IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 10 July 2024, accepted 29 August 2024, date of publication 2 September 2024, date of current version 10 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3453418

RESEARCH ARTICLE

Modulation Recognition Method for Wireless Signals Based on Joint Neural Networks

XUE WANG, JIAQI WANG, AND XINMIAO LU
The Higher Educational Key Laboratory for Measuring and Control Technology and Instrumentation of Heilongjiang Province, Harbin University of Science and Technology, Harbin 150080, China

Corresponding author: Xue Wang (wangxue@hrbust.edu.cn)

This work was supported in part by Heilongjiang Postdoctoral Foundation under Grant LBH-Z19168, and in part by Heilongjiang Provincial Natural Science Foundation of China under Grant YQ2022F014.

ABSTRACT In the realms of Internet of Things (IoT), satellite communication, and related scenarios, automatic modulation recognition is crucial for accurate signal demodulation. In complex communication environments, accurately identifying diverse modulation types is a challenging task. This paper introduces an automatic modulation recognition approach leveraging a joint neural network framework. The method integrates a flow-based collaborative training module for signal enhancement, a deep learning mechanism for feature extraction, and a two-dimensional sparse weighting mechanism. This method enhances the input signal through enhancement processing and strengthens attention to different dimensional features via a weighting mechanism, thereby suppressing irrelevant features with lower weights. The network architecture is optimized in terms of layer depth and connectivity to enhance modulation identification accuracy and model stability under non-ideal conditions. Experimental evaluations conducted on the RML2016.10a dataset across varying SNR demonstrate the method's robustness in low SNR environments and its effective recognition performance for high-order modulated signals compared to baseline models.

INDEX TERMS Automatic modulation recognition, joint neural network, signal enhancement module, weight sparsity, two-dimension sparse weighting mechanism.

I. INTRODUCTION

As we enter the era of the Internet of Things, an increasing variety of signals have become ubiquitous in our daily lives. In order to extract and exploit the information carried by these signals, it is necessary to recognize and demodulate the signal modulation first. Automatic modulation recognition (AMR) technology obtains crucial information such as bandwidth requirements, modulation parameters, and data transmission rates necessary for demodulation, through the examination of the modulation method. This technology provides technical support and assurance for signal reception and processing. This technology is initially applied in the military domain, such as radar signal reconnaissance and signal interception [1], [2]. Currently, it has found extensive applications in various fields including wireless communications, smart homes, intelligent transportation, and spectrum management [3].

The conventional approaches can be extensively sorted into 2 primary domains: likelihood-based identification approaches that rely on maximum likelihood theory [4], and statistical pattern recognition techniques utilize feature extraction [5]. However, both methods have inherent limitations. For instance, likelihood-based methods suffer from relatively high computational complexity, while feature-based methods heavily depend on the choice of features, the subjective nature of the task can pose challenges. The wireless communication landscape is evolving into a realm of ever-increasing intricacy and diversity, the aforementioned methods are unable to meet the demands of current wireless communication requirements. In 2006, Hinton et al. played a significant role in advancing deep learning (DL) [6], a technique applied to tackle various challenges in wireless communication [7]. It has demonstrated remarkable

The associate editor coordinating the review of this manuscript and approving it for publication was Tawfik Al-Hadhrami.

performance in the physical layer [8] and has been applied in many areas, including signal modulation recognition [9], channel estimation [10], interference coordination [11], and more.

These neural networks possess the innate ability to automatically optimize the extracted features with the aim of minimizing classification errors. Different AMR applications have utilized various neural network architectures, including CNN [12], DNN [13], RNN [14], LSTM [15], and GRU [16]. To enhance the recognition accuracy of AMR, many experts have conducted research focusing on signal features and noise elimination. A novel CNN based on spectral analysis is designed by Zeng et al. for modulation recognition [17]. This method utilizes the varying frequency characteristics over time for different modulation types. The results demonstrate significant performance improvement. However, it should be noted that this approach requires additional memory space due to its complexity. The literature proposes to map the extracted subset of features from modulated signals onto a graph and recognize the modulation type through a graph convolutional neural network [18]. Many modulation recognition methods have incorporated attention mechanisms to emphasize the salient features of the signal. Lin et al. introduce a time-frequency focus mechanism after the four-layer CNN to enhance the learning of recognition-relevant features by incorporating both temporal and frequency contexts [19]. Zhang W et al. use channel attention mechanism, spatial attention mechanism, and a four-layer CNN [20]. These studies mostly focus on capturing distinct features through attentional operations, but lack consideration of the contribution factors of weights. By utilizing the contribution factors of weights, it is possible to further suppress non-significant channels or pixels [21]. Signal distortion will have a certain impact on AMR results. The neural network used for AMR lacks the capability to eliminate signal distortion. The proposed correction module in the reference demonstrates the ability to effectively mitigate the influence of arbitrary frequency and phase distortions, even without pre-existing knowledge of these parameters [22]. Ke and Vikalo recommend a modulation framework building upon an LSTM denoising autoencoder, which simultaneously trains to restore distorted signals [23]. A signal enhancement module consisting of 11 dilated CNN layers is proposed to enhance the signal by multiplying the input spectrogram with the estimated ratio mask [19].

Despite the advancement of information technology, the susceptibility to interference from factors like noise remains a challenge that needs to be addressed. The different dimensional features of signals also have an impact on the recognition results to some extent. We introduce an innovative approach in this paper for identifying communication signal modulation using a joint neural network (JNNet). This method includes signal enhancement processing and signal modulation recognition modules. The recognition model incorporates, at its frontend, a collaborative training flow-based signal enhancement module, employed for the reconstruction, denoising, and enhancement of the original signal. This process effectively emphasizes relevant signals while mitigating the impact of noise on the recognition outcomes. The signal modulation recognition module consists of CNN, two-dimension sparse weighting mechanism (2DSWM), GRU and Dense layer. The introduced 2DSWM includes Channel-dimension weighting sub-block (CDWSB) and Spatial-dimension weighting sub-block (SDWSB), allowing the neural network to focus on the two-dimensions of signal features, allocating limited information processing resources to useful information. To consider the suppression of unimportant features, we introduced a regularization term in the loss function to achieve weight sparsity and promote the suppression of irrelevant features. In the basic framework of the neural network, we used two CNN layers with different convolution kernel sizes to focus on feature information of different sizes and improve network performance. Additionally, considering the temporal characteristics of the signal and the model's lightweight nature, we introduced the GRU to capture time-dependent features within the data. The outlined aspects are listed:

(1) To enhance the efficiency of AMR, instead of reconstructing the undistorted signal for signal modulation identification, we improve the existing signal to minimize the impact of environmental and channel variables. Through collaborative training, the flow-based signal enhancement module continuously optimizes the parameters via forward and backward propagation, processing the signal to achieve the enhancement of useful information.

(2) This paper integrates the 2DSWM into the joint neural network framework. The objective of this operation aims to emphasize the input characteristics' significant components. Additionally, we also consider the suppression of unimportant features, which makes the weights in the different dimensions sparse and efficiently utilizes limited resources.

(3) We suggest a joint neural network-based approach for AMR that is efficient. It employs two layers of CNN to extract features from input data at various scales. In addition, we consider the temporal properties of the signal by incorporating GRU.

(4) Under identical dataset and experimental conditions, we demonstrate the effectiveness of the signal enhancement module and the proposed method through ablation and comparative experiments.

The rest of the paper is organized as follows. In Section II, we discuss the signal model and analyze the AMR's problem formulation. Section III presents a comprehensive overview of the proposed AMR framework, including modules for signal enhancement and 2DSWM. Experimental simulations demonstrating the effectiveness of our approach are showcased in Section IV. In conclusion, Section V provides a summary of the discoveries.

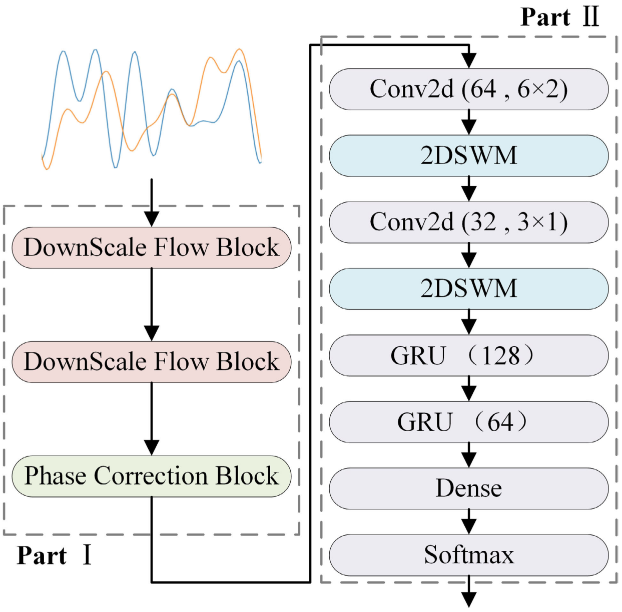


FIGURE 1. The proposed signal modulation recognition model framework.

II. SIGNAL MODEL AND AMR'S PROBLEM FORMULATION

A. THE SIGNAL MODEL IS DISCUSSED

The communication system under consideration in this paper consists solely of a transmitter and a receiver. The signal $y(t)$ received at time t:

$$y(t) = h(t)e^{j(2\pi f_0(t)+\varphi_0)}x(t) + \omega(t) \tag{1}$$

where $h(t)$ describes the pulse effect caused by the propagation of wireless signals through the channel. $f_0(t)$ and φ_0 respectively denote the deviation in frequency and phase carried by the signal upon reception. $x(t)$ describes a complex baseband signal. The noise referred to as $\omega(t)$ and it exhibits a Gaussian distribution with 0 mean and variance, and its power spectral density of $N_0/2$.

In the field of communication systems, it is common to express signals using I/Q(In-phase/Quadrature) signals for the purpose of simplifying data processing and modulation identification as $\mathbf{Y}^{L\times 2}$ with L symbols in a sampled signal

$$\mathbf{Y} = \begin{bmatrix} I\{y[1]\}, \ldots, I\{y[L]\} \\ Q\{y[1]\}, \ldots, Q\{y[L]\} \end{bmatrix}^T \tag{2}$$

This process performs an I/Q mapping decomposition of the signal. In the practical implementation of DL in AMR, the input signal typically consists of a real matrix containing I/Q components for subsequent identification. The entirety of information within a signal sample is typically encapsulated within its real and imaginary components. These components typically follow identical independent distributions, thereby obviating the necessity for normalization prior to their utilization as inputs for neural networks.

B. AMR'S PROBLEM STATEMENT

Recognizing modulation involves the challenge of distinguishing the modulation scheme used in a received signal from a set of N potential modulation schemes. First, the received signal sample matrix $\mathbf{Y}^{L\times 2}$ is mapped to an S-dimensional tensor e after layer by layer feature extraction in the joint neural network

$$\mathbf{Y}^{L\times 2} \rightarrow e \in \mathbf{Y}^S \tag{3}$$

Tensor e is the final extracted feature, which is processed by the Softmax function in the Dense layer. The raw data output by each node first exponentiated by the Softmax function, followed by normalization. Specifically, Softmax output $\mathcal{Y}_j$ of $j \in [1, S]$ th node

$$\mathcal{Y}_j = \frac{\exp(e_j)}{\sum_S \exp(e_S)} \tag{4}$$

where e_j represent the raw output of the j-th node. That is, the output of each node falls between 0 and 1, and the sum of the outputs of all nodes is equal to 1. To determine the identification result, the model calculates the probabilities for each modulation category. These probabilities are normalized so that their sum is equal to 1, and the highest probability corresponds to the identified modulation category. The final modulation recognition results are closely related to the features extracted by the neural network. In this paper, an attention mechanism is added between different layers of neural networks to help the model better self-adapt to the important features and information, and improve the model's efficiency and capacity to apply learned knowledge in various scenarios.

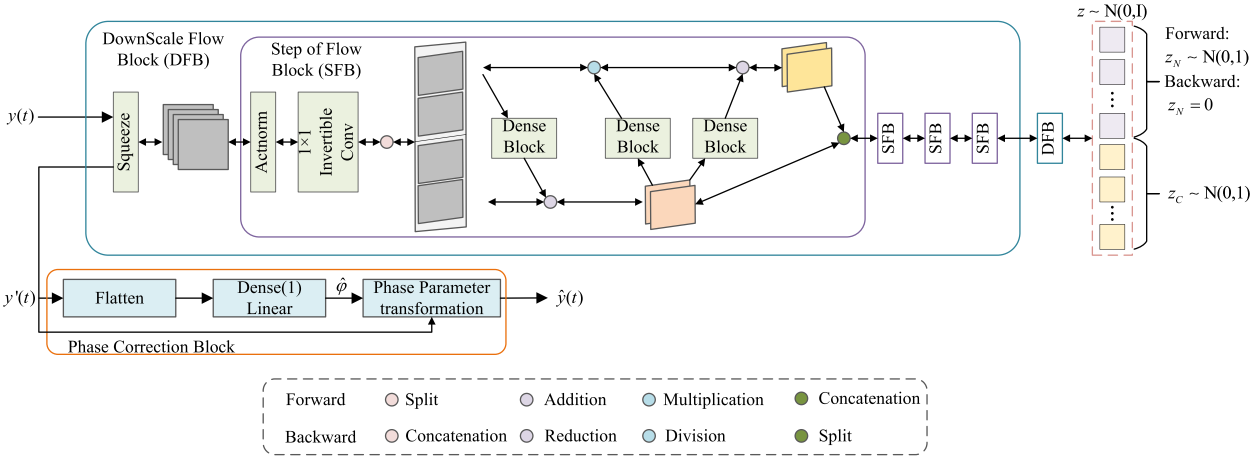


FIGURE 2. The overall architecture of the flow-based signal enhancement module within collaborative training.

III. MODULATION RECOGNITION FRAMEWORK

Our paper presents a novel approach to identifying modulations, which involves a flow-based signal enhancement module for collaborative training (Part I) in conjunction with a modulation recognition module that employs a joint neural network (Part II). Figure 1 provides an overview of the overall approach. We present a module designed to enhance signals in the preprocessing stage, aiming to reduce noise and rectify phase shifts. Moreover, through collaborative training among neural networks, the continual guidance of model parameter updates is pursued to attain optimal outputs. This leads to improved signal recognition accuracy even in scenarios with low SNR. We use a joint neural network composed of CNN, GRU, and Dense layers to achieve modulation recognition of wireless signals. 2DSWM is incorporated between different neural network layers to select features, emphasizing useful features while limiting unimportant ones.

A. FLOW-BASED SIGNAL ENHANCEMENT MODULE FOR COLLABORATIVE TRAINING

In real-world communication environments, signals are often accompanied by noise. Moreover, they are susceptible to interference, leading to the phenomenon of phase offset. To alleviate the negative effects of these factors on the AMR results and achieve signal enhancement, we propose a Flow-based signal enhancement module for collaborative training. The overall framework is illustrated in the accompanying Figure 2. We employ an invertible neural network based on normalizing flows to learn the distribution of signals [24], [25], supplemented by a neural network capable of estimating phase offset parameters [26]. Throughout the training phase, we fully utilize the forward and backward propagation mechanisms of the invertible neural network. By iteratively adjusting network parameters, the network is enabled to better fit the input signals. Furthermore, we employ a tightly integrated collaborative training strategy among modules, enabling mutual support and joint optimization of network parameter updates across the various modules. This training methodology contributes to enhancing the network's generalization ability, enabling it to achieve better performance when dealing with new signals. The overall model can be segmented into three components: DownScale Flow Block (DFB), Step of Flow Block (SFB), and Phase Correction Block (PCB). Next, we will provide a detailed explanation of the functions of each layer.

TABLE 1. The operations of Actnorm and 1 × 1 invertible convolutional layers.

Layer	Actnorm	1×1 Invertible Conv
Forward	$\mathbf{y}_{i+1} = \mathbf{s} \odot \mathbf{y}_i + \mathbf{b}$	$\mathbf{y}_{i+1} = \mathbf{W}\mathbf{y}_i$
Backward	$\mathbf{y}_i = (\mathbf{y}_{i+1} - \mathbf{b}) \odot \mathbf{s}^{-1}$	$\mathbf{y}_i = \mathbf{W}^{-1}\mathbf{y}_{i+1}$

DownScale Flow Block: To achieve better training efficacy, we incorporate two DFBs connected hierarchically within the signal enhancement module. Among these, the Squeeze layer selects elements from the latent variables within the original input data and combines the selected elements to form a new latent representation with fourfold channels. Following the Squeeze layer, the feature map undergoes scale reduction while an increase in the number of channels, thereby extracting richer and more abstract feature representations. After the Squeeze layer, we cascade four SFBs to increase the model's depth and intricacy, enhancing the richness of feature representation, and improving information propagation and feedback.

Step of Flow Block: The Actnorm layer performs normalization to ensure that the input data conforms to $N(0, 1)$ Gaussian distribution. 1×1 Invertible Convolutional layer, unlike traditional convolutional layers, possesses reversibility. The operations of these two layers can be represented as follows, as shown in Table 1. $\mathbf{Y}_i$ and $\mathbf{Y}_{i+1}$ represent latent representations. $\mathbf{s}$ and $\mathbf{b}$ denote scale and offset parameters, utilized for scaling and offsetting the normalized values. $\mathbf{W}$

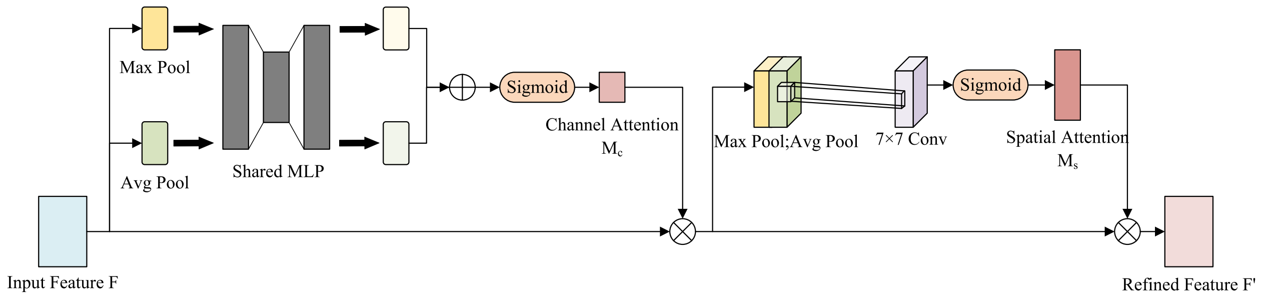


FIGURE 3. 2DSWM: Two-dimension sparse weighting mechanism.

TABLE 2. The operations of affine coupling layer (forward and backward).

Forward	Backward
$\mathbf{y}_i^a, \mathbf{y}_i^b = Split(\mathbf{y}_i)$	$\mathbf{y}_{i+1}^a, \mathbf{y}_{i+1}^b = Split(\mathbf{y}_{i+1})$
$\mathbf{y}_{i+1}^a = \mathbf{y}_i^a + g(\mathbf{y}_i^b)$	$\mathbf{y}_i^b = (\mathbf{y}_{i+1}^b - g(\mathbf{y}_{i+1}^a)) \odot g^{-1}(\mathbf{y}_{i+1}^a)$
$\mathbf{y}_{i+1}^b = g(\mathbf{y}_{i+1}^a) \odot \mathbf{y}_i^b + g(\mathbf{y}_{i+1}^a)$	$\mathbf{y}_i^a = \mathbf{y}_{i+1}^a - g(\mathbf{y}_i^b)$
$\mathbf{y}_{i+1} = Concat(\mathbf{y}_{i+1}^a, \mathbf{y}_{i+1}^b)$	$\mathbf{y}_i = Concat(\mathbf{y}_i^a, \mathbf{y}_i^b)$

represents a randomly initialized invertible matrix, which is continuously updated through backpropagation.

The remaining portion is designated as the affine coupling layer, where forward and backward propagations are sequentially depicted according to the model's cascade order, as illustrated in the following Table 2. Where, $Split(\cdot)$ and $Contact(\cdot)$ are performed along the channel dimension. $g(\cdot)$ represents a neural network composed of Dense Blocks.

Phase Correction Block: As a result of multiple factors within the wireless channel, the phase difference between the I and Q signals is not ideally 90°. This deviation in phase may potentially impact the final recognition outcome. Flatten layer and Dense layer are employed to transform the matrix of I/Q signal samples from multi-dimensional to one-dimensional vectors. Linear activation functions are utilized to generate an approximate phase parameter denoted as $\hat{\varphi}$. Let the coordinates of the original signal in the complex plane be denoted as $M(p, q)$, and the signal after phase offset be represented as $N(p', q')$. The phase offset encompasses both phase rotation and translation, as depicted by the following equation.

$$\begin{cases} p = Z\cos\alpha \\ q = Z\sin\alpha \end{cases}$$
$$\begin{cases} p' = Z\cos(\alpha+\sigma)+e \\ q' = Z\sin(\alpha+\sigma)+f \end{cases}$$
$$\rightarrow \begin{cases} p' = p\cos\sigma - q\sin\sigma + e \\ q' = q\cos\sigma + p\sin\sigma + f \end{cases}$$
$$\rightarrow \begin{bmatrix} p' \\ q' \end{bmatrix} = \begin{bmatrix} \cos\sigma & -\sin\sigma \\ \sin\sigma & \cos\sigma \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix} \tag{5}$$

Without loss of generality, the estimated phase parameter $\hat{\varphi}$ includes both phase rotation-related parameters $\begin{bmatrix} \tilde{a} & \tilde{b} \\ \tilde{c} & \tilde{d} \end{bmatrix}$ and phase shift-related parameters $\begin{bmatrix} \tilde{e} \\ \tilde{f} \end{bmatrix}$. The phase parameter transformer performs a parameter inverse transformation on the input signal according to phase shift parameter $\hat{\varphi}$. We obtain the corrected output signal $\hat{y}(t)$ as follows.

$$\begin{aligned} \hat{y}(t) &= y'(t)e^{-j\hat{\varphi}} \\ &= \begin{bmatrix} I\{y'(t)\}\cos\hat{\varphi} + Q\{y'(t)\}\sin\hat{\varphi} \\ I\{y'(t)\}\cos\hat{\varphi} - Q\{y'(t)\}\sin\hat{\varphi} \end{bmatrix} \end{aligned} \tag{6}$$

B. SE-JNNET: A JOINT NEURAL NETWORK FOR AMR

We introduce a AMR module that leverages a collaborative neural network architecture, in this section. This module is composed of tow CNN layers, followed by two GRU layers, and finally one Dense layer. 2DSWM is inserted after each CNN layer.

First, the modulated signals that have undergone signal enhancement processing are sequentially input into CNN layers with 64 and 32 channels, using convolutional kernels of size 6 × 2 and 3 × 1. These convolutional layers extract feature information of different sizes. The global contextual information is captured by the larger kernels, whereas the smaller kernels capture regional particulars. The utilization of ReLU activation functions in these two CNN layers aims to tackle the issue of gradient disappearance. The 2DSWM is inserted after each CNN layer focuses on the required feature information for this experiment, suppressing non-important features. This reduces redundancy and noise, simultaneously enhancing the model's emphasis on critical attributes, leading to an enhancement in the model's capability to represent features.

The fusion of inter-channel interaction information and spatial information in the convolution operation enables efficient extraction of features from the input data. To maximize the selection of key features that are crucial for recognition, this paper introduces the 2DSWM into the modulation recognition model, as indicated in Figure 3. The 2DSWM comprises two components primarily: CDWSB and CDWSB. In order to enhance the focus on meaningful features, we compute attention weights along the main degrees of the channel and the spatial axis respectively, which can focus more on the regions of interest while reducing parameters number and computational complexity. To suppress features that are not significant for recognition, we have introduced loss functions with $\ell 1$ regularization terms in the weight calculation for both attention sub-modules. By increasing the sparsity of the weights, the model is able to efficiently compute while maintaining performance.

The output feature F $\in$ $R^{V\times H\times C}$ of the previous convolutional layer is passed to the attention mechanism as input. Here, V and H express the vertical and horizontal dimensions correspondingly, and C represents the quantity of channels. First, CDWSB obtains the attention weight $\mathrm{M_c}$ on the channel dimension according to the input feature F. By calculating the product of $\mathrm{M_c(F)}$ and F, we obtain the refined feature $\mathrm{F_c}$ after the CDWSB. Next, $\mathrm{F_c}$ undergoes SDWSB in order to acquire knowledge of $\mathrm{M_s(F_c)}$. We obtain the refined feature $\mathrm{F_s}$ after the SDWSB module. The procedure is depicted as follows

$$\mathrm{F_c = M_c(F) \otimes F} \tag{7}$$

$$\mathrm{F_s = M_s(F_c) \otimes F_c} \tag{8}$$

In the CDWSB, the input F is passed through parallel Max Pooling (MP) and Global Average Pooling (GAP) layers to generate two different feature maps. These two feature maps are then used as inputs and transmitted through a shared network. The network that is shared includes a multi-layer perceptron with two layers with high connectivity and one hidden layer. The initial layer is comprised of C/r neurons, with r representing the decay factor, and it employs the ReLU activation function. Following this, the subsequent layer consists of C neurons. After passing through the shared network, the two feature maps are added together. The sum obtained is subsequently fed into a sigmoid activation function in order to produce the CDWSB weights $\mathrm{M_c}$. After acquiring the CDWSB weights, the initial feature map is subjected to multiplication with the feature map in order to function as input for the SDWSB module. In the SDWSB module, two feature maps are obtained by applying MP and GAP layers to the input. The two feature maps are combined by joining them together in the channel dimension. The concatenated result is further passed through a 7 × 7 convolutional layer to generate a two-dimensional SDWSB map. This map undergoes a sigmoid activation function, resulting in the production of SDWSB weights denoted as $\mathrm{F_s}$.

$$\mathrm{M_c(F) = \ Sigmoid(MLP(AvgPool(F){+}MaxPool(F)))} \tag{9}$$

$$\mathrm{M_s(F_c) = Sigmoid\{}f^{7\times 7}\mathrm{([AvgPool(F_c);\ MaxPool(F_c)])\}} \tag{10}$$

We incorporate a loss function with $\ell 1$ regularization terms in both the CDWSB and SDWSB to preserve weights that contribute significantly to the model's recognition performance while filtering out others. This further enhances the focus on important weights, resulting in a final weight vector with sparsity. By reducing the number of parameters while maintaining good performance. The following equations describe these loss functions.

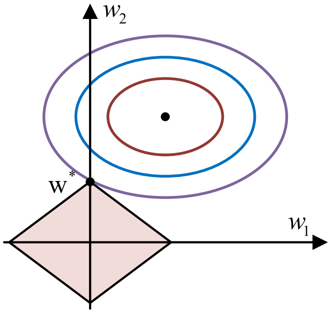


FIGURE 4. $\ell 1$ **regularization with two weights.**

$$Loss_c = L(\theta_c) + \lambda \sum\nolimits_{i=1}^{n} |w_{ci}| \tag{11}$$

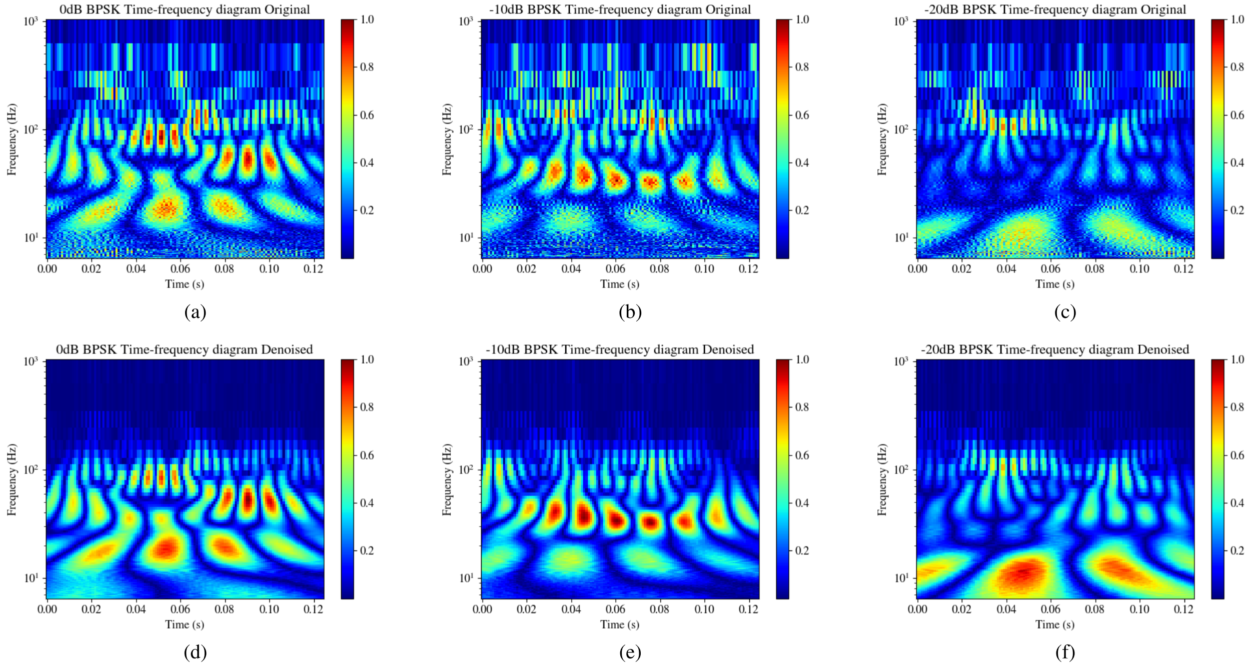
$$Loss_s = L(\theta_s) + \lambda \sum\nolimits_{i=1}^{n} |w_{si}| \tag{12}$$

where $Loss_c$ and $Loss_s$ represent the loss functions added to the both sub-modules, respectively. $L(\theta_c)$ and $L(\theta_s)$ represent the original loss functions without regularization terms. θ_c and θ_s represent the parameters of the model, including weights and biases. λ is a hyperparameter that governs the intensity of regularization. $\sum_{i=1}^{n} |w_i|$ is the regularization term, calculated as the total of the absolute magnitudes of all elements in the weight vector w.

Figure 4 shows a schematic diagram of $\ell 1$ regularization when considering two weights w_1 and w_2. The rectangular lines represent the range of values for the two weights in $\ell 1$ regularization. The colored circles represent the contour lines of the original loss function being optimized. When we add a regularization term to the original loss function, it is equivalent to impose a constraint on the original loss function. Subject to this constraint, we can find the lowest value of the original loss function. From another perspective, satisfying the regularization condition is actually finding the intersection point between the rectangular and circular regions, that is, simultaneously satisfying the constraints on the weights and minimizing the loss function. For $\ell 1$ regularization, the constraint region is a square, so the probability of the intersection point being a vertex of the circular region is high. Therefore, there is a high probability that one of the weights w_1 or w_2 will be zero, leading to sparsity in the solution obtained with $\ell 1$ regularization.

The $\ell 1$ regularization term is implemented to enhance the sparsity of attention weights, focusing the model's attention on useful features for the task, and reducing redundancy. This helps improve the feature representation and discriminative power of the model, further enhancing its performance and generalization ability.

Next are two GRU layers. The introduction of GRU helps to capture time-related features more effectively and speeds up the model's training process. Compared to RNN and LSTM, GRU introduces update gates and reset gates, which



(a) (b) (c)

(d) (e) (f)

FIGURE 5. Comparison of the spectrograms of the original and denoised BPSK signals at SNR values of 0dB, −10dB, and −20dB.

TABLE 3. Signal modulation recognition module architecture information.

Layer	Channels	Filter Size	2DSWM	Activation
Conv2d-1	64	6×2	Y	ReLu
Conv2d-2	32	3×1	Y	ReLu
GRU	128	-	-	-
GRU	64	-	-	-
Dense	11	-	-	Softmax

TABLE 4. RML2016.10a dataset information.

Dataset	RML2016.10a
Generation method	Gnuradio+Python
Total number of samples	220000
Modulation schemes	11 modulation schemes: 8PSK, BPSK, QPSK, CPFSK, GFSK, PAM4, 16QAM, 64QAM, AM-DSB, AM-SSB, WBFM
Range of SNR	[−20dB, 18dB], at intervals of 2 dB
Data format	IQ (In-phase and Quadrature) data format: 2x128
Environment of the channel	Additive Gaussian white noise, selective fading, center frequency offset, sampling rate offset

better control the flow of information and handle long-term dependencies, addressing the issues of vanishing and exploding gradients in long sequential data. Moreover, GRU has a simpler structure, fewer parameters, and is easier to train. The first GRU layer's output is also in sequence form. This enables subsequent layers to further analyze the output from the initial GRU layer and extract latent features from the input sequence. To generate the final classification results, we use a Dense layer as the concluding component within our model architecture. The output of this particular layer undergoes a softmax activation function, which transforms it into probabilities for each individual class. Based on this probability distribution, we can predict the category to which a sample belongs. The cross-entropy loss function plays a crucial role in directing the modifications of model parameters throughout the training process. Accuracy is also used as an evaluation metric to assess the model's classification performance. For the choice of optimizer, we utilized the Adam optimizer, a widely adopted adaptive learning rate optimization algorithm known for efficiently updating model parameters and accelerating convergence during training. This adaptability proves especially beneficial in handling complex tasks like AMR, ensuring robust and stable performance across diverse modulation types and dataset distributions. While considering alternative optimizers such as SGDm and RMSProp, our experimental evaluations consistently demonstrated Adam's superior suitability for our specific tasks and dataset characteristics. The relevant parameters for Part II are shown in Table 3.

IV. EXPERIMENTAL AND ANALYTICAL DISCUSSION

A. RELATED EXPERIMENTAL CONDITIONS SETTINGS

The dataset used in this paper is the open dataset RML2016.10a [27]. Table 4 presents pertinent details regarding the dataset. This particular dataset is generated by GNU Radio and exhibits good performance characteristics. The dataset, which contains 220,000 randomly partitioned modulated signals, is distributed with a ratio of 6:2:2 = training: validation: test. The categorical classification employs the

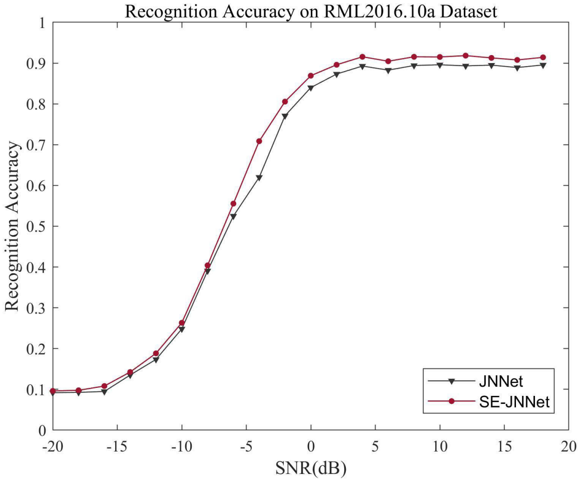


FIGURE 6. Comparison of recognition accuracy between JNNet and SE-JNNet.

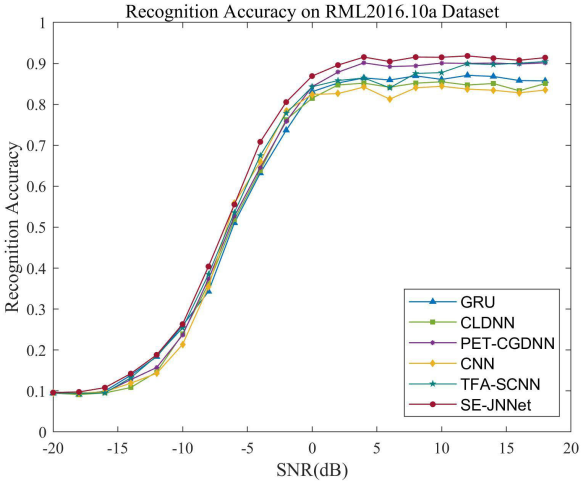


FIGURE 7. Assessment of modulation recognition precision across six different models.

cross-entropy loss function to minimize the disparity between predicted probabilities for class membership and actual class labels. Equation (13), where M represents the number of classes, $M = 11$ in this paper. $\mathcal{Y}_{ic}$ indicates a match between the true class of sample i and c by assigning a value of 1, and otherwise, it remains at zero. $\mathcal{P}_{ic}$ signifies the forecasted likelihood of observation sample i being associated with class c.

$$L = \frac{1}{N}\sum\nolimits_i L_i = -\frac{1}{N}\sum\nolimits_i \sum\nolimits_{c=1}^{M} \mathcal{Y}_{ic}\log(\mathcal{P}_{ic}) \qquad (13)$$

The Adam optimizer, in addition, is employed to expedite the model's convergence. To mitigate overfitting in the neural network, if there is no reduction in validation loss for 5 consecutive epochs, the learning rate is halved. Moreover, if there is no decrease in validation loss after 50 epochs, the training process is halted. The experiments are carried out utilizing the NVIDIA GeForce RTX 3080 platform, employing Keras with Tensorflow as the underlying framework.

B. SIGNAL ENHANCEMENT MODULE ABLATION EXPERIMENT

A spectrogram is a visual representation that combines both the temporal and spectral aspects of a signal, depicting the intensity of the signal using varying colors. As shown in Figure 5, we take the BPSK signals as examples to visualize the effect of the signal enhancement module. On the horizontal axis, time is depicted, while frequency is represented on the vertical axis. The color intensity signifies the normalized signal amplitude at each position within the time-frequency domain. The legend located at the right side of the spectrogram displays the normalized amplitude values of the signal, where darker shades represent comparatively elevated amplitudes in terms of time-frequency positioning. (a), (b), and (c) depict the time-frequency diagrams of the BPSK signal at 0dB-10dB, −20dB, respectively. (d), (e), and (f) depict the processed signals after the signal enhancement module. In the above three figures, noise typically appears as a diffused distribution on the spectrogram, which severely affects the clarity and distinguishability of the useful signal. In contrast, the useful signal tends to concentrate within a specific range of time and frequency on the spectrogram. After the denoising process, the signal contour becomes clearer, and the strength of the useful signal is enhanced. Most of the diffused noise intensity has been reduced or even eliminated. The aforementioned statement exemplifies the remarkable efficacy of the proposed signal enhancement module in mitigating noise and enhancing signal quality, thereby augmenting the discernibility and effectiveness of the signal.

The ablation experiments are proposed to assess the impact of the suggested signal enhancement module on modulation recognition outcomes, aiming to ascertain its efficacy. Two experiments are conducted: one with only the AMR module (JNNet), and another with the signal enhancement module added to the AMR module (SE-JNNet). The effectiveness of the signal enhancement module is discussed by observing the recognition accuracy on the dataset.

According to the comparison of recognition accuracy shown in Figure 6, it is clear that the model's ability to identify signals has been improved across a wide range of −20dB to 18dB upon integration of the signal enhancement module. The recognition accuracy has been improved by 0.44% to 3.95%. Clearly, the proposed model demonstrates improve robustness to noise, enabling better performance in various noisy environments for the recognition task. Therefore, the inclusion of the signal enhancement module has distinct advantages in enhancing the model's robustness and performance.

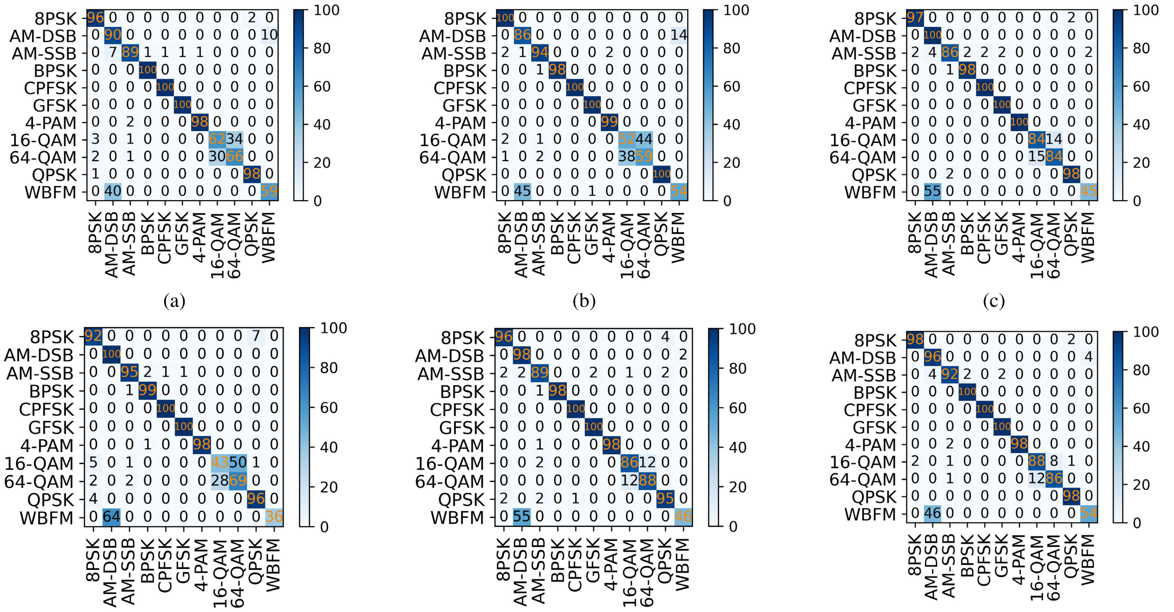
C. EXPERIMENTAL SIMULATION AND COMPARISON

1) BASELINE MODEL

The SE-JNNet model's performance is compared to the GRU2 [14], CLDNN [28], PET-CGDNN [26], CNN [29], and TFA-SCNN [30] approaches.

TABLE 5. Comparison of six models on the RML2016.10a dataset.

Model	-20dB	-10dB	0dB	10dB	18dB	Highest Accuracy
GRU2	9.45%	26.00%	83.09%	86.09%	85.73%	87.09%
CLDNN2	9.50%	23.95%	81.50%	85.50%	85.09%	85.50%
PET-CGDNN	9.50%	23.68%	84.32%	90.09%	90.14%	90.13%
CNN	9.61%	21.36%	82.41%	84.41%	83.50%	84.40%
TFA-SCNN	9.53%	25.41%	84.30%	87.75%	90.45%	90.36%
SE-JNNet	9.63%	26.33%	86.91%	91.50%	91.41%	91.83%



(a) (b) (c)

(d) (e) (f)

FIGURE 8. Confusion matrix of the six models at their highest accuracy: (a) GRU2 (SNR = 12dB, Highest accuracy = 87.09%); (b) CLDNN2 (SNR = 10dB, Highest accuracy = 85.50%); (c) PET-CGDNN (SNR = 18dB, Highest accuracy = 90.13%); (d) CNN (SNR = 10dB, Highest accuracy = 84.40%); (e) TFA-SCNN (SNR = 18dB, Highest accuracy = 90.36%); (f) SE-JNNet (SNR = 12dB, Highest accuracy = 91.83%).

The identical experimental platform and dataset are used to conduct experiments on these five models, as well as the SE-JNNet model. The experimental results are analyzed from two aspects, overall recognition accuracy and recognition accuracy of 11 modulation signals on the model. Specifically, analyze the model complexity and recognition accuracy.

2) MODEL COMPARISON ANALYSIS

To visually analyze the disparity in recognition accuracy between the SE-JNNet model and the baseline model across various SNR levels. The contrast of recognition accuracy curves for the six models in the range of −20dB to 18dB is illustrated in Figure 7. As the SNR increases, the accuracy of all six modulation types gradually improves. After the SNR exceeds 4dB, the accuracy tends to stabilize. In the range of −20dB to −14dB. There is not much difference in accuracy between the SE-JNNet model and the other baseline models. Nonetheless, in cases where the SNR exceeds −14dB. The proposed model surpasses the other models, exhibiting an accuracy enhancement ranging from 0.15% to 7.45%. The SE-JNNet model exhibits significantly better recognition performance compared to the other models, particularly when the SNR exceeds −4dB. Table 5 shows the level of accuracy in recognizing the six models at various SNR levels. SNR values include −20dB, −10dB, 0dB, 10dB, and 18dB. The table also includes the highest recognition accuracy achieved by each model. This method achieves the highest recognition accuracy without any prior knowledge or human involvement. It has an improvement of 4.74%, 6.33%, 1.70%, 7.43% and 1.47% compared to GRU2, CLDNN2, PET-CGDNN, CNN, and TFA-SCNN, respectively.

Figure 8 displays the confusion matrices for the 11 modulation types obtained by the six models when achieving their highest recognition accuracy. The confusion matrix uses rows to represent the true type labels and columns to represent the estimated type labels. The matrix contains values representing the precision of AMR, with the diagonal elements indicating the accuracy of each specific modulation type. Analysis shows that the proposed model exhibits the highest discriminability for complex modulation signals like 16QAM and 64QAM, which have a tendency to be confusion.

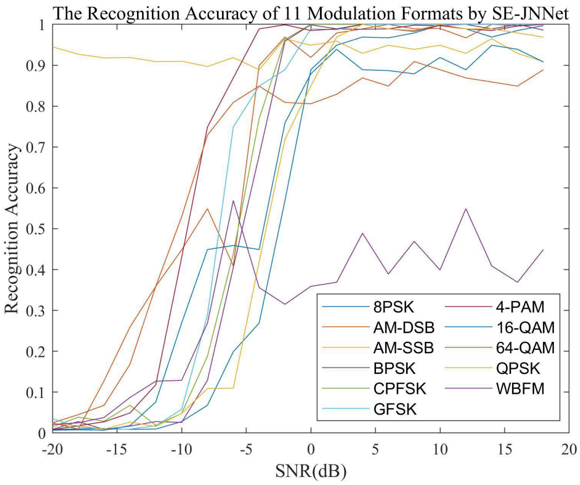


FIGURE 9. The performance of the SE-JNNet model in recognizing 11 modulation formats.

TABLE 6. Comparing the computational complexity of the six methods.

Model	Total params	Trainning time
GRU2	151,179	3s/epoch
CLDNN2	517,643	8s/epoch
PET-CGDNN	71,871	3s/epoch
CNN	858,123	6s/epoch
TFA-SCNN	104,792	3s/epoch
SE-JNNet	277,716	4s/epoch

Apart from the models' accuracy on the dataset, this research also compared the complexities of the six models. Table 6 showcases the total parameters and training time per epoch for the six models with the batch size of 1000.

The capability of the proposed model with regard to total parameters and training time per epoch may not be optimal. This is because, when introducing attention mechanisms, it is common to incorporate additional parameters to learn weight allocation for channels and spatial locations. As a result, the training time increases. Yet, these extra computational expenses are necessary to attain improved recognition accuracy. The proposed model demonstrates good performance in recognition accuracy. This compensates for the additional computational costs by improving the recognition performance.

3) PERFORMANCE OF THE SE-JNNet MODEL

Figure 9 displays the accuracy curves of the 11 modulation types considered in the proposed model, over the [−20dB, 18dB] SNR ratio range. The analysis reveals that the recognition accuracy for all 11 modulation types improves with increasing SNR and stabilizes after a certain threshold, with minimal fluctuations. The recognition accuracy for most modulation types approaches 100%. However, 16QAM and 64QAM, which use a set of symbols to represent multiple bits, have lower tolerance to noise during transmission, resulting in increased confusion between these two signals. The proposed model achieves recognition accuracies of over 90% for 16QAM and 64QAM, but its performance decreases when the SNR ratio drops below −10dB. It's important to mention that the recognition accuracy of WBFM (Wideband Frequency Modulation) falls below 60%. This can be explained by the origin of the WBFM modulation signal, which is obtained from a real-time audio dataset. The dataset may potentially encompass instances of silence or glitches in the speech signal. As a result, the WBFM-modulated signal may not contain valid information, making it challenging for the classifier to correctly identify it.

V. CONCLUSION

The present paper introduces an innovative approach to the recognition of wireless signal modulation, employing a joint neural network. The method consists of three modules: signal enhancement module, 2DSWM module, and modulation recognition module. These modules are cascaded and jointly trained to perform modulation recognition. Moreover, a flow-based signal enhancement module for collaborative training is employed to enhance signals. The signal enhancement processing module is introduced to address the noise impact and phase shift caused by communication environment and channel, minimizing the impact of extraneous data on the outcome of AMR. To enhance the efficiency of modulation recognition, we incorporate a 2DSWM module into our proposed approach. This module enhances the attention to different feature dimensions and suppresses unimportant features by considering their contribution factors, making more efficient use of limited resources. We cascade CNN, GRU and Dense layers to form our proposed joint neural network framework. Different CNN layers extract features of different sizes, and the time features of the signal are considered by two GRU layers, which additionally derive data characteristics using a relatively small number of parameters and training duration. The effectiveness of the SE-JNNet model is analyzed by modulation recognition accuracy and SNR on the RadioML2016.10a dataset. In the exact same experimental conditions, an evaluation is carried out to compare the proposed model with the baseline model. The findings suggest that the effectiveness of the suggested approach becomes evident when the SNR surpasses −14dB, highlighting its advantages.

REFERENCES

[1] M. Gupta and A. Mahla, "Electronic warfare: Issues and challenges for emitter classification," *Defence Sci. J.*, vol. 61, no. 3, pp. 228–234, May 2011.

[2] Z. Qu, C. Hou, C. Hou, and W. Wang, "Radar signal intra-pulse modulation recognition based on convolutional neural network and deep Q-learning network," *IEEE Access*, vol. 8, pp. 49125–49136, 2020.

[3] A. Vagollari, M. Hirschbeck, and W. Gerstacker, "An end-to-end deep learning framework for wideband signal recognition," *IEEE Access*, vol. 11, pp. 52899–52922, 2023.

[4] D. Zhu, V. J. Mathews, and D. H. Detienne, "A likelihood-based algorithm for blind identification of QAM and PSK signals," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 3417–3430, May 2018.

[5] L. Y. Uys, M. Gouws, J. J. Strydom, and A. S. J. Helberg, "The performance of feature-based classification of digital modulations under varying SNR and fading channel conditions," in *Proc. IEEE AFRICON*, Sep. 2017, pp. 198–203.

[6] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.

[7] D. Gündüz, P. de Kerret, N. D. Sidiropoulos, D. Gesbert, C. R. Murthy, and M. van der Schaar, "Machine learning in the air," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 10, pp. 2184–2199, Oct. 2019.

[8] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cognit. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[9] Y. Xu, D. Li, Z. Wang, Q. Guo, and W. Xiang, "A deep learning method based on convolutional neural network for automatic modulation classification of wireless signals," *Wireless Netw.*, vol. 25, no. 7, pp. 3735–3746, Oct. 2019.

[10] M. Soltani, V. Pourahmadi, A. Mirzaei, and H. Sheikhzadeh, "Deep learning-based channel estimation," *IEEE Commun. Lett.*, vol. 23, no. 4, pp. 652–655, Apr. 2019.

[11] Z. Zhu and A. K. Nandi, *Automatic Modulation Classification: Principles, Algorithms and Applications*. Hoboken, NJ, USA: Wiley, 2015.

[12] F. Liu, Z. Zhang, and R. Zhou, "Automatic modulation recognition based on CNN and GRU," *Tsinghua Sci. Technol.*, vol. 27, no. 2, pp. 422–431, Apr. 2022.

[13] W. Shi, D. Liu, X. Cheng, Y. Li, and Y. Zhao, "Particle swarm optimization-based deep neural network for digital modulation recognition," *IEEE Access*, vol. 7, pp. 104591–104600, 2019.

[14] D. Hong, Z. Zhang, and X. Xu, "Automatic modulation classification using recurrent neural networks," in *Proc. 3rd IEEE Int. Conf. Comput. Commun. (ICCC)*, Dec. 2017, pp. 695–700.

[15] S. Hou, Y. Fan, B. Han, Y. Li, and S. Fang, "Signal modulation recognition algorithm based on improved spatiotemporal multi-channel network," *Electronics*, vol. 12, no. 2, p. 422, Jan. 2023.

[16] L. Li, Y. Zhu, and Z. Zhu, "Automatic modulation classification using ResNeXt-GRU with deep feature fusion," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–10, 2023.

[17] Y. Zeng, M. Zhang, F. Han, Y. Gong, and J. Zhang, "Spectrum analysis and convolutional neural network for automatic modulation recognition," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 929–932, Jun. 2019.

[18] Y. Liu, Y. Liu, and C. Yang, "Modulation recognition with graph convolutional network," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 624–627, May 2020.

[19] S. Lin, Y. Zeng, and Y. Gong, "Modulation recognition using signal enhancement and multistage attention mechanism," *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 9921–9935, Nov. 2022.

[20] W. Zhang, Y. Sun, K. Xue, and A. Yao, "Research on modulation recognition algorithm based on channel and spatial self-attention mechanism," *IEEE Access*, vol. 11, pp. 68617–68631, 2023.

[21] Y. Liu, Z. Shao, Y. Teng, and N. Hoffmann, "NAM: Normalization-based attention module," in *Proc. 35th Conf. Neural Inf. Process. Syst.*, Sydney, NSW, Australia, 2021, pp. 1–5.

[22] K. Yashashwi, A. Sethi, and P. Chaporkar, "A learnable distortion correction module for modulation recognition," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 77–80, Feb. 2019.

[23] Z. Ke and H. Vikalo, "Real-time radio technology and modulation classification via an LSTM auto-encoder," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 370–382, Jan. 2022.

[24] J.-J. Huang and P. L. Dragotti, "WINNet: Wavelet-inspired invertible network for image denoising," *IEEE Trans. Image Process.*, vol. 31, pp. 4377–4392, 2022.

[25] Y. Liu, S. Anwar, Z. Qin, P. Ji, S. Caldwell, and T. Gedeon, "Disentangling noise from images: A flow-based image denoising neural network," *Sensors*, vol. 22, no. 24, p. 9844, Dec. 2022.

[26] F. Zhang, C. Luo, J. Xu, and Y. Luo, "An efficient deep learning model for automatic modulation recognition based on parameter estimation and transformation," *IEEE Commun. Lett.*, vol. 25, no. 10, pp. 3287–3290, Oct. 2021.

[27] T. J. O'Shea and N. West, "Radio machine learning dataset generation with GNU radio," in *Proc. GNU Radio Conf.*, 2016, pp. 1–6.

[28] X. Liu, D. Yang, and A. E. Gamal, "Deep neural network architectures for modulation classification," in *Proc. 51st Asilomar Conf. Signals, Syst., Comput.*, Oct. 2017, pp. 915–919.

[29] K. Tekbiyik, A. R. Ekti, A. Görçin, G. K. Kurt, and C. Keçeci, "Robust and fast automatic modulation classification with CNN under multipath fading channels," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–6.

[30] S. Lin, Y. Zeng, and Y. Gong, "Learning of time-frequency attention mechanism for automatic modulation recognition," *IEEE Wireless Commun. Lett.*, vol. 11, no. 4, pp. 707–711, Apr. 2022.

XUE WANG received the M.Sc. and Ph.D. degrees in information and communication engineering from Harbin Institute of Technology, in 2014 and 2018, respectively. She is currently a Lecturer with the School of Measurement and Control Technology and Communication Engineering, Harbin University of Science and Technology. Her research interests include wideband spectrum sensing techniques, broadband satellite communications, and satellite Internet of Things.

JIAQI WANG received the B.E. degree from the College of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan, China, in June 2019. She is currently pursuing the master's degree with the School of Measurement and Control Technology and Communication Engineering, Harbin University of Science and Technology. Her research interests include deep learning and pattern recognition.

XINMIAO LU received the M.Sc. and Ph.D. degrees in engineering from Harbin University of Science and Technology, in 2011 and 2017 respectively. He is currently an Associate Professor with the School of Measurement and Control Technology and Communication Engineering, Harbin University of Science and Technology. His research interests include wireless sensor networks and circuit fault diagnosis.