GAN-based Sample Expansion for Underwater Acoustic Signal

Hongbin Yang¹, Han Gu²*, Jinyong Yin¹ and Jian Yang¹

¹ Jiangsu Automation Research Institute, Lianyungang, Jiangsu, 222061, China
² College Of Marine Life And Fisheries, Jiangsu Ocean University, Lianyungang, Jiangsu, 222005, China
*Corresponding author’s E-mail: gu_han963@163.com

Abstract. A deep learning model has a large number of free parameters, which need to be effectively trained on a large number of samples to calculate the depth parameters. However, many special applications like underwater acoustic signal recognition cannot provide enough dataset to guarantee high performance. In addition, the original dataset adopts some formats, such as audio, which makes it difficult to capture features. To overcome these challenges, we propose a novel framework. Firstly, based on the evaluation of spectrum method, our framework selects the appropriate preprocessing method. Then it modifies GAN to generate samples, and establishes an independent classification network to ensure the quality of samples. Finally, our framework applies the existing classification network to evaluate performance and selects the best one for the LOFAR spectrum. The experimental results demonstrate that the proposed method can generate high-quality LOFAR spectrum and improve the prediction accuracy of the classification model significantly.

1. Introduction

Deep learning models have testified that the state-of-the-art performance can be achieved by using high-dimensional data in such fields as computer vision [1], speech recognition [2] and natural language processing [3]. Through the training on labeled data, these models optimize millions of parameters and achieve high performance. However, with the rapid development of computational power and algorithms to meet the needs of deep learning, training samples become the bottleneck of the classification and recognition gradually, especially the domain-specific datasets. Underwater acoustic signal (UAS) [4-5] is this type of data.

UAS is usually supplied to let professionals judge the type of target, and the original sample is in audio format. After audio frequency spectrum conversion, it may contain disturbing information (such as noise). LOFAR spectrum [6] is commonly used for UAS processing and target tracking. Therefore, this article attempts to apply LOFAR spectrum instead of audio spectrum for classifying objects when utilizing deep learning methods. Due to the limitation of confidentiality or acquisition difficulty, it is impossible to provide enough UAS samples for neural networks, so that the classification model cannot reach satisfactory results.

In order to break through the bottleneck of deep learning, data augmentation [7-8] has become one of the most popular techniques. Data augmentation amplifies the variation of target data by adding geometric transformations, color space transformations and kernel filters. However, when we train the target model on a small batch data set, the improvement of data augmentation is limited because the
data growth cannot generate real invisible targets, and the modification is small. To overcome the limitation of data augmentation, several means [9-11] have been proposed by utilizing generative adversarial networks (GAN) [12]. Through the adversarial learning of generator and discriminator, a fake data consistent with the real data distribution can be obtained in the end. Gabbay [13] proposed a StyleGan-based generator inversion, which can repair images or converting still images to animations. Anitha advanced LR-GAN [14] to generate image backgrounds and foregrounds recursively and individually, and acquired a splendid result.

This article focuses on the augmentation of UAS in LOFAR spectrum, and proposes a novel framework for the object classification. It also focuses on a performance assessment to quantify the quality of the generated samples. Then, our framework trained a neural network to check each sample. The experimental results demonstrated that the proposed method can generate the spectrum of UAS efficiently and improve the prediction accuracy of the classification model significantly. It can be included that the LOFAR spectrum image generated based on GAN is of great significance and can be applied to spectrum-related applications such as training deep learning classifiers and training seafarers. This paper is organized as follows: in section 2, the framework we will experiment to classify UAS are introduced. It includes UAS preprocessing and the architecture of our modified GAN. In section 3, we describe our analysis to the experimental results and verify the quality of the generated data. Our conclusions and future works are presented in section 4.

2. Materials and methods
At the beginning of our framework, the original samples are converted into audio spectrum samples, LOFAR spectrum samples, demon spectrum samples and other formats. Then, we utilize CNN to evaluate these types of samples respectively, and confirm that LOFAR spectrum can achieve higher classification accuracy. In the next process, our framework generates samples through a modified GAN model, and applies an independent classification network to evaluate the quality of the generated samples. The structure of whole framework is revealed in figure 1.

2.1 Data Preprocessing
LOFAR spectrum is a spectrogram formed by processing the signal from both the time domain and the frequency domain through STFT (short-Time Fourier Transform). It can detect the underwater target radiated signal and extract the signal under the condition of low signal-to-noise ratio. However, the signal will be interfered by background noise in the process of underwater complex environment transmission, which affects the extraction of target line spectrum, so it is necessary to denoise LOFAR spectrum.

In the process of preprocessing, the first step is to transform the original UAS into LOFAR signal by a STFT operation. According to the data energy interpolation, a characteristic frequency band is selected and the result is converted into a corresponding gray sample.

Then, we utilize a processing method based on background equalization, morphological de-noising and median filtering. By inhibit the random fluctuation of the underwater background noise, restore the line spectrum which has lower amplitude or being covered by background noise in LOFAR spectrum, retain the edge of the line spectrum. In the last step, we obtain gray-scale samples at regular intervals, and adjust the scale of LOFAR spectrum to 64 * 64 for model training.
2.2 Network Structure

Our framework presents a simple modification to InfoGAN, which can solve the problems of unconstrained, incontrollable and difficult to interpret the noise signal of the original GAN model. From the principle displayed in Eq 1, it can be seen that our framework also contains 2 parts: a generator network $G$ and a discriminator model $D$. In the process, the $G$ and $D$ will play a “max-min game” continuously until gaining the optimal result. Differently, the input vector $z$ is divided into two parts: an interpretable hidden variable $c$ and an incompressible noise $z'$. $I(c; G(z,c))$ is an information-theoretic regularization, which represents the mutual information between interpretable hidden variable $c$ and generator distribution $G(z, c)$. However, in the calculation of $I(c; G(z,c))$, the posterior $P(c|x)$ is not clear. Therefore, in the specific optimization process, we adopt the idea of variational inference and introduces the variational distribution $Q(c|x)$ to approximate $P(c|x)$. It is based on the alternative iteration of the lower bound of the optimal mutual information to achieve the final solution, so the objective function of InfoGAN converts into Eq 2. And $L_I(G,Q)$ is the variational lower bound of the mutual information $I(c; G(z,c))$. Figure 2 shows the main frame of our information adversarial network.

\[
\min_G \max_D V_I(D, G) = V(D, G) - \lambda I(c; G(z,c)) \tag{1}
\]

\[
\min_G \max_Q V_{\text{InfoGAN}}(D, G, Q) = V(D, G) - \lambda L_I(G,Q) \tag{2}
\]
The main part of the discriminator consists of one input layer, three convolution layers and two full connection layers. Considering the characteristics of UAS, our framework cancels the design of pool layer to make the generated samples closer to the real situation. The image channel number is 32, which is composed of one original data channel number and 31 types of training data. The first volume has 64 filters, the second volume has 128 filters and the third volume has 256 filters. According to the comprehensive evaluation, we set the size of each filter to 4*4 and the sliding step size to 2. After three-level convolution, it changes the dimension of the data into two dimensions to facilitate input as a fully connected layer, where the first dimension corresponds to the batch number and the second convolution data value corresponding to the sample. After the body part, it uses the sigmoid activation function to map the input value between 0 and 1 to get scalar.

The principal part of the generator is similar to the discriminator but in the opposite direction. It is composed of input layer, two fully connected layers and three deconvolution layers. The random noise vector \( z \) is selected by the input data layer data, and its vector length is obtained by the data layer length 62, plus the category 31. In the deconvolution part, the first deconvolution layer has 256 filters, the second deconvolution layer has 128 filters and the third one has 64 filters. Finally, the output of the generator network processing is a batch of image data. You can observe the change of the generated data by changing a certain dimension of \( c \), so as to further optimize the generated data set according to the interpretable information.

Classification network \( Q \) shares all convolution layers with \( D \), and only a full connection layer is added at the end to output \( Q(c|x) \). The new modified network structure is shown in Figure 3, where (a) corresponds to the discriminator, (b) corresponds to the generator, and (c) corresponds to the classifier.

3. Experiments and results
This paper has collected several public datasets of UAS to train our modified GAN. They are selected from the ShipsEar [15] and the San Francisco National Park Association. Our framework preprocesses
the original UAS and marks them as 31 types, each spectrum contains 1000 samples. Then we train these samples with modified GAN, and finally generate high-quality samples. After training, our method can not only generate all kinds of samples synthetically by utilizing G, but also generate specific spectral samples for specific types by changing one dimension of latent code. Figure 4 shows some of the original and generated samples, where (a) and (b) correspond to a single category and (c) and (d) correspond to comprehensive categories. As shown in Figure 4, the data generated by our network contains common feature lines, and in some cases, all feature lines are clearly displayed and have more prominent feature lines than the original training samples.

![Figure 4. Original and Generated samples](image)

To verify the quality of the generated samples, we utilize VGGish [16], which catches the best performance in ALEXNET [17] VGG, LENET [18] and VGGish, as the classification model. The samples generated by our framework need to be produced in large quantities for classifier training, and 3000 samples are produced for each category in this paper. Finally, it can be proved that the generated samples are of high quality through the generalization ability of the classification model. It mainly carried out the following 3 experiments:

First of all, 80% of the original samples are used for training, the remaining 20% are used for training, and 31 kinds of samples are classified and identified by the model. Then, an 80% original sample is trained as the training data and 20% generated data as the validation data of the classification model. Finally, the original sample and the generated sample are mixed into a comprehensive dataset, 80% of which is used for training and the remaining 20% for verification. The experimental results show that the performance of the model trained by mixed data is 15% higher than that of the same classification system trained only by the original data set. The overall results are presented in Table 1.

| Training set                  | Validation set        | Recognition rate |
|-------------------------------|-----------------------|------------------|
| original sample (80%)         | original sample (20%) | 77.3%            |
| original sample (80%)         | generated sample (20%)| 85.4%            |
| mixed sample (80%)            | mixed sample (20%)    | 92.9%            |

Compared with the original samples, most of the noises are preprocessed and the diversity of training samples is greatly increased. Therefore, applying the classification model trained by the generated samples and the original mixed samples, we can achieve a higher recognition rate on the verification dataset. It expresses that the model has high generalization capability and avoids the overfitting problem caused by a small amount of data to some extent.

4. Conclusion
In this paper, a novel UAS recognition framework is proposed. Firstly, preprocessing is carried out by denoising and spectrum conversion, and then LOFAR spectrum samples are generated by the modified
GAN. Finally, the classification accuracy is verified by using the tested and evaluated classification network. The experimental results show that our method can solve the problems of low sample quality and insufficient data set in UAS and other fields. By mixing the generated high-quality LOFAR spectrum with the original dataset to train the classification model, the best recognition rate can be obtained, and the accuracy and stability of UAS classification and recognition task execution can be improved. In the future work, we intend to use large-scale images to carry out comparative research with other Gan derived models, and at the same time carry out the research of target classification through small samples.

References
[1] Real, E. , Aggarwal, A. , Huang, Y. , & Le, Q. V. . (2018). Regularized evolution for image classifier architecture search.
[2] Zeyer, A. , Irie, K. , Schlüter, Ralf, & Ney, H. . (2018). Improved training of end-to-end attention models for speech recognition.
[3] Vaswani, A. , Shazeer, N. , Parmar, N. , Uszkoreit, J. , Jones, L. , Gomez, A. N. , Kaiser, L. , and Polosukhin, I. 2017. Attention is all you need. In Advances in Neural Information Processing Systems 30.
[4] Nishida, S. , Iwase, R. , Kawaguchi, K. , Matsuo, I. , & Akamatsu, T. . (2017). Real time detection and localization system for underwater acoustic signal with cable observatories in the west Pacific Ocean. Techno-ocean. IEEE.
[5] Li, Y. , & Chen, Z. . (2017). Entropy based underwater acoustic signal detection. International Bhurban Conference on Applied Sciences & Technology. IEEE.
[6] Yang, P. , Yuan, B. , & Zhou, S. . (2010). A Line Spectrum Estimation Method of Underwater Target Radiated Noise Base on the 1. International Conference on Innovative Computing & Communication & Asia-pacific Conference on Information Technology & Ocean Engineering. IEEE Computer Society.
[7] Fong, R. , & Vedaldi, A. . (2019). Occlusions for effective data augmentation in image classification.
[8] Baek, F. , Park, S. , & Kim, H. . (2019). Data augmentation using adversarial training for construction-equipment classification.
[9] Choi, E. , Biswal, S. , Malin, B. , Duke, J. , Stewart, W. F. , & Sun, J. . (2017). Generating multi-label discrete electronic health records using generative adversarial networks.
[10] Huang, S. W. , Lin, C. T. , Chen, S. P. , Wu, Y. Y. , & Lai, S. H. . (2018). AugGAN: Cross Domain Adaptation with Gan-based Data Augmentation. ECCV 2018: European Conference on Computer Vision.
[11] Gulrajani, I. , Ahmed, F. , Arjovsky, M. , Dumoulin, V. , & Courville, A. . (2017). Improved training of wasserstein gans.
[12] Goodfellow I J. , Pouget-Abadie J. , Mirza M. , et al, “Generative adversarial nets” International Conference on Neural Information Processing Systems. MIT Press, 2672-2680, (2014).
[13] Gabbay, A. , & Hoshen, Y. . (2019). Style generator inversion for image enhancement and animation.
[14] Yang, J. , Kannan, A. , Batra, D. , & Parikh, D. . (2017). Lr-gan: layered recursive generative adversarial networks for image generation.
[15] Santos-Dominguez, D. , Torres-Guijarro, S. , Cardenal-Lopez, A. , & Pena-Gimenez, A. . (2016). Shipsear: an underwater vessel noise database. Applied Acoustics, 113(dec.), 64-69.
[16] Hershey, S. , Chaudhuri, S. , Ellis, D. P. W. , Gemmeke, J. F. , Jansen, A. , & Moore, R. C. , et al. (2016). Cnn architectures for large-scale audio classification.
[17] Krizhevsky, A. , Sutskever, I. , & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. International Conference on Neural Information Processing Systems, (Vol.60, pp.1097–1105). Nevada, USA, Curran Associates Inc.
[18] Lecun, Y. (1998). Lenet-5, convolutional neural networks. Retrieved from: yann.lecun.com.