applied sciences

MDPI

Article

CNN-BiLSTM-DNN-Based Modulation Recognition Algorithm at Low SNR

Xueqin Zhang [1], Zhongqiang Luo [1,2,*] and Wenshi Xiao [1]

[1] School of Automation and Information Engineering, Sichuan University of Science and Engineering, Yibin 644000, China; 322085404213@stu.suse.edu.cn (X.Z.); 321085404115@stu.suse.edu.cn (W.X.)
[2] Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science and Engineering, Yibin 644000, China
* Correspondence: luozhongqiang@suse.edu.cn

Abstract: Radio spectrum resources are very limited and have become increasingly tight in recent years, and the exponential growth of various frequency-using devices has led to an increasingly complex and changeable electromagnetic environment. Wireless channel complexity and uncertainty have increased dramatically, and automated modulation recognition (AMR) performs poorly at low signal-to-noise ratios. It is proposed to use convolutional bidirectional long short-term memory deep neural networks (CNN-BiLSTM-DNNs) as a deep learning framework to extract features from single and mixed in-phase/orthogonal (I/Q) symbols in modulated data. The framework combines the capabilities of one- and two-dimensional convolution, a bidirectional long short-term memory network, and a deep neural network more efficiently, extracting characteristics from the perspective of time and space to enhance the accuracy of automatic modulation recognition. Modulation recognition experiments on the benchmark datasets RML2016.10b and RML2016.10a show that the average recognition accuracies of the proposed model from −20 dB to 18 dB are 64.76% and 62.73%, respectively, and the improvement ranges of modulation recognition accuracy are 0.29−5.56% and 0.32−4.23% when the signal-to-noise ratio (SNR) is −10 dB to 4 dB, respectively. The CNN-BiLSTM-DNN model outperforms classical models such as MCLDNN, MCNet, CGDNet, ResNet, and IC-AMCNet in terms of modulation type recognition accuracy.

Keywords: automatic modulation recognition; deep learning; convolutional neural network; bidirectional long short-term memory network; deep neural network

Citation: Zhang, X.; Luo, Z.; Xiao, W. CNN-BiLSTM-DNN-Based Modulation Recognition Algorithm at Low SNR. *Appl. Sci.* **2024**, *14*, 5879. https://doi.org/10.3390/app14135879

Academic Editor: Alessandro Lo Schiavo

Received: 29 April 2024
Revised: 2 July 2024
Accepted: 3 July 2024
Published: 5 July 2024

1. Introduction

The development of automatic modulation and recognition technology has always been a crucial focus in modern military communications [1,2]. This technology is vital in identifying enemy jamming signals and important military information during electronic surveillance and countermeasures operations [3]. By utilizing this technology, military forces can effectively devise targeted strategies for reconnaissance and counter-reconnaissance missions [4].

In the field of traditional modulation recognition methods, there are two main categories: algorithms based on likelihood functions [5,6] and algorithms based on signal features [7,8]. Likelihood function-based algorithms require the accurate modeling of unknown signals, treating the modulation recognition problem as a multi-hypothetical scenario. However, these algorithms rely heavily on prior information, demanding high precision and complex calculations for the likelihood function model. On the other hand, feature-based approaches involve extracting signal features and employing classifiers for modulation recognition. While this method offers lower computational complexity and greater practicality compared to likelihood function-based approaches, it necessitates manual feature extraction and classifier design, leading to recognition outcomes heavily influenced by expert experience in feature extraction.

Compared with the limitations of traditional modulation recognition methods, such as artificial feature analysis, complex algorithms, and large computational costs [9], the current deep learning (DL) AMR technology has brought new solutions to problems that cannot be overcome by traditional modulation recognition methods [10]. The advantage of modulation recognition with DL is that it can automatically extract the features of the signal and obtain the classifier by training the neural network without the need to do complex and difficult artificial features and classifier design, and it has higher recognition accuracy than the traditional modulation recognition method. One of the most common methods is to process the signal into in-phase (I) and quadrature (Q) components and extract them with deep networks. In their research, S. Hong and his team introduced a cutting-edge DL-powered AMR system tailored for detecting signals in OFDM systems. By leveraging convolutional neural networks to analyze IQ samples of OFDM signals, they achieved remarkable results [11]. Their experiments revealed that, with a signal-to-noise ratio of 10 dB, the accuracy of signal classification exceeded 90%. However, as the signal-to-noise ratio (SNR) dropped below 10 dB, the recognition accuracy experienced a sharp decline. Fei and colleagues [12] took a different approach by developing the CLRD model, a dual-channel fusion network that harnesses deep learning techniques to effectively identify signals. By simultaneously capturing temporal and spatial features of the signals, their model proved to be highly efficient in signal recognition. Wang X et al. [13] proposed a multi-stream neural network designed to extract features from various aspects of modulation signals, such as amplitude, phase, frequency, and raw data. Despite its innovative design, the recognition rate for certain modulation types fell short of expectations. Meanwhile, Lin S. and his team [14] combined a convolutional neural network with a time–frequency attention mechanism to extract crucial features from signals.

To address the problem of poor modulation recognition in environments with low SNRs, a cutting-edge automatic modulation recognition model leveraging deep learning techniques has been developed. This innovative model combines a convolutional neural network (CNN), a bidirectional long short-term memory (BiLSTM) network, a deep neural network (DNN), and an attention mechanism. The process begins with the preprocessing of the original IQ signal using a signal processing (SP) module, which effectively eliminates unwanted additive white Gaussian noise. Subsequently, the preprocessed I/Q multi-channel signals are separated into independent I and Q data streams. These three input data streams are then fed into the CNN to capture both multi-channel and single-channel spatial characteristics of the I/Q signals. The incorporation of a spatial attention mechanism ensures that the model focuses on crucial spatial regions while disregarding irrelevant ones. Moving forward, the BiLSTM network extracts temporal features, allowing the model to understand dependencies and prioritize different parts of the sequence effectively through a time attention mechanism. Finally, a fully connected layer is used to identify the modulation mode. To evaluate the way in which this was performed in the CNN-BiLSTM-DNN framework under challenging conditions with low signal-to-noise ratios, extensive testing was conducted on the RML2016.10b and RML2016.10a public datasets.

The main contributions of the CNN-BiLSTM-DNN framework proposed in this paper are as follows:

- We combine a CNN, BiLSTM, a DNN, and an attention mechanism in a hybrid neural network architecture to leverage their complementarity and synergy for extracting and classifying spatiotemporal features. The CNN is used to learn the spatial features of I/Q signals. The BiLSTM network can extract bidirectional time series features in the time dimension and effectively avoid the problems of gradient explosion and gradient vanishing, and fully connected (FC) deep neural networks achieve effective feature classification.
- The signal preprocessing (SP) module is used to process the original I/Q signal, which effectively filters out additive white Gaussian noise and lays a solid foundation for subsequent feature extraction.

- By including the attention mechanism module in the model, it is possible to elevate the model's representation capabilities, minimize the interference caused by invalid targets, enhance the target of interest's recognition effect, and ultimately elevate the model's overall performance.

The structure of this study is as follows: the signal model and the signal preprocessing methods used in this work are introduced in the second part. The structure of the convolutional bidirectional long short-term memory deep neural networks suggested in this paper is shown in the Section 3, along with a detailed explanation of each module's makeup and purposes. Section 4 describes and analyzes the experimental setup and results. The Section 5 concludes with a summary of the study and looks at the advantages and disadvantages of the recommended approach.

2. Signal Model and Signal Preprocessing

2.1. Signal Model

The single-input, single-output (SISO) system under consideration in this article can be written as follows [15]:

$$y(l) = Ae^{j(2\pi f_0 Tl+\theta_l)} \cdot \sum_{n=-\infty}^{\infty} x(n)h(lT - nT + t_0T) + g(l) \tag{1}$$

Here, $x(n)$ is the complex baseband symbol sequence transmitted by the transmitter modulated in a certain way, A is an amplitude factor, $h(\cdot)$ represents channel effects, T is the symbol duration, t_0 is the timing error, f_0 is the frequency offset, θ_l is the phase offset, and $g(l)$ denotes complex additive white Gaussian noise (AWGN). It is possible to store the received signal in a discrete IQ form with a sampling length of L to facilitate data processing and modulation recognition, represented as

$$Y = \begin{bmatrix} y_i \\ y_q \end{bmatrix} = \begin{bmatrix} \Re\{y[1]\}, \cdots, \Re\{y[L]\} \\ \Im\{y[1]\}, \cdots, \Im\{y[L]\} \end{bmatrix} \tag{2}$$

where y_i and y_q are the in-phase and quadrature components, respectively. Modulation recognition involves identifying the modulation mode of the transmitted signal $x(n)$ from Y, even though the signal's structural features may have been distorted. It is also possible to process the incoming signal further and transform it into different signal representations, including eye diagrams, spectral graphs, and higher-order cumulants. For different signal representation methods, it is necessary to design distinct modulation recognition models.

2.2. Signal Preprocessing

To filter out the mixed noise in the signal, the signal is denoised using the SP module [16]. The SP module design includes a pooling layer stack consisting of the MinPool1D layer and the AvgPool1D layer:

$$R'_m = AvgPool1D(MinPool1D)(R_m) \tag{3}$$

where R'_m represents the radio signal. MinPool and AvgPool are two operations that use a sliding window to smooth out the data. The SP module uses the MinPool1D layer for sliding window filtering, which is particularly effective for noise mitigation in low SNR conditions. Subsequently, the data are smoothed through the AvgPool1D layer to further eliminate noise and spikes introduced by MinPool operations. The combination of MinPool and AvgPool processing significantly reduces the impact of noise and improves the uniformity of feature representation. Therefore, the SP module is located in front of CNN-BiLSTM-DNN to reduce the noise mixed with signals with a low SNR and promote the uniformity of feature representation; it facilitates the extraction of features.

3. AMR Framework

A framework of CNN-BiLSTM-DNN is proposed, as shown in Figure 1. The CNN-BiLSTM-DNN model is composed of a CNN, BiLSTM, an attention mechanism, and a fully connected (FC) deep neural network, which can be used to extract and classify spatiotemporal features by using their complementarity and synergy. Incorporating an attention mechanism into the model can greatly enhance its ability to represent data, effectively filtering out irrelevant information and improving the recognition accuracy of the target of interest. This ultimately leads to a significant boost in overall model performance. To eliminate noise from the input signal, it is first filtered by the minimum pooling layer and the average pooling layer. Then, the I signal and the Q signal pass through two one-dimensional convolution layers, respectively, and then, the spliced signal passes through a two-dimensional convolution layer. Then, it fuses with the I/Q signal and finally extracts the spatial features of the signal through another two-dimensional convolution layer. Then, the signal passes through two bidirectional LSTM layers to extract the signal's time characteristics. Finally, the classification results are output by the fully connected layer. Three-channel input, spatial feature mapping, temporal feature extraction, and fully integrated classification make up the functional parts of the model.

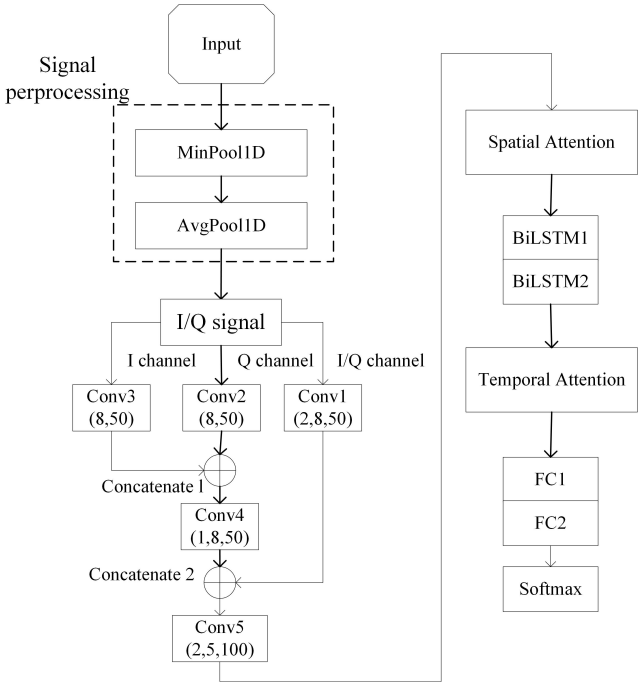

Figure 1. CNN-BiLSTM-DNN model frame diagram.

3.1. CNN

As a feedforward neural network, the CNN [17] is composed of multi-layer neural networks, and its neurons can respond to the partial coverage of surrounding units, which has obvious advantages in local feature extraction. Convolutional layers and pooling layers make up the majority of CNNs. The core component of the CNN, the convolutional layer, uses a convolutional filter to extract the spatial aspects of the signal from the input data to learn the features and convolve with the learnable kernel. The following is a representation of the convolution operation:

$$X(i,j) = \sum_{m}\sum_{n} x(m+i, n+i)\omega(m,n) + b \tag{4}$$

where $X(i,j)$ represents the convolution output, x is the input matrix, ω is the weight matrix of size $m \times n$, and b is the bias.

Two 1D convolutional layers (Conv2 and Conv3) and three 2D convolutional layers (Conv1, Conv4, and Conv5) make up the spatial feature extraction module, which refines features for BiLSTM by reducing noise and abstracting input data at a higher level. Initially, the I/Q multiplex signal is preprocessed and split into distinct I-channel and Q-channel data streams. These three data streams are then individually processed by Conv1, Conv2, and Conv3 to capture both multi-channel and single-channel characteristics of the I/Q signal. To maintain data integrity during modeling, Conv2 and Conv3 utilize zero padding. The outputs are then combined in Concatenate2 before being fed into Conv5 for spatial feature extraction. This multi-channel input structure effectively captures representation features at various scales and maximizes the utilization of information from I-channel, Q-channel, and I/Q multi-channel data.

3.2. BiLSTM

Recurrent neural networks (RNNs) have a special variation known as long short-term memory networks (LSTMs) [18], which include memory units, forgetting gates, input gates, and output gates, as shown in Figure 2. These elements work together to selectively store, retrieve, and discard information within the network. The input gate regulates which input data is stored in the memory unit, and the information flow from the memory cell to external sources is managed by the output gate. The forgetting gate determines whether certain information should be retained or discarded. The specific calculation process of LSTM is as follows:

$$F_t = \sigma(W_f \cdot [H_{t-1}, X_t] + b_f) \tag{5}$$

$$I_t = \sigma(W_i \cdot [H_{t-1}, X_t] + b_i) \tag{6}$$

$$\widetilde{C}_t = \tanh(W_c \cdot [H_{t-1}, X_t] + b_c) \tag{7}$$

$$C_t = F_t \cdot C_t - 1 + I_t \cdot \widetilde{C}_t \tag{8}$$

$$O_t = \sigma(W_o \cdot [H_{t-1}, X_t] + b_o) \tag{9}$$

$$H_t = O_t \cdot \tanh(C_t) \tag{10}$$

In these equations, the forget gate, input gate, and output gate's respective outputs are denoted by the letters F_t, I_t, and O_t. $\widetilde{C}_t$ and C_t represent the cell states at the current time step and the next time step, respectively. H_t denotes the final output value. σ is the sigmoid function. W_f, W_i, W_c, and W_o are the weight matrices of the forget, input gate, and output gates and the current time step cell state, respectively. b_f, b_i, b_c, and b_o represent the bias terms of the gates and the current time step cell state, respectively. $[H_{t-1}, X_t]$ signifies a vector composed of the output values from the previous time step and the current time step.

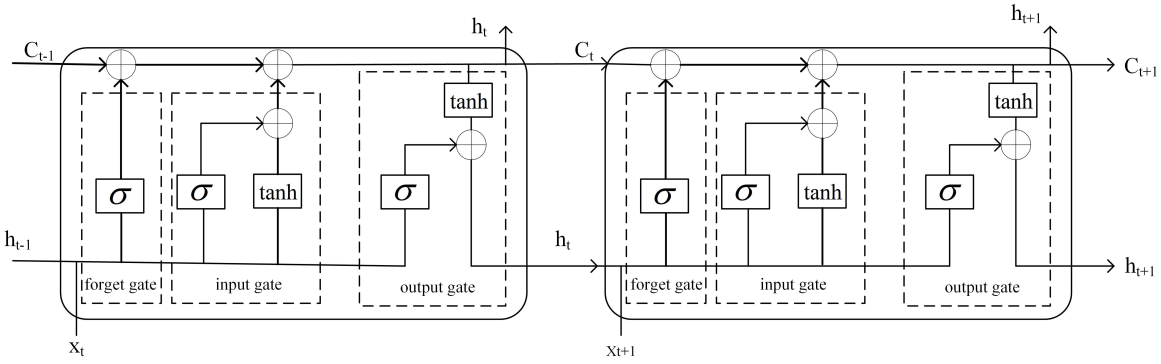

Figure 2. LSTM network model frame diagram.

The network architecture of the BiLSTM, which consists of two separate LSTMs, is depicted in Figure 3. The input sequence is extracted by two LSTMs, forward and backward,

respectively, and the extracted feature vectors are spliced as the final output features. The calculation process is as follows:

$$\vec{C}_t = LSTM(X_t, \vec{H}_{t-1}, \vec{C}_{t-1}) \tag{11}$$

$$\overleftarrow{C}_t = LSTM(X_t, \overleftarrow{H}_{t-1}, \overleftarrow{C}_{t-1}) \tag{12}$$

$$C_t = W^T \vec{C}_t + W^V \overleftarrow{C}_t \tag{13}$$

The forward LSTM and the backward LSTM cell states at time step t are represented by $\vec{C}_t$ and $\overleftarrow{C}_t$, respectively. W^T and W^V denote the weight coefficients of the forward LSTM and the backward LSTM, respectively.

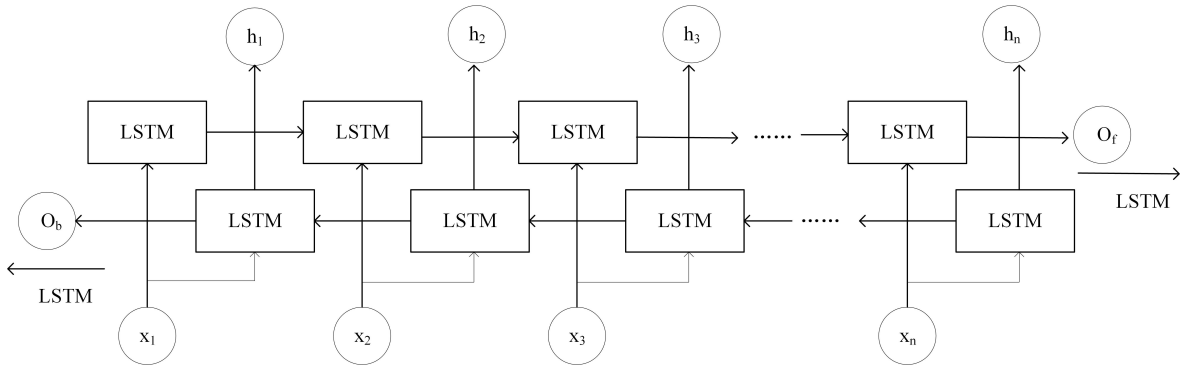

Figure 3. BiLSTM network model frame diagram.

While a single CNN network is adept at capturing the spatial intricacies of wireless signals, it falls short in capturing their temporal nuances. Inspired by the structure proposed in [19], we have integrated a series BiLSTM network behind the CNN to delve into the bidirectional time series features along the temporal axis. This innovative design comprises two BiLSTM layers with 128 units each, enabling efficient processing of sequence data and the extraction of time correlations. The gate mechanism within the BiLSTM effectively tackles gradient-related challenges, elevating classification accuracy to new heights.

3.3. Time Attention Mechanism

In deep learning, the time attention mechanism is a method for handling sequential input. It allows the model to assign different importance or attention to the information at different time steps when processing sequence data. With the help of this approach, the model may more effectively understand the dependencies and significance of various sequence segments and change its attention as necessary. During initialization, this layer creates two sets of weights and is used for calculating attention scores. Attention scores are computed using Formula (14):

$$e_{ij} = \tanh(Wx + b) \tag{14}$$

where x represents the input. Softmax and the weighted inputs index and normalize the attention score to obtain the attention weight a. To emphasize the key components of the input sequence, these attention weights are applied to the input. Finally, the weighted inputs are summed along the sequence axis to produce the final output of the layer.

By means of the aforementioned mechanism, the temporal attention mechanism is able to dynamically modify the weights in accordance with the significance of various segments of the input sequence. This allows the model to concentrate on the most pertinent information at every stage, thereby enhancing its processing efficiency and highlighting the salient features of the sequence data.

3.4. Spatial Attention Mechanism

The spatial attention mechanism allows the model to adaptively learn the attention weights of various regions by incorporating an attention module, whereas the temporal attention mechanism seeks to capture the significance of time series data. In this way, the model can focus more on important areas of the image and ignore unimportant areas. Using convolution operations and the sigmoid activation function to highlight certain regional traits, the spatial attention mechanism represents the significance of input features in the spatial dimension. The working mechanism is

$$attention = \sigma(\mathrm{con}v) \tag{15}$$

where $attention$ represents the attention weights obtained through an activation function, σ denotes the sigmoid activation function, and $conv$ is the result of the convolution operation.

To improve the quality of feature representation, we added two FC layers, each containing 128 neurons, along with scaled exponential linear unit (SeLu) activations to improve the network design. To combat overfitting, the dropout algorithm is judiciously employed. The output layer boasts Softmax activation with 10 neurons (11 under the RadioML2016.10a dataset), with each neuron corresponding to a distinct modulation scheme. Scaled exponential linear units (SeLus) are strategically incorporated to deepen the network's capacity. Dropout comes into play as a shield against overfitting, ensuring robust performance. The output layer determines the modulation mode of the modulated signal using Softmax.

4. Experiment and Conclusion Analysis

4.1. Experimental Data and Parameter Settings

The experiment was conducted using the well-known open-source benchmark datasets RadioML2016.10a and RadioML2016.10b [20]. Eleven widely used modulation signals, including WBFM, AM-DSB, AM-SSB, BPSK, CPFSK, GFSK, 4PAM, 16QAM, 64QAM, QPSK, and 8PSK, are included in the RadioML2016.10a dataset. Every modulation signal has a signal-to-noise ratio between -20 and 18 dB. A total of 220,000 modulated signals were generated with 1000 samples at 2 dB intervals. On the other hand, the 1,200,000 signals in the RadioML2016.10b dataset are an expansion of the RadioML2016.10a dataset and include 10 modulation signals (not AM-SSB). These datasets contain signals with lengths of 128 that combine the components of the real part (I) and the imaginary part (Q) to form complicated groups. These signals, which include Gaussian white noise, multipath fading, sampling rate shift, and center frequency shift, were generated in difficult propagation settings simulations.

Based on a single SNR for each modulation style, the signals in the dataset were randomly divided into training sets, verification sets, and test sets at a ratio of 6:2:2. Gradient updates with a 400-batch size were optimized using the Adam optimizer and the cross-entropy loss function. In cases where the verified loss did not decrease within five epochs, a multiplier of 0.8 was applied to enhance training efficiency. If the verification loss remained stagnant for 60 consecutive epochs, the training process was halted, and the model with the lowest verification loss was selected for predicting the modulation type of each test signal. All experiments were conducted on a Windows 10 operating system utilizing the Pycharm software platform. The hardware setup included a robust configuration featuring a 36-core Intel (R) Xeon (R) CPU E5-269 v4 @ 2.10GHz processor, 160 GB of memory, and an NVIDIA Titan Xp graphics card with a total video memory of 36 G.

4.2. Analysis of Experimental Results

4.2.1. Recognition Accuracy

The fundamental challenge of modulation signal classification recognition is one of classification, where classification accuracy is defined as the proportion of correctly predicted samples among all input samples, as stated in the Formula (16) [21]:

$$acc = \frac{tp + tn}{tp + tn + fp + fn} \tag{16}$$

where tp represents a sample labeled as correct and a model prediction as also correct. tn represents a sample that is labeled wrong and a model prediction that is also wrong, and fp represents a sample that is labeled wrong and a model prediction that is correct. fn represents a sample marked as correct but mispredicted by the model.

To validate the proposed algorithmic structure, a comparison was made with algorithms such as a modulation classification convolutional neural network (MCNet) [22], convolutional gated recurrent fully connected deep neural network (CGDNet) [23], residual network (ResNet) [24], multi-channel convolutional long short-term deep neural network (MCLDNN) [25], and improved convolutional neural network-based automatic modulation classification network (IC-AMCNet) [26]. Figure 4 shows the accuracy of each SNR of the six models under the RadioML2016.10b dataset. Ours in the figure represents the model proposed in this paper. The classification accuracy of different network structures can be clearly seen in the figure, and when SNR rises, so does its accuracy. The maximum recognition accuracy at 12 dB reaches 93.79%, the greatest of all comparison models. The CNN-BiLSTM-DNN network structure performs noticeably better in terms of recognition accuracy than other networks when the signal-to-noise ratio is low, ranging from −10 dB to 4 dB. This further validates the efficacy of the SP module denoising method and enhances modulated signal classification and recognition.

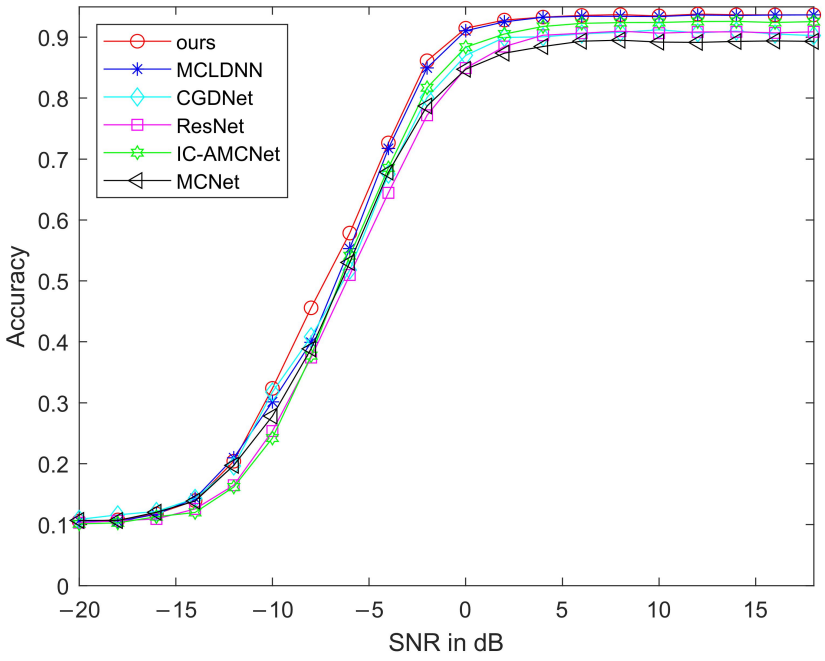

Figure 4. Comparison of recognition accuracies of different models in the RadioML2016.10b dataset.

To demonstrate our innovative model, we conducted simulations on the RadioML2016.10a dataset. According to Figure 5, the rate of recognition of our CNN-BiLSTM-DNN network structure surpasses that of competing algorithms, achieving an impressive 93.18% accuracy

at an SNR of 4 dB. The results captured in Tables 1 and 2 showcase the peak recognition accuracy and average performance of our experiment.

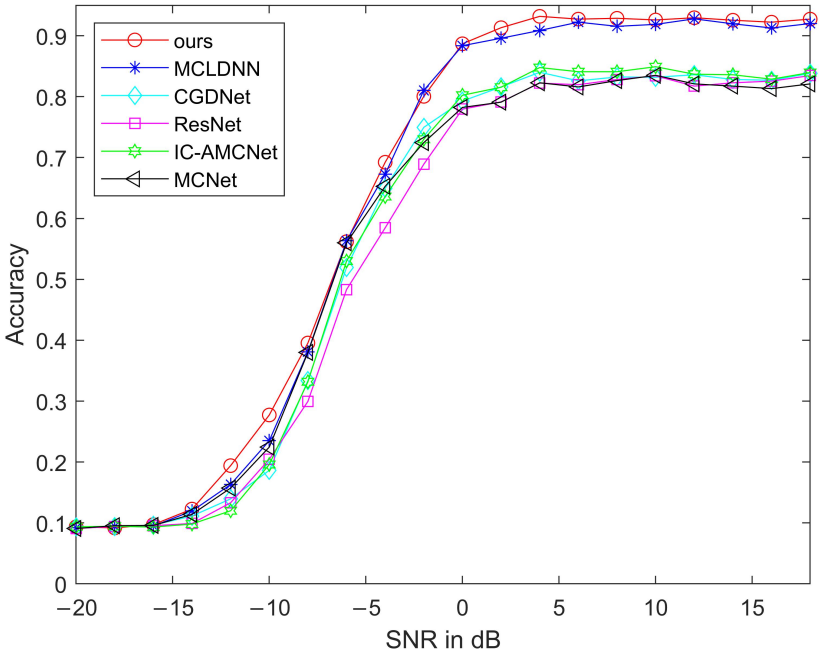

Figure 5. Comparison of recognition accuracies of different models in the RadioML2016.10a dataset.

Table 1. Experimental results of different models in the RadioML2016.10b dataset. (**bold**: best)

Model	Maximum Accuracy (%)	Average Accuracy (%)
CNN-BiLSTM-DNN	**93.79**	**64.76**
MCLDNN	93.67	64.09
MCNet	89.51	60.95
CGDNet	91.21	62.13
ResNet	90.99	60.78
IC-AMCNet	92.59	62.21

Table 2. Experimental results of different models in the RadioML2016.10a dataset. (**bold**: best)

Model	Maximum Accuracy (%)	Average Accuracy (%)
CNN-BiLSTM-DNN	**93.18**	**62.73**
MCLDNN	92.77	61.75
MCNet	83.50	56.20
CGDNet	84.00	56.21
ResNet	83.45	54.74
IC-AMCNet	84.91	56.30

Table 3 compares the modulation recognition results of different models at −10 dB, −8 dB, −6 dB, −4 dB, −2 dB, 0 dB, 2 dB, and 4 dB. The table's results indicate that the proposed CNN-BiLSTM-DNN and other models except MCLDNN have maximum increases of 8.16%, 8.15%, 6.87%, 8.19%, 8.94%, 6.75%, 5.43%, and 4.82%, respectively, under low signal-to-noise ratios.

Table 3. Classification and recognition results of different frameworks with low SNRs in the RadioML2016.10b dataset. (**bold**: best)

Model	Average Accuracy (%)							
	−10 dB	**−8 dB**	**−6 dB**	**−4 dB**	**−2 dB**	**0 dB**	**2 dB**	**4 dB**
CNN-BiLSTM-DNN	**32.37**	**45.58**	**57.85**	**72.65**	**86.13**	**91.48**	**92.81**	**93.30**
MCLDNN	30.14	40.02	55.30	72.10	85.47	91.14	92.52	93.26
MCNet	27.82	38.86	53.03	67.85	78.71	84.74	87.38	88.48
CGDNet	31.76	40.83	51.79	67.53	80.12	87.01	89.96	90.07
ResNet	25.36	37.43	50.98	64.46	77.19	84.93	88.47	90.33
IC-AMCNet	24.21	37.73	54.13	68.55	81.71	88.38	90.53	91.77

Table 4 compares the classification results of different frameworks under the RadioML2016.10a dataset. The overall performance of the CNN-BiLSTM-DNN architecture is superior to that of the MCLDNN model at low SNRs, despite the accuracy at −6 dB and −2 dB being 0.14% and 1% worse than that of the MCLDNN model. The experimental results indicate improvement in recognition accuracy at low SNRs.

Table 4. Classification and recognition results of different frameworks with low SNRs in the RadioML2016.10a dataset. (**bold**: best)

Model	Average Accuracy (%)							
	−10 dB	**−8 dB**	**−6 dB**	**−4 dB**	**−2 dB**	**0 dB**	**2 dB**	**4 dB**
CNN-BiLSTM-DNN	**27.73**	**39.55**	56.18	**69.23**	80.05	**88.68**	**91.32**	**93.18**
MCLDNN	23.50	38.09	**56.32**	67.27	**81.05**	88.36	89.59	90.86
MCNet	22.45	38.00	56.00	65.27	72.45	78.23	79.09	82.27
CGDNet	18.64	33.36	51.95	64.95	74.95	79.27	81.59	84.00
ResNet	20.45	29.95	48.32	58.45	68.91	77.95	79.14	82.23
IC-AMCNet	19.59	33.18	53.00	63.59	73.00	80.23	81.55	84.77

4.2.2. Confusion Matrix

Matrices are used to characterize the degree of confusion between different features [27]. In the confusion matrix, different shades of color indicate the effect of classification, and the darker the diagonal color, the better the classification effect.

The confusion matrices for several network types with an SNR of 0 dB are displayed in Figures 6 and 7. Both AM-DSB and WBFM have a quiescent period, during which the signal is almost zero, and the network can only randomly classify invalid signals; as can be seen from the graph, the misclassification of the modulation format is mainly between AM-DSB and WBFM. The prediction accuracy of CNN-BiLSTM-DNN for these two types of modulation formats is significantly better than those of other models. CNN-BiLSTM-DNN improves the classification performance of most modulation formats compared with other models at low SNRs. Furthermore, 16QAM is also easily confused with 64QAM because their modulation types are very similar, but they use different orders, which makes the network unable to identify the two modulation methods accurately. After denoising the I/Q samples by the SP module, the influence of the modulation type on noise is reduced, and the CNN-BiLSTM-DNN model accurately identifies 97% of 16QAM and 98% of 64QAM under the RadioML2016.10b dataset. Under the RadioML2016.10a dataset, the CNN-BiLSTM-DNN model accurately identifies 90% of 16QAM and 94% of 64QAM, and the confusion between the two is significantly reduced, which exceeds the ability of other models to identify these two modulation types.

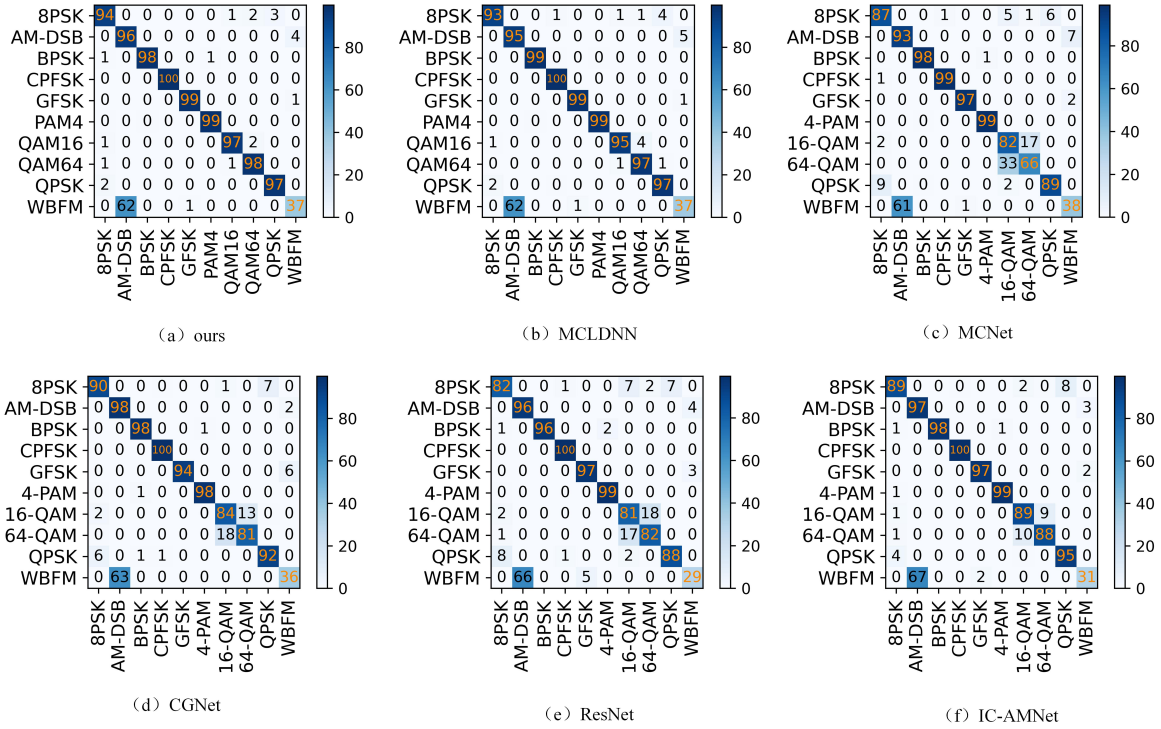

Figure 6. Confusion matrix of different models at 0 dB in the RadioML2016.10b dataset.

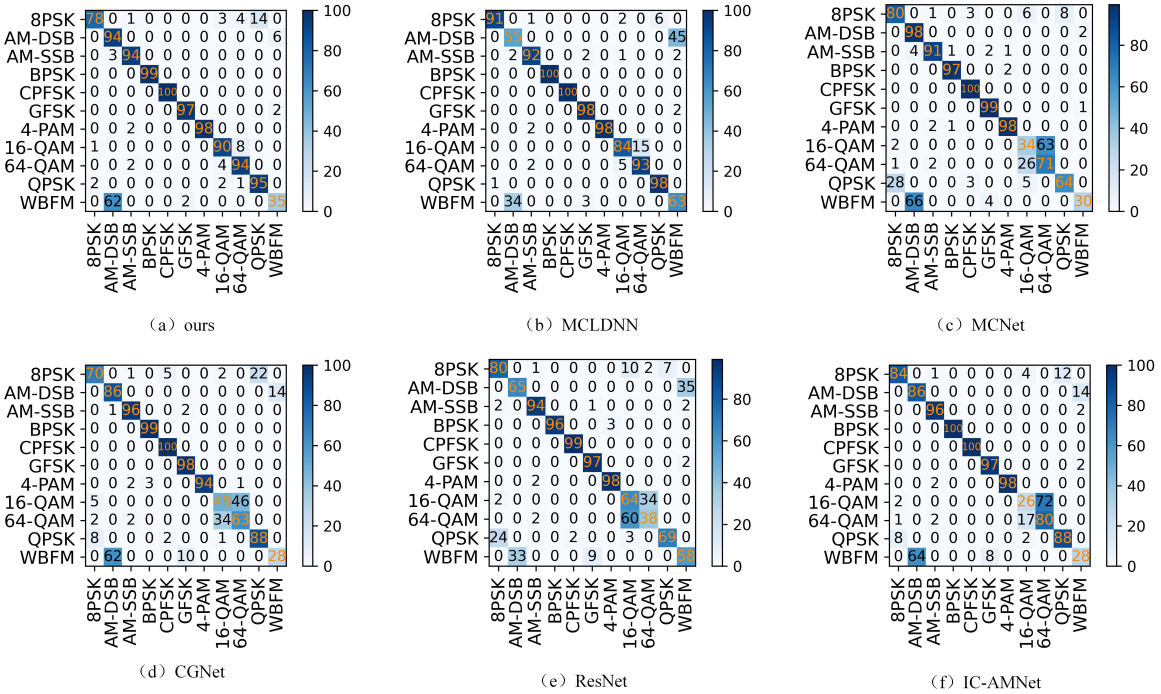

Figure 7. Confusion matrix of different models at 0 dB in the RadioML2016.10a dataset.

4.2.3. Ablation Experiment

To confirm the algorithm's efficacy, we carried out ablation experiments in which we experimented with models with attention mechanisms and models without attention mechanisms and compared the experimental results. All experiments used the conditions described in Section 4.1. Table 5 records the highest recognition accuracies and average accuracies of the study.

Table 5. Comparison of accuracy between models with and without an attention mechanism. (**bold**: best)

Dataset	Model	Highest Accuracy (%)	Average Accuracy (%)
RadioML2016.10b	With Attention	**93.79**	**64.76**
	Without Attention	93.53	64.25
RadioML2016.10a	With Attention	**93.18**	**62.73**
	Without Attention	91.32	60.98

The modulation recognition experiments of the proposed network model on the RML2016.10b and RML2016.10a datasets demonstrate that, from −20 dB to 18 dB, the average recognition accuracy of the model with an attention mechanism is 0.51% and 1.75% higher, respectively, than that of the model without an attention mechanism. The highest accurate recognition rate is increased by 0.26% and 1.86%, respectively, which shows that the model with an attention mechanism can identify the modulation mode better than the model without an attention mechanism.

4.2.4. Computational Complexity Comparison

Calculating the total number of parameters used in neural network training allows one to determine the complexity of a model. There are 806,392 parameters in the model overall, which affect the memory usage of the model and also determine its size. In Table 6, we present a comparison of computational complexity against other models using the RML2016.10a dataset. While our computational demands are higher, they are justified by improved average and peak recognition accuracies, as well as faster convergence speeds compared to alternative models. By leveraging the unique strengths of each component, our model excels at extracting and categorizing spatiotemporal features from complex datasets. This innovative approach addresses the limitations of traditional networks, which struggle to capture both spatial and temporal information simultaneously.

Table 6. Comparison of model computational complexity.

Model	Total Parameter Quantity	Training Time (Second/Epoch)	Training Epochs	Highest Accuracy (%)	Average Accuracy (%)
CNN-BiLSTM-DNN	806,392	30	69	93.18	62.73
MCLDNN	406,070	20	98	92.77	61.75
MCNet	121,611	15	101	83.50	56.20
CGDNet	124,933	7	188	84.00	56.21
ResNet	3,098,283	25	124	83.45	54.74
IC-AMCNet	1,263,882	6	235	84.91	56.30

5. Conclusions

Addressing the issue of low SNR modulation recognition accuracy, the SP module is used to denoise and preprocess the input signal to eliminate the noise influence on the I/Q data samples. A neural network model of CNN-BiLSTM-DNN is proposed, which is composed of a CNN, BiLSTM, an attention mechanism, and a fully connected deep neural network and uses their complementarity and synergy to extract and classify spatiotemporal features, which solves the problem of single spatial features and temporal features of sample signals extracted by traditional networks. The incorporation of an attention mechanism further enhances the model's ability to discern relevant patterns within the data, effectively filtering out noise and irrelevant information. The model performs noticeably better in recognizing and categorizing target signals of interest thanks to this focused attention on important details.

Experimental results indicate that the proposed method may effectively improve the modulation recognition accuracy for low SNR wireless communication signals. Modulation recognition experiments on the benchmark datasets RML2016.10b and RML2016.10a show that the average recognition accuracies of the proposed model from −20 dB to 18 dB are 64.76% and 62.73%, respectively, and the improvement ranges of modulation recognition accuracy are 0.29% to 5.56% and 0.32% to 4.23% when the SNR is −10 dB to 4 dB, respectively, and the computational complexity and training time are also increased. Future work will still be needed to streamline the algorithm and increase accuracy across a few perplexing modulation patterns in order to speed up training even more.

Author Contributions: Conceptualization: Z.L.; methodology: X.Z.; investigation: X.Z. and W.X.; writing–original draft preparation: X.Z. and W.X.; writing–review and editing: Z.L.; supervision: Z.L.; project administration: Z.L.; funding acquisition: Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 61801319; in part by Sichuan Science and Technology Program under Grant 2020JDJQ0061, 2021YFG0099; in part by Innovation Fund of Chinese Universities under Grant 2020HYA04001; in part by Engineering Research Center of Integration and Application of Digital Learning Technology, Ministry of Education under Grant 1321002; and in part by the 2022 Graduate Innovation Fund of Sichuan University of Science and Engineering under Grant Y2023288.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts to interest.

References

1. Guo, E.; Zhang, H.; Yang, L.; Liu, Y.; Peng, J.; Zhang, L. Radar radiation source signal recognition based on improved residual network. *Radio Eng.* **2022**, *52*, 2178–2185.
2. Ma, J.; Jing, Y.; Yang, Z.; Yang, H.; Wu, Z. ShuffleFormer: An efficient shuffle meta framework for automatic modulation classification. *Phys. Commun.* **2023**, *61*, 102226. [CrossRef]
3. Wu, M.; Gao, Y.; Tu, Y.; Qin, J.; Tang, Z.; Hu, F. Automatic modulation recognition of communication signals based on feature fusion and MACLNN. *Radio Eng.* **2022**, *52*, 1970–1976.
4. Xiao, W.; Luo, Z.; Hu, Q. A review of research on signal modulation recognition based on deep learning. *Electronics* **2022**, *11*, 2764. [CrossRef]
5. Hameed, F.; Dobre, O.A.; Popescu, D.C. On the likelihood-based approach to modulation classification. *IEEE Trans. Wirel. Commun.* **2009**, *8*, 5884–5892. [CrossRef]
6. Zheng, J.; Lv, Y. Likelihood-based automatic modulation classification in OFDM with index modulation. *IEEE Trans. Veh. Technol.* **2018**, *67*, 8192–8204. [CrossRef]
7. Han, L.; Gao, F.; Li, Z.; Dobre, O.A. Low complexity automatic modulation classification based on order-statistics. *IEEE Trans. Wirel. Commun.* **2016**, *16*, 400–411. [CrossRef]
8. Orlic, V.D.; Dukic, M.L. Multipath channel estimation algorithm for automatic modulation classification using sixth-order cumulants. *Electron. Lett.* **2010**, *46*, 1. [CrossRef]
9. Qu, Z.; Mao, X.; Deng, Z. Radar signal intra-pulse modulation recognition based on convolutional neural network. *IEEE Access* **2018**, *6*, 43874–43884. [CrossRef]
10. Downey, J.; Hilburn, B.; O'Shea, T.; West, N. Machine learning remakes radio. *IEEE Spectr.* **2020**, *57*, 35–39. [CrossRef]
11. Hong, S.; Zhang, Y.; Wang, Y.; Gu, H.; Gui, G.; Sari, H. Deep learning-based signal modulation identification in OFDM systems. *IEEE Access* **2019**, *7*, 114631–114638. [CrossRef]
12. Fei, S.C.; Zhang, C.P. Research on Modulation Recognition Method Based on Dual-Channel Fusion Network Model. *J. Shenyang Univ. Technol.* **2023**, *42*, 34–39+47.
13. Wang, X.; Liu, D.; Zhang, Y.; Li, Y.; Wu, S. A spatiotemporal multi-stream learning framework based on attention mechanism for automatic modulation recognition. *Digit. Signal Process.* **2022**, *130*, 103703. [CrossRef]
14. Lin, S.; Zeng, Y.; Gong, Y. Learning of time-frequency attention mechanism for automatic modulation recognition. *IEEE Wirel. Commun. Lett.* **2022**, *11*, 707–711. [CrossRef]

15. Zhang, F.X. Research on Automatic Modulation Recognition Technology Based on Deep Learning. *Univ. Electron. Sci. Technol. China* **2023**. [CrossRef]
16. Liang, J.; Li, X.; Liang, C.; Tong, H.; Mai, X.; Kong, R. JCCM: Joint conformer and CNN model for overlapping radio signals recognition. *Electron. Lett.* **2023**, *59*, e13006. [CrossRef]
17. Bouvrie, J. Notes on Convolutional Neural Networks. Technical Report, 2006. Available online: https://web-archive.southampton.ac.uk/cogprints.org/5869/1/cnn_tutorial.pdf (accessed on 3 May 2024)
18. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
19. Fang, C.; Sheng, Z.; Xia, M.; Zhou, H. Radar signal modulation recognition based on CNN-BiLSTM hybrid neural network. *Radio Eng.* **2024**, *54*, 1440–1445.
20. O'Shea, T.J.; West, N. Radio machine learning dataset generation with GNU radio. *Proc. GNU Radio Conf.* **2016**, *1*, 1.
21. Hu, G.; Li, P.; Lin, S.; Zong, B. Automatic modulation recognition algorithm based on phase transformation and CNN-BiLSTM. *Telecom Tech.* **2024**, 1–10. [CrossRef]
22. Zhang, F.; Luo, C.; Xu, J.; Luo, Y.; Zheng, F.-C. Deep learning based automatic modulation recognition: Models, datasets, and challenges. *Digit. Signal Process.* **2022**, *129*, 103650. [CrossRef]
23. 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]
24. Liu, X.; Yang, D.; El Gamal, A. Deep neural network architectures for modulation classification. In Proceedings of the 2017 51st Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 29 October–1 November 2017; pp. 915–919.
25. Xu, J.; Luo, C.; Parr, G.; Luo, Y. A spatiotemporal multi-channel learning framework for automatic modulation recognition. *IEEE Wirel. Commun. Lett.* **2020**, *9*, 1629–1632. [CrossRef]
26. Hermawan, A.P.; Ginanjar, R.R.; Kim, D.S. CNN-based automatic modulation classification for beyond 5G communications. *IEEE Commun. Lett.* **2020**, *24*, 1038–1041. [CrossRef]
27. Oikonomou, T.K.; Evgenidis, N.G.; Nixarlidis, D.G.; Tyrovolas, D.; Tegos, S.A.; Diamantoulakis, P.D.; Sarigiannidis, P.G.; Karagiannidis, G.K. CNN-Based Automatic Modulation Classification Under Phase Imperfections. *IEEE Wirel. Commun. Lett.* **2024**, *13*, 1508–1512. [CrossRef]