Automatic Modulation Classification Based on Multimodal Coordinated Integration Architecture and Feature Fusion

Xiao Zhang and Yun Lin(✉)

College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China
{zxiao,linyun}@hrbeu.edu.cn

Abstract. With the rapid advancement of the 5G wireless communication technology, automatic modulation classification (AMC) is not only faced with more complex communication environment, but also needs to deal with more modulation styles, which increases the difficulty of modulation recognition invisibly. However, most deep learning (DL)-based AMC approaches currently merely use time domain or frequency domain monomodal information and ignore the complementarities between multimodal information. To address the issue, we exploit a signal statistical graph domain-I/Q waveform domain multimodal fusion (SIMF) method to achieve AMC based on AlexNet, complex-valued networks and multimodal technology. The extracted multimodal features from signal statistical graph domain and I/Q waveform domain are fused to obtain a richer joint feature representation and T-SNE algorithm is used to visualize the extracted feature. Moreover, coordinated integration architecture was adopted to achieve mutual collaboration and constraints between multiple modalities to maintaining the unique characteristics and exclusivity of each modal. Simulation results demonstrate the superior performance of our proposed SIMF method compared with unimodal model and feature fusion model.

Keywords: Automatic modulation classification · 5G wireless communications · Multimodal deep learning · Feature fusion

1 Introduction

Recently, from 2020 onwards, 5G wireless communication networks will be introduced worldwide and entered the commercial development stage [1]. Some countries have launched research and exploration on 6th generation mobile networks (6G). 6G wireless communication networks are supposed to provide global reach, improved spectral/energy/cost performance, increased information, and improved protection [2, 3]. In order to meet the standards of wireless communication technology, as an important technology that cannot be ignored in non-cooperative communication, AMC technology plays an important role. However, AMC not only faces an increasingly complex communication environment, but also needs to deal with more modulation patterns, which

J. Xiong et al. (Eds.): MobiMedia 2021, LNICST 394, pp. 782–795, 2021.
https://doi.org/10.1007/978-3-030-89814-4_58

invisibly increases the difficulty of modulation recognition. Therefore, it is important and urgent to explore a more efficient AMC method.

DL has sparked widespread interest as a major achievement in artificial intelligence. Recently, DL has also become a research hotspot in the field of AMC [4–10], and provides a new way to improve recognition performance by virtue of multiple advantages such as pattern recognition and feature expression. The basic idea of the AMC based on the image recognition network model is to transform the signal recognition problem into an image recognition problem. Peng *et al*. [11] convert signals into three-channel images and studies DL models for AMC. In the sluggish and flat fading channels, G. Jajoo *et al*. [12] proposed a novel method focused on constellation structure to identify PSK and QAM modulation of different orders. Lin et al. [13] suggested the contour stellar image (CSI) approach for transforming signal waveforms into statistically significant pictures, which can migrate deep statistical information from the initial wireless signal waveforms.

The sum of information encoded in the phase of a signal is adequate to retrieve the remainder of the information encoded in its magnitude. Currently, most of the researchers interpret I/Q data with an ordered pair of real-valued numbers, and numerous architectures have been developed for this data format. [14–16]. Cheng *et al*. [17] proved that the complex-valued networks has certain advantages over the real-valued networks in the AMC, because it has a richer representation ability and better generalization characteristics. Tu *et al*. [18] proposed to apply complex-valued networks for AMC and it is proved that the performance of complex-valued networks in AMC is better than that of real-valued networks.

But as presented above, the majority of current DL-based approaches for AMC use monomodal information from single-dimensional domain [19, 20]. The general shortcoming of these AMC methods is that they do not make reasonable use of the existing multimodal information and ignore the complementarity among them. To address this problem, multimodal technology [9, 21, 22] has been applied to the field of AMC. Zhang et al. [23] proposed a multi-modality fusion model and attempts to integrate different modality features for the first time to improve performance in the area of AMC. Wu et al. [24] propose a CNN-based AMC method of multi-feature fusion that converts signals into cyclic spectra (CS) and constellation diagram (CD) of image representations. To realize AMC based on deep residual networks, Qi et al. [25] exploit a waveform-spectrum multimodal fusion method.

In this paper, we proposed a signal statistical graph domain-I/Q waveform domain multimodal fusion (SIMF) method to achieve AMC based on AlexNet, complex-valued networks and multimodal technology. This paper focuses on the use of multimodal information technology in AMC. The following is a list of the paper's key contributions.

1. We extract multimodal information from the original modulated signal dataset. Specifically, the first modality is the CSI of the signal statistical graph domain, and the second modality is the signal I/Q waveform domain. The two modalities are fused by multimodal fusion technology. We adopted the T-SNE algorithm to visualize the extracted features.

2. We adopt coordinated integration architecture to achieve mutual collaboration and constraints between multiple modalities to maintaining the unique characteristics and exclusivity of each modal.
3. The modulated signal datasets of different sampling points are established to evaluate our model performance. The experimental results prove that the classification accuracy of our proposed SIMF method is better than other methods.

The following is the structure of this paper: Signal modal and data preprocessing are covered in Sect. 2. We introduce our scheme in detail in Sect. 3. Section 4 shows the simulation results. Finally, the conclusions are drawn in Sect. 5.

2 Signal Modal and Preprocessing

2.1 Signal Model

In a communication system, the general equation of the signal received by the receiver can be expressed as

$$r(t) = h(t) * s(t) + n(t) \tag{1}$$

where $n(t)$ represents the additive white Gaussian noise (AWGN) with zero mean, $h(t)$ denotes the channel impulse response, $s(t)$ is modulated signals, and $r(t)$ is the received signals, $*$ is the convolution operation. The received signal $r(t)$ is usually transformed into a discrete version $r[n]$ and represented by I/Q format data. It is comprised of the in-phase component r_I and the quadrature component r_Q. Thus, the discrete signal $r[n]$ can be described

$$r[n] = r_I[n] + \widetilde{j} r_Q[n] \tag{2}$$

2.2 Signal Preprocessing

Most of the AMC methods are based on monomodal information from single depicting dimension and ignore the complementarities among the multimodal information of electromagnetic signal. In this article, we will study two modal representation approach of electromagnetic signal I/Q waveform domain and statistical graph domain. By taking advantage of the correlation between signal multimodal data and eliminating the redundancy among multimodal data we will learn a more accurate representation method of data features.

Signal I/Q Waveform Domain. The first modality is the I/Q vector which is obtained by the imaginary part and real part of the received signals. The i-th sample of signal I/Q vector is defined as

$$x_i^{I/Q} = \begin{bmatrix} x_i^I \\ x_i^Q \end{bmatrix} \tag{3}$$

where $x_i^I = \mathrm{Re}[r[n]]$, $x_i^Q = \mathrm{Im}[r[n]]$.

Signal Statistical Graph Domain. In this paper, we transform the original signal data into CSI as the second modality. The transformation method is as follows. In view of the original signal CD in different regions have different characteristics of the sample point density, density of using rectangular window function in the CD sliding on the graph, the statistics window in different regions of points, number of sampling points divided by the whole CD figure get normalized dot density value, the final size normalized dot density values mapped to different color, make the original signal CD figure into new CSI. The CSI estimation method can be outlined as follows (Fig. 1):

$$\rho(i,j) = \frac{\sum_{i=x_1}^{x_2} \sum_{j=y_1}^{y_2} \text{dots}(i,j)}{\sum_{x_1=W_0}^{W_1} \sum_{y_1=H_0}^{H_1} \sum_{i=x_1}^{x_2} \sum_{j=y_1}^{y_2} \text{dots}(i,j)} \tag{4}$$

Where W_0, H_0 are top left coordinate of CSI; W_1, H_1 are the lower right corner of CSI; x_0, y_0 are the upper left corner coordinates of the density window function; x_1 and y_1 the lower right corner coordinates of the density window function. Figure 2 shows the CSI of the twelve signals at 8 dB.

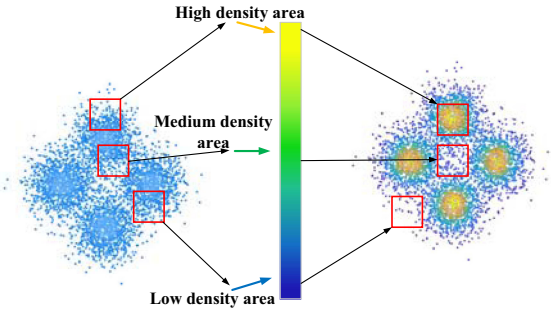


Fig. 1. CD converted to CSI

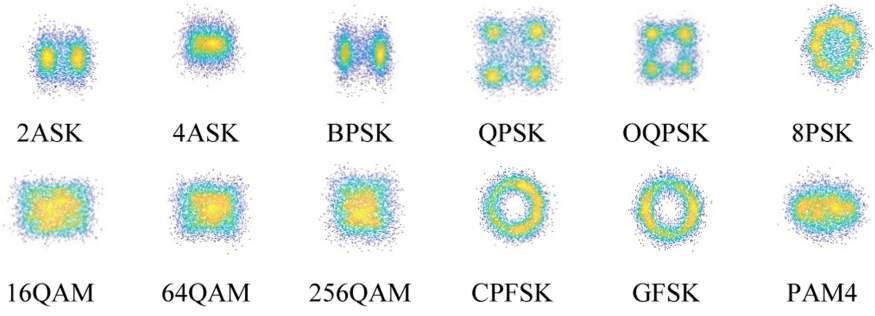


Fig. 2. Contour stellar images of twelve signals at 8 dB

CD as a binary image, it does not distinguish among pixels at a single sampling point and pixels at multiple sampling points. By contrast, colors and shapes of CSI can give us more specific information about the wireless signal they represent and provide finer grained features. It can retain the statistical characteristics of the signal even under the interference of noise.

3 The Proposed Modulation Classification Method

In this section, we'll go through the prototype and of the AlexNet networks as signal statistical graph domain features extraction and Complex-valued networks as I/Q waveform domain features extraction, then introduce the SIMF method we proposed.

3.1 Signal Statistical Graph Domain Features Extraction

Recently, DL has gradually become a hot topic in the teaching field of various subjects. Alex Krizhevesky et al. proposed a deeper, broader CNN model and won the most difficult visual object recognition challenge in the 2012 ImageNet Large Visual Recognition Challenge (ILSVRC). Therefore, AlexNet is employed as feature extractor to extract the acquirement of the powerful features.

The architecture of AlexNet is shown in Fig. 3. The first convolution layer uses Local Response Normalization (LRN) to perform convolution and maximum pooling, where 96 different acceptance filters of size 11 × 11 are used. The maximum pool operation is performed using a 3 × 3 filter with a step size of 2. The same action is performed in the second layer using 5 × 5 filters. 384, 384 and 296 feature maps were used in the third, fourth and fifth convolutional layers of 3 × 3. Finally, there are two fully connected layers (FC) and a softmax layer.

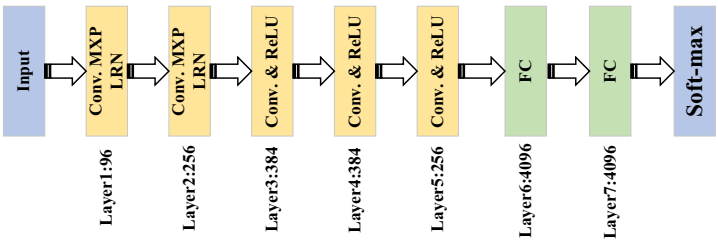


Fig. 3. Architecture of AlexNet

Transfer learning is adopted in our framework because it could accelerate and optimize the learning efficiency of models. We keep the parameters of the original AlexNet model and only set the number of neurons in the last two FC to 1024 to fit the size of our data set. We extract the deep features of CSI with AlexNet based on transfer learning. Figure 4 is the feature extraction process by AlexNet. In our work, the input of AlexNet networks are RGB image sets of CSI with size of 227 × 227 × 3 and the output of the last FC layers are 1024 deep CSI features we extracted.

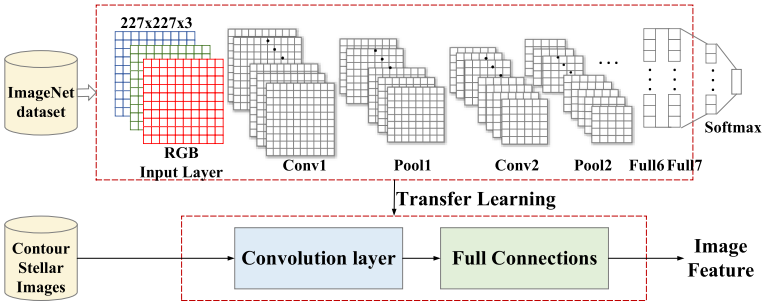


Fig. 4. The transfer process of the AlexNet

3.2 Signal I/Q Waveform Domain Features Extractions

Current mainstream deep learning technologies and architectures are all based on real number manipulation and representation, while some recent basic theoretical analysis shows that complex numbers have more expressive capabilities. In addition, electromagnetic signals mostly appear in the form of I/Q complex, and the phase information of electromagnetic signals can also represent the time-domain characteristics of electromagnetic signals. Therefore, the network with the expression and processing of complex forms can help to improve the electromagnetic signal recognition rate to some extent. Thus, we use complex-valued networks as signal I/Q waveform domain features extractions. Next, we will present the implementation of the complex-valued building blocks of the complex neural network.

Complex Convolutional Layer. In order to perform the equivalent operation equivalent to traditional real 2D convolution in complex number domain, in this scheme, the complex filter matrix $W = A + iB$ is convolved by the complex vector $\mathbf{h} = \mathbf{x} + i\mathbf{y}$. By using real numbers to simulate the operation of complex numbers, so that x and y are real matrices. Since the convolution operator is distributed, the vector h is convolved through the filter W, and the following can be obtained:

$$\mathbf{W} * \mathbf{h} = (\mathbf{A} * \mathbf{x} - \mathbf{B} * \mathbf{y}) + i(\mathbf{B} * \mathbf{x} + \mathbf{A} * \mathbf{y}) \tag{5}$$

Complex Batch Normalization. In deep learning models, batch normalization can accelerate the convergence speed of models and alleviate the problem of "gradient dispersion" in deep networks. The complex number normalization is specifically designed according to the complex number network, and its formula is as follows:

$$BN(\dot{x}) = \gamma \cdot \dot{x} + \beta \tag{6}$$

Where γ is variance, β is shift factor.

Complex Dense Layer. FC is used as a classifier in deep neural network. In order to make full use of the advantages of complex-valued networks, complex-FC is used at the end of complex-valued networks. Similar to the complex convolution kernel, the complex FC also carries out I/Q two-channel association information mining of electromagnetic

data through the alternating product of the real part imaginary part of the weight of the complex full connection layer and the real part imaginary part of the signal. Let $W = A + iB$ represents complex dense vector weight, and $\vec{s} = x + iy$ represents complex-valued input. Similar to the complex convolutional operation, we can get:

$$\mathbf{W} \cdot \vec{s} = (\mathbf{A} \cdot x - \mathbf{B} \cdot y) + i(\mathbf{B} \cdot x + \mathbf{A} \cdot y) \tag{7}$$

In our work, the overall structure of the complex-valued network we used as features extractor is shown in Fig. 5. We use the output of the penultimate FC layer as the deep features of the signal I/Q waveform domain.

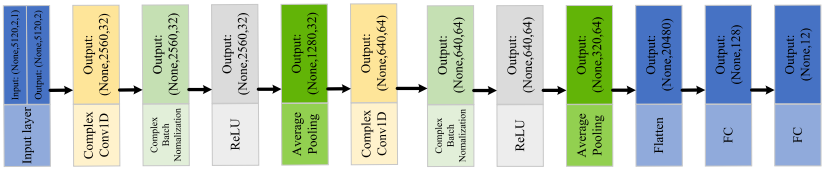


Fig. 5. The structure of Complex-valued model

3.3 The Proposed SIMF Method

In order to fuse the theories and methods of AMC, deep learning, complex-valued networks and multimodal deep learning, we proposed a framework based on the multimodal information fusion of signal statistical graph domain and I/Q waveform domain to achieve AMC. The fusion of multimodal information makes use of the complementarity of different modal information to realize the fusion of multi domain information to obtain more comprehensive and easily distinguishable features. Therefore, the final fusion feature representation is more discriminative than the single-mode representation, which can improve the classification accuracy and robustness of the model. Moreover, in order to make the separate single mode and the feature fusion multi-mode maintain the cooperative classification goal, we adopt the coordinated integration architecture to maximize the similarity of different models while maintaining the independent operation of each model.

Our proposed SIMF framework is shown in Fig. 6. Specifically, first, we convert the original signal into CSI. Then, AlexNet network is selected as the feature extractor of the signal statistical graph domain, and at the same time, the high-level features of the electromagnetic signal I/Q data are extracted using a complex-valued networks. The features of the two modals are fused by series splicing method. At last, the fusion features are input into the FC to map the classification results, thereby obtaining the multimodal model. Finally, the use of coordinated integration architecture to achieve mutual collaboration and constraints between multiple modalities is conducive to maintaining the unique characteristics and exclusivity of each modal. It takes into account the problems of modal coordination and feature fusion at the same time. Here, the fusion target of the two unimodal models of the signal statistical graph domain model and I /Q waveform

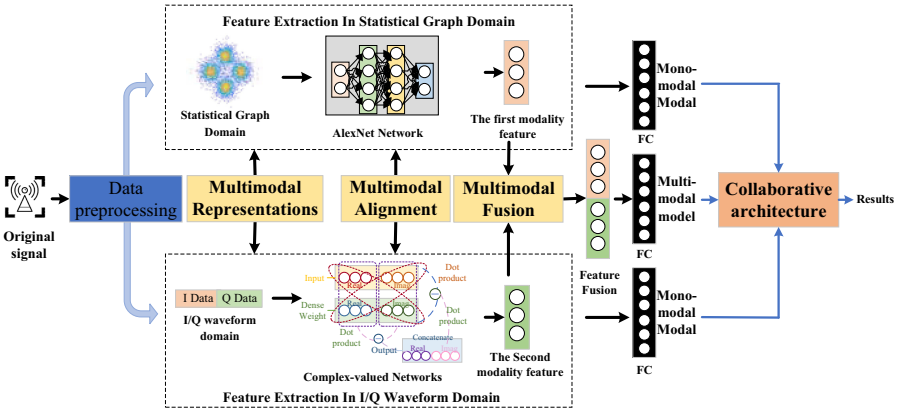


Fig. 6. Architecture of proposed SIMF model

domain and the feature fusion multimodal model are expressed by the same loss function, as shown in formula (9). In this process, the alignment of multimodal is ensured. In this way, the similarity structure between and within the modals is maintained, and at the same time the goal of mutual cooperation between the modals is achieved.

Next, we shall discuss the method of coordinated integration architecture in detail. Since the input of the model contains two modalities of the statistical graph domain and the I/Q waveform domain of same original signal data source, the penalty of the two predicted label distributions of two unimodal models must be taken into consideration. We use cross entropy to calculate the penalty of different modalities of predicted label distributions. Let $p_\theta(x_i)$ and $q_\theta(x_i)$ represent the output probability distributions of two unimodal models. The following is a description of cross entropy:

$$\text{Cross Entropy } = -\sum_{i=1}^{n} p_\theta(x_i)\ln(q_\theta(x_i)) \tag{8}$$

Let x_i^m (for $m \in \{1, 2\}$), $i \in \{1, 2, ...N\}$ denote the m-th modal information of the i-th example. x_i^c is the fusion features from the two unimodal modalities of the i-th example. $\Theta=\{\theta^c, \theta^m\}$ are obtained by training. N represents the number of training samples. In this way, the loss function of the feature fusion multimodal model and the two unimodal models can be represented by the same fusion target as follows:

$$LossFuction = -\frac{1}{N}\sum_{i=1}^{N} t_i \ln\left(p_{\theta^c}\left(x_i^c\right)\right) - \frac{1}{N}\sum_{m=1}^{2}\sum_{i=1}^{N} p_{\theta^c}(x_i^c)\ln\left(p_{\theta^m}\left(x_i^m\right)\right) \tag{9}$$

where $p_\theta(x_i)$ is probability distribution and is obtained by softmax function that is defined as:

$$p_\theta(x_i) = \text{softmax}(x_i) = \frac{1}{\sum_{k=1}^{K} e^{\theta_k^T x_i}}\left[e^{\theta_1^T x_i}, e^{\theta_2^T x_i}, \ldots, e^{\theta_k^T x_i}\right]^T \tag{10}$$

where K denotes the number of all classes, t_i is the true label probability distribution.

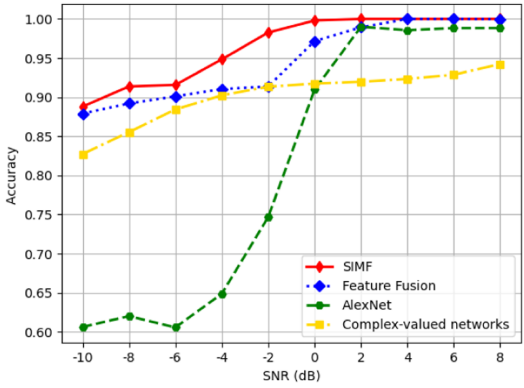
4 Experiments

4.1 Experimental Dataset

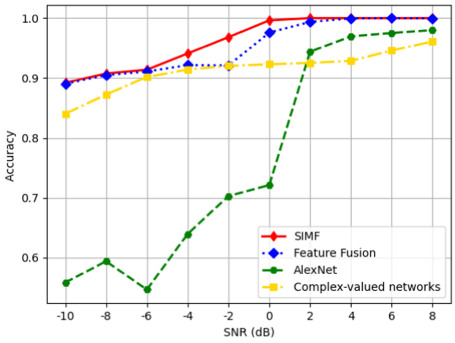
BPSK, QPSK, 16QAM, 64QAM, and 256QAM are the modulation types used in the traffic channel of 5G mobile communications, according to open protocol and specifications. Looking forward to dynamic spectrum access in 5G mobile communications, digital modulation recognition has more practical significance. Hence, in this paper, we use MATLAB to generate 12 digital modulated signals, including 2ASK, 4ASK, BPSK, QPSK, OPQSK, 8PSK, 16QAM, 64QAM, 246QAM, CPFSK, GFSK, PAM4. The noise environment considered in the experiment is additive white Gaussian noise (AWGN) with SNR from −10 dB to 8 dB and a stride of 2 dB. In this experiment, under each SNR, 1250 samples are generated for each modulated signal, among them are 1000 training samples and 250 test samples.

4.2 Results and Discussions

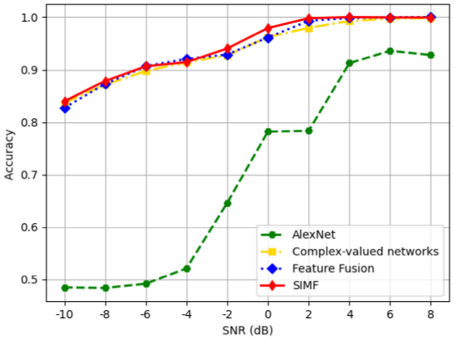
Figure 7 presents average accuracy of the proposed SIMF method, feature fusion method, AlexNet based unimodal model method and complex-valued networks based unimodal



(a) Signal sampling points = 5120



(b) Signal sampling points = 3072



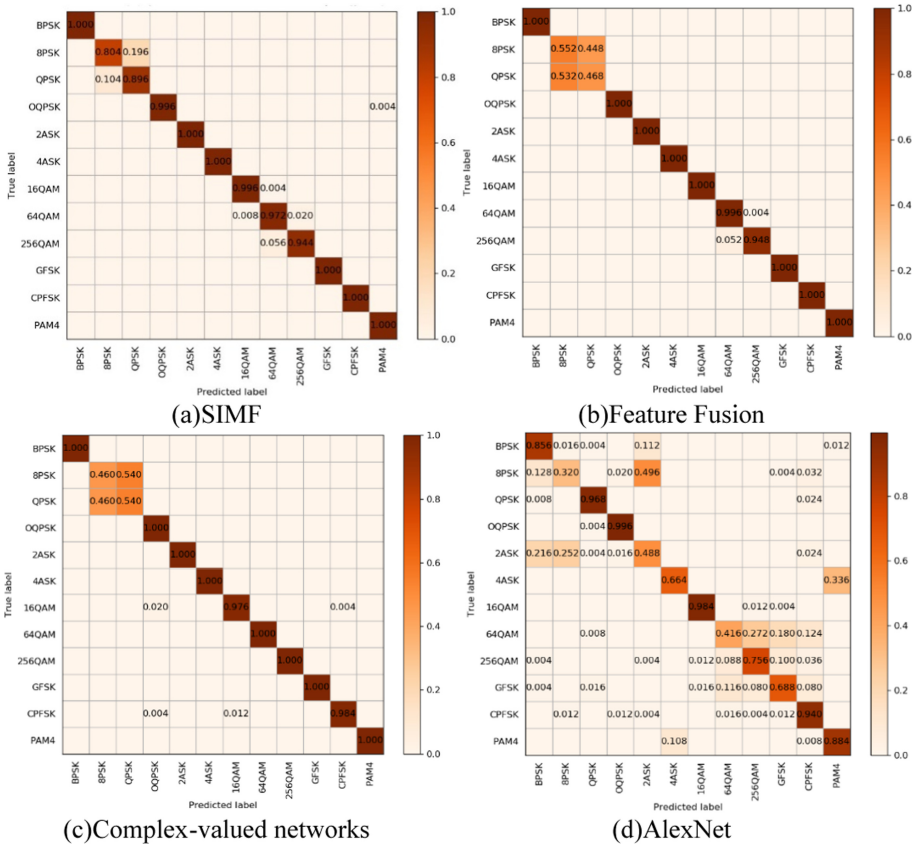
(c) Signal sampling points = 1024

Fig. 7. Average classification accuracy of different methods versus SNR under different signal sampling points.

model method versus signal to noise ratio (SNR) under different signal sampling points. In the experimental simulation, we set the signal sampling points as 5120, 3072 and 1024 respectively to explore the influence of different signal sampling points on the performance of different methods.

From the figures, the average classification accuracy of the SIMF method proposed in this paper is significantly better than the other three methods under the overall SNR for different signal sampling points, especially when the SNR is greater than 0 dB, the accuracy reaches 100%. Moreover, the average accuracy of the multimodal model with feature fusion is significantly higher than that of the other two unimodal models. This indicates that the multimodal fusion method can realize the complementarity among the multimodal information, so as to obtain a more comprehensive joint feature to improve the accuracy and model robustness. On the other hand, as the number of signal sampling points decreases, we can see that the AlexNet based unimodal model is greatly affected by it. The smaller the number of signal sampling points, the lower the average accuracy of the AlexNet based unimodal model. In addition, the performance of AlexNet based unimodal model is greatly affected by SNR, because its average accuracy is significantly reduced with the decrease of SNR. Under the condition of large number of signal sampling points and high SNR, the performance of AlexNet based unimodal model is better than that of complex-valued networks unimodal model. In addition, the signal sampling number has little effect on the complex-valued networks unimodal model, and better classification accuracy can be achieved under low SNR. Because the multimodal model can combine the advantages of two unimodal models, the accuracy of the multimodal model in the overall SNR is significantly improved. When the number of signal sampling points is 1024, the performance of the proposed SIMF method and the multimodal model is not significantly improved due to the limitations of the AlexNet based unimodal model.

Figure 8 shows the confusion matrixes of different AMC methods with SNR being −2 dB. It can be seen from Fig. 8(d) that the main errors of the AlexNet unimodal model occur between 8PSK and 2ASK, 4ASK and PAM4, as well as among 16QAM, 64QAM and 256QAM. This can be explained by the CSI diagrams of the dataset, since their CSI diagrams are similar. From Fig. 8(c), it can be seen that the main error of the complex-valued network unimodal model occurs between 8PSK and QPSK. In comparison, because the feature fusion multimodal model can achieve the complementary advantages of the two unimodal models, the probability of correct classification is significantly improved. Most importantly, we can see from Fig. 8 (a) that our proposed SIMF method has an obvious advantage in classification correctness probability compared with the other three models due to its combination of feature fusion and coordinated integration architecture.



(a)SIMF (b)Feature Fusion

(c)Complex-valued networks (d)AlexNet

Fig. 8. Confusion matrix of different methods under SNR $= -2$ dB

To better understand the classification performance of each kind of signal under different methods, we analyze the correct classification probability of twelve modulation types differs with SNR in Fig. 9. From the Fig. 9 we can see that the AlexNet unimodal model has some error in the correct classification of each modulated signal when the SNR is lower than 2 dB. This is because the CSI of the signals is greatly affected by the SNR. However, the classification accuracy of the complex-valued networks unimodal model for 8PSK and QPSK signals is always low, because complex-valued networks are unable to acquire enough features to differentiate these categories when only I/Q element is used as input. In comparison, our proposed SIMF model can achieve 100% accuracy when the SNR is greater than 0 dB, thanks to the combination of feature fusion and coordinated integration architecture.

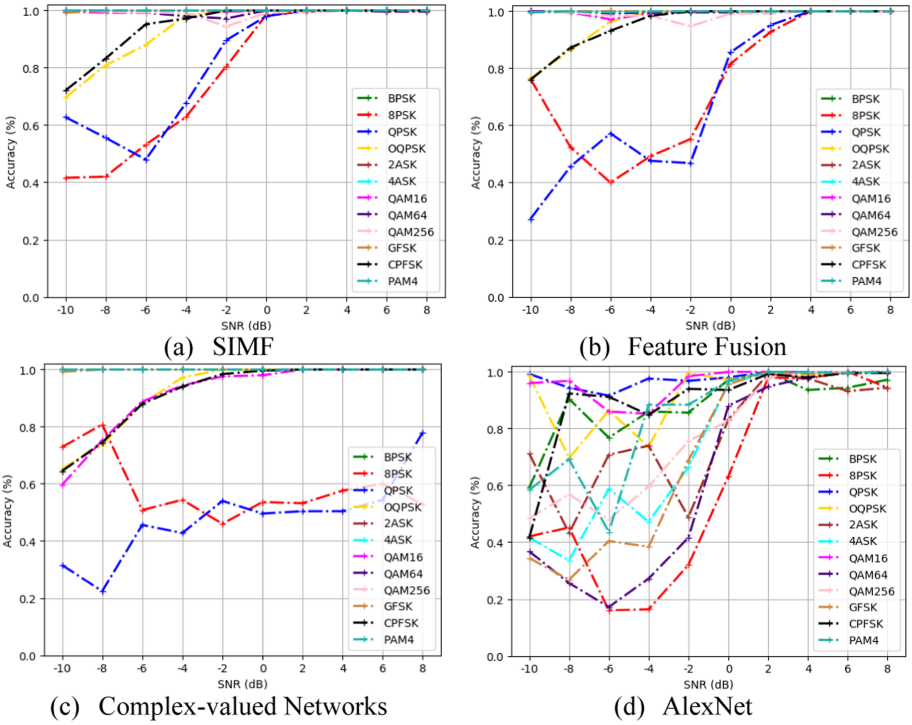


Fig. 9. Correct classification probability of different methods versus SNR

Figure 10 is a visualization of features extracted by different methods at SNR = 6 dB. The specific method is to extract the output of the penultimate FC of the networks, and use the T-SEN [26] method for dimensionality reduction and visual display. Each color represents a signal modulation type. It can be seen that the characteristic aliasing of each signal in the AlexNet unimodal model is serious and the distance between classes is short. However, both the complex-valued networks unimodal model and the feature fusion multimodal model have the aliasing of the two signal features (the signal features represented by orange and yellow). By comparison, the class spacing of each signal feature in the proposed SIMF method is relatively large and easy to distinguish, which reflects the advantages of the proposed method.

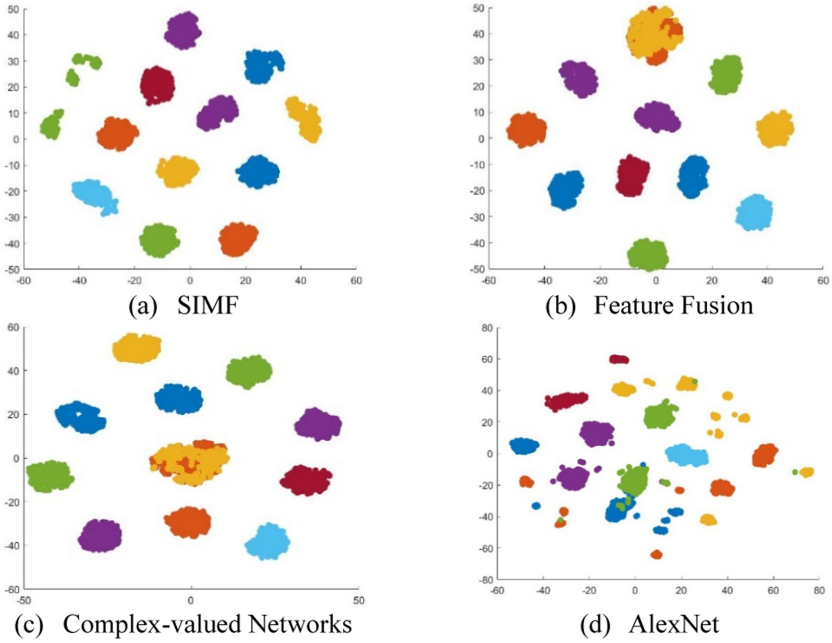


Fig. 10. Feature visualization of different methods at SNR = 6 dB

5 Conclusions

In order to take advantage of the complementarity between multimodal information and realize the conversion and fusion of multi-domain information, we proposed SIMF framework based on the multimodal information fusion of signal statistical graph domain and I/Q waveform domain to achieve AMC. We use series fusion to obtain a more informative joint feature representation of multimodal features. Furthermore, we use a coordinated integration architecture to achieve mutual cooperation and constraints between multiple modes, which is conducive to maintaining the unique characteristics and exclusivity of each modal. The final simulation results show that our proposed framework achieves superior performance than other unimodal models and multimodal models under the entire SNR. In further work, we will continue to explore other multimodal fusion methods to improve the robustness and performance of our model.

References

1. You, X., Zhang, C., Tan, X., et al.: AI for 5G: research directions and paradigms. Sci. China Inf. Sci. **62**(2), 1–13 (2019)
2. You, X., Wang, C.X., Huang, J., et al.: Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts. Sci. China Inf. Sci. **64**(1), 1–74 (2021)
3. Zhang, P., Kai, N.I.U., Hui, T., et al.: Technology prospect of 6G mobile communications. J. Commun. **40**(1), 141 (2019)
4. Wang, Y., Wang, J., Zhang, W., et al.: Deep learning-based cooperative automatic modulation classification method for MIMO systems. IEEE Trans. Veh. Technol. **69**(4), 4575–4579 (2020)

5. Tu, Y., Lin, Y., Wang, J., et al.: Semi-supervised learning with generative adversarial networks on digital signal modulation classification. Comput. Mater. Continua **55**(2), 243–254 (2018)
6. Lin, Y., Zhu, X., Zheng, Z., et al.: The individual identification method of wireless device based on dimensionality reduction and machine learning. J. Supercomput. **75**(6), 3010–3027 (2019)
7. Lin, Y., Tu, Y., Dou, Z.: An improved neural network pruning technology for automatic modulation classification in edge devices. IEEE Trans. Veh. Technol. **69**(5), 5703–5706 (2020)
8. Yun, L., Haojun, Z., Xuefei, M., et al.: Adversarial attacks in modulation recognition with convolutional neural networks. IEEE Trans. Reliab. **70**(1), 389–401 (2021). https://doi.org/10.1109/TR.2020.3032744
9. Lin, J., Wei, M.: Network security situation prediction based on combining 3D-CNNs and Bi-GRUs. Int. J. Perform. Eng. **16**(12), 1875–1887 (2020)
10. Restuccia, F., Melodia, T.: Physical-Layer Deep Learning: Challenges and Applications to 5G and Beyond. arXiv preprint arXiv:2004.10113 (2020)
11. Peng, S., Jiang, H., Wang, H., et al.: Modulation classification based on signal constellation diagrams and deep learning. IEEE Trans. Neural Netw. Learn. Syst. **30**(3), 718–727 (2018)
12. Jajoo, G., Kumar, Y., Yadav, S.K.: Blind signal PSK/QAM recognition using clustering analysis of constellation signature in flat fading channel. IEEE Commun. Lett. **23**(10), 1853–1856 (2019)
13. Lin Yun, T., Ya, D.Z., et al.: Contour stella image and deep learning for signal recognition in the physical layer. IEEE Trans. Cogn. Commun. Netw. **7**(1), 34–46 (2021). https://doi.org/10.1109/TCCN.2020.3024610
14. 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)
15. Xu, J., Luo, C., Parr, G., et al.: A spatiotemporal multi-channel learning framework for automatic modulation recognition. IEEE Wirel. Commun. Lett. **9**(10), 1629–1632 (2020)
16. O'Shea, T.J., Roy, T., Clancy, T.C.: Over-the-air deep learning based radio signal classification. IEEE J. Sel. Top. Sig. Process. **12**(1), 168–179 (2018)
17. Cheng, X., He, J., He, J., et al.: Cv-CapsNet: complex-valued capsule network. IEEE Access **7**, 85492–85499 (2019)
18. Tu, Y., Lin, Y., Hou, C., et al.: Complex-valued networks for automatic modulation classification. IEEE Trans. Veh. Technol. **69**(9), 10085–10089 (2020)
19. Meng, F., Chen, P., Wu, L., et al.: Automatic modulation classification: a deep learning enabled approach. IEEE Trans. Veh. Technol. **67**(11), 10760–10772 (2018)
20. Li, R., Li, L., Yang, S., et al.: Robust automated VHF modulation recognition based on deep convolutional neural networks. IEEE Commun. Lett. **22**(5), 946–949 (2018)
21. Baltrušaitis, T., Ahuja, C., Morency, L.P.: Multimodal machine learning: a survey and taxonomy. IEEE Trans. Pattern Anal. Mach. Intell. **41**(2), 423–443 (2018)
22. Zhang, C., Yang, Z., He, X., et al.: Multimodal intelligence: Representation learning, information fusion, and applications. IEEE J. Sel. Top. Sig. Process. **14**(3), 478–493 (2020)
23. 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. Netw. **5**(3), 469–478 (2019)
24. Wu, H., Li, Y., Zhou, L., et al.: Convolutional neural network and multi-feature fusion for automatic modulation classification. Electron. Lett. **55**(16), 895–897 (2019)
25. Qi, P., Zhou, X., Zheng, S., et al.: Automatic Modulation Classification Based on Deep Residual Networks with Multimodal Information. IEEE Trans. Cogn. Commun. Netw. **7**, 21–33 (2020)
26. Van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. J. Mach. Learn. Res. **9**(11), 1–48 (2008)