electronics

MDPI

Article

A Multi-Modal Modulation Recognition Method with SNR Segmentation Based on Time Domain Signals and Constellation Diagrams

Ruifeng Duan [1,2], Xinze Li [1,2], Haiyan Zhang [1,2,*], Guoting Yang [1,2], Shurui Li [3], Peng Cheng [4,5] and Yonghui Li [5]

[1] School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China; drffighting2019@bjfu.edu.cn (R.D.); lixinze0322@bjfu.edu.cn (X.L.); kitoygt2020@bjfu.edu.cn (G.Y.)
[2] Engineering Research Center for Forestry-Oriented Intelligent Information Processing of National Forestry and Grassland Administration, Beijing 100083, China
[3] School of Technology, Beijing Forestry University, Beijing 100083, China; lg13104660772@bjfu.edu.cn
[4] Department of Computer Science and Information Technology, La Trobe University, Melbourne, VIC 3086, Australia; peng.cheng@sydney.edu.au
[5] School of Electrical and Information Engineering, The University of Sydney, Sydney, NSW 2006, Australia; yonghui.li@sydney.edu.au
* Correspondence: zhyzml@bjfu.edu.cn

Abstract: Deep-learning-based automatic modulation recognition (AMR) has recently attracted significant interest due to its high recognition accuracy and the lack of a need to manually set classification standards. However, it is extremely challenging to achieve a high recognition accuracy in increasingly complex channel environments and balance the complexity. To address this issue, we propose a multi-modal AMR neural network model with SNR segmentation called M-LSCANet, which integrates an SNR segmentation strategy, lightweight residual stacks, skip connections, and an attention mechanism. In the proposed model, we use time domain I/Q data and constellation diagram data only in medium and high signal-to-noise (SNR) regions to jointly extract the signal features. But for the low SNR region, only I/Q signals are used. This is because constellation diagrams are very recognizable in the medium and high SNRs, which is conducive to distinguishing high-order modulation. However, in the low SNR region, excessive similarity and the blurring of constellations caused by heavy noise will seriously interfere with modulation recognition, resulting in performance loss. Remarkably, the proposed method uses lightweight residuals stacks and rich ski connections, so that more initial information is retained to learn the constellation diagram feature information and extract the time domain features from shallow to deep, but with a moderate complexity. Additionally, after feature fusion, we adopt the convolution block attention module (CBAM) to reweigh both the channel and spatial domains, further improving the model's ability to mine signal characteristics. As a result, the proposed approach significantly improves the overall recognition accuracy. The experimental results on the RadioML 2016.10B public dataset, with SNR ranging from −20 dB to 18 dB, show that the proposed M-LSCANet outperforms existing methods in terms of classification accuracy, achieving 93.4% and 95.8% at 0 dB and 12 dB, respectively, which are improvements of 2.7% and 2.0% compared to TMRN-GLU. Moreover, the proposed model exhibits a moderate parameter number compared to state-of-the-art methods.

Keywords: multi-modal; automatic modulation recognition; deep learning; constellation; lightweight

Citation: Duan, R.; Li, X.; Zhang, H.; Yang, G.; Li, S.; Cheng, P.; Li, Y. A Multi-Modal Modulation Recognition Method with SNR Segmentation Based on Time Domain Signals and Constellation Diagrams. *Electronics* **2023**, *12*, 3175. https://doi.org/10.3390/electronics12143175

Academic Editor: Yide Wang

Received: 6 June 2023
Revised: 9 July 2023
Accepted: 18 July 2023
Published: 21 July 2023

1. Introduction

With the rapid development of mobile communication and Internet of things (IoT) technology, spectrum resources are becoming increasingly scarce. Cognitive radio (CR) technology is often used to increase spectral usage efficiency, where signal detection and

automatic modulation recognition (AMR) serve as the foundation for spectrum sensing, spectrum understanding, and signal demodulation. After accurate signal classification and modulation recognition, secondary users (SUs) and primary users (PUs) can efficiently share the spectrum, thereby significantly improving the spectrum utilization [1]. In non-cooperative communication scenarios, AMR is also essential for signal interception or implementing interference [2,3]. Consequently, AMR has been a research focus in wireless communication, with widespread applications in many communication systems, including satellite communication systems [4].

AMR [5] is crucial for achieving dynamic spectrum allocation (DSA) and spectrum sharing (SS) by automatically estimating the modulation mode and other parameters from the received signals [6]. Traditional AMR methods are generally categorized into two types: likelihood-based (LB) methods [7,8] and feature-based (FB) methods [9]. The LB approach treats AMR as a multiple composite hypothesis-testing problem, and it is optimal in the Bayesian sense [10]. However, LB methods rely heavily on prior knowledge and also feature a high computational complexity and poor robustness. FB methods classify different modulation signals according to their statistical features, such as instantaneous statistics features, higher-order cumulants, various cyclic spectral features, and so on [11–16]. They are more practical, since no prior knowledge is required. However, FB methods are suboptimal and usually designed for certain modulation patterns, lacking strong generalization.

Due to the high complexity and limited generality of traditional techniques, researchers have explored new modulation recognition methods using machine learning algorithms such as decision trees (DT) [17], support vector machines (SVM) [18], and KNearest Neighbors (KNN) [13]. Machine learning is a branch of artificial intelligence that learns the intrinsic characteristics from data and makes predictions accordingly. It has been widely applied in lots of fields, such as image and speech recognition, natural language processing, predictive modeling, and so on. By using machine learning, researchers and practitioners can uncover hidden patterns and develop more efficient ways to solve complex problems. While these algorithms have achieved good performances, they have not significantly outperformed the traditional methods, particularly in terms of their generalization performance. Owing to the considerable growth of computing power, the emergence of large-scale datasets and advances in algorithm design, deep learning, as a branch of machine learning, can rapidly solve some complex and abstract intelligent tasks [19] and has achieved great success in many fields, such as natural language processing [20], computer vision [21], and recommendation systems [22], in recent years. Deep learning is able to automatically mine the potential features of data and classify them accordingly, without manually designing classification criteria based on the statistical characteristics of the data. Therefore, it is more adaptable to modern AMR tasks.

Recently, deep learning has obtained some achievements in AMR. O'Shea et al. [23] introduced convolutional neural networks (CNN) to classify modulation modes for the first time and proposed a four-layer CNN model. The experiment results showed that the shallow four-layer CNN not only performed better in terms of accuracy than traditional techniques, but also had significant flexibility in distinguishing different modulation modes. However, it is often difficult for a CNN to capture temporal information, which exists widely in modulation signals. As a result, AMR methods based on recurrent neural networks (RNN) were proposed in Ref. [24], aiming to take full advantage of RNNs' strong extraction ability for both the temporal and semantic information of modulation signals. As an improvement of RNN, long short-term memory networks (LSTM) [25] and bi-directional long short-term memory (BiLSTM) [26] are also used for modulation recognition, and improve the recognition accuracy compared to a sole CNN. However, limited to a simple network construction, the above-mentioned RNNs and CNN cannot achieve a satisfactory recognition accuracy.

In fact, increasing the number of CNN layers is beneficial to improving the network performance, but excessive CNN layers will cause gradient vanishing or an explosion problem, resulting in a significant performance loss. Introducing skip-connections between

these different layers can solve this problem well. The Residual Network (ResNet) including skip connections was also proposed by O'Shea [27] in 2018 to classify 24 kinds of modulation signals and achieve a high recognition accuracy. Huynh-The et al. [28] proposed the MCNet, which uses skip connections to establish three specific convolutional blocks consisting of different asymmetric convolutional kernels, in order to learn the spatial–temporal correlation of modulation signals and classify them accordingly. Similar network structures include MBNet [29], SCGNet [30], Ref-Net [31], and Chain-Net [32], etc. All these networks use an asymmetric convolutional structure to achieve a high recognition accuracy and keep a relatively low complexity. Nevertheless, the number and arrangement of convolutional kernels can be further optimized to improve the recognition performance. Wang Y [33] et al. proposed a lightweight network, named LightAMC, to implement modulation recognition in UAV-assisted systems. The LightAMC assigns each neuron a scaling factor, then selects those redundant neurons with small scaling factors, and removes them from the network by using pruning techniques. As a result, the LightAMC has a lower computational cost with more than an 80% reduction in the number of parameters compared to the original model. However, it ignores the information loss caused by convolution and down-sampling. Thus, its recognition performance cannot meet the requirements of high-accuracy applications.

The mixed network architecture is a very effective way of improving recognition performance, since it can make full use of the strengths of each type of network in signal feature extraction. Consequently, many works have recently explored the application of mixed networks in the field of AMR. Njoku et al. [34] constructed a hybrid neural network model consisting of a CNN, gate recurrent unit (GRU), and fully connected neural network to fully extract the feature information of modulation signals and achieved a good recognition performance. The model uses a modified GRU, which has fewer parameters than long and short-term memory gates (LSTM) and reduces the complexity. Moreover, it solves the problems of gradient vanishing by using rich skip connections in the CNN network. Yang et al. [35] proposed a novel dual-path modulation recognition network called IRLNet, which consists of an improved residual stack (IRS) and LSTM. The model first uses the IRS module and LSTM module to extract the feature information of modulation signals, and then uses the fully connected layer to classify them. On the RadioML 2016.10B dataset, its average recognition accuracy achieves 93% in SNR, ranging from 0 dB to 18 dB. Zheng Y [36], etc., proposed a transformer-based automatic classification recognition network improved by a Gate Linear Unit (TMRN-GLU), which combines the advantages of CNNs with a high efficiency of parallel operations and RNNs with a sufficient extraction of the global information of the temporal signal context. Its highest recognition accuracy is about 94% on the RadioML 2016.10B dataset when SNR varies from 0 dB to 18 dB. With the development of the transformer network structure, an efficient and robust pyramid Transformer AMR model, referred to as SigFormer-Net [37], uses a dual-attention block with parallel self-attention and scaling-attention layers for simultaneous global feature extraction and noise resistance learning. Its highest recognition accuracy is about 94.8%. However, the pyramid signal transformation is computationally intensive. Meanwhile, as with most networks, this method has a poor recognition performance in low SNR regions, which are typical in some complex communication scenarios. In order to improve its recognition accuracy in the low SNR range, An et al. [38] proposed a threshold autoencoder denoiser convolutional neural network (TADCNN), which consists of a threshold autoencoder denoiser (TAD) and a CNN. This method improved the classification accuracy by 70% in the low SNR range compared to a model without a denoiser. Nevertheless, it does not perform particularly well in the high SNR region, possibly because TADs bring damage to the modulation signals of high SNR.

The above-mentioned methods all use time domain I/Q signals to perform modulation recognition, which are not ideal for high-order modulation recognition. In fact, other forms of modulation signals also contain rich characteristic information. Inspired by the great success of deep learning in the field of computer vision, some researchers have used image data as AMR network inputs, such as constellation diagrams [39], spectrum diagrams [40],

signal eye diagrams, and vector diagrams [41], etc. Such methods directly utilize excellent network models in the field of computer vision, but there are still some problems, such as a high computational complexity, complex data conversion from time series signals to images, and a low recognition accuracy due to mismatches between the model and the data. Huang et al. [39] proposed an AMR method based on the network constellation matrix, which employs low-complexity preprocessing to obtain the constellation diagram and feeds it into the network. However, the classification accuracy for certain modulation methods is suboptimal. In Ref. [42], Y. Wang et al. adopted two cascaded CNN models, referred to as CNN1 and CNN2, respectively, to hierarchically identify eight modulation modes. After the preliminary modulation recognition in CNN1 using I/Q signals, the authors used constellation diagrams as the CNN2 input to further identify QAM16 and QAM64 and obtain a good performance.

By utilizing multiple forms of modulation signals in a joint manner, it is possible to more effectively extract key signal features, thereby enhancing the overall recognition performance. Han [43] et al. proposed an AF-RDGNN to establish and enhance the constellation diagram for AMR, then they [44] further proposed a double-branches-based AMR method that integrates I/Q signals and constellation diagram data. However, branch models and fusion methods are not well designed, and the computation of three-channel image data increases the complexity of the model. More importantly, this method does not consider the characteristics of constellations in different SNR regions, and the dual branch processing in the whole SNR range may deteriorate the recognition performance of low SNR regions. Consequently, the model structure and using the strategy of multi-modal data need to be further optimized.

Based on the above discussion, existing modulation recognition methods still exhibit some shortcomings, such as a low recognition accuracy, particularly for high-order modulation, or a high complexity. Although some methods have combined time domain I/Q signals and image data for modulation recognition, feature fusion has not been optimized, making it challenging to achieve an improved recognition performance. This paper proposes a multi-modal network architecture with SNR segmentation based on time-domain signals and constellation diagrams (M-LSCANet). The proposed modulation recognition method intelligently selects whether to enable the constellation processing branch according to the SNR values to achieve a high recognition accuracy in the whole SNR range and maintain a moderate complexity. The main contributions of this paper are summarized as follows.

(1) We propose a lightweight AMR neural network model, named M-LSCANet, that utilizes a multi-modal mechanism. The model integrates two distinct networks to extract the time-domain and constellation diagram features, allowing for more comprehensive mining of the signal phase-amplitude information. In particular, incorporating constellation features significantly improves the model's recognition performance in medium and high SNR regions. We also optimize the number of samples contained in each constellation diagram to ensure an optimal recognition performance. Additionally, sample outliers are eliminated during the training process to further improve this recognition performance.

(2) In the feature extraction stage, the incorporation of rich skip-connection structures and residual stacks with small convolutional kernels helps to mine the shallow and deep features of the signals simultaneously, thereby enhancing the model's feature extraction capabilities while reducing the network complexity. In the feature fusion stage, the CBAM module is utilized to accurately capture dependent information by recalibrating the channel and spatial weights. This enables the model to solve complex problems by selectively focusing on the most informative features at each layer. As a result, the CBAM module significantly enhances the identification ability of the model.

(3) Considering the blurring of constellation diagrams in low SNR regions, we propose an SNR segmentation strategy to further enhance the network recognition ability in

the low SNR range. Specifically, we jointly apply the time domain I/Q branch and constellation branch for modulation recognition when the SNR is above a certain threshold, but only use the I/Q branch when the SNR is low. This strategy yields a better performance across a wide SNR region and dramatically reduces the model complexity in the low SNR region. We choose the RadioML 2016.10B dataset as the benchmark for the simulation experiments, which is widely used for AMR tasks. The proposed method outperforms all existing AMR neural network models in terms of its classification accuracy over the entire SNR range from −20 dB to +18 dB.

This paper is organized as follows. Section 2 describes the basic signal system model. Section 3 presents the network architecture and principles of the proposed M-LSCANet and illustrates more details of its key modules. The simulation results and performance analysis are summarized in Section 4. Finally, Section 5 concludes the paper.

2. Signal Model

In this paper, we focus on cognitive radio and wireless communication systems, where the channel is complex and the modulation types are varied and unknown to the receiver. Therefore, the receiver must first perform modulation identification before demodulation and decoding.

A schematic diagram of a typical wireless communication system is shown in Figure 1. Analog or digital signals are sent to a modulator after source coding, encryption, and channel coding. The information bits generated by the previous stage are mapped into different kinds of modulation symbols in the modulator and then fed into a radio frequency (RF)-processing module. At the receiver, the signal from the receiving antenna is converted into a baseband signal after the RF processing. Then, the AMR module is used to classify the modulation type of the baseband signal, followed by demodulation and decoding.

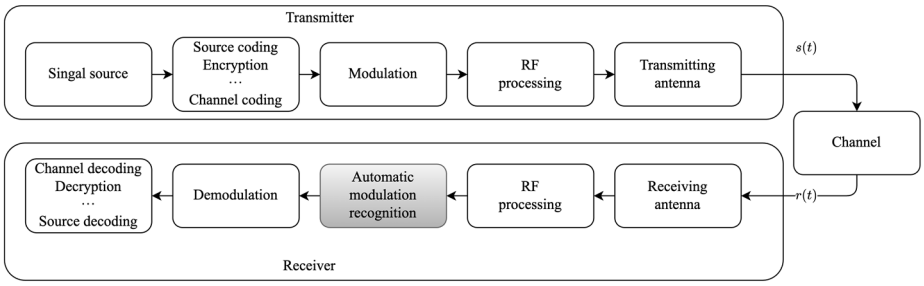


Figure 1. Schematic diagram of wireless communication system.

The received signal $r(t)$ can be expressed as

$$r(t) = s(t)^* c(t) + n(t) \tag{1}$$

where $s(t)$ is the modulation signal at time t, $c(t)$ is the impulse response of the wireless channel, and $n(t)$ is the additive white Gaussian noise (AWGN) with zero mean. When the modulation mode is different, the transmitted signal has different expressions. If the signal is modulated by amplitude shift keying (ASK), frequency shift keying (FSK), or phase shift keying (PSK), $s(t)$ is expressed as follows.

$$s(t) = \left[A_m \sum a_n g(t\text{-}nT)\right] \cos(2\pi(f_c + f_m)t + \phi_0 + \phi_m) \tag{2}$$

where A_m is the modulation amplitude, a_n is the symbol sequence, $g(\cdot)$ is the signal pulse, T is the symbol period, f_m is the modulation frequency, and f_c is the carrier frequency. The parameter ϕ_0 represents the initial phase and ϕ_m represents the modulation phase. In addition, quadrature amplitude modulation (QAM) is different from the modulation

mode mentioned above. This modulation mode has two quadrature carriers, the signal is modulated by a_n and b_n, and $s(t)$ is expressed as follows.

$$s(t) = \left[A_m \sum_n a_n g(t\text{-}nT)\right] \cos(2\pi f_c t + \phi_0) + \left[A_m \sum_n b_n g(t\text{-}nT)\right] \sin(2\pi f_c t + \phi_0). \quad (3)$$

At the receiver, $r(t)$ is expressed in I/Q format and sampled k times by the AD converter at a rate of $f_s = 1/T_s$. The real and imaginary parts of $r(t)$ represent the I and Q components, respectively. Specifically, $r(t)$ can be also expressed as

$$r(t) = \alpha(t) e^{j(2\pi f_0 t + \theta_0)} s(t) + n(t) \quad (4)$$

where $\alpha(t)$ represents the Rayleigh fading factor and f_0 and θ_0 represent the frequency and phase shifts introduced by different local oscillators and Doppler effects.

In this paper, we focus on automatic modulation recognition, corresponding to the shaded module in Figure 1. The received signals are complex with different modulation types, Rayleigh fading, noise, frequency offset, and phase offset. They do not go through carrier phase synchronization, channel equalization, and other preprocessing, but directly carry out modulation recognition. Based on the above analysis, the RadioML 2016B dataset generated using the GNU radio platform agrees well with the communication system model we describe. Consequently, we just adopt RadioML 2016B in the experiments, and the parameter definitions can be found in Ref. [23] in detail.

3. Proposed Network Model Structure

In this paper, the multi-modal data include the time domain I/Q components of signal and constellation diagrams. When the noise power is large enough, the difference between the constellations of different signals is not obvious, and the introduction and processing of a constellation will increase the complexity of the model but deteriorate its recognition performance. Based on the above analysis, a strategy of SNR segmentation processing is proposed. That is, only when the SNR is above a certain threshold is the multi-modal modulation identification performed, and when the SNR is low, only the branch of the I/Q components is enabled. In this way, the overall complexity of the training and validation can be greatly reduced and the recognition accuracy is optimal in the whole SNR range. The whole model structure includes a preprocessing module, SNR classification module, and modulation recognition module. The complete modulation recognition process is shown in Figure 2.

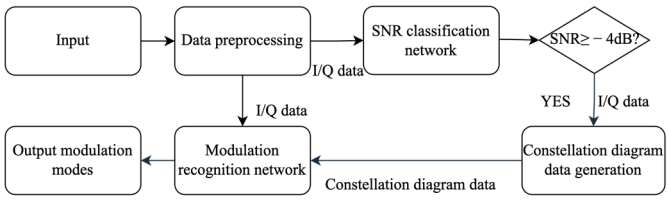


Figure 2. The complete process for automatic modulation recognition.

We directly adopt the SNR classification network, a simple two-layer CNN, proposed in Ref. [45], where the optimal SNR threshold of two-stage SNR classification is −4 dB for the same RadioML 2016B dataset. Through the classification network, we divide the dataset into medium and high SNR regions and the low SNR region. When modulation signals fall into the medium and high SNR regions, the I/Q data and constellation data are jointly used for modulation classification. Otherwise, we only use I/Q data to perform the modulation recognition.

The proposed modulation recognition network consists of the following parts: a time domain I/Q-signal-processing branch, constellation-diagram-processing branch, multi-modal fusion layer, and full connection layer. The whole structure is shown in Figure 3.

The time domain I/Q branch constructs three sets of convolutional blocks with different kernels to extract the signal features from shallow to deep based on the I/Q component. The constellation diagram branch extracts the signal's amplitude-phase information from the constellations using a residual convolution stack with fewer convolutional kernels.

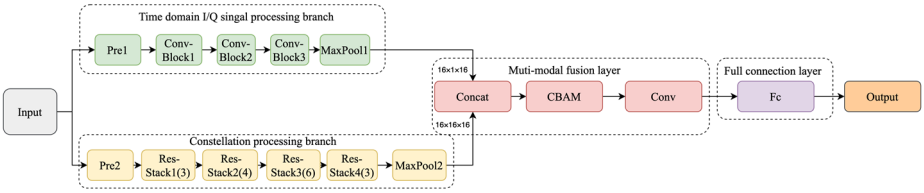


Figure 3. Description of the proposed multi-modal network architecture.

The multi-modal fusion layer realizes the feature fusion of two branches. It combines the features extracted from the time domain I/Q signals and the features extracted from the constellation diagrams. At the same time, the CBAM is also introduced to further explore the important information hidden in the space and channel, with a large weight to strengthen the important information and a small weight to ignore the irrelevant information. Moreover, the CBAM is able to constantly adjust the weights, thus improving the model's ability to deal with a complex problem. The full connection layer maps the outputs of multiple neurons to probabilities in [0,1] through the SoftMax activation function to achieve a multi-label classification. The detailed structure of each module is described below.

3.1. Time Domain I/Q-Signal-Processing Branch

The time domain I/Q-signal-processing branch extracts the signal's deep–shallow feature from the I/Q components. It uses a CNN architecture consisting of a preprocessing layer (Pre1), three blocks, and one pooling layer. The Pre1 is composed of two identical convolution-pooling layer combinations. The first convolutional layer processes the input I/Q data with the dimension of 2 × 1024, a convolutional kernel size of 3 × 3, and a padding size of (3, 1). To optimize the feature extraction, a pooling layer is introduced immediately after the convolutional layer to reduce the size of the feature map. The second convolution-pooling layer is the same as the first one. The feature maps output by the processing layer are sequentially transmitted to the first, second, and third convolution blocks, each of which consists of four convolution layers and four Batch norm layers. The structure of these convolution blocks is given in Figure 4. The details of each block are listed in Table 1 and the difference is the number of convolution kernels used. Each convolution block consists of two-dimension convolution (conv2d), two-dimension batch normalization (batchnorm2d), and activation function ReLU.

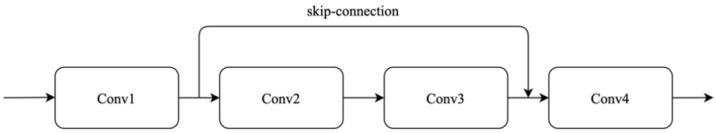


Figure 4. The architecture of Conv-Block1, Conv-Block2, and Conv-Block3.

Table 1. The details of each Conv-Block.

Conv-Name	Conv1	Conv2	Conv3	Conv4
Conv-Block1	Conv2d(32, 32, 1, 1, 0)	Conv2d(32, 8, 1, 1, 0)	Conv2d(8, 8, 1, 1, 0)	Conv2d(8, 32, 1, 1, 0)
Conv-Blcok2	Conv2d(32, 64, 1, 1, 0)	Conv2d(32, 16, 1, 1, 1)	Conv2d(16, 16, 1, 1, 1)	Conv2d(16, 64, 1, 1, 0)
Conv-Block3	Conv2d(64, 128, 1, 1, 0)	Conv2d(128, 32, 1, 1, 0)	Conv2d(32, 32, 1, 1, 0)	Conv2d(32, 16, 1, 1, 0)

Conv2d is defined as Conv2d(in_channels, out_channels, kernel_size, stride, padding).

Both the Conv1 and Conv4 in Figure 4 use a 1×1 kernel matrix to reduce the dimensionality of the feature-mapping channel. In addition, the Conv2 and Conv3 use asymmetric kernel matrices with kernel dimensions of 3×1 and 1×3, respectively, instead of 3×3; thus, the number of trainable parameters further decreases. The pooling layer behind Conv-block3 uses several 3×16 convolutional kernels to match the subsequent feature fusion operations.

3.2. Constellation-Diagram-Processing Branch

A ResNet architecture is used to extract the constellation diagram features. The constellation-diagram-processing branch consists of a preprocessing layer (Pre2) and four residual stacks, denoted as Res-stack1, Res-stack2, Res-stack3, and Res-stack4, in Figure 3. The Pre2 is similar to the Pre1 in the I/Q signal branch, and it includes one Conv2d with convolutional kernels of 3×3 and padding of 3×1, and one pooling layer. The Pre2 adjusts the dimensions of the constellation diagram to a resolution of 256 by 256, so that it fits into the residual stacks later. Each residual stack is composed of 3 to 6 cascaded ResNet modules and the number of modules in the four Res-stacks is 3, 4, 6, and 3, respectively. The single ResNet module is depicted in Figure 5. It needs to be noticed that the four Res-stacks use different numbers of convolution kernels to establish their ResNet modules. The number of convolution kernels is 16, 32, 32, and 16, respectively. The corresponding Res-stack1 consisting of three cascaded ResNet modules is shown in Figure 6. Therefore, a total of 16 ResNet modules with four different numbers of convolution kernels are stacked together to form a large residual network architecture to extract the feature extraction of the constellations.

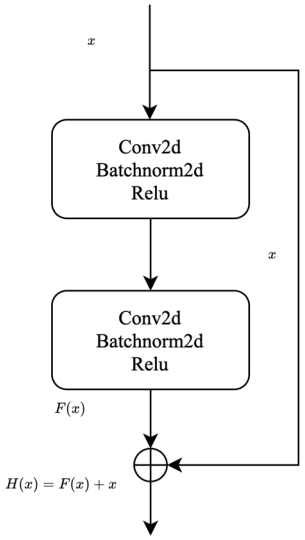


Figure 5. The architecture of ResNet module.

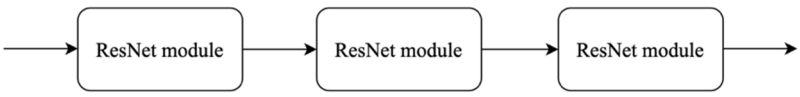


Figure 6. The structure of Res-stack1.

The diversified Res-stacks are able to extract multi-scale feature information to improve the recognition performance. Moreover, a large number of ResNet modules not only retains the strong feature extraction capability of a deep network, but also effectively alleviates gradient vanishing and explosion. All of these are beneficial for achieving a high recognition accuracy. Furthermore, four Res-stacks are obtained by modifying the original ResNet34, which reduces the number of convolutional kernels, thus guaranteeing a low complexity.

3.3. Multi-Modal Fusion Layer

This multi-modal fusion layer integrates the feature maps of the I/Q-signal-processing branch and constellation-diagram-processing branch to enhance the recognition performance. Firstly, this layer uses the concatenation operation [46] to fuse the feature information of the two branches. The constellation diagram expresses the static characteristics of the signal in the form of an image, and the time domain I/Q data display the dynamic characteristics of the signal amplitude over time. Therefore, after the front two branches' processing separately, the semantics of the feature information at the corresponding channels are different. In other words, the feature information extracted by the two branches is rich and diversified, and the concatenation operation can fully fuse the feature information from the time domains and constellations.

Although the feature maps after concatenation contain richer information, the contribution of the feature information of different channels and spaces to the modulation recognition is different. We introduce the CBAM [47] to improve the performance of the model, and the structure of CBAM is shown in Figure 7.

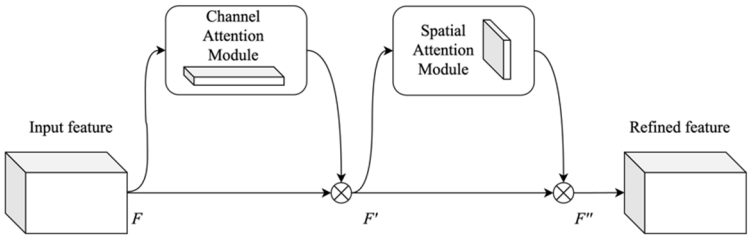


Figure 7. The architecture of CBAM.

Given an intermediate feature map $F \in \mathrm{R}^{\mathrm{C}\times\mathrm{H}\times\mathrm{W}}$ as input, the CBAM sequentially infers a 1D channel attention map $Mc \in \mathrm{R}^{\mathrm{C}\times1\times1}$ and 2D spatial attention map $Ms \in \mathrm{R}^{1\times\mathrm{H}\times\mathrm{W}}$.

The overall attention process can be summarized as

$$\begin{aligned} F' &= Mc(F) \otimes F, \\ F'' &= Ms(F') \otimes F\prime, \end{aligned} \tag{5}$$

where $\otimes$ denotes element-wise multiplication. During this multiplication, the attention values are broadcasted accordingly, i.e., the channel attention values are broadcasted along the spatial dimension, and vice versa.

As can be seen in Figure 7, in the CBAM, both a channel attention module (CAM) and spatial attention module (SAM) are adopted. On the basis of saving computing power, the CBAM gives the model weight on channel and space, recalibrates the weight on each channel and different space, and increases the weight of the important information to enhance the feature extraction capability of the model.

Finally, the feature map obtained from the CBAM is fed into a convolutional layer with a kernel of 3×3 to implement the feature extraction and dimension transformation. After that, the new feature map is transferred into the subsequent full connection (Fc) layer.

3.4. Full Connection Module

The full connection layer computes the data from the multi-modal fusion layer and finally outputs the classification results. The data are first fed into a dense layer with 128 units, and then the ReLU activation function is used, followed by a dropout layer with a dropout probability of 0.7. The ReLU operation is given by:

$$z = \mathrm{Re}LU(F'') = \max\{F'', 0\} \tag{6}$$

The regularization strategy of dropout is used to prevent some nodes from updating weights and to reduce the node dependencies between adjacent dense layers, and then the SoftMax activation is used to alleviate overfitting and force the nodes to be more

independent than usual. The SoftMax is a general method for multi-label classification tasks, and the operation is given by:

$$p[i] = SoftMax(z[i]) = \frac{e^{z[i]}}{\sum_j e^{z[j]}} \tag{7}$$

where $p[i]$ is the probability of the i-th label and $j \in \{1, 2, \ldots, L\}$, L is the number of labels. The label with the highest probability is the final predicted label of the model. The modulation modes are obtained based on the output value of the SoftMax function. In order to better describe the proposed multi-mode modulation recognition network model, we list the detailed model parameters in Table 2, and the computational process of the proposed M-LSCANet is shown in Algorithm 1.

Table 2. The detailed structure of model.

Module Type	Output Size	Description
Input-1	(1, 2, 1024)	In-phase and quadrature components of signal (I/Q data)
Pre-1	(32, 3, 256)	Includes 2 convolutions, 2 normalizations, and 1 pooling
Conv-Block-1	(32, 3, 256) (8, 3, 256) (8, 3, 256) (32, 3, 256)	4 convolutions
Conv-Block-2	(64, 3, 256) (16, 3, 256) (16, 3, 256) (64, 3, 256)	4 convolutions
Conv-Block-3	(128, 3, 256) (32, 3, 256) (32, 3, 256) (16, 3, 256)	4 convolutions
Maxpool-1	(16, 1, 16)	kernel size = (3, 16)
Input-2	(3, 256, 256)	Constellation diagram data
Pre-2	(8, 64, 64)	1 convolution, 1 normalization, and 1 pooling
Res-Stack-1	(16, 64, 64) × 3	Convolution(stride = 0)
Res-Stack-2	(32, 64, 64) × 4	Convolution(stride = 0)
Res-Stack-3	(32, 64, 64) × 6	Convolution(stride = 2)
Res-Stack-4	(16, 64, 64) × 3	Convolution(stride = 0)
Maxpool-2	(16, 16, 16)	kernel size = (4, 4)
Concatenation	(16, 17, 16)	Concat output of Maxpool 1 and Maxpool 2
CBAM(Attention)	(16, 17, 16)	Spatial attention and channel attention
Dense-1 Dense-2(Output)	(128) (10)	Full connection units Activation: SoftMax

In Algorithm 1, the architecture utilizes time domain I/Q data (X) and constellation diagrams (S) as inputs, and produces the predicted modulation mode as an output. The M-LSCANet comprises four primary modules, namely: SNR segmentation, the time domain I/Q-signal-processing branch, the constellation-diagram-processing branch, and the multi-modal fusion layer process. The SNR of the modulated signal is classified into categories greater than -4 dB or less than -4 dB. When the SNR is greater than or equal to -4 dB, the I/Q-signal-processing branch and constellation-diagram-processing branch are computed,

then the two branches are concatenated. Otherwise, only the I/Q-signal-processing branch is used. In the multi-modal fusion layer process, the CBAM is employed to adjust the weights of the channels and spaces. Finally, the Fc layer outputs the modulation mode. The proposed M-LSCANet wisely integrates the two-modal data to generate the predicted modulation modes.

Algorithm 1: Description of the proposed M-LSCANet

Input: original time domain I/Q data(X) and constellation diagram(S)
Output: modulation mode
Time domain I/Q signal processing branch:
for i = 1 to 3 **do**:
    $X_i = Conv\_i(X_{i-1})$
    $Conv\_i$ is the ith convolutional block described in Section 3.1
end for
SNR Segmentation:
SNR_val = CNN(X)
if $SNR\_val \geq -4$ dB
    Constellation diagram processing branch:
    for i = 1 to 4 **do**:
        $S_i = Res\_i(S_{i-1})$
        Res_i is the ith residual stack described in Section 3.2
    end for
    F = Concatenate (X_3, S_4, axis = 1)
else
    $F = X_3$
end if
Multi-modal fusion layer process:
F'' = Cbam (F)
Cbam represents the recalibration for the weights of the CBAM module, and the detailed calculation process can be referred to Equation (5).
Classification layer process:
Output = SoftMax (ReLU (F''))
ReLU and SoftMax are activation functions, and they are defined in Equations (6) and (7).

4. Simulation Results and Analysis

4.1. Dataset

4.1.1. Dataset's Parameters

This paper uses the public standard RadioML2016.10B DeepSig dataset as the benchmark for training and testing the performance of the proposed method. We also compare it with other state-of-the-art models. This dataset was given by Prof. O'Shen and has been one of the most recognized datasets used in AMR. It is generated by simulating a real communication environment with AWGN, selective fading (Rician and Rayleigh), center frequency offset (CFO), sample rate offset (SRO), and other channel interference. Each modulation mode under each SNR value contains 6000 pieces of two channels of I/Q data, and each piece of data has 2×128 samples. We split the training set and the test set by the ratio of approximately 4:1, and each group contains 4800 and 1200 pieces of data, respectively. The specific parameters of the dataset are listed in Table 3.

Table 3. RadioML2016.10B dataset parameter.

Parameter	Value
Modulation modes	Digital: BPSK [1], QPSK [1], 8PSK [1], QAM16 [2], QAM64 [2], GFSK [3], CPFSK [3], and PAM4 [4]; Analog: WBFM [5] and AM-DSB [5]
Number of datasets	1,200,000 pieces
SNR (dB)	−20:2:18
The number of signal samples for per piece	2×128
Channel model	Center frequency offset (CFO), Sample rate offset (SRO), AWGN, Rayleigh, Rician
Data source	GNU Radio simulated

[1] nPSK: *n*-order phase shift key. [2] QAMn: *n*-order quadrature amplitude modulation. [3] GFSK: Gaussian frequency shift key, CPFSK: continuous phase frequency shift key. [4] PAM4: *n*-order pulse amplitude modulation. [5] WBFM: wideband frequency modulation, AM-DSB: amplitude modulation of double side band.

4.1.2. The Generation of Constellation Diagrams

The dataset consisting of time domain I/Q signals has certain limitations when used to extract signal characteristics. Especially for QAM16, QAM64, and other high-order modulations, the characteristics expressed in the time domain I/Q components are not rich enough. In other words, the difference between different modulation modes is small, making the classification for high-order modulation rather difficult. In fact, as long as the SNR is not particularly low, the difference in the higher-order modulations in the constellations is very obvious; thus, we introduce constellation diagrams into the modulation recognition network to improve its performance.

In this paper, a constellation diagram is directly generated by taking the real and imaginary parts of the modulation signals as data on the X and Y axes, respectively. Because the number of samples at each SNR for each modulation signal is limited, when we use more samples to produce a high-resolution constellation diagram, the number of constellation diagrams will decrease, and vice versa. In order to ensure a high recognition accuracy, we need to make a compromise between the clarity of the constellation diagram and the number of images. Through the subsequent ablation experiments, we find that, when 1024 samples are used to generate one constellation diagram, the model gives a better recognition performance. That is to say, we use eight pieces of I/Q data containing 2×1024 samples for each modulation to generate one constellation diagram. The constellation diagram of each modulation mode at an SNR of 18 dB for the dataset is shown in Figure 8. From left to right, the corresponding modulation modes are 8PSK, AM-DSB, BPSK, CPFSK, GFSK, PAM4, QAM16, QAM64, QPSK, and WBFM.

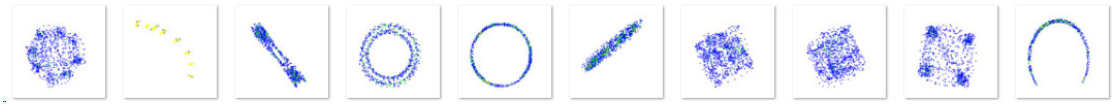


Figure 8. Constellation diagrams with 1024 samples at SNR = 18 dB.

4.2. Experimental Environment

We conduct the simulation using PyTorch as the deep learning development framework and Python3 as the programming language. The default argument Adam [48] is used as the optimizer for learning the hyper-parameters. Its task is to update the model parameters based on first-order and second-order moment estimates of the gradient values. The training rounds are set to 50 epochs and the size of each mini-batch is 64. The initial learning rate is 3×10^{-4} with a decrease to 0.1 times every 20 epochs. The specific hyper-parameter settings are shown in Table 4. The hardware platform has twelve cores i9-9920X @ 3.50 GHz CPU, 64 GB DDR4 RAM, and a single NVIDIA GeForce RTX 2080Ti GPU.

Table 4. Configuration for the simulation.

Type	Value
Optimizer	Adam (Default)
Epoch	50
Mini-batch	64
Learning Rate	3×10^{-4}
LRScheduler	StepLR
LRDropStep	20 epochs
LRDropFactor	0.1

4.3. Ablation Experiments and Analysis

4.3.1. SNR-Segmentation Multi-Modal Network Model

We perform a series of ablation experiments to evaluate the advantages of the proposed M-LSCANet model. In the first ablation experiment, we evaluate if the multi-modal feature fusion is efficient. We compare the modulation recognition performance of three different schemes: (1) using only the I/Q branch in the time domain; (2) using only the constellation diagram branch; and (3) combining the two branches, i.e., a multi-modal network model.

The recognition performances of the three schemes are shown in Figure 9. As can be seen from Figure 9, when only the constellation branch is used, the modulation recognition accuracy is almost the worst in the whole SNR range. When the SNR is higher than −6 dB, the multi-modal joint network achieves the highest recognition accuracy. In the same SNR region, when only using the I/Q-processing branch or constellation branch, the recognition accuracy decreases by 0.4~11.5% and 7.1~74.6%, respectively.

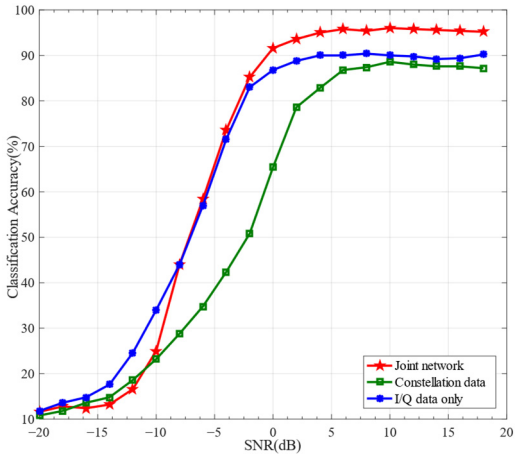


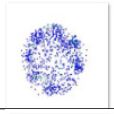
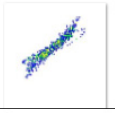
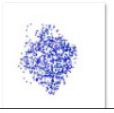
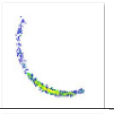
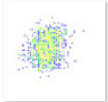
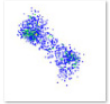
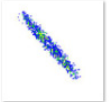
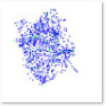
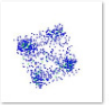
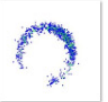
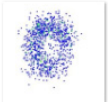
Figure 9. Recognition accuracy with different processing branches.

It is clear that multi-modal feature fusion can improve the performance of the AMR when the SNR is higher than −6 dB. However, it should be noticed that, when the SNR is lower than −6 dB, both the joint network and only constellation schemes are inferior to the only I/Q data scheme, and only using the time domain I/Q signals significantly improves the recognition accuracy. It seems that the introduction of constellation data to the low SNR region leads to a decrease in the recognition performance, despite an increase in the model complexity.

In order to investigate why the recognition performance of the multi-modal network is good at higher SNRs but poor at lower SNRs, we display the constellation diagrams for ten modes of modulation signals at different SNRs in Table 5. It should be noticed that, even for the same modulation mode, such as BPSK, PAM, and WBFM, the constellation diagrams are different at different SNR values. This is because the channel state varies dynamically over time, and modulation signals often undergo different fading and Gaussian noise at different times. This leads to diverse phase rotations and scaling changes in the constellation

diagrams, even for the same modulation mode. Table 5 displays the constellation diagrams of ten kinds of modulation signals at different time intervals. Thus, the phase rotations and scaling changes are also different even for the same modulation mode.

Table 5. Constellations of 1024 points at different SNRs.

SNR \ Mode	8PSK	AM-DSB	BPSK	CPFSK	GFSK	PAM4	QAM16	QAM64	QPSK	WBFM
12 dB										
6 dB										
0 dB										
−6 dB										

By comparing the constellations of ten modulation modes with several SNRs, as shown in Table 5, it is clear that, as the SNR continues to decrease, the difference between the constellation diagrams becomes smaller and smaller. As a result, it becomes almost impossible to distinguish between different modulation signals at an SNR below −6 dB. Therefore, introducing constellation data to the low SNR region to implement modulation recognition is likely to lead to a decline in the recognition accuracy.

Based on the above analysis, we propose an SNR-segmentation modulation recognition strategy for the proposed model structure. In the proposed method, we first divide the modulation signals into medium–high SNR and low SNR, with two segmentations according to the method of Ref. [45], where the same RML2016 dataset is used. In Ref. [45], the SNR classification model uses a CNN to extract the signal features and the Fixed K-means (FK-means) algorithm to cluster the extracted features, in order to accurately differentiate high and low SNR signals. Figure 10 displays the variation curve of the SNR classification accuracy with the SNR threshold. It is clear that the optimal SNR threshold value is −4 dB with an SNR classification accuracy of 97%.

From Figure 9, compared with the single time-domain I/Q scheme, although the recognition accuracy improvement of the joint network first appears when the SNR = −6 dB, its performance advantage is very limited when the SNR varies from −6 dB to −4 dB. Moreover, the SNR classification network depicted in Figure 10 fails to achieve the highest classification accuracy at an SNR threshold of −6 dB, and it experiences a decline compared to the peak. If −6 dB is selected as the SNR segmentation threshold, a larger proportion of the modulation signals will fall into the incorrect SNR region, resulting in a decrease in the modulation recognition accuracy. In contrast, at an SNR = −4 dB, the performance gain of the joint network becomes apparent and the SNR segmentation also provides the optimal accuracy. Furthermore, enabling joint networks early, when the SNR is less than −4 dB, will lead to a significant increase in the complexity. Jointly taking into account the advantages of the SNR classification network and multi-mode modulation recognition network, the SNR classification threshold is set to −4 dB to guarantee a high recognition accuracy and low computational complexity.

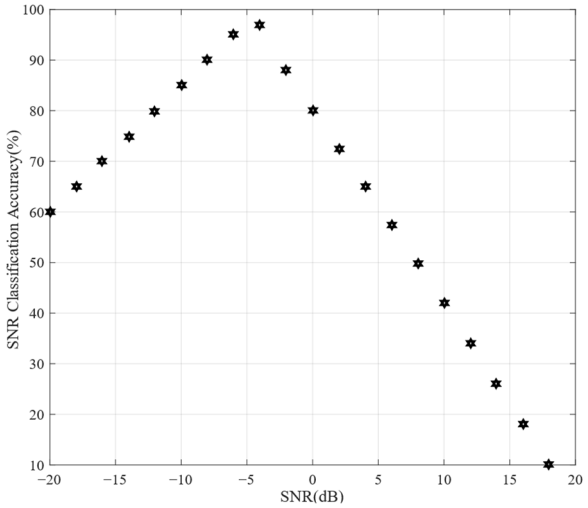


Figure 10. SNR classification accuracy versus SNR threshold.

In the proposed method, the SNR classification network is a two-layer CNN with about 10 k parameters; thus, adding SNR segmentation results in little increase in the computation complexity. When the SNR is judged to be lower than −4 dB, only the I/Q-processing branch is used to implement the AMR. Otherwise, the two branches are jointly used. For the modulation recognition model, we perform SNR-segmentation training and then test it in the entire SNR range. Figure 11 compares the recognition accuracy between the SNR-segmentation M-LSCANet and uniform M-LSCANet, where both processing branches are always used simultaneously throughout the SNR. From Figure 11, it can be seen that, when the SNR varies from −18 dB to −4 dB, the SNR-segmentation M-LSCANet improves the recognition accuracy by 5.9~32.5% compared to the uniform M-LSCANet. It is clear that the SNR segmentation strategy brings obvious improvement to the recognition performance. Unless otherwise stated, the M-LSCANet mentioned below refers to the SNR-segmentation scheme.

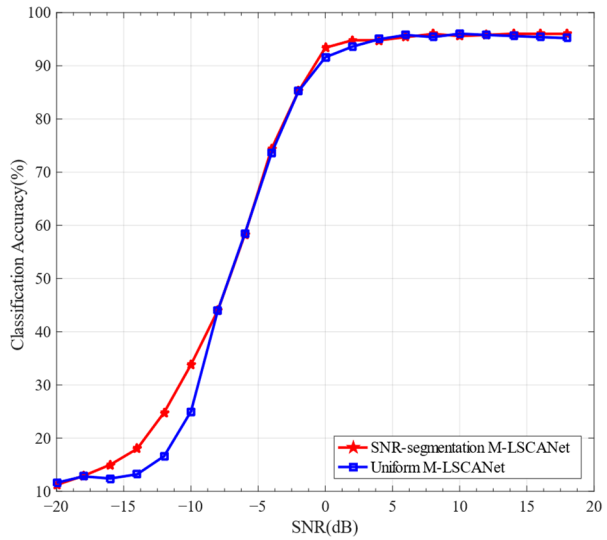


Figure 11. The classification performance comparison between the SNR-segmentation and uniform schemes.

4.3.2. Optimization of Sample Number for One Constellation Diagram

On the other side, the resolution and number of constellation diagrams also affect the recognition performance. Due to the limited number of dataset samples, the increase in the constellation diagram resolution means that the number of available constellation diagrams

will be reduced. According to the experience in the field of image detection, although high-resolution images can improve a model's recognition performance, a small number of images will also lead to insufficient training for the model, thus making it unable to achieve a better performance, and vice versa. Therefore, the optimal number of samples for each constellation diagram will be discussed in the second ablation experiment. We reconstruct the samples of the modulation signals by merging 4, 8, and 16 adjacent data pieces into one group. In this case, each group of data contains 2×512, 2×1024, and 2×2048 I/Q samples, respectively, and the constellation diagrams at an SNR of 18 dB with different sample sizes are given in Figures 8, 12 and 13. It can be seen that, when one constellation diagram contains more samples, it is clearer and can better express the characteristics of the modulation signals, but the number of available constellations will decrease.

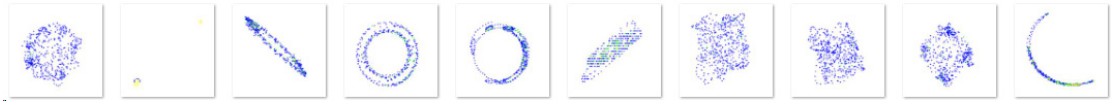


Figure 12. Constellation with 2×512 samples at SNR = 18 dB.

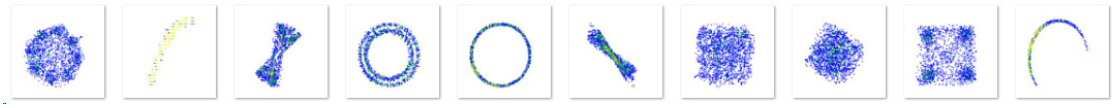


Figure 13. Constellation with 2×2048 samples at SNR = 18 dB.

We use three types of constellation diagrams with different samples in the proposed M-LSCANet, and the recognition accuracy is given in Figure 14. From Figure 14, the scheme using a constellation with 1024 samples achieves the best recognition performance, compared to the schemes using constellations with 512 samples and 2048 samples, where the average recognition accuracy under the entire SNR region improves by up to 2.0% and 3.6%, respectively. Thus, we use 2×1024 samples to generate one constellation diagram in the proposed model, and the corresponding time-domain I/Q data are also grouped into 2×1024 samples.

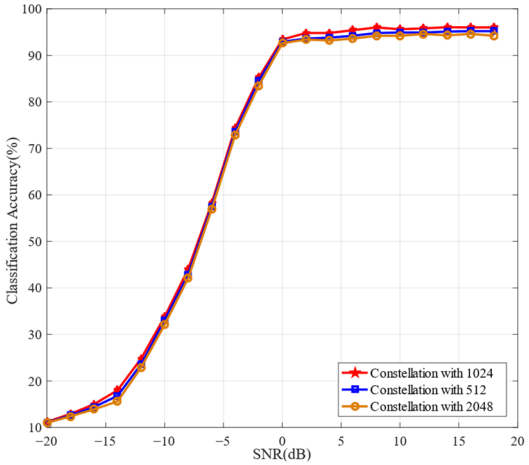


Figure 14. The classification accuracy comparison for one constellation with different sample numbers.

At the same time, we found that there are some irregular data in the dataset for the same modulation mode, which may be the beginning or end of the data in a certain time period. We regard the irregular data as "outliers", such as the yellow image of the 8PSK signals at an SNR of 18 dB shown in Figure 15.

Figure 15. Constellation diagrams of 8PSK at SNR = 18 dB.

In order to reduce the influence of these outliers on the model's training, we replace the outliers for the adjacent data of the same modulation signal during the training stage. However, during the test stage, the outliers are still retained, which is consistent with actual communication systems. The simulation results after eliminating the outliers are shown in Figure 16. It can be clearly seen that the replacement of these "outliers" improves the recognition accuracy of the M-LSCANet model by 0.2~1.3% in the entire SNR range. To some extent, the replacement of outliers is similar to the denoising of the modulation signals. This approach does not lead to an increase in complexity, but a slight improvement in the recognition accuracy.

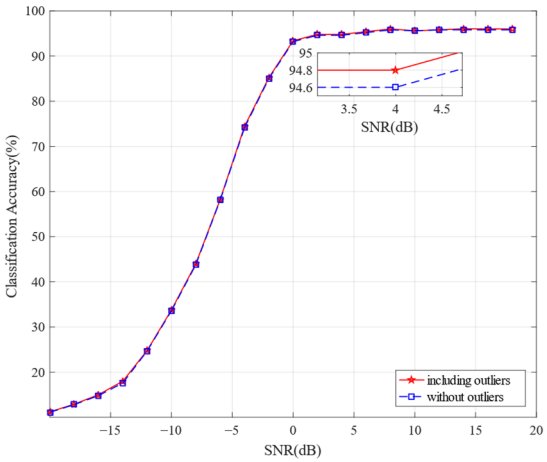


Figure 16. The classification accuracy of eliminating outliers in training data.

4.4. Classification Performance for Each Modulation Mode

To better verify the performance of the proposed method, Figure 17 illustrates the confusion matrices of M-LSCANet's classification accuracy across different SNR values at −6 dB, 0 dB, 6 dB, and 12 dB.

The overall accuracy of the four conditions is 58.2%, 93.4%, 95.8%, and 96.0%, respectively. As the SNR increases, the diagonal of the confusion matrix becomes clearer, indicating that each modulation mode is more accurately identified. It can be seen from Figure 17 that the digital modulation has a good classification performance, while the errors are mainly concentrated in the analog modulation modes, such as AM-DSB and WBFM. This phenomenon can be also observed in Ref. [23]. It can be attributed to the fact that the analog voice signal in RadioML2016.10B has periods of silence where only a carrier tone is present, making these two types of modulation signals indiscernible. Additionally, confusion between QAM16 and QAM64 occurs at an SNR of −6 dB, since the two modulation modes are very close to each other from the time domain waveform when there is strong noise interference. However, the proposed M-LSCANet effectively eliminates this confusion between QAM16 and QAM64 at the medium and high SNR ranges by introducing constellation data. Even if the SNR only reaches 0 dB, the recognition accuracy of QAM16 and QAM64 approaches 100%, which is far better than that of other methods. This proves the superiority of the proposed method model in classifying higher-order modulation modes. It is clear that the combination of time domain IQ signals and constellation diagrams can significantly improve the recognition performance of the model for each modulation mode.

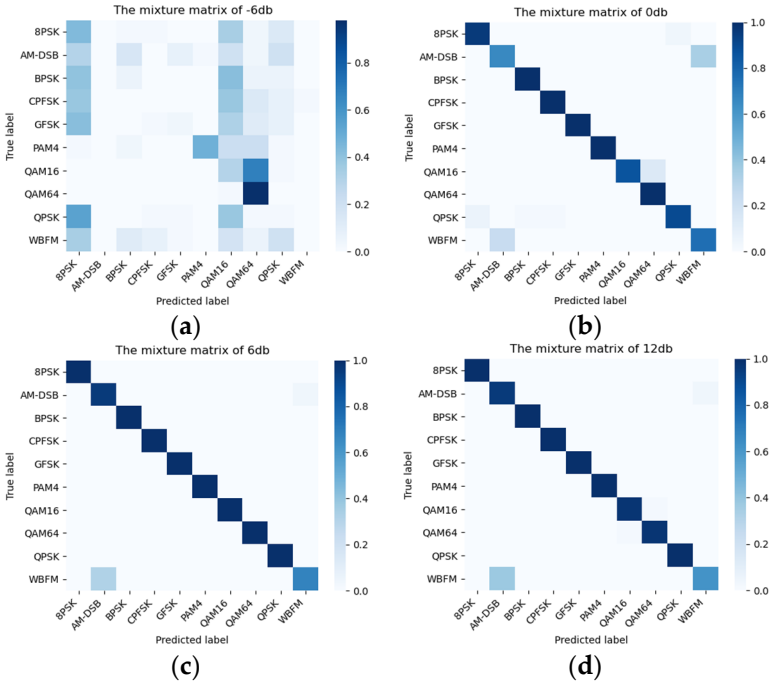


Figure 17. Confusion matrices of M-LSCANet. (**a**) SNR = −6 dB; (**b**) SNR = 0 dB; (**c**) SNR = 6 dB; and (**d**) SNR= 12 dB.

Figure 18 also shows the classification accuracy of M-LSCANet for all 10 modulation modes. As the SNR increases, the recognition accuracy of the digital modulations also increases. When the SNR is above 2 dB, the recognition accuracy of almost all the digital modulations achieves 100%. However, when the SNR is less than 0 dB, the recognition accuracy curves of QAM16 and QAM64 fluctuate, because they are confused with each other. This is consistent with the previous analysis. When it comes to the analog modulations, the recognition accuracy of AM-DSB and WBFM shows sharp fluctuations throughout the SNR range. The highest accuracy of WBFM and AM-DSB is only about 70% and 97%, respectively. A large discrepancy exists in the WBFM classification, which is likely to be recognized as AM-DSB. This discrepancy can also be attributed to the presence of the silence period in the analog voice signals, which is consistent with the previous analysis on the confusion matrix. The recognition accuracy of analog signals needs to be further improved.

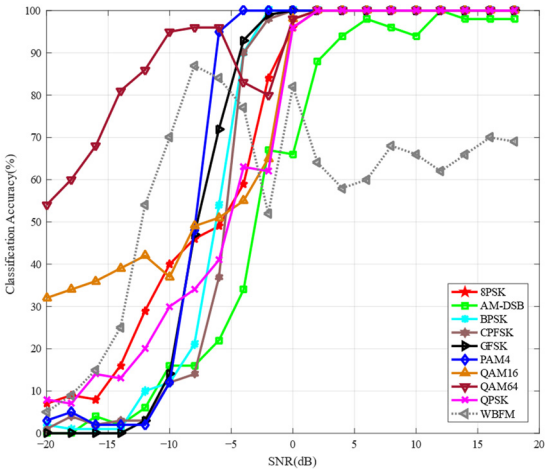


Figure 18. Classification accuracy of different kinds of modulation patterns vs. SNRs.

4.5. Performance Comparison of Different Methods

In order to further evaluate the advantages of the proposed method in this paper, we compared M-LSCANet with different AMR models, including ResNet, CLDNN, IRLNet, TMRN-GLU, and SigFormer-Net. The comparison of the recognition accuracy at different SNRs is shown in Figure 19, and the average and highest recognition accuracy of each model from −20 dB to 18 dB are presented in Table 6. At the same time, the number of parameters in each model is also listed in Table 6. From Figure 19 and Table 6, it is clear that our method achieves the best recognition accuracy across the entire SNR range. Especially in the high SNR region (SNR $\geq$ 0 dB) and low SNR region (SNR $\leq$ −10 dB), the performance advantage of the proposed method is more obvious. Compared to TMRN-GLU, SigFormer-Net, and IRLNet with good recognition performances, the proposed method improves the recognition accuracy by 1.0~28.5%, 0.9~30.4%, and 1.4~32.9%, respectively, in the whole SNR region. At an SNR of 6 dB, M-LSCANet achieves a 2.1%, 2.2%, and 2.6% higher recognition accuracy than TMRN-GLU, SigFormer-Net, and IRLNet, respectively.

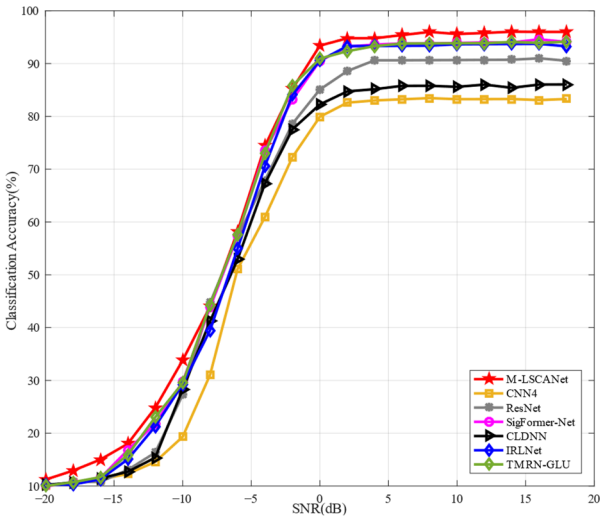


Figure 19. Recognition accuracy comparison with others on RadioML 2016.10B.

Table 6. The average and maximum recognition accuracy and number of model parameters of different networks.

Model	Average Recognition Accuracy	The Maximum Recognition Accuracy	Model Parameters Numbers
M-LSCANet	66.56%	96.04%	544 k [1] (SNR $\geq$ −4 dB) 63 k [1] (SNR < −4 dB)
TMRN-GLU [36]	66.05%	94.74%	103 k
SigFormer-Net [37]	65.77%	94.08%	440 k
IRLNet [35]	63.92%	93.74%	318 k
ResNet [27]	61.79%	91.00%	157 k
CLDNN [49]	59.06%	86.83%	163 k
CNN4 [50]	56.09%	83.46%	3263 k

[1] model parameters include SNR classification network with about parameters of 10 k.

In addition, CNN4 has a lower recognition accuracy than other models, which can be attributed to their poor capabilities for feature extraction. Although CLDNN and ResNet have a higher recognition accuracy than CNN4, their recognition accuracy is much lower than that of M-LSCANet, because these two models do not distinguish QAM modulations very well. Furthermore, the average recognition accuracy of the proposed method under all SNRs achieves 66.56%, the highest value of all the methods.

When it comes to the model parameter numbers, the proposed M-LSCANet has a moderate parameter number with the value of 544 k at the medium and high SNR regions and 63 k at the low SNR region, which are far less than that of CNN4. When the SNR is above -4 dB, the model parameter numbers of the proposed method are 5.3 times, 1.2 times, and 1.7 times that of TMRN-GLU, SigFormer-Net, and IRLNet, respectively, but when the SNR is smaller than -4 dB, the parameter numbers of the proposed method are only 61%, 14%, and 20% of them, respectively. Considering the excellent classification performance in the entire SNR region, the proposed method has remarkable advantages.

5. Conclusions

In this paper, we proposed M-LSCANet, a multi-modal neural network model with SNR segmentation, which integrated the time domain I/Q-processing branch and the constellation-diagram-processing branch. The proposed method featured rich skip connections, lightweight residual stacks, an attention mechanism, and SNR segmentation. The rich skip-connection structures and lightweight residual stacks enabled the model to fully extract the signal features of different layers, while maintaining a moderate complexity. Furthermore, the CBAM module was also introduced to enhance the feature extraction capability of the model. We also proposed the SNR segmentation strategy to enable the constellation branch only when the SNR was higher than or equal to -4 dB. In this case, the constellation feature was properly utilized, the recognition performance of the model in the low SNR region was further improved, and the computational complexity was considerably reduced.

Ablation experiments were carried out to optimize the number of samples for one constellation diagram and demonstrate the efficiencies of the multi-modal structure and SNR segmentation strategy. Additionally, we also eliminated the sample outliers of the signals during the training process, which was helpful for improving the recognition accuracy. The experiment results showed that the proposed method surpassed all the other methods in terms of its recognition accuracy. Due to its high classification accuracy and moderate complexity, the proposed M-LSCANet has a great potential for application in automatic modulation recognition. Due to the increasing complexity of electromagnetic environments, in further work, we will explore the optimization method of a modulation recognition network in the presence of interference signals.

Author Contributions: Conceptualization, R.D. and X.L.; methodology, R.D. and H.Z.; validation, R.D., X.L. and H.Z.; formal analysis, Y.L. and X.L.; investigation, G.Y. and S.L.; resources, R.D. and G.Y.; data curation, G.Y. and S.L.; writing—original draft preparation, R.D. and X.L.; writing—review and editing, X.L., Y.L. and P.C.; visualization, P.C. and S.L.; supervision, R.D. and H.Z.; project administration, R.D. and H.Z.; funding acquisition, R.D. All authors have read and agreed to the published version of the manuscript.

Funding: The work was supported by Beijing Natural Science Foundation (No. L202003) and National Natural Science Foundation of China (No. 31700479).

Data Availability Statement: We used open data set RadioML2016.10B and the link is: http://opendata.deepsig.io/datasets/2016.10/.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ke, D.; Huang, Z.; Wang, X.; Li, X. Blind detection techniques for non-cooperative communication signals based on deep learning. *IEEE Access* **2019**, *7*, 89218–89225. [CrossRef]
2. Zhang, Z.; Wang, C.; Gan, C.; Sun, S.; Wang, M. Automatic modulation classification using convolutional neural network with features fusion of SPWVD and BJD. *IEEE Trans. Signal Inf. Process. Over Netw.* **2019**, *5*, 469–478. [CrossRef]
3. Liang, Y.; Tan, J.; Dusit, N. Overview on intelligent wireless communication technology. *J. Commun.* **2020**, *41*, 1–17.
4. Jiang, J.; Wang, Z.; Zhao, H.; Qiu, S.; Li, J. Modulation recognition method of satellite communication based on CLDNN model. In Proceedings of the IEEE 30th International Symposium on Industrial Electronics (ISIE), Kyoto, Japan, 20–23 June 2021; IEEE: Piscataway, NJ, USA; pp. 1–6.

5. Zhu, Z.; Nandi, A.K. *Automatic Modulation Classification: Principles, Algorithms and Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
6. Zhaoyu, M.A.; Ffuli, H.A.N.; Zhidong, X.I.E. Modulation recognition technology of satellite communication signal system. *Acta Aeronaut. Astronaut. Sin.* **2014**, *35*, 3403–3414.
7. Su, W.; Xu, J.L.; Zhou, M. Real-time modulation classification based on maximum likelihood. *IEEE Commun. Lett.* **2008**, *12*, 801–803. [CrossRef]
8. Meng, F.; Chen, P.; Wu, L.; Wang, X. Automatic modulation classification: A deep learning enabled approach. *IEEE Trans. Veh. Technol.* **2018**, *67*, 10760–10772. [CrossRef]
9. Dobre, O.; Abdi, A.; Bar-Ness, Y.; Su, W. Survey of automatic modulation classification techniques: Classical approaches and new trends. *IET Commun.* **2007**, *1*, 137–156. [CrossRef]
10. Panagiotou, P.; Anastasopoulos, A.; Polydoros, A. Likelihood ratio tests for modulation classification MILCOM 2000 Proceedings. 21st Century Military Communications. *Archit. Technol. Inf. Super.* **2000**, *2*, 670–674.
11. Popoola, J.J.; Van Olst, R. A novel modulation-sensing method. *IEEE Veh. Technol. Mag.* **2011**, *6*, 60–69. [CrossRef]
12. Park, C.S.; Jang, W.; Nah, S.P.; Kim, D.Y. Automatic modulation recognition using support vector machine in software radio applications. In Proceedings of the 9th International Conference on Advanced Communication Technology, Phoenix Park, Republic of Korea, 12–14 February 2007; IEEE: Piscataway, NJ, USA, 2007; Volume 1, pp. 9–12.
13. Aslam, M.W.; Zhu, Z.; Nandi, A.K. Automatic modulation classification using combination of genetic programming and KNN. *IEEE Trans. Wirel. Commun.* **2012**, *11*, 2742–2750.
14. Park, C.S.; Choi, J.H.; Nah, S.P.; Jang, W.; Kim, D.Y. Automatic modulation recognition of digital signals using wavelet features and SVM2008. In Proceedings of the 10th International Conference on Advanced Communication Technology, Gangwon, Republic of Korea, 17–20 February 2008; IEEE: Piscataway, NJ, USA, 2008; Volume 1, pp. 387–390.
15. Das, D.; Anand, A.; Bora, P.K.; Bhattacharjee, R. *Cumulant Based Automatic Modulation Classification of QPSK, OQPSK, π/4-QPSK and 8-PSK in MIMO Environment International Conference on Signal Processing & Communications*; IEEE: Piscataway, NJ, USA, 2016.
16. Ling, Q.; Zhang, L.; Yan, W.; Kong, D. Hierarchical space–time block codes signals classification using higher order cumulants. *Chin. J. Aeronaut.* **2016**, *29*, 754–762. [CrossRef]
17. Sun, X.; Su, S.; Zuo, Z.; Guo, X.; Tan, X. Modulation classification using compressed sensing and decision tree–support vector machine in cognitive radio system. *Sensors* **2020**, *20*, 1438. [CrossRef]
18. Zhang, W. *Automatic Modulation Classification Based on Statistical Features and Support Vector Machine 2014 XXXIth URSI General Assembly and Scientific Symposium (URSI GASS)*; IEEE: Piscataway, NJ, USA, 2014; pp. 1–4.
19. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
20. Young, T.; Hazarika, D.; Poria, S.; Cambria, E. Recent trends in deep learning based natural language processing. *IEEE Comput. Intell. Mag.* **2018**, *13*, 55–75. [CrossRef]
21. Masita, K.L.; Hasan, A.N.; Shongwe, T. Deep learning in object detection: A review. In Proceedings of the International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD), Durban, South Africa, 6–7 August 2020; IEEE: Piscataway, NJ, USA; pp. 1–11.
22. Covington, P.; Adams, J.; Sargin, E. Deep neural networks for youtube recommendations. In Proceedings of the 10th ACM Conference on Recommender Systems, Boston, MA, USA, 15–19 September 2016; pp. 191–198.
23. O'Shea, T.J.; Corgan, J.; Clancy, T.C. *Convolutional Radio Modulation Recognition Networks International Conference on Engineering Applications of Neural Networks*; Springer: Cham, Switzerland, 2016; pp. 213–226.
24. Hong, D.; Zhang, Z.; Xu, X. Automatic modulation classification using recurrent neural networks. In Proceedings of the 2017 3rd IEEE International Conference on Computer and Communications (ICCC), Chengdu, China, 13–16 December 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 695–700.
25. Rajendran, S.; Meert, W.; Giustiniano, D.; Lenders, V.; Pollin, S. Deep learning models for wireless signal classification with distributed low-cost spectrum sensors. *IEEE Trans. Cogn. Commun. Netw.* **2018**, *4*, 433–445. [CrossRef]
26. Sun, J.; Shi, W.; Yang, Z.; Yang, J.; Gui, G. Behavioral modeling and linearization of wideband RF power amplifiers using BiLSTM networks for 5G wireless systems. *IEEE Trans. Veh. Technol.* **2019**, *68*, 10348–10356. [CrossRef]
27. O'Shea, T.J.; Roy, T.; Clancy, T.C. Over-the-air deep learning based radio signal classification. *IEEE J. Sel. Top. Signal Process.* **2018**, *12*, 168–179. [CrossRef]
28. Huynh-The, T.; Hua, C.H.; Pham, Q.V.; Kim, D.S. MCNet: An efficient CNN architecture for robust automatic modulation classification. *IEEE Commun. Lett.* **2020**, *24*, 811–815. [CrossRef]
29. Tunze, G.B.; Huynh-The, T.; Lee, J.M.; Kim, D.S. Multi-shuffled convolutional blocks for low-complex modulation recognition. In Proceedings of the 2020 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Islan, Republic of Korea, 21–23 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 939–942.
30. Tunze, G.B.; Huynh-The, T.; Lee, J.-M.; Kim, D.S. Sparsely connected CNN for efficient automatic modulation recognition. *IEEE Trans. Veh. Technol.* **2020**, *69*, 15557–15568. [CrossRef]
31. Huynh-The, T.; Hua, C.H.; Doan, V.S.; Kim, D.S. Accurate modulation classification with reusable-feature convolutional neural network. In Proceedings of the 2020 IEEE Eighth International Conference on Communications and Electronics (ICCE), Phu Quoc Island, Vietnam, 13–15 January 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 12–17.

32. Huynh-The, T.; Doan, V.S.; Hua, C.H.; Pham, Q.V.; Kim, D.S. Chain-Net: Learning deep model for modulation classification under synthetic channel impairment GLOBECOM 2020. In Proceedings of the 2020 IEEE Global Communications Conference, Taipei, Taiwan, 7–11 December 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.
33. Wang, Y.; Yang, J.; Liu, M.; Gui, G. LightAMC: Lightweight automatic modulation classification via deep learning and compressive sensing. *IEEE Trans. Veh. Technol.* **2020**, *69*, 3491–3495. [CrossRef]
34. Njoku, J.N.; Morocho-Cayamcela, M.E.; Lim, W. CGDNet: Efficient hybrid deep learning model for robust automatic modulation recognition. *IEEE Netw. Lett.* **2021**, *3*, 47–51. [CrossRef]
35. Yang, H.; Zhao, L.; Yue, G.; Ma, B.; Li, W. IRLNet: A Short-Time and Robust Architecture for Automatic Modulation Recognition. *IEEE Access* **2021**, *9*, 143661–143676. [CrossRef]
36. Zheng, Y.; Ma, Y.; Tian, C. TMRN-GLU: A Transformer-Based Automatic Classification Recognition Network Improved by Gate Linear Unit. *Electronics* **2022**, *11*, 1554. [CrossRef]
37. Su, H.; Fan, X.; Liu, H. Robust and Efficient Modulation Recognition with Pyramid Signal Transformer GLOBECOM 2022. In Proceedings of the 2022 IEEE Global Communications Conference, Rio de Janeiro, Brazil, 4–8 December 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1868–1874.
38. An, T.T.; Lee, B.M. Robust Automatic Modulation Classification in Low Signal to Noise Ratio. *IEEE Access* **2023**, *11*, 7860–7872. [CrossRef]
39. Huang, S.; Jiang, Y.; Gao, Y.; Feng, Z.; Zhang, P. Automatic modulation classification using contrastive fully convolutional network. *IEEE Wirel. Commun. Lett.* **2019**, *8*, 1044–1047. [CrossRef]
40. Xu, M.; Hou, J.; Wu, P.; Liu, Y.; Lv, Z. Convolutional neural networks based on time-frequency characteristic for modulation classification. *Comput. Sci.* **2020**, *47*, 175–179.
41. Zha, X.; Peng, H.; Qin, X. Modulation recognition method based on multi-inputs convolution neural network. *J. Commun.* **2019**, *40*, 30–37.
42. Wang, Y.; Liu, M.; Yang, J.; Gui, G. Data-driven deep learning for automatic modulation recognition in cognitive radios. *IEEE Trans. Veh. Technol.* **2019**, *68*, 4074–4077. [CrossRef]
43. Han, H.; Yi, Z.; Zhu, Z.; Li, L.; Gong, S.; Li, B.; Wang, M. Automatic Modulation Recognition Based on Deep-Learning Features Fusion of Signal and Constellation Diagram. *Electronics* **2023**, *12*, 552. [CrossRef]
44. Ma, P.; Liu, Y.; Li, L.; Zhu, Z.; Li, B. A Robust Constellation Diagram Representation for Communication Signal and Automatic Modulation Classification. *Electronics* **2023**, *12*, 920. [CrossRef]
45. Guo, Y.; Yao, W. Modulation Signal Classification and Recognition Algorithm Based on Signal to Noise Ratio Classification Network. *J. Electron. Inf. Technol.* **2022**, *44*, 3507–3515. (In Chinese)
46. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans. Med. Imaging* **2019**, *39*, 1856–1867. [CrossRef]
47. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, 8–14 September 2018; pp. 3–19.
48. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
49. West, N.E.; O'shea, T. Deep architectures for modulation recognition. In Proceedings of the 2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), Baltimore, MD, USA, 6–9 March 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–6.
50. Ramjee, S.; Ju, S.; Yang, D.; Liu, X.; Gamal, A.E.; Eldar, Y.C. Fast deep learning for automatic modulation classification. *arXiv* **2019**, arXiv:1901.05850.