Wenhan Li[1], Hua Chen[1], Wei Liu[1], Jiangong Wang[1], and Gaoming Xu[1]

[1]Affiliation not available

June 07, 2024

A ConvTransNet Model Based on I/Q-Language Mutual Learning and Supervised Learning for Automatic Modulation Recognition

Wenhan Li, Hua Chen, *Senior Member, IEEE*, Wei Liu, *Senior Member, IEEE*, Jiangong Wang, and Gaoming Xu

***Abstract*—Automatic modulation recognition (AMR) is an important signal classification technology in cognitive radio. As AMR advances, an increasing number of artificial neural networks are being employed in the field to enhance its performance. In order to further improve its performance, a ConTransNet model based on I/Q-language mutual learning and supervised learning is proposed in this work. First, a ConTransNet model is introduced to handle modulation signals. The model consists of two branches: one is CNN, and the other is transformer. To facilitate information exchange between the two branches, an information interaction module is introduced, implemented with a bridge connection. To enhance the model's performance, a training algorithm called I/Q-language mutual learning and supervised learning is designed. This method utilizes mutual supervision between the output of one branch of the ConTransNet model and the output of a language feature extraction model, while the other branch adopts supervised learning. Finally, through experimental comparisons with five other algorithms (CE-FuFormer, ConvLSTMAE, DAE, FEA-T, and MCLDNN), the effectiveness of the proposed method is validated.**

***Index Terms*—Automatic modulation recognition, artificial neural network, ConTransNet, I/Q-language mutual learning.**

I. INTRODUCTION

With the rapid development of integrated air, space, land, and sea networks, there is a growing shortage of spectrum resources. Moreover, the actual utilization rate of some allocated authorized spectrum is excessively low. Cognitive radio [1−7] is considered a viable solution to address this shortage, with automatic modulation recognition (AMR) being one of its crucial technologies. It identifies the type of received wireless signals, determines the signal category in the current spectrum, and consequently understands the current spectrum occupancy state.

Traditional AMR methods can be broadly categorized into two types: likelihood based (LB) [8−10] and feature based (FB) [11−14]. For LB methods, they often require explicit assumptions about the signal model, while obtaining accurate prior knowledge of signal models in real-world environments can be challenging. Due to the diversity and complexity of signals in the environment, achieving completely accurate signal models is often difficult. This limitation may restrict the practicality of LB methods, especially in complex and dynamic communication environments. For FB methods, the limitation lies primarily in the requirement for manual feature design, which may involve domain-specific knowledge and expertise, making it necessary to tailor features to different signals and application scenarios. This introduces subjectivity, potentially leading to lack of robustness in the face of diverse signals and rapidly changing communication environments. Moreover, the choice of specific features can impact the accuracy of modulation recognition, requiring different feature designs for different types of signals and noise, thereby increasing the complexity of applications. Thus, while LB and FB methods may excel in certain contexts, they can face practical challenges when dealing with complex, varied signals in the real world, necessitating consideration of other more robust approaches.

Artificial neural network (ANN) models, with their excellent feature extraction capabilities, have been widely applied in the field of AMR. In the domain of AMR, ANN models mainly consist of three types: convolutional NN (CNN) models [15−21], recurrent NN (RNN) models [22−27], and transformer models [28−33]. These ANN architectures play an important role in the accurate identification and classification of modulation schemes within wireless communication signals. The CNN models excel in capturing spatial relationships and patterns within signal data, while RNN models are adept at processing sequential dependencies over time. The innovative transformer models, originally designed for natural language processing tasks, have been successfully adapted to AMR, showcasing their effectiveness in capturing long-range dependencies and enhancing overall recognition performance. In the realm of research on AMR utilizing CNN, a CNN model was applied to modulation recognition in [15] and the input to this model consists of in-phase and quadrature (I/Q) signals. Experimental results

This work was supported in part by the Natural Science Foundation of China under Grant 62071264; in part by the Science and Technology Innovation 2025 Major Project of Ningbo (2022Z186). *(Corresponding author: Gaoming Xu)*

Wenhan Li, Hua Chen, and Gaoming Xu are with the Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo, Zhejiang, 315211, China. (e-mail: wenhan_li@163.com; chenhua@nbu.edu.cn; xugaoming@nbu.edu.cn).

Wei Liu is with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, UK. (e-mail: wliu.eee@gmail.com).

Jiangong Wang is with the Faculty of China Guard Academy, Ningbo, Zhejiang, 315211, China. (e-mail: hw12xian@hotmail.com).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org

demonstrate that, compared to traditional methods, CNN shows improved performance in terms of recognition accuracy. Subsequently, two variants of CNN were introduced in [16], namely ResNet and DenseNet. These variants were tested on the RadioML2016.10b dataset, demonstrating good recognition results. In order to capture multi-scale features, a CNN model employing a multi-scale module was proposed in [17], and experimental results proved the effectiveness of the model. Differing from [15−17], the spectrogram data was employed in [18] as the input and achieved satisfactory results. Additionally, the constellation data was utilized in [21] as the input and likewise achieved good performance. In the realm of research on AMR based on RNN, a CLDNN model [22] was designed to recognize modulated signals. This model contains a long short-term memory (LSTM) layer and three convolutional layers. Subsequently, an LSTM de-noising auto-encoder model was proposed in [23], taking amplitude and phase data as its input. This model learns the features of modulated signals by simultaneously training an auto-encoder and a classifier. In contrast, a gated recurrent unit (GRU) model was applied in [25] to learn I/Q features and received good recognition results. In [27], a dual-path structure was designed, including both parallel ResNeXt and GRU. For transformer-based AMR, a transformer model was introduced in [28], and experimental results demonstrated the feasibility of using the transformer model for recognizing modulated signals. Subsequently, the transformer model in [29] utilized convolutional layers to generate embeddings, followed by employing self-attention layers to capture the relationships between these embeddings. In contrast to previous single-scale embeddings, a transformer model utilizing cross-scale embedding layers was proposed in [32], where the cross-scale embedding layers were employed to capture features at different scales.

In order to improve the performance of the model for AMR, a ConvTransNet model based on I/Q-language mutual learning and supervised learning is proposed in this work. When the label of a signal is known, language can be used to describe the signal. For example, for a signal labeled as QAM16, it can be described as *'This is a QAM16 signal'*. Based on this, a method is proposed to learn signal features using natural language supervision. The contributions of this paper are as follows.

(1) A ConvTransNet model is introduced, which adopts a dual-path structure comprising CNN and transformer. To integrate the local features extracted by CNN and the global features extracted by transformer, an information interaction module is introduced and integrated into the ConvTransNet model.

(2) A method for I/Q-language mutual learning is proposed. The core of this method lies in utilizing I/Q features and language features simultaneously as supervisory signals, thereby encouraging the model to learn meaningful features. By promoting interaction between I/Q features and language features, the network can effectively comprehend information about modulated signals.

(3) The proposed method is validated for its performance on three public datasets, RadioML2016.10a, RadioML2016.10b and RadioML2018.01A, demonstrating that the proposed method outperforms other ones on three datasets.

The remaining sections are arranged as follows: Section II provides a detailed description of the modulation recognition problem, Section III elaborates on the proposed method, Section IV conducts experiments with detailed discussions, and finally, conclusions are drawn in Section V.

II. Signal Model

The process of modulation recognition is shown in Fig. 1. In a wireless communication system, the wireless signal reaching the receiver end can be represented as follows

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

where $x(t)$ is the transmitted signal, $h(t)$ denotes the channel impulse response, and $n(t)$ is additive white Gaussian noise with a mean of 0 and a variance of σ^2. At the receiver, the received signal undergoes amplification, down-conversion, and analog-to-digital conversion to obtain a discrete signal $y_d(l)$, given by

$$y_d(l) = \mathrm{Re}\left[y_d(l)\right] + j\,\mathrm{Im}\left[y_d(l)\right], l = 0,1,\ldots,L-1, \tag{2}$$

where $\mathrm{Re}\left[y_d(l)\right]$ is the real part of $y_d(l)$ and $\mathrm{Im}\left[y_d(l)\right]$ the imaginary part. The I/Q data is then extracted, and represented in matrix form as follows

$$\mathbf{y}_d = \begin{bmatrix} y_{dI}(0) & y_{dI}(1) & \cdots & y_{dI}(L-1) \\ y_{dQ}(0) & y_{dQ}(1) & \cdots & y_{dQ}(L-1) \end{bmatrix}, \tag{3}$$

where $y_{dI}(\cdot) = \mathrm{Re}\left[y_d(\cdot)\right]$ and $y_{dQ}(\cdot) = \mathrm{Im}\left[y_d(\cdot)\right]$. The purpose of modulation recognition is to determine the modulation scheme of $y_d(l)$.

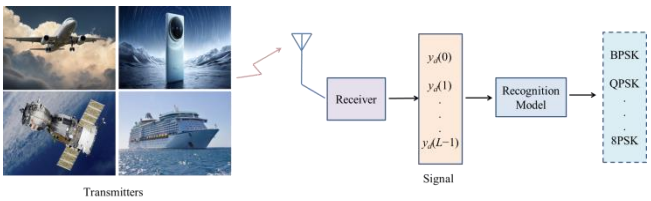

Fig. 1. The flow of modulation recognition.

III. Proposed Method

This section will first introduce the overall process of the proposed method in Fig. 2, followed by a detailed description of the subparts, including the ConvTransNet model, the language model, as well as two supervised learning methods: the I/Q-language mutual learning method and the supervised learning method.

The ConvTransNet model consists of two branches, one

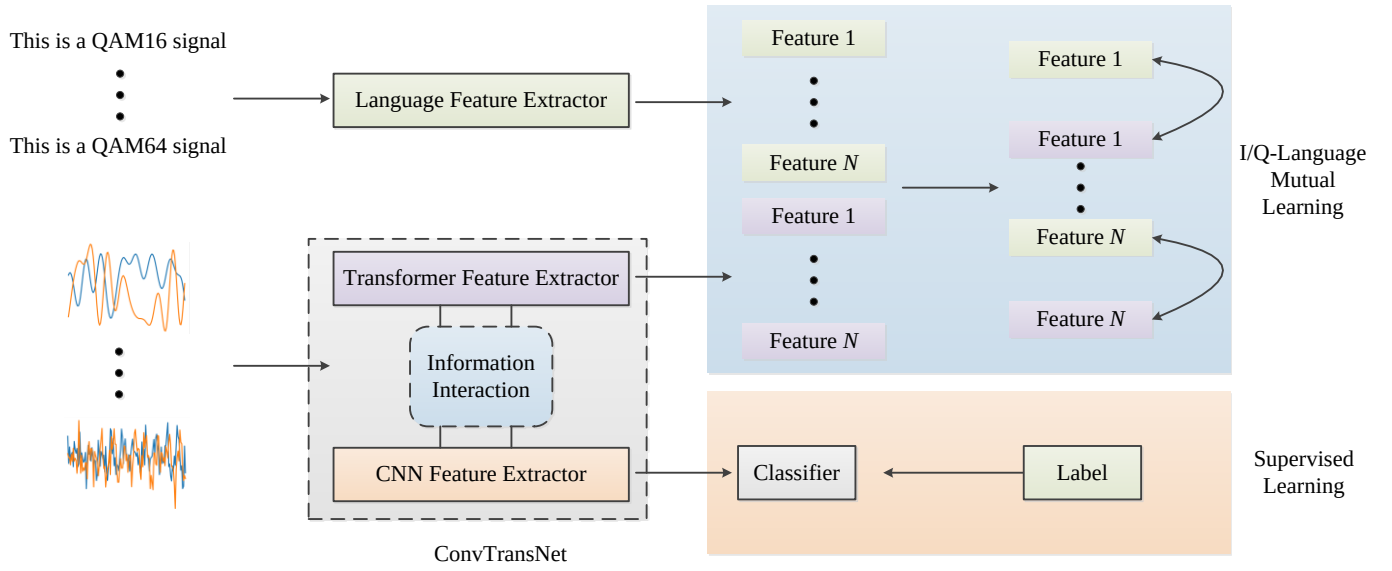

Fig. 2. The structure of the proposed method.

being the CNN architecture and the other being the transformer architecture. CNN is specifically employed to extract local features from I/Q signals, focusing on capturing details within local regions. Conversely, the transformer is adept at capturing global features, enabling a broader understanding of the entire input sequence. To ensure effective interaction between these complementary features, an information interaction module is introduced. The role of this module is to facilitate the fusion of local and global information, thereby enhancing the model's overall performance. The structure of ConvTransNet model is shown in Fig. 3.

In the CNN branch, every two convolutional layers form a residual structure, while in the transformer branch, each self-attention module consists of a self-attention layer and a fully connected layer, and these two parts constitute a residual structure. The local features extracted by the residual structure of the CNN branch can be represented as follows

$$F_{local}^{m} = \mathrm{Conv}_{3\times1}\left(\mathrm{Conv}_{3\times1}\left(S_{local}^{m-1}\right)\right) + S_{local}^{m-1}, \tag{4}$$

where $\mathrm{Conv}_{3\times1}(\cdot)$ represents a 3×1 convolutional kernel, and F_{local}^{m} is the output of the mth residual structure and S_{local}^{m-1} represents the input of the mth residual structure. The global features extracted by the residual structure of the transformer branch can be represented as

$$F_{global}^{m} = \mathrm{FC}\left(\mathrm{Attention}\left(S_{global}^{m-1}\right)\right) + S_{global}^{m-1}, \tag{5}$$

where $\mathrm{Attention}(\cdot)$ is a self-attention layer, F_2 is the input of the residual structure of the transformer branch, and $\mathrm{FC}(\cdot)$ denotes a fully connected layer. This information interaction module employs bridge connections, allowing features extracted by CNN to be passed to the transformer, while features extracted by the transformer can be passed back to CNN. The flow of information from transformer to CNN can be represented as follows

$$S_{local}^{m-1} = \mathrm{Conv}_{1\times1}\left(F_{global}^{m-1}\right) + F_{local}^{m-1}, \tag{6}$$

where $\mathrm{Conv}_{1\times1}(\cdot)$ is a 1×1 convolutional kernel. The flow of information from CNN to transformer can be represented by

$$S_{global}^{m-1} = \mathrm{Conv}_{1\times1}\left(F_{local}^{m-1}\right) + F_{global}^{m-1}, \tag{7}$$

where $\mathrm{Conv}_{1\times1}(\cdot)$ denotes a 1×1 convolutional layer. The process of feature extraction by the language model is illustrated in Fig. 4, consisting of two steps. Firstly, words in the text are mapped into a low-dimensional vector space to obtain embeddings for each word. Then, feature representation of the text is obtained through the output of the language model. The vocabulary is shown in Table I.

The key to the mutual learning method between I/Q signals and natural language lies in using them as mutual supervised signals. This method aims to leverage the interaction between I/Q signals and natural language, enabling the model to effectively learn features of modulated signals. Through this mutual supervision, it encourages the model to capture the underlying correlation between I/Q signals and natural language, thereby effectively understanding I/Q signals. Therefore, this method encourages the model to learn efficient feature representations. When the output of the language model serves as the supervised signal, this loss function of the ConvTransNet model can be expressed as:

$$Loss_{CT} = \frac{1}{N}\sum_{i=1}^{N}\left(F_{CT} - \mathrm{detach}(F_{language})\right)^2, \tag{8}$$

where F_{CT} denotes the output of the CNN branch or the transformer branch, $F_{language}$ is the output of the language model, and the use of $\mathrm{detach}(\cdot)$ indicates no gradient calculation. The loss function of the language model is given by:

$$Loss_{language} = \frac{1}{N}\sum_{i=1}^{N}\left(F_{language} - \mathrm{detach}(F_{CT})\right)^2. \tag{9}$$

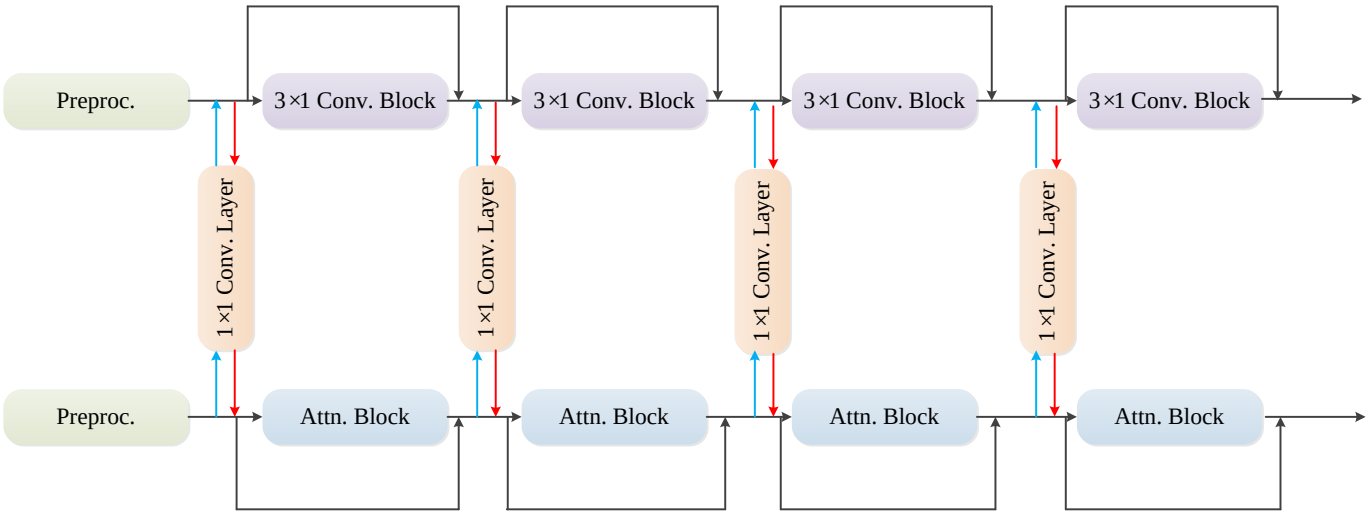

Fig. 3. The ConvTransNet model.

The loss function for I/Q-language mutual learning can be expressed as:

$$Loss_{I/Q-language} = Loss_{CT} + Loss_{language}. \quad (10)$$

In supervised learning, the loss function used is the cross-entropy loss function, which measures the difference between model predictions and true labels, guiding the model training process.

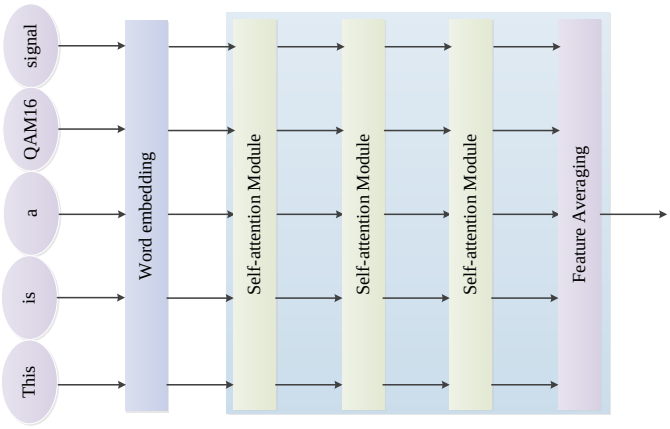

Fig. 4. The language model.

TABLE I
THE VOCABULARY

This	0	BPSK	8
is	1	CPFSK	9
signal	2	GFSK	10
a	3	PAM4	11
an	4	QAM16	12
8PSK	5	QAM64	13
AM-DSB	6	QPSK	14
AM-SSB	7	WBFM	15

IV. EXPERIMENTAL RESULTS AND ANALYSIS

To assess the performance of the proposed solution, experiments are conducted on three open-source datasets, RadioML2016.10a, RadioML2016.10b and RadioML2018.01A. Firstly, to validate the effectiveness of the proposed model structure, the proposed method is compared with CNN and transformer through experiments. Subsequently, to verify the role of I/Q-language mutual learning, experiments are conducted comparing the proposed with the ConvTransNet based on supervised learning. Then, to explore the robustness of the proposed method, confusion matrices are presented at different signal-to-noise ratios (SNRs). Finally, to validate its efficacy, comparisons are made with other modulation recognition solutions.

A. Experimental Dataset

The RadioML2016.10a dataset comprises 11 modulation types (16QAM, 64QAM, BPSK, QPSK, 8PSK, CPFSK, GFSK, 4PAM, AM-DSB, AM-SSB, and WBFM). Each type has 20 SNR conditions ranging from -18 dB to 20 dB, with a 2 dB interval. At each SNR level, there are 1000 samples for each modulation type. The RadioML2016.10b dataset also consists of 20 SNRs ranging from -18 dB to 20 dB with a 2 dB interval, but it lacks the AM-SSB type. Under each SNR level, there are 6000 samples for each modulation type. The RadioML2018.10A dataset comprises 24 types of modulated signals, namely OOK, 4ASK, 8ASK, BPSK, QPSK, 8PSK, 16PSK, 32PSK, 16APSK, 32APSK, 64APSK, 128APSK, 16QAM, 32QAM, 64QAM, 128QAM, 256QAM, AM-SSB-WC, AM-SSB-SC, AM-DSB-WC, AM-DSB-SC, FM, GMSK, and OQPSK. Each signal is evaluated across 26 different SNR levels ranging from -20 dB to 30 dB, with intervals of 2 dB. For each SNR, each signal comprises 4096 samples. At each SNR condition, we randomly select 50 samples from each signal type as the training set, and randomly choose 100 samples as the test set.

B. Experimental Platform and Settings of Hyper-parameters

The experimental software platform runs the Windows operating system, with Python programming language and PyTorch deep learning framework library. The hardware side includes an Intel i7-12700 CPU and a NVIDIA RTX 3080 GPU.

The hyper-parameters of the proposed method are shown in Table II. $3\times1\times16$ represents a convolutional kernel size of 3×1 with 16 kernels. Meanwhile, 16/16/16 denotes three fully connected layers generating queries, keys, and values, each with 16 neurons. 11/10/24 indicates the number of categories in RadioML2016.10a, RadioML2016.10b, and RadioML2018.01A.

TABLE II
THE MODEL STRUCTURE

CNN	Bridge Conn.	transformer
$3\times1\times16$	–	$3\times1\times16$
$3\times1\times16$, stride=2	–	$3\times1\times16$, stride=2
$\begin{bmatrix}3\times1\times16\\3\times1\times16\end{bmatrix}$	$1\times1\times16\leftarrow$ $\rightarrow1\times1\times16$	$\begin{bmatrix}16/16/16\\16\end{bmatrix}$
$\begin{bmatrix}3\times1\times16\\3\times1\times16\end{bmatrix}$	$1\times1\times16\leftarrow$ $\rightarrow1\times1\times16$	$\begin{bmatrix}16/16/16\\16\end{bmatrix}$
$\begin{bmatrix}3\times1\times16\\3\times1\times16\end{bmatrix}$	$1\times1\times16\leftarrow$ $\rightarrow1\times1\times16$	$\begin{bmatrix}16/16/16\\16\end{bmatrix}$
$\begin{bmatrix}3\times1\times16\\3\times1\times16\end{bmatrix}$	$1\times1\times16\leftarrow$ $\rightarrow1\times1\times16$	$\begin{bmatrix}16/16/16\\16\end{bmatrix}$
11/10/24	–	–

C. Experiment

To investigate the influence of the number of layers on ConvTransNet, the recognition rates of the proposed method are compared with 2, 4, 6, and 8 layers, respectively. The results show that the recognition rates of the proposed method vary with different number of layers. As the number of layers increases, the recognition rate also increases. However, the rate of increase is initially faster and then slows down. For example, the recognition rate with 4 layers is 2.2% higher than that with 2 layers, while the recognition rate with 8 layers is only 0.1% higher than that with 6 layers. Among the selected numbers of layers, the recognition rate is highest with 8 layers. Therefore, 8 layers are chose for further experiments.

Due to the presence of two branches in the proposed method, there are two ways to conduct both I/Q-language mutual learning and supervised learning. The first involves supervised learning on the CNN branch and I/Q-language mutual learning on the transformer branch, while the second involves supervised learning on the transformer branch and I/Q-language mutual learning on the CNN branch. To this end, the performance of models is compared by being trained using these two approaches, and the experimental results are shown in Table III. It can be seen that the performance of models trained using these two approaches is nearly identical. This could be because the information interaction module effectively interacts the features extracted from both branches.

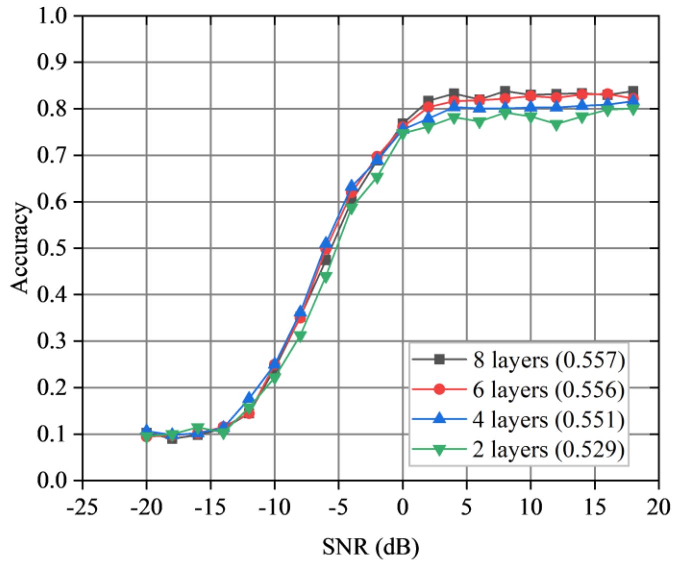

Fig. 5. Accuracy comparison with different number of layers.

To assess the effectiveness of the proposed model structure, it is compared with CNN and transformer models. Experimental results in Fig. 6 demonstrate that the proposed model outperforms both CNN and transformer models in terms of the average recognition rate. Specifically, the proposed model shows a 2.5% improvement over the CNN model and a 2.7% improvement over the transformer model. These results indicate a clear advantage of the proposed model on the RadioML2016.10a dataset. The superiority of the proposed model over the CNN model may stem from its integration of both local and global features. Unlike the CNN model, the proposed model emphasizes the integration of global features, enabling a more comprehensive understanding of the signals. Additionally, compared to the transformer model, the proposed model enhances feature diversity by integrating local features into global features. In summary, the proposed model exhibits a better performance on the RadioML2016.10a dataset by effectively integrating local and global features. These results also validate the rational of the model design.

TABLE III
OUTPUT RESULTS OF THE TWO BRANCHES

	-8 dB	-4 dB	0 dB	4 dB	8 dB	Avg.
1st	0.3526	0.6018	0.7682	0.8327	0.8382	0.557
2nd	0.3489	0.5997	0.7701	0.8352	0.8345	0.556

To demonstrate the effectiveness of I/Q-language mutual learning, a comparison is made between the proposed method and ConvTransNet using supervised learning, as shown in Table IV. It can be seen that the proposed method outperforms ConvTransNet using supervised learning in terms of

recognition performance. For example, at an 8 dB SNR, the recognition rate of the proposed method is 1.9% higher than that of ConvTransNet using supervised learning. The reason is, under the framework of supervised learning, I/Q-language mutual learning is integrated. I/Q-language mutual learning employs natural language to supervise I/Q signals. The combination of these two learning methods helps the model extract better features, thereby enhancing the model's recognition performance.

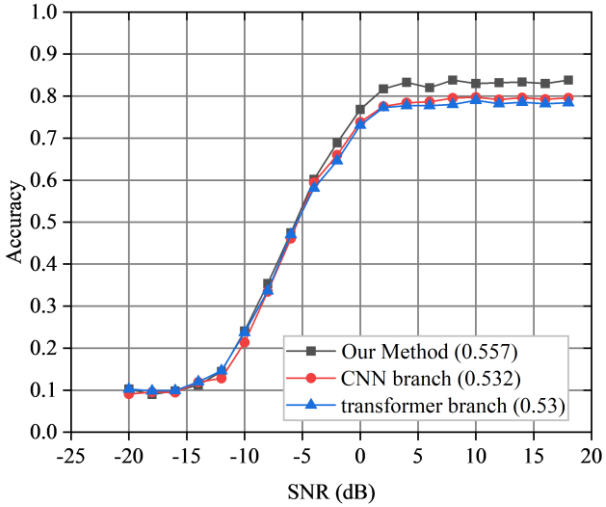

Fig. 6. Accuracy comparison of the proposed method and two branches.

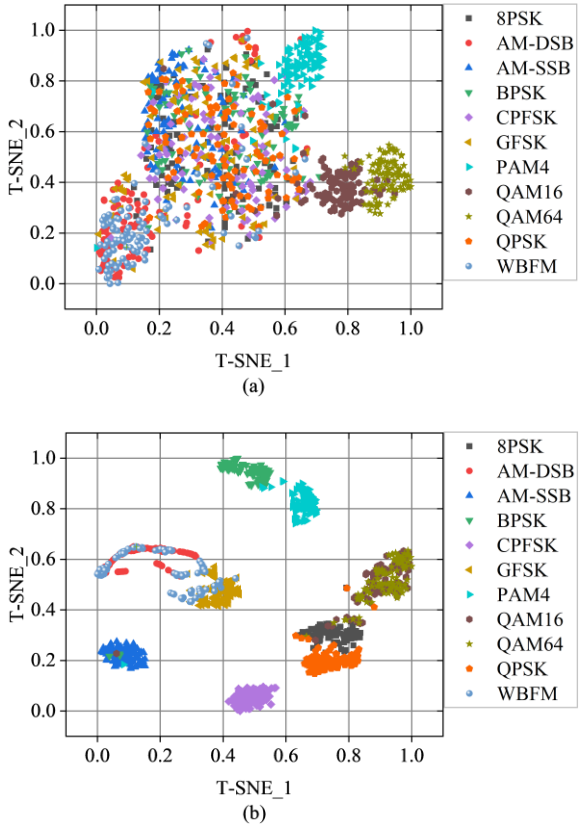

Fig. 7. Feature visualization on the RadioML2016.10a dataset: (a) -8 dB SNR, (b) 8 dB SNR.

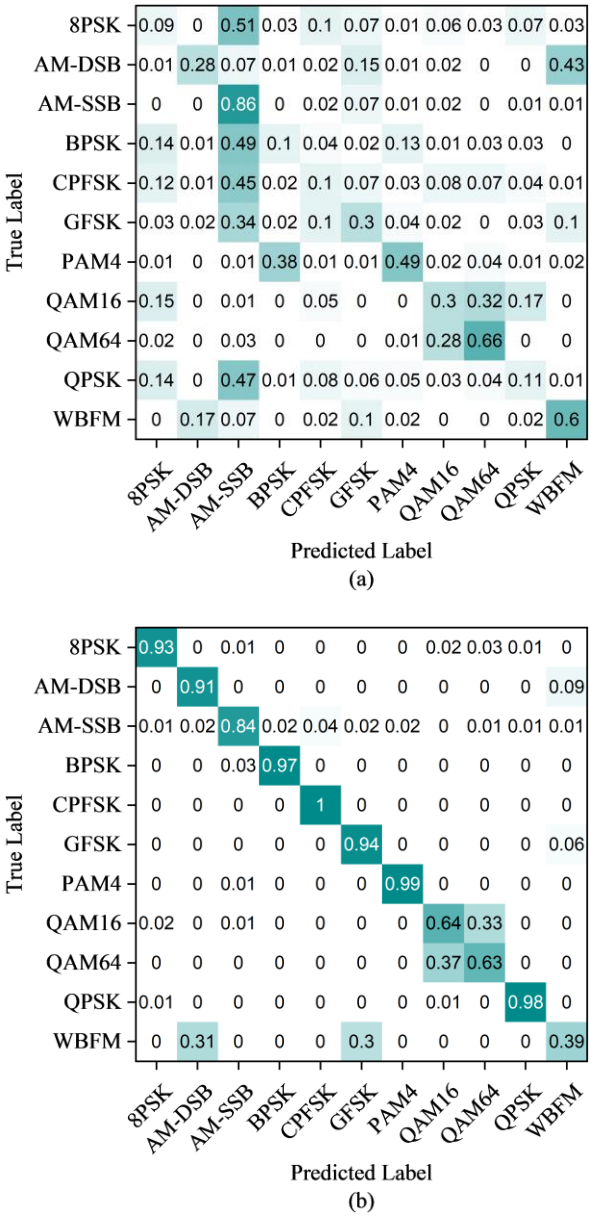

Fig. 8. The confusion matrices of SPDMFE on the RadioML2016.10a dataset: (a) -8 dB SNR, (b) 8 dB SNR.

To visually demonstrate the differences in signal characteristics at different SNRs, the t-distributed stochastic neighbor embedding (t-SNE) algorithm is employed to perform dimensionality reduction on signal features. The results are shown in Fig. 7. In Fig. 7(a), the outcome at an SNR of -8 dB is depicted. It can be observed that almost all signals exhibit overlapping phenomena, without clearly forming distinct clusters, which may be attributed to the relatively low SNR. Fig. 7(b) shows the results at an SNR of 8 dB, where most signals form separate clusters, including 8PSK, AM-SSB, BPSK, CPFSK, PAM4, and QPSK. However, the WBFM signal not only overlaps with the AM-DSB signal but also with the GFSK signal, possibly due to similarities between the WBFM and AM-DSB signal characteristics, as well as similarities with the GFSK signal characteristics. Additionally, there is an overlap between

QAM16 and QAM64, due to similarities in their signal features.

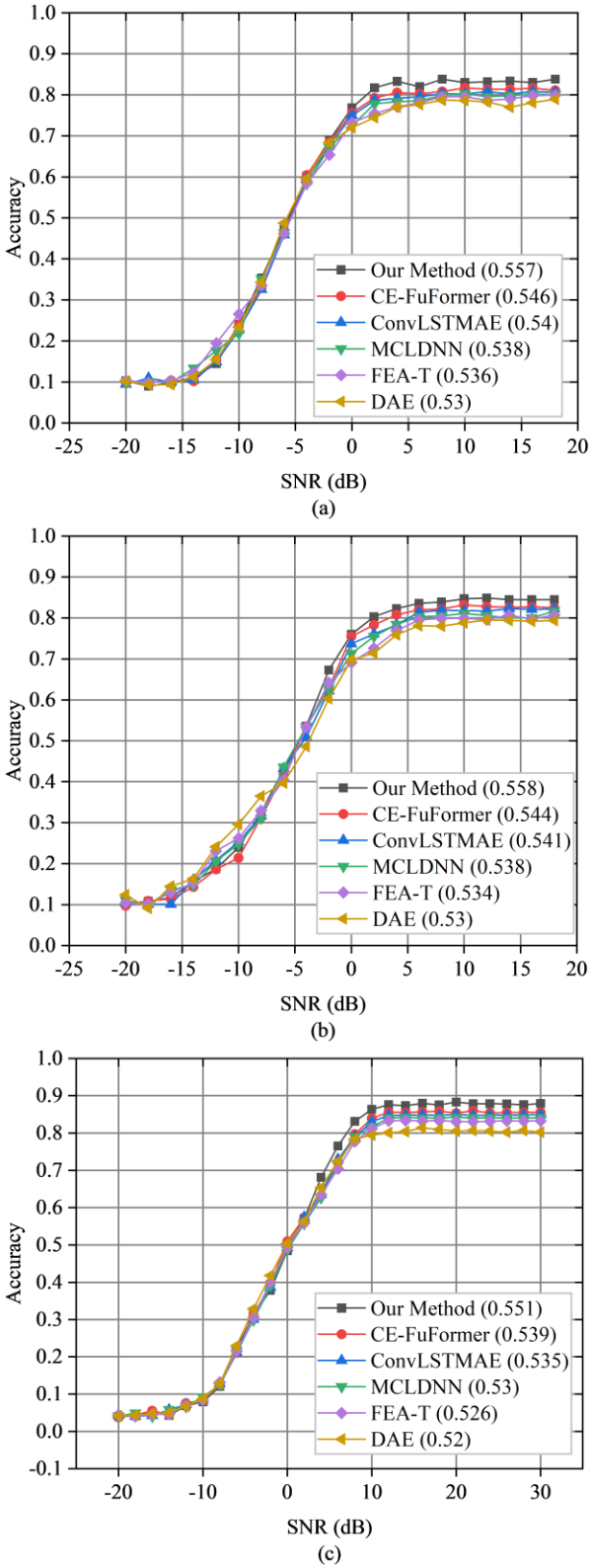

Fig. 9. Recognition accuracy of different models: (a) training on the RadioML2016.10a dataset, (b) training on the RadioML2016.10b dataset, (c) training on the RadioML2018.01A dataset.

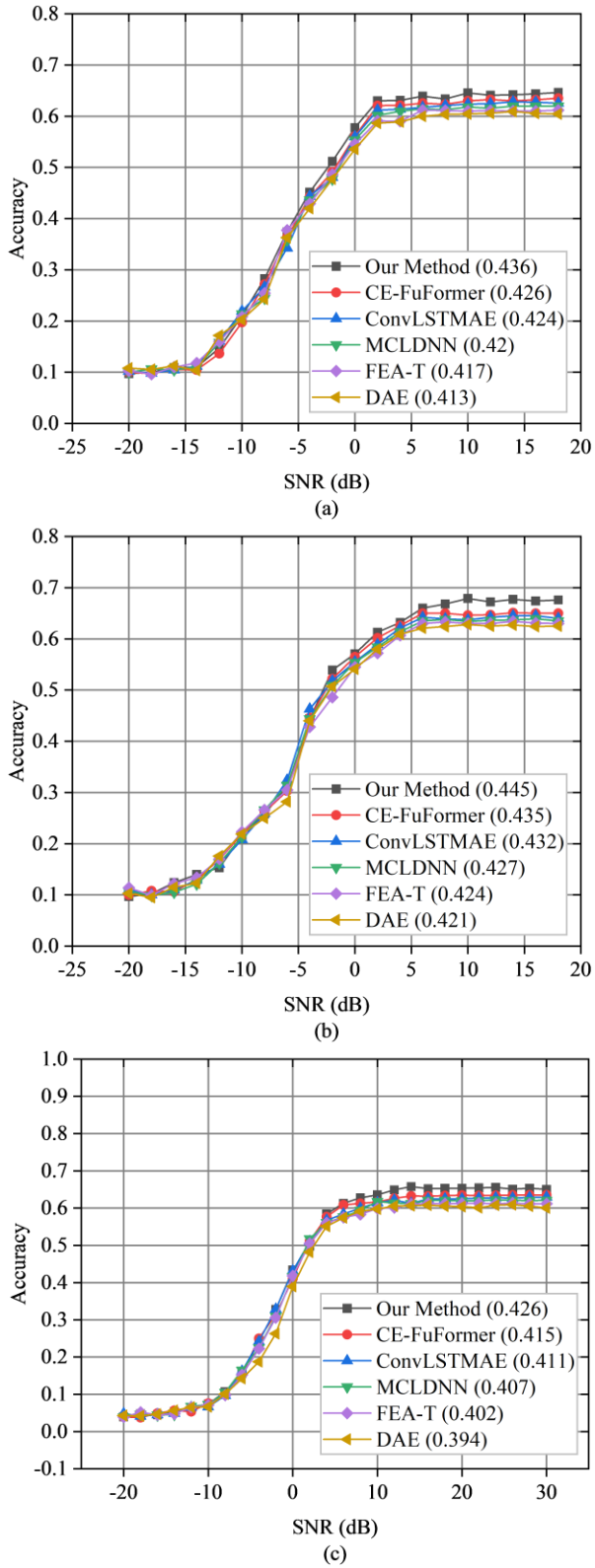

Fig. 10. Recognition accuracy of different models: (a) training on the RadioML2016.10a dataset, (b) training on the RadioML2016.10b dataset, (c) training on the RadioML2018.01A dataset.

To investigate the model's recognition performance under various SNRs, the confusion matrices are presented in Fig.

8(a) and 8(b). In Fig. 8(a), when the SNR is -8 dB, the model's recognition accuracy for most signals is below 60%, such as 8PSK, AM-DSB, BPSK, CPFSK, GFSK, PAM4, QAM16, and QPSK. This could be attributed to the lower SNR affecting the model's performance, resulting in inaccurate recognition of these signals. However, in Fig. 8(b), as the SNR increases to 8 dB, the recognition accuracy of most modulation schemes exceeds 80%, such as 8PSK, AM-DSB, AM-SSB, BPSK, CPFSK, GFSK, PAM4, and QPSK. However, under the same conditions, the recognition accuracy of QAM16, QAM64, and WBFM is relatively lower. Some QAM16 signals are incorrectly identified as QAM64 signals, and vice versa, as they both employ QAM modulation techniques, making it challenging for the model to distinguish between them. Additionally, some WBFM signals are misclassified as AM-DSB ones, while others are misclassified as GFSK signals. This could be due to the similarity between WBFM and AM-DSB or GFSK, making it difficult for the model to accurately differentiate them. These findings are crucial for understanding the model's recognition capabilities under different SNRs.

TABLE IV
ACCURACY RESULTS OF TWO LEARNING METHODS

	Mut. Learn.	No Mut. Learn.
0 dB	0.7682	0.7604
4 dB	0.8327	0.8173
8 dB	0.8382	0.8191
12 dB	0.8318	0.8145
16 dB	0.83	0.818
Avg.	0.557	0.543

To validate the performance of the proposed method, experimental comparisons are conducted with CE-FuFormer [32], ConvLSTMAE [35], DAE [23], FEA-T [33], and MCLDNN [34]. The results based on the RadioML2016.10a dataset, as shown in Fig. 9(a), indicate that the proposed method outperforms the others in terms of recognition accuracy. For instance, compared to the DAE algorithm, our method achieves an average improvement of 2.7% in recognition rate; compared to the CE-FuFormer algorithm, it improves by 1.1%. The advantage of our method over FEA-T and DAE may stem from the introduction of an information interaction module, which facilitates effective interaction between global and local features. Compared to CE-FuFormer, ConvLSTMAE, and MCLDNN, the advantage of our method lies in its simultaneous utilization of I/Q-language mutual learning and supervised learning during the training process. This joint training approach helps the model acquire better features, thereby enhancing its recognition capability. The experimental results in Fig. 9(b) and Fig. 9(c) are similar to those in Fig. 9(a), demonstrating that our method also outperforms others in recognition rates on the RadioML2016.10b and RadioML2018.01A datasets.

To further validate the performance of the proposed method, 10 samples are selected from each modulation scheme at each SNR to train the model. The experimental results are shown in Fig. 10, where it is evident that the proposed method results in a higher accuracy on both datasets than the other models. These comparative results demonstrate the effectiveness of the proposed method. In addition, the computational complexity of the proposed method is compared with others in terms of parameters, and floating point operations (FLOPs), and the results are listed in Table V, showing that the proposed method is the fastest with the least number of parameters.

TABLE V
COMPUTATIONAL COMPLEXITY OF DIFFERENT MODELS

Algorithm	Parameters(M)	FLOPs(M)
Our Method	0.014	1.8
CE-FuFormer	0.016	2.5
ConvLSTMAE	0.046	6
DAE	0.017	4.4
FEA-T	0.028	3.6
MCLDNN	0.023	4.1

V. CONCLUSION

In this paper, a ConTransNet model based on I/Q-language mutual learning and supervised learning is proposed for AMR. The effectiveness of the ConTransNet model structure and the feasibility of both I/Q-language mutual learning and supervised learning are validated on the RadioML2016.10a dataset. Furthermore, it is demonstrated that the proposed method outperforms five other methods (CE-FuFormer, ConvLSTMAE, DAE, FEA-T, and MCLDNN) on the RadioML2016.10a, RadioML2016.10b and RadioML2018.01A datasets. Compared to CE-FuFormer, ConvLSTMAE, and MCLDNN, the superiority of the proposed method lies in its utilization of both I/Q-language mutual learning and supervised learning for training the network, which enables it to learn better features and thus enhances the recognition performance of the network. Additionally, compared to FEA-T and DAE, the advantage of the proposed method lies in the incorporation of an information interaction module, facilitating the interaction between local features and global features. This feature interaction helps the model better understand I/Q signals and enhances the model's performance.

REFERENCES

[1] S. Haykin, ''Cognitive radio: Brain-empowered wireless communications,'' *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.
[2] Z. Qing and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 79–89, May 2007.
[3] A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proc. IEEE*, vol. 97, no. 5, pp. 894–914, May 2009.

[4] M. Liu, G. Liao, N. Zhao, H. Song, and F. Gong, "Data-driven deep learning for signal classification in industrial cognitive radio networks," *IEEE Trans. Ind. Informat.*, vol. 17, no. 5, pp. 3412–3421, May 2021.
[5] J. Ma, H. Liu, C. Peng, and T. Qiu, "Unauthorized broadcasting identification: A deep LSTM recurrent learning approach," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 9, pp. 5981–5983, Sep. 2020.
[6] T. M. Chiwewe and G. P. Hancke, "Fast convergence cooperative dynamic spectrum access for cognitive radio networks," *IEEE Trans. Ind. Informat.*, vol. 14, no. 8, pp. 3386–3394, Aug. 2018.
[7] T. M. Chiwewe, C. F. Mbuya, and G. P. Hancke, "Using cognitive radio for interference-resistant industrial wireless sensor networks: An overview," *IEEE Trans. Ind. Informat.*, vol. 11, no. 6, pp. 1466–1481, Dec. 2015.
[8] A. K. Nandi and E. E. Azzouz, "Algorithms for automatic modulation recognition of communication signals," *IEEE Trans. Commun.*, vol. 46, no. 4, pp. 431–436, Apr. 1998.
[9] W. Wei and J. M. Mendel, "Maximum-likelihood classification for digital amplitude-phase modulations," *IEEE Trans. Commun.*, vol. 48, no. 2, pp. 189–193, Feb. 2000.
[10] A. Polydoros and K. Kim, "On the detection and classification of quadrature digital modulations in broad-band noise," *IEEE Trans. Commun.*, vol. 38, no. 8, pp. 1199–1211, Aug. 1990.
[11] H.-C. Wu, M. Saquib, and Z. Yun, "Novel automatic modulation classification using cumulant features for communications via multipath channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 3098–3105, Aug. 2008.
[12] A. Swami and B. M. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 416–429, Mar. 2000.
[13] M. W. Aslam, Z. Zhu, and A. K. Nandi, "Automatic modulation classification using combination of genetic programming and KNN," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2742–2750, Aug. 2012.
[14] V. D. Orlic and M. L. Dukic, "Automatic modulation classification algorithm using higher-order cumulants under real-world channel conditions," *IEEE Commun. Lett.*, vol. 13, no. 12, pp. 917–919, Dec. 2009.
[15] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *Proc. Int. Conf. Eng. Appl. Neural Netw.(EANN)*, 2016, pp. 213–226.
[16] 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.
[17] H. Zhang, F. Zhou, Q. Wu, W. Wu and R. Q. Hu, "A Novel Automatic Modulation Classification Scheme Based on Multi-scale Networks," *IEEE Trans. Cognit. Commun. Netw.*, vol. 8, no. 1, pp. 97–110, Mar. 2022.
[18] 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.
[19] 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.
[20] 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.
[21] Y. Wang, M. Liu, J. Yang and G. Gui, "Data-driven Deep Learning for Automatic Modulation Recognition in Cognitive Radios," *IEEE Trans.Veh. Technol.*, vol. 68, no. 4, pp. 4074–4077, Apr. 2019.
[22] N. E. West and T. O'Shea, "Deep architectures for modulation recognition," in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw. (DySPAN)*, 2017, pp. 1–6.
[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] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin, "Deep learning models for wireless signal classification with distributed low-cost spectrum sensors," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 3, pp. 433–445, Sep. 2018.
[25] D. Hong, Z. Zhang, and X. Xu, "Automatic modulation classification using recurrent neural networks," in *Proc. 3rd IEEE Int. Conf. Comput. Commun.(ICCC)*, 2017, pp. 695–700.
[26] S. Huang et al., "Automatic modulation classification using gated recurrent residual network," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 7795–7807, Aug. 2020.
[27] L. Li, Y. Zhu and Z. Zhu, "Automatic modulation classification using ResNeXt-GRU with deep feature fusion," *IEEE Trans. Instru. Meas.*, vol. 72, pp. 1-10, 2023.
[28] S. Hamidi-Rad and S. Jain, "MCformer: A transformer based deep neural network for automatic modulation classification," *in Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2021, pp. 1–6.
[29] J. Cai, F. Gan, X. Cao, and W. Liu, "Signal modulation classification based on the transformer network," *IEEE Trans. Cogn. Commun. Netw.*, vol. 8, no. 3, pp. 1348–1357, Sep. 2022.
[30] W. Kong, Q. Yang, X. Jiao, Y. Niu, and G. Ji, "A transformer-based CTDNN structure for automatic modulation recognition," in *Proc. 7th Int. Conf. Comput. Commun. (ICCC)*, Dec. 2021, pp. 159–163.
[31] J. Li et al., "Automatic Modulation Recognition of Underwater Acoustic Signals Using a Two-Stream Transformer," *IEEE Internet Things J.*, doi: 10.1109/JIOT.2024.3367852.
[32] C. Zhao, J. Chen, X. Huang and Z. Wu, "A Cross-Scale Embedding Based Fusion Transformer for Automatic Modulation Recognition," *IEEE Commun. Lett.*, vol. 28, no. 1, pp. 68–72, Jan. 2024.
[33] Y. Chen, B. Dong, C. Liu, W. Xiong, and S. Li, "Abandon Locality: Frame-wise Embedding Aided Transformer for Automatic Modulation Recognition," *IEEE Commun. Lett.*, vol. 27, no. 1, pp. 327–331, Jan. 2023.
[34] J. Xu, C. Luo, G. Parr, and Y. Luo, "A Spatiotemporal Multi-channel Learning Framework for Automatic Modulation Recognition," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1629–1632, Oct. 2020.
[35] S. Yunhao, X. Hua, J. Lei, and Q. Zisen, "ConvLSTMAE: A Spatiotemporal Parallel Autoencoders for Automatic Modulation Classification," *IEEE Commun. Lett.*, vol. 26, no. 8, pp. 1804–1808, Aug. 2022.

Wenhan Li received the B.S. degree in internet of things engineering and the M.S. degree in information and communication engineering from Heilongjiang University, Harbin, China, in 2018 and 2021, respectively.

He is currently pursuing the Ph.D. degree with Ningbo University, Ningbo, China. His research interests include signal processing and deep learning.

Hua Chen (Senior Member, IEEE) received the M.Eng. degree and Ph.D. degree in Information and Communication Engineering from Tianjin University, Tianjin, China, in 2013 and 2017, respectively.

He is now as an Associate Professor in Faculty of Electrical Engineering and Computer Science, Ningbo University, China. His research interests include array signal processing, MIMO radar. He is currently an Associate Editor for Circuits, Systems, and Signal Processing.

Wei Liu (S'01-M'04-SM'10) received his BSc and LLB. degrees from Peking University, China, in 1996 and 1997, respectively, MPhil from the University of Hong Kong in 2001, and PhD from University of Southampton, UK, in 2003. He then worked as a postdoc first at Southampton and later at Imperial College London. In September 2005, he joined the Department of Electronic and Electrical Engineering, University of Sheffield, UK, first as a Lecturer and then a Senior Lecturer. Since September 2023, he has been a Reader at the School of Electronic Engineering and Computer Science, Queen Mary University of London, UK. He has published 390+ journal and conference papers, five book chapters, and two research monographs titled "Wideband Beamforming: Concepts and Techniques" (Wiley, March 2010) and "Low-Cost Smart Antennas" (Wiley, March 2019), respectively. His research interests cover a wide range of topics in signal processing, with a focus on sensor array signal processing and its various applications, such as robotics and autonomous systems, human computer interface, radar, sonar, and wireless communications.

He is a member of the Digital Signal Processing Technical Committee of the IEEE Circuits and Systems Society (Chair from May 2022) and the Sensor Array and Multichannel Signal Processing Technical Committee of the IEEE Signal Processing Society (Chair for 2021-2022), and an IEEE Distinguished Lecturer for the Aerospace and Electronic Systems Society (2023-2024). He also acted as an associate editor for IEEE Trans. on Signal Processing, IEEE Access, and Journal of the Franklin Institute (2021-2023), and currently he is an Executive Associate Editor-in-Chief of the Frontiers of Information Technology and Electronic Engineering.

Jiangong Wang was born in Heilongjiang, China, in 1982. He received the Ph.D. degree in Wireless communication from PLA University of Science and Technology.

He is currently serving as a lecturer at the China Coast Guard Academy. His research interests include blind signal processing and image object detection algorithms.

Gaoming Xu received the B.S. degree in Electronic and Information Engineering from Naval Aeronautical Engineering Academy, Qingdao, China, in 2007, and the M.S. and Ph.D. degree in Faculty of Electrical Engineering and Computer Science, Ningbo University (NBU), Ningbo, China, in 2010 and 2015.

He is currently working as an Associate Professor with Faculty of Electrical Engineering and Computer Science NBU. His main research interests are in the area of wireless communications, with a focus on signal processing, high efficiency RF power amplifiers design, analog and digital predistortion and nonlinear modeling.