MORPHOLOGICAL CLASSIFICATION OF S-PLUS GALAXIES USING A ENSEMBLE OF CONVOLUTIONAL NEURAL NETWORK

Natanael Magalhães Cardoso

Profa. Cláudia Mendes de Oliveira

Instituto de Astronomia, Geofísica e Ciências Atmosféricas, USP – SP

nauxmac@gmail.com

Objectives

Morphological classification is the cataloging of galaxies according to their visual appearance and it is linked to their physical properties. A morphological classification made through visual analysis is sensible to differences in evaluation among diferent human volunteers. For this reason, systematic, objective and easily reproducible classification of galaxies has been gaining importance since the astronomer Edwin Hubble created his famous classification method [1]. In this work, we combine accurate visual classifications of the GalaxyZoo[1] project with Deep Learning methods. The goal is to find an automatic and efficient technique at human performance level classification based on analysis of DR1 images of the Southern Photometric Local Universe Survey (S-PLUS) [2] and publish a catalog with not yet classified galaxies. For this, a Deep Learning model was developed through a set of four convolutional neural networks using the Stacking Ensemble technique [3].

Materials and Methods

The proposed model classifies the morphology of galaxies based on their images. The images of each of the 12 bands (details about each band in Table 1 of [2]) were obtained from the S-PLUS project database[2], grouped in 3 bands and converted to RGB color space. This reduction step allows the use of the information from the 12 bands, mapped into only 3 bands, causing a reduction in the training time of the models. As the proposed technique consists of a supervised training, in addition to the data to be classified, GalaxyZoo human classifications are also necessary so that the algorithm can relate the patterns between the images and their respective classes. The training, validation and testing sets were created, with approximately 30% of elliptical galaxies and 70% of spiral galaxies. Techniques such as random resampling and class weighting were used to address the imbalance. Another set, called *blind*, was created with galaxies not yet classified by GalaxyZoo and which will be classified by the model. All these sets have similar redshift distribution and magnitude.

To train the classifiers, convolutional neural networks with weights initialized from a pre-training on ImageNet[3] were used. The data pre-processing, the hyperparameter adjustment and the training were done for each classifier independently. Among the trained models, the ones that obtained the best classification performance are based on the InceptionResNetV2, EfficientNetB2, DenseNet-169 and DenseNet-201 architectures. To take full advantage of the potential of these four classifiers, another model (called meta-model) was trained to receive the predictions of each classifier and return a single prediction. Thus, a Ensemble of classifiers is formed. This procedure is summarized in the diagram in Figure 1.

---

[1] http://galaxyzoo.org
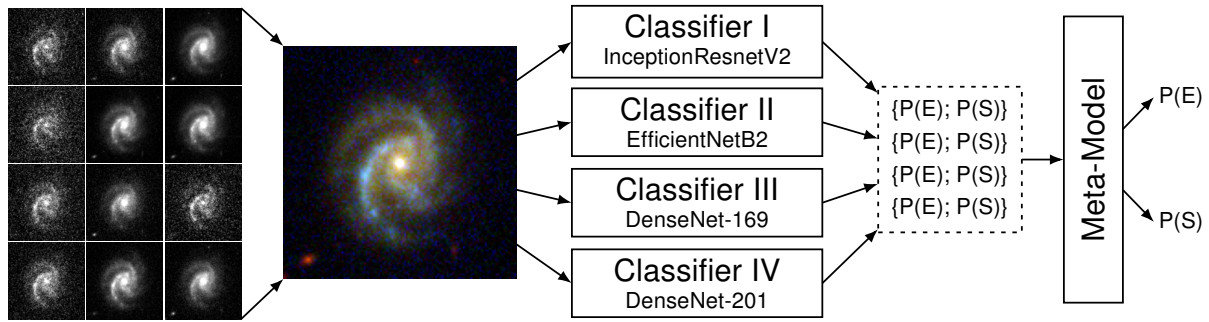[2] http://splus.cloud

[3] http://image-net.org

Figure 1: Architecture diagram. From left to right, the 12 images of each band are grouped into a single RGB image, which is the input for the individual classifiers. These classifiers have the function of extracting visual features from the image and returning the probability of being elliptical or spiral. The metamodel has the function of combining the predictions of the classifiers into a single, more robust prediction.

Results

One way to evaluate the model quantitatively is from the Receiver Operating Characteristic Curve (ROC) in the test set, as shown in Figure 2. From this graph it is possible to draw conclusions about the classification potential of the model. The area under the curve, whose maximum value is 1, indicates the probability of correct classification by the model. Furthermore, the faster the curve reaches the top, the greater the degree of separability of the model, that is, the greater the model's ability to distinguish be-

tween classes. As expected, the Ensemble (blue curve), which is the union of the four classifiers, outperforms the individual classifiers, reaching an AUC of 0.998.

Conclusions

From this project, a fully developed and operational Deep Learning model was obtained with high potential for morphological classification between elliptical and spiral galaxies from S-PLUS images, as well as a catalog with automated classifications of 2536 galaxies with magnitude less than 17 without classifications on Galaxy-Zoo, both publicly available[4]. In addition, a scientific paper about this project is in the final writing stages.

References

[1] E. P. Hubble. "Extragalactic nebulae". In: (Dec. 1926).

[2] C. M. de Oliveira et al. "The Southern Photometric Local Universe Survey (S-PLUS): improved SEDs, morphologies, and redshifts with 12 optical filters". In: (Aug. 2019).

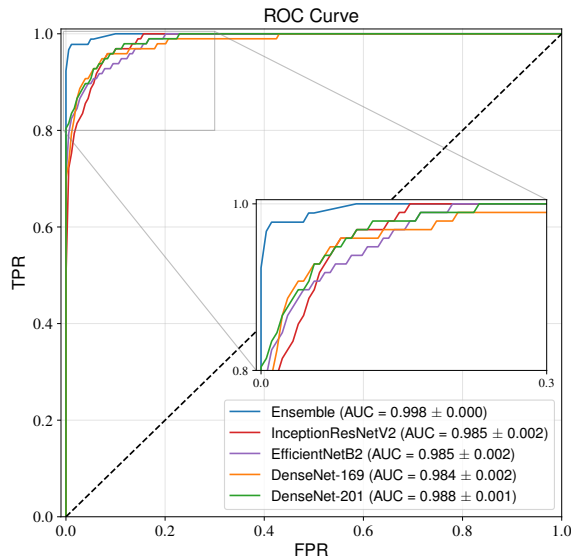[3] D. H. Wolpert. "Stacked generalization". In: (1992).

Figure 2: ROC curve for each classifier and for the Ensemble. Each curve represents the median of 60 measurements, the area under the curve (AUC) and its standard deviation are in the legend for each curve.

---

[4]http://natanael.net