IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 19, 2019, accepted December 2, 2019, date of publication December 6, 2019, date of current version January 6, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2958131

One-Dimensional Deep Attention Convolution Network (ODACN) for Signals Classification

SHUYUAN YANG [1], (Senior Member, IEEE), CHEN YANG [1], DONGZHU FENG [2], XIAOYANG HAO [1], AND MIN WANG [3]

[1]School of Artificial Intelligence, Xidian University, Xi'an 710126, China
[2]School of Aerospace Science and Technology, Xidian University, Xi'an 710126, China
[3]Key Laboratory of Radar Signal Processing, Xidian University, Xi'an 710126, China

Corresponding author: Shuyuan Yang (syyang2009@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61771380, Grant 61906145, Grant U1730109, Grant 91438103, Grant 61771376, Grant 61703328, Grant 91438201, Grant U1701267, and Grant 61703328, in part by the Equipment pre-research project of the 13th Five-Years Plan under Grant 6140137050206, Grant 414120101026, Grant 6140312010103, Grant 6141A020223, Grant 6141B06160301, and Grant 6141B07090102, in part by the Major Research Plan in Shaanxi Province of China under Grant 2017ZDXM-GY-103, and Grant 017ZDCXL-GY-03-02, in part by the Foundation of the State Key Laboratory of CEMEE under Grant 2017K0202B, Grant 2018K0101B, Grant 2019K0203B, and Grant 2019Z0101B, in part by the Science Basis Research Program in Shaanxi Province of China Grant 16JK1823, Grant 2017JM6086, and Grant 2019JQ-663, in part by the Key Scientific and Technological Innovation Team of Shaanxi Province under Grant S2020-ZC-TD-0084.

ABSTRACT Handcraft features are commonly used for signal classification, which is a time-consuming feature engineering. In order to develop a general and robust feature learning method for radio signals, a novel One-dimensional Deep Attention Convolution Network (ODACN) is proposed to automatically extract discriminative features and classify various kinds of signals. First, one-dimensional (1-D) sparse filters are designed to learn hierarchical features of raw signals. Second, an attention layer is constructed to weight and assemble feature maps, to derive more context-relevant representation. By using simple 1-D filtering, ODACN is characteristic of less parameters and lower computation complexity than traditional Convolutional Neural Networks (CNNs). Moreover, feature attention can mimic a succession of partial glimpses of humans and focus on context parts of signals, thus helps in recognizing signals even at low Signal-to-Noise Ratio (SNR). Some experiments are taken to classify 31 kinds of signals with different modulation and channel coding types, and the results show that ODACN can achieve accurate classification of very similar signals, without any prior knowledge and manual operation.

INDEX TERMS Signal classification, feature learning, one-dimensional convolution neural network, attention layer.

I. INTRODUCTION

In the non-cooperative communication system, the receiver needs to demodulate and decode signals to recover the transmitted information sequence. Thus, a blind recognition of modulation and channel coding type is necessary, which can be regarded as a pattern recognition task and has attracted interests of many researchers [1], [2]. Traditional methods first extract discriminative features of signals and then use them for the subsequent classification, where the features play the most important role in the recognition. Currently, a great deal of works have been done on the feature extraction of radio signals [3]–[16], such as signal statistics [3]–[6], Higher Order signal Statistics (HOS, including moments, cumulants and kurtosis) [7]–[9], Wavelet Transform (WT) [10], [11], spectral features [12], signal constellations [13], zero-crossings [14], multi-fractals [15] and radon transform [16], to distinguish various modulation types and constellations. HOS are most often adopted for digital modulation classification, especially cumulants. For example, A. Swami proposed hierarchical features based on the fourth-order cumulants [7]. P. Marchand combined the second order and fourth order Cyclic Cumulants' (CC) magnitudes to formulate features [17]. O. A. Dobre advanced the eighth order CC in [8], and the n^{th}-order warped CC magnitude was employed in [18] for QAM modulation classification.

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

In [18], the authors used spectral analysis and instantaneous time-domain parameters for feature extraction. In [3], [19], [20], Choi–Williams, Wigner Ville Distribution (WVD), and Quadrature Mirror Filter Bank (QMFB) were also explored for finding reliable features. For radar signals, the most often used features were carrier frequency, pulse width and pulse repetition interval [21]–[25]. Radon Wigner transform and radon QMFB techniques were employed in [21] and [22] for LPI signals. In [23], the modulation parameters were used as features of polyphase radar signals. In [24], Radon Ambiguity Transform (RAT) and Radon-WVD features were proposed via the Fractional Fourier transform (FrFT) for the classification. In [25], the authors addressed the feature extraction of advanced LPI radar via a Wigner-Ville distribution-Radon transform.

Although many works have explored various types of features for radio signal classification, most of them are designed handcraft and heavily depend on the domain knowledge and empirical trial. However, with the ever-increasing interferences and emitters, these features are not flexible and robust to real electromagnetic environment. That is, the handcrafted designed features will present degraded performance on the signals with varied types and heavy distortions. Consequently it is desired to develop an efficient and general-purpose method to automatically extract features for varied signals, such as communication, radar and navigation signals.

Since the pioneered work of Hinton, many deep learning (DL) models have been developed to automatically extract features of complex objects [31], [32]. In the signal processing field, deep learning technology is also explored for high-level abstraction of signals [26]–[29]. Inspired by the advances of DL technology in computer vision, several signal classification approaches have been developed, including convolution neural networks [33], [36], autoencoder networks [26] and recurrent neural networks [34]. However, most of them first formulate an empirical feature to represent signals, and then use deep networks for the classification. For example, in [26] an autoencoder network trained by a nonnegativity constraint algorithm is proposed, where the fourth-order cumulants of signals are used. Other works also calculate some time-frequency maps of signals, and then use the 2-D conventional operations to deal with the maps.

Considering these limitations, in this paper we design a new One-dimensional Deep Attention Convolutional Network (ODACN), to directly deal with raw radio signals for automatic feature learning and signal classification. One-dimensional (1-D) sparse filters are designed to automatically extract hierarchical features of signals. Moreover, an attention layer is constructed to weight and assemble feature maps to derive the most discriminative features. By using simple 1-D filtering and feature synthesis, ODACN is characteristic of less parameters and lower computation complexity than traditional Convolutional Neural Networks (CNNs). By using a group of 1-D filters in each layer of this deep neural network, ODACN can automatically learn hierarchical features of raw signals. Moreover, an attention layer is employed to weight and assemble feature maps to select the most discriminative features.

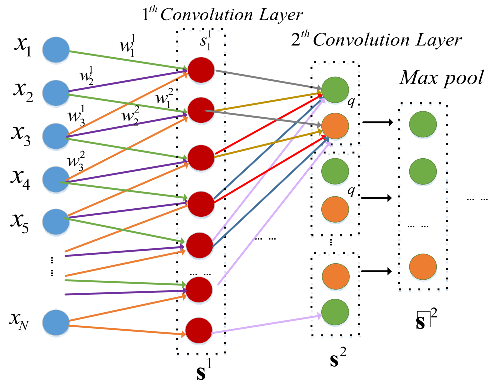


FIGURE 1. Schematic of 1-D convolution. In the first and second layer, the sizes of convolution kernels are 3 and 5 respectively. In each layer, only one convolution kernel is expounded.

Different with the available works on automatic modulation classification methods, the proposed ODACN has the following characteristics: 1) it provides a general-purpose and robust feature extraction framework for various kinds of radio signals, which can avoid a tedious and manual selection of features; 2) it replaces the 2-D convolution in CNN by 1-D convolution, so obtaining a lightweight network with fewer parameter and rapider implementation; 3) it employs attention mechanism to weight and assemble feature maps to select the most discriminative features, which can further reduce the computational burden and make features more representative. Some experiments are taken to classify 31 kinds of signals with different modulation and channel coding types, and the results show that ODACN can achieve accurate classification of very similar signals, without any prior knowledge and manual operation.

The rest part of this paper is organized as follows: section II describes the construction of the proposed ODACN. Section III discusses the experimental results. Section IV gives the conclusion.

II. ONE-DIMENSIONAL DEEP ATTENTION CONVOLUTION NETWORK (ODACN)

A. 1- D CONVOLUTION

The traditional CNNs are good at handling 2D data and have achieved tremendous success in many engineering fields [30], [31]. Thus signals are usually converted into 2D features for the further feature learning. For example, Grid Constellation Matrix (GCM) is used in [33], or some features are rearranged as 2D feature map [36]. Considering that radio signals are inherent one-dimensional, in this paper we construct 1-D sparse convolution layer for more efficient feature learning. In each layer, we utilize a group of relatively small filters than input images to reduce the memory cost.

Given a signal sequence $\{x_i\}_{i=1}^{N}$, and x_i denotes the i^{th} sample in the sequence. In the first layer, assume there are N input neurons and K output neurons. Then we construct a group of connected weights $\{w_m^k\}(m = 1, .., M; k = 1, .., K)$, where M is the length of the filter. Here w_m^k is the connected weight between the m^{th} input neuron and k^{th} output neuron, w_b^k is the bias of the k^{th} output neuron. Then the output layer can provide a group of features for the signal sequence, with the n^{th} feature being:

$$s_n^k = f\left(\sum_{i=1}^{M} w_i^k x_{i+n-1} + w_b^k\right) \tag{1}$$

Similar to 2-D CNNs, we use the same parameters for a group of neural connections to reduce the number of parameters, i.e., the same set of $\{w_1, w_2, \ldots, w_M\}$ is adopted for all the neurons. Thus in our network we use the sparse interactions, parameter sharing and equivalent representations to improve the computational efficiency and quality of features.

Denote the output features of the first layer as $\mathbf{s}^1 \in R^{1\times N}$, with its elements being s_n^k. Assume there are $B_i(i = 1, 2, \ldots)$ kernels in each convolution layer, the output of the first convolution layer is a feature matrix $\mathbf{S}^1 \in R^{B_1\times N}$. Then a pooling function replaces the features at a certain location with a summary statistic of the nearby outputs in each channel, and each channel is processed individually. In our proposed architecture, we apply *Max Pool* operation as the sampling value after under-sampling, which take the maximum value in a segment with length q. After the pooling operation that is executed on each channel individually, we can obtain a new feature map of the first layer $\tilde{\mathbf{S}}^1$. Multiple layers are cascaded to formulate a deep neural network, and the signal sequence is processed by the cascaded filters [32]. A max pooling operation can follow multiple convolution layers. For example, we can construct a network with the output filtering maps of each layer being,

$$\mathbf{S}^0(\{x_i\}_{i=1}^N) \rightarrow \mathbf{S}^1 \rightarrow \mathbf{S}^2 \rightarrow \tilde{\mathbf{S}}^2 \rightarrow \mathbf{S}^3 \rightarrow \tilde{\mathbf{S}}^3 \ldots\ldots \tag{2}$$

B. ATTENTION LAYER

In order to further improve the performance of the 1-D convolution network, and select the most discriminative features, in our work an attention layer is employed to weight and assemble the filtering maps. In [35], "attention" mechanism has been explored in deep neural networks for utilizing long-term dependencies of sequences. Inspired by it, we explore the dependencies among features using the saliency attention mechanism. That is, there are some "salient" features that are most important than the others. Consequently, casting a large weight for them to assemble features can not only reduce parameters of the network but also improve the discriminative of features.

In our method, denote the feature vector produced by the final *Max Pool* layer (the f^{th} layer)as $\tilde{\mathbf{S}}^f \in R^{B_f\times N_f}$, where $N_f = N/q^f$ and each row is $\tilde{\underline{\mathbf{s}}}_j^f \in R^{1\times N_f}(j = 1, .., B_f)$ that is called as a "state" in the attention layer. Then we calculate a "synthesis" vector $\mathbf{c}$ by the states sequence $\{\tilde{\underline{\mathbf{s}}}_j^f\}_{j=1}^{B_f}$,

$$\mathbf{c} = \sum_{j=1}^{B_f} \alpha_j \tilde{\underline{\mathbf{s}}}_j^f \tag{3}$$

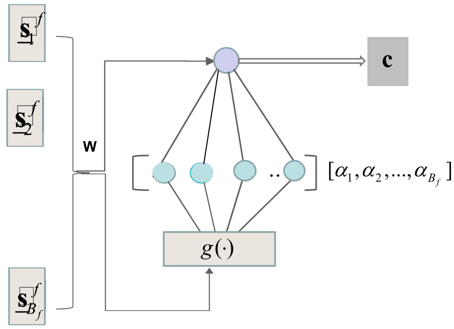


FIGURE 2. Schematic of attention layer. Vectors in the hidden state sequence are fed into a learnable system to produce a probability vector α.

The weights α_j are then calculated by $e_j = g(\tilde{\underline{\mathbf{s}}}_j^f \mathbf{w})$, where $g(\cdot)$ is a tansig function and $\mathbf{w} \in R^{N_f\times 1}$ is a set of learnable parameters. Tansig function can amplify the difference among neurons, when compared with Sigmoid function. In the attention layer, the weights in the attention layer are calculated by,

$$\alpha_j = \exp(e_j) / \sum_{k=1}^{B_f} \exp(e_k) \tag{4}$$

A schematic of this form of attention is shown in Fig.2, where the vectors in the hidden state sequence are fed into a learnable system to produce a probability vector α. The vector $\mathbf{c}$ is calculated as a weighted average, with weighting given by α.

C. ONE-DIMENSIONAL DEEP ATTENTION CONVOLUTIONAL (ODAC)

By cascading the above 1-D convolution layers and attention layers, we can construct an ODAC (One-dimensional Deep Attention Convolutional) module. First, raw radio signals are processed by several convolution layers, pooling layers and then the attention layer, to extract feature maps of signals. In the ODAC, six convolution layers and five pooling layers are adopted, as shown in Fig.3. In the layers, $B_1 = 64, B_2 = 64, B_3 = 128, B_4 = 128, B_5 = 256, B_6 = 256$ and $f = 6$, that is, 64, 64, 128, 128, 256 and 256 convolution kernels are used. In the pooling layer, maxpooling is employed with a step $q = 2$. The network structure of ODAC module is denoted as: $\mathbf{S}^1 \rightarrow \mathbf{S}^2 \rightarrow \tilde{\mathbf{S}}^1 \rightarrow \mathbf{S}^3 \rightarrow \tilde{\mathbf{S}}^3 \rightarrow \mathbf{S}^4 \rightarrow \tilde{\mathbf{S}}^3 \rightarrow \mathbf{S}^5 \rightarrow \tilde{\mathbf{S}}^4 \rightarrow \mathbf{S}^6 \rightarrow \tilde{\mathbf{S}}^5 \rightarrow \mathbf{S}^6 \rightarrow \mathbf{A}$, which is shown in Fig.3. In the figure, the extracted features of conv6 are input into the attention layer. Then the features are weighted by attention. Based on the ODAC, we construct the One-dimensional Deep Attention Convolution Network (ODACN) to classify signals via the learned features, as shown in Fig.4.

In the ODACN, three ODAC are employed to extract features of signals and then the features are cascaded to be processed by two dense layers. After the concatenated processing in a node, the feature maps can be fed into a fully connected layer (the SoftMax layer in Fig.4) for the subsequent classification.

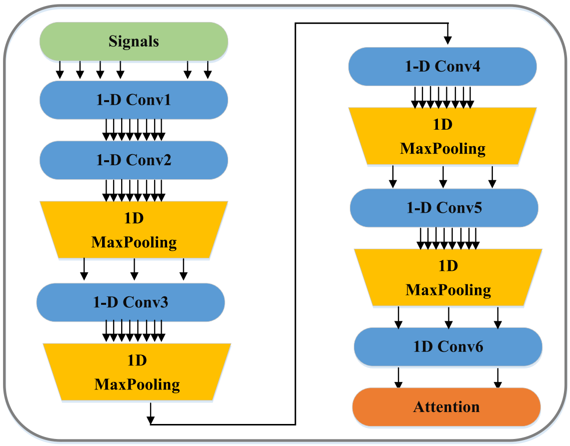


FIGURE 3. The structure of our proposed ODAC. The parameters in six convolution layers are: 64@1∗3, 64@1∗5, 128@1∗3, 128@1∗5, 256@1∗3, 256@1∗5. Five Max Pool layers are used, with $q = 2$ in each layer.

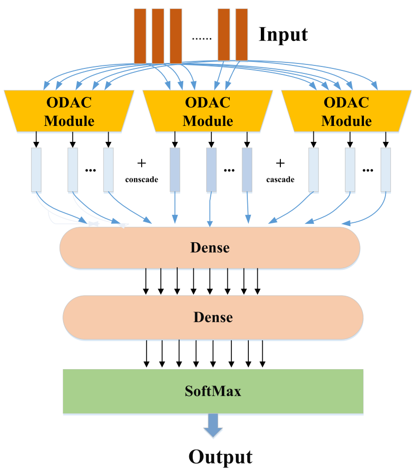


FIGURE 4. The structure of our proposed ODACN.

In the network training, denote the output of the network as a_i for the i^{th} signal sequence x_i, and denote the label of the i^{th} signal sequence as y_i. As soon as the ODACN is constructed, it is trained by a set of training samples, with the loss function being,

$$L = \sum_{i=1}^{N} \left[y_i \log a_i + (1 - y_i) \log(1 - a_i) \right] \tag{5}$$

In the training process, we update all the network parameters (including the convolution, attention and dense layers) according to the Stochastic Gradient Descent (SGD) algorithm. In SGD, the batch size is set as 64 and the learning ratio is set as 0.001. The maximum number of iterations is set as 40.

III. EXPERIMENTAL RESULTS

A. DATASET DESCRIPTION

In this section, we investigate the performance of ODACN by recognizing signals with different modulation and channel coding modes, which is shown in Table 1. Three kinds of coding and modulation types are considered for shortwave band signals, and four kinds of coding and five kinds of modulation types are considered for ultra shortwave bands. By a cross combination of band, channel coding with modulation, we can obtain 29 kinds of signal, which follow the naming rule: modulation type +band (digital) +channel coding (digital) + the scrambling code polynomial. The scrambler parameters are pseudo-random symbol scrambling and the scrambling code polynomial support 9 level. We set the interleaving parameters as matrix interleaving. Finally AM and FM radio signals are also included in the dataset. For shortwave signals, each kind of signal includes 25000 signals in the frame length of 10ms-150ms.

TABLE 1. The modulation and channel coding modes of signals.

Band	Modulation	Channel Coding	#Signals
Shortwave (1)	QPSK	Hamming code(0)	QPSK109,QPSK119 QPSK129,QPSK139 EPSK109,EPSK119 EPSK139,FSK109 FSK119,FSK139
	8PSK	None-systematic convolution code (code rate is 1/2)(1)	
	2FSK	None-systematic convolution code (code rate is 2/3)(2)	
Ultra Shortwave (2)	FM	Hamming code(0)	BPSK209, BPSK219, BPSK229, BPSK239, EPSK209, EPSK219, EPSK239, FSK209, FSK219, FSK229, FSK239, FM209, FM219, FM229, FM239, QPSK209 QPSK219, QPSK229, QPSK239, AM, FM
	2FSK	None-systematic convolution code (code rate is 1/2)(1)	
	QPSK	None-systematic convolution code (code rate is 2/3)(2)	
	BPSK	None-systematic convolution code (code rate is 3/4)(3)	

For the signal generation, the modulation rate takes a random number in the range of 50Bd-2.4kBd. For ultra shortwave signals, each kind of signal includes 15000 signal samples in the frame length of 10ms-800ms. The modulation rate takes a random number in the range of 1kBd-20kBd. The waveforms of 31 kinds of signals are shown in Fig.5, from which we can observe that it is very difficult to distinguish some signals for their similarity. The dataset is generated with GNU Radio and has varied signal-to-noise ratios in the range of [5 30]dB.

In the dataset, the SNRs of various kinds of signals are taken randomly in in the range of [5 30]dB, and the time-selective fading and frequency-selective fading channels are considered. For each signal, we extract successive 440 samples from signals with an interval of 100, to construct the training samples, which has about 4450000 samples. In the experiments, half of the dataset is randomly selected for the training and the rest is used for the testing. Moreover, we select 100 samples for each class in the testing dataset to formulate a validation dataset, to perform a model selection of the ODACN. In this section, we design a series of experiments on the dataset to investigate the performance of proposed ODACN. All the experiments are realized on a 64GB RAM HP Z840 workstation with dual E5-2630v CPUs and a NVIDIA GeForce GTX TITAN X GPU.

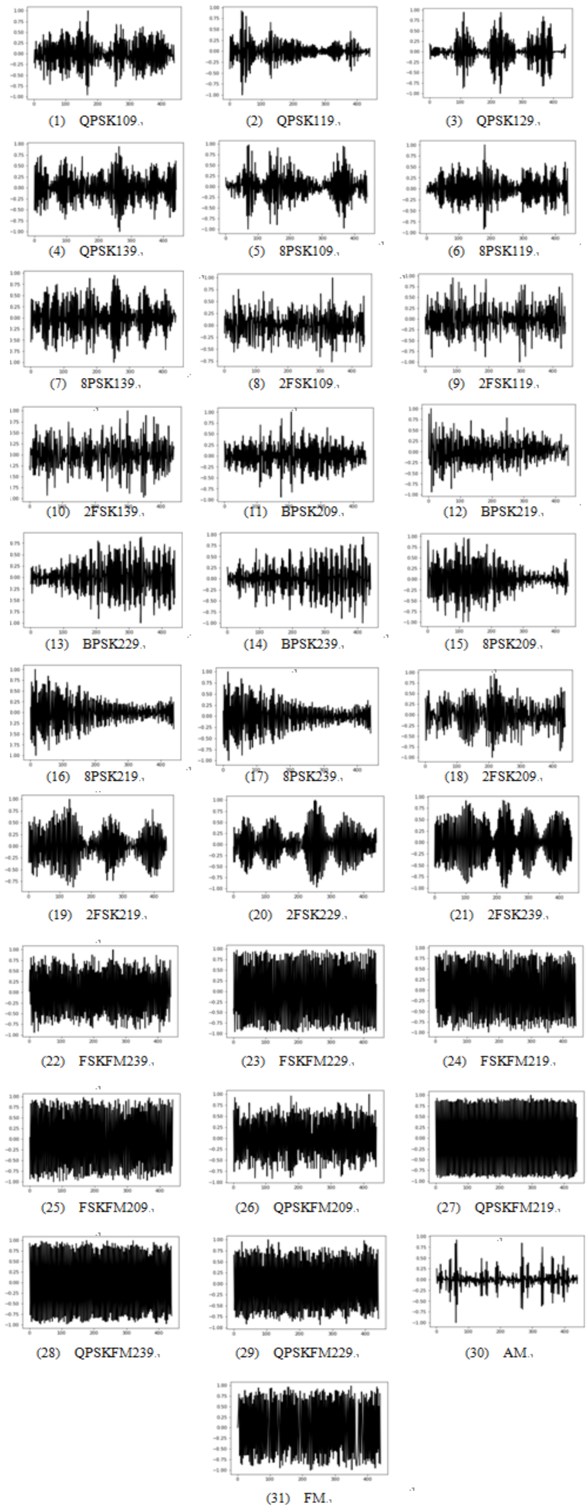


FIGURE 5. Waveforms of 31 kinds of radio signals.

B. TRAINING AND TESTING RESULTS

The training set is used to train the network shown in Fig.4 for thirty times, and it will consume about 600 seconds per epoch by using the platform described in section III.A. In the training, each fully connected layer is followed by dropout [30] at ratio of 0.5 to prevent the network from overfitting. The network is trained through a forward calculation and fine-tuning with a stochastic gradient decent algorithm to optimize the loss until the whole network converges. The variation of the training loss function and the validation loss function (red line) with the number of iterations, is shown in Fig.6. It can be seen that with the increase of the number of training iterations, the value of both training loss function and validation loss function decrease and eventually converge to a stable value within 40 iterations.

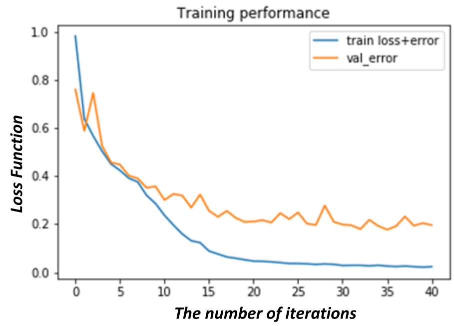


FIGURE 6. Variation of the loss function in the training.

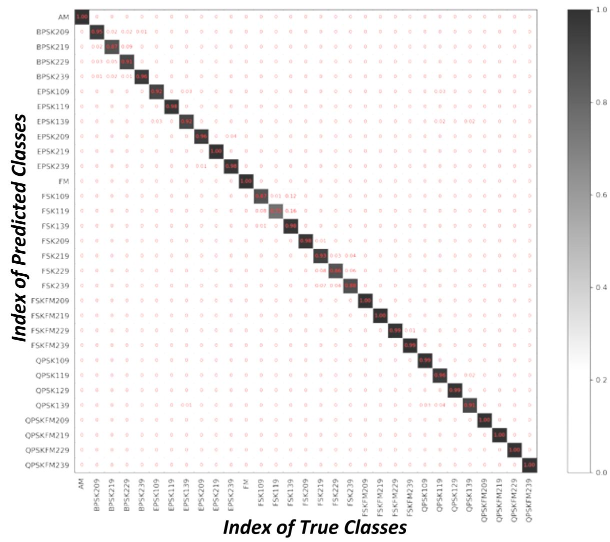


FIGURE 7. The confusion matrix of the classification results by ODACN.

When the training and validation completes, we use the trained network to predict the labels of signals in the testing dataset. The Overall Accuracy (OA) is used for evaluating the classification results of ODACN. Thirty independent experiments are taken and the average accuracy of the network on the 31 kinds of signals, and the average OA of the classification results by ODACN is 95.57%. The confusion matrix of one test is shown in Fig.7. The number in the percentage column represents the percentage of the left hand side signal type that is misclassified as the signal type on the right hand side. From it we can observe that for most of signals, the proposed network can achieve accurate classification, without any signals priors and handcraft operations.

TABLE 2. The Classification accuracy of ODACN.

Index / #ODAC	Epoch	OA(%)	AA(%)	KC	#Parameters
1	65	92.18	92.74	0.9301	572,783
2	121	94.32	94.83	0.9481	1,143,487
3	182	95.57	95.91	0.9562	1,714,191
4	239	95.63	95.67	0.9591	2,284,895

Making an analysis on the classification results we will find that a small portion of BPSK219 signals are misclassified as BPSK229 signals and a small portion of BPSK229 signals are misclassified as BPSK219 signals, as shown in Fig.8(a). Moreover, we can notice that FSK119 and FSK139 are more likely to be misclassified as each other, as shown in Fig.8(b). Fig.8(c) and Fig.8(d) give some misclassification of QPSK and EPSK signals respectively. From it we can observe fewer errors appear in QPSK and EPSK signals than that of BPSK and FSK signals.

Then we vary the number of ODAC channels in the ODACN, and analyze the influence of it on the classification accuracy. Thirty independent experiments are taken and the average OA, Average Accuracy (AA) and Kappa Coefficient (KC) are also calculated, which are shown in Table 2. Moreover, the number of network parameters is also shown in the table for a comparison.

From it we can observe that the increase of the ODAC channels will increase both the accuracy and the network size. In order to achieve rapid implementation, we take three ODAC channels in the constructed ODACN. From the table we can observe that ODACN only has less than 2 million parameters, which is remarkably than the two-dimensional CNNs. Thus our network is characteristic of low computational and storage complexity.

C. LEARNED FEATURES

In this section we investigate the extracted features of signals by our proposed network. Some features filtered by the first, second, third, fourth, fifth and sixth layer of ODACN, are plotted in the Fig.9(a), Fig.9(b), Fig.9(c), Fig.9(d). Fig.9(e) and Fig.9(f) respectively. From them we can observe that at the top of the constructed network, the filtered feature are similar and messy.

With the increase of the depth of layers, the extracted features become sparser and sparser and some channels with zero output are identified. Thus the constructed network is redundant and its size can be further reduced to obtain a lighter network. On the other hand, the filtered features become more abstract and regular with the increase of the layer depth, which indicates some intrinsic patterns existed in signals.

D. COMPARISON OF DIFFERENT METHODS

In this section we investigate the effectiveness of ODACN with the degraded signals whose SNR is lower than 3dB. In this section, we vary the SNR of the signals from

	BPSK209	BPSK219	BPSK229	BPSK239
BPSK209	0.95	0.02	0.02	0.01
BPSK219	0.02	0.87	0.09	0
BPSK229	0.03	0.05	0.91	0
BPSK239	0.01	0.02	0.01	0.96

(a) BPSK

	QPSK109	QPSK119	QPSK129	QPSK139
QPSK109	0.99	0	0	0
QPSK119	0	0.96	0	0.02
QPSK129	0	0	0.99	0
QPSK139	0.03	0.04	0	0.91

(b) QPSK

	FSK109	FSK119	FSK139	FSK209	FSK219	FSK229	FSK239
FSK109	0.87	0.01	0.12	0	0	0	0
FSK119	0.08	0.75	0.16	0	0	0	0
FSK139	0.01	0	0.98	0	0	0	0
FSK209	0	0	0	0.98	0	0	0
FSK219	0	0	0	0	0.93	0.03	0.04
FSK229	0	0	0	0	0.08	0.86	0.06
FSK239	0	0	0	0	0.07	0.04	0.88

(c)FSK

	EPSK109	EPSK119	EPSK139	EPSK209	EPSK219	EPSK239
EPSK109	0.92	0	0.03	0	0	0
EPSK119	0	0.98	0	0	0	0
EPSK139	0.03	0	0.92	0	0	0
EPSK209	0	0	0	0.96	0	0.04
EPSK219	0	0	0	0	1	0
EPSK239	0	0	0	0.01	0	0.98

(d)EPSK

FIGURE 8. The confusion matrixes of BPSK, QPSK, FSK and EPSK signals.

−20dB to 20dB, and the same number of signals are generated every 2dB to test the performance of ODACN. The constructed network is compared with some classic machine learning based classification methods and some state-of-the-art methods, including the Complement Naive Bayes (CNB) classifier [37], Support Vector Machine (SVM) [38], Contrastive Learning based Deep Neural Network (CLDNN) [33], Random Forest (RF) [39], Sparse Auto-Encoder (SAE) [41] and Convolution Neural Network (CNN) [40] are considered. In the comparison methods,

(a) features filtered by the first layer

(b) features filtered by the second layer

(c) features filtered by the third layer

(d) features filtered by the forth layer

(e) features filtered by the fifth layer

(e) features filtered by the sixth layer

FIGURE 9. Extracted Features by the first, second, third, fourth, fifth and sixth layer of ODACN.

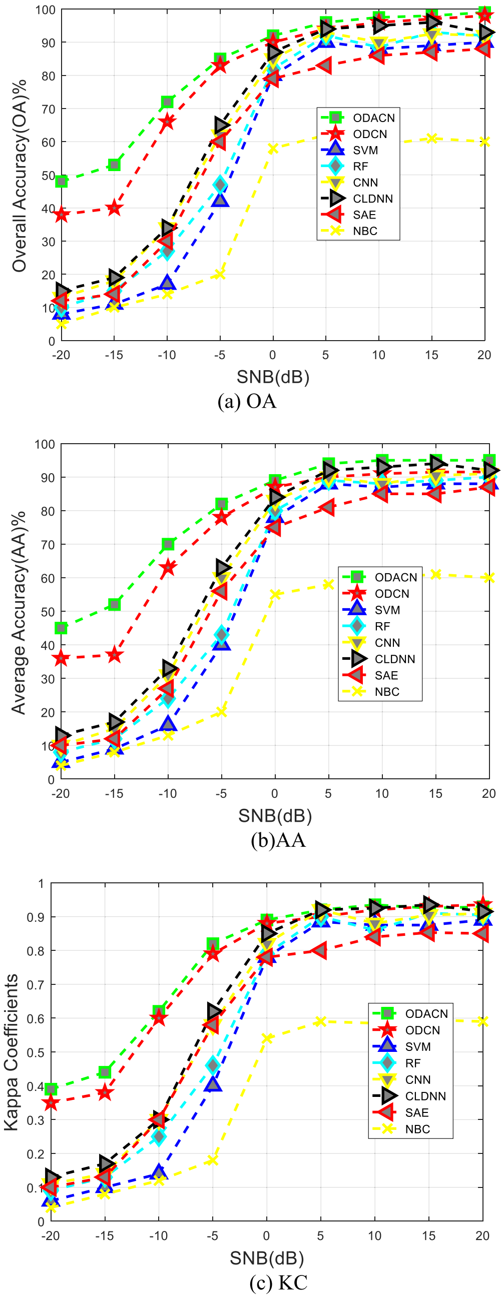


FIGURE 10. Comparison of ODACN with other methods.

the parameters are set according to the literatures' suggestion. Moreover, we delete the attention module in the ODAC in Fig.3, to construct a One-dimensional Deep Convolution Network (ODCN), to investigate the influence of the attention layer on the ODCAN. The comparative methods are trained using the same dataset, and we take the parameters suggested in the literatures. The Overall Accuracies (OAs), Average Accuracies (AAs) and Kappa Coefficients (KCs) of the classification results by different methods, are shown in Fig.10(a), Fig.10 (b) and Fig.10(c) respectively. From it we can observe that deep learning based methods are superior to traditional shallow machines, including SVM, RF and NBC.

Compared with other deep networks, our method can achieve higher OA, AA and KC, which invalidate the robustness of our constructed network. We can also observe that ODACN has remarkable improvement over ODCN at low SNRs, which also validate the effectiveness of attention module.

IV. CONCLUSION

In this paper we propose a novel ODACN for feature extraction and classification of raw radio signals. By 1-D designing convolution layer and attention layer, the network can directly deal with raw signals and extract discriminative features for classification. Several experiments are performed on the dataset with many kinds of similar signals, to investigate its classification accuracy and robustness. The results show that ODACN can identify signals with different modulation and channel coding types, and is characteristic of rapid training and good generalization. The comparison results with some traditional and deep learning based methods demonstrate its effectiveness and superiority to its counterparts. In future, we will consider more advanced CNNs such as siamese network for further improvement.

REFERENCES

[1] M. Bkassiny, S. K. Jayaweera, Y. Li, and K. A. Avery, "Wideband spectrum sensing and non-parametric signal classification for autonomous self-learning cognitive radios," *IEEE Trans. Wireless Commun.*, vol. 11, no. 7, pp. 2596–2605, Jul. 2012.

[2] V. Gouldieff, J. Palicot, and S. Daumont, "Blind modulation classification for cognitive satellite in the spectral coexistence context," *IEEE Trans. Signal Process.*, vol. 65, no. 12, pp. 3204–3217, Jun. 2017.

[3] E. E. Azzouz and A. K. Nandi, "Automatic identification of digital modulation types," *Signal Process.*, vol. 47, no. 1, pp. 55–69, Nov. 1995.

[4] A. K. Nandi and E. E. Azzouz, "Algorithms for automatic modulation recognition of communication signals," *IEEE Trans. Commun.*, vol. 46, no. 4, pp. 431–436, Apr. 1998.

[5] I. Druckmann, E. I. Plotkin, and M. N. S. Swamy, "Automatic modulation type recognition," in *Proc. IEEE Can. Conf. Elect. Comput. Eng.*, vol. 1, May 1998, pp. 65–68.

[6] K. E. Nolan, L. Doyle, D. O'Mahony, and P. Mackenzie, "Signal space based adaptive modulation for software radio," in *Proc. IEEE Wireless Commun. Netw. Conf. Rec. (WCNC)*, vol. 1, Mar. 2002, pp. 510–515.

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

[8] O. A. Dobre, Y. Bar-Ness, and W. Su, "Higher-order cyclic cumulants for high order modulation classification," in *Proc. IEEE Mil. Commun. Conf.*, vol. 1, Oct. 2003, pp. 112–117.

[9] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Selection combining for modulation recognition in fading channels," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, vol. 4, Oct. 2005, pp. 2499–2505.

[10] K. C. Ho, W. Prokopiw, and Y. T. Chan, "Modulation identification of digital signals by the wavelet transform," *IEE Proc.-Radar, Sonar Navigat.*, vol. 147, no. 4, pp. 169–176, Aug. 2000.

[11] D. Wu, X. Gu, and Q. Guo, "A new scheme of automatic modulation classification using wavelet and WSVM," in *Proc. 2nd Asia–Pacific Conf. Mobile Technol., Appl. Syst.*, Guangzhou, China, 2005, pp. 1–5.

[12] C. Schreyogg, K. Kittel, U. Kressel, and J. Reichert, "Robust classification of modulation types using spectral features applied to HMM," in *Proc. MILCOM*, Monterey, CA, USA, 1997, pp. 1377–1381.

[13] B. G. Mobasseri, "Constellation shape as a robust signature for digital modulation recognition," in *Proc. IEEE Mil. Commun.*, Atlantic City, NJ, USA, Oct./Nov. 1999, pp. 442–446.

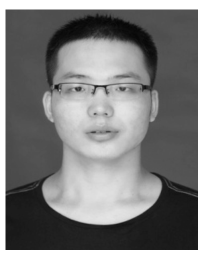
[14] S.-Z. Hsue and S. S. Soliman, "Automatic modulation classification using zero crossing," *IEE Proc. F, Radar Signal Process.*, vol. 137, no. 6, pp. 459–465, 1990.

[15] H. Tao, Z. Zheng-Ou, and L. Xi-Rong, "Modulation classification based on multifractal features," in *Proc. 6th Int. Conf. ITS Telecommun.*, Chengdu, China, vol. 1, 2006, pp. 152–155.
[16] S. L. Wood, M. J. Ready, and J. R. Treichler, "Constellation identification using the radon transform," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, 1988, pp. 1878–1881.
[17] P. Marchand, C. L. Martret, and J.-L. Lacoume, "Classification of linear modulations by a combination of different orders cyclic cumulants," in *Proc. IEEE Signal Process. Workshop Higher-Order Statist.*, Jul. 1997, pp. 47–51.
[18] O. A. Dobre, Y. Bar-Ness, and W. Su, "Robust QAM modulation classification algorithm using cyclic cumulants," in *Proc. IEEE Wireless Commun. Netw. Conf.*, vol. 2, Mar. 2004, pp. 745–748.
[19] J. Lunden, L. Terho, and V. Koivunen, "Classifying pulse compression radar waveforms using time frequency distributions," in *Proc. Conf. Inf. Sci. Syst.*, Mar. 2005, pp. 1–7.
[20] P. R. Milne and P. E. Pace, "Wigner distribution detection and analysis of FMCW and P-4 polyphase LPI waveforms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Orlando, FL, USA, May 2002, pp. IV-3944–IV-3947.
[21] D. B. Copeland and P. E. Pace, "Detection and analysis of FMCW and P-4 polyphase LPI waveforms using quadrature mirror filter trees," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Orlando, FL, USA, May 2002, pp. IV-3960–IV-3963.
[22] J. Lunden and V. Koivunen, "Automatic radar waveform recognition," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 1, pp. 124–136, Jun. 2007.
[23] O. K. Singh, N. Sarada, T. Srikanth, and T. R. Kishore, "Fingerprinting techniques for complex radars," in *Proc. EWCI*, 2012, pp. 466–472.
[24] O. K. Singh, N. Sarada, T. Srikanth, and T. R. Kishore, "Augmented identification system," in *Proc. EWCI*, 2014, pp. 295–300.
[25] T. O. Gulum, P. E. Pace, and R. Cristi, "Extraction of polyphase radar modulation parameters using a Wigner-Ville distribution-Radon transform," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, 2008, pp. 1505–1508.
[26] T. R. Kishore and K. D. Rao, "Automatic intrapulse modulation classification of advanced LPI radar waveforms," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 2, pp. 901–914, Apr. 2017.
[27] A. Ali and F. Yangyu, "Automatic modulation classification using deep learning based on sparse autoencoders with nonnegativity constraints," *IEEE Signal Process. Lett.*, vol. 24, no. 11, pp. 1626–1630, Nov. 2017.
[28] G. J. Mendis, J. Wei, and A. Madanayake, "Deep learning-based automated modulation classification for cognitive radio," in *Proc. IEEE Int. Conf. Commun. Syst. (ICCS)*, Shenzhen, China, Dec. 2016, pp. 1–6.
[29] G. Işık and H. Artuner, "Recognition of radio signals with deep learning neural networks," in *Proc. 24th Signal Process. Commun. Appl. Conf. (SIU)*, Zonguldak, Turkey, 2016, pp. 837–840.
[30] F. Paisana, A. Selim, M. Kist, P. Alvarez, J. Tallon, C. Bluemm, A. Puschmann, and L. DaSilva, "Context-aware cognitive radio using deep learning," in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw. (DySPAN)*, Piscataway, NJ, USA, Mar. 2017, pp. 1–2.
[31] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
[32] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1–9.
[33] S. Huang, Y. Jiang, Y. Gao, Z. Feng, and P. Zhang, "Automatic modulation classification using contrastive fully convolutional network," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1044–1047, Aug. 2019.
[34] D. Bahdanau, J. Chorowski, D. Serdyuk, P. Brakel, and Y. Bengio, "End-to-end attention-based large vocabulary speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Shanghai, China, Mar. 2016, pp. 4945–4949.
[35] C. Raffel and D. P. Ellis, "Feed-forward networks with attention can solve some long-term memory problems," in *Proc. Workshop Track (ICLR)*, 2016, pp. 1–6.
[36] J. H. Lee, K. Y. Kim, and Y. Shin, "Feature image-based automatic modulation classification method using CNN algorithm," in *Proc. Int. Conf. Artif. Intell. Inf. Commun. (ICAIIC)*, 2019, pp. 1–4.
[37] J. D. Rennie, L. Shih, J. Teevan, and D. R. Karger, "Tackling the poor assumptions of naive Bayes text classifiers," in *Proc. ICML*, vol. 3, 2013, pp. 616–623.
[38] J. Lv, L. Zhang, and X. Teng, "A modulation classification based on SVM," in *Proc. IEEE Int. Conf. Opt. Commun. Netw.*, Sep. 2017, pp. 1–3.
[39] Z. Zhang, Y. Li, and X. Zhu, "A method for modulation recognition based on entropy features and random forest," in *Proc. IEEE Int. Conf. Softw. Qual., Rel. Secur. Companion (QRS-C)*, Jul. 2017, pp. 243–246.
[40] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," 2016, *arXiv:1602.04105*. [Online]. Available: https://arxiv.org/abs/1602.04105
[41] A. Ali, F. Yangyu, and S. Liu, "Automatic modulation classification of digital modulation signals with stacked autoencoders," *Digit. Signal Process.*, vol. 71, pp. 108–116, Dec. 2017.

SHUYUAN YANG received the B.S. degree in electrical engineering and the M.S. and Ph.D. degrees in circuit and system from Xidian University, Xi'an, China, in 2000, 2003, and 2005, respectively. Her research interests are intelligent signal processing, machine learning, and image processing. She is currently a IET Senior Member.

CHEN YANG received the B.S. degree from the Xi'an University of Technology. He is currently pursuing the M.S. degree with the School of Artificial Intelligence, Xidian University. His research interests include machine learning and image processing.

DONGZHU FENG received the Ph.D. degree in electrical engineering from the Northwestern Polytechnical University, Xi'an, China, in 2006. She is currently working as an Associate Professor with Xidian University. Her research interests are intelligent signal and image processing.



XIAOYANG HAO received the B.S. degree in software engineering from Xidian University, Xi'an, China, in 2017, where he is currently pursuing the Ph.D. degree in computer science and technology with the School of Artificial Intelligence. His research interests include deep learning and intelligent signal analysis.

MIN WANG received the B.S. degree in auto control and the M.S. and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, China, in 2000, 2003, and 2005, respectively. He is currently working as an Associate Professor with the National Laboratory of Radar Signal Processing, Xidian University. His research interests include intelligent signal, image processing, and statistical learning.