Underwater Image Classification Using Deep Convolutional Neural Networks and Data Augmentation

Yifeng Xu
School of Marine Science and Technology
Northwestern Polytechnical University
Xi'an, Shaanxi, 710072
42805959@qq.com

Yang Zhang
Dalian Scientific Test and Control Technology Institute
Dalian, Liaoning, 116013
251822510@qq.com

Huigang Wang
School of Marine Science and Technology
Northwestern Polytechnical University
Xi'an, Shaanxi, 710072
wanghg74@nwpu.edu.cn

Xing Liu
School of Marine Science and Technology
Northwestern Polytechnical University
Xi'an, Shaanxi, 710072
xingliui86@nwpu.edu.cn

Abstract—**The classification accuracy of underwater image, which have the special image characteristic, is lower than the corresponding result of images in the air. A study was carried out to underwater image classification with deep convolutional neural networks and the classification ability was improved with two data augmentation methods. The experiments showed that the submarine image classification with the ILSVRC Championship GoogLenet model was still relatively low confidence. The classification probability can be improved by two augmentation methods. The first was optical transformation of raw data such as scale and aspect ratio augmentation and Color augmentation. The second was increasing the virtue data generated by Generative Adversarial Nets. The results of the study validated the effectiveness of two data augmentation methods. Especially, the generative adversarial nets approach gave a new path to increasing the train data.**

Index Terms—**Data Augmentation, Deep Leaning, Generative Adversarial Nets, Underwater Image Classification.**

I. INTRODUCTION

In the marine environment, optic image classification was an assistant method to sonar detection and had better resolution ability than sonar. The optic method was used to detect underwater close object precisely, which was widely used to such domains as wreck detection, disaster salvation, underwater treasure detection, underwater vehicle detection and so forth.

Traditional computer vision (CV) method to recognize object was called pattern recognition, which analyzed the features of images. In the underwater environment, however, images were fuzzy, which were seriously influenced by the underwater environment such as viewpoint variation, deformation, illumination conditions and background clutter. As a result, the classification and recognition precision by the traditional method was imperfect.

With Machine Learning sucessful application, deep convolutional neural network(CNN)[1] was becoming a hot research topic. CNN was a valid model that had recently been achieved considerable success and surpassed traditional CV methods in a variety of CV tasks such as the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Herein, we applied CNN to the special data: optic underwater images and discussed how to set the best net parameters according to the resolution features.

Deep CNN model needed to be trained on a huge number of labeled images to achieve satisfactory performance. The ImageNet 2012 database had 1000 classes over 15 million labeled high-resolution images. In the case, however, we only test the images related to the under marine environment. So we collected the underwater images from ImageNet2012 database[2] and generated more data based on these kinds of data. The former included 45 classes images related with the underwater environment from the ImageNet2012 database. For example, one of the selected classes was ID n04606251 and named "wreck".

After preparing the database, the fully open source calculation framework named CAFFE, afforded unobstructed access to deep architectures, was deployed. The Championship of ILSVRC 2014 competition, GoogLenet model[3], was modified to compute the confidence of the top 5 predicted categories. Since the particular of underwater images optic features, the confidence of classification had the space to be further improved. We used data augmentation, divided into two general classes, to achieve this purpose. The first was raw images optic transformation. The second was to generate the virtue data using the generate adversarial networks (GANs) .

The two main contributions of this paper are summarized as follows: (1)we collected and organized the database about marine environment; (2)we improved the classification accuracy using two data augmentation methods.

The rest of this paper is organized as follows. Related work about historically image classification critical models are proposed in Section Ⅱ.The methods such as convolutional layer, the framework of CNN model and GANs are showed in Section Ⅲ. Detailed experimental results are shown in Section Ⅳ. Several problems need to be further studied are put forwarded in Section Ⅴ. The paper concludes with Section Ⅵ.

978-1-5386-3142-3/17/$31.00 ©2017 IEEE

II. Related work

Starting with LeNet [1], convolutional neural network had become a standard structure in deep learning frameworks. Variants of this basic structure are prevalent in the image classification literature and have achieved the better results on the stand database such as MNIST, CIFAR and Imagenet. Increasing the number of layers and layer size[4] has become the trend. In addition, Dropout is used to address the problem of overfitting[5].

The model named AlexNet[6] which achieved the 2012 ImageNet competition champion, had 8 layers and the top 5 error was 16.4%. In the 2014 ImageNet competition, the model named VGG[7] won, which had 19 layers and achieved the 7.3% top 5 error. Later the 22 layers model named GoogLenet[3] achieved the top 5 error 6.7%. In the 2015, ResNet[8], 152 layers, the 3.57% top 5 error. Recently the densenet[9], 3.46% test top 1 error in Cifar10 database.

The research of image classification about undersea image was relatively less than that of optic image spread through the air. Automatic plankton image recognition[10] combines traditional invariant moment features and Fourier boundary descriptors with gray-scale morphological granulometries to form a feature vector capturing both shape and texture information of plankton images, achieved 95% classification accuracy on six plankton taxa taken from nearly 2,000 images.

III. The approach

A. convolutional layer

The convolution layer, the important element of deep learning framework, was to discover local conjunctions of features from the previous layer. At a convolution layer, the previous layer's feature maps are convolved with learnable kernels and put through the activation function to form the output feature map. Each output map combines convolutions with multiple input maps. In general, convolutional layer can be written as Eq. 1,

$$X_j^{\ell} = f\left(\sum_{i \in M_j} X_i^{\ell-1} * K_{ij}^{\ell} + b_j^{\ell} \right). \tag{1}$$

The parameters M_j, K_{ij}, b_j, f, and X_j respectively stand for input maps, a kernel, additive bias, activation function, and output maps at the l'th layer. Rectified Linear Unit (ReLU) ,the function $f(x) = \max(0, x)$, was the activation function.

B. the Pooling Layer

The max pooling layer, written as Eq.2, downsampled the input maps to decrease the computational work.

$$X_j^{\ell} = f(\beta_j^{\ell} down(X_j^{\ell-1}) + b_j^{\ell}). \tag{2}$$

The function *"down"* represents a sub-sampling function ,and maxes each distinct n-by-n block in the input image. β and b represents the multiplicative bias (weight) and an additive bias.

There were several ways to prevent overfitting such as L2/L1 regularization, max norm constrains and dropout. After comparing all kinds of regularization methods dropout, the most effective method, stand out.

C. the Framework of Convolutional Neural Networks

The ILSVRC14 competition champion named GoogLeNet[3] was modified for underwater objects classification. The framework of the model was in the following order: convolution(path size/stride: 7 × 7/2), max pooling(3 × 3/2), convolution(3 × 3/1), max pooling(3 × 3/2), inception, inception, max pooling (3 × 3/2), five inceptions, max pooling (3 × 3/2), inception, inception, avg pooling, dropout, linear, softmax. The inception module was the concatenation of a convolution(1 × 1), convolution(3 × 3), convolution(5 × 5) , and max pooling(3 × 3).

D. Generate Adversarial Network[11]

Generate adversarial mode includes two sub models: the generator (Abbreviated as G) and the discriminator (Abbreviated as D)[12]. G's aim is to generate virtue picture from scratch. To learn the G's distribution(pg) close to which of the original real data, random white noise variables pz(z) are defined as input, then represent a mapping to data space as G(z;θg), where G is a function represented by a multilayer perceptron with parameters θg. The sub model, D, also is a multilayer perceptron. D(x;θd) represents the probability that x came from the true data rather than the virtue generated data.

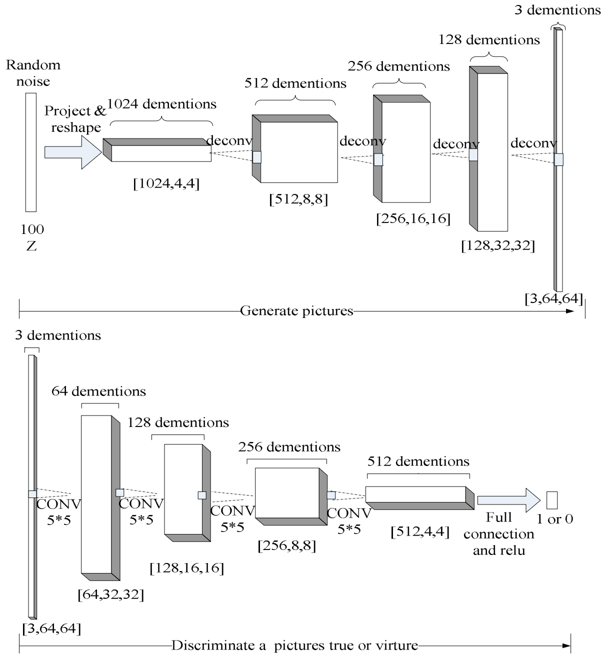


Fig. 1. the framework of generator and discriminator in GAN.

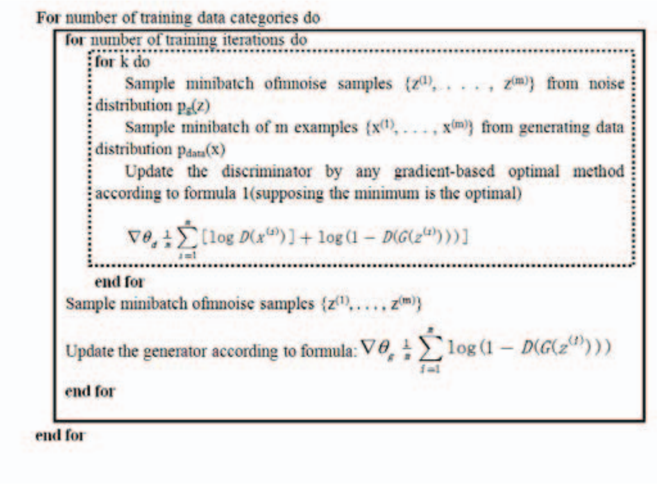
The model D outputs a single scalar 1 or 0 indicating real true data or virtually generated data. GANS trains D to maximize the probability of assigning the correct label to both true and virtue data. Simultaneously, it trains G to minimize the same probability. In other words, GANs follow a two-player minimax optimization process with value function V(D,G) generally written as formula $\min_G \max_D D(x) - D(G(z))$ [13]. In order to facilitate the calculation, the log function is added to D. Considering the input number for log function must be greater than zero, the specific formula of GAN has little change as follows Eq. 3

$$\max_{\theta_D} V(D,G) = \min_G \max_D \{ E_{x\sim pD}[\log D(x)] + E_{x\sim pG}[\log(1 - D(G(z)))] \}. \tag{3}$$

The framework of G and D was shown as figure 1.

E. Minibatch Stochastic Gradient Descent

Minibatch stochastic gradient descent[14] used in the training stage of generative adversarial nets to generate various categories of data. The number 2 was assigned to he hyper parameter k in our experiments , which defines the multiple of the D's number relative to the G's,.



```
For number of training data categories do
    for number of training iterations do
        for k do
            Sample minibatch ofmnoise samples $\{z^{(1)}, . . . , z^{(m)}\}$ from noise distribution $p_g(z)$
            Sample minibatch of m examples $\{x^{(1)}, . . . , x^{(m)}\}$ from generating data distribution $p_{data}(x)$
            Update the discriminator by any gradient-based optimal method according to formula 1(supposing the minimum is the optimal)
            $\nabla\theta_d \frac{1}{s}\sum_{i=1}^{s}[\log D(x^{(i)})] + \log(1 - D(G(z^{(i)})))]$
        end for
        Sample minibatch ofmnoise samples $\{z^{(1)}, . . . , z^{(m)}\}$
        Update the generator according to formula: $\nabla\theta_g \frac{1}{s}\sum_{i=1}^{s}\log(1 - D(G(z^{(i)})))$
    end for
end for
```

Fig. 2. Algorithm of minibatch stochastic gradient descent.

The pseudo-code in the dotted rectangle achieve the optimization of the discriminator. Those in the solid line rectangle complete synchronal optimization of the discriminator and generator. Because of many types of pictures, if all the varies of pictures are trained at the same time, the generated virtual picture will be chaotic and meaningless. So the process of generating pictures is done according to category. Finally, the pictures did not generated by the conventional optically transformation but by the GAN at the semantic layer.

IV. EXPERIMENTS AND RESULTS

To evaluate the accuracy of our approach for underwater images classification, 4 images showed in figure 3 were predicted by the CNN model described in part C of the approach section. The result of the top 5 results showed in table Ⅰ descending by the possibility.

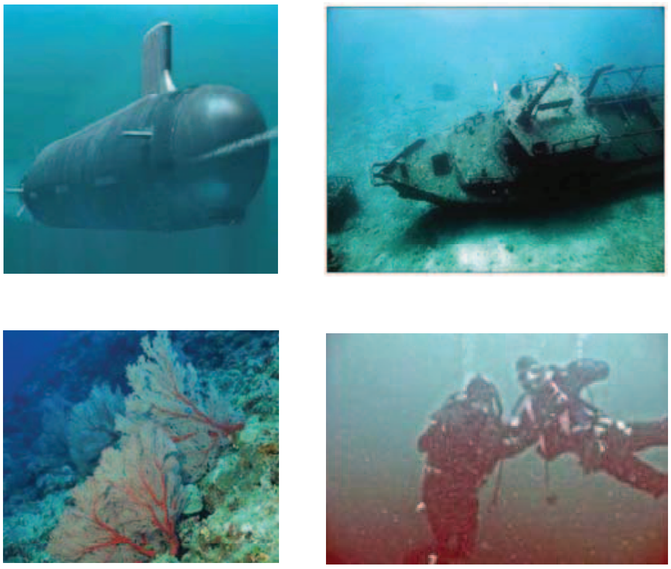


Fig. 3. samples of underwater images.

The Table Ⅰ told that the predicted probability of underwater images was low. The top n predicted score stand for the probability of predicting the category which's rank was n. The top1 predicted probability of picture 1 was merely 0.572. The possibility of predicting the correct category about picture 4 even just 0.014 and was top 4.

TABLE I. THE TOP 5 PREDICTING CLASS AND PROBABILITY

Images	**Top5 predicted and Ground Truth**					
Images	***Top1 predicted score***	***Top2 predicted score***	***Top3 predicted Score***	***Top4 predicted score***	***Top5 predicted score***	***Ground Truth***
Pic1 top, left	**submarine 0.572**	Great white shark 0.301	Dugong 0.075	Tiger shark 0.04	Airship 0.003	submarine
Pic2 top, right	**Wreck 0.858**	Tank 0.041	Haf track 0.032	Amphibian 0.02	Loggerhead 0.007	wreck
Pic3 bottom, left	**Coral reef 0.836**	lionfish 0.077	Spiny lobster 0.035	Sea anemone 0.011	Sea snake 0.01	Coral reef
Pic4 bottom, right	American lobster 0.445	King crab 0.39	Rock crab 0.022	**Scuba diver** 0.014	Spiny lobster 0.015	scubadiver

The main advantage of data augmentation was its great simplicity in increasing the train data.

TABLE II. THE PREDICTING PROBABILITY USING DATA AUGMENTATION

Images	The method of augmentation			
Images	***predicted class score***	***Raw data augmentation***	***Increase data by GAN***	***Two augmentations***
Pic1 top, left	submarine 0.572	submarine 0.582	submarine 0.603	submarine 0.620
Pic2 top, right	Wreck 0.858	Wreck 0.869	Wreck 0.898	Wreck 0.892
Pic3 bottom, left	Coral reef 0.836	Coral reef 0.847	Coral reef 0.876	Coral reef 0.885
Pic4 bottom, right	Scuba diver 0.014	Scuba diver 0.022	Scuba diver 0.025	Scuba diver 0.027

The first method was the conventional data augmentation method .It included scale and aspect ratio augmentation and color augmentation and used the photometric distortions from Andrew Howard[15]. This method doubled the 45 classes images related with the underwater environment. The result showed that it gave a better effection as the 3rd column in the tableⅡ.

The second data augmentation method tripled the raw data by GANs. One thousand and three hundreds pictures can be taken by every type of underwater images in the imagenet 2012 database. The method of GAN generated 3900 pictures aiming at every class. The method of GAN generates the similar pictures in the semantic level. A number of experiments has been taken to verify the theory that the above-mentioned methods will increase the probability. This results showed as the 4th column of Table Ⅱ. The results of two methods' joint utilization showed as the 5th column.

V. DISCUSSION

The model of optic image classification was still valid used on the underwater image classification. In the particular situation, however, the number of special kind of pictures was not enough, which resulted in the relatively low confidence of predicting class probability. By data augmentation, sufficient amouts of training data were generated to train an end to end network to perform the better results.

Because of the small underwater images sample sizes, a more significant result than data augematation may be achieved by increasing the new raw data collected from the real environment. It should be my future's work.

Another important recognition means is sonar under marine. But, the quantity of sonar images is much less than that of underwater images. The above GAN method can also be applied to increase the sonar images.

VI. CONCLUSIONS

This paper applied CNN, a general optic image classification model, to the underwater image classification. The results showed that the method of data augmentation could increase the predictive confidence. The traditional optical transform method, data augmentation, was effective. Moreover a new way of data augmentaion has paved by the GAN model.

ACKNOWLEDGMENT

This work was supported by the Natural Science Foundation of NSFC under the grant No. 61571369 and No. 61471299 and supported by the Fundamental Research Funds for the Central Universities No. 3102017jc06003 and supported by Science and Technology on Underwater Test and Control Laboratory under the grant No.YSOC261501.

REFERENCES

[1] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation Applied to Handwritten Zip Code Recognition," *Neural Computation*, vol. 1, no. 4. pp. 541–551, 1989.

[2] J. D. J. Deng, W. D. W. Dong, R. Socher, L.-J. L. L.-J. Li, K. L. K. Li, and L. F.-F. L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," *2009 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 2–9, 2009.

[3] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07–12–June, pp. 1–9, 2015.

[4] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014, vol. 8689 LNCS, no. PART 1, pp. 818–833.

[5] G. E. Dahl, T. N. Sainath, and G. E. Hinton, "Improving deep neural networks for LVCSR using rectified linear units and dropout," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2013, pp. 8609–8613.

[6] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015, vol. 07–12–June, pp. 3431–3440.

[7] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *Int. Conf. Learn. Represent.*, pp. 1–14, 2015.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Arxiv.Org*, vol. 7, no. 3, pp. 171–180, 2015.

[9] G. Huang, Z. Liu, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” *arXiv Prepr.*, pp. 1–12, 2016.

[10] X. Tang, W. Kenneth Stewart, L. Vincent, H. Huang, M. Marra, S. M. Gallager, and C. S. Davis, “Automatic Plankton Image Recognition,” *Artif. Intell. Rev.*, vol. 12, pp. 177–199, 1998.

[11] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, “Improved Techniques for Training GANs,” *Nips*, pp. 1–10, 2016.

[12] A. Radford, L. Metz, and S. Chintala, “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks,” *arXiv*, pp. 1–15, 2015.

[13] I. Goodfellow, J. Pouget-Abadie, and M. Mirza, “Generative Adversarial Networks,” *arXiv Prepr. arXiv …*, no. August 2016, pp. 1–9, 2014.

[14] P. Zhao and T. Zhang, “Accelerating Minibatch Stochastic Gradient Descent using Stratified Sampling,” *arXiv Prepr. arXiv1405.3080*, pp. 1–13, 2014.

[15] A. G. Howard, “Some Improvements on Deep Convolutional Neural Network Based Image Classification,” *arXiv Prepr. arXiv1312.5402*, pp. 1–6, 2013.