https://doi.org/10.15388/vu.thesis.114
https://orcid.org/0000-0001-7603-6382

VILNIUS UNIVERSITY
CENTER FOR PHYSICAL SCIENCES AND TECHNOLOGY

Jonas
BIALOPETRAVIČIUS

Analysis of Star Clusters with Convolutional Neural Networks

DOCTORAL DISSERTATION

Natural Sciences,
Physics (N 002)

VILNIUS 2020

This dissertation was written between 2015 – 2019 in Vilnius University. The research was supported by the Research Council of Lithuania.

Academic supervisor – **dr. Donatas Narbutis** (Vilnius University, Natural Sciences, Physics – N 002).

https://doi.org/10.15388/vu.thesis.114
https://orcid.org/0000-0001-7603-6382

VILNIAUS UNIVERSITETAS
FIZINIŲ IR TECHNOLOGIJOS MOKSLŲ CENTRAS

Jonas
BIALOPETRAVIČIUS

Žvaigždžių spiečių analizė konvoliuciniais neuroniniais tinklais

DAKTARO DISERTACIJA

Gamtos mokslai,
Fizika (N 002)

VILNIUS 2020

Disertacija rengta 2015 – 2019 metais Vilniaus universitete. Mokslinius tyrimus rėmė Lietuvos mokslo taryba.

Mokslinis vadovas – **dr. Donatas Narbutis** (Vilniaus universitetas, gamtos mokslai, fizika – N 002).

Abbreviations

ACS – Advanced Camera for Surveys, Hubble Space Telescope instrument;

ANN – Artifical Neural Network;

A_V – interstellar extinction in the V band;

CNN – Convolutional Neural Network;

HST – Hubble Space Telescope;

IMF – Initial Mass Function of a cluster's stars;

MW – Milky Way Galaxy;

PHAT – The Panchromatic Hubble Andromeda Treasury;

r_c – star cluster's core radius of the King model;

r_t – star cluster's tidal radius of the King model;

R_V – total-to-selective extinction factor;

WFC3 – Wide Field Camera 3, Hubble Space Telescope instrument.

Contents

Introduction

Star clusters are gravitationally bound groups of stars that form in molecular clouds. The vast majority of all stars are formed as parts of clusters (Lada & Lada 2003), therefore star clusters are invaluable objects of study for stellar dynamics and evolution. There have been numerous studies on star cluster detection and parameter inference. Prominent examples include Stochastically Lighting Up Galaxies (SLUG) (Krumholz et al. 2015), which is one of the most mature codes in stochastic cluster population simulation and inference, and The Panchromatic Hubble Andromeda Treasury (PHAT) survey (Dalcanton et al. 2012), which is one of the largest homogeneous star cluster studies based on the Hubble Space Telescope (HST) data.

Convolutional neural networks (CNNs) have been established as the go-to method for fast object detection and classification on natural images. As of now, a number of these tasks can be solved with a better accuracy than that of a human (Russakovsky et al. 2015). Examples of image recognition competitions where CNNs have recently driven progress include ILVSCR[1] (Russakovsky et al. 2015), PASCAL VOC[2] (Ren et al. 2017) and Microsoft COCO[3] (Lin et al. 2014).

The uptake of CNN-based methods has also been accelerating in the field of astronomy. Prominent examples include galaxy classification (Dieleman et al. 2015; Huertas-Company et al. 2018; Domínguez Sánchez et al. 2018), gravitational lensing (Petrillo et al. 2017; Lanusse et al. 2018; Pourrahmani et al. 2018) and transient detection (Cabrera-Vives et al. 2017; Lanusse et al. 2018; Sedaghat & Mahabal 2018). There has also been work on astronomical image reconstruction (Flamary 2016), exoplanet identification (Shallue & Vanderburg 2018), and point spread function modeling (Herbel et al. 2018).

[1] ImageNet Large Scale Visual Recognition Competition

[2] The PASCAL (Pattern Analysis, Statistical Modelling and Computational Learning) Visual Object Classes

[3] Common Objects in Context

This opens the door for astrophysical parameter inference on the exponentially increasing amount of sky survey data. However, star cluster parameter estimation has not yet been attempted with these methods, even though the field is an ideal candidate since accurate inference from imaging data is sorely needed. Until now, the vast majority of star cluster analysis was based on integral or resolved stellar photometry, which limits the amount of information that can be extracted from individual pixels of cluster images. The usefulness of such information has been shown in studies such as Whitmore et al. (2011).

This work focuses on the problem of star cluster detection and parameter inference using CNNs. A full pipeline is constructed that can perform star cluster analysis starting from multi-band sky survey data and concluding in evolutionary, structural, and environmental parameters with several metrics describing star cluster detection likelihoods.

The use of the methods created during the course of this work were demonstrated on PHAT data of the M31 galaxy (Dalcanton et al. 2012) as well as the M83 galaxy HST survey (Blair et al. 2014). Cluster age, mass, and size derivation was demonstrated on the M31 galaxy cluster images. Extinction derivation as well as the cluster detection pipeline was demonstrated on the M83 galaxy. The use of different surveys illustrated the applicability of the method for the analysis of data of any galaxy with a small amount of human supervision.

In Chapter 1 we present an overview of current star cluster detection and analysis methods. In Chapter 2 a CNN-based method to perform the inference of star cluster age, mass, and size is presented. In Chapter 3 the method is extended to allow the inference of extinction and a method to perform star cluster detection is introduced. In Chapter 4 the full pipeline for performing star cluster detection and parameter inference is described and used to analyse cluster candidates of the M83 galaxy.

Goal and tasks of the thesis

The goal of this thesis was to develop a method capable of simultaneously deriving evolutionary, structural, and environmental star cluster parameters, as well as localizing them in multi-band images.

To this end the following tasks were formulated:

1. Develop a CNN-based method to infer age, mass, size, and extinction parameters of star clusters.

2. Develop an approach that allows to account for degeneracies between the derived parameters.

3. Develop a method to identify the presence of clusters in multi-band photometric images of galaxies.

4. Develop codes for generating mock images of star clusters used for CNN training.

5. Apply the codes to the M83 galaxy Hubble Space Telescope images to demonstrate a full pipeline for star cluster search and parameter inference.

Statements of the doctoral thesis

1. A CNN-based method suitable to derive evolutionary (age, mass), structural (size), and environmental (extinction) parameters of star clusters and represent their degeneracies was developed, trainable on as little as 50,000 mock cluster image samples with minimal human supervision.

2. The degeneracies between age and extinction can be accounted for by using three HST photometric passbands (F336W, F438W, and F814W) for clusters of age $\log_{10}(t/\mathrm{yr}) < 8$ and mass between $\log_{10}(M/\mathrm{M}_\odot) = 3.5$ and 5.5 in the M83 galaxy.

3. A novel star cluster detection method was proposed, which achieves a recall of $\sim$90% for mock clusters of age $\log_{10}(t/\mathrm{yr}) = 7$ and mass $\log_{10}(M/\mathrm{M}_\odot) = 4$ in the M83 galaxy, utilizing the inference of the likelihood of a cluster's presence in a given field, as well as a proxy for the cluster's signal-to-noise.

4. The analysis of M83 cluster candidates with age $\log_{10}(t/\mathrm{yr}) \geq 8.5$ revealed the presence of more dense clusters near the galactic center, with densities decreasing from $\log(\rho_h/(\mathrm{M}_\odot \cdot \mathrm{pc}^{-3})) = 3.5$ at the distance of 0.5 kpc to $\log(\rho_h/(\mathrm{M}_\odot \cdot \mathrm{pc}^{-3})) = 2$ at the distance of 5 kpc from the galactic center.

5. Using images of the M83 galaxy in F336W, F438W, and F814W HST passbands, 3,380 star cluster candidates were detected and their age, mass, size, and extinction have been determined.

6. An age gradient of M83 cluster candidates with respect to the spiral arms of the galaxy is observed, with the younger population of cluster candidates ($\log_{10}(t/\mathrm{yr}) < 7$) peaking at $\sim$0.4 kpc leading the spiral arms and the older population ($7 \leq \log_{10}(t/\mathrm{yr}) < 7.7$) shifted towards $\gtrsim$0.7 kpc, which is consistent with the density wave theory.

Scientific novelty

1. For the first time a CNN was applied to the problem of star cluster detection and parameter inference.

2. A method to represent star cluster parameter degeneracies in the outputs of a CNN was developed and verified.

3. A method to account for multiple passbands of astronomical observations as well as information present in all individual pixels of the analysed image when performing star cluster parameter inference was presented.

4. A full pipeline for detecting star cluster candidates and inferring their evolutionary, structural, and environmental parameters has been developed and verified.

5. Using images of the M83 galaxy in F336W, F438W, and F814W HST passbands, 3,380 star cluster candidates were detected and their parameter analysis was performed in a self-consistent manner.

Contributions of the author

The adaptation of CNN-based methods for star cluster analysis, the simulation of star clusters, as well as the writing of all of the software created for this purpose was done by the author. The astronomical survey data preparation, analysis of results and the writing of papers was the result of teamwork between co-authors of the respective papers.

Publications included in the thesis

1. **J. Bialopetravičius**, D. Narbutis, V. Vansevičius, Deriving star cluster parameters with convolutional neural networks. I. Age, mass, and size,

Astronomy & Astrophysics, 2019, 621, A103, doi: 10.1051/0004-6361/201833833.

2. **J. Bialopetravičius**, D. Narbutis, Deriving star cluster parameters with convolutional neural networks. II. Extinction and cluster/background classification, Astronomy & Astrophysics, 2020, 633, A148, doi: 10.1051/0004-6361/201936185.

3. **J. Bialopetravičius**, D. Narbutis, Study of star clusters in the M83 galaxy with a convolutional neural network, The Astronomical Journal, 2021, in press, arXiv: 2010.11126, doi: 10.3847/1538-3881/abbf5.

Presentations at conferences

1. **J. Bialopetravičius**, D. Narbutis, V. Vansevičius, Deep learning based star cluster parameter inference, 2019, Machine Learning Prague, Poster Presentation.

2. **J. Bialopetravičius**, D. Narbutis, V. Vansevičius, Žvaigždžių spiečių parametrų nustatymas giliais konvoliuciniais neuroniniais tinklai, 2019, 9–oji LMA jaunųjų mokslininkų konferencija "Fizinių ir technologijos mokslų tarpdalykiniai tyrimai", Vilnius, Poster Presentation.

3. **J. Bialopetravičius**, D. Narbutis, V. Vansevičius, Deriving star cluster parameters with convolutional neural networks, 2019, Open Readings, Vilnius, Poster Presentation.

4. **J. Bialopetravičius**, Deep learning for object detection in space, 2017, Artificial Intelligence Group Meetup, Vilnius, Oral Presentation.

Chapter 1

Overview

1.1 Introduction

In the most general sense star clusters are groupings of stars formed in the same molecular cloud. Because of this the stars in a cluster are of very similar chemical composition and age. With some exceptions, such as the age spread of very young clusters, these values can be assumed to be constant throughout the cluster. In effect the member star properties vary solely as a function of their mass. The vast majority of all stars are formed in clusters (Lada & Lada 2003), therefore they are an invaluable subject of study for stellar dynamics and evolution.

Precisely defining a star cluster from an observational perspective can be quite difficult. As our sensors simply capture noisy projections of actual distributions of stars in space, the question of which stars belong to a cluster and which don't could be hard to answer. In some cases we may not even be able to account for enough stars to be reasonably sure if what we're observing is a cluster, or simply a density enhancement observed purely by chance. In light of this the Encyclopedia of Astronomy and Astrophysics offers to define a star cluster as "an obvious concentration of several stars or more above the surrounding stellar background, apparently localized in space and identifiable on visual or infrared images covering a suitable field of view" (Baugh 2001). This is not a precise definition and does not even attempt to distinguish gravitationaly bound and unbound systems. Nevertheless this gives us a starting point for further categorization.

In this chapter we present an overview of star cluster detection algorithms, most of which focus on resolved Milky Way (MW) clusters. In Section 1.2 we

overview the main cluster types, their properties, useful parametrizations and observational biases. Section 1.3 details cluster detection methods described in literature, mainly dealing with near-infrared surveys and their aggregations, such as MWSC[1] (Kharchenko et al. 2012) and PPMXL[2] (Roeser et al. 2010). In Section 1.4 the limitations of current methods are discussed and an alternative is proposed.

1.2 Star cluster properties

1.2.1 Types

Two main types of star clusters are observed in the Milky Way galaxy: open and globular. These are by no means the only types that are distinguished in observations. For example, searches for open star clusters, especially in the context of surveys in the infrared spectrum, almost unavoidably detect embedded clusters as well. However the question of how we can recognize their differences and categorize them will not be discussed in this work.

Open clusters are found in galaxies where star formation is still active, such as our own. In the case of MW (or spiral galaxies in general) these clusters mainly reside in the spiral arms, where gas densities are highest and most of the star formation happens. These are relatively recently formed systems populated by young, metal-rich stars, with metallicity values approximately in the range of $-0.7 < [\mathrm{Fe/H}] < 0.3$. This also reflects in their color-magnitude diagrams as most of the stars can still be found on the main sequence. The total masses of such clusters ranges roughly anywhere from $10\mathrm{M}_\odot$ to $10^4\mathrm{M}_\odot$(Baugh 2001), with some rare systems having substantially higher masses in the order of $10^5\mathrm{M}_\odot$ (Portegies Zwart et al. 2010a). Spatially, open clusters are relatively sparse systems. This property allows our instruments, at least in the case of observations in the MW, to resolve most of the individual stars. Nevertheless the observations are still troubled by interstellar extinction, as well as the difficult problem of star membership. As most of these clusters are lightly gravitationally bound, their appearance is usually fairly irregular. This not only makes them difficult to find, but also makes it hard to characterize them parametrically, with such parameters as the tidal radius of the King model (i.e. the radius at which the density profile reaches zero), for example. Usually the ap-

[1] Milky Way Star Clusters

[2] Position and Proper Motion eXtended-L

parent angular diameter is measured instead. Most open clusters are yet to be found, as the fact that they're localized in the Galactic disk means that they also have a high chance of being heavily obscured.

Globular clusters, in contrast, are massive, strongly gravitationally bound systems residing mainly in the galactic halo. The mass of these systems ranges anywhere from $10^4 M_\odot$ to $10^6 M_\odot$ and more, such as is the case with the M22 cluster. The stars in globular clusters are distributed roughly spherically, with their density increasing towards the cluster's center. This can often cause problems when trying to resolve individual stars, as the concentration is high enough to hide less luminous stars near the center of a cluster due to image resolution effects. They are composed of old, metal poor stars, with metallicity values roughly between $-2.3 < [Fe/H] < 0.2$ (Baugh 2001).

1.2.2 Spatial distribution and selection effects

Observations of open clusters, embedded in the galactic disk, are especially vulnerable to selection effects. The high density of interstellar dust blocks our view to most of these structures and so we mainly observe open clusters only a few kiloparsecs around the Sun. If we wish to study a complete open cluster sample we have to confine ourselves to an extremely close distance range. For example, the databases compiled by Kharchenko et al. (2005) and Piskunov et al. (2008) are likely complete only up to the 600 pc range. This excludes all of the clusters in the MW's major spiral arms, where cluster formation is more active than in our immediate neighborhood. Nevertheless, we currently know of approximately 3,000 open clusters in our galaxy (Kharchenko et al. 2013). The real number, due to the aforementioned difficulties in observing these clusters, is much larger. Since interstellar dust is most opaque to shorter wavelengths there has been progress in cluster detection in the infrared spectrum. For example the new catalog of Kharchenko et al. (2013) – MWSC – is (mostly) complete up to the 1.8 kpc range. Knowing the large potential number of undiscovered open clusters, increasing amounts of such observations will greatly benefit from automatic analysis.

The interstellar dust blocking our field of view isn't the only effect biasing catalog samples. Open clusters, at least in comparison to globular clusters, are already short lived. At low galactic latitudes clusters can get easily disrupted, especially considering their low mass and the fact that they are loosely gravitationally bound. Closer to the galactic center, where tidal forces are stronger and the chances of collision with molecular clouds are higher, these clusters

tend to get disrupted even earlier. All of this, in effect, produces a distribution skewed towards older clusters further away from the galactic center (Portegies Zwart et al. 2010a).

In comparison the samples of globular clusters are much more complete. The sparse field of the galactic halo allows us to probe most of this type of cluster. It is estimated that there are approximately 200 globular clusters in the MW galaxy, and we have good distance measurements to about 150 of them. Main issue for completeness is again due to extinction. Intuitively most of the globular clusters behind the galactic center cannot be detected. This also applies to the whole galactic bar, as shown by Nikiforov & Smirnova (2013).

The remaining undetected clusters are likely to be obscured by dust on the far side of the Galaxy.

1.2.3 Density distribution profiles

Density profiles are functions connecting the stellar density of a cluster to the distance from its center. They are usually empirically derived functions that allow us to infer a cluster's structural parameters. Density profiles are not used for the detection of clusters directly, but are often a good characteristic for highlighting any anomalous properties that might indicate bad cluster membership assignments.

The King density profile (King 1962) is one of the most widely used in literature. This model fits old globular MW clusters well, due to the truncation of the power law in the outer parts of the profile. This is due to the fact fact that as a cluster ages it keeps losing its outer stars due to tidal effects, while more massive stars slowly concentrate in the middle region. In such cases something like a simple Plummer profile predicts densities that are too high in the outer parts of the cluster. The King model can be characterized by three parameters – central surface brightness I_0, core radius r_c and the tidal radius r_t. The core radius is defined as the radius at which the surface brightness drops to a half of the central value. The King profile is constructed by approximating two power-law functions: $I(r) = k(\frac{1}{r} - \frac{1}{r_t})^2$ and $I(r) = \frac{I_0}{1+(r/r_c)^2}$. The result of this is:

$$I(r) = I_0 \left[\frac{1}{(1+(r/r_c)^2)^{\frac{1}{2}}} - \frac{1}{(1+(r_t/r_c)^2)^{\frac{1}{2}}} \right]^2. \tag{1.1}$$

The Elson-Fall-Freeman (EFF) model (Elson et al. 1987) was derived for fitting young (< 300 Myr) clusters in the Magellanic Clouds. These clusters

don't exhibit the same tidal truncations and in effect are better fit by continuous functions. The EFF profile is defined as:

$$I = I_0 \left(1 + \frac{r^2}{a^2}\right)^{-\gamma/2}, \tag{1.2}$$

where, again, I_0 is the central surface brightness, a is the characteristic radius and γ is the power-law slope.

1.3 Star cluster detection

1.3.1 As overdensities

Star clusters are relatively tightly packed collections of stars. One can then expect for a field of stars containing a cluster to have a higher count of point sources than its immediate neighbourhood. This is one of the most intuitive assumptions to exploit when designing automated detection methods. However, high relative densities by themselves cannot confirm the existence of a cluster and other methods have to be employed for verification. Color information, proper motions and profile fitting all could be used to select only highly probable clusters from a candidate list of overdensities.

Although the idea is sound, star clusters as well as background fields demonstrate a highly variable range of densities. This becomes increasingly relevant the closer we get to the galactic disk. Small open or embedded clusters often are barely visible on a dense field of galactic stars and dust. This requires robust methods that are capable of dealing with such difficult cases uniformly in our observation.

In essence overdensity based methods firstly need a way to estimate star densities, either in predefined regions (e.g. a rectangular grid) or in the immediate neighbourhood of point sources themselves. And secondly some sort of measure defining what counts as a deviation from the background is needed. For the first step there is a variety of graph algorithms that can be applied directly on point source catalogs. A thorough overview of methods that have been applied for the search or analysis of clusters has been done by Schmeja (2011). As noted by the author there isn't much difference between the algorithms when applied to real data, so simpler methods, such as binning of star counts, tend to be of more use. For completeness, however, all four types of algorithms are presented and discussed here.

Binning of star counts

The simplest method to estimate local density enhancements is binning stars in patches of the sky. The patches are usually chosen to be rectangular for ease of computation, sometimes having overlapping edges if robustness to noise is important. "Hot" bins, which have collected a large number of point sources, could then be selected as candidates for further analysis. Besides being very computationally efficient this method also allows us to look at the problem in a more statistically rigorous manner. As we're dealing with a large number of point sources, if we keep our bins small enough the number of stars that fall into a bin could be regarded as following the Poisson distribution. One example of such approach is the work of Mercer et al. (2005). Square bins of 0.01 degrees in size were used. If the average count of stars that fall into a bin is approximately constant throughout all bins, the counts can then be characterized by a single Poisson distribution. This allows us to simply select bins that have a sufficient number of stars as candidates. Although the homogeneity assumption doesn't hold in reality, as densities tend to increase towards the galactic disc and bar, Mercer et al. (2005) still found ~ 30 cluster candidates, half of which matched previously cataloged clusters. An improved, region-adaptive method is proposed to more realistically model fields of varying densities. The improved method assumes areas of $5' \times 5'$ (or $\sim$0.1 degrees) are homogeneous and therefore can be described by single Poisson distributions. The areas over denser fields tend to be more noisy, as the mean of the Poisson distribution is equal to its variance. In order to apply a global threshold of density these values need to be normalized. This is done by estimating the mean and variance of star counts with a circular median filter covering a $3' \times 3'$ area. The bins are then mean-subtracted and divided by the standard deviation. This effectively gives us values following the standard normal distribution. These values could then be searched for deviating densities as before. This simple improvement more than tripled the number of detections compared to the globally homogeneous model.

This method has been extensively applied to MW (Kirsanova et al. 2008; Kumar et al. 2006; Borissova et al. 2005), as well as other galaxies (Karampelas et al. 2009).

The nearest neighbour method

Nearest neighbour methods approximate the density around a point source with its distance to its j-th nearest neighbour. The density estimate, p_j, of said point is the inverse of this distance:

$$p_j = \frac{j-1}{S(r_j)} m, \tag{1.3}$$

where r_j is the distance of a given point source to it's j-th nearest neighbour, $S(r_j)$ is the area of a circle with radius r_j and m is the average mass. Varying j can have drastic effects on density estimates. Small values capture more localized structure, but are susceptible to noise, while large values remedy this in exchange for reduced sensitivity to small density variations. This method has been applied in search for proto-star clusters in the Small Magellanic Cloud by Schmeja et al. (2009). The value of $j = 20$ is used and points exceeding 3σ were selected as cluster candidates.

Voronoi tesselation

Voronoi tesselation divides space into convex polygons, such that all of the polygon edges are on equidistant lines between points. Density then can be estimated by the inverse area of the polygon the point lies on. Although not yet used for star clusters, Voronoi tesselation has been applied to the detection of galaxy clusters by van Breukelen et al. (2006). Much like in the case of the nearest neighbour method, sufficiently deviating densities can then be identified as cluster candidates.

Minimum spanning trees

In graph theory the minimum spanning tree (MST) is a tree connecting all vertices in a graph such that the combined sum of edge weight is minimal. When applied to point sources the edges are simply connections between nearby points (in the sense of Delaunay triangulated edges) weighted by Euclidean distances. Cutting off longer edges can then give you smaller sub-trees that hopefully make up clusters. Unfortunately, this method is not very robust to noise, but nevertheless has been used for cluster identification. Maschberger et al. (2010) used a MST algorithm for cluster identification in simulations. Meanwhile Beerer et al. (2010) has used MST to identify clusters in the star forming region of Cygnus-X in the infrared spectrum.

Though some effort has gone into using MST as a criteria for cluster selection its main use remains in analyzing and describing the structure of clusters. Being formed in molecular clouds of a fractal nature clusters themselves tend to retain some fractal structure. Sanchez & Alfaro (2009) analyzed 16 open clusters in the MW and found a correlation between the Q parameter ($Q = \bar{m}/\bar{s}$, where $\bar{m}$ is mean edge length and $\bar{s}$ is the normalized correlation length) of a cluster's MST and T/r_t, where T is age and r_t is the tidal radius.

Although not directly applicable to detection, methods like this could serve as another indicator of a cluster's existence.

1.3.2 From motions

The stars of a cluster all generally move together. Knowing the velocities of stars it should be possible to segment out these moving groups from the background. This does not, by itself, differentiate between moving associations and actual gravitationally bound clusters, but is still a strong indicator of cluster membership that can be later verified using other methods. Usually full 3D motions aren't available for a large number of object, so we have to make do with proper motions.

Scholz et al. (2015) attempted segmentation with proper motions using the PPXML catalog, which has proper motion information on nearly 900 million objects. The idea is similar to the previously outlined overdensity search – point sources are binned spatially. Scholz et al. (2015) chose 0.25×0.25 deg spatial bins. The main difference being – separate bins are used for objects with different proper motions. These were chosen to fall within the error bars of the majority of proper motion estimates – 15 mas/yr. Clusters can then be identified as "hot" bins of a specific proper motion, indicating a concentration of objects that are co-moving.

A new generation of methods is being developed for Gaia data, which in its first data release contains the positions of over a billion sources (Gaia Collaboration et al. 2016). Castro-Ginard et al. (2018) used positions, proper motions and parallaxes of the Tycho-Gaia Astrometric Solution (Michalik et al. 2015) to perform cluster search using the DBSCAN algorithm (Ester et al. 1996). The clustering algorithm DBSCAN is utilized to search for overdensities in the 5D space of positions, proper motions and parallaxes of point sources. The algorithm requires two parameters: ε and $minPts$. A hypersphere around each point source with radius ε is defined, and if the number of sources that are nearby is greater than or equal to the $minPts$ parameter (as measured by eu-

clidean distance), the points are considered a part of the same cluster. Each parameter is standardized separately, instead of using weights, when calculating the distance. These clusters are then verified by using an artificial neural network, which is trained on CMDs of previously found real clusters available in the Gaia data. The use of DBSCAN for object search in source catalogs is by no means new, being attempted by Caballero & Dinis (2008) for finding overdensities, Wilkinson et al. (2018); Xin-hua et al. (2014) for identifying cluster members. However, the utilization of an artificial neural network for verification of candidates is an important step in automating the whole pipeline, which is essential when working with sources of large amounts of data, such as Gaia.

1.3.3 The cluster membership problem

Detecting purely spatial density enhancements or correlations in proper motions is enough to determine cluster existence only in the most simple of cases. Massive globular clusters in sparse fields, for example, are already quite apparent in images and no further verification might be necessary. Nevertheless even in such cases accurate cluster star membership is useful for further analysis. For smaller clusters this is even more true. The limited number of members can have adverse effects on any parameter determination method. In some cases if accurate fits cannot be obtained, for example, in a CMD, the stars could even be discarded as a non-cluster. In this sense the estimation of parameters is an essential part of cluster detection. If we wish to search for small clusters in dense fields, simply thresholding on density deviations will produce a lot of false positives – random fluctuations in density concentration of background stars. Proper motions aren't always available for a sufficient number of stars, or might not be available at all. Even knowing proper motions is no guarantee that we'll be able to find a separation between background and cluster stars, if it exists, as their errors can be quite large, especially for distant, severely obscured objects.

In light of these difficulties statistical methods are preferable for this task. Not only can we factor noise and measurement errors directly into the model, but also reason about inclusion criteria with varying levels of confidence. These ideas are by no means novel. Vasilevskis et al. (1958) proposed a model where cluster star proper motions are assumed to be distributed by a circular Gaussian distribution, and the field stars – an elliptical Gaussian. Sanders (1971) extended this model in a more statistically rigorous manner, using maximum-likelihood estimation. However, modeling both field stars and

cluster members as Gaussians is problematic. Firstly, because field stars are rarely normally distributed, especially spatially and secondly, because extending these methods with more realistic assumptions quickly yields algorithms that are unwieldy to apply and lack universality.

Spatial and kinematic methods

More recently, Mercer et al. (2005) took a heavily spatial approach to this problem. Clusters are described by two dimensional Gaussians, the mean μ defining their position and the covariance matrix Σ describing their size and ellipticity. The probability density function of star membership in the sky area A at position $\mathbf{x} = (x, y)$ is defined as:

$$f(\mathbf{x}) = \tau_0 \frac{b(\mathbf{x})}{A} + \sum_{i=1}^{k} \left\{ \frac{\tau_i}{2\pi\sqrt{|\Sigma_i|}} \exp\left[-\frac{1}{2} (\mathbf{x} - \mu_i)^T \Sigma_i^{-1} (\mathbf{x} - \mu_i) \right] \right\}, \quad (1.4)$$

where the first term defines the density of the stellar background as described in Section 1.3.1. The second term is a sum of k Gaussians each describing a probable cluster or subcluster, where τ_i is the weight given to each mixture component (τ_0 being the background weight). With N stars being observed in the sky area A, approximately $\tau_0 N$ stars should belong to the background, and $\tau_i N$ stars – to the i-th cluster. The parameters of this model are then found using expectation-maximization (EM), which finds the set of parameters (τ, μ and Σ in this case) which maximize the likelihood of the model for all point sources in the sky area A:

$$L(X|\tau, \mu, \Sigma) = \prod_{j=1}^{N} f(X_j), \quad (1.5)$$

where X_j is the j-th point source.

Since k is a fixed parameter some of the Gaussians can describe non-cluster groupings or just statistical noise. Mercer et al. (2005) used the Bayesian information criterion (BIC) for the removal of these false clusters. Clusters are iteratively removed until the maximal BIC value, defined as $BIC = 2\ln(L) - m\ln(N)$, is found, where m is the number of free parameters. This allowed Mercer et al. (2005) to find 91 cluster candidates in the GLIMPSE survey, 59 of which were confirmed by visual inspection to be open or embedded clusters. A variant of this method has also been applied for clus-

ter detection in the UKIDSS (Solin et al. 2012) as well as the VISTA variables in the Vía Láctea (VVV) survey (Solin et al. 2014).

Cabrera-Cano & Alfaro (1990) proposed a non-parametric method that only assumes that separate groups (cluster members and non-members) exist, but places no constraints on their distributions. This is done using a kernel density estimation method. The probability density function f is estimated with kernel K at point (μ_x, μ_y, x, y) as:

$$f(\mu_x, \mu_y, x, y) = \frac{1}{n h_\mu^2 h_r^2} \sum_{i=1}^{n} K\left(\frac{\mu_x - \mu_{xi}}{h_\mu}, \frac{\mu_y - \mu_{yi}}{h_\mu}, \frac{x - x_i}{h_r}, \frac{y - y_i}{h_r}\right), \quad (1.6)$$

where h_μ and h_r are smoothing parameters. If we make the assumption that positional and kinematic variables are independent, we can factorize the function f, leaving us with $f(\mu_x, \mu_y, x, y) = f_\mu(\mu_x, \mu_y) f_r(x, y)$, where

$$f_\mu(\mu_x, \mu_y) = \frac{1}{n h_\mu^2} \sum_{i=1}^{n} K\left(\frac{\mu_x - \mu_{xi}}{h_\mu}, \frac{\mu_y - \mu_{yi}}{h_\mu}\right) \quad (1.7)$$

and

$$f_r(x, y) = \frac{1}{n h_r^2} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h_r}, \frac{y - y_i}{h_r}\right). \quad (1.8)$$

The kernel chosen by Cabrera-Cano & Alfaro (1990) was Gaussian, defined as $K(\xi) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\xi^2\right\}$. The only free parameters in these equations are h_μ and h_r. They were estimated from data in the maximum likelihood sense. The rest of the algorithm is simple: density estimates are calculated for each star, and those that match a set threshold are counted as cluster members.

Including photometry

Kharchenko et al. (2012, 2013) took a more elaborate approach to the ones discussed previously, putting less emphasis on spatial distribution and instead analyzing star movement and photometric fits. Although Mercer et al. (2005) used EM to iteratively fit their parameters, the method of Kharchenko et al. (2012, 2013) relies on iterative fitting in a more explicit manner. This is because some decisions cannot be stated in probabilistic terms and have to be continuously reevaluated. For example, fitting isochrones accurately requires knowing which stars belong to a cluster, but the isochrone fits themselves are used for calculating membership probabilities. The whole procedure then has to be repeated several times until some sort of convergence is achieved. First,

initial cluster position and size values are taken from catalogs. Only stars in an area r_a around a cluster were analyzed. This radius is defined as $r_a = r_{cl} + r_{add}$, where $r_{add} = 18$ arcmin and r_{cl} is taken from literature ($r_{cl} = r_{add}$ if it's not available). Then each star within this radius is characterized by its membership probability, which is made up of four separate components: spatial probability P_s, kinematic probability P_{kin} and two photometric probabilities P_{JH} and P_{JK}, derived from three near-infrared filters available to the authors (J, H and K_s). The spatial probability is simply defined as 1 within an outer cluster radius r_2 (determined by eye) and 0 everywhere else. The kinematic probability for the i-th star is defined as:

$$P_{kin}^i = \left\{ -\frac{1}{4} \left[\left(\frac{\mu_y^i - \bar{\mu}_y}{\varepsilon_{\mu^i}} \right)^2 + \left(\frac{\mu_y^i - \bar{\mu}_y}{\varepsilon_{\mu^i}} \right)^2 \right] \right\}, \tag{1.9}$$

where $\bar{\mu}$ is the mean proper motion of the whole cluster, μ^i is the proper motion of the i-th star and ε_{μ^i} is the proper motion error with a lower bound of 1.5 mas/yr. For the calculation of photometric probabilities isochrones are fit to two color-magnitude diagrams: $K_s, (J-H)$ and $K_s, (J-K_s)$. For a given K_s magnitude of a star, its assumed that it belongs to a cluster with probability 1, if its color is within the bounds of two smoothed isochrones: one fit for single stars, the other for unresolved binaries of equal mass. The points on these isochrones for a given $K_s$are denoted as c_b and c_r. If the star's color lies outside these bounds, the photometric probabilities are defined as:

$$P_c^i = \exp \left\{ -\frac{1}{2} \left[\frac{\triangle c^i}{\varepsilon_c^i} \right]^2 \right\}, \tag{1.10}$$

where $\triangle c^i = c^i - c_b$ if $c^i < c_b$ and $\triangle c^i = c^i - c_r$ if $c^i > c_r$. ε_c^i is the error of the stars color index c^i.

The final membership probability is then defined as $P = P_s \cdot \min\{P_{kin}, P_{JH}, P_{JK}\}$.

This method shows how more elaborate cluster parameters can be built directly into the member selection criteria.

1.4 Discussion

To date most of the cluster detection algorithms are heavily heuristic approaches. Usually they can be separated into two steps: candidate selection, for

finding probable clusters, and detailed analysis, which is a more labor intensive step for verifying found clusters and calculating star membership probabilities. The first step is done for a few reasons. First of all, performing a detailed analysis on each patch in the sky is computationally infeasible. The number of patches that need to be processed can easily be of the order of 10^6. This is exacerbated by the fact that some steps of cluster verification may need to be done by hand, which would make such an exhaustive pipeline completely impractical. Secondly, most algorithms operate under the assumption that our candidates are clusters with a high probability. This prevents simply extending these methods for the analysis of the whole field, as the vast majority of sky patches don't contain resolved clusters. In order to achieve higher detection rates we should strive to include as much information as possible in the candidate selection step without sacrificing performance. Assuming resolved point sources, generally three types of data are available to us: spatial, kinematic and photometric. As discussed previously, kinematic information is an easy way to identify clusters, with some authors even using this information to perform candidate selection. Unfortunately, even proper motions are not available for a large number of point sources, especially in deeper, less imaged regions of space. As object locations by themselves are a poor source of information, this leaves us with photometric measurements.

Ideally an algorithm would account for all information jointly and for each point source, modeling any dependencies intrinsically instead of using heuristic rules, possibly even leaving only the final verification step to humans. Lately, with increasing computational resources there has been a push for generative statistical models describing whole fields. However, finding the parameters (such as star positions) w.r.t. observations in such models is an intractable problem. There are ways to get approximate answers, such as replacing the posterior distribution with an approximation (as is done in Regier et al. 2015) or sampling from the posterior directly by Markov chain Monte-Carlo. Unfortunately, getting good quality solutions is still notoriously difficult even for relatively simple models. All effort so far has gone into solely modeling point sources and their parameters, sometimes including unresolved structures, such as galaxies. The results look promising for small-scale experiments, but it is unclear how difficult it would be to include more relational information, such as the concept of a cluster. To the best of the author's knowledge there has been no work on this so far. The posterior is difficult to calculate as is without imposing higher level variables that are heavily dependent on various

other parameters.

In the author's opinion a compromise is needed between the purely heuristic, hand-tuned methods and modeling everything as a part of a joint distribution. If we look at detection as a classification problem, discriminative models become a good alternative. We lose the ability to infer the values of variables in an unsupervised manner, but this can be remedied by large amounts of labeled data – cluster and non-cluster samples – which can be easily obtained using tools such as SimClust (Deveikis et al. 2008).

1.5 Conclusion

As a first step most star cluster detection methods found in literature rely on binning stars either spatially or, in some cases, kinematically. The purpose of this is finding density enhancements that are distinct from the surrounding field. In theory this could be easily extended with more sophisticated methods for density estimation. There has been some effort going in this direction, but the performance differences are found to be negligible by most authors.

Density enhancements alone cannot lead to reliable detection methods, as they can be caused by purely random fluctuations in star counts. This is especially true when looking for small clusters in dense fields – which is precisely where most of the as of yet undiscovered clusters are located. These dense candidates are usually verified by eye, before moving on to later stages of the pipeline.

Verifying and collecting the member stars of a cluster could be called the second step of detection. There has been effort in automating this procedure, usually by trying to probabilistically reason about the separation of cluster and field stars in spatial and kinematic terms. Kinematic information is an especially good indicator of cluster membership, but looking ahead a lot of significantly dust-obscured or distant observations are unlikely to have proper motions within any reasonable error bounds. Reasoning about other indicators of membership, such as photometric properties, although limited, has shown good results. A good example are the methods of Kharchenko et al., detailed in Section 1.3.

There is still a lot to be done in terms of aggregating all of the observational information in a plausible and effective manner. Current systems require a lot of human intervention in multiple stages of their pipelines, in effect negating some of benefits of reasoning in a probabilistic manner. The role of people in

judging the results is still far from being replaced, but with increasing amounts of observational data, automation has to play a bigger role in cluster detection. Unavoidably, machine learning methods will have to be an essential part of such solutions. Hand-coded cases and simple heuristics don't lend themselves well to problems of increasing complexity.

Increasingly convolutional neural network models have gained a foothold in computer vision research, oftentimes surpassing methods reliant on hand-coding and heuristics, and sometimes exceeding even human performance on said tasks. CNNs excel in discriminative tasks, which can be extended to encompass cluster detection, classification and parameter inference. Furthermore, CNNs directly use the information from all pixels of an observation instead of relying on compressed data, such as star catalogs. This makes them ideal for unresolved and semi-resolved star clusters as well as different levels of noise and cluster environmental factors, as modelling them only requires data instead of sophisticated, hand-crafted generative models.

Chapter 2

Inference of cluster age, mass, and size

One of the benefits of CNN architectures is that parameter inference tasks can be solved in much the same way as object type classification. This opens up the possibility to perform astrophysical parameter inference from star cluster images both accurately and efficiently, utilizing all of the available information in each pixel of an image. However, the method first has to be trained on either human-annotated or simulated mock observations.

The Panchromatic Hubble Andromeda Treasury (PHAT) (Dalcanton et al. 2012) provides high quality multi-band imaging data[1] along with human-annotated star cluster catalogs (Johnson et al. 2012, 2015). These catalogs have previously been analysed by Fouesneau et al. (2014) and de Meulenaer et al. (2017), who provided cluster age and mass estimates using integrated photometry. Based on resolved star photometry, Johnson et al. (2017) derived cluster ages and masses. Previously, Caldwell et al. (2009) had derived cluster parameters of massive clusters in M31 based on spectroscopy. Sizes were estimated from images using aperture photometry by Johnson et al. (2015). Therefore, the PHAT dataset provides a good basis for testing CNN-based methods.

In this chapter, we propose a CNN architecture to estimate the ages, masses, and sizes of star clusters. The network is trained on realistic mock observations, with backgrounds taken from the PHAT survey. The method is tested on a different set of artificial clusters. The chapter is organized into the following sections: Section 2.1 gives details about the PHAT survey data, Section 2.2 gives a summary of used deep learning based methods, Section 2.3

[1] https://archive.stsci.edu/prepds/phat/

describes the proposed CNN and its training methodology, Section 2.4 presents the results of testing the method on artificial clusters.

2.1 Data

The PHAT survey data is extensively described by Dalcanton et al. (2012). We have used stacked, defect-free mosaic ("brick") images, which are photometrically (pixel values are in counts per second) and astrometrically (with available world coordinate system information) calibrated. Multi-band images for four bricks (19, 20, 21, 22) were obtained from the PHAT archive[2]. The bricks were chosen to be located in the outer part of M31 disk for low stellar background contamination. Although the PHAT survey was conducted in six passbands, only three (F336W, F475W, F814W) were used in this work because of their high signal-to-noise ratio. The frames in these passbands were observed with WFC3 and ACS instruments, and have different PSFs as well as exposure times.

Each passband image of a brick was transformed into a tangential projection with a common scale (0.05 arcsec/pixel) and size, oriented in such a way that the pixel grids of the images are aligned with the north and east directions. First, the world coordinate systems of images were aligned and then pixel values were transformed using the reproject[3] package from Astropy[4], conserving flux. Bicubic interpolation was used, providing appropriately resampled images for CNN training purposes.

The brick images contain a lot of saturated stars and extended objects, which could introduce bias to the CNN training procedure. To deal with this, we masked out real clusters and galaxies listed in catalogs by Johnson et al. (2012, 2015) as well as stars brighter than 18 mag in the G passband of the Gaia catalog (Gaia Collaboration et al. 2016). These regions were omitted when generating backgrounds for artificial clusters.

To generate mock clusters of various ages, masses, and sizes, PARSEC isochrones[5], release 1.2S (Bressan et al. 2012), were used to sample stellar masses according to the Kroupa (2001) initial mass function. A fixed metallicity of $Z = 0.009$ and no interstellar extinction was assumed. From the

[2]https://archive.stsci.edu/prepds/phat/

[3]https://reproject.readthedocs.io/

[4]http://www.astropy.org/

[5]http://stev.oapd.inaf.it/cgi-bin/cmd

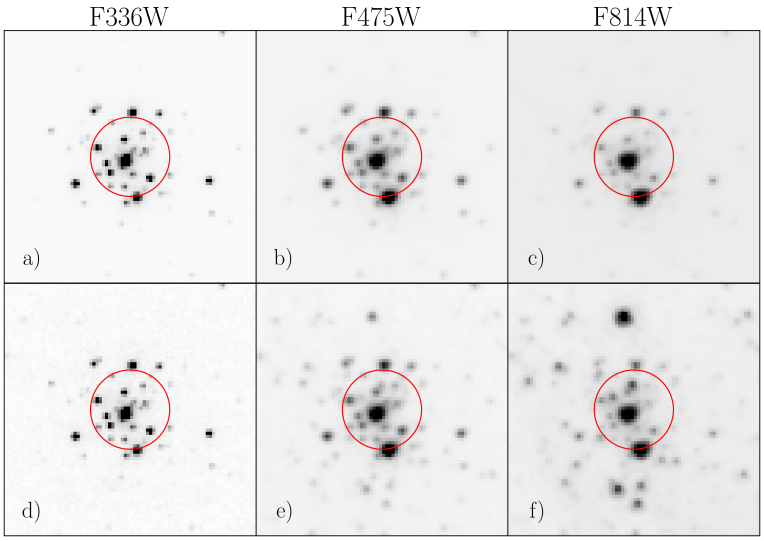

Figure 2.1. An example of a mock cluster drawn with GalSim, without (top) and with a random background from PHAT (bottom). Passbands are indicated on top of each panel. A cluster with typical parameters of $\log_{10}(t/\mathrm{yr}) = 8.0$, $\log_{10}(M/\mathrm{M}_{\odot}) = 3.0$, and $r_h = 0.6$ arcsec is shown. The red circle of r_h radius encloses half of the cluster's stars. The intensity of the images is normalized with the arcsinh function.

isochrones the absolute magnitudes were taken, stars were placed at a distance of M31 (McConnachie et al. 2005, 785 kpc) and their fluxes were transformed to HST counts per second using HST calibrations for ACS (Avila 2017) and WFC3 (Dressel 2012) cameras.

The spatial distribution of stars was sampled from the Elson-Fall-Freeman (EFF) model (Elson et al. 1987), placing the cluster in the center of the image.

TinyTim[6] PSFs (Krist et al. 2011) were used to draw the individual stars of the clusters. The PSFs used were 6×6 arcsec in size and drawn with the GalSim package (Rowe et al. 2015) in the coordinate system of aligned PHAT bricks. Artificial clusters were then placed on backgrounds taken from the transformed PHAT bricks. An example of a mock cluster without and with a background is shown in Fig. 2.1.

[6]http://tinytim.stsci.edu/cgi-bin/tinytimweb.cgi

2.2 Deep learning

2.2.1 Artificial neural networks

For a comprehensive review of artificial neural networks (ANNs) and the algorithms for their training, we refer the reader to Haykin (2009). Here, a short summary is presented as a basis for the methods used in this work.

The convolutional neural network used in this work is a type of ANN, which is a machine learning model comprising multiple artificial neurons, each taking a vector of inputs and producing a single output. The vector of inputs is multiplied by a vector of weights, summed and then passed through a non-linearity, called the activation function. One of the most common activation functions is the rectified linear unit (ReLU), defined as $f(x) = \max(0, x)$ (Nair & Hinton 2010).

An ANN can be trained using the gradient descent algorithm. This requires training samples, which consist of inputs to the network (or observations), and the expected outputs (targets). A loss function $\mathscr{L}$ is used to determine how well the outputs of the network match the expected outputs in the training set. The gradient of this loss function is computed for each sample with respect to the weights of the neural network, and then the weights are adjusted according to the gradient. In essence, the network's weights w_{t+1} at iteration $t+1$ are updated from the weights w_t at iteration t by the formula $w_{t+1} = w_t - \eta \nabla \mathscr{L}$, where η is called the learning rate and is used to control the speed of the learning process. Large values of the learning rate result in quick learning, but the model converges poorly due to large parameter steps near optima; meanwhile, small learning rates have the opposite effect. It is important to control this variable during training so that the training procedure converges.

In the most common form of an ANN, neurons are arranged in multiple layers, where each layer has some number of neurons taking inputs from the previous layer and producing activations for the next layer to process. Such an arrangement is called a feed-forward network. The algorithm used for optimizing such networks is called backpropagation, which provides a convenient way to calculate gradients at each layer using the chain rule. With this algorithm, the inputs are passed through the network, obtaining outputs. Gradients are then calculated and propagated backwards, adjusting the weights layer by layer so that the training error (loss) is minimized.

Since each neuron performs a non-linear mapping of its inputs, stacking many layers of neurons results in a deeply non-linear model, which is bene-

ficial in many real world tasks. However, backpropagation-based algorithms often result in gradients vanishing and learning stalling in the lower layers, or in gradients exploding, which results in divergence. Usually, these problems are solved by either minimizing the number of optimizable parameters with more restrictive neural network architectures, or restricting the possible values of the network's weights that can be obtained during training.

2.2.2 Convolutional neural networks

A convolutional neural network is one way to regularize the weights of a neural network via architectural constraints. In a regular ANN, each neuron takes input from every neuron in the lower layers. Meanwhile, in a CNN each layer consists of learnable convolutional filters with a small number of parameters that exploit the 2D structure of images. Each convolutional filter may be as small as a 3×3 matrix, resulting in 9 optimizable parameters. The filter is applied many times to its input by moving it by a step, called a stride, and a lot of outputs for the next layer are produced. Training such models results in increasingly more abstract feature detectors in the form of convolutional filters. The lower layers look for simple features, such as corners or edges of objects. Deeper layers combine these features to form more abstract concepts about objects present in an image, while discarding irrelevant information and noise. It seems that this hierarchical pipeline of feature extraction is essential to solving computer vision problems, as the neocortex of animals is known to work in a similar way (Kruger et al. 2013).

In classical computer vision, hierarchy was usually implemented by hand. Simple features were first detected and extracted, then combined into more complex aggregates, that can describe whole objects, and later processed with simple machine learning algorithms. These hand-engineered approaches, however, had subpar performance as real world data has a lot of noise, irregularities, irrelevant correlations and a high level of variation. Meanwhile, CNNs learn the necessary regularities from the data itself without any feature engineering. This is also why CNNs are such a favorable algorithm to apply to star clusters.

2.2.3 Residual networks

During the past few years, many variants of CNN architecture have been proposed. It has become quite common in literature to reuse these standardized

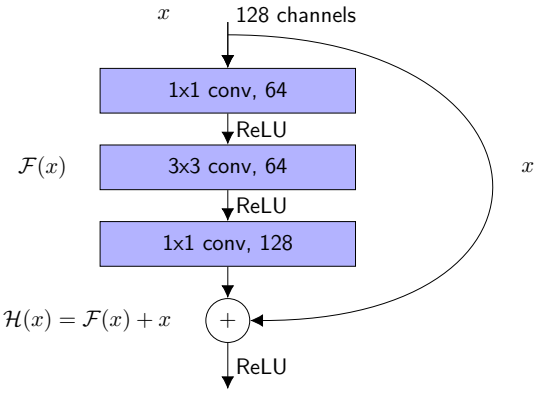

Figure 2.2. Example building block of a residual network (ResNet), consisting of three sequential convolutional layers. The input is a 128-channel activation map, which is passed through 64 1×1 convolutional filters. The filters extract a 64-channel feature map. These features, after applying a ReLU activation, are then passed through 64 3×3 convolutional filters. The purpose of the first layer is to compress the channels for the 3×3 convolutional layer, which results in less optimizable parameters. Then, the ReLU activation is applied again and the final 1×1 convolutional layer expands the number of channels back to 128. Finally, these outputs are summed with the inputs via a skip-connection and passed through a ReLU activation.

architectures in order to have a common ground of reference for measuring improvements.

One such architecture is called the ResNet (He et al. 2016). Convolutional layers in a regular CNN learn feature detectors, taking inputs from lower layers and mapping them to a representation, called a feature map. The idea of a ResNet is that instead of learning to produce a simple feature map, a residual mapping on top of the features of the previous layer is learned. Denoting the inputs of a layer as x, we're looking for a mapping $\mathscr{H}(x)$, such that $\mathscr{F}(x) = \mathscr{H}(x) - x$ (the residual function). This gives the desired mapping as $\mathscr{H}(x) = \mathscr{F}(x) + x$. This has a convenient implementation for CNNs as a skip-connection, which consists of simply taking the activations of a layer and adding it to a deeper layer in the network's graph (see Fig. 2.2).

The ResNet architecture has allowed He et al. (2016) to reach state-of-the-art results on a few standard image recognition datasets with a lower number of

network weights than the alternatives at the time. The training of ResNet type networks is stable regardless of network depth since good results are achieved with networks as shallow as 20 layers (ResNet-20) and as deep as 1202 layers (ResNet-1202). Such networks also seem to work well on object classification as well as detection and regression tasks. Therefore, they are ideal candidates for application in clusters.

2.3 Methods

2.3.1 CNN architecture

In this work, the ResNet-50 version was used as a basis for the constructed CNN. The network was adapted from the variant used for the ImageNet dataset (Russakovsky et al. 2015). The usual inputs of this network are 224×224 pixel natural photographic RGB images (3 channels), which get compressed very quickly in the lower layers to narrow feature mappings, because of the large stride by which the convolutional kernels are moved. Meanwhile, our inputs are 80×80 pixels in size (3 channels: F336W, F475W, F814W), but we operate in a low signal-to-noise ratio regime. Therefore, we reduced the stride of the earliest convolution operations, which should allow the network to extract more low-level information. The network was implemented with the TensorFlow package[7] and is depicted in Figs. 2.3 and 2.4.

Considering that star clusters have rotational symmetry, rotating an image should not impact what the network infers about it. One way to alleviate this problem is with a method proposed by Dieleman et al. (2015): the original image of a cluster is rotated by 90 degrees 3 times and passed through the same convolutional layers. The resulting layer activations are then averaged before the output of the network is computed (2nd block from the bottom in Fig. 2.3). This enforces the idea that a rotated image of a cluster should result in the same high-level representations in the network's activations, as the parameters do not depend on the rotation angle. The outputs of the network were represented as a separate neuron for each parameter (age, mass, and size).

2.3.2 CNN training

The training procedure consists of propagating the images through multiple layers of convolutions until the outputs (in our case – the three cluster

[7] https://www.tensorflow.org/

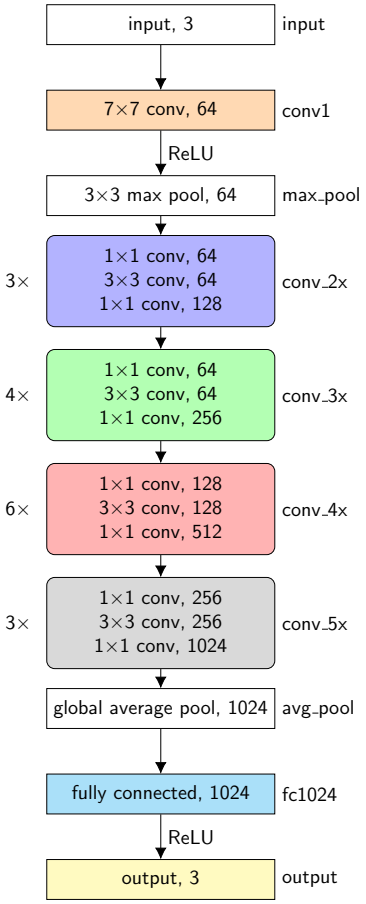

Figure 2.3. A block diagram of the used CNN. The input image of a cluster passes through the network top to bottom, with the output result being age, mass, and size. All blocks with sharp corners depict singular layers, while blocks with rounded corners are groupings of layers (see example in Fig. 2.2), with the number on the left indicating how many times the group is repeated sequentially and the name on the right corresponding to the layer names in Fig. 2.4. The blocks in color are parts of the network with optimizable parameters. The last number in each row is the number of output channels from that layer.

Layer	Output		Operation
	Size	Channels	
input	80×80	3	
conv1	80×80	64	7×7
max_pool	40×40	64	3×3 max pool /2
conv2_x	40×40	128	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 128 \end{bmatrix}\times3$
conv3_x	20×20	256	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times4$
conv4_x	10×10	512	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times6$
conv5_x	5×5	1024	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times3$
avg_pool	1×1	1024	global average pool
fc1024	1×1	1024	fully connected
output	1×1	3	linear

Figure 2.4. The designed 50-layer CNN, based on the ResNet architecture. The network's layers are listed top to bottom, starting from the images of clusters and with the final layer producing the cluster's age, mass, and size. The convolutional layers are actually groups of blocks depicted in Fig. 2.2, with the "_x" in the name standing as a placeholder for the block number. The size of the outputs of each layer, both in spatial dimensions and in channel count, are listed on the second and third columns. The last column lists the operations that each layer performs. The layers or blocks with a stride of 2 are: max_pool, conv3_1, conv4_1, and conv5_1; as can be seen when input and output sizes differ by 2.

parameters) are obtained. These outputs are directly determined by the network's structure and weights. The structure is fixed beforehand; however the weights start out random and need to be optimized in order to construct an accurate inferential model, mapping images of clusters to their parameters.

At each iteration of the training procedure, a batch of 64 clusters (which includes all 3 passband images and their 4 rotated variants) is taken and propagated through the network. The inferred output parameters are then compared to the expected outputs and all the weights are updated according to a loss function:

$$\mathscr{L}(targets, outputs) = \sum_{i=1}^{3} \text{smooth}_{\mathscr{L}_1}(targets_i - outputs_i), \tag{2.1}$$

where *targets* refers to the true values of a given cluster's parameters, while *outputs* refers to the parameters given by the network; i indicates parameter number (age, mass, and size).

For comparing these values and deriving the training gradient for each of the parameters, the following function was used:

$$\text{smooth}_{\mathscr{L}_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1, \\ |x| - 0.5 & \text{otherwise,} \end{cases} \tag{2.2}$$

which is a robust Manhattan distance based loss function (Girshick 2015); it is more resilient to large prediction outliers than the more commonly used mean-squared-error loss. The Adam optimizer (Kingma & Ba 2014) was used to compute the network's weight gradient updates.

The constructed CNN has $\sim$7 million parameters. Even with a large number of training samples, optimizing this many parameters is problematic. Memorizing the images or unimportant peculiarities in the training data can become much easier than learning actual generalized tendencies that produce the output parameters we seek, resulting in overfitting. To combat this, we used the dropout method (Hinton et al. 2012), which discards neurons randomly during training from the network with probability p_{dropout}. An example of this can be seen in Fig. 2.5. This reduces overfitting by preventing neurons from adapting to each other, in effect making the network more robust to missing outputs from neurons in other layers. This was only done during training with $p_{\text{dropout}} = 0.5$, in the last fully connected layer seen in Fig. 2.3 (2nd block from the bottom). For inference, the fully trained network with all of the neurons is used.

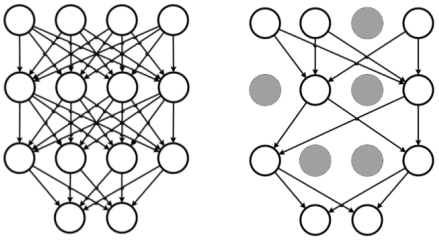

Figure 2.5. Illustration of the dropout method (Hinton et al. 2012). The left panel shows a fully connected feed-forward neural network, while the right panel shows the same network with some of the neurons disconnected by dropout.

2.3.3 Artificial clusters

Each cluster's age was sampled from the logarithmic range of $\log(t/\mathrm{yr}) = [6.6, 9.5]$ with a step of 0.05 dex; mass was sampled uniformly as a real number from the logarithmic range of $\log(M/\mathrm{M}_{\odot}) = [2.0, 4.0]$ to cover the majority of real young, low-mass M31 clusters studied by de Meulenaer et al. (2017). The cluster's star count surface density radial profile $\mu(r)$ was sampled from the EFF (Elson et al. 1987) model:

$$\mu(r) = \mu_0(1 + r^2/a^2)^{-\gamma/2}. \tag{2.3}$$

The parameters $a = [0.05, 6.4]$ arcsec and $\gamma = [2.05, 8.0]$ were sampled in a logarithmic space within the curves defined by constant r_h values (between 0.1 and 1.6 arcsec), as shown in Fig. 2.6. We define r_h as the radius of a circle on the sky enclosing half of a cluster's stars. These values at the assumed M31 distance (785 kpc) roughly correspond to real cluster sizes in M31 (Vansevičius et al. 2009; Narbutis et al. 2014). The setup was chosen to target low mass semi-resolved star clusters in order to demonstrate the CNN's capabilities in low signal-to-noise ratio conditions.

Samples of artificial clusters were generated with these parameters and combined with backgrounds of M31 stars. In order to accurately model photon noise the process is as follows. A random image of a background is selected and its photon count median value is determined. This median is then added to the image of an artificial cluster and then each pixel is sampled from a Poisson distribution, with its mean set to the pixel's value. The median is then subtracted back from this image and the real background image is added. The counts per second of the images were then transformed with the logarithmic

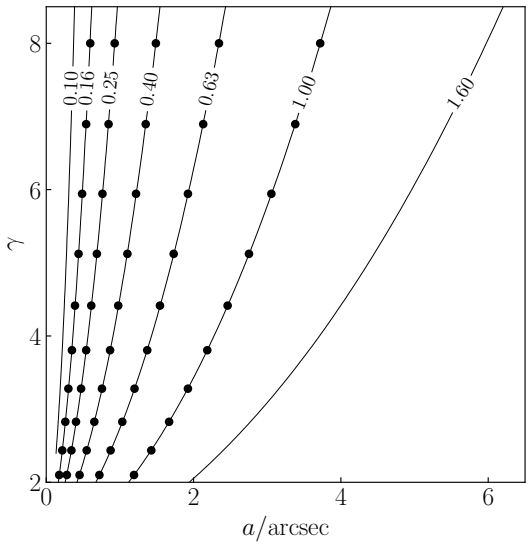

Figure 2.6. Used EFF parameter space and corresponding lines of constant r_h. Clusters for CNN training were sampled from lines, while dots correspond to the values of r_h used to construct a grid of the artificial test samples.

function $\log(x+1)$. The resulting images were 80×80 pixels in size, which correspond to 4×4 arcsec, or 15×15 pc at the distance of M31 (785 kpc).

Examples of the generated clusters with different ages, masses, and sizes, covering most of the parameter space, are shown in Fig. 2.7, the background is the same for all of the displayed clusters. For more examples drawn in all passbands available in PHAT data see Figs. A.1-A.4.

We generated 200,000 images of artificial clusters as a training sample for our CNN. A batch size of 64 images per training step was used. To ensure that the training procedure of our CNN would converge, we experimented with various learning rates, starting from $\eta = 0.1$, down to $\eta = 0.0001$. The learning rate of $\eta = 0.01$ gave the best performance on the validation set, so this was the value used for the final training of the network. A few different learning rate schedules were also tried out. Decreasing the learning rate twofold after every pass over the data gave the best results in our case. In order to control for overfitting we trained the network for 10 passes over the data while continuously monitoring its performance. The results on the validation set remained stable after about 3 passes, therefore, we chose to stop after 5 passes over the training data.

In addition, to test whether our training sample of 200,000 clusters is suf-

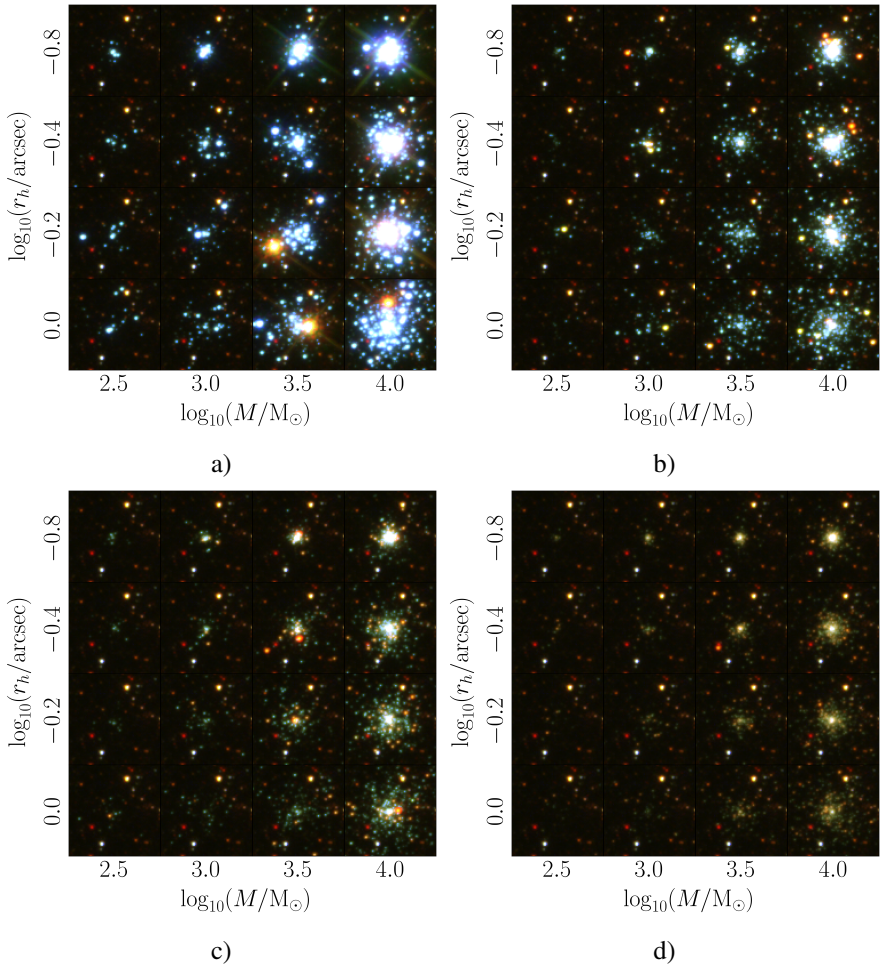

Figure 2.7. Examples of generated clusters on a real background image. The ages of all of the displayed clusters are: a) $\log(t/\mathrm{yr}) = 7.0$, b) 8.0, c) 8.5, d) 9.0. The masses and r_h values are varied as shown on the axes. The intensity scale of the images was normalized with the arcsinh function. The background is the same for all of the displayed clusters for clarity.

ficient for the network to generalize, we also trained our network with as little as 50,000 and as many as 400,000 cluster samples. During these experiments we didn't observe a significant difference between the inference results when trained with smaller or larger sample sizes.

2.4 Results and discussion

The data sets for testing the CNN were prepared by drawing from the same age and mass ranges as described in Section 2.3.3. However, for ease of analysis, the EFF model parameter γ and a pairs were chosen in such a way that the value of r_h would be equal to one of the following values: 0.16, 0.25, 0.4, 0.63, 1.0 arcsec (see Fig. 2.6). The procedure of image creation was also identical to that of the training data. The backgrounds for clusters were picked making sure that they would not overlap with the backgrounds used for the training set.

2.4.1 Parameter accuracy

To test the performance of the CNN, we built a bank of 10,000 artificial clusters by varying all 3 cluster parameters. The ages were sampled from the range of $\log(t/\mathrm{yr}) = [6.75, 9.25]$; mass was sampled from the range of $\log(M/\mathrm{M}_{\odot}) = [2.25, 3.75]$ and r_h, as shown in Fig. 2.6.

Differences between true and CNN-derived parameters vs. true parameter (age, mass and r_h) values of the testing set's artificial clusters are shown in Fig. 2.8. For the comparison in the figure the true ages are binned into 0.5 dex width bins, and the true masses into 0.3 dex width bins. The r_h value bins correspond to combinations of a and γ parameters denoted as dots in Fig. 2.6. The spread of the parameter differences within bins is displayed as box plots. The line in the middle of each box is the median difference between the true and derived values. Boxes extend from the 1st to the 3rd quartiles. Whiskers denote the range between the 2nd and the 98th percentiles. Anything above and below the whiskers is plotted as separate points.

Fig. 2.8a shows no significant age difference between the true and derived values. The typical standard deviation of the age difference is estimated to be $\sim$0.1 dex. The youngest clusters show a slight systematic bias towards older ages, as can be seen by the non-symmetric whiskers. This is also true for the oldest clusters in the test sample; only the bias is reversed. This could be explained as an age boundary effect due to the stellar isochrone age range limits

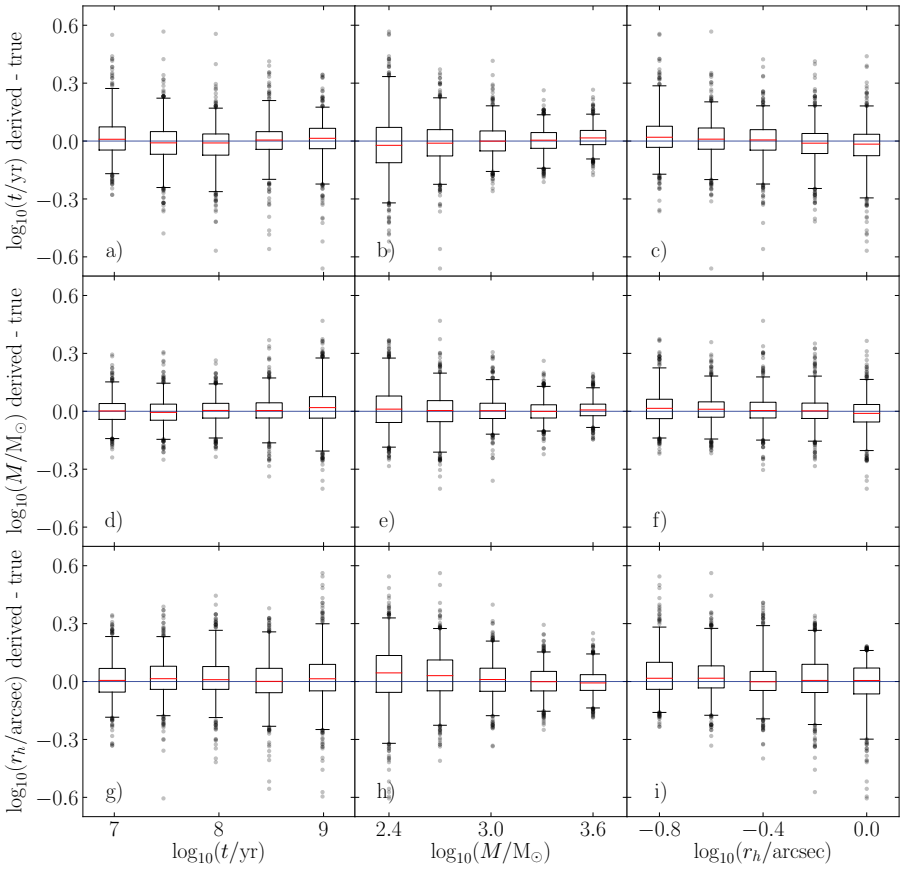

Figure 2.8. Differences between CNN derived and true parameters vs. true parameter (age, mass, and size) values of artificial clusters. The ages are binned into 0.5 dex width bins, and the masses into 0.3 dex width bins, while r_h bins correspond to EFF model parameters indicated by dots on the constant r_h values in Fig. 2.6. The widths of the boxes for age and mass correspond to half the widths of the bins. The spread of the parameter differences are displayed as box plots. The line in the middle of each box is the median error. Boxes extend from the 1st to the 3rd quartiles. The whiskers denote the range between the 2nd and 98th percentiles. Anything above and below the whiskers is plotted as separate points.

used for the CNN training. Fig. 2.8b shows the true and derived age value differences plotted against the true cluster masses. The median differences have no systematic shifts. However, the least massive clusters show a larger mass error spread, with standard deviation of the differences as high as $\sim$0.15 dex. The errors get systematically lower as we move towards the higher masses, stabilizing at a standard deviation of $\sim$0.05 dex for the most massive clusters. This could be explained by the low signal-to-noise ratio of $\log_{10}(M/\mathrm{M}_{\odot}) = 2.4$ clusters, as can be seen from Fig. 2.7. In Fig. 2.8c the error standard deviations are equal to $\sim$0.1 dex across the whole range of the true r_h values with a slight systematic bias for the least and most massive clusters.

Fig. 2.8d shows no systematic errors when deriving cluster mass as cluster age varies. However, the spread of errors does get larger towards older clusters, as they tend to be dimmer and of lower signal-to-noise. In Fig. 2.8e there are no systematic shifts; however errors are again larger for the lowest mass clusters. This could also be explained by the very low signal-to-noise ratio of $\log_{10}(M/\mathrm{M}_{\odot}) = 2.4$ clusters. In Fig. 2.8f the derived mass values do not seem to vary with changing r_h values and only a very slight bias can be observed for the smallest and biggest clusters.

Fig. 2.8g shows consistently derived r_h values with changing ages. Fig. 2.8h shows systematically larger derived r_h values for low mass clusters, as well as clearly larger errors. Fig. 2.8i shows a clear systematic trend deriving larger r_h values for small clusters, as well as lower values for large clusters. The shift in large clusters, with r_h as high as 1.0 arcsec, could be explained by the fact that a significant portion of a cluster's stars do not fit into the 4×4 arcsec images, the size of which is restricted by the chosen network architecture to minimize the background area for the majority of clusters.

2.4.2 Star cluster stochastic effects

In order to analyse the effects of star cluster stochasticity (initial mass function and star position sampling) on the CNN's inference results, we built a grid of clusters (200 per node) with fixed age, mass, and size parameters. The only random effects were background, star position, and mass sampling.

Fig. 2.9 shows three panels with different ages: $\log_{10}(t/\mathrm{yr}) = 7.0$, 8.0 and 9.0. Each black dot corresponds to a node of 200 clusters of fixed true mass and r_h, as denoted on the axes. Ellipses show the spread of the derived parameter values of these nodes. The arrows show the shift of the mean of the values and the boundary of the ellipses enclose 1σ of the 2D distribution ($\sim$39% of

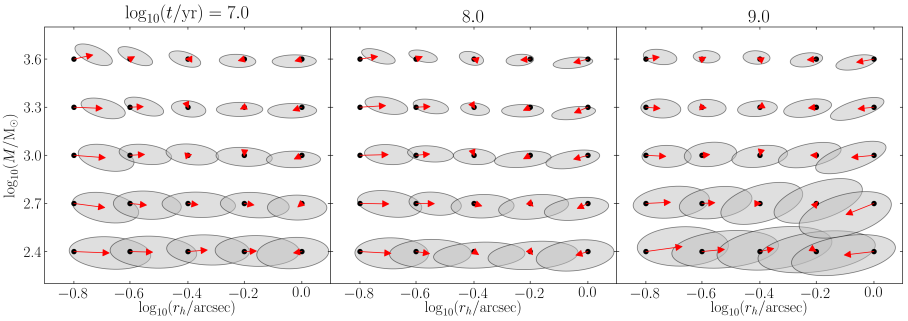

Figure 2.9. Tests results of CNN performance. Each black dot corresponds to the true parameters of 200 artificial clusters. The gray ellipses enclose one σ of the inferred values (accuracy), with the red arrows pointing to the means of the distributions (biases). The panels show mass vs. r_h for three different ages.

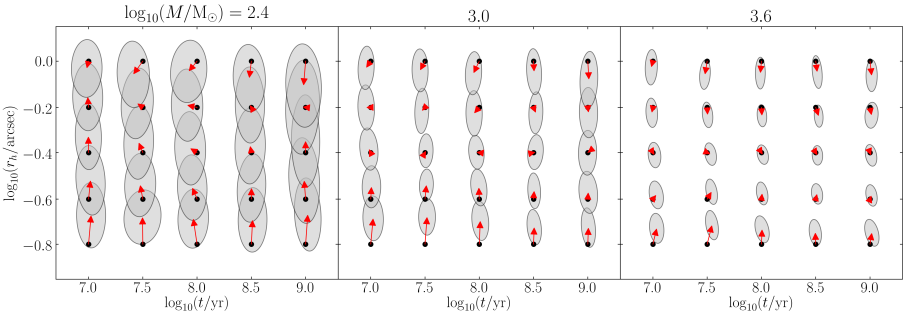

Figure 2.10. Same as Fig. 2.9, but panels show r_h vs. age for three different masses.

the clusters). The size of the ellipses at low masses are approximately three times larger than at high masses, especially so with the oldest ages. This is because of the low signal-to-noise ratio of the old low mass clusters, as can be seen in Fig. 2.7d. Low mass clusters tend to have slightly overestimated sizes. No significant correlated systematic shifts between the biases (red arrows) of derived mass and size can be seen.

Fig. 2.10 shows three panels with different masses: $\log_{10}(M/\mathrm{M}_\odot) = 2.4$, 3.0, 3.6. Black dots mark nodes of age and size as denoted on the axes. The spread of the derived age and size values is largest for the lower mass clusters, being at least three times larger for $\log_{10}(M/\mathrm{M}_\odot) = 2.4$ than for $\log_{10}(M/\mathrm{M}_\odot) = 3.6$. However, at a fixed mass, there's no correlated systematic shift between the derived cluster's mass and size biases. The slight

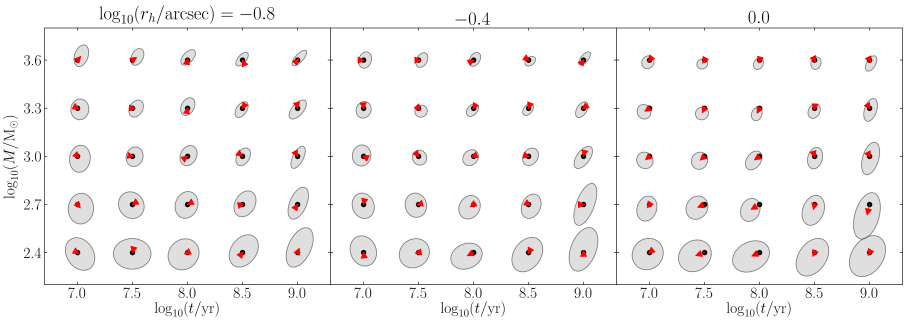

Figure 2.11. Same as Fig. 2.9, but panels show mass vs. age for three different r_h values.

overestimation of sizes can be seen again for the lowest mass clusters.

Fig. 2.11 shows three panels with different sizes: $\log_{10}(r_h/\mathrm{arcsec}) = -0.8$, -0.4, 0.0. Black dots mark nodes of age and mass as denoted on the axes. As seen in the previous figures, the spread of derived values tends to increase towards the lower mass clusters. However, there is no obvious influence of cluster size on the spread. A slight elongation of ellipses along the diagonal can be observed between the derived mass and age for the older grid clusters. This is because of the age-mass degeneracy, as the CNN has learned that older clusters have lower flux, and lower flux can be explained either by an older age or a lower mass. For $\log_{10}(t/\mathrm{yr}) \lesssim 8.0$ (as seen in Figs. 2.7a and 2.7b) stochastic effects are dominant; however for $\log_{10}(t/\mathrm{yr}) > 8.0$ (Figs. 2.7c and 2.7d) flux becomes the main factor in the mass-age correlation.

2.4.3 Overall CNN performance

The proposed CNN was verified on mock images of artificial clusters. It has demonstrated a high precision and no significant bias for semi-resolved clusters with ages between $\log_{10}(t/\mathrm{yr}) = 7.0$ and 9.0, and masses between $\log_{10}(M/\mathrm{M}_\odot) = 2.4$ and 3.6. Artificial cluster tests have demonstrated the effectiveness of CNNs in deriving star cluster parameters.

Even in the low mass regime it is possible to recover both age and mass as seen in Fig. 2.11. However, the plots show only one σ ($\sim$39%) of the 2D distributions, which means that for the lowest mass clusters five age groups can be identified. As mass increases, the number of identifiable groups increases. The same trend is also seen in Fig. 2.8b and holds regardless of cluster size (see Figs. 2.8c and 2.8f). Note that even though this is true on average, there

are still outliers with a parameter derivation error as large as 0.6 dex for the low mass clusters (Fig. 2.8b).

The low mass clusters shown in Fig. 2.7d are not visible to the naked eye. Nevertheless, the network still learns on the limited amount of signal and manages to produce valid parameter estimates. Even though the old age lowest mass clusters have been included, in real surveys they would be omitted because they would fall below the detection limit.

Meanwhile, for the cluster size, only three categories can be identified in the lowest mass regime as can be seen in Fig. 2.9. This is, however, unsurprising as the signal-to-noise ratio of these clusters is low (see Fig. 2.7). This is especially true for the oldest clusters. However, higher masses allow us to identify up to five size groups. Fig. 2.8h also confirms this finding.

Currently, evolutionary and structural parameters of semi-resolved clusters are estimated separately and in two completely different ways. Usually, for age and mass estimates a grid of models is constructed by varying the parameters of interest and comparing the resulting mock observations to real observations (Bridžius et al. 2008; de Meulenaer et al. 2013). Meanwhile, a common approach for structural parameters is Markov chain Monte Carlo sampling of the parameter space (Narbutis et al. 2015). Both of these methods essentially sift through a large number of possible observations to get a likely parameter estimate or the distribution over the parameters.

CNNs, while still requiring a sample of mock clusters for their training, are able to learn the properties of the system they're modeling within their weights and generalize better to new examples. This also allows the derivation of both evolutionary and structural parameters in a unified way, using all of the available information in image pixels.

The information registered in image pixels is affected by many additional factors, which were not included in the CNN training; the PSF shape is expected to be the most significant among them for cluster size. In this work, a fixed PSF, simulated for the center of the detector, was used. This was done for both the training cluster images and the test sets. In order to find the effects of variable PSFs on the network's performance, tests were done with the PSF taken from a corner of the detector; however the effects of this were negligible. In order to test the network's robustness to different views of the cluster, we also tried running the CNN on shifted and rotated cluster images. This, however, proved to have a negligible impact on the uncertainty of derived values.

Extinction and metallicity are also significant factors in age and mass

derivation (de Meulenaer et al. 2014, 2015a). Since for the first time we attempt to derive both evolutionary and structural parameters, we assumed zero extinction and fixed metallicity just to demonstrate CNN performance. At fixed extinction and metallicity the CNN demonstrates good accuracy for age and mass derivation.

In comparison to methods where any sort of fitting or sampling procedure needs to be performed, a CNN can obtain parameter inference results from an image very efficiently. All experiments in this work were performed using a GeForce GTX 1080 graphics card. The initial training procedure on the 200,000 clusters takes $\sim$6 hours; however this only needs to be run once. Inference runs much faster, taking 40 seconds per 10,000 images of all three passbands.

2.4.4 Tests on real clusters

To validate our method on real star clusters we took the sample of clusters used by de Meulenaer et al. (2017). From those clusters we selected only objects located on PHAT bricks 19-22. The CNN in our study was trained ignoring extinction entirely, therefore it is not possible to derive the parameters of significantly reddened clusters correctly. However, extinction estimates for the chosen clusters have been published by de Meulenaer et al. (2017). In order to run our experiments, we corrected the cluster images in each passband for the extinction effect by increasing their corresponding flux.

We then used our CNN to infer the cluster age and mass. Figure 2.12 shows a comparison between our results and those of de Meulenaer et al. (2017). A good agreement between the derived values can be seen. However, this result can only be used for a preliminary validation of the method. By increasing the flux of our real cluster sample to remove the effects of extinction, we change the flux of not only the cluster itself, but also of its background. On the other hand, the artificial clusters used in this work, while themselves drawn without extinction, are superimposed on real backgrounds which are affected by extinction. This makes any results on real clusters unreliable, introducing possible degeneracies. However, the preliminary results are promising and show a clear applicability to real clusters.

In order to deal with real clusters correctly our CNN needs to be trained on images with various extinction levels, as well as to be able to predict extinction in the same way it currently predicts age, mass, and size. However, for a reliable derivation of extinction more photometric filters are required.

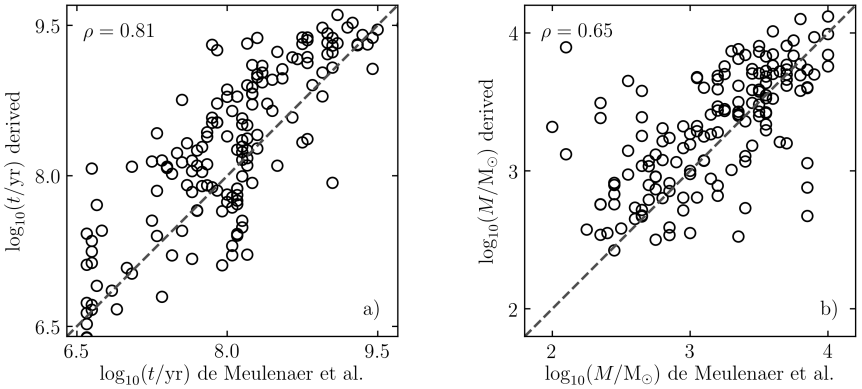

Figure 2.12. Comparison of ages (a) and masses (b) derived by de Meulenaer et al. (2017) and our CNN for 157 real PHAT clusters located in bricks 19-22. The correlation coefficient ρ between values is displayed in the top-left of each subfigure.

2.5 Conclusions

We have proposed a convolutional neural network based on the ResNet architecture for simultaneous derivation of both evolutionary and structural parameters of star clusters from imaging data. Artificial cluster images were combined with real M31 backgrounds observed with the HST and used for training the neural network.

The proposed CNN was verified on images of artificial clusters. It has demonstrated a high accuracy and no significant bias for semi-resolved clusters with ages between $\log_{10}(t/\mathrm{yr}) = 7.0$ and 9.0, masses between $\log_{10}(M/\mathrm{M}_\odot) =$ 2.4 and 3.6, and sizes between $\log_{10}(r_h/\mathrm{arcsec}) = -0.8$ and 0.0.

We have shown with artificial tests that CNNs can perform both structural and evolutionary star cluster parameter derivation directly from raw imaging data. This allows dealing with both unresolved and semi-resolved cases homogeneously, as well as utilizing multiple photometric passbands in an integrated manner.

Chapter 3

Extinction and cluster/background classification

In Chapter 2 we have implemented a CNN-based algorithm to simultaneously derive age, mass, and size of clusters in the low signal-to-noise regime. The algorithm was applied to M31 clusters, cataloged by The Panchromatic Hubble Andromeda Treasury (PHAT) survey. We have found that even when including information from all pixels and using accurate flux calibrations, interstellar extinction still plays a major role in influencing the results of parameter inference.

Numerous previous studies have explored physical parameter inference by taking into account the extinction problem, but were focused on the cases of resolved stellar or integrated cluster photometry. Among them are works by Bridžius et al. (2008), who used analytically integrated stellar luminosities, Fouesneau & Lançon (2010) and de Meulenaer et al. (2013, 2014), who used stochastically sampled stellar luminosities according to the stellar initial mass function (IMF), and SLUG, developed by Krumholz et al. (2015), which is one of the most mature codes in stochastic cluster population simulation and inference.

In this chapter we extend the CNN architecture proposed in Chapter 2 to allow the inference of a cluster's interstellar extinction directly from images. With an eye towards automated star cluster detection, we also explore indicators of cluster presence in images. The outputs of the network were modified to infer multiple cluster parameters jointly, which allows the degeneracies be-

tween them to be expressed in the outputs of the network, instead of relying on single-point estimates. This is especially useful when visualizing and dealing with age-extinction degeneracies.

We used the M83 galaxy HST survey (Blair et al. 2014), which covers the entire disk of this face-on galaxy in a number of passbands. This allows us to investigate the effects of extinction in a variety of dense and sparse environments. Previous studies of the M83 star cluster population were based on aperture photometry, such as Ryon et al. (2015) covering the whole galactic disk, Bastian et al. (2011) who studied a smaller part of the galaxy in detail, and Harris et al. (2001) covering its central region.

We trained the CNN on realistic mock observations and tested on mock clusters, as well as validated on the aforementioned real cluster catalogs.

We also experimented with the re-normalization of image fluxes for each passband separately when training the network, suggesting that precise photometric calibrations may not be necessary to derive star cluster parameters. This was done in the vein of Dieleman et al. (2015), where JPEG color images were used to classify galaxies, achieving reliable results. This brings the approach of analysis of astronomical images closer to the methods used on natural images, which rarely have accurate flux calibrations.

The chapter is organized as follows. Section 3.1 provides details about the M83 survey data, the mock cluster bank construction, the added new parameters, and training data preparation. Section 3.2 describes the proposed CNN and its training methodology. Section 3.3 presents the results of testing the method on mock as well as validating on real M83 clusters previously studied using integral photometry. Section 3.4 discusses the CNN parameter inference results in an astrophysical context.

3.1 Data

3.1.1 M83 mosaics

The M83 mosaic project data observed by the HST Wide Field Camera 3 (WFC3) (Blair et al. 2014; Dopita et al. 2010) was obtained from the Mikulski Archive for Space Telescopes[1]. We use stacked, defect-free mosaic images of 7 WFC3 fields, which are calibrated photometrically (pixel values are in counts per second) and astrometrically (with available world coordinate system

[1] https://doi.org/10.17909/T96888

information). The details of image processing are provided by Blair et al. (2014).

The mosaics cover the whole extent of the galaxy, from the dense center to its sparse outskirts where stellar background contamination is low. For the analysis we selected wide passband images that cover the whole galaxy without gaps: F336W, F438W, and F814W. All three mosaics are of the same size and in a tangential projection with a common scale (0.04 arcsec/pixel).

In Chapter 2 the M31 images were masked for saturated stars and extended objects in order to prevent unreliable CNN training. The distance to the M31 galaxy is 785 kpc (McConnachie et al. 2005), however the distance to M83 is 4.5 Mpc (Thim et al. 2003), therefore only a few saturated stars are visible. Because the area covered by extended objects in comparison to genuine stellar backgrounds is negligible, we decided to skip the masking step altogether and use all of the available mosaic area when selecting backgrounds for artificial clusters.

3.1.2 Mock cluster generation

Mock clusters were generated with different ages, masses, sizes, and affected by various levels of extinction. A fixed metallicity of $Z = 0.03$ (Hernandez et al. 2019) was assumed.

The procedure and the used parameter ranges are largely the same as in Section 2.3.3. The age of each cluster was sampled from the logarithmic range of $\log(t/\mathrm{yr}) = [6.6, 10.1]$ (with a step of 0.05 dex); mass was sampled uniformly from the logarithmic range of $\log(M/\mathrm{M}_\odot) = [3.5, 5.5]$ to cover the majority of M83 clusters studied by Bastian et al. (2011). Extinction was drawn from the range of $A_V = [0.0, 3.0]$ mag (with a step of 0.1 mag), by obtaining the extincted isochrones computed for Cardelli et al. (1989) extinction curve, with $R_V = 3.1$. We define r_h as the radius of a circle on the sky enclosing half of the stars of a cluster. The spatial distributions of stars were sampled from the Elson-Fall-Freeman (EFF) (Elson et al. 1987) profile:

$$\mu(r) = \mu_0(1 + r^2/a^2)^{-\gamma/2}. \tag{3.1}$$

The parameters a and γ were drawn uniformly from logarithmic ranges of $[0.04, 1.2]$ and $[2.05, 8.0]$ respectively, such that r_h is within the limits of $[0.04, 0.4]$ arcsec. These values at the assumed distance of M83 (Thim et al. 2003, 4.5 Mpc) roughly correspond to real cluster sizes in M83 (Bastian et al. 2011).

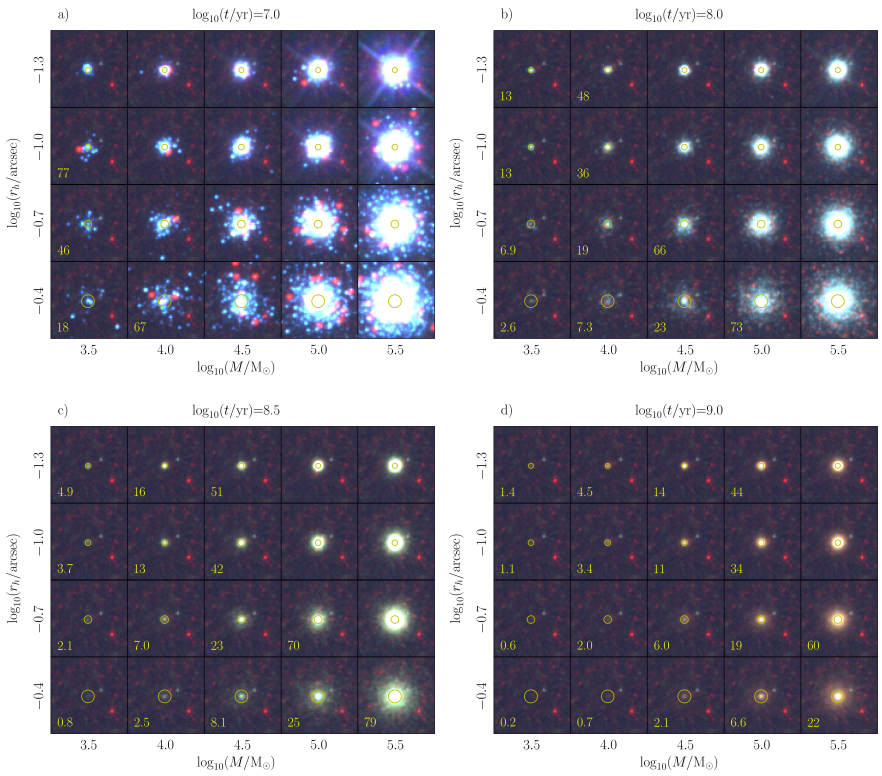

Figure 3.1. Examples of generated mock clusters placed on a real background image, which is the same for all panels. The ages of the displayed clusters are shown on top of each panel. The mass and size, r_h, values are varied as shown on the axes; extinction $A_V = 0$ mag. The intensity scale of the images was normalized with the arcsinh function within identical pixel value limits for each image. The yellow circles represent $r_{h,\mathrm{obs}}$ values (r_h convolved with the point spread function). The *visibility* (signal-to-noise proxy) value is displayed on the bottom-left of each image for fainter objects with $visibility < 100$. Image sizes are 64×64 pixels (2.6×2.6 arcsec) or $\sim 60 \times 60$ pc at the distance of M83. In the images the color red corresponds to passband F814W, green to F438W, and blue to F336W.

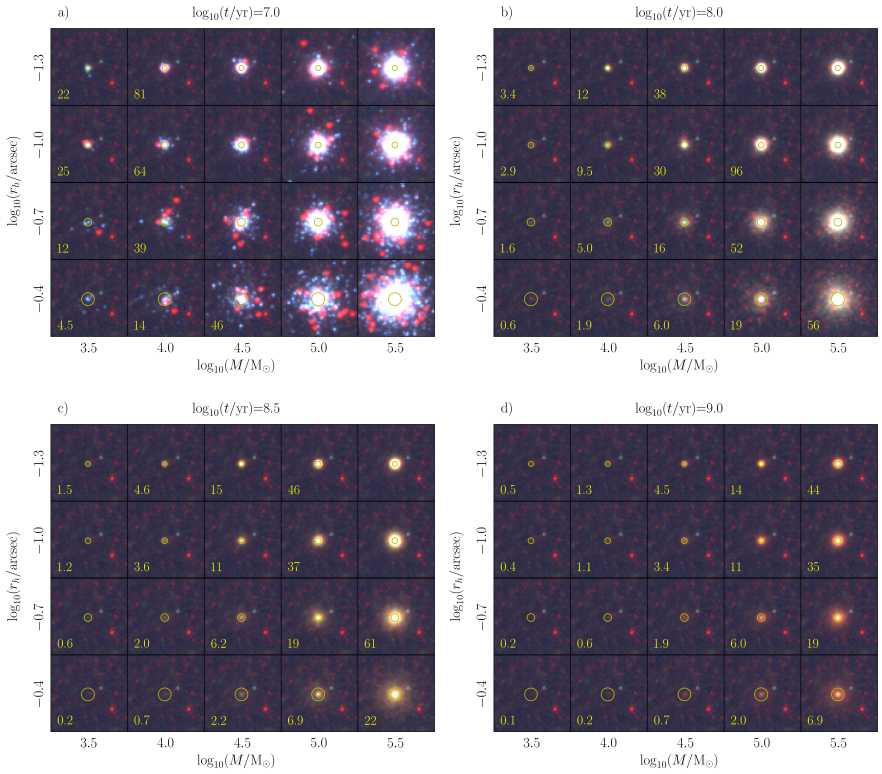

Figure 3.2. Same as Fig. 3.1, but with $A_V = 1$ mag.

The stars of the clusters were generated as follows. Given the initial mass, M, of a cluster, star masses were sampled according to the Kroupa (2001) IMF from Padova PARSEC isochrones[2] (Bressan et al. 2012, release 1.2S), obtaining the absolute star magnitudes for passbands F336W, F438W, and F814W. Then, the absolute magnitudes were transformed to apparent magnitudes at the distance of M83 (Thim et al. 2003, 4.5 Mpc) and converted to the WFC3 camera counts per second for the three passbands using calibrations provided by Dressel (2012). Finally, the 2D star spatial positions were generated by sampling star distances from the cluster's center according to the EFF profile (with given a and γ values) and then distributing them symmetrically around the center.

The GalSim package (Rowe et al. 2015) was used to draw the individual stars of the clusters using TinyTim-generated[3] point spread functions (PSFs) (Krist et al. 2011) for each of the three passbands. Every star in the cluster was drawn separately for each passband using the appropriate PSF scaled by the star's counts per second. For a single cluster this produces three images, which can then be visualized as either RGB pictures or given to a CNN as 3D (width $\times$ height $\times$ passband) arrays. Artificial clusters were then placed on backgrounds cut from the M83 mosaics. See Figs. 3.1 and 3.2 for examples of the generated mock clusters.

To explore the photometric properties of the cluster bank, we show integrated color-color and color-magnitude diagrams in Fig. 3.3. The magnitudes depicted were obtained solely from integrating the total flux of mock clusters and therefore are an idealized case which does not take into account the variations of background and spatial star positions. The only source of stochastic effects in such a case is IMF sampling. Panels are dedicated to illustrate the influence of age, extinction, and mass present in the bank. The effects of these parameters are in different directions in the color-color and color-magnitude space. The oldest clusters are red (panel a) and low-luminosity (panel e) objects. Clusters with high extinction are reddened (panel b), and the lowest mass clusters are faintest (panel g).

The last column (Fig. 3.3, panels d and h) shows distributions of star clusters filtered by mass (as specified on the color bar on top) and by extinction $A_V < 0.2$ mag. The simple stellar population (SSP) tracks centered on the specified masses are shown as black curves. In both color-color and color-

[2]http://stev.oapd.inaf.it/cgi-bin/cmd

[3]http://tinytim.stsci.edu/cgi-bin/tinytimweb.cgi

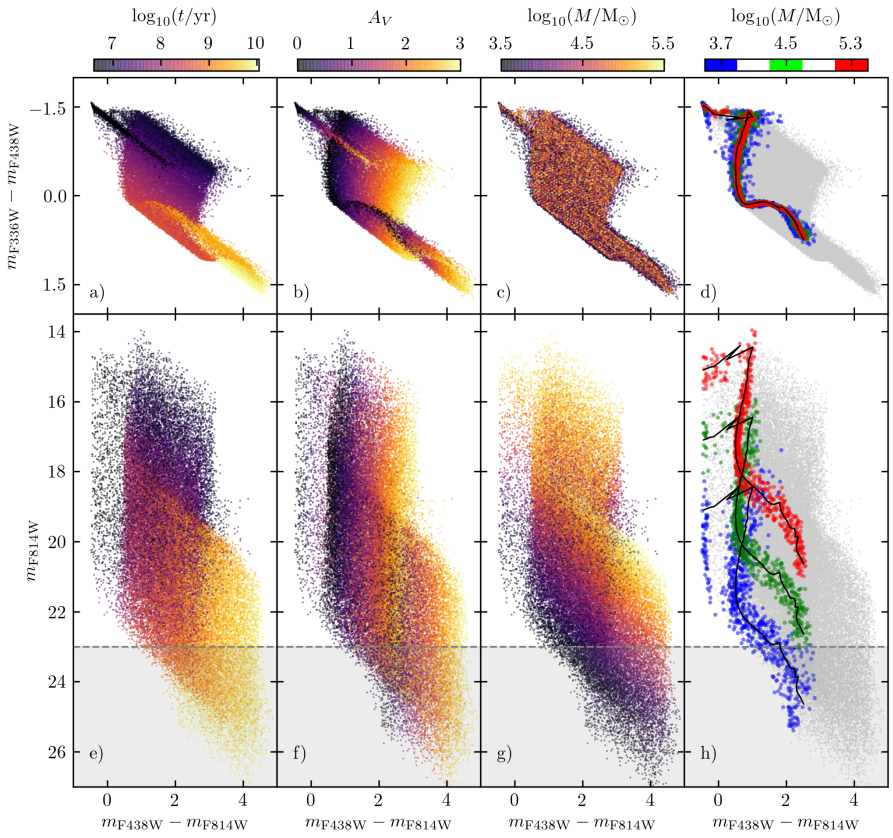

Figure 3.3. Integrated color-color and color-magnitude diagrams of 50,000 generated mock clusters. The color coding represents different ages (a, e), extinctions (b, f), and masses (c, g). Their values are as noted on the color bars on top. The last column (d, h) shows distributions of $A_V < 0.2$ mag star clusters filtered by three mass ranges as specified on the color bar on top. The simple stellar population tracks centered on the specified masses are shown as black curves. The shaded area below the dashed line represents the F814W magnitude limit used to filter out faint clusters.

magnitude space, it can be seen that lower mass clusters are more widely distributed due to the stochastic IMF sampling. The effects of mass on cluster magnitude can be seen again in Fig. 3.3h as vertical shifts of the SSP tracks.

This means that a point in color-color and color-magnitude space can't uniquely map to a point in cluster parameter space. This is worsened by stochastic IMF sampling effects and results in degeneracies with which any parameter inference method has to deal with. In cases like this any additional sources of information, such as individual image pixel values, are welcome.

Faint objects with $m_{\rm F336W} > 24$ mag, $m_{\rm F438W} > 23.5$ mag, and $m_{\rm F814W} > 23$ mag were excluded from the cluster bank due to their low signal, to mimic age/mass/extinction selection effects existing in magnitude-limited real cluster samples. For the F814W band this is illustrated by the shaded gray area in Fig. 3.3. See the lower-left corner of panel d in Figs. 3.1 and 3.2 for examples of such barely visible clusters.

3.1.3 Mock cluster properties

Samples of artificial clusters were generated with the described parameters and placed on real backgrounds of M83. In order to realistically model photon noise the following steps were applied. A cutout image of an M83 background from a random position in the mosaics is selected and its median value is determined. This median is then added to the image of an artificial cluster, multiplied by the exposure time to get photon counts, and then each pixel is sampled from a Poisson distribution, with its mean set to the value of the pixel. The median is then subtracted back from this image, the real background image is added and photon counts are transformed back to counts per second.

We also define a cluster *visibility* parameter, constructing it to approximate signal-to-noise in such a way that higher values would be assigned to clusters that stand out relative to their stochastic stellar backgrounds. It is defined as follows:

$$visibility = \frac{f_c}{n \cdot \sigma_b}, \tag{3.2}$$

where f_c is the integral flux of the cluster within its $r_{h,\rm obs}$, while σ_b is the standard deviation of the background's pixel values in a 25 pix (1 arcsec) radius aperture, and n is the number of pixels within $r_{h,\rm obs}$. Here $r_{h,\rm obs}$ is the cluster's r_h value increased to account for PSF size, which has the largest effect on the most compact clusters. A mock cluster with $visibility = 1$ has mean

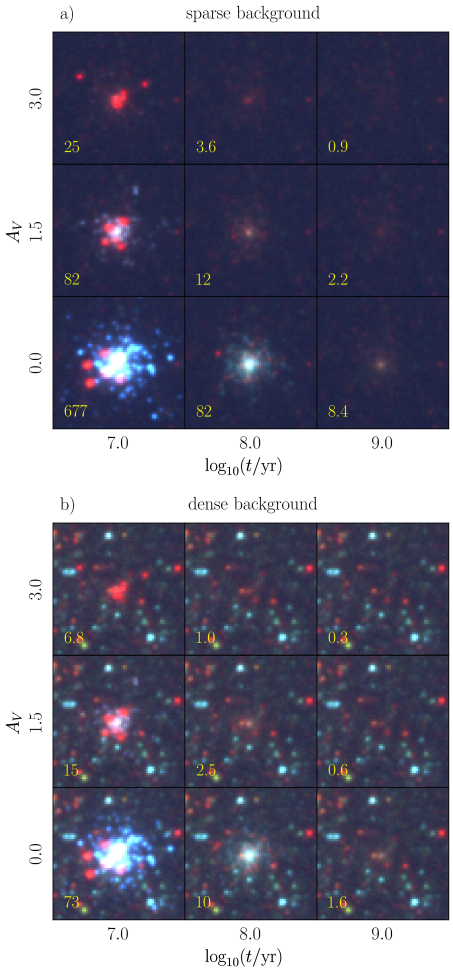

Figure 3.4. Examples of generated mock clusters with varying ages and extinctions on real background images. The masses of the clusters are $\log_{10}(M/\mathrm{M}_\odot) = 4.5$, their sizes are $\log_{10}(r_h/\mathrm{arcsec}) = -0.7$. The images are normalized as in Fig. 3.1. Top panel shows clusters superimposed on a sparse background, while the bottom panel shows the same clusters superimposed on a denser background. The *visibility* value is displayed on the bottom-left of each image. The sizes of the images are 64×64 pix.

flux per pixel approximately equal to the value of the standard deviation of a background it is placed on.

See Figs. 3.1 and 3.2 for a variety of *visibility* values of clusters, displayed as yellow text in the corner of each image, with A_V up to 1 mag. See Fig. 3.4 for samples of clusters with the full range of extinction (A_V up to 3 mag) used in this work to illustrate the effect of background crowding on *visibility*. It can be seen that the values of *visibility* correlate well with the ability to resolve clusters by eye – the best tool for cluster detection up to date.

Note that for real clusters it is not possible to infer properties of background covered by cluster's light, however by placing mock objects into backgrounds, we can compute *visibility* parameter beforehand and train the network to infer it from the data of real observations.

3.1.4 Training data preparation

To minimize the influence of photometric image calibration accuracy, the counts per second of each passband of a cluster's image were individually normalized to the mean of 0 and standard deviation of 1. They were then rescaled with the arcsinh function. The resulting images were 64×64 pixels in size, which correspond to 2.6×2.6 arcsec, or 60×60 pc at the distance of M83 (Thim et al. 2003, 4.5 Mpc). Examples of the generated clusters with different ages, masses, and sizes, and without extinction, covering most of the parameter space, are shown in Fig. 3.1. A series of different examples (star position and mass sampling), but with extinction $A_V = 1$ mag, are shown in Fig. 3.2. We generated 50,000 such images of mock clusters as a training sample for the CNN. The backgrounds have also been precomputed for efficiency resulting in 80,000 cutouts that were combined with the cluster images.

3.2 Convolutional Neural Network

3.2.1 Architecture

Following the work in Chapter 2, the ResNet-50 (He et al. 2016) architecture was used as a basis for our CNN. In addition, a series of modifications were made to it in order to accommodate the different survey images, the higher number of predicted parameters, as well as the degeneracies between them. See Figs. 3.5 and 3.6 for details on the structure of the modified CNN.

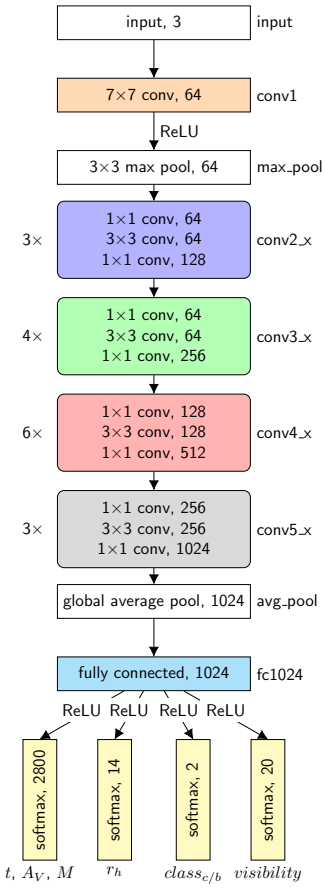

Figure 3.5. A block diagram of the CNN. The three-channel input image of a cluster passes through the network top to bottom, with the output result being age, extinction, mass, size, the cluster/background class $class_{c/b}$, and $visibility$. All blocks with sharp corners depict single layers, while blocks with rounded corners are groupings of layers, with the number on the left indicating how many times the group is repeated sequentially and the name on the right corresponding to the layer names in Fig. 3.6. The blocks with non-white backgrounds are parts of the network with optimizable parameters. The last number in each row is the number of output channels from that layer. "ReLU" indicates the locations in the network where rectified linear activations are applied between blocks.

Layer	Output		Operation
	Size	Channels	
input	64×64	3	
conv1	64×64	64	7×7
max_pool	32×32	64	3×3 max pool /2
conv2_x	32×32	128	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 128 \end{bmatrix}\times3$
conv3_x	16×16	256	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times4$
conv4_x	8×8	512	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times6$
conv5_x	4×4	1024	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times3$
avg_pool	1×1	1024	global average pool
fc1024	1×1	1024	fully connected
t, A_V, M r_h $class_{c/b}$ $visibility$	1×1	$20 \cdot 10 \cdot 14$ 14 2 20	softmax

Figure 3.6. The designed 50-layer CNN, based on the ResNet architecture. The layers of the network are listed top to bottom, starting from the images of clusters and with the final layer producing cluster's parameters. The convolutional layers are actually groups of blocks, with the "_x" in the name acting as a placeholder for the block number. The size of the outputs of each layer, both in spatial dimensions and in channel count, are listed on the second and third columns. The last column lists the operations that each layer performs. The layers or blocks with a stride of 2 are: max_pool, conv3_1, conv4_1, and conv5_1; as can be seen when input and output sizes differ 2 times. The last layer has 4 groups of softmax layers branching out in parallel, with the first predicting age, extinction, and mass, the second – cluster size, the third – the cluster/background class, and the fourth – $visibility$.

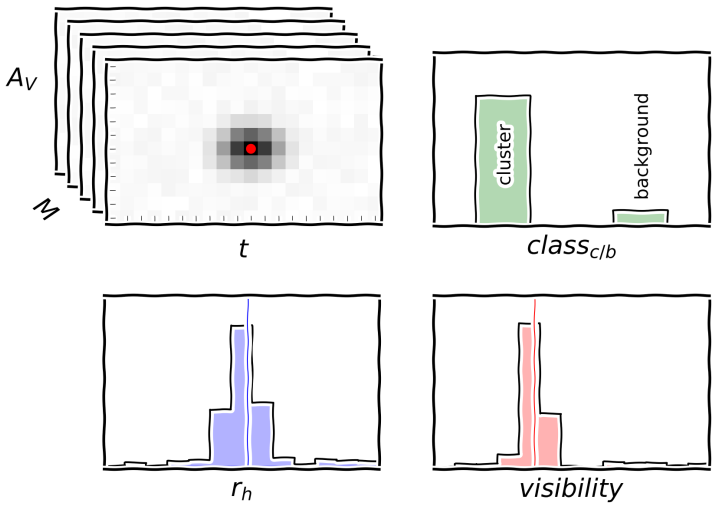

Figure 3.7. An illustration of the activations in the output layers of the CNN. While training the CNN, target activation values are provided as a 3D Gaussian distribution for age, extinction, and mass (centered on true values as denoted by the red dot) and as 1D Gaussian distributions for r_h (the blue line) and *visibility* (the red line). The cluster/background class is represented as a value of 0 or 1. During inference the network produces similar outputs, examples of which are depicted in Figs. 3.9 and 3.21.

In Chapter 2 we used a method by Dieleman et al. (2015) to rotate the input image multiple times and pass it through the same convolutional layers; to simplify the network we omitted this step. The input image size was decreased to 64×64 pixels to account for the smaller angular size of the clusters due to the more distant galaxy. Three input channels were used corresponding to the F336W, F438W, and F814W passbands.

In Chapter 2, the cluster's parameters were predicted via linear output layers by treating it as a regression problem. This meant that each parameter was predicted independently. However, due to age/extinction degeneracies and age/extinction/mass selection effects (shown in Fig. 3.3) this approach is no longer viable.

Therefore, we predict all of the parameters on a grid, with the positions on it corresponding to the parameter values. This essentially transforms the regression problem into classification, allowing the network to predict each parameter in multiple locations of the parameter space, properly representing some degenerate cases such as low-extinction and old-age being just as likely as high-extinction and young-age.

The network's output are 4 groups of layers branching out in parallel. The first group predicts age, extinction, and mass, the second – cluster size, the third – cluster/background class ($class_{c/b}$), and the fourth – cluster *visibility* (see bottom of Fig. 3.5).

Fig. 3.7 depicts the four output layer activations. We grouped age, extinction, and mass into a single output layer to allow the degeneracies between these parameters to be expressed in the network architecture itself. This was done by predicting them as activations on a 3D grid, with 20 bins for age, 10 for extinction, and 14 for mass. When flattened, this results in a softmax layer with $K = 2800$ neurons. For $class_{c/b}$ $K = 2$ neurons were used to encode the likelihood of a cluster's presence in the image. For the remaining parameters single-dimensional grids were used, resulting in $K = 14$ neurons for size and $K = 20$ for $visibility$.

Each of the four groups of output parameters were represented as softmax activations:

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}, \tag{3.3}$$

where $\vec{z}$ is the activations of a whole layer, and i specifies the index of a neuron (position on the parameter grid). The network was implemented with Keras[4]

[4]https://keras.io/

and TensorFlow[5] packages.

3.2.2 Training and inference

When training the network, we wish to infer both $class_{c/b}$, which indicates the presence of a cluster, and the cluster's astrophysical parameters at the same time. To that end learning the $class_{c/b}$ parameter was modeled as a simple binary classification task. The network is trained on batches of 512 images, half of which are images of backgrounds, and the other half are images of backgrounds combined with clusters as described in Section 3.1.2. For the images with only background in them we set $class_{c/b} = 0$, while for the samples with clusters we set $class_{c/b} = 1$.

In addition, for background images we zero out the training loss gradients for all cluster parameters. In effect this causes gradient updates to only be derived from the $class_{c/b}$ and *visibility* parameters, both of which are set to 0, indicating that the background contains no cluster. Training proceeds by sampling from M83 backgrounds ($\sim$25,000 images) and the cluster bank (50,000 mock clusters) separately, combining the cluster and background images on the fly, effectively giving us over 10^9 unique training samples.

The usual way to encode real-valued parameters as bins is called one-hot encoding. The parameter space is divided into bins and the bin at the position of the parameter's value is set to 1. This array is then passed as a target vector, $\vec{y}$, for the network. One-hot encoding is ideal for categorical classification, where only one of the target bins is true at a time. However, for binned real-valued parameters this has the unfortunate side-effect of penalizing bins far away from the target just as much as bins nearby to it. The way we solve this is by inserting a Gaussian distribution centered on the true value of the parameter (see Fig. 3.7). For the case of r_h and *visibility* this is a simple 1D Gaussian, with a standard deviation equal to 0.5 the width of a bin. For age, mass, and A_V a 3D Gaussian was used, with a standard deviation equal to 0.25 the width of a bin.

To obtain parameter estimates from this network we need a way to transform the network's output activations back into single-point estimates that can then be analyzed. The 1D and 3D histograms, depicted in Fig. 3.7, need to be "unfolded". This was done by finding the bin with the highest value in the histogram, which represents the most likely set of parameters inferred by

[5] https://www.tensorflow.org/

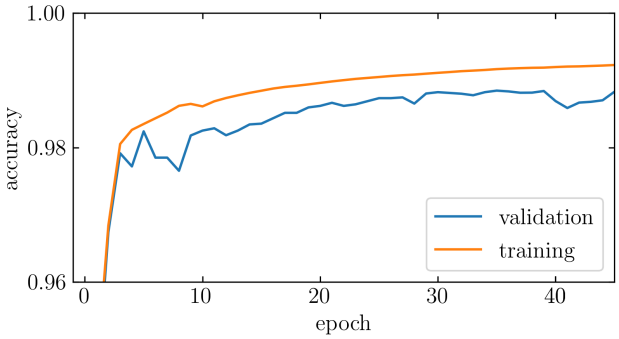

Figure 3.8. Binary classification accuracy metric for $class_{c/b}$ on the training (orange) and validation (blue) sets during the course of CNN training.

the network, and calculating a weighted average within a 3 bin radius. In effect this gives us an interpolated value in-between the bins making the output real-valued instead of discrete. Examples of inference results on mock and real clusters with both the raw activation outputs and the derived single-point estimates are shown in Figs. 3.9 and 3.21.

For computing the training gradients for the network the categorical cross-entropy loss function was used:

$$\mathscr{L} = -\sum_{i=1}^{K} y_i \log(\sigma(\vec{z})_i), \tag{3.4}$$

where $\sigma(\vec{z})_i$ is the neuron's activation as described in Eq. 3.3, and y_i is the target output for the given training cluster image.

The Adam optimizer (Kingma & Ba 2014) was used to calculate the gradients at each step of training. We experimented with various learning rates, starting from 0.1 down to 0.0001, with the learning rate decaying down to 0.0001 at the final iteration of training for all experiments. The learning rate of 0.01 gave the best performance on the mock validation set, so this was the value used for the final training of the network. The best CNN model was selected by picking the training iteration during which the CNN's loss was the lowest on the validation set. The training accuracy track of the $class_{c/b}$ parameter for the resulting model is shown in Fig. 3.8. This is the only parameter for which accuracy can be meaningfully calculated, because the other parameters are encoded as Gaussian distributions.

3.2.3 Output activations and stochastic effects

Three types of stochastic effects play a major role in the variation of CNN-inferred cluster parameters: 1) stellar mass sampling, 2) star spatial position randomization, and 3) background field.

We combine stellar mass and position sampling into one stochastic factor as a property of the cluster itself, while leaving the background choice as a property of its environment. We study both effects separately by a) generating different clusters with fixed parameters and placing them on the same background, and b) placing the same cluster on different backgrounds.

Fig. 3.9 displays the influence of stochastic effects on the inference results of mock cluster. Left column shows clusters with $\log_{10}(M/\mathrm{M}_{\odot}) = 5.0$, extinctions A_V = 0.5, 1.5, 2.5 mag, and $\log(t/\mathrm{yr})$ = 7.5, 8.5, 9.5. Right column shows clusters with $A_V = 0.5$ mag, $\log_{10}(M/\mathrm{M}_{\odot})$ = 4.0, 4.5, 5.0, and the same ages. Cluster sizes are fixed at $\log_{10}(r_h/\mathrm{arcsec}) = -0.6$ for all cases. The top row shows the results of inference when stellar IMF sampling and spatial positions are varied while holding the cluster parameters constant. The middle row shows the results of inference when background images are varied while using the same cluster image. The bottom row shows visualizations of the analyzed clusters in the same format as Fig. 3.4. The *visibility* parameter value is displayed on the bottom-left of each image. Note that the CNN predicts ages, masses, and extinctions as one 3D cube, while the outputs shown here are marginalized either over mass (left column) or extinction (right column).

In Fig. 3.9 it can be seen that the inference results for clusters with high visibility are all tightly packed for both types of stochastic effects (top and middle rows). This applies for both the spread of the CNN activation maps (grayscale) as well as the single-point estimates on different cluster images (magenta dots).

However, as clusters get fainter, and especially when they disappear into the background, the spread of activation maps (grayscale) as well as single-point estimates (magenta dots) gets wider. Background variability has a significantly larger influence on the parameter spread than stellar sampling effects.

It is worth noting that for old clusters CNN output activations are elongated, attempting to represent age/extinction degeneracies. For a small number of cases bimodal solutions are obtained. However, for cases where clusters are completely invisible both activations and single-point estimates end up tightly packed. This highlights the importance of the *visibility* parameter.

We note that less than 1% of the clusters show the bimodal distribution

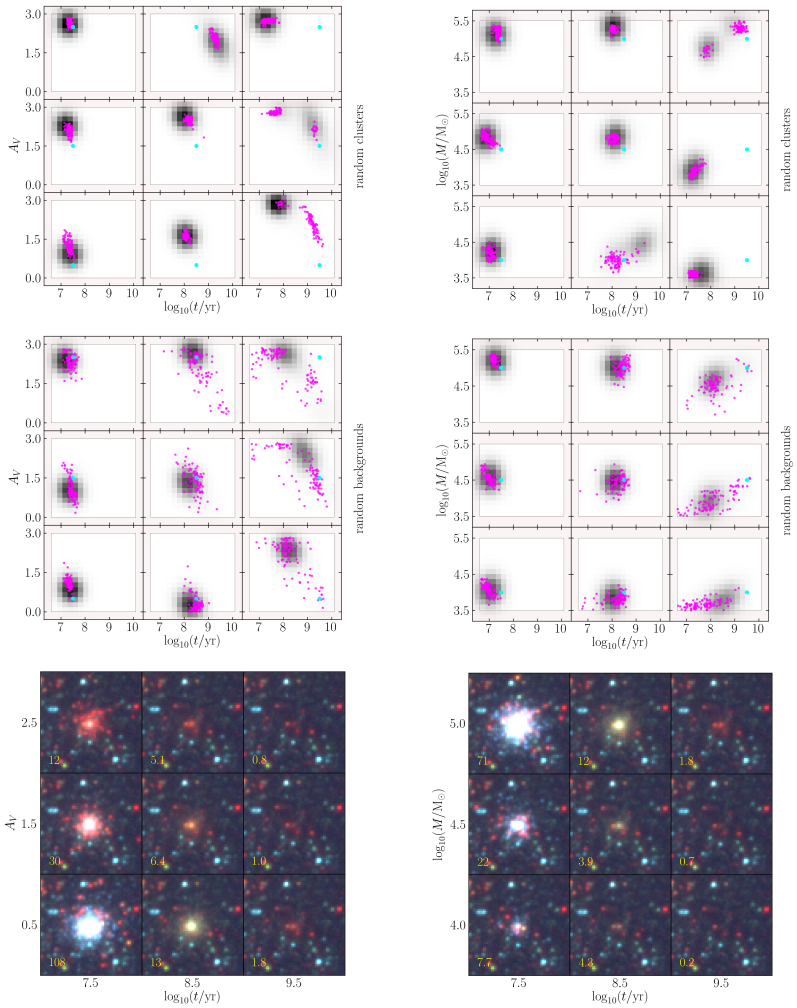

Figure 3.9. Influence of stochastic effects on mock cluster inference results. Left column shows clusters with $\log_{10}(M/\mathrm{M}_\odot) = 5.0$ and varied extinction and age. Right column shows clusters with $A_V = 0.5$ mag and varied mass and age. Cluster sizes are fixed at $\log_{10}(r_h/\mathrm{arcsec}) = -0.6$ for all cases. The top row shows the results of inference when stellar IMF sampling and spatial positions are varied while holding the cluster parameters constant. The middle row shows the results of inference when background images are varied while using the same cluster image. The cyan circles correspond to the true values of parameters. The grayscale colormaps are raw CNN outputs (activations over the parameter space) for one specific case, while magenta circles show 100 single-point estimates obtained for different random cases. The bottom row shows visualizations of the analyzed clusters in the same format as Fig. 3.4.

of activations. About 20% of the samples show an extended unimodal distribution, while the rest of the results are symmetric and unimodal. Therefore, selecting the highest activation and obtaining single-point estimates from it is a viable approach, as that captures most of the information present in the CNN outputs.

Fig. 3.9 also illustrates the possibility to quantify the uncertainties of single inference results either from the extent of activation maps or by sampling random backgrounds, adding them to a cluster's image, and re-running inference. However, the former produces tightly packed activation maps for some high-uncertainty samples, making them unreliable for low-visibility scenarios. The latter can also introduce additional effects depending on the used background sampling method, as well as the tendency to overestimate the uncertainties on real clusters, where the background effects would get doubled.

In subsequent sections the single-point estimates will be analyzed with respect to inferred parameter accuracy and the age/extinction degeneracy.

3.3 Results

3.3.1 Tests on mock clusters

To test the performance of the CNN, we built a separate bank of 5,000 artificial clusters. Their parameters were drawn from the same distributions as described in Section 3.1.2. The backgrounds for these mock clusters were also sampled from the used M83 mosaic, making sure that they are not the same as the backgrounds used for training. The inferred parameter values were obtained as described in Section 3.2.2.

Differences between CNN-derived single-point estimates of age, mass, extinction, and size vs. true parameters are shown in Fig. 3.10. The spread of errors is visualized as a hexagonal density map with the count bins scaled logarithmically in order to highlight the spread of outliers. Dashed lines represent the error bounds containing 95% of the inference results for each parameter. Note that because of magnitude cuts introduced in the mock cluster bank, discussed in Section 3.1.2, the parameter distributions aren't uniform. For example, there are relatively fewer low-mass old-age clusters. In all of the panels, the clusters that are classified as much younger than the true given values are shown as red points, while the clusters classified as much older are highlighted as blue.

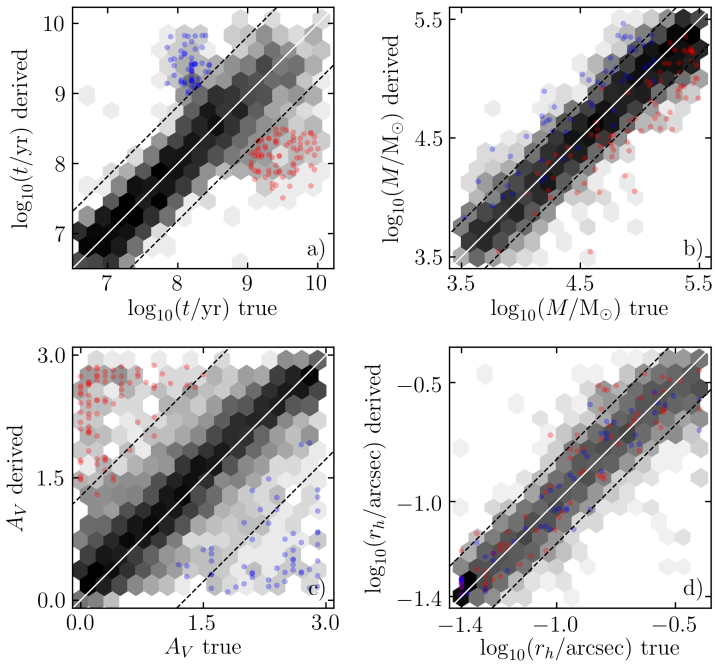

Figure 3.10. The true and derived parameter values of test mock clusters visualized as a hexagonal density map. The bins are scaled logarithmically. Panels show comparisons for a) age, b) mass, c) A_V, and d) r_h. Dashed lines highlight the area containing 95% of the clusters. Red dots represent the clusters that were misclassified as younger than the real age values, while blue dots represent clusters misclassified as older.

Fig. 3.10a shows no significant difference between the true and derived age values for $\log_{10}(t/\mathrm{yr}) < 8$ and the distribution for all ages is symmetrical along the diagonal. The 95% of all inference results deviate <0.9 mag from the true values, as shown by the dashed lines. Starting at $\log_{10}(t/\mathrm{yr}) = 8$ and above a large scatter in both directions – towards older and younger ages can be seen.

Fig. 3.10c shows the true and derived A_V values. The 95% of all inference results deviate <1.4 mag from the true values, as shown by the dashed lines. The highlighted blue and red clusters are classified as having significantly higher and lower extinction respectively. This can be explained by the age-extinction degeneracy, as older clusters with low extinction are hard to distinguish from younger clusters with high extinction, and vice-versa, when using only three photometric passbands.

In Fig. 3.3a-b the age-extinction degeneracy can be seen in the lower S-shaped part ($m_{F336W} - m_{F438W} > 0$) of the color-color distribution of clusters. Clusters older than $\log_{10}(t/\mathrm{yr}) = 8$ with high extinction can be located in the same color-color area as clusters with low extinction. These effects have also been observed when using analytically integrated stellar luminosities (Bridžius et al. 2008) and remain when stochastic effects of IMF sampling are included (de Meulenaer et al. 2014).

Fig. 3.10b shows the true and derived mass values. The 95% of all inference results deviate <0.4 dex from the true values, as shown by the dashed lines. Overall no systematic effects can be seen, however the clusters highlighted in red have slightly underestimated masses and vice-versa for the clusters highlighted in blue.

Fig. 3.10d shows the true and derived r_h values. No systematic effects can be seen. The 95% of all inference results deviate <0.2 dex from the true values, as shown by the dashed lines. However, for the smallest clusters the error spread is as low as ∼0.1 dex, while for the largest clusters the error spread goes up to ∼0.2 dex. This can be explained by the clusters with higher r_h having lower signal-to-noise, as their stars are spread out over a larger area in space.

Although in Fig. 3.10b due to the age-extinction degeneracy we observe underestimated and overestimated cluster masses, size errors shown in panel d show no such bias. This can be explained by mass being a function of a cluster's magnitude as can be seen in Fig. 3.3g, which makes the network mispredict its value if age and extinction are also mispredicted. However, size

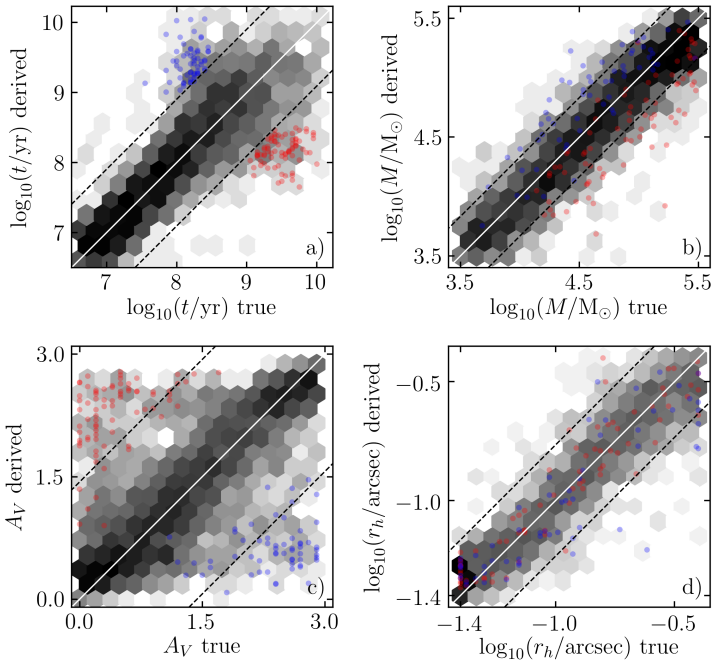

Figure 3.11. Same as Fig. 3.10, but using a CNN which was trained on mock clusters with simulated uncertainty of photometric calibrations.

has no impact on cluster magnitude or color.

As we use images normalized in a passband-independent manner, the influence of calibration accuracy to our method was also explored. Fig. 3.11 shows results obtained on the same dataset as Fig. 3.10, only with the CNN trained on images with background fluxes that were varied from image to image. The flux scaling factor was sampled independently for each passband as a Gaussian with a mean of 1 and a standard deviation of 0.2. After multiplying the background image flux by this factor the cluster images were added and the final images normalized as usual. This encourages the network to learn parameter inference regardless of whether the calibrations for backgrounds match mock clusters well. As can be seen when comparing Figs. 3.10 and 3.11, the inference results are very similar, only with the error spread increasing for each parameter by about 10%. This implies that accurate calibrations, while still associated with slightly more precise results, are not essential for a CNN to derive cluster parameters.

Fig. 3.12 shows the derived $class_{c/b}$ and $visibility$ values for the 5,000 test mock clusters, as well as a random sample of 5,000 M83 background images.

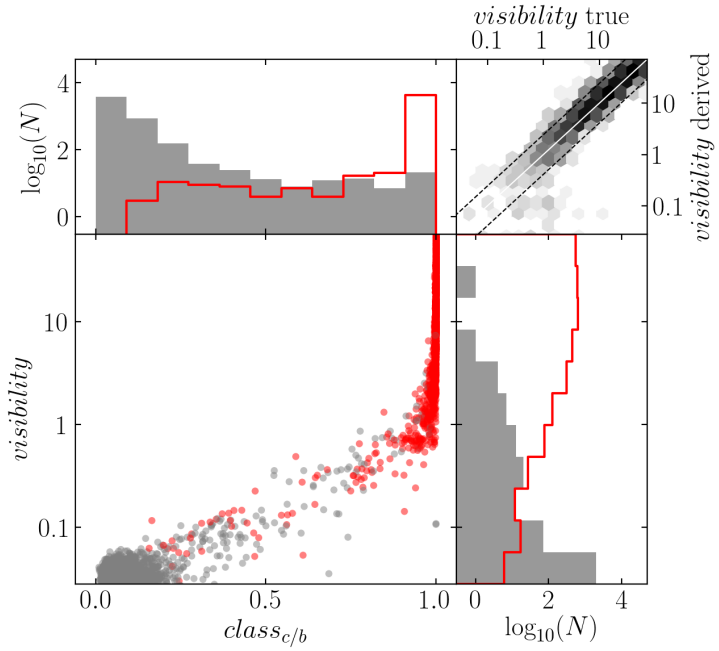

Figure 3.12. *Visibility* parameter values vs. $class_{c/b}$ for mock (red) and random M83 field (gray) samples. The histograms are marginalized logarithmic counts of samples for *visibility* (right) and $class_{c/b}$ (top). Top-right panel shows true vs. derived *visibility* parameters of mock clusters as a logarithmic density map; dashed lines outline the area containing 95% of the clusters.

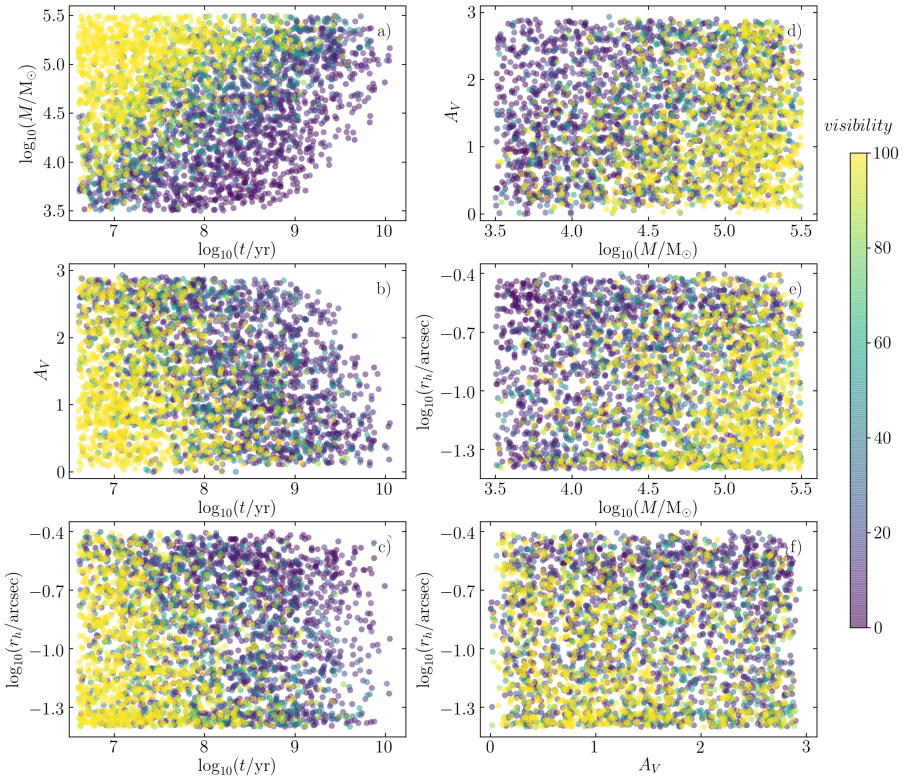

Figure 3.13. Inferred parameter distributions for the test cluster sample. The panels show the following parameter combinations: a) mass vs. age, b) extinction vs. age, c) size vs. age, d) extinction vs. mass, e) size vs mass, and f) size vs. extinction. Diagonal cutoffs in panels a) and b) are related to the cluster detection limit, applied as magnitude cuts, shown in Fig. 3.3. The color map represents the *visibility* parameter, which acts as a proxy for the selection effects in a magnitude-limited sample, while also taking into account variable cluster sizes and extinctions.

As can be seen in the histogram on top, the $class_{c/b}$ parameter is predicted as $>$ 0.5 for the vast majority of mock cluster images, and as < 0.5 for the majority of background images. This suggests that the fraction of background images that are classified as $class_{c/b} > 0.5$ are likely to correspond to real clusters. The *visibility* parameter is highly correlated with $class_{c/b}$, again showing high values for the majority of mock clusters and low values for the majority of backgrounds. The few remaining mock clusters with $class_{c/b} < 0.5$ have very low *visibility* values, which indicates faint, nearly invisible objects seen in Fig. 3.4.

Fig. 3.13 illustrates selection effects by showing the derived age, extinction, mass, and size parameters of the test mock clusters, with the color bar representing the derived *visibility* parameter value for each cluster. In Fig. 3.13a it can be seen that mass and age are correlated as expected when deriving the *visibility* parameter: clusters with lower mass and older ages tend to be less visible (this can also be seen in Figs. 3.1 and 3.2). The same is true for extinction (panels b and d), as higher extinctions tend to make cluster less visible, and size (panels c, d, and f), as more concentrated clusters stand out relative to their backgrounds.

Even though the cluster-related parameter inference results for background images have no inherit meaning, the CNN produces values for all of its output neurons regardless. Looking at these values can provide us with additional insights. For example, we would expect backgrounds to be classified as low-mass extended objects. Fig. 3.14 shows the derived parameters for the background images from Fig. 3.12, with dot size and color indicating $class_{c/b}$. Black dots are images with $class_{c/b}$ close to 0, while red circles are images with $class_{c/b}$ close to 1. As can be seen in Fig. 3.14e, the vast majority of the backgrounds are classified as low-mass extended objects as expected, with some probable cluster images being spread out more evenly through the parameter space. The derived age values of these images are spread out through the whole age range (panels a, b, and c), however extinctions are heavily correlated with ages as seen in panel b. As the network is trained to predict extinction and age values regardless of what the cluster's background looks like, there is no intuitive value that should be predicted for background images in this case. In effect the CNN avoids areas of age-extinction parameter space where the appearance of an observed object is either extremely blue (low-extinction low-age) or extremely red (high-extinction high-age), which can only be associated with genuine clusters, resulting in this diagonal effect.

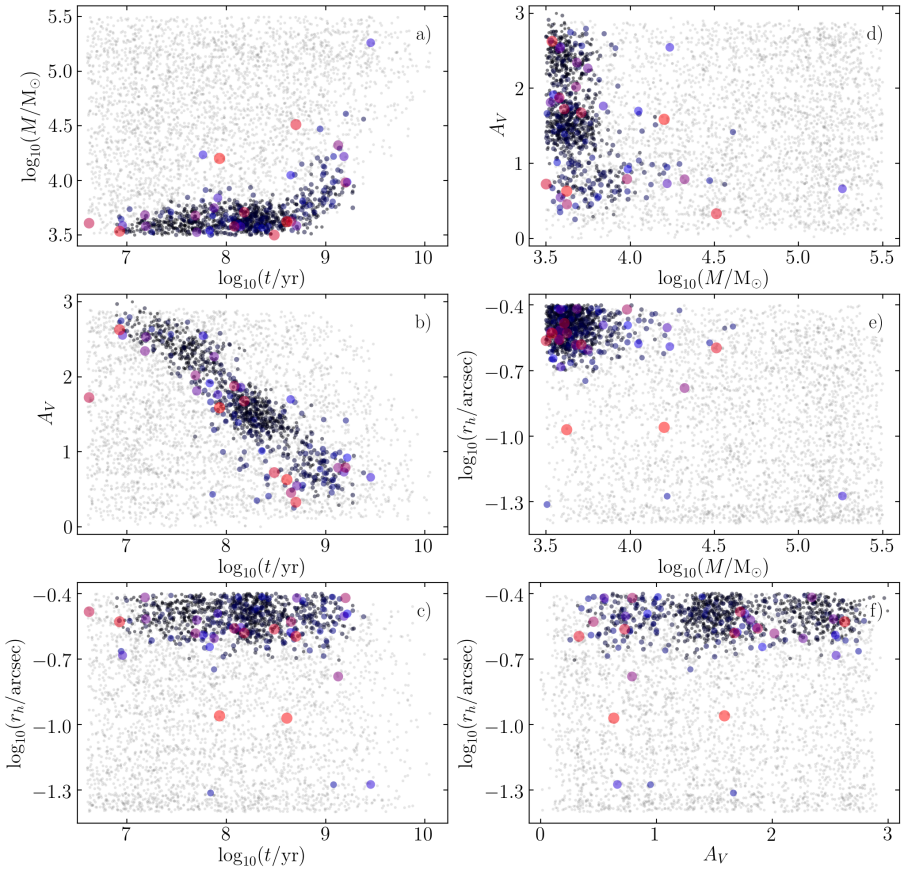

Figure 3.14. Same as in Fig. 3.13, but for 5,000 randomly spatially sampled M83 background images. For reference, derived mock cluster parameters are shown as faint gray dots. Real background samples are shown as black dots, that transition to blue and then to red. The dot size and color represents $class_{c/b}$. The blue dots are objects with $class_{c/b} > 0.5$ and the red dots are objects with $class_{c/b} > 0.99$.

The $class_{c/b}$ parameter was shown to be useful in differentiating between cluster and background images, while the *visibility* parameter is correlated well with those cluster parameter ranges which can show more confidently identified clusters. We conclude that these parameters can be useful indicators in star cluster search application.

3.3.2 Validation with cataloged clusters

To validate our method on real clusters we used three previous M83 HST star cluster studies which had published catalogs. This includes the study covering the whole galactic disk (7 WFC3 fields) by Ryon et al. (2015, R15), two WFC3 fields by Bastian et al. (2011, B11) and the galaxies central region by Harris et al. (2001, H01).

The study by Bastian et al. (2011) is comprised of 939 objects. We discarded objects with missing parameter values, leaving us with 889 of them to compare to the CNN inference results. Bastian et al. (2011) estimated the cluster age, mass, and extinction by comparing the integral photometry of the observed clusters to SSP models. Meanwhile, the sizes of clusters were estimated by fitting spatial models to F438W, F555W, and F814W band images. For this comparison we took the median value of these three size estimates. As the cluster magnitudes used by Bastian et al. (2011) were Galactic extinction corrected, we shift the A_V values of those objects by 0.3 mag[6]. This was done so that we could compare CNN-derived values directly, because we compute total extinctions for clusters regardless of the dust source.

Figs. 3.15 and 3.16 show a comparison between Bastian et al. (2011) and CNN-derived values. In Fig. 3.15 the red and blue dots represent clusters with significantly overestimated and underestimated extinction values respectively. They were defined as clusters that are outside the dashed lines in panel c, which represent the area containing 95% of mock cluster parameter derivations. This mirrors the situation with mock objects in Fig. 3.10, as the majority of clusters with overestimated extinctions end up with underestimated ages, and vice-versa for clusters with underestimated extinction values. These effects can again be attributed to the age-extinction degeneracy. In Fig. 3.16 the green dots represent images classified by the network as likely to be real clusters ($class_{c/b} \geq 0.5$), while the magenta dots are objects with $class_{c/b} < 0.5$. The vast majority of the objects are classified as likely clusters.

[6]https://irsa.ipac.caltech.edu/applications/DUST/

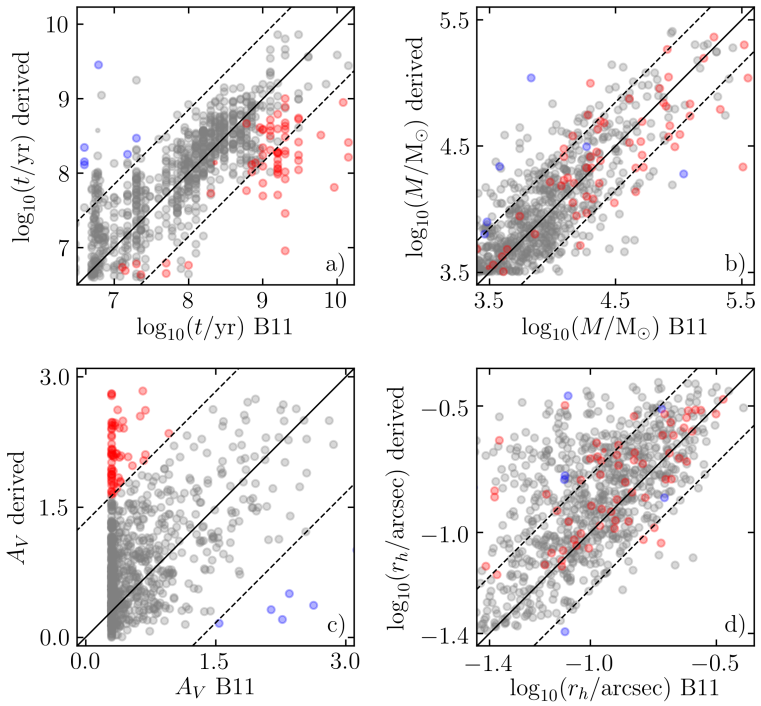

Figure 3.15. The comparisons for a) age, b) mass, c) A_V, and d) r_h derived by Bastian et al. (2011) and by CNN. Red dots represent clusters with overestimated, while blue dots represent clusters with underestimated extinction values with respect to Bastian et al. (2011). The remaining clusters are colored gray. Dashed lines outline the area containing 95% of the mock clusters.

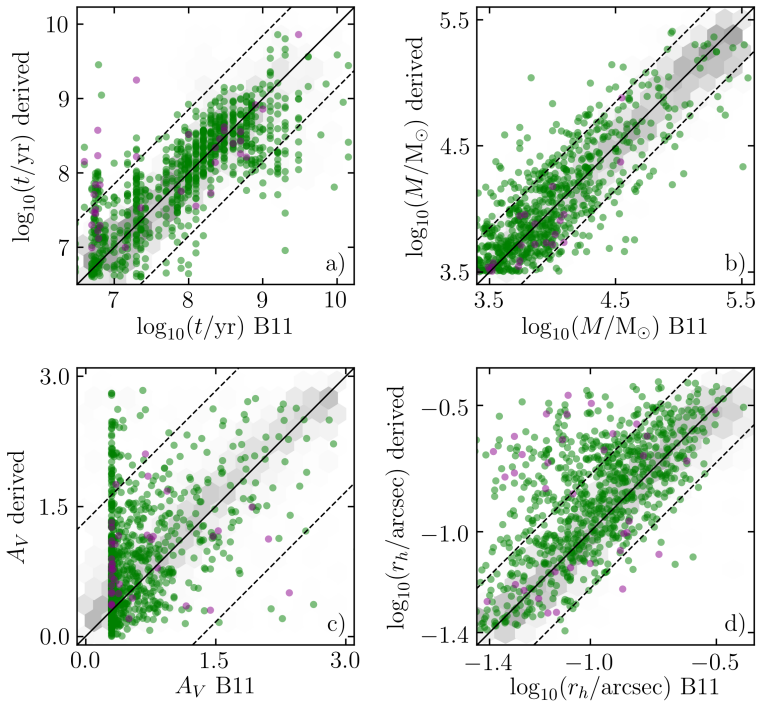

Figure 3.16. Same as Fig. 3.15, but with large green dots representing objects classified as likely clusters ($class_{c/b} \geq 0.5$), and magenta dots representing objects with $class_{c/b} < 0.5$. Mock clusters are displayed in the background as a hexagonal density maps with bins scaled linearly.

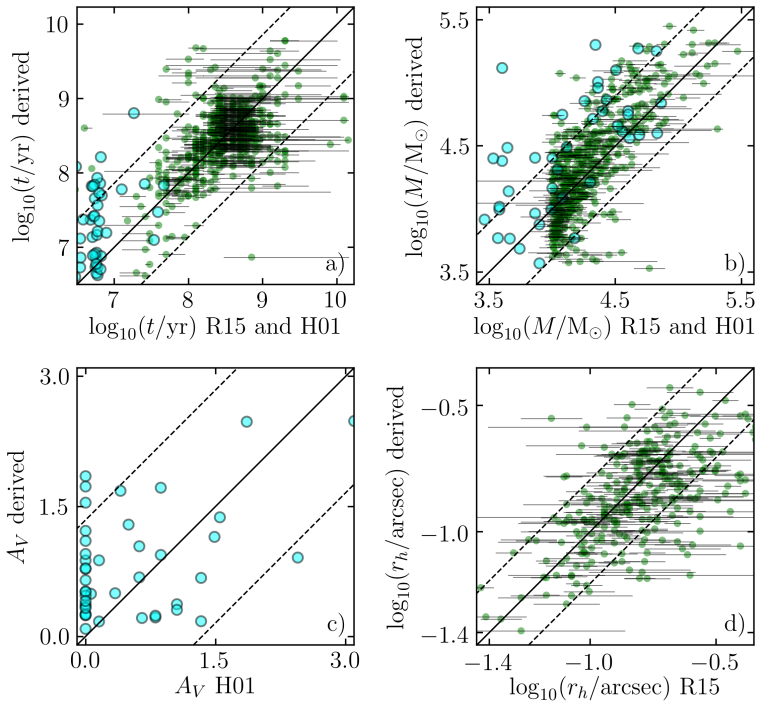

Figure 3.17. Same as Fig. 3.16, but with green dots representing objects from Ryon et al. (2015) and cyan dots representing objects from Harris et al. (2001). The horizontal bars denote minimum and maximum parameter values for age and mass, and statistical errors for size, as provided in the catalogs.

Overall the derived ages and masses show a fairly good agreement between Bastian et al. (2011) and CNN-derived values. Many of the objects have cataloged $A_V = 0$ mag values (shown as $A_V = 0.3$ mag in the figures, accounting for Galactic extinction). The CNN derives higher extinctions for some of these clusters, however, visual inspection has revealed that Galactic dust is unlikely to be the only source of extinction for the majority of them. The sizes show a good agreement for most of the objects, however, there is a subset of objects with somewhat overestimated values.

For the comparison with Ryon et al. (2015) we used 478 objects which had sizes obtained by 2D spatial model fitting as well as age and mass estimates derived using spectral energy distribution fitting. We also took 45 objects from Harris et al. (2001) with their age, mass, and extinction estimates obtained by comparing the cluster photometry to theoretical population synthesis models. Fig. 3.17 shows our results compared against both of these catalogs. Ryon

et al. (2015) objects are denoted as green dots with the parameter error bounds marked with black lines. Harris et al. (2001) objects are marked as large cyan circles. For both of these catalogs a reasonable agreement with the CNN-derived values can be seen, with only masses being slightly overestimated. However, there's some age estimate divergence over $\log_{10}(t/\mathrm{yr}) = 8$, which is similar to the situation in Fig. 3.15.

We have shown that the CNN is capable of deriving cluster parameters on real clusters by comparing our results with those of other authors. The agreements between the values are reasonable and follow the results obtained with mock clusters. However, due to the age-extinction degeneracy with the used 3 passbands, the results with clusters older than $\log_{10}(t/\mathrm{yr}) = 8$ are ambiguous and should be interpreted carefully.

3.4 Discussion

We have shown the applicability of a CNN-based method in deriving a variety of star cluster parameters from M83 mosaic images in terms of quantitative error analysis. However, the final aim for this method is to be of use in star cluster search and automatic catalog construction. To this end, a better look into the derived parameters is needed both in terms of each other, and their context in the galaxy. In this chapter we look at derived values of the Bastian et al. (2011) sample of objects in more detail.

Fig. 3.18 shows the inferred age, extinction, mass, and size parameters of the Bastian et al. (2011) object sample. The objects are colored as in Fig. 3.16, with mock results shown in the background. The clusters cover the whole parameter range well, with $class_{c/b} < 0.5$ samples being classified as expected: as low-mass objects (panel a). The minimal extinction line, with a large number of clusters around it, seen in panels b, d, and f, coincides with $A_V \sim 0.3$ mag, expected due to Galactic dust foreground in the direction of M83. Lines of constant density are shown in panel e. The majority of the objects fall between 10 and 1000 $\mathrm{M}_\odot/\mathrm{pc}^{-3}$, which is consistent with results for clusters of the M31 galaxy (Vansevičius et al. 2009).

Fig. 3.19 shows Bastian et al. (2011) objects marked on two fields of the M83 mosaic. Objects of $\log_{10}(t/\mathrm{yr}) < 7.4$ are marked as blue circles in panel a, $\log_{10}(t/\mathrm{yr}) < 8.6$ objects are marked as orange circles in panel b, and $\log_{10}(t/\mathrm{yr}) \geq 8.6$ objects are marked as red circles in panel c. Panel d shows all of the objects marked as dots, with $A_V < 1$ mag colored cyan, and $A_V \geq 1$ mag

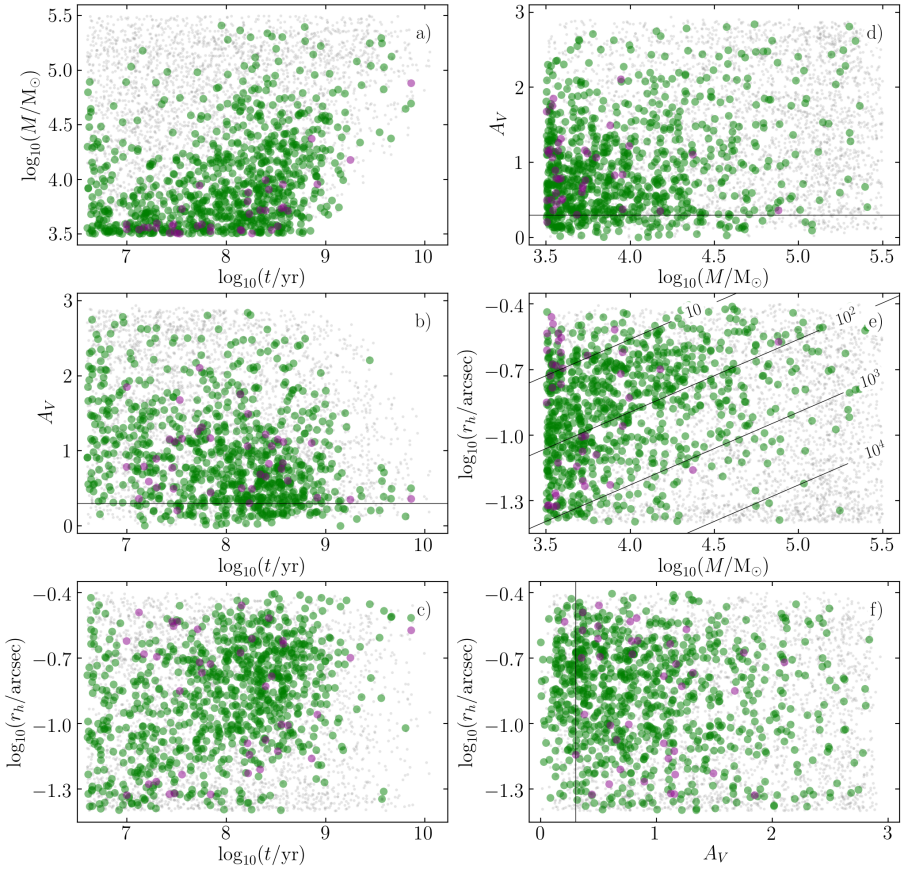

Figure 3.18. Same as Fig. 3.13, but for objects from Bastian et al. (2011), with values derived by CNN. The green circles represent objects classified as likely clusters, while magenta circles represent likely non-clusters, as in Fig. 3.16. For reference, the derived parameters of the mock cluster set are shown as faint gray points. In panel e lines show locations of clusters with the same density, varying from 10 to 10^4 $\mathrm{M}_\odot/\mathrm{pc}^{-3}$. In panels b, d, and f the solid black lines represent the amount of Galactic extinction in the direction of M83 ($A_V = 0.3$ mag).

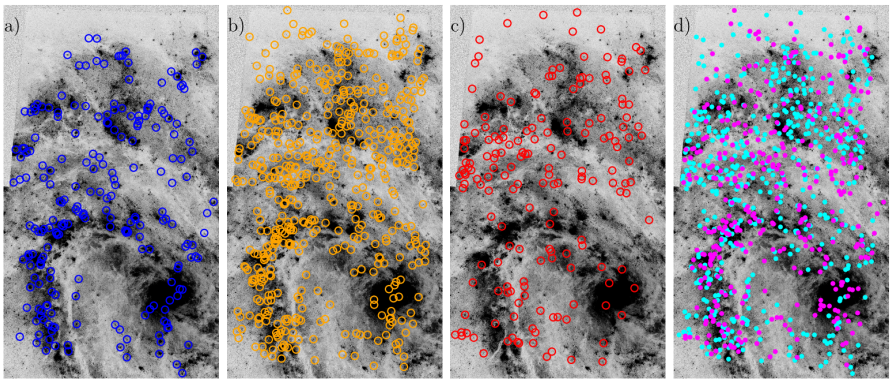

Figure 3.19. M83 mozaic of the F438W passband observations, overplotted with young – $\log_{10}(t/\mathrm{yr}) < 7.4$ (blue), intermediate – $7.4 \geq \log_{10}(t/\mathrm{yr}) < 8.6$ (yellow), and old – $\log_{10}(t/\mathrm{yr}) \geq 8.6$ (red) objects from Bastian et al. (2011). The last panel depicts objects with low – $A_V < 1$ mag (cyan) and high – $A_V \geq 1$ mag (magenta) extinctions.

colored magenta. The spatial distribution of objects is sensible, with young star clusters grouping around the galaxy's spiral arms, near the dust clouds where they were formed, and old clusters spread out more evenly throughout the galaxy, as they had more time to drift away. The extinction distributions are less clear-cut, however some crowding of high-extinction objects around dust-heavy regions can be seen, as is expected. The spatial distributions of age-selected clusters in Fig. 3.19 correspond well to the results obtained by Fouesneau et al. (2012) using UBVIHα fluxes to measure ages, masses, and extinctions in the central region of M83. Sánchez-Gil et al. (2019) has derived age maps for the M83 galaxy for stellar populations younger than 20 Myr, which corresponds to the lower age range of clusters in this work.

We studied clusters with masses $\log(M/\mathrm{M}_\odot) = [3.5, 5.5]$. This does not imply that only such clusters are detectable with the HST/WFC3 observations of M83. In fact, clusters of masses as low as $\log(M/\mathrm{M}_\odot) \sim 3$ have been studied by Whitmore et al. (2011) and Andrews et al. (2014). However, such clusters are dominated by stochastic effects of IMF sampling making the analysis of the effects of extinction problematic. The lower-limit of masses was selected to focus on the effects of extinction as well as to align with the range of clusters used by Bastian et al. (2011). The presented CNN classifies lower mass clusters as being on the lower-limit of this range.

Fig. 3.20 shows the binned mass distributions obtained with the CNN. The

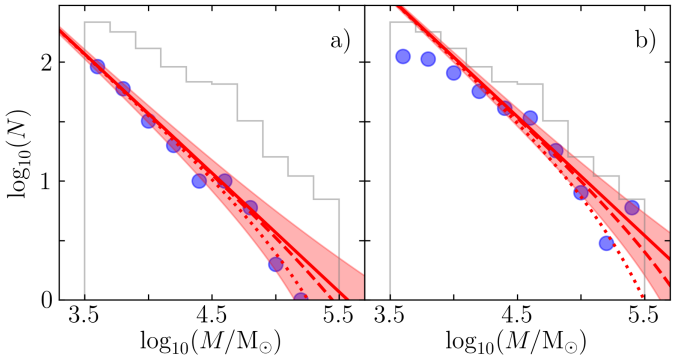

Figure 3.20. Cluster mass distributions for samples with ages $\log_{10}(t/\mathrm{yr}) < 7.7$ (a) and $7.7 < \log_{10}(t/\mathrm{yr}) < 8.7$ (b). Lines represent the power law mass distribution function of the form $dN/dM = A \cdot M^{-2} \cdot \exp(-M/M_*)$: $M_* = \infty$ (solid line, with the shaded area encompassing its Poisson standard deviation), $10^6\ M_\odot$ (dashed line), and $2.5 \cdot 10^5\ M_\odot$ (dotted line).

gray outline shows all of the cluster distributions, with blue dots representing clusters of $\log_{10}(t/\mathrm{yr}) < 7.7$ (panel a), and $7.7 \leq \log_{10}(t/\mathrm{yr}) < 8.7$ (panel b). The red lines represent Schechter type mass functions (Portegies Zwart et al. 2010b) with various amounts of truncation. The solid red line follows the non-truncated power law $dN/dM = A \cdot M^{-2}$, the dashed red line follows $A \cdot M^{-2} \cdot \exp(\frac{-M}{10^6})$, and the dotted red line follows $A \cdot M^{-2} \cdot \exp(\frac{-M}{2.5 \cdot 10^5})$. The power law distributions fit the data well for both of the age cuts, however, there is a lack of low-mass clusters ($\log_{10}(M/\mathrm{M}_\odot) \leq 4$) for the mid-age data sample. This is due to selection effects, with fewer star clusters being detectable at those ages (see Fig. 3.18a). Similar cluster mass distributions and selection effects have been found in M31 (Vansevičius et al. 2009) and M33 (de Meulenaer et al. 2015b) star cluster samples.

Fig. 3.21 show examples of inference results on 3 distinct Bastian et al. (2011) clusters chosen to illustrate the variety of CNN outputs (previously sketched in Fig. 3.7). The top figure shows a young, low-mass cluster. The inferred age and mass matches Bastian et al. (2011) parameters well. Extinction is derived to be slightly higher, however the value is very close when Galactic extinction is accounted for. The *visibility* parameter is derived to be $\sim$15, which corresponds well to similarly looking clusters in Fig. 3.2b.

In the middle, a cluster of $\log_{10}(t/\mathrm{yr}) \sim 8.3$, medium extinction and high mass is shown. The age, mass, and size correspond well to the values derived

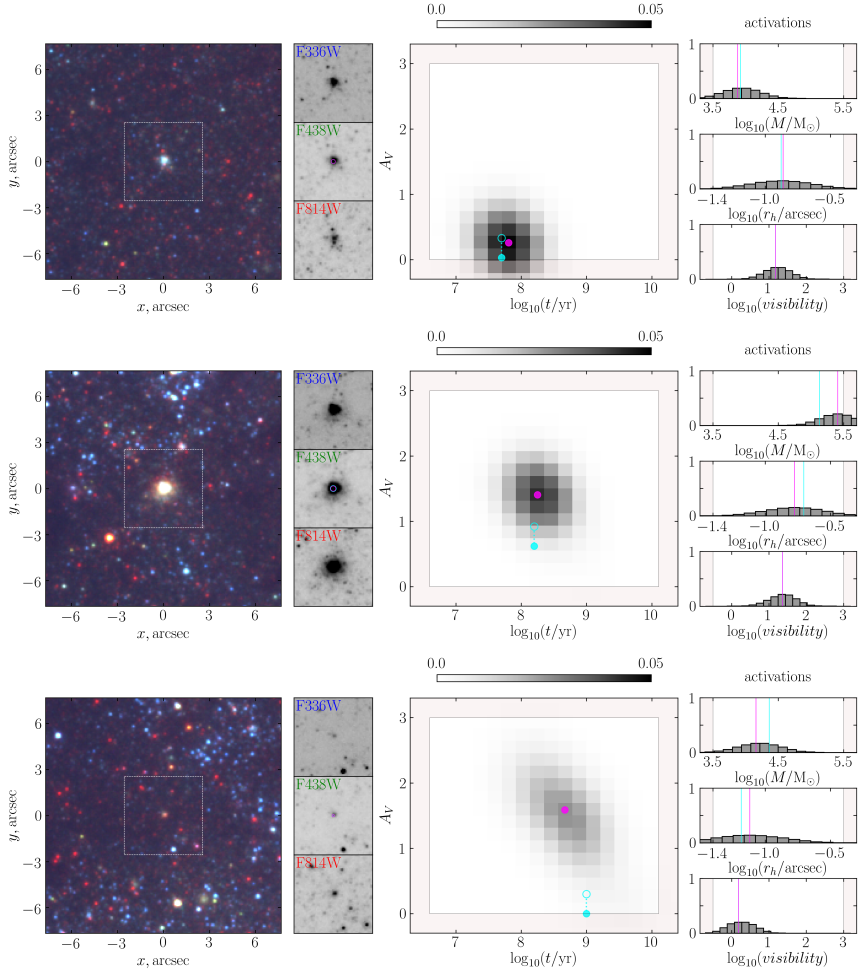

Figure 3.21. Examples of real clusters and their parameter distributions inferred by the CNN. The left panels show the color image of each cluster in a field of $7'' \times 7''$, with a $2.6'' \times 2.6''$ field used for inference highlighted with a dashed square. The remaining panels depict inference results, with the histograms showing the neural network's neuron activations for the given parameter. Age and extinction is depicted as a 2D activation map marginalized over mass to highlight the effects of age-extinction degeneracies, with the color bar on the top indicating CNN output neuron activation strength. Parameter values derived by Bastian et al. (2011) are marked in cyan, and the CNN-inferred values in magenta. Light-red shaded areas show parameter ranges where the CNN produces activations, but which were not covered by the clusters used in this work to deal with parameter boundary effects. The empty cyan circle in the age-extinction map represents the values obtained after Galactic extinction correction.

by Bastian et al. (2011). Extinction is derived as slightly higher, however it's still within the range of CNN's activations. The cluster is classified as brighter by the CNN, with $visibility \sim 25$.

On the bottom an older cluster is depicted. Its mass and size estimates correspond well, however extinction is overestimated in comparison to Bastian et al. (2011). Furthermore, the neuron activations show a diagonal pattern highlighting the age-extinction degeneracy which is hard to resolve with the used 3 passbands. However, the higher-extinction results are more likely as a significant amount of the field seen in the leftmost panel appears reddened, which suggest the presence of dust obscuring the cluster.

As detailed in Section 3.2.3, there exists some correlation between the spread of CNN output activations and the difficulty of inferring cluster parameters.

These results further validate the applicability of the CNN in deriving the parameters of star clusters in realistic scenarios. In addition, the $class_{c/b}$ and $visibility$ parameters act as accurate proxies for cluster presence in images.

3.5 Conclusions

We have extended the method introduced in Chapter 2 to infer cluster ages, masses, sizes, extinctions, as well as to account for the degeneracies between them. Additional parameters were added for identifying the presence of clusters on background images of M83, and judging their visibility (signal-to-noise).

To train this network a bank of mock clusters was generated utilizing three photometric passbands in the context of the M83 galaxy. The CNN was verified on mock images of artificial clusters with ages, $\log_{10}(t/\mathrm{yr})$, between 6.6 and 10.1, masses, $\log_{10}(M/\mathrm{M}_{\odot})$, between 3.5 and 5.5, sizes between 0.04 and 0.4 arcsec, and extinctions $A_V \leq 3$ mag. Parameters derived by CNN have shown a good agreement with the true parameters for $\log_{10}(t/\mathrm{yr}) < 8$, with higher age estimates being unreliable due to the age-extinction degeneracy.

Real cluster parameter inference tests were performed with three different M83 cluster catalogs from Bastian et al. (2011), Ryon et al. (2015), and Harris et al. (2001) and have shown consistent results.

We have demonstrated that a CNN can perform evolutionary (age, mass), structural (size), and environmental (extinction) star cluster parameter inference. In addition, the network is capable of giving an indication of cluster

presence in images. Therefore, the created CNN is a useful tool for further research in constructing a full pipeline of star cluster detection and parameter inference.

Chapter 4

M83 cluster population inference

This chapter focuses on the application of a CNN to star cluster candidate detection and parameter inference on the M83 galaxy HST data. A full pipeline was constructed that performs star cluster analysis starting from multi-band images and resulting in evolutionary, structural, and environmental parameter estimates for 3,380 cluster candidates. This sample was used to analyse the dependence of cluster parameters on their galactocentric distance and positions relative to the spiral arms.

The chapter is structured as follows. In Section 4.1 we present the HST image data and the mock clusters used in the training and testing of the CNN. In Section 4.2 we present how the CNN is applied to cluster search, as well as the artificial cluster tests used to measure its performance. Finally, in Section 4.3 we explore the results on the detected sample of cluster candidates through various astrophysical aspects, and in Section 4.4 we discuss the limitations of the method and some of the caveats of our results.

4.1 Data

We used M83 images described in Chapter 3, again only selecting F336W, F438W, and F814W mosaics, which cover the whole galaxy without gaps and allow homogeneous cluster search and parameter inference throughout the galaxy.

Fig. 4.1 shows the M83 mosaic comprised of three passbands used in this work, with superimposed mock clusters. For the training of a CNN a large

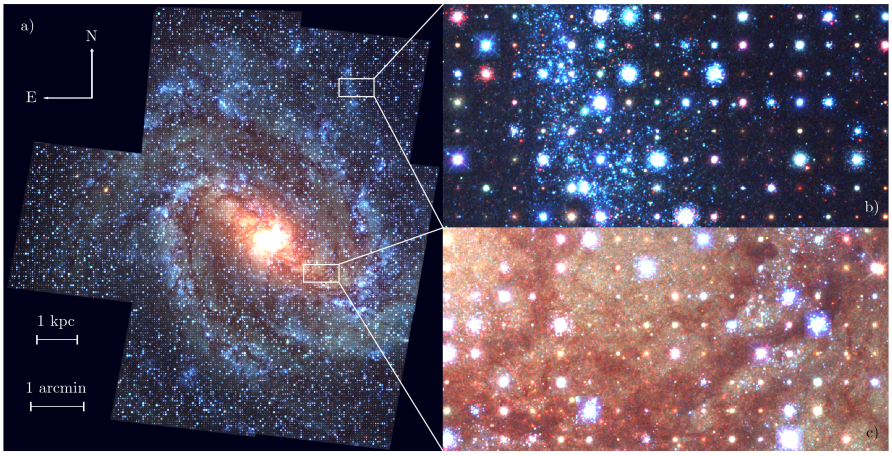

Figure 4.1. Panel a shows the mosaic of seven HST WFC3 fields of the M83 galaxy in three photometric passbands: F336W (blue), F438W (green), and F814W (red). Mock clusters used for artificial cluster tests ($\sim$26,000 in total) are superimposed on the image in a dense grid, placing a cluster every 64×64 pixels. Panels b and c show zoomed-in areas of the galaxy's outskirts and central region respectively. The superimposed mock clusters are of various luminosities and appearances due to the random sampling of their ages, masses, extinctions, and sizes.

Table 4.1. Star cluster parameters

Parameter	Notation	Value	Comment
	Sampled and inferred		
Age	$\log(t/\mathrm{yr})$	[6.6, 10.1]	discrete, step of 0.05 dex[a]
Mass	$\log(M/\mathrm{M}_\odot)$	[3.5, 5.5]	continuous
Extinction	A_V [mag]	[0, 3]	discrete, step of 0.1 mag[a]
Size	$\log(r_h/\mathrm{arcsec})$	$[-1.4, -0.4]$	continuous EFF parameter sampling
	Fixed		
Metallicity	Z	0.03	constant (Hernandez et al. 2019)
Distance	d [Mpc]	4.5	constant (Thim et al. 2003)
	Inferred		
Visibility	$visibility$	($\sim$0.1, $\sim$1,000)	continuous, a proxy for signal-to-noise
Cluster or background	$class_{c/b}$	[0, 1]	continuous, detection certainty
	Computed		
Density	$\log(\rho_h/(\mathrm{M}_\odot \cdot \mathrm{pc}^{-3}))$	[$\sim$0, $\sim$5]	continuous, computed within r_h

[a]From Padova isochrone bank with standard extinction law ($R_V = 3.1$).

number of mock cluster images with corresponding cluster parameters are required. Mock clusters are also needed to perform artificial cluster tests, which allow us to estimate the network's performance in an object search scenario. The procedure used for generating the mock clusters is the same as in Chapter 3, with four main randomly sampled parameters: age, mass, extinction, and size (see Table 4.1).

Only clusters of $m_{F336W} < 24$ mag, $m_{F438W} < 23.5$ mag, and $m_{F814W} < 23$ mag were selected for further use. See Figs. 3.1–3.4 for visualizations of mock clusters, as well as their parameter values.

4.2 Method

4.2.1 Convolutional neural network

We used a CNN based on the ResNet50 architecture (He et al. 2016), detailed in Chapter 3, capable of simultaneously inferring age, mass, extinction, size, $class_{c/b}$, and $visibility$ parameters (see Table 4.1). The CNN takes 64×64 pix images of three passbands as inputs and produces the aforementioned parameters as outputs. We simulated mock clusters as described in Chapter 3, placed them in randomly selected background fields of M83 and trained the network to distinguish between random stellar backgrounds and clusters, as

well as to simultaneously infer cluster parameters. The pixel values of each passband image were individually normalized to the mean of zero and standard deviation of one to minimize the influence of photometric image calibrations.

The continuous $class_{c/b}$ parameter gives us an indication of a cluster's presence in an image. Unlike the other parameters, it does not have a strict physical interpretation, and it's meant as a way for the CNN to express its certainty on whether the processed image contains a cluster. In contrast, *visibility* was developed as a proxy for signal-to-noise, and is defined as the ratio of a cluster's flux divided by the standard deviation of its background. This parameter lets us control the faintness of cluster candidates that we detect. Since $class_{c/b}$ works in the same way on all clusters regardless of their background or parameters, and the masses of our training clusters can go quite low, it becomes an unreliable measure when the clusters are very dim or drowned out by a dense background. Deciding whether objects like these should be included in a sample needs a separate source of information. For this purpose the *visibility* parameter is used. Lowering the values at which we cut off both of these parameters allows us to detect more of genuine clusters at the cost of more false positives.

4.2.2 Cluster search procedure

For cluster search (regardless of whether they're artificial or real) the network was applied on input images of fixed size, which were obtained by sliding a 64×64 pix window through the whole mosaic with a step of 1 pixel. In every such window the CNN inferred the parameters that indicate a cluster's presence (namely $class_{c/b}$ and *visibility*), as well as age, mass, extinction, and size. In effect, value-maps are obtained for the whole galaxy, where each pixel has an associated set of inferred parameters. Of particular note is the $class_{c/b}$ value-map, which can be interpreted as each pixel having an associated likelihood of there being a cluster centered on it. This approach is similar to Hausen & Robertson (2020), who perform source detection, segmentation and morphological pixel-by-pixel classification.

Star clusters of a wide variety of sizes exist and localizing their centers precisely can be difficult. For sparser clusters, two nearby pixels can easily be of nearly the same $class_{c/b}$ value and both of these pixels can be valid interpretations of where the center is located. In such scenarios, while trying to localize a cluster, simply taking pixels with high $class_{c/b}$ values is not enough. Instead, we first smoothed the $class_{c/b}$ map with a Gaussian kernel ($\sigma = 3$

pixels) and then searched for local peaks. This produces a large number of object candidates to work with.

Similarly, for each such object candidate, values of *visibility*, age, mass, extinction, and size parameters are obtained by taking several pixels around the peak of $class_{c/b}$ into account. This was done by computing a weighted average of the inferred parameters around a peak, using each pixel's $class_{c/b}$ value, as well as its distance from said peak in a 5×5 window, as the weight. This results in each object candidate having an associated set of inferred parameter values.

4.2.3 Artificial cluster tests

Artificial cluster tests were used to evaluate the CNN's detection performance. For that we generated ∼26,000 mock clusters and placed them on the mosaic in a regular grid. The result of this can be seen in Fig. 4.1. The mock clusters are spaced 64 pix apart from each other, covering the whole extent of the mosaic. Fig. 4.1b shows a sparser outskirt of the galaxy with an OB association. Most of the mock clusters can be easily seen in the sparser areas of this image, however, even bright clusters can be out-shined by the dense background of the OB association. Fig. 4.1c shows a central spiral-arm region of the galaxy, where fainter clusters are harder to see in the dense stellar background. The variety of mock cluster appearances are obtained due to the random sampling of cluster ages, masses, sizes, extinctions, as well the cluster's star masses and positions.

We performed the search for artificial clusters on the mock-cluster-filled mosaic and cross-matched the potential objects with the inserted mock cluster catalog, using a maximum distance of 3 pixels, in order to capture the mocks with a high confidence. This resulted in ∼23,000 objects displayed as blue points in Fig. 4.2. For reference, another sample of objects are displayed as black points and histograms, using a maximum distance of 6 pixels when crossmatching, which results in ∼24,600 objects. We repeated the same search procedure on the original M83 mosaic and found ∼200,000 object candidates with various $class_{c/b}$ and *visibility* values visualized as gray points in Fig. 4.2, which contain real cluster candidates we aim to select later.

In Fig. 4.2 it can be seen that the majority of recovered mock clusters have high $class_{c/b}$ and *visibility* values, but that is also true for a number of the original-mosaic detections. This is expected, as the gray points include real clusters of the M83 galaxy. We vary the thresholds for $class_{c/b}$ and *visibility*

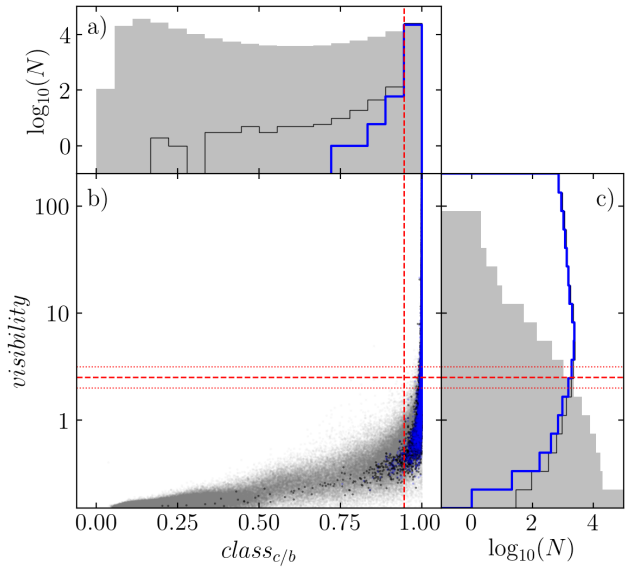

Figure 4.2. Detection results of star clusters with a CNN: 1) on the mosaic of M83 (gray, ∼200,000 object candidates) and 2) on the artificial cluster test mosaic seen in Fig. 4.1a (blue, ∼23,000 total recovered objects, which matched with the input list with a maximum distance of 3 pixels). Recovered ∼24,600 mock clusters with a less strict maximum cross-matching distance of 6 pixels are displayed as black histograms and dots. The histograms are logarithmic counts of objects marginalized over *visibility* (top) and $class_{c/b}$ (right). The dashed red line shows the chosen $visibility = 2.5$ and $class_{c/b} = 0.95$ thresholds, while the dotted red lines show $visibility = 2$ and 3.

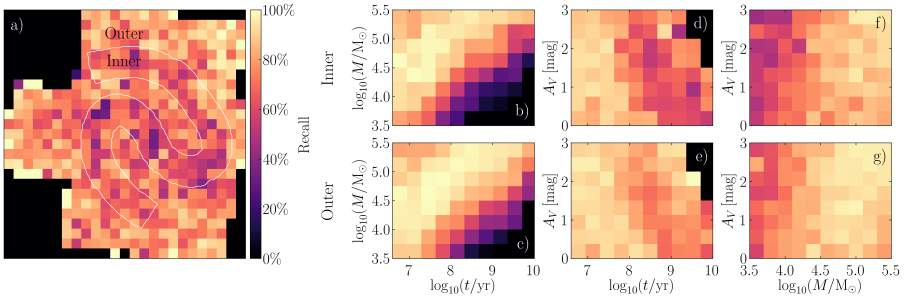

Figure 4.3. Recall (completeness) maps of artificial cluster tests. Each bin represents the number of objects that were recovered divided by the number of objects that were placed. Panel a shows recall of $\log_{10}(M/\mathrm{M}_\odot) < 4.5$ mock clusters as a function of their spatial position in the galaxy (as in Fig. 4.1), with the spiral arm regions outlined in white. Panels on the right show recall as a function of age and mass (b and c), age and extinction (d and e), as well as mass and extinction (f and g). The top row of panels (b, d, and f) show the recall within the spiral arms (inside the white outline in panel a), while the bottom row of panels (c, e, and g) show recall outside of the spiral arms. Note that the black bins signify areas where no artificial clusters were inserted.

to analyse recall (fraction of clusters that were found out of those that were inserted) as a function of spatial position in the galaxy and the mock cluster parameters in Fig. 4.3.

Fig. 4.3 shows recall maps of artificial clusters after applying the $class_{c/b}$ threshold of 0.95 and the *visibility* threshold of 2.5. In Fig. 4.3a the recall is shown spatially, for clusters with $\log_{10}(M/\mathrm{M}_\odot) < 4.5$. The spiral arm areas used for recall analysis are outlined in white. This divides the galaxy into two regions: inner (i.e. the area on the arms) and outer. For the inner regions the recall of the CNN is lower than for the outer regions, which is expected due to the relatively higher stellar background in and around the arms. A more detailed breakdown of recall is given in panels Fig. 4.3b-g.

In Fig. 4.3b and c the recall of clusters as functions of age and mass is shown. For both the inner and outer regions of the spiral arms recall decreases with older ages and lower masses. The effect is more pronounced for the inner region, because in a dense background faint clusters become much harder to find. Fig. 4.3d and e show recall as a function of age and extinction. The recall is broadly uniform, however, for the inner region lower values of recall can be observed for old and high extinction clusters. Fig. 4.3f and g show recall as a

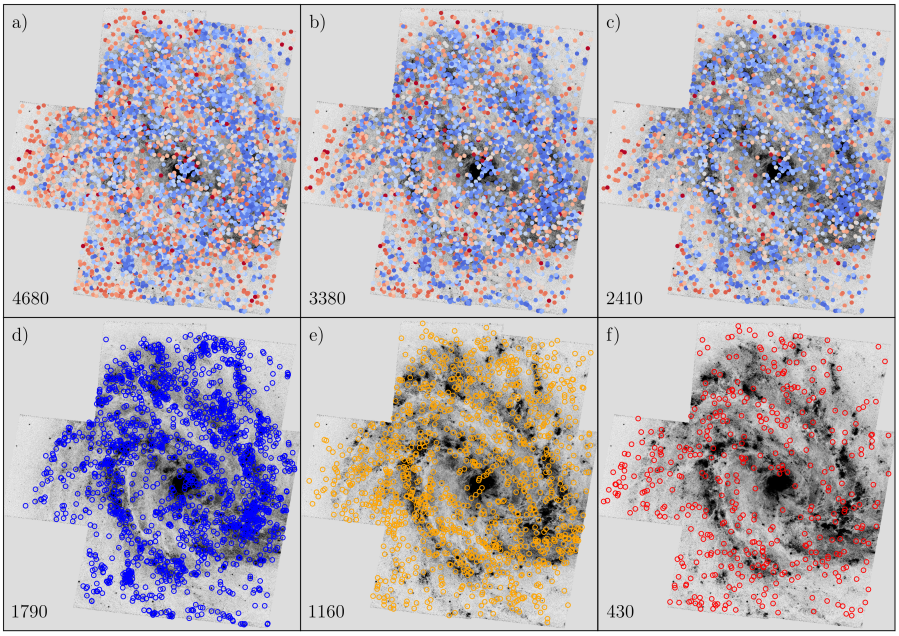

Figure 4.4. Detected cluster candidates on the M83 mosaic. The top row of panels depict different *visibility* thresholds used to filter out candidates: a) 2, b) 2.5, and c) 3, with blue indicating young clusters $\log_{10}(t/\mathrm{yr}) = 6.6$ and red – old clusters $\log_{10}(t/\mathrm{yr}) \geq 9$. The bottom row of panels depict clusters from panel b, split into ages of: d) $\log_{10}(t/\mathrm{yr}) < 7.5$, e) $7.5 \leq \log_{10}(t/\mathrm{yr}) < 8.5$, and f) $8.5 \leq \log_{10}(t/\mathrm{yr}) < 10.1$. The number of cluster candidates is displayed on the lower left of each panel.

function of mass and extinction. The decrease of recall occurs for low luminosity clusters of low masses and high extinctions. This is more pronounced on the spiral arms.

4.3 Results

4.3.1 Cluster selection

The ∼200,000 object candidates detected on the original M83 mosaic, depicted as gray dots in Fig. 4.2, were filtered using a *visibility* threshold of 2 and $class_{c/b}$ of 0.95. This resulted in 5,460 object candidates, which were inspected visually and obvious false positives were removed. This process left 4,680 cluster candidates with $visibility > 2$ and 3,380 cluster candidates with

our baseline threshold of $visibility > 2.5$. The visual inspection of possibly missed cluster candidates with $visibility > 2.5$ was estimated to be in line with artificial cluster tests seen in Fig. 4.3, however we did not include any human-detected clusters and analyse only CNN-detected candidates for consistency.

Fig. 4.4 shows these candidates on the M83 mosaic using several *visibility* thresholds, as well as highlighting the spatial distributions of cluster ages. The top row of panels in Fig. 4.4 depict different *visibility* thresholds, with panel b representing our baseline threshold, denoted as a red dashed line in Fig. 4.2. Increasing or decreasing the threshold by 0.5 produces a difference of $\sim$30% in the number of candidate counts.

The bottom row of panels in Fig. 4.4 depict clusters from panel b, grouped by age. Young cluster candidates seen in panel d are concentrated around the spiral arms and other star forming regions, while intermediate and older clusters in panels e and f are spread around relatively uniformly, with a somewhat higher concentration around spiral arms for the intermediate age clusters.

4.3.2 Inferred parameters

Fig. 4.5 shows the inferred age, extinction, mass, and size distributions of the cluster candidates. Diagonal cutoffs due to the *visibility*-limited detection can be seen in Fig. 4.5a for old low-mass clusters and in Fig. 4.5b for old high-extinction clusters. This was also seen for artificial clusters in Fig. 4.3. In addition, because of age-extinction degeneracies, noted in Chapter 3, arising due to the passbands that were used, parameter inference for old clusters can be unreliable.

In Chapter 3 we have compared the CNN inference results to catalogs of Bastian et al. (2011), Ryon et al. (2015), and Harris et al. (2001) showing a reasonably good agreement between the catalogs and the CNN results. The parameter distributions of cluster candidates seen in Fig. 4.5 also corresponds well to the Bastian et al. (2011) sample, inferred with the CNN and displayed in Fig. 3.18 of Chapter 3.

However, a numerous population of young-age low-mass extended cluster candidates (which by visual inspection could be classified as small associations of stars), is clearly visible in the bottom-left of Fig. 4.5a and the top-left of Fig. 4.5c. These are low density objects (Fig. 4.5e), with many of them having density values of $\rho_h < 100\,\mathrm{M}_\odot \cdot \mathrm{pc}^{-3}$. Mass and size relations of clusters in various galaxies are summarized in Fig. 9 of Krumholz et al. (2019), indicating that the majority of objects follow a $\rho_h \sim 100\,\mathrm{M}_\odot \cdot \mathrm{pc}^{-3}$ line and that the ranges

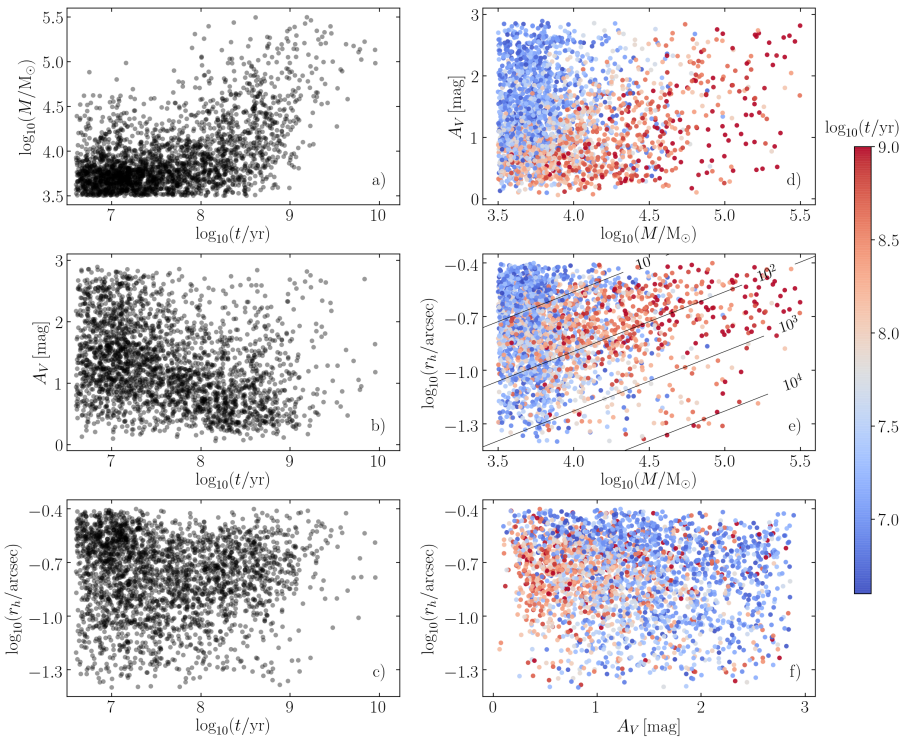

Figure 4.5. Parameter distributions of 3,380 cluster candidates detected by CNN on the M83 mosaic. Panels show combinations of: a) mass vs. age, b) extinction vs. age, c) size vs. age, d) extinction vs. mass, e) size vs. mass, and f) size vs. extinction. The colors in panels d, e, and f represent cluster ages, with blue indicating young clusters with $\log_{10}(t/\mathrm{yr}) = 6.6$ and red – old clusters with $\log_{10}(t/\mathrm{yr}) \geq 9$. The diagonal lines in panel e represent various cluster density $\rho_h/(\mathrm{M}_\odot \cdot \mathrm{pc}^{-3})$ levels.

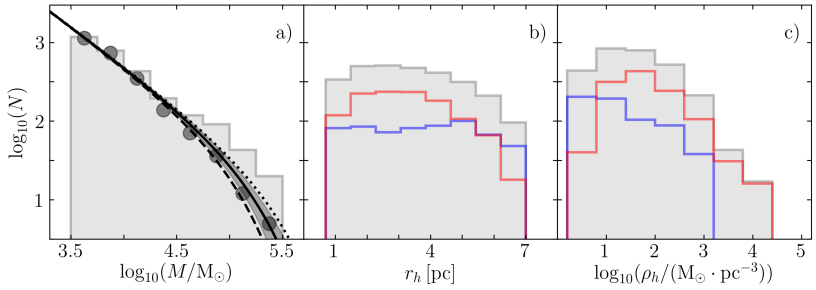

Figure 4.6. Mass (a), size (b), and density (c) histograms of the 3,380 cluster candidates. In panel a circles represent the distribution for clusters with ages $\log_{10}(t/\mathrm{yr}) < 8.5$, while the shaded gray histogram represents the whole cluster sample. Lines represent the Schechter mass distribution functions of the form $dN/dM = A \cdot M^{-2} \cdot \exp(-M/M_*)$, where M_* is: 1) 10^5 (dotted line), 2) $1.6 \cdot 10^5$ $M_\odot$ (solid line, with the shaded area encompassing its Poisson standard deviation), and 3) $2.5 \cdot 10^5$ $M_\odot$ (dashed line). In panels b and c the gray shaded histograms show the whole cluster sample, the blue outlines – young $\log_{10}(t/\mathrm{yr}) < 7$ and red – older $\log_{10}(t/\mathrm{yr}) \geq 7.7$ clusters.

of density are consistent with CNN results.

Although the definition of what observationally should be considered a cluster vs. an unbound association is not clear at such young ages (Bastian et al. 2012), these objects are detected by our CNN as cluster candidates. Comparing the cluster appearance of different catalogs one can see a variety of inclusion criteria and in our case this is decided by the CNN and the parameter ranges of mock clusters on which it was trained.

In addition, in both Bastian et al. (2012) Fig. 14 and Ryon et al. (2015) Fig. 3 a trend can be seen connecting cluster age and size. We can also see a similar effect in Chapter 3 Fig. 3.18, where parameters of the same catalog are derived via CNN. However, this is not present in our cluster candidate sample, as seen in Fig. 4.5 panel c, where a large number of young low-mass extended objects are observed.

Portegies Zwart et al. (2010b) discussed the importance of detection limits and their influence on the lack of low mass clusters in catalog construction. This is explored in Fig. 4.6a, where circles represent the mass distributions of clusters with ages $\log_{10}(t/\mathrm{yr}) < 8.5$. The superimposed lines are Schechter type mass functions (Portegies Zwart et al. 2010b), of the form $dN/dM = A \cdot M^{-2} \cdot \exp(-M/M_*)$, with different amounts of truncation, which

were previously derived by Bastian et al. (2012) on their sample of M83 clusters. The Schechter type function matches the data, explaining the numerous low-mass clusters seen in Fig. 4.5 and is consistent with Mok et al. (2020).

In Fig. 4.6b we show the distributions of sizes of young and older cluster candidates. The sizes are somewhat evenly distributed, with a drop for smaller and larger clusters. This distribution is similar to the one found for Galactic clusters based on Gaia observations by Sánchez et al. (2020) and consistent with previous results for M83 clusters (Ryon et al. 2015). In our sample, older clusters are smaller than younger clusters, which is as expected due to the fact that extended clusters are disrupted more easily due to various factors such as galactic tidal forces and passing through molecular clouds in the spiral arms. These effects can also be seen in Fig. 4.6c, with older clusters being more dense than younger clusters.

4.3.3 Distance from spiral arms

According to the density wave theory (Shu 2016), the spiral arms of galaxies are structures made up of objects and gas which move into them, slow down, and then move out again. In this model the spiral pattern itself rotates at a fixed angular velocity, while objects, such as star clusters, move at differing angular velocities, which depend on their distance to the galactic center. In this scenario one would expect star clusters inside the corotation radius to outpace the spiral arms, creating an age gradient with older clusters being observed further away from the arms than younger clusters. This is supported by observations, such as Miller et al. (2019), who measured the pitch angle of spiral arms for a sample of galaxies at several wavelengths, finding that it is systematically larger in blue and smaller in red passbands, claiming the results to be consistent with enhanced stellar light located downstream of a star forming region for a galactocentric radius smaller than the corotation radius. Sofue (2018) identified bow shock structures as wavy arrays of star forming regions distributed along the spiral arms of M83 and used the shape and orientation of HII region shock cones to estimate the position of the corotation radius coinciding to the one estimated by Hirota et al. (2014).

To analyze the distribution of parameters as a function of the distance from the spiral arms, we traced out dust lanes along the inner part of the arms as a reference. This follows the approach of Shabani et al. (2018), where cluster distances from dust lanes were used to analyze age gradients in three spiral galaxies. However, we used linear instead of azimuthal distances due to pro-

jection effects and the possibly warped disk of M83 (see Section 4.4 for more details).

Fig. 4.7 shows the counts of cluster candidates as a function of their shortest-line distances from the dust lanes of the spiral arms of the galaxy (see Fig. 4.9a). Note that for trailing (negative distance) cluster candidates we only collect objects up to about $\sim$1 kpc, before running into the leading (positive distance) cluster candidates of a different spiral arm (see Fig. 4.4a) and vice versa. This has to be kept in mind as any trends beyond those points will exhibit boundary effects, limiting the interpretation of this data within $\pm$1 kpc.

Fig. 4.7a shows 3 different cuts of ages, with blue indicating $\log_{10}(t/\mathrm{yr}) < 7$, orange – $7 \leq \log_{10}(t/\mathrm{yr}) < 7.7$, and red – $\log_{10}(t/\mathrm{yr}) \geq 7.7$. The blue and orange lines represent recently formed clusters, associated with the spiral arm, while the red line shows the field population. The peak for young clusters appears closer to the spiral arms, at $\sim$0.4 kpc in the leading direction, while for the oldest field population of clusters it stays approximately flat. This corresponds well to the result obtained for other spiral galaxies in Shabani et al. (2018).

Assuming the rotation curve speed of $\sim$160 km/s (Heald et al. 2016) and the spiral pattern speed of $\sim$110 km/s at the distance of $\sim$2.5 kpc (Zimmer et al. 2004), where the majority of our cluster candidates (that are in the vicinity of the spiral arms) are located, we expect objects to overtake and lead the arms at the rate of $\sim$50 pc per one million years. This is in good agreement with the peak we find of $6.6 \leq \log_{10}(t/\mathrm{yr}) < 7$ cluster candidates at $\sim$0.4 kpc and $7 \leq \log_{10}(t/\mathrm{yr}) < 7.7$ cluster candidates at $\sim$0.7 kpc. Note also that cluster candidates of $\log_{10}(t/\mathrm{yr}) \approx 8$ have already completed a full orbit around the galaxy, which supports our lower age cut of $\log_{10}(t/\mathrm{yr}) \geq 7.7$ as representing the field population of clusters.

Fig. 4.7b shows the extinctions of older clusters, with ages of $\log_{10}(t/\mathrm{yr}) > 8$. This age cut was chosen to select a sample of clusters that are already independent of the regions where they were formed, giving us an insight into the extinctions of the field population. The blue line indicates low extinction objects with $A_V < 1$ mag, while the red line indicates high extinction objects with $A_V \geq 1$ mag. The peak of the number of objects with high extinction occurs right on top of the spiral arm dust lanes. On the other hand, the number of low extinction objects gradually increases leading the arms, reaching the highest counts at the distance of $\gtrsim$1 kpc in front of the arms. This corresponds

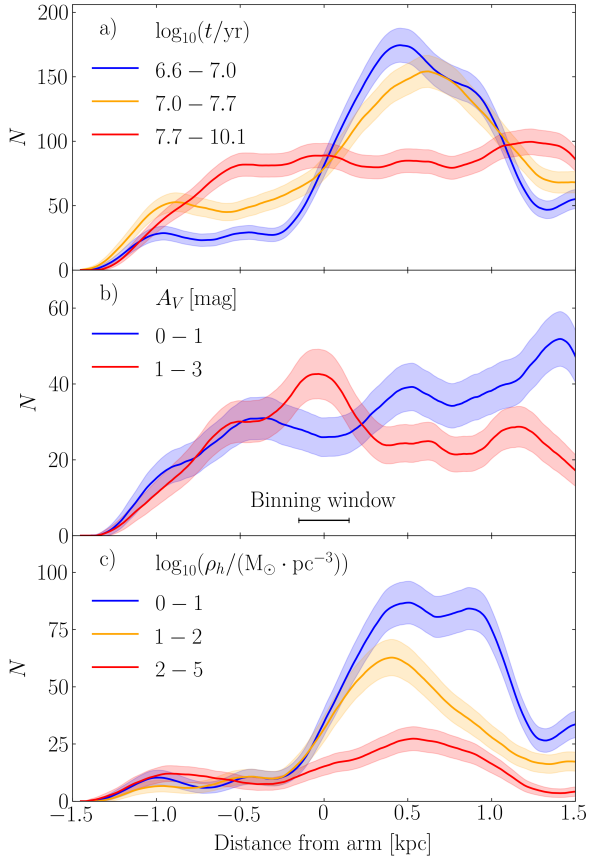

Figure 4.7. Counts of the detected cluster candidates as a function of their distance from the spiral arms of the galaxy with: a) 3 different cuts for age, b) 2 cuts for extinction for clusters with $\log_{10}(t/\mathrm{yr}) > 8$, and c) 3 cuts for density for clusters with $\log_{10}(t/\mathrm{yr}) < 7$. The binning window line on the bottom of panel b represents the bin width used to calculate the cluster count at each distance point. Negative distances indicate clusters trailing the spiral arms, while positive distances indicate clusters leaving and outpacing the arms. Poisson error bars are visualized as shaded areas around each curve.

to objects which are located in sparse areas of the arms and are, therefore, less obscured by dust.

Fig. 4.7c shows the densities of the population of younger clusters, with ages of $\log_{10}(t/\mathrm{yr}) < 7$ (corresponding to the blue line of panel a). This younger age cut was chosen to get the densities of clusters that were formed within the nearby spiral arm instead of being in the field population. The blue line indicates densities of $\log_{10}(\rho_h/(\mathrm{M}_\odot \cdot \mathrm{pc}^{-3})) < 1$, the orange line – $1 \leq \log_{10}(\rho_h/(\mathrm{M}_\odot \cdot \mathrm{pc}^{-3})) < 2$, and the red line – $\log_{10}(\rho_h/(\mathrm{M}_\odot \cdot \mathrm{pc}^{-3})) \geq 2$. There is a vague trend of higher density cluster counts peaking further away from the spiral arms than the lower density objects.

To verify that these trends did not arise due to selection effects, we took the mock clusters that were inserted in a uniform grid on the whole galaxy image, recovered them with the CNN and sliced their inferred parameter ranges in the same way as was done for Fig. 4.7. This way both the counts and the relevant parameters of the clusters are distributed uniformly over the galaxy and should result in curves without the peaks seen for real clusters. For these recovered mock clusters the ends of the distributions (for low negative and high positive distances) do tend to have a lower number of clusters, due to boundary effects of the maximum possible distance from the spiral arms. For the area in which we draw our results (approximately between -0.2 and 1.4 kpc) the recovered mock cluster distributions are flat. For age in particular, the recovered mock young cluster distributions are very similar to the real old cluster distributions, which is exactly what we would expect due to the old real cluster population being relatively uniformly distributed in the galaxy. Selection effects should not have a measurable impact on the trends found for cluster candidates.

4.3.4 Distance from galaxy center

Fig. 4.8 shows cluster distances from the galactic center against their density, mass, and size. In Fig. 4.8a it can be seen that older clusters are more dense closer to the galactic center, while for young clusters this effect is less pronounced. As density is a function of mass and size, we show their distributions in Fig. 4.8b and c. Both young and old clusters are more massive near the galactic center, possibly due to the fact that lower mass clusters break apart more easily in a denser environment. However, the average mass differences between these two populations get smaller the further away they are from the center. It can also be seen in Fig. 4.8c, as there are fewer diffuse clusters near the center of the galaxy, with the effect being more pronounced

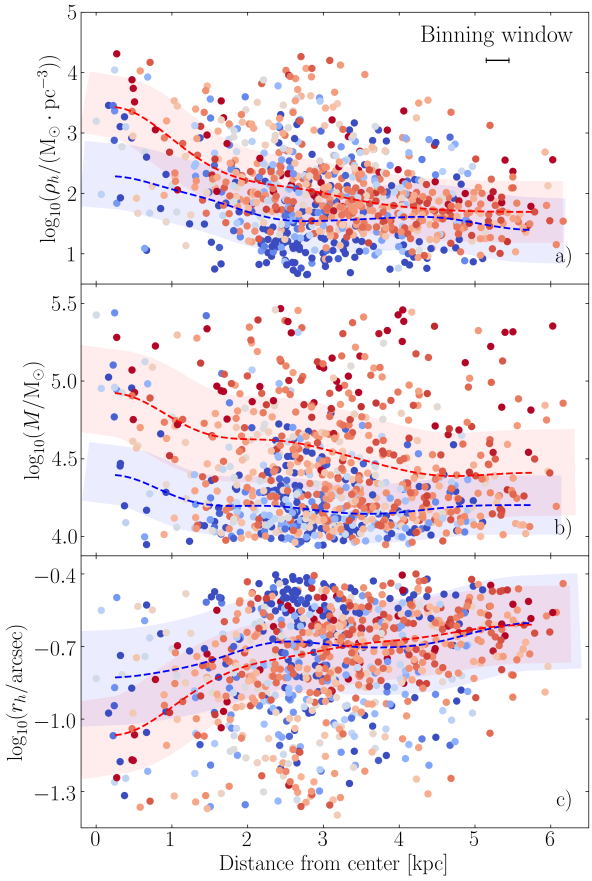

Figure 4.8. Densities (a), masses (b), and sizes (c) of the selected cluster candidates vs. their distance from the galactic center. Colors of the circles indicate ages, starting with $\log_{10}(t/\mathrm{yr}) \leq 7$ (blue) and up to $\log_{10}(t/\mathrm{yr}) \geq 9$ (red). The two dashed lines represent the mean parameter values for young ($\log_{10}(t/\mathrm{yr}) < 7.5$) and old ($\log_{10}(t/\mathrm{yr}) \geq 8.5$) cluster candidates, with the shaded area around the dashed lines representing the standard deviation of each parameter. The binning window line in the top right of panel a represents the width of the window used to calculate the means at each distance point.

for the older population. This is in agreement with Freeman et al. (2017), who have found that the surface density of molecular clouds varies with distance from the galactic center and shown that the mass of the most massive cluster decreases with distance.

We have also repeated the mock cluster test described in Section 4.3.3 to check whether selection effects have an impact on the results presented in this section as well and found no systematic biases.

4.3.5 Spatial distribution

Fig. 4.9 shows the spatial distribution of selected cluster candidates. The dust-band-traced spiral arms used for the analysis in Fig. 4.7 are displayed as black curves. In Fig. 4.9a a fraction of the selected cluster candidates are displayed with colors corresponding to their distance from the spiral arms, up to ± 1.5 kpc. The color blue indicates clusters trailing the arm, seen as negative distances in Fig. 4.7, and the color red indicates clusters leading the arm.

Fig. 4.9b shows the ages of the cluster candidates. A higher concentration of young clusters on the spiral arms and other star forming regions is observed, while older clusters are more evenly spread out in-between the arms and in the outer regions of the galaxy. Kim et al. (2012) analyzed the distributions of ages of resolved stars and found groupings of young stars along a spiral arm and more evenly distributed old stars. This closely corresponds to the distributions of our cluster candidate sample in the same region.

Fig. 4.9c shows the extinction of clusters more massive than $\log_{10}(M/\mathrm{M}_{\odot}) > 4$. There are more clusters with high extinction located on the spiral arms and other dust-rich areas of the galaxy. This can also be seen as a function of distance to the arm in Fig. 4.7b.

Fig. 4.9d shows the densities of the same higher-mass cluster sample as in Fig. 4.9c. Higher density clusters concentrate around the spiral arms of the galaxy and closer to its central region in general, as is shown in Figs. 4.7c and Fig. 4.8a and could be due to less dense clusters being broken up more easily in these crowded environments.

CNN results allowed us to identify populations of clusters with different properties on and within the spiral arms, as well as get consistent results of the age gradient of clusters within the corotation radius of M83. Such data could be used to test and constrain various hypotheses of cluster formation and disruption by comparing it to models of galaxy evolution.

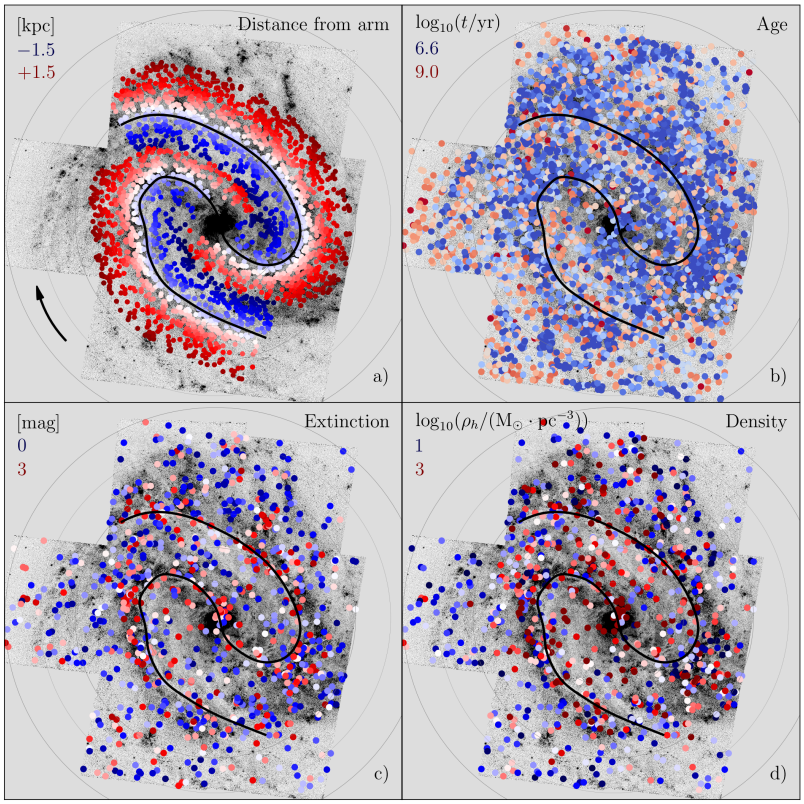

Figure 4.9. Spatial distribution of selected cluster candidates displayed on the F438W band. Panel a) shows the distance of cluster candidates from their nearest spiral arm as an increasing intensity of red for the leading arm areas, and blue for the trailing arm areas. Clusters are drawn up to the distance of 1.5 kpc. Panel b) shows the ages of clusters starting with $\log_{10}(t/\mathrm{yr}) = 6.6$ (blue) and up to $\log_{10}(t/\mathrm{yr}) \geq 9$ (red). Panel c) shows extinctions of $\log_{10}(M/\mathrm{M}_{\odot}) > 4$ cluster candidates, with the color map showing clusters starting at $A_V = 0$ mag (blue) up to $A_V = 3$ mag (red). Panel d) shows the same sample of clusters as panel c), only with the color coding corresponding to density values starting with $\log_{10}(\rho_h/(\mathrm{M}_{\odot} \cdot \mathrm{pc}^{-3})) < 1$ (blue) and going up to $\log_{10}(\rho_h/(\mathrm{M}_{\odot} \cdot \mathrm{pc}^{-3})) \geq 3$ (red). The faint concentric circles indicate distances from the galactic center, starting with 1 kpc and up to 6 kpc, with a step of 1 kpc. The arrow in the bottom left of panel a indicates the direction of rotation; the galactic disc is inclined with the NW part being closer to observer.

4.4 Discussion

Here we discuss the limitations of the method, as well as some of the caveats of our results and other relevant observations. The items are presented in no particular order.

- **Detection threshold.** The $class_{c/b}$ and $visibility$ parameter thresholds can be varied freely to select a different number of cluster candidates, as can be seen in Fig. 4.2. We repeated our experiments using cluster candidate samples obtained with different thresholds and observed no notable differences in the results. The distributions seen in Figs. 4.6–4.9 remain consistent with the presented analysis.

- **Parameter ranges.** They set the limits within which the CNN can produce predictions. Of particular note is cluster mass, which was constrained to $3.5 \leq \log(M/\mathrm{M}_{\odot}) \leq 5.5$, to cover the majority of M83 disc cluster population and is consistent with the mass ranges used by Shabani et al. (2018).

- **Mosaic area.** Our sample is effectively limited to the area within approximately 6 kpc from the center of the galaxy by the 7 WFC3 fields. However, this spatial extent is sufficient for our analysis and is consistent with the three galaxy coverage used by Shabani et al. (2018) for a similar study.

- **Possible interaction.** A deep photographic image of M83 has revealed "an enormous loop around NW quadrant of the galaxy" (Malin & Hadley 1997). Also, radio observations of its outskirts show a significant extent of a disturbed nature, suggesting a possible retrograde interaction (Heald et al. 2016). This is possibly due to a closest approach 1–2 Gyr ago of a neighbouring galaxy (Finlay 2014). In addition, by tracing the dust lanes as displayed in Fig. 4.9 we observed that the north-western spiral arm is slightly closer to the center than the south-eastern arm. This, in addition to the galaxy's large estimated inclination angle of $\sim$40 deg and its uncertainty (Heald et al. 2016), compelled us to measure cluster distances perpendicularly to the dust lanes of the arms and not along the direction of rotation as was done in Shabani et al. (2018).

- **Detection reliability.** Wright et al. (2017) have demonstrated that in combination human and CNN-based machine classification of transient

objects outperforms either one individually; humans tend to have lower false positive rates, but higher missed detection rates, while machine classification has the opposite trend. Therefore, we visually inspected the CNN-detected object candidates, removed possible false positives from the sample and counted objects potentially missed by the CNN. We estimated the false negative rate to be consistent with our artificial cluster tests and those of previous studies of limited samples by Bastian et al. (2011) and Ryon et al. (2015). Visual inspection also revealed that likely background galaxies (red extended objects) are not detected by the CNN as clusters. Note, Silva-Villa et al. (2014) compiled ∼1,800 clusters and associations of $\log_{10}(t/\mathrm{yr}) \lesssim 8.5$ and $\log(M/\mathrm{M}_{\odot}) \gtrsim 3.7$. We find ∼1,600 cluster candidates within these age and mass ranges.

- **Sample selection.** From the experience in doing visual selection of star clusters in the Andromeda galaxy (Narbutis et al. 2008), based on Subaru telescope observations, which effectively give a similar resolution as HST does in M83, it was noted, that higher selection thresholds are applied by humans to avoid including too many faint clusters. This results in a limited sample of low-mass clusters deviating from Schechter mass functions as shown in Vansevičius et al. (2009). A similar effect is seen in the catalog of Bastian et al. (2011) as shown in of Chapter 3. Examples of mock clusters reveal that such low visibility objects might be easily rejected from a visually compiled sample, but could be genuine star clusters.

- **Young low-mass candidates.** Comparing Fig. 4.5a with Fig. 3.14a, the high number of young low-mass cluster candidates of this work does not significantly overlap with random background inference results, which are concentrated at $\log_{10}(t/\mathrm{yr}) > 7.5$ and $\log(M/\mathrm{M}_{\odot}) < 3.75$. Even though in Fig. 3.14e a high number of backgrounds are identified as low-mass extended objects, they are not concentrated around young ages $\log_{10}(t/\mathrm{yr}) < 7$, while our cluster sample exhibiting the same properties in Fig. 4.5e are younger objects. In addition, in Fig. 3.14b the backgrounds are spread out diagonally in an area where one would expect the age-extinction degeneracy to take place, while in Fig. 4.5b the overlap with this area is minor. Krumholz et al. (2019) caution that all observations are prone to selection biases in favor of more compact clusters, as well as a survival bias. Clusters of age ∼10 Myr might be only weakly

bound, but these would be under-represented in cluster catalogs, because almost none would survive to reach ages more than a few tens of millions of years. Our CNN, however, detects numerous young extended objects.

4.5 Conclusions

We have presented a full pipeline of cluster search using a CNN, which is able to infer cluster parameters (age, mass, extinctions, and size), as well as provide inference reliability estimates. The CNN is trained on mock star clusters superimposed on images of real backgrounds and is applied to multi-band images as a sliding window, without using precise photometric flux calibrations. This produces object candidate detection certainty maps, which are then filtered to obtain any number of cluster candidates, depending on recall (completeness) tolerance, estimated by artificial cluster tests.

We applied the star cluster detection pipeline to collect 3,380 cluster candidates in HST WFC3 observations of M83 in F336W, F438W, and F814W passbands, and used this data to analyze the spatial distributions of cluster parameters (age, mass, extinction, size) w.r.t. the galaxy's spiral arms and its center.

We have shown that an age gradient w.r.t. the spiral arms can be observed inside the corotation radius of the galaxy. The younger population of cluster candidates peaks at $\sim$0.4 kpc leading the spiral arms, while the older population is shifted towards $\gtrsim$0.7 kpc, this finding being consistent with the density wave theory of spiral arms. We also find a greater proportion of high extinction clusters on the dust lanes of the spiral arms, as well as more dense older clusters towards the galactic center.

Conclusions of thesis

In Chapter 2 we have proposed a CNN based on the ResNet architecture for simultaneous derivation of evolutionary and structural parameters of star clusters from imaging data.

The proposed CNN was verified on mock images of artificial clusters in the context of HST observations of the M31 galaxy. It has demonstrated a high accuracy and no significant bias for semi-resolved clusters with ages between $\log_{10}(t/\mathrm{yr}) = 7.0$ and 9.0, masses between $\log_{10}(M/\mathrm{M}_{\odot}) = 2.4$ and 3.6, and sizes between $\log_{10}(r_h/\mathrm{arcsec}) = -0.8$ and 0.0.

We have shown with artificial tests that CNNs can perform both structural and evolutionary star cluster parameter derivation directly from raw imaging data. This allows dealing with both unresolved and semi-resolved cases homogeneously, as well as utilizing multiple photometric passbands in an integrated manner.

In Chapter 3 we have extended the method introduced in Chapter 2 to infer cluster ages, masses, sizes, and extinctions in tandem, as well as to account for the degeneracies between them. Additional parameters were added for identifying the presence of clusters on background images of M83 and judging their visibility (signal-to-noise ratio).

To train this network a bank of mock clusters was generated utilizing three photometric passbands in the context of HST observations of the M83 galaxy. The CNN was verified on mock images of artificial clusters with ages, $\log_{10}(t/\mathrm{yr})$, between 6.6 and 10.1, masses, $\log_{10}(M/\mathrm{M}_{\odot})$, between 3.5 and 5.5, sizes between 0.04 and 0.4 arcsec, and extinctions $A_V \leq 3$ mag. Parameters derived by the CNN have shown a good agreement with the true parameters for clusters with ages $\log_{10}(t/\mathrm{yr}) < 8$, with higher age estimates being unreliable due to the age-extinction degeneracy.

Real cluster parameter inference tests were performed with three different M83 cluster catalogs from Bastian et al. (2011), Ryon et al. (2015), and Harris

et al. (2001) and have shown consistent results.

We have demonstrated that a CNN can perform evolutionary (age, mass), structural (size), and environmental (extinction) star cluster parameter inference. In addition, the network is capable of giving an indication of cluster presence in images.

In Chapter 4 we have presented the full pipeline for performing star cluster candidate detection and their parameter inference.

We applied the star cluster candidate detection pipeline to collect 3,380 cluster candidates in HST WFC3 observations of M83 in F336W, F438W, and F814W passbands, and used this data to analyze the spatial distributions of cluster parameters (age, mass, extinction, size) with respect to the galaxy's spiral arms and its center. It was found that an age gradient with respect to the spiral arms can be observed inside the corotation radius of the galaxy. More dense cluster candidates were also found towards the galactic center.

The CNN-based approach naturally joins various techniques used in the study of star clusters – detection, inference of evolutionary parameters, and size estimation – into a single pipeline, which gives a coherent view of cluster populations.

Bibliography

Andrews, J. E., Calzetti, D., Chandar, R., et al. 2014, ApJ, 793, 4

Avila, R. J. 2017, Advanced Camera for Surveys Instrument Handbook for Cycle 25 v. 16.0

Bastian, N., Adamo, A., Gieles, M., et al. 2011, MNRAS, 417, L6

Bastian, N., Adamo, A., Gieles, M., et al. 2012, MNRAS, 419, 2606

Baugh, C. 2001, The Encyclopedia of Astronomy and Astrophysics (IOP Publishing Ltd)

Beerer, I. M., Koenig, X. P., Hora, J. L., et al. 2010, The Astrophysical Journal, 720, 679

Blair, W. P., Chandar, R., Dopita, M. A., et al. 2014, ApJ, 788, 55

Borissova, J., Ivanov, V. D., Minniti, D., Geisler, D., & Stephens, A. W. 2005, Astronomy & Astrophysics, 435, 95

Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127

Bridžius, A., Narbutis, D., Stonkutė, R., Deveikis, V., & Vansevičius, V. 2008, Baltic Astronomy, 17, 337

Caballero, J. A. & Dinis, L. 2008, Astronomische Nachrichten, 329, 801

Cabrera-Cano, J. & Alfaro, E. J. 1990, Astronomy and Astrophysics (ISSN 0004-6361), 235, 94

Cabrera-Vives, G., Reyes, I., Förster, F., Estévez, P. A., & Maureira, J.-C. 2017, ApJ, 836, 97

Caldwell, N., Harding, P., Morrison, H., et al. 2009, AJ, 137, 94

Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245

Castro-Ginard, A., Jordi, C., Luri, X., et al. 2018, A&A, 618, A59

Dalcanton, J. J., Williams, B. F., Lang, D., et al. 2012, ApJS, 200, 18

de Meulenaer, P., Narbutis, D., Mineikis, T., & Vansevičius, V. 2013, A&A, 550, A20

de Meulenaer, P., Narbutis, D., Mineikis, T., & Vansevičius, V. 2014, A&A, 569, A4

de Meulenaer, P., Narbutis, D., Mineikis, T., & Vansevičius, V. 2015a, A&A, 574, A66

de Meulenaer, P., Narbutis, D., Mineikis, T., & Vansevičius, V. 2015b, A&A, 581, A111

de Meulenaer, P., Stonkutė, R., & Vansevičius, V. 2017, A&A, 602, A112

Deveikis, V., Narbutis, D., Stonkute, R., Bridzius, A., & Vansevicius, V. 2008, Baltic Astronomy, 17, 351

Dieleman, S., Willett, K. W., & Dambre, J. 2015, MNRAS, 450, 1441

Domínguez Sánchez, H., Huertas-Company, M., Bernardi, M., Tuccillo, D., & Fischer, J. L. 2018, MNRAS, 476, 3661

Dopita, M. A., Blair, W. P., Long, K. S., et al. 2010, ApJ, 710, 964

Dressel, L. 2012, Wide Field Camera 3 Instrument Handbook for Cycle 21 v. 5.0

Elson, R. A. W., Fall, S. M., & Freeman, K. C. 1987, Astrophysical Journal, 323, 54

Elson, R. A. W., Fall, S. M., & Freeman, K. C. 1987, ApJ, 323, 54

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. 1996, in (AAAI Press), 226–231

Finlay, W. H. 2014, Concise Catalog of Deep-Sky Objects (Springer International Publishing Switzerland)

Flamary, R. 2016, ArXiv e-prints [arXiv:1612.04526]

Fouesneau, M., Johnson, L. C., Weisz, D. R., et al. 2014, ApJ, 786, 117

Fouesneau, M. & Lançon, A. 2010, A&A, 521, A22

Fouesneau, M., Lançon, A., Chandar, R., & Whitmore, B. C. 2012, ApJ, 750, 60

Freeman, P., Rosolowsky, E., Kruijssen, J. M. D., Bastian, N., & Adamo, A. 2017, MNRAS, 468, 1769

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2016, A&A, 595, A2

Girshick, R. 2015, in 2015 IEEE International Conference on Computer Vision (ICCV), 1440

Harris, J., Calzetti, D., III, J. S. G., Conselice, C. J., & Smith, D. A. 2001, The Astronomical Journal, 122, 3046

Hausen, R. & Robertson, B. E. 2020, ApJS, 248, 20

Haykin, S. S. 2009, Neural networks and learning machines, 3rd edn. (Upper Saddle River, NJ: Pearson Education)

He, K., Zhang, X., Ren, S., & Sun, J. 2016, in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770

Heald, G., de Blok, W. J. G., Lucero, D., et al. 2016, MNRAS, 462, 1238

Herbel, J., Kacprzak, T., Amara, A., Refregier, A., & Lucchi, A. 2018, ArXiv e-prints [arXiv:1801.07615]

Hernandez, S., Larsen, S., Aloisi, A., et al. 2019, ApJ, 872, 116

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. 2012, ArXiv e-prints [arXiv:1207.0580]

Hirota, A., Kuno, N., Baba, J., et al. 2014, PASJ, 66, 46

Huertas-Company, M., Primack, J. R., Dekel, A., et al. 2018, The Astrophysical Journal, 858, 114

Johnson, L. C., Seth, A. C., Dalcanton, J. J., et al. 2017, ApJ, 839, 78

Johnson, L. C., Seth, A. C., Dalcanton, J. J., et al. 2012, ApJ, 752, 95

Johnson, L. C., Seth, A. C., Dalcanton, J. J., et al. 2015, ApJ, 802, 127

Karampelas, a., Dapergolas, A., Kontizas, E., et al. 2009, Astronomy & Astrophysics, 497, 703

Kharchenko, N. V., Piskunov, A. E., Röser, S., Schilbach, E., & Scholz, R. D. 2005, Astronomy & Astrophysics, astro-ph, 403

Kharchenko, N. V., Piskunov, A. E., Schilbach, E., Röser, S., & Scholz, R.-D. 2012, Astronomy & Astrophysics, 543, A156

Kharchenko, N. V., Piskunov, A. E., Schilbach, E., Röser, S., & Scholz, R.-D. 2013, Astronomy & Astrophysics, 558, A53

Kim, H., Whitmore, B. C., Chandar, R., et al. 2012, ApJ, 753, 26

King, I. 1962, Astronomical Journal, 67, 471

Kingma, D. P. & Ba, J. 2014, ArXiv e-prints [arXiv:1412.6980]

Kirsanova, M. S., Sobolev, A. M., Thomasson, M., et al. 2008, Monthly Notices of the Royal Astronomical Society, 388, 729

Krist, J. E., Hook, R. N., & Stoehr, F. 2011, in Proc. SPIE, Vol. 8127, Optical Modeling and Performance Predictions V, 81270J

Kroupa, P. 2001, MNRAS, 322, 231

Kruger, N., Janssen, P., Kalkan, S., et al. 2013, IEEE Trans. Pattern Anal. Mach. Intell., 35, 1847

Krumholz, M. R., Fumagalli, M., da Silva, R. L., Rendahl, T., & Parra, J. 2015, MNRAS, 452, 1447

Krumholz, M. R., McKee, C. F., & Bland -Hawthorn, J. 2019, ARA&A, 57, 227

Kumar, M. S. N., Keto, E., & Clerkin, E. 2006, Astronomy & Astrophysics, 449, 1033

Lada, C. J. & Lada, E. a. 2003, Annual Review of Astronomy and Astrophysics, 73

Lanusse, F., Ma, Q., Li, N., et al. 2018, MNRAS, 473, 3895

Lin, T.-Y., Maire, M., Belongie, S., et al. 2014, in Computer Vision – ECCV 2014, ed. D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Cham: Springer International Publishing), 740

Malin, D. & Hadley, B. 1997, PASA, 14, 52

Maschberger, T., Clarke, C. J., Bonnell, I. A., & Kroupa, P. 2010, Monthly Notices of the Royal Astronomical Society, 404, 1061

McConnachie, A. W., Irwin, M. J., Ferguson, A. M. N., et al. 2005, MNRAS, 356, 979

Mercer, E. P., Clemens, D. P., Meade, M. R., et al. 2005, The Astrophysical Journal, 635, 560

Michalik, D., Lindegren, L., & Hobbs, D. 2015, A&A, 574, A115

Miller, R., Kennefick, D., Kennefick, J., et al. 2019, ApJ, 874, 177

Mok, A., Chandar, R., & Fall, S. M. 2020, ApJ, 893, 135

Nair, V. & Hinton, G. E. 2010, in Proceedings of the 27th International Conference on Machine Learning, ICML'10, 807–814

Narbutis, D., Bridžius, A., & Semionov, D. 2015, Baltic Astronomy, 24, 305

Narbutis, D., Semionov, D., Stonkutė, R., et al. 2014, A&A, 569, A30

Narbutis, D., Vansevičius, V., Kodaira, K., Bridžius, A., & Stonkutė, R. 2008, ApJS, 177, 174

Nikiforov, I. & Smirnova, O. 2013, Astronomische Nachrichten, 334, 749

Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2017, Monthly Notices of the Royal Astronomical Society, 472, 1129

Piskunov, A. E., Schilbach, E., Kharchenko, N. V., Roser, S., & Scholz, R. D. 2008, Astronomy & Astrophysics, 477, 165

Portegies Zwart, S. F., McMillan, S. L., & Gieles, M. 2010a, Annual Review of Astronomy and Astrophysics, 48, 431

Portegies Zwart, S. F., McMillan, S. L. W., & Gieles, M. 2010b, ARA&A, 48, 431

Pourrahmani, M., Nayyeri, H., & Cooray, A. 2018, ApJ, 856, 68

Regier, J., Miller, A., McAuliffe, J., et al. 2015, in 32nd International Conference on Machine Learning

Ren, S., He, K., Girshick, R., & Sun, J. 2017, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39, 1137

Roeser, S., Demleitner, M., & Schilbach, E. 2010, The Astronomical Journal, 139, 2440

Rowe, B. T. P., Jarvis, M., Mandelbaum, R., et al. 2015, Astronomy and Computing, 10, 121

Russakovsky, O., Deng, J., Su, H., et al. 2015, International Journal of Computer Vision, 115, 211

Ryon, J. E., Bastian, N., Adamo, A., et al. 2015, MNRAS, 452, 525

Sanchez, N. & Alfaro, E. J. 2009, The Astrophysical Journal, 696, 2086

Sánchez, N., Alfaro, E. J., & López-Martínez, F. 2020, MNRAS, 495, 2882

Sánchez-Gil, M. C., Alfaro, E. J., Cerviño, M., et al. 2019, MNRAS, 483, 2641

Sanders, W. L. 1971, Astronomy & Astrophysics, 14, 226

Schmeja, S. 2011, Astronomische Nachrichten, 332, 172

Schmeja, S., Gouliermis, D. A., & Klessen, R. 2009, Astronomy & Astrophysics, 694, 367

Scholz, R.-D., Kharchenko, N. V., Piskunov, A. E., Röser, S., & Schilbach, E. 2015, Astronomy & Astrophysics, 581

Sedaghat, N. & Mahabal, A. 2018, MNRAS, 476, 5365

Shabani, F., Grebel, E. K., Pasquali, A., et al. 2018, MNRAS, 478, 3590

Shallue, C. J. & Vanderburg, A. 2018, AJ, 155, 94

Shu, F. H. 2016, ARA&A, 54, 667

Silva-Villa, E., Adamo, A., Bastian, N., Fouesneau, M., & Zackrisson, E. 2014, MNRAS, 440, L116

Sofue, Y. 2018, PASJ, 70, 106

Solin, O., Haikala, L., & Ukkonen, E. 2014, Astronomy & Astrophysics, 562, A115

Solin, O., Ukkonen, E., & Haikala, L. 2012, Astronomy & Astrophysics, 542, 3

Thim, F., Tammann, G. A., Saha, A., et al. 2003, ApJ, 590, 256

van Breukelen, C., Clewley, L., Bonfield, D. G., et al. 2006, Galaxy clusters at $0.6 < z < 1.4$ in the UKIDSS Ultra Deep Survey Early Data Release

Vansevičius, V., Kodaira, K., Narbutis, D., et al. 2009, ApJ, 703, 1872

Vasilevskis, S., Klemola, A., & Preston, G. 1958, Astronomical Journal, 63, 387

Whitmore, B. C., Chandar, R., Kim, H., et al. 2011, ApJ, 729, 78

Wilkinson, S., Merín, B., & Riviere-Marichalar, P. 2018, A&A, 618, A12

Wright, D. E., Lintott, C. J., Smartt, S. J., et al. 2017, MNRAS, 472, 1315

Xin-hua, G., Li, C., & Zhen-jie, H. 2014, Chinese Astronomy and Astrophysics, 38, 257

Zimmer, P., Rand, R. J., & McGraw, J. T. 2004, ApJ, 607, 285

Appendices

Appendix A

Mock cluster samples

This section contains examples of mock clusters generated in F275W, F336W, F475W, F814W, F110W, and F160W passbands of the HST/WFC3 instrument, which match the filters available in the PHAT data. The mock clusters are displayed as false color images, with Fig. A.1 covering passbands F275W, F336W, and F475W, Fig. A.2 – passbands F336W, F475W, and F814W, Fig. A.3 – passbands F475W, F814W, and F110W, finally Fig. A.4 – passbands F814W, F110W, and F160W.

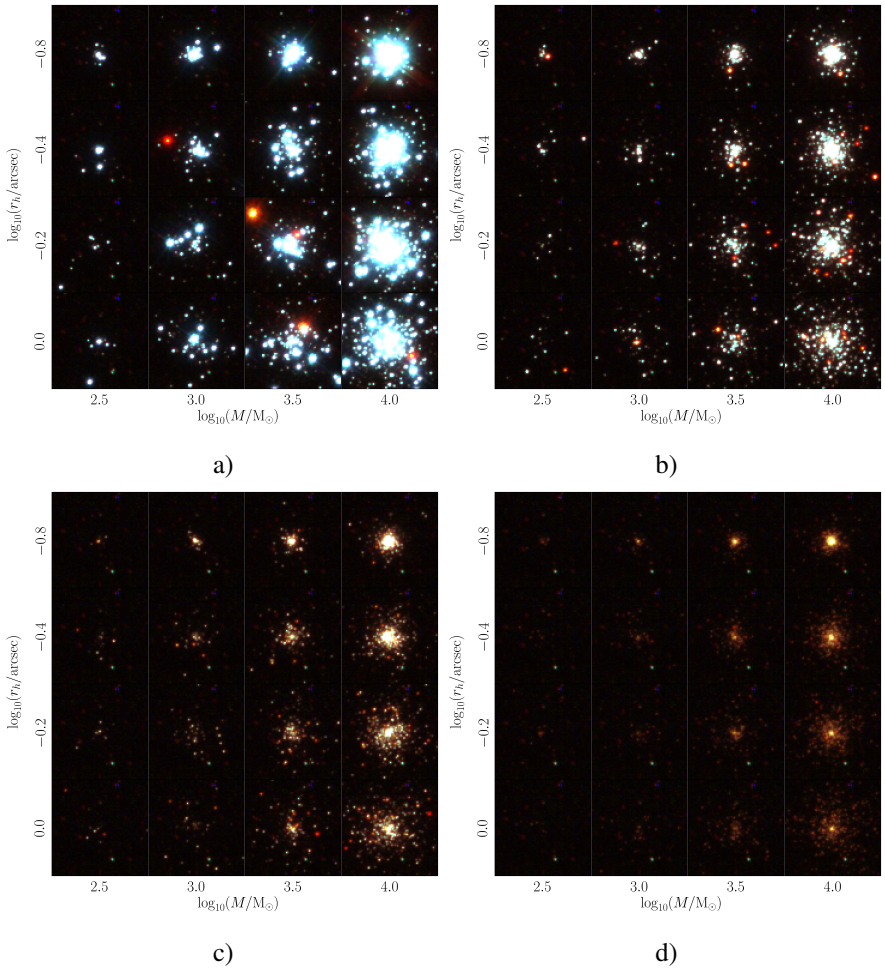

Figure A.1. Examples of generated clusters, without extinction, on a real background image, shown in passbands F275W, F336W, and F475W. The ages of all of the displayed clusters are: a) $\log(t/\mathrm{yr}) = 7.0$, b) 8.0, c) 8.5, d) 9.0. The masses and r_h values are varied as shown on the axes. The intensity scale of the images was normalized with the arcsinh function. The background is the same for all of the displayed clusters for clarity.

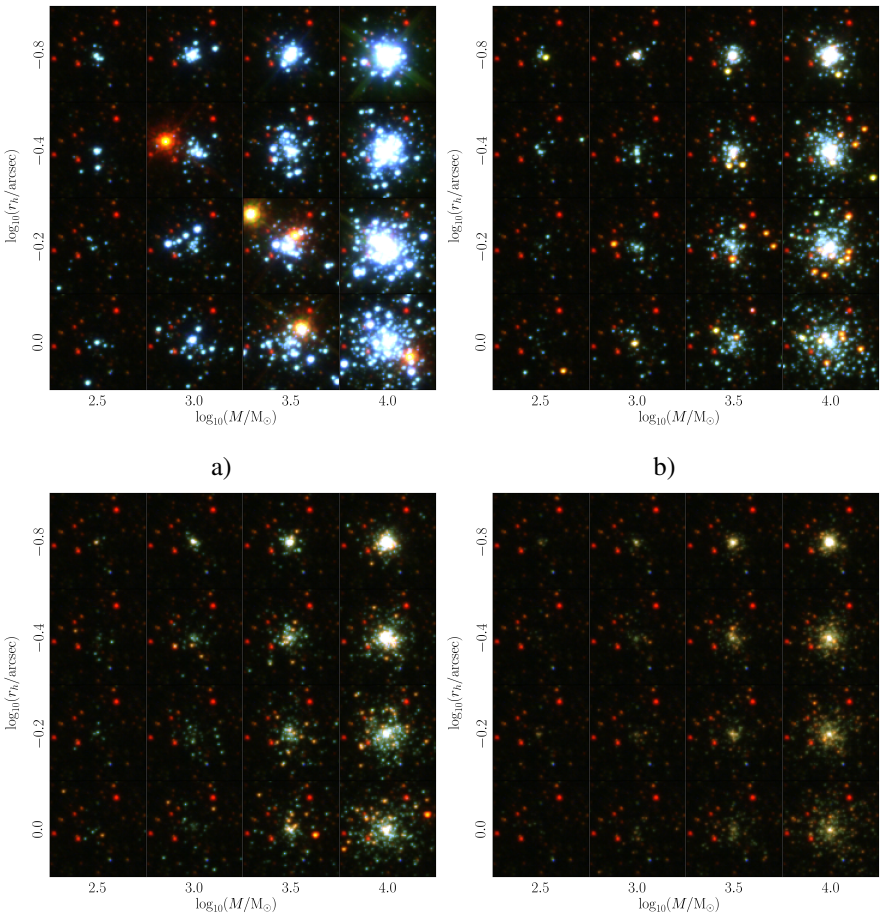

Figure A.2. Same as Fig. A.1, but shown in passbands F336W, F475W, and F814W.

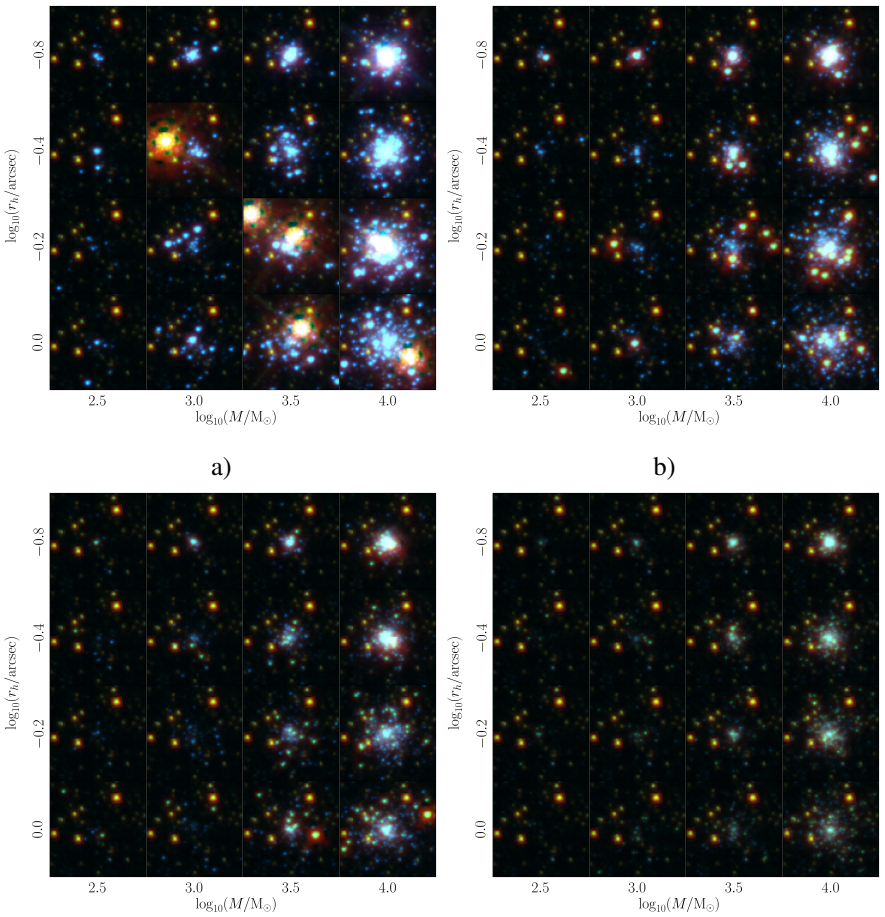

Figure A.3. Same as Fig. A.1, but shown in passbands F475W, F814W, and F110W.

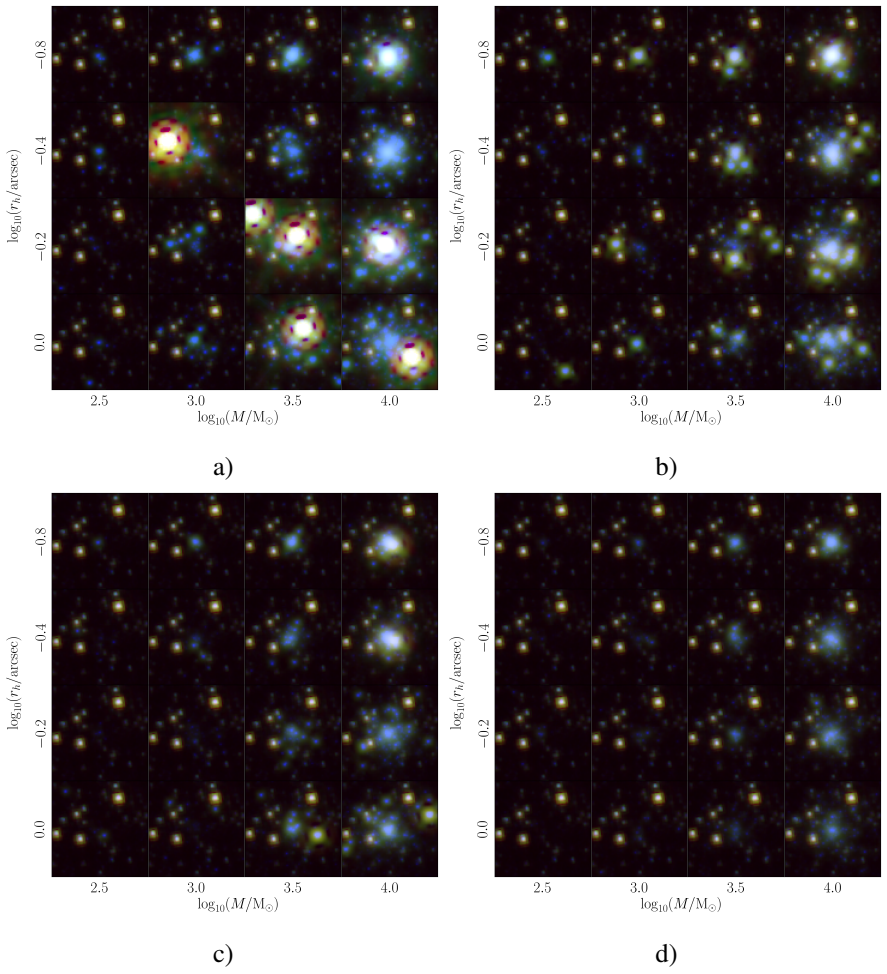

Figure A.4. Same as Fig. A.1, but shown in passbands F814W, F110W, and F160W.

NOTES

Vilnius University Press
9 Saulėtekio Ave., Building III, LT-10222 Vilnius
Email: info@leidykla.vu.lt, www.leidykla.vu.lt
Print run copies 16