



A deep learning based algorithm with multi-level feature extraction for automatic modulation recognition

Hang Zhang¹ · Ruihua Nie² · Minghui Lin¹ · Ruijuan Wu³ · Guo Xian³ · Xiaofeng Gong¹ · Qin Yu¹ · Ruisen Luo¹

Accepted: 2 August 2021 / Published online: 26 August 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Automatic modulation recognition is a critical challenge in the field of cognitive radio. In the process of communication, radio signals are modulated in various modes and are interfered by the complex electromagnetic environment. To cope with these problems and avoid manual selection of complex expert features, we propose a multi-level feature extraction algorithm based on deep learning to adequately exploit the hidden feature information of modulated signals. Our algorithm integrates the correlation between the channels of radio signals with convolutional neural networks and Bidirectional Long Short-term Memory (Bi-LSTM), and adopts the appropriate skip connection, which avoids the loss of valid information and achieves the complementarity between spatial and temporal features. In our model, the one-dimensional convolutional layer is specially utilized to enrich the feature representation of each sample point of in-phase and quadrature (I/Q) signals and emphasize the mutual influence of I channel (in-phase signal) and Q channel (quadrature signal). In addition, the label smoothing technique is used to improve the generalization ability of the model. Our proposed method is also of certain significance for other signal processing methods based on deep learning. Experiment results demonstrate that our algorithm outperforms the popular algorithms and is of higher robustness. Specifically, the proposed method improves the recognition accuracy, reaching 92.68% at high signal-to-noise ratio (SNR). In particular, it also reduces the difficulty of recognition for multiple quadrature amplitude modulation (MQAM) signals and significantly improves the recognition accuracy for 16QAM and 64QAM.

Keywords Automatic modulation recognition · Deep learning · Spatial features · Temporal features · Attention mechanism

1 Introduction

Automatic modulation recognition (AMR) is an essential technology in the field of wireless communication. In the process of non-cooperative communication, due to the unknown modulation type of the radio signal, it is necessary to determine the modulation type of the signal, so as to

adopt the corresponding demodulation mode for demodulation. AMR is widely applied in civilian fields [1, 2] and military fields such as military reconnaissance [3], electronic countermeasures [4] and so on. Blind recognition of modulation types of radio signals is a major challenge due to the lack of the prior knowledge and receiving parameters. In addition, the diversity of modulation types and the complex electromagnetic environment in the transmission process exacerbate the difficulty of AMR. Therefore, lots of researchers are trying to explore efficient and robust modulation recognition algorithms.

In general, modulation recognition methods are divided into two categories: likelihood-based and feature-based modulation recognition methods. The likelihood-based methods [5–8] require rich prior knowledge. By processing the likelihood function of the signals, the obtained likelihood ratio is compared with the setting threshold to achieve

✉ Ruisen Luo
rsluo@scu.edu.cn

¹ College of Electrical Engineering, Sichuan University, 24 South Section 1, One Ring Road, Chengdu 610065, China

² State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University, Chengdu 610065, Sichuan, China

³ Chengdu Dagongbochuang Information Technology Co., Ltd, Chengdu 610059, China

the modulation recognition. This method has high computational complexity and heavily relies on the thresholds. Besides, it is more sensitive to parameter deviations and has poor model robustness. The feature-based methods mainly include three steps: pre-processing, feature extraction, and classifier selection. Traditional feature extraction for modulated signals consists of the following methods, such as feature parameter extraction based on higher-order cumulants [9], cyclic spectral analysis [10], wavelet transform [11, 12], and transform analysis based on constellation diagrams [13]. Then these artificially extracted expert features are used to identify the modulation mode by a classifier, e.g., the decision tree or support vector machine [14, 15], etc. These extracted features belong to expert knowledge and require strong expertise. Meanwhile, they have a limited scope of application.

In recent years, with the rise of deep learning, some scholars have utilized the artificial neural networks (ANNs) with powerful feature extraction capabilities to avoid the cumbersome manual selection of features and achieve better performance of modulation recognition. Reference [16] compares the effects of decision-theoretic methods and ANNs on modulation recognition and finds that the overall accuracy of the ANN-based algorithm was higher than that of the decision-theoretic algorithm at 15 dB signal-to-noise ratio (SNR). Reference [17] converts the modulated signals into the form of constellation diagrams and uses deep learning methods to achieve high accuracy recognition at high SNR for two classes of signals, multiple phase shift keying (MPSK) and multiple quadrature amplitude modulation (MQAM). Recently, in order to facilitate the specification of the study and to compare the advantages and disadvantages of various algorithms, a researcher releases a radio modulated signal standard dataset RML2016.10a [18], and related research has gradually increased. An early proposed AMR model based on deep learning is a convolutional neural network (CNN) based algorithm [19]. O'Shea uses the appropriate convolutional layer and fully connected layer to extract the features of in-phase and quadrature (I/Q) signals, which proves the effectiveness of the CNN for modulation recognition. To eliminate the frequency offset and phase offset in the signal transmission process, reference [20] introduces a signal correction module and combines with the CNN, which can complete the modulation recognition task. Reference [21] makes a comparison between CNN, inception modules, residual networks, and convolutional long short-term deep neural networks (CLDNN) based methods on the RML2016.10a dataset. Results show that modulation recognition performance is not limited by network depth and the performance of CLDNN is better than that of other networks. Reference [22] proposes two ensemble CNN models for modulation recognition, which are utilized for identifying MQAM and other modulation types, respectively. This

approach is effective, but the model is complex and needs more processing time. The recurrent neural network (RNN) has been proved to be a powerful model for sequential data processing [23]. Considering the temporal characteristics of I/Q signals, reference [24] proposes to use an improved structure of RNN, the gated recurrent unit network (GRU) [25], to accomplish the modulation recognition. From another perspective, reference [26] proposes an automatic recognition algorithm based on long short-term memory (LSTM) at the benchmark dataset. It achieves a high accuracy closed to 90% above 0 dB SNR by taking the instantaneous amplitude and phase of the I/Q signal as input. It shows that the relationship between the I channel (in-phase signal) and Q channel (quadrature signal) has significant influence on automatic modulation recognition.

The existing modulation recognition methods based on neural networks mainly use CNN or RNN to complete modulation recognition through an elaborately designed network model. Most of these methods ignore the mutual information between the components of I/Q signals, and there is useful information loss during the network calculation process. In addition, the hidden information mining of I/Q signals is not sufficient, which also limits the performance of modulation recognition. Moreover, in the actual transmission process of radio signals, there are frequency shifts, multipath fading, and other complex environmental influences, which increase the difficulty of identifying some modulation types. To fully exploit the hidden information of I/Q signals and pay attention to the mutual influence between the I and Q components, we propose an algorithm based on the combination of spatial fusion features and temporal features of I/Q signals. Our contributions consist mainly of the following points:

- (1) The substantial correlations between the components of I/Q signal are emphasized in our proposed algorithm. A one-dimensional (1D) convolutional layer is used to extract the mutual information between the I channel and Q channel, and to expand the feature channel of each sample point to rich the representation of features.
- (2) The features of I/Q signals are extracted from multiple levels, to utilize the diverse features of I/Q signals. Firstly, the fused spatial features are extracted by a series of convolutional blocks and skip connections. Then, they are fed into the bidirectional long short-term memory (Bi-LSTM) with the attention mechanism to obtain the temporal features. The label smoothing technology is also utilized to improve the generalization ability of the model. It is demonstrated that the overall performance of our model is better than other popular models, especially in terms of the recognition accuracy of MQAM signals.

2 Related work

The effect of the modulation recognition algorithms based on feature engineering mainly depends on the choice of artificial features and classifiers. They are easily affected by expert knowledge and have weak generalization ability. In comparison, the deep learning model is composed of the multi-layer non-linear structure, which can achieve automatic mining of the potential features of the data, avoid the cumbersome manual feature selection, and has better effects. Therefore, the modulation recognition algorithm based on deep learning has great potential. Considering the differences between modulated signals, images and speech signals, how to make better use of the characteristics of neural networks and design a suitable algorithm for modulation recognition task is the current research focus.

In reference [27], higher order spectra features of modulated signals are constructed, and are input to neural networks for classification to achieve high recognition accuracy for three kinds of modulated signals. However, this method still needs to construct artificial features, and there are few kinds of modulation recognition types, which has certain limitations. Reference [28] first calculates the cyclic spectra of the modulated signal and then feeds them into a well-designed CNN, which enables the recognition of the modulation types of very high frequency signals. In [29, 30], modulated signals are firstly preprocessed to obtain the corresponding time–frequency distribution maps, and then the CNN is used for feature extraction from these maps, so as to achieve the recognition of several modulation types. Reference [31] combines the eye diagrams and vector diagrams of modulated signals, and then inputs them into the proposed CNN for feature extraction, which can recognize eight kinds of digital signals. These methods mentioned above all convert the modulated signals into other characteristic representations according to different modulation types, and then utilize the neural networks for modulation recognition. On the one hand, these above algorithms require additional signal preprocessing and lack real-time performance. On the other hand, these algorithms need to artificially determine preprocessing methods according to different modulation types, and the converted feature representation may have a loss of valid information, which has some limitations in applications. Therefore, a large number of researchers are exploring how to take advantage of the powerful feature extraction ability of neural networks to design an efficient algorithm for AMR. In reference [32], an improved CNN inspired by the AlexNet network [33] is designed for automatic feature extraction of I/Q signals to identify their modulation types. Reference [34, 35] proposes an algorithm based on CNN, which can recognize the modulation

of pulse repetition intervals. Referring to the development of CNN in image field, a specific CNN is designed in [36, 37] to achieve modulation recognition. Reference [38, 39] uses generative adversarial networks to improve the performance of modulation recognition, but it requires certain skills for the training of this network structure and can easily lead to unstable training results.

The above related researches illustrate the inherent feature diversity of modulated signals, prompting us to think about how to improve the performance of modulation recognition algorithms based on deep learning. In our algorithm, the intrinsic temporal and spatial characteristics of the modulated signals are fully considered, and the mutual information between the I/Q components is emphasized, which significantly improves the modulation recognition effect. The experiment results show its effectiveness.

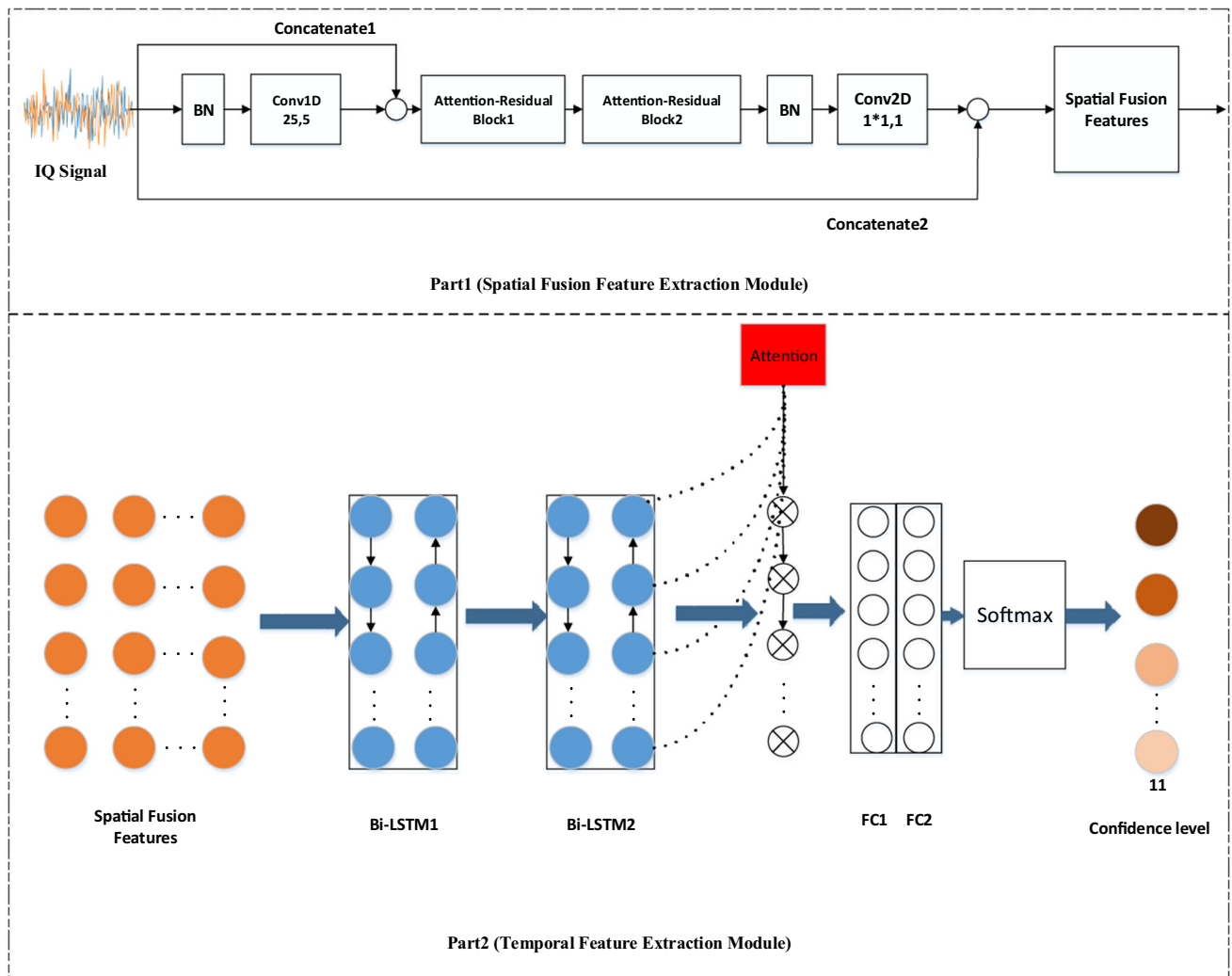
3 Methods

3.1 Our proposed model

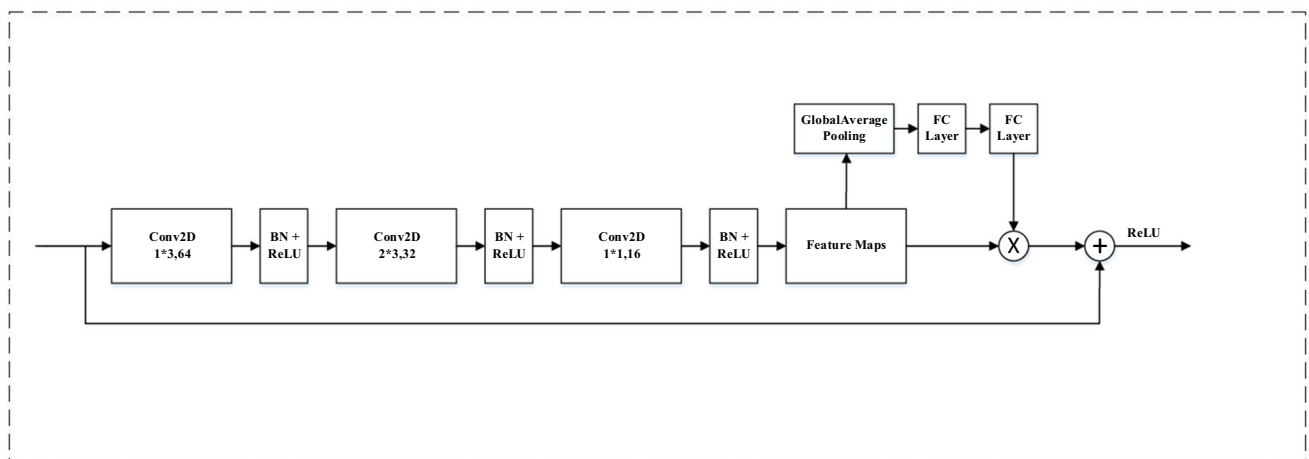
I/Q signals of different modulation types have their corresponding rules, and in the actual channel transmission process, there are complex effects such as multipath fading, frequency offset, and path loss. Therefore, to improve the robustness of modulation recognition and find an effective representation of features of modulated signals, a multi-level feature fusion algorithm is proposed in this paper, as shown in Fig. 1. Firstly, a 1D convolutional layer is adopted to explore the mutual information corrections between the I channel and the Q channel. Next, the high-order spatial fusion features extracted by subsequent convolutional blocks and skip connections method are sent to the Bi-LSTM for temporal feature extraction. Moreover, the attention mechanism is utilized to mine more important temporal characteristics. Through the integration of CNN and LSTM, it can achieve the complementarity between the spatial and temporal characteristics of the I/Q signals and attain more competitive recognition accuracy for multiple modulation signals.

3.2 Spatial fusion feature extraction module

In the field of communications, I/Q signals are composed of the quadrature component and the in-phase component. In this way, it is convenient for determining various characteristic information of the signal, such as instantaneous amplitude, phase, power, and so on. Reference [26] proposed a method to obtain high recognition accuracy by calculating the instantaneous amplitude and phase of I/Q signals and then sending them to the LSTM network for



(a) Our Proposed Model



(b) Attention-Residual Block

Fig. 1 The architecture of our proposed algorithm

feature extraction. It indicates that there is crucial correlations between the I and Q channel. Inspired by this, in order to sufficiently explore the mutual influence between the I/Q components and use the powerful feature extraction capabilities of the neural network, we propose an algorithm that combines spatial fusion feature extraction module and temporal feature extraction module.

The spatial fusion feature extraction module is shown in Fig. 1. *BN* represents the Batch Normalization layer. It is used to speed up the training process, and can also suppress the disappearance of the gradient and the internal covariate shift phenomenon [40]. For obtaining the mutual information between each component of the I/Q signals and extracting their hidden instantaneous characteristics, the normalized I/Q data with the size of 2×128 is fed into a 1D convolutional layer. In our experiment, the number of 1D convolution kernel is set to 25 and its size is 5. Thus, the output size of I/Q data by one-dimensional convolution operation is 25×128 . Then by Concatenate1 operation, the original I/Q signal and the extracted features are merged to get an array of size 27×128 , where 27 contains both the original I/Q sample points (2) and the hidden features (25) extracted by the 1D convolutional layer, and 128 represents the number of sample points in each frame. It not only avoids the loss of the original useful information, but also effectively combines the extracted I/Q mutual information. In this way, each I/Q sample point is expanded from the original two dimensions to more dimensions, which enriches their feature representation. Then the fused I/Q feature information is fed into two Attention-Residual Blocks to acquire the spatial features. The internal structure of this block is shown in Fig. 1. Its main structure is composed of three 2D convolutional layers. BN layer is added between each convolutional layer, and the ‘relu’ activation function is used to improve its nonlinear mapping ability. Furthermore, to improve the exploitation of useful information and enhance the feature extraction capability of the model, the channel attention mechanism [41] is introduced in our designed network. The implementation principle of this attention mechanism is shown in Fig. 2. It includes two parts: squeeze and excitation. Firstly, the extracted feature map U is compressed into the channel descriptor $z \in \mathbb{R}^c$ by

using the global average pooling (GAP). The element of c^{th} is calculated by the following formula in (1):

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i,j) \tag{1}$$

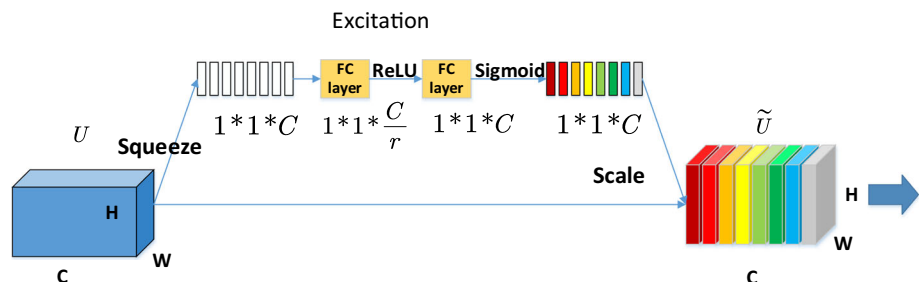
Therefore, each element of z has global receptive field corresponding to the feature map U . The excitation part is composed of two fully-connected layers (FC layer). The dimension of the fully-connected layer is controlled by the hyper-parameter r . The weights of the respective channels are learned by dimensional transformations of the two FC layers, and then the obtained weights are multiplied by the feature maps of its corresponding channel to readjust the feature map. It is a lightweight gating mechanism used to improve the network representation capabilities by establishing the channel-wise relationship.

Finally, the recalibrated feature maps and the previous inputs are added through the skip-connection to achieve residual learning [42] and avoid network degradation. Then the features extracted by the two Attention-Residual Blocks are input into a BN layer and a 2D convolutional layer with the size of 1×1 to realize cross-channel interaction and information integration. Through the Concatenate2 operation, the original IQ signal and the extracted spatial features are concatenated together to obtain the spatial fusion features. This not only takes into account the spatial features extracted by the network, but also avoids the potential loss of temporal features during the extraction process, so as to facilitate the feature extraction of later modules. The next step is to send them into the Part2 of the network for hidden temporal characteristics extraction.

3.3 Temporal feature extraction module

As an important branch of neural networks, RNN shows great advantages in time-series data processing. However, due to structural defects in traditional RNN, the problem of gradient disappearance is prone to occur during model training, and it is difficult to learn long-term dependencies. Therefore, a kind of improved RNN named LSTM [43] is used in this paper, which has a gating mechanism and can choose to update the content that needs to be remembered.

Fig. 2 A Squeeze-and-Excitation block [41]



To extract the intrinsic characteristics of modulated signals from multiple dimensions, the spatial fusion features obtained from the Part1 of our model are entered into the Part2 structure consisting of two Bi-LSTM layers and FC layers, which is expected to capture the temporal features embedded in the modulated signal. The advantage of Bi-LSTM is that it can extract the past and future correlation information of the sequence data, which can gain more complete temporal features. Since there is a large amount of information stored in Bi-LSTM, we add an attention mechanism [44] to the final output of the two Bi-LSTM layers to filter out irrelevant information and make the model focus on more important features. That is, one of the output vectors of the last Bi-LSTM layer is extracted and projected as a query vector using the FC layer, and then a dot product operation is performed with the output of Bi-LSTM. After that, the result is normalized by the ‘softmax’ function to get the values of the attention probability distribution. Next, the attention value, i.e., a vector, makes dot product with the output of the last Bi-LSTM layer to obtain the final feature vector. To map the extracted feature vector to a more easily separable space, two FC layers are introduced with the activation function, ‘selu’ in our model. The ‘dropout’ strategy is adopted to prevent overfitting, and finally input them to the FC layer with the ‘softmax’ activation function to get the confidence level of the I/Q signal corresponding to each kind of modulation types.

3.4 Label smoothing

Automatic modulation recognition is essentially a multi-classification task, therefore, the modulation type labels are generally encoded in one-hot vector and cross-entropy loss function is used. However, some of the modulation types are similar, e.g., 16QAM, 64QAM. Moreover, due to the influence of noise, it is easy to cause overfitting, and makes the model too confident of the classification result and leads to misclassification of certain complex modulation types. To improve this situation and enhance the generalization ability of our model, the label smoothing technique [45] is introduced in our algorithm. Besides, it encourages the representations of training examples from the same class to group in tight clusters to improve model calibration [46]. For the ground-truth label y corresponding to the training sample x , we use (2) to replace the label distribution $q(k|x) = \delta_{k,y}$.

$$q'(k|x) = (1 - \epsilon)\delta_{k,y} + \epsilon u(k) \quad (2)$$

where ϵ represents the smoothing parameter. In our experiment, it is set to 0.2. $u(k)$ denotes the distribution of labels. Since the number of each modulation types are equal in the experiment, we use the uniform distribution.

Let $u(k) = 1/K$, K is equal to number of label categories (11).

4 Experiments

4.1 Dataset

In fact, the received I/Q signals $y(t)$ can be given by (3):

$$y(t) = s(t) * h(t) + n(t) \quad (3)$$

where $s(t)$ represents the modulated signal, $h(t)$ represents the channel impulse response, $n(t)$ denotes the additive white Gaussian noise (AWGN) with zero mean.

In our experiment, the benchmark dataset RML2016.10a [18] generated by GNU Radio is used as the basis for different algorithm comparisons. It includes 11 kinds of common I/Q signals (BPSK, QPSK, 8PSK, 16QAM, 64QAM, BFSK, CPFSK WB-FM, AM-SSB, AM-DSB, and PAM4) with SNRs ranging from -20 dB to 18 dB with an interval of 2. The number of each type of modulation signals corresponding to the certain SNR is 1000. The entire dataset has a total of 220,000 modulated signals, and each I/Q signal has 128 complex floating-point samples. In the process of signal generation, in addition to adding noise, other factors such as center frequency offset, sampling rate offset, and multipath fading are also considered to resemble the real communication conditions.

4.2 Experimental details

4.2.1 Parameter setting

The neural network is implemented using Keras with the TensorFlow backend. All of the models are trained by the GPU, Nvidia GeForce RTX 2080. The batch size is 64. The initial learning rate is 0.001, and it would decay every 10 epochs during the training process. The dataset is divided by a ratio of 7:1:2, i.e., for each modulation type per SNR, 700 signals are randomly selected as the training set, 100 signals are considered as the validation set, and 200 are selected as the test set. The early-stopping technique is introduced in the training process, that is, if the recognition accuracy of validation set is not improved within 10 epochs, then the model would be stopped training. Adam is selected as the optimizer.

4.2.2 Verification experiments

To compare and evaluate our proposed model, it was made a comparison with the previous state of the art algorithm, [19, 21, 24, 26], respectively named VTCNN2, CLDNN, GRU2, LSTM-AP. Among them, VTCNN2, CLDNN, and

GRU2 took IQ raw signals as inputs directly and fulfilled the task of modulation recognition through well-designed models. LSTM-AP used the normalized instantaneous amplitude and phase series extracted from I/Q signals as inputs. We also visualized the attention mechanism introduced in our proposed model for the Bi-LSTM layer and analyzed the label smoothing techniques. In addition, the advantages and disadvantages of different algorithms in terms of computational complexity are respectively illustrated.

4.3 Results and discussion

4.3.1 The overall recognition accuracy

As shown in Fig. 3, a total of five algorithms are compared in the experiment. It can be found that when the SNR is over -6 dB, the recognition accuracy of our proposed algorithm is significantly better than that of other algorithms. Above 4 dB SNR, its recognition accuracy keeps approximately 92%. The maximum recognition accuracy can reach 92.68% at 16 dB. The average recognition accuracy is 91.3% from 0 dB to 18db SNR, which is improved by nearly 1% ~ 17% compared with other models. These results validate the stability and reliability of our proposed algorithm.

Figure 4 shows the effects of the label smoothing technique and the attention mechanism in Bi-LSTM on the modulation recognition. It shows that when these two techniques are introduced, above 0 dB SNR, the recognition accuracy of our algorithm is almost higher than that of the models without adding the label smoothing technique or the attention mechanism. Furthermore, even without these two optimization methods, the average recognition

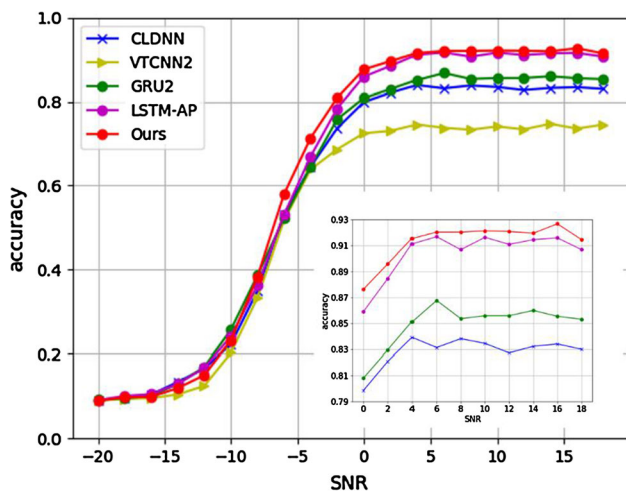


Fig. 3 The recognition accuracy of different algorithms from -20 dB to 18 dB SNRs. The figure at the bottom right corner shows the detailed information

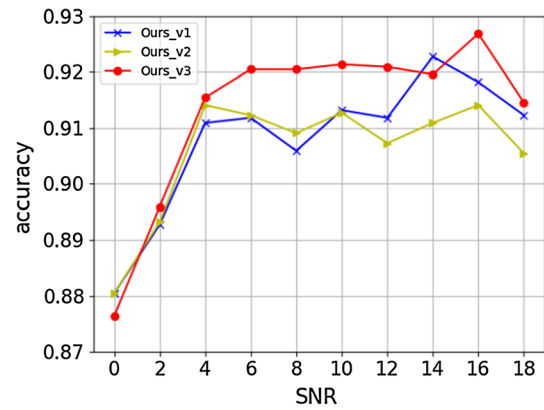


Fig. 4 The influences of label smoothing and attention mechanism techniques. *Ours_v3* denotes the holistic algorithm. *Ours_v2* denotes our algorithm without attention mechanism in the Bi-LSTM layer. *Ours_v1* denotes our algorithm without label smoothing technique

rate of our model over 0 dB SNR is still higher than that of the other aforementioned algorithms. It indicates that our algorithm, by fully exploiting implicit characteristics of I/Q signals from multiple dimensions, achieves the complementarity between different dimensional features. What's more, it is conducive to enhancing the robustness of our algorithm and implementing high recognition accuracy for different modulated signals.

4.3.2 Confusion matrix

To distinguish the difficulty of identifying each type of I/Q modulated signal by these models, we draw the confusion matrix generated by various models at 0 dB SNR. As demonstrated in Fig. 5, the horizontal axis represents the predicted modulation types, and the vertical axis represents the real modulation types. The diagonal reflects the recognition accuracy of each kind of modulation signals, and the darker color indicates higher recognition accuracy. Both WBFM and AM-DSB belong to analog modulation. The presence of silent periods in the analog audio signals leads to recognition difficulties between WBFM and AM-DSB. Besides, at lower SNRs, the model is prone to confuse 16QAM with 64QAM. This is because both 16QAM and 64QAM belong to the multiple quadrature amplitude modulation, and they have overlapped constellation mapping. Meanwhile, due to the noise disruption, it makes misrecognition more probable to occur at low SNRs. Although this is a common problem, our proposed model has obviously alleviated the difficulty of misrecognition. The accuracy of 16QAM and 64QAM by our model is higher than that of other models, which indicates that the misidentification of some modulation types at low SNR can be reduced by making full use of the correlation information of I/Q signals through multi-level feature fusion.

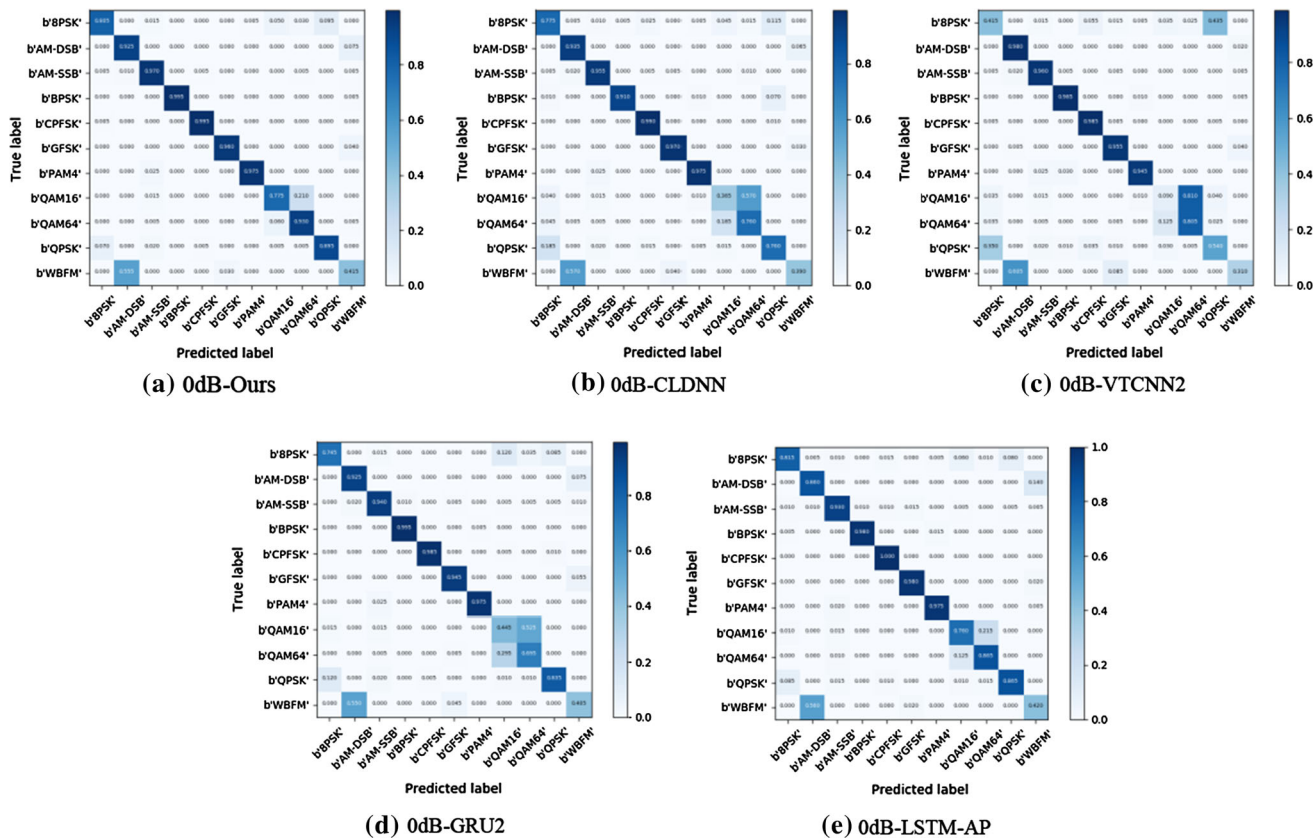


Fig. 5 The confusion matrix of **a** our proposed algorithm, **b** CLDNN, **c** VTCNN2, **d** GRU2, **e** LSTM-AP at 0 dB SNR

At high SNRs, since the modulated signal is almost free from noise pollution, it can better reflect its essential characteristics. Table 1 shows our model performs the highest average recognition rate of 91.5% for these 11 modulation types at 18 dB SNR. The recognition accuracy of our model for most of the modulation types is also higher than others. It further reveals the importance of multi-level feature mining, and it is conducive to improving the robustness of the modulation recognition algorithm.

4.3.3 Attention mechanism

It can be seen from Fig. 4 that introducing the attention mechanism can enhance the performance of our method. To further observe the role of attention mechanism played in the Bi-LSTM, we visualize the outputs obtained through the attention mechanism. As shown in Fig. 6, taking CPFSK and GFSK as examples, for different types of modulation signals, the attention distribution values from

Table 1 The recognition accuracy of different algorithms for each modulated types at 18 dB SNR, with the best performances are in bold

	VTCNN2 (%)	CLDNN (%)	GRU2 (%)	LSTM-AP (%)	Ours (%)
8PSK	36.0	87.5	96.5	99.5	98.5
AM-DSB	99.0	100.0	100.0	100.0	100.0
AM-SSB	98.0	95.5	93.5	94.0	97.5
BPSK	96.5	97.5	96.5	97.5	97.0
CPFSK	100.0	100.0	100.0	100.0	100.0
GFSK	96.0	97.0	100.0	100.0	100.0
PAM4	99.0	100.0	100.0	99.5	100.0
16QAM	14.0	36.0	40.0	85.0	86.0
64QAM	83.5	74.0	74.0	83.0	87.5
QPSK	70.0	88.5	98.0	98.5	99.0
WBFM	28.0	37.0	40.0	40.5	40.5
Average Accuracy	74.5	83.0	85.3	90.7	91.5

the Bi-LSTM have a significant temporal correlation. By weighting them with the extracted features, the main features of the corresponding modulated signals can be highlighted, and the redundant or unimportant features can be weakened, so as to improve the recognition accuracy of various types of modulated signals.

4.3.4 The computational complexity

Computational complexity is an indispensable index to evaluate the models. The average training time per epoch, the number of total training epochs, trainable parameters and classification time for mentioned models are compared in the experiment.

Figure 7 shows that the VTCNN2 model has the maximum number of trainable parameters, while the number of parameters in our model is similar to that of LSTM-AP, but larger than that of CLDNN and GRU2. This is because our algorithm extracts features of the IQ signals from multiple dimensions, and some parameters are appropriately added. The total number of our training epochs required is significantly lower than that of several other algorithms. Therefore, our model converges faster in a limited number of training epochs. To further reflect the time complexity of

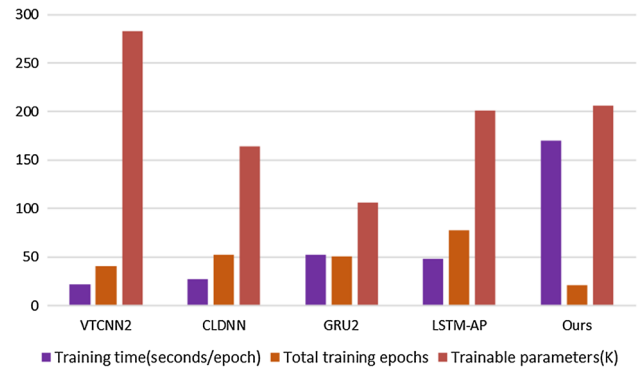
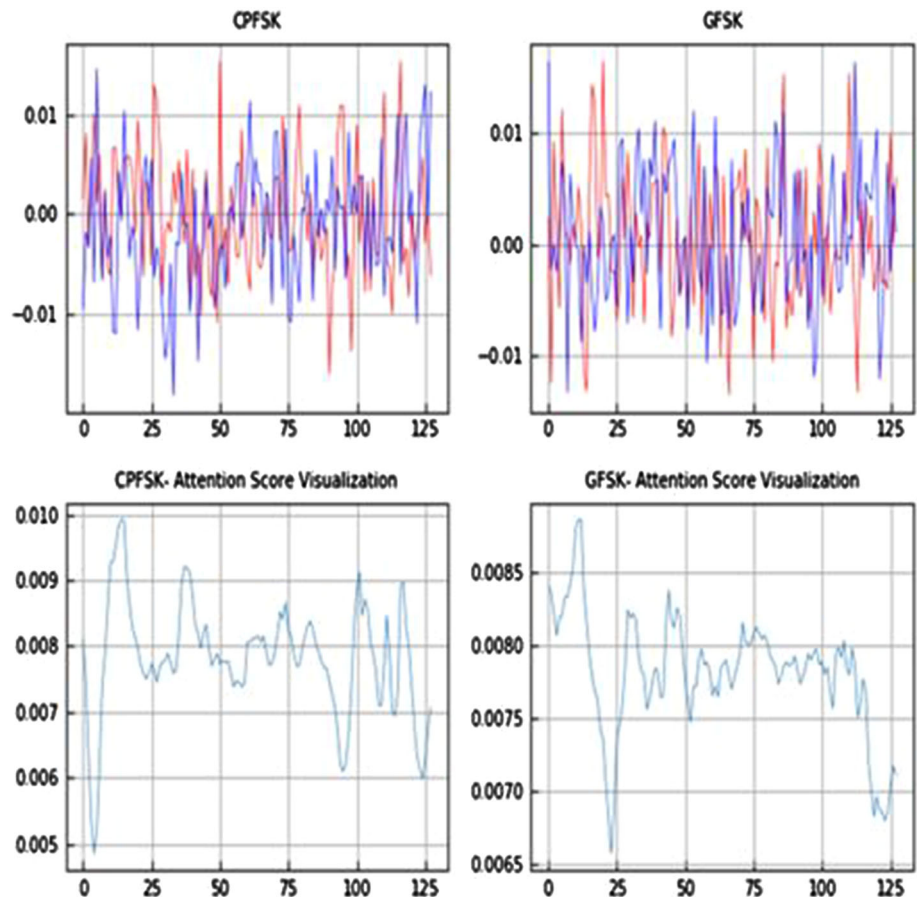


Fig. 7 The comparison of training time and trainable parameters of different models

different models, we recorded the time to classify a batch of samples. As shown in Fig. 8, our proposed model takes 2.59 s to identify 1024 examples, which is higher than the other models, but our model does not require extra running time to preprocess the IQ signal, like LSTM-AP. Compared with VTCNN2, CLDNN and GRU2, our proposed algorithm consumes a bit more time to extract the potential features of the modulated signal from higher levels, which enables the model to maintain a higher recognition accuracy and better robustness even at low SNRs.

Fig. 6 The Visualization of the Attention Mechanism



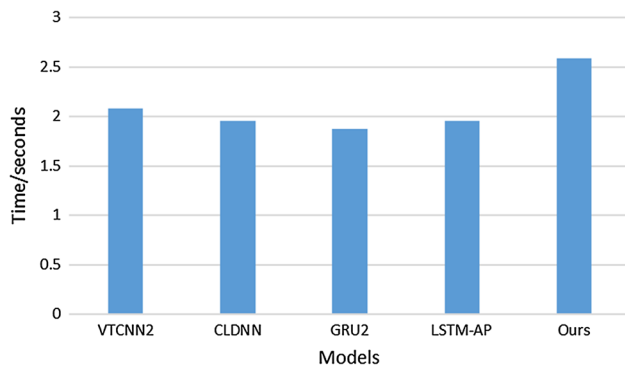


Fig. 8 Recognition time of AMR per batch (1024 samples)

5 Conclusion

Since the I/Q signals carry rich information, how to effectively extract the characteristics of the modulated signals has always been the focus of researches. Therefore, we propose a multi-level feature extraction algorithm that highlights the correlation between the I and Q channel. Our designed model could achieve complementation between spatial and temporal features and enrich the feature representation for modulated signals. The recognition accuracy of our proposed algorithm is higher than that of other models. Moreover, it also ameliorates the confusion situation between 16QAM and 64QAM. These results demonstrate the superiority of our algorithm. It also enriches the expression of modulation signal characteristics and the huge potential of deep learning for modulation recognition.

In general, the current modulation recognition algorithm based on deep learning is still over complex. In actual communication conditions, low time complexity and low computation is required. Thus, in the future, we would like to further investigate and design a miniaturized model to make feature extraction more effective.

Acknowledgements This work is supported by the National Key Research and Development Program of China (2019YFC1510705), the Sichuan Science and Technology Program (2020YFG0051), and the University-Enterprise Cooperation Projects (17H1199, 19H0355, 19H1121). We also thank Michael Tan of University College London for his writing suggestions.

References

1. Ali, A., & Hamouda, W. (2014). Spectrum monitoring using energy ratio algorithm for OFDM-based cognitive radio networks. *IEEE Transactions on Wireless Communications*, 14(4), 2257–2268.
2. Daskalakis, S. N., Correia, R., Goussetis, G., Tentzeris, M. M., Carvalho, N. B., & Georgiadis, A. (2018). Four-PAM modulation of ambient FM backscattering for spectrally efficient low-power

3. applications. *IEEE Transactions on Microwave Theory and Techniques*, 66(12), 5909–5921.
3. Li, P. (2019). Research on radar signal recognition based on automatic machine learning. *Neural Computing and Applications*, 32(7), 1959–1969.
4. Chen, Y., Yang, M., Long, J., Xu, D., & Blaabjerg, F. (2019). A DDS-based wait-free phase-continuous carrier frequency modulation strategy for EMI reduction in FPGA-based motor drive. *IEEE Transactions on Power Electronics*, 34(10), 9619–9631.
5. El-Mahdy, A. E., & Namazi, N. M. (2002). Classification of multiple M-ary frequency-shift keying signals over a Rayleigh fading channel. *IEEE Transactions on Communications*, 50(6), 967–974.
6. Wei, W., & Jerry M. M. (1995). A new maximum-likelihood method for modulation classification. *Conference Record of the Twenty-Ninth Asilomar Conference on Signals, Systems and Computers*. Vol. 2. IEEE.
7. Huan, C.-Y., & Polydoros, A. (1995). Likelihood methods for MPSK modulation classification. *IEEE Transactions on Communications*, 43(2/3/4), 1493–1504.
8. Sills, J. A. (1999). Maximum-likelihood modulation classification for PSK/QAM. MILCOM 1999. *IEEE Military Communications Conference Proceedings* (Cat. No. 99CH36341). Vol. 1. IEEE.
9. Swami, A., & Sadler, B. M. (2000). Hierarchical digital modulation classification using cumulants. *IEEE Transactions on communications*, 48(3), 416–429.
10. Gardner, W. A., & Chad, M. S. (1988). Cyclic spectral analysis for signal detection and modulation recognition. “MILCOM 88, 21st Century Military Communications-What’s Possible?”. *Conference record. Military Communications Conference*. IEEE.
11. Ho, K. C., W. Prokopiw, & Y. T. Chan. (1995). Modulation identification by the wavelet transform.” *Proceedings of MILCOM’95*. Vol. 2. IEEE.
12. Ho, K. C., Prokopiw, W., & Chan, Y. T. (2000). Modulation identification of digital signals by the wavelet transform. *IEEE Proceedings-Radar, Sonar and Navigation*, 147(4), 169–176.
13. Mobasseri, B. G. (2000). Digital modulation classification using constellation shape. *Signal processing*, 80(2), 251–277.
14. Avci, E., & Avci, D. (2009). Using combination of support vector machines for automatic analog modulation recognition. *Expert Systems with applications*, 36(2), 3956–3964.
15. Wang, C., Du, J., Chen, G., Wang, H., Sun, L., Xu, K., Liu, B., & He, Z. (2019). QAM classification methods by SVM machine learning for improved optical interconnection. *Optics Communications*, 444, 1–8.
16. Nandi, A. K., & Azzouz, E. E. (1998). Algorithms for automatic modulation recognition of communication signals. *IEEE Transactions on communications*, 46(4), 431–436.
17. Peng, S., Jiang, H., Wang, H., Alwageed, H., Zhou, Y., Sebdani, M. M., & Yao, Y. D. (2018). Modulation classification based on signal constellation diagrams and deep learning. *IEEE transactions on neural networks and learning systems*, 30(3), 718–727.
18. O’Shea, T. J., & Nathan W. (2016). Radio machine learning dataset generation with gnu radio. *Proceedings of the GNU Radio Conference*. Vol. 1. No. 1.
19. O’Shea, T. J., Johnathan C., & Charles T. C. (2016). Convolutional radio modulation recognition networks. “*International conference on engineering applications of neural networks*. Springer, Cham.
20. Yashashwi, K., Sethi, A., & Chaporkar, P. (2018). A learnable distortion correction module for modulation recognition. *IEEE Wireless Communications Letters*, 8(1), 77–80.
21. West, N. E., Tim, O. (2017). Deep architectures for modulation recognition. *2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*. IEEE.

22. Wang, Y., Liu, M., Yang, J., & Gui, G. (2019). Data-driven deep learning for automatic modulation recognition in cognitive radios. *IEEE Transactions on Vehicular Technology*, 68(4), 4074–4077.
23. Kawakami, K. (2008). Supervised sequence labelling with recurrent neural networks. *Ph. D. thesis*.
24. Hong, D., Zilong Z., & Xiaodong X. (2017). Automatic modulation classification using recurrent neural networks." *2017 3rd IEEE International Conference on Computer and Communications (ICCC)*. IEEE.
25. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. & Bengio, Y., (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.
26. Rajendran, S., Meert, W., Giustiniano, D., Lenders, V., & Pollin, S. (2018). Deep learning models for wireless signal classification with distributed low-cost spectrum sensors. *IEEE Transactions on Cognitive Communications and Networking*, 4(3), 433–445.
27. Ali, A. K., & Ergelebi, E. (2020). Automatic modulation recognition of DVB-S2X standard-specific with an APSK-based neural network classifier. *Measurement*, 151, 107257.
28. Li, R., Li, L., Yang, S., & Li, S. (2018). Robust automated VHF modulation recognition based on deep convolutional neural networks. *IEEE Communications Letters*, 22(5), 946–949.
29. Karra, K., Scott K., & Josh P. (2017). Modulation recognition using hierarchical deep neural networks." *2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*. IEEE.
30. Tian, X., Sun, X., Yu, X. and Li, X., (2019). Modulation pattern recognition of communication signals based on fractional low-order Choi-Williams distribution and convolutional neural network in impulsive noise environment." *2019 IEEE 19th International Conference on Communication Technology (ICCT)*. IEEE.
31. Zha, X., Peng, H., Qin, X., Li, G., & Yang, S. (2019). A deep learning framework for signal detection and modulation classification. *Sensors*, 19(18), 4042.
32. Xu, Y., Li, D., Wang, Z., Guo, Q., & Xiang, W. (2019). A deep learning method based on convolutional neural network for automatic modulation classification of wireless signals. *Wireless Networks*, 25(7), 3735–3746.
33. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 1097–1105.
34. Li, X., Huang, Z., Wang, F., Wang, X., & Liu, T. (2018). Toward convolutional neural networks on pulse repetition interval modulation recognition. *IEEE Communications Letters*, 22(11), 2286–2289.
35. Wei, S., Qu, Q., Wu, Y., Wang, M., & Shi, J. (2020). PRI modulation recognition based on squeeze-and-excitation networks. *IEEE Communications Letters*, 24(5), 1047–1051.
36. Hermawan, A. P., Ginanjar, R. R., Kim, D. S., & Lee, J. M. (2020). CNN-based automatic modulation classification for beyond 5G communications. *IEEE Communications Letters*, 24(5), 1038–1041.
37. Liu, R., Yunxin, G., & Shibing Z., (2020). Modulation recognition method of complex modulation signal based on convolution neural network." *2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*. Vol. 9. IEEE.
38. Li, M., Li, O., Liu, G., & Zhang, C. (2018). Generative adversarial networks-based semi-supervised automatic modulation recognition for cognitive radio networks. *Sensors*, 18(11), 3913.
39. Bu, K., He, Y., Jing, X., & Han, J. (2020). Adversarial transfer learning for deep learning based automatic modulation classification. *IEEE Signal Processing Letters*, 27, 880–884.
40. Ioffe, S., & Christian S. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift." *International conference on machine learning*. PMLR.
41. Hu, J., Li, S., & Gang S. (2018). Squeeze-and-excitation networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*.
42. He, K., Zhang, X., Ren, S. & Sun, J., (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*.
43. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735–1780.
44. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I., (2017). Attention is all you need. *arXiv preprint arXiv:1706.03762*.
45. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z., (2016). Rethinking the inception architecture for computer vision. *Proceedings of the IEEE conference on computer vision and pattern recognition*.
46. Müller, R., Simon K., & Geoffrey E. H. (2019). When does label smoothing help?." *Advances in Neural Information Processing Systems*.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Hang Zhang received the B.S. degree from the Southwest Petroleum University, Chengdu, China. He is currently pursuing the M.S. degree in Control Engineering at the College of Electrical Engineering, Sichuan University in Chengdu, China. His research interests include artificial intelligence and signal processing.



Ruihua Nie received his Ph.D. from Sichuan University, Chengdu, China. He is currently a professor in the State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University, Chengdu, China. His research interests include riverbed evolution and disaster prevention.



Minghui Lin is currently pursuing the B.S. degree in Automation at the College of Electrical Engineering, Sichuan University in Chengdu, China. Her research interests include signal processing and deep learning.



Xiaofeng Gong received Ph.D. degree from Zhejiang University, Hanzhou, China. He is recently working as professor in College of Electrical Engineering, Sichuan University. His research interests include radio spectrum detection and process control.



Ruijuan Wu received her M.S. degree from Sichuan University in Chengdu, China. She is working in the Chengdu Dagongbochuang Information Technology Co., Ltd., Chengdu, China. Her research interests include radio spectrum detection and signal processing.



Qin Yu is currently a supervisor of master degree in Sichuan University. Her research interests include intelligent systems, graphics and image processing.



Guo Xian received his M.S. degree from Sichuan University in Chengdu, China. He is working in the Chengdu Dagongbochuang Information Technology Co., Ltd., Chengdu, China. His current research interests include signal processing and machine learning.



Ruisen Luo received his Ph.D. from Zhejiang University, Hangzhou, China. He is currently an associate professor and a graduate student supervisor in College of Electrical Engineering, Sichuan University. His research interests include artificial intelligence and signal processing.