Automatic Modulation Classification with Multi-domain Feature Fusion

Guangyang Li[1], Xiaofeng Wang[1], Mengting Jiang[1], Yun Chen[1], Hengliang Liu[2], and Daying Quan[1(✉)]

[1] School of Information Engineering, China Jiliang University, Hangzhou 310018, China
{xfwang,qdy}@cjlu.edu.cn
[2] Jptek Corporation Limited Hangzhou, Hangzhou 310018, China

Abstract. Automatic Modulation Classification (AMC) is a crucial technology that empowers communication systems to adapt to different signal environments, thus ensuring efficient signal detection and demodulation in a wide range of civil and military applications. The use of Convolutional Neural Networks (CNNs), with the emergence of deep learning techniques, has become prevalent in AMC. However, the signal features that are extracted from a single domain by the CNNs may be missing some of the essential characteristics of the raw signal. In this paper, a dual-channel model combining a deep residual network and an improved AlexNet has been proposed for AMC. The model makes use of features of the radar signals from both the time domain and the time-frequency domain to improve the detection performance. In the experiment, the model achieved a recognition accuracy of 91.70% on the RadioML 2016.10A dataset when the signal-to-noise ratio (SNR) was 2 dB, demonstrating its effectiveness in AMC. Experimental results demonstrate our proposed model outperforms the state-of-the-art methods.

Keywords: Automatic modulation classification (AMC) · deep residual network · AlexNet · dual channel model

1 Introduction

Automatic modulation classification (AMC) is an important technology deployed at the receiver to identify the modulation types of various received signals. In the era of electronic countermeasures, AMC is a key step between signal detection and signal demodulation, which plays a vital role in non-cooperative communication systems [1]. With the rapid development of radio technologies, the diversity of signal modulation types has increased, and channel interference has significantly intensified, which makes it difficult for traditional algorithms [2] to achieve accurate modulation classification. In recent years, deep learning has emerged as an effective solution for AMC, due to its outstanding performance in various classification tasks.

The methods of AMC mainly include likelihood-based (LB) and feature-based (FB) approaches. Due to the high complexity of LB approaches, FB methods have received more attention, and various deep-learning models are used for feature extraction in

W. Yu and L. Xuan (Eds.): MLICOM 2023, LNICST 518, pp. 15–26, 2024.
https://doi.org/10.1007/978-3-031-71716-1_2

AMC. O'Shea et al. [3] used a convolutional neural network (CNN) and ResNet to identify the signals of 24 modulation types, proving the feasibility of applying deep learning methods to AMC. Tu et al. [4] proposed to apply complex networks to AMC, verifying that complex networks have greater potential in AMC than real networks. Ke et al. [5] introduced a framework utilizing LSTM denoising self-encoder to acquire stable features, which achieves high recognition accuracy on a low-cost computing platform. To extract more effective information from the raw data, Peng et al. [6] first converted the original data into a constellation map and then performed AMC through AlexNet. Furthermore, Zhang et al. [2] proposed a time-frequency analysis combined with a deep learning method, where data preprocessing such as time-frequency analysis is performed on the original in-phase/quadrature (I/Q) data, and the preprocessed data is sent to a convolutional network with residual blocks for AMC. By preprocessing the original signal, the key features of the signal can be effectively enhanced, and the classification accuracy is better than the original signal directly input to the network. In addition, fusing different types of features can usually extract more feature information. Xu et al. [7] suggested a method to extract features from single and combined I/Q signals and used a network structure of one-dimensional (1D) convolution, two-dimensional (2D) convolution, and long short-term memory (LSTM) to extract features from time and space more efficiently. This approach shows that the feature fusion method has a higher recognition accuracy than the single feature method. Zhang et al. [8] employed a dual-stream structure based on CNN-LSTM to fuse the features extracted from I/Q and amplitude/phase (A/P). This increased the diversity of features, thereby improving performance. However, most of the existing methods only extract signal features from a single perspective, such as time domain or time-frequency domain, and few methods comprehensively consider features from both time domain and time-frequency domain to improve the accuracy of modulation classification.

In this paper, we propose to simultaneously utilize signal features from both time and time-frequency domains for AMC, and design a dual-channel model for feature extraction. In the model, the time domain channel first acquires shallow features of original signals by basic mathematical operations and then extracts deep features through a deep residual network. Moreover, the time-frequency domain channel utilizes Wigner-Ville Distribution (WVD) to obtain time-frequency information of the signals, and AlexNet is used for deep feature mining. Finally, the features extracted by the dual channel model are fused to further improve the classification accuracy of AMC.

2 Signal Model and Preprocessing Method

2.1 Signal Model

The typical signal model received by the receiver can be expressed as

$$y(t) = x(t) * h(t) + n(t) \tag{1}$$

where $x(t)$ is the modulated signal without noise, $*$ and $h(t)$ are convolution operation and channel impulse response respectively, and $n(t)$ symbolizes the additive white Gaussian noise (AWGN). Sample the received signal interval by T_s, and the sampled signal can

be expressed as

$$y_L = \left[s_0, s_1, ..., s_k, ..., s_{L-1}\right] \tag{2}$$

where s_k is the sampled signal, with a total of L data points. s_k is the data of the k-th sampling point. Since $s_k = I_k + jQ_k$ is a complex signal, it can be further expressed as

$$y_L = \begin{bmatrix} I_0, I_1, ..., I_k, ..., I_{L-1} \\ Q_0, Q_1, ..., Q_k, ..., Q_{L-1} \end{bmatrix} \tag{3}$$

where I_k and Q_k are the in-phase and quadrature components of the k-th sampling point, respectively. j is a complex number unit.

2.2 Time Domain Channel Preprocessing

Raw signal pre-processing is one of the key methods for improving the accuracy of AMC, as the limited number of sampling points prevents the network from effectively extracting signal features and classifying different modal types. Figure 1 depicts the process of time-domain channel preprocessing. For easy understanding, we describe the proposed signal preprocessing method for the time domain channel in Table 1. First, the key features of signal amplitude and phase are added, and the data length is changed from $L \times 2$ to $L \times 4$. Secondly, the signal is reversed and spliced at the end of the signal to achieve data enhancement of the signal, and the data length is changed from $L \times 4$ to $2L \times 4$. Finally, considering the network convolution process, when the receptive field of the convolution is large, it is easy to ignore the data relationship characteristics of adjacent points. Therefore, the horizontal gradient and one-dimensional difference calculation are respectively performed on the I/Q signal, amplitude, and phase, and the data length is changed from $2L \times 4$ to $2L \times 12$, which further enhances the effective information of the signal. In the time domain channel, an IQ signal with the original length $L \times 2$ is transformed into data with the length $2L \times 12$ after preprocessing, which enhances the effective information input into the network.

2.3 Time-Frequency Domain Channel Preprocessing

Due to the influence of white Gaussian noise, the features obtained in the time domain are limited. Currently, the acquisition of signal features is also mainly concentrated in the transform domain [9]. By employing time-frequency analysis of signals, the influence of Gaussian white noise can be effectively reduced. This enables the amplification of differences between various modulated signals and facilitates the extraction of more effective features. Moreover, using time-frequency analysis, one-dimensional time-domain signals can be transformed into two-dimensional time-frequency images, which facilitates subsequent deep feature extraction by CNN.

As a common time-frequency analysis method, WVD is widely used in the field of information processing [10]. By correlating two specific displacement signals of the signal $s(t)$, WVD can be expressed as:

$$WVD_s(t, \omega) = \int_{-\infty}^{+\infty} s\left(t + \frac{\tau}{2}\right) s^*\left(t - \frac{\tau}{2}\right) e^{-j\omega\tau} d\tau \tag{4}$$

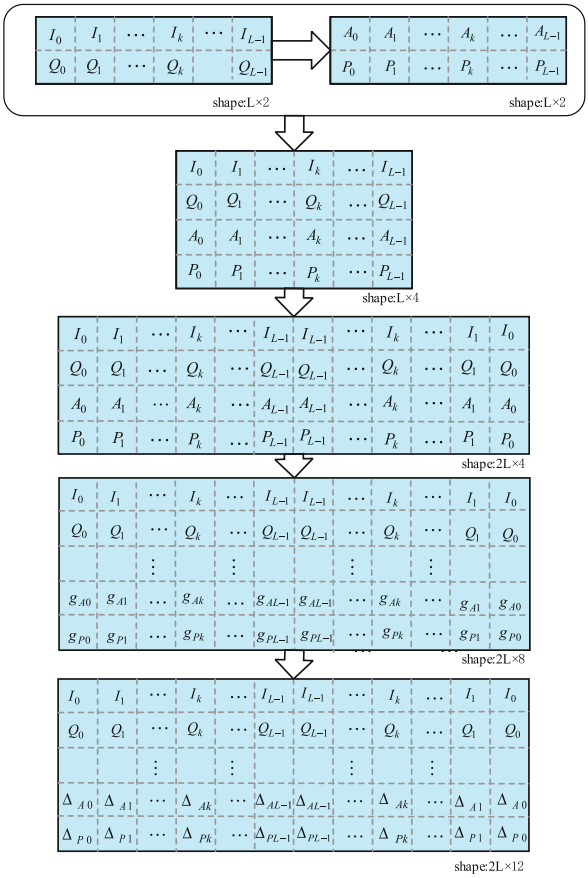

Fig. 1. Description of time domain channel preprocessing method

It can be seen that WVD is a Fourier transform of the instantaneous autocorrelation $s(t)$. As a global nonlinear distribution, WVD has no loss of resolution and has good time-frequency aggregation. Therefore, this paper selected WVD as the time-frequency transformation method for the time-frequency domain channel. Figure 2 shows the process of the WVD time-frequency transform.

To remove redundant information in time-frequency images and facilitate subsequent feature extraction, it is necessary to preprocess the WVD time-frequency images. Specifically, the time-frequency images obtained by WVD are first converted to grayscale to realize image enhancement. Secondly, the grayscale images are binarized by the OTSU algorithm [11], which is convenient for image feature discrimination. Then we perform a morphological opening operation on the binarized images to remove isolated burrs. Finally, the images are resized to 64 × 64 and normalized.

Table 1. Time domain channel preprocessing method

Preprocessing method	
1	Calculate the amplitude value $A_k = \sqrt{I_k^2 + Q_k^2}$ and the phase value $P_k = \tan^{-1}\left(\frac{I_k}{Q_k}\right)$ of each sample point s_k
2	Reverse signal, splice at the tail
3	According to the horizontal gradient formula $gradf = \frac{\partial f}{\partial x}$, calculate the horizontal gradient of IQ signal, amplitude and phase
4	According to the first-order difference formula $\Delta f(x_k) = f(x_{k+1}) - f(x_k)$, calculate the first-order difference of IQ signal, amplitude and phase
5	Splicing the calculated data to form data of length $2L \times 12$

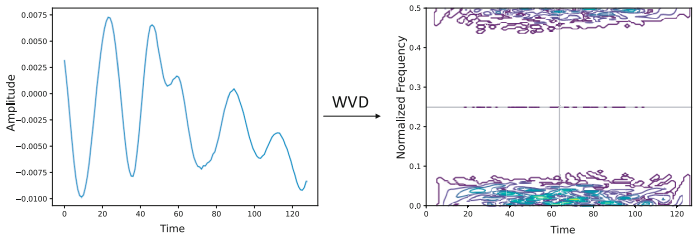

Fig. 2. WVD time-frequency transform

3 Structure of Model

3.1 Proposed Model

In this paper, we expand the feature perspective by performing shallow feature and deep feature extraction on the original signal from both time and time-frequency domains. To better extract the deep features of signals, we adopt a dual-channel architecture, which includes a time domain channel and a time-frequency domain channel. Since residual blocks can retain the integrity of the input information, a deep residual network is used in the time-domain channel to learn time-domain features. Moreover, due to the high recognition accuracy and fast convergence speed of AlexNet in image processing, AlexNet is used in the time-frequency domain channel to extract deep features from the processed time-frequency images. Finally, the features extracted from the two channels are fused and classified using the softmax classifier to achieve high-precision modulation recognition. Figure 3 shows the overall structure of the proposed model.

3.2 Time Domain Channel Network Architecture

The residual network is an enhanced version of the traditional CNN. It has been widely used in computer vision because of its powerful ability to extract deep features efficiently and achieve rapid convergence during training [12]. A basic residual module consists of

convolutional layers, batch normalization (BN) layers, activation functions, and identity mappings. Figure 4 shows the structure of a basic residual module.

In the foundational residual module, the convolution layer serves as the module's primary component and is responsible for extracting signal features. The BN layer's function is to convert the output into a standard normal distribution, prevent gradient explosion, and enhance the model's generalization capabilities. To address the issue of gradient vanishing, ReLU is used as the activation function in the residual model, which, to some extent, accelerates the model's convergence. After adopting the identity map, the network does not fit $H(x)$ directly but instead fits the residual map $F(x)$ through the stacked nonlinear layers, and the original map is reconstructed $F(x)+x$. By using identity mapping, the process of fitting becomes easier and it also allows for the introduction of the original information.

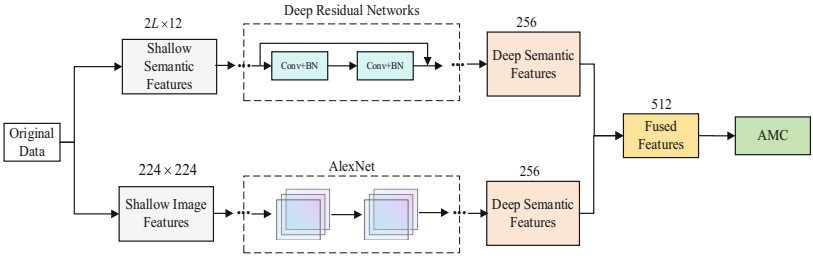

Fig. 3. Overall structure of the proposed model

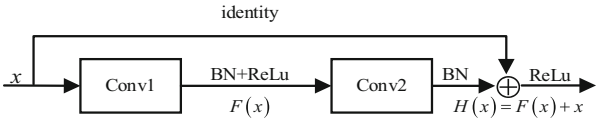

Fig. 4. Structure of a basic residual module.

A deep residual network is typically constructed by stacking multiple residual modules. Based on the deep residual network, we add convolution layers between the residual module and the residual module to further enhance the network's ability to extract distinguishable features. Figure 5 shows the network architecture of the time domain channel.

3.3 Time-Frequency Domain Channel Network Architecture

AlexNet is a classical deep neural network model, which has a good classification effect on image recognition [13]. The network mainly includes 5 convolutional layers, 3 max-pooling layers, and 3 fully connected layers. Here, we use it as the backbone network for the time-frequency channel. As mentioned above, we use a dual-channel model to fuse the features from this channel and the features from the time domain channel. Therefore, based on the original AlexNet, we appropriately adjust the number of convolution

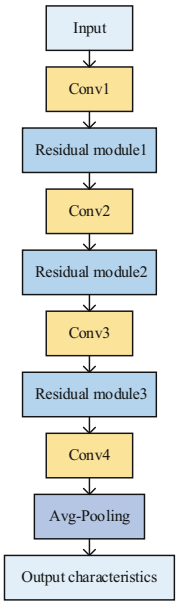

Fig. 5. Network architecture of time domain channel

kernels and the number of fully connected nodes in the network. Figure 6 shows the network architecture for the time-frequency domain channel. Major adjustments to AlexNet mainly include:

(1) Convert the 64 × 64 × 1 image to 67 × 67 × 1 by zero-padding;
(2) Use 64 convolution kernels with a size of 11 × 11, extracting signal features with a stride of 4, and use the ReLU activation function to make the value within a reasonable range;
(3) Use 3 * 3 maximum pooling layers with a step size of 2 to perform maximum pooling to reduce redundant information. The maximum pooling layer is connected after the first, second, and fifth convolutional layers respectively;
(4) The first fully connected layer uses 256 nodes, making the output feature size is 256;
(5) To further extract features and match the 256 features output by the semantic channel, there is another full connection with 256 nodes, and the final image channel output feature size is 256.

The deep features extracted from the time-frequency domain channel are fused with the deep features extracted from the time domain channel. Finally, AMC is realized with a softmax classifier on the fused features.

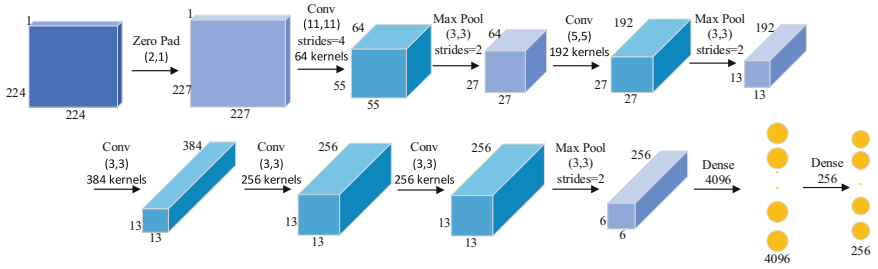

Fig. 6. Network architecture of time-frequency domain channel

4 Experiments

4.1 Dataset

In this paper, we have used the RadioML 2016.10A dataset as a benchmark to evaluate the proposed model. RadioML 2016.10A is generated by GNU Radio and contains 11 modulation type signals (8 digital, 3 analog) in the SNR range from −20 dB to 18 dB, including 8PSK, BPSK, CPFSK, GFSK, PAM4, 16QAM, 64QAM, QPSK, AM-DSB, AM-SSB, WBFM. Under the same SNR, each signal contains 1000 samples, and the data also contains random processes such as center frequency offset and sampling rate offset, so RadioML 2016.10A is widely used as a benchmark for AMC [14]. Here, we set a seed number to generate random sequences to randomize the RadioML 2016.10A dataset to generate the training and test sets.

As an option for model training, the dual-channel network selects the cross-entropy function as the loss function, the optimization function selects the Nadam function, the batch size is set to 32, and the epoch is set to 100. The initial learning rate is set to 0.002, and the learning rate is adjusted to 0.9 times the original value every 10 epochs. The performance is measured on a system equipped with a 3.50-GHz CPU, 32 GB RAM, and a single NVIDIA GeForce GTX 1050Ti GPU.

4.2 Performance Analysis

In the first experiment, we analyze the classification accuracy of 11 modulation types at different signal-to-noise ratios. Figure 7 shows the confusion matrix of 11 modulation types with SNR of −20 dB, −14 dB, −6 dB, and 18 dB, respectively. When the SNR drops to −20 dB, most of the signals are inaccurately labeled as AM-SSB, indicating that the network's classification ability is low at this point. At −14 dB SNR, the overall accuracy of recognition reaches 12.59%, with the most effective classification performance achieved for AM-DSB. With an SNR at −6 dB, the network achieves a classification accuracy of 57.52%. Notably, the classification outcomes for AM-SSB and PAM4 are particularly good, achieving accuracies of 98.45% and 94.85%, respectively. However, the classification outcome for 8PSK is still inadequate, measuring only 23.39%. At an SNR of 18 dB, the model's complete identification precision amounts to 92.95%. However, the analysis overall indicates a relatively weak classification accuracy of the model for WBFM. This could be attributed to the fact that the WBFM and AM-DSB signals

within the dataset were acquired from analog audio signals, which include silent periods that only contain carrier audio. Additionally, a misclassification between QAM64 and QAM16 was observed. The potential for confusion rises as adjacent modulation symbols in 16-QAM and 64-QAM have some similarity and noise disturbs the position of symbol points on the constellation diagram, resulting in shifts.

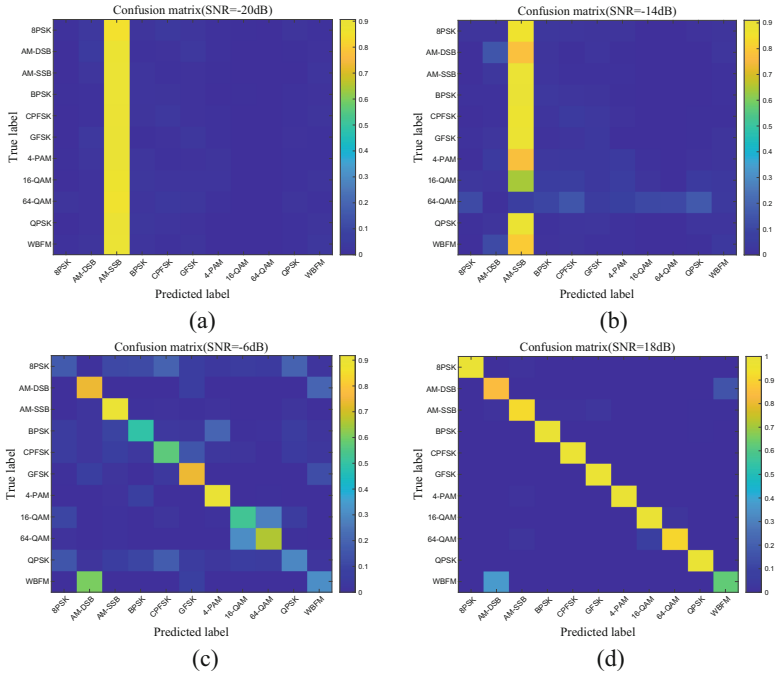

Fig. 7. Confusion matrix for 11 modulation types

4.3 Performance Comparison

In this experiment, we verify the superiority of the method proposed in this paper in terms of recognition accuracy. The difference in accuracy between the models on RadioML 2016.10A is shown in Fig. 8. We compare the proposed dual channel model with other approaches, including DAE [5], MCLDNN [7], CNN [15], MCNET [16], and IC-AMCNet [17]. When the SNR ranges from −20 dB to 18 dB, the proposed approach achieves the highest classification accuracy of 93.29% at 14 dB, while DAE, MCLDNN, CNN, MCNET and IC-AMCNet achieve classification accuracies of 84.31%, 86.63%, 89.45%, 84.32% and 83.34% respectively. At an SNR of 2 dB, the proposed dual-channel network produces a classification accuracy of 91.70%, which fluctuates around 92% when the SNR exceeds 2 dB. Meanwhile, for SNR values above −4 dB, the proposed dual channel model shows good performance compared to the other comparative models.

Among these comparative models, MCLDNN and CNN are multi-channel models. However, CNN only extracts the time domain features and fails to fully capture the frequency characteristics of the signal. In contrast, the dual channel model proposed combines time domain and time-frequency domain features. It provides more comprehensive information for the model and helps to more accurately identify various communication signal types. In MCLDNN, although time domain and time-frequency domain features are fused, no additional preprocessing is performed on time domain features. In contrast, in this paper, the time domain feature is further enriched by preprocessing. Such a processing strategy enables the model to better understand the time-domain features, resulting in a more significant performance improvement for signal recognition.

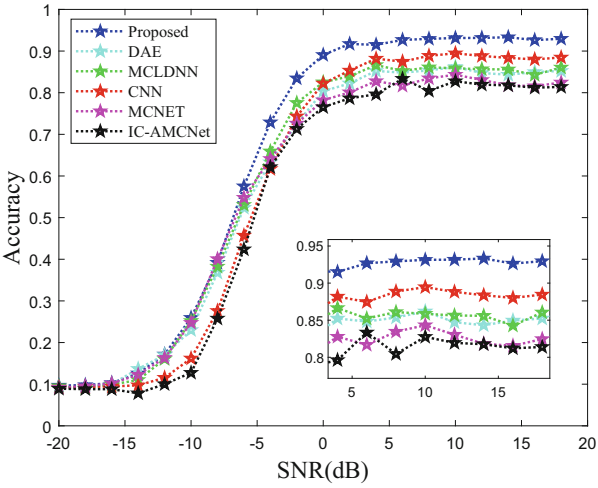

Fig. 8. Accuracy of different models on RadioML 2016.10A

4.4 Ablation Experiment

In the ablation experiment, we demonstrate that the improvement in recognition accuracy is due to the introduction of the two-channel model. As shown in Fig. 9, we compare the proposed dual channel model with the time domain channel and the time-frequency domain channel. In the figure, the highest recognition accuracy is achieved by our proposed dual-channel model, which suggests that the time domain and time-frequency domain features provide complementary information and that their combination results in a superior model. The recognition accuracy of the time-frequency domain channel is significantly lower, illustrating the importance of time-domain features.

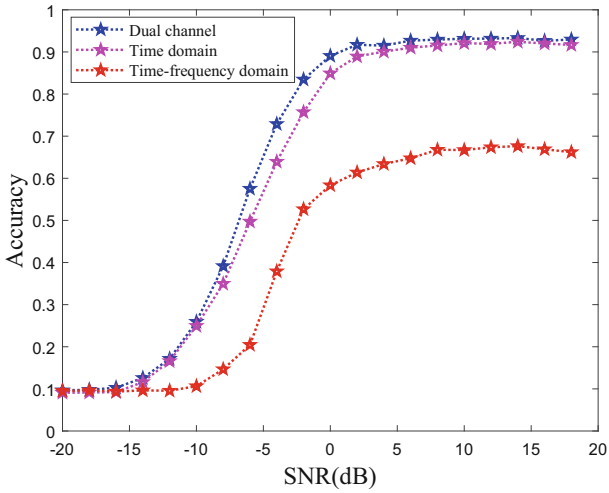

Fig. 9. Accuracy of dual channel model and single channels

5 Conclusion

This paper proposed a dual-channel network model that leverages time-domain and time-frequency domain features to achieve efficient classification of radar signal modulation. It achieves an overall classification accuracy of more than 91.70% under the condition of SNR = 2 dB, which provides better recognition performance compared to the other four algorithms. However, due to the utilization of a dual deep neural network, the proposed model still incurs relatively high computational costs. Fortunately, the emergence of deep learning pruning technology makes it possible to reduce the computational load of the algorithm. In our future work, we will focus on developing a pruning algorithm for this model.

Acknowledgment. This research was supported by Key Research and Development Projects in Zhejiang Province (No. 2022C01144).

References

1. Dobre, O., Abdi, A., Bar-Ness, Y., et al.: Survey of automatic modulation classification techniques: classical approaches and new trends. IEEE Commun. **1**(2), 137–156 (2007)
2. Zhang, H., Huang, M., Yang, J., et al.: A data preprocessing method for automatic modulation classification based on CNN. IEEE Commun. Lett. **25**(4), 1206–1210 (2021)
3. O'Shea, T., Roy, T., Clancy, T.: Over-the-air deep learning based radio signal classification. IEEE J. Sel. Top. Sig. Process. **12**(1), 168–179 (2018)
4. Tu, Y., Lin, Y., Hou, C., et al.: Complex-valued networks for automatic modulation classification. IEEE Trans. Veh. Technol. **69**(9), 10085–10089 (2020)
5. Ke, Z., Vikalo, H.: Real-time radio technology and modulation classification via an LSTM auto-encoder. IEEE Trans. Wireless Commun. **21**(1), 370–382 (2022)

6. Peng, S., et al.: Modulation classification based on signal constellation diagrams and deep learning. IEEE Trans. Neural Netw. Learn. Syst. **30**(3), 718–727 (2019)
7. Xu, J., Luo, C., Parr, G., et al.: A spatiotemporal multi-channel learning framework for automatic modulation recognition. IEEE Wireless Commun. Lett. **9**(10), 1629–1632 (2020)
8. Zhang, Z., Luo, H., Wang, C., et al.: Automatic modulation classification using CNN-LSTM based dual-stream structure. IEEE Trans. Veh. Technol. **69**(11), 13521–13531 (2020)
9. Alaskar, H.: Deep learning-based model architecture for time-frequency images analysis. Int. J. Adv. Comput. Sci. Appl. **9**(12), 486–494 (2018)
10. Wang, J., Mo, Z., Zhang, H., et al.: A deep learning method for bearing fault diagnosis based on time-frequency image. IEEE Access **7**, 42373–42383 (2019)
11. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. **9**(1), 62–66 (1979)
12. Duta, I., Liu, L., Zhu, F., et al.: Improved residual networks for image and video recognition. In: 25th International Conference on Pattern Recognition, Milan, Italy, pp. 9415–9422 (2021)
13. Rawat, W., Wang, Z.: Deep convolutional neural networks for image classification: a comprehensive review. Neural Comput. **29**(9), 2352–2449 (2017)
14. O'Shea, T., Corgan, J., Clancy, T.: Convolutional radio modulation recognition networks. In: Jayne, C., Iliadis, L. (eds.) EANN 2016. CCIS, vol. 629, pp. 213–226. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-44188-7_16
15. Zhang, Z., Wang, C., Gan, C., et al.: Automatic modulation classification using convolutional neural network with features fusion of SPWVD and BJD. IEEE Trans. Sig. Inf. Process. Over Netw. **5**(3), 469–478 (2019)
16. Huynh-The, T., Hua, C., Pham, Q., et al.: MCNet: an efficient CNN architecture for robust automatic modulation classification. IEEE Commun. Lett. **24**(4), 811–815 (2020)
17. Hermawan, A., Ginanjar, R., Kim, D., et al.: CNN-based automatic modulation classification for beyond 5G communications. IEEE Commun. Lett. **24**(5), 1038–1041 (2020)