Attention Mechanism Enabled Wireless Modulation Recognition Network

Xuelin Chen, Xinyu Tian, Qinghe Zheng*, Weiguang Wang

Abstract—**With the rapid development of communication systems, we are facing unprecedented challenges that require real-time and meticulous processing of massive amounts of wireless signals. For complex electromagnetic environments, signals can be usually affected by various interferences and noise, resulting in changes in the waveform, frequency, phase, and other characteristics of the signal. In response to this issue, we propose an efficient classifier based on convolutional neural network (CNN) with attention mechanism, aiming to improve the automatic modulation classification (AMC) accuracy in complex electromagnetic environments. The dynamic weighting function is achieved through attention mechanism, allowing the model to focus on the more important parts of the input signal, thereby improving AMC accuracy. Furthermore, the model is able to capture the global features of the signal rather than local features based on the global information perception ability of attention mechanism, which further enhances the model's generalization capability. In addition, the adaptive learning characteristics of attention mechanism are utilized to enable the model to adaptively adjust attention weights during the training process, thus to better adapt to different modulation schemes and parameter changes. This characteristic enables the model to maintain high classification accuracy in the face of complex electromagnetic environments and diverse modulation schemes. Finally, through the resource optimization function of attention mechanism, both the computational complexity and storage requirements of CNN can be effectively reduced, and the operational efficiency is improved. Using the proposed method that combines CNN and attention mechanism, AMC accuracy of over 90% under 11 different modulation schemes is achieved, with four schemes even achieving an accuracy rate of 100% at +18 dB. Even in harsh environments of -6dB, the accuracy remains above 52.82%, which is 9.4% higher than traditional methods. Experiments have demonstrated the effectiveness and reliability of the proposed method in AMC task in complex electromagnetic environments.**

Index Terms—**automatic modulation classification, attention mechanism, frequency-domain analysis, deep learning, wireless communication.**

Manuscript received May 24, 2024; revised January 9, 2025.

This work was supported in part by Shandong Provincial Natural Science Foundation under Grant No. ZR2023QF125.

Xuelin Chen is an undergraduate student of Shandong Management University, Jinan 250357, China. (email: 370345313@qq.com)

Xinyu Tian is a lecturer of Shandong Management University, Jinan 250357, China. (email: txy@sdmu.edu.cn,)

Qinghe Zheng is an associate professor of Shandong Management University, Jinan 250357, China. (corresponding author to provide phone: 89636095; fax: 89636095; e-mail: zqh@sdmu.edu.cn)

Weiguang Wang is an associate professor of Shandong Management University, Jinan 250357, China. (email: wwg@sdmu.edu.cn)

I. INTRODUCTION

IN the field of wireless communication, signal modulation recognition technology plays a crucial role. From a civilian perspective, it can not only supervise the legitimate operation of communication and prevent illegal encroachment and theft of resources, but also effectively avoid the radio interference, spectrum optimization, and the reliability and intelligence improvement of cognitive radio systems [1][2][3][4]. In the military field, the AMC is also a key technology for space electromagnetic countermeasures [5][6][7] and information acquisition [8][9][10][11], which is of great significance for ensuring the wireless communication security. As the core technology of wireless communication [12], AMC is help for ensuring communication quality [13], spectrum management [14], and signal monitoring [15][16][17][18]. However, the accurate identification of various modulation schemes in the complex real-world communication environments remains a highly challenging task [19][20][21].

At present, AMC is mainly divided into the following two categories: likelihood-based methods [22][23] and feature based methods [24] [25][26]. The likelihood-based method utilizes probability theory and statistical decision to infer the modulation scheme of the received signal based on its statistical characteristics. It is widely used in the performance evaluation of wireless communication system [27] and signal reconnaissance in electronic warfare [28]. The feature-based methods extract features representing modulation schemes from the signal for AMC, which is suitable for real-time demodulation of wireless communication systems and radar target recognition scenarios [29].

With the rapid development of deep learning technology, models such as convolutional neural networks (CNN) [30] [31][32][33] and Transformers [34][35][36][37] have shown strong potential in communication signal processing. CNN excels at capturing the spatial local features of signals while Transformer and its inherent attention mechanism emphasize understanding and modeling the global information of signals [38][39][40]. In response to the unique demand for AMC of radio signals [41][42][43][44], we propose a hybrid deep learning model that combines CNN and attention mechanisms in Transformer in this paper. The time-domain attention mechanism is used to capture the dynamic evolution features of signals over time, while the frequency-domain attention mechanism reveals the key patterns of signals in spectral distribution, thereby effectively processing the time-series and frequency-domain characteristics of communication signals. Finally, the proposed method has been validated and optimized on the widely used RadioML 2016.10A dataset, aiming to improve the accuracy and robustness of AMC, and provide more reliable signal monitoring and recognition technology for the sixth generation (6G) and beyond 6G wireless communication systems.

The rest of the paper is organized as follows. Section II introduces the proposed method for AMC. In Section III, we describe the experimental results and corresponding analysis. Finally, the conclusion and future work are summarized in Section IV.

II. The Proposed Method

The processing flow of the proposed CNN based method is consisted of the following steps: signal preprocessing, feature extraction, residual module, attention mechanism, global average pooling, fully connected, and classification module, as shown in Fig. 1. Firstly, the received signal is preprocessed to obtain the signal representation that adapts to subsequent processing requirements, such as noise suppression, signal smoothing, normalization, etc. Next, we utilize the powerful feature extraction capability of combining CNN feature extraction module with residual module, multi-scale features of the signal are extracted, and high-dimensional feature detail maps are output. Then we use both the time-domain attention and the frequency-domain attention mechanism to focus on the dynamic characteristics of signal evolution over time and learn the key patterns of signal distribution in the spectrum. Through the global average pooling module, the feature dimensionality and parameter can be reduced, and classification results through the classification module can be finally output. The following is a detailed introduction to each part.

A. Dataset and Preprocessing

● *Dataset*

The RadioML 2016.10A dataset is one of the widely used standard datasets in the field of signal modulation recognition, aiming at supporting and evaluating the AMC performance of various methods. This dataset includes two categories: analog modulation and digital modulation, covering a total of 11 modulation formats. Analog modulation schemes include: double-sideband amplitude modulation (AM-DSB), single-sideband amplitude modulation (AM-SSB), and broadband frequency modulation (WBFM), while digital modulation schemes contain 8-phase shift keying (8PSK), binary phase shift keying (BPSK), continuous phase shift keying (CPFSK), high frequency shift keying (GFSK), four level pulse amplitude modulation (PAM4), multiple levels of orthogonal amplitude modulation (16QAM, 64QAM), and orthogonal phase shift keying (QPSK).

All modulated signals are digitized at a sampling rate of 200KHz, with each symbol corresponding to 8 sampling points, ensuring sufficient sampling accuracy to capture signal features. The signal-to-noise ratio (SNR) in the dataset is set between -20 dB and +18 dB, adjusted every 2 dB to simulate a wide range of noisy environments. This setting enables the evaluation of the recognition performance of various methods under different noise levels.

At each specified SNR level, each modulation scheme has 1000 independent data samples. Each data sample contains 128 sampling points and is organized in a 2×128 matrix form, where two dimensions represent the real (I) and imaginary (Q) components of the radio signal, fully reflecting the complex characteristics of the received signal.

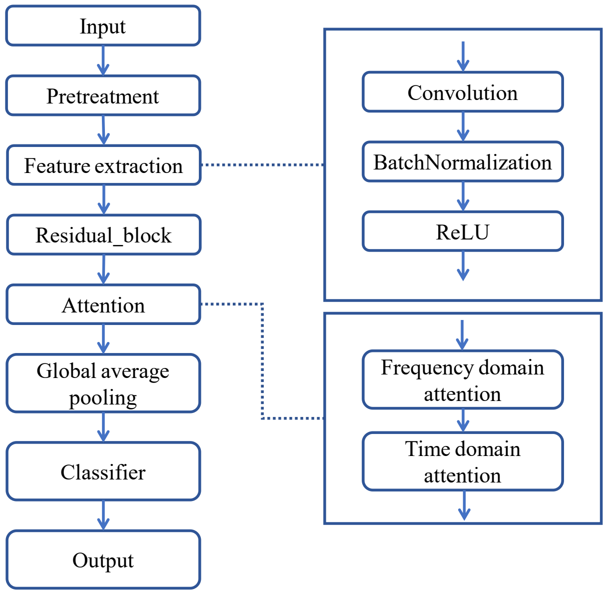

Fig. 1. Flow chart of modulation recognition method.

When the SNR is at +8 dB, the time-domain and frequency-domain plots of each modulation scheme can visually display their characteristics as shown in Fig. 2. These graphs are very helpful for understanding and analyzing the performance of different modulation methods under specific noise conditions.

The RadioML 2016.10A dataset provides researchers with a comprehensive and rich resource for developing and validating various AMC methods. By utilizing this dataset, the performance of a series of methods can be evaluated, compared and improved to promote the further development of AMC technology.

● *Pretreatment*

The deep learning network designed by combining CNN and Transformer has strong expressive power, and one important factor is the massive and diverse data support. In practical communication environments, not all datasets meet the requirements of complexity and diversity. Therefore, effective data denoising techniques can not only reduce noise interference, scale inconsistency or distribution differences in signal data, but also effectively avoid reducing the model overfitting problem. Therefore, we adopt data normalization to enhance the diversity of the dataset.

In order to better prepare the input feature matrix of the deep neural network model, we further implement the data normalization steps for data sequences containing complex components, i.e., in-phase (I) and orthogonal (Q) components, with particular attention to the processing of phase data.

Then we use mean elimination and amplitude adjustment: The initialization step involves subtracting the overall mean $E(a(n))$ from each data sample $a(n)$ to eliminate bias effects in the data and generate the intermediate variable $ac(n)$. Subsequently, amplitude standardization was achieved by dividing each element of $ac(n)$ by its maximum absolute value in its set, ensuring that the processed data $an(n)$ is constrained within a clear range, as expressed by

$$ac'(n) = a(n) - \mu_a \tag{1}$$

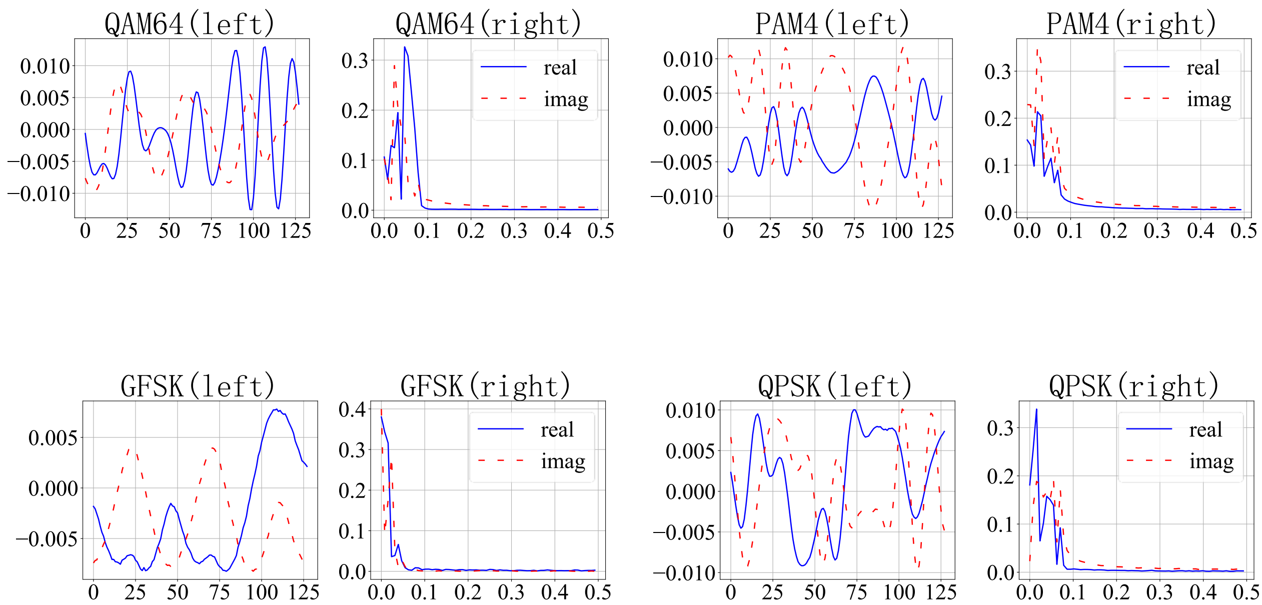

Fig. 2. Time domain distribution (left) and frequency domain distribution (right) of different modulation signals.

$$an(n) = \frac{ac'(n)}{\max(|ac'(n)|)} \tag{2}$$

where μ_a represents the average value of $a(n)$, which ensures the scale consistency of the data.

Normalization of phase data: When processing phase data, it is important to note that directly using the [0,2Π] range may lead to phase wrapping. To avoid this, phase values should be remapped to the [-Π, Π] range. This processing maintains the continuity and intuitiveness of phase information, avoiding unnecessary periodic repetition.

Amplitude and phase extraction: During the construction of the feature matrix, the amplitude A and phase of the signal are calculated using I/Q data ϕ, which can be computed by

$$A' = I^2 + Q^2 \tag{3}$$

$$\Phi' = a\tan 2(Q, I) \tag{4}$$

where the usage of the $a$tan2 function ensures the accurate calculation of all quadrant phases. By compared it to the traditional arctan function, it can provide the correct phase angle for the four quadrants.

B. Model Structure

● *CNN Feature Extraction Module*

The self-attention mechanism of Transformer can capture the dependency relationships between different positions in the input sequence, thereby better modeling the long-distance dependencies, which enables the model to possess better performance in processing sequence data, especially when capturing relationships between distant elements. However, due to the lack of handling of local details in the Transformer, it may overlook the local features in the sequence and the relationship between adjacent elements. To compensate for the shortcomings of Transformer in local detail processing, CNN and residual convolution are used in the front end of the network model to extract signal detail features to enhance the model's ability to capture local information and improve its ability to process local details.

To extract the multi-scale features of signals, we propose to use a feature extraction module as shown in Fig. 2. The feature extraction layer in CNN first uses the 2D convolution kernels of size 1×8 to expand the channel of input sequence signal to 32, and then uses batch normalization to accelerate the training process of the neural network. As the network parameters are constantly changing according to the gradient descent, the distribution of each network layer will change differently. Therefore, batch normalization normalizes the model to pay more attention to the overall distribution of the data, which helps the model learn more common features and improve its generalization ability. In deep neural networks, as the number of layers increases, the gradient may gradually decrease or even approach 0 during the back-propagation process, resulting in the problem of "gradient vanishing". This makes it difficult for the model to update parameters during the training process, thereby affecting the training effectiveness. Since ReLU activation function output 0 when the input is less than or equal to 0, and is able to output the input value itself when the input is greater than 0. This non-linear processing way enables ReLU to better handle non-linear problems and improve the model's representation ability. Therefore, incorporating a ReLU activation function helps the model better fit the complex data distribution and improve the model performance.

Originating from the problem of network degradation that deep neural networks may encounter during training, it may gradually decrease during back-propagation process, making it difficult for the model to update parameters during training, thereby affecting training effectiveness. We propose to use the residual block to solve this problem, as shown in Fig3. In the residual module, two layers of feature extraction are used to enhance the feature extraction capability. By using the skip connection, information is directly transmitted from shallow layers to deep layers, introducing more hierarchical features and improving the efficiency of the training process. This structure ensures the avoidance of gradient vanishing during backpropagation.

● *Attention Mechanism Module*

When processing signal data, it can cause critical temporal information to be overwhelmed, as wireless communication signals, particularly continuous time series signals, typically contain a large amount of information. Only a small portion of this information is crucial for specific processing tasks. However, due to the complexity and redundancy of signal data, this key temporal information may be submerged in a large amount of data, making it difficult for the model to accurately identify and extract them. In response to this issue, we use a time-domain attention mechanism module to solve the problem of key information flooding.

The core of the time-domain attention module is to extract key features in the time domain by introducing the multi head self-attention mechanism. To achieve this goal, we construct an architecture consisting of four Transformer encoders stacked together as shown in Fig. 4. Since Transformer was originally designed to process sequential data, we need to serialize and transform the two-dimensional feature map output by the channel attention module before inputting the features into Transformer encoder. The conversion process of serialization is as follows. We first calculate the query vector Q_i for each position i, key vector K_j, and value vector V_j, where j traverses all positions. Then we can calculate the attention weight a_{ij}, which reflects the attention of i to position j. The specific calculation is given by

$$a_{ij}=\frac{\exp\left(\frac{Q_i\cdot K_j}{\sqrt{d_k}}\right)}{\sum_{l=1}^{T}\exp\left(\frac{Q_i\cdot K_l}{\sqrt{d_k}}\right)} \tag{5}$$

where d_k is the dimension of the key vector k_j, which is used to scale the dot product attention score to avoid excessive gradients or small country sizes during the training process. T represents the total number of positions in the sequence.

After obtaining the attention weights, the value vectors can be weighted and summed based on these weights to obtain the output vectors for each position as shown in

$$o_i=\sum_{j=1}^{T}a_{ij}v_j \tag{6}$$

where o_i denotes the weighted combination of all positional information in the original input sequence, where the key information can receive more attention.

During conducting the above transformation, we adopt a specific strategy as follows. we first segment the feature maps into multiple segments along the timeline and then use the linear transformation to map these segments to high-dimensional space to create the segment level serialized representation. To ensure that these serialized elements retain their relative positional information in the original feature map, we further combine these serialized representations

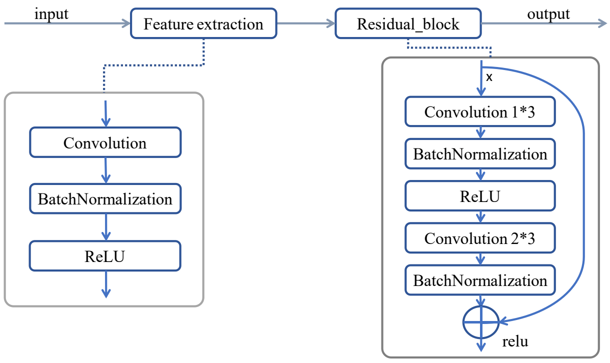

Fig. 3. Efficient feature extraction module.

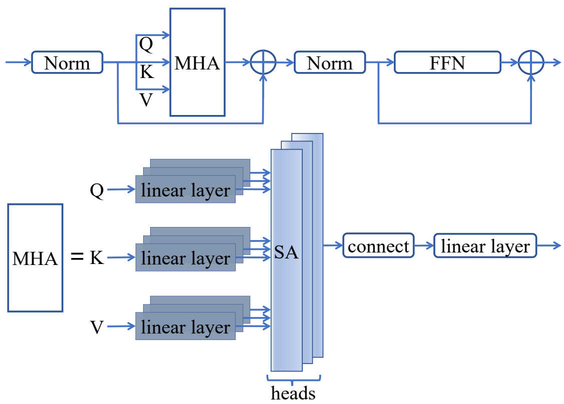

Fig. 4. Time domain attention mechanism.

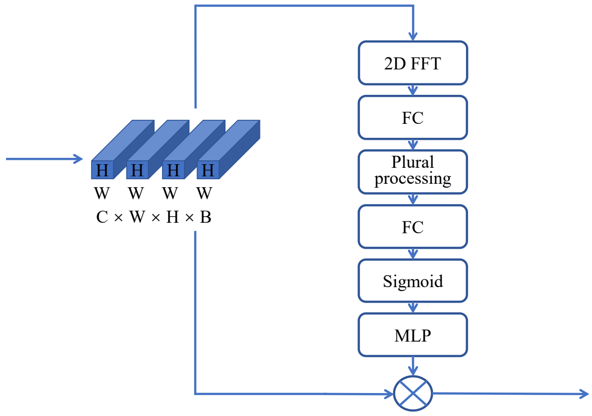

Fig. 5. Frequency domain attention mechanism.

with corresponding positional embedding vectors to obtain the final input sequence, which can be used as input to the Transformer encoder for subsequent processing. This design enables us to effectively convert two-dimensional feature maps into serialized data that Transformers can process, and capture key features in the time domain through self-attention mechanisms, providing powerful tools for signal analysis.

Signals can be usually affected by various noises and interferences during transmission and acquisition, which can mask useful information in the signal and cause interference in the model during signal processing, thereby reducing the performance. In this case, we further adopt the frequency domain attention mechanism, as shown in Fig. 5. By converting the signal from time domain to frequency domain, the model can analyze the characteristics of the frequency components in the signal and focus on those frequency components that contain useful information. Meanwhile, the frequency domain attention mechanism can also reduce the impact of noise and interference on the model performance, and improve the robustness of the model.

In the processing of radio signals, samples usually have different frequency components in frequency, which reflect the changing characteristics of the signal at different time scales. However, due to the complexity of the signal, it may contain a large amount of redundant information and noise, which are difficult to distinguish and remove in the time domain. Based on this issue of redundancy and noise, we adopt the frequency domain attention mechanism. Firstly, the input signal is preprocessed into a matrix that contains batch size, channels, height, and width. Then, the time-domain signal is converted to the frequency-domain after undergoing a two-dimensional Fourier transform, as calculated by

$$F(k,l)=\sum_{m=0}^{height-1}\sum_{n=0}^{width-1}X(m,n)\cdot e^{-j2\pi\left(\frac{mk}{height}+\frac{nl}{width}\right)} \tag{7}$$

where k and l are frequency indices, j is an imaginary unit, $e^{-j2\Pi}$ represents the complex exponential function to calculate the contribution of different frequency components.

Considering that the frequency domain representation is a complex matrix, the complex processing layer is shown in Eq. (8), followed by a multi-layer perceptron (MLP) as shown in Eq. (9). Finally, the frequency domain features and attention weights are multiplied as Eq. (10), and the calculated results show the weighted frequency domain features. The method based on frequency domain attention mechanism can reduce the weight of noises and redundant information, thereby achieving suppression of these components.

$$A_{real}, A_{imag} = \Pr ocessComplex(F) \tag{8}$$

where A_{real} and A_{imag} denote the processed real and imaginary features, respectively.

$$a = Soft\max\left(MLP\left(A_{combined}\right)\right) \tag{9}$$

where $A_{combined}$ is the input that combines the features of the real and imaginary parts, while the Softmax function ensures that the total weight is 1, achieving a probability distribution.

$$F_{att}\left(k,l\right)=a\left(k,l\right)\cdot F\left(k,l\right) \tag{10}$$

We apply the obtained attention weight a to the original frequency domain feature F to obtain the weighted frequency domain feature F_{att}. In this way, the contribution of important frequency components in the signal is amplified, while the contribution of noise and redundant information is relatively reduced, improving the model's attention and robustness to useful signal components.

● *Global Average Pooling and Classification Module*

In order to effectively reduce the input dimension of the classification module, a global average pooling operation is performed on the output values processed in the time and frequency domains, taking the average of all elements in each feature map to generate a single scalar value. Due to the strong robustness of the global average pooling to changes in the spatial position, it can also enhance the generalization capability of the model. After global average pooling, the probabilities of each category are obtained through linear mapping and Softmax function, thereby achieving AMC.

III. Experimental Results and Analysis

A. Experimental Configuration

To test the performance of the proposed method, the dataset RadioML 2016.10A dataset was used for evaluation. The dataset is consisted of 220000 signals, with 60% used as the training set, 20% used as the testing set, and 20% used as the validation set. Three deep learning models including ResNet, LSTM, and CNN are selected for comparative experiments. During the training process, the optimizer uses Adam, the loss function uses cross entropy, the initial learning rate is set as 0.0001, the factor is 0.5, the training batch size is 256, and the epoch is 100. When the loss on the validation set is not decreasing, the learning rate is reduced to half of the original until the training process is end.

B. Evaluating Indicator

We use commonly used evaluation indicators for AMC to quantitatively evaluate the performance of the model. These indicators include accuracy (A), accuracy (P), recall (R), F1 score ($F1$), and confusion matrix, which can be computed using Eqs. (11), (12), (13), and (14).

$$\mathrm{A}=\frac{TP+TN}{TP+TN+FP+FN} \tag{11}$$

$$\mathrm{P}=\frac{TP}{TP+FP} \tag{12}$$

$$R=\frac{TP}{TP+FN} \tag{13}$$

$$F1=\frac{2\times P\times \mathrm{R}}{\left(P+R\right)} \tag{14}$$

where A represents accuracy, TP represents the number of samples that are actually positive and predicted to be positive. FN denotes the number of samples that are actually positive and predicted to be negative. FP represents the number of samples that are actually negative and predicted to be positive. TN is the number of samples that are actually negative and predicted to be negative. $F1$ considers both accuracy and recall rate, and is a comprehensive indicator that reflects the overall performance of the network.

C. Experimental Results and Analysis

● *Experimental Result*

To observe the AMC efficiency of different modulation schemes in depth, the confusion matrix can be a powerful analytical tool and is shown in Fig. 6, which shows the confusion matrix of the proposed method under a series of SNRs. As the SNR gradually increases, the classification accuracy of modulated schemes also shows a significant improvement. Taking the extreme SNR levels of -6 dB and 18 dB as examples, the AMC accuracy of modulated signals in different wireless communication conditions can be clearly observed.

In the environment with low SNR of -6dB, communication conditions are relatively noisy, but the recognition rate of the modulated signal can still reach 52.82% even under such unfavorable conditions, indicating a certain adaptability of the modulation method in complex environments. In the

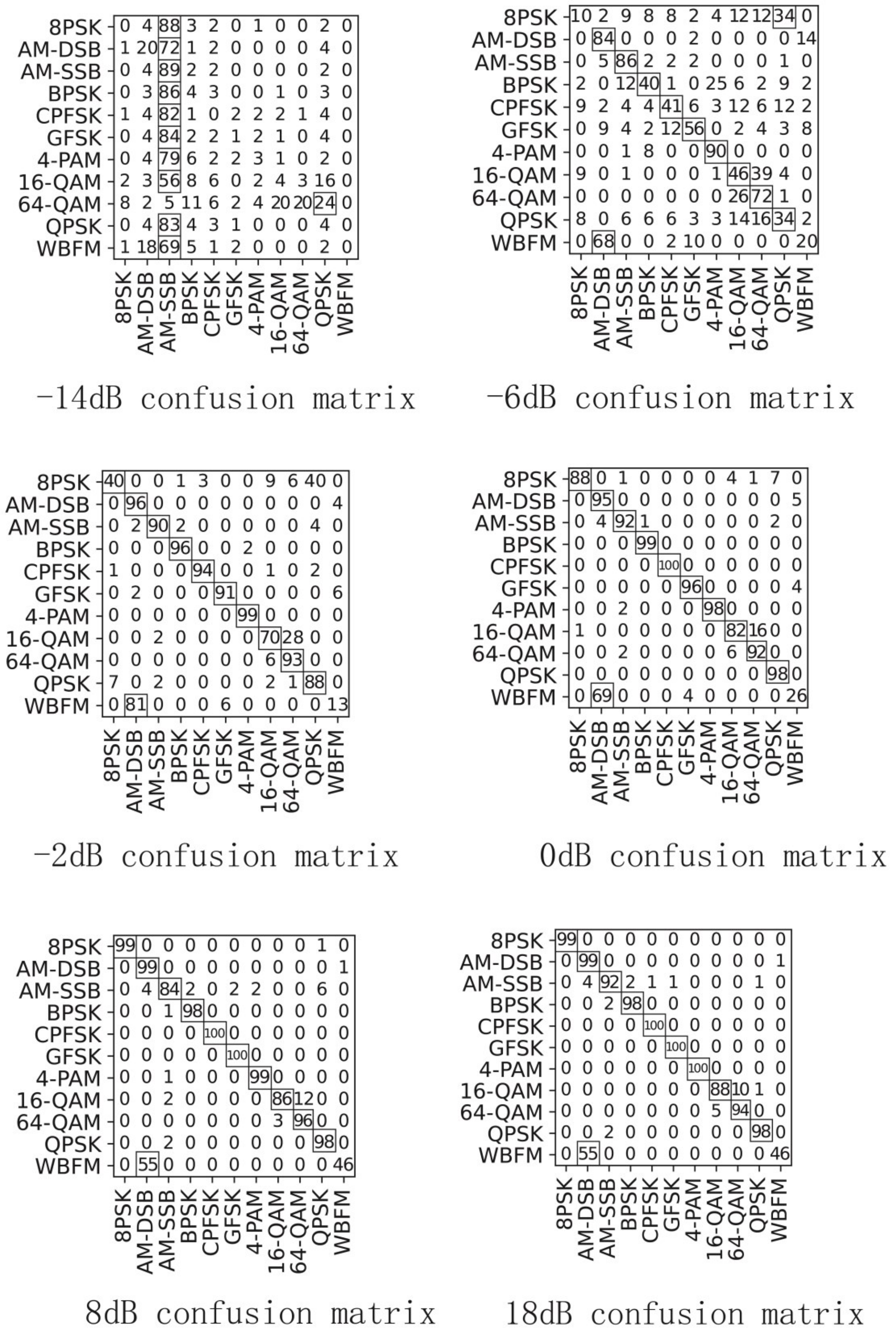

Fig. 6. Confusion matrix under different SNRs.

environment with a high SNR of 18dB, the communication quality has been significantly improved and the interference is reduced. At this time, the recognition rate of the modulated signal is as high as 92.18%, fully demonstrating the excellent performance of the proposed AMC method in clear and stable communication scenarios.

By conducting a detailed analysis of the recognition rates at these two extreme but representative SNR levels, we can gain a more comprehensive and in-depth understanding of the adaptability and robustness of various modulation schemes in different communication scenarios. This not only helps us choose the appropriate modulation method in actual communication systems, but also provides a strong data support for the optimization and improvement of future AMC technologies.

By analyzing the trend of accuracy changing with SNR, it can be found that the model exhibits high accuracy in identifying 8PSK, AM-DSB, AM-SSB, BPSK, CPFSK,

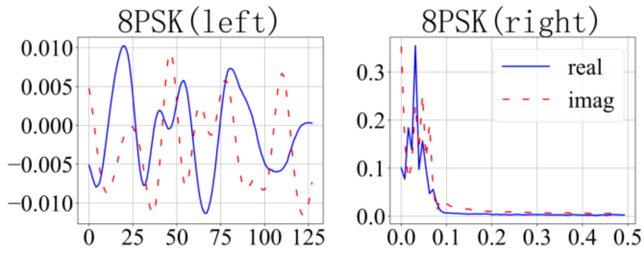

Fig. 7. Time domain distribution (left) and frequency domain distribution (right) of 8PSK.

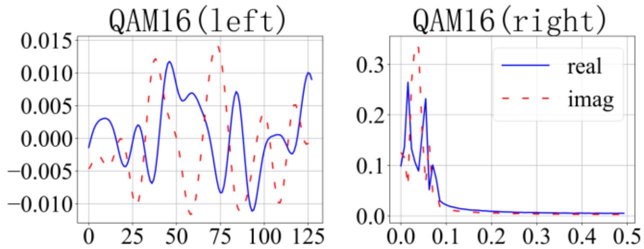

Fig. 8. Time domain distribution (left) and frequency domain distribution (right) of QAM16.

GFSK, 4-PAM, 16-QAM, 64-QAM, and QPSK signals. Due to the similarity between WBFM and AM-DSB signals, the model often misjudges WBFM signals and mistakenly identifies them as AM-DSB signals. Due to the inherent similarity between these two types of signals, the method proposed in this paper has limitations in feature extraction, resulting in misclassification of WBFM and AM-DSB. This question indicates that an important direction for future research is how to improve the model to better capture and distinguish subtle differences in identifying highly similar signals. In Fig. 7 and Fig. 8, we present the time-domain and frequency-domain plots of 8PSK and QAM16 modulated signals extracted from RML2016.10A dataset, respectively.

- *Contrast Test*

To further evaluate the performance of the proposed method, we have conducted comparative experiments on the RML2016.10a dataset. All experiments were conducted on the same dataset, hyper-parameter settings, and training strategy, and compared and analyzed with various types of methods.

In Fig. 9, we show the classification accuracy of LSTM, ResNet18, and KAN models in different modulation recognition tasks. The recognition ability of each method for specific modulated signals is observed under different SNR conditions. Specifically, the LSTM method performs poorly in the recognition of QAM16 signals, failing to effectively learn signal features at any SNR. In contrast, the overall recognition rate of ResNet18 network has been improved at various SNRs, indicating its advantage in feature extraction. However, ResNet18 still has limitations in distinguishing AM-DSB and QAM64 modulation schemes with the low recognition accuracy, indicating that its generalization ability on complex or high-order modulation signals still needs further optimization.

The KAN neural network shows significant shortcomings in the recognition of 8PSK, QAM64, QPSK, WBFM, and QAM16 signals, with generally low recognition rates. This indicates that KAN has significant limitations in extracting and distinguishing these signal features, especially when faced with similar or similar modulation patterns, it is often difficult to make accurate judgments, and even leads to misidentification or inability to recognize. This phenomenon highlights the challenges of existing methods in processing

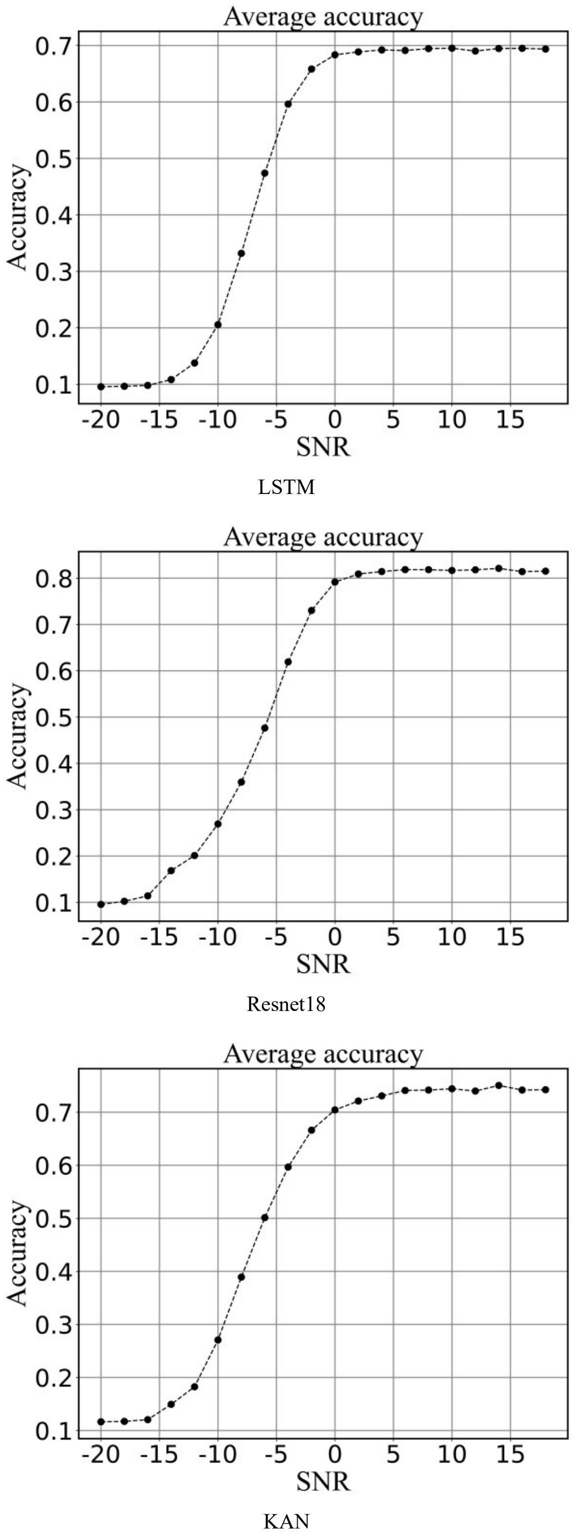

Fig. 9. The accuracy of different modulation recognition.

complex radio signals, limiting their effectiveness in radio signal analysis and applications. Further research is needed to address these shortcomings and enhance the robustness and identification capabilities of methods to adapt to diverse and complex wireless communication environments.

In Fig. 10, we show the accuracy performance of our proposed method in the same task. As shown in Fig. 10, the method proposed in this paper exhibits significant advantages in recognition accuracy for multiple modulation signals under different SNR conditions. Except for the high similarity in modulation characteristics between WBFM signal and AM-DSB signal, which results in low recognition

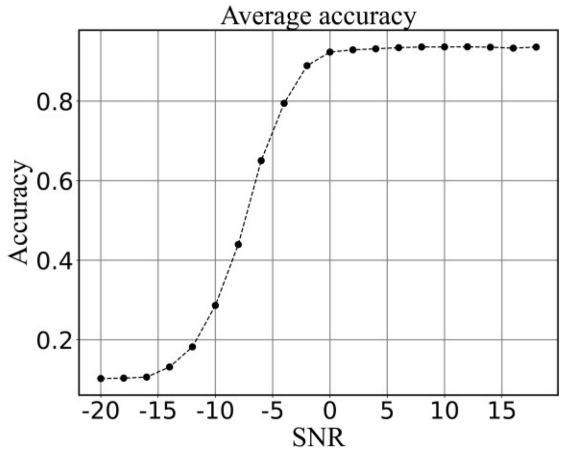

Fig. 10. The accuracy of different modulation recognition.

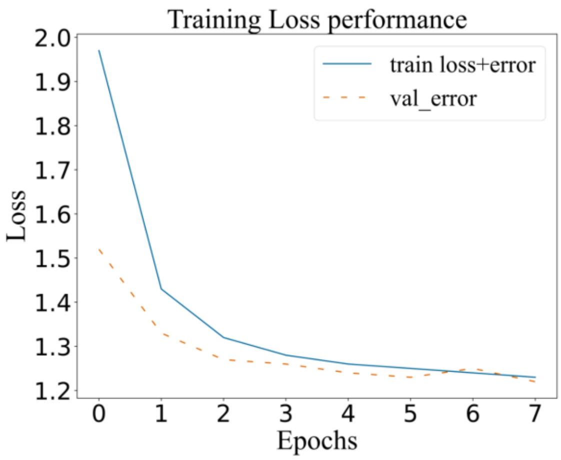

Fig. 11. The accuracy of different modulation recognition.

TABLE I
THE MODULATION RECOGNITION ACCURACY OF DIFFERENT MODELS

Models	Parameters/KB	Accuracy/%
Transformer	705	63.4
LSTM-CNN	460	67.5
ResNet34	994	72.2
ResNet18	730	71.4
LSTM	206.4	62.7
KAN	140.2	66.3
ours	508	78.5

accuracy at any SNR, the recognition accuracy of the other ten modulation signals exceeds 90%, and in some cases even approaches 100%. Even in extremely noisy environment with a SNR of -4 dB, the overall recognition accuracy of our proposed method can still remain above 80%, demonstrating the strong robustness of the method at low SNRs. When the SNR is increased to 0 dB or above, the overall recognition accuracy of the method is significantly improved, almost exceeding 90% across the entire line, approaching the optimal performance. This result fully demonstrates the wide adaptability and excellent performance of the proposed method in complex signal environments, providing reliable technical support for efficient modulation recognition in wireless communication. Further analysis shows that the proposed method can effectively capture and distinguish subtle features of different modulation signals, especially when facing high-order modulation schemes, demonstrating the excellent recognition ability and potential for application in modern wireless communication systems.

In Fig. 11, we present the training loss curve of the proposed model, from which it can be clearly observed that the method quickly converges in a short period of time and reaches a stable state within the initial epoch. Specifically, during the training process, the loss value showed a smooth downward trend with small fluctuations, indicating that the method has strong robustness and stability in parameter adjustment. As the epoch gradually increases, the loss value further converges to a stable value, verifying the significant advantages of our proposed method in terms of convergence speed and accuracy.

This fast convergence feature not only reduces the time cost of model training, but also effectively avoids the occurrence of overfitting problems, further proving the efficiency and robustness of our proposed method in handling complex modulation signal tasks. The experimental results have clearly demonstrated the superior performance in the model optimization process, especially in the early training stage. Its fast reaching of stability enables the model to enter the optimal performance state more quickly, laying a solid foundation for subsequent signal recognition tasks. This performance fully demonstrates the effectiveness of the method proposed in this paper and provides strong theoretical support for its promotion in practical applications.

In Table I, we report the number of parameters and recognition accuracy of six different deep learning models on the RML2016.10A dataset. The results indicate that although our proposed method has increased the number of parameters compared to Transformer, LSTM, ResNet18/34, and KAN models, it has shown significant performance in parameter optimization, with only two-thirds of the total number of parameters compared to ResNet18/34. This optimization not only reflects the efficiency of our method in structural design, but also provides potential solutions for applications with limited real-time processing and computing resources. In terms of recognition rate, the method proposed in this study performs well, with a recognition accuracy of up to 78.5%. This result is significantly better than ResNet18's recognition accuracy of 71.4%, and improves the recognition accuracy by 12.2% compared to the KAN model. Compared to LSTM, our proposed method has improved the AMC accuracy by 15.8%, which further confirms the outstanding performance and effectiveness of the proposed method in modulation recognition tasks. These results not only demonstrate the excellent performance of the proposed method in parameter control and AMC accuracy, but also validate the practical value and application potential in the field of AMC. Our research results indicate that this method has unique advantages and competitiveness in AMC, laying a solid foundation for future modulation recognition research and practical applications.

IV. CONCLUSION

This paper delves into an AMC method based on the fusion of CNN and attention mechanism. Through the effective integration of carefully designed network architecture and attention mechanism, the proposed deep learning method has shown significant performance improvement in AMC task. Compared with existing methods, the recognition accuracy of the proposed method can reach 80.18% when SNR > -2dB, and the average classification accuracy can even reach up to 92.18% at SNR of 18 dB. Compared with other algorithms, the performance improvement is as high as 12.2%. This result fully demonstrates the effectiveness and superiority of the proposed method in AMC task.

In the practical wireless communication application, some challenges and limitations have been found. Especially in environments with extremely low SNRs, the erosion effect of noise on radio time-frequency features becomes particularly significant. This erosion effect not only damages the integrity of the signal, but also seriously affects the expression of its time-domain characteristics, making the classifier face huge difficulties in feature learning and decision-making process, and the misjudgment rate sharply increases. It is particularly important to develop more advanced denoising methods in response to this issue. In the field of AMC, the denoising technology has always been a hot and difficult research topic. With the continuous deepening of research, we strive to explore more efficient and robust denoising methods. These methods not only need to be able to effectively remove noise and retain useful information of the signal, but also adapt to different communication environments and changes in noise distributions, thereby demonstrating stronger generalization ability and robustness in practical applications.

The AMC method based on CNN and attention mechanism proposed in this paper has shown significant performance improvement. However, in the communication environments with extremely low SNRs, the erosion effect of noise on the features of radio time-frequency images is still a problem that needs to be solved. In the future, we will continue to conduct in-depth research on denoising techniques and optimization methods to address the complex and ever-changing wireless environment and noise challenges, further improving the performance and practicality of the AMC methods.

REFERENCES

[1] F Zhang, C Luo, J Xu, et al., "Deep learning based automatic modulation recognition: Models, datasets, and challenges," Digital Signal Processing, vol. 129, no. 103650, 2022.

[2] Q. Zheng et al., "A real-time constellation image classification method of wireless communication signals based on the lightweight network MobileViT," Cognitive Neurodynamics, vol. 18, pp. 650-671, 2024.

[3] R. Zhou, F. Liu, C. W. Gravelle, "Deep learning for modulation recognition: A survey with a demonstration," IEEE Access, vol. 8, pp. 67366-67376, 2020.

[4] Q. Zheng et al., "MobileRaT: A lightweight radio transformer method for automatic modulation classification in drone communication systems," Drones, vol. 7, no. 10, 596, 2023.

[5] Q. Zheng et al., "DL-PR: Generalized automatic modulation classification method based on deep learning with priori regularization," Engineering Applications of Artificial Intelligence, vol. 122, 106082, 2023.

[6] Q. Zheng et al., "A real-time transformer discharge pattern recognition method based on CNN-LSTM driven by few-shot learning," Electric Power Systems Research, vol. 219, 109241, 2023.

[7] S. Saponara, A. Elhanashi, and Q. Zheng, "Recreating fingerprint images by convolutional neural network autoencoder architecture," IEEE Access, vol. 9, pp. 147888-147899, 2021.

[8] X. Zhu et al., "Explainable prediction of loan default based on machine learning models," Data Science and Management, vol. 6, no. 3, pp. 123-133, 2023.

[9] T. Zhang, C. Shuai, and Y. Zhou. "Deep learning for robust automatic modulation recognition method for IoT applications," IEEE Access, vol. 8, pp. 117689-117697, 2020.

[10] P. Cai, C. Zhang, and J. Chai, "Forecasting hourly PM2. 5 concentrations based on decomposition-ensemble-reconstruction framework incorporating deep learning algorithms," Data Science and Management, vol. 6, no. 1, pp. 46-54, 2023.

[11] W. Xiao, Z. Luo, and Q. Hu, "A review of research on signal modulation recognition based on deep learning," Electronics, vol. 11, no. 17, 2764, 2022.

[12] M. Hu et al., "ETA-rPPGNet: Effective time-domain attention network for remote heart rate measurement," IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-12, 2021.

[13] Q. Zheng et al., "Application of wavelet-packet transform driven deep learning method in PM2. 5 concentration prediction: A case study of Qingdao, China," Sustainable Cities and Society, vol. 92, 104486, 2023.

[14] M. Jasim et al., "A survey on spectrum management for unmanned aerial vehicles (UAVs)," IEEE Access, vol. 10, pp. 11443-11499, 2021.

[15] S. Peng, S. Sun, and Y. Yao, "A survey of modulation classification using deep learning: Signal representation and data preprocessing," IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 12, pp. 7020-7038, 2021.

[16] A. Vaswani et al., "Attention is all you need," Advances in Neural Information Processing Systems (NIPS), 2017.

[17] M. Pan et al., "Radar HRRP target recognition model based on a stacked CNN–Bi-RNN with attention mechanism," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-14, 2021.

[18] Q. Zheng et al., "CLMIP: cross-layer manifold invariance based pruning method of deep convolutional neural network for real-time road type recognition," Multidimensional Systems and Signal Processing, vol. 32, no. 1, pp. 239-262, 2021.

[19] Z. Zhang et al., "Automatic modulation classification using CNN-LSTM based dual-stream structure," IEEE Transactions on Vehicular Technology, vol. 69, no. 11, pp. 13521-13531, 2020.

[20] Q. Zheng et al., "Fine-grained image classification based on the combination of artificial features and deep convolutional activation features," IEEE/CIC International Conference on Communications in China (ICCC), Qingdao, China, 2017.

[21] A. Pandey, D. Wang, "Dense CNN with self-attention for time-domain speech enhancement," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, pp. 1270-1279, 2021.

[22] Q. Zheng et al., "Robust automatic modulation classification using asymmetric trilinear attention net with noisy activation function," Engineering Applications of Artificial Intelligence, vol. 141, 109861, 2025.

[23] T. J. O'Shea, J. Corgan, T. Clancy, "Convolutional radio modulation recognition networks," Engineering Applications of Neural Networks: 17th International Conference, Aberdeen, UK, pp. 213-226, 2016.

[24] Q. Zheng et al., "Static hand gesture recognition based on Gaussian mixture model and partial differential equation," IAENG International Journal of Computer Science, vol. 45, no. 4, pp. 569-583,2018.

[25] X. Zhuang, Y. Ying, and A Chen., "A combined forecasting method for intermittent demand using the automotive aftermarket data," Data Science and Management, vol. 5, no. 2, pp. 43-56, 2022.

[26] Y. Wang et al., "LightAMC: Lightweight automatic modulation classification via deep learning and compressive sensing," IEEE Transactions on Vehicular Technology, vol. 69, no. 3, pp. 3491-3495, 2020.

[27] S. Zhong et al., "An autoencoder framework with attention mechanism for cross-domain recommendation," IEEE Transactions on Cybernetics, vol. 52, no. 6, pp. 5229-5241, 2020.

[28] Q. Zheng et al., "A bilinear multi-scale convolutional neural network for fine-grained object classification," IAENG International Journal of Computer Science, vol. 45, no. 2, pp. 340-352, 2018.

[29] N. Jiang et al., "A municipal PM2. 5 forecasting method based on random forest and WRF model," Engineering Letters, vol. 28, no. 2, pp. 312-321, 2020.

[30] Q. Zheng et al., "Reconstruction error based implicit regularization method and its engineering application to lung cancer diagnosis," Engineering Applications of Artificial Intelligence, vol. 139, 109439, 2025.

[31] Y. Wang et al., "LightAMC: Lightweight automatic modulation classification via deep learning and compressive sensing," IEEE Transactions on Vehicular Technology, vol. 69, no. 3, pp. 3491-3495, 2020.

[32] P. Qi et al., "Automatic modulation classification based on deep residual networks with multimodal information," IEEE Transactions on Cognitive Communications and Networking, vol. 7, no. 1, pp. 21-33, 2020.

[33] Q. Zheng et al., "Rethinking the role of activation functions in deep convolutional neural networks for image classification," Engineering Letters, vol. 28, no. 1, pp. 80-92, 2020.

[34] Q. Zheng et al., "Spectrum interference-based two-level data augmentation method in deep learning for automatic modulation classification," Neural Computing and Applications, vol. 33, no. 13, pp. 7723-7745, 2021.

[35] Y. Kumar et al., "Automatic modulation classification based on constellation density using deep learning," IEEE Communications Letters, vol. 24, no. 6, pp. 1275-1278, 2020.
[36] X. Huang et al., "Feature extraction of search product based on multi-feature fusion-oriented to Chinese online reviews," Data Science and Management, vol. 5, no. 2, pp. 57-65, 2022.
[37] A. Pathy et al., "Design and implementation of blind modulation classification for asynchronous MIMO-OFDM system," IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-11, 2021.
[38] Y. Wang et al., "Multi-task learning for generalized automatic modulation classification under non-Gaussian noise with varying SNR conditions," IEEE Transactions on Wireless Communications, vol. 20, no. 6, pp. 3587-3596, 2021.
[39] D. Pierpaolo et al., "Overview on intrusion detection systems design exploiting machine learning for networking cybersecurity," Applied Sciences, vol. 13, no. 13, 7507, 2023.
[40] J. Cai et al., "Semi-supervised radar intra-pulse signal modulation classification with virtual adversarial training," IEEE Internet of Things Journal, vol. 11, no. 6, pp. 9929-9940, 2023.
[41] N. Pathak and V. Bajaj, "Automatic modulation classification using bimodal parallel multichannel deep learning framework for spatial multiplexing MIMO system," Physical Communication, vol. 59, 102071, 2023.
[42] Xu J et al., "A spatiotemporal multi-channel learning framework for automatic modulation recognition," IEEE Wireless Communications Letters, vol. 9, no. 10, pp. 1629-1632, 2020.
[43] X. Tian, Q. Zheng, and N. Jiang, "An abnormal behavior detection method leveraging multi-modal data fusion and deep mining," IAENG International Journal of Applied Mathematics, vol. 51, no. 1, pp. 92-99, 2021.
[44] Y. Lin et al., "Adversarial attacks in modulation recognition with convolutional neural networks," IEEE Transactions on Reliability, vol. 70, no. 1, pp. 389-401, 2020.

Xuelin Chen is pursuing his bachelor's degree at Shandong Management University. His research interests include machine learning, deep learning, and pattern recognition.

Xinyu Tian is a Lecturer of the Shandong Management University. She received the B.E. degree from Shandong Jiaotong University in 2014. He received the M.E. degree from Shandong University in 2018. Her research interests include signal processing, pattern classification, computer vision, and intelligent computing.

Qinghe Zheng is an Associate Professor of the Shandong Management University. He received the B.E. degree from Xi'an University of Posts and Telecommunications in 2014. He received the M.E. and Ph.D. degrees from Shandong University in 2018 and 2022, respectively. He has co-authored about 50 peer reviewed scientific journal articles and holds 11 patents. He is a member of IEEE and Guest Editors of several journals, including Sensors, Symmetry, and Electronics. He is also TPC members of some international academic conferences, including ICAML 2020, ICBTA 2020, BDMIP 2020, IVPAI 2020, ICCAES 2021. His research interests include image processing, pattern recognition, deep learning, and edge computing.

Weiguang Wang is an Associate Professor of the Shandong Management University. He has presided one special research project on social science planning in Shandong Province, one scientific research plan project for higher education institutions in Shandong Province, one philosophy and social science planning project in Jinan, and participated in two projects of the Shandong Province Development Plan of Safety Production Science and Technology. He has published more than 20 papers, and obtained 4 utility model patents and 5 software copyrights. His research interests include information security and intelligent algorithms.