IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 16, 2020, accepted March 12, 2020, date of publication March 16, 2020, date of current version July 6, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2981130

Deep Learning for Robust Automatic Modulation Recognition Method for IoT Applications

TINGPING ZHANG[1], CONG SHUAI[1], AND YARU ZHOU[2], (Student Member, IEEE)
[1]School of Information Science and Engineering, Chongqing Jiaotong University, Chongqing 400074, China
[2]College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

Corresponding author: Tingping Zhang (ztp@cqjtu.edu.cn)

ABSTRACT In the scenarios of non-cooperative wireless communications, automatic modulation recognition (AMR) is an indispensable algorithm to recognize various types of signal modulations before demodulation in many internet of things applications. Convolutional neural network (CNN)-based AMR is considered as one of the most promising methods to achieve good recognition performance. However, conventional CNN-based methods are often unstable and also lack of generalized capabilities under varying noise conditions, because these methods are merely trained on specific dataset and can only work at the corresponding noise condition. Hence, it is hard to apply these methods directly in practical systems. In this paper, we propose a CNN-based robust automatic modulation recognition (RAMR) method to recognize three types of modulation signals, i.e., frequency shift key (FSK), phase shift key (PSK), and quadrature amplitude modulation (QAM). The proposed method is trained on a mixed dataset for extracting common features under varying noise scenarios. Simulation results show that our proposed generalized CNN-based architecture can achieve higher robustness and convenience than conventional ones.

INDEX TERMS Automatic modulation recognition, deep learning, convolutional neural network, the Internet of Things.

I. INTRODUCTION

Automatic modulation recognition (AMR) is an essential technology in non-cooperative communication systems or internet of things applications for demodulation tasks of unknown signals, and it has various applications in military and civilian strategies [1]–[9]. In the aspect of modern military applications, AMR is a key step to recover the intercepted signals in electronic warfare (EW) [10]. In civilian scenarios, AMR can also act as an assistance to interference signal analysis or spectrum sensing [11] in heterogeneous network communications [12]–[16] and direction of arrival estimation [17]–[19].

There are two common AMR methods are mainly based on likelihood and features, respectively. In the likelihood-based method, AMR can be formulated as a hypothesis testing problem [5]. It is necessary to design correct likelihood function to evaluate likelihood for each modulation type within hypothesis pooling. Then, the likelihoods of each modulation type are compared to make a final decision. However, likelihood-based AMR methods excessively depend on channel state information (CSI). In the feature-based method, AMR is modeled as a pattern recognition problem and it consists of three steps: pre-processing, feature extraction, and classifier design [20]. Various AMR methods have been developed using instantaneous features (or signal spectral-based features), wavelet transform-based features, high-order statistics-based features, cyclic spectrum analysis-based features and so on. To realize modulation type classification by extracted features, they usually adopt classifiers, such as support vector machine (SVM), decision tree (DT), k-nearest neighbor (KNN) and multilayer perception (MLP).

In recent years, deep learning (DL) is considered as a powerful tool, because it is expert in automatic feature extraction from huge amounts of data, instead of the complex and difficult design of man-made features [21]–[24]. For this reason, DL has been successfully applied in network traffic prediction [25]–[28], physical layer wireless techniques [30]–[35] and internet-of-things [31], [36]–[39]. In addition, DL has been applied in multiple input and multiple output (MIMO) [40], non-orthogonal multiple access (NOMA), and cognitive radio (CR). For example, Paper [41] proposed a fast

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

beam forming technology for downlink MIMO based on unsupervised learning. Paper [30] applied a long short-term memory (LSTM) network into typical NOMA system for enhancing spectral efficiency. Papers [42], [43] introduced deep learning into resource allocation in CR and achieved great success.

Moreover, state-of-the-art DL-based AMR methods have been developed in recent two years. Paper [44] proposed a convolutional neural network (CNN)-based AMR, which is realized by training CNN on the in-phase and quadrature (IQ) component of signals. Paper [45] transformed the modulated signals into constellation diagrams, and then generative adversarial network (GAN) was applied to distinguish these constellation diagrams. Paper [46] proposed a lightweight and fast CNN-based AMR method for edge devices on their previous works [45], [47]. Paper [48] proposed combined *IQ* sample-based CNN and constellation diagram-based CNN method to recognize different modulation types.

Although these CNN-based AMR methods have been proposed to demonstrate better performance than traditional methods, most of them are trained by dataset with single signal-to-noise (SNR). It means that these CNNs can just achieve satisfying performance at the corresponding single SNR rather than all SNR scenarios. These independent CNNs are termed as the fixed CNNs, and are hard to be generalized. If we adopt these CNNs in practical applications, we must train various CNNs with dataset collected from different SNR conditions, and choose correct CNN-models according to actual communication environments, which is not convenient.

In this paper, a generalized CNN-based automatic modulation recognition (RAMR) method with higher generalized capability under varying noise conditions is proposed, and it achieves higher convenience and robustness for actual project applications at a slight cost of recognition accuracy. Compared with fixed CNN-based AMR methods, our proposed method has two obvious advantages, which are listed as follows:

1) PRECISE SNR ESTIMATION IS UNNECESSARY IN OUR PROPOSED METHOD

Fixed CNN-based AMR methods rely on precise SNR estimation, because fixed CNNs are trained on samples with single SNR. For multiple CNNs-based solution, precise SNR estimation is essential to assist systems in selecting correct CNN from trained fixed CNNs. If SNR cannot be estimated precisely, these methods may be ineffective. Unlike these conventional methods, generalized CNN is trained on a mixed dataset containing different signals with SNR $\in \{-5$ dB, 0 dB, 5 dB$\}$. Unknown signals with SNR ranging from -5 dB to 5 dB can be recognized by the same CNN. It means that we just need to determine whether the SNR is in the range from -5 dB to 5dB before recognizing, and then directly apply generalized CNN without other operations.

2) OUR PROPOSED METHOD CONSUMES LESS DEVICE MEMORY

When SNR is ranging from -5 dB to 5 dB with interval 1 dB, far more than one fixed CNN should be trained for responding to different SNR condition in fixed CNN-based AMR methods. However, our proposed method just need to train one CNN in actual project applications. In the case of the same network structures, our method just requires less device memory than fixed CNN-based AMR methods.

The rest of this paper is organized as follows. Section II includes system model, signal model and dataset. In Section III, we propose a generalized CNN-based AMR methods. In Section IV, various experiment results are provided to compare their performance under AWGN and Rayleigh channels, respectively. Finally, we conclude this paper.

II. SYSTEM MODEL, SIGNAL MODEL AND DATASET

A. SYSTEM MODEL

A typical non-cooperative communication system is considered, where transmitters transmit digital modulation signals through wireless channel with multipath fading and additive white Gaussian noise (AWGN), and receiver does not get any priori information about modulation types, symbol rates and so on. Receiver detected these signals, and then the system makes preprocessing, including down conversion, low pass filtering and analog-to-digital conversion. After preprocessing, we can get baseband signals, which are fed into an AMR module to identify modulation types. AMR-based receiver is shown in Fig. 1.

In this paper, we focus on the feature-based AMR methods, and they generally consist of three steps: processing, feature extraction and classification [20]. In traditional AMR methods, the most difficult part is the design of effective manmade features, and classifiers are usually based on machine learning or simple threshold detection, which is shown in Fig. 2(a). Unlike the traditional methods, DL methods, e.g., CNN or recurrent neural network (RNN), can simultaneously achieve feature extraction and classification. Moreover, the proposed DL-based RAMR method can get rid of complex and different manmade feature design. The framework of the DL-based RAMR method is depicted as Fig. 2(b).

B. SIGNAL MODEL

C. DATASET GENERATION

The received sequence is defined as R, and is shown as

$$R = \{r(1), r(2), \ldots, r(N)\}, \tag{1}$$

which is a complex matrix with dimension $1 \times N$ and the value of N is 128. Then, we separate real part $real(R)$ and imaginary part $imag(R)$ of R. Next, $real(R)$ and $imag(R)$ are combined into a matrix with dimension $2 \times N$, and it is one sample for training or testing. In addition, the real part and imaginary part are also in-phase (I) component and

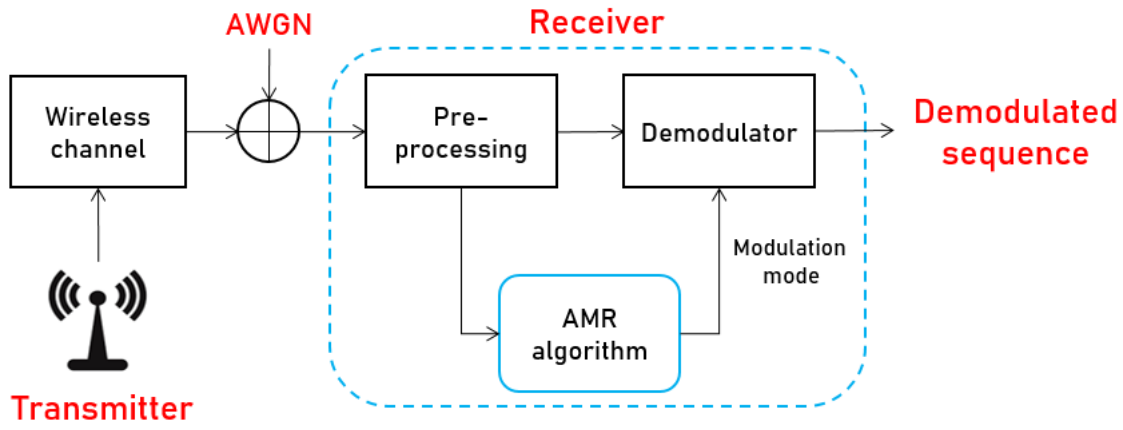

FIGURE 1. AMR-based receiver in a non-cooperative communication system.

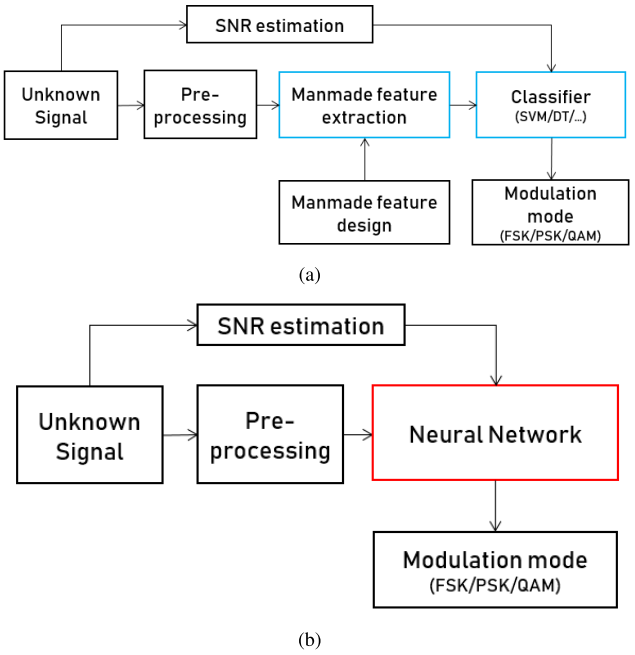

FIGURE 2. Feature-based AMR methods: (a) The structure of traditional AMR method. (b) The structure of the proposed DL-based AMR method.

quadrature (Q) component of signal, respectively. So this training or test sample is denoted as IQ sample, which is shown in (2).

$$IQ = \begin{bmatrix} real(R) \\ imag(R) \end{bmatrix} = \begin{Bmatrix} real[r(1)], real[r(2)], \ldots, real[r(N)] \\ imag[r(1)], imag[r(2)], \ldots, imag[r(N)] \end{Bmatrix}. \quad (2)$$

In this paper, we prepare two independent datasets for training and testing, and the number of each modulation type at each SNR can reach up to 6000 in both training dataset and test dataset. DL-based AMR method applies these dataset for training and testing, respectively. In order to compare DL-based RAMR method with traditional AMR method fairly, the training and test of typical machine learning-based classifiers depend on extracted manmade features from training dataset and testing dataset, respectively.

III. CNN-BASED AMR METHODS AND OUR PROPOSED RAMR METHOD

In this section, a generalized CNN-based RAMR method is proposed with better generalization performance under varying noise condition. The better generalization performance benefits from a mixed training dataset and different training strategies, so it is stated from two aspects of neural network structure, and training and testing methods. In addition, traditional AMR methods, based on classical manmade features and typical machine learning-based classifiers, are first introduced as a comparison.

A. TRADITIONAL AMR METHODS

For the purpose of highlighting the performance of DL-based AMR method, we adopt one of traditional AMR methods as a comparison. The structure of traditional AMR method is shown in Fig. 2(a). Unknown signal must pass through pre-processing and the system needs to estimate SNR for the choice of suitable classifier. Then we can extract manmade features from pre-processed signal, and suitable classifier can recognize modulation type. In this paper, we choose instantaneous features [49] and high-order cumulants (HOC) [49] as classical manmade features, which are listed as follows.

1) INSTANTANEOUS FEATURES

The instantaneous features are mostly extracted from instantaneous amplitude and instantaneous phase, which can respectively represent as:

$$a(n) = \sqrt{real^2[r(n)] + imag^2[r(n)]}, \quad (3)$$

$$\theta(n) = artan \frac{imag[r(n)]}{real[r(n)]}. \quad (4)$$

The four instantaneous features in this paper has been applied in paper [49], and they are respectively:

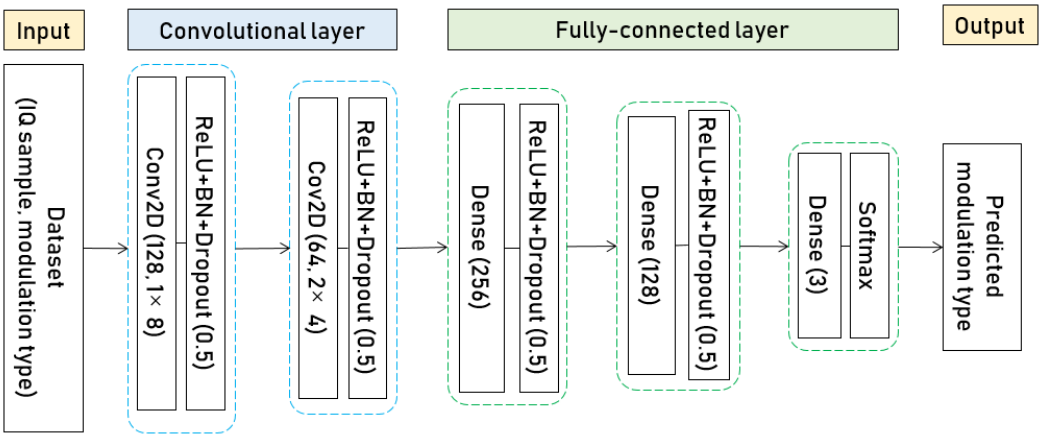

FIGURE 3. The structure of generalized CNN method.

Maximum value of the power spectral density of the normalized-centered instantaneous amplitude:

$$\gamma_{max} = max\left\{\left|DFT\left[{a_c}^2(n)\right]\right|\right\}/N, \tag{5}$$

where $a_c(n) = \frac{a(n)}{m_a} - 1$, $m_a = \frac{1}{N}\sum_{n=1}^{N} a(n)$, N is the number of sampling points, and DFT(·) is discrete Fourier transform.

2) HOC

We apply the estimation of second-order, fourth-order, and sixth-order cumulants, and combine them into seven feature values, which are proposed in [20]. The estimation value of the O-th moment $\hat{M}_{mk}$ can be denoted as $\hat{M}_{mk} = \frac{\sum_{n=1}^{N} r^{m-k}(n) r^{*k}(n)}{N}$, where $O = m + k$ and $r^*(n)$ is the conjugate of $r(n)$. So the estimation values of HOC can be describe as:

$$\hat{C}_{20} = \hat{M}_{20}, \tag{6}$$
$$\hat{C}_{21} = \hat{M}_{21}, \tag{7}$$
$$\hat{C}_{40} = \hat{M}_{40} - 3\hat{M}_{20}^2, \tag{8}$$
$$\hat{C}_{41} = \hat{M}_{41} - 3\hat{M}_{21}\hat{M}_{20}, \tag{9}$$
$$\hat{C}_{42} = \hat{M}_{42} - \left|\hat{M}_{20}\right|^2 - 2\hat{M}_{21}^2, \tag{10}$$
$$\hat{C}_{60} = \hat{M}_{60} - 15\hat{M}_{20}\hat{M}_{40} + 30\hat{M}_{20}^3, \tag{11}$$
$$\hat{C}_{63} = \hat{M}_{63} - 9\hat{M}_{42}\hat{M}_{21} - 6\hat{M}_{21}^3, \tag{12}$$

The seven feature values based on HOC are shown as follows:

$$f_1 = \frac{\left|\hat{C}_{40}\right|}{\left|\hat{C}_{42}\right|}, \quad f_2 = \frac{\left|\hat{C}_{41}\right|}{\left|\hat{C}_{42}\right|}, f_3 = \frac{\left|\hat{C}_{42}\right|}{\left|\hat{C}_{21}\right|^2}, f_4 = \frac{\left|\hat{C}_{60}\right|}{\left|\hat{C}_{21}\right|^3},$$
$$f_5 = \frac{\left|\hat{C}_{63}\right|}{\left|\hat{C}_{21}\right|^3}, \quad f_6 = \frac{\left|\hat{C}_{60}\right|^2}{\left|\hat{C}_{42}\right|^3}, f_7 = \frac{\left|\hat{C}_{63}\right|^2}{\left|\hat{C}_{42}\right|^3}. \tag{13}$$

We combine the eleven feature values from instantaneous features and HOC into a eleven-dimension feature vector $\left[\gamma_{max}, \sigma_{aa}, \sigma_{ap}, \sigma_{dp}, f_1, f_2, f_3, f_4, f_5, f_6, f_7\right]$, and two typical machine learning-based classifiers of SVM and DT are applied to recognize the modulation type of this signal, relying on these feature vectors. Hence, these traditional AMR methods are adopted as SVM-based AMR method and DT-based AMR method, respectively.

B. THE PROPOSED CNN-BASED RAMR METHOD

1) THE STRUCTURE OF GENERALIZED CNN METHOD

The CNN consists of four parts: input, output, convolutional layer and fully-connected layer, and its structure is depicted in Fig. 3. Input is dataset with samples and corresponding labels. Output is a probability distribution, which contains the possibility of each modulation type. Convolutional layer is to automatically extract features and contains two convolutional layers. In convolutional layer, the convolutional kernel size 1×4 and 1×8, and it is designed according to input data: IQ sample, the dimensionality of which is 2×128. The fully-connected layer is fundamentally a classifier with three dense layers.

Rectified linear unit ($ReLU$) plays the role of activation function following behind each available layers except the last dense layer, where $Softmax$ is applied. Assuming x_i is the output of the i-th neuron in certain layer, the function of $ReLU$ and $Softmax$ can be described as follows:

$$f_{ReLU}(x_i) = \max(0, x_i). \tag{14}$$
$$f_{Softmax}(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}. \tag{15}$$

In addition, we also introduce dropout layers to avoid overfitting and Adam as the optimizer. Categorical cross entropy (CCE) function is applied as loss function, considering that AMR is essentially a muti-class classification task. CCE loss function is also called $Softmax$ loss function, due to that CCE is a combination of $Softmax$ activation and cross entropy

loss function. The function of CE is shown as

$$y_{CE} = -\sum_{C} y_c log(\widehat{y}_c), \tag{16}$$

where y_c is the ground truth vector, and it can be achieved through one-hot encoding of sample label. y_c is the predicted vector. C is the number of the samples' types. On the basis of *CE* and *Softmax*, the *CCE* loss function is given as follows.

$$y_{CCE} = -\sum_{C} y_c log(f_{Softmax}\,(x_i)), \tag{17}$$

2) TRAINING AND TEST STRATEGIES

In Section I, we have introduced the weak generalized capability of existing fixed CNN-based AMR methods. In this paper, we propose a generalized CNN-based RAMR method, and it has generalized ability to against varying noise condition with different SNRs. The structure of generalized CNN and fixed CNN are the same, which has been shown in Fig. 3, and the differences between them are training and testing strategies.

The fixed CNN is often trained on data with single SNR, but SNR of received signals is unchangeable. So we must train various fixed CNNs for varying noise conditions. In reality, fixed CNN-based AMR can be applied in this system, the structure of which is shown in Fig. 4(a). When receiving unknown signal, on the one hand, system should make preprocessing and convert signals into *IQ* samples; on the other hand, system also need to precisely estimate the SNR. Then, system chooses certain trained CNN model according to the estimated SNR, and the *IQ* samples are fed into chosen CNN to be recognized.

Fixed CNN-based AMR method requires precise SNR estimation to select a trained model, but the generalized CC-based RAMR method only needs to determine whether SNR is in the range of −5 dB to 5 dB. Hence, the training and test method of the generalized CNN is different. Before training, we need to proportionally mix three datasets with SNRs of −5 dB, 0 dB and 5 dB. Then, the mixed dataset is divided into training dataset and validation dataset by 7 : 3 in random. Training dataset is fed into generalized CNN for training and validation dataset is applied to measure the performance of trained CNN after each epoch. The trained generalized CNN can be employed into modulation type recognition of unknown signals with SNR ranging from −5 dB to 5 dB.

The detailed training steps are given in TABLE 1. It is worth noting that early-stopping is adopted into training to avoid overfitting, and it means that if the performance of the generalized CNN cannot be improved on validation dataset, training will stops automatically. We choose five epochs for AWGN channel and ten epochs for Rayleigh fading channel, as intermittent cycle, respectively. The curves of accuracy and cross entropy loss changing in training generalized CNN are shown in Fig. 5. The epoch times for training generalized CNN are set as 40 in AWGN channel, but 55 training epoch times are required for generalized CNN in Rayleigh fading channel. The loss of training generalized CNN cannot be improved after the 35 epoch and 45-th epoch for AWGN channel and Rayleigh fading channel, respectively. Hence, the training is terminated.

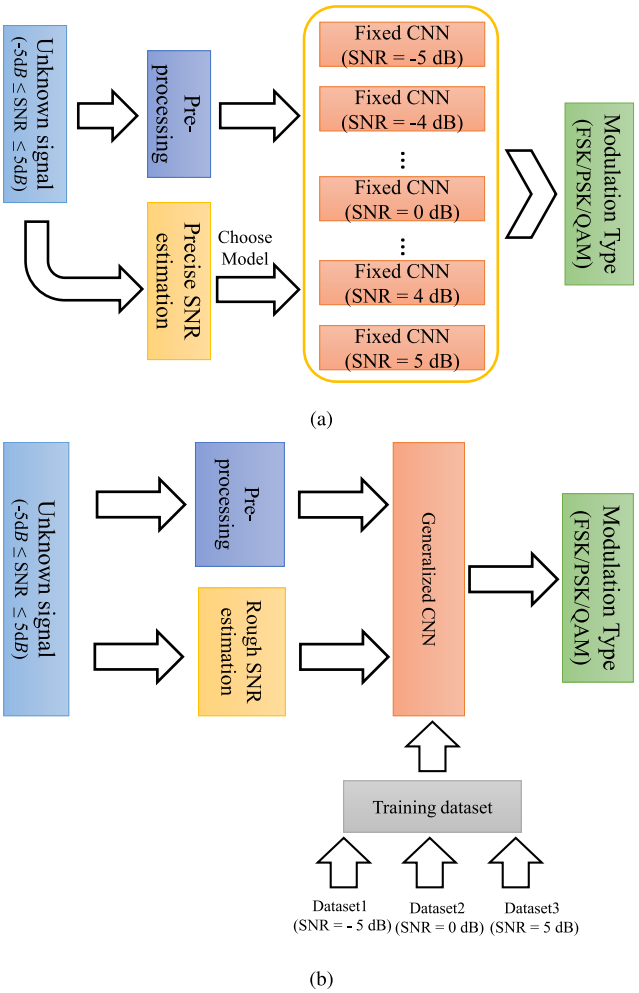

FIGURE 4. The flow chart of actual application: (a) Fixed CNN-based AMR method where SNR = *j* dB) is trained on samples with SNR = *j* dB; (b) The proposed CNN-based RAMR method where is trained on mixed dataset with three SNR regimes: {−5 dB, 0 dB, 5 dB}.

TABLE 1. The training of the proposed CNN-based RAMR method.

Algorithm: The proposed CNN-based RAMR method
Input: Transmitted data
Output: Generalized CNN
Step 1: Generate wireless channel and add AWGN. Create three datasets containing FSK, PSK and QAM baseband signals with SNR $\in\{-5\text{dB}, 0\text{dB}, 5\text{dB}\}$.
Step 2: Mix three datasets proportionally, and randomly divided the mixed dataset into training dataset and validation dataset by 7 : 3.
Step 3: Construct Generalized CNN according to Fig. 3. In addition, initialize CNNs weight, and set learning rate = 0.001.
Step 4: Train Generalized CNN on training dataset and update CNNs weight according to Adam until validation loss is not improved in setting epochs. Setting epoch times are 5 in AWGN channel and 10 in Rayleigh fading channel.
Return: Generalized CNN

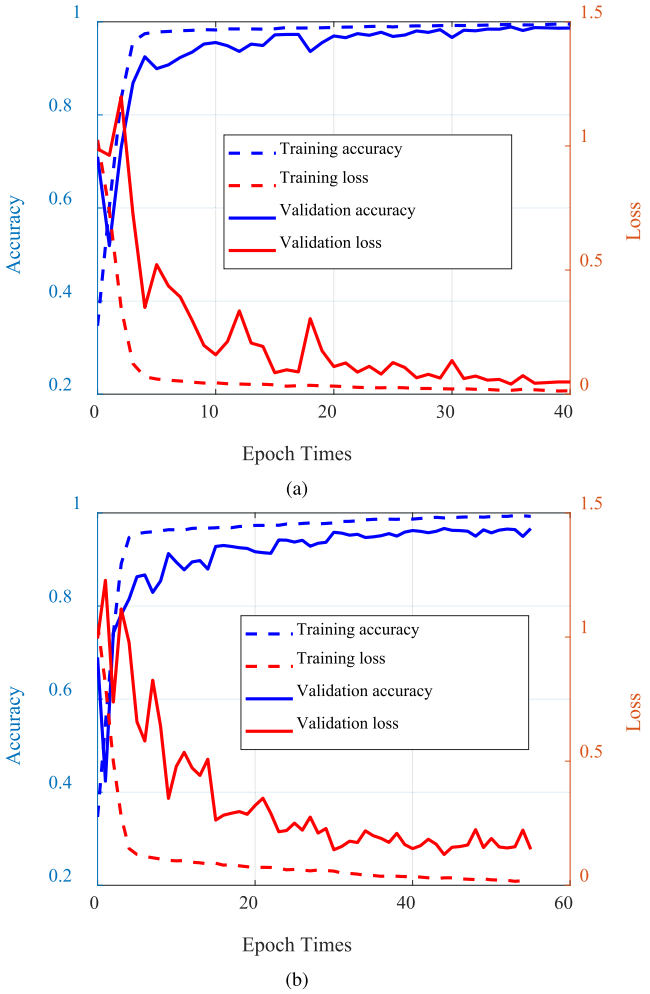

FIGURE 5. The changing of accuracy and cross entropy loss in training generalized CNN: (a) AWGN channel, (b) Rayleigh fading channel.

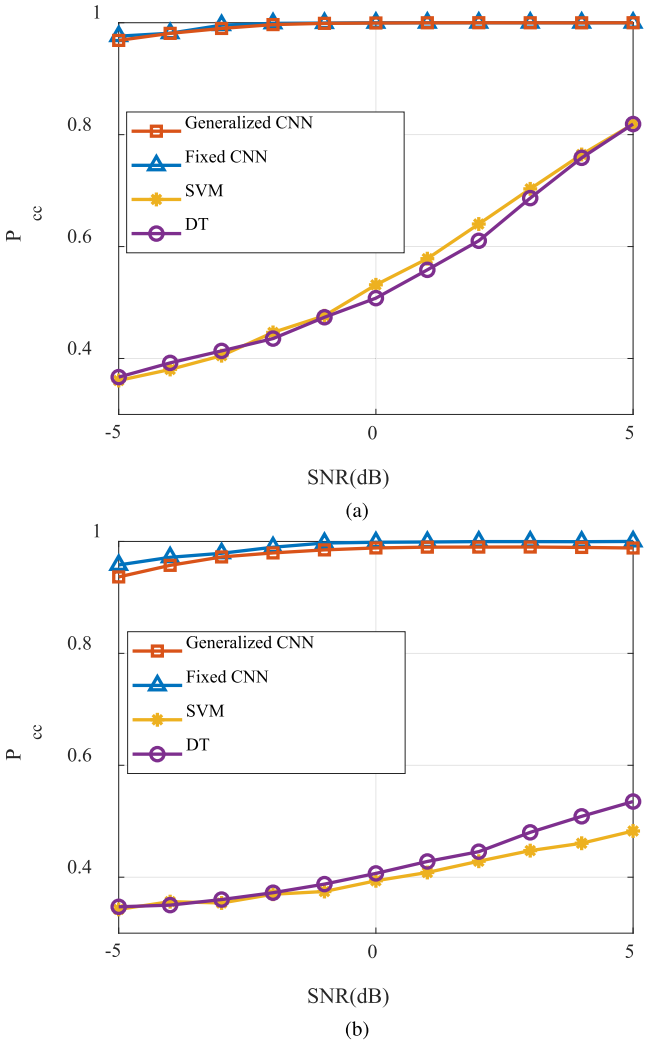

FIGURE 6. The performance of various AMR methods: (a) AWGN channel, (b) Rayleigh fading channel.

Generalized CNN-based RAMR method can be applied into actual applications according to Fig. 4(b). Compared Fig. 4(a) and Fig. 4(b), there are two differences between generalized CNN-based RAMR method and fixed CNN-based AMR method. The first difference is that we just need to train one CNN model in generalized CNN-aided system. The second one is that generalized CNN-based system just needs to make rough SNR estimation rather than precisely estimate SNR. That is to say, fixed CNN-based AMR method requires precise SNR estimation to select a trained model, but generalized CNN-based RAMR method only needs to determinate whether SNR is in the range of -5 dB to 5 dB.

3) IMPLEMENTATION

The experiment, especially network trimming, requires powerful computing resources, so it conducted in the platform with one NVIDIA GTX 1080Ti GPU and the implementation of the network relies on Keras 2.2.2 with Tensorflow 1.10 and Python 3.6.5 as the backend. Moreover, Matlab R2018a is applied to build our datasets.

IV. EXPERIMENT RESULTS

A. PERFORMANCES OF VARIOUS AMR METHODS

In this paper, correct classification probability P_{CC} is adopted to measure the performances of these AMR methods, which is given as (18)

$$P_{cc}(j) = \frac{S_{cc}(j)}{S(j)}, \tag{18}$$

where $S(j)$ is the number of all test samples at SNR=j dB (where $j \in [-5$ dB, 5 dB]) and $S(j) = 18000$. $S_{cc}(j)$ is the number of test samples at SNR=j dB, which are correctly classified. These performances are measured on the independent IQ testing dataset, and the test data of the traditional AMR methods are a set of feature vectors extracted from the IQ test dataset.

The performances of various AMR methods in AWGN channel and Rayleigh channel are depicted in Fig. 6. From these experiment results, we can observe that AMR algorithms, based on SVM or DT, have unsatisfactory

performances in the two channels (i.e, AWCN channel and Rayleigh channel). Here, both fixed CNN and generalized CNN-based RAMR algorithm can work perfectly in the two kinds of channels. The P_{cc} in AWGN channel can reach almost 100% when SNR $>$ -3 dB, and similar performances can be achieved in Rayleigh channel when SNR > -1 dB.

For the purpose of evaluating the performances of the two CNN-based AMR algorithms, $P_{cc}^{gap}(j)$ is proposed to measure the performance gap, which is defined as follows.

$$P_{cc}^{gap}(j) = P_{cc}^{fixedCNN}(j) - P_{cc}^{generalizedCNN}(j), \tag{19}$$

where $P_{cc}^{fixedCNN}(j)$ and $P_{cc}^{generalizedCNN}(j)$ are the correct classification probabilities of fixed CNN and generalized CNN for SNR $= j$ dB ($j \in [-5$ dB, 5 dB]), respectively. The calculated $P_{cc}^{gap}(j)$ is shown in Fig. 7. There are almost no performance loss in AWGN channel, and performance loss in Rayleigh channel ranges from 1% to 2%, and most of losses are around 1%. In addition, the method cannot promise the optimal performance at each SNR, but presents a global performance optimization for SNR ranging from -5 dB to 5 dB. Thus, the performance losses at varying SNR are different.

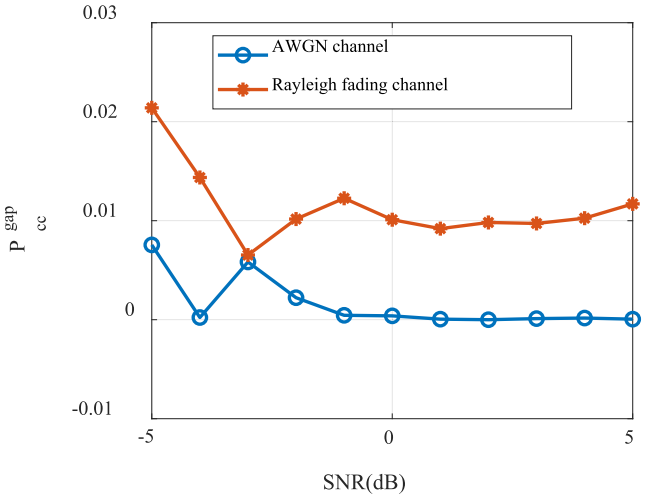

FIGURE 7. The performance gap between fixed CNN and generalized CNN.

B. GENERALIZED CAPABILITIES UNDER VARYING NOISE CONDITION

To compare the generalization capabilities between the fixed CNN and generalized CNN under varying noise conditions, we depict three curves of fixed CNN trained at different SNRs (i.e., -5 dB, 0 dB, 5 dB) in Fig. 8, and we tested them at SNRs ranging from -5 dB to 5 dB. These three fixed CNNs perform well the testing when SNRs are close to j dB, but the training SNRs performance gets worse at other SNRs, which means that the fixed CNN does not have higher robustness and generalization capabilities. On the contrary,

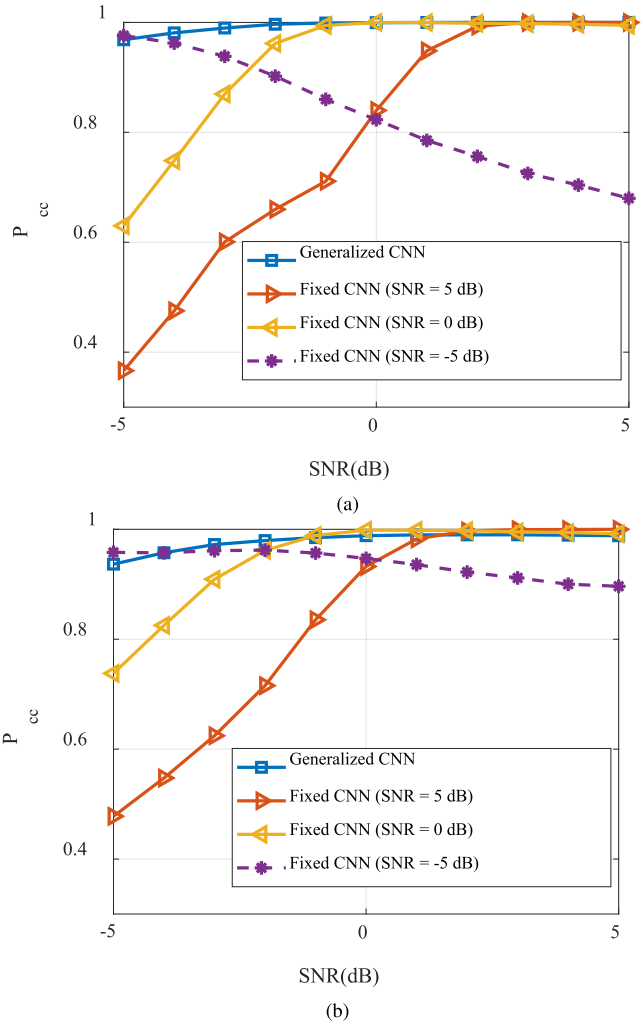

FIGURE 8. Generalized capabilities under varying noise condition: (a) AWGN channel. (b) Rayleigh fading channel.

Fig. 8 demonstrates that generalized CNN can work effectively at all the testing SNRs.

V. CONCLUSION

In this paper, we have proposed a generalized CNN-based RAMR method with better robustness under varying noise conditions. Compared with the traditional AMR methods, the classification accuracy of the proposed method is far beyond not only in AWGN channel but also in Rayleigh fading channel. Besides, it is more robust than fixed CNN-based AMR method at the cost of negligible performance loss, because the generalized CNN, trained by a mixed *IQ* dataset containing received signals with SNRs of -5 dB, 0 dB and 5 dB, can be applied to recognize modulation types of signals with uncertain SNR from -5 dB to 5 dB, while the fixed CNN is trained on fixed SNR and can just be applied to identify signal modulation types for precisely estimated SNR. Moreover, our proposed method is more

convenient for deployment than the fixed CNN-based AMR methods, because precise SNR estimation is unnecessary in our method. We just need to determine whether SNR is within the range from −5 dB to 5 dB before recognizing. In addition, we just have to train one generalized CNN model rather than many fixed CNN, which means fewer device memory assumption. Hence, our proposed generalized CNN-based RAMR method is meaningful for practical applications.

REFERENCES

[1] V. Orlic and M. 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.
[2] L. Haring, Y. Chen, and A. Czylwik, "Automatic modulation classification methods for wireless OFDM systems in TDD mode," *IEEE Trans. Commun.*, vol. 58, no. 9, pp. 2480–2485, Sep. 2010.
[3] A. Abdelmutalab, K. Assaleh, and M. El-Tarhuni, "Automatic modulation classification using hierarchical polynomial classifier and stepwise regression," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Apr. 2016, vol. 17, no. 10, pp. 1–5.
[4] S. Hakimi and G. A. Hodtani, "Optimized distributed automatic modulation classification in wireless sensor networks using information theoretic measures," *IEEE Sensors J.*, vol. 17, no. 10, pp. 3079–3091, May 2017.
[5] F. Meng, P. Chen, L. Wu, and X. Wang, "Automatic modulation classification: A deep learning enabled approach," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 10760–10772, Nov. 2018.
[6] F. Wen, J. Shi, and Z. Zhang, "Joint 2D-DOD, 2D-DOA, and polarization angles estimation for bistatic EMVS-MIMO radar via PARAFAC analysis," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 1626–1638, Feb. 2020.
[7] W. Fangqing, J. Shi, and Z. Zhang, "Direction finding for bistatic MIMO radar with unknown spatially colored noise," *Circuits Syst. Signal Process.*, to be published, doi: 10.1007/s00034-019-01260-5.
[8] F. Wen, C. Mao, and G. Zhang, "Direction finding in MIMO radar with large antenna arrays and nonorthogonal waveforms," *Digit. Signal Process.*, vol. 94, pp. 75–83, Nov. 2019.
[9] F. Wen, X. Xiong, and Z. Zhang, "Angle and mutual coupling estimation in bistatic MIMO radar based on PARAFAC decomposition," *Digit. Signal Process.*, vol. 65, pp. 1–10, Jun. 2017.
[10] R. Schroer, "Electronic warfare," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 18, no. 7, pp. 49–54, Jul. 2003.
[11] 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.
[12] H. Liao, Z. Zhou, X. Zhao, L. Zhang, S. Mumtaz, A. Jolfaei, S. H. Ahmed, and A. K. Bashir, "Learning-based context-aware resource allocation for edge computing-empowered industrial IoT," *IEEE Internet Things J.*, to be published, doi: 10.1109/JIOT.2019.2963371.
[13] Z. Zhou, X. Chen, and B. Gu, "Multi-scale dynamic allocation of licensed and unlicensed spectrum in software-defined HetNets," *IEEE Netw.*, vol. 33, no. 4, pp. 9–15, Jul./Aug. 2019.
[14] Z. Zhou, Y. Guo, Y. He, X. Zhao, and W. M. Bazzi, "Access control and resource allocation for M2M communications in industrial automation," *IEEE Trans Ind. Informat.*, vol. 15, no. 5, pp. 3093–3103, May 2019.
[15] G. Gui, H. Sari, and E. Biglieri, "A new definition of fairness for non-orthogonal multiple access," *IEEE Commun. Lett.*, vol. 23, no. 7, pp. 1267–1271, Jul. 2019.
[16] Z. Zhou, J. Feng, Z. Chang, and X. Shen, "Energy-efficient edge computing service provisioning for vehicular networks: A consensus ADMM approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 5087–5099, May 2019.
[17] F. Wen, Z. Zhang, and G. Zhang, "Joint DOD and DOA estimation for bistatic MIMO radar: A covariance trilinear decomposition perspective," *IEEE Access*, vol. 7, pp. 53273–53283, Apr. 2019.
[18] F. Wen, Z. Zhang, K. Wang, G. Sheng, and G. Zhang, "Angle estimation and mutual coupling self-calibration for ULA-based bistatic MIMO radar," *Signal Process.*, vol. 144, pp. 61–67, Mar. 2018.
[19] F. Wen, X. Xiong, J. Su, and Z. Zhang, "Angle estimation for bistatic MIMO radar in the presence of spatial colored noise," *Signal Process.*, vol. 134, pp. 261–267, May 2017.
[20] Z. Wu, S. Zhou, Z. Yin, B. Ma, and Z. Yang, "Robust automatic modulation classification under varying noise conditions," *IEEE Access*, vol. 5, pp. 19733–19741, 2017.
[21] H. Gacanin, "Autonomous wireless systems with artificial intelligence: A knowledge management perspective," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 51–59, Sep. 2019.
[22] H. Gacanin, E. Perenda, and R. Atawia, "Self-deployment of non-stationary wireless systems by knowledge management with artificial intelligence," *IEEE Trans. Cogn. Commun. Netw.*, vol. 5, no. 4, pp. 1004–1018, Dec. 2019.
[23] N. Kato, B. Mao, F. Tang, Y. Kawamoto, and J. Liu, "Ten challenges in advancing machine learning technologies towards 6G," *IEEE Wireless Commun. Mag.*, to be published. doi: 10.1109/MNET.001.1900476.
[24] J. Sun, J. Wang, L. Guo, J. Yang, and G. Gui, "Adaptive deep learning aided digital predistorter considering dynamic envelope," *IEEE Trans. Veh. Technol.*, to be published, doi: 10.1109/TVT.2020.2974506.
[25] Z. M. Fadlullah, F. Tang, B. Mao, N. Kato, O. Akashi, T. Inoue, and K. Mizutani, "State-of-the-art deep learning: Evolving machine intelligence toward tomorrow's intelligent network traffic control systems," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2432–2455, 4th Quart., 2017.
[26] B. Mao, Z. M. Fadlullah, F. Tang, N. Kato, O. Akashi, T. Inoue, and K. Mizutani, "Routing or computing? The paradigm shift towards intelligent computer network packet transmission based on deep learning," *IEEE Trans. Comput.*, vol. 66, no. 11, pp. 1946–1960, Nov. 2017.
[27] X. Sun, S. Ma, Y. Li, D. Wang, Z. Li, N. Wang, and G. Gui, "Enhanced echo-state restricted boltzmann machines for network traffic prediction," *IEEE Internet Things J.*, vol. 7, no. 2, pp. 1287–1297, Feb. 2020.
[28] B. Mao, F. Tang, Z. M. Fadlullah, N. Kato, O. Akashi, T. Inoue, and K. Mizutani, "A novel non-supervised deep-learning-based network traffic control method for software defined wireless networks," *IEEE Wireless Commun.*, vol. 25, no. 4, pp. 74–81, Aug. 2018.
[29] F. Tang, Z. M. Fadlullah, B. Mao, and N. Kato, "An intelligent traffic load prediction-based adaptive channel assignment algorithm in SDN-IoT: A deep learning approach," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 5141–5154, Dec. 2018.
[30] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8440–8450, Sep. 2018.
[31] H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, "Deep-learning-based millimeter-wave massive MIMO for hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3027–3032, Mar. 2019.
[32] J. Wang, X. Zhang, Q. Gao, H. Yue, and H. Wang, "Device-free wireless localization and activity recognition: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6258–6267, Jul. 2017.
[33] H. Huang, S. Guo, G. Gui, Z. Yang, J. Zhang, H. Sari, and F. Adachi, "Deep learning for physical-layer 5G wireless techniques: Opportunities, challenges and solutions," *IEEE Wireless Commun.*, vol. 27, no. 1, pp. 214–222, Feb. 2020.
[34] J. Sun, W. Shi, Z. Yang, J. Yang, and G. Gui, "Behavioral modeling and linearization of wideband RF power amplifiers using BiLSTM networks for 5G wireless systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10348–10356, Nov. 2019.
[35] G. Gui, F. Liu, J. Sun, J. Yang, Z. Zhou, and D. Zhao, "Flight delay prediction based on aviation big data and machine learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 140–150, Jan. 2020.
[36] X. Sun, G. Gui, Y. Li, R. P. Liu, and Y. An, "ResInNet: A novel deep neural network with feature reuse for Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 1, pp. 679–691, Feb. 2019.
[37] M. Liu, J. Yang, and G. Gui, "DSF-NOMA: UAV-assisted emergency communication technology in a heterogeneous Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5508–5519, Jun. 2019, doi: 10.1109/JIOT.2019.2903165.
[38] F. Tang, B. Mao, Z. M. Fadlullah, and N. Kato, "On a novel deep-learning-based intelligent partially overlapping channel assignment in SDN-IoT," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 80–86, Sep. 2018.
[39] W. Kang and D. Kim, "Poster abstract: DeepRT: A predictable deep learning inference framework for IoT devices," in *Proc. IEEE/ACM 3rd Int. Conf. Internet-Things Design Implement. (IoTDI)*, Apr. 2018, pp. 279–280.
[40] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8549–8560, Sep. 2018.

[41] H. Huang, Y. Peng, J. Yang, W. Xia, and G. Gui, "Fast beamforming design via deep learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 1065–1069, Jan. 2020.
[42] M. Liu, T. Song, and G. Gui, "Deep cognitive perspective: Resource allocation for NOMA-based heterogeneous IoT with imperfect SIC," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2885–2894, Apr. 2019.
[43] M. Liu, T. Song, J. Hu, J. Yang, and G. Gui, "Deep learning-inspired message passing algorithm for efficient resource allocation in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 641–653, Jan. 2019.
[44] 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.
[45] B. Tang, Y. Tu, Z. Zhang, and Y. Lin, "Digital signal modulation classification with data augmentation using generative adversarial nets in cognitive radio networks," *IEEE Access*, vol. 6, pp. 15713–15722, 2018.
[46] Y. Tu and Y. Lin, "Deep neural network compression technique towards efficient digital signal modulation recognition in edge device," *IEEE Access*, vol. 7, pp. 58113–58119, 2019, doi: 10.1109/ACCESS.2019.2913945.
[47] Y. Lin, Y. Tu, Z. Dou, and Z. Wu, "The application of deep learning in communication signal modulation recognition," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Oct. 2017, pp. 1–5.
[48] 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.
[49] A. Hazza, M. Shoaib, S. A. Alshebeili, and A. Fahad, "An overview of feature-based methods for digital modulation classification," in *Proc. 1st Int. Conf. Commun., Signal Process., Appl. (ICCSPA)*, Feb. 2013, pp. 1–6.

TINGPING ZHANG received the B.S., M.S., and Ph.D. degrees in computer science from Chongqing University, Chongqing, China, in 2000, 2007, and 2017, respectively. She is currently an Associate Professor with the School of Information Science and Engineering, Chongqing Jiaotong University. She has published several research articles in scholarly journals in these research areas. Her main research interests include artificial intelligence, deep learning, and wireless communications.

CONG SHUAI is currently pursuing the M.S. degree with the School of Information Science and Engineering, Chongqing Jiaotong University. His research interest includes image processing.

YARU ZHOU (Student Member, IEEE) received the B.S. degree in communication engineering from Changzhou University, in 2019. She is currently pursuing the master's degree with the Nanjing University of Posts and Telecommunications. Her research interest includes deep learning and its application in wireless communications.