xi}'|} d^2\xi' \tag{2.23}$$

In the case of circular symmetric lens, the light deflection can be reduced to a one-dimension problem: the deflection angle points toward the center of the symmetry with a modulus of:

$$\hat{\alpha}(\xi) = \frac{4GM(\xi)}{c^2 \xi} \tag{2.24}$$

where $M(\xi)$ is the mass enclosed within a radius of ξ: $M(\xi) = 2\pi \int_0^{\xi} \Sigma(\xi')\xi' d\xi'$.

2.2.2 Lens equation

From the lensing geometry shown in Fig. 2.4, in the small-angle approximation, $\hat{\boldsymbol{\alpha}}$, $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are related through the well-known *lens equation*, geometrically deduced from Fig. 2.4:

$$\boldsymbol{\theta} D_s = \boldsymbol{\beta} D_S + \hat{\boldsymbol{\alpha}} D_{LS} \tag{2.25}$$

where D_{LS} is the angular diameter distance between the lens and the source plane. By defining the reduced deflection angle as $\boldsymbol{\alpha}(\boldsymbol{\theta}) = \hat{\boldsymbol{\alpha}}(\boldsymbol{\theta}) D_{LS}/D_S$, lens equation can be rewritten as:

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\beta}) \tag{2.26}$$

Typically, lens equation is expressed in a dimensionless form, by introducing a length scale on the lens plane ξ_0 and the corresponding length scale on the source plane, $\eta_0 = \xi_0 D_S/D_L$. In

this way, it is possible to define two dimensionless vectors: $\boldsymbol{x} = \boldsymbol{\xi}/\boldsymbol{\xi_0}$, $\boldsymbol{y} = \boldsymbol{\eta}/\boldsymbol{\eta_0}$, so that the dimensionless deflection angle is written as: $\boldsymbol{\alpha}(\boldsymbol{x}) = \frac{D_L D_{LS}}{\xi_0 D_S}\hat{\boldsymbol{\alpha}}(\xi_0 \boldsymbol{x})$. By replacing these definition, Eq. 2.25 can be written as:

$$\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{\alpha}(\boldsymbol{x}) \tag{2.27}$$

In order to describe the lensing geometry, a quantity called *Einstein radius*, θ_E, is typically used, since it represent a natural angular-scale for gravitational lensing (Narayan & Bartelmann 1996), e.g. sources closer than θ_E are significantly magnified, or, the typically angular separation between multiple images is $2\theta_E$. It can be derived from the lens equation, by assuming circular symmetric lens with an arbitrary mass profile, i.e. by combining Eq. 2.26 with Eq. 2.24:

$$\beta(\theta) = \theta - \frac{D_{LS}}{D_S D_L}\frac{4GM(\theta)}{c^2\theta} \tag{2.28}$$

where the reduced deflection angle has been used, together with the relation $\xi = \theta D_L$. By imposing that the source lies on the optical axis ($\beta = 0$), the source is deformed into a ring, whose radius is the so-called Einstein Radius:

$$\theta_E = \left[\frac{4GM(\theta_E)}{c^2}\frac{D_{LS}}{D_L D_S}\right]^{1/2} \tag{2.29}$$

2.2.3 Lensing potential

Typically, the *lensing potential* is used to describe the light ray bending due to a gravitational lens, which is defined as the projection on the lens-plane of the Newtonian gravitational potential (Meneghetti 2019):

$$\hat{\psi}(\boldsymbol{\theta}) = \frac{D_{LS}}{D_L D_S}\frac{2}{c^2}\int_{-\infty}^{+\infty}\phi(D_L\boldsymbol{\theta}, z)dz \tag{2.30}$$

whose dimensionless counterpart is:

$$\psi(\boldsymbol{x}) = \frac{D_L^2}{\xi_0^2}\hat{\psi}(\boldsymbol{\theta}) \tag{2.31}$$

Given this definition, it is possible to show that the lensing potential fulfils these conditions (Meneghetti 2019):

- the gradient of the lensing potential is the reduced deflection angle:

$$\nabla_{\boldsymbol{\theta}}\hat{\psi}(\boldsymbol{\theta}) = \boldsymbol{\alpha}(\boldsymbol{\theta}) \tag{2.32}$$

- the *convergence*, κ, describes an isotropic magnification of the source (i.e. an isotropic focusing of the light rays), it is defined as a dimensionless surface density:

$$\kappa(\boldsymbol{\theta}) = \frac{\Sigma(\boldsymbol{\theta})}{\Sigma_{\rm cr}} \quad \text{with} \quad \Sigma_{cr} = \frac{c^2}{4\pi G}\frac{D_S}{D_L D_{LS}} \tag{2.33}$$

  where $\Sigma_{\rm cr}$ is the so-called critical surface density. The laplacian of the lensing potential is twice the convergence:

$$\Delta_{\boldsymbol{\theta}}\hat{\psi}(\boldsymbol{\theta}) = 2\kappa(\boldsymbol{\theta}) \tag{2.34}$$

- the *shear*, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$, is a pseudo-vector, which introduces an anisotropy, by causing a stretch of the source along a certain direction, its components are defined as:

$$\gamma_1(\boldsymbol{\theta}) = \frac{1}{2}\left(\frac{\partial^2}{\partial\theta_1^2} - \frac{\partial^2}{\partial\theta_2^2}\right)\hat{\psi}(\boldsymbol{\theta}) \tag{2.35}$$

$$\gamma_2(\boldsymbol{\theta}) = \frac{\partial^2\hat{\psi}(\boldsymbol{\theta})}{\partial\theta_1\partial\theta_2} = \frac{\partial^2\hat{\psi}(\boldsymbol{\theta})}{\partial\theta_2\partial\theta_1} \tag{2.36}$$

Thus, the deflection angle and the image distortion parameters (i.e. convergence and shear) can be inferred from the lensing potential.

2.2.4 First and second order lens mapping

This section is dedicated to the description of the first and second order lens equation approximation. The section is extracted from Meneghetti (2019).

In theory, the distorted image of a source due to a gravitational lens can be determine by solving the lens equation taking into account a large number of light rays. However, if the source is much smaller than the angular dimension on which the physical properties of the lens change, the relation between the source and image can locally be linearised, i.e. the image distortion can be described by a Jacobian matrix. Thus, let's assume that this condition is satisfied and let's consider a point on lens plane at $\boldsymbol{\theta}_0$, where the deflection angle is $\boldsymbol{\alpha}_0$, corresponding to a point on the source plane at $\boldsymbol{\beta}_0 = \boldsymbol{\theta}_0 - \boldsymbol{\alpha}_0$; for a infinitesimally near point on the lens plane located at $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + d\boldsymbol{\theta}$, the deflection angle is $\boldsymbol{\alpha} \simeq \boldsymbol{\alpha}_0 + \frac{d\boldsymbol{\alpha}}{d\boldsymbol{\theta}}d\boldsymbol{\theta}$, which corresponds to a point on the source plane located at $\boldsymbol{\beta} = \boldsymbol{\beta}_0 + d\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}$. Thus, the vector $\boldsymbol{\beta} - \boldsymbol{\beta}_0$ is mapped through:

$$\boldsymbol{\beta} - \boldsymbol{\beta}_0 = \left(I - \frac{d\boldsymbol{\alpha}}{d\boldsymbol{\theta}}\right)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \tag{2.37}$$

So, the Jacobian matrix, A, describing the image deformation can be expressed as:

$$A \equiv \frac{\partial\boldsymbol{\beta}}{\partial\boldsymbol{\theta}} = \left(\delta_{ij} - \frac{\partial\alpha_i(\boldsymbol{\theta})}{\partial\theta_j}\right) = \left(\delta_{ij} - \frac{\partial^2\hat{\psi}(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right) \tag{2.38}$$

where the indices i and j, $\in (1, 2)$, label the components of $\boldsymbol{\theta}$ or $\boldsymbol{\alpha}$ on the source plane. Eq. 2.38 can be rewritten in terms of convergence and shear:

$$A = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix} \tag{2.39}$$

This Jacobian matrix can be expressed as a combination of the isotropic and anisotropic part. In order to split these components, the shear tensor is written in a diagonal form, by applying a rotation:

$$\Gamma = \begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & -\gamma_1 \end{pmatrix} = \gamma \begin{pmatrix} \cos 2\varphi & \sin 2\varphi \\ \sin 2\varphi & -\cos 2\varphi \end{pmatrix} \tag{2.40}$$

where $\pm\gamma = \pm\sqrt{\gamma_1^2 + \gamma_2^2}$ are the eigenvalues of the shear tensor and φ is the direction of the eigenvectors with respect to the axis θ_1, corresponging to the direction along which the source is stretched. Thus, the Jacobian matrix (Eq. 2.39) can be expressed as:

$$A = (1 - \kappa)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \gamma \begin{pmatrix} \cos 2\varphi & \sin 2\varphi \\ \sin 2\varphi & -\cos 2\varphi \end{pmatrix} \tag{2.41}$$

This equation specifies the contribution of the convergence and the shear to the source image distortion: convergence isotropically focuses the light ray producing an image rescaled by a constant factor $1/(1+\kappa)$; while, due to the shear, the source intrinsic shape is stretched along the direction φ and shrunk along the perpendicular one.

So, if the "small-source" condition is satisfied, by also imposing that the origin of the source and lens reference frames coincide with the source and image positions, i.e. $\boldsymbol{\beta}_0 = \boldsymbol{\theta}_0 = (0,0)$, the components of $\boldsymbol{\beta}$ can be expressed as a linear combination, i.e. a linear mapping between θ_i and β_i elements through the Jacobian matrix:

$$\beta_i \simeq \sum_j \frac{\partial \beta_i}{\partial \theta_j}\theta_j = \sum_j A_{ij}\theta_j \quad i,j \in (1,2) \tag{2.42}$$

Since gravitational lensing does not involve emission or absorption of photons, neither it changes the photon momenta, the surface brightness is conserved (Liouville theorem) despite the light bending. So, the changing in the solid angle subtended by the source, due to the lensing, implies a magnification (or a demagnification) of the flux. The inverse of the Jacobian matrix $M = A^{-1}$ is called the magnification tensor and its determinant is the so-called *magnification*:

$$\mu = \det M = \frac{1}{\det A} = \frac{1}{(1-\kappa)^2 - \gamma^2} \tag{2.43}$$

while the magnification tensor eigenvalues describe the amplification in the tangential and radial directions:

$$\mu_t = \frac{1}{1-\kappa-\gamma} \qquad \mu_r = \frac{1}{1-\kappa+\gamma} \tag{2.44}$$

the tangential and radial critical lines are the locus of points on the lens plane for which the magnification goes to infinity, i.e. by imposing, respectively, $1/\mu_t = 1-\kappa-\gamma = 0$ and $1/\mu_r = 1-\kappa+\gamma = 0$. An image near to the tangential critical line is tangentially distorted, while an image close to the radial critical line is stretched perpendicularly to this line.

Finally, the first order lens mapping (Eq. 2.42) can be extended by including the second order terms:

$$\begin{aligned} \beta_i &\simeq \sum_j \frac{\partial \beta_i}{\partial \theta_j}\theta_j + \frac{1}{2}\sum_j\sum_k \frac{\partial^2 \beta_i}{\partial \theta_j \partial \theta_k}\theta_j\theta_k \\ &= \sum_j A_{ij}\theta_j + \frac{1}{2}\sum_j\sum_k D_{ijk}\theta_j\theta_k \end{aligned} \tag{2.45}$$

where:

$$D_{ijk} = \frac{\partial^2 \beta_i}{\partial \theta_j \partial \theta_k} = \frac{\partial A_{ij}}{\partial \theta_k} \tag{2.46}$$

In conclusion, it should be underlined that given the lensing potential and the angular diameter distances D_L, D_S and D_{LS}, the lensing geometry is completely defined, since all the quantities which characterise the image distortion can be deduced from the potential.

2.3 Cluster lens models

This section briefly outlines the public software `LensTool` (Kneib et al. 1996; Jullo et al. 2007; Jullo & Kneib 2009) used by Bergamini et al. (2019, 2021b) and Caminha et al. (2019) to fit the cluster lens models, from which we extracted the lensing potential used to generate the galaxy-galaxy strong-lens dataset with which we trained the neural networks (see Chap. 6).

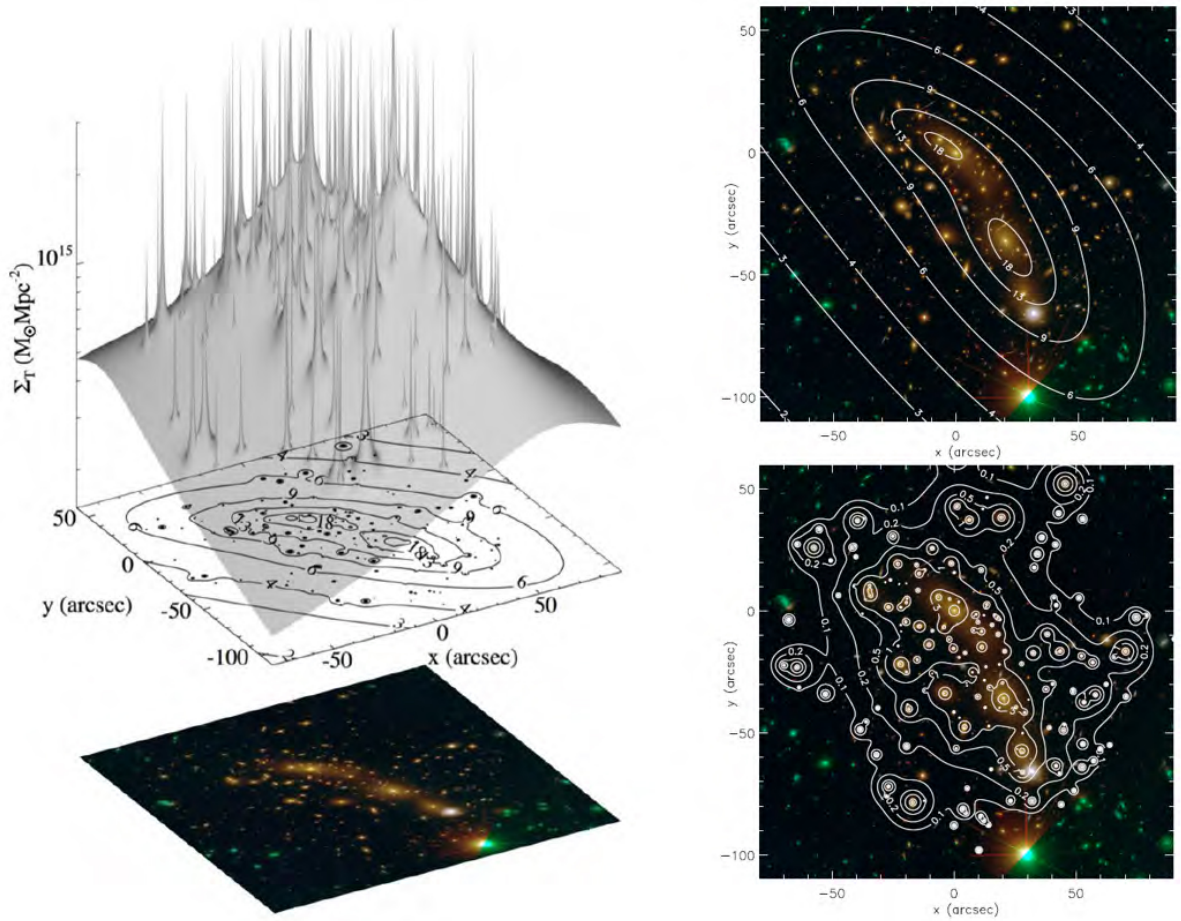


Figure 2.5: *Left panel*: total surface mass density, Σ_T, in the inner region of MACS J0416-2403. *Right panels*: decomposition of the total surface mass density into the two extended dark matter halos contributions (top right panel) and cluster galaxies components (bottom right panel). The contour level on the lens plane are in units of $10^{14} M_\odot\,\mathrm{Mpc}^{-2}$.

`LensTool` performs a parametric fit to constrain the total mass distribution of a galaxy cluster, by exploiting the observed multiple image positions as constraints for the lens model. In particular, the best-fit model is determined with a Bayesian Markov chain Monte Carlo approach, minimising the lens-plane likelihood:

$$\mathcal{L} = \prod_{i=1}^{N} \frac{1}{\prod_{j=1}^{n_i} \Delta x_{ij} \sqrt{2\pi}} e^{-\chi_i^2/2} \quad \text{with} \quad \chi_i^2 = \sum_{j=1}^{n_i} [x^j_{\rm obs} - x^j(\boldsymbol{W})]^2 \tag{2.47}$$

where N is the number of sources and n_i is the number of multiple images associated to the i-est source (a.k.a *family*), Δx_{ij} are the observed position uncertainties; $x^j_{\rm obs}$ is the position of the j-est image on the lens plane, while $x^j(\boldsymbol{W})$ is its corresponding predicted position, given the model parameter set $\boldsymbol{W}$.

Typically, the total mass distribution is parametrised by combining three components (Natarajan et al. 2017; Bonamigo et al. 2018; Bergamini et al. 2021a,b; Caminha et al. 2017a,b, 2019):

$$\phi_{\rm tot} = \sum_{i=1}^{N_h} \phi_i^{\rm halo} + \sum_{j=1}^{N_{\rm gal}} \phi_j^{\rm gal} + \sum_{k=1}^{N_{\rm sl}} \phi_k^{\rm shear+los} \tag{2.48}$$

where the first term accounts for the smooth cluster-scale halos (dark Matter content combined with the baryonic intra-cluster gas and intra-cluster light contributions); the second component

describes the clumpy sub-halos related to the cluster member galaxies (dark Matter and baryons); the third sum takes into account massive structures in the cluster outskirts and massive halos along the light-of-sight. So, in this equation, N_h, $N_{\rm gal}$ and $N_{\rm sl}$ are the number of cluster-scale halos, the number of cluster members and the shear plus line-of-sight contributions, respectively.

Since the sub-halos are described with a circular dPIE profiles (see Eq. 2.11), each cluster member introduces two additional free parameters (σ_0 and $r_{\rm cut}$), resulting in $2N_{\rm gal}$ extra free parameters, which can not be constrained. To solve this problem, the velocity dispersion and the truncation radius for each galaxy are measured by assuming the following scaling relation (Brainerd et al. 1996; Jullo et al. 2007):

$$\sigma_{0,j}^{\rm gal} = \sigma_0^{\rm ref}\left(\frac{L_j}{L_0}\right)^{\alpha} \tag{2.49}$$

$$r_{{\rm cut},j}^{\rm gal} = r_{\rm cut}^{\rm ref}\left(\frac{L_j}{L_0}\right)^{\beta} \tag{2.50}$$

where L_0 is a reference luminosity, L_j, $\sigma_{0,j}$ and $r_{{\rm cut},j}^{\rm gal}$ are the luminosity, the velocity dispersion and the truncation radius of the j-est cluster member; in this way the additional set of free parameters is composed by $\sigma_0^{\rm ref}$, $r_{\rm cut}^{\rm ref}$ (i.e. the reference velocity dispersion and the reference truncation radius), together with the scaling relation slopes α and β.

These lens modelling techniques produce deflection and magnification maps, as well as mass distribution maps which include the sub-halo and cluster halo components (see Fig. 2.5).

Chapter 3

Machine Learning approach

3.1 Data Mining in the realm of astronomy

As introduced in Chap. 1, it is evident that nowadays, and even more in the next feature, the volumes of data provided by upcoming dedicated surveys cannot be dealt with traditional methods. The modern multi-wavelength, multi-epoch, high-dimensionality, heterogeneous, Peta-scaled datasets require automatic, scalable and reliable methods that allow the exploration, mining and, in the final analysis, the inferring of knowledge. In this context Data Mining and Machine Learning (hereafter ML) techniques have become a necessity. This prompted the scientific community to develop general purpose, web-based and distributed tool-infrastructure able not only to face the data-intensive challenge, but also to offer opportunities of collaborations between astronomers and data-scientists (e.g. the Virtual Observatory, Djorgovski & Williams 2005; Graham et al. 2005; Pasian et al. 2012; or DAME, Djorgovski et al. 2012).

Machine Learning, sometimes treated as a branch of artificial intelligence, consists in the development of algorithms that allow an automatic adaptation of computers capabilities to solve the assigned problem based on empirical data: such techniques embed the intrinsic data-driven learning capability to explore huge amounts of multi-dimensional data by searching for hidden correlations within the data parameter space. A ML method can be thought as a "learner" which exploits examples of data to comprehend characteristics, discerns patterns, uncover anomalies and associations from the unknown underlying data distributions (Bishop 2006; Kamath 2009). Such collection of examples is typically called Knowledge Base (hereafter KB): a large ensemble of examples used to train the method and test its performance. The KB can be represented as a table in which each sample could be: a vector of features, a curve describing the variability over the time, a spectra or an image. In theory, the KB should be large enough to equally cover, qualitatively and quantitatively, all the parameter space; in practice, however, some regions of the parameter space will be undersampled, or even unparameterised, regions could be characterised by extremely different signal-to-noise ratio, resulting into a heterogeneous space; these are common problems in the context of astronomy, in which observation are affected by the instrumental characteristics and limits (e.g. photometric depth, spatial or spectral resolution, differences between throughput effectiveness of filter pass-bands or between the camera Field of View sizes), by also operating in a not-controlled environment. Such difficulties, if not properly contained, could lead to the building of uncompleted, rich of missing data and highly-heterogeneous KBs, which only partially represent the real distribution and the resulting trained model will not gain a sufficient degree of generability (Batista & Monard 2003; Marlin 2008; Parker 2010).

Despite these difficulties, the number of ML applications in the context of astronomy is

exploded in the last decade[1]. The ML application spectrum has rapidly expanded over time, for example: photometric redshift have been measured by training Multi-Layer Perceptron networks (Brescia et al. 2013; Cavuoti et al. 2012, 2015; Schmidt et al. 2020); regressors, such as Random Forest, Multi-Layer Perceptron and Support Vector Machine, have been applied to estimate the galaxy star formation rate (Delli Veneri et al. 2019) or to infer dark matter halo properties (von Marttens et al. 2021); ML-classifiers have photometrically identified astronomical sources, such as globular clusters or Quasar by disentangling from other objects, using Random Forest, Multi-Layer Perceptron, Principal Component Analysis and Neural-Gas models (Brescia et al. 2012, 2015; D'Abrusco et al. 2016; Nakoneczny et al. 2019; Angora et al. 2019); clustering models have been applied to various astronomical datasets: identification and characterisation of X-ray sources from spectra (Hojnacki et al. 2007), automatically segmenting and labelling of sources in galaxy clusters (Hocking et al. 2015), classification of asteroids and clustering of chemical species from Mars spectral images (Galluccio et al. 2008). Further astronomical examples can be found in Sec. 3.4 referred to the Deep Learning approach, in Sec. 5.1 referred to cluster member selection and in Sec. 6.1 referred to the galaxy-galaxy strong lenses identification.

3.2 Data Mining paradigms and functionalities

Nowadays a plethora of ML-based algorithms exist, that exploit different paradigms:

- *Supervised learning*: it is a type of algorithm that maps an input to an output, by inferring a function from a domain in $\mathbb{R}^n$ to a domain in $\mathbb{R}^m$, where $m < n$. The KB is composed by *pairs*, consisting of an input sample and a desired output vector. In astronomy, supervised methods have been deployed to address a plethora of problems, e.g. searching of quasar candidates (Abraham et al. 2012; Brescia et al. 2015), measuring the galaxy photometric redshifts (Cavuoti et al. 2012; Sadeh et al. 2016), star formation rate estimation (Delli Veneri et al. 2019), fitting galaxy surface brightness profile (Tuccillo et al. 2018).

- *Unsupervised learning*: it is the task of learning patterns from unlabelled data, by clustering samples assuming a certain metric. Typically, such kind of algorithms produce a tree-like structures across the parameter space. Unsupervised algorithms have found several applications in the realm of astronomy, for example: transient detection in light curves (Webb et al. 2020), source segmentation and labelling in galaxy clusters (Hocking et al. 2015), identification of globular clusters (D'Abrusco et al. 2016; Angora et al. 2019).

- *Self-supervised learning*: it is an approach with which unlabelled samples are encoded into a low-dimension space. It consists of an *encoder*, which maps data from a domain $\in \mathbb{R}^n$ into a latent domain $\in \mathbb{R}^m$, where $m \ll n$, and a *decoder* which maps the opposite function (i.e., from the latent space to the original domain). As examples in astronomy, these *autoencoders* have been used to: map interstellar dust (Thorne et al. 2021), recover the spectral energy distribution of galaxies (Frontera-Pons et al. 2017), emulate thermal Sunyaev-Zeldovich maps of galaxy clusters (Rothschild et al. 2021), deblend overlapping galaxies (Reiman & Göhre 2019).

- *Reinforcement learning*: it is a technique with which a model (in this context called *agent*) learns how it should act to maximise a certain goal. The agent learns to make a sequence of decision in order to achieve a goal in an uncertain, potentially complex environment.

[1] Between 2006 and 2011 fewer than 50 papers regarding ML application in astronomy were published per year, whereas in the 2018 the number is grown to 300 per year with a 50% increasing respect to the 2016 (Acquaviva 2020).

As examples, reinforcement learning methods have been applied to control adaptive optics for astronomical systems (Nousiainen et al. 2021) and to calibrate hyperparameters in data pipelines of radio telescope (Yatawatta & Avruch 2021).

Within these paradigms, several functionalities can be defined:

- *Classification*: it is the process with which samples are divided into m groups by exploiting the relation between the input and its label. A classifier learns a mapping between input space $X \in \mathbb{R}^n$ to a labelled set $Y \in \mathbb{R}^m$, where m is the number of classes.

- *Regression*: it is the procedure in which one or more floating values are fitted for each item, by searching a mapping between the input domain $\in \mathbb{R}^n$ to a domain $\in \mathbb{R}^m$, where $m < n$ is the number of values fitted for each sample. The regressor can either exploit a prior assumption on the data distribution, or, in the absence of such well-defined function, it can automatically search for a statistical correlation between the two domains.

- *Clustering*: it is a strategy based on the natural splitting of the dataset into groups, without any previous labelling, by performing a self-adapting mechanism. The method maps the input domain $X \in \mathbb{R}^n$ to a domain $W \in \mathbb{R}^n$, preserving the input topology and, in the same time, simplifying the dataset representation. The number of clusters could be set by the user or automatically determined by minimising a certain metric.

- *Dimensional reduction*: it is the reducing of the number of involved dimensions (i.e. the number of features composing the dataset). It could based on a posterior analysis of the feature informative contribution (i.e. a *feature selection* technique, Guyon & Elisseeff 2003) or on a data transformation which linearly, or non-linearly, projects the dataset into a space with fewer dimensions (Jolliffe 2011; Guyon & Elisseeff 2006).

3.3 Artificial Neural Networks

Inspired by biological behaviour, Artificial Neural Network (hereafter ANN) architecture is based on a collection of artificial neurons (called *perceptrons*, Minsky & Papert 1969; Rojas 1996; Hassabis et al. 2017), arranged in several layers, where each neuron takes as input the signal coming from neurons belonging to the previous layer; such as biological neurons, the variation of the synaptic connection sensibility (with respect to a certain input signal) is correlated to the learning mechanism (Hebb 1949). During the training, these connection sensibilities among layers (i.e. the weights) are iteratively adapted through a forward-backward mechanism. After the training, ANNs define a non-linear relation between the input and output spaces, which is encoded within the weight matrices. An illustrative representation of a perceptron in shown in Fig. 3.1: an artificial neuron performs a weighted sum of inputs ($\{x_i\}_{i=1}^n$), which is added to a bias term, describing the neuron's resting state. Thus, the output of a perceptron is:

$$y(\mathbf{x}; \mathbf{w}) = \sigma\left(\sum_{i=1}^{n} w_i x_i + b \right) \tag{3.1}$$

where the weights $\{w\}_{i=1}^n$ represent the sensibility of connections, and σ is the so-called *activation function* describing the neuron reaction to the incoming stimuli.

As stated before, these artificial neurons are typically organised in layers, within which there is no communication between neurons, but each layer receives the information from the previous layer, propagating it towards the next layer, as shown in Fig. 3.2. In such architectures, layers are nominally distinguished in: (i) input layer, (ii) hidden layer, (iii) output layer. Hidden layers

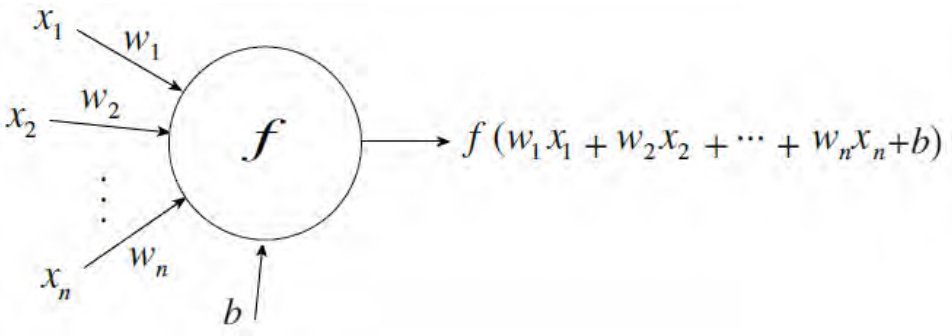


Figure 3.1: Schematic representation of an artificial neuron. Image from Rojas (1996).

transform input by combining features in order to extract an abstract representation of them. The output layer provides a supplementary transformation fulfilling the required task. Finally, a layer in which *all* neurons are connected with *all* neurons of the following layer are called *fully-connected* (or *dense*) layer. For example, Multi-Layer Perceptron (MLP, Hastie et al. 2001) is one the most typical and widely used feed-forward fully-connected network.

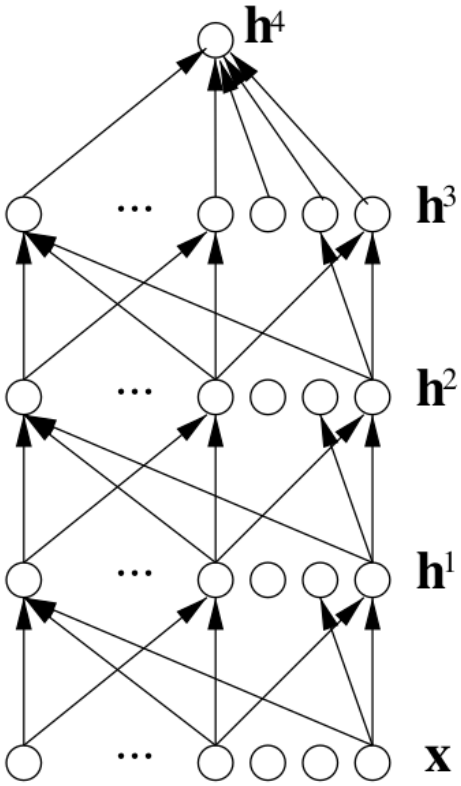


Figure 3.2: Architecture of a multi-layer network with three hidden layers. Hidden layers are denoted as $\mathbf{h}^m$, $m = 1, 2, 3$, $\mathbf{h}^4$ is the output layer and $\mathbf{x}$ is input layer. Image from Bengio (2009).

The output of a fully-connected layer can be expressed as:

$$\mathbf{h}^k = \sigma\left(\mathbf{b} + W^k \cdot \mathbf{h}^{k-1}\right) \tag{3.2}$$

where $\mathbf{b}$ is the vector of biases (composed by n elements, where n is number of neuron in the k-est layer) and W^k is the weight matrix related to the k-est layer, whose elements w_{ij}^k are the connections between the i-est neuron in the $(k-1)$-est layer and j-est neuron in the k-est layer.

Finally, concerning the activation function σ in Eq. 3.1 and 3.2, despite the varieties of such functions, all of them share a threshold behaviour implying that only intense signal will be

propagated. In the extreme case in which neuron transmits information through a step activation function, the network is composed only of excitatory and inhibitory neurons (Minsky & Papert 1969). In all the other cases, the neuron acts as a *threshold gate* capable of implementing logical functions of n arguments (Rojas 1996). A setup of the widely-used activation functions is shown in Tab. 3.1.

Activation function	Equation
Step[1,5]	$\sigma(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases}$
Rectified Linear Unit (ReLU)[1,5]	$\sigma(x) = max(0, x)$
Parametric ReLU (PReLU)[2,3]	$\sigma(x) = \begin{cases} \alpha x & x \leq 0 \\ x & x > 0 \end{cases}$
Exponential Linear Unit (ELU)[4,6]	$\sigma(x) = \begin{cases} \alpha(e^x - 1) & x \leq 0 \\ x & x > 0 \end{cases}$
Softplus[1,7,8]	$\sigma(x) = \ln(1 + e^x)$
Sigmoid[1,6,7,8]	$\sigma(x) = \frac{1}{1 + e^{-x}}$
Hyperbolic tangent[1,6]	$\sigma(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
Softmax[1,5,6]	$\sigma_i(\mathbf{x}) = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}} \quad i = 1, \ldots, M \quad \mathbf{x} = (x_1, \ldots, x_M)$

Table 3.1: List of activation functions commonly used in Neural Networks and their equation. [1]Szandała (2020); [2]He et al. (2015b); [3]Maas et al. (2013); [4]Clevert et al. (2015); [5]Bishop (2006); [6]Goodfellow et al. (2016); [7]Glorot et al. (2011); [8]Dugas et al. (2000).

With the exception of softmax, all other activation functions listed in Tab. 3.1 can be optionally used[2] to mimic a biological neuron behaviour by adding a non-linearity to the perceptron response. PReLU is a parametric version of the rectifier, the leakage coefficient (i.e. α) could be included in the learning parameter set (He et al. 2015b), whereas if it is maintained fixed during the training (typically with a value of ~ 0.01) the rectifier takes the name of Leaky ReLU (LeReLU). The presence of the leakage coefficient allows: (i) a small, non-zero gradient also when the unit is saturated and not active, (ii) a gain of the convergence with the increase of the units (Maas et al. 2013). Both ELU and Softplus are non-linear and continuous variants of the rectifier (Dugas et al. 2000). ELU forces to make the mean activations closer to zero, which speed up the learning (Clevert et al. 2015). Finally, softmax activation function (a.k.a. softargmax) is a logistic function which maps a vector of M floating values into M probabilities (M is the number of classes involeved in the classification task):

$$\sigma : \mathbf{x} \in \mathbb{R}^M \rightarrow \mathbf{p} \in [0, 1]^M, \text{with} \sum_{j=1}^{M} p_j = 1$$

[2]If no activation function is set, then the neural response is linear.

Respect to the other activations, softmax is only used for a classifier as activation function for all the neurons in the output layer, since its output can be interpreted as an estimator of $P(Y = i|\mathbf{x})$, where Y is the class associated with the input $\mathbf{x}$ and i is an integer number $i = 1, \ldots, M$ labelling the i-est class (Bengio 2009).

3.4 Deep Learning approach

Although multi-layer ANNs are universal approximators (Hornik et al. 1989), performance of ANNs is strongly dependent on the chosen ensemble of features with which data are represented (Bengio et al. 2012), this limits ML algorithms to process natural data in their raw form and it has forced the scientific community to design feature extractor pipelines that transform raw data into a suitable internal representation (Lecun et al. 2015). In order to expand the scope and the application fields of Machine Learning, it was necessary to make learning methods less dependent on feature engineering by developing algorithm able to identify and disentangle the underlying explanatory factors hidden in the observed environment (Schmidhuber 2014). Deep Learning (hereafter DL) algorithms allow a network to be fed with raw data and to automatically discover the representations needed to solve the problem. In fact, DL techniques are *representation-learning* methods (Bengio et al. 2012) composed by tens (sometimes hundreds) levels of representation, obtained by assembling non-linear modules that transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level. With the composition of enough such transformations, very complex functions can be learned, resulting into a network automatically able to extract meaningful features from raw data which become the input vector to any standard ML model. Even if DL methods had been used for decades (the first applications date back to 1989 with the earliest convolutional networks implemented by LeCun et al. 1989), the advent of modern Graphic Processing Units (GPUs) has made possible a real revolution (Oh & Jung 2004; Schmidhuber 2014), improving computing times by orders of magnitude (GPU computational costs are $\sim$ 700 times lower, Simard et al. 2005) and causing an unprecedented impact on our society by influencing an ever so broad spectrum of disciplines: from self-driving vehicles (Huval et al. 2015) to medical image analysis (Cireşan et al. 2013; Litjens et al. 2017), from image and object recognition[3] (Krizhevsky et al. 2012; Lu & Tang 2014) to natural language processing (Collobert & Weston 2008; Gonzalez-Dominguez et al. 2014; Wu et al. 2016), from financial applications (Kleanthous & Chatzis 2020) to AI gamer[4] (Campbell et al. 2002; Silver et al. 2016; Silver et al. 2018); DL algorithms have even bordered into the world of art[5] (Gatys et al. 2015; Smith & Leymarie 2017; Agüera y Arcas 2017). An interesting historical overview of the main stages in the DL developments can be found in Schmidhuber (2014).

Clearly, also the number of DL astronomical applications is exploded: over the last fifteen years more than 800 referred articles exploiting DL methods have been published[6]. For example:

[3] In 2014, for the first time, the human-level performance in face verification has been surpassed by DL algorithms (Lu & Tang 2014).

[4] In 2018 an AI gamer called *AlphaGo Zero* achieved superhuman performance in the game of Go, chess and shogi, convincingly defeated any world champion program, with a self-play training based on reinforcement learning (Silver et al. 2018).

[5] AIVA (Artificial Intelligence Virtual Artist, `https://www.aiva.ai/`) is an AI composer created in 2016 and specialised in classical and symphonic composition. It is the first AI composer recognised by a professional artists and editors association (SACEM). While *Edmond de Belamy* is AI painter based on generative adversarial network (Goodfellow et al. 2014a), in 2018 one of its artwork has been sold for $432'500 (`https://www.nytimes.com/2018/10/25/arts/design/ai-art-sold-christies.html`).

[6] This number has been computed by counting the number of referred papers including "Deep Learning" in the abstract in the NASA/ADS archive from 2005 to 2021.

deep neural networks have been used to estimate photometric redshift directly from Sloan Digital Sky Survey images (Hoyle 2016; Pasquet-Itam & Pasquet 2018), to classify radio galaxies from FIRST images (Aniyan & Thorat 2017; Lukic et al. 2018; Wu et al. 2019a), or to detect exoplanets from Kepler light curves (Pearson et al. 2018); deep learning methods were also able to retrieve the full set of parameters of one-component Sérsic models exploiting HST images reproducing GALFIT results $\sim$ 3000 faster (Tuccillo et al. 2017, 2018; Li et al. 2021a) and to predict cluster masses from Chandra X-ray images (Ho et al. 2019; Ntampaka et al. 2015, 2016, 2019); deep generative algorithms (Goodfellow et al. 2014a) can simulate galaxy images trained with HST 0.030″/pixel images (Lanusse et al. 2021), or deblend overlapping galaxies (Reiman & Göhre 2019); finally, region-based deep networks (Girshick et al. 2013) were able to automatically identify and deblend sources in multi-band DECam images (Reiman & Göhre 2019).

3.5 Convolutional Neural Networks

For the historical point of view, the first convolutional network was the *neocognitron* which is a self-organiser multi-layer ANN introduced by Fukushima (1980) exploiting convolutional and down-sampling layers, inspired by the physiologist work by Hubel & Wiesel (1959), which showed how the cat visual cortex is composed by retinal ganglion cells, which individually respond to patterns of light stimuli, defining a light *receptive field* (already observed by Kuffler 1953) with separate excitatory and inhibitory regions. Finally, the translation invariance introduced in the time delay neural network (Waibel et al. 1989) combined with the typical forward-backwards propagation led to the implementation of the current paradigm of Convolutional Neural Networks (CNNs, LeCun et al. 1989, 1990, 1998).

CNN represents one of the most widely-used supervised technique among the Deep Neural Networks, specialised for processing data characterised by a grid topology, e.g. images, that can be treated as a grid of pixels (Goodfellow et al. 2016), whose peculiarity is an ensemble of receptive fields which trigger the neuron activity. The receptive field is represented by a small matrix (called as kernel or filter), which connects two consecutive layers through a convolution operation. Similar to the adaptation mechanism imposed by supervised machine learning, the kernels are modified during the training. The idea behind CNN is a convolution-subsampling chain mechanism: deep networks are characterised by tens of layers (in some cases hundreds, as proposed by He et al. 2015a and Xie et al. 2016), where at each depth level, the convolution acts as a filter, emphasising (or suppressing) some properties; while the subsampling (often called pooling) makes sure that only essential information is moved towards the next layer. At each level, the processed input is transformed into an ensemble of feature maps, resulting into a hierarchically learning of high-level representations of data input: lower layers maintain a faithful representation of the image, while deeper level includes the extraction of abstract patterns capturing progressively larger deformations of the input (Mahendran & Vedaldi 2014). Remarkably, this behaviour emerges naturally in the learned network without any mechanism directly encouraging this extraction.

As previously said, CNNs are organised as a hierarchical series of layers, based on convolution and pooling operations. Convolution kernel is represented by a 4D matrix $\mathbf{W}$, where the element $W_{i,j,k,l}$ is the connection weight between the output unit i and the input unit j, with an offset of k rows and l columns. This kernel is convoluted with the input signal and adapted during the training. Given an input $\mathbf{X}$, whose element $X_{i,j,k}$ represents an observed data value of the

channel i at row j and column k, the neuron activity can be expressed as (Goodfellow 2010):

$$\begin{aligned} Z_{ijk} &= c(\mathbf{W}, \mathbf{X}, s)_{ijk} + b_i = \\ &= \sum_{lmn} X_{l,(j-1)\times s+m,(k-1)\times s+n} W_{ilmn} + b_l \qquad (3.3) \\ Z_{ijk} &\leftarrow p(\mathbf{Z}, d)_{ijk} \qquad (3.4) \\ Z_{ijk} &\leftarrow \sigma(\mathbf{Z}, \{a\}_q)_{ijk} \qquad (3.5) \end{aligned}$$

where $c(\mathbf{W}, \mathbf{X}, s)$ is the convolution operation between the input $\mathbf{X}$ and the kernel $\mathbf{W}$ with stride s; b is an addend that acts as bias; $p(\mathbf{Z}, d)$ is the pooling operation with down-sampling factor d; $\sigma(\mathbf{Z}, \{a\}_q)$ is the activation function characterised by the set of hyper-parameters $\{a\}_q$.

Unlike traditional artificial neural networks (e.g. Multi-Layer Perceptron), where all neurons of two consecutive layers are fully connected among them, the connection among neurons, performed with convolutional layers, is *sparse* and *shared*, i.e. the interaction between neurons belonging to different layers is limited to a small fraction and such interaction filter is shared for all neurons in a certain receptive field. This reduces the number of operations, the memory requirements and, thus, the computing time. Moreover the repeating of the same kernel over the input grid allows a little translation equivariance of the network.

The pooling function reduces the dimension, by replacing the network output at a certain location with a summary statistic of nearby outputs (Goodfellow et al. 2016). Typically down-sampling is performed by applying a max-pooling filter to (usually non-overlapping) subregions of the input (Zhou & Chellappa 1988), but other functions may also be used, e.g. Global Average pooling (Lin et al. 2013) or Spatial Pyramid pooling (He et al. 2015). Pooling reducing the number of parameters to learn and provides basic translation invariance to the internal representation of the network (Goodfellow et al. 2016).

3.6 Data flow: network training, validating and testing

A complete ML experiment is structured into three phases, each of them exploits an independent fraction of the KB. Typically, the KB is split into three not-overlapping sets: the training set (generally $60 - 80\%$ of the KB), validation set ($5 - 15\%$), test set ($10 - 30\%$).

During the training stage, the model parameters are iteratively adjusted according to a predefined *loss function* (a.k.a *cost function*): for supervised learning this function is a comparison between the output predicted by the model and "truth", for unsupervised learning this metric is related to network capability to map the input. The time required to process all the examples in the training set is called *epoch*. Typically, a large number of epochs is necessary to fit the model parameters allowing a convergence to the best possible configuration.

The validation set can be used for two main, mutually excluding, reasons: (i) *at the end of the training*, to chose the best hyper-parameter setup between different possible configurations, (ii) *during the training* (actually, at the end of each epoch), to update some hyper-parameters or to introduce reguralisation techniques (e.g. early-stopping criteria).

After the network has been trained and validated, the test set is used to measure the performance of the model, to understand if the network is able to apply the knowledge gained from the training on samples never processed before, achieving a good level of generalisation. For this reason the test set should cover the whole KB reproducing the same training parameter set. Network performance are measured through a set a metrics, listed and discussed in Sec. 3.6.6.

Although the canonical training-testing split (briefly described at the beginning of the section) is widely used within the ML context, it is correctly applied only when samples are uniformly distributed in the parameter space and when the KB size is so large to ensure that the randomly

extraction of testing examples equally cover the whole space reflecting the training distribution. As already stated, this cannot be typically achieved in the context of astronomy. The simple way with which is possible to get over this problem is the usage of a k-fold split, which is also exploited to estimate prediction error (Hastie et al. 2001).

In the next sections I show how the network learns from the training set through the definition of the loss function, the adaptation rule and the chose of an optimisier (Secs. 3.6.1 and 3.6.2); in Secs. 3.6.2 and 3.6.3 I introduce the overfitting and underfitting problems and how these tasks can be handle by modifying the loss function, take advantage of specific layers, by also exploiting the validation set, involving data augmentation techniques and the inclusion of adversarial examples. Finally, in Sec. 3.6.5 I describe the adopted k-fold strategy to split the KB into validation, test and training set, while in Sec. 3.6.6 I show the statistical estimators used to measure network performances on the test set.

3.6.1 Loss function and adaptation mechanism

For a supervised classification problem, the most common loss function is the *binary cross-entropy* (Goodfellow et al. 2016):

$$\mathcal{H}(y, \bar{y}) = -\frac{1}{N}\sum_{i=1}^{N} \bar{y}(\mathbf{x}_i) \cdot \log(y(\mathbf{x}_i)) + (1 - \bar{y}(\mathbf{x}_i)) \cdot \log(1 - y(\mathbf{x}_i)) \tag{3.6}$$

where $\bar{y}$ is the target and y is the output of the final layer; N is the number of extracted samples. Such cost function can be generalised for M classes:

$$\mathcal{H}(y, \bar{y}) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} \bar{y}_j(\mathbf{x}_i) \cdot \log(y_j(\mathbf{x}_i)) \tag{3.7}$$

From this latter equation is clear that cross-entropy is the negative log-likelihood divided by N conditionally independent samples. Some other loss functions are:

- Mean Squared Error (MSE):

$$MSE(y, \bar{y}) = \frac{1}{N}\sum_{i=1}^{N}(y(\mathbf{x}_i) - \bar{y}(\mathbf{x}_i))^2 \tag{3.8}$$

- Mean Absolute Error (MAE):

$$MAE(y, \bar{y}) = \frac{1}{N}\sum_{i=1}^{N}|y(\mathbf{x}_i) - \bar{y}(\mathbf{x}_i)| \tag{3.9}$$

- Kullback–Leibler divergence (D_{KL}):

$$\mathcal{D}_{KL}(y, \bar{y}) = \sum_{i=1}^{N} y(\mathbf{x}_i) \log\left(\frac{y(\mathbf{x}_i)}{\bar{y}(\mathbf{x}_i)}\right) \tag{3.10}$$

These functions can be indifferently used for regression or classification. Since KL divergence can be written as $\mathcal{D}_{KL}(y, \bar{y}) = \mathcal{H}(y, \bar{y}) - \mathcal{H}(y)$, it can be interpreted as a dissimilarity measure between the true and the predicted distributions (as the cross-entropy) by subtracting the *self* cross-entropy $\mathcal{H}(y)$ which could not be identically zero (Kullback & Leibler 1951).

Thus, during the training, samples extracted from the train set are propagated through the network, while weights and biases are adapted along with a backward flow in order to minimise the cost function. Backpropagation is the understanding of how changing the weights and biases in a network changes the cost function. Ultimately, this means computing the partial derivatives $\partial C/\partial w$ and $\partial C/\partial b$, i.e. the derivatives of the cost function respect to the weights and biases; such process can be schematised as follow[7] (Nielsen 2015):

1. Compute the adaptation for the output layer L:

$$\delta^L = \nabla_h C \odot \sigma'(z^L) \tag{3.11}$$

   where h is the output of the L-est layer, $\sigma'(z^L)$ is the derivatives of the neuron respond respect to the involved weights or biases, and $\odot$ denotes the element-wise product;

2. for all the other layers $l = L-1, \ldots, 1$ compute:

$$\delta^l = ((w^{l+1})^T \delta^{l+1} \odot \sigma'(z^l) \tag{3.12}$$

3. the weights and biases adaptation is performed by estimating:

$$\frac{\partial C}{\partial b^l_j} = \delta^l_j \tag{3.13}$$

$$\frac{\partial C}{\partial w^l_{jk}} = h^{l-1}_k \delta^l_j \tag{3.14}$$

4. finally, the weights and biases updating is modulated by a factor, called *learning rate*, which represents the intensity of the modification:

$$\mathbf{w} \leftarrow \mathbf{w} + \eta \nabla_{\mathbf{w}} C \qquad \mathbf{b} \leftarrow \mathbf{b} + \eta \nabla_{\mathbf{b}} C \tag{3.15}$$

It should be noted that the gradient of the loss function is an average estimation of the gradients calculated for the *extracted* training examples, in fact, the training set is split into several *batches*, each of them includes a relatively small number of examples extracted from the whole training set.

3.6.2 Regularisation techniques

In the Machine Learning context, overfitting and underfitting are two central challenges. Method performances are determined by the simultaneous reduction of both the training loss and the difference between the training and test error; these factors affect the method generalisation capabilities, which determine the amount of underfitting or overfitting.

Overfitting is a constant and ubiquitous problem in the ML context. It can be defined as the tendency of an algorithm to extract more information than necessary to capture the noise-less signal from the training set (Bashir et al. 2020), which results to a model hyper-specialisation on

[7] Here the notation is simplified: just a simple connection between the j-est and k-est elements is considered for the l-est layer, using the notation valid for a fully connected layer with N-dimensional input and M-dimensional output whose weights are encoded into an $N \times M$ weights matrix (i.e. w_{jk} for $j = 1, \ldots, N$ and $k = 1, \ldots, M$). In any case, it should be remarked that, for a convolutional layer connecting N-dimensional input and M-dimensional output, the weights are encoded into a 4D matrix (i.e. w_{jkmn} where j and k label the input and output dimension, whereas m and n label the vertical and horizontal offset). So, the Eq. 3.14 must be repeated $\forall m, n \in H \times W$ where H and W are the filter height and width. In both cases biases are MD vectors, i.e. with the output dimension.

the learning set preventing the gain of generalisation. This may happen due to a variety of reasons: a not appropriate number of samples in the training set, an excessively long learning phase, a trivial solution in the training set characterised by a too fast convergence (Huesmann et al. 2021). On the other hand, too simple model could not be able to extract abstract properties during the training phase; in this case, the model is *underfitted*; it could be the result of insufficient capacity, insufficient training, or insufficient information retention (Bashir et al. 2020).

Both problems are related to the model *capacity*, defined as the network capability to fit a wide variety of functions (Goodfellow et al. 2016). Overfitting occurs for large capacity models, it is usually identified by observing the evolution of the training and valid loss: assuming the independence between these two sets, a diverging trend is suggesting an high generalisation error and an inappropriate large variance, implying, therefore, the model overfitting (Zhang et al. 2018); on the contrary, underfitting takes place when a model is below its optimal capacity, it could result in a more or less constant performance with the ongoing of the training, even if the simultaneous reduction of both training and valid loss does not imply the lack of underfitting. In fact no underfitting predictor can be universally applicable (Sehra et al. 2021). Nevertheless, this does not rule out probabilistic bounds on the likelihood of underfitting.

Regularisation is an ensemble of techniques aimed at reducing the generalisation error without influencing the training error (Goodfellow et al. 2016), i.e. it prevents overfitting by reducing the model capacity. Generally, a model can be regularised by adding a penalty term, called regulariser, to the cost function (Bishop 2006). Common regularisation methods are: L2 and L1 parameter regularisation (Nowlan & Hinton 1992) and dropout (Srivastava et al. 2014). A overview on regularisation functions, together with a performative comparison, can be found in Parkes et al. (2021).

The L2 parameter norm penalty (commonly known as weight decay) drives the weights closer to zero, by adding a regularisation term to the cost function:

$$C' = C + \frac{1}{2}\lambda_2 \sum_{i=1}^{N} w_i^2 \tag{3.16}$$

where N is the number of involved weights and λ_2 is coefficient which modulates the strength of the decay. L2 regularisation is also known as Tikhonov regularisation (Groetsch 1984) or ridge regression (Tibshirani 2013). The addition of the weight decay term modifies the learning rule to shrink the weight vector by a factor which linearly scales with weights (i.e. $(1 - \eta\lambda_2)\mathbf{w}$), before performing the usual gradient update. The usage of L2 regularisation allows the model to constrain the weights when the input has a higher variance, by preserving adaptation along directions which significantly contribute to minimising the loss function and affecting only weight changes in the other directions (Goodfellow et al. 2016).

Formally, the L1 regularisation term added to the cost function is defined as:

$$C' = C + \lambda_1 \sum_{i=1}^{N} |w_i| \tag{3.17}$$

where, again, λ_1 determines the intensity of the regularisation and N is the number of involved weights. The effect of this regularisation is different from that of L2. Specifically, the L1 regularisation contribution to the adaptation mechanism no longer scales linearly with each w_i, instead it is a constant factor equals to $\lambda_1 \operatorname{sign}(w_i)$, applied to each involved w_i. This results in a more sparse solution implying that some parameters have an optimal value equals to zero. Due to this sparsity property, L1 regularisation has been used as a feature selection mechanism (Tibshirani 2013; Hara & Maehara 2016; Hara & Maehara 2017; Hastie et al. 2001), since features corresponding to weights identically zero may safely be discarded.

Dropout is a computationally inexpensive widely-used regularisation technique. It consists in the random masking of weights (i.e. it randomly deletes rows from the weight matrix). This function prevents units from co-adapting, reduces significantly overfitting and gives major improvements over other regularisation methods (Hinton et al. 2012; Goodfellow et al. 2013; Warde-Farley et al. 2013; Srivastava et al. 2014). Dropout can be thought as a bagging method which exponentially creates many ensembles of subnetworks sharing their parameters (Goodfellow et al. 2016). Dropout has been shown to be equivalent to L2 regularisation, after that the weight decay coefficients, corresponding to each input feature, have been individually scaled by their Fisher information (Wager et al. 2013).

A common way with which overfitting can be prevented is the application of an *early stopping* regularisation criterion (Prechelt 1997; Raskutti et al. 2011). Such technique exploits the valid set to iteratively measure the generalisation error. Given this collection of measurements, it is possible to impose a semi-empirical early stopping to the learning mechanism, by comparing the generalisation error at the current epoch with error evaluated at the previous epochs. Finding the epoch which minimises the generalisation error can be interpreted as the finding an approximate solution to the bias-variance tradeoff phenomena (von Luxburg & Schoelkopf 2008; Hastie et al. 2001), in fact stopping too early reduces variance but increases bias; on the contrary, stopping too late enlarges variance though reduces bias (Yao et al. 2007). Solving this bias-variance tradeoff leads to an early stopping rule. Such rules search for the minimum of the valid loss function assuming that this latter function has a convex (U-like) shape. One way with which an early stopping criteria can be implemented is through the creation of a *best-model checkpoint*, i.e. the saving of the model when a generalisation error minimum is found, by dumping the whole parameter ensemble. This parameter coping leads to a negligible memory cost, but it prevents a too early stopping. Indeed the form of the valid function is unknown, for some architectures (e.g. Residual Networks, He et al. 2015a) the training and valid losses have a descending trait characterised by a step-like behaviour: the loss function alternates periods during which it is practically constant with moments in which it rapidly converges. In these cases, it is advisable to force the model to overfit and to choose a posteriori the epoch that guarantees the best compromise between bias and variance.

3.6.3 Data augmentation

The best way with which a ML-model can generalise its learning is to train it on larger dataset (Goodfellow et al. 2016). Clearly, in practice, the amount of available data can be limited, thus, in order to get around this problem, artifacts can be added to training set (Cui et al. 2015; Poole et al. 2014; Jannik Bjerrum 2017; Mikołajczyk & Grochowski 2018). In the DL context, the simplest way with which artifacts can be produced is through image rotations, vertical and horizontal flipping (Perez & Wang 2017). The inclusion of these images in the training set also offered the possibility to make the network invariant to these operations (Goodfellow et al. 2016; Shorten & Khoshgoftaar 2019). Of course, this approach is not applicable when objects in the images have a uniquely defined spatial orientation (e.g. alphabet letters or numeric digits), but works as an advantage for astronomical images as there is no specified spatial orientation for the observed sources. On the other hand, an uncontrolled augmentation could introduce false correlations among the training samples, therefore, typically, only a fraction of sources have been subject to these transformations.

Even with large datasets, a model can overfit when the problem is trivial. Especially in classification tasks, the repetition of the same slightly-warped sample, the lack of borderline or interloper objects, the building of to crispy categorical datasets, could lead to overfitting or, more commonly, in the inability to predict the membership for samples slightly different

from the training examples. This can be observed by testing the network on training samples on which a faint noise is added, producing the so-called *adversarial* examples (Szegedy et al. 2013). Such adversarial samples are so similar to the originals that an human observer is not able to distinguish between the original and the adversarial example, but the network entirely mistakes the prediction. Although this is not properly overfitting, the model turns out to be incapable to generalise. In order to solve this problem adversarial samples can be added to training set (Goodfellow et al. 2014b). Adversarial examples can be generated by augmenting the dataset with the inclusion of noised version of original images (Goodfellow et al. 2014b), by warping images (Zhao et al. 2020) or by directly perturbing the feature nodes (Kong et al. 2020). In the context of astronomy, the limited amount of labelled data, the heterogeneous and the (typically) low signal-to-noise ratio prevent to apply image deformations or the addition of noise (for the same reason the cropping of astronomical images is avoided). In this work, given such complicated uncontrollable environment, in order to introduce adversarial examples, very complex samples have been added to the datasets. These examples are objects whose membership inferring is sophisticated and it is typically deduced by combining photometric and spectroscopic information together with human experience, or, at least, whose classification is not so crispy and it depends on flexible criteria. Specifically, in this thesis we addressed two problems: the identification of cluster members (Chap. 5) and of galaxy-galaxy strong-lenses in galaxy clusters (Chap. 6). Concerning cluster galaxy membership, we built the KB by exploiting spectroscopic information, by assuming as member a galaxy with a cluster rest-frame velocity separation within $\pm 3000\,\mathrm{km\,s^{-1}}$. In this way, adversarial examples are represented by those objects weakly bound to the cluster, which are indistinguishable from members based on their photometry and colors, but are correctly classified only through a dynamical analysis. With regard to the strong-lenses classification, the KB positive class is composed by simulated lensing events, while the negative class is represented by spectroscopic members which do not reveal any strong-lensing feature based on visual classification. Adversarial examples also include faint injected sources or lenses with small Einstein radii, in these cases arc or ring-like features are masked by the lens galaxy, resulting into images indistinguishable from the negative class.

3.6.4 Optimisation

Once the loss function has been set together with the regularisers, in order to apply an update on weights and biases, the network requires an optimiser. The optimisation is the process with which the generalisation error is minimised. Such error is evaluated on the training set, that underlies an unknown probability distribution; for such reason this error is also called as *empirical risk* (Goodfellow et al. 2016), defined as:

$$\mathbb{E}_{(\mathbf{x},y)\sim\hat{p}_{\mathrm{data}}(\mathbf{x},y)}[C(f(\mathbf{x};\mathbf{w},\mathbf{b}),y)] = \frac{1}{N}\sum_{i=1}^{N} C(f(\mathbf{x}^{(i)};\mathbf{w},\mathbf{b}),y^{(i)}) \tag{3.18}$$

where N is the number of training samples, C is the selected cost function evaluated by combining the network output given the current weights and biases $f(\mathbf{x};\mathbf{w},\mathbf{b})$ and the ground truth y, on the empirical distribution of the data, $\hat{p}_{\mathrm{data}}(\mathbf{x},y)$. Eq. 3.18 is stating that the optimiser minimises the expectation value of the loss function taken across the empirical data distribution.

Non-convex loss functions, such as for ANNs, might have many local minima, a common issue that must be addressed by the optimiser. Local minima could pose a serious problem when they correspond to high costs (i.e. high values of the loss function) compared to the cost of global minimum (Brady et al. 1989; Sontag & Sussmann 1989; Floudas & Gounaris 2009). Although this problem continues to interest various research areas, many authors claimed that very large network architectures are not affected by this problem, since the local minima have

low costs, comparable with the global minimum; therefore, it is not important to find the real global minimum, rather a minimum with a sufficiently low cost, trying to avoid the saddle points which are more frequent than high-cost local minima (Saxe et al. 2013; Dauphin et al. 2014; Goodfellow et al. 2014c). Besides saddle points and local minima, there are other points with zero gradient: maxima (that typically do not attract optimisers) and flatten regions (in which gradient and Hessian are all zero).

Optimisation algorithms are based on the Stochastic Gradient Descendent (SGD, Bishop 2006; Goodfellow et al. 2016). An unbiased estimation of the gradient can be obtained by exploiting an average of gradients evaluated on a batch of m samples extracted from the training set (Bottou & Bousquet 2008). In the simplest formulation SGD updates weights and biases using a fraction of loss function gradient:

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \eta \nabla C(\boldsymbol{\theta}_t) \tag{3.19}$$

where $\boldsymbol{\theta}_t$ is parameter ensemble to optimise (i.e. weights and biases) at the current epoch t, η is the learning rate labelling the magnitude of the adaptation. Since the SGD estimator introduces a source of noise due to random sampling of m examples, it is a good practice to gradually decrease hyper-parameter η with the ongoing of the training through a linear decay (Wu et al. 2019b).

In order to accelerate SGD, it is possible to add a *momentum* (Polyak 1964). This algorithm is designed to accumulate a decaying averaged movement by summing the past gradients and impose an adaptation in this cumulative direction. The momentum is inspired from the physical analogy: assuming unit mass, the velocity may also be regarded as the momentum of a particle. Formally, this algorithm introduces a variable $\mathbf{v}$ which plays the role of a velocity (i.e. it imposes a direction and an intensity to the parameter movement in its space), which is computed from an exponentially decaying average of the accumulated gradients:

$$\mathbf{v} \leftarrow \alpha \mathbf{v} - \eta \nabla_{\boldsymbol{\theta}} \left(\frac{1}{m} \sum_{i=1}^{m} C(f(\mathbf{x}^{(i)}; \boldsymbol{\theta}), y^{(i)}) \right) \tag{3.20}$$

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \mathbf{v} \tag{3.21}$$

where α is an hyper-parameter determines the contributions of the previous estimated gradients. A variant of the momentum algorithm was introduced by Sutskever et al. (2013) inspired by Nesterov's accelerated gradient method (Nesterov 1983, 2003). Updating rules 3.20 - 3.21 is rewritten as:

$$\mathbf{v} \leftarrow \alpha \mathbf{v} - \eta \nabla_{\boldsymbol{\theta}} \left(\frac{1}{m} \sum_{i=1}^{m} C(f(\mathbf{x}^{(i)}; \boldsymbol{\theta} + \alpha \mathbf{v}), y^{(i)}) \right) \tag{3.22}$$

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \mathbf{v} \tag{3.23}$$

The difference with the standard momentum is *where* the gradient is estimated: in the Nesterov momentum the gradient is evaluated *after* the velocity application and it can be thought as a *correction* to the standard momentum (Goodfellow et al. 2016).

The momentum algorithms try to minimise the path through the minimum by averaging the previous gradients, paying the price of adding an extra hyper-parameter to an already large hyper-parameter space. Considering that learning rate significantly affects model performances, several optimisers use an independent learning rate for each parameter and automatically update these learning rates during the training. The most used self-adapting optimisers are Adagrad (Duchi et al. 2011), RMSProp (Hinton et al. 2012), Adam (Kingma & Ba 2014) and Adadelta (Zeiler 2012).

In the AdaGrad (Adaptive Gradient, Duchi et al. 2011) algorithm, the learning rates of all parameters are individually adapted with a scaling factor which is inversely proportional to the square root of the sum of *all* the previous squared values of the gradient. In this way, parameters with largest derivative have a learning rate which rapidly decreases, and vice versa. The adaptation rule can be summarised as:

$$\mathbf{g} \leftarrow \nabla_\theta \left(\frac{1}{m} \sum_{i=1}^{m} C(f(\mathbf{x}^{(i)}; \boldsymbol{\theta}), y^{(i)}) \right) \tag{3.24}$$

$$\mathbf{r} \leftarrow \mathbf{r} + \mathbf{g} \odot \mathbf{g} \tag{3.25}$$

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \frac{\eta}{\delta + \sqrt{\mathbf{r}}} \odot \mathbf{g} \tag{3.26}$$

where δ is a small constant used for numerical stability, typically 10^{-7}, $\mathbf{g}$ is the gradient; in the Eq. 3.26 both square root and division are applied element-wise. Even if AdaGrad exploits some theoretical properties, empirically, however, the collection of all historical squared gradients can turn out in a excessive decrease of the learning rates (Goodfellow et al. 2016), indeed when AdaGrad algorithm is applied on non-convex loss function, the learning path might pass through different structures by finally arriving in a locally convex region with a negligible value of the learning rate. RMSProp (Hinton 2012) is a batch version of RProp (Resilient backPropagation, Riedmiller & Braun 1993) algorithm, which solves this problem by including an exponentially decaying average to emphasise the contribution of more recent gradient respect to extreme past gradients. The adaptation rule of RMSProp can obtained by replacing Eqs. 3.24 - 3.26 with:

$$\tilde{\boldsymbol{\theta}} \leftarrow \boldsymbol{\theta} + \alpha \mathbf{v} \tag{3.27}$$

$$\mathbf{g} \leftarrow \nabla_\theta \left(\frac{1}{m} \sum_{i=1}^{m} C(f(\mathbf{x}^{(i)}; \tilde{\boldsymbol{\theta}}), y^{(i)}) \right) \tag{3.28}$$

$$\mathbf{r} \leftarrow \rho \mathbf{r} + (1 - \rho) \mathbf{g} \odot \mathbf{g} \tag{3.29}$$

$$\mathbf{v} \leftarrow \alpha \mathbf{v} - \frac{\eta}{\delta + \sqrt{\mathbf{r}}} \odot \mathbf{g} \tag{3.30}$$

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \mathbf{v} \tag{3.31}$$

where also Nesterov's momentum has been included[8]; respect to the AdaGrad algorithm, RMSProp has a further hyper-parameter ρ, which is the decay rate used to discard past history.

Adam (Kingma & Ba 2014) is another self-adaptive learning rate optimisation algorithm, whose name derive from the expression *adaptive moments*. It can be thought as variant of RMSProp where first and second order momentum estimations are embedded and both moments are bias-corrected to take into account their initialisation to zero. Adam updating rule can be summarised as:

[8] Nesterov's momentum can be added also in AdaGrad by combining Nesterov Eqs. 3.22 and 3.22 with AdaGrad optimisation rules 3.24 - 3.26.

$$\mathbf{g} \leftarrow \nabla_\theta \left(\frac{1}{m} \sum_{i=1}^{m} C(f(\mathbf{x}^{(i)}; \boldsymbol{\theta}), y^{(i)}) \right) \tag{3.32}$$

$$t \leftarrow t + 1 \tag{3.33}$$

$$\mathbf{s} \leftarrow \rho_1 \mathbf{s} + (1 - \rho_1)\mathbf{g} \tag{3.34}$$

$$\mathbf{r} \leftarrow \rho_2 \mathbf{r} + (1 - \rho_2)\mathbf{g} \odot \mathbf{g} \tag{3.35}$$

$$\hat{\mathbf{s}} \leftarrow \frac{\mathbf{s}}{1 - \rho_1^t} \tag{3.36}$$

$$\hat{\mathbf{r}} \leftarrow \frac{\mathbf{r}}{1 - \rho_2^t} \tag{3.37}$$

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \eta \frac{\hat{s}}{\delta + \sqrt{\mathbf{r}}} \tag{3.38}$$

where ρ_1 and ρ_2 are the decay rates for moment estimations, $\mathbf{s}$ and $\mathbf{r}$ are the biased first and second moments, while $\hat{\mathbf{s}}$ and $\hat{\mathbf{r}}$ are bias-corrected first and second moments. Despite the theoretical properties of Adam, it is not robust to the hyper-parameter setup which is critical for the algorithm performance. AMSGRad (Reddi et al. 2019) try to circumvent this problem by combining the historical collection of past steps (as AdaGrad) with the unbiased estimation of the first and second moments. The adaptation rule is modified by adding a selection criteria on the second order momentum: $\hat{\mathbf{s}} \leftarrow \max(\hat{\mathbf{s}}^t, \hat{\mathbf{s}}^{t-1})$, i.e. by preserving the maximum between the current second-order momentum and the previous step estimation. This slightly modification allows a stable convergence by also preventing Adam pitfalls.

Another variant of AdaGrad is AdaDelta (Zeiler 2012) algorithm, which combines; (i) an accumulation of previous steps over a restricted window (Schaul et al. 2012), (ii) a parameter updating which involves an approximation of Hessian diagonal (Becker & Lecun 1989), (iii) a gradual reduction of the learning rates near minima based on the achieved amount of epochs (Robbins & Monro 1951). The adaptation rule can be expressed as:

$$\mathbf{g}_t \leftarrow \nabla_\theta \left(\frac{1}{m} \sum_{i=1}^{m} C(f(\mathbf{x}^{(i)}; \boldsymbol{\theta}), y^{(i)}) \right) \tag{3.39}$$

$$\mathbf{r}_t = \rho \mathbf{r}_{t-1} + (1 - \rho)\mathbf{g} \odot \mathbf{g} \tag{3.40}$$

$$\Delta\boldsymbol{\theta}_t = -\eta_t \frac{\sqrt{\frac{1}{w} \sum_{k=t-w-1}^{t-1} (\Delta\boldsymbol{\theta})_k^2 + \delta}}{\sqrt{\mathbf{r}_t + \delta}} \odot \mathbf{g}_t \tag{3.41}$$

$$(\Delta\boldsymbol{\theta})_t^2 \leftarrow \rho \sqrt{\frac{1}{w} \sum_{k=t-w-1}^{t-1} (\Delta\boldsymbol{\theta})_k^2} + (1 - \rho)(\Delta\boldsymbol{\theta})_t^2 \tag{3.42}$$

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \Delta\boldsymbol{\theta}_t \tag{3.43}$$

where w is the window size, ρ is the decay rates for the second order momentum, η_t is the current value of the learning rate gradually reduced as the optimiser approaches to a minimum. Besides the exponentially decaying of the accumulated momentum (Eq. 3.40), the algorithm maintains the w most recent squared parameter updates $(\Delta\boldsymbol{\theta})_k^2$ exponentially decreased with the same decay rate ρ used for the momentum. The computed parameter modification (Eq. 3.41) is an approximation of the Newton's method used to find local minima and maxima (Davidon 1968) through the Hessian:

$$\Delta\boldsymbol{\theta}_t \propto H_t^{-1}\mathbf{g}_t \sim -\frac{1}{|diag(H_t)| + \delta} \odot \mathbf{g}_t \tag{3.44}$$

since the gradient has been computed as an average over m samples, Eq. 3.41 is actually involving the root mean square of the gradient for each examples, according to the definition given in Eq. 3.18:

$$\mathbf{g}_t = \nabla_\theta \left(\frac{1}{m} \sum_{i=1}^{m} C(f(\mathbf{x}^{(i)}; \boldsymbol{\theta}), y^{(i)}) \right) = \frac{1}{m} \sum_{i=1}^{m} \nabla_\theta C(f(\mathbf{x}^{(i)}; \boldsymbol{\theta}), y^{(i)}) \equiv \mathbb{E}[\mathbf{g}]_t \tag{3.45}$$

$$\begin{aligned} \implies \mathbf{g}_t^2 &\equiv \mathbb{E}[\mathbf{g}^2]_t = \mathbb{E}[\mathbf{g} \odot \mathbf{g}]_t = \\ &= \frac{1}{m} \sum_{i=1}^{m} \nabla_\theta C(f(\mathbf{x}^{(i)}; \boldsymbol{\theta}), y^{(i)}) \odot \nabla_\theta C(f(\mathbf{x}^{(i)}; \boldsymbol{\theta}), y^{(i)}) = \\ &= \frac{1}{m} \sum_{i=1}^{m} \left(\nabla_\theta C(f(\mathbf{x}^{(i)}; \boldsymbol{\theta}), y^{(i)}) \right)^2 \equiv \mathrm{RMS}[\mathbf{g}]_t \end{aligned} \tag{3.46}$$

and, given the momentum definition (Eq. 3.40), under the assumption of locally smooth diagonal curvature, the reverse of the diagonal Hessian is approximated with only RMS measures of the $\mathbf{g}$ and $\Delta\boldsymbol{\theta}$:

$$\frac{1}{diag(H_t)} \propto \frac{\sqrt{\frac{1}{w} \sum_{k=t-w-1}^{t-1} (\Delta\boldsymbol{\theta})_k^2 + \delta}}{\sqrt{\mathbf{r}_t + \delta}} \sim \frac{\mathrm{RMS}[\Delta\boldsymbol{\theta}]_{t-1}}{RMS\,[\mathbf{g}]_t} \sim \frac{1}{(\Delta f)_t^2 / (\Delta\boldsymbol{\theta})_t^2} \sim \frac{1}{\partial^2 f / \partial \boldsymbol{\theta}_t^2} \tag{3.47}$$

since this approximation is always positive (Becker & Lecun 1989), it ensures that the updating direction follows the negative gradient.

Adadelta showed robustness with respect to different types of data inputs, number of layers, hyper-parameter setup and network architectures with performance at least comparable with other optimisers and negligible computational cost (Zeiler 2012). It has been efficiently adopted to solve various problems: classification of variable sources (Kim et al. 2021), identification of post-merged galaxies (Bickley et al. 2021), prediction of morphological galaxy types (Cavanagh et al. 2021), synthetic galaxy image generation based on cross-survey mapping (Buncher et al. 2021), even to invert nonlinear Schrödinger equation (Wang & Li 2021), just to cite some recent results. In this work, we tested different architectures, by also varying the hyper-parameter setup and the optimiser, finding, at the end of ~ 100 different networks training, that Adadelta turned out to be the best-choice as optimiser.

Finally, *batch normalisation* (Ioffe & Szegedy 2015) is a recent innovations in deep learning optimisation; even if it is actually not an optimisation algorithm, it improves the optimisation with an adaptive reparametrisation of very deep model, making, sometimes, dropout unnecessary (Goodfellow et al. 2016); although it looks like a regulariser, it is not at all, since it depend on parameters which are adapted during the training. Such technique can be applied to normalise any layer output: let $\mathbf{X} = x_i^k$ the input of the batch normalisation layer (i.e. the output of the previous layer), where $i = 1, \ldots, m$ labels the extracted samples and $k = 1, \ldots, D$ labels the dimension of the input vector, each dimension is independently transformed with:

$$\mu_{BN} \leftarrow \frac{1}{m} \sum_{i=1}^{m} x_i \quad \sigma_{BN}^2 \leftarrow \frac{1}{m} \sum_{i=1}^{m} (x_i - \mu_{BN})^2 \tag{3.48}$$

$$\hat{x}_i \leftarrow \frac{x_i - \mu_{BN}}{\sqrt{\sigma_B^2 N + \delta}} \tag{3.49}$$

$$y_i \leftarrow \gamma \hat{x}_i + \beta \equiv \mathrm{BN}_{\gamma,\beta}(x_i) \tag{3.50}$$

where δ is a small constant value used for numerical stability. Thus, the mean and variance are computed and used for normalise the input (for each dimension), then, a scaling and shifting are

applied where γ and β coefficients are adapted during the training as any other fitted parameter; so that, updating directions imposed by the gradient will never force the increasing of standard deviation or mean.

3.6.5 Knowledge Base partitioning and augmenting

As already introduced at the beginning of this section, in this work we opted for a stratified k-fold partitioning (Hastie et al. 2001; Kohavi 1995) to handle the training and test phases. In this way, the test set fully covers the input parameter space: the whole data set was split into $k = 10$ non-overlapping folds, of which, iteratively, one extracted subset was used as a blind test set, while the others were taken as a training set. Such an approach has several advantages: *(i)* increase of the statistical significance of the test set; *(ii)* the blind test is performed only on original images; and *(iii)* it ensures a complete coverage of both training and test sets, keeping them well-separated at the same time.
Before the k-fold splitting we randomly extracted a small sample of sources (10% of the data set), reserved as validation set during the training phase in order to control the gradual reduction of the learning rate on the plateau of the cost function (Bengio 2012) and an early stopping regularisation process (Prechelt 1997; Raskutti et al. 2011).

The data preparation flow is depicted in Fig. 3.3: (*i*) the dataset is composed by multi-band images; (*ii*) a fraction of sources (10%) is extracted as validation set; (*iii*) the remaining samples are split into $k = 10$ folds without overlapping; (*iv*) for each of them, a fraction of samples is augmented through cutout rotations and flips; (*v*) the training sets are built by concatenating $k - 1$ folds (composed by the original images and the artefacts) and the learning is evaluated on the k-th fold (without artefacts), acting as blind test; (*vi*) finally, the model performances are evaluated on the whole training set, obtained by stacking all its (test) folds. Concerning the augmentation, the cutouts have been rotated around the three right angles and flipped with respect to the horizontal and vertical axes (an example of such process is shown in Fig. 3.4). Given the considerable number of model parameters to fit ($\sim 10^5$), deep learning networks require an adequate amount of samples, in order to avoid overfitting (Cui et al. 2015; Perez & Wang 2017). However, an uncontrolled augmentation could introduce false correlations among the training samples. Therefore, only a fraction of sources have been subject to these transformations. The resulting augmentation factor is computed as $1 + 5 * u$ times the original dimension of the training set (where u is the fraction of samples involved into the augmentation process). Obviously, such augmentation process involves only the training images.

3.6.6 Statistical estimators

In order to assess the model classification performances, we chose the following statistical estimators: *average efficiency* (among all classes, abbreviated as *AE*), *purity* (also know as *positive predictive value* or *precision*, abbreviated as *pur*), *completeness* (also known as *true positive rate* or *recall*, abbreviated as *comp*), and F1-score (a measure of the combination of purity and completeness, abbreviated as *F1*). Such set of statistical estimators is directly derived from the classification confusion matrix (Stehman 1997).

In a binary confusion matrix, as in the example shown in Table 3.2, columns indicate the class objects as predicted by the classifier, while rows refer to the true objects per class. The main diagonal terms contain the number of correctly classified objects for each class, while the terms FP and FN report the amount of, respectively, False Positives and False Negatives.

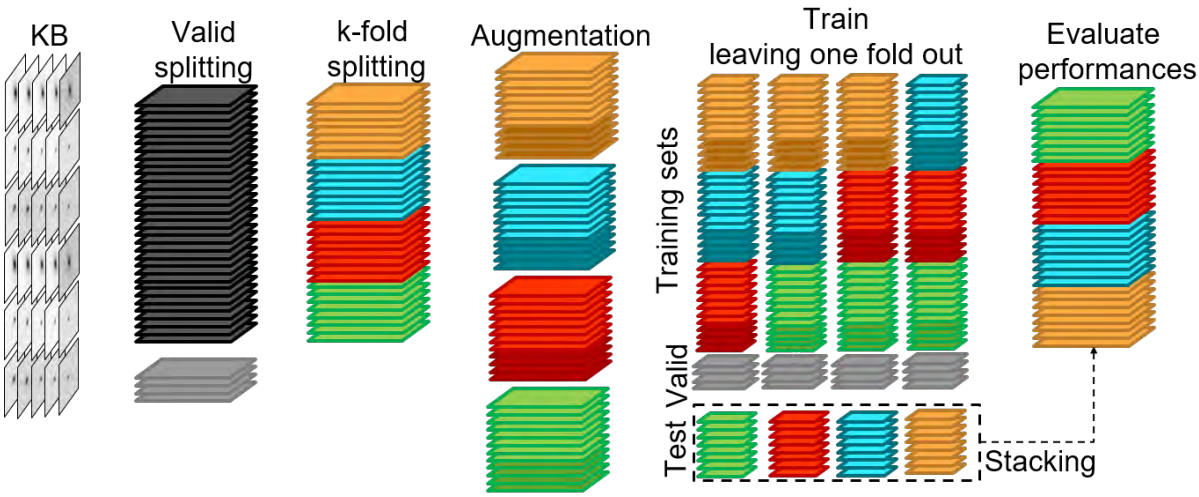


Figure 3.3: Data preparation flow: from the whole dataset (i.e. the knowledge base) a validation set is extracted. The rest of the dataset is split through a k-fold partitioning process (in this image, we simplified the figure assuming $k = 4$ folds, while in reality we used $k = 10$). The training samples are then arranged, by permuting the involved augmented folds, while the test samples do not include the artefact images generated by the augmentation process. These sets are finally stacked in order to evaluate the global training performances.

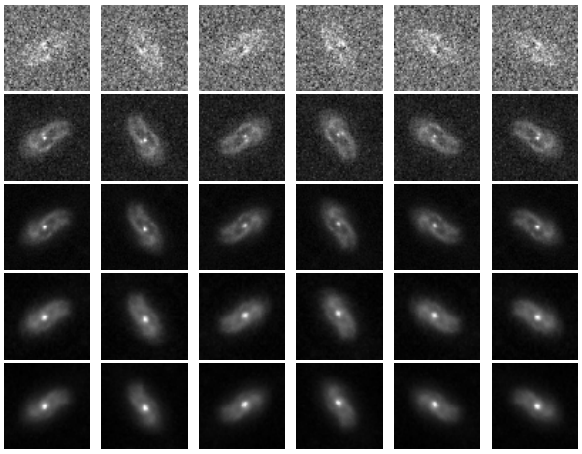


Figure 3.4: Data augmentation example for a cluster member at redshift $z = 0.531$ (e.g. within the gravitational potential of the galaxy cluster MACS J1149+2223). Five HST bands are represented from the top to the bottom ($F435$, $F606$, $F814$, $F105$, $F140$). The first column shows the original cut-out, while the three rotations (90°, 180°, 270°) are reported in columns $2-4$. The two vertical and horizontal flips are shown in the last two columns.

		Predictions	
		positive	negative
True	positive	**TP**	**FN**
	negative	**FP**	**TN**

Table 3.2: Generic confusion matrix for a binary classification problem. In a confusion matrix, columns indicate the number of objects per class, as predicted by the classifier, while rows are referred to the true (known) objects per class. Hence, the main diagonal terms report the number of correctly classified objects for each class. While, the terms FP and FN count, respectively, the False Positives and False Negative quantities.

Therefore, the derived estimators are computed as:

$$\text{AE} = \frac{TP + TN}{TP + FP + TN + FN} \tag{3.51}$$

$$\text{pur} = \frac{TP}{TP + FP} \tag{3.52}$$

$$\text{comp} = \frac{TP}{TP + FN} \tag{3.53}$$

$$\text{F1} = 2 \cdot \frac{\text{pur} \cdot \text{comp}}{\text{pur} + \text{comp}} \tag{3.54}$$

$$\tag{3.55}$$

The *AE* is the ratio between the sum of the correctly classified objects (for all the involved classes) and the total amount of objects; it describes an average evaluation weighted on all involved classes. The *purity* of a class is the ratio between the correctly classified objects and the sum of all objects assigned to that class (i.e. the predicted membership); it measures the precision of the classification. The *completeness* of a class is the ratio between the correctly classified objects and the total amount of objects belonging to that class (i.e. the *true* membership), it estimates the sensitivity of the classification. By definition, the dual quantity of purity is the 'contamination', a measure which indicates the amount of misclassified objects for each class. Finally, the *F*-measure provides a way to combine purity and completeness into a single measure; respect to the Eq. 3.54, it can be expressed in a more general (weighted) form (Chinchor 1992):

$$F_\beta = (1 + \beta^2)\frac{\text{pur} \cdot \text{comp}}{(\beta^2 \cdot \text{pur}) + \text{comp}} \qquad \beta > 0 \tag{3.56}$$

where β is the relative importance given to the completeness over the purity. In this work we used $\beta = 1$, i.e. we equally weighted purity and completeness, resulting into an *F1-score* which is the harmonic mean between purity and completeness. The *F1-score* allows to include the information about the gap between purity and completeness: indeed, a method with *pur*=*comp*= 0.5 has an higher *F1* than a method with *pur*= 0.2 and *comp*= 0.8.

Among these estimators, completeness and purity are the most interesting estimators, suitable for measuring the quality of the classification performed by any method. The completeness, in fact, measures the capability to extract a *complete* set of candidates of a given class, while purity estimates the capability of selecting a *pure* set of candidates (thus, minimising the contamination). Therefore, the classification quality is usually based on either one of such two estimators or their combination, depending on the specific interest of an experiment (D'Isanto et al. 2016). The statistical evaluation was completed by also using the Receiver Operating Characteristic curve (ROC, Hanley & McNeil 1982), which is a diagram where the True Positive Tate (TPR, i.e. the completeness rate) is plotted versus the False Positive Rate (FPR, i.e. the contamination rate, which corresponds to 1− purity) by varying a membership probability threshold. The

model performances are measured in terms of the area under the curve (AUC), thus providing an aggregate measure of performance across all possible classification thresholds: an area of 1 represents a perfect classification, while an area of 0.5 indicates a useless result (akin to a toss of a coin). Examples of ROC curves can be found in Fig. 5.7, Fig. 5.8, Fig. A.1 and Fig. 6.8.

3.7 Implemented architectures

In this work several CNN architectures have been implemented: the Visual Geometry Group network (hereafter VGG, Simonyan & Zisserman 2014), Residual Network (hereafter ResNet, He et al. 2015a), Inception Network (herafter GNet[9], Szegedy et al. 2014) and Residual Inception Network (hereafter ResGNet, Szegedy et al. 2016). All these architectures have been implemented through `keras` (Chollet et al. 2015) and `tensorflow` (Abadi et al. 2015) exploiting their mutual integration. Both of them are open-source Python libraries, allowing the automatic handling of the Graphic Processing Unit (GPU), achieving a huge gain in terms of computational cost. In this work the experiments were performed with an NVIDIA GPU Titan Xp and an NVIDIA GPU Quadro P5000, requiring $\sim$ 20 up to $\sim$ 50 minutes to complete the training (on a single fold, see Sec. 3.6.5; this means that a complete training and test experiment, i.e. performed on 10 folds, requires $\sim$ 200 up to $\sim$ 500 minutes). We performed hundreds of experiments, by varying the models, the architecture configuration, by exploring the hyper-parameter and the feature space, searching for both the best network setup and the optimal physical space. At the end of thousands of hours of training we found that the implemented VGG-like networks turned out to be the best model, showing a high resilience to the hyper-parameter variation and training configuration, it resulted the network with the best performances and the great generalisation capabilities (a comparison between the networks is shown in Tab. A.1 in appendix A, regarding the cluster member identification, and in Tab. B.1 in appendix B, for what concern the galaxy-galaxy strong-lens classification), furthermore it is the model which requires the lowest computing time and memory resources. For all these reasons, the other implemented networks are briefly outlined in Sec. 3.7.2, while VGG is widely described in Sec. 3.7.1, and, for the same reasons, the analysis of the performances and behaviour, classification capabilities and limits are referred only to the VGG.

3.7.1 Visual Geometry Group network

VGG (Visual Geometry Group) network has been presented in the work by Simonyan & Zisserman (2014), where the authors analyse the model dependence on the depth building and comparing network with 11, 13, 16 and 19 layers, making the VGG19 a standard reference. Besides the comparison between architectures with different depth, the peculiarities of VGG are: (i) the usage of only 3×3 kernel filters and (ii) the chain of several convolutional layers with the same number of filters, interspersed by a max-pooling. The exclusive adoption of small filters allows the building of such deep model, constraining, at the same time, the number of parameters and, so, saving memory. The repeated application of many convolutional layers before the pooling operation allows to exploit translation invariance, by replacing a single layer with a mini-network composed by many smaller layers (Szegedy et al. 2015). For example, the capability of capture correlation between signals of a 7×7 convolutional layer can be recovered by a reduction of the geometric size of the filters and the concatenation of three 3×3

[9] In their original work, Szegedy et al. (2014) chose the name GoogleLeNet which is the name with which the network has been submitted to the ILSVRC14 competition, as a homage to Yann LeCuns pioneering LeNet 5 network presented in LeCun et al. (1989). In this work the name has been simplified as GNet to avoid the weighting of text and tables.

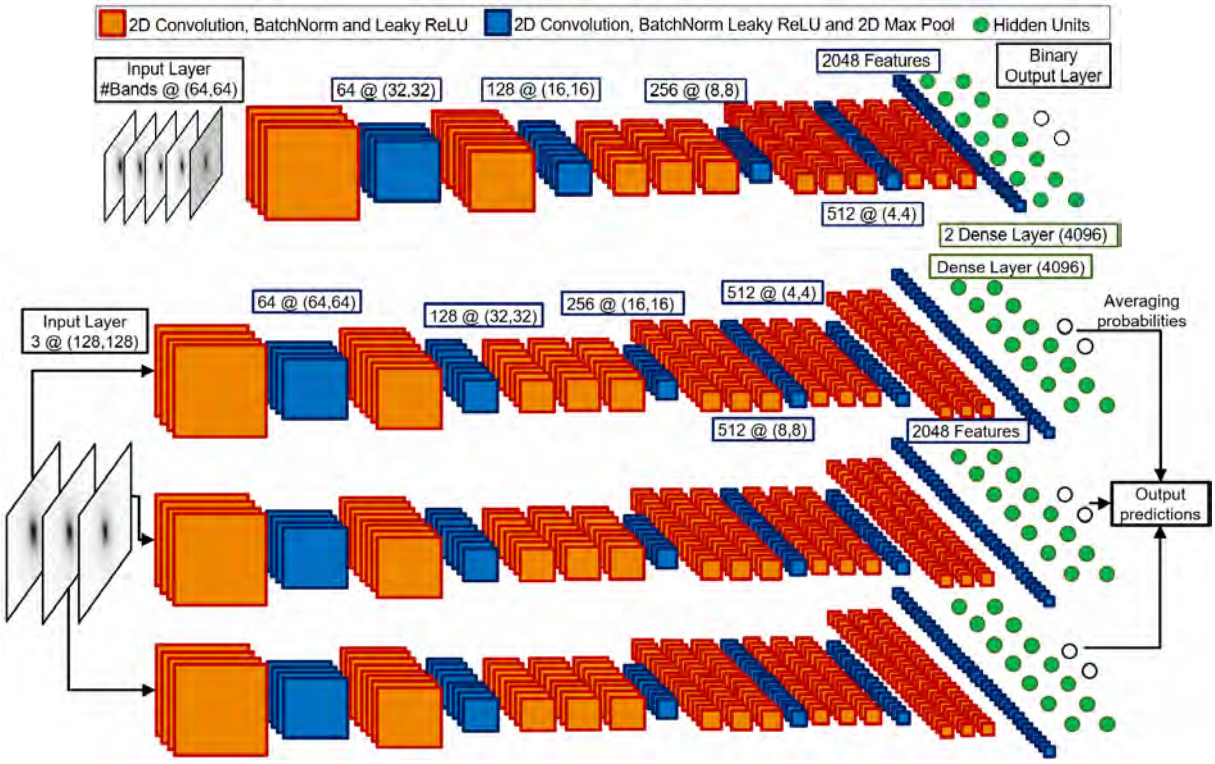


Figure 3.5: Streamlined representation of the VGG-like architectures used in this work. The VGG is shown in the top of the panel, while the SC-VGG is displayed in the bottom of the panel. Orange and blue items describe two different block operations, respectively: (i) convolution, batch normalisation and activation function, (ii) convolution, batch normalisation, activation function and pooling. The simultaneous reduction of the square dimensions and their increasing amount intuitively represent the abstraction process typical of a CNN. Green circular units are arranged in order to describe the fully connected (i.e. dense) layers. The dimensions of the feature maps are reported for each pooling operation, together with the number of features extracted by the CNN.

convolutional layers. In this example also the computing time is reduced: since the three layers have the same number of channels C (i.e. the number of 3×3 filter kernels) the amount of parameters is $3(3^2C) = 27C$, while for a single 7×7 convolutional layer this amount grows to $7^2C = 49C$, i.e. 81% more. Same reason can be done when a 5 layer is replaced by two 3×3 layers, resulting into a saving of 94% of parameters.

In this work, a VGG-like network has been implemented; respect to the the original work by Simonyan & Zisserman (2014), the proposed model has been set with: (i) a LeackyReLU activation function for each neural units, (ii) a batch normalisation layer after each convolutional layer, (iii) a dropout regulariser between the dense layers. Moreover, only for the identification of galaxy-galaxy strong-lenses (see Chap. 6), we implemented a VGG in which each channel is independently propagated toward the output layer, obtaining a membership probabilities for each channel; the probability vectors related to the positive class (i.e. the GGSL) are averaged to compute a global output, whereas the probability vector for the negative class is calculated as a complementary vector, i.e. $1 - Pr(y = \text{GGSL})$; these vectors are used to estimate the binary cross-entropy (Eq. 3.6), which is used to during the back-propagation phase. Such single channel network has been called Single Channel VGG (hereafter, SC-VGG). A simplified layout of the implemented VGGs is shown in Fig. 3.5, where the VGG is depicted at the top of the panel, while the SC-VGG is presented at the bottom of the figure, this latter refers to the dataset whose samples are images with a side of 128 pixels, whereas for the dataset with cutout side of 256 pixels, the last operational block (convolution chain combined with the MaxPool2D operation) is

VGG for CLMs identification			VGG for GGSLs identification		
Layer	Output Shape	Params #	Layer	Output Shape	Params #
Input layer	(64, 64, NC)	0	Input layer	(128, 128, 3)	0
Conv2D LReLU	(64, 64, 64)	6976	Conv2D LReLU	(128, 128, 64)	6976
Conv2D LReLU	(64, 64, 64)	36928	Conv2D LReLU	(128, 128, 64)	36928
Max Pool2D	(32, 32, 64)	0	Max Pool2D	(64, 64, 64)	0
Conv2D LReLU	(32, 32, 128)	73856	Conv2D LReLU	(64, 64, 128)	73856
Conv2D LReLU	(32, 32, 128)	147584	Conv2D LReLU	(64, 64, 128)	147584
Max Pool2D	(16, 16, 128)	0	Max Pool2D	(32, 32, 128)	0
Conv2D LReLU	(16, 16, 256)	295168	Conv2D LReLU	(32, 32, 256)	295168
Conv2D LReLU	(16, 16, 256)	590080	Conv2D LReLU	(32, 32, 256)	590080
Conv2D LReLU	(16, 16, 256)	590080	Conv2D LReLU	(32, 32, 256)	590080
Conv2D LReLU	(16, 16, 256)	590080	Conv2D LReLU	(32, 32, 256)	590080
Max Pool2D	(8, 8, 256)	0	Max Pool2D	(16, 16, 256)	0
Conv2D LReLU	(8, 8, 512)	1180160	Conv2D LReLU	(16, 16, 512)	1180160
Conv2D LReLU	(8, 8, 512)	2359808	Conv2D LReLU	(16, 16, 512)	2359808
Conv2D LReLU	(8, 8, 512)	2359808	Conv2D LReLU	(16, 16, 512)	2359808
Conv2D LReLU	(8, 8, 512)	2359808	Conv2D LReLU	(16, 16, 512)	2359808
Max Pool2D	(4, 4, 512)	0	Max Pool2D	(8, 8, 512)	0
Conv2D LReLU	(4, 4, 512)	2359808	Conv2D LReLU	(8, 8, 512)	2359808
Conv2D LReLU	(4, 4, 512)	2359808	Conv2D LReLU	(8, 8, 512)	2359808
Conv2D LReLU	(4, 4, 512)	2359808	Conv2D LReLU	(8, 8, 512)	2359808
Conv2D LReLU	(4, 4, 512)	2359808	Conv2D LReLU	(8, 8, 512)	2359808
Max Pool2D	(2, 2, 512)	0	Max Pool2D	(4, 4, 512)	0
Flatten	(2048)	0	Conv2D LReLU	(4, 4, 512)	2359808
Dense LReLU	(4096)	8392704	Conv2D LReLU	(4, 4, 512)	2359808
Leaky ReLU	(4096)	0	Conv2D LReLU	(4, 4, 512)	2359808
Dropout	(4096)	0	Conv2D LReLU	(4, 4, 512)	2359808
Dense LReLU	(4096)	16781312	Max Pool2D	(2, 2, 512)	0
Leaky ReLU	(4096)	0	Flatten	(2048)	0
Dropout	(4096)	0	Dense LReLU	(4096)	8392704
Dense	(2)	8194	Dropout	(4096)	0
Output Layer	(2)	0	Dense LReLU	(4096)	16781312
			Dropout	(4096)	0
			Dense	(2)	8194
			Leaky ReLU	(4096)	0
			Output Layer	(2)	0

Table 3.3: VGGs configuration. The columns specify the layer operation, the shape of the output and the number of parameters to fit. The output shape of a layer is a 4D matrix, but, since the first dimension is the fixed size of the input data batch, we do not mention this number to prevent confusion. To avoid the weighted of the text we merged the Conv2D and Leaky ReLU layer as a one layer (as well as for Dense and Leaky ReLU) and we omitted the batch normalisation layer that always follows the Conv2D operation. The total amount of trainable parameters is larger than 45M. The last dimension of the input layer is the involved number of channels (i.e. the number of photometric bands used), a quantity depending on the specific experiment (see sections 5.2 and 6.2). Left columns refer to the network used for cluster members identification (see Chap. 5, while right columns describe the VGG adopted for the strong lensing events classification (see Chap. 6). In this latter work we also trained the net with cutout whose size is 256 pixels, in this case the last operational block)convolutions chain combined with the MaxPool2D) is repeated again before the flatten layer, always with 512 filters. The implementation of SC-VGG (used for the strong-lensing work) consists in the iteration of this VGG for three times: excluding the final probabilities averaging, the only difference is the input layer, whose dimension is $[(128, 128, 1)] \times 3$ or $[(256, 256, 1)] \times 3$.

repeated again before the flatten layer, always with 512 filters. A configuration layout is shown in Tab. 3.3.

Concerning the other hyper-parameters, at the end of all experiments we found that AdaDelta is the best-choice as optimiser (it allowed an performance improvement up to 5%). To avoid memory loss, the network has been trained with input data batches of size equals to 64 patterns for the cluster member identification, while, for what concern the strong-lenses classification it has been reduce to 32 patterns.

3.7.2 Other implemented network

Network performance can be improved by increasing its sizes, which includes both expanding the depth (i.e. the number of layers) and the width (i.e. the number of units in each level). However, enlarging the architecture dimension means an increasing of the amount of parameters, which makes the network more prone to overfitting, in particular when the training involves restricted dataset. This expansion weights down the computing and ends up to require a dramatically amount of computing resources. Moreover, if the added capacity is not sufficient, then computing resources and times are unnecessarily wasted. The fundamental way of solving these issues is to move from fully connected to sparsely connected layers, even inside convolutions (Arora et al. 2013). Inception and residual networks try solve this problem by constructing sophisticated topology algorithms that approximate sparse structures. This is achieved by building networks with repeated *inception modules* and *residual blocks*.

Inception Networks (Szegedy et al. 2014) exploit the concept of Network in Network proposed by Lin et al. (2013) which inserted a micro fully connected network within a deep architecture by abstracting data within the receptive field. This architecture is implemented as chain a *inception modules*, each of them processes the input through several parallel path characterised by different convolutional filters and sub-sampling, by concatenating the outputs. In this way, this network embeds a sort of hyper-parameter optimisation, since the most adapted filter combinations automatically produce the feature maps that allow to solve the problem.

Residual Networks (He et al. 2015a) are based on the idea that a certain network layer should be able to extract feature maps carried at least the same information (necesary to solve the problem) carried by input itself. Thus, this architecture is implemented as a sequence of *residual blocks* in which the input is processed by a set of convolution and pooling units, but it is also propagated directly to the output (though the so-called *short-cut*) and summed to the resulting extracted feature maps. An evolution of this architecture is the Residual NeXt (Xie et al. 2016), based on which the residual block perform an *aggregated transformation*: an N-dimensional feature map is split into M lower dimension (N/M) features maps, which are independently processed by the network and finally summed together and to the propagated input.

Finally, Residual Inception Networks (Szegedy et al. 2016) combine the the previous described architectures. These models are composed as a chain of *inception residual modules*, in which the large usage of different convolutional filters and subsampling is combined to the aggregated transformation and input short-cut propagation.

3.8 Benchmark methods

The results achieved by the application of CNN to the cluster member selection (Chap. 5) has been compared with two two techniques based on photometric catalogues (see Sec. 5.3.3): a Random Forest (RF, Breiman 2001) and a Bayesian Method (briefly described in Grillo et al. 2015). Here we summarise the main feature and functionality of these two methods.

A Bayesian classifier is a model able to minimise the error probability (Devroye et al. 1996), defined as: $L(g) = P[g(X) \neq Y]$, where (X, Y) are pair values $\in \mathbb{R}^d \times \{1, \ldots, M\}$ (i.e. Y is the ensemble of class labels related to the manifold X), g is a classifier (i.e. a function $g : x \in X \subseteq \mathbb{R}^d \rightarrow y \in \{1, \ldots, M\}$), L is an application mapping g into probabilities. The minimal probability error is denoted $L^* = L(g^*)$, that can be written as:

$$g^* = \underset{g:\mathbb{R}^d \rightarrow \{1,\ldots,M\}}{\operatorname{argmin}} P[g(X) \neq Y]$$

Given a classical linear model $\bar{y}_i = \sum_{j=1}^{p} x_{ij}\theta_j$, $i = 1, \ldots, n$, the method estimates $\{\theta\}_i^p$ in order to minimise a coherent combination of the residuals $r_i = y_i - \bar{y}_i$. The implemented method exploits a minimum covariance determinant method (Rousseeuw 1984), which is based on the minimisation of the median of squared residuals.

Random forest is a machine learning classifier consisting of a collection of tree-structured classifiers $\{h(x, \Theta_k), k = 1, ...\}$ where the $\{\Theta_k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x. The generalisation error for this algorithm depends on the strength of single trees and from their correlations through the raw margin functions. To improve the model accuracy by keeping trees strength, the correlation between trees is decreased and bagging with a random selection of features is adopted. Bagging, or bootstrap aggregating, is a method designed to improve the stability and accuracy of machine learning algorithms. It also reduces variance and helps to avoid overfitting. In this work, we used the RF provided by Scikit-Learn python library (Pedregosa et al. 2011).

Chapter 4

Observational data

The works presented in the following chapters are based on extensive observations of the core of several massive clusters, which provided large volumes of spectroscopic and photometric data characterised by an unprecedented quality. Particularly, imaging have been acquired with the Hubble Space telescope (HST) and with the Subaru telescope, while spectroscopy were obtained with the VIMOS multi-object panoramic spectrograph and with the integral field spectrograph MUSE (Multi Unit Spectroscopic Explorer), both of them part of Very Large Telescope (VLT) programs. In this chapter, we outline the used instrumentation and the corresponding imaging and spectroscopic observational surveys.

4.1 Imaging

4.1.1 Hubble Space Telescope imaging

Hubble Space Telescope[1] has been widely recognised as one of the most productive machine ever built. It was lauched in 1990 in a stable orbit at 600 km of height, with a revolution time around 96 minutes. It provides images with an angular resolution of 0.050″.

In this work, the used images were acquired by the Advanced Camera for Survey (ACS) and the Wide Field Camera 3 (WFC3), both of them have of a large set of broad and narrow band-pass filters, covering a spectral range from the UV ot the near-IR, whose transmission curves are shown in Fig. 4.1.

WFC3[2] consists of two working channels: the UVIS channel, which covers the ultraviolet wavelength range, approximately 2000−4000 Å, and the NIR channel, over the range 0.9−1.7 μm. In this work, we used only NIR channel. Its detector is a 1k×1k HgCdTe array, with a pixel scale of 0.13″/pixel and a FoV of 123″ × 137″. This camera replaced the WFPC2 in 2009 during the servicing mission 4.

The ACS[3] detector consists of two 2k×4k CCDs, which are butted together to create an effective FoV of 202″ × 202″, with a pixel scale of 0.049″/pixel. It covers a spectral range from 3700 Å up to 11000 Å. This camera replaced the HST Faint Objects Camera during the service mission 3B in 2002. Originally ACS had two independent working channels: Wide Field Camera (WFC) and High Resolution Camera (HRC); however after an electric fault in 2007, only the WFC was recovered during the service mission 4 in 2009.

The images were calibrated, reduced and then combined into mosaics with a pixel scale of 0.030″ and 0.065″ (see Koekemoer et al. 2007, 2011).

[1] http://www.stsci.edu/hst
[2] http://www.stsci.edu/hst/wfc3
[3] http://www.stsci.edu/hst/acs/

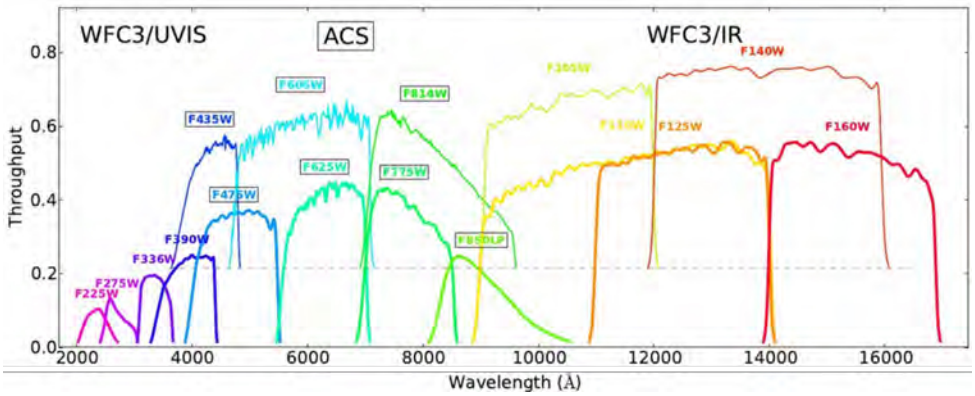


Figure 4.1: Transmission curves of HST/ACS and HST/WFC3 filters. Figure from Monna et al. (2014).

4.1.2 Subaru imaging

Subaru telescope is a ground-based instrument built in Hilo, Hawaii. Its primary mirror has a effective diameter of 8.2m made of ULE (ultra-low thermal expansion glass), with a focal length of 15m. In this work, the used images were acquired with the Subaru Prime Focus Camera (Suprime-Cam, Miyazaki et al. 2002), which consists in a mosaiac of ten 2k×4k CCDs with a pixel scale of 0.20″/pixel, covering a FoV of 34′ × 27′. The Suprime-Cam have two photometric band systems for broad-band filters: the Johnson-Morgan-Cousins system, i.e. b, v, r_c, i_c (Johnson & Morgan 1953; Cousins 1978; Bessell 1990), and the Sloan Digital Sky Survey (SDSS) system, i.e. g', r', i', z' (Fukugita et al. 1996). The transmission curves are shown in Fig. 4.2.

4.2 Spectroscopic data

The VLT (Very Large Telescope) is a telescope facility operated by the European Southern Observatory on Cerro Paranal in Chile, it consists of an array of four Ritchey-Chrétien Unit Telescopes (UT), whose primary mirrors have a diameter of 8.2m, and four auxiliary movable telescopes (AT), with mirrors of 1.8m of diameter. The UTs are equipped with an adaptive optic system, which correct for the aberrations introduced by the atmospheric turbulence. The maximum FoV of an UT is 27′. The eight VLT telescopes can work independently or in a

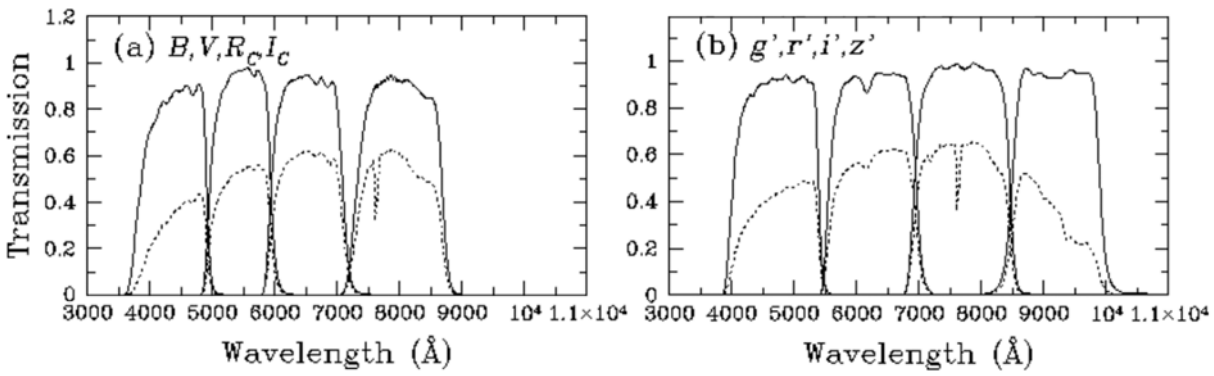


Figure 4.2: Transmission curves of Subaru Surprime-Cam standard filters (solid line). The dotted lines indicate the combined responses by considering the CCD quantum efficiency, the throughput of the prime focus corrector, the reflection of the primary mirror and atmospheric absorption. Figure from Miyazaki et al. (2002).

combination mode by summing their collecting area and acting as a single telescope of 16m of diameter.

The spectroscopic data used in this work are measured with MUSE[4] IFS (Bacon et al. 2012, 2014, 2015), located at the Nasmyth B focus of the VLT/UT4. It is a second generation instrument, which exploits the imporved spatial resolution of adaptive optics: by combining a set of image slicers and 24 spectrographs, it generates a data cube, whose FoV is $1' \times 1'$, with a pixel scale of $0.2''$/pixel (operating with the Wide Field Mode); thus a total amount of $(60/0.2)^2 = 90 \cdot 10^3$ spectra, with a wavelength range included in 4650 – 9300 Å, a spectral sampling of 1.25Å/pixel and a spectral resolution of $R(\lambda = 4650\,\text{Å}) \sim 1750$ (or $R(\lambda = 9300\,\text{Å}) \sim 3750$).

MUSE can observe with two operating modes: (*i*) WFM (Wide Field Mode), which is the mode with which observations are carried out to produce the data used in this work, and (*ii*) NFM (Narrow Field Mode). With the WFM, MUSE observes with the maximum FoV, i.e. $1' \times 1'$ sampled in pixels of $0.2'' \times 0.2''$. The whole MUSE FoV is divided in sub-fields which are sent to 24 IFUs, that collect a spectrum for each one of the $90 \cdot 10^3$ pixels; then, the MUSE data reduction software arranges these spectra in a data cube, i.e. a tree-dimensional matrix, whose first two dimensions represent the observed FoV and the spectrum lies on the third dimension. Fig. 4.3 shows a layout of the data cube (3 partially overlapped observations of the M1206 core, $z = 0.439$), obtained by extracting frames at 20 wavelengths.
With the WFM, MUSE observes in a reduced area of $7.5'' \times 7.5''$, sampled in pixels of $0.025'' \times 0.025''$.

4.3 Galaxy Clusters Programs

Data used in this thesis were collected as part of several galaxy cluster surveys: CLASH, CLASH-VLT, HFF and RELICS. These surveys provided extensive and high quality spectroscopic and photometric data for 69 galaxy clusters spanning a redshift range from $z = 0.189$ (Abell 383) to $z = 0.972$ (SPT-CLJ0615-5746). In this work, CLASH, CLASH-VLT and HFF data are used to train the convolutional neural networks, i.e. to build the Knowledge Base, while RELICS data are finally processed by the trained networks to test their achieved level of generalisation.
In this section we outline the primary scientific goals and features of these surveys.

4.3.1 CLASH survey

The Cluster Lensing And Supernova survey with Hubble (CLASH[5], Postman et al. 2012a), whose Principal Investigator (P.I.) is Marc Postman, was one of the three selected HST Multy-Cycle Treasury Programs in 2011. CLASH observed 25 massive galaxy clusters using HST panchromatic imaging (16 filters of WFC3 and ACS cameras, covering the wavelength range 2000 – 17000Å), for a total of 524 HST orbits. Furthermore, these clusters have been observed with the Advanced CCD Imaging Spectrometer of Chandra[6]. A selection of 15 CLASH clusters is shown in Fig. 4.4. The four clusters, observed by also the Subaru telescope, which are used in this thesis (and which are part of CLASH observation) are shown in Fig. 4.5. These coloured images were produced with the `Trilogy` code (Coe et al. 2012), by combining HST/WFC3 and HST/ACS filters from the optical to the near-infrared, and by combining Subaru Suprime-Cam *b*, *v*, r_c, i_c (i' for Abell 209) and z'.

[4] https://www.eso.org/sci/facilities/develop/instruments/muse.html
[5] https://archive.stsci.edu/prepds/clash/
https://www.stsci.edu/~postman/CLASH/Home.html
[6] http://cxc.harvard.edu/cal/Acis/

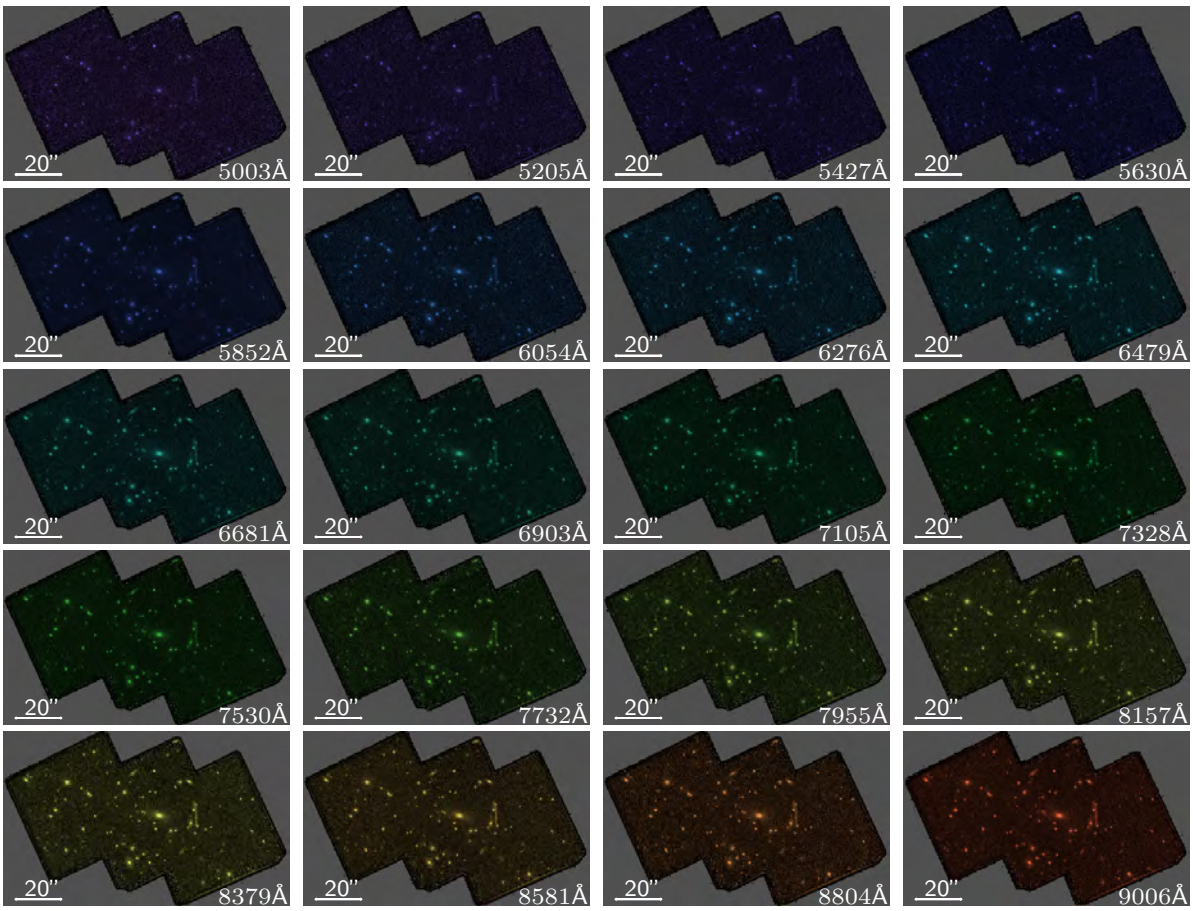


Figure 4.3: Layout of MUSE data cube (actually, 3 partially overlapped observations of the M1206 core, $z = 0.439$) represented as a set of images extracted at 20 wavelengths. The corresponding central wavelength is quoted in each panel. Images are coloured according to this wavelength, from violet to red.

CLASH program was completed in 2003, but its data products still supply the scientific community, by accomplishing its main science goals: revealing and characterising distant lensed galaxies at $z \geq 7$ (e.g., Zheng et al. 2012; Coe et al. 2013; Bouwens et al. 2014), mapping and studying the dark matter distribution in galaxy clusters (e.g., Umetsu et al. 2014; Merten et al. 2015; Sartoris et al. 2014), understanding dark energy and testing cosmological paradigms by detecting supernovae (e.g., Riess et al. 2018; Gómez-Valent & Amendola 2018), study the internal structure and the evolution of cluster galaxies (e.g., Postman et al. 2012b; Connor et al. 2017; Fogarty et al. 2017).

4.3.2 CLASH-VLT program

Built on CLASH survey, the "Dark Matter Mass Distributions of Hubble Treasury Clusters and the Foundations of ΛCDM Structure Formation Models" program (hereafter CLASH-VLT[7], Rosati et al. 2014) is an ESO-Large Program approved in 2014 (P.I. Piero Rosati), which consists of an comprehensive spectroscopic campaign on 13 CLASH clusters, accessible from the VLT. Spectroscopic observations have been carried out with the Visible wide field Imager and Multi-Object Spectrograph (VIMOS), instelled on the UT3 of the VLT. VIMOS has a spectral range of 3600 – 10000 Å and a spectral resolution between 13 Å and 28 Å. The program was completed in 2016, after 225 hours of observations, finalising its original goals: confirming at least 500

[7] https://sites.google.com/site/vltclashpublic/home

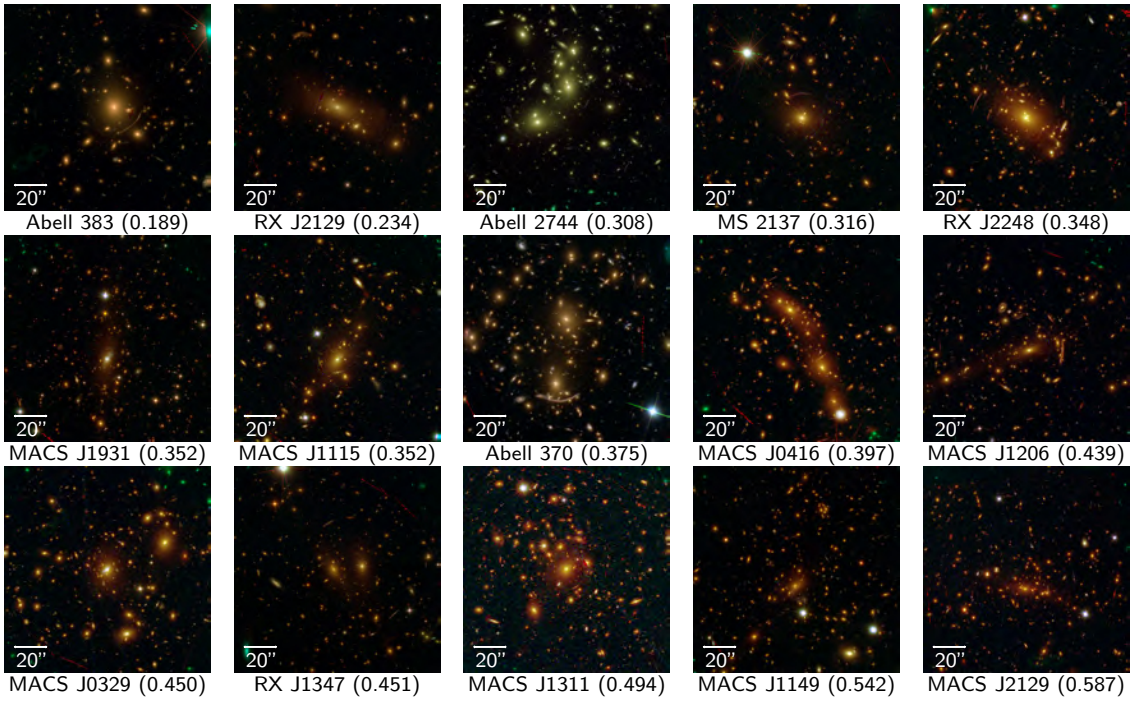


Figure 4.4: Colour-composite images of the 15 CLASH clusters included in our analysis, obtained by combining HST/WFC3 and HST/ACS filters from optical to near IR. Images are squared cut-outs, $\sim 130''$ across, centred on the cluster core. For each cluster, the central redshift is reported between brackets.

members per cluster, over $3-5$ Mpc, for accurate cluster mass reconstruction from dynamical analysis (e.g., Balestra et al. 2016; Annunziatella et al. 2017; Girardi et al. 2015; Jiménez-Teja et al. 2018); measuring redshift of over 200 magnified galaxies out to $z \sim 7$, with which constraining cluster strong lensing model (e.g., Pizzuti et al. 2016; Parry et al. 2016; Grillo et al. 2018; Bonamigo et al. 2018; Mercurio et al. 2021). An impressive example of the spectroscopic coverage of CLASH-VLT data is shown in Fig. 4.6, where spectroscopic confirmed sources are plotted on Subaru Suprime-Cam r_c band field of MACSJ0416.1-2403.

4.3.3 Hubble Frontier Fields survey

The Hubble Frotier Fields survey (HFF[8], P.I. Matt Mountain) used the HST/ACS and HST/WCF3 cameras to produce the deepest observations of the core of clusters (Lotz et al. 2014, 2017; Koekemoer et al. 2014). In particular, HFF provided ultra-deep observations (5σ point-source depth of $F814 = 29.1$ mag, i.e. 1.5 magnitudes deeper than CLASH) of six clusters (which are shown in Fig. 4.6), for a total of 840 orbits (i.e. 160 per cluster), in three ACS filters ($F435$, $F606$, $F814$) and in four WFC3 bands ($F105$, $F125$, $F140$, $F160$). The six clusters were selected for their high-magnification lensing properties, for the large set of spectroscopic confirmed multiple images systems, for the available magnification maps and for absence of bright stars in the FoV.

The primary science goals of HFF are the exploration of the high-redshift universe by revealing galaxy populations of $z = 5-10$ (e.g., Kikuchihara et al. 2020; Vanzella et al. 2021; Meyer et al. 2021); the studying of the stellar populations of faint galaxies and the characterisation of high-z star-forming galaxies (e.g., Bhatawdekar et al. 2019; Vanzella et al. 2019; Furtak et al. 2021); set the stage of the James Webb Space Telescope (e.g., Ryan & Reid 2016; Vanzella et al. 2017a).

[8] https://outerspace.stsci.edu/display/HPR/HST+Frontier+Fields

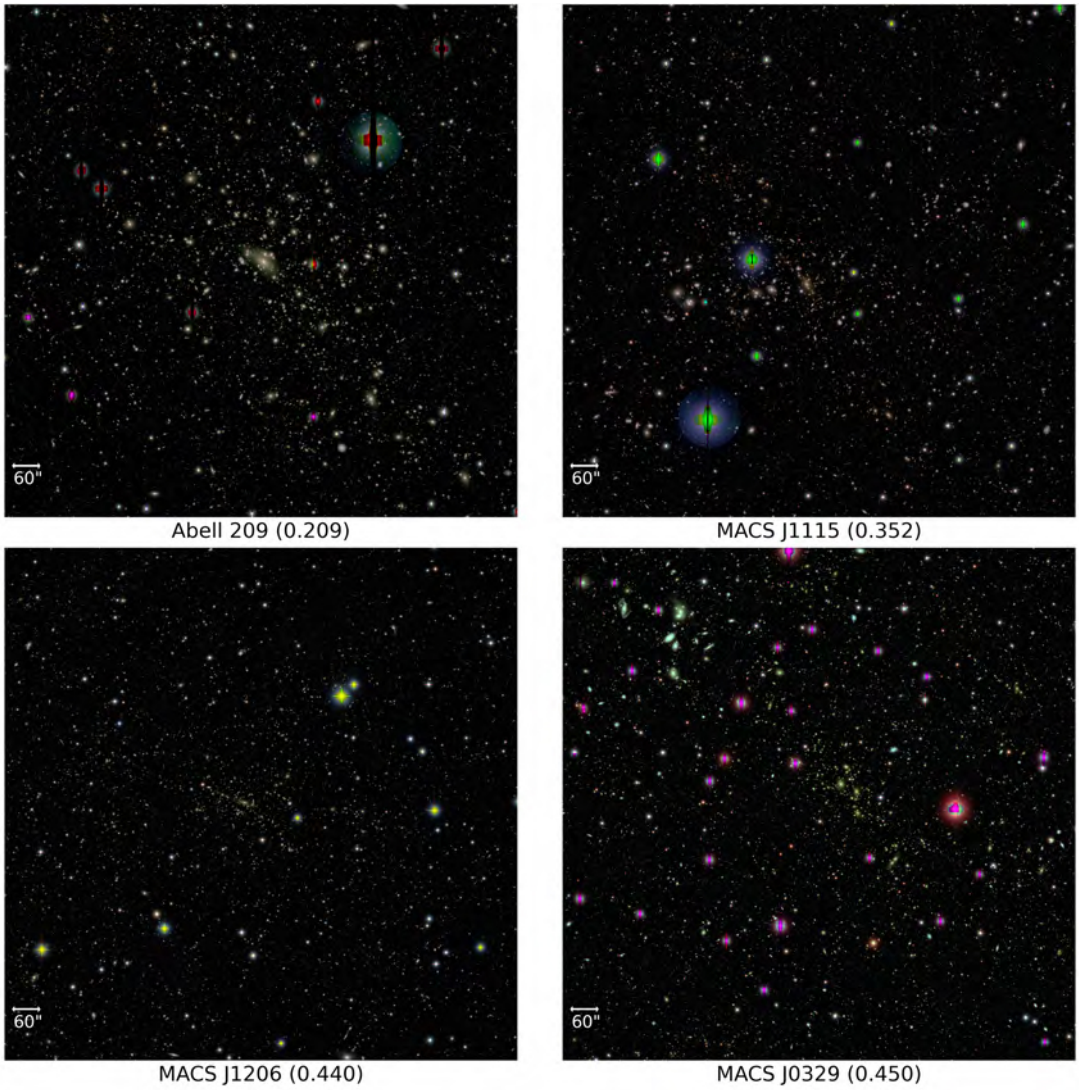


Figure 4.5: Colour-composite images of the 4 Subaru clusters, obtained by combining Subaru Suprime-Cam filters b, v, r_c, i_c (i' for Abell 209) and z'. Images are squared cutouts $\sim 20'$ across, centred on the cluster core. For each cluster, the central redshift is reported between brackets.

4.3.4 RELICS survey

The Reionization Lensing Cluster Survey (RELICS[9], P.I. Dan Coe) extends the CLASH and HFF strategy through a wider, shallower survey of 41 galaxy clusters ($z \in [0.182, 0.97]$), observed for a total of 188 HST orbits (5 per cluster), using the same 7 HFF filters, i.e. spanning a wavelength range $4000 - 17000$ Å (Coe et al. 2019). RELICS data products are less deep than CLASH and HFF (5σ point-source depth of $F160 = 26.5$ mag, i.e. 1.0 and 2.2 magnitude shallower than CLASH and HFF), however observations covered a larger area. 21 clusters have been selected among the most massive clusters known based on *Planck* PSZ2 estimations, and 20 additional clusters based on observed or suggesting exceptional lensing strength. None of them had existing HST infrared imaging. A selection of 33 RELICS clusters, used in this thesis, is shown in Fig. 4.8.

RELICS was mainly designed to: search for brightly lensed high-redshfit galaxies in the epoch of reionisation (e.g., Salmon et al. 2018, 2020; Strait et al. 2020); extend robust strong-lens model to other clusters (e.g., Acebron et al. 2018; Cerny et al. 2018); improve the precision of

[9] https://relics.stsci.edu/

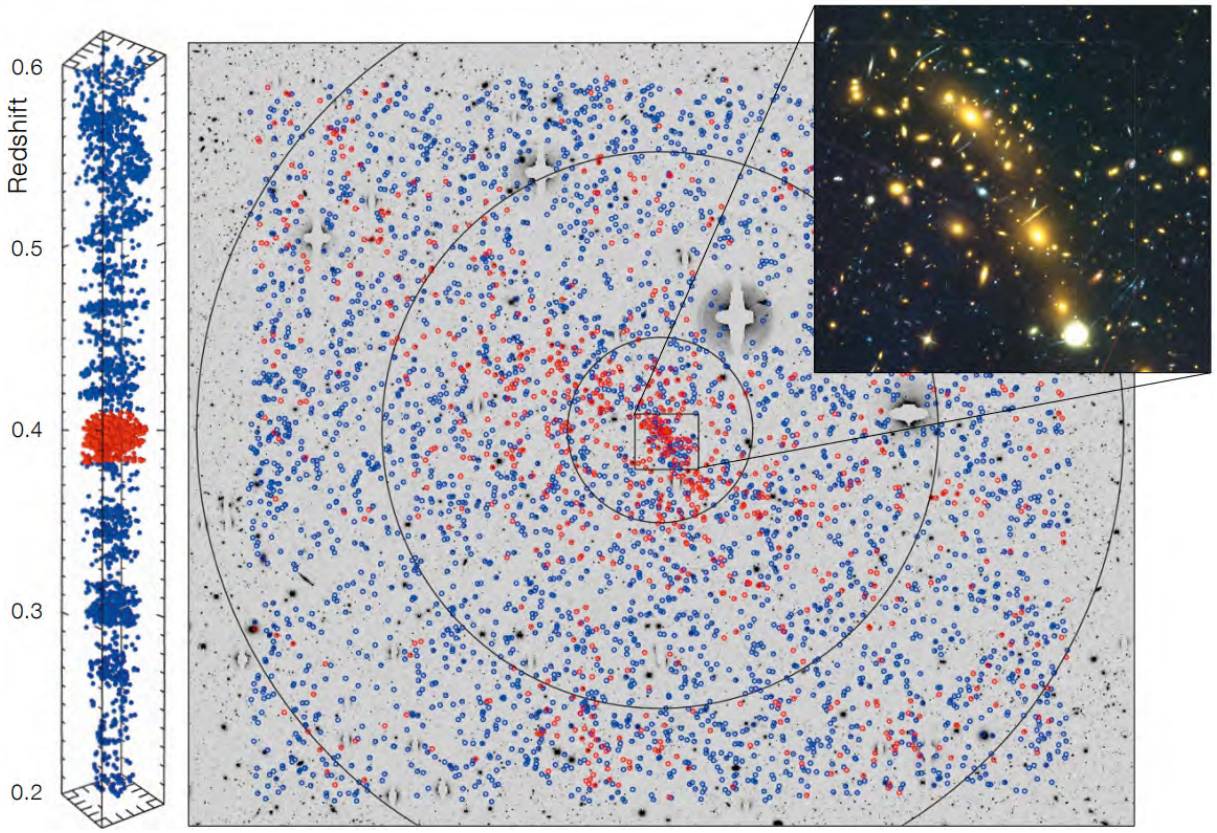


Figure 4.6: Spatial distribution of galaxies on the Subaru Suprime-Cam r_c band field of MACSJ0416.1-2403 ($29' \times 25'$). Red circles indicate the 880 confirmed cluster members (with rest-frame velocity in $\pm 3000\,\mathrm{km\,s^{-1}}$ from the cluster central redshift, $z = 0.396$), blue circles represent the other 3307 sources with $z \in (0.02, 4.15)$. Large black circles mark 1, 3 and 5 Mpc radii from the northern BCG. The HST colour composite image ($2.0' \times 1.8'$) shows the view of the core. On the left, the 3D source distribution in z (restricted in $0.2 - 0.6$). Figure from Rosati et al. (2014).

mass scaling relations and tighten limits on the dark matter particle cross section, by combing lensing, X-ray observations and SZ studies (e.g., Sayers et al. 2019).

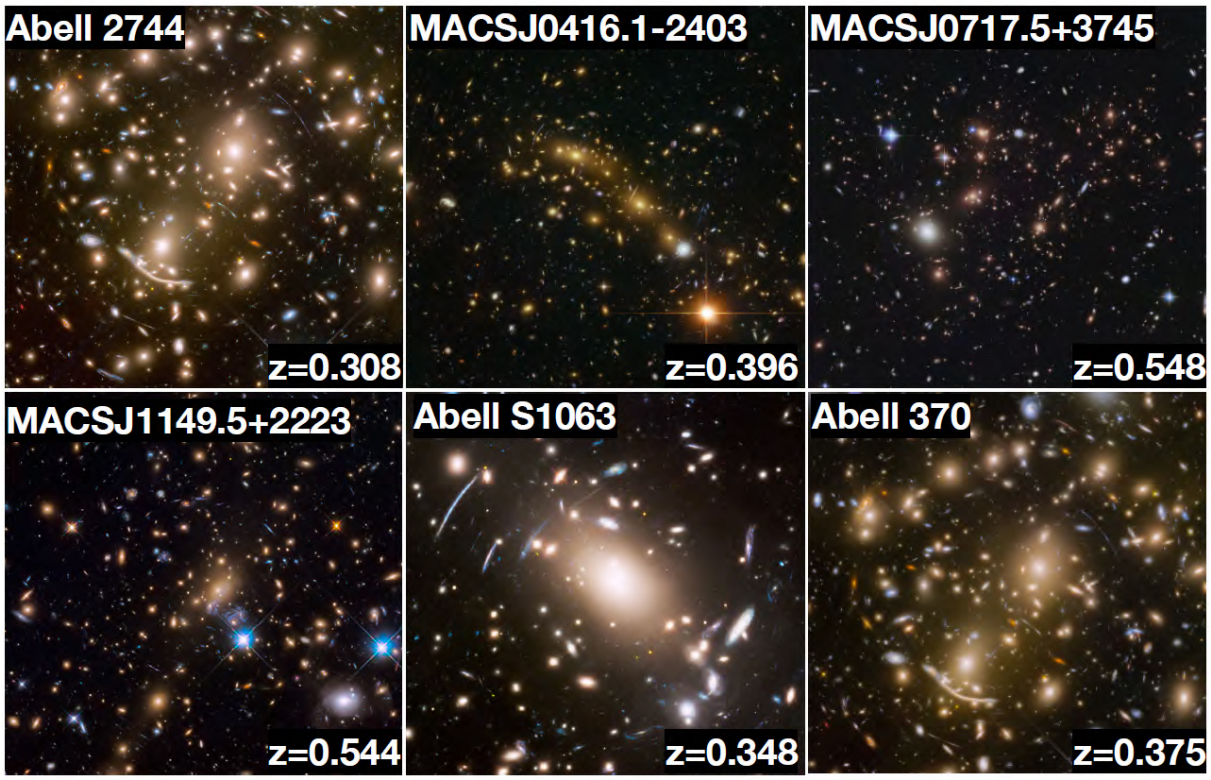


Figure 4.7: Colour composite image of the 6 HFF clusters. The cluster ID and central redshift are written in white. Credit: https://esahubble.org/.

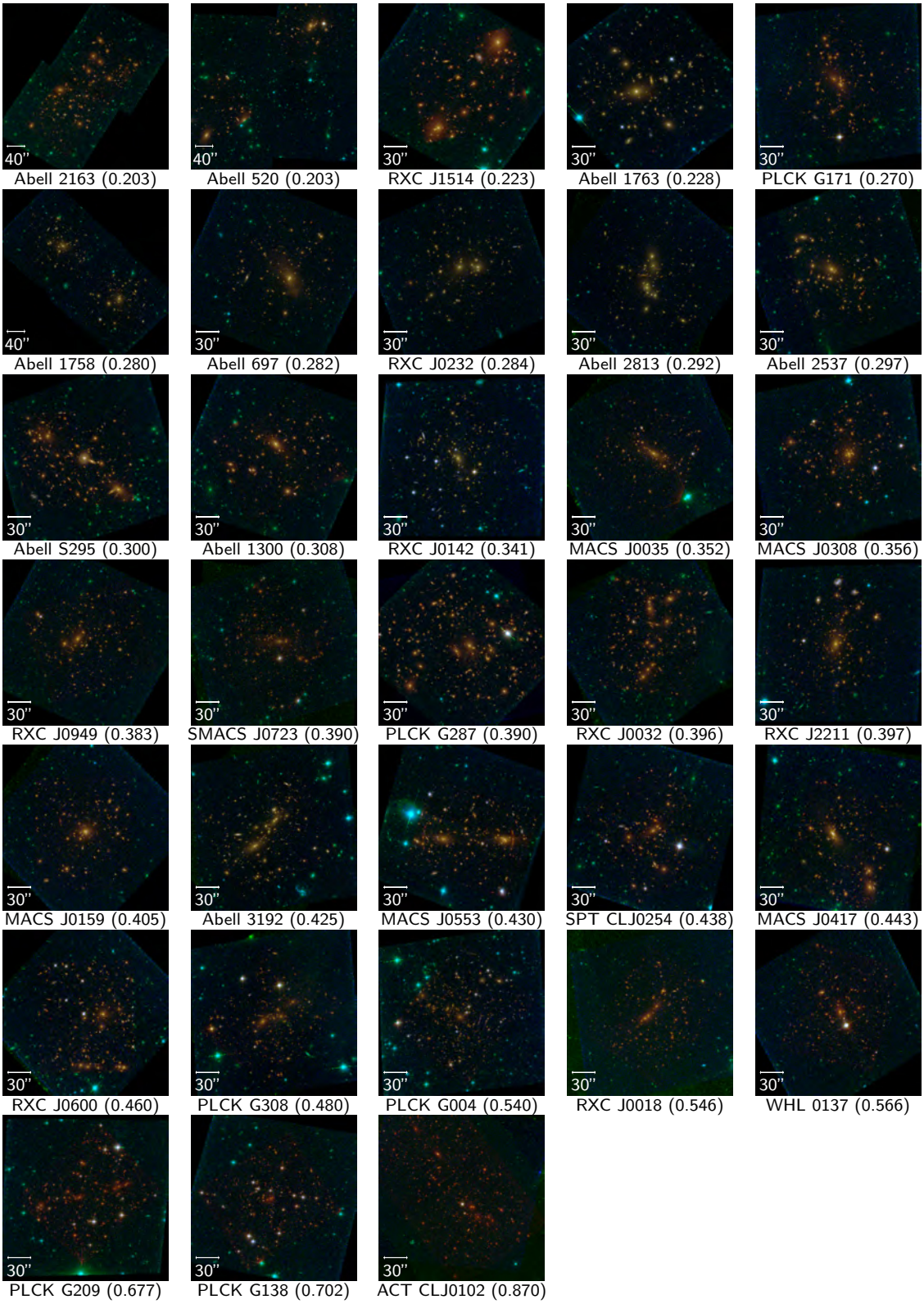


Figure 4.8: Colour-composite images of the 33 RELICS clusters used as run set, obtained by combining HST bands from optical to near IR. The images are squared cutouts, $\sim 220''$ across (excluded for Abell 2163, Abell 520 and Abell 1758, for which the cutouts are $\sim 380''$ across), centred on the cluster core. For each cluster, the central redshift is reported between brackets.

Chapter 5

Identification of Cluster Members

In this chapter I present the identification of cluster member population in galaxy clusters exploiting CNN capabilities to classify objects based on imaging data alone. As introduced in section 3.7, we test classification capabilities of four architectures, whose comparison is shown in Tab. A.1, however, since performances are strictly comparable, to avoid the weighting of the text, we restrict the following analysis to the results achieved by the application of the VGG network (see Sec. 3.7.1), i.e. the model which required the least use of computational resources, with the shortest computing time and with the smallest hyper-parameter space. This chapter is largely extracted from Angora et al. (2020).

5.1 Aim of the work

As introduced in Sec. 2, galaxy clusters are massive systems made of hundreds of galaxies bound by dark matter, such galaxies are called cluster members (hereafter CLMs). Disentangling these galaxy members from background and foreground sources is an essential step in the measurement of physical properties of galaxy clusters, e.g. the galaxy luminosity and the galaxy stellar mass functions (Annunziatella et al. 2014, 2016, 2017; Mercurio et al. 2016); furthermore, high-precision strong lensing models require the simultaneous identification of background multiple images and galaxy members in order to study the cluster mass distribution (e.g. Caminha et al. 2017b, 2019; Medezinski et al. 2016; Lagattuta et al. 2017), separate the sub-halo population from cluster projected total mass distribution (e.g. Grillo et al. 2015; Bergamini et al. 2019), test structure-formation models, cold dark matter paradigm and to constrain cosmological parameters (e.g. Diemand & Moore 2011; Meneghetti et al. 2020; Grillo et al. 2016, 2018).

Cluster members can be identified through spectroscopic measurements. However obtaining a complete sample of spectroscopic members is an expensive and time-consuming task, which can be simplified and accelerated thanks to the use of a limited amount of spectroscopic information combined with other techniques. In recent yeas, the CLM selection has been addressed in several ways: by exploiting the the members' red-sequence in colour-magnitude diagrams, aided by spectroscopic measurements (Caminha et al. 2019); by estimating photometric redshifts of sources in galaxy clusters with a Bayesian method (Molino et al. 2017, 2019) or using ML techniques (Lopes & Ribeiro 2020); by training a Multi-Layer-Perceptron with a quasi-Newton approach, used as galaxy member classifier (Biviano et al. 2013; Cavuoti et al. 2015; Brescia et al. 2013); or by fitting a multivariate normal distribution in a multi-dimensional colour space combining spectroscopic members and field galaxies (Grillo et al. 2015). Besides the necessity of spectroscopic information, all these methods require accurate photometric measurements, which are difficult to obtain with standard photometric techniques in galaxy clusters, due to the strong contamination from bright cluster galaxies, including the brightest cluster galaxies, and

the intra-cluster light (Molino et al. 2017).

In this work, we bypassed this problem by avoiding the extraction of photometry-based features (i.e. the typical information provided by software like SExtractor, e.g. magnitudes corresponding to different apertures, galaxy size measurements, such as semi-major and semi-minor axis lengths, Kron or Petrosian apertures), designing a CNN able to select cluster members exploiting only Hubble Space Telescope panchromatic images combined with a large spectroscopic coverage. We studied the network dependence on the involved filters, on the member redshift, magnitude and colours, on the training setup and sizes, on the cutout crowding; we also performed a comparison with two different photometric approaches (i.e. a Bayesian method and a Random Forest); we exploited the trained network to identify new candidate members in galaxy clusters and test this classification by estimating the cumulative projected number of cluster members and the differential number density profiles. Finally, we also trained the CNN using the SUBARU imaging, by performing a comparison between ground-based wide-field data and images acquired by space-born telescope with smaller FoV.

5.2 Knowledge Base

In order to build the Knowledge Base suitable to address the CLM identification, we used the spectroscopic information based on CLASH-VLT VIMOS program (ESO 200h Large Program 186.A-0798, "Dark Matter Mass Distributions of Hubble Treasury Clusters and the Foundations of ΛCDM Structure Formation Models", PI: P. Rosati; Rosati et al. 2014), combined with archival observations carried out with the MUSE spectrograph (Bacon et al. 2014). We exploit such measurements to label the training set, by defining as CLM a source having rest-frame velocity separation $|v| \leq 3000\,\mathrm{km\,s^{-1}}$, respect to the cluster central velocity (Grillo et al. 2015; Caminha et al. 2016, 2017a). On the contrary, non-cluster-members (NCLMs) are those having greater differences in velocity. Formally:

$$y_i = \begin{cases} 1, & \text{if} \quad |v| = c\,|z_i - z_{cl}|/(1+z_{cl}) \leq 3000\,\mathrm{km\,s^{-1}} \\ 0, & \text{otherwise} \end{cases} \tag{5.1}$$

where y_i is the label related to the i-est source (CLM: 1, NCLM: 0), c is the speed of light (in $\mathrm{km\,s^{-1}}$), z_i and z_{cl} are the source and the cluster redshift, respectively. In this work, we built the KB by stacking all members belonging to clusters spanning a reshift range $0.18 - 0.59$. The 15 involved clusters are shown in Fig. 4.4 in Chap. 4. Images were acquired by the HST ACS and WFC3 cameras. To build the dataset, we opted for images with spatial resolution of $0.065''$. Among the 16 available HST filters, we considered bands covering the spectral range $4000\,\text{Å} - 16000\,\text{Å}$, that is, the optical and NIR bands, excluding the UV filters for which the signal-to-noise ratio (S/N) of faint CLMs was too low. This results into 12 available bands (7 optical bands and 5 NIR bands). As detailed below, we test 3 different filter combinations: (*i*) including all the available bands, (*ii*) using only the optical filters, (*iii*) selecting 3 optical and 2 near-infrared bands (i.e. the HFF filters). These configurations are summarised in Tab. 5.1. For each spectroscopic source within the HST images (excluding the BCGs), we extracted a squared cut-out with a side of $\sim 4''$ (64 pixels), centered on the source position. This size is large enough to include most of the member light as suggested by the galaxy effective radii estimated by Tortorelli et al. (2018) in RX J2248-4431[1] ($z = 0.346$), who found a value in the range $(0.08'', 1.16'')$, with a median of $0.31''$. A selection of the sources composing the dataset is shown in Fig. 5.1, where objects were extracted from the field of view of four clusters:

[1] Also known as Abell S1063

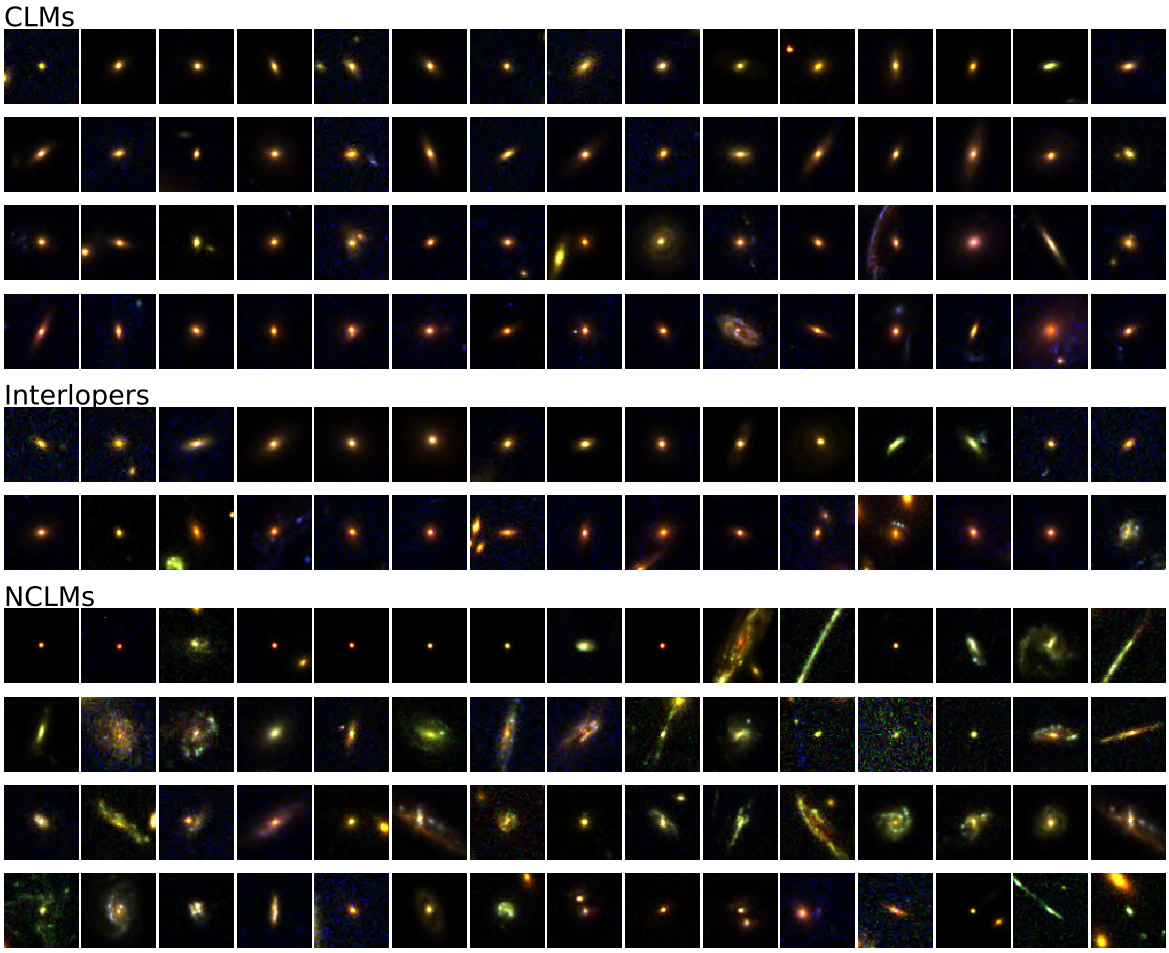


Figure 5.1: Examples of RGB cutouts of cluster members, interlopers and non-members extracted from HST images ($F435$, $F606$, $F814$). To emphasise fainter sources, images have been stretched by clipping values within $\pm 3\sigma$ and then normalised. Cutouts are $\sim 4''$ across.

RX J2248-4431 (R2248, $z = 0.346$), MACS J0416-2403 (M0416, $z = 0.397$), MACS J1206-0847 (M1206, $z = 0.439$), and MACS J1149+2223 (M1149, $z = 0.542$). n

We did not apply any magnitude or colour thresholds, nor did we use photometric information to train the network, at least non directly, indeed, magnitude is encoded within the cutouts presented to the network, whereas colour-like features are automatically computed in the first convolutional layer by algebraically summing the involved bands (see Eq. 3.3); the photometry information together with the cluster rest-frame velocity separation are shown in Fig. 5.2. Given the member red-sequence dependence on cluster redshift, we exploited the correlation between the Balmer break and the normalised colour (a.k.a. corrected colour or differential colour), as shown in Girardi et al. (2015), defined as the difference between the observed colour and the colour-magnitude relation, that is:

$$(F814 - F160)_{\rm norm} = (F814 - F160)_{\rm obs} - [\text{colour-magnitude}(F814)] \tag{5.2}$$

the colour-magnitude relation was fitted for each cluster with spectroscopic confirmed members, using a robust linear regression (Cappellari et al. 2013), that takes into account a possible intrinsic data scatter and clips outliers, adopting the least trimmed squares technique (Rousseeuw & Driessen 2006). In this way, the red-sequence is centered around the zero for members belonging to clusters at different redshift (as shown in the bottom left panel of Fig. 5.2); furthermore, it is possible to define a global limit to separate bluer and redder members, regardless of the specific

cluster red-sequence. Avoiding photometry thresholds allows us to: (*i*) increase the KB, (*ii*) neglect any kind of photometry measurements and (*iii*) introduce adversarial examples to prevent overfitting (see section 3.6.3). Particularly, this latter condition is achieved by "contaminating" the KB by also including small fraction of faint and blue members, together with *interlopers*, as shown in the panels of Fig. 5.2:

- faint members: $\lesssim 10\%$ ($\lesssim 4\%$) of CLMs have $F814 \geq 24$ mag (≥ 25 mag);
- blue members: $\lesssim 20\%$ ($\lesssim 6\%$) of CLMs have $(F606 - F814)_{\rm norm} \leq -0.15$ mag (≤ -0.5 mag);
- interlopers: typically red early-type galaxies, photometrically indistinguishable from CLMs (see also Fig. 5.1), weakly bound to the cluster, $\lesssim 23\%$ of NCLMs have $|v| \in (3000, 6000)\,{\rm km\,s^{-1}}$.

We found that, in absence of these adversarial sources, networks are quickly prone to overfitting and lose any generalisation capability. On the other hand, these are the most challenging cases for the network and they affect the metrics evaluated on the test set, enlarging the number of False Positives and False Negatives (as discussed in Sec.5.4). However, the benefits outweighs the costs: given the limited number of sources, these adversarial examples make the network able to efficiently predict the membership of sources extracted from other clusters (e.g. RELICS clusters, Sec. 5.5.1), which does not happen for the same network trained on dataset from which these objects have been, even partially, excluded.

Due to different pointing strategies and to the fields of view of HST ACS and WFC3 cameras, many sources do not have a complete photometric coverage, especially in the IR range. As a result, these objects with missing information were not useful for the training process (Batista & Monard 2003; Marlin 2008; Parker 2010). With the aim of maximising the number of training samples and, at the same time, of researching the optimal filter combination, we have chosen four different band configurations:

- *ACS*: only the seven optical bands (i.e. $F435$, $F475$, $F606$, $F625$, $F775$, $F814$, $F850$) were included in the training set, obtaining 1603 CLMs and 1899 NCLMs;
- *ALL*: the training set involved all twelve bands (i.e. the seven optical bands and the five IR bands $F105$, $F110$, $F125$, $F140$, $F160$), thus reducing the number of objects to 1156 and 1425, respectively for CLMs and NCLMs, due to the rejection of missing data;
- *Mixed*: we selected five bands, corresponding to the filters available in the Hubble Frontier Fields survey, covering the optical-IR range, namely, $F435$, $F606$, $F814$, $F105$, $F140$, respectively. This includes 1249 CLMs and 1571 NCLMs;
- *Mixed**: same band combination as in the previous case (*mixed*), but including two further clusters, namely, Abell 2744 (A2744) and Abell 370 (A370), for which only HFF imaging were available. This set is composed of 1629 CLMs and 2161 NCLMs.

In practice, the three configurations, *ACS*, *ALL* and *mixed*, share the same clusters, while exploring different spectral information by varying the number of sources. The *mixed** configuration considers an augmented cluster data set by including additional spectroscopic members. A summary of the cluster sample and the spectroscopic data sets is given in Tab. 5.1.

As explained in Sec. 3.6.5, we opted for a stratified k-fold strategy to handle the split of dataset, preceded by a sample extraction to build the validation set. In particular, we extracted 10% of samples reserved for the validation phase, the number of folds has been set to $k = 10$,

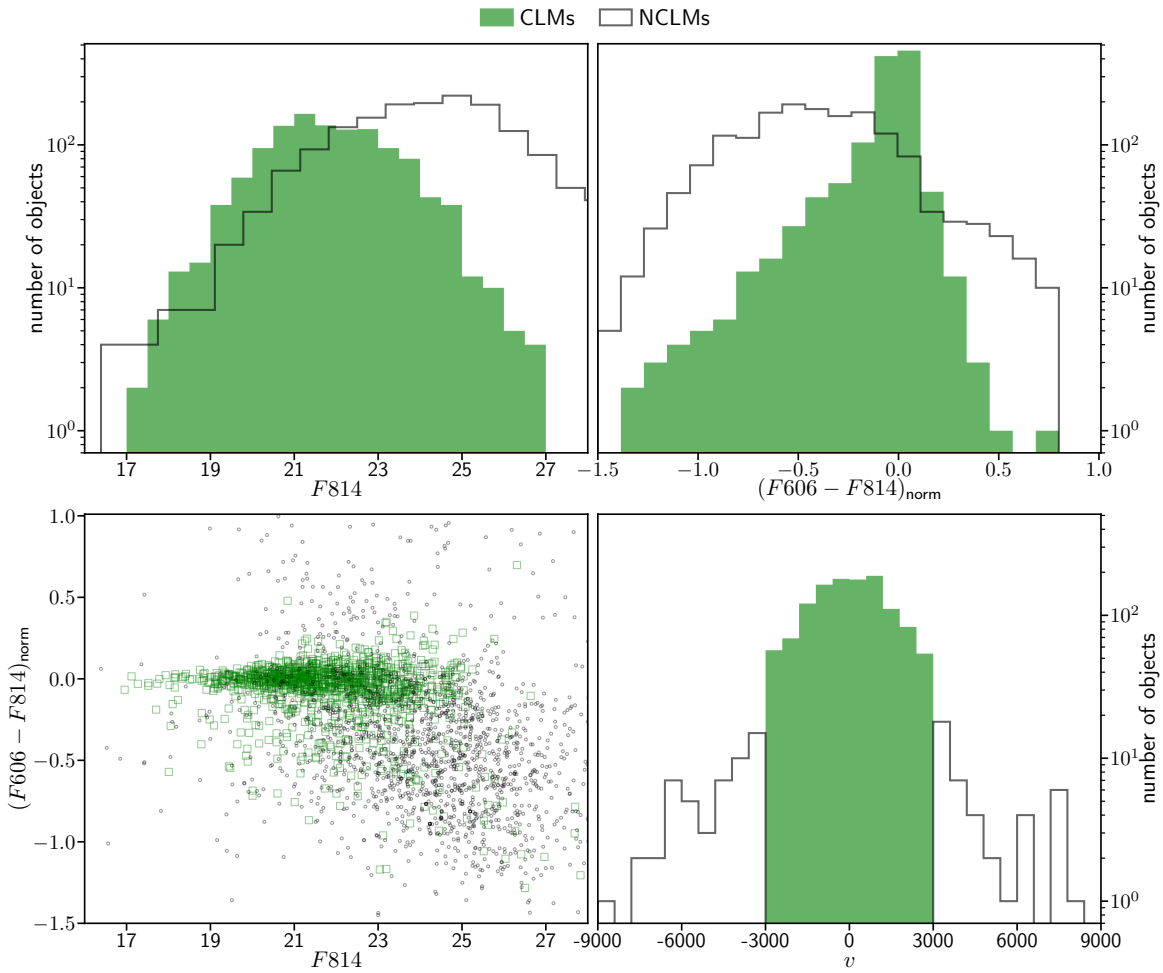


Figure 5.2: Knowledge Base layout for CLMs (coloured in green) and NCLMs (coloured in grey): magnitude and colour logarithmic distributions are shown in the upper panels, the colour-magnitude relation is plotted in bottom left panel, while the cluster velocity separation logarithmic distribution is shown in the bottom right panel (constrained within ±9000 km s^{-1}).

the 15%[2] of each training fold has been augmented through rotations and flips, resulting into an augmentation factor equals to 1.75. We studied fluctuations of performances by evaluating the results for each fold, while we measured the network global classification capability by stacking the all test folds. Obviously, the test set does not contain augmented objects. Moreover, the validation and the k-fold splitting are applied independently for each cluster involved in the training phase, in this way, samples are extracted for each cluster FoV, ensuring that the k-est fold is populated by objects extracted from each cluster proportionally to the number of available spectroscopic sources, that is, providing adequate coverage of the training set respect to the test set. The statistical estimators adopted to estimate network performances are listed in Sec. 3.6.6. In this work, we are interested to find the best compromise between purity and completeness. Unless otherwise specified, we assume a membership probability threshold of 0.50.

[2]This percentage of sources to augment has been empirically determined by measuring the network performance and its propensity to overfit as a function of the augmented sources; we found that this fraction represents a good trade-off, balancing the need to increase the training set size avoiding the network overfitting.

					*mixed** (*mixed*)		*ACS*		*ALL*		
Cluster		$z_{cluster}$	z_{min}	z_{max}	CLMs	NCLMs	CLMs	NCLMs	CLMs	NCLMs	ref
Abell 383	A383	0.188	0.176	0.200	59	51	91	79	59	51	(1, 2)
RX J2129+0005	R2129	0.234	0.222	0.246	47	124	66	132	40	118	(3, 1)
Abell 2744	A2744	0.308	0.288	0.331	156[(a)]	279[(a)]	only frontier-field bands				(4, 1)
MS 2137-2353	MS2137	0.316	0.303	0.329	45	49	70	80	45	49	(3, 1)
RX J2248-4431[(b)]	R2248	0.346	0.332	0.359	131	112	203	166	117	86	(5, 1)
MACS J1931-2635	M1931	0.352	0.338	0.365	68	97	80	110	65	96	(3, 1)
MACS 1115+0129	M1115	0.352	0.338	0.365	78	69	116	111	62	55	(3, 1)
Abell 370	A370	0.375	0.361	0.389	224[(a)]	311[(a)]	only frontier-field bands				(6, 1)
MACS J0416-2403	M0416	0.397	0.382	0.410	237	277	266	287	227	230	(7, 8, 9, 1)
MACS J1206-0847	M1206	0.439	0.425	0.454	172	216	226	242	149	203	(10, 1)
MACS J0329-0211	M0329	0.450	0.435	0.464	74	76	104	104	66	73	(3, 1)
RX J1347-1145	R1347	0.451	0.438	0.467	56	107	71	120	56	107	(3, 1)
MACS J1311-0310	M1311	0.494	0.477	0.507	52	54	69	95	52	54	(3, 1)
MACS J1149+2223	M1149	0.542	0.527	0.558	141	237	149	270	129	202	(11, 12, 1)
MACS J2129-0741	M2129	0.587	0.571	0.603	89	102	92	103	89	101	(1, 3)
	TOTAL				1629 (1249)	2161 (1571)	1603	1899	1156	1425	

Table 5.1: Cluster sample description. The name and short name of the clusters, their redshift and their spectroscopic range to identify CLMs (i.e. $|v| < 3000\,\mathrm{km\,s^{-1}}$, as state in Eq. 5.1) are reported in the first 5 columns. The four band configurations, described in Sec. 5.2, are listed in columns 6 to 11. The references for each cluster can be found in the last column.
[(a)] Different spectroscopic data sets are described in the text. The case *mixed* is similar to the *mixed** one, with the only difference that it does not include the two clusters A2744 and A370. Numbers between brackets in the bottom row refer to the the *mixed* configuration.
[(b)] The cluster RX J2248.7−4431 is also known as Abell S1063.
(1) Rosati & Clash-VLT Team 2020; (2) Monna et al. 2015; (3) Caminha et al. 2019; (4) Mahler et al. 2018; (5) Caminha et al. 2016; (6) Lagattuta et al. 2019; (7) Grillo et al. 2015; (8) Balestra et al. 2016; (9) Caminha et al. 2017a; (10) Caminha et al. 2017b; (11) Grillo et al. 2016; (12) Treu et al. 2016.

5.3 Experiments

In this section, we describe several experiments designed to test and compare CNN performances. Specifically, with the band configurations described in the previous section (see also Tab. 5.1), we performed the following tests or experiments:

- *EXP1*: efficiency of the DL approach by stacking the data of all the clusters in terms of:
  - *EXP1a*: global evaluation;
  - *EXP1b*: redshift-dependence, namely separating CLMs into redshift bins;
- *EXP2*: magnitude or colour dependence, by stacking data of a group of three clusters and varying their redshift range through:
  - *EXP2a*: separating bright and faint sources;
  - *EXP2b*: separating red and blue galaxies;
- *EXP3*: a comparison of performances of our image-based CNN technique with other approaches, based on photometric measurements of field and cluster galaxies;

Practically, with the *EXP1* we evaluated CNN global performances embracing a traditional approach based on a training-test split applied on the whole member population; while in the *EXP2* we excluded three clusters (A370, MS2137 and M0329) from the training set, which act as as blind test clusters performing a sort of "stress test"; finally, with the *EXP3* we compare CNN classification capabilities with two photometric-based techniques.

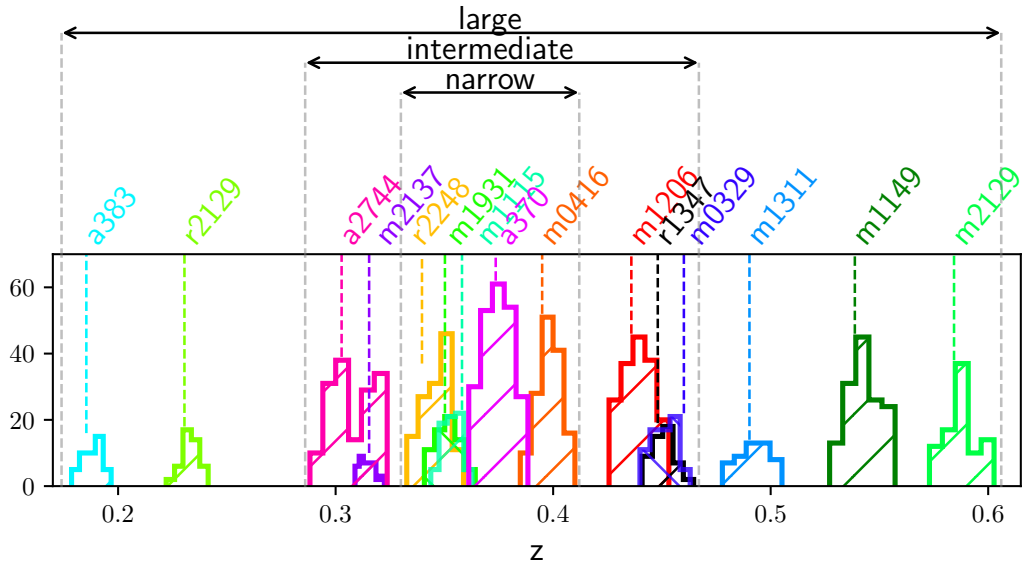


Figure 5.3: Redshift distribution of 1629 spectroscopic members used for the *EXP2* configuration. The three clusters A370 (z = 0.375, 224 CLMs), MS2137 (z = 0.316, 45 CLMs) and M0329 (z = 0.450, 74 CLMs) are used as blind test set.

Class	*%*	*mixed*	*ACS*	*ALL*	*mixed**
	AE	86.7	87.4	87.7	**89.3**
CLM	*pur*	83.1	85.0	86.4	**88.3**
	compl	88.4	**88.5**	86.4	86.7
	F1	85.6	86.7	86.4	**87.4**
NCLM	*pur*	**90.0**	89.9	88.9	**90.0**
	compl	85.5	86.7	88.9	**91.2**
	F1	87.7	88.3	88.9	**90.6**

Table 5.2: CNN percentage performances in the *EXP1* experiment. The performances are related to the four band configurations (see Sec. 5.2) and expressed in terms of the statistical estimators described in Sec. 3.6.6. The overall best results are highlighted in bold.

5.3.1 *EXP1*: Combination of all clusters

At the first stage, we evaluated the global efficiency of a DL approach including all the available clusters, regardless of their redshift (ranging between 0.2 and 0.6), by exploring different combinations of photometric bands (as described in Sec. 5.2) and assembling the data set by stacking the information from all the images extracted from our cluster sample. We wanted to verify that DL models, given their intrinsic generalisation capabilities, were able to learn how to disentangle cluster members from non-member (foreground or background) sources, independently from the cluster redshift (*EXP1a*). This although their members have different characteristics, such as apparent magnitudes or sizes, and also different signal-to-noise ratio at a fixed apparent magnitude, due to the different image depths. The results are shown in Fig. 5.4 and Table 5.2, as a function of the band configuration (described in Sec. 5.2), while performances for each cluster are presented in Tab. 5.3.

Concerning NCLMs, we found similar values of the average efficiency (87% – 89%), the purity (stable around ∼ 90%) and the F1-score (with variations within 1.5%), regardless of band

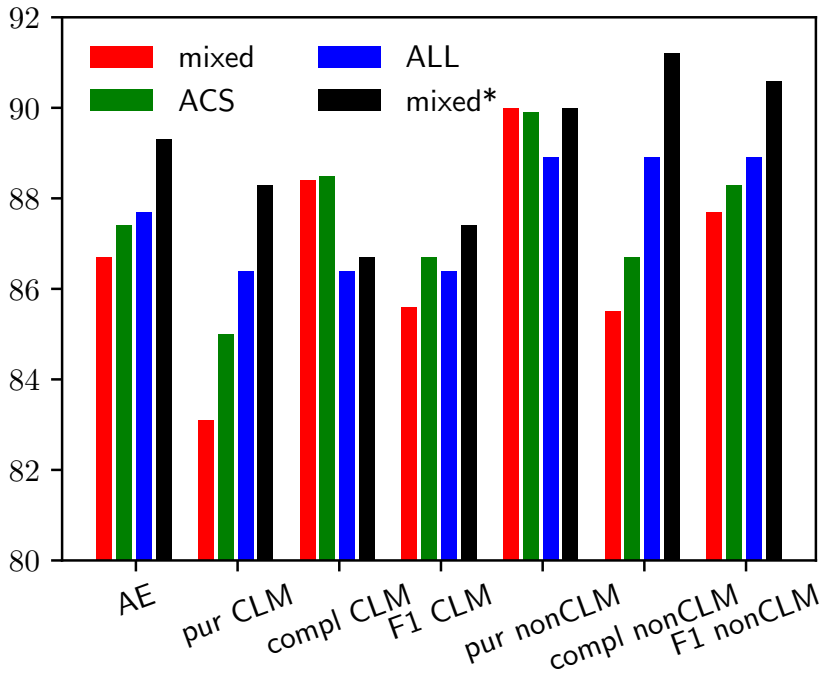


Figure 5.4: Performance percentages of the CNN in the *EXP1* experiment with the four band configurations (see Sec. 5.2) in terms of the statistical estimators described in Sec. 3.6.6

configuration. On the other hand, the CLM identification was, in general, characterised by larger variation (83% – 91%) in the statistical estimators. By exploring performances achieved for each cluster (see Tab. 5.3), the CNN trained with the *mixed* and *ACS* configurations tend to be more complete than pure for the CLM classification, while the CNN appear to be more balanced between purity and completeness when trained on *ALL* and *mixed** band configurations.

With the *mixed** configuration, CNN achieved the best performances for CLM and it was also very stable in terms of NCLM, reaching an overall efficiency of $\sim$ 89%. This analysis is confirmed by also analysing the results achieved for each cluster (see Tab. 5.3): CNN trained with the *mixed** configuration covers $\sim$ 42% of the best performances (followed by the 30% covered by *ACS*), moreover, even when it is not the best configuration, by averaging the cross-compared couples of the same estimator, we measured a difference with the corresponding best configuration of 2.0 ± 0.3%, where differences of more than 5% occur in 9% of cases. On the other hand, by comparing the CNN results achieved with the *mixed** configuration with the second best configuration, we found an average difference of 0.3 ± 0.5, implying that the CNN trained on the *mixed** dataset has performances comparable with the best combination of all the other configurations; whereas by repeating this reckoning for the other configurations, we measure an average differences of 4.3 ± 0.5%, 3.1 ± 0.5%, 3.4 ± 0.5%, respectively for *mixed*, *ACS* and *ALL* band configuration.

These performance discrepancies can be due to the different sizes of the datasets or to the peculiar filter combinations. In order to disentangle the dependence on these two terms, we performed an additional experiment using the common sample of sources (i.e. the *ALL* dataset), exploring how the band combination affects the results. The statistical estimators related to this experiment are listed in Tab. A.2. We found that all the cross-differences are strictly comparable within 2%, with a maximum efficiency of 87.7% and a maximum CLM F1-score of 86.4%, suggesting that the dimension of dataset (i.e. the sampling and the covering of the parameter space) contributes more than the combination of filters, at least for the band configurations considered in this work. This behaviour can be also deduced from the performance increasing between the *mixed* and *mixed** configuration; indeed, these two datasets share the same samples

		A383 z = 0.188				R2129 z = 0.234				A2744 z = 0.308			
		mixed	ACS	ALL	mixed*	mixed	ACS	ALL	mixed*	mixed	ACS	ALL	mixed*
	AE	77.0	81.8	78.3	**83.0**	89.7	91.6	**93.7**	92.3				93.6
	pur	77.2	82.9	82.5	**86.3**	76.5	84.6	**86.5**	84.4				95.3
CLM	compl	81.5	**82.9**	75.0	81.5	90.7	**91.7**	88.9	88.4				86.5
	F1	79.3	82.9	78.6	**83.8**	83.0	**88.0**	87.7	86.4	only mixed*			90.7
	pur	76.7	**80.6**	74.4	79.6	**96.2**	95.6	**96.2**	95.5				92.8
NCLM	compl	71.7	80.6	82.1	**84.8**	89.3	91.6	**95.3**	93.8				97.6
	F1	74.2	80.6	78.0	**82.1**	92.6	93.6	**95.8**	94.6				95.2
		MS2137 z = 0.316				R2248 z = 0.346				M1931 z = 0.352			
		mixed	ACS	ALL	mixed*	mixed	ACS	ALL	mixed*	mixed	ACS	ALL	mixed*
	AE	83.7	81.5	88.2	**88.4**	89.5	86.5	**90.2**	88.1	84.0	86.0	84.9	**90.0**
	pur	80.0	79.7	85.7	**89.7**	88.6	85.2	**90.7**	88.3	91.3	85.3	83.6	**100.0**
CLM	compl	87.8	81.0	**90.9**	85.4	92.4	91.3	**92.5**	89.8	67.7	**80.6**	78.0	75.8
	F1	83.7	80.3	**88.2**	87.5	90.5	88.1	**91.6**	89.1	77.8	82.9	80.7	**86.2**
	pur	87.8	83.1	**90.9**	87.2	**90.6**	88.3	89.5	87.9	80.8	**86.4**	85.7	85.4
NCLM	compl	80.0	81.9	85.7	**91.1**	86.1	80.7	**87.2**	86.1	95.5	89.9	89.7	**100.0**
	F1	83.7	82.5	88.2	**89.1**	**88.3**	84.3	88.3	87.0	87.5	88.1	87.6	**92.1**
		M1115 z = 0.352				A370 z = 0.375				M0416 z = 0.397			
		mixed	ACS	ALL	mixed*	mixed	ACS	ALL	mixed*	mixed	ACS	ALL	mixed*
	AE	88.1	84.9	89.6	**92.5**				88.9	90.3	90.0	91.5	**92.2**
	pur	85.7	82.5	90.9	**91.8**				85.8	92.4	90.3	**95.7**	93.3
CLM	compl	93.0	89.5	89.3	**94.4**				87.6	87.1	**88.8**	86.8	87.1
	F1	89.2	85.8	90.1	**93.1**	only mixed*			86.7	89.7	89.5	91.0	**91.5**
	pur	91.2	87.9	88.2	**93.4**				89.5	88.6	89.7	88.1	**91.7**
NCLM	compl	82.5	80.0	90.0	**90.5**				87.9	93.3	91.1	96.1	**96.9**
	F1	86.7	83.8	89.1	**91.9**				88.6	90.9	90.4	91.9	**94.3**
		M1206 z = 0.439				M0329 z = 0.450				R1347 z = 0.451			
		mixed	ACS	ALL	mixed*	mixed	ACS	ALL	mixed*	mixed	ACS	ALL	mixed*
	AE	87.7	**90.3**	87.4	89.7	81.6	81.9	83.3	**85.0**	**91.2**	90.7	89.7	89.9
	pur	83.7	89.8	84.2	**89.9**	76.9	76.8	79.1	**83.3**	79.7	**81.6**	80.4	81.0
CLM	compl	89.7	**90.2**	86.7	86.5	89.6	**91.5**	88.3	91.0	**100.0**	96.9	93.8	92.2
	F1	86.6	**90.0**	85.4	88.2	82.8	83.5	83.5	**87.1**	**88.7**	88.6	86.5	86.2
	pur	**91.3**	90.8	89.9	89.6	87.9	89.5	88.1	**90.0**	**100.0**	97.9	96.2	95.6
NCLM	compl	86.2	90.4	88.0	**92.3**	73.9	72.3	**78.8**	78.3	86.6	87.0	87.5	**88.7**
	F1	88.7	90.6	88.9	**90.9**	80.3	80.0	83.2	**83.7**	**92.8**	92.2	91.7	92.0
		M1311 z = 0.494				M1149 z = 0.542				M2129 z = 0.587			
		mixed	ACS	ALL	mixed*	mixed	ACS	ALL	mixed*	mixed	ACS	ALL	mixed*
	AE	77.1	**82.5**	75.8	78.1	85.9	**90.7**	88.0	89.4	85.5	**86.4**	84.9	86.1
	pur	72.7	**80.3**	75.0	76.0	74.5	**83.3**	80.5	82.3	85.9	87.3	91.0	**91.3**
CLM	compl	**85.1**	77.8	78.3	80.9	**94.5**	92.6	91.5	91.3	82.7	**83.1**	75.3	77.8
	F1	78.4	**79.0**	76.6	78.4	83.3	**87.7**	85.6	86.6	84.3	**85.2**	82.4	84.0
	pur	82.9	**84.1**	76.7	80.4	**96.1**	95.6	94.0	94.5	85.3	**85.6**	81.0	82.7
NCLM	compl	69.4	**86.0**	73.3	75.5	80.8	**89.7**	85.7	88.3	88.0	89.2	93.4	**93.5**
	F1	75.6	**85.1**	75.0	77.9	87.8	**92.6**	89.7	91.3	86.6	87.4	86.7	**87.8**

Table 5.3: CNN percentage performances evaluated for each cluster and for each band configuration related to the *EXP1* experiment.

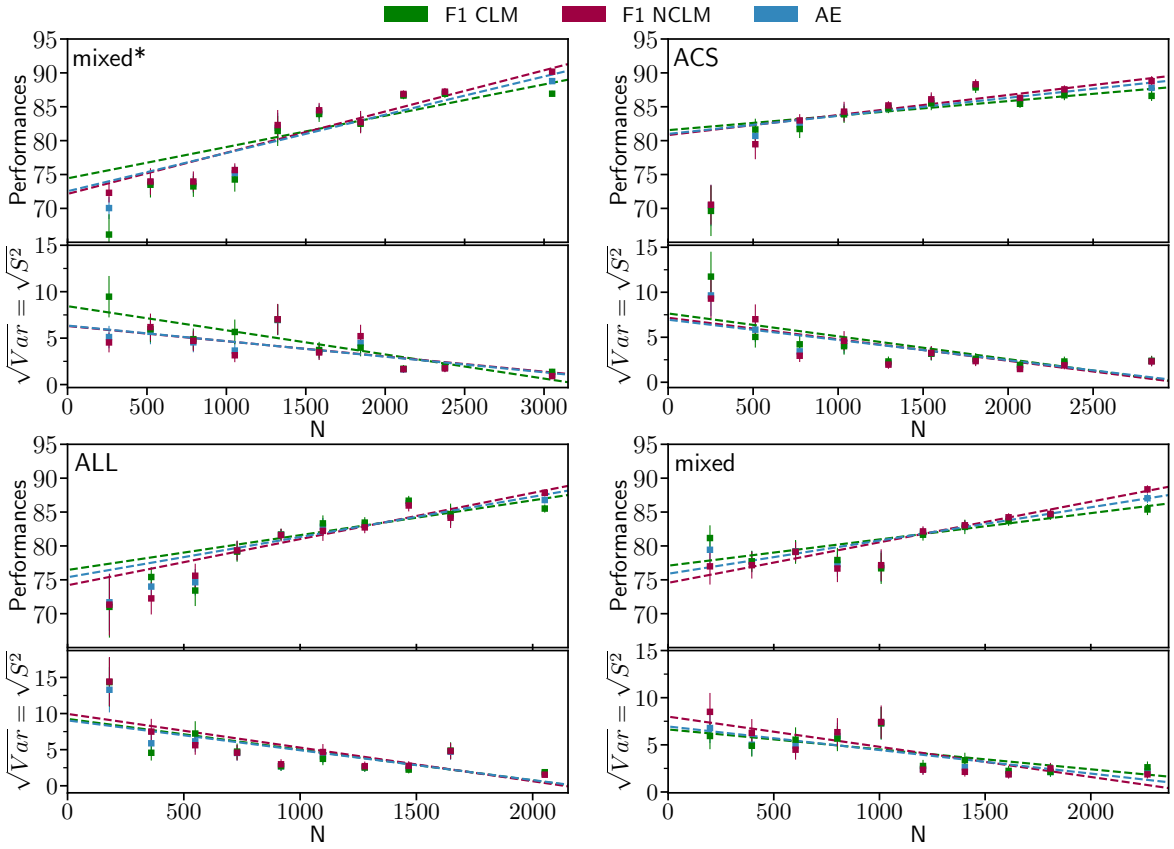


Figure 5.5: Comparison among the four band configurations (see Sec. 5.2), in terms of F1 score and average efficiency (AE) percentages (top panels), together with their square root of variances (bottom panels), as the number of spectroscopic sources in the training set increases (*EXP1*). In all panels, the linear best-fit trends are displayed as dashed lines. Due to the k-fold approach, performances have been averaged over the 10 folds, i.e. the x-axis shows the dimension of the training set, thus, the k-est fold used as test set has a size of $N/9$.

related to 13 common cluster and the only difference is the inclusion of two additional clusters in the *mixed** configuration. With an addition of 380 CLMs, we measured an improvement of $\sim 2\%$ in term of F1-score and efficiency. Such difference is particularly evident by looking at Tab. 5.3: the CNN trained with the *mixed* configuration covers just the 13% of the best metrics (respect to the 41% related to the *mixed**) with an average difference between all the cross-compared estimators $2.7 \pm 0.6\%$.

The dependence on the dataset dimension is also shown in Fig. 5.5, where performances and their fluctuations are displayed as function of the involved number of samples. For each configuration, we split the knowledge space into ten disjointed subsets, which have been progressively merged in order to build a dataset with which CNN has been trained and tested, always using the k-fold approach. As expected, there is a clear improvement of the classification capabilities as the number of sources increases (an accuracy gain of $\sim 2.3\%$ for an increment of 500 sources). Furthermore, fluctuations of these estimators tend to be better constrained for a large set of objects, stabilising around 3% when the number of samples is ≥ 2000 and showing an average reduction of $\sim 9\%$ by quadrupling the number of sources. Although this evident dependence, it must be emphasised that increasing of the available filters supplies, at least in part, the lack of samples: the usage of the complete set of bands (i.e. *ALL*), which is also the smallest dataset, allows to compensate the restrict number of spectroscopic examples, achieving a good trade-off between purity and completeness, as it is clear from Tab. A.2. Thus,

		k-fold	global	k-fold	global	k-fold	global	k-fold	global	k-fold	global
		$z_{CLM} \in (0.18, 0.32)$		$z_{CLM} \in (0.32, 0.37)$		$z_{CLM} \in (0.37, 0.41)$		$z_{CLM} \in (0.41, 0.46)$		$z_{CLM} \in (0.46, 0.60)$	
	AE	86.4 ± 1.1	86.2	89.0 ± 1.2	89.2	88.8 ± 1.4	88.6	88.1 ± 1.0	87.9	89.6 ± 1.3	89.6
	pur	84.9 ± 2.6	84.1	87.0 ± 1.7	86.9	87.9 ± 1.6	87.3	87.1 ± 1.0	87.0	87.7 ± 2.0	87.3
NCLM	compl	89.6 ± 1.6	89.2	92.1 ± 1.8	92.4	90.3 ± 1.5	90.3	89.5 ± 1.6	89.2	92.8 ± 0.9	92.7
	F1	86.9 ± 0.9	86.6	89.3 ± 1.2	89.5	89.0 ± 1.3	88.8	88.2 ± 1.0	88.1	90.0 ± 1.2	89.9
	pur	89.3 ± 1.1	88.5	91.7 ± 1.7	91.8	90.2 ± 1.5	89.9	89.4 ± 1.5	88.9	92.5 ± 0.9	92.2
CLM	compl	83.1 ± 3.2	83.1	85.9 ± 2.1	86.0	87.3 ± 1.8	86.9	86.6 ± 1.2	86.6	86.5 ± 2.2	86.5
	F1	85.7 ± 1.4	85.7	88.5 ± 1.3	88.9	88.6 ± 1.4	88.4	87.9 ± 1.0	87.8	89.2 ± 1.4	89.2

Table 5.4: Statistical estimators measured in each redshift bin for the *EXP1b* experiment. Due to the k-fold approach, the performances are reported as pairs of mean and error (evaluated on the 10 folds) and as a single global value.

the CNN classification capabilities seem to be largely dependent on dataset dimension once an optimal band configuration has been set. Conversely, by increasing the dataset size over a critical value (between 1500 and 2000 sources), performances do not improve further, whereas it is the combination of filters which plays a critical role.

In order to quantify the dependence on filter combination, we performed an additional experiment in which we used the same sources related to the *mixed** dataset and we explored all possible filter combinations. Results are outlined in Tab. A.3, from which is clear the performance increasing with the enlarging of the involved bands: excluding the usage of the only $F435$ filter (which has the lowest S/N ratio), by averaging efficiencies and F1-scores of experiments involving the same number of filters, we measured accuracy equals to 78.9%, 83.7%, 86.1%, 87.5%, 89.3%, and F1-score equals to 76.6%, 81.7%, 84.1%, 85.7%, 87.4%, by moving from experiments involving just one band to experiments with all the 5 bands related to the *mixed** configuration. These performances correspond to a relative increasing of 6%, 3%, 2%, 2% for both accuracy and F1-score, suggesting a saturation of performance; indeed, by cross-comparing all the efficiencies related the experiments involving 3 filters with those involving 4 bands, we measure an average relative increasing of 1.6%, to compare with the relative increasing between 2 and 3 filters experiments equals to 3.0%, and between 1 and 2 filters experiments equals to 6.2%. By using just one filter (always excluding the experiment involving only $F435$ band), performances are typically $\lesssim 80\%$, with an CLM F1-score ranging in $(73\%, 79\%)$ and efficiency in $(76\%, 81\%)$, implying that even if a single band encodes morphology information, this is not enough to allow the cluster member separation, also suggesting that the use of colours (or rather, a filter weighted combinations) improves significantly the network classification capabilities. It is worth underling that the ACS filter combination (i.e. $F435$, $F606$ and $F814$) carries enough information to identify members, showing performances higher than any other 3 filter configurations and than when just one WFC3 band is added to this combination (i.e. by comparing $F435, F606, F814$ with $F435, F606, F814, F105$ and $F435, F606, F814, F140$, we measure a average relative reduction of $\lesssim 0.4\%$), such configuration becomes optimal only when both $F105$ and $F140$ are included. This analysis confirms the requirement to exploit the optimal combination between ACS and WFC3 bands in order to disentangle CLM from background and foreground sources.

Finally, since the training set used in this study was composed of galaxies spanning a large redshift range, as part of *EXP1*, we investigated whether any dependence on redshift is present. To this aim, the CLM redshift range was split into five equally populated bins (~ 280 samples) and, to complete the knowledge space, we extracted without repetitions an appropriate number of objects from the NCLM population. The network has been trained within each ensemble adopting the k-fold approach, using only the *mixed** band combination. The performances and fluctuations related to the *mixed** band combination are graphically shown in Fig. 5.6, while

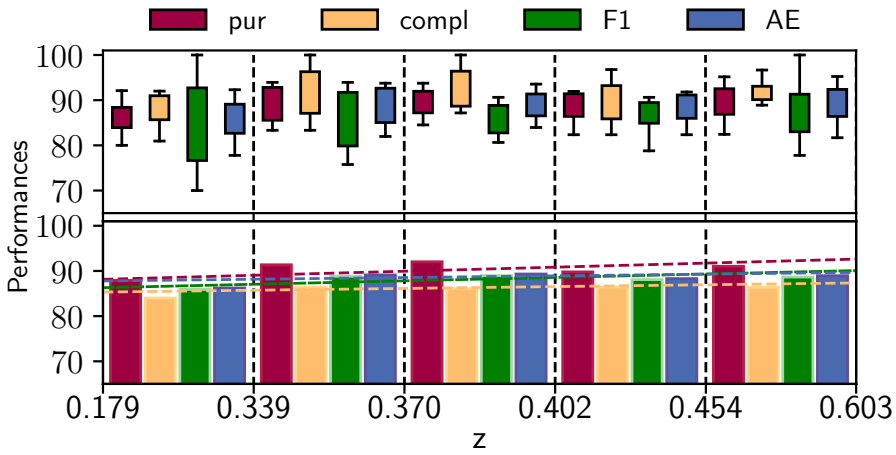


Figure 5.6: Percentages of CNN classification results for the four statistical estimators, CLM purity (*pur*, in red), completeness (*compl*, in orange) and F1-score (in green), together with the average efficiency (*AE*, in blue), measured as a function of CLM redshift range (*EXP1b*). The top panel describes their fluctuation in each bin (evaluated on the 10 folds), with the boxes delimiting the 25th and 75th percentiles (first and third quartile) and error bars enclosing the maximum variations. The bottom panel shows these metrics globally evaluated in each bin by stacking the 10 folds (these performances are also listed in Tab. 5.4, see columns named as "global"), together with the best-fit lines.

details on the metrics are given in Table 5.4, in which, we have specified the fluctuation of estimators as an error estimated on the ten folds. Despite the dissimilarities between galaxies at different depths, the CNN did not seem to be affected by the CLM redshifts. In fact, CNN performances achieved in different redshift bins were all comparable, with a dispersion included within $0.04 - 1\sigma$ for the 65% of cross-compared estimator pairs and a mean separation of $\sim 0.8\sigma$.

Since the *mixed** band combinations provided the best results, all further experiments in the next sections refer to this band configuration.

5.3.2 *EXP2*: Selection of clusters as blind test set

A second set of experiments was devoted to the study of the CNN capability to predict cluster membership of sources belonging to clusters that are not included in the training set, that is, avoiding having member galaxies belonging to the same cluster populating both the training and test sets (as has been done for *EXP1*). Thus, we considered A370 ($z = 0.375$), MS 2137-2353 (MS2137, $z = 0.316$), and MACS J0329-0211 (M0329, $z = 0.450$) as blind test clusters, while the remaining clusters were organised into three different training sets based on different redshift ranges, as shown in Fig. 5.3. Specifically:

- *Narrow*: clusters with redshift $0.332 \leq z \leq 0.412$ (514 CLMs, 555 NCLMs)
- *Intermediate*: clusters with redshift $0.286 \leq z \leq 0.467$ (898 CLMs, 1157 NCLMs)
- *Large*: clusters with redshift $0.174 \leq z \leq 0.606$ (1286 CLMs, 1725 NCLMs)

The training set configurations were mostly organised to identify CLMs in A370. This is the most significant test bench since it includes 535 spectroscopic sources and it is in the middle of CLM redshift range. The other two clusters, MS2137 and M0329, were chosen as additional test sets located at redshift lying outside the *narrow* and *intermediate* ranges, while remaining well within the *large* training set.

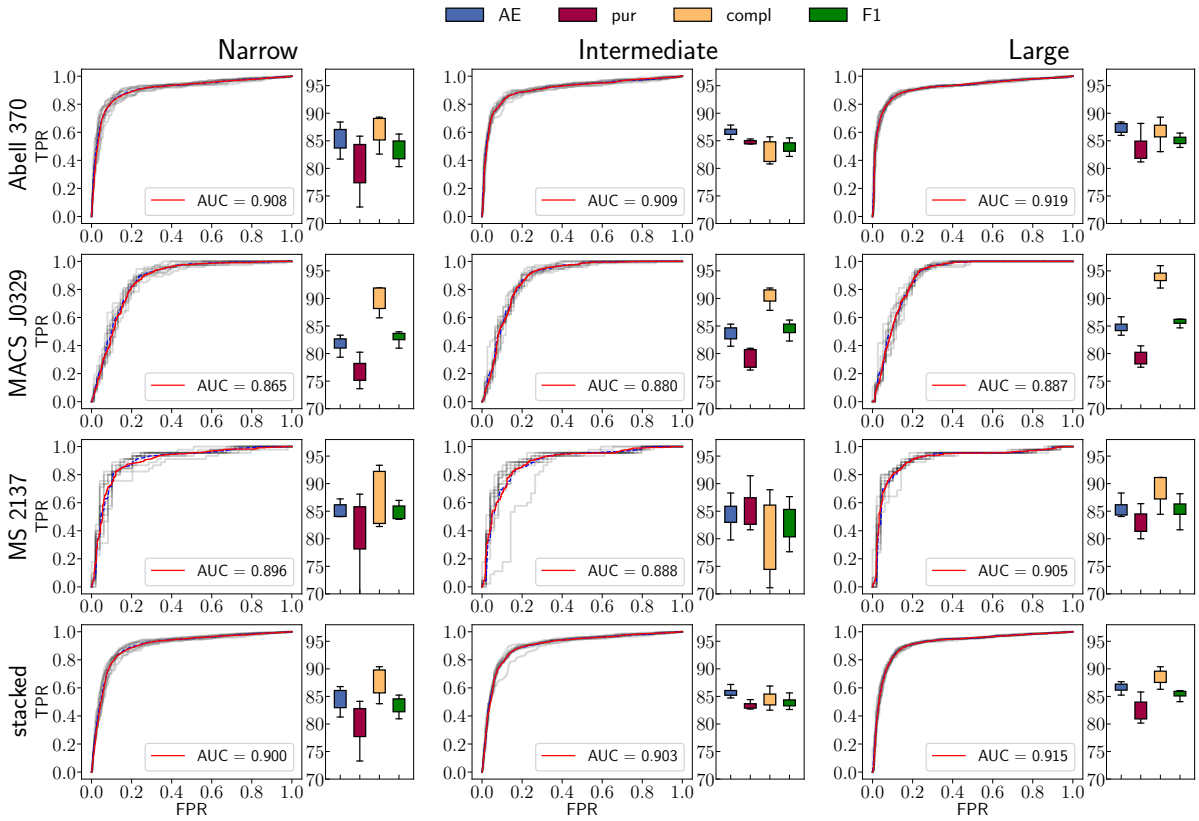


Figure 5.7: Summary of the *EXP2* experiment. The statistical performances for the three clusters (A370, M0329 and MS2137) are reported in each row, while results for the three training configurations (i.e. *narrow*, *intermediate* and *large*) are organised by column. The global performances achieved by stacking together the three clusters are reported in the bottom row. For each test set, we display the ROC curves (grey lines refer to the performances achieved by any training fold, while the main trend is emphasised in red, together with its AUC score); the box plots represent the fluctuation of measured estimators related to the CLMs, together with the average efficiency measured for both classes. As in Fig. 5.6, such boxes delimit the 25th and 75th percentiles, while error bars enclose the maximum variations.

The results are shown in Fig. 5.7 and detailed in Table 5.5. They show that: (*i*) the *large* training set reached best results in most cases, with an average improvement between 1.1% and 4.3% with respect to the *intermediate* case; (*ii*) the *narrow* training ensemble provided, in most cases, the worst results, showing a lower trade-off between purity and completeness, particularly evident (larger than 3σ) for A370 and M0329; (*iii*) a reduction of performance fluctuations with the increasing of training set ($\sim$ 50% moving from the *narrow* to the *large* training set). This confirmed that the best performances were reached by extending the knowledge base, that is, when the CM training sample covers the largest available redshift range.

Moreover, with the *EXP2* we evaluated the network capabilities to identify members in clusters that were not involved in the training phase, thus, this experiment estimates the performance loss when we apply a trained network to an unseen cluster. By comparing performances achieved on these three clusters between *EXP1* and *EXP2* experiments (i.e. by measure the differences between the statistical estimators listed in Tab. 5.5 and in Tab. 5.3), we found an average drop of $\sim$ 2% both in terms of accuracy and CLM F1-score.

We also analysed the CNN classification performances separately on bright and faint (*EXP2a*) galaxies, as well as on red and blue galaxies (*EXP2b*). The magnitude values adopted to split the CLM into equally sized samples are $F814$=22.0, 21.7, and 21.6 mag for A370, M0329, and MS2137, respectively. For the analysis of the colour dependence, we used the normalised colour

		stacked				A370		
		Narrow	*Intermediate*	*Large*		*Narrow*	*Intermediate*	*Large*
	AE	84.5 ± 0.6	85.5 ± 0.4	**86.6 ± 0.3**	AE	85.4 ± 0.7	86.6 ± 0.3	**87.4 ± 0.3**
CLM	pur	79.6 ± 1.2	**83.2 ± 0.2**	82.5 ± 0.6	pur	80.3 ± 1.4	84.5 ± 0.2	**83.9 ± 0.7**
%	comp	87.6 ± 0.8	83.9 ± 0.8	**88.5 ± 0.4**	comp	86.5 ± 0.7	83.1 ± 0.6	**86.6 ± 0.6**
	F1	83.3 ± 1.2	83.6 ± 0.2	**85.4 ± 0.6**	F1	83.3 ± 1.4	83.8 ± 0.2	**85.1 ± 0.7**
		M0329				MS2137		
		Narrow	*Intermediate*	*Large*		*Narrow*	*Intermediate*	*Large*
	AE	81.7 ± 0.5	83.5 ± 0.5	**84.8 ± 0.3**	AE	84.1 ± 1.1	82.4 ± 2.3	**85.4 ± 0.7**
CLM	pur	76.9 ± 0.7	**79.2 ± 0.5**	**79.2 ± 0.4**	pur	81.5 ± 1.9	**84.4 ± 1.8**	82.3 ± 1.0
%	comp	90.0 ± 0.6	90.4 ± 0.4	**93.9 ± 0.4**	comp	87.6 ± 1.5	77.1 ± 4.8	**88.9 ± 0.7**
	F1	82.9 ± 0.7	84.4 ± 0.5	**85.9 ± 0.4**	F1	84.2 ± 1.9	80.0 ± 1.8	**85.4 ± 1.0**

Table 5.5: Percentage performances on a blind test set related to the *EXP2* experiment. Performances have been displayed by splitting between the three test clusters: A370 ($z = 0.375$), MS2137 ($z = 0.316$), M0329 ($z = 0.450$) and by their stacking. Best results are emphasised in bold. For ease of reading, only statistics related to the CLM class are reported, together with the average efficiency (AE), which refers to both classes.

	stacked				A370			
%	bright	faint	redder	bluer	bright	faint	redder	bluer
pur	**85.9 ± 0.4**	82.2 ± 0.8	**91.0 ± 0.5**	79.4 ± 0.9	**88.4 ± 0.7**	83.6 ± 0.9	**90.5 ± 0.7**	79.8 ± 1.0
compl	**95.2 ± 0.7**	81.4 ± 1.0	**95.2 ± 0.6**	75.7 ± 1.0	**96.8 ± 0.7**	80.8 ± 1.2	**93.9 ± 0.4**	77.4 ± 1.2
F1	**90.3 ± 0.4**	81.7 ± 0.8	**93.1 ± 0.7**	77.6 ± 0.8	**92.4 ± 0.7**	82.1 ± 0.9	**92.2 ± 0.8**	78.6 ± 0.9

	M0329				MS2137			
%	bright	faint	redder	bluer	bright	faint	redder	bluer
pur	80.7 ± 0.6	**81.1 ± 1.7**	**88.3 ± 0.9**	74.4 ± 1.2	**90.8 ± 1.0**	76.7 ± 1.5	**87.5 ± 0.3**	72.0 ± 1.3
compl	**98.0 ± 1.0**	85.1 ± 0.6	**95.1 ± 0.6**	78.6 ± 0.8	**88.9 ± 1.2**	80.0 ± 0.9	**90.6 ± 0.6**	76.2 ± 1.0
F1	**89.3 ± 0.5**	83.0 ± 1.7	**91.7 ± 0.7**	76.5 ± 1.0	**89.7 ± 1.1**	78.3 ± 1.2	**89.0 ± 0.4**	74.1 ± 1.1

Table 5.6: Statistical performances of the CNN model in *EXP2a* and *EXP2b*. Performances have been displayed by splitting between the three test clusters: A370 ($z = 0.375$), MS2137 ($z = 0.316$), M0329 ($z = 0.450$) and by their stacking. Best results are emphasised in bold.

$(F814 - F160)_{\text{norm}}$, defined in Eq. 5.2. By applying the correction for the colour-magnitude sequence, we found that blue members can be defined as galaxies having $(F814 - F160)_{\text{norm}} < -0.160, -0.165, -0.157$ for A370, M0329, and MS2137, respectively. Both experiments (*EXP2a* and *EXP2b*) were performed using the *large* redshift configuration.

The results of the CLM identification are shown in Table 5.6. Concerning *EXP2a*, all the statistical estimators indicated a very good performance of the method, although with a lower efficiency in identifying faint objects, as expected, due to the reduced S/N ratio of fainter members. In fact, brighter members were detected with higher completeness (90% – 98%) and purity (81% – 91%), with a significant F1 score improvement (89% – 92%), when compared to fainter members (completeness: 80% – 85%; purity: 77% – 85%; F1 score: 78% – 83%), obtaining remarkable results for A370, in which purity and completeness of CLMs are $\sim 88\%$ and $\sim 97\%$, respectively. Nevertheless, fainter CLMs were identified with an acceptable F1 score ($\sim 80\%$).

The experiment *EXP2b*, also showed good performances of the method for both red and blue objects, although the colour dependence of the results was evident. In particular, red galaxies were classified with a mean F1 score of $\sim 91\%$, decreasing down to $\sim 77\%$ for blue objects. The results reflect the underlying similarity between blue members and background objects, which implies that they cannot be separated easily. This was confirmed by the analysis of false positives

		R2248 $z = 0.346$					M0416 $z = 0.397$			
		CNN	RF	Bayesian	Δ		CNN	RF	Bayesian	Δ
	AE	**88.1**	86.5	85.9	1.6	AE	**92.2**	89.2	87.1	3.0
	pur	**88.3**	87.7	80.9	0.6	pur	**93.3**	93.0	84.6	0.3
CLM	compl	89.8	87.7	**96.1**	-6.3	compl	87.1	86.5	**91.2**	-4.1
	F1	**89.1**	87.7	87.8	1.3	F1	**91.5**	89.7	87.8	1.8
	pur	87.9	85.1	**94.4**	-6.5	pur	89.0	84.5	**90.0**	-1.0
NCLM	compl	**86.1**	85.1	74.4	1.0	compl	**96.9**	92.3	82.7	4.6
	F1	**87.0**	85.1	83.2	1.9	F1	**91.5**	88.3	86.2	3.2
	μ_Δ		-0.91 ± 1.42			μ_Δ		1.11 ± 1.12		
		M1206 $z = 0.439$					M1149 $z = 0.542$			
		CNN	RF	Bayesian	Δ		CNN	RF	Bayesian	Δ
	AE	**89.7**	87.9	85.0	1.8	AE	**89.4**	86.9	85.5	2.5
	pur	89.9	**90.4**	80.2	-0.5	pur	**82.3**	78.8	71.8	3.5
CLM	compl	86.5	81.9	**91.2**	-4.7	compl	91.3	88.5	**98.0**	-6.7
	F1	**88.2**	85.9	85.3	2.3	F1	**86.6**	83.4	82.9	3.2
	pur	89.6	86.3	**90.8**	-1.2	pur	94.5	92.7	**98.6**	-4.1
NCLM	compl	92.3	**92.9**	79.4	-0.6	compl	**88.3**	86.0	78.4	2.3
	F1	**90.9**	89.7	84.7	1.2	F1	**91.3**	83.4	87.4	3.9
	μ_Δ		-0.24 ± 0.90			μ_Δ		0.66 ± 1.60		

Table 5.7: Comparison between our image-based CNN model and two different photometric catalogue-based approaches, referred to the *EXP3* experiment. The comparison involves two different model: a Random Forest and a Bayesian method, applied on photometric tabular information of four clusters: R2248 ($z = 0.346$), M0416 ($z = 0.397$), M1206 ($z = 0.439$) and M1149 ($z = 0.542$). Last column, Δ, shows the difference between CNN estimators and the best between the two photometric approaches (see Eq. 5.3), while rows μ_Δ list the averages among these Δs for each cluster.

and false negatives discussed in Sec. 5.4.

5.3.3 Comparison with photometric approaches (*EXP3*)

This section is dedicated to a comparison of the classification performance of cluster members using the image-based DL method described above along with two different techniques based on photometric catalogues. The first is a random forest classifier[3] (RF, Breiman 2001) and the second one is a photometry-based Bayesian model (BM) described in Grillo et al. (2015), which has already been applied in order to enlarge the cluster member selection, including galaxies without spectroscopic information, for four clusters: R2248, M0416, M1206, and M1149 (Grillo et al. 2015; Caminha et al. 2016, 2017b; Treu et al. 2016). Both methods are briefly described in Sec. 3.8. Both techniques critically use multi-band photometric information, for example, magnitudes and colours.

In this experiment, CNN was trained with the *mixed** filter set (see Sec. 5.2). We focused on the results obtained by these three methods on four galaxy clusters: R2248, M0416, M1206, and M1149. The statistical estimators are detailed in Table 5.7 and shown in Fig. A.1 as ROC curves for each involved cluster, while in Fig. 5.8 performances are summarised by combining the results from the four clusters based on their ROC curves (upper left panel), the trade-off between purity and completeness (right panel), and the usual statistical estimators (bottom left panel). The

[3]Random forest has been developed by our team by exploiting Scikit-Learn python open source library (Pedregosa et al. 2011).

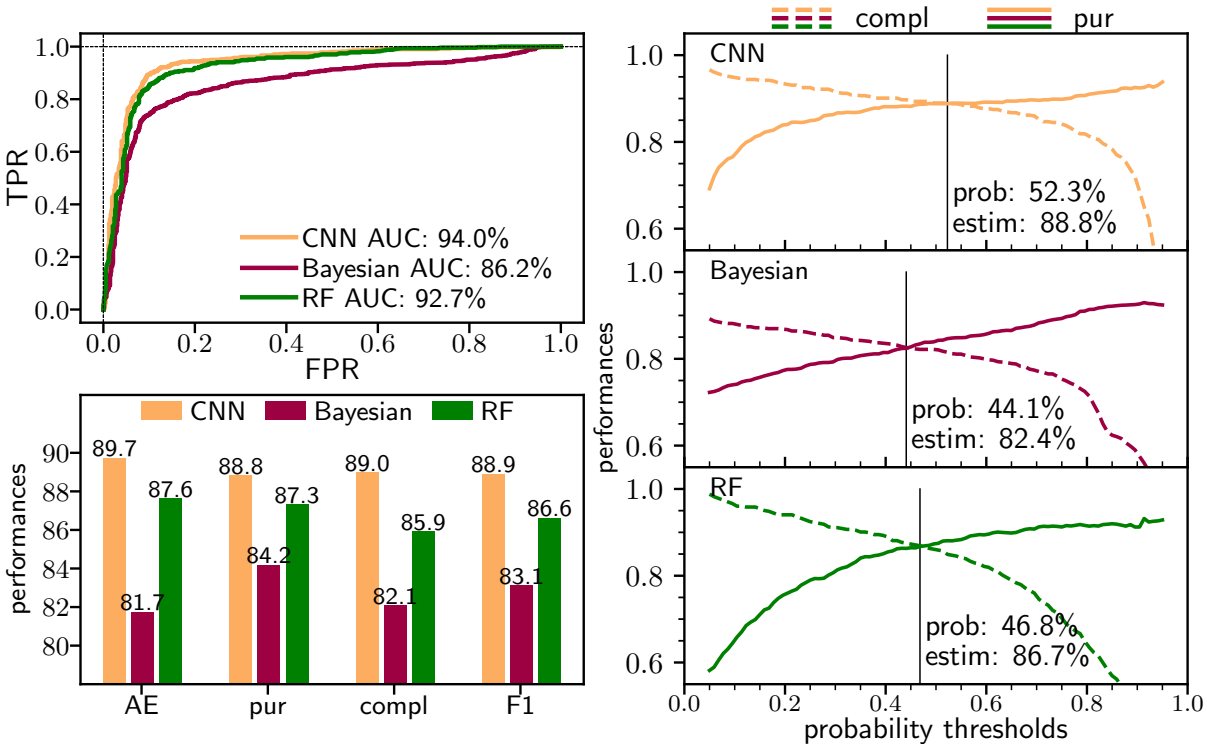


Figure 5.8: Comparison among the image-based CNN and two photometric catalogue-based approaches, namely, a random forest and Bayesian method (*EXP3*), by combining results from the four clusters (R2248, M0416, M1206, M1149). Upper left panel shows the ROC curves for the three methods with measured Area Under the Curve (AUC). The right panel reports the trends of purity and completeness as a function of the probability thresholds used to obtain the ROC curves, where, for each diagram, we mark the intersection between such curves, i.e. the probability for which completeness and purity are equal. Bottom left panel shows the differences between the three methods based on the statistical estimators described in Sec. 3.6.6.

photometric techniques show an average efficiency around 86−89%, with some values $\gtrsim$ 96% for the Bayesian approach, although the F1 scores always remain between 83% and 88%. The CNN confirmed its ability to detect CLMs with an F1-score between 87% and 91%. The upper left panel in Fig. 5.8 shows that globally CNN reaches an AUC of $\sim$ 94%, which is $\sim$ 8% higher than the Bayesian method, while exhibiting the sharpest rise and the highest plateau. This means that for the CNN method there is a larger probability range in which the performances remain stable, while for the other methods a fine-tuning of the probability value is needed to balance purity and completeness. These behaviours are also represented in the right panels of Fig. 5.8, in term of trade-off between purity and completeness as a function of a membership probability threshold: CNN completeness remains between 95 and 80% for a large probability range, dropping down for $Pr \geq 85\%$, RF completeness has a sharply descending trend, falling below 80% for $Pr \geq 65\%$, while the BM shows a more constant and lower decreasing completeness with values ranging between 90 and 70% for $Pr \leq 70\%$; a complementary, similar reasoning can be done for the purity curves. Moreover CNN performances are characterised by a larger probability range in which purity and completeness are comparable: pur/compl $\in (0.95, 1.05)$ for $Pr \in (30\%, 70\%)$; whereas this range is $Pr \in (30\%, 50\%)$ for the BM and $Pr \in (40\%, 50\%)$ for the RF. As result, the CNN purity-completeness cross-over occurs at $\sim$ 89%: $\sim$ 6% better than the Bayesian classifier and $\sim$ 2% better than RF. This confirms a more balanced behavior of CNN with respect to photometric methods. A summary of the results is shown in the bottom left panel of Fig. 5.8, where the differences among the CNN and the two photometric methods are measured using

the four statistical estimators. The CNN performances were overall near 90% and remained consistently higher than those of photometric-based methods, with a gain between 1.5% and 3.1% respect to the RF, and between 4.6% and 8.0% respect to the BM.

Analysing the results achieved for each cluster (see Tab. 5.7 and Fig. A.1), we found that CNN has a stable behaviour regardless of the considered cluster with AUC values $\geq 92.7\%$ and efficiency ranges in 88.1% and 92.2%; also the RF classifier has stable performances, although always minor than CNN: difference between AUC values range within 0.1% and 2.6%, while differences between efficiencies are constrained between 1.6% and 2.5%; the Bayesian classifier is the method with the largest performance fluctuations: the AUC value drops to $\sim 66\%$ in the case of R2248, but reaches the $\sim 95\%$ for M1149 (which is the best score among the AUCs related to this cluster), efficiency is stable around $85.9 \pm 0.4\%$, however it shows the largest trade-off between completeness and purity, which differences (completeness − purity) range between 6.6% and 27.2% (to compared with the $(-1.5, 9.0)\%$ and $(-8.5, 9.7)\%$, respectively for the CNN and RF). In order to emphasise the CNN capabilities compared to the other methods, we also computed the differences:

$$\Delta_{\text{estim}} = \text{estim}_{CNN} - max\{\text{estim}_{RF}, \text{estim}_{Bayesian}\}, \text{estim} \in [\text{pur, compl, F1, AE}] \tag{5.3}$$

which is a kind of stress test for the CNN since we estimated the difference between CNN metrics and the corresponding *maximum* scores achieved by RF or BM. All these differences are listed in the last column of Table 5.7, together with the average among these Δs for each cluster (rows μ_Δ). Despite such test disadvantages the CNN, we found that CNN outperforms the best combination of the other two methods in the 64% of cases, with Δs oscillates between −6.7% and +4.6%, suggesting that the CNN performances are at least comparable with the best combination of the results achieved by the two photometric approaches.

Finally, we analysed the common predictions among the three methods, both in terms of correctly classified and misclassified sources, separately for CLMs and NCLMs. Such results are graphically represented in Fig. A.2. All three methods share $\sim 76\%$ of their commonalities (i.e. summing of correct and incorrect predictions), of which, $\sim 97\%$ (i.e. 74.6% with respect to the whole set of common sources) were correctly classified. Common true positives and true negatives (i.e. CLMs and NCLMs that have been correctly classified) were $\sim 75\%$. The CNN and Bayesian method shared the largest fraction of predictions $\sim 90\%$ (of which $\sim 93\%$ were correct) with respect to the joint classification of CNN and RF ($\sim 82\%$); this implied that RF had a significant fraction of uncommon predictions ($\sim 14\%$).

Concerning the misclassified objects, the methods shared $\sim 2\%$ of incorrect predictions, of which: $\sim 1\%$ of CLMs were common false negatives (FNs, i.e. CLMs sources wrongly predicted as NCLMs), while 2.5% were common false positives (FPs, i.e. NCLMs sources wrongly predicted as CLMs). The CNN exhibited the least fraction of misclassifications (about 10%). The CNN showed a percentage of FNs larger than BM (10% versus 7%), which, in turn, had a wider FP rate (11% versus 17%). Therefore, although CNN and Bayesian methods shared a significant fraction of incorrect predictions (85% of common misclassifications, suggesting the existence of a fraction of sources for which the membership is particularly complex for both of them), these two models exhibit a different behaviour: the CNN tended to produce more pure than complete CLMs samples, whereas the BM showed the opposite, which is in agreement with what is reported in Table 5.7.

5.4 False Positive and False Negative analysis

In the previous sections we studied the impact on the classification performances due to the amount of data available: the *EXP1* (Sec. 5.3.1) enabled an analysis of the trade-off between

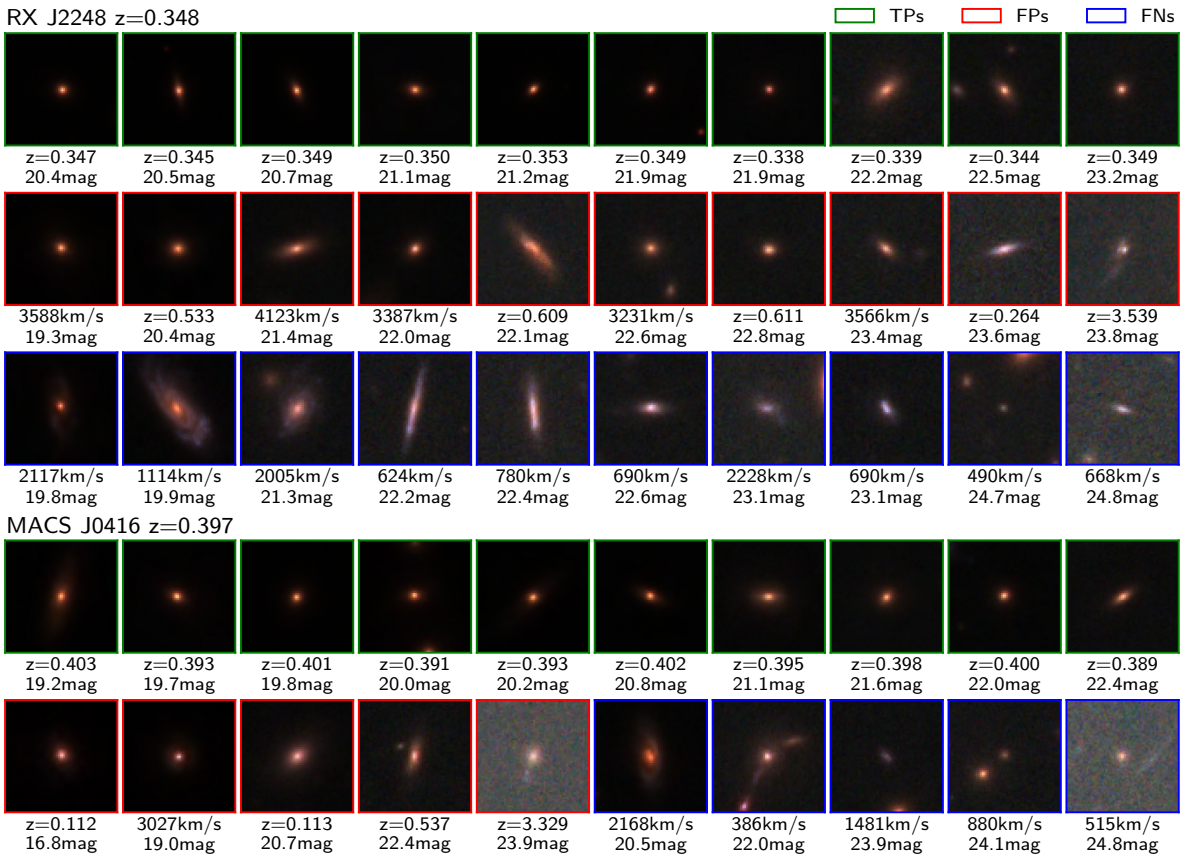


Figure 5.9: Ensemble of object cut-outs with a size of 64 pixels ($\sim 4''$), corresponding to some specific CNN predictions in the clusters R2248 (first three rows) and M0416 (last two rows). The True Positives (TPs), i.e. the CLMs correctly identified, are shown on first and fourth row with green boxes, while False Positive and False Negative (FPs and FNs) are shown on the second, third and fifth row, framed by red and blue boxes, respectively. The images were obtained by combining five HST bands: $F435$, $F606$, $F814$, $F105$, $F140$. The figure shows sources in the $F814$ band with a magnitude $F814 \leq 25$ mag. TPs are shown together with their spectroscopic redshift, while FNs together with their cluster rest-frame velocity separation. For convenience, in the case of FPs, their cluster velocity separations are quoted when within $\pm 9000\,\mathrm{km\,s^{-1}}$, otherwise their redshift is shown.

the information carried by the imaging bands and the number of samples in the dataset; while the *EXP2* is configured as a "stress" test, with which we explored the CNN to predict the membership for sources in clusters whose members (and non-members) have been completely excluded from the training set, with this experiment we also investigated the dependence of member classification performance on the magnitudes and colours. In this section we specifically analyse CNN predictions, by investigating the False Positive and False Negative distributions, trying to understand the causes of such misclassifications. In the first place we studied the CNN classification dependence on the cutout crowding, then, we inspected the False Positive and False Negative dependence on their magnitude and colour. In this analysis we used the CNN predictions trained with *mixed** band configuration (see Sec. 5.2).

For easy reading, we specify the definition of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN), already given in Sec. 3.6.6, within the context of this work. We assume the cluster member as the *positive* class, thus, the TPs are CLMs correctly classified, FPs are NCLMs classified as CLMs, FNs are CLMs classified as NCLMs, and, finally, TNs are NCLMs correctly classified. A short sample of TPs, FPs and FNs in R2248 and M0416, and in M1206 and M1149 are shown in Fig. 5.9 and Fig. 5.10, respectively.

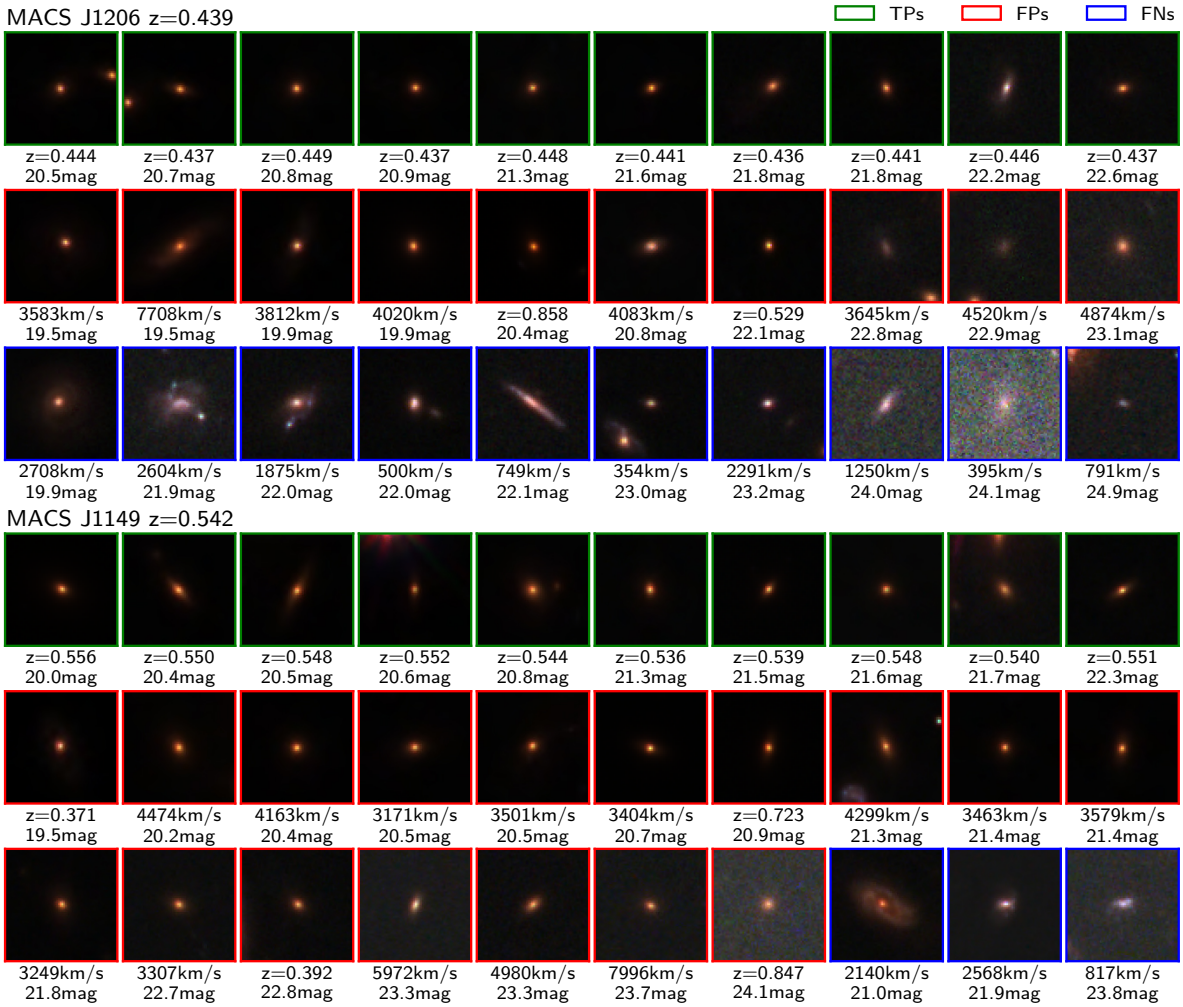


Figure 5.10: Same of Fig. 5.9 for the cluster M1206 (first three rows) and M1149 (last three rows).

A critical aspect of the classification of members within the central cluster region is the impact of crowding. Therefore, we specifically focused on the DL ability to predict cluster membership in such circumstances (see a few examples of cut-outs in Figs. 5.9 and 5.10). We introduced a contamination index (CI) for each cut-out, defined as:

$$CI = \sum_{i=1}^{N_c} 1/(d_i \cdot F814_i) \tag{5.4}$$

where N_c is the number of contaminants in the cut-outs, d_i is the distance in arcsec between the central source and $i-$th contaminant, while $F814_i$ is the magnitude of the contaminating source. The indices for cut-outs without contaminants were set to zero. Then, we normalised this index in the [0, 1] interval. Fig. 5.11 shows that the four contamination index distributions of, respectively, TPs, TNs, FPs and FNs mostly overlapped and followed the same trend. In fact, the 48% of FNs and 28% of FPs had a non-zero contamination index, as well as the 31% and 43% of TNs and TPs. The lack of a correlation between the contamination index and incorrect prediction rates (FPs and FNs) suggests that the source crowding did not significantly affect the CNN classification efficiency.

By analysing the FP and FN rows in Figs. 5.9 and 5.10, we can see an interesting dichotomy: FPs appear as red galaxies, while FNs as blue; in addition, the FPs have $F814 < 24$ mag, whereas FNs are found also down to $F814 \sim 25$. In order to quantify such behaviours, we analysed the

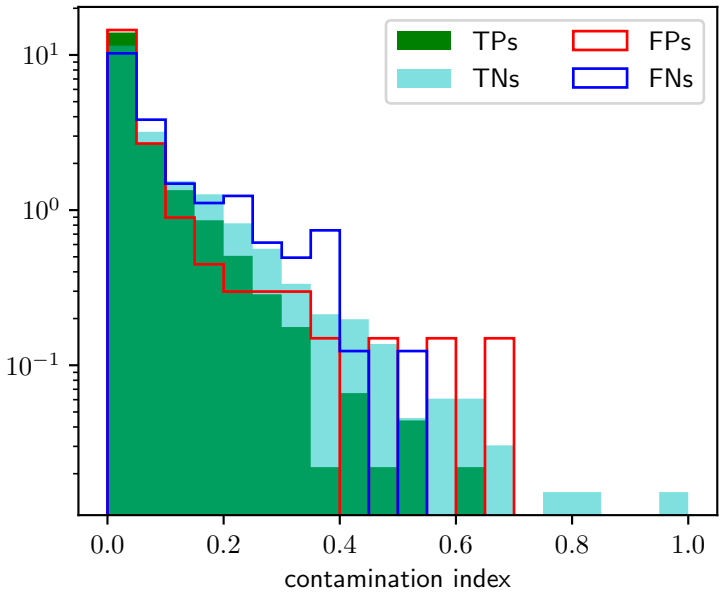


Figure 5.11: Logarithmic distribution of the *contamination index* (see Eq. 5.4) for True Positives (TPs, green), True Negatives (TN, cyan), False Positives (FP, red) and False Negatives (FN, blue). The distribution includes all available clusters.

distribution of FPs and FNs in terms of: *(i)* the $F814$ magnitude for both FPs and FNs (left panel in Fig. 5.12); *(ii)* the correlation between the CNN incorrect predictions and normalised colours $(F606 - F814)_{\rm norm}$ (right panel in Fig. 5.12). These results are summarised in Table 5.8.

Left panel in Fig. 5.12 and Col. 4 in Table 5.8 showed that almost all CLMs fainter than $F814 = 25$ mag (representing a small fraction with respect to the total, see Col. 2 in Table 5.8) were FNs. This was not due to any failure on the part of the model, but, rather, to the poor sampling of such objects within the parameter space available to train the model. This was also confirmed when comparing the percentage of FPs and FNs with respect to the percentage of CLMs and NCLMs in Table 5.8 as a function of magnitude. In fact, Table 5.8 showed that the model tried to reproduce the distribution in terms of fractions of CLMs for FPs, and in terms of the fraction of NCLMs for FNs.

We analysed the correlations between the CNN incorrect predictions and colours. These distributions are shown in the right panel of Fig. 5.12 using the normalised colour $(F606 - F814)_{\rm norm}$, while, in Table 5.8 the misclassification percentages are summarised. Also in this case, the distributions of FPs and FNs, as a function of colours, are mimicking, respectively, the distributions of CLMs and NCLMs in Table 5.8. Very blue sources ($(F606 - F814)_{\rm norm} < -0.5$) populated only 5.8% of CLMs and represented the $\sim$ 35.4% of incorrect predictions, which is very similar to the fraction of very blue sources in the population of NCLMs (i.e. 43.2%). Conversely, redder sources were typically correctly classified, showing a FN rate of 16.6%. Moreover, from the fraction of FN/CLMs, we observed that almost all the blue cluster members were wrongly classified as NCLMs (see Col. 4 in Table 5.8 and right panel in Fig. 5.12).

Regarding FPs, there was not a real classification problem with faint and very blue objects, whose rates in terms of CLMs were, respectively, 3.4% and 5.8%, corresponding to 2.2% and 4.3% of incorrect predictions, respectively. From Table 5.8, it was also evident that within red misclassifications, FPs were more frequent than FNs (29.5% versus 16.6%), reproducing the distributions of CLMs (39.2%) and NCLMs (15.4%), respectively. Finally, we found that a fraction of FPs are interlopers (identified as those sources having $|v| \in (3000, 6000)\,{\rm km\,s^{-1}}$): even if they represent just the 4% of the whole NCLM set, 56.2% of them are FPs covering up to 30% of the whole FP ensemble; this behaviour is due to the similarity between members and

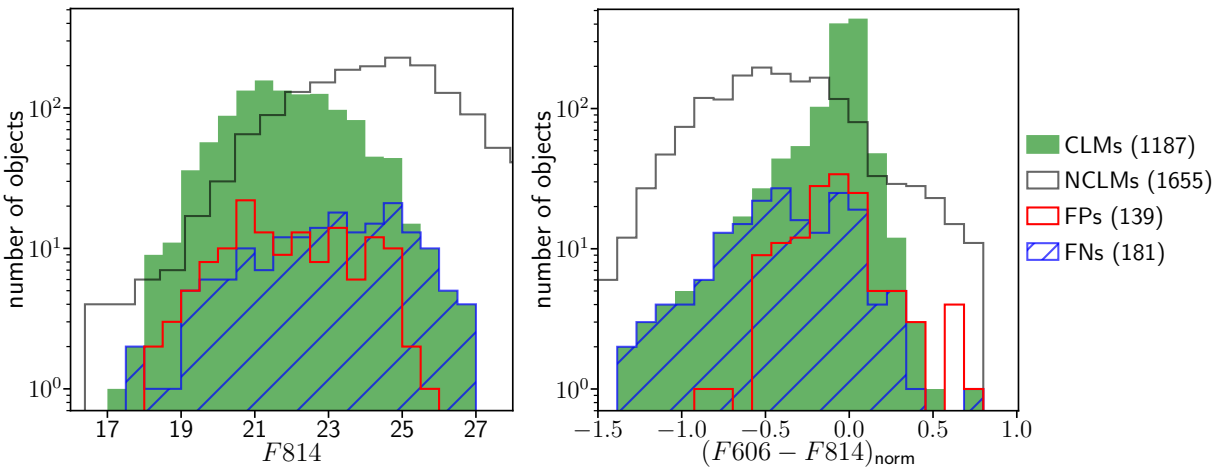


Figure 5.12: Magnitude (left panel) and colour (right panel) logarithmic distributions of FPs (red) and FNs (blue), overlapped to the CLM (green) and NCLM distributions (open grey), for the fifteen clusters (*stacked*) included in our study. The number of objects for each plotted distribution is quoted between brackets in the legend. The normalised colour $(F606 - F814)_{\rm norm}$ is obtained by applying the correction for the colour-magnitude relation for each cluster (Eq. 5.2). Tab. 5.8 outlines such results.

interlopers (as partially shown in Fig 5.9 and in Fig. 5.10 looking for objects whose velocity separation is lager than 3000 km s^{-1}), indeed 83.6% of interlopers are red galaxies, of which 47.5% are FPs.

Fig. 5.13 shows the colour-magnitude relation of CLMs (green squares), overlapping the FP (red cross), FN (blue cross) and NCLM (grey circle) distributions. It emphasises the CLMs undersampling of the blue and faint region, together with the large concentration of FNs among bluer and fainter sources (see blue crosses). Among all the FNs, $\sim$ 35% are very blue ($(F606 - F814)_{\rm norm} < -0.5$), $\sim$ 40% of these had $F814 > 25$ mag, suggesting that in the bluer region the FNs follows the NCLM distribution, while among FPs, $\sim$ 64% of them are red ($(F606 - F814)_{\rm norm} > -0.1$), but only $\sim$ 1% of these have magnitude fainter than $F814 > 25$ mag. On the other hand, $\sim$ 35% of all FPs were within the yellow contours, which refer to the 1σ colour-magnitude relation, indicating that they were on the red sequence.

As introduced in Sec. 5.2, we did not apply any photometric thresholds, in this way we exploited faint and blue members together with red interlopers as adversarial examples to prevent the model overfitting and increasing the network generalisation capabilities. In order to understand the impact of these examples, we report, in Tab. 5.9, the statistical estimators for the stacked sample considering either the whole sample, by removing faint and very blue objects (i.e. $F814 > 25$ & $(F606 - F814)_{\rm norm} < -0.5$) and, only for the sake of completeness, by also excluding interlopers (i.e. 3000 km s^{-1} $< |v| \leq$ 6000 km s^{-1}). As already stated (Sec. 5.2), the exclusion (even partial) of these objects from the whole KB causes an apparent significant improvement of the performances, but the network loses any capability to detect members in other clusters, i.e. the network is prone to overfit in the absence of these adversarial sources. However, it is worth quantifying their influence in term of misclassifications, by evaluating the statistical estimators on a test set from which they have been excluded. By comparing these results (Tab. 5.9), we observed a relevant increase of the completeness (for the stacked sample, it goes from 84.8% to 90.8%). This was mainly motivated by the sensible reduction of the FNs amount, which, by definition, had a higher impact on the completeness, rather than on other estimators. In fact, the purity showed a smaller improvement, going from 87.9% to 88.4%. Whereas, by also removing interlopers from the test set, we observed a significant increasing of purity (from 88.4% to 94.5%), due to the exclusion of a large fraction of FPs ($\sim$ 56%).

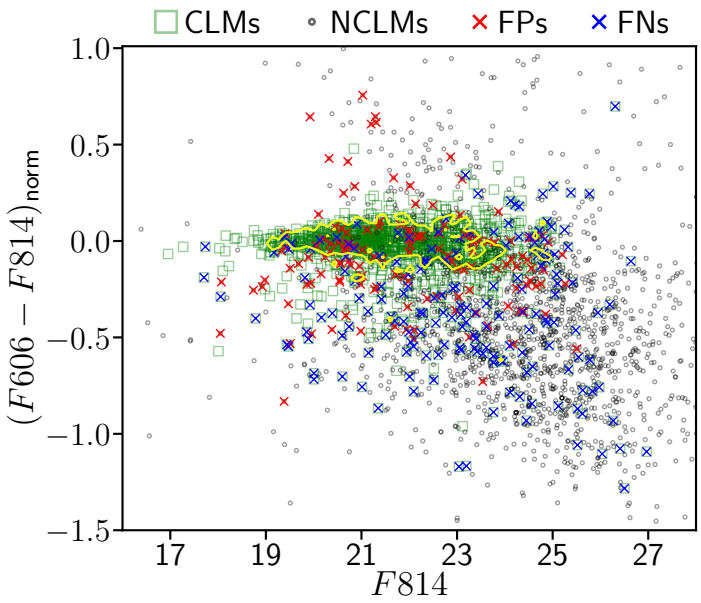


Figure 5.13: Colour-magnitude relation for the CLMs (green squares), with the overlapped distributions of False Positives (FPs, red crosses), False Negatives (FNs, blue crosses) and NCLMs (grey circles), for the sample of fifteen clusters (*stacked*). The yellow contour delimits the red-sequence at 1σ confidence level. Colours reported on the y-axis are corrected for the mean red-sequence of each cluster (see Sec. 5.2).

In summary, the FNs were mainly blue and faint CLMs, while FPs were typically red galaxies weakly bound to cluster. This was expected, given their under-representation in the dataset and the similarity between blue member with NCLMs and between interlopers with CLMs. We note, in fact, that we were mapping a population of cluster members in the central and highest density region of clusters, dominated by a high fraction of bright and red members. Nevertheless, the simple exclusion of fainter sources with $F814 > 25$ and $(F606 - F814)_{\rm norm} < -0.5$ improved the CNN performance.

Similar performances in terms of the distribution of false positives and negatives for sources with $F814 > 25$ and $(F606 - F814)_{\rm norm} < -0.5$ were obtained by the random forest classifier and the photometry-based Bayesian method. By comparing the behaviour of these three models on four clusters (R2248, M0416, M1206 and M1149), we found that the rate of blue FN is 28% for the Bayesian method and 25% for the random forest versus the 20% for the CNN. The rate of faint FN is 1% for the random forest and 6% for the Bayesian method versus the 5% of CNN. For what concerns FPs, the CNN, being the purest method, preserved the lowest contamination for both bluer and fainter members, with only four NCLMs classified as CLMs, compared with the 12 and 24 NCLMs for the Bayesian method and the random forest, respectively. This comparison, while it confirms the good performances of the CNN, also shows that the three methods have comparable efficiencies in the faint and blue region of the parameter space, which is likely due to undersampling of members in this region of the knowledge base, as pointed out above. This is due to the fact that the population of galaxies in the densest central cluster regions is brighter and redder than that of the less dense and outer cluster regions (see Annunziatella et al. 2014; Mercurio et al. 2016 for the specific study of M1206). Clearly, an improvement of the model's performances would require including member galaxies in the outer cluster regions and balancing the number of bluer and fainter members. In our case, even if the spectroscopic data cover more than two cluster virial radii, multi-band HST imaging with sufficient depth is only available in the central cluster regions.

	CLMs	FPs	FPs/NCLMs
Total Number	1187	139	0.084
$F814$ <25.0	96.6%	97.8%	0.131
$F814$ ≥25.0	3.4%	2.2%	0.005
$(F606-F814)_{\rm norm} \leq -0.5$	5.8%	4.3%	0.008
$(F606-F814)_{\rm norm} \leq -0.25$	13.7%	23.0%	0.033
$(F606-F814)_{\rm norm} < 0.0$	60.8%	70.5%	0.070
$(F606-F814)_{\rm norm} \geq 0.0$	39.2%	29.5%	0.161
	NCLMs	FNs	FNs/CLMs
Total Number	1655	181	0.152
$F814$ <25.0	62.7%	79.0%	0.125
$F814$ ≥25.0	37.3%	21.0%	0.950
$(F606-F814)_{\rm norm} < -0.5$	43.2%	35.4%	0.928
$(F606-F814)_{\rm norm} \leq -0.25$	64.6%	56.1%	0.656
$(F606-F814)_{\rm norm}$ <0.0	84.6%	83.4%	0.209
$(F606-F814)_{\rm norm}$ ≥0.0	15.4%	16.6%	0.065

Table 5.8: Summary of False Positive and False Negative distributions. Fractions of CLMs (Col. 2), False Positives (FPs) (Col. 3) and the ratio of FPs to NCLMs (Col. 4) as a function of magnitude (*second and third row*) and colours (*fourth to sixth row)*. The total number of spectroscopic CLMs and FPs are quoted in the first row. Fractions are computed only for sources whose $F814$ and $F606$ magnitudes are available (∼ 84% of the whole dataset). Similar fractions for NCLMs, FNs (False Negatives) and FNs/CLMs are quoted in the bottom half of the table. This table can be compared with Fig. 5.12 and Fig. 5.13 where FPs and FNs are plot as a function of magnitude $F814$ and normalised colour $(F606-F814)_{\rm norm}$.

5.5 Selection of member candidates

The experiments described in the previous sections are mostly focused on the classification efficiency and limits of the image-based CNN approach and evaluating its dependence from observational parameters such as redshift, number of CLM, photometric band compositions, magnitude and colour. In this section, we are mainly interested in evaluating the degree of generalisation capability of the trained CNN in classifying new sources as cluster members, a step process that is commonly referred to as *run* in the ML context. The training set was constructed by combining all clusters with the *mixed** band configuration. In order to maximise the parameter space sampling, we did not use the k-fold approach, instead we exploited the whole spectroscopic source ensemble to train the network, by just excluding the validation set used for the regularisation processes (see Sec. 3.6.5), i.e. we did not apply any training-testing split.

Similarly to what was done to build the knowledge base (see Sec. 5.2), for the *run* set we used squared cut-outs ∼ 4″ across, centered on the source positions as extracted by SExtractor (Bertin & Arnouts 1996), once spectroscopic sources have been excluded from the SExtractor catalogues. Thus, the *run* set was composed by 16156 unknown sources. The CNN identified a total of 1418 members with $F814 \leq 25$ mag, which is approximately the magnitude limit of the spectroscopic members (only ∼ 3% of spectroscopic members has $F814 > 25$). Within this CLM candidate sample: 158 (i.e. 11%) sources have normalised colour ≤ −0.25, while just 27 (i.e. 2%) are very blue sources with normalised colour ≤ −0.5, such quantities are comparable with the fractions shown in Tab. 5.8. Finally, we found that ∼ 48% of candidate CLMs have membership probabilities larger than 90%. The magnitude ($F814$) distribution and the colour-magnitude relations ($(F606-F814)_{\rm norm}$ versus $F814$) of both spectroscopic and

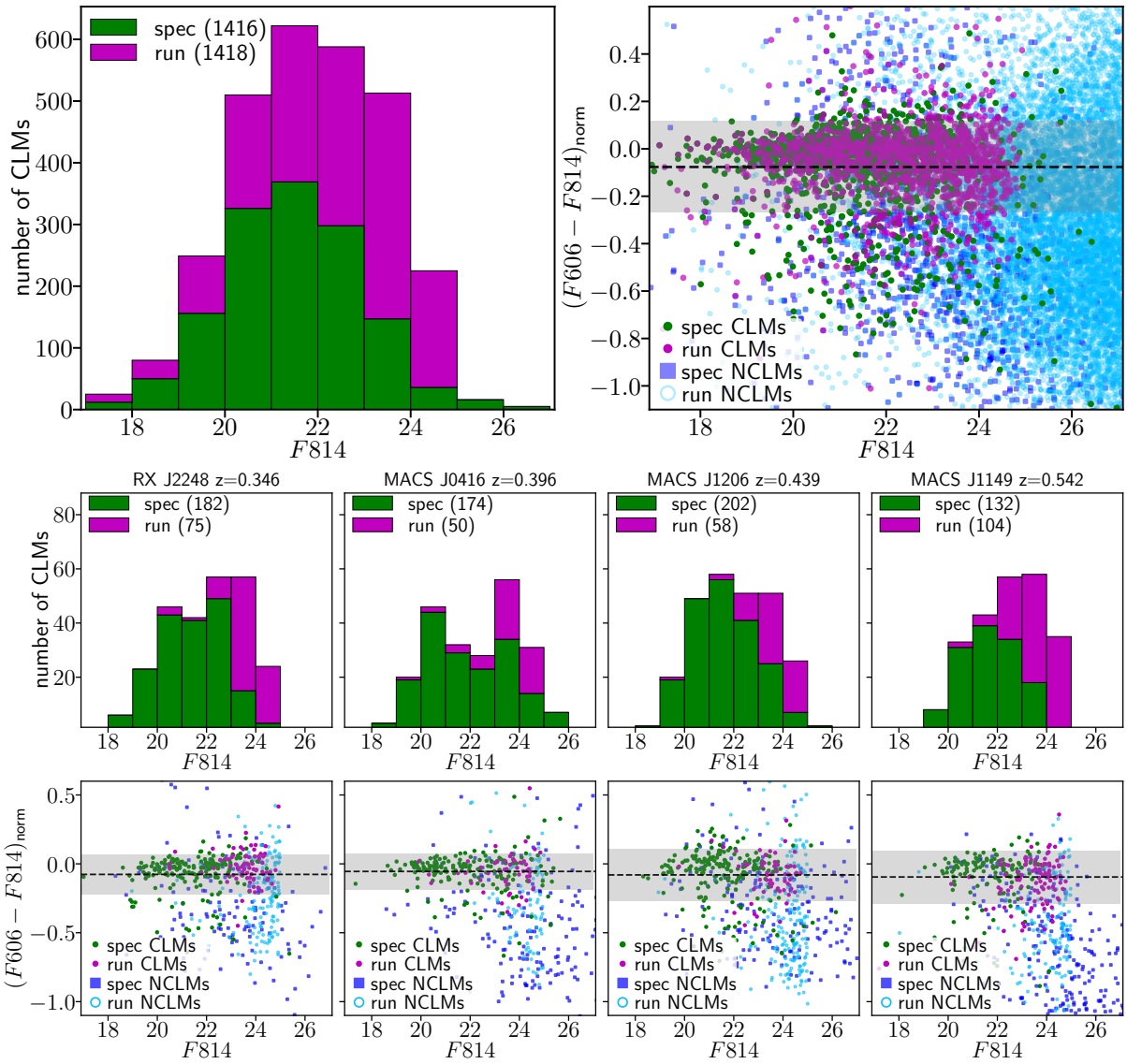


Figure 5.14: CNN membership predictions (*run*) together with spectroscopic sources: the $F814$ distribution is shown in the upper left panel and in the second row panels; the normalised colour − magnitude sequence is shown in the upper right panel and in the last row. The magnitude distributions and the colour-magnitude sequences are plotted: (*i*) by stacking members of all the involved clusters (first row panels) and (*ii*) individually for four cluster: R2248, M0416, M1206 and M1149 (second and third row panels). Spectroscopic CLMs are shown in green, candidate members in purple, spectroscopic NCLMs with blue squares and candidate NCLMs with open cyan circle. We only plot identified members with $F814 \leq 25$ mag. The grey region within the CM diagrams limits the area corresponding to $\pm 1\sigma$ from the median (dashed horizontal line) of $(F606 - F814)_{\rm norm}$.

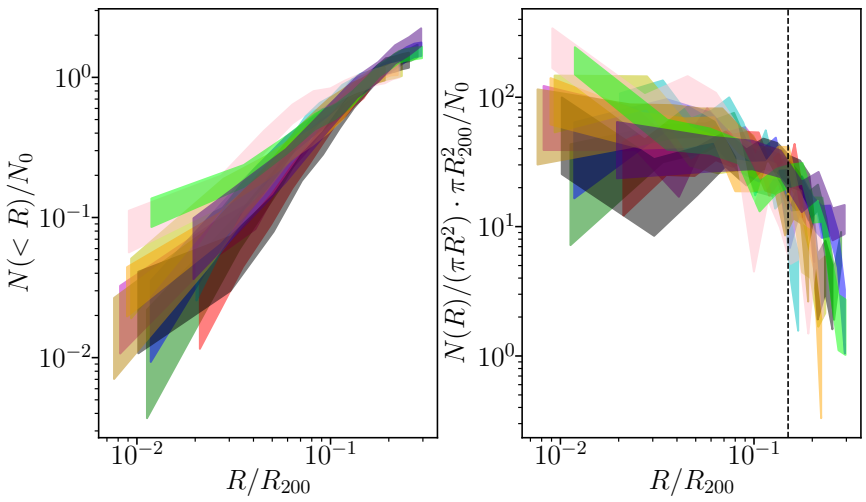
	Complete sample	$F814 < 25.0$ & $(F606 - F814)_{\rm norm} \geq -0.5$	$F814 < 25.0$ & $(F606 - F814)_{\rm norm} \geq -0.5$ & $(\lvert v\rvert \leq 3000 \vee \lvert v\rvert > 6000)$
true CLMs	1187	1100	1100
pred CLMs	1145	1130	1057
TPs	1006	999	999
FPs	139	131	58
FNs	181	101	101
pur	87.9%	88.4%	94.5%
compl	84.8%	90.8%	90.8%
F1	86.3%	89.6%	92.6%

Table 5.9: Comparison among CNN performances considering the whole sample (Col. 2), by removing sources with $F814 \geq 25$ and $(F606 - F814)_{\rm norm} < -0.5$ (Col. 3) and by additionally excluding interlopers (Col. 4).

predicted CLMs are shown in Fig. 5.14 both by stacking all clusters (first row panels) and by splitting the predictions made in the FoV of four clusters: R2248, M0416, M1206, and M1149 (second and third row panels). The magnitude distributions indicate that the CNN was able to complete the spectroscopic CLMs sample down to $F814 = 25$. This was also confirmed by the analysis of the colour-magnitude diagrams, which show that the photometrically identified CLMs complete the spectroscopic red-sequence at $F814 < 25$, emphasising the CNN capability to disentangle CLMs from background objects. We counted also the number of recognised CLMs within 1, 2, and 3σ from the median of normalised colour $(F606 - F814)_{\rm norm}$, respectively equal to 1156, 1336 and 1382 (i.e. 81.5%, 94.2% and 97.5%).

Finally, to further validate the galaxy cluster selection, we used both spectroscopic members and candidate CLMs identified by CNN to estimate the cumulative projected number of cluster members and the differential number density profiles (Fig. 5.15). According to our previous analysis, we excluded candidate CLMs with $F814 > 25$ mag, where only $\sim 3\%$ of spectroscopic members were present. To properly compare profiles of clusters with different virial masses, we computed the values R_{200} from of the values of M_{200c} obtained by Umetsu et al. (2018) with independent weak lensing measurements[4]. We then computed all profiles as a function of the projected radius in units of R_{200} and rescaled them by the number of members, N_0, found within the radius $R/R_{200} = 0.15$ in each cluster. In Fig. 5.15, we showed the cumulative projected number and the normalised projected number density profiles of cluster members after applying such renormalisations, where the shaded areas correspond to 68% confidence levels. Interestingly, we found that the radial distributions of all clusters followed a universal profile, including M0416, which is an asymmetric merging cluster. Our profiles, properly normalised, indicated the self-similarity of mass cluster profile expected in N-body simulation due to the scale-free nature of gravity (Kravtsov & Borgani 2012). We noted that a similar relation among rescaled projected mass profiles was found in Bonamigo et al. (2018) and Caminha et al. (2019), using strong lensing modelling. This result confirms that our methodology was able to identify the CLM population with a high degree of purity and completeness. Based on this accurate CNN membership, we are planning to analyse the luminosity and mass functions of a larger sample of clusters to study the ensemble variance of the galaxy number profiles. We expect our method to yield more robust results than those obtained using photometric redshifts.

[4]We note again that $R_{200} = \left(\frac{2G}{H(z_{cl})^2}\frac{M_{200c}}{200}\right)^{\frac{1}{3}}$, where $H(z_{cl})$ is the Hubble constant computed at the cluster redshift.



	spec	run	N_0	$R_{200}[Mpc]$
M1206	202	58	156	2.0±0.1
R2248	182	75	185	2.3±0.2
M0416	174	50	134	1.9±0.1
M1149	132	104	171	2.4±0.2
A383	82	102	168	1.9±0.3
R2129	68	103	143	1.8±0.2
M2137	63	100	116	1.9±0.2
M1115	116	156	219	2.2±0.2
M1931	79	64	119	1.9±0.2
R1347	57	145	178	2.7±0.1
M0329	104	199	189	1.9±0.1
M1311	65	116	74	1.3±0.0

Figure 5.15: Cumulative (left) and normalised (right) projected number of CLMs for 12 clusters (with available M_{200c} in Umetsu et al. 2018), including spectroscopic CLMs and candidate members identified by CNN (limited to $F814 \leq 25$ mag). The areas correspond to the 68% confidence level regions. All profiles are normalised by the number N_0 of members with $R < 0.15\,R_{200}$ in all clusters. The number of spectroscopic, CNN-identified members ("run"), N_0 values and the adopted values of R_{200} are quoted in the right-side legend. The dashed line in the right panel corresponds to $R = 0.15R_{200}$.

5.5.1 Selection of members in RELICS clusters

As final application of our trained network, we performed a *run* on 33 RELICS clusters (Coe et al. 2019), with central redshift ranging in $(0.20, 0.87)$. A layout of the involved RELICS clusters is shown in Fig. 4.8 in Chap. 4. Globally, the run set is composed by 39586 sources. Tab. 5.10 summarises the number of objects in each cluster, the amount of CLMs and NCLMs identified by the CNN, together with information about the photometric distribution of the candidate members. We restricted the candidate CLM sample to sources with $F814 \leq 25$ mag, identifying 5988 objects as cluster members. The resulting member magnitude distribution and the normalised colour-magnitude relation for both candidate CLMs and NCLMs are shown in Fig. 5.16. By comparing this colour-magnitude sequence with the one in the upper right panel of Fig. 5.14, members appear to be spreading out the same colour range: 86.4% of CLMs are within 1σ from the median colour of spectroscopic members (98.2% within 3σ); the CNN classified as members 602 sources with $F814 \in [24, 25]$ ($\sim$ 10%); a large fraction ($\sim$ 50%) of candidate members are red (i.e. $(F606 - F814)_{\rm norm} \geq 0$ mag), while just $\sim$ 6.5% are very blue objects with $(F606 - F814)_{\rm norm} < -0.25$ mag. Such quantities are comparable with the spectroscopic set of members (see Tab. 5.8), given the FPs and FNs analysis carried in Sec. 5.4: the oversampling of red members is partially due to the interlopers which increase the FP rate in the redder region, while the undersampling of blue members can be traced back to the high FN rate in the bluer region of parameter space.

5.6 Cluster member selection with ground-based images

In this section we present the identification of cluster members achieved by exploiting images acquired by SUBARU Suprime-Cam, comparing, in this way, CNN classification capabilities carried out with two different instruments. With the upcoming data of the Vera Rubin Observatory, we wanted to explore the network behavior in presence of ground-based wide field images, analysing the trade-off between the increasing of the available sources, due to the expanding

Cluster		$z_{cluster}$	N	CLMs	NCLMs	CLM with $F814 \geq 24$	CLMs with $(F606 - F814)_{\text{norm}}$ ≥ 0	< 0	< -0.25	< -0.5
A2163	Abell2163	0.203	1904	251	1653	7	144	107	1	0
A520	Abell520	0.203	1875	119	1756	3	65	54	3	0
R1514	RXC J1514.9-1523	0.223	862	113	749	8	60	53	3	0
A1763	Abell1763	0.228	1083	158	925	13	81	77	1	0
P171	PLCK G171.9-40.7	0.270	691	176	515	1	88	88	9	1
A1758	Abell1758a	0.280	2827	278	2549	27	137	141	3	0
A697	Abell697	0.282	861	151	710	6	80	71	4	0
R0232	RXC J0232.2-4420	0.284	1026	156	870	11	82	74	1	0
A2813	Abell2813	0.292	1040	163	877	16	97	66	1	0
A2537	Abell2537	0.297	733	167	566	11	84	83	9	0
AS295	AbellS295	0.300	1005	180	825	23	91	89	10	0
A1300	Abell1300	0.308	999	141	858	11	63	78	5	0
R0142	RXC J0142.9+4438	0.341	1038	167	871	19	87	80	11	0
M0035	MACS J0035.4-2015	0.352	1063	144	919	18	77	67	5	0
M0308	MACS J0308.9+2645	0.356	845	218	627	14	103	115	10	1
R0949	RXC J0949.8+1707	0.383	1060	194	866	19	101	93	7	1
P287	PLCK G287.0+32.9	0.390	1210	246	964	31	119	127	6	0
SM0723	SMACS J0723.3-7327	0.390	1114	142	972	10	66	76	11	2
R0032	RXC J0032.1+1808	0.396	932	239	693	29	120	119	16	2
R2211	RXC J2211.7-0350	0.397	925	222	703	24	107	115	14	1
M0159	MACS J0159.8-0849	0.405	1038	185	853	27	100	85	7	1
A3192	Abell3192	0.425	1168	210	958	33	96	114	16	1
M0553	MACS J0553.4-3342	0.430	1399	192	1207	29	98	94	10	5
S0254	SPT-CLJ0254-5857	0.438	1172	156	1016	19	74	82	18	1
M0417	MACS J0417.5-1154	0.443	1050	176	874	23	83	93	16	2
R0600	RXC J0600.1-2007	0.460	1231	254	977	31	138	116	15	0
P308	PLCK G308.3-20.2	0.480	1090	204	886	17	99	105	25	2
P004	PLCK G004.5-19.5	0.540	2065	205	1860	23	94	111	30	9
R0018	RXC J0018.5+1626	0.546	1191	222	969	27	102	120	33	12
W0137[(a)]	WHL 0137-08	0.566	1046	170	876	30	70	100	26	9
P209	PLCK G209.79+10.23	0.677	1040	150	890	15	74	76	11	6
P138	PLCK G138.61-10.84	0.702	1016	56	960	3	25	31	20	12
A0102	ACT-CLJ0102-49151	0.870	1987	183	1804	24	88	95	34	14
TOTAL			39586	5988	33598	602	2993	2995	391	82
FRACTION [%]				15.1	84.9	10.0	50.0	50.0	6.5	1.4

Table 5.10: Summary of the run performed on 33 RELICS clusters. The name, the short name of the clusters and their redshift are reported in the first 3 columns; the amount of sources in the run set, the identified CLMs and NCLMs are listed in columns 4 to 6; columns from 7 to 11 show the number of faint CLMs ($F814 \geq 24$), the amount of CLMs with $(F606 - F814)_{\text{norm}}$ respectively ≥ 0, < 0, < -0.25 and < -0.5 mag.

[(a)] The cluster WHL 0137-08 is also known as WHL J24.3324-8.477.

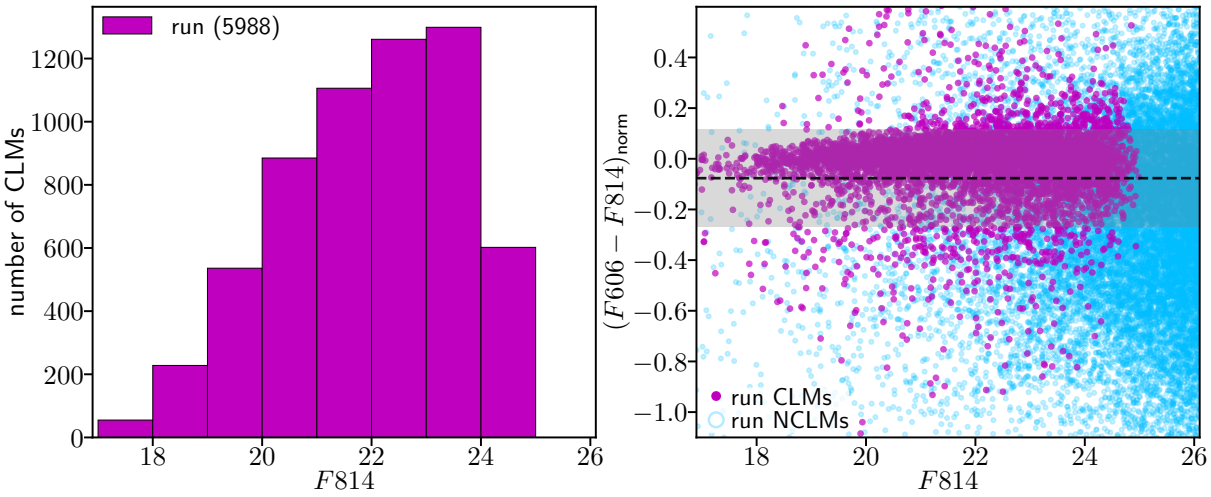


Figure 5.16: CNN membership predictions (*run*) on the RELICS clusters, represented as (i) CLMs distribution of $F814$ magnitudes (left panel), (ii) normalised colour - magnitude sequence for both CLMs and NCLMs (right panel). Candidate CLMs are shown in purple, while candidate NCLMs with open cyan circle. We only plot identified members with $F814 \leq 25$ mag. The grey region within the CM diagrams limits the area corresponding to $\pm 1\sigma$ from the median (dashed horizontal line) of $(F606 - F814)_{\rm norm}$, estimated from the spectroscopic CLM sample.

of the FoV sizes (SUBARU FoV is $\sim 30'$, with respect to the HST FoV of $\sim 3'$), and the degradation of the image quality, due to the seeing and to different sampling (SUBARU images have a sampling of $0.200''$/pixel, with respect to the HST sampling of $0.065''$/pixel).

In order to maximise the number of filters available for all the involved clusters, we restrict the analysis to 4 clusters: Abell 209 (A209, with central redshfit $z_{cl} = 0.209$), MACS J1115+0129 (M1115, with central redshift $z_{cl} = 0.352$), MACS J1206-0847 (M1206, with central redshift $z_{cl} = 0.439$) and MACS J0329-0211 (M0329, with central redshift $z_{cl} = 0.450$). For these clusters, B, V, R, I and Z bands are all available. These four clusters are shown in Fig. 4.5 in Chap. 4.

Similarly to what has been done for HST images (see Sec. 5.2), in order to label the dataset, we used the extensive spectroscopic information obtained by combining CLASH-VLT VIMOS programme with MUSE archival observations, identifying as CLM a source having rest-frame velocity separation $|v| \leq 3000\,{\rm km\,s^{-1}}$ (see Eq. 5.1). Concerning the image extraction, for each spectroscopic source, we extracted a squared cutout centred on source position with the same side used for HST cutouts, i.e. equals to $4''$, corresponding to 20 pixels. A sample of the dataset is shown in Fig. 5.17, where sources extracted from the involved clusters are split between members, interlopers and non-members. By comparing these cutouts with the ones drawn out from HST images (Fig 5.1), it is clear the reduction of the image quality: SUBARU images appear as blurred versions of the HST cutouts. The number of samples for each involved cluster is shown in Tab. 5.11.

In order to train the network, we opted for stratified k-fold approach, following the same method applied for the HST images (see Sec. 5.2), as described in Sec. 3.6.5: 10% of sample are reserved for the validation set, the rest of which has been partitioned into $k = 10$ non-overlapped folds, 15% of each training fold has been augmented through rotations and flips (with a resulting augmentation factor equals to 1.75). Finally, since the number of NCLMs significantly exceeds the number of CLMs, before the k-fold splitting, we extracted subset of them (equals to the 120% of the CLMs set), in order to balance the classes.

Within this experiment, we trained the CNN individually for each cluster and by stacking all the involved clusters, testing, in this way, the network capabilities to generalise the learning

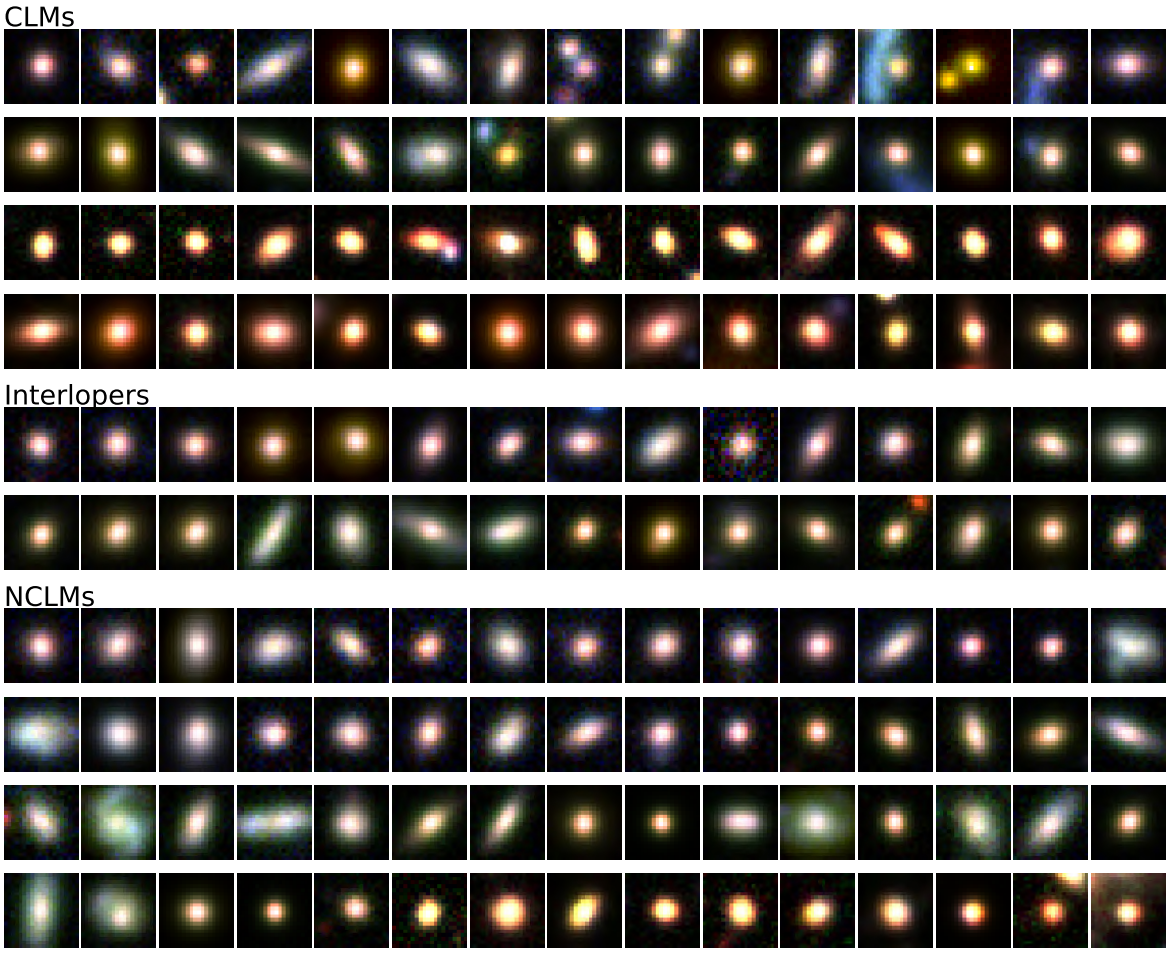


Figure 5.17: Examples of RGB cutouts of cluster members, interlopers and non-members extracted from SUBARU images (B, V, R bands). To emphasise fainter sources, images have been stretched by clipping values within $\pm 3\sigma$ and then normalised. Cutouts are $4''$ across.

over the considered cluster redshift range $(0.20, 0.46)$. The results are shown in Tab. 5.12. Concerning the global performance, achieved by stacking all the involved clusters, by comparing the results with those carried out with HST data (see Tab. 5.2), we measured a decreasing of $\sim 6\%$, $\sim 4\%$ and $\sim 7\%$, respectively for average efficiency, CLM and NCLM F1-score. Despite this reduction, the network is able to identify members with an acceptable trade-off between purity and completeness, characterised by an F1-score ranges in $\sim 82\%$, $\sim 85\%$. Respect to the network behaviour trained with HST images, in this experiment we found that the CNN is more complete than pure, with differences between $\lesssim 1\%$ and $\sim 8\%$. Regarding the performed comparison between the "cluster-by-cluster" training and the global training, despite the photometric and morphological differences between members at different redshift, we measure an average 5% improvement with the enlarging of the KB, evaluated by cross-comparing all the statistical estimator couples. As already observed for the analysis carried out with HST images, even this experiment confirms the classification capabilities increasing with the enlarging of the KB, particularly the involving of members belonging to clusters at different redshift improves the network generalisation degree respect to an individual "cluster-by-cluster" training.

The network misclassifications (i.e. False Positives and False negatives) are shown in Fig. 5.18 in terms of magnitude and normalised colour logarithmic distributions (left and central panels), as well as in terms of the normalised colour - magnitude relation (right panel). While, a summary of FP and FN percentages as a function of magnitude and colour is reported in

	Abell 209 (A209)	MACS J1115+0129 (M1115)	MACS J0329-0211 (M1206)	MACS J1206-0847 (M0329)	TOTAL
$z_{cluster}$	0.209	0.352	0.4390	0.450	
z_{min}	0.196	0.338	0.425	0.435	
z_{max}	0.221	0.365	0.454	0.464	
CLM	996	724	672	508	2900
NCLM	1485	1767	2168	1317	4569

Table 5.11: SUBARU cluster sample description. The cluster central redshifts and the spectroscopic range to identify CLMs are listed in the first, second and third row. The numbers of CLMs and NCLMs for each involved cluster are reported in the fourth and fifth row.

	%	A209	A209*	M1115	M1115*	M1206	M1206*	M0329	M0329*	STACKED
	AE	86.3	**86.5**	76.6	**81.9**	80.0	**82.5**	73.2	**84.8**	84.1
	pur	83.7	**85.7**	72.9	**78.7**	76.1	**79.0**	71.6	**81.0**	81.5
CLM	compl	**88.6**	86.0	81.1	**84.9**	84.6	**86.1**	72.7	**89.1**	86.3
	F1	**86.1**	85.8	76.8	**81.7**	80.1	**82.4**	72.2	**84.9**	83.9
	pur	**89.0**	87.2	80.8	**85.2**	84.5	**86.3**	74.7	**89.1**	86.8
NCLM	compl	84.3	**87.0**	72.5	**79.1**	75.9	**79.3**	73.7	**80.9**	82.2
	F1	86.6	**87.1**	76.4	**82.0**	79.9	**82.6**	74.2	**84.8**	84.4

Table 5.12: CNN percentage performances with SUBARU images. Statistical estimators are related to the individual training of each cluster and to the global training achieved by staking all the four clusters (then evaluated separately for each cluster, flagged with an asterisk). The best results for each cluster are highlighted in bold.

Tab. 5.13. As it can be seen from the distributions in Fig. 5.18, FPs and FNs do not gather in specific regions of the parameter space, rather they reflect the CLMs and NCLMs distributions, at least down to $i = 23$ mag and $(v-i)_{\rm norm} > -0.5$ mag; indeed, as expected, there is an excess of FNs in the member faint end ($i \geq 23$) and in the bluest region of the parameter space. ($(v-i)_{\rm norm} < -0.5$ mag). Such behaviour is also evident in the colour-magnitude diagram (right panel in Fig. 5.18). Concerning the FPs, their colour distribution nearly coincides with that of CLMs, suggesting that FPs are galaxies photometrically similar to the members (for both red and blue regions). More in general, FPs strictly follow the NCLM trends as shown in Col. 6 of Tab. 5.13: the fraction FP/NCLM is almost constant in the considered magnitude and colour ranges ($\sim$ 20%). Finally, regarding the interlopers, only 50 of them (i.e. $\sim$ 30%) have been

	CLMs	**NCLMs**	FP	FN	FP/NCLM	FN/CLM	Interlopers (FP)
Total Number	2020	2538	517	241	20.4%	11.9%	173 (123)
i <23.0	90.9%	82.3%	84.5%	78.0%	20.9%	10.2%	82.7% (82.1%)
i ≥23.0	9.1%	17.7%	15.5%	12.0%	17.8%	28.8%	17.3% (17.9%)
$(v-i)_{\rm norm} \leq -0.5$	20.0%	36.6%	39.8%	42.3%	21.0%	25.2%	54.3% (73.2%)
$(v-i)_{\rm norm} \leq -0.25$	34.8%	61.2%	62.7%	68.5%	20.8%	23.5%	69.4% (72.5%)
$(v-i)_{\rm norm} < 0.0$	69.1%	81.8%	83.7%	86.3%	20.9%	14.9%	62.4% (72.5%)
$(v-i)_{\rm norm} \geq 0.0$	30.9%	18.2%	16.3%	13.7%	18.2%	5.3%	37.6% (62.5%)

Table 5.13: Fractions of CLMs (Col. 2), NCLMs (Col. 3), FPs (Col. 4), FNs (Col. 5), FPs to NCLMs ratio (Col. 6), FNs to CLMs ratio (Col. 7) and interlopers (Col. 8) as a function of magnitude (*second and third row*) and colours (*fourth to seventh row)*. The total number of spectroscopic samples are quoted in the first row. The fractions of false positive interlopers have been reported between brackets in the last column. These numbers are computed only for sources whose i and v magnitudes are available ($\sim$ 70% of the whole dataset).

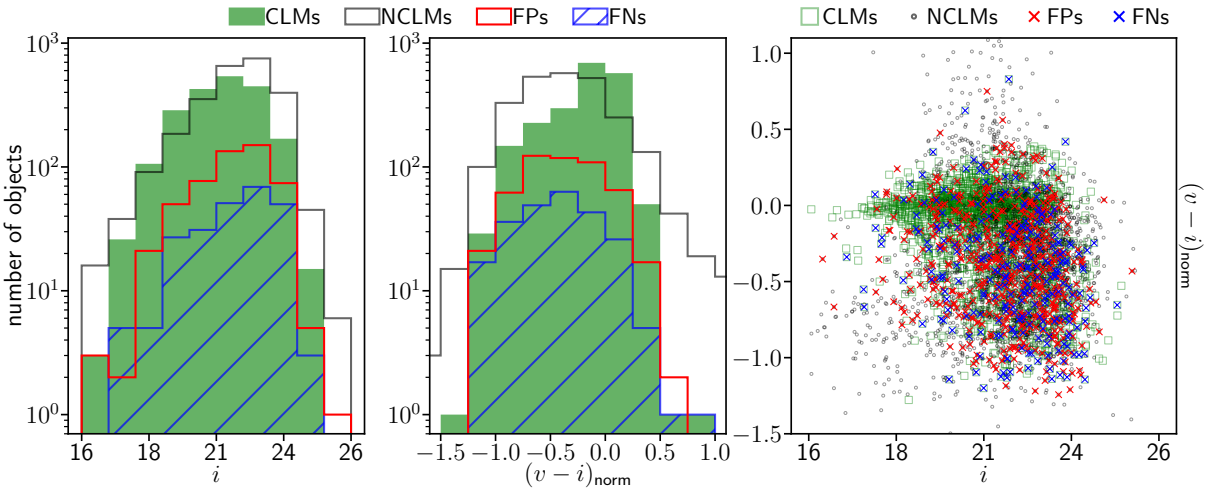


Figure 5.18: Magnitude (left panel) and colour (central panel) logarithmic distributions, together with colour-magnitude relation (right panel), of FPs (red) and FNs (blue), overlapped to the CLM (green) and NCLM (grey) distributions for the four clusters involved in our analysis. The normalised colour $(v-i)_{\rm norm}$ is obtained by applying the correction for colour-magnitude relation for each cluster (Eq. 5.2). Tab. 5.13 outlines such results.

correctly classified, resulting that 22% of FPs are interlopers.

Respect to CNN performances achieved exploiting HST imaging, characterised by strong dichotomies (faint vs. bright and blue vs red sources, see Figures 5.12, 5.13 and Tab. 5.8), in this experiment, which involved wide-field ground-based images, we have not found any similar behaviour: although the performances are lower, miscassification appears to be distributed in the whole parameter space. Another example is represented by interlopers: even if $\sim$ 70% of them are FPs, they populate both the red and blue region of the parameter space ($\sim$ 45% of them are red galaxies, while in the HST experiments the interloper red population composed the $\sim$ 84% of the whole interloper ensemble). Thus, respect to experiments involving HST data, the performance reducing achieved with SUBARU supreme-cam imaging is due to the different quality of the images. Indeed, the CNN trained with HST image was able to detect up to the 90% of the member population by disentangling them from background and foreground sources, which can not be separated with worse resolution imaging; however, the larger source sampling achieved by SUBURU, due to its wider FoV, allows to build KB with which the CNN achieved more uniform results, in terms of magnitude and colour distributions.

5.7 Conclusions

In this work, we carried out a detailed analysis of CNN capabilities to identify members in galaxy clusters, disentangling them from foreground and background objects, based on imaging data alone. Such a methodology, therefore, avoided the time consuming and challenging task of building photometric catalogues in cluster cores, since it just requires the position of sources, on which cutouts are extracted. We used optical-NIR high quality HST images, supported by MUSE and CLASH-VLT spectroscopic observations of fifteen clusters, spanning the redshift range $z_{\rm cluster} = (0.19, 0.60)$. The redshift for a large fraction of the spectroscopic objects ($\sim$ 80%) has been estimated using MUSE data cubes, so the source set should not be affected by bias in the selection. We used this extensive spectroscopic coverage to build a KB by combining CLMs and NCLMs. We performed several experiments by consecutively varying the HST band combinations and the set of training clusters to study the dependence of DL efficiency on (i)

band configuration, training size and cluster redshift (*EXP1*); (*ii*) the magnitude and colour of cluster galaxies (*EXP2*). We also compared the CNN performance with other methods (random forest and Bayesian model), based on photometric measurements (*EXP3*). The main results can be summarised as follows:

- Despite members belonging to clusters spanning a wide range of redshift, the CNN achieved a purity-completeness rate $\gtrsim 90\%$, showing a stable behaviour and a remarkable generalisation capability over a relatively wide cluster redshift range (Sec. 5.3.1).

- The CNN efficiency was maximised when a large set of sources was combined with HST passbands, including both optical and infrared information. The robustness of the trained model appeared reliable even when a subset of clusters was moved from the training to the blind test set, causing a small drop ($< 5\%$) in performance. As predictable, we found that CNN performs better on bright sources than for faint objects, as well as the unsuccess rate is higher for the blue galaxies than for red ellipticals. However, the results maintained the purity, completeness and F1 score greater than 72% (Table 5.6 in Sec. 5.3.2).

- By using images, rather than photometric measurements, the CNN technique was able to identify CLMs with the lowest rate of contamination and the best trade-off between purity and completeness, when compared to photometry-based methods, which instead require a critical fine-tuning of the classification probability.

- The false negatives, that is, the NCLMs wrongly classified as CLMs were mainly blue and faint. This was simply the result of their under-sampling in the training dataset, as well as their similarity with NCLMs. However, by excluding sources with $F814 > 25$ mag and $(F606 - F814)_{\rm norm} < -0.5$ mag, the CNN performance improved significantly gaining more 6% and 1%, respectively in term of completeness and purity (Sec. 5.4). These performances reflected the capability of the CNN to classify unknown objects, from which a highly complete and pure magnitude limited sample of candidate CLMs could be extracted for several different applications in the study of the galaxy populations and mass distribution of galaxy clusters via lensing techniques.

- The usage of ground-based wide-field imaging (e.g. acquired by SUBARU Suprime-Cam, Sec. 5.6) lead to a decreasing of performance, up to 10%, with F1-scores $\sim 84\%$; however, the involving of sources in the outskirt of clusters allows the sampling of the bluer region of the parameter space, resulting into performance less dependent on colour.

In this work, we limit our analysis to the center of galaxy clusters, where we can exploit an extensive spectroscopic coverage and the complete optical-NIR spectral information; however, this method can be extended in external regions, which host a larger fraction of blue members that could be misclassified by the network (since it has been mainly trained on red elliptical galaxies, as already mentioned). To solve this issue, we will apply other DL methods, such as deep auto-encoders (Goodfellow 2010) and conditional generative adversarial networks (Mirza & Osindero 2014), on wide-field ground-based lower resolution images to complement high quality HST imaging in cluster outskirts. Moreover, this method can be fruitfully extended to upcoming survey facilities, such as Vera Rubin Observatory (Ivezić et al. 2019) and Euclid (Laureijs et al. 2011). Finally, the results obtained in this work encourage the exploration of Deep Learning techniques to automatically estimate galaxy properties, such as stellar masses, sizes or magnitudes.

Chapter 6

Identification of Galaxy-Galaxy Strong Lenses in clusters

In this chapter I present the adopted strategy to identify the Galaxy-Galaxy Strong Lenses (GGSLs) with CNNs exploiting a dataset whose samples are generated by combining simulations with the complexity of observed clusters. This chapter will be the subject of a forthcoming publication.

6.1 Introduction

Strong gravitational lensing is a very powerful tool for studying galaxy evolution and cosmology. For example strong lenses have been exploited to analyse galaxies structures and their evolution (Treu & Koopmans 2002; Auger et al. 2010; Sonnenfeld et al. 2013) or to constrain the stellar initial mass function of galaxies (Spiniello et al. 2012; Sonnenfeld et al. 2019). Strong lensing acts as unique tool to reveal high-redshift magnified sources, otherwise undetectable (Vanzella et al. 2020, 2021), to measure cosmological parameters exploiting time delay observations (Grillo et al. 2018; Millon et al. 2020), to test cosmological paradigms (Diemand & Moore 2011; Meneghetti et al. 2020). Moreover, the multiple images detection in galaxy clusters (together with galaxy members identification) allows to build lensing models with which is possible to study the cluster mass distribution, the sub-halo populations and dark matter halos (e.g., Caminha et al. 2017b, 2019; Bergamini et al. 2019, 2021b).

Despite all these applications, strong lenses are rare, since they require the alignment of the lens with a foreground source. This low number of known events makes it difficult to use any Machine Learning approach to search for GGSL. Indeed, the most common technique for the GGSL identification has long been the visual inspection of candidates, selected via a colour-magnitude selection or with a semi-automated searching for arc-shaped and ring-shaped features (Le Fèvre & Hammer 1988; Jackson 2008; Sygnet et al. 2010; Pawase et al. 2014; Seidel & Bartelmann 2007; Gavazzi et al. 2014; Sonnenfeld et al. 2018). Recently, Sonnenfeld et al. (2020) proposed a crowdsourcing to find lenses using images of galaxies acquired within the Hyper Suprime-Cam survey (Aihara et al. 2018; Miyazaki et al. 2018), together with simulated lenses and non-lenses, to "train" $\sim$ 6000 volunteer citizens in the lens identification, their answers are then collected into a Bayesian framework to obtain a lens-probability. Given the large redshift range at which GGSLs are searched, their different morphology, colours and magnitudes the automatic finding of lens through artificial networks requires large simulated dataset (e.g., Metcalf et al. 2019). Recent attempts to exploit Deep Learning approaches have been proposed by Jacobs et al. (2019a,b), which searched for lensing events within the Dark Energy Survey (Dark Energy Survey Collaboration et al. 2016), and by Petrillo et al. (2017,

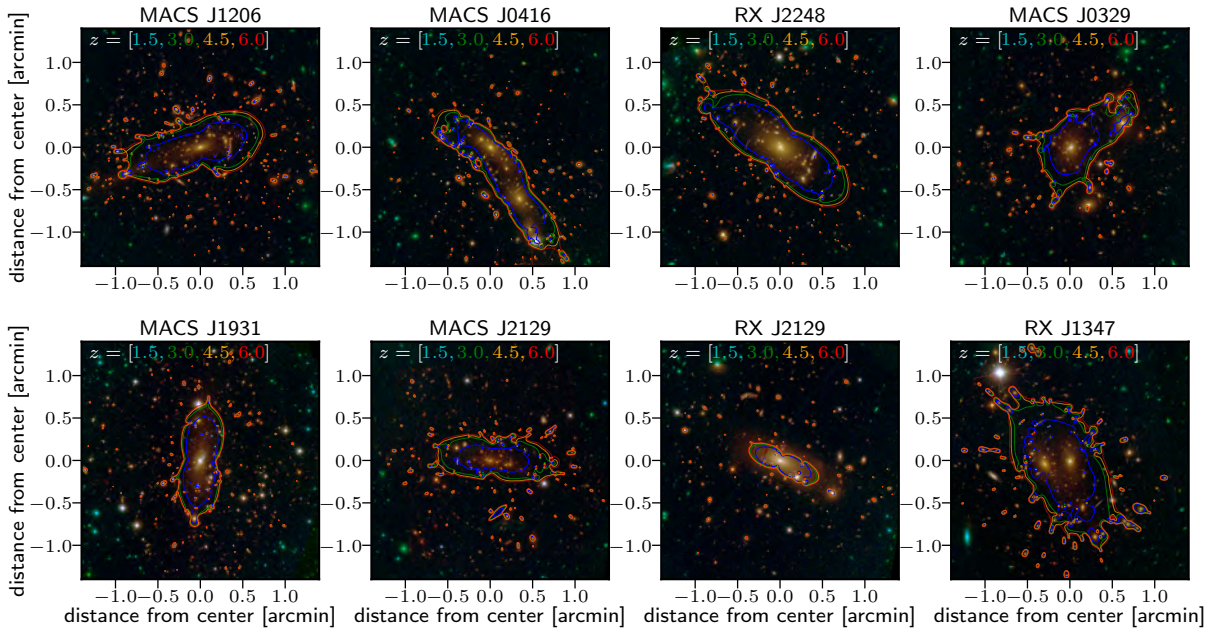


Figure 6.1: The involved clusters (also listed in Tab. 6.1). The tangential critical lines corresponding to 4 different redshfits, $z = [1.5, 3.0, 4.5, 6.0]$ (respectively coloured in cyan, green, orange and red), are overlapped to the cluster FoV.

2019b); Li et al. (2020, 2021b), which explored the Kilo Degree Survey (de Jong et al. 2015). Both works exploited Convolutional Neural Networks trained on images simulated by adding arcs and rings on the top of selected Luminous Red Galaxy cutouts.

In this work, we present a novel approach based on the injection of a Sérsic profile galaxy (Sérsic 1963, 1968) behind an observed HST member galaxy: by exploiting fitted cluster lens models, we tracked the light rays from the source plane to the lens plane, to the observer, obtained an GGSL example within the complexity of a galaxy cluster.

6.2 Building the dataset

As mentioned above, the restricted number of known GGSL events prevents from training any Machine Learning model, therefore, in order to build the Knowledge Base, we simulate a large dataset involving a variety of strong lensing configurations, by keeping the complexity and heterogeneity of the observed imaging data. In this work, we simulate GGSLs in 8 clusters with a redshift range (0.2 – 0.6), using three ACS bands, $F435$, $F606$, $F814$, from the HFF survey, when observations were available, from CLASH otherwise, as listed in Tab. 6.1.

To simulate strong lenses, we use the (reduced) deflection angle maps computed by tacking the gradient of the potential map (Eq. 2.32), once a cluster lens model has been fitted. The lens models were developed with the public software `LensTool` (Kneib et al. 1996; Jullo et al. 2007; Jullo & Kneib 2009), which determines the best-fit cluster lens model with a Bayesian Markov chain Monte Carlo approach, by minimising the difference between the multiple images observed positions and their predicted positions given the set of model parameter (see Sec. 2.3). In this way, given the cluster redshift (i.e. the distance from the cluster lens plane) and source redshift (i.e. the distance from the source plane), the tangential and radial critical lines are found by imposing that the magnification of the images goes to infinity (i.e. setting the denominators in Eq. 2.44 equal to zero). The magnification is estimated by computing the convergence (with Eq. 2.34) and the components of the shear (Eqs. 2.35 and 2.36).

Cluster		$z_{cluster}$	Survey	$M^{(a)}_{200c}[10^{14}M_{\odot}]$	N_{img}	$\Delta_{rms}['']$	ref
RX J2129+0005	R2129	0.234	CLASH	7.78±2.43	22	0.20	(1)
RX J2248-4431$^{(b)}$	R2248	0.346	HFF	19.81±5.97	55	0.55	(2)
MACS J1931-2635	M1931	0.352	CLASH	11.62±8.84	19	0.38	(1)
MACS J0416-2403	M0416	0.397	HFF	11.43±2.66	182	0.40	(3)
MACS J1206-0847	M1206	0.439	CLASH	15.05±3.20	82	0.46	(2)
MACS J0329-0211	M0329	0.450	CLASH	12.70±2.19	23	0.24	(1)
RX J1347-1145	R1347	0.451	CLASH	35.40±5.05	20	0.36	(1)
MACS J2129-0741	M2129	0.587	CLASH	N/A	38	0.56	(1)

Table 6.1: Description of the cluster sample involved in the GGSL simulation. The name, the short name of the clusters and their redshift are reported in the first 3 columns. The fourth column shows the surveys from which images are extracted. N_{img} (Col. 5) is the number of multiple images used to constrain the model, Δ_{rms} is the root-mean-squared separation between the observed and the predicted multiple images positions. The reference lens model for each cluster can be found in the last column. $^{(a)}$ The halo mass measured by Umetsu et al. (2018).
$^{(b)}$ The cluster RX J2248.7−4431 is also known as Abell S1063.
(1) Caminha et al. 2019; (2) Bergamini et al. 2019; (3) Bergamini et al. 2021b.

In this work, we exploit the deflection maps computed from the cluster lens models fitted by Bergamini et al. (2019, 2021b) and Caminha et al. (2019) in 8 clusters, spanning a redshift range $\in (0.23, 0.59)$, listed in Tab. 6.1 and shown in Fig. 6.1, where the tangential critical lines corresponding to 4 different redshifts are overlapped to the cluster FoV. By exploiting these high-precision "macro" models, we use the complex cluster mass distribution (see e.g. the surface mass density in Fig. 2.5, from Grillo et al. 2015 for the M0416 cluster), which affects the formation of the lensing event. We note that in simulations of GGSL in the field (e.g., Petrillo et al. 2019a,b; Li et al. 2021b; Gentile et al. 2022), the effect of a cluster/group environment is often reproduced with an external shear added to a local galaxy mass model. In our case, we model the whole cluster deflection field by taking into account all the physical components that actually contribute to the cluster potential model: the cluster-scale mass component due to the dark matter cluster halo and the smooth hot-gas mass traced by Chandra X-ray data, the clumpy component of cluster galaxies (dark matter and baryons in each sub-halo), foreground-structures in the outer cluster region and line-of-sight mass contributions (Bergamini et al. 2019, 2021b). As a consequence, the Einstein radius associated to each galaxy member is larger than what it would have been in absence of the cluster deflection field. As an example, to quantify the effect of the cluster potential on a single galaxy, we used the lens model of M0416 (Bergamini et al. 2021b) to measure the mass of each galaxy within 3″, with and without the large-scale mass contribution, finding that a typical ratio between these masses is 3.8 ± 1.0, implying that the (median) mass contribution of the galaxy alone is just ∼ 26%.

An example of the simulation process is graphically shown in Fig. 6.2: given the source and the cluster redshift, we determined all the critical lines using our deflection maps (central panel in Fig. 6.2); galaxy lenses are chosen by selecting critical lines associated to galaxy members; the galaxy position on the source plane (β) is randomly extracted within a buffer surrounding the corresponding caustic (whose width is set to $0.5r_e$, left panel in Fig. 6.2). The brightness is computed assuming a Sérsic profile (Sérsic 1963, 1968), $I_s(\beta)$ (see Sec. 2.1.3), and, since the surface brightness is conserved (i.e. $I(\vec{\theta}) = I_s(\vec{\beta})$), the observed and the intrinsic surface brightness can be related using the lens equation (Eq. 2.26). So, the source image on the lens plane is reconstructed through ray-tracing: the ray positions $\vec{\theta}_{ij}$ on the lens plane are defined with a squared grid and the surface brightness $I_s(\vec{\beta}_{ij})$ is assigned to the element centred in $\vec{\theta}_{ij}$,

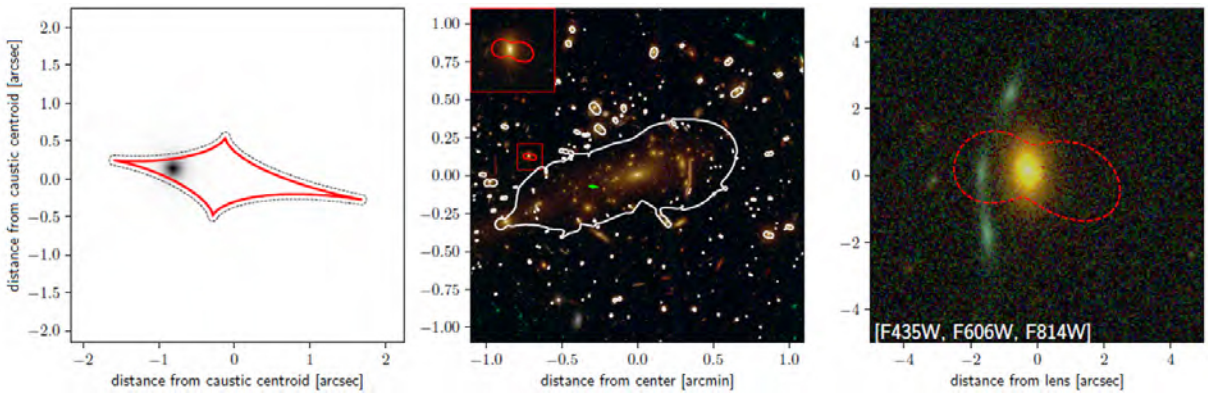


Figure 6.2: Example of GGSL simulation. *Left panel*: source plane at $z = 2.5$ where the caustic (in red) is plotted, together with the buffer (black dotted line) delimiting the injection region and the injected source ($F814 = 26.3$ mag, Sérsic index $n = 1.5$ and effective radius $r_{eff} = 0.14''$); the cutout is $\sim 4''$ across. *Central panel*: cluster lens plane, i.e. the cluster M1206 ($z = 0.439$) with all the critical lines (in white), among them, the selected critical line emphasised in the upper left $\sim 1''0$ cutout (coloured in red); the green spot indicates the position of the caustic on the lens plane; the cutout is $\sim 2'$ across. *Right panel*: simulated GGSL event results into an arc-shaped object partially surrounding the galaxy member; the critical line is overlapped to the image (red dotted line); the cutout is $\sim 10''$ across.

Parameter	Symbol	Extraction description
Position (source plane)	y_s	Extracted within a buffer around the caustic (width $0.5r_e$)
Source magnitude	m_{F814}	Sampled from CDF, $P(i_+)$, COSMOS + HST fields
Source redshift	z_s	Sampled from CDF, $P(z\|\Delta i_+)$, COSMOS
Effective Radius	r_e	$r_e = 2.54$ kpc, $z \leq 1$ $r_e = B(1+z)^\beta$, $z > 1$ (Shibuya et al. 2015)
Sérsic index	n	Extracted within $(1.0, 2.0)$
axis ratio	q	Extracted within $(0.2, 1.0)$
position angle	φ	Extracted within $(0.0, \pi)$

Table 6.2: List of Sérsic parameters and their adopted value range, for the injected sources.

by mapping this latter to the corresponding position on the source plane using the second order approximation of the lens equation (Eq. 2.45). The resulting image is convolved with the point spread function (PSF) and coadded to the observed cluster galaxy (right panel in Fig. 6.2); finally, a squared cutout centered on the centroid of the critical curve is extracted. The PSFs used in this work are measured by Tortorelli et al. (2018), as a combination (obtained with the software developed by Paolillo et al. 2011) of PSFs extracted from an ensemble of selected point-like sources in the field (HFF ACS bands of R2248 and M1149), with empirical PSFs made available within the PSF library developed by Anderson & King (2006). In this way, the lens galaxy is automatically selected by choosing the critical line.

Another example is displayed in Fig. 6.3, where the same source has been injected in the centroid of caustics at different redshifts, showing the typical reddening and thickening with the increasing of depth.

As already stated, we assumed that the injected source is a galaxy whose surface brightness is described by a Sérsic profile (see Sec. 2.1.3 and Eqs. 2.1, 2.6- 2.8), parametrised through an attribute tuple composed by: Sérsic index (n), effective radius (r_e), axis ratio (q), position angle (φ), which are added to the source redshift (z_s), its position on the source plane ($y_s = (y_{s_1}, y_{s_2})$) and the spectral flux density (F_ν) which is converted into instrumental count rate (N_e): for HST ACS system, the conversion between flux density and count rate is computed through (Ryon

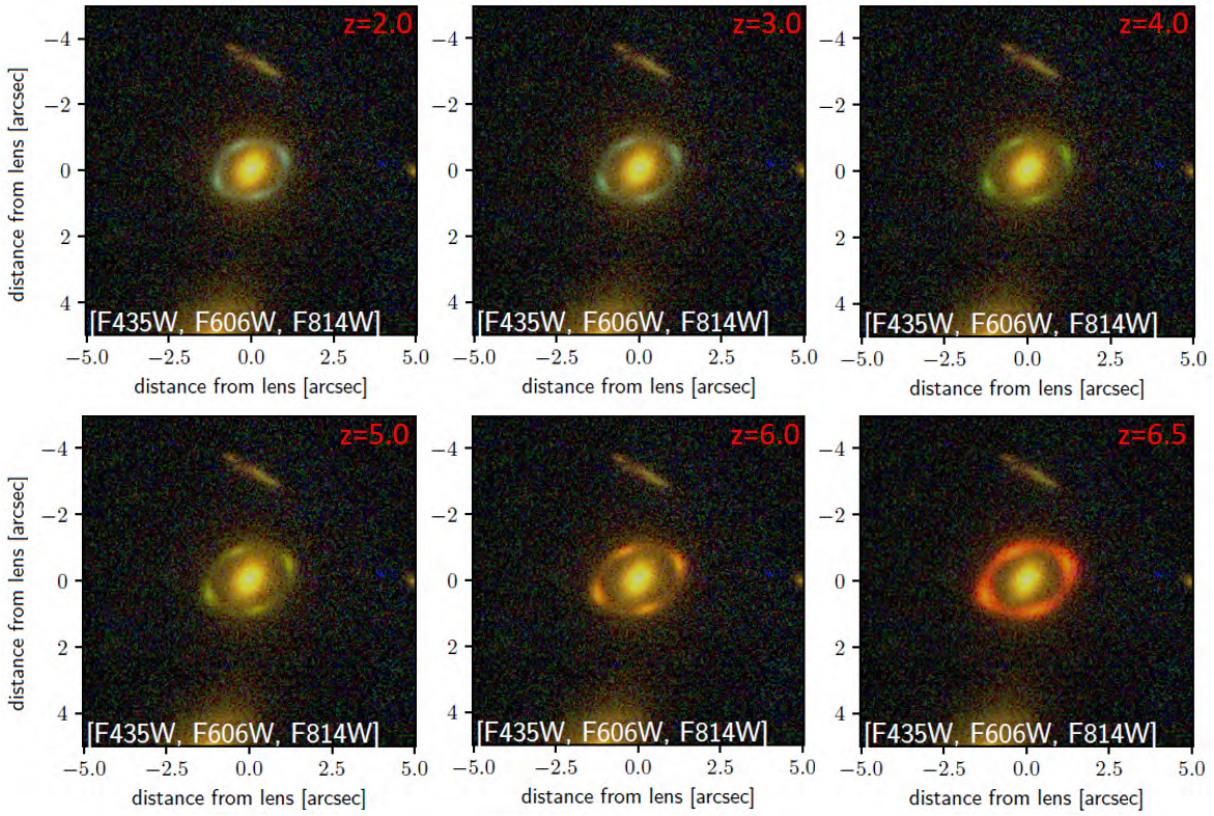


Figure 6.3: Example of a simulated Einstein Ring, obtained by injecting the same source in the centroid of caustics at different redshift. The source brightness has been computed assuming a Sérsic profile ($F814 = 26.0$ mag, Sérsic index $n = 1.0$, $r_{\rm eff} = 0.1''$).

2021):

$$
\begin{aligned}
mag_{AB} &= -2.5 \times \log F_\nu - 48.60 \\
mag_{AB} &= -2.5 \times \log N_e + \mathrm{ZP} \\
\mathrm{ZP} &= -2.5 \times \log(\mathrm{PHOTFLAM}) - 5 \times \log(\mathrm{PHOTPLAM}) - 2.4079
\end{aligned}
$$

where mag_{AB} is magnitude in the AB system in a certain passband in any of the ACS system, ZP is the instrumental zeropoint which is calculated with the header keywords: PHOTFLAM (the inverse sensitivity) and PHOTPLAM (pivot wavelenght). Thus the conversion is given by:

$$\log N_e = (\mathrm{ZP} + 48.6)/2.5 \times \log F_\nu \tag{6.1}$$

In order to estimate the flux density we combine the `BPZ` (Bayesian photo-z) public code[1] described in Benítez (2000), which includes various Spectral Energy Distributions (SEDs) and filter transmission curves of different instruments, with the `Barak` package[2] used to handle both SEDs and passband filters, to extract spectral flux density given the passband together with the galaxy SED, redshift and magnitude. All the available SEDs are shown in Fig. B.1. The SED used in this work is a starburst template from Kinney et al. (1996) (plotted in bottom central panel of Fig. B.1 with a red line). Thus, given the source magnitude in the $F814$ band and its redshift, the SED is redshifted and normalised by matching the flux equivalent to the given magnitude in the given passband; finally, the magnitude is converted into a flux density. We find

[1] https://www.stsci.edu/~dcoe/BPZ/
[2] https://nhmc.github.io/Barak/

that with this process we could generated realistic GGSL examples, however, we are planning to inject galaxies described with different SEDs and to include extinction correction.

In Tab. 6.2 we list the parameter set which allows to generate the source: the axis ratio and the position angle values are randomly extracted from a uniform distribution constrained between $[0.2, 1.0]$ and $[0, \pi]$, respectively. The Sérsic index is also extracted from a uniform distribution between $[1.0, 2.0]$, imposed to approximate late-type galaxy profiles. In order to reproduce the observed universe, we avoid a uniform sampling of the other required parameters. Rather, concerning the source magnitude and redshift, we exploit the COSMOS 2015 catalogue (Scoville et al. 2007; Laigle et al. 2016), completed with HST fields observations, to estimate the redshift probability density function (PDF) within a magnitude range, i.e. $p(z|\Delta m)$, from whose Cumulative Distribution Functions (CDFs) we extract a magnitude and a redshift that will be assigned to the injected galaxy. The COSMOS catalogue contains PSF-matched photometry and photometric redshifts for $\sim 8 \times 10^5$ objects covering an area of $\sim 2\text{deg}^2$. As shown in Fig. 6.4, we estimate the number counts (i.e. the number of galaxies per square degree per magnitude bin) and compare with the number counts taken from Capak et al. (2007), obtained by combining observation in the $F814$ band of Hubble Deep Field North and South (Williams et al. 1996; Metcalfe et al. 2001). We interpolate these counts down to $i_+ = 29$, and, in order to complete the number counts beyond the COSMOS depth limit, we compute the difference between COSMOS and Capak et al. (2007) number counts. We selecte 6 magnitude ranges (with i_+ limits $= \{22, 24, 25, 26, 27, 28, 29\}$ mag), in each of which we have fitted a redshift PDF, $p(z|\Delta i_+)$, by approximating it with a simple function of the form (see, e.g., Lombardi & Bertin 1999; Lombardi et al. 2005):

$$p(z|\Delta i_+) = \begin{cases} Az^2 e^{-z/z_0} & \text{for } i_+ \in [22, 24) \quad (6.2) \\ Az^2 e^{(-z/z_0)^{1/2}} & \text{otherwise} \quad (6.3) \end{cases}$$

The PDFs fitted in each magnitude bin are shown in the right panels of Fig. 6.4. In order to extract a relation between the fraction of sources and their magnitude, we estimate the CDF for number counts, $P(i_+)$, and redshift, $P(z|\Delta i_+)$. So, given the global number of galaxies to inject, we use the CDF $P(i_+)$ to extract a magnitude, $\hat{i}_+$, assigned to each galaxy, and, for each considered bin, we exploit the cumulative distribution, $P(z\,|\,\hat{i}_+ \in \Delta i_+)$, to extrapolate a redshift.

Finally, regarding the source effective radius, we exploit the relation describing the redshift evolution of galaxy physical sizes studied by Shibuya et al. (2015) approximated with a function of the form: $r_e = B(1+z)^\beta$ (fitted by combining galaxy radii estimated in the UV and optical bands). However, after comparing this relation with the effective radii measured by Tortorelli et al. (2018), we found a significant discrepancy for low-z galaxies. Therefore, to constrain galaxy sizes at lower redshift, we adopt this relation only for $z > 1$, while for $z \leq 1$ we opt for a constant function whose value is set to the effective radius evaluated in $z = 1$:

$$R_e(z)\ [\text{kpc}] = \begin{cases} 2^\beta B = 2.54 & \text{for } z \leq 1 \\ B(1+z)^\beta & \text{for } z > 1 \end{cases} \quad (6.4)$$

Given such relation, by assuming a deviation $\sigma = 0.25$kpc (obtained by Shibuya et al. 2015 analysis) for the whole redshift range, we extract the r_e corresponding to a given z within $(-1.5\sigma, \sigma)$ (bottom right panel in Fig. 6.5), we opted for these thresholds by analysing the resulting $R_e(z)$ limits: since we are also interested in finding background galaxies with size $R_e \sim 0.5$ kpc (close to the PSF size, for $z > 4$, as shown in the bottom right panel of Fig. 6.5), we set R_e in the range $(-1.5\sigma, \sigma)$.

The simulation process is carried out through the `PyLensLib` (Meneghetti 2021) and can be summarised as follow:

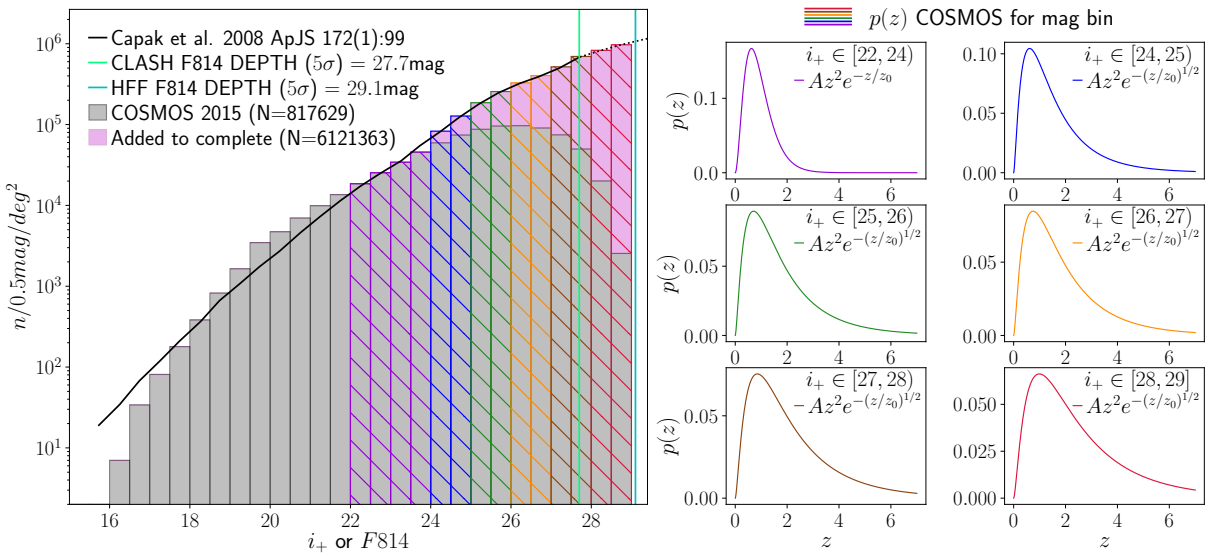


Figure 6.4: *Left panel*: number counts estimated from COSMOS catalogue (grey bars), compared with Capak et al. (2007) (black line), together with the 5σ HFF and CLASH $F814$ depth limit (cyan and mint-green vertical lines) with the counts added to complete the distribution (light magenta area), the 6 magnitude ranges within the redshift PDFs have been estimated (coloured hatched bars), showed in the second and third column panels.

1. Inputs: the deflection angle map, the lens redshift (i.e. the cluster central redshift, $z_{cluster}$) and the number of objects to simulate;

2. the m_{F814} magnitudes are extracted from the magnitude CDF, $P(i_+)$, obtained using the COSMOS catalogue completed with HST fields observations;

3. for each magnitude bin, the source redshift (z_s) is derived from the corresponding CDF, $P(z|\Delta i_+)$, approximated with Eq. 6.2 or 6.3, imposing a minimum value for z_s equals to $z_{\text{cluster}} + 0.4$[3];

4. given the source redshift, the effective radius is extracted within $(-1.5\sigma, \sigma)$ from the $r_e(z_s)$, computed according to Eq. 6.4;

5. the axis ratio, the position angle and the Sérsic index are extracted from uniform distributions constrained between $[0.2, 1.0]$, $[0, \pi]$ and $[1.0, 2.0]$, respectively;

6. the starburst SED is redshifted at z_s and normalised using the selected m_{F814} magnitude; this SED is used to extract the flux density corresponding to the involved bands, which in turn are converted into count rates;

7. given the source and lens redshift, together with the deflection angle maps, the critical lines are computed from convergence and shear (Eqs. 2.34, 2.35, 2.36) by imposing that magnification goes to infinity (Eq. 2.43);

8. the critical lines with $\theta_E \leq 0.2''$ or $\theta_E > 5''$ are excluded (to avoid the inclusion of the main, largest, critical line, and the very small-scaled galaxies); then a probability is assigned to each critical line as a function of its θ_E (i.e. larger critics have more chance to be selected);

[3]Following the same approach carried out by Meneghetti et al. (2020), we measured the lensing cross-section for the involved galaxy clusters, finding that this goes to zero when z_s gets close to $z_{cluster}$: it is negligible for $z_s \lesssim z_{cluster} + 0.1$, while at $z_s = z_{cluster} + 0.4$ it has more than doubled its first non-null value.

9. once the critical line has been identified (which is associated to the lens), the corresponding caustic is computed with the second order approximation lens equation (Eq. 2.45) and the galaxy position in source plane ($\vec{\beta}$) is extracted within a buffer surrounding the caustic (whose width is $0.5 r_e$);

10. the surface brightness is computed by assuming a Sérsic profile (Eqs. 2.1, 2.6- 2.8), in each $\vec{\beta}_{ij}$ and assigned to the corresponding element $\vec{\theta}_{ij}$ (mapped with Eq. 2.45);

11. the resulting image is convolved with a PSF and summed to the HST $F814$ image;

12. finally, the image in the other bands is simulated by rescaling the count rates:

$$X_F = X_{F814} \cdot (N_F / N_{F814}) \quad \text{for } F \in \{F606, F435\} \tag{6.5}$$

    where F labels the other involved ACS bands ($F606$ and $F435$), N_F is the count rate in a certain bandpass, X_{F814} is the non-convoluted image in the $F814$ filter and X_F is the output image in the F band, which is finally convolved with the corresponding PSF;

13. a squared cutout centred on critical line centroid is extracted for each simulated galaxy.

Fig. 6.5 layouts the features for the $\sim$ 3000 injected galaxies: the top panels show the fraction of number counts (split into the 6 involved bins) and the whole normalised redshift distribution, while the normalised redshift distribution for each magnitude bin is shown in the second and third row panels; the bottom right panel displays the r_e evolution with z_s, overlapped to the Eq. 6.4 (together with the z_s and r_e distribution); finally, to validate our simulation, we plot the relation $\theta_E(F160)$ in the bottom right panel (together with $F160$ and θ_E distribution): Einstein radii are computed by measuring the area of the selected critics and the corresponding cluster member $F160$ magnitude is associated with these radii. It can be strongly compared with the scaling relation for the cluster member velocity dispersion, $\sigma^{\rm CLM}$ (see Eq. 2.49), once the Einstein radius has been written as a function of the cluster member $F160$ magnitude, $m_i^{\rm CLM} = m_{F160}$, by assuming a SIS profile (see Eq. 2.9 described in Sec. 2.1.4):

$$\theta_{E,i} = 4\pi \left(\frac{\sigma_v^{\rm ref}}{c} \right)^2 \left(\frac{D_{LS}}{D_S} \right) 10^{0.8\alpha (m_{F160}^{\rm ref} - m_i^{\rm CLM})} \tag{6.6}$$

where $m_{F160}^{\rm ref}$ is the $F160$ reference magnitude, chosen to be that of the BCG, $\sigma_v^{\rm ref}$ is a free parameter of the lens model; α is the slope in the scaling relation (Eq. 2.49), its value has been fitted when velocity dispersion measurements were available (i.e. for the cluster models provided by Bergamini et al. 2019 and Bergamini et al. 2021b), taken from the literature in the other cases. The modest discrepancy between the scaling relations (plotted in the bottom left panel of Fig. 6.5) is due to the multiple sampling of the same cluster member, which results to be associated with several critical lines, obtained by varying the redshift of the injected galaxy.

In order to complete the Knowledge Base, it is necessary to select an adequate number of non-GGSLs (i.e. the negative class for the classification problem). To perform this selection, by exploiting the spectroscopic information obtained by combining CLASH-VLT VIMOS programme with MUSE archival observations (as done for the cluster member identification, see Section 5.2), we extract squared $10''$ across cutouts centred on the cluster member positions (with rest-frame velocity separation $|v| \leq 5000\,{\rm km\,s^{-1}}$) belonging to 16 clusters, listed in Tab. 6.3. Since some of these cutouts could be contaminated by strong lensing events, this member ensemble has been submitted for a visual inspection to our science team, through the Google Forms Service[4]

[4] https://developers.google.com/apps-script/reference/forms

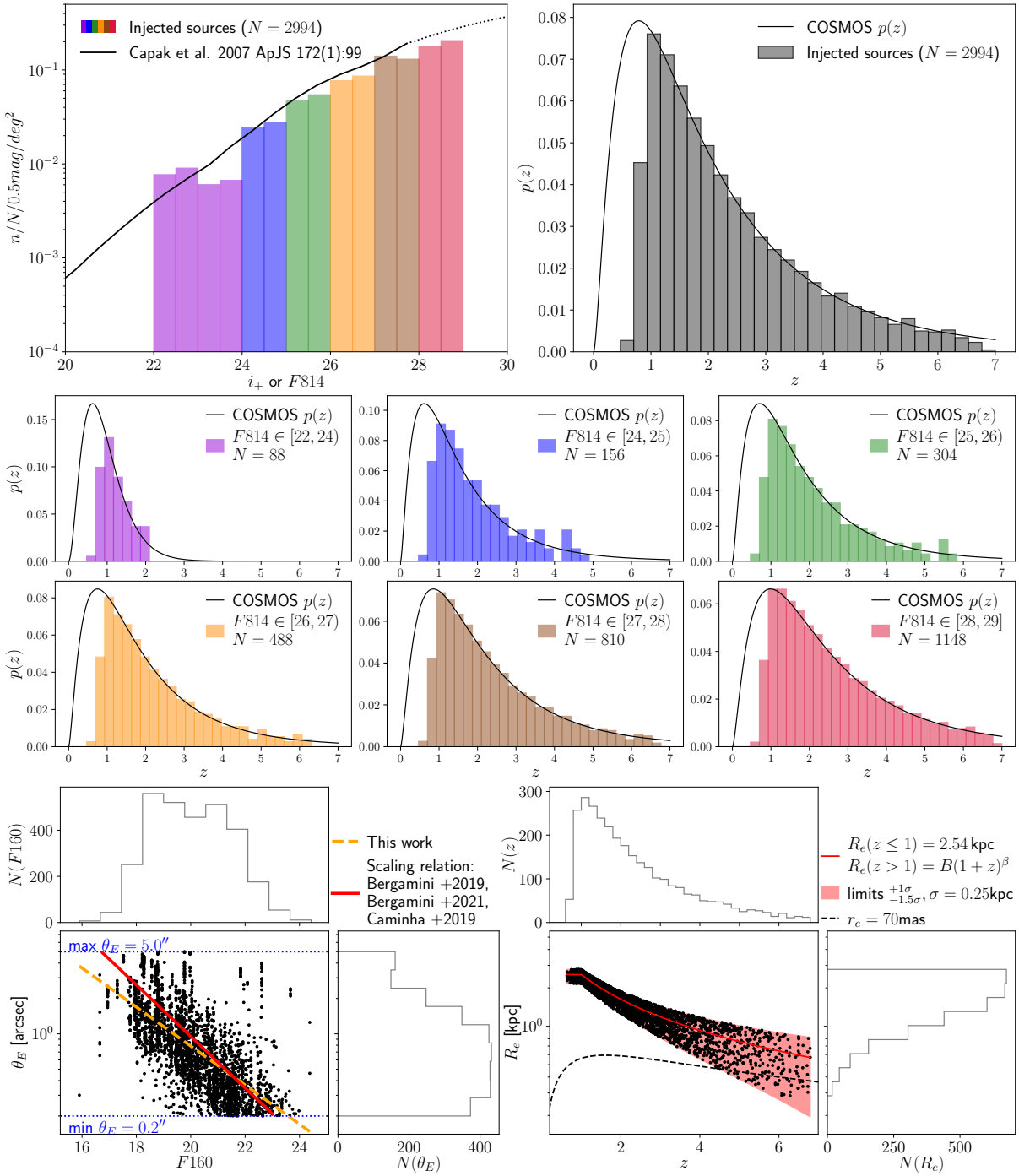


Figure 6.5: GGSL set layout. The number counts as a function of magnitude, compared with Capak et al. (2007), is shown in the upper left panel. The extension (dashed line) at faint magnitudes reproduces HST deep fields' counts. The normalised redshift distribution for the whole sample is plotted in the upper right panel (together with the PDF fitted from COSMOS photometric redshift), while panels in the second and third rows show the normalised redshift distribution for each involved magnitude interval (together with the PDFs fitted from COSMOS photometric redshift). The missing of low-z galaxies in these plots is due to the adopted selection on the source redshift, i.e. $z_s > z_{\rm cluster} + 0.4$. The resulting scaling relation, i.e. θ_E vs. $F160$ (in orange), compared with Bergamini et al. (2019, 2021b) and Caminha et al. (2019) relation (in red), is shown in the bottom left panel, while the bottom right panel represents the adopted relation for the redshift evolution of R_e (constant for $z \leq 1$, taken from Shibuya et al. 2015 for $z > 1$), together with the upper and lower limits within which R_e is extracted (light red area) and the $0.070''$ threshold, under the which the source size is indistinguishable from the PSF, after the convolution.

Figure 6.6: Examples of RGB cutouts of GGSLs and non-GGSL, by combining $F435$, $F606$, $F814$ bands. GGSL cutouts are sorted in order of increasing θ_E (on columns) and $F814$ (on rows). To emphasise fainter sources, images have been stretched by clipping values within $\pm 3\sigma$ and then normalised. Cutouts are $\sim 9''$ across; red and green squares enclose an area with a side of $\sim 4''$ and $\sim 8''$, respectively. z_s, θ_E value and $F814$ magnitude are shown at the bottom of each GGSL cutout.

Cluster		$z_{cluster}$	N	Cluster		$z_{cluster}$	N
Abell 383	A383	0.188	70	Abell 370	A370	0.375	172
Abell 209	A209	0.209	75	MACS J0416-2403	M0416	0.397	120
RX J2129+0005	R2129	0.234	51	MACS J1206-0847	M1206	0.439	147
Abell 2744	A2744	0.308	126	MACS J0329-0211	M0329	0.450	66
MS 2137-2353	MS2137	0.316	52	RX J1347-1145	R1347	0.451	44
RX J2248-4431[(b)]	R2248	0.346	178	MACS J1311-0310	M1311	0.494	53
MACS J1931-2635	M1931	0.352	28	MACS J1149+2223	M1149	0.542	130
MACS 1115+0129	M1115	0.352	96	MACS J2129-0741	M2129	0.587	45

Table 6.3: Description of the cluster sample involved in the non-GGSL selection. The name of the clusters and their redshift are reported in the first 3 columns. The fourth column shows the number of non-GGSLs identified through visual inspection.
[(a)] The cluster RX J2248.7−4431 is also known as Abell S1063.

developed using the `Google Apps Script` platform[5]. In this way, 7 expert astronomers were called to identify strong lensing events within the presented cutouts. To help the identification, image was separately shown in three bands ($F435$, $F606$, $F814$) and by combining these filters into an RGB cutout; moreover the member $F814$ magnitude was provided and any spectroscopic source (with $z_s \geq z_{\rm cluster} + 0.1$) within the cutout has been circled. The answers have been then collected, by assigning to each galaxy a score of +1, +0.5 or −1, when an astronomer identifies a reliable GGSL, a less likely GGSL or a non-GGSL, respectively. In this way, we obtain a dataset of non-GGSLs as pure as possible, by removing from the KB even suspicious GGSLs. Furthermore, we also exclude cutouts which are contaminated (e.g. by a star or in the nearby of a brighter galaxies) or corrupted (due to cuts of the FoV or overlapping between different bands). All these scores are finally averaged so that at end of this process, we visually inspected ∼ 2000 galaxies: ∼ 1500 selected as non-GGSL, while ∼ 280 cutouts were excluded and ∼ 320 were judged as candidate GGSLs (including both reliable and doubtful GGSLs).

As a result, the Knowledge Base is composed by ∼ 3000 GGSLs and ∼ 1500 non-GGSLs. This difference between the amount of images in each class makes the problem unbalanced. Actually, this unbalancing is intentional (indeed we simulated two times the number of non-GGSLs), since we wanted to ensure a diversity among the simulated events, in order to cover as wide a parameter space as possible. Thus, the pre-processing phase has been performed differently for the two classes, in order to compensate their different sample sizes:

- the valid set is obtained by extracting the 10% of GGSL and the 20% of non-GGSL, resulting into ∼ 300 GGSLs and non-GGSLs;

- k-fold splitting (as always with $k = 10$) applied for each of these two sets (after the valid sample exclusion), i.e. the training sets are composed by ∼ 2500 GGSLs and ∼ 1100 non-GGSLs;

- the training folds are independently augmented involving the 10% of GGSLs and the 50% of non-GGSLs, the resulting augmentation factors are, 1.5 and 3.5, respectively. Thus, the both classes are populated with ∼ 3800 training examples.

We build two datasets by extracting cutouts with side of 128 pixels (∼ 4″) and 256 pixels (∼ 8″). Hereafter, we refer to these two datasets as *EXP-A* and *EXP-B*, respectively. A sample of simulated GGSLs and non-GGSLs is shown in Fig. 6.6, where limits of both ∼ 4″ and ∼ 8″

[5] https://developers.google.com/apps-script/overview

cutouts are marked, respectively, as red and green squares; moreover GGSL cutouts are sorted in order of increasing the galaxy-lens θ_E (on columns) and the source intrinsic magnitude $F814$ (on rows). Besides the typical arc and ring like features, several GGSL images do not reveal any strong lensing event (i.e. the appear as non-GGSL), this mainly happens when: (*i*) the injected sources are faint, so the lens galaxy masks the GGSL signal (38% of sources have $F814 > 28$ mag), (*ii*) for small-scale lenses with limited θ_E, so the source is hidden by lens galaxy halo (35% of lenses have $\theta_E < 0.5''$), (*iii*) a combination of the previous two cases (13% of examples have both $F814 > 28$ mag and $\theta_E < 0.5''$). Although these cutouts represent the most challenging cases for the classifier, they act as adversarial examples (see Sec. 3.6.3) preventing network overfitting and allowing the network to gain a high degree of generalisation, indeed, we also performed some experiments by removing faint sources and small-scale lenses, even though networks achieved nearly perfect results, they were not able to identify strong lensing events when tested on real data, probing an insufficient generalisation capabilities, i.e. the networks overfitted. Finally, it should be noted that, since critical lines are formed around galaxy members, the same cutout could represent both a GGSL (with the addition of an injected source) or a non-GGSL (without the injection).

6.3 Experiments

This section is dedicated to the experiment description and the result analysis. As already stated in Sec. 3.7, we test classification capabilities of different architectures, whose comparison is shown in Tab. B.1; from which is clear that, among the involved networks, two of them (the VGG and SC-VGG) show a stable behaviour for both GGSL and non-GGSL (abbreviated as NGGSL in tables and figures, to avoid the weighted of the text), for both the cutout configurations (with size $\sim 4''$ and $\sim 8''$). Moreover, networks behaved differently when tested on real images, clearly suggesting that the VGG and SC-VGG gained a larger degree of generalisation than the other architectures. As in the case of cluster member identification (see Chapter 5), we restrict the following analysis to the results achieved by the application of these two networks (see Section 3.7.1).

6.3.1 CNN performance comparison

A summary of performance is shown in Tab. B.1 and graphically in Fig. 6.7. Unless otherwise specified, performances are computed by assuming a probability threshold of 0.50. In general terms, CNNs correctly classified at least 80% of sources, with accuracy peaks of $\sim$ 90%. By exploring these results, it is evident the performance dropping between *EXP-A* and *EXP-B* (which is clear even for the other implemented architectures, Tab. B.1). All models seem to be affected by the thumbnail sizes: when trained with $8''$ cutouts, networks show an average efficiency around 80%, whereas they achieve accuracy > 87% with smaller cutouts, with differences ranges between 8% to 10%; suggesting that CNNs lose their identification capabilities with increasing of the physical size of the image. This suggests that larger images could include neighbours which contaminate the cutouts and affect network performances, whereas wide-separation arcs or strongly asymmetric lensing features could be (partially) lost with smaller cutouts. Such dependence on the Einstein Radius (θ_E) and other parameters will be discussed in further details below.

Respect to the GGSL identification, all CNNs appear more pure than complete (even by including other architectures, Tab. B.1), with purity-completeness differences ranging between 1.4% up to 11.3%, particularly enhanced for the *EXP-B* configurations. Concerning the negative class, i.e. the non-GGSLs, networks reveal the opposite behaviour with purity-completeness

		EXP-A (∼ 4″ cutouts)				*EXP-B* (∼ 8″ cutouts)			
Class	*%*	VGG	VGG*	SC-VGG	SC-VGG*	VGG	VGG*	SC-VGG	SC-VGG*
	AE	87.7	89.7	89.4	89.7	79.4	81.4	79.2	81.7
GGSL	*pur*	93.4	90.4	93.1	91.5	89.6	89.2	90.2	88.0
	compl	88.6	92.5	91.7	93.4	78.4	86.9	78.9	87.2
	F1	91.0	91.5	92.4	92.5	84.0	88.1	84.5	87.6
NGGSL	*pur*	76.7	88.5	81.4	89.6	72.7	74.3	73.1	77.5
	compl	85.4	85.4	84.1	84.1	81.7	81.7	79.3	79.3
	F1	81.1	86.9	82.8	86.8	77.2	78.0	75.7	78.4

Table 6.4: Performance comparison between the VGG architectures, related to the two cutout configurations (*EXP-A* and *EXP-B*, described in Section 6.2) and expressed in terms of the statistical estimators described in Section 3.6.6. Performances are also evaluated by removing faint sources and small-scaled lenses (i.e. $F814 > 28$ mag and $\theta_E < 0.25''$), marked with an asterisk.

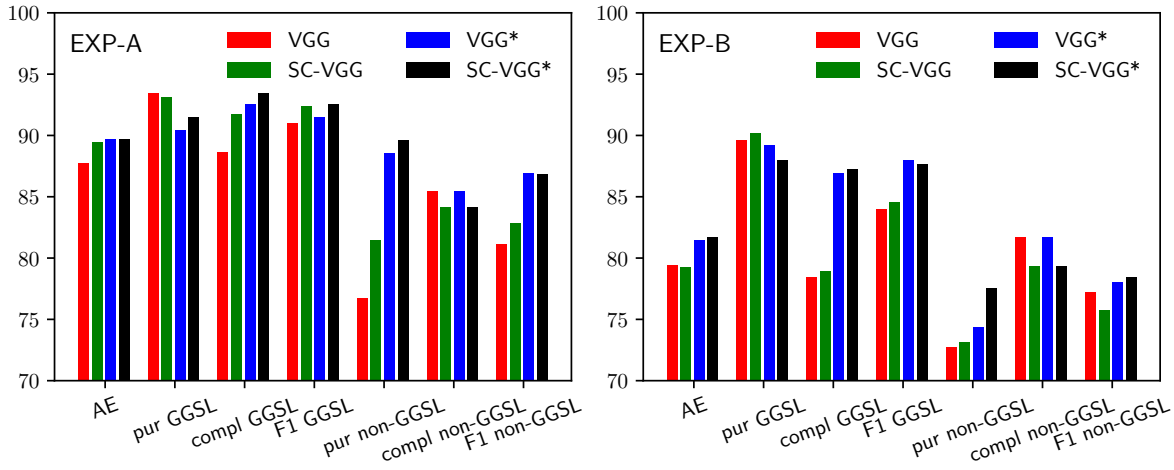


Figure 6.7: Comparison between VGGs performance on *EXP-A* and *EXP-B* dataset configuration, in terms of statistical estimators described in Section 3.6.6.Performances are also evaluated by removing faint sources and small-scaled lenses (i.e. $F814 > 28$ mag and $\theta_E < 0.25''$), marked with an asterisk in the legends.

differences fluctuating in $(-9.0\%, -2.7\%)$. Such typical dichotomy turns out when, at least for a fraction of examples, there is not clear distinction between classes, even when they are inspected by experts. In this work, such boundary class encloses fainter sources and small-scaled lenses, whose intersection represents 7% of the GGSL set. As already stated, these cutouts prevent model overfitting by acting as adversarial examples (Sec. 3.6.3), but, in order to quantify their effect on model performances, we also measure the statistical estimators by removing them from the test set, these re-estimated performances are reported in Tab. 6.4 and Fig. 6.4, marked with an asterisk. Clearly, non-GGSL completeness is not affected by this modification, while non-GGSL purity increases up to $\sim$ 12% for the *EXP-A* configurations, $\sim$ 4% for the *EXP-B* configurations, with an F1 score improving of 4% and 2%, respectively for *EXP-A* and *EXP-B* configurations. Regarding the GGSL identification we find a more balanced trade-off between purity and completeness: respect to a purity maximum decreasing of $\sim$ 3%, we measure a maximum completeness increasing of 10%, with an average F1 improvement ranging between 0.1%, for the more stable model (i.e. the SC-VGG trained with *EXP-A* configuration), and 4.1%, related to the volatile network (i.e. the VGG trained with *EXP-B* configuration). The unequal impact of this test-example exclusion, respect to the cutout configurations, confirms the larger performance fluctuation of the *EXP-B* experiments, regardless for the adopted model.

Due to this difference, we focus the analysis on the results achieved by the VGG and SC-VGG

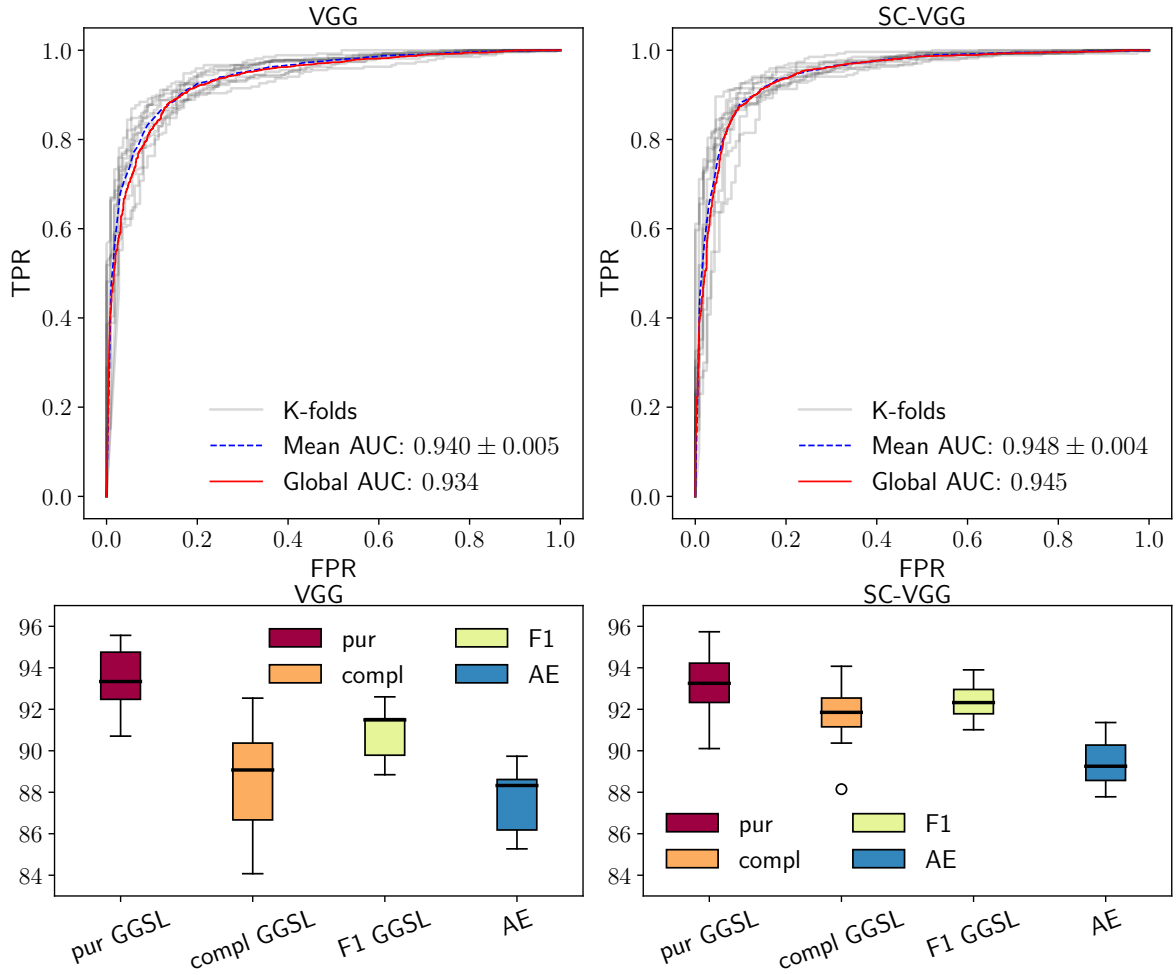


Figure 6.8: Comparison between VGG and SC-VGG models, by evaluating performance fluctuations over the $k = 10$ folds. *Top panels*: ROC curves, where grey lines represent the results achieved on each fold, blue line is the mean between these latter, red lines refer to the global stacked performance (the AUC values are shown in legend). *Bottom panels*: box plots, where the boxes delimit the 25th and 75th percentile, i.e. first and third quartile, Q_1, Q_3, with $IQR = Q_3 - Q_1$ is the so-called interquartile range, and error bars standardly enclose $(Q_1 - 1.5 \cdot IQR, Q_3 + 1.5 \cdot IQR)$ (corresponding to the 90.3% of data, i.e. within $\pm 2.698\sigma$ values). Black horizontal lines are the medians.

with the *EXP-A* configuration.

Fig. 6.8 and Tab. B.2 show a comparison between VGG and SC-VGG models in terms of performance fluctuations. Indeed, due to the adpoted k-fold approach, performances were also evaluated in each fold. From ROC curves (top panels in Fig. 6.8), models appear comparable, even if SC-VGG achieves results slightly better (with an AUC increasing $\sim$ 1%). However, their different classification capabilities are evident by exploring the fluctuation of the statistical estimators (bottom panel in Fig. 6.8 and Tab. B.2). Both of them have similar GGSL purity, in terms of median ($\sim$ 93.3%), first and third quartile ($Q_1 \sim 92.4\%$, $Q_3 \sim 94.5\%$) and inter-quartile range $IQR = Q_3 - Q_1 \sim 2.1\%$, with differences spanning in $(-0.6\%, 0.4\%)$. Larger differences occur for the other GGSL metrics: SC-VGG performances show an overall significant improvement in terms of completeness (median: +2.8%, first and third quartile: +4.5%, +2.1%, "minimum" and "maximum" variation: +6.3%, +1.6%), which in turn is reflected into a F1 score gain, ranging in (+0.8%, +3.8%). Concerning the non-GGSL metrics (only listed in Tab. B.2), SC-VGG achieves lager purity values, while VGG shows better completeness, even if the average purity increasing of SC-VGG is $\sim$ 4.7%, while the average gain in term of completeness obtained

by the VGG is $\sim$ 1.0%. Looking at non-GGSL F1 scores, SC-VGG achieves the best results, varying in (+0.8%, +2.2%), with an average improvement of 1.6%. In term of inter-quartile range, even if both models seem to reproduce the same trend, characterised by GGSL metrics more constrained than the ones related to non-GGSL, SC-VGG shows more reliable IQRs for almost all the estimators, with relative increasing ranging in (7%, 62%) and an average of 31%. Thus, based on this analysis, SC-VGG is the architecture characterised by the best trade-off between purity and completeness for both GGSL and non-GGSL (92.4%, 82.8% vs. 91.0%, 81.1%), most stable results evaluated on $k = 10$ folds ($\langle IQR \rangle_{SC-VGG} = 2.1\%$ vs $\langle IQR \rangle_{VGG} = 3.4\%$), particularly accentuated for the GGSL completeness, with respect of which it shows a relative improvement of 62%.

We also perform an experiment using a single band. We restricted this analysis only to the *EXP-A* cutout configuration. The results are outlined in Tab. B.4, where we summarise the comparison between the network trained and tested using one band ($F435$, $F606$ and $F814$ independently) with the predictions carried out by the VGG and SC-VGG models. Although performances reproduce the VGG and SC-VGG trends (i.e. a high GGSL purity, around 91%, and a low NGGSL purity, about 76%), the use of a single band does not provide any improvement: we found a reduced accuracy (between 0.5% and 3.3%), a GGSL F1-score (0.2% − 6%) and a NGGSL F1-score (1.8% − 5.2%). This leads to an average performance reduction of 1.8%, 1.3% and 3.5%, respectively for the average efficiency, GGSL and NGGSL F1-scores. Despite this reduction, performances suggest that GGSL can be also classified using just a single band, as already shown by Petrillo et al. (2017, 2019a); Li et al. (2021b), which exploited the optical images from the Kilo-Degree Survey (de Jong et al. 2015).

Finally, since we have selected the non-GGSLs through a visual inspection (see Sec. 6.2, we compare the predictions made by neural networks with the outcomes of this visual selection. As pointed out, since our aim is to produce a non-GGSL catalogue as pure as possible, the resulting set of GGSL candidates is strongly contaminated. Thus, these sets of candidates are more suited to perform a comparison on the identification of non-GGSLs than of GGSLs. However, by considering the $\sim$ 1800 visually inspected sources, we measure an high fraction of non-GGSL predicted in common ($\sim$ 95%), while the percentage of common GGSLs is just $\sim$ 35%. By taking as candidate GGSLs only those sources with a CNN probability ≥ 0.75, we found that all the 105 candidates are classified as GGSLs by both neural networks and astronomers.

6.3.2 False Negative and False Positive analysis

In this section we specifically analyse CNN predictions, by investigating the False Positive and False Negative distributions, trying to understand the causes of such misclassifications, by inspecting correlations with: (*i*) galaxy-lens magnitude and colour[6] (respect to the False Positives), (*ii*) redshift, intrinsic magnitude, source effective radius, size of the galaxy-lens, ratio between the source and distance between the source and the caustic centroid normalised to the caustic size (respect to the False Negatives). In this analysis we use the predictions of both VGG and SC-VGG.

For easy reading, we specify the definition of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN), already given in Sec 3.6.6, within the context of this work. We assume the GGSL as the positive class, thus, the TPs are strong-lenses correctly classified, FPs are non-GGSLs classified as GGSLs, FNs are GGSLs classified as non-GGSL, and, finally, TNs are non-GGSLs correctly classified.

[6]Similarly to what as done for the cluster member identification work (Chap. 5), we used the normalised colour to bring together galaxies at different redshift (see Eq. 5.2).

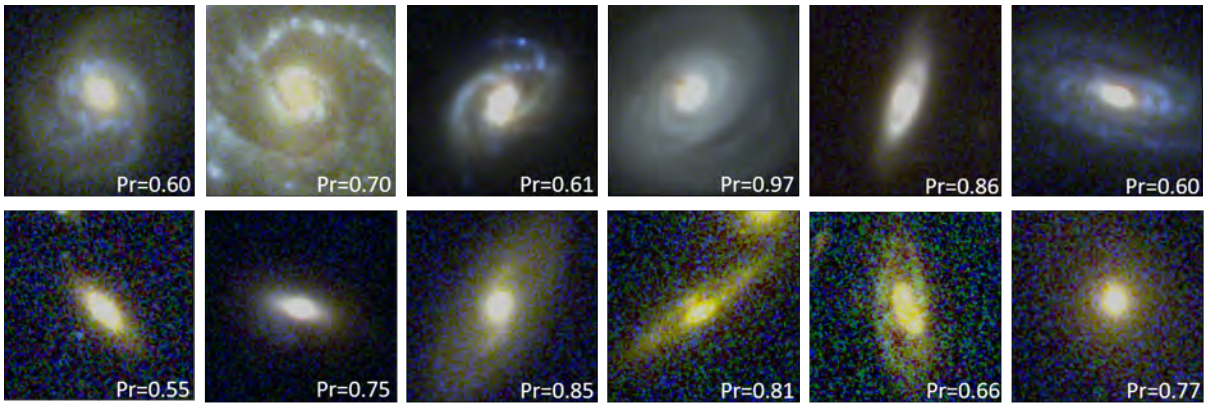


Figure 6.9: Selection of False Positives common to both VGG and SC-VGG models (they represent the 8% and 7% of the FP ensemble, respectively, referred to VGG and SC-VGG). The probability of belonging to the GGSL class is shown in each thumbnail (referred to the SC-VGG model). Cutouts are $\sim 4''$ across.

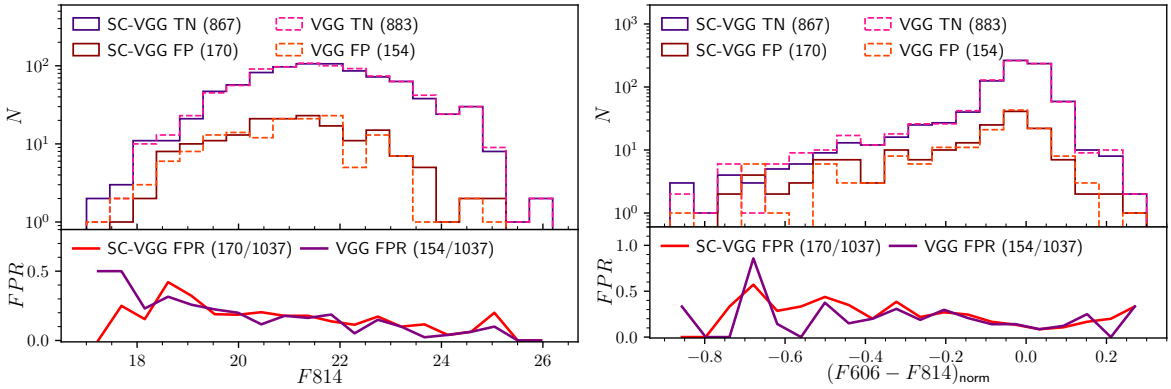


Figure 6.10: True Negative (TN) and False Positive (FP) analysis related to the VGG and SC-VGG performances, as a function of the galaxy lens features: $F814$ magnitude (left panel), $(F606-F814)_{\rm norm}$ normalised colour (right panel). In both panels, TN rates are plotted with purplish lines, while FP rates with reddish lines, in both cases, solid for VGG, dotted for SC-VGG. The FP Ratio ($FPR = FP/(TN+FP)$) for each involved features is plotted in the bottom of each panel (as purple line for VGG, red for SC-VGG).

Concerning the non-GGSL mistakenly classified as strong-lenses, a selection of False Positives common to both VGG and SC-VGG models is displayed in Fig. 6.9, while Fig. 6.10 shows the TN and FP rate, together with the False Positive Ratio ($FPR = FP/(TN+FP)$) as a function of the cluster member photometry: $F814$ magnitude (left panel) and the normalised colour (right panel), also summarised in Tab. B.3. The number of False Positives follows both the non-GGSL magnitudes and colours distribution for $F814 > 19$ and $(F606-F814)_{\rm norm} > -0.5$, characterised by an approximately constant FPR with a median of ~ 0.20 and ~ 0.16, respectively for the $F814$ and colour dependence. There are two FP excesses, respectively, in the brighter and bluer part of the parameter space. The FP increasing up to $\sim 7\%$ (with FPR spikes around 0.9 and 0.5 for $(F606-F814)_{\rm norm} \sim -0.7$, related to VGG and SC-VGG results) for $(F606-F814)_{\rm norm} < -0.5$ is reasonable: (*i*) disc galaxies, with a red bulge surrounding by blueish spiral-like structures, are confused as strong-lenses (see Fig. 6.9); (*ii*) blue galaxies are under-represented in the KB, since we have sampled the core of clusters (where cluster lens models have been fitted), populated by red members. VGG and SC-VGG dependence on colours appears strictly comparable within 1%. Models have similar trends also with regard to the $F814$ magnitude: an about constant FPR ratio of ~ 0.16 for $F814 > 19.5$, with a FP excess in the brightest region. This can be due to

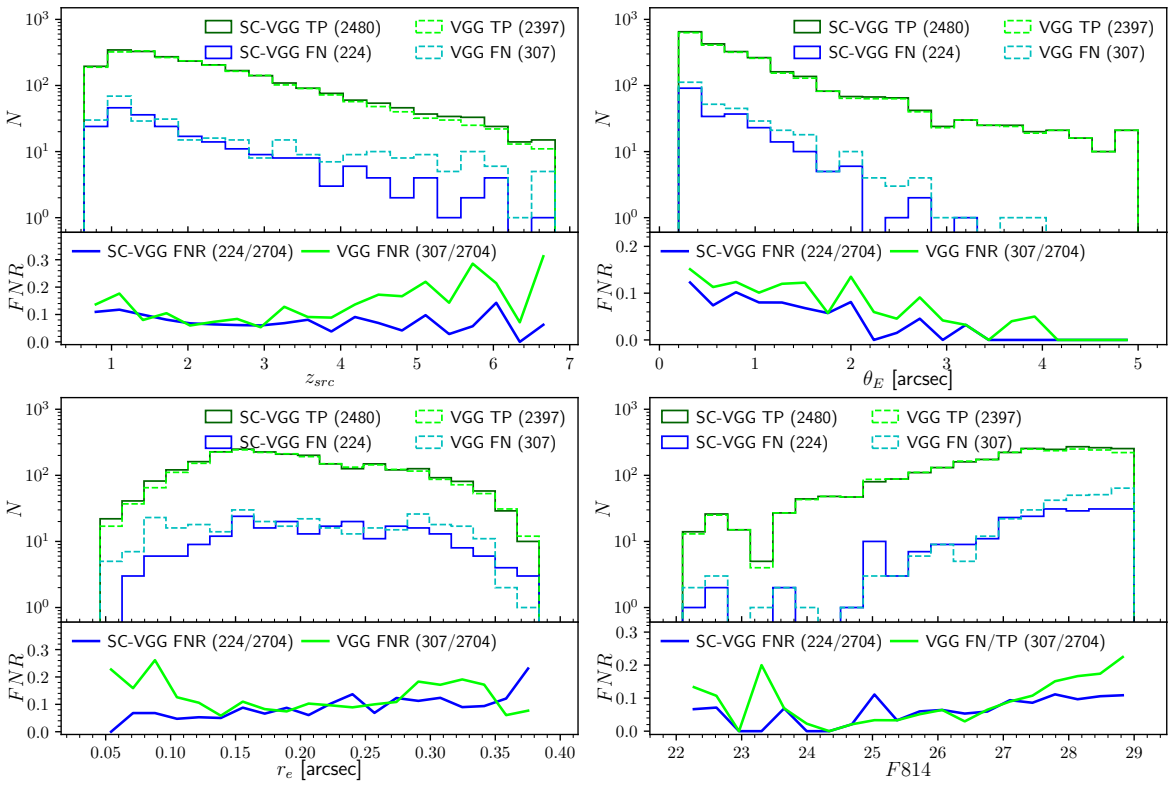


Figure 6.11: True Positive (TP) and False Negative (FP) analysis, related to the VGG and SC-VGG performances, as a function of: source redshift (z_{src}, upper left panel), lens dimension (θ_E, upper right panel), source effective radius (r_e, central left panel), source intrinsic $F814$ magnitude (central right panel), source. In all panels, TP rates are plotted with green lines, while FP rates with blueish lines, in both cases, solid for SC-VGG, dotted for VGG. The FN Ratio ($FNR = FN/(TP + FN)$) for each involved features is plotted in the bottom of each panel (as green line for VGG, blue for SC-VGG).

the embedded lensed features in the training set. Indeed, as already discussed, several lens-like structures are hidden by galaxy halo, so, in presence of bright galaxies, the networks suggest the existence of a lens event blended by the galaxy. FPRs are also shown in Fig. B.2 binned on colour-magnitude diagrams, which show a decreasing FPR gradient in the faint-red direction. FPRs related to VGG and SC-VGG are comparable for large parts of the parameter space: 84% of this space are characterised by differences between ratios ranging $\in [-0.25\%, 0.25\%]$.

Regarding the strong-lenses misclassified as non-GGSLs, Fig. 6.11 shows the True Positive and False Positive rate, together with the False Negative Ratio ($FNR = (FN/(TP + FN))$ as a function of: source redshift (upper left panel), galaxy-lens size (measured in term of θ_E, upper right panel), source effective radius (r_e, bottom left panel), source intrinsic $F814$ magnitude (bottom right panel), also summarised in Tab. B.3. The number of False Negatives decreases with θ_E, characterised by FNR ratio which becomes $\lesssim 0.06$ for $\theta_E \geq 3''$ and zero for $\theta_E \geq 4''$, therefore, interestingly, also strongly asymmetric lensing features with large θ_E are correctly classified, even if they are partially lost outside the cutout. On the other hand, FNs are associated with small scaled galaxy-lens: a fraction of FN larger than 10% gathers at $\theta_E < 0.5''$. At the same time, misclassifications increase with the source $F814$ magnitude showing a FN fraction ~ 0.10 for $F814 \geq 27$ mag. The VGG and SC-VGG FN ratios are similar (0.05) down to $F814 = 27$, but they diverge in the faint end: SC-VGG shows a constant FNR of ~ 0.10, while the VGG FNR continues to increase up to 0.20 for $F814 \geq 28$ mag. The combined dependence on galaxy-lens Einstein radius and source $F814$ magnitude is shown in the top panels of Fig. B.3, where a FNR increasing gradient clearly points toward the small θ_E and faint magnitude region for both VGG

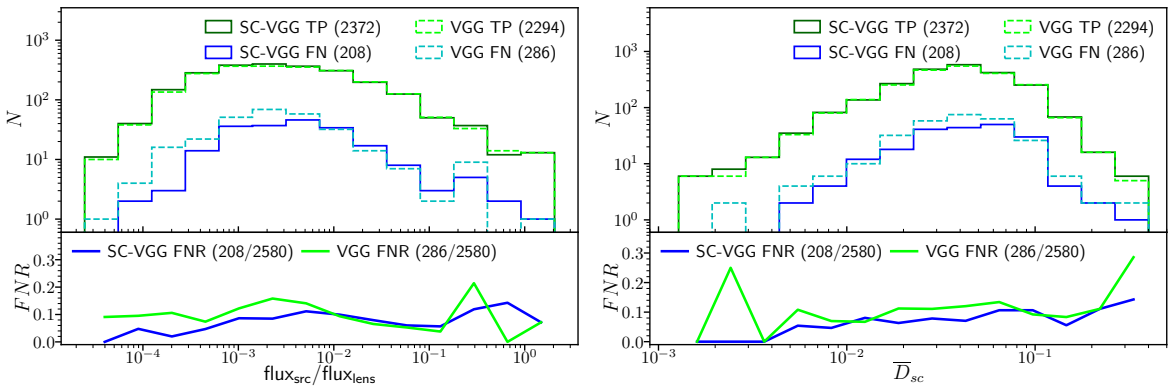


Figure 6.12: True Positive (TP) and False Negative (FP) analysis, related to the VGG and SC-VGG performances, as a function of ratio between source and lens flux (left panel), and distance between the source and caustic centroid normalised to the caustic size ($\overline{D}_{sc}$, right panel). In all panels, TP rates are plotted with green lines, FP rates with blue and cyan lines; solid and dotted lines refer to the SC-VGG and VGG models (see text). Bottom panels show the FN Ratio ($FNR = FN/(TP + FN)$) for each parameter (green line for VGG, blue for SC-VGG).

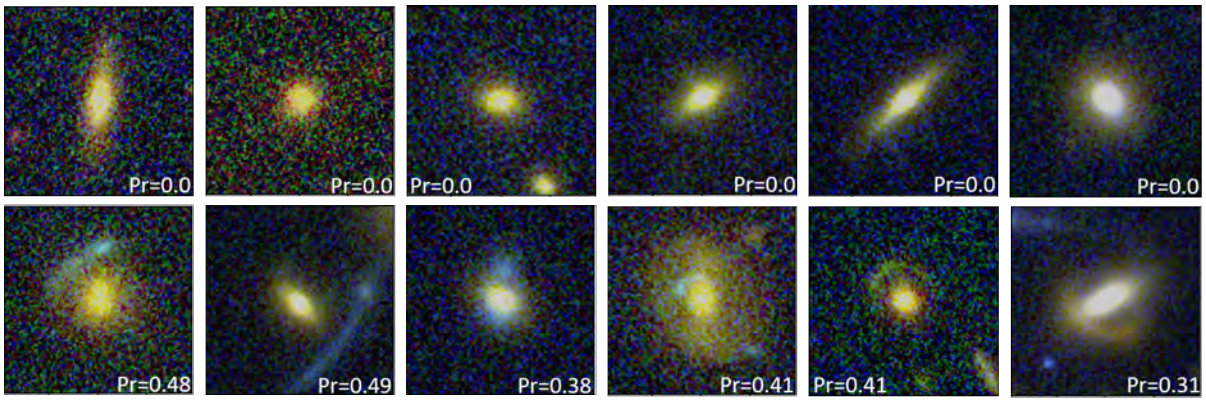


Figure 6.13: Selection of False Negative common to both VGG and SC-VGG models. The probability of belonging to the GGSL class is shown in each thumbnail (referred to the SC-VGG model). Cutouts are $\sim 4''$ across.

and SC-VGG. Model performances differ significantly in this part of the parameter space, where the VGG FN ratio grows up to 0.3 (twice the SC-VGG FNR).

Analogously, the VGG and SC-VGG False Negative dependencies on the source redshift are comparable for $z < 3$ with ratios ~ 0.10, representing the 70% of the whole FN set, but the VGG FNR significantly diverges by increasing the redshift up to ~ 0.21 for $z \geq 6$, respect to the SC-VGG whose FNR remains about constant, ~ 0.07, for $z \geq 3$. This VGG divergence with z, solved by the SC-VGG, can be due to the *drop out* (i.e. the source emits in some bands, but not in others): by increasing z, the source reddening causes the dropping out of the lens event from the image at lower wavelengths toward the image at higher wavelengths. With the VGG, convolutions at first layer combine images corresponding to several filters, mixing and fuzzing the signals, instead, by using the single-channel approach, only filters which carry information (useful to disentangle GGSL from non-GGSL) contribute to the classification, while dropped out bands do not affect the membership. Finally, concerning the dependence on the source effective radius, models have similar behavior for $r_e \in (0.14'', 0.27'')$ with a FNR ~ 0.10, they differ significantly for $r_e \leq 0.14$ where the VGG FNR ratio is ~ 2.5 times larger than that of SC-VGG. Here, we are observing the same dependence on z_{src} parametrised through Eq. 6.4, which can

be evinced from bottom panels in Fig. B.3: small effective radii, $< 0.15''$, are associated with sources at higher redshift, whose classification is affected by the drop out problem.

We also analyse (see Fig. 6.12) the False Negative dependence on the ratio between the source and the flux ratio, together with the distance between the source and caustic centroid normalised to the caustic size ($\overline{D}_{sc}$, i.e. the distance on the source plane divided by the circular radius of the caustic). We do not find any significant dependence on the flux ratio (left panel in Fig. 6.12): the FNR is roughly constant with values around 10%, VGG and SC-VGG FNRs are similar for flux ratios ≥ 0.01 (25% of the GGSL sample), with differences within 2% and a mean of 0.01 ± 0.02%; while for flux ratios < 0.01 (75% of the GGSL set), the difference is slightly more pronounced with a mean $FNR_{VGG} - FNR_{SC-VGG} = 0.06 \pm 0.01$. On the other hand, there is a slight increase of the FNR as a function of $\overline{D}_{sc}$ (right panel in Fig. 6.12): the SC-VGG FNR moves from ~ 5% for $\overline{D}_{sc} < 0.02$, up to ~ 9% for $\overline{D}_{sc} \geq 0.04$, similarly the VGG FNR grows from ~ 7% for $\overline{D}_{sc} < 0.02$, up to ~ 12% for $\overline{D}_{sc} \geq 0.04$.

Finally, a selection of False Negatives is shown in Fig. 6.13, split into two groups: (*i*) adversarial examples on the first row, with source $F814 \in (27.5, 29.0)$ mag and $\theta_E \in (0.10'', 0.25'')$, and (*ii*) "clear" GGSLs on the second row, with visible arc-like features. All the adversarial FNs have probabilities (to be a GGSL) equal to zero, while "clear" could be retrieved by reducing the GGSL probability threshold: Fig. 6.14 shows purity and completeness as a function of the probability threshold. This diagram can be used to select the desired purity or completeness level, particularly, by setting the probability threshold corresponding to the intersection between purity and completeness (i.e. $Pr = 0.25$), the True Positive Rate improves of 3.5%, by paying a price of 1.8% in terms of purity. Clearly, this low probability value increases the number of sources that have to be inspected, which could become a problem in the next future when the number of galaxies will be more than 2 orders of magnitudes larger than the one characterising this work.

The analysis carried out in this section can be compared with other studies even if based on different datasets. For example, we can compare our results with Petrillo et al. (2019a) and Gentile et al. (2022), who trained CNNs on a simulated dataset built with an hybrid approach: arc and ring like features are superimposed on real galaxy images, selected from KiDS (de Jong et al. 2015) and VOICE (Vaccari et al. 2016) surveys, respectively. Specifically, we can compare the False Negative distribution in terms of θ_E (top right panel in Fig. 6.11) and the ratio between the source and the lens fluxes (left panel in Fig. 6.12) with the same distributions obtained by these authors (see Fig. 3 in Petrillo et al. and Fig. 6 in Gentile et al.). With respect to these works, we find similar FNR ranges, especially when compared with Petrillo et al. (2019a), whose FNR as a function θ_E varies in the $4\% - 15\%$ range, while in Gentile et al. FNR values are larger ($10\% - 35\%$). On the other hand, we do not find a similar dependence on the flux ratio, as we measure a FNR roughly constant ($\sim 10\%$), in contrast to the clear decreasing trend found by Gentile et al. (FNR within 60% and 5%) and Petrillo et al. (FNR within 35% and 1%).

6.4 Searching for strong-lenses in galaxy clusters

The experiments described in the previous sections are mostly focused on the classification efficiency and limits of the image-based CNN approach with simulated lenses and evaluating its dependence from observational parameters, such as magnitude and colour, or lens features, such as Einstein radius. In this section, similar to what has been done for the cluster member selection (Sec. 5.5), we are mainly interested in evaluating the degree of generalisation achieved by the trained CNNs in classifying real sources as GGSLs. As introduced in Sec. 5.5, this step process is commonly referred to as *run* in the ML context. In order to maximise the parameter space sampling, we did not use the k-fold approach, instead we exploited the whole source ensemble to train the network, by just excluding the validation set used for the regularisation processes

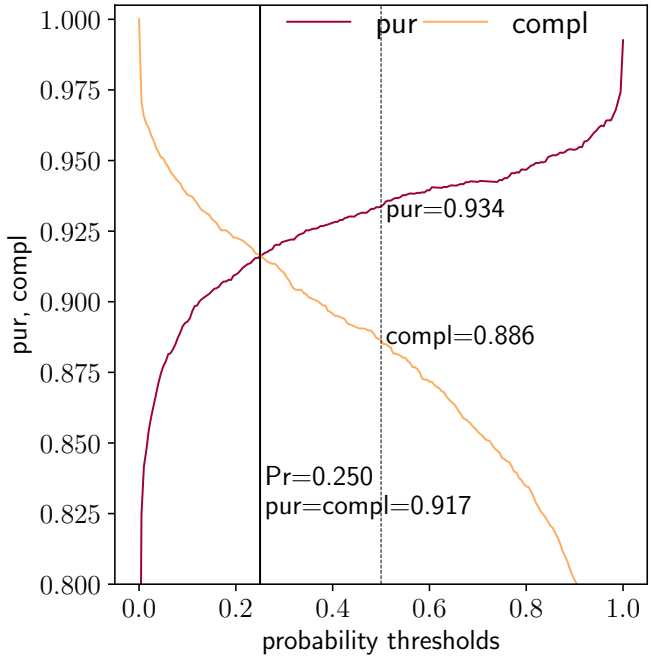


Figure 6.14: Purity (red) and completeness (orange) as a function of GGSL probability threshold. Vertical lines correspond to the purity-completeness intersection at $Pr = 0.25$ (as a solid line) and to the classical threshold at $Pr = 0.50$ (as a dotted line). Purity and completeness values at $Pr = 0.25$ and at $Pr = 0.5$ are reported in the panel.

(see Sec. 5.2), i.e. we did not apply any training-testing split. Firstly, we perform a *run* on 33 already known GGSLs in galaxy clusters to explore the CNN capabilities to find, at least, the known strong-lenses; then we process the networks on a large ensemble of galaxies searching for GGSLs around members in CLASH and HFF galaxy clusters; finally we extended this run phase to the RELICS clusters. We visually inspect the candidate GGSLs with respect to the classification probability.

6.4.1 Performance with real strong-lenses

The confirmed GGSLs used in this test are listed in Tab. 6.5 and shown in Fig. 6.15 and Figs. B.4 - B.7. They are extracted from different works (Desprez et al. 2018; Vanzella et al. 2017b; Diego et al. 2015; Smith et al. 2005; Caminha et al. 2016; Bergamini et al. 2021b, 2019), except for four strong-lensing events whose cutouts are centered on cluster galaxies around the giant arc in M1206 (a.k.a. "snake-arc"), to test CNN generalisation capabilities presenting types of completely different forms (see last four rows in Fig. B.7). These table and figures also show the VGG and SC-VGG predictions, where classification results are organised according to two membership probability thresholds: $Pr > 0.5$, $Pr \in [0.2, 0.5]$ and $Pr < 0.2$, respectively called as True Positive (TP), quasi True Positive (qTP) and False Negative (FN) in Tab. 6.5. On 33 processed GGSLs, 24 of them are common classified by both models, of which 22 are correctly classified. By considering as correct also prediction with $Pr \in [0.2, 0.5]$, VGG True Positive are 28, i.e. 85% (26 by excluding qTPs, i.e. 79%), whereas SC-VGG scores 27 TPs, i.e. 82% (25 by excluding qTPs, i.e. 76%). All typical lenses, with arc or ring like features have been correctly classified (see, for example, the Einstein rings, fourth row in Fig. B.7 or the third and forth row in Fig. B.6), even the strong-lensing events related to the snake-arc in M1206 (see last four rows in Fig. B.7) are predicted as GGSLs by both models with high probabilities ($0.7 - 1.0$). Concerning the common FNs, both models failed the identification of the GGSL in A383 (named

RA	DEC	Cluster	Image ref	VGG	SC-VGG	Reference
22.9577568	-13.6032558	A209	A1	TP	TP	Desprez et al. (2018)
22.9648793	-13.6363138	A209	A2	TP	TP	Desprez et al. (2018)
42.0113589	-3.5480288	A383	B1	FN	FN	Desprez et al. (2018)
52.4201304	-2.2216321	M0329	C1	TP	TP	Desprez et al. (2018)
64.0340808	-24.0667448	M0416	E1	FN	FN	Vanzella et al. (2017b)
64.0284705	-24.085668	M0416	E2	qTP	FN	Desprez et al. (2018)
64.0170899	-24.0895541	M0416	E3	TP	TP	Diego et al. (2015)
64.04275	-24.06316	M0416	E4	TP	TP	Bergamini et al. (2021b)
64.03262	-24.06838	M0416	E5	TP	TP	Bergamini et al. (2021b)
64.03250	-24.07849	M0416	E6	TP	TP	Bergamini et al. (2021b)
64.02442	-24.08106	M0416	E7	TP	FN	Bergamini et al. (2021b)
168.956259	1.4974098	M1115	F1	TP	TP	Desprez et al. (2018)
177.403888	22.426630	M1149	D1	TP	FN	Smith et al. (2005)
177.393135	22.411336	M1149	D2	TP	qTP	Smith et al. (2005)
206.896032	-11.7536032	R1347	H1	FN	TP	Desprez et al. (2018)
206.87105	-11.76684	R1347	H2	TP	TP	Caminha et al. (2019)
22.4287798	0.1080707	R2129	I1	TP	TP	Desprez et al. (2018)
342.1557424	-44.5459123	R2248	G1	qTP	TP	Desprez et al. (2018)
342.1633643	-44.5297236	R2248	G2	FN	qTP	Desprez et al. (2018)
342.18205	-44.54035	R2248	G3	TP	TP	Caminha et al. (2016)
342.18867	-44.54015	R2248	G4	TP	TP	Caminha et al. (2016)
342.16691	-44.53483	R2248	G5	TP	TP	Caminha et al. (2016)
342.17554	-44.53558	R2248	G6	FN	TP	Caminha et al. (2016)
67.4020771	-2.8713932	M0429	J1	TP	TP	Desprez et al. (2018)
67.3892478	-2.8741192	M0429	J2	TP	FN	Desprez et al. (2018)
116.2121685	39.4598681	M0744	K1	TP	TP	Desprez et al. (2018)
181.56667	-08.80478	M1206	L1	TP	TP	Bergamini et al. (2019)
181.56532	-08.80608	M1206	L2	TP	TP	Bergamini et al. (2019)
181.55309	-08.79486	M1206	L3	TP	TP	Bergamini et al. (2019)
181.54490	-08.80259	M1206	L4	TP	TP	
181.54494	-08.80180	M1206	L5	TP	TP	
181.54482	-08.80064	M1206	L6	TP	TP	
181.54460	-08.79928	M1206	L7	TP	TP	
			N_{TP}	26	25	
		TOTAL	N_{FN}	5	6	
			N_{qTP}	2	2	

Table 6.5: Catalogue of known GGSLs processed by both VGG (fifth column) and SC-VGG (six column) networks, together with a reference to the images shown in Fig. 6.15 and Figs. B.4 - B.7. GGSL references are quoted in last column, except for the last four objects which are parts of the giant arc in M1206 (see specifically Fig. B.7). According to the membership probability computed by the CNN models, classification results are organised as: True positive (TP, $Pr > 0.5$), quasi True Positive (qTP, $Pr \in [0.2, 0.5]$), False Negative (FN, $Pr < 0.2$).

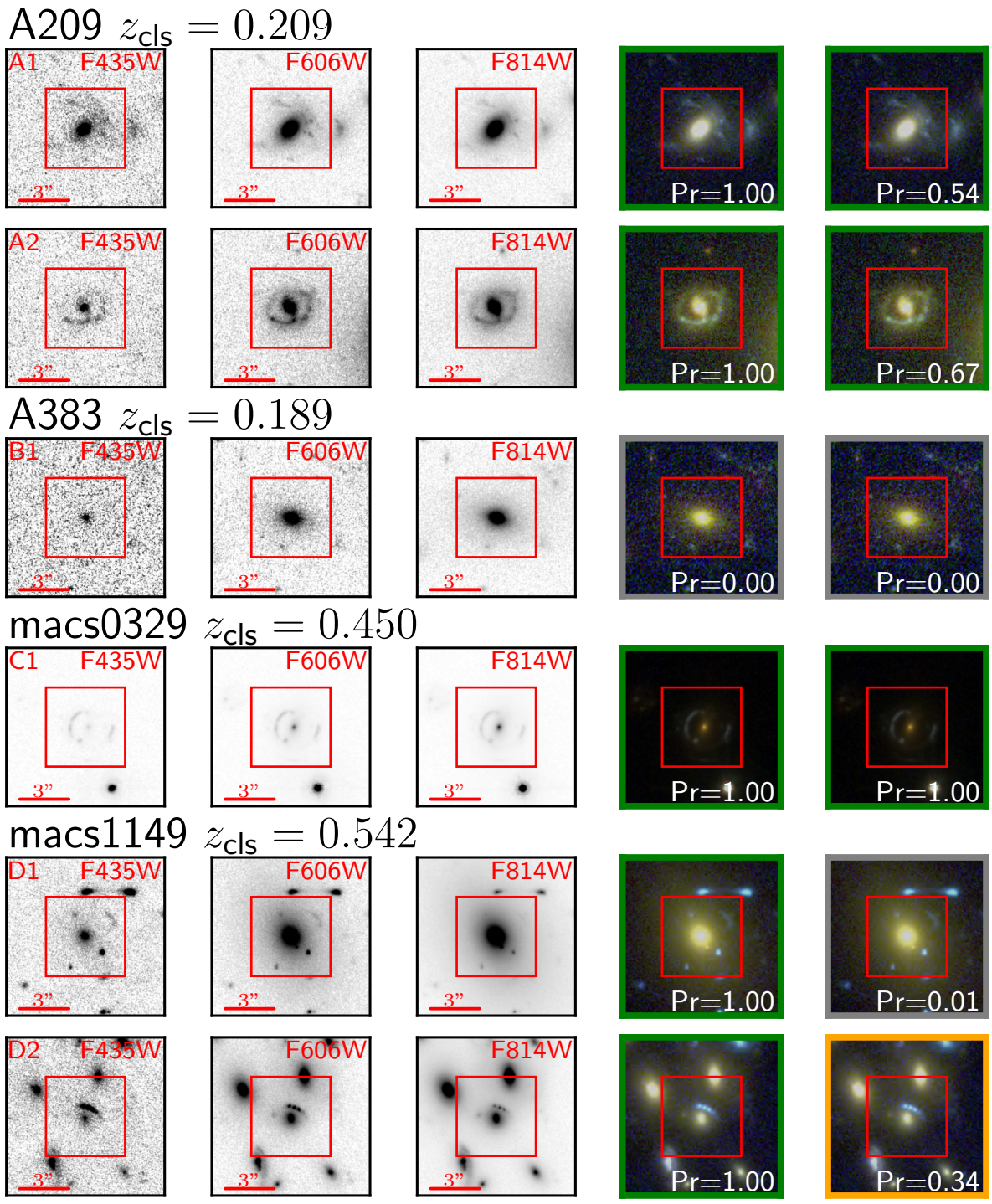


Figure 6.15: Ensemble of GGSLs processed by both VGG (fourth column) and SC-VGG (fifth column) models, listed in Tab. 6.5. Cutouts are $\sim 8''$ across. The inner red square ($\sim 4''$ across) represents the actual image processed by the network. RGB cutouts on the last two columns are obtained by combining $F435$, $F606$, $F814$ bands (shown on the first three column), they are surrounding by a box coloured in green, orange or grey according to their classification probability: $Pr > 0.5$, $Pr \in [0.2, 0.5]$ or $Pr < 0.2$, respectively. Here, images are extracted from 4 clusters: A209, A383, M0329 and M1149; the rest of the processed GGSLs are shown in Figs. B.4 - B.7 in appendix B.

				VGG		SC-VGG		Common		
Cluster	$z_{cluster}$	E[N]	N	Pr≥ 0.2	Pr> 0.5	Pr≥ 0.2	Pr> 0.5	Pr≥ 0.2	Pr> 0.5	Fig. ref
A383	0.189		77	6	5	12	7	4	4	B.8
A209	0.209		84	8	3	10	3	4	1	B.8
R2129	0.234	0.75	58	11	10	18	9	10	6	B.10
A2744	0.308		140	8	6	15	8	6	3	B.10
MS2137	0.316		54	8	4	13	9	5	1	B.10
R2248	0.346	0.96	206	18	11	27	11	13	6	6.16
M1931	0.352	1.03	37	9	7	10	7	7	5	B.9
M1115	0.352		119	13	12	16	6	11	4	B.9
A370	0.375		203	18	15	26	11	10	6	B.9
M0416	0.397	0.95	147	26	21	33	15	20	10	B.9
M1206	0.439	2.96	234	52	41	65	39	42	27	6.16
M0329	0.450	1.09	89	15	10	19	14	12	7	B.8
R1347	0.451	3.67	53	15	13	14	8	11	7	B.10
M1311	0.494		59	2	1	8	2	2	0	B.8
M1149	0.542		149	19	19	22	8	11	5	B.8
M2129	0.587	1.34	64	16	15	21	11	15	9	B.10

Table 6.6: Summary of the GGSL *run* process performed on 16 CLASH and HFF clusters. The expected number of GGSLs is shown in Col. 3 (only for cluster whose lens model is available), estimated by following Meneghetti et al. (2020). The number of processed cluster galaxies is listed in Col. 4. The amounts of sources classified as GGSLs with probability ≥ 0.2 and > 0.5 identified by the VGG model are listed in Col. 5 and Col. 6, identified by the SC-VGG model in Col. 7 and Col. 8, while their intersections are shown on Col. 9 and Col. 10. Figure reference for each cluster is reported in the last column.

B1, third row in Fig. 6.15), classified by Desprez et al. (2018) through visual inspection, which should be a faint Einstein cross, partially outside the cutout; and the E1 source in M0416 (first row in Fig. B.4), named as ID.14 by Vanzella et al. (2017b), spectroscopically confirmed. This latter misclassification could be imputed to the configuration of the lens composed by two cluster galaxies, such configurations have not been expressly included in the training set, even if they occurred by chance (< 0.01% of lenses are composed by two or more members). Other partially misclassified sources are G2 and G6 images in R2248 (third and last rows in Fig. B.5), H1 image in R1347 (first row in Fig. B.6) and J2 image in M0429 (fifth row in Fig. B.4), whose classification were affected by the unique morphology of these examples, which have not been reproduced in the training set.

6.4.2 Searching for GGSLs in CLASH and HFF galaxy clusters

As second step, we perform the *run* on the 16 involved CLASH and HFF clusters, by extracting a squared cutout with a side of $\sim 4''$ centred on the cluster member position, selected with a cluster rest-frame velocity separation of $\pm 5000\,\mathrm{km\,s^{-1}}$. The number of sources classified as GGSLs by both the VGG and SC-VGG, together with their common predictions, are listed in Tab. 6.6, where we have also reported the expected number of GGSLs in galaxy clusters (with available lens model), computed by following the same approach as the one carried out by Meneghetti et al. (2020). Based on the redshift distribution of sources extrapolated from COSMOS2015, we estimate the lensing cross-section at each z_s, $\sigma(z_s)$, and the number density of sources, $n(z_s)$. Thus, we determine the number of expected GGSLs for cluster as $N = \int \sigma(z_s) n(z_s) dz_s$; as previously done, we exclude critical lines with $\theta_E \leq 0.2''$ and $\theta_E > 5''$ from the computation of the lensing cross-section.

The union between VGG and SC-VGG objects classified as GGSLs (with probability ≥ 0.2)

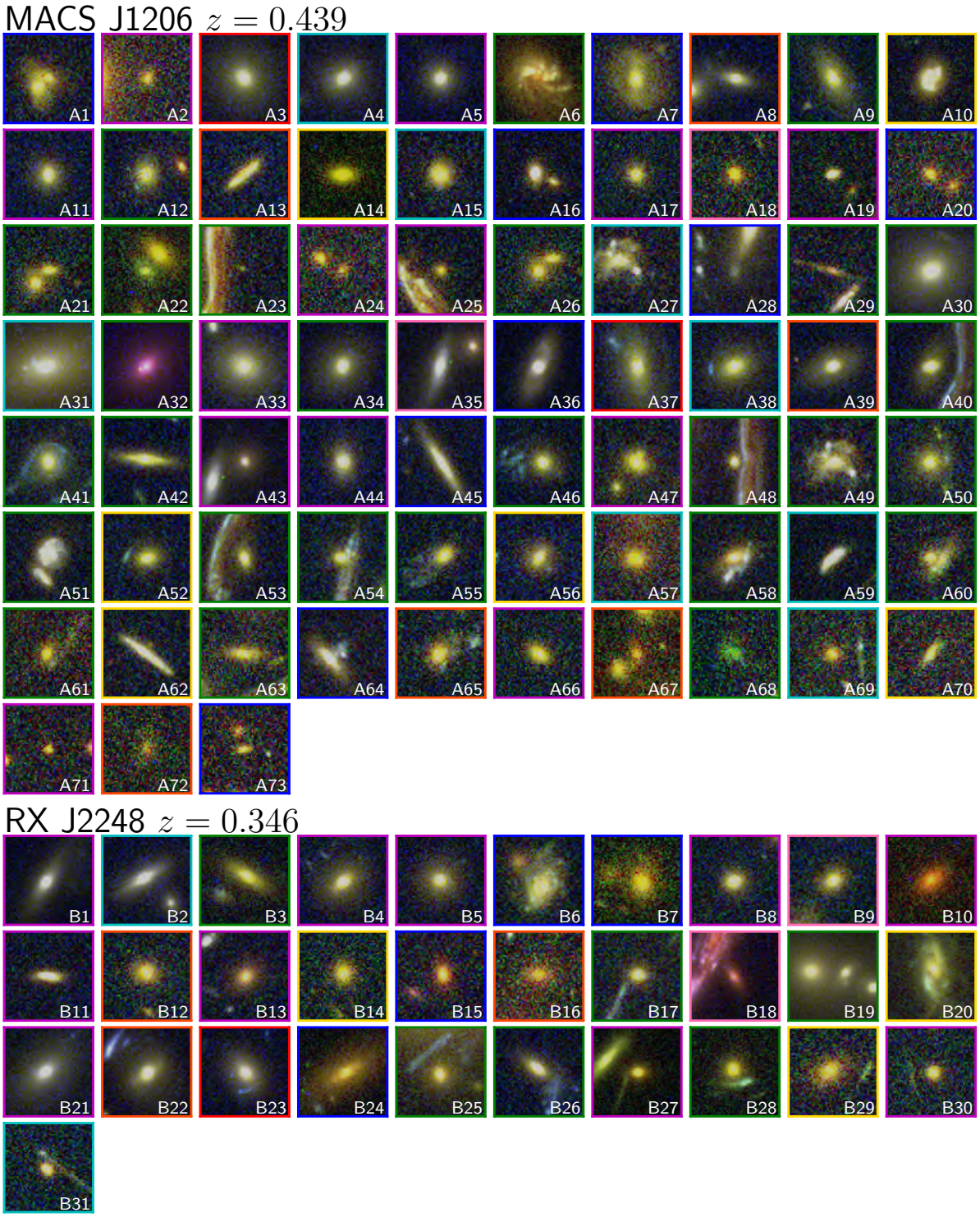


Figure 6.16: Identified GGSLs by both the models, by searching them around cluster galaxies, related to M1206 and R2248. Each cutout is surrounding by a coloured square according to the follow scheme based on the GGSL probability:
(*i*) $Pr(VGG)$, $Pr(SC\text{-}VGG) > 0.5 \rightarrow$green;
(*ii*) $Pr(VGG)$, $Pr(SC\text{-}VGG) \in [0.2, 0.5] \rightarrow$orange;
(*iii*) $Pr(VGG) \in [0.2, 0.5] \wedge Pr(SC\text{-}VGG) > 0.5 \rightarrow$red;
(*iv*) $Pr(SC\text{-}VGG) \in [0.2, 0.5] \wedge Pr(VGG) > 0.5 \rightarrow$yellow;
(*v*) $Pr(VGG) \in [0.2, 0.5] \wedge Pr(SC\text{-}VGG) < 0.2 \rightarrow$pink;
(*vi*) $Pr(SC\text{-}VGG) \in [0.2, 0.5] \wedge Pr(VGG) < 0.2 \rightarrow$magenta;
(*vii*) $Pr(VGG) > 0.5 \wedge Pr(SC\text{-}VGG) < 0.2 \rightarrow$cyan;
(*viii*) $Pr(SC\text{-}VGG) > 0.5 \wedge Pr(VGG) < 0.2 \rightarrow$blue.

is listed in Tabs. B.5 - B.7, and shown in Fig. 6.16 and Figs. B.8 - B.10, where, each cutout is surrounding with a coloured square, according to the probability to be a GGSL measured by both the networks; based on this colour scheme predictions have been split into:

- Pr(VGG) $\wedge$ Pr(SC-VGG) > 0.5: both models predicts the same membership with higher probability (green);
- Pr(VGG) $\wedge$ Pr(SC-VGG) $\in [0.2, 0.5]$: for both the models the source has a certain (lower) probability to be a GGSL (orange);
- Pr(VGG)$\in [0.2, 0.5] \wedge$ Pr(SC-VGG)> 0.5: the VGG predicts the membership with a probability lower than that predicted by the SC-VGG, however greater than 0.2 (red);
- Pr(SC-VGG)$\in [0.2, 0.5] \wedge$ Pr(VGG)> 0.5: the SC-VGG predicts the membership with a probability lower than that predicted by the VGG, however greater than 0.2 (yellow);
- Pr(VGG)$\in [0.2, 0.5] \wedge$ Pr(SC-VGG)< 0.2: there is a slight conflict between models predictions, however both models predict the membership with probability less than 0.5 (pink);
- Pr(SC-VGG)$\in [0.2, 0.5] \wedge$ Pr(VGG)< 0.2: there is a slight conflict between models predictions, however both models predict the membership with probability less than 0.5 (magenta).
- Pr(VGG)$> 0.5 \wedge$ Pr(SC-VGG)< 0.2: there is a conflict between models predictions, the object is classified as GGSL by the VGG and as non-GGSL by the SC-VGG (cyan);
- Pr(SC-VGG)$> 0.5 \wedge$ Pr(VGG)< 0.2: there is a conflict between models predictions, the object is classified as GGSL by the SC-VGG and as non-GGSL by the VGG (blue);

With this division, we can analyse *all* the networks predictions (i.e. the union), keeping them separated and, thus, allowing to compare the models behaviour. Globally, models share 101 classifications with Pr> 0.5 (which represent more than a half of the whole prediction sets) and 183 with Pr≥ 0.2 (equals to the 75% and 56% of the candidate GGSLs identified by VGG and SC-VGG, respectively). The medians (referred to prediction with Pr> 0.5) are 11 and 17, respectively for the VGG and SC-VGG, whereas it is 5.5 by considering only the common candidate GGSLs. By exploring the intersection between the two candidate sets, both with Pr> 0.5 (i.e. the green cutouts in Fig. 6.16 and Figs. B.8 - B.10), we find that $\sim$ 60% of these cutouts are characterised by the typical strong-lensing morphology, presenting more or less pronounced arc and ring like structures (some of them have been already processed as true GGSL in Sec. 6.4.1). By visual inspecting this ensemble of GGSL candidates, we find that it is contaminated by disc galaxies ($\sim$ 4%) and by isolated elliptical galaxies, which recall the adversarial examples ($\sim$ 15%). By reducing the probability threshold and the level of agreement between models, the degree of contaminators increases: by considering Pr(VGG) $\vee$ Pr(SC-VGG) $\geq 0.2 -$\{Pr(VGG) $\wedge$ Pr(SC-VGG)>0.5\} (i.e. the orange, red and yellow cutouts in Fig. 6.16 and Figs. B.8 - B.10), we find $\sim$ 5% of disc like galaxies and $\sim$ 30% of ellipticals, whereas $\sim$ 20% of cutouts show strong-lensing features. By examining predictions whose probabilities are less than 0.5 for both models (i.e. pink and magenta cutouts in Fig. 6.16 and Figs. B.8 - B.10, corresponding to [Pr(VGG) $\in [0.2, 0.5] \wedge$ Pr(SC-VGG)$< 0.2] \vee$ [Pr(SC-VGG) $\in [0.2, 0.5] \wedge$ Pr(VGG)< 0.2]) we measure similar percentages respect to ellipticals and late-types ($\sim$ 28% and $\sim$ 7%, respectively), while the likely GGSL fraction is significantly decreased ($\sim$ 7%). These fractions are strictly comparable with that estimated by considering the last two cases,

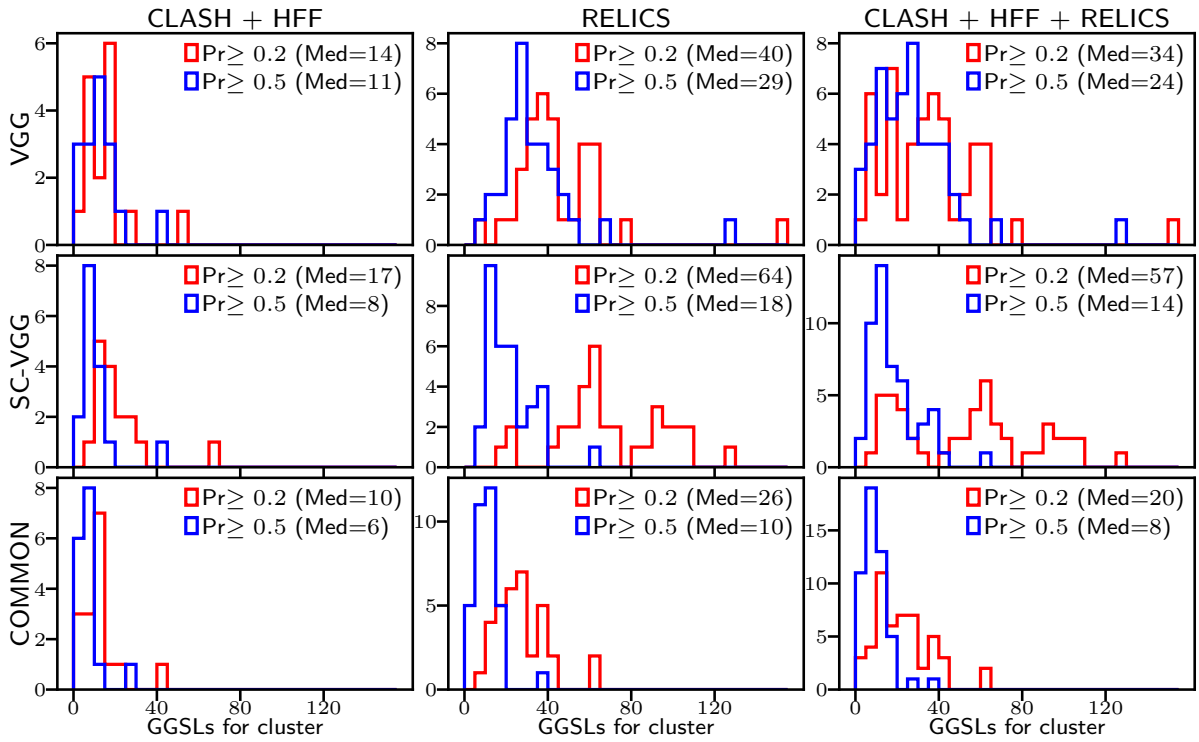


Figure 6.17: Distribution of the number of GGSLs for cluster, identified by VGG (top panles), by SC-VGG (middle panels) and by tacking their common predictions (bottom panels). All diagrams show the distribution related to the GGSL candidates with Pr≥ 0.2 (in red) and with Pr> 0.5 (in blue). Median of distributions is labelled in the legend.

when networks predictions are not in agreement ([Pr(VGG)> 0.5 ∧ Pr(SC-VGG)< 0.2] ∨ [Pr(SC-VGG)> 0.5 ∧ Pr(VGG)< 0.2], corresponding to blue and cyan cutouts in Fig. 6.16 and Figs. B.8 - B.10).

Thus, GGSL candidates appear reliable when they have been classified by both CNNs in the same way, otherwise their soundness falters. This is evident by looking at Fig. 6.17, where the distributions of the number of GGSLs for cluster is plotted respect to: (*i*) the VGG and SC-VGG classification (top and middle panels) and by taking their common predictions (bottom panels); (*ii*) by varying the involved survey, CLASH and HFF clusters (on the first column), RELICS clusters (on the second one, whose classification is discussed in Sec. 6.4.3), while the combinations of all the surveys are plotted on the third column. There is a clear excess of GGSL number for cluster due to the False Positives (disc galaxies and, especially, adversarial like galaxies), which increases when the probability threshold is reduced. However, by considering the common predictions with $Pr > 0.5$, the number of GGSLs per cluster (with a median values of 6 to 10), is somewhat higher than the expected value based on the cross-section inferred from the lens model (listed in Tab. 6.6), typically $\lesssim 3$. This confirms the need to perform a visual inspection of the candidates found by the CNNs to derive a set of GGSLs with high purity.

6.4.3 Searching for GGSLs in RELICS galaxy clusters

As final application of our trained networks, similarly to what has been done for the member identification (see Sec. 5.5.1), we perform a *run* on 33 RELICS clusters (Coe et al. 2019), with central redshift ranging in $(0.20, 0.87)$. A layout of the involved RELICS clusters is shown in Fig. 4.8 in Chap. 4. With the aim of searching GGSLs around cluster galaxies, we exploit the members classified by the CNN (see Tab. 5.10). Given this classification, we extract a squared

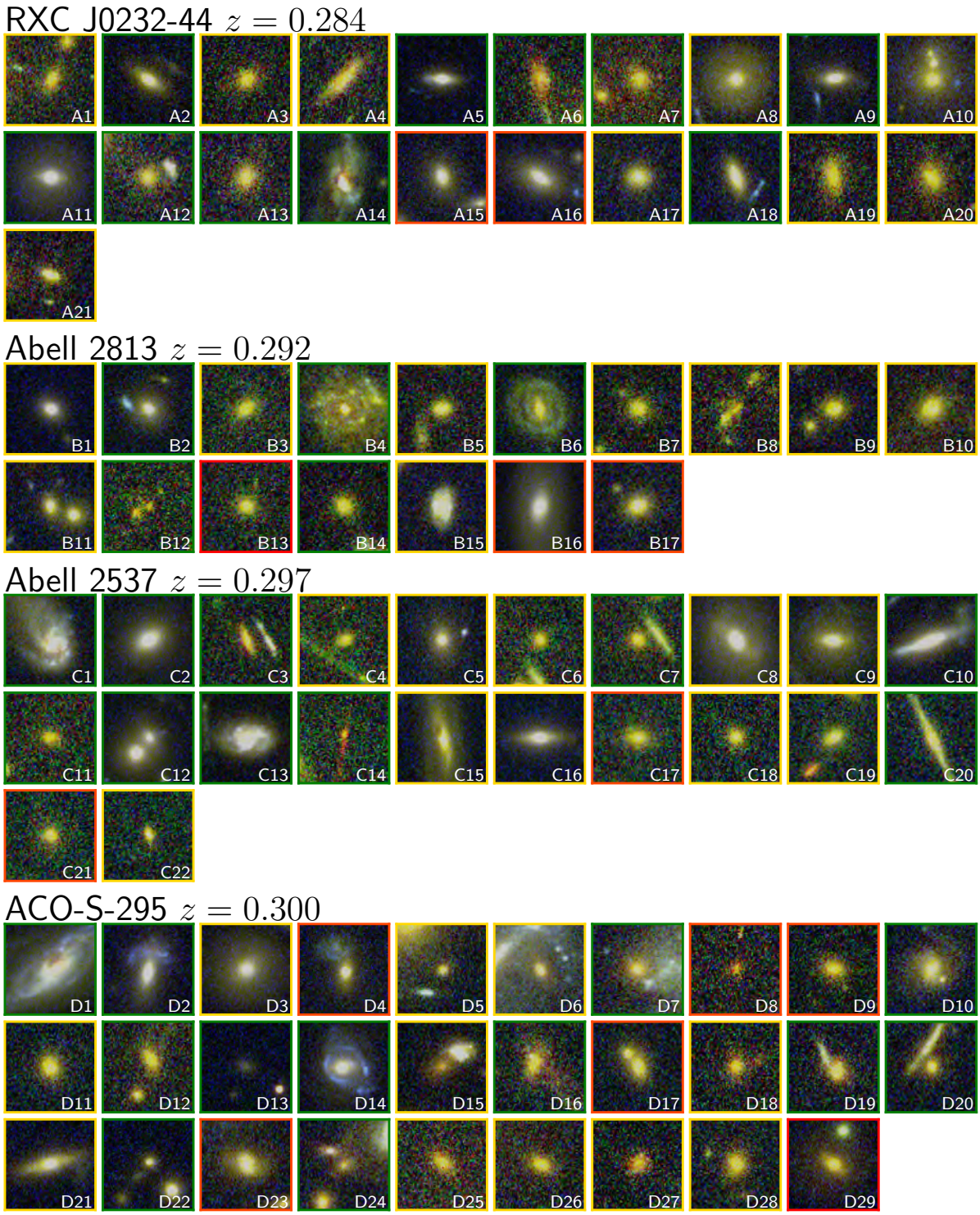


Figure 6.18: Identified GGSLs by both the models, by searching them around RELICS cluster galaxies detected with CNN (see Sec. 5.5.1), related to R0232, A2813, A2537, A295. Each cutout is surrounding by a coloured square according to the follow scheme based on GGSL probability:
(*i*) $Pr(VGG)$, $Pr(SC\text{-}VGG) > 0.5 \rightarrow$green;
(*ii*) $Pr(VGG)$, $Pr(SC\text{-}VGG) \in [0.2, 0.5] \rightarrow$orange;
(*iii*) $Pr(VGG) \in [0.2, 0.5] \wedge Pr(SC\text{-}VGG) > 0.5 \rightarrow$red;
(*iv*) $Pr(SC\text{-}VGG) \in [0.2, 0.5] \wedge Pr(VGG) > 0.5 \rightarrow$yellow.

				VGG		SC-VGG		Common		
Cluster		$z_{cluster}$	N	Pr≥ 0.2	Pr> 0.5	Pr≥ 0.2	Pr> 0.5	Pr≥ 0.2	Pr> 0.5	Fig. ref
A665	Abell 665	0.182	46	8	6	15	5	5	1	B.11
A2163	Abell 2163	0.203	251	64	44	105	34	42	16	B.11
A520	Abell 520	0.203	119	38	29	57	23	26	12	B.11
R1514	RXC J1514-15	0.223	113	29	21	40	13	16	6	B.11
A1763	Abell 1763	0.228	158	41	31	58	16	21	9	B.12
P171	PLCK G171-40	0.270	176	28	17	60	14	14	1	B.12
A1758	Abell 1758	0.280	278	61	46	99	36	38	17	B.12
A697	Abell 697	0.282	152	39	22	49	12	19	4	B.12
R0232	RXC J0232-44	0.284	156	31	27	61	21	21	10	6.18
A2813	Abell 2813	0.292	163	35	26	57	11	17	5	6.18
A2537	Abell 2537	0.297	170	44	38	52	16	22	10	6.18
AS295	Abell s295	0.300	180	39	29	83	32	29	12	6.18
A1300	Abell 1300	0.308	141	32	23	51	17	22	8	B.13
R0142	RXC J0142+44	0.341	167	31	20	61	12	13	4	B.13
M0035	MACS 0035-20	0.352	144	27	19	46	11	15	4	B.13
M0308	MACS 0308+26	0.356	218	33	22	68	12	19	6	B.13
R0949	RXC J0949+17	0.383	195	42	33	72	22	29	12	B.13
P287	PLCK G287+32	0.390	246	57	44	104	36	37	14	B.14
SM0723	SMACS 0723-73	0.390	142	35	26	57	11	26	8	B.14
R0032	RXC J0032+18	0.396	239	52	39	101	32	29	10	B.14
R2211	RXC J2211-03	0.397	222	59	38	92	27	33	16	B.15
M0159	MACS 0159-08	0.405	187	49	33	64	16	28	10	B.15
A3192	Abell 3192	0.425	210	56	34	99	37	35	15	B.15
M0553	MACS 0553-33	0.430	193	34	28	64	29	23	10	B.16
S0254	SPT-CLJ0254-58	0.438	157	43	29	74	21	31	8	B.16
R0600	RXC J0600-20	0.460	254	61	46	106	38	37	15	B.19
P308	PLCK G308-20	0.480	205	63	54	92	19	36	10	B.17
P004	PLCK G004-19	0.540	205	77	66	126	61	61	36	B.17
R0018	RXC J0018+16	0.546	222	152	129	86	20	64	14	B.18
W0137	WHL J0137-08	0.566	172	35	28	63	12	23	7	B.19
P209	PLCK G209+10	0.677	150	44	37	67	17	28	8	B.19
P138	PLCK G138-10	0.702	56	23	13	20	6	13	6	B.14
A0102	ACT-CLJ0102-49	0.870	190	58	44	94	23	42	13	B.18
S0615	SPT-CLJ0615-57	0.972	47	17	11	22	11	12	7	B.16

Table 6.7: Summary of the GGSL *run* process performed on 33 RELICS clusters. The number of processed cluster galaxies is listed in Col. 4. The amounts of sources classified as GGSLs with probability ≥ 0.2 and > 0.5 identified by the VGG model are listed in Col. 5 and Col. 6, identified by the SC-VGG model in Col. 7 and Col. 8, while their intersections are shown on Col. 9 and Col. 10. Figure reference for each cluster is reported in the last column.

cutout with a side of ∼ 4″ centred on the candidate member position (using HST images with a sampling of 0.030″/pixel). The number of GGSLs for cluster identified by both VGG and SC-VGG, together with their common predictions, is listed in Tab. 6.7; while all the candidate GGSLs with Pr(VGG) ∧ Pr(SC-VGG)> 0.2 are shown in Fig. 6.18 and Figs. B.11 - B.19, and listed in Tabs. B.8 - B.15. We restrict the analysis to this subset of cutouts, using the same colour scheme adopted for CLASH and HFF (only for Pr(VGG) ∧ Pr(SC-VGG)> 0.2, i.e. the first four items in Sec. 6.4.2), based on which, we found:

- by considering predictions with Pr> 0.5 shared by both models (i.e. green cutouts in Fig. 6.18 and Figs. B.11 - B.19), we identify a fraction ∼ 25% of candidate GGSL with evident strong-lensing features, ∼ 17% of ellipticals and ∼ 8% of disc like galaxies;

- for all the other examples (i.e. orange, yellow and red cutouts in Fig. 6.18 and Figs. B.11 - B.19), there is a large fraction of isolated elliptical galaxies (∼ 45%), a negligible fraction

of late-types ($\sim 2\%$), whereas the cutouts, showing visible arc and ring like features, are just $\sim 4\%$.

These considerations can be also deduced by comparing plots in Fig. 6.17: the number of GGSLs for RELICS cluster is too high to be reliable, but this False Positive excess is strongly softened by exploiting common predictions between models, with a decreasing factor $\gtrsim 2$: by considering probability thresholds of 0.5 and by taking the intersection between models candidate GGSLs, we measure a median value for all the HST clusters ~ 8.

6.5 Conclusions

In this chapter, we presented an approach for the galaxy-galaxy strong-lenses identification in galaxy clusters by training CNNs with realistic simulations embedded in the complexity of observed data. Simulated examples have been produced by exploiting robust and accurate cluster lens models, allowing to introduce the contribution of the overall cluster mass density distribution, while the non-lenses have been selected by using MUSE and CLASH-VLT spectroscopic observations, combined with a visual selection performed by expert astronomers. We performed several experiments, involving several CNNs and focusing on the results achieved by the two best models (VGG and SC-VGG), exploring the CNNs dependence on: (*i*) the cutout sizes and measuring the impact of the adversarial examples (Sec. 6.3.1), (*ii*) on the source redshift, magnitude, effective radius and on lens colours, magnitude and Einstein radius, by studying the False Positive and False Negative distributions (Sec. 6.3.2). We tested the CNNs capabilities to recognise at least the known GGSL and to find candidate strong-lenses in galaxy clusters, by processing hundreds of cutouts centred on spectroscopic confirmed members (CLASH and HFF clusters) or candidate members (selected by CNN in RELICS clusters). The main results can be summarised as follows:

- we develop of a methodology able to simulate truthful GGSLs, by sampling parameters (e.g., source magnitude, redshift and effective radius) from observed data (Sec. 6.2);
- CNNs are able to identify simulated GGSLs with a F1-score ranging in $84\% - 92\%$, with purity typically higher than completeness; although adversarial examples increase the False Negative rate, their presence ensures the prevention from model overfitting (Sec. 6.3.1);
- False Negatives typically gather at high redshift, in magnitude faint-end and for small-scaled lenses, while False Positive rate increase for bluer region of the parameter space (Sec. 6.3.2);
- Networks are able to identify $\gtrsim 80\%$ of the processed real GGSLs (Sec. 6.4.1), and to produce a reliable selection of candidate GGSLs, that, however, have to be post-processed by visually inspecting the cutouts in order to remove False Positives and extract a purer set of candidates (Secs. 6.4.3 and 6.4.2);

With the analysis carried out in this chapter, we tried to explore all the networks predictions, enlarging as possible as the set of candidate GGSLs, even if with this approach we included a large fraction of False Positives. We also underlined that classifications are more reliable when we consider the intersection between models predictions with higher probabilities; although we measured a significant purity rate increased, in this way several strong-lenses have been misclassified, resulting into a higher False Negative rate. There is not an always corrected way, it is dependent on the specific goal of the work. Here, we tried to explore both the solutions, studying strengths and weaknesses of these complementary strategies.

In the next future, we are planning to improve the the GGSL simulation, for example, by injecting background galaxies with substructures (blobs generated with small-scale Sérsic profiles), as has been done by Petrillo et al. (2019a,b) and Gentile et al. (2022), or by injecting real high-z galaxies extracted from the Hubble Ultra-Deep Field (Rafelski et al. 2015). We also plan to extend our analysis to other clusters whose lens model has been recently constructed, e.g. PSZ1 G311.65-18.48 (Pignataro et al. 2021), or SDSS J1029+2623 (Acebron et al. 2021). Moreover, given the expected impressive number of galaxy clusters which will be observed with upcoming surveys, we intend to explore other neural network architectures, such as Faster Region CNN (Ren et al. 2015) and Masked Region CNN (He et al. 2017) to automatically detect GGSLs by processing directly the whole cluster FoVs.

Chapter 7

Cross-correlation tool for 3D spectroscopy

In this chapter, I present the implementation and the application of a GPU-optimised cross-correlation tool (XC-tool), built to measure redshifts in an automated fashion from the MUSE (Multi Unit Spectroscopic Explorer) data cubes (see Sec. 4.2), which is an essential information for the applications described in the previous chapters. The GPU-based approach allows us to complete all the cross-correlation operations, at the pixel level, in few tens of seconds (to compare with other MPI strategies which require $\sim$ 50 minutes), which makes this approach suitable for the upcoming surveys, characterised by huge volume of data. This method has been validated with simulations and tested by comparing its performance with conventional methods which heavily rely on manual analysis. Such a tool is an important component of the work flow at the basis of this thesis, specifically it can be used to automatically measure redshifts of cluster galaxies and multiple images which are essential to build the lens models. In addition, a large sample of spectroscopic member galaxies is critical for training CNN in identifying clusters members. The latter allow the input cutouts to be selected for the CNN-based search for GGSL described in the previous chapter. As an additional application of this new cross-correlation tool, we show how we can efficiently and automatically reconstruct velocity maps of lensed disk galaxies. This information can be used to model the effect of perturbers on the cluster mass distribution and is interesting in itself since one can derive rotation curves out to $z \sim 2$.

7.1 Introduction

The usage of cross-correlation to estimate radial velocities by cross-correlating the source spectrum with a target template spectrum has long been employed (e.g., Griffin 1967; Simkin 1974; Tonry & Davis 1979). Over the time, many different approaches have been developed: employing Fast Fourier Transform (e.g., Tonry & Davis 1979; Heavens 1993; Statler 1995; Torres et al. 2007), or by decomposing galaxy spectra into several orthogonal template (e.g., Glazebrook et al. 1998), however the computational resources available to the modern astronomy has made Fourier techniques no longer essential and cross-correlation can be performed in the Real domain (e.g., Zucker 2003; Garilli et al. 2010) avoiding the significant limitation which characterised Fourier-based approaches: the equal weight of each pixel, regardless for the signal-to-noise ratio, which turns out into a method without flexibility (Kelson et al. 2003). However, several issues continue to be not completely solved: the selection of templates (e.g., observed vs. synthetic), the combination of measurements coming from different templates, the determination of the "true" peak of the cross-correlation function (peaks detection and relevance of secondary peaks), the reliability of the metrics used to evaluate the similarity between spectra and template.

Typically, spectroscopic samples are composed by galaxies pre-selected on photometric criteria (Noll et al. 2004; Balestra et al. 2010; Le Fèvre et al. 2013), however this selection acts as

a sort of bias on the resulting spectroscopic sample. Moreover, multi-object spectrographs have a limited degree of freedom in choosing sources in simultaneous observation and the prefigured slit mask significant reduce source fluxes (Herenz et al. 2017). Integral field spectroscopy (IFS) can overcame these problems: instead of individual targets, this instrument maps a contiguous area in the sky, providing spectral information for the whole FoV. The IFS MUSE at ESO Very Large Telescope (described in Sec. 4.2) is a powerful instrument able to perform a blind survey for faint and high-redshfit galaxies (Bacon et al. 2014; Caillier et al. 2014), allowing measurements for hundreds of galaxies inside a single MUSE FoV of 1 arcmin^2 (Bacon et al. 2015).

In this work, we present a GPU-based implementation of the cross-correlation between the whole MUSE data cube and a set of reference templates. The tool has been developed in `python` exploiting the open source library `tensorflow`, which allows a significant computing time improvement ($\sim$ 100 times faster). We validated the cross-correlation tool by processing a simulated galaxy (Sec. 7.3.1) and by estimating velocity maps of Refsdal host, which supports a direct comparison with the analysis carried by Grillo et al. (2016) (see Sec. 7.3.2). Finally, we estimate redshift for 274 sources in the (northern) MUSE FoV of M0416 cluster (Sec. 7.3.3).

7.2 Implementation

In this work, the spectral cross-correlation is implemented as a vectorised convolution (similar to Das 1991). The cross-correlation (XC) between two function s and t is defined as:

$$[s \star t](x) = \int_{-\infty}^{+\infty} s(\xi) t(x + \xi) d\xi \tag{7.1}$$

In signal theory, the function s e t are called *signal* and *impulse response*. Here they represent the spectrum and the reference template as a function of the wavelength λ, which is the integration variable, while x is the value of z at which the template is redshifted. So, this equation can be written as:

$$c(z) \equiv [s \star t](z) = \int_{-\infty}^{+\infty} s(\lambda)\, t(\lambda \cdot (1 + z)) d\lambda \tag{7.2}$$

or, by discretising it:

$$c(z_k) = \sum_{j \in \Lambda} s_j\, t_j(z_k) \tag{7.3}$$

where we have assumed that the cross-correlation is evaluated at several redshift z_k, so the result is denoted as $c(z_k)$; s_j and $t_j(z_k)$ are the spectrum and template (redshifted to z_k) at the wavelength "pixel" j, which varies in Λ, representing the integration domain, i.e. the MUSE wavelength range (4650Å– 9300Å). Another criterion to express similarity between two patterns is the Pearson product-moment correlation coefficient (Pearson 1895), based on which Eq. 7.3 can be written as:

$$c(z_k) = \frac{\sum_{j \in \Lambda} [s_j - \langle s \rangle][t_j(z_k) - \langle t(z_k) \rangle]}{\sqrt{\sum_{j \in \Lambda} [s_j - \langle s \rangle]^2 \sum_{j \in \Lambda} [t_j(z_k) - \langle t(z_k) \rangle]^2}} \tag{7.4}$$

where we have introduced the spectrum and template mean, $\langle s \rangle$ and $\langle t(z_k) \rangle$, and a normalisation factor, used to measure the similarity between unscaled signals. Such expression is also known as normalised cross-correlation. Finally, we can include in Eq. 7.4 the weights. In the general case, they act on both spectrum and template; in this work, weights are computed from noise in data and act only on the spectra. Thus, the normalised weighted cross-correlation is defined as:

$$c(z_k) = \frac{\sum_{j \in \Lambda} w_j [s_j - \langle s \rangle][t_j(z_k) - \langle t(z_k) \rangle]}{\sqrt{\sum_{j \in \Lambda} w_j} \sqrt{\sum_{j \in \Lambda} w_j [s_j - \langle s \rangle]^2 \sum_{j \in \Lambda} [t_j(z_k) - \langle t(z_k) \rangle]^2}} \tag{7.5}$$

where the spectra mean $\langle s \rangle$ is computed weighted:

$$\langle s \rangle = \frac{\sum_{j \in \Lambda} w_j s_j}{\sum_{j \in \Lambda} w_j} \tag{7.6}$$

This measure can be expressed in a matrix form. Let's assume that the wavelength range is composed by M elements, i.e. $j = 1, \ldots, M$, and the redshift range is sampled with an equal-step grid of element z_k, i.e. $k = 1, \ldots, K$, thus, the template can be arranged in a matrix T_{jk} with shape (M, K) whose elements are $T_{jk} = t_j(z_k) - \langle t(z_k) \rangle$. Correspondingly, the N spectra (i.e. one for each spatial pixel) can be organised as a matrix S_{ij} with $i = 1, \ldots, N$ and $j = 1, \ldots, M$, whose shape is (N, M) and whose elements are $S_{ij} = s_{ij} - \langle s_i \rangle$. In the same way the weights can be represented as matrix W_{ij} with shape (N, M) if each spectrum has its own weight vector, or as a row vector W_j with shape $(1, M)$ if all spectra share the same weights. With these assumptions Eq. 7.5 can be written as:

$$C_{ik} = \frac{\sum_{j=1}^{M} W_{ij} S_{ij} T_{jk}}{\sqrt{\sum_{j=1}^{M} W_{ij}} \sqrt{\sum_{j=1}^{M} W_{ij} S_{ij}^2} \sqrt{\sum_{j=1}^{M} T_{jk}^2}} \tag{7.7}$$

Thus, the results of the cross-correlation is represented with a matrix C_{ik} with $i = 1, \ldots, N$ and $k = 1, \ldots, K$, i.e. a cross-correlation as a function of z, for each pixel. In the general case, the three elements in the denominator are three vectors: $\sqrt{\sum_{j=1}^{M} W_{ij}} = \overline{W}_i$ (with shape $(N, 1)$), $\sqrt{\sum_{j=1}^{M} W_{ij} S_{ij}^2} = \overline{S}_i$ (with shape $(N, 1)$) and $\sqrt{\sum_{j=1}^{M} T_{jk}^2} = \overline{T}_k$ (with shape $(1, K)$), so Eq. 7.7 can be expressed in a matrix form:

$$C = (W \odot S) \cdot T \oslash [(\overline{W} \odot \overline{S}) \cdot \overline{T}] \tag{7.8}$$

where $\odot$ and $\oslash$ denote the element-wise product and division.

Since we are interested in finding and saving an arbitrary number of solutions, instead of just the maximum of the XC-function, we implement a fast peak-detection algorithm based on a *max* pooling sub-sampling. Given N cross-correlation function $C_i(z)$ and the number of peaks to save n_z, the algorithm performs a sub-sampling with a fixed window of size $\Delta z = 0.002$, the resulting ensemble of vectors, Z_i with $i = 1, \ldots, N$, is compared with C_i: indices n corresponding to values for which $Z_i(n) = C_i(n)$ are the desired maxima. To sort these indices, the method computes the corresponding peaks and returns the first n_z. The algorithm can be summarised as follow:

$$\begin{aligned} Z_i = &\leftarrow \text{max\_pool}(C_i, \Delta z) \\ \text{maxima}_i &\leftarrow C_i == Z_i \\ \text{maxima}_i &\leftarrow \text{cast}(\text{maxima}_i, \text{float}) \\ \text{peaks}_i &\leftarrow C_i \odot \text{maxima}_i \\ \text{maxima}_i &\leftarrow \text{argsort}(\text{peaks}_i, n_z) \end{aligned} \tag{7.9}$$

where the index i labels the i-est spatial pixel, *cast* is the type conversion operation from integer to floating values, and *argsort* indicates the process used to sort peaks (in descending order), which currently selects only the first n_z peak. Although the previous algorithmic steps are described separately for each i, operations are computed with matrices, optimising the GPU computing (computational costs are listed in Tab. 7.1).

The complete procedure with which the tool processes the data cube is summarised as follow:

0. Input: MUSE data cube (with shape (M, H, W), i.e. wavelength range, height and width), reference template, redshift range ($\Delta z = [z_{min}, z_{max}]$), redshift sampling ($dz$), weights (it can be a single "spectrum" shared for all pixels, or it can be computed from the variance map as $W = 1/\sigma^2$);

N spectra	dz	Δz	z size (K)	time [sec]
(484,522) = 252648	10^{-3}	(0.000, 3.000)	$3 \cdot 10^3$	20
(484,522) = 252648	10^{-4}	(0.000, 3.000)	$3 \cdot 10^4$	85
(484,522) = 252648	10^{-5}	(3.200, 3.300)	10^4	45
(484,522) = 252648	$5 \cdot 10^{-6}$	(6.130, 6.168)	$7.6 \cdot 10^3$	30

Table 7.1: Cross-correlation computing times by varying the z sampling. Times refer to the amount of seconds required by the tool to compute the XC-functions, i.e. to solve Eq. 7.8 for all the involved pixels. The computing has been performed on a NVIDIA GPU Titan Xp.

1. Applying a "running mean" on the whole cube, independently of each wavelength step, which consists in a spatial smoothing, performed as an *average* pooling with a window size of 0.6″ (i.e. 3 pixels);

2. defining the wavelength grid $\Lambda = \{\lambda\}_j$ with $j = 1, \ldots, M$, deduced from the cube header, and a redshift grid z_k with $k = 1, \ldots, K$ such that $K \cdot dz = \Delta z$;

3. optionally, the cube is continuum subtracted by fitting a 5th degree Chebyshev polynomial;

4. cube flattening: the cube is reshaped into a matrix with shape $(N = H \cdot W, M)$, so the i-est row corresponds to the spatial pixel $h = \text{floor}(i/W)$, $w = \text{mod}(i, h)$;

5. computing the template matrix T_{jk}: the k-est column is the redshifted template at $z_k = z_{min} + k \cdot dz$, estimated by interpolating fluxes on a shifted wavelength grid $\lambda(z_k) = \lambda' \cdot (1 + z_k)$, where λ' is the original wavelength grid on which the template is defined; fluxes are then evaluated on the defined wavelength grid $\Lambda = \{\lambda\}_{j=1}^{M}$, i.e. the redshifted template has the same spectral resolution as the data cube;

6. cross-correlation is performed according to Eq. 7.8, by extracting a batch of spectra $B < N$, to avoid memory loss. The resulting redshfit map is built by taking the first peak of the XC-function, evaluated with the peak detection algorithm (Eqs. 7.9). Optionally, n_z solutions are saved, corresponding to the first n_z peaks. Peaks at the edge of the z range are discarded;

7. Optionally: if a catalogue of sources (composed by position and an aperture) is provided, the XC-tool estimates the redshift for each source by computing an average of the spectra within the given aperture and by extracting the first n_z peaks, producing a catalogue with n_z solutions for each source.

The tool has been optimised to be processed on GPU. Respect to the same version, always implemented in `python` exploiting `multiprocessing` library (McKerns et al. 2012) to handle the CPU-parallelisation, which required $\sim$ 50 minutes, our tool completes the computation in just $\sim$ 20 seconds (using the same hyper-parameters), which means a boost of 150.

7.3 Cross-correlation applications

In this work, we use the implemented XC-tool to estimate redshift of galaxies in the MUSE FoV. However, before carrying out these measurements we validate the tool capacity using a simulated galaxy (Sec. 7.3.1) and comparing with Grillo et al. (2016) the velocity maps estimated for 'Refsdal' host, a lensed galaxies at $z = 1.4888$ (Sec. 7.3.2).

Concerning the templates used in this work, we cross-correlate MUSE data cube with 3 different templates (listed in Fig. 7.1):

- Early-type galaxy (ETG) template: it has been generated by stacking spectra of 63 cluster galaxies belonging to 7 clusters (R2129, R2248, M0416, M1206, M1311, M1149 and M2129) spanning a redshift range $(0.23, 0.59)$, with 2 magnitude fainter than the BCG, $F814 > F814_{BCG} + 2$, and with normalised colour $(F606 - F814)_{\rm norm} > -0.2$mag. The spectra stacking is weighted according to their S/N ratio (as described in Gobat et al. 2008; Nantais et al. 2013);

- Irregular template (Im): it is taken from VANDELS template ensemble (Le Fèvre et al. 2013);

- Emitter template: it has been generated by stacking spectra of line emitter galaxies (used to in the lens model of Bergamini et al. 2021b), spanning a redshift range $(1.7, 3.9)$, whose main features have been studied and outlined by Vanzella et al. (2021). Each main line has been independently staked and fitted with a Gaussian curve (with a bimdal Gaussian curve for double lines or couple of emission and absorption lines, or with an asymmetric Gaussian for the Lyman-α).

We use only these three templates, even if they are not representative of all galaxy types, because we are mostly interested in searching for cluster members and emitter galaxies. Clearly using a larger set of templates should improve the tool performance, however one needs to deal with the difficult task of separating degenerate solutions in an automated fashion. For this reason, a small template set was used in this preliminary development phase.

7.3.1 Performance with simulation

As first step, we are interested in evaluating the tool performance by using a simulated galaxy dominated by [OII]λ3726.2, 3729.1 emission doublet, like system-12 (as named and studied by Bergamini et al. 2021b) and 'Refsdal' host (studied by Grillo et al. 2016).

In order to simulate this galaxy, we start by cross-correlating two sub-cubes, extracted on the two images[1] of system-12 (RA = 4:16:080.75, DEC = − 24:04:02.18; RA = 4:16:09.91, DEC = − 24:04:16.81), with a template simply constituted by a bimodal Gauassian with peaks centred on (3726.2, 3729.1) Å and a FWHM of 1.5Å. To determine a work area, we select pixels with velocity separation within $\pm 100\,\mathrm{km\,s^{-1}}$ and split them into 6 velocity bins (see central panels of Fig. 7.2). We extract a spectrum in each bin by averaging all the belonging pixels; spectra are shifted at rest-frame, stacked by averaging and used to fit a bimodal Gaussian (see right panel in Fig. 7.2). As results, we obtain an empirical [OII] template.

Such empirical template is used to modulate a simulated galaxy spectrum. By assuming a PIEMD mass density profile (see Eq. 2.10 described in Sec. 2.1.4), galaxy rotational velocity can be expressed as:

$$v(R) = v_0 \left[1 - \frac{\arctan\left(R/R_c\right)}{R/R_c} \right]^{1/2} \tag{7.10}$$

where R_c is a core radius, and v_0 is the circular velocity, imposed to be $100\,\mathrm{km\,s^{-1}}$. These velocities are computed by assuming a galaxy size $R = 10 \cdot R_c$ (i.e. $R/R_c \leq 10$), then they are projected along the line of sight $v(R)\cos\theta$, where θ is angle between the 3D galaxy radius and the projected radius. So, a spectrum is assigned to each R by shifting the [OII] empirical template according to the radial velocity $v(R)\cos\theta$ and redshifted at $z_0 = 1.4888$ (i.e. Refsdal redshift).

[1] Actually, these are three multiple images, two of them are merging speculatively at the turn of the critical line, resulting as single image, top panels in Fig. 7.2 (see Bergamini et al. 2021b).

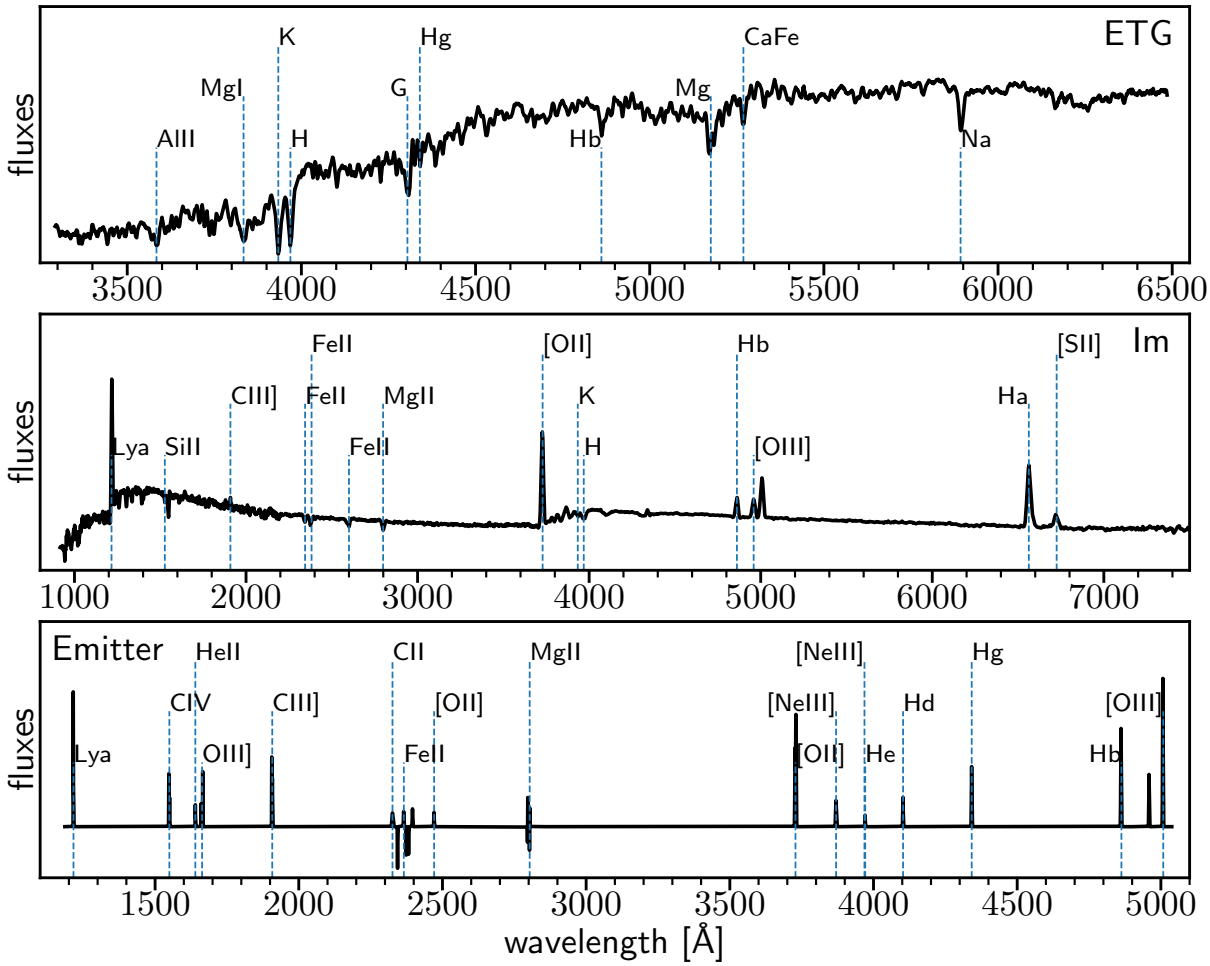


Figure 7.1: Reference templates used for the cross-correlation with the main absorption or emission lines. The early-type galaxy template (ETG, top panel) has been generated by stacking spectra of cluster galaxies spanning a redshift range (0.23 – 0.59). The irregular galaxy template (Im, middle panel) has been taken from VANDELS template ensemble (Le Fèvre et al. 2013). The template for emitters has been obtained by stacking spectra of galaxies with redshift ∈ (1.6 – 6.1) from Bergamini et al. (2021b), guided by the analysis carried out by Vanzella et al. (2021).

The cube, describing the simulated galaxy, is computed as:

$$\text{cube}(R,\theta) = \text{template}_{[\text{OII}]}\left(z = z_0 + \frac{v(R)\cos\theta}{c}(1+z_0)\right) \tag{7.11}$$

We finally add a Gaussian noise estimated in 50 circular apertures (0.8″ diameter) in the surrounding of Refsdal, by preserving the dependence on wavelengths (i.e. mean and standard deviation are estimated for wavelength step). A comparison between the simulated galaxy and Refsdal spectra is depicted in Fig. 7.3, in terms of [OII]λ3726.2, 3729.1 double lines, which shows the S/N similarity between real spectra (randomly extracted from Refsdal host) and simulated spectra.

The simulated cube is cross-correlated with the empirical [OII] template, and the resulting redshift map is converted into a (rest-frame) velocity map, which is compared with the velocity map obtained from the simulated cube. A summary of this comparison is displayed in Fig. 7.4, in terms of: difference between velocity maps (top panels), pixel-by-pixel difference distribution (bottom left panel) and velocity profiles as a function of distance from the center, i.e. velocities averaged within circular annuli (bottom right panel). These figures show that differences are

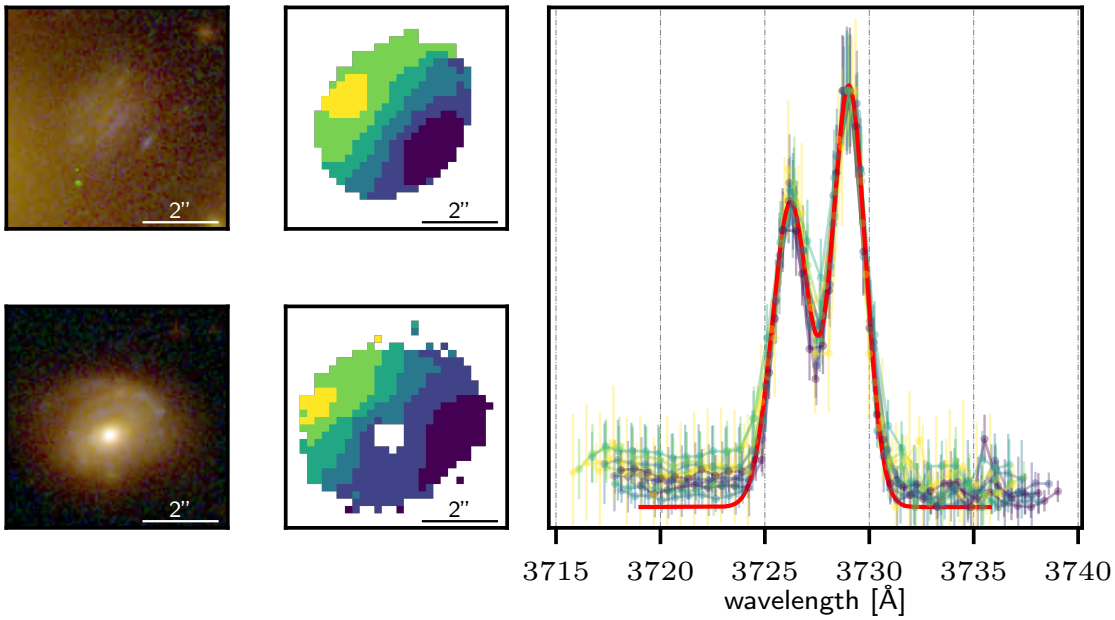


Figure 7.2: In the central panels we show the velocity maps of two multiply lensed galaxies in M0416 (system-12) (left column) obtained via our cross-correlation tool. On the right panel, we plot the [OII]λ3726.2, 3729.1 doublet used as a cross-correlation template. This was derived by stacking and shifting in rest-frame the MUSE spectra in six velocity bins of both galaxies, and then by fitting a bimodal Gaussian (in red).

small, with a $\sigma \sim 3\,\mathrm{km\,s^{-1}}$, and not-biased ($\mu = -0.1 \pm 0.2\,\mathrm{km\,s^{-1}}$), while larger discrepancies characterise the inner region with $\Delta v \in (5, 10)\,\mathrm{km\,s^{-1}}$ for $R \leq 0.5''$, although the discrepancies are constrained in 1.3σ.

7.3.2 Velocity estimation comparison

As a second step, we perform a comparison between the velocities estimated in 'Refsdal' host for 62 knots, belonging to 18 families, identified by Treu et al. (2016) and used by Grillo et al. (2016), together with other multiple images, to constrain the lens model of M1149. The redshift map is computed by cross-correlating two sub-cubes (whose centres are: RA, DEC = (177.40343, 22.40239), and RA, DEC = 177.39916, 22.39606, with sides: 10″ and 16.8″, respectively for the North-East and the South-West image) with the empirical [OII] template, using the variance map associated to the MUSE data cube as weights. The velocity maps are shown in top panels of Fig. 7.5, together with the 64 knots, that can be compared with those estimated by Grillo et al. (2016), showed in the middle panels of Fig. 7.5. Note that we are not using a Boolean mask to select the galaxy contour, instead it is automatically determined by selecting pixels with velocity separation in $\pm 100\,\mathrm{km\,s^{-1}}$, which results in some pixels that scattered out of the galaxy image.

For each family, we estimate:

- the rest-frame velocity difference between all cross-compared couple of knots belonging to same family:

$$\Delta v_n = v_i - v_j \quad \forall i \neq j = 1, \ldots, N_k \tag{7.12}$$

  where N_k is number of knots belonging the k-est family and n labels the entries. The resulting distribution is displayed in left bottom panel of Fig. 7.5;

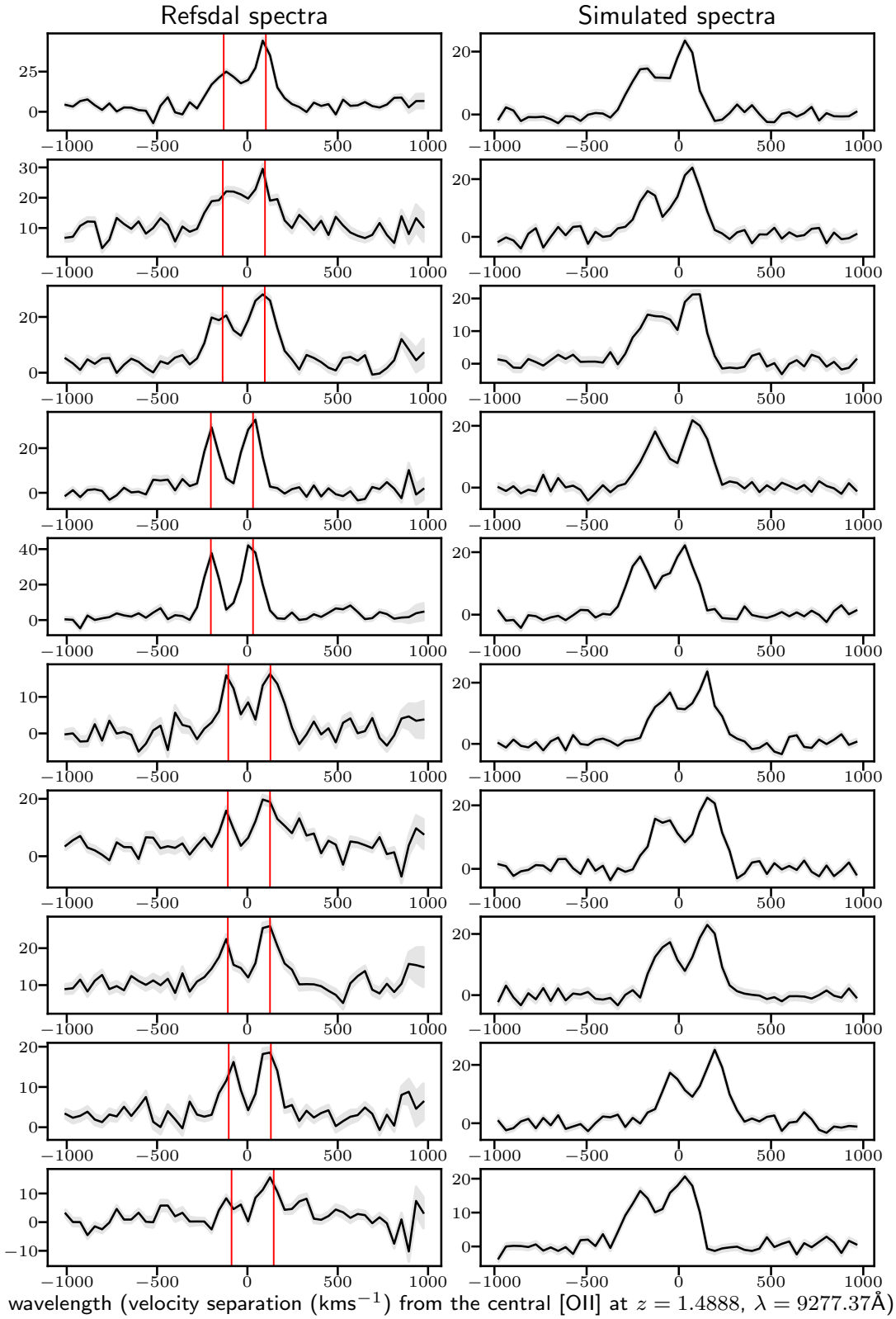


Figure 7.3: Comparison between Refsdal (left panel) and simulated galaxy (right panels) in term of [OII] doublet. Wavelengths are expressed as velocity separation from the center of [OII] redshifted at $z = 1.4888$. [OII]λ3726.2, 3729.1 lines are overlapped (in red) on the real spectra. The grey area represents 1σ signal fluctuation. Spectra on the same row are not related.

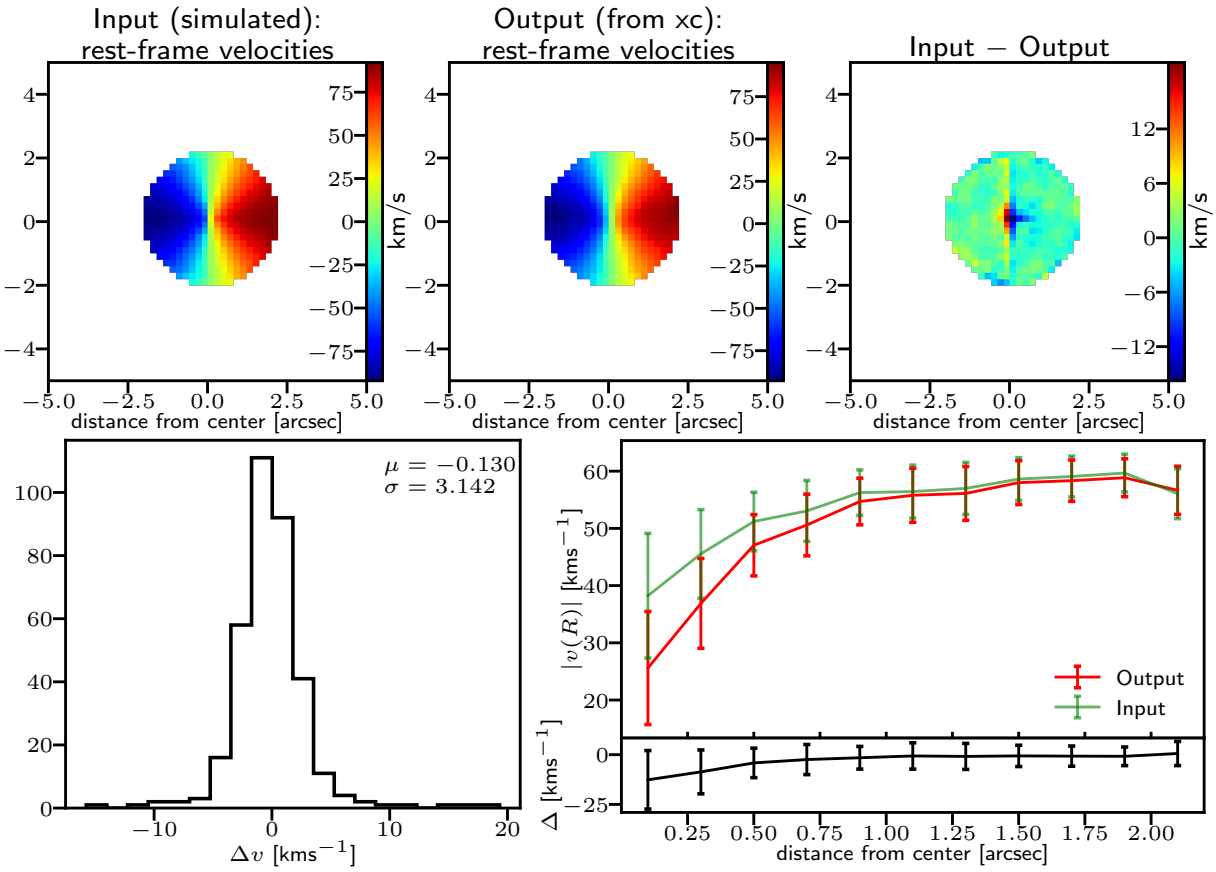


Figure 7.4: *Top panel*: comparison of simulated velocity maps ("input") with those obtained with the cross-correlation tool ("output"). The rightmost panel shows the difference between these two. *Bottom panel*: the pixel-by-pixel velocity difference is shown to the left, while the velocity profiles as a function of the distance from the center are plotted in the right panel, for both input (green) and output (red), together with their difference (bottom axes).

- the standard deviation of rest-frame velocities computed for each of the 18 families:

$$\Delta v_k = RMS\left(\{v_i\}_{i=1}^{N_k}\right) \quad k = 1, \ldots, N_F \tag{7.13}$$

where N_F in the number of families, and, again, N_k is the number of knots belonging the k-est family. The resulting distribution is plotted in right panel of Fig. 7.5 (and can be directly compared with Fig. 9 in Grillo et al. 2016).

The cross-compared velocity distribution (left bottom panel in Fig. 7.5) shows that the cross-correlation is actually able to measure velocities, at the pixel level, with a good degree of accuracy without biases, since the distribution is centred around zero, with a $\sigma = 7.6\,\mathrm{km\,s^{-1}}$ and with maximum variation constrained in $\pm 20\,\mathrm{km\,s^{-1}}$. The RMS distribution for each family (right bottom panel in Fig. 7.5) confirms this level of accuracy, since all the standard deviations are with $10\,\mathrm{km\,s^{-1}}$, with two peaks, the first around $2\,\mathrm{km\,s^{-1}}$, the second around $\sim 4\,\mathrm{km\,s^{-1}}$. By comparing this distribution with the same provided by Grillo et al. (2016) (see their Fig. 9), we find an improvement of a factor ~ 3, since their estimated velocity dispersions increase up to $30\,\mathrm{km\,s^{-1}}$, with a mean of $10.5\,\mathrm{km\,s^{-1}}$.

As final example, to underline the XC-tool capabilities to estimate velocity map for large FoVs in few seconds, we measure the velocity map of the snake arc and its counterpart image, in M1206 cluster. As done before, we extract two sub-cubes (centred in RA, DEC = (181.54491, −8.80130) and in RA, DEC = 181.54656, −8.79553, respectively with shape of $21'' \times 7''$) and $10'' \times 10''$),

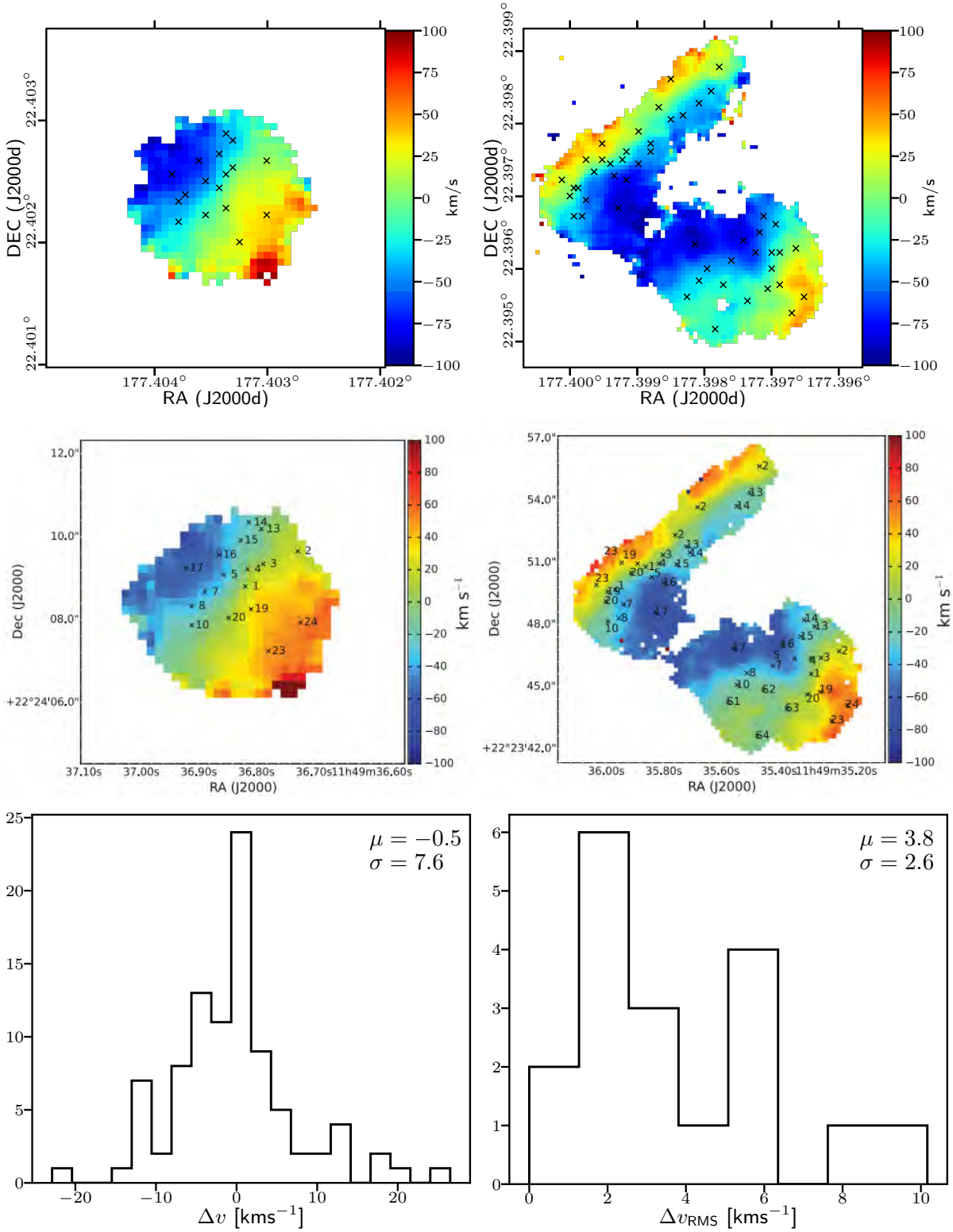


Figure 7.5: Rest-frame velocities of Refsdal host, together with the 62 knots (marked as crosses), are plotted on top panels, fully comparable with velocity maps estimated by Grillo et al. (2016), plotted in the middle panels. Bottom left panel shows the distribution of the velocity difference between all cross-compared couple of knots belonging to same family, while the distribution of the velocity RMS, computed for each family, is displayed in the bottom right panel (to compare with Fig. 9 in Grillo et al. 2016).

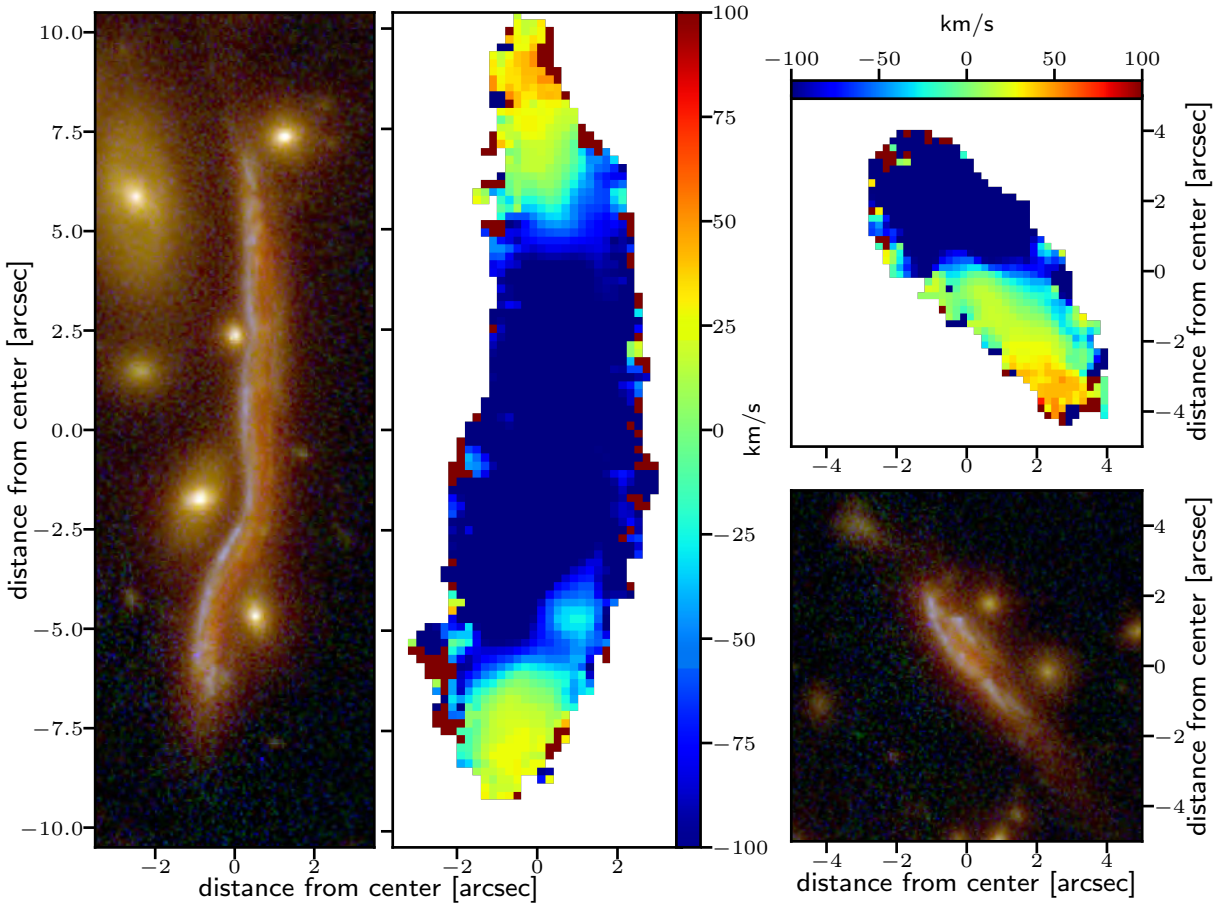


Figure 7.6: Rest-frame velocity maps for snake arc and its counterpart image in M1206, together with the corresponding HST cutouts. The snake arc velocity map and its corresponding HST imaging are shown in middle and left panels (image center is RA, DEC = (181.54491, −8.80130). The counterpart image and the estimated velocity map are shown in bottom and top right panels (image center is RA, DEC = 181.54656, −8.79553).

which were cross-correlated with the [OII] empirical template, using the MUSE variance map as weight. The estimated velocity maps, together with the corresponding HST imaging, are shown in Fig. 7.6. We measure a median redshift of $z = 1.03680$, which can be compared with the previous measurement $z = 1.0369$ (Caminha et al. 2017b).

7.3.3 Redshift measurement

In the previous sections, cross-correlations have been performed with a relatively narrow redshift range, typically constrained in some tens of thousands of km s^{-1} from the known redshift target. Here we evaluate the capabilities to estimate redshift of sources spanning a larger interval $0 - 7$. We restrict the analysis to the North-East MUSE pointing of M0416 galaxy cluster, called as MUSE Deep Lensed Field (MDLF). We exploited the spectroscopic catalogue (which have been measured with EZ software developed by Garilli et al. (2010)), whose intersection with the MDLF consists in 274 sources (by also removing stars); these measurements are assumed as ground truth.

The cross-correlation is performed between the whole data cube and three reference templates: early-type galaxy (ETG) template, irregular (Im) template and emitter template (shown in Fig. 7.1). Even if this ensemble has been set to represent the involved 274 galaxies, we note that the usage of other templates could help to map lines, however due to the complex handling

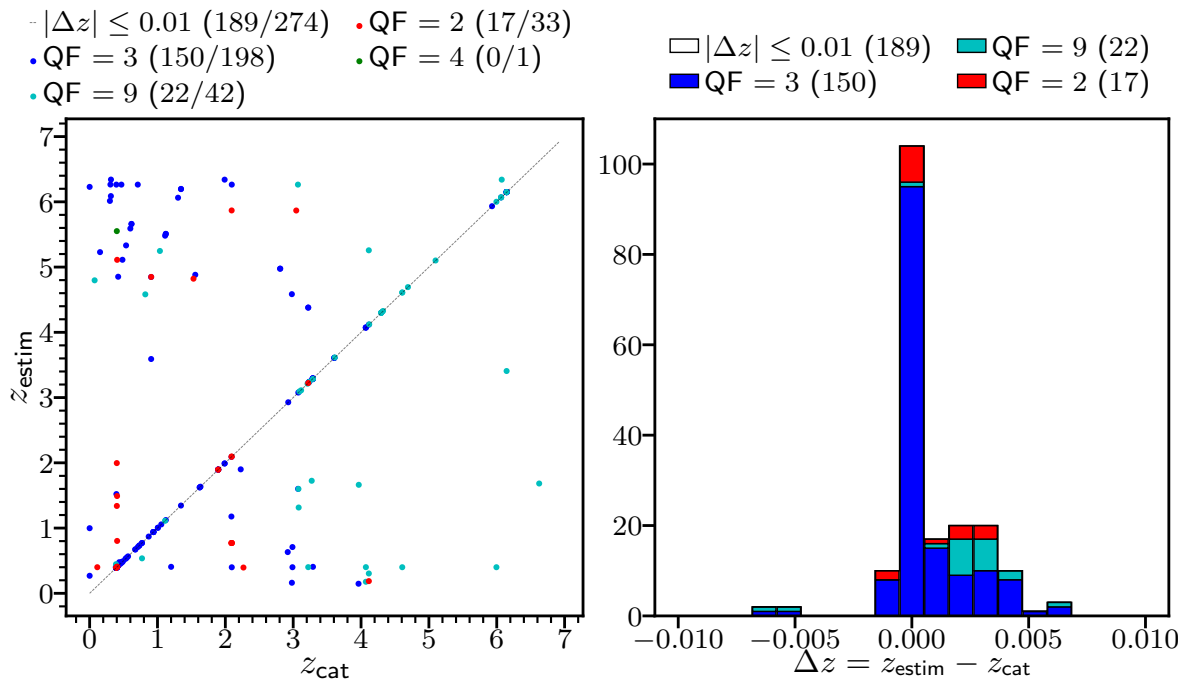


Figure 7.7: Redshift estimated with the cross-correlation. *Left panel*: catalogue redshift (z_{cat}, assumed as ground truth) vs. estimated redshift (z_{estim}, automatically with the XC-tool); diagonal represents points with $z_{cat} = z_{estim}$. *Right panel*: $\Delta z = z_{estim} - z_{cat}$ distribution for the 189 source (70%) with $|\Delta z| \leq 0.01$. In both panels results are organised into 4 classes according to the catalogue quality flags: *likely* (QF = 2, red), *secure* (QF = 3, blue), *single-line* (QF = 9, cyan), *from literature* (QF = 4, green, D. Kelson, priv. comm.). Numbers for classes are quoted in legends. In the left panel, numbers between brackets represent the fraction of sources with $z_{estim} \geq 0.01$.

of multiple solution combinations, we avoided to introduce other templates. To weight the cross-correlation we use the noise as a function of wavelength provided by Vanzella et al. (2021), estimated by averaging fluxes in 600 non-overlapped apertures (of 0.8″ diameter) extracted on positions not intercepting evident sources (see Fig. 2 in Vanzella et al. 2021). For each template we vary the redshift range:

- ETG template: $\Delta z = [0.1, 1.2]$;
- Im template: $\Delta z = [0.1, 2.5]$;
- Emitter template: $\Delta z = [0.4, 6.7]$.

The redshift ranges have been imposed based on the templates wavelength limits. For all the three cross-correlation, we set a sampling in z of 10^{-4}. Finally, after having cross-correlated the MDLF data cube with each template, we extract a cross-correlation function for each spectroscopic source (by averaging spectra within a fixed aperture of 0.8″ diameter) and measure the tool capabilities to automatically estimate redshifts. Since we involve three templates, we assume as solution the redshift corresponding to the higher XC-peak.

Fig. 7.7 outlines the results in terms of z_{cat} vs. z_{estim}, i.e. ground truth vs. estimated (left panel), and $\Delta z = z_{estim} - z_{cat}$ distribution for the 189 source (70%) with $|\Delta z| \geq 0.01$ (right panel). In this figure, results are organised into 4 classes according to the catalogue quality flags: *likely* (QF = 2, red), *secure* (QF = 3, blue), *single-line* (QF = 9, cyan), *from literature* (QF = 4, green, D. Kelson, priv. comm.). Globally, the tool was able to recover the redshift for 189 sources, these measurements differ from $z_{cat} \leq 0.01$, of which: 150 (79%) have QF = 3, i.e. multiple lines have been detected; 17 (9%) have QF = 2, i.e. 80% reliability, 22 (12%) have QF = 9,

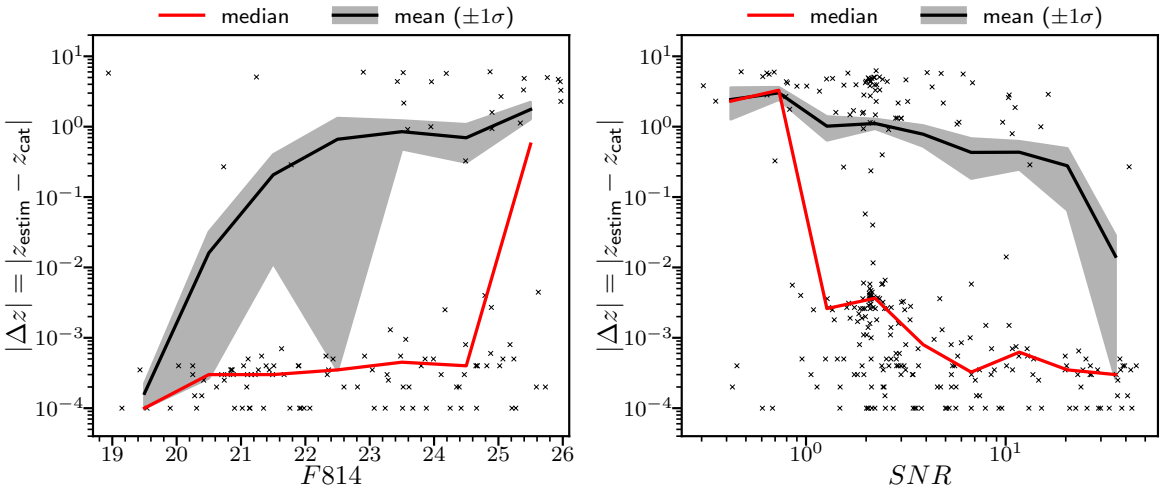


Figure 7.8: Estimated redshifts with the cross-correlation, represented in terms of $|\Delta z| = |z_{\rm estim} - z_{\rm cat}|$ as a function of $F814$ magnitude (left panel) and Signal-to-Noise Ratio (SNR, right panel). The mean, median and 1σ deviation of these distributions are represented as black, red lines and grey areas. In the left panel, only sources with available magnitude are included (119/274 sources).

i.e. just a line has been detected. Thus, the tool was able to recover 76% of sources with QF = 3, 52% of sources with QF = 2 and 52% of sources with QF = 9. By inspecting these 189 estimations, we found that for 122 (65%) sources $|\Delta z| = |z_{\rm estim} - z_{\rm cat}| \leq 0.001$, while 98 (52%) sources are measured with $|\Delta z| \leq 0.0005$. Moreover, in Fig. 7.8 we plot the $|\Delta z|$ dependence on the $F814$ magnitude (left panel, only for source with available magnitude, i.e. 119/274) and on the Signal-to-Noise Ratio (SNR, right panel), showing the mean, median and 1σ deviations. These plots show that the gap from $z_{\rm cat}$ increases for faint and low-SNR sources. Indeed, by analysing the magnitude dependence, we find that the median of $|\Delta z|$ is $5 \cdot 10^{-4}$ for $F814 \geq 23$, and $3 \cdot 10^{-4}$ for $F814 < 23$. Conversely, the $F814$ median values are 24.5 mag for $|\Delta z| \geq 0.01$, and 22.2 mag for $|\Delta z| < 0.01$. On the other hand, the median of $|\Delta z|$ is $3.6 \cdot 10^{-3}$ for $SNR \leq 2$, and $|\Delta z| = 8 \cdot 10^{-4}$ for $SNR > 2$. Alternatively, the median value of $SNR(\Delta z \geq 0.01)$ is 2.16, whereas $SNR(\Delta z < 0.01)$ is 2.90.

As example, two sources, whose redshift has been correctly measured, are shown in Fig. 7.9, where the result of cross-correlation are shown for an elliptical and a lensed emitter galaxy: their spectrum is overlapped to the main lines, which were redshifted at $z_{\rm estim}$, and the cross-correlation function is shown together with the first 10 peaks. Considering that these measurements are carried out in a completely automatic way, in few tens of seconds, the cross-correlation tool is promising.

Regarding the incorrect measurements, from the left panel in Fig. 7.7, they seem to split into two clouds: (*i*) 33 sources at lower redshift $z_{\rm cat} \lesssim 2.5$, to which an higher redshfit has been assigned $z_{\rm estim} \gtrsim 4$, representing the 44% of the incorrect measurements; (*ii*) 24 sources with $z_{\rm cat} \in (2, 4.5)$ whose $z_{\rm estim}$ have been estimated in $(0.1, 2)$, they are the 28% of the incorrect measurements.

Figs. 7.10 - 7.12 show the three examples of incorrect redshfit estimations. These figures outlines: the spectrum with the main lines redshifted to the corresponding $z_{\rm cat}$, the XC function with the first 10 peaks, the spectrum with the main lines redshifted to the corresponding $z_{\rm estim}$. In Fig. 7.10 is shown a faint source near to an elliptical: its spectrum is clearly contaminated by the elliptical halo, the cross-correlation with the emitter template is able to recover a reliable measure ($z_{\rm estim} = 3.295$ vs. $z_{\rm cat} = 3.291$), however the peak with the ETG template prevails. Even if the redshift measure is incorrect, the tool is actually working, since it recovers all redshifts in the field: the early-type galaxy (fitted with the ETG template) and the emitter galaxy (fitted with

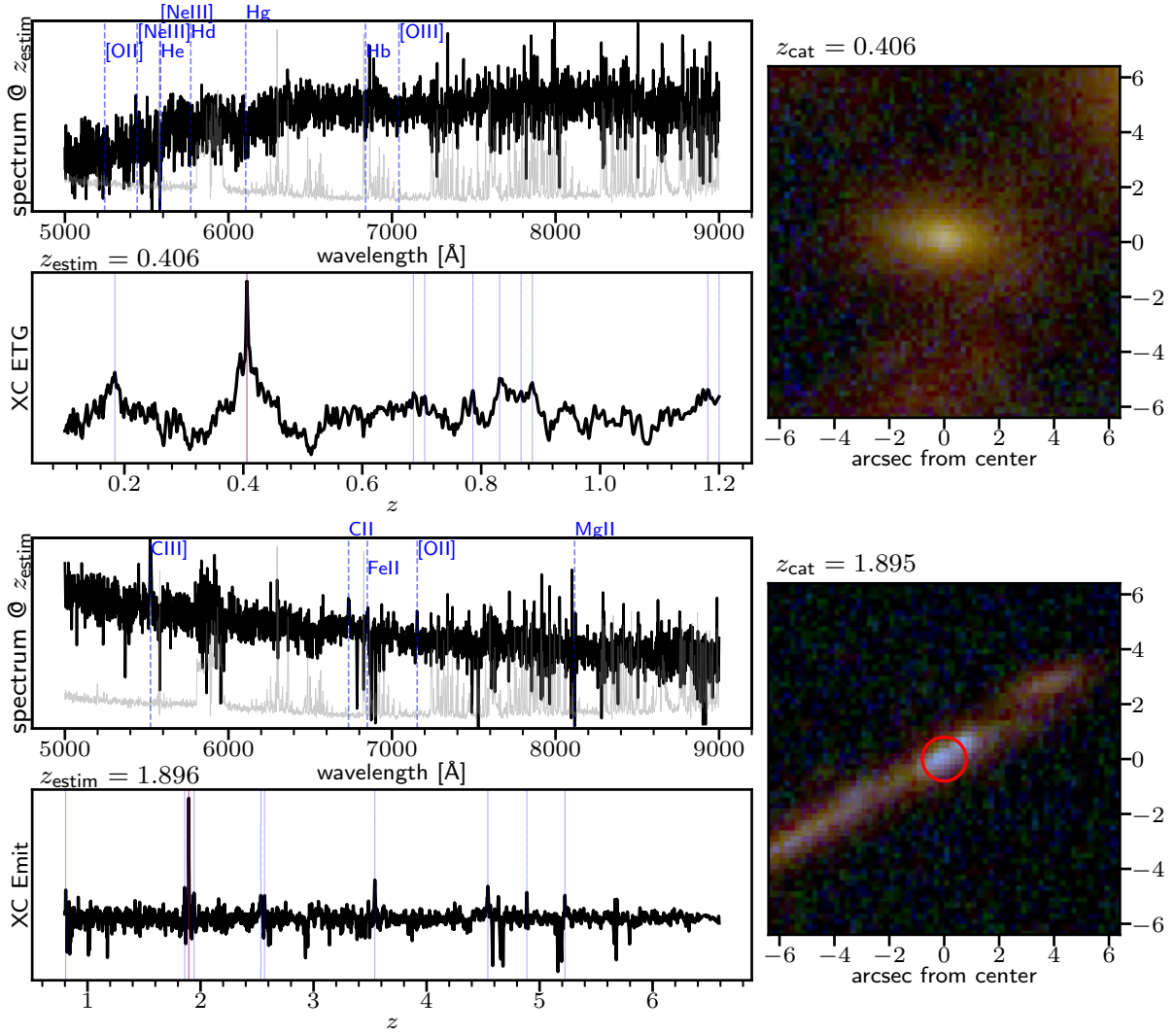


Figure 7.9: Cross-correlation results related to two correct redshift measurements: an elliptical (RA, DEC = 64.0375677, −24.0727686) and a lensed emitter galaxy (RA, DEC = 64.0411523, −24.061843). For both of them: the spectrum is plotted in top panels, overlapped to the main lines (redshifted at $z_{estim} = \{0.406, 1.896\}$), together with the noise (used a weight); the HST cutout ($\sim 13''$ across) is displayed in the right panels; the cross-correlation performed respectively with ETG and emitter template is plotted in the bottom panels (where the first 10 peaks are marked with vertical blue lines, the first of which, i.e. the solution, in red).

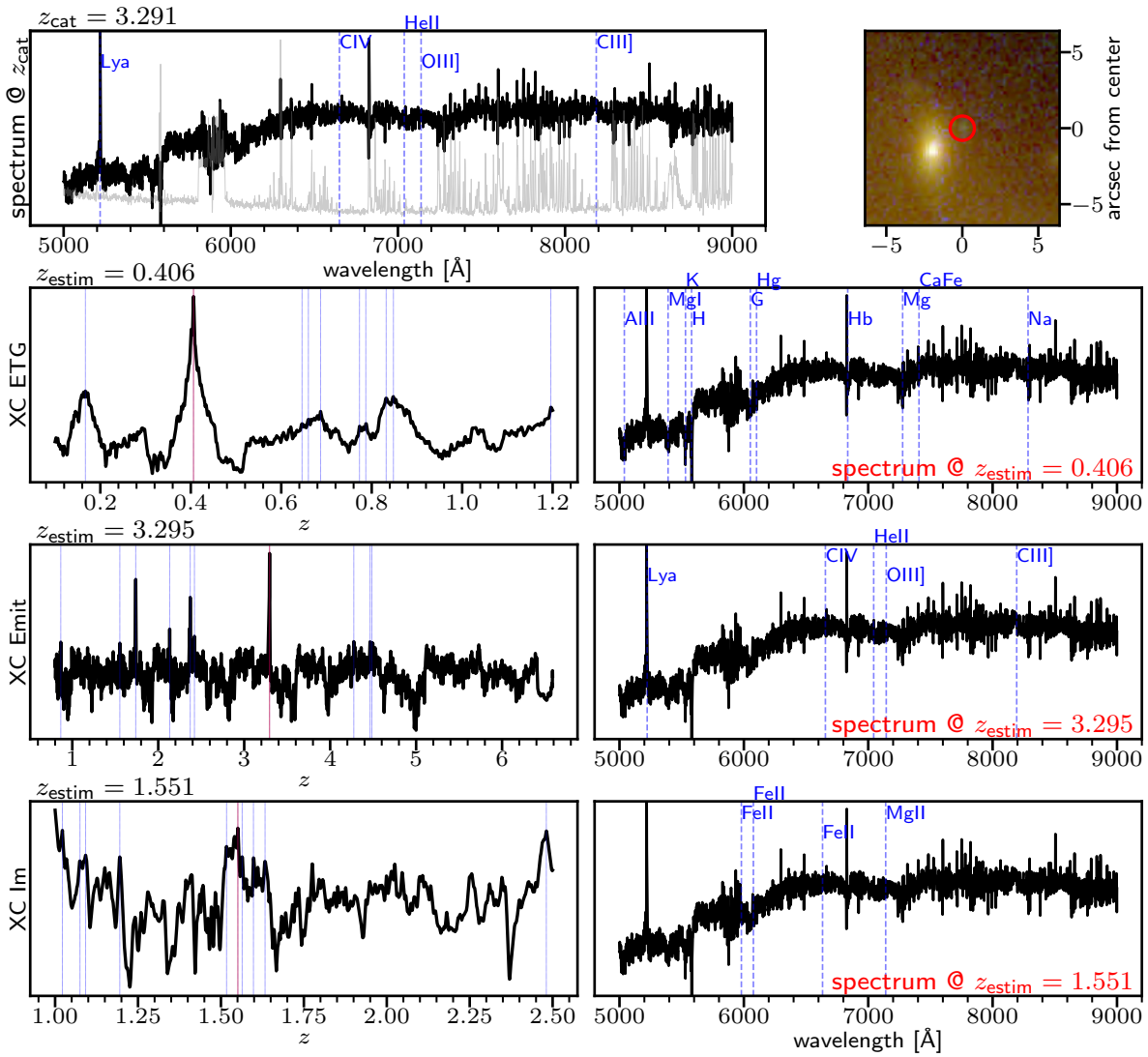


Figure 7.10: Cross-correlation results related to an incorrect redshift measurement. Source (at RA, DEC = 64.038515, −24.065965) spectrum is plotted in top panel, overlapped to the main lines redshifted at $z_{cat} = 3.291$, together with the noise (used a weight). An HST cutout (∼ 13″ across) is displayed in the top right panel, where the source is surrounded by a red circle. Second, third and fourth raw panels show the cross-correlation function (on left), with the first 10 peaks marked as vertical blue lines (the first of which in red), and the spectrum overlapped to the main lines redshifted at $z_{estim} = \{0.406, 3.295, 1.551\}$ (on right), i.e. the solutions, respectively for the cross-correlation performed with ETG, Im and Emitter template.

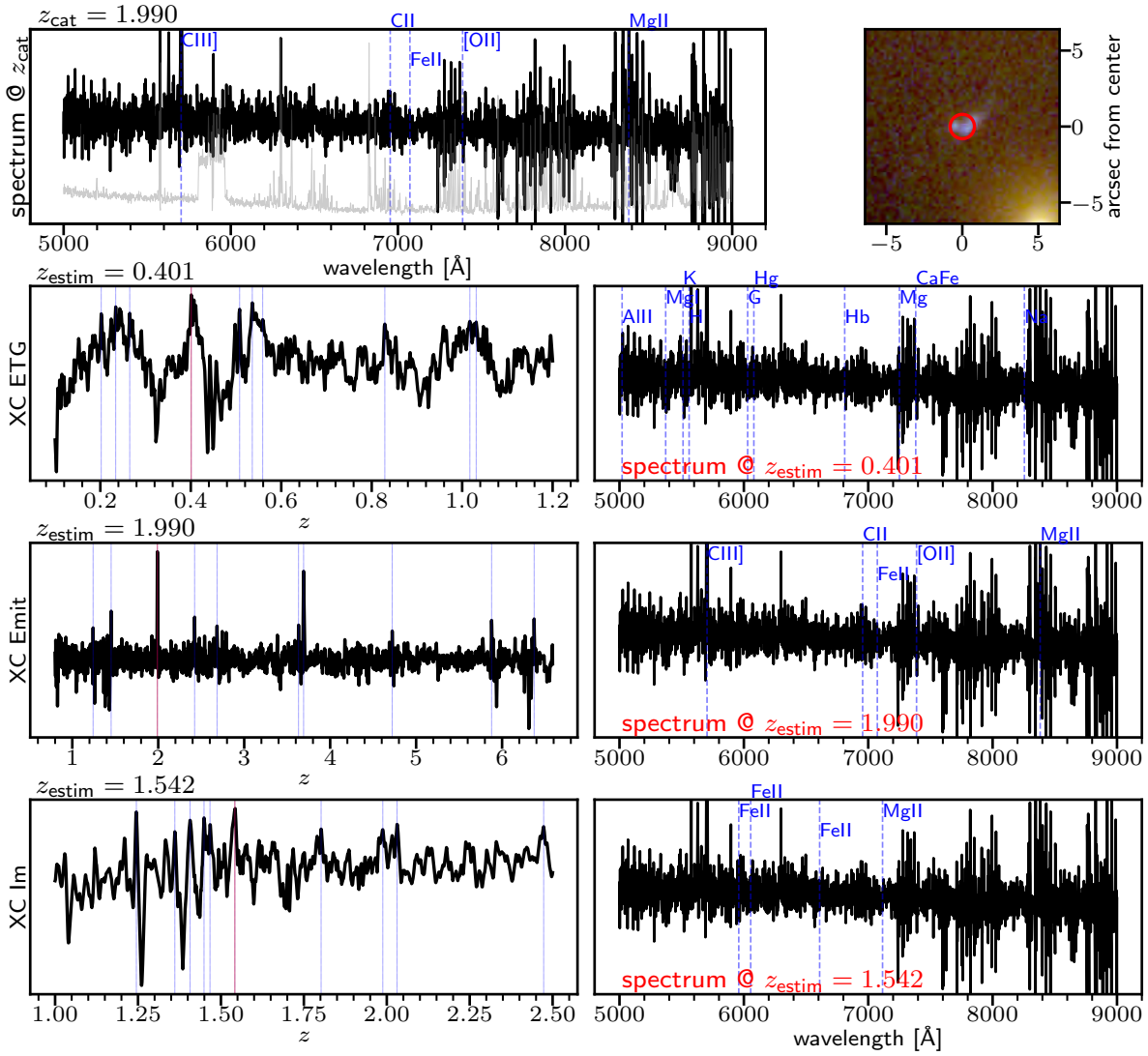


Figure 7.11: Same as Fig. 7.10, referred to source at RA, DEC = 64.035187, −24.07099.

the emitter template). A similar example is shown in Fig. 7.11, where the peak corresponding to the XC with the ETG template determines the redshift, but in contrast to the previous case, the spectrum is not contaminated despite the proximity with an early-type galaxy. Indeed, the XC function related to the ETG template (second left panel in Fig. 7.11) is characterised by several peaks of similar height. Based on a posterior visual analysis of the XC function, this solution will be undoubtedly discarded, however here we are evaluating the tool capability to automatically measure redshifts, without any human support. Finally, Fig. 7.12 displays a scenario in which neither of the first peaks is a reliable measurement, although the solution ($z = 2.806$) is associated with the second best peak of the XC-function with the emitter template. We should note, however, that the redshift in the published catalogue was determined using information from an associated multiple image source with secure redshift. We analysed all the incorrect estimations, by also exploring other metrics to evaluate the reliability of the measure (e.g. the χ^2 between spectrum and template), but we have not find a criterion with which is possible to automatically estimate redshift of all sources. Redshifts for galaxies with higher S/N are typically well estimated, but faint sources require the knowledge of experts, able to discriminate between false XC-peaks, helped with RGB images and graphic interfaces which allow to easily explore cross-correlation solutions.

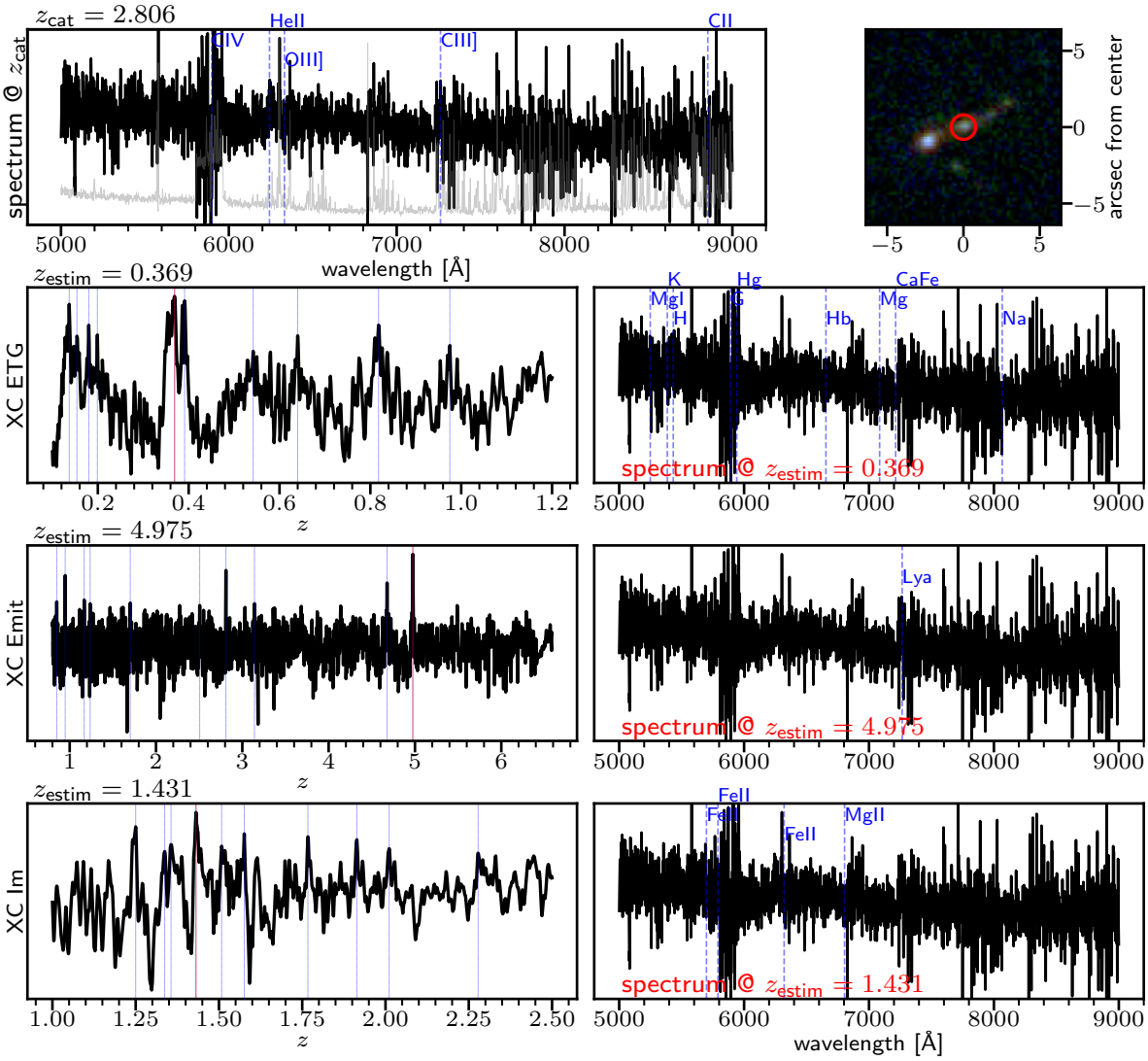


Figure 7.12: Same as Fig. 7.10, referred to source at RA, DEC = 64.0446872, −24.0576737.

7.4 Conclusions

In this chapter, we presented a GPU-optimised cross-correlation tool developed to be applied on the whole MUSE data cubes, whose primary features is the low computing cost (few tens of seconds), more than 100 times faster than executions times based on conventional multi-thread CPUs. The XC-tool can be used to measure galaxy velocity maps by processing large region of MUSE FoV or to automatically measure redshifts of all sources in the data cube.

The main results of this work can be outlined as follows:

- by processing a data cube representing a simulated galaxy (based on MUSE observation of M0416 and M1149 cluster), we find that the measured velocity map is comparable with the simulated velocity distributions, $\sigma_{\Delta v} \sim 3\,\mathrm{km\,s^{-1}}$ (Sec. 7.3.1);

- by estimating velocities for the 62 identified knots in Refsdal host, the cross-compared velocity difference distribution has a $\sigma_{\Delta v} = 7.6\,\mathrm{km\,s^{-1}}$ while the RMS evaluated for each source family has a $\sigma_{\Delta v_{RMS}} = 2.6\,\mathrm{km\,s^{-1}}$, with an improvement respect to the work carried out by Grillo et al. (2016) of a factor ~ 3 (Sec. 7.3.2);

- the automated redshift measurements, by exploring the range $z \in (0, 7)$, allows to recover

the redshift for the 70% of the involved galaxies; while the estimation for the remaining part is affected by elliptical halo contamination, faintness of the sources and low S/N ratio (Sec. 7.3.3).

Despite the difficulties to measure automatically the redshift, the exceptional computing power allows to explore the MUSE data cube in few tens of seconds, providing a fast and flexible method that can be re-processed easily by varying the hyper-parameter configuration (e.g. wavelength domain, redshift range and sampling). However, the presence of false cross-correlation peaks, the intra-cluster light which contaminates source spectra and the noise characterising a significant fraction of high-z and/or faint galaxies prevent from a completely automated measuring. So, in order to better estimate redshift in the cluster complex environment, the tool must be guided by expert astronomers. For these reasons, we are working on a web-application which combines the computing power of GPUs, the facilities provided by Graphic User Interface and the know-how of astronomers into a single functional framework. Indeed, given the complexity of the problem, the usage of graphic interfaces has become standard for an redshift measurements (e.g., see Garilli et al. 2010; Herenz et al. 2017). The impressive computing time represents a fundamental advantage of the tool, since it opens the possibility to process future large IFU datasets.

Finally, as briefly discussed during the visual inspection of the source shown in Fig. 7.10, the tool is able to measure the redshift of both the contaminant early-type galaxy and the generally blue emitter galaxy, disentangling background and foreground contributions (at least when strong emission lines are superimposed onto the absorption spectrum of an early-type galaxy). In principle, the cross-correlation tool could be suitably modified to search for emitters in the vicinity of luminous galaxies. With this method, it should also be possible to blindly search for strong-lenses, or to confirm classification performed with other techniques (e.g. CNNs) by extracting spectra in the halo of a lens candidate.

Chapter 8

Conclusions and future perspectives

In recent years, dedicated image campaigns, such as CLASH (Postman et al. 2012a), Hubble Frontier Fields (Lotz et al. 2017) and RELICS (Coe et al. 2019), have provided high quality panchromatic observations for more than 60 galaxy clusters, which have demonstrated to be a golden mine for the scientific community. Additionally, intensive spectroscopic programs, such as CLASH-VLT (Rosati et al. 2014), GLASS (Treu et al. 2015; Schmidt et al. 2014), and comprehensive VLT/MUSE observations have offered a three-dimensional view of several clusters, by providing spectra for many thousands of galaxies. These new datasets have driven a remarkable progress in our ability to exploit strong gravitational lensing to investigate the mass distribution of galaxy clusters and to unveil distant faint galaxies.

In this PhD thesis, we applied Machine Learning techniques to address three interconnected issues in gravitational lensing studies of clusters: the identification of cluster members, the search for galaxy-galaxy strong lenses around cluster galaxies and the development of a novel cross-correlation technique to extract redshifts from VLT/MUSE data cubes.

We exploited the extensive spectroscopic coverage, the quality of HST multi-band imaging and new high precision cluster lens models (provided by Bergamini et al. 2019, 2021b; Caminha et al. 2019) to build a large Knowledge Base, on which convolutional neural networks have been trained and evaluated by measuring specific performance metrics. Specifically, we explored the possibility to use deep neural networks (*i*) to recognise cluster members from background and foreground objects using a large sample of spectroscopically confirmed sources as training set, (*ii*) to identify strong-lenses in galaxy clusters by simulating realistic images of background sources using the accurate knowledge of their lensing deflection fields. Finally, we developed a GPU-optimised cross-correlation tool directly processing MUSE data cubes to efficiently measure galaxy redshifts and velocity maps. This overall methodology can be used to tackle a number of important issues when studying galaxy clusters and their galaxy populations. For example: cluster member classification allows one to analyse the cluster assembly history, as well as, the dependence of galaxy properties (e.g. structural and stellar population parameters) on the environment; the identification of GGSL candidates provides further constraints on the galaxy total mass, and on the structure of dark matter sub-halos when combined with velocity dispersion measurements. Finally, we note that the *mining* of information from Integral Fields Units (particularly MUSE) data on clusters remains significantly unexplored in data cubes, so that a 3D cross-correlation tool, such as the one developed in this work, can be the best method to discover faint line emitters, and extract redshift and kinematic information across the entire field.

In detail, the main results of this thesis can be summarised as follows:

- the image-based selection of cluster members, based on an adequate spectroscopic survey of a limited sample of clusters as a training base, can be considered a valid alternative to

photometry-based methods, working directly on multi-band imaging data in counts, thus circumventing the time-consuming process of multi-band magnitude measurements. We showed that 90% of the galaxy member population can be recovered, with an acceptable purity-completeness trade-off, down to $F814 = 25$ mag, and a superior performance when compared to photometric techniques. Misclassifications are confined in regions of the parameter space which are not well sampled by the training set (blue faint-end of the galaxy population). The first application of this method on wide-area ground based observations of galaxy clusters also showed very promising performances, albeit not as high as in HST data;

- by injecting thousands of realistic galaxy-galaxy strong lensing events in HST images utilising high-precision lens models, our deep neural networks are capable to efficiently identify galaxy-scale lenses in crowded cluster fields, with a favorable trade-off between false negatives ($\sim$ 10%) and false positives (7%). Such trained networks are then utilised to systematically search for lenses in archival HST observations of cluster cores, by detecting more than 400 GGSL candidates in 50 galaxy clusters;

- the GPU-based cross-correlation tool developed to analyse MUSE 3D spectroscopy data has shown exceptional computing performance: the entire MUSE data cube ($90 \cdot 10^3$ spectra) can be cross-correlated with a set of spectral templates in a less than a minute. Several tests on cluster fields, with redshifts measured with standard non-automated techniques, have shown a success rate of 70% in recovering galaxy redshifts. The tool needs to be further developed introducing an appropriate trade-off between automation and user interaction, specifically in the low S/N regime. In addition, this method is capable to reconstruct galaxy velocity maps (for example of lensed galaxies) fully automatically with very a high accuracy as verified with simulations and on real data of spatially extended lensed late type galaxies.

These three methodologies can be combined in an end-to-end process. Indeed, using already trained convolutional neural networks we identified cluster galaxies in archival RELICS clusters (Sec. 5.5.1) and then searched for strong-lensing features around these galaxies (Sec. 6.4.3). We plan to exploit the cross-correlation tool to spectroscopically confirm these galaxy-scale lens candidates where MUSE observations are available.

Furthermore, the generalisation capability of convolutional networks makes them both versatile and reusable tools. In fact, the convolution layers of a trained deep network can be reused as *shared* layers in larger models, such as the Faster Region CNN (Ren et al. 2015) and Masked Region CNN (He et al. 2017), which exploit kernel weights to extract multidimensional information suitable to perform object detection. Such architectures have already found interesting astrophysical applications, for example, in the identification of radio sources (Wu et al. 2019a) and the automatic deblending of astronomical sources (Burke et al. 2019). We are planning to explore this region-based architectures to automatically detect members and strong-lenses in galaxy clusters.

In future works, we will also explore other promising deep learning architectures, such as deep auto-encoders (Goodfellow 2010) and conditional generative adversarial networks (Mirza & Osindero 2014), to integrate the ground-based lower resolution images with the high quality of HST images in cluster fields. These generative techniques can be exploited to overcome the problem of missing data, thus increasing the size of the training set with a more homogeneous sampling of the entire parameter space, or to deblend the arc and ring features from the lens galaxy, and, thus, to estimate the lensed galaxy parameters (e.g. effective radius or Sérsic index).

As for the cross-correlation tool, we are working to include the tool within a wider framework, combining observations coming from different surveys, by developing a software which

incorporates the GPU computing power within a Graphic User Interface and support a larger level of interaction between the tool and astronomers. Finally, given the tool manageable computing costs, we are planning to implement a sort of "self" cross-correlation, which can be use to: detect multiple images belonging to same family in MUSE FoVs, source deblending, identification of line emitting galaxies, and improve redshift measuring by processing spectra extracting within region with irregular shapes (instead of circular), which often characterise lensed galaxies. The methodology developed in this thesis can be extended beyond the HST imaging data with a relatively modest effort and promises to have important applications with the next generation facilities. These include the analysis of deep high source density fields with the James Webb Space Telescope (Gardner et al. 2006), and more prominently the exploitation of the upcoming survey machines such as the Vera Rubin Observatory (Ivezić et al. 2019) and Euclid (Laureijs et al. 2011). The latter will cover very large sky areas (with Peta-scale data volume), discovering up to 10^5 galaxy clusters and groups and will require automated machine learning based methodologies to exploit their astrophysical and cosmological content.

Bibliography

Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, software available from tensorflow.org

Abell, G. O. 1958, ApJS, 3, 211

Abraham, S., Philip, N. S., Kembhavi, A., Wadadekar, Y. G., & Sinha, R. 2012, MNRAS, 419, 80

Acebron, A., Cibirka, N., Zitrin, A., et al. 2018, ApJ, 858, 42

Acebron, A., Grillo, C., Bergamini, P., et al. 2021, arXiv e-prints, arXiv:2111.05871

Acquaviva, V. 2020, in Panchromatic Modelling with Next Generation Facilities, ed. M. Boquien, E. Lusso, C. Gruppioni, & P. Tissera, Vol. 341, 88–98

Agüera y Arcas, B. 2017, Arts, 6

Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, PASJ, 70, S4

Anderson, J. & King, I. R. 2006, PSFs, Photometry, and Astronomy for the ACS/WFC, Instrument Science Report ACS 2006-01

Angora, G., Brescia, M., Cavuoti, S., et al. 2019, MNRAS, 490, 4080

Angora, G., Rosati, P., Brescia, M., et al. 2020, A&A, 643, A177

Aniyan, A. K. & Thorat, K. 2017, ApJS, 230, 20

Annunziatella, M., Biviano, A., Mercurio, A., et al. 2014, A&A, 571, A80

Annunziatella, M., Bonamigo, M., Grillo, C., et al. 2017, ApJ, 851, 81

Annunziatella, M., Mercurio, A., Biviano, A., et al. 2016, A&A, 585, A160

Arora, S., Bhaskara, A., Ge, R., & Ma, T. 2013, arXiv e-prints, arXiv:1310.6343

Auger, M. W., Treu, T., Gavazzi, R., et al. 2010, ApJ, 721, L163

Bacon, R., Accardo, M., Adjali, L., et al. 2012, The Messenger, 147, 4

Bacon, R., Brinchmann, J., Richard, J., et al. 2015, A&A, 575, A75

Bacon, R., Vernet, J., Borisova, E., et al. 2014, The Messenger, 157, 13

Balestra, I., Mainieri, V., Popesso, P., et al. 2010, A&A, 512, A12

Balestra, I., Mercurio, A., Sartoris, B., et al. 2016, ApJS, 224, 33

Bartelmann, M. & Schneider, P. 2000, Physics Reports, 340

Bashir, D., Montanez, G. D., Sehra, S., Sandoval Segura, P., & Lauw, J. 2020, arXiv e-prints, arXiv:2010.06076

Batista, G. E. A. P. A. & Monard, M. C. 2003, Applied Artificial Intelligence, 17, 519

Becker, S. & Lecun, Y. 1989, in Proceedings of the 1988 Connectionist Models Summer School, San Mateo, ed. D. Touretzky, G. Hinton, & T. Sejnowski (Morgan Kaufmann), 29–37

Bekki, K., Couch, W. J., Shioya, Y., & Vazdekis, A. 2005, MNRAS, 359, 949

Bell, E. F., Wolf, C., Meisenheimer, K., et al. 2004, ApJ, 608, 752

Bengio, Y. 2009, Found. Trends Mach. Learn., 2, 1

Bengio, Y. 2012, CoRR, abs/1206.5533

Bengio, Y., Courville, A., & Vincent, P. 2012, arXiv e-prints, arXiv:1206.5538

Benítez, N. 2000, ApJ, 536, 571

Bergamini, P., Agnello, A., & Caminha, G. B. 2021a, A&A, 648, A123

Bergamini, P., Rosati, P., Mercurio, A., et al. 2019, A&A, 631, A130

Bergamini, P., Rosati, P., Vanzella, E., et al. 2021b, A&A, 645, A140

Bertin, E. & Arnouts, S. 1996, Astrophysics and Space Science, 117, 393

Bessell, M. S. 1990, PASP, 102, 1181

Bhatawdekar, R., Conselice, C. J., Margalef-Bentabol, B., & Duncan, K. 2019, MNRAS, 486, 3805

Bickley, R. W., Bottrell, C., Hani, M. H., et al. 2021, MNRAS, 504, 372

Bishop, C. M. 2006, Pattern Recognition and Machine Learning (Information Science and Statistics) (Secaucus, NJ, USA: Springer-Verlag New York, Inc.)

Biviano, A., Rosati, P., Balestra, I., et al. 2013, A&A, 558, A1

Bleem, L. E., Stalder, B., de Haan, T., et al. 2015, ApJS, 216, 27

Blumenthal, G. R., Faber, S. M., Primack, J. R., & Rees, M. J. 1984, Nature, 311, 517

Blumenthal, G. R., Pagels, H., & Primack, J. R. 1982, Nature, 299, 37

Bonamigo, M., Grillo, C., Ettori, S., et al. 2018, ApJ, 864, 98

Bond, J. R., Szalay, A. S., & Turner, M. S. 1982, Phys. Rev. Lett., 48, 1636

Borgani, S. & Kravtsov, A. 2011, Advanced Science Letters, 4, 204

Borne, K., Accomazzi, A., Bloom, J., et al. 2009, in Astronomy, Vol. 2010, astro2010: The Astronomy and Astrophysics Decadal Survey

Bottou, L. & Bousquet, O. 2008, in Advances in Neural Information Processing Systems, ed. J. Platt, D. Koller, Y. Singer, & S. Roweis, Vol. 20 (Curran Associates, Inc.)

Bouwens, R. J., Bradley, L., Zitrin, A., et al. 2014, ApJ, 795, 126

Brady, M., Raghavan, R., & Slawny, J. 1989, IEEE Transactions on Circuits and Systems, 36, 665

Brainerd, T. G., Blandford, R. D., & Smail, I. 1996, ApJ, 466, 623

Breiman, L. 2001, Mach. Learn., 45, 5

Brescia, M., Cavuoti, S., Amaro, V., et al. 2018, in Data Analytics and Management in Data Intensive Domains, ed. L. Kalinichenko, Y. Manolopoulos, O. Malkov, N. Skvortsov, S. Stupnikov, & V. Sukhomlin (Cham: Springer International Publishing), 61–72

Brescia, M., Cavuoti, S., D'Abrusco, R., Longo, G., & Mercurio, A. 2013, ApJ, 772, 140

Brescia, M., Cavuoti, S., & Longo, G. 2015, MNRAS, 450, 3893

Brescia, M., Cavuoti, S., Paolillo, M., Longo, G., & Puzia, T. 2012, MNRAS, 421, 1155

Buncher, B., Sharma, A. N., & Carrasco Kind, M. 2021, MNRAS, 503, 777

Burke, C. J., Aleo, P. D., Chen, Y.-C., et al. 2019, MNRAS, 490, 3952

Caillier, P., Accardo, M., Adjali, L., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9150, Modeling, Systems Engineering, and Project Management for Astronomy VI, ed. G. Z. Angeli & P. Dierickx, 91500D

Caminha, G. B., Grillo, C., Rosati, P., et al. 2016, A&A, 587, A80

Caminha, G. B., Grillo, C., Rosati, P., et al. 2017a, A&A, 600, A90

Caminha, G. B., Grillo, C., Rosati, P., et al. 2017b, A&A, 607, A93

Caminha, G. B., Rosati, P., Grillo, C., et al. 2019, A&A, 632, A36

Campbell, M., Hoane, A., & hsiung Hsu, F. 2002, Artificial Intelligence, 134, 57

Capaccioli, M. 1989, in World of Galaxies (Le Monde des Galaxies), ed. J. Corwin, Harold G. & L. Bottinelli (New York, NY: Springer US), 208–227

Capak, P., Aussel, H., Ajiki, M., et al. 2007, ApJS, 172, 99

Cappellari, M., Scott, N., Alatalo, K., et al. 2013, MNRAS, 432, 1709

Carroll, S., Carroll, S., & Addison-Wesley. 2004, Spacetime and Geometry: An Introduction to General Relativity (Addison Wesley)

Cavanagh, M. K., Bekki, K., & Groves, B. A. 2021, MNRAS, 506, 659

Cavuoti, S., Brescia, M., De Stefano, V., & Longo, G. 2015, Experimental Astronomy, 39, 45

Cavuoti, S., Brescia, M., Longo, G., & Mercurio, A. 2012, A&A, 546, A13

Cerny, C., Sharon, K., Andrade-Santos, F., et al. 2018, ApJ, 859, 159

Chiang, Y.-K., Overzier, R., & Gebhardt, K. 2013, ApJ, 779, 127

Chinchor, N. 1992, in Proceedings of the 4th Conference on Message Understanding, MUC4 '92 (USA: Association for Computational Linguistics), 22–29

Chollet, F. et al. 2015, Keras, https://keras.io

Ciotti, L. & Bertin, G. 1999, A&A, 352, 447

Cireşan, D. C., Giusti, A., Gambardella, L. M., & Schmidhuber, J. 2013, in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013, ed. K. Mori, I. Sakuma, Y. Sato, C. Barillot, & N. Navab (Berlin, Heidelberg: Springer Berlin Heidelberg), 411–418

Clevert, D.-A., Unterthiner, T., & Hochreiter, S. 2015, arXiv e-prints, arXiv:1511.07289

Coe, D., Salmon, B., Bradač, M., et al. 2019, ApJ, 884, 85

Coe, D., Umetsu, K., Zitrin, A., et al. 2012, ApJ, 757, 22

Coe, D., Zitrin, A., Carrasco, M., et al. 2013, ApJ, 762, 32

Collobert, R. & Weston, J. 2008, in Proceedings of the 25th International Conference on Machine Learning, ICML '08 (New York, NY, USA: Association for Computing Machinery), 160–167

Congdon, A. B. & Keeton, C. 2018, Principles of Gravitational Lensing: Light Deflection as a Probe of Astrophysics and Cosmology (Springer International Publishing)

Connor, T., Donahue, M., Kelson, D. D., et al. 2017, ApJ, 848, 37

Cousins, A. W. J. 1978, Monthly Notes of the Astronomical Society of South Africa, 37, 8

Cowie, L. L. & Songaila, A. 1977, Nature, 266, 501

Cui, X., Goel, V., & Kingsbury, B. 2015, IEEE/ACM Trans. Audio, Speech & Language Processing, 23, 1469

D'Abrusco, R., Cantiello, M., Paolillo, M., et al. 2016, ApJ, 819, L31

Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, MNRAS, 460, 1270

Das, P. K. 1991, Optical Signal Processing: Fundamentals (Berlin, Heidelberg: Springer)

Dauphin, Y., Pascanu, R., Gulcehre, C., et al. 2014, arXiv e-prints, arXiv:1406.2572

Davidon, W. C. 1968, The Computer Journal, 10, 406

de Jong, J. T. A., Verdoes Kleijn, G. A., Boxhoorn, D. R., et al. 2015, A&A, 582, A62

de Vaucouleurs, G. 1948, Annales d'Astrophysique, 11, 247

Delli Veneri, M., Cavuoti, S., Brescia, M., Longo, G., & Riccio, G. 2019, MNRAS, 486, 1377

Desprez, G., Richard, J., Jauzac, M., et al. 2018, MNRAS, 479, 2630

Devroye, L., Györfi, L., & Lugosi, G. 1996, Stochastic Modelling and Applied Probability, Vol. 31, A Probabilistic Theory of Pattern Recognition (Springer), 1–638

Diego, J. M., Broadhurst, T., Benitez, N., Lim, J., & Lam, D. 2015, MNRAS, 449, 588

Diemand, J. & Moore, B. 2011, Advanced Science Letters, 4, 297

D'Isanto, A., Cavuoti, S., Brescia, M., et al. 2016, Monthly Notices of the Royal Astronomical Society, 457, 3119

Djorgovski, S. G., Donalek, C., Mahabal, A., et al. 2006, arXiv e-prints, astro

Djorgovski, S. G., Longo, G., Brescia, M., et al. 2012, in American Astronomical Society Meeting Abstracts, Vol. 219, American Astronomical Society Meeting Abstracts #219, 145.12

Djorgovski, S. G. & Williams, R. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 345, From Clark Lake to the Long Wavelength Array: Bill Erickson's Radio Science, ed. N. Kassim, M. Perez, W. Junor, & P. Henning, 517

Dressler, A. 1980, ApJ, 236, 351

Dressler, A. 1984, in Astrophysics and Space Science Library, Vol. 111, Clusters and Groups of Galaxies, ed. F. Mardirossian, G. Giuricin, & M. Mezzetti, 117

Dressler, A. & Gunn, J. E. 1983, ApJ, 270, 7

Driver, S. P., Allen, P. D., Graham, A. W., et al. 2006, MNRAS, 368, 414

Duchi, J., Hazan, E., & Singer, Y. 2011, J. Mach. Learn. Res., 12, 2121

Dugas, C., Bengio, Y., Bélisle, F., Nadeau, C., & Garcia, R. 2000, in Proceedings of the 13th International Conference on Neural Information Processing Systems, NIPS'00 (Cambridge, MA, USA: MIT Press), 451–457

Dyson, F. W., Eddington, A. S., & Davidson, C. 1920, Philosophical Transactions of the Royal Society of London Series A, 220, 291

Einstein, A. 1916, Annalen der Physik, 354, 769

Elíasdóttir, Á., Limousin, M., Richard, J., et al. 2007, arXiv e-prints, arXiv:0710.5636

Floudas, C. A. & Gounaris, C. E. 2009, J. of Global Optimization, 45, 3–38

Fogarty, K., Postman, M., Larson, R., Donahue, M., & Moustakas, J. 2017, ApJ, 846, 103

Frontera-Pons, J., Sureau, F., Bobin, J., & Le Floc'h, E. 2017, A&A, 603, A60

Fukugita, M., Ichikawa, T., Gunn, J. E., et al. 1996, AJ, 111, 1748

Fukushima, K. 1980, Biological Cybernetics, 36, 193

Furtak, L. J., Atek, H., Lehnert, M. D., Chevallard, J., & Charlot, S. 2021, MNRAS, 501, 1568

Galluccio, L., Michel, O., Bendjoya, P., & Slezak, E. 2008, in American Institute of Physics Conference Series, Vol. 1082, Classification and Discovery in Large Astronomical Surveys, ed. C. A. L. Bailer-Jones, 165–171

Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, Space Sci. Rev., 123, 485

Garilli, B., Fumana, M., Franzetti, P., et al. 2010, PASP, 122, 827

Gatys, L. A., Ecker, A. S., & Bethge, M. 2015, arXiv e-prints, arXiv:1508.06576

Gavazzi, R., Marshall, P. J., Treu, T., & Sonnenfeld, A. 2014, ApJ, 785, 144

Gavazzi, R. & Soucail, G. 2007, A&A, 462, 459

Genel, S., Vogelsberger, M., Springel, V., et al. 2014, MNRAS, 445, 175

Gentile, F., Tortora, C., Covone, G., et al. 2022, MNRAS, 510, 500

Girardi, M., Mercurio, A., Balestra, I., et al. 2015, A&A, 579, A4

Girshick, R., Donahue, J., Darrell, T., & Malik, J. 2013, arXiv e-prints, arXiv:1311.2524

Glazebrook, K., Offer, A. R., & Deeley, K. 1998, ApJ, 492, 98

Glorot, X., Bordes, A., & Bengio, Y. 2011, in Proceedings of Machine Learning Research, Vol. 15, Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, ed. G. J. Gordon, D. B. Dunson, & M. Dudík (Fort Lauderdale, FL, USA: PMLR), 315–323

Gobat, R., Rosati, P., Strazzullo, V., et al. 2008, A&A, 488, 853

Gómez, P. L., Valkonen, L. E., Romer, A. K., et al. 2012, AJ, 144, 79

Gómez-Valent, A. & Amendola, L. 2018, J. Cosmology Astropart. Phys., 2018, 051

Gonzalez-Dominguez, J., Lopez-Moreno, I., & Sak, H. 2014, in Interspeech

Goodfellow, I., Bengio, Y., & Courville, A. 2016, Deep Learning (MIT Press), http://www.deeplearningbook.org

Goodfellow, I. J. 2010, Technical Report: Multidimensional, Downsampled Convolution for Autoencoders, Tech. rep., Université de Montréal

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., et al. 2014a, arXiv e-prints, arXiv:1406.2661

Goodfellow, I. J., Shlens, J., & Szegedy, C. 2014b, arXiv e-prints, arXiv:1412.6572

Goodfellow, I. J., Vinyals, O., & Saxe, A. M. 2014c, arXiv e-prints, arXiv:1412.6544

Goodfellow, I. J., Warde-Farley, D., Mirza, M., Courville, A., & Bengio, Y. 2013, arXiv e-prints, arXiv:1302.4389

Graham, A. W. 2001, AJ, 121, 820

Graham, M. J., Djorgovski, S. G., Mahabal, A. A., et al. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert, 394

Griffin, R. F. 1967, ApJ, 148, 465

Grillo, C., Karman, W., Suyu, S. H., et al. 2016, ApJ, 822, 78

Grillo, C., Rosati, P., Suyu, S. H., et al. 2018, ApJ, 860, 94

Grillo, C., Suyu, S. H., Rosati, P., et al. 2015, ApJ, 800, 38

Groetsch, C. 1984, The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind, Research Notes in Mathematics Series (Pitman Advanced Pub. Program)

Gruen, D., Brimioulle, F., Seitz, S., et al. 2013, MNRAS, 432, 1455

Gunn, J. E. & Gott, J. Richard, I. 1972, ApJ, 176, 1

Guyon, I. & Elisseeff, A. 2003, J. Mach. Learn. Res., 3, 1157

Guyon, I. & Elisseeff, A. 2006, An Introduction to Feature Extraction, ed. I. Guyon, M. Nikravesh, S. Gunn, & L. A. Zadeh (Berlin, Heidelberg: Springer Berlin Heidelberg), 1–25

Hanley, J. A. & McNeil, B. J. 1982, Radiology, 143, 29

Hara, S. & Maehara, T. 2016, arXiv e-prints, arXiv:1611.05940

Hara, S. & Maehara, T. 2017, in 31st AAAI Conference on Artificial Intelligence, AAAI 2017, 1985–1991

Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. 2017, Neuron, 95, 245

Hastie, T., Tibshirani, R., & Friedman, J. 2001, The Elements of Statistical Learning, Springer Series in Statistics (New York, NY, USA: Springer New York Inc.)

He, K., Gkioxari, G., Dollár, P., & Girshick, R. 2017, arXiv e-prints, arXiv:1703.06870

He, K., Zhang, X., Ren, S., & Sun, J. 2015a, arXiv e-prints, arXiv:1512.03385

He, K., Zhang, X., Ren, S., & Sun, J. 2015b, arXiv e-prints, arXiv:1502.01852

He, K., Zhang, X., Ren, S., & Sun, J. 2015, IEEE Transactions on Pattern Analysis & Machine Intelligence, 37, 1904

Heavens, A. F. 1993, MNRAS, 263, 735

Hebb, D. O. 1949, The organization of behavior : a neuropsychological theory / D.O. Hebb (Wiley New York), xix, 335 p. :

Herenz, E. C., Urrutia, T., Wisotzki, L., et al. 2017, A&A, 606, A12

Hey, T. 2012, in E-Science and Information Management, ed. S. Kurbanoğlu, U. Al, P. L. Erdoğan, Y. Tonta, & N. Uçak (Berlin, Heidelberg: Springer Berlin Heidelberg), 1–1

Hinton, G. E. 2012, Neural networks for machine learning, Coursera, video lectures

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. 2012, arXiv e-prints, arXiv:1207.0580

Ho, M., Rau, M. M., Ntampaka, M., et al. 2019, ApJ, 887, 25

Hocking, A., Geach, J. E., Davey, N., & Sun, Y. 2015, arXiv e-prints, arXiv:1507.01589

Hoekstra, H., Bartelmann, M., Dahle, H., et al. 2013, Space Sci. Rev., 177, 75

Hojnacki, S. M., Kastner, J. H., Micela, G., Feigelson, E. D., & LaLonde, S. M. 2007, ApJ, 659, 585

Hornik, K., Stinchcombe, M., & White, H. 1989, Neural Networks, 2, 359–366

Hoyle, B. 2016, Astronomy and Computing, 16, 34

Hubble, E. & Humason, M. L. 1931, ApJ, 74, 43

Hubble, E. P. 1926, ApJ, 64, 321

Hubble, E. P. 1936, Realm of the Nebulae, Dover books on science (Dover Publications)

Hubel, D. H. & Wiesel, T. N. 1959, The Journal of Physiology, 148, 574

Huesmann, K., Garcia Rodriguez, L., Linsen, L., & Risse, B. 2021, arXiv e-prints, arXiv:2104.06153

Huval, B., Wang, T., Tandon, S., et al. 2015, arXiv e-prints, arXiv:1504.01716

Ioffe, S. & Szegedy, C. 2015, arXiv e-prints, arXiv:1502.03167

Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111

Jackson, N. 2008, MNRAS, 389, 1311

Jacobs, C., Collett, T., Glazebrook, K., et al. 2019a, ApJS, 243, 17

Jacobs, C., Collett, T., Glazebrook, K., et al. 2019b, MNRAS, 484, 5330

Jannik Bjerrum, E. 2017, arXiv e-prints, arXiv:1703.07076

Jiang, L., Wu, J., Bian, F., et al. 2018, Nature Astronomy, 2, 962

Jiménez-Teja, Y., Dupke, R., Benítez, N., et al. 2018, ApJ, 857, 79

Johnson, H. L. & Morgan, W. W. 1953, ApJ, 117, 313

Jolliffe, I. 2011, Principal Component Analysis, ed. M. Lovric (Berlin, Heidelberg: Springer Berlin Heidelberg), 1094–1096

Jullo, E. & Kneib, J. P. 2009, MNRAS, 395, 1319

Jullo, E., Kneib, J. P., Limousin, M., et al. 2007, New Journal of Physics, 9, 447

Kamath, C. 2009, Scientific Data Mining: A Practical Perspective (USA: Society for Industrial and Applied Mathematics)

Kassiola, A. & Kovner, I. 1993, ApJ, 417, 450

Kauffmann, G., White, S. D. M., Heckman, T. M., et al. 2004, MNRAS, 353, 713

Kelson, D. D., Martini, P., & Mulchaey, J. S. 2003, Optimal Measurements of Redshifts using the Weighted Cross-Correlation, https://code.obs.carnegiescience.edu/Algorithms/realcc/view

Kikuchihara, S., Ouchi, M., Ono, Y., et al. 2020, ApJ, 893, 60

Kim, D.-W., Yeo, D., Bailer-Jones, C. A. L., & Lee, G. 2021, arXiv e-prints, arXiv:2106.00187

Kingma, D. P. & Ba, J. 2014, CoRR, abs/1412.6980

Kinney, A. L., Calzetti, D., Bohlin, R. C., et al. 1996, ApJ, 467, 38

Kleanthous, C. & Chatzis, S. 2020, Knowledge-Based Systems, 188, 105048

Kneib, J. P., Ellis, R. S., Smail, I., Couch, W. J., & Sharples, R. M. 1996, ApJ, 471, 643

Koekemoer, A. M., Aussel, H., Calzetti, D., et al. 2007, ApJS, 172, 196

Koekemoer, A. M., Avila, R. J., Hammer, D., et al. 2014, in American Astronomical Society Meeting Abstracts, Vol. 223, American Astronomical Society Meeting Abstracts #223, 254.02

Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36

Kohavi, R. 1995, in Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2, IJCAI'95 (San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.), 1137–1143

Kong, K., Li, G., Ding, M., et al. 2020, arXiv e-prints, arXiv:2010.09891

Kormendy, J. 1982, Saas-Fee Advanced Course, 12, 115

Kravtsov, A. V. & Borgani, S. 2012, Annual Review of Astronomy and Astrophysics, 50, 353

Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2012, in Advances in Neural Information Processing Systems 25, ed. F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Curran Associates, Inc.), 1097–1105

Kuffler, S. W. 1953, Journal of Neurophysiology, 16, 37, pMID: 13035466

Kullback, S. & Leibler, R. A. 1951, The Annals of Mathematical Statistics, 22, 79

Lagattuta, D. J., Richard, J., Bauer, F. E., et al. 2019, MNRAS, 485, 3738

Lagattuta, D. J., Richard, J., Clément, B., et al. 2017, MNRAS, 469, 3946

Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, ApJS, 224, 24

Lanusse, F., Mandelbaum, R., Ravanbakhsh, S., et al. 2021, MNRAS, 504, 5543

Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv e-prints, arXiv:1110.3193

Le Fèvre, O., Cassata, P., Cucciati, O., et al. 2013, A&A, 559, A14

Le Fèvre, O. & Hammer, F. 1988, ApJ, 333, L37

Lecun, Y., Bengio, Y., & Hinton, G. 2015, Nature, 521, 436

LeCun, Y., Boser, B., Denker, J., et al. 1990, Handwritten Digit Recognition with a Back-Propagation Network, ed. D. Touretzky, Vol. 2 (San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.), 396–404

LeCun, Y., Boser, B., Denker, J. S., et al. 1989, Neural Comput., 1, 541

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998, Proceedings of the Institute of Radio Engineers, 86, 2278

Li, R., Napolitano, N. R., Roy, N., et al. 2021a, arXiv e-prints, arXiv:2111.05434

Li, R., Napolitano, N. R., Spiniello, C., et al. 2021b, ApJ, 923, 16

Li, R., Napolitano, N. R., Tortora, C., et al. 2020, ApJ, 899, 30

Limousin, M., Kneib, J.-P., & Natarajan, P. 2005, MNRAS, 356, 309

Lin, M., Chen, Q., & Yan, S. 2013, arXiv e-prints, arXiv:1312.4400

Liske, J., Lemon, D. J., Driver, S. P., Cross, N. J. G., & Couch, W. J. 2003, MNRAS, 344, 307

Litjens, G., Kooi, T., Ehteshami Bejnordi, B., et al. 2017, arXiv e-prints, arXiv:1702.05747

Lombardi, M. & Bertin, G. 1999, A&A, 342, 337

Lombardi, M., Rosati, P., Blakeslee, J. P., et al. 2005, ApJ, 623, 42

Longair, M. S. 2008, Galaxy formation, 2nd edn., Astronomy and Astrophysics Library (Springer Berlin Heidelberg)

Lopes, P. A. A. & Ribeiro, A. L. B. 2020, MNRAS, 493, 3429

Lotz, J., Mountain, M., Grogin, N. A., et al. 2014, in American Astronomical Society Meeting Abstracts, Vol. 223, American Astronomical Society Meeting Abstracts #223, 254.01

Lotz, J. M., Koekemoer, A., Coe, D., et al. 2017, ApJ, 837, 97

Lu, C. & Tang, X. 2014, arXiv e-prints, arXiv:1404.3840

Lukic, V., Brüggen, M., Banfield, J. K., et al. 2018, MNRAS, 476, 246

Maas, A. L., Hannun, A. Y., & Ng, A. Y. 2013, in ICML Workshop on Deep Learning for Audio, Speech and Language Processing

Mahendran, A. & Vedaldi, A. 2014, arXiv e-prints, arXiv:1412.0035

Mahler, G., Richard, J., Clément, B., et al. 2018, MNRAS, 473, 663

Marlin, B. 2008, PhD thesis, Department of Computer Science, University of Toronto

McKerns, M. M., Strand, L., Sullivan, T., Fang, A., & Aivazis, M. A. G. 2012, arXiv e-prints, arXiv:1202.1056

Medezinski, E., Umetsu, K., Okabe, N., et al. 2016, ApJ, 817, 24

Meneghetti, M. 2019, Introduction to Gravitational Lensing, https://www.ita.uni-heidelberg.de/

Meneghetti, M. 2021, Introduction to Gravitational Lensing, 1st edn., Lecture Notes in Physics (Springer International Publishing)

Meneghetti, M., Davoli, G., Bergamini, P., et al. 2020, Science, 369, 1347

Mercurio, A., Annunziatella, M., Biviano, A., et al. 2016, in The Universe of Digital Sky Surveys, ed. N. R. Napolitano, G. Longo, M. Marconi, M. Paolillo, & E. Iodice, Vol. 42, 225

Mercurio, A., Rosati, P., Biviano, A., et al. 2021, arXiv e-prints, arXiv:2109.03305

Merten, J., Meneghetti, M., Postman, M., et al. 2015, ApJ, 806, 4

Metcalf, R. B., Meneghetti, M., Avestruz, C., et al. 2019, A&A, 625, A119

Metcalfe, N., Shanks, T., Campos, A., McCracken, H. J., & Fong, R. 2001, MNRAS, 323, 795

Meyer, R. A., Laporte, N., Ellis, R. S., Verhamme, A., & Garel, T. 2021, MNRAS, 500, 558

Mikołajczyk, A. & Grochowski, M. 2018, in 2018 International Interdisciplinary PhD Workshop (IIPhDW), 117–122

Millon, M., Galan, A., Courbin, F., et al. 2020, A&A, 639, A101

Minsky, M. & Papert, S. 1969, Perceptrons (Cambridge, MA: MIT Press)

Mirza, M. & Osindero, S. 2014, arXiv e-prints, arXiv:1411.1784

Miyazaki, S., Komiyama, Y., Kawanomoto, S., et al. 2018, PASJ, 70, S1

Miyazaki, S., Komiyama, Y., Sekiguchi, M., et al. 2002, PASJ, 54, 833

Mo, H., van den Bosch, F. C., & White, S. 2010, Galaxy Formation and Evolution (Cambridge University Press)

Molino, A., Benítez, N., Ascaso, B., et al. 2017, MNRAS, 470, 95

Molino, A., Costa-Duarte, M. V., Mendes de Oliveira, C., et al. 2019, A&A, 622, A178

Monna, A., Seitz, S., Greisel, N., et al. 2014, MNRAS, 438, 1417

Monna, A., Seitz, S., Zitrin, A., et al. 2015, MNRAS, 447, 1224

Nakoneczny, S., Bilicki, M., Solarz, A., et al. 2019, A&A, 624, A13

Nantais, J. B., Rettura, A., Lidman, C., et al. 2013, A&A, 556, A112

Narayan, R. & Bartelmann, M. 1996, arXiv e-prints, astro

Natarajan, P., Chadayammuri, U., Jauzac, M., et al. 2017, MNRAS, 468, 1962

Nesterov, Y. 1983, Proceedings of the USSR Academy of Sciences, 269, 543

Nesterov, Y. 2003, Introductory Lectures on Convex Optimization: A Basic Course, Applied Optimization (Springer US)

Newman, A. B., Treu, T., Ellis, R. S., & Sand, D. J. 2011, ApJ, 728, L39

Newman, A. B., Treu, T., Ellis, R. S., et al. 2009, ApJ, 706, 1078

Newton, I. 1704, Opticks (Dover Press)

Nielsen, M. 2015, Neural Networks and Deep Learning (Determination Press)

Noll, S., Mehlert, D., Appenzeller, I., et al. 2004, A&A, 418, 885

Nousiainen, J., Rajani, C., Kasper, M., & Helin, T. 2021, Optics Express, 29, 15327

Nowlan, S. J. & Hinton, G. E. 1992, Neural Computation (NECO), 4, 473

Ntampaka, M., Trac, H., Sutherland, D. J., et al. 2015, ApJ, 803, 50

Ntampaka, M., Trac, H., Sutherland, D. J., et al. 2016, ApJ, 831, 135

Ntampaka, M., ZuHone, J., Eisenstein, D., et al. 2019, ApJ, 876, 82

Nulsen, P. E. J. 1982, MNRAS, 198, 1007

Oemler, Augustus, J. 1974, ApJ, 194, 1

Oh, K.-S. & Jung, K. 2004, Pattern Recognition, 37, 1311

Pacaud, F., Clerc, N., Giles, P. A., et al. 2016, A&A, 592, A2

Paolillo, M., Puzia, T. H., Goudfrooij, P., et al. 2011, ApJ, 736, 90

Parker, R. 2010, Missing Data Problems in Machine Learning (VDM Verlag)

Parkes, A. I., Camilleri, J., Hudson, D. A., & Sobey, A. J. 2021, arXiv e-prints, arXiv:2106.03428

Parry, W. G., Grillo, C., Mercurio, A., et al. 2016, MNRAS, 458, 1493

Pasian, F., Brescia, M., & Longo, G. 2012, in Science: Image in Action, ed. B. Zavidovique & G. Lo Bosco, 230–240

Pasquet-Itam, J. & Pasquet, J. 2018, A&A, 611, A97

Pawase, R. S., Courbin, F., Faure, C., Kokotanekova, R., & Meylan, G. 2014, MNRAS, 439, 3392

Pearson, K. 1895, Proceedings of the Royal Society of London Series I, 58, 240

Pearson, K. A., Palafox, L., & Griffith, C. A. 2018, MNRAS, 474, 478

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, J. Mach. Learn. Res., 12, 2825

Peebles, P. J. E. 1982, ApJ, 263, L1

Perez, L. & Wang, J. 2017, arXiv e-prints, arXiv:1712.04621

Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2019a, MNRAS, 482, 807

Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2017, MNRAS, 472, 1129

Petrillo, C. E., Tortora, C., Vernardos, G., et al. 2019b, MNRAS, 484, 3879

Pignataro, G. V., Bergamini, P., Meneghetti, M., et al. 2021, arXiv e-prints, arXiv:2106.10286

Pizzuti, L., Sartoris, B., Borgani, S., et al. 2016, J. Cosmology Astropart. Phys., 2016, 023

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A27

Polyak, B. 1964, USSR Computational Mathematics and Mathematical Physics, 4, 1

Poole, B., Sohl-Dickstein, J., & Ganguli, S. 2014, arXiv e-prints, arXiv:1406.1831

Postman, M., Coe, D., Benítez, N., et al. 2012a, ApJS, 199, 25

Postman, M., Lauer, T. R., Donahue, M., et al. 2012b, ApJ, 756, 159

Prechelt, L. 1997, in Neural Networks: Tricks of the Trade, volume 1524 of LNCS, chapter 2 (Springer-Verlag), 55–69

Press, W. H. & Schechter, P. 1974, ApJ, 187, 425

Prugniel, P. & Simien, F. 1997, A&A, 321, 111

Rafelski, M., Teplitz, H. I., Gardner, J. P., et al. 2015, AJ, 150, 31

Raskutti, G., Wainwright, M. J., & Yu, B. 2011, in 2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 1318–1325

Reddi, S. J., Kale, S., & Kumar, S. 2019, arXiv e-prints, arXiv:1904.09237

Reiman, D. M. & Göhre, B. E. 2019, MNRAS, 485, 2617

Ren, S., He, K., Girshick, R., & Sun, J. 2015, arXiv e-prints, arXiv:1506.01497

Renzini, A. 2006, ARA&A, 44, 141

Rettura, A., Martinez-Manso, J., Stern, D., et al. 2014, ApJ, 797, 109

Riedmiller, M. & Braun, H. 1993, in IEEE International Conference on Neural Networks, 586–591 vol.1

Riess, A. G., Rodney, S. A., Scolnic, D. M., et al. 2018, ApJ, 853, 126

Robbins, H. & Monro, S. 1951, The Annals of Mathematical Statistics, 22, 400

Rojas, R. 1996, Neural Networks - A Systematic Introduction (Berlin: Springer-Verlag)

Rosati, P. 2018, Nature Astronomy, 2, 944

Rosati, P., Balestra, I., Grillo, C., et al. 2014, The Messenger, 158, 48

Rosati, P. & Clash-VLT Team. 2020, A&A, in preparation

Rothschild, T., Nagai, D., Aung, H., et al. 2021, arXiv e-prints, arXiv:2110.02232

Rousseeuw, P. J. 1984, Journal of the American Statistical Association, 79, 871

Rousseeuw, P. J. & Driessen, K. 2006, Data Min. Knowl. Discov., 12, 29–45

Ryan, R. E., J. & Reid, I. N. 2016, AJ, 151, 92

Rykoff, E. S., Rozo, E., Busha, M. T., et al. 2014, ApJ, 785, 104

Ryon, J. E. 2021, Advanced Camera for Surveys Instrument Handbook for Cycle 29 v. 20.0

Sadeh, I., Abdalla, F. B., & Lahav, O. 2016, PASP, 128, 104502

Salmon, B., Coe, D., Bradley, L., et al. 2020, ApJ, 889, 189

Salmon, B., Coe, D., Bradley, L., et al. 2018, ApJ, 864, L22

Sartoris, B., Biviano, A., Rosati, P., et al. 2014, ApJ, 783, L11

Saxe, A. M., McClelland, J. L., & Ganguli, S. 2013, arXiv e-prints, arXiv:1312.6120

Sayers, J., Montaña, A., Mroczkowski, T., et al. 2019, ApJ, 880, 45

Schaul, T., Zhang, S., & LeCun, Y. 2012, arXiv e-prints, arXiv:1206.1106

Schmidhuber, J. 2014, arXiv e-prints, arXiv:1404.7828

Schmidt, K. B., Treu, T., Brammer, G. B., et al. 2014, ApJ, 782, L36

Schmidt, S. J., Malz, A. I., Soo, J. Y. H., et al. 2020, MNRAS, 499, 1587

Schneider, P. 2006, Extragalactic Astronomy and Cosmology (Springer Berlin Heidelberg)

Schneider, P., King, L., & Erben, T. 2000, A&A, 353, 41

Scoville, N., Aussel, H., Brusa, M., et al. 2007, ApJS, 172, 1

Sehra, S., Flores, D., & Montanez, G. D. 2021, arXiv e-prints, arXiv:2102.02850

Seidel, G. & Bartelmann, M. 2007, A&A, 472, 341

Sérsic, J. L. 1963, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 6, 41

Sérsic, J. L. 1968, Atlas de Galaxias Australes (Cordoba, Argentina: Observatorio Astronomico)

Shan, H., Kneib, J.-P., Tao, C., et al. 2012, ApJ, 748, 56

Shibuya, T., Ouchi, M., & Harikane, Y. 2015, ApJS, 219, 15

Shorten, C. & Khoshgoftaar, T. 2019, Journal of Big Data, 6

Silver, D., Huang, A., Maddison, C. J., et al. 2016, Nature, 529, 484

Silver, D., Hubert, T., Schrittwieser, J., et al. 2018, Science, 362, 1140

Simard, P. Y., Steinkrau, D., & Buck, I. 2005, in Eighth International Conference on Document Analysis and Recognition (ICDAR'05)(ICDAR), Vol. 00, 1115–1119

Simkin, S. M. 1974, A&A, 31, 129

Simonyan, K. & Zisserman, A. 2014, arXiv e-prints, arXiv:1409.1556

Smith, G. P., Kneib, J.-P., Smail, I., et al. 2005, MNRAS, 359, 417

Smith, G. W. & Leymarie, F. F. 2017, Arts, 6

Sonnenfeld, A., Chan, J. H. H., Shu, Y., et al. 2018, PASJ, 70, S29

Sonnenfeld, A., Jaelani, A. T., Chan, J., et al. 2019, A&A, 630, A71

Sonnenfeld, A., Treu, T., Gavazzi, R., et al. 2013, ApJ, 777, 98

Sonnenfeld, A., Verma, A., More, A., et al. 2020, A&A, 642, A148

Sontag, E. D. & Sussmann, H. J. 1989, Complex Systems, 3, 91

Sparke, L. S. & Gallagher, John S., I. 2007, Galaxies in the Universe: An Introduction, 2nd edn. (Cambridge University Press)

Spiniello, C., Trager, S. C., Koopmans, L. V. E., & Chen, Y. P. 2012, ApJ, 753, L32

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. 2014, J. Mach. Learn. Res., 15, 1929

Statler, T. 1995, AJ, 109, 1371

Stehman, S. V. 1997, Remote Sensing of Environment, 62, 77

Stock, D., Meyer, S., Sarli, E., et al. 2015, A&A, 584, A63

Strait, V., Bradač, M., Coe, D., et al. 2020, ApJ, 888, 124

Sunyaev, R. A. & Zeldovich, Y. B. 1972, Comments on Astrophysics and Space Physics, 4, 173

Sutskever, I., Martens, J., Dahl, G., & Hinton, G. 2013, in Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28, ICML'13 (JMLR.org), III–1139–III–1147

Sygnet, J. F., Tu, H., Fort, B., & Gavazzi, R. 2010, A&A, 517, A25

Szandała, T. 2020, arXiv e-prints, arXiv:2010.09458

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. 2016, arXiv e-prints, arXiv:1602.07261

Szegedy, C., Liu, W., Jia, Y., et al. 2014, arXiv e-prints, arXiv:1409.4842

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. 2015, arXiv e-prints, arXiv:1512.00567

Szegedy, C., Zaremba, W., Sutskever, I., et al. 2013, arXiv e-prints, arXiv:1312.6199

Thorne, B., Knox, L., & Prabhu, K. 2021, MNRAS, 504, 2603

Tibshirani, R. J. 2013, Electron. J. Statist., 7, 1456

Tonry, J. & Davis, M. 1979, AJ, 84, 1511

Torres, G., Latham, D. W., & Stefanik, R. P. 2007, ApJ, 662, 602

Tortorelli, L., Mercurio, A., Paolillo, M., et al. 2018, MNRAS, 477, 648

Treu, T., Brammer, G., Diego, J. M., et al. 2016, ApJ, 817, 60

Treu, T., Ellis, R. S., Kneib, J.-P., et al. 2003, ApJ, 591, 53

Treu, T. & Koopmans, L. V. E. 2002, ApJ, 575, 87

Treu, T., Schmidt, K. B., Brammer, G. B., et al. 2015, ApJ, 812, 114

Tuccillo, D., Huertas-Company, M., Decencière, E., & Velasco-Forero, S. 2017, in Astroinformatics, ed. M. Brescia, S. G. Djorgovski, E. D. Feigelson, G. Longo, & S. Cavuoti, Vol. 325, 191–196

Tuccillo, D., Huertas-Company, M., Decencière, E., et al. 2018, MNRAS, 475, 894

Umetsu, K. 2013, ApJ, 769, 13

Umetsu, K. & Broadhurst, T. 2008, ApJ, 684, 177

Umetsu, K., Broadhurst, T., Zitrin, A., Medezinski, E., & Hsu, L.-Y. 2011, ApJ, 729, 127

Umetsu, K., Medezinski, E., Nonino, M., et al. 2014, ApJ, 795, 163

Umetsu, K., Sereno, M., Tam, S.-I., et al. 2018, ApJ, 860, 104

Umetsu, K., Tada, M., & Futamase, T. 1999, Progress of Theoretical Physics Supplement, 133, 53

Vaccari, M., Covone, G., Radovich, M., et al. 2016, in The 4th Annual Conference on High Energy Astrophysics in Southern Africa (HEASA 2016), 26

Vanzella, E., Calura, F., Meneghetti, M., et al. 2019, MNRAS, 483, 3618

Vanzella, E., Calura, F., Meneghetti, M., et al. 2017a, MNRAS, 467, 4304

Vanzella, E., Caminha, G. B., Rosati, P., et al. 2021, A&A, 646, A57

Vanzella, E., Castellano, M., Meneghetti, M., et al. 2017b, ApJ, 842, 47

Vanzella, E., Meneghetti, M., Caminha, G. B., et al. 2020, MNRAS, 494, L81

von Luxburg, U. & Schoelkopf, B. 2008, arXiv e-prints, arXiv:0810.4752

von Marttens, R., Casarini, L., Napolitano, N. R., et al. 2021, arXiv e-prints, arXiv:2111.01185

Wager, S., Wang, S., & Liang, P. 2013, arXiv e-prints, arXiv:1307.1493

Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., & Lang, K. 1989, Acoustics, Speech and Signal Processing, IEEE Transactions on, 37, 328

Wang, Y. & Li, Z. 2021, arXiv e-prints, arXiv:2107.08593

Warde-Farley, D., Goodfellow, I. J., Courville, A., & Bengio, Y. 2013, arXiv e-prints, arXiv:1312.6197

Webb, S., Lochner, M., Muthukrishna, D., et al. 2020, MNRAS, 498, 3077

Weinmann, S. M., van den Bosch, F. C., Yang, X., & Mo, H. J. 2006, MNRAS, 366, 2

Williams, R. E., Blacker, B., Dickinson, M., et al. 1996, AJ, 112, 1335

Wu, C., Wong, O. I., Rudnick, L., et al. 2019a, MNRAS, 482, 1211

Wu, Y., Liu, L., Bae, J., et al. 2019b, arXiv e-prints, arXiv:1908.06477

Wu, Y., Schuster, M., Chen, Z., et al. 2016, arXiv e-prints, arXiv:1609.08144

Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. 2016, arXiv e-prints, arXiv:1611.05431

Yang, Y., Zabludoff, A. I., Zaritsky, D., Lauer, T. R., & Mihos, J. C. 2004, ApJ, 607, 258

Yao, Y., Rosasco, L., & Caponnetto, A. 2007, Constr. Approx, 289

Yatawatta, S. & Avruch, I. M. 2021, MNRAS, 505, 2141

Zeiler, M. D. 2012, arXiv e-prints, arXiv:1212.5701

Zhang, C., Vinyals, O., Munos, R., & Bengio, S. 2018, arXiv e-prints, arXiv:1804.06893

Zhao, L., Liu, T., Peng, X., & Metaxas, D. 2020, arXiv e-prints, arXiv:2010.08001

Zheng, W., Postman, M., Zitrin, A., et al. 2012, Nature, 489, 406

Zhou & Chellappa. 1988, in IEEE 1988 International Conference on Neural Networks, Vol. 2, 71–78

Zucker, S. 2003, MNRAS, 342, 1291

Zwicky, F. 1937, ApJ, 86, 217

Appendix A

Complementary tables and figures for cluster member identification

In this appendix I report completing tables and figures, which integrate the work on the cluster member identification. The following tables are referred to the *EXP1* (see Sec. 5.3.1): the comparison between the involved architectures by varying the band configuration is report in Tab. A.1; Tab. A.2 shows the VGG performance achieved with the same dataset by varying the band configuration; CNN results by permuting all the involved bands are listed in Tab. A.3. Concerning the *EXP3* (see Sec. 5.3.3): Fig. A.1 and Fig. A.2 show the comparison between the CNN and the benchmark methods (outlined in Sec. 3.8), respectively in term of ROC curves and common predictions; finally, the sky distribution of spectroscopic CLM and candidate members identified by the CNN (see Sec. 5.5 is shown in Fig. A.3, overlapped to the FoV of four clusters.

		*mixed**				*mixed*			
Class	*%*	VGG	ResNet	Gnet	GResNet	VGG	ResNet	Gnet	GResNet
	AE	**89.3**	88.7	88.1	88.4	86.7	86.4	87.0	**87.1**
	pur	88.3	85.7	86.8	**90.0**	83.1	**84.2**	80.4	81.2
CLM	*compl*	86.7	**87.5**	85.5	84.4	88.4	87.1	**92.0**	88.5
	F1	**87.4**	86.9	86.1	87.1	85.6	85.6	**86.2**	84.8
	pur	**90.0**	85.2	89.2	89.7	**90.0**	88.6	83.0	89.0
NCLM	*compl*	91.2	**95.7**	91.0	90.0	85.5	86.1	**88.3**	84.6
	F1	**90.6**	90.4	90.1	89.8	**88.3**	87.3	85.7	85.8

		ACS				*ALL*			
Class	*%*	VGG	ResNet	Gnet	GResNet	VGG	ResNet	Gnet	GResNet
	AE	**87.4**	85.9	87.1	86.5	**87.7**	86.9	**87.7**	87.2
	pur	85.0	83.1	**85.2**	84.7	86.4	86.7	**87.1**	84.6
CLM	*compl*	**88.5**	85.4	86.7	85.8	**86.4**	82.4	85.3	85.9
	F1	**86.7**	84.2	85.9	85.3	**86.4**	84.5	86.2	85.3
	pur	89.9	90.0	**91.6**	90.1	88.9	88.5	88.7	**89.0**
NCLM	*compl*	**86.7**	85.4	84.7	84.8	88.9	**89.4**	88.5	88.7
	F1	**88.3**	87.0	88.1	87.4	**88.9**	**88.9**	88.6	88.8

Table A.1: Performance comparison between the involved architectures (see Sec. 3.7), related to the four band configurations (see Sec. 5.2) and expressed in terms of the statistical estimators described in Sec. 3.6.6. The overall best results are highlighted in bold.

Class	*%*	*mixed*	*ACS*	*ALL*
	AE	84.9	86.4	**87.7**
CLM	*pur*	84.8	**86.8**	86.4
	compl	80.2	81.6	**86.4**
	F1	82.5	84.1	**86.4**
NCLM	*pur*	85.0	86.1	**88.9**
	compl	**90.1**	86.7	88.9
	F1	87.5	88.1	**88.9**

Table A.2: CNN percentage performances achieved with the same dataset by varying the filter configuration (*EXP1*). The performances are related to the band configurations (see Sec. 5.2) and expressed in terms of the statistical estimators described in Sec. 3.6.6. The overall best results are highlighted in bold.

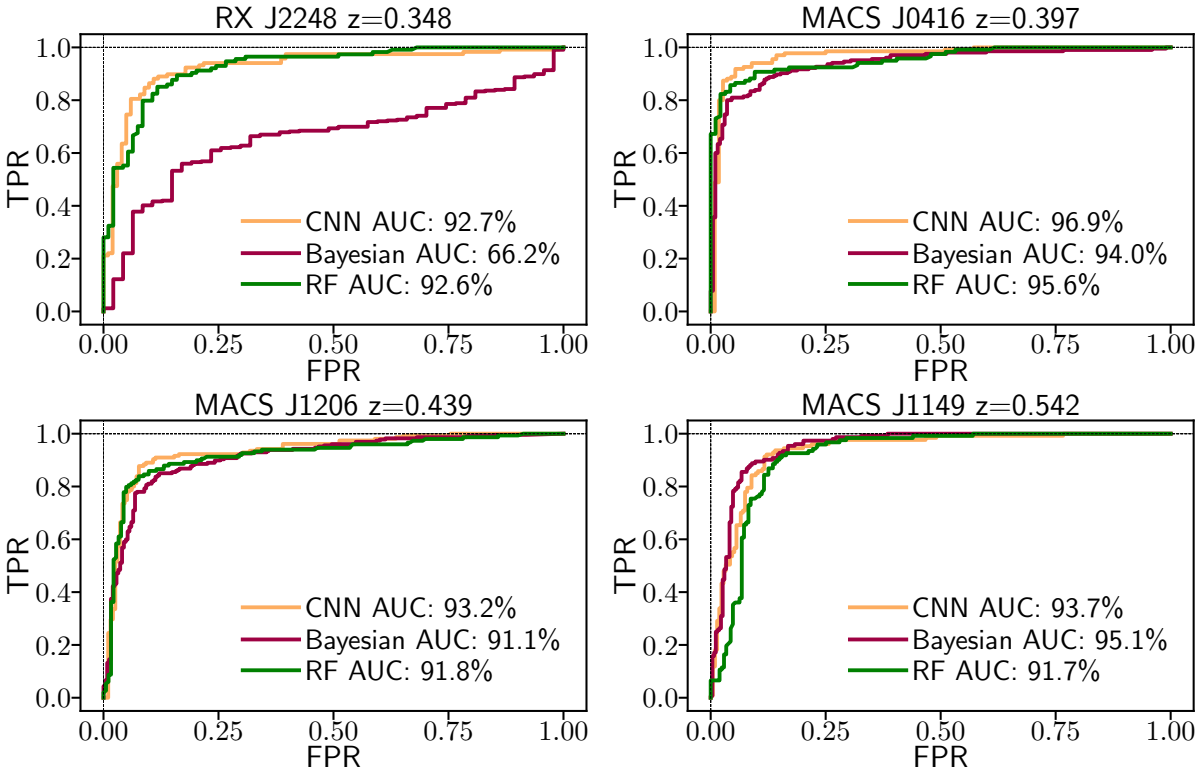


Figure A.1: Comparison between the image-based CNN and two photometric catalogue-based approaches, RF and Bayesian method (*EXP3*) in term of ROC curves for the four clusters: R2248 (top-left panel), M0416 (top-right panel), M1206 (bottom left panel), M1149 (bottom right panel).

			CLM	
Involved filters	*AE*	*pur*	*compl*	*F1*
F435	43.0	43.0	80.0	50.1
F606	75.8	69.6	77.4	73.3
F814	80.6	74.8	82.8	78.6
F105	79.2	73.8	80.0	76.8
F140	80.0	75.0	80.2	77.5
F435_F606	87.1	83.4	87.3	85.3
F435_F814	86.3	82.3	86.7	84.4
F435_F105	85.6	81.1	86.7	83.8
F435_F140	85.3	81.0	85.8	83.3
F606_F814	85.0	81.0	85.0	83.0
F606_F105	85.2	80.3	86.8	83.4
F606_F140	83.1	78.6	83.4	80.9
F814_F105	80.3	76.0	79.3	77.6
F814_F140	80.1	73.8	83.1	78.2
F105_F140	79.3	73.4	81.4	77.2
F435_F606_F814	89.0	87.0	87.6	87.3
F435_F606_F105	88.7	86.3	87.8	87.0
F435_F606_F140	88.6	85.9	88.0	86.9
F435_F814_F105	85.7	82.2	85.1	83.6
F435_F814_F140	85.8	84.3	82.4	83.3
F435_F105_F140	86.8	83.2	86.7	84.9
F606_F814_F105	84.7	82.0	82.5	82.3
F606_F814_F140	85.0	81.6	84.0	82.8
F814_F105_F140	81.0	75.0	83.7	79.1
F435_F606_F814_F105	88.5	84.9	**89.1**	87.0
F435_F606_F814_F140	88.6	85.2	88.8	87.0
F606_F814_F105_F140	84.3	80.4	83.9	82.1
F435_F606_F105_F140	88.6	85.6	88.3	86.9
F435_F814_F105_F140	87.3	84.9	85.8	85.3
F435_F606_F814_F105_F140	**89.3**	**88.3**	86.7	**87.4**

Table A.3: CNN percentage performances achieved by exploring all possible filter combinations for the *mixed** band configuration (see Sec. 5.2) related to the *EXP1*. The performances are expressed in terms of the statistical estimators described in Sec. 3.6.6; for ease of reading, only statistics related to the CLM class are reported, together with the average efficiency (AE), which refers to both classes. The overall best results are highlighted in bold. The last row, referring to the complete usage of all the bands, corresponds to the results achieved with the *mixed** configuration, shown in Tab. 5.2.

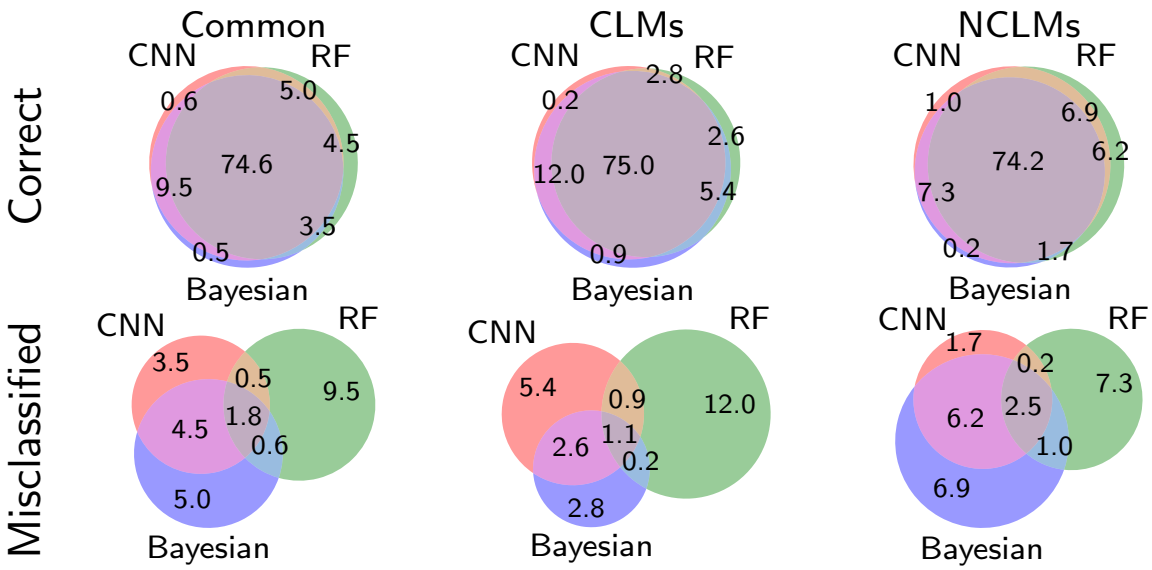


Figure A.2: Venn diagrams reporting the percentages of membership predictions performed by three different methods (CNN, RF, and BM), measured on the common blind test set, obtained by combining the four clusters R2248, M0416, M1206, and M1149 (*EXP3*). On the columns, the common areas refer to the available shared sources, respectively, 460 CLMs and 519 NCLMs). On the rows, common predictions are split between correct and incorrect classifications. Global commonalities can be derived by summing values on the rows.

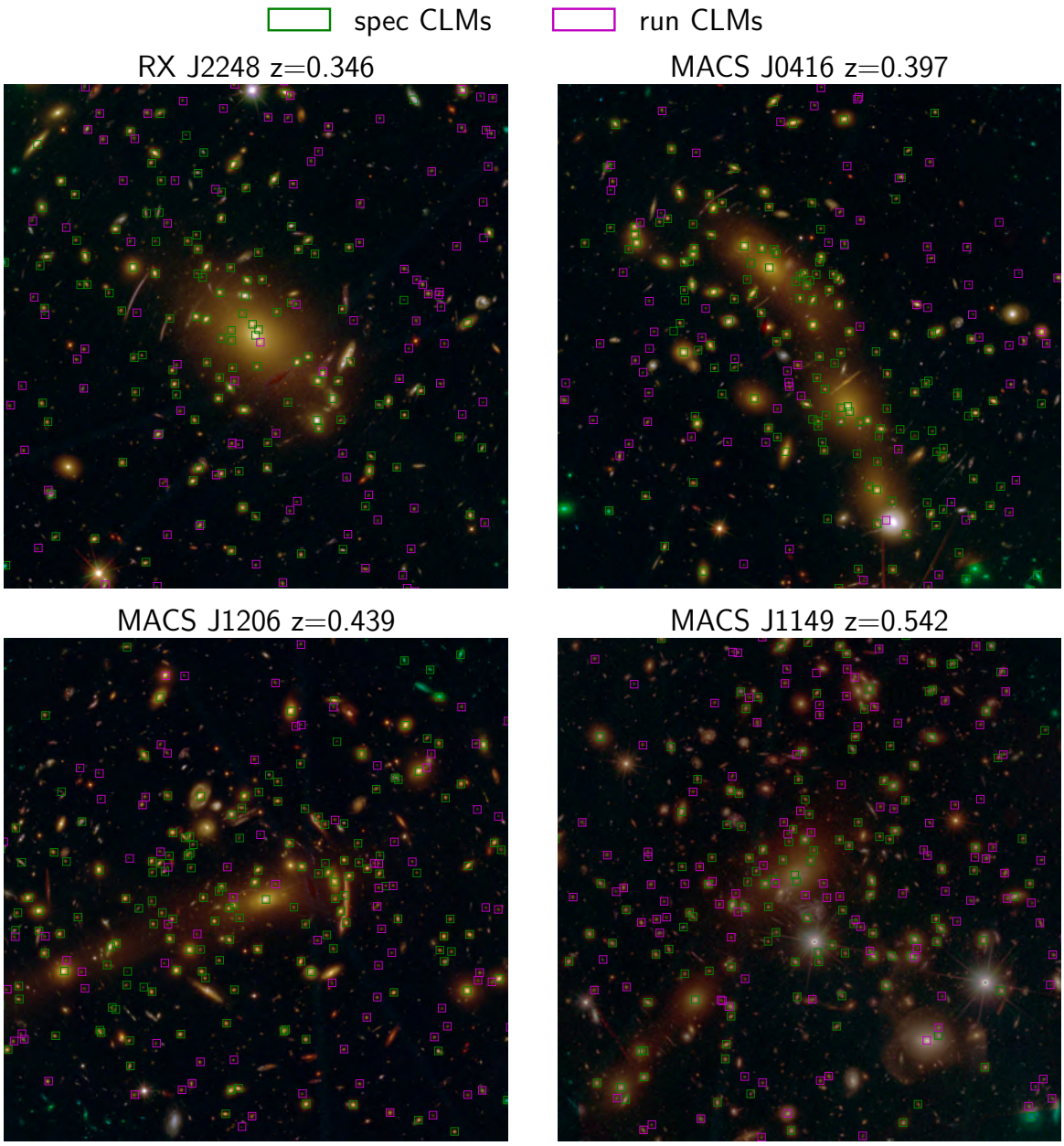


Figure A.3: CNN member selection (marked with open magenta squares) obtained with the *run* set, together with the spectroscopic CLMs (marked with open green squares), in the core of the four clusters R2248 (z = 0.346), M0416 (z = 0.397), M1206 (z = 0.439) and M1149 (z = 0.542). All images are 130 arcsec across.

Appendix B

Complementary tables and figures for GGSL identification

In this appendix, we report completing tables and figures, which integrate the work on the galaxy-galaxy strong-lenses identification. Fig. B.1 shows all the available SEDs, marked in red the starburst template used for the galaxy injection (see Sec. 6.2). Tab. B.1 summarises the comparison between the different involved deep networks for both cutout configurations (see Sec. 6.3.1). Tab. B.2 outlines the comparison between VGG and SC-VGG models by evaluating the distribution of performances evaluated in each fold (discussed in Sec. 6.3.1), while the same comparison in term of False Positive and False Negatives distribution is displayed in Tab. B.3, which is graphically shown in as 2D histograms of (*i*) the False Positive to True Negative ratios in a normalised colour $(F606 - F814)_{\rm norm}$ vs $F814$ magnitude diagram (Fig. B.2) and (*ii*) the False Negative to True Positive ratios in a two diagrams: θ_E vs $F814$ (upper panels in Fig. B.3) and r_e vs $z_{\rm src}$ (bottom panels in Fig. B.3). Tab. B.4 summarises the results achieved by using a single band with the *EXP-A* cutout configuration, together with a comparison with the VGG and the SC-VGG models.

Figs. B.4-B.7 show the classification performed by the CNNs on a set of known real galaxy-galaxy strong-lenses (continuing of Fig. 6.15), whose results are discussed in Sec. 6.4.1. Tabs. B.5-B.7 list all the candidate GGSLs identified by networks with Pr> 0.2, whose cutouts are shown in Fig. 6.16 and in Figs. B.8- B.10 (discussed in Sec. 6.4.1). Finally, the candidate GGSLs selected by networks with Pr> 0.2 are listed in Tabs. B.8-B.15 and shown in Fig. 6.18 and Figs.B.11-B.19 (discussed in 6.4.3).

		EXP-A (∼ 4″ cutouts)					*EXP-B* (∼ 8″ cutouts)				
Class	*%*	VGG	ResNet	Gnet	GResNet	SC-VGG	VGG	ResNet	Gnet	GResNet	SC-VGG
	AE	87.7	76.7	79.1	66.8	**89.4**	79.4	75.5	78.2	**79.5**	79.2
	pur	**93.4**	84.5	87.2	83.8	93.1	89.6	88.2	89.5	87.1	**90.2**
GGSL	*compl*	88.6	81.5	82.1	65.1	**91.7**	78.4	75.0	76.0	78.8	**78.9**
	F1	91.0	83.0	84.6	73.2	**92.4**	84.0	81.1	82.7	82.9	**84.5**
	pur	76.7	60.5	63.5	46.7	**81.4**	72.7	56.7	52.8	67.4	**73.1**
NGGSL	*compl*	**85.4**	65.4	72.1	70.9	84.1	**81.7**	76.7	73.4	81.0	79.3
	F1	81.1	62.8	67.5	56.3	**82.8**	**78.0**	65.4	63.1	74.2	75.7

Table B.1: Performance comparison between the involved architectures (see section 3.7), related to the two cutout configurations, *EXP-A* and *EXP-B* (see section. 6.2) and expressed in terms of the statistical estimators described in Section 3.6.6. The overall best results are highlighted in bold.

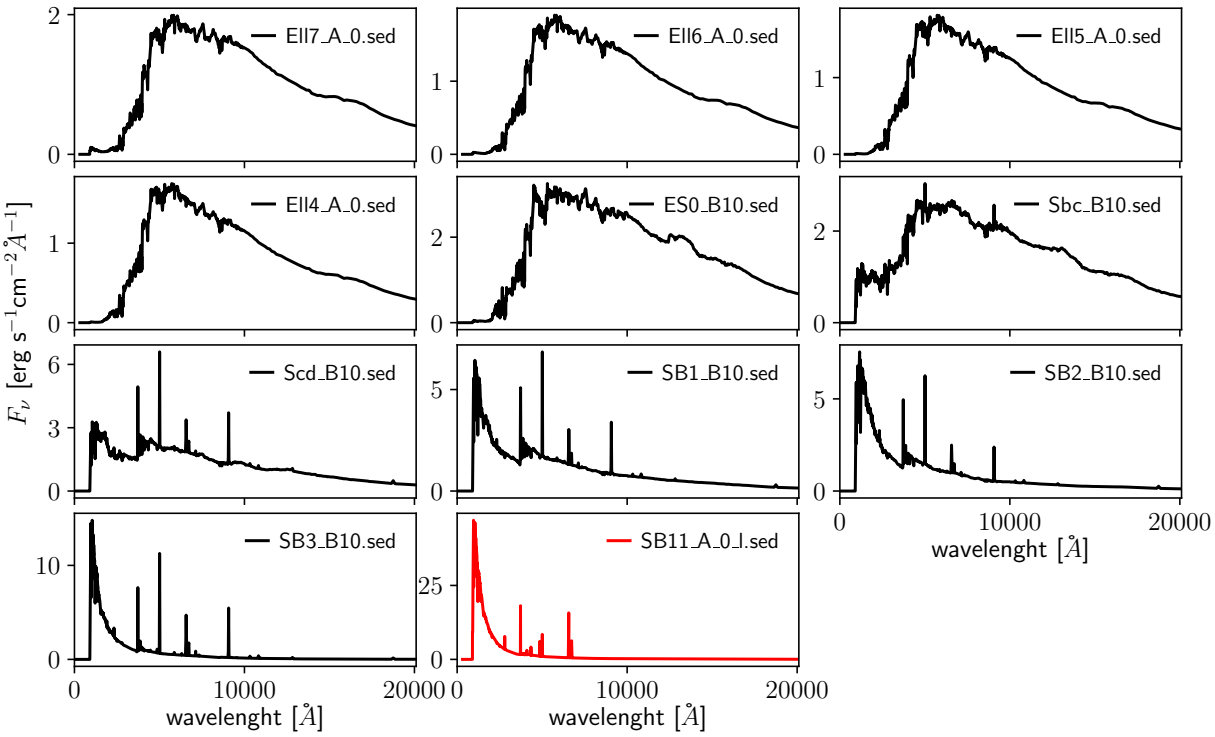


Figure B.1: All available SEDs within the `Barak` package (https://nhmc.github.io/Barak/). The one used in this work is a starburst template from Kinney et al. (1996), plotted in bottom central panel in red).

		median		Q_1		Q_3	
	[%]	VGG	SC-VGG	VGG	SC-VGG	VGG	SC-VGG
	AE	**88.3**	**89.3**	86.2	**88.6**	88.6	**90.3**
	pur	**93.3**	**93.3**	**92.5**	92.3	**94.8**	94.2
GGSL	*compl*	89.1	**91.9**	86.7	**91.2**	90.4	**92.5**
	F1	91.5	**92.3**	89.8	**91.8**	91.5	**93.0**
	pur	77.6	**81.9**	74.0	**79.7**	79.0	**82.9**
NGGSL	*compl*	**85.6**	84.2	**82.3**	81.7	**88.1**	87.2
	F1	91.5	**92.3**	89.8	**91.8**	91.5	**93.0**
		IQR		$Q_1 - 1.5 \cdot IQR$		$Q_3 + 1.5 \cdot IQR$	
	[%]	VGG	SC-VGG	VGG	SC-VGG	VGG	SC-VGG
	AE	2.4	**1.7**	85.3	**87.8**	89.7	**91.4**
	pur	2.3	**1.9**	**90.7**	90.1	95.6	**95.7**
GGSL	*compl*	3.7	**1.4**	84.1	**90.4**	92.5	**94.1**
	F1	1.8	**1.2**	88.8	**91.0**	92.6	**93.9**
	pur	5.0	**3.2**	70.5	**76.1**	81.8	**85.6**
NGGSL	*compl*	5.8	**5.4**	**79.7**	77.6	**90.7**	**90.7**
	F1	1.8	**1.2**	88.8	**91.0**	92.6	**93.9**

Table B.2: Comparison between VGG and SC-VGG models, by evaluating performance fluctuations over the $k = 10$ folds. Q_1 and Q_3 are the first and third quartile (i.e. delimit 25th and 75th percentile). The inter-quartile range (IQR) is the difference $Q_3 - Q_1$, while the range ($Q_1 - 1.5 \cdot IQR$, $Q_3 + 1.5 \cdot IQR$) delimits the data within $\pm 2.698\sigma$. Average efficiency and GGSL estimators are graphically shown in the bottom panels of Fig. 6.8. Best results are highlighted in bold.

		VGG		SC-VGG	
	GGSL	FN	FN/TP	FN	FN/TP
Total Number	2704	307	0.128	224	0.090
$F814 \geq 28.0$	31.1%	52.1%	0.235	38.8%	0.115
$F814 \geq 27.0$	61.0%	83.1%	0.183	73.2%	0.110
$F814 < 27.0$	39.0%	17.9%	0.052	26.8%	0.060
$\theta_E < 0.5''$	32.2%	41.4%	0.171	46.9%	0.137
$\theta_E \geq 0.5''$	67.8%	58.6%	0.109	53.1%	0.069
$z_{src} \geq 5$	5.9%	11.4%	0.282	4.9%	0.074
$z_{src} \geq 4$	12.9%	20.8%	0.225	10.7%	0.074
$z_{src} \geq 3$	25.5%	31.9%	0.166	20.1%	0.070
$z_{src} < 3$	74.5%	68.1%	0.116	70.9%	0.098
		VGG		SC-VGG	
	NGGSL	FP	FP/TN	FP	FP/TN
Total Number	1037	154	0.174	170	0.196
$F814 < 19.5$	9.6%	16.8%	0.263	15.3%	0.356
$F814 \geq 19.5$	90.4%	83.2%	0.158	84.7%	0.181
$(F606 - F814)_{\text{norm}} < -0.5$	3.9%	6.5%	0.333	7.6%	0.481
$(F606 - F814)_{\text{norm}} \geq -0.5$	96.1%	93.5%	0.169	92.3%	0.187

Table B.3: Summary of False Positive and False Negative distributions. Fractions of GGSL (Col. 2), False Negative (FN) (Col. 3) and FN to GGSL ratio (Col. 4) as a function of source magnitude (second to fourth row), galaxy-lens θ_E (fifth and sixth row) and source redshift (seventh to eighth row). The total number of spectroscopic GGLSs and FNs are quoted in the first row. Similar fractions for NGGSLs, FP (False Positive) and FP/NGGSL ratio are quoted in the bottom half of the table. Classification metrics are split between VGG and SC-VGG network.

		EXP-A (~ 4″ cutouts)				
Class	*%*	VGG	SC-VGG	*F*435	*F*606	*F*814
	AE	87.7	89.4	87.2	86.1	86.8
GGSL	*pur*	93.4	93.1	91.8	91.1	91.5
	compl	88.6	91.7	89.8	88.8	89.3
	F1	91.0	92.4	90.8	89.9	90.4
NGGSL	*pur*	76.7	81.4	77.5	75.4	76.6
	compl	85.4	84.1	81.3	79.9	80.9
	F1	81.1	82.8	79.3	77.6	78.7

Table B.4: Performance comparison by also including network trained with a single band, related to the *EXP-A* cutout configuration (see section. 6.2) and expressed in terms of the statistical estimators described in Section 3.6.6. Performances are compared with the ones carried out by the VGG and the SC-VGG networks.

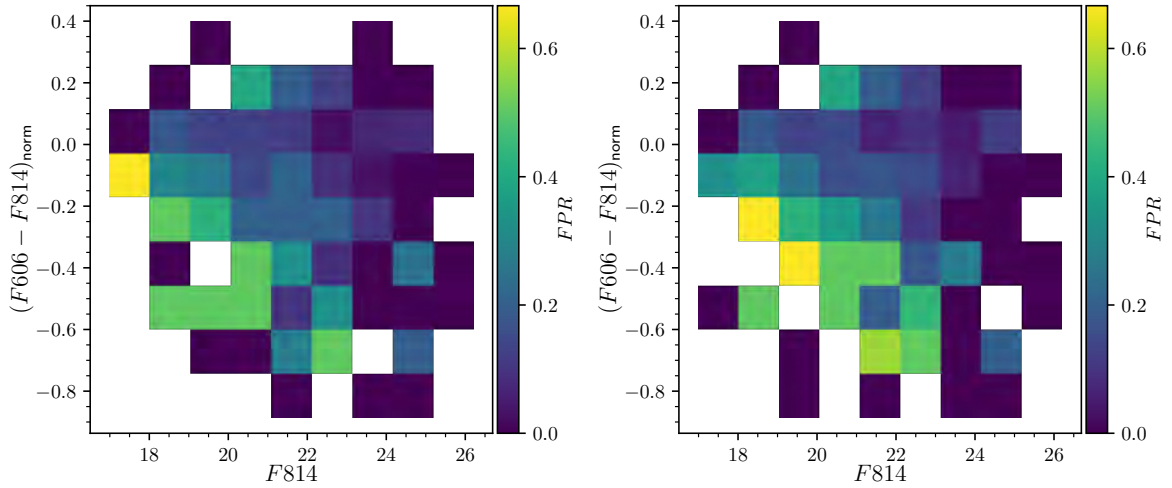


Figure B.2: False Positive to True Negative ratios as a function of normalised colour $(F606 - F814)_{\mathrm{norm}}$ and $F814$ magnitude related to VGG (left panel) and SC-VGG (right panel) performances. Parts of the parameter space with zero True Negative are left white.

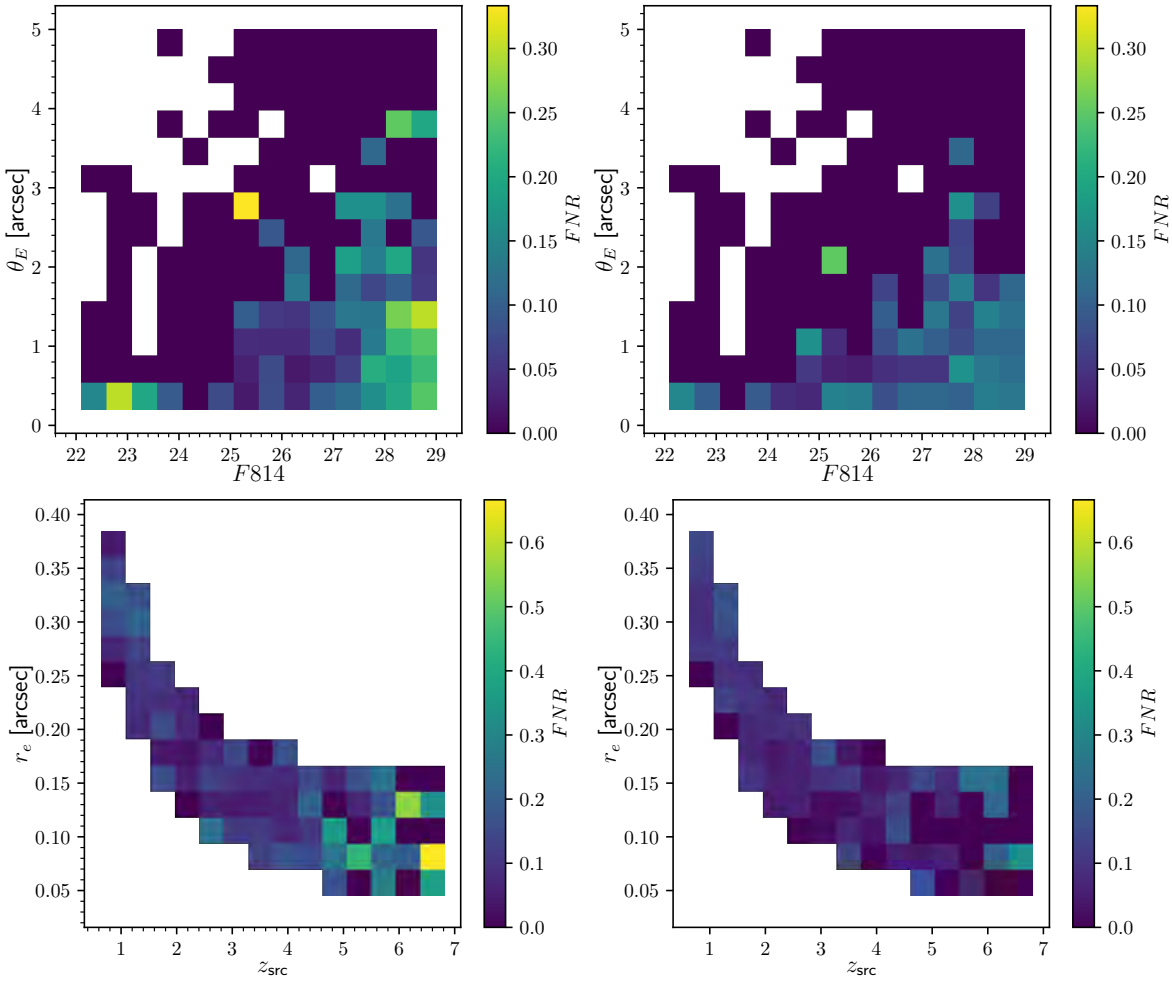


Figure B.3: False Negative Ratio (FNR) as a function of: (*i*) galaxy-lens size θ_E and source $F814$ magnitude (top panels), (*ii*) source effective radius r_e and redshift z_{src} (bottom panels). Performances are related to VGG (left panels) and SC-VGG (right panels). Parts of the parameter space with zero True Positive are left white.

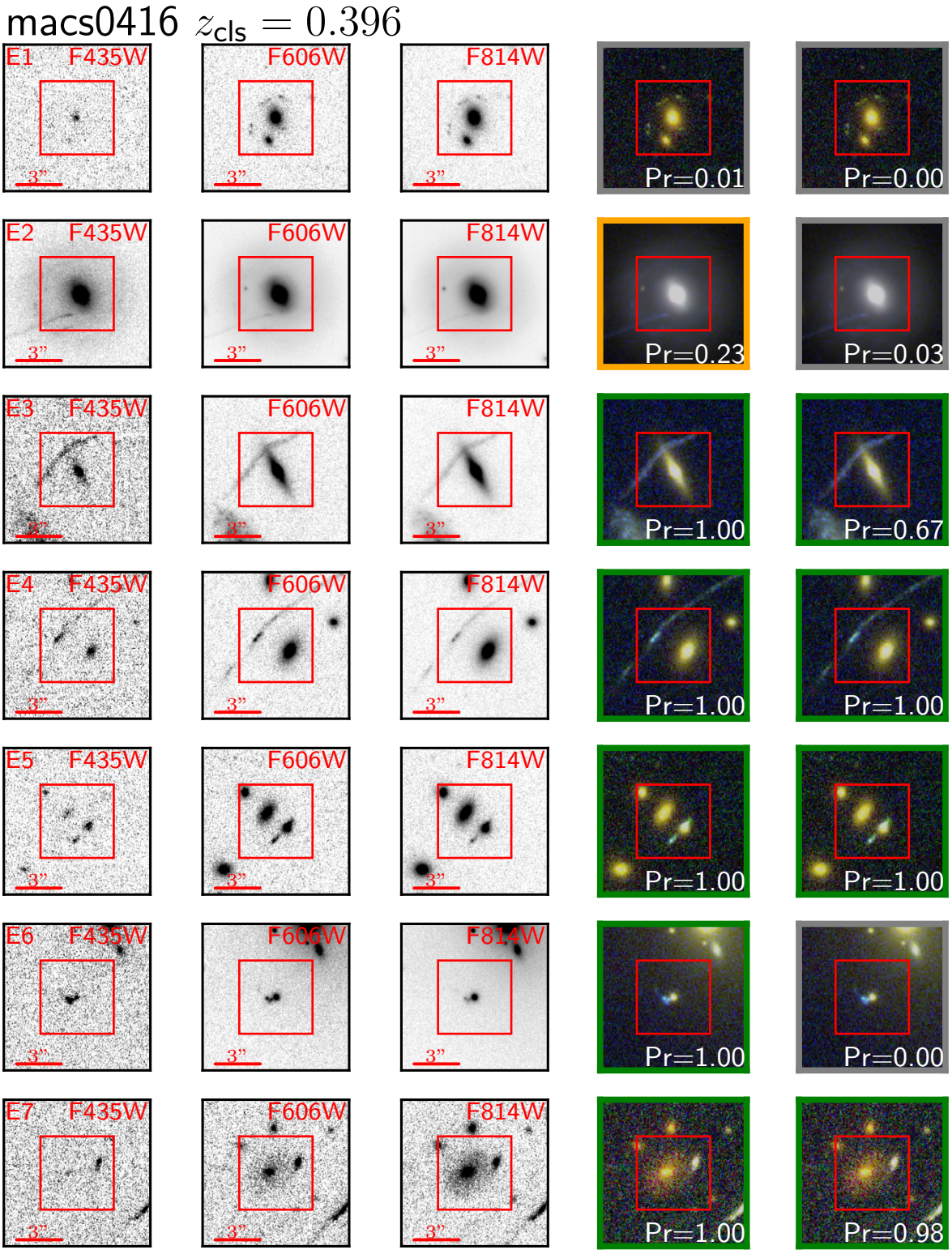


Figure B.4: Continuing of Fig. 6.15, related to M0416 galaxy cluster.

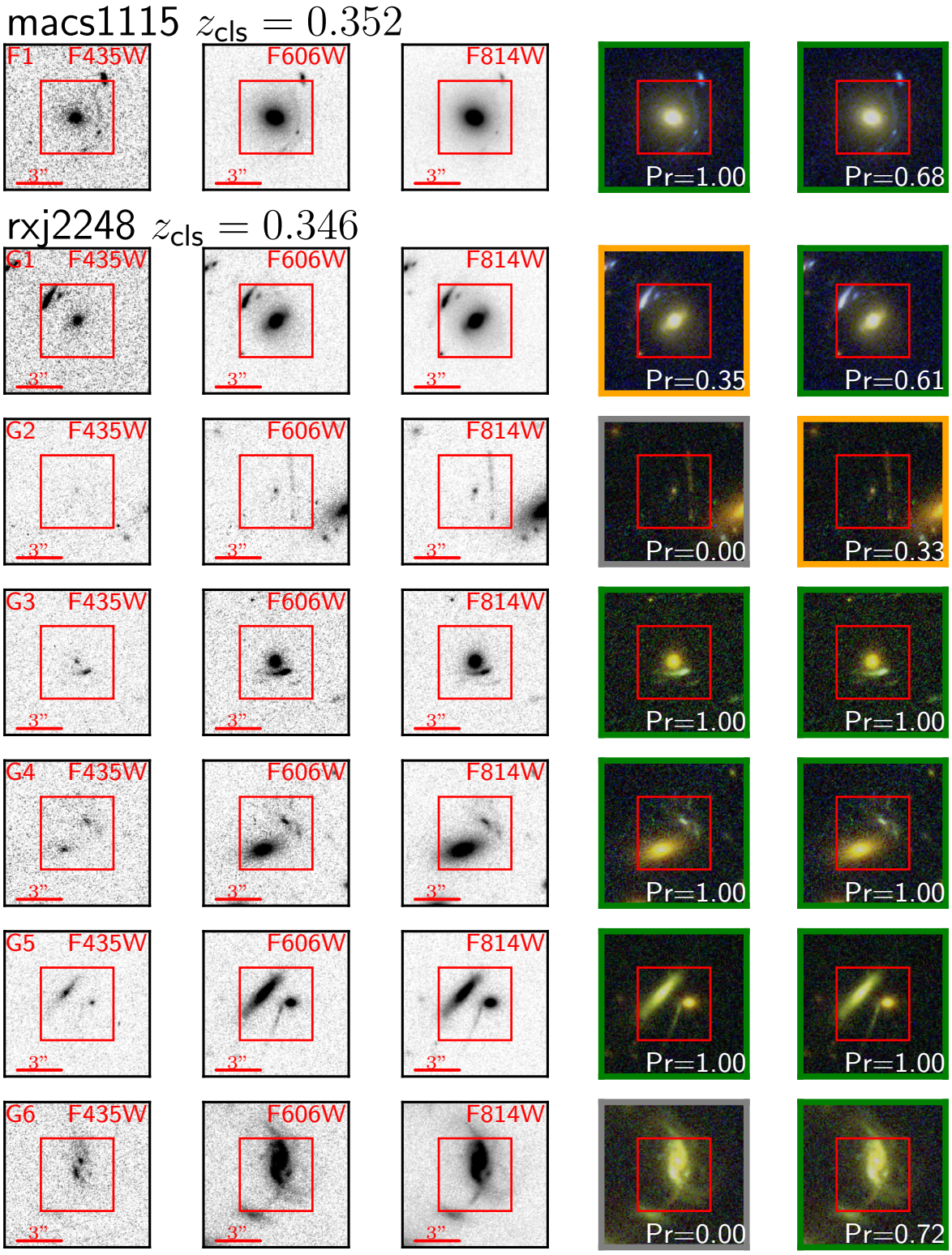


Figure B.5: Continuing of Fig. B.4, related to M1115, R2248 galaxy clusters.

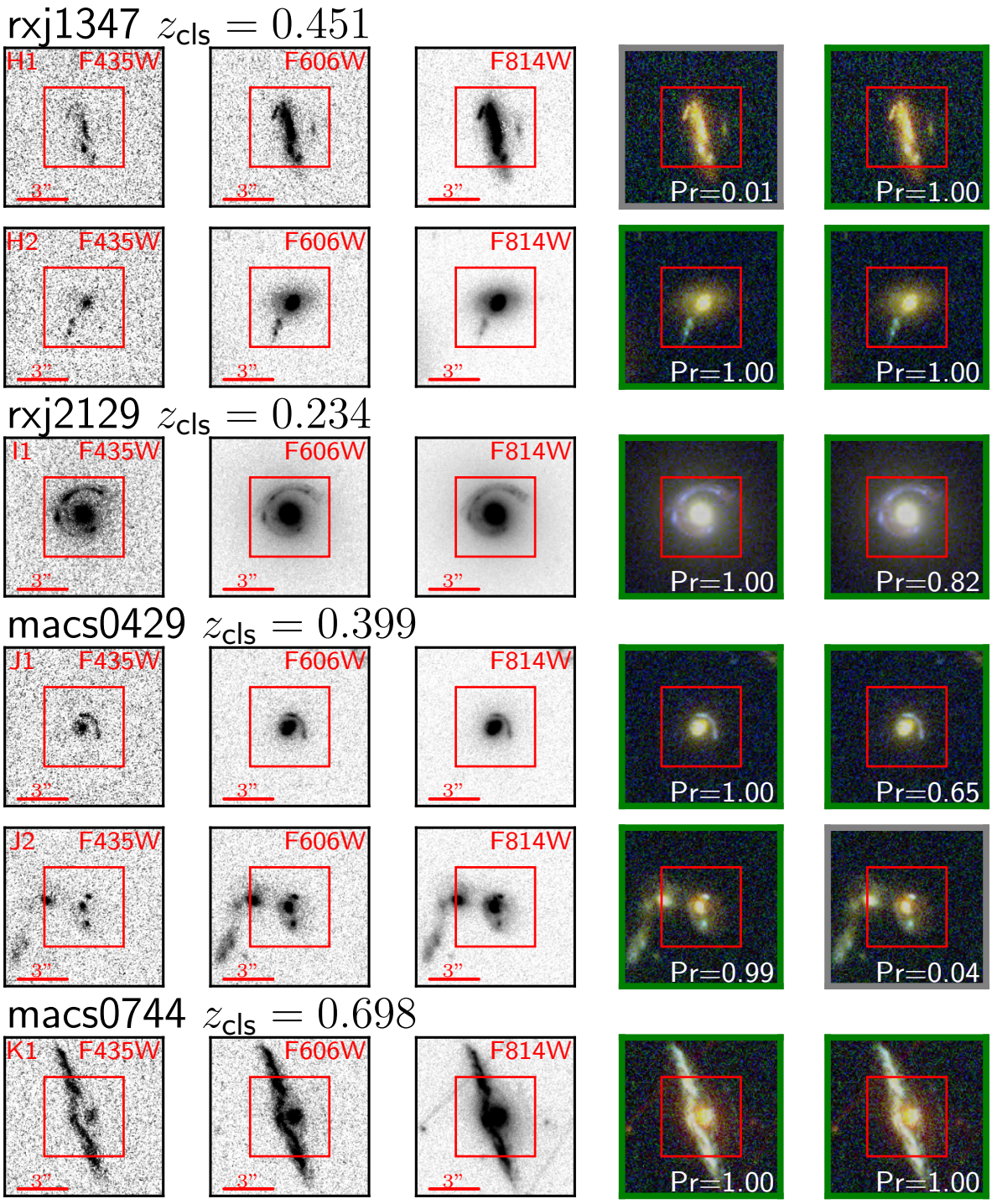


Figure B.6: Continuing of Fig. B.4, related to R1347, R2129, M0429, M0744 galaxy clusters.

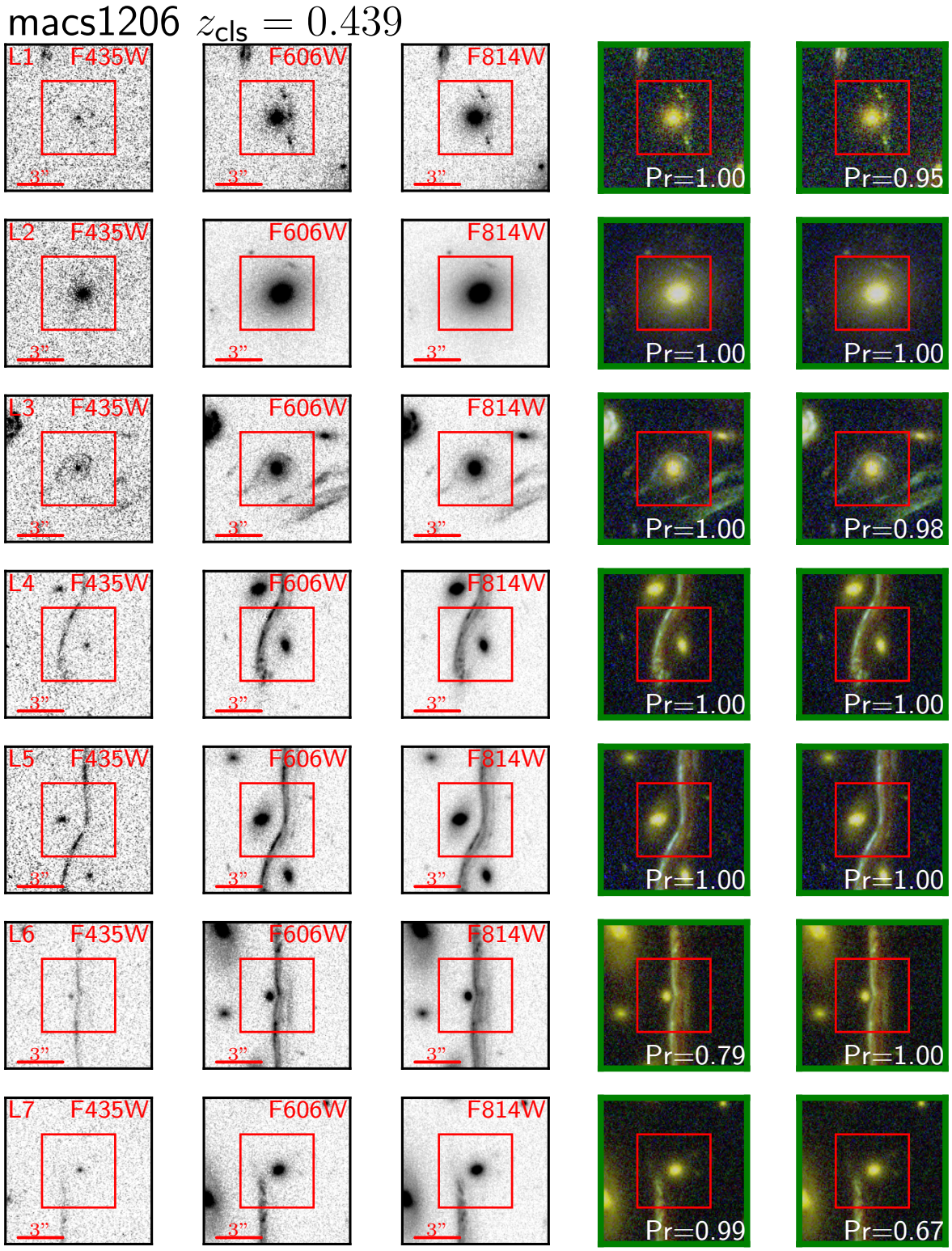


Figure B.7: Continuing of Fig. B.6, related to M1206 galaxy clusters.

RA	DEC	Cluster	Pr_{VGG}	$Pr_{SC\text{-}VGG}$	ref	Fig.	RA	DEC	Cluster	Pr_{VGG}	$Pr_{SC\text{-}VGG}$	ref	Fig.
181.549205	-8.810293	M1206	0.001	0.549	A1	6.16	181.569892	-8.797327	M1206	0.383	0.205	A65	6.16
181.555360	-8.804530	M1206	0.000	0.224	A2	6.16	181.566006	-8.797562	M1206	0.001	0.330	A66	6.16
181.536011	-8.807485	M1206	0.290	0.567	A3	6.16	181.549147	-8.797627	M1206	0.209	0.234	A67	6.16
181.533720	-8.806617	M1206	0.513	0.023	A4	6.16	181.547794	-8.794433	M1206	1.000	0.655	A68	6.16
181.557092	-8.806018	M1206	0.012	0.206	A5	6.16	181.564275	-8.802208	M1206	1.000	0.006	A69	6.16
181.520680	-8.822387	M1206	1.000	0.711	A6	6.16	181.559851	-8.801577	M1206	0.962	0.313	A70	6.16
181.543662	-8.790532	M1206	0.002	0.621	A7	6.16	181.541915	-8.799983	M1206	0.000	0.331	A71	6.16
181.552620	-8.814487	M1206	0.347	0.339	A8	6.16	181.552532	-8.796535	M1206	0.256	0.399	A72	6.16
181.559438	-8.790606	M1206	0.958	0.696	A9	6.16	181.542348	-8.797811	M1206	0.002	0.653	A73	6.16
181.519260	-8.810741	M1206	0.740	0.459	A10	6.16	342.185399	-44.518624	R2248	0.000	0.204	B1	6.16
181.523813	-8.795791	M1206	0.007	0.201	A11	6.16	342.161400	-44.555896	R2248	0.999	0.003	B2	6.16
181.536463	-8.792662	M1206	0.988	0.792	A12	6.16	342.141917	-44.525964	R2248	0.995	0.959	B3	6.16
181.525265	-8.789044	M1206	0.232	0.313	A13	6.16	342.148934	-44.523247	R2248	0.000	0.339	B4	6.16
181.549753	-8.793907	M1206	0.946	0.366	A14	6.16	342.184490	-44.543188	R2248	0.000	0.322	B5	6.16
181.547707	-8.810972	M1206	0.761	0.028	A15	6.16	342.203796	-44.542198	R2248	0.002	0.978	B6	6.16
181.538071	-8.811822	M1206	0.001	0.741	A16	6.16	342.210759	-44.540675	R2248	0.001	0.649	B7	6.16
181.553472	-8.809823	M1206	0.013	0.240	A17	6.16	342.178906	-44.524710	R2248	0.000	0.328	B8	6.16
181.551389	-8.792406	M1206	0.339	0.002	A18	6.16	342.184434	-44.517662	R2248	0.210	0.001	B9	6.16
181.550373	-8.792018	M1206	0.000	0.334	A19	6.16	342.152487	-44.506223	R2248	0.000	0.294	B10	6.16
181.561787	-8.808694	M1206	0.016	0.601	A20	6.16	342.174439	-44.546221	R2248	0.000	0.276	B11	6.16
181.556131	-8.797638	M1206	0.748	1.000	A21	6.16	342.204449	-44.528966	R2248	0.270	0.336	B12	6.16
181.549053	-8.797094	M1206	0.908	0.555	A22	6.16	342.152796	-44.554330	R2248	0.064	0.403	B13	6.16
181.544433	-8.801504	M1206	1.000	0.778	A23	6.16	342.204135	-44.537097	R2248	0.512	0.316	B14	6.16
181.561526	-8.808836	M1206	0.002	0.231	A24	6.16	342.201799	-44.563313	R2248	0.165	0.543	B15	6.16
181.546796	-8.795032	M1206	0.133	0.335	A25	6.16	342.223745	-44.533249	R2248	0.305	0.492	B16	6.16
181.556286	-8.797781	M1206	0.995	0.821	A26	6.16	342.191888	-44.529672	R2248	0.702	1.000	B17	6.16
181.541007	-8.792347	M1206	1.000	0.023	A27	6.16	342.173658	-44.532778	R2248	0.491	0.000	B18	6.16
181.562070	-8.804915	M1206	0.032	0.906	A28	6.16	342.215863	-44.518431	R2248	0.998	1.000	B19	6.16
181.565468	-8.793209	M1206	0.759	0.667	A29	6.16	342.175508	-44.535471	R2248	0.882	0.399	B20	6.16
181.565257	-8.806081	M1206	0.985	0.924	A30	6.16	342.211458	-44.525615	R2248	0.000	0.434	B21	6.16
181.550620	-8.800939	M1206	0.976	0.002	A31	6.16	342.155753	-44.545904	R2248	0.397	0.380	B22	6.16
181.539516	-8.791749	M1206	0.867	0.799	A32	6.16	342.154672	-44.539330	R2248	0.248	0.507	B23	6.16
181.527058	-8.780598	M1206	0.000	0.273	A33	6.16	342.161945	-44.529975	R2248	0.001	0.892	B24	6.16
181.540476	-8.788438	M1206	0.984	0.683	A34	6.16	342.215524	-44.519505	R2248	0.562	0.564	B25	6.16
181.561870	-8.804334	M1206	0.322	0.003	A35	6.16	342.142738	-44.531344	R2248	1.000	0.666	B26	6.16
181.565419	-8.795550	M1206	0.013	0.838	A36	6.16	342.166626	-44.534812	R2248	0.017	0.419	B27	6.16
181.530345	-8.785732	M1206	0.458	0.511	A37	6.16	342.182025	-44.540375	R2248	1.000	0.989	B28	6.16
181.553034	-8.780057	M1206	1.000	0.072	A38	6.16	342.162678	-44.538171	R2248	1.000	0.400	B29	6.16
181.567840	-8.800822	M1206	0.231	0.415	A39	6.16	342.192743	-44.519772	R2248	0.005	0.237	B30	6.16
181.545168	-8.801791	M1206	1.000	0.971	A40	6.16	342.170862	-44.506796	R2248	0.999	0.021	B31	6.16
181.553098	-8.794855	M1206	1.000	0.977	A41	6.16	64.041046	-24.069426	M0416	0.013	0.307	C1	B.9
181.545643	-8.786723	M1206	0.924	0.666	A42	6.16	64.029739	-24.083373	M0416	0.955	0.338	C2	B.9
181.561485	-8.804095	M1206	0.101	0.352	A43	6.16	64.044409	-24.067011	M0416	0.881	0.201	C3	B.9
181.528202	-8.786448	M1206	0.002	0.329	A44	6.16	64.029518	-24.079546	M0416	0.695	0.000	C4	B.9
181.573679	-8.802283	M1206	0.000	0.697	A45	6.16	64.044856	-24.073521	M0416	0.667	0.772	C5	B.9
181.544327	-8.815659	M1206	1.000	0.912	A46	6.16	64.055655	-24.060280	M0416	0.819	0.722	C6	B.9
181.553805	-8.800057	M1206	0.069	0.336	A47	6.16	64.057109	-24.051496	M0416	0.001	0.385	C7	B.9
181.544916	-8.800652	M1206	0.992	0.999	A48	6.16	64.038517	-24.062071	M0416	0.318	0.263	C8	B.9
181.541230	-8.792218	M1206	0.544	1.000	A49	6.16	64.017648	-24.090606	M0416	0.297	0.631	C9	B.9
181.566661	-8.804784	M1206	1.000	0.993	A50	6.16	64.048703	-24.064571	M0416	0.000	0.246	C10	B.9
181.572924	-8.787941	M1206	0.970	0.713	A51	6.16	64.050218	-24.057200	M0416	0.884	0.329	C11	B.9
181.530203	-8.809179	M1206	1.000	0.323	A52	6.16	64.020895	-24.073343	M0416	0.099	0.824	C12	B.9
181.544783	-8.802598	M1206	1.000	0.657	A53	6.16	64.022290	-24.072602	M0416	0.157	0.389	C13	B.9
181.567096	-8.803197	M1206	1.000	1.000	A54	6.16	64.027765	-24.061003	M0416	0.003	0.668	C14	B.9
181.539554	-8.816781	M1206	0.999	1.000	A55	6.16	64.033839	-24.080561	M0416	0.913	0.912	C15	B.9
181.524215	-8.785849	M1206	0.803	0.384	A56	6.16	64.025874	-24.075780	M0416	0.821	0.663	C16	B.9
181.545579	-8.800887	M1206	0.836	0.030	A57	6.16	64.059507	-24.054116	M0416	0.943	0.333	C17	B.9
181.563979	-8.795352	M1206	1.000	0.990	A58	6.16	64.032249	-24.065751	M0416	0.488	0.283	C18	B.9
181.556528	-8.791484	M1206	0.871	0.000	A59	6.16	64.042450	-24.063194	M0416	1.000	0.712	C19	B.9
181.537685	-8.805377	M1206	1.000	1.000	A60	6.16	64.032458	-24.068488	M0416	1.000	0.722	C20	B.9
181.547738	-8.802707	M1206	0.917	0.999	A61	6.16	64.033000	-24.074319	M0416	0.000	0.333	C21	B.9
181.568355	-8.798285	M1206	0.998	0.246	A62	6.16	64.030996	-24.078539	M0416	1.000	0.000	C22	B.9
181.552373	-8.794385	M1206	0.914	0.871	A63	6.16	64.025719	-24.084333	M0416	0.908	0.002	C23	B.9
181.562071	-8.802550	M1206	0.018	1.000	A64	6.16	64.041305	-24.071338	M0416	1.000	1.000	C24	B.9

Table B.5: List of candidate GGSLs identified by both models, with Pr(VGG) $\vee$ Pr(SC-VGG) ≥ 0.2.

RA	DEC	Cluster	Pr_{VGG}	$Pr_{SC\text{-}VGG}$	ref	Fig.	RA	DEC	Cluster	Pr_{VGG}	$Pr_{SC\text{-}VGG}$	ref	Fig.
64.017085	-24.089554	M0416	1.000	0.673	C25	B.9	168.980906	1.498636	M1115	0.000	0.315	E17	B.9
64.033445	-24.069097	M0416	0.979	0.256	C26	B.9	168.960956	1.514954	M1115	0.974	0.000	E18	B.9
64.032755	-24.068255	M0416	0.999	0.334	C27	B.9	292.964810	-26.580178	M1931	0.363	0.221	F1	B.9
64.042850	-24.062220	M0416	0.263	0.350	C28	B.9	292.973015	-26.557051	M1931	0.843	0.338	F2	B.9
64.032505	-24.078488	M0416	1.000	0.000	C29	B.9	292.941558	-26.599137	M1931	0.836	0.051	F3	B.9
64.050881	-24.060269	M0416	0.987	0.944	C30	B.9	292.947821	-26.576614	M1931	0.242	0.008	F4	B.9
64.024525	-24.081041	M0416	0.000	0.665	C31	B.9	292.963189	-26.571103	M1931	0.000	0.407	F5	B.9
64.041830	-24.062824	M0416	0.071	0.434	C32	B.9	292.956828	-26.575842	M1931	0.083	0.857	F6	B.9
64.050419	-24.061315	M0416	0.468	0.000	C33	B.9	292.946030	-26.568565	M1931	0.953	0.669	F7	B.9
64.046889	-24.075730	M0416	0.841	0.729	C34	B.9	292.949739	-26.582414	M1931	1.000	0.824	F8	B.9
64.023351	-24.076460	M0416	0.000	0.548	C35	B.9	292.950788	-26.577951	M1931	1.000	1.000	F9	B.9
64.024450	-24.080421	M0416	0.019	0.334	C36	B.9	292.954051	-26.584634	M1931	0.023	0.805	F10	B.9
64.047569	-24.069879	M0416	0.023	0.321	C37	B.9	292.962925	-26.572257	M1931	0.990	0.975	F11	B.9
64.027852	-24.079065	M0416	0.038	0.338	C38	B.9	292.966129	-26.580228	M1931	0.634	0.956	F12	B.9
39.973218	-1.599127	A370	0.000	0.666	D1	B.9	52.434135	-2.193304	M0329	0.985	0.908	G1	B.8
39.975147	-1.576869	A370	1.000	0.005	D2	B.9	52.435281	-2.196621	M0329	0.995	0.992	G2	B.8
39.962614	-1.562538	A370	0.147	0.263	D3	B.9	52.440005	-2.219678	M0329	0.000	0.667	G3	B.8
39.964630	-1.580285	A370	0.000	0.253	D4	B.9	52.394249	-2.209243	M0329	0.061	0.586	G4	B.8
39.981797	-1.570870	A370	0.008	0.474	D5	B.9	52.429369	-2.205997	M0329	0.008	0.615	G5	B.8
39.964958	-1.570948	A370	0.003	0.690	D6	B.9	52.405259	-2.218065	M0329	0.345	0.345	G6	B.8
39.980065	-1.582286	A370	0.803	0.302	D7	B.9	52.431373	-2.193189	M0329	0.000	0.335	G7	B.8
39.972213	-1.580356	A370	0.948	0.994	D8	B.9	52.426689	-2.205520	M0329	0.006	0.516	G8	B.8
39.978672	-1.561884	A370	0.005	0.336	D9	B.9	52.423219	-2.196231	M0329	1.000	0.463	G9	B.8
39.962863	-1.578398	A370	0.497	0.000	D10	B.9	52.426735	-2.207306	M0329	0.020	0.548	G10	B.8
39.981980	-1.571269	A370	0.009	0.718	D11	B.9	52.405343	-2.184214	M0329	0.298	0.676	G11	B.8
39.972699	-1.586730	A370	0.002	0.349	D12	B.9	52.428625	-2.193656	M0329	1.000	0.358	G12	B.8
39.971816	-1.583978	A370	0.767	0.000	D13	B.9	52.420547	-2.191693	M0329	1.000	0.610	G13	B.8
39.975173	-1.597391	A370	0.900	0.000	D14	B.9	52.414471	-2.202243	M0329	0.999	0.681	G14	B.8
39.963190	-1.565781	A370	1.000	0.662	D15	B.9	52.424938	-2.172465	M0329	0.393	0.012	G15	B.8
39.970596	-1.569273	A370	0.001	0.333	D16	B.9	52.426019	-2.188714	M0329	0.994	0.932	G16	B.8
39.972445	-1.584585	A370	0.998	0.380	D17	B.9	52.429977	-2.194287	M0329	0.448	0.001	G17	B.8
39.949592	-1.581789	A370	1.000	0.017	D18	B.9	52.438711	-2.199283	M0329	1.000	0.646	G18	B.8
39.959741	-1.573407	A370	0.851	0.802	D19	B.9	52.412965	-2.208764	M0329	0.208	0.017	G19	B.8
39.971814	-1.574785	A370	0.931	0.205	D20	B.9	52.426923	-2.211980	M0329	0.536	0.333	G20	B.8
39.963605	-1.553493	A370	0.379	0.090	D21	B.9	52.444865	-2.185521	M0329	0.999	1.000	G21	B.8
39.972376	-1.584299	A370	0.033	0.333	D22	B.9	52.430995	-2.195969	M0329	0.000	0.585	G22	B.8
39.964048	-1.566449	A370	0.000	0.463	D23	B.9	177.417479	22.399804	M1149	0.780	0.053	H1	B.8
39.970894	-1.584598	A370	0.011	0.401	D24	B.9	177.403577	22.396382	M1149	0.016	0.233	H2	B.8
39.967938	-1.584447	A370	1.000	0.014	D25	B.9	177.381690	22.396987	M1149	0.006	0.201	H3	B.8
39.962167	-1.562983	A370	0.295	0.288	D26	B.9	177.399049	22.422468	M1149	0.956	0.000	H4	B.8
39.961728	-1.560650	A370	0.004	0.503	D27	B.9	177.410707	22.385895	M1149	0.004	0.295	H5	B.8
39.955672	-1.576411	A370	0.951	0.010	D28	B.9	177.391681	22.390617	M1149	0.003	0.417	H6	B.8
39.985276	-1.580475	A370	0.962	0.841	D29	B.9	177.390966	22.401682	M1149	0.001	0.332	H7	B.8
39.978620	-1.591035	A370	1.000	0.667	D30	B.9	177.399004	22.418394	M1149	0.917	0.005	H8	B.8
39.970143	-1.580750	A370	1.000	1.000	D31	B.9	177.390152	22.403889	M1149	0.515	0.191	H9	B.8
39.974087	-1.594507	A370	0.195	0.326	D32	B.9	177.409508	22.392842	M1149	0.997	0.067	H10	B.8
39.968082	-1.577159	A370	0.000	0.331	D33	B.9	177.412881	22.393652	M1149	0.000	0.380	H11	B.8
39.973630	-1.559746	A370	0.038	0.992	D34	B.9	177.381092	22.401135	M1149	0.993	0.105	H12	B.8
168.960621	1.527189	M1115	0.006	0.587	E1	B.9	177.398745	22.398530	M1149	0.845	0.005	H13	B.8
168.944813	1.477052	M1115	0.213	0.669	E2	B.9	177.384125	22.393850	M1149	0.016	0.331	H14	B.8
168.961762	1.508663	M1115	0.601	0.221	E3	B.9	177.399824	22.397257	M1149	1.000	0.349	H15	B.8
168.956298	1.483837	M1115	0.001	0.257	E4	B.9	177.397787	22.395452	M1149	1.000	0.332	H16	B.8
168.957402	1.523481	M1115	0.000	0.327	E5	B.9	177.403692	22.389108	M1149	0.003	0.338	H17	B.8
168.954699	1.517676	M1115	1.000	0.333	E6	B.9	177.392881	22.397102	M1149	0.842	0.572	H18	B.8
168.966242	1.498636	M1115	1.000	0.472	E7	B.9	177.387703	22.393751	M1149	0.990	0.558	H19	B.8
168.956259	1.497409	M1115	1.000	0.613	E8	B.9	177.398603	22.398080	M1149	0.998	0.279	H20	B.8
168.978360	1.504300	M1115	1.000	0.698	E9	B.9	177.386621	22.413649	M1149	0.916	0.610	H21	B.8
168.952017	1.510496	M1115	0.894	0.140	E10	B.9	177.394534	22.400634	M1149	0.940	0.735	H22	B.8
168.980347	1.509714	M1115	0.996	0.362	E11	B.9	177.398456	22.405363	M1149	1.000	0.311	H23	B.8
168.968954	1.472671	M1115	1.000	0.678	E12	B.9	177.386380	22.413708	M1149	0.010	0.699	H24	B.8
168.960031	1.495261	M1115	0.000	0.207	E13	B.9	177.392878	22.418198	M1149	0.985	0.028	H25	B.8
168.992819	1.487158	M1115	0.991	0.629	E14	B.9	177.402882	22.402009	M1149	0.004	0.524	H26	B.8
168.978868	1.480593	M1115	0.942	0.266	E15	B.9	177.395075	22.389852	M1149	0.993	0.362	H27	B.8
168.949297	1.511959	M1115	0.999	0.453	E16	B.9	177.404015	22.402132	M1149	0.999	0.335	H28	B.8

Table B.6: Continuing of Tab. B.5.

RA	DEC	Cluster	Pr_{VGG}	$Pr_{SC\text{-}VGG}$	ref	Fig.	RA	DEC	Cluster	Pr_{VGG}	$Pr_{SC\text{-}VGG}$	ref	Fig.
177.382598	22.416507	M1149	1.000	1.000	H29	B.8	322.420801	0.103281	R2129	0.010	0.383	M6	B.10
177.402162	22.382830	M1149	0.000	0.643	H30	B.8	322.417389	0.094434	R2129	0.501	0.382	M7	B.10
41.997696	-3.555797	A383	0.026	0.698	I1	B.8	322.410915	0.092051	R2129	0.392	0.733	M8	B.10
42.004492	-3.550612	A383	0.046	0.288	I2	B.8	322.411309	0.096838	R2129	0.804	0.650	M9	B.10
42.020755	-3.552819	A383	0.992	0.658	I3	B.8	322.415970	0.097734	R2129	0.996	0.680	M10	B.10
42.005684	-3.501878	A383	0.007	0.597	I4	B.8	322.420880	0.098992	R2129	0.998	0.937	M11	B.10
42.014319	-3.559952	A383	0.951	1.000	I5	B.8	322.402667	0.070389	R2129	1.000	0.164	M12	B.10
42.024630	-3.546054	A383	0.003	0.353	I6	B.8	322.398374	0.062152	R2129	0.005	0.245	M13	B.10
42.040794	-3.540184	A383	0.407	0.048	I7	B.8	322.428787	0.108113	R2129	1.000	0.909	M14	B.10
42.002519	-3.534629	A383	0.000	0.439	I8	B.8	322.439126	0.078839	R2129	0.002	0.309	M15	B.10
42.015529	-3.511954	A383	1.000	1.000	I9	B.8	322.437007	0.119642	R2129	0.035	0.648	M16	B.10
42.010044	-3.533805	A383	0.000	0.327	I10	B.8	322.417106	0.083945	R2129	0.935	0.997	M17	B.10
42.020457	-3.522868	A383	0.151	0.811	I11	B.8	322.410033	0.081391	R2129	0.915	0.338	M18	B.10
42.008426	-3.529415	A383	1.000	1.000	I12	B.8	322.420240	0.098401	R2129	0.998	1.000	M19	B.10
42.013228	-3.535131	A383	0.613	0.065	I13	B.8	206.908966	-11.745779	R1347	0.001	0.568	N1	B.10
42.022289	-3.513336	A383	0.001	0.328	I14	B.8	206.883957	-11.731207	R1347	0.984	0.384	N2	B.10
22.963918	-13.615755	A209	0.000	0.243	J1	B.8	206.877701	-11.749572	R1347	0.445	0.298	N3	B.10
22.983180	-13.611237	A209	1.000	0.993	J2	B.8	206.873688	-11.733620	R1347	0.823	0.662	N4	B.10
22.977588	-13.617790	A209	0.214	0.002	J3	B.8	206.879166	-11.761577	R1347	0.389	0.010	N5	B.10
22.996970	-13.620002	A209	0.451	0.011	J4	B.8	206.883957	-11.773227	R1347	0.678	0.067	N6	B.10
22.959937	-13.598104	A209	0.236	0.946	J5	B.8	206.887988	-11.755872	R1347	0.985	0.997	N7	B.10
22.967728	-13.603752	A209	0.245	0.000	J6	B.8	206.884064	-11.760151	R1347	0.645	0.502	N8	B.10
22.988337	-13.590658	A209	0.012	0.279	J7	B.8	206.890745	-11.754540	R1347	0.987	0.843	N9	B.10
22.962118	-13.625705	A209	0.013	0.218	J8	B.8	206.877594	-11.752647	R1347	1.000	0.398	N10	B.10
22.997745	-13.596058	A209	0.868	0.066	J9	B.8	206.894897	-11.750385	R1347	1.000	0.295	N11	B.10
22.964874	-13.585267	A209	0.010	0.215	J10	B.8	206.886490	-11.770810	R1347	1.000	0.812	N12	B.10
22.962026	-13.601133	A209	0.000	0.297	J11	B.8	206.855698	-11.758474	R1347	0.999	0.197	N13	B.10
22.984309	-13.612992	A209	0.349	0.270	J12	B.8	206.878204	-11.774970	R1347	1.000	0.086	N14	B.10
22.957748	-13.603227	A209	0.988	0.367	J13	B.8	206.892150	-11.747934	R1347	1.000	1.000	N15	B.10
22.957086	-13.603233	A209	0.023	0.666	J14	B.8	206.884262	-11.745551	R1347	0.070	0.351	N16	B.10
197.763155	-3.184823	M1311	0.004	0.313	K1	B.8	206.882604	-11.740795	R1347	0.047	0.367	N17	B.10
197.739113	-3.174417	M1311	0.060	0.318	K2	B.8	206.893573	-11.750476	R1347	1.000	0.838	N18	B.10
197.770163	-3.186733	M1311	0.383	0.364	K3	B.8	3.585191	-30.394666	A2744	0.000	0.315	O1	B.10
197.751655	-3.192362	M1311	0.061	0.432	K4	B.8	3.581389	-30.393933	A2744	0.991	0.676	O2	B.10
197.763034	-3.175674	M1311	0.000	0.522	K5	B.8	3.578947	-30.394119	A2744	0.000	0.426	O3	B.10
197.741295	-3.160086	M1311	0.000	0.332	K6	B.8	3.604396	-30.384960	A2744	0.064	0.244	O4	B.10
197.761433	-3.173610	M1311	0.085	1.000	K7	B.8	3.578347	-30.389466	A2744	0.000	0.648	O5	B.10
197.769542	-3.172347	M1311	0.904	0.351	K8	B.8	3.603449	-30.416761	A2744	0.234	0.001	O6	B.10
322.350964	-7.697149	M2129	0.003	0.411	L1	B.10	3.578733	-30.384224	A2744	1.000	0.918	O7	B.10
322.370408	-7.692092	M2129	0.073	0.240	L2	B.10	3.593028	-30.382969	A2744	0.095	0.657	O8	B.10
322.356745	-7.692505	M2129	0.031	0.667	L3	B.10	3.586835	-30.384628	A2744	0.893	0.331	O9	B.10
322.351922	-7.689095	M2129	0.876	0.479	L4	B.10	3.571033	-30.409597	A2744	0.001	0.626	O10	B.10
322.353041	-7.666496	M2129	0.969	0.968	L5	B.10	3.579079	-30.400092	A2744	0.183	0.262	O11	B.10
322.346174	-7.693583	M2129	0.534	0.519	L6	B.10	3.582929	-30.399708	A2744	0.239	0.997	O12	B.10
322.357111	-7.689092	M2129	0.354	0.332	L7	B.10	3.601247	-30.404884	A2744	0.000	0.334	O13	B.10
322.348909	-7.690223	M2129	1.000	0.465	L8	B.10	3.608083	-30.387064	A2744	0.690	0.314	O14	B.10
322.349711	-7.692455	M2129	0.754	0.923	L9	B.10	3.598171	-30.404825	A2744	0.997	0.003	O15	B.10
322.354946	-7.696953	M2129	0.780	0.002	L10	B.10	3.581606	-30.399107	A2744	1.000	0.998	O16	B.10
322.364550	-7.684812	M2129	0.999	0.364	L11	B.10	3.591216	-30.386703	A2744	0.013	0.618	O17	B.10
322.358791	-7.691052	M2129	0.010	0.458	L12	B.10	325.063367	-23.640846	MS2137	0.004	0.268	P1	B.10
322.354951	-7.691770	M2129	1.000	0.518	L13	B.10	325.057296	-23.643997	MS2137	0.390	0.560	P2	B.10
322.359630	-7.690811	M2129	0.616	0.215	L14	B.10	325.046757	-23.666493	MS2137	0.826	0.962	P3	B.10
322.374608	-7.693138	M2129	1.000	0.994	L15	B.10	325.060821	-23.653246	MS2137	0.601	0.026	P4	B.10
322.363898	-7.683846	M2129	1.000	0.844	L16	B.10	325.058598	-23.687639	MS2137	0.008	0.329	P5	B.10
322.354745	-7.700537	M2129	1.000	1.000	L17	B.10	325.055221	-23.647320	MS2137	0.838	0.422	P6	B.10
322.361265	-7.696433	M2129	0.001	0.992	L18	B.10	325.064809	-23.643127	MS2137	0.137	0.667	P7	B.10
322.346094	-7.686202	M2129	1.000	0.671	L19	B.10	325.058077	-23.674354	MS2137	0.466	0.556	P8	B.10
322.374405	-7.688749	M2129	0.998	0.354	L20	B.10	325.045623	-23.661151	MS2137	0.113	0.998	P9	B.10
322.345472	-7.694764	M2129	0.948	0.998	L21	B.10	325.055548	-23.684806	MS2137	0.058	0.642	P10	B.10
322.346402	-7.687102	M2129	0.006	0.314	L22	B.10	325.050656	-23.662811	MS2137	1.000	0.002	P11	B.10
322.430260	0.081579	R2129	0.058	0.501	M1	B.10	325.043977	-23.640896	MS2137	0.120	0.656	P12	B.10
322.430535	0.077667	R2129	0.086	0.463	M2	B.10	325.041368	-23.649098	MS2137	0.001	0.616	P13	B.10
322.403545	0.104669	R2129	0.005	0.287	M3	B.10	325.069874	-23.693262	MS2137	0.221	0.004	P14	B.10
322.431283	0.099556	R2129	0.887	0.210	M4	B.10	325.061574	-23.651216	MS2137	0.013	0.284	P15	B.10
322.419355	0.080291	R2129	0.014	0.391	M5	B.10	325.084015	-23.658056	MS2137	0.312	0.920	P16	B.10

Table B.7: Continuing of Tab. B.6.

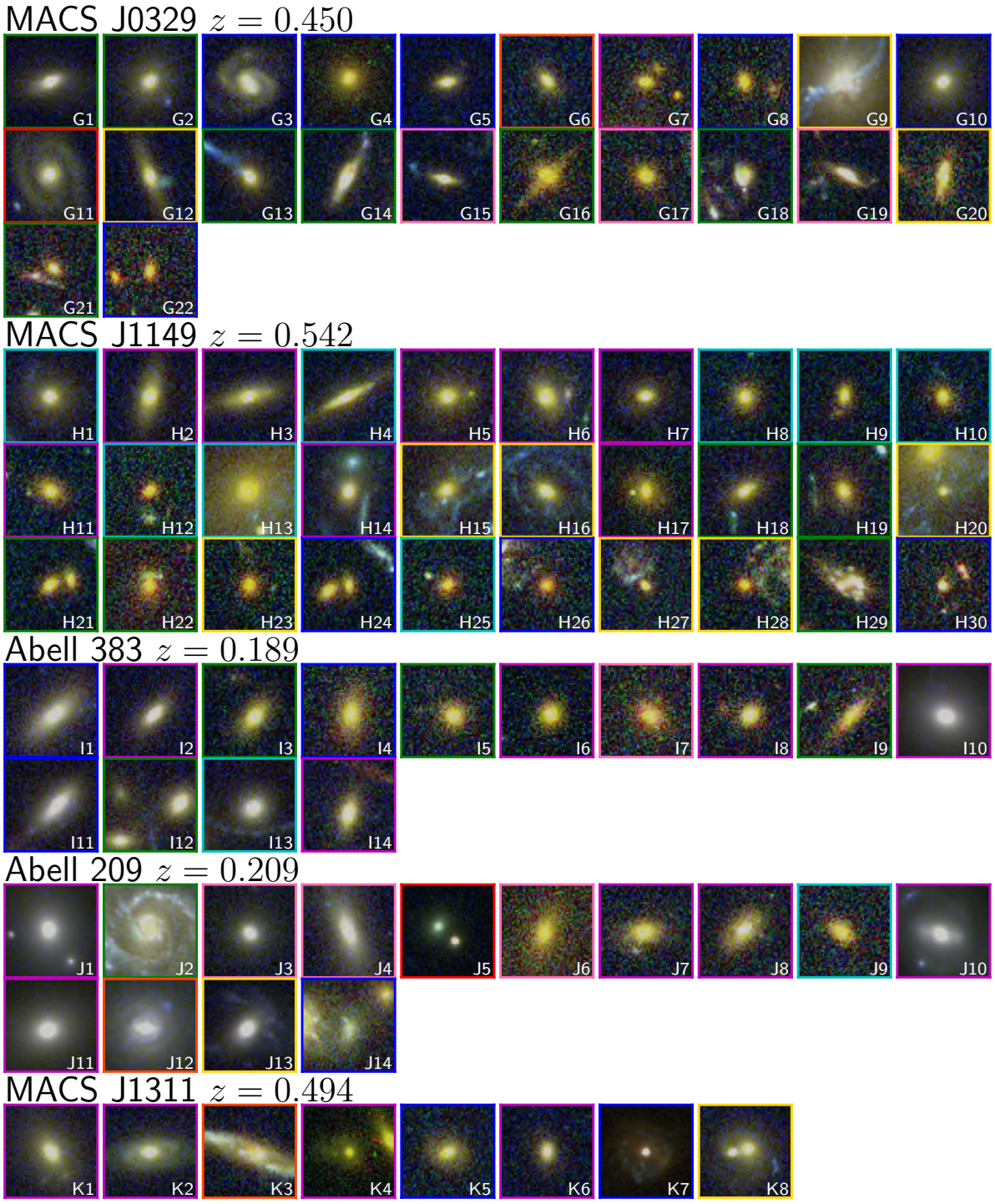


Figure B.8: Continuing of Fig. 6.16. Identified GGSLs by both the models, by searching them around cluster galaxies, related to M0329, M1149, A383, A209 and M1311. Each cutout is surrounding by a coloured square according to the follow scheme based on the GGSL probability:
(*i*) $Pr(VGG)$, $Pr(SC\text{-}VGG) > 0.5 \rightarrow$green;
(*ii*) $Pr(VGG)$, $Pr(SC\text{-}VGG) \in [0.2, 0.5] \rightarrow$orange;
(*iii*) $Pr(VGG) \in [0.2, 0.5]$ & $Pr(SC\text{-}VGG) > 0.5 \rightarrow$red;
(*iv*) $Pr(SC\text{-}VGG) \in [0.2, 0.5]$ & $Pr(VGG) > 0.5 \rightarrow$yellow;
(*v*) $Pr(VGG) > 0.5$ & $Pr(SC\text{-}VGG) \leq 0.2 \rightarrow$cyan;
(*vii*) $Pr(SC\text{-}VGG) > 0.5$ & $Pr(VGG) \leq 0.2 \rightarrow$blue;
(*ix*) $Pr(VGG) \in [0.2, 0.5]$ & $Pr(SC\text{-}VGG) > 0.5 \rightarrow$pink;
(*ix*) $Pr(SC\text{-}VGG) \in [0.2, 0.5]$ & $Pr(VGG) > 0.5 \rightarrow$magenta.

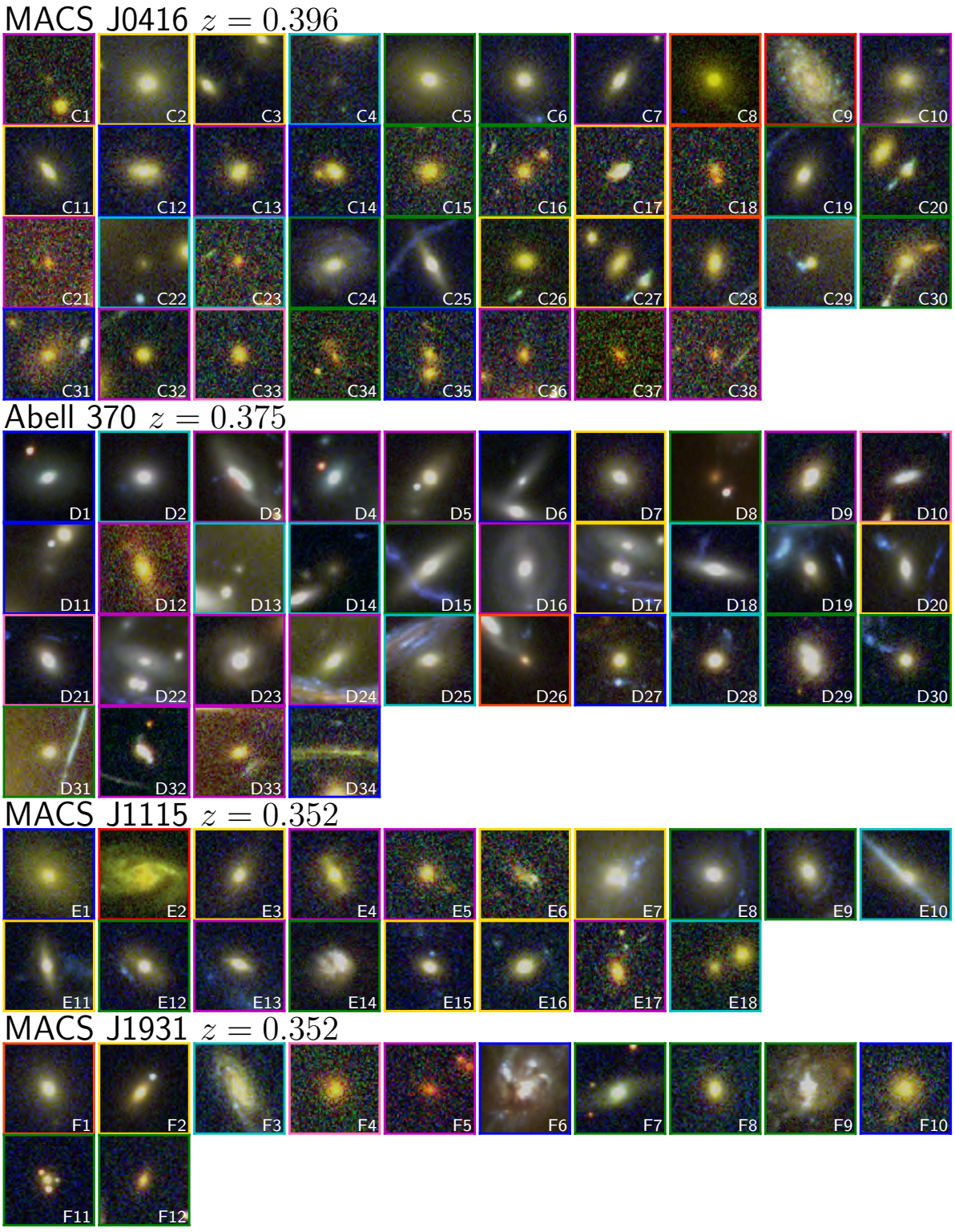


Figure B.9: Continuing of Fig. B.8, related to M0416, A370, M1115, M1931.

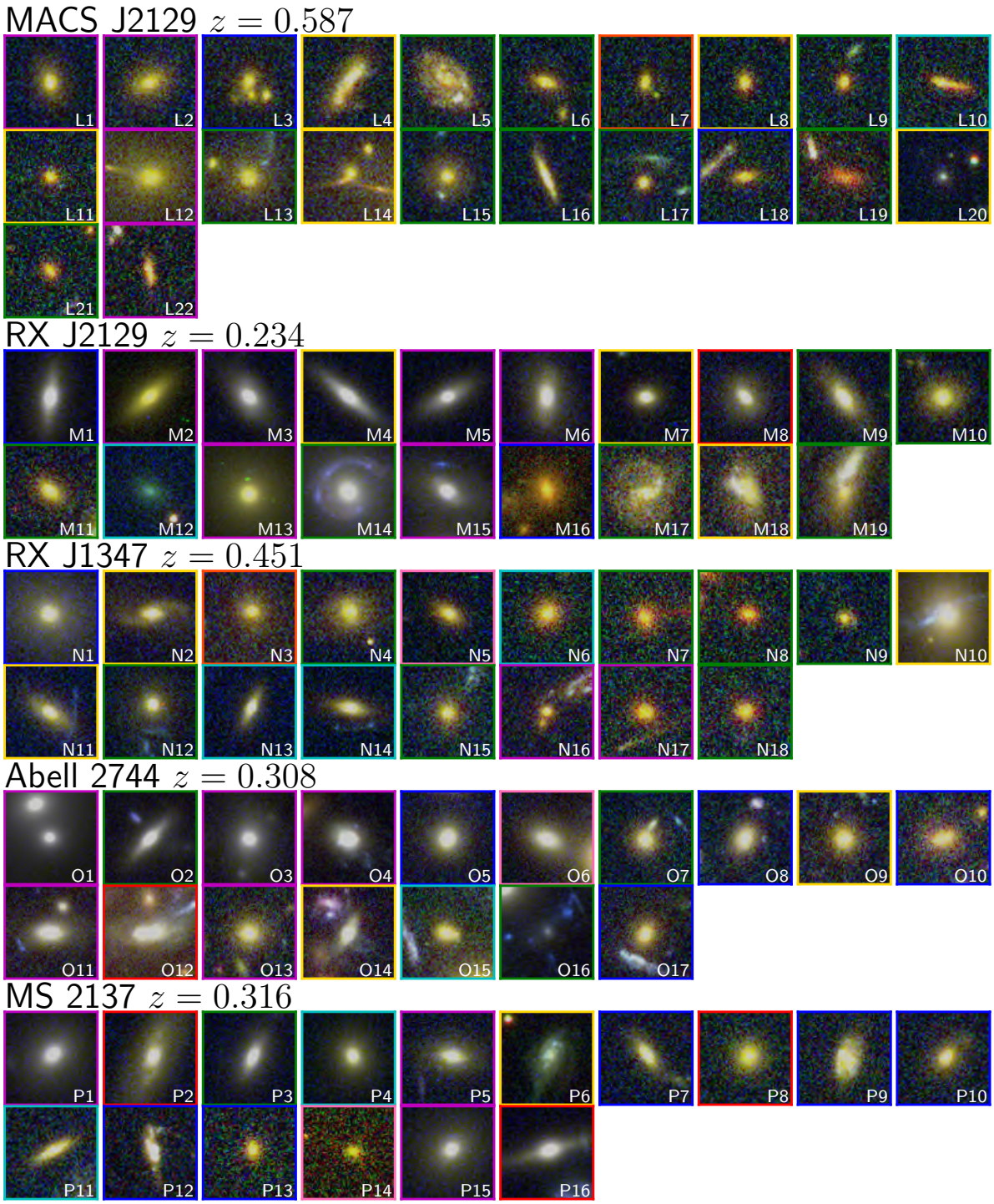


Figure B.10: Continuing of Fig. B.9, related to M2129, R2129, R1347, A2744 and MS2137.

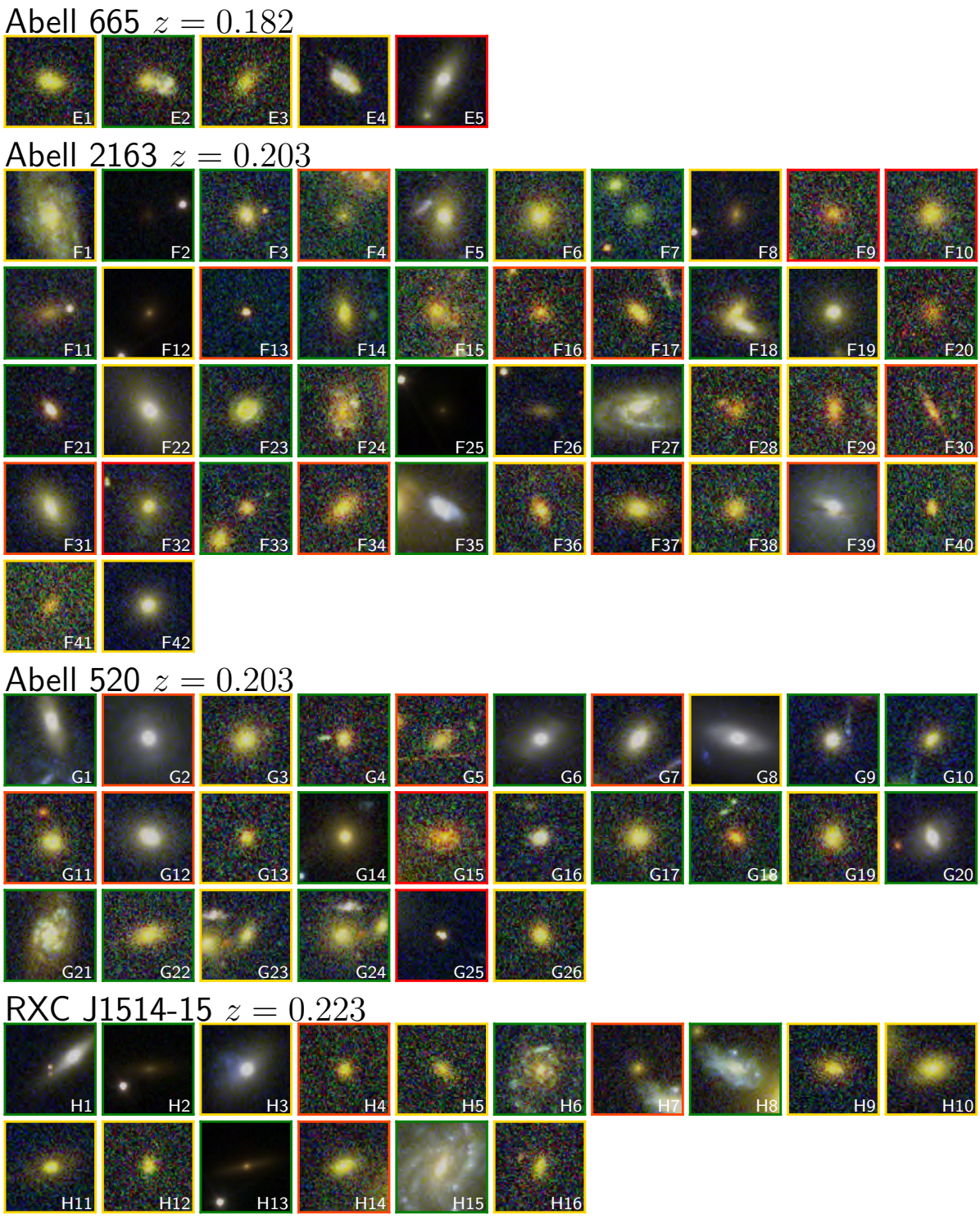


Figure B.11: Continuing of Fig. 6.18. Identified GGSLs by both the models, by searching them around RELICS cluster galaxies detected with CNN (see Sec. 5.5.1), related to A665, A2163, A520 and R1514. Each cutout is surrounding by a coloured square according to the follow scheme based on GGSL probability:
(*i*) $Pr(VGG)$, $Pr(SC\text{-}VGG) > 0.5 \rightarrow$green;
(*ii*) $Pr(VGG)$, $Pr(SC\text{-}VGG) \in [0.2, 0.5] \rightarrow$orange;
(*iii*) $Pr(VGG) \in [0.2, 0.5] \wedge Pr(SC\text{-}VGG) > 0.5 \rightarrow$red;
(*iv*) $Pr(SC\text{-}VGG) \in [0.2, 0.5] \wedge Pr(VGG) > 0.5 \rightarrow$yellow.

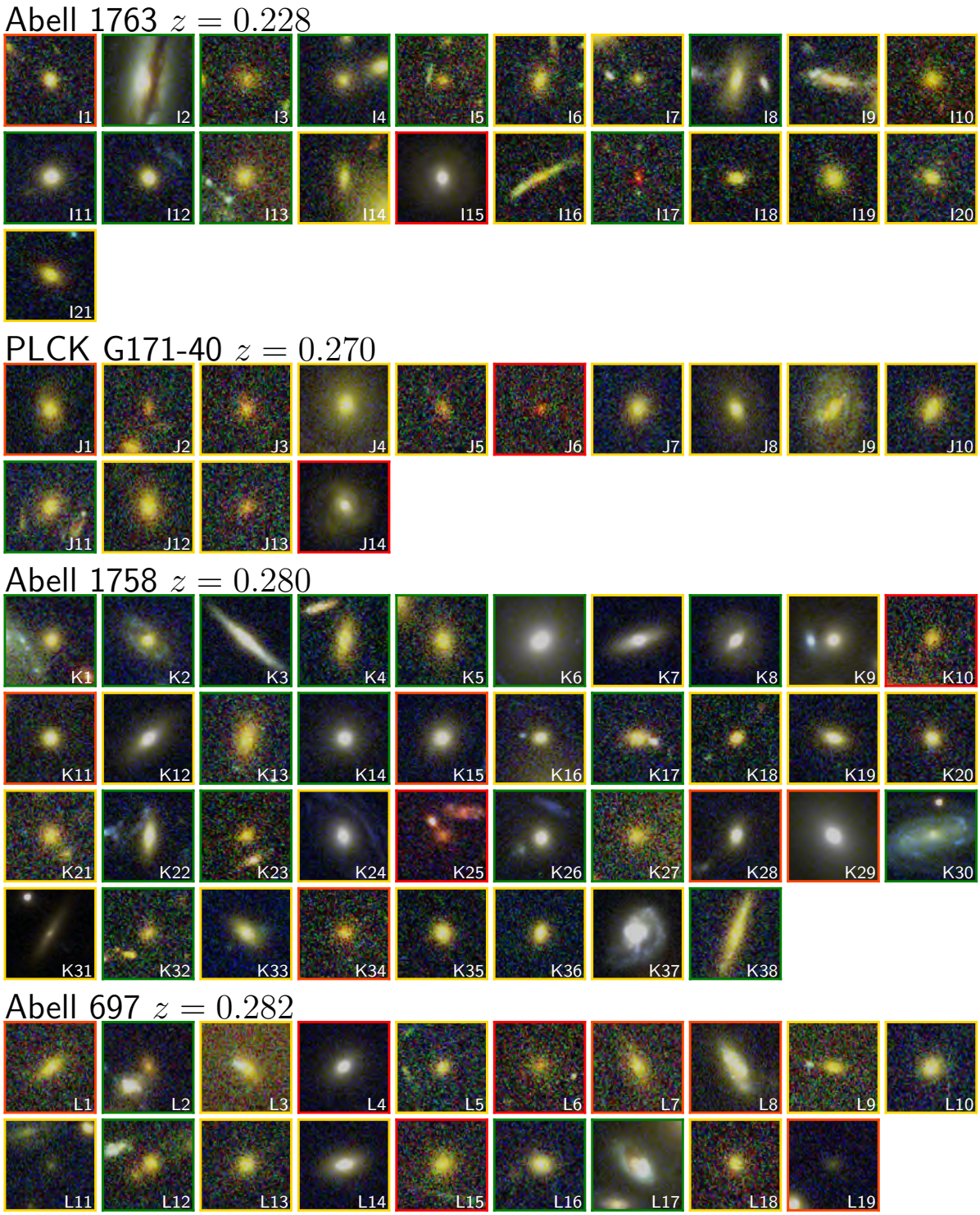


Figure B.12: Continuing of Fig. B.11, related to A1763, P171, A1758 and A697.

Figure B.13: Continuing of Fig. B.12, related to A1300, R0142, M0035, M0308 and R0949.

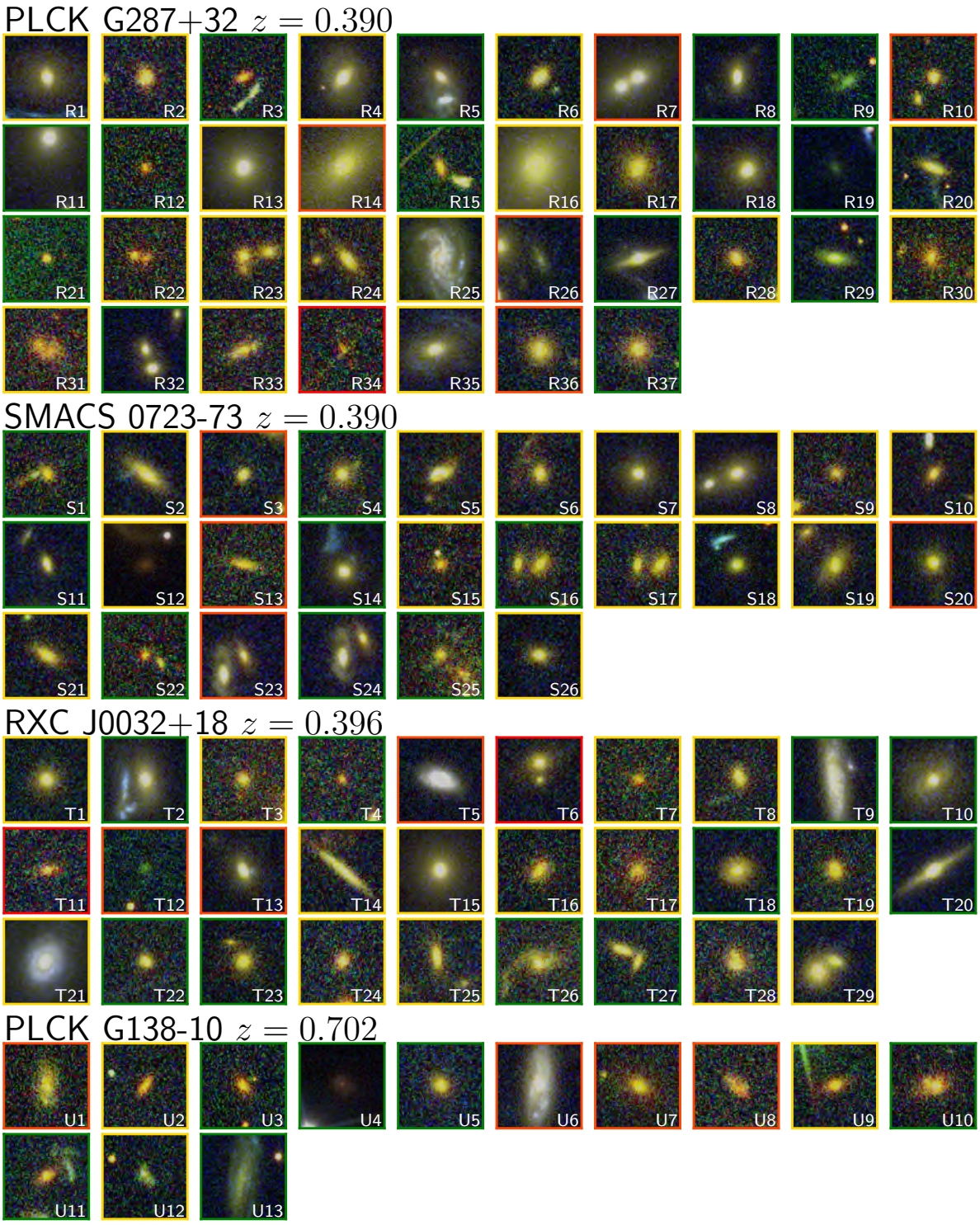


Figure B.14: Continuing of Fig. B.13, related to P287, SM0723, R0032 and P138.

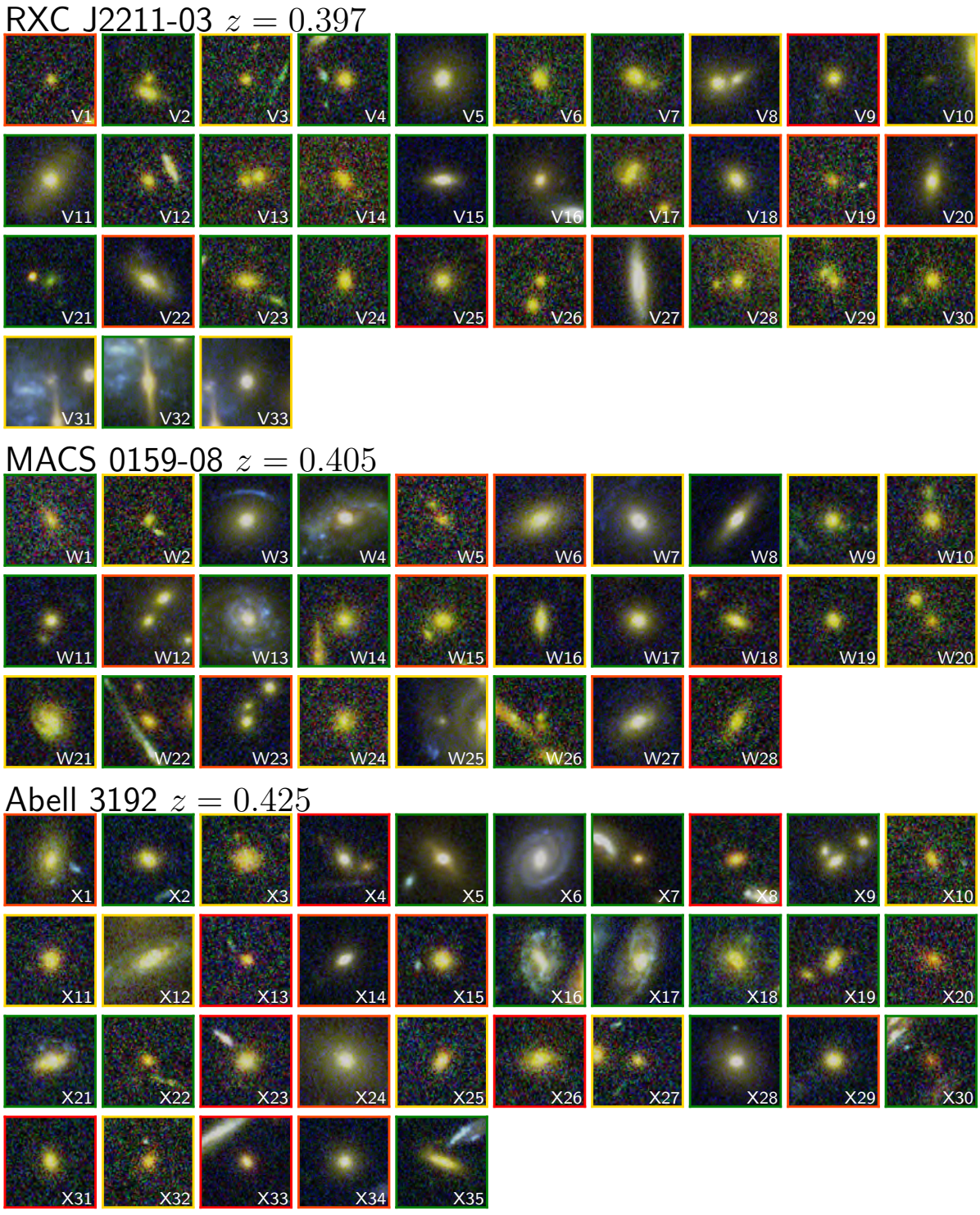


Figure B.15: Continuing of Fig. B.14, related to R2211, M0159 and A3192.

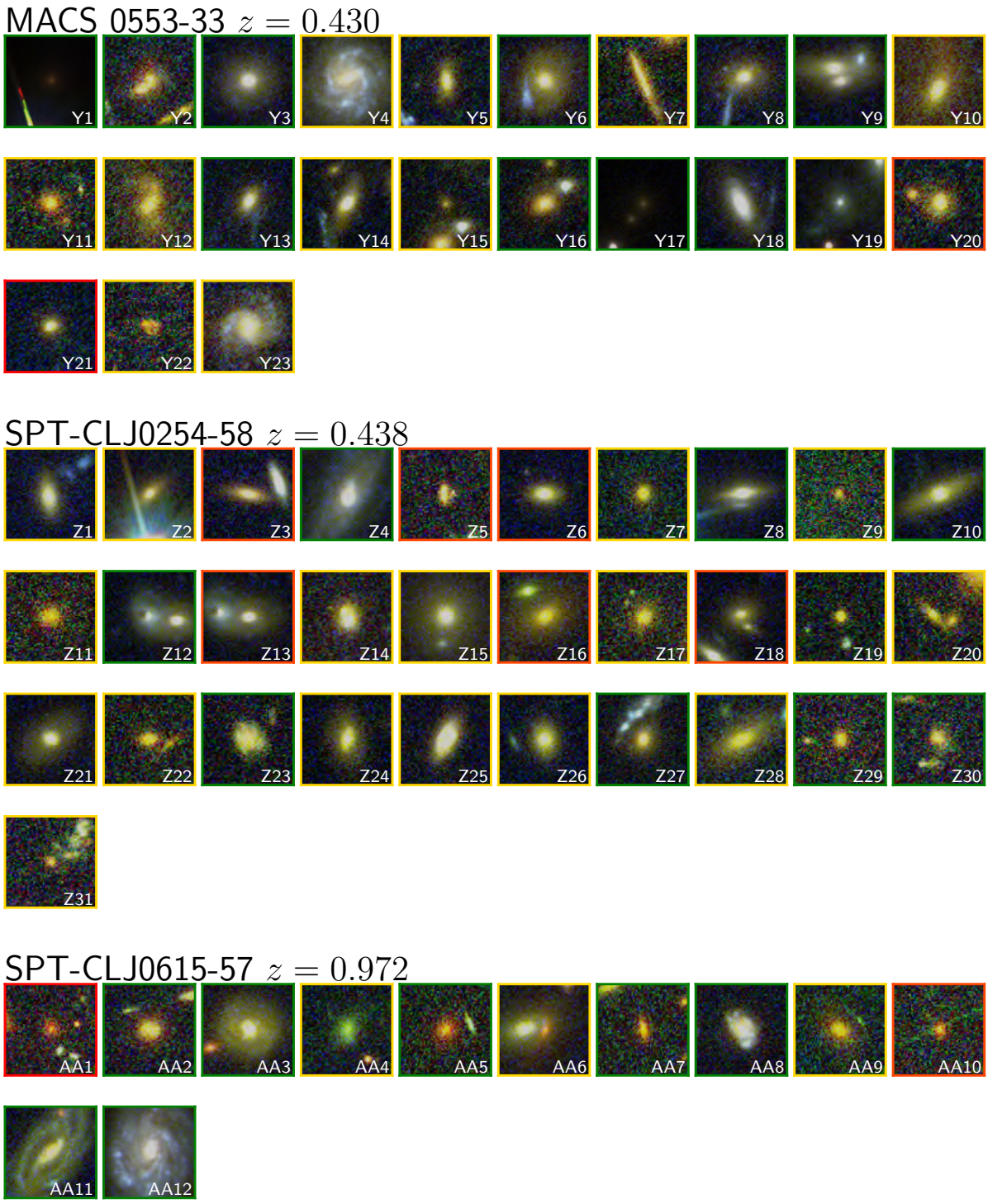


Figure B.16: Continuing of Fig. B.15, related to M0553, S0254, S0615.

Figure B.17: Continuing of Fig. B.16, related to P308 and P004.

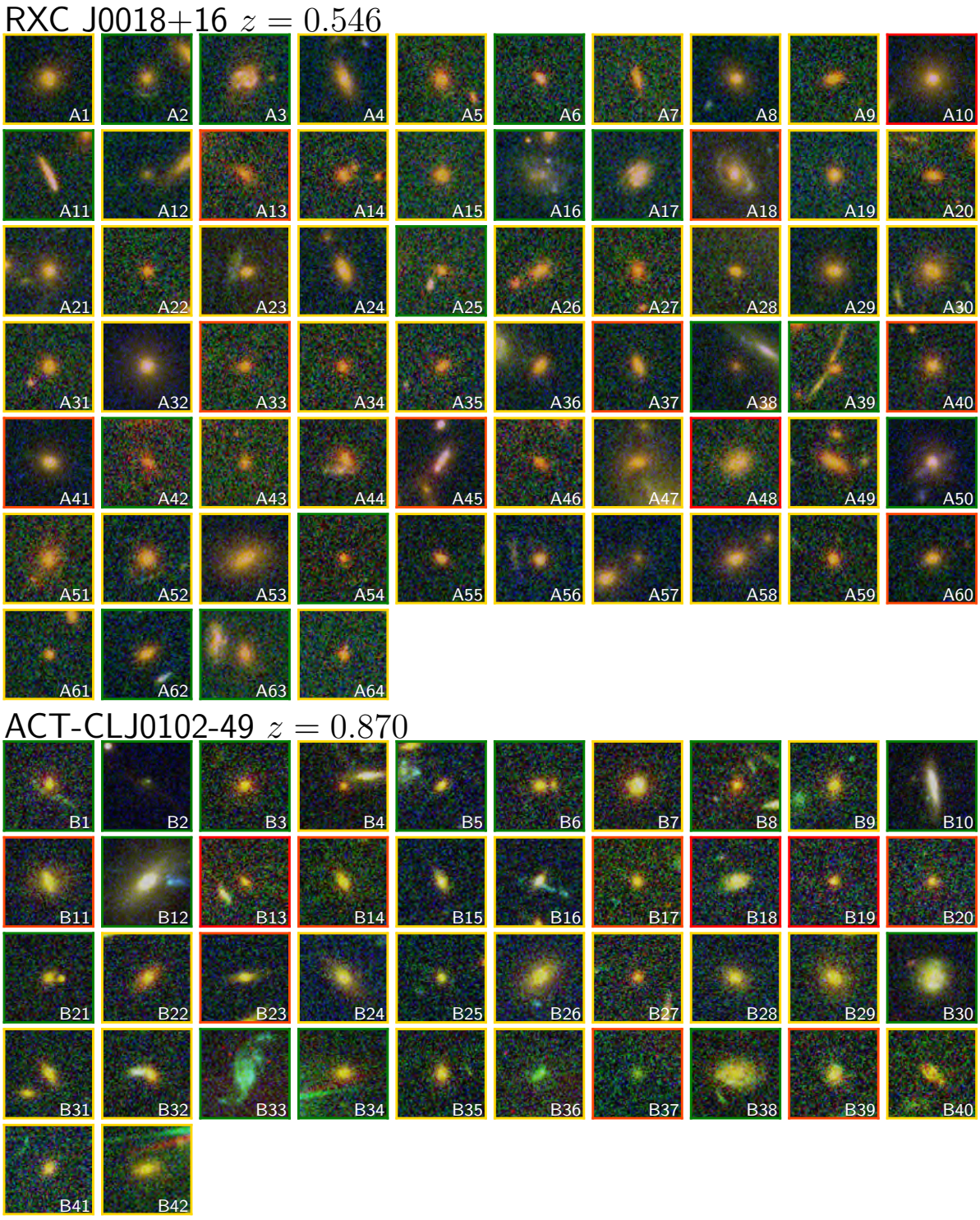


Figure B.18: Continuing of Fig. B.17, related to R0018 and A0102.

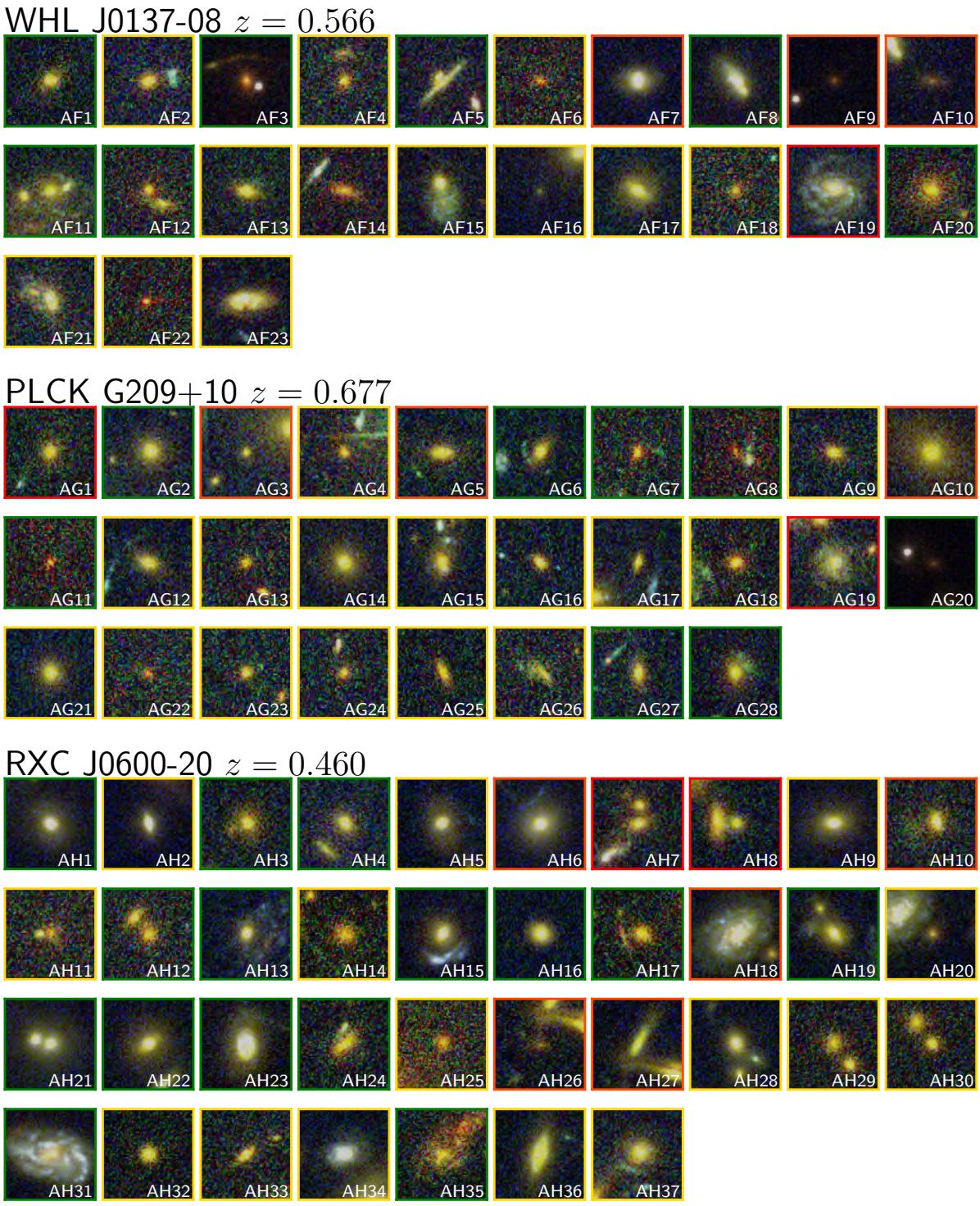


Figure B.19: Continuing of Fig. B.18, related to WHL0137, P209 and R0600.

RA	DEC	Cluster	Pr_{VGG}	$Pr_{SC\text{-}VGG}$	ref	Fig.	RA	DEC	Cluster	Pr_{VGG}	$Pr_{SC\text{-}VGG}$	ref	Fig.
38.083543	-44.329022	R0232	0.977	0.230	A1	6.18	41.399409	-53.025286	AS295	0.933	0.328	D5	6.18
38.076415	-44.327149	R0232	0.952	0.597	A2	6.18	41.373087	-53.025589	AS295	0.999	0.396	D6	6.18
38.073241	-44.328398	R0232	0.995	0.350	A3	6.18	41.380877	-53.025005	AS295	0.974	0.665	D7	6.18
38.081053	-44.327592	R0232	0.936	0.332	A4	6.18	41.389771	-53.025044	AS295	0.315	0.294	D8	6.18
38.070998	-44.329991	R0232	0.947	0.665	A5	6.18	41.356002	-53.032220	AS295	0.294	0.333	D9	6.18
38.087439	-44.330440	R0232	0.987	0.651	A6	6.18	41.349989	-53.025540	AS295	0.505	0.687	D10	6.18
38.073960	-44.329866	R0232	0.999	0.654	A7	6.18	41.348735	-53.025400	AS295	0.513	0.404	D11	6.18
38.093396	-44.331208	R0232	0.971	0.338	A8	6.18	41.377118	-53.028160	AS295	1.000	0.853	D12	6.18
38.089018	-44.332592	R0232	1.000	0.986	A9	6.18	41.403151	-53.026813	AS295	0.868	0.682	D13	6.18
38.080842	-44.333526	R0232	0.999	0.346	A10	6.18	41.405686	-53.027130	AS295	1.000	0.849	D14	6.18
38.079770	-44.333807	R0232	1.000	0.820	A11	6.18	41.361073	-53.026270	AS295	0.999	0.335	D15	6.18
38.085996	-44.333707	R0232	0.996	0.693	A12	6.18	41.364551	-53.025670	AS295	0.921	0.756	D16	6.18
38.062041	-44.333774	R0232	0.821	0.711	A13	6.18	41.353387	-53.029324	AS295	0.212	0.269	D17	6.18
38.077651	-44.334615	R0232	0.674	0.599	A14	6.18	41.371924	-53.028152	AS295	0.744	0.332	D18	6.18
38.088500	-44.333941	R0232	0.453	0.335	A15	6.18	41.366768	-53.029584	AS295	0.859	0.625	D19	6.18
38.096277	-44.333996	R0232	0.272	0.376	A16	6.18	41.367887	-53.028023	AS295	1.000	1.000	D20	6.18
38.096880	-44.334218	R0232	0.570	0.304	A17	6.18	41.399633	-53.028507	AS295	0.935	0.338	D21	6.18
38.078660	-44.334681	R0232	1.000	0.666	A18	6.18	41.394021	-53.029050	AS295	0.987	0.633	D22	6.18
38.091609	-44.336589	R0232	0.996	0.396	A19	6.18	41.369750	-53.029485	AS295	0.308	0.385	D23	6.18
38.063360	-44.334814	R0232	0.741	0.330	A20	6.18	41.363116	-53.029203	AS295	0.992	0.999	D24	6.18
38.106702	-44.335152	R0232	0.994	0.221	A21	6.18	41.383817	-53.029615	AS295	0.920	0.286	D25	6.18
10.852736	-20.628240	A2813	0.981	0.201	B1	6.18	41.407159	-53.030058	AS295	0.701	0.331	D26	6.18
10.868791	-20.599029	A2813	1.000	0.546	B2	6.18	41.346021	-53.030689	AS295	0.977	0.323	D27	6.18
10.865267	-20.600064	A2813	0.819	0.367	B3	6.18	41.388060	-53.030980	AS295	0.504	0.333	D28	6.18
10.841784	-20.601531	A2813	1.000	0.581	B4	6.18	41.396259	-53.030953	AS295	0.264	0.617	D29	6.18
10.861113	-20.601913	A2813	0.938	0.345	B5	6.18	127.751205	65.860738	A665	0.723	0.264	E1	B.11
10.865811	-20.602559	A2813	1.000	1.000	B6	6.18	127.737079	65.864931	A665	1.000	0.962	E2	B.11
10.836914	-20.602395	A2813	1.000	0.332	B7	6.18	127.730490	65.860703	A665	0.602	0.329	E3	B.11
10.863904	-20.603331	A2813	0.782	0.348	B8	6.18	127.726286	65.858298	A665	0.985	0.288	E4	B.11
10.845211	-20.603161	A2813	1.000	0.408	B9	6.18	127.725784	65.857161	A665	0.389	0.647	E5	B.11
10.861435	-20.603005	A2813	0.999	0.356	B10	6.18	243.928079	-6.176388	A2163	0.992	0.256	F1	B.11
10.848932	-20.603527	A2813	0.592	0.326	B11	6.18	243.932567	-6.166514	A2163	0.997	0.895	F2	B.11
10.838214	-20.605300	A2813	0.964	0.982	B12	6.18	243.927604	-6.181065	A2163	0.767	0.667	F3	B.11
10.840793	-20.605032	A2813	0.412	0.536	B13	6.18	243.954509	-6.114815	A2163	0.289	0.340	F4	B.11
10.853951	-20.604483	A2813	0.999	0.529	B14	6.18	243.943015	-6.116479	A2163	0.943	0.725	F5	B.11
10.855255	-20.604495	A2813	0.629	0.326	B15	6.18	243.963589	-6.116862	A2163	0.550	0.333	F6	B.11
10.844802	-20.604287	A2813	0.321	0.213	B16	6.18	243.945983	-6.117801	A2163	0.999	0.665	F7	B.11
10.866275	-20.606122	A2813	0.357	0.421	B17	6.18	243.939783	-6.115412	A2163	0.995	0.333	F8	B.11
347.082112	-2.173065	A2537	1.000	0.874	C1	6.18	243.953630	-6.118431	A2163	0.438	0.655	F9	B.11
347.082648	-2.173957	A2537	0.954	0.503	C2	6.18	243.953180	-6.118344	A2163	0.275	0.784	F10	B.11
347.093627	-2.173993	A2537	1.000	1.000	C3	6.18	243.965554	-6.120163	A2163	1.000	0.992	F11	B.11
347.098979	-2.175405	A2537	0.999	0.333	C4	6.18	243.964979	-6.121074	A2163	0.996	0.291	F12	B.11
347.098264	-2.175023	A2537	0.653	0.464	C5	6.18	243.944383	-6.123981	A2163	0.254	0.333	F13	B.11
347.097265	-2.177083	A2537	0.982	0.323	C6	6.18	243.954813	-6.121941	A2163	0.903	0.653	F14	B.11
347.077677	-2.174581	A2537	0.999	1.000	C7	6.18	243.945840	-6.122522	A2163	0.999	0.855	F15	B.11
347.097907	-2.175957	A2537	0.974	0.441	C8	6.18	243.936479	-6.122404	A2163	0.250	0.333	F16	B.11
347.086477	-2.176348	A2537	0.537	0.209	C9	6.18	243.944732	-6.122504	A2163	0.223	0.301	F17	B.11
347.081802	-2.176433	A2537	0.903	0.652	C10	6.18	243.948987	-6.122425	A2163	0.877	0.799	F18	B.11
347.082295	-2.177589	A2537	1.000	0.560	C11	6.18	243.945703	-6.123422	A2163	0.559	0.283	F19	B.11
347.082749	-2.178636	A2537	0.988	0.949	C12	6.18	243.957849	-6.123865	A2163	0.999	0.666	F20	B.11
347.092148	-2.177220	A2537	0.995	0.957	C13	6.18	243.941699	-6.123722	A2163	0.990	0.516	F21	B.11
347.089517	-2.177938	A2537	0.976	0.620	C14	6.18	243.944075	-6.125182	A2163	0.504	0.273	F22	B.11
347.079016	-2.178462	A2537	0.998	0.334	C15	6.18	243.933384	-6.124473	A2163	0.980	0.830	F23	B.11
347.086675	-2.177875	A2537	0.999	0.367	C16	6.18	243.933122	-6.124808	A2163	0.999	0.688	F24	B.11
347.086522	-2.178336	A2537	0.383	0.333	C17	6.18	243.964522	-6.124531	A2163	0.969	0.685	F25	B.11
347.097725	-2.178192	A2537	0.999	0.255	C18	6.18	243.967581	-6.124480	A2163	0.998	0.334	F26	B.11
347.088423	-2.178213	A2537	0.982	0.339	C19	6.18	243.958084	-6.125638	A2163	0.676	0.668	F27	B.11
347.080590	-2.179580	A2537	1.000	0.649	C20	6.18	243.932159	-6.125774	A2163	0.948	0.207	F28	B.11
347.078155	-2.179578	A2537	0.383	0.220	C21	6.18	243.947271	-6.125733	A2163	0.657	0.228	F29	B.11
347.087689	-2.180594	A2537	0.837	0.382	C22	6.18	243.966202	-6.126714	A2163	0.266	0.314	F30	B.11
41.369295	-53.039381	AS295	0.972	0.895	D1	6.18	243.951838	-6.126754	A2163	0.445	0.496	F31	B.11
41.385382	-53.023733	AS295	1.000	0.845	D2	6.18	243.941327	-6.123020	A2163	0.470	0.790	F32	B.11
41.399157	-53.023974	AS295	0.994	0.349	D3	6.18	243.960555	-6.127892	A2163	0.999	0.744	F33	B.11
41.393147	-53.023839	AS295	0.231	0.205	D4	6.18	243.951900	-6.128254	A2163	0.251	0.370	F34	B.11

Table B.8: List of candidate GGSLs identified by both models, with Pr(VGG) $\wedge$ Pr(SC-VGG) ≥ 0.2.

RA	DEC	Cluster	Pr_{VGG}	$Pr_{SC\text{-}VGG}$	ref	Fig.	RA	DEC	Cluster	Pr_{VGG}	$Pr_{SC\text{-}VGG}$	ref	Fig.
243.969074	-6.129009	A2163	1.000	0.647	F35	B.11	203.834135	41.011804	A1763	0.335	0.652	I15	B.12
243.963453	-6.129161	A2163	0.688	0.280	F36	B.11	203.809154	41.011884	A1763	1.000	0.327	I16	B.12
243.965468	-6.126715	A2163	0.214	0.246	F37	B.11	203.835735	41.010798	A1763	1.000	0.667	I17	B.12
243.966164	-6.131243	A2163	1.000	0.376	F38	B.11	203.829070	41.011168	A1763	0.651	0.208	I18	B.12
243.965383	-6.130550	A2163	0.263	0.425	F39	B.11	203.829807	41.015022	A1763	0.974	0.403	I19	B.12
243.939866	-6.130051	A2163	0.955	0.332	F40	B.11	203.849334	41.009271	A1763	0.987	0.248	I20	B.12
243.957728	-6.132009	A2163	0.894	0.333	F41	B.11	203.846265	41.009587	A1763	0.997	0.289	I21	B.12
243.961024	-6.131954	A2163	0.595	0.246	F42	B.11	48.237654	8.394189	P171	0.484	0.329	J1	B.12
73.502484	2.892460	A520	1.000	0.998	G1	B.11	48.232350	8.388718	P171	0.986	0.425	J2	B.12
73.515839	2.892225	A520	0.347	0.311	G2	B.11	48.231096	8.387971	P171	0.653	0.274	J3	B.12
73.506102	2.894669	A520	0.999	0.333	G3	B.11	48.241684	8.388159	P171	0.575	0.241	J4	B.12
73.511918	2.872859	A520	0.559	0.934	G4	B.11	48.232614	8.387442	P171	0.902	0.333	J5	B.12
73.516587	2.879256	A520	0.241	0.332	G5	B.11	48.243548	8.386929	P171	0.483	0.976	J6	B.12
73.519485	2.877890	A520	0.833	0.608	G6	B.11	48.242611	8.387193	P171	0.898	0.393	J7	B.12
73.508703	2.882102	A520	0.221	0.472	G7	B.11	48.241917	8.386927	P171	0.896	0.399	J8	B.12
73.518811	2.878598	A520	0.980	0.233	G8	B.11	48.233890	8.386564	P171	0.822	0.403	J9	B.12
73.512245	2.880953	A520	1.000	0.980	G9	B.11	48.232002	8.385988	P171	0.599	0.330	J10	B.12
73.510044	2.879602	A520	1.000	0.570	G10	B.11	48.232394	8.385528	P171	0.993	0.722	J11	B.12
73.522595	2.883650	A520	0.232	0.321	G11	B.11	48.233435	8.385559	P171	0.630	0.304	J12	B.12
73.531135	2.884223	A520	0.451	0.202	G12	B.11	48.233702	8.384986	P171	0.934	0.333	J13	B.12
73.532224	2.884880	A520	0.939	0.261	G13	B.11	48.238218	8.385716	P171	0.365	0.727	J14	B.12
73.508372	2.883982	A520	0.951	0.853	G14	B.11	203.227378	50.507703	A1758	0.714	0.586	K1	B.12
73.528543	2.884857	A520	0.469	0.697	G15	B.11	203.245075	50.506892	A1758	0.998	0.578	K2	B.12
73.503075	2.885970	A520	0.706	0.264	G16	B.11	203.144731	50.574991	A1758	1.000	0.998	K3	B.12
73.518635	2.886463	A520	0.739	0.664	G17	B.11	203.162039	50.575119	A1758	0.994	0.997	K4	B.12
73.517729	2.885292	A520	0.998	0.879	G18	B.11	203.175074	50.576332	A1758	0.994	0.817	K5	B.12
73.515037	2.888292	A520	0.997	0.339	G19	B.11	203.185998	50.571825	A1758	0.865	0.588	K6	B.12
73.531860	2.886364	A520	0.996	0.786	G20	B.11	203.189520	50.569973	A1758	0.633	0.277	K7	B.12
73.525684	2.887077	A520	0.998	0.688	G21	B.11	203.137141	50.569222	A1758	0.992	0.675	K8	B.12
73.513571	2.886791	A520	0.984	0.666	G22	B.11	203.187317	50.568311	A1758	0.996	0.286	K9	B.12
73.520903	2.886179	A520	0.911	0.315	G23	B.11	203.177243	50.568650	A1758	0.209	0.610	K10	B.12
73.537001	2.889209	A520	0.832	0.646	G24	B.11	203.180469	50.568280	A1758	0.465	0.360	K11	B.12
73.527925	2.889526	A520	0.433	0.654	G25	B.11	203.164827	50.566739	A1758	0.817	0.228	K12	B.12
73.521153	2.888601	A520	0.901	0.332	G26	B.11	203.186184	50.567327	A1758	0.875	0.554	K13	B.12
228.765673	-15.360110	R1514	1.000	0.995	H1	B.11	203.161545	50.567085	A1758	0.980	0.668	K14	B.12
228.765944	-15.360266	R1514	0.985	0.513	H2	B.11	203.134210	50.567240	A1758	0.262	0.369	K15	B.12
228.757889	-15.363546	R1514	0.868	0.327	H3	B.11	203.144768	50.567397	A1758	1.000	0.207	K16	B.12
228.761182	-15.361645	R1514	0.479	0.282	H4	B.11	203.191053	50.566101	A1758	0.992	0.627	K17	B.12
228.762870	-15.362066	R1514	0.883	0.292	H5	B.11	203.170178	50.564441	A1758	0.631	0.486	K18	B.12
228.761116	-15.364888	R1514	1.000	0.778	H6	B.11	203.170577	50.562903	A1758	0.563	0.317	K19	B.12
228.763191	-15.363550	R1514	0.299	0.333	H7	B.11	203.134941	50.563609	A1758	0.999	0.204	K20	B.12
228.762808	-15.365690	R1514	1.000	0.668	H8	B.11	203.150043	50.563574	A1758	0.869	0.442	K21	B.12
228.763984	-15.364214	R1514	0.863	0.319	H9	B.11	203.149268	50.564453	A1758	1.000	0.842	K22	B.12
228.762746	-15.364774	R1514	0.812	0.450	H10	B.11	203.164299	50.558240	A1758	0.850	0.585	K23	B.12
228.755525	-15.363511	R1514	0.704	0.451	H11	B.11	203.175381	50.561865	A1758	0.994	0.260	K24	B.12
228.768402	-15.363776	R1514	0.570	0.360	H12	B.11	203.140136	50.562191	A1758	0.300	0.993	K25	B.12
228.756827	-15.363892	R1514	0.997	0.686	H13	B.11	203.164682	50.562525	A1758	1.000	0.781	K26	B.12
228.753447	-15.364636	R1514	0.215	0.366	H14	B.11	203.188410	50.561612	A1758	0.518	0.558	K27	B.12
228.746458	-15.365964	R1514	0.652	0.724	H15	B.11	203.152112	50.561224	A1758	0.418	0.332	K28	B.12
228.743536	-15.367332	R1514	0.717	0.344	H16	B.11	203.150525	50.560493	A1758	0.218	0.207	K29	B.12
203.815783	41.011039	A1763	0.209	0.334	I1	B.12	203.142701	50.560807	A1758	1.000	0.983	K30	B.12
203.823442	41.009372	A1763	0.620	0.992	I2	B.12	203.183297	50.560480	A1758	0.716	0.316	K31	B.12
203.824574	41.007607	A1763	0.999	0.692	I3	B.12	203.169830	50.560359	A1758	0.872	0.614	K32	B.12
203.809801	41.014865	A1763	0.880	0.556	I4	B.12	203.148680	50.560605	A1758	1.000	0.437	K33	B.12
203.808563	41.014640	A1763	1.000	0.667	I5	B.12	203.185150	50.559935	A1758	0.438	0.215	K34	B.12
203.820694	41.009933	A1763	0.977	0.333	I6	B.12	203.143698	50.560142	A1758	0.814	0.346	K35	B.12
203.834272	41.013873	A1763	0.941	0.473	I7	B.12	203.134934	50.559376	A1758	0.992	0.464	K36	B.12
203.837263	41.013997	A1763	0.669	1.000	I8	B.12	203.180912	50.557989	A1758	0.883	0.324	K37	B.12
203.812413	41.013948	A1763	0.998	0.333	I9	B.12	203.175744	50.558895	A1758	0.837	0.895	K38	B.12
203.805959	41.013636	A1763	0.809	0.269	I10	B.12	130.751119	36.383217	A697	0.464	0.346	L1	B.12
203.851870	41.013330	A1763	0.983	0.623	I11	B.12	130.748605	36.381625	A697	0.998	0.796	L2	B.12
203.810994	41.012405	A1763	0.999	0.739	I12	B.12	130.741750	36.379195	A697	0.897	0.299	L3	B.12
203.824297	41.015710	A1763	0.998	0.730	I13	B.12	130.750966	36.379115	A697	0.361	0.501	L4	B.12
203.846413	41.012665	A1763	0.507	0.289	I14	B.12	130.758225	36.379297	A697	0.999	0.332	L5	B.12

Table B.9: Continuing of Tab. B.8.

RA	DEC	Cluster	Pr_{VGG}	$Pr_{SC\text{-}VGG}$	ref	Fig.	RA	DEC	Cluster	Pr_{VGG}	$Pr_{SC\text{-}VGG}$	ref	Fig.
130.762025	36.377412	A697	0.260	0.528	L6	B.12	47.238189	26.781248	M0308	1.000	1.000	P1	B.13
130.747593	36.376134	A697	0.452	0.304	L7	B.12	47.245838	26.781023	M0308	0.667	0.450	P2	B.13
130.748646	36.375835	A697	0.240	0.333	L8	B.12	47.235160	26.779966	M0308	0.965	0.332	P3	B.13
130.737197	36.376042	A697	0.602	0.353	L9	B.12	47.243672	26.780568	M0308	0.965	0.289	P4	B.13
130.739249	36.375289	A697	1.000	0.285	L10	B.12	47.234077	26.777638	M0308	0.996	0.344	P5	B.13
130.725808	36.375084	A697	0.833	0.342	L11	B.12	47.245308	26.777551	M0308	0.998	0.250	P6	B.13
130.740900	36.374686	A697	0.988	0.667	L12	B.12	47.240736	26.777097	M0308	0.563	0.667	P7	B.13
130.739241	36.374455	A697	0.976	0.334	L13	B.12	47.242694	26.776556	M0308	0.330	0.252	P8	B.13
130.724229	36.374141	A697	0.950	0.255	L14	B.12	47.241061	26.776545	M0308	0.966	1.000	P9	B.13
130.727404	36.374074	A697	0.225	0.565	L15	B.12	47.240113	26.775889	M0308	1.000	0.322	P10	B.13
130.730753	36.373980	A697	1.000	0.614	L16	B.12	47.231850	26.775639	M0308	0.999	0.335	P11	B.13
130.751409	36.373914	A697	0.704	0.628	L17	B.12	47.242354	26.775036	M0308	0.923	0.665	P12	B.13
130.762202	36.373712	A697	0.969	0.349	L18	B.12	47.228422	26.776195	M0308	0.995	0.727	P13	B.13
130.743802	36.364267	A697	0.494	0.274	L19	B.12	47.231809	26.774892	M0308	0.383	0.294	P14	B.13
172.972198	-19.929100	A1300	0.980	0.438	M1	B.13	47.224539	26.774908	M0308	0.395	0.423	P15	B.13
172.981823	-19.902731	A1300	0.828	0.202	M2	B.13	47.244640	26.774713	M0308	0.321	0.220	P16	B.13
172.974797	-19.906521	A1300	1.000	0.469	M3	B.13	47.238096	26.774197	M0308	0.306	0.333	P17	B.13
172.983333	-19.907735	A1300	0.999	0.270	M4	B.13	47.237336	26.773844	M0308	0.927	0.294	P18	B.13
172.987749	-19.907636	A1300	0.999	0.803	M5	B.13	47.229172	26.773672	M0308	1.000	0.703	P19	B.13
172.970618	-19.907926	A1300	0.625	0.632	M6	B.13	147.472323	17.139095	R0949	0.922	0.674	Q1	B.13
172.987746	-19.908371	A1300	1.000	0.663	M7	B.13	147.468292	17.138444	R0949	0.999	0.549	Q2	B.13
172.971213	-19.910747	A1300	0.538	0.342	M8	B.13	147.474248	17.138188	R0949	0.345	0.227	Q3	B.13
172.967673	-19.912232	A1300	0.997	0.933	M9	B.13	147.466839	17.137507	R0949	0.804	0.667	Q4	B.13
172.987576	-19.910593	A1300	0.882	0.589	M10	B.13	147.462137	17.137186	R0949	1.000	0.429	Q5	B.13
172.975181	-19.911377	A1300	0.978	0.474	M11	B.13	147.469564	17.137098	R0949	0.991	0.366	Q6	B.13
172.982521	-19.912162	A1300	0.312	0.692	M12	B.13	147.469043	17.136818	R0949	0.992	0.973	Q7	B.13
172.961008	-19.912426	A1300	0.993	0.642	M13	B.13	147.464849	17.136532	R0949	0.968	0.605	Q8	B.13
172.972642	-19.909413	A1300	1.000	0.999	M14	B.13	147.456091	17.136123	R0949	0.901	0.252	Q9	B.13
172.991050	-19.912786	A1300	0.915	0.445	M15	B.13	147.463530	17.136079	R0949	0.587	0.523	Q10	B.13
172.965870	-19.912038	A1300	0.999	0.370	M16	B.13	147.468843	17.135241	R0949	0.964	0.333	Q11	B.13
172.979752	-19.913286	A1300	0.726	0.326	M17	B.13	147.452568	17.134959	R0949	0.956	0.535	Q12	B.13
172.962195	-19.913371	A1300	0.932	0.279	M18	B.13	147.467276	17.134854	R0949	0.924	0.347	Q13	B.13
172.977756	-19.913425	A1300	0.415	0.518	M19	B.13	147.453441	17.134149	R0949	0.787	0.439	Q14	B.13
172.973807	-19.914043	A1300	0.998	0.471	M20	B.13	147.459301	17.134340	R0949	1.000	1.000	Q15	B.13
172.974242	-19.912522	A1300	0.852	0.577	M21	B.13	147.450586	17.133682	R0949	0.724	0.259	Q16	B.13
172.960913	-19.914074	A1300	0.245	0.463	M22	B.13	147.467889	17.132991	R0949	0.564	0.332	Q17	B.13
25.733187	44.658451	R0142	0.933	0.561	N1	B.13	147.464238	17.133526	R0949	0.997	0.328	Q18	B.13
25.738049	44.655190	R0142	0.256	0.335	N2	B.13	147.478119	17.132445	R0949	0.994	0.664	Q19	B.13
25.739779	44.655876	R0142	0.998	0.646	N3	B.13	147.461082	17.132416	R0949	0.474	0.468	Q20	B.13
25.729237	44.654171	R0142	0.990	0.417	N4	B.13	147.469684	17.132223	R0949	0.292	0.333	Q21	B.13
25.736988	44.653946	R0142	0.534	0.322	N5	B.13	147.466597	17.132210	R0949	0.962	0.353	Q22	B.13
25.733071	44.652208	R0142	0.309	0.656	N6	B.13	147.465909	17.131664	R0949	0.246	0.335	Q23	B.13
25.741553	44.651228	R0142	1.000	0.224	N7	B.13	147.447388	17.131799	R0949	0.999	0.666	Q24	B.13
25.720645	44.650318	R0142	0.997	0.629	N8	B.13	147.445330	17.131481	R0949	0.999	0.667	Q25	B.13
25.739717	44.650971	R0142	0.415	0.372	N9	B.13	147.457065	17.131177	R0949	0.351	0.350	Q26	B.13
25.741456	44.650383	R0142	0.677	0.360	N10	B.13	147.465073	17.131176	R0949	0.424	0.619	Q27	B.13
25.738044	44.650559	R0142	1.000	0.333	N11	B.13	147.468775	17.130714	R0949	0.738	0.967	Q28	B.13
25.746599	44.649857	R0142	0.298	0.358	N12	B.13	147.473401	17.130036	R0949	0.519	0.361	Q29	B.13
25.727291	44.649652	R0142	1.000	0.674	N13	B.13	177.699693	-28.090042	P287	0.968	0.331	R1	B.14
8.854833	-20.239799	M0035	0.996	0.282	O1	B.13	177.698819	-28.060667	P287	0.965	0.328	R2	B.14
8.855933	-20.243171	M0035	0.897	0.350	O2	B.13	177.704937	-28.061015	P287	1.000	0.926	R3	B.14
8.858709	-20.243636	M0035	0.792	0.705	O3	B.13	177.708342	-28.062141	P287	0.918	0.426	R4	B.14
8.864167	-20.244825	M0035	1.000	0.336	O4	B.13	177.699184	-28.062859	P287	0.993	0.666	R5	B.14
8.852255	-20.245320	M0035	0.999	0.531	O5	B.13	177.705402	-28.063905	P287	1.000	0.480	R6	B.14
8.867574	-20.246947	M0035	1.000	0.333	O6	B.13	177.704149	-28.064528	P287	0.325	0.322	R7	B.14
8.860331	-20.248645	M0035	0.736	0.395	O7	B.13	177.714712	-28.064546	P287	0.569	0.555	R8	B.14
8.884869	-20.248035	M0035	0.492	0.593	O8	B.13	177.707035	-28.062522	P287	1.000	0.982	R9	B.14
8.862643	-20.246181	M0035	0.447	0.231	O9	B.13	177.709675	-28.064669	P287	0.236	0.334	R10	B.14
8.855061	-20.248869	M0035	0.419	0.532	O10	B.13	177.717635	-28.065714	P287	0.996	1.000	R11	B.14
8.852004	-20.249465	M0035	0.435	0.982	O11	B.13	177.701623	-28.065965	P287	0.510	0.934	R12	B.14
8.868308	-20.242858	M0035	0.891	0.638	O12	B.13	177.704161	-28.066537	P287	0.650	0.253	R13	B.14
8.865948	-20.252209	M0035	0.557	0.758	O13	B.13	177.710545	-28.066542	P287	0.444	0.248	R14	B.14
8.856458	-20.252347	M0035	0.825	0.395	O14	B.13	177.704865	-28.066455	P287	0.742	0.985	R15	B.14
8.869537	-20.253199	M0035	0.984	0.355	O15	B.13	177.699844	-28.066534	P287	0.790	0.309	R16	B.14

Table B.10: Continuing of Tab. B.9.

RA	DEC	Cluster	Pr_{VGG}	$Pr_{SC\text{-}VGG}$	ref	Fig.	RA	DEC	Cluster	Pr_{VGG}	$Pr_{SC\text{-}VGG}$	ref	Fig.
177.714458	-28.067332	P287	0.822	0.404	R17	B.14	8.063187	18.145563	R0032	0.983	0.537	T18	B.14
177.703455	-28.067592	P287	1.000	0.890	R18	B.14	8.056441	18.145435	R0032	0.963	0.380	T19	B.14
177.698174	-28.067561	P287	0.973	0.581	R19	B.14	8.051886	18.144986	R0032	0.791	0.588	T20	B.14
177.716704	-28.067598	P287	1.000	0.300	R20	B.14	8.053920	18.145246	R0032	0.638	0.275	T21	B.14
177.702759	-28.067849	P287	1.000	0.858	R21	B.14	8.045466	18.145230	R0032	1.000	0.577	T22	B.14
177.713576	-28.067318	P287	0.963	0.435	R22	B.14	8.048316	18.145290	R0032	0.994	0.646	T23	B.14
177.718427	-28.067836	P287	0.989	0.251	R23	B.14	8.051599	18.142515	R0032	0.800	0.380	T24	B.14
177.706995	-28.068124	P287	1.000	0.236	R24	B.14	8.050777	18.142065	R0032	0.999	0.334	T25	B.14
177.702324	-28.068199	P287	0.947	0.401	R25	B.14	8.033421	18.144492	R0032	0.995	0.667	T26	B.14
177.705192	-28.068764	P287	0.238	0.318	R26	B.14	8.059431	18.145121	R0032	0.918	0.703	T27	B.14
177.715230	-28.068991	P287	0.786	0.970	R27	B.14	8.044874	18.139102	R0032	0.883	0.415	T28	B.14
177.714032	-28.068670	P287	0.669	0.274	R28	B.14	8.065822	18.144855	R0032	0.993	0.378	T29	B.14
177.711964	-28.068809	P287	1.000	0.921	R29	B.14	36.793718	49.028778	P138	0.334	0.305	U1	B.14
177.718060	-28.069205	P287	0.748	0.343	R30	B.14	36.777962	49.028105	P138	0.698	0.321	U2	B.14
177.706413	-28.069512	P287	1.000	0.339	R31	B.14	36.785239	49.027930	P138	0.958	0.613	U3	B.14
177.709392	-28.069886	P287	0.994	0.645	R32	B.14	36.779104	49.026812	P138	0.987	0.593	U4	B.14
177.698591	-28.065546	P287	0.897	0.333	R33	B.14	36.774338	49.023918	P138	0.997	0.569	U5	B.14
177.726546	-28.070000	P287	0.311	0.669	R34	B.14	36.778074	49.024234	P138	0.400	0.361	U6	B.14
177.692986	-28.070934	P287	0.826	0.334	R35	B.14	36.766373	49.020475	P138	0.469	0.323	U7	B.14
177.691826	-28.071195	P287	0.428	0.354	R36	B.14	36.802899	49.019861	P138	0.406	0.333	U8	B.14
177.702485	-28.070230	P287	0.995	0.553	R37	B.14	36.790663	49.019515	P138	1.000	0.309	U9	B.14
110.829775	-73.433967	SM0723	0.986	0.666	S1	B.14	36.777108	49.017547	P138	0.961	0.626	U10	B.14
110.809605	-73.435490	SM0723	0.838	0.254	S2	B.14	36.777036	49.016390	P138	1.000	0.666	U11	B.14
110.838529	-73.435266	SM0723	0.409	0.235	S3	B.14	36.771894	49.016999	P138	0.997	0.286	U12	B.14
110.836723	-73.436255	SM0723	0.825	0.662	S4	B.14	36.760974	49.016506	P138	1.000	0.637	U13	B.14
110.837537	-73.436923	SM0723	0.584	0.295	S5	B.14	332.944317	-3.806315	R2211	0.328	0.213	V1	B.15
110.835860	-73.437334	SM0723	0.957	0.432	S6	B.14	332.947055	-3.806606	R2211	0.961	0.966	V2	B.15
110.853149	-73.437837	SM0723	0.951	0.308	S7	B.14	332.946578	-3.807178	R2211	1.000	0.333	V3	B.15
110.812462	-73.439979	SM0723	0.803	0.293	S8	B.14	332.942118	-3.807942	R2211	1.000	0.673	V4	B.15
110.840551	-73.440620	SM0723	0.991	0.334	S9	B.14	332.944407	-3.808375	R2211	0.954	0.665	V5	B.15
110.806765	-73.440961	SM0723	0.893	0.338	S10	B.14	332.946068	-3.808871	R2211	0.697	0.338	V6	B.15
110.852630	-73.441596	SM0723	0.951	0.658	S11	B.14	332.951242	-3.810859	R2211	0.907	0.522	V7	B.15
110.849733	-73.442273	SM0723	0.797	0.200	S12	B.14	332.939099	-3.810894	R2211	0.788	0.230	V8	B.15
110.805396	-73.442262	SM0723	0.348	0.297	S13	B.14	332.950487	-3.810783	R2211	0.436	0.699	V9	B.15
110.865293	-73.442315	SM0723	1.000	0.971	S14	B.14	332.942567	-3.811655	R2211	0.940	0.250	V10	B.15
110.856499	-73.441232	SM0723	1.000	0.348	S15	B.14	332.942149	-3.811989	R2211	0.892	0.517	V11	B.15
110.886405	-73.442762	SM0723	0.700	0.667	S16	B.14	332.953150	-3.812144	R2211	1.000	0.593	V12	B.15
110.828065	-73.443223	SM0723	0.719	0.335	S17	B.14	332.948313	-3.812906	R2211	1.000	0.667	V13	B.15
110.851143	-73.443188	SM0723	1.000	0.415	S18	B.14	332.947132	-3.813275	R2211	0.748	0.556	V14	B.15
110.849419	-73.443369	SM0723	0.558	0.226	S19	B.14	332.941042	-3.813978	R2211	0.857	0.725	V15	B.15
110.840216	-73.443452	SM0723	0.430	0.325	S20	B.14	332.937471	-3.813614	R2211	0.964	0.514	V16	B.15
110.832129	-73.443359	SM0723	0.919	0.218	S21	B.14	332.935563	-3.813725	R2211	1.000	0.998	V17	B.15
110.829746	-73.443423	SM0723	0.896	0.666	S22	B.14	332.948259	-3.813784	R2211	0.264	0.321	V18	B.15
110.833941	-73.443915	SM0723	0.433	0.333	S23	B.14	332.948100	-3.814232	R2211	0.258	0.476	V19	B.15
110.794150	-73.444223	SM0723	0.605	0.630	S24	B.14	332.939898	-3.814730	R2211	0.392	0.331	V20	B.15
110.810849	-73.444244	SM0723	0.646	0.690	S25	B.14	332.938427	-3.815240	R2211	0.645	1.000	V21	B.15
110.903105	-73.444390	SM0723	0.983	0.399	S26	B.14	332.945707	-3.814471	R2211	0.278	0.225	V22	B.15
8.036738	18.149779	R0032	0.571	0.243	T1	B.14	332.955343	-3.814889	R2211	1.000	0.654	V23	B.15
8.038601	18.149456	R0032	1.000	0.978	T2	B.14	332.933061	-3.815107	R2211	0.992	0.507	V24	B.15
8.046262	18.148888	R0032	0.985	0.318	T3	B.14	332.940491	-3.814992	R2211	0.302	0.666	V25	B.15
8.055383	18.147672	R0032	0.999	0.667	T4	B.14	332.948160	-3.816034	R2211	0.401	0.316	V26	B.15
8.052479	18.147878	R0032	0.399	0.292	T5	B.14	332.941155	-3.815918	R2211	0.414	0.396	V27	B.15
8.050332	18.147844	R0032	0.283	0.772	T6	B.14	332.949352	-3.816439	R2211	0.888	0.667	V28	B.15
8.055651	18.147573	R0032	0.990	0.334	T7	B.14	332.937394	-3.816302	R2211	0.999	0.331	V29	B.15
8.034867	18.147451	R0032	0.995	0.338	T8	B.14	332.934683	-3.817371	R2211	0.988	0.307	V30	B.15
8.053493	18.147262	R0032	0.658	0.894	T9	B.14	332.934958	-3.816604	R2211	1.000	0.367	V31	B.15
8.057944	18.146889	R0032	0.724	0.860	T10	B.14	332.934967	-3.816757	R2211	1.000	0.994	V32	B.15
8.058874	18.147045	R0032	0.374	0.509	T11	B.14	332.950377	-3.816667	R2211	1.000	0.201	V33	B.15
8.059349	18.147042	R0032	0.299	0.301	T12	B.14	29.969893	-8.813222	M0159	0.999	0.671	W1	B.15
8.043436	18.146547	R0032	0.294	0.255	T13	B.14	29.963165	-8.813345	M0159	0.688	0.339	W2	B.15
8.046783	18.146416	R0032	0.981	0.409	T14	B.14	29.944651	-8.813286	M0159	1.000	0.692	W3	B.15
8.039010	18.146019	R0032	0.972	0.296	T15	B.14	29.960498	-8.813415	M0159	1.000	0.928	W4	B.15
8.052291	18.146047	R0032	0.992	0.325	T16	B.14	29.955329	-8.813938	M0159	0.339	0.352	W5	B.15
8.065649	18.145925	R0032	0.994	0.334	T17	B.14	29.961945	-8.814059	M0159	0.391	0.427	W6	B.15

Table B.11: Continuing of Tab. B.10.

RA	DEC	Cluster	Pr_{VGG}	Pr_{SC-VGG}	ref	Fig.	RA	DEC	Cluster	Pr_{VGG}	Pr_{SC-VGG}	ref	Fig.
29.969867	-8.815333	M0159	0.999	0.303	W7	B.15	88.339031	-33.693556	M0553	1.000	0.883	Y8	B.16
29.956619	-8.815032	M0159	0.967	0.669	W8	B.15	88.337511	-33.693345	M0553	0.984	0.981	Y9	B.16
29.963584	-8.814993	M0159	0.999	0.315	W9	B.15	88.332790	-33.694259	M0553	0.544	0.305	Y10	B.16
29.946252	-8.815569	M0159	0.826	0.465	W10	B.15	88.362085	-33.695670	M0553	0.960	0.342	Y11	B.16
29.960936	-8.815975	M0159	1.000	0.668	W11	B.15	88.338518	-33.696419	M0553	0.844	0.299	Y12	B.16
29.950093	-8.816147	M0159	0.297	0.317	W12	B.15	88.342484	-33.696519	M0553	1.000	0.668	Y13	B.16
29.959910	-8.817069	M0159	1.000	0.755	W13	B.15	88.334992	-33.696018	M0553	0.984	0.381	Y14	B.16
29.965087	-8.817241	M0159	0.979	0.676	W14	B.15	88.337029	-33.697135	M0553	0.573	0.300	Y15	B.16
29.955378	-8.818280	M0159	0.214	0.432	W15	B.15	88.357152	-33.697193	M0553	0.576	0.769	Y16	B.16
29.951784	-8.816701	M0159	0.953	0.410	W16	B.15	88.333265	-33.697596	M0553	0.977	0.840	Y17	B.16
29.941304	-8.819185	M0159	0.740	0.524	W17	B.15	88.357210	-33.697889	M0553	0.999	0.828	Y18	B.16
29.972749	-8.818741	M0159	0.362	0.210	W18	B.15	88.362645	-33.698459	M0553	0.878	0.355	Y19	B.16
29.963686	-8.820711	M0159	0.638	0.351	W19	B.15	88.359834	-33.699040	M0553	0.452	0.394	Y20	B.16
29.962907	-8.818324	M0159	1.000	0.347	W20	B.15	88.360638	-33.699873	M0553	0.251	0.509	Y21	B.16
29.963878	-8.818891	M0159	0.661	0.305	W21	B.15	88.324509	-33.699484	M0553	0.992	0.336	Y22	B.16
29.958514	-8.819400	M0159	1.000	1.000	W22	B.15	88.357361	-33.699504	M0553	1.000	0.343	Y23	B.16
29.962156	-8.818672	M0159	0.268	0.299	W23	B.15	43.574005	-58.927629	S0254	1.000	0.208	Z1	B.16
29.951979	-8.820668	M0159	0.619	0.446	W24	B.15	43.575589	-58.930225	S0254	1.000	0.366	Z2	B.16
29.942636	-8.820841	M0159	1.000	0.381	W25	B.15	43.570747	-58.933812	S0254	0.270	0.333	Z3	B.16
29.953744	-8.818570	M0159	0.670	0.551	W26	B.15	43.579072	-58.933953	S0254	1.000	0.916	Z4	B.16
29.972130	-8.822329	M0159	0.458	0.295	W27	B.15	43.578162	-58.933132	S0254	0.251	0.308	Z5	B.16
29.962355	-8.819547	M0159	0.387	0.642	W28	B.15	43.570004	-58.931458	S0254	0.244	0.346	Z6	B.16
59.726259	-29.909934	A3192	0.225	0.347	X1	B.15	43.565137	-58.933097	S0254	0.678	0.366	Z7	B.16
59.728679	-29.912681	A3192	1.000	0.666	X2	B.15	43.555985	-58.933565	S0254	0.904	0.569	Z8	B.16
59.719663	-29.913937	A3192	0.541	0.356	X3	B.15	43.580422	-58.934486	S0254	0.901	0.295	Z9	B.16
59.720985	-29.913443	A3192	0.241	0.656	X4	B.15	43.565955	-58.934592	S0254	0.993	0.646	Z10	B.16
59.723888	-29.915539	A3192	0.604	0.614	X5	B.15	43.564410	-58.935735	S0254	0.947	0.354	Z11	B.16
59.716934	-29.914192	A3192	1.000	0.987	X6	B.15	43.570150	-58.935499	S0254	0.921	0.674	Z12	B.16
59.724025	-29.914276	A3192	0.878	0.620	X7	B.15	43.558560	-58.935629	S0254	0.358	0.360	Z13	B.16
59.732350	-29.913922	A3192	0.262	0.728	X8	B.15	43.578367	-58.935657	S0254	0.705	0.333	Z14	B.16
59.719884	-29.913063	A3192	0.996	0.992	X9	B.15	43.574890	-58.935780	S0254	0.866	0.266	Z15	B.16
59.724292	-29.914442	A3192	0.998	0.268	X10	B.15	43.551766	-58.936129	S0254	0.318	0.372	Z16	B.16
59.718992	-29.915672	A3192	0.866	0.302	X11	B.15	43.556949	-58.939099	S0254	0.973	0.264	Z17	B.16
59.736897	-29.915681	A3192	0.870	0.460	X12	B.15	43.564599	-58.940267	S0254	0.444	0.331	Z18	B.16
59.724900	-29.913664	A3192	0.245	0.844	X13	B.15	43.561956	-58.939889	S0254	0.945	0.239	Z19	B.16
59.718010	-29.915815	A3192	0.386	0.260	X14	B.15	43.561417	-58.939778	S0254	0.991	0.352	Z20	B.16
59.728948	-29.914017	A3192	0.424	0.282	X15	B.15	43.547532	-58.939242	S0254	0.841	0.258	Z21	B.16
59.709115	-29.917094	A3192	1.000	0.999	X16	B.15	43.563404	-58.940977	S0254	0.874	0.380	Z22	B.16
59.725469	-29.917758	A3192	0.985	0.877	X17	B.15	43.548926	-58.939262	S0254	1.000	0.965	Z23	B.16
59.722855	-29.918200	A3192	1.000	0.819	X18	B.15	43.561321	-58.940513	S0254	0.855	0.276	Z24	B.16
59.734213	-29.914696	A3192	0.991	0.573	X19	B.15	43.570944	-58.938469	S0254	0.524	0.331	Z25	B.16
59.727098	-29.919150	A3192	0.720	0.607	X20	B.15	43.580027	-58.937881	S0254	0.990	0.335	Z26	B.16
59.730697	-29.918834	A3192	1.000	0.653	X21	B.15	43.576554	-58.940151	S0254	1.000	1.000	Z27	B.16
59.737041	-29.919278	A3192	0.806	0.666	X22	B.15	43.577089	-58.939562	S0254	0.896	0.495	Z28	B.16
59.738681	-29.919427	A3192	0.337	0.678	X23	B.15	43.571522	-58.937704	S0254	0.944	0.664	Z29	B.16
59.737525	-29.920146	A3192	0.409	0.221	X24	B.15	43.586093	-58.940390	S0254	1.000	0.550	Z30	B.16
59.722413	-29.919727	A3192	0.974	0.316	X25	B.15	43.566343	-58.941888	S0254	0.934	0.427	Z31	B.16
59.721118	-29.920113	A3192	0.378	0.679	X26	B.15	93.986844	-57.803065	S0615	0.202	0.616	AA1	B.16
59.733235	-29.919868	A3192	1.000	0.337	X27	B.15	93.976855	-57.799307	S0615	1.000	0.718	AA2	B.16
59.730215	-29.919435	A3192	0.973	0.659	X28	B.15	93.974158	-57.799626	S0615	0.997	0.576	AA3	B.16
59.728156	-29.919593	A3192	0.289	0.344	X29	B.15	93.977906	-57.797646	S0615	1.000	0.446	AA4	B.16
59.734750	-29.919779	A3192	0.950	0.665	X30	B.15	93.974835	-57.797083	S0615	0.972	0.551	AA5	B.16
59.723670	-29.919979	A3192	0.420	0.557	X31	B.15	93.954261	-57.793606	S0615	0.989	0.326	AA6	B.16
59.731197	-29.920052	A3192	0.721	0.404	X32	B.15	93.953270	-57.793209	S0615	0.872	0.521	AA7	B.16
59.731973	-29.920883	A3192	0.205	0.728	X33	B.15	93.958712	-57.795099	S0615	1.000	1.000	AA8	B.16
59.735798	-29.920785	A3192	0.445	0.283	X34	B.15	93.945177	-57.791392	S0615	1.000	0.361	AA9	B.16
59.725362	-29.921053	A3192	1.000	0.989	X35	B.15	93.991599	-57.790049	S0615	0.202	0.332	AA10	B.16
88.331720	-33.688137	M0553	1.000	0.835	Y1	B.16	93.963460	-57.789227	S0615	1.000	0.986	AA11	B.16
88.345715	-33.690849	M0553	0.985	0.667	Y2	B.16	93.952538	-57.788006	S0615	1.000	0.668	AA12	B.16
88.341234	-33.691706	M0553	1.000	0.835	Y3	B.16	229.688934	-81.514530	P308	0.568	0.216	AB1	B.17
88.329983	-33.693396	M0553	0.990	0.264	Y4	B.16	229.683893	-81.487553	P308	0.998	0.938	AB2	B.17
88.331393	-33.692017	M0553	0.997	0.365	Y5	B.16	229.733949	-81.486810	P308	1.000	0.347	AB3	B.17
88.334965	-33.692342	M0553	1.000	0.830	Y6	B.16	229.697560	-81.488074	P308	1.000	0.366	AB4	B.17
88.328215	-33.692657	M0553	0.911	0.215	Y7	B.16	229.695916	-81.489011	P308	0.761	0.239	AB5	B.17

Table B.12: Continuing of Tab. B.11.

RA	DEC	Cluster	Pr_{VGG}	$Pr_{SC\text{-}VGG}$	ref	Fig.	RA	DEC	Cluster	Pr_{VGG}	$Pr_{SC\text{-}VGG}$	ref	Fig.
229.762961	-81.490642	P308	0.999	0.260	AB6	B.17	289.267327	-33.514765	P004	1.000	0.335	AC35	B.17
229.733966	-81.491011	P308	1.000	0.680	AB7	B.17	289.269847	-33.514431	P004	0.965	0.882	AC36	B.17
229.677095	-81.491134	P308	0.816	0.333	AB8	B.17	289.271915	-33.514639	P004	0.998	0.666	AC37	B.17
229.760154	-81.491500	P308	0.263	0.635	AB9	B.17	289.259681	-33.514345	P004	0.896	0.348	AC38	B.17
229.705934	-81.492081	P308	0.879	0.222	AB10	B.17	289.267509	-33.515316	P004	0.541	0.333	AC39	B.17
229.707530	-81.492375	P308	0.377	0.655	AB11	B.17	289.271121	-33.517015	P004	0.804	0.674	AC40	B.17
229.726832	-81.492266	P308	0.283	0.321	AB12	B.17	289.262945	-33.515016	P004	0.778	0.334	AC41	B.17
229.737264	-81.492243	P308	0.916	0.535	AB13	B.17	289.278089	-33.515303	P004	0.895	0.993	AC42	B.17
229.675872	-81.492711	P308	0.534	0.428	AB14	B.17	289.274790	-33.514919	P004	0.985	0.766	AC43	B.17
229.702828	-81.493225	P308	0.986	0.331	AB15	B.17	289.284722	-33.513645	P004	0.933	0.941	AC44	B.17
229.760383	-81.493963	P308	0.840	0.667	AB16	B.17	289.273189	-33.514407	P004	1.000	0.669	AC45	B.17
229.728141	-81.494126	P308	0.779	1.000	AB17	B.17	289.284455	-33.514593	P004	1.000	0.955	AC46	B.17
229.733976	-81.495567	P308	0.903	0.335	AB18	B.17	289.265681	-33.516325	P004	0.439	0.699	AC47	B.17
229.775275	-81.495998	P308	0.942	0.969	AB19	B.17	289.274868	-33.516479	P004	0.997	0.980	AC48	B.17
229.721945	-81.496847	P308	0.999	0.667	AB20	B.17	289.264398	-33.517122	P004	0.965	0.329	AC49	B.17
229.657885	-81.496314	P308	0.996	0.338	AB21	B.17	289.271746	-33.513807	P004	0.890	0.535	AC50	B.17
229.693787	-81.496756	P308	0.300	0.308	AB22	B.17	289.265373	-33.517447	P004	0.987	0.666	AC51	B.17
229.742174	-81.498625	P308	0.986	0.334	AB23	B.17	289.262628	-33.517478	P004	0.667	0.439	AC52	B.17
229.736026	-81.499647	P308	0.966	0.333	AB24	B.17	289.260920	-33.517860	P004	0.786	0.999	AC53	B.17
229.730087	-81.497850	P308	0.456	0.333	AB25	B.17	289.263008	-33.517980	P004	0.955	0.210	AC54	B.17
229.635020	-81.497320	P308	0.977	0.321	AB26	B.17	289.262888	-33.518037	P004	0.660	0.364	AC55	B.17
229.768562	-81.497856	P308	0.926	0.333	AB27	B.17	289.265225	-33.519544	P004	0.999	0.724	AC56	B.17
229.699976	-81.498355	P308	1.000	0.318	AB28	B.17	289.267405	-33.519569	P004	1.000	0.998	AC57	B.17
229.695387	-81.497873	P308	0.946	0.646	AB29	B.17	289.295169	-33.518503	P004	1.000	0.799	AC58	B.17
229.691409	-81.498933	P308	0.983	0.336	AB30	B.17	289.275504	-33.518683	P004	0.228	0.430	AC59	B.17
229.689462	-81.499706	P308	0.986	0.339	AB31	B.17	289.276079	-33.518743	P004	0.991	0.821	AC60	B.17
229.808439	-81.498408	P308	0.677	0.368	AB32	B.17	289.259485	-33.520921	P004	0.356	0.224	AC61	B.17
229.671246	-81.497870	P308	0.988	0.629	AB33	B.17	4.638111	16.417528	R0018	0.800	0.270	AD1	B.18
229.766630	-81.498558	P308	0.942	0.409	AB34	B.17	4.640558	16.416644	R0018	1.000	0.667	AD2	B.18
229.743960	-81.499291	P308	0.847	0.289	AB35	B.17	4.628445	16.418617	R0018	1.000	0.666	AD3	B.18
289.268263	-33.502597	P004	1.000	0.996	AC1	B.17	4.636089	16.418924	R0018	0.539	0.451	AD4	B.18
289.264746	-33.503127	P004	0.877	0.534	AC2	B.17	4.634200	16.418606	R0018	0.987	0.334	AD5	B.18
289.265208	-33.504694	P004	0.683	0.652	AC3	B.17	4.631823	16.449216	R0018	0.999	0.665	AD6	B.18
289.267787	-33.505382	P004	0.766	0.994	AC4	B.17	4.631542	16.448977	R0018	1.000	0.298	AD7	B.18
289.280296	-33.506429	P004	0.883	0.537	AC5	B.17	4.636183	16.448870	R0018	0.999	0.335	AD8	B.18
289.261098	-33.506541	P004	0.843	0.971	AC6	B.17	4.640629	16.448155	R0018	0.999	0.311	AD9	B.18
289.266844	-33.508021	P004	0.973	0.245	AC7	B.17	4.630036	16.448047	R0018	0.451	0.626	AD10	B.18
289.277188	-33.508067	P004	0.911	0.562	AC8	B.17	4.649228	16.448333	R0018	1.000	0.565	AD11	B.18
289.278724	-33.508175	P004	1.000	0.999	AC9	B.17	4.648767	16.448261	R0018	0.997	0.353	AD12	B.18
289.271212	-33.507800	P004	0.247	0.259	AC10	B.17	4.637701	16.448067	R0018	0.319	0.314	AD13	B.18
289.269597	-33.509748	P004	0.683	0.332	AC11	B.17	4.634978	16.448334	R0018	0.997	0.244	AD14	B.18
289.272744	-33.510036	P004	0.947	0.341	AC12	B.17	4.642577	16.448321	R0018	1.000	0.332	AD15	B.18
289.259615	-33.510396	P004	0.314	0.443	AC13	B.17	4.650537	16.447164	R0018	0.997	0.660	AD16	B.18
289.269460	-33.510761	P004	0.759	0.314	AC14	B.17	4.651958	16.448059	R0018	1.000	0.669	AD17	B.18
289.278255	-33.510990	P004	0.209	0.433	AC15	B.17	4.651472	16.447559	R0018	0.283	0.347	AD18	B.18
289.259296	-33.511166	P004	0.935	0.664	AC16	B.17	4.638962	16.447779	R0018	0.997	0.290	AD19	B.18
289.260395	-33.511607	P004	0.468	0.742	AC17	B.17	4.627094	16.447631	R0018	0.865	0.320	AD20	B.18
289.271911	-33.512456	P004	1.000	1.000	AC18	B.17	4.651420	16.447051	R0018	0.988	0.268	AD21	B.18
289.277592	-33.511932	P004	0.558	0.691	AC19	B.17	4.646250	16.446962	R0018	0.999	0.324	AD22	B.18
289.266263	-33.511943	P004	0.312	0.939	AC20	B.17	4.626265	16.446785	R0018	1.000	0.344	AD23	B.18
289.273956	-33.512782	P004	1.000	0.667	AC21	B.17	4.647046	16.446509	R0018	0.849	0.333	AD24	B.18
289.277111	-33.513385	P004	1.000	0.743	AC22	B.17	4.632581	16.446228	R0018	1.000	0.966	AD25	B.18
289.284706	-33.512975	P004	0.388	0.987	AC23	B.17	4.639124	16.445602	R0018	0.992	0.333	AD26	B.18
289.256348	-33.512940	P004	0.848	0.952	AC24	B.17	4.629930	16.446205	R0018	1.000	0.388	AD27	B.18
289.289576	-33.513562	P004	0.761	0.399	AC25	B.17	4.652814	16.446108	R0018	0.775	0.256	AD28	B.18
289.261938	-33.513941	P004	0.587	0.999	AC26	B.17	4.639442	16.445460	R0018	0.941	0.289	AD29	B.18
289.262687	-33.513591	P004	0.997	0.360	AC27	B.17	4.627114	16.445818	R0018	0.937	0.242	AD30	B.18
289.260361	-33.513339	P004	0.927	0.692	AC28	B.17	4.622135	16.445808	R0018	0.984	0.321	AD31	B.18
289.286884	-33.515284	P004	0.650	0.666	AC29	B.17	4.629878	16.443151	R0018	0.591	0.233	AD32	B.18
289.271067	-33.515473	P004	0.986	0.320	AC30	B.17	4.627585	16.445328	R0018	0.321	0.333	AD33	B.18
289.270644	-33.516327	P004	0.980	0.674	AC31	B.17	4.629517	16.444072	R0018	0.998	0.333	AD34	B.18
289.253847	-33.514585	P004	0.692	0.331	AC32	B.17	4.639300	16.450342	R0018	0.961	0.333	AD35	B.18
289.280101	-33.514731	P004	0.887	0.667	AC33	B.17	4.628272	16.445334	R0018	0.881	0.269	AD36	B.18
289.266953	-33.514313	P004	1.000	0.526	AC34	B.17	4.625257	16.444819	R0018	0.496	0.333	AD37	B.18

Table B.13: Continuing of Tab. B.12.

RA	DEC	Cluster	Pr_{VGG}	$Pr_{SC\text{-}VGG}$	ref	Fig.	RA	DEC	Cluster	Pr_{VGG}	$Pr_{SC\text{-}VGG}$	ref	Fig.
4.632423	16.444267	R0018	1.000	0.717	AD38	B.18	15.723411	-49.272209	A0102	0.981	0.998	AE38	B.18
4.643235	16.445059	R0018	1.000	0.979	AD39	B.18	15.762875	-49.271909	A0102	0.351	0.357	AE39	B.18
4.624582	16.444181	R0018	0.208	0.281	AD40	B.18	15.727086	-49.271864	A0102	1.000	0.333	AE40	B.18
4.635673	16.445253	R0018	0.476	0.313	AD41	B.18	15.758277	-49.271854	A0102	0.997	0.319	AE41	B.18
4.647043	16.445444	R0018	1.000	0.661	AD42	B.18	15.735657	-49.271786	A0102	1.000	0.202	AE42	B.18
4.641571	16.443584	R0018	1.000	0.333	AD43	B.18	24.346045	-8.437080	W0137	0.611	0.657	AF1	B.19
4.620435	16.443779	R0018	0.749	0.333	AD44	B.18	24.343247	-8.436651	W0137	1.000	0.333	AF2	B.19
4.639015	16.443877	R0018	0.379	0.212	AD45	B.18	24.345894	-8.438249	W0137	0.988	0.999	AF3	B.19
4.646416	16.443746	R0018	0.950	0.321	AD46	B.18	24.341976	-8.439406	W0137	0.734	0.332	AF4	B.19
4.650429	16.443751	R0018	0.996	0.341	AD47	B.18	24.341967	-8.440675	W0137	1.000	0.991	AF5	B.19
4.650545	16.443261	R0018	0.381	0.538	AD48	B.18	24.356057	-8.441498	W0137	0.999	0.342	AF6	B.19
4.646573	16.449744	R0018	0.897	0.400	AD49	B.18	24.355217	-8.441927	W0137	0.270	0.383	AF7	B.19
4.621677	16.443658	R0018	0.856	0.626	AD50	B.18	24.363809	-8.442644	W0137	0.994	0.667	AF8	B.19
4.626820	16.443231	R0018	0.998	0.218	AD51	B.18	24.372027	-8.442642	W0137	0.398	0.355	AF9	B.19
4.630465	16.442819	R0018	0.998	0.222	AD52	B.18	24.357117	-8.442711	W0137	0.325	0.413	AF10	B.19
4.645386	16.443097	R0018	0.748	0.316	AD53	B.18	24.345480	-8.443148	W0137	0.995	1.000	AF11	B.19
4.629645	16.444763	R0018	1.000	0.537	AD54	B.18	24.345509	-8.443452	W0137	0.998	0.670	AF12	B.19
4.646007	16.442448	R0018	0.870	0.281	AD55	B.18	24.344954	-8.443348	W0137	0.993	0.310	AF13	B.19
4.641977	16.442762	R0018	1.000	0.332	AD56	B.18	24.343053	-8.443251	W0137	0.953	0.341	AF14	B.19
4.626080	16.442053	R0018	1.000	0.333	AD57	B.18	24.360774	-8.443541	W0137	0.982	0.333	AF15	B.19
4.638662	16.442014	R0018	1.000	0.339	AD58	B.18	24.363839	-8.444494	W0137	0.731	0.313	AF16	B.19
4.658798	16.442134	R0018	0.968	0.382	AD59	B.18	24.354650	-8.444528	W0137	0.686	0.349	AF17	B.19
4.629318	16.442080	R0018	0.282	0.333	AD60	B.18	24.357497	-8.444641	W0137	0.772	0.332	AF18	B.19
4.652313	16.441750	R0018	0.997	0.215	AD61	B.18	24.344602	-8.445021	W0137	0.232	0.962	AF19	B.19
4.647441	16.442342	R0018	1.000	0.986	AD62	B.18	24.366638	-8.445276	W0137	0.945	0.998	AF20	B.19
4.628606	16.440959	R0018	1.000	0.664	AD63	B.18	24.358397	-8.445070	W0137	0.682	0.309	AF21	B.19
4.628057	16.441327	R0018	1.000	0.239	AD64	B.18	24.372270	-8.445366	W0137	0.662	0.431	AF22	B.19
15.747497	-49.288363	A0102	0.998	0.569	AE1	B.18	24.356285	-8.445614	W0137	0.999	0.334	AF23	B.19
15.751036	-49.283926	A0102	0.790	0.540	AE2	B.18	110.599394	7.429031	P209	0.247	0.556	AG1	B.19
15.756001	-49.285065	A0102	0.505	0.536	AE3	B.18	110.594804	7.429163	P209	0.502	0.668	AG2	B.19
15.754326	-49.283525	A0102	0.997	0.336	AE4	B.18	110.597103	7.426830	P209	0.327	0.390	AG3	B.19
15.741512	-49.283777	A0102	0.996	0.634	AE5	B.18	110.600199	7.426543	P209	0.816	0.263	AG4	B.19
15.748365	-49.282203	A0102	0.663	0.561	AE6	B.18	110.599546	7.423701	P209	0.396	0.498	AG5	B.19
15.749764	-49.287315	A0102	0.983	0.325	AE7	B.18	110.598643	7.423538	P209	1.000	0.672	AG6	B.19
15.757187	-49.283757	A0102	0.987	1.000	AE8	B.18	110.593232	7.423150	P209	0.998	0.502	AG7	B.19
15.736601	-49.281107	A0102	0.992	0.332	AE9	B.18	110.596109	7.422303	P209	1.000	0.669	AG8	B.19
15.758512	-49.280717	A0102	0.820	0.660	AE10	B.18	110.604339	7.422300	P209	0.810	0.333	AG9	B.19
15.728561	-49.281072	A0102	0.252	0.360	AE11	B.18	110.603650	7.420963	P209	0.431	0.298	AG10	B.19
15.749822	-49.280469	A0102	1.000	0.659	AE12	B.18	110.602119	7.420704	P209	0.840	0.667	AG11	B.19
15.757079	-49.280365	A0102	0.439	0.662	AE13	B.18	110.600499	7.419679	P209	0.971	0.334	AG12	B.19
15.729662	-49.280268	A0102	0.251	0.239	AE14	B.18	110.600980	7.417047	P209	0.990	0.420	AG13	B.19
15.750815	-49.279652	A0102	0.987	0.226	AE15	B.18	110.602728	7.419524	P209	0.685	0.496	AG14	B.19
15.736055	-49.281462	A0102	1.000	0.333	AE16	B.18	110.596634	7.420050	P209	0.927	0.351	AG15	B.19
15.727456	-49.278875	A0102	0.204	0.326	AE17	B.18	110.609209	7.418805	P209	0.778	0.216	AG16	B.19
15.750832	-49.277345	A0102	0.446	0.644	AE18	B.18	110.602065	7.416801	P209	1.000	0.331	AG17	B.19
15.746601	-49.277344	A0102	0.391	0.629	AE19	B.18	110.611788	7.416489	P209	1.000	0.491	AG18	B.19
15.750003	-49.277332	A0102	0.447	0.461	AE20	B.18	110.611167	7.416341	P209	0.249	0.719	AG19	B.19
15.741019	-49.278046	A0102	0.879	0.674	AE21	B.18	110.611373	7.415681	P209	1.000	0.545	AG20	B.19
15.730908	-49.277157	A0102	0.964	0.303	AE22	B.18	110.599822	7.414025	P209	0.586	0.309	AG21	B.19
15.732312	-49.276890	A0102	0.216	0.330	AE23	B.18	110.587345	7.414312	P209	1.000	0.428	AG22	B.19
15.722926	-49.276564	A0102	0.821	0.292	AE24	B.18	110.611868	7.413887	P209	0.925	0.456	AG23	B.19
15.741058	-49.276350	A0102	0.996	0.257	AE25	B.18	110.612921	7.412865	P209	0.998	0.362	AG24	B.19
15.761061	-49.275841	A0102	0.997	0.351	AE26	B.18	110.613358	7.411222	P209	0.989	0.308	AG25	B.19
15.752898	-49.281052	A0102	0.906	0.334	AE27	B.18	110.614711	7.412055	P209	0.943	0.321	AG26	B.19
15.736098	-49.275362	A0102	0.925	0.328	AE28	B.18	110.584016	7.417427	P209	1.000	0.942	AG27	B.19
15.747847	-49.275324	A0102	0.625	0.272	AE29	B.18	110.579786	7.412825	P209	1.000	0.778	AG28	B.19
15.746534	-49.274617	A0102	0.601	0.888	AE30	B.18	90.027605	-20.118237	R0600	0.504	0.623	AH1	B.19
15.758474	-49.274466	A0102	0.927	0.352	AE31	B.18	90.049379	-20.115928	R0600	0.986	0.317	AH2	B.19
15.741352	-49.274333	A0102	0.655	0.340	AE32	B.18	90.035957	-20.116439	R0600	0.900	0.667	AH3	B.19
15.755926	-49.274297	A0102	1.000	0.564	AE33	B.18	90.051412	-20.116087	R0600	0.992	0.661	AH4	B.19
15.734411	-49.274219	A0102	1.000	0.649	AE34	B.18	90.037106	-20.117656	R0600	0.592	0.436	AH5	B.19
15.740798	-49.274493	A0102	1.000	0.463	AE35	B.18	90.044546	-20.117673	R0600	0.307	0.221	AH6	B.19
15.719006	-49.274424	A0102	0.870	0.289	AE36	B.18	90.042037	-20.116914	R0600	0.264	0.667	AH7	B.19
15.734381	-49.272584	A0102	0.331	0.333	AE37	B.18	90.042653	-20.117302	R0600	0.357	0.905	AH8	B.19

Table B.14: Continuing of Tab. B.13.

RA	DEC	Cluster	Pr_{VGG}	$Pr_{SC\text{-}VGG}$	ref	Fig.	RA	DEC	Cluster	Pr_{VGG}	$Pr_{SC\text{-}VGG}$	ref	Fig.
90.045772	-20.117582	R0600	0.576	0.245	AH9	B.19	90.048186	-20.118354	R0600	0.994	0.627	AH24	B.19
90.044559	-20.116850	R0600	0.255	0.328	AH10	B.19	90.029621	-20.118742	R0600	0.688	0.332	AH25	B.19
90.038981	-20.117781	R0600	0.996	0.330	AH11	B.19	90.057160	-20.118681	R0600	0.424	0.335	AH26	B.19
90.025945	-20.116590	R0600	0.944	0.826	AH12	B.19	90.028746	-20.118705	R0600	0.210	0.373	AH27	B.19
90.039903	-20.116757	R0600	0.972	0.999	AH13	B.19	90.030092	-20.119584	R0600	0.966	0.423	AH28	B.19
90.054457	-20.117200	R0600	0.988	0.390	AH14	B.19	90.051747	-20.121539	R0600	1.000	0.365	AH29	B.19
90.051865	-20.117657	R0600	1.000	0.996	AH15	B.19	90.050168	-20.118739	R0600	0.773	0.335	AH30	B.19
90.056770	-20.119379	R0600	0.963	0.621	AH16	B.19	90.026315	-20.116969	R0600	1.000	0.660	AH31	B.19
90.036921	-20.116975	R0600	0.982	0.579	AH17	B.19	90.026262	-20.116329	R0600	0.894	0.468	AH32	B.19
90.033996	-20.118433	R0600	0.418	0.324	AH18	B.19	90.046736	-20.119705	R0600	0.794	0.216	AH33	B.19
90.034041	-20.119145	R0600	0.946	0.667	AH19	B.19	90.046411	-20.119161	R0600	0.987	0.347	AH34	B.19
90.036136	-20.117680	R0600	0.975	0.243	AH20	B.19	90.043497	-20.119097	R0600	0.866	0.759	AH35	B.19
90.027826	-20.116841	R0600	0.688	0.748	AH21	B.19	90.052706	-20.118990	R0600	1.000	0.342	AH36	B.19
90.039183	-20.117243	R0600	0.593	0.667	AH22	B.19	90.039637	-20.119513	R0600	0.998	0.351	AH37	B.19
90.032899	-20.115873	R0600	0.954	0.507	AH23	B.19							

Table B.15: Continuing of Tab. B.14.