

Deep neural network-based real time fish detection method in the scene of marine fishing supervision

Junpeng Li^a, Kaiyan Zhu^{a,*}, Fei Wang^b and Fengjiao Jiang^a

^a*Journal School of Information Engineering, Dalian Ocean University, Dalian, China*

^b*School of Software, Dalian Jiaotong University, Dalian, China*

Abstract. Overfishing of marine fishery is a serious threat to fishery ecological security. Fishing regulation is one of the main ways to maintain marine fishery ecology. In order to improve the intelligence of fishing supervision system, a real time fish detection method based on YOLO-V3-Tiny-MobileNet was proposed. Aiming at the problems of shallow network layers and insufficient feature extraction ability in YOLO-V3-Tiny network, the proposed network takes YOLO-V3-Tiny as baseline and combines it with MobileNet. The proposed network is pre-trained by VOC2012 dataset, and then retrained and tested on Kaggle_NCFM (The Nature Conservancy Fisheries Monitoring) dataset. The experimental results show that the proposed method has superior performance in parameters number, mean average precision and detection performance, compared with other methods. Compared with the monitoring method of fishing vessel detection on shore supervision, the real time monitoring method can give timely warning to the fishing vessel operators, which is more conducive to fishery ecological protection.

Keywords: Marine fishery, fishing supervision, real time fish detection, YOLO-V3-Tiny, MobileNet

1. Introduction

At present, the phenomenon of overfishing in marine fisheries is still very serious, and the existing regulatory methods are also facing many difficulties. The state of world fisheries and aquaculture in 2018 was released by the Food and Agriculture Organization of the United Nations, which pointed out that since the 1970s, the situation of global wild fish overfishing has deteriorated at least three times. The global fishing quantity has approached the limit value of sustainable development of fisheries, and about 90% of wild fish are facing overfishing. In order to prevent overfishing and protect the ocean, ecological environment, many countries have formulated relevant laws and regulations, but illegal and dis-

ordered fishing is still very serious. Illegal and disordered marine fishing may lead to the gradual loss of species regeneration ability of some fish species, causing great damage to the marine ecological environment and seriously threatening the sustainable development of marine ecology. Fishery regulation is one of the main ways to maintain marine fishery ecology. The traditional fishing supervision methods mainly include the supervision of on-board observers, sonar detection of the number of undersea fish, and artificial analysis after fishing video captured by fishing boat camera. These methods are highly dependent on people, and may have problems such as high cost, low efficiency and even regulatory inaction. In order to ensure the ecological safety of marine fish and realize the continuous monitoring of fishing vessels for a long time, many countries have begun to use the electronic monitoring camera of fishing vessels combined with computer vision technology to supervise the fishing situation of fishing vessels [1, 2]. At

*Corresponding author. Kaiyan Zhu, Journal School of information engineering, Dalian Ocean University, Dalian 116023, China. E-mail: zhukaiyan@dlou.edu.cn.

present, most of the methods of fish fishing supervision are mainly explored from the perspective of shore supervision. This kind of method belongs to the post supervision method. These methods can only punish the illegal behavior after the violation is found, but can not give the operation warning to the on-site operators of fishing vessels in time.

In order to solve these problems, this paper proposes a deep neural network named YOLO-V3-Tiny-MobileNetwork (YOLO-V3-Tiny-MobileNet), which can has good feature extraction ability with low computational complexity. Compared with other related methods, this method has the best comprehensive effect, can realize the real time fishing supervision on fishing vessels, and provide reliable technical support for the intelligent upgrading of fishing supervision system.

The rest of the paper is organized as follows. Section II gives a brief introduction about the related works. Section III describes the structure of YOLO-V3-Tiny-MobileNet. Section IV introduces the experimental process. Section V reports the results of the proposed method on the Nature Conservancy Fisheries Monitoring dataset Kaggle. NCFM.

2. Related works

In recent years, with the rapid development of artificial intelligence and machine learning, deep learning, which can automatically extract the features of objects, has become a great success in the field of target recognition and target detection. Researchers have gradually applied deep learning to the field of fish recognition and detection. Ding [3] extracted image features with CNN, and used softmax layer as classification layer, and achieved 96.51% recognition accuracy on fish4 knowledge dataset [4]. Qin [5] used CNN to extract image features and SVM as classifier, which achieved 98.64% accuracy on f4k fish dataset. Based on the idea of transfer learning, Siddiqui [6] respectively used VGG-net, Res-net and other pre training models as universal feature extractors, and used support vector machine as classifier. When classifying and recognizing fish in underwater video images collected from the coast of Western Australia, the accuracy rate of the optimal model reached 94.3%. By transfer learning, Wang [7] respectively used VGG16, ResNet50 and other pre-trained models as universal feature extractors, and softmax layer as classifier to identify four aquatic animals, including fish, shrimp, shellfish and crab, and the optimal

model recognition rate reached 97.4%. Yuan [8] proposed an underwater fish detection method based on migration learning and fast R-CNN, with a precision rate of 98%. Li [9] proposed a method for underwater fish detection based on transfer learning and YOLO, improves the speed of target detection, with an accuracy rate of 93%, and realizes real time detection of underwater fish.

However, there are not many systematic theoretical research papers focusing on fish identification methods with deep learning in the complex scene of fishing supervision. In 2019, the author proposed a fish identification method specifically for the complex scene of marine fishing supervision. The transfer learning and model fusion method can effectively solve the problem of marine fish identification in complex scenes [10]. In order to supervise the fishing vessel operation in time and protect fishery ecology better, this paper proposes a real time fish detection method, which takes YOLO-V3-Tiny as the baseline and combines it with MobileNet to form the network structure. The improved network is pre-trained by VOC2012 dataset, and then retrained and tested on NCFM dataset. Compared with other related methods, this method has advantages on model size, parameters number, recognition accuracy and convergence speed.

3. Structure of YOLO-V3-Tiny-MobileNet

YOLO-V3-Tiny is simplified on the basis of YOLO-V3 and is specially designed for deployment on low-performance computing platforms. YOLO-V3-Tiny uses a network composed of convolution and maximum pooling as the backbone network for feature extraction, and adopts $13 * 13$ and $26 * 26$ resolution YOLO output layers. The advantage of this method is that the network is relatively simple and the amount of calculation is small. However, the backbone network of YOLO-V3-Tiny is shallow, so it is difficult to extract higher-level semantic features, and the detection accuracy is not enough. Therefore, this paper improves the backbone network of YOLO-V3-Tiny by combining MobileNet [11], which improves the detection accuracy on the premise of ensuring the detection speed and lightweight.

The overall structure of YOLO-V3-Tiny-MobileNet is shown in Fig. 1. The second part of the network retains the structure of YOLO-V3-Tiny detection part, and replaces the original feature extraction network with MobileNet. Compared with

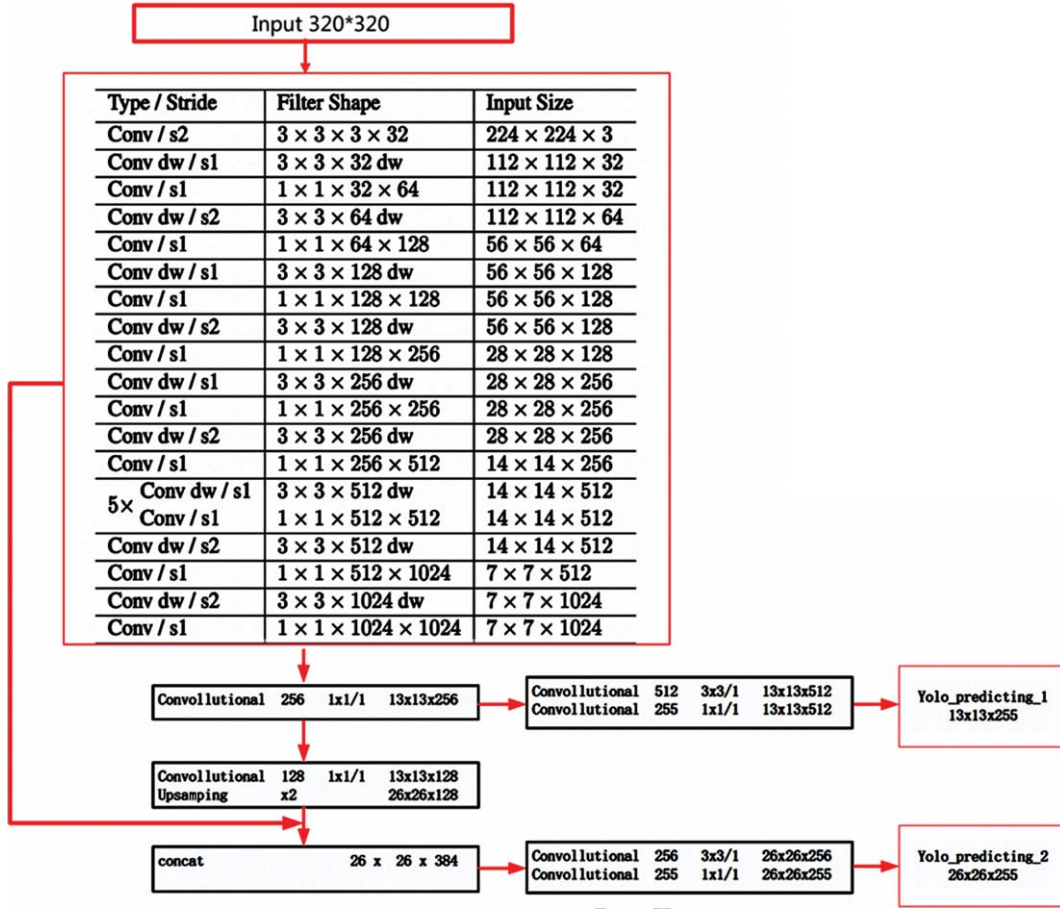


Fig. 1. Structure of YOLO-V3-Tiny-MobileNet.

the original feature extraction network, MobileNet has a deeper network layer, while the parameters and computational complexity are still low. The feature of depth-wise and point-wise in MobileNet reduces the amount of convolution calculation, and is equipped with a reduction factor to adjust the relationship between calculation speed and accuracy. The number of the improved network parameters are 3.17 million, while that of the original feature extraction network is 8.6 million.

4. The results and analysis

The establishment of neural network model for automatic learning target features is often complex and time-consuming. Therefore, in the field of image recognition and detection, transfer learning is often used to speed up the learning efficiency. This paper compares and analyzes the recognition effect

of direct training and transfer learning. For direct training, NCFM fish dataset labeled by LabelImg is used to train the target detection network directly. For transfer learning, the pre-training is carried out on VOC2012 dataset, and then the obtained pre-trained model is used to fine tune and optimize on NCFM dataset. For convenience of expression, YOLO-V3-Tiny-Pre, YOLO-V3-MobileNet-pre, YOLO-Lite-pre and YOLO-V3-Tiny-MobileNet-Pre are used to represent YOLO-V3-Tiny network, YOLO-V3-MobileNet, YOLO-Lite network and YOLO-V3-Tiny-MobileNet after transfer learning pre-trained, respectively.

The experiments in this section are conducted on NVIDIA Tesla P40, which is a GPU device with powerful computing performance. In addition to using NCFM fish dataset, VOC2012 dataset is chosen for pre-training. There are 20 kinds of objects in VOC2012, and 21 types of background, including human, animal, car and indoor objects. The size of the

Table 1
Comparison of AP and mAP

Catego-ry\Model	YOLO-V3-Tiny	YOLO-V3-Tiny-Pre	YOLO-V3-MobileNet	YOLO-V3-MobileNet-Pre	YOLO-Lite	YOLO-Lite-Pre	YOLO-V3-Tiny-MobileNet	YOLO-V3-Tiny-MobileNet-Pre
ALB	0.4871	0.4112	0.3698	0.4456	0.3237	0.3166	0.533	0.518
BET	0.6696	0.5071	0.4829	0.6185	0.4772	0.5259	0.6221	0.6907
DOL	0.3004	0.637	0.2533	0.4094	0.2000	0.2444	0.3785	0.4042
LAG	0.6619	0.8977	0.5625	0.7375	0.4427	0.7969	0.9726	0.9901
SHAKE	0.5776	0.7564	0.5820	0.6826	0.6329	0.5400	0.7338	0.7982
YFT	0.5797	0.5716	0.5421	0.5936	0.5497	0.4767	0.7212	0.7102
mAP	0.5461	0.6302	0.4654	0.5812	0.4383	0.4834	0.6602	0.6853

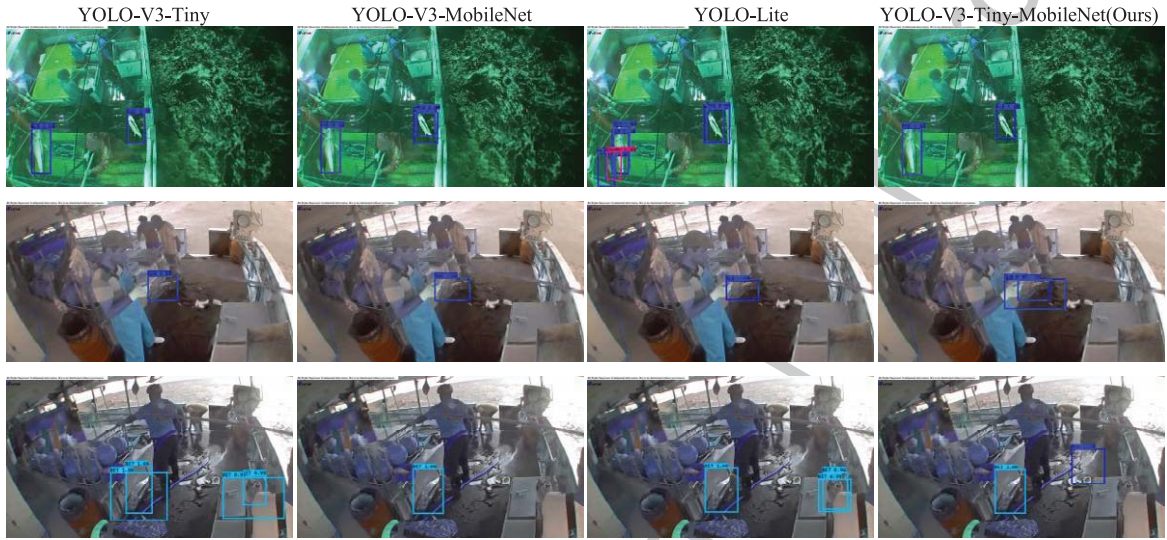


Fig. 2. Detection results comparison of our method and other methods with direct training.

dataset is about 2 G, including 5717 training images and 13609 targets.

4.1. Model size

For lightweight real-time detection model, the space occupation and parameter scale of the model are very important. The size of YOLO-V3-MobileNet, YOLO-V3-Tiny, YOLO-Lite, YOLO-V3-Tiny-MobileNet are 96.9MB, 33.1MB, 1.18MB and 23.7MB respectively. The space occupied by YOLO-V3-Tiny-MobileNet is smaller than that of YOLO-V3-Tiny and YOLO-V3-MobileNet.

4.2. Mean average precision (mAP)

Mean average precision (mAP) is one of the important indexes to measure model accuracy in target detection task. It is calculated based on AP (average precision). AP is calculated by precision rate and recall rate. Strictly, the average precision is precision

averaged across all values of recall between 0 and 1 [12]. The accuracy of each level pinterp is calculated by interpolation [13].

Table 1 compares AP and mAP of above eight models, which is YOLO-V3-Tiny, YOLO-V3-MobileNet, YOLO-Lite, YOLO-V3-Tiny-MobileNet and their pre-trained models. It can be seen that the mAP of pre-trained methods are higher than that of corresponding directly trained methods. Meanwhile, the Map of YOLO-V3-Tiny-MobileNet-Pre on NCFM dataset is 0.6853, which is higher than other methods. It also has good performance on individual categories.

4.3. Detection results analysis

In this section, detection results of the proposed method are compared with other related methods. The detection results are shown in Figs. 2 and 3. Among them, Fig. 2 shows the detection results of YOLO-V3-Tiny, YOLO-V3-MobileNet, YOLO-Lite and YOLO-V3-Tiny-MobileNet directly trained

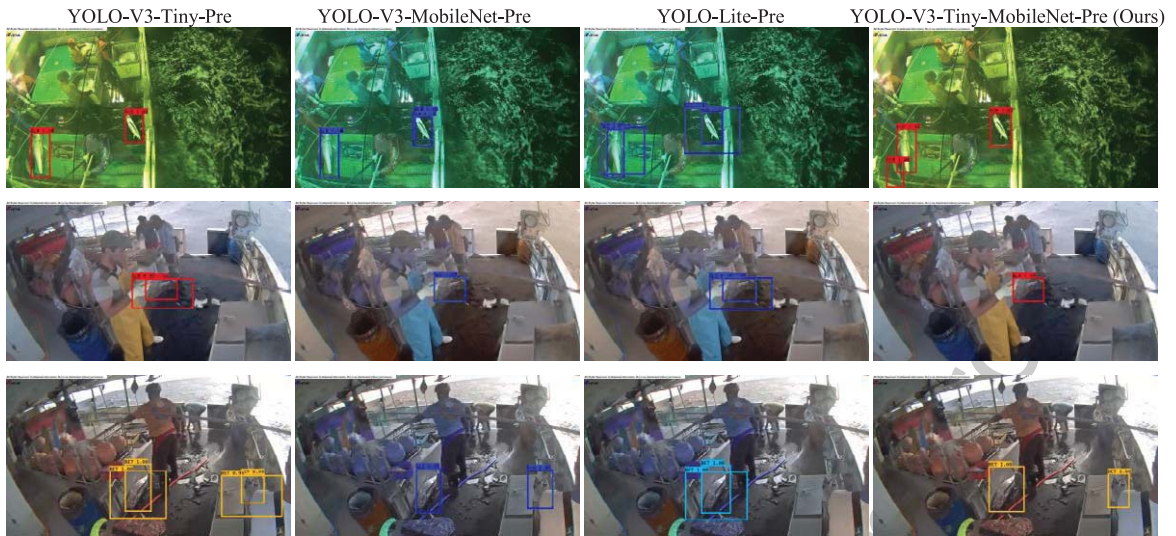


Fig. 3. Detection results comparison of our method and other methods with transfer training.

by NCFM fish dataset. Figure 3 shows the detection results of above methods after transfer learning training.

In Fig. 2, it can be found that each model has the phenomenon of missing detection or false detection. For example, the fish in the lower left corner of the first image is missed detected, the gloves in the second image are recognized as fish, and the faucet or water stains on the ground in the third image are recognized as fish.

Figure 3 shows the detection results of the transfer learning training models. Except YOLO-V3-Tiny is an open-source pre-trained model, the other three models are pre-trained on the VOC2012 dataset. Then the pre-trained model is fine-tuned on the NCFM dataset. As can be seen from Fig. 3, in the case of occlusion, YOLO-V3-Tiny-MobileNet-Pre (ours) has better detection effect and more accurate target frame, while the other three models may miss fish detection and mistake gloves or faucets for fish. Compared with Figs. 2 and 3, it can be concluded that YOLO-V3-Tiny-MobileNet-Pre can better solve the problem of missed and false detection in longline fishing video.

5. Conclusions

This paper proposes a real-time monitoring method for fishing, which can better protect fishery resources compared with other shore monitoring methods. A real-time fish detection method based on YOLO-V3-

Tiny-MobileNet is proposed. Aiming at the problem of low recognition accuracy caused by the shallow network layers of YOLO-V3-Tiny feature extraction, the proposed model combines YOLO-V3-Tiny with MobileNet, is pre-trained by VOC2012 dataset, and then retrained and tested on NCFM dataset. At last, the paper compares and analyzes the model storage size, mAP and detection performance with the existing algorithms. It shows that the proposed method is outstanding in many aspects such as mAP, detection performance, it also has relatively reasonable model size and detection speed. The proposed method can better solve the problems of missed detection and false detection in longline fishing video. So, the proposed method provides an efficient and feasible idea for the real-time fish detection in the marine fishing supervision scene, and lays a good foundation for the intelligent upgrading of the fishing supervision system.

Acknowledgments

This project was supported by Liaoning Province Natural Science Foundation Program (No. 2019-ZD-0731), Liaoning Province Education Department Science Research Program (No. QL201913).

References

- [1] A.T.M.V. Helmond, L.O. Mortensen and S.P.H. Kristian, Electronic monitoring in fisheries: lessons from global

- experiences and future opportunities, *Fish and Fisheries* **21**(1) (2020), 162–189.
- [2] D.C. Bartholomew, J.C. Mangel and J. Alfaro-Shigueto, Remote electronic monitoring as a potential alternative to on-board observers in small-scale sheries, *Biological Conservation* **219** (2018), 35–45.
- [3] G. Ding, Y. Song and J. Guo, Fish recognition using convolutional neural network, *Oceans Conference*, Anchorage, USA, 2017, 1–4.
- [4] B.J. Boom, P.X. Huang and J. He, Supporting ground-truth annotation of image datasets using clustering, *21st International Conference on Pattern Recognition*, 2012, 1542–1545.
- [5] H. Qin, X. Li and J. Liang, Deep fish: accurate underwater live fish recognition with a deep architecture, *Neurocomputing* **187** (2016), 49–58.
- [6] S.A. Siddiqui, A. Salman and M.I. Malik, Automatic fish species classification in underwater videos: exploiting pre-trained deep neural network models to compensate for limited labelled data, *ICES Journal of Marine Science* **75**(1) (2017), 374–389.
- [7] K.L. Wang and H.C. Yuan, Aquatic animal image classification method based on transfer learning, *Journal of Computer Applications* **38**(5) (2017), 1304–1308.
- [8] H.C. Yuan and S. Zhang, Detection of underwater fish based on faster R-CNN and image enhancement, *Journal of Dalian Ocean University* **35**(4) (2020), 612–619.
- [9] Q.Z. Li, Y.B. Li and J. Niu, Real-time detection of underwater fish based on improved Yolo and transfer learning, *Pattern Recognition and Artificial Intelligence* **32**(3) (2019), 193–203.
- [10] J.P. Li, K.Y. Zhu and S. Yang, Ocean fish recognition in complex scene based on transfer learning, *Computer Applications and Software* **36**(09) (2019), 168–174.
- [11] A.G. Howard, M. Zhu, B. Chen, et al., MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications (2017).
- [12] M. Everingham, G.L. Van and C.K. Williams, The pascal visual object classes (VOC) challenge, *International Journal of Computer Vision* **88**(2) (2010), 303–338.
- [13] G. Salton and M.J. McGill, Introduction to modern information retrieval, *Information Processing and Management* **19**(6) (1983).