Spectrum Analysis and Convolutional Neural Network for Automatic Modulation Recognition

Yuan Zeng, Meng Zhang, Fei Han, Yi Gong, *Senior Member, IEEE*, and Jin Zhang

***Abstract*—Recent convolutional neural networks (CNNs)-based image processing methods have proven that CNNs are good at extracting features of spatial data. In this letter, we present a CNN-based modulation recognition framework for the detection of radio signals in communication systems. Since the frequency variation with time is the most important distinction among radio signals with different modulation types, we transform 1-D radio signals into spectrogram images using the short-time discrete Fourier transform. Furthermore, we analyze statistical features of the radio signals and use a Gaussian filter to reduce noise. We compare the proposed CNN framework with two existing methods from literature in terms of recognition accuracy and computational complexity. The experiments show that the proposed CNN architecture with spectrogram images as signal representation achieves better recognition accuracy than existing deep learning-based methods.**

***Index Terms*—Modulation recognition, convolutional neural network, time-frequency analysis, noise reduction.**

I. Introduction

In COMMUNICATION systems, transmitted signals are generally modulated with different modulation methods for efficient data transmission. As an intermediate process between signal detection and signal demodulation, automatic modulation recognition (AMR) provides modulation information of signals and plays a key role in practical civilian and military applications, such as cognitive radio, signal recognition, threat assessment and spectrum monitoring.

In the last few decades, a large number of algorithms have been developed for AMR. In general, AMR algorithms can be categorized into two classes: likelihood based method and feature based method. The likelihood based approaches use probability theory, hypothesis testing theory and a proper decision criterion to solve modulation recognition problems [1], while feature based approaches perform feature extraction and classification. In feature based approaches, recognition performance is proportional to the number of manually designed features. Various statistical features of the instantaneous amplitude, phase and frequency have been used to classify modulation types, such as high-order statistics (HOS) [2] and cyclostationary characteristics [3]. For the classification process, the existing classifiers include decision tree algorithms [4] and machine learning algorithms, such as support vector machine [5] and artificial neural network [6].

Recently, deep learning as a powerful machine learning method has achieved great success in image classification and speech recognition, etc. Deep learning based method uses a cascade of multiple layers of nonlinear processing units for feature extraction and transformation. It can automatically optimize the extracted features for minimizing classification error. Deep learning approaches have also been applied in modulation recognition. The paper [7] surveys the emerging applications of deep learning in radio signal processing and uses GNU radio to generate an open data set of modulated signals with in-phase and quadrature (IQ) information for modulation recognition. O'Shea *et al.* [8] study the adaptation of convolutional neural networks (CNNs) to the data set in [7] and compares the recognition performance of the proposed CNN against those of the expert cyclic moment features based methods. Later, the paper [9] makes a comparison between CNN, residual networks, inception modules, convolutional long short-term deep neural networks based on the data set in [7], and experiment results show that modulation recognition performance is not limited by network depth. Furthermore, the paper [10] proposes a pre-processing signal representation that leverages the IQ information and HOS feature of the modulated signals to improve the recognition performance of their presented deep learning architectures.

CNNs are good at extracting features of spatial data and have shown significant results in image processing, such as image classification and semantic segmentation. Images of spectrum features have been exploited to modulation recognition. The paper [11] uses ambiguity function (AF) images as signal representation and performs modulation recognition using fine-tuning stacked sparse autoencoder [12]. The method in [13] was proposed to automatically recognize modulation types using spectral correlation function (SCF) as signal representation and deep belief network as classifier. In this letter, we present a time-frequency analysis of the modulated signals, where one dimensional signals are transformed into two dimensional spectrogram images by using the short-time discrete Fourier transform, and we design a CNN architecture to

Manuscript received December 25, 2018; revised February 13, 2019; accepted February 14, 2019. Date of publication February 18, 2019; date of current version June 19, 2019. This work was supported in part by National Natural Science Foundation of China under Grant 61531009 and Shenzhen Science and Technology Program under Grant JCYJ20170817110410346. The associate editor coordinating the review of this paper and approving it for publication was X. Chu. *(Corresponding author: Yi Gong.)*

Y. Zeng is with the Academy for Advanced Interdisciplinary Studies, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: zengy3@sustech.edu.cn).

M. Zhang, F. Han, and Y. Gong are with the Shenzhen Engineering Laboratory of Intelligent Information Processing for IoT, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: gongy@sustech.edu.cn).

J. Zhang is with the Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: zhangj4@sustech.edu.cn).

Digital Object Identifier 10.1109/LWC.2019.2900247

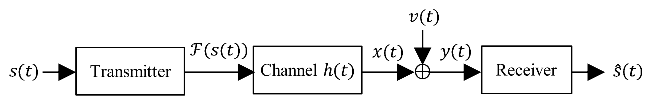

Fig. 1. A simple communication system with a transmitter and a receiver connected through a channel.

recognize modulations automatically. We name the algorithm spectrum CNN (SCNN). In addition, we use a Gaussian filter to suppress noise, and refer to the method as SCNN2. We evaluate the recognition performance of the proposed methods using the public data set in [7] with 11 common used modulation types, and compare the recognition accuracy of the SCNN2 with those of the methods in [8] and [10]. Furthermore, we evaluate the effectiveness of the three representation methods: the spectrogram image, AF image and SCF image using CNN based recognition performance comparison. Moreover, the computational complexity of the methods in experiments is analyzed in terms of memory consumption, learned parameters and training time per signal.

II. Time-Frequency Analysis and Noise Reduction

A. Time-Frequency Analysis

Let us consider a simple communication system in Fig. 1, which consists of a transmitter, a channel and a receiver. Let $s(t)$ denote a transmission symbol that to be transmitted to the receiver. The transmission symbol $s(t)$ is first converted into transmission signal using a modulation function $\mathcal{F}$, and the signal then is transmitted to the receiver via a communication channel $h(t)$. Let $y(t)$ denote the observed signal of the transmission symbol $s(t)$ at the receiver. The received signal $y(t)$ is then given as $y(t) = x(t)+v(t)$, where $x(t) = \mathcal{F}(s(t)) * h(t)$ is the received clean signal and $v(t)$ is the additive white Gaussian noise. Given the observed signal $y(t)$, modulation recognition aims to predict the modulation function $\mathcal{F}$, and thus to provide modulation information for estimating the transmitted symbol $s(t)$ from the observed signal $y(t)$. Let $y(n)$ denote the discrete-time observed signal at time-sampling index n. $y(n)$ can be obtained by sampling the continuous-time signal $y(t)$ at time $\frac{n}{f_s}$, i.e., $y(n) = y(t)|_{t=n/f_s}$, and $-\infty < n < +\infty$.

In this letter, we use spectrogram as a visual representation of the spectrum of frequencies of the observed signals, which vary with time. The spectrogram is obtained by computing the squared magnitude of the short-time discrete Fourier transform (STFT) of the observed signals. Let $w(\cdot)$ denote a window function of length J and let K be the window shift. The observed signals are windowed and transformed into the frequency domain by applying the STFT, that is,

$$Y(f, m) = \sum_{n=mK+1}^{mK+J} y(n)w(n - mK)e^{-j\omega_f(n-mK)}, \quad (1)$$

where $Y(f,m)$ is the DFT coefficient at frequency-bin index f and discrete-time frame index m, and $\omega_f = 2\pi f/J$ is the discrete frequency variable at f. The spectrogram is then given by $\tilde{Y}(f, m) = |Y(f, m)|^2$. $\tilde{Y}(f, m)$ is a mixed time-frequency representation of $y(n)$, since each location on $\tilde{Y}(f,m)$ corresponds to a point in frequency and time.

B. Noise Reduction

Performance of modulation recognition can be disturbed severely by the additive noise $v(t)$. Since transmitted messages are base-band signal while power spectral density of noise is independent of frequency and uniformly distributed throughout the frequency domain, performing noise reduction directly on spectrogram images with Gaussian filter only produces blurred images and has limited capability of frequency rejection. Here, we use low-pass filter to reduce noise of the signal $y(t)$ before the time-frequency analysis. To reduce noise of the observed signals, we design a Gaussian filter, that is, $\hat{x}(n) = y(n)G(n)$, where $\hat{x}(n)$ is the filtered signal and the Gaussian filter $G(n)$ is given by $G(n) = \frac{1}{\sqrt{2\pi}}e^{-\frac{n^2}{2}}$.

III. Convolutional Neural Network for Modulation Recognition

Fig. 2 shows our proposed CNN architecture, which is a neural network with many levels of non-linearities allowing them to represent a highly non-linear classification function that maps spectrogram features to modulation methods. It consists of one input layer, 4 convolutional layers, where the first three layers have max pooling. After the convolutions, it is followed by one densely connected layer and finally a softmax layer. ReLU is used for all activation functions. The input is the images with dimension of 100 × 100 × 3. The number of filters used in ith convolutional layer is 64, 32, 12, 8 with dimension of 3 × 3, 3 × 3, 3 × 3 and 3 × 3, respectively. The size of zero-padding and stride are 1 and 1, respectively. The pooling size of the max-pooling layer is (2, 2). The fully connected layer consists 128 neurons. The design of network architecture first consider the used architectures in related work [8] and [10], and the recommend setup of high parameters, such as the number of layer, filter and filter size, in image classification. Later, the network architecture is determined by performing multiple experiments with different high parameters and comparing their recognition accuracy. The output of the network is the estimated modulation method of the input spectrogram image. Moreover, the network is trained using stochastic gradient descent to minimize the cross-entropy loss function [14]: $\mathbf{w}^* = \mathrm{argmin}_{\mathbf{W}} \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}(\mathbf{w}; \mathbf{x}^i, t^i)$ with the number of training examples N, the right labels t^i and the predict labels x^i.

IV. Experiments

A. Data

This experiment uses the data set RadioML2016.10a in [7] as the basic data set. It considers 11 modulation methods: BPSK, QPSK, 8PSK, 16QAM, 64QAM, BFSK, CPFSK, PAM4, WB-FM, AM-SSB and AM-DSB, which are widely used in practical communication systems and operates on both discrete binary alphabets and continuous alphabets. The data set considers 20 different signal-to-noise ratios (SNRs) varying from −20 dB to 18 dB, and 1000 signals per modulation mode per SNR. Each signal consists of 128 samples, and each sample contains the real and imaginary parts.

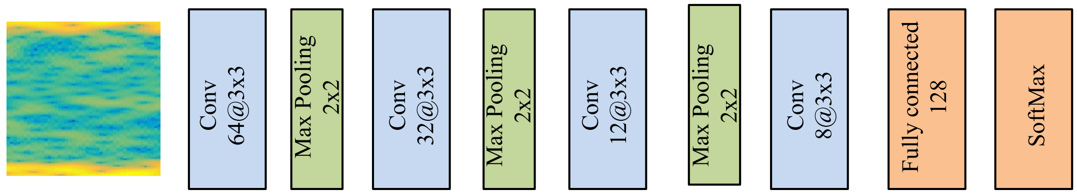

Fig. 2. CNN architecture for modulation recognition.

In this letter, the signals are transformed into spectrogram images using frame-based processing, with a frame length of 40 samples and a 90%-overlapping Hann window. Our spectrogram images are generated using 'spectrogram' function in MATLAB R2017b and saved as $200 \times 200 \times 3$ RGB images. Later, we use nearest interpolation to re-size the image resolution of spectrogram image from $200 \times 200 \times 3$ to $100 \times 100 \times 3$. We refer to the proposed framework as SCNN. To reduce noise, we use the Gaussian filter with size 7 to pre-process the observed signal $y(n)$, and we refer to the SCNN method with noise reduction pre-processing as SCNN2.

B. Experimental Setup

To evaluate the recognition performance of the proposed framework, we consider two experiments in this section. First, we compare the recognition accuracy of the SCNN and SCNN2 with two methods from [8] and [10], which we refer to as CNNR-IQ and CNNR-IQFOC, respectively. Both the CNNR-IQ and CNNR-IQFOC are proposed for classifying modulation modes using the same data set as in this letter. Moreover, we compare the computational complexity of the SCNN2 with those of the CNNR-IQ and CNNR-IQFOC in terms of memory consumption, number of learned parameters and training time per signal.

Later, to evaluate the effectiveness of the proposed spectrum analysis of the modulated signals, we further use the ambiguity function image and spectral correlation function image as signal representation, which we refer to as AF-CNN and SCF-CNN, respectively. We compare the recognition performance of the proposed SCNN2 with those of the AF-CNN and SCF-CNN versus different SNRs.

Our experiments randomly select 700 signals per modulation mode per SNR as training data, and the remaining signals are divided into validation data (100 signals per modulation per SNR) and test data (200 signals per modulation per SNR). Specifically, we train classification model per SNR with 700×11 images. We normalize all images before processing. The learning rate starts with 0.0005 and is divided by a factor of 10 every 100 iterations. We stop the training process when the validation loss is not decreased within 15 iterations, and we save the trained model with the smallest validation loss. The recognition accuracy of the SCNN2 versus training steps is shown in Fig. 3. Later, the trained model is used to predict the modulation type of each test image. The CNN based experiments are implemented using Tensorflow based Keras and Nvidia TITAN X GPU.

C. Experiment Results

Fig. 4 shows the recognition accuracy comparison between the SCNN, SCNN2, CNNR-IQ and CNNR-IQFOC versus different SNRs. The results in Fig. 4 show that the recognition accuracy of the SCNN2 is around 4% lower than those of the SCNN method at the 18 dB SNR, but around 2% higher than those of the SCNN when SNR is below -4 dB, since the noise reduction algorithm have limited capability to improve SNR when signals are severely distorted and close to clean. In addition, the recognition accuracy improvement is around 10% when SNR is between -4 dB and 16 dB. Furthermore, we observe that both the SCNN2 and CNNR-IQFOC obtain better recognition accuracy than the CNNR-IQ. Specifically, the recognition performance of the SCNN2 is around 5% higher than those of the CNNR-IQ when the SNR is below -4 dB, and is around 15% to 20% higher than those of the CNNR-IQ when SNR is above -2 dB. Moreover, the recognition performance of the SCNN2 and the CNNR-IQFOC is very similar when the SNR level is lower than -8 dB. The SCNN2 gets around 5% lower accuracy than the CNNR-IQFOC at the -8 dB, -6 dB and -4 dB SNR. However, the SCNN2 gets around 8% higher accuracy than the CNNR-IQFOC when SNR is between 2 dB and 18 dB. It indicates that the SCNN2 generally has similar or slightly worse performance than the CNNR-IQFOC when the SNR is lower than -2 dB, but performs better than the CNNR-IQFOC when the SNR is above -2 dB.

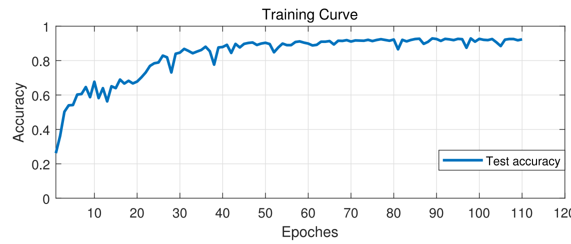

Fig. 3. Recognition accuracy of SCNN2 versus training epochs at 10dB SNR.

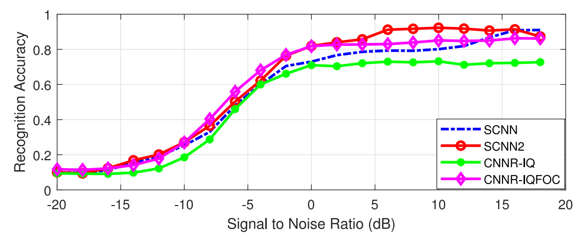

Fig. 4. Recognition accuracy comparison between the SCNN, SCNN2, CNNR-IQ and CNNR-IQFOC versus SNR.

Next, we compare the recognition performance of the SCNN2 with those of the AF-CNN and SCF-CNN. The experiment results in Fig. 5 show that the SCNN2 performs better than both the AF-CNN and the SCF-CNN. Specifically, the SCNN2 gets around 15% higher accuracy than the SCF-CNN

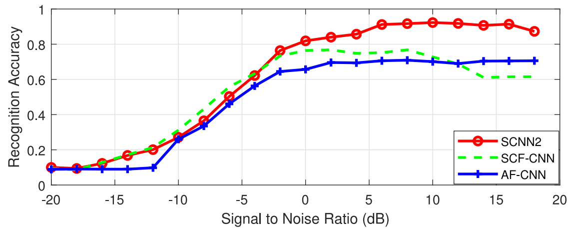

Fig. 5. Recognition accuracy comparison between the SCNN2, SCF-CNN and AF-CNN versus SNR.

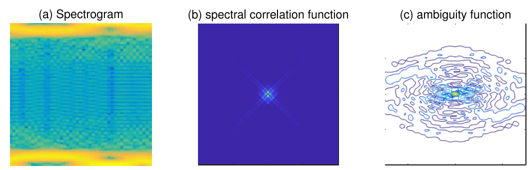

Fig. 6. Spectrum features of a BPSK modulated signal at 10 dB SNR.

TABLE I
COMPUTATION COMPLEXITY COMPARISON BETWEEN THE SCNN2, CNNR-IQ AND CNNR-IQFOC

	SCNN2	CNNR-IQ	CNNR-IQFOC
Memory(k)	941	21	185
Learned parameters(k)	199	532	3675
Training time(ms)	7.14	5.15	8.43

and gets around 20% higher accuracy than the AF-CNN when the SNR is above -2 dB. The recognition performance of the SCNN2 and the SCF-CNN is very similar when the SNR level is lower than -8 dB. This can be explained that the recognition performance of learning based classification method is proportional to the diversity of the input data, and the spectrogram analysis in the SCNN2 provides richer time-frequency representation of the signal than the ambiguity function image in the AF-CNN and the spectral correlation function image in the SCF-CNN, see Fig. 6.

D. Computational Complexity

The computational complexity of the SCNN2, CNNR-IQ and CNNR-IQFOC is evaluated by comparing memory consumption, the number of learned parameters and the average training time per signal. As shown in Table I, the SCNN2 requires much more memory than the CNNR-IQ and the CNNR-IQFOC, since the input data size of the SCNN2, CNNR-IQ and CNNR-IQFOC are $100 \times 100 \times 3$, 2×128, and 3×128, respectively. However, the number of learned parameters of the SCNN2 is smaller than those of the reference methods, since the SCNN2 uses pooling access and small number of filters in each convolutional layer, while the network architecture of the CNNR-IQ did not include pooling layer and the CNNR-IQFOC uses larger number of filters in each convolutional layer. In addition, the training time of the SCNN2 is bigger than those of the CNNR-IQ, but smaller than those of the CNNR-IQFOC.

V. CONCLUSION

In this letter, we presented a time-frequency analysis of modulated radio signals and designed a novel spectrum analysis based convolutional neural network (SCNN) framework for automatic modulation recognition. We applied the short-time discrete Fourier transform to the observed signals and used spectrogram images as the input of the SCNN. Experiment results demonstrated the impressive performance of the proposed SCNN method and showed the effectiveness of the introduced noise reduction approach, which is referred to as SCNN2. Moreover, experiments on the comparison between the SCNN2 and two deep learning based methods (CNNR-IQ and CNNR-IQFOC) from literature demonstrated the better recognition capability of the proposed SCNN2. Furthermore, experiments on the comparison between the SCNN2 and two other spectrum analysis based methods (AF-CNN and SCF-CNN) demonstrated the effectiveness of the spectrogram image. The computational complexity analysis indicated that the SCNN2 requires more memory but less learned parameters than the reference methods, and the training time per signal of the SCNN2 is bigger than those of the CNNR-IQ but smaller than those of the CNNR-IQFOC.

REFERENCES

[1] W. Wei and J. M. Mendel, "Maximum-likelihood classification for digital amplitude-phase modulations," *IEEE Trans. Commun.*, vol. 48, no. 2, pp. 189–193, Feb. 2000.
[2] A. Swami and B. M. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 416–429, Mar. 2000.
[3] L. Hong, "Classification of BPSK and QPSK signals in fading environment using the ICA technique," in *Proc. 37th Southeastern Symp. Syst. Theory (SSST)*, Tuskegee, AL, USA, Mar. 2005, pp. 491–494.
[4] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 21, no. 3, pp. 660–674, May/Jun. 1991.
[5] B. Scholkopf, K. Tsuda, and J. Vert, *Advanced Application of Support Vector Machines*. London, U.K.: MIT Press, 2004, p. 275.
[6] R. Lippmann, "An introduction to computing with neural nets," *IEEE ASSP Mag.*, vol. 4, no. 2, pp. 4–22, Apr. 1987.
[7] T. J. O'Shea and N. West, "Radio machine learning dataset generation with GNU radio," in *Proc. GNU Radio Conf.*, vol. 1, 2016, pp. 1–6.
[8] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *Proc. Int. Conf. Eng. Appl. Neural Netw.*, vol. 629. Aberdeen, U.K., 2016, pp. 213–226.
[9] N. E. West and T. O'Shea, "Deep architectures for modulation recognition," in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw. (DySPAN)*, Mar. 2017, pp. 1–6.
[10] M. Zhang, Y. Zeng, Z. Han, and Y. Gong, "Automatic modulation recognition using deep learning architectures," in *Proc. 19th IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2018, pp. 1–5.
[11] A. Dai, H. Zhang, and H. Sun, "Automatic modulation classification using stacked sparse auto-encoders," in *Proc. IEEE 13th Int. Conf. Signal Process. (ICSP)*, Chengdu, China, Nov. 2016, pp. 248–252.
[12] Q. V. Le *et al.*, "On optimization methods for deep learning," in *Proc. 28th Int. Conf. Int. Conf. Mach. Learn. (ICML)*, Bellevue, WA, USA, 2011, pp. 265–272.
[13] 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.
[14] C. M. Bishop, *Pattern Recognition and Machine Learning* (Information Science and Statistics). Heidelberg, Germany: Springer-Verlag, 2006.