IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 29, 2019, accepted March 8, 2019, date of publication April 4, 2019, date of current version April 29, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2909348

Radar Target Recognition Based on Feature Pyramid Fusion Lightweight CNN

CHEN GUO, HAIPENG WANG, TAO JIAN, YOU HE, AND XIAOHAN ZHANG
Institute of Information Fusion, Naval Aviation University, Yantai 264001, China

Corresponding author: Haipeng Wang (whp5691@163.com)

This work was supported by the National Natural Science Foundation of China under Grant 61471379, Grant 61790551, Grant 61531020, and Grant 61671463.

ABSTRACT In order to improve the accuracy and robustness of radar target recognition under low SNR conditions, a novel radar high range resolution profile (HRRP) target recognition method based on feature pyramid fusion lightweight CNN is proposed in this paper. The proposed method combines the multi-scale space theory with a deep convolutional neural network. Because of the local connection characteristic of convolutional kernel, feature extracted by CNN mainly focus on the local information of the target. To make full use of both the local and global information in the target HRRP, multi-scale representation of the HRRP with different Gaussian kernels is introduced to construct the multi-channel input of the model. The generalization performance is improved by reducing the parameters of the proposed model with a depthwise separable convolution feature extraction block. Simultaneously, feature pyramid fusion is adapted to take full advantage of the features extracted by each block, which effectively improves the stability of the model and the training efficiency. The experimental results show that the multi-scale representation of the HRRP contributes to robust feature extraction. Meanwhile, the proposed feature pyramid fusion lightweight CNN can effectively prevent over-fitting and improve the stability of the model.

INDEX TERMS Radar target recognition, multi-scale representation, lightweight CNN, feature pyramid fusion, noise robust.

I. INTRODUCTION

In practical applications, the training High Range Resolution Profiles (HRRP) of the radar target are mainly the high signal-to-noise ratio (SNR) data obtained under cooperative conditions or generated by electromagnetic simulation software. However, the targets to be recognized are mostly non-cooperative, and the SNR of HRRP is usually low due to the influence of the distance between the radar and the target and strong background noise. Under low SNR conditions, the target HRRP is strongly disturbed by noise, resulting in a significant reduction in target recognition accuracy. Therefore, it is of great significance to study the robust feature extraction method of target HRRP under low SNR conditions. In response to the above problems, many scholars have conducted extensive research. In [1], multi-task factor analysis method is applied to extract the robust features of HRRP. The method proposed in [2] modifies the Gaussian model of the training data according to the SNR of the testing samples, and solves the mismatch problem between the training samples and the noisy testing samples. In [3], the OMP algorithm is used to extract the strong scattering points of the target, and the Hausdorff distance is used to perform the scattering point matching. In [4], sparse representation of the target HRRP is performed by the K-SVD algorithm, thereby improving the robustness of recognition accuracy. A stable dictionary learning method is proposed in [5] with structured sparse regularization and robust loss function via marginalizing dropout to extract the robust features of the target HRRP. In [6], the robustness of the features is improved by fusing the sparse features of the target multi-scale space. The features extracted by these methods are artificially designed, which cannot accurately represent the complete information of the target and calls for certain prior knowledge. In the practical applications, however, the prior information is variable, or even unknown. More importantly, under low SNR conditions, these shallow methods can hardly extract the essential features of the targets. Therefore, how

The associate editor coordinating the review of this manuscript and approving it for publication was Hasan S. Mir.

to automatically extract the robust features of the target is a difficult problem in radar target recognition.

In recent years, deep learning algorithms have been widely used in computer vision, such as target detection [7], [8] and target classification [9]–[13]. The features usually have good robustness because of the data-driven and automatic extraction properties of deep learning algorithms. HRRP contains massive information of the target such as the structure and intensity of the scattering point et al. At present, some deep models have been applied to the field of radar target recognition. The algorithm proposed in [14] adopts the HRRP average profile to construct the robust features of the objective function, and Stacked Auto-Encoder (SAE) is applied to extract the robust features of the target. Based on [14], a radar target recognition method based on Stacked Corrective Auto-Encoder is proposed in [15]. This method employs the Mahalanobis distance criterion and the average HRRP in each frame to construct the objective function to further improve the robustness of the features. In [16], Sparse Denoising Auto-Encoder and Multi-layer Perceptron (SDAE&MLP) is combined to recognize the target HRRP. The model has better de-noising performance because of the noise added during training process. In [17], robust features of the target is extracted by Robust Variational Auto-Encoder which adopts the HRPR average profile to constrain the features. In [18], an adaptive feature learning model is proposed for HRRP sequences, which can adaptively extract the features of HRRP sequences based on the sequence length and the complexity of a single HRRP. The deep learning algorithm currently applied to radar HRRP target recognition technology is mainly based on Stacked Auto-Encoder. Stacked Auto-Encoder is an unsupervised feature extraction method, which is unable to fully utilize label information of the targets. It adopts greedy layer-wise training method, and the extracted features are prone to fail as the number of layers increases. In order to solve the above problems, a supervised target recognition method based on CNN is proposed.

The traditional CNNs only adopt the output features of the deepest convolutional layer for target classification and recognition. The output of each layer in the deep network is the target feature. The shallow features mainly include information such as contours and edges, whereas, the deep features are mostly advanced semantic information. Additionally, feature extracted by CNN mainly focus on the local information of the target because of the local connection characteristic of convolutional kernel. In order to make full use of the features extracted by each layer as well as the global and local information of the target, a radar target recognition method based on Feature Pyramid Fusion Lightweight CNN (FPFL-CNN) is proposed by applying the concept of Scale Invariant Feature Transform (SIFT) [19], [20] on the traditional deep CNN model. The multi-scale representation of HRRP is employed to construct multi-channel input of the model, then features of the target is extracted by the lightweight CNN with the depthwise separable convolution BLOCK. Finally the features of each BLOCK are fused to obtain the target recognition result. The proposed method has the following characteristics:

1. The proposed model is a supervised learning model driven by data, which can automatically extract the deep features of the target after training process.
2. The multi-scale representation of HRRP is adopted to form the multi-channel input of the proposed model, so as to extract the robust features which can take both the detail and structural information of the target into consideration.
3. Compared with the traditional CNN, a lightweight CNN is designed based on the depthwise separable convolutional layer and performs great advantage on the computational efficiency and generalization.
4. The proposed model makes full use of the features of each layer and improves the robustness and convergence speed of the proposed model via feature pyramid fusion.

The rest of this paper is organized as follows: Section II gives a specific description of the proposed method. In the Section III, the dataset and experimental steps are described in detail. The effectiveness and superiority of the proposed method are verified by simulations in the Section IV. At last, Section V gives a summary on the proposed method.

II. FEATURE PYRAMID FUSION LIGHTWEIGHT CNN

The proposed method firstly utilizes the Gaussian kernel function to obtain the multi-scale representation of HRRP as the input of the model, and then employs the feature pyramid lightweight CNN model to extract the features of the target. The proposed CNN model is mainly composed of four separable convolution feature extraction BLOCKs, which are numbered as BLOCK 1, 2, 3 and 4. The outputs of the first three BLOCKs are used as inputs to BRANCH 1, 2, and 3, respectively, and each BRANCH adopts a depthwise convolution to downsample the input vector. The output of BRANCH 1, 2, and 3 and the output of BLOCK 4 are concatenated to form a new feature vector, each channel of the feature vector is fused by pointwise convolution, and the obtained feature vector is expanded into a one-dimensional vector to connect with the fully connected layer. Finally, the recognition result is achieved by the output layer. The overall diagram of the proposed model is shown in Fig. 1, where, DC represents depthwise convolution, PC represents pointwise convolution, P represents pooling layer with stride 2, FC represents the fully connected layer and O represents output layer. The fully connected layer and output layer adopted as classifier are traditional neural networks. The proposed model is end-to-end, that is, the recognition result can be directly obtained by feeding the data into the trained model.

A detailed description and analysis of the proposed method will be presented as the following aspects: 1) Multi-scale representation of HRRP data, 2) Construction of separable depthwise convolution feature extraction BLOCK, 3) Specific method of feature pyramid fusion, 4) The computational

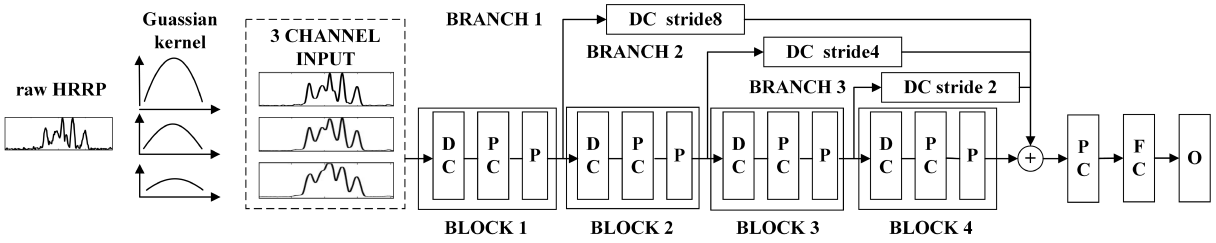


FIGURE 1. The structure of the FPFL-CNN.

complexity of the proposed model and other deep models.

A. CONSTRUCTION OF MULTI-CHANNEL INPUT BASED ON MULTI-SCALE REPRESENTATION OF HRRP

According to scale-space theory, the multi-scale representation of the signal can be derived by convoluting with Gaussian kernel of different parameters [21]. Small scale signals are prominent in detail and local information, while structural and global features will be extracted from large ones. Therefore, compared with single-scale signals, it is easier to obtain the essential features by comprehensively utilizing the multi-scale information of the signal. However, most of the existing researches only extract the features of the original scale HRRP, which mainly focus on the local information and has poor generalization performance. To solve the above problem, multi-scale representation of the HRRP is employed for feature extraction.

It can be known from [22] that the Gaussian kernel is the only linear kernel to realize multi-scale representation of the signal. By convoluting the signal with Gaussian kernel, the high frequency components in the signal can be filtered, some details of the signal are discarded, while the global information remains unchanged. The convolution formula of Gaussian kernel and HRRP is:

$$L(x,\sigma) = G(x,\sigma) \otimes I(x) \tag{1}$$

where, $G(x,\sigma) = a\exp(-\frac{x^2}{2\sigma^2})$ is one-dimensional Gaussian kernel function, a and σ represent the amplitude and scale of the Gaussian kernel respectively, $I(x)$ represents the input signal, and $L(x,\sigma)$ is the signal after the Gaussian convolution.

In order to make the proposed model focus more on the global information, three Gaussian kernels are selected, of which width are 3, 5, 7, and the scale σ is $\sigma_0, 2^{1/1}\sigma_0, 2^{2/1}\sigma_0$ respectively, where σ_0 equals to 1. The results of Gaussian convolution with a single HRRP are shown in Fig. 2. The red line indicates the result of Gaussian convolution and the blue line indicates the original HRRP. It can be seen from Fig. 2 that as the scale parameter σ increases, the detail information is reduced, while the structural information is better preserved. At last, the HRRPs with different scales are normalized and concatenated into a three-channel data to serve as the input of the proposed model.

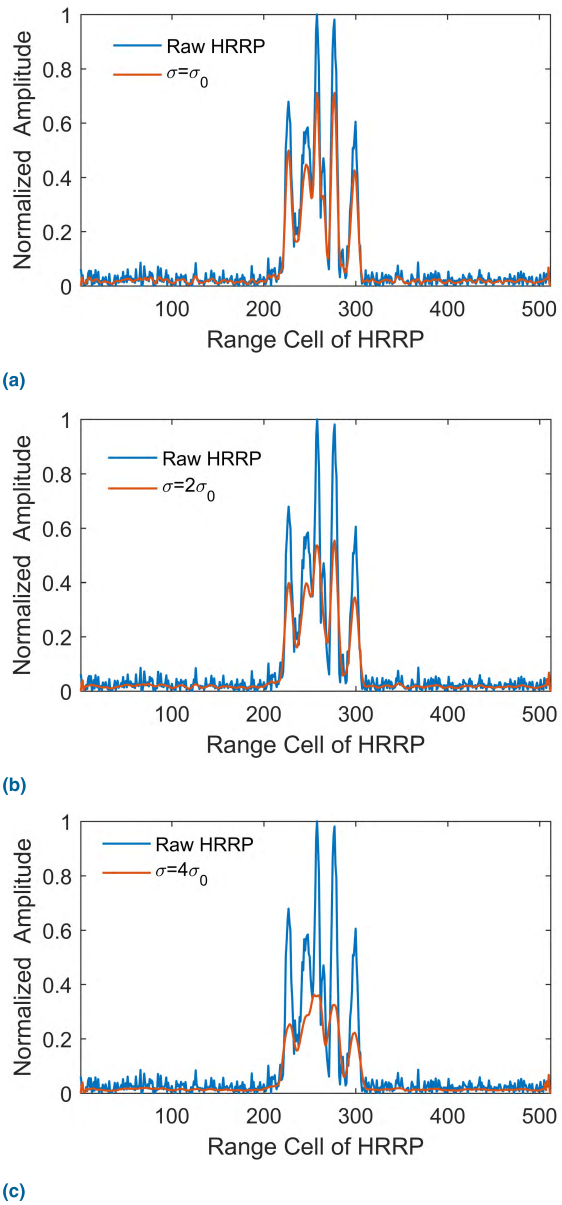


FIGURE 2. The Gaussian convolution results of the HRRP (a) $\sigma = \sigma_0$ (b) $\sigma = 2^{1/1}\sigma_0$ (c) $\sigma = 2^{2/1}\sigma_0$.

B. DEPTHWISE SEPARABLE CONVOLUTION FEATURE EXTRACTION BLOCK

As is known to all, features are extracted by convolution kernel of the convolutional layer in CNN. The convolution

kernel can be regarded as a sliding window correlating with the corresponding part of the input vector. The entire input vector is traversed according to a certain stride to obtain the output feature vector. The formula for the convolution operation is:

$$\begin{cases} \boldsymbol{x}_j^l = f(\boldsymbol{u}_j^l) \\ \boldsymbol{u}_j^l = \sum\limits_{i \in M_j} \boldsymbol{x}_i^{l-1} \otimes \boldsymbol{k}_{ij}^l + \boldsymbol{b}_j^l \end{cases} \tag{2}$$

where, $\otimes$ represents convolution operation, $\boldsymbol{u}_j^l$ and $\boldsymbol{x}_j^l$ are the raw activation and output of the jth channel in the convolutional layer l respectively. $f(\cdot)$ is the activation function and ReLU function is adopted. $\boldsymbol{k}_{ij}^l$ is the convolution kernel of the jth channel in the convolutional layer l corresponding to the ith input feature vector, $\boldsymbol{b}_j^l$ is the offset term of the jth channel in the convolutional layer l.

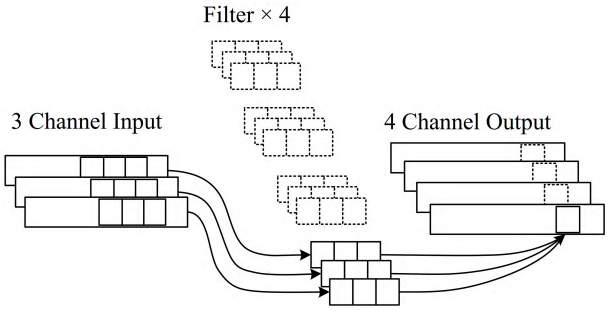


FIGURE 3. Schematic diagram of tradition convolution.

Fig. 3 gives a detailed description on the feature extraction process of convolution kernel in a single convolution layer, where the input and output are one-dimensional vector with 3 and 4 channels respectively, the size of the convolution kernel is 3×1. The number of channels in a single convolutional kernel and input vector are the same, as well as the number of convolutional kernels and the channels of the output vector. The output of the convolutional kernel is derived by adding output of each channel, which is achieved by convolving with the corresponding channel of the input vector. The output of each convolution kernel corresponds to a single channel of the output.

The number of parameters in the convolutional layer is determined by the convolutional kernel. Suppose the size of the convolution kernel is $n_k \times 1$, the number of channels of the input and output feature are n_i and n_o, respectively, then the number of the parameters is $n_k n_i n_o$.

Since the convolutional kernel has characteristics of local connection and weight sharing, the parameters of the traditional CNN are small in magnitude, but the training efficiency is low because the CNN needs to cache the feature vectors of each layer. Simplifying the structure and improving the training efficiency are problems to be solved. Therefore, the proposed model adopts a new feature extraction BLOCK, which is composed of a depthwise separable convolutional layer and a pooling layer. The depthwise separable convolutional layer is mainly responsible for feature extraction, and the pooling layer for reducing the redundancy of features.

Depthwise separable convolution decomposes a complete convolution operation into two steps, that is, depthwise convolution (DC) and pointwise convolution (PC) [23] . Depthwise convolution is responsible for extracting the features of each input channel and pointwise convolution is responsible for fusing the features of each channel. By combining the above two, it is available to decouple the spatial information and depth information of the features. Fig. 4 describes the specific operation of depthwise convolution. The input is a one-dimensional vector with 3 channels. The size of convolutional kernel is 3×1. The depthwise convolution is different from the traditional convolution, of which the number of channel in the depthwise convolutional kernel is always 1. Each convolution kernel is only convoluted with the single channel of input, so the number of channels in the output and input feature are the same.

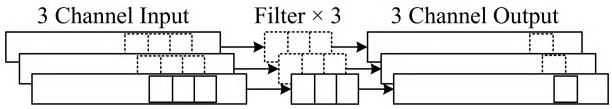


FIGURE 4. Schematic diagram of depthwise convolution.

Depthwise convolution only performs independent convolution operations on each channel of the input, and does not fuse the feature information of different channels at the same spatial position. Therefore, pointwise convolution is connected to the depthwise convolution to combine the features of each channel into new ones. Pointwise convolution is a special case of traditional convolution. The size of the convolutional kernel is fixed to 1×1. Performing pointwise convolution on multi-channel features is equivalent to obtaining a new feature vector by weighted sum of each channel.

The parameters in the depthwise separable convolutional layer is the sum of ones in depth convolution and pointwise convolution. Suppose the size of the convolution kernel (depthwise convolution) is $n_k \times 1$, and the number of channels of the input/output feature is n_i and n_o respectively, then number of parameters in depthwise separable convolution layer is $n_i n_k + n_i n_o$.

The structure of the depthwise separable convolution feature extraction BLOCK is shown in Fig. 5. It consists of a depthwise convolution layer, a pointwise convolution layer, and a pooling layer. The size of the depthwise convolution kernel is fixed to 3×1 [24]. The stride of the pooling layer is 2. After the features are down-sampled by the pooling layer, the number of channels remains unchanged, and the dimension becomes one-half of the original one.

C. DESCRIPTION OF FEATURE PYRAMID FUSION METHOD

The traditional CNN consists of convolutional layer, pooling layer, fully connected layer, etc. Traditional CNN only utilize the features extracted by the last convolutional layer for target recognition, and obtains more advanced semantic features by

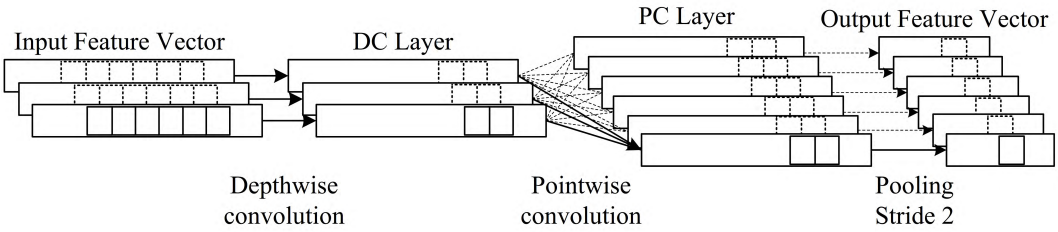


FIGURE 5. Schematic diagram of feature extraction block based on depthwise convolution.

increasing the depth of the model to improve the recognition accuracy [25]. However, the features extracted by other convolutional layers are all available and have not been fully utilized. In order to make full use of the features extracted by each layer, a feature pyramid fusion method is proposed. The schematic diagram of the feature pyramid is shown in Fig. 6.

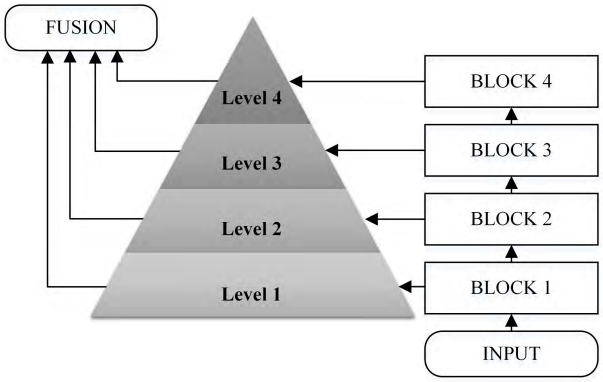


FIGURE 6. Schematic diagram of feature pyramid.

In Fig. 6, Level 1∼4 represent the output feature vectors of BLOCK 1∼4 respectively. The feature vector of Level $i+1$ is obtained by convoluting and down-sampling the feature vector of Level i with BLOCK i, and its dimension is one-half of Level i. Therefore, the feature vector combination of Level 1∼4 can be called a feature pyramid. The specific method of feature pyramid fusion is as follows:

(1) The feature vectors of the feature pyramid levels 1∼3 are down-sampled so as to match the dimension of features in Level 4. It is prone to cause partial loss of effective information by utilizing the pooling layer to downsample the shallow features. Therefore, we adopt the depthwise convolution to downsample the feature vectors of Level 1∼3, as shown in Fig. 1. The sizes of depthwise convolution kernel corresponding to BRANCH 1, 2 and 3 are 9×1, 5×1, 3×1, and the strides are 8, 4, and 2 respectively.

(2) Concatenate the features of each Level. Assume the number of channels in the Level 1∼4 is $c_1 \sim c_4$, the channel of the feature vector after concatenating is $\sum_{i=1}^{4} c_i$.

(3) The pointwise convolution is used to fuse each channel of the concatenated vector.

TABLE 1. The structure Of the proposed model.

STRUCTURE	NOF	OUTPUT
BLOCK 1, 3×1, 16	140	512×3×16
BLOCK 2, 3×1, 32	2144	256×1×32
BLOCK 3, 3×1, 32	4288	128×1×32
BLOCK 4, 3×1, 64	8384	64×1×64
BRANCH 1, 9×1, 16	288	64×1×16
BRANCH 2, 5×1, 32	320	64×1×32
BRANCH 3, 3×1, 32	192	64×1×32
PC, 1×1, 64	32768	64×1×64
POOLING, STRIDE 4	0	16×1×64
TOTAL	48524	

The final recognition result can be obtained by inputting the fusion vector into the fully connected layer.

D. COMPARISON OF VARIOUS MODEL STRUCTURES AND COMPUTATIONAL COMPLEXITY

The complexity of the model is analyzed from two aspects, one is the structural composition, and the other is the computational complexity. The traditional CNN and autoencoder model (AE) are selected as the comparison models. The structure and number of parameters in each layer of feature extraction part are shown in Tables 1-3, where, NOF represent the number of parameters and the offset of each layer is ignored.

As can be seen from Table 1-3, the numbers of parameters of the proposed model and the traditional CNN are much smaller than autoencoder model. Among them, the parameters of the proposed model is smaller than that of the traditional CNN because of the depthwise separable convolution layer.

Since the computational complexity of the model is a linear addition of the computational complexity of each layer, therefore only the computational complexity of a single layer is analyzed. Suppose that the number of training samples is N, the input and output feature dimensions of the single layer are D and T respectively, and then the single layer computational complexity of a single hidden layer in AE is $O(NDT)$, where DT is the number of parameters of the hidden layer.

For the two CNN models, suppose the size of the convolution kernel is n_k, and the number of input and output channels are n_i and n_o respectively. The computational complexity of the convolutional layer and the depthwise separable

TABLE 2. The structure of CNN.

STRUCTURE	NOF	OUTPUT
CONV, 3×1, 16	96	512×1×32
CONV, 3×1, 16	3072	512×1×32
POOLING	0	256×1×32
CONV, 3×1, 32	6144	256×1×64
CONV, 3×1, 32	12288	256×1×64
POOLING	0	128×1×64
CONV, 3×1, 64	24576	128×1×128
POOLING	0	64×1×128
CONV, 1×1, 64	16384	64×1×128
POOLING, STRIDE 4	0	16×1×128
TOTAL	62560	

TABLE 3. The structure of AE.

STRUCTURE	NOF	OUTPUT
HIDDEN LAYER I	307200	600
HIDDEN LAYER II	120000	200
HIDDEN LAYER III	20000	100
TOTAL	447200	

convolutional layer are $O(NDn_k n_i n_o)$ and $O(ND(n_k n_i + n_i n_o))$ respectively, where, the corresponding number of parameters are $n_k n_i n_o$ and $n_k n_i + n_i n_o$.

It can be seen from Tables 1∼3 that $n_k n_i n_o > n_k n_i + n_i n_o > T$ in general, and the number of neural network layers based on convolution kernels is larger than that of AE models, therefore the corresponding computational complexity is greater than that of AE model. However, the computational complexity of the proposed model is smaller than that of the traditional CNN due to the fewer parameters.

In summary, the proposed model has a small number of parameters and relatively low computational complexity, thus belongs to a lightweight CNN.

III. DATASET AND EXPERIMENTAL PROCEDURE

A. DESCRIPTION OF DATASET

Solidworks3D software is utilized to establish a 1:1 model of seven ship targets, and CST electromagnetic simulation software is employed to simulate the corresponding HRRP [26]. The structural parameters of the seven ship targets are shown in Table 4. Considering the symmetry of the ship target and the computational efficiency of the CST simulation software, the CST simulation parameters are set as follows: azimuth angle is −90∼90 degrees, pitch angle is 0 degree, angle stride is 1 degree; the center frequency of the radar is 10 GHz, the bandwidth is 100 Mhz, the polarization mode is vertical polarization, and the frequency sampling points are 256. Use the default optimal mesh analysis size of the software and select ray tracing algorithm to solve the problem. Finally, the simulation yields 181 azimuthal angle HRRP data for seven ship targets.

The target HRRP is obtained by converting the Radar Cross Section (RCS) of range cell simulated from the CST with radar equation. Since the simulation background is simple,

TABLE 4. Structure parameters of 7 kinds of ship targets.

NO.	LENGTH (M)	WIDTH (M)	DRAUGHT DEPTH (M)
SHIP 1	182.8	24.1	8.1
SHIP 2	172.8	16.8	6.5
SHIP 3	153.8	20.4	6.3
SHIP 4	135	16.8	4.5
SHIP 5	121	17.6	4.3
SHIP 6	102.2	16.5	4.2
SHIP 7	89.3	12.1	4.0

the HRRP data obtained by the simulation can be regarded as noise-free data. In order to simulate the actual scene, the noise is superimposed on the raw HRRP. Usually, the SNR of training data is high and set to 20dB. In order to test the recognition performance of the target HRRP under different SNR conditions, the SNR of the testing data is set as 0 dB, 5 dB, 10 dB, 15 dB, and 20 dB respectively. The SNR is defined as the power ratio of the signal to the noise. The formula is as follows

$$SNR = 10\log(\frac{P_s}{P_n}) \tag{3}$$

where, P_s and P_s represents the average power of the signal and noise respectively, and the unit of SNR is dB.

The proposed model is data-driven and requires a lot of training data. However, the existing data is insufficient, so it is necessary to perform data augmentation on HRRP. The specific method of data augmentation is to add the noise $n_i \sim N(0, \sigma_i^2)$ with the Gaussian distribution to the original HRRP data 20 times according to the set SNR, and the noise power σ_i^2 is calculated according to eq.(3). The number of training data available by this method is $7 \times 181 \times 20 = 25340$. Similarly, the test dataset with different SNR can be obtained. It is worth noting that the training data and the test data have no overlapping parts.

B. EXPERIMENTAL ENVIRONMENT AND PROCEDURE

Experiments are carried out in the 64-bit win7 system. The software is mainly based on deep learning architecture of Keras and python development environment Sublime Text 3. The hardware is based on Intel (R) Core (TM) i7-7700K @ 3.60GHz CPU and one NVIDIA GTX 1070 GPU, with CUDA8.0 accelerating computation. The training process of the proposed model is performed according to Table 5.

The well trained model can extract the features of the testing samples and output the corresponding recognition results automatically.

IV. SIMULATION RESULTS AND ANALYSI

This section is divided into two parts. The first part simulates and analyzes the recognition accuracy of the proposed model and the comparison model under different SNR conditions. The second part validates the effectiveness of improvement on traditional CNN model proposed by FTFL-CNN.

TABLE 5. The training process of the proposed model.

Input	Training samples
Output	FPFL-CNN
Step 1	Model Initialization, construct the proposed model according to Fig. 1 and Table 1, and initialize the parameters. The epoch is set to 300, the batch size is 200, the initial learning rate is 0.01, and the learning rate is halved for every 50 epochs.
Step 2	Forward propagation, calculate the loss function of the data during each iteration. The Large Margin Cosine Loss is adopted as the loss function [27], whose formula is $$L=-\frac{1}{m}\sum_{i=1}^{m}\log\frac{e^{s(\cos(\theta_{W_{y_i},x_i})-a)}}{e^{s(\cos(\theta_{W_{y_i},x_i})-a)}+\sum_{j\neq y_i}e^{s\cos(\theta_{W_j,x_i})}} \quad (4)$$ where m is the number of training data in a batch, x_i represents the feature of the sample i in the fully connected layer, y_i represents the label of the sample i, $W_j \in \mathrm{R}^d$ represents the jth column of the weight matrix in the fully connected layer, which is also corresponding to the jth class. n represents the total number of target categories, $\cos(\theta_{W_j,x_i})$ represents the cosine value between W_j and x_i, s and a are hyper parameters. The role of s is to ensure the convergence of the loss function and a is the margin to increase the distance between two classes.
Step 3	Back propagation, use the chain rule[28] to calculate the gradient, and update the parameters with stochastic gradient descent method .
Step 4	Repeat **Step** 2 and **Step** 3 until the loss converges and no longer falls. End the training process and use the test data to verify the validity of the model.

A. RECOGNITION ACCURACY OF PROPOSED MODEL AND COMPARISON MODEL UNDER DIFFERENT SNR CONDITIONS

In this section, six comparison models are selected, which are three classical deep learning networks Convolutional Neural Network (CNN), Sparse Auto-Encoder (SAE), Denoising Auto-Encoder (DAE) and three classical shallow algorithms K-SVD, LSVM and PCA. Among the deep models, the structure of CNN is shown in Table 2, meanwhile the structure of SAE and DAE is listed in Table 3. The number of dictionary atoms and sparsity coefficient of are 600 and 100, respectively. LSVM directly classifies the original data. PCA reduces the HRRP dimension to 100 and then classifies it with LSVM. Under different conditions of SNR, simulation experiments are carried out on each model. The recognition accuracy is shown in Fig. 7.

As shown in Fig. 7 that the recognition accuracy of the proposed model is higher than all the other models under different SNR conditions. Among the shallow algorithms, K-SVD has the highest average recognition accuracy and the smallest slope under different SNR conditions, which indicates that the extracted features have better robustness. The SNR has a great influence on the recognition accuracy of LSVM and PCA. Under the condition of high SNR, both LSVM and PCA can have good performance, but under the low SNR condition, the recognition accuracy of the two has a drastic decreasing.

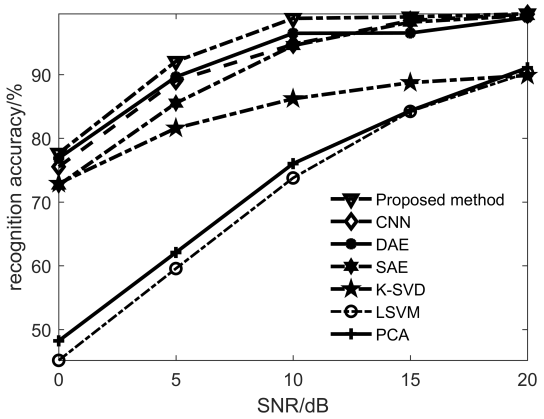


FIGURE 7. Recognition Accuracy of Different Models under Different SNR Conditions.

The recognition accuracy of deep model under different SNR conditions is better than that of the shallow models. The average recognition accuracy of SAE is the lowest but still about 6% higher than that of K-SVD, which indicates that the features extracted by deep models have good robustness and can better express the essential features of the target. Among them, the recognition accuracy of DAE under low SNR is higher than SAE and CNN, mainly because adding the random noise to the input during training can improve the robustness of the extracted features.

B. ANALYSIS ON THE ADVANTAGES OF THE PROPOSED MODEL

Compared with CNN, the proposed model has mainly improved in three aspects. In order to verify the effectiveness of these three aspects, the univariate experimental method is adopted. Influence of the three aspects on the model is simulated respectively based on the proposed model.

1) ANALYSIS OF RECOGNITION ACCURACY UNDER DIFFERENT CHANNEL COMBINATIONS

In order to verify the influence of different multi-scale representation combinations on the robustness of recognition accuracy, the original HRRP data and multi-scale representation

TABLE 6. Recognition accuracy of six channel combinations under different SNR conditions.

SNR	5DB		0DB	
	ACC /%	SD /%	ACC /%	SD /%
C 1	89.53	1.3	74.83	2.45
C 2	90.63	1.05	76.54	1.36
C 3	90.35	0.77	74.99	2.05
C 4	91.02	1.09	75.38	2.15
C 234	**92.06**	**0.41**	**77.65**	**0.88**
C 1234	91.65	0.64	76.95	1.46

TABLE 7. Recognition results of the proposed model and CNN with branches under different SNR conditions.

ACC /%	0DB	5 DB	10 DB	15 DB	20 DB
FPFL-CNN	77.65	92.06	98.80	99.03	99.52
FP CNN	78.17	91.54	98.34	98.77	99.45

TABLE 8. Recognition accuracy of models with different pyramid levels.

MODEL	FPFL-CNN	B12	B 13	B 23	B 1	B 2	B 3	NB
ACC /%	92.06	85.78	85.71	84.55	85.19	83.27	83.09	76.57
SD /%	0.43	0.69	1.22	3.1	1.37	3.37	5.59	10.51

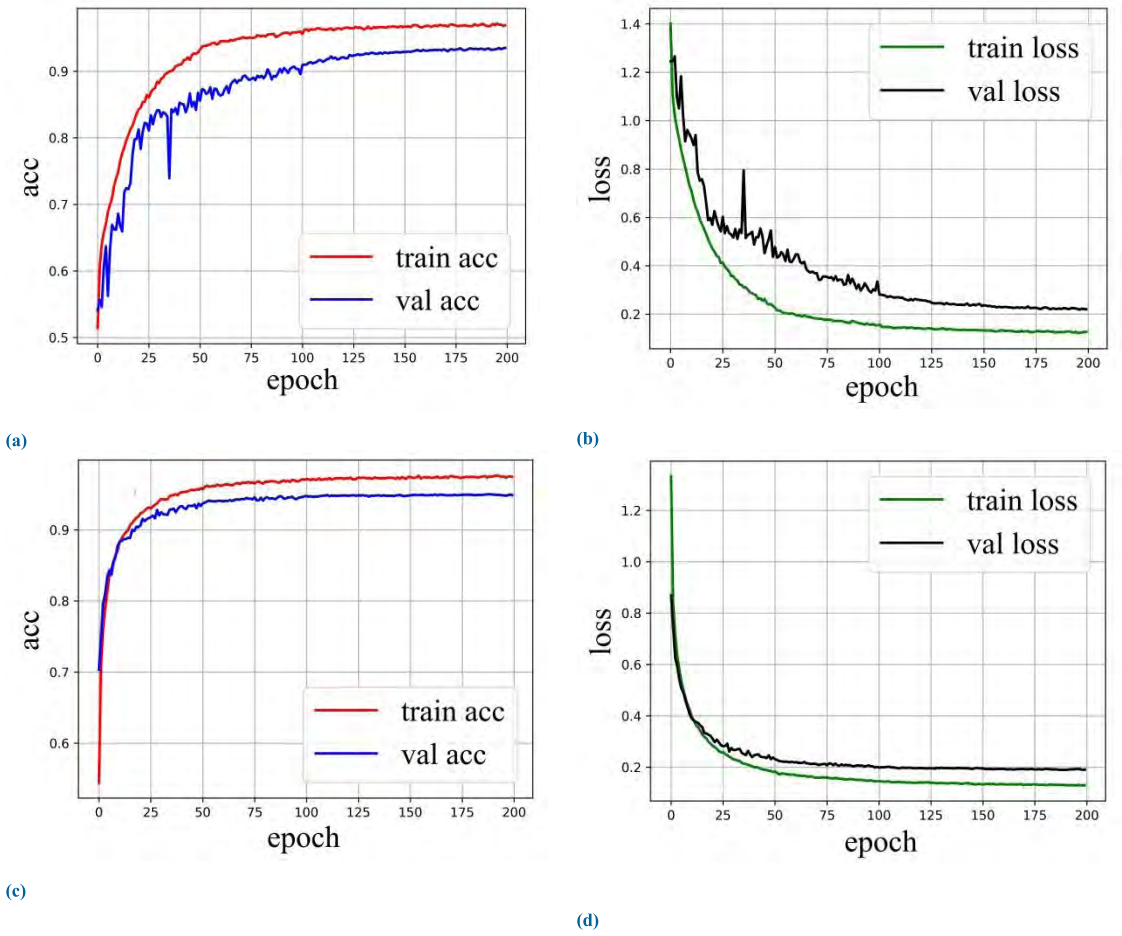


FIGURE 8. The accuracy curve and loss curve of the FPFL-CNN and model with no branches(a) The accuracy curve of the model with no branches (b) The loss curve of the model with no branches (c) The accuracy curve of the proposed method (d) The loss curve of the proposed method.

of HRRP obtained from Section II, Part A numbered as Channels 1∼4 are arranged into six combinations, which are Channel 1 only (C 1), Channel 2 only (C 2), Channel 3 only (C 3), Channel 4 only (C 4), Channel 234 (C 234) and Channel 1234 (C 1234). Since the recognition performance is already very good at SNR = 10, only the case of SNR = 0 and SNR = 5 are considered in this section. To validate the stability of the recognition results, 50 simulation experiments are performed for each combination. The experimental results are shown in Table 6. The average recognition accuracy is represented by ACC, and the standard deviation of the recognition accuracy is represented by SD.

It can be seen from Table 6 that under the two SNR conditions, the average recognition accuracy of C 234 is the highest, and the standard deviation of accuracy is the smallest, indicating that the multi-scale representation of HRRP is beneficial for the proposed model to extract robust features and improve recognition accuracy.

2) THE DIFFERENCE BETWEEN THE DEPTHWISE SEPARABLE CONVOLUTIONAL LAYER AND THE TRADITIONAL CONVOLUTIONAL LAYER

In order to discover the difference between the depthwise separable convolutional layer and the traditional convolutional layer, Feature pyramid CNN (FP CNN) is adopted as the comparison model by replacing the depthwise separable convolutional layer of the proposed model with the traditional convolutional layer. The experimental results of the two under different SNR conditions are shown in Table 7.

It can be seen from Table 7 that the average recognition accuracy of the FPFL-CNN is slightly larger than that of the FP CNN. The depthwise separable convolution can

greatly reduce the parameters of the feature extraction layer, while decoupling the spatial information and depth information in the feature. Especially under the premise of insufficient data, FPFL-CNN is the better choice for preventing over-fitting.

3) THE INFLUENCE OF FEATURE PYRAMID FUSION ON RECOGNITION ACCURACY

In order to verify the influence of feature pyramid fusion from different levels on the recognition accuracy, seven combinations are selected to form the comparison model. They are the six models obtained by the permutation and combination of BRANCH 1, 2, and 3, and the model with no branch (NB). The experimental results are shown in Table 8, where, B represents BRANCH. The SNR of the testing data is 5dB.

It can be seen from Table 8 that the recognition performance of the proposed model is the best, with the highest recognition accuracy and the smallest standard deviation, whereas the model with no branches has the lowest recognition accuracy and the largest standard deviation due to the occasional non-convergence of loss function. Among the three models with only a single branch, the model with BRANCH 1 has the highest recognition accuracy and the smallest standard deviation, and the model with BRANCH 3 has the lowest recognition accuracy and the largest standard deviation. Among the three models with two branches, the model containing BRANCH 1 and 2 has the highest recognition accuracy and the smallest standard deviation, and the model containing BRANCH 2 and 3 has the lowest recognition accuracy and the largest standard deviation. In general, the average recognition accuracy of the models with two branches is better than the model with a single branch, and the performance of the model containing the BRANCH 1 is better than those without the BRANCH 1 in the aspect of average recognition accuracy, standard deviation and stability. It can be summarized from the above experimental results that fusion of the features from each layer can improve the recognition accuracy and robustness of the model, and the shallow features may have a greater impact on recognition accuracy.

In order to compare and analyze the convergence speed of the model with no branches and the proposed model, the recognition accuracy curve and loss curve in the training process are shown in Fig. 8, where acc means recognition accuracy.

It can be seen from Fig. 8 (a) and Fig. 8 (b) that for the model with no branches, the accuracy curve and the loss curve fluctuate sharply in the first 100 epochs, and slowly converge to a stable value between 100 and 200 epochs.

As shown in Fig. 8 (c) and Fig. 8 (d) that for the proposed model, the accuracy curve (the loss curve) rises (drops) rapidly in the first 50 epochs with little fluctuation and both quickly converge to a stable value from epoch 50 to epoch 100. Compared with the model with no branches, the convergence speed of the proposed model is greatly improved.

V. CONCLUSION

A radar target recognition method based on feature pyramid fusion lightweight CNN is proposed in this paper. Compared with traditional CNN, the proposed model has mainly improved in three aspects. Learn from the concept of SIFT method, multi-scale representation of HRRP is adopted as the model input. In order to reduce the parameters as well as the risk of over-fitting while ensuring the recognition accuracy, the traditional convolutional layer is replaced by the depthwise separable convolutional layer. Finally, feature pyramid fusion is performed to fully utilize the features extracted by each depthwise separable convolutional feature extraction BLOCK. The experimental results show that multi-scale representation of HRRP contributes to better extraction of local and global information of the target and improve the robustness of the features. Feature pyramid fusion can take full advantage of the feature extracted by each BLOCK, improve the stability of the proposed model, and boost the convergence speed.

REFERENCES

[1] L. Du, H. Liu, P. Wang, B. Feng, M. Pan, and Z. Bao, "Noise robust radar HRRP target recognition based on multitask factor analysis with small training data size," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3546–3559, Jul. 2012.

[2] M. Pan, L. Du, P. Wang, H. Liu, and Z. Bao, "Noise-robust modification method for Gaussian-based models with application to radar HRRP recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 558–562, May 2013.

[3] L. Du, H. He, L. Zhao, and P. Wang, "Noise robust radar HRRP target recognition based on scatterer matching algorithm," *IEEE Sensors J.*, vol. 16, no. 6, pp. 1743–1753, Mar. 2016.

[4] B. Feng, L. Du, H.-W. Liu, and F. Li, "Radar HRRP target recognition based on K-SVD algorithm," in *Proc. IEEE CIE Int. Conf. Radar*, Oct. 2011, pp. 642–645.

[5] H. W. Liu, B. Feng, B. Chen, and L. Du, "Radar high-resolution range profiles target recognition based on stable dictionary learning," *IET Radar, Sonar Navigat.*, vol. 10, no. 2, pp. 228–237, 2016.

[6] W. Dai, G. Zhang, and Y. Zhang, "HRRP classification based on multi-scale fusion sparsity preserving projections," *Electron. Lett.*, vol. 53, no. 11, pp. 748–750, 2017.

[7] R. Girshick. (2015). "Fast R-CNN." [Online]. Available: https://arxiv.org/abs/1504.08083

[8] S. Ren, K. He, R. Girshick, and S. Jian, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.

[10] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556

[11] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi. (2016). "Inception-v4, inception-ResNet and the impact of residual connections on learning." [Online]. Available: https://arxiv.org/abs/1602.07261

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

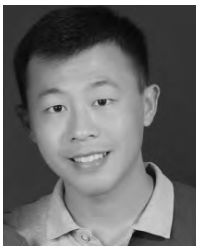
[13] G. Huang, Z. Liu, L. V. der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2016, pp. 2261–2269.

[14] B. Feng, B. Chen, and P. H. Wang, "Feature extraction method for radar high resolution range profile targets based on robust deep networks," *J. Electron. Inf. Technol.*, vol. 36, no. 12, pp. 2949–2955, 2014.

[15] B. Feng, B. Chen, and H. Liu, "Radar HRRP target recognition with deep networks," *Pattern Recognit.*, vol. 61, pp. 379–393, 2017.

[16] H. Yan, Z. Zhang, G. Xiong, and W. Yu, "Radar HRRP recognition based on sparse denoising autoencoder and multi-layer perceptron deep model," in *Proc. IEEE UPINLBS*, Nov. 2016, pp. 283–288.
[17] Y. Zhai, B. Chen, H. Zhang, and Z. Wang, "Robust variational auto-encoder for radar HRRP target recognition," in *Proc. Int. Conf. Intell. Sci. Big Data Eng.*, vol. 10559, 2017, pp. 356–367.
[18] X. Peng, X. Gao, Y. Zhang, and X. Li, "An adaptive feature learning model for sequential radar high resolution range profile recognition," *Sensors*, vol. 17, no. 7, p. 1675, 2017.
[19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[20] T. Lindeberg, "Scale invariant feature transform," *Scholarpedia*, vol. 7, no. 5, pp. 2012–2021, 2012.
[21] A. P. Witkin, "Scale-space filtering," in *Proc. Int. Joint Conf. Artif. Intell.*, vol. 2, 1983, pp. 1019–1022.
[22] J. Babaud, A. P. Witkin, M. Baudin, and R. O. Duda, "Uniqueness of the Gaussian kernel for scale-space filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 1, pp. 26–33, Jan. 1986.
[23] F. F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.
[24] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 2818–2826.
[25] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.
[26] C. Guo, Y. He, H. Wang, T. Jian, and S. Sun, "Radar HRRP target recognition based on deep one-dimensional residual-inception network," *IEEE Access*, vol. 7, pp. 9191–9204, 2019.
[27] H. Wang *et al.* (2018). "CosFace: Large margin cosine loss for deep face recognition." [Online]. Available: https://arxiv.org/abs/1801.09414
[28] L. Ambrosio and G. Dal Maso, "A general chain rule for distributional derivatives," *Proc. Amer. Math. Soc.*, vol. 108, no. 3, pp. 691–702, 1990.

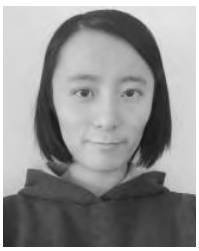
CHEN GUO received the M.S. degree from the National University of Defense Technology, China, in 2015. She is currently pursuing the Ph.D. degree with Naval Aviation University, China. She is also with Naval Aviation University. Her research interests include radar signal processing, machine learning, and deep learning.



HAIPENG WANG received the Ph.D. degree from Naval Aviation University, in 2012, where he is currently an Associate Professor. His research interests include intelligent perception and fusion, and big data technology and application. He also serves as a Reviewer of the several distinguished journals as *IET Radar, Sonar & Navigation* and the IEEE Aerospace and Electronic Systems Society.

TAO JIAN received the Ph.D. degree from Naval Aviation University, in 2010, where he is currently an Associate Professor. His research interests include radar signal processing and target recognition.

YOU HE received the Ph.D. degree from Tsinghua University, in 1997. He was a Visiting Scholar with the Brunswick University of Technology, Germany. He is currently a Full Professor and a Ph.D. Tutor of Naval Aviation University. His research interests include information fusion theory and technology, equipment simulation, and big data technology and application. He is an IET Fellow.



XIAOHAN ZHANG received the M.S. degree from the Aviation University of Air Force, China, in 2017. She is currently pursuing the Ph.D. degree with Naval Aviation University, China. She is also with Naval Aviation University. Her research interests include SAR target detection and deep learning.