drones

MDPI

Article

Hybrid Data Augmentation and Dual-Stream Spatiotemporal Fusion Neural Network for Automatic Modulation Classification in Drone Communications

An Gong [1], Xingyu Zhang [1], Yu Wang [2,*], Yongan Zhang [1] and Mengyan Li [3]

[1] College of Computer Science and Technology, China University of Petroleum (East China), Qingdao 266580, China; 19930048@upc.edu.cn (A.G.); z21070263@s.upc.edu.cn (X.Z.); s21070021@s.upc.edu.cn (Y.Z.)
[2] College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China
[3] College of Communication Engineering, Hebei University of Science and Technology, Shijiazhuang 050091, China; limengyan2@nsfocus.com
* Correspondence: yuwang@njupt.edu.cn or 1018010407@njupt.edu.cn

Abstract: Automatic modulation classification (AMC) is one of the most important technologies in various communication systems, including drone communications. It can be applied to confirm the legitimacy of access devices, help drone systems better identify and track signals from other communication devices, and prevent drone interference to ensure the safety and reliability of communication. However, the classification performance of previously proposed AMC approaches still needs to be improved. In this study, a dual-stream spatiotemporal fusion neural network (DSSFNN)-based AMC approach is proposed to enhance the classification accuracy for the purpose of aiding drone communication because SDDFNN can effectively mine spatiotemporal features from modulation signals through residual modules, long-short term memory (LSTM) modules, and attention mechanisms. In addition, a novel hybrid data augmentation method based on phase shift and self-perturbation is introduced to further improve performance and avoid overfitting. The experimental results demonstrate that the proposed AMC approach can achieve an average classification accuracy of 63.44%, and the maximum accuracy can reach 95.01% at SNR = 10 dB, which outperforms the previously proposed methods.

Keywords: deep learning; attention mechanism; data augmentation; automatic modulation classification (AMC); spatiotemporal feature fusion; drone communication

check for updates

Citation: Gong, A.; Zhang, X.; Wang, Y.; Zhang, Y.; Li, M. Hybrid Data Augmentation and Dual-Stream Spatiotemporal Fusion Neural Network for Automatic Modulation Classification in Drone Communications. *Drones* **2023**, *7*, 346. https://doi.org/10.3390/drones7060346

Academic Editor: Emmanouel T. Michailidis

Received: 17 April 2023
Revised: 20 May 2023
Accepted: 23 May 2023
Published: 25 May 2023

1. Introduction

Automatic modulation classification (AMC) is the process of identifying the modulation types of communication signals, which has been widely applied in various communication systems for enhancing communication efficiency, ensuring security, and enabling safe and reliable drone operations [1–4]. In drone communications, it can also play an important role for distinguishing signals between drones, detecting unauthorized devices or signals, enabling automated control for optimal communication performance, and so on [5,6].

Traditional AMC methods can be classified into two categories: likelihood-based (LB) methods and feature-based (FB) methods. LB methods [7,8] typically involve significant computational complexity or require prior knowledge about the channel or noise [9]. FB methods [10,11] are based on expert features derived from time–frequency analysis and statistical theory. However, it is difficult to process massive signal samples in parallel, and the classification accuracy does not meet expectations. Moreover, with the rapid development of communication technologies, the electromagnetic environment has become increasingly complex, modulation schemes have become more diverse, and signal density has increased rapidly. As a result, traditional AMC methods are becoming increasingly unable to meet current application requirements.

In recent years, deep learning (DL) has been applied in wireless communication fields [12–18], and there are many DL-based AMC methods for addressing the challenges encountered in physical-layer wireless communications, including AMC [19–27]. In 2016, Shea et al. [28] first used convolutional neural networks for modulation classification for the first time, demonstrating the feasibility of neural networks in the field of AMC. In 2019, Ramjee et al. [29] extended the existing research by proposing three architectures for AMC: CLDNN, LSTM, and ResNet. The authors reported an classification accuracy of 90% in scenarios with high SNR using their proposed approach [29]. However, it is worth noting that despite incorporating both spatial and temporal features in the CLDNN model, the authors did not fully exploit the potential synergy between these features. This observation suggests that there is room for further improvement and optimization in utilizing spatial and temporal information effectively within the CLDNN architecture. In their 2020 study on automatic modulation classification (AMC), Zhang et al. introduced a novel dual-stream model that combines the strengths of CNN and LSTM networks [9]. This model diverges from previous research by simultaneously incorporating both temporal and spatial features of the data, thereby effectively capturing feature interactions and spatiotemporal characteristics inherent in complex raw time signals. However, despite the utilization of a dual-stream architecture, it should be noted that the LSTM+CNN structure employed in this study may still be considered relatively simplistic, limiting its ability to fully exploit and extract more refined and effective features. In 2021, Liao et al. [30] presented a novel and efficient end-to-end learning model for automatic modulation classification that learnt from time domain in-phase and quadrature data [30]. The model demonstrated improved accuracy and reduced training and prediction time. However, since the authors only utilized IQ data for training, the accuracy improvement was limited. Chang et al. proposed a multitask learning-based deep neural network, for modulation classification [31]. This network effectively integrates I/Q data and A/P data, achieving high classification accuracy. However, the model encounters a challenge in distinguishing between WBFM and AM-DSB at high SNR. Table 1 presents an evaluation of the strengths and weaknesses of the related work.

Table 1. Related Works.

Related Works	DNN	Strength	Weakness
Shea et al. [28]	CNN	Innovative use of deep learning for modulation recognition achieved significant accuracy improvement compared to traditional methods.	Only explored the application of CNN for modulation recognition.
Ramjee et al. [29]	CLDNN	Both time and spatial features were extracted from IQ signals, resulting in a more diverse feature set.	Using only IQ data for feature extraction results in insufficiently diverse feature sets.
	ResNet	Further exploration was conducted on top of CNN.	The time feature of IQ data was neglected.
	LSTM	RNNs were utilized to extract time information from IQ signals for modulation recognition.	The absence of convolutional neural networks (CNNs) for spatial feature extraction is a limitation.
Zhang et al. [9]	CNN-LSTM	Features were extracted separately from IQ and AP data for modulation recognition.	Extracting temporal features from the spatial features extracted by CNN may have an impact on the accuracy of modulation recognition.
Liao et al. [30]	SCRNN	The accuracy of the model is ensured while reducing the required training time.	Extracting features solely from IQ data limits the diversity of the features.
Chang et al. [31]	MLDNN	Explored the interaction features and temporal information of both IQ and AP data, which resulted in a significantly improved accuracy rate.	The WBFM modulation scheme is highly susceptible to misclassification as AM-DSB.
	CGDNN2	The parameter estimator and the parameter transformer were introduced, resulting in a significant reduction in the model's parameter count.	The design of the model architecture lacks significant innovation, impeding the extraction of improved features.

In this study, we propose an AMC method using hybrid data augmentation and a dual-stream spatiotemporal fusion neural network (DSSFNN), where the former is to expand training samples to prevent model overfitting, while the latter is a parallel architecture to extract the spatiotemporal features for high classification performance. In detail, the spatial feature extraction branch is responsible for IQ data, while the temporal feature extraction branch is designed for AP data. The features extracted from these branches are fused for modulation classification. The contributions of the paper are listed as follows:

- We propose a hybrid data augmentation method based on phase shift and self-perturbation, which can effectively expand training samples without introducing additional information.
- We propose a DSSFNN structure for AMC, which can extract features from both the spatial and temporal dimensions of data. Compared to the single-dimensional feature extraction method, the features extracted by DSSFNN are more diverse and effective, which improve the accuracy of AMC.

The remaining parts of this paper are as follows: Section 2 elaborates on the problem formulation. In Section 3, a detailed description of the proposed AMC method is provided, including the data augmentation technique and the dual-stream spatiotemporal fusion neural network (DSSFNN) architecture. Section 4 presents the simulation results and analysis. Finally, in Section 5, the conclusions drawn from the study are presented, highlighting the contributions of the proposed method.

2. Problem Formulation

2.1. Signal Model

The complex baseband signal model [32] can be represented equivalently without losing generality as follows:

$$x(t) = s(t) * h(t) + n(t),\ t \in [1, T] \tag{1}$$

where the received signal is represented by $x(t)$, $s(t)$ represents the modulated signal, the channel is represented using $h(t)$, and $n(t)$ represents zero-mean complex AWGN with a bilateral power spectral density of $N_0/2$ [32].

2.2. Dual-Stream Data

In this paper, an IQ signal along with the AP data were used as the training data for model training. In general, signal reception equipment can be used to receive signals in a communication channel and store them in the format of IQ data. The AP data can then be obtained from the IQ data using mathematical formulas. This approach is commonly used in the field of digital signal processing for wireless communication systems. The model for storing signal data in IQ format is shown as follows:

$$x[i] = x_I[i] + jx_Q[i] \tag{2}$$

where, $x_I[i]$ and $x_Q[i]$ represent the real and imaginary parts of the IQ signal of the i-th signal, respectively. By decomposing IQ data into in-phase and quadrature components and calculating their amplitudes and phases, one can obtain corresponding AP data. This process can be represented as:

$$A_i = \sqrt{I_i^2 + Q_i^2} \tag{3}$$

$$\varphi_i = \arctan\left(\frac{Q_i}{I_i}\right) \tag{4}$$

where I_i and Q_i represent the real and imaginary parts of the IQ signal, A_i represents the amplitude of the ith data in AP data, and φ_i represents the phase of the ith data in AP data.

2.3. Problem Description

Modulation classification is the process of determining the modulation scheme used by a received signal based on the sampled signal sequence $x = [x(1), x(2), \cdots, x(T)]$, from a candidate set $M = [M_1, M_2, \cdots, M_N]$ of N modulation schemes.The deep-learning-based modulation classification scheme can be represented as [32]:

$$\hat{M} = \arg\max f(M_i|x; W), M_i \in M \tag{5}$$

where $\hat{M}$ represents the predicted value of the modulation classification type, M_i represents the true value of the modulation classification type, M represents the set of modulation schemes [32], $f(W)$ represents the deep learning model that maps the signal sample x to the modulation classification type $\hat{M}$, and W represents the parameter weights of the model. The deep-learning-based modulation classification scheme can be simplified as the task of obtaining a high-precision deep learning model $f(W)$.

3. Our Proposed Robust AMC Method

3.1. The Framework of the Proposed Method

Our proposed robust AMC method based on hybrid data augmentation and dual-stream spatiotemporal fusion neural network is illustrated in Figure 1.

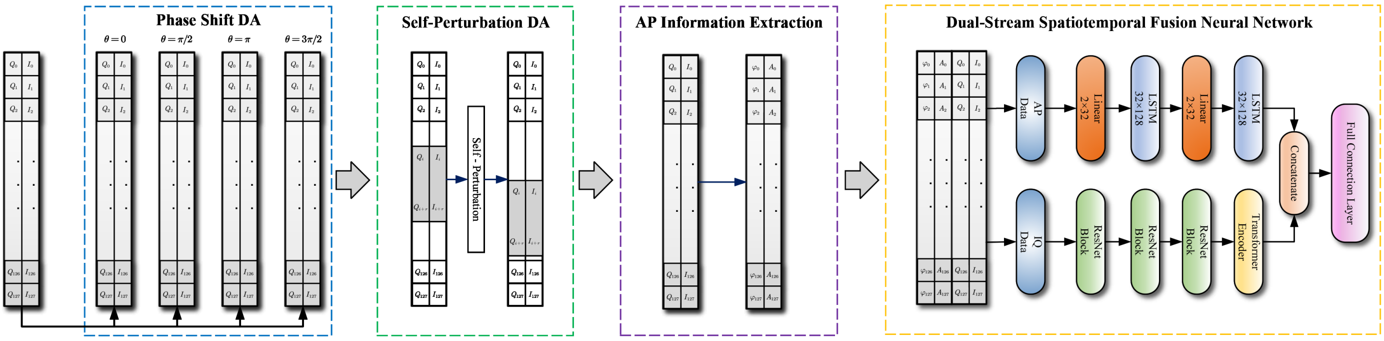


Figure 1. The overall architecture of automatic modulation classification method based on hybrid data augmentation and dual-stream spatiotemporal fusion neural network.

The proposed scheme consists of three key parts: hybrid data augmentation, AP information extraction, and dual-stream spatiotemporal fusion neural network. The hybrid data augmentation includes phase transformation data augmentation and self-perturbation data augmentation. The IQ data is fed into the model as the raw input, which first goes through the hybrid data augmentation part. This increases the amount of data and makes the data features more rich and diverse. After that, the augmented data are used to extract amplitude and phase information. The IQ data and AP data after data augmentation will be fed into the dual-stream spatiotemporal fusion neural network for modulation classification. The IQ data are fed into the spatial feature extraction branch for extracting spatial features. The AP data are fed into the temporal feature extraction branch to extract temporal features. The spatial features and temporal features will be fused at the end and fed into a fully connected layer for classification.

3.2. Hybrid Data Augmentation

In real-world scenarios, due to the complex electromagnetic environment, the received signals by the receiver are often not as satisfactory as expected. At the same time, deep learning models often fail to extract good features and are prone to overfitting due to insufficient training samples. To enhance the robustness and generalization ability of the trained deep learning model, a data augmentation algorithm is proposed in this paper. The proposed algorithm performs phase transformations on the original data to generate data samples at different phases, thereby increasing the quantity of the training data and effectively preventing the occurrence of model overfitting. Next, the augmented data will be subjected to self-perturbation data augmentation, a method that enhances data diversity and helps the model learn different features.

3.2.1. Phase-Shift Data Augmentation

Phase transformation is a simple and effective data augmentation method in the field of modulation classification. By varying the phase angle, data can be obtained at different phase angles, thereby achieving the purpose of data augmentation. The phase-shift data augmentation process can be represented as [33]:

$$\begin{bmatrix} R(\tilde{x}) \\ \mathcal{L}(\tilde{x}) \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} R(x) \\ \mathcal{L}(x) \end{bmatrix} \tag{6}$$

where x denotes the original data, $\tilde{x}$ denotes the augmented data, $R()$ and $\mathcal{L}()$ represent the operations performed on the real and imaginary parts, and θ takes the values 0, $\frac{\pi}{2}$, π, and $\frac{3\pi}{2}$.

3.2.2. Self Perturbation Data Augmentation

Data augmentation through self-perturbation refers to the process of randomly cropping a portion of the data and then splicing it back into the remaining data. Assuming the data to be augmented with self-perturbation are denoted as D, the remaining data after cropping are denoted as D_{cut}, the length of the remaining data after cropping is denoted as $|D_{cut}|$, and a random segment taken from the original data is denoted as s. The process of self-perturbation data augmentation can be represented as follows:

$$D_{OutPut} = D_{cut[0:p]} + s + D_{cut[p:|D_{cut}|]} \tag{7}$$

where D_{OutPut} represents the output of the self-perturbation data augmentation algorithm, and p represents a random position within D_{cut}.

The self-perturbation algorithm involves cropping some parts of a sequence and adding them to random positions, which expands the data while enriching its features in automatic modulation classification. This approach has the potential to improve the model's generalization performance—its ability to perform well on data outside the training set. Some advantages of this algorithm include:

- Enhancing data diversity: The self-perturbation algorithm enhances the robustness and generalization ability of a model by adding new variations to the dataset. This augmentation of data diversity can enable the model to better capture the features of the dataset and improve its accuracy.
- Reducing overfitting: Overfitting is a common problem in machine learning, and the self-perturbation algorithm can reduce the risk of overfitting by increasing the size of the dataset. This is because training the model on more data can help to better learn the true distribution of the dataset.
- Simplicity and ease of implementation: The self-perturbation algorithm is relatively simple to implement, requiring only a small amount of manipulation on the original data. Compared to other complex data augmentation techniques, self-perturbation algorithm has lower implementation costs and higher practicality.
- No introduction of additional information: The self-perturbation algorithm achieves data augmentation by cropping parts of the original data and then splicing them together. This approach ensures that no additional information is introduced into the data. In contrast, adding noise as a form of data augmentation introduces additional information that may sometimes affect classification performance.

3.3. Dual-Stream Spatiotemporal Fusion Neural Network

3.3.1. Spatial Feature Extraction Module for IQ Data

The proposed AMC method includes a spatial feature extraction module for IQ data, which is based on ResNeXt [34] and a self-attention mechanism [35]. ResNeXt is used to compute the real and imaginary parts of the IQ signal, which are then used to extract features from the spatial dimension of the IQ data [34]. The self-attention mechanism is employed to weight the feature maps and enhance the discriminability of the features [35]. This spatial feature extraction module plays a crucial role in the overall AMC method, as it enables the extraction of informative features from the IQ data, which are then used for modulation classification. ResNeXt can be represented as:

$$y = H(x) + F(x) \tag{8}$$

where x denotes the input data, $H(x)$ represents the mapping function, and $F(x)$ refers to the residual block. The residual block can be expressed as:

$$y = F(x, \{W_i\}) + x \tag{9}$$

In the residual block represented above, $F(x, \{W_i\})$ denotes the mapping function, and W_i represents the weight parameters. To improve the performance and efficiency of the network, ResNeXt introduces grouped convolutions into the residual block $F(x, \{W_i\})$. The input data x are divided into several groups with the same number of channels; then, a convolution operation is performed on each group of data. Finally, the convolution results of each group are merged. Grouped convolutions can be expressed as:

$$Y_i = \sum\nolimits_{j \in Group_i} K_j * X_j \tag{10}$$

where the notation X represents the input data, Y represents the output data, K denotes the convolution kernel, $Group$ denotes the partitioning of the input data into multiple groups, i denotes the i-th group, and j denotes the channel within each group [34].

At the final stage of the spatial feature extraction module, a self-attention mechanism was employed [35]. The purpose is to feed the extracted features into a self-attention mechanism, with the aim of computing weights to enhance the more salient features. The calculation process of the self-attention mechanism can be divided into three parts: computing the attention scores, computing the weighted sum, and computing the output. For any element x_i in the input data, the formula for calculating its attention scores a_i with respect to other elements can be expressed as:

$$a_i = soft\max\left(\frac{q_i K^T}{\sqrt{d_k}}\right) \tag{11}$$

where $q_i, K \in \mathbb{R}^{d_k}$ represents the elements in the input sequence, which represent queries and keys, and d_k represents the dimensionality. The attention score a_i represents the relevance between the ith element in the input sequence and other elements. The function $soft\max$ is used to normalize the attention scores. Using the attention scores a_i, each element in the input data can be weighted and summed to obtain a weighted sum z, which can be represented as:

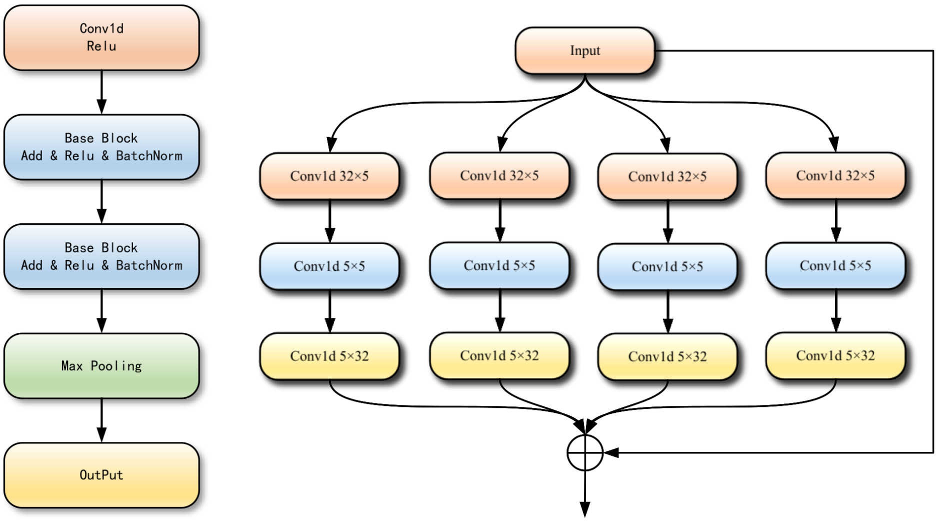
$$z = \sum_{i=1}^{n} a_i v_i \tag{12}$$

where $v_i \in \mathbb{R}^{d_v}$ represents the value of the i-th element in the input sequence, and d_v is the dimension of the value representation. The output sequence y is obtained by applying a linear transformation and a non-linear activation to the weighted sum z:

$$y_i = \text{Re}LU(W_o z + b_o) \tag{13}$$

where $W_o \in \mathbb{R}^{d_h \times d_v}$ and $b_o \in \mathbb{R}^{d_h}$ are weight matrices and bias vectors used for linear transformation. d_h denotes the dimensionality of the output representation, and ReLU is a non-linear activation function.

In the spatial feature extraction branch of DSSFNN, three ResNet blocks are stacked, each consists of a one-dimensional 2 × 32 convolutional layer, two base blocks, and a max pooling layer. After each base block in the spatial feature extraction branch of DSSFNN, a ReLU activation layer and a one-dimensional batch normalization layer are added to prevent overfitting. We made appropriate adjustments to the grouping convolutions proposed by Xie et al. in 2017 [34]. Specifically, we designed four branches for each base block, with each branch composed of three convolutional layers. After passing through the three convolutional layers in each branch, the data from the four branches are combined. In addition, we also added a shortcut connection between the input and output of the base block, which can help reduce the occurrence of gradient explosion and vanishing problems while deepening the network, as well as accelerate the convergence speed of the model. The detailed structures of the ResNet block and base block are illustrated in Figure 2.



(**a**) ResNet block (**b**) Base block

Figure 2. ResNet block and base block architecture.

We incorporated a Transformer Encoder with a self-attention mechanism as the core into the end of the spatial feature extraction branch. By utilizing the characteristics of the self-attention mechanism, the feature weights of the extracted features are calculated, which enhances the core features and accelerates convergence while improving the accuracy of modulation classification. The excellent performance of the self-attention mechanism can mainly be attributed to the following aspects:

- Global information: the self-attention mechanism [35] can consider the entire sequence of information while processing the information at each position.
- Interpretability: the self-attention mechanism [35] can increase the interpretability of the model by assigning different weights to information from different positions.
- Addressing long-range dependencies: the self-attention mechanism can solve the problem of long-range dependencies, where the model is capable of correctly processing distantly related contextual information.
- Powerful feature representation capability: the self-attention mechanism can fuse information from different positions to obtain powerful feature representation capability.

3.3.2. Time Feature Extraction Module for AP Data

The time feature extraction module for AP data is constructed based on the LSTM model [36]. The LSTM model is capable of extracting temporal features from both phase and amplitude information, while addressing the issues of gradient vanishing and exploding in traditional RNN models. The LSTM architecture comprises a memory cell and three gating components, namely, the input gate, output gate, and forget gate.

Due to the limited feature information contained in low-dimensional word vectors, in order to enhance the LSTM's ability to extract temporal information, the input data need to be first expanded with word vector extensions in the time feature extraction module. Suppose the length of the input AP data is T; then, the shape of the input data X is $T \times 2$. The word vector expansion can be represented as follows:

$$Y = WX + b \tag{14}$$

where W represents a linear layer weight matrix of size $2 \times E$, b represents a bias vector, E represents the dimensionality of the extended word vectors, and Y denotes the extended word vectors. With word vector expansion, the input data are expanded from $T \times 2$ to $T \times E$, as shown in Figure 3.

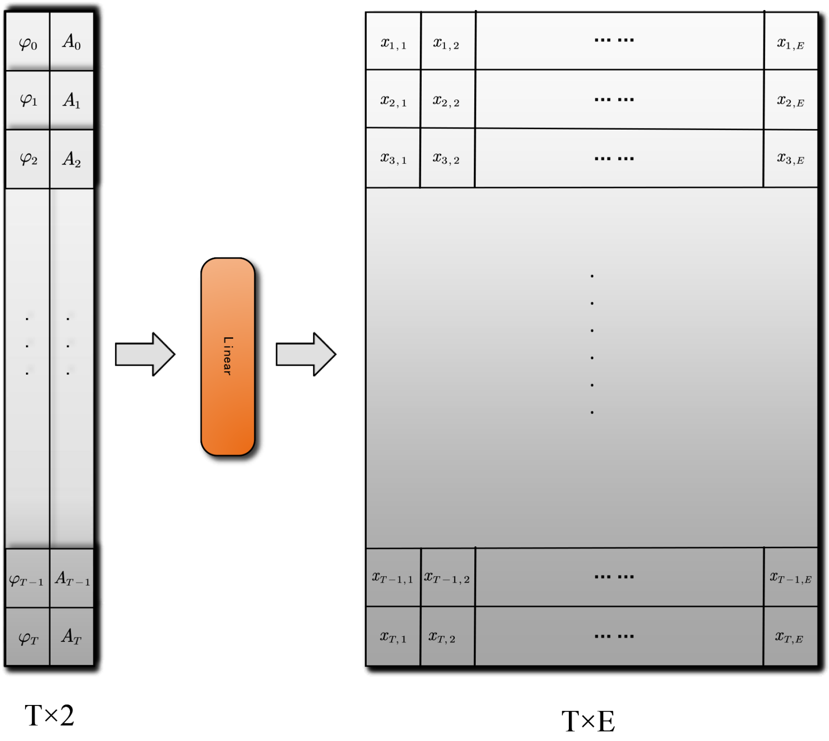

φ_0 A_0
φ_1 A_1
φ_2 A_2
φ_{T-1} A_{T-1}
φ_T A_T
Linear
$x_{1,1}$ $x_{1,2}$ …… $x_{1,E}$
$x_{2,1}$ $x_{2,2}$ …… $x_{2,E}$
$x_{3,1}$ $x_{3,2}$ …… $x_{3,E}$
$x_{T-1,1}$ $x_{T-1,2}$ …… $x_{T-1,E}$
$x_{T,1}$ $x_{T,2}$ …… $x_{T,E}$
T×2
T×E

Figure 3. Word vector expansion.

The time feature extraction module consists of two layers of LSTM. After passing through the first layer of LSTM, the model will extract the data outputted by the output gate of the last LSTM unit. The extracted data are then subjected to another word vector expansion operation and fed into the next layer of LSTM for further temporal feature extraction. Finally, the outputted information is sent to the feature fusion module for feature fusion.

3.3.3. Spatiotemporal Feature Fusion Mechanism

In the feature fusion module, we concatenated the temporal and spatial features and then fed the concatenated feature vector as input to a linear layer for further processing. Specifically, assuming the dimensionality of the temporal features is d_t and the dimensionality of the spatial features is d_s, we concatenate them along the feature dimension to obtain a new feature vector of dimensionality $d_{t+s} = d_t + d_s$. The resulting new feature vector obtained by concatenating the temporal features and spatial features is fed into a linear layer, where it undergoes a linear transformation to obtain a new feature vector that can be used for further task processing. This process can be represented as follows:

$$H = \text{Linear}([x_t, x_s]) \tag{15}$$

where the notation $[x_t, x_s]$ denotes the concatenation of temporal and spatial features, while $\text{Linear}(\cdot)$ represents the linear transformation applied to this concatenated feature vector.

3.3.4. Loss Functions and Optimization Algorithms

In this paper, the cross-entropy loss function is adopted as the objective function to solve the multi-class classification problem. As for the optimizer, the AdamW optimizer is used with a training cycle of 128 and a learning rate of 0.001.

4. Experimental Results

The experimental results of the proposed automatic modulation classification method (AMC) are presented in this section of the paper, which includes an evaluation of the proposed hybrid data augmentation algorithm for the dual-stream spatiotemporal fusion neural network (DSSFNN) model. The classification accuracy is evaluated with and without the hybrid data augmentation algorithm. This study also investigates the optimal architecture of the DSSFNN model, evaluating the necessity of the number of LSTM layers, ResNet structure, and self-attention mechanism. The performance of the proposed approach is compared with other state-of-the-art models in terms of classification accuracy, and the findings indicate that the proposed method surpasses the performance of the existing methods. Additionally, the classification performance of the DSSFNN model on different modulation types under various SNR conditions is analyzed, showing that the proposed method is effective in real-world scenarios.

4.1. Simulation Environment, Parameters, and Performance Metrics

To demonstrate the performance of the proposed AMC scheme, the dataset used in this paper to evaluate the proposed scheme is the RML2016.10a open radio machine learning dataset. This dataset contains 220,000 samples comprising 11 modulation types, each with 20 SNR levels ranging from −20 dB to 18 dB [28]. Each sample includes two signal components, I and Q, each with 128 samples per component. During the experiments, 70% of the data sets were randomly assigned to the training set, while 15% were assigned to the validation set and 15% were assigned to the test set. In our experiments, the training environment employed a Windows 11 operating system, with an NVIDIA RTX 3060 GPU utilized for training the models. Python was used as the programming language, and the deep learning models were constructed using the PyTorch framework.

4.2. Ablation Experiment of Hybrid Data Augmentation Algorithm

Figure 4 is referenced in this paper to present comparative results between the DSSFNN model trained by the hybrid data augmentation method and other methods. The outcomes exhibit a notable improvement in the classification accuracy of the DSSFNN model trained by the hybrid data augmentation approach as compared to the model trained without it. This finding highlights the effectiveness of the hybrid data augmentation method in improving the classification accuracy of AMC.

When -20 dB $\leqslant$ SNR $\leqslant$ 18 dB, the mean classification accuracy of the DSSFNN model trained without hybrid data augmentation is 60.90%. The DSSFNN model trained solely with the self-perturbation data augmentation scheme attained a mean accuracy rate of 62.52%. The mean classification accuracy attained by the DSSFNN model trained exclusively by the phase-shift data augmentation scheme is 62.60%. The DSSFNN model trained with the data augmentation approach proposed in this paper achieved a mean classification accuracy of 63.44%. The proposed hybrid data augmentation approach improved the mean classification accuracy of DSSFNN by 2.54%. The experimental outcomes validate the efficacy of the suggested approach in improving the classification accuracy of the DSSFNN model.

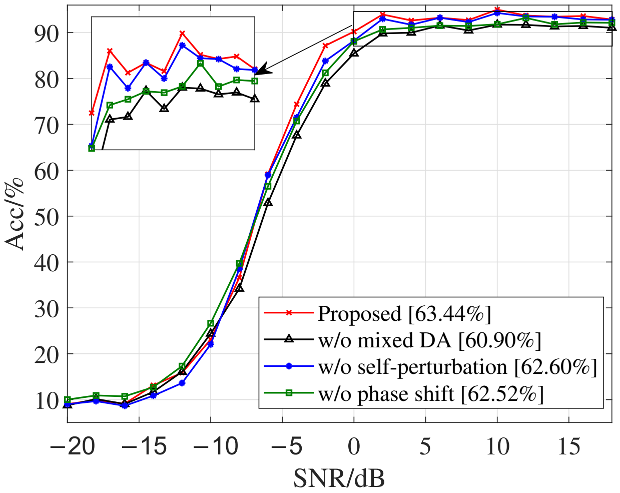


Figure 4. Performance comparison of DSSFNN modulation classification under various data augmentation methods.

4.3. Ablation Experiment of DSSFNN

Figure 5 illustrates the changes in modulation classification accuracy of the DSSFNN model under different structures of the base block. It can be observed from the graph that the highest classification accuracy of 93.75% is achieved when the number of grouped convolutions is two. When the number of grouped convolutions is set to four, the highest classification accuracy achieved by the model is 95.01%. On the other hand, when the number of grouped convolutions is set to six, the model's highest classification accuracy is 93.81%. When the number of grouped convolutions exceeded four, the accuracy of the model decreased slightly. The experimental results suggest that increasing the number of branches in the DSSFNN model leads to a slight decrease in the classification accuracy after a certain threshold. Therefore, in this study, the number of grouped convolutions in the base block of the DSSFNN model was set to four.

In addition, we also conducted experiments on DSSFNN without using group convolutions. The results show that the DSSFNN model with group convolutions achieved significantly higher accuracy compared to the one without group convolutions.

Figure 6 illustrates the variation in the DSSFNN modulation classification accuracy for different numbers of LSTM layers and after deleting the Transformer Encoder. When the Transformer Encoder was removed from the DSSFNN model, a significant decrease in accuracy was observed. At a high signal-to-noise ratio (SNR), the average accuracy of the model was only 90.89%. When a single layer of LSTM was used, the DSSFNN model achieved an average accuracy of 92.00%. Compared with the models without attention block and with a single LSTM layer, the proposed DSSFNN model in this paper improved the average accuracy by 1.91% and 0.8%, respectively, under high signal-to-noise ratio conditions. The experimental results demonstrate that using a double-layer LSTM in the DSSFNN model can achieve the best classification accuracy. Moreover, the presence or absence of a self-attention mechanism has a significant impact on the accuracy of the DSSFNN model. Table 2 shows the results of the ablative experiments of DSSFNN under a high SNR case.

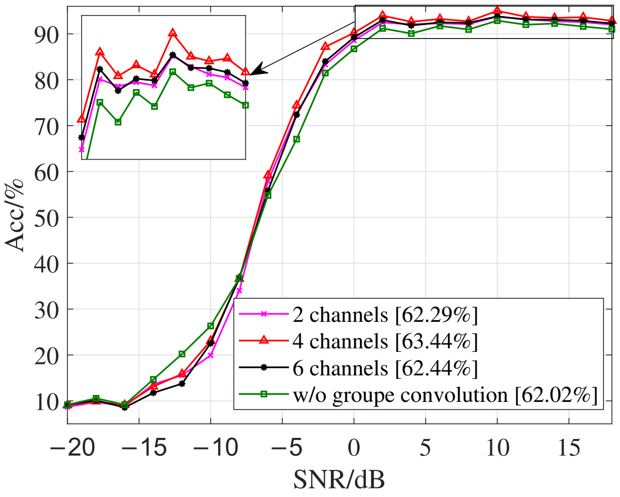


Figure 5. Performance comparison of DSSFNN modulation classification using different base block architectures.

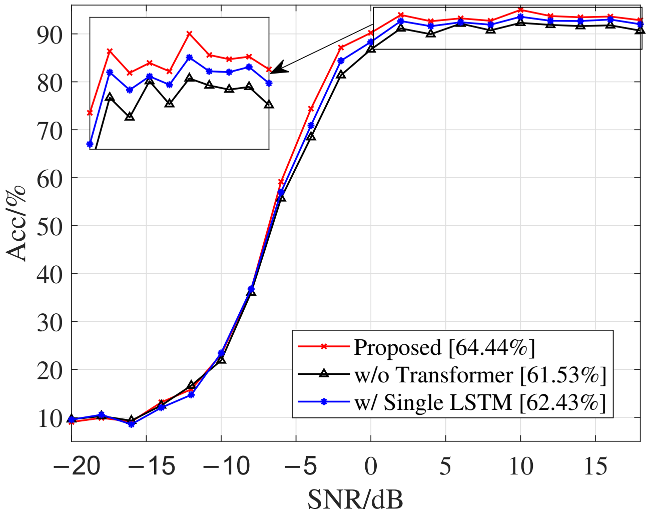


Figure 6. Performance comparison of different LSTM layer numbers and self-attention mechanisms on the accuracy of DSSFNN model.

Table 2. DSSFNN experimental results of ablation study under different SNR scenarios.

Models	**Results under Different SNR Scenarios**										**Average ACC**	**Max ACC**
	0	**2**	**4**	**6**	**8**	**10**	**12**	**14**	**16**	**18**		
DSSFNN	90.21	93.95	92.64	93.24	92.73	95.01	93.72	93.46	93.62	92.85	93.14	95.01
DSSFNN w/o DA	85.44	89.81	89.98	91.55	90.45	91.73	91.68	91.33	91.44	91.04	90.45	91.73
DSSFNN w/ phase-shift DA	88.24	92.98	91.70	93.24	92.28	94.28	93.50	93.45	92.86	92.80	92.54	94.28
DSSFNN w/ self-perturbation DA	88.08	90.68	91.04	91.51	91.43	91.81	93.20	91.80	92.20	92.13	91.39	93.20
DSSFNN w/ 2 channels	88.54	92.46	92.06	92.28	92.10	93.75	93.16	92.73	92.54	91.99	92.17	93.75
DSSFNN w/ 6 channels	89.21	93.01	91.82	92.48	92.38	93.81	93.09	93.07	92.84	92.24	92.40	93.81
DSSFNN w/o grouped convolutions	86.73	91.17	90.07	91.72	90.94	92.87	91.99	92.23	91.59	91.01	91.03	92.87

4.4. Our Proposed Method vs. Existing Methods

Figure 7 illustrates a comparison between the proposed approach and existing modulation classification methods, including ResNet [29], CLDNN [29], CNN-LSTM [9], SCRNN [30], CNN4 [31], MLDNN [31], CGDNN [31], and the proposed DSSFNN model. Among the models compared, MCLDNN achieved the lowest classification accuracy. Specifically, when 0 dB $\leqslant$ SNR $\leqslant$ 18 dB, the mean accuracy achieved by MCLDNN was 81.06%, with a maximum classification accuracy of 81.74%. ResNet attained a mean accuracy of 82.82% at an SNR of $[0, 18]$ dB. The average classification accuracy of CNN4 was 83.36%, with a maximum accuracy of 84.8%. The average accuracy of the LSTM-CNN dual-stream model was only 60.97%, which partially demonstrates the effectiveness of the proposed model in this paper. Comparatively, under high SNR circumstances, SCRNN, CLDNN, and CGDNN demonstrated exceptional performance. The average classification accuracy of SCRNN was 89.92%. The average classification accuracy of CLDNN was 90.61%. The average classification accuracy of CGDNN2 was 91.65%. The modulation classification scheme proposed in this paper achieves the utmost classification accuracy. Specifically, at high SNR, the proposed modulation classification scheme achieved an accuracy rate of 93.14%, with the highest accuracy rate being 95.01%. The experimental results indicate that the modulation classification scheme proposed in this paper can achieve a relatively advanced level and outperforms other schemes.

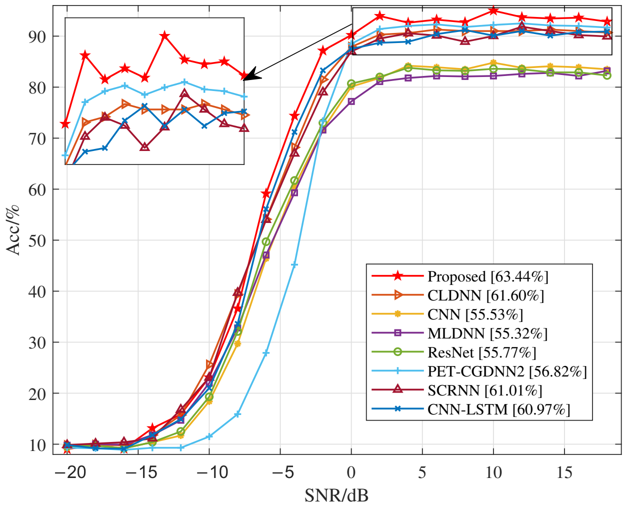
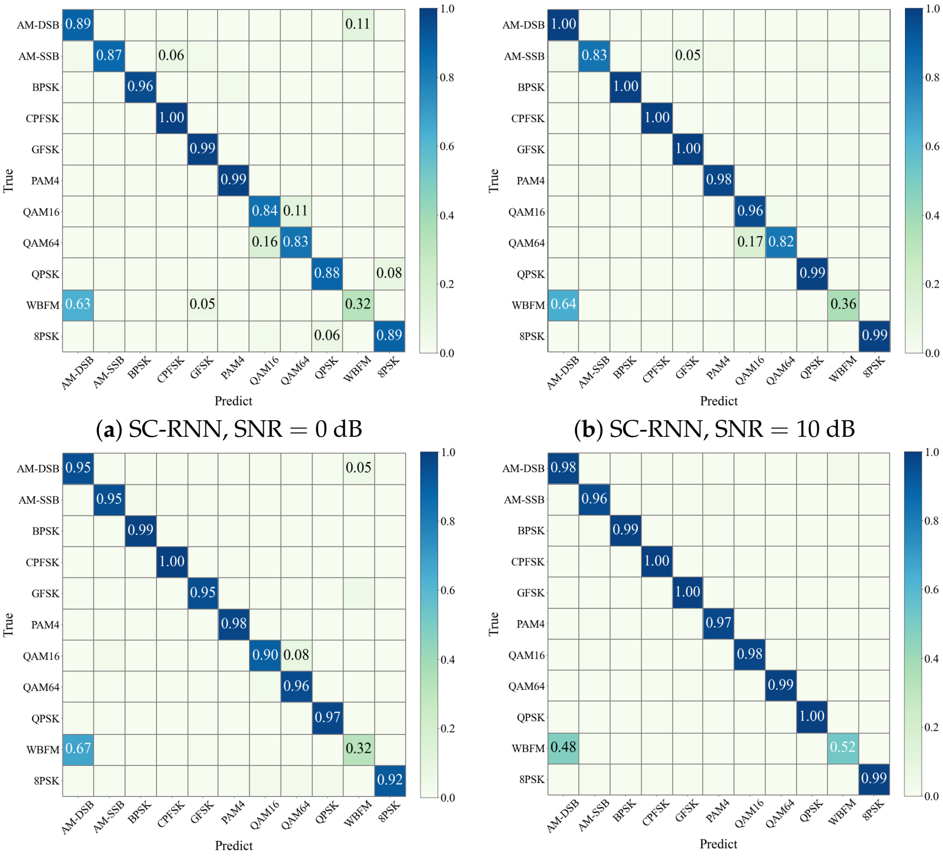


Figure 7. Comparison of modulation classification results between DSSFNN and other models.

Figure 8 shows a comparison of classification results obtained by the proposed modulation classification scheme and SCRNN at SNR = 10 dB. Based on the figure, it can be observed that at aaa, the proposed modulation classification scheme in this paper shows significant improvement compared to SCRNN in the classification of QAM64 and WBFM modulation schemes, with an increase of 17% and 16%, respectively. However, the modulation classification accuracy of the proposed scheme for WBFM modulation style is still not high enough. Therefore, how to improve the classification accuracy of deep learning models for WBFM modulation style is a research problem worthy of further investigation.



(**a**) SC-RNN, SNR = 0 dB (**b**) SC-RNN, SNR = 10 dB

(**c**) Our proposed, SNR = 0 dB (**d**) Our proposed, SNR = 10 dB

Figure 8. Modulation classification comparison of SC-RNN and proposed model under different SNR scenarios.

5. Conclusions

This paper proposed a robust AMC method based on data augmentation and deep learning models, which achieves high-precision classification of signal modulation methods. Firstly, a hybrid data augmentation method was selected to augment the original data. By using data augmentation, the trained model can have higher robustness and generalization ability and can effectively suppress overfitting during the training process. Additionally, it should be noted that AMC plays a crucial role in drone communication due to the requirement of reliable data transmission between drones and ground stations. Next, this paper proposed a novel AMC method, DSSFNN, which adopts a parallel design to extract both temporal and spatial features separately and fuses them for high-precision classification of most modulation schemes. A method for spatial and temporal feature extraction, based on two independent branches and dual-stream components, was developed. This approach ensures diversity between spatial and temporal features while preventing the extraction of features from previously extracted ones, effectively eliminating potential factors that may degrade model performance. By conducting experiments on the publicly available dataset RML2016.10a and comparing with existing models, the proposed modulation classification scheme in this paper achieved the highest classification accuracy, reaching an advanced level in terms of accuracy.

Author Contributions: Conceptualization, A.G. and X.Z.; methodology, X.Z.; software, X.Z.; validation, A.G., X.Z. and Y.W.; formal analysis, Y.Z. and M.L.; investigation, X.Z.; resources, A.G.; writing—original draft preparation, X.Z.; writing—review and editing, A.G. and Y.W.; project administration, A.G.; funding acquisition, A.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are openly available in [RML2016.10a] at [28].

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

References

1. Gui, G.; Liu, M.; Tang, F.; Kato, N.; Adachi, F. 6G: Opening New Horizons for Integration of Comfort, Security, and Intelligence. *IEEE Wirel. Commun.* **2020**, *27*, 126–132. [CrossRef]
2. Huang, H.; Liu, M.; Gui, G.; Haris, G.; Adachi, F. Unsupervised learning-inspired power control methods for energy-efficient wireless networks over fading channels. *IEEE Trans. Wirel. Commun.* **2022**, *21*, 9892–9905. [CrossRef]
3. Ohtsuki, T. Machine learning in 6G wireless communications. *IEICE Trans. Commun.* **2022**, *106*, 75–83. [CrossRef]
4. Xu, Y.; Gui, G.; Gacanin, H.; Adachi, F. A Survey on Resource Allocation for 5G Heterogeneous Networks: Current Research, Future Trends, and Challenges. *IEEE Commun. Surv. Tutor.* **2021**, *23*, 668–695. [CrossRef]
5. Yang, J.; Gu, H.; Hu, C.; Zhang, X.; Gui, G.; Gacanin, H. Deep complex-valued convolutional neural network for drone recognition based on RF fingerprinting. *Drones* **2022**, *6*, 374. [CrossRef]
6. Azari, M.M.; Sallouha, H.; Chiumento, A.; Rajendran, S.; Vinogradov, E.; Pollin, S. Key technologies and system trade-offs for detection and localization of amateur drones. *IEEE Commun. Mag.* **2018**, *56*, 51–57. [CrossRef]
7. Haring, L.; Chen, Y.; Czylwik, A. Automatic modulation classification methods for wireless OFDM systems in TDD mode. *IEEE Trans. Commun.* **2010**, *58*, 2480–2485. [CrossRef]
8. Xu, J.L.; Su, W.; Zhou, M. Likelihood-ratio approaches to automatic modulation classification. *IEEE Trans. Syst. Man Cybern.* **2010**, *41*, 455–469. [CrossRef]
9. Zhang, Z.; Luo, H.; Wang, C.; Gan, C.; Xiang, Y. Automatic modulation classification using CNN-LSTM based dual-stream structure. *IEEE Trans. Veh. Technol.* **2020**, *69*, 13521–13531. [CrossRef]
10. Huang, S.; Lin, C.; Xu, W.; Gao, Y.; Feng, Z.; Zhu, F. Identification of active attacks in Internet of Things: Joint model-and data-driven automatic modulation classification approach. *IEEE Internet Things J.* **2020**, *8*, 2051–2065. [CrossRef]
11. Dobre, O.A.; 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]
12. 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]
13. 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]
14. Krichen, M.; Mihoub, A.; Alzahrani, M.Y.; Adoni, W.Y.H.; Nahhal, T. Are Formal Methods Applicable To Machine Learning and Artificial Intelligence? In Proceedings of the 2022 2nd International Conference of Smart Systems and Emerging Technologies (SMARTTECH), Riyadh, Saudi Arabia, 9–11 May 2022; pp. 48–53.
15. Raman, R.; Gupta, N.; Jeppu, Y. Framework for Formal Verification of Machine Learning Based Complex System-of-Systems. *Insight* **2023**, *26*, 91–102. [CrossRef]
16. Tu, Y.; Lin, Y.; Zha, H.; Zhang, J.; Wang, Y.; Gui, G.; Mao, S. Large-scale real-world radio signal recognition with deep learning. *Chin. J. Aeronaut.* **2022**, *35*, 35–48. [CrossRef]
17. Huang, H.; Peng, Y.; Yang, J.; Xia, W.; Gui, G. Fast beamforming design via deep learning. *IEEE Trans. Veh. Technol.* **2020**, *69*, 1065–1069. [CrossRef]
18. Guan, G.; Zhou, Z.; Wang, J.; Liu, F.; Sun, J. Machine learning aided air traffic flow analysis based on aviation big data. *IEEE Trans. Veh. Technol.* **2020**, *69*, 4817–4826.
19. Zhang, X.; Zhao, H.; Zhu, H.B.; Adebisi, B.; Gui, G.; Gacanin, H.; Adachi, F. NAS-AMR: Neural architecture search based automatic modulation recognition method for integrating sensing and communication system. *IEEE Trans. Cogn. Commun. Netw.* **2022**, *8*, 1374–1386. [CrossRef]
20. Fu, X.; Gui, G.; Wang, Y.; Gacanin, H.; Adachi, F. Automatic modulation classification based on decentralized learning and ensemble learning. *IEEE Trans. Veh. Technol.* **2022**, *71*, 7942–7946. [CrossRef]
21. Fu, X.; Gui, G.; Wang, Y.; Ohtsuki, T.; Adebisi, B.; Gacanin, H.; Adachi, F. Lightweight automatic modulation classification based on decentralized learning. *IEEE Trans. Cogn. Commun. Netw.* **2022**, *8*, 57–70. [CrossRef]
22. Wang, Y.; Gui, G.; Gacanin, H.; Adebisi, B.; Sari, H.; Adachi, F. Federated learning for automatic modulation classification under class imbalance and varying noise condition. *IEEE Trans. Cogn. Commun. Netw.* **2022**, *8*, 86–96. [CrossRef]
23. Wang, Y.; Gui, G.; Ohtsuki, T.; Adachi, F. Multi-task learning for generalized automatic modulation classification under non-Gaussian noise with varying SNR conditions. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 3587–3596. [CrossRef]
24. Zheng, Q.; Zhao, P.; Li, Y.; Wang, H.; Yang, Y. Spectrum interference-based two-level data augmentation method in deep learning for automatic modulation classification. *Neural Comput. Appl.* **2020**, *33*, 7723–7745. [CrossRef]
25. Zheng, Q.; Zhao, P.; Wang, H.; Elhanashi, A.; Saponara, S. Fine-grained modulation classification using multi-scale radio transformer with dual-channel representation. *IEEE Commun. Lett.* **2022**, *26*, 1298–1302. [CrossRef]
26. Hou, C.B.; Liu, G.W.; Tian, Q.; Zhou, Z.C.; Hua, L.J.; Lin, Y. Multi-signal modulation classification using sliding window detection and complex convolutional network in frequency domain. *IEEE Internet Things J.* **2022**, *9*, 19438–19449. [CrossRef]

27. Qi, P.; Zhou, X.; Ding, Y.; Zhang, Z.; Zheng, S.; Li, Z. FedBKD: Heterogenous federated learning via bidirectional knowledge distillation for modulation classification in IoT-edge system. *IEEE J. Sel. Top. Signal Process.* **2023**, *17*, 189–204. [CrossRef]
28. O'Shea, T.J.O.; Corgan, J.; Clancy, T.C. Convolutional radio modulation recognition networks. In Proceedings of the Engineering Applications of Neural Networks: 17th International Conference, EANN 2016, Aberdeen, UK, 2–5 September 2016; pp. 213–226.
29. 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.
30. Liao, K.; Zhao, Y.; Gu, J.; Zhang, Y.; Zhong, Y. Sequential convolutional recurrent neural networks for fast automatic modulation classification. *IEEE Access* **2021**, *9*, 27182–27188. [CrossRef]
31. Chang, S.; Huang, S.; Zhang, R.; Feng, Z.; Liu, L. Multitask-learning-based deep neural network for automatic modulation classification. *IEEE Internet Things J.* **2021**, *9*, 2192–2206. [CrossRef]
32. Guo, L.; Wang, Y.; Hou, C.; Lin, Y.; Zhao, H.; Gui, G. Ultra Lite Convolutional Neural Network for Automatic Modulation Classification. *arXiv* **2022**, arXiv:2208.04659.
33. Huang, L.; Pan, W.; Zhang, Y.; Qian, L.; Gao, N.; Wu, Y. Data augmentation for deep learning-based radio modulation classification. *IEEE Access* **2019**, *8*, 1498–1506. [CrossRef]
34. Xie, S.; Girshick, R.; Dollar, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1492–1500.
35. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, A.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
36. Graves, A. Long short-term memory. In *Supervised Sequence Labelling with Recurrent Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 37–45.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.