Galaxy segmentation using U-Net deep-learning algorithm

T. Rey Deutsch[1], L.A. Bignone[2] & S.E. Pedrosa[2]

[1] *Facultad de Ciencias Exactas y Naturales, UBA, Argentina*
[2] *Instituto de Astronomía y Física del Espacio, CONICET–UBA, Argentina*

Contact / tomasreydeutsch@gmail.com

Resumen / La automatización en la segmentación de imágenes es crucial para estudiar la morfología de galaxias en relevamientos de gran escala. En este trabajo utilizamos el conjunto de datos de Galaxy Zoo 3D para entrenar una serie de redes neuronales convolucionales capaces de detectar brazos espirales en imágenes de galaxias. Se contruyeron seis modelos de aprendizaje profundo según el diferente grado de confianza que se tiene para la región marcada como brazo. Utilizamos redes neuronales con arquitectura U-Net, capaces de generar máscaras binarias de brazos espirales con un alto grado de precisión. Esto permite, no solo identificar qué galaxias tienen brazos espirales, sino también ubicarlos a nivel de los píxeles que ocupan y medir su tamaño relativo para seis grados de certeza distintos.

Abstract / Automation in image segmentation is crucial to study the morphology of galaxies from large-scale surveys. In this work we use the Galaxy Zoo 3D dataset to train a series of convolutional neural networks for spiral arms detection in galaxy images. Six different deep-learning models were built according to the levels of confidence for the region marked as an arm. Using an architecture called U-Net, we trained an algorithm capable of generating spiral arms binary masks over a new set of images with high precision. This allows, not only to identify which galaxies have spiral arms, but to easily position the pixels from the spiral arms and measure their relative size for six different degrees of certainty.

Keywords / galaxies: spiral — methods: numerical — techniques: image processing

1. Introduction

Being able to analyze in a fast and efficient way massive amount of data has become of utmost importance in the era of Big Data. The advent of extensive surveys will deliver millions of quality images of galaxies, and the morphological study of these immense datasets can no longer be carried out only via human inspection. However, deep-learning algorithms can perform this task automatically and in reasonable times with the use of GPUs. This is specially advantageous for extremely labor-intensive tasks, such as semantic segmentation of images.

In this work, we focus on using a convolutional neural network (CNN or neural network) to perform image segmentation of galaxies in order to detect spiral arms. Our goal is that, after training these machine learning algorithms, we will be able to predict which pixels in a galaxy image belong to a spiral arm. This information can be used in future works to correlate the presence of spiral arms with star formation rates for different regions of galaxies, along with other physical properties.

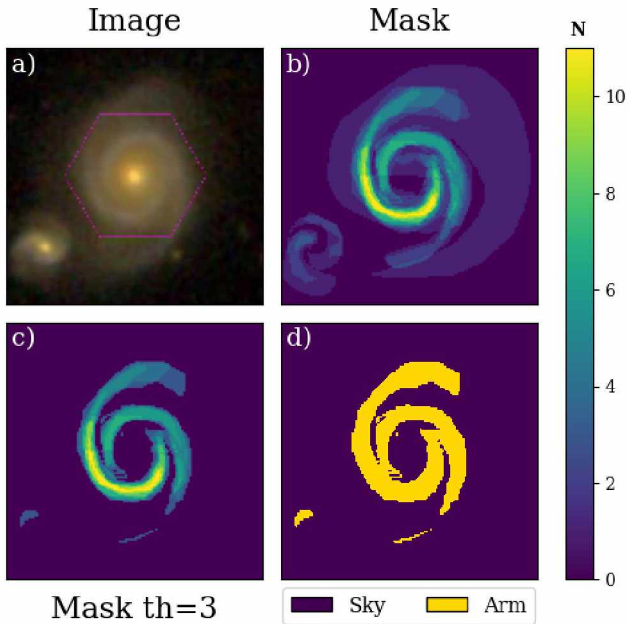

Figure 1: Example of a galaxy from the dataset. MaNGA ID: 1-575229. a) Original image shown to the volunteers. b) Mask voted by the volunteers. c) Mask with pixels that have 3 or more votes. d) Final binary mask used to train the model with $th = 3$.

2. Dataset: Galaxy Zoo 3D

In order to train the neural network for spiral arms detection, the first step was to obtain a proper dataset. The Galaxy Zoo 3D (GZ:3D) is a project that aimed to study the morphology and internal structure (Masters et al., 2021) for ∼30 000 target galaxies of the Mapping Nearby Galaxies at Apache Point Observatory (MaNGA) survey (Bundy et al., 2015), part of the Sloan

Digital Sky Survey, or SDSS IV, (Blanton et al., 2017). To accomplish this, images of these galaxies were shown to volunteers, who noted down the the positions for center of the galaxy, foreground stars, bars and spiral-arms. Each galaxy was annotated by 15 volunteers, completing a total of 29 813 images classified, 7 193 ($\sim 24\%$) of which had at least one pixel marked as belonging to a spiral arm.

With this information, we were able to generate training masks for spiral arms detection, as seen in Figure 1. Images b) and c) show the original mask acquired from GZ:3D and the mask obtained considering only pixels with 3 or more votes towards the spiral arm category, respectively. The colormap used represents the amount of votes each pixel received (N), as indicated by the bar on the right. However, in this work, we train our CNNs with binary masks to test their performance. These masks are created by choosing a fixed threshold value (th) as shown in Sub-figures c) and d). If a pixel has a number of votes $N \geq th$, then it belongs to a spiral arm, and we assign it a value of 1 (yellow pixels in the example). If $N < th$, then the pixel value is set to 0 (violet pixels).

We constructed six training datasets, each one of them with a respective value of $th = 1, 2, 3, 4, 5$ and 6. For the larger values of th, the binary masks used for training have less annotated pixels but more confidence that they belong to a spiral arm. Each dataset was used to train an independent neural network from scratch, resulting in six models, each with different sensitivities to the presence of spiral arms.

3. Neural Networks Architecture: U-Net

All our models use the image of a galaxy as an input and produce a mask with the segmented spiral arms as an output (if there are any). The specific architecture used for our neural networks, U-Net (Ronneberger et al., 2015), was developed in 2015 and is widely used for image segmentation. Previous work has shown that this architecture can be used particularly for segmentation of simulated galaxies (Bekki, 2021) but has never been used in real images. All the models trained in this work have 38 hidden layers (Figure 2) with a total of 34.5 million parameters, which are adjusted during training using the Sparse Categorical Cross-Entropy (SCCE) as loss function.

The original images that were used in the GZ:3D project are customs cutouts of images from the SDSS-I/II legacy imaging described in the target selection of the MaNGA project (Wake et al., 2017). The six models described were trained using a 15GB RAM GPU from the Google Colab free cloud service, which allowed each model to take less than three hours to be trained.

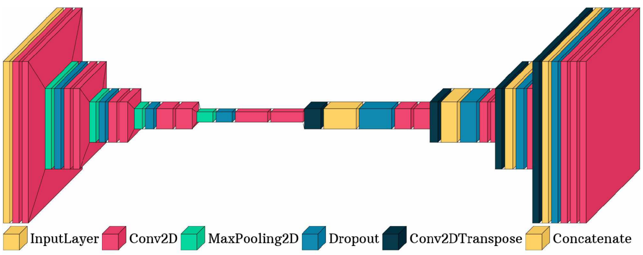

Figure 2: Representation of the U-Net architecture used in this work. Each type of layer has a different color and the relative size between them is respected.

4. Training

The original dataset was divided into a training set with 75% of the images (22 360 in total) and a test set with the remaining 25% (7 453 images), which is used to evaluate the performance of each neural network.

Every model was trained for a fixed number of epochs. In each epoch our deep-learning method separates our training dataset in batches and predicts the potential location of spiral arms for every image in the batch. These predictions are compared with the ground truth annotated by the volunteers and the binary cross-entropy loss is minimized using back-propagation. The top graph in Figure 3 shows an example with the evolution of the loss function at every epoch during the training process of a model (blue line). We also track the accuracy of the model during training for each epoch. Just as we are looking to minimize the loss function we are also looking to maximize the accuracy. Figure 3 also shows the evolution of these metrics for the test set.

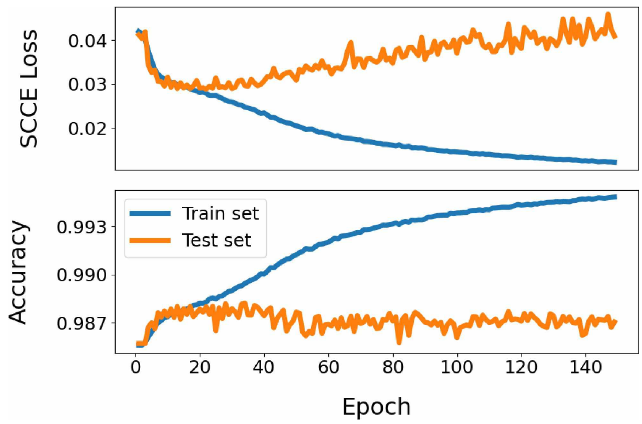

Figure 3: Evolution of the loss function and the accuracy for an example model with $th = 3$, as a function of the epoch. In blue, the training curves. In orange, the progress of these functions with test data that was not used for training.

This test set is used to assess the performance of the trained model with new data (not used for training). We can see that around the 18th epoch the training and the test curves separate. All the models exhibit a similar behavior. This is the point where the neural network starts overfitting the data and the model cannot improve its performance even if it is trained for a larger number of epochs. By that point, models already exhibit a high level of accuracy ($\sim 98\%$) on the test set. We visually inspect each training curve to determine the optimal epoch from where to extract the final snapshot for our models.

5. Predictions

After training, we analyzed the performance of the predictions made by the models. Figure 4 shows an example of a galaxy with two spiral arms where we show the ground truth binary masks and the binary masks predicted by the neural networks for the six different values of th. It can be seen that our U-Nets detect both of the spiral arms in every case and that they are in general agreement with the masks annotated by the volunteers. As the threshold value increases, both the ground truth mask and the predicted mask span a smaller area, as expected since bigger values of th means that only the pixels with more votes are being selected.

In order to evaluate the quality of a prediction and therefore to assess the performance of the models, we analyzed different metrics: accuracy, precision and the Jaccard score. All of them are defined from taking the voted mask as the truth labels, the mask from the neural network as the predicted labels and calculating the confusion matrix. Figure 5 shows as an example the distribution of these metrics for the same model with $th = 3$ from before. We focus on this particular model because it is precisely at the middle range in terms of how selective it is in determining spiral arms pixels. The masks from the test set used to build the histograms are only those where the volunteers and U-Net agreed that there was a spiral arm, since, in most cases (73% of the total 7 453), they both agreed that there was none. This left a total of 1 347 galaxies with their score computed (18%) for this particular th.

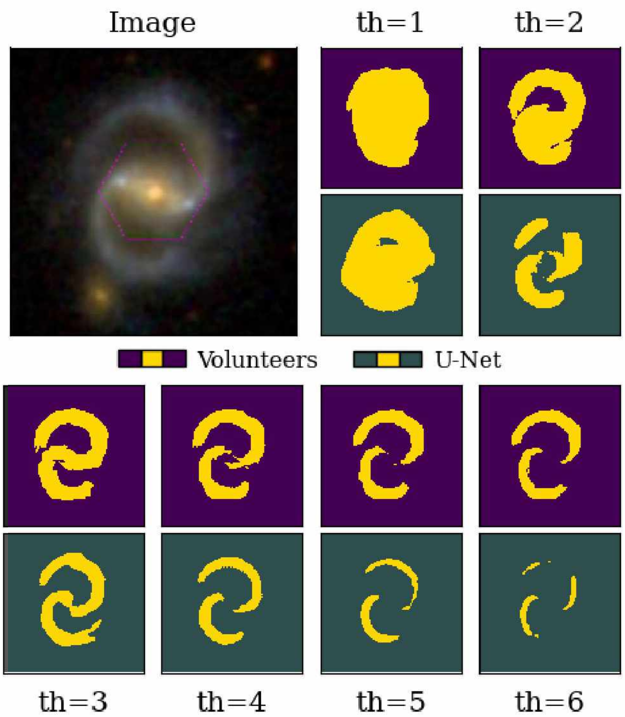

Figure 4: Example of a galaxy with two spiral arms, MaNGA ID: 1-590142. For every model with different th we plot the ground truth binary mask voted by the volunteers and the one predicted by our neural networks from the original image.

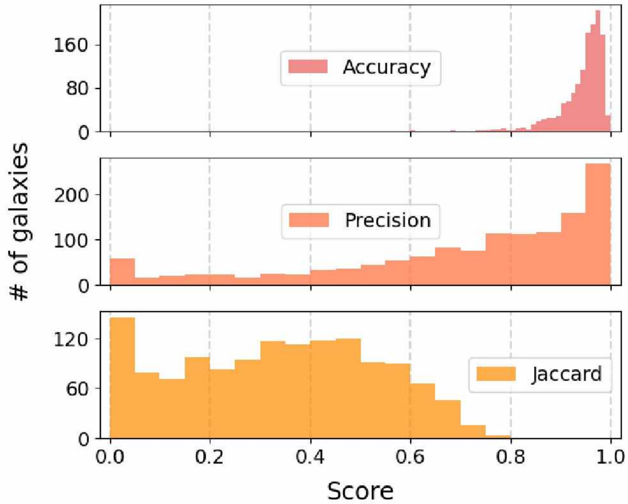

Figure 5: Distributions of the 1 347 images used to asses the neural networks performance. The model shown as an example is the one trained with $th = 3$.

The high and narrow peak value in the accuracy distribution is due to the fact that most of the pixels in both masks belong to the background sky (violet pixels in Figures 1 and 4) and are correctly classified. On the other hand, we can see from the peak of the precision distribution that there is also a high proportion of pixels predicted as belonging to a spiral arm that are well classified. Finally, the Jaccard score (TP/[TP+FN+FP]) indicates that the correctly classified pixels represent, in average, the 34% of the region that results from joining the spiral arms of the volunteers and U-Net. Table 1 shows the mean values for this example.

Table 1: Mean values of the metrics from the example in Fig. 5 with their respective standard deviations.

Accuracy	Precision	Jaccard Score
0.95 ± 0.04	0.71 ± 0.27	0.34 ± 0.20

6. Conclusions

We were able to train six convolutional neural networks that identify spiral arms with different levels of confidence and a high value of precision. We expect to use these models in future works to study the physical properties of spiral armed galaxies over new images.

References

Bekki K., 2021, A&A, 647, A120
Blanton M.R., et al., 2017, AJ, 154, 28
Bundy K., et al., 2015, ApJ, 798, 7
Masters K.L., et al., 2021, MNRAS, 507, 3923
Ronneberger O., Fischer P., Brox T., 2015, U-net: Convolutional networks for biomedical image segmentation
Wake D.A., et al., 2017, AJ, 154, 86