Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 18, No. 2, 2025

Automatic Modulation Recognition Based on Deep Learning

Xiaoting Wang

Southwest Minzu University, Chengdu 610000, China

Abstract: Automatic modulation recognition is an important aspect of wireless communication. In recent years, the rapid development of deep learning technology has provided new solutions for modulation recognition. Deep learning-based modulation recognition has strong feature extraction and classification capabilities, resulting in higher recognition accuracy compared to traditional detection methods. However, the commonly used neural networks currently all face the problem of low recognition accuracy under low signal-to-noise ratio (SNR). To address this issue, this paper proposes a hybrid neural network model based on multi-channel input, which utilizes three channels of input: I/Q signals, time-frequency (T-F) distribution matrix, and signal-to-noise ratio. By incorporating SNR to introduce an environmental perception mechanism and a gated fusion module, the model's ability to understand the features of complex sequence information is enhanced. In addition, a phased block based on the Transformer architecture is employed to learn local and global features by combining tokens of different scales. Experimental results on the open-source dataset RML2016.10A indicate that the proposed method outperforms the current state-of-the-art modulation recognition methods, with an average accuracy of 47.60% at signal-to-noise ratios from -20dB to 0dB, and an overall average recognition accuracy of 68.52%.

Keywords: Automatic Modulation Recognition; Gated Fusion, Multi-channel fusion; Mixed Neural Network.

1. Introduction

Modulation recognition can provide key information for real-time response and decision-making in electronic warfare in the military field, and can be applied to spectrum monitoring, interference identification, and situational awareness in the civilian field. Accurate modulation recognition enables systems to better adapt to dynamic changes in the spectrum, enhancing the intelligence and flexibility of communication systems. Therefore, how to achieve automatic modulation recognition has become an important topic in current research. Deep learning-based [1] automatic modulation recognition (DL-AMR) avoids the errors brought by feature selection due to its end-to-end learning capability, compared to traditional methods [2,3] that rely on manually extracting feature parameters, and it shows stronger generalization ability and adaptability in complex environments [4].

Due to O'Shea et al. [5,7] demonstrating the superior performance of Convolutional Neural Networks (CNN) on complex-valued time-domain radio signals, several DL-AMR methods have emerged that use raw signals as input, paired with CNNs or Recurrent Neural Networks (RNN)[8,10].However, CNNs lack temporal sensitivity and exhibit local inductive bias, while RNNs lack spatial sensitivity and require more computational resources, leading to performance limitations. Subsequent research has mainly focused on improvements based on networks and inputs.

Currently, the Transformer model is regarded as the preferred model for natural language processing tasks [11]. Yin Zhan et al. introduced attention mechanisms in modulation recognition, effectively extracting feature information from I/Q sequences under low signal-to-noise ratios, achieving a 50% recognition accuracy when the SNR is greater than -5 dB. However, due to the Transformer network's focus on long-range dependencies [12], it has limited capability in perceiving local information in modulation recognition tasks, making it challenging to achieve good AMR performance when applied alone. Literature [13,15] fully exploits the advantages of the Transformer model by combining it with CNN to construct variant models that enhance recognition accuracy. Yang Jingya et al. [16] implemented a more lightweight modulation recognition mechanism under the CNN-Transformer model. Niu Ruiting et al. [17] effectively improved the accuracy of high-noise signals using lightweight neural networks while maintaining the same level of accuracy.

In summary, the above model still has limitations in low SNR environments. To address this issue, this paper proposes a hybrid neural network model based on multi-channel input, primarily processing I/Q signals, T-F distribution matrices, and SNR as three input channels. After gated fusion, these inputs are sent to a Bi-directional LSTM (Bi-LSTM) [18] for time series modeling, fully utilizing the potential relationships between multi-channel features. Finally, through a stage-wise block based on the Transformer architecture, the network learns to combine local and global features of different scales, enabling it to fully acquire rich semantic representations. The contributions of this study can be summarized as follows:

A multimodal hybrid network framework for AMR tasks based on I/Q signals, T-F distribution matrices, and SNR multi-channel inputs is proposed, and the effectiveness of multimodal inputs in identifying modulation types is validated;

Designed a gate fusion module based on SNR environmental awareness, dynamically allocating weight information to effectively fuse three-channel features based on the environmental information provided by SNR data.

A stage-wise block based on the Transformer architecture has been proposed to effectively extract fused local and global features.

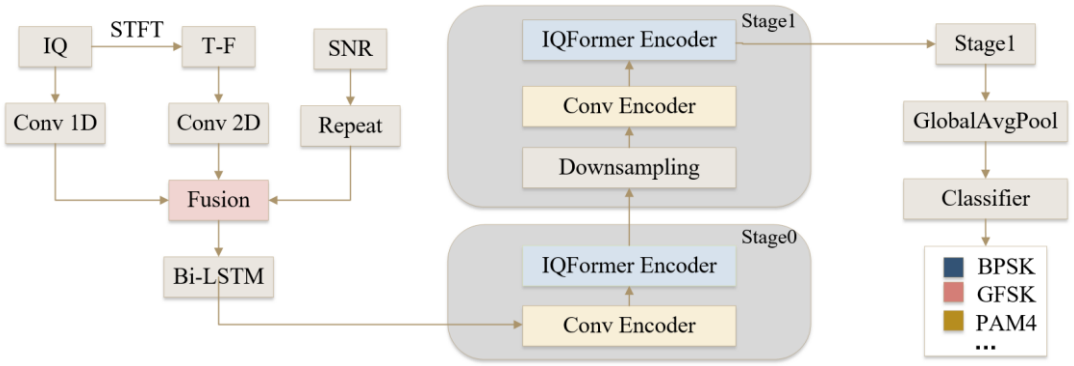

Fig.1 MC-IQFormer Model Structure

2. Model framework

The multi-channel hybrid neural network model structure is shown in Fig.1, consisting of four parts: input preprocessing, gated fusion, Bi-LSTM temporal modeling, and staged feature extraction. The input preprocessing module is responsible for processing the information from the I/Q signal, T-F distribution matrix, and SNR three channels. The gated fusion module is responsible for dynamically allocating weights based on SNR information to merge features. Then, the Bi-LSTM aligns the time-varying trends and captures the temporal dependencies of the signals. Finally, local and global feature extraction and fusion are performed through staged blocks, followed by classification output.

(1) Input preprocessing

The essence of modulation is to load information into the amplitude or phase of the signal waveform. There may be issues with classification errors in the model due to imbalances in amplitude and phase in the I/Q channels. Therefore, the proposed model considers using multiple data to construct a multi-channel data parallel extraction structure, capturing the spatiotemporal correlation of the signal more comprehensively.

Previous studies [19] have shown that both I/Q signal data and T-F distribution matrix data are effective methods for mapping signal features. Furthermore, the SNR channel provides direct information about signal quality to the model, enabling it to adjust feature extraction and classification strategies based on the signal-to-noise ratio. This contextual information enhances the model's adaptability to different noise environments. In the proposed MC-IQFormer model, I/Q signal data and SNR information are used as inputs. The input I/Q signal data is transformed into T-F distribution matrix data using short-time Fourier transform, followed by one-dimensional and two-dimensional convolutions. The dimensionally expanded SNR data is then combined in a fusion module. This enhances the model's capability to understand complex sequential information, thereby improving the model's modulation recognition performance.

(2) Gated fusion

Compared to simple feature concatenation, the gating mechanism offers more efficient feature utilization, capable of adaptively handling signal features under different SNR conditions, making full use of the complementarity between time-domain and frequency-domain information, and enhancing the model's generalization ability in various scenarios.

Through the gating fusion mechanism, the SNR channel helps the model to integrate the features of the IQ and time-frequency distribution matrix more precisely. Additionally, the optimization of the fusion process improves overall performance. The specific design structure is shown in Fig.2, where the preprocessed I/Q signal data, time-frequency distribution matrix, and SNR data are first concatenated in channels, followed by fusion operations. It consists of two parts: the main fusion branch and the gating branch. The main fusion branch undergoes a convolution operation for feature transformation, using the GELU non-linear transformation to enhance expressive capacity, followed by a second convolution layer to further extract fused features. The gating branch runs in parallel with the main fusion branch, learning gating weights from the same concatenated features, using the Sigmoid activation function to ensure that the gating weights are within the range of [0,1] to control the flow of information, and finally performing weighted fusion.

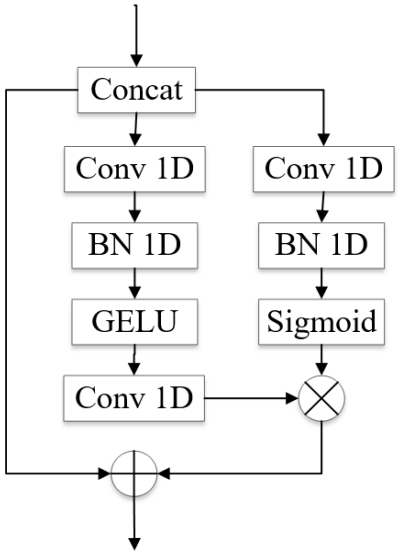

Fig.2 Block Diagram of Gate-Controlled Fusion Structure

Gate-controlled fusion performs multi-channel integration to achieve complementary information from three types of features. Compared to single-channel, it has a stronger feature expression capability. Additionally, residual connections are used to ensure that important information is not completely gated out, alleviating the gradient vanishing problem in deep networks, which helps the model converge to the optimal solution more quickly.

(3) Bi-LSTM Time Series Modeling

Wireless signals have obvious temporal characteristics, and Bi-LSTM can capture the dependencies of the signal at different time steps. Therefore, after gated fusion and before

the phased IQFormerEncoder, a Bi-LSTM module is designed to model the temporal characteristics of the fused features, capturing the temporal dependencies of the signal sequence.

Bi-LSTM performs feature dimension optimization, learning higher-level feature representations, while using a 2-layer stack to learn features at different levels of abstraction, providing stronger nonlinear expressiveness. Moreover, it captures local temporal dependencies and complements the subsequent staged IQFormerEncoder.

(4) Stage-wise feature extraction

From Fig.1, it can be seen that the model enters the Stage block for phased feature extraction after passing through the Bi-LSTM. Each Stage block is an independent block, which internally mixes the ConvEncoder for local feature extraction and the IQFormerEncoder for global feature modeling, achieving a progressive feature extraction that transitions from local features to global features.

The ConvEncoder focuses on capturing local feature space patterns, with the specific design structure shown in Fig.3. It uses depthwise separable convolution to capture the local textures and edge features of the signal, requiring fewer parameters and being computationally efficient, while also introducing residual connections to alleviate the gradient vanishing problem.

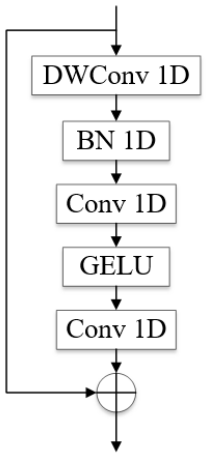

Fig.3 Conv Encoder Structure Diagram

The IQFormerEncoder is designed to establish a global feature representation and includes the LocalRepresentation module for local representations, a global attention module, and a feedforward network (FCN) module, as shown in Fig.4. The LocalRepresentation module first performs local feature extraction, then captures long-distance dependencies through an efficient additive attention mechanism for global feature extraction, and finally processes through the feedforward network. This implementation improves noise resistance by fusing features from different scales.

The attention mechanism calculates attention weights through queries and keys, allowing the model to focus on the correlations at different time steps of the input sequence. The specific principle is shown in Fig.5, where the input x is projected into queries and keys, calculates the clicks of queries with weight parameters, generates attention weights, and then performs a weighted sum with queries to generate global features, which are then combined with keys to produce the final output.

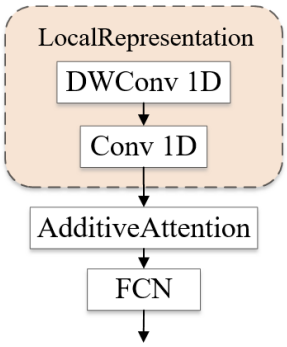

Fig.4 IQFormer Encoder Structure Diagram

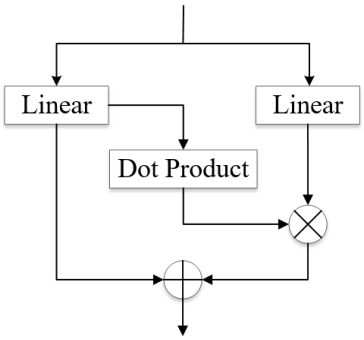

Fig.5 Attention Mechanism Structure Diagram

3. Experiment and Result Analysis

(1) Dataset

The RML2016.10A dataset is a widely used dataset for radio signal modulation recognition, particularly suitable for modulation classification tasks in wireless communication. The signals in the dataset are generated through simulation, mimicking various modulation methods and different communication environments. It includes 11 modulation modes such as 8PSK, BPSK, QAM16, QAM64, QPSK, WBFM, CPFSK, GFSK, AMDSB, AM-SSB, and PAM4, with a signal-to-noise ratio range of -20 to 18 dB, and a step size of 2 dB. For each modulation mode, the number of samples at each signal-to-noise ratio is 1000. It also simulates different channel environments, including AWGN, multipath fading, sample rate offset, and carrier frequency offset, similar to real-world conditions. These features make the dataset suitable for training and evaluating models in various noise environments.

(2) Implementation details

All experiments in this study were conducted in an environment based on the PyTorch-GPU deep learning framework, which has high efficiency and good compatibility. The experiments were performed on a server using a GeForce RTX 4090 GPU. The software environment utilized Python 3.10 and Anaconda as development tools, providing an efficient and stable operating environment for the experiments. The hyperparameter settings are shown in Table 1.

Table 1. Hyperparameter Settings

Batch size	256
Eval batch size	400
Learning rate	0.001
Early stop	10

(3) Model comparison

The performance of five SOTA DL-AMR models was compared, including PET-CGDNN [23], MCLDNN [24], AMC-NET [25], FEA-T [26], and IQFormer. The comparison primarily focused on the number of parameters and recognition accuracy, including the highest and average recognition accuracy under high, low, and overall signal-to-noise ratios, to assess the model's performance. Table 2 lists the experimental evaluation results on the RML2016.10A dataset, where the MC-IQFormer algorithm outperforms the compared SOTA algorithms in overall accuracy. From the table, it can be seen that the recognition accuracy of the MC-IQFormer algorithm is higher than that of the best results of the SOTA algorithms by 0.77%, 7.27%, and 4.33% for high, low, and overall signal-to-noise ratios, respectively.

Fig.6 describes the recognition accuracy of all models at different SNR levels in more detail. It can be seen that at low signal-to-noise ratios, the signal is almost drowned out by noise, and the performance of the five SOTA DL-AMR models is relatively consistent. However, the recognition accuracy of the MC-IQFormer model, which includes an SNR-aware mechanism, is significantly higher than that of the other models. As the SNR increases, feature learning becomes easier, and recognition accuracy gradually improves. When the SNR is greater than or equal to 0dB, the recognition accuracy of MC-IQFormer is basically consistent with that of IQFormer, both slightly higher than the other models.

Table 2. Comparison of Model Performances

Model	Parameters	Highest Accuracy	SNR (dB)		Overall	Inference Time (ms/sample)
			-20-0	0-18		
PET-CGDNN	0.07M	90.77%	36.21%	89.45%	60.38%	0.4099
MCLDNN	0.41M	92.86%	37.27%	91.68%	61.93%	0.7775
AMC-NET	0.47M	92.82%	38.56%	91.32%	62.40%	0.6861
FEA-T	0.17M	90.09%	37.31%	88.54%	60.55%	0.2716
IQFormer	0.35M	93.90%	40.33%	93.15%	64.19%	0.7114
MC-IQFormer	0.35M	94.54%	47.60%	93.92%	68.52%	0.5047

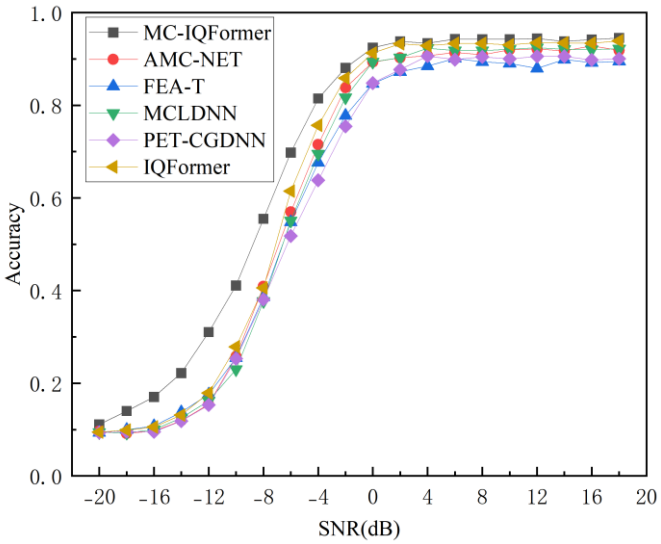

Fig.6 Comparison experiment

(4) Melting experiment

In this section, we conduct ablation experiments to evaluate the contribution of the proposed modules and input features to the model.

First, ablation experiments were conducted on the multi-channel input by removing the I/Q signal data input, the T-F distribution matrix, and the SNR environmental information respectively. Table 3 shows the results of the ablation experiments. It can be seen that when the I/Q signal is removed, the accuracy of the model decreases the most, and removing the T-F distribution matrix and SNR also leads to a decrease in model accuracy, demonstrating the effectiveness of multi-channel input and the contribution of different input features to the model. Fig.7 details the recognition accuracy of different ablation methods under different SNRs.

Table 3. Multichannel Ablation Results

Model	Highest Accuracy	SNR (dB)		Overall
		-20-0	0-18	
w/o I/Q	91.54%	36.77%	88.32%	60.26%
w/o T-F	94.50%	45.63%	93.63%	67.34%
w/o SNR	93.81%	41.35%	93.13%	64.75%
MC-IQFormer	94.54%	47.60%	93.92%	68.52%

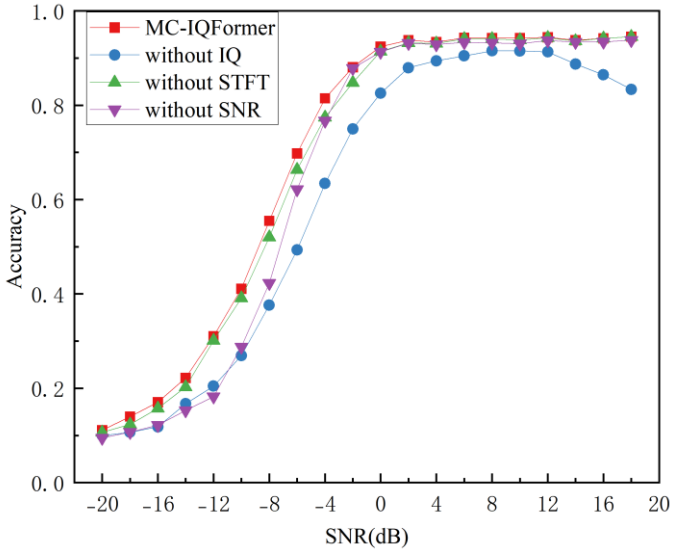

Fig.7 Multi-channel ablation

In addition, an ablation experiment was conducted on the representation of SNR in MC-IQFormer. When SNR is treated as a global variable, only I/Q signal data and T-F distribution matrix data are merged in the fusion part, with

SNR being dynamically assigned weights as global environmental information, without being integrated as a feature; when SNR is treated as a local feature, the fusion part integrates I/Q signal data, T-F distribution matrix, and SNR, adaptively assigning weights based on SNR. A comparison between treating SNR as a global variable and as a local feature is presented in Table 4, and Fig.8 provides a more detailed view of accuracy under different SNR conditions. It can be observed that SNR as a local feature has slightly higher accuracy than SNR as a global vector.

Table 4. Ablation results of SNR representation

Model	Highest Accuracy	SNR (dB)		Overall
		-20-0	0-18	
SNR Local	94.54%	47.60%	93.92%	68.52%
SNR Global	92.95%	46.96%	92.66%	67.56%

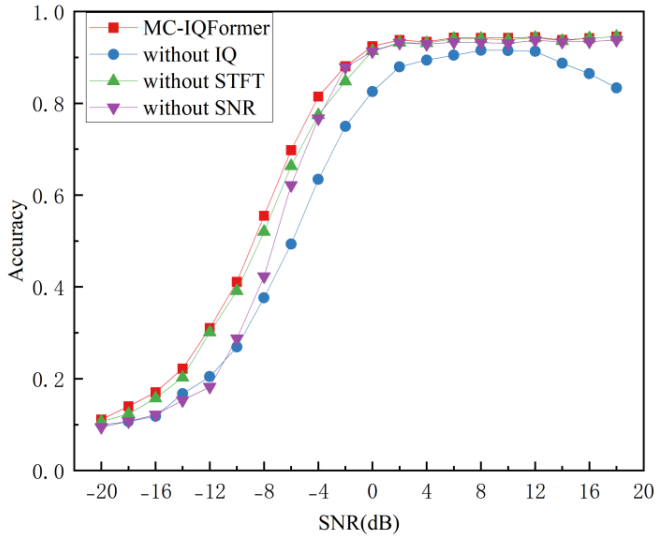

Fig.8 Comparison and Ablation of SNR Representation Methods

Finally, ablation experiments were conducted on the remaining modules, specifically removing the gated fusion module, the Bi-LSTM module, and the Stage module. Table 5 lists the ablation results, demonstrating that the removal of these three modules led to a decline in recognition accuracy for high, low, and overall SNR compared to MC-IQFormer, proving the effectiveness of the model.

Table 5. Ablation Results 3

Model	Highest Accuracy	SNR (dB)		Overall
		-20-0	0-18	
w/o MK	94.36%	46.24%	93.82%	67.72%
w/o Bi-LSTM	94.13%	46.08%	92.80%	67.21%
w/o Stage	93.72%	46.26%	92.41%	67.08%
MC-IQFormer	94.54%	47.60%	93.92%	68.52%

4. Conclusion

This article addresses the problem of low recognition accuracy under low signal-to-noise ratios (SNR) and improves the IQFormer model by proposing a multi-channel hybrid neural network model. The proposed MC-IQFormer incorporates an SNR environmental awareness mechanism, gaining additional contextual information through the inclusion of SNR channel inputs, which helps the model better understand the input data and perform dynamic weight allocation. Experimental results on the public dataset RML2016.10A demonstrate that the proposed MC-IQFormer method can achieve higher recognition accuracy under low SNR compared to state-of-the-art (SOTA) methods.

Acknowledgements

The authors gratefully acknowledge the support of the 2025 Graduate Innovative Research Project of Southwest Minzu University.

References

[1] Weaver C,Cole C A,Krumland R B,et al.THE AUTOMATIC CLASSIFICATION OF MODULATION TYPES BY PATTERN RECOGNITION.[C]. 1969.

[2] Liedtke F F. Computer simulation of an automatic classification procedure for digitally modulated communication signals with unknown parameters[J]. Signal Processing, 1984, 6(4): 311-323.

[3] Azzouz E E, Nandi A K. Automatic identification of digital modulation types[J]. Signal Processing, 1995, 47(1): 55-69.

[4] Yin Zan, Wang Chaojie, Cheng Ziheng, et al. An automatic modulation recognition algorithm based on attention mechanism convolutional neural network model [J]. Journal of Electromagnetic Waves and Applications, 2023, 38(05): 773-779.

[5] T. Huynh-The, C.-H. Hua, Q.-V. Pham, and D.-S. Kim, “Mcnet: An efficient cnn architecture for robust automatic modulation classification,” IEEE Communications Letters, vol. 24, no. 4, pp. 811–815, 2020.

[6] A. P. Hermawan, R. R. Ginanjar, D.-S. Kim, and J.-M. Lee, “Cnn-based automatic modulation classification for beyond 5g communications,” IEEE Communications Letters, vol. 24, no. 5, pp. 1038–1041, 2020.

[7] Z. Chen, H. Cui, J. Xiang, K. Qiu, L. Huang, S. Zheng, S. Chen, Q. Xuan, and X. Yang, “Signet: A novel deep learning framework for radio signal classification,” IEEE Transactions on Cognitive Communi cations and Networking, vol. 8, no. 2, pp. 529–541, 2021.

[8] D. Hong, Z. Zhang, and X. Xu, “Automatic modulation classification using recurrent neural networks,” in 2017 3rd IEEE International Conference on Computer and Communications (ICCC), 2017, pp. 695– 700.

[9] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin, “Deep learning models for wireless signal classification with distributed low cost spectrum sensors,” IEEE Transactions on Cognitive Communica tions and Networking, vol. 4, no. 3, pp. 433–445, 2018.

[10] Z. Ke and H. Vikalo, “Real-time radio technology and modulation classification via an lstm auto-encoder,” IEEE Transactions on Wireless Communications, vol. 21, no. 1, pp. 370–382, 2021.

[11] Tang Bochuan, Palidan Tuerxun, Bai Jiexin, et al. Land cover classification method for remote sensing images combining CNN and Transformer. Microelectronics and Computer, 2024, 41(04): 64-73. DOI: 10.19304/J.ISSN1000-7180.2023.0240.

[12] Wen Q, Zhou T, Zhang C, et al. Transformers in Time Series: A Survey[C]//Thirty-Second International Joint Conference on Artificial Intelligence. 2023: 6778-6786.

[13] Kong W, Yang Q, Jiao X,et al. A Transformer-based CTDNN Structure for Automatic Modulation Recognition[C]//2021 7th

International Conference on Computer and Communications (ICCC). 2021: 159-163.

[14] Hamidi-Rad S, Jain S. MCformer: A Transformer Based Deep Neural Network for Automatic Modulation Classification[C]//2021 IEEE Global Communications Conference (GLOBECOM). 2021: 1-6.

[15] Su H, Fan X, Liu H. Robust and Efficient Modulation Recognition with Pyramid Signal Transformer[C]//GLOBECOM 2022 - 2022 IEEE Global Communications Conference. 2022: 1868-1874.

[16] Yang Jingya, Qi Yanli, Zhou Yiqing, et al. CNN-Transformer lightweight intelligent modulation recognition algorithm [J]. Journal of Xi'an University of Electronic Science and Technology, 2023, 50(03): 40-49. DOI: 10.19665/j.issn1001-2400.2023.03.004.

[17] Niu Ruiting, Yan Tianfeng, Gao Rui, et al. Modulation recognition based on deep learning TCNN-MobileNet under low signal-to-noise ratio [J]. Computer Engineering, 2024, 50(07): 204-215. DOI: 10.19678/j.issn.1000-3428.0068243.

[18] S. Shabanian, D. Arpit, A. Trischler, and Y. Bengio, " Variational bi-lstms," arXiv preprint arXiv:1711.05717, 2017.

[19] M. Shao, D. Li, S. Hong, J. Qi and H. Sun, "IQFormer: A Novel Transformer-Based Model With Multi-Modality Fusion for Automatic Modulation Recognition," in IEEE Transactions on Cognitive Communications and Networking, vol. 11, no. 3, pp. 1623-1634, June 2025, doi: 10.1109/TCCN.2024.3485118.

[20] Y. Guo, D. Zhong, H. Sun, Z. Jiang, L. Ye, Z. Deng, and H. Liu, " Semiamr: Semi-supervised automatic modulation recognition with corrected pseudo-label and consistency regularization," IEEE Transactions on Cognitive Communications and Networking, vol. 10, no. 1, pp. 107– 121, 2024.

[21] Wu Changcheng, Sun Xiaochuan, Yu Jike, et al. A Lightweight Modulated Signal Recognition Method Based on Enhanced Multi-Scale Feature Fusion [J/OL]. Telecommunications Technology, 1-10 [2025-08-07]. https://doi.org/10.20079/j.issn.1001-893x.240613002.

[22] Peng Yulin. Research on Perception Communication Performance Based on Intelligent Algorithms [D]. University of Electronic Science and Technology, 2024.

[23] F. Zhang, C. Luo, J. Xu, and Y. Luo, " An efficient deep learning model for automatic modulation recognition based on parameter estimation and transformation," IEEE Communications Letters, vol. 25, no. 10, pp. 3287–3290, 2021.

[24] J. Xu, C. Luo, G. Parr, and Y. Luo, " A spatiotemporal multi-channel learning framework for automatic modulation recognition," IEEE Wire-less Communications Letters, vol. 9, no. 10, pp. 1629– 1632, 2020.

[25] J. Zhang, T. Wang, Z. Feng, and S. Yang, " Amc-net: An effective network for automatic modulation classification, " in 2023 IEEE In-ternational Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023, pp. 1– 5.

[26] Y. Chen, B. Dong, C. Liu, W. Xiong, and S. Li, "Abandon locality: Frame wise embedding aided transformer for automatic modulation recognition," IEEE Communications Letters, vol. 27, no. 1, pp. 327– 331, 2023.