2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)

Automatic Modulation Recognition using Deep Learning Architectures

Meng Zhang, Yuan Zeng, Zidong Han and Yi Gong, *senior member, IEEE*
Southern University of Science and Technology, Shenzhen, China
email: 11649162@mail.sustc.edu.cn, zengy3, hanzd, gongy@sustc.edu.cn

***Abstract*—In this paper, we present an automatic modulation recognition framework for the detection of radio signals in a communication system. The framework considers both a deep convolutional neural network (CNN) and a long short term memory network. Further, we propose a pre-processing signal representation that combines the in-phase, quadrature and fourth-order statistics of the modulated signals. The presented data representation allows our CNN and LSTM models to achieve 8% improvements on our testing dataset. We compare the recognition accuracy of the proposed recognition methods with existing methods under various SNR values. Experimental results show that our methods perform better than the existing methods.**

***Index Terms*—Modulation recognition, convolutional neural network, long short term memory, signal representation.**

I. Introduction

To achieve efficient data transmission, transmitted signals are generally modulated using different modulation methods in a communication system. As an intermediate process between signal detection and signal demodulation, modulation recognition is an important technology to provide modulation information of signals for further signal demodulation and decoding in practical applications, such as cognitive radio, signal recognition, threat assessment and spectrum monitoring.

In general, modulation recognition methods can be divided into two categories: decision theory based method and statistical pattern recognition based method. The decision theory based methods [1]–[3] use probability theory, hypothesis testing theory and a proper decision criterion to solve modulation recognition problem. However, since this method requires the prior information of received signals, it has been proven to be time-consuming and inefficient when a large number of modulation types coexist. As an alternative, since modulation recognition can be formulated as a pattern recognition problem, statistical pattern recognition approaches are able to learn a recognition model with multiple unknown parameters based on sufficiently large training data. Thus, by providing complete information of different modulations in training data and considering some expert features simultaneously, pattern recognition approaches are more robust and efficient. Pattern recognition methods normally focus on feature extraction and classification algorithms. The existing feature extraction methods use instantaneous amplitude, frequency and phase [4], high-order statistics (HOS) [5], cyclostationary characteristics [6], while the existing classifiers include decision tree [7], support vector machine (SVM) [8] and artificial neural network [9].

Recently, deep learning is well studied in image classification [10], object detection [11] and speech recognition [12]. It has also attracted great attention and been applied in modulation recognition. In [13], the authors propose to convert modulated signals into constellation diagrams and feed them to the Alexnet model [10] to perform classification. Based on high order statistics and the extreme learning machine (ELM), the paper [14] proposes a signal recognition algorithm for blind digital modulation identification. The paper [15] uses cyclic spectrum to pre-process signals, and then compares the recognition performance of a deep autoencoder network with those of the three algorithms including SVM, naive Bayes and neural network. The paper [16] surveys the emerging applications of machine learning in radio signal processing domain and uses GNU Radio [16] to generate an open dataset with raw In-Phase and Quadrature (IQ) information for modulation recognition. The paper [17] studies the adaptation of convolutional neural networks (CNN) to the dataset in [16] and compares the modulation recognition performance of the proposed CNN against those of the expert cyclic moment features based methods. Furthermore, in [18], the authors make a comparison between CNN, residual networks, inception modules, convolutional long short-term deep neural networks(CLDNN) based on the dataset in [16], and experimental results show that modulation recognition performance is not limited by network depth.

In this work, we assess the applicability of the CNNs and the Long Short Term Memory (LSTM) networks [19] for modulation recognition, since the CNNs are good at extracting features of spatial data and the LSTM networks have been proved to behave well in sequence data. In addition, instead of only using IQ information of signal as the input data of networks, we propose a pre-processing signal representation that leverages the IQ information and HOS feature of the modulated signals, to improve modulation recognition performance. The idea of using three channel information of signal as input of networks is inspired by the application of deep learning in computer vision such as RGB images classification. Specifically, we combine the raw IQ data and Fourth order Cumulants (FOC) together to represent the modulated signal, since the higher-order cumulants can characterize the shape of the distribution of received noisy signals and the FOC of received signals are usually employed as the features for modulation recognition. We evaluate the recognition performance of the proposed frameworks by using public dataset with 11 common used modulation methods. Moreover, the recognition performance of the proposed frameworks is compared with those of the state-of-the-art CNN based method in literature [17].

The rest of the paper is organized as follows. In Section II the problem statement and system model are discussed. Section III introduces the CNN and LSTM network architecture for modulation recognition. In Section IV, experimental results are given. Finally, conclusion and discussion are drawn in Section V.

II. Problem Statement and Signal Representation

In this section, we first briefly introduce the modulated model in a simple communication system, and then we present a signal representation used in this work.

A. Modulated Signal Model

Let us consider a wireless communication system in Fig. 1, which consists of a transmitter, a channel and a receiver. Let $y(t)$ denote a continous-time signal at receiver. The received signal $y(t)$ is generally given by

$$y(t) = f(s(t)) * h(t) + n(t), \tag{1}$$

where $s(t)$ is the signal to be transmitted with time index t, f is the transmitter function, $h(t)$ is the channel response and $n(t)$ is the additive noise. Given the received signal $y(t)$, modulation recognition aims to predict the modulation type of f, and thus to provide modulation information for estimating the source signal $s(t)$ from $y(t)$. However, the mathematical model of $y(t)$ in real world communication systems is more complex. In this work, the received signal $y(t)$ is generated in harsh simulated propagation environments, and we consider 11 modulations including BPSK, QPSK, 8PSK, 16QAM, 64QAM, BFSK, CPFSK, PAM4, WB-FM, AM-SSB, and AM-DSB.

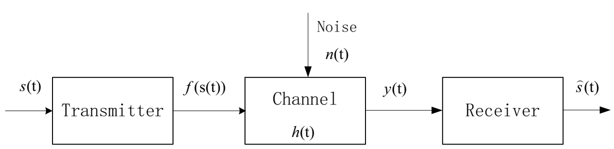

Fig. 1. A simple wireless communication system

B. Modulated signal representation

Second order cumulants, such as correlation function and power spectrum, generally cannot be used to deal with signal processing problems related to either nonlinearity, non-Gaussianity, or non-minimum phase systems. Unlike second order cumulants, high-order cumulants are usually used to solve these kinds of signal processing problems since they contain the phase information of the underlying linear system. The method of digital modulation recognition taking advantage of high-order statistics as the feature can effectively suppress the white Gaussian noise. In this work, in addition to using the IQ components as the representation of the received signal $y(t)$, we also use high-order cumulants of $y(t)$. Here, we give a brief introduction of the high-order cumulants.

Let $y(n)$ denote the discrete-time received signal at time sampling index n. The second-order cumulants of the complex-valued received signal $y(n)$ can be designed in two different equations, which are $C_{20} = E[y^2(n)]$ and $C_{21} = E[|y(n)|^n]$, where $E(\cdot)$ is the expectation function. This work uses the fourth-order cumulants of the sampled received signal $y(n)$. The fourth-order cumulants of the complex-valued signal $y(n)$ is defined as

$$C_{40} = cum(y(n), y(n), y(n), y(n)), \tag{2}$$

and

$$C_{41} = cum(y(n), y(n), y(n), y^*(n)), \tag{3}$$

and

$$C_{42} = cum(y(n), y(n), y^*(n), y^*(n)), \tag{4}$$

where C_{40}, C_{41} and C_{42} denote three representations of fourth-order cumulants depending on placement of conjugation and $cum(\cdot)$ denotes the cumulant function.Specifically, the joint cumulant function of ω, x, y and z is expressed by:

$$\begin{aligned} cum(\omega, x, y, z) = E(\omega xyz) - E(\omega x)E(yz) \\ -E(\omega y)E(xz) - E(\omega z)E(xy), \end{aligned} \tag{5}$$

Thus the fourth-order cumulants C_{40}, C_{41} and C_{42} can be expressed by the fourth and second-order statistics of $y(n)$.

III. Methods

To classify modulation types, we investigate the use of the CNN and the LSTM for modulation recognition in this section.

A. CNN for Modulation Recognition

The input of the network is the sequence data consisting of the IQ and FOC, and the output is the corresponding modulation type. Fig. 2 shows the architecture of the network. The network performs convolution between the feature maps of the previous layer and a set of learnable filters to extract data features. We then apply RELU $(\max(0, x))$ activations [10], pooling layer to perform a down-sampling operation, fully connected layer to connect all the features and send the output value to the classifier. To reach the output of the network, the input data is fed to 2 convolutional layers followed by 2 max-pooling layers respectively, then the extracted features are feed into one fully connected layer and the softmax classifier. The number of filters used in ith convolutional layer is 256, 128 with dimension of 1×8 and 2×8, respectively. After each convolutional layer, max-pooling layer with the pool size (1,2) is used to perform the feature dimension. Then one fully connected layer consisting of 256 neurons with RELU activation function is used to send the calculated value to the softmax classifier. At the same time, dropout is applied after fully connected layer to reduce over-fitting.

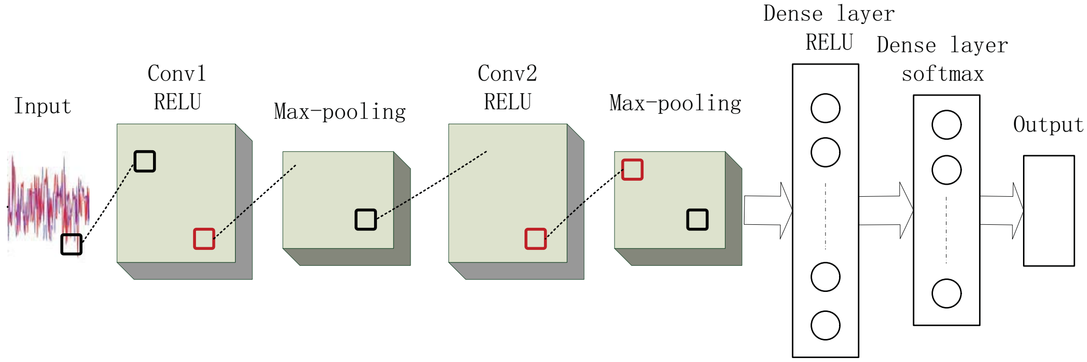

Fig. 2. CNN architecture for modulation recognition

B. LSTM for Modulation Recognition

RNNs are generally used to deal with the interrelated problems between inputs by introducing a directional loop. The LSTM network is a special RNNs, which is used to solve the long term dependencies problems. The architecture of a common LSTM is composed of a memory cell, an input gate, an output gate and a forget gate. The forget gate is responsible for deciding what information we will discard from the cell state. The input gate aims to determine which values will be updated to the cell state. The output gate controls the extent to which the value in the cell is used to compute the output activation of the LSTM unit. The final output of the LSTM is determined by the output gate and the cell state.

The working mechanism of the LSTM in this work can be described as

$$f_t = \sigma_g(\mathbf{W}_f x_t + \mathbf{U}_f h_{t-1} + \mathbf{b}_f), \tag{6}$$

$$i_t = \sigma_g(\mathbf{W}_i x_t + \mathbf{U}_i h_{t-1} + \mathbf{b}_i), \tag{7}$$

$$o_t = \sigma_g(\mathbf{W}_o x_t + \mathbf{U}_o h_{t-1} + \mathbf{b}_o), \tag{8}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \sigma_h(\mathbf{W}_c x_t + \mathbf{U}_c h_{t-1} + \mathbf{b}_c), \tag{9}$$

$$h_t = o_t \odot \sigma_t(c_t), \tag{10}$$

where $\odot$ denotes the Hadamard product, $\mathbf{b}_f$, $\mathbf{b}_i$, $\mathbf{b}_o$ and $\mathbf{b}_c$ are the bias vectors, $\mathbf{W}$ and $\mathbf{U}$ are the weight matrices of the input vector x_t and output h_{t-1} of the previous layer respectively. In addition, c_t is the cell state, and h_t represents the output of the LSTM unit, f_t, i_t, o_t are the activation vectors of the forget gate, input gate, and output gate, respectively. σ_g and σ_h denote the activation function of sigmoid function and tanh function, respectively. The network architecture for modulation recognition is illustrated in Fig. 3. The input data is first fed to 3 convolution layers, where each convolution layer is followed by the RELU activations, for feature extracting. Furthermore, each convolution layer uses 50 filters with dimension of 1×9. We then use one LSTM layer which returns a 250 dimensions vectors with tanh activation function for further processing. After that, one fully connected layer consisting of 250 neurons with RELU activation function is used before the softmax classifier.

IV. Experiments

A. Data

In this experiment, we use the RadioML2016.10a [16] dataset as the basic dataset. We choose it because it is open and available, and the amount and types of modulation data fits our work in this paper. The dataset RadioML2016.10a contains 11 modulation methods: BPSK, QPSK, 8PSK, 16QAM, 64QAM, BFSK, CPFSK, PAM4, WB-FM, AM-SSB, and AM-DSB, which are widely used in wireless communication systems. In addition, the dataset considers 20 different signal-to-noise ratios (SNRs) varying from -20dB to 18dB, and 1000 signals per modulation mode and per SNR. Each signal with the dimension of 2×128 consists of the real and imaginary parts, and we refer to this representation as IQ. We consider two signal representations in our experiments: the IQ, and the combination of the IQ and the extracted FOC features (IQ-FOC) with dimension of 3×128.

B. Experimental Setup

In this section, we evaluate the recognition performance of the CNN and LSTM using the IQ and IQ-FOC, and we refer to the methods as CNN-IQ, CNN-IQFOC, LSTM-IQ and LSTM-IQFOC, respectively. We first compare the recognition performance of the CNN-IQ, CNN-IQFOC, LSTM-IQ and LSTM-IQFOC methods under different noise levels. Later, we compare the recognition performance of the presented method CNN-IQFOC and the LSTM-IQFOC with a method from the literature [17], which we refer to as CNNR-IQ. The CNNR-IQ method in [17] is the state-of-the-art deep learning based automatic modulation recognition method for classifying such large number of modulation types, which uses IQ information as signal representation and the CNN network architecture without pooling layer for modulation recognition. In our experiments, we randomly select 700 signals per modulation mode per SNR in the dataset 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). In addition, our experiments are implemented by using Nvidia GTX 1050 GPU.

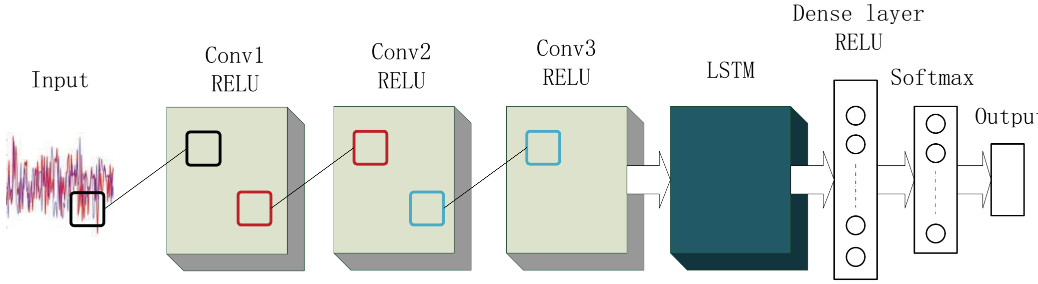

Fig. 3. LSTM network architecture for modulation recognition

C. Experimental Results

Fig. 4 shows the recognition performance comparison in terms of recognition accuracy versus different SNRs between the CNN-IQ, CNN-IQFOC, LSTM-IQ and LSTM-IQFOC. Fig. 4 shows that the recognition accuracy of the LSTM-IQFOC is around higher 2% than those of the CNN-IQFOC when SNR is above -5dB, but the recognition performance of the LSTM-IQ and the CNN-IQ is almost the same at all SNR levels, since the LSTM network behaves better than the CNN framework for detecting the correlation between sequence data. In addition, the results in Fig. 4 show that the IQ-FOC signal representation based recognition methods perform better than the IQ signal representation based methods, where the recognition accuracy of the the CNN-IQFOC and LSTM-IQFOC is higher than those of the CNN-IQ and LSTM-IQ. Specifically, the recognition accuracy of the CNN-IQFOC and LSTM-IQFOC is a roughly 8% higher than those of the CNN-IQ and LSTM-IQ when SNR is above 0dB.

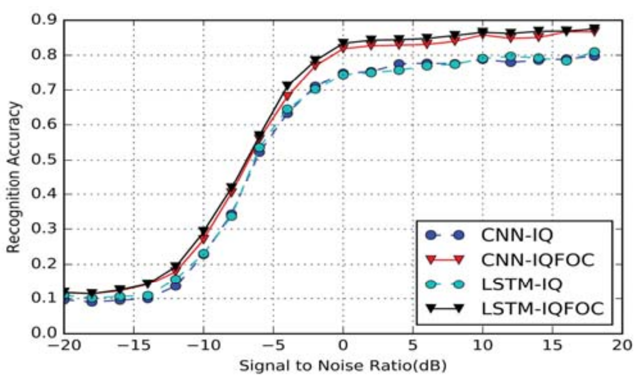

Fig. 4. Recognition accuracy versus SNR

Next, we compare the recognition performance of the proposed CNN-IQFOC and LSTM-IQFOC method with those of the reference method CNNR-IQ in the literature [17]. The experimental results are shown in Fig. 5. We observe that the recognition performance of the CNN-IQFOC, the LSTM-IQFOC and CNNR-IQ is very similar when the SNR level is lower than -16dB. However, the recognition accuracy of the CNN-IQFOC and LSTM-IQFOC is around 15% higher than those of the CNNR-IQ method when the SNR is above 0dB. This indicated that the signal representation and network architecture such as filter size and whether or not the network containing the pooling layers affect the recognition performance significantly.

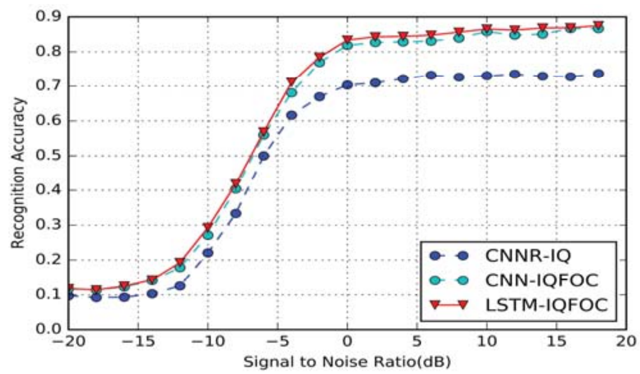

Fig. 5. Comparative performance evaluation of the CNN-IQFOC, the LSTM-IQFOC and CNNR-IQ versus SNR

Note that although we only compare recognition performance of the CNN-IQFOC, the LSTM-IQFOC and CNNR-IQ, the experimental results in Fig. 4 show that both the CNN-IQ and LSTM-IQ perform better than the CNNR-IQ due to the different model architectures. In addition, all experimental results in Fig. 4 and Fig. 5 show that all deep learning methods do not learn the features well from the data when the SNR lever is lower than -8dB. However, the proposed IQ-FOC based methods gradually have a better recognition performance, since the combination of raw IQ and FOC provides more robust features for learning algorithms.

V. Conclusion and Future Work

In this paper, we presented a deep CNN network and a deep LSTM network for modulation recognition of radio signals. In addition, we proposed IQ-FOC data representation of modulated signals, where the raw IQ data and the Fourth order Cumulants (FOC) of signals are combined. We compared the modulation recognition performance for different data representation and concluded that our proposed IQ-FOC representation enabled both the CNN and the LSTM models to obtain high classification accuracy and it was more robust to noise. In addition, the experimental results showed that the LSTM network behaved a little better than the CNN, since the LSTM contains the interconnect between hidden layer nodes which allows it to exhibit dynamic temporal behavior for a time sequence. Moreover, experiments on the comparison between the proposed algorithm and a CNN method (CNNR-IQ) from literature demonstrated the better recognition capability of the proposed network architectures and the effectiveness of the proposed IQ-FOC representation.

The dataset used in this paper is generated via GNU radio with the GNU Radio Dynamic Channel Model hierarchical block. In our future work, we intend to conduct a wireless communication system to collect and build our dataset using practical signals. In addition, all experimental results showed that the recognition accuracy of all deep learning based recognition algorithms is low in extreme noisy scenarios. Ongoing research investigates how the recognition accuracy can be improved in extremely noisy scenarios.

References

[1] Wen Wei and J. M. Mendel, "Maximum-likelihood classification for digital amplitude-phase modulations," *IEEE Transactions on Communications*, vol. 48, no. 2, pp. 189–193, Feb 2000.

[2] J. A. Sills, "Maximum-likelihood modulation classification for psk/qam," in *MILCOM 1999. IEEE Military Communications. Conference Proceedings (Cat. No.99CH36341)*, 1999, vol. 1, pp. 217–220 vol.1.

[3] Chung-Yu Huan and A. Polydoros, "Likelihood methods for mpsk modulation classification," *IEEE Transactions on Communications*, vol. 43, no. 2/3/4, pp. 1493–1504, Feb 1995.

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

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

[6] Liang Hong, "Classification of bpsk and qpsk signals in fading environment using the ica technique," in *Proceedings of the Thirty-Seventh Southeastern Symposium on System Theory, 2005. SSST '05.*, March 2005, pp. 491–494.

[7] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 3, pp. 660–674, May 1991.

[8] Bernhard Sch?lkopf, Koji Tsuda, and Jean-Philippe Vert, *Advanced Application of Support Vector Machines*, pp. 275–275, MIT Press, 2004.

[9] R. Lippmann, "An introduction to computing with neural nets," *IEEE ASSP Magazine*, vol. 4, no. 2, pp. 4–22, Apr 1987.

[10] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., pp. 1097–1105. Curran Associates, Inc., 2012.

[11] Ross Girshick, "Fast r-cnn," Santiago, Chile, 2015, vol. 2015 International Conference on Computer Vision, ICCV 2015, pp. 1440 – 1448, Convolutional networks;Detection accuracy;Open sources;Region-based;Test time;Training and testing;.

[12] A. Graves, A. r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 6645–6649.

[13] S. Peng, H. Jiang, H. Wang, H. Alwageed, and Y. D. Yao, "Modulation classification using convolutional neural network based deep learning model," in *2017 26th Wireless and Optical Communication Conference (WOCC)*, April 2017, pp. 1–5.

[14] X. Liu, C. Zhao, P. Wang, Y. Zhang, and T. Yang, "Blind modulation classification algorithm based on machine learning for spatially correlated mimo system," *IET Communications*, vol. 11, no. 7, pp. 1000–1007, 2017.

[15] J. Li, L. Qi, and Y. Lin, "Research on modulation identification of digital signals based on deep learning," in *2016 IEEE International Conference on Electronic Information and Communication Technology (ICEICT)*, Aug 2016, pp. 402–405.

[16] T. O'Shea and N. West, "Radio machine learning dataset generation with gnu radio," in *Proceedings of the GNU Radio Conference*, 2016, vol. 1.

[17] T. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," Aberdeen, United kingdom, 2016, vol. 629, pp. 213 – 226.

[18] N. E. West and T. O'Shea, "Deep architectures for modulation recognition," in *2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, March 2017, pp. 1–6.

[19] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.