

Research Article

Research on Distributed Multisensor Spectrum Semantic Sensing and Recognition in Internet of Things Environment

Xiguo Liu ^{1,2}, Jing Zhang,³ Min Liu,² Zhongyang Mao ^{1,2} and Changbo Hou³

¹Aviation Communication Teaching and Research Section, Naval Aeronautic University, Yantai 264001, China

²State Key Laboratory of Signal and Information Processing of Shandong Province, Yantai 264001, China

³College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China

Correspondence should be addressed to Zhongyang Mao; freedom_mzy@163.com

Received 21 January 2022; Accepted 23 March 2022; Published 20 May 2022

Academic Editor: Xingsi Xue

Copyright © 2022 Xiguo Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The development of 5G technology has brought about a new era of Internet of Thing (IoT), and at the same time, electromagnetic spectrum monitoring and sensing have also ushered in huge challenges. Digital modulation recognition technology is an important content of electromagnetic spectrum sensing. In the increasingly complex wireless communication transmission environment, especially in noncooperative communication, it becomes more and more difficult to receive target signals and accurately extract effective semantic information from diverse modulation signals of the electromagnetic spectrum. At this stage, with the rapid development of network information and wireless communication technology, within a prescribed distance, the IoT built by many sensors has attracted wide attention from people in related fields. This paper proposes a distributed collaborative sensing spectrum semantic recognition architecture for communication signals based on feature fusion. Perform wireless communication and transmission between multiple sensors to form a self-organizing network to cooperatively sense signal semantic information, and extract the signal features of each sensor in the distributed network structure. Finally, the extracted sensor features are semantically analyzed and modeled, and the effective features are fused to complete the entire perception and recognition process. Even if the channel environment of a small number of receiving nodes deteriorates in a complex transmission environment, the signal quality features can still be accurately extracted, the classification and recognition effect like or higher than the best channel state performance can be achieved, and the fault tolerance of the system can be effectively improved. It can also enhance the performance of spectral semantic information sensing and recognition in the IoT environment.

1. Introduction

The rapid development of technology in the communication field has opened a new era of IoT with the emergence of 5G technology, followed by richer application scenarios and increasingly complex electromagnetic environments, which has brought spectrum monitoring management and electromagnetic spectrum sensing utilization huge challenge. In the complex electromagnetic environment where everything is interconnected, accurate perception of semantic spectrum information and identification of signal modulation methods can provide important information for communication networking, etc., thereby effectively improving spectrum utilization efficiency. Typical spectrum semantic

sensing and recognition (SMSR) methods are mainly divided into traditional likelihood function-based decision-making [1, 2], feature extraction-based pattern recognition [3, 4], and other methods, as well as deep learning (DL) methods that have emerged recently [5]. Traditional methods mainly extract specific features manually, and the recognition effect largely depends on manual experience, which leads to poor recognition performance and fewer recognition types. With the advent of AlphaGo in 2015, more and more researchers are focusing on DL. It has achieved good results in classification tasks with its outstanding feature extraction capabilities. O'shea and West [6] built a simulated communication model through GNU Radio and collected 11 communication signals, using a Convolutional Neural Network (CNN) to

extract signal features from In-phase and Quadrature (I/Q) components. The DL method has no expert-derived features, showing a great accuracy improvement over traditional statistical methods. Wu et al. [7] improved the CNN and added a Long and Short-Term Memory (LSTM) structure, which improved the network's feature extraction ability for signal timing and increased the signal recognition accuracy to 80%. Wu et al. [8] combined the cyclic spectrogram and constellation diagram and simulated it on the public dataset. When the signal-to-noise ratio (SNR) is 0 dB, the accuracy rate reaches 80% and the training time is shortened. In the next few years, researchers used more complex DL architectures [9–13], using extracted feature inputs and neural network pruning to improve operating efficiency. With the increasingly complex electromagnetic environment of signals, DL has gradually become the mainstream algorithm in spectrum semantic sensing (SMS) and modulation recognition algorithms relying on powerful feature extraction capabilities and robustness. Although communication signal modulation recognition technology has gradually matured and the results have become more abundant [14], with the rapid development of wireless communication technology, signal transmission scenarios have become increasingly diversified, and application requirements have become increasingly updated [15], all of which promote the improvement of modulation methods. Therefore, modulation recognition technology always needs to be constantly updated according to changes in application scenarios and application requirements.

In the actual wireless communication environment, the single-node SMS technology is easily affected by factors such as multipath effects, hidden terminals, and path loss and cannot obtain correct sensing results. At this stage, with the rapid development of network information and wireless communication technology, within a prescribed distance, the IoT built by a large number of nodes has attracted wide attention from people in related fields, and the number of network devices and sensors deployed in the physical environment is rapidly increasing. The increase also brings new challenges to the wireless SMS and recognition, and the research on the distributed network architecture [16] of the combination of multiple receivers arises at the historic moment. Distributed multisensor node wireless SMSR technology can be divided into data layer-based fusion, feature layer-based fusion, and decision-making layer-based fusion schemes. Zhang et al. [17] proposed that the automatic modulation recognition scheme based on multisensor signal fusion can provide higher reliability than single-sensor signals. Dulek [18] proposed a classifier based on online and distributed expectation maximization, which can achieve a classification and recognition effect similar to the best channel state performance. Distributed recognition technology is widely used in optical fiber vibration sensing recognition [19, 20]; Sun et al. [21] developed an improved deep learning method based on a serial fusion feature extraction model for an optical fiber distributed vibration sensing system which can automatically extract and identify effective features. Distributed fusion schemes based on the feature layer mostly use artificial features to achieve [22–24], but in noncoopera-

tive communication scenarios, the received signal is usually a weak signal, which makes it difficult to obtain accurate feature expression. Although the fusion scheme based on the data layer can enhance the received signal strength to a certain extent, it is often necessary to perform centralized calculation and processing on the data of each node in the fusion center, which causes the fusion center to be overloaded. The distributed recognition architecture based on the decision-making layer needs to clarify the influence factors of each node on the final decision. Although the decision result can improve the recognition performance, it needs to know the prior information such as the SNR of the signal at the receiving end of each receiver node. It is not conducive to signal recognition in noncooperative communication scenarios. Therefore, in order to improve the performance of spectrum semantic perception and recognition in complex electromagnetic environments such as noncooperative communication scenarios, this study uses the outstanding feature extraction capabilities of DL methods to propose a distributed collaborative recognition scheme based on feature fusion. The main contributions of this study are as follows:

- (1) Build a distributed multisensor signal reception scene, and set up a transmitter and multiple receiver nodes for signal reception. Model the spectral semantic information in distributed scenarios, and simulate the transmission of communication signals in different state channels through simulation experiments. The specific scene settings are introduced in the next section. In addition, the method proposed in this paper can to a certain extent solve the problem of inability to perform accurate SMSR recognition modulation recognition in noncooperative communication confrontation scenarios due to weak received signals
- (2) The use of DL algorithms is to realize the feature semantic information extraction of multireceiver node signals, the more accurate feature expression can be obtained by fusion of multinode features' semantic information, and the dimensionality reduction of the fused features is performed through the classifier to complete the distributed and coordinated communication signal recognition. It can eliminate the uncertainty of communication signal recognition caused by poor channel conditions to a certain extent

2. Materials and Methods

Under the IoT, the rapid development and comprehensive use of digital communication technology have brought huge challenges to the task of electromagnetic spectrum sensing. Recognition of communication signals, as one of the key technologies for electromagnetic spectrum monitoring, is of great significance in both military and civilian fields. However, most of the current researches are limited to a single node. Channel conditions and received signal strength will directly affect recognition performance. A single node

is affected by environmental changes and has poor fault tolerance performance, which leads to its signal recognition effect; when the channel environment is poor, the effect will be poor. With the development of wireless sensor networks, spectral semantic sensing, signal estimation, and recognition algorithms have received more and more attention. Distributed deployment of multiple sensors in the monitoring area forms a wireless communication-based self-organizing network system, which can realize multisensor collaboration to sense the sensing objects in the detection area and finally send the collected semantic information to the control center for further processing. The use of multinode data fusion, feature fusion, and other methods can greatly eliminate the ambiguity of unknown signals. When the channel conditions of a small number of sensor nodes deteriorate, the recognition probability can still be finally maintained. In this study, the communication signal features extracted by each sensor node are fused, and the fused features' semantic information is used to identify the signal modulation mode, to complete the identification process of the entire distributed algorithm. Figure 1 shows the distributed cooperative signal sensing and recognition framework based on feature fusion in this study, which can be divided into three modules, the distributed signal receiving module, the CLDNN feature semantic extraction module, and the fusion classification and recognition module.

The specific process is shown in Algorithm 1

2.1. Distributed Signal Reception. Through the research and development of networking information systems and autonomous sensing and intelligent information equipment, the electronic equipment system is developing towards decentralization, networking, and distributed coordination, which greatly improves the level of electronic warfare. The network communication system based on the distributed concept is promoting the development of combat intelligence. The distributed scenario of this study is shown in Figure 2.

In the distributed scenario built in this paper, there are one transmitter and multiple receivers. The number of receivers needs to be determined according to the actual scene requirements, and different numbers of receivers often affect the final recognition performance. The transmitting end signal transmits the communication signal to each receiving end through different channels. After the signal features of each receiver are extracted, the features semantic information of different receivers can be analyzed and fused in the fusion classification center.

In the signal receiving module, the receiver converts the received signal into a baseband modulated signal through digital downconversion and other processing. The signal model received by each signal in the AWAG channel can be expressed as

$$x_k(n) = h_k s(n) + \text{noise}, \quad n = 1, 2, \dots, N. \quad (1)$$

Among them, $s(n)$ is the signal sequence sent by the transmitter, h_k represents the channel coefficient, $x_k(n)$ is the signal at the receiving end, and noise is Gaussian white noise. However, due to the influence of distance and other

factors in actual signal transmission, the signals of different receiver nodes have different delays and other factors. It can be further expressed as

$$x_k(n) = h_k s(n - D_k) + \text{noise}, \quad n = 1, 2, \dots, N, \quad (2)$$

where D_k represents the channel transmission delay of the k th receiver. The received signal quality of different receivers mainly depends on h_k and D_k .

For the popular digital receivers on the market, especially the software radio platform (SDR), the received communication signal is often a baseband I/Q complex sequence. Therefore, it is very necessary to start with the baseband I/Q data to perform modulation recognition on the signal. In this study, a vector is used to represent the received complex signal sequence with noise, and the signal model received by the k th receiver can be expressed as

$$X_k = [x(1), x(2), \dots, x(n)], \quad n = 1, 2, \dots, N. \quad (3)$$

Feature extraction is performed on the received signals of each receiver node and then sent to the fusion classification center for fusion analysis and modeling of feature semantic information to further complete the recognition.

2.2. Feature Extraction and Analysis. In recent years, DL methods have stood out among many machine learning methods by their neural network architecture, algorithms, and optimization technologies. They have been widely used in machine vision and speech recognition and have achieved a series of breakthroughs. DL is a method that effectively uses a data-driven approach to extract features and accurately identify it. Compared with manual feature design and extraction, DL algorithms can effectively extract the shallow features and implicit features of the data, while also saving time. The natural attributes of big data in the communication field have led scholars to explore the possibility of applying DL to the communication field, such as the use of deep neural networks for modulation recognition and radar waveform recognition. Therefore, this study will also use the method based on DL for the feature extraction of the distributed collaborative recognition of communication signals.

CNN is mostly used in the field of image processing. In recent years, they have also been widely used in the field of communication for automatic modulation recognition of signals. The convolutional layer extracts the local features of the data through the convolution kernel; the LSTM network is a special recurrent neural network designed to avoid long-term dependence problems. The difference between the deep neural network (DNN) and recurrent neural network (RNN) and CNN is that DNN specifically refers to a fully connected neuron structure and does not include convolutional units or temporal associations. It can convert the extracted features into a feature space making the output easier to classify.

CNN is good at reducing frequency changes, LSTM is good at time modeling, and DNN is suitable for mapping features to more separable spaces. Therefore, CNN, LSTM,

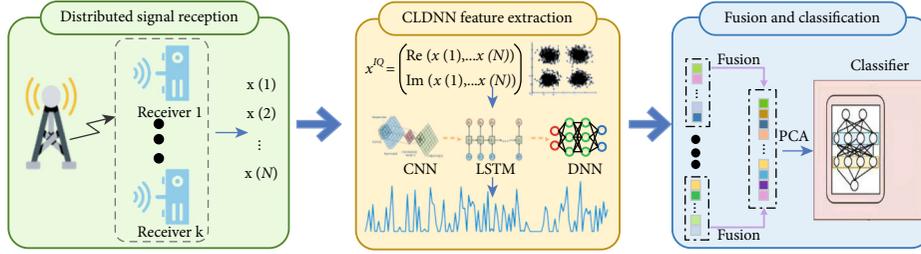


FIGURE 1: Distributed cooperative signal recognition framework based on feature fusion. This framework includes a distributed signal reception module, a CLDNN feature extraction module, and a feature fusion and classification module.

Multireceiver distributed collaborative identification process

Step 1: Signal reception: the data received by each receiver is a baseband modulated complex sequence of length N , $X_k = [x(1), x(2), \dots, x(N)]$;

Step 2: Signal preprocessing: extract the real and imaginary parts of the complex baseband signal sequence as I and Q channels:

$$X_k = \text{Re}(X_k) + j \cdot \text{Im}(X_k) = I_k + Q_k;$$

$$\text{Will be stored as a two-dimensional matrix: } X_k^{IQ} = \begin{pmatrix} \text{Re}[x(1), \dots, x(N)] \\ \text{Im}[x(1), \dots, x(N)] \end{pmatrix};$$

Step 3: Feature extraction: input the processed data of each node into the feature extraction network for training, map the data to the high-dimensional feature space, and extract the feature vector after training;

Step 4: Feature fusion: fusion of the feature vectors of different receivers in step 3, dimensionality reduction processing of high-dimensional features, retaining the main features with differences, removing redundant features, and obtaining the fused feature vector;

Step 5: Classification output: send the fused feature vector to the classifier for classification and recognition, and output the recognition result.

ALGORITHM 1: Multireceiver distributed collaborative identification process.

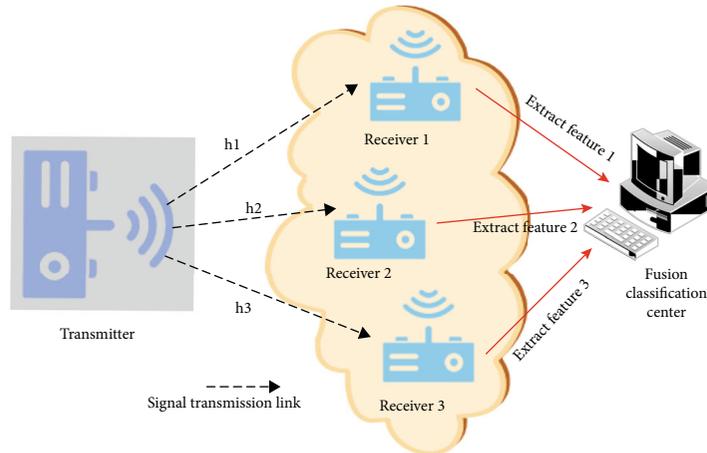


FIGURE 2: Distributed signal transmission. This scene includes one signal transmitter and multiple signal receivers (the number of receivers can be set according to the actual scene; this figure sets three for illustration). Parameter h represents the transmission parameters in each channel.

and DNN are complementary in modeling capabilities. Sainath et al. [25] use the complementarity of CNN, LSTM, and DNN to build a CLDNN network model for speech signal recognition, which has an improved effect compared to the three models used alone. There is a natural similarity between speech signal and communication signal, or it can be said that speech signal is a kind of communication signal.

In terms of data representation, it is a discrete correlation sequence in the time domain, but the digital modulation signal data is included in natural language processing. In addition to the same information that it carries, the more important thing is its modulation information. Different from natural language processing, its modulation information is only related to the current symbol and a few adjacent

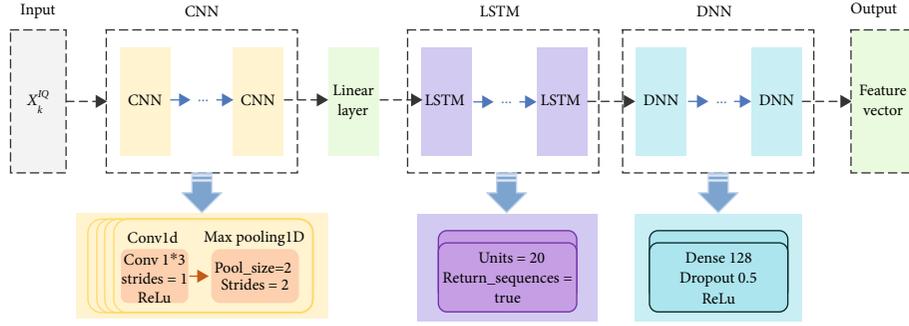


FIGURE 3: CLDNN feature extraction structure. The network contains 7 CNN layers, two LSTM layers, and two DNN layers.

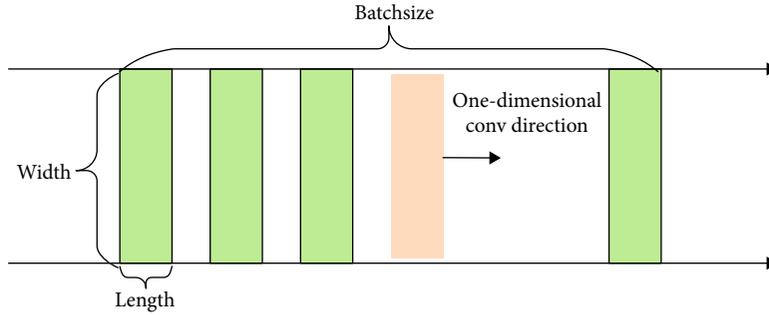


FIGURE 4: One-dimensional convolution diagram.

TABLE 1: LSTM and DNN network parameter settings.

Layers	Input: CNN feature ($\mathbb{R}^{M \times 512}$) Kernel parameters	Output shape
LSTM	Units = 20	(None, M , 20)
LSTM	Units = 20	(None, 20)
Dense+ReLU	128	(None, 128)
Dropout	0.5	(None, 128)
Dense+ReLU	128	(None, 128)
Dropout	0.5	(None, 128)

Output: CLDNN feature (Dimension = 128)

TABLE 2: Simulation parameter setting.

Dataset parameter setting	
Modulation	8 classes (BPSK, QPSK, 8PSK, 16PSK, 16QAM, 64QAM, 256QAM, PAM4)
Sample rate and length	100 kHz and 1024
SNR (dB) range	-20 : 2 : 18
Training dataset	$6400 \times 1024 \times 2$ (6400 instances)
Test dataset	$1600 \times 1024 \times 2$ (1600 instances)

symbols, not a real-time sequence. With obvious timing features, this paper builds a CLDNN network model that is more suitable for signal modulation recognition based on [21] for feature extraction. Its structure is shown in Figure 3.

In traditional feature extraction methods, only one channel of I/Q data is usually used for processing. In order to fully extract the subtle features' semantic information of the received signal, this study converts the received complex number sequence into I/Q two channels of data:

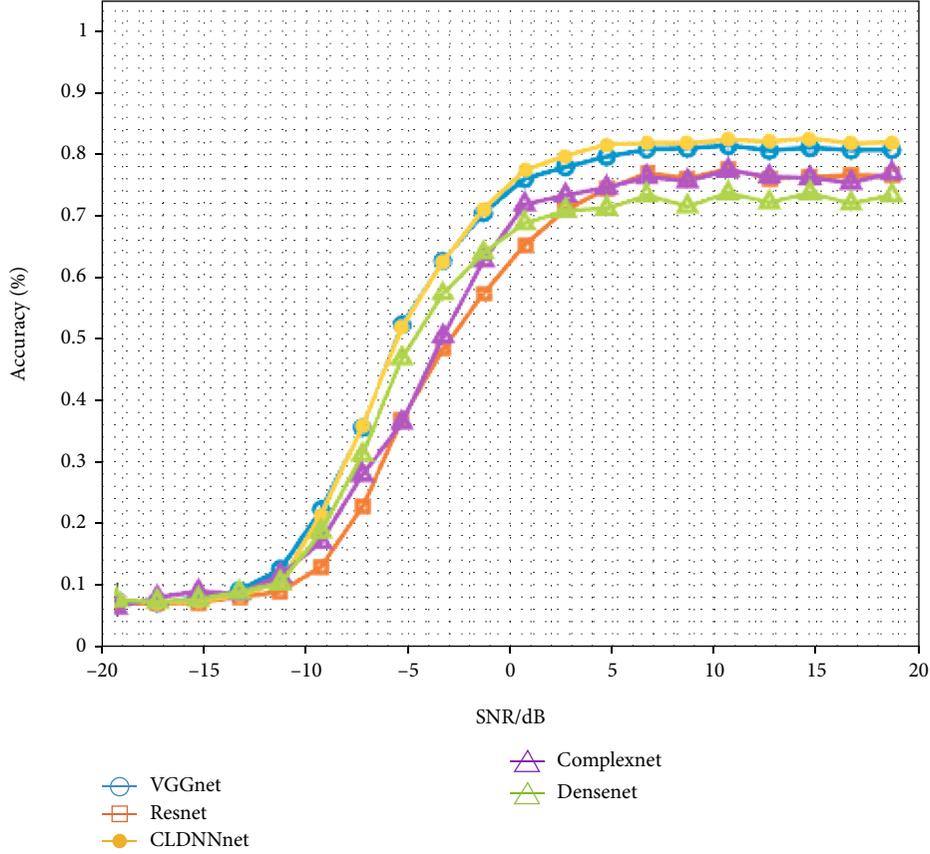
$$X_k = \text{Re}(X_k) + j \cdot \text{Im}(X_k) = I_k + Q_k, \quad (4)$$

where I_k is the real part of the complex modulated signal and Q_k is the imaginary part. The I/Q two-way data is stored as a two-dimensional matrix as the input of the feature extraction network, that is, $\text{Input} = [N \times 2]$, where N is the

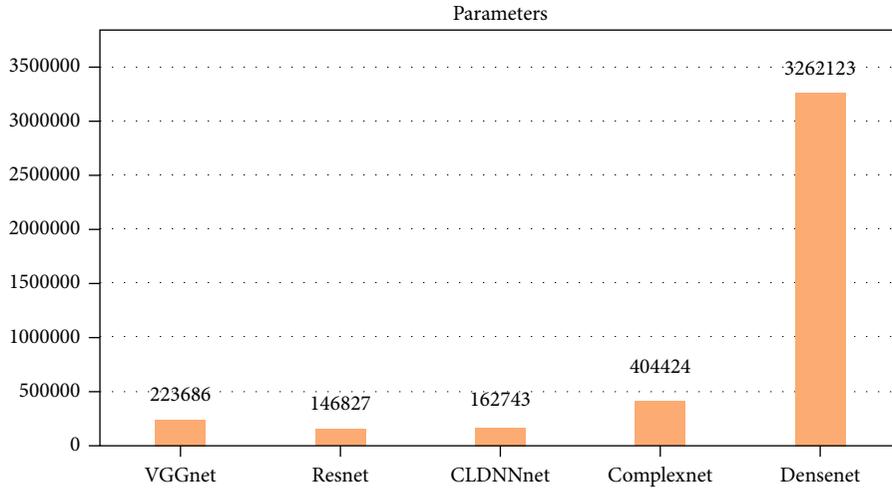
signal length, as shown in

$$X_k^{IQ} = \begin{pmatrix} \text{Re}[x(1), \dots, x(N)] \\ \text{Im}[x(1), \dots, x(N)] \end{pmatrix}. \quad (5)$$

Input the two-dimensional matrix sequence of the digital modulation signal into the CNN network, and complete the feature extraction and dimensionality reduction at this stage after convolution and pooling operations. To adapt to the features of the input I/Q sequence, this study uses the one-dimensional convolution kernel commonly used in sequences to replace the traditional two-dimensional convolution kernel to extract features, as shown in Figure 4, and the expression of the one-dimensional convolution kernel



(a) The recognition accuracy of different models



(b) Comparison of parameters of different models

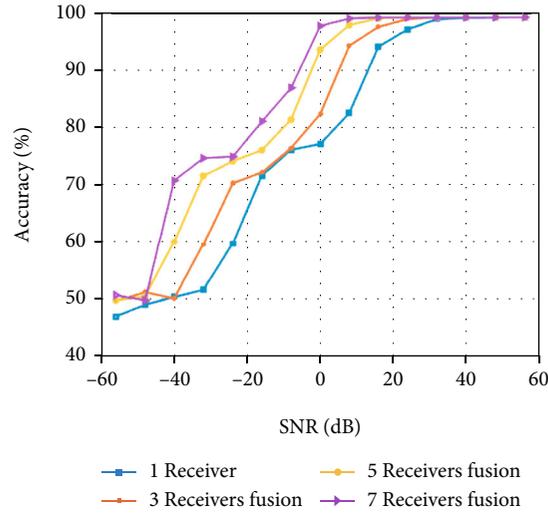
FIGURE 5: Comparison of classification performance of different network models. The experimental results are based on the RadioML2016.10a dataset.

is shown in

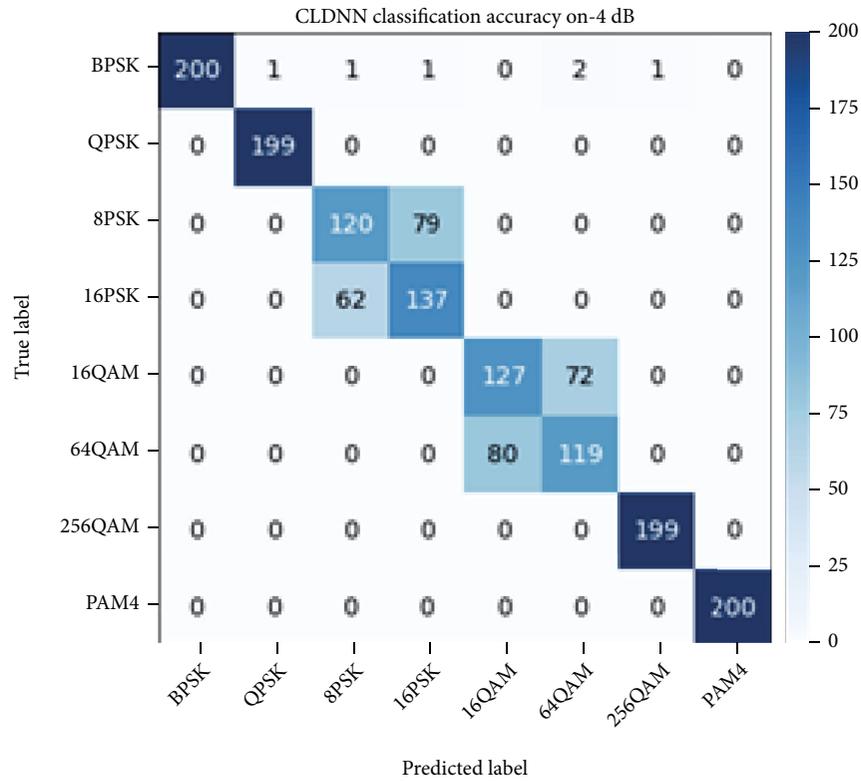
$$y_j^l = f\left(\sum_{i=1}^{M_j} \omega_{i,j} \times x_i^{l-1} + b_j^l\right), \quad j = 1, 2, \dots, N, \quad (6)$$

where y_j^l is the first feature of the layer, $\omega_{i,j}$ is the weight

value of the j feature of the i layer, b_j^l is the offset of the j feature in the l layer, N is the number of feature maps in the l layer, M represents the size of the one-dimensional convolution kernel, $f(\cdot)$ represents the activation function, and $f(\cdot) = \max(0, \cdot)$. In this study, the main features of the modulated signal are extracted by 7 layers of convolution kernels with a size of 1×3 . The number of convolution kernels are 64, 128, 128, 256, 256, and 512, respectively.



(a) Recognition results under different SNRs



(b) 7 receivers cooperatively identify confusion matrix under -4 dB

FIGURE 6: Feature fusion recognition performance of different numbers of receivers: (a) the accuracy change curve of a single receiver and different numbers of receivers under cooperative recognition at -14-14 dB and (b) the effect of cooperative recognition by 7 receivers under -4 dB.

To reduce the size of the model and increase the calculation speed, this study uses a one-dimensional maximum pooling method with a step size of 2 to downsample the convolutional feature map. It should be referred to as

$$y_j^l = f\left(\text{down}\left(y_j^{l-1}\right)\right). \quad (7)$$

Among them, y_j^l and y_j^{l-1} represent the feature map of the l and $l-1$ layer, respectively, and $\text{down}(\cdot)$ represents the downsampling.

After the input data vector $N \times 2$ is convolved and pooled, the feature space at this time can be expressed as $\mathbb{R}^{M \times 512}$. The long input information is converted into a shorter high-level feature sequence as the input of the LSTM

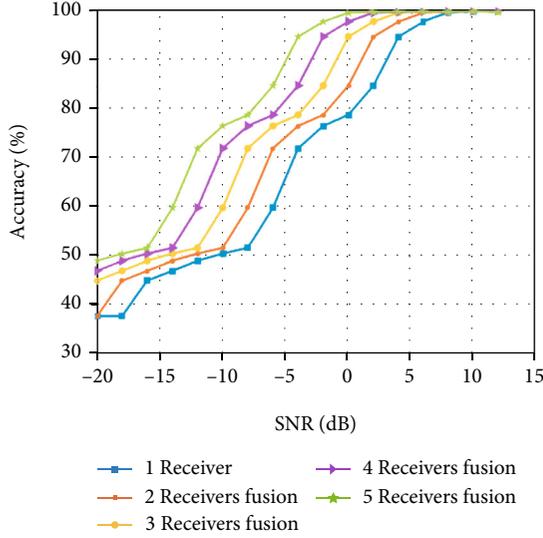


FIGURE 7: Recognition results under different SNRs (take the SNR of receiver 1 as the abscissa).

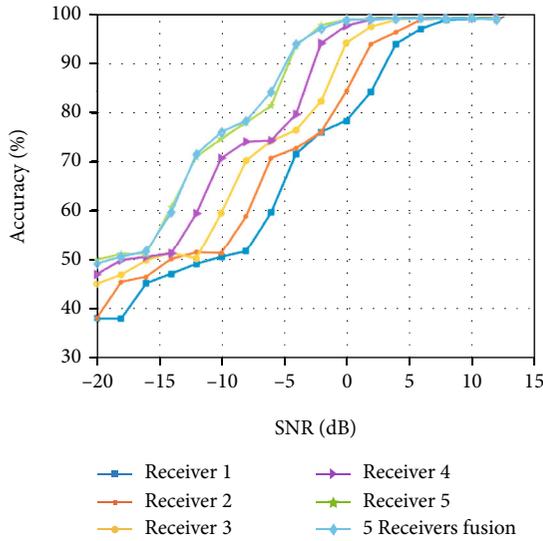


FIGURE 8: Comparison of collaborative recognition effect between different receivers and multireceivers.

module, using its features to learn the features of several adjacent symbols and input into the DNN network module, and the extracted features are mapped to a feature space that is easier to separate, and the final feature vector is obtained. The LSTM and DNN network parameter settings are shown in Table 1.

2.3. Fusion and Classification. The previous section introduced how to extract the features of the digital modulation signals of different receiver nodes. In order to make full use of the semantic information of different receivers in the communication network and further improve the accuracy of signal recognition, next, this section will introduce how to fuse feature semantics and collaborative identification of data from different receiver nodes.

Feature-level fusion refers to a fusion method that is completed by integrating or combining features from all nodes. Its purpose is to use the complementarity of each single node semantic information to synthesize the extracted features into a feature that is more discriminative than the input feature.

This study uses the aforementioned method to complete the feature extraction of the received data and finally extracts 128-dimensional features for each receiver node's data as the feature vector of the signal sequence of the node, and the resulting fusion feature vector is as follows:

$$F_{\text{fusion}} = F_1(X_1^{IQ}) \oplus F_2(X_2^{IQ}) \oplus \dots \oplus F_k(X_k^{IQ}). \quad (8)$$

Among them, F_{fusion} represents the feature vector after fusion, the size of each sample is 1×128 , $F_k(\cdot)$ represents the data feature vector of each receiver obtained using the feature extraction method in this study, and \oplus represents the fusion operation on the feature vector of each node, and the size of the fused feature vector is $1 \times (128 * k)$.

After extracting effective feature vectors from single-node receiver data and analyzing them, perform feature vector fusion on multinode feature data. The obtained feature fusion vector can reduce the influence of channel quality and signal strength on the extracted features and can represent more modulation information about the signal than the vector extracted by a single node. But at the same time, multiple nodes also increase the complexity of problem analysis. Therefore, it is necessary to process the fusion features, so as to reduce the feature parameters while retaining the effective features to the greatest extent and complete the comprehensive analysis of the feature data. In summary, it is necessary to reanalyze the closely related feature parameters, eliminate redundant feature quantities, and finally realize the information contained in each feature with fewer comprehensive feature parameters. This study uses the same network model for feature extraction on different receiver node data, so the Principal Component Analysis (PCA) algorithm, which is often used in high-dimensional vector analysis, is used to process the fused feature data.

Assuming that the number of receiver nodes is k , the feature of each node has a dimension of m , the data of each node can be expressed as X^1, X^2, \dots, X^k , the data feature of each node can be expressed as $X^k = [x_1^k, x_2^k, \dots, x_m^k]^T$, and the error in the sample mapping process can be expressed as

$$\text{error} = \frac{1}{k} \sum_{i=1}^k \|X^i - X_{\text{map}}^i\|^2. \quad (9)$$

Among them, X_{map}^i represents the new feature after mapping, and the dimension remains unchanged. The process of PCA is as follows.

(1) Feature normalization

Normalize the training samples to obtain the training parameters and then normalize the test samples.

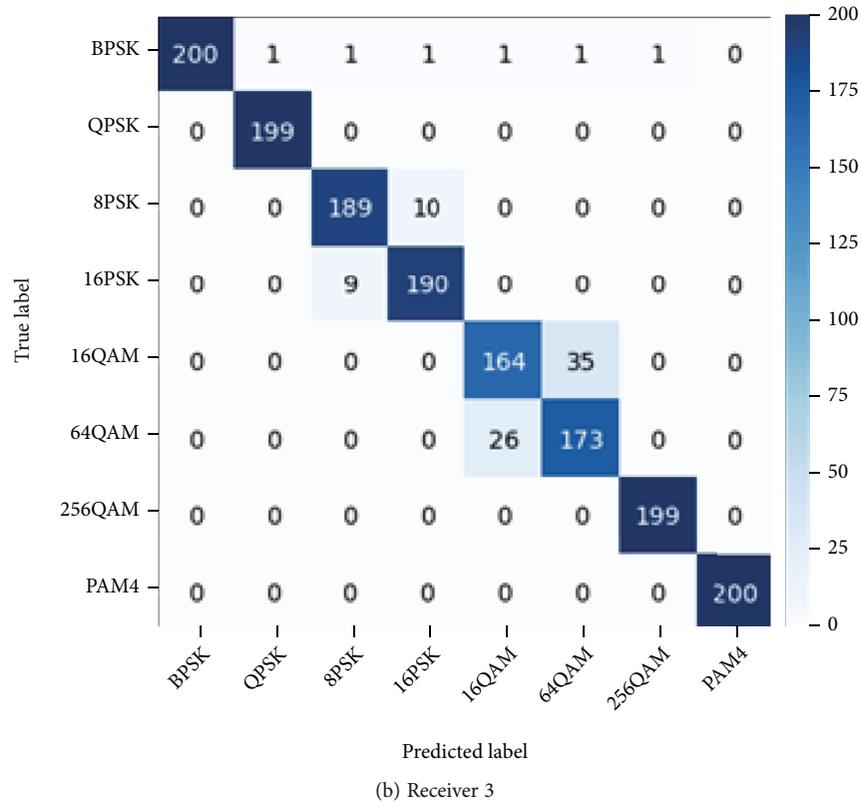
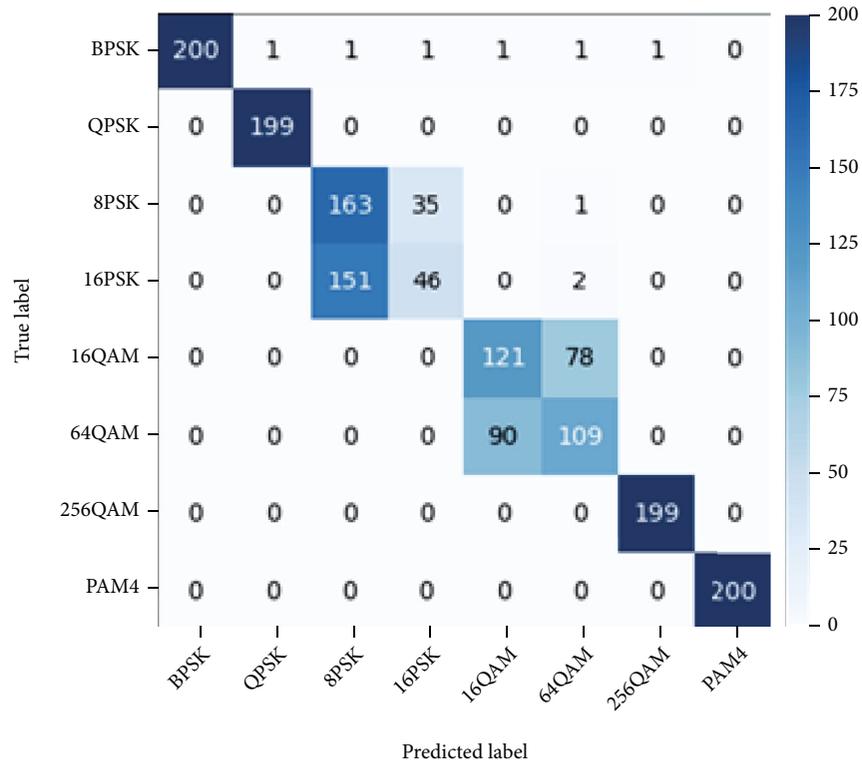


FIGURE 9: Continued.

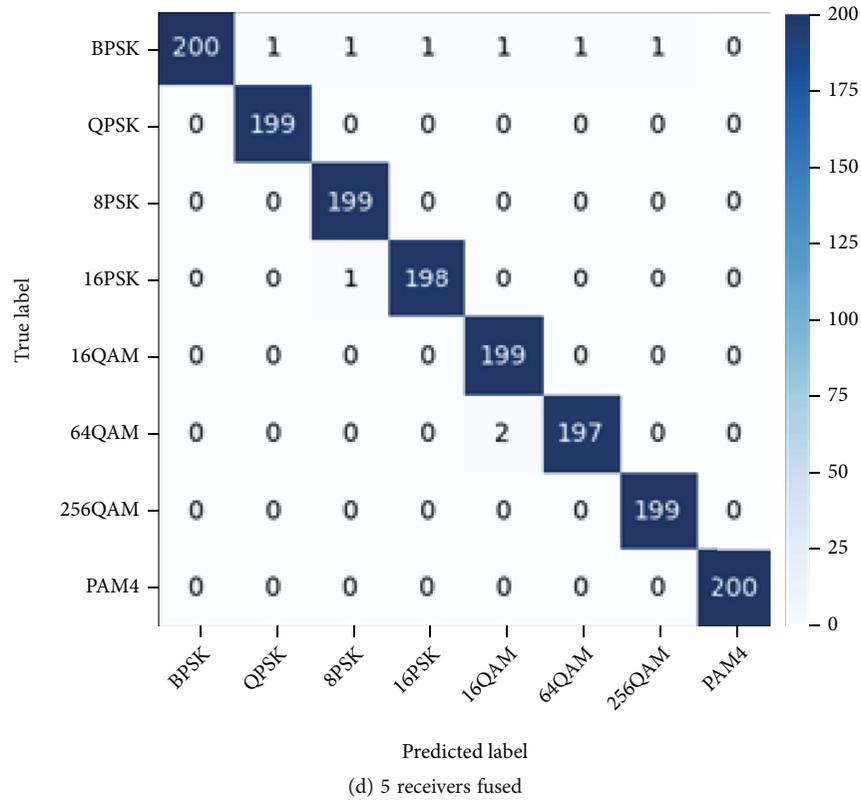
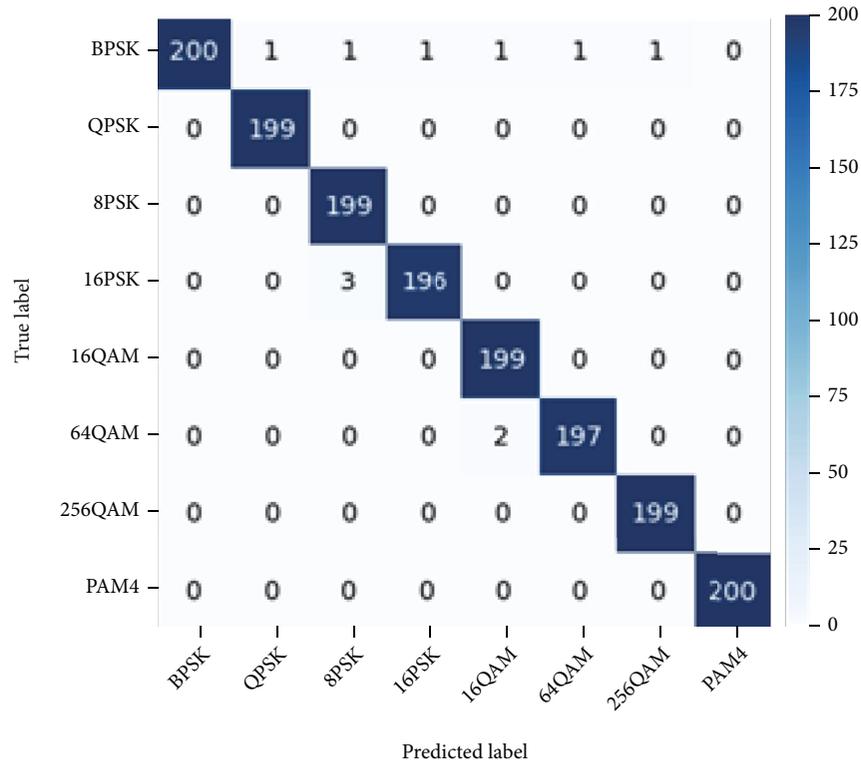


FIGURE 9: Recognition confusion matrix of different receivers under 0 dB. (a) is the recognition effect of receiver 1, receiver 1 is the transmission situation with the worst channel condition, (b) is the recognition effect of receiver 3, and (c) is the recognition effect of receiver 5. Receiver 5 is the best channel transmission situation, and (d) is the effect of collaborative identification by 5 receivers.

TABLE 3: Parameter settings for the new dataset.

Parameter	Dataset 1	Dataset 2
Path delays	[0 1e-4]	[0 1e-5]
AveragePathGains	[1 4]	[-3 3]
MaximumDopplerShift (Hz)	20	10
Phase offset (degree)	[-90 90]	[-90 90]
Frequency offset (Hz)	[0 100]	[0 100]
Fading model	Jakes	Jakes
Rician K -factor	3	3
SNR (dB)	-20:2:20	-20:2:20

(2) Calculate the covariance matrix of the sample

$$\text{Cov} = \frac{1}{k \sum_{i=1}^m (X^i) \cdot (X^i)^T}. \quad (10)$$

Use the singular value decomposition method to calculate the eigenvalues and eigenvectors of the covariance matrix:

$$[U, S, V] = \text{SVD}(\text{Cov}), \quad (11)$$

wherein U is a dimensionality reduction matrix, which means that all the eigenvectors corresponding to the covariance matrix correspond to the eigenvalues one-to-one, and its dimension is $m * m$, and if the first d column of the matrix is selected, the sample features will be reduced to d dimensionality.

(3) Dimensionality reduction analysis

All nodes' data samples can be expressed as $X = [X^1, X^2, \dots, X^k]$, and the dimensionality reduction feature matrix is obtained according to the rules shown below:

$$Z = X \cdot U_d. \quad (12)$$

Among them, the dimension of X is $k \times m$, and the dimension U_d is $m \times d$; then, the matrix Z dimension after dimensionality reduction analysis is $k \times d$. The size of the dimensionality reduction error mainly depends on the selection of d . The larger the value of d , the more the feature vectors in the representation of U , which can retain the features of the original features, and the smaller the error, but the redundant features will also be retained, and the amount of calculation will also be reduced. To retain the system's 99% uncertainty, the determination of the d can refer to

$$\frac{1/k \sum_{i=1}^k \|X^i - X_{map}^i\|^2}{1/k \sum_{i=1}^k \|X^i\|^2} \leq 0.01. \quad (13)$$

Through the above steps, the eigenvalues of the principal components can be determined, and the new feature space

after dimensionality reduction and the new fusion feature vector can be obtained.

The feature vector after the dimensionality reduction process reduces the redundancy of the feature semantics and retains the main semantic information. At this time, the feature vector dimension becomes 30. Input the fused dimensionality reduction feature vector into the classifier for classification. Then, the modulation method of the current sample can be obtained.

In machine learning, the function of the classifier is to judge the class to which a new observation belongs on the basis of the labeled training data. In this paper, after extracting the data features of different nodes through the deep learning model, the classifier can be used to complete the category judgment of the unknown sample data. Commonly used classifiers generally include K -nearest neighbors (KNN), decision tree classifiers, and support vector machines (SVM). SVM cannot rely on statistical methods, thus simplifying the usual classification and regression problems, and can find key samples that are critical to the task, so this method is used in this paper for the final classification and recognition.

3. Results and Discussion

3.1. Experimental Dataset. In this section, a number of experiments are carried out to evaluate the performance of this model and algorithm. The dataset used in the experiment is generated by simulation, and its parameters are shown in Table 2. In actual communication scenarios, low-order modulated signal features are easier to extract, but high-order modulated signals are usually used in practice. Recognize modulation modes including confusing modulation mode signals as dataset Φ . Each modulation type under each SNR includes 1000 instances, 80% of which are selected as the training set and 20% as the test set.

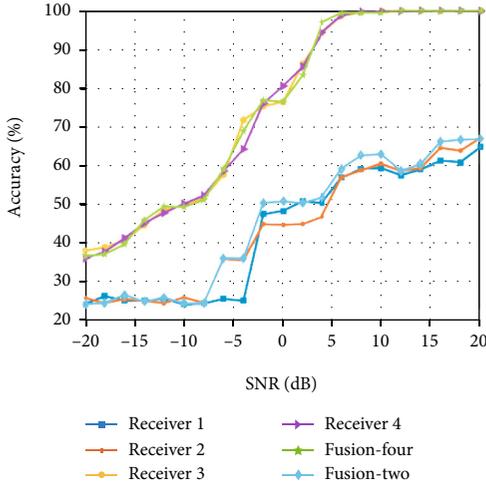
$$\Phi = \{\text{BPSK}, \text{QPSK}, \text{8PSK}, \text{16PSK}, \text{16QAM}, \text{64QAM}, \text{256QAM}, \text{PAM4}\}. \quad (14)$$

3.2. Experimental Settings

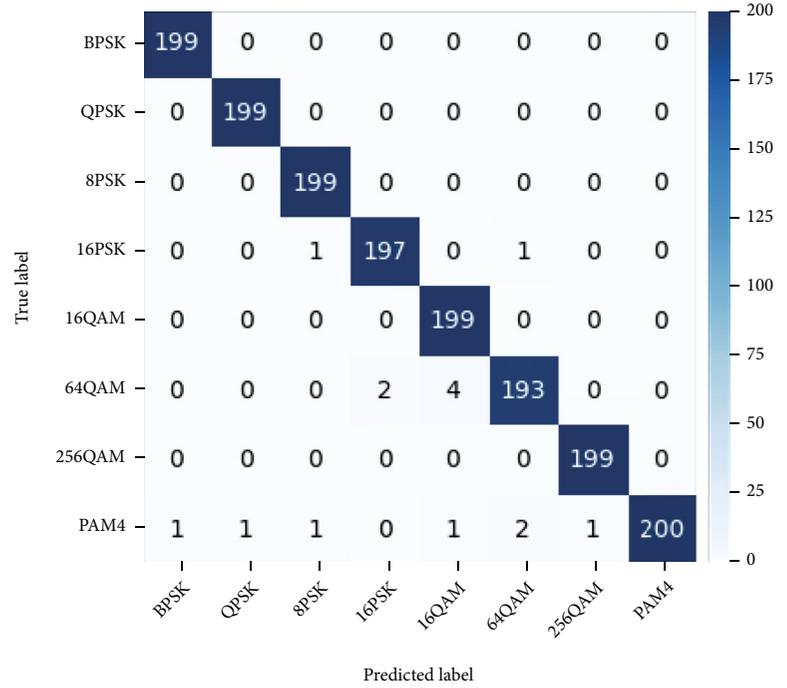
3.2.1. Comparison of Feature Extraction Network Models. All experiments in this study are carried out in the TensorFlow framework, and the GPU accelerator used is GeForce RTX 2060. In all simulation experiments, the training model adopts the adaptive moment estimation optimizer, and the learning rate is set to 0.0001 to evaluate the training parameters. In order to prevent overfitting, this paper adds an early stop mechanism during the training process. When the loss function is iterated 30 times and when it is not falling, it can be considered that the model training is completed and tested.

To verify the effectiveness of the feature extraction module in this study, several different network models are trained and tested on the RadioML2016.10a dataset. The test results are shown in Figure 5.

As can be seen from Figure 5(a), on the open-source dataset RadioML2016.10a dataset, the CLDNN network



(a) Recognition accuracy rate change curve



(b) Recognition of fusion-four under 8 dB

FIGURE 10: Recognition performance under different channel conditions. In (a), fusion-four is the variation curve of the cooperative recognition results of four receivers with different channel transmission status with SNR. Fusion-two represents the variation of the cooperative recognition results of receivers 1 and 2 with poor channel conditions.

model (CLDNNnet) network exhibits the best recognition performance. In terms of parameters, it can be seen from Figure 5(b) that ResNet and CLDNNnet have the least amount of network parameters, saving computing resources, and CLDNNnet is better than ResNet in terms of recognition performance. Therefore, considering comprehensively, the subsequent experiments in this paper extract different node features through the CLDNNnet.

3.2.2. Distributed Collaborative Recognition Results and Analysis. In the actual sensor network, due to the influence of factors such as the transmission distance and the distribution location of the monitoring nodes, the signal energy and quality at each receiving node are different. In the simulation experiment environment, it is mainly reflected in the different SNR of the received signal of each node. Therefore, in this experiment, each channel is set to be independent, and the signal quality of different receivers under different channel state transmission conditions is simulated with different sizes of Gaussian white noise. The experimental verification is divided into the following two cases:

- Keep the average SNR of each node the same, and test the recognition effect of feature fusion of different numbers of receiver signals
- Node 1 represents the node with the worst channel quality, and its SNR, namely, SNR1 varies from -20 to 12 dB, and the remaining nodes increase by 2 dB

on the basis of node 1, namely, $\text{SNR}_2 = \text{SNR}_1 + 2$, $\text{SNR}_3 = \text{SNR}_1 + 4$, ...

Figure 6(a) illustrates the recognition performance of different numbers of receivers under the same average SNR. It can be seen from the figure that with the increase in the number of receivers, the recognition effect after feature fusion is continuously improving. The recognition performance is improved significantly under the low SNRs. In the case of 7-receiver cooperative recognition, the average recognition accuracy of the 8 kinds of modulated signals can reach 81.3% at -4 dB, which is nearly 10% higher than the single-receiver recognition accuracy. This is conducive to realize modulation recognition of weak and poor-quality communication signals in noncooperative communication scenarios. It can be seen from the confusion matrix in Figure 6(b) that the recognition error mainly comes from the confusion of 8PSK and 16PSK, 16QAM, and 64QAM which have similar features.

Under the setting of experiment (b), the variation curve of the cooperative recognition accuracy rate of different numbers of receivers under -20-12 dB is shown in Figure 7. With the increase in the number of receivers, the accuracy rate is continuously improving, and the improvement is obvious when the SNR is low. It can be found from Figure 8 that in the signal transmission process of different channels, under the conditions of low SNR or high SNR, the fusion node identification effect always tends to identify the receiver node with the highest performance. In a

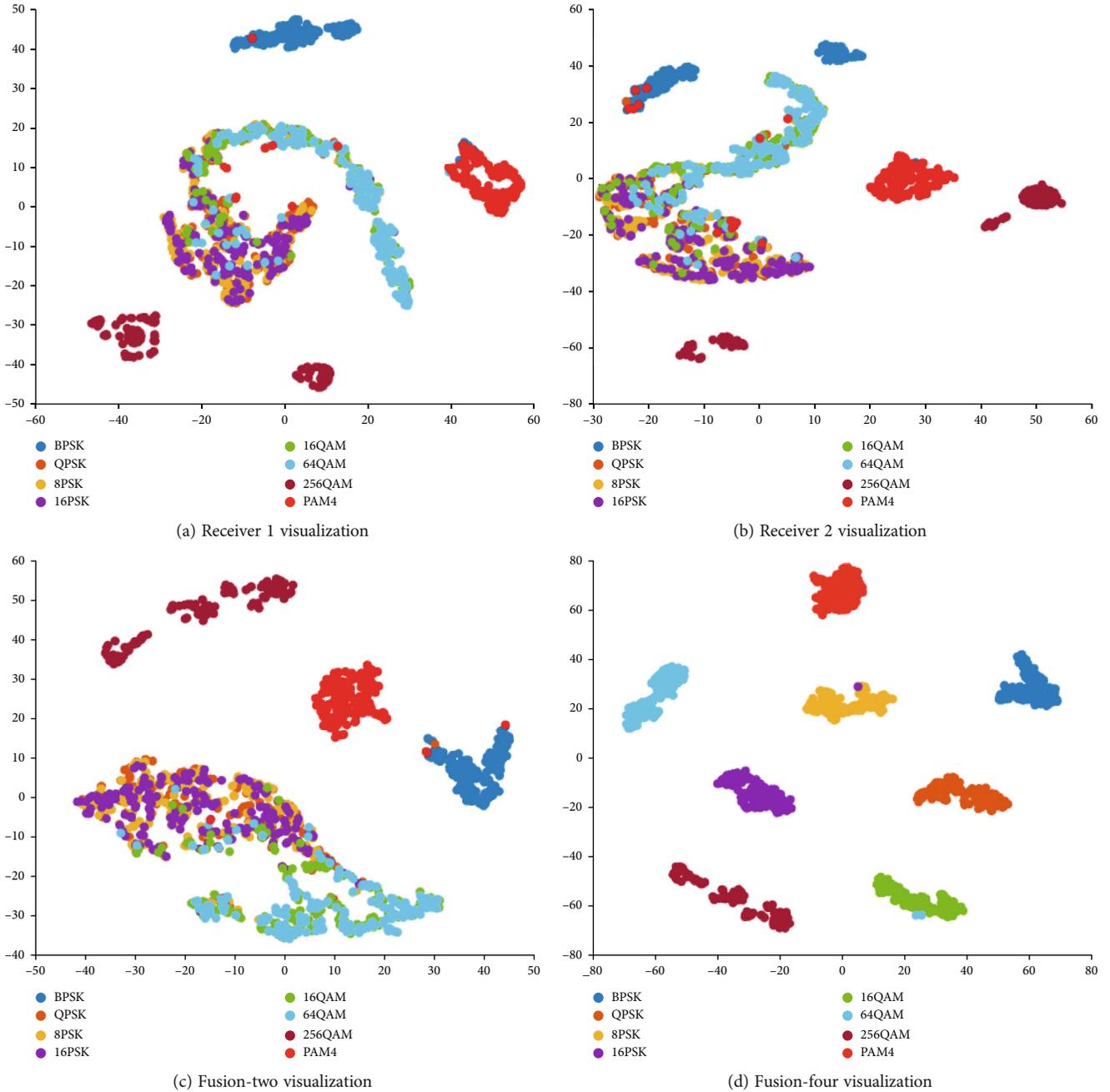


FIGURE 11: Feature cluster visualization.

distributed sensor network, the fusion recognition of the features of different receiver nodes can eliminate the ambiguity of unknown signals to a great extent. When the condition of the transmission channel of a small number of signals deteriorates, it can still maintain a high recognition probability in the end.

In Figure 9, receiver 1 is the transmission with the worst channel conditions. The recognition accuracy rate under 0 dB is only 77.31%. There is serious aliasing between 8PSK and 16PSK and between 16QAM and 64QAM. A good receiver 3 still has a serious aliasing phenomenon between 16QAM and 64QAM. This is because despite the improved

channel conditions, for signals with similar characteristics, the model is still unable to accurately discriminate between confusing samples. In receiver 5, due to good channel conditions and high received signal quality, an average recognition accuracy rate of 99.3% is achieved at 0 dB. In spite of the poor channel conditions, the average recognition accuracy rate of 0 dB still reaches 99.44% when five receivers are cooperatively recognized, and there is basically no recognition confusion. The distributed architecture of the sensor network uses the complementarity of the data features of each receiver to synthesize the extracted features into a feature that is more discriminative than the input feature,

thereby greatly improving the recognition effect in harsh environments without affecting performance of other receivers.

3.2.3. Comparison of Different Channel Transmission Conditions. In actual signal transmission on different channels, different magnitudes of frequency offset, phase offset, multipath fading, etc. are often generated. In order to further explore the identification method of distributed multireceiver node feature fusion in different channel environments, on the basis of the above dataset, different degrees of frequency offset, phase offset, and multiple fading delays are added to affect the channel. In order to facilitate comparison, experiments are carried out by adding noise of the same magnitude to each node. The parameters of the further generated dataset are shown in Table 3.

Datasets 1 and 2 are generated by simulation in Rayleigh fading channels. By adding different degrees of fading coefficients and delays, the situation of receivers placed at different distances from the transmitter is further simulated, representing the signals at both ends of receivers 1 and 2, respectively. In addition, two datasets with only frequency offset or phase offset are set as the channel simulation with better transmission status, receiver 3 and receiver 4, respectively.

Although there are a few cases of severe channel deterioration in the distributed system architecture, the recognition performance has been improved after fusion, and the result after fusion always tends to go to the node with the best performance. Even compared to the best receiving node, there is still a certain degree of improvement. As shown in Figure 10, the recognition performance of receiver 1 and receiver 2 is very unsatisfactory compared with receivers 3 and 4, but the fusion effect is significantly improved.

The extracted features of different receiver signals are visualized after dimensionality reduction, and the effect of feature extraction can be further observed. The effect of data feature visualization under 8 dB is shown in Figure 11. Because the transmitter signal is transmitted in a complex channel environment, it is affected by a variety of factors, resulting in poor signal quality received by receivers 1 and 2 and difficulty in feature extraction. In Figures 11(a) and 11(b), it can be found that the signal features of the two have serious aliasing phenomenon, and the clustering effect is poor. It is difficult to obtain accurate feature expressions, resulting in poor classification and recognition performance. When the features of the two nodes are fused, there is still serious aliasing, and it is difficult to achieve better recognition results, as shown in Figure 11(c). But when the signal features of the receiving end with better transmission are added, as shown in Figure 11(d), there is basically no aliasing. The signal features of each modulation type can be clearly distinguished, and the signal features of the same modulation are strongly aggregated, so that better recognition results can be achieved. This experiment further illustrates that although the distributed collaborative recognition framework based on feature fusion can improve the signal recognition performance to a certain extent, when the signal transmission channels are severely

deteriorated, the recognition performance will not be greatly improved.

4. Conclusions

Aiming at the problem that it is difficult to accurately extract signal features and perception signal semantic information from complex channel transmission in noncooperative communication scenarios, this paper uses the excellent feature extraction performance of DL algorithms to propose a recognition architecture for multisensor distributed cooperative sensing spectrum semantics based on feature fusion. The distributed wireless sensor network monitors the electromagnetic spectrum to realize the semantic information of wireless communication signals. The simulation experiment proves that the distributed collaborative sensing and recognition architecture in this paper can solve the problems of single-node signal transmission, which is difficult to accurately analyze the spectrum semantic information under complex channel conditions, poor adaptability to the environment, and low recognition performance. In particular, in the confrontation scenario of noncooperative communication, the algorithm in this paper is more conducive to the feature extraction of weak signals and realizes spectral semantic perception and recognition. However, this solution still cannot guarantee a good recognition effect in a more complex communication signal transmission environment with higher recognition accuracy requirements. Since the datasets in this experiment are all generated under simulation conditions, in order to further verify the effectiveness of the algorithm, the next research will be applied to the perception recognition of the measured datasets on this basis. In addition, this paper realizes the fusion scheme based on the feature layer. How to coordinate the recognition of the data layer, feature layer, and decision layer without increasing the data computing load is an important research direction.

Data Availability

The RadioML2016.10a used in this paper is an open-source dataset in Reference [6]. Other datasets used in this article can be obtained by contacting the email zhangjing03@hr-beu.edu.cn.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (grant number 62001137), the Natural Science Foundation of Heilongjiang Province (grant number JJ2019LH2398), and the Foundation of Key Laboratory of Signal and Information System of Shandong Province.

References

- [1] J. L. Xu, W. Su, and M. Zhou, "Likelihood-ratio approaches to automatic modulation classification," *IEEE Transactions on Systems Man and Cybernetics Part C (Applications and Reviews)*, vol. 41, no. 4, pp. 455–469, 2011.
- [2] M. Mansouri, R. Baklouti, M. F. Harkat, M. Nounou, H. Nounou, and A. B. Hamida, "Kernel generalized likelihood ratio test for fault detection of biological systems," *IEEE Transactions on Nanobioscience*, vol. 17, no. 4, pp. 498–506, 2018.
- [3] A. Hazza, M. Shoaib, and S. A. Alshebeili, "An overview of feature-based methods for digital modulation classification," in *2013 1st international conference on communications, signal processing, and their applications (ICCSPA)*, pp. 1–6, Sharjah, United Arab Emirates, 2013.
- [4] M. L. D. Wong and A. K. Nandi, "Automatic digital modulation recognition using spectral and statistical features with multi-layer perceptrons," in *Proceedings of the Sixth International Symposium on Signal Processing and its Applications*, pp. 390–393, Kuala Lumpur, Malaysia, 2001.
- [5] C. Hou, Y. Li, X. Chen, and J. Zhang, "Automatic modulation classification using KELM with joint features of CNN and LBP," *Physical Communication*, vol. 45, no. 3, article 101259, 2021.
- [6] T. J. O'shea and N. West, "Radio machine learning dataset generation with gnu radio," *Proceedings of the GNU Radio Conference*, vol. 1, no. 1, 2016.
- [7] Y. Wu, X. Li, and J. Fang, "A deep learning approach for modulation recognition via exploiting temporal correlations," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications*, pp. 1–5, Kalamata, Greece, 2018.
- [8] H. Wu, Y. Li, L. Zhou, and J. Meng, "Convolutional neural network and multi-feature fusion for automatic modulation classification," *Electronics Letters*, vol. 55, no. 16, pp. 895–897, 2019.
- [9] Y. Lin, Y. Tu, and Z. Dou, "An improved neural network pruning technology for automatic modulation classification in edge devices," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 5, pp. 5703–5706, 2020.
- [10] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *International conference on engineering applications of neural networks*, pp. 213–226, Cham, 2016.
- [11] Y. Lin, M. Wang, X. Zhou, G. Ding, and S. Mao, "Dynamic spectrum interaction of UAV flight formation communication with priority: a deep reinforcement learning approach," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 3, pp. 892–903, 2020.
- [12] S. Ramjee, S. Ju, D. Yang, X. Liu, A. E. Gamal, and Y. C. Eldar, "Fast deep learning for automatic modulation classification," 2019, <http://arxiv.org/abs/1901.05850>.
- [13] Y. Tu, Y. Lin, J. Wang, and J. U. Kim, "Semi-supervised learning with generative adversarial networks on digital signal modulation classification," *CMC-Computers Materials & Continua*, vol. 55, no. 2, pp. 243–254, 2018.
- [14] C. Hou, X. Zhang, and X. Chen, "Electromagnetic signal feature fusion and recognition based on multi-modal deep learning," *International Journal of Performability Engineering*, vol. 16, no. 6, pp. 941–949, 2020.
- [15] X. Li, F. Dong, S. Zhang, and W. Guo, "A survey on deep learning techniques in wireless signal recognition," *Wireless Communications and Mobile Computing*, vol. 2019, 12 pages, 2019.
- [16] Q. Cheng, Y. Yang, and X. Gui, "Disturbance signal recognition using convolutional neural network for DAS system," in *13th international conference on measuring technology and mechatronics automation (ICMTMA)*, pp. 278–281, Beihai, China, 2021.
- [17] Y. Zhang, N. Ansari, and W. Su, "Multi-sensor signal fusion based modulation classification by using wireless sensor networks," *Wireless Communications and Mobile Computing*, vol. 15, no. 12, pp. 1621–1632, 2015.
- [18] B. Dulek, "An online and distributed approach for modulation classification using wireless sensor networks," *IEEE Sensors Journal*, vol. 17, no. 6, pp. 1781–1787, 2017.
- [19] C. Lyu, Z. Huo, X. Cheng, J. Jiang, A. Alimasi, and H. Liu, "Distributed optical fiber sensing intrusion pattern recognition based on GAF and CNN," *Journal of Lightwave Technology*, vol. 38, no. 15, pp. 4174–4182, 2020.
- [20] J. Li, B. Lu, Y. Wang, X. Liu, Q. Bai, and B. Jin, "Distributed optical fiber vibration sensor for the identification of pipeline leakage using relevant vector machine," in *Optics Frontiers Online 2020: Distributed Optical Fiber Sensing Technology and Applications*, vol. 11607, p. 116070J, Virtual, Online, China, 2021.
- [21] Z. Sun, K. Liu, J. Jiang et al., "Optical fiber distributed vibration sensing using grayscale image and multi-class deep learning framework for multi-event recognition," *IEEE Sensors Journal*, vol. 21, no. 17, pp. 19112–19120, 2021.
- [22] T. Wimalajeewa, J. Jagannath, P. K. Varshney, A. Drozd, and W. Su, "Distributed asynchronous modulation classification based on hybrid maximum likelihood approach," in *MILCOM 2015-2015 IEEE Military Communications Conference*, pp. 1519–1523, Tampa, FL, United States, 2015.
- [23] J. L. Xu, W. Su, and M. Zhou, "Distributed automatic modulation classification with multiple sensors," *IEEE Sensor Journal*, vol. 10, no. 11, 2010.
- [24] B. Dulek, O. Ozdemir, P. K. Varshney, and W. Su, "Distributed maximum likelihood classification of linear modulations over nonidentical flat block-fading Gaussian channels," *IEEE Transactions on Wireless Communications*, vol. 14, no. 2, pp. 724–737, 2015.
- [25] T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, "Convolutional, long short-term memory, fully connected deep neural networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4580–4584, Brisbane, QLD, Australia, 2015.