Deep Learning-based Cooperative Automatic Modulation Classification Method for MIMO Systems

Yu Wang [1], Juan Wang [1], Jie Yang [1], and Wei Zhang [1]

[1]Affiliation not available

October 30, 2023

Abstract

Automatic modulation classification (AMC) is one of the most essential algorithms to identify the modulation types for the non-cooperative communication systems. Recently, it has been demonstrated that deep learning (DL)-based AMC method effectively works in the single-input single-output (SISO) systems, but DL-based AMC method is scarcely explored in the multiple-input multiple-output (MIMO) systems. In this correspondence, we propose a convolutional neural network (CNN)-based cooperative AMC (Co-AMC) method for the MIMO systems, where the receiver equipped with multiple antennas cooperatively recognizes the modulation types. Specifically, each received antenna gives their recognition sub-results via the CNN, respectively. Then, the decision maker identifies the modulation types with the recognition sub-results and cooperative decision rules, such as direct voting (DV), weighty voting (WV), direct averaging (DA) and weighty averaging (WA). The simulation results demonstrate that the Co-AMC method, based on the CNN and WA, has the highest correct classification probability in the four cooperative decision rules. In addition, the CNN-based Co-AMC method also performs better than the high order cumulants (HOC)-based traditional AMC methods, which shows the effective feature extraction and powerful classification capabilities of the CNN.

Deep Learning-based Cooperative Automatic Modulation Classification Method for MIMO Systems

Yu Wang, *Student Member, IEEE*, Juan Wang, *Student Member, IEEE*, Wei Zhang, *Member, IEEE*,
Jie Yang, *Member, IEEE*, and Guan Gui, *Senior Member, IEEE*

Abstract—Automatic modulation classification (AMC) is one of the most essential algorithms to identify the modulation types for the non-cooperative communication systems. Recently, it has been demonstrated that deep learning (DL)-based AMC method effectively works in the single-input single-output (SISO) systems, but DL-based AMC method is scarcely explored in the multiple-input multiple-output (MIMO) systems. In this correspondence, we propose a convolutional neural network (CNN)-based cooperative AMC (Co-AMC) method for the MIMO systems, where the receiver equipped with multiple antennas cooperatively recognizes the modulation types. Specifically, each received antenna gives their recognition sub-results via the CNN, respectively. Then, the decision maker identifies the modulation types with the recognition sub-results and cooperative decision rules, such as direct voting (DV), weighty voting (WV), direct averaging (DA) and weighty averaging (WA). The simulation results demonstrate that the Co-AMC method, based on the CNN and WA, has the highest correct classification probability in the four cooperative decision rules. In addition, the CNN-based Co-AMC method also performs better than the high order cumulants (HOC)-based traditional AMC methods, which shows the effective feature extraction and powerful classification capabilities of the CNN.

Index Terms—Automatic modulation classification, multiple-input multiple-output (MIMO), deep learning (DL), convolutional neural network (CNN), cooperative decision.

I. INTRODUCTION

Automatic modulation classification (AMC) is one the most critical technologies in both non-cooperative communication systems and cognitive radio (CR)-aided systems [1]–[3]. AMC can strengthen the cognition capabilities of the communication systems via identifying modulation types of unknown signals, and it is widely applied into the military and civilian domains [1]. Various AMC methods have been proposed for single-input single-output (SISO) systems [4]–[10] and multiple-input multiple-output (MIMO) systems [11]–[14]. These AMC

Y. Wang, J. Wang, J. Yang and G. Gui are with the College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China (E-mails: {1018010407, 1219012920, jyang, guiguan}@njupt.edu.cn).

W. Zhang is with the School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China (E-mail: zhang-w@njupt.edu.cn).

methods can be classified into two categories, which are based on decision theoretic and pattern recognition [11], respectively. Decision theoretic-based AMC methods rely on the perfect channel statement information (CSI) and the precise noise variance [15].

However, pattern recognition-based AMC methods can get rid of the restrictions of the prior information. These AMC methods can be divided into two sub-methods of the feature extraction and classification. In these methods, the common artificial features includes high order statistics (HOS), cyclic feature [9] and so on. The HOS is the one of the most widely applied features for both SISO and MIMO systems [10]–[14], because the HOS with the k-th ($k \geq 3$) order can effectively suppress additive white Gaussian noise (AWGN). In addition, support vector machine (SVM), decision tree (DT) and traditional artificial neural network (ANN) are generally adopted as the classifier [11]–[14].

In recent years, deep learning (DL) has been combined with various communication technologies, such as network traffic control [16]–[18], non-orthogonal multiple access (NOMA) [19], [20], beamforming [21] and AMC [2], [4], due to the powerful classification and prediction capabilities of DL [22]–[25]. However, DL-based AMC methods are designed primarily for SISO systems. In [4], T. OShea and J. Hoydis firstly proposed a CNN-based AMC methods for SISO systems and achieve the far beyond performances than the traditional AMC methods. B. Tang, *et al.* in [5] proposed a novel semi-supervised AMC method via generative adversarial network (GAN) and constellation diagrams for the single-antenna point-to-point communication. What's more, S. Hu *et al.* [6] proposed a stacked long short-term memory (LSTM)-based AMC method for the wireless communication systems with non-Gaussian noise, and achieved greater performance than the some previously proposed AMC methods.

There are few researches introducing deep neuron network, such as CNN or LSTM, for the AMC in MIMO systems. In this paper, the CNN and cooperative decision rules are proposed for the cooperative AMC (Co-AMC) method in the MIMO systems, different from [11], [12], where AMC methods are based on HOC and ANN. In detail, the CNN is trained on the dataset from all of the received antennas, and then the CNN gives the recognition sub-results, based on each received antennas. Combining the recognition sub-results together, the decision maker identifies the modulation types via the cooperative decision rules. Our simulation results

demonstrate that the CNN-based Co-AMC method performs better than the HOC and ANN-based AMC method. In addition, we also compare different cooperative decision rules of direct voting (DV), weighty voting (WV), direct averaging (DA) and weighty averaging (WA), and the WA-based Co-AMC method shows the best classification performances in the four cooperative decision rules.

II. MIMO SIGNAL MODEL, DATASET AND AMC DESCRIPTION

A. MIMO Signal Model

Here, a typical MIMO system is considered with N_t transmitting antennas and N_r receiving antennas ($N_r \geq N_t$). Assuming that the MIMO channel is a flat fading and time-invariant channel, and the sampling in the receiver strictly follows Nyquist sampling theorem without any carrier frequency offset and phase offset, the received symbol vector in the k-th observation moment can be written as follows,

$$\boldsymbol{y}_k = \boldsymbol{H}\boldsymbol{x}_k + \boldsymbol{n}_k, \tag{1}$$

where $\boldsymbol{y}_k = [y_k(1), y_k(2), \cdots, y_k(N_r)]^T$ represents the received baseband symbol vector with dimension $N_r \times 1$; $\boldsymbol{H}$ is the MIMO channel and it obeys circular symmetric complex normal distribution with zero mean and unit variance, i. e. $\boldsymbol{H} \sim \mathcal{CN}(0, \boldsymbol{I}_{N_r \times N_t})$; $\boldsymbol{x}_k = [x_k(1), x_k(2), \cdots, x_k(N_t)]^T$ is the ($N_t \times 1$) baseband modulation symbol vector.
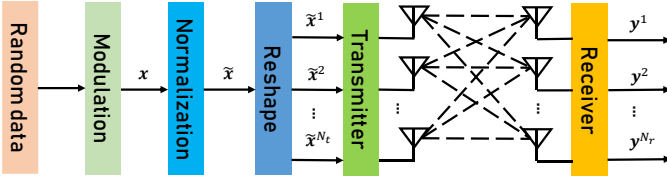

Fig. 1. The process of dataset generation which includes four steps in the transmitter: random data generation, modulation, normalization, and reshape.

B. Dataset Generation

The specified dataset generation is shown in Fig. 1. We generate a random data sequence for modulation, and then the modulation complex-value baseband signal $\boldsymbol{x}$, whose dimension is $1 \times N$, is normalized to $\widetilde{\boldsymbol{x}}$ with unit power, in order to fairly distinguish from the signals with different modulation types. Next, $\widetilde{\boldsymbol{x}}$ is reshaped into $[\widetilde{\boldsymbol{x}}^1, \widetilde{\boldsymbol{x}}^2, \cdots, \widetilde{\boldsymbol{x}}^{N_t}]^T$, with dimension $N_t \times N/N_t$, and $\widetilde{\boldsymbol{x}}^i = [x_1(i), x_2(i), \cdots, x_{N/N_t}(i)]^T, i \in [1, N_t]$. It is noted that $\widetilde{\boldsymbol{x}}_i$ represents the N/N_t continuous transmitted symbols at the i-th transmitter antenna. When the transmitted symbols pass through the MIMO channel, the received complex-value baseband signal at the j-th transmitter antenna can be presented by $\boldsymbol{y}^j = [y_1(j), y_2(j), \cdots, y_{N/N_t}(j)]^T, j \in [1, N_r]$. Then, we need to extract the real part and imaginary part of $\boldsymbol{y}^j$ and combine them together into a matrix with dimension $N/N_t \times 2$, which is the sample in the dataset for the CNN.

TABLE I
THE STRUCTURE OF CNN, INCLUDING FIVE LAYERS: TWO
CONVOLUTIONAL LAYERS AND THREE FULLY-CONNECTED LAYERS.

Type	Structure
Convolutional layer	Conv1D (128, 8) + BN + ReLU + Dropout (0.5)
Convolutional layer	Conv1D (64, 4) + BN + ReLU + Dropout (0.5)
Fully-connected layer	Dense (256) + BN + ReLU + Dropout (0.5)
Fully-connected layer	Dense (128) + BN + ReLU + Dropout (0.5)
Fully-connected layer	Dense ($|\mathcal{M}|$) + Softmax

C. AMC Description

AMC is modeled as a typical close-set classification problem under maximum a posteriori (MAP) criterion and a modulation candidate pool with the limited modulation types. When receiving the $\boldsymbol{y}^j$ at the j-th antenna, the AMC can be described as follows,

$$\hat{m}_n^j = \arg\max_{n \in [1, |\mathcal{M}|]} P(m_n|\boldsymbol{y}^j), j \in [1, N_r], \tag{2}$$

where $\hat{m}_n^j$ represents the predicted modulation type under the received signal at the j-th antenna; m_n and $\mathcal{M} = \{m_n\}_{n=1}^{|\mathcal{M}|}$ are the real modulation type and the modulation candidate pool, respectively, where $|\mathcal{M}|$ is the number of modulation types in $\mathcal{M}$; $\boldsymbol{P}^j = \{P(m_n|\boldsymbol{y}^j)\}_{n=1}^{|\mathcal{M}|}, j \in [1, N_r]$ is the probability distribution function (PDF) with the inputting of $\boldsymbol{y}^j$, and it is also the output of the neural network, when inputting $\boldsymbol{y}^j$. In this paper, we consider a classical modulation candidate pool $\mathcal{M}$ = {BPSK, QPSK, 8PSK, 16QAM} [1], [12].

III. THE PROPOSED CNN-BASED CO-AMC METHOD

A. CNN-based Co-AMC Method

The structure of the CNN-based Co-AMC method is shown in Fig. 2. The CNN-based Co-AMC method contains two part: the CNN and the decision maker. After the received signals in each antenna are fed into the CNN in turn, the decision maker cooperatively gives the final predicted modulation type on the basis of the predicted PDFs $\{\boldsymbol{P}^j\}_{j=1}^{N_r}$, which are given by CNN.
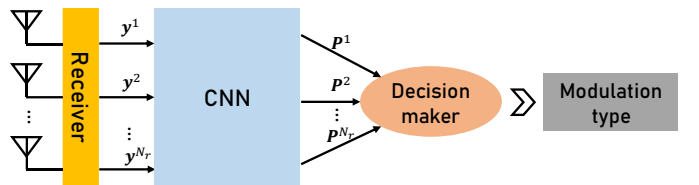

Fig. 2. The structure of CNN-based Co-AMC method, containing CNN and a cooperative decision maker.

1) CNN structure and loss function: In this paper, a five-layer CNN model is considered with two convolutional layers and three fully-connected layers. The structure of CNN is depicted in Tab. I, where "Conv1D" is a typical convolutional layer and "Dense" is a general fully-connected layer. What's more, batch normalization (BN) is applied to accelerate the training, and BN and dropout can also prevent overfitting.

2) Training and loss function: The centralized training is applied, which means that only one CNN is trained, based on the mixed received signals from different received antennas, rather than multiple CNNs are trained for the multiple corresponding received antennas. Based on the training dataset $\{s_i, l_i\}_{i=1}^{S}$, which contains S training samples with their corresponding one-hot coded-labels, the loss function can be described as

$$\mathcal{L}(f_{CNN}, \boldsymbol{\theta}; \{s_i, l_i\}_{i=1}^{S}) = -\frac{1}{S}\sum_{i=1}^{S} l_i log(f_{CNN}(\boldsymbol{\theta}; s_i)) \\ + \lambda J(f_{CNN}, \boldsymbol{\theta}), \quad (3)$$

where f_{CNN} and $\boldsymbol{\theta}$ are the mapping function and parameters of the CNN, respectively, and $f_{CNN}(\boldsymbol{\theta}; \cdot)$ is actually $\{P(m_n|\cdot)\}_{n=1}^{|\mathcal{M}|}$; The first term of the final loss function is a typical empirical loss function for the classification problem; The second term is just the structural loss function to constrain the model complexity, and BN and dropout can be considered as the hidden structural loss function. In addition, λ is to balance the two loss functions. Adaptive moment estimation (Adam) as the optimizer is to minimize the loss function.

3) Test and various cooperative decision rules: Test is divided into two phases: The trained CNN outputs the N_r PDFs, based on the test samples from N_r received antennas, and then these PDFs $\{\boldsymbol{P}^j\}_{j=1}^{N_r}$ are fed into the decision maker, relying on the cooperative decision rules, to cooperatively decide the modulation type.

Here, we introduce two kinds of the cooperative decision rules: voting method and averaging method. Voting method is based on the decided modulation types, given by each antenna, and obeys the rule of that the final decided modulation type is the modulation type with the majority voting, while the averaging method is to calculate the average of PDFs achieved by all of the receiving antennas, and obeys the rule of that the modulation type with the highest probability is the final predicted modulation type.

In addition, each kind method is divided into two sub-kinds: direct method and weighty method. In the direct method, each antenna is fair, and the decision at each antenna is equally important, but in the weighty method, each antenna has a weight, and the weights are generally different. In this paper, we apply the the normalized validation accuracy of each antenna as each received antenna's weight.

In detail, we test the trained CNN on the validation dataset, and then calculate the each antenna's accuracy based on their corresponding validation dataset, which is denoted as acc^j, $j \in [1, N_r]$. Thus, the normalized validation accuracy $\widetilde{acc}^j = \frac{acc^j}{\sum_{i=1}^{N_r} acc^i}$, $j \in [1, N_r]$ is set as the corresponding weight of each received antenna W^j, $j \in [1, N_r]$.

The specified descriptions of these methods are shown as follows, where the direct voting (DV) method and the weighty voting (WV) are listed in Algorithm 1, and the direct averaging (DA) method and the weighty averaging (WA) are given in Algorithm 2.

Algorithm 1 The Co-AMC based on the voting method.

Input: Test sample $\{\boldsymbol{y}^j\}_{j=1}^{N_r}$ and the trained CNN;
Output: The predicted modulation type;
1: for $j = 1 : N_r$:
    Give the $\hat{m}_n^j$ by Eq. (2);
    end
2: if choosing direct voting (DV) method,
    $\hat{\boldsymbol{M}} = \frac{1}{N_r}\sum_{i=1}^{N_r} \hat{m}_n^j$;
    end
3: if choosing weighty voting (WV) method,
    Calculate the validation accuracy at each receiver antenna acc^j;
    Set $W^j = \frac{acc^j}{\sum_{i=1}^{N_r} acc^i}$;
    $\hat{\boldsymbol{M}} = \sum_{j=1}^{N_r} W^j \cdot \hat{m}_n^j$;
    end
4: $\hat{m}_n^{Voting} = \arg\max_{n \in [1, |\mathcal{M}|]} \hat{\boldsymbol{M}}(n)$;
5: **return** The predicted modulation type $\hat{m}_n^{Voting}$.

Algorithm 2 The Co-AMC based on the averaging method.

Input: Test sample $\{\boldsymbol{y}^j\}_{j=1}^{N_r}$ and the trained CNN;
Output: The predicted modulation type;
1: for $j = 1 : N_r$:
    Give the $\boldsymbol{P}^j = [P(m_1|\boldsymbol{y}_j), P(m_2|\boldsymbol{y}_j), ..., P(m_{|\mathcal{M}|}|\boldsymbol{y}_j)]$;
    end
2: if choosing direct average (DA) method,
    $\hat{\boldsymbol{P}} = \frac{1}{N_r}\sum_{j=1}^{N_r} \boldsymbol{P}^j$;
    end
3: if choosing weighty average (WA) method,
    Calculate the validation accuracy at each receiver antenna acc^j;
    Set $W^j = \frac{acc^j}{\sum_{i=1}^{N_r} acc^i}$;
    $\hat{\boldsymbol{P}} = \sum_{j=1}^{N_r} W^j \cdot \boldsymbol{P}^j$;
    end
4: $\hat{m}_n^{Averaging} = \arg\max_{n \in [1, |\mathcal{M}|]} \hat{\boldsymbol{P}}(n)$;
5: **return** The predicted modulation type $\hat{m}_n^{Averaging}$.

B. Review of Traditional AMC Method

A traditional AMC method, based on high order cumulants (HOC) and ANN [11], [12], is introduced for the comparison of the CNN-based Co-AMC method, and its structure is shown in Fig. 3. In this paper, the HOC of fourth order, $C_{4k}, k \in \{0, 1, 2\}$, is considered, and its theoretical values are shown in Tab. II. In addition, ANN is chosen as the classifier. The traditional method also applies the centralized training and cooperative decision, which are same with the CNN-based Co-AMC method. For the fair comparison with the CNN-based Co-AMC method, the ANN structure has three fully-connected layers with the same parameters with the three fully-connected layers in the CNN.

TABLE II
THE THEORETICAL VALUES OF THE FOURTH ORDER HOC [11], [12].

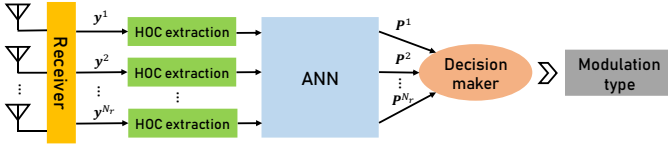| C_{4k} \ Type	BPSK	QPSK	8SPK	16QAM
C_{40}	-2	-1	0	-0.68
C_{41}	-2	0	0	0
C_{42}	-2	-1	-1	-0.68

Fig. 3.   The structure of HOC and ANN-based traditional AMC method.

IV. SIMULATION RESULTS

Both CNN in the Co-AMC method and ANN in the traditional AMC method are based on Keras and are trained on the GPU. In addition, the simulation of MIMO system, MIMO signal and HOC extraction are relied on Matlab. We prepare 20000 samples per SNR per type for training, which are divided into two parts for training and validation by the ratio of 7:3, respectively, and 10000 samples per SNR per type for test. In addition, There are two classification performance metrics of the correct classification probability P_{cc}^{snr} at snr dB, where $snr \in [-10, 10]$ dB and the average correct classification probability P_{cc}^{ave}. Hence they are defined as follows,

$$P_{cc}^{snr} = \frac{S_{correct}^{snr}}{S_{test} \cdot |\mathcal{M}|} \times 100\%, \tag{4}$$

$$P_{cc}^{ave} = \frac{\sum_{snr=-10}^{10} S_{correct}^{snr}}{S_{test} \cdot |\mathcal{M}| \cdot N_{snr}} \times 100\%, \tag{5}$$

where S_{test} is the number of test samples of each type; $S_{correct}^{snr}$ is the samples classified correctly at snr dB; N_{snr} are the number of SNR for test, and it is equal to eleven in this paper.

A. Performance Comparison vs. Cooperative Rules

The performances of different cooperative rules-based Co-AMC methods are shown in Fig. 4, where "AMC" is the average correct correct classification probability of four received antennas, and other other Co-AMC methods are based on four cooperative decision rules, respectively. It can be observed that the Co-AMC methods have higher classification performance than AMC method without cooperative decision rules. What' more, the averaging method is always better than the voting method, and the weighty method performs better than the direct method, which can be also demonstrated with P_{cc}^{ave} in Tab. III.

In this paper, the weights in the weighty methods are decided by the each validation accuracy of each antenna, and the weight can describe the quality for the modulation classification. It means that the high weights will be assigned the antenna with the high validation accuracy, while the low weights will be given to the antenna with the low validation accuracy, which is different from the same-weight-assignment method in the direct methods. Thus, the weighty methods has the better performance.

Besides, it is noted that we consider that the number of the received antennas is four, which means that decision maker has four voter in the voting methods, but the number of modulation types in $\mathcal{M}$ is also four. Thus, it is possible that multiple modulation types have the maximum votes at the same time in the DV method, and we have to randomly choose a modulation type as the predicted modulation type.
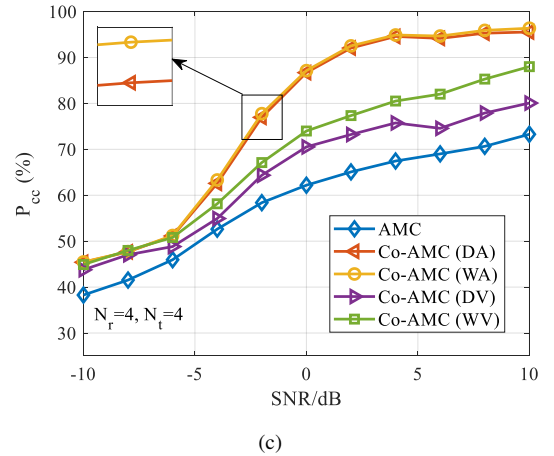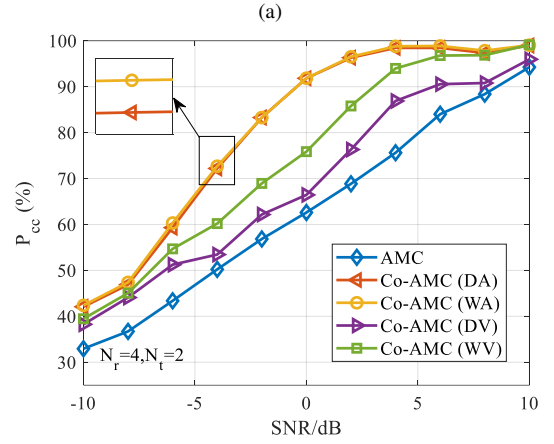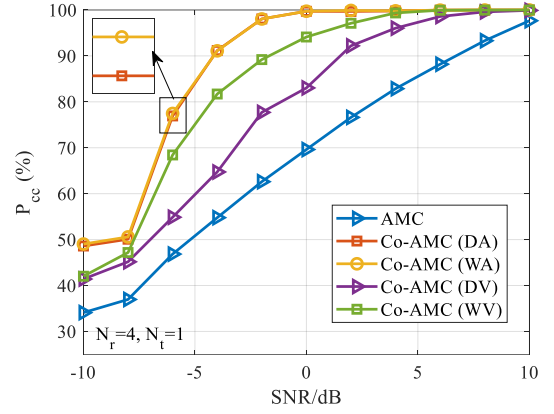

(a)

(b)

(c)

Fig. 4.   The classification performances of CNN-based Co-AMC method under different cooperative decision rules. (a) $N_r = 4$, $N_t = 1$; (b) $N_r = 4$, $N_t = 2$; (c) $N_r = 4$, $N_t = 4$;

B. Comparing with Traditional AMC Method

The WA-based Co-AMC method has the best performance in four Co-AMC methods. Thus, we also consider to apply the WA method into the HOC and ANN-based traditional AMC method. The classification performances of Co-AMC method and traditional AMC method under different transmitted antennas and the same received antennas are shown in Fig. 5. Compared with the HOC and ANN-based traditional AMC methods, the CNN-based Co-AMC methods have the better classification performances under the different kinds of MIMO

TABLE III
THE AVERAGE CORRECT CLASSIFICATION PROBABILITY OF THE CO-AMC
METHOD UNDER VARIOUS COOPERATIVE DECISION RULES.

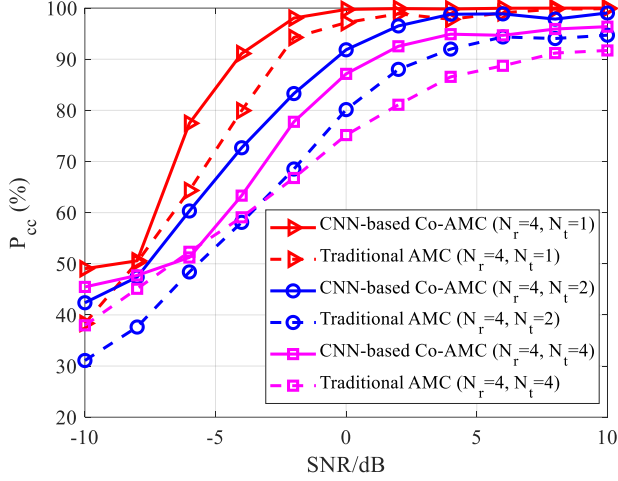| P_{cc}^{ave} (%) $\quad$ (N_r, N_t) Method	(4, 1)	(4, 2)	(4, 4)
AMC	67.59	63.07	58.58
Co-AMC (DV)	77.57	68.74	64.63
Co-AMC (WV)	83.55	74.25	68.74
Co-AMC (DA)	87.63	80.46	76.56
Co-AMC (WA)	87.79	80.82	77.00
Traditional AMC (WA)	83.62	71.54	70.41

Fig. 5.   The comparison of the correct classification performances of CNN-based Co-AMC method and HOC and ANN-based traditional AMC method. These two methods are based on the decision maker with the WV method.

antennas. It is demonstrated that the CNN can effectively modulation features for classification than the manmade features, such as HOC features.

V. CONCLUSION

In this correspondence, we propose the CNN-based Co-AMC methods for MIMO systems. Four cooperative decision rules of direct voting (DV), weighty voting (WV), direct averaging (DA), and weighty averaging (WA) methods are applied for the multiple received antennas to cooperatively decide the final modulation type. The simulation results demonstrate that the voting methods has weaker performances than the averaging methods in the CNN-based Co-AMC method. Besides, the weighty methods, where the weights are decided by the validation accuracy of each antenna, is better than the direct methods, which assign the same weights for each antenna. What' more, the CNN-based Co-AMC method performs better than the HOC and ANN-based traditional AMC methods under the condition of the same cooperative rule. The results shows that the CNN has the capability to extract the more effective features for the modulation classification than the traditional artificial feature designing methods.

REFERENCES

[1] 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, 2018.

[2] 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, 2019.

[3] M. Liu, *et al.*, "Deep learning-inspired message passing algorithm for efficient resource allocation in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 641–653, Jan. 2019.

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

[5] B. Tang, *et al.*, "Digital signal modulation classification with data augmentation using generative adversarial nets in cognitive radio networks," *IEEE Access*, vol. 6, pp. 15713–15722, 2018.

[6] S. Hu, Y. Pei, P. Lian and Y. Liang, "Deep neural network for robust modulation classification under uncertain noise conditions," *IEEE Trans. Veh. Technol.*, doi: 10.1109/TVT.2019.2951594.

[7] Y. Tu, Y. Lin, J. Wang and J. Kim, "Semi-supervised learning with generative adversarial networks on digital signal modulation classification,"*Comput. Mater. Continua*, vol. 55, no. 2, pp. 243-254, May. 2018.

[8] Y. Tu, and Y. Lin, "Deep neural network compression technique towards efficient digital signal modulation recognition in edge device," *IEEE Access*, vol. 7, no. 1, pp. 58113–58119, 2019.

[9] B. Ramkumar, "Automatic modulation classification for cognitive radios using cyclic feature detection," *IEEE Circuits Syst. Mag.*, vol. 9, no. 2, pp. 27–45, 2009.

[10] A. Swami, and B. M. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 416–429, 2000.

[11] K. Hassan, I. Dayoub, W. Hamouda, C. Nzeza and M. Berbineau, "Blind digital modulation identification for spatially-correlated MIMO systems," *IEEE Trans. Wirel. Commun.*, vol. 11, no. 2, pp. 683–693, 2011.

[12] M. Abdelbar, *et al.*, "Cooperative modulation classification of multiple signals in cognitive radio networks," in *IEEE ICC*, Sydney, Australia, June 10–14, 2014, pp. 1483–1488.

[13] S. Kharbech, *et al.*, "Blind digital modulation identification for time-selective MIMO channels," *IEEE Wirel. Commun. Lett.*, vol. 3, no. 4, pp. 373–376, 2014.

[14] D. Das, P. K. Bora and R. Bhattacharjee, "Blind modulation recognition of the lower order PSK signals under the MIMO keyhole channel," *IEEE Commun. Lett.*, vol. 22, no. 9, pp. 1834–1837, 2018.

[15] F. Hameed, *et al.*, "On the likelihood-based approach to modulation classification," *IEEE Trans. Wirel. Commun.*, vol. 8, no. 12, pp. 5884–5892, 2009.

[16] N. Kato, *et al.*, "The deep learning vision for heterogeneous network traffic control: proposal, challenges, and future perspective," *IEEE Wirel. Commun. Mag.*, vol. 24, no. 3, pp. 146–153, 2016.

[17] Z. Md. Fadlullah, *et al.*, "State-of-the-art deep learning: evolving machine intelligence toward tomorrow's intelligent network traffic control systems," *IEEE Commun. Surveys and Tuts.*, vol. 19, no. 4, pp. 2432–2455, 2017.

[18] L. Zhao, J. Wang, J. Liu and N. Kato, "Routing for crowd management in smart cities: a deep reinforcement learning perspective," *IEEE Commun. Mag.*, vol. 57, no. 4, pp. 88–93, 2019.

[19] G. Gui, *et al.*, "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8440–8450, 2018.

[20] H. Huang, *et al.*, "Deep learning for physical-Layer 5G wireless techniques: opportunities, challenges and solutions," *IEEE Wirel. Commun.*, doi: 10.1109/MWC.2019.1900027.

[21] H. Huang, *et al.*, "Fast beamforming design via deep learning," *IEEE Trans. Veh. Technol.*, doi: 10.1109/TVT.2019.2949122.

[22] 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.*, doi: 10.1109/TVT.2019.2954094.

[23] Z. Md. Fadlullah, F. Tang, B. Mao, J. Liu and N. Kato, "On intelligent traffic control for large-scale heterogeneous networks: a value matrix-based deep learning approach," *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2479–2482, 2018.

[24] F. Tang, Z. Md. 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, 2018.

[25] Z. Md. Fadlullah, B. Mao, F. Tang and N. Kato, "Value iteration architecture based deep learning for intelligent routing exploiting heterogeneous computing platforms," *IEEE Trans. Computers*, vol. 68, no. 6, pp. 939–950, 2019.