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

Ziqiang Zheng
College of Information Science and Engineering
Ocean University of China
Qingdao 266100, China
Email: zhengziqiang1@gmail.com

***Abstract*—Fish recognition is one of the most significant application of computer vision in fish survey, and the fish classification is crucially required. Different from the best known and the most well-investigated object detection - face detection, fish usually occupy a smaller area in an image than the face of one person. Besides some pictures are taken under water and the pictures are not clear and dark for the light is absorbed by water. For these two reasons, it's a challenge for us to detect and classify the fish. The Kaggle competition 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 this paper, 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 proposed a local region based the fish area modelling approch so that the local feature can be modelled. In order to obtain more image information and realize imbalanced datasets classification, we have done some data augmentation. In the futrue, we will use the py-faster-rcnn to do the detection and achieve better performance.**

I. Introduction

Nearly half 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 avluable for us to do the the fish detection and classification. It's a waste of time and money for humans to do it. We research to achieve the fish detection and classificaion 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.

In this paper, we introduced two methods to improve the accuracy. In consideration of the imbalance of the dataset, we got a new picture 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. By rotating the fish images, we can improve the robustness of detection. We doubled our datasets by this way and the accuracy increases two percent. On the other hand, 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].

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

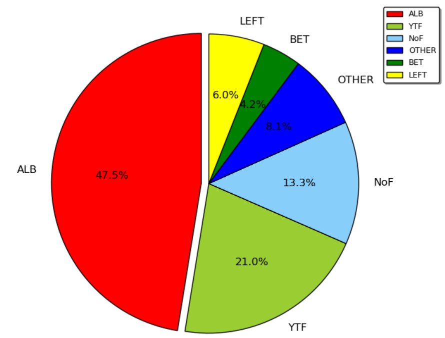
database	model	iteration	accuracy
full	alexnet	70k	0.94
full	googlenet	40k	0.942
full	vgg16	34k	0.44

TABLE III
KAGGLE DATABASE

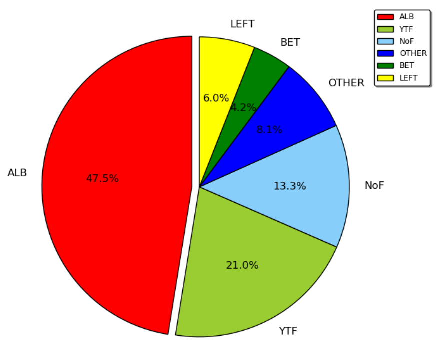
database	model	iteration	accuracy
full+rotating	alexnet	78k	0.9673
full+rotating	googlenet	80k	0.964
full+rotating	vgg16	20k	0.45

TABLE IV
KAGGLE DATABASE

database	model	iteration	accuracy
full+rotating+mosaic	alexnet	60k	0.9703
full+rotating+mosaic	googlenet	56k	0.954



(a) Training database



(b) Valing database

Fig. 1. Data distribution of Kaggle

The second method is to pick up some pictures to do image processing. We have made a mosaic 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 pictures with the pictures with mosaic. In this method, We modified the Caffe[2] to get the pooled image data to search the more significant area in classification. 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 finetune our pre-trained model and it performs better when testing.

Recently, several papers have brought some methods for fish detection and fish classifaction[3]. Spatial constraints are are introduced by dividing the image to regions and learning the feature region[4]. In our method, we use the image processing method to dividing the image to feature regions and background regions. And we used the mosaic images to lead our model to fit the feature regions and ignore the other

(a) the raw image

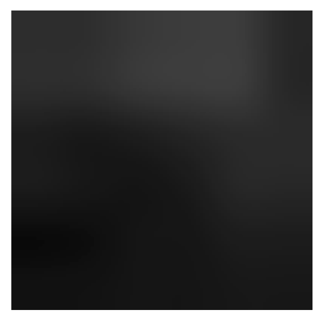
(b) image with mosaic

Fig. 2. We make a mosaic of the area where the fish are located because the mosaic 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 realize the system performance and improve the robust.

regions.[5]

The architecture of our model used in the experiment is the convolutional neural network(CNN). And CNN is usually composed of convolutional layer, pooling layer and fully connected layer. CNN has achieved remarkable performance in image classification recently. Compared with feature-designed extraction, it can explore more abstract and high-level information by deep neural network. So we can apply it in fish classification.

(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.

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 propertion. We can see the distribution of the dataset is imbalanced. The Table II illustrates the benchmarks. We can see the AlexNet and the GoogLeNet achived 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 method actually worked out and the accuracy has increased about two percent. Basing this, we have made a mosaic of the region of fish. The accuracy has also increased in the Table IV, which proves that our methods are effective.

III. Conclusion

We have used two methods to help our model to detect the fish and avoid the over-fitting. We have done the data augmentation and found that it can improve the accuracy and help avoid over-fitting. And comparing the raw image and the mosaic image can help our model to detect the fish. We have used both the two methods simultaneously 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.