Astronomy & Astrophysics manuscript no. aanda
October 8, 2020

The search for galaxy cluster members with deep learning of panchromatic HST imaging and extensive spectroscopy

G. Angora[1,3⋆], P. Rosati[1,2,9], M. Brescia[3], A. Mercurio[3], C. Grillo[4,5,6], G. Caminha[7], M. Meneghetti[2], M. Nonino[8], E. Vanzella[2], P. Bergamini[2], A. Biviano[8], and M. Lombardi[4]

[1] Department of Physics and Earth Science of the University of Ferrara, Via Saragat 1, I-44122 Ferrara, Italy
[2] INAF - Astronomical Observatory of Bologna, via Gobetti 93/3, 40129 Bologna, Italy
[3] INAF - Astronomical Observatory of Capodimonte, Salita Moiariello 16, I-80131 Napoli, Italy
[4] Department of Physics of the University of Milano, via Celoria 16, I-20133 Milano, Italy
[5] Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Lyngbyvej 2, DK-2100 Copenhagen, Denmark
[6] INAF - IASF Milano, via A. Corti 12, I-20133 Milano, Italy
[7] Kapteyn Astronomical Institute, University of Groningen, Postbus 800, 9700 AV Groningen, The Netherlands
[8] INAF - Astronomical Observatory of Trieste, via G. B. Tiepolo 11, I-34131, Trieste, Italy
[9] INFN, Sezione di Ferrara, Via Saragat 1, 44122 Ferrara, Italy

Received xxx; accepted xxx

ABSTRACT

Context. The next generation of extensive and data-intensive surveys are bound to produce a vast amount of data, which can be efficiently dealt with using machine-learning and deep-learning methods to explore possible correlations within the multi-dimensional parameter space.

Aims. We explore the classification capabilities of convolution neural networks (CNNs) to identify galaxy cluster members (CLMs) by using Hubble Space Telescope (HST) images of fifteen galaxy clusters at redshift $0.19 \lesssim z \lesssim 0.60$, observed as part of the CLASH and Hubble Frontier Field programmes.

Methods. We used extensive spectroscopic information, based on the CLASH-VLT VIMOS programme combined with MUSE observations, to define the knowledge base. We performed various tests to quantify how well CNNs can identify cluster members on ht basis of imaging information only. Furthermore, we investigated the CNN capability to predict source memberships outside the training coverage, in particular, by identifying CLMs at the faint end of the magnitude distributions.

Results. We find that the CNNs achieve a purity-completeness rate $\gtrsim 90\%$, demonstrating stable behaviour across the luminosity and colour of cluster galaxies, along with a remarkable generalisation capability with respect to cluster redshifts. We concluded that if extensive spectroscopic information is available as a training base, the proposed approach is a valid alternative to catalogue-based methods because it has the advantage of avoiding photometric measurements, which are particularly challenging and time-consuming in crowded cluster cores. As a byproduct, we identified 372 photometric cluster members, with mag(F814)<25, to complete the sample of 812 spectroscopic members in four galaxy clusters RX J2248-4431, MACS J0416-2403, MACS J1206-0847 and MACS J1149+2223.

Conclusions. When this technique is applied to the data that are expected to become available from forthcoming surveys, it will be an efficient tool for a variety of studies requiring CLM selection, such as galaxy number densities, luminosity functions, and lensing mass reconstruction.

Key words. galaxies: clusters: general, galaxies: photometry, galaxies: distances and redshifts, techniques: image processing, methods: data analysis

1. Introduction

Over the past decade, the field of astrophysics has been experiencing a true paradigmatic shift, moving rapidly from relatively small data sets to the big data regime. Dedicated survey telescopes, both ground-based and space-borne, are set to routinely produce tens of terabytes of data of unprecedented quality and complexity on a daily basis. These volumes of data can be dealt with 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 visualisation (Brescia et al. 2018). The paradigms of 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.

Here, we explore the application of ML techniques in the context of studies of galaxy clusters, more specifically, to identify cluster members (CLMs) based on imaging data alone. In fact, obtaining a highly complete sample of spectroscopic members is an extremely expensive and time-consuming task, which can be simplified and accelerated thanks to the use of a limited amount of spectroscopic information in training ML methods.

Disentangling CLMs from background and foreground sources is an essential step in the measurement of physical properties of galaxy clusters, measuring, for example, the galaxy luminosity and stellar mass functions (e.g. Annunziatella et al. 2016, 2017), in addition to studies of the cluster mass distribution via strong and weak lensing techniques (e.g. Caminha et al. 2017b, 2019; Lagattuta et al. 2017; Medezinski et al. 2016).

⋆ e-mail: gius.angora@gmail.com