AstroYOLO: Learning Astronomy Multi-Tasks in a Single Unified Real-Time Framework

Nodirkhuja Khujaev*
Department of Computer Science and Engineering
UNIST, South Korea
nodirkh@unist.ac.kr

Roman Tsoy*
Department of Industrial Engineering
UNIST, South Korea
romatsoy26@unist.ac.kr

Seungryul Baek
Artificial Intelligence Graduate School (AIGS)
UNIST, South Korea
srbaek@unist.ac.kr

Abstract

In this paper, we proposed a single unified real-time pipeline that jointly performs two tasks: star vs. galaxy detection and smooth type vs. disk type galaxy classification. To achieve the goal, we introduced a model which have two different classification heads sharing the same backbone: The first classification head is used to classify detected objects from the astronomical images into star vs. galaxy; while the second classification head is used to further classify whether the galaxy is smooth or disk type. As the backbone, we used YOLOX architecture, add two classification heads upon it and train them using two heterogeneous datasets: (1) the star vs. galaxy detection dataset which have images including star and galaxy objects and corresponding bounding box and class labels, (2) the smooth vs. disk type classification dataset having galaxy images and their corresponding labels. To prevent the catastrophic forgetting when learning two heads and a backbone, we performed the alternative training between two tasks and also applied data augmentation such as mosaic and mix-up methods. The final model achieved 77.2% accuracy on the smooth vs. disk type classification task, and 65.6 mAP score on star vs. galaxy detection task.

1 Introduction

The task of recognizing astronomical objects such as star and galaxy has been popularly studied and solved as the application of deep learning algorithms [1, 2]. In this paper, we try to combine such efforts based on the single unified model to well capture the hierarchical task: we first (1) detect stars and galaxies from the whole astronomical images, and (2) classify the detected galaxies into either smooth or disk type based on the shape. Previously, smooth/disk types were classified from the center-cropped galaxy images [2]; while the proposed pipeline is more convenient and practical since we can automatically discriminate galaxies from the whole astronomical image and then further classify their shapes. The challenges lie in the insufficient data; and we combined two heterogeneous datasets (ie. whole astronomical images having star and galaxy objects, cropped galaxy images with disk-/smooth types), to train the network. We constituted our framework based on the YOLO [9], especially their YOLO-X [3] version. We compare the proposed framework with other alternative algorithms that may be used to to achieve the same goal. Via the comparison, we demonstrate

*equal contribution

Machine Learning and the Physical Sciences Workshop, NeurIPS 2023.

the superiority of the proposed framework for achieving our goal: multiple astronomy tasks using partially available datasets.

2 Related work

Object detection Object detection is the most prominent and fundamental task in computer vision. Over the past decades, the object detection field has undergone significant progress towards building more accurate and faster algorithms. Early-stage algorithms [11] relied on utilizing shallow learning models and sliding window techniques. While these algorithms were costly and had substantial flaws, they laid a substantial foundation for the subsequent research. With the introduction of convolutional neural networks (CNN) [7, 6], the efficiency and accuracy of object detectors has been cardinally improved and newer algorithms utilizing CNN began to emerge. The R-CNN [4, 10, 5] series of CNN-based models introduced region proposal networks (RPN) using CNN and combined both classification and detection tasks together. One-stage detectors like YOLO[9] and SSD[8] have gained popularity for their real-time performance. Our work is closely related to YOLO architecture, specifically the YOLOX[3] version of the YOLO-series detectors. We utilize YOLOX's ability to perform detection and classification decoupled from each other as it is discussed in Section 3.

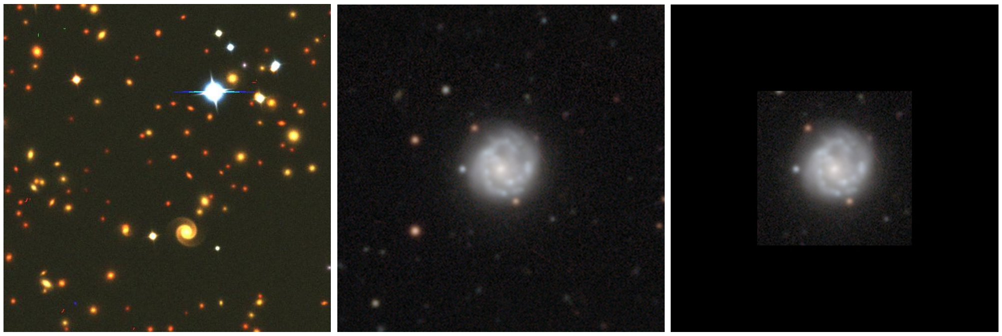
Astrophysics and computer vision Astronomy and computer vision have formed a powerful partnership in recent years, revolutionizing the way we explore the cosmos. One prominent application is the task of classifying celestial objects such as stars and galaxies. Through the integration of advanced computer vision techniques, astronomers can efficiently process vast datasets and identify these celestial entities with remarkable precision. One of the example would also be detecting smooth galaxy vs disk galaxy types. Among the notable developments in this field is the paper "Astro RCNN"[1], which has introduced a novel approach to object recognition using astronomical images, particularly galaxies and stars. This paper exemplifies the fusion of astronomy and computer vision, demonstrating how new deep learning techniques enhance our understanding of the universe. This paper evaluates performance of galaxy vs star task based on Mask RCNN[5] architecture as base framework, and was able to achieve very high performance on both star and galaxy detection. In addition, the model is robust for overlapping objects and mask occlusion, which allows to apply clean deblends for significantly blended objects.

3 Methodology

Our framework builds upon the YOLO [9], specifically YOLOX [3]. We chose YOLOX as the backbone, since it is able to secure the real-time speed as well as the accuracy is good. While, decoupled head in YOLOX enables training the regression and classification head separately from each other. We further used two types of datasets $D = \{D_{sg}, D_{ds}\}$, where D_{sg} is the dataset having $512 \times 512 \times 3$-dimensional RGB images of small objects (i.e. star and galaxy) in the universe which is simulated by the PhoSim project [1], and D_{ds} consists of $424 \times 424 \times 3$-dimensional RGB images selected from Galaxy Zoo DECaLS project [12], which consist of a single object in the center of the picture and represents either a smooth type or disk type galaxy. The example raw images for D_{sg} and D_{ds} are illustrated in the Fig. 1(a) and Fig. 1(b), respectively. The data processing and architectural details and proposed training techniques are detailed in the remainder of the section.

Data processing. We visualized the example samples of D_{sg} and D_{ds} in Fig. 1: Samples in D_{sg} have both star and galaxy objects inside; while samples in D_{ds} have the galaxy object whose type is either smooth or disk. We frequently observed that D_{ds} involves small objects other than the main galaxy, which may act as the noise and hinder the robust recognition of the main galaxy. To tackle the issue, processed the raw images of D_{ds} in Fig. 1(b) to make them as images in Fig. 1(c), by cropping the center patches of the raw images.

Muti-head architecture. To tackle multiple tasks altogether, our framework extends the YOLOX architecture [3] to have three different heads as in Fig. 2. Three heads are directly attached to the YOLOX backbones and we used the identical architecture for each individual head. The first head is used to detect useful objects (ie. star vs. galaxy) from the entire images, the second and third heads are used to achieve the star vs. galaxy classification and smooth vs. disk-type galaxy classification, respectively.



(a) Raw images of D_{sg}. (b) Raw images of D_{ds}. (c) Processed images of D_{ds}.

Figure 1: Examples of two types of data that we used.

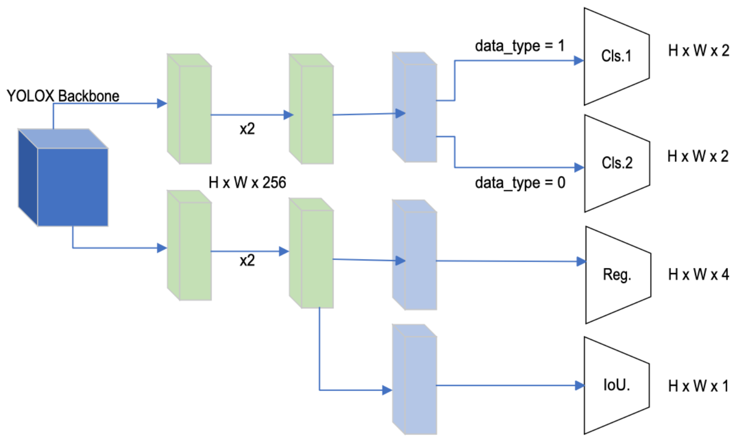


Figure 2: Astro-YOLO architecture having three heads: 1 detection head, 2 classification heads for star vs. galaxy classification and smooth vs. disk-type galaxy classification.

Iterative training. Our end goal is to train a model, which is able to detect useful objects (ie. stars and galaxies) from the entire astronomy images, and classify the objects into stars and galaxies first and then further classify galaxies into smooth and disk types.

To achieve this, we attempted the iterative training of our model: During the iterative training, we first trained the detection head and the 1st classification head using samples of D_{sg} and then train the 2nd classification head using D_{ds} next. Since the detection head and and 1st classification head are trained with D_{sg}; while the 2nd classification head is trained with D_{ds}, we found that naïvely applying the iterative training makes the model forget the previous iteration's knowledge.

To relieve such catastrophic forgetting, we proposed the alternating training scheme: Firstly, we freezed the backbone of YOLOX after training it on the entire D_{sg}. After freezing the backbone, we halved the learning rate of the model and trained it on the entire D_{ds}. This way, the bounding boxes generated for the D_{sg} would not be forgotten after training the model on images from D_{ds}. We only unfreeze the backbone whenever the model was trained on D_{sg}. The freezing scheme is effective; while we observed that when we freeze the backbone during training on D_{ds}, the accuracy becomes limited.

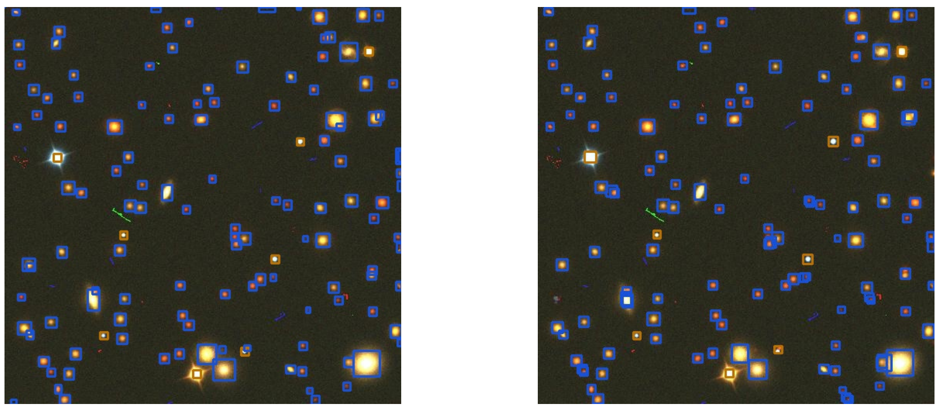
To further tackle the issue, we proposed to sample the equal number of sub-samples from D_{sg} and D_{ds} and train our model without freezing any parts. On each iteration, we randomly sampled either $1,000$ data from D_{sg} and $1,000$ data from D_{ds}. Then, we apply the iterative training by training our detection head and the 1st classification head using sampled data from D_{sg} and training the 2nd classification head using sampled data from D_{ds}. After iterative training, we observed that the model was able to discriminate the majority of bounding boxes correctly for images from either dataset, without suffering from the catastrophic forgetting.

Data augmentation Since D_{ds} data are originally developed for the classification task, we observed that naïvely applying it to our iterative training results in the less robust results for the 2nd classification head. To solve the issue, we additionally applied the mosaic augmentation to extend the classification data towards the detection task. We combined four independent images of D_{ds} to make the new image whose distribution is close to that of D_{sg}.

4 Experiments

In this section, we summarized our results using the proposed YOLOX-based multi-task architecture.

The ground-truth and obtained results are denoted in Fig. 3(a) and Fig. 3(b), respectively. In each map, orange boxes denote the star (label 0) and blue boxes denote the galaxy (label 1). As seen in Fig. 3, our results are good for detecting stars and galaxies; while the inference time for each of the image is only 0.007-0.008 seconds. Making evaluation based on mAP on test set, mAP was 65.7.



(a) Ground truth on star vs galaxy detection. (b) Inference on star vs. galaxy detection.

Figure 3: Ground truth and inference on synthetic data

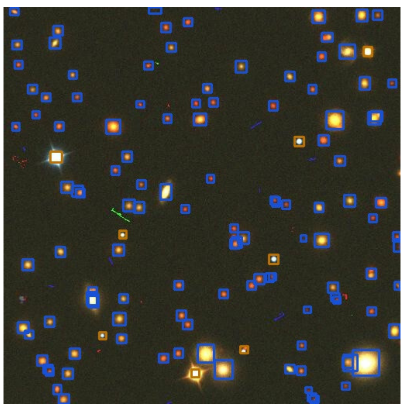
In order to evaluate the performance of the galaxy type classification head which was trained on smooth/disk type dataset (D_{ds}), the final model is used to make inference on the test set. Due to the gap in the feature space captured from both D_{sg} and D_{ds}, we observed that in D_{ds}, some unnecessary small objects are detected. In order to prevent detection of such small objects, the following post-processing approach was applied during inference: if two bounding boxes detected have IOU more than 90%, smaller one is removed, so that we left only 1 detected bounding box with the corresponding label. Evaluating the performance on the classification accuracy of final model on D_{ds}, it is around 73.4%. Evaluation metrics for the final model can be seen in the following table:

Table 1: Evaluation

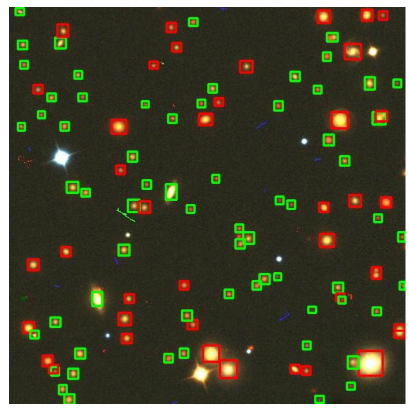
Training		
Methods	mAP(50-90)(D_{sg} data)	Accuracy(D_{ds} data) (%)
Astro-RCNN	49	-
YOLOX(only star vs. galaxy dataset)	65.7	-
YOLOX(only smooth vs. disk dataset)	-	76.8
Training(catastrophic forgetting problem)	0	60.7
Iterative training(frozen backbone)	60.6	59.5
Iterative training(unfrozen backbone)	65.6	77.2

Compared to Astro-RCNN, our final model performs much better. Comparing to the performance of YOLOX models trained only on 1 dataset, YOLOX model trained on a single unified framework (unfrozen backbone) (denoted as 'YOLOX(only star vs. galaxy dataset)') shows similar performance (65.6 mAP on D_{sg} dataset & 77.2 % on D_{ds} dataset) compared to individual YOLOX models, showing 65.7 mAP and 76.8 % respectively.

We additionally perform ablative studies for other alternatives that can achieve the multi-task training. The first baseline is denoted as 'Training (catastrohphic forgetting problem)' that encounters the catastrophic forgetting problem. We only achieve the 0 mAP for star vs. galaxy detection; while we could obtain some accuracy for smooth vs. disk-type classification task using the baseline. The second baseline is denoted as 'Iterative training(frozen backbone)' and we freeze the backbone when training the 2nd classification head, while unfreezing the backbone when training the 1st classification head and detection head. The model was able to accomplish both tasks; but the accuracy is somehow limited. The third baseline is denoted as 'Iterative training(unfrozen backbone)' and we didn't freeze any backbones; while we used the small-size batches for two benchmark datasets. It allowed us to obtain the best performance. Fig. 4 visualizes our results on two tasks: (1) star vs. galaxy classification, (2) smooth and disk-type galaxy classification. Note that, two heads' results are obtained in the parallel manner; while we visualized the second task's results only for boxes classified as 'galaxy' from the first task's results.



(a) Inference using 1st classification head.



(b) Inference using 2nd classification head.

Figure 4: Final results of inference: Orange and blue denote stars and galaxies, respectively in (a), red and green denote smooth and disk types, respectively in (b).

5 Conclusion

The proposed model with multi-heads enables the recognition of multiple attributes of the astronomical objects effectively. Mainly we introduced the effective iterative training scheme to train the model on two heterogeneous datasets. The resulting architecture is a multi-headed architecture that jointly performs both star vs. galaxy detection and smooth vs. disk-type classification from the astronomical images.

6 Acknowledgements

This work was supported by IITP grants (No. 2020-0-01336 Artificial intelligence graduate school program (UNIST) 10%; No. 2021-0-02068 Artificial intelligence innovation hub 90%), all funded by the Korean government (MSIT).

References

[1] Colin J Burke, Patrick D Aleo, Yu-Ching Chen, Xin Liu, John R Peterson, Glenn H Sembroski, and Joshua Yao-Yu Lin. Deblending and classifying astronomical sources with mask r-cnn deep learning. *Monthly Notices of the Royal Astronomical Society*, 490(3):3952–3965, 2019.

[2] H Domínguez Sánchez, M Huertas-Company, M Bernardi, S Kaviraj, JL Fischer, TMC Abbott, FB Abdalla, J Annis, S Avila, D Brooks, et al. Transfer learning for galaxy morphology from

one survey to another. *Monthly Notices of the Royal Astronomical Society*, 484(1):93–100, 2019.

[3] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun. Yolox: Exceeding yolo series in 2021. *arXiv preprint arXiv:2107.08430*, 2021.

[4] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.

[5] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[8] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pages 21–37. Springer, 2016.

[9] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[10] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.

[11] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, volume 1, pages I–I. Ieee, 2001.

[12] Mike Walmsley, Chris Lintott, Tobias Géron, Sandor Kruk, Coleman Krawczyk, Kyle W Willett, Steven Bamford, Lee S Kelvin, Lucy Fortson, Yarin Gal, et al. Galaxy zoo decals: Detailed visual morphology measurements from volunteers and deep learning for 314 000 galaxies. *Monthly Notices of the Royal Astronomical Society*, 509(3):3966–3988, 2022.