Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 3995331, 10 pages
https://doi.org/10.1155/2022/3995331

Research Article

Supervised Contrastive Learning-Based Modulation Classification of Underwater Acoustic Communication

Daqing Gao, Wenhui Hua, Wei Su, Zehong Xu, and Keyu Chen

Information and Communication Engineering, Xiamen University, Xiamen, China

Correspondence should be addressed to Wei Su; suweixiamen@xmu.edu.cn

Received 12 January 2022; Accepted 21 February 2022; Published 21 March 2022

Academic Editor: Hamada Esmaiel

Modulation parameters are very significant to underwater target recognition. But influenced by the severe and time-space varying channel, most currently proposed intelligent classification networks cannot work well under these large dynamic environments. Based on supervised contrastive learning, an underwater acoustic (UWA) communication modulation classifier named UMC-SCL is proposed. Firstly, the UMC-SCL uses a simply convolutional neural networks (CNN) to identify the presence of the UWA signals. Then, the UMC-SCL uses ResNet50 as an encoder and updates the network by supervised contrastive learning loss function, which can effectively use the category information and make the eigenvector distribution of the same category more concentrated. Then, the classifier uses the feature vector output by the encoder to distinguish the final modulation categories. Finally, extensive ocean, pool, and simulation experiments are done to verify the performance of the UMC-SCL. Without any prior information, the average classification accuracy for MPSK and MFSK can reach 98.6% at 0 dB and is increased by 6% compared to the benchmark algorithm under low SNR.

1. Introduction

With the development of UWA communication technology, more and more ocean applications have installed UWA communication equipments. Through modulation classification can explore the influence of ocean multipath and Doppler effect and more effectively assistant target identification, signal identification, interference identification, and spectrum management.

In general, conventional modulation classification algorithms can be divided into two categories: likelihood-based and feature-based methods [1]. The likelihood-based method requires a large amount of prior information and computation, which makes it unsuitable for harsh noncooperative UWA communication. On the contrary, feature-based method has gradually become the mainstream method due to its low computational complexity and no dependence on prior information.

Feature-based methods consist of two parts: feature extraction and classifier. In [2], multiscale reverse dispersion entropy and grey relational degree features are used to improve the classification performance of ship-radiated noise. In [3–5], support vector machine (SVM) is used to distinguish wireless signals. In [6], high order cumulant features are put into SVM based on mixture kernel function to classify the digital signals. Wei et al. [7] use a SVM based on hybrid features, cyclostationary, and information entropy to classify the modulation types, including BPSK, QPSK, 2FSK, 4FSK, and MSK. By this means, the parameter extraction process is complicated, and the capacity is low. Even if more training data is added, the classification performance cannot always be improved [8]. For recent years, deep learning [9–16] has shown excellent performance in image feature extraction, speech recognition, and natural language processing and has been successfully used on acoustic signal sets [17, 18]. However, in modulation classification area, it is mainly used in the electromagnetic communication.

In [9], long-short term memory (LSTM) is used to classify the modulation schemes for a distributed wireless spectrum sensing network. Li et al. [10] use the I/Q data to classify signal directly through deep neural networks (DNN). In [11, 12], adaption of deep learning to the complex temporal signal domain is studied, and first proposed a CNN-based classifier to solve the problem of excessive

parameters in DNN. In [13], AlexNet and GoogLeNet are used to classify the constellation of the signal samples. Huang et al. [14] introduce a novel cascaded CNN that cascade two-block CNN to identify MPSK and MQAM hierarchically. Wang et al. [15] propose a hierarchical CNN scheme to more accurately classify the higher-order QAM signals. Liu et al. [16] combine CNN with long short-term memory (LSTM) architecture into DNN and increase the accuracy rate by 13.5% compared with original CNN. In these classic end-to-end neural networks, cross-entropy loss is the most widely used loss function to achieve the purpose of updating network weights. However, the cross-entropy loss function also lacks robustness to noisy tags [19, 20] and may have marginality [21, 22], leading to reduce generalization performance. The traditional end-to-end supervised training methods focus on the final classification accuracy rather than the quality of the features extracted from the UWA data. As a result, when the signal-to-noise ratio (SNR) becomes low, the accuracy of traditional methods will drop sharply and cannot work well. In recent years, the renaissance of contrastive learning has led to major advances in self-supervised performance learning [23–25]. When there is no available label, the data is augmented through its own cropping and flipping, and the encoder is updated through the self-supervised loss function. Although it can alleviate the disadvantages of traditional networks to a certain extent, it cannot learn from the other samples in the same category. As a result, self-supervised contrastive learning methods are not suitable for UWA data with different SNR.

In this paper, from the perspective of representation learning, we extract features with high discrimination through supervised contrastive learning [26] to support the normal classification tasks in harsh UWA channel and propose a novel classification framework named UMC-SCL. We first distinguish between valid signal and ocean noise through a simply CNN. Then, the supervised contrastive learning module will learn from the valid modulation signal and update the encoder network by supervised contrastive learning loss function. Go through this module, the features of the same category are as close as possible, and the features of different categories are as far away as possible. Therefore, we can achieve the purpose of classification only by using a fully connected layer. Finally, we verify the superiority of the proposed method through extensive ocean, pool, and simulation experiments and use principal component analysis (PCA) to visualize the output features for interpretability. Compared with the known traditional supervised networks, the proposed method greatly improves the classification accuracy under low SNR without any prior information and parameter extraction process.

2. System Model

2.1. Signal Model. The UWA communication channel is one of the most challenging wireless communication media known to human. The medium space of underwater sound propagation is very complicated, with high attenuation, long time delay, strong multipath, and high Doppler effect.

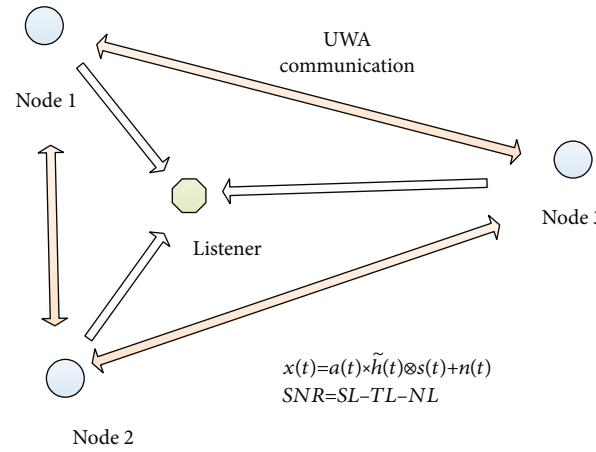


Figure 1: The progress of UWA communication.

Figure 1 shows the basic process of UWA communication. $\tilde{h}(t)$ is the energy normalized impulse response of UWA channel, $s(t)$ is the original signal, and $n(t)$ is the ocean noise. $a(t)$ is related to SNR. Node 1, Node 2, and Node 3 communicate with each other. The listener can intercept their communication signals from the sea water. SL is emitting sound source level, TL is propagation loss, and NL is the background noise level [27].

2.2. UWA Data Sources. In order to make the research result more applicable, we have constructed a complete data set through actual ocean experiments, pool experiments, and simulation experiments that are close to the reality.

2.2.1. Ocean Data. The ocean data are collected in Wuyuan Bay, Xiamen, China. As shown in Figure 2, the sound source T_{x1} and the receiving hydrophone R_{x1} are placed in the shallow sea near the footpath, with a depth of 5 m and a communication distance of 60 m.

We send and receive signals at four different times of the day. During the experiment, there are some activities such as yachts, fishing boats, and other activities that introduce a lot of man-made noise. Besides, dozens of plank road bridge piers between the sending and receiving ends make the reflection effect more significant.

2.2.2. Pool Data. Ocean experiments are costly, and the data acquisition is difficult. In order to increase the richness of the dataset, we further conduct pool experiments. The pool is located in UAC laboratory in Xiamen University. Figure 3(a) is the photo of the pool. The pool has a length of 25 m and a width of 5 m. It is divided into deep water area (depth = 1.5 m) and shallow water area (depth = 1.15 m). Figure 3(b) is the distribution of transmitter and receivers for pool experiment. T_x is the sound source, and R_x is the hydrophone. The distances between R_{x1}, R_{x2}, R_{x3}, and T_x are 3 m, 6 m, and 12 m, respectively, and the depth is 1 m. When T_x sends a signal, the sound rays will be attenuated by water and reflected on the pool wall.

Figure 2: The scene map of ocean experiment.

2.2.3. Simulation Data. Constructing a good simulation UWA channel is the basis for carrying out practical experiments. Figure 4 shows the sound ray propagation in a shallow sea channel. The sound ray will be reflected by the sea surface and bottom during propagation. Moreover, the speed of sound in seawater changes with temperature, salinity, and water depth, causing sound rays to be refracted. The speed of sound can be described according to the following formula [27].

$$c = 1449.2 + 4.6T - 0.055T^2 + 0.00029T^3 + (1.34 - 0.01T)(S - 35) + 0.016Z, \quad (1)$$

where T is temperature in, S is salinity in ppm, and Z is the depth of seawater in m.

In UWA communication, the impulse response can be assessed by beam tracing for typical acoustic communication frequencies. The basic path loss of the received signal that traveling through the UWA channel is given by [28].

$$A(l) = A_0 l^k \alpha^l, \quad (2)$$

where A_0 is a scaling constant, l is the traveling distance of sound ray, k is the spreading factor, and α is the absorption coefficient which is closely related to the frequency of sound waves and can be obtained by Thorp's empirical formula as

$$\alpha = \frac{0.11f^2}{1+f^2} + \frac{44f^2}{4100+f^2} + 2.75 \times 10^{-4} f^2 + 0.003, \quad (3)$$

where the units of α and f are dB/km and kHz, respectively. The impulse response of the multipath channel can be expressed as the summary of the transfer function of each path

$$\bar{H}(f) = \sum_p \bar{H}_p(f) e^{-j2\pi f \bar{\tau}_p} = \sum_p \frac{\Gamma_p}{\sqrt{A(\bar{l}_p)}} e^{-j2\pi f \bar{\tau}_p}, \quad (4)$$

where Γ_p, τ_p, and $\bar{l}_p$ are, respectively, the cumulative reflection coefficient of the surface and bottom, propagation delay, and the propagation distance of the p-th path. Generally speaking, an ideal surface can be modeled by a reflection coefficient $\gamma_s = -1$, while the bottom reflection can be modeled by

$$\gamma_b(\theta_p) = \begin{cases} \dfrac{\rho_p \sin\theta_p - \rho\sqrt{(c/c_b)^2 - (\cos^2\theta_p)}}{\rho_p \sin\theta_p + \rho\sqrt{(c/c_b)^2 - (\cos^2\theta_p)}}, & \cos\theta_p \le \dfrac{c}{c_b}, \\ 1, & \text{otherwise} \end{cases} \quad (5)$$

where θ_p is the grazing angle associated with the p-th propagation path and ρ and c are the nominal density and the speed of sound in water ($\rho = 1000\,\text{kg/m}^3$ and $c = 1500\,\text{m/s}$). ρ_p and c_b (calculated by Equation (1)) are the density and the speed of sound in bottom. The propagation delay of p-th path can be simple calculated as

$$\bar{\tau}_p = \frac{\bar{l}_p - \bar{l}_0}{c}, \quad (6)$$

where $\bar{l}_0$ is the direct distance from the sender to the receiver. In order to get a tractable, simple channel model, we examine an approximation to the function. Taking $p = 0$ as the reference path and $\bar{H}_0(f)$ as the impulse function corresponding to $\bar{l}_0$, the impulse function of the receiving end can be further expressed as

$$\bar{H}_p(f) = \frac{\Gamma_p}{\sqrt{(\bar{l}_p/\bar{l}_0)^k \alpha^{\bar{l}_p - \bar{l}_0}}} \bar{H}_0(f). \quad (7)$$

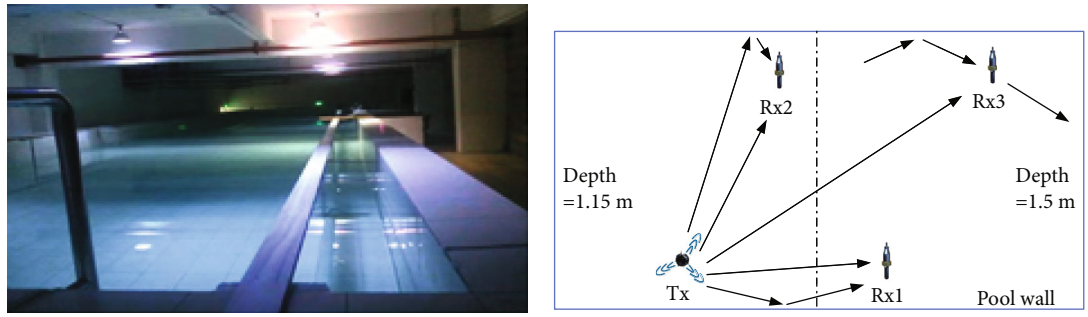
3. Supervised Contrastive Learning-Based Modulation Classification

A large number of studies have proved that DNN is superior to SVM. In the field of UWA modulation classification, the application of DNN is still scarce and all use end-to-end supervised methods. However, when the SNR becomes low, the accuracy will drop sharply. In response to this problem, we use supervised contrastive learning to narrow the feature distance between the same category and expand the distance between different categories, so as to improve the classification accuracy of modulation schemes under low SNR.

3.1. Classification System Model. As shown in Figure 5, in Step 1, the signals received by the receiver may be useful signal or useless ocean noise. In Step 2, the input signals are recognized through a simple two convolutional layers and a fully connected layer. Conv11 × 32 means the channel number is 32, and the size of convolutional kernels is 11 × 11.

If the input signal is useful signal, it will be transported to supervised contrastive learning module for further classification; if it is ocean noise, it will be discarded.

In Step 3, supervised contrastive learning loss function is used to update the backbone network (ResNet50) to extract features from UWA data and then put the features into



(a) (b)

Figure 3: The real pool (a) vs. top view of equipment distribution for pool experiment (b).

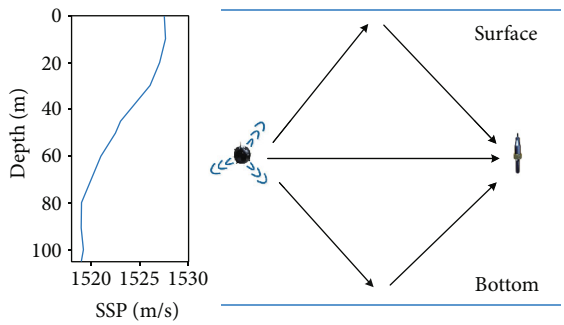


Figure 4: Sound ray propagation in shallow sea channel.

classifier for classification. By this means, the influence of ocean noise can be effectively eliminated, and the classification accuracy at low SNR will be significantly improved. In the following content, the specific network architecture will be given in details.

3.2. Backbone Network. The backbone network of supervised contrastive learning in this paper is ResNet50. ResNet50 is a residual CNN with 50 layers. It directly skips several layers and introduces the output of a certain layer into the input part of the following data layer, which overcomes the problems of low learning efficiency and ineffective improvement of accuracy due to the deepening of the network. Another two important operations in the network are batch normalization and ReLU. Batch normalization is aimed at converting the input data to an output data distribution with a variance of 1 and a mean of 0 to improve the speed of network optimization. ReLU is a nonlinear activation function. It makes the output of some neurons be 0, so as to improve the sparsity and avoid the overfitting phenomenon of the network.

In traditional supervised end-to-end CNN, as shown in Figure 6, the output of the classifier is used as the only indicator to update the network. The most widely used loss function is the cross-entropy loss function, and the expression is

$$L_{oss} = \frac{1}{N}\sum_{i} L_i = -\frac{1}{N}\sum_{i}\sum_{c=1}^{M} y_{ic}\log\left(p_{ic}\right), \tag{8}$$

where N is the number of samples and M is the number of label categories. If the true category of sample i is equal to c, then $y_{ic} = 1$; otherwise, $y_{ic} = 0$. p_{ic} is the predicted probability that the sample i belongs to the corresponding category.

3.3. Supervised Contrastive Learning. Supervised contrastive learning effectively utilizes the category label information, making the feature points from the same category closer than the points from different categories. Different from self-supervised learning [24], the positive samples are other samples in the same category. As shown in Figure 7, the progress is divided into two training stages. The first stage focuses on the training of the encoder and uses the supervised contrastive learning loss function to update the encoder. The second stage focuses on the training of the classifier using the feature output by the encoder and using the cross-entropy loss function to update.

In self-supervision, the function of the two converters is to flip or crop the input picture so that the two newly generated images can be used as the positive samples. Due to the high complexity of UWA data, cropping or flipping the time domain signal will destroy its original characteristics. Since the label information is known, the supervised contrastive learning takes all the samples from the same class in the batch as positive samples and compares them with the negative samples in the rest of the batch. The loss function becomes

$$L^{\text{sup}} = \sum_{i=1}^{2N} L_i^{\text{sup}}, \tag{9}$$

where

$$L_i^{\text{sup}} = \frac{-1}{2N_{\tilde{y}_i} - 1}\sum_{j=1}^{2N} 1_{i\neq j}\cdot 1_{\tilde{y}_i=\tilde{y}_j}\cdot\log\frac{\exp\left(z_i\cdot z_j/\tau\right)}{\sum_{k=1}^{2N} 1_{i\neq k}\cdot\exp\left(z_i\cdot z_k/\tau\right)}, \tag{10}$$

where i is the blind UWA data and z_i represents the feature generated by the backbone network. z_j represents the feature that comes from the same category with data i, and z_k represents the feature generated by backbone network that is different from data i. τ is a scalar temperature parameter larger than 0. $\tilde{y}_i$ is the category label of i. To update the network parameters under the constraint of the loss function, the

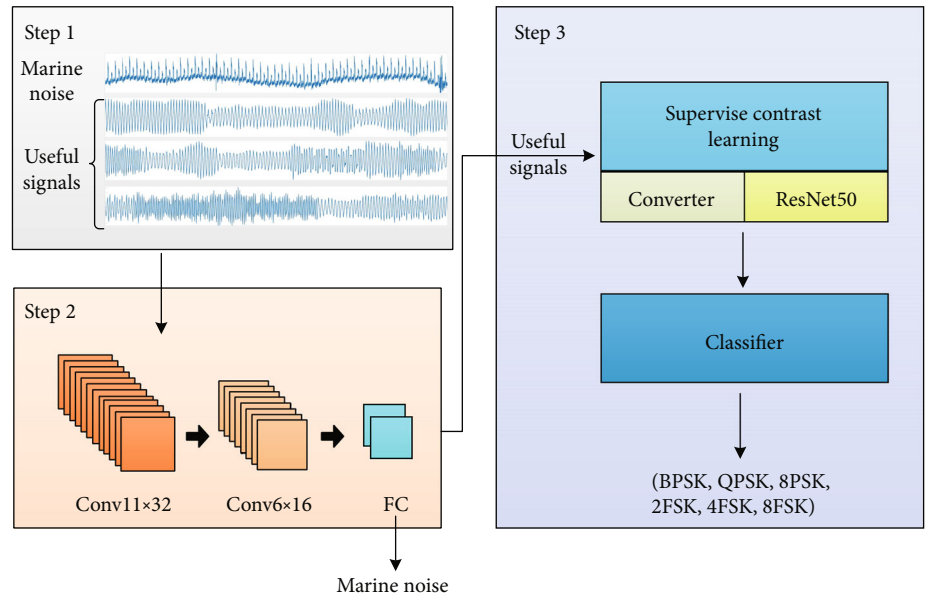


Figure 5: Flow chart of supervised contrastive learning-based modulation classification scheme.

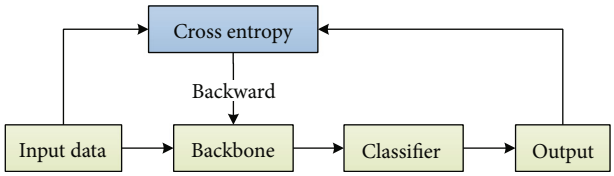


Figure 6: Schematic diagram of traditional supervised end-to-end training progress.

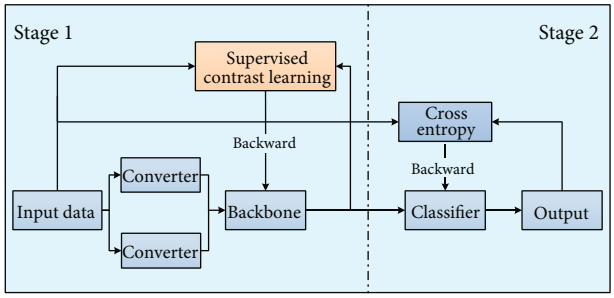


Figure 7: Schematic diagram of supervised contrastive learning structure.

feature output from backbone network will have the following characteristics:

(1) The sum of the cosine distance between the feature vectors of all other samples in the same category and the feature vectors of the sample i, the larger the better

(2) The sum of the cosine distance between the feature vectors of the sample in different categories and the feature vectors of the sample i, the smaller the better

The classifier in the second stage is a simple fully connected layer. It uses the 2048-dimensional standardized feature output by the encoder to classify the modulation schemes. It should be mentioned that the parameters of the encoder are frozen in the second stage. Therefore, whether the encoder can obtain excellent features after training plays a decisive role.

Algorithm 1 describes the update process of the supervised contrastive learning.

4. Experiments and Results

In this section, the details of the experiments are explained. We also evaluate the modulation classification performance of the proposed method and compare it with the existing methods. In order to analyze the algorithm performance more intuitively, we use PCA to visualize the features to provide the interpretability of the proposed method.

4.1. Dataset Generation. The original modulation signals are generated through the MATLAB simulation platform. The candidate modulation set is given by

$$M = \{\text{BPSK, QPSK, 8PSK, 2FSK, 4FSK, 8FSK}\}. \quad (11)$$

Table 1 shows the parameter setting of different modulation schemes. The ocean noise is actually collected in the Wuyuan Bay sea area. After passing through the ocean channel, the pool channel, and the simulation channel, the data with the characteristics of multipath fading and Doppler frequency shift is obtained. On this basis, Gaussian white noise with different SNR is superimposed on the obtained data through MATLAB. In this paper, the intraband SNR is used to evaluate the performance of the proposed algorithm. It can be calculated as

$$\text{SNR} = 10\log_{10}\left(\frac{F_s}{B_s}\right) + \text{SNR}_{\text{Gaussian}}(dB), \quad (12)$$

Input: Encoder training: batch size 32, initial learning rate α=5e − 2, epoch E_p=100, τ=0.07
Classifier network training: batch size 128, initial learning rate $\alpha' = 1e - 3$, epoch E_p'=100
Output: Backbone network parameter θ, The Classifier network Θ.
//Encoder training
1: **for** epoch =1:E_p
2: sample a batch of data, update α as described in Section $III - D$
3: Backbone encodes m into F.
4: calculates loss L^{sup} (10)
5: update θ with $\theta \longleftarrow \theta - \alpha \cdot \nabla_\theta L^{sup}$
6: **end for**
//Classifier network training
7: **for** epoch = 1:E_p' **do**
8: Freeze encoder parameter, update α' as described in Section $III - D$
9: Classifier network decodes F into result
10: calculates loss L_{oss} (8)
11: update Θ with $\Theta \longleftarrow \Theta - \alpha' \cdot \nabla_\Theta L_{oss}$
12: **end for**
//Finish training
Return the parameter θ, Θ

Algorithm 1: Two-stage training of supervised contrastive learning.

Table 1: Parameter setting of the modulation signals.

Modulation type	2/4/8PSK	2/4/8FSK
Modulation point	$\theta_m = 2\pi m/M$ $m = 0, 1, \cdots, M-1$	$f_m = 11\,\text{kHz} + 2\,\text{mkHz}$ $m = 0, 1, \cdots, M-1$
Sample frequency	66 kHz	
Symbol width	1 ms	

where F_s is the sampling frequency and B_s is the bandwidth of the signal.

In Step 2, it is aimed at distinguishing the ocean noise and the useful signals. The train set consists of 2,000 ocean noise samples and 2,000 modulation signal samples with different SNR. The corresponding test set is 800 samples per category. In Step 3, the training set of supervised contrastive learning consists the data with different SNR after noise pollution. Among them, 550 samples of each modulated signal are generated from -9 dB to 9 dB every 2 dB, 250 samples of which are used as the training set and 300 samples are used as the test set. Therefore, the training set contains 15,000 samples with different SNR, and the test set of each SNR contains 1,800 samples.

4.2. Experimental Implements. In the ocean and pool experiments, NI USB-6259 Pinout capture card is used to convert the digital signal to analog signal at the transmitter and convert the analog signal to digital signal at the receiver. JYH500A power amplifier and Type-2692-0S2 charge amplifier are used to amplify the transmitted signal and the received signal, respectively. WBT22-1107 transducer which can convert the analog electrical signal to acoustic signal is used to send and receive signal in the water. Besides, the experiments are performed on computing server equipped with an Intel(R) Core(TM) i7-9700K 3.6GHz CPU, a NVIDIA GeForce RTX 2060 SUPER GPU, "Pytorch" and "Python" programming language, the CUDA 10.1 and CUNDD software. The optimizer of ResNet50 is "Adam," and the learning rate is 0.05 and decays to 10% of the original learning rate every 30 epochs.

4.3. Experiment Results

4.3.1. Simulation Results. The simulation experiment is carried out under the simulation UWA channel. In the noise distinction stage, the distinction between ocean noise and useful signals is obvious, especially in the frequency domain. Even when the SNR is -6 dB, the classification accuracy can still achieve 100%. Therefore, it can be explained that the simple convolutional network of Step 2 can well eliminate the influence of marine noise. In the Step 3, Figure 8 gives the classification accuracy of six modulation schemes. In general, the classification accuracy of six modulation signals increases with the increase of SNR and can achieve an average accuracy of 98.84% at 0 dB. When the SNR decreases to -6 dB, the recognition of 8PSK is the most difficult, and the confusion of modulation categories is mainly concentrated on QPSK and 8PSK.

4.3.2. Actual Ocean and Pool Experiment Results. Due to the difficulty and high cost of obtaining ocean data, in practical experiments, we mix pool data with the ocean data to increase the richness of training set, so that the trained encoder and classifier can better fit the distribution characteristics of UWA data. The result of Step 2 in practical experiments is the same as mentioned in the previous simulation part. In Step 3, using the feature output by the encoder, the classification accuracy of the single fully connected layer is shown in Figure 9. For MPSK, its information is modulated in phase, so its characteristics in the time domain are not as obvious as MFSK. When the SNR is -6 dB, the average accuracy of MPSK is 79.7%, while MFSK can achieve a high

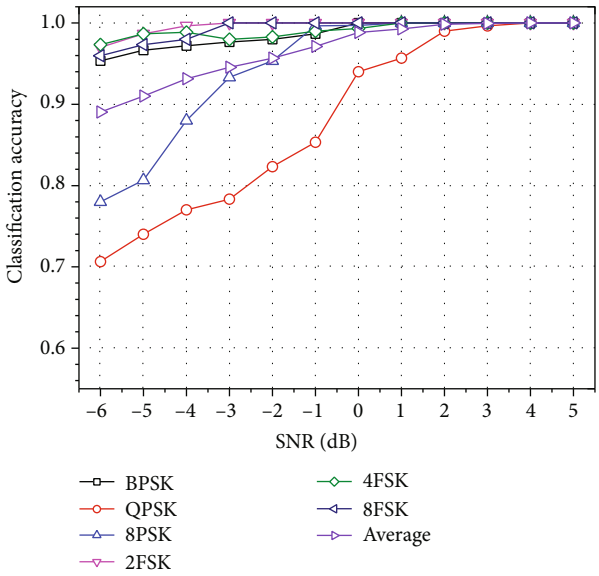


Figure 8: Classification accuracy of simulation experiments versus SNR.

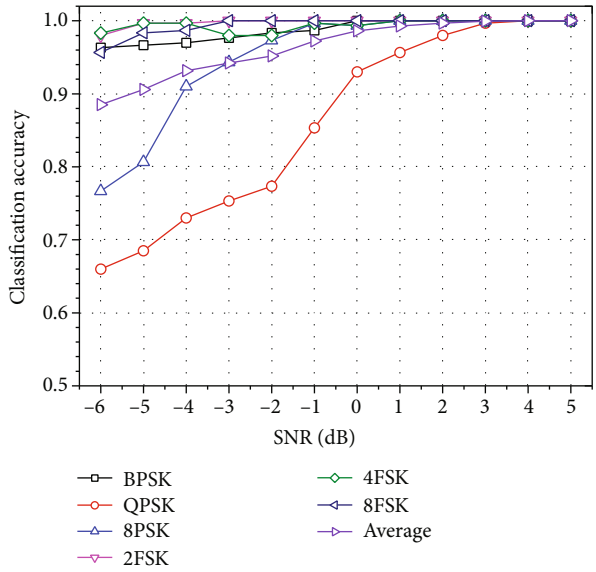


Figure 9: Classification accuracy of ocean and pool experiments versus SNR.

accuracy of 97.3%. When the SNR increases to 0 dB, the average accuracy of six types of modulation signals can reach 98.6%.

Classification performance results of six modulation categories at -6 dB and 0 dB are presented using confusion matrix in Figure 10. In each modulation category, 300 tests are implemented. When SNR is -6 dB, BPSK, 2FSK, 4FSK, and 8FSK have achieved high classification accuracy through supervised contrastive learning. However, since QPSK and 8PSK are relatively similar in modulation phase, they are easy to be confused. There are 102 QPSK samples that are mistaken for 8PSK and 70 8PSK samples that are mistaken for QPSK. When the SNR reaches 0 dB, except for the slightly larger classification error of QPSK, the recognition accuracy of other modulation schemes almost reaches 100%.

4.3.3. Accuracy Comparison. To verify the superiority of the proposed method in this paper, the performance is investigated by making comparisons with four relevant algorithms in recent years; the comparison algorithms are as follows:

(1) Algorithm 1 based on ResNet50 using constellation density as feature [29]

(2) Algorithm 2 based on AlexNet using 3-channel image as feature [13]

(3) Algorithm 3 based on VGGNet using original gray image as feature [30]

(4) Algorithm 4 based on SE-Net using the features in time domain, frequency domain, and time-frequency domain [31]

Figure 11 presents the average classification accuracy of five algorithms versus SNR. The average accuracy is obtained by averaging the classification performance of six modulation categories. As shown in Figure 11, the following observations can be made.

(1) For all five algorithms, the modulation classification performance improves with an increasing SNR value

(2) Given the same SNR, in addition to the proposed algorithm, the other four algorithms will have a sharp decay on the classification accuracy when the SNR becomes low

(3) The proposed supervised contrastive learning algorithm has strong adaptability to low SNR UWA modulation signals and outperform all other algorithms. When the SNR is -6 dB, the accuracy of our proposed method is 6% higher than the benchmark algorithm [29]

4.3.4. PCA for Interpretability. PCA can reduce a set of n -dimensional vectors to k-dimension through orthogonal transformation. That is, k unit orthogonal basis is selected, so that the original n-dimensional data is represented by this group of basis. For high-dimensional data, first make the mean of the input vector to 0 and then use the covariance to represent the correlation between vectors a and b. The covariance is calculated as

$$\mathrm{Cov}(a,b) = \frac{1}{n}\sum_{i=1}^{n}(a_i-\mu_a)(b_i-\mu_b) = \frac{1}{n}\sum_{i=1}^{n}a_i b_i. \tag{13}$$

For mn-dimensional vectors $\{a_1, a_2, \cdots a_m\}$, the matrix X is composed of

$$X = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}. \tag{14}$$

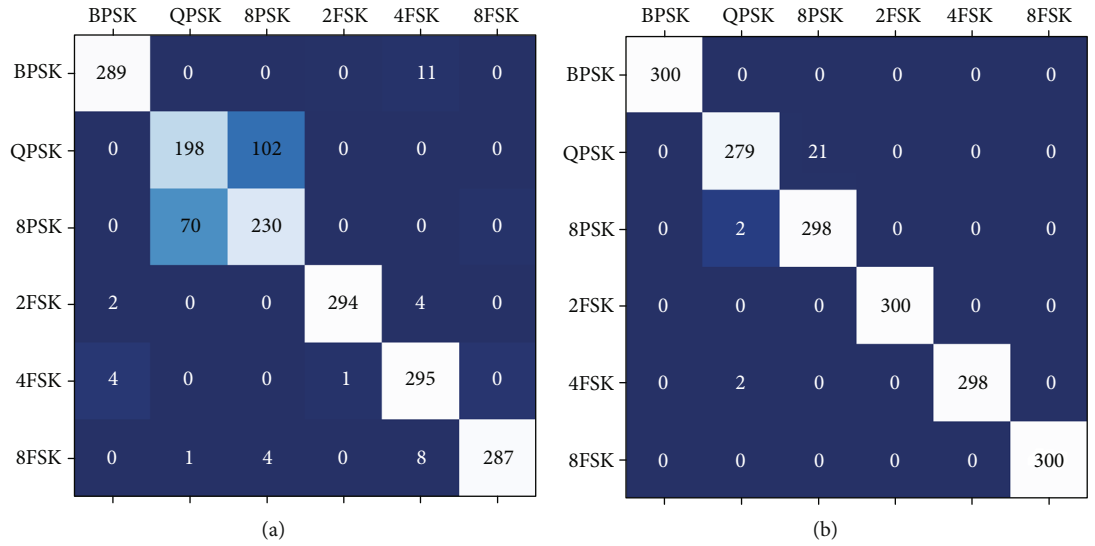


Figure 10: Confusion matrix under different SNR based on supervised contrastive learning.

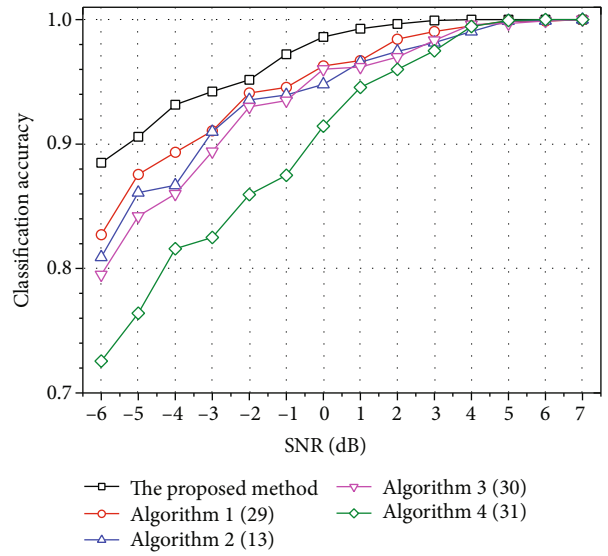


Figure 11: Classification accuracy of the proposed algorithm and the comparison algorithm versus SNR.

The covariance matrix C is

$$C = \frac{1}{n}XX^T, = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} a_{1i}{}^2 & \cdots & \frac{1}{n}\sum_{i=1}^{n} a_{1i}a_{mi} \\ \vdots & \ddots & \vdots \\ \frac{1}{n}\sum_{i=1}^{n} a_{mi}a_{1i} & \cdots & \frac{1}{n}\sum_{i=1}^{n} a_{mi}{}^2 \end{pmatrix}. \quad (15)$$

It can be seen that the diagonal of the matrix C is the variance of the vectors, and the other elements are the covariances between different vectors. Supposing $Y = PX$ is the vector of the original data X projected to the low-dimensional space, P is the transformation matrix, and D is the covariance matrix of Y, there is the following equation

$$D = \frac{1}{n}YY^T = \frac{1}{n}(PX)(PX)^T = PCP^T. \quad (16)$$

In order to enable the transformed low-dimensional vectors to represent more original information, we hope that they are not correlated with each other; that is, the covariance is equal to 0. Therefore, the matrix D should be a diagonal matrix. According to the relevant knowledge of linear algebra, the matrix P should be the eigenvector matrix of matrix C, and it should be arranged from top to bottom according to the size of the corresponding eigenvalues. Select the matrix P_k composed of the first k rows of matrix P, and obtain a matrix Y_k with k-dimensional vectors. Taking $k = 3$,

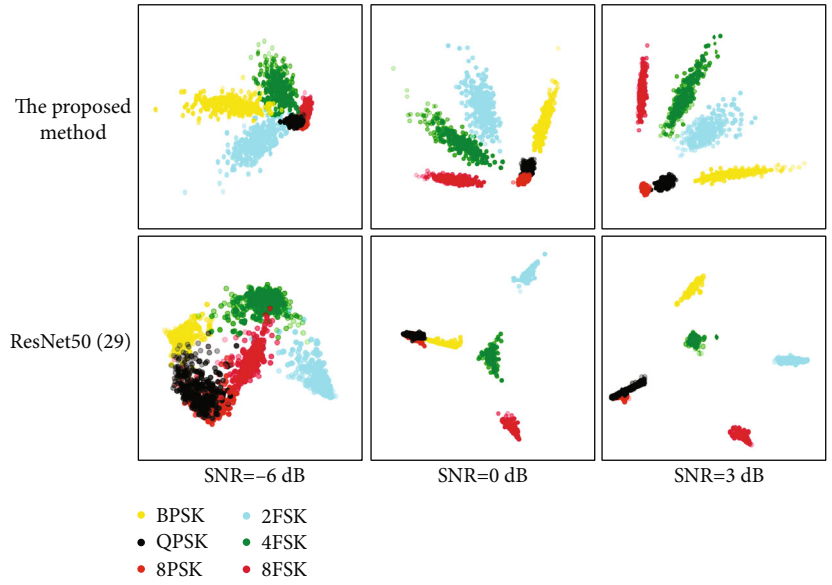


Figure 12: Feature point distribution after dimensionality reduction by PCA.

the high-dimensional feature outputs by the network are presented in a 3-dimensional plane. Figure 12 shows a 3-dimensional space cross-sectional view of the feature point distributions extracted by different networks.

It is easy to see that the features extracted by the supervised contrastive learning method have a higher degree of discrimination and better classification effect under low SNR. When SNR is -6 dB, the features extracted by ResNet50 [29] are overlapped. In contrast, the features extracted by the proposed method, except that the features of QPSK, 8PSK, and 8FSK, have some overlap; the feature distributions of the other three modulation signals are concentrated and easy to distinguish. What is more, with the increase of the SNR, the feature point distribution boundaries of different modulation schemes become clearer and clearer.

5. Conclusion

In this paper, we are the first to propose a novel modulation classification scheme based on supervised contrastive learning. Firstly, the useful signals and ocean noise will be distinguished in the first module. Secondly, the encoder ResNet50 in the supervised contrastive learning module will learn the input UWA data under the guidance of the supervised contrastive learning loss function to update the network. By this means, the distance between feature vectors in the same category but with different SNR will be minimized, and the distance between feature vectors of different categories will be expanded as much as possible. Then, the classifier recognizes the modulation scheme according to the feature output by the encoder. Finally, the ocean, pool, and simulation experimental results verify the superiority of the proposed method. Compared with the existing researches, the experimental verification in this paper is more complete. The proposed method eliminates the complex parameter extraction process and does not require any prior information. When the SNR is 0 dB, the average accuracy can achieve 98.6%. Compared to the benchmark algorithm, the accuracy at -6 dB is improved by 6%. Moreover, we use PCA to visualize the feature distribution, which can intuitively analyze the superiority of the proposed algorithm.

Data Availability

The data used to support the findings of this study were supplied by Daqing Gao under license and so cannot be made freely available. Requests for access to these data should be made to Daqing Gao (dqgao@stu.xmu.edu.cn).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (62071400, 62071402).

References

[1] A. Ali and Y. Fan, "k-sparse autoencoder based automatic modulation classification with low complexity," *IEEE Communications Letters*, vol. 21, no. 10, pp. 2162–2165, 2017.

[2] Y. Li, S. Jiao, and B. Geng, "A comparative study of four multiscale entropies combined with grey relational degree in classification of ship-radiated noise," *Applied Acoustics*, vol. 176, no. 4, article 107865, 2021.

[3] K. Tekbyk, Z. Akbunar, A. R. Ekti, G. K. Kurt, and A. Grin, "On the investigation of wireless signal identification using spectral correlation function and SVMs," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, Marrakesh, Morocco, 2019.

[4] S. Zhou, Z. Wu, Z. Yin, and Z. Yang, "Noise-robust feature combination method for modulation classification under

fading channels," in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, Chicago, IL, USA, 2018.

[5] J. Huang and R. Diamant, "Adaptive modulation for long-range underwater acoustic communication," *IEEE Transactions on Wireless Communications*, vol. 19, no. 10, pp. 6844–6857, 2020.

[6] X. U. Wen and B. Wang, "A signal modulation classification method based on SVM," *Computer Engineering*, vol. 8695, no. 1, pp. 117–136, 2013.

[7] Y. Wei, S. Fang, and X. Wang, "Automatic modulation classification of digital communication signals using svm based on hybrid features, cyclostationary, and information entropy," *Entropy*, vol. 21, no. 8, p. 745, 2019.

[8] J. Fan, Y. Ren, X. Luo, and J. Joung, "Iterative carrier frequency offset estimation scheme for faster-than-nyquist signaling systems," *IEEE Photonics Technology Letters*, vol. 32, no. 18, pp. 1203–1206, 2020.

[9] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin, "Deep learning models for wireless signal classification with distributed low-cost spectrum sensors," *IEEE Transactions on Cognitive Communications and Networking*, vol. 4, no. 3, pp. 433–445, 2018.

[10] W. Li, Z. Dou, C. Wang, and Y. Zhang, "Signal modulation classification based on deep belief network," in *2019 IEEE Globecom Workshops (GC Wkshps)*, Waikoloa, HI, USA, 2019.

[11] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *International Conference on Engineering Applications of Neural Networks*, Springer, 2016.

[12] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[13] S. Peng, H. Jiang, H. Wang et al., "Modulation classification based on signal constellation diagrams and deep learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 3, pp. 718–727, 2019.

[14] J. Huang, S. Huang, Y. Zeng, H. Chen, S. Chang, and Y. Zhang, "Hierarchical digital modulation classification using cascaded convolutional neural network," *Journal of Communications and Information Networks*, vol. 6, no. 1, pp. 72–81, 2021.

[15] Y. Wang, M. Liu, J. Yang, and G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive radios," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 4074–4077, 2019.

[16] X. Liu, D. Yang, and A. E. Gamal, "Deep neural network architectures for modulation classification," in *2017 51st Asilomar Conference on Signals, Systems, and Computers*, pp. 915–919, Pacific Grove, CA, USA, 2017.

[17] L. Zhang, D. Wang, C. Bao, Y. Wang, and K. Xu, "Large-scale whale-call classification by transfer learning on multi-scale waveforms and time-frequency features," *Applied Sciences*, vol. 9, no. 5, p. 1020, 2019.

[18] S. Yu, K. J. Palmer, M. A. Roch, E. Fleishman, and H. Klinck, "Deep neural networks for automated detection of marine mammal species," *Scientific Reports*, vol. 10, no. 1, p. 607, 2020.

[19] S. Sukhbaatar, J. Bruna, M. Paluri, L. Bourdev, and R. Fergus, *Training Convolutional Networks with Noisy Labels*, Computer Science, 2015.

[20] Z. Zhang and M. R. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[21] G. Elsayed, D. Krishnan, H. Mobahi, K. Regan, and S. Bengio, "Large margin deep networks for classification," 2018, http://arxiv.org/abs/1803.05598.

[22] W. Liu, Y. Wen, Z. Yu, and M. Yang, *Large-Margin Softmax Loss for Convolutional Neural Networks*, ICML, 2016.

[23] O. Henaff, "Data-efficient image recognition with contrastive predictive coding," in *International Conference on Machine Learning*, pp. 4182–4192, PMLR, 2020.

[24] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9729–9738, Seattle, WA, USA, 2020.

[25] R. D. Hjelm, A. Fedorov, S. Lavoie-Marchildon et al., "Learning deep representations by mutual information estimation and maximization," 2018, http://arxiv.org/abs/1808.06670.

[26] P. Khosla, P. Teterwak, C. Wang et al., "Supervised contrastive learning," 2020, http://arxiv.org/abs/2004.11362.

[27] R. J. Urick, *Principles of Underwater Sound 3rd Edition*, McGraw-Hill Book Company, 1983.

[28] P. Qarabaqi and M. Stojanovic, "Statistical characterization and computationally efficient modeling of a class of underwater acoustic communication channels," *IEEE Journal of Oceanic Engineering*, vol. 38, no. 4, pp. 701–717, 2013.

[29] Y. Kumar, M. Sheoran, G. Jajoo, and S. K. Yadav, "Automatic modulation classification based on constellation density using deep learning," *IEEE Communications Letters*, vol. 24, no. 6, pp. 1275–1278, 2020.

[30] D. Sun, Y. Chen, J. Liu, Y. Li, and R. Ma, "Digital signal modulation recognition algorithm based on VGGNet model," in *2019 IEEE 5th International Conference on Computer and Communications (ICCC)*, Chengdu, China, 2019.

[31] Q. Qu, S. Wei, H. Su, M. Wang, and X. Hao, "Radar signal recognition based on squeeze-and-excitation networks," *IEEE International Conference on Signal, Information and Data Processing (ICSIDP)*, 2019, Chongqing, China, 2019, 2019.