HRRP Target Recognition Based on the Dual-mode Gram Angle Field Features and the Multi-level CNN

2021 IEEE 6th International Conference on Signal and Image Processing (ICSIP) | 978-1-6654-3904-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICSIP52628.2021.9688620

Erchun Xu[1,2], Gang Xiong[1,2*], Shuning Zhang[3] and Huichang Zhao[3]
[1.] School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China
[2.] Institute of Navigation and Sensing, Shanghai Jiao Tong University, Shanghai, China
[3.] School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, China
gxiong@sjtu.edu.cn, xuerchun@sjtu.edu.cn

***Abstract*—Aiming at the Carrier-Free Ultra-Wide-Band Radar (CF-UWBR) under the near-range conditions, a novel target recognition method based on the Gram Angle Field (GAF) feature extraction and the Inception V3 network is proposed in this paper. The one-dimensional High-Resolution Range Profile (HRRP) of CF-UWBR is highly sensitive to observation angle so that the performance of direct recognition using the original time-domain HRRP data is not satisfying. In response to this problem, the GAF method is employed to convert the HRRP data into a two-dimensional image firstly, to maintain and highlight the correlation features between the adjacent time points in the polar coordinate domain. Furthermore, the Inception V3 network is selected for the target recognition based on the dual-mode GAF features, due to the excellent performance with the fewer computing resources. The experimental results on the actual measured CF-UWBR data set indicated that the proposed method is superior in performance to the methods based on the Support Vector Machine (SVM), the Gated Recurrent Unit (GRU) network, or the Convolutional Neural Network (CNN) with the original HRRP image. Especially when the dual-mode GAFs are used, the recognition accuracy of 97.63% can be achieved, about 1.95% higher than the STOA.**

***Keywords*—GAF features, CNN network, UWB, HRRP target recognition**

I. INTRODUCTION

Carrier-Free Ultra-Wide-Band Radar (UWBR) is a short-range detection system with a working distance ranging from several meters to a hundred meters, and the target is in the quasi-near-field range. Compared with the traditional radar signal and image recognition [1], [2], UWBR can provide richer electromagnetic spectrum features for target identification, including one-dimensional High-Resolution Range Profile (HRRP), the time-domain echo signal, the signal spectrogram, and the other features. The robustness and high resolvability of the extracted features are critical to the performance of the radar target recognition. Generally, radar target recognition models can be roughly divided into two categories: the classic machine learning algorithms-based method and the Deep Learning (DL) based method. A large number of effective feature extraction and recognition methods have been proposed in the classic machine learning methods, including the Decision Tree (DT) [3], the Random Forest (RF) [4], and the Support Vector Machine (SVM) [5]. In recent years, deep neural networks provided novel tools for solving various tasks in image understanding and recognition. Among them, one-dimensional and two-dimensional deep networks are mainly used in the field of radar target recognition. Wang [6] introduced the radar target recognition method based on the one-dimensional - deep network and obtained a higher recognition accuracy than the traditional recognition methods. Currently commonly used one-dimensional deep learning models include Convolutional Neural Network (CNN) [7], Recurrent Neural Network (RNN) [8], and Deep Belief Network (DBN) [9], etc. The research indicated that the higher recognition accuracy can be achieved by the one-dimensional DL methods than the DT, MLP, and other traditional methods.

Besides, the one-dimensional HRRP data can be transformed into the image, and further the image recognition method based on the CNN can be utilized. The visualization method of the HRRP data includes the data superposition and arrangement, the direct 2D-HRRP image, as well as the time-frequency presentation, etc. When directly processing 1D-HRRP as a 2D image mentioned in [8], some information will be lost due to down-sampling and binarization, while the blank area in the HRRP image will lead to a large amount of 2D convolution operation. Also, the time-frequency presentation method will introduce too much calculation load to meet the requirements of an online processing system. The Gram Angle Field (GAF) algorithm [10] is introduced in this paper, to improve the interpretability of the conversion from one-dimensional time series to two-dimensional images, meanwhile, lay the foundation for the effectiveness of feature extraction. The two-dimensional convolutional neural network is widely used in recognition tasks [11] such as Alex Net [12], Caffe Net [13], VGGNet [14], GoogLeNet [15] and ResNet [16] and other network structures. Based on the comprehensive consideration of network performance and computing efficiency, the Inception V3[17] model is chosen in this paper for the HRRP target recognition. The first part of this paper introduces the development of the field and the existing methods, the second part explains the feature extraction method and network structure used in this paper, the third part describes the experimental process and analysis of the results in detail and the fourth part gives the conclusion of this paper. The main contributions of this paper are as follows:

(1) Utilizing the GAF method to quickly convert one-dimensional time series into two-dimensional images, while retaining the timing information of the original signal and

avoiding the problem of the high sensitivity of HRRP images, making it extremely effective with the CNN network.

(2) Utilizing the Inception V3 network model to recognize the UWBR target additionally with the help of the multi-level convolution method to learn the rich depth features of the UWBR target, ensuring the effectiveness of the recognition.

II. METHODOLOGY

A. Two-dimensional Transformation HRRP Based on GAF Method

The Gram Angle Field evolved based on the GAF matrix using the polar coordinate system instead of the typical Cartesian coordinate system to represent the time series. In the GAF matrix, each element is the cosine of the sum of the paired time series values (GASF) or the difference of the sine values of the paired time series values (GADF) in the polar coordinate system.

The basic principle of GAF processing involves containing GAF features and representing the original input sequence by polar coordinates instead of rectangular coordinates. The definition of GAF is as follows.

Firstly, the input sequence $X=\{x_1,x_2,...,x_n\}$ is normalized as

$$\tilde{x}_i=\frac{(x_i-\max(X))+(x_i-\min(X))}{\max(X)-\min(X)}. \tag{1}$$

The representation of X in polar coordinate system is:

$$\begin{cases}\phi=\arccos(\tilde{x}_i),-1\le\tilde{x}_i\le 1,\tilde{x}_i\in X\\ r=\frac{t_i}{N},t_i\in N\end{cases}, \tag{2}$$

where t_i is the timestamp of the sequence, and N is used to adjust the radial span of the polar coordinate. Its feasibility lies in monotonicity on the $\cos\phi,\phi\in[0,\pi]$ .Therefore, given an input sequence, the result obtained from this encoding method is unique. Since the input one-dimensional HRRP is a real number signal, the mapping range is [0, π]. By combining the original sequence information and the correlation between each point in the polar coordinate system, the GAF feature based on the one-dimensional HRRP is defined as $\langle u_i,u_j\rangle=\cos(\phi_i+\phi_j)$.

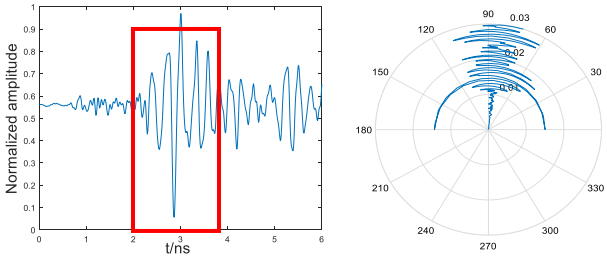
GAF matrix can be redefined as the inner product operation as the following formula:

$$GASF=\begin{pmatrix}\cos(\phi_1+\phi_1) & \cos(\phi_1+\phi_2) & \dots & \cos(\phi_1+\phi_n)\\ \cos(\phi_2+\phi_1) & \dots & \dots & \cos(\phi_2+\phi_n)\\ \dots & \dots & \dots & \dots\\ \cos(\phi_n+\phi_1) & \dots & \dots & \cos(\phi_n+\phi_n)\end{pmatrix} \tag{3}$$

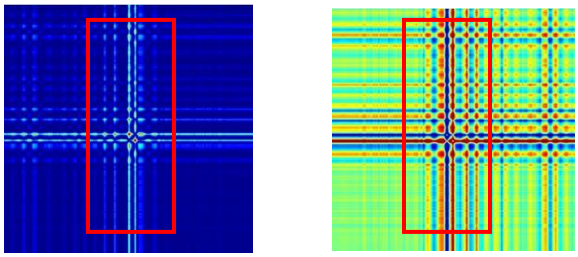
and

$$GADF=\begin{pmatrix}\sin(\phi_1-\phi_1) & \sin(\phi_1-\phi_2) & \dots & \sin(\phi_1-\phi_n)\\ \sin(\phi_2-\phi_1) & \dots & \dots & \sin(\phi_2-\phi_n)\\ \dots & \dots & \dots & \dots\\ \sin(\phi_n-\phi_1) & \dots & \dots & \sin(\phi_n-\phi_n)\end{pmatrix}. \tag{4}$$

Based on the above discussion, GAF defines a single-channel feature focusing on the correlation between sequence points. In addition, according to the definition of the main diagonal, different targets can be distinguished from the high-dimensional features learned from the deep neural network. Compared with the ordinary echo signal images, the GAF feature map focuses on a specific part of the input signal corresponding to the part in the one-dimensional HRRP data of the original model. The positional relationship retains the time dependence of the HRRP data, which is beneficial to the subsequent recognition.



(a) Original HRRP data (b) HRRP in polar coordinate domain



(c) GASF feature (d) GADF feature

Fig. 1. The dual-mode GAF Transformation of 1D-HRRP.

Fig.1 (a)~(d) shows the radar one-dimensional HRRP, the transformation of HRRP in the polar coordinate system, the corresponding GASF and GADF features, respectively. It can be seen from figure 1 (a) that the echo signal pulses are mainly concentrated in the framed interval, exactly corresponding to the framed interval in the GAF feature. It shows that the GAF feature well reflects the location information of the strong scattering point and the information of time series. In a two-dimensional image, the signal can be completely mapped through different features such as colors, points, and lines at the corresponding locations in GASF and GADF becoming great help of recognition.

B. Multi-level Convolutional Network Design

GoogleNet's Inception architecture is widely used in the fields of computer vision such as image recognition. It performs well under strict constraints under memory and computing budget. The computational cost of Inception is much lower than VGGNet or its higher-performance successor network. This allows the Inception network to be used in big data scenarios, say, when a large amount of data needs to be processed at a reasonable cost or when memory or computing power is inherently limited, such as in image classification, image retrieval, and image generation scene. Before GoogLeNet was

invented, the mainstream of network structure direction was deeper and wider network but there were some disadvantages such as leading to overfitting, great computational complexity, and gratitude vanishing.

The Inception model improved basic feature extraction units to achieve better feature extraction performance. The Inception structure proposes the famous Batch Normalization (BN) method. BN is a very effective regularization method, speeding up the training of large convolutional networks a lot and increasing classification accuracy as well. When BN is used in a certain layer of a neural network, it normalizes the internal mini-batch data, so that the output is normalized to the normal distribution of $N(0,1)$, and the internal covariate shift is reduced. During the training of traditional deep neural networks, the changing of distribution of each layer's input makes training difficult. We can only use a small learning rate to solve this problem before utilizing BN. The number of iterations required to achieve the performance without BN is only 1/14, meanwhile, the training time is greatly shortened. Furthermore, we can abandon the dropout setting if BN is utilized.

BN enables the Inception structure to achieve faster speed and higher accuracy during training. There is a problem with the Inception module structure, that is, if the output of the previous layer directly inputs to the convolutional layer of the next layer without any processing, the number of feature maps of the previous layer will be large. Targeting at solving this problem, the Inception V3 network has two main modification based on the conventional inception structure:

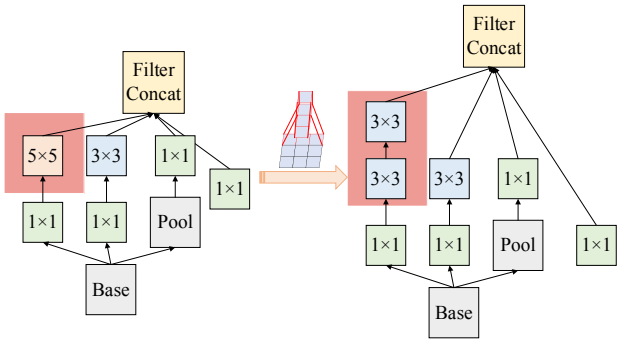


Fig. 2. The inception module convolution resolution

- Inception v3 also proposes a decomposition method. As can be seen from Fig.2, a 3x3 convolution can be a 1x3 convolution stacked with 3x1 convolution to obtain the same output. Similar operations could be applied to 5x5 and 7x7, significantly reducing the convolution kernel parameters and the amount of calculation.
- The appendix classifier uses auxiliary classifiers, adding a classifier in the middle layer of the network, calculating auxiliary distortion and gradient, and then adding gradient of this layer to the gradient of the conventional network, which effectively solves the phenomenon of gradient disappearance.

Due to the superiority of GoogleNet Inception v3 over the previous generations and its outstanding performance in computing performance and efficiency, the classification model of this article will be based on its design.

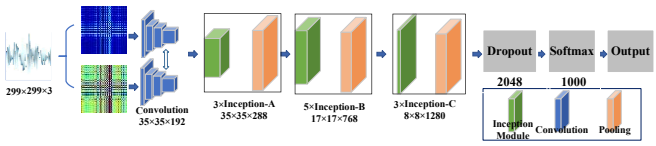


Fig. 3. Overall network structure

TABLE I. INCEPTION V3 NETWORK STRUCTURE PARAMETERS

Type	Batch size/stride Or remarks	Input size
Conv	3×3/2	299×299×3
Conv	3×3/1	149×149×32
Conv padded	3×3/1	147×147×32
Pool	3×3/2	147×147×64
Conv	3×3/1	73×73×64
Conv	3×3/2	71×71×80
Conv	3×3/1	35×35×192
3×Inception	Defined by parameters	35×35×288
5×Inception	Defined by parameters	17×17×768
3×Inception	Defined by parameters	8×8×1280
Pool	8×8	8×8×2048
Linear	Logits	1×1×2048
Softmax	Classifier	1×1×1000

Fig.3 shows the structure of the network, and Table 1 defines the parameters. Define the network structure layer by layer according to the network parameters and network structure diagram.

III. EXPERIMENT

A. Data Set Description

The actually measured data set is divided into 4 categories: bush, car, grass, and ground echo signals propagated from the environments amplified to 10dBm by narrow pulses generated by an Arbitrary Waveform Generator (AWG) with a pulse width of 2-3ns.

For each collected sample, there are many pulse-echo signals needed to be normalized, filtered and denoised, and sliced to obtain standardized samples before subsequent classification work. After observing the spectrograms of the four types of targets, it is found that the spectrum distributions of different targets vary greatly. Dissimilar spectrum distributions also represent various characteristics of different types of targets. Based on this, a low-pass filter is applied according to the target scene spectrum for divergent distributions, setting different cut-off frequencies to filter and denoise the signal. Since the echo signal has a broadening effect during the receiving process, 6ns of data are retained as samples when the echo signal slice is intercepted. After a series of preprocessing, a total of 4634 samples are obtained.

The actually measured samples have the following characteristics: (1) In the actual scene, there is a lot of random clutter in the ground measured echo signals, and the signal-to-noise ratio fluctuates greatly over time; (2) The distance between classes is small since different types UWBR echo signals of the

targets have great similarities in the time-domain waveforms; (3) The distance inside the class is large because the echo signals of the same scene or target have a large spread of time-domain characteristics. Therefore, if the traditional feature extraction and classifier design methods are used to recognize the above data sets, it will be difficult to achieve desired results. The experiment results on based the proposed GAF and multi-level CNN is presented in the following subsets. Table 2 and Table 3 define experimental parameters and configuration.

TABLE II. HARDWARE AND SOFTWARE CONFIGURATION OF THE EXPERIMENT

Equipment	Configuration
CPU	Intel(R) Xeon(R) CPU E5-2620
GPU	NVIDIA TITAN RTX
Operating system	Windows 10 Pro
Program language	Python 3.7.9
Deep learning framework	Pytorch 1.6.0
IDE	Vscode

TABLE III. EXPERIMENTAL PARAMETER SETTING

Parameter	Experimental parameter setting
Dataset partitioning	Training set 70%,validation set 20%,test set10%
loss function	Mean Square Error(MSE)
optimizer	Adam（β_1=0.9，β_2=0.999）
Learning rate	0.0003
Batch size	32

B. *Experimental Results and Analysis*

TABLE IV. RECOGNITION RESULTS BASED ON DUAL-MODE GAF FEATURES

Epochs	50				200			
GASF+GADF	Bush	Car	Grass	Ground	Bush	Car	Grass	Ground
Recall	93.06%	93.5%	98.15%	99.07%	92.99%	94.9%	97.45%	96.82%
Precision	87.39%	90.58%	94.22%	94.27%	96.05%	98.68%	99.35%	99.35%
Accuracy	94.08%				95.24%			
Epochs	500				1000			
GASF+GADF	Bush	Car	Grass	Ground	Bush	Car	Grass	Ground
Recall	92.5%	93.2%	96.07%	95%	97.1%	98.9%	98.91%	98.91%
Precision	94.53%	95.6%	98.53%	97.79%	98.17%	96.81%	98.56%	98.56%
Accuracy	**97.63%**				97.42%			

TABLE V. RECOGNITION RESULTS BASED ON ORIGINAL 2D IMAGES

Epochs	50				200			
2D HRRP	Bush	Car	Grass	Ground	Bush	Car	Grass	Ground
Recall	91.67%	87.96%	93.52%	93.46%	78.21%	97.44%	97.44%	97.47%
Precision	63.06%	92.23%	85.59%	92.59%	89.71%	91.57%	1	98.72%
Accuracy	78.45%				94.83%			
Epochs	500				1000			
2D HRRP	Bush	Car	Grass	Ground	Bush	Car	Grass	Ground
Recall	62.86%	94.3%	92.14%	96.4%	84.06%	99.28%	95.65%	96.4%
Precision	88.89%	97.1%	95.56%	95.7%	82.86%	96.48%	97.78%	97.1%
Accuracy	94.4%				**95.9%**			

Table 4 and Table 5 are the experimental results of inputting both the GAF feature map and image of the original time-domain echo waveform with the Inception V3 model. We selected the maximum number of iterations as 50, 200, 500, and 1000 respectively to compare and analyze their performance via the following evaluation standards :

The commonly used evaluation measures include accuracy, precision,and recall.

The accuracy can be expressed as

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \tag{5}$$

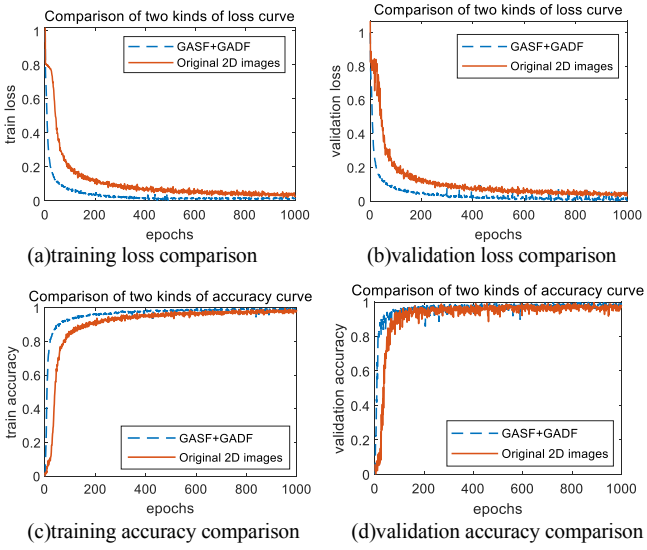
The precision can be expressed as

$$precision = \frac{TP}{TP + FP}. \tag{6}$$

The recall can be expressed as :

$$recall = \frac{TP}{TP + FN}. \tag{7}$$

Different iteration times influence the gradient descent direction and the update of network model parameters. It can be seen from Table 4 that the recognition effect achieved by using GAF features as input samples combining with the CNN network is satisfying with the accuracy rate can be maintained above 94.08%. Compared on the accuracy, the network was convergence at iterations 500. However, the recognition rate of the network using the original 2D time-domain echo image as input can only reach 78.45% at iterations 50, indicating that the network is under-fitting and the convergence process is slow. Furthermore, the accuracy just reached 95.9% at 1000 iterations. Considering both the recognition accuracy and the network convergence speed, the methods proposed in this article had better results. From multiple indicators like recall and precision, it can be found that the use of GAF features has a positive effect whatever the complexity of recognition models.



(a)training loss comparison (b)validation loss comparison

(c)training accuracy comparison (d)validation accuracy comparison

Fig. 4. Loss and accuracy comparison

Comparing the loss function and accuracy of the two input methods under the optimal model, the results are as follows:

Fig.4 shows the training loss, validation loss, training accuracy, validation accuracy varies with the increase of the iteration number. At the beginning of training, since the network

is under-fitting, the training loss and validation loss are higher. As the iterations increase, the training loss continues to decrease while validation loss occasionally fluctuates. However, after several iterations, the training loss and validation loss tend to stabilize and gradually converge. It can be seen from Fig.4 that in the initial stage of network training, because the gradient descent may fall on the local minimum of the non-optimal solution, validation accuracy fluctuates, but after the iterations increase, both training accuracy and validation accuracy tend to be stable and better. Compared with the results based on dual-mode GAF features or the 2D HRRP image as input, it can be found that the network trained using GASF+GADF features as input shows greater results both on the training set and the validation set.

TABLE VI. COMPARISON OF ACCURACY OF DIFFERENT METHODS

Methods	1D-SVM	1D-BiGRU	inception V3		
			2D HRRP	GASF	GASF+GADF
Accuracy	76.62%	95.68%	95.90%	96.60%	97.63%

Experiments under the same experimental conditions, Table 6 compared the proposed method with the traditional SVM method, the Inception V3 network recognition method based on 2D time-domain waveform images, and the Inception V3 network recognition method based on GAF feature map and methods based on 1D GRU network. Among them, the 1D Bidirectional-GRU (BiGRU) network adopts a standard structure, the batch size is 32, the maximum number of iterations is 5000 and the learning rate is 0.0001. It can be seen that the recognition accuracy with 97.63% can be achieved based on the proposed GASF+ GADF-Inception V3 method, which is higher than that of the SVM method by 21.01%, the 1D Bi-GRU method by 1.95%, the Inception V3 with the 2D-HRRP image by 1.73%, and the Inception V3 with the 2D-HRRP image by 1.03%, respectively.

According to the above experiments, we can conclude: (1) GAF features can well characterize the dependence and correlation of the time series information of UWBR targets. (2) The use of the Inception V3 network has the characteristics of a multi-level convolution structure. With the usage of the BN method and convolution splitting method, which greatly reduces the amounts of parameters, significantly improved the calculation efficiency, in the meantime, effectively alleviated the over-fitting situation.

IV. CONCLUSION

In this paper, a novel CF-UWBR target recognition method based on the dual-mode GAF features extraction and the Inception V3 network is proposed. Experiment results indicate that the proposed method has excellent feasibility and reliability in the recognition of UWBR targets. However, the method proposed in this paper still needs to be further studied and improved, including (1) For GAF features, this article only considers GASF features and GADF features, labeled separately as samples but ignoring feature fusion. (2) The Inception network structure can be optimized based on the GAFs features of the CF-UWBR HRRP on more suitable for CF-UWBR. In addition, the data set covers the single target environment and only four categories are considered. The number of classes can be further enriched.

ACKNOWLEDGMENT

We are grateful to the anonymous reviewers for their valuable comments and suggestions. This work is supported by the National Natural Science Foundation of China (NSFC. 62071293), the Equipment pre-research area fund project (61404130221), and the key laboratory fund project (61421060105).

REFERENCES

[1] Yan H Q, Zhang Z H, Xiong G, Yu W X. “Radar HRRP recognition based on sparse denoising autoencoder and multi-layer perceptron deep model”,2016 Fourth International Conference on Ubiquitous Positioning, Indoor Navigation and Location Based Services (UPINLBS), 2016.11.

[2] Xiong G, Xi Y, Chen D and Yu W, "Dual-Polarization SAR Ship Target Recognition Based on Mini Hourglass Region Extraction and Dual-Channel Efficient Fusion Network”, IEEE Access, vol. 9, pp. 29078-29089, 2021.

[3] Meng X C, Jiang T G, Da W B,” Design and implementation of radar RCS target recognition system based on decision tree”.2019, pp:40-64.

[4] Hu H P, Li Y Y, Bai Y P,” Synthetic Aperture Radar Target Recognition Based on Convolutional Neural Network and Random Forest”,2019.

[5] Jiang L B, Zhou X L, Che L,” Small sample human motion recognition based on carrier-free UWB radar”,2020.

[6] Wang J,Zheng T,Lei P,et al.” A review of depth learning in radar”[J]. Journal of radar,2018,7(4):395-411.

[7] Xiang Q, Wang X D, Li R, et al. ”HRRP image recognition of ballistic target based on DCNN”,2020.

[8] Xu B,Chen B,Liu H W,et al.” Radar high-resolution resolution and distance image identification based on attention loop neural network model”[J]. Journal of Electronics and Information Technology, 2016,38(12):2988-2995.

[9] Wang X,Zhou Y P,Zhou D Q,et al.” Low intercept probability radar signal recognition based on deep confidence network and bispectral diagonal slice”[J]. Journal of Electronics and Information Technology, 2016,38(11):2972-2976.

[10] Wang Z and Oates T, “Imaging time-series to improve classification and imputation," in Proc. 24th Int. Joint Conf. Artif. Intell. (IJCAI), Buenos Aires, Argentina, Jul. 2015, pp. 3939-3945.

[11] Zou Q, Ni L, Zhang T, et al. ”Deep learning based feature selection for remote sensing scene classification”[J]. IEEE Geoscience and Remote Sensing Letters,2015,12(11) : 2321 -2325.

[12] Dang Y, Zhang J X, Deng K Z, et al. “Classification and evaluation of surface coverage of remote sensing images based on deep learning Alex Net”[J]. Journal of Geoinformatics,2017,19(11) : 1530 -1537.

[13] Krizhevsky A, Sutskever I, Hinton G E. “Image Net classification with deep convolutional neural networks”[C].Advances in Neural Information Processing Systems,2012: 1097 -1105.

[14] Simonyan K, Zisserman A. “Very deep convolutional networks for large scale image recognition”[EB/OL].2019: 1409- 1556.

[15] Szegedy C, Liu W, Jia Y, et al. “Going deeper with convolutions”[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE,2015: 1-9.

[16] He K, Zhang X, Ren S, et al. “Deep residual learning for image recognition”[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.IEEE,2016: 770-778.

[17] Szegedy C, Liu W, Jia Y Q, et al. “Going deeper with convolutions”[C]. Conference on Computer Vision and Pattern Recognition.2015: 1- 9.