Separating Overlapping galaxies with Mask R-CNN
CS230: Final Project Report

Sowmya Kamath
Department of Physics
Stanford University
sowmyak@stanford.edu

Abstract

This research note summarizes the results of an implementation of Mask R-CNN to perform object detection and segmentation of galaxies in overlapping galaxy images. The results are promising with the network being able to attain a detection mean Average precision of 0.9. Further improvements can be done on the architecture to improve performance on semi-transparent objects like galaxies without hard boundaries.

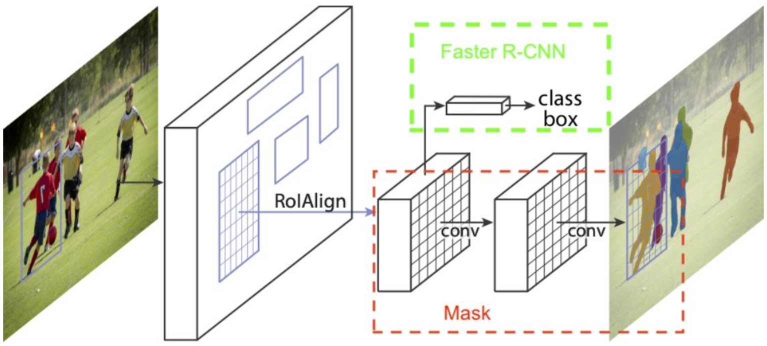

Figure 1: Mask R-CNN framework consisting of Faster R-CNN (green) that performs classification and bounding box prediction and the mask layer (red) that generates the segmentation mask [3].

1 Introduction

Upcoming ground based telescopes like the Large Scale Synoptic Telescope(LSST) coming up in 2020 will observe the sky at resolutions previously not possible. Galaxies too dim to be observed earlier will now be visible, leading to increased overlapping of observed galaxies. Most scientific measurements require individual isolated galaxy images, making it a vital task to correctly separate each galaxy image. Convolution Neural Networks (CNN) are well posed to tackle this problem and perform object detection and separation. Here we present the analysis and results of using Mask Region-based CNN (Mask R-CNN here-after) [3] architecture to perform detection and instance segmentation of galaxies in images of overlapping galaxy pairs.

2 Related Work

Object detection and instance segmentation is an active area of research in computer vision applications. These advances have been largely driven by architectures like Fully Convolution Network (FCN) Fast/Faster R-CNN [1, 7]. In the Region-based CNN (R-CNN) approach [2], the model divides the image into smaller candidate regions and then perform CNN detection independently on each Region of Interest(RoI). Faster R-CNN extends this further with an initial stage that proposes candidate object bounding boxes and a second stage that extracts features using RoIPool from each candidate box and performs classification and bounding-box regression.

An alternate model, fully convolutional instance segmentation (FCIS) [4], predicts a set of position-sensitive output channels fully convolutionally and simultaneously address object classes, boxes, and masks, making the system fast. But, as noted in [3], FCIS exhibits systematic errors on overlapping instances and creates spurious edges. Since our dataset will largely comprise of overlapping objects, Mask R-CNN is better suited for the task.

3 Methodology

3.1 Mask R-CNN

Mask R-CNN comprises of a FPN [5] backbone that predicts RoI, with a Faster R-CNN branch that perform classification and bounding box regression in parallel to a mask layer that generates segmentation masks (Figure 1). The mask branch is a small FCN applied to each RoI, predicting a binary segmentation mask in a pixel-to-pixel manner. This analysis used Tensorflow, Keras implementation of the architecture [1].

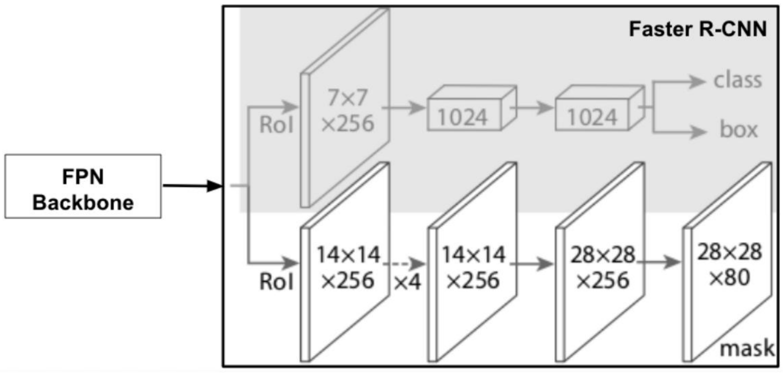

Figure 2: Mask R-CNN head architecture. Numbers denote spatial resolution and channels. "×4" denotes a stack of four consecutive convs [3].

3.2 Loss Function

During training, we define a multi-task loss on each sampled RoI as:

$$L = L_{\text{cls}} + L_{\text{box}} + L_{\text{mask}} \tag{1}$$

The classification loss $L_{\text{cls}}(u, p) = \log(p_u)$ is the log loss for each class u with network predicted probability p. The box loss, L_{box}, is defined over a tuple of true bounding-box regression targets that denote the center of the box and dimensions for class u, $v = (v_x, v_y, v_w, v_h)$, and a predicted tuple $t_u = (t_x^u, t_y^u, t_w^u, t_h^u)$, again for class u is :

$$L_{\text{box}} = \sum_{i \in x,y,w,h} \text{smooth}_{L_1}(t_i^u - v_i) \tag{2}$$

[1]https://github.com/matterport/Mask_RCNN

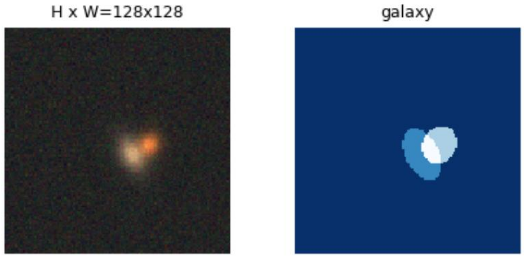

Figure 3: Training data: Input image of overlapping galaxies (left). Segmentation mask for each galaxy (right).

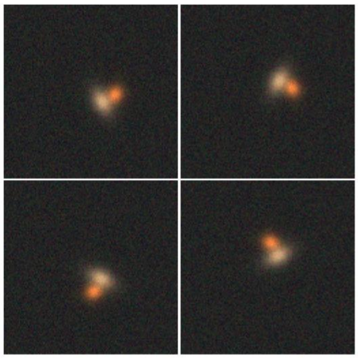

Figure 4: Augmented training data

in which

$$\mathrm{smooth}_{L_1}(x) = \begin{cases} 0.5x^2, & \text{if } |x| < 1. \\ |x| - 0.5, & \text{otherwise.} \end{cases} \tag{3}$$

To each of the predicted masks, a per-pixel sigmoid is applied, and defined L_{mask} as the average binary cross-entropy loss:

$$L_{\text{mask}} = -(y \log(p) + (1-y)\log(1-p)) \tag{4}$$

3.3 Dataset

The complete dataset comprises of 20,000 simulated overlapping galaxy images with varying degrees of overlap. They were simulated using the modular galaxy image simulation toolkit `GalSim`[2]. Segmentation maps were created by selection regions with pixel values above a threshold value for each galaxy. An example image of an overlapping galaxy and the segmentation is shown in Figure 3 (left panel) showing the overlapping image and the segmentation map for each galaxy.

The dataset was then divided into 90% training set, 5% validation and 5% test set as shown in Table 1. Taking advantage of rotational symmetry of galaxy profiles, data augmentation was performed by rotating the images + mirroring the image, in order to increase the training by four as shown in Figure 4.

3.4 Network Training

The network was initialize with weights pre-trained on the Ms COCO dataset. Since our overlapping images vastly differ from the MS COCO dataset [6], we retrain the weights. Initial training of the training set revealed that a major contribution to the loss function arose from the Mask-CNN bounding box loss indicating that the architecture head required more training than the backbone. We proceeded to perform training in multiple stages to improve performance-with the initial training on just the head, keeping the backbone weights fixed and then proceeding to fine-tune all the layers Table 2.

[2]https://github.com/GalSim-developers/GalSim

Table 1: Dividing dataset

Dataset	Training	Validation	Test
20,000	18,000(×4)	1,000	1,000
	(72,000)		
	90%	5%	5%

Table 2: Training Steps, Mini-batch size: 64

Layers trained	epochs	learning rate
Head	1-15	0.001
4+	16-20	0.005
all	20-60	0.001
all	60-80	0.0001

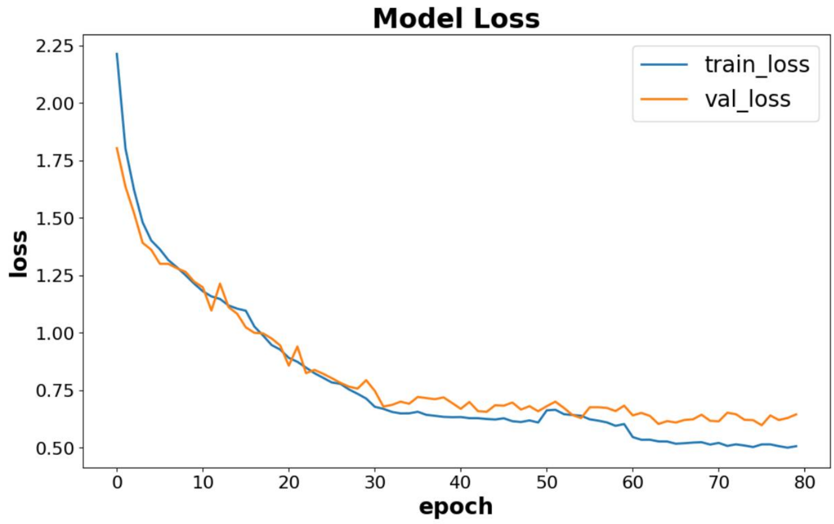

Figure 5: Training set loss(blue) and validation set loss(yellow) per epoch during the 80 training epochs.

The dataset image size is 120 × 120, much smaller than the 800 × 800 for training the MS COCO dataset. Thus the RPN anchor sizes were reduced in proportion in order to be able to detect features with scales desired here.

4 Results

Results from the training over 80 epochs is shown in Figure 5. The blue and yellow lines plot the training set and validation set loss respectively for each training epoch. Initially the training and validation loss decreases with time. After epoch 50 the validation loss remains constant even though the training set loss decreases indicating that training longer will not result in improved performance on the test set.

A summary of the detection performance of the network on the test set are shown in Table 3. While the network is able to detect galaxies with a precision of 75%, it also has a significant number of false positives.

Table 3: Results of evaluation on test set

Precision	Recall	F1 score
0.75	0.625	0.682

An example of a successful detection is shown in Figure 6. The left panel shows the truth image of a two galaxy blend input to the network. Overlaid on it is the segmentation mask and bounding mask, which were not input to the network. The right panel shows

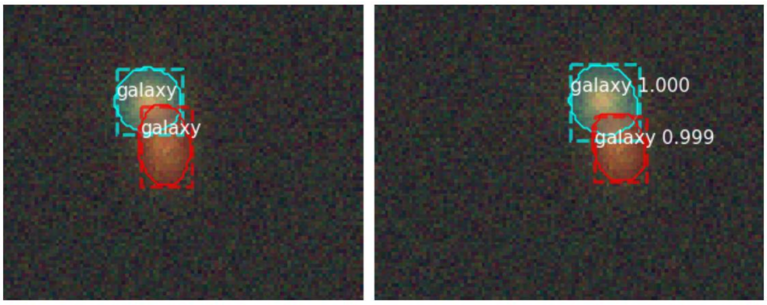

Figure 6: Successful detection: galaxy overlap image with true segmentation mask and box(left) and network output mask and box(right).

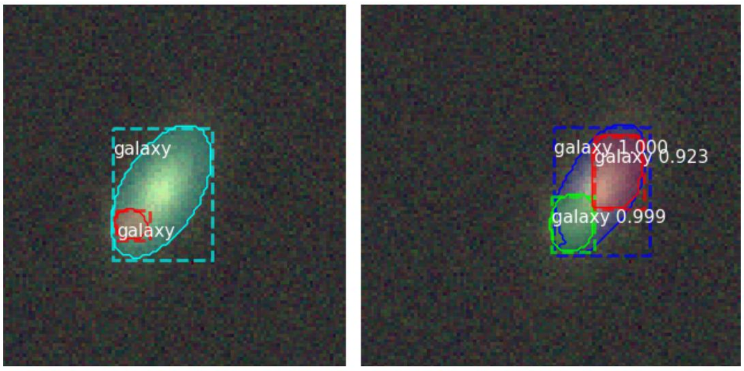

Figure 7: Unsuccessful detection: galaxy overlap image with two galaxies(left) is detected as being composed of three galaxies by the network (right).

the bounding box and segmentation mask for the two galaxies detected by the network. The number on the right panel next to the bounding box is the detection probability.

5 Conclusion

The network was successfully trained and was able to detect individual galaxies in the overlapping pair images along with a segmentation mask and bounding box. The results look promising, however, the detection precision level can be improved as required for scientific analysis.

6 Future Work

The fundamental challenge of using instance segmentation here is that galaxies do not have sharp edges; the brightness dropping with distance from the center. This makes it difficult to develop segmentation maps that capture the exact boundary of the object. In addition unlike ordinary objects, galaxies are relatively transparent. Thus in an overlapping pair the foreground galaxy will not occult the background galaxy, rather the pixels in the overlapping region will include brightness from both objects.

To address these challenges of the model for our application we propose to modify the network for future applications. Instead of generating segmentation maps per object we propose to dissect the final layer the mask generating layer, provide that as input to another network along with the input overlap image, to generate isolated image per object.

The model was tested on simple instances of two-galaxy overlaps that had smooth elliptical profiles. However, real telescope images would involve multiple objects including different classes of galaxies a well as stars. The network can be further trained in these cases to obtain a more useful application for scientific analysis.

References

[1] R. Girshick. Fast R-CNN. *ArXiv e-prints*, April 2015.

[2] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *ArXiv e-prints*, November 2013.

[3] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask R-CNN. *ArXiv e-prints*, March 2017.

[4] Y. Li, H. Qi, J. Dai, X. Ji, and Y. Wei. Fully Convolutional Instance-aware Semantic Segmentation. *ArXiv e-prints*, November 2016.

[5] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie. Feature Pyramid Networks for Object Detection. *ArXiv e-prints*, December 2016.

[6] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár. Microsoft COCO: Common Objects in Context. *ArXiv e-prints*, May 2014.

[7] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *ArXiv e-prints*, June 2015.