IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 20 February 2024, accepted 22 April 2024, date of publication 25 April 2024, date of current version 2 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3393481

RESEARCH ARTICLE

Data-Transform Multi-Channel Hybrid Deep Learning for Automatic Modulation Recognition

MENG QI, NIANFENG SHI, GUOQIANG WANG, AND HONGXIANG SHAO

Luoyang Institute of Science and Technology, Luoyang 471000, China
Henan Province Engineering Research Center of Industrial Intelligent Vision, Luoyang 471023, China
Henan Key Laboratory of Green Building Materials Manufacturing and Intelligent Equipment, Luoyang 471023, China

Corresponding author: Meng Qi (iemengq@lit.edu.cn)

ABSTRACT Automatic modulation recognition (AMR) is an essential topic of cognitive radio, which is of great significance for the analysis of wireless signals and is one of the current research hotspots. Traditional AMR approaches predominantly utilize raw in-phase/quadrature symbols (I/Q), amplitude/phase (A/P), or pre-processed data (e.g., high-order cumulates, spectrum images, or constellation diagrams) as inputs for the recognition model. However, it is difficult to achieve superior performance with only a single type of data as input. This paper proposes a novel multi-channel hybrid learning framework that integrates convolutional layers, Long Short-Term Memory (LSTM) layers, fully connected layers and classification layers. The model is built for modeling spatial-temporal correlations from four signal cues (including I/Q signals, A/P signals, I, and Q signals), which aims to explore various differences and leverage the complements from multiple data-form. Two functions employed during the data conversion process further enhance the non-linear representational capacity of the model, thereby boosting the recognition accuracy of the model. Experimental results demonstrate that the proposed framework effectively addresses the classification challenges of QAM16 and QAM64. For the RAML2016A dataset, our model achieves an impressive recognition accuracy of 95% at an SNR of 0 dB. Extensive experiments indicate that the proposed framework outperforms other current networks in terms of recognition accuracy.

INDEX TERMS Automatic modulation recognition (AMR), data-driven, hybrid learning, recognition accuracy.

I. INTRODUCTION

Automatic Modulation Recognition (AMR) enables receivers to automatically detect the modulation scheme of signals in non-cooperative communication wireless communication systems and has a wide range of applications in the fields of dynamic spectrum access (DSA), cognitive radio (CR), and signal surveillance (SS). In DSA, knowledge of the presence of a primary user (PUs) is a prerequisite for CR users to achieve wireless access without causing harmful interference. Considering the complex channel environment and other potential radio emitters, focusing only on the occupancy of the frequency band of interest is not sufficient to identify the PU's signal from nearby radio interference. Automatic modulation classification distinguishes the modulation type of the received radio signals and can be used as a stair to understand the type of communication scheme and the presence or absence of the target PU. Due to factors such as noise, multipath fading, frequency selectivity, and time-varying channels in real-world environments, AMR is a significantly challenging task. Traditional AMR methods are categorized into two types: likelihood theory-based AMR (LB-AMR) and feature-based AMR (FB-AMR) [1]. The LB-AMR is essentially a composite hypothesis testing method. It uses the probability density function of a random signal to establish a hypothesis, thus determining the cost function, obtaining a test statistic, and then comparing it with an appropriate threshold to form a verdict criterion.

The associate editor coordinating the review of this manuscript and approving it for publication was Mauro Fadda.

The FB-AMR originates from the classical pattern recognition theory, which is essentially a mapping relationship and essentially maps the original signal space to the feature space and then to the target space. However, LB-AMR methods have high computational complexity, and the performance of FB-AMR depends on the rationality and completeness of the manually crafted feature space.

Deep learning-based AMR (DL-AMR), which uses an end-to-end structure to automatically extract features, has achieved higher recognition accuracy with reasonable computational complexity, achieving a revolutionary breakthrough. Researchers have recently proposed various DL-AMR methods that outperform traditional LB-AMR and FB-AMR. Reference [1] provides a detailed analysis of AMR research in SISO and MIMO communication systems, categorizing deep learning frameworks into two types based on input data: raw I/Q data and pre-processed data, and presenting the recognition performance of various neural network structures. Reference [2] proposes a hybrid deep learning framework that integrates 1D convolution, 2D convolution, and LSTM layers, which achieves promising performance. Reference [3] proposes an automatic modulation classification model that combines the residual neural network (ResNet) and the long short-term memory network (LSTM), achieving 92% classification accuracy on the RML2016B dataset at 18 dB SNR. Reference [4] presents a multi-scale network for AMR and proposes a new loss function combining the center loss and cross entropy loss. Reference [5] proposes a multi-scale convolutional network model based on I/Q sequences and five statistical features called MSNet-SF. Reference [6] proposes a hybrid neural network based on CNN and GRU using multiple statistical features as input data. Reference [7] proposes a feed-forward attention neural network based on ResNet and LSTM called RLADNN. Reference [8] suggests a CNN-LSTM network based on signal periodic features called the Intra-InterNet network (IIN-Net). Reference [9] presents a spectral CNN model based on a time-frequency attention mechanism for AMR called TFA-SCNN. Reference [10] proposes a AMR model based on CNN and spatial self-attention mechanisms.

The above studies utilize raw in-phase/quadrature data or original I/Q data combined with statistical features as inputs. Additionally, many researchers have proposed using amplitude and phase as inputs for recognition models. Reference [11] proposes an auto-encoder recognition model that uses standardized amplitude and phase data as inputs. Its encoder consists of a 2-layer LSTM that transforms inputs into hidden state vectors, and its decoder is a shared dense layer. The model employs mean squared error as the reconstruction loss and classification cross-entropy as the classification loss, combining both as the loss function. Reference [15] presents an AMC model based on a 2-layer LSTM structure, with L2-normalized amplitude data and normalized phase data as inputs. Experiments show that even a simple LSTM structure can achieve satisfactory recognition accuracy using amplitude and phase instead of IQ samples. Reference [16] notes that models using A/P as input data significantly outperform those using I/Q when SNR is high, while the results are reversed when SNR is low. Therefore, A/P data and I/Q data complement each other well. Consequently, a multi-task deep neural network (MLDNN) fusing A/P and I/Q is proposed. The MLDNN has a novel backbone, which is comprised of CNN modules, bidirectional gated recurrent unit modules (BiGRU), and single-step attention fusion modules. Unlike the traditional one, which only uses the output of the last step of the RNN, the MLDNN can fully utilize the output information of all BiGRU layers. To fully extract features from A/P and I/Q data, [17] proposes a dual-stream fusion network structure with two channels, each inputting A/P or I/Q data and using a cascade structure of 3-layer CNN and 2-layer LSTM. The outputs of both channels are fused through a fully connected layer, and the classification results are generated through a Softmax classification layer. Reference [18] introduces a complex CNN structure based on residual attention. The input to this structure is A/P and I/Q data, cascaded with five residual attention-based convolutional layers after preprocessing the convolutional module, followed by a gap layer, a fully connected layer, and a softmax classification layer. Although these models achieve commendable recognition accuracy, they exhibit high time and space complexity due to their overcomplexity.

Inspired by the above studies, we draw the following conclusions. First, for AMC, the simple CNN structure is inferior to LSTM, while the well-designed CNN-LSTM structure outperforms the CNN and LSTM structures. Second, in terms of input data formats, A/P is superior to I/Q, whereas the combination of I/Q and A/P with a well-designed network structure may yield even better performance. Therefore, this paper proposes a novel multi-channel neural network structure that combines CNN and LSTM, denoted as DMHNN. The convolutional layers can fully explore features in adjacent spatial dimensions of the data, while LSTM can effectively extract temporal features from sequential data. The model includes a data converter that is capable of transforming raw I/Q signals into A/P signals. The two functions used in the data conversion process further increase the nonlinear characterization capability of the whole model, which in turn improves the recognition accuracy of the model.

The arrangement of the other sections in this paper is as follows. Section II contains a detailed description of the proposed multi-channel neural network structure DMHNN, including the data converter, multi-channel CNN layers, dual-layer LSTM, and fully connected layers, along with the parameter settings for DMHNN. Section III briefly introduces the experimental environment, such as the dataset, benchmark algorithms, and software and hardware environments. Section IV details the recognition accuracy and confusion matrix of DMHNN and other models, conducts ablation

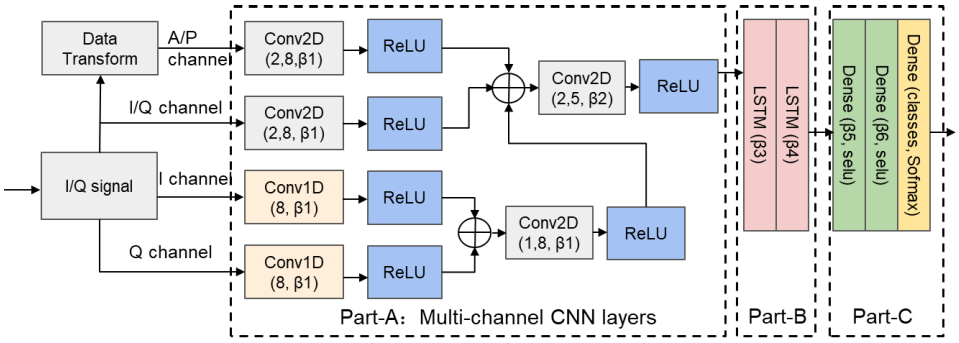


FIGURE 1. Structure of DMHNN.

experiments and cross-dataset validation and compares the time and space complexity of the models.

II. DATA-TRANSFORM MULTI-CHANNEL HYBRID DEEP LEARNING

The deep learning framework proposed in this paper is called the data-driven Multi-channel Hybrid Neural Network (DMHNN), as shown in Fig. 1. It contains a data converter that transforms raw in-phase/quadrature signals into Amplitude/Phase signals, and the subsequent network layers are all driven by either the I/Q signal or the A/P signal. DMHNN mainly consists of three parts: the CNN segment (Part-A), the RNN segment (Part-B), and the DNN segment (Part-C). Part-A consists of two channels, one channel is driven by I/Q signal and the other channel is driven by A/P signal. Part-B consists of two LSTM layers. Part-C consists of 2 layers of fully connected network and one layer of Softmax classification layer.

A. DATA CONVERTER

This paper concentrates on single-input single-output communication systems, where one antenna is used for the transmitter and another for the receiver. For the receiver, it is assumed that the low-pass equivalent of the bandpass signal is a complex signal $x_l(t) = x_i(t) + x_q(t)$, where the real part $x_i(t)$ is referred to as the I signal and the imaginary part as the Q signal $x_q(t)$. I/Q signals are the input data for DMHNN.

The data converter converts I/Q signals in Cartesian coordinates into A/P signals in polar coordinates (r, θ), it can be expressed by

$$r(t) = \sqrt{x_i{}^2(t) + x_q{}^2(t)}, \tag{1}$$

$$\theta(t) = \arctan(\frac{x_q(t)}{x_i(t)}), \tag{2}$$

According to the above equation, $r(t)$ and $\theta(t)$ are obtained from the I signal and Q signal as independent variables. Conversely, the I signal and Q signal can also be obtained from $r(t)$ and $\theta(t)$. Hence, the two representations are equivalent. It is worth mentioning that although the data converter module is within the DMHNN model, it has no learnable parameters and will be set as non-trainable and prohibited from participating during the process of backpropagation. Moreover, we recommend that it be placed outside the DMHNN model when training the model and then inside the model when deploying it. This approach not only saves training time but also maintains the model's traditional form, i.e., using only traditional I/Q signals as input data.

B. PART-A: MULTI-CHANNEL CNN LAYERS

CNNs exploit convolutional kernels to explore features in the spatial dimensions of data fully. Part A primarily uses convolutional operations to extract spatial features adjacent to each other between the data, consisting of four channels: Channel-A, Channel-B, Channel-C, and Channel-D. Among them, the inputs of Channel-A and Channel-B are I/Q signals and A/P signals, respectively. They used two-dimensional convolutional operations to extract the spatial features inside two types, respectively, and their output feature maps are denoted as MapA and MapB. The inputs of Channel-C and Channel-D are I and Q signals, respectively, and a 1D convolutional operation is used to extract unique spatial features of the I and Q signals. After completing 1D convolution, Channel-C and Channel-D are concatenated into a 2-dimensional vector, and then 2D convolution is performed to output the feature map, denoted as MapCD. This structure is conducive to fully extracting features between I and Q signals. Finally, the three channels, MapA, MapB, and MapCD, are combined and concatenated to form a new 2D vector, and then 2D convolution is performed, and the output feature map is recorded as Map4. At this point, all channels are aggregated. All convolutional modules are equipped with ReLU activation functions.

C. PART-B: TWO-LSTM LAYERS

The advantage of RNNs lies in their ability to exploit the chain-like connections between nodes to extract temporal features from sequence data. LSTM is a classical RNN model that can effectively capture temporal characteristics of time-series data, as shown in Fig. 2. Each LSTM unit consists of forget gates, input gates, and output gates. The formulae for

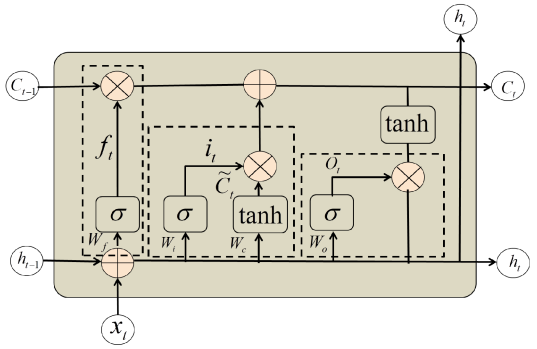


FIGURE 2. LSTM unit structure.

the three gates are given by

$$f^t = \sigma(W_f \cdot [x^t, h^{t-1}] + b_f) \tag{3}$$
$$i^t = \sigma(W_i \cdot [x^t, h^{t-1}] + b_i) \tag{4}$$
$$o^t = \sigma(W_o \cdot [x^t, h^{t-1}] + b_o) \tag{5}$$

where $\sigma()$ represents the sigmoid activation function, W represents the weight, b represents the bias, x^t represents the input at moment t, h^{t-1} represents the output at moment t-1. Using f^t and i^t updating the memory cells C^{t-1} to obtain C^t, the specific formula for the output h^t are given by

$$\tilde{C}^t = \tanh(W_c \cdot [x^t, h^{t-1}] + b_c) \tag{6}$$
$$C^t = f^t \cdot C^{t-1} + i^t \cdot \tilde{C}^t \tag{7}$$
$$h^t = \tanh(C^t) \cdot o^t \tag{8}$$

D. PART-C: FULLY-CONNECTED LAYERS

Fully connected DNNs can use dense connections to extract individual features from data, typically employed in the classification layer. Part-C consists of a 3-layer fully connected network, with 10 and 11 nodes in the final layer for RML2016.10a and RML2016.10b, respectively, depending on the number of modulation types. To better characterize the nonlinear characteristics of the target, the first two fully connected layers employ a variant of ReLU called the SELU function, given as

$$SELU(x) = \lambda \begin{cases} x & x > 0 \\ \alpha e^x - \alpha & x \le 0 \end{cases} \tag{9}$$

To avoid overfitting, dropout layers have been added to the first two fully linked layers. When the neural network is trained, a random subset of neurons is chosen, and during that iteration, they are not allowed to take part in either forward inference or backward propagation. In this paper, we set the dropout to 0.2, i.e., 20% of neurons are prohibited at each iteration.

The activation function for the third fully connected layer is Softmax, given as

$$Soft\ \max(x) = \frac{e^{x_i}}{\sum_i e^{x_i}}. \tag{10}$$

Categorical cross-entropy is used as the loss function, given as

$$Loss = -\sum_{k=1}^{K} y_i \log \hat{y}_i, \tag{11}$$

where y_i represents the true label value, which is in the range 0,1, when y_i equals 1, it indicates that the current sample belongs to the ith class. $\hat{y}_i$ is the ith component of the output value of the Softmax classification layer of the model, which ranges from 0 to 1, representing the probability that the model predicts that the current sample belongs to the ith class.

E. PARAMETER SETTINGS

Fig. 1 illustrates the number of convolutional kernels in the CNN layers, the number of units in the LSTM layers, and the number of nodes in the FC layers. The number of convolutional kernels is β_1=25, except for the last 2D convolutional module, which has β_2=50, and the number of cells in the two LSTM layers are β_3=100 and β_4=64, respectively, and the number of junctions in the first two fully-connected layers are β_5=64 and β_6=64, respectively. In ablation experiments, we increase or decrease the value of β to compare model performance across various sizes and identify the model with the best cost-performance ratio.

We set the kernel dimensions of the two 1D convolutional modules and the 2D convolution modules that follow them to (1,8). The convolutional kernels for the two 2D convolutional modules are set to (2,8), and the final 2D convolutional kernels are set to (2,5). For all the convolution modules, the default step size is used, i.e., (1, 1) for 2D convolution and 1 for 1D convolution. Padding is used in all convolutional modules to maintain the output size. Additionally, a glorot uniform was used to initialize the weight matrix of the convolution modules. The model employs the Adam optimizer, which is one of the most widely used optimizers.

III. DATA-TRANSFORM MULTI-CHANNEL HYBRID DEEP LEARNING

To verify the effectiveness of the DMHNN architecture, we trained and tested it on various datasets and compared it with the current state-of-the-art AMR network.

A. DATASETS

The experiments in this paper used two open-source datasets: RadioML2016.10a and RadioML2016.10b. RadioML2016.10a is a synthetic dataset developed by the Institute of Radio Communications of the Italian National Research Council for wireless signal modulation recognition. It contains 11 modulations (8 digital and 3 analogue) over 20 signal samples with signal-to-noise ratios (SNRs) that vary by 2 dB, from -20 dB to 18 dB, with noise signals sourced from additive white Gaussian noise (AWGN). For a total of 220,000 signal samples, each modulation type includes 1,000 signal samples with different SNRs. RadioML2016.10b

TABLE 1. Division of the data set.

Dataset	Modulation Schemes	SNR(dB)	Number of samples				Total Sample Size
			total number	training set	validation set	test set	
RML2016.10a	11 classes	20 types (-20dB∼18dB)	1000	600	200	200	220000
RML2016.10b	10 classes		6000	3600	1200	1200	1200000

TABLE 2. Recognition accuracy of common recognition models at 0 dB SNR on RML2016.10a dataset.

Moduation	CGDNet[13]	MCNet[14]	PET-CGDNN[12]	MCLDNN[2]	LSTMDAE[11]	RLADNN[7]	DMHNN
8PSK	80.00%	78.00%	82.00%	94.00%	77.00%	94.00%	94.00%
AM-DSB	98.00%	90.00%	66.00%	91.00%	93.00%	62.00%	97.00%
AM-SSB	94.00%	93.00%	90.00%	94.50%	93.00%	94.00%	94.00%
BPSK	99.00%	96.00%	100%	99.50%	95.00%	100%	99.00%
CPFSK	100%	98.00%	98.00%	100%	99.00%	99.00%	100%
GFSK	98.00%	96.00%	99.00%	96.00%	98.00%	98.00%	100%
PAM4	98.00%	98.00%	98.00%	98.50%	98.00%	98.00%	99.00%
QAM16	47.00%	24.00%	75.00%	92.00%	85.00%	94.00%	95.00%
QAM64	54.00%	81.00%	80.00%	88.00%	90.00%	95.00%	97.00%
QPSK	78.00%	69.00%	85.00%	97.00%	94.00%	96.00%	97.00%
WBFM	16.00%	44.00%	48.00%	35.50%	41.00%	61.00%	39.00%
Average	78.45%	78.63%	83.68%	89.64%	87.55%	90.04%	91.90%

is a wireless communication signal classification dataset funded by the Defense Advanced Research Projects Agency (DARPA) of the United States, containing 10 modulation types with 20 different SNRs. Each modulation type contain 6,000 signal samples at different SNRs, totaling 1.2 million signal samples. All samples are randomly divided into training set, validation set and test set in a 6:2:2 ratio based on the modulation styles and SNR. To ensure fairness, we set a random seed (default value of 2016) to ensure the certainty of the dataset division.

B. SOFTWARE AND HARDWARE ENVIRONMENT

The experimental host runs on the Windows 10 operating system, equipped with an I7-10700K CPU, DDR4 16GB RAM, and an RTX 3060 graphics card with 12GB of VRAM. It has installed software such as CUDA (version 12.2.79), CUDNN (version 11.4), and Anaconda (version 1.10), with modules like Spyder (version 5.3.3) and Keras (version 2.10) installed within the Anaconda environment.

C. TRAINING PROCESS

In order to obtain the optimal model, we maintain the model with the lowest validation loss as the current optimal model. Furthermore, we set the number of training epochs to a relatively large number and employ an early stopping mechanism. The early stopping mechanism means that when the model satisfies certain conditions, the model is considered to have converged, the model training is ended, and the model is saved. In all the experiments in this paper, the training will end when the validation loss still does not decrease after 30 epochs. Additionally, the initial learning rate is set to 0.001, which is changed by the validation loss. The learning rate is 80% lower if, after five epochs, there is no improvement in performance. The minimum learning rate is set to 10^{-7}.The batch size is set to 400.

IV. ANALYSIS OF EXPERIMENTAL RESULTS

A. PERFORMANCE COMPARISON USING THE RML2016.10a DATASET

In this section, we compare the recognition accuracy of DMHNN and other recognition models on the RML2016.10a dataset. Fig. 3 presents the recognition accuracy of our proposed model DMHNN and the current state-of-the-art models on the RML2016.10a dataset as SNR changes. The experimental results are all tested in the runtime environment described in Section III. As the same environment is used, such as the same dataset split, the same optimal model saving mechanism, and the same learning rate decreasing method, Table 2 intuitively compares the recognition efficiency of DMHNN and other models. It is worth mentioning that there are certain differences between the running results under the current configuration and the data provided in the original papers, which may be due to differences in the running environment, parameter configuration, random seeds, etc. Moreover, all data in Table 2 are rounded to retain only 2 digits after the decimal point. When calculating the average recognition accuracy, the original data were first summed and then rounded.

As can be seen from Fig. 3, DMHNN achieves significantly better recognition performance than the other models both at low SNR and at high SNR. Except for SNR of -18 dB, LSTMDAE achieves slightly higher performance than DMHNN. The recognition accuracy of DMHNN is the highest of all other SNRs. The recognition accuracy of DMHNN is no less than 91.91% at SNR not less than 0 dB and as high as 93.56% at SNR 16 dB. Under different SNRs, DMHNN achieved an average recognition accuracy of up to 64.94%, which is 4.09%, 4.66%, 4.97%, 8.18%, and 9.12% higher than LSTMDAE, MCLDNN, PET-CGDNN, MCNet, and CGDNet, respectively. Table 2 shows the recognition accuracies of our proposed model DMHNN with the current

TABLE 3. Average recognition accuracy of DMHNN with other recognition models on RML2016.10b dataset.

Moduation	CGDNet	MCNet	PET-CGDNN	MCLDNN	LSTMDAE	DMHNN
Average	62.47%	60.56%	63.75%	64.47%	59.96%	65.19%

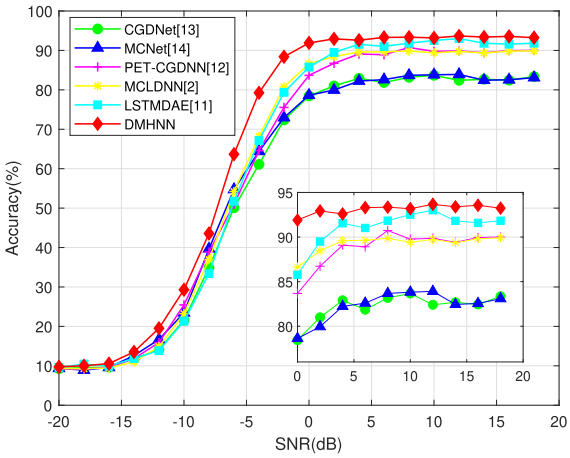


FIGURE 3. Recognition accuracy of common recognition models at different SNRs on RML2016.10a dataset.

highest-level model for different modulation types at 0 dB SNR for the RML2016.10a dataset. It should be noted in particular that although some papers provide source code, even if the network structure is easily reproducible without disclosing the source code, most of the data we obtained are slightly worse than those provided by the original papers, considering that many factors, such as random seeds, hyperparameters, initialization methods, etc., can affect the recognition accuracy. For fairness, Table 2 uses data provided in the original papers as much as possible if they are available.

As can be seen from Table 2, DMHNN has significantly better recognition accuracy than other models for different modulation types at 0 dB SNR on the RML2016.10a dataset. The model achieves the highest recognition accuracy in seven modulation types: 8PSK, CPFSK, GFSK, PAM4, QAM16, QAM64, and QPSK. In particular, the recognition accuracy for CPFSK and GFSK reaches as high as 100%. For QAM16 and QAM64, the model has made a new breakthrough, with accuracies as high as 95% and 98%, respectively, which are significantly better than those of MCLDNN and RLADNN. In terms of average accuracy, DMHNN is as high as 91.9%, which is 1.86%, 4.35%, 2.26%, 8.23%, 13.27%, and 13.45% higher than RLADNN, LSTMDAE, MCLDNN, PET-CGDNN, MCNet, and CGDNet, respectively. The confusion matrices of DMHNN and other recognition models at -2 dB SNR on the RML2016.10a dataset are shown in Fig. 4. Overall, DMHNN's recognition accuracy is significantly better than other models. Analysis of the confusion matrices given in Fig. 4 shows that CGDNet, MCNet, and PET-CGDNN have difficulty in distinguishing between QAM16 and QAM64 modulations when the SNR is −2dB and MCLDNN and LSTMDAE significantly improve the ability to classify QAM16 and QAM64. Compared to other recognition models, DMHNN further enhances the ability to classify QAM16 and QAM64. The recognition accuracy of DMHNN is acceptable for all other modulation types except WBFM. It is important to note that the recognition accuracy of all current models for WBFM needs to be improved, and many WBFM samples are incorrectly classified as AM-DSB signals.

B. PERFORMANCE COMPARISON USING THE RML2016.10b DATASET

To thoroughly evaluate the performance of our model, we compared the recognition accuracy of DMHNN with other recognition models on the RML2016.10b dataset. All models were run in the environment described in Section III.

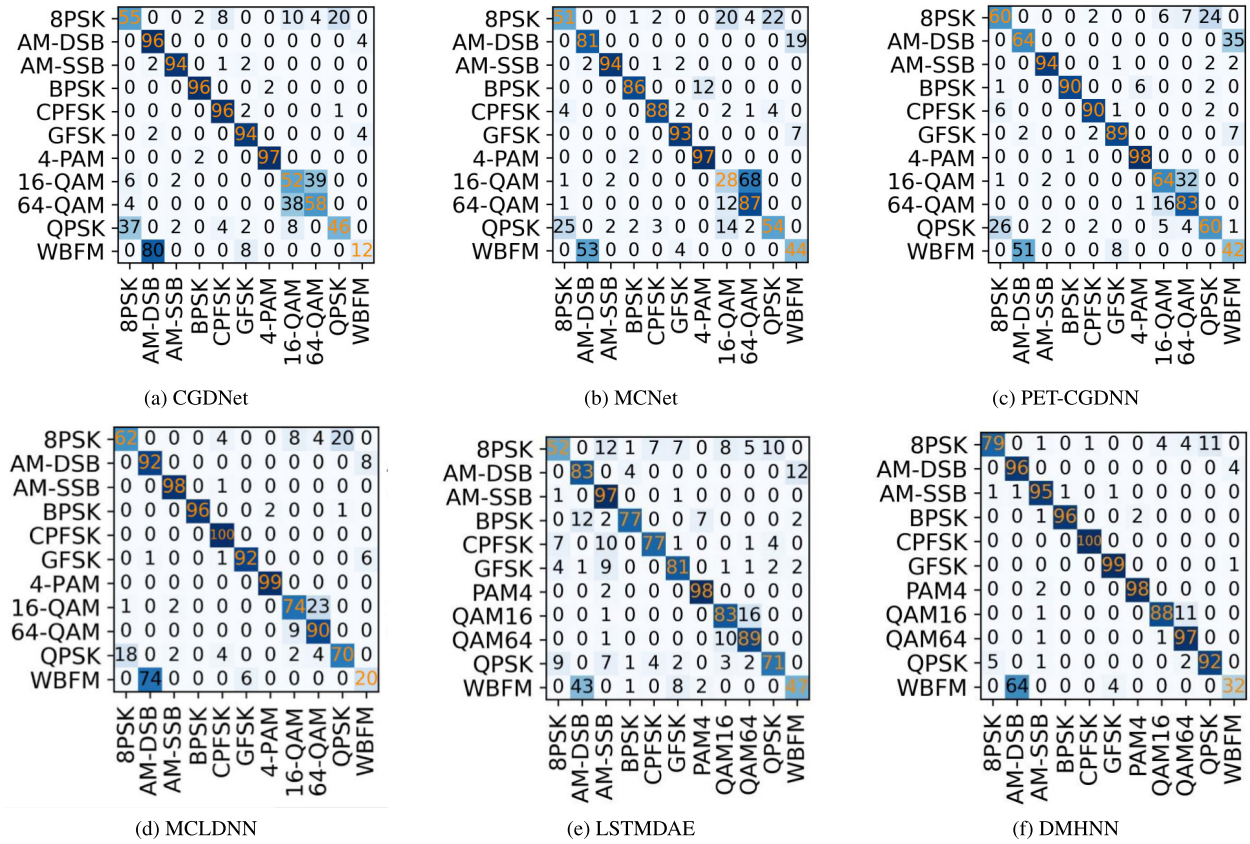
Analysing Fig. 5, it is clear that DMHNN consistently exhibits superior recognition accuracy across different SNR levels, outperforming other recognition models. When SNR is greater than or equal to 0 dB, the recognition accuracy of DMHNN is much higher than PET-CGDNN, CGDNet, and MCNet, and marginally higher than MCLDNN and LSTMDAE. Between -8dB and 0dB SNR, the recognition accuracy of DMHNN is significantly higher than the other five models. This indicates that the DMHNN model performs well for low SNR.

Table 3 compares the average recognition accuracy of DMHNN with other recognition models on the RML2016.10b dataset. The average recognition accuracy of DMHNN is significantly higher than all other models, which is 2.72%, 4.63%, 1.44%, 0.72%, and 5.23% higher than CGDNet, MCNet, PET-CGDNN, MCLDNN, and LSTMDAE, respectively. Combining the experimental results on the RML2016.10a dataset, it is clear that the average recognition accuracy of DMHNN on the RML2016.10b dataset is slightly better than MCLDNN. However, the recognition accuracy of DMHNN on the RML2016.10a dataset is significantly better than MCLDNN. Therefore, DMHNN exhibits better generalization performance.

C. ABLATION EXPERIMENTS

We compressed and amplified β_i, i=1, 2, 3, 4, 5, 6 based on the structure of DMHNN, and obtained four compressed networks (DMHNN-Ti, i=1, 2, 3, 4) and three amplified networks (DMHNN-Fi, i=1, 2, 3), respectively. The eight network sizes of DMHNN and its compressed and amplified networks are shown in Table 4.

The results of the ablation experiments are given in Fig. 6. Fig. 6 (a) shows that, for the RML2016.10a dataset, DMHNN performs better than compressed networks, and that performance decreases noticeably as network size is reduced.



(a) CGDNet (b) MCNet (c) PET-CGDNN

(d) MCLDNN (e) LSTMDAE (f) DMHNN

FIGURE 4. Confusion matrix of DMHNN and other recognition models at −2 dB SNR on RML2016.10a dataset.

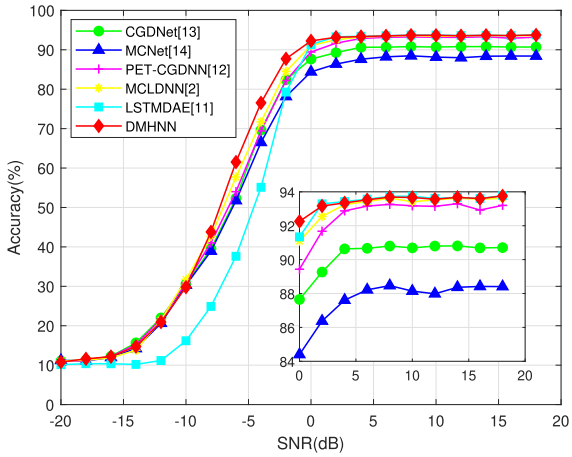


FIGURE 5. Recognition accuracy with SNR for DMHNN and other recognition models on RML2016.10b dataset.

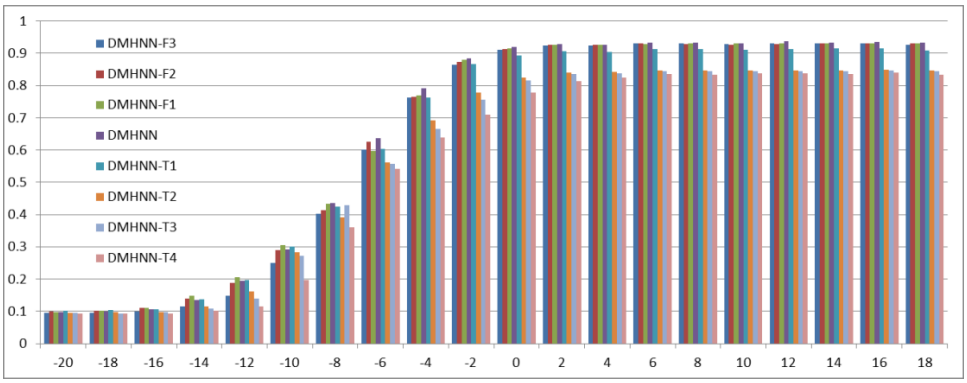
TABLE 4. Network dimensions.

	β_1	β_2	β_3	β_4	β_5	β_6
DMHNN-F3	100	200	400	256	256	256
DMHNN-F2	75	150	300	192	192	192
DMHNN-F1	50	100	200	128	128	128
DMHNN	25	50	100	64	64	64
DMHNN-T1	18	37	75	48	48	48
DMHNN-T2	12	25	50	32	32	32
DMHNN-T3	6	12	25	16	16	16
DMHNN-T4	3	6	12	8	8	8

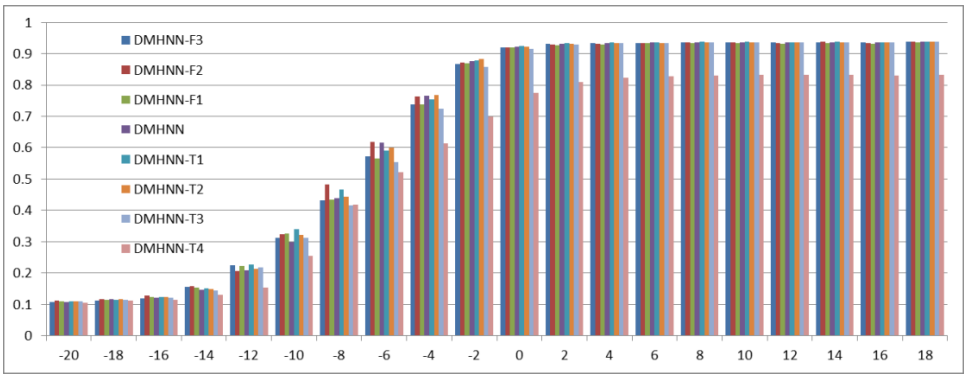
At SNR greater than or equal to −8 dB, DMHNN performs slightly better than DMHNN-F1 and significantly better than DMHNN-F2 and DMHNN-F3. At SNR less than or equal to −10 dB, DMHNN is slightly worse than DMHNN-F1 and DMHNN-F2, but better than DMHNN-F3. Thus, reducing network size results in significant performance degradation, while increasing network size does not achieve significant performance improvements. For the RML2016.10b dataset, as shown in Fig. 6 (b), DMHNN performs similarly to the compressed networks such as DMHNN-T1, DMHNN-T2, and DMHNN-T3, and significantly better than DMHNN-T4. DMHNN also outperforms all the expanded networks. These findings indicate that DMHNN is the optimal network size, with comprehensive performance that surpasses other network sizes.

D. COMPUTATIONAL COMPLEXITY

Table 5 compares the computational complexity of DMHNN with other recognition models. We used three metrics to reflect computational complexity, which is the number of trainable parameters, the training duration per epoch, and



(a) RML2016.10a dataset



(b) RML2016.10b dataset

FIGURE 6. Ablation experiments.

TABLE 5. Computation complexity.

Models	CGDNet	MCNet	PET-CGDNN	MCLDNN	DMHNN
Trainable parameters	124676	121226	71742	406070	156206
Training time(sec)/epoch	132	138	134	142	131
Prediction time (us)/sample	5.3	6.6	8	16	4.4

the inference time per sample. Regarding the quantity of parameters that can be trained, DMHNN has only 38% of the parameters of MCLDNN. Compared to CGDNet and MCNet, DMHNN has approximately 25.3% and 28.9% more parameters, respectively. PET-CGDNN has the smallest parameters. In terms of training duration, the difference in training duration between DMHNN and other models is not significant. In terms of inference time, DMHNN is significantly lower than other models. Because it has a relatively small amount of parameters and a relatively low network depth. Therefore, DMHNN is better suited for deployment in resource-constrained environments with limited inference capabilities than other recognition models.

V. CONCLUSION

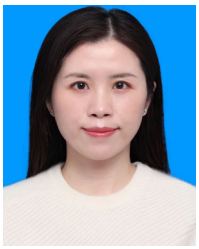
In this paper, we propose a novel hybrid multi-channel neural network structure of CNN and LSTM, referred to as DMHNN. The convolutional layers are able to fully exploit the features of spatial dimensions of neighboring data, while LSTM is able to extract temporal features from sequential data efficiently. The model includes a data converter that transforms raw in-phase/quadrature input into amplitude/phase data. The two functions used in the data conversion process further enhance the nonlinear characterization capability of the whole model, thereby improving its recognition accuracy. The experimental simulation results demonstrate that our model solves the classification challenges associated with QAM16 and QAM64. For RAML2016A, the classification accuracy of both is as high as 95% at an SNR of 0 dB. Extensive experiments show that the network structure proposed in this paper outperforms other current networks in terms of recognition accuracy.

REFERENCES

[1] F. Zhang, C. Luo, J. Xu, Y. Luo, and F.-C. Zheng, "Deep learning based automatic modulation recognition: Models, datasets, and challenges," *Digit. Signal Process.*, vol. 129, Sep. 2022, Art. no. 103650.

[2] 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.

[3] M. M. Elsagheer and S. M. Ramzy, "A hybrid model for automatic modulation classification based on residual neural networks and long short term memory," *Alexandria Eng. J.*, vol. 67, pp. 117–128, Mar. 2023.
[4] 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.
[5] K. Liu and F. Li, "Automatic modulation recognition based on a multiscale network with statistical features," *Phys. Commun.*, vol. 58, Jun. 2023, Art. no. 102052.
[6] F. Liu, Z. Zhang, and R. Zhou, "Automatic modulation recognition based on a multiscale network with statistical features," *Tsinghua Sci. Technol.*, vol. 27, no. 2, pp. 422–431, Apr. 2022.
[7] B. Wang, S. Zhang, and Y. Zhu, "Modulation signal recognition based on feed-forward attention mechanism," *Electron. Lett.*, vol. 59, no. 14, Jul. 2023.
[8] F. Zhou, J. Li, and Y. Wang, "An improved CNN-LSTM network for modulation identification relying on periodic features of signal," *IET Commun.*, vol. 17, no. 18, pp. 2097–2106, Sep. 2023.
[9] 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.
[10] 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.
[11] 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.
[12] 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.
[13] J. N. Njoku, M. E. Morocho-Cayamcela, and W. Lim, "CGDNet: Efficient hybrid deep learning model for robust automatic modulation recognition," *IEEE Netw. Lett.*, vol. 3, no. 2, pp. 47–51, Jun. 2021.
[14] T. Huynh-The, C.-H. Hua, Q.-V. Pham, and D.-S. Kim, "MCNet: An efficient CNN architecture for robust automatic modulation classification," *IEEE Commun. Lett.*, vol. 24, no. 4, pp. 811–815, Apr. 2020.
[15] 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. Cognit. Commun. Netw.*, vol. 4, no. 3, pp. 433–445, Sep. 2018.
[16] S. Chang, S. Huang, R. Zhang, Z. Feng, and L. Liu, "Multitask-learning-based deep neural network for automatic modulation classification," *IEEE Internet Things J.*, vol. 9, no. 3, pp. 2192–2206, Feb. 2022.
[17] Z. Zhang, H. Luo, C. Wang, C. Gan, and Y. Xiang, "Automatic modulation classification using CNN-LSTM based dual-stream structure," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13521–13531, Nov. 2020.
[18] T. Huynh-The, Q.-V. Pham, T.-V. Nguyen, T. T. Nguyen, D. B. D. Costa, and D.-S. Kim, "RanNet: Learning residual-attention structure in CNNs for automatic modulation classification," *IEEE Wireless Commun. Lett.*, vol. 11, no. 6, pp. 1243–1247, Jun. 2022.
[19] P. Ghasemzadeh, S. Banerjee, M. Hempel, and H. Sharif, "A novel deep learning and polar transformation framework for an adaptive automatic modulation classification," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13243–13258, Nov. 2020.
[20] T. Wang, Y. Hou, H. Zhang, and Z. Guo, "Deep learning based modulation recognition with multi-cue fusion," *IEEE Wireless Commun. Lett.*, vol. 10, no. 8, pp. 1757–1760, Aug. 2021.
[21] P. Qi, X. Zhou, S. Zheng, and Z. Li, "Automatic modulation classification based on deep residual networks with multimodal information," *IEEE Trans. Cognit. Commun. Netw.*, vol. 7, no. 1, pp. 21–33, Mar. 2021.



MENG QI received the B.S. degree in communication engineering from Changsha University of Science and Technology, in 2006, and the M.S. degree in communications and information system from Zhengzhou University, China, in 2010. She is currently a Lecturer with Luoyang Institute of Science and Technology. Her research interests include artificial intelligence, machine learning, and deep learning in biomedical signal processing.

NIANFENG SHI received the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, China, in 2012, and the Ph.D. degree in computer application from Chinese Academy of Sciences. He was a Visiting Scholar with the University of Southern Queensland. He is currently a Computer Science Professor with Luoyang Institute of Science and Technology. His research interests include pattern recognition and computer education.

GUOQIANG WANG received the Ph.D. degree in engineering from Dalian University of Technology, in 2008. He is currently a Computer Science Professor with Luoyang Institute of Science and Technology. His research interests include pattern recognition, image processing, and deep learning.

HONGXIANG SHAO received the B.S. degree from Southwest University of Science and Technology, Mianyang, China, in 2007, and the Ph.D. degree in communications and information system from the Institute of Communications Engineering, PLA University of Science and Technology, in 2018. He is currently a Lecturer with Luoyang Institute of Science and Technology. His research interests include opportunistic spectrum access, learning theory, game theory, and optimization techniques.