Article

Deep Learning-Based Multi-Feature Fusion for Communication and Radar Signal Sensing

Ting Li [1], Tian Liu [1,*], Zhangli Song [2], Lin Zhang [2,*] and Yiming Ma [2]

[1] Southwest China Institute of Electronic Technology, Chengdu 610036, China; liting142@126.com
[2] National Key Laboratory of Wireless Communications, University of Electronic Science and Technology of China (UESTC), Chengdu 611731, China; 202221220234@std.uestc.edu.cn (Z.S.); 202121220235@std.uestc.edu.cn (Y.M.)
[*] Correspondence: liutian139@139.com (T.L.); linzhang1913@gmail.com (L.Z.)

Abstract: Recent years witness the rapid development of communication and radar technologies, and many transmitters are equipped with both communication and radar functionalities. To keep track of the working state of a target dual-functional transmitter, it is crucial to sense the modulation mode of the emitted signals. In this paper, we propose a deep learning-based intelligent modulation sensing technique for dual-functional transmitters. Different from existing modulation sensing methods, which usually focus on communication signals, we take both communication and radar signals into consideration. Typically, the dominant features of communication signals lie in the time domain, while those of radar signals lie in both time and frequency domains. To enhance the sensing accuracy, we first exploit real and complex value convolution operations to extract both time-domain and frequency-domain features of emitted signals from the target transmitter. Then, we fuse the extracted features by assigning proper weights with the attention mechanism. Simulation results reveal that the proposed technique can improve the sensing accuracy by up to 4% on average compared with benchmarks.

Keywords: communication and radar signals; modulation sensing; deep learning; multi-feature fusion

1. Introduction

In recent years, with the rapid development of communication and radar technologies, many transmitters are equipped with both communication and radar functionalities [1,2]. To keep track of the working state of a target dual-functional transmitter, it is crucial to sense the modulation mode of the emitted signal. Concretely, for a high-rate data transmission purpose, a higher order modulation mode is preferred [3], such as quadrature amplitude modulation 16 (QAM16). For the target detection with a longer distance, the pulse compression modulation (PCM) is a better choice than linear frequency modulation (LFM), due to PCM signals having a higher range resolution and a higher signal-to-noise ratio (SNR) than LFM signals [4]. Therefore, it is possible to identify the working state of a target dual-functional transmitter by sensing the modulation mode of emitted signals.

Modulation sensing is adopted to enhance the communication efficiency in a wide range of cooperative and non-cooperative scenarios. To be specific, modulation sensing can be used to enhance the communication efficiency in multi-point cooperative transmission scenarios and can also be used to understand the signals of non-cooperative transceivers for competition purpose, e.g., cognitive radio and electronic warfare [5–7]. With the radio environment becoming increasingly complicated, how to effectively and accurately sense the modulation modes of the emitted signals from target transceivers has become a hot research direction. In general, modulation sensing techniques mainly include traditional and deep learning (DL) methods. Traditional methods could be roughly divided into two

groups: likelihood-based (LB) and feature-based (FB) [8]. In particular, LB methods transform a modulation sensing problem into a multi-hypothesis testing problem and correctly identify corresponding modulation modes by setting appropriate discriminant thresholds. The disadvantages of these methods lie in that it is difficult to determine proper discriminant thresholds and the performance is typically sensitive to the radio environment [9]; meanwhile, the computational complexity is high. FB methods extract some physical or statistical characteristics, such as cyclic moments or higher order cumulants [10], and adopt proper classifiers, such as support vector machines or K-nearest neighbor, to sense the modulation modes of emitted signals. In practical applications, these methods heavily depend on manual analysis [11], and the performance is unacceptable when modulated signals are complicated [12].

Recently, data-driven DL methods have attracted a lot of attention of the researchers in academia [13–17] and strongly promoted the development of modulation sensing technology. In particular, O'Shea et al. propose a convolution neural network (CNN) architecture and demonstrate that DL methods outperform several traditional methods in terms of sensing accuracy and inference speed [18]. Jakob et al. propose a complex value convolution (CVC) method to extract complex feature of signals, which is proved to improve the sensing accuracy [19]. Lin et al. introduce the frequency sequence and the time-frequency attention (TFA) mechanism into CNN to further boost the sensing accuracy [20]. Godwin et al. adopt multiple skip connections in a modified CNN architecture, which is proved to improve the stability of the network [21]. Ayman et al. propose a modulation recognition method leveraging temporal correlation of features, which integrates CNN and LSTM models [22]. Zhang et al. suppose that different features are complementary, and propose a CNN based multi-delay feature fusion scheme which performes well at low SNRs [23]. In general, DL methods typically consist of two successive stages. In the first stage, amounts of representations of received signals are fed into DL methods to train the neural networks, which is time-consuming but conducted off-line. In the second stage, well-trained neural networks are utilized to sense the modulation modes of received signals, which is conducted online and time-saving. During training, massive simulation data required in DL methods are readily available. Compared with traditional methods, DL methods typically have a lower computational complexity but can achieve better performance. Meanwhile, DL methods avoid manual feature selections and can be applied in more complicated scenarios.

We notice that most existing works focus on communication signals while ignoring radar signals and cannot be applied in sensing the working state of dual-functional transmitters. Meanwhile, the main features of communication signals typically lie in the time domain, while those of radar signals lie in both time and frequency domains. Thus, the modulation sensing algorithm for dual-functional transmitters needs to be carefully designed to accurately identify the working state of a dual-functional target transmitter. In this paper, we propose a DL-based intelligent modulation sensing technique. In particular, to fully and succinctly represent the emitted signals from the target transmitter, we first exploit real value convolution (RVC) and CVC nerual networks to extract multiple features from both time and frequency domains. Further, we fuse the extracted features by assigning proper weights with the attention mechanism. Simulation results reveal that the proposed technique can improve the sensing accuracy by up to 4% on average compared with benchmarks.

The structure of this paper is as follows: Section 1 introduces the background and applications of modulation sensing technology and compares DL algorithms with traditional methods; Section 2 describes two readily available signal representations in both time and frequency domains for the proposed technique; Section 3 introduces the structure and theory of the proposed algorithm in detail and analyzes the time and space complexity; Section 4 analyzes the performance of the proposed algorithm by conducting simulation experiments; and the conclusion is given in Section 5.

2. Signal Representation

The radio signal from a dual-functional target transmitter can be typically represented in both time and frequency domains. In particular, the time-domain signal conveys the variation pattern of the signal amplitude in terms of the time, while the frequency-domain signal can directly reflect the energy distribution in terms of frequency. There exists complementary information between different representations of signals [7], meaning that inputting both time and frequency domain sequences could extract the complementary features between both domains, thus enhancing the sensing performance. In this section, we introduce two readily available representations of a received signal in both time and frequency domains.

2.1. Time-Domain Representation

A received complex signal in the time domain can be represented as

$$r(t) = x(t) * h(t) + n(t),　\tag{1}$$

where $x(t)$ represents the modulated complex signal emitted from the target dual-functional transmitter, $h(t)$ represents the channel impulse response, $*$ represents the convolution operation, and $n(t)$ represents the additive complex noise at the receiver. Upon receiving the signal, the receiver converts it into a baseband discrete form with an analog-to-digital converter. Without abusing the notations, we denote the discrete form of the received complex signal as $r[n]$, which is usually decomposed into two orthogonal components:

$$r[n] = r_I[n] + j\,r_Q[n],　\tag{2}$$

where n is the sampling index, j is the imaginary unit, and $j^2 = -1$, $r_I[n]$ and $r_Q[n]$ represent the in-phase (I) and quadrature (Q) components of the receiving complex signal, respectively. Since the IQ sequence is the raw data obtained at the receiver, it is most readily available as the input for DL methods.

2.2. Frequency-Domain Representation

A typical method to obtain the frequency-domain sequence of a complex signal is to conduct the discrete fourier transform (DFT) operation to the corresponding time-domain sequence, i.e.,

$$R[k] = \mathrm{DFT}(r[n]) = \sum_{n=0}^{N-1} r[n] W_N^{kn}, \ \ k = 0, 1, \cdots, N-1,　\tag{3}$$

where N is the DFT length, $W_N^{kn} = e^{-j\frac{2\pi kn}{N}}$, and $R[k]$ is the complex sequence in frequency domain.

It is noted that DFT has a variety of fast algorithms, collectively known as fast fourier transform (FFT), which can effectively reduce the signal processing complexity. The FFT algorithm is widely used in both industry and academia, and there are many mature and simple implementation schemes. For the convenience of subsequent processing, we adopt the FFT the same length as the time-domain sequence.

3. Proposed Modulation Sensing Algorithm

As mentioned above, time-domain and frequency-domain representations of signals convey different features; meanwhile, RVC and CVC can extract different features from time-domain and frequency-domain signals. In particular, RVC can extract independent features of the real and imaginary parts of signals, while CVC can extract interaction features between both parts [19,24]. In this section, we propose an intelligent modulation sensing algorithm by first extracting multiple features of time-domain and frequency-domain signals and then fusing the extracted features with the attention mechanism. Finally,

a classifier is used to map the fused features to the sensing result. In the following section, we will first provide the algorithm structure and then elaborate on the algorithm designs.

3.1. Algorithm Structure

As shown in Figure 1, the proposed algorithm has three main modules including feature extraction, feature fusion, and classifier. In particular, the feature extraction module transforms the input sequences into multiple feature flows by conducting both CVC and RVC operations, the feature fusion module deeply fuses multiple feature flows into one fused feature flow with the attention mechanism, and the classifier module maps the fused features to the sensing results.
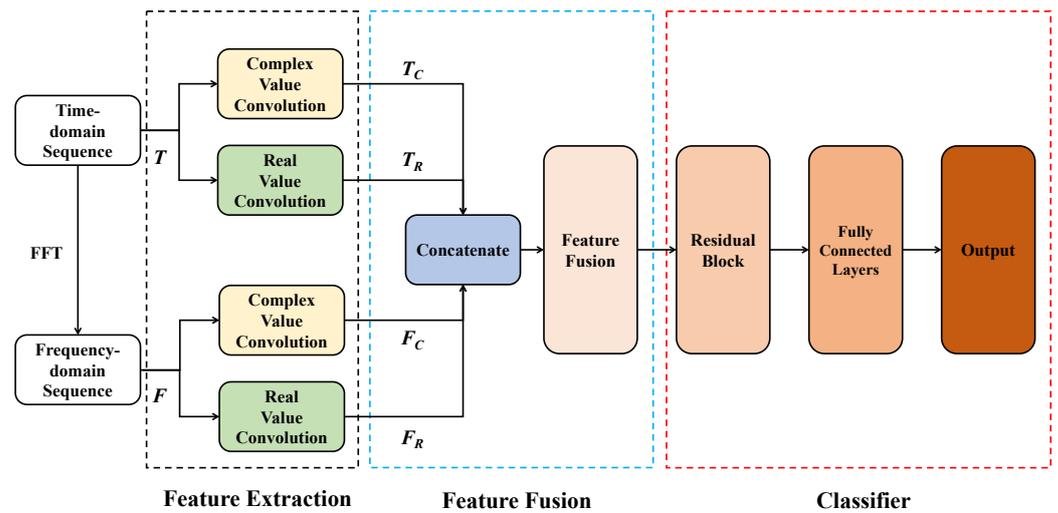
Figure 1. Proposed algorithm structure.

3.2. Feature Extraction

The time-domain sequence, denoted as T in Figure 1, could be expressed as Equation (2), while the frequency-domain sequence, denoted as F, could be obtained by conducting an FFT operation. Since the methods of processing T and F are identical, we only elaborate on the procedure of processing T for simplicity.

As T is a complex vector, the CVC operation [25] can be expressed as

$$T * h = (T_r + jT_i) * (h_r + jh_i) = (T_r * h_r - T_i * h_i) + j(T_r * h_i + T_i * h_r), \quad (4)$$

where h is the complex convolution filter, h_r and h_i represent real and imaginary parts of h, and T_r and T_i represent real and imaginary parts of T, respectively.

As neural networks only support the calculation of real values [26], a complex value sequence can be represented as real and imaginary parts. Then, the CVC operation can be rewritten as

$$T * h = \begin{bmatrix} \mathrm{Re}(T * h) \\ \mathrm{Im}(T * h) \end{bmatrix} = \begin{bmatrix} h_r & -h_i \\ h_i & h_r \end{bmatrix} * \begin{bmatrix} T_r \\ T_i \end{bmatrix}. \quad (5)$$

The CVC operation needs two real value convolution filters h_r and h_i, meaning that a CVC layer in the neural network could be implemented by two RVC layers. Without special instructions, 1×3-sized filters are used in all convolution layers in the proposed algorithm.

By performing RVC and CVC in terms of T and F, four feature flows can be extracted, which are denoted as T_R, T_C, F_R, and F_C, respectively, in Figure 1. The dimension of each feature flow is $C \times 2 \times N$, where C denotes the number of channels.

3.3. Feature Fusion

It should be noted that there exits some complementary information among different feature flows, which may have different contributions to the sensing performance. In other words, different feature flows need to be carefully fused according to their contributions to the sensing result. Four feature flows in Figure 1 have distinct effects on the sensing accuracy, but it is challenging to quantify these effects and obtain the optimal weights with closed forms. So, motivated by the successful application in multi-modal feature fusion [24], we also adopt the attention mechanism to automatically learn the optimal weights to enhance the performance.

The overall structure of the feature fusion module is shown in Figure 2. In particular, four feature flows are first concatenated in the dimension of channel and then input into the feature fusion module, i.e.,

$$CF = \left[T_R, T_C, F_R, F_C \right],\tag{6}$$

where CF denotes the concatenated features, and the dimension is $4C \times 2 \times N$.

Upon receiving CF, the feature fusion module processes it with a depthwise separable convolution (DSC) layer [27], which aims to reduce parameters in the neural network. Then, the output of DSC layer is fed into an attention block, which distributes proper weights to different features. In particular, a DSC layer consists of a channel-wise convolution layer and a 1×1 convolution layer, which is used to change the dimension of the feature map. In the paper, we reduce the channel dimension of features to C. Here, a skip connection is introduced to prevent gradient vanishing.
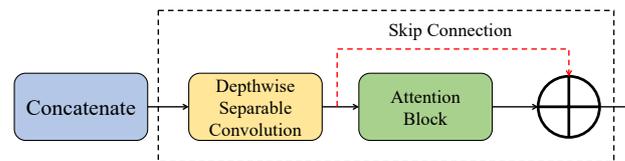
Figure 2. Feature fusion module structure.

The structure of the attention block is shown in Figure 3. In particular, the input features are first processed by a squeeze layer [28] and an activation (Sigmoid) layer to calculate proper weights of different feature channels. The meaning of proper weights is that the weights could help the neural network to focus on significant features and ignore interference, which are obtained automatically rather than artificially assigned. In this paper, the channel attention mechanism is adopted to assign weights according to the importance of different features, thus improving sensing accuracy. In particular, the squeeze layer is to calculate the global average value of the features in the dimension of the channel, whose dimension is equal to the channel dimension of the input feature map, i.e., C. It can be expressed as

$$Y_i = \frac{1}{W \times H} \sum_{j=1}^{W} \sum_{k=1}^{H} X_{i,j,k}, \quad i = 1, \cdots, C,\tag{7}$$

where X denotes the input feature map, and the subscripts represent indexes of the channel width and height, respectively. The Sigmoid function could be expressed as

$$y = \frac{1}{1 + e^{-x}}\tag{8}$$

Then, the attention block outputs the fused features after multiplying pixels in each channel by the corresponding weight.
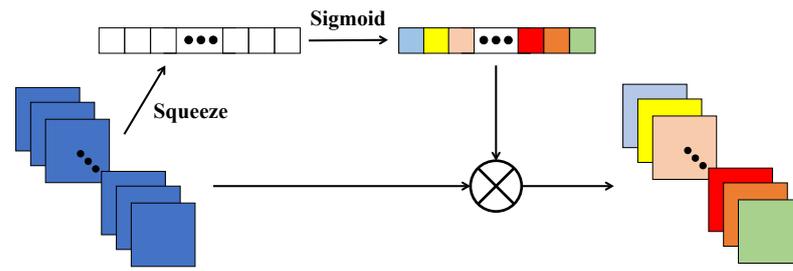
Figure 3. Attention block structure. Dark blue squares represent the input features, other colors represent corresponding weights of different channels.

3.4. Classifier

Upon obtaining the fused feature flow, the residual blocks and fully connected (FC) layers are designed to further extract the dominant features contained in the fused feature flow and map the dominant feature to the sensing result.

3.4.1. Residual Block

The structure of the residual block is shown in Figure 4a, in which the input is processed by two residual units [29] and a maxpool layer in a sequential order. In particular, the kernel size of the maxpool layer is set to 2×2, and the stride is set to $(2, 2)$ to prevent overfitting, which can be expressed as

$$y_h = Max(x_{0,2h}, x_{0,2h+1}, x_{1,2h}, x_{1,2h+1}), \tag{9}$$

where the first dimension of the subscript represents the index of the width of the feature map, and the second dimension represents the index of the height. Since the width of input feature map is 2, the first dimension of the output feature map is ignored.

The structure of residual unit is shown in Figure 4b, in which FF denotes the input features and $f(FF)$ denotes the result by processing FF with two cascaded RVC layers.

(**a**) Residual block
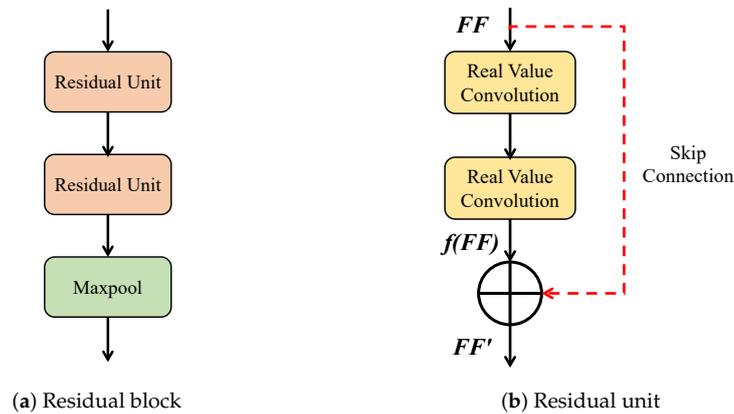
(**b**) Residual unit

Figure 4. Residual block and residual unit structures.

Here, a skip connection is also adopted to prevent a vanishing gradient in each residual unit. Then, the output FF' of each residual unit can be expressed as

$$FF' = FF + f(FF) \tag{10}$$

3.4.2. FC Layer

The feature map ultimately needs to be mapped to the correct modulation mode, so a flatten layer is needed after the residual block. In particular, the feature map tends to be multi-dimensional, while the input of the FC layer is a vector. The function of the flatten

layer is to arrange pixels of the multi-dimensional feature map into a vector in a certain order, which may be different for different DL frameworks. Then, several FC layers are adopted to change the number of neurons to map to the probability that the signal belongs to each modulation mode. The equation of FC layer could be expressed as

$$y = f_{act}(wx + b), \tag{11}$$

where x is the input feature vector, w and b denote weights and bias, respectively, and f_{act} denotes the activation function. Except for the output layer which uses Softmax, other FC layers in the proposed algorithm adopt Selu as the activation function, which is expressed as

$$f(x) = \lambda \begin{cases} x & \text{if } x > 0 \\ \alpha(e^x - 1) & \text{if } x \leq 0 \end{cases} \tag{12}$$

where $\lambda = 1.0507$, $\alpha = 1.6733$.

We optimize the number of residual blocks [29], the number of FC layers, and the number of neurons in each layer [24] through extensive cross-validations. In particular, the number of residual blocks is one, while the numbers of neurons in the FC layers are 128 and 64, respectively.

3.5. Complexity Analysis

The time and space complexity of the proposed algorithm is dominated by the calculation and parameters in RVC, CVC, DSC, and FC layers.

3.5.1. Time Complexity

Time complexity is typically measured by multiply–accumulate operations [30]. The time complexity of each convolution layer depends on the size and number of convolution kernels. In particular, to obtain a pixel in the output feature map, the number of multiply–accumulate operations in each channel is equal to the size of the convolution kernel. Therefore, the time complexity of RVC operations could be expressed as

$$\mathcal{O}\Big(\sum_l m_l^h m_l^w k_l^h k_l^w c_{l-1} c_l\Big), \tag{13}$$

where l denotes the index of the RVC layer, m_l^h and m_l^w denote the height and width of the output feature map, k_l^h and k_l^w denote the height and width of the convolution kernel, and c_l denotes the channel number of the l-th convolution layer, respectively.

As a CVC layer is implemented by two RVC layers, the time complexity of CVC is twice of that in Equation (13). Meanwhile, the time complexity of the DSC layer consists of two parts: channel-wise convolution and 1×1 convolution, i.e.,

$$\mathcal{O}\big(m_l^h m_l^w k_l^h k_l^w c_l + m_l^h m_l^w c_l c_{l+1}\big) \tag{14}$$

The calculation of the FC layer, i.e., Equation (11), is equivalent to multiplying weights matrix w with input vector x and then adding bias vector b. Therefore, the time complexity of FC layers could be obtained by summing up the calculations in each layer, i.e.,

$$\mathcal{O}\Big(\sum_l n_{l-1} n_l\Big), \tag{15}$$

where n_l denotes the number of neurons in the l-th FC layer.

Therefore, the overall time complexity of the proposed algorithm can be obtained by summing up the time complexity of each module, i.e.,

$$\mathcal{O}\Big(3\sum_l m_l^h m_l^w k_l^h k_l^w c_{l-1} c_l + \sum_l n_{l-1} n_l + (m_l^h m_l^w k_l^h k_l^w c_l + m_l^h m_l^w c_l c_{l+1})\Big) \tag{16}$$

3.5.2. Space Complexity

Space complexity is typically measured by the number of trainable parameters, which are mainly distributed in RCV, CVC, DSC, and FC layers. In particular, the trainable parameters of RVC consist of weights and bias, and therefore, the space complexity of RVC can be expressed as

$$\mathcal{O}\Big(\sum_l (k_l^h k_l^w c_{l-1} c_l + c_l)\Big) \tag{17}$$

As a CVC layer is implemented by two RVC layers, the space complexity of CVC is twice of that in Equation (17); meanwhile, the space complexity of DSC is

$$\mathcal{O}\big(k_l^h k_l^w c_l + c_l c_{l+1} + c_l + c_{l+1}\big) \tag{18}$$

In the FC layer, the weight matrix and bias vector are both trainable variables, so the space complexity of FC layers could be expressed as

$$\mathcal{O}\Big(\sum_l (n_{l-1} n_l + n_l)\Big) \tag{19}$$

Therefore, the overall space complexity of the proposed algorithm is

$$\mathcal{O}\Big(3\sum_l (k_l^h k_l^w c_{l-1} c_l + c_l) + \sum_l (n_{l-1} n_l + n_l) + k_l^h k_l^w c_l + c_l c_{l+1} + c_l + c_{l+1}\Big) \tag{20}$$

4. Simulation Results and Analysis

4.1. Dataset and Algorithm Architecture

Although there exist some public datasets on modulation sensing, they only include communication signals and ignore radar signals. To build the dataset for fair comparisons, we refer to communication signals of the public dataset RML2016.10a in [18] and radar signals in [4]. Meanwhile, we also refer to the corresponding parameter settings and signal generation processes. In particular, we consider eight typical modulation modes of communication signals including binary phase shift keying (BPSK), quadrature PSK (QPSK), 8PSK, continuous phase frequency shift keying (CPFSK), Gaussian FSK (GFSK), QAM16, QAM64, and analog modulation (AM), as well as two representative modulation modes of radar signals including LFM and PCM. The simulated channel environment considers two cases, Gaussian channel and Rice channel, with SNRs ranging from -5 dB to 15 dB. Thereinto, K-factor of the Rice channel is 4, and 8 sinusoids are used to model the fading process. By considering the relative motion between the target dual-functional transmitter and the receiver, we set the maximum value of Doppler shift to be 20 kHz. The bandwidth of the signal is 100 MHz, the carrier frequency is 2.4 GHz, and the up-sample ratio is 8. Generated signals have 1024 sampling points per frame, and 128 continuous points are randomly selected in the dataset.

The dataset includes 2000 frames with 128 samples of each modulation type per SNR in each channel, and the ratio of the dataset for training, validation, and testing is 3:3:4. In addition, our simulation experiments are implemented by using Nvidia RTX 3080Ti GPU based on Pytorch 2.0.1. Through extensive cross-validations and end-to-end supervised training, we optimize the hyperparameters as follows: the optimizer adopted is Adam, whose initial learning rate is 0.001, and the loss function is categorical crossentropy; epochs are set to 30, while the training process would be terminated if the validation loss fails to decrease within 5 epochs. Meanwhile, the model with minimum validation loss is chosen for sensing the modulation modes of testing signals, because it exhibits the best generalization performance. The architecture of the proposed algorithm in Figure 1 is summarized in Table 1. It is known that a larger scale neural network has a stronger learning and representation ability. In real-world scenarios with large-scale datasets, we can still use the framework in Figure 1 and balance the scalability and efficiency of the proposed algorithm by adapting the number of layers or the number of neurons in each layer.

Table 1. Model layout.

Layer	Output Dimensions
Input	2×128
CVC/RVC	$16 \times 2 \times 128$
Concatenate	$64 \times 2 \times 128$
DSC	$16 \times 2 \times 128$
Attention Block	$16 \times 2 \times 128$
Residual Block	$16 \times 1 \times 64$
FC/Selu	128
FC/Selu	64
Output/Softmax	10

4.2. Ablation Experiment

As mentioned previously, our hypothesis is that complementary features can be extracted from time-domain sequence *T* and frequency-domain sequence *F*; meanwhile, CVC and RVC could extract different features from the same sequence. To demonstrate this idea, we design an ablation experiment [31]. In the experiment, an input sequence (*T* or *F*), feature extraction operation (CVC or RVC), or attention block is removed from the proposed algorithm. The number of filters in the feature extraction module of each ablated algorithm is doubled to ensure that both numbers of trainable parameters and neurons are close to the proposed algorithm. The results in the Gaussian channel are shown in Figure 5.
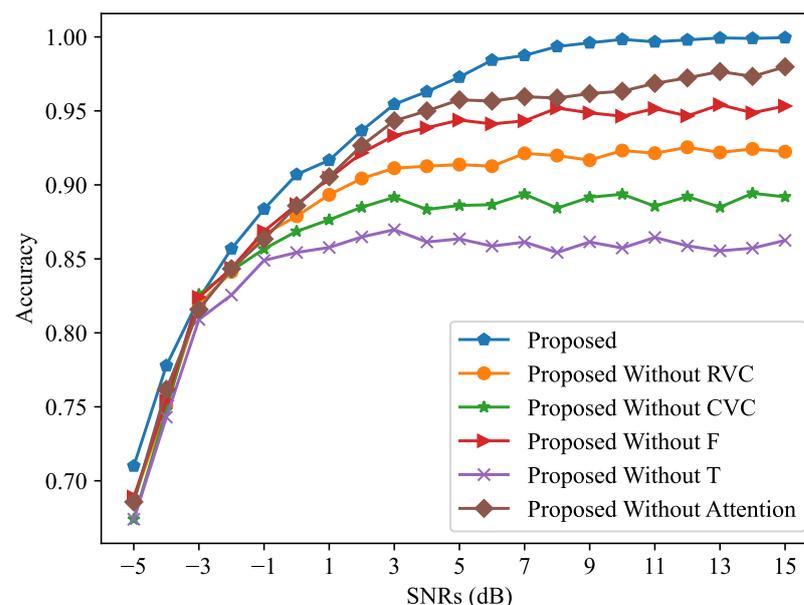
Figure 5. Ablation experiment results.

From Figure 5, the sensing accuracy of the proposed algorithm increases rapidly from 71.2% to 99.1% as SNR grows from −5 dB to around 8 dB and then gradually converges to 1 as the SNR further grows. The sensing accuracy of the ablated algorithm using only *T* (proposed without *F*) as the input converges to around 94.2%, while the sensing accuracy of the ablated algorithm using only *F* merely converges to 85.6%. It can be concluded that the time-domain features are more important to improve the sensing accuracy compared with frequency-domain features. This is because we adopt eight up-samples to process each symbol in the time domain, and there exist time correlations among different sample points, making it easier for the neural network to extract features and enhance the sensing accuracy, while there is little correlation in the frequency-domain signals. Meanwhile, the sensing accuracy can be further improved by properly fusing multiple features in both time and frequency domains. Similarly, we can also conclude that the CVC operation is

able to extract more information than the RVC operation, and the sensing accuracy can be further enhanced by properly combining CVC and RVC operations. In addition, the sensing accuracy without the attention block decreases around 2% compared with the proposed algorithm, indicating that the attention mechanism can help the neural network pay more attention to important features, thus enhancing the performance.

4.3. Algorithms Comparison

To demonstrate the advantages of the proposed algorithm, we compare the sensing accuracy of the proposed algorithm with benchmarks, which are widely adopted in modulation sensing [7,24], including dual-stream CNN-LSTM [9], TFA CNN [20], CNN-LSTM [32], and improved CNN [33]. In particular, dual-stream CNN-LSTM takes IQ and amplitude-phase sequences as inputs and utilizes CNN and LSTM layers to sense the modulation modes. TFA CNN adopts multiple attention modules to explore features in spectrogram images, while CNN-LSTM sensing corresponding modulation modes by using CNN and LSTM layers to extract features in IQ sequences and fourth order cumulants. In addition, improved CNN processes the IQ sequence by adopting convolution and dense layers. The numbers of trainable parameters of algorithms in this experiment are shown in Table 2, and the convergence curves in the Gaussian channel are shown in Figure 6.

Table 2. Algorithm parameters.

Algorithm	Parameters
Dual-stream CNN-LSTM	613 K
TFA CNN	104 K
CNN-LSTM	467 K
improved CNN	159 K
Proposed	145 K
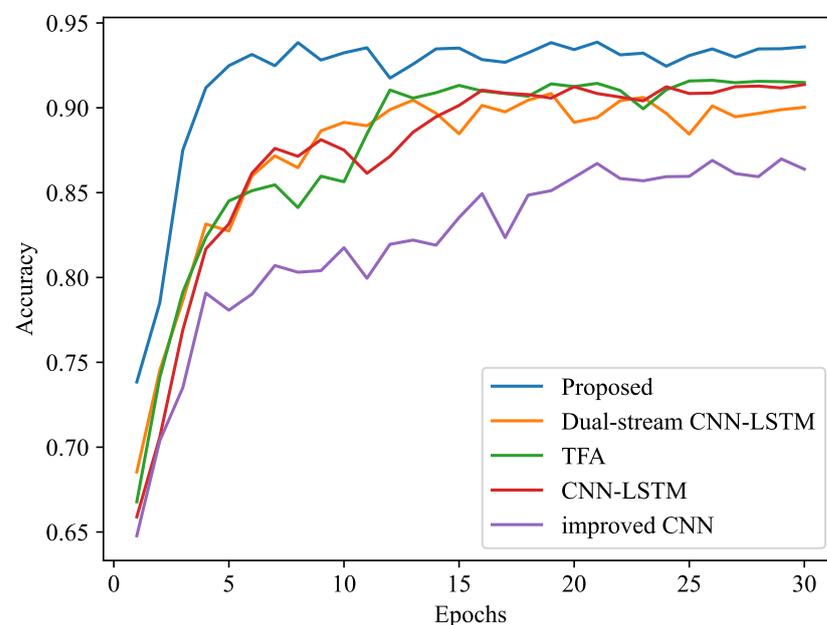
Figure 6. Convergence curves of different algorithms.

From Figure 6, it can be observed that the sensing accuracy of the improved CNN increases gradually from around 65% to 86% after 30 epochs. The sensing accuracy of dual-stream CNN-LSTM, TFA, and CNN-LSTM is comparable, increasing from around 67% to around 90% after 30 epochs; although, there exist some fluctuations during the training phases, and it does not increase any more. Meanwhile, we observe that the sensing accuracy of the proposed algorithm can converge to around 94% after only five epochs. Therefore,

the proposed algorithm is superior to the benchmarks in terms of both convergence rate and converged sensing accuracy. Moreover, the inference time of each algorithm is shown in Table 3. The inference speed of the proposed algorithm can reach 14.7 μs per frame, which is much faster than benchmarks. It is reasonable since this algorithm has fewer parameters and can be computed in parallel in the feature extraction module, while the algorithms with LSTM layers have slower inference speed due to time dependence.

Table 3. Inference time of different algorithms.

Algorithm	Inference Time (μs/Frame)
Dual-stream CNN-LSTM	78.4
TFA CNN	21.8
CNN-LSTM	47.4
improved CNN	19.8
Proposed	14.7

Figures 7 and 8 illustrate the sensing accuracy of different algorithms with different SNRs in the Gaussian and Rice channels, respectively. In general, the proposed algorithm performs better than benchmarks in any SNR condition, which is between −5 dB and 15 dB. From Figure 7, in the low and medium SNR regions, i.e., SNR is between −5 dB and 5 dB, the sensing accuracy of the proposed algorithm is around 5% higher than those of CNN-LSTM and improved CNN, and is slightly higher than those of dual-stream CNN-LSTM and TFA. In the high SNR region, i.e., the SNR is between 5 dB and 15 dB, the sensing accuracy of the proposed algorithm approaches 1 quickly and is around 8% higher than that of the improved CNN and is around 4% higher than those of the dual-stream CNN-LSTM, CNN-LSTM, and TFA algorithms. In fact, this is reasonable since the proposed algorithm can extract more features and fuse them in a better way to enhance the sensing accuracy compared with benchmarks. In the low SNR region, the signal is seriously impaired by the noise, resulting in ambiguous features for the neural network to learn, so the advantages of the proposed algorithm are not as obvious as those in the high SNR region.
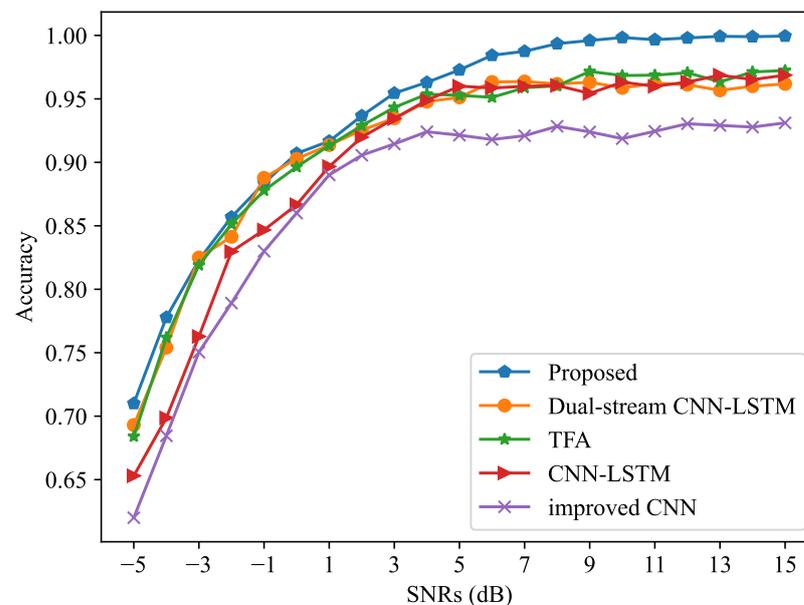
Figure 7. Sensing accuracy of different algorithms in the Gaussian channel.
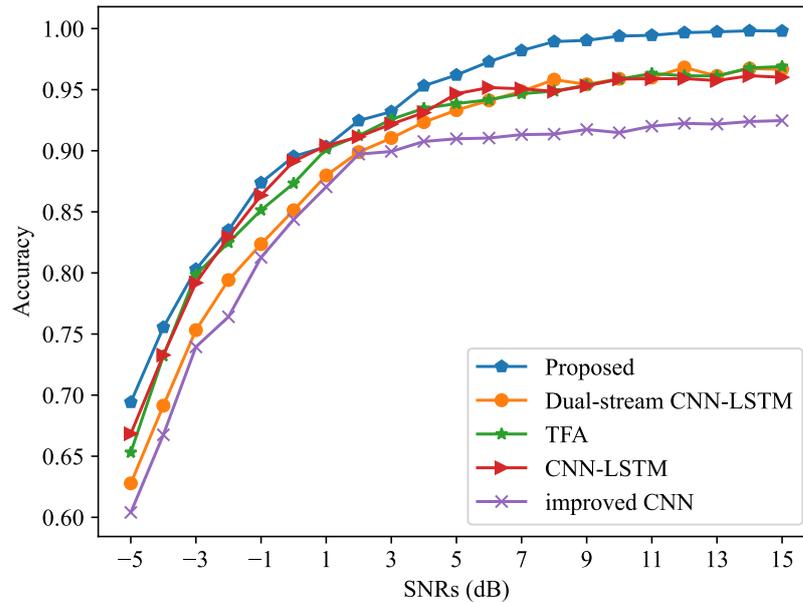
Figure 8. Sensing accuracy of different algorithms in the Rice channel.

From Figure 8, it can be observed that in the Rice channel, the sensing performance curves of all algorithms have a similar trend as those in the Gaussian channel. The sensing accuracy of the proposed algorithm is still about 4% higher than the benchmarks. Nevertheless, due to the impairment of the Rice channel on the signal, the sensing performance of all algorithms in the low and median SNR regions (lower than 10 dB) slightly degrades, e.g., about 1.5% lower on average.

4.4. Length Comparison

Figure 9 shows the performance with different sample points of the proposed algorithm. It can be observed from the figure that a larger number of samples generally leads to a higher sensing accuracy. This is because as the sequence length increases, the information of the signals becomes more abundant and features become more obvious, which are easier for the neural network to extract.
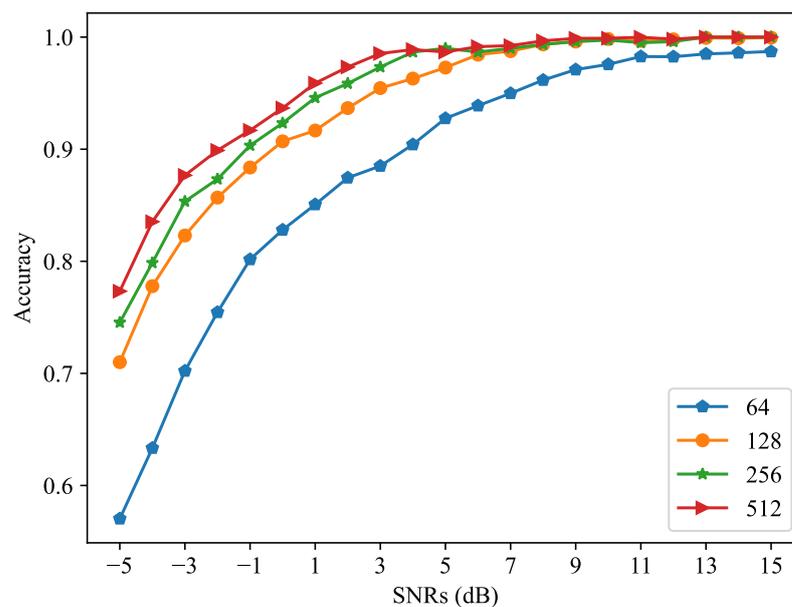
Figure 9. Sensing accuracy with different sample points.

4.5. Confusion Matrices

To study the sensing accuracy of different modulation modes, we provide confusion matrices of different algorithms when the SNR is 5 dB in Figure 10. Generally, the proposed algorithm can accurately distinguish communication and radar signals and has almost perfect sensing capability in terms of BPSK, CPFSK, GFSK, AM, PCM, and LFP when the SNR is 5 dB. Compared with benchmarks, the proposed algorithm has a higher sensing accuracy for QPSK, 8PSK, QAM16, QAM64, PCM, and LFM signals. Especially for the QAM signals, it is challenging for benchmarks to distinguish QAM16 and QAM64 signals. The proposed algorithm can improve the sensing accuracy of QAM16 by 5%, 9%, 8%, and 32%, respectively, compared with dual-stream CNN-LSTM, TFA CNN, CNN-LSTM, and improved CNN. In addition, both QAM16 and QAM64 are high order modulation modes, and the SNR required to distinguish them is higher than the SNRs of other modulation modes. Therefore, the major challenge lies in the correct sensing of high-order modulation modes in the low SNR region. Generally, the proposed algorithm outperforms the benchmarks for the sensing of both communication and radar signals.
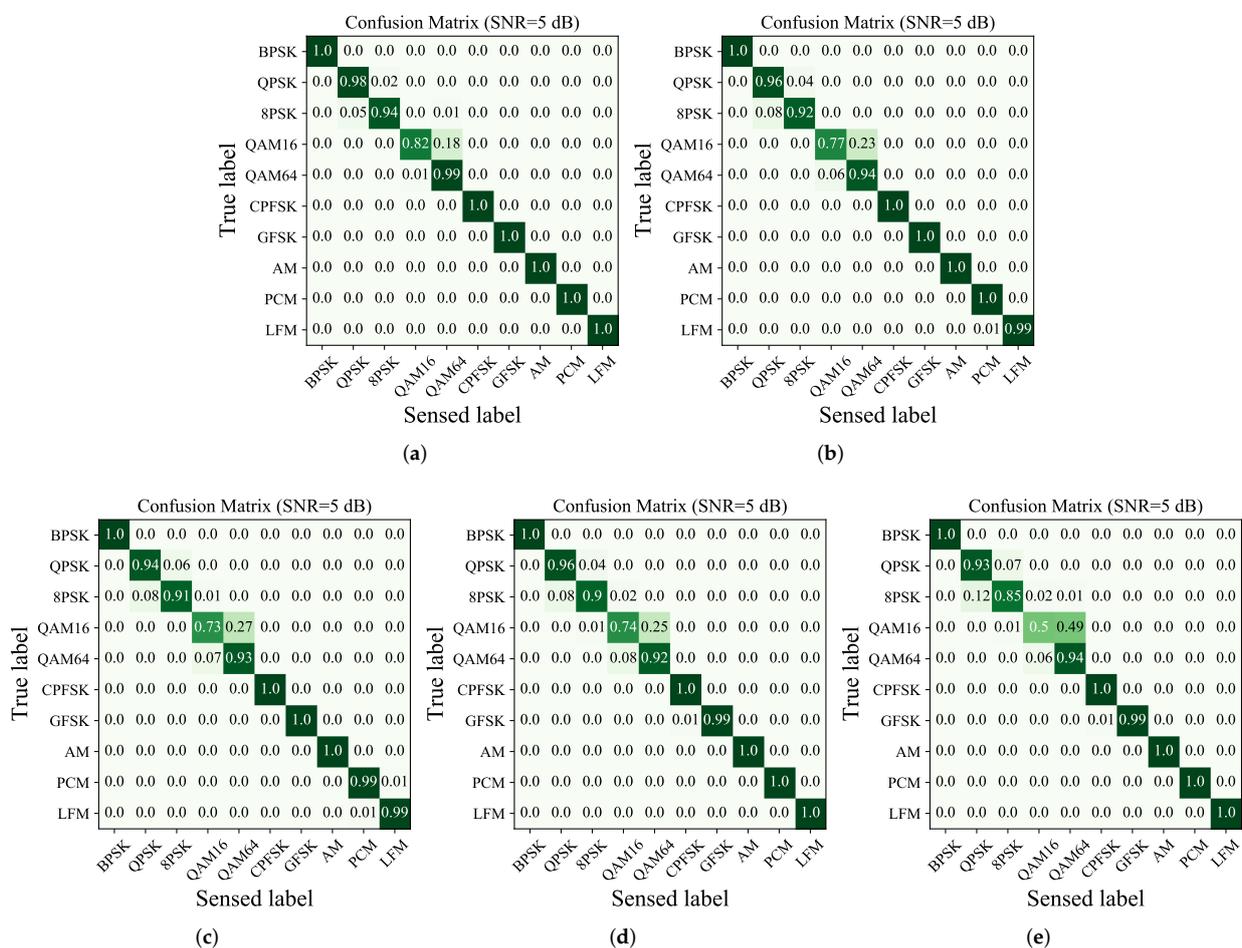
Figure 10. Confusion matrices. (**a**) proposed; (**b**) dual-stream CNN-LSTM in [9]; (**c**) TFA CNN in [20]; (**d**) CNN-LSTM in [32]; and (**e**) improved CNN in [33].

5. Conclusions

In this paper, we first introduce the background and applications of modulation sensing technology and analyze and compare DL methods with traditional methods. Therewith, two readily available representations of signals in both time and frequency domains are introduced. Then, a DL-based multi-feature fusion technique is proposed to sense the modulation modes to acquire the working state of dual-functional transmitters with both

communication and radar capabilities. In particular, the proposed technique first conducts both RVC and CVC operations to time-domain and frequency-domain sequences of received signals to explore multiple features. Then, the attention mechanism is leveraged to adaptively fuse the features by assigning proper weights. Finally, a classifier is adopted to sense different modulation modes based on fused features. Simulation results reveal that the proposed technique can generally improve the sensing accuracy by up to 4% on average compared with benchmarks.

Author Contributions: Software, Y.M.; Writing—original draft, T.L. (Ting Li) and Z.S.; Writing—review and editing, T.L. (Tian Liu); Supervision, L.Z. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Data available on request from the authors.

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

References

1. Ahmed, A.; Zhang, Y.D.; Himed, B. Distributed dual-function radar-communication MIMO system with optimized resource allocation. In Proceedings of the 2019 IEEE Radar Conference (RadarConf), Boston, MA, USA, 22–26 April 2019; pp. 1–5. [CrossRef]
2. Huang, J.; Zhang, X.; Wang, X.; Abdelhak, M.Z. Transmit sparse array beamformer design for dual-function radar communication systems. In Proceedings of the 2023 IEEE International Radar Conference (RADAR), Sydney, Australia, 6–10 November 2023; pp. 1–6. [CrossRef]
3. Prol, F.S.; Ferre, R.M.; Saleem, Z.; Välisuo, P.; Pinell, C.; Lohan, E.S.; Elsanhoury, M.; Elmusrati, M.; Islam, S.; Celikbilek, K.; et al. Position, navigation, and timing (PNT) through low earth orbit (LEO) satellites: A survey on current status, challenges, and opportunities. *IEEE Access* **2022**, *10*, 83971–84002. [CrossRef]
4. Li, H.; Qi, H.; Yu, T.; Wang, J. The influence of pulse compression on radar detection range under different interference conditions. *Aerosp. Electron. Warf.* **2023**, *39*, 1–3. (In Chinese)
5. 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]
6. Al-Hraishawi, H.; Chougrani, H.; Kisseleff, S.; Lagunas, E.; Chatzinotas, S. A survey on nongeostationary satellite systems: The communication perspective. *IEEE Commun. Surv. Tutor.* **2023**, *25*, 101–132. [CrossRef]
7. Peng, S.; Sun, S.; Yao, Y.-D. A survey of modulation classification using deep learning: Signal representation and data preprocessing. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 7020–7038. [CrossRef]
8. Xu, J.; Sun, X.; Liu, L.; Li, Y.; Wang, Y.; Zhang, T. Modulation recognition based on instantaneous feature fusion. In Proceedings of the 2023 3rd International Symposium on Computer Technology and Information Science (ISCTIS), Chengdu, China, 7–9 July 2023; pp. 632–638. [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. Wang, Y.; Liu, Z.; Xu, Y. Application of frequency-domain features and high-order cumulants in ANN-based communication modulation recognition. In Proceedings of the 2022 IEEE 4th International Conference on Civil Aviation Safety and Information Technology (ICCASIT), Dali, China, 12–14 October 2022; pp. 1288–1293. [CrossRef]
11. Zheng, S.; Qi, P.; Chen, S.; Yang, X. Fusion methods for CNN-based automatic modulation classification. *IEEE Access* **2019**, *7*, 66496–66504. [CrossRef]
12. Hou, C.; Liu, G.; Tian, Q.; Zhou, Z.; Hua, L.; Lin, Y. Multisignal modulation classification using sliding window detection and complex convolutional network in frequency domain. *IEEE Internet Things J.* **2022**, *9*, 19438–19449. [CrossRef]
13. Du, Y.; Qi, N.; Wang, K.; Xiao, M.; Wang, W. Intelligent reflecting surface-assisted UAV inspection system based on transfer learning. *IET Commun.* **2024**, *18*, 214–224. [CrossRef]
14. Qi, N.; Wang, W.; Xiao, M.; Jia, L.; Jin, S.; Zhu, Q.; Tsiftsis, T.A. A learning-based spectrum access stackelberg game: Friendly jammer-assisted communication confrontation. *IEEE Trans. Veh. Technol.* **2021**, *70*, 700–713. [CrossRef]
15. Jia, L.; Qi, N.; Chu, F.; Fang, S.; Wang, X.; Ma, S.; Feng, S. Game-theoretic learning anti-jamming approaches in wireless networks. *IEEE Commun. Mag.* **2022**, *5*, 60–66. [CrossRef]
16. Nguyen, T.-H.; Park, H.; Park, L. Recent studies on deep reinforcement learning in RIS-UAV communication networks. In Proceedings of the 2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Bali, Indonesia, 20–23 February 2023; pp. 378–381. [CrossRef]
17. Che, Y.; Lin, F.; Liu, J. Deep reinforcement learning in M2M communication for resource scheduling. In Proceedings of the 2021 World Conference on Computing and Communication Technologies (WCCCT), Dalian, China, 23–25 January 2021; pp. 97–100. [CrossRef]

18. O'Shea, T.J.; Corgan, J.; Clancy, T.C. Convolutional radio modulation recognition networks. In Proceedings of the 17th International Conference Engineering Applications of Neural Networks, Aberdeen, UK, 2–5 September 2016; pp. 213–226. [CrossRef]

19. Krzyston, J.; Bhattacharjea, R.; Stark, A. Complex-valued convolutions for modulation recognition using deep learning. In Proceedings of the 2020 IEEE International Conference on Communications Workshops (ICC Workshops), Dublin, Ireland, 7–11 June 2020; pp. 1–6. [CrossRef]

20. 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]

21. 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, Republic of Korea, 21–23 October 2020; pp. 939–942. [CrossRef]

22. Emam, A.; Shalaby, M.; Aboelazm, M.A.; Bakr, H.E.A.; Mansour, H.A.A. A comparative study between CNN, LSTM, and CLDNN models in the context of radio modulation classification. In Proceedings of the 2020 12th International Conference on Electrical Engineering (ICEENG), Cairo, Egypt, 7–9 July 2020; pp. 190–195. [CrossRef]

23. Zhang, Y.; Yan, W.; Zhang, L.; Ma, L. Automatic space-time block code recognition using convolutional neural network with multi-delay features fusion. *IEEE Access* **2021**, *9*, 79994–80005. [CrossRef]

24. Huang, C.; Ji, M.; Zhang, H.; Luo, R. A multi-level complex feature mining method based on deep learning for automatic modulation recognition. In Proceedings of the 2022 3rd International Conference on Information Science, Parallel and Distributed Systems (ISPDS), Guangzhou, China, 22–24 July 2022; pp. 335–339. [CrossRef]

25. Chatterjee, S.; Tummala, P.; Speck, O.; Nürnberger, A. Complex network for complex problems: A comparative study of CNN and complex-valued CNN. In Proceedings of the 2022 IEEE 5th International Conference on Image Processing Applications and Systems (IPAS), Genova, Italy, 5–7 December 2022; pp. 1–5. [CrossRef]

26. Sun, Z.; Xu, X.; Pan, Z. SAR ATR using complex-valued CNN. In Proceedings of the 2020 Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), Dalian, China, 14–16 April 2020; pp. 125–128. [CrossRef]

27. Xiao, C.; Yang, S.; Feng, Z. Complex-valued depthwise separable convolutional neural network for automatic modulation classification. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–10. [CrossRef]

28. Ren, Y.; Huo, W.; Pei, J.; Huang, Y.; Yang, J. Automatic modulation recognition for overlapping radar signals based on multi-domain SE-ResNeXt. In Proceedings of the 2021 IEEE Radar Conference (RadarConf21), Atlanta, GA, USA, 8–14 May 2021; pp. 1–6. [CrossRef]

29. O'Shea, T.J.; Tamoghna, R.; Clancy, T.C. Over-the-air deep learning based radio signal classification. *IEEE J. Sel. Top. Signal Process* **2018**, *12*, 168–179. [CrossRef]

30. He, K.; Sun, J. Convolutional neural networks at constrained time cost. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 5353–5360. [CrossRef]

31. Chen, J.; Wong, W.-K.; Hamdaoui, B.; Elmaghbub, A.; Sivanesan, K.; Dorrance, R.; Yang, L.L. An analysis of complex-valued CNNs for RF data-driven wireless device classification. In Proceedings of the ICC 2022—IEEE International Conference on Communications, Seoul, Republic of Korea, 16–20 May 2022; pp. 4318–4323. [CrossRef]

32. Zhang, M.; Zeng, Y.; Han, Z.; Gong, Y. Automatic modulation recognition using deep learning architectures. In Proceedings of the 2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Kalamata, Greece, 25–28 June 2018; pp. 1–5. [CrossRef]

33. Hermawan, A.P.; Ginanjar, R.R.; Kim, D.-S.; Lee, J.-M. CNN-based automatic modulation classification for beyond 5G communications. *IEEE Commun. Lett.* **2020**, *24*, 1038–1041. [CrossRef]