

Radar HRRP Target Recognition Model Based on a Stacked CNN–Bi-RNN With Attention Mechanism

Mian Pan^{ID}, Ailin Liu, Yanzhen Yu, Penghui Wang, *Member, IEEE*, Jianjun Li^{ID}, Yan Liu, Shuaishuai Lv^{ID}, and He Zhu^{ID}

Abstract—The range resolution of high-resolution wideband radar is much smaller than the target size. Its echo signals tend to be diverse and sensitive to small changes of targets. Therefore, it is difficult to capture and distinguish the features in radar signals. In this article, we propose a radar target recognition pipeline based on a deep nested neural network. The framework consists of three parts: The translation sensitivity of the training data is first addressed in the preprocessing section. The second step is to obtain an embedded representation of the radar echo signals by the combination of the adjustment layer, convolutional neural network (CNN), and the squeeze and excitation (SE) block. Finally, the target is recognized through inputting embedded representation as a time sequence into the stacked bidirectional recurrent neural network (bi-RNN) based on an attention mechanism. Compared with the traditional methods, the proposed deep nested neural network extracts and takes advantage of the features of radar echo signals more effectively, including the envelope features and local physical structural features. The experimental results based on the test data indicate that the proposed method has a great advantage over other methods in the case of large data sets as well as small training data sets and is robust to the small translation of test samples and noises, exhibiting high engineering practical value.

Index Terms—Bidirectional recurrent neural network (bi-RNN), convolutional neural network (CNN), deep neural network, high-resolution range profiles (HRRPs), radar.

I. INTRODUCTION

THE resolution of high-resolution wideband radar is much smaller than the target size. The echo is also called the 1-D high-resolution range profile (HRRP) of the target [1]. The

Manuscript received April 19, 2020; revised October 17, 2020 and December 20, 2020; accepted January 11, 2021. Date of publication February 11, 2021; date of current version December 6, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61871164, Grant 61701379, and Grant 61871170, and Grant 61805060; in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LQ19E070003; and in part by the National Science and Technology Key Laboratory Foundation under Grant 6142401200201. (Corresponding author: H. Zhu.)

Mian Pan, Ailin Liu, Yanzhen Yu, Yan Liu, and Shuaishuai Lv are with the School of electronics and information, Hangzhou Dianzi University, Hangzhou 310018, China (e-mail: roy1022@foxmail.com; lainn@foxmail.com; yuyanzhen03@163.com; liuyan_1028@163.com; lvshuai@hdu.edu.cn).

Penghui Wang is with the National Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China (e-mail: wangpenghui@mail.xidian.edu.cn).

Jianjun Li is with the School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China (email: lijjcan@gmail.com)

He Zhu is with the School of electronics and information, Hangzhou Dianzi University, Hangzhou 310018, China, and also with the Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Hangzhou 310024, China (e-mail: hezhu@hdu.edu.cn).

Digital Object Identifier 10.1109/TGRS.2021.3055061

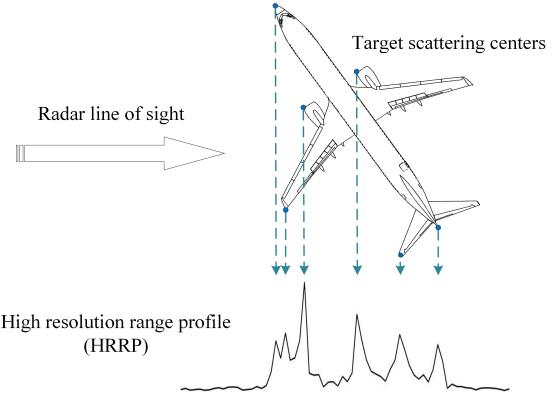


Fig. 1. Schematic of HRRP of radar target.

HRRP is the amplitude of the coherent sum of the complex time returning from the target scatterer along the radar line of sight, as shown in Fig. 1. The target scattering centers of a plane target are represented by blue circles in Fig. 1. The HRRP data contain structural information, such as the radial size of the target and the distribution of scattering points, which is extremely valuable for classification and recognition.

Therefore, the HRRP-based target recognition method has gradually become a hot topic in the radar automatic target recognition field [2]–[7].

For the majority HRRP target recognition system, feature extraction and utilization are a key portion [8]. The original HRRP data have large dimensions, which makes it an inadequate representation of the inherent attributes of the object. The main work of feature extraction is to provide useful information for subsequent recognition tasks through linear or nonlinear transformation. The effective features can not only thoroughly represent the data but also distinguish the differences between different categories to improve the accuracy of recognition.

The traditional feature extraction method can be divided into two classes, including the dimensional-reduction class and the transformer class. 1) Dimensional-reduction class, such as: a. subspace approximation model [9]–[12]; b. sparse representation [13]–[17]; c. scatter matching algorithm [18]; and d. manifold learning [19], [20] perform dimension reduction of the HRRP signal to obtain effective recognition features under certain assumptions. (2) Transformer class, such as bi-spectrum [21], spectrogram [22], and fast Fourier transform (FFT) amplitude feature [11], [23] projects HRRP signals to the frequency domain. Then the frequency features are

modeled and recognized. These methods have achieved good recognition performance in experiments. However, most of the feature extraction methods are unsupervised and lossy, which means that these methods do not focus well on the maximum separability feature. Moreover, part of the separability feature is lost during the extraction process, which is not favorable to the recognition. Second, the features extracted in these methods are highly dependent on the researchers' knowledge and experience accumulation of HRRP data, which makes it difficult to achieve satisfactory results in some cases lacking *a priori* information.

In recent years, deep learning-based methods have been introduced into the field of radar target recognition. The HRRP recognition method based on deep learning can be roughly divided into the following three classes: 1) encoder-decoder-based deep neural models [24]–[29]; 2) convolutional neural network (CNN) [30]–[35]; and 3) recurrent neural network (RNN) [36]. These end-to-end methods use supervised learning to automatically extract the separable features of the samples, which improves the deficiencies of traditional models in feature extraction. However, they still need to be improved. The encoder-decoder structure and CNN method extract the envelope feature of the HRRP while the sequence correlation between HRRPs that reflects the characteristics of the target physical structure units is ignored. The envelope feature here is also called the time-domain feature or regarded as a single-channel image feature, which is composed of the amplitude of the radar target echo on each range cell. The deep learning method based on an RNN uses the sequence correlation existing in HRRP to model and describe the physical structure characteristics of the target, but there are three shortcomings in the modeling process. First, the extracted features are highly redundant, which induces more difficulties for the subsequent RNN modeling [36]. Second, the input dimension and duration in time-domain segmentation are entangled. It is difficult to adjust these two parameters independently. Third, RNN only uses the information at the current moment and before without considering the information after the current moment. The envelope features contained in the HRRP cannot be fully exploited.

This article proposes an end-to-end radar HRRP target recognition model based on a deep nested neural network to solve the problems in the deep learning models. The method first overcomes the sensitivity in HRRP samples and introduces a dynamic adjusting layer. Then a deep nested neural network is used to build the model of the target structure in the HRRP signal. In the proposed model, dynamic adjustment layers, convolutional blocks, and squeeze and excitation (SE) blocks [37] are embedded in a stacked bidirectional RNN (bi-RNN) model based on attention mechanism. The dynamic adjustment layer is used to enhance those relatively small distance units that contain multiseparate information. The convolutional block and SE block could extract HRRP's envelope features while retain and enhance its local information. Stacked bi-RNN based on attention mechanism is used to abstract and describe the physical structure characteristics of the target contained in HRRP and integrate the multilayer physical structure characteristics for recognition. Compared

with the previous methods, the model proposed in this article has the following advantages.

- 1) *Nested Neural Network*: In the nested neural network, the CNN is adopted to extract the more representative envelope features in HRRP, which can effectively avoid the entanglement of input dimensions with duration as well as high redundancy in the time-domain segmentation method. Compared with the traditional unidirectional RNN model, the prior physical structure characteristic of the radar target can be better re-utilized in the bidirectional model, including the structural information around the current time.
- 2) *Two-Level Adjustment of Envelope Features and Local Physical Structure Features*: The weights of different channels are adjusted with distinct extracted envelope features of HRRP at the output of CNN by SE block when extracting the embedded representation in this model. Since different levels of features reflecting the local physical structure of the target are obtained, the importance of the features is adjusted by using a multilevel attention model. With this two-level adjustment, the model highlights the separable features and suppresses their relevant features to achieve better recognition performance.
- 3) *Dynamic Adjustment Layer*: As stated above, some useful separable features may not be well learned due to the relative amplitude issue [22]. Therefore, their influence on the subsequent classifiers is weak. To tackle this problem, a dynamic range adjustment layer has been added to adjust the relative amplitude of different range cells automatically. Compared with some two-stage methods that manually find the optimal power, the dynamic adjustment layer empowers the model to have end-to-end characteristics.

This article is organized as follows. In Section II, the related research results of deep learning in the field of HRRP recognition are briefly introduced. The framework and specific details of the target recognition method are proposed in Section III. In Section IV, the algorithm's training and test processes are given. The effectiveness of the algorithm is proposed in Section V, and it is verified by experiments based on measured data. Finally, the conclusion is outlined in Section VI.

II. BRIEF REVIEW OF RELATED DEEP NEURAL NETWORKS

In recent years, deep neural network has been widely used in language modeling [38], [39], speech recognition [40], signal enhancement [41], image recognition [42] and detection [43], [44] due to its strong representational ability. In the field of radar signal processing, deep learning models such as CNN [30]–[35] and RNN [36] are introduced in many works of literature for target recognition. Below we briefly introduce how CNN and RNN are used in the field of radar HRRP target recognition.

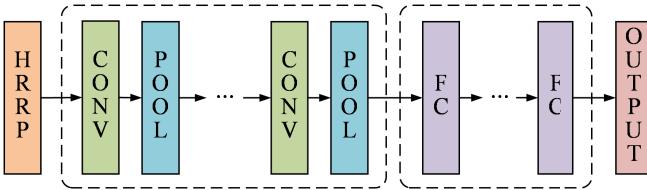


Fig. 2. HRRP recognition using CNN.

A. CNN [30]–[35], [45]

The CNN uses the local correlation of adjacent range cells in the HRRPs, and the HRRPs are hierarchically abstracted through the convolutional layer to obtain its high-level feature representation gradually. Finally, this representation is further processed through the fully connected (FC) layer to obtain classification results.

A typical framework of HRRP target recognition based on CNN is shown in Fig. 2. It consists of several convolutional layers, pooling layers, FC layers, and an output layer. First, the convolutional layer uses a sliding convolution kernel to extract local features of HRRP, and each convolution layer contains multiple different convolution kernels. Then, the pooling layer immediately after the convolution layer reduces the calculation amount of the model by suppressing the redundant information in the extracted features. Finally, high-level features are sent to the output layer whose neuron number is set to the total number of target categories. The softmax function is adopted to obtain classification results.

B. RNN [36], [46], [47]

RNN is a widely used neural network model. This model has memory characteristics and can process sequence data. Therefore, it is very suitable for modeling the target physical structure characteristics contained in HRRPs.

A typical RNN for radar HRRP target recognition is shown in Fig. 3. Its structure consists of an input layer at the bottom, a hidden layer in the middle, and an output layer at the top. The network first extracts the HRRP sequence features by time-domain segmentation or transformation; then uses this feature as the input of the RNN to sequentially calculate the hidden states at each time point; finally, the hidden states are calculated to obtain the output at each time point. After obtaining the classification features, similar to the CNN classification method, the softmax function is used to obtain the final classification result.

III. DEEP LEARNING FRAMEWORK FOR HRRP RECOGNITION

The framework of HRRP target recognition based on deep learning is explained in this section. As shown in Fig. 4, the framework can be divided into two parts: data preprocessing and deep nested neural network. In the first part, the sensitivity problems of HRRP are tackled, and the data set is augmented; in the second part, **CNN is used to extract the overall envelope features first**, and then the **stacked bi-RNN models targets' physical structure characteristics in the previous features**. Finally, the target recognition is performed through a multilevel attention model and a softmax function.

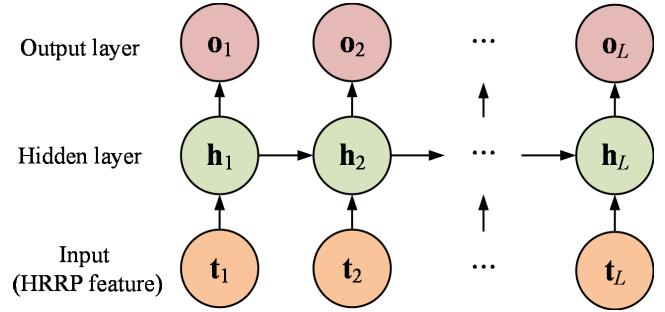


Fig. 3. HRRP recognition using RNN.

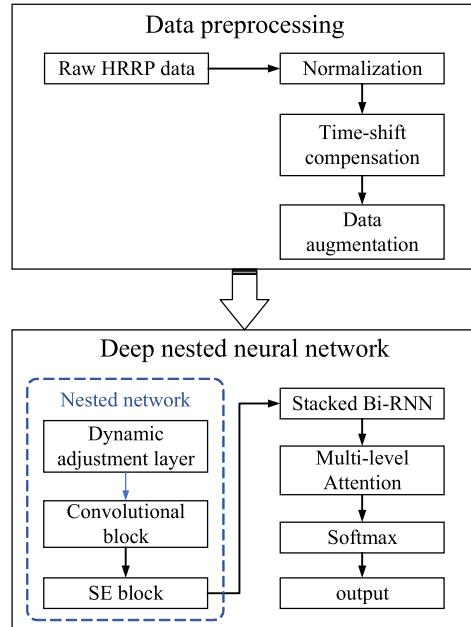


Fig. 4. Radar target recognition framework based on deep learning.

A. Raw HRRP Preprocessing

In this part, the intensity sensitivity and translation sensitivity of HRRP will be eliminated. The intensity of HRRP is determined by factors such as radar transmit power, target distance, radar antenna gain, and radar receiver gain. The collected data will have an inconsistent intensity problem. This problem is also called intensity sensitivity. Before using HRRP for target recognition, the original HRRP echo is processed by l_2 normalization to improve the intensity sensitivity problem. If the raw HRRP sample x_{raw} can be expressed as $x_{\text{raw}} = [x_1, x_2, \dots, x_M]$, where M is the total number of range cells contained in the HRRP, the intensity normalized HRRP sample $x_{\text{normalization}}$ can be expressed as

$$x_{\text{normalization}} = \frac{x_{\text{raw}}}{\sqrt{\sum_{m=1}^M x_m^2}}. \quad (1)$$

HRRP is intercepted from the radar echo data through the range window. During the interception, the position of the range image recorded in the range gate is not fixed, which leads to the translation sensitivity of HRRP. To have a uniform standard for training and testing, a center-of-gravity alignment

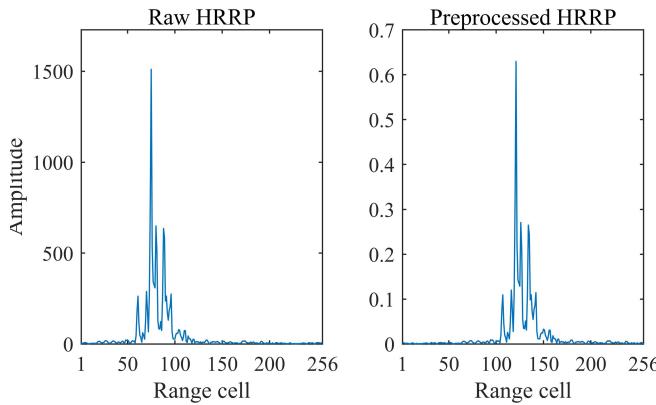


Fig. 5. Comparison of raw HRRP samples and preprocessed HRRP samples.

method is used to eliminate translation sensitivity. Center-of-gravity alignment by translating the HRRP so that its center of gravity is near the center of the HRRP, where the center of gravity g_{HRRP} of \mathbf{x}_{raw} is calculated as follows:

$$g_{\text{HRRP}} = \frac{\sum_{m=1}^M m x_m}{\sum_{m=1}^M x_m}. \quad (2)$$

The original HRRP sample and the HRRP sample after intensity normalization and center-of-gravity alignment are shown in Fig. 5. We can observe that the original HRRP sample has a large amplitude, and its center of gravity is to the left. After the intensity normalization and the center-of-gravity alignment method, the sample amplitude distribution is between 0 and 1, and the center of gravity is adjusted near the center point.

After dealing with the sensitivity problem, the training data set is expanded to improve the generalization ability of the recognition system and reduce its demand for the number of HRRP training samples. This article uses two methods for data expansion: 1) adding very small random noise to the HRRP sample, where the energy of the added random noise is generally lower than 1/100 of the signal energy; 2) shift the center of gravity of HRRP signal by several range cells randomly. By setting the parameters of the two methods, the size of the data set can be easily expanded to 50 to 100 times the size of the original HRRP data set.

B. Deep Nested Neural Network

The deep nested neural network mainly contains three parts: the nested network, stacked bi-RNN, and multilevel attention model. Next, the specific content of the three parts in the deep nested neural network will be discussed.

1) *Nested Network*: Due to the high information redundancy of the sequence features manually extracted by time-domain segmentation, and the mutual constraint between the data dimension and sequence length at each time point, the subsequent deep neural network recognition effect is limited. A nested network has been used instead of directly time-domain segmentation to avoid this problem. The nested network can not only automatically extract features from HRRP samples, but also easily control the dimension of features input to subsequent RNN by adjusting the convolution

kernel parameters. Besides, the sequence features extracted by the nested network retain the target structure information contained in the original echo, so that it can be used for subsequent RNN modeling. Therefore, the sequence features extracted by the CNN have good separability.

Because the amplitude of each range cell in the HRRP echo is quite different, sending the data directly into the convolutional block in the nested network will cause the model to focus too much on the range cell with the larger amplitude. However, the range cells with smaller amplitudes may also contain strongly separable features. If they are ignored, it is not good for radar target recognition. Therefore, the problem of excessively large range cell amplitude differences within the HRRP is first solved by dynamically adjusting the layer. The output $\mathbf{x}_{\text{dynamic}}$ of the dynamic adjustment layer can be expressed as

$$\mathbf{x}_{\text{dynamic}} = \begin{bmatrix} \mathbf{x}_{\text{dynamic}}^1 \\ \mathbf{x}_{\text{dynamic}}^2 \\ \vdots \\ \mathbf{x}_{\text{dynamic}}^D \end{bmatrix} \quad (3)$$

where D is the number of channels of the dynamic adjustment layer, and the i th dynamic adjustment channel $\mathbf{x}_{\text{dynamic}}^i$ can be expressed as

$$\mathbf{x}_{\text{dynamic}}^i = \mathbf{x}_{\text{pre}}^{\alpha_i} = [x_{\text{pre},1}^{\alpha_i}, x_{\text{pre},2}^{\alpha_i}, \dots, x_{\text{pre},M}^{\alpha_i}] \quad (4)$$

where α_i is the coefficient of the power transform and $x_{\text{pre},m}$ is the m th element in the preprocessed HRRP sample \mathbf{x}_{pre} .

Then, entering $\mathbf{x}_{\text{dynamic}}$ into the convolutional block. The convolutional block consists of three steps: convolution, batch normalization (BN), and ReLU. The working process of the convolution step is shown in Fig. 6. The input $\mathbf{x}_{\text{dynamic}}$ is convolved to obtain its features $\mathbf{F} = \{\mathbf{F}(i)\}_{i=1}^K$, where K represents the number of channels of the feature, $\mathbf{F}(i)$ represents the i th channel in \mathbf{F} , each channel contains L output values. $\mathbf{F}(i)$ can be obtained from

$$\mathbf{F}(i) = \mathbf{x}_{\text{dynamic}} \otimes \mathbf{Kernel}(i) + \mathbf{b}_{Ki} \quad (5)$$

where \otimes represents a convolution operation, $\mathbf{Kernel}(i)$ represents an i th convolution kernel, and \mathbf{b}_{Ki} represents an offset corresponding to the i th convolution kernel.

To make the training process more stable and the overall network easy to converge, a BN layer has been introduced after convolution. The BN layer normalizes inputs by calculating their mean and variance in the mini-batch. If a mini-batch contains N_{mini} HRRP samples, then the output of the mini-batch data through the convolution layer is recorded as $\mathbf{F}^n = \{\mathbf{F}^n(k)\}_{k=1}^{N_{\text{mini}}}$, where \mathbf{F}^n represents the convolution output corresponding to the n th HRRP sample. Each element in $\mathbf{F}^n = \{\mathbf{F}^n(k)\}_{k=1}^{N_{\text{mini}}}$ is transferred to $\tilde{\mathbf{F}}^n = \{\tilde{\mathbf{F}}^n(k)\}_{k=1}^{N_{\text{mini}}}$ by BN. $\tilde{\mathbf{F}}^n(k, l)$ can be expressed as

$$\tilde{\mathbf{F}}^n(k, l) = \frac{\gamma^k (\mathbf{F}^n(k, l) - E(\mathbf{F}^n(k, l)))}{\sqrt{Var(\mathbf{F}^n(k, l)) + \varepsilon}} + \beta^k \quad (6)$$

where $\mathbf{F}^n(k, l)$ and $\tilde{\mathbf{F}}^n(k, l)$, respectively, represent the l th element in the k th channel of the convolution layer corresponding to the n th HRRP sample before and after BN, γ^k and β^k are trainable parameters of the k th channel, ε is a small number,

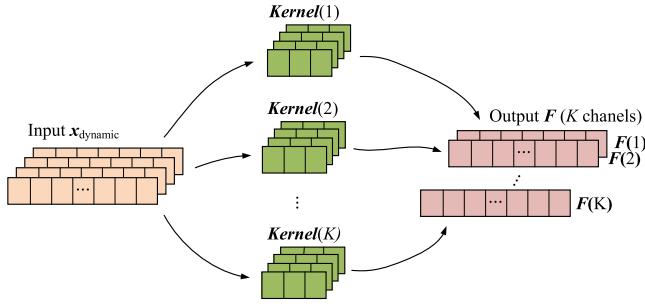


Fig. 6. Schematic of the convolutional step in a convolutional block (take input with four channels as an example).

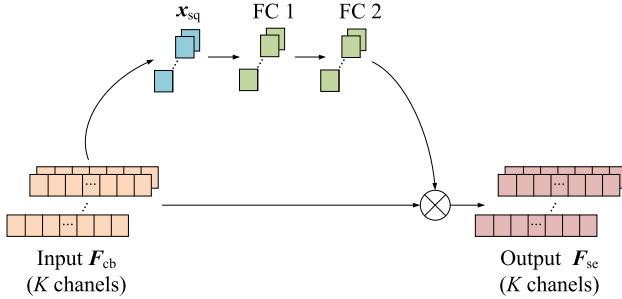


Fig. 7. Schematic of the SE block.

$E(\cdot)$ is the mean operation, $\text{Var}(\cdot)$ is the variance operation, and the mean and variance of $F^n(k, l)$ can be calculated by the following formula:

$$E(F^n(k, l)) = \frac{\sum_{n=1}^{N_{\min}} \sum_{l=1}^L F^n(k, l)}{N_{\min} L} \quad (7)$$

$$\text{Var}(F^n(k, l)) = \frac{\sum_{n=1}^{N_{\min}} \sum_{l=1}^L (F^n(k, l) - E(F^n(k, l)))^2}{N_{\min} L}. \quad (8)$$

ReLU function is applied to nonlinear activation of each element in $\tilde{F}^n\}_{n=1}^{N_{\min}}$ after BN to obtain $F_{cb}\}_{n=1}^{N_{\min}}$ whose elements can be expressed as

$$F_{cb}^n(k, l) = \max(0, \tilde{F}^n(k, l)). \quad (9)$$

Finally, the SE block is used for channel adjustment F_{cb} . The SE block can learn the global information of convolutional feature channels to selectively emphasize convolutional channels that contain more separability information and suppress less useful convolutional channels.

The implementation steps of the SE block are shown in Fig. 7. First, calculating the average value of each channel F_{cb} in turn and arrange it into a vector $x_{sq} = [x_{sq}(1), \dots, x_{sq}(K)]$. The element $x_{sq}(i)$ in the vector can be expressed as

$$x_{sq}(i) = \frac{1}{L} \sum_{l=1}^L F_{cb}(i, l) \quad (10)$$

where $F_{cb}(i, l)$ represents the l th element of the i th channel of the feature F_{cb} , and $F_{cb} \in \hat{\mathbb{R}}^{K \times L}$.

Next, passing x_{sq} through two layers of full connection, and the output can be obtained by the following formula:

$$x_{FC_i} = f(W_{FC_i} x_{FC_{i-1}} + b_i) \quad (11)$$

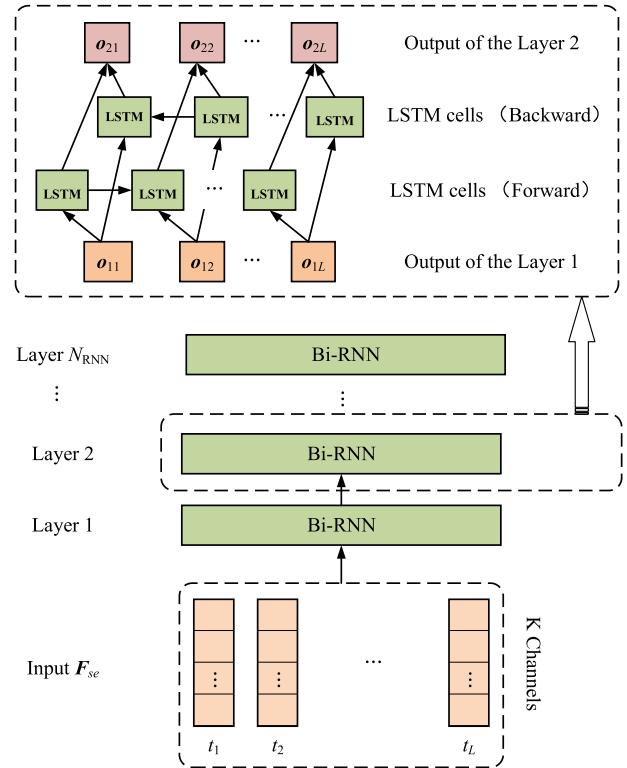


Fig. 8. Stacked bi-RNN structure diagram.

where W_{FC_i} and b_i are the weight matrix and the offset matrix of the i th FC layer, respectively, $i = 1, 2$. In the first full connection, enter $x_{FC_{i-1}}$ as x_{sq} and the activation function $f(\cdot)$ is ReLU; in the second full connection, the activation function $f(\cdot)$ uses the Sigmoid function.

Finally, the convolution feature F_{cb} is selectively adjusted to obtain the output F_{se} of the SE block

$$F_{se} = F_{cb} \odot x_{FC_2} \quad (12)$$

where \odot means that each element in each channel F_{cb} is multiplied by the corresponding element of the x_{FC_2} .

2) *Stacked Bi-RNN With Multilevel Attention Mechanism:* In the traditional radar target recognition methods implemented using RNN, HRRP data can only be input in one direction, resulting in the input of the current time being only related to the past input. Also, the shallow model has limited ability to describe the overall structural characteristics of the target. This article uses the stacked bi-RNN model to solve this problem. The model structure is shown in Fig. 8, where bi-RNN divides the HRRP data into two independent RNN models in forward and reverse directions and stitches their outputs together so that the input at the current moment can be connected with the entire inputs at any time point in the sequence. Besides, multiple bi-RNNs have been stacked to a suitable depth, which can better extract the high-level physical structure features.

In Fig. 8, the output F_{se} of the SE block is considered as a time series with duration L and dimensions K at each time point and inputted to the stacked bi-RNN. Stacked bi-RNN can be built based on long short-term memory (LSTM) units [47], [48]. Among them, the output vector o_{il} correspond-

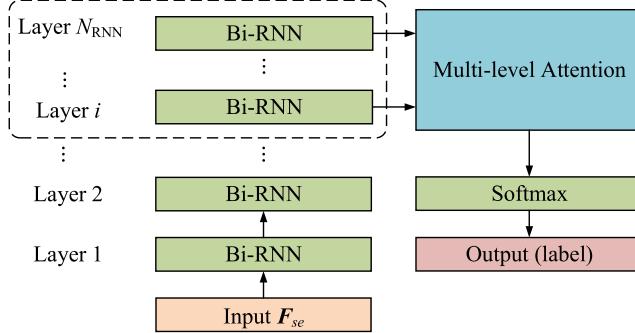


Fig. 9. Radar HRRP target recognition framework based on multilevel attention model.

ing to the l th time point of the i th bi-RNN is a concatenation of the forward and backward RNN output \mathbf{o}_{iFl} and \mathbf{o}_{iBl} at the l th time point, that is, $\mathbf{o}_{il} = [\mathbf{o}_{iFl}; \mathbf{o}_{iBl}]$. \mathbf{o}_{iFl} is calculated as follows:

$$\mathbf{o}_{iFl} = \tanh(s_{iFl})\mathbf{q}_{iFl} \quad (13)$$

where $\tanh(\cdot)$ represents a hyperbolic tangent activation function, and s_{iFl} and \mathbf{q}_{iFl} are the state units and output gates of the i th bi-RNN corresponding to the l th time point of the forward RNN, respectively. For detailed calculation methods, please refer to the literature [48]. The calculation method for \mathbf{o}_{iBl} is similar to that of \mathbf{o}_{iFl} and is omitted for brevity.

Then, we apply a multilevel attention mechanism to synthesize the outputs of each layer of the stacked bi-RNN and a softmax layer to obtain the classification results, both of which are shown in Fig. 9. **The multilevel attention mechanism can not only complement the outputs that reflect the characteristics of different levels of the structure but also highlight the output weights that are typical of the target structure and suppress the output weights vulnerable to noise.**

Applying a multilevel attention mechanism, the feature \mathbf{F}_{ATT} used for classification can be obtained by weighting the output \mathbf{o}_{il} of each layer

$$\mathbf{F}_{ATT} = \sum_{i=N_{RNN}-N_{ATT}+1}^{N_{RNN}} \sum_{l=1}^L \alpha_{il} \mathbf{o}_{il} \quad (14)$$

where α_{il} represents the weight of the output \mathbf{o}_{il} , N_{RNN} is the total layer number of the stacked bi-RNN, and N_{ATT} represents the number of layers used by the multilevel attention mechanism. The weight α_{il} is calculated as follows:

$$\alpha_{il} = \frac{\exp(\mathbf{U}_i \tanh(\mathbf{W}_i \mathbf{o}_{il}))}{\sum_{i=N_{RNN}-N_{ATT}+1}^{N_{RNN}} \sum_{j=1}^L \exp(\mathbf{U}_i \tanh(\mathbf{W}_i \mathbf{o}_{ij}))} \quad (15)$$

where \mathbf{U}_i and \mathbf{W}_i are parameters corresponding to the output of the i th layer.

C. Recognition

The features \mathbf{F}_{ATT} is classified through the softmax layer. The probability that the test HRRP sample \mathbf{x}_{test} corresponds to the i th target in the target set can be expressed as the following:

$$P(i|\mathbf{x}_{test}) = \frac{\exp(\mathbf{F}_s(i))}{\sum_{c=1}^C \exp(\mathbf{F}_s(c))} \quad (16)$$

where c is category number, C represents the total number of targets, and $\mathbf{F}_s(i)$ refers to the i th element in the vector \mathbf{F}_s . $\mathbf{F}_s = \mathbf{W}_s \mathbf{F}_{ATT}$, and \mathbf{W}_s is the weight matrix of the vector \mathbf{F}_s . The test HRRP sample \mathbf{x}_{test} is classified into c_0 with the maximum target probability:

$$c_0 = \operatorname{argmax}_i P(i|\mathbf{x}_{test}). \quad (17)$$

D. Cost Function

In the proposed model, the cross-entropy loss is used as the cost function. The cost of each training sample \mathbf{x}_{train} can be expressed as

$$L = - \sum_{i=1}^C z(i) \log P(i|\mathbf{x}_{train}) \quad (18)$$

where $z(i)$ represents the i th element in the true label \mathbf{z} of the training sample, z is a one-hot vector, and $P(i|\mathbf{x}_{train})$ represents the probability that the training sample \mathbf{x}_{train} corresponds to the i th target.

IV. TRAINING AND TESTING PROCEDURE

The overall training and testing process of the method proposed in this article is shown in Algorithm 1. In the training phase, the original HRRP data are first preprocessed, and a deep nested neural network is established with its parameters initialized randomly. Then the mini-batch-based BP algorithm is adopted to update the parameters of the network. Finally, the network parameters are saved. In the test phase, also each test HRRP sample is preprocessed first and then substituted into the trained network for forwarding propagation. Finally, the test HRRP sample is classified into the category with the greatest probability.

V. EXPERIMENTS AND RESULTS

A. Experimental Data Set

The data set used in this article contains three types of aircraft targets measured by an ISAR radar. This data set are the most used practical data sets in this field [3], [9]–[11], [13], [15], [18], [21], [24]–[26], [35], [36]. The specific parameters of the radar and aircraft targets are shown in Table I. The radar operates on the C-band, with a pulse repetition frequency of 400 Hz and a signal bandwidth of 400 MHz. The targets of the three types of aircraft are “An-26” aircraft, “Cessna” aircraft, and “Yark-42” aircraft. Among them, the “An-26” aircraft is medium-sized propeller aircraft, the “Cessna” aircraft is a small-sized jet aircraft, and the “Yark-42” aircraft is a large-sized jet aircraft. Each HRRP sample in the data set contains 256 range cells.

The projections of the trajectories of the three aircraft on the ground are shown in Fig. 10, which are segmented. Among them, the HRRP data of An-26 and Cessna are divided into seven segments, and the HRRP data of Yark-42 aircraft are divided into five segments. To verify the generalization ability of the recognition method, the training data of this article include the fifth and sixth segmentation of the An-26, the sixth

Algorithm 1 Whole Process of The Proposed Framework

Stage	Code
Preprocessing	<p>Input: HRRP data $\{\mathbf{x}_{c,n}\}_{c=1,n=1}^{C,N_c}$, where C represents the total number of targets, N_c represents the number of HRRP samples corresponding to the target c.</p> <p>For $c = 1$ to C</p> <p> For $n = 1$ to N_c</p> <p> Apply formulas (1) and (2) to HRRP sample $\mathbf{x}_{c,n}$ for intensity normalization and center-of-gravity alignment preprocessing respectively</p> <p> Add minimal random noise to the HRRP sample and randomly shift the center of gravity of HRRP signal by several range units</p> <p> End</p> <p>End</p> <p>Output: preprocessed data set $\{\hat{\mathbf{x}}_{c,n}\}_{c=1,n=1}^{C,\tilde{N}_c}$, \tilde{N}_c represents the number of HRRP samples corresponding to target c after data expansion.</p>
Training	<p>Input: data set $\{\hat{\mathbf{x}}_{c,n}\}_{c=1,n=1}^{C,\tilde{N}_c}$, random initialization parameter set $\Phi = \{\theta_{DA}, \theta_{CNN}, \theta_{SE}, \theta_{RNN}, \theta_{ATT}\}$, and learning rate l_r. θ_{DA}, θ_{CNN}, θ_{SE}, θ_{RNN}, and θ_{ATT} respectively represent the parameters of the dynamic adjustment layer, the convolutional block, the SE block, the stacked bi-RNN layer, and the multi-level attention mechanism layer.</p> <p>For $epoch = 1$ to $Cycle$ ($Cycle$ is the total number of data set cycles)</p> <p>Shuffle the data set and divide it into K mini-batches</p> <p>For $batch = 1$ to K</p> <p> (Forward propagation process)</p> <p> Calculate the characteristics after dynamic adjustment according to formulas (3) and (4)</p> <p> Calculate the features obtained after the convolutional block according to formula (5)</p> <p> The characteristics after BN are calculated according to formulas (6)–(8)</p> <p> Calculate the features after ReLU according to formula (9)</p> <p> Calculate the features obtained by SE block according to formulas (10)–(12)</p> <p> Calculate the output of a stacked bi-RNN according to formula (13)</p> <p> Calculate the weight of attention according to formulas (14) and (15)</p>

Algorithm 1 Continued. Whole Process of The Proposed Framework

Stage	<p>Calculate the output label distribution by equation (16) (Backward propagation process)</p> <p>Calculate the loss function according to formula (18)</p> <p>Calculate the stochastic gradient of each parameter in the parameter set in turn and use the gradient descent method for backpropagation.</p> <p>End</p> <p>End</p> <p>Output: the proposed deep neural network Φ with specific parameters.</p>
Testing	<p>Input: test data \mathbf{x}_{test}, trained network parameters Φ</p> <p>According to formulas (1) and (2), normalize the intensity of the test sample \mathbf{x}_{test} and align the center of gravity.</p> <p>Substitute the trained network Φ for forwarding propagation to get its output label distribution.</p> <p>Get the classification result of the test sample according to formula (17).</p> <p>Output: The classification result.</p>

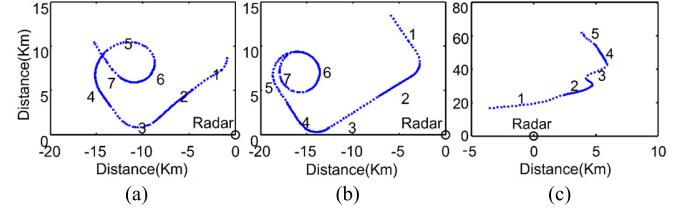


Fig. 10. Flight paths of three targets projected on the ground: (a) An-26, (b) Cessna, (c) Yark-42.

TABLE I
THREE AIRCRAFT PARAMETERS

The plane	Length (m)	Width (m)	Height (m)
An-26	23.80	29.20	9.83
Cessna	14.40	15.90	4.57
Yark-42	36.38	34.88	9.83

and seventh segmentation of the Cessna, and the second and fifth segmentation of the Yark-42. The remaining data are used as test data, and the training data contain information about all the aspects of the target.

B. Introduction to the Models and Comparative Models

1) *Parameter Settings of the Proposed Model:* The hyper-parameters of the proposed model are determined

regarding some previous similar studies [22], [49], and the results of multiple verification experiments.

The dynamic adjustment layer of the model proposed in this article contains four channels, whose parameters α are set to 0.3, 0.7, 1, and 1.4, respectively; the convolutional block contains 96 convolution kernels, each of which has a size of 1×16 and a step size of 8; the stacked bi-RNN is a stack of five bi-RNNs. The outputs of the last three bi-RNNs are used based on the multilevel attention mechanism.

2) Comparison Methods:

a) *Template matching method* [21]: Traditional recognition methods need to divide the training data into dozens of small parts according to the target-aspect to avoid the motion through range cells (MTRC) phenomenon [10], [49]. Each piece is called a “frame.” Template Matching Method (TMM) first divide the training data of An-26, Cessna, and Yark-42 into 50, 50, and 35 frames, respectively; and then the average profile of each frame is calculated and saved as a template; finally, the Euclidean distance between the test sample and each frame template is calculated, and the test sample is classified under the criterion of minimum Euclidean distance.

b) *Adaptive Gaussian Classifier and Gaussian Mixture Model* [49]: The Adaptive Gaussian Classifier (AGC) and Gaussian Mixture Model (GMM) are implemented through the scikit-learn [50] python library. In the implementation process, it is assumed that the AGC and the GMM model are independent in each dimension, and the GMM is a mixture of three Gaussian components.

c) *Factor analysis model* [10]: The factor analysis (FA) model is also implemented through the scikit-learn python library. The data preprocessing method of the FA model is consistent with that of the TMM, and the hidden dimension of the FA model is set to 30.

d) *SVM* [51]: LIBSVM [52] is used to implement linear and nonlinear SVM models. The preprocessing steps of the SVM model include normalization and center-of-gravity alignment. The kernel function of the nonlinear SVM model is RBF.

e) *Deep neural network*: The deep neural network includes three typical networks: CNN, RNN, and stacked AE. The CNN consists of three convolutional layers and two FC layers. The number of convolution kernels of each convolution layer is 8, 16, and 32 from the bottom to the upper layer. The size of the convolution kernel is 1×16 , and the step size is two. The FC layer contains 300 and 3 neurons, respectively. The implementation of RNN is based on LSTM cells, and its input sequence is extracted from HRRP samples according to the time-domain segmentation method [36]. The input sequence length is 31, and each time point is a 16-D vector. The stacked AE model is a stack of five AEs, where the number of neurons in each layer is 300, 600, 900, 2000, and 3.

C. Experimental Results Using All Training Data

The training process of this experiment used all HRRP training samples for a total of 137880. Among them, there are 51200 samples of An-26 aircraft, 51200 samples of Cessna

			Proposed model (ours)			Stacked AE			
			An-26	Cessna	Yark-42	An-26	Cessna	Yark-42	
	An-26	97.04	1.92	1.04		An-26	95.10	3.95	0.95
Cessna		1.09	98.26	0.66		Cessna	9.48	89.08	1.48
Yark-42		0.01	0.16	99.83		Yark-42	2.72	0.66	96.61
	An-26	Cessna	Yark-42			An-26	Cessna	Yark-42	
	CNN			RNN			Non-linear SVM		
	An-26	91.88	3.48	4.63	An-26	91.23	4.55	4.22	
Cessna		13.17	85.86	0.97	Cessna	5.58	93.78	0.64	
Yark-42		0.72	1.59	97.70	Yark-42	1.25	2.57	96.18	
	An-26	Cessna	Yark-42		An-26	Cessna	Yark-42		
	FA model			Linear SVM			GMM		
	An-26	91.44	1.13	7.43	An-26	94.95	2.23	2.81	
Cessna		4.64	94.62	0.74	Cessna	7.27	92.20	0.53	
Yark-42		2.00	0.20	98.80	Yark-42	1.11	0.22	98.67	
	An-26	Cessna	Yark-42		An-26	Cessna	Yark-42		
	TMM			AGC					
	An-26	82.22	12.36	5.42	An-26	90.41	1.58	8.02	
Cessna		10.84	89.01	0.16	Cessna	16.67	81.66	1.67	
Yark-42		6.39	15.72	77.89	Yark-42	0.10	0.30	99.60	
	An-26	Cessna	Yark-42		An-26	Cessna	Yark-42		

Fig. 11. Confusion matrix comparison of different methods using all training data.

aircraft, and 35840 sample aircraft of Yark-42. The recognition rate of each aircraft and the average recognition rate are used as evaluation indexes. Moreover, we also consider the balance between the recognition rates of each aircraft.

Table II shows the recognition performance of different models for each target. It can be seen that the proposed model performs best in all models. Comparing CNN, RNN, and stacked AE three typical deep neural network methods, the proposed method has an average recognition rate of 4.65% higher than the best performing RNN model among them; compared with the traditional methods like linear SVM, non-linear SVM, FA, AGC, and GMM, the model proposed in this article has an average recognition rate of 3.11% higher than the best performing nonlinear SVM. Compared with the TMM, the model proposed in this article has an average recognition of 15.34% higher.

TABLE II
RECOGNITION RESULTS OF DIFFERENT METHODS (%)

Methods	An-26	Cessna	Yark-42	Average
The proposed model	97.04	98.26	99.83	98.38
Stacked AE	95.10	89.08	96.61	93.60
CNN	91.88	85.86	97.70	91.81
RNN	91.23	93.78	96.18	93.73
FA	91.44	94.62	98.80	94.95
Nonlinear SVM	94.95	92.20	98.67	95.27
Linear SVM	92.09	84.85	97.03	91.32
GMM	94.98	81.33	99.09	91.80
TMM	82.22	89.01	77.89	83.04
AGC	90.41	81.66	99.60	90.56

After analyzing the overall recognition performance of each method, the confusion matrix of each method is shown in Fig. 11. As can be seen, the accuracy of the proposed method is relatively average for all types of aircraft recognition. Among them, Yark-42 aircraft with the highest recognition accuracy and An-26 aircraft with the lowest recognition accuracy differ by only 2.79%, which shows that the method proposed can model the characteristics of the three types of aircraft in a more balanced manner. That is because the proposed model can use not only the latent physical structure information of the target shared by the training data set but also utilizes both the envelope features and the sequence features of HRRP to extract better separable features. Other models, more or less, have the problem of imbalance in recognition ability among various types. For example, based on deep learning stacked AE, although the average recognition rate of the three types of aircraft reached 93.59%, its recognition rate for Cessna is only 89.08%, which was 7.53% lower than that of Yark-42; moreover, from the confusion matrix of stacked AE, it can be found that Cessna and An-26 misjudge each other more, which shows that stacked AE is insufficient in describing the characteristics of these two types of aircraft. The same problem is evident in CNN, linear SVM, GMM, AGC, and TMM models. If the recognition rate of each type of aircraft is higher than 90% as engineering availability conditions, these models will be challenging to apply to actual engineering.

D. Ablation Study

We evaluate each module's contribution to the proposed model by examining the recognition rate when removing different modules. The detailed results are shown in Table III. The average recognition rate of the complete model (Model X) is 98.38%. From the comparison, the following conclusions can be drawn:

1) *Multilayer Attention Mechanism (Model I Versus Model X)*: After removing the multilevel attention mechanism in the model, the average recognition rate decreases by 1.05% to 97.35%, indicating that multilevel attention can make the model focus more on the useful target structure information for recognition and achieve different levels of structural

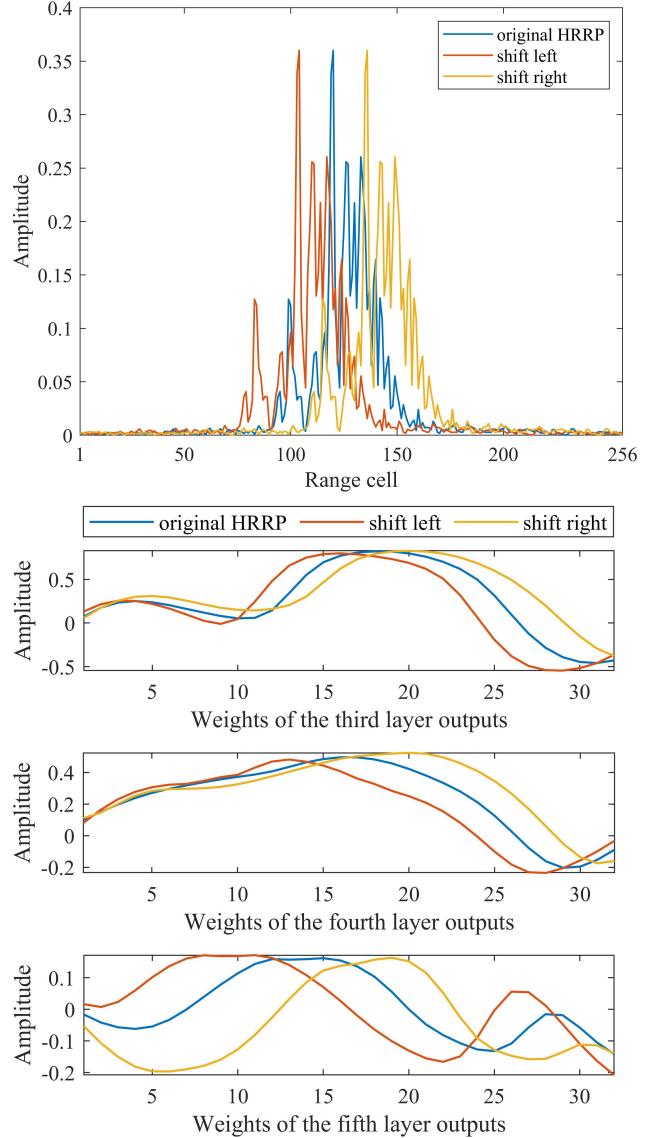


Fig. 12. Schematic of the multilevel attention mechanism for translation compensation principle.

features complement each other and are helpful for final recognition.

2) *Stack (Model II Versus Model X)*: When the stacked bi-RNN becomes a single bi-RNN, the model's average recognition rate is reduced by 1.53% to 96.85% (0.48% lower than the model I). This result indicates that the deep network formed by stacking can better extract the high-level features of the HRRP samples that are beneficial to classification.

3) *Bi-RNN (Model III Versus Model X)*: After changing the model's bi-RNN structure to traditional unidirectional RNN, the average recognition rate drops by 0.93% to 97.45%. Compared with the unidirectional RNN, which can only use the structure information at and before the current time, bi-RNN can better use the envelope features in HRRP.

4) *Dynamic Adjustment Layer (Model IV Versus Model X)*: The dynamic adjustment layer is removed in Model IV. At this time, some features with good separability are not well used due to the high dynamic range of amplitude, and thus the

TABLE III
COMPARISON OF THE RECOGNITION PERFORMANCE OF THE PROPOSED MODEL BY APPLYING ALL TRAINING DATA
AND REMOVING SOME MODULES IN THE MODEL (%)

Number	DAL	Nested CNN	SE block	Uni-RNN	Bi-RNN	Stack	MA	An-26	Cessna	Yark-42	Average
I	√	√	√		√	√		96.11	97.87	98.02	97.33
II	√	√	√		√		NS	95.82	96.04	98.68	96.85
III	√	√	√	√		√	√	95.72	97.00	99.63	97.45
IV		√	√		√	√	√	96.14	96.19	99.14	97.16
V	√	√			√	√	√	96.12	96.23	98.76	97.04
VI	√		√		√	√	√	95.25	96.31	99.40	96.99
VII		√		√				95.20	94.28	98.62	96.03
VIII				√				91.23	93.78	96.18	93.73
IX		CNN*						91.88	85.86	97.70	91.81
X	√	√	√		√	√	√	97.04	98.26	99.83	98.38

DAL: dynamic adjustment layer, Uni-RNN: traditional unidirectional RNN, Bi-RNN: bi-directional RNN, MA: multi-layer Attention, NS: not supported (MA cannot be implemented for one layer bi-RNN), CNN*: normal CNN (the nested CNN cannot be implemented alone)

average recognition rate of the model decreases by 1.22% to 97.16%.

5) *SE-Block (Model V Versus Model X)*: Again, removing the SE block in the model, the average recognition rate drops by 1.34% to 97.04%. The result indicates the channel importance adjustment can highlight the beneficial features and suppress the redundant features, which is conducive to the subsequent stacked bi-RNN modeling of the target.

6) *Nested CNN (Model VI Versus Model X or Model VII Versus Model VIII)*: We remove the nested CNN module from the original model and replace it with the time-domain segmentation method. The recognition rate drops to 96.03% (2.35% lower than the complete model). Using the nested CNN to extract sequence features can solve big redundancy problems in the sequence features extracted by time-domain segmentation and the mutual constraint of the data dimension and sequence length at each time point of input. The comparison between Model VII and Model VIII also demonstrates the effectiveness of nested CNN.

E. Robustness of the Model to Test Sample Translation

Due to the inevitability of peak fluctuations in the acquisition process of HRRP, even when the observations of the same target are adjacent, the use of the center-of-gravity alignment method sometimes has several range cell offsets. Therefore, it is of great engineering significance that the model is robust to the test sample's small-scale translation. Fig. 12 illustrates that the extracted attention feature F_{ATT} in this article will not change significantly with the translation of HRRP. In principle, the output of bi-RNN will be shifted according to the shift of HRRP samples. Suppose the weight of the multilayer attention mechanism also shifts according to the output of the bi-RNN. In that case, the extracted attention feature of the shifted HRRP sample is similar to the original HRRP sample. Consequently, the final recognition result is robust. In Fig. 12, we first show the left-shifted, untreated, and right-shifted HRRP examples at the top. Then we observe in the bottom three subgraph that the assignment of weights for the last three layers follows the shift of HRRP samples. This result shows that our model is robust

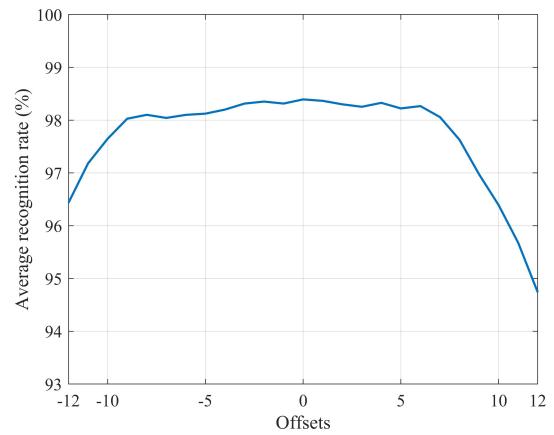


Fig. 13. Average recognition performance of the proposed model when testing HRRP samples with different degrees of offset.

to translation. Therefore, when the test sample undergoes a small range of translation, the multilevel attention mechanism can well compensate for the effect of translation.

The average recognition performance of the model proposed in this article is shown in Fig. 13 when the test samples are offset to different degrees. It can be seen from Fig. 13 that the proposed model is very robust to the translation of the test sample. When the offset of the test HRRP sample is less than five range units, the recognition rate of the proposed model remains basically unchanged; as the further offset increases, the recognition performance of the model slightly decreases. Even if the test HRRP sample offset reaches 12 range cells, the average recognition rate of the proposed model can still exceed 94%.

F. Recognition Results With Different Training Sample Sizes

In the case of noncooperative target recognition, it is difficult to ensure that enough HRRP samples can be collected for training, so the small sample learning problem is often encountered. A robust recognition model should still maintain good recognition performance with a small training set.

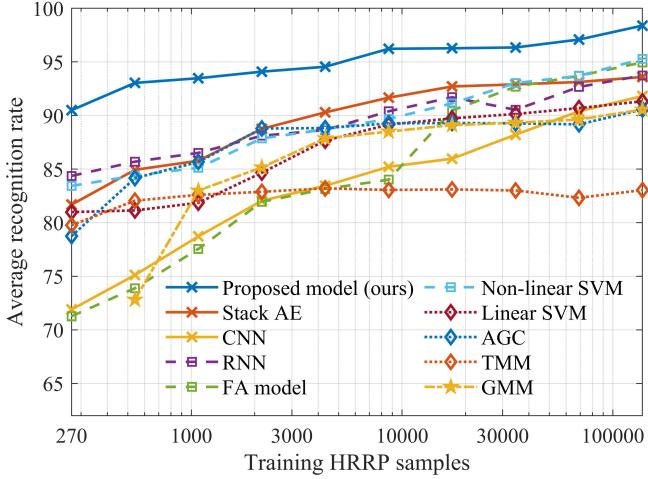


Fig. 14. Average recognition rates of various methods on different training sample sizes.

To evaluate the impact of training set size on the recognition results, a total of eight training sample sets of different sizes are used in the experiments in this section. Since the traditional model divides the training samples into 135 frames during the recognition process, a training sample set is used that is a multiple of 135 times to make it available to all models. The training sample set sizes used in the experiments are $135 \times 2^{n_{\text{train}}}$, and $n_{\text{train}} = 1, 2, \dots, 8$, respectively. The recognition performance of various methods under different training sample sets is shown in Fig. 14.

As can be seen from Fig. 14, the recognition effect of traditional recognition methods is more sensitive to the decrease of training sample size, because, in these models, the HRRP data are divided into many frames with each modeled independently. The proposed method trains the data in all frames collectively and can discover the similarity features contained in the HRRP samples in different frames. Compared to other methods, our method is more robust under small sample conditions. Even in the extreme case where the number of training samples is only 270, the proposed model can still reach an average recognition rate of 90.49%.

When the number of training samples is only 270, the confusion matrix of each method is shown in Fig. 15. In the Gaussian mixture model, the number of HRRP training samples contained in each frame must be greater than the number of mixed Gaussians, so this model cannot be used in this set of experiments. Similar to Section V-C, traditional methods and some deep learning methods, such as Gaussian Model, SVM, CNN, etc., are also very unbalanced in the recognition performance of the three types of aircraft. Especially for the An-26 and Cessna aircraft, the misjudgment is more serious. This shows that when the training sample set is small, the comparison model is difficult to model three types of aircraft. On the contrary, the method proposed in this article performs well with a small number of samples. For the An-26, Cessna, and Yark-42 aircraft, the recognition performance is still very balanced, and the correct recognition rate of all types of aircraft is higher than 90%.

Proposed model (ours)			Stacked AE				
An-26	90.59	4.35	4.06	An-26	83.74	6.00	10.26
Cessna	9.10	90.47	0.43	Cessna	27.30	70.48	2.21
Yark-42	1.32	7.25	90.42	Yark-42	7.99	1.18	90.82
An-26	An-26	Cessna	Yark-42 <th>An-26</th> <td>Cessna</td> <td>Yark-42</td>	An-26	Cessna	Yark-42	
CNN			RNN				
An-26	70.43	15.23	14.34	An-26	79.17	5.91	14.91
Cessna	29.66	66.51	3.84	Cessna	12.06	83.58	4.36
Yark-42	10.66	10.59	78.75	Yark-42	2.40	7.28	90.33
An-26	An-26	Cessna	Yark-42 <th>An-26</th> <td>Cessna</td> <td>Yark-42</td>	An-26	Cessna	Yark-42	
FA model			Non-linear SVM				
An-26	64.91	21.83	13.27	An-26	79.79	7.73	12.48
Cessna	21.23	78.09	0.68	Cessna	21.40	77.70	0.90
Yark-42	20.05	9.09	70.86	Yark-42	5.60	1.59	92.81
An-26	An-26	Cessna	Yark-42 <th>An-26</th> <td>Cessna</td> <td>Yark-42</td>	An-26	Cessna	Yark-42	
Linear SVM			AGC				
An-26	75.83	9.20	14.98	An-26	83.41	3.64	12.95
Cessna	21.84	76.92	1.23	Cessna	37.88	56.35	5.77
Yark-42	7.30	2.47	90.22	Yark-42	0.12	3.41	96.47
An-26	An-26	Cessna	Yark-42 <th>An-26</th> <td>Cessna</td> <td>Yark-42</td>	An-26	Cessna	Yark-42	
TMM							
An-26	82.20	11.97	5.83				
Cessna	13.44	86.32	0.24				
Yark-42	10.53	18.71	70.76				
An-26	An-26	Cessna	Yark-42 <th data-cs="3" data-kind="parent"></th> <th data-kind="ghost"></th> <th data-kind="ghost"></th>				

Fig. 15. Confusion matrix comparison of different models at 270 training data sets.

The detailed results of different methods with only 135 training samples are shown in Table IV. If the total number of HRRP training samples further drops to 135, the average recognition rate of our proposed model can still be maintained at around 90.40%. In contrast, the average recognition rate of other models has a significant decline compared with 270 training samples. The gap in the recognition rate has further widened. The average recognition rate of the compared stacked AE, CNN, and RNN model are 12.60%, 24.91%, and 9.09% lower than the proposed model, respectively. Traditional methods such as Nonlinear-SVM, linear SVM, and TMM are 13.58%, 14.61%, and 12.10% lower than the proposed model, respectively. Moreover, the recognition rate of our model in each aircraft category is also in the leading position. In summary, the proposed model is more robust than the

TABLE IV
RECOGNITION RESULTS OF DIFFERENT METHODS
WITH ONLY 135 TRAINING SAMPLES (%)

Methods	An-26	Cessna	Yark-42	Average
The proposed model	89.87	87.60	93.73	90.40
Stacked AE	69.66	70.18	93.57	77.80
CNN	64.70	59.20	72.56	65.49
RNN	79.13	85.38	79.41	81.31
Nonlinear SVM	68.40	68.94	93.13	76.82
Linear SVM	66.84	67.59	92.95	75.79
TMM	87.56	79.64	67.70	78.30

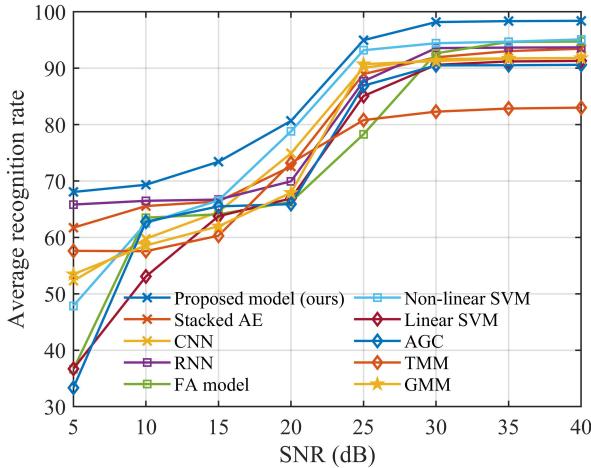


Fig. 16. The average recognition rate of various methods under different SNR.

comparison model as the number of samples declines, which extends its engineering practicability.

G. Recognition Results of Test Samples With Different SNRs

In actual engineering, the HRRP used in the test phase is often obtained in noncooperative situations, and it is hard to guarantee the high signal-to-noise ratio (SNR). To test the influence of the SNR on the recognition results, we generated the test HRRP samples with different SNR by artificially adding Gaussian white noise to produce eight types of test sample sets with different SNR from 5 to 40 dB. The SNR can be calculated as

$$\text{SNR} = 10 \times \log_{10} \left(\frac{\sum_{m=1}^M P_{x_m}}{M \times \sigma_n^2} \right) \quad (19)$$

where P_{x_m} denotes the power of the raw test HRRP data in the m th range cell, M is the total number of range cells contained in the HRRP, and σ_n^2 denotes the power of noise.

The recognition performance of various methods for different SNR test sets is shown in Fig. 16. As seen from Fig. 16, the recognition performance of various methods is good when the SNR is high ($\text{SNR} \geq 30 dB). When the SNR is 30 dB, our approach has an average recognition rate of 98.18%, which is 4.61% higher than the second-ranked nonlinear SVM method. The recognition performance of all methods decreases as the SNR decreases. Although our method exhibits the same trend, it has always been in a leading position. When the$

SNR of the test sample is reduced from 30 to 5 dB, our recognition rate remains at 68.06%, while the recognition rates of other methods are distributed between 30% and 67%. The SVM-based method, the Gauss-based FA model, and the AGC quickly dropped below 50%, indicating a loss of the recognition ability. In summary, the model proposed in this article is more robust to noise and has better engineering practicability than other comparison methods.

VI. CONCLUSION

This article proposes a deep nested neural network for radar HRRP target recognition. In this network, nested CNN, stacked bi-RNN, dynamic adjustment layer, SE block, and multilayer attention mechanism have been applied to avoid problems such as highly redundant input sequences, mutual constraints between the dimensions of each time point and the total length of the sequences, and limited model description capabilities in past RNN models. Second, the model is very robust to small-scale translations that occur in the test sample. Third, when the proposed model is trained with less than 4000 HRRP samples, the average recognition rate of the proposed model is also significantly better than other comparable models. And it is important to point out that when the number of training samples is only 270, the recognition rate of the proposed model for various types of aircraft can remain above 90%. Fourth, the proposed method in this article is more robust to noise than other comparison methods. In summary, the model proposed in this article is very effective and stable in practical application.

REFERENCES

- [1] M. Xing, "Properties of high-resolution range profiles," *Opt. Eng.*, vol. 41, no. 2, p. 493, Feb. 2002.
- [2] J. Pei, Y. Huang, W. Huo, Y. Zhang, J. Yang, and T.-S. Yeo, "SAR automatic target recognition based on multiview deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2196–2210, Apr. 2018.
- [3] L. Du, P. Wang, H. Liu, M. Pan, F. Chen, and Z. Bao, "Bayesian spatiotemporal multitask learning for radar HRRP target recognition," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3182–3196, Jul. 2011.
- [4] K. El-Darymli, E. W. Gill, P. McGuire, D. Power, and C. Moloney, "Automatic target recognition in synthetic aperture radar imagery: A state-of-the-art review," *IEEE Access*, vol. 4, pp. 6014–6058, 2016.
- [5] B. Ding, G. Wen, X. Huang, C. Ma, and X. Yang, "Target recognition in synthetic aperture radar images via matching of attributed scattering centers," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 7, pp. 3334–3347, Jul. 2017.
- [6] Z. Zhang, Z. Tian, and M. Zhou, "Latern: Dynamic continuous hand gesture recognition using FMCW radar sensor," *IEEE Sensors J.*, vol. 18, no. 8, pp. 3278–3289, Apr. 2018.
- [7] A. Karine, A. Toumi, A. Khenchaf, and M. El Hassouni, "Target recognition in radar images using weighted statistical dictionary-based sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2403–2407, Dec. 2017.
- [8] R. A. Mitchell and J. J. Westerkamp, "Robust statistical feature based aircraft identification," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, no. 3, pp. 1077–1094, Jul. 1999.
- [9] L. Shi, P. Wang, H. Liu, L. Xu, and Z. Bao, "Radar HRRP statistical recognition with local factor analysis by automatic Bayesian Ying-Yang harmony learning," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 610–617, Feb. 2011.
- [10] L. Du, H. Liu, and Z. Bao, "Radar HRRP statistical recognition: Parametric model and model selection," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1931–1944, May 2008.
- [11] 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.

- [12] P. López-Rodríguez, D. Escot-Bocanegra, R. Fernández-Recio, and I. Bravo, "Non-cooperative target recognition by means of singular value decomposition applied to radar high resolution range profiles," *Sensors*, vol. 15, no. 1, pp. 422–439, 2014.
- [13] H. 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, Feb. 2016.
- [14] D. Zhou, "Radar target HRRP recognition based on reconstructive and discriminative dictionary learning," *Signal Process.*, vol. 126, pp. 52–64, Sep. 2016.
- [15] 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*, Chengdu, China, Oct. 2011, pp. 642–645.
- [16] W. Xiong, G. Zhang, S. Liu, and J. Yin, "Multiscale kernel sparse coding-based classifier for HRRP radar target recognition," *IET Radar, Sonar Navigat.*, vol. 10, no. 9, pp. 1594–1602, Dec. 2016.
- [17] L. Li and Z. Liu, "Noise-robust HRRP target recognition method via sparse-low-rank representation," *Electron. Lett.*, vol. 53, no. 24, pp. 1602–1604, 2017.
- [18] 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.
- [19] Y. Jiang, Y. Han, and W. Sheng, "Target recognition of radar HRRP using manifold learning with feature weighting," in *Proc. IEEE Int. Workshop Electromagn., Appl. Student Innov. Competition (iWEM)*, Nanjing, China, May 2016, pp. 1–3.
- [20] H. Zhang, D. Ding, Z. Fan, and R. Chen, "Adaptive neighborhood-preserving discriminant projection method for HRRP-based radar target recognition," *IEEE Antennas Wireless Propag. Lett.*, vol. 14, pp. 650–653, 2015.
- [21] L. Du, H. Liu, Z. Bao, and M. Xing, "Radar HRRP target recognition based on higher order spectra," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2359–2368, Jul. 2005.
- [22] M. Pan, L. Du, P. Wang, H. Liu, and Z. Bao, "Multi-task hidden Markov modeling of spectrogram feature from radar high-resolution range profiles," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, p. 86, Dec. 2012.
- [23] J. P. Zwart, R. van der Heiden, S. Gelsema, and F. Groen, "Fast translation invariant classification of HRR range profiles in a zero phase representation," *IEE Proc.-Radar, Sonar Navigat.*, vol. 150, no. 6, pp. 411–418, Dec. 2003.
- [24] M. Pan, J. Jiang, Q. Kong, J. Shi, Q. Sheng, and T. Zhou, "Radar HRRP target recognition based on t-SNE segmentation and discriminant deep belief network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 9, pp. 1609–1613, Sep. 2017.
- [25] P. Mian, J. Jie, L. Zhu, C. Jing, and Z. Tao, "Radar HRRP recognition based on discriminant deep autoencoders with small training data size," *Electron. Lett.*, vol. 52, no. 20, pp. 1725–1727, Sep. 2016.
- [26] B. Feng, B. Chen, and H. Liu, "Radar HRRP target recognition with deep networks," *Pattern Recognit.*, vol. 61, pp. 379–393, Jan. 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320316302254>
- [27] C. Du, B. Chen, B. Xu, D. Guo, and H. Liu, "Factorized discriminative conditional variational auto-encoder for radar HRRP target recognition," *Signal Process.*, vol. 158, pp. 176–189, May 2019.
- [28] 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.*, 2017, pp. 356–367.
- [29] 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. 4th Int. Conf. Ubiquitous Positioning, Indoor Navigat. Location Based Services (UPINLBS)*, Nov. 2016, pp. 283–288.
- [30] 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.
- [31] C. Guo, H. Wang, T. Jian, Y. He, and X. Zhang, "Radar target recognition based on feature Pyramid fusion lightweight CNN," *IEEE Access*, vol. 7, pp. 51140–51149, 2019.
- [32] L. Yuan, "A time-frequency feature fusion algorithm based on neural network for HRRP," *Prog. Electromagn. Res. M*, vol. 55, pp. 63–71, 2017.
- [33] O. Karabayir, O. M. Yucedag, M. Z. Kartal, and H. A. Serim, "Convolutional neural networks-based ship target recognition using high resolution range profiles," in *Proc. 18th Int. Radar Symp. (IRS)*, Prague, Czech Republic, Jun. 2017, pp. 1–9.
- [34] C. Zhao, X. He, J. Liang, T. Wang, and C. Huang, "Radar HRRP target recognition via semi-supervised multi-task deep network," *IEEE Access*, vol. 7, pp. 114788–114794, 2019.
- [35] J. Wan, B. Chen, B. Xu, H. Liu, and L. Jin, "Convolutional neural networks for radar HRRP target recognition and rejection," *EURASIP J. Adv. Signal Process.*, vol. 2019, no. 1, p. 5, Dec. 2019.
- [36] B. Xu, B. Chen, J. Wan, H. Liu, and L. Jin, "Target-aware recurrent attentional network for radar HRRP target recognition," *Signal Process.*, vol. 155, pp. 268–280, Feb. 2019.
- [37] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.
- [38] M. Peters *et al.*, "Deep contextualized word representations," Feb. 2018, *arXiv:1802.05365*. [Online]. Available: <https://arxiv.org/abs/1802.05365>
- [39] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*. [Online]. Available: <http://arxiv.org/abs/1810.04805>
- [40] D. Amodei *et al.*, "Deep speech 2: End-to-end speech recognition in English and Mandarin," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 173–182.
- [41] G. Morrone, S. Bergamaschi, L. Pasa, L. Fadiga, V. Tikhanoff, and L. Badino, "Face landmark-based speaker-independent audio-visual speech enhancement in multi-talker environments," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 6900–6904.
- [42] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, vol. 25, no. 2, pp. 1097–1105.
- [43] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 386–397, Feb. 2020.
- [44] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*. [Online]. Available: <http://arxiv.org/abs/1804.02767>
- [45] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proc. IEEE Int. Symp. Circuits Syst.*, May/Jun. 2010, pp. 253–256.
- [46] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.
- [47] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- [48] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [49] L. Du, H. Liu, Z. Bao, and J. Zhang, "A two-distribution compounded statistical model for radar HRRP target recognition," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2226–2238, Jun. 2006.
- [50] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12 no. 10, pp. 2825–2830, 2012.
- [51] W. Xiao-Dan and W. Ji-Qin, "Support vector machine for HRRP classification," in *Proc. 7th Int. Symp. Signal Process. Appl.*, vol. 1. Paris, France, Jul. 2003, pp. 337–340.
- [52] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, Apr. 2011, doi: [10.1145/1961189.1961199](https://doi.org/10.1145/1961189.1961199).



Mian Pan received the B.S. degree in electronic engineering from Xidian University, Xi'an, China, in 2007, and the Ph.D. degree in signal processing from Xidian University, Xi'an, China, in 2013.

He is an Associate Professor at Hangzhou Dianzi University, Hangzhou, China.

His research interests include machine learning, radar signal processing, pattern recognition, audio, video, and image processing algorithms.



Ailin Liu received the B.S. degree in electronic information from Hangzhou Dianzi University, Hangzhou, China, in 2018, where she is pursuing the master's degree in electronic science and technology.

Her major research interests are radar signal processing and deep learning.



Yanzhen Yu received the B.S. degree in electronic information science and technology from the He'nan University of Technology, Zhengzhou, China, in 2017, and the M.S. degree in electronics and communication engineering from Hangzhou Dianzi University, Hangzhou, China, in 2020.

Her major research interests are image processing, radar signal processing, and machine learning.



Yan Liu received the B.S. degree in electronic engineering from Xidian University, Xi'an, China, in 2006, and the Ph.D. degree in signal and information processing from Xidian University, Xi'an, China, in 2012.

She is working at Xidian University and Hangzhou Dianzi University as an Associate Professor. Her major research interests are remote sensing image processing, pattern recognition, and radar moving target detection.



Penghui Wang (Member, IEEE) received the B.S. degree in communication engineering from the National University of Defense Technology (NUDT), Changsha, China, in 2005, and the Ph.D. degree in signal processing from Xidian University, Xi'an, China, in 2012.

He is an Associate Professor with the National Laboratory of Radar Signal Processing, Xidian University. His research interests include radar signal processing and automatic target recognition.



Shuaishuai Lv was born in Kaifeng, He'nan Province, China in 1986. He received the B.E. and Ph.D. degrees from the School of Automation, Northwestern Polytechnical University, in 2011 and 2016, respectively.

Since 2016, he has been a Lecturer with the College of Electronics and Information, Hangzhou Dianzi University. He is the author of more than 20 articles, and more than 10 inventions. His research interests include intelligent control, remote sensing image processing, and target detection based on deep learning.



Jianjun Li received the B.S. degree in information engineering from Xidian University, Xi'an, China, and the M.S. and Ph.D. degrees in electrical and computer from the University of Western Ontario and the University of Windsor, Canada, in 1996 and 2000, respectively.

He is working at Hangzhou Dianzi University as a Chair Professor. His research interests include computer vision algorithms, microelectronics, audio, video, and image processing algorithms and implementation.



He Zhu was born in 1990 in Shandong province. He received the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2017.

He visited Marburg University in Germany from 2016 to 2017. He is an Associate Professor with Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing, China.

His main research field includes signal processing, deep learning, image processing, photodetectors, infrared image sensor, optical metamaterial, etc.

At present, his interests are silicon-based room temperature infrared detectors, machine learning in sensor design and intelligent sensors based on edge computing.