Deep convolutional neural network and data augamentation for fish detection and classifaction

Ziqiang Zheng, Chao Wang, Zhibin Yu*, Haiyong Zheng, Weiwei Wang
College of Information Science and Engineering
Ocean University of China
Qingdao 266100, China
Email: zhengziqiang1@gmail.com
*Corresponding author: yuzhibin@ouc.edu.cn

***Abstract*—Nowadays, as a sub topic of computer vision and fishery industry, fish recognition is still a challenging work not only because of various kinds of fish, but also because of the complex background of images. In this paper, we aim to classify different fish images obtained from cameras of fishing vessels. Fish detection is different with the best known and the most well-investigated object detection - face detection. Fish has more different shapes than human faces. And fish always take only a small part of the whole image. For these two reasons, it's a challenge for us to detect and classify the fish. Our work is done with Kaggle dataset. The Kaggle dataset aims to detect and classify the species of fish. The competition provides a train dataset which contains six species of tuna fish, fish but not tuna and no fish. Eight target categories are available in the dataset. The dataset is critically imbalanced, the Albacore tuna has thousands of images while the Opah has only sixty images. In order to overcome these problems we introduced two methods to improve our accuracy for the fish classification. Our approach can avoid over-fitting caused by the imbalance of training dataset. We employ the AlexNet, GoogLeNet, and VGGNet neural network for the classification. As a result of the small dataset, the VGGNet architecture performs worse than AlexNet. We propose a local region based the fish area modeling approach so that the local feature can be modeled. In order to obtain more image information and realize imbalanced datasets classification, we have done some data augmentation. In the future, we will use the py-faster-rcnn to do the detection and achieve better performance.**

I. Introduction

Nearly half the population of the world depends on seafood for their main source of protein. And fish play an important role in our marine ecosystem. So it's valuable for us to do the fish detection and classification. We research to achieve the fish detection and classification automatically, which has drawn increasing attention.

In the western and central pacific, where 60% of the world's tuna is caught, illegal, unreported , and unregulated fishing practices are threatening marine ecosystems. The Nature Conservancy is inviting the Kaggle community to develop the algorithms to automatically detect and classify species of tunas. The Kaggle dataset includes Albacore tuna, Bigeye tuna, Yellowfin tuna, Mahi, Opah, Shark, Other(meaning that there are fish present but not in the above six categories), and No Fish(meaning that no fish in the picture). Each image has only one fish category. Fig. 1 shows the imbalance of the Kaggle dataset.

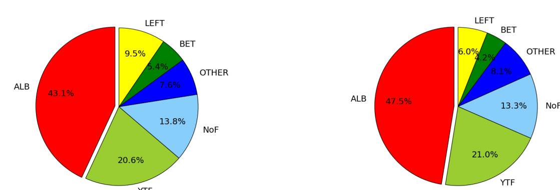
TABLE I
Kaggle Database

classes	Total	Training	Valiation
ALB	45.5%	43.1%	47.5%
YFT	19.4%	20.6%	21.0%
NoF	12.3%	13.8%	13.3%
OTHER	7.9%	7.6%	8.1%
BET	5.2%	5.4%	4.2%
Sum	90.3%	90.5%	94.0%

TABLE II
Singal model result on full data

database	model	accuracy
full	alexnet	0.94
full	googlenet	0.942
full	vgg16	0.44

In this paper, we introduce two methods to improve the classification accuracy. In consideration of the imbalance of the dataset, we got a new image by rotating the selected image. Through this method, we can increase the number of some species of fish images, which can help avoid the over-fitting, and rotating the fish images can improve the robustness of detection. We doubled our datasets by this way and the accuracy increased two percent. And we will employ the faster-rcnn to achieve the detection and classification. There have been less efforts made on solving fish detection and classification simultaneously[1].



(a) Training database (b) Valing database

Fig. 1. Data distribution of Kaggle

The second improvement is based on an image preprocess-

TABLE III
KAGGLE DATABASE

database	model	accuracy
full+rotating	alexnet	0.9673
full+rotating	googlenet	0.964
full+rotating	vgg16	0.45

TABLE IV
KAGGLE DATABASE

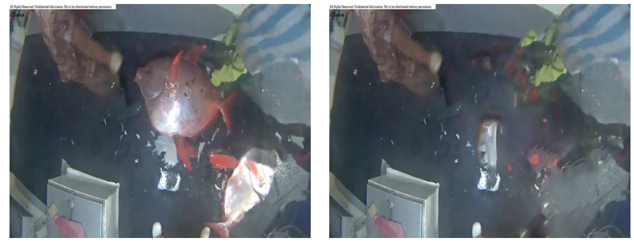
database	model	accuracy
full+rotating+mask	alexnet	0.9703
full+rotating+mask	googlenet	0.954

ing method. In order to force the neural networks focusing on fish not the vessel, we have made a mask of the area where the fish are located, and we used both the processed images and the raw images to train our model. Fig. 2 shows the differences between the two images. This method can force our neural network to concentrate on the area where the fish present. And the features are learned from comparing the raw images with the images with mask. In order to prove our assumption, we test one fish image and get the image data from the last pooling layer. The higher value area in the pooling layer means higher contribution to the final decision. Through comparing the pooling gray image with the raw image, we found that the area of fish is brighter than other areas, which proves what we expected that the fish area make great contribution in the classification. Fig. 3 illustrates the image information of the region of fish and other regions without fish. We can see that the region of fish is brighter than other regions and the pooling image data can confirm that the learned features are from fish. In addition, we use the processed images to fine-tune our pre-trained model and it performs better when testing.

The architecture of our model used in the experiment is the convolutional neural network(CNN), which has achieved remarkable performance in image classification and object detection recently. We employ the CNN architecture for the fish classification. But there still exists the problem that our model can not distinguish between foreground and background. Our methods can prevent the CNN architecture model from concentrating on the boat region rather than the fish region.

II. EXPERIMENTS

The Table I indicates the details of the Kaggle Database. The first column is the name of fish and the numbers in the rest columns mean that the proportion of this class in all fish. And the Fig. 1 is also the another form of fish proportion. We can see the distribution of the dataset is imbalanced. The Table II illustrates the benchmarks. We can see the AlexNet and the GoogLeNet achieved greater performance than the VGGNet. The Table III shows the accuracy that we have rotated some images to get new images. From the table we can see this



(a) the raw image (b) image with mask

Fig. 2. We make a mask of the area where the fish are located because the mask can force our model concentrate on the fish area. So it can help our convolutional nueral network detect the fish. we have trained our model useing both the two images at the same time, which can improve the robustness.

(a) the image (b) the pooling image

Fig. 3. The region of fish are brighter than other regions, which can prove that the fish has more influence on the classification.

method actually worked out and the accuracy has increased about two percent. Basing this, we have made a mask of the region of fish. The accuracy has also increased in the Table IV, which proves that our methods are effective.

III. CONCLUSION

In this paper, we used two methods to improve deep neural networks to detect the fish more efficiently. The experiment results show that our methods can classify fish from images captured on a camera of fishing vessels well. And the classification accuracy has increased more than two percent.

REFERENCES

[1] H. Qin, X. Li, J. Liang, Y. Peng, and C. Zhang, "Deepfish: accurate underwater live fish recognition with a deep architecture," *Neurocomputing*, 2015.

[2] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.

[3] M.-C. Chuang, J.-N. Hwang, and K. Williams, "A feature learning and object recognition framework for underwater fish images," *IEEE Transactions on Image Processing*, vol. 25, no. 4, pp. 1862–1872, 2016.

[4] K. Anantharajah, Z. Ge, C. McCool, S. Denman, C. Fookes, P. Corke, D. Tjondronegoro, and S. Sridharan, "Local inter-session variability modelling for object classification," in *IEEE Winter Conference on Applications of Computer Vision*. IEEE, 2014, pp. 309–316.

[5] Z. Ge, C. McCool, C. Sanderson, and P. Corke, "Modelling local deep convolutional neural network features to improve fine-grained image classification," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 4112–4116.