IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 8, 2019, accepted June 1, 2019, date of publication June 11, 2019, date of current version July 1, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2921988

Automatic Modulation Classification Using Compressive Convolutional Neural Network

SAI HUANG [1], (Member, IEEE), LU CHAI [1], ZENING LI [2], DI ZHANG [3], (Member, IEEE), YUANYUAN YAO [4], (Member, IEEE), YIFAN ZHANG [1], AND ZHIYONG FENG [1], (Senior Member, IEEE)

[1] Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2] Daniel Felix Ritchie School of Engineering and Computer Science, University of Denver, Colorado 80208, USA
[3] School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China
[4] School of Information and Communication Engineering, Beijing Information Science and Technology University, Beijing 100101, China

Corresponding author: Zhiyong Feng (fengzy@bupt.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFF0301202, in part by the National Natural Science Foundation of China under Grant 61801052, Grant 61525101, and Grant 61227801, in part by the Supplementary and Supportive Project for Teachers at Beijing Information Science and Technology University under Grant 5029011103, and in part by the Key Research and Cultivation Project at Beijing Information Science and Technology University under Grant 5211910926.

ABSTRACT The deep convolutional neural network has strong representative ability, which can learn latent information repeatedly from signal samples and improve the accuracy of automatic modulation classification (AMC). In this paper, a novel compressive convolutional neural network (CCNN) is proposed for AMC, where different constellation images, i.e., regular constellation images (RCs) and contrast enhanced grid constellation images (CGCs), are generated as network inputs from received signals. Moreover, a compressive loss constraint is proposed to train the CCNN, which aims at capturing high-dimensional features for modulation classification. Additionally, CCNN utilizes intra-class compactness and inter-class separability to enhance the classification and robustness performance for the different orders of modulations. The simulation results demonstrate that CCNN displays superior classification and robustness performance than existing AMC methods.

INDEX TERMS Automatic modulation classification, compressive loss constraint, deep convolutional neural network, multiple constellation images.

I. INTRODUCTION

With the rapid development of wireless communications and their corresponding services in recent years, various communications standards and technologies such as the fifth generation (5G) or dynamic spectrum access emerge [1]. Consequently, the radio environment is becoming increasingly disordered, making it more difficult to enforce the spectrum monitoring, which ensures the normal operation of wireless services and strengthens spectrum management. Automatic modulation classification (AMC) can improve such situations in spectrum monitoring [2] as it enables the supervision of interference signals via identifying the modulation format of the received signal in a fading channel. Moreover, it can be applied in many military and civilian applications, such as the identifying of interference signals and jammers, and the bridge between signal detection and demodulation [3].

Generally, there are two main methodologies accounting for AMC, i.e., likelihood-based (LB) methods and feature-based (FB) methods [4]. In LB, the likelihood functions of different hypotheses are firstly calculated using received signal symbols, and then compared to make final classification decisions. The LB is viewed as the optimal classifier in performance as it maximizes the probability of correct classification when both channel models and noises are given [5]. However, unknown parameters should be addressed in actual non-cooperative scenarios. Average likelihood ratio test (ALRT) calculates the integral of the probability density function (PDF) of unknown parameters to deal

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

with the unknown parameter problem [6]. However, ALRT suffers from inaccuracy and high computational complexity when considering model mismatches or increased unknown parameters. Generalized likelihood ratio test (GLRT) utilizes a maximum likelihood estimator to estimate the unknown parameters [7], which performs lower computational complexity, but this biased classifier results in performance degradation of nested modulations. Hybrid likelihood ratio test (HLRT) calculates the likelihood of each signal sample belonging to each alphabet symbol and then makes decisions by the maximum likelihood principle [8]. In a word, LB classifiers achieve optimal performance in a Bayesian sense, but the high computational complexity and the weak robustness to model mismatches limit the real-time implementation with low cost in practice.

Compared with LB classifiers, FB methods extract signal features to identify the modulation formats with lower computational complexity. Moreover, FB methods are widely utilized in practice due to the strong robustness to model mismatches [9]. There are some widely utilized features including high-order statistics, cyclostationary spectrum, *etc.*, [10]. Swami *et al.* proposed a fourth-order cumulant-based hierarchical modulation classification framework to identify the signal formats. Better robustness to phase offset is achieved, but minimum distance criterion for decision making leads to sub-optimal classification performance [11]. Huang *et al.* proposed a cumulant-based maximum likelihood classification scheme, where the sample estimate of cumulant is utilized for classification and classification decision is made by maximizing the asymptotic distribution function of the cumulant [12]. Gardner *et al.* first implemented cyclostationary-based modulation classification method and it performs superior performance at low signal-to-noise (SNR) sense by exploiting the discrepancies on the cyclic spectrum features [13]. However, the computational complexity is high and the symbol rate of the received signal should be known. Kim *et al.* constructed a cyclic domain feature using the maximum of cyclostationary spectrum over frequency, aiming at reducing the computational complexity [14]. However, cyclostationary-based AMC is unable to deal with the classification problem of M-ary phase shift keying (M-PSK) and M-ary quadrature amplitude modulation (M-QAM). In short, FB methods conduct modulation classification with low complexity and perform good robustness to model mismatches but yield suboptimal classification performance. Moreover, since different features have their own characteristic and applicability, manual feature selection and decision thresholds that rely on the experience may largely influence classification performance. Therefore, this paper focuses on improving the classification accuracy of FB methods and avoiding experience-based feature selection.

Recently, machine learning technique can be viewed as a feasible approach for FB classification methods since it considers a versatile feature set and utilizes prior information and feature engineering to improve performance and computational efficiency [15-18]. Zhou *et al.* proposed a Gaussian support vector machine (SVM) and cumulant-based modulation classification method [19], which performs superiority in low SNR scenarios, but it is only applicable to binary classification problem. Aslam *et al.* proposed a k-nearest neighbors (KNN)-based AMC scheme, where genetic programming is utilized for feature selection [20]. It performs acceptable classification performance, but the efficiency between the improved accuracy and increased complexity needs to be refined. Moreover, the number K in KNN should be manually predefined. In [21], a random forest-based modulation classification system was built, where a variety of digital and analog modulation schemes under various SNR environments can be distinguished. However, the classification among M-PSK and M-QAM is still a challenge as well as the robustness to channel distortion and noise can be further improved. In a word, the above machine learning-based AMC methods need to be tested with new features in order to quantify the performance and re-tune their parameters accordingly, illustrating weak model generalization.

The deep learning based neural network with strong representative ability can be applied in AMC for higher classification accuracy [22-24]. In [25], the authors utilized recurrent neural network (RNN) to exploit the temporal sequence characteristics of received signals for modulation classification. However, the spatial correlation factor is not taken into account in this study. In [26], the authors proposed an unsorted deep neural network (DNN) to identify the signal modulation formats with low computational complexity, but the absence of the convolutional operation makes it difficult to extract high-dimensional features. Likewise, the authors in [27] proposed a convolutional long short-term deep natural network (CLDNN) to achieve better classification accuracy. To our best of knowledge, CLDNN is widely used in voice processing involving raw time-domain signals, but its application in wireless signals is still rare. In [28], convolutional neural network was used for AMC to extract features from baseband signal samples. However, the absence of simulations of model mismatches makes this method less convincing.

Moreover, the loss function in neural network quantifies the amount by which the prediction values deviate from the actual ones [29]. It largely affects the convergence speed of the neural network and the discriminative ability of classification. In [30], the authors proposed auxiliary loss to train the network for obtaining effective features. The study of [31] suggested calculating the contractive loss by coupling samples, which performs a prominent feature generalization, but it needs extra sample pre-selection, aggravating the computational complexity.

In this paper, we propose the multiple constellation image-based AMC method using deep compressive convolutional neural network (CCNN) to effectively implement the modulation classification. The contributions of this paper are listed as follows.

TABLE 1. Symbol table.

Symbol	Definition
h	Rayleigh fading channel coefficients
f_0	The frequency offset
θ	The phase offset
N_s	The length of signal samples
$x_r(n)$	The nth signal sample extracted from the constellation H_r
$w(n)$	The additive white Gaussian noise(AWGN)
γ	The signal-to-noise ratio
p_f	The value of pixel points
L_s	The softmax loss
L_c	The compressive loss
N	The number of training samples in a min-batch
$\boldsymbol{z}_i$	The ith training sample vector of CCNN
$y^{(i)}$	The class label of $\boldsymbol{z}_i$
$\boldsymbol{W}_j$	The weights of the neural network
$\boldsymbol{f}(\boldsymbol{z}_i)$	The normalized feature extracted in the ith class
$\boldsymbol{c}_{y^{(i)}}$	The feature center of the $y^{(i)}$ class
$\boldsymbol{c}_n$	The current class center
$\mathbf{1}\{\cdot\}$	The indicator function
$\delta(\cdot)$	The impulse response function
φ	The learning rate of class centers
S	The total number of candidate modulations
$\mathbf{M}$	The candidate modulation format set
P_{cc}	Probability of correct classification
$\|\|\cdot\|\|_2$	The Euclidean norm

1) On the basis of regular constellation images (RCs), contrast enhanced grid constellation images (CGCs) in the form of density are added as the inputs of CCNN. With RCs and CGCs, manually feature selection is avoided and density features are considered in a comprehensive way.
2) Small convolutional kernels, which can extract high-dimensional features directly for identifying signal formats, are exploited to increase the ability of learning representation and prevent over-fitting in CCNN.
3) The proposed compressive loss updates the network parameters and class centers, and enhances the discrepancy of obtained deep representations by minimizing the intra-class variations and maximizing the inter-class separability.
4) Simulation results demonstrate that CCNN performs superior classification and robustness performance than existing classification methods, such as RNN and DNN.

The remainder of this paper is organized as follows. Section II shows the signal model and signal features. In Section III, we illustrate the network structure of CCNN and the proposed compressive loss. Section V exhibits simulation results and the conclusion of this paper is given in Section VI.

The notations used in this paper are summarized in Table 1.

II. SIGNAL MODEL AND FEATURES EXTRACTION

A. SIGNAL MODEL

Assume that radio frequency (RF) signals are received by the single-antenna receiver and then transformed to baseband signals, which can be formatted as follows.

$$y(n) = he^{j(2\pi f_0 n+\theta)}x_r(n) + w(n), \quad n = 1, \ldots, N_s \qquad (1)$$

where h denotes the Rayleigh fading channel coefficient, f_0 and θ stand for the frequency and phase offset, respectively, N_s denotes the length of signal samples, $x_r(n)$ refers to the nth complex sample extracted from the candidate constellation H_r and $w(n)$ is additive white Gaussian noise (AWGN) with mean zero and variance σ_w^2. In this case, the received signal-to-noise ratio (SNR) is defined as $\gamma = |h|^2/\sigma_w^2$.

B. FEATURE EXTRACTION

In order to extend the representation dimension of RCs, CGCs that reflect the density of constellation points are added as the input of the network. In this case, multiple constellation images are composed of two parts: RCs and CGCs.

For RCs, we plot the real and imaginary parts of the received signals by transforming the signal samples into two-dimensional scatter diagrams, which can indicate detailed characteristics of the modulated signal.

For CGCs, we consider the two-dimensional probability distribution of signal samples on the basis of RCs. Since the limited pixels in one image can result in the feature deficiency in RCs, we equally cut the image blocks into grids in line with the predetermined size G. The value in each grid is defined as the ratio of the sample number located in the corresponding grid region and the total signal sample number, which is titled the grid constellation images (GCs). Then, we suitably expand the regions of pixel gray-values by increasing the contrast of GCs according to the contrast-enhanced method. Therefore, the highlighted regions in CGCs imply that signal samples are notably dense in these regions. Since the pixel region of gray-scale images is in a range of 0-255, it is easy to enhance the image contrast by gray-scale expanding. The pixel conversion rule of the contrast-enhanced method is given in Eq. 2. Furthermore, we enlarge the original interval $p_1 = 0.3, p_2 = 0.7$ to $q_1 = 0.1, q_2 = 1$, alternating each pixel into p_f.

$$p_f = \left\{ \begin{array}{cc} 255p_1 & p < p_1 \\ \dfrac{q_2 - q_1}{p_2 - p_1} \times (p - p_1) \times 255 + 255q_1 & p_1 < p < p_2 \\ 255p_2 & p > p_2 \end{array} \right\} \qquad (2)$$

A generation diagram is conducted in FIGURE 1 to illustrate the multiple constellation images of three widely-used modulation formats (i.e. Quadrature PSK (QPSK), 16-QAM, 64-QAM) at specific SNRs, where RCs and CGCs are sequenced. For different modulations under the same SNR, RCs are used to provide signal original minutiae features. For the same modulation under different SNRs, CGCs in the form of constellation density are effective at low SNR. Therefore, multiple constellation images aggregate the original constellation points and density characteristics.

III. CCNN BASED CLASSIFICATION METHOD

A. STRUCTURE OF CCNN

The CCNN structure aims at extracting deep representations and identifying the signal format. As shown in

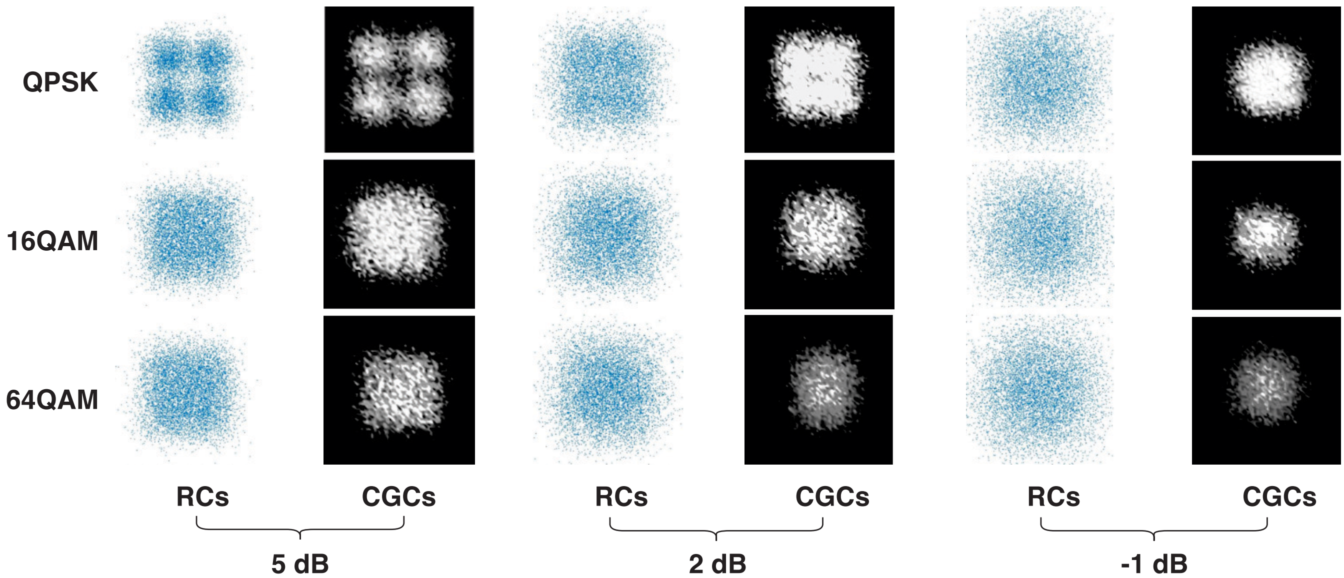

FIGURE 1. Multiple constellation images for three modulations under different SNRs.

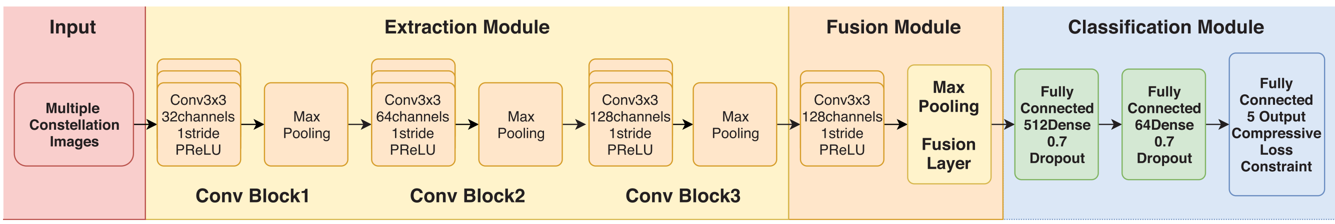

FIGURE 2. The structure of CCNN.

FIGURE 2, the proposed network consists of three modules, i.e., extraction module, fusion module and classification module. To begin with, the extraction module extracts high-dimensional features from multiple constellation images. Next, the fusion module unifies the feature vectors of different lengths. Finally, with compressive loss constraint, classification module continually classifies the signal modulation types using fused features. As multiple constellation images are diverse from real photos that are well processed by traditional neural networks, our proposed CCNN structure has its own characteristics in the following aspects. First, with the consideration of the small size of inputs, we set the stride step as one. Hence the classification module can derive enough feature dimensions. Second, we utilize 3×3 kernel size to actualize sparse connectivity among input and output units, intending to reduce computational complexity and improve network nonlinearity.

Incorporating accuracy and efficiency, the extraction module includes three blocks, where each block consists of three convolutional layers followed by a batch normalization, a parametric rectified linear unit (PReLU) activation function and a 2×2 max-pooling layer. Note that "Conv3x3-32,3" denotes 3 convolutional layers with 3×3 convolutional filters and 32 channels). 3×3 convolutional filters adopted in the first two layers bring two advantages. For one thing, continuous convolutional layers with small kernels can enhance nonlinearity and get sufficient specific receptive fields. Besides, small kernels decrease the dimensionality of network parameters, aiming at reducing the computational complexity. The padding and the stride step of the kernel are set as 1 and the stride step of the third convolutional layer is set as 3 to reduce the input spare information. The number of the last hidden layer is limited to 2, which indicates that the dimension of the deep feature is reduced to 2. This is beneficial to two-dimensional visualization. Moreover, PReLU activation function and maxpool layers are involved to introduce sparsity and enhance nonlinearity.

The fusion module, composed of the convolutional layer and the fusion layer, unifies the form of feature representations for the subsequent classification module. In order to reduce feature redundancy, fusion layer is added for down-sampling, where maxpooling is utilized in our network model. The dropout is set as 0.7 to avoid overfitting [32] and increase the robustness to model mismatches.

In order to reduce the number of output channels and improve the computational efficiency, there are 3 fully connected layers in modulation classification module, whose channel numbers are 512, 64, 5, respectively. Note that the neurons number at the output layer is equal to the modulation types.

B. COMPRESSIVE LOSS

A novel auxiliary supervision constraint named compressive loss is added to the original loss function of CNN network. CCNN aims at enhancing the discriminative ability of representations by minimizing the intra-class variations and maximizing the inter-class separability. The loss function of CCNN is given by

$$\begin{aligned} L &= L_s + \lambda L_c \\ &= L_s + \lambda_1 L_{c1} + \lambda_2 L_{c2} \end{aligned} \tag{3}$$

where the first term L_s denotes the softmax loss and the second term L_c denotes the compressive loss. The latter L_c consists of two parts: L_{c1} can compact intra-class features and L_{c_2} disperses inter-class features. The weights λ_1 and λ_2 are designed to balance the two parts. Moreover, the softmax loss function maps the output of neurons to (0,1) values, and the sum of these values is 1 [33], which is given by

$$\begin{aligned} L_s &= -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k}\mathbf{1}\left\{y^{(i)}=j\right\}\log P(y^{(i)}=j|z_i)\right] \\ &= -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k}\mathbf{1}\left\{y^{(i)}=j\right\}\log\frac{e^{W_j^T z_i}}{\sum_{l=1}^{k}e^{W_l^T z_i}}\right] \end{aligned} \tag{4}$$

where m and k are the number of samples and modulation hypotheses, respectively, z_i denotes the ith training sample, $y^{(i)}$ is the label of z_i, $P(y^{(i)}=j|z_i)$ is the probability where the label of z_i is j and $\mathbf{1}\{\cdot\}$ is the indicator function.

In order to increase the discriminative power of the deeply learned features, the first term of compressive loss is defined as follows.

$$L_{c1} = \frac{1}{2}\sum_{i=1}^{N}\left\|\boldsymbol{f}(z_i)-\boldsymbol{c}_{y^{(i)}}\right\|_2^2 \tag{5}$$

where N is the number of training samples in a min-batch, $\boldsymbol{f}(z_i)$ denotes the normalized feature extracted in the ith class and $\boldsymbol{c}_{y^{(i)}}$ denotes the $y^{(i)}$'s class center of deep features. Moreover, L_{c1} can get the center of features from each class and reduce the distance among the feature vectors and their corresponding centers. Therefore, the deep features in the same class can be pulled towards the class center.

However, L_{c1} only considers intra-class compactness, and inter-class separation is not taken into consideration. If class centers are initialized with small variance, the discriminating distances among different classes cannot be far enough.

On this base, the item L_{c2} is given as follows.

$$L_{c2} = \sum_{r=1}^{k}\sum_{\substack{q=1\\ q\neq r}}^{k}\frac{1}{\|\boldsymbol{c}_r-\boldsymbol{c}_q\|_2^2+\delta} \tag{6}$$

where k denotes the number of classes and $\boldsymbol{c}_r$ denotes the center of the rth class, δ can prevent the denominator from being zero. L_{c2} can penalize small distance among different class centers and boost the representation discriminative ability. Obviously, the compressive loss with L_{c1} and L_{c2} can consider intra-class compactness and inter-class separability simultaneously. It penalizes the distances of training samples to their corresponding class centers and the sum of distances among all class centers. For the classification in a sub-class, minimizing compressive loss means compacting the feature vectors. For the classification among different modulation classes, minimizing compressive loss means increasing the Euclidean distance of different modulation features.

During training process, the softmax loss and compressive loss are utilized jointly to optimize the network and update the class centers. We train the weights and biases of the network to minimize the loss L and the derivative of which with respect to z_i is given by

$$\begin{aligned} \frac{\partial L}{\partial z_i} &= \frac{\partial L_s}{\partial z_i} + \lambda\cdot\frac{\partial L_c}{\partial z_i} \\ &= \frac{1}{m}\left[\sum_{i=1}^{m}\frac{1}{P(y^{(i)}=j|z_i)}\cdot\frac{\partial P(y^{(i)}=j|z_i)}{\partial z_i}\right] \\ &\quad + \lambda\cdot(z_i-\boldsymbol{c}_{y^{(i)}}) \end{aligned} \tag{7}$$

Moreover, during the training process, the class center $\boldsymbol{c}_n$ is trained according to the partial derivative of compressive loss, which can be formulated as follows.

$$\frac{\partial L_c}{\partial \boldsymbol{c}_n} = \lambda_1\frac{\partial L_{c1}}{\partial \boldsymbol{c}_n} + \lambda_2\frac{\partial L_{c2}}{\partial \boldsymbol{c}_n} \tag{8}$$

where

$$\frac{\partial L_{c1}}{\partial \boldsymbol{c}_n} = \frac{\sum_{i=1}^{m}\delta(y^{(i)}=j)(\boldsymbol{c}_j-z_i)}{1+\sum_{i=1}^{m}\delta(y^{(i)}=j)} \tag{9}$$

$$\frac{\partial L_{c2}}{\partial \boldsymbol{c}_r} = \frac{-2(\boldsymbol{c}_r-\boldsymbol{c}_q)}{\left(\|\boldsymbol{c}_r-\boldsymbol{c}_q\|_2^2+\delta\right)^2} \tag{10}$$

$$\frac{\partial L_{c2}}{\partial \boldsymbol{c}_q} = \frac{2(\boldsymbol{c}_r-\boldsymbol{c}_q)}{\left(\|\boldsymbol{c}_r-\boldsymbol{c}_q\|_2^2+\delta\right)^2} \tag{11}$$

$$\frac{\partial L_{c2}}{\partial \boldsymbol{c}_n} = \frac{1}{k-1}\sum_{r=1}^{k}\sum_{\substack{q=1\\ q\neq r}}^{k}\begin{cases}\dfrac{\partial L_{c2}}{\partial \boldsymbol{c}_r}, & n=r\\[2mm] \dfrac{\partial L_{c2}}{\partial \boldsymbol{c}_q}, & n=q\end{cases} \tag{12}$$

In Eqs. (11)-(14), $\delta(y^{(i)}=j)$ is the impulse response function and $\boldsymbol{c}_n$ is the current class center. In each iteration, the center can be updated as follows.

$$\boldsymbol{c}_n^{t+1} = \boldsymbol{c}_n^{t} + \varphi\frac{\partial L_c}{\partial \boldsymbol{c}_n} \tag{13}$$

where $\boldsymbol{c}_n^t$ denotes the class center in tth iteration. The scaler φ is added to control the learning rate for a steady training process.

On this base, we train the weights of CCNN using stochastic gradient descent (SGD) and back propagation algorithm [34]. The training process ends when the loss tends to be stable. In summary, compressive loss is proposed to fine tuning softmax module in the training process. By minimizing the compressive loss, CCNN can make the features more discriminative.

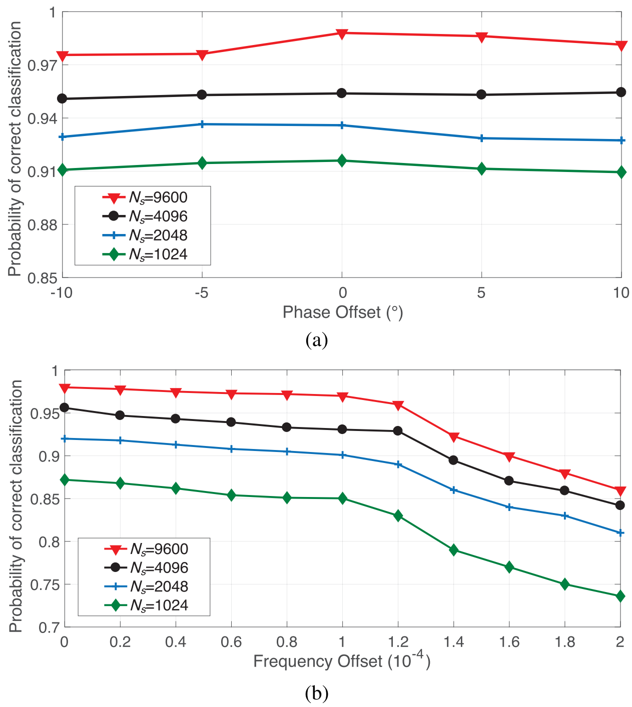

FIGURE 3. The robustness to phase offset and frequency offset of CCNN under different settings of sample lengths. (a) The robustness to phase offset of CCNN. (b) The robustness to frequency offset of CCNN.

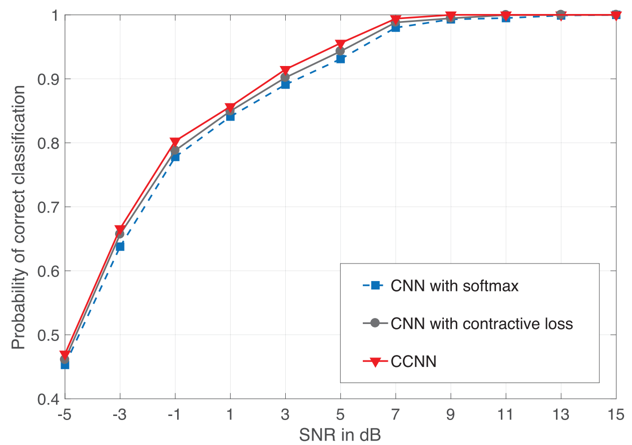

FIGURE 4. The performance comparison of convolutional neural network using different loss functions.

IV. SIMULATIONS

In this section, simulations are conducted to illustrate the superiority and offset robustness of CCNN. The candidate modulation set is $\mathbf{M} = \{BPSK, QPSK, 8PSK, 16QAM, 64QAM\}$. Aiming at reducing the computational complexity and improving the generalization ability, the pixel size of multiple constellation images is down-sampled to 80×80 and all pixels are normalized. Multiple constellation images are obtained by received signals with various pre-defined sample lengths, i.e., 1024, 2048, 4096, 9600. In the training stage, we collect 40000 RCs and CGCs respectively for each modulation at a given SNR as training data. In the classification stage, 10000 RCs and CGCs in different assumed conditions are used to obtain the average probability of correct classification P_{cc}, which is utilized as the performance metric and given by

$$P_{cc} = \sum_{s=1}^{S} P(\hat{H} = H_s|H_s)P(H_s), \quad H_s \in \mathbf{M} \tag{14}$$

where S is the total number of candidate modulations, $P(H_s)$ is the prior probability and we assume the prior probability of each modulation is equal. Additionally, $P(\hat{H} = H_s|H_s)$ denotes the probability that the modulation is distinguished as H_s under hypothesis H_s.

The robustness of CCNN versus phase offset and frequency offset with various sample lengths is shown in FIGURE 3. The SNR is set as 3 dB. In FIGURE 3(a), the range of phase offset is set from $-10°$ to $10°$ with a step of $5°$. It is obvious that, for same sample length, the CCNN performs consistently within the range of phase offset, which indicates better robustness to phase offset. Moreover, we achieve better classification performance by using more signal samples, proving the asymptotic behavior of CCNN. FIGURE 3(b) shows the classification performance versus normalized frequency offset. The x-axis ranges from 0 to 2×10^{-4} (corresponding to a maximum rotation of $180°$) with a step of 2×10^{-5}. It can be observed that the accuracy reduces by 1% from 0 to 10^{-4}, but decreases severely beyond 10^{-4}, illustrating that the region of robustness to frequency offset ends at 10^{-4}.

In FIGURE 4, to demonstrate the performance improvement carried by the compressive loss constraint, we compare the performance among CCNN with three different losses, i.e., softmax loss [33], contractive loss [29] and compressive loss. The sample length is 4096. We can see from FIGURE 4 that CCNN outperforms the others under each SNR. The compressive loss achieves 0.5 dB, 1 dB gains over contractive loss and softmax at 95% P_{cc}. This further proves that the compressive loss facilitates the network to learn discriminative features, thus improving the classification performance. The reason is that CCNN considers not only intra-class compaction but also inter-class separability.

FIGURE 5 reveals the performance comparisons between using multiple constellation images and only RCs. The sample length is set as 4096 and the other settings are the same as FIGURE 4. From the simulation result, the gain of using multiple constellation images is 0.5 dB over RCs at 90% P_{cc}, illustrating the effectiveness of multiple constellation images. This is because CGCs can provide more constellation characteristics, especially under low SNR scenarios, enhancing the characterization performance of neural networks.

FIGURE 6 is an overall method comparison among CCNN, RNN [25], DNN [26] and CNN with softmax [33] under different SNRs. The sample length is set as 4096. FIGURE 6 illustrates that CCNN outperforms the others for the entire SNR range. Specifically, CCNN yields 1.5 dB,

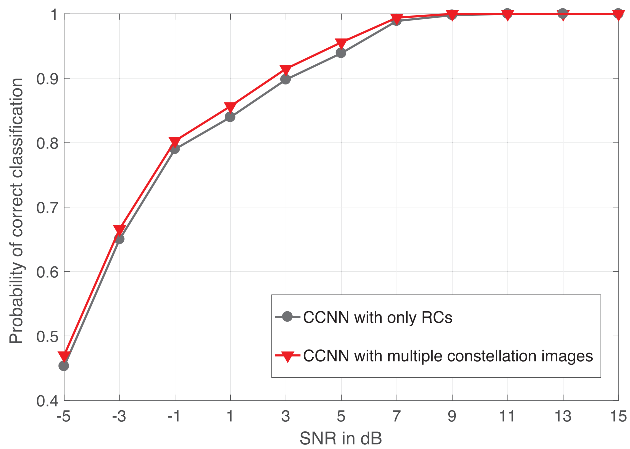

FIGURE 5. The performance comparison of CCNN using different inputs.

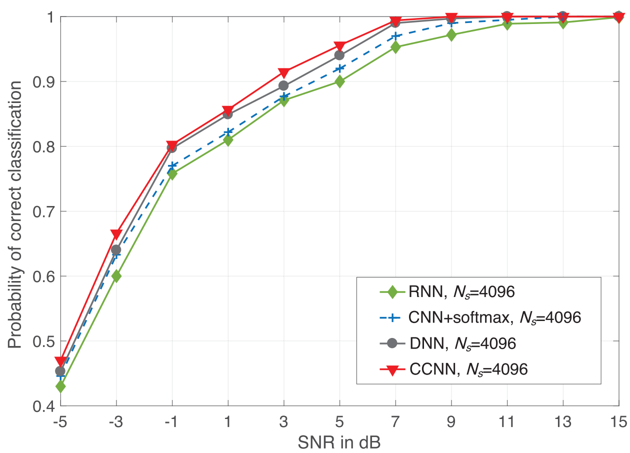

FIGURE 6. The performance comparison among CCNN and other AMC methods.

1.8 dB and 2 dB gains over RNN, DNN and CNN with softmax at 95% P_{cc}, illustrating the superiority of the CCNN. The reasons are from two aspects. Firstly, compared to RNN and DNN, CCNN can extract more discriminative features from input multiple constellation images, which are time-independent. Secondly, it is obvious that the original softmax loss function limits the representation learning ability. Moreover, it is noted that the training time of CCNN, RNN and DNN are 720 seconds (s), 810s and 900s, respectively using the same dataset and equipment (a GTX1080 GPU), illustrating the lower computational complexity of CCNN.

V. CONCLUSION

In this paper, a novel neural network-based AMC method named CCNN is proposed. Firstly, multiple constellation images, including RCs and CGCs, are utilized as the input of the network. Secondly, a CNN with small convolutional kernels is conducted to extract discriminative features and make classification decisions. Finally, a novel compressive loss is designed to enforce the intra-class compactness and the inter-class separability, which makes contributions to the performance improvement and saving training time. Simulations verify the superiority and robustness of CCNN.

REFERENCES

[1] E. Pateromichelakis and K. Samdanis, "A graph coloring based inter-slice resource management for 5G dynamic TDD RANs," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kansas City, MO, USA, May 2018, pp. 1–6.
[2] S. Huang, Y. Jiang, X. Qin, Y. Gao, Z. Feng, and P. Zhang, "Automatic modulation classification of overlapped sources using multi-gene genetic programming with structural risk minimization principle," *IEEE Access.*, vol. 6, pp. 48827–48839, Aug. 2018.
[3] S. Huang, Y. Yao, Z. Wei, Z. Feng, and P. Zhang, "Automatic modulation classification of overlapped sources using multiple cumulants," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6089–6101, Jul. 2016.
[4] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: Classical approaches and new trends," *IET Commun.*, vol. 1, no. 2, pp. 137–156, Apr. 2007.
[5] F. Hameed, O. A. Dobre, and D. Popescu, "On the likelihood-based approach to modulation classification," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 5884–5892, Dec. 2009.
[6] A. Ramezani-Kebrya, I.-M. Kim, D. I. Kim, F. Chan, and R. Inkol, "Likelihood-based modulation classification for multiple-antenna receiver," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3816–3829, Sep. 2013.
[7] A. Polydoros and K. Kim, "On the detection and classification of quadrature digital modulations in broad-band noise," *IEEE Trans. Commun.*, vol. 38, no. 8, pp. 1199–1211, Aug. 1990.
[8] P. Panagiotou, A. Anastasopoulos, and A. Polydoros, "Likelihood ratio tests for modulation classification," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Los Angeles, CA, USA, Oct. 2000, pp. 670–674.
[9] S. Huang, Y. Jiang, Y. Gao, Z. Feng, and P. Zhang, "Automatic modulation classification using contrastive fully convolutional network," *IEEE Wireless Commun. Lett.*, to be published.
[10] Z. Zhu and A. K. Nandi, *Automatic Modulation Classification: Principles, Algorithms and Applications*. Hoboken, NJ, USA: Wiley, 2015.
[11] A. Swami and B. M. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 416–429, Mar. 2000.
[12] S. Huang, Y. Yao, Y. Xiao, and Z. Feng, "Cumulant based maximum likelihood classification for overlapped signals," *Electron. Lett.*, vol. 52, no. 21, pp. 1761–1763, Oct. 2016.
[13] W. A. Gardner and C. M. Spooner, "Cyclic spectral analysis for signal detection and modulation recognition," in *Proc. Mil. Commun. Conf. (MILCOM)*, San Diego, CA, USA, Oct. 1988, pp. 419–424.
[14] K. Kim, I. A. Akbar, K. K. Bae, J.-S. Um, C. M. Spooner, and J. H. Reed, "Cyclostationary approaches to signal detection and classification in cognitive radio," in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw. (DySPAN)*, Dublin, Ireland, Apr. 2007, pp. 212–215.
[15] Z. Zhou, *Machine Learning*. Beijing, China: Tsinghua Univ. Press, 2016.
[16] T. Wang, C.-K. Wen, H. Wang, F. Gao, T. Jiang, and S. Jin, "Deep learning for wireless physical layer: Opportunities and challenges," *China Commun.*, vol. 14, no. 11, pp. 92–111, 2017.
[17] X. Sun, G. Gui, R. Liu, Y. Li, 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.
[18] 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.
[19] X. Zhou, Y. Wu, and B. Yang, "Signal classification method based on support vector machine and high-order cumulants," *Wireless Sensor Netw.*, vol. 2, no. 1, pp. 48–57, 2010.
[20] M. W. Aslam, Z. Zhu, and A. K. Nandi, "Automatic modulation classification using combination of genetic programming and KNN," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2742–2750, Aug. 2012.
[21] K. Triantafyllakis, M. Surligas, G. Vardakis, and S. Papadakis, "Phasma: An automatic modulation classification system based on Random Forest," in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw. (DySPAN)*, Piscataway, NJ, USA, Mar. 2017, pp. 1–3.
[22] Q. Mao, F. Hu, and Q. Hao, "Deep learning for intelligent wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2595–2621, 4th Quart., 2018.

[23] W. Li, H. Liu, Y. Wang, Z. Li, Y. Jia, and G. Gui, "Deep learning-based classification methods for remote sensing images in urban built-up areas," *IEEE Access*, vol. 7, pp. 36274–36284, Dec. 2019.

[24] 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.

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

[26] S. Peng, H. Jiang, H. Wang, H. Alwageed, Y. Zhou, M. M. Sebdani, and Y.-D. Yao, "Modulation classification based on signal constellation diagrams and deep learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 3, pp. 718–727, Mar. 2019.

[27] X. Liu, D. Yang, and A. El Gamal, "Deep neural network architectures for modulation classification," in *Proc. 51st Asilomar Conf. Signals, Syst., Comput. (ACSSC)*, Pacific Grove, CA, USA, Oct. 2017, pp. 915–919.

[28] S. C. Hauser, W. C. Headley, and A. J. Michaels, "Signal detection effects on deep neural networks utilizing raw IQ for modulation classification," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Baltimore, MD, USA, Oct. 2017, pp. 121–127.

[29] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Proc. 28th Annu. Conf. Adv. Neural Inf. Process. Syst.*, Montréal, QC, Canada, Dec. 2014, pp. 1988–1996.

[30] P. T. De Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Ann. Oper. Res.*, vol. 133, no. 1, pp. 19–67, Jan. 2005.

[31] R. Hadsell, S. Chopra, and Y. Lecun, "Dimensionality reduction by learning an invariant mapping," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New York, NY, USA, Jun. 2006, pp. 1735–1742.

[32] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *Engineering Applications of Neural Networks* (Communications in Computer and Information Science), vol. 629, no. 1. Cham, Switzerland: Springer, Feb. 2016, pp. 213–226.

[33] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Columbus, OH, USA, Jun. 2014, pp. 1891–1898.

[34] Y. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient backprop," in *Neural Networks: Tricks of the Trade*, vol. 1524. Berlin, Germany: Springer, 2012, pp. 9–48.

SAI HUANG is currently with the Department of Information and Communication Engineering, Beijing University of Posts and Telecommunications, as a Lecturer, and serves as the Academic Secretary of the Key Laboratory of Universal Wireless Communications, Ministry of Education, China. He is the Reviewer of international journals such as the IEEE Transactions on Vehicular Technology, the IEEE Wireless Communications Letters, the IEEE Transactions on Cognitive Communications and Networking, and international conferences such as the IEEE ICC and the IEEE GLOBECOM. His research interests include machine learning assisted intelligent signal processing, statistical spectrum sensing and analysis, fast detection and depth recognition of universal wireless signals, millimeter wave signal processing, and cognitive radio network.

LU CHAI received the B.S. degree from Beijing University of Posts and Telecommunications (BUPT), where he is currently pursuing the M.S. degree. His research interests include deep learning algorithms, and universal signal detection and identification.

ZENING LI is currently pursuing the M.S. degree in data science with the University of Denver. His research interests include parallel and distributed computing, and data mining.

DI ZHANG (S'13–M'17) received the Ph.D. degree from the Waseda University, Tokyo, Japan, in 2017. He is an Assistant Professor with the Zhengzhou University, Zhengzhou, China. He was a Senior Researcher with the Seoul National University, Seoul, Korea, from 2017 to 2018. He visited the National Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, Beijing, China, from 2015 to 2017, and the National Chung Hsing University, Taichung, Taiwan, in 2012. He serves as the Editor of the KSII Transactions on Internet and Information Systems, has served as the Guest Editor of the IEEE Network, IEEE Access, the IET Intelligent Transport Systems, TPC Member of many IEEE and ACM flagship conferences, such as ACM MobiHoc, the IEEE ICC, and the IEEE WCNC. His research interests include 5G wireless networks, the Internet of Things, and telemedicine.

YUANYUAN YAO received the Ph.D. degree in information and communication engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 2017. Since 2017, she has been with the School of Information and Communication Engineering, Beijing Information Science and Technology University, Beijing, China, as a Lecturer. Her research interests include stochastic geometry and its applications in large-scale wireless networks, energy harvesting, and wireless communication.

YIFAN ZHANG received the B.S., M.S., and Ph.D. degrees from Beijing University of Posts and Telecommunications (BUPT). He is a Researcher with the School of Information and Communication Engineering, BUPT. His current research interests include signal modulation classification, signal processing, cognitive radio network implementation, and optimization algorithm in wireless networks.

ZHIYONG FENG received the B.S., M.S., and Ph.D. degrees in information and communication engineering from Beijing University of Posts and Telecommunications (BUPT), Beijing, China. She is a Full Professor. She is the Director of the Key Laboratory of Universal Wireless Communications, Ministry of Education. She is a Technical Advisor of NGMN, the Editor of IET Communications, and KSII Transactions on Internet and Information Systems, the Reviewer of the IEEE TWC, the IEEE TVT, and the IEEE JSAC. She is active in ITU-R, IEEE, ETSI, and CCSA standards. Her main research interests include wireless network architecture design and radio resource management in 5th generation mobile networks (5G), spectrum sensing and dynamic spectrum management in cognitive wireless networks, universal signal detection and identification, and network information theory.