Modulation Recognition of MIMO Systems Based on Dimensional Interactive Lightweight Network

Sileng Aer, Xiaolin Zhang*, Zhenduo Wang, and Kailin Wang
College of Information and Communication Engineering, Harbin Engineering University
Harbin, 150001, China
[e-mail: aersileng@hrbeu.edu.cn, zhangxiaolin@hrbeu.edu.cn, zhenduowang@hrbeu.edu.cn,
529653295@hrbeu.edu.cn]
*Corresponding author: Xiaolin Zhang

Abstract

Automatic modulation recognition is the core algorithm in the field of modulation classification in communication systems. Our investigations show that deep learning (DL) based modulation recognition techniques have achieved effective progress for multiple-input multiple-output (MIMO) systems. However, network complexity is always an additional burden for high-accuracy classifications, which makes it impractical. Therefore, in this paper, we propose a low-complexity dimensional interactive lightweight network (DilNet) for MIMO systems. Specifically, the signals received by different antennas are cooperatively input into the network, and the network calculation amount is reduced through the depth-wise separable convolution. A two-dimensional interactive attention (TDIA) module is designed to extract interactive information of different dimensions, and improve the effectiveness of the cooperation features. In addition, the TDIA module ensures low complexity through compressing the convolution dimension, and the computational burden after inserting TDIA is also acceptable. Finally, the network is trained with a penalized statistical entropy loss function. Simulation results show that compared to existing modulation recognition methods, the proposed DilNet dramatically reduces the model complexity. The dimensional interactive lightweight network trained by penalized statistical entropy also performs better for recognition accuracy in MIMO systems.

1. Introduction

Automatic modulation recognition is an important technique for analyzing unauthorized users in non-cooperative communications. In addition, the accurate recognition of modulation types provides essential technical support for parameter estimation, demodulation, and interference of received signals.[1]-[4] Therefore, the concept of modulation recognition has been widely studied since proposed.

The existing automatic modulation classification techniques mainly include two kinds of methods. i.e. the maximum likelihood detection method and the other **is** the feature classification method. In [5], J.B.Tamakuwala propose a maximum-likelihood-based modulation recognition method. However, the computational complexity is too high and there are more requirements for the acquisition of prior information. The author of [6] propose a novel decision-theoretic approach to modulation classification for MIMO systems, where the channel state information (CSI), the number of transmit antennas, and the noise variance could be unknown. This approach reduces the need for prior information in the maximum likelihood method, but the computational complexity is still not effectively addressed. Therefore, more and more scholars solve the problem of modulation recognition by studying pattern recognition methods that could eliminate the limitation of prior information. The pattern recognition method mainly includes two processes: feature extraction and feature classification. In the paper thesis [7] and [8], the feature extraction includes instantaneous features, high-order cumulant features (HOC), and cyclic spectral features. The HOC can effectively overcome the additive white Gaussian noise (AWGN). The cyclic spectral features also have good noise immunity due to the cyclic smooth and stable characteristic. In addition, some scholars have also made important contributions to the research under non-Gaussian noise environments.[9]-[10] The authors use the adaptive weight myriad filter to preprocess the impulse noise in the received signal and propose an evolutionary neural network based on the quantum elephant herding algorithm (QEHA) to classify the modulated signal.[11] These methods usually have a good classification effect in single-in single-out (SISO) systems [12]-[13], but the space-time aliasing caused by the MIMO channel seriously degrades the effectiveness of the classical methods. Therefore, the traditional pattern recognition method can no longer adequately meet the needs of modulation recognition in the MIMO systems.

Recently, deep learning (DL) has been widely used in modulation recognition due to its robust feature extraction and accurate classification capabilities. T.OShea and J.Hoydis in [14] first combine the Convolutional Neural Network (CNN) network with the modulation recognition algorithm and achieved far better performance than traditional algorithms. This is a milestone advancement in the research process of modulation recognition problems. Different from conventional pattern recognition methods, feature extraction is developed from artificial features to neural network features. The sampled data in the modulation recognition includes mainly the in-phase/quadrature (I/Q) information of the signal and the constellation map information of the amplitude-phase mapping. In [15]-[18], Long Short-Term Memory (LSTM) and Convolutional Long Short-Term Deep Neural Network (CLDNN) are proposed to take signal I/Q information as input. Y.Mao proposed a graph neural network based on constellation map information that possesses excellent performance for phase modulation signal recognition in literature [19]. However, all these methods require a large number of training samples and the model is very computationally intensive. Existing networks all use SISO system received signal as a sample data set for recognition. The modulation characteristics of MIMO systems received signals are severely fading and can not be reliably classified by the network. Therefore, the modulation recognition problem for MIMO systems

requires the design of a novel network that can not only recognize the signal modulation method with high accuracy but also keep the low complexity of the network.

In response to above questions, Y. Wang proposed a cooperative decision algorithm based on the CNN network in [20]. The cooperative decision algorithm effectively solves the problem of large errors in judging the modulation type. But it also generates recognition divergence among different sub-results, which is very limited to the improvement of recognition in low SNR environments, and it is also difficult to achieve 100% accuracy in high SNR environments. In [21], the CNN-based zero-forcing (ZF) equalization method was proposed for MIMO systems. However, this approach is strongly influenced by channel prior information estimation accuracy.

Summarizing the above problems, this paper designs a lightweight network with little influence from prior information by fully considering the two critical factors of classification accuracy and algorithm complexity of the MIMO system. The input samples of the lightweight network are the I/Q information of the signals received by all antennas. Additionally, it is found in [18]-[23] that adding an attention module to a deep learning network can significantly improve classification accuracy. The attention module can effectively guide the network to extract the network features needed for classification, just as humans consciously learn essential knowledge. Therefore, this paper designs a two-dimensional interactive attention mechanism (TDIA). The TDIA module is used to extract the time series interaction information and the channel interaction information. Different channels correspond to different receiving antennas. Then, the dimensional interactive lightweight network (DilNet) designed in this paper introduces depth-wise separable convolution (DSC) to reduce the complexity of the residual structure. The DSC drastically reduces the complexity of the algorithm at the cost of small classification accuracy, but it greatly improves the lightness of the DilNet. Finally, the DilNet is trained through the penalized statistical entropy loss function. The simulation results show that the DilNet can accurately classify the modulation methods and greatly reduce the complexity. The main contributions of this paper are as follows.

- The TDIA extracts the cooperative features of the received signals from different antennas. The experiment results show that the TDIA module can efficiently extract the interaction information of spatial dimension and channel dimension. The network features would have two-dimensional interaction characteristics, which could significantly improve the classification accuracy of digitally modulated signals in MIMO systems.

- We designed a dimensional interactive lightweight network (DilNet) with four residual layers and reduced the complexity of convolutional operations by depth-wise separable convolution (DSC). The TDIA module is embedded in the DilNet network. Compared with existing networks, this network has higher classification accuracy and lower complexity. But it usually requires more training epoch times to obtain a better classification model.

- We define a novel way of calculating the loss function. This method uses penalized statistical entropy to calculate the incremental output information for each residual layer and weigh it. Compared with the method of calculating loss function by cross-entropy, the proposed method could improve the classification accuracy of DilNet significantly.

2. MIMO Signal Model And Dataset Generation

In **Fig. 1**, the MIMO system consists of six processes: random generation of symbols, baseband modulation, power normalization, signal format reconstruction, antenna transmission, and antenna reception. The receiver guarantees random reception of all transmitted signals.
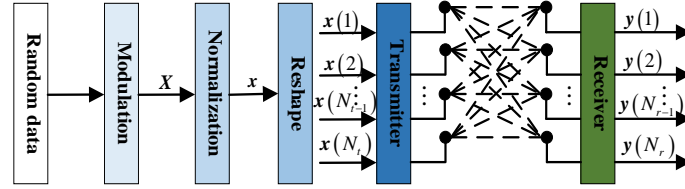
Fig. 1. MIMO system signal generation

2.1 MIMO System And Signal Generation

Digital signal modulation obtains different modulated signals by controlling the carrier wave frequency, amplitude, and phase characteristics. The signal modulated by different modulation methods not only has the same information as the original signal but also enhances the transmission efficiency. The mathematical models of modulation types in this paper are shown below.

Phase-shift keying (PSK) signal is a phase modulated signal. And the PSK signal modulated by using a rectangular pulse ($g(.)$) can be expressed as

$$X_{MPSK}(t) = \left[\sum_n g(t - nT_s)\right]\cos\left(2\pi f_c t + \varphi_n\right) \tag{1}$$

where φ_n represents the phase of the n-th symbol in the signal. f_c is the carrier frequency, T_s is the symbol period.

Quadrature amplitude modulation (QAM) signals are phase and amplitude modulated signals that can be expressed as

$$S_{MQAM}(t) = \left[\sum_n a_n g(t - nT_s)\right]\cos\left(2\pi f_c t + \varphi_n\right) \tag{2}$$

where a_n represents the amplitude of the n-th symbol and other parameters have the same meaning as (1).

It is assumed that the MIMO channel has N_t transmit antennas and N_r receive antennas $(N_t \leq N_r)$, MIMO channel is a flat-fading and time-invariant channel. The received signal at the time k-th can be expressed as

$$y_k = \mathbf{R}x_k + \mathbf{G}_k \tag{3}$$

where $x_k = [x_k(1), x_k(2),..., x_k(N_t)]^T$ is the modulated signal vector $(N_t \times 1)$, $y_k = [y_k(1), y_k(2),..., y_k(N_r)]^T$ is the received baseband signal vector $(N_r \times 1)$, G_k is additive white Gaussian noise.

$$\mathbf{R} = \begin{bmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,N_t} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,N_t} \\ & & \cdots & \\ r_{N_r,1} & r_{N_r,2} & \cdots & r_{N_r,N_t} \end{bmatrix} \tag{4}$$

where $\mathbf{R}$ is the $(N_r \times N_t)$ MIMO channel matrix with mean equal to 0 and variance equal to 1, and r_{N_r,N_t} represents the channel parameter. In addition, the matrix $\mathbf{R}$ obeys a complex Gaussian distribution.

2.2 Dataset Generation

This paper generates the data set with modulation types (2PSK, 4PSK, 8PSK, 16QAM) according to the MIMO system signal model. The sample dataset generation process is shown in **Fig. 1**. The vector of modulated signals can be given as $X = [X_1, X_2, \ldots, X_N]$, N represents the length of the signal vector. To reasonably distinguish X from different modulation types, the power of the signal is normalized as x. It is assumed that the number of transmitting antennas is N_t. The normalized signal x can be reshaped as $[x(1), x(2), \ldots, x(N_t)]^T$, with dimension $N_t \times N / N_t$. The signal transmitted by one of the antennas can be expressed as $x(j) = [x_1(j), x_2(j), \ldots, x_{N/N_t}(j)]^T$, $j \in [1, N_t]$. The j-th transmitter antenna, $x(j)$ has the N / N_t continuously transmitted symbols. Continuous signals of length N / N_t transmitted by N_t antennas pass through a MIMO channel, where the MIMO channel model R obeys the complex Gaussian distribution. The noise environment is additive white Gaussian noise. The signal will be received by N_r antennas and the baseband signal will be obtained by frequency conversion. The receiver sampled signal can be expressed $y = [y(1), y(2), \ldots, y(N_r)]^T$, with dimension $N_r \times (f_D \times N / N_t)$. The i-th antenna receiver sampled signal $y(i) = [y_1(i), y_2(i), \ldots, y_{f_D \times N/N_t}(i)]$, $i \in [1, N_r]$, with length $f_D \times N / N_t$. f_D is the number of sampling points for per symbol.

In the MIMO system, the signals received by different antennas correspond to different channel dimensions, and the in-phase and quadrature parts of the signals correspond to the spatial dimensions. The format of the sample dataset is $Y = y, y \in R^{N_r \times 2 \times \left(f_D \times \frac{N}{N_t} \right)} \in R^{C \times W \times H}$, $C = N_r$ is the number of sample channels (number of receiving antennas), $W = 2$ is the sample width (the in-phase and quadrature information of the received signal) and $H = f_D \times N / N_t$ is the sample length (received signal length). Because the signals transmitted by the transmit antennas are modulated in the same way, extracting the cooperative features of the signals received by all antennas helps to correctly classify the signal modulations.

3. Proposed Dimensional Interactive Lightweight Network Method

3.1 Two-dimensional Interactive Attention Mechanism

In MIMO system, the signals are mainly affected by environmental noise and space-time aliasing from different transmit antenna signals. Due to the transmit antennas being modulated in the same way, space-time aliasing will lead to severe fading of signal amplitude and phase variation features. Antenna diversity is the use of two receive channels, their fading effects are uncorrelated. It is very unlikely that both will experience the same deep fade point at the same time. When combining the signals from the two antennas, the degree of fading can be reduced. Space-diversity approach can overcome space-selective fading. Therefore, the classification of the modulation in the MIMO system needs to consider the received signals of all antennas.

Liang proposes a two-dimensional attention mechanism module that can effectively extract channel and spatial features. [18] This approach verifies the positive impact of the attention mechanism module on feature extraction. Therefore, this paper designs a TDIA module to help the network extract the collaborative interaction features of signals from different antennas.

The interaction information extracted by the TDIA module includes spatial interaction information among different sampling points, channel interaction information among different receiving antennas, and channel spatial across dimensions interaction information.
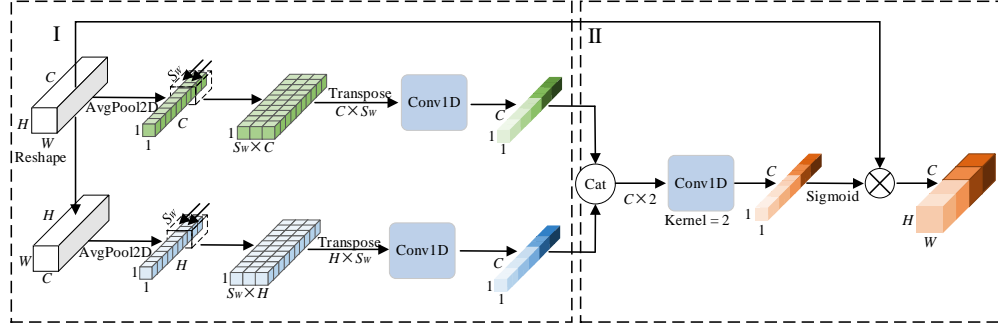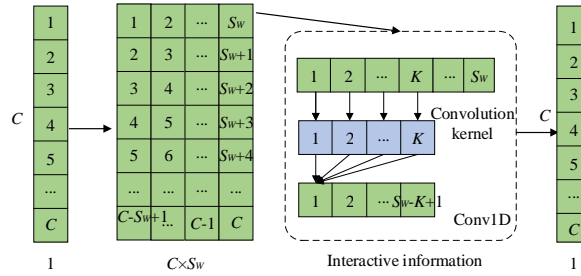

Fig. 2. Two-dimensional interactive attention mechanism

Fig. 3. The principle of sliding window extraction of interactive information.

In **Fig. 2**, TDIA consists of two parts I and II. I is to extract the interaction feature of channel and spatial from the sampled signals through two branches, respectively. In the first branch, the spatial information is averagely pooled by the (5) to obtain independent channel information.

$$Y_{gap} = \frac{1}{WH} \sum_{w=1}^{W} \sum_{h=1}^{H} Y_{wh} \tag{5}$$

where $Y_{gap} \in R^{C \times 1 \times 1}$ represents the global average pooling of the input feature $Y \in R^{C \times W \times H}$. Different channel information is extracted by the sliding window method, and the sliding window size is S_W. The S_W adjacent channel information is placed in the same spatial dimension through sliding window processing in **Fig. 3**. The feature size of the channel information becomes $Y_{S_W} \in R^{S_W \times C \times 1}$, and transpose converts the dimension to $Y_{S_W} \in R^{C \times S_W}$. Each channel contains the information of K adjacent channels. Extracting the interaction information of adjacent channels can be done by convolution, and the specific calculation is as follows:

$$g(Y^{C \times S_w^n}) = \begin{bmatrix} Y_{(n,n+k)}^1 \\ Y_{(n,n+k)}^2 \\ \cdots \\ Y_{(n,n+k)}^C \end{bmatrix} \otimes \begin{bmatrix} \alpha_1^1 & \cdots & \alpha_K^1 \\ \alpha_2^2 & \cdots & \alpha_{K+1}^2 \\ & \cdots & \\ \alpha_{C-K+1}^C & \cdots & \alpha_C^C \end{bmatrix}^T \tag{6}$$

Equation (6), $g(Y^{C \times S_w^n})$ represents the n-th extraction of interaction information in traversing the window of size S_W, $\otimes$ is the matrix operation symbol for the same channel $\left(Y_{(n,n+K)}^C \times \left[\alpha_{C-K+1}^C \ \ldots \ \alpha_C^C \right]^T \right)$. The interaction information of each channel under the same window is extracted by one-dimensional convolution (Conv1D), α is the convolution weight parameter, and K is the convolution kernel size. The interaction information of K adjacent channels can be extracted by one convolution.

$$Y^{C \times (S_W - K+1)} = f_{(1,(S_W - K+1))} \left(g \left(Y^{(C \times S_W)} \right) \right) \tag{7}$$

where $f_{(1,(S_W - K+1))}$ is the statistics of all interaction information in the compressed window after one traversal, the feature size of the channel information becomes $Y \in R^{C \times (S_W - K+1)}$. The interaction information of S_W adjacent channels can not be extracted by one traversal of convolution when the sliding window is larger than the convolution kernel. Loop traversal until output channel interaction feature size is $Y \in R^{C \times 1}$, complete S_W channel interaction information extraction.

In the second branch of **Fig. 2**, the input feature size is reshaped to $Y \in R^{H \times W \times C}$ and the spatial information is compressed into $Y_{gap} \in R^{H \times 1 \times 1}$ by average pooling. At this time, the feature size extracted by the sliding window is $Y_{S_W} \in R^{H \times S_W}$. After traversing all the information of the window through the convolution operation, the feature size becomes $Y \in R^{H \times (S_W - K+1)}$. Unlike the first branch, the output feature $Y \in R^{H \times 1}$ after loop traversal expands the feature size to $Y \in R^{C \times 1}$ by the convolution operation with kernel size 1 when the lengths of H and C are different.

$$Y^{C \times 2} = cat \left(Y^C, Y^H \right) \tag{8}$$

The work of part II is the interaction of channel and spatial features. In formula (8), the channel interaction information and spatial interaction information obtained by the two branches in part I are spliced (cat) as $Y^{C \times 2} \in R^{C \times 2}$. Next, the interactive information of splicing feature is extracted by Conv1D. The feature contains a column of spatial information and a column of channel information, so only one convolution is needed to complete the interaction of spatial and channel information. The output feature size is $Y^{C \times 1} \in R^{C \times 1}$.

$$Y_{weight} = \delta \left(Y^{C \times 1} \right) \tag{9}$$

$$Y_{out} = Y \times Y_{weight} \tag{10}$$

Finally, the interactive features are output through the sigmoid activation function δ. The dimension size of the output feature is expanded from $Y_{weight} \in R^{C \times 1}$ to $Y_{weight} \in R^{C \times 1 \times 1}$ and applied to the input information as a weight. $Y_{weight} \in R^{C \times 1 \times 1}$ and $Y \in R^{C \times W \times H}$ are the same in the first dimension. Therefore, they are multiplied together in the first dimension. The output feature size of TDIA is $Y_{out} \in R^{C \times W \times H}$, and the feature size does not change compared to the input feature. However, the output of the TDIA module realizes channel interaction, spatial interaction, and channel-space interaction by adding weight information. We obtain weighted output features with two-dimensional interaction characteristics and realize the extraction of cooperative features of received signals from different antennas. In section 4, we will fully verify the effectiveness of the TDIA module through comparative experiments.

3.2 Dep-thwise Separable Convolution

The difference between depth-wise separable convolution (DSC) and standard convolution is that the standard convolution has only one step of convolution processing. However, DSC consists of two steps: depth-wise convolution and point-wise convolution. Depth-wise convolution only processes spatial information within one channel, while point convolution is a standard convolution with a convolution kernel size of 1×1, which is a centralized process of channel information. The DSC completes one channel and spatial information process by concatenating depth-wise convolution and point-wise convolution.
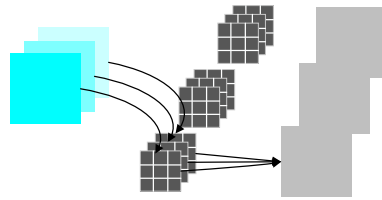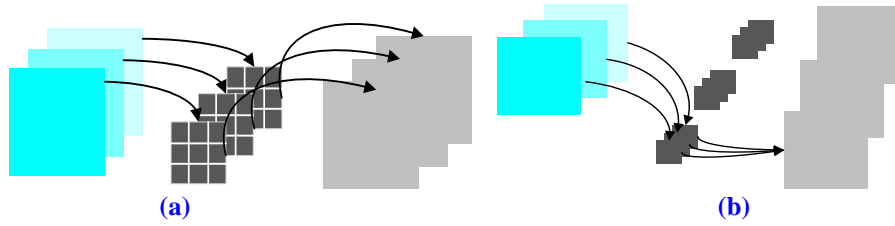

Fig. 4. Standard convolution

(a)   **(b)**

Fig. 5. Depth-wise separable convolution **(a)** Depth-wise convolution **(b)** Point-wise convolution

By comparing **Fig. 4** and **Fig. 5**, we can find the different structures of the depth separable convolution and the standard convolution. The size of the input feature is $Y \in R^{C_i \times W_i \times H_i}$ and output feature size is $Y \in R^{C_o \times W_o \times H_o}$ by the standard convolution. When the size of the convolution kernel for extracting spatial features is $D_W \times D_H$, the standard convolution parameter and the calculation amount are $C_i \times C_o \times D_W \times D_H$ and $C_i \times C_o \times D_W \times D_H \times W_o \times H_o$, respectively. In the case where the input feature size and output feature size are the same, the parameter of depth-wise convolution is $C_i \times D_W \times D_H$ and the parameter of point-wise convolution is $C_i \times C_o$. The calculation amount of depth-wise convolution is $C_i \times D_W \times D_H \times W_o \times H_o$ and the calculation amount of point-wise convolution is $C_i \times C_o \times W_o \times H_o$. Therefore, the parameter ratio of depth-wise separable convolution and standard convolution can be expressed as

$$P_{D/S} = \frac{C_i \times D_W \times D_H + C_i \times C_o}{C_i \times C_o \times D_W \times D_H} = \frac{1}{C_o} + \frac{1}{D_W \times D_H} \tag{11}$$

the calculation ratio can be expressed as:

$$C_{D/S} = \frac{C_i \times D_W \times D_H \times W_o \times H_o + C_i \times C_o \times W_o \times H_o}{C_i \times C_o \times D_W \times D_H \times W_o \times H_o} = \frac{1}{C_o} + \frac{1}{D_W \times D_H} \tag{12}$$

It can be seen from (10) and (11) that the DSC method divides the convolution into depth-wise convolution and point-wise convolution, which reduces the number of convolution parameters and calculations. We will apply DSC to the network in this paper in a reasonable way.

3.3 Dimensional Interactive Lightweight Network Structure

The residual structure usually has an excellent performance in classification problems. [27]-[29] Based on the residual structure, this paper designs a dimensional interactive lightweight network (DilNet) for modulation recognition in MIMO systems. The residual structure can efficiently improve the classification accuracy of the network, but high complexity is often difficult to practice. Therefore, the depth-wise separable convolution is introduced to replace the standard convolution in the residual structure to reduce the network complexity. In addition, the performance of deep neural network models for image recognition in signal recognition is limited. The number of model channels grows with the number of network layers. However, the network features of the signal are relatively easy to extract. Increasing the number of channels in the hidden layer is not conducive to network convergence. Therefore, the number of network channels proposed in this paper decreases as the depth of the model increases.
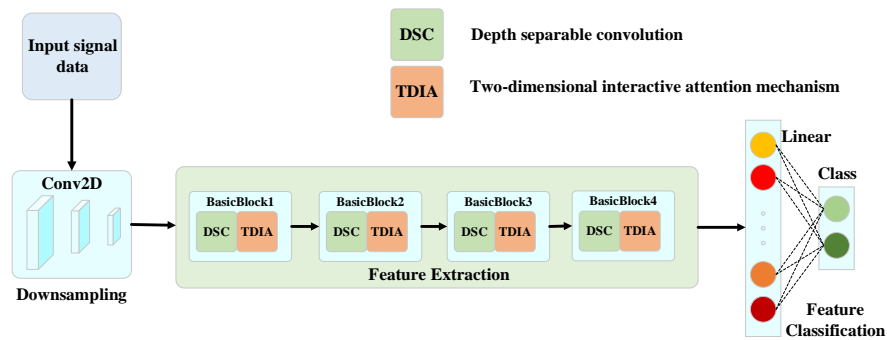
Fig. 6. Dimensional interactive lightweight network structure

The proposed network consists of downsampling, feature extraction, and feature classification, which can be seen in **Fig. 6**. Among them, downsampling is composed of two-dimensional standard convolution (Conv2D) to prevent serious information loss during downsampling. As the core part of the network, feature extraction consists of residual block and TDIA module, which is used to extract efficient cross-dimensional interaction features. The feature classifier outputs the classification results through the fully connected layer.

(a)   **(b)**

Fig. 7. (a) Standard residual block **(b)** DSC residual block

In **Fig. 7 (a)** and **Fig. 7 (b),** the difference between the DSC residual block and the standard residual is that the DSC residual block uses depth-wise convolution to process spatial informat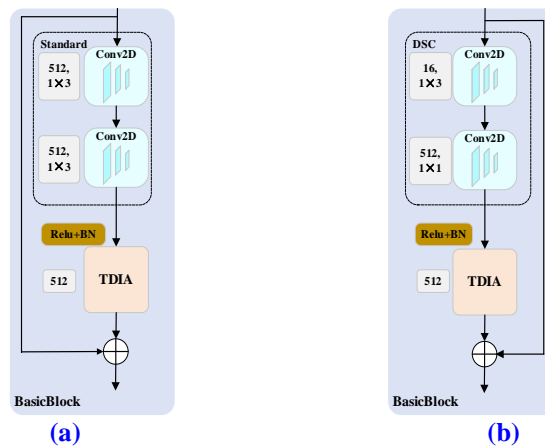ion. The number of channels is not dilated in the first two-dimensional convolution (Conv2D) ensuring that the output feature sizes are the same as standard convolutions. Standard convolution processes spatial information and completes the dilation of channel information in the first Conv2D, which are calculated according to formulas (11) and (12) as follows:

$$P_{D/S} = \frac{C_i \times D_W \times D_H + C_i \times C_o}{\left(C_i + C_o\right) \times C_o \times D_W \times D_H} = \left(\frac{C_i}{C_i + C_o}\right)\left(\frac{1}{C_o} + \frac{1}{D_W \times D_H}\right) \tag{13}$$

$$C_{D/S} = \frac{C_i \times D_W \times D_H \times W_o \times H_o + C_i \times C_o \times W_o \times H_o}{\left(C_i + C_o\right) \times C_o \times D_W \times D_H \times W_o \times H_o} = \left(\frac{C_i}{C_i + C_o}\right)\left(\frac{1}{C_o} + \frac{1}{D_W \times D_H}\right) \tag{14}$$

It can be seen from (12) and (13) that compared with the standard residual block, the DSC residual block designed in this paper has significantly reduced parameters and calculations. Then, the TDIA module is embedded behind the DSC, which makes the features extracted by each DSC residual block have two-dimensional interactive characteristics.

Table 1. The structural parameters of dimensional interactive lightweight network

No.	Structure
1	Conv2D (channel(N_r , 16), kernel (7×7)) +BN+Relu
2	DSC (channel(16,512), kernel (1×3)) +BN+Relu+TDIA
3	DSC (channel(512,256), kernel (1×3)) +BN+Relu+TDIA
4	DSC (channel(256,128), kernel (1×3)) +BN+Relu+TDIA
5	DSC (channel(128,64), kernel (1×3)) +BN+Relu+TDIA
6	Linear($|M|$) + Softmax

In **Table 1**, we provide the specific layout of DilNet, where Conv2D is standard convolutional layer. BN is to prevent overfitting, Linear is fully-connected layer and Relu is activation function.

3.4 Penalized Statistical Entropy

Table 2. Penalized Statistical Entropy

Algorithm 1 The Penalized Statistical Entropy Method.
Input Y is the sample to be classified.
Output loss^j is the loss value after the j-th iteration.
1: N is the number of one batch samples; L is the number of residual layers; $\left(p_{ni}^j(l)\right)'$ is the prediction result of the j-th iteration and the l-th residual layers; $p_{ni}^j(l)$ is the probability.
2: **for** $l = 1$ **to** L **do**
3: The l-th residual layer extracts features are classified by the fully connected layer, where the probability distribution is $p_{ni}^j(l) = \dfrac{e^{\left(p_{ni}^j(l)\right)'}}{\sum\limits_{n=1}^{N} e^{\left(p_{ni}^j(l)\right)'}}$.
4: Calculate the loss function: $\text{loss}_l^j = \dfrac{1}{N} \times \text{loss}_l^{(j-1)} \times \sum\limits_{n=1}^{N} \sum\limits_{i=1}^{I} y_{ni}^j(l) \log\left(p_{ni}^j(l)\right)$.
5: **end for**

6: The statistical loss function: $\text{loss}^{j} = -\dfrac{1}{L}\sum_{l=1}^{L}\text{loss}_{l}^{(j-1)}$

7: **return** loss^{j}

As one of the calculation methods of loss function, cross-entropy is widely used in training modulation recognition models. The softmax function usually converts the network output into a probability distribution. This process can be defined as

$$p_{ni} = \frac{e^{(p_{ni})'}}{\sum_{n=1}^{N} e^{(p_{ni})'}} \tag{15}$$

where $(p_{ni})'$ is the network prediction, p_{ni} is the predicted probability after processing by the softmax function. i is one of the classification result labels of sample n-th .

$$\text{loss} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{I} y_{ni} \log(p_{ni}) \tag{16}$$

where N is the total number of samples, I is the number of all classification categories. y_{ni} is the true sample distribution probability.

The cross-entropy loss function can effectively statistics output information, but only calculate the cross-entropy of the output layer. In the lightweight network, the feature extraction of each residual layer after adding the attention module will have an important impact on the classification results due to the small number of network layers. It is necessary to gather loss information for the output features of each residual layer. Therefore, this article proposes penalized statistical entropy to calculate the loss function, which can be designed in **Table 2**.

loss_{l}^{j} represents the cross-entropy of the l-th network layer during the j-th round of iteration. If the loss of a certain layer suddenly increases, it will affect the loss of the corresponding layer in the next iteration process. We define this problem as the unqualified feature extraction effect of this layer, and use the loss to punish the corresponding layer in the next iteration. $\left[\text{loss}_{1}^{j-1},\ \text{loss}_{2}^{j-1},\dots,\ \text{loss}_{L}^{j-1}\right]$ is the penalty term for the hidden layer that needs statistical penalty loss information during the j-th iteration. The network avoids finding locations that lead to penalties to reduce the loss. When the loss of each layer is minimized by this method, the loss statistics of the entire network will also be reduced, each layer can extract efficient classification features, and the recognition rate of the network will also be significantly improved.

4. Simulation Results

Table 3. The simulation parameters of modulation signals

No.	Parameter	Value
N_s	Symbol length	128
f_d	Symbol rate	500 symbol/s
f_c	Carrier frequency	2000 Hz
f_s	Sampling frequency	16000 Hz
f_D	**Number of sampling points for per symbol**	16
SNR	Signal-to-noise ratio	[-10,10]

Table 4. The environment parameters of DilNet

No.	Parameter	Value
Batch_size	Number of samples taken in one training	32
CPU	Central processing unit	i7-11800
RAM	Random access memory	DDR4 16GB
GPU	Graphics processing unit	RTX3060 6GB
N	Number of samples per modulated signal per SNR	10000
Ratio	The ratio of training set and test set	4:1
S_W	Sliding window size for TDIA modules	5
K	Conv1D kernel size in TDIA module	3

We fully verify the performance of DilNet proposed in this paper for modulation classification in MIMO system. Types of modulated signals include{2PSK, 4PSK, 8PSK, 16QAM }. We experiment on two cases in the MIMO channel, one is that the number of antennas at the transmitting is equal to the receiving $N_t = 4$, $N_r = 4$, and the other is that the number of antennas at the transmitting is less than the receiving $N_t = 2$, $N_r = 4$.

The noise environment in a MIMO channel is Gaussian white noise, and it is defined as follows:

$$SNR = 10\lg(\frac{E_s}{E_G}) \tag{17}$$

where SNR is the signal-to-noise ratio, E_s is signal power, and E_G is noise power. The two classification performance evaluation rules can be expressed as:

$$P_{cc}^{snr} = \frac{S_{correct}^{snr}}{S_{test}} \times 100\% \tag{18}$$

$$P_{cc}^{ave} = \frac{\sum_{snr=-10}^{10} S_{correct}^{snr}}{S_{test} \times N_{snr}} \tag{19}$$

Where P_{cc}^{snr} is the accuracy of a single SNR. P_{cc}^{ave} is the average accuracy of all SNRs. S_{test} is the number of samples in the testset, $S_{correct}^{snr}$ is the number of correctly classified samples in the testset; N_{snr} is the number of SNRs. In addition, we also verify the DilNet classification accuracy in the absence of SNR priors, and DilNet excellent performance in computational complexity. The simulation parameters of modulation signals and the environment parameters of DilNet is presented in **Table 3** and **Table 4**.

4.1 Performance Comparisons of TDIA

In this section, we compare the proposed TDIA module with other attention modules. In the attention module comparison, we guarantee the criterion that the embedding modules are in the same network.
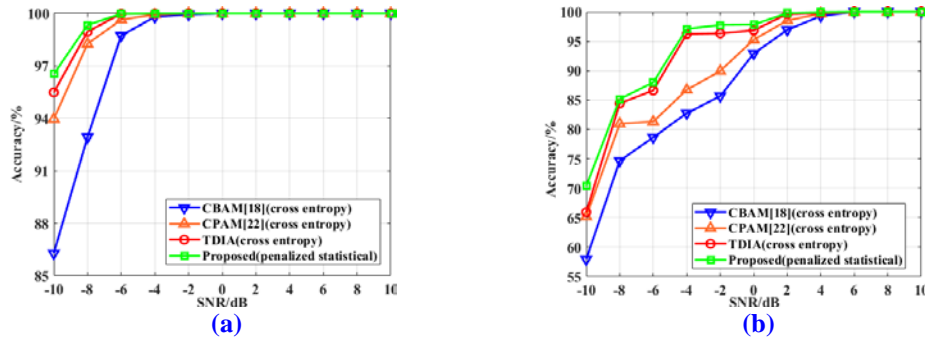
Fig. 8. The classification performance of different modules: **(a)** $N_t = 2$ $N_r = 4$ **(b)** $N_t = 4$ $N_r = 4$

Fig. 8(a) and **Fig. 8(b)** show the classification performance of the same network embedded with different attention modules. When the number of transmit antennas is less than the number of receive antennas, the classification accuracy of digitally modulated signals is higher. It can be observed that the TDIA module has higher classification performance than CBAM and CPAM modules. The CBAM module can help the network extract effective features from the channel and space dimensions, but it can not extract interactive information. The modulation recognition of MIMO system needs to make cooperative decisions on the received signals of all antennas. Therefore, it is very important to extract the cross-dimensional interaction information. This is the reason for the poor performance of the CBAM module. The CPAM interacts with all channel and spatial information, resulting in information redundancy. This is an important reason why the recognition accuracy is lower than TDIA. Besides, the DilNet is trained by using the penalty statistical entropy loss function. Compared with to the TDIA network trained by the cross-entropy loss function, the experiment results show that the penalty statistical entropy can improve the overall recognition accuracy of the signal. This proves that the supervision of the hidden layer is effective. The classification accuracy is improved by about 1%-5% at low SNRs.

4.2 Performance Comparisons of DilNet with Other Automatic Modulation Classification Method

In this section, we compare the proposed network with typical modulation recognition methods. In the network comparison, we follow the same guidelines for the number of network layers and the parameters of the hidden layer. Using the characteristic that the high-order cumulant (HOC) of Gaussian noise is 0, the interference of noise can be effectively reduced. Therefore higher-order cumulant features are used to distinguish modulated signals.[12] The traditional HOC feature extraction method is as follows:

$$M_{pq} = E[y_j(t)^{p-q} y_j^*(t)^q] \qquad (20)$$

where $y_j(t)$ is the signal received by the j-th antenna at time t , M_{pq} is p-order mixing moment, and , $E[.]$ is the mathematical expectation function. Next, HOC features will be extracted based on the mixing moments.

$$
\begin{aligned}
C_{40} &= M_{40} - 3M_{20}^2 \\
C_{41} &= M_{41} - 3M_{20}M_{21} \\
C_{42} &= M_{42} - |M_{20}|^2 - 2M_{21}^2
\end{aligned}
\qquad (21)
$$

According to (21) the theoretical value of the HOC can be calculated. The theoretical values of the four signals are shown in **Table 5.**

It can be seen from **Table 5** that the HOC features can distinguish four kinds of modulation signals. This traditional method usually uses the support vector machine (SVM) as the classifier, so this paper will also compare it.

Table 5. Higher order cumulant theoretical value

Type	2PSK	4PSK	8PSK	16QAM
C_{40}	-2	-1	0	-0.68
C_{41}	-2	0	0	0
C_{42}	-2	-1	-1	-0.68

The performance of different typical methods is shown in **Fig. 9 (a)** and **Fig. 9 (b)**, the classification accuracy of the traditional method based on the HOC feature is significantly lower than that of the convolution feature method extracted by the DL, and it can not be accurately classified at high SNRs when $N_t = N_r$. Because the signal is severely fading in the MIMO channel, the traditional features are affected and distorted. The Co-proposed method is to calculate each CNN model independent output results and make a cooperative decision. However, cooperative features are not extracted in the feature extraction process. There is a lack of information compensation among different antennas. Therefore, the DL methods for extracting cooperative features are significantly better than the Co-proposed method. In addition, comparing the lightweight network proposed in this paper with the other three existing networks (Resnet10, CNN, MobilenetV2), it can be found that the classification accuracy of MobilenetV2, a typical lightweight network, is lower than that of Resnet10 and CNN networks. The reason for the above situation is that existing lightweight network methods reduce network complexity by sacrificing classification accuracy. In this paper, the method of embedding the attention module and training the network with penalized statistical entropy is used to compensate for the classification accuracy without greatly increasing the complexity of the network. The experimental results show that the lightweight network proposed in this paper has the best classification accuracy.
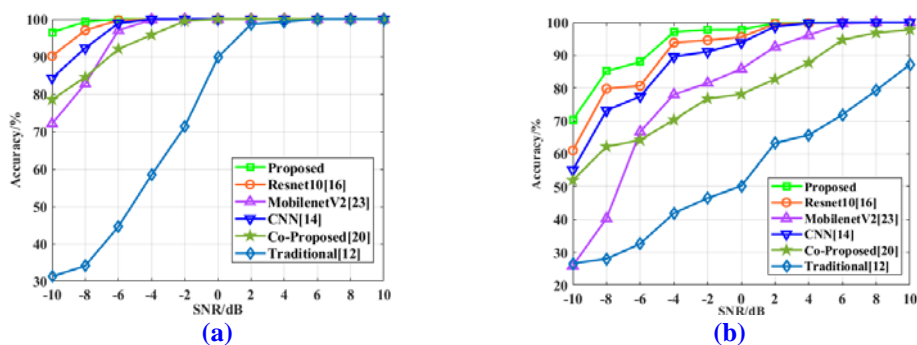

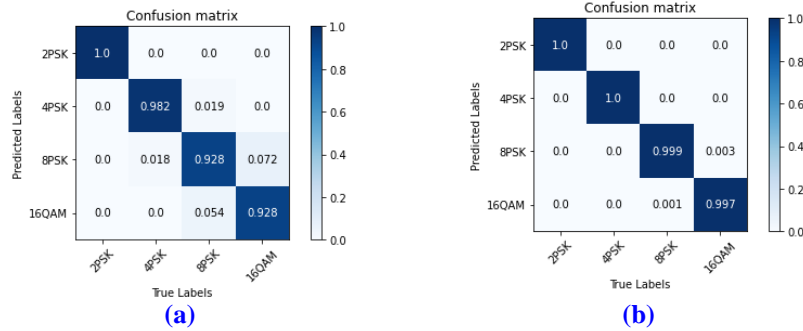**Fig. 9.** The classification performance of different mothods: **(a)** $N_t = 2$ $N_r = 4$ **(b)** $N_t = 4$ $N_r = 4$

Fig. 10. $N_t = 2$ $N_r = 4$ **,** DilNet confusion matrix for the modulation classification. **(a)**
SNR $= -10$dB **(b)** SNR $= -6$dB
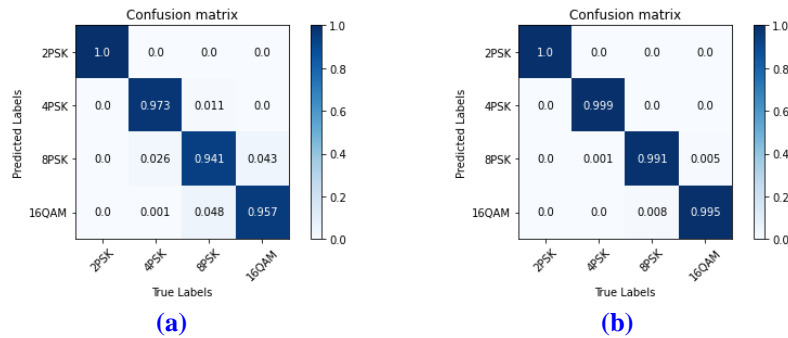
Fig. 11. $N_t = 4$ $N_r = 4$ **,** DilNet confusion matrix for the modulation classification. **(a)** SNR $= -2$dB
(b) SNR $= 2$dB

In order to observe the effect of DilNet on each signal classification, **Fig. 10** and **Fig. 11** show experimental results through confusion matrix. By observing **Fig. 10 (a)** and **Fig. 11 (a)** that 8PSK and 16QAM signals have lower classification accuracy. There are many types of phases of these two signals, and the phase fading is severe in the MIMO channel transmission process. The phase characteristics are unclear and difficult to distinguish. By observing **Fig. 10 (b)** and **Fig. 11 (b)** that when the SNR is greater than -6dB and -2dB, all signals have classification accuracy greater than 99%.

4.3 Performance of DilNet in Mixed SNR and Small Sample Datasets

We compare the performance of some classical modulation recognition methods in the absence of SNR priors, where SNR $\in [-10, 10]$ dB and step size 2 takes 11 SNRs in this paper. We mix all SNR samples for training, the average recognition rate P_{cc}^{ave} can be calculated by (18). P_{cc}^{snr} is the performance of the mixed SNR model at each SNR. The training dataset in the mixed SNR dataset is 20% of the total sample dataset. The ratio of training set, test set, and validation set is 4:1:4. The performance of different typical methods is shown in **Table 6** and **Table 7**.

Table 6. $N_t = 2$ $N_r = 4$, the performance of different typical methods in mixed SNR

Method	P_{cc}^{-10}	P_{cc}^{-8}	P_{cc}^{-6}	P_{cc}^{-4}	P_{cc}^{-2}	P_{cc}^{0}	P_{cc}^{2}	P_{cc}^{4}	P_{cc}^{6}	P_{cc}^{8}	P_{cc}^{10}	P_{cc}^{ave}
Proposed	**0.897**	**0.968**	**0.989**	0.998	0.998	**1**	1	1	1	1	1	**0.986**
Resnet10	0.863	0.967	0.987	**0.999**	**0.999**	0.999	1	1	1	1	1	0.983
MobilenetV2	0.736	0.892	0.973	0.995	0.998	0.998	1	1	1	1	1	0.963
CNN	0.849	0.914	0.978	0.997	0.998	0.999	1	1	1	1	1	0.976
CO-Proposed	0.728	0.806	0.909	0.925	0.994	0.999	0.999	1	1	1	1	0.942
Traditional	0.210	0.245	0.30	0.450	0.585	0.700	0.965	0.950	0.975	0.985	0.995	0.669

Table 7. $N_t = 4$ $N_r = 4$, the performance of different typical methods in mixed SNR

Method	P_{cc}^{-10}	P_{cc}^{-8}	P_{cc}^{-6}	P_{cc}^{-4}	P_{cc}^{-2}	P_{cc}^{0}	P_{cc}^{2}	P_{cc}^{4}	P_{cc}^{6}	P_{cc}^{8}	P_{cc}^{10}	P_{cc}^{ave}
Proposed	0.607	0.764	0.785	0.907	**0.926**	**0.941**	0.979	**0.994**	**0.998**	**0.998**	**0.999**	**0.900**
Resnet10	**0.608**	**0.774**	**0.793**	**0.916**	0.907	0.912	**0.987**	0.991	0.994	0.994	0.998	0.897
MobilenetV2	0.460	0.669	0.718	0.831	0.782	0.804	0.881	0.926	0.955	0.983	0.988	0.818
CNN	0.578	0.718	0.76	0.862	0.844	0.863	0.963	0.969	0.979	0.987	0.995	0.865
CO-Proposed	0.476	0.578	0.622	0.653	0.692	0.767	0.778	0.777	0.842	0.934	0.975	0.736
Traditional	0.265	0.265	0.235	0.345	0.265	0.36	0.545	0.545	0.525	0.496	0.615	0.406

In **Table 6** and **Table 7**, the experiment results show that the DilNet has the highest classification accuracy. When $SNR \in [-8, -2]$ dB, the traditional and MobilenetV2 methods show the phenomenon of the accuracy decreases. Because the traditional method can hardly meet the requirements of MIMO system signal recognition through traditional feature classification. The MobilenetV2 method only seeks to reduce the complexity of the network, which leads to a decrease in the ability of network feature extraction. Therefore, there is no guarantee that the classification accuracy will increase with the increase of SNR. While reducing the network complexity, DilNet ensures the efficiency of feature extraction through the TDIA module and penalized statistical entropy. DilNet has good performance in the absence of SNR priors. The average recognition rate of DilNet is higher than Resnet10, but the recognition rate of some SNRs is lower than Resnet10. Because reducing the complexity of the model through DSC will also affect the model classification ability. Besides, the performance of CO-Proposed in mixed SNR also proves that the method of cooperative feature classification is superior to cooperative decision.

Table 8 and **Table 9** exhibit the recognition accuracy of different methods in the small sample dataset. When the training dataset is larger than 3% of the sample dataset, DilNet has the highest recognition rate. However, the recognition rate of DilNet is lower than that of Resnet10, when the training dataset is less than 3% of the sample dataset. In the process of reducing the weight of the network, a part of the classification accuracy needs to be sacrificed. When the samples are sufficient, the classification accuracy can be compensated by efficient supervised training methods and reasonable attention module assistance. When the number of samples is small, the disadvantage of DilNet is difficult to eliminate. Simulation experiments show that when the number of samples in the dataset is larger than 5%, DilNet can guarantee classification accuracy.

Table 8. $N_t = 2$ $N_r = 4$, the performance of different typical methods in small sample datasets

Sample size	3%	5%	10%	20%
Proposed	0.937	**0.962**	**0.974**	**0.986**
Resnet10	**0.952**	0.960	0.967	0.983
MobilenetV2	0.923	0.928	0.943	0.963
CNN	0.932	0.954	0.961	0.976
CO- Proposed	0.898	0.915	0.931	0.942
Traditional	0.379	0.418	0.602	0.669

Table 9. $N_t = 4$ $N_r = 4$, the performance of different typical methods in small sample datasets

Sample size	3%	5%	10%	20%
Proposed	0.788	**0.837**	**0.861**	**0.900**
Resnet10	**0.813**	0.821	0.859	0.897
MobilenetV2	0.692	0.735	0.783	0.818
CNN	0.752	0.812	0.849	0.865
CO- Proposed	0.657	0.689	0.702	0.736
Traditional	0.309	0.327	0.375	0.406

4.4 Performance of DilNet on Computational Complexity and Convergence

Table 10 shows the computational complexity of different methods. Parameters is model parameter size, and FLOPs are the amount of computation required by the model to classify a single sample. The model train and test times are the average training and testing times per epoch. The number of transmitting antennas does not affect the analysis of the complexity regular, so this paper chooses to analyze the case of $N_t = 4$ $N_r = 4$.

Table 10. Computational complexity performance

Method	Train(s)	Test(s)	Parameters(MB)	FLOPs(MB)
CPAM	36.8	2.7	3.11	144.64
CBAM	23.1	1.8	1.78	54.08
TDIA	**16.9**	**1.5**	**1.48**	**46.72**
Resnet10	85.3	5.8	28.49	1071.34
CNN	**14.7**	**1.3**	6.02	301.44
MobilenetV2	36.9	2.5	4.52	428.48
Proposed	18.1	1.6	**1.49**	**47.04**

TDIA completes the interaction of channel and spatial information in the channel dimension and effectively reduces the complexity of the attention module through dimensional reshaping and Conv1D. Compared with CBAM and CPAM, the time complexity and space complexity of TDIA are significantly reduced. The penalty statistical entropy training network is based on the output information of the fully connected layer for statistics of each hidden layer loss. Therefore, time and space complexity are slightly improved compared to cross-entropy. Besides, compared with Resnet10 with the same number of residual layers, our network reduces the computational time and space complexity by 78.78% and 94.75%. The network proposed in this paper also has obvious advantages compared to CNN and MobilenetV2. However, the train and test time is higher than that of the CNN network, the residual structure is usually an important reason for the increased time complexity.
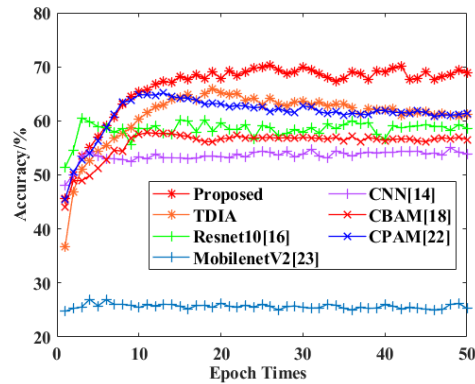
Fig. 12. The training accuracy of the different methods, where the number of training epoch times increases.

When the number of training epoch times increases, the training accuracy of models will be changed as shown in **Fig. 12**. Compared with CPAM, Resnet10, and CNN, the TDIA module requires more training epoch times to obtain the best training model. However, the TDIA module is prone to overfitting, resulting in a decrease in classification accuracy, when there are too many training epoch times. After adding penalized statistical entropy to our network, the problem of overfitting is solved, and the training accuracy is significantly improved. This experiment proves that the DilNet requires more training epoch times to obtain a better trained model. We will do further research on this issue in future work.

5. Conclusion

This paper proposes a dimensional interactive lightweight network for modulation recognition in MIMO systems. The lightweight network based on the TDIA module is used to extract cooperate features, and the network is trained by penalized statistical entropy. The experiment results demonstrate that the TDIA module has the best classification accuracy and the lowest computational complexity among three attention modules in the same network structure. When the lightweight network chooses the penalty statistical entropy as the loss function, training results are significantly improved on the original basis, with a slight extra computation needed. Additionally, the lightweight network also performs better than traditional HOC method, the cooperative modulation recognition method, and other neural networks in mixed SNR and small-sample datasets.

Acknowledgment

References

[1] Dhamyaa H. Al-Nuaimi, Ivan A. Hashim, Intan S. Zainal Abidin, Laith B. Salman, and Nor Ashidi Mat Isa, "Performance of Feature-Based Techniques for Automatic Digital Modulation Recognition and Classification—A Review," *electronics.*, vol. 8, no. 12, pp. 1407, Nov 2019. Article (CrossRef Link)

[2] C. Lin, W. Yan, L. Zhang, and Y. Wang, "An Overview of Communication Signals Modulation Recognition," *J. China. Academy. Electron. Inform. Technol.*, vol. 16, no. 11, pp. 1074-1085, Nov 2021. Article (CrossRef Link)

[3] J. He, and W. Zhang, "Communication Signal Modulation Recognition Technology and Its Development," *High. Technol. lett.,* vol. 26, no. 2, pp. 157-165, Feb 2016. Article (CrossRef Link)

[4] Y. Wang, J. Yang, M. Liu, and G. Gui, "LightAMC: Lightweight Automatic Modulation Classification Via Deep Learning and Compressive Sensing," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 3491-3495, March 2020. Article (CrossRef Link)

[5] J. B. Tamakuwala, "New Low Complexity Variance Method for Automatic Modulation Classification and Comparison with Maximum Likelihood Method," in *Proc. of 2019 International Conference on Range Technology (ICORT)*, Balasore, India, pp. 1-5, February. 2019. Article (CrossRef Link)

[6] M. Turan, M. Öner, and H. A. Çırpan, "Joint Modulation Classification and Antenna Number Detection for MIMO Systems," *IEEE Commun. lett.*, vol. 20, no. 1, pp. 193-196, Jan. 2016. Article (CrossRef Link)

[7] X. Zhao, C. Guo, and J. Li, "Mixed Recognition Algorithm for Signal Modulation Schemes by High-order Cumulants and Cyclic Spectrum," (in Chinese), *J. Electron. Inform. Technol.*, vol. 38, no. 3, pp. 674–680, Mar 2016. Article (CrossRef Link)

[8] C. Wu, and G. Feng, "New Automatic Modulation Classifier Using Cyclic-spectrum Graphs with Optimal Training Features," *IEEE Commun. lett.*, vol. 22, no. 6, pp. 1204-1207, June 2018. Article (CrossRef Link)

[9] X. Yan, G. Liu, H. Wu, G. Zhang, Q. Wang, and Y. Wu, "Robust Modulation Classification Over α-Stable Noise Using Graph-Based Fractional Lower-Order Cyclic Spectrum Analysis," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 2836-2849, January. 2020. Article (CrossRef Link)

[10] T. V. R. O. Câmara, A. D. L. Lima, B. M. M. Lima, A. I. R. Fontes, A. D. M. Martins, and L. F. Q. Silveira, "Automatic Modulation Classification Architectures Based on Cyclostationary Features in Impulsive Environments," *IEEE Access,* vol. 7, pp. 138512-138527, Sep. 2019. Article (CrossRef Link)

[11] H. Gao, S. Wang, Y. Su, H. Sun, and Z. Zhang, "Evolutionary Neural Network based on Quantum Elephant Herding Algorithm for Modulation Recognition in Impulse Noise," *KSII Trans. Internet. Inform. Systems.*, vol. 15, pp. 2356-2376, July. 2021. Article (CrossRef Link)

[12] X. Zhang, J. Sun, and X. Zhang, "Automatic Modulation Classification Based on Novel Feature Extraction Algorithms," *IEEE Access,* vol. 8, pp. 16362-16371, 2020. Article (CrossRef Link)

[13] D. Das, P. K. Bora, and R. Bhattacharjee, "Blind Modulation Recognition of the Lower Order PSK Signals under the MIMO Keyhole Channel," *IEEE Commun. lett.*, vol. 22, no. 9, pp. 1834-1837, Sept. 2018. Article (CrossRef Link)

[14] T. O'Shea, and J. Hoydis, "An Introduction to Deep Learning for The Physical Layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563-575, Dec. 2017. Article (CrossRef Link)

[15] S. Hu, Y. Pei, P. P. Liang, and Y. Liang, "Deep Neural Network for Robust Modulation Classification under Uncertain Noise Conditions," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 564-577, Jan. 2020. Article (CrossRef Link)

[16] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air Deep Learning Based Radio Signal Classification," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 168-179, Feb. 2018. Article (CrossRef Link)

[17] Y. Wang, G. Gui, T. Ohtsuki, and F. Adachi, "Multi-task Learning for Generalized Automatic Modulation Classification under Non-gaussian Noise with Varying SNR Conditions," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3587-3596, June 2021. Article (CrossRef Link)

[18] Z. Liang, M. Tao, L. Wang, J. Su, and X. Yang, "Automatic Modulation Recognition Based on Adaptive Attention Mechanism and ResNeXt WSL Model," *IEEE Commun. lett.*, vol. 25 pp. 2953-2957, Sept 2021. Article (CrossRef Link)

[19] Y. Mao, Y. -Y. Dong, T. Sun, X. Rao, and C. -X. Dong, "Attentive Siamese Networks for Automatic Modulation Classification Based on Multitiming Constellation Diagrams," *IEEE Trans. Neural Netw. Learn Systems.*, pp. 1-15, 2021. Article (CrossRef Link)

[20] Y. Wang, J. Wang, W. Zhang, J. Yang, and G. Gui, "Deep Learning-based Cooperative Automatic Modulation Classification Method for MIMO Systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4575-4579, April 2020. Article (CrossRef Link)

[21] Y. Wang, J. Gui, Y. Yin, J. Wang, and G. Gui, "Automatic Modulation Classification for MIMO Systems via Deep Learning and Zero-Forcing Equalization," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 5688-5692, May 2020. Article (CrossRef Link)

[22] J. Fu, J. Liu, and H. Tian, "Dual Attention Network for Scene Segmentation," in *Proc. of IEEE/CVF. Conf. Computer Vision. Pattern Recognition*, pp. 3141-3149, June 2019. Article (CrossRef Link)

[23] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. -C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *Proc. of IEEE/CVF. Conf. Computer Vision. Pattern Recognition*, pp. 4510-4520, June 2018. Article (CrossRef Link)

[24] S. Kalluri and G. R. Arce, "Adaptive Weighted Myriad Filter Algorithms for Robust Signal Processing in /Spl Alpha/-Stable Noise Environments," *IEEE Trans Signal Process.*, vol. 46, no. 2, pp. 322-334, Feb. 1998. Article (CrossRef Link)

[25] W. Shi, D. Liu, X. Cheng, Y. Li, and Y. Zhao, "Particle Swarm Optimization-Based Deep Neural Network for Digital Modulation Recognition," *IEEE Access*, vol. 7, pp. 104591-104600, 2019. Article (CrossRef Link)

[26] C. Zhang, S. Yu, G. Li, and Y. Xu, "The Recognition Method of MQAM Signals Based on BP Neural Network and Bird Swarm Algorithm," *IEEE Access*, vol. 9, pp. 36078-36086, 2021. Article (CrossRef Link)

[27] K. Bu, Y. He, X. Jing, and J. Han, "Adversarial Transfer Learning for Deep Learning Based Automatic Modulation Classification," *IEEE Signal Process Lett.*, vol. 27, pp. 880-884, 2020. Article (CrossRef Link)

[28] C. -F. Teng, C. -Y. Chou, C. -H. Chen, and A. -Y. Wu, "Accumulated Polar Feature-Based Deep Learning for Efficient and Lightweight Automatic Modulation Classification with Channel Compensation Mechanism," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 15472-15485, Dec. 2020. Article (CrossRef Link)

[29] L. Li, J. Huang, Q. Cheng, H. Meng, and Z. Han, "Automatic Modulation Recognition: A Few-Shot Learning Method Based on the Capsule Network," *IEEE Wireless Commun Lette.*, vol. 10, no. 3, pp. 474-477, March 2021. Article (CrossRef Link)

Aersileng was born in 1998. He received his B.E. degree from the Taiyuan University of Technology, Taiyuan, Shanxi, P. R. China, in June 2020. He is currently a postgraduate student in Harbin Engineering University. His current research interests include Deep learning, intelligent computing, modulation classification, and signal processing in non-cooperative wireless communication environment.

Xiaolin Zhang was born in 1971. He graduated from the College of Information and Communication Engineering, in 1995, and the Ph.D. degree from Harbin Engineering University, Harbin, China, in 2007. He is currently an Associate Professor. He has successively presided over complete Postdoctoral Fund in China and Postdoctoral Fund in Heilongjiang Province, as the technical director successively participated in national, provincial, and lateral projects. His research interests are communication signal detection and recognition, and broadband digital communication systems. He is a member of the China Communication Association.

Zhenduo Wang (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in information and communication engineering from the Harbin Institute of Technology, in 2012, 2014, and 2019, respectively. From September 2017 to September 2018, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC, Canada. He is currently an Associate Professor with the College of Information and Communication Engineering, Harbin Engineering University, Harbin, China. He has authored or coauthored more than 20 research articles and patents. His research interests include 5G waveforms and transform domain communication systems.

Kailin Wang was born in 1997. He received his B.E. degree from the Guilin University Of Electronic Technology, Guilin, Guangxi, P. R. China, in July 2020. He is currently a postgraduate student in Harbin Engineering University. His current research interests include Cognitive Radio Spectrum sensing digital signal processing in Alpha Stable noise background environment.