IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 3, 2021, accepted March 12, 2021, date of publication March 18, 2021, date of current version March 30, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3067070

Underwater Communication Signal Recognition Using Sequence Convolutional Network

YAN WANG[1], YIHENG JIN[2], HAO ZHANG[3,4], (Senior Member, IEEE), QIAN LU[5], CONGHUI CAO[6], ZHANLIANG SANG[7], AND MEI SUN[1]

[1]School of Physics and Electronic Engineering, Taishan University, Tai'an 271016, China
[2]School of Science and Information Science, Qingdao Agricultural University, Qingdao 266109, China
[3]Open Studio for Marine High Frequency Communications, Pilot National Laboratory for Marine Science and Technology, Qingdao 266237, China
[4]Department of Electrical Engineering, Ocean University of China, Qingdao 266100, China
[5]Department of Computer Science and Technology, Qingdao University, Qingdao 266071, China
[6]School of Physics and Information Engineering, Jianghan University, Wuhan 430056, China
[7]Department of Technical Engineering, CRRC Qingdao Sifang Company Ltd., Qingdao 266111, China

Corresponding author: Hao Zhang (zhanghao@ouc.edu.cn)

This work was supported in part by the Marine S and T Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology, Qingdao, under Grant 2018SDKJ0210, in part by the Scientific Research Startup Foundation of Taishan University under Grant Y-01-2020016, and in part by the Foundation of State Key Laboratory of Acoustics, Chinese Academy of Sciences under Grant SKLA201903.

ABSTRACT Automatic modulation recognition (AMR) is one of the essential parts in the intelligent communication system. In the underwater acoustic communication, it is a challenging work that promptly and easily recognizes the signal modulation schemes by conventional methods. The deep neural network method is a good solution to the problem, which creates a better recognition effect. The packets of data that are fed to the familiar neural network is constant. However, the packets of signal data on the communication course consistently change, which seriously reflects on the signal recognition veracity. A novel deep learning network with the sequence convolutional network in this paper is proposed, which is composed of one-dimensional sequence convolution of residual network modules and the variable convolution kernel range. By extracting the time-domain signal characteristics, the affection of various signal packets can be mitigated. In experiments, the employed network not only has more concentrated on the modulation recognition veracity, but also owns a lower parameter quantity and a shorter training time, which indicates ideal recognition results in the underwater communication environment. Moreover, it is more valuable to the real underwater communication system.

INDEX TERMS Deep learning, convolutional neural network, modulation recognition, underwater acoustic communication.

I. INTRODUCTION

The task of AMR, often considered as the signal recognition, may mainly include classifying individual signal arguments of modulation schemes to identify the communication style, which is imposed between the transceivers on the application scene. Customarily, the signal recognition has discovered more intentions in the military and civilian context [1]. It may be noticed that the signal recognition work, according to different application requirements, could be broadly used in the non-cooperative field and the network security field. The military context contains the test, analysis and identification of unknown signal modulations from the potential source of the adversary communication, which has a critical impact on electronic warfare, signal surveillance, and communication interception [2]. As a comparatively new study field, the cognitive radio (CR) has found more real applications in the civilian context, which has been regarded as the concrete form of software defined radio (SDR) [3]. It presents the solution that is the client-oriented and scalable management of the communication resource. The purpose of CR is to fully develop the reusable wireless resource at the greatest extent, applying agile, alterable and reconstructed software defined transceivers. CR rebuilds these transmission arguments based on the actual wireless environment in both time and frequency domain. The information of arguments can be flowed

The associate editor coordinating the review of this manuscript and approving it for publication was Qinghua Guo.

between CR transceivers through the wireless channel. It is important to note that this way requires some extra communication arguments to decrease the system effectiveness. AMR provides an executable solution to the inefficient way. At the receiver, the AMR algorithm provides a substituted plan for retrieving these arguments from the received signal.

Water has a low absorption ratio in low-frequency acoustic signals, and it is more evident in the shallow sea environment. Actually, the beneficial transmission mechanism relies on the sound wave in the underwater communication, which is disadvantageously obstructed by several situations, such as reflecting and scattering multi-path interference, external and internal surrounding noise, salinity and temperature variation, etc., [4]. The spreading speed is the extremely low mode of around 1500 m/s. The underwater acoustic channel associates two aspects of attributes: one is a bad condition of the physical link comparing the terrestrial mobile radio communication with the worst-case performance, the other is a large delay of the wireless beam comparing the satellite transmission [5], [6]. Although the lower frequency better carries on the acoustic transmission course, the communication resources are greatly constrained in the available bandwidth, may normally be from a few tens of hertz to a few kilohertzs. These statuses make the job of the underwater signal modulation recognition extremely difficult.

The most favorite modulation recognition methods are Likelihood-based (LB) [7]–[10] and feature-based (FB) [11]–[15]. The former depends on the likelihood function of received signals; the latter takes on the modulation signal characteristics. LB methods are enabled with a proportion verification between two assumptions to issue the honest judgment of different modulation schemes, which create opportunities for the better recognition performance with taking more cut-and-try struggles to select suitable judgments. With more instinctive methods, LB would discover the maximum likelihood among all candidates, called the maximum likelihood (ML) recognizer, it does not need carefully designed judgments and is easier to fulfil. LB methods are on behalf of the optimal recognizer in the circumstance of the prior channel state knowledge.

Since LB recognizers produce the best recognition result, their high calculation consumption has an influence on the actual applied deployment, which drives the generation of FB recognizers. FB methods take the sub-optimal effectiveness, and have a much lower calculation wastage. The time and frequency characteristics in FB are well extracted, which are adjusted to the modulation recognition of both analogue and digital signals. In the actual recognition course, FB recognizers need the judgment strategy formality, and miscellaneous modulation schemes will be decomposed into several branches of subgroups, where the recognized modulations could be separated from the other. The most classical result-making forms in FB are to set the judgment at each decision-identifying point. Normally, a given state information of channel or a Gaussian white noise of signals is considered to determine the judgment. The most popular method with the high-order statistics characteristics mainly cover the moduli of time and cumulant, which is appropriate for the modulation recognizer. On the basis of the wavelet-based characteristics, FB recognizer undertakes the waveform function for the measure essence, and the signal cyclic characteristics specify the cyclostationary analysis as the underlying classification type.

Deep learning (DL) as a part of the machine learning has a variety of applications in image recognition, speech recognition, and natural language understanding [16]. DL has been considered as indispensable tools for worth in many areas of the wireless communication, such as RF signal processing [17], radio resource allocation [18], [19], radio control [20]–[22], MIMO detection [23], [24], channel estimation [24]–[26] and IoT detection [27], [28]. In the terrestrial environment, DL for AMR primarily refers to the network methods, such as the convolutional neural network (CNN), the recurrent neural network (RNN) and the network of the composite structural configuration with both CNN and RNN. The deep full-connected feedforward network showed a satisfactory effect for AMR in fading channels [29]. The modified CNN accomplishes a more obvious classification quality comparing to 2-layer CNN and 32-layer ResNet [30], [31]. The extensible neural network with many hidden layers earns a more correct probability of AMR [32]. The CLDNN (Convolutional Long Short-term Deep Neural Network) analyzes the structure design affected by different hyper parameters, and there are excellent recognition outcomes comparing with CNN, ResNet and DenseNet [33]. The separate network structure comprised of CNN and long-short term memory (LSTM) belongs to the fusion pattern in the two paths of the structure, which has powerful function for the AMR problem [34]. There are few studies of AMR with DL in the underwater acoustic communication. The deep sparse automatic encoder handles better in the random signal conflict, and is qualified for recognizing the four modulation schemes [35]. The DL method distinguishes five underwater signal modulation schemes, and there are the better results than the statistic's method [36]. The mixed-structure network consisting of LSTM and CNN learn the multidimensional signal characteristics, which exceeds the impulse noise disturbance to certify the AMR work competence [37].

It can be seen from these pursuits of literatures, which highlights that the proper network structure can improve the recognition rate of AMR. This paper considers the sequence convolutional network (SCNet), an innovative structure of one-dimensional sequence convolution (1DSC), for the underwater acoustic signal modulation recognition. In the network form, the recognition rate of acoustic signal modulations is mainly improved by stacking one-dimensional convolutional residual network modules and the variable convolution kernel range (CKR) in 1DSC. The complexity of underwater communication environment makes it difficult to distinguish the modulations of received signals. Through SCNet, more recognition characteristics of time sequence signals can be obtained to promote the modulation recognition

effect. The signal packets sent in the communication course are constantly changing. SCNet can adapt to the influence of various signal packets and achieve the effective modulation recognition. During the signal sequence processing, the extraordinary cost of too many computing resources and too many parameter quantities are two main obstacles normally encountered in the ordinary RNN and the sophisticated structure of the deep network, which causes the entire system to run inefficiently. SCNet can capitalize on the parallel manipulation of the 1DSC structure to solve problems. The contributions of this paper are mainly as follows:

(1) The recognition effect of SCNet is verified, and it can identify the various signal modulation schemes in harsh underwater communication conditions, which adapts to the different packets of the recombination signals.
(2) SCNet takes in the method of the variable CKR in varying 1DSC layers to enrich the types of the extracted signal characteristics and further heighten the recognition accuracy, which has a great favor to the efficiency improvement with no extra calculation amount.
(3) The employed network is authenticated by the simulation experiment with the real underwater channel arguments. There is also the comparison result of the diverse cross-layer connection modes and the variable CKR.

The remaining parts of this paper is organized as follows. The communication signal model is given in Section II along with the basic CNN and RNN structure. In Section III, there are the details of SCNet structure and the various packets of the signal dataset to the modulation recognition. Section IV discusses the modulation signal dataset produced at great length, and the modulation recognition performance is evaluated in the real underwater acoustic channel. Finally, Section V provides a summary of the paper.

II. SIGNAL AND NETWORK MODELS

A. SIGNAL MODEL

There are several principal kinds of factors influencing the underwater communication, such as multi-path, time delay, doppler and additive white Gaussian noise (AWGN), etc. Fig. 1 shows the underwater channel model [35], which is the most representative form of received signals

$$y(t) = h(t,\theta) \otimes x(t) + n(t) = \sum_{i=1}^{I} p_i(t)x(t-\theta_i(t)) + n(t) \tag{1}$$

where $x(t)$ is the transmitted signal, $n(t)$ is AWGN, $h(t,\theta)$ is the channel arguments, including multi-path and time delay,

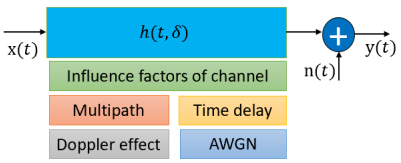


FIGURE 1. The underwater acoustic channel model.

$p_i(t)$ is the ith path signal attenuation and $\otimes$ denotes the signal convolution operation. $\theta_i(t)$ is the ith path signal time delay, I is the multi-path number, and there is a similar doppler scaling factor α at all paths, $\theta_i(t) \approx \theta_i - \alpha t$ [38]. Transmitting signals can be digital (e.g. phase-shift keying) or analog (e.g. phase amplitude modulation).

B. NETWORK MODEL

In the multi-layer deep learning network, each layer acts as the characteristic's extractor. The neurons of the network layer extract feature vectors that input from the previous layer, and map them into a new vector space for further learning, which can be one, two or three dimension convolution. While the one-dimensional convolution is used in the natural language processing (NLP) [39], 2D and 3D convolution are used in the image processing [40], 2D for the single image, 3D for multiple images and videos. The usual 2D convolution operation is shown in Fig. 2. The green and yellow cube represents two data planes, and the blue cube conducts the convolution results. Eight cubes, including four green cubes and four yellow cubes in the transparent black box, are connected to the blue result cube by the convolution calculator.

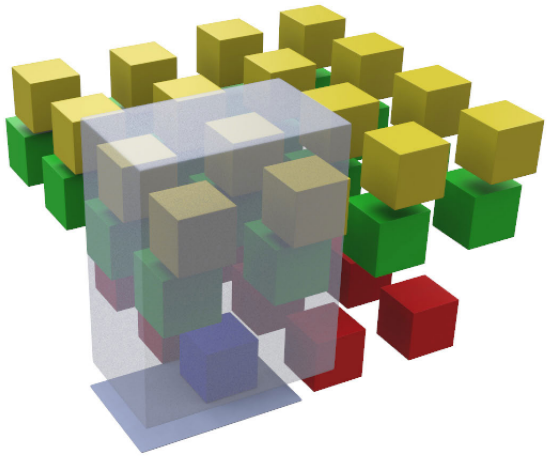


FIGURE 2. 2D convolution operation.

There is only a simple superposition of CNN layers, which is easy to encounter the degradation phenomenon [31]. It demonstrates that the certain network scale can properly play a considerable role instead of blindly expanding the restriction. When further deepening the network layers, the recognition result will become worse. The network learns an excess of characteristics of the training dataset, and the recognition accuracy decreases gradually and tends to be saturated. As a result, there is a limited number of network layers, and more hidden data characteristics cannot be ulteriorly extracted to improve the recognition effect. When the network is degraded, it shows that the shallow network can achieve a better training effect than the deep network. By analyzing values transferred between the network layers, it is always that values learned from the front layer are transferred

to the back layer, and the effect in the deeper network should be at least no worse than that of the shallow network.

Through the characteristics' delivery learned within network layers, it can overcome problems to some degree, which caused by deepening the network. From the perspective of information theory, there is the existence of processing data inequalities, and the data information contained in the signal characteristics' map make the layer-by-layer decrease in the forward transport. The identity mapping in the residual network structure ensures that the back layer in the network must involve more data information than that in the front layer. Based on the idea using the identity mapping to connect the varying layers in the network, the layer number is further added for a better recognition ability, which is the residual network module in Fig. 3, where Conv2D represents two-dimensional convolution layer, ReLU represents an activation function layer that uses the relu function [41], BN represents the batch normalization layer that is the input data standardization technique, Add represents the Add layer that attaches the two paths in the residual network module.

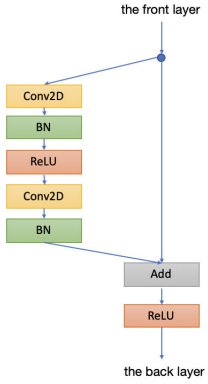


FIGURE 3. Basic residual network module.

Communication signal dataset is a kind of the time sequence data, and the signal data before and after actually have the inevitable correlation. RNN is an ordinary neural network for coping with the time sequence shown in Fig. 4. RNN performs well in almost all sequence problems, including speech recognition, machine translation and handwriting recognition, etc. In the application, there is a serious problem in the internal design of RNN. Since RNN can only handle one-time step at a time, the next step must wait until the previous step is completed. It means that RNN cannot do massive and parallel processing like CNN, which is extremely computationally intensive because all intermediate results must be saved before the entire task runs.

When CNN handles data, it regards data as a two-dimensional matrix. Moving to the time sequence, it can be considered as one-dimensional object ($1 \times n$ vector). Through the multi-layer network structure, a large enough receptive field can be obtained to extract more signal data information to achieve better results. CNN supports overlay layers to obtain the advanced recognition characteristics of data. The

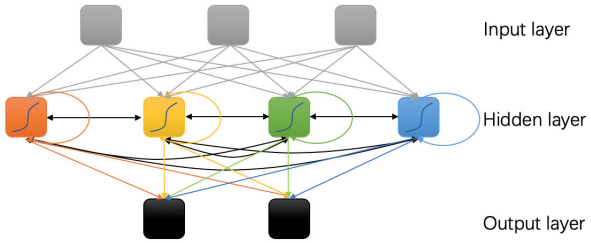


FIGURE 4. RNN structure.

calculation in CNN does not depend on the previous time information, so each calculation is independent and can be paralleled to get the utmost out of the hardware system power. Meanwhile, the reverse propagation of one-dimensional CNN structure is unlike RNN, thus avoids the problem of gradient vanishment and gradient explosion that often occurs in RNN, especially for the long input training data [42], [43]. Thanks to the impressive speciality of the large-scale parallel processing in CNN structure, which can be carried out no matter how deep the network to raise the computational efficiency.

III. SCNet NETWORK STRUCTURE

Due to the special propagation carrier in underwater acoustic communication environment, it is tough to recognize the received signal modulation schemes. In order to obtain more advanced recognition characteristics of the signal dataset, the designment of the deep network structure is essential. However, when network layers are accumulated continuously, common insurmountable problems make it extremely exhausting to train the deep network, which can be better addressed by the network structure optimization of SCNet.

A. NETWORK STRUCTURE

1) STRUCTURE DESIGN

This paper adopts the design form of stacking multiple residual modules, which are connected in series in Fig. 5. It can effectively overcome the network model problems caused by the multiple layers. CNN can extract low/middle/high-level features. The deeper the network is, the more abstract the features are, and the more distinguishing information is. In the prediction task of the time-series dataset, the network achieves the ability to acquire the high-dimensional input spatial features through deeper layers. The more multi-level SCNet has the stronger stability requirement, and it is necessary to avoid the gradient problem (vanishing gradient and exploding gradient) caused by the deeper network structure. There is the solution for the problem to initialize the weight parameters and adopt the Batch Normalization regularization layer, and the deeper network can be trained. When the gradient problem is solved, another problem will arise, which is the network degradation. As the layer number of the network increases, the accuracy rate on the training set becomes saturated or even decreases. Theoretically, the solution space of the deep network includes the solution space of the shallow

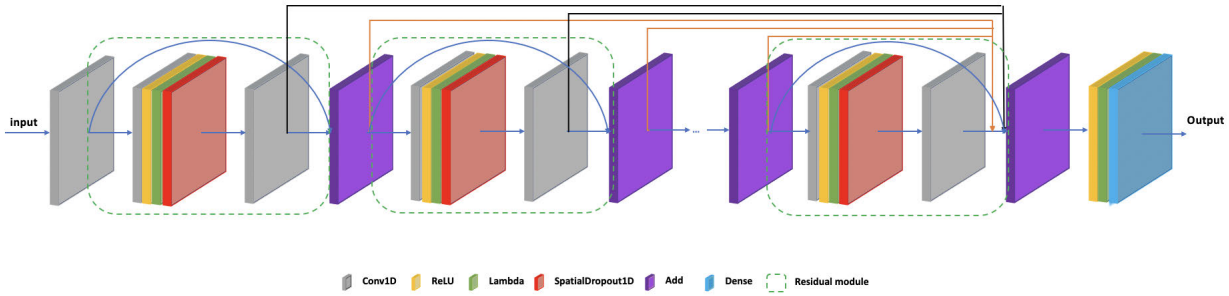


FIGURE 5. SCNet structure.

network, so the performance of the deep network should be greater than or equal to the shallow network. However, both the training error and the testing error of the deep network are larger than those of the shallow network in the training results. Although the solution space of the shallow network is contained in the solution space of the deep network, the random gradient descent strategy is used in training, which often gets the local optima instead of the global local optima. Obviously, the solution space of the deep network is more complex, and the optima cannot be obtained by the random gradient descent strategy. The network degradation problem can be rectified to connect the redundant layer of the network by the identity mapping of the residual structure, which regards the deep network as the shallow network. Each residual module (green dotted line block diagram) is composed of Conv1D, ReLU, Lambda and SpatianlDropout1D plus the identity mapping. Conv1D represents the one-dimensional convolution layer, and weights after the convolution operation are all handled by Lambda, which is the weight normalization layer. In addition, after each Lambda layer in the residual module, Spatianl-Dropout1D is added to strengthen the generalization ability of the trained network. In SCNet structure, ReLU and Add represents the equivalent functionality like the above basic residual network module. The residual modules are also connected by the Add layer to establish the construction of the deep network. The internal network only transfer values learned by the residual module to get the limited recognition effect. For better acquiring more hidden signal characteristics, it is a practical way to find more learned values of middle layers, which are further supplemented by transmitting more information gained Conv1D (black solid line) and Add (orange solid line). This will help to promote a higher talent that extracts signal characteristics. In the final output, Add, ReLU and Lambda serve as the complete network, which uniformly deals with the cross-layer values passed by Conv1D and Add. Dense outputs the final recognition results.

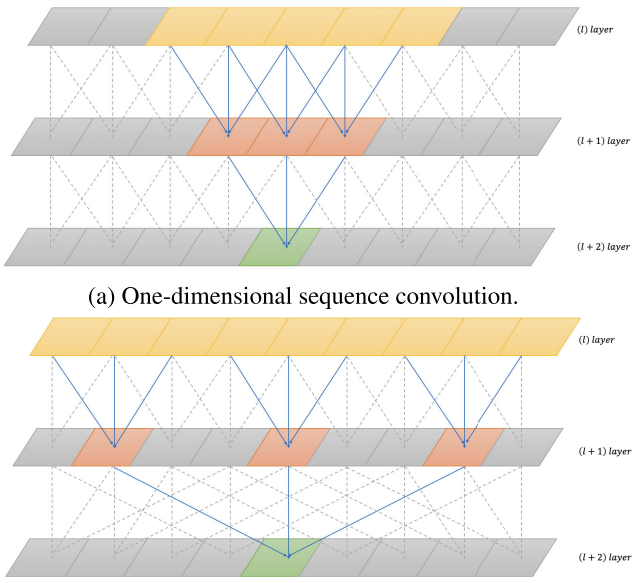
The formula of SCNet structure is

$$\mathcal{S}_M = \sum_{m=1}^{M} [s_m + \mathcal{G}(r_m)] + \bigwedge_{n \in N} \mathcal{Z}_n \tag{2}$$

where $\mathcal{S}_M$ represents the output of the employed network. s_m represents the signal characteristics learned by the residual module, m represents the number of residual modules, $m = 0, 1, 2, \cdots, M$. When $m = 0$, r_0 represents the input original signal data, M represents the total number of residual modules. $s_m = w_m \times r_{m-1} + b_m$, w_m represents the weight, r_{m-1} represents the input from the front layer, and b_m represents the deviation. $\mathcal{G}(r_m)$ represents the learned residual part, $\mathcal{G}(r_m) = r_m - r_{m-1}$, which is the residual identity mapping operation. $\bigwedge_{n \in N}(\cdot)$ represents the selection method, which is the Conv1D layer or the Add layer in the employed network. $\mathcal{Z}(\cdot)$ represents to choose a cross-layer approach, and $N = 1, 2, 3$ represents three ways. The cross-layer connection mode of the Conv1D layer, the Add layer, and the Conv1D layer with the alliance of the Add layer is corresponding to $n = 1$, $n = 2$, $n = 3$, respectively.

2) INTERNAL STRUCTURE

The convolution layer of SCNet combines 1DSC in Fig. 6 (a) with the variable CKR in Fig. 6 (b), and each square in layers represents a neuron, which contains the fixed



(a) One-dimensional sequence convolution.

(b) The variable convolution kernel range.

FIGURE 6. Internal structure.

convolution kernel size. ReLU, Lambda, SpatianlDropout1D and Add in intermediate layers are temporarily ignored for the clear description. The time sequence convolution operation requires that the prediction of y_t at the t time can only be determined by the input from x_{t-1} to x_1. The requirements for 1DSC can be achieved to limit the sliding window of the convolution, which can be realized by simply shifting the convolution output at several time steps. The purpose of 1DSC is to ensure that the prediction at the previous time step will not deal with the future data information. It can be guaranteed that the output at the t time step will only be obtained by the $t-1$ time step and the previous time step of the convolution operation. In contrast to the fact that RNN cannot directly predefine the data object to length. 1DSC can ascertain the data to be trained as a sequence by the length information defined in advance, and it assures the parallel processing efficiency. In the training course, the convolution prediction of all past time steps can be parallel, and their input and labeled true values are known, which has the enhanced capacity over RNN in the execution speed. Simultaneously, the simplified network of SCNet with 1DSC is more productive than the sophisticated network structure.

1DSC requires a larger convolution kernel to extend the receptive field to enrich the signal recognition information. Nonetheless, it needs to increase increases the computational complexity by the convolution operation of the broader scope. When the signal modulation schemes is classified, a larger convolution kernel will also cause the redundant local signal characteristics to be acquired, resulting in an unsatisfactory final recognition effect. Therefore, it is necessary to widen CKR without adding the redundant signal information for covering the whole signal dataset in 1DSC. The biggest advantage of the variable CKR is to extend the receptive field progressively, which does not insert the blank data between the sequence convolution but skips over some existing data. It is equivalent to keep the unchanged input and add some weights with zero values to the convolution kernel. Where the calculation amount is basically unchanged, the sequence range of signals observed by the network is fortified. If the stride length of the general convolution operation is exaggerated, it can also extend the receptive field. When the convolution stride is greater than 1, it will have the down sampling impression, and the output sequence range will be reduced. The signal data at the previous sequence cannot be covered in the condition, and the time sequence analysis cannot be finished.

It can be seen in Fig. 6 (b) that the convolution form of the variable CKR is very similar to the sequence convolution in Fig. 6 (a), and the biggest difference is the dilation of the convolution kernel. There are no empty holes in convolution kernel windows used by the sequential convolution, and the data involved in the convolution operation are closely connected together. As the layer number increases, the convolution kernel window will become larger and larger in the variable CKR, and more signal data will be skipped in the convolution kernel window. Through the variable CKR, the receptive field of the back convolution layer can be more extensive, so that more signal history information can be introduced. While the neuron in the second layer can see 3 neurons in the first layer, each neuron in the output layer can see 9 neurons in the first layer. If the output layer needs to remember a longer length, the network layer number needs to be augmented by the corresponding layer number. The advantage of the variable CKR is that there is no information loss contrasting to the pooling operation [44]. As the receptive field is enlarged, each convolution output contains a wide range of rich history information. When the deep convolution network structure is accepted, it ensures that the convolution kernel covers all the inputs in the effective history information.

1DSC and the variable CKR is translated into the formula expression

$$D^{(l)}(a) = \mathcal{C}^{(l)}(\sum_{e^{(l)}}^{E^{(l)}} d_{e^{(l)}}(k^{(l)}) \times \cdots \mathcal{C}^{(1)}(\sum_{e^{(1)}=1}^{E^{(1)}} d^{(1)}_{e^{(1)}}(k^{(1)}) \times \mu_{a-(E^{(1)}-e^{(1)})})) \quad (3)$$

$D^{(l)}(\cdot)$ represents the (l) layer output of neurons selected in the (l) convergence layers, μ represents the input sequence, and a represents the time sequence number according to neurons in the network layer. $\mathcal{C}^{(l)}(\cdot)$ is the function of the (l) layer in the network to select inputs for the sequence convolution operation in the previous layer. $e^{(l)}$ is the sequence number of neurons selected in the (l) layer, $e^{(l)} = 1, 2, \cdots, E^{(l)}$, $E^{(l)}$ is the total neuron number in the relating layer. $d^{(l)}_{e^{(l)}}(\cdot)$ represents the 1DSC operation in the (l) layer. The receptive field in each layer of the variable CKR is $k(l), k^{(1)} = 1, k^{(2)} = 3, \cdots, k^{(L)} = K, K = 3^L$, which exponentially multiplies by 3 to extend the receptive field range. As shown in Fig. 6 (b), k is 1, 3, 9. Assume that 9 neurons of the first layer are $\mu_a, \mu_{a-1}, \mu_{a-2}, \mu_{a-3}, \mu_{a-4}, \mu_{a-5}, \mu_{a-6}, \mu_{a-7}, \mu_{a-8}$, The last output of the third layer is $D^{(3)}$. The dependent 1DSC operation outputs of the first layer and the second layer are

$$\begin{aligned} D^{(1)} = (&d^{(1)}_1(k^{(1)}), d^{(1)}_2(k^{(1)}), d^{(1)}_3(k^{(1)}), \\ &d^{(1)}_4(k^{(1)}), d^{(1)}_5(k^{(1)}), d^{(1)}_6(k^{(1)}), \\ &d^{(1)}_7(k^{(1)}), d^{(1)}_8(k^{(1)}), d^{(1)}_9(k^{(1)})) \end{aligned} \quad (4)$$

$$D^{(2)} = (d^{(2)}_1(k^{(2)}), d^{(2)}_2(k^{(2)}), d^{(2)}_3(k^{(2)})) \quad (5)$$

Substituting Equ. 4 and Equ. 5 into Equ. 3 has

$$\begin{aligned} D^{(3)} = &d^{(3)}_1(k^{(3)}) \\ &\times \{d^{(2)}_1(k^{(2)}) \times [d^{(1)}_1(k^{(1)}) \cdot \mu_{a-8} + d^{(1)}_2(k^{(1)}) \cdot \mu_{a-7} \\ &+ d^{(1)}_3(k^{(1)}) \cdot \mu_{a-6}] \\ &+ d^{(2)}_2(k^{(2)}) \times [d^{(1)}_4(k^{(1)}) \cdot \mu_{a-5} + d^{(1)}_5(k^{(1)}) \cdot \mu_{a-4} \\ &+ d^{(1)}_6(k^{(1)}) \cdot \mu_{a-3}] \\ &+ d^{(2)}_3(k^{(2)}) \times [d^{(1)}_7(k^{(1)}) \cdot \mu_{a-2} + d^{(1)}_8(k^{(1)}) \cdot \mu_{a-1} \\ &+ d^{(1)}_9(k^{(1)}) \cdot \mu_a]\} \end{aligned} \quad (6)$$

B. SIGNAL PROCESSING

The familiar CNN requires to import a constant magnitude of data in the network. However, the solution is not applicable to the recognition of signal modulations. Data with the constant magnitude cannot fully exploit the information profoundly embedded signals. When the signal dataset is split to different magnitude of packets, there is a powerful genius to extract key recognition thresholds in the modulated signals. The different magnitude of packets matching up the signal formats are represented as

$$\Psi_J(l) = \sum_{j=1}^{J-1} \psi(j) + \psi_l(J) \tag{7}$$

where a byte composes of eight binary signal elements, and several bytes form a packet in terms of the demand for the input training data. $\Psi(\cdot)$ is the total signals in the current location. j is the packet number, J is the starting packet in the new position, and $\psi(\cdot)$ is how many chosen signal bytes according to the packet number. l is the present accessing number of the signal bytes. The employed network treats $\Psi_J(l)$ as the input format of classified signals, and the different modulation schemes are efficiently recognized by SCNet. The recognition efficacy of DL model is better safeguarded by the signal data normalization, and each signal packet data are distributed to the unit vector for fitting the feature diversities between the simulation experiment and the live environment. The data format is the real vector with 32 bits of the floating-point type. It is saved as a 2-dimension eigenvector with the vogue utility of numpy in the DL ecosystem, which is stored as $P \times I$ form. P represents the input packet, which is equivalent to $\Psi_J(l)$. I represents the input dimension. SCNet uses Conv1D in the network design, and I is set to 1, which corresponds to the one dimension data entry mode.

IV. EXPERIMENT

A. SIGNAL DATASET AND TRAINING PARAMETER SETTING

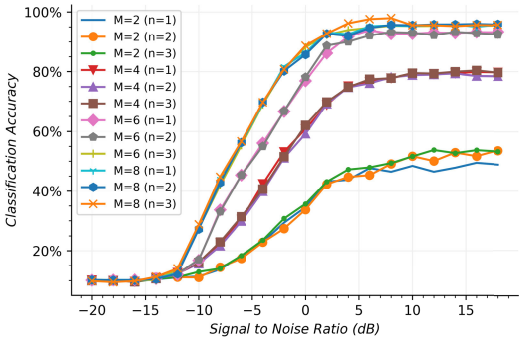
The employed network method is evaluated by the simulation experiment with the signal dataset, which establishes on the real channel arguments in the shallow underwater environment [45]. The underwater testing environment has an average depth of about 53 meters. There are approximately 1.5 kilometers away from the distance between the transducer and the hydrophone, and they are placed at the depth of 38 meters in water. The attenuation factor is around 0.01875 $dB/wavelength$, the velocity of sound is roughly 1574 m/s, and the density of water reaches 1.268 g/cm^3. The communicational channel exhibits the typical underwater sparse features. Ten signal modulation schemes are considered, such as binary phase-shift keying (BPSK), quadrature PSK (QPSK) and 8PSK, 16 quadrature amplitude modulation (16QAM), 32QAM, single-sideBand (SSB), frequency modulation (FM), pulse amplitude modulation (PAM), 4 frequency-shift keying (4FSK), 8FSK. The doppler shift is set to 5×10^{-3} [5]. The carrier frequency is 10 kHz, and the symbol transmission rate is 1000 symbol/s. SNRs are from -20 dB to 20 dB, and the packets are 81, 243, 729 and 2187, respectively. The sample rate of the maximum deviation is 15 Hz in the random fashion, which is 1 Hz per sample in the sample rate during the standard offset drift process. The sampling clock drift is in the random fashion. The sampling frequency deviation between the transmitter and receiver usually results in the frequency offset before received signals are demodulated. The deviation also leads to the performance degradation of the time synchronizer algorithm at the receiver. In order to achieve the better synchronization in the communication system, the symbol clock synchronization module frequently searches the fast fourier transform (FFT) window to ensure the synchronization information. It will take up too many computing resources, resulting in the significant decline in the communication system performance. Through the setting method, the simulation experiment can be more realistically close to the actual communication situation. The Rayleigh distribution does duty for the time fading model, and there is 20 cosines in the frequency selective fading simulation. The additive noise is assumed to be band-limited, zero mean white Gaussian noise, and the random seed number is 14631, which is generated by the noise generator. The filter of the raised cosine pulse-shaping is the 0.35 roll-off factor.

The original data are txt format and mp3 format files (two files are half of each), and they are converted into the binary data through coding, which is modulated by various modulation schemes to be sent. It ensures that the final modulation recognition effect has nothing to do with sent contents. The distribution of continuous data in the dataset may be similar, and the sending contents may influence the recognition outcome. The uncertainties of influence factors are taken into account, and the training dataset and the testing dataset are randomly and dispersedly formed by the data index serial numbers. Through the use of the index, it is not necessary to scan the entire dataset, and directly locates the signal data record that meets the index number, which greatly speeds up the query of the captured data. Contemporarily, the manner pledges that the recognition effect of the trained network model is independent of the distribution and specific content of the original txt and mp3 data. The number of training vectors is 5000 for each modulation scheme, and there are 5000 testing vectors in one of the modulation schemes. The batch size of the employed network is set to the input time sequence number, corresponding to the packets. In the training course, the optimizer, the loss function and the early stopping number are taken into consideration to refine the learning effect of networks. The Adam optimizer [46] has been adopted. It is different from the stochastic gradient descent (SGD) [47], which maintains the fixed single learning rate to update all the weights. Adam combines the advantage of adaptive gradient (AdaGrad) [48] and root mean square propagation (RMSProp) [49]. AdaGrad maintains the learning rate for each parameter to improve the performance in the sparse gradients, and RMSProp adaptively pursues the learning rate ground on the nearest magnitude of the weight

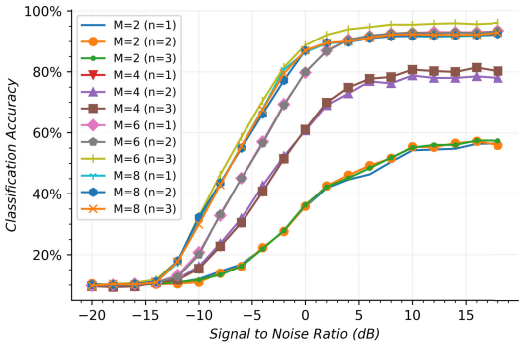
gradient for each parameter. Adam not only calculates the adaptive parameter learning rate of the first-order moment, but also makes full use of the second-order moment of the gradient. The initial value in the moving average is close to 1, and the order moment deviation estimate is close to zero. In the issue, the deviation is improved for calculating the estimated deviation. The common loss functions include mean squared error (MSE) [50], binary cross entropy (BCE) [51] and categorical cross entropy (CCE) [52]. MSE assumes that the error matches the Gaussian distribution, which cannot be satisfied in the AMR task to result in bad performances. In the perspective of information theory, the minimizing results of the cross entropy are in good agreement with the effects of the maximum likelihood, so that the cross entropy is used for the loss function. BCE only compresses and restores the output, and the output of the sigmoid activation function is transformed into the range between 0 and 1. Obviously, it is difficult to deal with the multi-classification problem in this way. The output value of CCE is a one-hot vector, and the output activation function is Softmax, which limits the output range of each dimension to (0, 1). The output sum of all dimensions is 1, which represents the probability distribution. This method is in accordance with the compatibility of the modulation signal classification problem. The loss function chooses the CCE function. The early stopping mechanism [53] is an inconspicuous form of regularization. It hardly needs to change the training process, the value of the objective function or the set of the legitimate parameters. It means that the early stopping is easier to use without affecting the learning mechanics of the network. This is different from the weight-decay used in the regularized L1/L2 methods [53]. The method directly adds the weight values to the cost function, which is controlled by a reasonable range during the training. In the training process, the problem of the weight-decay setting must be brought into focus. Otherwise, the network is easy to fall into the local minimum value, and cannot correctly obtain the global minimum value to achieve the optimal effect of the network. The number of early stopping mechanism is set to 5 for the improved generalization performance.

B. EXPERIMENT PERFORMANCE

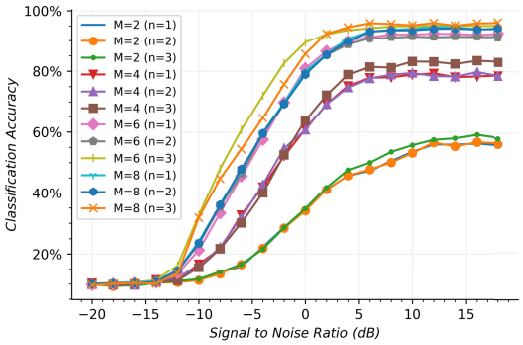
In Fig. 7, there is the recognition performance with the employed network. M represents the residual module number, and n represents the cross-layer connection mode. As M grows from 2 to 8, CKR used by Conv1D in the residual module grows exponentially with a base of 3, which is the increase of CKR from 1 to 3^6. At SNR < -15, there is a similar recognition result to about 10% in 4 forms of input packets. With the increase of SNRs, the recognition performance dramatically improves. In Fig. 7 (d), $M = 6$ $(n = 3)$ has the best effect from -15 dB to 0 dB at 2187 packets, which is an average of around 44.4%, 27.2% and 1.5% more than $M = 2$ $(n = 1) \sim (n = 3)$, $M = 4$ $(n = 1) \sim (n = 3)$, and $M = 8$ $(n = 1) \sim (n = 3)$, respectively. In the same SNR range, the other three packets forms play out in similar



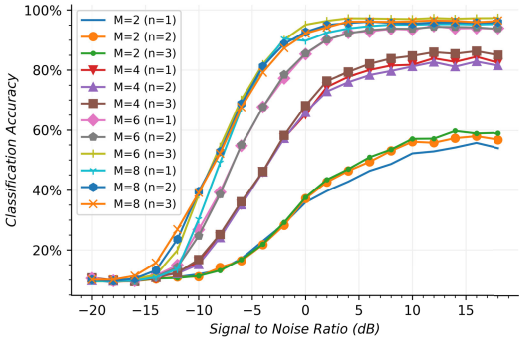
(a) Modulation recognition accuracy at 81 packets.



(b) Modulation recognition accuracy at 243 packets.



(c) Modulation recognition accuracy at 729 packets.

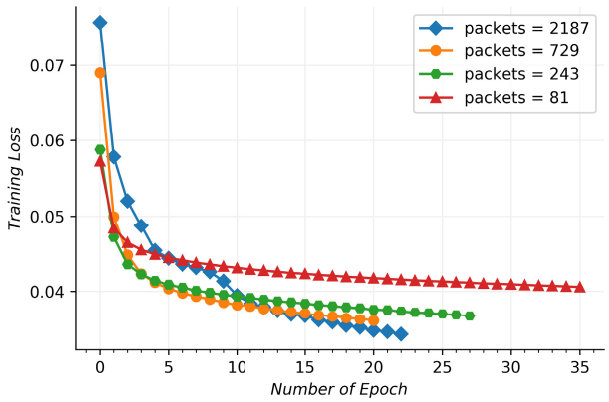


(d) Modulation recognition accuracy at 2187 packets.

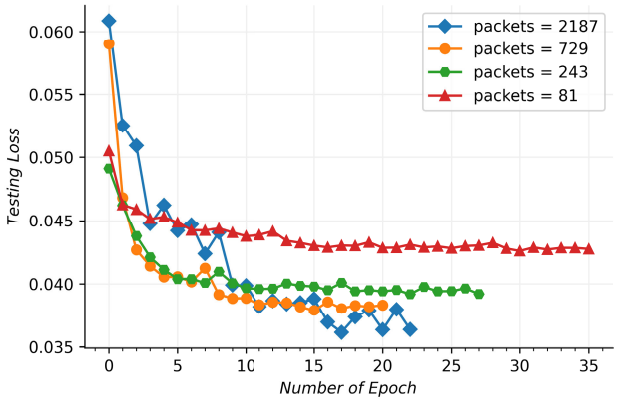
FIGURE 7. The recognition performance of the different number of residual modules and the cross-layer connection modes at various packets.

fashion in Fig. 7 (a) ∼ (c). In Fig. 7 (a), $M = 6$ $(n = 3)$ is, on average, about 29.3%, 20.2% higher than $M = 2$ $(n = 1) \sim (n = 3)$, $M = 4$ $(n = 1) \sim (n = 3)$, respectively. $M = 6$ $(n = 3)$ and $M = 8$ $(n = 1) \sim (n = 3)$ show the high degree of recognition consistency, and the former is slightly almost 0.3% lower than the latter. In Fig. 7 (b), $M = 2$ $(n = 1) \sim (n = 3)$, $M = 4$ $(n = 1) \sim (n = 3)$, and $M = 8$ $(n = 1) \sim (n = 3)$ average approximately 32.4%, 13.9% and 2.1% less than $M = 6$ $(n = 3)$, respectively. A similar event occurs in Fig. 7 (c), and $M = 2$ $(n = 1) \sim (n = 3)$, $M = 4$ $(n = 1) \sim (n = 3)$ and $M = 8$ $(n = 1) \sim (n = 3)$ is nearly 33.2%, 14.3% and 7.3% fewer than $M = 6$ $(n = 3)$ on average, respectively. The trend of recognition results between $M = 6$ $(n = 3)$ and $M = 8$ $(n = 1) \sim (n = 3)$ are closer at SNR ≥ 0. $M = 6$ $(n = 3)$ is averagely 3.6%, 1.0% and 1.8% better than $M = 8$ $(n = 1) \sim (n = 3)$ at 243, 729 and 2187 packets, and $M = 6$ $(n = 3)$ is averagely 0.3% worse than $M = 8$ $(n = 1) \sim (n = 3)$ at 81 packets. When the employed network is fed by the sufficient signal data, the recognition performances are promoted to some extent. The more interactions the residual module has, the better the recognition effect will be. It is the dominant consideration that the deeper network and the enlarged kernel size in layers obtain the advanced signal characteristics. When the residual module number increases to a certain extent, the recognition effect does not continue to grow at $M = 8$. It is chiefly because too deep network suffers from the degradation problem. In this case, there is a similar recognition result between three cross-layer connection modes between $(n = 1)$, $(n = 2)$ and $(n = 3)$ in 4 forms of input packets. Under the condition of $M = 2$ and $M = 4$, three cross-layer connection modes have the small difference of recognition performance at SNR ≤ 0, and $M = 6$ $(n = 3)$ is 42.4%, 15.6% higher than $M = 2, M = 4$ on average in 4 forms of input packets. When there are few signal characteristics transferred from layers in the shallow network, the performance is improved to some degree. From −15 dB to 5 dB, $M = 6$ $(n = 3)$ has better recognition effect about 8.5%, 9.9%, 9.6% and 10.7% than the separate mode of $(n = 1)$ and $(n = 2)$ at 81, 243, 729 and 2187 packets, which is approximately 2.4%, 2.8%, 3.2% and 3.1% higher than the two connection modes at SNR ≥ 5 dB. It is shown that the recognition effect can be further improved by transferring more signal characteristics from layers in the appropriate depth of SCNet.

In the training and testing course of various packets, there is the convergence situation in Fig. 8. In th vertical axis, Training Loss in Fig. 8 (a) is the loss result that the employed network is fed by the training dataset to learn the signal recognition characteristics, and Testing Loss in Fig. 8 (b) is the loss result that the trained network is verified by the signal testing dataset. Two losses are computed from the categorical crossentropy function. Number of Epoch in horizontal axis is the number of epoches, which means the network passes throughout the overall dataset and returns once. During the training course, Training Loss at packets = 81 is average poorer than the other three by more than 0.56. When packets



(a) Training convergence course



(b) Testing convergence course

FIGURE 8. Convergence situation in the training and testing course.

are 2187, Training Loss achieves the best convergence. With the rise of epoches, losses have a sustained reduction, which proves the employed network can acquire the essential signal characteristics to complete the training course. Testing Loss has a nearly trend to Training Loss. The testing course is slightly more volatile than the training course, which can finally converge to complete the verification. The trained model is definitely optimized for the training dataset, and the sample distribution of the testing dataset is not exactly the same as that of the training set, which will produce oscillatory results. The situation is similar to the training course, and epoches of the testing course constantly decline when packets boost. It shows that the employed network can apply to a variety of packets.

According to the recognition performance of four kinds of input packets in Fig. 7, SCNet selects the optimal network form with M = 6 (n = 3) to introduce more contrast to the recognition effect in different packets. Fig. 9 illustrates the modulation recognition accuracy in various packets, which is gradually increased with the expansion of packets at SNR $\geqslant$ −16 dB. The packets of 243 is above average 1.6% higher than the packets of 81. The packets of 729 has the similar

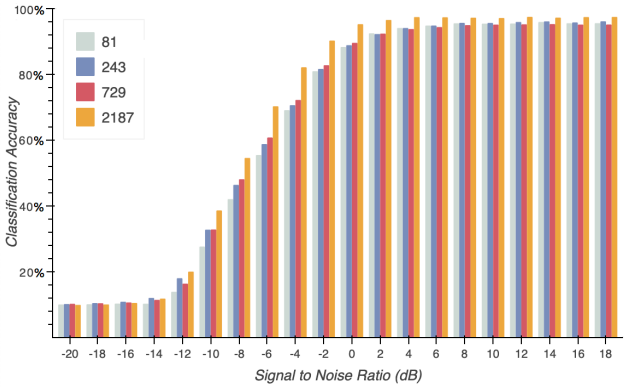
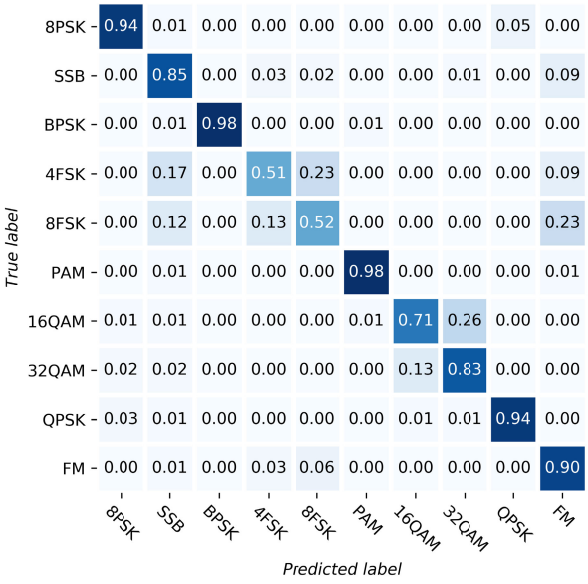


FIGURE 9. Modulation recognition accuracy with various packets.

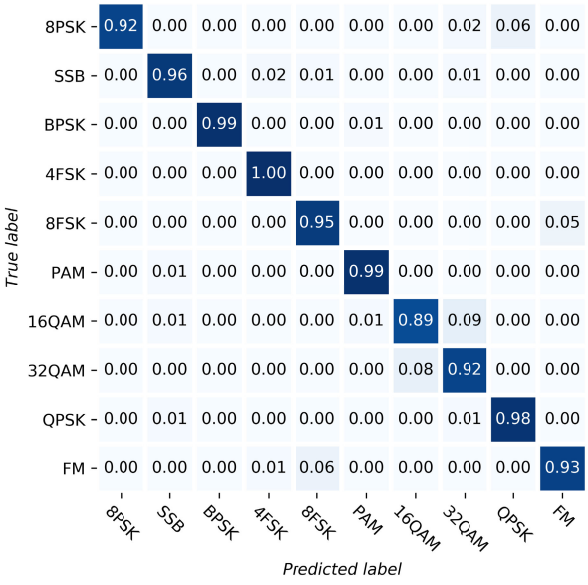
recognition effect to that of the packets of 243, and the discrepancy between them does not exceed 0.7%. The packets of 2187 have a certain improvement to the packets of 729, and the former was about average 3.6% higher than the latter. The employed network can realize the effective modulation recognition in the various packets. Comparing with the various packets, it is helpful for improving the recognition effect by inputting more signal data. The better effects mainly come from that more signal data contain more hidden recognition characteristics, which benefit the employed network to learn more precise classification thresholds for the favorable results.

Fig. 10 shows the recognition effect of different modulation schemes at typical SNRs of the 2187 packets. In Fig. 10 (a), the recognition rates of 4FSK and 8FSK are low at SNR $= -4$ dB. 4FSK and 8FSK misrecognize each other, which are also misrecognized as SSB and FM. The low SNR seriously interferes with the analog signal waveforms, which is easily confused with the results of the poor performance. At the same time, 16QAM and 32QAM are poorly recognized, and the modulated signal constellations between the two modulations have a higher likeness to obtain bad performance. As SNR increases to 0 dB in Fig. 10 (b), The recognition effect of 4FSK and 8FSK have an effective increase, which are expressively improved by 49%, 48% in the recognition rate, respectively. At the SNR, the employed network can easily distinguish between 16QAM and 32QAM, and other modulation schemes can also implement the better recognition effect. As SNRs grow, the employed network acquires more signal characteristics and realizes ideal recognition effects.

In Fig. 11, the employed network compared with the typical solutions is as follow: LSTM [54] and GRU [55] are the familiar form of RNN; ResNet [56], DensenNet [57], SENet [58] are the sophisticated structure of deep residual neural network with more layer-by-layer connections; PnasNet [59] is the irregular structure generated by reinforcement learning method; Deep neural network (DNN) [29], extensible neural network (ENN) [32], Sparse Autoencoder [35] are in the light of the regular one-dimensional fully connected layer



(a) SNR = -4 dB



(b) SNR = 0 dB

FIGURE 10. Recognition results of different modulations.

to construct the network; the improved CNN [30], CLDNN (Convolutional Long Short-term Deep Neural Network) [60], fusion neural network (FNN) [34] are made up of CNN and RNN; K Nearest Neighbors [13], Random Forest [15], support vector machines (SVM) [14] are frequently used ML methods in AMR work. All networks are trained and verified at packets $= 2187$. Although the recognition effect of SENet, FNN and the improve CNN is around 2.7%, 4.1%

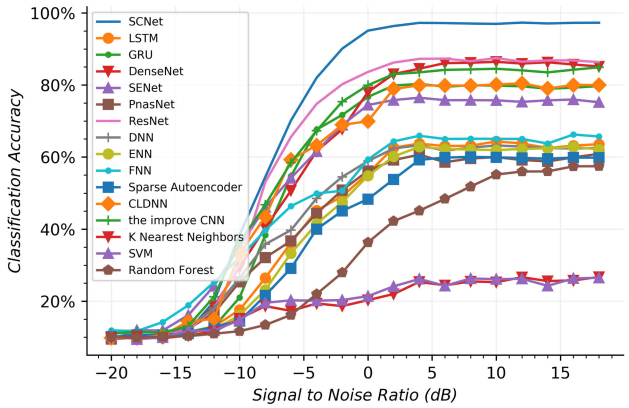


FIGURE 11. Recognition effects between different networks.

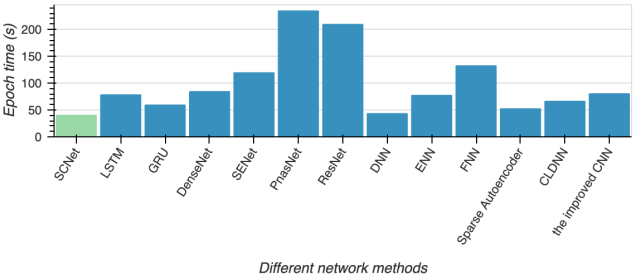


FIGURE 12. Training time of each epoch.

and 1.3% higher than SCNet from −20 dB to −11 dB, their maximal recognition rates are less than 25.5%, which hardly work at the low SNRs. The compared networks except SENet and FNN have almost the same recognition effect to SCNet from −20 dB to −11 dB. When SNRs are tightened up, the recognition effects continue to rise on all network methods, especially the effect is most significant in SCNet. From −11 dB to 0 dB, SCNet is adequately higher than other network methods in the recognition rate, which is more around 33.4%, 16.7%, 17.1%, 14.9%, 30.7%, 7.1%, 27.7%, 35.6%, 25.3%, 38.6%, 15.4%, 11.0%, 53.5%, 52.3% and 50.4% than LSTM, GRU, DenseNet, SENet, PnasNet, ResNet, DNN, ENN, FNN, Sparse Autoencoder, CLDNN, the improve CNN, K Nearest Neighbors, SVM and Random Forest on average, respectively. SCNet gains more unattractive signal characteristic due to the benefit of the network structure. There is a great improvement in the recognition effect of all networks after SNR = 0 dB, and SCNet also has a better performance than other five networks. LSTM, GRU, DenseNet, SENet, PnasNet, ResNet, DNN, ENN, FNN, Sparse Autoencoder, CLDNN, the improve CNN, K Nearest Neighbors, SVM and Random Forest approximately mean lower 19.1%, 17.4%, 11.6%, 21.3%, 37.4%, 10.3%, 34.5%, 35.0%, 31.9%, 38.1%, 17.3%, 13.0%, 71.8%, 71.5% and 44.9% than SCNet, respectively, which are less valid than SCNet. It is due to that SCNet has a greater ability to extract the plentiful recognition information with the variable CSK, which performs better than the compared network methods.

The trained network model is deployed to perform the AMR work, and it is necessary to consider the hardware resource limitations on the communication system. The smaller network model can run better in the limited hardware resources of communication terminals. The training time of the network model is one of the important factors that reflect the sensible design of the network structure. A more reasonable network structure can greatly shorten training time, improve training efficiency, and facilitate engineering applications. The conventional ML methods of K Nearest Neighbors, Random Forest and SVM do not find a function to predict all samples like DL methods, which control the error of the prediction function. For this reason, the parameter quantities and each epoch time are investigated in the DL methods. SCNet compares the parameter quantities with other networks in Table 1, which were acquired on CPU i5, GPU 2080ti, ubuntu 16.04 and tensorflow version 1.14. SCNet includes 1DSC, which has the easy way to decrease the parameter quantities. In the case, there are minimal values in SCNet, which is less than 1/2, 1/18, 1/15 and 1/29 of ENN, FNN, LSTM and GRU, respectively. The improve CNN, DenseNet, SENet, ResNet and PnasNet have the parameter quantities about 132 times, 207 times, 274 times, 468 times and 551 times than SCNet, respectively. The complex network has a larger number of parameters, meanwhile, the recognition effect is not as good as that of SCNet. There is little difference in the parameter quantities compared to SCNet, DNN, Sparse Autoencoder and CLDNN, and the total accurate rate of SCNet is more excellent than that of DNN, Sparse Autoencoder and CLDNN. In Fig. 12, Epoch time is the training time of each epoch. SCNet owns the shortest time, and DNN, Sparse Autoencoder, GRU, CLDNN, ENN, LSTM, the improved CNN, DenseNet, SENet, FNN, ResNet and PnasNet have around 1.1 times, 1.3 times, 1.5 times, 1.7 times, 1.9 times, 2.0 times, 2.0 times, 2.1 times, 2.9 times 3.3 times, 5.2 times and 5.8 times each epoch time of SCNet,

TABLE 1. The parameter quantities of different networks.

Network method	Parameter quantities
SCNet	153,930
LSTM	2,383,370
GRU	4,546,058
DenseNet	31,872,370
SENet	42,281,344
PnasNet	84,939,820
ResNet	71,990,730
DNN	171,150
ENN	345,210
FNN	2,833,051
Sparse Autoencoder	169,514
CLDNN	258,568
the improved CNN	20,326,666

respectively. Each epoch time concerns on the parameter quantities. The lower the parameter quantities are, the faster the training speed is, and the shorter each epoch time gets. Another major factor is that the CNN structure adopted by SCNet can perform the parallel manipulation, which is more efficient than LSTM, GRU and FNN that need to save intermediate states. The brief and efficient network structure designed in SCNet also shows better results than the complex network structure, including the improve CNN, ResNet, DenseNet, SENet and PnasNet, and it is also more efficient than the fully connected layer networks, such as DNN, ENN and Sparse Autoencoder. The employed network is remarkable on lower parameter quantities and shorter training time.

V. CONCLUSION

The paper considered the modulation recognition of underwater acoustic communication signals. It is hard to carry out high recognition accuracy in the tough condition of the underwater communication. The employed network with 1DSC and the variable CKR provides more advantages than the traditional neural network in recognition performance, which not only has the low parameter quantities but also has the short training time. It is also robust to the various signal packets. This shows that the trained network can be deployed and plugged in the actual underwater communication scenario with the restricted condition, which can also be afforded to the other signal recognition situation for the underwater communication. In the future, we will study the deep neural network in the orthogonal frequency division multiplexing (OFDM) environment to recognize the underwater communication signals effectively.

REFERENCES

[1] E. Demirors, G. Sklivanitis, T. Melodia, S. N. Batalama, and D. A. Pados, "Software-defined underwater acoustic networks: Toward a high-rate real-time reconfigurable modem," *IEEE Commun. Mag.*, vol. 53, no. 11, pp. 64–71, Nov. 2015.

[2] Y. A. Eldemerdash, O. A. Dobre, and M. Oner, "Signal identification for multiple-antenna wireless systems: Achievements and challenges," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1524–1551, 3rd Quart., 2016.

[3] G. Miao, N. Himayat, and G. Y. Li, "Energy-efficient link adaptation in frequency-selective channels," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 545–554, Feb. 2010.

[4] M. Stojanovic and J. Preisig, "Underwater acoustic communication channels: Propagation models and statistical characterization," *IEEE Commun. Mag.*, vol. 47, no. 1, pp. 84–89, Jan. 2009.

[5] A. C. Sing, J. K. Nelson, and S. S. Kozat, "Signal processing for underwater acoustic communications," *IEEE Commun. Mag.*, vol. 47, no. 1, pp. 90–96, Jan. 2009.

[6] B. Tomasi, L. Toni, P. Casari, L. Rossi, and M. Zorzi, "Performance study of variable-rate modulation for underwater communications based on experimental data," in *Proc. OCEANS MTS/IEEE SEATTLE*, Sep. 2010, pp. 1–8.

[7] Q. Shi and Y. Karasawa, "Noncoherent maximum likelihood classification of quadrature amplitude modulation constellations: Simplification, analysis, and extension," *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1312–1322, Apr. 2011.

[8] A. Abdi, O. A. Dobre, R. Choudhry, Y. Bar-Ness, and W. Su, "Modulation classification in fading channels using antenna arrays," in *Proc. IEEE MILCOM Mil. Commun. Conf.*, Oct. 2004, pp. 211–217.

[9] P. Panagiotou, A. Anastasopoulos, and A. Polydoros, "Likelihood ratio tests for modulation classification," in *Proc. MILCOM 21st Century Mil. Commun., Archit. Technol. Inf. Superiority*, Oct. 2000, pp. 670–674.

[10] V. G. Chavali and C. R. C. M. da Silva, "Maximum-likelihood classification of digital amplitude-phase modulated signals in flat fading non-Gaussian channels," *IEEE Trans. Commun.*, vol. 59, no. 8, pp. 2051–2056, Aug. 2011.

[11] S. Mihandoost and M. Chehel Amirani, "Automatic modulation classification using combination of wavelet transform and GARCH model," in *Proc. 8th Int. Symp. Telecommun. (IST)*, Sep. 2016, pp. 484–488.

[12] O. A. Dobre, M. Oner, S. Rajan, and R. Inkol, "Cyclostationarity-based robust algorithms for QAM signal identification," *IEEE Commun. Lett.*, vol. 16, no. 1, pp. 12–15, Jan. 2012.

[13] M. A. Hazar, N. Odabasioglu, T. Ensari, Y. Kavurucu, and O. F. Sayan, "Performance analysis and improvement of machine learning algorithms for automatic modulation recognition over Rayleigh fading channels," *Neural Comput. Appl.*, vol. 29, no. 9, pp. 351–360, May 2018.

[14] D. A. Amoedo, W. S. da Silva Junior, and E. B. de Lima Filho, "Parameter selection for SVM in automatic modulation classification of analog and digital signals," in *Proc. Int. Telecommun. Symp. (ITS)*, Aug. 2014, pp. 1–5.

[15] K. Triantafyllakis, M. Surligas, G. Vardakis, and S. Papadakis, "Phasma: An automatic modulation classification system based on random forest," in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw. (DySPAN)*, Mar. 2017, pp. 1–3.

[16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

[17] S. Zheng, S. Chen, L. Yang, J. Zhu, Z. Luo, J. Hu, and X. Yang, "Big data processing architecture for radio signals empowered by deep learning: Concept, experiment, applications and challenges," *IEEE Access*, vol. 6, pp. 55907–55922, 2018.

[18] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, and N. D. Sidiropoulos, "Learning to optimize: Training deep neural networks for interference management," *IEEE Trans. Signal Process.*, vol. 66, no. 20, pp. 5438–5453, Oct. 2018.

[19] P. V. R. Ferreira, R. Paffenroth, A. M. Wyglinski, T. M. Hackett, S. G. Bilen, R. C. Reinhart, and D. J. Mortensen, "Multi-objective reinforcement learning-based deep neural networks for cognitive space communications," in *Proc. Cognit. Commun. Aerosp. Appl. Workshop (CCAA)*, Jun. 2017, pp. 1–8.

[20] M. A. Wijaya, K. Fukawa, and H. Suzuki, "Intercell-interference cancellation and neural network transmit power optimization for MIMO channels," in *Proc. IEEE 82nd Veh. Technol. Conf. (VTC-Fall)*, Sep. 2015, pp. 1–5.

[21] O. Naparstek and K. Cohen, "Deep multi-user reinforcement learning for dynamic spectrum access in multichannel wireless networks," in *Proc. GLOBECOM - IEEE Global Commun. Conf.*, Dec. 2017, pp. 1–7.

[22] M. A. Wijaya, K. Fukawa, and H. Suzuki, "Neural network based transmit power control and interference cancellation for MIMO small cell networks," *IEICE Trans. Commun.*, vol. E99.B, no. 5, pp. 1157–1169, 2016.

[23] N. Samuel, T. Diskin, and A. Wiesel, "Deep MIMO detection," in *Proc. IEEE 18th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2017, pp. 1–5.

[24] X. Yan, F. Long, J. Wang, N. Fu, W. Ou, and B. Liu, "Signal detection of MIMO-OFDM system based on auto encoder and extreme learning machine," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 1602–1606.

[25] D. Neumann, T. Wiese, and W. Utschick, "Deep channel estimation," in *Proc. WSA 21th Int. ITG Workshop Smart Antennas*, 2017, pp. 1–6.

[26] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8549–8560, Sep. 2018.

[27] A. A. Diro and N. Chilamkurti, "Distributed attack detection scheme using deep learning approach for Internet of Things," *Future Gener. Comput. Syst.*, vol. 82, pp. 761–768, May 2018.

[28] M. Lopez-Martin, B. Carro, A. Sanchez-Esguevillas, and J. Lloret, "Conditional variational autoencoder for prediction and feature recovery applied to intrusion detection in IoT," *Sensors*, vol. 17, no. 9, p. 1967, Aug. 2017.

[29] J. Lee, J. Kim, B. Kim, D. Yoon, and J. Choi, "Robust automatic modulation classification technique for fading channels via deep neural network," *Entropy*, vol. 19, no. 9, p. 454, Aug. 2017.

[30] Y. Xu, D. Li, Z. Wang, Q. Guo, and W. Xiang, "A deep learning method based on convolutional neural network for automatic modulation classification of wireless signals," *Wireless Netw.*, vol. 25, no. 7, pp. 3735–3746, Oct. 2019.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[32] G. Qing Yang, "Modulation classification based on extensible neural networks," *Math. Problems Eng.*, vol. 2017, pp. 1–10, Oct. 2017.
[33] X. Liu, D. Yang, and A. El Gamal, "Deep neural network architectures for modulation classification," 2017, *arXiv:1712.00443*. [Online]. Available: http://arxiv.org/abs/1712.00443
[34] D. Zhang, W. Ding, B. Zhang, C. Xie, H. Li, C. Liu, and J. Han, "Automatic modulation classification based on deep learning for unmanned aerial vehicles," *Sensors*, vol. 18, no. 3, p. 924, Mar. 2018.
[35] H. Yang, S. Shen, J. Xiong, and X. Zhang, "Modulation recognition of underwater acoustic communication signals based on denoting & deep sparse autoencoder," in *Proc. Inter-Noise Noise-Con Congr. Conf.*, vol. 253, no. 3, 2016, pp. 5506–5511.
[36] C. LI, Q. Zhou, X. Han, J. Yin, and M. Shao, "Underwater non-cooperative communication signal recognition with deep learning," *The J. Acoust. Soc. Amer.*, vol. 142, no. 4, p. 2732, 2017.
[37] D. Li-Da, W. Shi-Lian, and Z. Wei, "Modulation classification of underwater acoustic communication signals based on deep learning," in *Proc. OCEANS MTS/IEEE Kobe Techno-Oceans (OTO)*, May 2018, pp. 1–4.
[38] B. Li, S. Zhou, M. Stojanovic, L. Freitag, and P. Willett, "Multicarrier communication over underwater acoustic channels with nonuniform Doppler shifts," *IEEE J. Ocean. Eng.*, vol. 33, no. 2, pp. 198–209, Apr. 2008.
[39] H. Lim, J. Park, and Y. Han, "Rare sound event detection using 1D convolutional recurrent neural networks," in *Proc. Detection Classification Acoustic Scenes Events Workshop*, Nov. 2017, pp. 80–84.
[40] I. Lecomte, P. L. Lavadera, C. Botter, I. Anell, S. J. Buckley, C. H. Eide, A. Grippa, V. Mascolo, and S. Kjoberg, "2(3)D convolution modelling of complex geological targets beyond—1D convolution," *First Break*, vol. 34, no. 5, pp. 99–107, May 2016.
[41] A. Fred Agarap, "Deep learning using rectified linear units (ReLU)," 2018, *arXiv:1803.08375*. [Online]. Available: http://arxiv.org/abs/1803.08375
[42] Y. Hu, A. Huber, J. Anumula, and S.-C. Liu, "Overcoming the vanishing gradient problem in plain recurrent networks," 2018, *arXiv:1801.06105*. [Online]. Available: http://arxiv.org/abs/1801.06105
[43] R. Pascanu, T. Mikolov, and Y. Bengio, "Understanding the exploding gradient problem," 2012, *arXiv:1211.5063*. [Online]. Available: https://arxiv.org/abs/1211.5063
[44] M. Cheung, J. Shi, O. Wright, L. Y. Jiang, X. Liu, and J. M. F. Moura, "Graph signal processing and deep learning: Convolution, pooling, and topology," *IEEE Signal Process. Mag.*, vol. 37, no. 6, pp. 139–149, Nov. 2020.
[45] Y. Zhang, T. Wu, Y. Zakharov, and J. Li, "MMP-DCD-CV based sparse channel estimation algorithm for underwater acoustic transform domain communication system," *Appl. Acoust.*, vol. 154, pp. 43–52, Nov. 2019.
[46] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980
[47] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proc. COMPSTAT*. Springer, 2010, pp. 177–186.
[48] M. D. Zeiler, "ADADELTA: An adaptive learning rate method," 2012, *arXiv:1212.5701*. [Online]. Available: http://arxiv.org/abs/1212.5701
[49] T. Kurbiel and S. Khaleghian, "Training of deep neural networks based on distance measures using RMSProp," 2017, *arXiv:1708.01911*. [Online]. Available: http://arxiv.org/abs/1708.01911
[50] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.
[51] D. H. Murphree and C. Ngufor, "Transfer learning for melanoma detection: Participation in ISIC 2017 skin lesion classification challenge," 2017, *arXiv:1703.05235*. [Online]. Available: http://arxiv.org/abs/1703.05235
[52] S. Chambon, M. N. Galtier, P. J. Arnal, G. Wainrib, and A. Gramfort, "A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 4, pp. 758–769, Apr. 2018.
[53] L. Rice, E. Wong, and J. Z. Kolter, "Overfitting in adversarially robust deep learning," 2020, *arXiv:2002.11569*. [Online]. Available: http://arxiv.org/abs/2002.11569
[54] 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 Trans. Cognit. Commun. Netw.*, vol. 4, no. 3, pp. 433–445, Sep. 2018.
[55] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1539–1548, Feb. 2018.
[56] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 168–179, Feb. 2018.
[57] R. Deng and S. Liu, "Relative depth order estimation using multi-scale densely connected convolutional networks," *IEEE Access*, vol. 7, pp. 38630–38643, 2019.
[58] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
[59] C. Liu, B. Zoph, M. Neumann, J. Shlens, W. Hua, L.-J. Li, L. Fei-Fei, A. Yuille, J. Huang, and K. Murphy, "Progressive neural architecture search," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 19–34.
[60] X. Liu, D. Yang, and A. E. Gamal, "Deep neural network architectures for modulation classification," in *Proc. 51st Asilomar Conf. Signals, Syst., Comput.*, Oct. 2017, pp. 915–919.

YAN WANG was born in Shandong, China, in 1982. He received the B.S. degree from the College of Information Science and Engineering, Jinan University, Jinan, in 2004, and the M.S. degree from the College of Information Science and Engineering, Ocean University of China, Qingdao, in 2008, where he is currently pursuing the Ph.D. degree in intelligent information and communication system. He is also a Lecturer with the School of Physics and Electronic Engineering, Taishan University. His research interests include wireless communication systems, underwater acoustic communication systems, and UWB systems.

YIHENG JIN was born in Shandong, China, in 1992. She received the B.S. and Ph.D. degrees from the College of Information Science and Engineering, Ocean University of China, Qingdao, in 2013 and 2019, respectively. She is currently a Lecturer with the School of Science and Information Science, Qingdao Agricultural University. Her research interests include wireless communication systems, BeiDou navigation satellite systems, array signal processing, and UWB systems.

HAO ZHANG (Senior Member, IEEE) was born in Jiangsu, China, in 1975. He received the B.S. degree in telecom engineering and industrial management from Shanghai Jiao Tong University, China, in 1994, the M.B.A. degree from the New York Institute of Technology, New York, NY, USA, in 2001, and the Ph.D. degree in electrical engineering from the University of Victoria, Victoria, BC, Canada, in 2004. From 1994 to 1997, he was an Assistant President of ICO Global Communications Company, China. He is currently a Professor with the Department of Electrical Engineering, Ocean University of China. He is also an Adjunct Professor with the University of Victoria. His research interests include wireless signal recognition, cooperative networks communications, ultra-wideband systems, MIMO wireless systems, and spread spectrum communications.

QIAN LU received the Ph.D. degree from the Ocean University of China. She is currently an Assistant Professor with the Department of Computer Science and Technology, Qingdao University. Her research interests include network security, wireless communication, and cyber physical systems.

CONGHUI CAO was born in Shanxi, China, in 1992. She received the B.S. degree in communication engineering from Jianghan University, Wuhan, China, in 2014, and the Ph.D. degree in intelligent information and communication system from the Ocean University of China, Qingdao, China, in 2019. She is currently a Lecturer with Jianghan University. Her research interests include wireless signal recognition and cooperative network communication.

ZHANLIANG SANG was born in Shandong, China, in 1978. He received the B.S. degree from the College of Electrical and Electronic Engineering, Shandong University of Technology, Zibo, in 2001, and the M.S. degree from the College of Information Science and Engineering, Ocean University of China, Qingdao, China, in 2008. He is currently a Senior Engineer with CRRC Qingdao Sifang Company Ltd., Qingdao. His main research interests include edge computing, FPGA, and embedded systems.

MEI SUN was born in Hubei, China, in 1982. She received the B.S. degree in physics from Hubei Normal University, Huangshi, China, in 2003, and the M.S. and Ph.D. degrees in acoustics from the Institute of Acoustics, Chinese Academy of Sciences, Beijing, China, in 2008 and 2011, respectively. She is currently an Associate Professor with Taishan University, Tai'an, China. Her main research interest includes the vector sound field properties in the ocean and its applications.