This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2971586, IEEE Access

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Modulated Autocorrelation Convolution Networks for Automatic Modulation Classification based on Small Sample Set

DUONA ZHANG[1], WENRUI DING[2], CHUNHUI LIU[2], HONGYU WANG[1], AND BAOCHANG ZHANG[3].

[1]School of Electronics and Information Engineering, Beihang University, Beijing, China (e-mail: zhangduona@buaa.edu.cn, buaawanghongyu@buaa.edu.cn)
[2]Unmanned Systems Research Institute, Beihang University, Beijing, China (e-mail: ding@buaa.edu.cn, liuchunhui2134@buaa.edu.cn)
[3]School of Automation Science and Electrical Engineering, Beihang University, Beijing, China and Shenzhen Academy of Aerospace Technology, Shenzhen 100083, China (e-mail: bczhang@buaa.edu.cn)

Corresponding author: Chunhui Liu (e-mail: liuchunhui2134@buaa.edu.cn) and Baochang Zhang (e-mail: bczhang@buaa.edu.cn).

This work was supported by the National Defense Basic Scientific Research Project (Grant no. JCKY2017601C006), the Beijing Natural Science Foundation (Grant no. 4204102), National Natural Science Foundation of China (Grant no. 61672079) and Shenzhen Science and Technology Program (Grant no. KQTD2016112515134654).

ABSTRACT For modulation classification, hand-crafted approaches can generalize well from a few samples, yet deep learning algorithms require millions of samples to achieve the superior performance with purely data-driven manner. However for many practical problems only with small sample set (SSS) available, there still remains a challenge for deep learning. In this paper, we employ deep learning to solve the modulation classification task in a more practical setting, particularly suffering from the SSS problem and with low signal-to-noise ratios (SNRs). Novel modulated autocorrelation convolution networks (MACNs) are introduced to capture periodic representation for automatic modulation classification (AMC). In MACNs, modulated communication signals are classified with the periodic local features under an auto-correlation convolution criterion. Modulation filters are utilized to enhance the capacity of the convolution filters and compress the model. On a challenging SSS learning task in low SNRs, MACNs achieve state-of-the-art performance that outperforms the existing algorithms for AMC, while compressing the size of required storage space of convolutional filters by a factor of 8 compared with convolution neural networks (CNNs).

INDEX TERMS Autocorrelation convolution, modulation filters, small sample set, automatic modulation classification, deep learning.

I. INTRODUCTION

SOFTWARE Defined Radio (SDR) has attracted a great deal attention given its ability to provide effective solutions to multipurpose communication devices. Automatic modulation classification (AMC) of SDR has itself led to a growth in interest of classification tasks in the field of the communication [1]. Despite remarkable results in artificial intelligence and machine learning for AMC, two aspects of practical reasons have limited the application of deep learning [2] algorithms on it. First, for classical signal processing algorithms, modulation features are extracted by hand-craft methods to achieve noise reduction, whereas deep learning algorithms learn robust representations in a purely data-driven pattern. Although the effect of deep learning algorithms is superior to hand-craft features, they suffer from poor performance in the application with low signal-to-noise ratios (SNRs). Second, deep learning algorithms have drawn much attention on their capability to extract high-quality feature representations directly from the raw data, thereby achieving favorable performance on object, speech and networking recognition benchmarks [3]–[10]. However, they generally require millions of samples to learn the models of a satisfactory performance which makes them impractical for small sample size (SSS) problem. Particularly, radio signals

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2971586, IEEE Access

are intangible waves with complicated interference which are difficult to collect in a real-world environment.

In the initial stage of development, AMC is accomplished by likelihood-based methods with the prior knowledge of signals. The likelihood function is calculated for all candidate modulations to make decision by maximum likelihood ratio test while minimizing the probability of misclassification. Likelihood-based methods are optimal in terms of classification performance, however they suffer from computational complexity problem. Afterwards, AMC technologies depend heavily on hand-crafted filters for signal processing to extract modulation features. Statistical modulation features such as higher order statistics [11] and cycle-stationary moments [12] are the most widely used features to achieve robust classification for signals with periodic components. Then machine learning or decision theory can be utilized to translate the features into a modulation label. Popular approaches include support vector machines (SVMs) [13], decision trees (DTrees) [14], neural networks (NNs) [15] and ensemble approaches which compromise classifiers to improve performance. Recently, convolution neural networks (CNNs) [16] and long short term memory networks (LSTMs) [17] which rely on back-propagation to optimize large parametric neural network models have become mainstream with the advancement of deep learning.

Despite deep learning algorithms have achieved a superior performance for AMC, they generally require millions of samples in a purely data-driven pattern. The challenges are still significant for deep learning in many real life cases. To cope with SSS and low SNR in AMC, we settle the problems by model compression and essential feature extraction of communication [18]–[20]. Modulation and autocorrelation are rolled into CNNs to construct a novel framework, termed modulated autocorrelation convolution networks (MACNs), which capture periodic representation of communication signals with refining model parameters. On one hand, autocorrelation convolution increases the representational capabilities of the periodic signals and suppresses thermal noise. On the other hand, the compression based on modulation filters can allow us to improve the filter efficiency during the convolution procedure, enabling to complete the training process with a few samples.

The MACNs introduced in this paper are capable of learning a large class of modulation modes with a few samples in low SNRs as shown in Fig. 1. Modulation modes are represented as simple yet essential features—that is, MACNs expressed as modulation modes in a periodic local regularity. Meanwhile, we incorporate modulation filters into convolution so as to significantly reduce the number of parameters. Both the modulation filter and autocorrelation convolution can be jointly optimized and obtained in an end-to-end learning framework, leading to a compact and exclusive deep learning architecture. Thanks to the essential representation with low model complexity, such an architecture is less prone to be overfitting and suitable for periodic communication signals. On a challenging SSS learning task in low SNRs, MACNs reduce the required storage space of CNNs by a factor of 8, while achieving the best performance so far, as compared to the existing algorithms for AMC. In summary, the contributions of this paper are as follows:

(1) MACNs can take advantages of both model-based and data-driven methods to improve the representational capabilities of modulation. A new framework is proposed for AMC to capture periodic characteristics via autocorrelation convolution that is insensitive to thermal noise. The experimental results indicate that MACNs achieve state-of-the-art performance.

(2) The modulation filters are employed to compress the model to avoid over-fitting and improve the filter efficiency during the convolution procedure, leading to a new architecture to calculate CNNs. As a result, the modulated networks compress the convolutional filters by a factor of 8, compared with the standard CNNs, while the performance is comparable to the original networks.

The rest of this paper is organized as follows. Section II describes the details of the modulated autocorrelation convolution networks. The experimental results are presented in Section III. Finally, we give the conclusion in Section VI.

II. MODULATED AUTOCORRELATION CONVOLUTION NETWORKS

We design the architectures in MACNs based on the autocorrelation convolution and modulation filters. Autocorrelation convolution is particularly designed to increase the representational capabilities of the periodic signals in the end-to-end framework. The convolutional feature maps are processed by correlation at multiple intervals which leads to the random noise suppression even with low SNRs. To alleviate the disturbance caused by overfitting, the model compression based on the modulation filters is deployed to improve the convolution efficiency. With two measures mentioned above, MACNs achieve advanced performance with a few samples in low SNRs.

A. AUTOCORRELATION CONVOLUTION IN MACNS

In order to enhance CNNs to learn periodic representation, we introduce novel autocorrelation convolution layers in MACNs. Two aspects are considered: the convolutional feature maps are intensified based on periodic learning; the random thermal noise is suppressed by correlation which is distinct from periodic signal. Autocorrelation is a mathematical representation of the degree of similarity (Pearson Correlation Coefficient) between a given time series and a lagged version of itself over successive time intervals. It is the same as calculating the correlation between two different time series, except autocorrelation uses the same time series twice: once in its original form and once lagged one or more time periods. For a time series $Y = \{y(1), \cdots, y(t)\}$, the autocorrelation coefficient is given by ($\tau < t$):

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2971586, IEEE Access

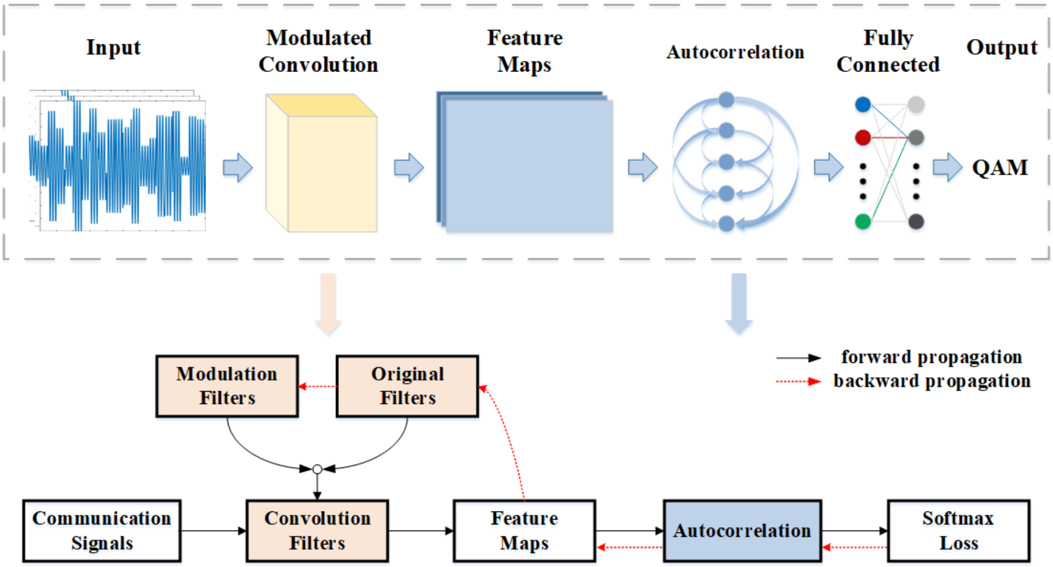

FIGURE 1. Pipeline of the modulated autocorrelation convolution networks (MACNs). MACNs are implemented by the autocorrelation convolution and the modulation filters. As a result, MACNs achieve state-of-the-art performance that outperforms the existing algorithms for automatic modulation classification (AMC) in low signal-to-noise ratios (SNRs), while compressing the convolutional filters by a factor of 8 compared with convolution neural networks (CNNs).

$$\phi(\tau) = \frac{1}{(t-\tau)\sigma^2}\sum_{\tau=1}^{t-\tau}(y(t)-\mu)\cdot(y(t+\tau)-\mu) \quad (1)$$

where μ and σ^2 are the mean and variance of the time series. To simplify the deep structure, the autocorrelation in this paper can be estimated as:

$$\phi(\tau) = \frac{1}{N}\sum_{n=0}^{N-1} c(n)\cdot c(n+\tau) \quad (2)$$

where $c(n)$ is the convolutional feature map of communication signal, τ is lag number ($\tau = 1, 2, \cdots, N-1$) and n is time step of discrete feature map. The property of $\phi(\tau)$ is that $\phi(\tau)$ shows a peak value when $c(n)$ is similar with $c(n+\tau)$. If $c(n)$ has a period of T, then $\phi(\tau)$ has peaks at gT where g is an integer. Essentially, autocorrelation coefficient is symmetric in τ, that is, $\phi(\tau) = \phi(-\tau)$. Accordingly, only half the result of $\phi(\tau)$ is needed ($\tau \geq 0$) whose dimension is same as the convolutional feature map.

As mentioned above, $c(n)$ is the convolutional feature map of communication signal $s(n)$ that is given by:

$$c(n) = C(s(n)) = C(x(n) + w(n)) \quad (3)$$

where C is convolution procedure. $x(n)$ is a clean communication signal and $w(n)$ is the background noise. In this case, we have an autocorrelation coefficient given by:

$$\begin{aligned}\phi(\tau) &= C(s(n)\cdot s(n+\tau)) \\ &= C(\phi_{xx}(\tau) + 2\phi_{xw}(\tau) + \phi_{ww}(\tau))\end{aligned} \quad (4)$$

where $\phi_{xx}(\tau)$ is the autocorrelation coefficient of $x(n)$, $\phi_{xw}(\tau)$ is the cross-correlation function of $x(n)$ and $w(n)$, $\phi_{ww}(\tau)$ is the autocorrelation coefficient of $w(n)$. For communication signals, $x(n)$ does not correlate with $w(n)$, then $\phi_{xw}(\tau) = 0$. Furthermore, $w(n)$ is uncorrelated, then $\phi_{ww}(\tau) = 0$. In such a case, the relation:

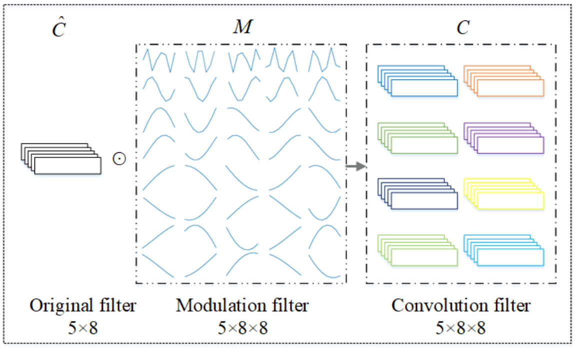

FIGURE 2. Model compression based on the modulation filters.

$$\phi(\tau) = C(\phi_{xx}(\tau)) \quad (5)$$

is valid. Based on these properties, the autocorrelation coefficient provides robust performance against noise with periodicity in low SNRs. Moreover, autocorrelation convolution has the ability of learning local characteristics which inherits from the unique network structure of CNNs.

B. MODEL COMPRESSION OF MACNS WITH MODULATION FILTERS

Owing to the one-dimensional waveform, the $2D$ filters across all convolution layers with the size of $K \times W$ are deployed, which have K planes and each of the planes is a $1 \times W$-sized $1D$ filter. For multiple convolution filters, the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2971586, IEEE Access

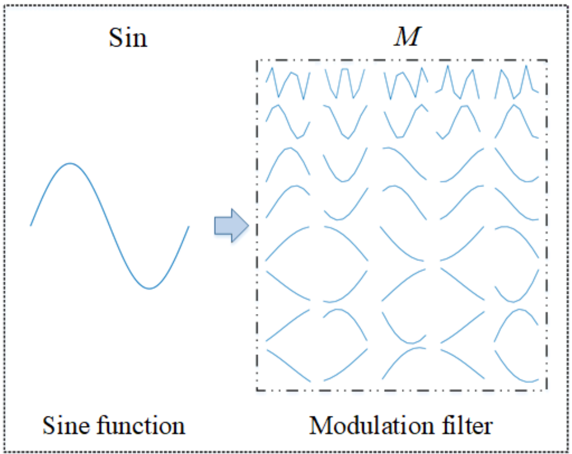

FIGURE 3. Generation of modulation filters.

size of convolution filter in MACNs is $K \times W \times N$. Note that the dimensions of K and W are exchanged in practice, by doing so, we can easily implement MACNs in the Tensorflow platform. To reduce the number of parameters, we introduce a modulated process based on the modulation filters (Fig. 2). The modulation filter is a constant matrix serving as the weight of the original convolution filters, which is also with the size of $K \times W \times N$. The operation $\odot$ is defined as:

$$C = \hat{C} \odot M = \sum_{j}^{K} \hat{C} \cdot M_j \tag{6}$$

where C is the convolution filter in MACNs, $\hat{C}$ is the original filter, M is the modulation filter, M_j is the jth plane of the modulation filter, $\cdot$ is the element-wise multiplication operator, also named Schur product operation. Note that the size of $\hat{C}$ is $K \times W$. In this paper, W is defined as 5, K and N are both defined as 8 and each layer shares only one modulation filter, leading to significant reduction of the network model. In addition, the operation $\odot$ results in a new matrix which is elaborated as:

$$Q_{ij} = \hat{C} \cdot M_j \tag{7}$$

$$Q_i = \{Q_{i1}, \cdots, Q_{iN}\} \tag{8}$$

As mentioned above, the high-performance modulation filter is the critical success factor. In essence for modulation, the operation of convolution and correlation is similar, the convolution operation is defined as:

$$\varphi(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) \cdot y(\tau - n) \tag{9}$$

Compared to Eq. 2, the convolution operation is almost identical to the cross-correlation operation which has similar attributes, but the filter is reversed in convolution. To achieve the correlation between convolution filter and signal waveform, the modulation filter is pre-defined based on sine function for the original filter. Meanwhile, multiple initial phases and sampling frequencies are utilized to generate the modulation filter. The modulation filter is calculated as (Fig. 3):

$$m(t) = \sum_{t=0}^{2\pi} \sin(2\pi f t + \theta) \cdot \delta\left(t - n\frac{2\pi}{f_s}\right) \tag{10}$$

where f_s is sampling rate, θ is initial phase and δ is sampling pulse. In this paper, $\theta = 0, 2/5\pi, 4/5\pi, 6/5\pi, 8/5\pi$ and $f_s = 4f, 8f, 16f, 32f$.

The advantage of modulation filter is that the number of input and output channels in each feature map are the same, making the filter to be replicated and implemented easily.

C. DATASET GENERATION APPROACH

The datasets are generated for MACNs investigation by building upon the open-source software development toolkit GNU Radio [21] and the universal software radio peripheral (USRP) [22] B210 SDR.12 different digital modulators are used that cover a range of single carrier modulation schemes. Several propagation scenarios are considered in the context of this work. Accordingly, the communication signal $s(n)$ in Eq. 3 can be expressed as:

$$s(n) = (A_c + jA_s)\, e^{j[2\pi(f_c + \Delta f_c)t + \theta]} \cdot \delta(t - \varepsilon - nT) \tag{11}$$

where A_c and A_s are the amplitudes on in-phase and quadrature branch of the signal respectively, f_c and Δf_c are the carrier frequency and its offset, θ is the initial phase, δ is the digital sampling pulse and ε is the timing offset. The background noise is additive white Gaussian noise (AWGN) in this paper and the signal-to-noise ratio (SNR) is -20∼10 dB.

We use the wireless channel on the 845 MHz unmanned aerial vehicle (UAV) band, the symbol rate on 200 KHz and off-tune the signal by around 1 MHz to avoid direct current (DC) impairment. The sampling rate in this work is 10 MHz. When modeling a wireless channel, many compact stochastic models for propagation effects can be used [23]. Primary impairments in wireless channel include:

(1) Carrier frequency offset: carrier frequency offset dues to local oscillators and motion.

(2) Delay Spread: delay spread dues to delayed reflection, diffraction and diffusion of emissions on multiple paths (Rayleigh fading in this paper).

(3) Thermal Noise: additive white-noise impairment at the receiver dues to physical device (AWGN in this paper).

For each sample in the data sets, we independently draw a value for each of the variables shown in Table 1. This results in a random channel initialization for each sample.

The specific modulations considered within the dataset types are as follows:

OOK, 4ASK, 8ASK, 2FSK, 4FSK, 8FSK, BPSK, QPSK, 8PSK, 16QAM, 64QAM, 128QAM.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2971586, IEEE Access

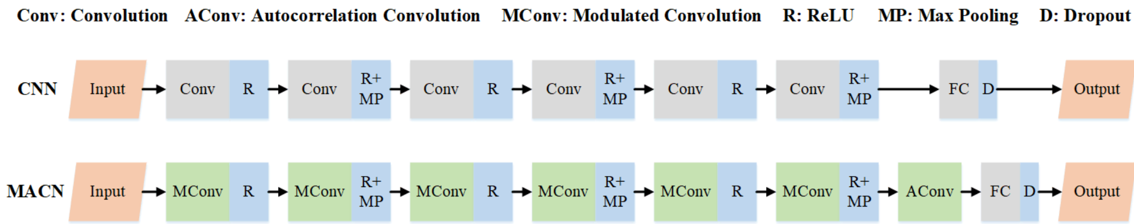

FIGURE 4. Network architectures of CNNs and MACNs.

TABLE 1. Random Variable Initialization

Random Variable	Distribution
Δf_c	$N\left(0, \sigma_{clk}\right)$
θ	$U\left(0, 2\pi\right)$
ε	$N\left(0, \sigma_{clk}\right)$
Multi-path Fading	$\sum_i \delta\left(t - Rayleigh_i\left(\tau\right)\right)$

The dataset in this work is publicly available in https://github.com/bczhangbczhang.

D. IMPLEMENTATION DETAILS

In the experiments, we use MACNs with six modulated convolutional layers on the dataset. MACNs are effective for the periodic signal and the modulation process can be applied to any CNNs. The size of each modulation filter and convolutional filter is $5 \times 8 \times 8$. We adopt Max-pooling and rectified linear unit (ReLU) after the convolution layers, and a dropout layer after the fully connected layer is used. The Adam optimizer [24] is used and the initial value of learning rate is set to 0.001 which attenuates 0.05 every 200 cycles. Fig. 4 shows the details of the network architectures of MACNs in this paper.

III. EXPERIMENTS

In this section, the results of the experiments are included to compare MACNs with classical approaches. The experiments are implemented with few training samples (1, 5, 10, 20 in this paper) for each modulation mode (described in section 2.C). Meanwhile, the impairments of the wireless channel are taken into consideration in the experiments.

A. PARAMETERS EVALUATION

The number of convolution filters and layers in CNNs are related to the performance of modulation mode classification. In order to achieve the optimal parameters, the effects of filter number and convolution layer are evaluated on the dataset for CNNs without autocorrelation. With different values of parameters, the performance of the CNNs is shown in Fig. 5 and Fig. 6.

B. SSS LEARNING RESULTS

MACNs are evaluated in various SSS learning tasks. The experimental results are summarized by Fig. 7 $\sim$ Fig. 10, and the modulation modes of all behavioral experiments are described in section 2.C. In this paper, we choose 20, 10,

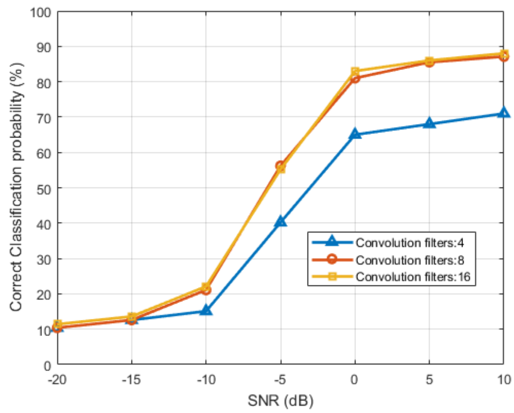

FIGURE 5. Classification accuracy of CNNs with convolution filters 4, 8 and 16.

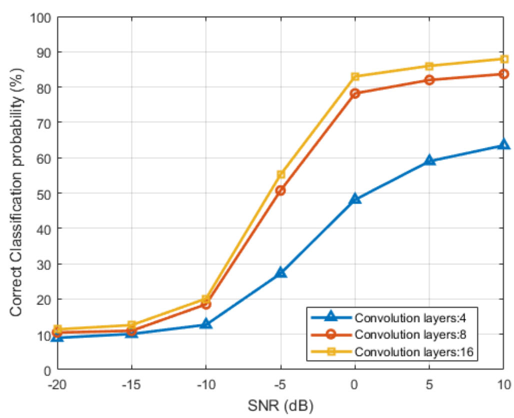

FIGURE 6. Classification accuracy of CNNs with convolution layers 4, 8 and 16.

5 and 1 training samples per modulation learning tasks to evaluate MACNs for the SSS problem with SNR from -20 dB to 10 dB. MACNs achieve average classification accuracy of 53.85%, 45.57%, 40.62%, and 32.16% with 20, 10, 5, and 1 training samples per modulation respectively, and obtain 59.19% on large dataset with 4096 training samples per modulation.

Table 2 shows the parameters and delay results of MACNs and CNNs with convolution filter size of 5. With the little difference in time delay, the parameters of the MACNs are

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2971586, IEEE Access

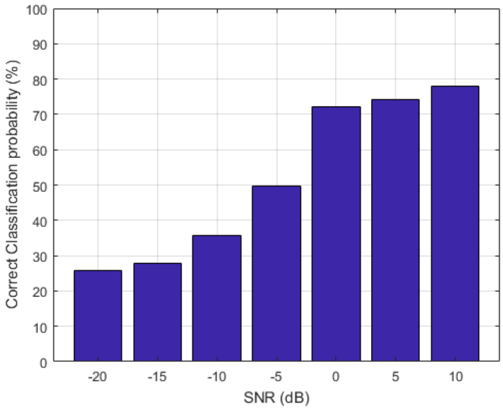

FIGURE 7. 20 training samples per modulation learning results of MACNs.

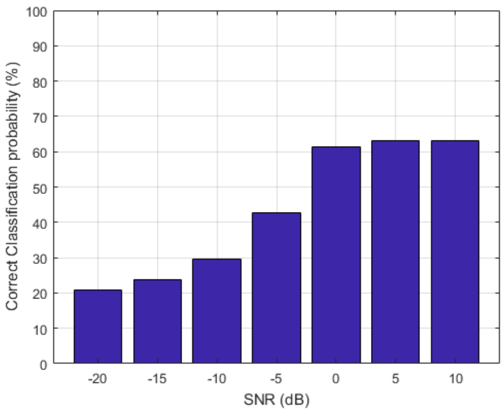

FIGURE 8. 10 training samples per modulation learning results of MACNs.

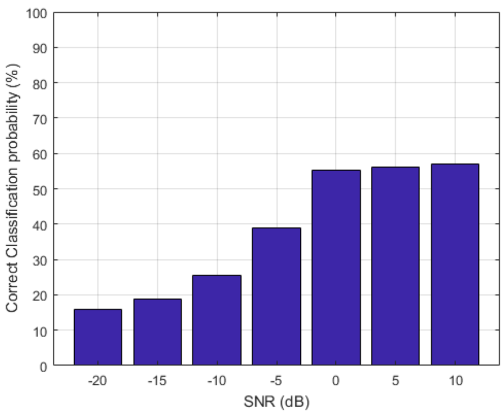

FIGURE 9. 5 training samples per modulation learning results of MACNs.

much fewer, but the performance is better.

Owing to large amounts of parameters, conventional CNNs suffer from overfitting caused by the SSS problem, which affects the generalization performance of our model on the new data. To address this issue, few efficient methods

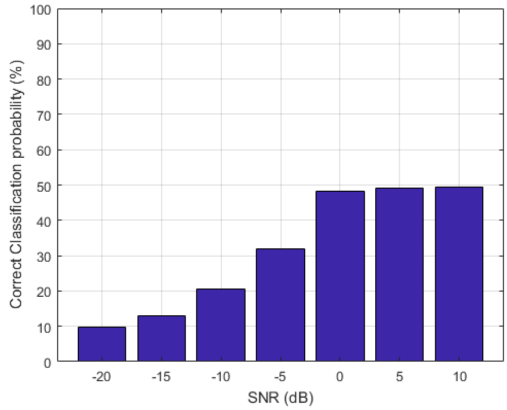

FIGURE 10. 1 training sample per modulation learning results of MACNs.

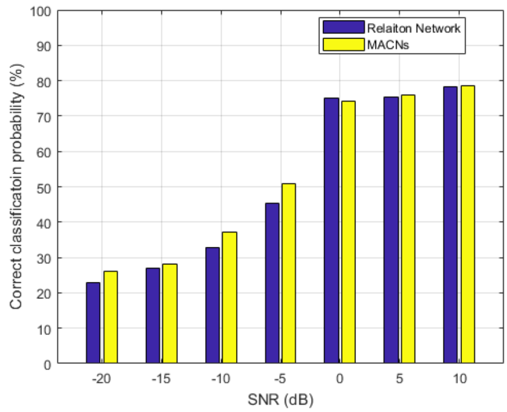

FIGURE 11. Comparison between MACNs and the relation network with 20 training samples for each modulation mode.

are introduced for the SSS problem. Fig. 11 shows the comparison between MACNs and the relation network [25] with 20 training samples for each modulation mode.

C. COMPARISON RESULTS ON VARIOUS CLASSIFICATION ALGORITHMS

After a long period of development, there are many mature techniques for AMC as described in section 1. Owing to the inapplicability of the popular deep learning algorithms for SSS learning, MACNs are compared with conventional algorithms in the case of sufficient samples (4096 samples per modulation mode). Fig. 12∼Fig. 14 show the performance with channel impairments of MACNs in comparison with high-order moments (C40), wavelets (db10) [26], CNNs and long short-term memory networks (LSTMs) based on the same dataset with SNRs -20∼10 dB. Apparently, MACNs outperform the previous algorithms that rely on prior knowledge of communication or purely data-driven patterns. The basic nature of hand-craft features is open-loop operation which is limited by degradation in performance caused by interference. By contrast, deep learning solves this problem

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2971586, IEEE Access

TABLE 2. Parameters and Delay of MACNs and CNNs with Convolution Filter Size of 5

	Convolution filters	Parameters (M)	Training time (s)	Testing time (s)
CNNs (Conv layers 3)	8 16	1.540 3.084	0.032 0.041	0.011 0.014
CNNs (Conv layers 6)	8 16	1.986 4.072	0.048 0.067	0.016 0.019
MACNs (Conv layers 3)	8 16	0.202 0.390	0.054 0.071	0.012 0.015
MACNs (Conv layers 6)	8 16	0.291 0.576	0.093 0.166	0.029 0.033

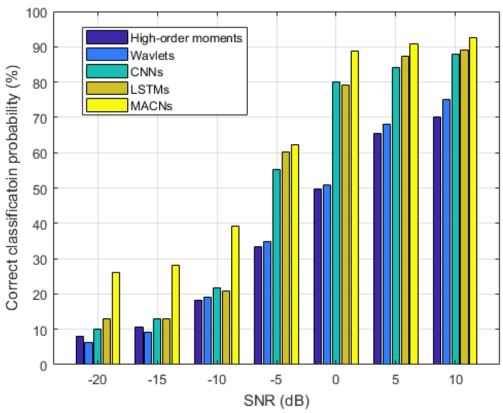

FIGURE 12. Comparison results on various classification algorithms.

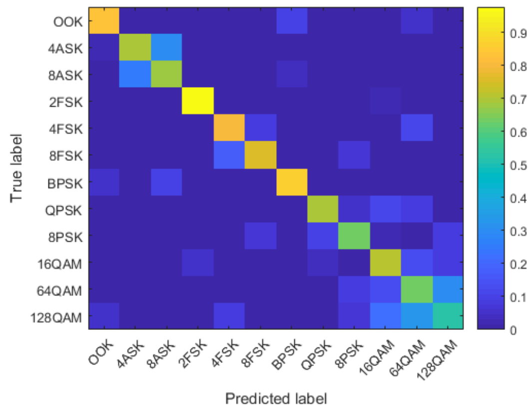

FIGURE 14. Confusion matrix of MACNs with SNR -10dB.

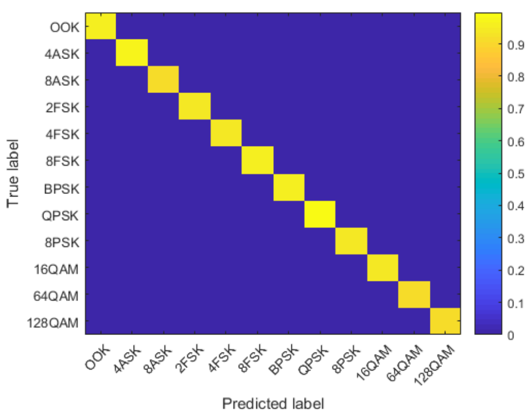

FIGURE 13. Confusion matrix of MACNs with SNR 10dB.

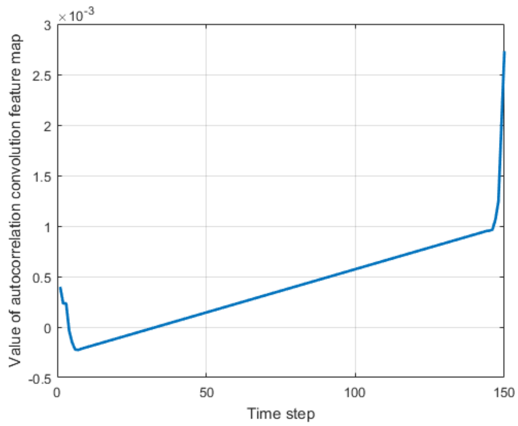

FIGURE 15. The 128QAM signal feature map of MACNs with SNR 0 dB.

with back propagation. Although the conventional neural networks achieve high performance on object and speech recognition benchmarks, they ignore the periodic feature in communication. In this paper, a periodic learned framework with back propagation is more applicable to modulation classification.

Table 3 shows the classification result of MACNs for each modulation mode with SNR 5dB and -5dB. The experimental results indicate that the higher-order phase modulations are confused in low SNRs.

Fig. 15 and Fig. 16 show the autocorrelation convolution feature map of MACNs with the same 128QAM signal at SNR 0 dB and 10dB. The approximate results confirm that MACNs can suppress thermal noise.

D. CHANNEL IMPAIRMENTS

In real world scenario, wireless channels are impaired by a number of factors. While AWGN is widely used for wireless signals in simulation, the factors described in section 2.C are present almost universally. It is necessary to consider the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2971586, IEEE Access

TABLE 3. Classification Result of MACNs for Each Modulation Mode with SNR 5dB and -5dB

	OOK	4ASK	8ASK	2FSK	4FSK	8FSK	BPSK	QPSK	8PSK	16QAM	64QAM	128QAM
5dB (%)	95.1	95.6	95.6	93.1	91.0	89.6	92.1	89.6	89.9	90.0	85.3	84.6
-5dB (%)	75.3	72.2	72.0	69.8	63.3	59.9	65.3	60.0	54.1	52.6	51.2	50.9

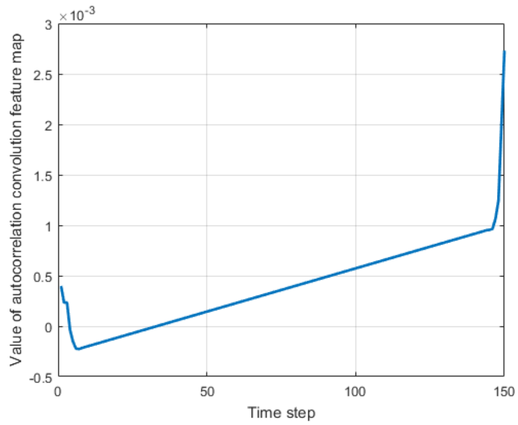

FIGURE 16. The 128QAM signal feature map of MACNs with SNR 10 dB.

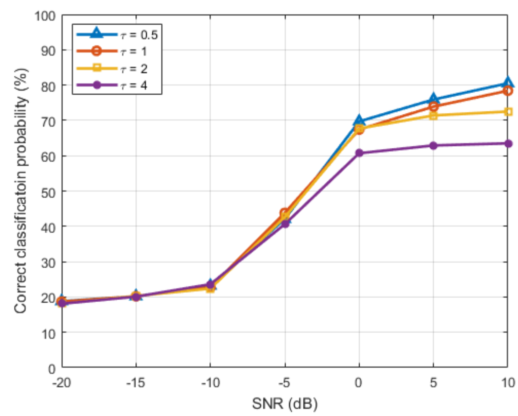

FIGURE 18. Multi-path fading results of MACNs with SNR -20∼10 dB.

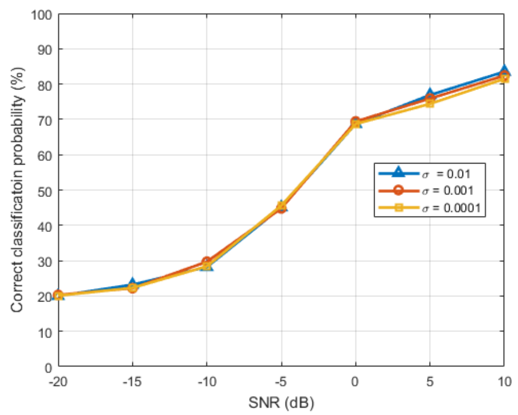

FIGURE 17. Frequency offset results of MACNs with SNR -20∼10 dB.

effects of such impairments to the performance of MACNs. In Fig. 17 and Fig. 18, we show the performance of MACNs under the considered impairment model. This includes background noise (AWGN), frequency offset (σ_{clk}) and multi-path fading ($Rayleigh_i\,(\tau)$).

IV. CONCLUSION

We have developed a novel deep learning model, termed MACNs, which capture periodic representation for SSS in AMC. MACNs are implemented by the autocorrelation convolution and the proposed filter modulation operation. In MACNs, the signals are represented as periodic local features which best classify modulation modes under an autocorrelation convolution criterion. Modulation filters are used to enhance the capacity of the convolutional filters and compress the model. As a result, MACNs achieve advanced performance that outperforms state-of-the-art AMC algorithms in low SNRs and compress the convolutional filters by a factor of 8 compared with CNNs.

As a general learning framework, MACNs can be utilized for efficient feature extraction to obtain periodic characteristics, which represent the essential attribute of time-series signal in the real world. We will explore the possibility to deal with the signals in the UCR time series archive [27] by MACNs and further enhance the model performance in the future.

ACKNOWLEDGMENT

This work was supported by the National Defense Basic Scientific Research Project (Grant no. JCKY2017601C006), the Beijing Natural Science Foundation (Grant no. 4204102), National Natural Science Foundation of China (Grant no. 61672079) and Shenzhen Science and Technology Program (Grant no. KQTD2016112515134654).

APPENDIX

1) Back-propagation Updating

In MACNs, the original filters $\hat{C}$ are learned and updated. We define the gradient of the original filters $\hat{C}$ as $\gamma_{\hat{C}}$:

$$\gamma_{\hat{C}} = \frac{\partial L}{\partial \hat{C}} = \frac{\partial L}{\partial Q} \cdot \frac{\partial Q}{\partial \hat{C}} = \sum_j \frac{\partial L}{\partial Q_{ij}} \cdot M_j \tag{12}$$

$$\hat{C} \leftarrow \hat{C} - \eta \gamma_{\hat{C}} \tag{13}$$

where L is the loss function, η is the learning rate. The above derivations show that MACNs are learnable with the back-propagation algorithm.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2971586, IEEE Access

REFERENCES

[1] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: classical approaches and new trends," IET communications, vol. 1, no. 2, pp. 137–156, 2007.

[2] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural networks, vol. 61, pp. 85–117, 2015.

[3] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," IEEE Transactions on Pattern Analysis & Machine Intelligence, no. 4, pp. 509–522, 2002.

[4] A. Hannun, C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, R. Prenger, S. Satheesh, S. Sengupta, A. Coates et al., "Deep speech: Scaling up end-to-end speech recognition," arXiv preprint arXiv:1412.5567, 2014.

[5] G. Ding, Y. Guo, K. Chen, C. Chu, J. Han, and Q. Dai, "Decode: deep confidence network for robust image classification," IEEE Transactions on Image Processing, 2019.

[6] G. Wu, J. Han, Z. Lin, G. Ding, B. Zhang, and Q. Ni, "Joint image-text hashing for fast large-scale cross-media retrieval using self-supervised deep learning," IEEE Transactions on Industrial Electronics, 2018.

[7] G. Wu, J. Han, Y. Guo, L. Liu, G. Ding, Q. Ni, and L. Shao, "Unsupervised deep video hashing via balanced code for large-scale video retrieval," IEEE Transactions on Image Processing, vol. 28, no. 4, pp. 1993–2007, 2018.

[8] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescapé, "Mobile encrypted traffic classification using deep learning: Experimental evaluation, lessons learned, and challenges," IEEE Transactions on Network and Service Management, 2019.

[9] P. Wang, F. Ye, X. Chen, and Y. Qian, "Datanet: Deep learning based encrypted network traffic classification in sdn home gateway," IEEE Access, vol. 6, pp. 55 380–55 391, 2018.

[10] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescapè, "Mimetic: Mobile encrypted traffic classification using multimodal deep learning," Computer Networks, vol. 165, p. 106944, 2019.

[11] A. Abdelmutalab, K. Assaleh, and M. El-Tarhuni, "Automatic modulation classification based on high order cumulants and hierarchical polynomial classifiers," Physical Communication, vol. 21, pp. 10–18, 2016.

[12] W. A. Gardner and C. M. Spooner, "Signal interception: performance advantages of cyclic-feature detectors," IEEE Transactions on Communications, vol. 40, no. 1, pp. 149–159, 1992.

[13] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," IEEE Intelligent Systems and their applications, vol. 13, no. 4, pp. 18–28, 1998.

[14] J. R. Quinlan, "Induction of decision trees," Machine learning, vol. 1, no. 1, pp. 81–106, 1986.

[15] S. Haykin, Neural networks: a comprehensive foundation. Prentice Hall PTR, 1994.

[16] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in International conference on engineering applications of neural networks. Springer, 2016, pp. 213–226.

[17] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin, "Deep learning models for wireless signal classification with distributed low-cost spectrum sensors," IEEE Transactions on Cognitive Communications and Networking, vol. 4, no. 3, pp. 433–445, 2018.

[18] Y. Cheng, D. Wang, P. Zhou, and T. Zhang, "A survey of model compression and acceleration for deep neural networks," arXiv preprint arXiv:1710.09282, 2017.

[19] H. Kim and K. G. Shin, "In-band spectrum sensing in cognitive radio networks: energy detection or feature detection?" in Proceedings of the 14th ACM international conference on Mobile computing and networking. ACM, 2008, pp. 14–25.

[20] Y. Wang and Q. Yao, "Few-shot learning: A survey," arXiv preprint arXiv:1904.05046, 2019.

[21] E. Blossom, "Gnu radio: tools for exploring the radio frequency spectrum," Linux journal, vol. 2004, no. 122, p. 4, 2004.

[22] O. Holland, H. Bogucka, and A. Medeisis, Opportunistic spectrum sharing and white space access: The practical reality. John Wiley & Sons, 2015.

[23] A. Goldsmith, Wireless communications. Cambridge university press, 2005.

[24] S. Bock, J. Goppold, and M. Weiß, "An improvement of the convergence proof of the adam-optimizer," arXiv preprint arXiv:1804.10587, 2018.

[25] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1199–1208.

[26] W. Dan, G. Xuemai, and G. Qing, "A new scheme of automatic modulation classification using wavelet and wsvm," 2005.

[27] H. A. Dau, A. Bagnall, K. Kamgar, C.-C. M. Yeh, Y. Zhu, S. Gharghabi, C. A. Ratanamahatana, and E. Keogh, "The ucr time series archive," arXiv preprint arXiv:1810.07758, 2018.

DUONA ZHANG received the B.S. degree in Electronic and Information Engineering from Changchun University of Science and Technology, Changchun, China, in 2014, and the M.S. degree in Electronic and Communication Engineering from Beihang University, Beijing, China, in 2017. He is currently a Ph.D. student at the Communication and information system, Beihang University. His current research interests include signal processing, pattern recognition and machine learning.

WENRUI DING has a doctorate degree in Electrical and Information Engineering from Beihang University. She is currently in charge of information transmission and processing data link in the Unmanned System Research Institute in Beihang University. Her current research interests include the command and control of UAV, image processing, and pattern recognition.

CHUNHUI LIU received the B.S. degree, the M.S. degree and the Ph.D. from Beihang University in 2006, 2009 and 2017 respectively. He is now a research assistant in Key laboratory of Advanced Technology of Intelligent Unmanned Flight System of Ministry of Industry and Information Technology at Beihang University. His research interests include computer vision and signal processing.

HONGYU WANG received B.S. degree (2019) and is currently a master candidate with School of Electronics and Information Engineering, Beihang University, Beijing, China. Her research interests focus on the area of few shot learning and Automatic Modulation Recognition.

BAOCHANG ZHANG received the B.S., M.S. and Ph.D. degrees in Computer Science from Harbin Institute of the Technology, Harbin, China, in 1999, 2001, and 2006, respectively. From 2006 to 2008, he was a research fellow with the Chinese University of Hong Kong, Hong Kong, and with Griffith University, Brisban, Australia. Currently, he is an associate professor with the Science and Technology on Aircraft Control Laboratory, School of Automation Science and Electrical Engineering, Beihang University, Beijing, China. He was supported by the Program for New Century Excellent Talents in University of Ministry of Education of China. His current research interests include pattern recognition, machine learning, face recognition, and wavelets.