TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 17/18 pp422–431
DOI: 10.26599/TST.2020.9010057
Volume 27, Number 2, April 2022

Automatic Modulation Recognition Based on CNN and GRU

Fugang Liu, Ziwei Zhang, and Ruolin Zhou*

Abstract: Based on a comparative analysis of the Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks, we optimize the structure of the GRU network and propose a new modulation recognition method based on feature extraction and a deep learning algorithm. High-order cumulant, Signal-to-Noise Ratio (SNR), instantaneous feature, and the cyclic spectrum of signals are extracted firstly, and then input into the Convolutional Neural Network (CNN) and the parallel network of GRU for recognition. Eight modulation modes of communication signals are recognized automatically. Simulation results show that the proposed method can achieve high recognition rate at low SNR.

Key words: modulation recognition; deep learning; Gated Recurrent Unit (GRU); Convolutional Neural Network (CNN)

1 Introduction

With the development of wireless communication technology, modulation classification has been widely used in electronic countermeasures, cost reduction of cooperative communication, electronic reconnaissance, and other fields. Automatic modulation recognition technology can be roughly divided into three categories: Likelihood-Based (LB) signal recognition methods based on hypothesis testing, Feature-Based (FB) methods based on feature extraction, and deep learning-based methods[1]. LB recognition has two methods: Average Likelihood Ratio Test (ALRT) and Generalized Likelihood Ratio Test (GLRT). ALRT takes unknown variables as random variables and calculates the likelihood function by computing the average value. GLRT calculates the probability density function of the input signal on the basis of the maximum likelihood estimation of unknown quantity and determines the modulation mode accordingly[2–4]. The LB classification method can theoretically obtain the optimal classification performance, but it requires substantial prior knowledge and a considerable amount of computation. The recognition method of FB relies on features, such as instantaneous feature[5], High-Order Cumulant (HOC) feature[6, 7]. Artificial neutral networks[8], Support Vector Machine (SVM)[9], Decision Trees (DTs)[10–12], and k-nearest neighbor[13], are used in FB methods. The recognition effect of these methods is not optimal, but the required prior knowledge and computation power are remarkable improved in comparison with the first method. However, the selection of features greatly affects the recognition rate of signals.

In recent years, deep learning theory has been widely used due to its effective feature extraction capability in image processing; thus, many researchers have begun to apply deep learning to modulation recognition[14]. For example, the Long Short-Term Memory network (LSTM) is directly used to identify signals[15], or a Convolutional Neural Network (CNN) is used to identify the orthogonal and in-phase components of the signal[16] for blind recognition; however, the recognition accuracy of this method is poor at low Signal-to-Noise Ratio (SNR). Later, some researchers used deep learning theory to introduce a priori knowledge to a certain extent, conducted feature transformation on the target signal, extracted the signal constellation[17] and time-frequency energy graph[18], and used a CNN to identify

• Fugang Liu and Ziwei Zhang are with the Department of Electronics and Information Engineering, Heilongjiang University of Science and Technology, Harbin 150022, China. E-mail: liufugang_36@163.com; zhangziwei_2020@163.com.
• Ruolin Zhou is with the Department of Electrical and Computer Engineering, University of Massachusetts, Dartmouth, MA 02747, USA. E-mail: ruolin.zhou@umassd.edu.
* To whom correspondence should be addressed.
Manuscript received: 2020-10-29; accepted: 2020-11-23

classification. The recognition effect is better than the previous blind recognition, but the recognition rate is still not high enough when the SNR is low. The feature maps of some signals are similar or identical to those of other signals, such as the Quadrature Phase Shift Keying (QPSK) signal, and the offset QPSK signal constellations[19] are completely consistent; thus, the recognition method based on deep learning and constellation map alone is not effective.

On the basis of the above conditions and problems, modulation recognition techniques of eight kinds of digital signals, including M-ary Phase Shift Keying (MPSK), M-ary Amplitude Shift Keying (MASK), M-ary Frequency Shift Keying (MFSK), and M-ary Quadrature Amplitude Modulation (MQAM), are studied. After extracting the features of signals (e.g., HOC, SNR estimation, and cyclic spectrum), the feature data obtained are input into the CNN and Gated Recurrent Unit (GRU) for recognition; then, the recognition results of the two neural networks are input into the recognition system designed in this study for the final decision. Experimental results show that the recognition rate of this method is significantly improved in comparison with other methods at low SNR.

The paper is organized as follows: Section 2 briefly introduces the features of eight kinds of signals. Section 3 analyzes related deep learning models. Section 4 demonstrates the feasibility of using CNN and GRU to recognize eight wireless signals using MATLAB. Section 5 summarizes the paper.

2 Analysis of Signal Characteristic Parameters

2.1 HOC characteristics

Under the assumption that the signal received by the receiver has undergone carrier synchronization, symbol timing, and matched filtering, and the channel noise is Gaussian white noise, the symbol synchronous sampling complex signal sequence[6] obtained at the output is

$$x(t) = s(t) + n(t) = \sqrt{A}\sum_{k}^{g}\mu_k\sqrt{En}\,\lambda(t-nT_s)\exp[\mathrm{j}(2\pi f_c+\theta_c)]+n(t) \tag{1}$$

where $k = 1, 2, \ldots, g$, and g is the length of the transmitted code element sequence; A is the unknown amplitude factor; μ_k represents the code element sequence; $\lambda(t)$ is the transmitted code element waveform; T_s is the width of the code element; f_c is the carrier frequency; θ_c is the carrier phase; En is the signal energy; $n(t)$ is the zero-mean complex Gaussian white noise; $x(t)$ is the signal received at the receiving end, and $s(t)$ is the signal at the transmitting end.

For zero-mean stationary random process $X(t)$, the p-order mixing moment and k-order HOC are defined as follows[20]:

$$M_{pq} = E\left[X(t)^{p-q}\,X^*(t)^q\right] \tag{2}$$

$$C_{kx}(\tau_1,\tau_2,\ldots,\tau_{k-1}) = \mathrm{Cum}\left(x(t), x(t+\tau_1),\ldots,x(t+\tau_{k+1})\right) \tag{3}$$

where E is the expected operator, q is the position of conjugation, τ is the delay parameter, $X^*(t)$ is the conjugate transpose of $X(t)$, $\mathrm{Cum}(\cdot)$ is the operator of HOC, and $C_{kx}(\cdot)$ is the k-order cumulant of $x(t)$. Table 1 is a list of the derivation values of HOC for wireless communication digital modulation signals.

As shown in Table 1, appropriate characteristic parameters can be designed in accordance with the HOCs of eight kinds of signals. By using Eq. (4), three feature parameters (i.e., F_1, F_2, and F_3) are extracted,

$$F_1 = \frac{|C_{40}|}{C_{42}},\ F_2 = \frac{|C_{41}|}{C_{42}},\ F_3 = \frac{|C_{63}|^2}{|C_{42}|^3} \tag{4}$$

The three signal features extracted from HOC have strong robustness at low SNR. Theoretically, 4PSK and 4ASK can be identified from eight signals by setting an appropriate threshold, and the remaining six signals can be divided into three groups: 16QAM and 64QAM, 2FSK and 4FSK, and 2PSK and 2ASK.

2.2 Wavelet transform characteristics

Wavelet transform can represent local time-varying states in the time and frequency domains, and then reflect the subtle change of signals. During the process of signal modulation, the frequency, amplitude, and phase parameters of a signal often change between different code elements, and the wavelet transform can well reflect the local transient information of the signal and detect

Table 1 Derivation values of HOC for wireless digital modulation signals.

Signal	$\lvert C_{40}\rvert$	$\lvert C_{41}\rvert$	$\lvert C_{42}\rvert$	$\lvert C_{60}\rvert$	$\lvert C_{63}\rvert$
2ASK	$2En^2$	$2En^2$	$2En^2$	$16En^3$	$16En^3$
4ASK	$1.36En^2$	$1.36En^2$	$1.36En^2$	$8.32En^3$	$8.32En^3$
2PSK	$2En^2$	$2En^2$	$2En^2$	$16En^3$	$16En^3$
4PSK	En^2	0	En^2	0	$4En^3$
2FSK	0	0	En^2	0	$4En^3$
4FSK	0	0	En^2	0	$4En^3$
16QAM	$0.68En^2$	0	$0.68En^2$	0	$2.08En^3$
64QAM	$0.62En^2$	0	$0.62En^2$	0	$1.08En^3$

the abrupt part of the signal by scaling and shifting. The Continuous Wavelet Transform (CWT) of a signal is defined as[21]

$$\mathrm{CWT}(l,\xi)=\int s(t)\Psi_l^*(t)\mathrm{d}t=\frac{1}{\sqrt{l}}\int s(t)\Psi^*\left(\frac{t-\xi}{l}\right)\mathrm{d}t \tag{5}$$

where $\Psi^*(\cdot)$ is the wavelet basis function, l is the scale factor, and ξ is the displacement factor. The wavelet basis function selected in this study is complex Gaussian wavelet cgau4. First, the received modulated signal is wavelet transformed. After this operation, the signal waveform generates many small burrs, and then the burrs are filtered via median filtering. Lastly, the variance of the signal is calculated, and the wavelet transform feature of the signal is obtained. This feature reflects the stability of the amplitude of the wavelet transform of various signals.

2.3 Instantaneous characteristics

For the real signal $\hat{x}(t)$, Hilbert transformed signal is $\hat{y}(t)$. It can be represented by

$$\hat{y}(t)=\frac{1}{\pi t}\otimes\hat{x}(t)=\frac{1}{\pi}\int_{-\infty}^{\infty}\left(\frac{\hat{x}(\tau)}{t-\tau}\right)\mathrm{d}t \tag{6}$$

and $s(t)=\hat{x}(t)+\mathrm{j}\hat{y}(t)$. $\hat{x}(i)$ and $\hat{y}(i)$ are the discrete values after sampling. Instantaneous amplitude $A(i)$, instantaneous phase $\phi(i)$, and instantaneous frequency $f(i)$[22] are in the following:

$$A(i)=\left(\hat{x}^2(i)+\hat{y}^2(i)\right)^{\frac{1}{2}} \tag{7}$$

$$\phi(i)=\tan^{-1}\left[\frac{\hat{y}(i)}{\hat{x}(i)}\right] \tag{8}$$

$$f(i)=\frac{1}{2\pi T}\left[\phi(i)+\phi(i-1)\right] \tag{9}$$

where T is the sampling period.

The standard deviation of the absolute value of the instantaneous frequency of nonweak signals is assumed to be σ_{af}, and the standard deviation of the instantaneous phase nonlinear component of the zero-center nonweak signal segment is σ_{dp}[5]. Their expressions are shown in the following:

$$\sigma_{af}=\sqrt{\frac{1}{c}\left(\sum_{a(i)>a_u(i)}f_B^2(i)\right)-\left(\frac{1}{c}\sum_{a(i)>a_u(i)}\left|f_B^2(i)\right|\right)^2} \tag{10}$$

where σ_{af} is used to distinguish signals with absolute frequency information, $f_B(i)$ is the nonlinear component of $f(i)$, $a(i)$ is the amplitude of the signal, $a_u(i)$ is an amplitude judgment threshold level for judging the weak signal segment, and $a(i)>a_u(i)$ means selecting non-weak signal segment from the signal. The absolute value of the normalized instantaneous frequency of the zero center of 2FSK is a constant, and its standard deviation is zero. However, this system is implemented under noisy environments; thus σ_{af} of 2FSK is a relatively small constant. By contrast, the absolute value of the instantaneous frequency of 4FSK is not zero; thus σ_{af} of 4FSK is greater than 2FSK. Therefore, 2FSK and 4FSK can be identified with an appropriate threshold.

$$\sigma_{dp}=\sqrt{\frac{1}{c}\left[\sum_{a(i)>a_u(i)}\phi_B^2(i)\right]-\left[\frac{1}{c}\sum_{a(i)>a_u(i)}\phi_B(i)\right]^2} \tag{11}$$

where σ_{dp} can be used to distinguish signals with different direct phase information, c is the number of non-weak signals, and $\Phi_B(i)$ is the nonlinear component of $\Phi(i)$.

For example, 2ASK does not contain direct phase information; thus σ_{dp} is approximately equal to 0. The 2PSK signal contains direct phase information (its instantaneous phase is 0 or π); thus σ_{dp} is greater than 0, and 2ASK and 2PSK can be separated by characteristic σ_{dp}.

2.4 Analysis of SNR estimation

The n-order statistic of Gaussian white noise is nonzero, except for the second-order statistic; thus, the second-order statistic of the signal contains some noise information, but the fourth-order statistic does not. For MPSK signals in complex Gaussian channels, the estimated value of SNR is[23]

$$\widehat{SNR}=\frac{\sqrt{2M_2^2-M_4}}{M_2-\sqrt{2M_2^2-M_4}} \tag{12}$$

The actual expressions of the second-order and fourth-order moment method (M2M4) are as follows:

$$M_2=\frac{1}{N}\sum_{n=0}^{N-1}|x(n)|^2,\ M_4=\frac{1}{N}\sum_{n=0}^{N-1}|x(n)|^4 \tag{13}$$

where N is the length of the signal, and $x(n)$ is the received signal. This method only needs to calculate the envelope of the signal to obtain the estimated value of the SNR without signal modulation information, and it can blindly estimate the SNR.

2.5 Cyclic spectrum characteristics

Communication signals usually have cyclic stationarity, and cyclic spectral density function is an important tool for analyzing stationary signals. The cyclic spectrum of communication signals has a large nonzero value at the point where the cyclic frequency is not zero,

whereas the cyclic spectrum of stationary noise is mainly concentrated at the point where the cyclic frequency is zero. Stationary noise has zero or considerably small value at nonzero cyclic frequency. Therefore, the cyclic spectrum can well suppress noise or interference. Most modulated signals can be modeled as a cyclically stationary random process, and the cyclic spectrum of signals can be expressed as[24]

$$S_x^{\alpha}(f)=\int_{-\infty}^{\infty}R_x^{\alpha}(\tau)\,\mathrm{e}^{-\mathrm{j}2\pi f\tau}\mathrm{d}\tau \tag{14}$$

where α is the cyclic frequency, f is the spectral frequency, and $R_x^{\alpha}(\cdot)$ is the cyclic autocorrelation of signals. Common cyclic spectrum estimation algorithms use the frequency- and time-domain smoothing methods. Considering that the algorithm based on the time-domain smoothing cannot achieve a high-frequency resolution, this study uses the discrete frequency-domain smoothing algorithm to estimate the cyclic spectrum with the following expression[25]:

$$S_{x\Delta t}^{a}(t,f)_{\Delta f}=\frac{1}{L}\sum_{v=-(G-1)/2}^{(G-1)/2}\frac{1}{\Delta t}\times x_{\Delta t}(t,f+\alpha/2+vF_n)x_{\Delta t}^{*}(t,f-\alpha/2+vF_n) \tag{15}$$

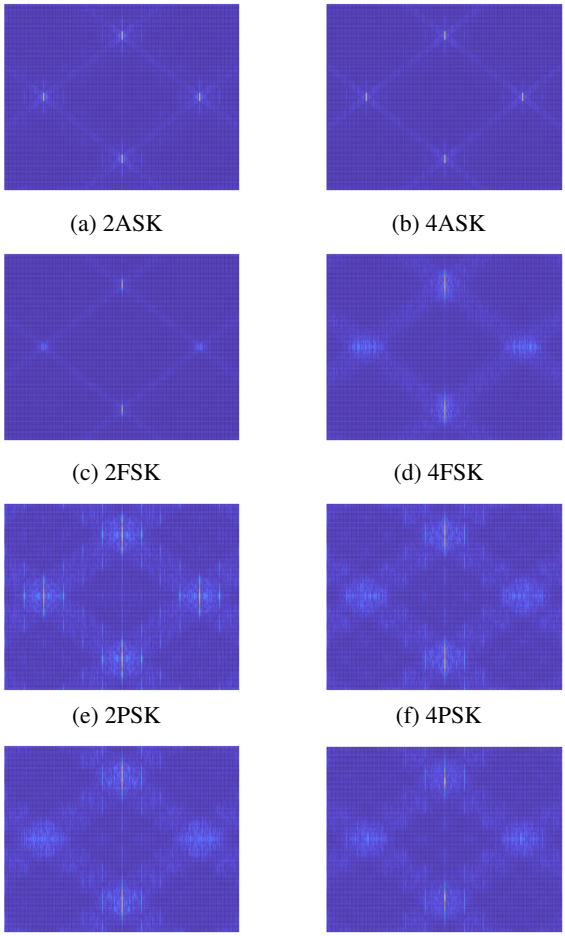
$$x_{\Delta t}(t,f)\triangleq\sum_{K=-G/2}^{G/2-1}\eta_{\Delta t}(KT_s)\,x(t-KT_s)\,\mathrm{e}^{-\mathrm{j}2\pi f(t-KT_s)} \tag{16}$$

where $\eta_{\Delta t}$ is the truncated window function, G is the total number of sampling points, L is the length of the smoothing window in the frequency domain, T_s is the sampling interval, Δf is the frequency range corresponding to the smoothing window in the frequency domain, $\Delta f = LF_n = L/(NT_s)$, the truncated time length is $\Delta t = T = NT_s$, and the frequency resolution of the discrete fourier transformation is $F_n = 1/NT_s$.

As can be seen from Fig. 1, the 2-Dimensional (2-D) cyclic spectrum graphs of MQAM, 4PSK, 2ASK, and 4ASK are similar, and the recognition effect without image enhancement is poor, while the in-class recognition and inter-class recognition of MFSK are both good.

3 CNN and GRU

CNNs are a feedforward neural network consisting of convolution computation and a deep structure. As shown in Fig. 2, the basic structural unit of CNNs[26] includes a convolutional layer, an activation layer, a pooling layer[27], and a full connection layer. The convolution



(a) 2ASK (b) 4ASK (c) 2FSK (d) 4FSK (e) 2PSK (f) 4PSK (g) 16QAM (h) 64QAM

Fig. 1 Cyclic spectrums of 8 signals (at 5 dB SNR).

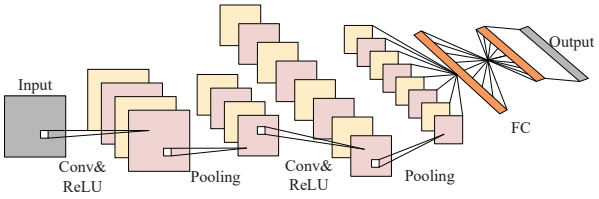


Fig. 2 Basic structural unit of CNNs.

function is for extracting image features where each neuron in the input and the previous layer of local receptive field is linked together, along with the network gradually deepening. The function of the activation layer is to perform nonlinear transformation on the extracted features, activate the extracted features, and make the feature mapping have displacement invariance. The pooling layer, also known as the lower sampling layer, can effectively reduce the parameters in the network and prevent the occurrence of overfitting. In general, one or more full connection layers are connected at the end of the CNN. Each neuron in the full connection layer is connected with each neuron in the upper layer. The full

connection layer acts as a classifier in the entire CNN and maps the learned features to the sample marker space.

Recurrent Neural Network (RNN) is a kind of neural network used to process sequence data. RNNs remember the previous information and apply it to the calculation of the current output. The nodes between the RNNs' hidden layers are no longer connected, the input of the hidden layer includes not only the output of the input layer, but also the output of the hidden layer in the last moment. In RNNs, gradient disappearance and gradient explosion easily occur during the training process, thus leading to the unsatisfactory effect of RNNs in practical application. Hochreiter and Schmidhuber proposed LSTM with memory function[28]. It consists of three parts: the input, forget, and output gates[29]. The input data are judged by whether it is needed in accordance with the algorithm. If needed, data are retained; otherwise, data are forgotten.

The emergence of the LSTM solves the inability of RNNs to handle long-distance dependencies. Cho et al.[30] proposed an improved GRU neural network of LSTM in 2014. Its basic structure is shown in Fig. 3, and its expressions are as follows:

$$z_t = \sigma\left(W_z \times [h_{t-1}, x_t]\right) \tag{17}$$

$$r_t = \sigma\left(W_r \times [h_{t-1}, x_t]\right) \tag{18}$$

$$\tilde{h}_t = \tanh\left(W_{\tilde{h}} \times [r_t \times h_{t-1}, x_t]\right) \tag{19}$$

$$h_t = (1-z_t) \times h_{t-1} + z_t \times \tilde{h}_t \tag{20}$$

where $\sigma(\cdot)$ and $\tanh(\cdot)$ are the sigmoid and hyperbolic tangent functions, respectively; x_t and h_{t-1} represent the input of the current cell, and h_t represents the output of the current cell; z_t and r_t are the states of the update and reset gates, respectively; $\tilde{h}_t$ is the output candidate set; $W_{\tilde{h}}$, W_r, and W_z are the weights of the candidate set, update gate, and reset gate, respectively. In this study, the GRU neural network is adopted, as shown in Fig. 3. Compared with LSTM, GRU simplifies the three doors into reset and update doors, reduces the number of parameters required in the calculation process, reduces the time required for training, and speeds up convergence. When the dataset is large, the effect is improved because of the numerous LSTM parameters. However, in the application of small-batch datasets, GRU can achieve the same processing effect as LSTM under the condition of short training time. In practical applications, data that can be used to train the network model are usually relatively small. Therefore, this study adopts the GRU network for modulation recognition.

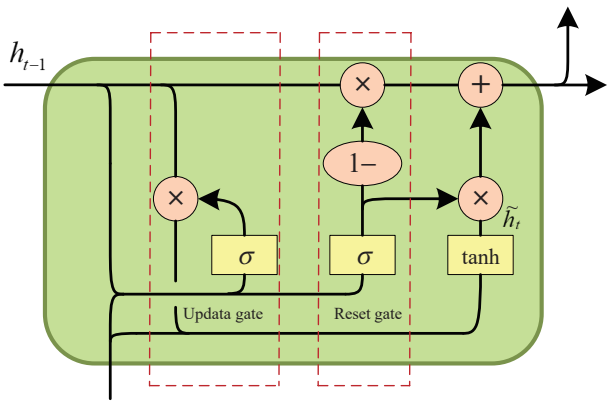


Fig. 3 Basic structural unit of GRU.

4 Implementation of the Modulation Identification System

4.1 Data generation

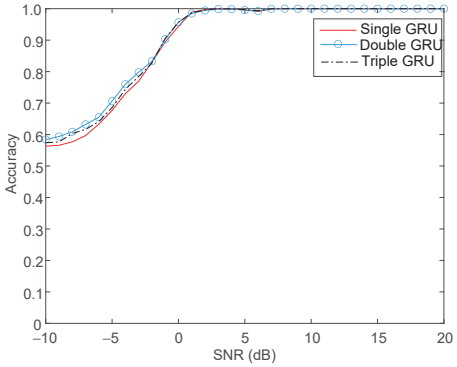
Data in this study were simulated on MATLAB software platform. There were 256 symbols in the signal and the carrier frequency was 4000 Hz, the sampling frequency was 16 000 Hz, the noise was Gaussian white noise, and the SNR was –10 dB to 20 dB. Eight modulated signals (2PSK, 2ASK, 2FSK, 4PSK, 4ASK, 4FSK, 16QAM, and 64QAM) were generated, with a total of 4960 groups of data and 620 groups of each signal (this paper used small batch datasets that are closer to the real data). After preprocessing, the signal generated by MATLAB was superimposed with Gaussian white noise. Several features, such as HOC, wavelet transform, instantaneous, and SNR estimation features, were extracted, and their cyclic spectrum was calculated. The obtained cyclic spectrum was transformed into 2-D grayscale images, and each image size was set as 150 pixel × 200 pixel. The dataset obtained was allocated at a ratio of 1 : 1 in accordance with the training and test sets.

The neural network was built and trained using the Deep Learning Toolbox provided by MATLAB. The platform for model training was the Win10 system, the CPU was an Intel i7-9700KF, and the graphics card was the NVIDIA RTX2070S.

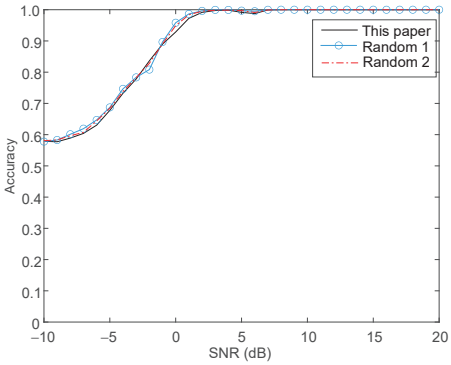
4.2 Design and experiment

The training set was placed into the GRU with different parameters and structures for training. Each case was trained 100 times, and the average value of the recognition rate obtained after 100 tests was taken for comparison.

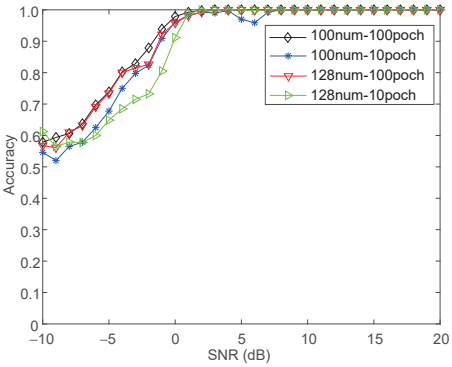
As shown in Fig. 4a, the recognition rate of double-layered GRU is better than single- and three-layered



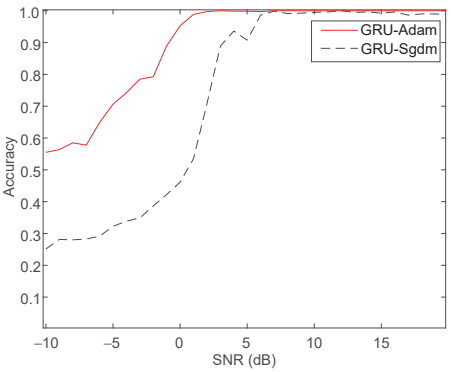
(a) Recognition rate in different layers



(b) Recognition rate under different feature orders



(c) Recognition rate for different number of nodes in hidden layers and iterations round

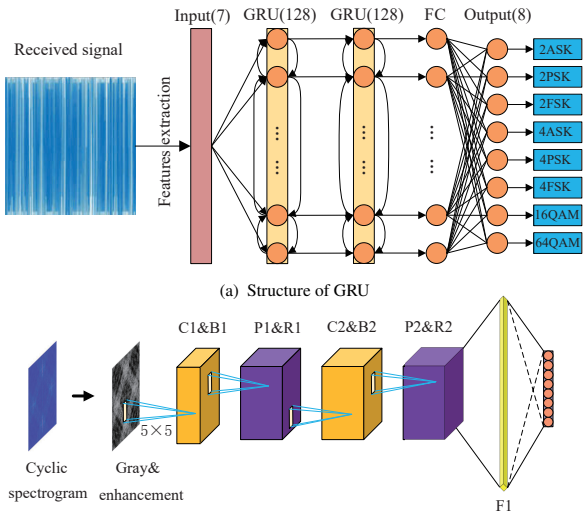


(d) Recognition rate under different gradient descent algorithms

Fig. 4 Recognition rate of signals on the basis of different GRU parameters.

GRUs. The recognition rate of single- and three-layered GRUs is lower than that of the double-layered GRU due to the insufficient depth of the neural network and the disappearance of gradient. As shown in Fig. 4b, the order of input features in this paper is HOC feature, instantaneous feature, wavelet transform feature, and SNR estimation feature. Randoms 1 and 2 are the random order of the above features. For the same signal, the order of input characteristic parameters of the training set has almost no influence on the recognition rate, and the difference of recognition rate of the three sequence datasets floats within the normal range. In Fig. 4c, the numbers of nodes in different GRU hidden layers and iteration rounds are compared, where 100 num-100 poch represents the numbers of nodes and iterations of the hidden layer are both 100. When the number of nodes in hidden layers and iteration rounds are both 100, the comprehensive recognition rate of the eight signals is optimal. In Fig. 4d, Adam and Sgdm are two gradient descent algorithms, the gradient descent algorithm with Adam has faster convergence speed and higher accuracy during training.

On the basis of the above comparison, new types of GRU networks and CNNs are designed in this study. The GRU is shown in Fig. 5a; the network consists of two GRU layers, a Full Connection (FC) layer, and a SoftMax layer. The numbers in input (7), GRU (128), and output (8) mean the numbers of nodes in the layers. The new type of CNN is shown in Fig. 5b; it consists of convolutional layer C1, batch normalization layer B1, activation layer R1, and pooling layer P1 to form



(a) Structure of GRU

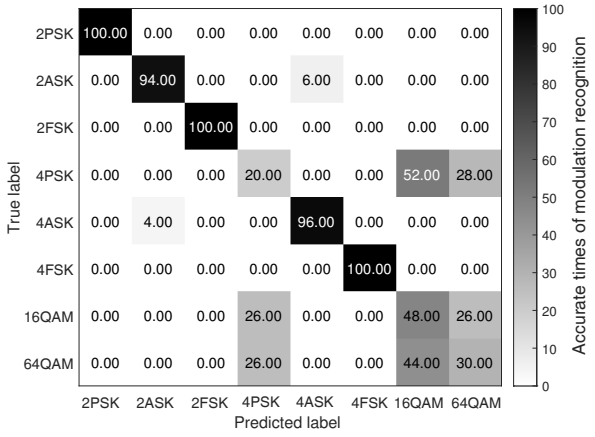
(b) Structure of CNN

Fig. 5 Structural diagrams of the GRU and CNN.

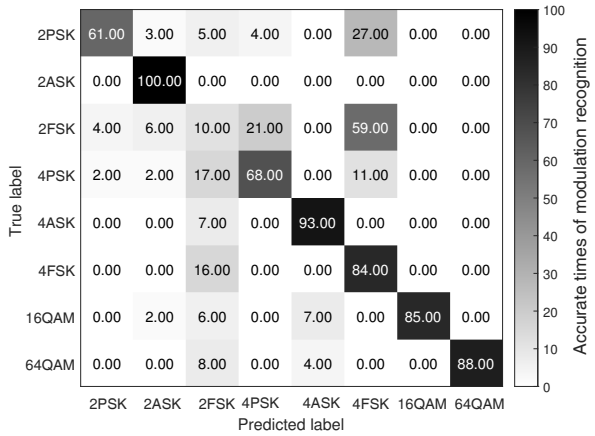
the first round of convolution combination layer. The network structure consists of two of such convolution combination layers; after two convolution combination layers, the full connection layer F1 and the SoftMax layer are connected.

The 800 test sets of cyclic spectra generated by eight signals under –5 dB and the corresponding test sets of seven characteristic parameters were input into the trained CNN and GRU models, respectively; the confusion matrix is shown in Fig. 6.

As shown in Fig. 6, the darker the color in the confusion matrix, the greater the number of times the modulation method is judged, and vice versa. The recognition rate of the three signals of CNN based on the cyclic spectrum is poor due to the similarity between the 2-D cyclic spectrum of MQAM and that of 4PSK. A certain confusion also exists between 2ASK and 4ASK. The recognition effects of 2PSK and MFSK are better. Compared with CNN, the GRU based on feature extraction has better intraclass and interclass recognition



(a) CNN confusion matrix



(b) GRU confusion matrix

Fig. 6 Confusion matrix of eight signals based on CNN and GRU.

of MQAM; moreover, the intraclass recognition of MASK is good. However, the interclass recognition of MFSK and MASK is poor.

4.3 Implementation of the algorithm

Based on the recognition characteristics of above two neural networks, this study summarizes the recognition methods based on feature extraction GRU and cyclic spectrogram CNN. The specific steps are described as follows:

Step 1: The receiver receives signals of enough length, conducts sampling and signal preprocessing, and uses the M2M4 method to perform a blind estimation of SNR.

Step 2: Extract the characteristic parameters of the signal (i.e., HOC, wavelet transform, instantaneous, and SNR estimation features) and arrange the characteristic parameters in a certain order.

Step 3: Extract the three-dimensional circular spectrum of the signal, transform the 3-D diagram into a 2-D top view, and simplify data. Then, perform gray processing and histogram equalization on 2-D color diagram to enhance the image, and thus further simplify the data and enhance the robustness of the data.

Step 4: Input the characteristic parameters and the 2-D gray map of the cyclic spectrum into the GRU and CNN designed in this study, respectively, and obtain two recognition results.

Step 5: Make the final decision according to the effects of two kinds of network recognition: when the predicted result of CNN is MFSK and 2PSK, the predicted result is directly output; when CNN predicts MASK, MQAM, and 4PSK, the GRU-predicted result is taken as the output result.

4.4 Simulation results

The generated datasets of characteristic parameters were input into GRU, Back Propagation (BP) neural network, and SVM. In accordance with the values of six characteristic parameters (no SNR estimation features), appropriate threshold values were set to form the DT recognition mode. Then, six features of the estimation without SNR were input into GRU. The generated cyclic spectrum was input into the CNN designed in this study as a dataset. The above method and the method in this study are statistically compared with each other in terms of the comprehensive recognition rate of eight signals with an SNR of -10 dB to 20 dB. The results are shown in Fig. 7.

The method based on DT has the worst recognition rate when the SNR is low, and the recognition rate is

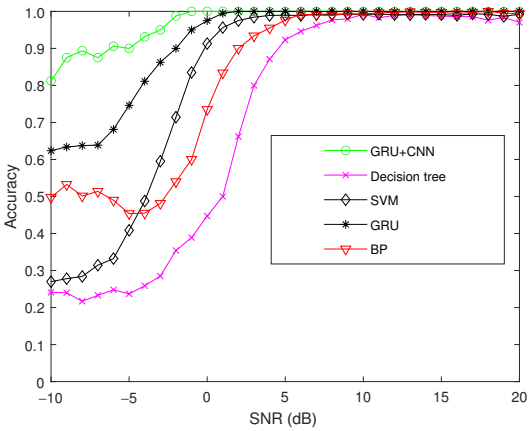
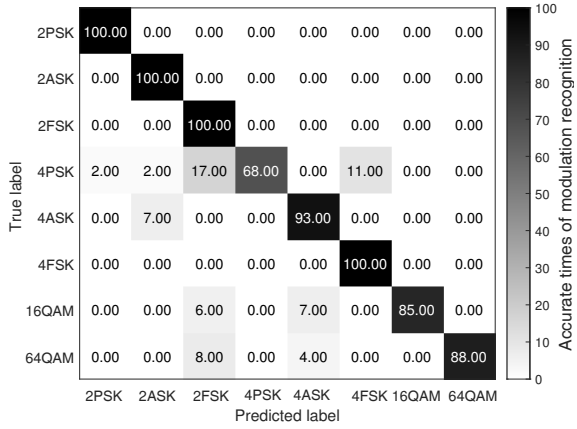


Fig. 7 Synthetic recognition rate of 8 signals under 5 methods.

close to 100% when just the SNR is 10 dB. The reason for poor result is that with the reduction of SNR, each characteristic parameter experiences serious distortion, and the modulation method based on a single threshold is seriously affected. The recognition rate of SVM-based and BP neural network is still not high when the SNR is low, especially when the SNR is less than −5 dB. However, the recognition rate of SVM is better than that of the BP neural network when the SNR is between –5 dB and 5 dB, and the recognition rate of both is close to 100% when the SNR is more than 5 dB. The feature extraction method based on GRU is superior to other traditional recognition methods based on statistical patterns, thus proving that the feature extraction method based on GRU is more accurate and reliable than the traditional recognition methods based on statistical patterns. The algorithm proposed in this study still has a high recognition rate at low SNR, i.e., above 80% at –10 dB, above 90% at –8 dB, and 100% at –1 dB.

Figure 8 shows the confusion matrixes when the SNRs of this method are –5 and –10 dB, respectively. As shown in Fig. 8a, this method combines the recognition characteristics of the two networks, and the recognition rate of the network is significantly improved in comparison with that of running only the CNN or GRU network. As shown in Fig. 8b, this method still



(a) SNR= −5 dB

(b) SNR= −10 dB

Fig. 8 Confusion matrix of 8 signals based on the proposed method.

has a relatively stable recognition performance at a low SNR.

Table 2 shows the compared recognition rate of feature parameters without SNR estimation. Table 3 shows the compared recognition rate of feature parameters with SNR estimation. After adding SNR features, the recognition rate of the four methods is improved to some extent in comparison with that without SNR features. The method in this study reaches 100% recognition rate earlier after adding SNR estimation features. The SVM-based method is not as effective as the BP neural network method in the case of low SNR. However, with the improvement of SNR, the SVM-

Table 2 Recognition rate when SNRs are not estimated.

Method	SNR (dB)								
	−10	−8	−6	−4	−2	0	2	4	6
SVM	0.2700	0.2840	0.3320	0.4878	0.7142	0.9130	0.9725	0.9890	0.9888
BP	0.3812	0.4487	0.4462	0.4300	0.5157	0.6613	0.7538	0.9363	0.9587
GRU	0.5212	0.5388	0.5887	0.6387	0.7925	0.9088	0.9763	0.9925	0.9963
Our proposed	0.6875	0.7188	0.7813	0.8125	0.8750	0.9688	0.9963	1.0000	1.0000

Table 3 Recognition rate when SNRs are estimated.

Method	SNR (dB)								
	−10	−8	−6	−4	−2	0	2	4	6
SVM	0.2840	0.2900	0.3395	0.4983	0.7188	0.9195	0.9770	0.9902	0.9895
BP	0.4975	0.5012	0.4900	0.4550	0.5400	0.7350	0.9000	0.9550	0.9925
GRU	0.6238	0.6375	0.6813	0.8113	0.9000	0.9750	1.0000	1.0000	1.0000
Our proposed	0.8125	0.8938	0.9063	0.9313	0.9875	1.0000	1.0000	1.0000	1.0000

based method has more advantages than the BP neural network. The recognition performance of the method based on the GRU network is still better than that of the traditional method based on SVM and BP neural network, even without the addition of SNR estimation features. The method presented in this study is superior to the other three methods, regardless of whether or not SNR estimation features are added.

5 Conclusion

To solve the low recognition rate of traditional modulation recognition methods at low SNR and small dataset in actual applications, this study designs and optimizes a GRU network structure and proposes a GRU recognition method based on feature extraction. This method incorporates SNR estimation into feature extraction and has a certain degree of improvement in comparison with traditional modulation recognition at low SNR. In addition, this study proposes a new modulation recognition method, which combines GRU based on feature extraction and CNN based on cyclic spectrum. The recognition rate is more than 90% when SNR is –6 dB and 100% when SNR is –1 dB. Future work will improve the accuracy of the estimation of nonconstant modulus signal by M2M4. The algorithm of SNR estimation can also be further optimized.

Acknowledgment

This work was partially supported by Major Scientific and Technological Achievements Transformation Project of Heilongjiang Province in 2019 (No. CG20A007).

References

[1] R. Zhou, F. Liu, and C. W. Gravelle, Deep learning for modulation recognition: A survey with a demonstration, *Behav. Ecol. Sociobiol.,* vol. 8, pp. 67366–67376, 2020.

[2] L. Y. Uys, M. Gouws, J. J. Strydom, and A. S. J. Helberg, The performance of feature-based classification of digital modulations under varying SNR and fading channel conditions, in *Proc. 2017 IEEE AFRICON*, Cape Town, South Africa, 2017, pp. 198–203.

[3] M. H. Shah and X. Dang, An effective approach for low-complexity maximum likelihood based automatic modulation classification of STBC-MIMO systems, *Frontiers of Information Technology & Electronic Engineering,* vol. 21, no. 3, pp. 465–475, 2019.

[4] G. Qian, P. Wei, Z. Ruan, and J. Lu, A low-complexity modulation classification algorithm for MIMO–OSTBC system, *Circuits Systems & Signal Processing*, vol. 36, no. 6, pp. 2622–2634, 2019.

[5] E. Moser, M. K. Moran, E. Hillen, D. Li, and Z. Wu, Automatic modulation classification via instantaneous features, in *Proc. 2015 National Aerospace and Electronics Conference*, Dayton, OH, USA, 2015, pp. 218–223.

[6] G. B. Markovic, V. S. Sokolovic, and M. L. Dukic, Distributed hybrid two-stage multi-sensor fusion for cooperative modulation classification in large-scale wireless sensor networks, *Sensors,* vol. 19, no. 19, pp. 4339–4362, 2019.

[7] S. Kharbech, I. Dayoub, M. Zwingelstein-Colin, and E. P. Simon, Blind digital modulation identification for MIMO systems in railway environments with high-speed channels and impulsive noise, *IEEE Transactions on Vehicular Technology*, vol. 67, no. 8, pp. 7370–7379, 2018.

[8] A. K. Nandi and E. E. Azzouz, Algorithms for automatic modulation recognition of communication signals, *IEEE Transactions on Communications*, vol. 46, no. 4, pp. 431–436, 1998.

[9] X. Zhang, J. Sun, and X. Zhang, Automatic modulation classification based on novel feature extraction algorithms, *IEEE Access*, vol. 8, pp. 16362–16371, 2020.

[10] A. Yi, H. Liu, L. Yan, L. Jiang, Y. Pan, and B. Luo, Amplitude variance and 4th power transformation based modulation format identification for digital coherent receiver, *Optics Communications*, vol. 452, pp. 109–115, 2019.

[11] A. K. Ali and E. Erelebi, Algorithm for automatic recognition of PSK and QAM with unique classifier based on features and threshold levels, *ISA Transactions*, vol. 102, pp. 173–192, 2020.

[12] L. Zhao, H. Xu, S. Bai, and C. Bai, Modulus mean square-based blind hybrid modulation format recognition for orthogonal frequency division multiplexing-based elastic optical networking, *Optics Communications*, vol. 445, pp. 284–290, 2019.

[13] M. W. Aslam, Z. Zhu, and A. K. Nandi, Automatic modulation classification using combination of genetic programming and KNN, *IEEE Transactions on Wireless Communications*, vol. 11, no. 8, pp. 2742–2750, 2012.

[14] T. J. O'Shea, J. Corgan, and T. C. Clancy, Convolutional radio modulation recognition networks, in *Proc. 17th*

International Conference on Engineering Applications of Neural Networks, Aberdeen, UK, 2016, pp. 213–226.

[15] N. Daldal, O. Yildirim, and K. Polat, Deep long short-term memory networks-based automatic recognition of six different digital modulation types under varying noise conditions, *Neural Computing and Applications*, vol. 31, no. 6, pp. 1967–1981, 2019.

[16] J. Shi, S. Hong, C. Cai, Y. Wang, H. Huang, and G. Gui, Deep learning-based automatic modulation recognition method in the presence of phase offset, *IEEE Access*, vol. 8, pp. 42841–42847, 2020.

[17] S. Peng, H. Jiang, H. Wang, H. Alwageed, Y. Zhou, M. M. Sebdani, and Y. Yao, Modulation classification based on signal constellation diagrams and deep learning, *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 3, pp. 718–727, 2019.

[18] 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.

[19] X. Zha, H. Peng, X. Qin, G. Li, and S. Yang, A deep learning framework for signal detection and modulation classification, *Sensors*, vol. 19, no. 18, pp. 4042–4063, 2019.

[20] X. Sun, S. Su, Z. Huang, Z. Zuo, X. Guo, and J. Wei, Blind modulation format identification using decision tree twin support vector machine in optical communication system, *Optics Communications*, vol. 438, pp. 67–77, 2019.

[21] G. Lu, K. Zhang, S. Huang, Y. Zhang, and Z. Feng, Modulation recognition for incomplete signals through dictionary learning, in *Proc. 7th Wireless Communications and Networking Conference*, San Francisco, CA, USA, 2017, pp. 1–6.

[22] K. C. Ho, W. Prokopiw, and Y. T. Chan, Modulation identification of digital signals by the wavelet transform, *IEE Proceedings—Radar, Sonar and Navigation*, vol. 147, no. 4, pp. 169–176, 2002.

[23] E. Karastergios, M. A. K. Sumanasena, and B. G. Evans, Simple SNR estimator for mobile fading channels, *Electronics Letters*, vol. 39, no. 2, pp. 244–245, 2003.

[24] W. A. Gardner and C. M. Spooner, Signal interception: Performance advantages of cyclic-feature detectors, *IEEE Transactions on Communications*, vol. 40, no. 1, pp. 149–159, 1992.

[25] W. A. Gardner, Measurement of spectral correlation, *Behav. Ecol. Sociobiol.*, vol. 34, no. 5, pp. 1111–1123, 1986.

[26] S. Kong, M. Kim, L. M. Hoang, and E. Kim, Automatic LPI radar waveform recognition using CNN, *IEEE Access*, vol. 6, pp. 4207–4219, 2018.

[27] R. Xin, J. Zhang, and Y. Shao, Complex network classification with convolutional neural network, *Tsinghua Science and Technology*, vol. 25, no. 4, pp. 447–457, 2018.

[28] S. Hochreiter and J. Schmidhuber, Long short-term memory, *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[29] Y. Liu, S. Dong, M. Lu, and J. Wang, LSTM based reserve prediction for bank outlets, *Tsinghua Science and Technology*, vol. 24, no. 1, pp. 77–85, 2019.

[30] K. Cho, V. M. Bart, D. Bahdanau, and Y. Bengio, On the properties of neural machine translation: Encoder-decoder approaches, *Computer Science,* vol. 99, pp. 103–111, 2014.

Fugang Liu received the PhD degree in communication and information system from Harbin Engineering University in 2013. He is now a professor at Heilongjiang University of Science and Technology. His research interests include direction of arrival estimation of wideband signals, wireless communication, and deep learning.

Ziwei Zhang received the BS degree from Shandong University of Technology, China in 2014. Now he is a master student at the Department of Electronics and Information, Heilongjiang University of Science and Technology, China. His main research interests include modulation recognition of communication signals.

Ruolin Zhou received the PhD degree in electrical engineering from Wright State University, USA in 2012. She has been a professor at University of Massachusetts, Dartmouth, MA, USA since 2018. Her research interests include intelligent radio and RF systems, AI-enabled spectrum learning and management, and security of real-time embedded systems.