university of groningen

faculty of science and engineering

The influence of Galactic morphology on Dark Matter detecting neural networks

Author:
Lorenzo DE CESARE
(s4711025)

Supervisor:
asst. prof. Jelle AALBERS
Second examiner :
dr. Emanuela DIMASTROGIOVANNI

Bachelor's Thesis
To fulfill the requirements for the degree of
Bachelor of Science in Physics
at the University of Groningen

July 9, 2024

Contents

Abstract

Determining the distribution dark matter structures, allows for the cosmological model of our universe to be refined, testing the widely accepted Λ cold dark matter model. Specifically, studying the distribution of small dark matter 'clumps', also known as sub-halos, in strongly gravitationally lensed systems seems to be an incredibly promising marker. Recent studies on this topic have focused on combining simulated lensed images with neural networks, in order to test the potential of the network in inferring the sub-halo mass distribution in these strongly lensed system. One of the most recent such studies is that by Wagner-Cerena et al [1], the study focuses on the development of the neural network and simulation methods. In this thesis, the dependence of this neural network on galaxy morphology is tested, in order to verify the reliability of the neural network. Through testing of the network performance with various morphological markers, it is established that galaxy morphology does not have any effect on the neural network developed by Wagner-Cerena et al [1].

1 Introduction

Cosmology is one of the most fundamental subjects in all of physics, dictating exactly how our universe evolved into what it currently is. The basis of cosmological theory serves as the basis for all of modern theoretical physics, as it details the manner in which the fundamental laws of nature became what they are today. One of the main components in cosmological theories is dark matter, a mysterious type of matter which continues to elude the direct detection of modern physicists.

Various theories on the nature of this mysterious dark matter exist, and each one of them characterise a different cosmological model, as the nature of this dark matter is crucial in the description of our universe. For now, the most prominent theory of dark matter is the Cold dark matter (CDM) model. This model gives rise to the most prominent model of cosmology, the ΛCDM model. The ΛCDM model or concordance model, is incredibly well established, however, more evidence is required to properly narrow down whether the CDM model should continue to serve as the basis of modern cosmological research [2][3]. Evidence for cosmological theories is unfortunately, quite rare, as the times and scales in which the predictions of cosmological theories become significant are often too large to be able to test [3].

Fortunately, there are some pieces of evidence on scales which are observable, one of these which could allow to differentiate between the possible models of cosmology are low mass($< 10^{10} M_{\odot}$) dark matter halos, or in other words clumps of dark matter. The various models of cosmology predict different formation mechanisms for dark matter within the universe, and this leads to differences in the distributions of these dark matter clumps as shown in figure (1) [4][5][6]. Detecting these halos would allow a constraint to be put on the dark matter halo mass, restricting the possible cosmological models as has already been done in the past [7]. In order to be able to effectively utilise these low mass halos as an indicator, it is necessary to find a method of analysing them.

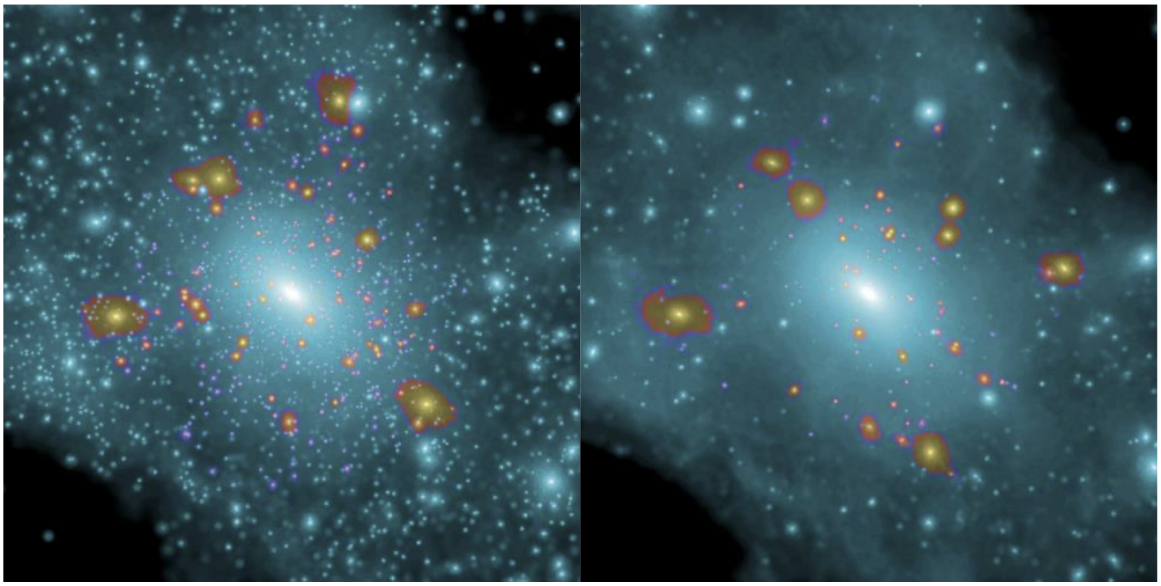

Figure 1: This image shows the simulated differences in halo formation in two different CDM and WDM models. These simulations specifically show the dark matter halos inside dwarf galaxies such as the Large Magellanic cloud. Image taken from [8]

Detection using light proves difficult, since it is not known whether dark matter emits electromagnetic radiation, and if it does then it is an incredibly small amount [9]. More precisely, methods involving direct detection by observing the light emitted rely on the absorption of light by baryons in the halo and being able to link the luminous emissions to the dark matter distribution. This process requires very intricately and precisely modeling baryonic physics, a step which creates large uncertainties[10] .

Due to the uncertainty of detection by direct emission via light, the most prominent method to detect these low mass halos is via strong gravitational lensing. Strong gravitational lensing is a phenomenon first predicted by Einstein from his theory of general relativity [11], the idea is that an incredibly massive object causes the light from an object behind it to be distorted. This phenomenon causes the observer to see a highly distorted version of the original image. The low mass($< 10^{10}M_{\odot}$) halos are observed as sub-halos of larger lenses. The presence of these sub-halos cause perturbations in the final lensed image, which can then be detected by analysing the image[1]. This method also gives a complete gravitational profile of the sub-halo, allowing for a precise determination of it's mass distribution. Furthermore, with the number of observable strong lenses growing, as more powerful telescopes become operational, the potential of this method in determining dark matter parameter constraints is evident [12] [13].
When studying strong lensing, the main parameter of interest is the mass distribution of the sub-halos, also known as the sub-halo mass function (SHMF) [1] [14]. The sub-halo mass function, as stated above is essentially a function describing the mass distribution of these smaller sub-halos. An example of a sub-halo mass function, and how different dark matter models can apply different constraints on these distributions is shown below:

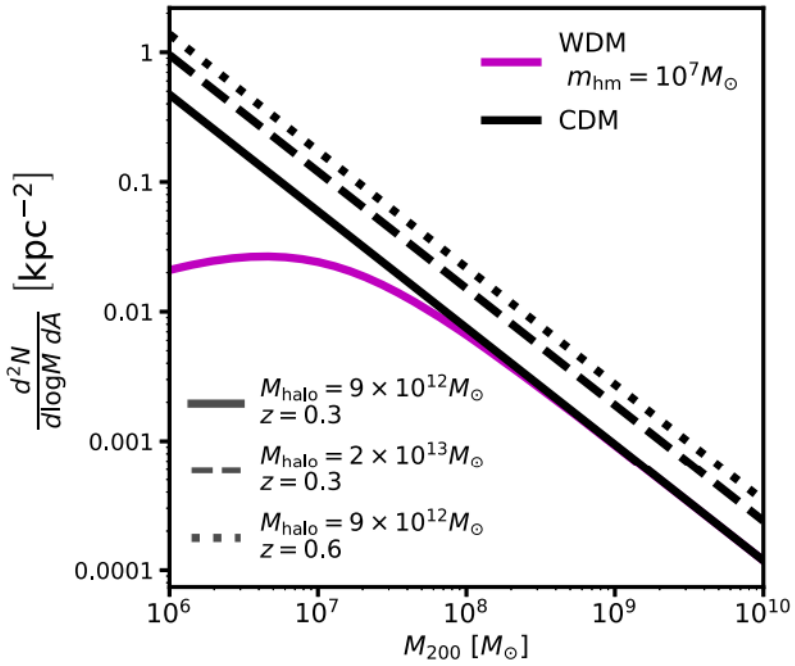

Figure 2: Diagram showing the sub-halo mass function (on the y-axis) as a function of halo mass, given in solar masses. Two different dark matter models are shown Warm dark matter (WDM) and cold dark matter (CDM). Three different redshift values are shown for the cold dark matter model as the time of formation effects the SHMF. The m_{hm} variable is called the half-mode mass, a parameter utilised to describe the interactions of dark matter. Image taken from Gilman et al ([15])

Constraining the value of the SHMF could be an incredibly important tool in being able to find deviations from the CDM and hence test the long standing concordance model of cosmology [5] [6]. There are two main approaches for determining the SHMF, direct detection and statistical detection [10] [1]. Direct detection consists of detecting the signal of the individual sub-halos in a strongly lensed image, returning quantities such as the amount, mass and position of the sub-halos in the strong lensing system [10] [16]. However, this method of detection is difficult to scale to more massive sub-halos. This scaling results in the model ignoring the thousands of sub-halos with masses of $10^7{}_\odot$-$10^9{}_\odot$. In addition, to be able to determine the SHMF from one single detection means using significant assumptions about the population of these sub-halos. These reasons are why more and more statistical detection is becoming the standard method of approaching strongly lensed systems [17] [1].

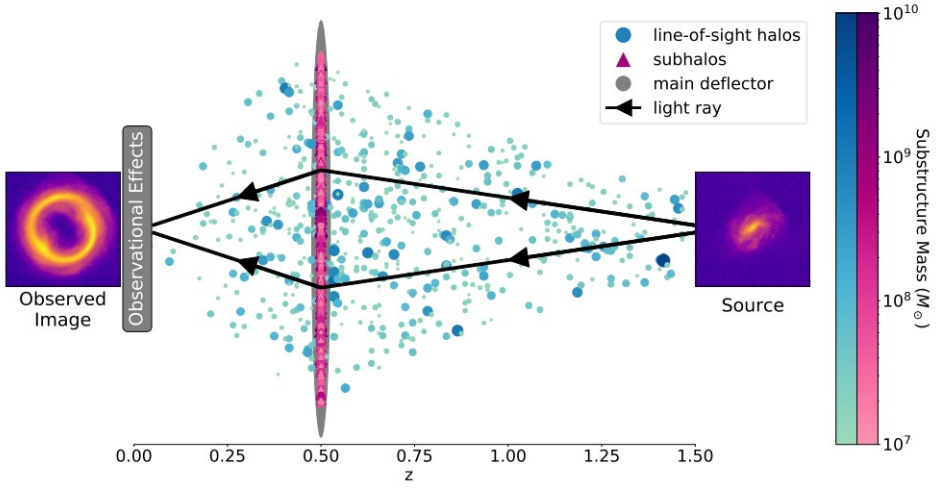

Figure 3: Diagram showing components of the strong gravitationally lensed systems analysed throughout this thesis. The image on the right is the source galaxy, the light from the galaxy is then distorted by the strong lens giving the final observed image. The scale on the bottom shows the redshift values as the image shows a highly compressed version of the actual physical process taking place. Image taken from [1]

Statistical detection is based on modelling the parameters of a population of sub-halos within the system rather than directly detecting the signal of each individual sub-halo. This method of modelling reduces the amount of free parameters in the system, reducing the overall computing power required. This allows for more massive sub-halos, and in general for more focus on parameters, such as the SHMF, which are of most interest for being able to determine the model of cosmology present [1][18].

Within the statistical detection framework, there are two main methods to model and subsequently analyse the dark matter sub-halo population. The most common method is the approximate Bayesian computation (ABC), in which data is generated by utilising prior probability on the parameter space, a more thorough description of the ABC method can be found in the papers originally responsible for developing this method [19][20] . However, this method has significant downsides, such as the fact that ABC inference cannot be scaled to large datasets. This means that each individual lens requires a significant number of simulations to yield meaningful results. Furthermore, it means that analysing image datasets using the ABC method, requires reducing the images hence discarding some of the data. Due to these problems, the main method used in recent years shifted towards utilising neural

networks as density estimators [21][22][23].

Neural network based density estimators utilise a neural network which has been trained on a specific training set, to then predict certain parameters based on the input data. The way in which the neural density estimators work is highly favorable, as they allow for a detection method which is not only completely scalable, but which also makes full use of the information in the given image. This method has been put to the test several times recently, most notably in the papers by [1] [24], these papers have shown just how effective this method can be in constraining the SHMF.

Specifically, in this thesis, the focus will be on the paper published Wagner-Cerena et al [1], which produced significant results, and showed how powerful this method can be in analysing strongly lensed systems. In early iterations of the neural network developed in this paper, it was noted that most of the worst performing galaxies seemed to be spirals. Exploring the motives behind this under-performance could prove crucial in determining the limits of the neural network developed so far. There are several possible explanations as to why certain galaxies perform better within the neural network than others. One of the most interesting explanations, and the one which will be investigated in this thesis is the morphological profile of the specific galaxy. More precisely, the expectation is that how smooth a galaxy is could have an impact on the predicting power of the neural network. This is hypothesised, as the high luminosity regions which characterise not smooth (or 'clumpy') galaxies, could lead to an effect where the chaotic morphology of the galaxy is construed with the effect of several sub-halos bending the light of the galaxy[1].

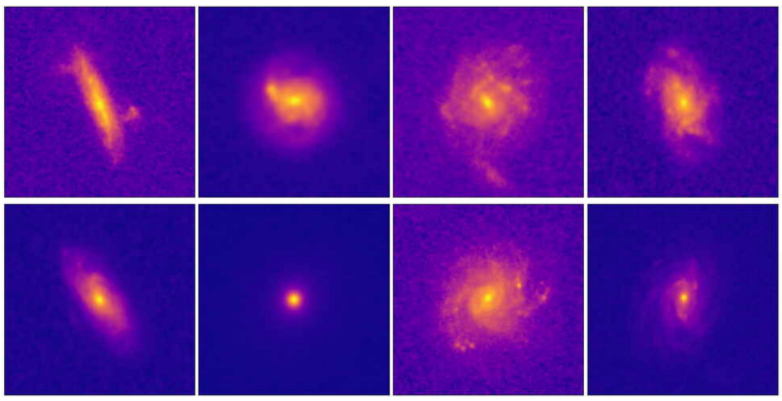

Figure 4: Mosaic of various of the galaxies analysed in this thesis. The images are plotted utilising flux per pixel values of the galaxies, the brighter areas show the areas with higher flux per pixel. The galaxy images are made utilising the *galsim* package, which contains images of the HST made using the advanced camera for surveys(ACS) in the F814W filter .The variety in the morphological structure of the galaxies is evident.

More specifically, the focus is put on two main parameters which are related to the 'clumpiness' of a galaxy, namely, the Sersic index (n) and the CAS parameter for clumpiness (denoted by S) [25][26]. In essence, the following questions will be the focus of this thesis: How does simulation based inference respond to different galaxy morphologies ? Does the Sersic index of a galaxy hinder the effectiveness of the neural density estimator method ? How does the neural network respond to galaxies of differing clumpiness ? To do this, the simulation methods and neural network utilised in the paper [1] will be used. Subsequently, a relation will be established between morphological parameters and the performance of the network.

2 Background Literature

2.1 Strong Gravitational lensing

The basis of this thesis is built upon the simulations carried out in the paper ([1]), and hence, the theory behind the methods utilised to simulate the lenses is the same as that explored in the paper. Both the simulation methods and the neural network used to analyse the simulations in this thesis are directly taken from the paper by Wagner-Cerena et al ([1]) . More specifically, in the paper to simulate the strongly lensed images, the components of the strong lens, such as the main deflector, the sub-halos and the line of sight halos, were separated and distinct parameterizations were used in order to physically represent each of them.

2.1.1 Main Deflector

The most significant component in a strong lensing system is of course, the main deflector. To represent the main deflectors in the simulations, the mass parameterization utilised is the power-law elliptical mass distribution with external shear. This type of profile is thoroughly described in the works of Tressore et al ([27]) and Kormann et al ([28]) the fundamental equation of the PEMD is :

$$\kappa(x,y) = \frac{3-\gamma_{lens}}{2}\left(\frac{\theta_E}{\sqrt{q_{lens}x^2+\frac{y^2}{q_{lens}}}}\right)^{\gamma_{lens}-1} \tag{1}$$

Where, x,y are the x-y coordinates defined along the major and minor axis of the main deflector, γ_{lens} the logarithmic slope, θ_E the Einstein radius and q_{lens} the minor major axis ratio[28]. Due to the fact that in order to obtain x and y values we must define a coordinate axes along the minor and major axes of the main deflector, there are in fact, 3 additional parameters to fully describe the profile. These three parameters are, the x-y positions of the center x_{lens} and y_{lens}, and the main deflector rotation angle ϕ_{lens}.
This is gives the surface mass density of the lens κ, which allows for a full description of the main deflector in simulations. The external shear component of the PEMD is added on later, and is described by the parameters γ_{ext} and ϕ_{ext} [29]. An issue arises from using ϕ_{lens} and ϕ_{ext}, since they are cyclic parameters (i.e. since they are limited to a range 0 to 2π they repeat) , to avoid this issue, the angles are expressed in terms of eccentricity and ellipticity of Cartesian coordinates[29]:

$$e_1 = \frac{1-q_{lens}}{1+q_{lens}}cos(2\phi_{lens}) \tag{2}$$

$$e_2 = \frac{1-q_{lens}}{1+q_{lens}}sin(2\phi_{lens}) \tag{3}$$

$$\gamma_1 = \gamma_{lens}cos(2\phi_{ext}) \tag{4}$$

$$\gamma_2 = \gamma_{lens}sin(2\phi_{ext}) \tag{5}$$

These equations allow for the cyclic parameters to be expressed in quantities which are not cyclic, but rather standard non repeating variables.
The last crucial step in parameterizing the main deflector, is the mass of the main deflector, m_{host}, which is described by the M_{200c} definition as established in the paper by White ([30]).

2.1.2 Sub-halos

The sub-halos are the main focus of this paper, as it is the sub-halo mass distribution that is the most interesting for comparing cosmological models. The parameterization for the mass distribution of the sub-halos inside the strong lensing system is:

$$\frac{d^2N_{sub}}{dAdm_{sub}} = \Sigma_{sub}\frac{m_{sub}^{\gamma_{sub}}}{m_{pivot,sub}^{\gamma_{sub}+1}} \tag{6}$$

Where, Σ_{sub} is the normalization of the sub-halo mass function (SHMF), m_{sub} the sub-halo mass, which is once again modelled according to the work of White ([30]), and dA the differential area. This model was made uniquely for the simulations utilised in the paper by Wagner-Cerena et al ([1]), however it is based off of the parameterization utilised by Gilman et al([15]). The sub-halos are generated within a specific mass range set by the boundaries $\{m_{min,sub}, m_{max,sub}\}$[1].
Equation (6) specifically describes the number density of the sub-halos, and how the mass of the functions is distributed in the simulations. The individual sub-halos also have a specific model in the simulations however, whose model is given by a truncated Navarro-Frenk-White (NFW) radial density profile, a model taken from Baltz et al [31]. The NFW profile is essentially a profile for the mass distribution of dark matter given by:

$$\rho_{\rm NFW}(r) = \frac{\rho_{\rm sub}}{\frac{r}{r_{s,\rm sub}}\left(1+\frac{r}{r_{s,\rm sub}}\right)^2}\frac{r_t^2}{r^2+r_t^2} \tag{7}$$

Where ρ_{NFW}, is the density function, $r_{s,sub}$ is the scaling radius, ρ_{sub} is the amplitude of the NFW function, r is the radial position and r_t is the truncation radius. The traditional NFW profile is a great representation, however it leads to potential infinities, which is why a truncated NFW profile is utilised. This type of profile limits the radius to the truncation radius which stops the function from reaching infinity[31].
Both the $r_{s,\rm sub}$ and the ρ_{sub} quantities are quantities which are calculated utilising the functions for the sub-halo mass m_{sub} and the sub-halo concentration c_{sub}. The sub-halo mass is simply calculated from equation 6, while the concentration of the sub-halo is calculated using the median mass-concentration relation established in Gilman et al ([32]). The equation for the concentration and the details behind how the variables in the NFW truncated profile relate to the mass-concentration are specified in the appendix B and in the original paper [1] .

In order to make the simulations less computationally expensive, the sub-halos themselves are only generated within a radius of 3 times the Einstein radius of the main deflector. This has a negligible effect on the actual signal, and hence is only done to reduce computation expenditure[1]. Another crucial aspect is that the sub-halos are constrained with z coordinates in the range of [-R_{200c}, R_{200c}], where R_{200c} is the minimum radius in which the enclosed halo mass has a density 200 times ρ_{crit}. Where, ρ_{crit} is the critical density of the universe at the redshift corresponding to the halo. Outside of this radius halos are considered line of sight halos, another important component of the simulations.

A more detailed description of the details behind the theory underlining the simulation of the sub-halo parameters is found in the paper in which these simulations are originally developed ie Wagner-Cerena et al ([1]).

2.1.3 Line of sight halos

Line of sight halos were ignored in many early galaxy-galaxy lensing studies[33], as their contribution was thought to be negligible compared to the signal produced by the sub-halos.

However, recent studies [10][16] have shown that their contributions could be much larger than previously thought, comparable to the signal of sub-halos. Due to the fact that fundamentally line of sight halos and sub-halos are very similar, LOS halos are also individually described by a truncated NFW profile[1]. In general, the theory behind simulating the individual LOS halos is almost identical to the theory simulating the sub-halos. The distribution of the LOS halos differs however, as the LOS halos in general are represented by a modified Sheth-Tormen halo mass function[34]:

$$\frac{d^2N_{\rm los}}{dV\ dm_{\rm los}} = \delta_{\rm los}(1 + \xi_{\rm 2\ halo}(r, m_{\rm host}, z_{\rm host})) \times \left[\frac{d^2N_{\rm los}}{dV\ dm_{\rm los}}\right]_{\rm ST}. \quad (8)$$

In this equation, δ_{los} is a scaling parameter for the LOS mass function. The ξ_{2halo} factor comes from the two-point correlation factor, which essentially measures the clustering of the halos, it is a function depending on the redshift, mass and radius of the host. Finally the last term is the classical Sheth-Tormen function as described in the original paper by R.K Sheth. et al([34]) [1].

In order to properly represent the line of sight halos, they need to be distributed over a range of redshifts, from $[z_{min,los}\ ;\ z_{source}]$, over this range the redshifts are separated into bins of size $\Delta_{z,los}$. In each one of these bins equation (8) is applied, with the equation shifting for each bin due to the dependence of ξ_{2halo} on the redshift [1].

Furthermore, there is also a constraint placed on the x and y coordinates of line of sight halos based on the value of the redshift. This constraint is given by the fact that these halos are by definition constrained by the line of sight. Hence, the x and y coordinates are constrained by a cone following the manner in which the light from the source travels [1].

2.1.4 Sources

Another important aspect in the simulations utilised by Wagner-Cerena et al ([1]) is the sources utilised. The sources are particularly important because unlike previous studies [35][36] the study on which this thesis is based utilises real sources taken from the Hubble space telescope (HST) [1].

More specifically, the sources are taken from a set of real galaxy images taken by the HST COSMOS survey using the advanced camera for surveys (ACS) F814W filter between October 2003 and June 2005. From this larger dataset of images, those of the COSMOS dataset, included in the *GALSIM* package are utilised [37] [38]. These images are then restricted based on parameters such as the minimum cutout pixel size, faintest apparent magnitude, maximum redshift, which allow for only the most appropriate images to be selected. In total this gives 2262 different sources to be lensed [1].
The light from the source is then obtained by using the *Astropy* interpolation package to preform a linear interpolation of the light. Throughout this process, the model keeps the size and absolute magnitude of the galaxy the same as that of the original COSMOS image, this is done by scaling the magnitudes based on the redshift values of the source, and then converting the electron count values of the ACS from the original image to electron counts for the target detector.

2.1.5 Datasets

All of the variables involved in the generation of the images, allow for a multitude of different configurations for image generation. In the paper ([1]), three main different datasets are identified: the training dataset, the validation dataset and the test datasets.

The training and validation datasets have the same fundamental configurations, with the only difference between the two being the galaxies used, while all of the 20 test datasets have a number of differences, mainly the SHMF, for which:

$$\mu_{sub,training} = 2 \cdot 10^{-3}, \qquad \mu_{sub,test} = 2 \cdot 10^{-4} \cdot i. \tag{9}$$

In which μ_{sub} indicates the mean value of the SHMF and i is the number of the test dataset, i.e. the fifth dataset would have a mean of $1 \cdot 10^{-3}$. The difference in galaxies utilised arises from the fact that the training dataset was utilised for training the xResnet34 neural network, while the validation and test datasets were used to obtain data. This separation is important, as it means that when the neural network is used to obtain data, it is ran on images that it has not seen before, and hence is unfamiliar with. Out of the 2262 original galaxies, 2163 are reserved for the training dataset while 99 are reserved for the validation and test datasets [1] .

2.2 Galaxy Morphology

The other main aspect of this thesis aside from galaxy-galaxy lenses is galaxy morphology. Traditionally, galaxy morphology and classification has been a very subjective process, this is evident from the fact that the main classification models, i.e. Hubbles tuning fork [39] and De Vaucoleurs system [40], are both deeply reliant on human interpretation [41].

The main focus of this thesis is the inhomogenetiy of these various galaxy types, as the main hypothesis is that more inhomogenous galaxies could cause the prediction process of the neural network to be less accurate. In order to establish whether there is a relation between this inhomogenetiy or 'clumpiness' it is crucial to establish more quantitative galactic parameters that can be used to systematically classify galaxies. These quantities are necessary as the classical classification models, are difficult to apply computationally due to having relatively approximate definitions, depending largely on interpretation. The two classification models chosen for this quantitative analysis are the CAS classification model[25] and the Sersic profile[42], both of these provide a precise quantitative manner of classifying galaxies, and crucially, they allow for inhomogenous galaxies to be distinguished.

2.2.1 CAS metrics

The main concept for this thesis is the 'clumpiness' of a galaxy, by clumpiness it is meant a galaxy which has high intensity regions outside of the main central bulge, the more of these regions which a galaxy has the more 'clumpy' a galaxy is. This is the main concept as, the hypothesis of this thesis is that the 'clumpiness' of a galaxy could cause the neural network some problems, causing it to give less accurate Σ_{sub} values. The most common index utilised to quantify the clumpiness of a system is the clumpiness metric of the CAS morphological classification model, a model first developed by Conselice et al ([25]).

The essence behind the CAS clumpiness metric is to isolate the high frequency elements (the 'clumps') by subtracting a smoothed version of the image from the original image. The reasoning behind this is that the smoothed image leaves the non clumpy parts of the galaxy unchanged, while smoothing out the high frequency elements that are of interest. The result is then an image in which only the high frequency components are highlighted[25]. The precise computational definition of the high frequency clumpiness parameter is [25]:

$$S = 10 \sum_{x,y=1,1}^{N,N} \frac{(I_{x,y} - I^{\sigma}_{x,y}) - B_{sub,x,y}}{I_{x,y}} \tag{10}$$

Where $I_{x,y}$ is the flux of the original galaxy image, $I^{\sigma}_{x,y}$ the flux of the galaxy when put through a smoothing filter of width σ, $B_{sub,x,y}$ is the flux of a region of the background of size equal to the galaxy subtracted by a smoothed version of that same background, while N is the size of the galaxy in pixels.[25][43]
In essence, first the smoothed and unsmoothed images are subtracted to obtain a map of the high frequency clumps, and then subsequently, the background is subtracted in order to remove any noise. Subsequently, the clumpiness parameter is normalized in order to achieve a value of S which lies between 0 and 1. [25]

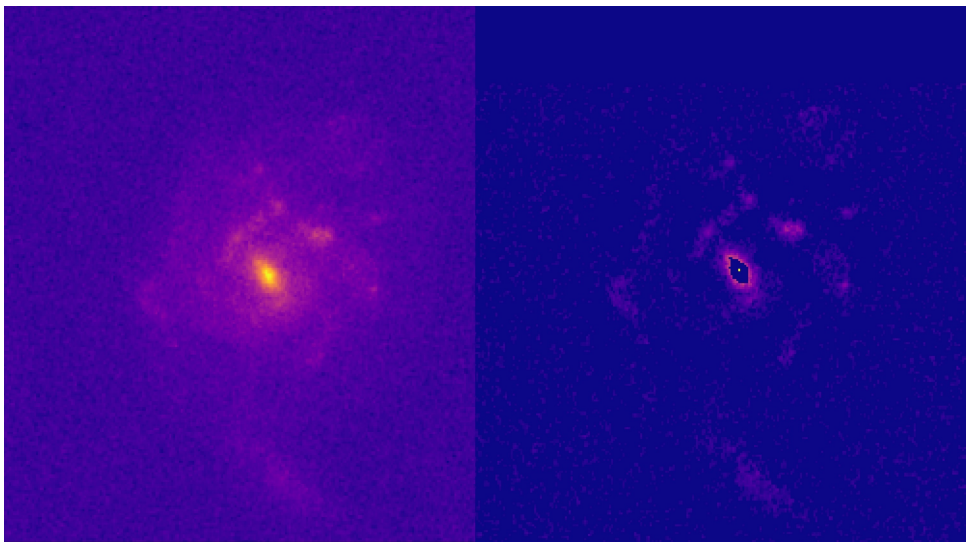

Figure 5: Comparison of a galaxy before and after the reduction process used to find CAS clumpiness as described in the paper by Conselice et al [25]. The plots are made utilising the flux per pixel values, utilising a linear normalization. The brighter areas representing the high intensity regions. The reduced image leaves behind only the main clumps outside of the central bulge

To obtain the smoothed image, it is necessary to apply a Gaussian kernel smoothing to the image, the width of this smoothing is given by:

$$\sigma_{smoothing} = 0.3 \cdot r(\eta) \tag{11}$$

The $r(\eta)$ denotes the Petrosian radius(i.e. The radius at which the ratio of the local surface brightness in a ring of radius r to the mean surface brightness within r is equal to η), the specific value of eta can be varied based on exactly what portion of the galaxy is probed, and in which region the clumps are assumed to be located. Before preforming this calculation for S, it is necessary to set all the negative

flux values from the numerator to 0, and crucially to ignore the central bulge, as this would cause an increase in S due to the high frequency nature of the bulge. [44][25] [43]

2.2.2 Sersic Index

The other parameter utilised in this experiment to classify galaxies based on their morphology is the Sersic index (n) which is a specific component of the Sersic profile of a galaxy, given by[45]:

$$I(R) = I_0 \cdot e^{-kR^{1/n}} \tag{12}$$

In which I is the intensity of the emitted light, R is the distance from the center and k is a constant given by the half-light radius of the galaxy(ie the radius which encompasses half of the galaxies total light emissions) [45]. The index dictates how luminous the central bulge is relative to the rest of the galaxy. This is incredibly useful in differentiating between ellipses and spirals, as in most cases the elliptical galaxy will be composed of almost uniquely its central bulge, while spiral galaxies tend to be less bulge dominated, thus spirals will have a much lower Sersic index in general [46].

More importantly for the purposes of this thesis, the Sersic index be used to distinguish between 'clumpiness' of galaxies, as more clumpy galaxies will have more luminosity being generated by areas which are not the central bulge. This is because 'clumps' are simply areas of high light intensity, and having high light intensity patches means that the bulge luminosity to total luminosity ratio decreases. This working principle has already been applied in several papers such as Elmgreen et al. ([26]) where a clear relation was found between clumpier galaxies and lower Sersic indices. The Sersic index is less reliable in being able to classify clumpiness compared to the CAS metric, as the Sersic parameter is not specifically designed to quantify the clumpiness of a galaxy, and hence is influenced by other factors, for example, in general more luminous galaxies will have larger Sersic indices even if they are not as clumpy as some lower luminosity galaxies. [47]

3 Methods

3.1 Simulations

The first step in being able to establish whether the clumpiness effects the performance of the neural network is to generate the images. Two different datasets were utilised throughout these simulations, as obtaining the original dataset utilised in the paper [1] proved impossible. The first dataset is composed of 12370 images generated utilising the training configuration galaxies. The other dataset is generated utilising the validation configuration, and contains a total of 99 different galaxies. Using a different dataset than the training dataset could prove useful, as since the network is trained on the training dataset it is 'familiar' with the galaxies used.

After the generation of these images, the xResnet34 neural network was run on the images, allowing the neural network to make predictions on the parameters of interest. To run the network, the *bendorbreak* package was utilised, in similar fashion to how it was implemented in the original paper by Wagner-Cerena et al [1] .

After the neural network is ran, all of the necessary data is available. The truth values are extracted from the metadata files generated with the images, while the predicted values and errors in the predicted values are found in the npy files generated by the neural network.

More specifically, the metadata of the images is used to extract the catalog index of the galaxy and the true SHMF values, while the network files are used to extract the predicted SHMF values and the errors in these predicted values. The galaxy index values are particularly important because they allow for more metadata to be obtained via the COSMOS galaxy catalog contained within the *GalSim* package, this allows for the half light radius to be determined, a step which will be crucial for the clumpiness calculation.

Using the values of the true Σ_{sub} and the predicted Σ_{sub}, it is possible to then determine the systematic error, which is ultimately one of the two manners in which the neural network performance will be analysed. The specific equation utilised to obtain the systematic error will be the following:

$$\sigma_{systematic} = < \Sigma_{sub,prediction} - \Sigma_{sub,truth} > \tag{13}$$

Additionally, as another measure of the performance of a certain galaxy within the network, the random error, or spread in the values is also calculated. This is done, as for certain galaxies it could be the case that although the systematic error is low, the spread is large, hence measuring only the systematic error would not be an adequate measure of the performance of said galaxy in the network. The spread is calculated utilising the root-mean square error:

$$\sigma_{rms} = \sqrt{< (\Sigma_{pred} - \Sigma_{truth})^2 >} \tag{14}$$

Before performing any calculations however, the galaxy index values are used to restructure the data of both the predicted and true Σ_{sub} values. This is done in order to be able to group images of the same galaxy together, as in order to reduce the number of total data points, images of the same galaxies are averaged giving the mean performance of said galaxy. The intent behind averaging these images

together is to allow for a more precise description of how one specific galaxy will perform in the network, as if only one image per galaxy was used then the galaxy's performance could be misrepresented . Furthermore, this method also allows to avoid plotting all 10k images and having a significant number of overlapping clumpiness values with differing error values.

Subsequently, it is necessary to shift the predicted SHMF values as, as highlighted in the paper [1] the neural network predicts the SHMF rather poorly, resulting in a distribution clumped around the mean value of $2 \cdot 10^{-3}$. This is shown in figure(6, where instead of a distribution resembling the one for Σ_{true}, the predicted values are clumped around the mean. This step also serves as a test to verify whether the utilised dataset is compatible with the neural network or not, as if the predicted values are significantly off of the expected mean value the neural network is not predicting the values as it should.

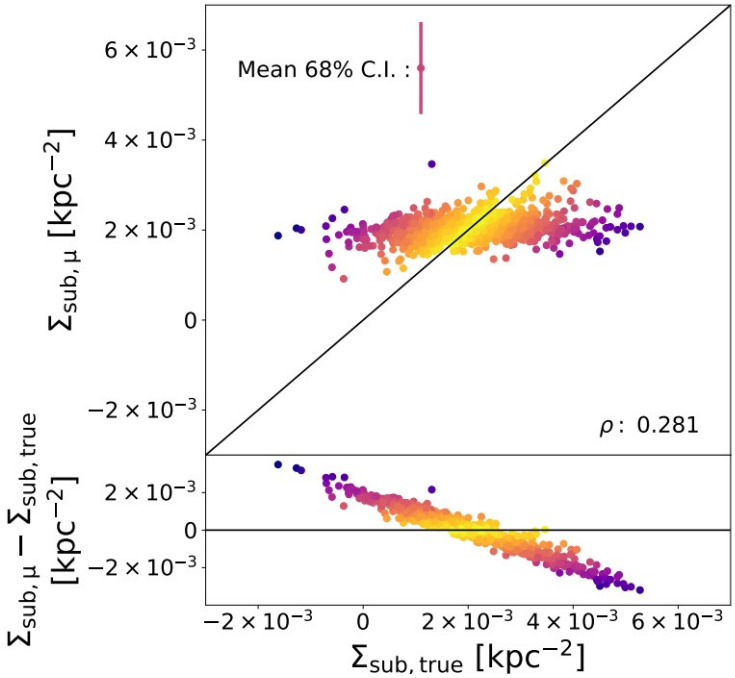

Figure 6: Graph from the paper by Wagner-Cerena et al [1]. In the upper section, the predicted SHMF are plotted against the true SHMF.On the bottom the difference between true and predicted values are plotted against the truth values.The colormap represents the differences between the truth value of the points and the mean.

In order to fix the prediction of the neural network, the values of the predicted SHMF are shifted by modifying the line of best fit to be equal to a straight line of slope 1. This is done by adding the difference between the original line of best fit and a line of slope 1, this process is then repeated until the line of best fit overlaps with the line of slope 1. To apply this process to the points as well, the points are associated with a segment of the best fit line, and then the same shift that would be applied to that segment is applied to all the points in that segment.

Finally, each image is analysed to determine the nature of the galaxy with which the image was generated. This is done in two different methods, as throughout this project, two different morphological parameters are used to try and establish a connection between neural network performance and galaxy

morphology.

3.1.1 CAS Parameters

The most significant of the two morphological parameters is the clumpiness index as defined in the CAS classification system. In order to calculate the *COSMOSclass* function is utilised, as not only does it allow for the plotting of images, but it has a built in smoothing function, which allows for smoothed images to be obtain rather straightforwardly. The smoothing is crucial, as in the equation for the clumpiness index:

$$S = 10\frac{\sum_{x,y=1,1}^{N,N} I_{x,y} - I_{x,y}^{\sigma} - B_{sub,x,y}}{\sum_{x,y=1,1}^{N,N} I_{x,y}} \tag{15}$$

Both $B_{sub,x,y}$ and $I_{x,y}^{\sigma}$ require smoothing. The smoothing in the COSMOSClass dictionary, also allows for modification of the size of the smoothed area, which is necessary for the procedure to obtain the smoothed image. In the original procedure for the clumpiness metric, a smoothing size equal to $0.3 \cdot r(\eta)$, the $r(\eta)$ value can be modified to best isolate certain clumps. The simplest way to do this is to let η be equal to 0.5, and utilise the half-light radius. However, using the half light radius resulted in clumpiness values which were too high for all kinds of galaxies, meaning that the high intensity clumps were not being isolated. To fix this issue, the factor of 0.3 was instead replaced with a factor of 0.05, this has a similar effect to utilising a smaller Petrosian radius, giving values in the range described by [25]. This specific choice of coefficient was given by the method utilised in the paper by Hambleton et al [43].

To obtain the background image, a part of the background is taken, and then this part is stretched so as to be the same size as the image of the galaxy. This step is necessary, as otherwise it is impossible to subtract the background image from the image of the galaxy.

Before subtracting the smoothed and unsmoothed image some steps are taken. First of all, it is necessary to convert the image into flux per arcsecond. This is because the definition of the clumpiness index relies on the fact that the image is expressed in flux per arcsecond, while the image provided by the *drawsource* function of the COSMOSCatalog is expressed in flux per pixel. To convert from one to the other, it is sufficient to note that all the generated images are generated with 0.03 arcseconds per pixel[36], hence:

$$I_{x,y}^{arcsecond} = 0.03 \cdot I_{x,y}^{pixel} \tag{16}$$

In which $I_{x,y}^{arcsecond}$ is the image in flux per arcseconds and $I_{x,y}^{pixel}$ is the image in flux per pixel.
Subsequently, low size and high noise images are filtered out, this is because this effects the clumpiness metric, as a galaxy with very few 'clumps' could be found to have large clumpiness values simply because of the significant amounts of noise in the background. In theory the B term in 15 should remove this effect, but at very low resolutions (i.e less than 100 pixels), the background noise is significant enough that even subtracting the background does not completely eliminate the noise. The specific size 'limit' of more than 100 pixels in size, as for values below this limit, when galaxy images were plotted, all of the images were entirely covered by background noise.

Finally, the negative flux values of the images must be replaced by 0, as these values are simply a result of over-subtraction and the central bulge must be removed. The removal of the central bulge is necessary in the process of calculating the CAS clumpiness index as leaving it would leave a high intensity region which would contribute to the final clumpiness value. In order to remove the center, the pixels with the highest flux were removed, to make sure that the number of pixels corresponded to the center and did not include other high intensity areas, the half light radius was used once again. The half light radius is a good representation of the radius of the central bulge region of the galaxy, and hence removing all of the pixels within $2 \cdot r_{half-light}$ is a good estimate of removing the central bulge.

3.1.2 Sersic Index

The other variable utilised to quantify the 'clumpiness' of a galaxy is the Sersic index. The process for determining the Sersic index is far simpler than that for determining the CAS clumpiness index. More specifically, to obtain the values of the Sersic index, the galsim package is utilised as it automatically carries out a Sersic fit on the galaxy images, and returns a value for the Sersic index. Each of these galaxy images is indexed by the previously determined galaxy index, hence simply extracting the specific Sersic index corresponding to the galaxy source index gives the final value.

In order to filter out the galaxies for which the Sersic model may not have been applied perfectly, a filter is applied utilising another in built variable, namely, the FIT_MAD_S variable. This value describes the median absolute deviation from the Sersic fit, hence filtering with this variable allows for the galaxies with a poor fit to be excluded.

3.2 Data Analysis

After having obtained all the values for the required variables, it is necessary to analyse the data and determine whether there is a link between clumpiness and the performance of the neural network or whether the two are independent.

Several methods are used to try and establish the relationship between the clumpiness index and the performance of the neural network. The first method will be to simply plot the best and worst performing galaxies in terms of the systematic and root mean square error. This serves simply as an initial visual test as to whether differing galaxy shapes may effect the error values. Although this method is more of a qualitative comparison, it still could prove valuable, as morphological analysis of galaxies has historically always been a very vision based topic. This is due to the previously mentioned concept that galaxy classification is something incredibly dependent on interpretation[41]. Hence, plotting the images could allow for patterns to be found that an analytical system could not.

In order to plot these images, the *drawsource*, function is utilised, allowing for galaxy images to be plotted. The final plots are of galaxy images with a colormap given by the normalized counts per pixel found in the original image. This process will be carried out for the 4 best and 4 worst performing galaxies, i.e. the 4 galaxies with the largest and smallest error values. In addition, before plotting all the negative flux values are set to 0, as negative flux values are simply a result of the image pipeline

and have no physical meaning.

The main statistical method utilised to determine whether there is a relation between the two variables is a scatter graph along with two separate correlation. This method involves simply plotting the direct clumpiness values against the error values, a line of best fit with a large slope would show a relationship between the two and hence establish whether the two values are related. The fitting of the line of best fit is carried out utilising the $np.polyfit$ function from the numpy package, this function was chosen as opposed to the scipy $curvefit$ function as the numpy function was easier to implement into the code, changing from one function to the other should yield no significant differences.

In conjugation with the line of best fit, the Pearson and Spearman correlation coefficients between the data was established. This was done by utilising the $scipy.stats$ library, which calculated not only the correlation coefficients, but also the p-values associated with these correlation coefficients. The p-values essentially give the data being distributed in the manner it is given that the null hypothesis is true. This means that the smallest the p-value the greater the likelihood of there being a correlation. The general accepted threshold for p-value is of 0.05 [48]. The coefficients themselves measure the strength of a linear correlation (Pearson) and a non-linear correlation(Spearman), these values range from -1 to 1. A value of 1 indicates there is a strong positive correlation while -1 indicates a strong negative correlation [49]. The error in the coefficients was given by the definition of the standard error [50] :

$$\sigma = \sqrt{\frac{1-r^2}{n-2}} \tag{17}$$

3.3 Error analysis

A crucial part throughout the analysis of the obtained data is determining the errors of said data before plotting.

The Sersic index as previously stated, has an error value in built in the metadata of the *galsim* COSMOS images. This error is simply given by the FIT_MAD_S variable, which gives the median absolute deviation between the Sersic fit and the image. Since each point plotted will be an individual galaxy, and the Sersic error is constant for each galaxy, the error values can simply be extracted from the metadata and directly utilised.

For the other variables more work is required in order to derive the error values, as the equations for the errors are not easily derivable.

As previously stated, two main performance parameters utilised throughout this thesis, the systematic error and the spread. In order to determine the error in the systematic error, the formula for the standard error of a mean was utilised. This is because the final points plotted will be the points given by the mean of a set of images of the same galaxy. Hence, in order to determine the error in this final mean error, the standard error formula will be utilised:

$$error = \frac{\sigma_{std}}{\sqrt{N}} \tag{18}$$

In which N is the total number of images of that one galaxy and σ_{std} is the standard deviation of the

set of systematic errors.

For the other parameter, defining a formula for the error proves significantly more difficult. As such, the bootstrap technique for determining errors is implemented. This method consists of estimating the standard error values utilising small datasets, such as the set of images of each individual galaxy. In essence the small dataset is resampled a large number of times, specifically 10000 for these errors, and the standard error is then calculated using the standard deviation. Using this method allows for the errors of quantities whose error values are difficult to calculate to be calculated.

The clumpiness index proves even more complicated, as the dataset size for the clumpiness index is only 1, as images of the same galaxy will clearly have the same clumpiness value. This impedes the utilisation of the bootstrap method. Therefore, the calculation of the error is based off of the limited pixel resolution and the limited flux values given in the metadata parameters. More specifically, an uncertainty in the I and B values in the formula 10 equal to the resolution error in the flux values. This error is then propagated utilising the partial derivative error propagation formula:

$$\Delta S = \sqrt{(\Delta I \cdot \frac{\partial(S)}{\partial I})^2 + (\Delta B \cdot \frac{\partial(S)}{\partial B})^2} \tag{19}$$

This then gives the final error value of the CAS clumpiness parameter.

4 Results

4.1 Training dataset

First of all, the training dataset was tested. This dataset contained 2163 different galaxies in 12370 different lensed images. The bias in predicting the SHMF was the first variable tested, this bias was calculated by taking the average error between the predicted SHMF and the true SHMF for each galaxy.

4.1.1 Bias

Initially, the bias was plotted against the CAS clumpiness index yielding figure (7). This figure shows a line of best fit with a slightly decreasing slope, potentially indicating a slight decrease of bias as the clumpiness index increases. However, the error in the line of best fit was important, seemingly indicating that there might be no relation at all, as the graph gave a slope of -0.2 with an error of ± 0.1.

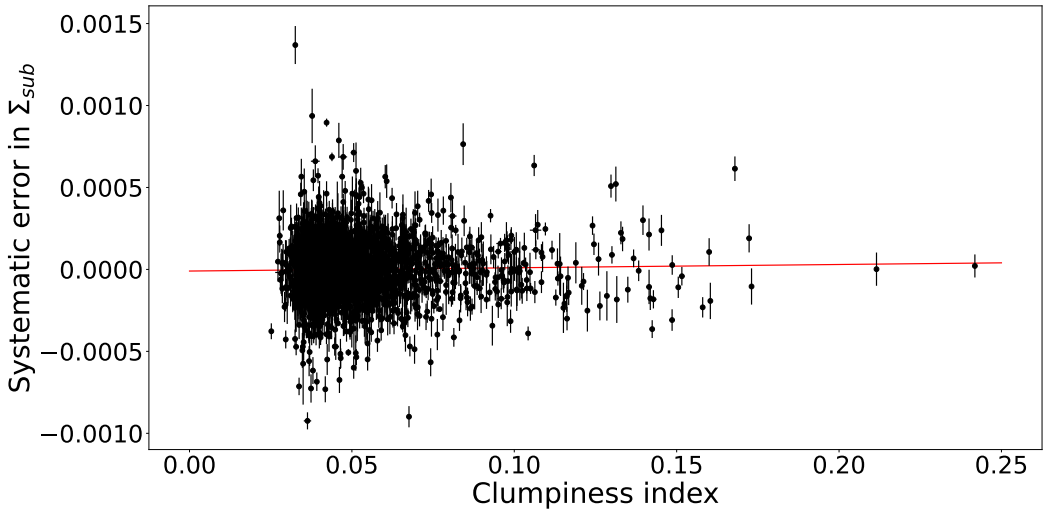

Figure 7: Plot showing the SHMF bias against the clumpiness index, the red line is the line of best fit. Each individual point represents one single galaxy, for which multiple images are generated within the dataset, the bias is plotted utilising the formula described in the Methods section.

The non-correlation between the values is further supported by the obtained p-values for the Pearson and Spearman correlation coefficients. These p-value both lie above the typical threshold utilised to establish correlations of 0.05

$$p_{Pearson} = 0.26, \qquad p_{Spearman} = 0.27 \tag{20}$$

While the p-values are low and may indicate at best a weak relation, they are not significant enough for a clear relation between the clumpiness index and the bias.

The Sersic index similarly showed no relation to the bias in the SHMF. This is shown clearly in the graph (8), for which the best fit line is almost completely flat. Specifically, the line has a slope of $-5 \cdot 10^{-3} \pm 7 \cdot 10^{-3}$

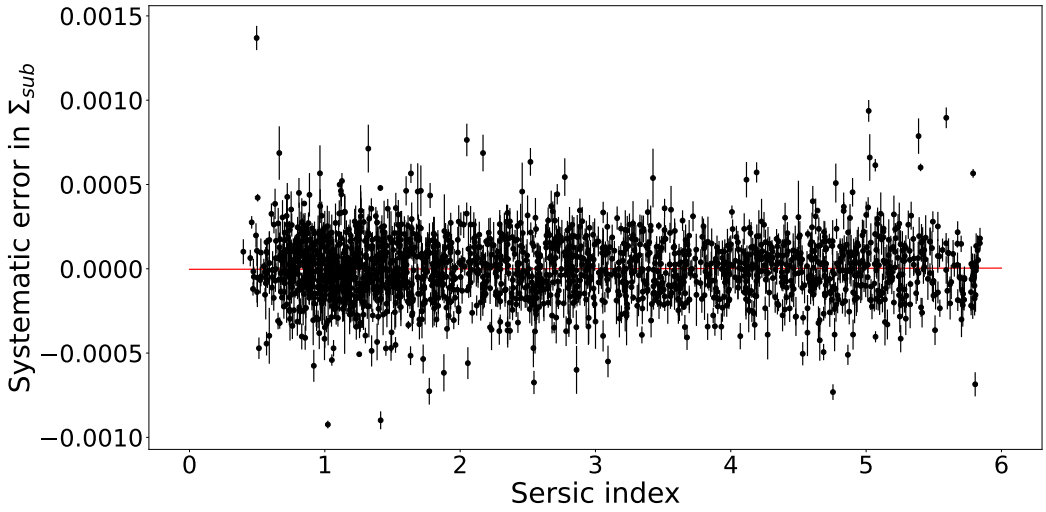

Figure 8: Plot showing the SHMF bias against the Sersic index, the red line is the line of best fit. Each point is a single galaxy within the training dataset. The Sersic indices are obtained utilising the metadata of each galaxy image.

This is further consolidated by the p-values of the correlation coefficients, which yielded values even higher than those obtained in the plots of the CAS clumpiness index:

$$p_{Pearson} = 0.56, \qquad p_{Spearman} = 0.63 \tag{21}$$

These values are even larger than those established for the CAS clumpiness index indicating no relation at all between the galactic morphology and the bias.

After this an attempt at establishing a relationship was made by plotting the galaxies with the highest difference between the predicted and true values. The images plotted were the following:

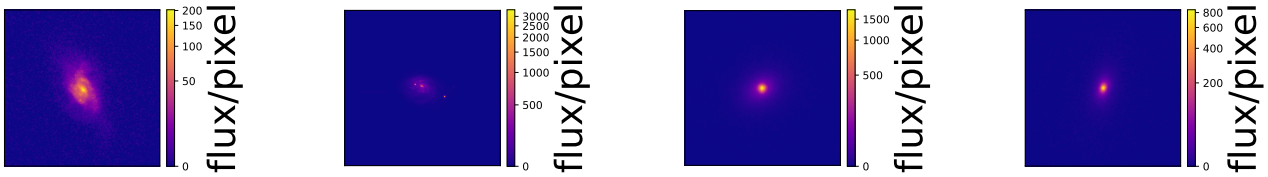

Figure 9: Image plots made utilising the normalized flux per pixel values, as indicated on the colorbar. Since the colorbar is normalized for each image the values are different for each image. The 4 galaxies shown are the 4 galaxies with the largest differences between the predicted and true SHMF values

The images show the intensity of light radiated by different parts of the galaxies, with the colorbar on the right of the images indicating the normalized flux/pixel as observed by the HST.

This was also done with the best performing galaxies to see if any evident differences between the two groups were visible.

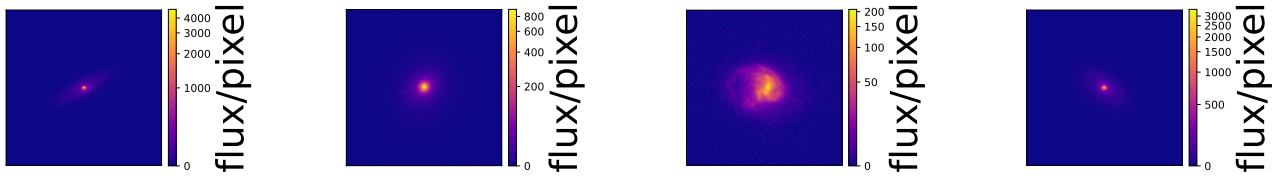

Figure 10: Image plots made utilising the normalized flux per pixel values, as indicated on the colorbar. Since the colorbar is normalized for each image the values are different for each image. The 4 galaxies shown are the 4 galaxies with the smallest differences between the predicted and true SHMF values.

The plotted images of the best and worst performers (9 and 10) give a similar conclusion as that drawn from the other indicators, namely, no relation can be established, as there is a variety of galaxy shapes in both the best and worst performers.

4.1.2 Random error

Similarly to the bias, the random error also seemed to show no clear relation to any of the morphological parameters.

For the clumpiness index, this is visible in the graph, as the line of best fit is very nearly flat indicating at most a very weak relation

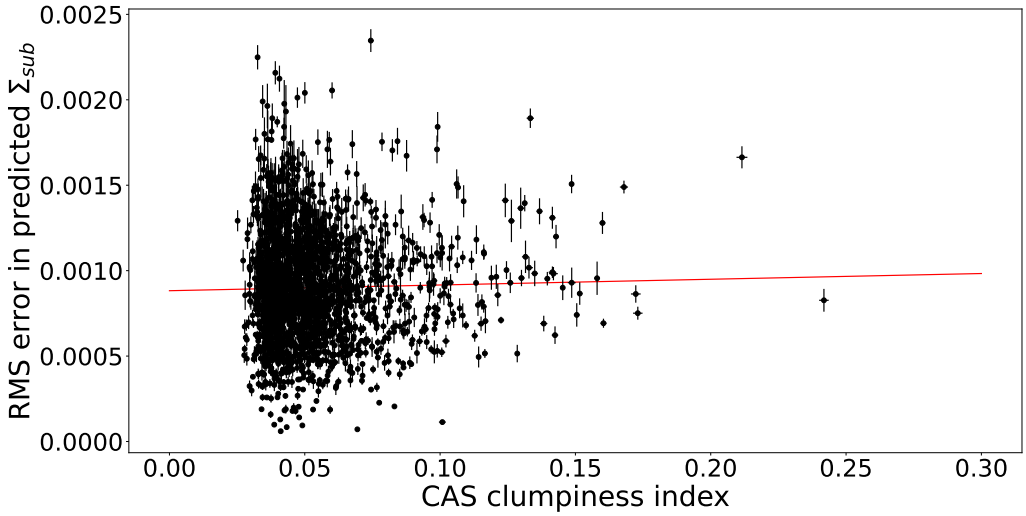

Figure 11: Plot showing the SHMF spread values, as described in the methods section against the clumpiness index, the red line is the line of best fit. The line of best fit is calculated utilising the $np.polyfit$ function.

This non correlation is consolidated by the p-values, which once again lie significantly above the threshold required to establish a clear correlation between the two quantities:

$$p_{Pearson} = 0.31, \qquad p_{Spearman} = 0.49 \tag{22}$$

The correlation between the Sersic index and the spread seems to be exactly the same as that between the clumpiness index. Specifically, the sersic index shows no clear correlation at all, this is shown in figure (12):

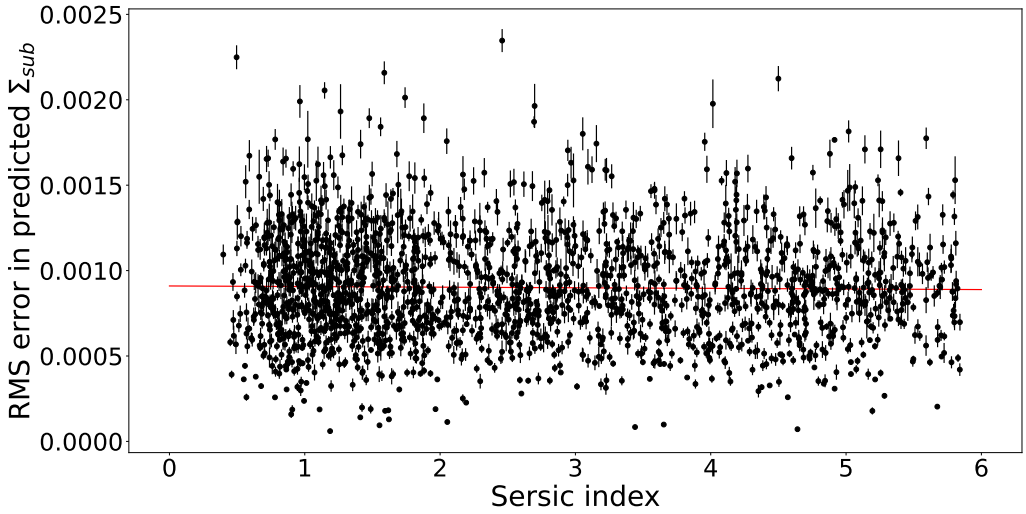

Figure 12: Plot showing the SHMF spread values, as described in the methods section against the Sersic indices, the red line is the line of best fit. The line of best fit is calculated utilising the $np.polyfit$ function.

The line of best fit is almost completely flat, and no relationship at all is visible in the data. This is also supported by the p-values and correlation coefficients which gave values of:

$$p_{Pearson} = 0.46, \qquad p_{Spearman} = 0.63 \tag{23}$$

Interestingly, the galaxy images seem to show some sort of relation between morphology and performance as shown in the figures below:

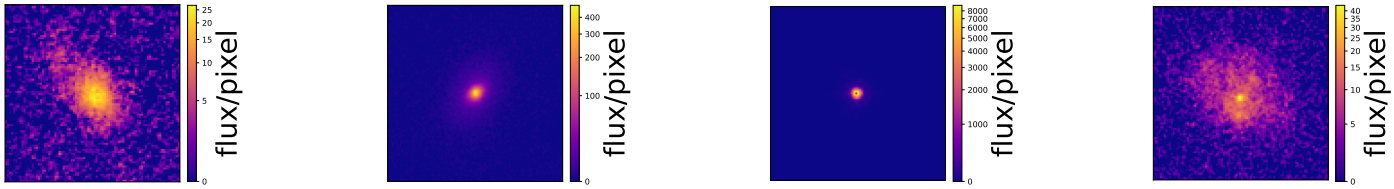

Figure 13: Image plots made utilising the normalized flux per pixel values, as indicated on the colorbar. Since the colorbar is normalized for each image the values are different for each image. The 4 galaxies shown are the 4 galaxies with the largest spread values.

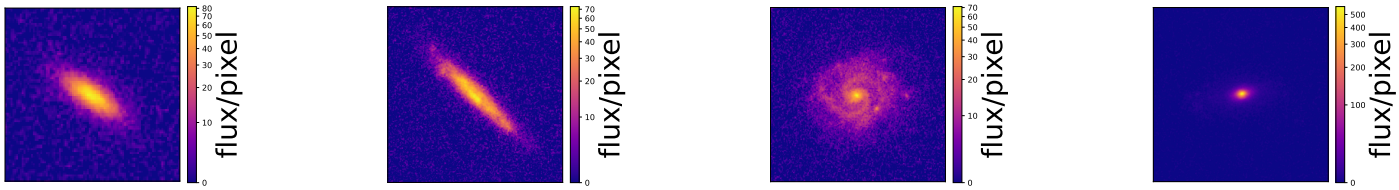

Figure 14: Image plots made utilising the normalized flux per pixel values, as indicated on the colorbar. Since the colorbar is normalized for each image the values are different for each image. The 4 galaxies shown are the 4 galaxies with the smallest spread values.

The images in contrast to the plots, do seem to show some relation, as the best performers seem to all be elliptical galaxies with very little high intensity regions outside the central bulge while the worst performers are all somewhat 'clumpy'.

4.2 Validation dataset

After testing the dataset generated with the training configuration, the dataset generated with the validation dataset was tested. This dataset with 99 galaxies yielded a total of 5735 images, with approximately 50 images generated per each galaxy. This dataset contained galaxies which the neural network had never analysed before however, in spite of this the results still showed no potential correlations.

4.2.1 Bias

For the bias, the graph yielded a best fit line which was essentially flat. This is visible in figure (15). Much like the training dataset graph, the slope had an error as significant as the slope itself. Specifically, a slope of -0.4 and an error in the slope of ± 0.3, showing no clear relation between the clumpiness index and the bias of the data :

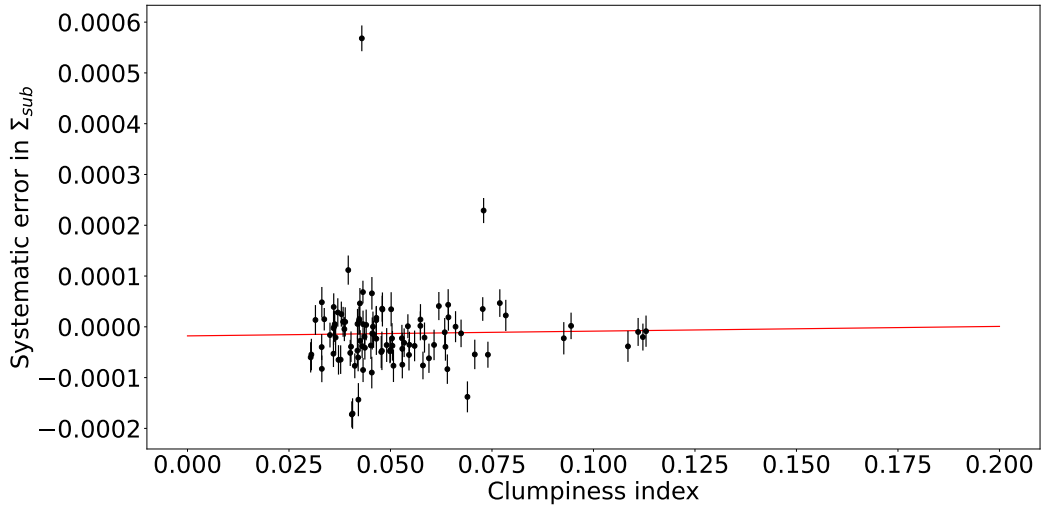

Figure 15: Plot showing the SHMF systematic against the clumpiness index, the red line is the line of best fit. bias is calculated as described in the methods section. Each point is a galaxy in the validation dataset

This non-correlation was supported by the obtained p-values, for both correlation coefficients, which were much larger than the p-values obtained from the training dataset:

$$p_{Pearson} = 0.37, \qquad p_{Spearman} = 0.46 \tag{24}$$

These values indicated no correlation whatsoever between the CAS clumpiness index and the bias in the predicted SHMF, as they were significantly above the typical threshold of 0.05 required to establish that there is a relationship between the two quantities.

Similarly, the Sersic index also seems to show no relation at all to the bias of the SHMF in the validation dataset, this is evidenced by the graph (16) which shows how the line of best fit between the two variables has a slope which is essentially completely flat:

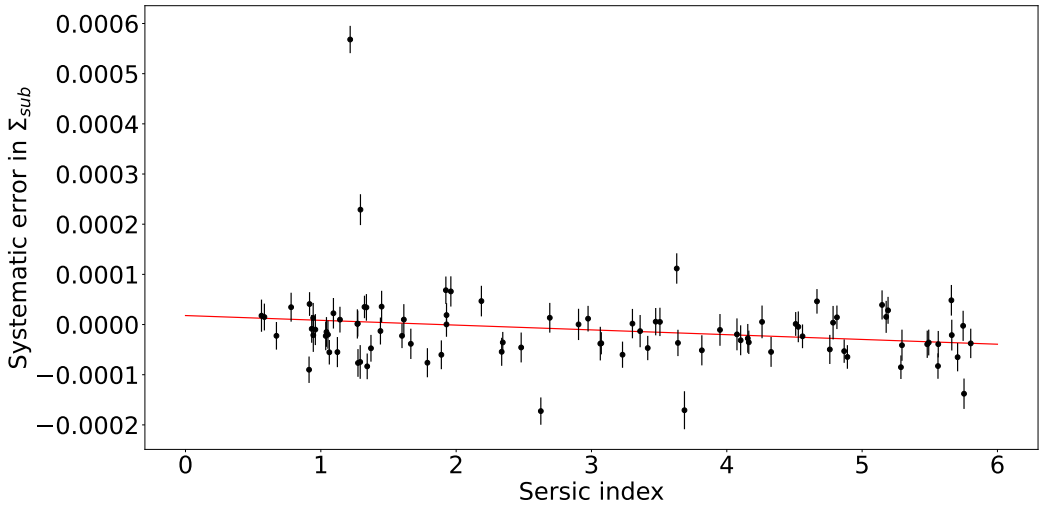

Figure 16: Plot showing the SHMF systematic against the Sersic index, the red line is the line of best fit. bias is calculated as described in the methods section. Each point is a galaxy in the validation dataset.

Once again, as with the clumpiness index, this non correlation was also highlighted in the p-values for both Pearson and Spearman correlation coefficients:

$$p_{Pearson} = 0.09, \qquad p_{Spearman} = 0.12 \tag{25}$$

While these values were smaller than the values obtained for the CAS clumpiness index, they were both above the required threshold of 0.05, showing a weak correlation at best. It should be noted that there was one extreme outlier, which could have caused some issues in the final fit. In order to further examine whether the Sersic index and the bias are related this outlier should be excluded, and it could result in a weaker correlation.

The image plots also indicated no apparent relationship between galaxy morphology and bias in the validation dataset, as both the worst and best performing galaxies were all incredibly 'smooth' elliptical galaxies:

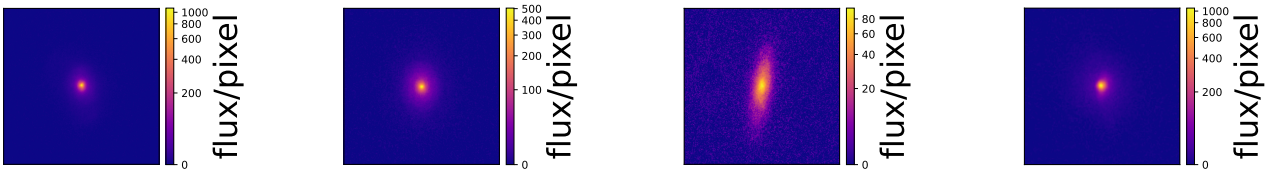

Figure 17: Image plots made utilising the normalized flux per pixel values, as indicated on the colorbar. Since the colorbar is normalized for each image the values are different for each image. The 4 galaxies shown are the 4 galaxies with the largest differences between the predicted and true SHMF values.

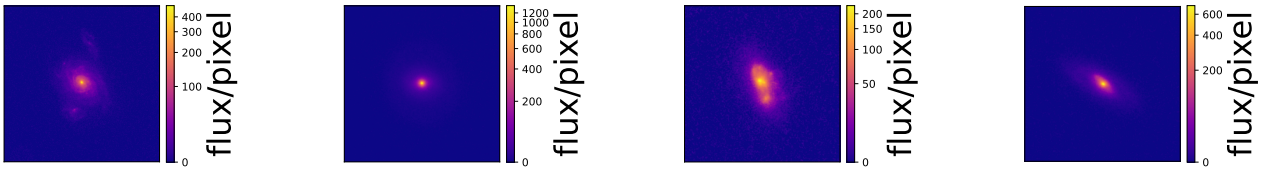

Figure 18: Image plots made utilising the normalized flux per pixel values, as indicated on the colorbar. Since the colorbar is normalized for each image the values are different for each image. The 4 galaxies shown are the 4 galaxies with the smallest differences between the predicted and true SHMF values.

4.2.2 Random error

Results for the random error of the validation dataset seemed to show the opposite relation to that found in the training dataset, although once again the low galaxy number effected the final results. The initial graph of the clumpiness index against the spread (19), showed a clearly decreasing slope indicating the opposite to the relationship found in the training dataset:

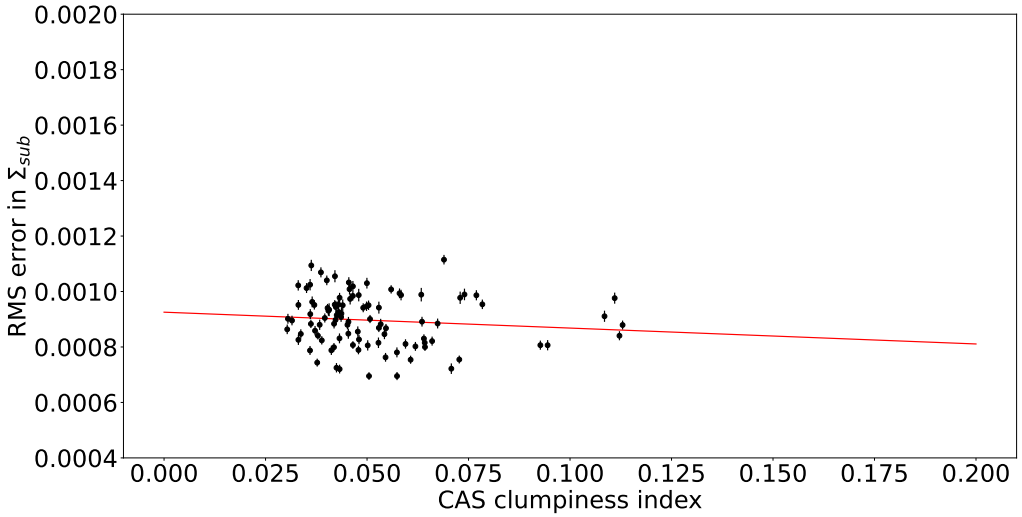

Figure 19: Plot showing the SHMF spread values, as described in the methods section against the clumpiness index, the red line is the line of best fit. The line of best fit is calculated utilising the $np.polyfit$ function.

However, the best fit line had an error in the slope of the same order as the slope itself, showing no correlation between the random error and the clumpiness index. This was was further supported by the values obtained for the Spearman and Pearson p-values, which were:

$$p_{Pearson} = 0.32, \qquad p_{Spearman} = 0.44 \tag{26}$$

The p-values obtained were much like the bias significantly above the value of 0.05 needed to establish a correlation.

Additionally, the Sersic index graph also indicated no relationship between the two variables, resulting in a graph with a flat best fit line, and supporting Pearson and Spearman coefficients which indicated no relationship between quantities:

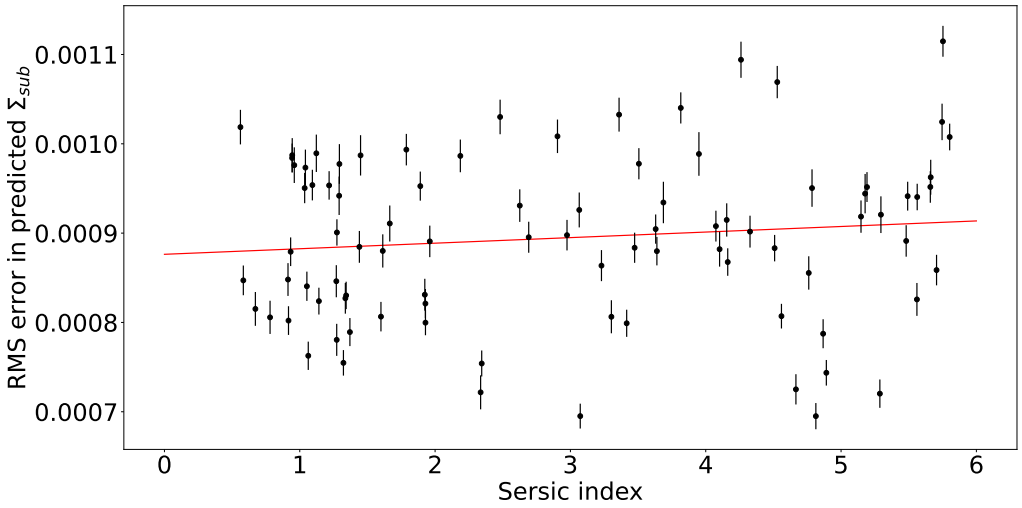

Figure 20: Plot showing the SHMF spread values, as described in the methods section against the Sersic index, the red line is the line of best fit. The line of best fit is calculated utilising the $np.polyfit$ function

While the best fit line is slightly increasing, the p-values of the correlation coefficients are once again significantly too high to establish any relation:

$$p_{Pearson} = 0.42, \qquad p_{Spearman} = 0.68 \tag{27}$$

After this process, once again, the worst and best images were plotted, with the background noise removed in order to show the galaxies better:

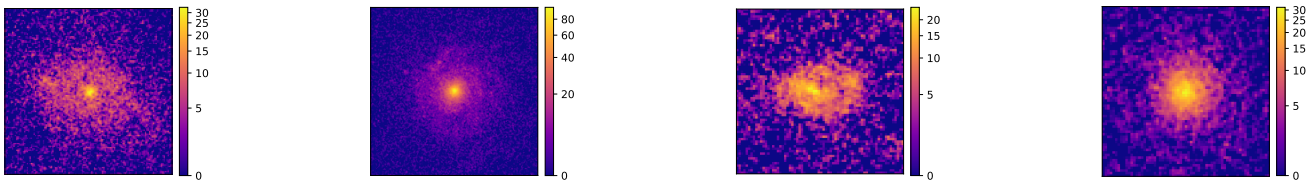

Figure 21: Image plots made utilising the normalized flux per pixel values, as indicated on the colorbar. Since the colorbar is normalized for each image the values are different for each image. The 4 galaxies shown are the 4 galaxies with the smallest spread values.

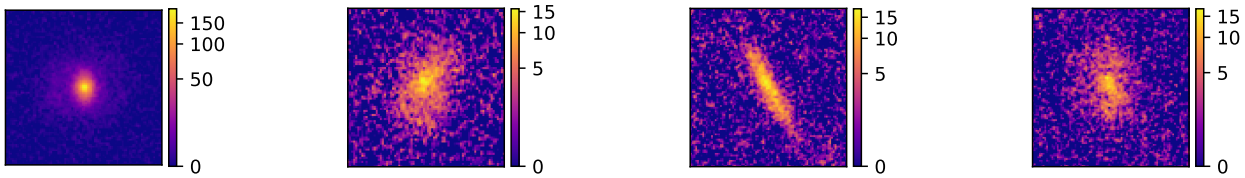

Figure 22: Image plots made utilising the normalized flux per pixel values, as indicated on the colorbar. Since the colorbar is normalized for each image the values are different for each image. The 4 galaxies shown are the 4 galaxies with the largest spread values.

All sets of images show the same conclusion, like the p-values and the slope of the graph, there is no relation between the galactic structure and the random error of the sub-halo mass function.

5 Discussion

Overall, the neural network developed by Wagner-Cerena et al ([1]) was thoroughly tested utilising 2 different configurations, over 2000 different galaxies and a total of roughly 15000 images. Utilising all of these different images, an attempt was made to establish whether there the morphological structure of a galaxy effects the performance of the neural network or not.

None of the potential correlations investigated exhibited any significant correlations. This was exemplified by the p-values of the scatter plots, which all were significantly above the threshold of 0.05 required in order to establish that a correlation between the two variables exists. This is exemplified by the fact that the smallest p-value obtained in all possible datasets, and parameter combinations was of 0.09

A promising result is obtained when analysing the image plots of the galaxies with the largest and smallest spreads (13, 14), as they seem to follow a trend. However the graphs and p-values indicate no such relation meaning that perhaps the extreme under and over-performers do follow a morphological pattern, but the galaxies between these extremes do not clear relation, hence resulting in low p-values and best fit line slopes. Furthermore, the images form only a dataset of 8 galaxies, which is not enough to properly establish any sort of correlation.

The validation dataset showed what appeared to be slightly more clear correlations, based on the slopes shown in figures (15) and (19). However, the low number of galaxies within the datasets led to higher p-values, meaning that no significant relation between the variables could be established. The only significant exception to these higher p-values was the graph in figure(16), which showed a potential relation between bias and Sersic index, although the p-values were still above the needed threshold of 0.05. In addition, the presence of an outlier in some of these graphs could have caused the larger slopes, further showing that these results are not significant.

The Sersic index showed no clear relation under most conditions, indicating that the morphological structure of the galaxy is essentially irrelevant in the performance of the neural network. This is highlighted by the graphs (8) and (12) which has a completely flat relation curve. However the results shown in figure (16) do have some of the lowest p-values out of all the plots made, and while these are still above 0.05, they could potentially indicate a weak correlation. This could warrant further research into the relations. The only issue with this plot is there is a couple of clear outliers which could cause issues in the final fit, as in total there are only 99 galaxies.

There could be several secondary explanations as to why the data yielded no results, despite there seemingly being a relation in earlier iterations of the neural network developed in the paper ([1]). The most evident issue with the obtained results is that the original training data could not be retrieved, and hence the original dataset which yielded the results presented in the paper wasn't analysed.

Furthermore, the clumpiness indices have large overlap values, this is due to the fact that the majority of galaxies will lie within a clumpiness value of 0.08 as shown in the paper by Conselice et al ([25]), this could skew the results as there is not a large variation in the clumpiness values in the training dataset. This is evident as the maximal clumpiness value is of only 0.25 with a minimum value of 0.02 this close concentration of values could skew the final lines of best fit, especially if the relation between clumpiness and error is smaller than expected. This could also explain the anomaly found

in figures (13 and 14) as while the figures do seem to indicate some dependence on morphological structure, this could be distorted in the graphs by the small variety of clumpiness values which could cause a flattening of the best fit lines. This is particularly true for the training dataset, which contained only 99 galaxies, and hence some outliers could have shifted the slopes significantly.
Furthermore, the great majority of the data obtained was obtained utilising the training dataset, by definition the neural network has already been tested on these galaxies, hence it is possible that any initial difficulties due to the presence of high intensity clumps could have been overcome.

Another issue is the error of the CAS clumpiness parameter. The calculation of the error of the clumpiness parameters was done simply by calculating the resolution uncertainty, but there are several other sources of error in the clumpiness values which were not accounted for. For example, the size of the half light radius and its effect on the smoothed image. While these errors are almost impossible to account for when utilising the datasets used, if a dataset utilising a variety of filters of the same image were utilised, a bootstrap method could be implemented. Furthermore, in general the utilisation of different color filters, could allow for more precise CAS clumpiness parameter calculations to be made, as it would allow for a more precise mapping of the clumpiness of a galaxy.

The Sersic index on the other hand shows clearly that there is no relation, as there is a significant spread of the original values, and the only significant limitation on this index is that it's value in the metadata parameters is limited to a range within 0 and 6, hence it is impossible to see how extremely large Sersic index galaxies perform. The Sersic index itself is at best a weak indicator of the smoothness of a galaxy, and hence, it is unsurprising that no relation is found given that even a more refined parameter such as the CAS clumpiness parameter was unable to show any sort of relation between the values. It could still prove useful to utilise a more varied dataset with more established Sersic indices in order to conclusively say that no evident relation between morphological structure and network performance is found.

In future tests of relations of this sort, perhaps it is more beneficial to utilise more advanced classification systems, which take in to account more elements than the two parameters utilised. One example of what could be utilised is neural networks trained to classify galaxies, for example that developed by S.Goutam et al [51].

In essence, almost all results indicate that there is no relationship between galaxy morphology and performance of the neural network. Some results, specifically those found in figure (11), figure (16) and figures (13 and 14) indicate there could potentially be a relation between network performance and galaxy morphology, however, the data is not enough and more research is needed.

6 Conclusion

6.1 Summary of Main Contributions

In conclusion, through the simulations carried out in this thesis, it was determined that no clear relation exists between galaxy morphology and performance of the neural network developed by Wagner-Cerena et al [1]. A possible relation between galaxy morphology and combined error was observed in image plots, although this was not confirmed by best fit lines or image plots. The validation dataset, containing unseen galaxies, showed potentially stronger relationships than those in the training configuration, however, due to the small amount of galaxies no clear relation.

The original dataset utilised in the paper by Wagner-Cerena et al[1] was unavailable, and hence, the data obtained throughout this thesis was obtained utilising a dataset generated utilising the training configuration with over $11 \cdot 10^3$generated images and a dataset generated utilising the validation configuration with over $5 \cdot 10^3$ images.

6.2 Future Work

Further testing of the relationship between the galaxy structure and performance of the neural network could be made utilising both the original training dataset, in order to utilise the same data as the original paper. Additionally, tests could be carried out utilising a vaster range of galaxies, with more differing CAS clumpiness values and Sersic index values, to further explore the relationship, and established whether the anomaly observed in images (13) and (14), represents a real relation between performance and galaxy morphology.

Furthermore, a more efficient method of noise reduction in the original images could be utilised, hence allowing to properly establish whether the galaxy shape is responsible for errors, as this could also allow for more precise CAS clumpiness values to be obtained.

A more advanced method of galaxy classification could also be utilised, for example recently developed neural networks [51] could be utilised to classify galaxy morphologies more accurately. This would allow for a more precise and definite relationship between network performance and galaxy morphology to be established

Bibliography

[1] S. Wagner-Carena, J. Aalbers, S. Birrer, E. O. Nadler, E. Darragh-Ford, P. J. Marshall, and R. H. Wechsler, "From images to dark matter: End-to-end inference of substructure from hundreds of strong gravitational lenses," *The Astrophysical Journal*, vol. 942, p. 75, Jan. 2023.

[2] S. Dodelson and F. Schmidt, "1 - the concordance model of cosmology," in *Modern Cosmology (Second Edition)* (S. Dodelson and F. Schmidt, eds.), pp. 1–19, Academic Press, second edition ed., 2021.

[3] I. Debono and G. F. Smoot, "General relativity and cosmology: Unsolved questions and future directions," *Universe*, vol. 2, no. 4, 2016.

[4] J. F. Navarro, C. S. Frenk, and S. D. M. White, "The structure of cold dark matter halos," *The Astrophysical Journal*, vol. 462, p. 563, May 1996.

[5] J. S. Bullock and M. Boylan-Kolchin, "Small-scale challenges to the cdm paradigm," *Annual Review of Astronomy and Astrophysics*, vol. 55, p. 343–387, Aug. 2017.

[6] P. Bode, J. P. Ostriker, and N. Turok, "Halo formation in warm dark matter models," *The Astrophysical Journal*, vol. 556, p. 93–107, July 2001.

[7] A. V. Macciò and F. Fontanot, "How cold is dark matter? Constraints from Milky Way satellites," *Monthly Notices of the Royal Astronomical Society: Letters*, vol. 404, pp. L16–L20, 05 2010.

[8] E. D. Jahn, L. V. Sales, A. Wetzel, M. Boylan-Kolchin, T. K. Chan, K. El-Badry, A. Lazar, and J. S. Bullock, "Dark and luminous satellites of LMC-mass galaxies in the FIRE simulations," *Monthly Notices of the Royal Astronomical Society*, vol. 489, pp. 5348–5364, 09 2019.

[9] A. Arbey and F. Mahmoudi, "Dark matter and the early universe: A review," *Progress in Particle and Nuclear Physics*, vol. 119, p. 103865, July 2021.

[10] S. Vegetti, G. Despali, M. R. Lovell, and W. Enzi, "Constraining sterile neutrino cosmologies with strong gravitational lensing observations at redshift z 0.2," *Monthly Notices of the Royal Astronomical Society*, vol. 481, pp. 3661–3669, 09 2018.

[11] M. Bartelmann, "Gravitational lensing," *Classical and Quantum Gravity*, vol. 27, p. 233001, Nov. 2010.

[12] T. E. Collett, "The population of galaxy–galaxy strong lenses in forthcoming optical imaging surveys," *The Astrophysical Journal*, vol. 811, p. 20, Sept. 2015.

[13] N. Dalal and C. S. Kochanek, "Direct detection of cold dark matter substructure," *The Astrophysical Journal*, vol. 572, p. 25–33, June 2002.

[14] S. Mao and P. Schneider, "Evidence for substructure in lens galaxies?," *Monthly Notices of the Royal Astronomical Society*, vol. 295, pp. 587–594, 04 1998.

[15] D. Gilman, S. Birrer, A. Nierenberg, T. Treu, X. Du, and A. Benson, "Warm dark matter chills out: constraints on the halo mass function and the free-streaming length of dark matter with eight quadruple-image strong gravitational lenses," *Monthly Notices of the Royal Astronomical Society*, vol. 491, p. 6077–6101, Dec. 2019.

[16] A. Tsang, A. Çağan Şengül, and C. Dvorkin, "Substructure detection in realistic strong lensing systems with machine learning," 2024.

[17] S. Wagner-Carena, J. W. Park, S. Birrer, P. J. Marshall, A. Roodman, and R. H. Wechsler, "Hierarchical inference with bayesian neural networks: An application to strong gravitational lensing," *The Astrophysical Journal*, vol. 909, p. 187, Mar. 2021.

[18] Y. Hezaveh, N. Dalal, G. Holder, T. Kisner, M. Kuhlen, and L. P. Levasseur, "Measuring the power spectrum of dark matter substructure using strong gravitational lensing," *Journal of Cosmology and Astroparticle Physics*, vol. 2016, p. 048–048, Nov. 2016.

[19] S. Birrer, C. Welschen, A. Amara, and A. Refregier, "Line-of-sight effects in strong lensing: putting theory into practice," *Journal of Cosmology and Astroparticle Physics*, vol. 2017, p. 049–049, Apr. 2017.

[20] D. B. Rubin, "Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician," *The Annals of Statistics*, vol. 12, no. 4, pp. 1151 – 1172, 1984.

[21] G. Papamakarios, D. C. Sterratt, and I. Murray, "Sequential neural likelihood: Fast likelihood-free inference with autoregressive flows," 2019.

[22] J.-M. Lueckmann, P. J. Goncalves, G. Bassetto, K. Öcal, M. Nonnenmacher, and J. H. Macke, "Flexible statistical inference for mechanistic models of neural dynamics," 2017.

[23] S. Mohamed and B. Lakshminarayanan, "Learning in implicit generative models," 2017.

[24] J. Pearson, J. Maresca, N. Li, and S. Dye, "Strong lens modelling: comparing and combining Bayesian neural networks and parametric profile fitting," *Monthly Notices of the Royal Astronomical Society*, vol. 505, pp. 4362–4382, 05 2021.

[25] C. J. Conselice, "The relationship between stellar light distributions of galaxies and their formation histories," *The Astrophysical Journal Supplement Series*, vol. 147, p. 1–28, July 2003.

[26] D. M. Elmegreen and B. G. Elmegreen, "The onset of spiral structure in the universe," *The Astrophysical Journal*, vol. 781, p. 11, Dec. 2013.

[27] N. Tessore and R. Benton Metcalf, "The elliptical power law profile lens," *Astronomy amp; Astrophysics*, vol. 580, p. A79, Aug. 2015.

[28] R. Kormann, P. Schneider, and M. Bartelmann, "Isothermal elliptical gravitational lens models," *Astronomy and Astrophysics*, vol. 284, pp. 285–299, 03 1994.

[29] C. R. Keeton, C. S. Kochanek, and U. Seljak, "Shear and ellipticity in gravitational lenses," *The Astrophysical Journal*, vol. 482, p. 604–620, June 1997.

[30] M. White, "The mass of a halo," *Astronomy amp; Astrophysics*, vol. 367, p. 27–32, Feb. 2001.

[31] E. A. Baltz, P. Marshall, and M. Oguri, "Analytic models of plausible gravitational lens potentials," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, p. 015–015, Jan. 2009.

[32] D. Gilman, X. Du, A. Benson, S. Birrer, A. Nierenberg, and T. Treu, "Constraints on the mass–concentration relation of cold dark matter haloes with 11 strong gravitational lenses," *Monthly Notices of the Royal Astronomical Society: Letters*, vol. 492, pp. L12–L16, 11 2019.

[33] S. Vegetti, L. V. E. Koopmans, A. Bolton, T. Treu, and R. Gavazzi, "Detection of a dark substructure through gravitational imaging," *Monthly Notices of the Royal Astronomical Society*, vol. 408, pp. 1969–1981, 10 2010.

[34] R. K. Sheth, H. Mo, and G. Tormen, "Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes," *Monthly Notices of the Royal Astronomical Society*, vol. 323, no. 1, pp. 1–12, 2001.

[35] L. Perreault Levasseur, Y. D. Hezaveh, and R. H. Wechsler, "Uncertainties in parameters estimated with neural networks: Application to strong gravitational lensing," *The Astrophysical Journal Letters*, vol. 850, p. L7, Nov. 2017.

[36] A. M. Koekemoer, H. Aussel, D. Calzetti, P. Capak, M. Giavalisco, J. Kneib, A. Leauthaud, O. Le Fevre, H. J. McCracken, R. Massey, B. Mobasher, J. Rhodes, N. Scoville, and P. L. Shopbell, "The cosmos survey: Hubble space telescope advanced camera for surveys observations and data processing," *The Astrophysical Journal Supplement Series*, vol. 172, p. 196–202, Sept. 2007.

[37] R. Mandelbaum, C. M. Hirata, A. Leauthaud, R. J. Massey, and J. Rhodes, "Precision simulation of ground-based lensing data using observations from space: Ground-based image simulation," *Monthly Notices of the Royal Astronomical Society*, vol. 420, p. 1518–1540, Dec. 2011.

[38] B. Rowe, M. Jarvis, R. Mandelbaum, G. M. Bernstein, J. Bosch, M. Simet, J. E. Meyers, T. Kacprzak, R. Nakajima, J. Zuntz, H. Miyatake, J. P. Dietrich, R. Armstrong, P. Melchior, and M. S. S. Gill, "Galsim: The modular galaxy image simulation toolkit," 2015.

[39] E. P. Hubble, "Extragalactic nebulae.," , vol. 64, pp. 321–369, Dec. 1926.

[40] G. De Vaucouleurs, *Classification and Morphology of External Galaxies*, pp. 275–310. Berlin, Heidelberg: Springer Berlin Heidelberg, 1959.

[41] K. L. Masters, C. J. Lintott, R. E. Hart, S. J. Kruk, R. J. Smethurst, K. V. Casteels, W. C. Keel, B. D. Simmons, D. O. Stanescu, J. Tate, and S. Tomi, "Galaxy Zoo: unwinding the winding problem – observations of spiral bulge prominence and arm pitch angles suggest local spiral galaxies are winding," *Monthly Notices of the Royal Astronomical Society*, vol. 487, pp. 1808–1820, 04 2019.

[42] J. L. Sérsic, "Influence of the atmospheric and instrumental dispersion on the brightness distribution in a galaxy," *Boletin de la Asociacion Argentina de Astronomia La Plata Argentina*, vol. 6, pp. 41–43, Feb. 1963.

[43] K. M. Hambleton, B. K. Gibson, C. B. Brook, G. S. Stinson, C. J. Conselice, J. Bailin, H. Couchman, and J. Wadsley, "Advanced morphological galaxy classification: a comparison of observed and simulated galaxies," *Monthly Notices of the Royal Astronomical Society*, vol. 418, pp. 801–810, 11 2011.

[44] G. Djorgovski and H. Spinrad, "Toward the application of metric size function in galactic evolution and cosmology," *The Astrophysical Journal*, vol. 251, pp. 417–423, 11 1981.

[45] A. W. Graham, *Elliptical and Disk Galaxy Structure and Modern Scaling Laws*, p. 91–139. Springer Netherlands, 2013.

[46] M. Vika, B. Vulcani, S. P. Bamford, B. Häußler, and A. L. Rojas, "Megamorph: classifying galaxy morphology using multi-wavelength sérsic profile fits," *Astronomy amp; Astrophysics*, vol. 577, p. A97, May 2015.

[47] N. Caon, M. Capaccioli, and M. D'Onofrio, "On the shape of the light profiles of early-type galaxies," *Monthly Notices of the Royal Astronomical Society*, vol. 265, p. 1013–1021, Dec. 1993.

[48] G. Di Leo and F. Sardanelli, "Statistical significance: p value, 0.05 threshold, and applications to radiomics—reasons for a conservative approach," *European Radiology Experimental*, vol. 4, p. 18, Mar 2020.

[49] P. Schober, C. Boer, and L. A. Schwarte, "Correlation coefficients: Appropriate use and interpretation," *Anesthesia & Analgesia*, vol. 126, no. 5, 2018.

[50] A. L. Bowley, "The standard deviation of the correlation coefficient," *Journal of the American Statistical Association*, vol. 23, no. 161, pp. 31–34, 1928.

[51] G. Sarker, "A new method of galaxy classification using optimal convolution neural network," in *Artificial Intelligence* (A. A. Sk, T. Turki, T. K. Ghosh, S. Joardar, and S. Barman, eds.), (Cham), pp. 284–293, Springer Nature Switzerland, 2022.

[52] P. J. E. Peebles, *The large-scale structure of the universe*. 1980.

Appendices

A Additional Graphs and results

In addition to the data included in the Results section, some additional results were collected. These consists entirely of the coefficient values(Pearson and Spearman) for the graphs plotted in the main results section. These values were excluded from the main paper as the p-values were large enough that these coefficients have no statistical significance. In essence the p-values alone are enough to show that the Pearson and Spearman fits showed no correlation between the variables tested.

A.1 Bias training data

The pearson and spearman coefficient values for the plots were:

$$K_{Pearson} = 0.02 \pm 0.02, \qquad K_{Spearman} = 0.02 \pm 0.02 \tag{28}$$

$$K_{Pearson} = 0 \pm 0.02, \qquad K_{Spearman} = 0 \pm 0.02 \tag{29}$$

The errors in the values were significantly higher than the values themselves hence the coefficients are essentially meaningless, showing only that there is no relation at all.
Similarly the clumpiness index plots showed values equal to:

$$K_{Pearson} = -0.02 \pm 0.02, \qquad K_{Spearman} = 0 \pm 0.02 \tag{30}$$

$$K_{Pearson} = 0 \pm 0.02, \qquad K_{Spearman} = 0 \pm 0.02 \tag{31}$$

A.2 Spread training data

Once again, the correlation coefficients yield values smaller than the errors both for the two CAS index graphs. Specifically for the systematic error:

$$K_{Pearson} = -0.02 \pm 0.02, \qquad K_{Spearman} = 0 \pm 0.02 \tag{32}$$

The Sersic index values were essentially the same, again highlighting no clear relation:

$$K_{Pearson} = 0.01 \pm 0.02, \qquad K_{Spearman} = 0 \pm 0.02 \tag{33}$$

A.3 Bias validation data

As with the training dataset, the correlation coefficients contain errors as large as the coefficients themselves, indicating that there is no relation at all between the values. For the clumpiness index:

$$K_{Pearson} = 0 \pm 0.1, \qquad K_{Spearman} = 0.1 \pm 0.1 \tag{34}$$

while for the sersic indices:

$$K_{Pearson} = -0.2 \pm 0.1, \qquad K_{Spearman} = -0.2 \pm 0.1 \tag{35}$$

A.4 Spread validation data

The Random error data from the validation dataset also had correlation coefficient values which were essentially insignificant for the CAS clumpiness index:

$$K_{Pearson} = -0.1 \pm 0.1, \qquad K_{Spearman} = 0 \pm 0.1 \tag{36}$$

Once again the same was true for the sersic index plots:

$$K_{Pearson} = 0.1 \pm 0.1, \qquad K_{Spearman} = 0 \pm 0.1 \tag{37}$$

B Statistical tools

The relations needed to express the concentration and radius variables required in the NFW function are given by[32][1]:

$$c_{\mathrm{sub}}(m,z) = c_0(1+z)^{\zeta}\left(\frac{\nu(r_{\mathrm{peak}}(m_{\mathrm{sub}}),z_{\mathrm{sub}})}{\nu(r_{\mathrm{peak}}(m_{\mathrm{pivot,conc}}),0)}\right)^{-\beta}, \tag{38}$$

$$r_{\mathrm{peak}}(m_{\mathrm{sub}}) = \left(\frac{3m_{\mathrm{sub}}}{4\pi\rho_{m,0}}\right)^{1/3}. \tag{39}$$

In the first equation, the concentration c is defined by the normalization factor c_0, ζ a parameter characterising the redshift power-law function, β is the value of the peak of this function and the ν variable is the peak height function. In the second equation, the variable $\rho_{m,0}$ is the matter density at redshift zero. To determine the peak height function the method utilised in the paper by Peebels et al[52].