Underwater object Images Classification Based on Convolutional Neural Network

Keqing Zhu[1,2]
1:Institute of Acoustics,
Chinese Academy of Sciences
2:University of Chinese Academy of Sciences,
Beijing, China
e-mail:zhukeqing@mail.ioa.ac.cn,

Jie Tian*
Institute of Acoustics,
Chinese Academy of Sciences
Beijing, China
e-mail: tianjie@mail.ioa.ac.cn,

Haining Huang*
Institute of Acoustics,
Chinese Academy of Sciences
Beijing, China
e-mail: hhn@mail.ioa.ac.cn

***Abstract*—In order to solve the problem of underwater object images classification under the condition of insufficient training data, a novel underwater object images classification method based on Convolutional Neural Network(CNN) is proposed. Firstly, an advanced method of Markov random field-Grabcut algorithm is adopted to segment images into two regions: shadow and sea-bottom. Then, considering the character of the dataset, a CNN is constructed referring to Alexnet structure, consisting of two parts with different functions: convolutional part and classification part. At last, the CNN is trained to classify three different shapes of underwater objects(cylinder, truncated cone and sphere) utilizing the transfer learning approach. The method is applied to synthetic aperture sonar(SAS) datasets for validation. Comparing with Support Vector Machine(SVM) and CNN which only use trial dataset, the proposed method can achieve a better accuracy.**

Keywords-underwater object classification, CNN, transfer learning

I. INTRODUCTION

It is a really complex problem to achieve reliable accuracy in underwater object images classification [1]-[4]. Firstly, due to the high cost of underwater experiments, obtaining sufficient object images datasets for training a CNN [5]-[7] remains a challenge [8]. In addition, because of small size of mine-like object(MLO)[9]-[11], it is difficult to discriminate between the objects and rocks.

In this paper, an improved underwater object classification method based on CNN is presented to solve these problems by using transfer learning method[12]-[15]. A CNN is trained to classify three different shapes of underwater objects(cylinder, truncated cone and sphere). To test this method, synthetic aperture sonar (SAS) [16]-[18] datasets were used.

Generally, image classifying process mainly consists of two steps: feature extraction and classification. For feature extraction, an advanced method of Markov random field-Grabcut algorithm [19] is adopted to segment images into two regions: shadow and sea-bottom. For classification, considering the character of the dataset, a CNN is constructed referring to Alexnet structure[20]. The network is divided into two parts with different functions: convolutional part and classification part. In the training process, transfer learning is realized by sharing the weight of convolutional part. Namely, the convolutional part is trained by simulated dataset and then used to learn features from trial sonar images through convolutional operation.

Transfer learning is a new machine learning method that uses existing knowledge to solve different but related problems. It is intended to improve a predictive function for target task through transfering existing knowledge obtained from other task [14]. Compared with traditional machine learning method, it relaxes the requirements of the dataset for building a good classifer.

Convolutional neural network (CNN) is one of the most popular neural networks used in image recognition and computer vision systems today. It is a improved multilayer perceptrons designed to improve the accuracy of recognition or prediction based on shared-weights architecture and translation invariance characteristics. In recent years, it has been widely used in speech recognition [21], medical diagnosis [22], pedestrian detection [23], [24], natural language processing [25] and many other fields.

II. RELATED WORK

Over the years, a lot of underwater objects classification methods have been reported.

In [26], the underwater object image is segmented into three regions: sea-bottom, shadow, and background. They construct a feature space by extracting several features, such as shape descriptors and statistics descriptors. According to the characters of the feature space, a kind of statistical classifier is designed for classification task.

Different from [26], [27] use support vector machine (SVM), a existing model , to classify underwater object images after extracting a variety of features. Although above classification methods yielded excellent classification performances, these approaches based on traditional "shallow" architectures have a drawback, that is, when the performance plateau is hit, it cannot be improved by the addition of more training data.

In [1], a CNN is trained to accomplish feature extraction and classification of underwater objects. The CNN is trained with the underwater object images. The result has greatly improved compared with relevance vector machine(RVM). This work demonstrated the tremendous potential of using deep convolutional neural networks on sonar imagery for underwater object classification. However, the training of CNN requires a large amount of images and quite enormous computing cost.

III. FEATURE EXTRACTION

For feature extraction, we adopts mainly image segmentation. To be exact, the image is divided into two regions: shadow and sea-bottom. Image segmentation is a key pre-processing technology in sonar image classification system. Accurate segmentation can simplify or change the representation of an image into something that is more meaningful and easier to analyze. There are many methods for sonar images segmentation, such as active contours [28] and Markov random fields [29]. In this paper, an improved method of Markov random field—Grabcut algorithm is used to segment underwater object images.

In Grabcut, the input image can be expressed as $x=(x_1,...,x_n)$ .The process of image segmentation can be viewed as a process of solving the transparency value α of each pixel, where $\alpha=(\alpha_1,...,\alpha_n), \alpha_i \in \{0,1\}, i=1,...,n$, $\alpha_i=1$ for background and $\alpha_i=0$ for foreground. Each GMM is a mixture of K Gaussian models (usually $K=5$). Each pixel corresponds to a parameter k_n , $k_n \in \{1,...,K\}$. The Gibbs energy function can be expressed as:

$$E(\alpha,k,\theta,x)=U(\alpha,k,\theta,x)+V(\alpha,x) \quad (1)$$

Where $U(\alpha,k,\theta,x)$ is defined as data term and $V(\alpha,x)$ is defined as smoothness term. The value of parameter θ is determined by the minimum of data term $U(\alpha,k,\theta,x)$:

$$\theta=\arg\min_{\theta} U(\alpha,k,\theta,x) \quad (2)$$

The process of Grabcut algorithm comprised of four stages: a) initializing the background region T_B and the unknown region T_U by simply interacting with the user, b) establishing a GMM for each of the two regions, c) initializing the GMMs using the k-means algorithm, d) optimizing the parameters k_n , θ and α iteratively. The final segmentation result is obtained as the energy E gradually decreasing until converging to the minimum. The segmentation result of shipwrecks image is shown in Fig. 1.

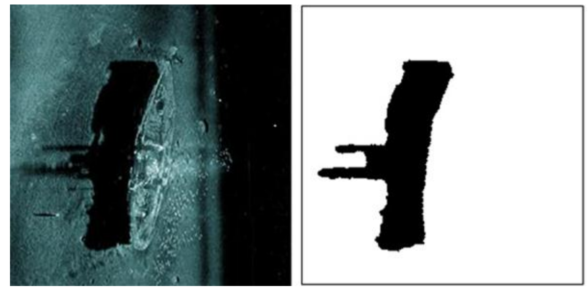

Figure 1. Shipwrecks SAS image (Left) and its segmentation result (Right)

IV. CLASSIFICATION

A. Convolutional Neural Network

A convolutional neural network is organized in two types layers with different topographic structures: convolutional and pooling. Convolutional layers accomplish learning image features by applying a convolution operation to the input, passing the result to the next layer. Pooling layers reduce the spatial dimensions of the input through a form of nonlinear down-sampling. In this paper, a convolutional neural network is constructed refers to the Alexnet structure, as shown in Fig. 2. The network consists of 8 layers: the first 5 layers are convolutional layers. The last 3 layers are full-connected layers, which are used to realize the high-level reasoning in the neural network. The first, second and fifth convolutional layer are followed by a pooling layer respectively.

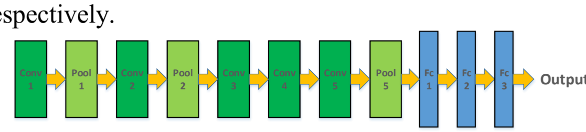

Figure 2. Structure of Alexnet

In forward propagation, CNN uses the convolutional layers and pooling layers to extract image features and reduce image dimension respectively. Full-connected layers multiply the input by a weight matrix and then add a bias vector. In the backward propagtion, a back propagation (BP)[30] algorithm is used to adjust the weight of neurons by calculating the gradient of the loss function. [31] gives the implementation of convolutional neural networks in detail and small snippets of MATLAB code as well.

Compared with other neural networks, convolutional neural networks have the following advantages: a)the excellent invarience of translation, rotation and scaling, b) learning features of the images automatically during the recognition process, c) simplified through weights sharing, increasing the efficiency of the parameters passing.

First, assuming the samples needs to be divided into C classes. So the cost function of the n-th sample is:

$$E^N=\frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{c}\left(t_k^n-y_k^n\right)^2 \quad (3)$$

Where t_k^n represents the expected value corresponding to the k-th neuron in the n-th sample. y_k^n represents the actual output value of the k-th neuron. When the network propagates backwards, the error describes the sensitivity of each neuron, which is defined as:

$$\frac{\partial E}{\partial b} = \frac{\partial E}{\partial u}\frac{\partial u}{\partial b} = \delta \tag{4}$$

The back propagation process can be described as follows:

$$\delta^l = (W^{l+1})^T \delta^{l+1} \bullet f'(u^l) \tag{5}$$

Note that the sensitivity of the output layer is different from the previous layers:

$$\delta^L = f'(u^L) \bullet (y^n - t^n) \tag{6}$$

Finally, gradient descent method is adopted to update weights according to the partial derivative of the cost function to the network parameters:

$$\frac{\partial E}{\partial W^l} = x^{l-1}(\delta^l)^T \tag{7}$$

The weights are also updated by:

$$\Delta W^l = -\eta \frac{\partial E}{\partial W^l} \tag{8}$$

B. *Transfer learning method*

Transfer learning aims to store knowledge gained and applying it to a different but related problem. More precisely, to cope with the problems such as the samples shortage, labels missing and low quality of the samples in domain A, we can transfer knowledges from domain B to improve the classification effect. Note that there exists some relationship, explicit or implicit, between the feature spaces of the two domains. Domain A is defined as target domain while domain B is defined as source domain. Due to the lack of sufficient training dataset problem in underwater object images classification task, it is very likely to bring overfitting problems. Therefore, transfer learning is an appropriate method for our imperfect dataset problem.

In this paper, a CNN is construted considering the character of the dataset. There are many weight parameters in the fully connected layer and numerous training data are needed. In order to fit the dataset better and make training process efficiency, the number of fully connected layers is changed from 3 to 2. Since we use grayscale images(1 channel) for training instead of RGB images(3 channels), the pooling layer behind the fifth convolutional layer is removed. The CNN structure constructed is shown in Figure 3.

The CNN is divided into two parts with different functions: convolutional part and classification part. The convolutional part contains 5 convolutional layers and 2 pooling layers while the classification part consists of 2 fully connected layers. Note that the last fully connected layer is the output layer as well, which can be viewed as a softmax classifier. In the training process, transfer learning is realized by sharing the weight of convolutional part. Namely, the convolutional part is trained by simulated dataset and then used to extract abstract feature from trial sonar images.

V. EXPERIMENT

A. *Dataset and Network Settings*

In order to test the improved underwater object classification method based CNN, trial SAS dataset is used as target domain, which is composed of three types of underwater object images, including 89 for cylindrical, 45 for truncated conical and 30 for spherical, one of them as shown in Fig. 4.

Simulated dataset is used as source domain, as shown in Fig. 5. There is a total of 930 images in simulated dataset, including 430 for cylinderical, 250 for trancated conical, and 250 spherical.

Before the network training, 30% of each types of samples are randomly selected to compose a testing dataset and the rest compose a training dataset. In order to meet the requirements of the CNN input, all images are processed to the size: 256×256. At the same time, Grabcut algorithm is used to segmentate all images. The network parameters setting of the CNN as shown in the TABLE I.

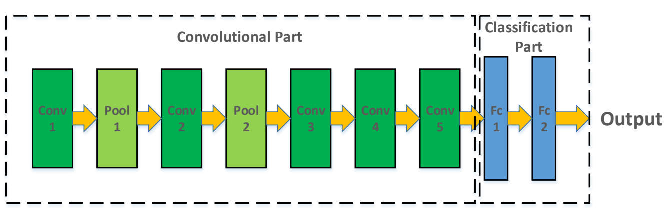

Figure 3. Structure of CNN constructed in this paper

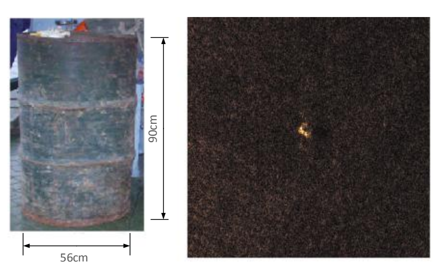

Figure 4. cylinderical object(Left) and its SAS image(Right)

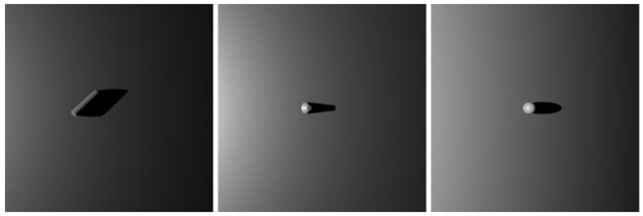

Figure 5. Simulated images of cylinderical(Left),trancated conical(Middle) and spherical object(Right)

TABLE I. THE PARAMETER SETTINGS OF CNN

Layer	Num of Convolutional kernel	Size of Convolutional kernel	Stride
Conv1	**64**	**11×11**	**4×4**
Conv2	**192**	**5×5**	**1×1**
Conv3	**192**	**3×3**	**1×1**
Conv4	**256**	**3×3**	**1×1**
Conv5	**256**	**3×3**	**1×1**

B. *Experiment A*

In Experiment A, the trial dataset was input into the network in Fig. 3 for training. Firstly, 938 batches were randomly selected from the training set of the trial sonar image dataset and the number of samples for each batch was 64. Then they were entered the above network for training. The accuracy of experiments was about 82.22%.

C. *Experiment B*

In Experiment B, simulated dataset was input into the same network's convolutional part and trial dataset into the classification part for training respectively. The accuracy of improved CNN was 91.11%, which is improved by 8.89%.

In addition, we plotted the learning curves of CNN and improved CNN in training process, as shown in Fig. 6. The learning curve of improved CNN converges faster, more stable than the previous method (CNN). Besides, compared to the traditional classifier SVM, the classification effect is more obviously improved, as shown in TABLE II.

TABLE II. THE ACCURACY OF SVM,CNN AND IMPROVED CNN

	SVM	CNN	Improved CNN
Cylinder	92.59%	77.78%	96.30%
Sphere	88.89%	77.78%	92.31%
Truncated cone	38.46%	100%	77.78%
Total	77.55%	82.22%	91.11%

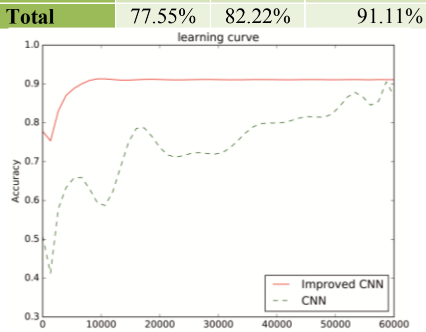

Figure 6. Learning curve of CNN and improved CNN

D. *Experiment C*

Taking into account of the factor that different training methods may affect the classification results, Experiment C was designed. The detailed training parameter settings of the three different training methods and their accuracy are shown in TABLE III. It was found that different training methods have few influence on classification accuracy, though the original method is slightly higher than another two methods.

This may be because the fully connected layer of the network's classification part acts as a "firewall" in the fine-tuning process, which makes the model's transfer ability enhanced. To obtain an exact theoretical explanation, further research is expected to be done.

TABLE III. 3 TYPES OF TRAINING METHOD PARAMETER SETTINGS AND ACCURACY

		Original training method	**Training method 1**	**Training method 2**
Dataset	**Convolutional part**	Simulated images	Simulated images	Simulated &Trial images
	Classification part	Trial images	Simulated &Trial images	Simulated &Trial images
Learning rate	**Convolutional part**	0.001	0.001	0.001
	Classification part	0.001	0.001	0.001
Number of iteration	**Convolutional part**	938	938	938
	Classification part	938	938	938
Accuracy	—	91.11%	90.48%	88.89%

VI. DISCUSSION

In this paper, a underwater object classification method approach is proposed to solve the problem of lack of sufficient training data by introducing transfer learning approach. As the preparation of classification, Grabcut algorithm is used to segment underwater object images. The design of our CNN refers to Alexnet architecture.

The experiment results proved that the proposed method achieves good performance in classification of underwater objects. At the same time, transfer learning method is applied to help train a good classifer when imperfect dataset is given. Namely, a CNN is pre-trained with a auxiliary simulated dataset. This may provide a reference for some target classification problems under the condition of insufficient training data.

ACKNOWLEDGMENT

The shipwrecks SAS image is downloaded from the website: https://stellwagen.noaa.gov/maritime/sass2010.html. The authors would like to thank NOAA/SBNMS and Applied Signal Technology.

REFERENCES

[1] D. P. Williams, "Underwater target classification in synthetic aperture sonar imagery using deep convolutional neural networks," 2016 23rd International Conference on Pattern Recognition (ICPR),

[2] T. W. Kim, S.-C. Yu, and J. Yuh, "Neural network-based underwater image classification for Autonomous Underwater Vehicles," IFAC Proceedings Volumes, vol. 41, no. 2, pp. 15991–15995, Jan. 2008.

[3] B. Marchand, N. Saito and H. Xiao, "Classification of Objects in Synthetic Aperture Sonar Images," 2007 IEEE/SP 14th Workshop on Statistical Signal Processing, Madison, WI, USA, 2007, pp. 433-437.

[4] Bello, Martin G., and Gerald J. Dobeck. "The use of texture measures in improving mine classification performance." OCEANS 2003. Proceedings. Vol. 2. IEEE, 2003.

[5] CUN Y L, BOSER B, DENKER J S, et al. Handwritten digit recognition with a back-propagation network[C].Neural Information Processing Systems, 1990: 396–404.

[6] Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep Learning[M]. MIT Press, 2016.

[7] Bengio Y. Learning deep architectures for AI[J]. Foundations and trends® in Machine Learning, 2009, 2(1): 1-127.

[8] J. McKay, I. Gerg, V. Monga and R. G. Raj, "What's mine is yours: Pretrained CNNs for limited training sonar ATR," OCEANS 2017 - Anchorage, Anchorage, AK, 2017, pp. 1-7.

[9] A. Sinai, A. Amar, and G. Gilboa, "Mine-Like Objects detection in Side-Scan Sonar images using a shadows-highlights geometrical features space," in OCEANS 2016 MTS/IEEE Monterey, 2016, pp. 1–6.

[10] C. Barngrover, A. Althoff, P. DeGuzman, and R. Kastner, "A Brain Computer Interface (BCI) for the Detection of Mine-Like Objects in Sidescan Sonar Imagery," IEEE Journal of Oceanic Engineering, vol. 41, no. 1, pp. 123–138, Jan. 2016.

[11] C. Barngrover, R. Kastner, and S. Belongie, "Semisynthetic Versus Real-World Sonar Training Data for the Classification of Mine-Like Objects," IEEE Journal of Oceanic Engineering, vol. 40, no. 1, pp. 48–56, Jan. 2015.

[12] Zhuang FZ, Luo P, He Q, Shi ZZ. Survey on transfer learning research. Ruan Jian Xue Bao/Journal of Software,2015,26(1):26-39 (in Chinese).

[13] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," Journal of Big Data, vol. 3, no. 1, p. 9, 2016.

[14] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," IEEE Trans. on Knowl. and Data Eng., vol. 22, no. 10, pp. 1345–1359, 2010.

[15] H. C. Shin et al., "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," in IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1285-1298, May 2016.

[16] M. P. Hayes and P. T. Gough, "Synthetic Aperture Sonar: A Review of Current Status," IEEE Journal of Oceanic Engineering, vol. 34, no. 3, pp. 207–224, Jul. 2009.

[17] P. T. Gough and D. W. Hawkins, "A short history of synthetic aperture sonar," in Geoscience and Remote Sensing Symposium Proceedings, 1998. IGARSS '98. 1998 IEEE International, 1998, vol. 2, pp. 618–620 vol.2.

[18] P. T. Gough, M. P. Hayes, "Synthetic aperture sonar: The past the present and the future", Proc. Inst. Acoust., vol. 26, pp. 81-87, Sep. 2004.

[19] Rother, Carsten, Vladimir Kolmogorov, and Andrew Blake. "Grabcut: Interactive foreground extraction using iterated graph cuts." ACM transactions on graphics (TOG). Vol. 23. No. 3. ACM, 2004.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.

[21] ZHANG Qingqing, LIU Yong, PAN Jielin, et al. Continuous speech recognition by convolutional neural networks[J]. Chinese Journal of Engineering, 2015, 37(9):1212–1217.

[22] Hoochang S, Roth H R, Gao M, et al. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning[J]. IEEE Transactions on Medical Imaging, 2016, 35(5):1285-1298.

[23] Hou, Ya-Li, et al. "Multispectral pedestrian detection based on deep convolutional neural networks." Signal Processing, Communications and Computing (ICSPCC), 2017 IEEE International Conference on. IEEE, 2017.

[24] Dominguez-Sanchez A, Cazorla M, Orts-Escolano S. Pedestrian Movement Direction Recognition Using Convolutional Neural Networks[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, PP(99):1-9.

[25] P. Lauren, G. Qu and P. Watta, "Convolutional neural network for clinical narrative categorization," 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, 2017, pp. 2001-2008.

[26] Fandos R, Zoubir A M. Optimal Feature Set for Automatic Detection and Classification of Underwater Objects in SAS Images[J]. IEEE Journal of Selected Topics in Signal Processing, 2011, 5(3):454-468.

[27] Couillard M, Fawcett J A, Myers V, et al. Support vector machines for classification of underwater targets in sidescan sonar imagery[J]. 2008.

[28] S. Reed, Y. Petillot and J. Bell, "An automatic approach to the detection and extraction of mine features in sidescan sonar," in *IEEE Journal of Oceanic Engineering*, vol. 28, no. 1, pp. 90-105, Jan 2003.

[29] M. Mignotte, C. Collet, P. Perez and P. Bouthemy, "Sonar image segmentation using an unsupervised hierarchical MRF model," in IEEE Transactions on Image Processing, vol. 9, no. 7, pp. 1216-1231, Jul 2000.

[30] Werbos, Paul J. "Applications of advances in nonlinear sensitivity analysis." System modeling and optimization. Springer, Berlin, Heidelberg, 1982. 762-770.

[31] BOUVRIE J. Notes on convolutional neural networks[C].Neural Nets, 2006