An Underwater Acoustic Target Recognition Method Based on Transfer Learning

Xiaozhuo Yang [1, 2, a], Huapeng Yu [1, b,*], Hanmin Sheng [2, c], Wenlong Zeng [2, d], Qingyuan He [1, e], and Junyang Tu [1, 2, f]

[1] National Innovation Institute of Defense Technology, Chinese Academy of Military Science, China;

[2] School of Automation Engineering, University of Electronic Science and Technology of China, China.

[a] 202222060218@std.uestc.edu.cn, [b, *] hpyu_qtxy@163.com, [c] shenghanmin@hotmail.com,

[d] 202122060438@std.uestc.edu.cn, [e] 15285550382@163.com, [f] tthales@foxmail.com

Abstract. Underwater acoustic target recognition(UATR)is a challenging task Due to the high cost of sampling data, it is difficult to build a large-scale dataset.In this study,a new method based on transfer learning with a VGG16 model(Transfer-VGG16) is developed in which three-dimensional(3-D)data is used as the input.Use the dataset obtained in real scenarios to validate the proposed method.Experiments have shown that Transfer-VGG16 model works well when the observation data is scarce.

Keywords: Underwater acoustic target recognition(UATR); Transfer learning; Deep learning.

1. Introduction

Owing to the intricate nature of underwater acoustic environments and the ongoing advancements in underwater acoustic countermeasures,The constraints of expert experience systems relying on conventional spectrum analysis techniques in UATR are becoming increasingly apparent.In the past few years, deep learning has achieved significant advancements in domains such as image recognition and speech recognition, and underwater acoustic recognition is no exception to this trend[3].Xiaoping Song et al.[2]propose an approach using Convolutional Neural Network (CNN) and Long Short-Term Memory(LSTM) to mine intrinsic features, and Wasserstein Generative Adversarial Network-Gradient Penalty(WGAN-GP) is used to enhance the training samples data.

Nonetheless, the efficacy of automatic feature extraction in deep learning is considerably impacted by the quantity and caliber of training samples, typically mandating a substantial amount of high-quality training samples as a foundation[3].Underwater acoustic signal has special characteristics, mainly in:Due to the significant influence of the marine environment on the sample signal, the signal quality is poor, and it is difficult to obtain a large amount of data.So traditional manual feature extraction methods are suitable for UATR,such as mel-spectrum,MFCC,GFCC[4,5,6],short-time fourier transform(STFT)[7],and so on.Biao Wang et al. [8]proposed an underwater target recognition method using an AMNet network,which uses STFT amplitude spectrum as network inputs,introduces attention structure,and couples with the multibranch structures to highlight important information hidden in the time-frequency maps.The drawback of traditional manual extraction methods is that they may lose fixed features.To address this shortfall, there has been a growing trend in the adoption of feature extraction techniques that rely on the fusion of multiple features[9,10].

Transfer learning has found extensive application in diverse domains of pattern recognition, including image processing and signal analysis.That can improve the recognition rate of models by reducing interference caused by a lack of prior knowledge.It has evolved into several subfields of research, encompassing domain adaptation,model-transfer,and domain generalization[11,12,13,14,22].Ke. et al.[15]proposed a transfer learning with resonance-based sparsity signal decomposition(RSSD)method to classify ship-radiated noise.However,pre-training necessitates vast amounts of data.In order to solve the problems,transfer learning with a VGG16 model method(Transfer-VGG16) is proposed.

In this work, we evaluated the effectiveness of our proposed method on Shipsear dataste[19](11 class) and Our private dataset(5 class) obtained in real scenarios.The contributions of this work are as follows.

1. We pre-trained VGG16 model on the source domain dataset,then fine-tuning the model with target domain dataset.Compared with other model like CLDNN,Resnet and VGG16,the recognition accuracy has been improved.
2. Mel-spectrum and its first-order and second-order incremental matrix be elongated to increase data smoothness and solved the problem of inconsistent data dimensions amidst the source and target domains.

The second section describes the data processing methods and Transfer-VGG16 models used for UATR.The third section introduces various comparative experiments.Finally,this paper concluded in the fourth section.

2. Proposed Method

This section offers a comprehensive overview of the concept and implementation of the method utilized in this study, which employs the model-based transfer learning approach.

2.1 Data Processing

Research has shown that mel-spectrum features exhibit good performance in UATR[4].We begin to use a invariable length window function to frame the time-domain signal,and then use the Fourier transform(FFT) of the framed part to obtain a frame FFT spectrums.This window is shifted throughout the entire time interval to provide a set of frame FFT spectrum and then splicing all of the frame FFT spectrum together.The result is the STFT amplitude spectrum,As shown in Eq.1.Where N means the length of the window function, x(k) means time-domain signal.

$$STFT_x(m,n) = \sum_{\{k=0\}}^{\{N-1\}} x(k)w(k-m)e\text{^}\{-\frac{j2\pi nk}{N}\}$$

The STFT amplitude spectrum undergoes filtration using Mel filter banks, followed by conversion to a mel-spectrum utilizing a logarithmic scale and integration operation, as depicted in Eq.2,where f means frequency.

$$Mel(f) = 2590 \times \log_{10}(1 + \frac{f}{700})$$

We introduce the first-order incremental matrix and second-order incremental matrix of mel-spectrum, which increases data smoothness and is suitable for models processing three-dimensional data,as shown in Fig.1

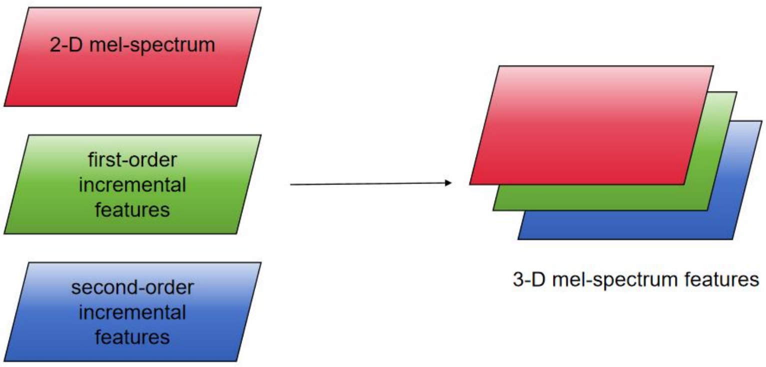

Fig. 1 Reshaping the 2-D mel-spectrogram into the 3-D data

2.2 Transfer Learning with a VGG16 model

In this study,the objective of transfer learning is to reduce the interference of the lack of prior knowledge on the model and achieve accurate recognition of the target.

Model-transfer mainly involves pre-training and fine-tuning,which limiting model flexibility and avoiding overfitting.It means pre-training all of the layers using a large amount of dataset from the source domain,then transferring the model parameters,and finally fine-tuning the layers except pre-trained layers using a small amount of target domain dataset.

The VGG model has well performance in image recognition and speech recognition[8,20,21,22].VGG16 is a heavyweight model composed of multiple convolutional and pooling layers stacked alternately,and finally classified using fully connected layers,and the model structure is shown in Figure.2 and Figure.3.

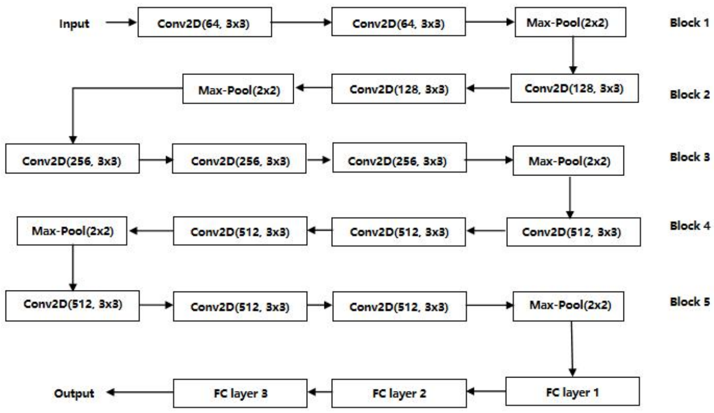

Fig. 2 Structure of Transfer-VGG16 module

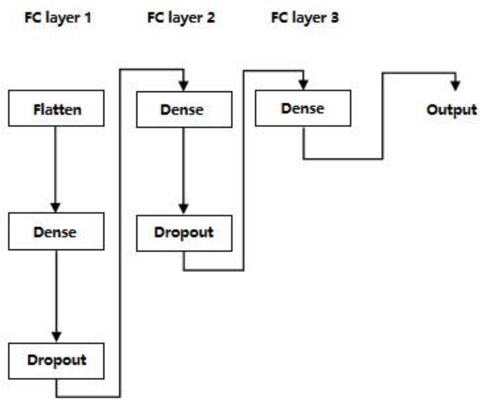

Fig. 3 Structure of Fully Connected Layer,all layers use ReLU activation,except the last layer that uses a Sigmoid

In this work,The VGG16 model pre-trained on the ImageNet[18]dataset,and then use the underwater acoustic dataset to fine-tuning.ImageNet is a large-scale dataset with 1.2 million images.This method saving time and resources required for pre-training.The pre-trained model has superior performance in feature extraction of 3-D data,this solves the difficulty of obtaining a large amount of data for pre-training.

As shown in Figure.1,through data processing,we reshaped mel-spectrogram into 3-D data,solving the problem of inconsistent data dimensions.Due to the lack of overlapping feature and label spaces between the source and target domains,we did not transfer all of the model parameters of the pre-trained model.

Research[17]has shown that the higher the similarity between the feature spaces of the source and target domains, the more layers' parameters available for transferring.And for models used for classification tasks,the parameters of deep convolutional layers have a significant impact on the output results.We transferred the model parameters of block 1,block 2 and block 3 of the pre-trained model and fine-tuning the layers except transferred layers to avoiding overfitting.

3. Experiments and Results

3.1 Experiments Setup and Dataset Describe

In order to validate the method,we used both the public dataset ShipsEar and our private dataset(Miniset) to demonstrate the performance of transfer-VGG16 and compared it with CLDNN,Resnet50 models and VGG16 model.We use 70% of the samples for training, 30% for prediction.We configured the batch size to be 64 and set the learning rate to 0.1. Additionally,we chose the cross entropy function as the loss function.The TensorFlow version used for this implementation is 2.7.0.

The ShipsEar dataset was captured during the autumn of 2012 and summer of 2013, on the Atlantic coastline in the northwestern region of Spain, more specifically in the vicinity of the port of Bigo in Ria de Vigo.The recording experiment used an SR-1 digital hybrid hydrophone,which has a flat response in the frequency range of 1 Hz to 28 kHz,a sampling rate of 52734 Hz, and resolution of 24 bits.The ShipsEar dataset comprises a collection of 90 audio files of varying durations, encompassing recordings of 11 distinct types of ship noise as well as one type of underwater background noise.

Miniset recording a total of 3 types of ship noise and 2 types of underwater background noise:Class A,Class C and Class D are three types of ship noise,Class B is the type of underwater background noise and Class E is the type of underwater biological noise.

For evaluating the proposed method on ShipsEar dataset,we allocated 50% and 5% of the data for training and testing respectively. All the samples were resampled at a frequency of 22.5 kHz.Additionally, to augment the sample count,we segmented the samples into 1-second intervals.When training our model with only 5% of the data,we allocated 820 and 621 samples for the test and train sets respectively.Conversely,when training with 50% of the data,we used 8450 and 621 samples for the train and test sets respectively.In the miniset evaluation, we utilized the entire dataset,which comprised of 244 and 90 samples for the train and test sets respectively.

3.2 Results and Discussion

For deep neural networks,the parameters of deep convolutional layers have a greater impact on the output results,while the parameters of shallow convolutional layers have a smaller impact on the output results.When the source and target domains have non-overlapping feature and label spaces, transferring too few parameters can lead to inadequate performance. Conversely, transferring an excessive amount of parameters can have a detrimental effect on task performance.The relationship between the accuracy in Miniset for the number of transferred blocks' parameters in Figure 4.Where N means the number of transferred blocks.When N = 3,it means transferring parameters from block 1 to block 3.The performance of the model is the worst when transferring parameters of all the blocks.and the improvement in accuracy is most significant when transferring parameters of block 1 to block 3.

Figure 5 illustrates the accuracy curves of the four models on the Miniset.Transfer-VGG16 demonstrates faster convergence compared to the other models,accompanied by higher accuracy.Transfer-VGG16 reaches a state of convergence and balance around the 40th iteration.Notably,Transfer-VGG16 outperforms Resnet50 by 6.6%.On the other hand,Resnet50 achieves convergence by the 50th epoch,which is comparatively slower than Transfer-VGG16.

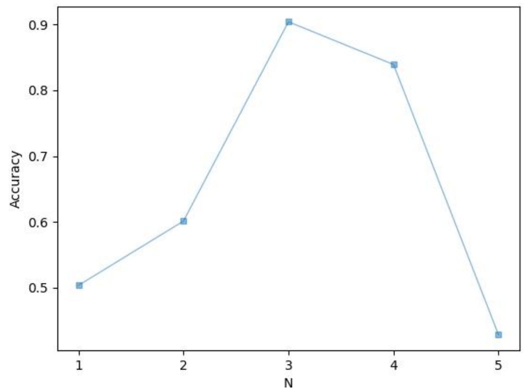

Fig.4 Block N at which model is pre-trained

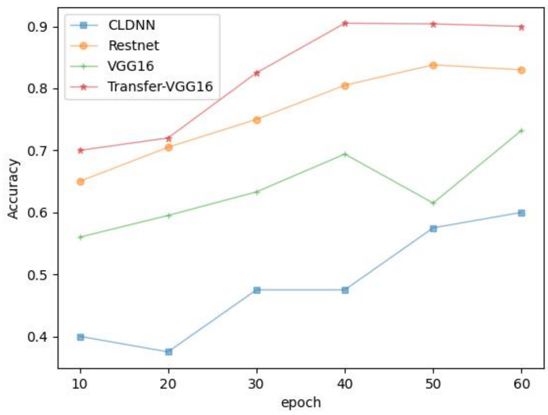

Fig. 5 Models accuracy curve

The purpose of this experiment is to demonstrate the performance of the proposed method in underwater recognition tasks,especially when data is scarce.The recognition results of each model are shown in Table 1.It depicts four models in terms of accuracy,precision,recall and F1-score.In ShipsEAR with 50% samples,we compared CLDNN, Resnet,and VGG16 with transfer-VGG16.The results indicate that increasing the depth of the model appropriately has a certain improvement in accuracy.Then we compared VGG16 with our transfer-VGG16.The result proved the effectiveness of our proposed method.

As shown in Table 2,in ShipsEAR with 5% samples,our proposed transfer learning method has a significant improvement in recognition accuracy.Compared to VGG16,Transfer-VGG16 has a 9.6% improvement in accuracy.

The performance of the model on Miniset is shown in Table 3,and it shows that Transfer-VGG16 performs the best.Compared with other models,Transfer-VGG16 still performs well when the data is scarce.It can be found that Transfer-VGG16 outperforms other models in each performance evaluation.This serves as validation for the efficacy of our proposed method.

Table 1. Performance of Models in ShipsEar with 50% Samples

Model	Precision	Recall	F1-score	Accuracy
CLDNN	81.2%	83.4%	81.1%	81.9%
Resnet50	87.8%	88.2%	88.0%	88.2%
VGG16	86.1%	86.8%	86.4%	87.0%
Transfer-VGG16	90.2%	87.8%	88.9%	89.9%

Table 2. Performance of Models in ShipsEar with 5% Samples

Model	Precision	Recall	F1-score	Accuracy
CLDNN	43.0%	44.7%	43.4%	47.5%
Resnet50	72.3%	72.7%	72%	72.5%
VGG16	77.6%	77.8%	77.5%	77.9%

Transfer-VGG16	87.5%	86.6%	86.9%	87.5%

Table 3. Performance of Models in Miniset

Model	Precision	Recall	F1-score	Accuracy
VGG16	71.3%	74.4%	73.2%	73.5%
CLDNN	72.8%	74.5%	72.6%	73.9%
Resnet50	83.6%	82.3%	84.9%	83.8%
Transfer-VGG16	89.5%	91.7%	90.1%	90.4%

As shown in Figure.6, we present the classification results of transfer-VGG16 in the Miniset.We provide the accuracy of the model for each type of data through the confusion matrix.It can be seen that transfer-VGG16 has the excellent performance for Class B(underwater background noise) and Class D(ship noise),both reaching over 90%.The accuracy of Class A is only 72%, indicating serious noise interference in the data.

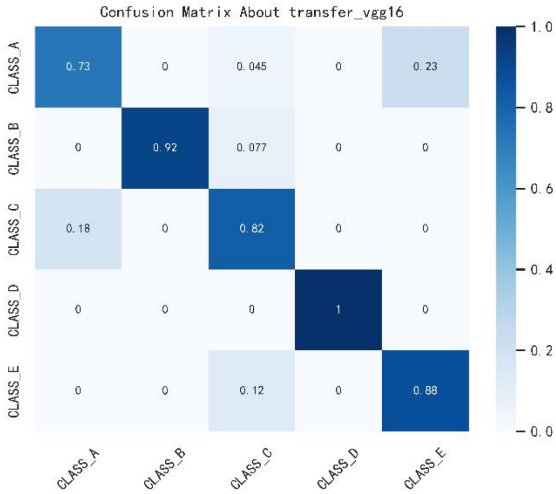

Fig. 6 Confusion matrix in the Miniset with Transfer-VGG16

4. Conclusion

In this paper,we proposed a new UATR method based on transfer learning with a VGG16 model.We smoothed the data by introducing first-order and second-order incremental matrix of mel-spectrum, and solved the problem of inconsistent feature dimensions.Experiments have shown that compared with other model,Transfer-VGG16 works well when the observation data is scarce.

References

[1] Valdenegro-Toro, M. Best practices in convolutional networks for forward looking sonar image recognition. 2017.Fangfang. Research on power load forecasting based on Improved BP neural network. Harbin Institute of Technology, 2011.

[2] Xiaoping, S., C.Jinsheng,and G. Yuan.A New Deep Learning Method for Underwater Target Recognition Based on One-Dimensional Time-Domain Signals, in 2021 OES China Ocean Acoustics (COA). 2021, IEEE. p. 1048-1051.

[3] Li, D., et al. Data augmentation method for underwater acoustic target recognition based on underwater acoustic channel modeling and transfer learning. Applied Acoustics, 2023. 208.

[4] C. Liu, F. Hong, H. Feng, and M. Hu. Underwater acoustic target recognition based on dual attention networks and multiresolution convolutional neural networks, in Proc. OCEANS, Sep. 2021,pp. 1 – 5.

[5] Wang W, Li S, Yang J, et al. Feature extraction of underwater target in auditory sensation area based on MFCC. IEEE/OES China Ocean Acoustics (COA) 2016;2016:1 – 6.

[6] M. Jeevan, A. Dhingra, M. Hanmandlu, and B. Panigrahi. Robust speaker verification using GFCC based I-vectors, in Proc. Int. Conf. Signal, Netw., Comput., Syst. Cham, Switzerland: Springer, 2017,pp. 85–91.

[7] Q. Zhang, L. Da, Y. Zhang, and Y. Hu. Integrated neural networks based on feature fusion for underwater target recognition, Appl. Acoust.,vol. 182, Nov. 2021, Art. no. 108261.

[8] Wang, B., et al.,An Underwater Acoustic Target Recognition Method Based on AMNet, in IEEE Geoscience and Remote Sensing Letters. 2023. p. 1-5.

[9] F. Liu, T. Shen, Z. Luo, D. Zhao, and S. Guo. Underwater target recognition using convolutional recurrent neural networks with 3-D mel-spectrogram and data augmentation, Appl. Acoust., vol.178, Jul. 2021,Art. no.107989.

[10] F. Hong, C. Liu, L. Guo, F. Chen, and H. Feng.Underwater acoustic target recognition with a residual network and the optimized feature extraction method, Appl. Sci., vol. 11, no. 4, p. 1442, Feb. 2021.

[11] Shao L,Zhu F,Li X.Transfer Learning for Visual Categorization:A Survey.IEEE Trans Neural Networks Learn Syst 2015;26(5):1019–34.

[12] Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, et al. A Comprehensive Survey on Transfer Learning. Proc IEEE 2021;109(1):43–76.

[13] Mittal S, Srivastava S, Jayanth JP. A Survey of Deep Learning Techniques for Underwater Image Classification. IEEE Transactions on Neural Networks and Learning Systems, 2022, Early Access Article.

[14] Nguyen CT, Van Huynh N, Chu NH, Saputra YM, Hoang DT, Nguyen DN, et al.Transfer Learning for Wireless Networks: A Comprehensive Survey. Proc IEEE 2022;110(8):1073–115.

[15] X. Ke, F. Yuan, and E. Cheng, Underwater acoustic target recognition based on supervised feature-separation algorithm, Sensors, vol. 18,no. 12, p. 4318, Dec. 2018.

[16] Wang W, Li S, Yang J, et al. Feature extraction of underwater target in auditory sensation area based on MFCC. IEEE/OES China Ocean Acoustics (COA) 2016;2016:1–6.

[17] Yosinski J, Clune J, Bengio Y, and Lipson H.How transferable are features in deep neural networks. In Advances in Neural Information Processing Systems 27 (NIPS’14), NIPS Foundation, 2014.

[18] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li,and Li Fei-Fei. Imagenet: A large scale hierarchical image database. In Computer Vision and Pattern Recognition,2009. CVPR 2009. IEEE Conference on, pages 248–255.IEEE, 2009. 6

[19] D. Santos-Domínguez, S. Torres-Guijarro, A. Cardenal-López, and A. Pena-Gimenez, “ShipsEar: An underwater vessel noise database,” Appl. Acoust., vol. 113, pp. 64–69, Dec. 2016.

[20] Chinta, B. and M. M, EEG-dependent automatic speech recognition using deep residual encoder based VGG net CNN. Computer Speech & Language, 2023. 79.

[21] Thepade, S.D., et al., Face presentation attack identification optimization with adjusting convolution blocks in VGG networks. Intelligent Systems with Applications, 2022. 16

[22] Paymode, A.S. and V.B. Malode, Transfer Learning for Multi-Crop Leaf Disease Image Classification using Convolutional Neural Network VGG. Artificial Intelligence in Agriculture, 2022. 6: p. 23-33.