Asia Pacific Science Press

Journal of Advances in Engineering and Technology
ISSN(O):3005-7507
ISSN(P):3005-7493
Vol. 3 No.2 (2026)

Automatic Modulation Classification using Residual Connection and Bottle-neck Transformers with Trend-aware Self-Attention

Geng Wang[1,3]*, Luming Li[2], Xin Chen[1,3]

1. Beijing Institute of Satellite Information Engineering, Beijing, 100095, China
2. Beijing Institute of Spacecraft System Engineering, Beijing, 100094, China
3. State Key Laboratory of Space Information System and Integrated Application, Beijing, 100095, China

Corresponding author: Geng Wang, wg312wg@gmail.com

Abstract: With the increase of modern communication equipment, the demand for AMC (Automatic Modulation Classification) in communication systems is increasing. Since deep learning was introduced into AMC, people have been working on improving the recognition accuracy and robustness of AMC. To achieve this, the study proposes the design of a ResDBoTTA (Residual DSC Bottleneck Transformers with Trend-aware Attention) AMC Network for communication signal modulation pattern classification. The DSC (Depthwise Separable Convolution) used in the model can significantly reduce the model parameters. The introduced time trend-aware self-attention mechanism can eliminate the influence of abnormal noise. Finally, global deep convolution is applied to enhance recognition accuracy. To verify the classification performance of ResDBoTTA AMC Net, relevant simulation studies are carried out. Experimental results demonstrate that ResDBoTTA AMC Net achieves superior recognition performance compared to existing technologies.

Keywords: Automatic Modulation Classification; Deep Learning; Depthwise Separable Convolution; Bottleneck Transformers
Published: Apr 20, 2026
DOI: https://doi.org/10.62177/jaet.v3i2.1262

1. Introduction

1.1 Research Background and Importance

In modern wireless communication systems, Automatic Modulation Classification (AMC) is one of the core technologies for signal demodulation and identification. Widely used in spectrum monitoring[1], electronic countermeasures, MIMO systems[2], Internet of Things (IoT)[3], next-generation 6G networks[4] and other fields. The goal is to quickly and accurately identify the modulation mode of the signal by analyzing the time-frequency characteristics of the received signal, thus laying the foundation for sub-sequent signal demodulation and information extraction. However, as the communi-cation environment becomes increasingly complex (for example, due to multi-path fading and non-Gaussian noise interference[5]), traditional AMC methods that are based on manual feature extraction (for example, high-order statistics[6], constellation diagram analysis[7] and wavelet transform[8], etc.) are faced with problems relating to insufficient feature robustness and limited generalization capability. The performance is shown to be especially susceptible to significant degradation when the signal-to-noise ratio is at a low level[9].

In the past few years, deep learning has become a novel approach to the challenges of AMC. O'Shea et al.[10] pioneered the application of convolutional neural networks (CNN) to the AMC task, attaining an 89% accuracy rate for 12 modulation types. This study sub-stantiated the efficacy of CNN in facilitating local feature extraction. Subsequently, the re-searchers mitigated the gradient vanishing problem of deep networks and improved training stability by introducing residual connectivity (ResNet) for modulated signal iden-tification[11]. However, conventional CNN face challenges in modeling the long-range temporal dependence of signals and exhibit sensitivity to local noise. To address these limitations, Zhou F et al.[12] Zhang put forward a proposal for a CNN-LSTM network modulation identification scheme. This method combines the advantages of CNN and LSTM networks, using the periodic characteristics of modulated signals and employing a recurrent network to capture timing information. This approach significantly enhances the identification accuracy of signals in complex electromagnetic environments. Hamza et al.[13] designs a bidirectional long short-term memory (BiLSTM) model to classify signals with different digital modulations.

This paper proposes a novel network architecture that fuses residual connected depth-separable convolution with a trend-aware self-attention mechanism. This architec-ture is referred to as ResDBoTTA (Residual DSC Bottleneck Transformers with Trend-aware Attention) AMC Network. It utilizes residual-connected deep separable con-volution (DSC) [14] to enhance the practicality of deployment on industrial devices, and introduces a trend-aware self-attention mechanism and global deep convolutional feature reconstruction. The network gives us fresh perspectives on AMC systems that have excel-lent modulation recognition capabilities, strong stability and minimal computation. The major contributions of this work include the following:

A Depth separable convolution residual block (DSC residual block) is applied. This type of block addresses the gradient vanishing problem by employing jump connec-tion, and it reduces the computational complexity by integrating Depthwise Separa-ble Convolution (DSC) to facilitate efficient multi-scale feature extraction.

A time trend-aware multi-head self-attention mechanism (MHSA-TA) was intro-duced. The one-dimensional causal convolution was incorporated into the mul-ti-head self-attention (MHSA) module in BOT, thereby establishing a novel MHSA-TA module. This module employs local context information to discern the temporal trend of the signal, suppress noise interference on attention weights, and augment the model's noise-resistance.

Feature reconstruction using global deep convolution (GDConv) is proposed: instead of GPA, we use the GDConv module to spatially reconstruct the feature map through global deep convolution. This approach enables the adaptive learning of feature im-portance at varying locations, thereby enhancing the classification accuracy.

The proposed ResDBoTTA will be validated by evaluating its performance on the publicly available RadioML2016.10a and RadioML2018a dataset. The experimental results demonstrate that ResDBoTTA exhibits a high level of performance in these two datasets, achieving classification accuracies of 93.23% and 97.43% at a 12 dB signal-to-noise ratio. These results indicate that ResDBoTTA significantly surpasses the performance of existing methods. The ablation experiments further validate the advantages of the temporal trend-aware attention mechanism and GDConv.

2. Related Works

This section reviews early AMC methods and further reviews the recent research of AMC.

2.1 Traditional AMC methods

Most of the early AMR methods for electromagnetic signals are designed for Gaussi-an noise channels, and they are mainly based on the likelihood based (LB) modulation identification method. The LB algorithm optimizes the model identification accuracy ac-cording to the likelihood functions of different modulated signals (e.g., mixed likelihood ratio, average likelihood ratio[15], generalized likelihood ratio[16]), and it can be consid-ered that the LB algorithm ensures that the modulation identification results can get the optimal value under the Bayes' minimum misclassification cost criterion. It can be con-sidered that the LB algorithm ensures that the modulation identification result can get the optimal value under the Bayes minimum error cost criterion, but when there are more types of modulation signals, it directly leads to the increase of computational complexity of the algorithm, and it is sensitive to the mismatch of the model and the parameter devia-tion. With the advancement of technology, feature based (FB) modulation recognition methods have gradually become the

mainstream. FB algorithms have low computational complexity and can achieve better recognition results through reasonable design of feature extraction. Feature based recognition algorithms generally include feature extraction and classification network, and many researchers have studied various types of features for extraction, including high-er-order cumulative quantities (HOC)[17], cyclic smoothness features[18], approximate entropy[19], etc., and the feature extraction algorithms can be used in a variety of ways. However, in the case of identification and classification of mul-tiple types of modulated signals, the above methods generally suffer from the problems of cumbersome manual feature extraction and high computational complexity, and their recognition effect is poor in low signal-to-noise ratio environments.

2.2 Recent AMC Method Based on Deep Learning

After CNN and LSTM, the Transformer[20] model has been shown to further en-hance feature modeling capability by capturing long-range dependencies through the Self-Attention mechanism (Self-Attention). Srinivas A et al.[21] initially proposed BoTNet (Bottleneck Transformer), a fusion network of CNN and Attention. This network replaces the 3x3 convolution operation in the middle of the Bottleneck layer, as implemented in ResNet50, with Multi-Head Self-Attention. This modification serves to reduce the amount of parameters by reducing the number of feature map channels in the middle layer. S. Lin et al.[22] proposed a spectral CNN model built on the AMC Time-Frequency Attention mechanism to correlate the time and frequency domains of the signal. W. Zhang et al.[23] proposed an AMC model based on CNN and Spatial Self-Attention mechanism to en-hance the connection between channel and space. Liang Z et al.[24] used multi-head at-tention in combination with CNNS to facilitate modulation recognition, thereby capturing the relationship between IQ signals. M Qi et al.[25] increases the amplitude and phase characteristics of the signal while using the IQ signal. Y. Feng et al.[26] added the Gramian angular field as a third feature input in addition to the two IQ signals used for AMC. Zhang et al.[27] combine CNN with Transformer and performed a sliding window process to fuse the IQ information prior to signal input. R. Duan et al.[28] used multi-modal inputs such as constellation diagrams to improve accuracy at high SNRs, and used only IQ signals at low SNRs to ensure accuracy. Z. Luo et al.[29] utilized a combination of CNN, LSTM and transformer algorithms to achieve optimal outcomes by employing mul-timodal input signals. W. Kong et al.[30] used the Transformer for self-supervised pre-training, after fine-tuning using a small number of labeled signal samples to accom-plish AMC.

In general, CNN Self-Attention Mechanism models have achieved good performance for AMC. Nonetheless, these methodologies continue to exhibit deficiencies. BoTNet solves the problem of gradient vanishing during feature extraction in deep networks, but the model parameter scale remains a potential area for enhancement. Moreover, the tradition-al multi-head self-attention mechanism is sensitive to noise, and is prone to erroneous al-location of attention weights under outliers or local disturbances, thus reducing model robustness.

3. Signal Model and Proposed Model

3.1 Signal Model

In this part, the signal model for automatic modulation classification is shown. In wireless communication systems, the received signal is usually affected by multipath fading, noise interference, and nonlinear distortion. Assume that the received signal passes through an additive Gaussian white noise (AWGN) channel, its baseband equivalent signal model can be described as:

$$\mathbf{y}(t) = \mathbf{h}(t) * \mathbf{s}(t) + \mathbf{n}(t) \tag{1}$$

where s(t) is the transmitted modulated signal, h(t) is the channel impulse response, n(t) is the noise obeying the complex Gaussian distribution $\mathcal{CN}(0, \sigma^2)$, and σ^2 is the noise power.

For discrete sampled signals, assume a symbol rate of B. The sampling interval is $T_s = 1/B$ and the received signal can be expressed as:

$$\mathbf{y}[n] = \sum_{k=0}^{L-1} \mathbf{h}[k]\mathbf{s}[n-k] + \mathbf{n}[n] \tag{2}$$

where L is the channel memory length. The mathematical form of the modulated signal s[n] depends on the modulation

method. Take the example of common modulation types:

PSK (Phase Shift Keying):

$$\mathbf{s}[n] = e^{j(2\pi m/M+\phi_0)}, m \in \{0,1,\ldots,M-1\} \tag{3}$$

QAM (Quadrature Amplitude Modulation):

$$\mathbf{s}[n] = a_m + jb_m \tag{4}$$

(a_m, b_m) is the coordinates of the constellation points.

FSK (Frequency Shift Keying):

$$\mathbf{s}[n] = e^{j2\pi f_m n T_s} \tag{5}$$

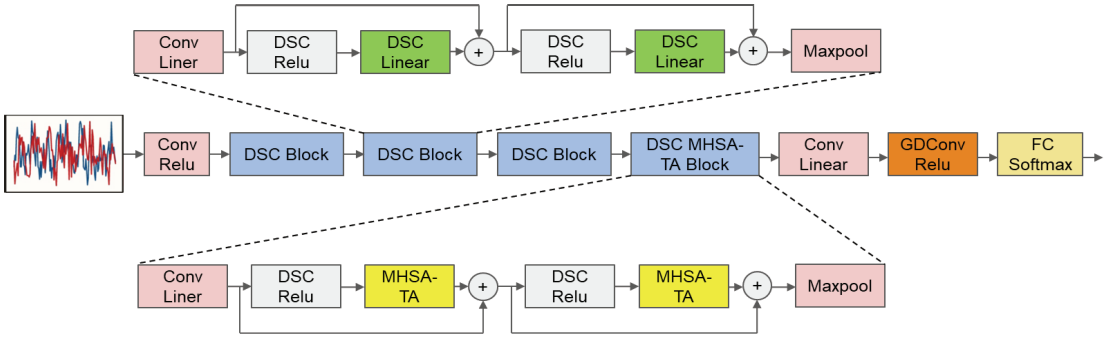
f_m is the discrete frequency point.

The datasets signal format used in the experiments is I/Q signals. The objective of modulation identification is to predict the kind of s[n] based on the received IQ signal.

3.2 ResDBOTTA model structure

The ResDBoTTA model proposed in this paper consists of the following three main modules: a feature extraction module using deep separable convolution with residual concatenation and a multi-head self-attention mechanism based on temporal trend sensing, a feature reconstruction module using a global deep convolution algorithm, and a classification module. The specific structure is shown in Figure 1.

Figure 1: The structure of ResDBOTTA model



3.3 Feature Extraction Layers

3.3.1 Residual Connected DSC Block

The jump connection method of residual network species can not only solve the problem of gradient vanishing, but also improve the network performance. The ResDBoTTA proposed in this paper designs a DSC residual block by introducing the jump connection method while using DSC instead of ordinary convolution. In order to extract the features of the signal, we first go through a linear convolution, then through two DSC residuals in parallel, and finally through a maximum pooling operation. Fig. 2 shows their structure.

The DSC operation process is comprised of two steps. Initially, the deep convolutional layer performs the convolutional operation on each channel of the input individually. Subsequently, the dot convolutional layer fuses all the channels. The operation of the deep convolutional layer is shown in Fig. 2. The deep convolution kernels are single-channel, and that the number of kernels is equal to the inputs. The height and width are still 5×5. The point convolution layer is characterized by a convolution kernel height and width of 1 standard convolution layer, a single convolution kernel size of 1×1×3, and a single point convolution kernel operation of the input and the output in the height and width of the dimension of the same, but the output into a single channel of the form of 8×8×1, as illustrated in Figure 2.

3.3.2 Time trend-aware multi-head self-attention mechanism

The conventional self-attention mechanism calculates the similarity between the query and the key by point-by-point (point-wise) without leveraging contextual information. In the event that the observation at a given time step is an outlier, it will incorrectly match the relevant points. This results in an inaccurate weight assignment for self-attention. As illustrated in the left panel of Fig. 3, point A is incorrectly matched with point B, rather than being matched with the more similar point C. In this study, we propose a temporal trend-aware attention mechanism (MHSA-TA) to extract the local context information of

each node so that each node has the ability to perceive the context, and then construct an attention matrix by capturing the local trend. The proposed MHSA-TA model contributes to more accurate predictions and is more stable than the traditional attention mechanism.

Figure 2: DSC residual block and DSC operation

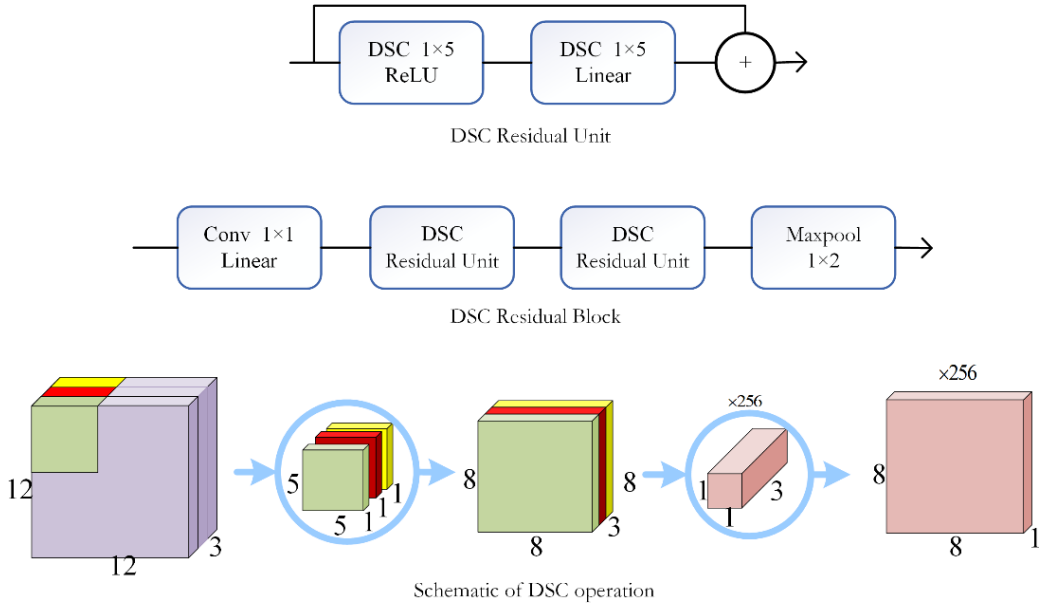
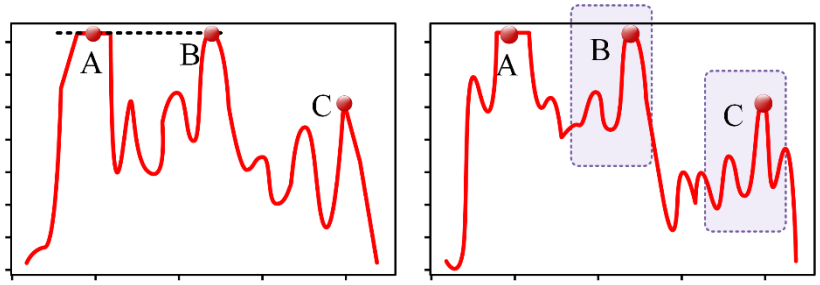


Figure 3: Conventional self-attention (left) and time-trend-aware attention (right)



The ResDBoTTA network employs a temporal trend-aware multi-head based self-attention mechanism module in the last residual block, replacing the conventional convolution operation. This modification enhances the modulation pattern recognition anti-noise capability, thereby functioning as a signal smoothing filter. The module has been shown to increase the robustness of the model in the face of noise and disturbances by reducing its sensitivity to outliers or extreme values. Furthermore, the module has been shown to reduce the mis allocation of weights and improve the generalization of the model over unknown data. In modulation recognition, time-trending self-attention ensures that the model maintains a high recognition accuracy even in the presence of poor signal quality.

MHSA-TA is a variant of MHSA that replaces the query and key projection operations in MHSA with 1D convolution, as shown in Fig. 4. The use of convolution here allows the contextual relationships of the time domain signals to be extracted as inputs to the MHSA, enabling our model to capture trends hidden in the time data series and to ignore anomalous changes. The definition of MHSA-TA is as follows.

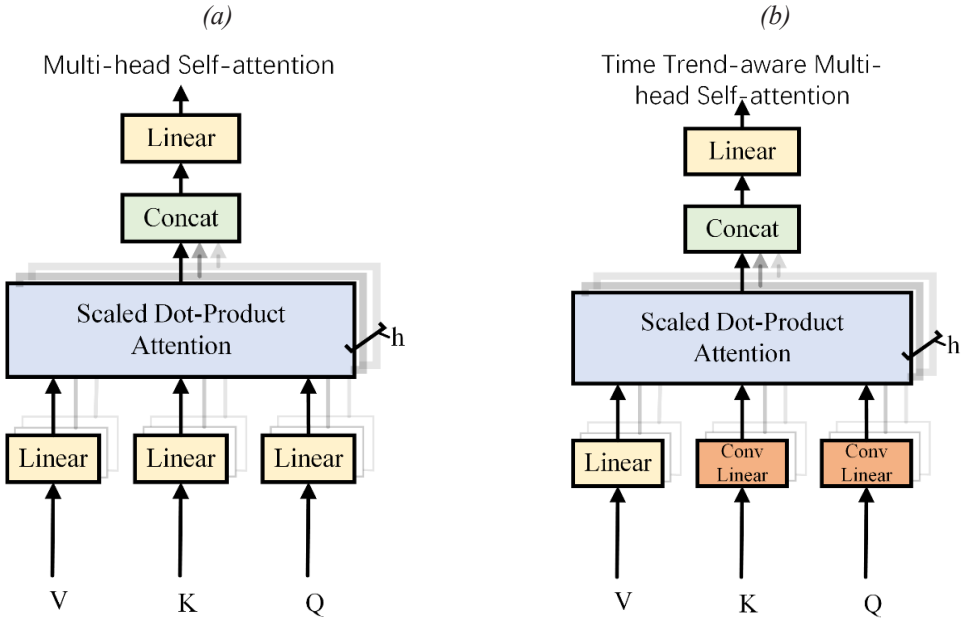
For an input $X \in \mathbb{R}^{N\times T\times C}$, where N represents the number of sequences, T denotes the sequence length, and C is the number of channels. The context relationship is captured by two 1D causal convolution of $1 \times k$. These layers generate the query Q and key K representations. Additionally, a separate 1×1 convolution layer is employed to generate the value V.

$$\begin{aligned} \mathbf{Q} \in \mathbb{R}^{N\times T\times C} &= \mathbf{Conv}_{1\times k}(\mathbf{X}) \\ \mathbf{K} \in \mathbb{R}^{N\times T\times C} &= \mathbf{Conv}_{1\times k}(\mathbf{X}) \\ \mathbf{V} \in \mathbb{R}^{N\times T\times C} &= \mathbf{Conv}_{1\times 1}(\mathbf{X}) \end{aligned} \tag{6}$$

The matrix of Q, K, V is reshaped by the reshape operation to obtain a representation in the form of multi-head attention:

$$\begin{aligned}\mathbf{Q} \in \mathbb{R}^{N\times h\times T\times d} &= \mathbf{reshape}(\mathbf{Q})\\ \mathbf{K} \in \mathbb{R}^{N\times h\times T\times d} &= \mathbf{reshape}(\mathbf{K})\\ \mathbf{V} \in \mathbb{R}^{N\times h\times T\times d} &= \mathbf{reshape}(\mathbf{V})\end{aligned} \tag{7}$$

Figure 4: The multi-head attention structure (a) and Time trend-aware multi-head self-attention structure (b)



Then, the attention matrix is computed in a localized contextual and normalized by softmax:

$$\mathbf{A} \in \mathbb{R}^{N\times h\times T\times T} = \mathbf{softmax}\left(\frac{\mathbf{QK}^{\boldsymbol{T}}}{\sqrt{d_k}}\right) \tag{8}$$

where h represents the number of attention heads, d represents the number of channels per attention head, and $C = h \times d$. Finally, V is weighted and summed and restored to the same shape as the input:

$$\begin{aligned}\mathbf{Y} \in \mathbb{R}^{N\times h\times T\times d} &= \mathbf{A}\cdot\mathbf{V}\\ \mathbf{Y} \in \mathbb{R}^{N\times T\times C} &= \mathbf{reshape}(\mathbf{Y})\end{aligned} \tag{9}$$

$\mathrm{Y} \in \mathbb{R}^{\mathrm{N}\times\mathrm{T}\times\mathrm{C}}$ is the output of the MHSA-TA.

3.4 Feature Reconstruction

After DSC operation, in the feature reconstruction part, ResDBoTTA proposes to utilize GDConv operation to learn the importance of features at different locations, after which the classification ability of reconstructed features is enhanced by ReLU activation function and bias term. The global deep convolutional layer is a deep convolutional layer with a convolutional kernel dimension equal to the size of the input feature map. The output of the global deep convolutional layer can be expressed as:

$$\widetilde{G}_m = \sigma_{\mathrm{ReLU}}\left(\sum_{i,j}\widetilde{K}_{i,j,m}\cdot\widetilde{F}_{i,j,m} + \tilde{b}_m\right) \tag{10}$$

In Eq. (10) $\widetilde{\mathrm{F}}$ is the feature map (Last Feature Map, LFM) with size $\mathrm{W}\times\mathrm{H}\times\mathrm{M}$; $\widetilde{\mathrm{K}}$ is a global deep convolution kernel with size $\mathrm{W}\times\mathrm{H}\times\mathrm{M}$; $\tilde{\mathrm{b}}$ is the bias term; $\widetilde{\mathrm{G}}$ is the output classification feature vector of size $1\times1\times\mathrm{M}$; (i,j) denotes the spatial location index; m denotes the channel index. Looking further, the size of the global deep convolution kernel $\widetilde{\mathrm{K}}$ is equal to the size of $\widetilde{\mathrm{F}}$ in the input LFM, The feature reconstruction operation occurs between the channels corresponding to $\widetilde{\mathrm{K}}$ and $\widetilde{\mathrm{F}}$. For a particular channel m, the values of $\widetilde{\mathrm{K}}_{\mathrm{m}}$ and $\widetilde{\mathrm{F}}_{\mathrm{m}}$ at the same position (i,j) are first multiplied, then all dot product results are summed plus a bias term $\tilde{\mathrm{b}}_{\mathrm{m}}$, and finally a classification feature is obtained by the output of the ReLU activation function.

Adding an FC layer activated by Softmax (normalized exponential function) after feature reconstruction transforms the predictions into probabilistic outputs.

3.5 ResDBOTTA model parameters

ResDBoTTA is implemented based on three parts: the DSC residual block, the MHSA module and the global deep convolutional feature reconstruction module. The specific configuration of the network is shown in Table 1. Four DSC residual blocks are deployed in the feature extraction part, the MHSA module is utilized in the fourth DSC residual block instead of the original convolution. The residual units are convolved and passed through the MHSA module before being summed up, and finally go through the maximum pooling. A layer of standard convolution is deployed at the front and back of the four DSC residual blocks. The first two DSC residual blocks are designed with 32 convolution kernels to decrease both the parameter count and computational load. And in order to enhance the feature dimension and extract sufficient information from the current input for feature extraction, the number of convolution kernels in the first standard convolutional layer and the last two DSC residual blocks are both designed to be 64, i.e., from 32 classes of features to 64 classes of features.

Table 1: ResDBoTTA parameter configuration

Network layer	Output dimension	function
Input	2×1024×1	--
First Conv 2×5 ReLU	1×1024×64	Feature Extraction
DSC Residual Block	1×512×32	
DSC Residual Block	1×256×32	
DSC Residual Block	1×128×64	Feature Extraction
DSC MHSA-TA Block	1×64×64	
Conv 1×1 Linear	1×16×64	
GDConv 1×16 ReLU	1×64	Feature Reconstruction
FC Softmax	24	Classifications

The loss function uses the categorical cross-entropy loss function categorical_crossentropy that works better with Softmax for classification tasks, and its definition can be expressed as:

$$Loss = -\sum_{i=1}^{output} y_i \cdot \log \hat{y}_i \tag{11}$$

From Eq. (11), we can see that the result is 0 when y_i is 0 and the correct result output exists when and only when y_i is 1. This loss function is mainly used for comparing the two probability distributions, and in the back side it is connected to Softmax to rescale the model output so that it has the correct attributes, and the network finally outputs the result through an FC layer activated by Softmax.

4. Experimental Results and Discussion

4.1 Dataset Description

The DeepSig RadioML2016.10A[31] and RadioML2018.01A[32] datasets were used in this paper, respectively. The DeepSig RadioML 2016.10A dataset contains 8 digital modulation methods and 3 analog modulation methods, with a total of 220,000 samples. The signal-to-noise ratio ranges from -20dB to 18dB with 2dB spacing. The data format is 2×128 IQ signals.

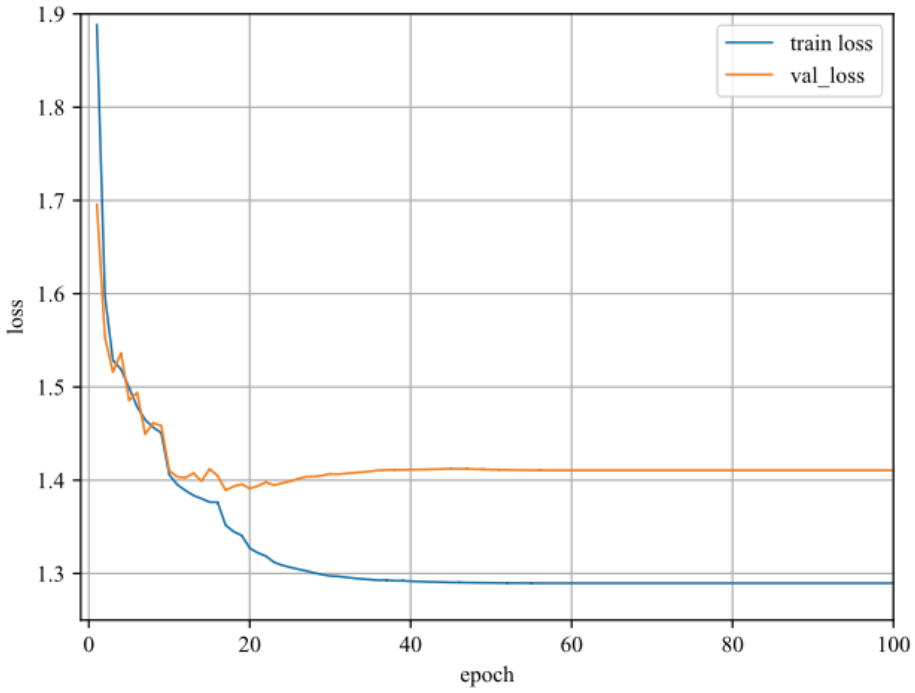
The DeepSig RadioML 2018.01A dataset contains a total of 24 signal modulation types, and there are 4096 data for each modulation type at each signal ratio in the range of signal ratios from -20dB to 30dB with a step value of 2dB, for a total of 2555904 data. The signal data is divided into I and Q and the data format is (1024,2). By sampling, 2064384 data points with signal-to-noise ratios from -20dB to 20dB are used, of which 1548288 are training samples and 516096 are test samples.

4.2 Experimental Settings

The ResDBoTTA network is trained with epoch of 100 and batch_size size of 512. Lazy Adam is used as the model optimizer during the training process instead of the commonly used SGD optimizer because in this structure the SGD optimizer tends to fall into local minima and has a slower rate of descent. Lazy Adam can handle sparse updates more efficiently. The ResDBoTTA network employs a categorical cross-entropy loss function, categorical_crossentropy, which is used to measure

the extent to which the model's predicted value f(x) differs from the true value y. It guides the direction of the next step of training by calculating the gap between the forward computed result and the true value in each iteration. The loss function of the ResDBoTTA network when training the RadioML 2018.01.A datasets is shown in Figure. 5.

Figure 5. Loss function for RadioML 2018.01.A datasets



4.3 Baseline Models

Table 2. Baseline models of RadioML 2016.10A datasets

Datasets	Model	Parameters
RadioML 2016.10A	CNN	0.85M
	ResNet	3.1M
	MCLDNN	0.41M
	1DCMMPF	0.17M
	CNN-atten	0.24M
	ResDBOTTA	0.17M

CNN[32] model: The classical CNN architecture proposed by O'Shea et al. contains 3 convolutional and 2 fully connected layers, with the input being a joint time-domain amplitude-phase representation of the I/Q signal. The model achieves 89% accuracy on 12 classes of modulated signals, but the timing dependence is not considered.

ResNet[11] model: A deep convolutional network with residual connections is used to mitigate the gradient vanishing problem by jump connections. The model has a classification accuracy of 93.2% (SNR = 20 dB) for 24 classes of signals on the RML2018 dataset.

MCLDNN[33]: The article proposes a spatio-temporal multi-channel learning framework. This model is able to receive different types of input data and process different spatio-temporal features through a multi-channel model.

CNN-atten[24]: A radio signal recognition method based on the combination of convolutional neural network (CNN) and Self-Attention Mechanism is proposed in this paper. The accuracy of recognition is enhanced in low signal-to-noise conditions.

1DCMMPF[34]: This article proposes an AMC model that combines the outputs of multiple CNN models to improve the accuracy and robustness of modulation recognition through parallel fusion.

The comparison baseline model of data set RadioML 2018.01A is shown in the Table 3:

Table 3. Baseline models of RadioML 2018.01A datasets

Datasets	Model	Parameters
RadioML 2018.01A	CNN	1.7M
	ResNet	21M
	MCNET	0.13M
	PETCGDNN	0.075M
	CNN-atten	0.25M
	ResDBOTTA	0.18M

MCNET[35]: The MCNET proposed in the article is an improved CNN modulation recognition model that can reduce the consumption of computational resources while ensuring high classification accuracy.

PETCGDNN[36]: The model proposed in the article combines graph convolutional network (GCN), temporal convolutional network (TCN) and pyramid structure for multilevel feature learning. The accuracy of recognition is improved and the number of parameters is reduced.

The same models as in the previous RadioML2016.10A experiment are CNN-atten, ResNet and CNN.

4.4 Comparative Experiments of Different Networks

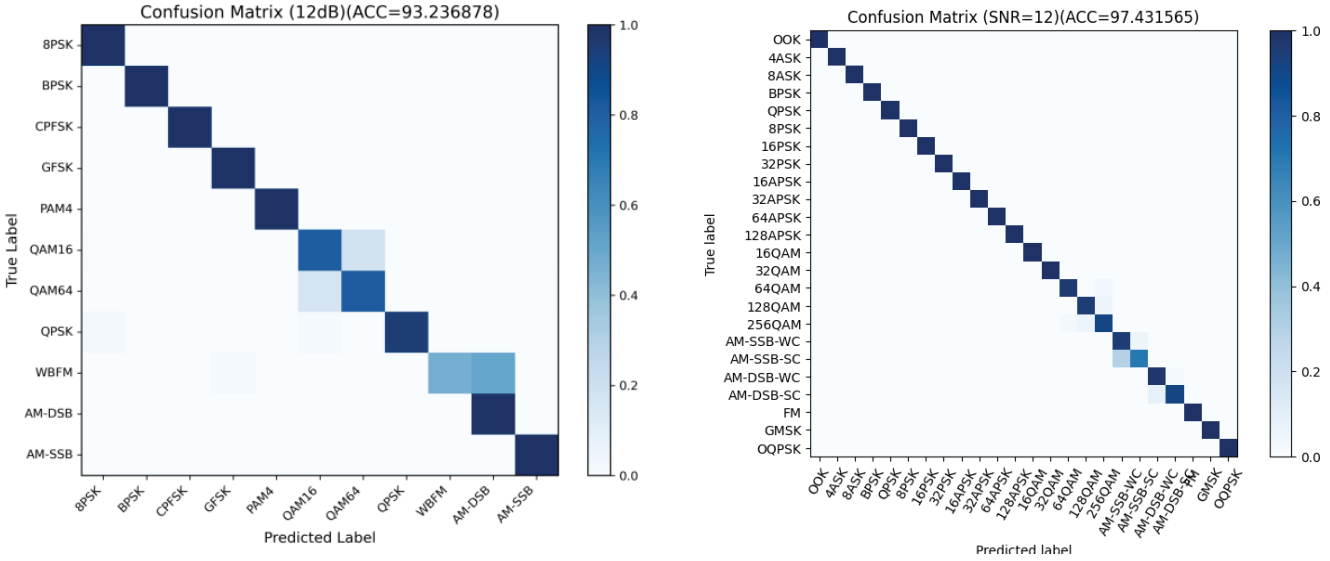
As shown in Figure 6, the confusion matrix shows the results of the proposed model in identifying the RML2016.10A (a) and RML2018.01A (b) datasets at a signal-to-noise ratio of 12dB. (c) and (d) respectively describe the recognition accuracy curves of the comparative test with the baseline model on the RML2016.10A datasets and the RML2018.01A datasets. The following sections (e) and (f) present the recognition accuracy of the ResDBoTTA for different modulation types on the RML2016.10A and RML2018.01A datasets, respectively. As shown in Figure 6(c), for the RML2016.10A dataset, when the signal-to-noise ratio is greater than 2 dB, the recognition accuracies of all the compared models reach more than 80%, among which MCLDNN, cnn-atten, and resdbot all exceed 90% accuracy.

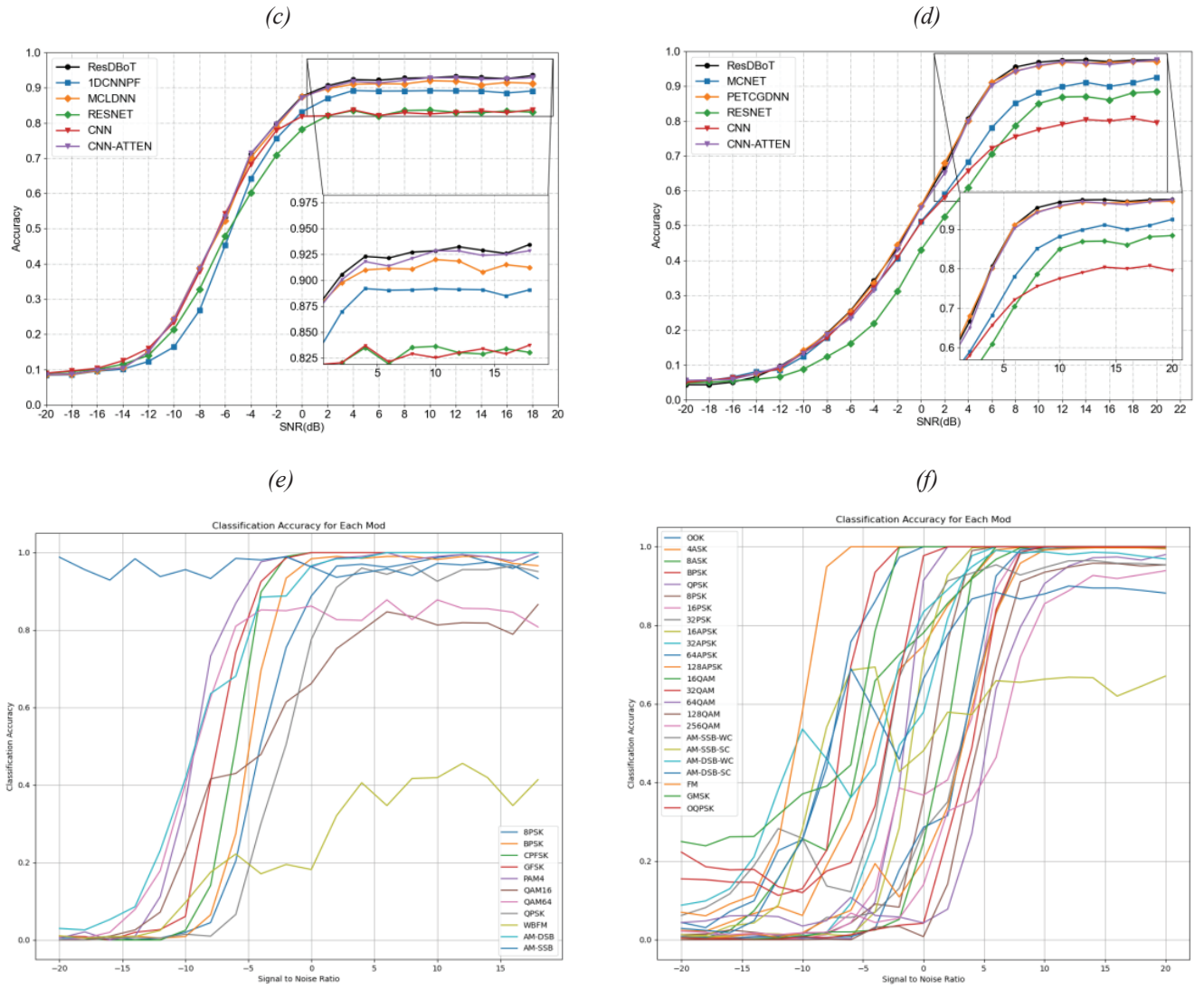
For the RML2018.01A dataset in Figure. 6(d), at 10 dB, all except CNN achieve over 85% accuracy. As demonstrated in Figure. 6(e) and Figure. 6(f), the proposed model demonstrates commendable recognition accuracy for various modulations, except for two modulations: WBFM and AM-SSB-SC.

As shown in Figure. 6(a) and Figure. 6(b), the ResDBoTTA network achieved better recognition results in the high signal-to-noise ratio interval, with 93.23% recognition accuracy for the RML2016.10A dataset and 97.43% for the RML2018.01A dataset under the condition of a signal-to-noise ratio of 12 dB. Under the condition of higher signal-to-noise ratio, the model proposed in this paper has higher recognition accuracy compared with other models.

Figure 6. Experiments results

(a) *(b)*





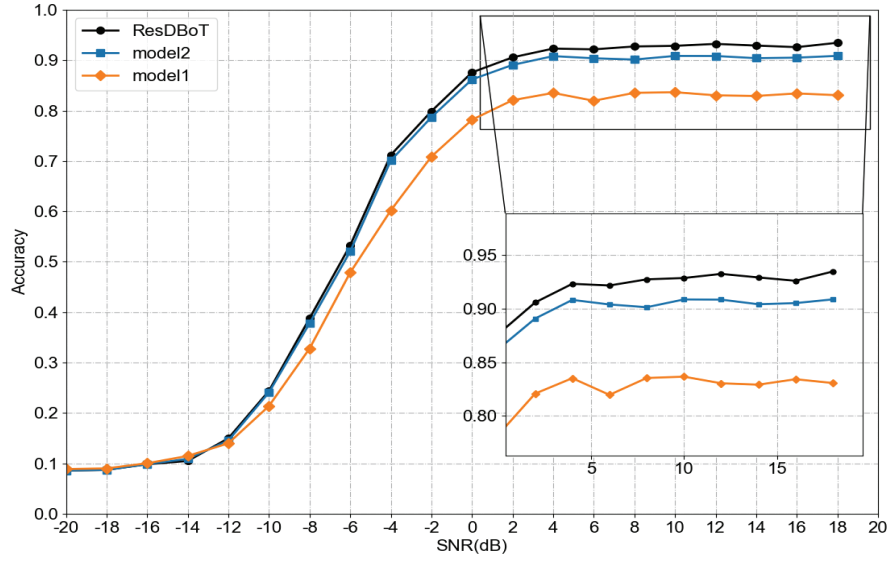
4.5 Ablation Experiment

An ablation study was conducted on the RadioML2018.01A dataset to test the performance of the module using two models:

1st model: Using the ResDBoTTA model without multi-head attention.

2nd model: Added multi-head attention to the 1st model, but no time trend-aware.

Figure 7. Recognition accuracy of the three models



Based on the results of the ablation experiments in Figure. 7, we can get the following conclusion. All models based on the DSC residual block structure have good recognition accuracy. The multi-head self-attention mechanism gives a great improvement in accuracy compared with 1st model and 2nd model. On this basis, the trend-aware self-attention mechanism can further improve the recognition accuracy in the case of high signal-to-noise ratio.

Conclusion

In this study, a new ResDBoTTA net is proposed to recognize the modulation of communication signals. The proposed ResDBoTTA model consists of three parts: depth-separable convolutional residual block to achieve multi-scale feature extraction, time trend-aware self-attention mechanism, and feature reconstruction using global depth convolution. The self-attention mechanism using time trend awareness helps to substantially improve the modulation pattern recognition accuracy of the ResDBoTTA model. In order to examine the modulation pattern recognition capability of ResDBoTTA net, a detailed experimental analysis is conducted and the results are examined from several aspects. The experimental results show that the modulation pattern recognition accuracy of ResDBoTTA net is improved compared to the existing techniques. Further research can be done in the future to reduce the model complexity while ensuring the accuracy.

Funding

No

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

[1] Mohsen, S., Ali, A. M., & Emam, A. (2023). Automatic modulation recognition using CNN deep learning models. Multimedia Tools and Applications, 83(3), 7035–7056. https://doi.org/10.1007/s11042-023-15814-y

[2] Das, D., Bora, P. K., & Bhattacharjee, R. (2021). Automatic modulation classification over MIMO amplify and forward (AF)-relay fading channels. Physical Communication, 47, 101399. https://doi.org/10.1016/j.phycom.2021.101399

[3] Chen, Q., Meng, W., Han, S., Li, C., & Chen, H.-H. (2022). Robust task scheduling for delay-aware IoT applications in civil aircraft-augmented SAGIN. IEEE Transactions on Communications, 70(8), 5368–5385. https://doi.org/10.1109/tcomm.2022.3186997

[4] Moulay, H., Djebb ar, A. B., Dehri, B., & Besseghier, M. (2024). Dendrogram-based heterogeneous learners for automatic modulation classification in DSTBC-OFDM systems. Physical Communication, 62, 102241. https://doi.org/10.1016/j.phycom.2023.102241

[5] Li, J., Chen, Q., Long, Z., Wang, W., Zhu, H., & Wang, L. (2021). Spectrum sensing with non-Gaussian noise over multi-path fading channels towards smart cities with IoT. IEEE Access, 9, 11194–11202. https://doi.org/10.1109/access.2021.3051719

[6] Dobre, O. A., Abdi, A., Bar-Ness, Y., & Su, W. (2007). Survey of automatic modulation classification techniques: Classical approaches and new trends. IET Communications, 1(2), 137. https://doi.org/10.1049/iet-com:20050176Wang,

[7] D., Zhang, M., Li, J., Li, Z., Li, J., Song, C., & Chen, X. (2017). Intelligent constellation diagram analyzer using convolutional neural network-based deep learning. Optics Express, 25(15), 17150. https://doi.org/10.1364/oe.25.017150

[8] Hassan, K., Dayoub, I., Hamouda, W., & Berbineau, M. (2010). Automatic modulation recognition using wavelet transform and neural networks in wireless systems. EURASIP Journal on Advances in Signal Processing, 2010(1). https://doi.org/10.1155/2010/532898

[9] Parmar, A., Chouhan, A., Captain, K., & Patel, J. (2024). Deep multilevel architecture for automatic modulation classification. Physical Communication, 64, 102361. https://doi.org/10.1016/j.phycom.2024.102361

[10] O'Shea, T. J., Corgan, J., & Clancy, T. C. (2016). Convolutional radio modulation recognition networks. In Engineering Applications of Neural Networks (pp. 213–226). https://doi.org/10.1007/978-3-319-44188-7_16

[11] Liu, X., Yang, D., & El Gamal, A. (2017). Deep neural network architectures for modulation classification. In 2017 51st Asilomar Conference on Signals, Systems, and Computers (pp. 1–5). https://doi.org/10.1109/acssc.2017.8335483

[12] Zhou, F., Li, J., & Wang, Y. (2023). An improved CNN-LSTM network for modulation identification relying on periodic features of signal. IET Communications, 17(18), 2097–2106. https://doi.org/10.1049/cmu2.12682

[13] Hamza, M. A., Alghamdi, A. M., Alzahrani, J. S., Alharbi, M. T., & Al-Turki, Y. (2022). Optimal bidirectional LSTM for modulation signal classification in communication systems. Computers, Materials & Continua, 72(2), 3055–3071. https://doi.org/10.32604/cmc.2022.024490

[14] Zhu, Z., Sun, D., Gong, K., Wang, W., & Sun, P. (2021). A lightweight CNN architecture for automatic modulation class ification. Electronics, 10(21), 2679. https://doi.org/10.3390/electronics10212679

[15] Wei, W., & Mendel, J. M. (2000). Maximum-likelihood classification for digital amplitude-phase modulations. IEEE Transactions on Communications, 48(2), 189–193. https://doi.org/10.1109/26.823550

[16] Panagiotou, P., Anastasopoulos, A., & Polydoros, A. (2000). Likelihood ratio tests for modulation classification. In MILCOM 2000 Proceedings. In 21st Century Military Communications. Architectures and Technologies for Information Superiority (Vol. 2, pp. 670–674). https://doi.org/10.1109/milcom.2000.904013

[17] Wu, H.-C., Saquib, M., & Yun, Z. (2008). Novel automatic modulation classification using cumulant features for communications via multipath channels. IEEE Transactions on Wireless Communications, 7(8), 3098–3105. https://doi.org/10.1109/twc.2008.070015

[18] Ramkumar, B. (2009). Automatic modulation classification for cognitive radios using cyclic feature detection. IEEE Circuits and Systems Magazine, 9(2), 27–45. https://doi.org/10.1109/mcas.2008.931739

[19] Pawar, S. U., & Doherty, J. F. (2011). Modulation recognition in continuous phase modulation using approximate entropy. IEEE Transactions on Information Forensics and Security, 6(3), 843–852. https://doi.org/10.1109/tifs.2011.2159000

[20] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008.

[21] Srinivas, A., Lin, T.-Y., Parmar, N., Shlens, J., Abbeel, P., & Vaswani, A. (2021). Bottleneck transformers for visual recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 16519–16529). https://ieeexplore.ieee.org/abstract/document/9577771

[22] Lin, S., Zeng, Y., & Gong, Y. (2022). Learning of time-frequency attention mechanism for automatic modulation recognition. IEEE Wireless Communications Letters, 11(4), 707–711. https://doi.org/10.1109/lwc.2022.3140828

[23] Zhang, W., Sun, Y., Xue, K., & Yao, A. (2023). Research on modulation recognition algorithm based on channel and spatial self-attention mechanism. IEEE Access, 11, 68617–68631. https://doi.org/10.1109/access.2023.3292408

[24] Liang, Z., Tao, M., Xie, J., Yang, X., & Wang, L. (2022). A radio signal recognition approach based on complex-valued CNN and self-attention mechanism. IEEE Transactions on Cognitive Communications and Networking, 8(3), 1358–1373. https://doi.org/10.1109/tccn.2022.3179450

[25] Qi, M., Shi, N., Wang, G., & Shao, H. (2024). Data-transform multi-channel hybrid deep learning for automatic modulation recognition. IEEE Access, 12, 59113–59121. https://doi.org/10.1109/access.2024.3393481

[26] Feng, Y., Duan, R., Li, S., Cheng, P., & Liu, W. (2025). A dual-branch network with feature assistance for automatic modulation recognition. IEEE Signal Processing Letters, 32, 701–705. https://doi.org/10.1109/lsp.2025.3527901

[27] Zhang, W., Xue, K., Yao, A., & Sun, Y. (2024). CTRNet: An automatic modulation recognition based on transformer-CNN neural network. Electronics, 13(17), 3408. https://doi.org/10.3390/electronics13173408

[28] Duan, R., Zhang, S., Wang, X., Li, Y., Liu, W., & Feng, Y. (2023). A multi-modal modulation recognition method with SNR segmentation based on time domain signals and constellation diagrams. Electronics, 12(14), 3175. https://doi.org/10.3390/electronics12143175

[29] Luo, Z., Xiao, W., Zhang, X., Zhu, L., & Xiong, X. (2024). RLITNN: A multi-channel modulation recognition model combining multi-modal features. IEEE Transactions on Wireless Communications, 23, 1–15. https://doi.org/10.1109/twc.2024.3478752

[30] Kong, W., Jiao, X., Xu, Y., Zhang, B., & Yang, Q. (2023). A transformer-based contrastive semi-supervised

learning framework for automatic modulation recognition. IEEE Transactions on Cognitive Communications and Networking, 9(4), 950–962. https://doi.org/10.1109/tccn.2023.3264908

[31] O'Shea, T. J., Roy, T., & Clancy, T. C. (2018). Over-the-air deep learning based radio signal classification. IEEE Journal of Selected Topics in Signal Processing, 12(1), 168–179. https://doi.org/10.1109/jstsp.2018.2797022

[32] O'Shea, T. J., & West, N. (2016). Radio machine learning dataset generation with GNU Radio. Proceedings of the GNU Radio Conference, 1(1). https://pubs.gnuradio.org/index.php/grcon/article/view/11

[33] Xu, J., Luo, C., Parr, G., & Luo, Y. (2020). A spatiotemporal multi-channel learning framework for automatic modulation recognition. IEEE Wireless Communications Letters, 9(10), 1629–1632. https://doi.org/10.1109/lwc.2020.2999453

[34] Perenda, E., Rajendran, S., & Pollin, S. (2019). Automatic modulation classification using parallel fusion of convolutional neural networks. IEEE Transactions on Vehicular Technology, 69(9), 9825–9837. https://doi.org/10.1109/tvt.2020.3000148

[35] Huynh-The, T., Hua, C.-H., Pham, Q.-V., & Kim, D.-S. (2020). MCNet: An efficient CNN architecture for robust automatic modulation classification. IEEE Communications Letters, 24(4), 811–815. https://doi.org/10.1109/lcomm.2020.2968030

[36] Zhang, F., Luo, C., Xu, J., Luo, Y., & Zheng, F.-C. (2022). Deep learning based automatic modulation recognition: Models, datasets, and challenges. Digital Signal Processing, 129, 103650. https://doi.org/10.1016/j.dsp.2022.103650