

Underwater target recognition method based on convolution residual network

Yuechao Chen*, Shuanping Du, HengHeng Quan, and Bin Zhou

Science and Technology on Sonar Laboratory, Hangzhou Applied Acoustics Research Institute, Hangzhou, 310023, China

Abstract. The underwater target radiated noises usually have characteristics of low signal to noise ratio, complex signal components and so on. Therefore the recognition is a difficult task and powerful recognition method must be applied to obtain good results. In this paper, a recognition method for underwater target radiated noise time-frequency image based on convolutional neural network with residual units is proposed. The principles and characteristics of the convolutional residual network are analyzed and three basic convolutional residual units are put forward. Then three convolutional residual network models with very deep structure are established based on basic convolutional residual units and some normal convolution layers. The number of the hidden layers is 50, 100 and 150 respectively and softmax algorithm is used as the top classifier. The wavelet transform is adopted to generate time-frequency images of the underwater target radiated noises with frequency band of 10~200Hz, thus ensuring the accuracy of local structure of the image, then the above three models can be used to recognize the images. The experimental data of two types of targets were processed. The results are as follows. As the number of training time increases, the training loss shows a convergence trend and the recognition accuracy of test data gradually increases to more than 90%. In addition, the top-level output has obvious separability. The final recognition accuracies of the three convolutional residual networks are all over 93% and higher than that of normal convolutional neural network with 5 layers. As the number of layers increases, the recognition accuracy of the convolutional residual network increases to a certain extent, illustrating the increase of layer number can improve the processing effect. The analysis results show that the convolution residual network can extract features with separability through deep structure and achieve effective underwater target recognition.

1 Introduction

Underwater target recognition is the key technology to improve the intelligentization of underwater acoustic equipment. It has been a hot topic in the field of underwater acoustic signal processing for a long time. The traditional underwater target recognition is mainly realized by extracting separability features and designing classifier to recognize the separability features [1,2]. Because of the complexity of the marine environment and the specificity of the underwater acoustic channel, the features that can reflect the nature of the target are often the results of the combination of various original features according to the degree of contribution and correlation. Therefore, separability features extracting is a difficult problem for underwater target recognition so far.

In recent years, artificial intelligence, big data and other technologies have been developing rapidly. And the deep learning is a hot research direction in the field of artificial intelligence. The concept was first put forward by G. Hinton in 2006 [3]. Compared with traditional machine learning methods, the number of hidden layers in deep learning model is greatly increased, which greatly improves the ability of complex

computing. Since put forward, the deep learning has attracted wide attention. Not only the theoretical algorithm has been constantly updated, but also the actual models have been used in many situations such as speech signal process and image recognition [4-10]. In the field of underwater acoustic signal processing, some scholars have tried to apply deep learning algorithms such as convolutional neural network (CNN) and deep belief network (DBN) to underwater target recognition and get some effect [11-14]. Compared with the usual speech signal and image, underwater acoustic signal has lower signal-to-noise ratio and more complex components, so the deep learning model should be optimized according to the characteristics of underwater acoustic signal.

Many studies show that increasing the depth of neural network can effectively improve the computing ability. The AlexNet, VGGNet, and Google Inception Net successively won classification project champions in ILSVRC (ImageNet Large Scale Visual Recognition Challenge). The number of network layers of the above models is 8, 19 and 22 respectively, and their recognition ability is improved in turn. However, the deepening of the number of networks has also brought problems such as difficulties in training [15]. In 2015, K. He from

* Corresponding author: chenyc_715@163.com

Microsoft Research Institute proposed residual neural network and the application of model with 152 layers again refreshed the record rate of ILSVRC identification project. The structure of residual neural network can speed up the training of neural network with deep layers and improve the training accuracy of the model. Now it has become a research hotspot in deep learning [16-18].

In this paper, convolution residual network models with different layers were constructed and then used for time-frequency image recognition of underwater target radiated noises. The research results can provide a basis for the application of deep convolutional neural network in underwater acoustic signal recognition.

2 Methods

2.1 Convolutional neural network

Convolutional neural network is a kind of artificial neural network. Its early model is called as neurocognitive machine. It is a biophysical model inspired by the neural mechanism of vision system. Convolutional neural network can be regarded as a special multilayer perceptron designed for two-dimensional object recognition. It has the characteristics of local connection and weight sharing. Traditional recognition algorithms need to extract data features before recognition. But convolutional neural network can directly input pictures into network and automatically extract features. The model has a high degree of invariability to the deformation of the image.

Fig.1 shows a general structure diagram of convolutional neural network. The hidden layers of convolutional neural network mainly consist of convolution layer and pooling layer. The convolution layer is used to extract features by translating a number of convolution kernels on original image. Each feature is a feature map.

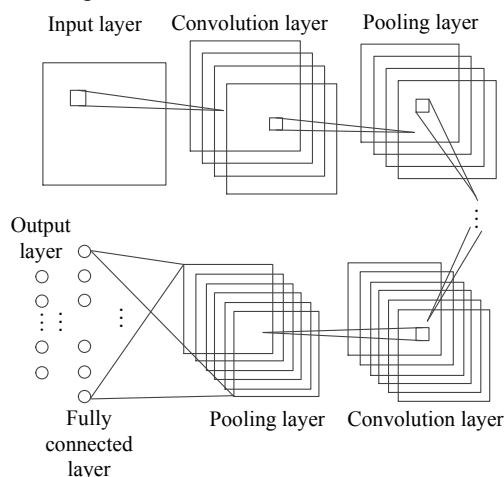


Fig. 1. General structure of convolutional neural network.

The pooling layer decreases the parameters to be learned and then reduces the complexity of the network. The most common pooling methods are maximum pooling and average pooling.

2.2 Convolution residual network

In general, depth plays an important role in convolutional neural network. However, when information is transferred between traditional convolution layer or fully connected layer, there will exist the problem of information loss. So the deeper the network is, the more difficult it is to train. For the traditional convolutional neural network, the accuracy will be increased gradually and reached saturation in the process of increasing depth. Then the accuracy rate will decline with the continuing increase of depth. This is not caused by the overfitting problem, because the error not only increases on the test set but also increases in the training set.

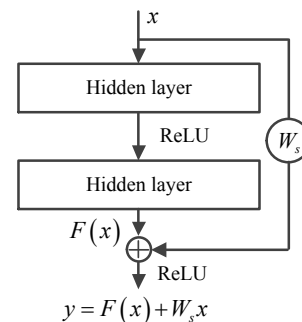


Fig. 2. Basic schematic diagram of residual module.

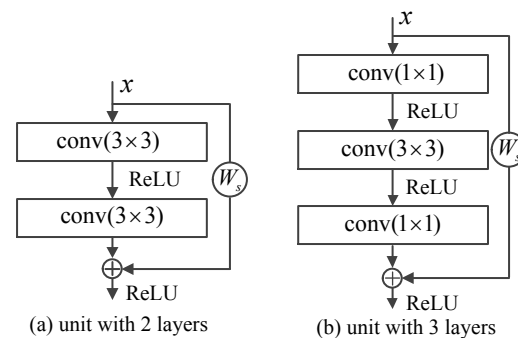


Fig. 3. Two residual learning units.

Convolution residual network solves this problem in a certain extent through inputting information to output directly. Then the integrity of the information is protected. The network only needs to learn the difference between input and output, and the learning goal is simplified. Fig.2 shows the basic principle frame diagram. Set the input and expected output of a neural network module are x and y respectively, then

$$y = F(x) + W_s x \quad (1)$$

where $F(x)$ is the residual learning function. There are two computing layers in the diagram. That is, a convolution operation is performed on the x first, which is activated based on ReLU function, and then a convolution operation is performed on the activation result. $W_s x$ is the direct input of x and W_s is a weight matrix set to match the dimensions of $F(x)$. If the dimensions of x and $F(x)$ are the same, then W_s is 1. At this time, the objective learning function of the neural

network model is $F(x) = y - W_s x$, that is, the learning target is the residual between the output and the input. Fig.3 shows two commonly used residual learning units. Fig.3(a) shows the 2-layer structure residual learning module, which contains 3×3 convolution products with the same number of output channels. Fig.3(b) shows the 3-layer structure residual learning module. Both layers outside are 1×1 convolutions and the middle layer is 3×3 convolutions.

3 Experiment and result

3.1 Data pre-processing

The raw data to be identified are two kinds of experimental signals of underwater target radiated noises. The sampling frequency is 5000Hz. Fig.4 shows some random selected segments. It can be seen that the signal to noise ratio is very low and the composition of the signal is complex. The components that can reflect intrinsic properties of targets can be extracted from underwater target radiated noise in frequency domain. In general, these separability characteristics mainly exist in the low frequency spectrum of 10-200Hz range. Therefore, the original signal is preprocessed by time-frequency transform to obtain time-frequency images that can highlight signal separability as input data for deep learning model.

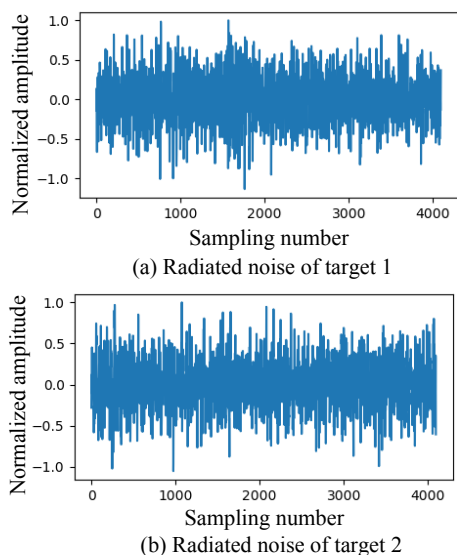


Fig. 4. Underwater target radiated noises.

At present, the commonly used time-frequency transform methods include short time Fourier transform, Wigner-Ville distribution, wavelet transform and so on, in which wavelet transform is the improvement of short time Fourier transform. By setting $f(t)$ as an arbitrary function in the $L^2(R)$ space, its continuous wavelet transform can be expressed as

$$WT_f(a, \tau) = \langle f(t), \psi_{a,\tau}(t) \rangle = \frac{1}{\sqrt{a}} \int_R f(t) \overline{\psi\left(\frac{t-\tau}{a}\right)} dt \quad (2)$$

where $\psi_{a,\tau}(t)$ is wavelet basis function, a is stretching factor, τ is translation factor. The wavelet transform maps the signal from one-dimensional time domain to two-dimensional time-frequency domain. By adjusting a and τ , the wavelet coefficients with multi-resolution can be obtained, thereby realizing the localized time-frequency analysis of the signal. At present, the wavelet transform has been applied in the ship radiated noise feature extraction and achieved relatively good results [19,20].

The time-frequency images of the target signals were generated based on the wavelet transform. The wavelet basis is Complex Morlet (CMOR). It is a complex sinusoidal modulation Gaussian wave. The resolution of finally acquired time-frequency image is 256×256 and the frequency is 10~200Hz, so that the time-frequency image can highlight the separability as much as possible. The number of samples is 8500. Fig.5 shows the generated time-frequency image results.

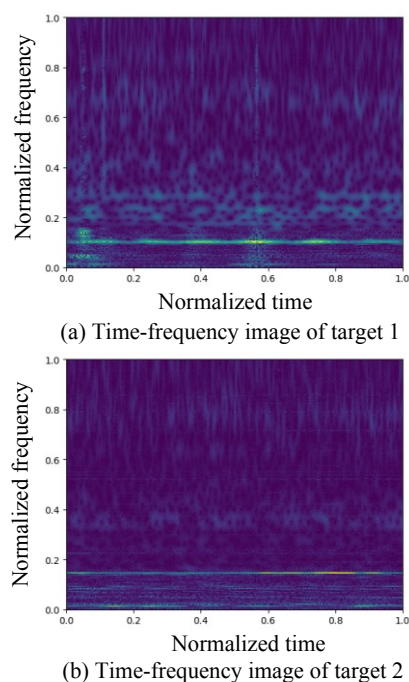


Fig. 5. Time-frequency image of target signal.

3.2 Underwater Target Recognition

Convolution residual network models were established to identify the time-frequency images of target signals, including 50-layer model, 101-layer model, and 152-layer model. All three models were first processed by ordinary convolution layer with kernel size of 7×7 and step length of 2. After a maximum pooling layer with size of 3×3 and step length of 2, the output features with size of 64×64 can be obtained. The convolution layer and the pooling layer are respectively represented as C_{layer} and P_{layer} . Then the models were designed using four different residual modules, which are represented as $R_{module1}$, $R_{module2}$, $R_{module3}$ and $R_{module4}$ respectively. The residual module was generated based on the three-layer structure residual learning units shown in Fig.3(b). By

combining different numbers of residual learning units, convolution residual network model with different number of hidden layers can be obtained. The size of the output characteristic results is 8×8 . Table 1 shows the specific convolution residual network model parameters. Finally, the softmax classifier was used to classify the feature results. The softmax is a multi-class classifier commonly used in deep learning and machine learning [21].

Table 1. Specific structure of convolution residual network models.

	Model with 50 layer	Model with 101 layer	Model with 152 layer
C_{layer}	$7 \times 7, 64, \text{step length } 2$	$7 \times 7, 64, \text{step length } 2$	$7 \times 7, 64, \text{step length } 2$
P_{layer}	$3 \times 3, \text{step length } 2$	$3 \times 3, \text{step length } 2$	$3 \times 3, \text{step length } 2$
$R_{module1}$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
$R_{module2}$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
$R_{module3}$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
$R_{module4}$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$

The above three kinds of convolution residual network models were used to process the time-frequency images of the target signal, in which the number of training samples is 6800 and the number of test samples is 1700. The size of each training data batch as well as that of test data batch is 32. The training time and testing time is 500 and 50 respectively.

Fig.6(a) shows the training error when training the 50-layer model. It can be seen that as the number of training time increases, the error gradually decreases and the result shows a convergence trend. Based on the test data, the accuracy rate of the model in the training process is tested. The change curve is shown in Fig. 6(b). It can be seen that the accuracy rate gradually increases to more than 90%.

The test data is processed based on the trained 50-layer model. Fig.7 shows the top-level output. It can be seen that most of the two types of targets have been separated.

The above three convolution residual network models with different layers and ordinary convolutional neural network with 5 layers were used to identify the test data. The ordinary convolutional neural network

contained three convolution layers and two fully connected layers.

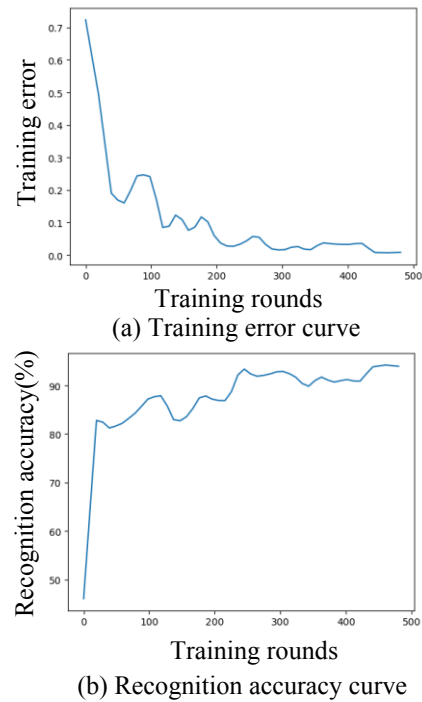


Fig. 6. Analysis of training process of convolution residual network.

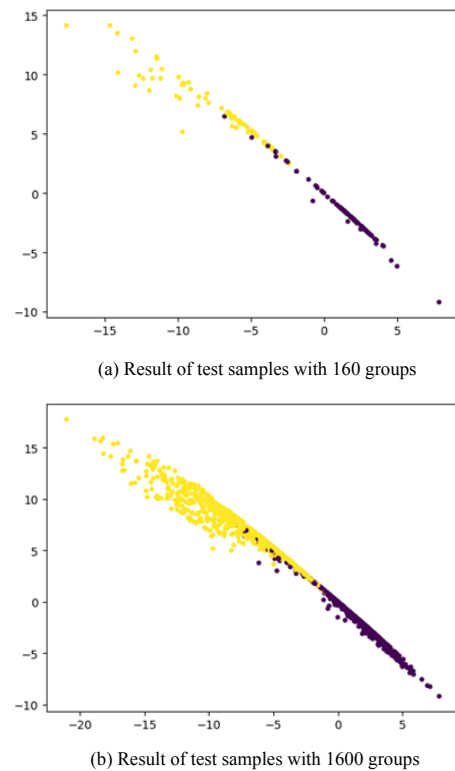


Fig. 7. Top-level output of convolution residual network.

Table 2 shows the recognition accuracy. It can be seen that the three convolution residual network models have higher recognition accuracy than that of the ordinary convolutional neural networks. For the three

convolution residual network models, the accuracy of the 101-layer model and the 152-layer model is higher than that of the 50-layer model.

This shows that the increase in the number of layers can improve the complex computing capacity of the model to a certain extent, thereby improving the target recognition accuracy.

Table 2. Recognition accuracy of each method.

Models	Convolution residual network			convolutional neural network
	50-layer model	101-layer model	152-layer model	
Accuracy/%	93.13	95.20	95.85	88.50

4 Conclusions

The underwater target radiation noise identification method based on convolution residual network was studied in this paper. Based on the analysis of the convolution residual network principle, three kinds of convolution residual networks with different layers were established. Two types of underwater target experimental signal time-frequency images were generated based on wavelet transform and then were identified. The results show that the recognition accuracy of convolution residual network models is higher than that of ordinary convolutional neural network, and the increase of layer number of convolutional residual network can improve the recognition accuracy to a certain extent. This shows that the convolution residual network with deep structure can realize the effective recognition of underwater targets.

This study was funded by the National Natural Science Foundation of China [Grant No.61701450] and the Young Elite Scientists Sponsorship Program by CAST [Grant No.2017QNRC001]. We also wish to express our gratitude to the Science and Technology on Sonar Laboratory in Hangzhou Applied Acoustics Research Institute for financially supporting this work.

References

1. S. Wang, X. Zeng, *Appl. Acoust.*, **78**, 68-76(2014)
2. D. Li, M. R. Azimi-Sadjadi, M. Robinson, *IEEE. T. Neural. Networ.*, **15**, 189-194(2004)
3. G. Hinton, R. R. Salakhutdinov, *Science*, **313**, 504-507(2006)
4. D. L. K. Yamins, J. J. Dicarlo, *Nat. Neurosci.*, **19**, 356-365(2016)
5. J. Schmidhuber, *Neural. Networks*, **61**, 85-117(2015)
6. B. Alipanahi, A. DeLong, M. T. Weirauch, B. J. Frey, *Nat. Biotechnol.*, **33**, 831-838(2015)
7. I. Lenz, H. Lee, A. Saxena, *Int. J. Robot. Res.*, **34**, 705-724(2014)

8. D. Ciresan, U. Meier, J. Masci, J. Schmidhuber, *Neural. Networks*, **32**, 333-338(2012)
9. G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, et al, *Med. Image. Anal.*, **42**, 60-88(2017)
10. M. Gong, J. Zhao, J. Liu, Q. Miao, L. Jiao, *IEEE. T. Neur. Net. Lear*, **27**, 125-138(2017)
11. G. Hu, K. Wang, Y. Peng, M. Qiu, J. Shi, L. Liu, *Comput. Intel. Neurosc.*, **3**, 1-10(2018)
12. C. Li, Q. Zhou, X. Han, J. Yin, M. Shao, *J. Acoust. Soc. Am.*, **142**, 2732-2732(2017)
13. Y. Chen, X Xu, *IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, Xiamen, China, (2017)
14. S. Kamal, S. K. Mohammed, P. R. S. Pillai, M. H. Supriya, *Ocean Electronics (SYMPOL)*, 48-54, 2013
15. F. Zhou, L. Jin, J. Dong, *Chinese Journal of Computers*, **40**, 1-23(2017)
16. K. He, X. Zhang, S. Ren, J. Sun, *2016 IEEE Conference on Computer Vision and Pattern Recognition*, 770-778, Las Vegas, USA, (2016)
17. H. Wen, J. Shi, W. Chen, Z. Liu, *Sci Rep*, **8**, 1-17 (2018)
18. S. Pouyanfar, S. C. Chen, M. L. Shyu, *IEEE International Conference on Multimedia and Expo*, 373-378, Hong Kong, China, (2017)
19. H. Ou, J. S. A. Iii, V. L. Syrmos, *J. Acoust. Soc. Am.*, **125**, 2578(2009)
20. X. Li, Y. Peng, L. Lin, Z. Lin, *Acta Acustica*, **1**, 63-67(2004)
21. P. K. Raj, D. S. Sheeja, *Jour of Adv Research in Dynamical & Control Systems*, **1**, 139-148(2017)