

Feature Fusion Convolutional Network Based Automatic Modulation Classification

Jiajie Mai, Shanchuan Ying, Nanxin Wang, Sai Huang
{maijiajie@bupt.edu.cn, yingshanchuan@bupt.edu.cn, nancingwang@bupt.edu.cn,
huangsai@bupt.edu.cn}

International School, Beijing University of Posts and Telecommunications, Beijing, China
Key Laboratory of Universal Wireless Communications, Ministry of Education

Abstract. Automatic modulation classification (AMC) technology, which utilizes to classify different kinds of signals using various expert features, typically the constellation map and the cyclic spectrum density graph, plays a significant role in spectrum monitoring and radio communication. Aiming at the commonly used features-based (FB) approach in practice, we propose a novel AMC model that jointly utilizes the constellation map and cyclic spectrum density graph as the signal features in this paper. In order to provide a solution for the model, we propose a feature-based and supervised network in Single in Single out (SISO) system to identify seven kinds of signals, which is called Feature Fusion Convolutional Network (FFCN). Simulation results are provided to show that FFCN can classify different types of signals with high probability. Typically when signal to noise ratio (SNR) is low, the success rate of the proposed network is 2% higher on average, which proves that our network has better performance.

Keywords: automatic modulation classification, residual neural network, constellation, cyclic spectrum density graph

1 Introduction

Along with the nowadays rapid development of the wireless communications, mountainous technologies like software radio and the fifth generation (5G) wireless systems arise, and radio communication is playing an increasingly essential role in the advanced area [1]. Meanwhile, spectrum monitoring, which ensures the normal operation of wireless services and strengthens the spectrum management, gets harder and the public have paid more and more attention on it [2] [3]. In the whole process of modulation format categories, automatic modulation classification (AMC) is vitally needed due to its strong ability to improve the performance of classifying signal, typically for signals corrupted by different kinds of noise and fading [4]. Not only used in modulation classification, but AMC also takes responsibility for identifying the interference signal, filtering unhealthy and dangerous information in many security applications for decades, which acts as the bridge between receive signal and signal demodulation.

Generally, there are two basic method of the AMC algorithm: likelihood-based (LB) and features-based (FB) approach [5]. Swami *et al.* achieved AMC using the approach of LB [6] and Ghasemzadeh *et al.* employed FB to recognize the modulations [7]. The LB method is used to calculate the likelihood equation of the received signal in different modulation modes.

When the channel and noise are fixed, the algorithm can maximize the probability of correct classification so it is regarded as the best classifier in the scenario. Despite of its excellent performance, it may cause the high computational complexity and poor performance when handling small-scale or sparse data. Compared to the likelihood-based approaches, the FB method extracts the signal features involving cumulant, wavelet feature, cyclostationary characteristic, *etc.*, to classify the format with lower computational complexity. Although the performance of FB is worse than LB in many scenario, FB reduces the computational complexity and achieve the low time cost and high robustness. As a result, FB is widely implemented in practical applications.

There are many contributions on AMC using classifiers belonging to the supervised machine learning. The authors of [8] utilized Support Vector Machine (SVM) to classify the signal, while the authors of [9] employed k-NearestNeighbor (KNN). Rather than using classifiers, Kim *et al.* used Deep Neural Network to achieve the goal [10]. The contributions above demonstrates that the machine learning has been used widely in AMC [11] [12]. Residual neural network (ResNet) [15] is a sort of supervised machine learning and can identify patterns in data according to diverse inputs such as examples, direct experience, or instruction. In this article, we employ ResNet18 which is a ResNet whose number of layers with weights is 18. After unceasing training, ResNet is able to make better decisions in the future based on the learning experience.

Regarding to the approaches of processing received discrete data by antennas, the constellation has advantages of recognizing Quadrature Amplitude Modulation (QAM) signal while the cyclic spectrum density has better performance when recognizing Frequency Shift Keying (FSK) signal. Huang *et al.* employed the constellation maps [13] and Ma *et al.* take advantages of the cyclic spectrum density maps to process the received data [14], respectively. However, we can not know which modulation the received signal belongs to in advance, so it is necessary to find a network that has relatively great performance whatever the modulation type of the received signal is. Therefore, we need to take into account their own advantages of the two approaches of processing data and to combine them together.

In order to provide a optimal solution for automatic modulation classification, we propose a Feature Fusion Convolutional Network (FFCN) based on modified ResNet, which can apply what has been learned in the past to newly coming data using labeled examples. The main contribution of our work are summarized as follows.

- 1) In contrast to traditional methods, we propose a novel AMC model based on the discrete signal processed by both constellation map and cyclic spectrum density graph. We normalize the heterogenous data from the two maps into the same specification and combine it in the same dimension, which allows estimating modulation format by taking advantages of the two approaches.
- 2) According to the form of input, we modify Resnet18 in the proposed FFCN. Depending on the efficiency and maximum a posteriori, cross entropy is selected as our loss function to promote the classification ability of ResNet18 to the maximum extent.
- 3) Extensive simulations are carried out and it is proven that the proposed network is able to automatically learn more discriminative features for recognition performance

improvement due to the advantage of powerful capability in extracting features of matrix-shaped data.

The rest of the paper is organized as follows. Section II presents the detailed introduction of the two ways to process signal data and the system model. The network structure and data combining methods are showed in Section III. Section IV indicates the simulation results and Section V draws a conclusion of this paper.

2 SYSTEM AND SIGNAL MODEL

There are two parts in this section. In the first part, the system model is outlined to show the whole process of AMC. In the latter part, we introduce the additive Gaussian noise (AWGN) channel and seven sorts of signal which would be classified.

2.1 System Model

The overall structure of signal transmission and reception is shown in Fig. 1. Firstly, different signals with different modulation format are generated in the source and transmitted to the receiver via the antenna after modulation. And in actual situation of spectrum monitoring radio detection, it usually does not require a large number of devices, so we assume that AMC is applied to a single receive, that is, the cooperation between different devices in different environments is not considered then the receiver can complete the task independently. Then, in the part of the receiver, after gaining the signal from a specified channel by the antenna, AMC technology is utilized to recognize the signal's modulation format through two main step that are feature extraction and classification respectively. Eventually, an estimated modulation format will be given to the demodulator to obtain the exact information contained.

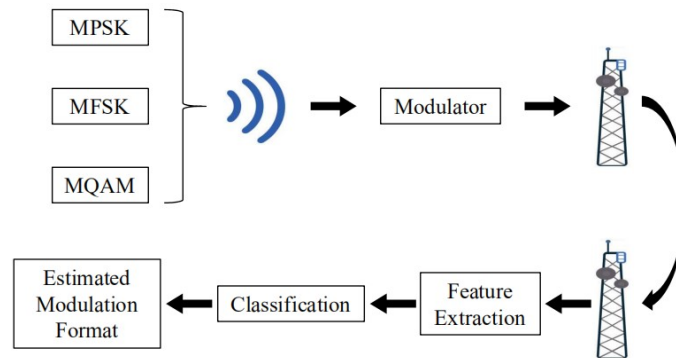


Fig. 1. System model

2.2 Signal Model

In this paper, there are seven kinds of signals would be classified: Binary Phase Shift Keying (BPSK), Quadrature Phase Shift Keying (QPSK), Binary Frequency Shift Keying (BFSK), Quadrature Frequency Shift Keying (QFSK), 16QAM, 32QAM and 64QAM. Initially, we provide the expression of baseband signals and analyze the noise. The received baseband signal can be given as follows.

$$r(n) = h e^{2\pi f_0 n + \theta_0} x_r(n) + w(n) \quad (1)$$

where $n = 1, 2, \dots, N$. And h is the channel parameter that follows Rayleigh distribution and is invariant during the identification process f_0 and θ_0 denote the frequency and phase offset. $x(n)$ represents the n^{th} complex symbol with unit average power, while $w(n)$ represents the noise. As far as the noise concerned, we choose the additive Gaussian noise. The reason of it is that we prefer synthetic noise to real noise, which brings the advantage that it's easy to analyze the problem and design the algorithm and AWGN is a simple and good approximation simulation to deal with complex situations without knowing the actual noise distribution. The based function of Gaussian Distribution is given as follows.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (2)$$

where μ represents the mean value, and σ represents the standard deviation.

The seven involved signals can be divided into three sorts: Multiple Phase Shift Keying (MPSK), Multiple Frequency Shift Keying (MFSK) and Multiple Quadrature Amplitude Modulation (MQAM). We then discuss the expressions of the three sorts of signals in discrete forms after they are received by antennas.

1) The transmission signal of MPSK modulation can be expressed in complex numbers, and the expression is:

$$x_z(n) = A \sum_m \exp\{j(2\pi f_c n + \phi_i)\} P(n - mT_z) \quad (3)$$

In the expression above, A represents the amplitude having a determined value. $P(n)$ represents the baseband pulse which has a duration of T_s , whose wave form depends on the selection of modeling filters. And f_c is the center frequency. Moreover, the value of ϕ which represents the corresponding phase of baseband signal should satisfy:

$$\phi_i \in \left\{ \frac{2\pi \times 1}{M}, \dots, \frac{2\pi \times M}{M} \right\} \quad (4)$$

where M is the mode of the modulation.

2) MFSK: The transmission signal of MFSK modulation can be expressed in complex numbers, and the expression is:

$$x_z(n) = A \sum_m \exp\{j(2\pi f_m n)\} P(n - mT_z) \quad (5)$$

And the frequency of the sub-carriers, f_i , should satisfy:

$$f_i \in \{f_1, \dots, f_M\} \quad (6)$$

In the expressions above, A and M also represents the amplitude having a determined value and the mode of the modulation respectively. The rectangular pulse, denoted as $P(n)$, has a duration of T_s , and f_{space} represents the frequency space, whose expression is $f_{space} = f_{i+1} - f_i, i = 1, \dots, M-1$. Furthermore, the center frequency, denoted as f_c , has the following expression.

$$f_c = \frac{1}{M} \sum_{i=1}^M f_i \quad (7)$$

3) MQAM: The signals can be expressed as:

$$S_{MQAM} = a_{ic} g_T(n) \cos(w_c n) - a_{is} g_T(n) \sin(w_c n) \quad (8)$$

In the expression above, a_{ic} and a_{is} are the in-phase and orthogonal component of the amplitude respectively, both of whose values are integers ranging from 1 to \sqrt{M} . Additionally, $g_T(n)$ is the pulse of symbols formed by baseband filters.

3 PROPOSED METHOD FOR MODULATION CLASSIFICATION

In Section 3.1, we introduce two sorts of data, constellation and cyclic spectrum density, which act as the heterogeneous input of ResNet18, and we also deliver how to formulate and normalize them into the specific matrix forms. Then, the structure of ResNet18 is given in Section 3.2. Eventually in Section 3.3, a brief introduction will be given about the chosen loss function, which is employed to optimize the whole performance.

3.1 Heterogeneous Input Data

Constellation. In the field of digital communication, digital signals are often represented on a complex plane to visually represent signals and the relationship between signals. The constellation has two axes. The horizontal X-axis is related to the in-phase carrier and the vertical Y-axis is related to the orthogonal carrier. After receiving diverse signals, we restore them to raw IQ sequences and organize which into the constellation map responding to Real and imaginary axis.

Fig. 2 shows the normalized constellation images for the input data. It demonstrates that the difference between BFSK and QFSK is scarcely possible to identify no matter at any signal to noise ratio (SNR), which is why we select the cyclic spectral density to make up this defect.

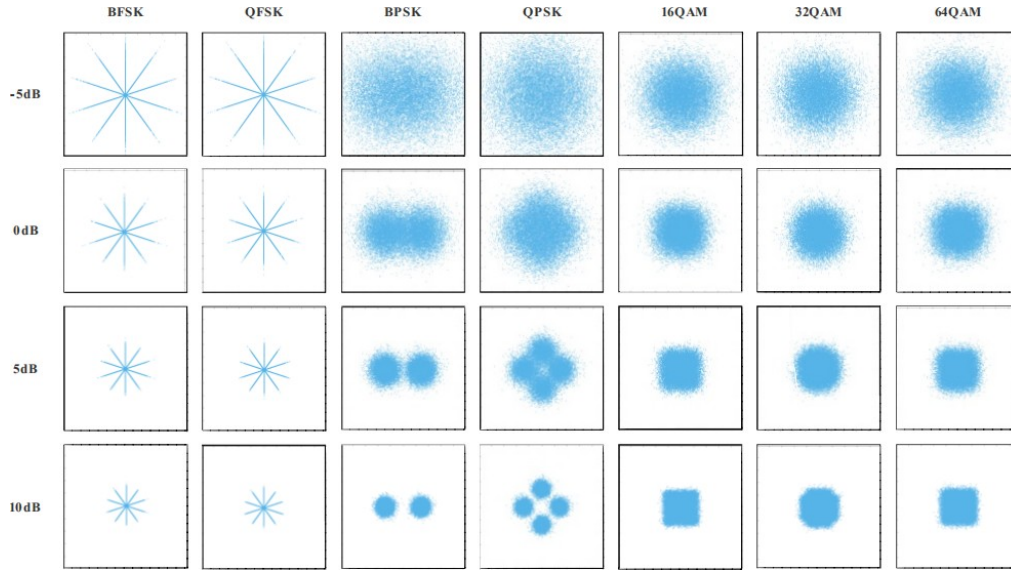


Fig. 2. Constellation map of seven types of modulations versus SNR

Cyclic Spectrum Theory. If the mean and autocorrelation function of random signal $x(t)$ is a periodic function of time, cycle is T , then $x(t)$ is called generalized cycle stationary signal. Its average $m_x(t)$ and autocorrelation function $R_x(t, \tau)$ should respectively meet the function below.

$$\begin{aligned} m_x(t) &= m_x(t + T) \\ R_x(t, \tau) &= R_x(t + T, \tau) \end{aligned} \quad (9)$$

Expand the periodic function $R_x(t, \tau)$ with Fourier series, and we have:

$$\begin{aligned} R_x(t, \tau) &= \sum_{m=-\infty}^{\infty} R_x^\alpha(\tau) e^{jm \frac{2\pi}{T} t} \\ &= \sum_{m=-\infty}^{\infty} R_x^\alpha(\tau) e^{j2\pi \alpha t} \end{aligned} \quad (10)$$

where $\alpha = \frac{m}{T}$ and the Fourier coefficient $R_x^\alpha(\tau)$ is:

$$\begin{aligned}
R_x^\alpha(\tau) &= \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} R_x(t, \tau) e^{-j2\pi\alpha t} dt \\
&= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) x^*(t - \tau) e^{-j2\pi\alpha t} dt \\
&= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-j2\pi\alpha t} dt
\end{aligned} \tag{11}$$

The coefficient $R_x^\alpha(\tau)$ represents the cyclic autocorrelation intensity of the frequency α , which is also a function of τ referred to as the cyclic (self) correlation function. The frequency α is called the cyclic frequency of the signal $x(t)$, which can be multiple in cyclostationary signals (including zero cycle frequency and non-zero cycle frequency). Among them, the zero cycle frequency corresponds to the stationary part of the signal, and only the non-zero cycle frequency can indicate the signal's cyclic stability. If $\alpha = 0$, then $R_x^0(\tau) = R_x(\tau)$ is the autocorrelation function of the stationary signal.

Additionally, the Fourier transform of $R_x^\alpha(\tau)$ is known as the Cyclic spectral density (or spectral correlation density function).

$$S_x^\alpha(f) = \int_{-\infty}^{\infty} R_x^\alpha(\tau) e^{-j2\pi f\tau} d\tau \tag{12}$$

Fig. 3 shows the normalized cyclic spectrum density maps of all types of signals at different SNR that the height of the 3D model represents the magnitude of the density value. It indicates that the difference between BPSK, BFSK, QPSK, QFSK is obvious and easy to recognize whatever the SNR is. However, in the aspect of QAM signals we can hardly distinguish whether the signal is QPSK, 16QAM, 32QAM or 64QAM whatever the SNR is, which is the disadvantage using cyclic spectrum density to classify the QAM signals and it has been proven that the features provided by the constellation can successfully complement.

Process of Formulation and Normalization. After obtaining the images of the two formats, rational process of formulation and normalization is needed to implement the complementation of the heterogeneous data. Firstly, we formulate the two kinds of images into 50×50 matrix in disparate approach. In terms of the constellation, we fill out each grid of the matrix with the number of IQ points involved in every single grid, after which we normalize the matrix through the function given below.

$$X_{Nor} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{13}$$

where X_{Nor} is the normalized number of IQ points. X_{max} and X_{min} represent the maximum and minimum number of IQ points respectively, and then X is the number of IQ points in a certain

grid. Consequently, we obtain one of the heterogeneous input of ResNet18.

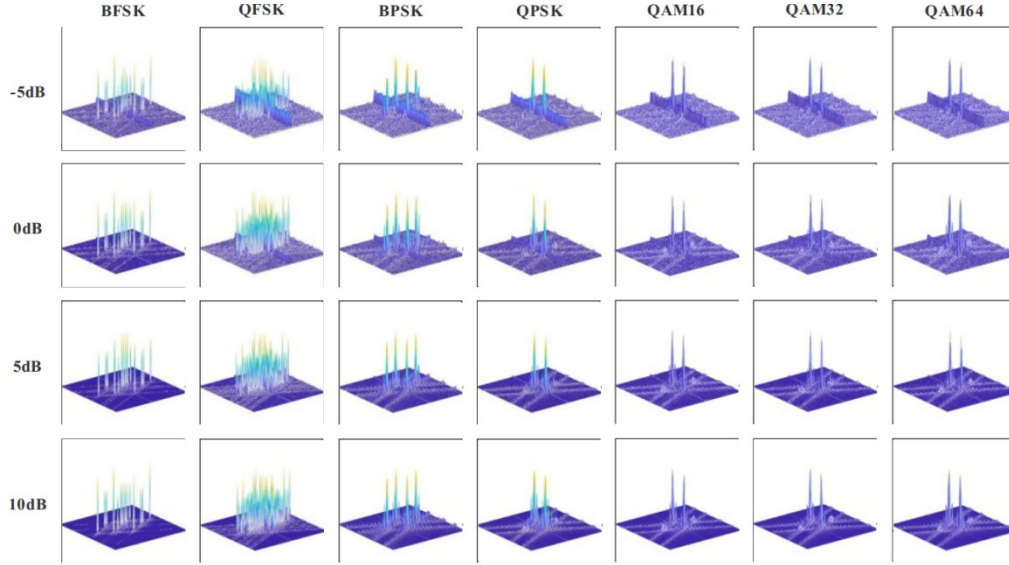


Fig. 3. Cyclic spectrum density map of seven types of modulations versus SNR

In terms of Cyclic Spectrum Density, the data has been normalized through been divided by the maximum value when the map is generated. We calculate the each average value of 50×50 grids based on the number of sample points in a certain grid, after which we acquire the other heterogeneous input.

3.2 Structure of Modified ResNet18

In this model, we prefer ResNet18 to be the classifier. The architecture of the ResNet18 model is shown in Fig. 4, which consists of an input layer on the left and 17 convolution layer followed by an average pool using softmax. Moreover, detailed introduction of the residual blocks are displayed in Table 1. The sample matrix of heterogeneous data is first sent to the input layer, and after various non-linear operation of the hidden layers, estimated modulation formats will be given in the final stages.

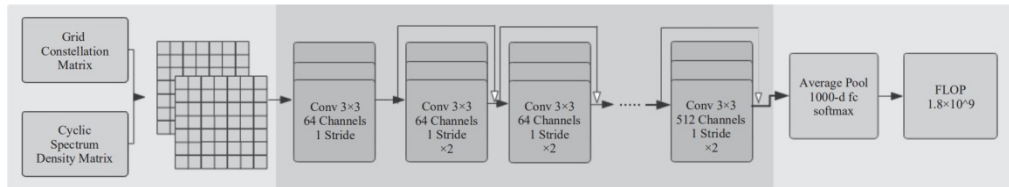


Fig. 4. Structure of modified ResNet18
Table 1. DETAILS OF THE RESNET18 STURCTURE

	Residual Block	Output Size
conv_1	$\begin{bmatrix} 3 \times 3 & 64 \\ 3 \times 3 & 64 \end{bmatrix} \times 2$	56×56
conv_2	$\begin{bmatrix} 3 \times 3 & 128 \\ 3 \times 3 & 128 \end{bmatrix} \times 2$	28×28
conv_3	$\begin{bmatrix} 3 \times 3 & 256 \\ 3 \times 3 & 256 \end{bmatrix} \times 2$	14×14
conv_4	$\begin{bmatrix} 3 \times 3 & 512 \\ 3 \times 3 & 512 \end{bmatrix} \times 2$	7×7

To make the network more practical and perform well, we modify it on two points. The first and foremost one is the modification before the second convolution layer. In the initial design of ResNet18, the first step in network operation is to pass the feature through a 7×7 convolution layer and a 3×3 maximum pooling layer. The significance of this block is to reduce the scale of the input feature, which is normally conducive to the process of feature extraction. If it is for large-sized data such as high-resolution pictures, this structure can definitely achieve good outcomes. But since the quality and length of the signal received in a single time, the number of sample points generated is usually too small to build such a large matrix. Therefore, we changed the original block to a convolution layer with stride equals to 1, just like the convolution network followed. In this way, not only can we prevent loss of information due to downsizing, we can also implement the process of feature extraction, which is crucial to the classification. The next modification is selection of loss function, which we will introduce in the next section.

Due to the sufficiently good performance of ResNet18, we do not make changes to its main framework. The consecutive convolution layers aim at extracting features, which play a crucial role in the process of feature extraction. And the link between two convolution layers is the rectified linear unit (ReLU) activation function. The most creative and incredible structure is definitely the residual block, which is implemented by the shortcut connection and we can operate an element-wise addition between the input and output of this block. This operation does not append extra parameters and calculations to the network, but greatly improve the performance of this model in terms of training effect and speed.

3.3 Loss Function

To enhance the recognition of this model, we use cross entropy function as our loss function, which is the derivative of relative entropy. Relative entropy, also known as KL divergence, can be used to measure the difference between two distribution if we have two separate probability distributions $p(x)$ and $q(x)$ for the same random variable x . In the context of machine learning, $D_{KL}(p||q)$ is often called the information gain achieved if p is used instead of q . The function of KL divergence is given below.

$$\begin{aligned}
D_{KL}(p\|q) &= \sum_{i=1}^n p(x_i) \log\left(\frac{p(x_i)}{q(x_i)}\right) \\
&= \sum_{i=1}^n p(x_i) \log(p(x_i)) - \sum_{i=1}^n p(x_i) \log(q(x_i)) \\
&= -H(p(x)) + \left[- \sum_{i=1}^n p(x_i) \log(q(x_i)) \right]
\end{aligned} \tag{14}$$

The first part of the last equation happens to be the entropy of p and the latter part of the equation is the cross entropy. In machine learning, we need to evaluate the difference between label and predictions. KL divergence is definitely a suitable choice. Additionally, due to the invariance of the first part, $-H(p(x))$, in the optimization process, we only need pay attention to the cross entropy.

4 NUMERICAL RESULTS

In this section, simulations are conducted to verify the performances of our proposed FFCN. The modulation set is $M = \{BPSK, QPSK, BFSK, QFSK, 16QAM, 32QAM, 64QAM\}$, and the calculation of the average probability of correct classification P_{cc} is showed below.

$$P_{cc} = \sum_{z=1}^S P(\hat{H} = H_z | H_z) P(H_z) \tag{15}$$

where $H_z \in M$. In this equation above, S is the total number of candidate modulations and H_z is the prior probability, and it is assumed that the prior probability of each modulation is identical. Furthermore, $P(\hat{H} = H_z | H_z)$ represents the probability that the modulation is recognized as H_z under the hypothesis H_z .

Fig. 5 shows the probability of correct identification for seven modulation models when SNR increases from -5 to $10dB$ with FFCN. The seven modulation modes include BPSK, QPSK, BFSK, QFSK, 16QAM, 32QAM and 64QAM. From the chart we can see that no matter which sorts of signals, the success rate of them continuously increase and remain stable at approximately 97% eventually at a SNR of 10 . Moreover, FFCN shows its outstanding performance in the aspect of BPSK and QPSK, with the figure of over 95% when the SNR is around 5 . To sum up, we can draw a conclusion that the FFCN performs well in classifying all the seven kinds of signals.

Fig. 6 shows the probability of correct identification for 3 kinds of features with different dimensions or data-processing approach when SNR increases from -5 to $10dB$ with FFCN. In the proposed model, the two channel of heterogeneous data take a role of the input. From the graph its clear that the novel type of input outperforms both the constellation and the cyclic spectrum density throughout the range of SNR. More specifically, the success rate of FFCN is

over 90% when SNR is increased to 2 eventually. Additionally, the performance of cyclic spectrum density is better than constellation when SNR is low, while it is inverse when SNR is high. It is proven that to employ the combination of heterogenous data is feasible and has better performance than to adopt single data.

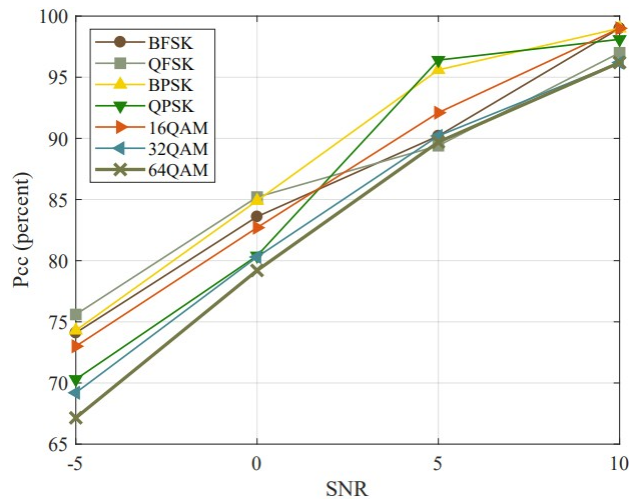


Fig. 5. Performance comparison of different kinds of modulation versus SNR

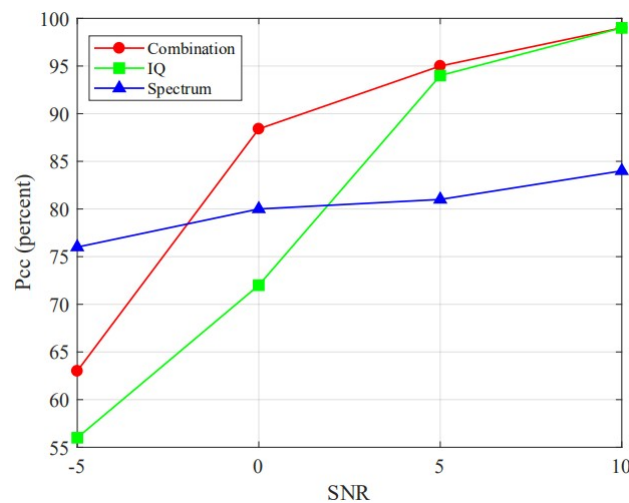


Fig. 6. Performance comparison of different kinds of inputs for seven types of modulations versus SNR

Fig. 7 shows the probability of correct identification for 4 diverse network (classifier) when SNR increases from -5 to 10dB with the same input of FFCN. From the illustration, when SNR is high, It can be observed that the success rate of all the four classifiers are similar because of the low noise level. In the condition of SNR 0 and 5, the figure of FFCN is outstanding at 95% and 88% respectively, which demonstrate the advantages of ResNet18. However, in the condition of low SNR (-5), the performance of ResNet18 appears slightly mediocre with the figure of around 63%, followed by SVM and RF with almost the same

success rate (60%) and KNN owns the best performance.

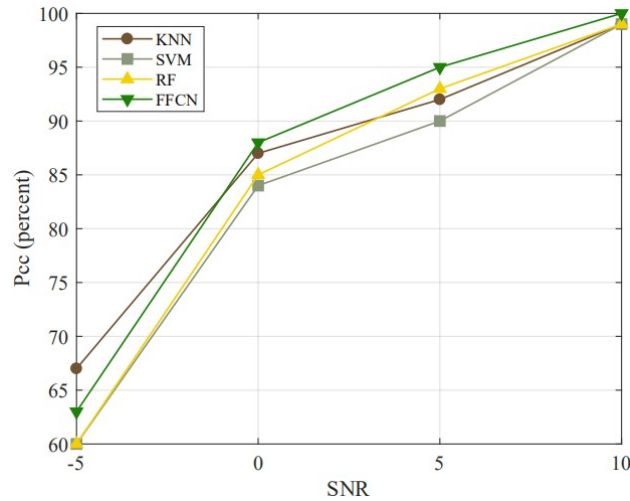


Fig. 7. The probability of correct classification for seven types of modulations versus SNR with different classifiers

5 CONCLUSIONS

This paper proposed a novel AMC model in SISO system using supervised learning and FB approach, which was based on the discrete heterogeneous data processed by both constellation map and cyclic spectrum density graph. In contrast to the traditional modulation classification method, we utilized the modified ResNet18 to optimize the unsatisfactory performance of the FB approach that accomplishes the prediction according to the original input matrices after normalization and combination. Due to the potent ability of ResNet18 in the aspect of classification, more discriminative features could be detected and the probability of correct identification could be improved. After extensive simulation, it had been proven that with the suitable operation between the data processed by two methods, the FFCN showed an outstanding performance compared to several popular classifier used today like KNN, SVM and RF, especially at low SNR.

6 ACKNOWLEDGEMENT

This work was supported in part by the National Key Research and Development Program of China under Grant (2018YFF0301202, 2019YFB1804400), in part by the National Natural Science Foundation of China under Grant (61801052), and in part by the Beijing Natural Science Foundation under Grants (4202046).

References

- [1] Y. Huo, X. Dong, W. Xu, and M. Yuen, A.: Cellular and WiFi co-design for 5G user equipment. 2018 IEEE 5G World Forum (5GWF), pp. 256–261 (2018)
- [2] A. G. Fragkiadakis, E. Z. Tragos, and I. G. Askoxylakis, A.: A survey of security threats and detection techniques in cognitive radio networks. *IEEE Communications Surveys Tutorials*, vol. 15, no. 1, pp. 428–445 (2013)
- [3] Y. Lin, X. Zhu, Z. Zheng, Z. Dou, and R. Zhou, A.: The individual identification method of wireless device based on dimensionality reduction and machine learning. *Journal of Supercomputing*, vol. 75, no. 6, pp. 3010 – 3027 (2019)
- [4] S. Huang, Y. Yao, Z. Wei, Z. Feng, and P. Zhang, A.: Automatic modulation classification of overlapped sources using multiple cumulants. *IEEE Transactions on Vehicular Technology*, vol. 66, no. 7, pp. 6089–6101 (2017)
- [5] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, A.: Survey of automatic modulation classification techniques: classical approaches and new trends. *IET Communications*, vol. 1, no. 2, pp. 137–156 (2007)
- [6] A. Swami and B. M. Sadler, “Hierarchical digital modulation classification using cumulants,” *IEEE Transactions on Communications*, vol. 48, no. 3, pp. 416–429 (2000)
- [7] P. Ghasemzadeh, S. Banerjee, M. Hempel, and H. Sharif, A.: Performance evaluation of feature-based automatic modulation classification. 12th International Conference on Signal Processing and Communication Systems (ICSPCS), Dec 2018, pp. 1–5 (2018)
- [8] C. Park, J. Choi, S. Nah, W. Jang, and D. Y. Kim, A.: Automatic modulation recognition of digital signals using wavelet features and SVM. 10th International Conference on Advanced Communication Technology, vol. 1, Feb 2008, pp. 387–390 (2018)
- [9] M. W. Aslam, Z. Zhu, and A. K. Nandi, “Automatic modulation classification using combination of genetic programming and KNN,” *IEEE Transactions on Wireless Communications*, vol. 11, no. 8, pp. 2742–2750 (2012)
- [10] B. Kim, J. Kim, H. Chae, D. Yoon, and J. W. Choi, A.: Deep neural network-based automatic modulation classification technique. International Conference on Information and Communication Technology Convergence (ICTC), Oct, pp. 579–582. (2016)
- [11] S. Huang, L. Chai, Z. Li, D. Zhang, Y. Yao, Y. Zhang, and Z. Feng, A.: Automatic modulation classification using compressive convolutional neural network. *IEEE Access*, vol. 7, pp. 79636–79643 (2019)
- [12] Y. Tu, Y. Lin, J. Wang, and J.-U. Kim, A.: Semi-supervised learning with generative adversarial networks on digital signal modulation classification. *Computers, Materials and Continua*, vol. 55, no. 2, pp. 243 – 254 (2018)
- [13] S. Huang, Y. Jiang, Y. Gao, Z. Feng, and P. Zhang, A.: Automatic modulation classification using contrastive fully convolutional network. *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1044–1047 (2019)
- [14] J. Ma and T. Qiu, A.: Automatic modulation classification using cyclic correntropy spectrum in impulsive noise,” *IEEE Wireless Communications Letters*, vol. 8, no. 2, pp. 440–443 (2019)
- [15] Targ, Sasha, Diogo Almeida, and Kevin Lyman, A.: Resnet in resnet: Generalizing residual architectures arXiv preprint arXiv:1603.08029 (2016).