EasyChair Preprint

№ 9293

EfficientNet-YOLOv5: Improved YOLOv5 Based on EfficientNet Backbone for Object Detection on Marine Microalgae

Rongsheng Wang, Yukun Li, Yaofei Duan and Tao Tan

November 9, 2022

EfficientNet-YOLOv5: Improved YOLOv5 Based on EfficientNet Backbone for Object Detection on Marine Microalgae

1st Rongsheng Wang
The Faculty of Applied
Macao Polytechnic University
Macao, China
p2213046@mpu.edu.mo

2nd Yukun Li
The Faculty of Applied
Macao Polytechnic University
Macao, China
p2212990@mpu.edu.mo

3th Yaofei Duan
The Faculty of Applied
Macao Polytechnic University
Macao, China
p2213964@mpu.edu.mo

4th Tao Tan*
The Faculty of Applied
Macao Polytechnic University
Macao, China
taotanjs@gmail.com

***Abstract*—Object detection has been a well-known task in deep learning. In IEEE UV 2022 "Vision Meets Algae" Object Detection Challenge, the dimension of the image in the marine microalgae is too large, but the object is too small compared with the images. Additionally, the number of images in each category differs greatly, which brings a great challenge to object detection. We propose EfficientNet-YOLOv5 to solve the two problems mentioned above. Based on YOLOv5, we replaced the Backbone of YOLOv5 with EfficientNet. To further strengthen our proposed EfficientNet-YOLOv5, we offer a variety of useful tricks, such as offline and online data augmentation, multi-scale testing, multi-model ensembled, and LabelSmooling. Extensive experiments on marine microalgae have shown that EfficientNet-YOLOv5 has good performance. It also has very strong interpretability in the marine microalgae scenario. On the marine microalgae test dataset, we used only the EfficientNet-YOLOv5 model and obtained an online score of 44.73%. Compared with the baseline model(scored 42.38%), EfficientNet-YOLOv5 improved by 2.35%. In model ensembled, we received an online score of 50.683% using the ensembled model of EfficientNet-YOLOv5 and YOLOv5s for detection. Overall, our model obtained a considerable improvement in detection accuracy.**

***Index Terms*—YOLOV5, EfficientNet, Marine Microalgae**

I. Introduction

Object detection techniques have been widely used in many scenarios such as plant protection, wildlife protection, and urban surveillance. In this paper, we test the performance of object detection models on marine microalgae and provide guiding directions for the above-mentioned numerous applications. In recent years, significant progress has been made in object detection tasks using deep convolutional neural networks. Several well-known benchmark datasets, such as MSCOCO and PASCAL VOC, have made significant contributions to the growth of object detection application. However, most of the previously proposed deep convolutional neural networks are designed for natural scene images. There are two main problems in using the models directly for the object detection task of marine microalgae images. One is the large variation in the size of different marine microalgae(Fig.1), and the second is the wide range in the number of various microalgae classes. The above two problems make the target detection of marine microalgae images very challenging.

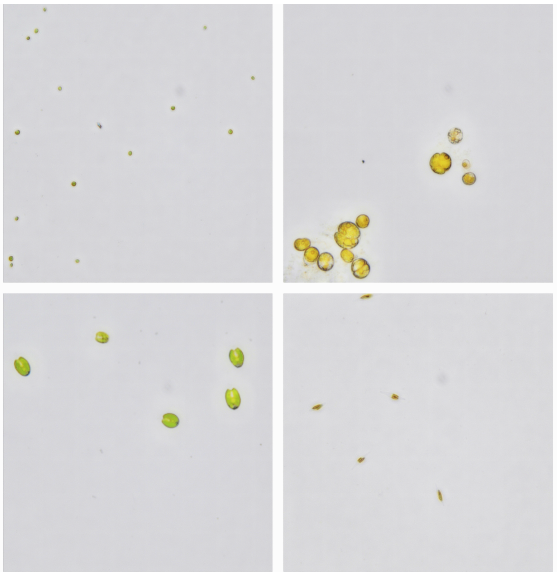

Fig. 1. Explaining the drastic size change of marine microalgae images.

In the object detection task, the YOLO series play an important role in one-stage detectors. To solve the above-mentioned two problems, we propose an improved model, EfficientNet-YOLOv5 based on YOLOv5.Fig.2 describes an overview of the detection pipeline using EfficientNet-YOLOv5. We use EfficientNet and Path Aggregation Network(PANet) as the Backbone and Neck of EfficientNet-YOLOv5, respectively. While the EfficientNet-YOLOv5 Head portion also adheres to the original version. Compared with YOLOv5, our improved EfficientNet-YOLOv5 can handle marine microalgae images better. We use the bag of tricks to enhance YOLOv5's performance even further. In particular, we use data augmentation during training to help the model cope with abrupt changes in object size. We also add multi-scale testing and multi-model

TaoTan is the corresponding author (e-mail: taotanjs@gmail.com).

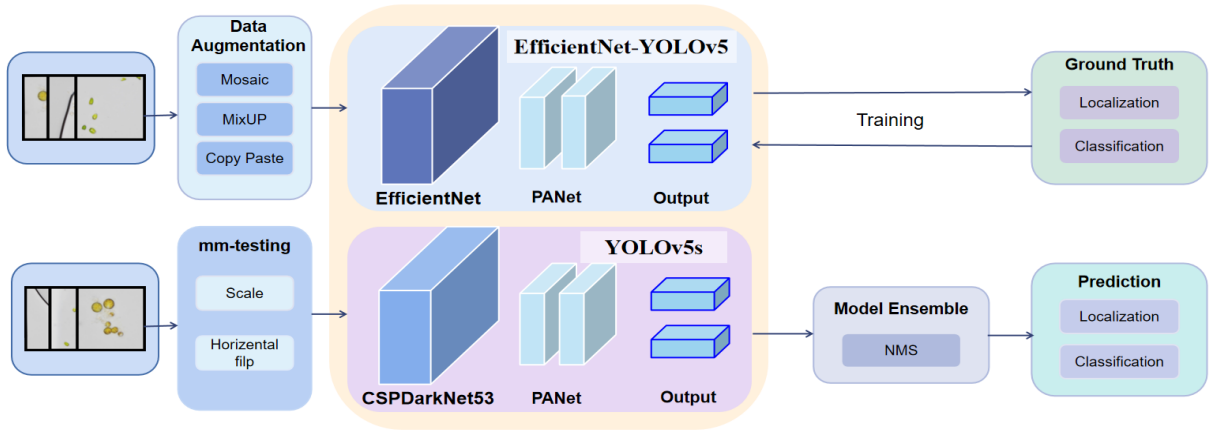

Fig. 2. Workflow overview using EfficientNet-YOLOv5. Compared to the original version, we replaced the Backbone of EfficientNet with YOLOv5. We also used data augmentation, multi-scale testing,multi-model ensembled, and other tricks to make EfficientNet-YOLOv5 more powerful.

ensembled to the inference process to obtain more convincing detection results.

Our contributions are as follows:

1. We replace the original YOLOv5 backbone with EfficientNet to enhance the recognition and classification of the model.

2. We provide a beneficial bag of tricks for object detection tasks in marine microalgae images.

3. On the marine microalgae test-challenge dataset, our proposed EfficientNet-YOLOv5 achieves 44.73%, and by ensembling with YOLOv5s, the score can reach 50.683%. In the IEEE UV 2022 "Vision Meets Algae" Object Detection Challenge, we achieved 42nd place and have a 7.5% gap compared with the 1st place score.

II. Related Work

A. Data Augmentation

The effectiveness of data augmentation is to expand the dataset, so that the model has higher robustness to the images obtained from different environments. Photometric distortions and geometric distortions are wildly used by researchers. As for photometric distortion, we adjusted the hue, saturation and value of the images. In dealing with geometric distortion, we add random scaling, cropping, translation, shearing, and rotating. In addition to the above mentioned global pixel augmentation methods, there are some more unique data augmentation methods i.e. MixUp, Copy Paste and Mosaic.

In EfficientNet-YOLOv5, we use a combination of MixUp, Mosaic,Copy Paste and traditional methods in data augmentation.

B. Model Ensemble

Model Ensemble is the process of combining the outcomes of several different models' predictions to produce more accurate results. Ensemble is based on the notion that distinct good models, trained independently, may perform well for various reasons. In conclusion, each model looks at the data from a different perspective to make predictions. And as a result, only captures a portion of the truth. These truths can be combined to create a more accurate view of the data.

C. Object Detection

Object detection usually consists of three structures. Firstly, the CNN-based Backbone is used to extract image features. Secondly, the Head, is used to predict the object's category and bounding box of this object. In addition, there are some Neck layers between the Backbone and Head. These layers mainly included are as follows.

1. Backbone: Frequently used Backbones include Darknet53, CSPDarknet53, and so on, rather than networks of our design. This is due to the fact that these networks have demonstrated their powerful feature extraction capabilities for classification and other problems. However, researchers need to fine-tune the Backbone to make it more suitable for specific tasks.

2. Neck: Better use of the features extracted by the Backbone. It reprocesses and rationalizes the feature maps extracted by Backbone in different stages.

3. Head: As a classification network, the Backbone is unable to perform the localization task, while the head is designed to be responsible for detecting the location and class of objects through the feature maps extracted from the backbone network.

III. EfficientNet-YOLOv5

A. Overview of Detection

The four models of YOLOv5 are the YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x. Generally, YOLOv5 uses the CSP-Darknet53 architecture, with an SPP layer acting as the backbone, PANet serving as the neck, and YOLO the detecting head. A variety of freebies and discounts are offered to better improve the architecture. We choose it as our baseline since it is the most prominent and practical one-stage detector.

When we train the model based on the marine microalgae using data augmentation strategies (Mosaic, MixUp, and Copy Paste). We find that the results of YOLOv5s are marginally worse than YOLOv5m(YOLOv5s online score is 1% lower than YOLOv5m), while the training computational cost of the YOLOv5m is higher than that of the YOLOv5s. Furthermore such an expense is not worth weighed against the performance improvement, so we opt to employ YOLOv5s in our pursuit of the best detection performance.

B. *EfficientNet-YOLOv5*

The framework of EfficientNet-YOLOv5 is shown in Figure 3. Through a comprehensive analysis of this dataset, it contains many unbalanced classes and small objects. Therefore, we improved the model of YOLOv5 for marine microalgae, replaced the backbone of YOLOv5 with EfficientNet to make YOLOv5 have a more effective detection module. Tan et al. [1] presented the EfficientNet network, which enables the network model to be tuned for accuracy and efficiency.

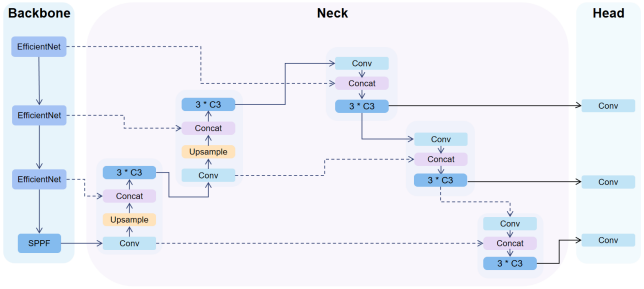

Fig. 3. EfficientNet-YOLOv5 architecture. a) Backbone with EfficientNet as YOLOv5. b) Neck using category PANet structure. c) YOLO Head.

IV. Experiment

A. *Experimental details*

Base environment. We improved EfficientNet-YOLOv5 on Pytorch 1.12.1, and all our models were trained and tested using NVIDIA RTXA4000 GPUs.

Transfer Learning. We use part of the pre-trained model from YOLOv5s and YOLOv5m because EfficientNet-YOLOv5 and YOLOv5 share most of the Head and part of the Backbone, many weights can be transferred from YOLOv5s and YOLOv5m to EfficientNet-YOLOv5, and by using these weights we can save a lot of training time.YOLOv5 is a pre-trained object detection on the MSCOCO datasets.

Hyperparameter settings. We train the model on the training set for only 100 epochs, with the first three epochs used for warm-up. We use SGD optimizer for training and use 1e-2 as the initial learning rate. The input image size for our model is very large, and the image's long side is 1280 pixels, resulting in a batch size of only 8.

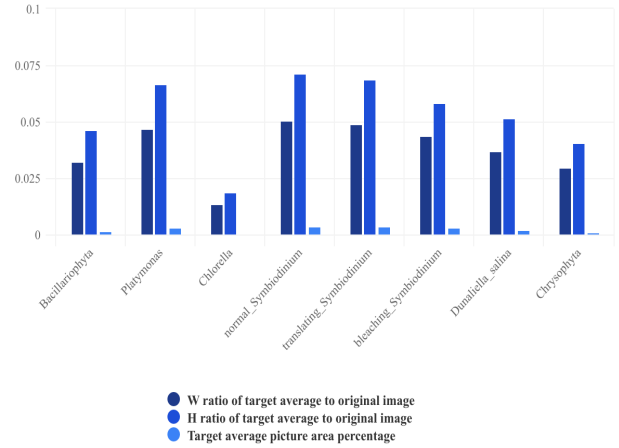

Fig. 4. Percentage of object sizes under different categories in the dataset.

Data Augmentation. Based on previous engineering experience, it is very important to analyze datasets, such as unbalanced data categories and a small amount of data(Fig.4). Therefore, we re-enhance the data by offline data enhancement (Add Random Pixels, Gaussian noise, Cutout, Random Rectangle Occlusion, Gaussian Blur, Motion Blur, Adaptive Histogram Equalization Horizontal Flip, Vertically Flip, Equal Scaling, Random Translation, Strengthen the edge, Random Brightness, Max pooling, Average pooling, Random crop, and padding) are expanded, and the final enhanced training set is 1433 and the validation set is 142.

Test-Time Augmentation(TTA). Data Augmentation is a technique often used to improve performance and reduce generalization errors when training neural network models for computer vision problems. Image data augmentation can also be applied to test datasets when using the model for prediction to allow the model to make predictions for multiple different versions of the image. The predictions of the enhanced images can be averaged to obtain better prediction performance.

B. *Ablation experiment*

Our performance improvement process is shown in Fig.5.

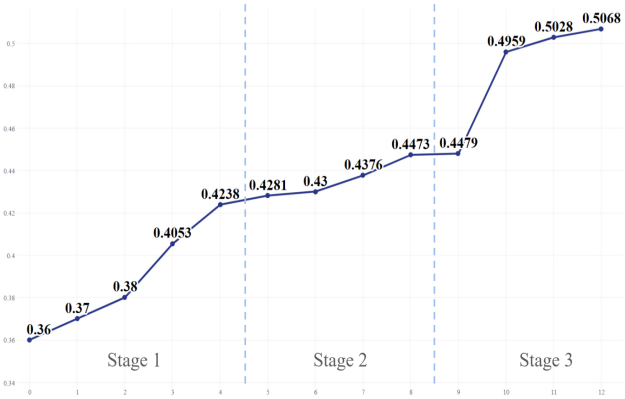

Fig. 5. Online Score Change Chart.

TABLE I
IMPROVED TABLE.

Stage	No.	model	score
stage1	1	YOLOv5s	36%
	2	YOLOv5m	37% (↑ 1 %)
	3	YOLOv5s(+multi-scale,LabelSmoothing)	38% (↑ 1 %)
	4	YOLOv5s(+Mixup=0.1,Copy Paste=0.1,Mosaic=1.0)	40.53% (↑ 2.53%)
stage2	5	YOLOv5s(+Offline Data Augmentation Dataset)	42.38% (↑ 1.85 %)
	6	YOLOv5m(+Offline Data Augmentation Dataset)	42.81% (↑ 0.43 %)
	7	YOLOv5s(+epochs)	43% (↑ 0.19 %)
	8	YOLOv5s(+mixup0.5,Decoupled Head)	43.76% (↑ 0.76 %)
stage3	9	YOLOv5s(+EfficientNet,-Decoupled Head)	44.73% (↑ 0.97 %)
	10	EfficientNet-YOLOv5s,YOLOv5s-DH,YOLOv5s(NO.5)	44.79% (↑ 0.06 %)
	11	EfficientNet-YOLOv5s,YOLOv5s-DH,YOLOv5s(NO.5),YOLOv5s-2(No.7)	49.59% (↑ 4.8 %)
	12	EfficientNet-YOLOv5s,YOLOv5s-2(NO.7)	50.28% (↑ 0.69 %)
	13	EfficientNet-YOLOv5s,YOLOv5s-2(Reduce Confidence,Raise IOU)	50.683% (↑ 0.403 %)

Stage 1: In this stage, we use the provided dataset and split the dataset into an 8:2 ratio. Adding data augmentation and tuning some parameters improves the model performance from 36% to 40.53%.

Stage 2: In this stage, we use the offline data augmentation mentioned in section 5.1 to obtain the new dataset and experiment extensively on YOLOv5s, EfficientNet-YOLOv5, and finally, our proposed EfficientNet-YOLOv5 scored 44.73% online performance on the new dataset.

Stage 3: In this stage, we take full advantage of the excellent performance of different models and combine several different model integrations to obtain better scores. Finally, the integrated YOLOv5s and EfficientNet-YOLOv5 models scored 50.683% online.

V. CONCLUSION

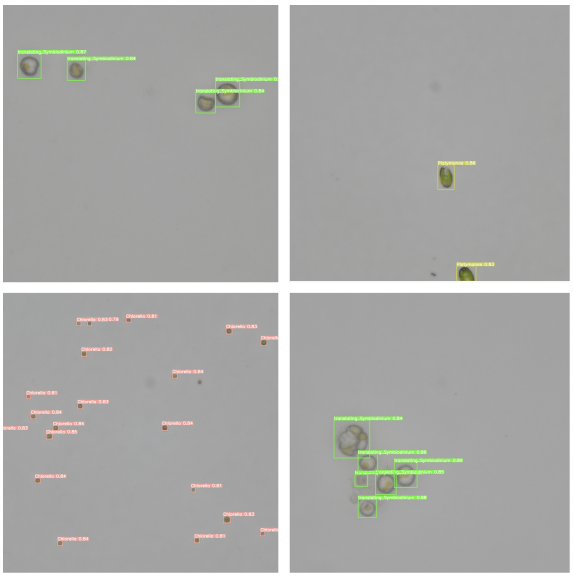

Fig. 6. YOLOv5s and EfficientNet-YOLOv5 ensemble prediction results for marine microalgae.

In this paper, we added some cutting-edge techniques namely EfficientNet, and some tricks to YOLOv5 to improve and form a state-of-the-art detector called EfficientNet-YOLOv5, which is particularly good at object detection in marine microalgae images. we obtained 50.683% online. Our experiments show that the ensemble of EfficientNet-YOLOv5 and YOLOv5s achieves state-of-the-art performance in the marine microalgae test dataset. We have tried a large number of features and used some of them to improve the accuracy of the object detector. We hope this paper will help developers and researchers to have a better experience in the analysis and processing of marine microalgae(Fig.6).

VI. FUTURE WORK

A. Mapping Image

The marine microalgae dataset is the Main challenge in object detection because it is unbalanced. For imbalanced data, reconstruction is a good strategy. For some objects with tiny numbers over the entire dataset, the features learned by the model are insignificant. To improve the training, we used mapping techniques(Fig.7) to do data augmentation, which increases their number on the image to enhance the training.

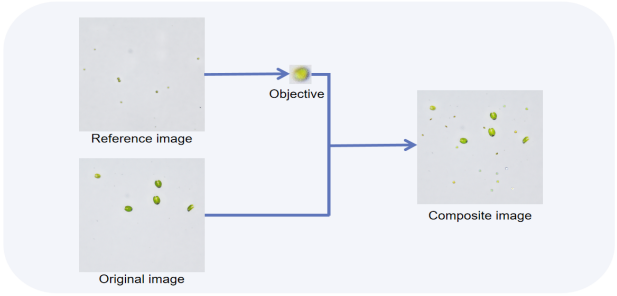

Fig. 7. Overview of the mapping augmentation process for datasets.

B. Cascaded Classification Network

Marine microalgae can be located and classified effectively using the enhanced EfficientNet-YOLOv5 model based on YOLOv5, but a common issue is the object's tiny size and the absence of visible classification features. Therefore, we believe that using EfficientNet-YOLOv5 and cascading a classification network for secondary subclassification is a more effective solution.

The specific implementation is as follows: EfficientNet-YOLOv5 is used for localization and classification, a cascaded classification network for secondary subclassification of detected objects. The subclassification results are combined with the EfficientNet-YOLOv5 localization and classification results to produce the final results (Fig.8).

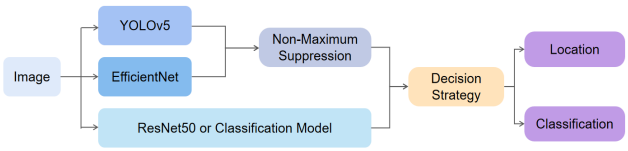

Fig. 8. Methodological process of cascade classifier for localization and classification.

REFERENCES

[1] Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." International conference on machine learning. PMLR, 2019.