Multisignal Modulation Classification Using Sliding Window Detection and Complex Convolutional Network in Frequency Domain

Changbo Hou, Guowei Liu, Qiao Tian, Zhichao Zhou, Lijie Hua, and Yun Lin, *Member, IEEE*

***Abstract*—With the development of the Internet of Things (IoT), the IoT devices are increasing day by day, resulting in increasingly scarce spectrum resources. At the same time, many IoT devices are facing inevitable malicious attacks. The cognitive Radio-enabled IoT (CR-IoT) is proposed as an effective method for spectrum resource allocation and risk monitoring in the IoT. The signal detection and modulation recognition are the key technologies for CR-IoT, addressing the problem of multisignal detection and automatic modulation classification (AMC) is one of the prerequisites for realizing secure dynamic spectrum access. Based on sliding window and deep learning (DL), this study proposes a multisignal frequency domain detection and recognition method. The frequency spectrum of the time-domain overlapping signal is obtained through the fast Fourier transform (FFT), and the frequency spectrum is segmented based on the signal energy detection method. Finally a complex convolutional neural network (CNN) is constructed for the identification of signal spectrum information. The proposed method can recognize 264 time-domain aliasing and frequency-closed signals with an accuracy of 97.3% under the influence of −2 dB corresponding to the noise of the calibration signal. In addition, the proposed method eliminates the influence of bandwidth, which can effectively detect and recognize the signal types of each component in the frequency band. This method has wide applicability and provides an effective scheme for the IoT cognitive technology.**

***Index Terms*—Automatic modulation classification (AMC), complex convolutional neural network (CNN), Internet of Things (IoT) cognitive, multisignal sensing.**

I. Introduction

AS AN extension of the information world to the physical world, the Internet of Things (IoT) is currently facing the problem of spectrum resource allocation caused by diverse access devices and the explosive growth of traffic [1]–[3]. Meanwhile, due to its open characteristics, devices connected to the IoT are easily targeted by attacks [4]–[6]. The realization of spectrum sensing and risk monitoring is essential to promote the development of the IoT, and the proposal of the cognitive radio-enabled IoT (CR-IoT) provides the possibility. Signal detection and automatic modulation classification (AMC) are key technologies in CR-IoT [7]–[9], which are the important parts of the process of accurate perception and decision making for the surrounding electromagnetic environment characteristics, and they have a broad application prospect in the CR-IoT [10]–[12], such as reliable communication [13]–[15], interference location, abnormal information monitoring, modulation and demodulation [16], [17], etc.

At present, the research on AMC of the single signal has made a lot of achievements and it mainly include the following two categories: 1) classical signal modulation recognition algorithms and 2) modulation recognition algorithms based on deep learning (DL). The classical modulation recognition algorithms can be divided into two categories: 1) maximum-likelihood (ML) hypothesis testing methods based on decision theory [18]–[20] and 2) pattern recognition methods based on feature-based (FB) [21]–[23]. The ML method derives the likelihood function based on the statistical characteristics of the signal model, and uses the likelihood function for measuring the similarity between the signal and the known classes to determine the modulation mode of the signal. This type of methods is simple to implement, but depends on the prior information of the signal and has high computational complexity. Its weak robustness to the environment and high computational complexity are difficult to meet the real-time implementation needs of IoT devices. The core of the FB method is to extract signal features for classification in a high-dimensional space, and then the classifiers used for signal classification. According to the difference of transform domain, it can be divided into time-domain analysis method [21], frequency-domain analysis method [22], and other domain analysis method [23]. Compared with the ML method, the FB method is more robust to model mismatches. However, the method cannot handle complex modulation signals and still needs to achieve a balance between accuracy and computing time. These methods cannot handle complex modulation signals and need to achieve a balance between accuracy and computing time. In CR-IoT, ensuring accuracy while reducing computational complexity plays an important role in the stability and reliability of the system, and the emergence of the DL provides the possibility for this. To

Manuscript received 12 March 2022; accepted 10 April 2022. Date of publication 15 April 2022; date of current version 23 September 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62001137; in part by the Natural Science Foundation of Heilongjiang Province under Grant JJ2019LH2398; and in part by the Fundamental Research Funds for the Central Universities under Grant 3072021CF0805. *(Corresponding author: Qiao Tian.)*

Changbo Hou is with the College of Information and Communication Engineering and the Key Laboratory of Advanced Marine Communication and Information Technology, Ministry of Industry and Information Technology, Harbin Engineering University, Harbin 150001, China.

Guowei Liu, Qiao Tian, Zhichao Zhou, Lijie Hua, and Yun Lin are with the College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China (e-mail: tianqiao@hrbeu.edu.cn).

Digital Object Identifier 10.1109/JIOT.2022.3167107

overcome the shortcomings of classical modulation recognition algorithms, researchers have applied DL to the field of cognitive radio [24]–[27]. Huang *et al.* [27] proposed a novel gated recurrent residual neural network (GrrNet), the amplitude/phase (A/P) characteristics of the signal are extracted through ResNet, and the gated recurrent unit (GRU) in the model is responsible for extracting the time information to ensure that it can process signals of any length. The experimental results verify the classification performance and robustness of GrrNet. Zhang *et al.* [26] proposed a two-way structure to achieve AMC. This method combines the spatial feature extraction capabilities of convolutional neural network (CNN) and the advantages of long short-term memory (LSTM) in time series processing to extract the in-phase/quadrature (I/Q) format and the amplitude/phase features of the signal. Finally, the interaction between different features improves the performance of recognition. The communication signal recognition method based on complex neural network is suitable for the signal I/Q waveform domain [24]. The existing research prove that the complex network has better model generalization ability than partial the real network. Using the complex network for the recognition and classification of communication signal I/Q waveform domain, which can obtain higher recognition accuracy. DL has excellent classification performance on single-component signal modulation recognition tasks. However, considering that there are more than one devices working in the IoT scenarios simultaneously, multisignal recognition is still an important factor hindering the development of CR-IoT.

A. Related Multisignal Recognition Work

To separate and recognize two-component or multicomponent mixed signals, some key problems must be overcome. Because the prior information of aliased signals is often unknown, when the receiver receives multiple signals, it is necessary to separate the received signals effectively to ensure subsequent and accurate separation. Due to the particularity of communication signals, the traditional single channel separation technology for audio signals cannot be directly transferred into the separation of communication signals. For example, computational auditory scene analysis is mainly for low-frequency signals, and it is powerless for high-frequency signal processing [28]. Signal separation can be carried out using basis function decomposition combined with data training methods, which can only separate the speech signal with the characteristics of this basis function [29]. Therefore, it is necessary to realize the separation of single channels according to the differences between the signal and channel. In [30], a feasible signal separation technique based on fractional Fourier transform (FRFT) was proposed, which supports up to nine kinds of two-component superimposed mixed signals for radar emitter signals. The received signals are decomposed into several components by FRFT. For the first component, GoogLeNet based on transfer learning is used for recognition, and the second component uses variational mode decomposition to eliminate noise and extracted fusion features to improve recognition accuracy. The recognition accuracy of the single signal is 96.23%, and that of the two-component signal is 72% when the signal-to-noise ratio (SNR) is 0 dB. This method aims at limited scenes, and when the signal components increase, the separation becomes difficult. Yuan *et al.* [31] proposed a DL framework for blind detection of Morse code in broadband signals. A multisignal detection module was used to extract features from the broadband spectrum, and then the desired narrowband signal was obtained by detection and location. Finally, the extracted signal features are sent to deep Morse based on CNN for detection and recognition of the narrowband signal. The proposed method had higher accuracy and the best stability in comparison with the traditional method.

There has been some research on the recognition and segmentation of aliased signals based on the image features of signals [32], [33]. In [32], first, the input mixed signals are transformed into a 2-D time–frequency image by short-time Fourier transform (STFT), and then the feature extraction and recognition are carried out using a deep CNN (DCNN). Meanwhile, the authors proposed a repeated selection strategy based on grid segmentation and filtering. The method is not only superior to SVM classifier in single signal recognition but also better in mixed signal. The accuracy of the classification is not less than 92% at 0 dB with an accuracy of greater than 98% at 5 dB. However, this method can only support the signal segmentation and recognition when the aliasing degree is low. Qu *et al.* [33] proposed a scheme to support modulation recognition of single-component and two-component radar signals based on the above method [32]. The authors change the kernel function of time-frequency transform, and use convolution de-noising autoencoder (CDAE) to suppress the noise of time–frequency images. This method can recognize eight kinds of common radar modulation signals, such as LFM, BPSK, FSK, and SFM, and the total PSR of the two-component radar signal can reach 94.83% when the SNR is 6 dB. However, the number of signals recognized by this method is limited, and it is difficult to locate the concerned signals.

B. Contributions

In summary, the above methods provide feasible ideas for multisignal recognition, but most of them only focus on dual-component signals, which cannot meet the scenario that a great many devices are connected in parallel of the CR-IoT. This study proposes a sliding window-based multisignal frequency domain detection and recognition scheme, and the innovations and contributions for the detection and recognition of multicomponent wireless communication signals in IoT scenarios are as follows.

1) For frequency bands, a detection and recognition method for multisignal in the frequency domain is proposed. This method can be used when many single-component signals are aliased in the time domain and close in the frequency domain. The proposed method can not only locate the frequency band of the signals but also make full use of the frequency and phase characteristics of the signals to recognize the signal modulation type in the frequency band.

TABLE I
SUMMARY OF ABBREVIATIONS

Abbreviations	Notations
IoT	Internet of Things
CR-IoT	Cognitive Radio-enabled IoT
AMC	Automatic Modulation Classification
DL	Fast Fourier Transform
CNN	Convolutional Neural Network
ML	Maximum Likelihood
FB	Feature Based
A/P	Amplitude/Phase
GRU	Gated Recurrent Unit
LSTM	Long Short Term Memory
I/Q	In-phase/Quadrature
FRFT	FRactional Fourier transform
SNR	Signal-to-Noise Ratio
STFT	Short-Time Fourier transform
DCNN	Deep Convolutional Neural Network
CDAE	Convolution De-noising Auto-Encoder
DQN	Deep Q Network
AWGN	Additive White Gaussian Noise
ZCR	Zero Crossing Rate
SPWVD	Smooth Pseudo Wigner-Ville Distribution
BA	Baseband Adjustment

2) The signal preprocessing method eliminates the recognition problem caused by the difference between the bandwidth. The baseband adaptive method is used to offset the error of the direct bandwidth inconsistency between the test data and the training data, which expands the application scope of the model. The trained model can effectively recognize the signals with different baseband frequencies.

This study is organized as follows. Section II introduces the methods used in this study, and presents the proposed multisignal sensing recognition model and algorithm flow. The results of related experiments are discussed in Section III. In Section IV, the conclusions and summarizes future research directions are given. Moreover, the summary of abbreviations is listed in Table I.

II. SIGANL MODEL AND METHOD

As illustrated in Fig. 1, the method proposed mainly includes three modules: 1) spectrum preprocessing; 2) signal detection; and 3) modulation recognition. After the frequency spectrum of the multicomponent time-domain aliasing signal is obtained by fast Fourier transform (FFT), the frequency spectrum is segmented based on the inductive detection of signal energy, and the signal segment data are reserved for modulation recognition. To make full use of the spectrum information after FFT, the Complex-ResNet is constructed, which is suitable for simultaneous real and imaginary part inputs.

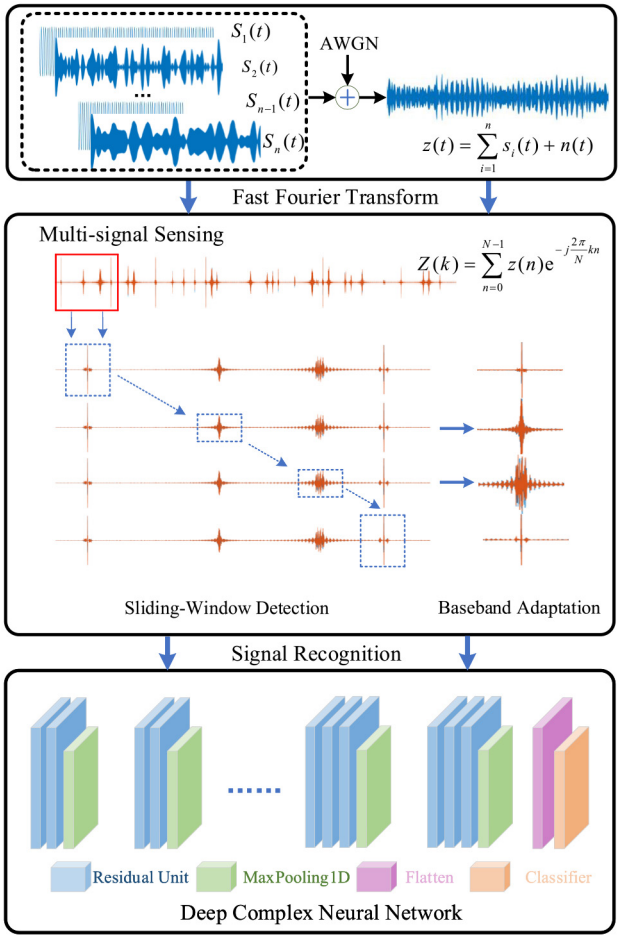


Fig. 1. Proposed multisignal sensing method.

Meanwhile, for eliminating the recognition error caused by the baseband rate, the baseband adaptive method is introduced to adapt the signal to the trained complex CNN.

A. Signal Model

In this study, signal modulation recognition in additive white Gaussian noise (AWGN) channels is considered, and data sets containing different SNR environments are established. In general, the expression of single-component modulation signals is shown as follows:

$$y(t) = s(t) + n(t) \tag{1}$$

where $n(t)$ is the white Gaussian noise with zero mean and variance of σ_n^2 and it is independent of the modulation signal; and $s(t)$ is the modulation signal to be recognized.

For the digital modulation signal, when the signal is MFSK, MPSK, or MASK, the signal model $s(t)$ can be expressed as

$$s(t) = \left[A_m \sum_n a_n g(t - nT_s)\right] \cos(2\pi(f_c + f_m)t + \phi) \tag{2}$$

where $g(t)$ is a rectangular pulse of duration T_s and A_m is the amplitude of the signal, $a_n \in \{\pm 1, \pm 3, \ldots, \pm(M-1)\}$.

In addition to digital modulation signals, AM, DSB, SSB, and FM are also added in this study for the universality of

the proposed method. AM and DSB can be, respectively, expressed as

$$s_{AM}(t) = A(1 + m(t))\cos(\omega_c t + \theta_c) \tag{3}$$

$$s_{DSB}(t) = m(t)A\cos(\omega_c t + \theta_c). \tag{4}$$

In expressions (3) and (4), $m(t)$ is the baseband signal, and ω_c and θ_c are carrier frequency and initial phase, respectively. Single sideband modulation is different from the former two, and the expression is as follows:

$$\begin{aligned} s_{USB}(t) &= Am(t)\cos(\omega_c t + \theta_c) - Am(t)\sin(\omega_c t + \theta_c) \\ s_{LSB}(t) &= Am(t)\cos(\omega_c t + \theta_c) + A\hat{m(t)}\sin(\omega_c t + \theta_c) \end{aligned} \tag{5}$$

where $\hat{m}(t)$ is the Hilbert transform of $m(t)$.

The mathematical expression of FM modulation signal is as follows:

$$\begin{aligned} \theta(t) &= 2\pi k_{FM}\int m(t)dt \\ s_{FM}(t) &= A\cos(\omega_c t + \theta(t)) \\ \Delta\theta_{\max} &= \frac{k_{FM}m_a}{f_m}. \end{aligned} \tag{6}$$

k_{FM} is called the frequency offset constant. The modulation index of $\Delta\theta_{\max}$ for frequency modulation is also the maximum phase shift.

The multicomponent signal model can be expressed as

$$x(t) = \sum_{i=1}^{n} y_i(t) + n(t) \tag{7}$$

where $y_i(t)$ represents any signal mentioned above, and n represents the number of component signals. Meanwhile, the multicomponent signal has the following assumptions.

1) $x(t)$ consists of K mutually independent signals $y_i(t)$ and noise $n(t)$.
2) Different $y_i(t)$ are completely aliased in the time domain, while the spectrum can be divided or close to each other in the frequency domain.
3) For the recognition model, the number of components K in $x(t)$ is unknown, meanwhile, the parameters of component signal $y_i(t)$, such as symbol rate, carrier frequency, and initial phase, are unknown.

B. Sliding Window Detection Based on Signal Spectrum Energy

As mentioned earlier, accurate segmentation of multiple signals in the frequency domain can effectively ensure the accuracy of modulation recognition. Therefore, based on FFT, this study proposes a detection method of signal spectrum energy with a sliding window. This method takes the short-time energy and the short-time average zero-crossing rate (ZCR) as parameters, and uses the dual-threshold detection algorithm to determine whether the signal exists with recording its endpoint position.

The short-time energy reflects the energy of the sample in the current window, and the formula is as follows:

$$E(i) = \frac{1}{N}\sum_{n=1}^{N} |x_i(n)|^2 \tag{8}$$

TABLE II
DIFFERENT SEGMENTS OF THE MULTICOMPONENT SIGNAL SPECTRUM

Segments	Detection threshold
Silent zone	$E(n) < E_{low}$ or $Zcr(n) < Z_{low}$
Transition zone	$E(n) > E_{low}$ or $Zcr(n) > Z_{low}$
Signal zone	$E(n) > E_{high}$ or $Zcr(n) > Z_{high}$

where $x_i(n)(n = 1, 2, \ldots, N)$ is the sample of the central sensing area, and the length is N. The short-time energy represents the energy distribution of the signals in the frequency domain, but the center frequency of the signals cannot be accurately calibrated, so the short-time average ZCR is introduced

$$Zcr(n) = \sum_{m=-\infty}^{\infty} |\mathrm{sgn}[x(m)] - \mathrm{sgn}[x(m-1)]|\omega(m). \tag{9}$$

In (11), $\mathrm{sgn}(n)$ is a symbolic function

$$\mathrm{sgn}[x(n)] = \begin{cases} 1, & x(n) \geq 0 \\ -1, & x(n) < 0 \end{cases} \tag{10}$$

meanwhile, $\omega(n)$ is the rectangular window function

$$\omega(n) = \begin{cases} 1/2, & 0 \leq n \leq N-1 \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

After obtaining the short-time energy $E(n)$ and the short-time ZCR $Zcr(n)$ of the spectrum samples in the sliding window, the multicomponent signal spectrum is divided into the silent zone, the transition zone, and the signal zone, as listed in Table II.

When the window is sliding, if $E(n) > E_{\text{low}}$ or $Zcr(n) > Z_{\text{low}}$ in the central sensing zone, it is determined that the window enters the transition zone from the silent zone, and the starting point position is recorded. When $E(n) > E_{\text{high}}$, it is judged that the segment is in the signal zone, then the signals are saved as a set window by window

$$S_i \leftarrow \left\{S_i^1, S_i^2, \ldots, S_i^k\right\}. \tag{12}$$

Until $E(n) < E_{\text{low}}$ or $Zcr(n) < Z_{\text{low}}$, it is determined that the sample reenters the silent zone, and the end position is recorded. In the above formula, S_i^k is a signal segment, and S_i is a set of signal data between the beginning frequency point and end frequency point, which is used as the input of the subsequent recognition model.

The window function is defined as follows:

$$f(x) = \begin{cases} \alpha, & 0 < x < N \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

where α is a dynamic scale factor, the size is automatically adjusted with the signal spectrum energy maximum value $s(x)_{\max}$ input to the window function, and the relationship between α and $s(x)_{\max}$ satisfies the formula

$$\alpha \times |s(x)_{\max}| = \beta \tag{14}$$

β is the normalized input value of the recognition network.

For the same type of digital modulation signals, although the spectrum bandwidth of different signals is different, their

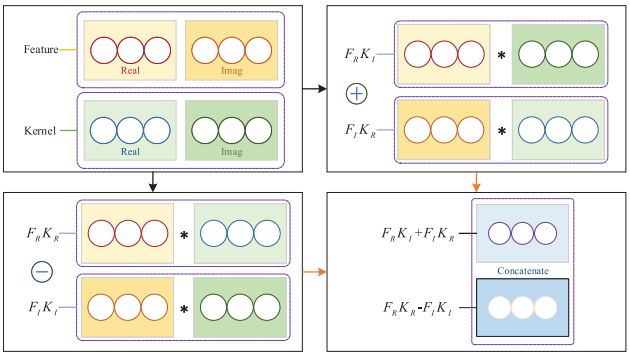


Fig. 2. Principle of complex convolution operation.

amplitude–frequency characteristics are close to the same, which called baseband invariance in this study. After the signal is detected by the sliding window, calculate the average value of the interval between the main lobe extremum $S_{pm}(x)$ and the adjacent side lobe extremum $S_{ps}(x)$

$$\Delta S_I^k = \frac{S_{pm}(x) - S_{ps}(x)}{n} \tag{15}$$

where n is the number of points between $S_{pm}(x)$ and $S_{ps}(x)$. Knowing the spectral bandwidth range ΔS_I of the training set, by measuring the average value $\Delta S_I^1, \Delta S_I^2, \ldots, \Delta S_I^k$ of the interval between k samples at the symbol rate R_b, the positively correlated parameter $\Delta S_I{}'$ can be obtained

$$\Delta S_I{}' = \frac{\Delta S_I^1, \Delta S_I^2, \ldots, \Delta S_I^k}{k}. \tag{16}$$

Interpolate the sample with $\Delta S_I / \Delta S_I{}'$ as the ratio, and then the signal to be detected can be adjusted to the range ΔS_I.

C. Proposed Complex-ResNet

The signal is represented by the complex number after FFT, for complex input data, the using of complex convolution can fully extract the sample amplitude and frequency characteristics. The process of the CNN to implement complex convolution operations is illustrated in Fig. 2. F_I and F_R represent the imaginary part and real part of the feature map and K_I and K_R correspond to the imaginary part and real part of the convolution kernel, respectively. The input vector is convolved through a complex weight matrix

$$\begin{bmatrix} \mathrm{Re}\{\hat{\boldsymbol{s}}\} \\ \mathrm{Im}\{\hat{\boldsymbol{s}}\} \end{bmatrix} = \begin{bmatrix} \mathrm{Re}\{\boldsymbol{w}\} & -\mathrm{Im}\{\boldsymbol{w}\} \\ \mathrm{Im}\{\boldsymbol{w}\} & \mathrm{Re}\{\boldsymbol{w}\} \end{bmatrix} * \begin{bmatrix} \mathrm{Re}\{\boldsymbol{x}\} \\ \mathrm{Im}\{\boldsymbol{x}\} \end{bmatrix}. \tag{17}$$

In (17), $\mathrm{Re}\{\cdot\}$ and $\mathrm{Im}\{\cdot\}$ represent the real and imaginary parts of a complex number, respectively. $\boldsymbol{w}$ is the complex weight matrix and $\boldsymbol{x}$ is the input complex vector.

Different from the network construction method in [24], to obtain a recognition model with higher identification accuracy and better robustness, we build a residuals network based on complex numbers by referring to the CNN architecture-ResNet.

The structure of the Complex-ResNet model is illustrated in Fig. 3. Its basic unit is the complex residual unit on the left of the figure, which is composed of a complex convolution layer and a complex batch normalized layer. Two complex residuals, the convolutional layer and the pooling layer, constitute a complex residuals block. The Complex-ResNet model contains five complex residuals blocks with similar structures and different parameters, which are responsible for feature extraction and finally put into the fully connected network of the model for classification.

The weight initialization the of CNN can help the model converge more quickly and accurately. The weight of complex networks can be expressed as

$$\mathbf{W} = \mathrm{Re}\{\mathbf{W}\} + i\mathrm{Im}\{\mathbf{W}\}. \tag{18}$$

According to Trabelsi *et al.* [24], if $\mathbf{W}$ is symmetrically distributed around 0, its variance can be estimated by the single parameter σ of Rayleigh distribution. Therefore, according to the Rayleigh distribution, the amplitude of the initial complex weight is $\mathbf{W}$, the expectation is 0, the variance is $2\sigma^2$, and the parameter σ is automatically adjusted with different neural network structures.

D. Multisignal Detection and Recognition

The multicomponent aliasing signals are composed of hundreds of single-component signals mentioned above. Since the signal spectrum data involves different modulation types and there is no prior information about the signal frequency distribution, if the Complex-ResNet is directly used for recognition, as the number of sampling points and signals increases, the computational complexity will increase significantly, and the recognition accuracy will also decrease. Therefore, to recognize the modulation types in the presence of signals without an estimation of signal components, this study proposes an adaptive baseband preprocessing method in the frequency domain based on signal energy induction detection. Baseband adjustment (BA) can greatly improve the applicability of the classification and recognition network. During recognition, adjusting the spectrum bandwidth of the detected signal to be consistent with the training data can eliminate the impact of different bandwidth.

The specific process of multisignal detection and recognition algorithm based on sliding window is illustrated in the Fig. 4. After the sliding window segments, the signal spectrum according to the segmentation rule, the signal segment set S_i^k on the signal spectrum is obtained. Subsequently, each group S_i in S_i^k is taken out and identified by Complex-ResNet. Suppose there are multiple spectrum sets in S_i

$$S_i = \left\{ S_i^1, S_i^2, \ldots, S_i^n \right\} \tag{19}$$

the recognition result of the signal segment is obtained after inputting Complex-ResNet

$$P_i = \left\{ P_i^1, P_i^2, \ldots, P_i^n \right\} \tag{20}$$

where P_i^n is the recognition result output by the classifier. For different signal data in the same signal segment, the deviation of the window truncation position will lead to misjudgment, so the voting algorithm is added to filter the output results and improve the recognition accuracy.

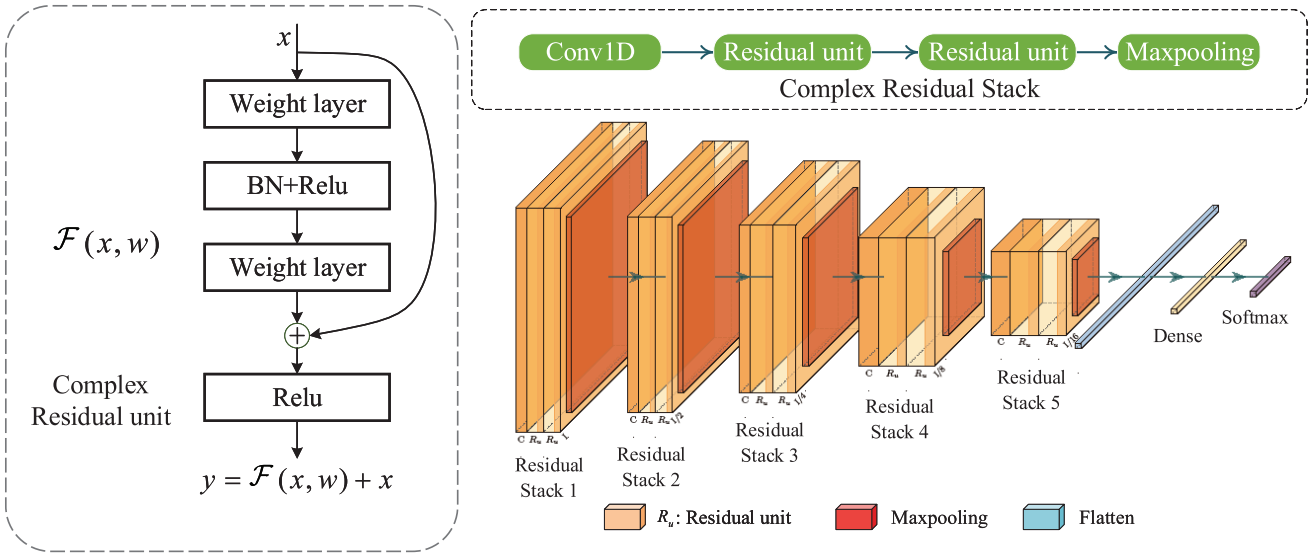


Fig. 3. Complex-Resnet network architecture.

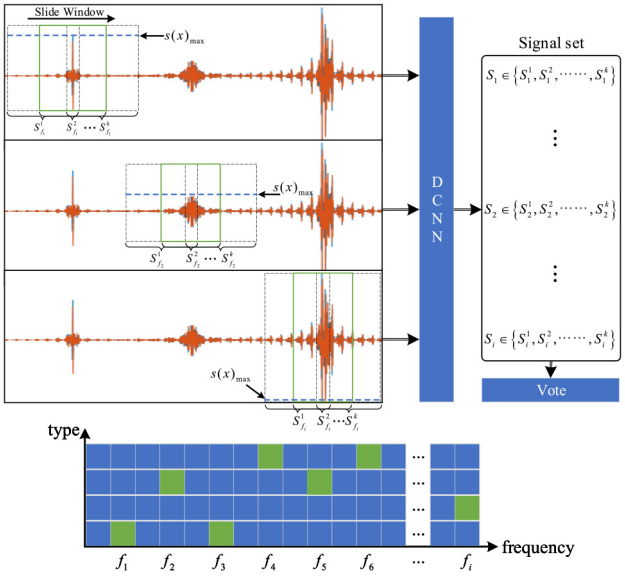


Fig. 4. Process of multisignal detection and recognition.

TABLE III
SIMULATION SIGNAL PARAMETERS

Signal type	Parameter	Range
ASK	Bandwidth T_b	1.5-30 MHz
	Carrier frequency f_0	0.1-0.2 ms
QPSK	T_b	1.5-30 MHz
	f_0	0.1-0.2 ms
2FSK	$f_1 f_2$	1.5-30 MHz
	T_b	0.1-0.2 ms
OFDM	f_0	1.5-30 MHz
	T_b	0.1-0.2 ms
	Frequency interval Δf	5-10 kHz
AM	f_0	1.5-30 MHz
	Modulation frequency m	3-15 kHz
SSB	f_0	1.5-30 MHz
	m	3-15 kHz
DSB	f_0	1.5-30 MHz
	m	3-15 kHz
FM	f_0	1.5-30MHz
	m	3-15 kHz
	Largest deviation f_{max}	7 kHz

III. Result Analysis

The specific parameters of the signals are listed in the following Table III. To test the performance of the recognition model, Complex-ResNet trains on a data set composed of eight types of signals. Meanwhile, the performance of the multicomponent signal detection and recognition method proposed in this study is tested.

A. Recognition Performance of the Complex-ResNet

The data set is generated from a signal after FFT, which is composed of real and imaginary parts, and the SNR is from −6 to 10 dB with a step of 2. The number of signals in each type under a single SNR is 1600, which are divided into training and test set according to 4:1. The architecture of a single signal is [1024, 2], to verify the recognition performance of the proposed Complex-ResNet, a CNN was constructed with reference to the CNN proposed by O'Shea *et al.* [17] and compared under the same data set. Meanwhile, the signals are transformed into time–frequency images through smooth pseudo wigner-ville distribution (SPWVD), and VGG16 and ResNet are used for identification images, all generated images are transformed resized to (224, 224, 3) to the default receptive field of VGG16 and ResNet. Experiments have been carried

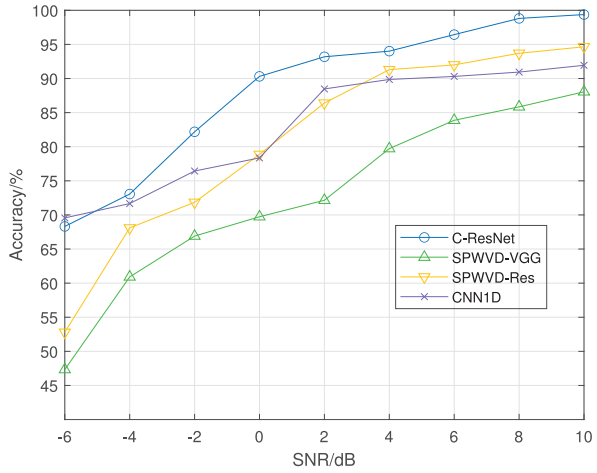


Fig. 5. Average recognition accuracy of different methods.

TABLE IV
COMPARISON OF MODEL COMPLEXITY AND TRAINING TIME

Medel	Traning Time/s	Parameters(k)
C-ResNet	4.2	432
CNN1D	3.7	185
VGG16	211.5	21140
ResNet	225.9	23553

out under various SNR data set, and the average recognition accuracy of different methods is illustrated in Fig. 5 as follows.

It can be seen from Fig. 5 that the average recognition accuracy of the model increases as the SNR increases. When the SNR is high, such as 4 and 6 dB, the recognition accuracy of our method is 94.0% and 96.4%, respectively. The method proposed in this study can improve the recognition effect when the SNR is low, and compared with the other three methods, the average recognition accuracy under the high SNR is also improved.

To analyze the recognition performance of C-ResNet in detail, the confusion matrix of different methods at 4 dB is given. As illustrated by Fig. 6, when using VGG and ResNet, the accuracy of recognizing AM signals is only 40.5% and 72%, but it has been significantly improved in FFT+C-ResNet. In contrast, the recognition accuracy of the proposed scheme is highly competitive results.

Table IV summarizes the computational complexity analysis of different methods, summarizing the network capacity and the average inference time for running an epoch. C-ResNet and CNN1D have fewer parameters than VGG16 and ResNet. Because CNN1D has fewer convolution operations, it takes the lead in execution speed, which is 3.7 s. C-ResNet followed closely behind with 4.2 s. The above results proved that C-ResNet not only has the advantage of short execution time but also has good recognition ability.

The robustness of the BA method is tested in Fig. 7, the training set bandwidth is 0–10 kHz, and the bandwidth ranges of the verification signals are 0–10, 10–20, 20–30, 30–40, and 40–50 kHz. It can be observed that the increase in bandwidth degrades the recognition accuracy of the model. Under the same SNRs, compared with the recognition model that does not use the BA method, the model using this method has better robustness. When SNR is −6 dB, as the bandwidth increases, the recognition accuracy of C-ResNet has a significant downward trend, while the downward trend of C-ResNet+BA method in the bandwidth of 20–30, 30–40, and 40–50 kHz is more gentle, and it has a higher recognition accuracy.

In brief, the Complex-ResNet has high recognition accuracy on data sets with different SNRs and can be used as a detection and recognition model for aliased signals.

B. Detection Performance

For the frequency band where there is a signal, the type and specific location of the signal are determined jointly by the recognition model and the voting algorithm. The experiment tested the detection of different amplitude signals under different SNR. E in Fig. 8 is the detection threshold E_{high}, which is the minimum short-time energy in the detected signals, E_{low} is set to 0.05, and the number of signals is 300. To better reflect the accurate detection performance of the signal, the experiment adopts the recall rate as an indicator.

Because the large number of signals contained in the aliased signals, the SNR values of the aliased signals are unsuitable in this experiment. Once the SNR of the aliased signal is determined, the energy of the noise is unchanged and the energy of the aliased signal is the superposition of the energy of all single-component signals, so the impact of noise on the single signal will be relatively large. To control the degree of noise interference in the aliased signal, the single-component signal with the largest signal energy in the aliased signal is used as the calibration signal, and noise is added to the aliased signal based on the SNR of the calibration signal to ensure that the intercepted signal is affected. This study constructs a single-component signal data set containing various SNRs, and randomly extracts each signal composition from the data set with different SNRs in the data sets used in the above experiment.

As illustrated in Fig. 8, the recall rate increases as the SNR and detection threshold increase. Meanwhile, the recall rate of the signal with higher energy under −6 dB is even greater than the recall rate of the signal with smaller energy under 6 dB. The influence of the detection threshold on the detection performance is generally higher than the SNR on the detection performance. Therefore, in the subsequent experiments, the detection threshold $E >= 0.3$ was taken to ensure that the signal could be detected.

C. Multisignal Detection and Recognition Model Experiment Results

While keeping the trained model unchanged, experiments are carried out on the multicomponent aliased signal under the influence of noise with different energies, and the performance of the multisignal detection and recognition method is studied. The definition of SNR is consistent with Section III-B, and the

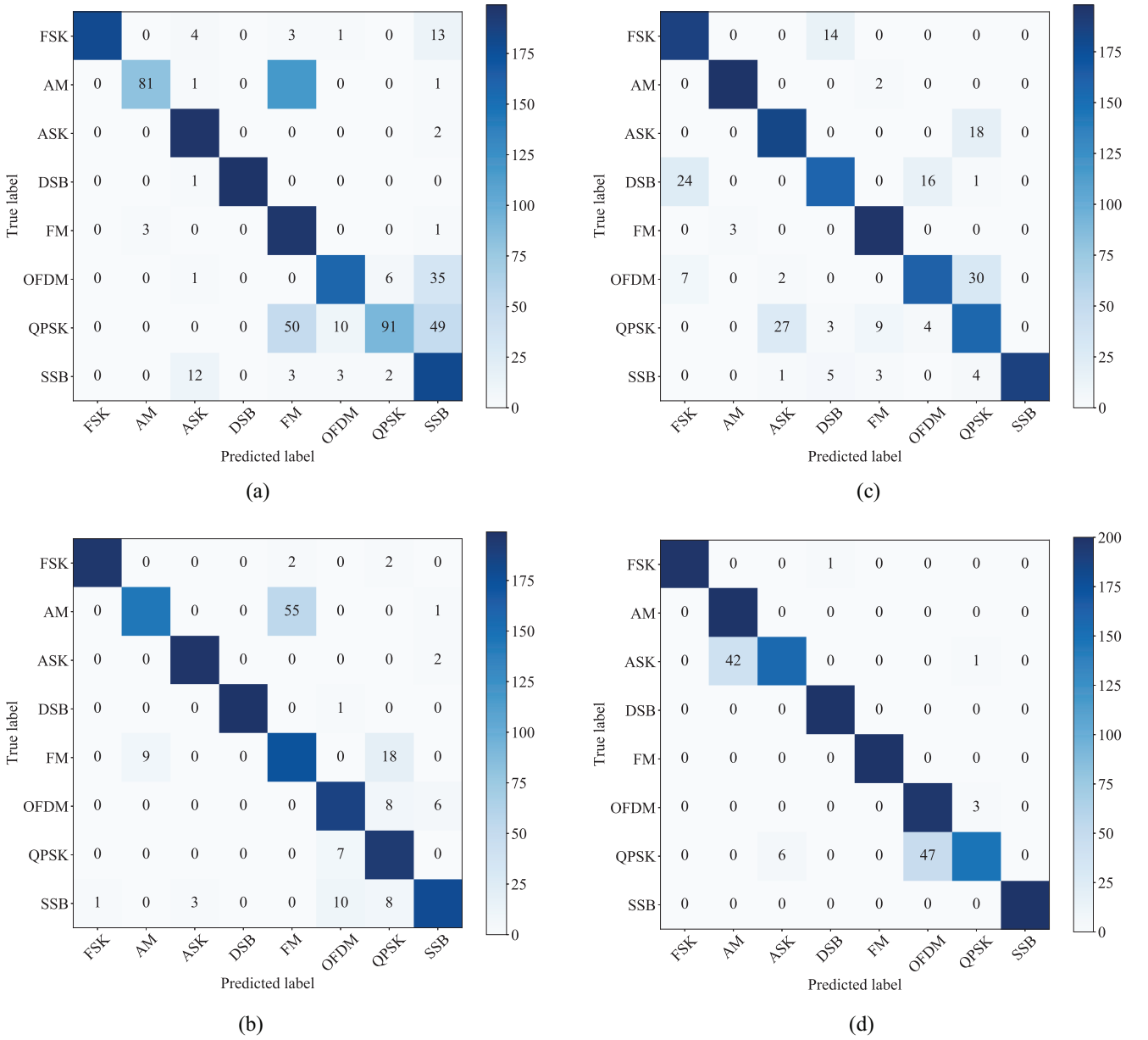


Fig. 6. Confusion matrices of different methods at 4 dB. (a) SPWVD+VGG. (b) SPWVD+Res. (c) FFT+CNN1D. (d) FFT+C-Res.

frequency band of the aliased multisignal is from 3 to 30 MHz. To maintain consistency, the frequency band was divided into 264 segments, then a certain type of modulation signal is randomly added to each frequency point. In the aliased signal, 33 numbers of each type of signal were added with random bandwidth (0–20 kHz), which constitute time-domain aliasing and frequency-domain separable or close multicomponent signals.

After the intercepted signals are obtained, they are output as images with the target detection and recognition methods, such as SSD, Yolo, and other models are used. However, the recognition accuracy of these methods is not satisfactory. Although the signal can be accurately framed, the accuracy of the signal category prediction is extremely low, even less than 50%. Compared with images, the method is more concerned with the amplitude and phase characteristics of the signal, and the recognition rate obtained is also higher. Table V lists the average recognition accuracy of the multisignal detection model after adding different noise to the aliased signal according to the SNR of the calibration signal with experiments.

It can be seen from the table that the performance of multisignal alias detection and recognition is consistent with Section III-A. After adding the noise corresponding to the SNR of the calibration signal, the recognition accuracy with the decreasing of the noise energy is added. When noise with an SNR of −4 dB is added, the detection and recognition effect of multiple signals is the worst with an average recognition accuracy rate of 82.0%. Meanwhile, the average recognition rate can reach 100.0% when the noise corresponding to 4 dB is added. Analyzing and comparing the recognition rate of various signals under the influence of noise with different energies, it can be seen that the recognition rate of ASK, AM, and SSB

TABLE V
RECOGNITION PERFORMANCE OF MULTISIGNAL DETECTION MODEL

Types	-4dB	-2dB	0dB	2dB	4dB	6dB
ASK	87.8%	97.0%	97.5%	97.6%	100.0%	100.0%
QPSK	92.3%	94.9%	95.1%	100.0%	100.0%	100.0%
FSK	98.0%	98.6%	98.6%	100.0%	100.0%	100.0%
OFDM	92.7%	100.0%	100.0%	100.0%	100.0%	100.0%
AM	80.5%	92.7%	100.0%	100.0%	100.0%	100.0%
SSB	4.9%	95.1%	97.6%	100.0%	100.0%	100.0%
DSB	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
FM	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Average accuracy	82.0%	97.3%	98.6%	99.7%	100.0%	100.0%

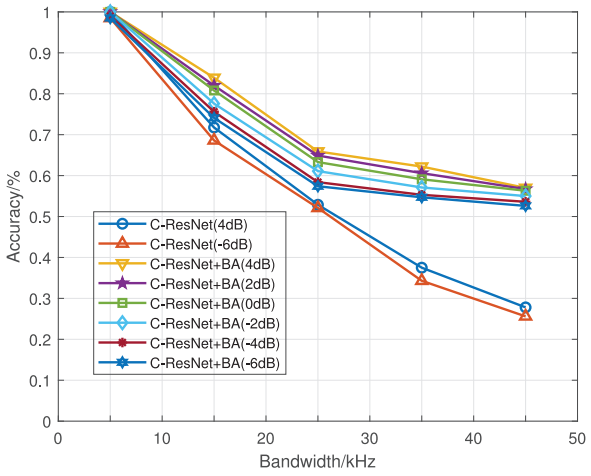


Fig. 7. Robustness to bandwidth change of proposed BA method.

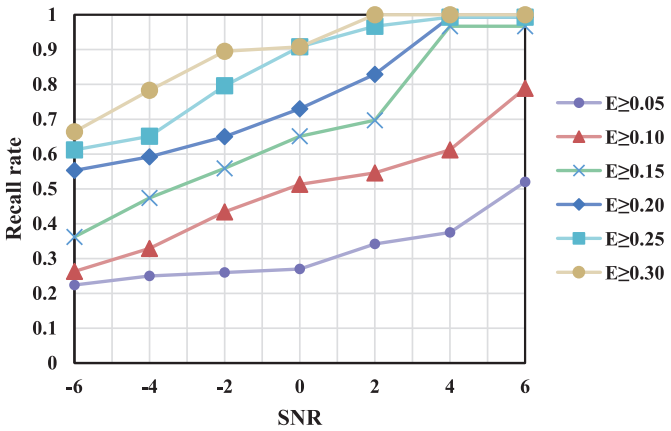


Fig. 8. Detection performance under different detection thresholds.

is not high at low SNRs. The reason lies not only in the high spectral similarity after FFT of each type signals in low SNRs but also signals greatly fluctuate near the main lobe, which makes it difficult to locate the sliding window and incomplete signal truncation.

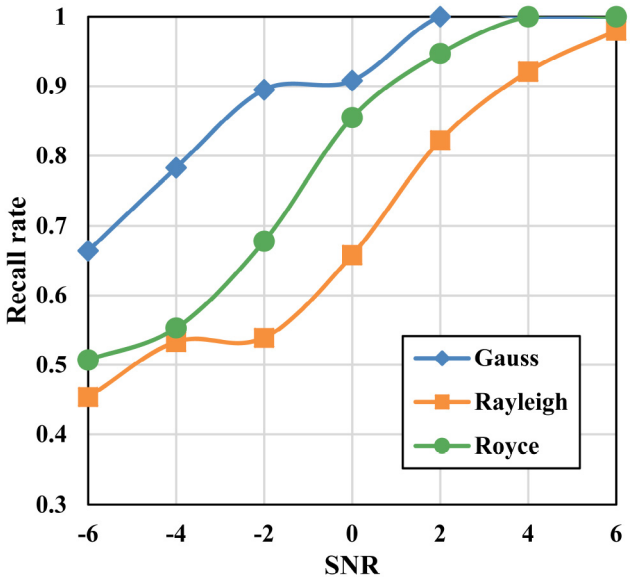


Fig. 9. Recognition performance under different fading.

Considering the fading effect in wireless communication, we construct two data sets, including Rayleigh channel and Rice channel, for testing to test the effect of fading on our proposed method. The same as the experimental conditions in Table V, it can be seen from the Fig. 9 that the recall rate of the model to the signal will decrease under either the Rice channel or the Rayleigh channel, and when the SNR increases, the accuracy rate can remain consistent.

Furthermore, Fig. 10 shows the frequency-domain distribution and recognition probability of various signals in the multicomponent signal, which adds the corresponding noise with the calibration signal in 0 dB. The subpicture contains the true values of the signals and the results of model prediction. For OFDM, AM, DSB, and FM signals, the predicted results of the model are consistent with the true value. At the same time, no false alarms are generated in other frequency points.

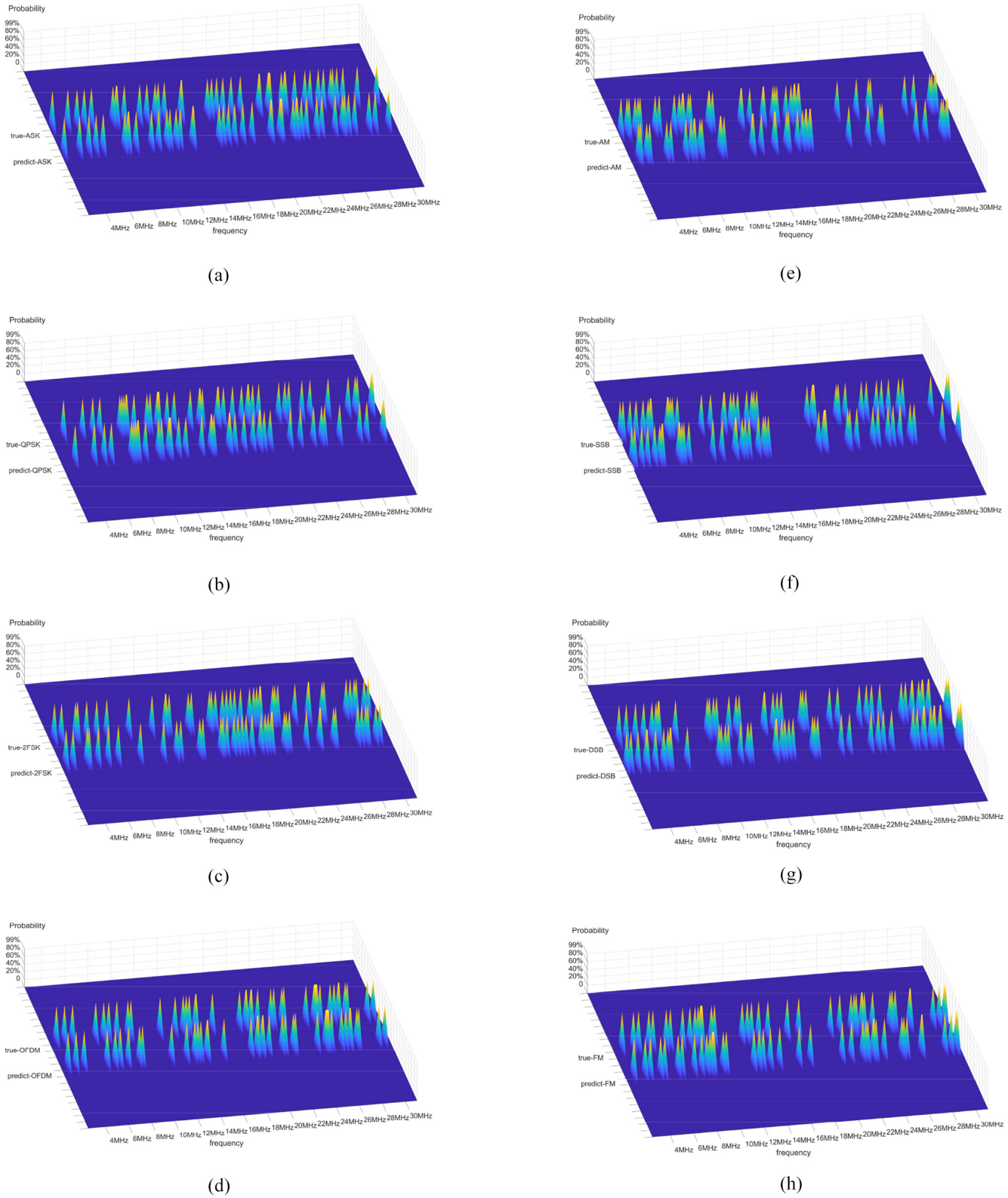


Fig. 10. Distribution of various signals and recognition probability. (a) ASK. (b) QPSK. (c) FSK. (d) OFDM. (e) AM. (f) SSB. (g) DSB. (h) FM.

But for ASK, QPSK, FSK, and SSB, although our method can locate the signals' position, it occasionally causes misjudgment. Meanwhile, some signals also have the probability of being recognized at other frequency points in low SNRs, and the probability of being recognized at several frequency points is high, such as SSB. Comparing the frequency points of the QPSK and ASK signals, the multisignal detection and recognition model is easy to confuse one type of signal with the other types, which is also consistent with the experimental results in Section IV-A.

IV. Conclusion

This study proposes a multicomponent signal detection and recognition method based on Complex CNN for the CR-IoT. When the number of signals is unknown, this method uses the amplitude and phase information of the signals after FFT, detecting and recognizing the aliased signals which are aliasing in the time domain and closely adjacent to the frequency domain. The Complex-ResNet used in this study can effectively recognize multiple types of signals, and the average recognition accuracy rate at 4 dB can reach 94.01%. The simulation results prove the effectiveness of the method, where the average recognition rate can reach 97.3% under the influence of the noise corresponding to the SNR of the calibration signal at -2 dB. Compared with other methods, this method can detect and recognize many signals and ensure high accuracy. However, the actual signals are greatly affected by noise, and the available information may be more submerged in the noise. When aliasing occurs in the frequency domain, the signal cannot be accurately detected by the sliding window, resulting in a reduced recognition rate. How to complete the accurate detection and signal recognition in very low SNR environments and the frequency domain aliasing are the problems that must be addressed in future research. For signal types that can be easily confused, analyzing the black-box characteristics of DL model through explainable artificial intelligence (XAI) techniques [34], [35], and improving the reliability of the model recognition results to reduce the probability of false positives are also directions that need to be further considered in this research.

References

[1] Y. Zhang, S. Chen, Y. Zhou, Y. Fang, and C. Qian, "Monitoring bodily oscillation with RFID tags," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3840–3854, Apr. 2019.

[2] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *Proc. ICML*, 2011. [Online]. Available: https://icml.cc/2011/papers/399_icmlpaper.pdf

[3] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescapè, "Mimetic: Mobile encrypted traffic classification using multimodal deep learning," *Comput. Netw.*, vol. 165, Dec. 2019, Art. no. 106944.

[4] F. Restuccia, S. D'Oro, and T. Melodia, "Securing the Internet of Things in the age of machine learning and software-defined networking," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4829–4842, Dec. 2018.

[5] L. Da Xu, W. He, and S. Li, "Internet of Things in industries: A survey," *IEEE Trans. Ind. Informat.*, vol. 10, no. 4, pp. 2233–2243, Nov. 2014.

[6] S. Sicari, A. Rizzardi, L. A. Grieco, and A. Coen-Porisini, "Security, privacy and trust in Internet of Things: The road ahead," *Comput. Netw.*, vol. 76, pp. 146–164, Jan. 2015.

[7] S. Huang *et al.*, "Automatic modulation classification using gated recurrent residual network," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 7795–7807, Aug. 2020.

[8] N. Bitar, S. Muhammad, and H. H. Refai, "Wireless technology identification using deep convolutional neural networks," in *Proc. IEEE 28th Annu. Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, 2017, pp. 1–6.

[9] Y. Lin, Y. Tu, Z. Dou, L. Chen, and S. Mao, "Contour Stella image and deep learning for signal recognition in the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 1, pp. 34–46, Mar. 2021.

[10] Z. Zhang, W. Zhang, H.-C. Chao, and C.-F. Lai, "Toward belief function-based cooperative sensing for interference resistant industrial wireless sensor networks," *IEEE Trans. Ind. Informat.*, vol. 12, no. 6, pp. 2115–2126, Dec. 2016.

[11] Y. Xing, A. Hu, J. Zhang, J. Yu, G. Li, and T. Wang, "Design of a robust radio-frequency fingerprint identification scheme for multimode LFM radar," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 10581–10593, Oct. 2020.

[12] Y. Liu, X. Kuai, X. Yuan, Y.-C. Liang, and L. Zhou, "Learning-based iterative interference cancellation for cognitive Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 7213–7224, Aug. 2019.

[13] M. H. Kapourchali and B. Banerjee, "Unsupervised feature learning from time-series data using linear models," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3918–3926, Oct. 2018.

[14] T. Yu, X. Wang, and A. Shami, "Recursive principal component analysis-based data outlier detection and sensor data aggregation in IoT systems," *IEEE Internet Things J.*, vol. 4, no. 6, pp. 2207–2216, Dec. 2017.

[15] V. Angelakis, I. Avgouleas, N. Pappas, E. Fitzgerald, and D. Yuan, "Allocation of heterogeneous resources of an IoT device to flexible services," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 691–700, Oct. 2016.

[16] Y.-C. Liang, K.-C. Chen, G. Y. Li, and P. Mahonen, "Cognitive radio networking and communications: An overview," *IEEE Trans. Veh. Technol.*, vol. 60, no. 7, pp. 3386–3407, Sep. 2011.

[17] 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.

[18] W. Wei and J. M. Mendel, "Maximum-likelihood classification for digital amplitude-phase modulations," *IEEE Trans. Commun.*, vol. 48, no. 2, pp. 189–193, Feb. 2000.

[19] F. Hameed, O. A. Dobre, and D. C. Popescu, "On the likelihood-based approach to modulation classification," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 5884–5892, Dec. 2009.

[20] P. Panagiotou, A. Anastasopoulos, and A. Polydoros, "Likelihood ratio tests for modulation classification," in *Proc. 21st Century Mil. Commun. Archit. Technol. Inf. Superiority (MILCOM)*, vol. 2, 2000, pp. 670–674.

[21] A. K. Nandi and E. E. Azzouz, "Algorithms for automatic modulation recognition of communication signals," *IEEE Trans. Commun.*, vol. 46, no. 4, pp. 431–436, Apr. 1998.

[22] C. L. Haley and M. Anitescu, "Optimal bandwidth for multitaper spectrum estimation," *IEEE Signal Process. Lett.*, vol. 24, no. 11, pp. 1696–1700, Nov. 2017.

[23] H. Yousefi, S. Askari, G. A. Dumont, and Z. Bastany, "Automated decomposition of needle EMG signal using STFT and wavelet transforms," in *Proc. 21st Iranian Conf. Biomed. Eng. (ICBME)*, 2014, pp. 358–363.

[24] C. Trabelsi *et al.*, "Deep complex networks," 2018, *arXiv:1705.09792*.

[25] H. Wu, Y. Li, L. Zhou, and J. Meng, "Convolutional neural network and multi-feature fusion for automatic modulation classification," *Electron. Lett.*, vol. 55, no. 16, pp. 895–897, 2019.

[26] Z. Zhang, H. Luo, C. Wang, C. Gan, and Y. Xiang, "Automatic modulation classification using CNN-LSTM based dual-stream structure," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13521–13531, Nov. 2020.

[27] S. Huang, C. Lin, W. Xu, Y. Gao, Z. Feng, and F. Zhu, "Identification of active attacks in Internet of Things: Joint model- and data-driven automatic modulation classification approach," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 2051–2065, Feb. 2021.

[28] M. Wu, D. Wang, and G. J. Brown, "A multipitch tracking algorithm for noisy speech," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 3, pp. 229–241, May 2003.

[29] G.-J. Jang, T.-W. Lee, and Y.-H. Oh, "Single-channel signal separation using time-domain basis functions," *IEEE Signal Process. Lett.*, vol. 10, no. 6, pp. 168–171, Jun. 2003.

[30] J. Gao, L. Shen, and L. Gao, "Modulation recognition for radar emitter signals based on convolutional neural network and fusion features," *Trans. Emerg. Telecommun. Technol.*, vol. 30, no. 12, p. e3612, 2019.

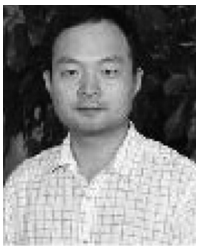
[31] Y. Yuan, Z. Sun, Z. Wei, and K. Jia, "DeepMorse: A deep convolutional learning method for blind morse signal detection in wideband wireless spectrum," *IEEE Access*, vol. 7, pp. 80577–80587, 2019.

[32] Z. Liu, L. Li, H. Xu, and H. Li, "A method for recognition and classification for hybrid signals based on deep convolutional neural network," in *Proc. Int. Conf. Electron. Technol. (ICET)*, 2018, pp. 325–330.

[33] Z. Qu, C. Hou, C. Hou, and W. Wang, "Radar signal intra-pulse modulation recognition based on convolutional neural network and deep Q-learning network," *IEEE Access*, vol. 8, pp. 49125–49136, 2020.

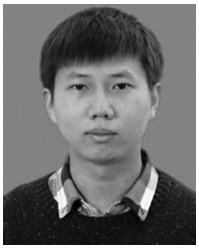
[34] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

[35] A. Nascita, A. Montieri, G. Aceto, D. Ciuonzo, V. Persico, and A. Pescapé, "Xai meets mobile traffic classification: Understanding and improving multimodal deep learning architectures," *IEEE Trans. Netw. Service Manag.*, vol. 18, no. 4, pp. 4225–4246, Dec. 2021.



Changbo Hou received the B.S. and M.S. degrees from the College of Information and Communication Engineering, Harbin Engineering University, Harbin, Heilongjiang, China, in 2008 and 2011, respectively.

He is currently an Associate Professor with the College of Information and Communication Engineering, Harbin Engineering University, where he is also a Doctor with the Key Laboratory of In-Fiber Integrated Optics, Ministry Education of China. His research interests include wideband signal processing, image processing, and deep learning.

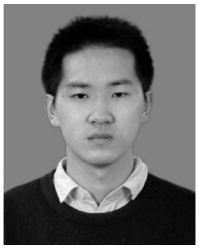


Guowei Liu received the B.S. degree from the College of Information and Communication Engineering, Harbin Engineering University, Harbin, Heilongjiang, China, in 2020, where he is currently pursuing the master's degree in electronic communication engineering.

His research interests include edge computing, deep learning, and signal processing.

Qiao Tian received the B.S. and Ph.D. degrees from Harbin Engineering University, Harbin, China, in 2014 and 2019, respectively.

She is currently a Lecturer with the College of Computer Science and Technology, Harbin Engineering University. Her current research interests include information and system security, architecture performance optimization, infinite communication, and artificial intelligence.



Zhichao Zhou received the B.S. degree in electronic information engineering from Harbin Engineering University, Harbin, Heilongjiang, China, in 2019.

His research interests include edge computing, deep learning, and signal processing.

Lijie Hua received the B.S. degree from the College of Information and Communication Engineering, Harbin Engineering University, Harbin, Heilongjiang, China, in 2019, where she is currently pursuing the master's degree in electronic communication engineering.

Her research interests include signal processing, image processing, and deep learning.

Yun Lin (Member, IEEE) received the B.S. degree from Dalian Maritime University, Dalian, China, in 2003, the M.S. degree from Harbin Institute of Technology, Harbin, China, in 2005, and the Ph.D. degree from Harbin Engineering University, Harbin, in 2010.

He was a Research Scholar with Wright State University, Dayton, OH, USA, from 2014 to 2015. He is currently a Full Professor with the College of Information and Communication Engineering, Harbin Engineering University. He had published more than 200 international peer-reviewed journal/conference papers, such as the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE INTERNET OF THINGS JOURNAL, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING, TR, INFOCOM, GLOBECOM, ICC, VTC, and ICNC. His current research interests include machine learning and data analytics over wireless networks, signal processing and analysis, cognitive radio and software-defined radio, artificial intelligence, and pattern recognition.

Prof. Lin is serving as an Editor for the IEEE TRANSACTIONS ON RELIABILITY, *Digital Communications and Networks*, *KSII Transactions on Internet and Information Systems*, and *International Journal of Performability Engineering*. In addition, he served as the General Vice Chair of VTC-2021 Fall, the General Chair of ADHIP 2020, the TPC Chair of MOBIMEDIA 2020, ICEICT 2019, and ADHIP 2017, and the TPC Member of GLOBECOM, ICC, ICNC, and VTC. He had successfully organized several international workshops and symposia with top-ranked IEEE conferences, including INFOCOM, GLOBECOM, DSP, and ICNC.