# Research on a Convolutional Neural Network Method for Modulation Waveform Classification

Ren Guo-Xi

*Abstract*—**Modulated signal recognition is difficult but essential for applications like cognitive radio, intelligent communication, radio supervision, and electronic countermeasure. Current modulation recognition models lack comprehensiveness and typicality of various signals and primarily rely on artificial feature extraction. In this study, a convolutional neural network (CNN)-based method for modulated signal recognition is proposed. The proposed method converts modulation recognition into image identification. To increase the acuity of CNN for learning time-frequency features, channel attention and spatial attention are further introduced based on the fused features. Eight different types of modulated signals, including Rect, LFM, Barker, GFSK, CPFSK, B-FM, DSB-AM, and SSB-AM, are used in the experiments. The recognition rate of the proposed model is greater than 85% when the SNR (signal-to-noise ratio) is greater than -10dB, and it ranges from 92% to 98% when the SNR is 0dB. The recognition rate of the proposed method outperforms the two other comparison methods, CNN without an attention mechanism and LSTM.**

*Index Terms*—**modulation recognition, convolutional neural network, feature fusion, attention mechanism, STFT, SPWVD**

## I. Introduction

MODULATION recognition is to detect the type of the modulated signal. It has a wide range of practical applications. For cognitive radio, modulation recognition aids sub-users in understanding the detailed signal information of the primary user [1,2]. It is an important part in the overall understanding of the wireless environment. For intelligent communication, modulation recognition can not only reduce the transmission overhead of modulated information but also enable blind demodulation of the signal [3,4]. For radio supervision, modulation recognition is useful in determining whether the signal meets radio management regulations [5,6]. For electronic countermeasure, modulation recognition is a prerequisite for effective implementation of radio jamming and spoofing [7,8].

Modulation recognition algorithms are mainly divided into two categories: maximum likelihood hypothesis testing (MLHT) methods based on decision theory [9,10] and pattern recognition (PR) methods based on feature extraction [11,12]. The MLHT methods employ Bayesian theory of probability and hypothesis testing to perform the modulation recognition. Although the classification results of MLHT methods are optimal in terms of Bayesian estimation, they are computationally intensive and mostly target baseband signals. However, the signal is a carrier modulated in most application scenarios. As a result, MLHT methods are less widely used and will not be discussed in the subsequent paper. PR methods are widely used in modulation recognition. An evaluation of artificial intelligence algorithms for identifying modulation format was published in literature [13]. Various identification techniques, including k-nearest neighbors (KNN), support-vector machines (SVM), were used. The SVM algorithm achieved the most robust performance based on the experimental results. A simple approach based on photonic reservoir computing was presented in literature [14]. It was used for modulation format identification (MFI) in the field of optical fiber communication. The final simulation results demonstrated that this technique was capable of accurately identifying modulation formats with a precision of 95%. The aforementioned techniques recognize modulations by extracting shallow features, which heavily rely on the experience of researchers. These techniques have a limited range of applications and are difficult to extract accurate features. As a result, the automatic extraction of deep features has become a research hotspot.

Deep learning is trained to create a multi-layer neural network using lots of sample data [15]. It extracts the hidden features of the sample and reduces the difficulty of designing features manually. Convolutional Neural Network (CNN), as a typical deep learning method, has been applied for modulation recognition. CNN was used in literature [16] to identify the eye diagram of signals. The experiment result could achieve 100% recognition rate for four modulated signals in a wide SNR range. Wavelet transformation and STFT were applied to the signal in literature [17]. And manifold learning method was employed to reduce the high dimension and extract the recognition feature. An automatic radar waveform recognition was proposed in literature [18] using Wigner-Ville distribution to extract the features of the original signal. Smoothed pseudo Wigner-Ville distribution and density peaks clustering for blind model estimation were used in literature [19]. The experimental results demonstrated the effectiveness of their proposed methods.

The combination of time-frequency analysis and CNN are a good attempt for modulation recognition [20]. It improves the accuracy and robustness of the recognition model. Different time-frequency analysis methods have their inherent characteristics. Short-time Fourier transform (STFT) is suitable for multicomponent signal analysis. Wigner-Ville distribution (WVD) is a two-dimensional distribution of signal energy in the time-frequency domain. WVD has a good time-frequency cohesiveness. However, WVD is heavily affected by crossing terms and noise. In order to overcome the shortcomings of STFT and WVD, the time-frequency maps of STFT and WVD are fused. The

Manuscript received January 19, 2023; revised May 28, 2023.

Ren Guo-Xi is an associate professor of Pingdingshan Polytechnic College, Pingdingshan, 467003, China (e-mail: 119887419@qq.com).

fused maps are used as the CNN input. Based on this, channel attention and spatial attention are introduced into CNN to learn important information for modulation recognition.

The rest of this paper is organized as follows. Section 2 explains the basic concepts of modulated signals and time-frequency characteristics. Section 3 introduces the CNN. In Section 4, the improved CNN is proposed based on multi-feature fusion and attention mechanism. The experimental results are discussed in Section 5. Finally, the conclusions are given in Section 6.

## II. BASICS OF MODULATION RECOGNITION

### A. Modulated signal

In this paper, eight commonly used modulation signals are selected for the experiments. There are three radar modulation signals: rectangular (Rect), linear frequency modulation (LFM), and barker code (Barker). There are also five communication modulation signals: Gaussian frequency shift keying (GFSK), continuous phase frequency shift keying (CPFSK), broadcast frequency modulation (B-FM), double side band amplitude modulation (DSB-AM), and single side band amplitude modulation (SSB-AM). The baseband waveform domain of the signal is as follows:

$$y(t)=\sum_i x_i f(t-iT)+n(t) \tag{1}$$

where $n(t)$ denotes the channel noise, which is usually assumed to be additive Gaussian white noise. $f(t)$ denotes the equivalent filters, such as shaping filter, channel filter, and matched filter. $x_i$ denotes the transmitter symbol sequence.

### B. Time-frequency analysis

Currently, the main time-frequency transform methods are short-time Fourier transform (STFT), Wigner-Ville distribution (WVD), pseudo Wigner-Ville distribution (PWVD), smoothed pseudo Wigner-Ville distribution (SPWVD), etc. STFT has low time-frequency resolution but no cross-term effect. Meanwhile WVD has serious cross-term and high time-frequency resolution. PWVD and SPWVD have some suppression effect on cross-term by adding a window function. However, the time-frequency resolution of PWVD and SPWVD decreases significantly compared with WVD.

(1) Short-time Fourier transform (STFT)

Given a window function $g(t)$ with a short time width. Let the window function $g(t)$ slide on the $t$-axis. Then the STFT of the signal $x(t)$ is defined as follows:

$$STFT(t,f)=\int x(\tau)g(\tau-t)e^{-jf\tau}\mathrm{d}\tau \tag{2}$$

where the physical meaning of equation (2) is the Fourier transform of the signal $x(t)$ multiplied by a $t$-centered analysis window $g(\tau-t)$. $STFT(t,f)$ is a function of both time and frequency.

STFT is a linear representation of the signal. STFT has no interference of cross-terms. It is suitable for multicomponent signal analysis. The resolution performance of STFT depends heavily on the window function and width. The actual signals are generally non-stationary. STFT, on the other hand, assumes that the signal is approximately smooth within the width of the window function. The type of window function is usually chosen as a low-pass one, such as Gaussian window, Hanning window, etc. The time-frequency aggregation of the window function is also determined once it has been selected. According to the uncertainty principle, the time width and bandwidth of window function cannot be arbitrarily small at the same time. As a result, its aggregation of time and frequency is poor. For non-stationary signals, STFT has no self-interference term. But the aggregation of STFT is poor.

(2) Smoothed pseudo Wigner-Ville distribution (SPWVD)

Wigner-Ville distribution is one of the most basic nonlinear representations. The WVD of the signal $x(t)$ is defined as follows:

$$W(t,f)=\int x^*(t-\frac{\tau}{2})(t+\frac{\tau}{2})e^{-jwt}\mathrm{d}\tau \tag{3}$$

WVD has good time-frequency resolution. Set

$$x(t)=x_1(t)+x_2(t) \tag{4}$$

Then the follows can be obtained.

$$\begin{aligned}W(t,w)&=\frac{1}{2\pi}\int x^*(t-\frac{\tau}{2})(t+\frac{\tau}{2})e^{jwt}\mathrm{d}\tau\\&=\frac{1}{2\pi}\int\left(x_1^*(t-\frac{\tau}{2})+x_2^*(t-\frac{\tau}{2})\right)\cdot\\&\quad\left(x_1(t+\frac{\tau}{2})+x_2(t+\frac{\tau}{2})\right)e^{jwt}\mathrm{d}\tau\\&=W_{11}(t,w)+W_{12}(t,w)+W_{22}(t,w)+W_{21}(t,w)\end{aligned} \tag{5}$$

where

$$W_{12}(t,w)=\frac{1}{2\pi}\int x_1(t-\frac{\tau}{2})x_2(t+\frac{\tau}{2})e^{jwt}\mathrm{d}\tau \tag{6}$$

$$W_{21}(t,w)=\frac{1}{2\pi}\int x_2(t-\frac{\tau}{2})x_1(t+\frac{\tau}{2})e^{jwt}\mathrm{d}\tau \tag{7}$$

$$W_{12}(t,w)=W_{21}{}^*(t,w) \tag{8}$$

From the above equation, WVD can be regarded as the Fourier transform of the signal time autocorrelation function. Since it is the quadratic time-frequency of the signal, there are serious cross-terms in WVD when performing time-frequency analysis on multiple signals. Many existing time-frequency analysis methods are basically a compromise between multi-portion cross-term suppression and signal time-frequency aggregation. In fact, even for the single component signal, WVD has its own interference term.

Although the WVD and PWVD have strong time-frequency focus, they are both quadratic time-frequency analysis methods. They are vulnerable to the interference of cross-terms. The performance of parameter estimation is seriously affected. Therefore, the SPWVD algorithm is used. The window is added simultaneously in the time-frequency domain, which suppresses the interference of cross-terms. SPWVD is defined as follows:

$$\begin{aligned}&\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x(u+\frac{\tau}{2})x^*(u-\frac{\tau}{2})\varphi(\tau,v)\cdot\\&\qquad e^{-j2\pi(tv+tf-uv)}\mathrm{d}u\mathrm{d}v\mathrm{d}\tau\end{aligned} \tag{9}$$

where $\varphi(\tau,v)$ is the smoothing function in the time frequency domain.

## III. CONVOLUTIONAL NEURAL NETWORK

CNN is a typical feed-forward neural network. It can be trained to adjust the weights and biases of the neurons. In recent years, several CNN structures have been developed, such as LeNet, AlexNet and GoogleNet. CNN mainly consists of convolutional layers, pooling layers and fully connected layers.

In the convolution layer, the feature map of the previous layer is convoluted with the convolution kernel. Then the output feature map is formed by the activation function. The process is computed as follows:

$$S_i^j = f(\sum_{l \in M_i} S_l^{j-1} * \boldsymbol{w}_{li}^j + \boldsymbol{b}_l^j) \tag{10}$$

where $M_i$ is the set of input feature maps. $j$ is the $j$th layer of the network; $S_l^{j-1}$ is the $l$th input feature map. $w$ is the corresponding convolution kernel weight matrix; $b$ is the bias matrix. $S_i^j$ is the $i$th feature map of the $j$th layer. $f()$ is the activation function. * is the convolution operation.

The Sigmoid function is a nonlinear activation function that is commonly used in neural networks. However, due to its gradient vanishing problem, the ReLU function is commonly used in deep learning models. The mathematical expression is as follows:

$$f(n) = \max(0, n) \tag{11}$$

Using the ReLU function can not only avoid gradient disappearance of back-propagating but also reduce the computational cost of the CNN. The ReLU function can also make the output of some neurons zero, i.e., the network becomes more sparse. As a result, the problem of overfitting is avoided.

The pooling layer can reduce the feature map dimension of the previous layer. Thus, the spatial dimension of the input convolution layer is reduced. The overfitting is effectively controlled. And the computational cost is reduced.

The common pooling functions are maximum pooling and mean pooling. In the maximum pooling, the maximum value of the restricted region is selected as the new feature. In the average pooling, the average value of the same region is calculated as the new feature. The maximum pooling reflects the most significant features. And the average pooling smoothes the region and selects the smoothed features. After $m \times m$ pooling, the number of the feature map is unchanged. But the size is reduced to 1/m of the original one.

The feature map matrix is transformed into an one-dimensional feature vector in the fully-connected layer, which is treated as the input of classifier. Its model can be summarized as follows:

$$P = f(\boldsymbol{b}_0 + \boldsymbol{w}_0 \cdot \boldsymbol{f}_v) \tag{12}$$

The Softmax function is usually used as the activation function $f()$ in multiclassification problems.

The difference between the predicted value $z$ and the actual value $y$ is obtained by forward propagation. The difference is passed to the CNN model by the Back Propagation (BP) algorithm. The loss function of the model is minimized by adjusting the weights and biases among the network layers. The common type of loss functions is mean-squared error (MSE).

In the process of error back propagation, the gradient descent method is used to update the network parameters. The adaptive parameters $w$ and $b$ are adjusted layer by layer by calculating the derivatives of the MSE. The error in each layer is reduced to a minimum. The calculation method is as follows:

$$\begin{cases} w' = w - \beta \dfrac{\partial E}{\partial w} \\ b' = b - \beta \dfrac{\partial E}{\partial b} \end{cases} \tag{13}$$

where $w'$ and $b'$ are the updated weights and biases. $w$ and $b$ are the existing weights and biases. $\beta$ is the learning rate, which is used to control the step size of the weight update. The network may fall into local optimum if $\beta$ is too large. On the contrary, the training time of the network may increase if $\beta$ is too small.

## IV. Improved CNN

### A. *Multi-feature fusion*

The modulated signal is passed through STFT and SPWVD respectively. And a STFT feature map and a SPWVD feature map are obtained. The two feature maps are fed into two independent CNNs for feature extraction. The convolutional kernels of different sizes are included in each single-channel CNN, which can extract various modulation pattern features from different images. Then the two feature vectors extracted from each single CNN are stitched together through the fusion layer. A high-dimensional feature vector is formed and fed into the fully connected layer. The single-channel CNN of extracting modulation features can be divided into eight hidden layers. The input layer is STFT feature map or SPWVD feature map. In order to improve the extraction of abstract features and enhance the SNR of the image, five convolutional layers are introduced. Abstract features, such as edge, texture, and shape, are extracted from the sample data by low-level convolution. Then new features are generated by high level convolution for subsequent recognition. To avoid losing the key information in the feature reduction process and to compensate for the lack of image feature invariance, three pooling layers are adopted. The input feature map is divided into several non-overlapping matrices. The maximum value of each matrix region is output. The computational complexity of the feature map is reduced, as is the data size. As a result, good displacement robustness is obtained by the features. To reduce the possibility of gradient disappearance, five normalization layers and ReLU activation layers are introduced in these eight hidden layers. Finally, the features of the last pooling layer are extracted to complete the subsequent feature fusion work.

To make full use of the different channel information, the features extracted from two independent CNNs are fused. The commonly used feature fusion methods are superposition fusion and vector stitching. The vector stitching method is used in this study. Suppose the existing feature vectors are $v_1 \in R^n$ and $v_2 \in R^m$, which are stitched in the same dimension. The final fused feature vector is $v = [v_1 + v_2] \in R^{n+m}$.

The feature fusion strategy can not only improve the feature representation but also avoid the negative impact of redundant feature on the information. As a result, the feature description of the modulated signal is rich and comprehensive.

### B. *Attention mechanism*

With the increasing application of deep learning in image classification, the attention mechanism is increasingly used to optimize the deep network structure. It is more like the way the human eye sees things. It allows the network to learn in a more focused way and improves the network learning ability. Similar to human observation, the network learns in a more focused way. Attention mechanism usually includes channel attention (CA) and spatial attention (SA).

For a given feature map $X$ with $L \times H \times W \times C$. $L$, $H$ and $W$ represent the spatial dimension of the feature map, while $C$ represents the number of channels. The principle of CA is shown in equations (14) to (16).

$$C_M = FC\left(\mathrm{ReLU}\left(FC\left(MaxPooling(X)\right)\right)\right) \tag{14}$$

$$C_A = FC\left(\delta\left(FC\left(AvgPooling(X)\right)\right)\right) \tag{15}$$

$$X_C = \mathrm{Sigmoid}(C_M + C_A)X \tag{16}$$

where $MaxPooling()$ and $AvgPooling()$ represent global maximum pooling and global average pooling in the spatial direction, respectively. ReLU() and Sigmoid() are the activation functions. $FC()$ is the fully connected layer.

For a given feature map $X$ with $L \times H \times W \times C$, the principle of SA is shown in equations (17) to (19).

$$S_M = MaxPooling(X) \tag{17}$$

$$S_A = AvgPooling(X) \tag{18}$$

$$X_S = \mathrm{Sigmoid}\left(Conv\left(Concat(S_M, S_A)\right)\right) \tag{19}$$

CA can be treated as channel weight. Channels containing important information are weighted heavily, while channels containing unimportant information are weighted less. The CA feature vector is input to each channel of the input image in the form of a broadcast. The CA feature map of the output is obtained. The important features of the channel and spatial can be focused on using CA and SA separately. And the unimportant features are filtered out. The attention mechanism is inserted in each convolutional block shown in Figure 1. The CA is first used to emphasize what the feature is, according to the intuitive interpretation. The SA is then used to emphasize the location of the feature.

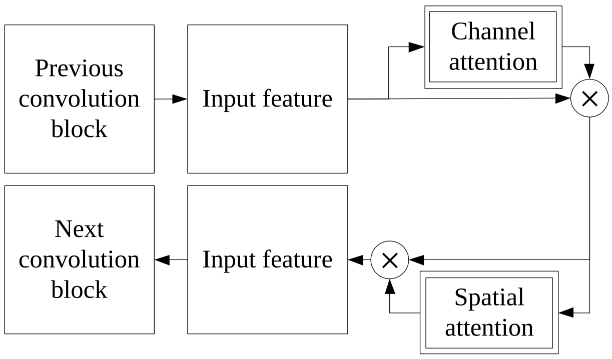



Fig. 1. Attention mechanism flow

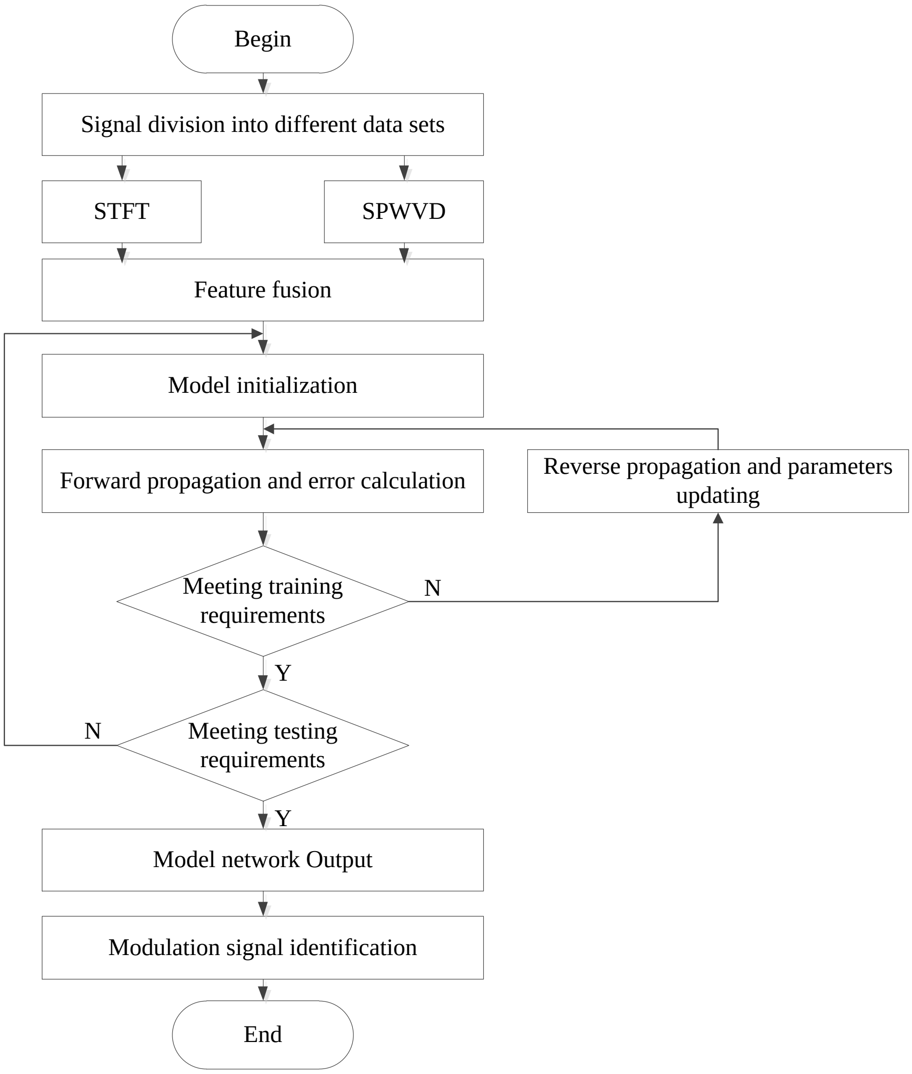



Fig. 2. Modulation recognition flow

### *C. Modulation recognition flow*

In this paper, the time-frequency features of the modulated signal are extracted by STFT and SPWVD. Then the two time-frequency features are fused. And fused features are input into a CNN model with the attention mechanism for supervised training. The trained model is used for modulation recognition. The steps for identifying are as follows, shown in Figure 2.

(1) The data set composed of modulated signals is divided into training set, validation set and testing set.

(2) The modulated signals are calculated by STFT and SPWVD respectively. The STFT based and SPWVD based time-frequency maps are obtained. The two time-frequency maps are fused using the method in Section 4.1. The input feature maps for CNN are obtained.

(3) The network parameters are initialized. The training parameters are set, such as the number of epochs, the size of batch, and the learning rate.

(4) The fused feature maps are fed into the CNN with attention mechanism in a batch manner. After the forward propagation, the error between the desired output and the actual output is calculated.

(5) The error is back-propagated with BP algorithm to update the network parameters layer by layer.

(6) Steps (4) and (5) are repeated until the iteration is completed or until the network accuracy requirement is met. Then the ideal CNN model is obtained.

(7) Testing samples are set to the trained model for estimating the model performance. The accuracy whether meeting the actual requirements is judged.

(8) The output network is used for modulation recognition.

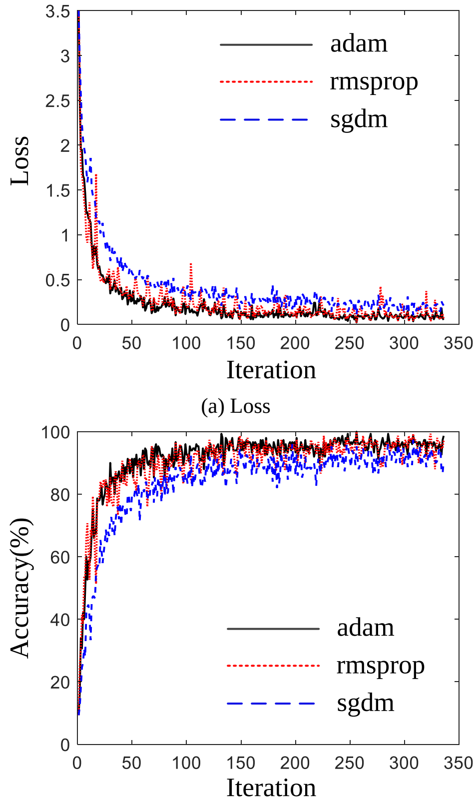
## V. EXPERIMENTS

### *A. Experimental setup*

In this experiment, eight commonly used modulated signals are utilized to verify the performance of the proposed method. The eight modulated signals are Rect, LFM, Barker, GFSK, CPFSK, B-FM, DSB-AM, and SSB-AM. The noise environment is additive Gaussian white noise. The experiments are all trained in batches. The training set, validation set and testing set is divided into multiple batches. The size of each training batch is set to 32, i.e., 32 samples are input for processing in each batch.

### *B. Training process analyzing*

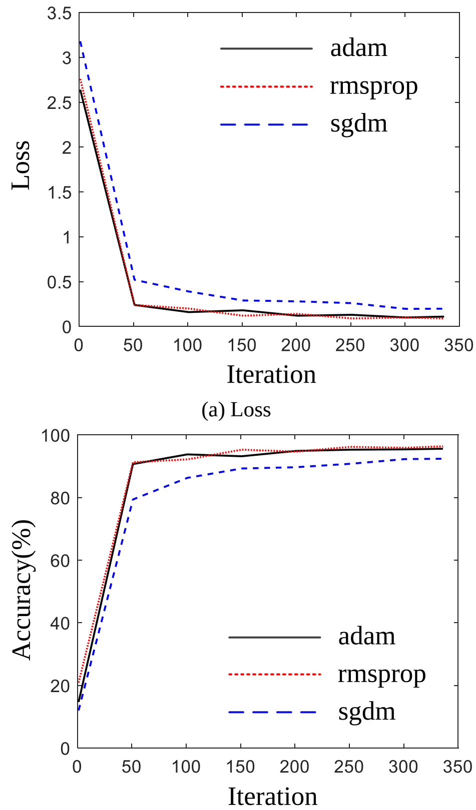
Considering the gradient error during the backpropagation of the network, experiments are conducted to find a training optimizer applicable to CNN. The three commonly used optimizers, stochastic gradient descent with momentum (sgdm), root mean square prop (rmsprop) and adaptive moment estimation (adam), are selected for the experiment. To evaluate the training performance of the recognition models, the training and validation dataset is used. The training results are shown in Figure 3. The loss value and accuracy variation curves of the training samples corresponding to the three optimizers are shown in Figure 3(a) and Figure 3(b), respectively. It can be seen that the accuracy corresponding to the three optimizers fluctuates and increases as the loss value decreases.




(a) Loss

(b) Accuracy

Fig. 3. Training loss and accuracy




(a) Loss

(b) Accuracy

Fig.4. Validating loss and accuracy

The training results tend to stabilize after 150 training epochs. All three optimizers can achieve more than 80% accuracy. And a lower loss value during the stabilization phase can result in a higher accuracy. After training and stabilization, the order of loss values from smallest to largest is adam, rmsprop, sgdm, while the accuracy is the opposite. It indicates that the adam optimizer has the best performance for the training samples. The validation sample loss values and accuracy variation curves corresponding to the three optimizers are shown in Figure 4(a) and Figure 4(b), respectively. Combining the experimental results of training samples and validation samples, adam is selected as the model optimizer in this experiment. The adam is used to obtain a model with better generalization performance.

### *C. The performance of feature fusion*

To verify the effectiveness of the feature fusion method, the traditional feature extraction method and the fusion feature extraction method proposed in this paper are used respectively. The conventional feature extraction method uses SPWVD for the modulated waveform. The feature fusion method uses SPWVD and STFT for the modulated waveform respectively. The confusion matrix of recognition results corresponding to the single feature and the mixed feature is shown in Figure 5 with a SNR of 0dB. It can be seen that for the seven waveforms of Barker, LFM, Rect, B-FM, CPFSK, GFSK, and DSB-AM, the recognition accuracy of the two feature extraction methods is relatively close. And the accuracy of the feature fusion method is slightly better than the single feature extraction method. However, for the SSB-AM recognition task, the feature fusion method is significantly better than the single feature extraction method. The recognition accuracy of the fusion method is 97.2%, while the single feature extraction method is only 68.5%. The reason is that 31.2% of SSB-AMs are incorrectly identified as DSB-AMs.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Barker | 99.8% | | 1.1% | | | | 0.6% | |
| LFM | | 99.5% | 0.3% | | | | | |
| Rect | | 0.5% | 98.3% | | | | | |
| B-FM | | | | 99.5% | | | | 0.3% |
| CPFSK | | | | 0.5% | 99.0% | | | |
| GFSK | | | | | 1% | 100% | | |
| DSB-AM | 0.2% | | 0.3% | | | | 94.5% | 31.2% |
| SSB-AM | | | | | | | 4.9% | 68.5% |
| | Barker | LFM | Rect | B-FM | CPFSK | GFSK | DSB-AM | SSB-AM |

(a) Single feature

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Barker | 100% | | 1.4% | | | | 0.4% | |
| LFM | | 99.7% | | | | | | |
| Rect | | 0.3% | 98.6% | | | | | |
| B-FM | | | | 99.7% | | | | |
| CPFSK | | | | 0.3% | 99.2% | | | |
| GFSK | | | | | 0.8% | 100% | | |
| DSB-AM | | | 0.3% | | | | 95.1% | 2.8% |
| SSB-AM | | | | | | | 4.5% | 97.2% |
| | Barker | LFM | Rect | B-FM | CPFSK | GFSK | DSB-AM | SSB-AM |

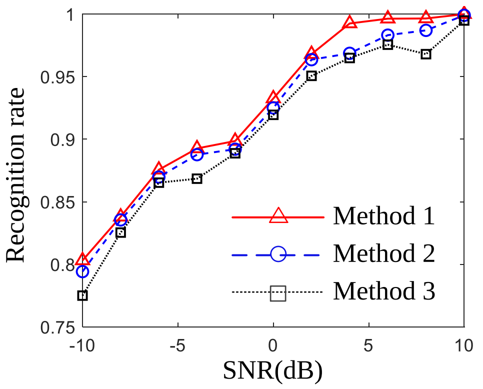
(b) fusion feature

Fig. 5. Confusion matrix with different feature

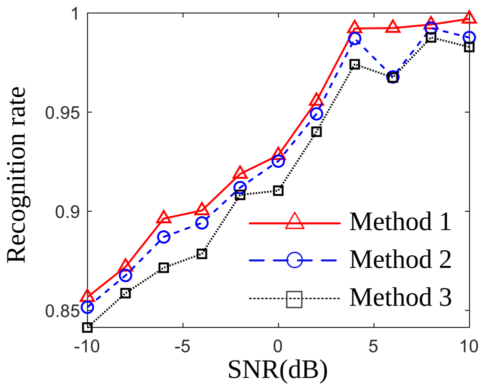
### *D. The performance of different methods*

To verify the effectiveness of the proposed method (Method1), CNN without attention mechanism (Method2) and LSTM (long short-term memory) method (Method3) are used as the comparison methods. The three methods are trained to classify eight types of modulated signals. The experiments were carried out for each modulated signal at 2 dB intervals under the conditions of SNR from -10dB to 10dB. The average recognition rate of the three methods under different SNR conditions is shown in Figure 6.

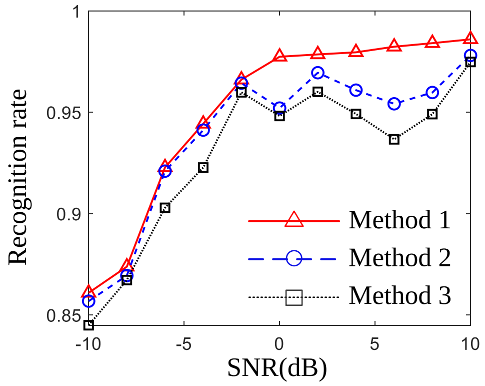
Seen from Figure 6, the recognition rate of all the three methods increases as the value of SNR grows. The recognition rate of the three methods increases slowly when the SNR is greater than 2dB. Compared with the other two methods, the recognition rate of the proposed method is the highest. For Barker, LFM, CPFSK, and SSB-AM, the recognition rate of the three methods decreases significantly when the SNR is less than 2dB. For Rect, B-FM, GFSK, and DSB-AM, the recognition rate decreases significantly when the SNR is less than 0dB. The recognition rate of the LSTM method is lower than the proposed method and the CNN method without attention mechanism. Compared with CNN, LSTM method has some limitations in identifying the time-frequency features of modulated signals. LSTM cannot outstandingly reflect the deep feature information of each modulation. The recognition rate of the CNN method without attention mechanism is lower than the proposed method. In particular, the recognition rate of the proposed method is significantly higher than that of the CNN method without attention mechanism at high SNR. The main reason is that the attention mechanism is introduced into the proposed method. Different regions can obtain different weight. Then the network becomes optional for training. After the processing of channel and spatial attention mechanism, strong modulated signal characteristics are obtained.



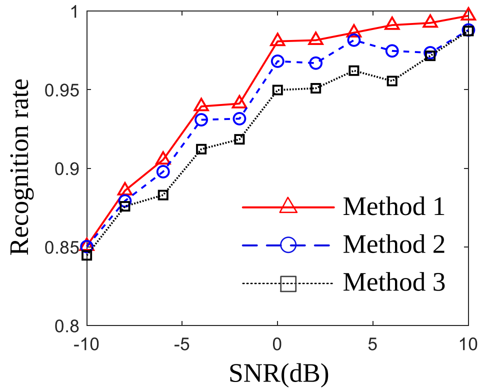

(a) Barker



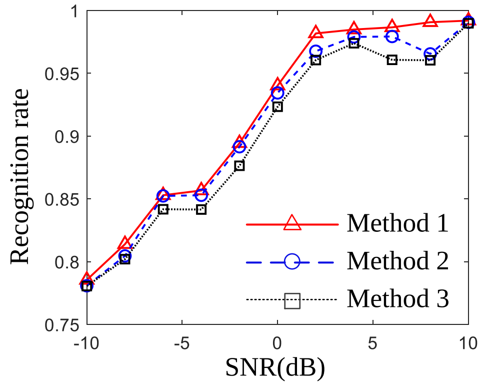

(b) LFM



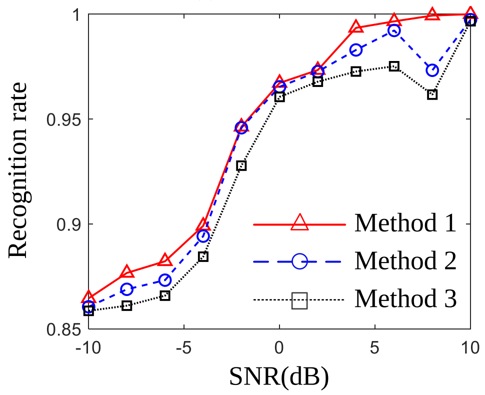

(c) Rect




(d) B-FM




(e) CPFSK




(f) GFSK

Fig. 6. Recognition rate of different methods under different SNR

## VI. CONCLUSION

A modulation recognition model based on convolutional neural network is proposed in this paper. The modulated signal recognition problem is transformed into an image recognition problem by time-frequency map of modulated signals. The STFT and WVD based time-frequency maps are fused using vector stitching. The proposed model introduces an attention mechanism to optimize the selection of different channels for the same feature map. All the spatial locations of the same feature map are reassigned to improve the learning ability. The experimental results show that the proposed model has a high recognition rate for eight modulated signals, namely Rect, LFM, Barker, GFSK, CPFSK, B-FM, DSB-AM and SSB-AM, under the condition of low SNR. The proposed model can lay a solid foundation for the subsequent signal baseband data demodulation and network station binning.

## REFERENCES

[1] Wang T, Hou Y, Zhang H, "Deep Learning Based Modulation Recognition with Multi-Cue Fusion," *IEEE Wireless Communications Letters*, vol. 10, no. 8, pp1757-1760, 2021.
[2] Xu Z, Petrunin I, Tsourdos A, "Identification of Communication Signals using Learning Approaches for Cognitive Radio Applications, " *IEEE Access*, vol. 8, no. 1, pp128930-128941, 2020.
[3] Marey M, Mostafa H, Alshebeili S A, "Blind Modulation Identification Algorithm for Two-Path Successive Relaying Systems," *IEEE Wireless Communications Letters*, vol. 10, no. 11, pp2369-2373, 2021.
[4] Jiang X, Hao M, Yan L, "Blind and Low-complexity Modulation Format Identification based on Signal Envelope Flatness for Autonomous Digital Coherent Receivers," *Applied Optics*, vol. 61, no. 20, pp5991-5997, 2022.
[5] Ahn S, "Implementation of a Noise-Shaped Signaling System through Software-Defined Radio," *Applied Sciences*, vol. 12, no. 2, pp1-6, 2022.
[6] Yang L, Xu H, Bai C, "Joint Modulation Format Identification and OSNR Estimation Method based on Trajectory Information Analysis," *Optics Communications*, vol. 507, no. 3, pp1-5, 2022.
[7] Zou C, Yu C, Guo J, "Improving Robustness and Defense with Signal Augmentation for Modulation Identification," *Journal of Physics Conference Series*, vol. 1616, pp012111, 2020.
[8] Wang Y, Zhang H, Xu L, "Adoption of Hybrid Time Series Neural Network in the Underwater Acoustic Signal Modulation Identification," *Journal of the Franklin Institute*, vol. 357, no. 18, pp13906-13922, 2020.
[9] Liang W, Shihong X U, "Research on Maximum-Likelihood Modulation Identification Algorithms for OFDM Systems," *Journal of Xinyang Normal University (Natural Science Edition)*, vol. 29, no. 04, pp1-3, 2016.
[10] Shi Q, "Automatic Modulation Identification Based on the Probability Density Function of Signal Phase," *IEEE Transactions on Communications*, vol. 60, no. 4, pp1033-1044, 2012.
[11] Wu J, Lin J, Tian B, "A Signal Modulation Identification Method Based on Neural Network," *2020 IEEE International Conference on Artificial Intelligence and Computer Applications*, 2020.
[12] Wang Y, Zhang H, Xu L, "Adoption of Hybrid Time Series Neural Network in the Underwater Acoustic Signal Modulation Identification," *Journal of the Franklin Institute*, vol. 357, no. 18, pp13906-13922, 2020.
[13] Souza A, Sutili T, JHD Cruz J, "Artificial Intelligence Techniques Evaluation for Modulation Format Identification in Optical Networks," *Journal of Microwaves, Optoelectronics and Electromagnetic Applications*, vol. 20, no. 4, pp1-3, 2021.
[14] Cai Q, Guo Y, Li P, "Modulation Format Identification in Fiber Communications using a Single Dynamical Node based Artificial Neural Network," *Photonics Research*, vol. 9, no. 1, pp1-3, 2020.
[15] Sim Kok Swee, Lim Choon Chen, Tan Shing Chiang, and Toa Chean Khim, "Deep Convolutional Neural Network for SEM Image Noise Variance Classification," *Engineering Letters*, vol. 31, no. 1, pp328-337, 2023.
[16] WANG D S, ZHANG M, LI Z, "Modulation Format Recognition and OSNR Estimation using CNN-based Deep Learning," *IEEE Photonics Technology Letters*, vol. 29, no. 19, pp1667-1670, 2017.
[17] Ji Q, Feng B, Yun L, "Identifying of Digital Signals Based on Manifold Learning," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 9, no. 2, pp127-134, 2016.
[18] Nkechinyere N J, Morocho-Cayamcela M E, Lim W, "Automatic Radar Waveform Recognition using the Wigner-Ville distribution and AlexNet-SVM," *The KICS 2020 Summer Conference*, 2020.

[19] Hu Z, Ma H, "Blind Modal Estimation using Smoothed Pseudo Wigner-Ville Distribution and Density Peaks Clustering," *Measurement Science & Technology*, no. 10, pp31, 2020.
[20] Soukaina Mouhib, Houda Anoun, Mohammed Ridouani, and Larbi Hassouni, "Global Big Data Maturity Model and its Corresponding Assessment Framework Results," *IAENG International Journal of Applied Mathematics*, vol. 53, no. 1, pp393-404, 2023.

**Ren Guo-Xi** (1979.4-) is an associate professor of Pingdingshan Polytechnic College. His research direction is computer application.