University of La Laguna

Investigating the morphologies of slow and fast rotators with deep learning

Author:
Alejandro López Morales

Supervisor:
Marc Huertas-Company

September, 2020

Abstract

The use of artificial intelligence and more precisely of Convolutional Neural Networks for the morphological classification of galaxies has been shown to be extremely reliable. In the last years, detailed kinematic studies of local early-type galaxies have revealed that a significant fraction of them seem to be rotationally supported. However, they seem to present no morphological signatures in their stellar morphology. In this work we apply machine learning to investigate the relation between kinematics and morphology of early-type galaxies. We use to that purpose SDSS imaging and spectroscopy from Manga together with simulated galaxies from the TNG100 cosmological simulation. We show that Neural Networks can distinguish fast and slow rotating early-type galaxies with 78% accuracy based exclusively on their stellar morphology even if no apparent signatures are visible to the human eye.

Acknowledgements

I am grateful to my supervisor Marc Huertas-Company for his help and availability even in times of lockdown, when contact and updating was much more complicated. I also thank my fellow masters Fernando, Alberto, Xavier and Manolo for their help and coworking throughout the course and the period of work on the final master's project.

Contents

1 Resumen en español **3**

2 Introduction **4**

3 Galaxy classification **6**

3.1 Morphology 6

3.2 Kinematics 7

4 Methods **9**

4.1 Data used in this work 9

4.1.1 Observations: Manga 9

4.1.2 Simulations: TNG100 9

4.2 Neural networks 10

4.3 Convolutional Neural Networks (CNNs) 11

4.3.1 Data augmentation 12

4.4 Evaluation tools 12

4.5 Interpretating the network decisions: Integrated gradients 14

5 Results **14**

5.1 Observations: MaNGA 14

5.1.1 Slow-Fast rotator, all galaxies 15

5.1.2 Slow-Fast rotator, early-type data 19

5.1.3 Integrated gradients 20

5.2 TNG100 (Illustris) 24

5.2.1 Integrated gradients 25

6 Model optimization and performance improvement **30**

6.1 Regularization methods 30

6.2 Observations 31

6.2.1 Color 31

6.2.2 Normalization 31
6.2.3 Image size 33
6.3 Simulations (No noise) 35
6.3.1 Normalization 35
6.3.2 Size 35

7 Conclusion and future work **39**

1 Resumen en español

Las galaxias se han clasificado históricamente según su aspecto visual, es decir su morfología, llegando así a subdivisiones como las de galaxias de tipo temprano (galaxias elípticas y lenticulares dominadas por un bulbo, población estelar vieja y poco gas) y tipo tardío (galaxias espirales e irregulares con masas menores y alta formación estelar). En la época moderna se han desarrollado métodos para clasificarlas según su dinámica, produciéndose así una división entre sistemas soportados por rotación o dispersión. Uno de los resultados importantes de estos estudios, es que una fracción importante de galaxias de tipo temprano presentan altas componentes de rotación en sus centros, lo que ha motivado llevar a cabo una división entre rotadores rápidos y lentos dentro de esta familia de galaxias. Los parámetros usados para llevar a cabo esta separación son datos de la velocidad y dispersión de las estrellas obtenidos por espectrógrafos de campo integral, describiendo parámetros como V/σ o λ_R que dependen de los parámetros comentados. En estudios posteriores se ha llegado a la conclusión de que no hay manera (al menos hasta ahora) descubierta de poder llevar a cabo esta clasificación de galaxias de tipo temprano según su rotación usando información visual, es decir morfología de las galaxias, ya que con una inspección visual no hay diferencia aparente que nos haga saber el tipo de rotación que tienen. En este punto es donde tiene interés el planteamiento de usar inteligencia artificial para ver si es capaz de encontrar información morfológica visual que nosotros a simple vista no podemos. Para llevar a cabo este estudio hemos usado las redes neuronales convolucionales (CNNs), las cuales mediante convoluciones son capaces de extraer características de imágenes y asociarlas a una clase, en este caso rotadores rápidos y lentos. La manera de diagnosticar qué tan buenos son nuestros resultados será mediante las curvas ROC, que nos dan el desempeño del sistema a diferentes rangos de discriminación, es decir, cuando variamos a partir de qué valor consideramos una clase u otra; también de esta curva se obtiene el AUC (aérea debajo de la curva), que es algo como la probabilidad de la red de hacer una clasificación correcta, que usaré llamándola como porcentaje de acierto. También usaremos un método bastante eficaz llamado gradientes integrados, de los que se obtienen los llamados mapas de atributos que se usan para ver qué pixeles tiene más en cuenta la red neuronal para llegar a asociar una clase u otra. Para este estudio usaremos observaciones procedentes del survey de MaNGA y simulaciones procedentes del TNG100, permitiéndonos así ver si la red trata y aprende de manera semejante con ambas bases de datos.

Los primeros resultados se obtienen usando datos de observaciones de galaxias de tipo tardío y temprano, siendo este un problema sencillo para la red ya que prácticamente todas las galaxias de tipo tardío son rotadores rápidos, obteniendo así un porcentaje del 92% de aciertos. Cuando se usan estos datos pero solo con las galaxias de tipo temprano el porcentaje de acierto disminuye hasta un 78%. Si se analizan las imágenes de las galaxias que se clasifican correcta y erróneamente vemos que galaxias con mayor cantidad de elementos en imagen como fuentes de luz o distribuciones de valores de pixeles no comunes son aquellas que se clasifican erróneamente, y en cambio las que se clasifican correctamente vemos que presentan apenas diferencia para el tipo de rotador, siendo la elipticidad la única característica que aparentemente diferencia a los rotadores lentos de los rápidos, siendo mayor para estos últimos. Si estudiamos un diagrama $\lambda_R - \epsilon$ se ve que la elipticidad de las galaxias tiene un papel importante a la hora de clasificar erróneamente galaxias con rotación rápida en lentas. Si en este caso se estudian los píxeles en los que la red se centra más vemos que trabaja de manera muy óptima, no teniendo en cuenta el ruido ni fuentes de luz cercanas al objeto de estudio y fijándose sobretodo en el bulbo galáctico.

Cuando usamos datos de simulaciones con ruido (para que sean lo más similares a las observaciones del SDSS) obtenemos un porcentaje del 70 %, un valor que es sensiblemente inferior al de las observaciones y por tanto es necesario saber por qué. Cuando estudiamos la dependencia de las clasificaciones con la elipticidad obtenemos un resultado semejante al caso anterior en el caso del diagrama $\lambda_R - \epsilon$, pero a la hora de hacer un análisis visual no queda clara la dependencia. Es cuando se analizan los gradientes integrados cuando se ve algo que puede dar una explicación a estos resultados; a diferencia de las observaciones, la red a parte de fijarse en el bulbo de la galaxia tiene en cuenta también características a su alrededor. Una manera de dar una explicación mas

clara es haciendo un estudio con las galaxias simuladas pero esta vez sin ruido. En efecto, los primeros resultados ya mejoran algo y arrojan una posibilidad de acierto del 72 % y de analizar los mapas de los gradientes integrados se deriva en efecto que la red se fija, al igual que la de las observaciones en los bulbos de las galaxias, teniendo las estructuras externas de las galaxias un peso mucho menor.

Una vez obtenidos estos resultados preliminares es de gran utilidad hacer "experimentos" en los que se estudian que factores pueden hacer variar nuestros resultados. Como si la información de color tiene importancia; el resultado obtenido para filtros rgi y r del SDSS son prácticamente similares por lo que no es un parámetro de suma importancia. La normalización de los datos también se ha demostrado que es un tema muy importante a la hora de que la red pueda obtener resultados adecuados, pudiendo llegar a tener un rango de valores dede el 78 % hasta prácticamente el 50 % para diferentes normalizaciones usando los mismos datos y parámetros en la red. Por otra parte, si hacemos un zoom al núcleo de la galaxia, que es donde la rotación debería ser más relevante, ¿el resultado será mejor o peor?, los resultados nos demuestran que para el caso de las observaciones prácticamente no varía el resultado pero para las simulaciones pasa de un AUC de 0.72 a 0.79, lo que las iguala las predicciones obtenidas con las observaciones.

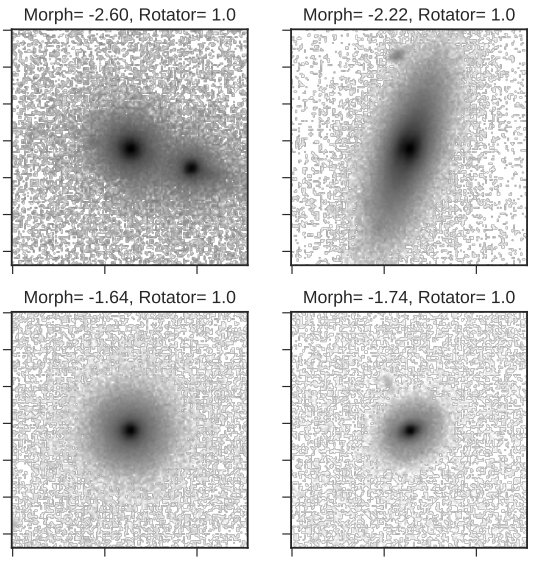
En conclusión, independientemente de observaciones o simulaciones, vemos que la red neuronal es capaz de encontrar algo de información en la morfología y por lo tanto puede clasificar correctamente algunas galaxias, pero está lejos de obtener resultados notables, por lo que es necesario continuar con el estudio, aumentando la cantidad de datos y casos.

2 Introduction

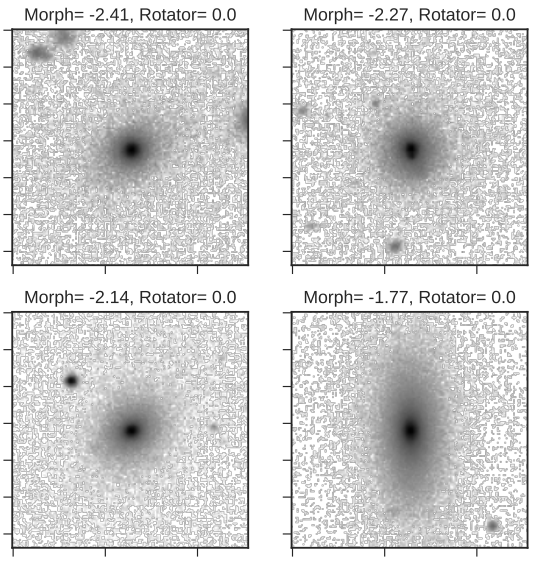
When galaxies were first discoveres, they were classified bases on their visual appearance, obtaining in the 1920s the so-called Hubble sequence (Hubble, 1926), where galaxies are divided into different types according to their morphological appearance and without taking into account any other characteristics. Nowadays this classification is still used, but advances in observation and the theory of galactic evolution have made us think that the Hubble sequence might not be the most suitable. Deep muilti-object spectroscopy and more recently integral field spectroscopy has enabled to study the kinematics of stars in statistical samples of galaxies. One key result of these first studies was that a significant fraction of visual early-type galaxies presented large fractions of rotation (Emsellem et al., 2011) and was possible to create a new classification: the fast and slow rotator galaxies which would be characterized by the measures of the velocity and velocity dispersion, obtaining a new parameter called the spin (Emsellem et al., 2007), (Emsellem et al., 2011) that will be described in the next sections. Several studies like Krajnović et al. (2013) have shown that there is no strong correlation between the morphology and the motion for the early-type galaxies. Although recent studies have shown that there exists noticeable differences in the stellar populations of slow and fast rotators (Bernardi et al. (2019)), as can be seen in Figure 1 (where Rotator type = 0 are slow rotators and Rotator type = 1 are fast rotators and for Morph < 2 are early-type galaxies and Morph > 2 are late-type galaxies) in which the morphology in early-types galaxies does not change with the rotator type.

The continuous increase in surveys of galaxies has allowed us to make much more accurate classifications and has opened up the possibility of being able to use big data tools such as artificial intelligence to carry out studies in many fields of research, such as astrophysics, where we have a lot of data, thanks to surveys such as those of the SDSS (Sloan Digital Sky Survey) and have been demonstrated that is a very efficient tool to estimate visual features from the morphology of galaxies (Dieleman et al. (2015b); Huertas-Company et al. (2015a); Domínguez Sánchez et al. (2018))

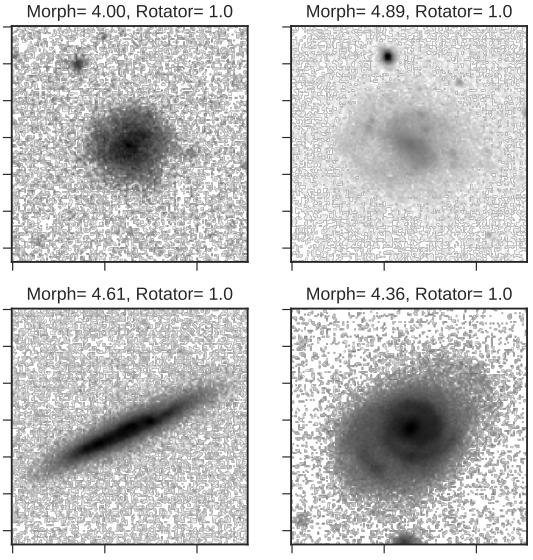
This project aims to use artificial intelligence to investigate if there is any link between kine-



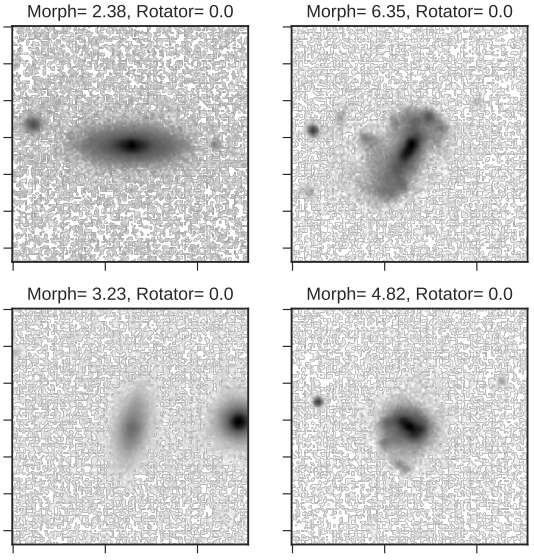
(a) Early-type fast rotator



(b) Early-type slow rotator



(c) Late-type fast rotator



(d) Late-type slow rotator

Figure 1: Examples of early and late type galaxies classified as slow and fast rotators. early type galaxies do not represent any morphological quality visible at least to the human eye that would allow us to know what type of rotator it is. On the other hand, in the late type galaxies it is possible to distinguish relevant characteristics, like that the fast rotators have well defined disks, some of them even with structures like arms, common qualities of rotation supported galaxies and the slow rotators are mostly irregular galaxies that dont have well structured disks.

matics and morphology in early-type systems. The question we are going to ask is: can a neural network distinguish between slow and fast rotators based only on the stellar morphology?

3 Galaxy classification

In this work we want therefore to solve a classification problem, that is to find a correlation between morphological and kinematic classification, so it is necessary to know the parameters that lead to these divisions.

3.1 Morphology

The morphological type describes the visual appearance of a galaxy. The basis of the morphological classification was made by Hubble (Hubble, 1926) and is represented in the so-called the Hubble sequence (figure 2).

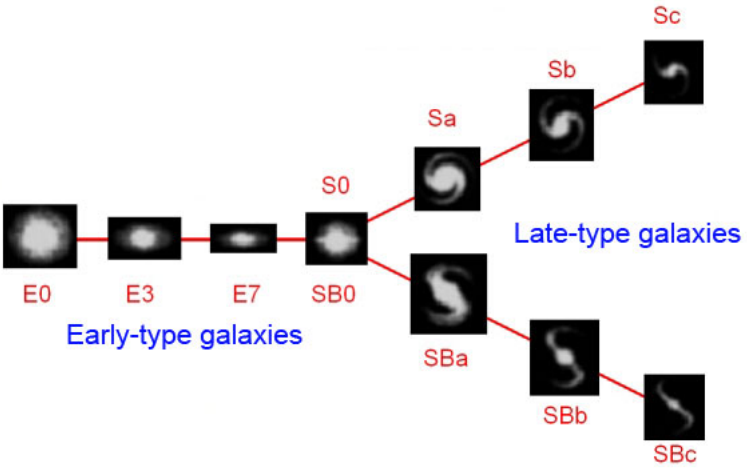


Figure 2: Modern representation of the Hubble sequence with the classification between early and late type galaxies. The early-type galaxies do not have defined disks in which structures like arms can be distinguished; we will assume that the lenticular galaxies S0 form part of this group. In the other hand, late-type galaxies, have a well defined bulge that can be shaped as a bar, surrounded by a disc with well-defined arms

Early-type galaxies can be elliptical or lenticulars. Ellipticals present a smooth light distribution, are usually more massive, have older star population, and are gas-poor. Lenticulars tend to present a featureless disk around the dominant bulge. Late-type galaxies consist of a flattened disk, with stars forming typically a spiral structure (although not always visible due to resolution effects), and a central concentration of stars known as the bulge, which is similar in appearance to an elliptical galaxy. They typically have higher star formation rates than early-type galaxies and therefore bluer colors in broadband photometry. They are also gas-richer and less massive that the early-type ones.

The first way in which the morphological classification was carried out was through a merely visual analysis and there were several authors who contributed in this field such as Hubble (1926) or Morgan (1958). The finally established classification was the Hubble classification, that was later revised by de Vaucouleurs, 1959. The need of more quantitative morphologies pushsed researchers to look for parameters that correlate with morphology. The measurement of the central

concentration or asymmetry of the surface brightness profiles of galaxies (Abraham et al., 1996) are some examples of those widely used parameters. The Gini coefficient of the galaxy light distribution also correlates with morphology (Abraham et al., 2003). However, visual inspection has remained an important channel to classify galaxies. With the increase of samples sizes, approaches based on citizen science such at the the Galaxy Zoo emerged (Lintott et al., 2008). This allowed the community to build large training samples for modern artificial intelligence algorithms. In the past years, supervised deep learning has rapidly become the state-of-the art approach to galaxy morphology (Dieleman et al. (2015a); Huertas-Company et al. (2015b); Fischer et al. (2019)).

3.2 Kinematics

With the advent of efficient spectroscopic surveys, statistical samples of the dynamics of stars in galaxies became possible. This studies started in the 80s using new detectors for measuring velocity dispersions (Kormendy & Illingworth, 1982). In the 90s the CCD developing gave us more detailed data about the galaxy dynamics. Then the integral field spectrographs became the main tool for the study of galaxy kinematics, because of the spatial information it provides. Numerous surveys have been now carried out using this technique, such as SAURON (Bacon et al. (2001); de Zeeuw et al. (2002); Emsellem et al. (2004)), CALIFA (Sánchez et al., 2012) or MaNGA (Bundy et al., 2015) which were the first large surveys. All this technical and scientific advances produced a new division between galaxies based on their kinematics. We essentially distinguish galaxies supported by rotation from those supported by dispersion. The parameter V/σ, (where V is the radial velocity and σ the velocity dispersion), quantifies the relative amount of rotation in a galaxy. In the local universe, this parameter for early-type galaxies is roughly ranged from 0 to 1 due to the high velocity dispersion, but in late-type galaxies V/σ can take much higher values due to the low velocity dispersion, for example taking typical values of a disk where $V \approx 200$ km/s and $\sigma \approx 30$ km/s we get a $V/\sigma \approx 7$. This means that the higher is the V/σ the radial velocity os more relevant, so the system is supported by rotation.

It was Emsellem (Emsellem et al., 2011) who, using SAURON's data, concluded that the Hubble type cannot serve as a proxy for the kinematics of the early-type galaxies. Although early type galaxies have always been though to be supported by dispersion, a significant fraction of them still have rotation. A new kinematic parameter was introduced to quantify the total velocity structure using the spatial information given by the IFUs, which is given by the expression:

$$\lambda_R = \frac{\langle R|V|\rangle}{\langle R\sqrt{V^2+\sigma^2}\rangle} \tag{3.1}$$

where R is the distance to the galactic center. This expression allows us to distinguish galaxies that have close V/σ values, but different velocity structures, obtaining a quantitative reference. The galaxies are now classified into a slow-fast rotator depending on their λ_R values.

In this work, we will use the catalogs collected in (Fischer et al., 2019) bases on MaNGA. From these catalogs we get information about the spin parameter λ_R, its galaxy morphological type, and its axis ratio, which is b/a from we can define the parameter epsilon as $\epsilon = 1 - b/a$, also called as ellipticity. Here is where the threshold that will separate the region of rotators is defined as:

$$Y_{th} = 0.31\sqrt{1 - b/a} \tag{3.2}$$

The way to represent the division between rotator types is by plotting the Y_{th} in the plane λ_R-ϵ. This is shown in Figure 3 where is shown that almost all the late-type galaxies are classified as fast-rotators. However, early-type galaxies are more equally distributed among the two types of rotators.

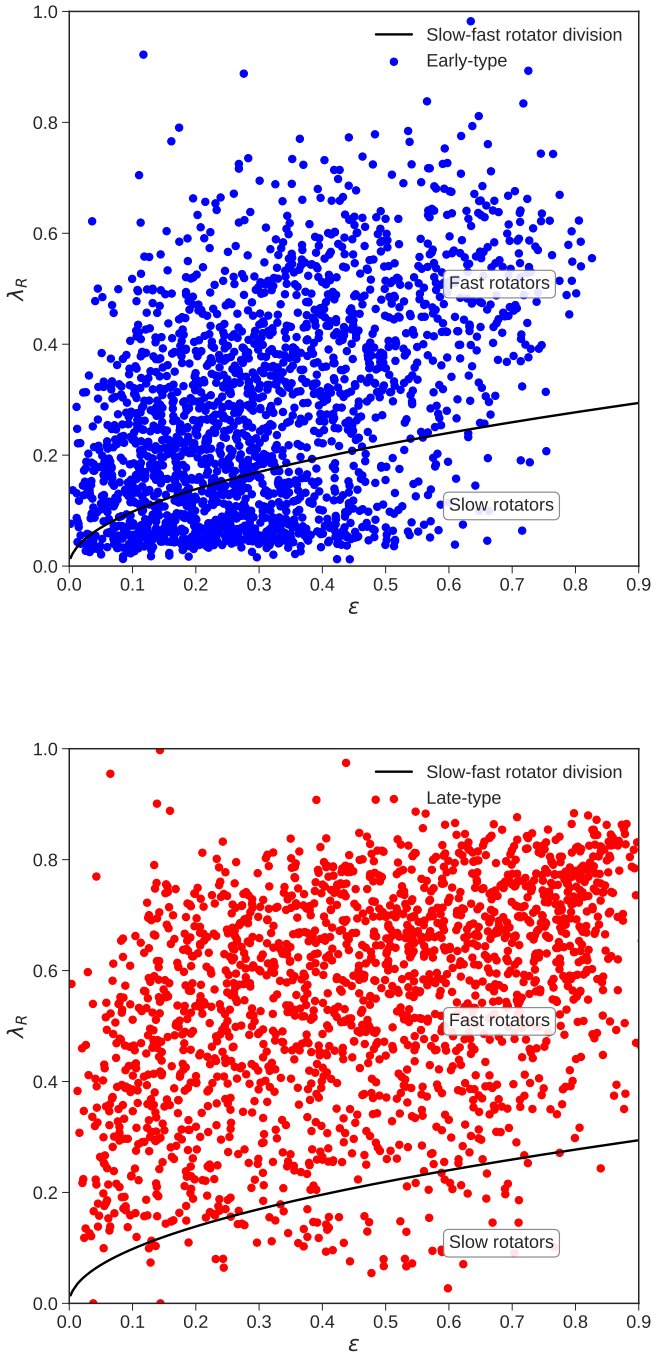


Figure 3: Fast-slow rotator distributions from the MaNGA sample for early and late-type galaxies used in this work. This is a typical $\lambda_R - \epsilon$ diagram where a variable dependent on velocity and velocity dispersion (λ_R) is represented against ellipticity (ϵ). Also is represented the curve Y_{th} that sets the division line between rotator types. In the first figure, with blue color the early type galaxies are represented, you can see that they are distributed all over the plane, taking λ_R and ϵ values in all the range, and being slightly limited in the value of ellipticity not exceeding 0.8. In the second figure, with red color is the distribution of late type, which are unequally distributed not like the early-type ones. We see that most is in the area of fast rotators and the values of ellipticity can take higher values.

4 Methods

4.1 Data used in this work

4.1.1 Observations: Manga

MaNGA, the abbreviation for the Mapping Nearby Galaxies at Apache Point Observatory is a survey obtained by the Sloan Digital Sky Survey (SDSS). MaNGA uses 17 simultaneous integral field units (IFUs) to map the spectra across 10000 nearby galaxies. The IFUs allows us to obtain very detailed information about the composition, velocity, dispersion velocity, being this measurements given by the catalogs (Fischer et al., 2019). In particular we will use the MaNGA Morphology Deep Learning DR15 catalog, a morphological catalog of MaNGA galaxies obtained with Deep learning models (Fischer et al., 2019) for MaNGA DR15 galaxies. The MaNGA DR15 catalogue provides photometric parameters obtained from Sersic and Sersic+Exponential fits to the 2D surface brightness profiles. The catalog correspond to 4672 PLATE-IFU entries including the g, r and i band. The estimate of the structural paremeters of the galaxies was made using a fitting algorithm called PyMorph (Vikram et al., 2010).

4.1.2 Simulations: TNG100

The IllustrisTNG project (Nelson et al. (2018); Pillepich et al. (2017); Naiman et al. (2018); Springel et al. (2017)) is an ongoing series of large, cosmological magnetohydrodynamical simulations of galaxy formation performed by the moving-mesh code AREPO (Springel (2010); Pakmor et al. (2011); Pakmor et al. (2015)). In this work we will use the version TNG100 which follows the evolution of 2 x 1820^3 resolution elements in a cube measuring $75h^{-1} \approx 100$ Mpc . We have used the snapshot 99 which is the one that has information about the kinematics at redshift=0, with ≈ 12000 galaxies. It is restricted to subhalos with stellar mass $M^* > 3.4$ x $10^8 M_\odot$. The initial conditions of the simulations (Pillepich et al., 2017) were set at z = 127 using the Zeldovich aproximation (Zel'Dovich (1970); Shandarin & Zeldovich (1989)). Using the cosmological parameters given by the matter density $\Omega_m = \Omega_{dm} + \Omega_b = 0.3089$ with $\Omega_b = 0.0486$, cosmological constant $\Omega \leq$ mbda 0.6911, Hubble constant $H_0 = 100h$ km s^{-1} with h = 0.6774.

In addition, in order to properly compare with observations, we create mock images from the simulations output. We use the radiative transfer code SKIRT to convert mass to light (Baes et al., 2011) and model dust emission. We include PSF and noise to match the SDSS S/N and resolution.

4.2 Neural networks

The basic unit of a neural network is the perceptron, which is a mathematical function that models the biological neurons. These artificial neurons have a structure like the one represented in Figure 4

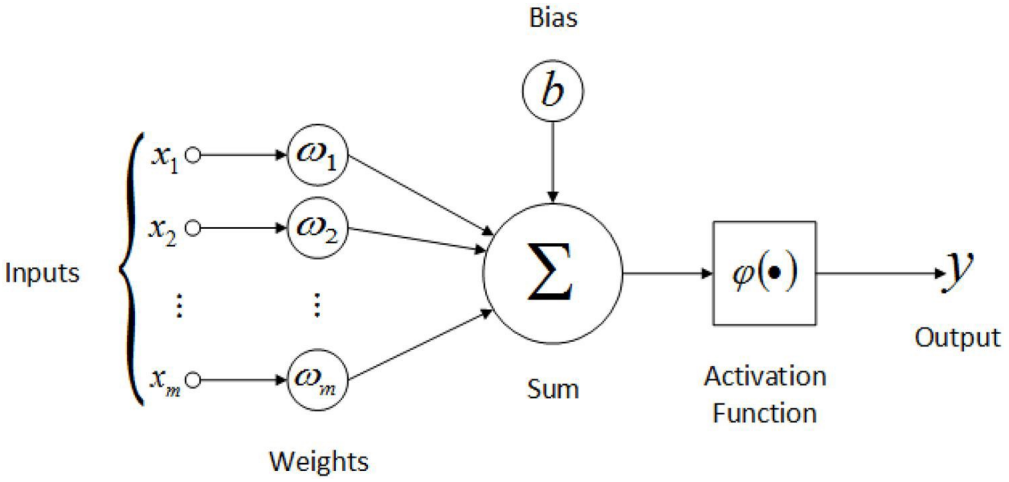


Figure 4: Artificial neuron diagram (Bursac et al., 2019). There are several inputs to which a certain weight is associated. The weighted sum of all these values is made, a bias is added and an activation function is applied.

Neurons have inputs $x_1, x_2, ..., x_m$ that may have been generated by the output of a previous neuron. Each input is then multiplied by its weight $w_1, w_2, ..., w_m$, which is the parameter that the model will adjust to learning. These results are added up and a Bias is applied. The resulting expression is passed through an activation function φ, which is used to add non-linearity to the neural networks. In our case we use the "sigmoid" function, commonly used for classification problems. It is defined as:

$$\varphi(x) = \frac{1}{1 + e^{-x}}$$

, and defined in the range $(0, 1)$. The output of the perceptron is therefore given by:

$$y = \varphi(Bias + \sum_{j=0}^{m} w_j x_j) \tag{4.1}$$

A set of neurons form a network that works by establishing initial weights, predicting output, and comparing it with an expected value. The error obtained adjusts the weights for the next iteration and is repeated until the error reaches a minimum. This means that a neural network solves a minimization problem. This is done by gradient descent since NNs are by definition differentiable. The algorithm that performs this efficiently is called backpropagation (Lecun, 1988).

An artificial neural network consists of an input layer, hidden layers, and an output layer formed by neurons. In this project, we will use convolutional neural networks (CNN), which are a type of deep neural network (DNN) i.e. a model that learn features from the data , that is also a type of artificial neural network (ANN).

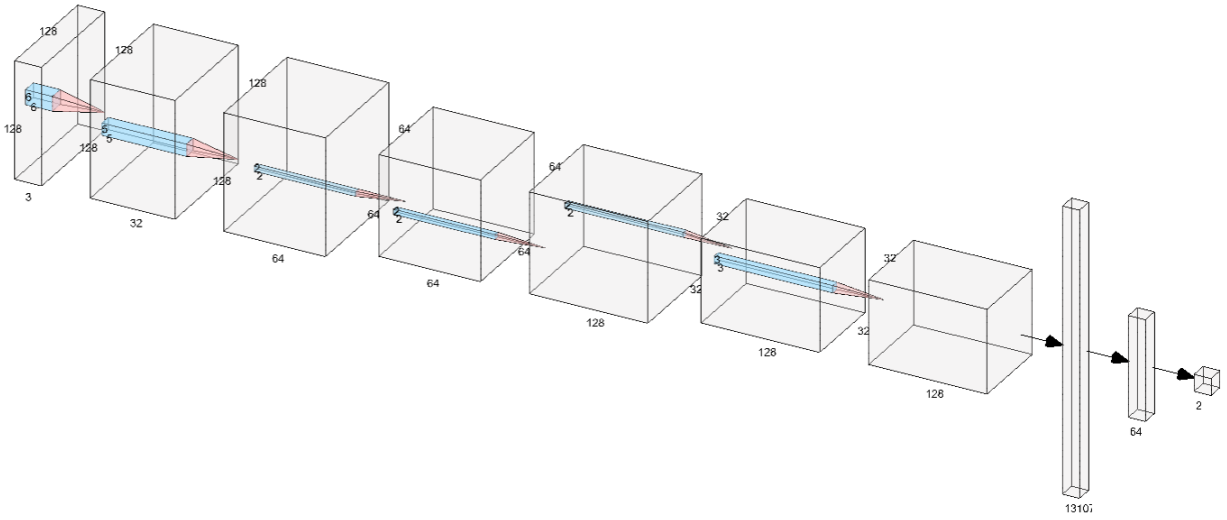


Figure 5: CNN structure that we will use where are represented the width, height and depth of each layer of the CNN. There are cubes in which the depth increases but not the height and width, these are the convolutions. When the height and width decreases but not the depth is that our data have passed the layer of pooling, finally reaching the flatten layers that put all the accumulated values in those data cubes in a one-dimensional array of 131072 and a dense layer that compresses all these data in an even smaller array with a size of 64, finally applying a sigmoid function that will allow us to make the binary classification

4.3 Convolutional Neural Networks (CNNs)

The CNNs are neural networks applied to data with spatial correlations such as images. They are capable of automatically finding features in the images which are by definition transnational invariant. To do so, the images are filtered with a variety of different filters. The properties of the filters are encoded in the weights of the neural network and are therefore learned.

When we work with CNNs, we use a particular data structure, it is a 4-D tensor where we have information on the image number, the size of the image, and the filter. The convolutional neural networks are composed by:

- Input layer: An image with a defined size with a depth given by the filters
- Hidden layers:
  - Convolutions: will process the output of neurons that are connected in input "local regions", i.e. nearby pixels, calculating the product scaling between their weights (pixel value) and a small region to which they are connected in the input volume extracting feature maps by varying the dimensions of the input tensor
  - Pooling: reduce the dimensions of the data by combining pixels. This is where we filter the image with a learnable kernel.
- Output layers: flat the data, converting the tensor on a 1-D array, then this array can be reduced using the dense layers, ending with the application of an activation function, in our case the called "sigmoid".

This structure is shown in Figure 5, which represents the network model I used in this project and which I will detail below.

The input layer will be different depending on the data we use, in Figure 5 a 128 x 128 pixel image with three filters is assumed as input (basically throughout the project this will be the standard size and the number of filters may vary). After the input layer, we have our first convolution, which gives us a depth of 32 (i.e. we filter the image with 32 different filters of size 128x128), and later we have another convolution that increases it again to 64. By having so many convolutions, what we will do is make the network capable of distinguishing more and more complex patterns. Next, we apply a max-pooling that reduces the size of the image to 64x64, this helps to reduce the computational cost and accumulate the characteristics of the maps generated by the previous convolutions; it also helps capturing larger scale correlations in the image. After this, another convolution is applied followed by another max-pooling and then another convolution, getting a tensor with a depth of 128 with images of 32 x 32 pixels. It is at this point when we enter the fully connected layers (that are normal ANNs, who are in charge of carrying out the classification), where we can use the information obtained in the convolutional layers to classify the image in a label, the first step is to apply a flatten layer that will convert the tensor we had into a vector with a size of 131072, a dense layer will be applied converting it in an array with 64 values and finally it will be applied a "softmax" function that will allow us to make the binary classification.

In all cases, we have divided the sample between train (80 % of the sample) and test (20 % of the sample), and the model is always trained with the train sample and evaluated with the test one.

4.3.1 Data augmentation

Deep learning is based on the large amount of data, and due that our samples are small we artificially augment it, this is the method called Data Augmentation. In each iteration we introduce perturbations like off-center, turning, flattening, or zooming, these changes increase the number of images that we use with the modification of our common database.

4.4 Evaluation tools

We use standard diagnostic tools like the ROC (Receiver Operating Curve) to assess the network performance. The ROC illustrates the ability of a binary classifier system as its discrimination threshold is varied. The discrimination threshold is defined in our case as the value at which a class is chosen over another. Recall that the output of the network is the result of applying a sigmoid function. It ranges therefore between 0 and 1, which can be interpreted as a probability. In Figure 6 we show an arbitrary ROC curve in which we also represent the different thresholds. It plots the true positive rate (TPR) also called sensitivity against the false positive rate (FPR) that are mathematically defined as:

$$TPR = \frac{TP}{TP + FN} \tag{4.2}$$

$$FPR = \frac{FP}{FP + TN} \tag{4.3}$$

where TP, TN, FP, FN are the true positives, true negatives, false positives, false negatives respectively, In our case the TP are fast rotators classified correctly, TN are slow rotators classified correctly, FP are fast rotators classified wrongly and FN are slow rotators classified wrongly.

The closer our curve is to the upper left corner, the better the quality, as the ratio of true

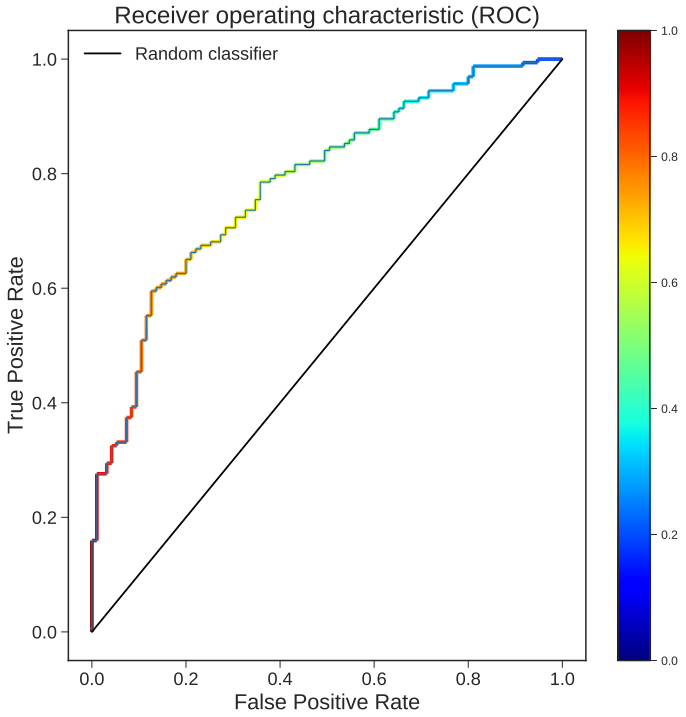


Figure 6: ROC curve with threshold representation. If we reduce the threshold, it means that more elements are classified as positive so the TPR and the FPR will increase.

positives is close to 1, or 100 percent. ROC curves look like Figure 7. Another way to evaluate our models using these diagnostic tools is by using the AUC (area under the ROC curve). The AUC provides an aggregate measurement of performance at all possible classification thresholds. One way to interpret the AUC is as the probability that the model will rank a random positive example higher than a random negative example. For example, in Figure 7, the AUC of the random classifier is 0.5, the AUC of the Less good is 0.65, and the AUC of the good result is 0.91.

For general purposes, ROC curves have an optimal point that would be the threshold value for which false positives are minimized by keeping a reasonable number of true positives. This point is located in the change of slope of the curve, and the threshold associated with it is the one that will give us a better result. For example in Figure 7 this change of slope is located at the point where the True Positive rate is 0.6 and a False Positive rate is 0.15, having a threshold of ≈ 0.75. The results that we will present will always be those associated with these "optimal points" with their threshold value.

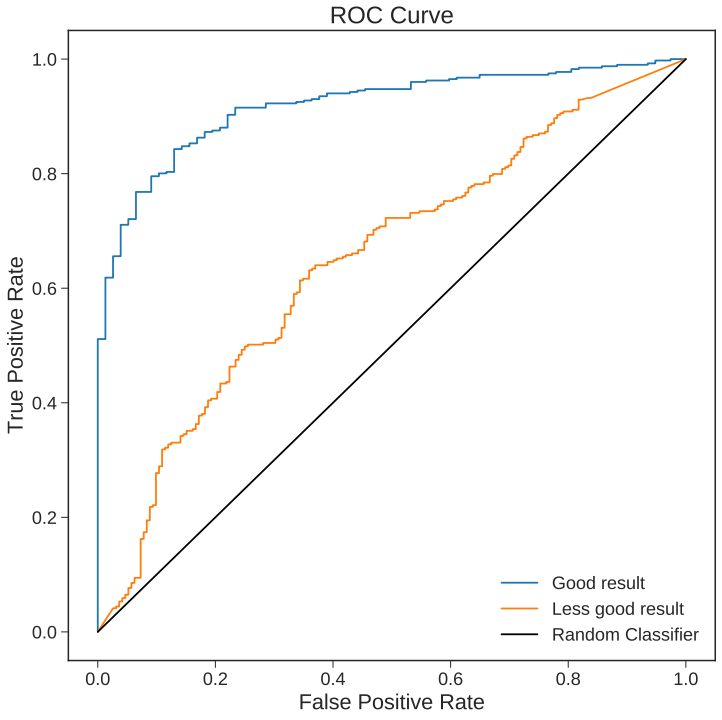


Figure 7: Example with three ROC curves, for random, good and less good classifier where it is clear that the higher the maximum ROC curve, the better the machine learning.

4.5 Interpretating the network decisions: Integrated gradients

One main issue with neural networks is their lack of interpretability. Attribution techniques help digging into the network decisions although the information is still rather limited. In this work we will use a technique called integrated gradients which introduced by Sundararajan et al. (2017). This technique, in a few words, will allow us to see in which pixels the model has been based to carry out a prediction. This can potentially help discovering patterns or elements that were not visible with a simple visual inspection. The results that we get are attribution maps. The computation has been made using the package by Ancona et al. 2017.

5 Results

The progress of the work will be carried out by performing studies on the two datasets, i.e., observations and simulations. As previously emphasized, the main objective is to understand if a neural network can classify early-type galaxies into fast or slow rotators with only the information of their optical morphology.

5.1 Observations: MaNGA

We will first focus on observational data from the MaNGA survey where we have a total of 4672 galaxies of which 2255 are early-type.

5.1.1 Slow-Fast rotator, all galaxies

The first step will be to check that the model and data structure are correct and that there are no matching errors. To that purpose we start with a simple classification task. We classify slow and fast rotators using all types, i.e. including late-type systems (Figure 3). As we have seen in the introduction, the majority of late-type galaxies are fast rotators and have a much better-differentiated structure than the early-type, with a well-defined bulge and spiral arms, so it should be easy for the convolutional network to make a classification, as elliptical galaxies will have less weight in the classification.

The percentage of fast-slow rotators we have is represented in Figure 8 where Rotator type = 0 are slow rotators and Rotator type = 1 are fast rotators (that is the way that the network works with the different classes); The difference in numbers between fast and slow rotators is very large, due in part to the fact that as we have mentioned most of the late type galaxies are fast rotators. Such large population differences can cause classification problems because the model loses generality as it learns more from the data that are in the majority.. I will not go into detail about the parameters used in this example, since as mentioned, it is a simple check.

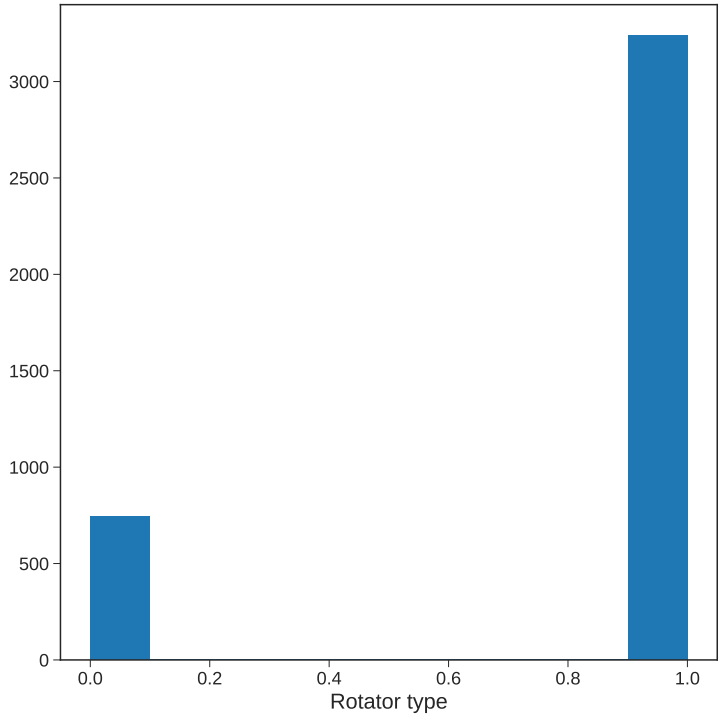


Figure 8: Histogram representing the number of slow rotators (Rotator type = 0) and fast rotators (Rotator type = 1) in the SDSS's early and late-type galaxy database. There are more than three times more fast than slow rotators.

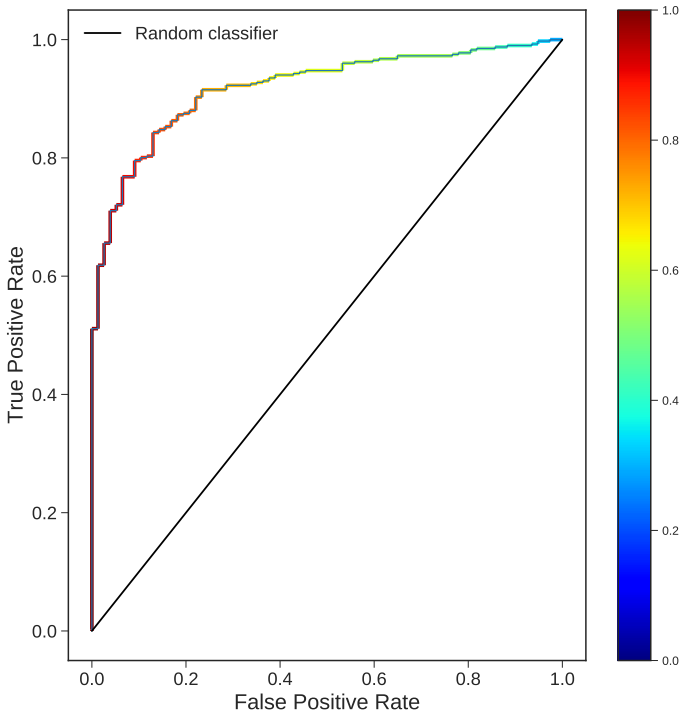
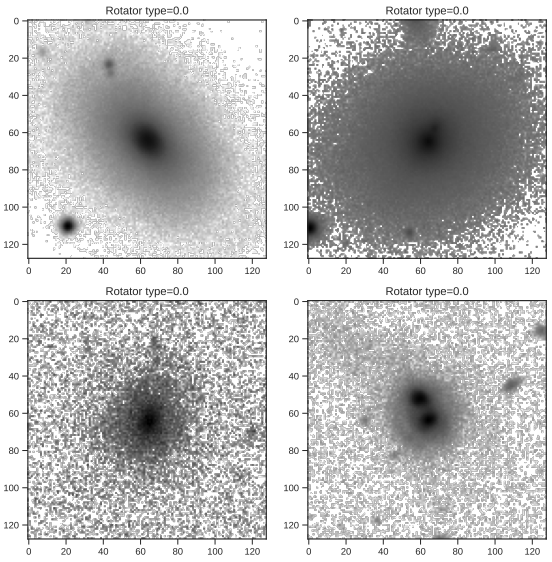


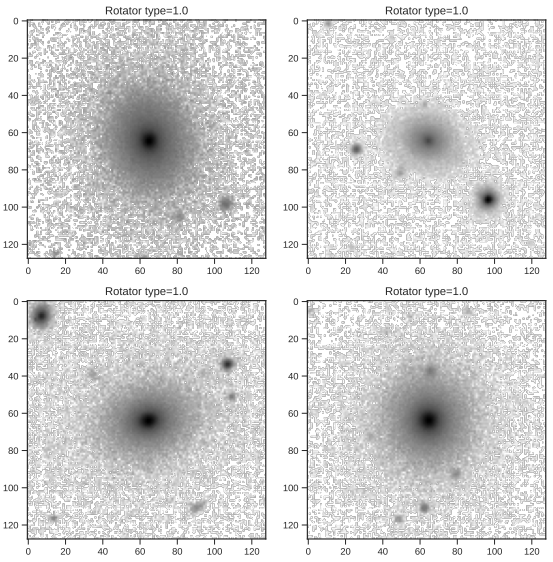
Figure 9: ROC curve for late and early type galaxies on the SDSS survey. The curve is close to the top left corner which reflects the fact that the classification is reasonably good as expected, obtaining an AUC of 0.92.

As expected, in Figure 9 we see that we have a good classification, with an AUC of 0.92, which could be improved, but as it was for evaluation purposes we are conformed with that result. This ensures us that our model is adequate, that we are not having errors when matching the images (this is an important issue because to get these results we are getting data from at least 4 different catalogs) and that the consideration made at the beginning that late-type galaxies are easier to classify. The optimal point in this model have a threshold of 0.75

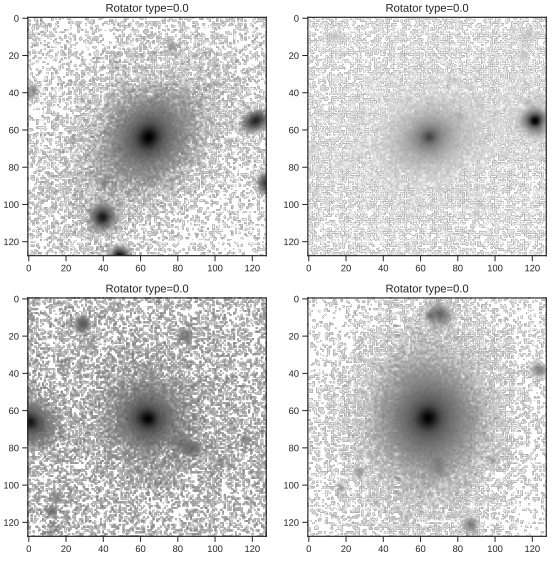
Next, we will make an inspection of which images have been correctly classified and which have not (Figure 10), trying to see if we can find some correlation from their visual structure and representing the $\lambda_R - \epsilon$ diagram to analyze patterns (Figure 11).



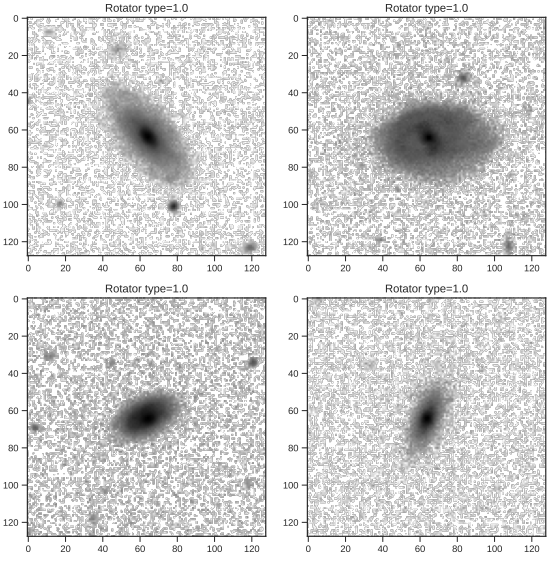
(a) Galaxies kinematically classified as slow rotators but morphologically classified as fast rotators



(b) Galaxies kinematically classified as fast rotator but morphologically classified as slow rotator

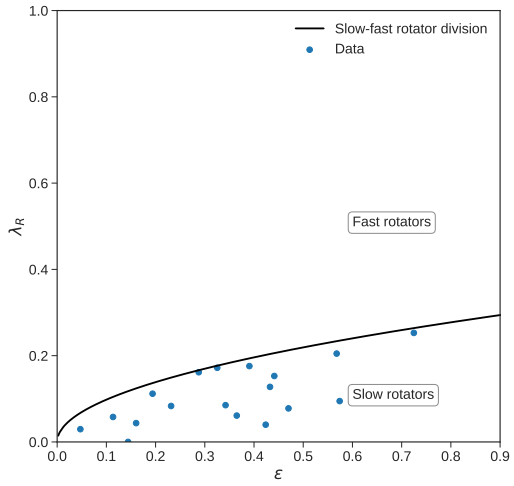


(c) Galaxies classified both kinematically and morphologically as slow rotators

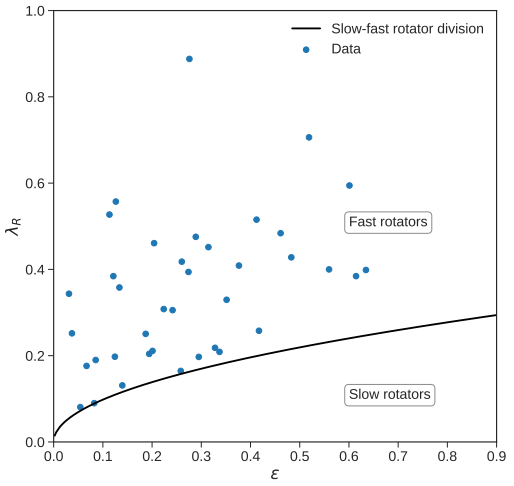


(d) Galaxies classified both kinematically and morphologically as fast rotators

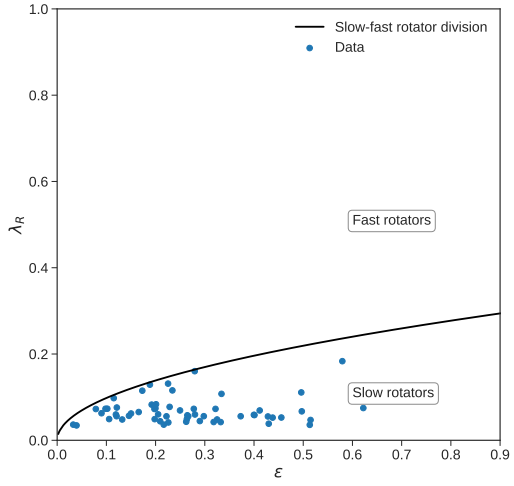
Figure 10: Visual examples of the morphology of galaxies that have been correctly and wrongly classified are represented, where type 0 rotator refers to slow rotator and 1 to fast rotator. In the wrongly classified galaxies (Figure 10a and 10b) we can see that all of them are early type, which confirms the variety of rotation types of this population and therefore the difficulty to classify those systems. If we focus on the case of well classified galaxies we see that the examples of slow rotators are characterized by early-type galaxies with low ellipticity and the fast rotators by galaxies with well described disks.



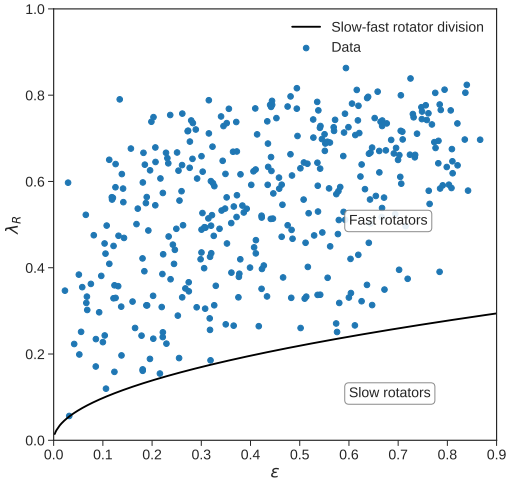
(a) Galaxies kinematically classified as slow rotator but morphologically classified as fast rotator



(b) Galaxies kinematically classified as fast rotator but morphologically classified as slow rotator



(c) Galaxies classified both kinematically and morphologically as slow rotators



(d) Galaxies classified both kinematically and morphologically as fast rotators

Figure 11: Representation on $\lambda_R - \epsilon$ diagram for wrong and correct classifications. The interest we have in studying these distributions is to know if there is any measurable characteristic that allows us to understand why the classifications, in this case we talk about ellipticity. The number of points we have is less than those represented in the Figure 1 because we are representing the data used for the neural network test. The galaxies that were really fast rotators and were classified as slow we see that they are those that have a low ellipticity, instead we see that those that were slow rotators and were classified as fast extend in a greater amount of possible ellipticities. As for the well classified ones, the fast ones have a slightly higher accumulation for higher ellipticities, and for the slow ones there is no population for high ellipticities.

5.1.2 Slow-Fast rotator, early-type data

With these results, it is time to work with only early-type galaxies. The number of data we will have for this case will be a third of the previous size and the fast-slow rotator distribution (Figure 12). As we expected because we don't have late-type galaxies, the number of fast rotators has dropped significantly, but still the number of fast rotators is more than twice the number of slow rotators. This again reflects the large variety of kinematical properties of early-type systems. If we run the model with the same characteristics as the previous one we will have a result like the one in Figure 13, with an AUC of 0.78. It is as expected worse than the previous case, but it it is still better than a random classifier. The optimal point in this model have a threshold of 0.8

A similar analysis to the one performed previously with all the data is necessary, since in this case, only with early-type galaxies is where we have interest in what produces the bad and good classifications. In Figure 14 we have visual examples of the morphology of galaxies that have been correctly and wrongly classified. We see that the miss-classified galaxies can have non common structures like other objects close to them. When we see the correct classifications there is not too much difference between fast and slow rotators, we see perhaps a hint that fast-rotators are more elongated, also we have an example of fast rotator (Figure 10d) where the galaxy have some structure, it means that it is just in the division of early-late type. With the Figure 15 where we have a diagram $\lambda_R - \epsilon$ there is a clear dependence with the ellipticity. The network can wrongly classify fast rotator galaxies as slow rotator when the ellipticity is low. In the other hand, the is no clear pattern of how the network wrongly classifies slow rotators as fast rotators, because the distribution of ϵ stays still in a low range (≈ 0.5).

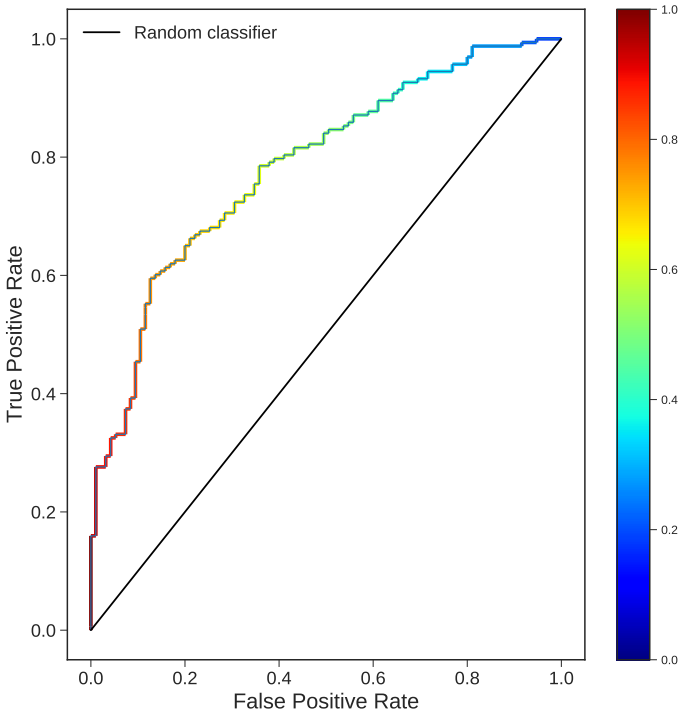


Figure 13: ROC curve for early type galaxies on the SDSS survey. The result is far of a random classification and gets an accuracy of 0.78 aprox.

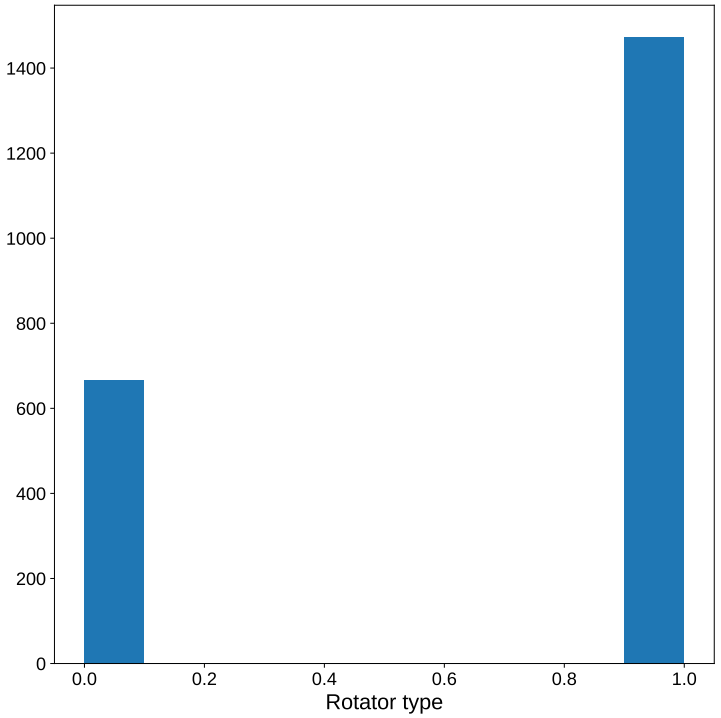
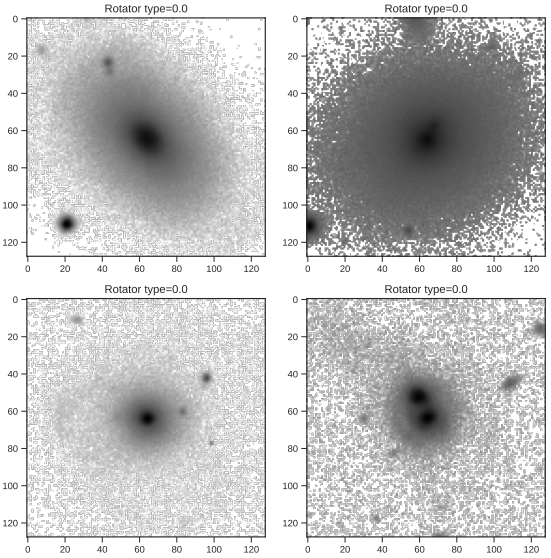


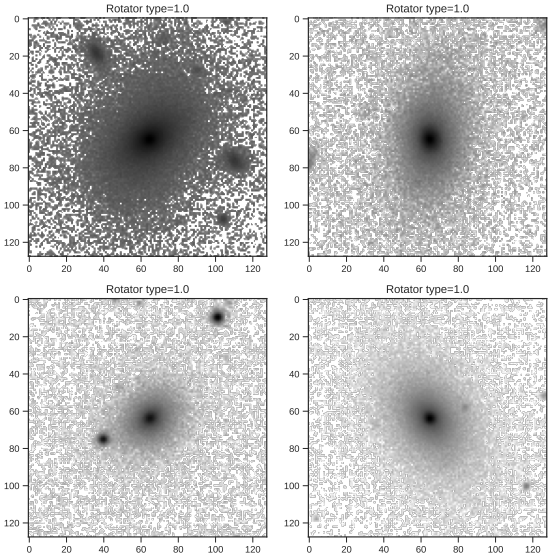
Figure 12: Histogram representing the number of slow rotators (Rotator type = 0) and fast rotators (Rotator type = 1) in the SDSS's early-type galaxy database. There are more than twice as many fast rotators as slow ones.

5.1.3 Integrated gradients

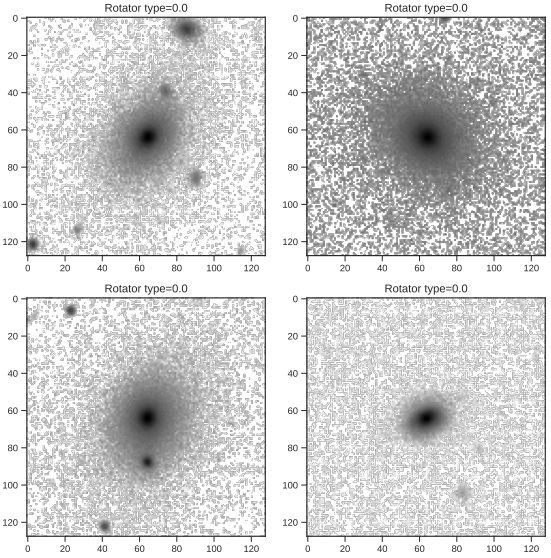
In order to get more insight into the network decisions, we use the integrated gradients method, to see if the network finds some kind of structure that can provide physical information. The integrated gradients show us the pixels that the neural network has taken most into account to learn. In Figure 16 the attributions maps are represented, and we can see that the network ignores almost all the noise and is also capable of ignoring almost all light sources from other galaxies that are located next to them. We also observe that the most important pixels are near the center, i.e. the bulge which is what we expected, because if some rotation exists, it will be on the center of the galaxy. However it is difficult to translate into a physical explanation.



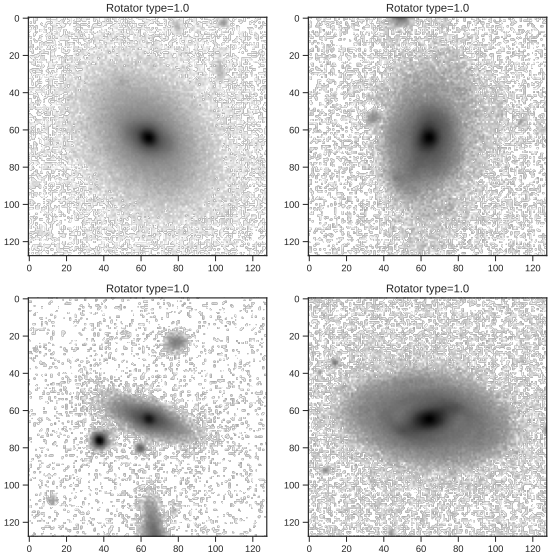
(a) Galaxies kinematically classified as slow rotator but morphologically classified as fast rotator



(b) Galaxies kinematically classified as fast rotator but morphologically classified as slow rotator

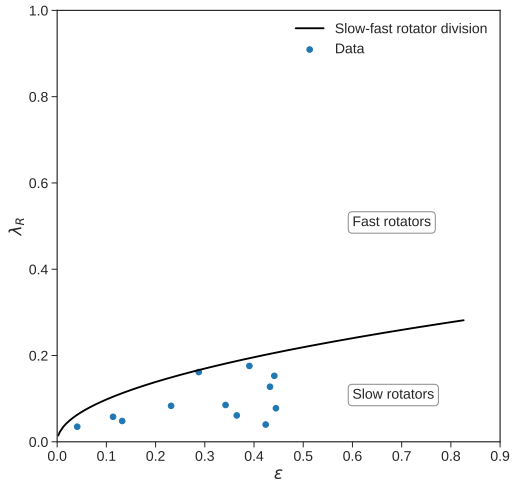


(c) Galaxies classified both morphologically and kinematically as slow rotators

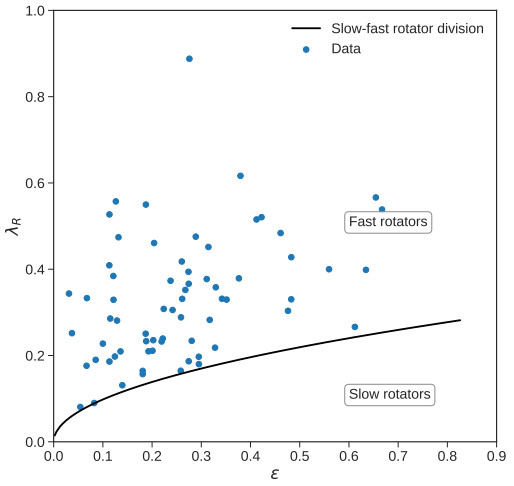


(d) Galaxies classified both morphologically and kinematically as fast rotators

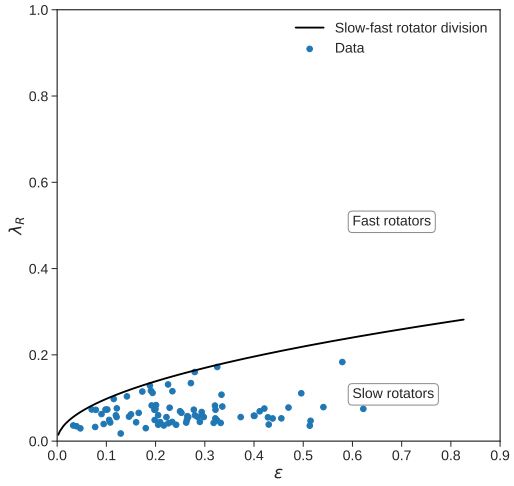
Figure 14: Visual examples of the morphology of galaxies that have been correctly and wrongly classified are represented, where type 0 rotator refers to slow rotator and 1 to fast rotator. We see that those galaxies that are miss-classified can have nearby objects and characteristic structures like rings of star formation, but in terms of morphological structure, the galaxies we observe are very similar among them. For example, in the well classified fast rotators we see a more elliptical structure than in the slow rotators.



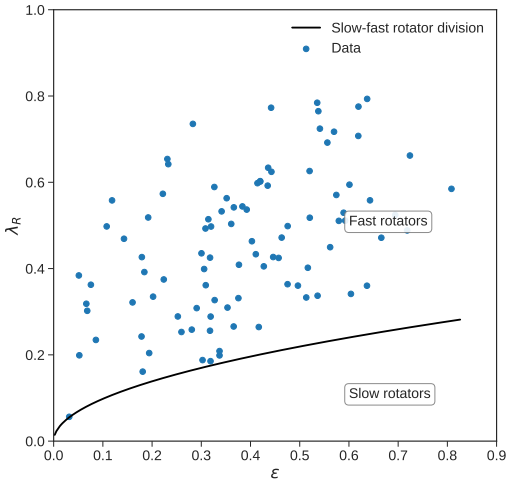
(a) Galaxies kinematically classified as slow rotator but morphologically classified as fast rotator



(b) Galaxies kinematically classified as fast rotator but morphologically classified as slow rotator



(c) Galaxies classified both kinematically and morphologically as slow rotators



(d) Galaxies classified both kinematically and morphologically as fast rotators

Figure 15: epresentation on $\lambda_R - \epsilon$ diagram for wrong and correct classifications. The data is clearly distributed in a way that you can see that the network makes many errors when classifying fast rotators into slow ones, these having a distribution of low ellipticities. Confuses a few fast slow with fast ones, and correctly classifies a significant number of galaxies. We see that the slow rotators have a greater proportion of galaxies in lower ellipticities, unlike the fast rotators, have it in higher ellipticities

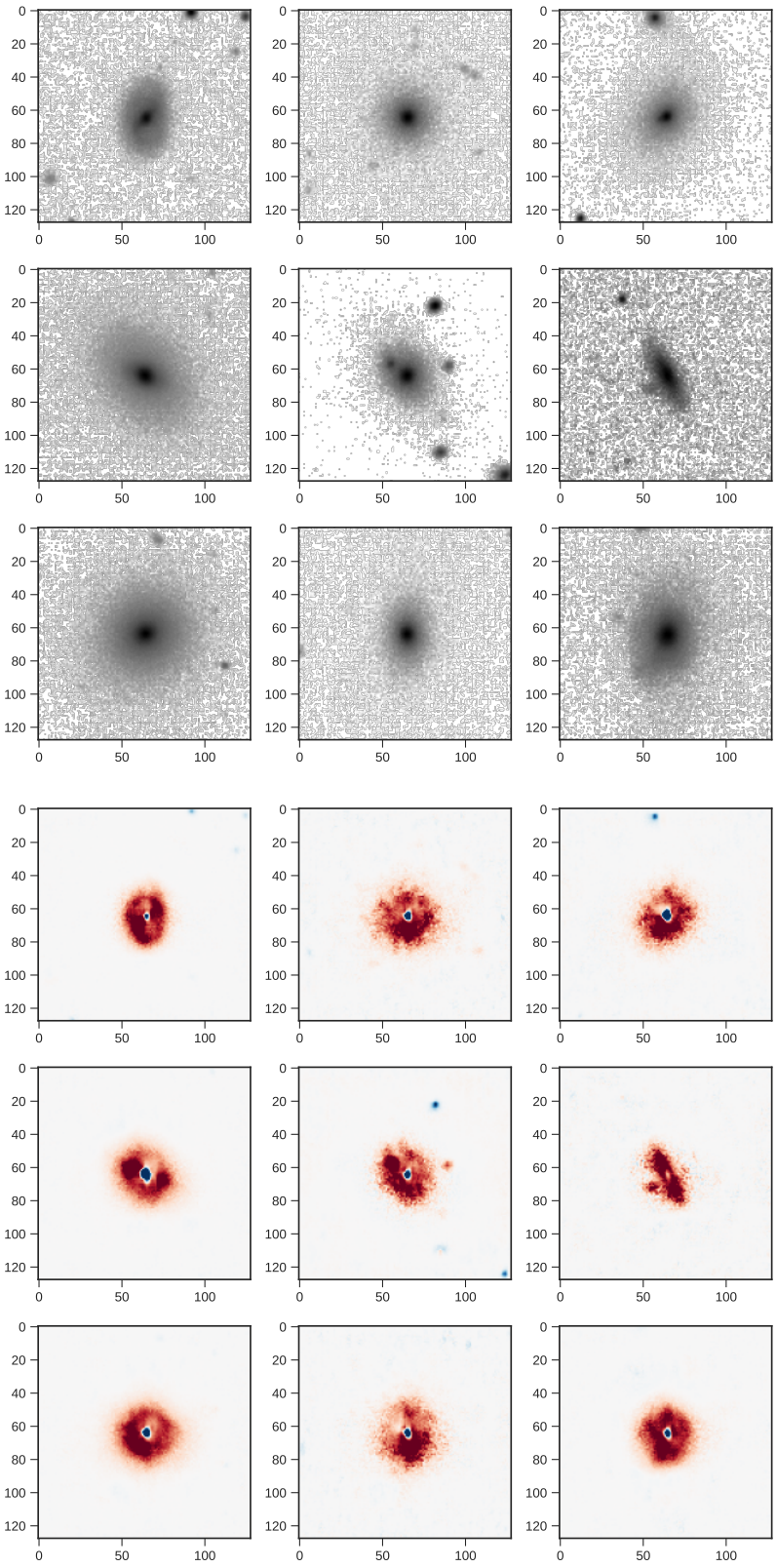


Figure 16: The top panel shows SDSS early type galaxies in a logarithmic normalization with a size of 128x128. The bottom panel shows the attribution maps of the same images computed through the integrated methods. These maps are normalized so they just reflect variations on the values.

5.2 TNG100 (Illustris)

The results of the network being trained with the MaNGA data have provided us with results that give us an idea of the extent to which the network is capable of performing a correct classification. We now investigate how things behave in simulations. There are different reasons to investigate simulations. Firstly, we can investigate if the properties of fast and slow rotators are well reproduced by state-of-the art simulations. Secondly, simulations allow us to explore noiseless images, so we can investigate whether there are signatures of the rotation status at low surface brightness not probed by the SDSS imaging. Finally, since our samples are small, we can use simulated images to increase the size of our training set and see if this improves the performance of the CNNs.. We will use the data of early-type galaxies from the Illustris TNG100 simulations with noise to try to match as closely as possible the environment of the observed galaxies. Noiseless images will be explored in Section 6.3.

The distribution of galaxies (Figure 18, Figure 17) is very similar to that of the SDSS. This is already interesting because it means that the dynamical properties of galaxies at z=0 are well reproduced by TNG. Huertas-Company et al. (2019) showed that the optical morphologies are realistic. We see here that if the kinematic properties are also well reproduced. This allows us to move forward in the analysis. If we study the ROC curve obtained (Figure 19), we have a curve similar to the previous ones but with a lower AUC, in the range of 0.68 versus 0.78 of those obtained for the observations and the optimal point in this model have a threshold of 0.8. There seems to be a drop in accuracy, although the values are in the same range. We will analyze this a bit further.

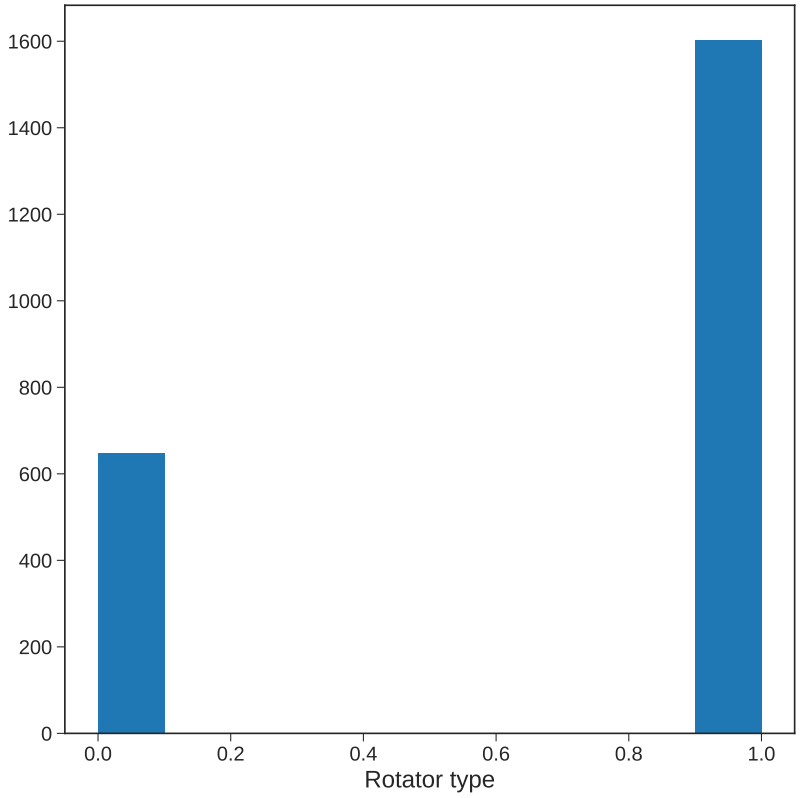


Figure 17: Histogram representing the number of slow rotators (Rotator type = 0) and fast rotators (Rotator type = 1) in the TNG100's early-type galaxy database. There are more than twice as many fast rotators as slow ones.

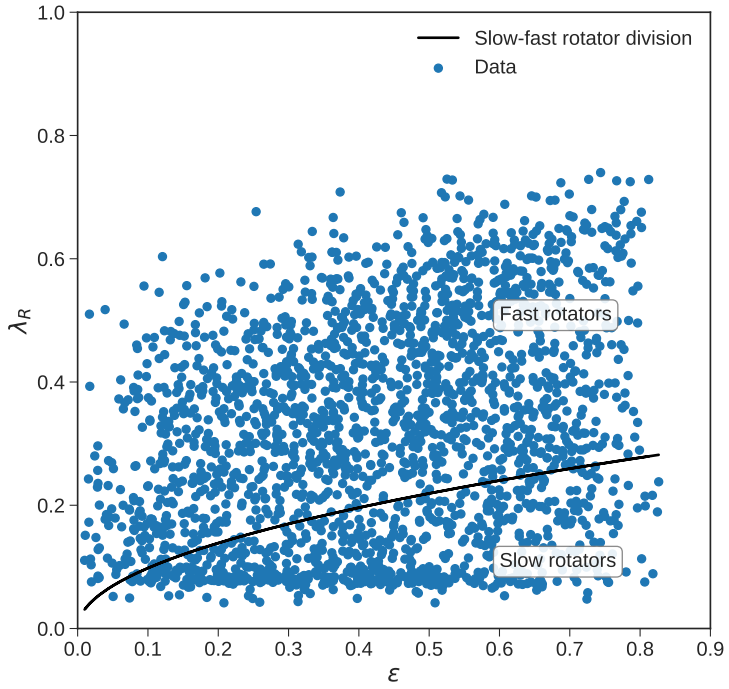


Figure 18: Fast-slow rotator distribution from the TNG100 simulation at z = 0. This distribution for early-type galaxies has a lower λ_R maximum than for the MaNGA data and is uniformly distributed in the fast and slow rotator groups.

The way to analyze this data will be the same as with the other survey, we will analyze the cases for correct and wrong classifications to see if we find characteristics similar to those we obtained from the observations. If this happens it would be interesting, since it would mean that the network classifies at least similarly with both databases. From the Figures 20 and 21 you can think that the network classifies much better fast rotators than slow ones, since in the latter it makes many mistakes and performs few correct classifications.

5.2.1 Integrated gradients

Studying the attribute maps of the gradients in integrated in this case has an important interest, since we obtained significantly different results from those obtained from the observations. In the Figure 22 the difference can be clearly seen. When in the observations the network was fixed practically only in the center and did not take into account external features, in this case the network takes into account many other factors, which make it very difficult to obtain any common characteristic. The structures further away from the bulge become more important. It should be remembered that the normalization used in this case is the same as for the observations.

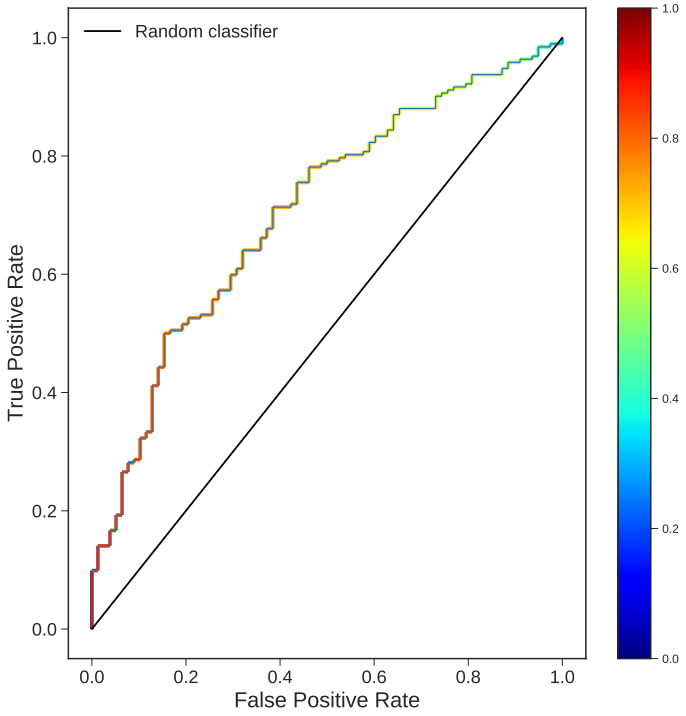
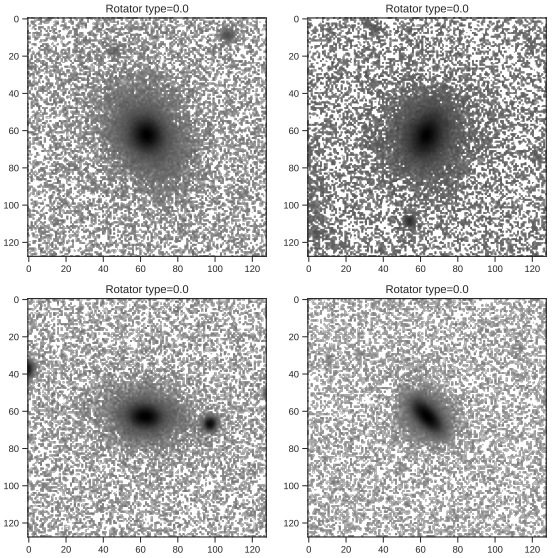
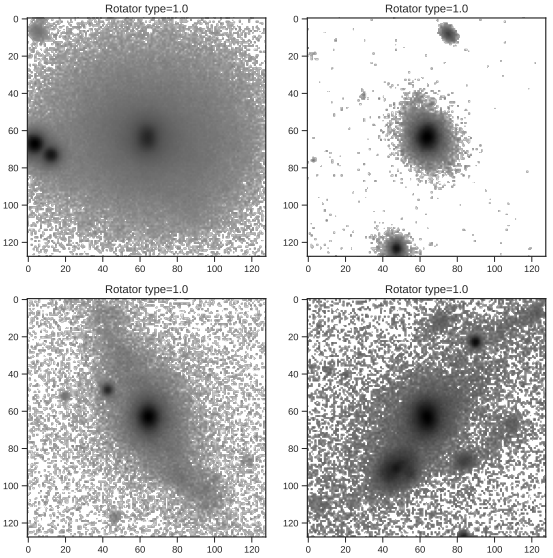


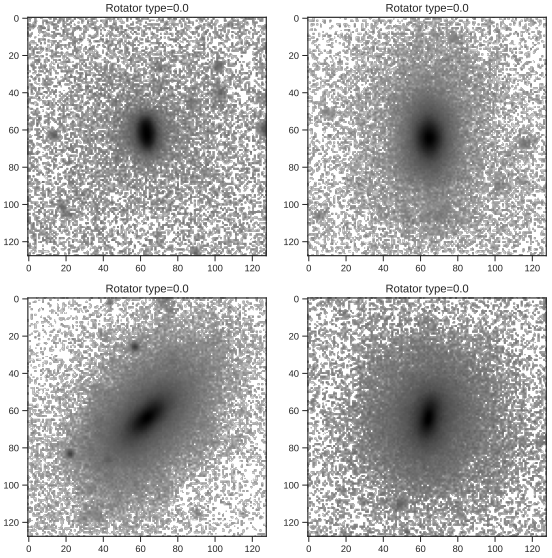
Figure 19: ROC curve for early-type galaxies for the TNG100 data. It is a worse result that the obtained for the observations, getting an AUC of 0.7



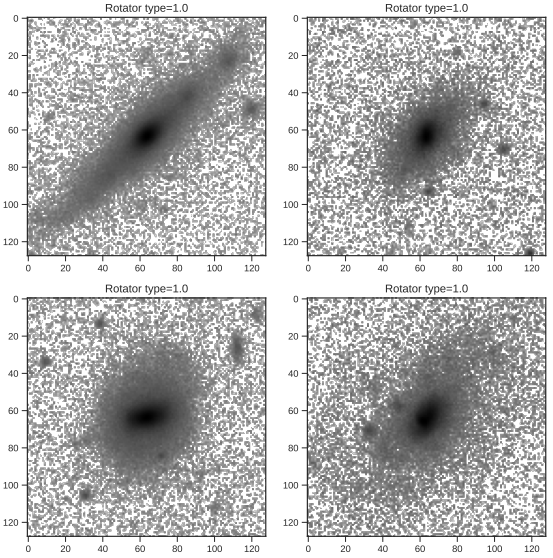
(a) Galaxies kinematically classified as slow rotator but morphologically classified as fast rotator



(b) Galaxies kinematically classified as fast rotator but morphologically classified as slow rotator

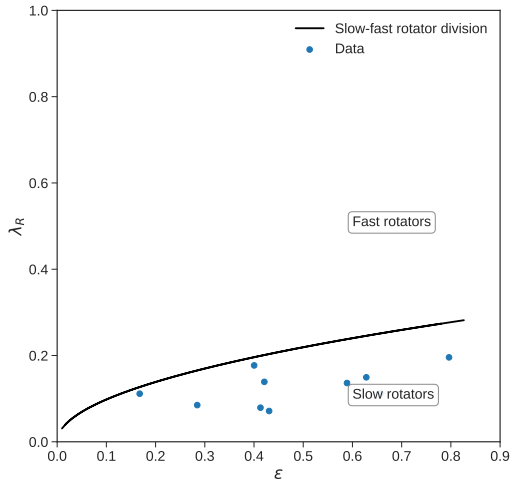


(c) Galaxies classified both kinematically and morphologically as slow rotators

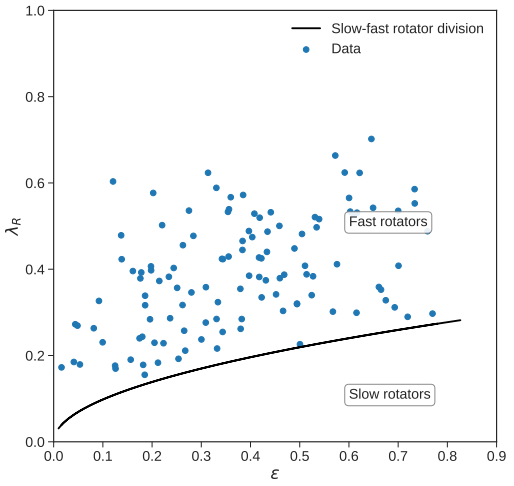


(d) Galaxies classified both kinematically and morphologically as fast rotators

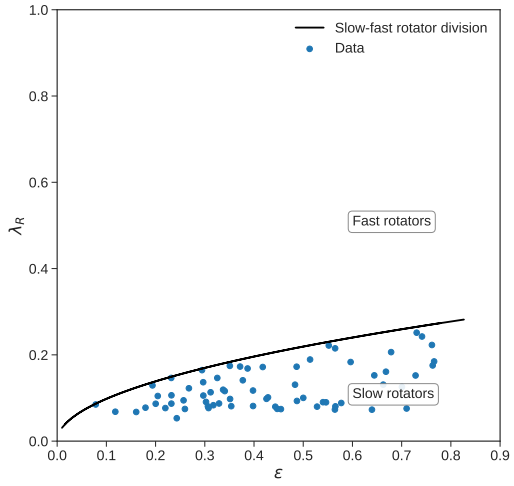
Figure 20: Visual examples of the morphology of galaxies that have been correctly and wrongly classified are represented, where type 0 rotator refers to slow rotator and 1 to fast rotator. We see that those galaxies that are miss-classified can have nearby objects and characteristic structures like rings of star formation or that they have a particular distribution of values that differentiates them from the rest, but in terms of morphological structure, the galaxies we observe are very similar among them. In the well classified fast rotators we see a more elliptical structure than in the slow rotators.



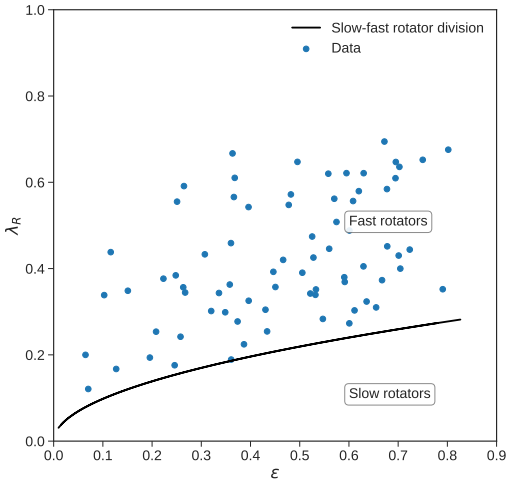
(a) Galaxies kinematically classified as slow rotator but morphologically classified as fast rotator



(b) Galaxies kinematically classified as fast rotator but morphologically classified as slow rotator



(c) Galaxies classified both kinematically and morphologically as slow rotators



(d) Galaxies classified both kinematically and morphologically as fast rotators

Figure 21: Representation on $\lambda_R - \epsilon$ diagram for wrong and correct classifications. The data is clearly distributed in a way that you can see that the network makes many errors when classifying fast rotators into slow ones, these having a distribution of low ellipticities. Confuses a few slow with fast ones, and correctly classifying a significant number of galaxies. We see that the slow rotators have a greater proportion in lower ellipticities, unlike the fast rotators, have it in higher ellipticities.

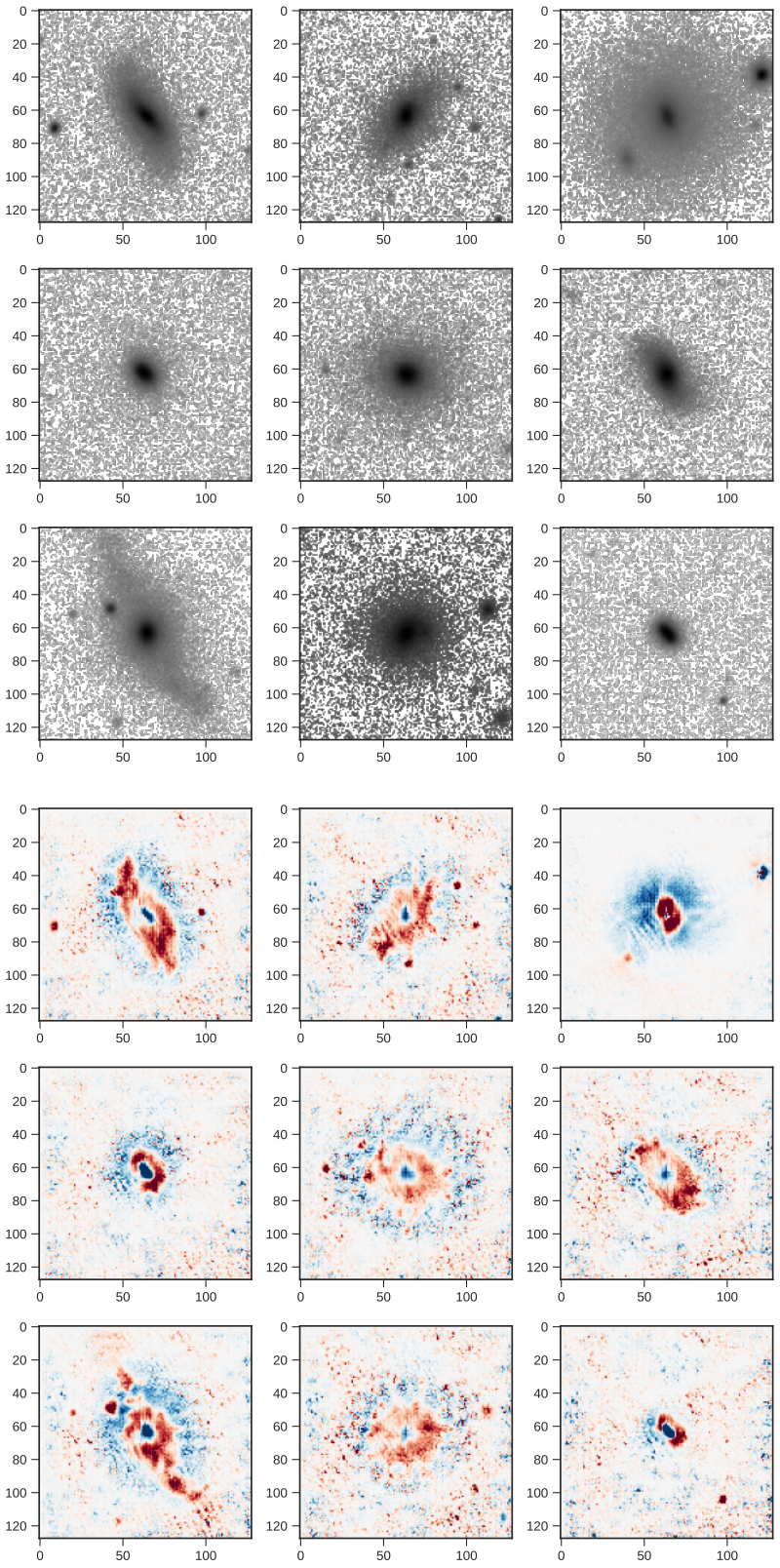


Figure 22: The top panel shows TNG100 early type galaxies in a logarithmic normalization with a size of 64x64. The bottom panel shows the attribution maps of the same images computed through the integrated methods. These maps are normalized so they just reflect variations on the values

6 Model optimization and performance improvement

In this section, we describe multiple experiments we have performed to better understand the network behavior and try to optimize its performance. In particular we try to optimize the network training to see how this affects the results. We also explore if color is a relevant information as well as if the region of the galaxy on which the networks focuses is important. We perform these experiments for both observations and simulations. For the simulations, we additionally explore the impact of noise since we can train on noiseless images.

6.1 Regularization methods

Apart from the commented structure of the CNNs, which are transformations and convolutions that produce changes in the sizes of the tensors, there are very important methods, called regularization methods that help the neural networks in the training process and avoid the risk of overfitting. There are numerous methods but we will focus on two, the Dropout (Srivastava et al., 2014) and the Batch Normalization (Ioffe & Szegedy, 2015) and we will see which one gives better results and is the one we will use in the rest of the project. The Dropout will randomly deactivate a percentage of the neurons in each hidden layer, according to a previously defined discard probability, which in our case is 25%. What this means is that no neurons will memorize part of the input; which is precisely what happens when we have overfitting. Batch Normalization consists of adding an extra step, usually between the neurons and the activation function, with the idea of normalizing the output activations. It was presented as a solution to reduce the Internal Covariate Shift, defined as the change in the distribution of network activations due to the change in network parameters during training.

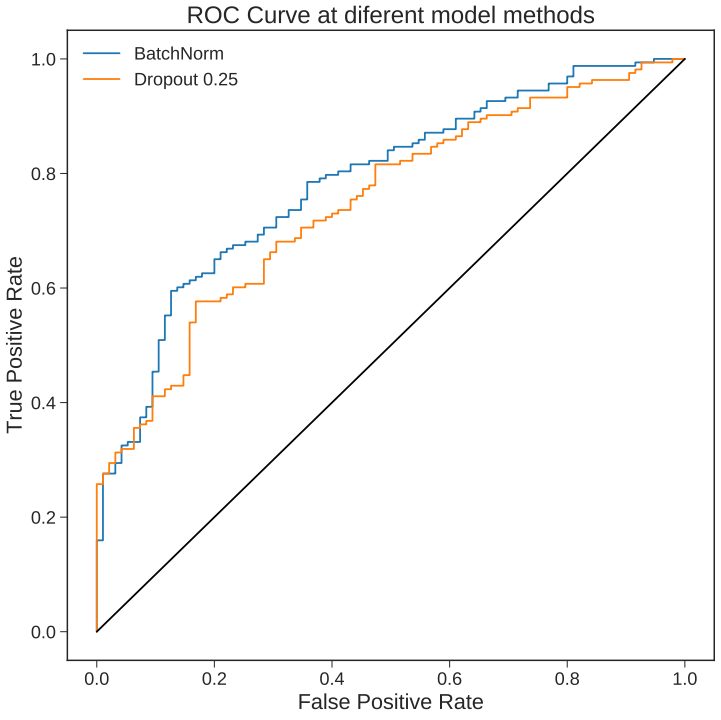


Figure 23: ROC curve for Dropout(0.25) and Batch Normalization. The difference of learning by varying the method of regularization is low, but we will use the BatchNorm because the result it gives is slightly better.

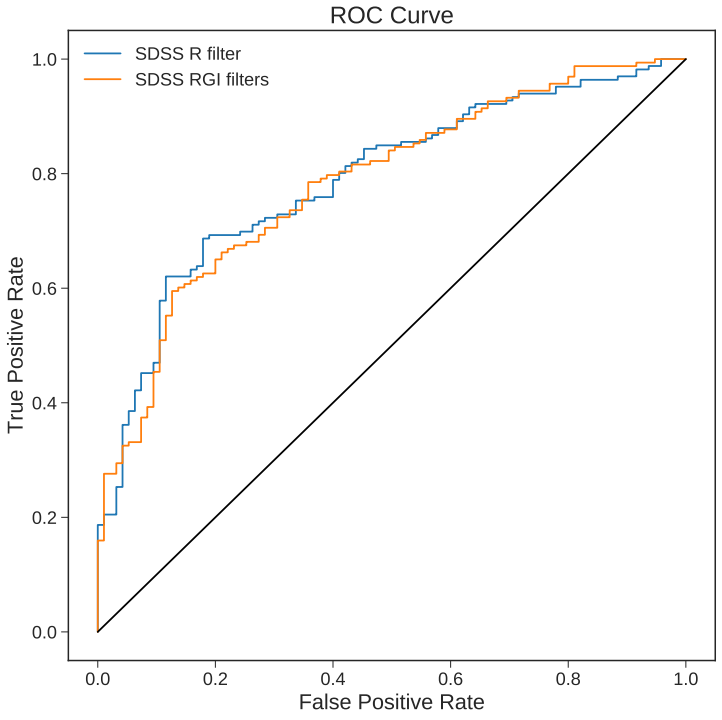


Figure 24: ROC curve for early type galaxies with different filters on the SDSS survey. The result is basically the same, so the information given by the color is non relevant.

6.2 Observations

6.2.1 Color

Is it relevant to have color information in our images when making a kinematic classification?. In Figure 24 ROC curves are compared for a case where the images were entered into the network with only one r (6231Å) filter and for another case with three r(6231Å), g, (4470Å), i (4625Å) filters (i.e. there is color information in the images). The result is quite obvious, the results are practically the same. Surprisingly, it seems that color information is not taken into account by the network. Recent work has shown however, that slow and fast rotators present age differences in their stellar populations. This is not seen here perhaps because the resolution is too coarse, so we could say that this parameter is hardly important for our network when it comes to learning.

6.2.2 Normalization

Finding the most appropriate normalization for our data is very important because although CNNs can distinguish very fine details, the calculations that are made in the neurons are products of the pixel values and weights, so if we have a distribution of pixel values not optimized for our problem we can get bad results, an example of this is the Figure 25 If our 4-D tensor in which our data are stored as the shape:

$$\mathbf{T} = T_{i,x,y,k} \tag{6.1}$$

where i is the image number, x and y are the image size and k is the filter. These variables in

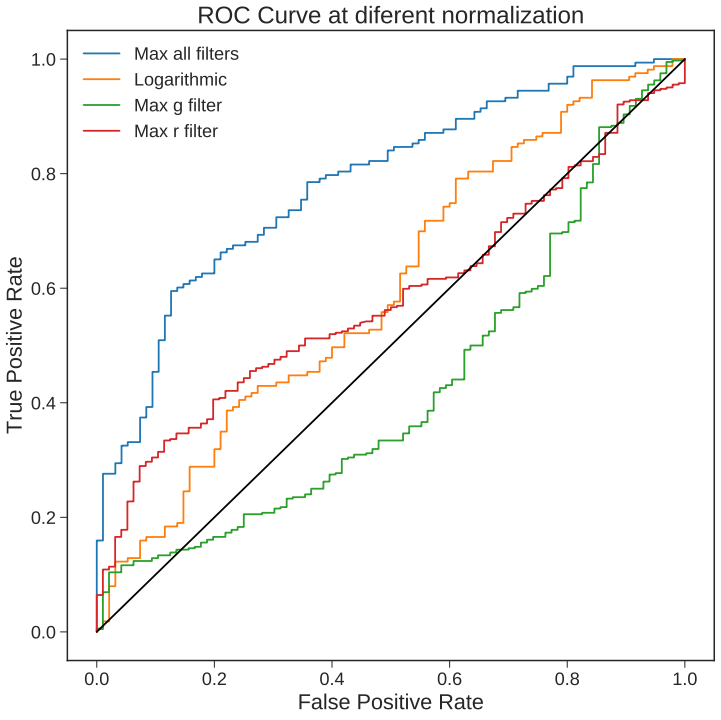


Figure 25: ROC curve for early type galaxies with different normalizations. Can be a large difference in accuracy depending of the normalization because relevant structures can be blurred or the difference in pixel values may not be enough for the network to work properly

our particular case of the SDSS survey are defined between the values: $i \in (0, N]$ (N the number of images), $x, y \in (0, 128]$ and $k \in (0, 1, 2)$ with $0, 1, 2$ the filters g,r,i respectively. We can describe our normalization methods as:

$$\text{Max all filters} \rightarrow T'_{N,x,y,k} = \frac{T_{N,x,y,k}}{T_{N,x,y,k_{max}}} \tag{6.2}$$

$$\text{Logarithmic} \rightarrow T'_{N,x,y,k} = \frac{\log T_{N,x,y,k}}{\log T_{N,x,y,k_{max}}} \tag{6.3}$$

$$\text{Max g filter} \rightarrow T'_{N,x,y,k} = \frac{T_{N,x,y,k}}{T_{N,x,y,0_{max}}} \tag{6.4}$$

$$\text{Max r filter} \rightarrow T'_{N,x,y,k} = \frac{T_{N,x,y,k}}{T_{N,x,y,1_{max}}} \tag{6.5}$$

apart from these filters, other more complex ones imported from different python packages were used, but the one with the best result was also the "Max all filters". As you can see in the figure, some normalizations make our network almost a random classification, this can happen because when modifying the data we are eliminating relevant information for learning.

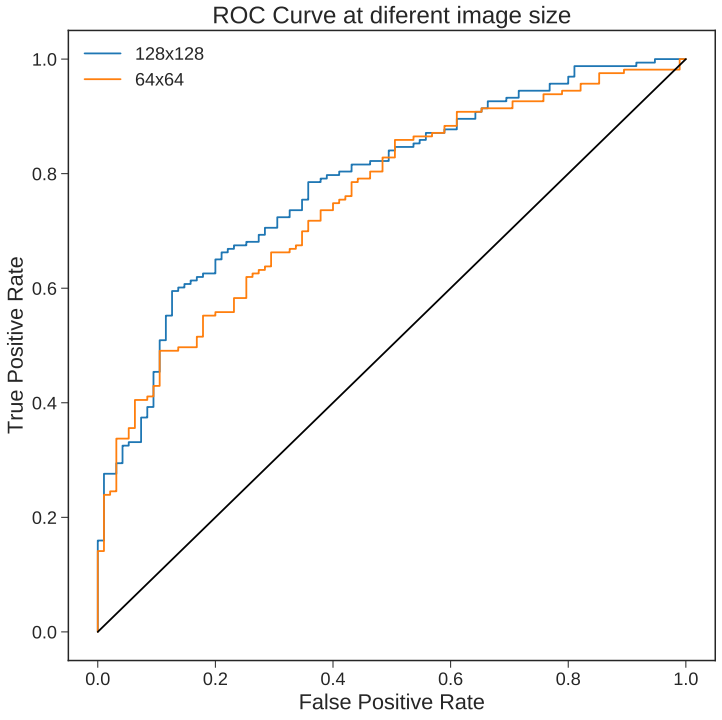


Figure 26: ROC curve for early type galaxies with different sizes. There is not clear difference, so we can ensure that most of the information acquired by the network is from the center

6.2.3 Image size

We have developed all the work with images of a certain size of 128x128 pixels, but it is possible to study for example what results we would have if we zoomed in and focused more on the core. Since the rotation is determined using the central parts, we want to test if we measure any improvement when forcing the CNN to focus on the central regions. Note that the results represented in Figure 26 are not just a resize, what we did was to crop a 64x64 image from the center, so the resolution stays still.

The results are clear and show that looking at the center produces practically the same result as having a larger view, this can be explained as most of the relevant information when studying the type of rotation that an early-type galaxy has is in the center. More data is needed to be sure to make such a statement, this information will come from the integrated gradients represented in Figure 27 where it is clear that the network is focused in the center and avoids almost all noise and nearby light sources. The similarity of results and maps make us think that in the outer zones of the galaxies there is no relevant information when studying the type of rotation.

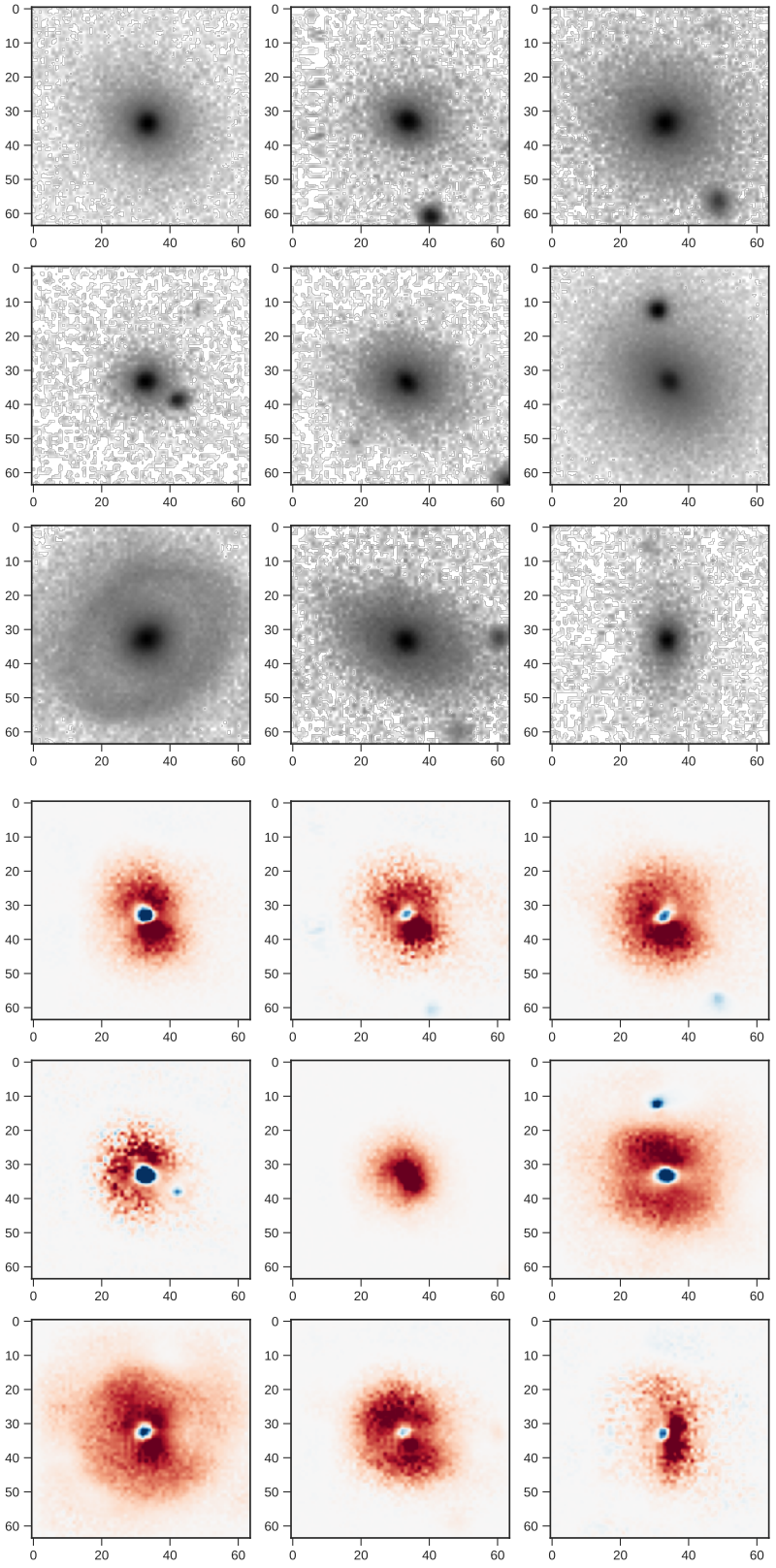


Figure 27: The top panel shows SDSS early type galaxies in a logarithmic normalization with a size of 64x64. The bottom panel shows the attribution maps of the same images computed through the integrated methods. These maps are normalized so they just reflect variations on the values

6.3 Simulations (No noise)

We have seen that the results on the noisy images of TNG100 provide similar results as for the observations. Observations, are necessarily affected by noise. It could be that there exist low surface brightness features which remain hidden by the noise and are still relevant to infer kinematics. We explore here noiseless images.

6.3.1 Normalization

The first data give us better results than in the case with noise, obtaining an AUC of 0.72 for the best normalization, compared to 0.68 obtained for the case with noise . The Figure 28 shows some results for some normalizations, among them the green line where no normalization was applied (I wanted to see if the simulated images themselves already had some kind of processing), and the logarithmic line that in this case gave the best result, being practically the same as the "Norm all filters" described before. Why are these results worse than the ones in the SDSS catalog?. One option could be that we are using less data, currently, we have about 1200 galaxies and with SDSS we had about 4500, or that maybe there is some kind of structure that the network finds in the observed galaxies and not the simulated ones. Anyway, a couple of more experiments are needed to get a clear idea.

6.3.2 Size

Applying the zoom procedure to the center of the galaxies we obtain a different result than expected from what we know from the previous experiment Figure 29. The AUC for the 128x128 case is 0.72 and for the 64x64 case is 0.79, this is that we have a better result when we focus the bulge of the galaxy, and we even surpassed the highest AUC we had obtained for the SDSS which was 0.78. If we study the maps of attributes calculated through the integrated gradients for the case of images with a size of 128x128 (Figure 30) we see that the network is focused in a much more precise way in the structure of the galaxy, and in the cases in which it seems that the background becomes important it is due to the fact that the galaxy extends visually until practically the limit of the image. When we study the maps for a size of 64x64 where we obtained better results we see that the background (that really is part of the center of the galaxy) becomes much more important and around the center more complex structures are formed. Another interesting point is that in some cases, the center of the image lacks impotence, and these are the edges where the network is focused.

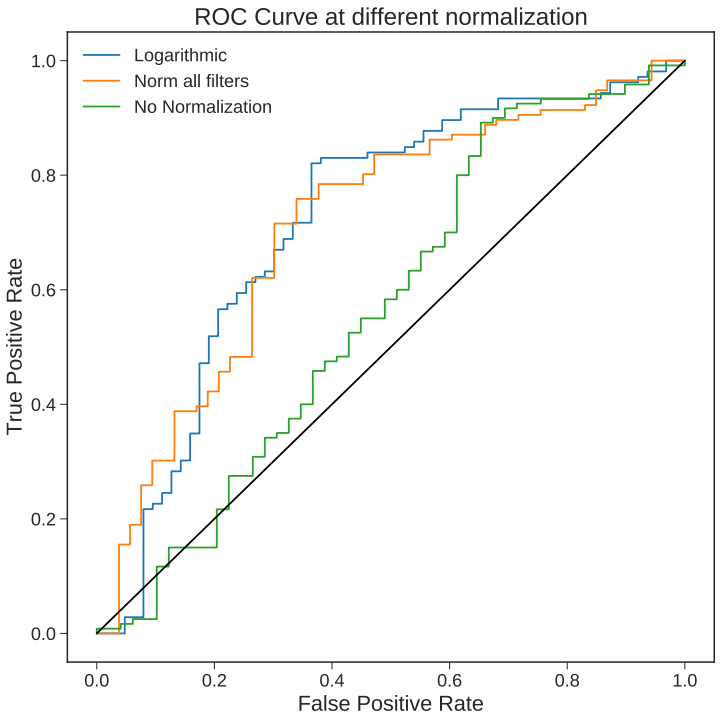


Figure 28: ROC curve for early type galaxies with different normalization, obtaining very similar results for a logarithmic and max normalization, and showing what we obtain almost a random classificator if we don't use normalization.

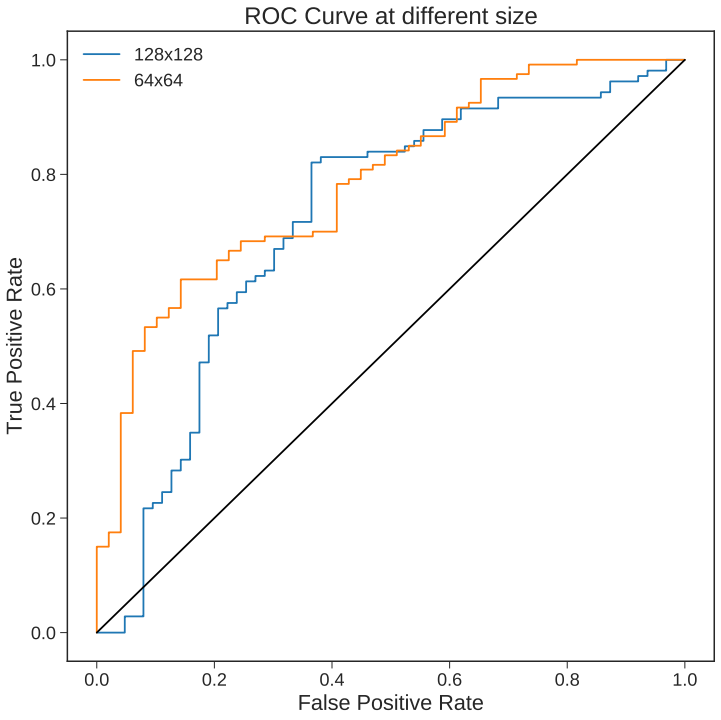


Figure 29: ROC curve for early type galaxies with different size. There is a difference between the model predictions when we change the size of the images, obtaining a better result with images that are focuses on the bulge of the galaxy.

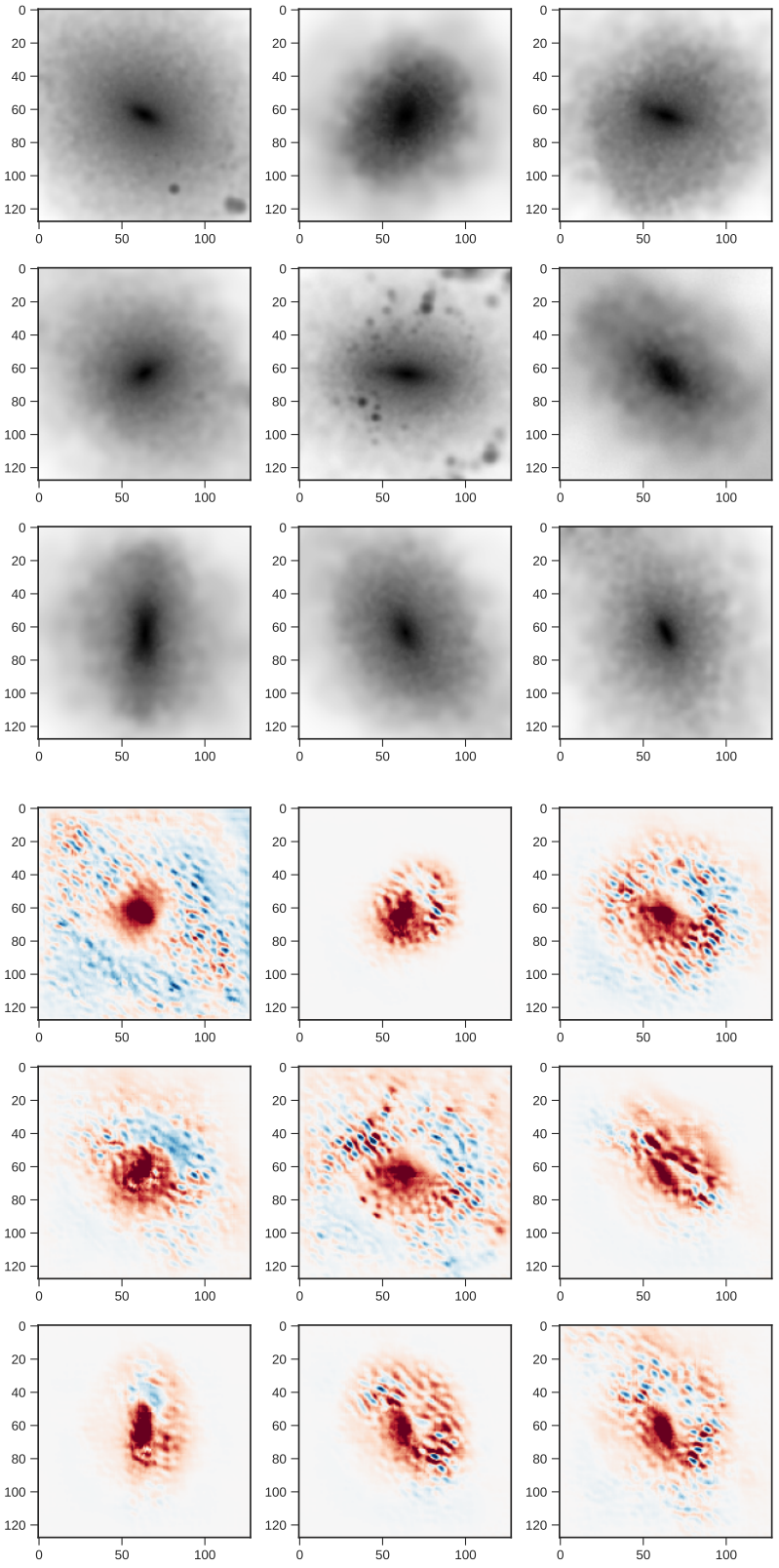


Figure 30: The top panel shows SDSS early type galaxies in a logarithmic normalization with a size of 128x128. The bottom panel shows the attribution maps of the same images computed through the integrated methods. These maps are normalized so they just reflect variations on the values

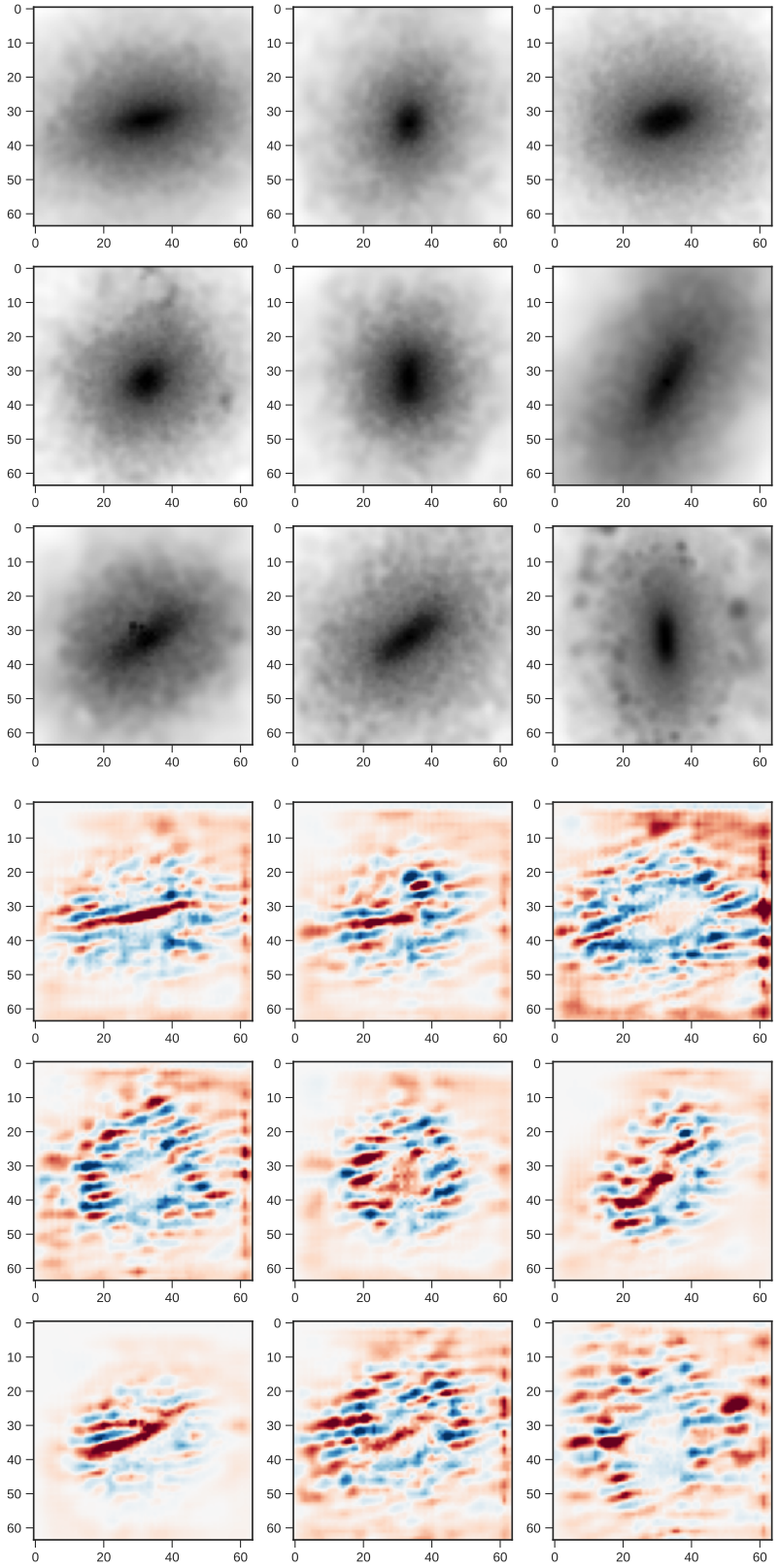


Figure 31: The top panel shows SDSS early type galaxies in a logarithmic normalization with a size of 64x64. The bottom panel shows the attribution maps of the same images computed through the integrated methods. These maps are normalized so they just reflect variations on the values

7 Conclusion and future work

In this work we have analyzed the morphological properties of slow and fast rotators in the local universe. We have used to that purpose observations from the Manga Survey of 4672 galaxies, as well as simulations with 2250 galaxies. It has become clear that the galaxies that fall into the late type group, are mostly fast rotators and the network is able to verify this correctly. When we talk about early type galaxies, where the separation of fast and slow rotators makes sense, both for observations and for simulations, and taking into account that the network is centered and gives more weight to different structures in each case, we obtain very similar results, which are around the AUC of 0.78 which according to Hosmer & Lemeshow (1989) means an acceptable discrimination, being an AUC from 0.5 to 0.7 poor and 0.8-0.9 excellent. It has also been seen that ellipticity is a parameter that can cause the network to do wrong classifications, which is something we expected, because is a visual characteristic that of course the network will take into account. However, it is not the only parameter the network is using since several round systems are properly classified. Also has been shown that elements like the color or the noise are not relevant

As for the future work that can continue this study, many paths are open. The observed and simulated galaxies have shown to have different results, and thanks to the attribute maps calculated through the integrated gradients it is possible to see the difference in the pixels where the network puts more weight, which opens the door to analyze more deeply the reason of these differences. Different "experiments" have been carried out in the project in which parameters such as normalization or image size were modified, but there are still points to be taken into account as if our CNN structure is the most optimal one or whether the iterations we are using are the right ones, that is, to make studies of model convergence, to see how far the maximum precision can go without overfitting, these studies requiring time and greater computing power.

Also the amount and distribution of data we are using can have a lot of influence on the results obtained. It has been shown that in all cases, the number of fast rotators has been substantially higher than the number of slow rotators, which can cause loss of generalization of the model. Also the number of data we are working with in each model does not exceed 5000, if we had for example 100 times more data the model would learn better?

There is also another option that could cancel all the previous ones, it is possible that there is not enough morphological information in early type galaxies and we are in a point where even if we used more means we could not reach models that would make us much better classifications.

References

Abraham R. G., Tanvir N. R., Santiago B. X., Ellis R. S., Glazebrook K., van den Bergh S., 1996, , 279, L47

Abraham R. G., van den Bergh S., Nair P., 2003, , 588, 218

Ancona M., Ceolini E., Öztireli C., Gross M., 2017, arXiv e-prints, p. arXiv:1711.06104

Bacon R., et al., 2001, , 326, 23

Baes M., Verstappen J., De Looze I., Fritz J., Saftly W., Vidal Pérez E., Stalevski M., Valcke S., 2011, , 196, 22

Bernardi M., Domínguez Sánchez H., Brownstein J. R., Drory N., Sheth R. K., 2019, , 489, 5633

Bundy K., et al., 2015, , 798, 7

Bursac M., Milosevic D., Mitrović K., 2019

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

Dieleman S., Willett K. W., Dambre J., 2015b, Monthly Notices of the Royal Astronomical Society, 450, 1441–1459

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

Emsellem E., et al., 2004, , 352, 721

Emsellem E., et al., 2007, , 379, 401

Emsellem E., et al., 2011, , 414, 888

Fischer J. L., Domínguez Sánchez H., Bernardi M., 2019, , 483, 2057

Hosmer D. W., Lemeshow S., 1989, Applied logistic regression. Wiley

Hubble E., 1926, Contributions from the Mount Wilson Observatory / Carnegie Institution of Washington, 324, 1

Huertas-Company M., et al., 2015b, , 221, 8

Huertas-Company M., et al., 2015a, The Astrophysical Journal Supplement Series, 221, 8

Huertas-Company M., et al., 2019, Monthly Notices of the Royal Astronomical Society, 489, 1859–1879

Ioffe S., Szegedy C., 2015, arXiv e-prints, p. arXiv:1502.03167

Kormendy J., Illingworth G., 1982, , 256, 460

Krajnović D., et al., 2013, , 432, 1768

Lecun Y., 1988, in Touretzky D., Hinton G., Sejnowski T., eds, Proceedings of the 1988 Connectionist Models Summer School, CMU, Pittsburg, PA. Morgan Kaufmann, pp 21–28

Lintott C. J., et al., 2008, , 389, 1179

Morgan W. W., 1958, , 70, 364

Naiman J. P., et al., 2018, Monthly Notices of the Royal Astronomical Society, 477, 1206–1224

Nelson D., et al., 2018, , 475, 624

Pakmor R., Bauer A., Springel V., 2011, Monthly Notices of the Royal Astronomical Society, 418, 1392–1401

Pakmor R., Springel V., Bauer A., Mocz P., Munoz D. J., Ohlmann S. T., Schaal K., Zhu C., 2015, Monthly Notices of the Royal Astronomical Society, 455, 1134–1143

Pillepich A., et al., 2017, Monthly Notices of the Royal Astronomical Society, 475, 648–675

Sánchez S. F., et al., 2012, , 538, A8

Shandarin S. F., Zeldovich Y. B., 1989, Reviews of Modern Physics, 61, 185

Springel V., 2010, , 48, 391

Springel V., et al., 2017, Monthly Notices of the Royal Astronomical Society, 475, 676–698

Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R., 2014, J. Mach. Learn. Res., 15, 1929–1958

Sundararajan M., Taly A., Yan Q., 2017, arXiv e-prints, p. arXiv:1703.01365

Vikram V., Wadadekar Y., Kembhavi A. K., Vijayagovindan G. V., 2010, , 409, 1379

Zel'Dovich Y. B., 1970, , 500, 13

de Vaucouleurs G., 1959, Handbuch der Physik, 53, 275

de Zeeuw P. T., et al., 2002, , 329, 513