



Mask galaxy: Morphological segmentation of galaxies

H. Farias ^{a,b,*}, D. Ortiz ^b, G. Damke ^{b,c,d}, M. Jaque Arancibia ^{b,d}, M. Solar ^a

^a Departamento de Informatica, Universidad Tecnica Federico Santa Maria, Santiago, Chile

^b Instituto de Investigacion Multidisciplinaria en Ciencia y Tecnologia, Universidad de La Serena, Raul Bitran 1305, La Serena, Chile

^c AURA Observatory in Chile, Avda. Juan Cisternas 1500, La Serena, Chile

^d Departamento de Fisica y Astronomia, Universidad de La Serena, Avda. Juan Cisternas 1200, La Serena, Chile



ARTICLE INFO

Article history:

Received 13 February 2020

Accepted 20 August 2020

Available online 28 August 2020

Keywords:

Computer-vision

Mask R-CNN

Deep Learning

Morphological-segmentation

ABSTRACT

The classification of galaxies based on their morphology is instrumental for the understanding of galaxy formation and evolution. This, in addition to the ever-growing digital astronomical datasets, has motivated the application of advanced computer vision techniques, such as Deep Learning. However, these models have not been implemented as single pipelines that replicate detection, segmentation and morphological classification of galaxies directly from images, as it would be made by experts. We present the first implementation of an automatic machine learning pipeline for detection, segmentation and morphological classification of galaxies based on the Mask R-CNN Deep Learning architecture. This state-of-the-art model of Instance Segmentation also performs image segmentation at the pixel level, which is a recurrent need in the astronomical community. We achieve Mean Average Precision (mAP) of 0.93 in the morphological classification of Spiral or Elliptical galaxies for a set of 239,639 objects from the Galaxy Zoo sample and JPEG images from the Sloan Digital Sky Survey. As a direct use of segmentation, we test the model for deriving centroids of extended sources, reaching a precision better than 1.0 arcsecond. We also test the network under additive Gaussian noise. We find that the Mask R-CNN network is able to perform with accuracy over 92% for a distribution scale of 76.5 counts.

The repository with the model code is in the following url: https://github.com/hfarias/mask_galaxy
 © 2020 Elsevier B.V. All rights reserved.

Contents

1. Introduction.....	3
2. Dataset creation.....	3
2.1. Dataset labels: The Galaxy Zoo datasets and morphological classes	3
2.2. Dataset images: The Sloan Digital Sky Survey	4
2.3. Dataset masks: Automated data annotation	4
3. Deep learning architecture.....	4
3.1. Mask R-CNN	5
3.2. Loss function	7
4. Model training	8
4.1. Learning strategy	8
4.2. Data augmentation	9
5. Results.....	10
5.1. Classification and location	10
5.2. Comparison of network and human classification (GZ1 and GZ2)	11
5.3. Instance segmentation	11
5.4. Determination of galaxy centroids: A mask R-CNN application test case.....	14
5.5. Automatic classification, location and segmentation pipeline	14
6. Contribution and future work.....	16
6.1. Future work	16
Declaration of competing interest.....	17

* Corresponding author at: Departamento de Informatica, Universidad Tecnica Federico Santa Maria, Santiago, Chile.
 E-mail address: humberto.farias@usm.cl (H. Farias).

Acknowledgments	17
References	17

1. Introduction

Galaxy morphology is, arguably, the major indicator of the physical processes that drive the evolution of galaxies. Hence, morphological classification of galaxies is essential for the understanding of the formation and evolution of galaxies in the universe. The foundations of morphological classification were proposed by Hubble (1926) as a purely descriptive system. Traditionally, the classification has been performed either visually by experts or through the automated extraction of features; this is, the measurement of proxies such as concentration index, surface brightness profile, color, etc., which correlate with different morphological characteristics (e.g., Conselice, 2003). However, both methods have pros and cons. For instance, experts can reach high accuracy in recognizing structures and shapes, but can only inspect a very limited amount of data. On the other hand, features can be obtained massively by computers, but results are not always satisfactory. Nevertheless, the advent of the Charged Coupled Devices (CCDs) and the increase of computational power propelled the use of computational techniques for the analysis of astronomical data. For over two decades, the identification and segmentation of individual and extended sources in astronomical images has relied on semi-automated software, mainly SExtractor (Bertin and Arnouts, 1996). The advantages of this approach are the capacity to classify extended sources (usually galaxies) in large volumes of data although at the risk of using proxies of morphology, which do not necessarily correlate with morphological types.

Recent advances in Machine Learning and Computer Vision techniques have demonstrated satisfactory results for the identification and automated classification of astronomical objects in images (for a summary, see Baron, 2019). Among the various types of data generated in astronomy, large area surveys are especially useful for the application and testing of such machine learning models, partly because the observations generated by a survey exhibit a certain level of homogeneity in their characteristics and in the technical configurations of the instruments from which they were generated. Some good examples of this are the Sloan Digital Sky Survey (SDSS hereafter; York et al., 2000) and the Gaia mission (Brown et al., 2016).

This decade will see a series of astronomical mega-projects coming into operation producing complex data whose dimensionality and volume will exceed any current scale. This requires the application of a new generation of machine learning models, more robust and faster than those currently applied. These projects include the ground-based Vera C. Rubin Telescope and its survey (LSST, Ivezić et al., 2011), space-based observatories as the Wide Field Infrared Survey Telescope (WFIRST, Spergel et al., 2015) and James Webb Space Telescope (JWST, Gardner et al., 2006). The LSST will generate about 36 Terabytes (TB) of data per night, and the data will have a sequential relationship between images, from extended and point sources. This will require the application of models that also incorporate sequential relationships in their architecture. The speed of classification is also relevant, as the LSST data stream will deliver a 15 GB image every 15 s. The requirement is even more stringent if we consider that one of the scientific objectives of the LSST is to generate alerts for transients in less than 1 min after the shutter is closed. It is expected that the LSST will generate 10,000,000 alerts every night, thus prompt classification is a characteristic that must be considered in the new proposed models.

When narrowing the problem to the morphological classification of extended sources, such as galaxies, the approaches that have shown the most outstanding results are those based on the latest advances in computational vision techniques. The architecture of these networks bases its operation on treating astronomical images as three-channel (RGB) images, even though the multi-band data of the surveys may be more complex and rich. In addition, current models usually address classification only by indicating to which category a certain astronomical source belongs to. Finally, the most advanced models can identify and localize a galaxy by incorporating a bounding box over the astronomical source (e.g., González et al., 2018), but it is not possible to indicate whether or not a certain pixel belongs to the identified source; i.e., the level of source segmentation is not reached.

Our proposal aims to implement a unique pipeline for classification, location, and segmentation of galaxies, according to their morphology. This pipeline is based on a Mask R-CNN network (He et al., 2017) that allows to incorporate a pixel-level segmentation of an object already classified and also located by the model. The proposed model is based on the fact that there is no need for a color adjustment to the dataset images. This contribution is organized as follows: we present the creation of the dataset using images from the SDSS, the Galaxy Zoo 1 (Lintott et al., 2008, 2011) labels, and the Galaxy Zoo 2 (Willett et al., 2013) labels in Section 2. Then, we introduce the definition of the Deep Learning architecture in Section 3. In Section 4, we show a detailed analysis of the learning strategy based on the transfer of learning, differential learning rates, and data augmentation. The general results are presented in Section 5 and, finally, the contribution of this work and future steps are addressed in Section 6.

2. Dataset creation

One of the considerations for the use of convolutional neural network (CNN), in general, is the need of a training dataset of images able to satisfy the following three requirements. First, the objects in the dataset have to contain a label indicating the class to which each object belongs to. Second, CNNs require all classes to be equally represented in the dataset; and third, the need to split the dataset into training and validation sub-sets. This last requirement implies that the dataset needs to contain many images – usually several thousands. In the case of Mask R-CNN, it is also necessary to incorporate an additional element: the mask. This last element is a key factor for the architecture, because it provides the segmentation shape (and implicitly the location) of the objects.

In the context of classification of astronomical sources, there is no validated, standardized and prepared dataset for Deep Learning training. To overcome this limitation, we need to create a dataset of images, labels and masks. Below we describe the steps for the creation of the training set.

2.1. Dataset labels: The Galaxy Zoo datasets and morphological classes

We utilize the catalog of morphological classifications from the Galaxy Zoo 1 (GZ1, hereafter), and the Galaxy Zoo 2 (hereafter GZ2) datasets for the network training and validation. These datasets have been largely used in the astronomical community because they fulfill the requisites of volume and labeled data required for training (e.g., Barchi et al., 2020). Both GZ1 and GZ2 were Citizen Science projects, where approximately 10^5 and $8 \times$

10^4 volunteers performed morphological classification of around 9×10^5 and 3×10^5 galaxies, for GZ1 and GZ2, respectively, using RGB images presented through a web portal.¹ The astronomical images came from the SDSS, which carried out a photometric campaign of a large portion of the northern sky in five photometric bands. The color images presented in GZ1 and GZ2 were composed using the (*g, r, i*) photometric bands, mapped as the JPEG (R,G,B) channels, respectively, following the prescription described in Lupton et al. (2004).

The GZ1 morphological classification scheme includes Spiral (clockwise, counter-clockwise and edge-on), Elliptical, Merger, and Star or Don't Know (e.g., artifact) labels. A second more complex classification scheme is presented in the GZ2 dataset. Because the latter scheme is a more detailed classification, we intersect the GZ1 and GZ2 catalogs of galaxies with spectroscopic redshifts, using the *dr7objid* label. The merged catalog contains 239,639 objects. For training and validation, we utilize only the labels from the GZ1 catalog. However, we assess the network performance using galaxies assigned to the same class (Spiral or Elliptical) in both catalogs.

It is important to note that the Galaxy Zoo samples only include the nearest, brightest, and largest systems. As a result, we will not train the network to detect star-like objects. Both Galaxy Zoo projects applied several cuts to ensure that the sub sample of SDSS only contains fine morphological features that can be resolved and classified by volunteers. Specifically, the sample of galaxies was pre-filtered, using the following parameters from the SDSS Data Release 7: a Petrosian half-light magnitude brighter than 17.0 in the *r*-band after Galactic extinction correction was applied; the Petrosian radius lies between (*petroR90*) $> 3[\text{arcsec}]$, and redshift between $0.0005 < z < 0.25$ (Willett et al., 2013).

To minimize uncertainties in the classification, we generate an intermediate catalog that only includes objects classified as either Spiral (S) or Elliptical (E), and whose debiased fraction of votes is greater than 80%. This results in a catalog with 33,809 and 81,406 objects, classes S and E, respectively. The remaining 124,424 objects are labeled as Uncertain (U), following the GZ1 classification. The final dataset was created by randomly selecting 12,000 objects from this intermediate dataset. This dataset is made up of 6000 spiral-type galaxies and 6,000 elliptical-type galaxies, for the purpose of having a balanced dataset. Galaxies labeled as Uncertain (U) were not included in the dataset because they do not constitute a morphological class, but they correspond to galaxies that did not obtain 80% of the votes to be assigned as elliptical or spiral. The galaxies marked as U will be used in Section 4.1 to validate the capacity of the model to classify and compared against the ability of a human classifier.

The 12,000 objects were divided as follows: Training set (8000 objects), Validation set (2000 objects) and Test set (2000 objects). These three subsets are composed evenly by the two classes (spiral and elliptical).

2.2. Dataset images: The sloan digital sky survey

After the creation of a balanced dataset of labels from the Galaxy Zoo catalogs, we generate the respective images that correspond to each labeled object. We utilize images from the SDSS, accessed through the SkyServer (Szalay et al., 2002) SDSS DR15 Finding Chart Tool web service. For each labeled object, we download a JPEG image centered at the object coordinates. We set the scale to match the native pixel scale of the SDSS imager ($0.396127 [\text{arcsec/pix}]$), and the image width and height to 256 pixels, which equals 101.4 arcseconds. The images provided by the web service are created by the SDSS through the ImgCutout web service (Nieto-Santisteban et al., 2004), which pre-processes the SDSS data to produce 24-bit JPEG images on demand.

2.3. Dataset masks: Automated data annotation

The final step to build the dataset is the creation of the masks. The mask serves the purpose of delineating the labeled objects in an image that will be used after for training and validation. The mask creation can be either a manual or an automated process. Several works use manual annotation tools, such as VGG Image Annotator (Dutta and Zisserman, 2019), LabelBox (Anon, 2020d), LabelMe (Torralba et al., 2010), COCO UI (Anon, 2020b) or RectLabel (Anon, 2020a). These tools allow the user to create the mask for each labeled object in an image manually. This process is illustrated in the middle panel of Fig. 1, where a galaxy is marked manually over a SDSS image using the VGG tool.

Fig. 1 presents the results of the automated mask creation method, using the center of the image as a reference. The masks were created using an automated annotation process, based on image processing algorithms from the Scikit-Image (Van der Walt et al., 2014) package in Python. The object, located in the center of the image, was identified using the Otsu method. This method is used in computer vision to perform segmentation by dividing an image into significant regions or objects, and then delivering a binary image as a result which is interpreted as a boolean mask. The basic idea behind the Otsu method is to select an optimal gray level threshold value that will separate the background and objects of interest in an image based on their gray level distribution. In addition to the method described above, we used the imantics library (Anon, 2020c) to perform a semantic analysis of the image and to have greater precision in the segmentation of the object.

3. Deep learning architecture

Applications of Deep Learning are delivering promising results in astronomy, especially in object recognition methods that apply computer vision models to identify objects in digital images. These techniques are grouped into four categories: Image Classification, Object Localization, Object Detection and Instance Segmentation. This categorization is based on the results that each method can achieve. In Fig. 2, we illustrate the result of each of these categories. The first category addresses one of the central problems in computer vision: predicting the class or type of an object in an image. The first applications of computer vision in astronomy correspond to Image Classification (e.g., Sandler et al., 1991; Odewahn et al., 1992; Dieleman et al., 2015; Banerji et al., 2010). The methods in the second category (Object Localization) answer the question if a certain object is in an image, indicating its location with a bounding box additionally.

The other two categories, Object Detection and Instance segmentation, represent the state of the art of Object in image recognition methods. Object Detection methods also indicate the class of an object as Image Classification methods do, showing its location with a bounding box as well. Within Object Detection we find two main families of architectures: Region Proposals (R-CNN; Girshick et al., 2014), Fast R-CNN (Girshick, 2015), Faster R-CNN (Ren et al., 2015) and You Only Look Once (YOLO; Redmon et al., 2016; Redmon and Farhadi, 2017). For instance, AstroCV is an implementation based on YOLO architecture for morphological classification of galaxies which emphasizes speed when performing this task.

The most powerful category is the Instance Segmentation. The methods in this category combine the characteristics of Object Detection with those of Semantic Segmentation. This is an intermediate category that determines what pixels in the image belong to a given class. The objective of the instance segmentation is to segment or separate each “instance” of a class in an image at pixel level. A key difference between Object Detection and Instance Segmentation is that the bounding box used

¹ Currently at <https://www.zooniverse.org/projects/zookeeper/galaxy-zoo/>.

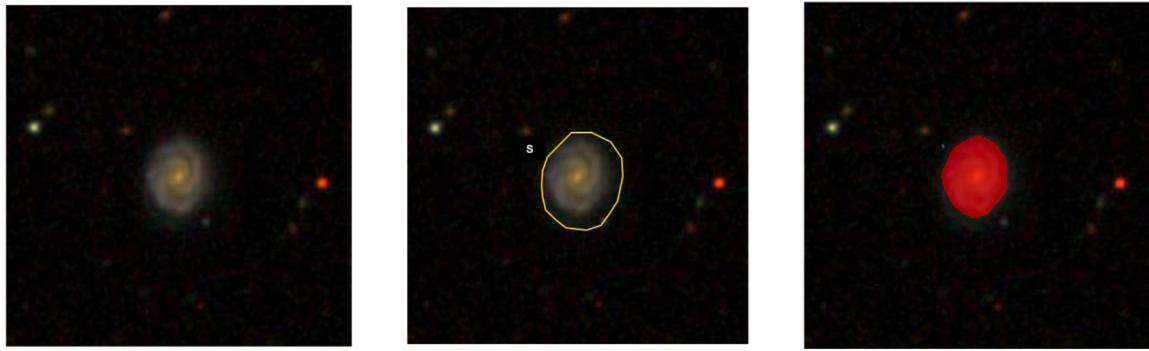


Fig. 1. Creation of a training mask over an image. Left: Original RGB image obtained from the SDSS SkyServer, centered at the Equatorial coordinates of the object listed by Galaxy Zoo. Center: Mask created following a manual process using the VGG Image Annotator tool. Red dots are marked by the user, while green lines are linearly interpolated between the dots by the tool. Right: Mask created by an automatic process (red) overlaid on the input image. The mask was calculated using image manipulation libraries in the Python programming language, for the central object in the image. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

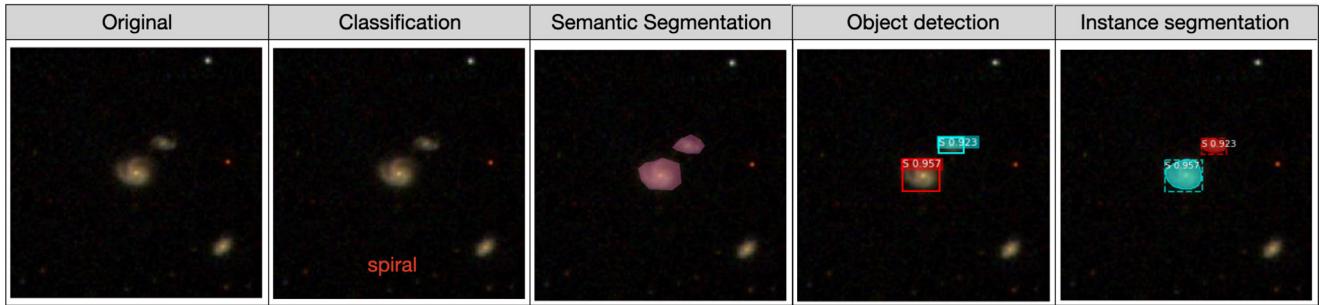


Fig. 2. Examples of the output of four different Computer Vision models applied to an astronomical image of galaxies.

for defining an object in the former is complemented with a segmentation mask in the latter. This last characteristic is a relevant feature for the morphological classification of galaxies, because it makes possible to extract each object separately without any background. This is not possible, for example, in semantic segmentation-based methods.

There are two especially outstanding architectures in Instance Segmentation: U-Net (Ronneberger et al., 2015) and Mask R-CNN (He et al., 2017). U-net is fully-convolutional network architecture. In other words, the network does not have a fully-connected layer. It owes its name to the way its layers are organized. This form is the result of network consistency, that is, a contracting path and an expansive path. For example, an architecture based on U-Net was used to mitigate radio frequency interference (RFI) signals in radio data (Akeret et al., 2017). The results, in terms of its precision-recall curve, are up to the standard method used by radio astronomers.

Mask R-CNN extends Faster R-CNN by adding the functionality to pixel-level image segmentation. It is based on two stages: Firstly, it scans the image to generate proposals. This process seeks to find areas with a high likelihood of containing an object; then it classifies these areas (proposals). Secondly, bounding boxes and masks are overlaid on the original image. Mask R-CNN is the state of the art in Instance Segmentation. A major reason to choose Mask R-CNN instead of U-NET-based architecture is whether or not the scientific goals of the problem need the polygonal mask that Mask R-CNN generates. To put it another way, whether or not it is necessary to identify each pixel (and its position) belonging to a feature. We present an example of its application to astronomical data in Section 5.4.

Arguably, the standard tool used for segmentation of extended sources is SExtractor. Although, this software requires expert knowledge for its use, given the multiple parameters that

must be set for successful results. Moreover, recent development in deep learning models are providing better results than SExtractor. For instance, Boucaud et al. (2019) applied a model to solve galaxy deblending, that is, the separation of overlapping galaxies. Their results indicate that the proposed model outperforms SExtractor by a factor of 4.

3.1. Mask R-CNN

The Mask R-CNN is the state of the art of the Region-based Convolutional Neural Networks (R-CNN). The goal of the R-CNN family architecture is to solve the problem of object detection. To achieve this purpose, broadly speaking, the process is divided into: Generate the Region of Interest (RoI); proposals; and, classify these RoIs. In particular, finding the RoIs presents high computational complexity. In order to deal with this complexity and solve the problem, a series of architectures based on R-CNN have been presented.

These architectures are Fast R-CNN, Faster R-CNN and finally Mask R-CNN. The main difference among them is the substantial increase in the speed of training and inference in the newer networks. Moreover, Mask R-CNN is the evolution of Faster R-CNN that adds the segmentation capability to the model. Thus, in addition classification and localization, the network adds instance segmentation as a third output to the model, as shown in Fig. 3. The first part of the Mask R-CNN architecture is a pre-trained Convolutional Neural Network on image classification tasks. This network is called the Backbone and it is responsible of generating the space of features (hereinafter feature map) of the model. We use a Residual Deep Neural Network (Resnet) (He et al., 2016) with 101 layers and the MS-COCO dataset (Lin et al., 2014) pre-trained weights as the backbone. The feature map acts as an input for the following stages of the model. In Fig. 4 we

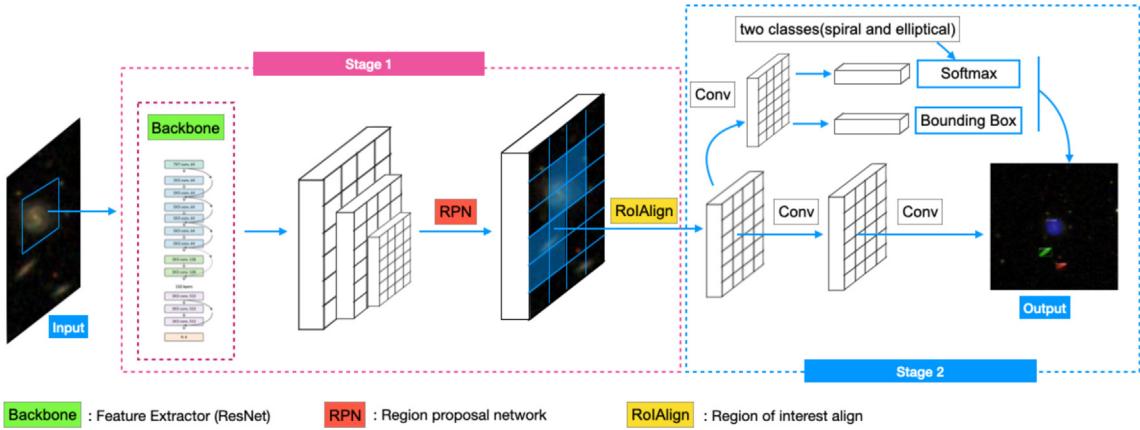


Fig. 3. Mask R-CNN model architecture with two stages: First, it scans the image to generate proposals. Later it classifies these areas (proposals). Finally, bounding boxes and masks are created on the original image. Diagram based on He et al. (2017).

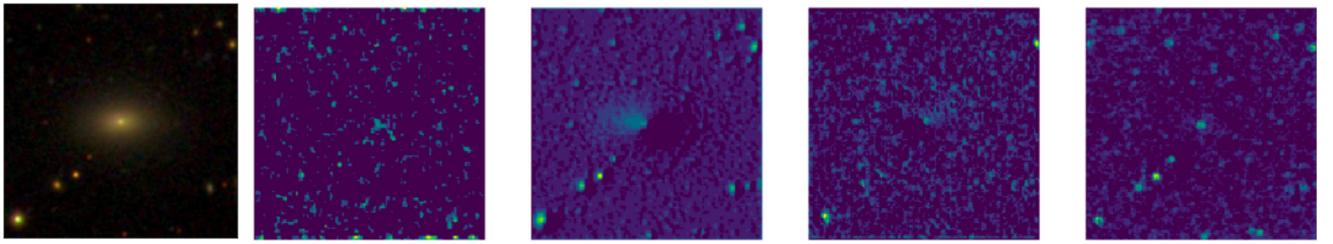


Fig. 4. Backbone Feature Map. The leftmost panel shows an input image passed to the network. The other four panels show different feature maps produced by the backbone in block 2 of a resnet-101. In the second map (center panel) the three galaxies are clearly visible, whereas the smallest sources are clearly highlighted in the rightmost panel.

show the feature maps of the backbone. It can be seen how different layers are sensitive, after training with our dataset, to astronomical objects at different spatial scales.

The next stage in the model is the Region Proposal Network (RPN). A RPN is a lightweight fully-convolutional network that operates as a binary classifier. The process consists of running a small $n \times n$ spatial window on the feature map of each image to generate a set of potential regions called *proposals*. These proposed regions or proposals have different sizes because they depend on the size of the spatial sliding window used to generate them. The size of such a sliding window defines an anchor box. A network has many anchor boxes that have predefined locations and scales relative to images. The size of the anchor boxes is closely related to the nature of the problem. In particular, we utilize anchors of sizes {4, 8, 16, 32, 128} pixels. We present the results of applying different anchor sizes to an image in the left panel of Fig. 5. The three galaxies are clearly identified in this panel.

Any model that aims to classify galaxies should consider their different sizes. For instance, a Milky Way-like galaxy is one hundred thousand light years in diameter approximately, whereas dwarf galaxies are a few percent of that size. On the other hand, galaxies such as M100 are approximately 52.5 million light years diameter. Furthermore, galaxies will exhibit a particular angular size according to their distances. Following this rationale, the size of the anchors that we use considers the size of the input images and the variety of sizes of the galaxies that could be found in such images. In particular, we choose small anchors considering the small size of the objects (galaxies) in our sample, the pixel scale of the data plus the small size (256×256 pixels) of the training images as well.

The RPN generated anchors at three different scales and three different aspect ratios. Each of these anchors has an objective ness base score of two classes: object/no-object. Finally, non-maximum suppression (NMS) is applied to anchors in order to

remove boxes that have not overlapped. In practice, a value is calculated using an Intersection over Union (IoU) metric and boxes that do not reach a predefined threshold are removed. IoU measures the overlap between 2 boundaries, as can be seen in Fig. 6. IoU is the relationship between the box that represents the correct position of an object (ground truth), and another box that represents the prediction of the model. IoU is defined as a fraction represented in Eq. (1) and represented as boxes B1 and B2 in Fig. 6. A threshold of 0.8 IoU was used in the model to filter RPN proposals.

$$\text{IoU} = \frac{B1 \cap B2}{B1 \cup B2} \quad (1)$$

Next, the RPN proposals that pass the binary classification process (object/no-object) go to the next stage with the name of ROI. The left panel of Fig. 5 shows the RPNs contributing to the training (the solid line) and those that do not contribute (the dotted lines). After applying Non-max Suppression, there are fewer anchors than the number of candidates. These filtered anchors are renamed as Regions of Interest (Rois). These Rois go through two processes called RoI Classifier and Bounding Box Regressor. The result of these processes generates two outputs for each RoI: The Class of the object and the Bounding Box Refinement. The Classifier, unlike the RPN, incorporates the real class that the object belongs to; in this case, whether it is an elliptical or a spiral galaxy. In the case of the Bounding Box Regressor, it is necessary to make a size adjustment because of the different sizes that the boxes generated by RPN have. Also, the pixel-level segmentation stage requires much more fine-grained alignment with the bounding box. Therefore, it is necessary to obtain fixed-size boxes first. For this purpose, the model incorporates the stage called The RoIAlign layer. The RoIAlign mechanism is based on cropping a part of an original feature map selected in a bounding

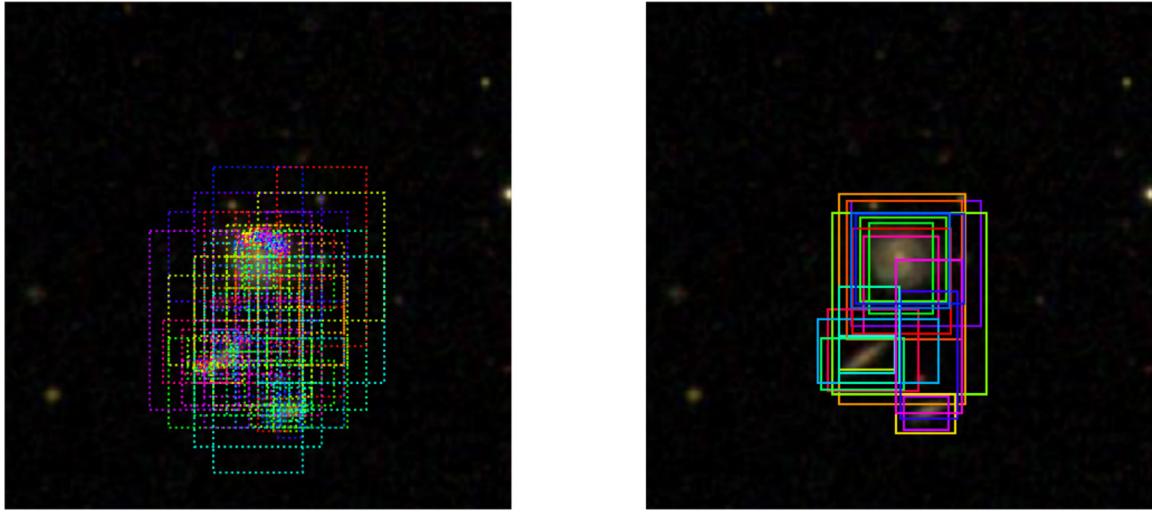


Fig. 5. RPN refinement process: This stage of the model transforms the boxes proposed by the RPN into a ROI. Left: Anchors before refinement. Right: Refined anchors after non-max suppression result in RoIs.

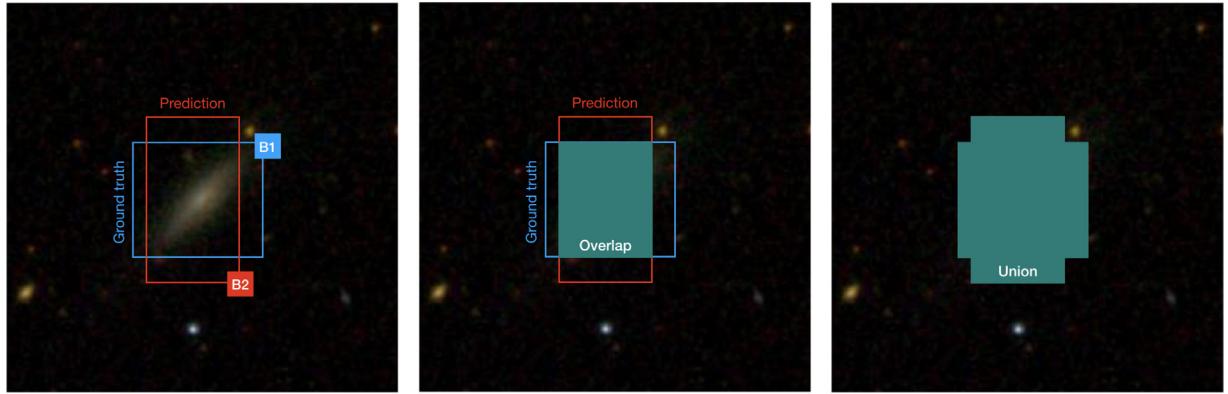


Fig. 6. Intersection over Union (IoU) visual representation.

box and resizing it to a fixed-size. [Table 1](#) lists the two techniques used for this size change: square and crop.

Finally, as indicated before, a unique element of the Mask R-CNN architecture for the objective of this work is the ability to segment the model. The mask is generated by a fully convolutional network receiving as input the ROI selected by the ROI Classifier. In other words, all the regions that contain objects of the classes that exceed the minimum parameterization accuracy and with bounding boxes of the defined size. An important aspect to consider for this model is the particularity that astronomical sources do not exhibit sharp edges. In this sense, the masks achieve the pixel level and is represented as floating numbers, achieving thus more precision to whether or not a pixel belongs to the classified object. This is a relevant aspect for the automatic segmentation of astronomical sources generated by the model. [Fig. 7](#) shows the step-by-step procedure followed by the entire model being applied to an SDSS image. The final result is verified for the classification, location and segmentation of the three galaxies present in the image. All this process occurs within a single automatic pipeline based on machine learning.

The model that we present was developed in the Python programming language, using the Tensorflow framework and the Keras library. The implementation is based on the [Abdulla \(2017\)](#) library. We summarize the configuration of the networks and their architectures in [Table 1](#), used in the training. One class network with which the domain adaptation of MS-COCO was

carried out in order to have as a base a model capable of recognizing galaxies. In the case of two class network it is the final architecture of the model, it is based on transfer learning from the weights of One class network. We offer a detailed description in the next section.

3.2. Loss function

The Mask R-CNN architecture, as already explained, enables the classification, location, and segmentation implemented as single automated process. Therefore, the error function of this model includes these three tasks, as defined in Eq. (2). This loss function is applied to each sampled ROI, where \mathcal{L}_{cls} is the log loss function over two classes (Elliptical and Spiral). The \mathcal{L}_{box} term captures the error generated by the bounding box regression.

$$\mathcal{L} = \mathcal{L}_{cls} + \mathcal{L}_{box} + \mathcal{L}_{mask} \quad (2)$$

These errors expressed in the (\mathcal{L}_{cls} , \mathcal{L}_{box}) terms are the ones that evaluate the classification and location tasks. Regarding the \mathcal{L}_{mask} term, it provides a gauge of the segmentation error. Eq. (3) defines the expression to calculate this error, which is the average binary cross-entropy loss. This error is only calculated in ROIs where the class of the identified object corresponds to the ground truth class of that object.

$$\mathcal{L}_{mask} = -\frac{1}{m^2} \sum_{1 \leq i,j \leq m} [y_{ij} \log \hat{y}_{ij}^k + (1 - y_{ij}) \log (1 - \hat{y}_{ij}^k)] \quad (3)$$

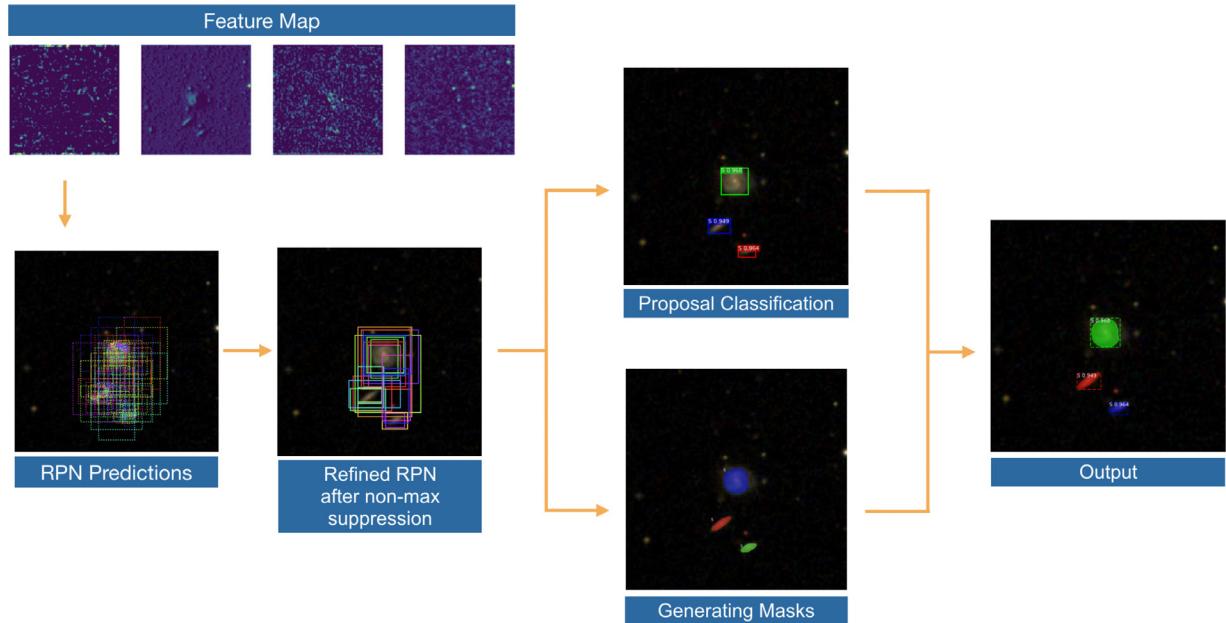


Fig. 7. The two-stage Mask R-CNN model is summarized in four steps.

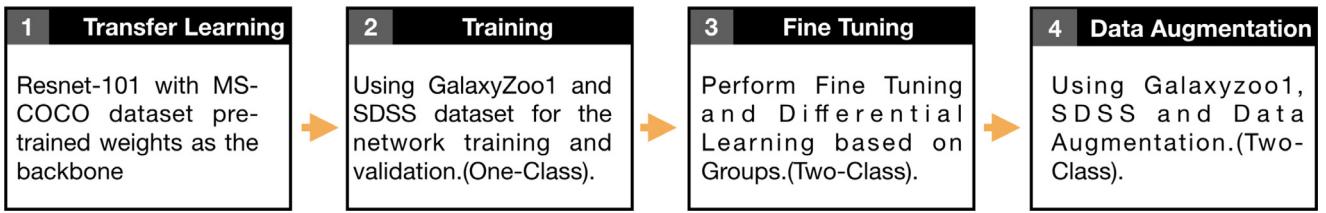


Fig. 8. Training strategy involves four steps.

4. Model training

As indicated in Section 3.1, the backbone is a key element of the Mask R-CNN architecture, because it makes possible to reuse an already highly specialized architecture in the detection of features in images. In this work, we used a Residual Deep Neural Network (Resnet) to utilize the transfer learning technique in the first stage, from MS-COCO (Lin et al., 2014), and in a second iteration, as will be seen in this chapter, from a network trained by the authors in the recognition of galaxies (not including morphology). Taking advantage of this characteristic of the model, it was applied a training strategy based on transfer Learning and Differential learning rates.

4.1. Learning strategy

Traditionally, the training of CNNs requires a dataset of thousands of images with a label that indicates their respective class; for example, in the case of galaxies, its morphological type. In the case of Mask R-CNN, this need can be mitigated by the backbone, which is part of the Mask R-CNN architecture. The backbone is a convolutional neural network responsible for generating the space of features of the model. Consequently, the training strategy is based on transfer learning and differential learning rates and is divided into four stages as shown in Fig. 8. The first stage (transfer learning) is focused on performing a domain adaptation of Resnet-10, moving from weights based on MS-COCO to those of the dataset created from Galaxy Zoo 1 (labels) and SDSS (images). The second stage is focused on executing a deep adjustment of the weights of the selected backbone, enabling this network to

Table 1
Learning configurations of the training process.

Configuration	One class network	Two class network
Number of classes(*)	2	3
Pool size	7	7
Backbone	resnet101	resnet101
Backbone strides	{8, 16, 32, 64, 128}	{4, 8, 16, 32, 64}
Activation	Relu	Relu
Detection max instances	400	400
Image resize mode	square	crop
Image shape	256 × 256	256 × 256
Learning momentum	0.9	0.9
Learning rate	1e-03	1e-03
Max gt instances	300	200
Mini mask	false	true
ROI positive ratio	0.33	0.33
RPN anchor scales	[8, 16, 32, 64, 128]	[8, 16, 32, 64, 128]
RPN nms threshold	0.7	0.8
Train ROIs per image	250	128
Steps per epoch	100	1000
Use RPN ROIs	true	true
(*) Number of classes including background		

identify and segment galaxies. As a third step, a strategy of deep adjustment of the backbone weights is maintained. In addition, a Data Augmentation pipeline was incorporated into the training process. At this stage, the morphology of the galaxies was incorporated into the parameter space that the network had to learn. Finally, a new dataset of 12,000 images and a fine-tuning learning

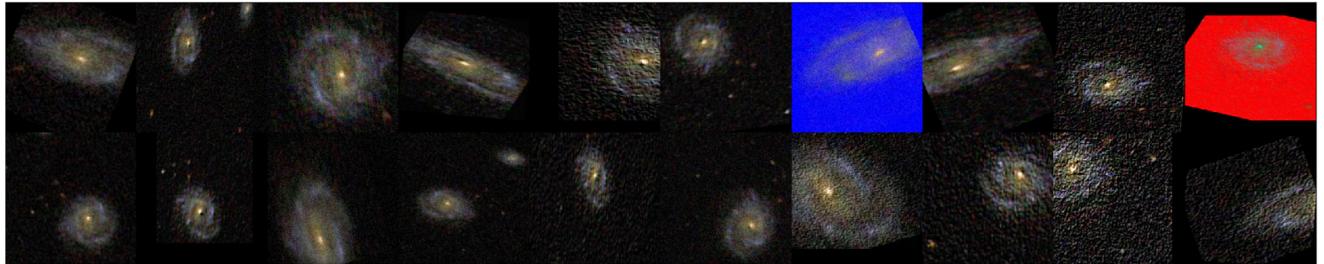


Fig. 9. Examples of different data augmentation techniques included in the imgaug library and applied to the same image of a spiral galaxy. As mentioned in Section 4.2, only techniques that are possible in an astronomical context were used for this work.

Table 2

Learning configurations of the training process.

Stage	Layer trained	Epoch	Learning rate	Train steps	Train steps
2	heads	10	1e-03	100	20
2	5+	20	1e-04	100	20
2	3+	30	1e-05	100	20
2	all	40	1e-06	100	20
3	heads	10	1e-03	1000	200
3	5+	20	1e-04	1000	200
3	3+	30	1e-05	1000	200
3	all	40	1e-06	1000	200
4	heads	15	1e-04	1000	200
4	5+	25	1e-05	1000	200
4	all	35	1e-06	1000	200
4	all	50	1e-06	1000	200

strategy based on differential learning rates were selected from the knowledge base. A detail of the steps is described below.

Stage one – A Resnet with 101 layers was used for the backbone of the network with the MS-COCO dataset pre-trained weights. This dataset was selected intending to harness the power of generalization of this model. In this case, the dataset had 12,000 randomly selected images, but it kept the representation of the balanced classes (elliptical and spiral). In the space of parameters that the network had to learn, it was not included the morphological information of the galaxies, the only indication was that it was a galaxy. The objective of this stage is to make a domain adaptation of Resnet-101 trained to identify 81 objects that are part of MS-COCO.

Stage two – An intense training strategy was put into effect to achieve a model of the recognition of galaxies between various astronomical sources. Table 2 shows the training strategy focused on using softer learning rates in the first layers and more intense rates in the intermediate layers, leaving the highest learning rates for the final layers. This strategy has been applied previously in other works, for instance (Yosinski et al., 2014). They suggest that the first layers are specialized in the detection of shapes and edges. Hence, at this level the adjustment needed is not deep, because the visible galaxies are also composed of shapes and edges. The most complex filters reside at the medium level and are specialized in the detection of parts of objects. Finally, the outermost layers of Resnet are specialized in recognizing complete objects from different shapes and sizes. This is where it is more important for the backbone weights to be adjusted so that the model can recognize galaxies, among other astronomical sources.

Stage three – This stage had a fine-tuning-based approach as a transfer learning strategy in complement to Differential Learning Rates. The learning rate was reduced by $10 \times$ per layer, as seen in Table 2, following a strategy of Differential Learning Rates based on groups. This is because at this stage the morphology is incorporated into the parameter space that the network must

learn. It is necessary to remove fully connected layers to be able to re-train the parts of the network responsible for creating the feature maps of the model.

This new network was initialized with the weights calculated in stage two. In the training, there was no adjustment to the weights of the first layers, but adjustments with a fine rate are permitted in the intermediate layers. Only in the case of the output layers the model was adjusted with greater intensity in order to achieve the morphological segmentation of galaxies according to the information incorporated into the model.

The difference with stage one is that it is not necessary to make a domain adaptation because it was already based on a model capable of recognizing galaxies.

Stage four – In stage four, the training also followed the strategy of Differential Learning Rates. In this case, there is an increase in the number of epochs with softer rates, as seen in Table 2. At this stage, a data augmentation pipeline was incorporated. This is a common technique in Deep Learning used to obtain a better generalization of the model. Basically, it consists of increasing the amount of training data using images from the same dataset. This pipeline was constrained to the physical characteristics of the problem, which will be presented in detail next.

4.2. Data augmentation

The training and validation datasets are able to satisfy the needs of the problem in terms of volumes and a balanced representation of examples for each class. As indicated, the use of transfer learning based on the backbone of the network and the training strategy also provide a base knowledge to train the model. As indicated by some works (e.g., Mikołajczyk and Grochowski, 2018), the use of Data Augmentation helps to improve the ability to generalize the model. In the Data Augmentation process, we used the imgaug (Jung et al., 2019) library with 122 possible image augmenters and augmentation techniques. These Data Augmentation techniques include Affine transformations, perspective transformations, contrast changes, Gaussian noise, dropout of regions, hue/saturation changes, cropping/padding, blurring, among others. But, as it can be seen in Fig. 9, not all augmentation techniques have scientific relevance for the characteristic of astronomical images.

Specifically, Fig. 9 makes clear that some techniques, such as contrast changes, are not applicable because they affect the chromatic characteristics of the data in an unrealistic shape in the context of astronomical data. Thus, these type of techniques do not add a value to the dataset and were not used. On the other hand, there are geometric transformations, such as Affine, which enable rotations that can be performed given the scientific nature of the problem. There are also techniques in this category, such as PiecewiseAffine, that work on the basis of fixing anchor points and then randomly moving the neighboring pixels around this point. This leads to morphological distortions of the galaxies in the images.

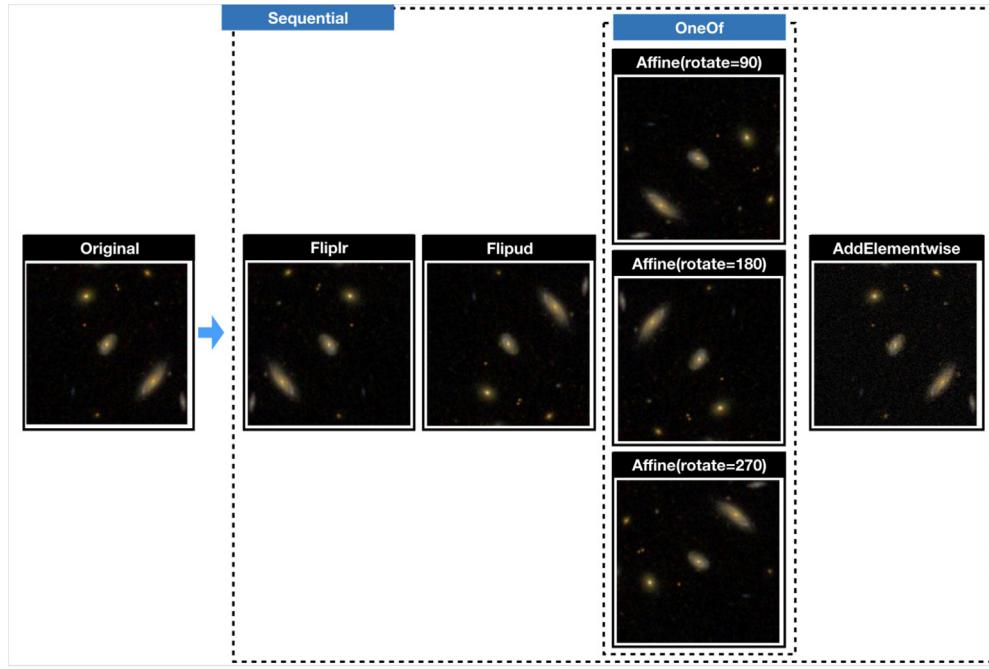


Fig. 10. Data Augmentation pipeline.

Fig. 10 displays the Data Augmentation Pipeline applied to the dataset. This pipeline consists of a sequence of four techniques: Flplr, Flipud, OneOf (rotates in different grades) and AddElementWise. The first of these techniques is Flplr. It consists of performing a horizontal mirror of the image, in this case it was applied randomly to 50% of all input images. In the case of Flipud, the probability parameterization is the same, it only differs in the rotation, which is vertical. The OneOf augmenter was incorporated to increase the geometric heterogeneity of the images. It is intended to always perform exactly one of the techniques that groups this block. The OneOf includes also three possible configurations of the rotation technique (90, 180 or 270 degrees) of the images. Finally, we have the arithmetic technique called AddElementwise, which aims to create new images by adding slightly different pixels to its neighbors. With the incorporation of this Data augmentation Pipeline, the dataset was considerably increased, allowing it to go from a mAP of 0.81 to one of mAP of 0.93.

5. Results

In this section, we present the relevant findings of this work. First, we evaluate the performance of the model using the classical machine learning metrics for classification tasks. Next, we evaluate the results obtained by Instance Segmentation applied to JPEG SDSS images. After, we test the application of the network to perform astrometric measurements on a random subset of Galaxy Zoo galaxies as an example of an application of the unique segmentation capabilities offered by Mask R-CNN.

5.1. Classification and location

The metrics used to assess the performance in the classification of a machine learning model are precision, recall, F1-score, accuracy and error rate. To calculate these metrics, it is necessary to count the results of the model on the validation sub-sample. **Fig. 11** shows confusion matrices, which is a standardized way to summarize the prediction results on classification. There are four possible outcomes, which are defined as:

Table 3
Statistics of the model.

Metric	Spiral galaxies (1)	Elliptical galaxies (2)
TP (True positive)	975	987
TN (True negative)	986	975
FP (False positive)	14	24
FN (False negative)	24	14
ACC (Accuracy)	0.98	0.98
ERR (Error rate)	0.019	0.019
F1	0.98	0.981

- TN/True Negative: Case when the entry datum is negative and the model predicts negative.
- TP/True Positive: Case when the entry datum is positive and the model predicts positive.
- FN/False Negative: Case when the entry datum is positive and the model predicts negative.
- FP/False Positive: Case when the entry datum is negative and the model predicts positive.

The classification accuracy is given by Eq. (4). Following this definition, the model achieves a value of 98%. This, in addition to the fact that the classes in the training dataset were balanced, enables us to estimate the predictive capacity of the model.

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

As shown in **Table 3**, TN reaches a total of 975 cases for spiral galaxies, and 987 cases for elliptical galaxies. This describes a layered model of correct discerning, in terms of the individual entries.

However, accuracy by itself is not sufficient to assert whether the model is a precise morphological classifier of galaxies. Precision and recall metrics are also necessary for this purpose. In global terms: Precision, as defined in Eq. (5), allows one to evaluate the quality of the model rating. The precision reaches 97% – this is the percentage of cases where the model is successful when in the correct galaxy class. Then we have the recall, which indicates what percentage of the positive class the model has

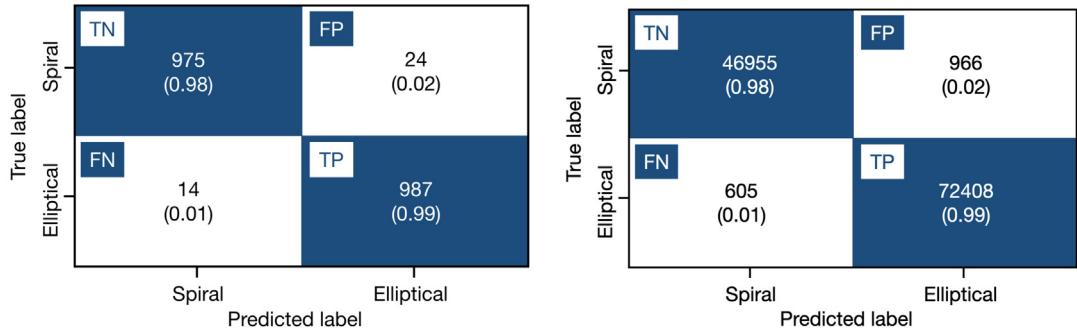


Fig. 11. Confusion matrices of the prediction results over image classification. Left: Metrics of the model for the validation subsample and the GZ1 classification labels. This sample required a minimum score of 0.8. Right: Same as Left panel, but for the remaining 120,934 galaxies that correspond to the galaxies in common between the GZ1 and GZ2 samples.

been able to identify correctly. Recall is defined in Eq. (6). Our model reaches 99%. These values were achieved by setting a classification threshold of $\text{IOU} = 0.8$. Thus, both metrics are usually in tension. Since a binary classifier seeks recall and precision, being both equally important, the metric F1 was calculated. This metric reaches a value of 98%. With these results, the classification power of our model is validated.

$$\text{precision}(p) = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

$$\text{recall}(r) = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

In the case of the evaluation of an object detection model, the element used is the Intersection over Union (IoU) metric. IoU applies the model as an evaluation metric. In addition to the classification indicators, it is necessary to verify the localization power of the model since it is difficult for the ground truth and prediction to match exactly. Having a metric that measures this relationship is important, for it is essential to set an IoU threshold to indicate whether the prediction is a true positive or a false positive. The IoU threshold is normally set at 0.5. IoU, and it does not consider the class to which the located object belongs. As the proposed model is a morphological classifier, it is necessary to incorporate this information into the evaluation. This is why the metric used is Mean Average Precision (mAP). The network achieved a value of 0.93 of mAP with all validation dataset. This result validates the power of the Architecture, based on Mask R-CNN for the automatic processes of automatic classification, location and segmentation of galaxies in a single pipeline based on deep learning.

5.2. Comparison of network and human classification (GZ1 and GZ2)

Having validated the classification capabilities of the model according to the usual metrics, next we compare the classification capabilities of our model against those of human classifiers. As we have indicated before, the motivation behind the present work is to generate a single pipeline for automatic classification and segmentation of galaxies in the context of the current volume of data generated in astronomy.

Figs. 12 and 13 summarize the accuracy of our model against the GZ1 scores for galaxies classified as Spiral or Elliptical, respectively. In particular, Fig. 12 compares the accuracy values of the model against the GZ1 scores for Spirals, with the requirement that the GZ2 class also is S. For the cases where the model agrees with the GZ classifications (shown in blue color), we divided the sample between ($\text{GZ1}_{\text{score}} > 0.8$) – which usually is referred as the *clean* sample in GZ1 (upper left panel) – and $\text{GZ1}_{\text{score}} > 0.8$ (lower left panel). The model returns a high accuracy for the clean sample, with median values over 97% and a low 1- σ dispersion.

The network performs well even for the second sample with lower GZ1 scores. In both cases, it is clear how accuracy increases monotonically as the GZ1 score increases. We interpret this as an indication of good agreement between the metric given by network accuracy and the scores of human classifiers. On the contrary, the upper right panel displays the values where the model is in disagreement with the GZ classification. However, it is important to note that there are two relevant differences compared to the previous panels. First, the median value of the accuracy, for GZ scores within the clean sample, is much lower than the accuracy for galaxies in agreement (upper panels). Second, the 1- σ values are also much broadly distributed around the mean, and lower. After a visual inspection of many misclassified galaxies, we conclude that the model actually fails in most of the cases; nevertheless, there are cases where the model is correct and GZ is not – one case is shown in the lower right panel.

Fig. 13 presents results in a similar fashion as Fig. 12, but for galaxies classified as Elliptical in both GZ1 and GZ2. In this case, we again see the correspondence and monotonical increase of accuracy with GZ1 score. However, the median value of accuracy is lower, for any GZ1 score, compared to the S class in Fig. 12. Furthermore, the dispersion is larger. In the case of misclassified galaxies (upper right panel) and after visual inspection of many galaxies, we find that the network fails in most cases; however, the fraction of cases where the network is right is larger compared to the equivalent case shown in Fig. 12. An example of this is shown in the images within the lower right panel. Anyway, the accuracy reported by the network for misclassified galaxies is still lower than the accuracy reported for the correctly classified ones (upper left panel). This behavior was also observed in the previous figure.

Some additional insight is provided by the distributions shown in the histograms of both figures. The distributions of accuracy values for objects classified correctly are skewed and narrowly peaked at values larger than 0.90. On the contrary, the accuracy distribution for misclassified galaxies resembles a uniform distribution, approximately between 0.6 and 0.95, with no peak. For both morphological classes, the misclassified galaxies are small compared to the ones classified correctly, as it was shown in Fig. 11. In general, we find that misclassified galaxies had a bright star near or over the galaxy.

Taking into consideration the performance of the model, shown above, we consider that the network provides clean samples for accuracy values over 0.9 for both classes, which would add around 48% more galaxies to the clean sample of the GZ1.

5.3. Instance segmentation

To assess the performance of an instance segmentation model, it is required to evaluate the mask. Fig. 14 displays the powerful

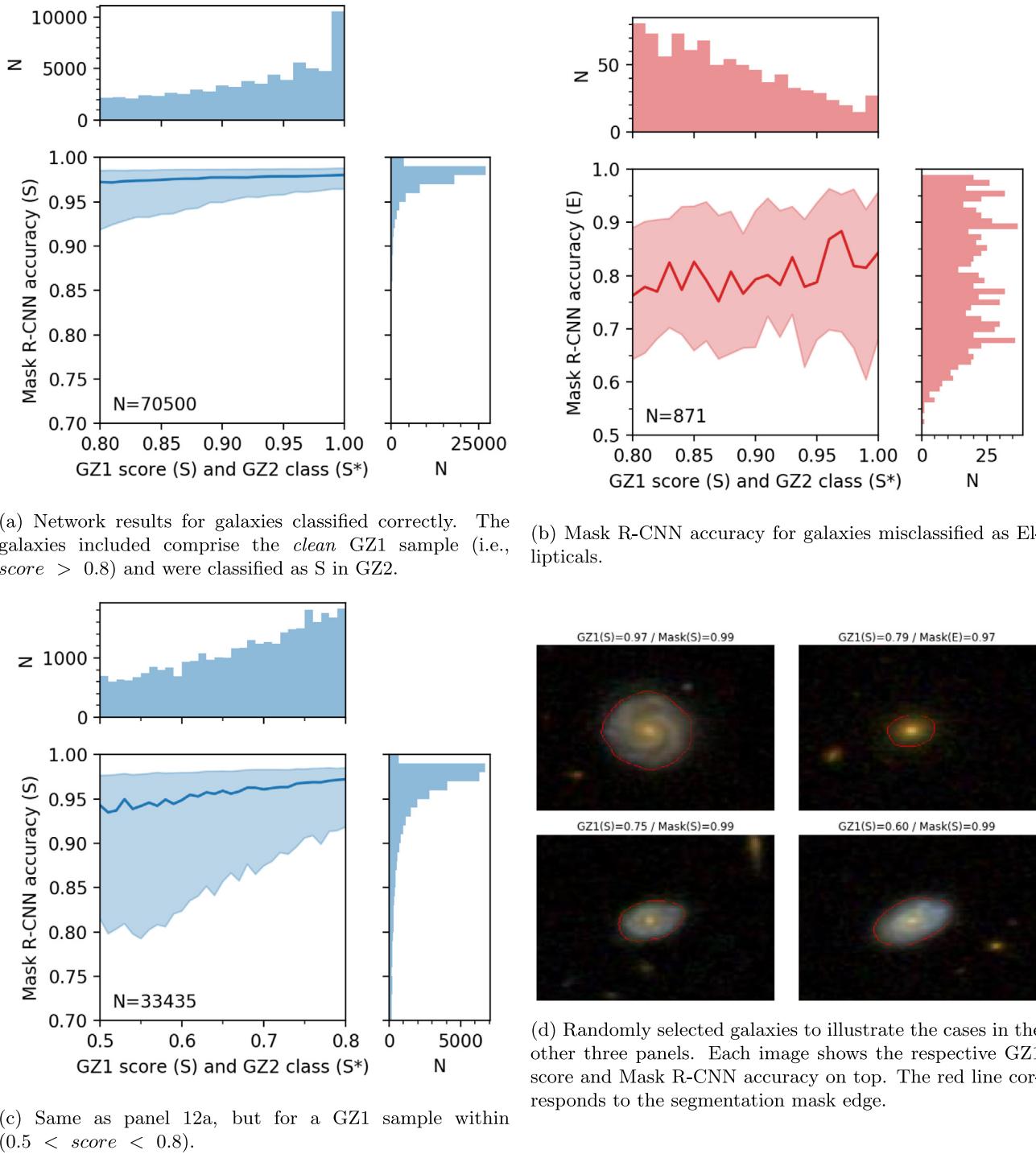
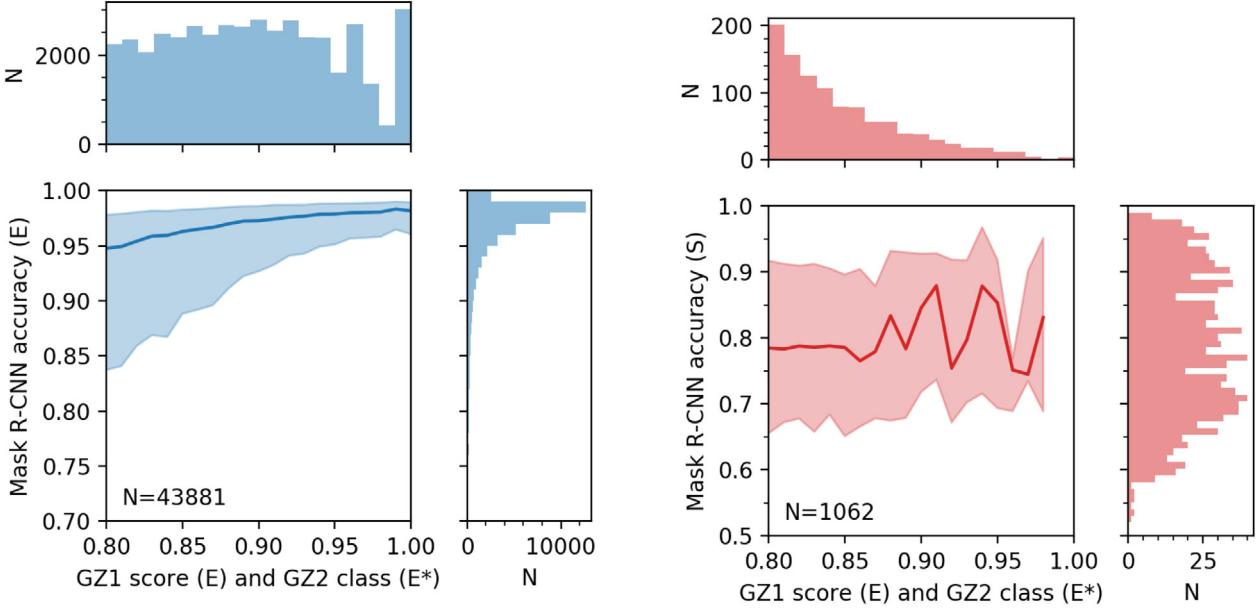


Fig. 12. Mask R-CNN accuracy performance compared to galaxies classified as Spiral in both GZ1 and GZ2. In each plot, the thick line represents the median Mask R-CNN score for 1% bins in the GZ1 score for the S class. The shaded region displays the $1 - \sigma$ percentiles. The histograms show the distribution of the sample included within each plot, whose total counts are given by the value of N . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

results in the ability to automatically segment the shape of the galaxies. The green contour corresponds to the mask entered as input to the model, whereas the red contour corresponds to the prediction made by the network. In addition, accuracy and IoU are displayed. The latter corresponds to how much the predicted boundary overlaps with the real object boundary. The left panel of Fig. 14 shows how the proposed mask, created using the imantics and Scikit-Image, includes an unresolved source that is not part of the spiral galaxy analyzed. On the other hand, using the automatic

pipeline proposed, the classification accuracy achieved is 97%. The segmentation generated by the model is superior, it fits almost exactly to the contour and shape of the galaxy, achieving 97% in the object localization indicator. The right panel of Fig. 14 shows a better case of the quality of the segmentation. The masking process through Scikit-Image generated three isolated masks, but they are considered part of the same galaxy, but the model correctly segments only the analyzed elliptical galaxy.



(a) Network results for galaxies classified correctly. The galaxies included comprise the *clean* GZ1 sample (i.e., $score > 0.8$) and were classified as E in GZ2.

(b) Mask R-CNN accuracy for galaxies misclassified as Spirals.

(c) Same as panel 13a, but for a GZ1 sample within $0.5 < score < 0.8$.

(d) Randomly selected galaxies to illustrate the cases in the other three panels. Each image shows the respective GZ1 score and Mask R-CNN accuracy on top. The red line corresponds to the segmentation mask edge

Fig. 13. Mask R-CNN accuracy performance compared to galaxies classified as Elliptical in GZ1 and GZ2. In each plot, the thick line represents the median Mask R-CNN score for a 1% bin in the GZ1 score for E class. The shaded region displays the 1- σ percentiles. The histograms show the distribution of the sample included within each plot, whose total counts are given by the value of N . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

As we mentioned in Section 2, no color correction nor level adjustments were applied to the images of the dataset. This is a relevant factor because it impacts in the ability to generalize the model. Fig. 18 shows the response of the network to an example image with marginal background calibration (first panel). Furthermore, we test the performance of the model to increasing chroma noise drawn from a Gaussian distribution with scales {12.75, 25.5, 51.0, 76.5}, for panels two to four}. Even under

a Gaussian noise with scale 76.5, the model is able to locate, segment and classify the object as a spiral galaxy with an accuracy of 0.92.

Additionally, we validate the ability to generalize the model in terms of not being limited to the size of the dataset images. We downloaded a 1000×1000 JPEG image from the SDSS SkyServer centered at $(\alpha, \delta) = (309.42896 \text{ deg}, -1.1696352 \text{ deg})$ with a pixel scale of 0.360 [arcsec/pixel]. The results obtained were 82

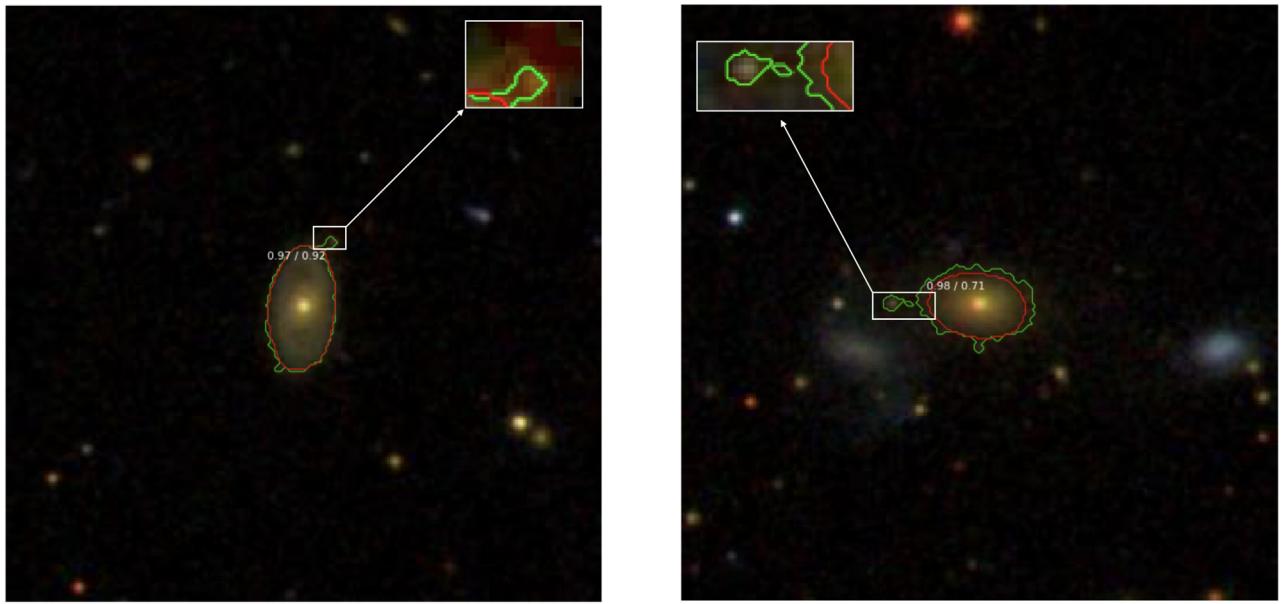


Fig. 14. Image Segmentation. Both panels show Ground Truth (GT) and detections. The green contour corresponds to the GT, whereas the red contour corresponds to the output made by the network. Numbers in white font display Accuracy/IoU, respectively, measured by the model. Left panel: Example for a S galaxy. Right panel: Example of an E galaxy. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

galaxies detected, segmented and classified, as shown in Fig. 15. It is noteworthy to indicate that the model provides the location of the galaxies found and that this information can be incorporated into a catalog.

5.4. Determination of galaxy centroids: A mask R-CNN application test case

There is no doubt that the information provided by a mask, generated at the pixel level, opens several applications in astronomy. Here we test the calculation of centroids as a direct application of the mask over JPEG data. Even though most applications would require to work with data in the FITS format – which goes beyond the scope of this contribution, we shall see that results in this test case are satisfactory even using JPEG images.

We picked 3000 galaxies randomly and evaluated them through the network. As part of the output, we obtained a mask (a 256×256 boolean array) for each image in the dataset. Next, we applied the mask over the RGB image of the galaxy, leaving only the pixels from the object. It would be possible to measure any linear combination of the three channels. However, we only used the G channel in the JPEG, which corresponds to the r band data. This band is used by the SDSS to define the apertures for deriving astrometric information in their catalogs. We measured galaxy centroids through the `CENTER_OF_MASS` function of the Scipy's `NDIMAGE` library. The results are summarized in Table 4, whereas Fig. 16 shows the respective distributions along the X and Y axis of the images. It is expected that the centroids would fall precisely at the center of the images $(x, y) = (127.5, 127.5)$ pixels because we use the center of the galaxies, as reported by the SDSS, to download our images. We find median positions at $(x, y) = (127.3, 128.5)$ with MAD 0.27 and 0.34 pixels along the X and Y axis, respectively. We find a good agreement for this simple method that is sensitive to bright sources, showing that these masks derived from lightweight JPEG images may prove useful for astronomical purposes. For instance, the difference between the median of the centroids and the image center correspond to a difference of 0.08 and 0.39 arcseconds, well below the typical seeing disk at Apache Point Observatory. Finally, the method is

Table 4

Statistics of Centroids for $N = 3000$ galaxies. The numbers presented are measured after a 5-sigma cut. All values are reported in pixel units.

	Mean	Median	Standard deviation	Median absolute deviation
X_{centroid}	126.93	127.3	1.31	0.27
Y_{centroid}	128.7	128.5	0.92	0.34

also consistent. The Median Absolute Deviation, in both image axes, is around 0.11 arcsec.

5.5. Automatic classification, location and segmentation pipeline

One of the premises of the proposal was the implementation of a machine learning pipeline for morphological classification of galaxies. This, under the context that classifying whether or not a galaxy has a certain morphology does not address the entire complexity of the problem. Furthermore, it is the location of this galaxy within a flat as the new scientific mega-projects such as the LSST, E-ELT, are captured.

Additionally, another problem is the location of the galaxy, in highly complex scenarios such as the 3.2 gigapixels of the LSST or other mega scientific projects such as the E-ELT. The location, as indicated, can be addressed with computer vision techniques, such as object detection and localization. The proposed model includes the generation of a bounding box around the galaxies, automatically locating these on the plane. With the results achieved in accuracy and other metrics presented in Table 3, the state of the art is achieved in terms of classification and object detection applied to the morphological classification of galaxies.

To achieve an automatic pipeline, another task must be included: Segmentation. The proposal includes automatic segmentation of galaxies tailored to their individual morphology. The result is visually presented as a mask, and the model returns this segmentation as a set of coordinates that can be automatically incorporated into a catalog. These results position the proposal as the first to achieve these three tasks (classification, location and segmentation). In Fig. 17 it can be appreciated that the model is an automatic pipeline of classification and morphological segmentation of galaxies. The first image corresponds to a galaxy

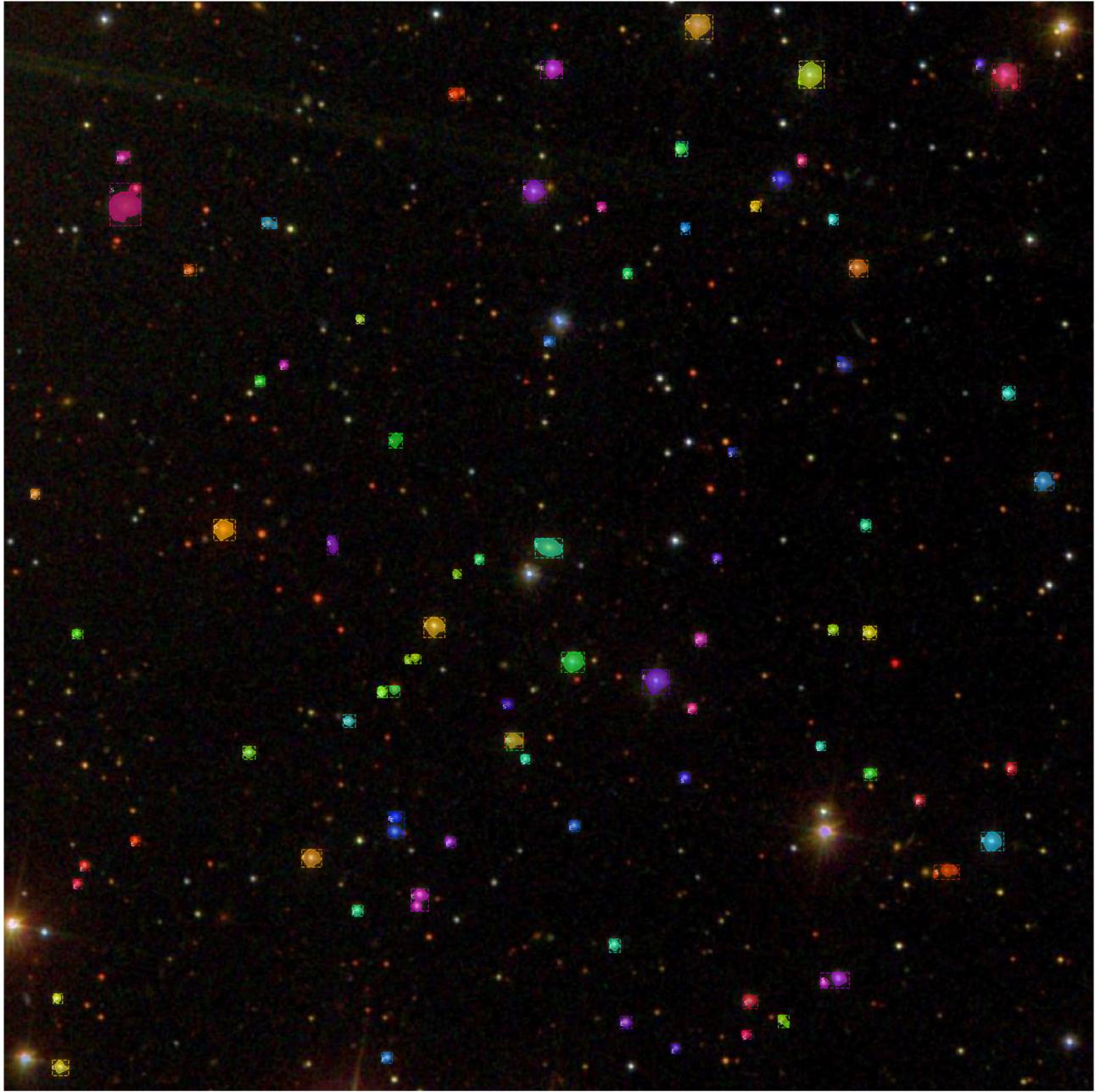


Fig. 15. Result of the application of the model to a query to the Skyserver. No color correction was applied. The result was 82 galaxies, correctly morphologically segmented.

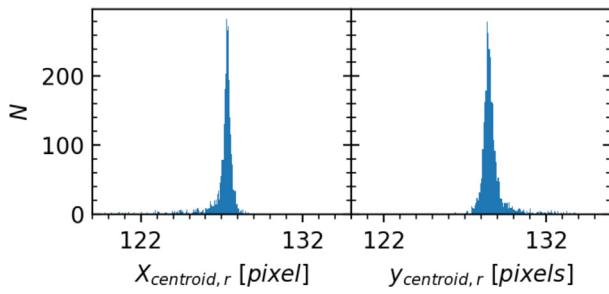


Fig. 16. Centroids derived for a random sample of 3000 galaxies. The left and right panels show the distribution for the X and Y centroids, respectively.

to be identified. The second shows the input of the model for training (tag and mask). The third corresponds to the feature map generated by the backbone (Resnet) of the network. The third corresponds to the result of classification and the location of the model. Finally, the model incorporates the segmentation of the galaxy.

The global model reaches the accuracy of 93% of mAP. It is clear that there are many iterations to increase this percentage, especially when dealing with problems such as separation of galaxies with overlapping, galaxies of a very large size, among other problems.

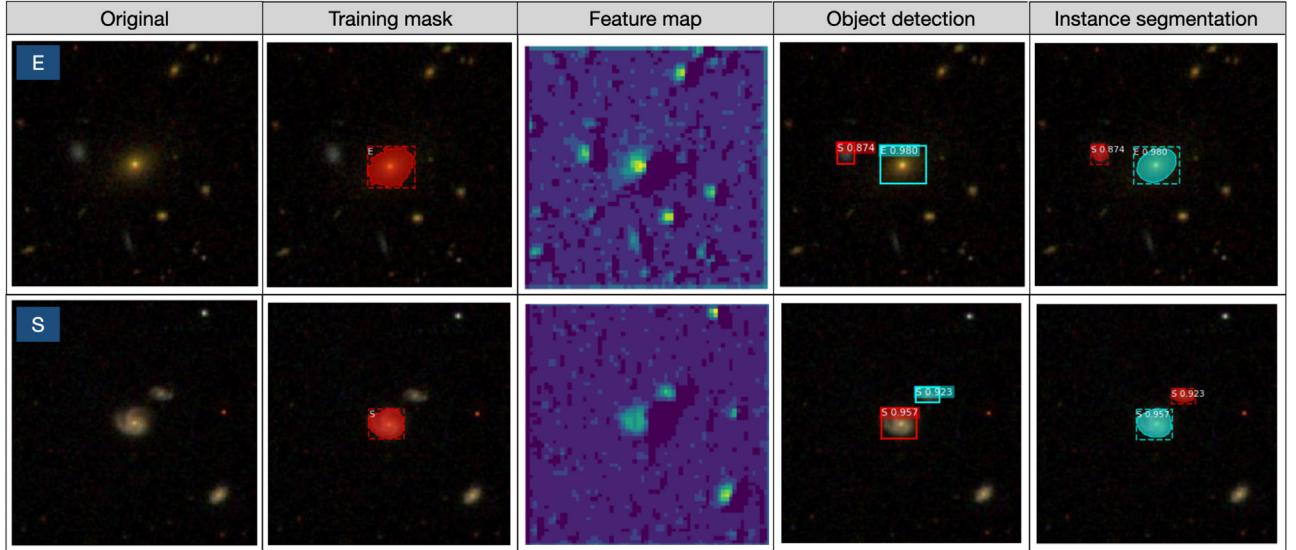


Fig. 17. Automatic classification, location and segmentation pipeline applied to elliptical-type and spiral-type of galaxies.

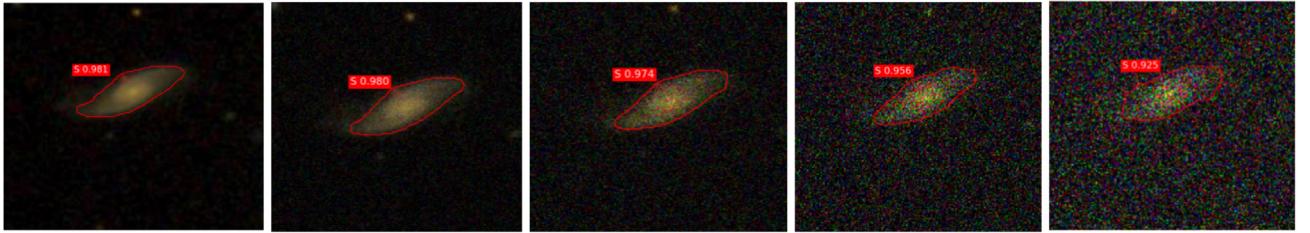


Fig. 18. Response of the Mask R-CNN to chroma noise in the background. The leftmost panel shows an image included in the validation dataset. The other three panels correspond to the same image after adding chroma Gaussian noise with scales {12.75, 25.5, 51.0 and 76.5}, respectively, from left to right.

6. Contribution and future work

We have validated the premise of the implementation of a single pipeline for detection, segmentation and morphological classification of galaxies. The pipeline delivers results for analysis in the form of images with segmented sources, masks and tabular data. These data can be joined to existing catalogs. The pipeline application on the entire Galaxy Zoo 2 catalog (243,480 SDSS images) was performed in approximately 2.5 h. This process was run on an Amazon Web Services (AWS) p2.8XLARGE instance. This instance has 8 NVIDIA K80 GPUs and 488 GB of RAM.

Semantic object segmentation and morphological classification are common tasks in astronomy. Typically, object segmentation is achieved with tools such like SExtractor. However, morphological classification requires the implementation of a different suit of tools, which also depend on the approach (using descriptive quantities as proxies of morphology or CNNs over images). In this work, we have shown that both stages can be performed by this model through an automatic process using an architecture based on the state of the art of instance segmentation, associated with a training strategy based on transfer learning and differential learning rates.

6.1. Future work

Considering the data scale of the upcoming surveys, such as the LSST, it becomes necessary to advance both in accuracy and inference speed. In the case of accuracy, effort should be made in passing from e-channel RGB images to an input based on the multi-band photometric observations currently available. However, this increases the computational complexity due to the

increase in dimensionality, so additional work should be made to advance inference speed. In this sense, the development of new techniques on dimensionality reduction while preserving the multidimensional structures and relationships present in these observations is required.

A second line of research is the implementation of the new architectures for working in FITS format. The first major contribution in this scope was published recently by [Hausen and Robertson \(2020\)](#), who use the U-Net architecture. Raw data from a telescope in FITS format comes as 16 bits integer data, allowing for 2^{16} or 65536 distinct values per image, which of course corresponds to a single photometric filter. After, these data are reduced as floating point data. On the other hand, JPEG images comprise three channels of 2^8 bits each, so there is a loss of information by quantizing values when converting FITS to JPEG (roughly speaking, similar to creating a histogram of 256 bins), although the dimensionality of the data is increased from a single photometric band to 3 bands. As a result, the advantage of working with multidimensional FITS data is the astrophysical information contained the combined bandpasses or filter along the spectrum, which considers the fact that most current surveys have at least 5 filters (SDSS), 6 (LSST) or including 12 bands (S-PLUS [Mendes de Oliveira et al., 2019](#)). The large ecosystem of architectures have not been tested yet under the proposed constraints.

Another relevant factor to consider are transfer learning strategies. As detailed in Section 4 is a key tool for deep learning model training. In models that have the same input dimensionality (JPG) this is a problem that has widely resulted. The challenge is how to perform this task by increasing the dimensionality of the input. There is work ([Dodge and Karam, 2016](#)) that validates as a good

practice training models with low resolution images (JPG), it helps the model's ability to classify when it later receives high resolution images. This same approach guides the completion of future work. In a scenario where a transfer learning is carried out from the current model to a larger model. This last model is the one that would receive as input the astronomical data in its original format.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank the two anonymous referees, for their helpful comments that contributed to improve this manuscript.

HFA acknowledges the financial support of UTFSM, Chile, through the Incentive Program for Scientific Initiation (IPSI).

MJA acknowledges the financial support of DIDULS/ULS, Chile, through the project PI192135.

References

- Abdulla, W., 2017. Mask r-cnn For Object Detection and Instance Segmentation on Keras and Tensorflow. GitHub Repository.
- Akeret, J., Chang, C., Lucchi, A., Refregier, A., 2017. Radio frequency interference mitigation using deep convolutional neural networks. *Astronom. Comput.* 18, 35–39.
- Anon, 2020a. Rectlabel An image annotation tool to label images for bounding box object detection and segmentation. <https://rectlabel.com/>. (Accessed 24 January 2020).
- Anon, 2020b. Coco-ui Front end user interfaces that are used to annotate COCO dataset. <https://github.com/tylin/coco-ui>. (Accessed 24 January 2020).
- Anon, 2020c. Imantics image semantics. <https://github.com/jsbroks/imantics>. (Accessed 24 January 2020).
- Anon, 2020d. Labelbox annotate data to build and ship artificial intelligence applications. <https://labelbox.com/>. (Accessed 24 January 2020).
- Banerji, M., Lahav, O., Lintott, C.J., Abdalla, F.B., Schawinski, K., Bamford, S.P., Andreescu, D., Murray, P., Raddick, M.J., Slosar, A., et al., 2010. Galaxy Zoo: reproducing galaxy morphologies via machine learning. *Mon. Not. R. Astron. Soc.* 406 (1), 342–353.
- Barchi, P.H., de Carvalho, R.R., Rosa, R.R., Sautter, R.A., Soares-Santos, M., Marques, B.A.D., Clua, E., Gonçalves, T.S., de Sá-Freitas, C., Moura, T.C., 2020. Machine and deep learning applied to galaxy morphology - A comparative study. *Astronom. Comput.* 30, 100334. doi:[10.1016/j.ascom.2019.100334](https://doi.org/10.1016/j.ascom.2019.100334).
- Baron, D., 2019. Machine learning in astronomy: a practical overview. arXiv preprint [arXiv:1904.07248](https://arxiv.org/abs/1904.07248).
- Bertin, E., Arnouts, S., 1996. SExtractor: Software for source extraction. *Astron. Astrophys. Suppl. Ser.* 117 (2), 393–404.
- Boucaud, A., Huertas-Company, M., Heneka, C., Ishida, E.E.O., Sedaghat, N., de Souza, R.S., Moews, B., Dole, H., Castellano, M., Merlin, E., Roscani, V., Tramacere, A., Killedar, M., Trindade, A.M.M., 2019. Photometry of high-redshift blended galaxies using deep learning. *Mon. Not. R. Astron. Soc.* doi:[10.1093/mnras/stz3056](https://doi.org/10.1093/mnras/stz3056), URL <https://doi.org/10.1093/mnras/stz3056>. stz3056.
- Brown, A.G., Vallenari, A., Prusti, T., De Bruijne, J., Mignard, F., Drimmel, R., Babusiaux, C., Bailer-Jones, C., Bastian, U., Biermann, M., et al., 2016. Gaia Data Release 1-Summary of the astrometric, photometric, and survey properties. *Astron. Astrophys.* 595, A2.
- Conselice, C.J., 2003. The relationship between stellar light distributions of galaxies and their formation histories. *Astrophys. J. Suppl. Ser.* 147 (1), 1.
- Dieleman, S., Willett, K.W., Dambre, J., 2015. Rotation-invariant convolutional neural networks for galaxy morphology prediction. *Mon. Not. R. Astron. Soc.* 450 (2), 1441–1459.
- Dodge, S., Karam, L., 2016. Understanding how image quality affects deep neural networks. In: 2016 Eighth International Conference on Quality of Multimedia Experience. QoMEX, IEEE, pp. 1–6.
- Dutta, A., Zisserman, A., 2019. The via annotation software for images, audio and video. In: Proceedings of the 27th ACM International Conference on Multimedia. In: MM '19, ACM, New York, NY, USA, ISBN: 978-1-4503-6889-6, doi:[10.1145/3343031.3350535](https://doi.org/10.1145/3343031.3350535), <https://doi.org/10.1145/3343031.3350535>.
- Gardner, J.P., Mather, J.C., Clampin, M., Doyon, R., Greenhouse, M.A., Hammel, H.B., Hutchings, J.B., Jakobsen, P., Lilly, S.J., Long, K.S., et al., 2006. The James Webb space telescope. *Space Sci. Rev.* 123 (4), 485–606.
- Girshick, R., 2015. Fast r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448.
- Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587.
- González, R.E., Muñoz, R.P., Hernández, C.A., 2018. Galaxy detection and identification using deep learning and data augmentation. *Astronom. Comput.* 25, 103–109. doi:[10.1016/j.ascom.2018.09.004](https://doi.org/10.1016/j.ascom.2018.09.004).
- Hausen, R., Robertson, B.E., 2020. Morpheus: A deep learning framework for the pixel-level analysis of astronomical image data. *Astrophys. J. Suppl. Ser.* 248 (1), 20. doi:[10.3847/1538-4365/ab8868](https://doi.org/10.3847/1538-4365/ab8868).
- He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.
- Hubble, E., 1926. No. 324. extra-galactic nebulae.. In: Contributions from the Mount Wilson Observatory, Vol. 324. Carnegie Institution of Washington, pp. 1–49.
- Ivezic, Z., Axelrod, T., Brandt, W., Burke, D., Claver, C., Connolly, A., Cook, K., Gee, P., Gilmore, D., Jacoby, S., et al., 2011. Large Synoptic Survey Telescope: From Science Drivers to Reference Design. Technical Report, SLAC National Accelerator Lab., Menlo Park, CA (United States).
- Jung, A.B., Wada, K., Crall, J., Tanaka, S., Graving, J., Yadav, S., Banerjee, J., Vecsei, G., Kraft, A., Borovec, J., Valentini, C., Zhydenko, S., Pfeiffer, K., Cook, B., Fernández, I., Chi-Hung, W., Ayala-Acevedo, A., Meudec, R., Laporte, M., et al., 2019. Imgaug. <https://github.com/aleju/imgaug>. Online; (Accessed 25 September 2019).
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft coco: Common objects in context. In: European Conference on Computer Vision. Springer, pp. 740–755.
- Lintott, C., Schawinski, K., Bamford, S., Slosar, A., Land, K., Thomas, D., Edmondson, E., Masters, K., Nichol, R.C., Raddick, M.J., et al., 2011. Galaxy Zoo 1: data release of morphological classifications for nearly 900 000 galaxies. *Mon. Not. R. Astron. Soc.* 410 (1), 166–178.
- Lintott, C.J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Radick, M.J., Nichol, R.C., Szalay, A., Andreescu, D., et al., 2008. Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan digital sky survey. *Mon. Not. R. Astron. Soc.* 389 (3), 1179–1189.
- Lupton, R., Blanton, M.R., Fekete, G., Hogg, D.W., O'Mullane, W., Szalay, A., Wherry, N., 2004. Preparing red-green-blue images from CCD data. *Publ. Astron. Soc. Pac.* 116 (816), 133.
- Mendes de Oliveira, C., Ribeiro, T., Schoenell, W., Kanaan, A., Overzier, R.A., Molino, A., Sampedro, L., Coelho, P., Barbosa, C.E., Cortesi, A., Costa-Duarte, M.V., Herpich, F.R., Hernand ez-Jimenez, J.A., Placco, V.M., Xavier, H.S., Abramo, L.R., Saito, R.K., Chies-Santos, A.L., Ederoclite, A., Lopes de Oliveira, R., Gonçalves, D.R., Akras, S., Almeida, L.A., Almeida-Fernandes, F., Beers, T.C., Bonatto, C., Bonoli, S., Cypriano, E.S., Vinicius-Lima, E., de Souza, R.S., Fabiano de Souza, G., Ferrari, F., Gonçalves, T.S., Gonzalez, A.H., Gutiérrez-Soto, L.A., Hartmann, E.A., Jaffe, Y., Kerber, L.O., Lima-Dias, C., Lopes, P.A.A., Menendez-Delmestre, K., Nakazono, L.M.I., Novais, P.M., Ortega-Minakata, R.A., Pereira, E.S., Perrottoni, H.D., Queiroz, C., Reis, R.R.R., Santos, W.A., Santos-Silva, T., Santucci, R.M., Barbosa, C.L., Siffert, B.B., Sodré, L., Torres-Flores, S., Westera, P., Whitten, D.D., Alcaniz, J.S., Alonso-García, J., Alencar, S., Alvarez-Cand al, A., Amram, P., Azanha, L., Barbá, R.H., Bernardinelli, P.H., Borges Fernandes, M., Branco, V., Brito-Silva, D., Buzzo, M.L., Caffer, J., Campillay, A., Cano, Z., Carvano, J.M., Castejon, M., Cid Fernandes, R., Dantas, M.L.L., Daflon, S., Damke, G., de la Reza, R., de Melo de Azevedo, L.J., De Paula, D.F., Diem, K.G., Donnerstein, R., Dors, O.L., Dupke, R., Eikenberry, S., Escudero, C.G., Faifer, F.R., Farias, H., Fernandes, B., Fernandes, C., Fontes, S., Galarza, A., Hirata, N.S.T., Katena, L., Gregorio-Hetem, J., Hernández-Fernández, J.D., Izzo, L., Jaque Arancibia, M., Jatenco-Pereira, V., Jiménez-Teja, Y., Kann, D.A., Krabbe, A.C., Labayru, C., Lazzaro, D., Lima Neto, G.B., Lopes, A.a.R., Magalhães, R., Makler, M., de Menezes, R., Miralda-Escudé, J., Monteiro-Oliveira, R., Montero-Dorta, A.D., Muñoz-Elgueta, N., Nemmen, R.S., Nilo Castellón, J.L., Oliveira, A.S., Ortiz, D., Pattro, E., Pereira, C.B., Quint, B., Riguccini, L., Rocha Pinto, H.J., Rodrigues, I., Roig, F., Rossi, S., Saha, K., Santos, R., Schnorr Müller, A., Sesto, L.A., Silva, R., Smith Castelli, A.V., Teixeira, R., Telles, E., Thom de Souza, R.C., Thöne, C., Trevisan, M., de Ugarte Postigo, A., Urrutia-Viscarra, F., Veiga, C.H., Vika, M., Vitorelli, A.Z., Werle, A., Werner, S.V., Zaritsky, D., 2019. The Southern Photometric Local Universe Survey (s-PLUS): improved SEDs, morphologies, and redshifts with 12 optical filters. *Mon. Not. R. Astron. Soc.* 489 (1), 241–267. doi:[10.1093/mnras/stz1985](https://doi.org/10.1093/mnras/stz1985).

- Mikołajczyk, A., Grochowski, M., 2018. Data augmentation for improving deep learning in image classification problem. In: 2018 International Interdisciplinary PhD Workshop (IIPhDW), pp. 117–122. <http://dx.doi.org/10.1109/IIPHDW.2018.8388338>.
- Nieto-Santisteban, M.A., Szalay, A.S., Gray, J., 2004. Imgcutout, an engine of instantaneous astronomical discovery. In: Ochsenbein, F., Allen, M.G., Egret, D. (Eds.), *Astronomical Data Analysis Software and Systems (ADASS) XIII*. In: *Astronomical Society of the Pacific Conference Series*, vol. 314, p. 666.
- Odewahn, S.C., Stockwell, E., Pennington, R., Humphreys, R.M., Zumach, W., 1992. Automated star/galaxy discrimination with neural networks. In: *Digitalized Optical Sky Surveys*. Springer, pp. 215–224.
- Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788.
- Redmon, J., Farhadi, A., 2017. YOLO9000: better, faster, stronger. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7263–7271.
- Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In: *Advances in Neural Information Processing Systems*. pp. 91–99.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 234–241.
- Sandler, D., Barrett, T., Palmer, D., Fugate, R., Wild, W., 1991. Use of a neural network to control an adaptive optics system for an astronomical telescope. *Nature* 351 (6324), 300–302.
- Spergel, D., Gehrels, N., Baltay, C., Bennett, D., Breckinridge, J., Donahue, M., Dressler, A., Gaudi, B., Greene, T., Guyon, O., et al., 2015. Wide-field infrared survey telescope-astrophysics focused telescope assets WFIRST-AFTA 2015 report. arXiv preprint [arXiv:1503.03757](https://arxiv.org/abs/1503.03757).
- Szalay, A.S., Gray, J., Thakar, A.R., Kunszt, P.Z., Malik, T., Raddick, J., Stoughton, C., vandenBerg, J., 2002. The sdss skyserver: public access to the Sloan digital sky server data. In: Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data. In: *SIGMOD '02*, Association for Computing Machinery, New York, NY, USA, ISBN: 1581134975, pp. 570–581. doi:10.1145/564691.564758, <https://doi.org/10.1145/564691.564758>.
- Torralba, A., Russell, B.C., Yuen, J., 2010. Labelme: Online image annotation and applications. *Proc. IEEE* 98 (8), 1467–1484.
- Van der Walt, S., Schönberger, J., Nunez-Iglesias, J., Boulogne, F., Warner, J., Yager, N., Gouillart, E., Yu, T., 2014. Scikit-image: image processing in Python. *PeerJ* 2, e453.
- Willett, K.W., Lintott, C.J., Bamford, S.P., Masters, K.L., Simmons, B.D., Casteele, K.R., Edmondson, E.M., Fortson, L.F., Kaviraj, S., Keel, W.C., et al., 2013. Galaxy Zoo 2: detailed morphological classifications for 304 122 galaxies from the Sloan Digital Sky Survey. *Mon. Not. R. Astron. Soc.* 435 (4), 2835–2860.
- York, D.G., Adelman, J., Anderson, Jr., J.E., Anderson, S.F., Annis, J., Bahcall, N.A., Bakken, J.A., Barkhouser, R., Bastian, S., Berman, E., Boroski, W.N., Bracker, S., Briegel, C., Briggs, J.W., Brinkmann, J., Brunner, R., Burles, S., Carey, L., Carr, M.A., Castander, F.J., Chen, B., Colestock, P.L., Connolly, A.J., Crocker, J.H., Csabai, I., Czarapata, P.C., Davis, J.E., Doi, M., Dombeck, T., Eisenstein, D., Ellman, N., Elms, B.R., Evans, M.L., Fan, X., Federwitz, G.R., Fischelli, L., Friedman, S., Frieman, J.A., Fukugita, M., Gillespie, B., Gunn, J.E., Gurbani, V.K., de Haas, E., Haldeman, M., Harris, F.H., Hayes, J., Heckman, T.M., Hennessy, G.S., Hindsley, R.B., Holm, S., Holmgren, D.J., Huang, C.-h., Hull, C., Husby, D., Ichikawa, S.-I., Ichikawa, T., Ivezić, v., Kent, S., Kim, R.S.J., Kinney, E., Klaene, M., Kleinman, A.N., Kleinman, S., Knapp, G.R., Korienek, J., Kron, R.G., Kunzst, P.Z., Lamb, D.Q., Lee, B., Leger, R.F., Limmongkol, S., Lindenmeyer, C., Long, D.C., Loomis, C., Loveday, J., Lucinio, R., Lupton, R.H., MacKinnon, B., Mannery, E.J., Mantsch, P.M., Margon, B., McGehee, P., McKay, T.A., Meiksin, A., Merelli, A., Monet, D.G., Munn, J.A., Narayanan, V.K., Nash, T., Neilsen, E., Neswold, R., Newberg, H.J., Nichol, R.C., Nicinski, T., Nonino, M., Okada, N., Okamura, S., Ostriker, J.P., Owen, R., Pauls, A.G., Peoples, J., Peterson, R.L., Petrick, D., Pier, J.R., Pope, A., Pordes, R., Prosaio, A., Rechenmacher, R., Quinn, T.R., Richards, C.T., Richmond, M.W., Rivetta, C.H., Rockosi, C.M., Ruthmansdorfer, K., Sandford, D., Schlegel, D.J., Schneider, D.P., Sekiguchi, M., Sergey, G., Shimasaku, K., Siegmund, W.A., Smee, S., Smith, J.A., Snedden, S., Stone, R., Stoughton, C., Strauss, M.A., Stubbs, C., SubbaRao, M., Szalay, A.S., Szapudi, I., Szokoly, G.P., Thakar, A.R., Tremonti, C., Tucker, D.L., Uomoto, A., Vanden Berk, D., Vogeley, M.S., Waddell, P., Wang, S.-i., Watanabe, M., Weinberg, D.H., Yanny, B., Yasuda, N., SDSS Collaboration, 2000. The Sloan digital sky survey: Technical summary. *Astronom. J.* 120 (3), 1579–1587. doi:10.1086/301513.
- Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks? In: *Advances in Neural Information Processing Systems*. pp. 3320–3328.