Inception Resnet V2-ECANet Based on Gramian Angular Field Image for Specific Emitter Identification

Zibo Ma[1], Chengyu Wu[1(✉)], Chen Zhong[2], and Ao Zhan[1]

[1] School of Information Science and Engineering, Zhejiang Sci-Tech University, Hangzhou 310018, People's Republic of China
jerry916@zstu.edu.cn

[2] School of Computer Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, People's Republic of China

Abstract. In this paper, we seek to efficiently and accurately identify the specific emitter with the DroneRF dataset. Firstly, we convert one-dimensional data to a Gramian Angular Field (GAF) image showing a spatial correlation, and add three kinds of noise to the original GAF image to prevent overfitting. Secondly, we propose an Inception-Resnet-V2 model based on the attention mechanism ECANet, which can improve the training effect obviously. Finally, we verify the validation and accuracy of the proposed model with GAF in the experimental results. Test accuracy of Inception Resnet V2-ECANet reaches 99.5% by using the same training set of five networks through 10-fold cross-validation.

Keywords: Specific Emitter Identification · Gramian Angular Field · Convolutional Neural Network · Inception Resnet V2-ECANet

1 Introduction

Specific emitter identification (SEI) is an essential research content in the field of electromagnetic spectrum confrontation [1], and is widely used in military applications in electronic reconnaissance, intelligence, communication confrontation, and other fields [2]. SEI is to compare the radio frequency fingerprint (RFF) in the signals received by the receiver to detect and identify the radiation source.

In [3], CNN is utilized to classify the Radio Frequency (RF) signals transmitted during the real-time communication session between the UAV and its controller. In [4], the authors use Discrete Fourier Transform (DFT) to extract features from the low-frequency RF signal of UAV to flight controller communication in the dataset, utilize the XGBoost model, and use 10-fold cross-validation

Supported by the Fundamental Research Funds of Zhejiang Sci-Tech University under grant 2021Q029.

X. Jiang (Ed.): MLICOM 2022, LNICST 481, pp. 26–38, 2023.
https://doi.org/10.1007/978-3-031-30237-4_3

to evaluate the model. The average accuracy of UAV detection, UAV type identification, and UAV flight mode identification is 99.96%, 90.73%, and 70.09%, respectively. The RF signal is subjected to discrete DFT as the input of the network in [5]. The intrinsic feature maps of RF signals are gathered from three different types of UAVs through the deep convolution layer. In [6], two Fully Connected Neural Networks (FCNNs) and two CNNs were mentioned and trained. In [7], the authors propose using high-performance CNN to detect and classify UAVs effectively. In [8], an integrated learning method for improving UAV identification and detection is proposed, which is composed of KNN and XGBoost learning methods. In [9], the authors prove the effectiveness of the selection of compressed sensing technology and multi-channel random demodulation for data sampling. Multi-stage DL UAV identification and detection method are proposed based on the difference in communication signals between UAV and controller in unusual situations. Two neural network structures (DNN and CNN) are constructed by the DL algorithm and are evaluated and verified by 10-the fold cross-validation method. The accuracy of detecting UAVs, and identifying UAV types and their flight modes is above 99%.

In [10], the author introduces a DL model of a multi-channel one-dimensional convolutional neural network (1-DCNN). This model consists of two parts, one is a feature extractor for learning features from the original data, and the other is Multilayer Perceptron (MLP) which performs a classification function through the features learned in the first part. On this basis, the RF signal is transformed into the frequency band of the spectrum for network training. The performance of the 1-DCNN proposes in the literature is significantly better than that of the DNN technology for UAV type and state identification introduced in [11]. The average accuracy of the first model for the two classifications is 100%. The second model for the four classifications is 94.6%, and the last model for every class is 87.4%.

In [12], the VGG16 neural network is used to evaluate the classification results of UAVs by Logistic Regression (LR), Support Vector Machine (SVM), and random forest. The results demonstrate that the network input for higher power spectrum density (PSD) image classification accuracy is higher. The accuracy rate of UAV detection is 100%. UAV types classification is 88.6%, and ten categories containing flight modes are 87.3%. In [13], the author regards UAV detection as an image classification problem. PSD, Spectrogram, Histogram, and Raw IQ constellation diagram are available in network input to CNN ResNet50 for feature extraction and classification. The results are verified by cross-validation and independent dataset, and the accuracy of the input PSD diagram in 10 classifications exceeds 91%.

In summary, the current research on the detection of UAVs and the identification of UAV types is better. So the contemporary task requires an in-depth study of the UAV flight pattern classification tasks. In [4,7–9,11–13] all carry out similar tasks for UAV flight modes. To identify the different flight modes of the three UAVs more accurately based on the theories and methods in the above literature, the CNN model for effectively identifying the different flight modes of the three

UAVs is obtained by changing the input form of the network model, expanding the dataset, and training multiple neural networks for multiple verifications.

In this paper, we explore extracting the spatial correlation with CNN for the input data firstly, and convert each one-dimensional signal data to a GAF image, by adding three kinds of noise to the GAF image to prevent overfitting. Then, we propose an improved Inception-Resnet-V2 network with three ECANet attention mechanisms to optimize the network performance. Finally, we study the proposed network with VGG16, Resnet18, Resnet50, Inception-Resnet-V2, and Inception Resnet V2-ECANet. The results demonstrate that the proposed network can achieve higher accuracy than others.

2 The Network Structure Formation

The designed network structure is given in Fig. 1, which is composed of data pre-processing, improved Inception Resnet V2-ECANet model, 10-fold cross-entropy verification, and classification.

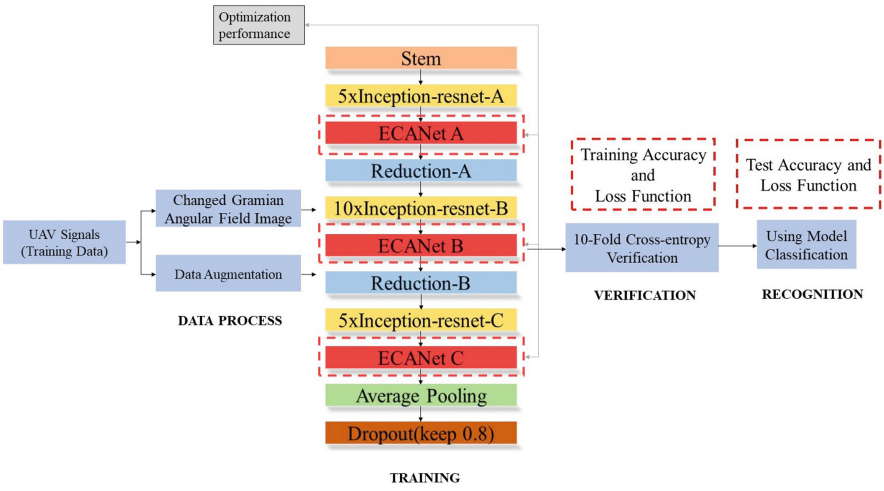


Fig. 1. Overall Structure

The spatial connection of images is based on the correlation between pixels. The local pixel correlation is accurate, while the far pixel correlation is weak. The image of GAF is the calculation of eccentric covariance between features, which expresses the correlation between data features and reflects the tightness between data. It is reasonable to convert the one-dimensional signal data into the GAF image as the input of the CNN, which avoids the complex process and effectively extracts the corresponding features from a large number of data. The increase in depth is not hard to cause over-fitting due to the insufficient training

data, which makes the model unable to complete sufficient training. To ensure that the training data is vast and the sample of each type of data is balanced through data enhancement, the quality and quantity of the original data are improved without a substantial increase in the dataset.

In order to better fit the characteristic network layer, it is necessary to be deep enough to ensure the reasonable use of calculation effective and reasonable training time. The deeper the network is, the more intricate features will be learned through a nonlinear expression. This network is improved based on the Inception Resnet V2 network. In addition, under the current constraints of computer computing resources, an attention mechanism is absolutely a necessary means to improve the effectiveness. ECANet is an improvement attention mechanism. The original Inception-Resnet-V2 is improved by adding the ECANet module.

2.1 GAF Image

GAF image is the visualization of the Gram matrix which is regarded as an eccentric covariance matrix between features. Then Gram calculates the correlation between two features to upgrade the classification accuracy of the model [14]. The main process of conversion is to convert a one-dimensional signal sequence into a polar coordinate representation and generate a GAF matrix with trigonometric function.

Step 1: First of all, the dimension reduction of Piecewise Aggregation Approximation (PAA) piecewise aggregation approximation is needed. PAA always satisfies the lower bound condition. The formula is shown as (1), (2), (3), and the sequence length n is used to aggregate the time series by taking the average value of point M.

$$\bar{x}_i = \frac{M}{n} \sum_{j=n/M(i-1)+1}^{(n/M)i} x_j \tag{1}$$

$$\bar{y}_i = \frac{M}{n} \sum_{j=n/M(i-1)+1}^{(n/M)i} y_j \tag{2}$$

$$D_{PAA}(\bar{X}, \bar{Y}) \equiv \sqrt{\frac{n}{M}} \sqrt{\sum_{i=1}^{M} (\bar{x}_i - \bar{y}_{i)}} \tag{3}$$

Step 2: Normalization is carried out to reduce the signal value between [−1, 1] or [0, 1], and the formula is as follows:

$$\tilde{x}_{-1}^{i} = \frac{x_i - \max(X) + (x_i - \min(X))}{\max(X) - \min(X)} \tag{4}$$

Or

$$\tilde{x}_{0}^{i} = \frac{x_i - \min(X)}{\max(X) - \min(X)} \tag{5}$$

Step 3: Use the timestamp N as the radius, the scaling value in Step 2 is used as the angle chord to generate the polar coordinates to provide the angle to get the new time series X. The formula is as follows :

$$\phi = \arccos(\tilde{x}_i), -1 \leq \tilde{x}_i \leq 1, \tilde{x}_i \in \tilde{X} \tag{6}$$

$$r = \frac{t_i}{N} \tag{7}$$

t is a timestamp that is a constant factor to adjust the polar coordinates.

Step 4: Generate GASF and calculate the value of each pixel. The formula is as following:

$$GASF = \cos(\phi_i + \phi_j) \tag{8}$$

After the above four steps, a total of 22700 grid GAF graphs can be obtained. In Fig. 2, GAF images of ten signals in the dataset are shown. One signal for each type is viewed as a display.

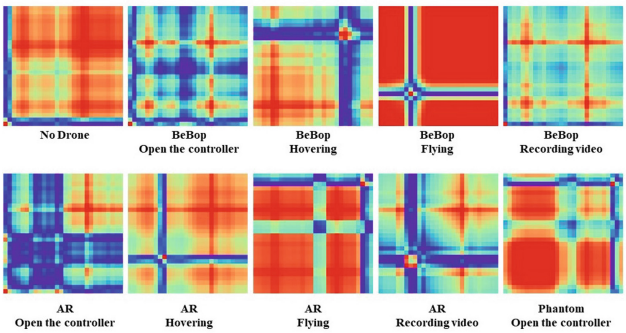
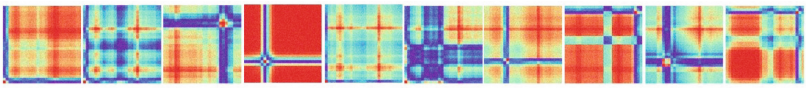


Fig. 2. Representation of each type of UAV signal converted to a GAF

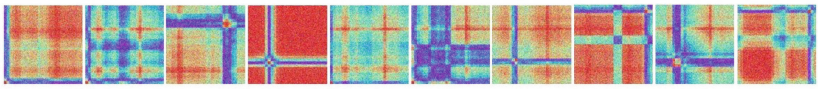
2.2 Noise Adding for GAF Images

Based on prior experience, a network with deeper layers is needed to extract subtle features. However, the small amount of original data is not difficult to cause over-fitting, so it is necessary to carry out data enhancement and expansion to increase the number and diversity of training samples. By using the image processing method, more data images are generated on the limited data images to reinforce the generalization ability of the model. The method is to add random noise, Gaussian noise, and salt and pepper noise to 22700 images converted from one-dimensional data to two-dimensional images. As showm in Fig. 3.

Random noise is generated by the accumulation of a large number of random fluctuations generated randomly in time. The excessive number of noise points

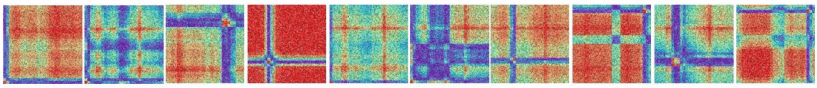


(a) GAF Image after Adding Random Noise



$z = 0.1\ \sigma = 0.5$

(b) GAF Image after Adding Gaussian Noise with Mean of 0.1 and Variance of 0.5



$SNR = 0.8$

(c) GAF Image Adding Salt And Pepper Noise With SNR=0.8

Fig. 3. Three noises are added to each signal for data enhancement.

will lead to a larger gap between the original image and the image after adding noise. As shown in Fig. 3(a).

The probability density of Gaussian noise obeys Gaussian distribution. The probability density expression is:

$$P(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(z-\bar{z})^2/2\sigma^2} \tag{9}$$

where σ is the standard deviation of z, and $\bar{z}$ is the mean. The brightness of the image is determined by the mean value. The larger the variance, the more dispersed the data, and the more noise, as shown in Fig. 3(b).

Salt and pepper noise in the image of random black-white pixels is a kind of noise caused by signal pulse intensity. Signal-to-noise ratio SNR needs to be specified, followed by random access to each pixel position to be added noise, and finally, the pixel value is specified as 255 or 0, as shown in Fig. 3(c).

2.3 The Improved Inception Resnet V2-ECANet

The improved inception Resnet V2-ECANet consists of three parts, namely the Resnet [15]module, Inception [16], and ECANet [17], one of the three attention mechanisms. The main principle is to upgrade the Inception-Resnet-V2 [18]network and add the ECANet module of attention mechanism after three Inception Resnet modules in the Inception-Resnet-V2 network.

As shown in Fig. 4, the input size of the Stem module in the main structure is 3×3 in the Inception-Resnet-V2. Three convolutions, maximum pooling, two convolutions, and final stacking are required after input. Replacing stems with convolutive pooled sequential connection is intended to design deeper networks.

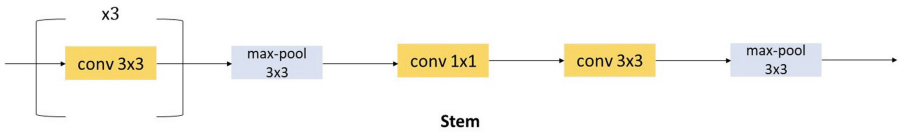


Fig. 4. Stem Module in Main Structure

Three Inception-Resnet blocks in the backbone structure are the most important. The module structure is given in Fig. 5. The Inception-resnet module introduces ResNet layers in the Inception module. The different structure helps to accelerate convergence and prevent gradient dispersion. The Inception module solves the use and pooling operation of the network self-determined filter. The module uses $1 \times n$ and $n \times 1$ instead of $n \times n$ convolution to reduce the amount of calculation.

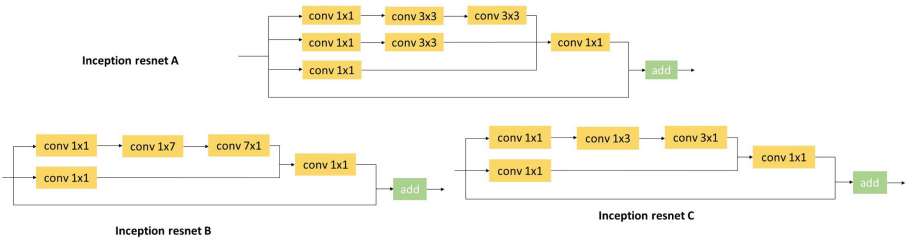


Fig. 5. Inception-resnet module in Main Structure

ECANet module is inserted after each Inception-Resnet module, as showed in Fig. 6. ECANet is an improvement of SENet [19]. ECANet utilizes the satisfactory cross-channel capability of convolution.

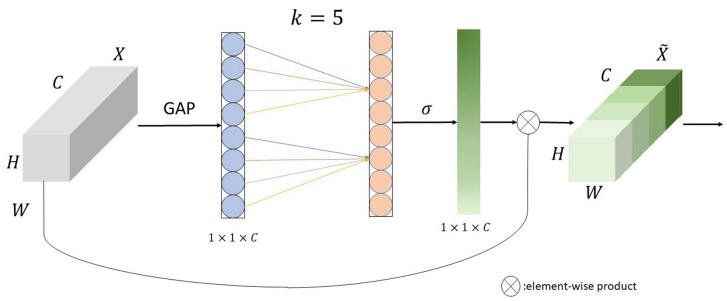


Fig. 6. ECANet Network Structure

The convolution kernel k of 1D convolution is called the coverage rate of cross-channel interaction, which is extremely significant for the number of channels to be considered in the weight calculation of the attention mechanism. When the model is implemented, select the convolution kernel size k. The coverage of a cross-channel interaction is proportional to the number of filters C. There exists a mapping φ between k and C.

$$C = \phi(k) \tag{10}$$

$\varphi(k)$ can be regarded as a linear function, then

$$\phi(k) = r * k - b \tag{11}$$

C is usually $2 \wedge n$, let $n = r * k - b$, then

$$C = \phi(k) = 2^{(r*k-b)} \tag{12}$$

$$k = \psi(C) = \left| \frac{\log_2(C)}{r} + \frac{b}{r} \right|_{odd} \tag{13}$$

where $= \frac{\log_2(C)}{r} + \frac{b}{r}$, $|t|_{odd}$ denotes the nearest odd number to t and $r = 2, b = 1$.

There are also two Reduction modules in the main structure, as shown in Fig. 7. The convolution structure of the observable module is used many times. One of the reasons can be reduced dimension operation. The second reason is that after adding nonlinear incentives, the expression ability of the network is enhanced. On the premise of keeping the scale of the feature map unchanged, the nonlinear characteristics are greatly increased. It can increase the overall network depth and extract more detailed features.

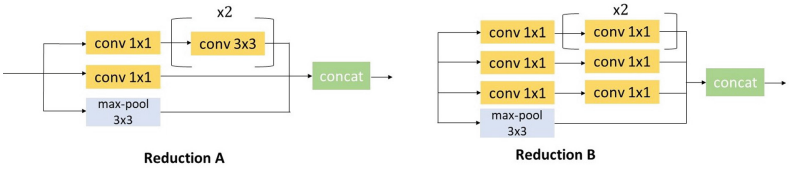


Fig. 7. Reduction Module in Main Structure

In addition, the core structure of the network is more complex, with more parameters. The global average pooling layer greatly reduces the network parameters and prevents overfitting. The Dropout layer is also designed to prevent overfitting.

This shows that since ECANet does not change the dimensions of input and output after increasing. ECANet is inserted after each module. However, in order not to increase network complexity as much as possible, attention mechanism is introduced to optimize network performance only after three important Inception-Resnet modules in the network structure.

3 Experiment

The model of training and testing DL methods needs an appropriate dataset. In [20], an open-source DroneRF dataset generated in the laboratory is constructed from three brands of UAV (AR, Bepop, Phantom), collected four flight modes: opening and connecting to the controller, automatic hovering, no video recording flight, and video recording flight signal respectively, and collected the radio frequency background signal without UAV, as shown in Table 1. The DroneRF dataset has taken advantage of by many papers and has a high use value. The dataset is stored in Comma Separated Values (CSV) format to facilitate loading. The network calculation loss is large, so it is necessary to train in a configured GPU environment. In the experimental setting, we utilize Tensorflow 2.5 with GPU RTX3060Ti in Ubuntu 18.04.

Table 1. The specific content of the DroneRF dataset.

UAV Detection	UAV Type	UAV Flight Mode	Segments
No Drone	No Drone	No Drone	4100
Drone	Bebob	Opening the controller	2100
Drone	Bebob	Hovering	2100
Drone	Bebob	Flying	2100
Drone	Bebob	Recording video	2100
Drone	AR	Opening the controller	2100
Drone	AR	Hovering	2100
Drone	AR	Flying	2100
Drone	AR	Recording video	1800
Drone	Phantom	Opening the controller	2100

The RF signals of UAVs in the dataset are divided into three levels, which are used to train three different neural networks, as shown in Fig. 8. The first network is utilized to determine whether the UAV exists, which is determined as a binary classification problem. The second network is used to determine the type of UAV, which is determined by four classification problems. The third network is used to identify the UAV flight mode and determine the appropriate class of problems. Given the high accuracy of the two and four classifications in the existing research results, that is, the full text mainly discusses the ten categories.

This experiment verifies the performance of Resnet18, Resnet50, Inception V3, Inception-Resnet-V2, and Inception Resnet V2-ECANet by a Tenfold cross validation. The cross-entropy loss function is as follows:

$$Loss = -\sum_{i=1}^{10} y_i \times \log_2(\hat{y}_i) \tag{14}$$

In (14), y_i represents the type i probability of the true label, and $\hat{y}_i$ represents the type i probability predicted by the model.

The flight pattern classification ability of the five networks in the dataset is 20 epochs of the same iteration and the same dataset are applied. The test accuracy is 96.76%, 94.26%, 51.84%, 99.4% and 99.5%, respectively. As shown in Fig. 9, clearly illustrates the specific training accuracy and test accuracy of each network for 20 epochs, and its training loss function and test loss function.

It is observed from Fig. 9 that the proposed Inception Resnet V2-ECANet network and Inception-Resnet-V2 have reached stability after 20 training epochs. Inception Resnet V2-ECANet has reached 99.5% accuracy and remained stable after the fourth training epoch, while Inception-Resnet-V2 has reached 99.4% and remained stable after the eighth training epoch.

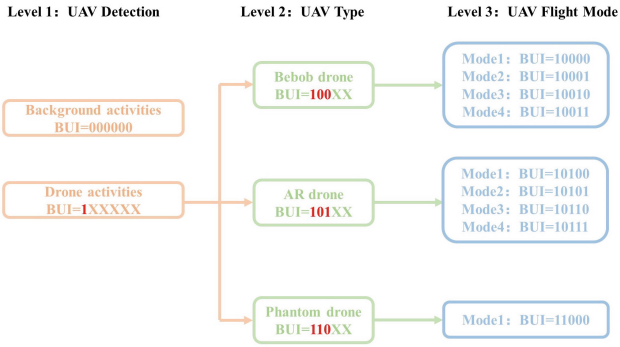


Fig. 8. Signal Distribution for Three Classifications

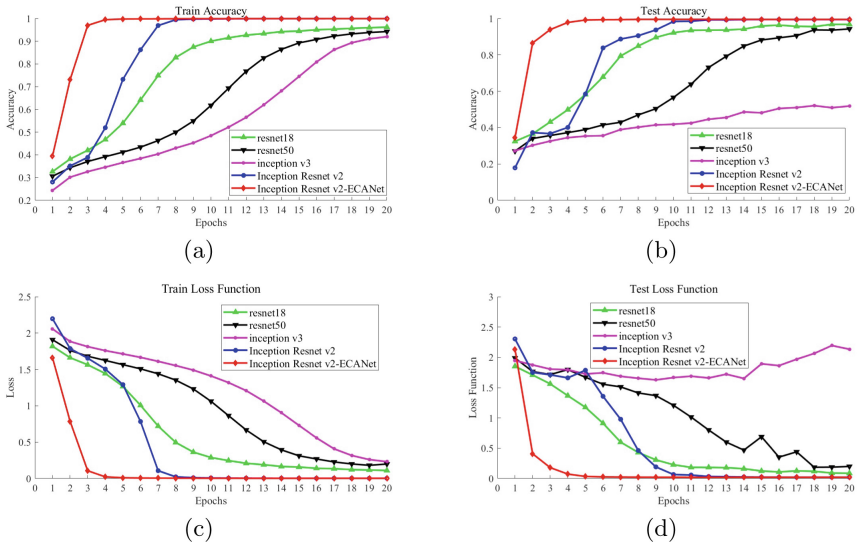


Fig. 9. Accuracy and Loss performance of the proposed network compared with others.

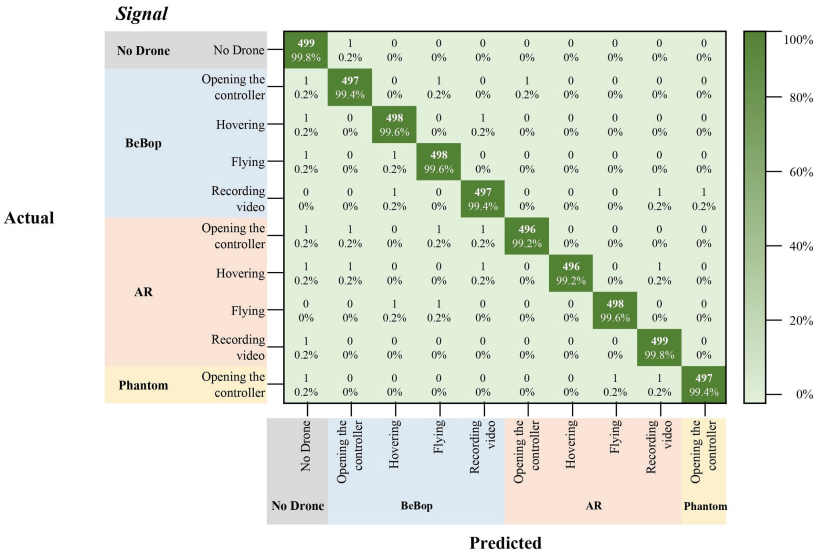


Fig. 10. Confusion Matrix of Prediction Model. (Color figure online)

Figure 10 represents a confusion matrix used to discriminate UAV flight modes. The row of the matrix represents the prediction category, and the column of the matrix represents the actual category. The cells showing the diagonal of dark green represent the correct classification, and the color becomes lighter. It represents the proportion of the wrong classification. There are TP (True Positive), FP (False Positive), FN (False Negative), and TN (True Negative) four parameters in the index.

In addition to the setting of the learning rate, after repeated verification, the over-fitting phenomenon in the model training will also lead to the uneven distribution of the ability of the model to learn various signals due to the random selection of data sets as training samples. It will be difficult for the model to extract more subtle features due to the high similarity of training samples and the small sample size of the training set. The solution of data enhancement is adopted to make the training set samples of each type of signal reach 10 million, and a total of 100,000 images are used as the training set for this experiment.

4 Conclusion

In this paper, the one-dimensional data signal is converted into a two-dimensional GAF image as the input form of the network. Owing to the great influence of the attention mechanism on image classification, the ECANet attention mechanism is added to the proposed network to upgrade the training accuracy. The confusion matrix shows that the proposed inception Resnet V2-ECANet network model with a complex structure has the best prediction effect. The trained Inception

Resnet V2-ECANet network effectively classifies the nine RF signals and RF scene signals of UAV flight mode, and the accuracy rate can reach 99.5%. In future work, straightforward complexity models should be utilized to reduce the training time without reducing the model performance for SEI.

References

1. Huang, G., Yuan, Y., Wang, X., Huang, Z.-T.: Specific emitter identification for communications transmitter using multi-measurements. Wirel. Pers. Commun. **94**(3), 1523–1542 (2017)
2. Sa, K., Lang, D., Wang, C., Bai, Yu.: Specific emitter identification techniques for the Internet of Things. IEEE Access **8**, 1644–1652 (2020)
3. Al-Emadi, S., Al-Senaid, F.: Drone detection approach based on radio-frequency using convolutional neural network. In: 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), pp. 29–34 (2020)
4. Medaiyese, O.O., Syed, A., Lauf, A.P.: Machine learning framework for RF-based drone detection and identification system. In: 2021 2nd International Conference on Smart Cities, Automation & Intelligent Computing Systems (ICON-SONICS), pp. 58–64 (2021)
5. Akter, R., Doan, V.-S., Lee, J.-M., Kim, D.-S.: CNN-SSDI: convolution neural network inspired surveillance system for UAVs detection and identification. Comput. Netw. **201**, 108519 (2021)
6. Basan, E.S., Tregubenko, M.D., Mudruk, N.N., Abramov, E.S.: Analysis of artificial intelligence methods for detecting drones based on radio frequency activity. In: 2021 XV International Scientific-Technical Conference on Actual Problems of Electronic Instrument Engineering (APEIE), pp. 238–242 (2021)
7. Huynh-The, T., Pham, Q.-V., Nguyen, T.-V., Costa, D.B.D., Kim, D.-S.: RF-UAVNET: high-performance convolutional network for RF-based drone surveillance systems. IEEE Access **10**, 49696–49707 (2022)
8. Nemer, I., Sheltami, T., Ahmad, I., Ul-Haque Yasar, A., Abdeen, M.A.R.: RF-based UAV detection and identification using hierarchical learning approach. Sensors **21**(6), 1947 (2021)
9. Mo, Y., Huang, J., Qian, G.: Deep learning approach to UAV detection and classification by using compressively sensed RF signal. Sensors **22**(8), 3072 (2022)
10. Allahham, Mhd.S., Khattab, T., Mohamed, A.: Deep learning for RF-based drone detection and identification: a multi-channel 1-D convolutional neural networks approach. In: 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), pp. 112–117 (2020)
11. Al-Sa'd, M.F., Al-Ali, A., Mohamed, A., Khattab, T., Erbad, A.: RF-based drone detection and identification using deep learning approaches: an initiative towards a large open source drone database. Futur. Gener. Comput. Syst. **100**, 86–97 (2019)
12. Swinney, C.J., Woods, J.C.: Unmanned aerial vehicle flight mode classification using convolutional neural network and transfer learning. In: 2020 16th International Computer Engineering Conference (ICENCO), pp. 83–87 (2020)
13. Swinney, C.J., Woods, J.C.: Unmanned aerial vehicle operating mode classification using deep residual learning feature extraction. Aerospace **8**(3), 79 (2021)
14. Wang, Z., Oates, T.: Imaging time-series to improve classification and imputation, pp. 3939–3945 (2015)

15. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016)
16. Szegedy, C., et al.: Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9 (2015)
17. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: ECA-Net: efficient channel attention for deep convolutional neural networks, pp. 11531–11539 (2020)
18. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-V4, inception-ResNet and the impact of residual connections on learning, pp. 4278–4284 (2017)
19. Jie, H., Shen, L., Albanie, S., Sun, G., Enhua, W.: Squeeze-and-excitation networks. IEEE Trans. Pattern Anal. Mach. Intell. **42**(8), 2011–2023 (2020)
20. Allahham, M.H.D.S., Al-Sa'd, M.F., Al-Ali, A., Mohamed, A., Khattab, T., Erbad, A.: DroneRF dataset: a dataset of drones for RF-based detection, classification and identification. Data Brief **26**, 104313 (2019)