Automatic Modulation Classification Using Convolutional Recurrent Attention Network

Yuzheng Yang[1], Sai Huang[1,2], Shuo Chang[1], Hua Lu[2](✉), and Yuanyuan Yao[3]

[1] Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 10087, People's Republic of China
{yangyuzheng,huangsai,changshuo}@bupt.edu.cn

[2] GuangDong Communications and Networks Institute, Guangdong 510700, People's Republic of China
luhua@gdcni.cn

[3] School of Information and Communication Engineering, Beijing Information Science and Technology University, Beijing 100101, People's Republic of China
yyyao@bistu.edu.cn

Abstract. The development of wireless communication technology is much faster than the pace of its security. The interference such as adversarial attack degrades the accuracy and efficiency of communication environment. Automatic modulation classification (AMC) is viewed as an effective method to discover and identify the modulation mode of wireless signal corrupted by noise and interference. This paper proposes a novel modulation classification framework using Convolutional Recurrent Attention Network (CraNET) which is mainly composed of the convolution block and long short term memory network (LSTM) based attention block. Convolution block extracts the signal features in the feature extraction module. In the weighting module, the LSTM based attention block selectively weights the extracted features to weaken the content that has no contribution to the performance improvement. Extensive simulation verifies that the CraNET based modulation classification method performs higher accuracy and superior robustness than that of other existing methods.

Keywords: Automatic modulation classification · Attention block · Adversarial attack · Convolution block · Long short term memory network

This work was supported in part by the Guangdong Key Field *R&D* Program under Grant (2018B010124001), in part by the National Key Research and Development Program of China under Grants (2018YFF0301202, 2019YFB1804404, 2020YFB1807602), in part by the Beijing Natural Science Foundation under Grants (4202046, 19L2022), in part by the National Natural Science Foundation of China under Grant (61801052,61941102), and in part by the Industrial Internet Innovation and Development Project of Ministry of Industry and Information Technology under Grant (TC200H031).

J. Xiong et al. (Eds.): MobiMedia 2021, LNICST 394, pp. 736–747, 2021.
https://doi.org/10.1007/978-3-030-89814-4_54

1 Introduction

With the booming development of wireless communication technology [1], both the number and variety of wireless devices are increasing rapidly. At the same time, the way of wireless communication has become more diversified. However, rapid technological innovation has broken the stability of technological security. Diversified wireless devices and communication methods promote the development of interference technology such as adversarial attacks [2] to interfere with normal communication, which seriously degrades the accuracy and efficiency of communication environment, and causes unpredictable security risks. Compared with the traditional interference, this kind of interference is not noise, but is more invasive than noise and more difficult to distinguish the resistance interference. Automatic modulation classification (AMC), which is regarded as the basis link of spectrum sensing technology [3], identifies the modulation mode of wireless signal corrupted by noise and interference. AMC is mainly composed of three major technologies: data acquisition, feature extraction and modulation classification [4–6], and it is a common method in the field of anti active malicious interference such as jamming and spoofing.

AMC is divided into two main directions: likelihood based approach (LB) and feature extraction based approach (FB) [7]. LB classifiers first calculate the likelihood of the received signals under the hypotheses of different modulation schemes and then validate modulation schemes with the maximum likelihood (ML). The recognition accuracy of LB can achieve optimal performance in the Bayesian sense via minimizing the probability of misclassification. However, it will be degraded in the complex and changeable signal data, priori knowledge such as the probability distribution function of the signal also needs to be known in advance. FB method makes up the defect of LB method. FB method extracts some statistical information about the input signal for modulation classification [8–10]. FB method is a sub-optimal method, and it is difficult to deal with the situation of low SNR channel. However, it has the advantages of low computational complexity and requiring few prior knowledge about the received modulated signals. Thus, this paper aims at solving problems induced by FB method and improving the performance of FB method.

In order to achieve better performance, the neural network, which is robust to channel and noise uncertainties, is introduced into signal modulation classification. O'Shea *et al.* proposed a convolution neural network, which is used to identify the modulation formats [11]. However, no changes have been made to the structure of the traditional convolution neural network, which is firstly utilized for image classification. Therefore, it can not considers the sequential features of wireless signal, the recognition accuracy of this network is also limited. Sun *et al.* proposed a new approach to the automatic modulation classification of cochannel signals [12], which improved CNN based network structure and achieved certain results. However, the proposed network only considers the classification of the digital modulation signal, and the classification of the analog modulations is not considered. Mendis *et al.* [13] proposed an AMC framework which consists of spectrum correlation function (SCF) based feature characterization mechanism and deep belief network (DBN) based identification schem. However, the robustness of the network depends on the SCF image feature. Yao *et al.* [14] proposed an innovative joint model using CNN-LSTM network. In this model, QAM classification accuracy improved by Haar-wavelet Crest Searching. However, other modulation types' classification accuracy has not been improved due to the limitations of LSTM in processing long sequences. Tu Ya *et al.* [15] pointed out that DL model requires many training data to combat with over-fitting, then they extended Genera-tive Adversarial Networks (GANs) to the semi-supervised learning to overcome the problem. Yun Lin *et al.* [16] pointed out that waveforms in the physical layer may not be suitable for the prevalent classical DL models. They also evaluated the security problems caused by adversarial attacks to modulation recognition with CNN [17]. The results showed that CNN was highly vulnerable to adversarial attacks.

To solve the problems mentioned above, in this paper, we propose a CraNET based modulation classification method which can be divided into three parts: feature extraction module, weighting module and modulation classification module. CraNET's structure is mainly composed of convolution block and long short term memory network (LSTM) based attention block. Convolution block extracts the signal features in the feature extraction module. Different from LSTM which only outputs the last hidden state of LSTM layer, the proposed LSTM based attention block uses all hidden states to calculate the output in the weighting module. The contribution of each hidden state to the final output is enhanced or weakened by weighted summation. Moreover, the modulation classification decision is made in the modulation classification module using the softmax function. Extensive simulation verifies that the CraNET based modulation classification method performs higher accuracy and superior robustness than that of other existing methods.

The rest part of this paper is organized as follows: Sect. 2 introduces the signal model and modulation classification framework. Section 3 presents the CraNET structure, and introduces the process of classification. Simulation results are given in Sect. 4, which includes the comparison with other network performance. Section 5 summarizes the content of this paper.

2 Problem Statement

2.1 Signal Model

Since the normal communication will be shut down or transmit the signal on another carrier when it is interfered by the adversarial attack, the single antenna receiver receives the signal from one source at a given time. Hence the received baseband signal is given by:

$$r(n) = he^{(2\pi f_0 n+\theta_0)} s_k(n) + w(n), n = 1, 2, ..., N \quad (1)$$

where h is the Rayleigh channel coefficient, s_k is the complex baseband envelope of the received signal generated from the k-th modulation hypothesis $H_k \in$

$(H_1, H_2, ..., H_K)$, N denotes the total number of signal symbols, $w(n)$ is noise with its mean and variance are 0 and σ^2 respectively, and it is used as additive white Gaussian noise (AWGN). f_0 denotes the frequency offset and θ_0 denotes the phase offset. Since the average power of received signal is normalized, we define the SNR as:

$$\gamma_{SNR} = \frac{|h|^2}{\sigma_w^2} \tag{2}$$

2.2 CraNET Based Modulation Classification Framework

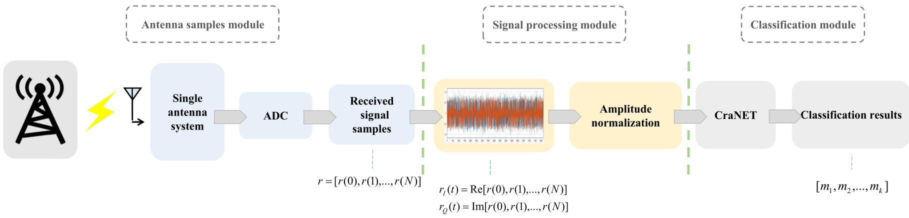

Fig. 1. CraNET based modulation classification framework

Figure 1 presents the CraNET based modulation classification framework, which consists of three modules: antenna samples module, signal processing module and classification module. The signal is received through the single-input and single-output (SISO) system, and the antenna samples module first receives the original signal by sampling, then converts the band-pass signal into baseband signal. Signal processing module divides the baseband signal into in-phase and quadrature (I/Q) channels and normalizes the signal. In the classification module, CraNET is first trained by the labeled data and is utilized to classify the modulation format of the unlabeled data. The in-phase and quadrature components are expressed as $I = Re[r(n)]$ and $Q = Im[r(n)]$. The complex representation of signal is given by:

$$r(n) = r_I(n) + r_Q(n) = \mathrm{Re}[r(n)] + j * \mathrm{Im}[r(n)] \tag{3}$$

Since the received signal may have great difference in amplitude, the signal amplitude is normalized. After the signal amplitude is normalized, a channel in complex signal is given by:

$$r_{I/Q}(n) = \frac{[r_{I/Q}(0), r_{I/Q}(1), ..., r_{I/Q}(N)]}{\max\{abs[r_{I/Q}(0), r_{I/Q}(1), ..., r_{I/Q}(N)]\}} \tag{4}$$

In the classification module, the well-trained classifier learns signal features by extracting I/Q features. Finally, the classifier gives the score value corresponding to each signal and selects the label corresponding to the maximum score as the output. The classification module outputs the classification results M according to the maximum probability value, and K is the number of a set of candidate modulations which is $M \in \{H_1, H_2, ..., H_K\}$.

3 CraNET Based Modulation Classification

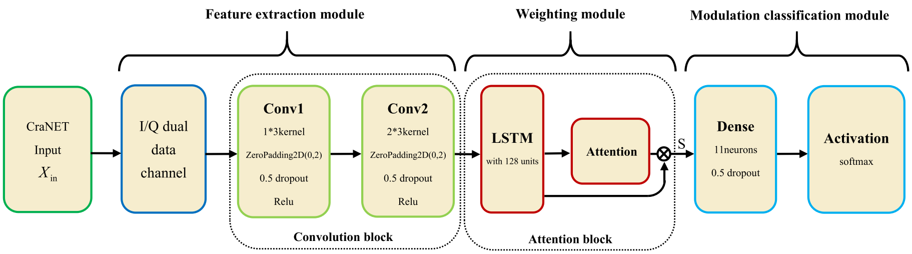

Fig. 2. CraNET based network structure

This section mainly introduces the CraNET structure, and then illustrates the classification process of the CraNET.

Figure 2 shows the overall network structure, which is divided into three parts: feature extraction module, weighting module and modulation classification module. The feature extraction module extracts the features of I/Q channels of the signal by convolution block which has two convolution layers. Then, these features are converted to hidden states and weighted by attention block. Finally it outputs s as the representation vector to the modulation classification module. The modulation classification module is composed of dense layer with softmax function, and the classification decision is finally conducted by using the representation vector. Since the convolutional neural network can find the intrinsic features of the data, it should be placed at the front of the network structure. Then the attention block can find the connection between the context information of the input sequence, and selectively focuses on the specific frame-related information of the input sequence. The proposed network takes full consideration of the data characteristics of the signal and combines the advantages of the attention mechanism.

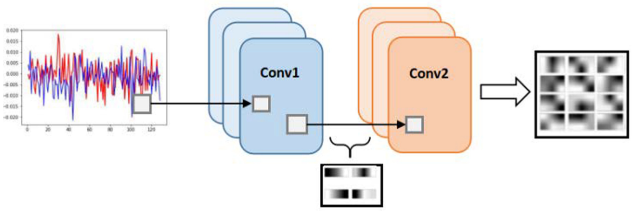

Fig. 3. CNN based network structure

3.1 CNN for Feature Extraction

Figure 3 shows the structure of CNN. Feature extraction is a crucial step in the modulation signal processing. In this module, both the analog signal and digital signal are considered for the experimental data set by the automatically features learning ability of CNN. CraNET first needs to find the intrinsic expression of the data through the convolutional network. The modulation recognition process in most wireless communication systems can be regarded as linear mixing, time shift, rotation, scale scaling and invariance with convolution. Convolutional neural networks have the characteristics of shift invariance in image processing problems [18, 19], thus, the convolution layer is used in the first two layers of the whole classification model.

Table 1. Parameters of CNN

Input: I/Q signal matrix (Dimension:1*2*N)				
Layers	**Kernel Size**		**Padding**	**Stride Step**
Conv1	256@(1*3)	(0*2)	1	0.5
Conv2	80@(2*3)	(0*2)	1	0.5
Output: representation tensor (Dimension:80*2*N)				

Parameters of CNN as shown in Table 1. To balance the tradeoff between accuracy and efficiency, two convolution layers are used in the convolution block. The first convolution layer uses 256 convolution kernels of size 1 * 3, the second convolution layer uses 80 convolution kernels of size 2 * 3, and both layers use ReLU function as the activation function. Meanwhile, the weight $||W||_2$ of two norms are added to prevent the parameter value too large and overfitting in the process of parameter updating of convolution layer, and the dropout method is used to improve the generalization ability of the network. The first layer is used to get the edge and gradient detection of the input data in one dimension, and the second layer integrates the features extracted by the first layer through a larger convolution kernel. The convolution block converts the I/Q data into a representation tensor and passes it to the weighting module.

3.2 Attention Mechanism for Weighting

Figure 4 shows the calculation process of attention weight distribution. First, input the representation tensor $(x_1,...,x_T)$ into the LSTM layer after feature extraction, then the LSTM cell transforms the representation tensor into hidden states. Attention block takes the hidden states $h_1,...,h_T$ into the softmax function, and the attention distribution α_t is given out by the softmax function. The final output vector s is calculated by the weighted sum of all hidden states.

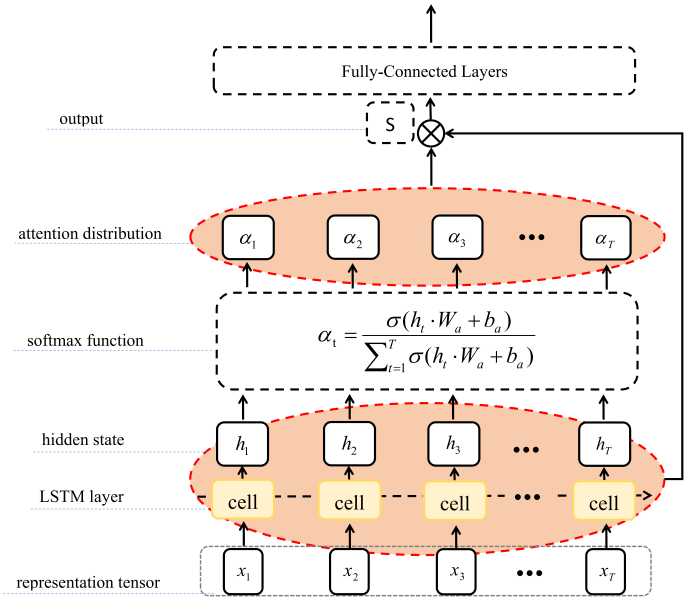

Fig. 4. Attention mechanism based network structure

The attention mechanism is based on LSTM network, and LSTM network is actually a RNN network. Denote the hidden states at the T time steps of an LSTM layer as h_1,...,h_T, where T is the number of time steps. The hidden state h_t is obtained by x_t and h_{t-1}. Meanwhile, T is the first dimension of the representation tensor (x_1,...,x_T) and its value is 80. Unit is the hidden neuron of the gate structure in the LSTM cell, and its value is equal to the length of the output vector of the LSTM cell. To summarize the final output of LSTM layer, a feasible method is to concatenate all hidden states as one vecter and then output $H = (h_1,...,h_T)$. In this way, the extracted information from all time steps will be used.

Attention mechanism is introduced to measure the importance of all hidden states. The contribution of each hidden state to the final output is enhanced or weakened by weighted summation. In this paper, the proposed model uses a self-attention mechanism to adaptively derive the final output of LSTM layer using all hidden states. First, LSTM layer maps the input representation tensor (x_1,...,x_T) to a sequence of hidden states h_1,...,h_T, then sends h_t to a softmax activation function through a shared time-distributed neural network layer with attention weight matrix W_a and bias b_a:

$$u_t = h_t \cdot W_a + b_a \tag{5}$$

The attention distribution α_t is calculated by sending the output u_t through a softmax activation layer:

$$\alpha_t = \frac{\sigma(u_t)}{\sum_{t=1}^{T} \sigma(u_t)} = \frac{\sigma(h_t \cdot W_a + b_a)}{\sum_{t=1}^{T} \sigma(h_t \cdot W_a + b_a)} \tag{6}$$

The final output vector s is the weighted sum of all hidden states H, its vector dimension is the same as h_t:

$$s = \sum_{t=1}^{T} \alpha_t h_t \tag{7}$$

By adding attention mechanism, the network achieves a better trade-off between increasing the amount of information that sent to the next layer and reducing the number of parameters that need to be trained in the next layer.

In the modulation classification module, the obtained s is fed into the fully connected layer as the final output of the attention block and the probability of all hypothetical modulation modes is calculated by the softmax function, the hypothetical modulation mode with the largest probability is selected as the final recognized modulation mode. Modulation signal labels $y \in \{1, 2, ..., K\}$ have K categories. When given a sample x, w_k is the weight vectors of the k-th class, the conditional probability of the k-th class for softmax regression prediction is:

$$p(y = k|x) = \frac{\exp(w_k^T x)}{\sum_{k=1}^{K} \exp(w_k^T x)} \tag{8}$$

Finally, the classification results are output by selecting the maximum statistics, which is given as follows:

$$k = \arg \max_{1 \le k \le K} p(y = k|x) \tag{9}$$

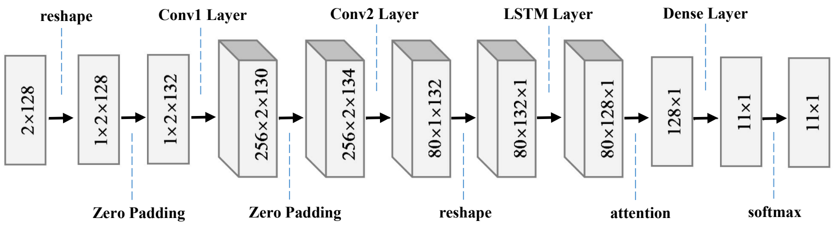

Fig. 5. Data-driven structure of CraNET

4 Simulation

In this section, the recognition and classification of signal features are carried out by constructing neural networks, and the training process of network learning is accelerated by tensorflow-gpu framework. The candidate modulation set M contains six kinds of modulation modes, which is given by M = *[AM-DSB, AM-SSB, BPSK, CPFSK, GFSK, PAM4]*. The signal data set is taken from RadioML [20] and SNR ranges from −20 dB to 18 dB with a step of 2 dB. Each sample vector has a size of 2 × 128. The Data-driven structure of CraNET is shown in Fig. 5. Adam optimization algorithm is used to replace the traditional random gradient descent algorithm. During the training, the loss value of the verification set is monitored by the Earlystopping in the callbacks, and the maximum number of tolerance is set to 8 times. The performance of the neural network used for modulation signal recognition is evaluated by classification accuracy to verify the performance and robustness of CraNET. The average classification accuracy is given by:

$$P_{CA} = \sum_{k=1}^{K} P(\hat{H} = H_k | H_k) P(H_k) \tag{10}$$

where K is the number of category, H_k is the hypothesis of k-th class.

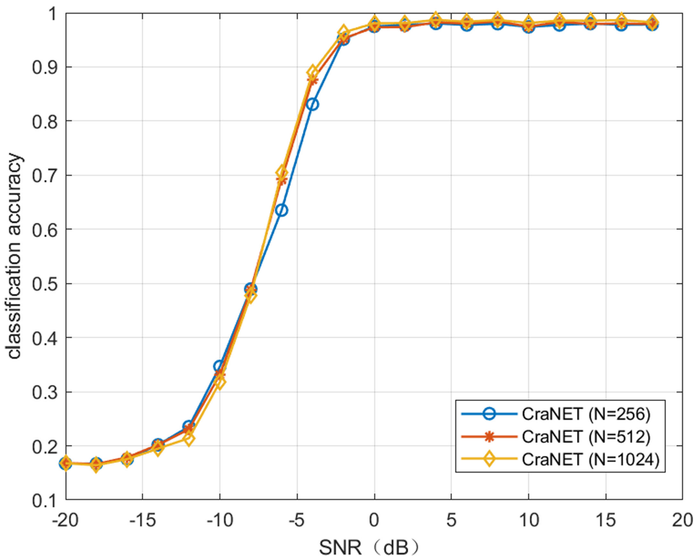

Fig. 6. The performance of CraNET with different batch size

Figure 6 shows the accuracy on different batch size set as 256, 512 and 1024 of CraNET, respectively. The batch size parameter determines the direction of gradient descent, and the direction of gradient descent is one of the influencing factors of classification accuracy. If the dataset is sufficient, the gradient calculated by training with small batch data is almost the same as that by training with large batch data. Thus, it can be found that the classification performance increases slightly with the batch size. It can be observed that SNR is close to 1.0 when it is above 0 dB, and CraNET has good performance with the accuracy more than 0.8 above −5 dB. Therefore, CraNET has better performance under normal SNR condition.

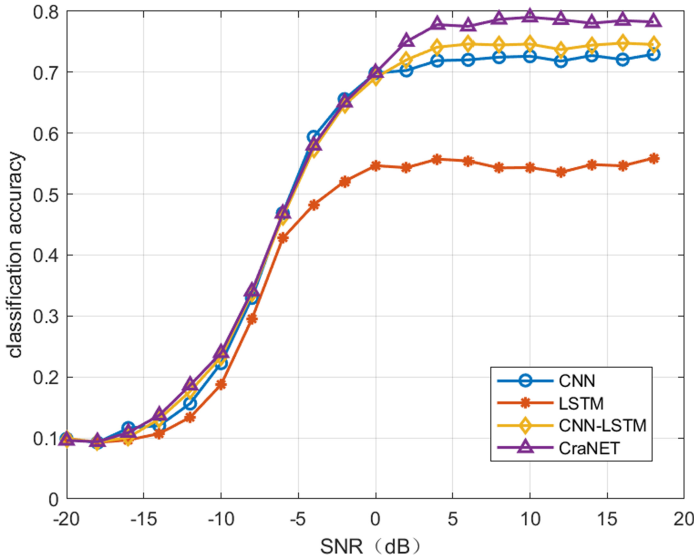

Fig. 7. The performance comparison with other AMC methods

Figure 7 compares the accuracy of CraNET with that of CNN, LSTM and CNN-LSTM network after training on the same data set. The data set contains eleven modulation modes of RadioML, which is given by *M = [8PSK, AM-DSB, AM-SSB, BPSK, CPFSK, GFSK, PAM4, QAM16, QAM64, QPSK, WBFM]*. Since the data set is generated through real channel simulation. The influence of various external factors is strengthened, and the situation is closer to the real modulation signal, which makes the accurate recognition rate decrease. Due to the increase of candidate types, the classification performance also declined. Through the curve in the graph, we can see that the accurate recognition rate of all classification network models is low when the SNR is lower than 0 dB, and the accurate recognition rate tends to a high constant value gradually when the SNR is higher than 0 dB. Overall, the accurate recognition rate of CraNET is always higher than that of other networks above −5 dB, and the highest accurate recognition rate of CraNET is about five percent higher than that of the CNN. Therefore, it is verified that CraNET has better performance than other networks.

5 Conclusions

This paper proposes a novel modulation classification framework using Convolutional Recurrent Attention Network which can be divided into three parts: feature extraction module, weighting module and modulation classification module. CraNET's structure is mainly composed of convolution block and long short term memory network (LSTM) based attention block. Convolution block extracts the signal features in the feature extraction module. Different from LSTM which only outputs the last hidden state of LSTM layer, the proposed LSTM based attention block uses all hidden states to calculate the output in the weighting module. The contribution of each hidden state to the final output is enhanced or weakened by weighted summation. Moreover, the modulation classification decision is made in the modulation classification module using the softmax function. CraNET is compared with other networks by Monte Carlo simulation to verify the superiority and robustness of CraNET.

References

1. Jain, A.K., Acharya, R., Jakhar, S., Mishra, T.: Fifth generation (5g) wireless technology "revolution in telecommunication". In: 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), pp. 1867–1872 (2018)
2. Sadeghi, M., Larsson, E.G.: Adversarial attacks on deep-learning based radio signal classification. IEEE Wirel. Commun. Lett. **8**(1), 213–216 (2019)
3. Dobre, O.A., Abdi, A., Bar-Ness, Y., Su, W.: Survey of automatic modulation classification techniques: classical approaches and new trends. IET Commun. **1**(2), 137–156 (2007)
4. Tian, X., Chen, C.: Modulation pattern recognition based on resnet50 neural network. In: 2019 IEEE 2nd International Conference on Information Communication and Signal Processing (ICICSP), pp. 34–38 (2019)
5. Fragkiadakis, A.G., Tragos, E.Z., Askoxylakis, I.G.: A survey on security threats and detection techniques in cognitive radio networks. IEEE Commun. Surv. Tutor. **15**(1), 428–445 (2013)
6. Gouldieff, V., Palicot, J., Daumont, S.: Blind modulation classification for cognitive satellite in the spectral coexistence context. IEEE Trans. Signal Process. **65**(12), 3204–3217 (2017)
7. Huang, S., Yao, Y., Wei, Z., Feng, Z., Zhang, P.: Automatic modulation classification of overlapped sources using multiple cumulants. IEEE Trans. Veh. Technol. **66**(7), 6089–6101 (2017)
8. Huan, C.-Y., Polydoros, A.: Likelihood methods for MPSK modulation classification. IEEE Trans. Commun. **43**(2/3/4), 1493–1504 (1995)
9. Chang, D., Shih, P.: Cumulants-based modulation classification technique in multipath fading channels. IET Commun. **9**(6), 828–835 (2015)
10. Karami, E., Dobre, O.A.: Identification of SM-OFDM and AL-OFDM signals based on their second-order cyclostationarity. IEEE Trans. Veh. Technol. **64**(3), 942–953 (2015)
11. O'Shea, T.J., Corgan, J., Charles Clancy, T.: Convolutional radio modulation recognition networks. arXiv preprint arXiv:1602.04105 (2016)

12. Sun, J., Wang, G., Lin, Z., Razul, S.G., Lai, X.: Automatic modulation classification of cochannel signals using deep learning. In: 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP), pp. 1–5 (2018)
13. Mendis, G.J., Wei, J., Madanayake, A.: Deep learning-based automated modulation classification for cognitive radio. In: 2016 IEEE International Conference on Communication Systems (ICCS), pp. 1–6 (2016)
14. Yao, T., Chai, Y., Wang, S., Miao, X., Bu, X.: Radio signal automatic modulation classification based on deep learning and expert features. In: 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), vol. 1, pp. 1225–1230 (2020)
15. Tu, Y., Lin, Y., Wang, J., Kim, J.U.: Semi-supervised learning with generative adversarial networks on digital signal modulation classification. Comput. Mater. Continua **55**(2), 243–254 (2018)
16. Lin, Y., Tu, Y., Dou, Z., Chen, L., Mao, S.: Contour Stella image and deep learning for signal recognition in the physical layer. IEEE Trans. Cogn. Commun. Netw., 1 (2020)
17. Lin, Y., Zhao, H., Ma, X., Tu, Y., Wang, M.: Adversarial attacks in modulation recognition with convolutional neural networks. IEEE Trans. Reliab., 1–13 (2020)
18. Wang, W.G.L., Li, X.: Deep convolutional neural network and its application in image recognition of road safety projects. Int. J. Perform. Eng. **15**(8), 2182 (2019)
19. Songyin, B.X., Pan, D.W., En, F., Pan, F.: Traffic sign detection via efficient ROI detector and deep convolution neural network. Int. J. Perform. Eng. **16**(10), 1566 (2020)
20. O'Shea, T.J., West, N.: Radio machine learning dataset generation with gnu radio. In: Proceedings of the 6th GNU Radio Conference (2016)