Research Article

Radar Signal Recognition and Localization Based on Multiscale Lightweight Attention Model

Weijian Si,^{1,2} Jiaji Luo^{ID},^{1,2} and Zhian Deng^{ID}^{1,2}

¹College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China

²Key Laboratory of Advanced Marine Communication and Information Technology, Ministry of Industry and Information Technology, Harbin 150001, China

Correspondence should be addressed to Zhian Deng; dengzhian@hrbeu.edu.cn

Received 3 November 2021; Accepted 25 January 2022; Published 14 February 2022

Academic Editor: Abdellah Touhafi

Copyright © 2022 Weijian Si et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The recognition technology of the radar signal modulation mode plays a critical role in electronic warfare, and the algorithm based on deep learning has significantly improved the recognition accuracy of radar signals. However, the convolutional neural networks became increasingly sophisticated with the progress of deep learning, making them unsuitable for platforms with limited computing resources. ResXNet, a novel multiscale lightweight attention model, is proposed in this paper. The proposed ResXNet model has a larger receptive field and a novel grouped residual structure to improve the feature representation capacity of the model. In addition, the convolution block attention module (CBAM) is utilized to effectively aggregate channel and spatial information, enabling the convolutional neural network model to extract features more effectively. The input time-frequency image size of the proposed model is increased to 600×600 , which effectively reduces the information loss of the input data. The average recognition accuracy of the proposed model achieves 91.1% at -8 dB. Furthermore, the proposed model performs better in terms of unsupervised object localization with the class activation map (CAM). The classification information and localization information of the radar signal can be fused for subsequent analysis.

1. Introduction

Radar is widely deployed on the current battlefield and has progressively become the dominant key technology in modern warfare as a result of the continuous improvement of radar technology [1–4]. Therefore, recognizing the modulation type of enemy radar signals rapidly and accurately can effectively obtain battlefield information and situation and provide decent support for subsequent decision-making. It is significantly vital in the field of electronic warfare.

Traditional radar signal recognition methods rely on handcraft features extraction [5–11]. However, these methods lack flexibility and are computationally inefficient. With the continuous development of radar technology, radar signal parameters have become more complex, and radar signals have become more concealed. In addition, the widespread application of radar equipment and the rapid increase in the number of radiation sources have made the electromagnetic environment on the battlefield more com-

plicated. Therefore, the traditional radar signal recognition method cannot effectively recognize radar signal modulation in the complex electromagnetic environment.

Deep learning has progressed rapidly in the last few years, and it has been extensively used in a wide range of traditional applications. Radar signal recognition methods based on convolutional neural networks have surpassed traditional recognition methods based on handcraft feature extraction [2, 7, 12–14]. The convolutional neural networks automatically extract the deep features of objects through supervised learning and have significant generalization performance [15]. Besides, the convolutional neural network has a hierarchical structure, and the model structure and parameters can be adjusted arbitrarily, which significantly reduces labor costs and is more convenient to use. To achieve higher recognition accuracy, the convolutional neural network becomes larger and the model structure tends to be more complicated. On the other hand, it is also important to strike a balance between recognition speed and computational efficiency in practical

applications. Therefore, this paper proposes a new multiscale lightweight structure, ResXNet, which has the advantages of lightweight and high computational efficiency.

In addition, the convolution block attention module (CBAM) attention mechanism is utilized in the model proposed in this paper. CBAM is a lightweight general attention module that can be seamlessly integrated into any convolutional neural network structure and trained end-to-end together with the basic convolutional neural networks [16–18]. The convolutional layer can only capture local feature information, ignoring the context relationship of features outside the receptive field. The CBAM significantly improves the feature representation capability of the model by enhancing or suppressing specific features in the channel and spatial dimension. And its calculation and memory overhead can be ignored.

At the same time, this paper also investigates the application of object localization based on class activation mapping (CAM) in radar signal recognition. Class activation mapping is a weakly supervised localization algorithm that locates the object position in a single forward pass, which improves the interpretability and transparency of the model and helps researchers build trust in the deep learning models [19, 20].

Therefore, this paper proposes a multiscale ResXNet model based on grouped residual modules and further improves the recognition accuracy of the model through the CBAM attention module. The ResXNet lightweight attention network model proposed in this paper is based on grouped convolution and constructs a hierarchical connection similar to residuals within a single convolution block. ResXNet can expand the size of the receptive field and improve the multiscale feature representation ability. The grouped residual convolutional layers effectively reduce the number of parameters while also improving the generalization performance. In addition, CAM is used to obtain the localization information of the radar signal in the time-frequency image, and the classification information and localization information of the radar signal can be fused for subsequent analysis.

2. Related Work

2.1. Radar Signal Classification. Traditional radar signal recognition methods usually rely on handcrafted feature extraction, such as cumulants, distribution distance, spectral correlation analysis, wavelet transform, and time-frequency distribution features [21]. Machine learning algorithms, such as clustering algorithms [22], support vector machines [9, 23], decision trees [7], artificial neural networks [5], and graph models [24], are used to classify radar signals according to the extracted features. However, the traditional radar signal recognition method is inefficient since it relies significantly on manual feature extraction and selection. And it is affected by noise easily, and the recognition performance substantially decreases in low SNR.

Convolutional neural networks have found their way into the field of radar signal classification as the development of deep learning. The radar signal recognition based on convolutional neural networks first converts one-dimensional

time-domain radar signals into two-dimensional time-frequency images through time-frequency analysis and then automatically extracts features from time-frequency images of different radar signals by training convolutional neural networks. The radar signal recognition based on the convolutional neural network significantly improves the recognition accuracy in low SNRs.

Kong et al. [3] proposed a convolutional neural network (CNN) for radar waveform recognition and a sample averaging technique to reduce the computational cost of time-frequency analysis. However, the input size of this method is small, and there is a significant loss of information. Hoang et al. [13] introduced a radar waveform recognition technique based on a single shot multibox detector and a supplementary classifier, which achieved extraordinary classification performance. However, this method requires much manual annotation, and the computational efficiency is low. Wang et al. [12] proposed a transferred deep learning waveform recognition method based on a two-channel architecture, which can significantly reduce the training time and the size of the training dataset, and multiscale convolution and time-related features are used to improve the recognition performance. On the other hand, the transfer learning method requires a large convolutional neural network as a pretraining model, which has a high computational cost and is incompatible for embedded platforms or platforms with limited computing resources.

2.2. Convolutional Neural Network. In 1995, LeNet created the history of deep convolutional neural networks. AlexNet [25], the first deep convolutional neural network structure, achieved breakthrough success in image classification and recognition applications in 2012. The VGG [26] model modularizes the convolutional neural network structure, increases the network depth, and uses 3×3 small-size convolution kernels. Experiments show that expanding the receptive field by increasing the depth of the convolutional neural network can effectively improve performance [27]. The GoogLeNet model utilizes parallel filters with varied convolution kernel sizes to increase the feature representation ability and recognition performance [15]. ResNet [28] presents a 152-layer deep convolutional neural network that incorporates the identity connection into the convolutional neural network topology, alleviating the vanishing gradient problem.

With the continuous development of deep learning, the depth of the convolutional neural network is deeper, the calculation is more sophisticated, and the requirements for hardware are higher to achieve higher accuracy. As a result, the construction of small and efficient convolutional neural networks has gained more attention. DenseNet [29] connects the output of each layer to each subsequent layer. All previous layers serve as inputs for each convolutional layer, and the output features serve as inputs for all subsequent layers. DenseNet enables the network to extract features on a larger scale and alleviate the vanishing gradient problem. MobileNet [30] employs a depthwise separable convolution to build lightweight convolutional neural networks. The advantages of MobileNet include a tiny model,

lower latency, lower computing complexity, and higher inference efficiency. It can easily match the requirements for platforms with limited computing resources and embedded applications.

Grouped convolution was first introduced in AlexNet [25] for distributing the convolutional neural network model over multiple GPU resources. ResNeXt [31] found that grouped convolution can reduce the number of parameters and simultaneously improve the accuracy. Channel-wise convolution is a special case of grouped convolution in which the number of groups is equal to the number of channels. The channel-wise convolutions are components of depth separable convolution [30].

3. Proposed Method

The overview of the proposed algorithm is depicted in Figure 1. The proposed algorithm first converts the radar signal into a time-frequency image through time-frequency analysis. The ResXNet model proposed in this paper is presented for radar signal recognition, and the CAM is utilized for signal localization in time-frequency images. The ResNet model is composed of a grouped residual module and a CBAM attention mechanism.

3.1. Radar Signal Processing. In this paper, the radar signal interfered by additive white Gaussian noise can be expressed as

$$y(t) = x(t) + N(t) = Ae^{j\theta(t)} + N(t), \quad (1)$$

where $x(t)$ signifies complex radar signal samples and $N(t)$ stands for additive Gaussian white noise (AWGN) with zero mean value and variance σ^2 . A represents the nonzero constant amplitude, and $\theta(t)$ denotes the instantaneous phase of the radar signal. The inherent difference between radar signals of different modulation types is the frequency variation over time. The one-dimensional radar signal is transferred into two-dimensional time-frequency images (TFIs) through time-frequency analysis. The pattern in the time-frequency image corresponds to the frequency variation with time.

The instantaneous phase $\theta(k)$ consists of instantaneous frequency $f(k)$ and the phase function $\phi(k)$, which determine the modulation type of the radar signal. The instantaneous phase $\theta(k)$ is defined as

$$\theta(k) = 2\pi f(k)(kT_s) + \phi(k). \quad (2)$$

Eight LPI radar waveforms considered in this paper are grouped into two categories, FM (frequency modulation) and PM (phase modulation). In the FM, the instantaneous frequency $f(k)$ varies while the phase $\phi(k)$ is constant; in the PM, the phase $\phi(k)$ varies while the instantaneous frequency $f(k)$ is constant [3], as defined in Table 1.

The Choi-Williams distribution (CWD) based on the time-frequency distribution of Cohen's class has the advantages of high resolution and cross-term suppression. The resolution of the time-frequency analysis can be modified

by adjusting the parameters of its exponential kernel function. The time-frequency image of radar signals based on Choi-Williams distribution is defined as

$$C(t, \omega) = \frac{1}{2\pi} \iiint_{\infty} f(\theta, \tau) x(s + \tau/2) x^*(s - \tau/2) e^{-j(\theta t + \omega \tau - s\theta)} ds d\tau d\theta, \quad (3)$$

where t and ω denote frequency and time axes, respectively. And $f(\theta, \tau)$ is the exponential kernel function of the Choi-Williams distribution. The kernel function is regarded as a low-pass filter that can suppress cross-terms effectively.

$$f(\theta, \tau) = \exp \left[\frac{-\theta^2 \tau^2}{\sigma} \right]. \quad (4)$$

Figure 2 depicts the time-frequency images of different radar signals by Choi-Williams distribution considered in this paper. The time-frequency images visualize the frequency variation over time and thus recognize the radar signals effectively. Before feature extraction, time-frequency images are normalized to reduce the influence of the bandwidth of distinct radar signals. The time-frequency images are transformed to gray images as follows:

$$G(x, y) = \frac{I(x, y) - \min(I(x, y))}{\max(I(x, y)) - \min(I(x, y))}, \quad (5)$$

where $I(x, y)$ indicates the time-frequency image by Choi-Williams distribution, $G(x, y)$ is the gray images, and (x, y) denotes each pixel in the time-frequency images. The gray time-frequency images contain significant components and information of radar signals.

3.2. ResXNet Module. The traditional convolutional neural network expands the receptive field size of the model by simply stacking convolutional layers. This strategy, on the other hand, increases the size of the model and the number of parameters, making the training of the model increasingly complex. The development of advanced model structures reveals a tendency toward improving the receptive field size and multiscale learning capability of the models while maintaining lightweight.

A multiscale model based on group convolution is proposed in this paper. On the premise of keeping the model lightweight, the receptive field of each convolutional layer is increased to improve the capabilities of feature extraction and multiscale representation. The proposed model structure is a modular design, which can flexibly adjust the size and parameters of the model.

Grouped convolution significantly reduces the size of the model and the number of parameters. Grouped convolution splits the input feature maps along the channel dimension into several feature map subsets, and each branch can only use a subset of the feature map and cannot use the entire input feature map [27]. The ResXNet proposed in this paper employs channel-wise convolution to each branch, and each branch takes the complete input feature map as input. 1×1

FIGURE 1: The overview of radar signal classification and localization based on ResXNet and CAM.

TABLE 1: LPI radar waveform.

Modulation type	$f(k)$	$\phi(k)$
LFM	$f_0 + (B/\tau_{pw})(kT_s)$	Constant
Costas	f_j	Constant
BPSK	Constant	0 or π
Frank	Constant	$(2\pi/M)(i-1)(j-1)$
P1	Constant	$-(\pi/M)[(M-(2j-1))][(j-1)M+(i-1)]$
P2	Constant	$-(\pi/2M)[2i-1-M][2j-1-M]$
P3	Constant	$(\pi/\rho)(i-1)^2$
P4	Constant	$(\pi/\rho)(i-1)^2 - \pi(i-1)$

channel-wise convolution compresses the number of channels of the input feature map, and then a 3×3 convolution with the same number of channels is used for feature extraction. Then, the output feature maps of each branch are concatenated, and features maps of different scales are fused using channel-wise convolution.

As shown in Figure 3, the same feature map is input to each branch, and the 1×1 convolution in each branch is used to compress the number of channels of the feature map, and each branch contains a 3×3 convolutional layer for feature extraction. K_i and C_i represent 3×3 convolution

operation and 1×1 channel-wise convolution, respectively. The 3×3 convolution input of the i th branch is the summation of C_i and K_{i-1} ; thus, the output feature map y_i can be expressed as

$$y_i = \begin{cases} K_i(C_i(x)), & i = 1, \\ K_i(C_i(x) + y_{i-1}), & 2 \leq i \leq s, \end{cases} \quad (6)$$

where x is the input feature map, and the convolution operation of each branch extracts features from the input

FIGURE 2: Time-frequency images of different radar waveforms.

FIGURE 3: Schematic representation of ResXNet modules ($S = 4$).

features and the output of the preceding branches. Therefore, the multibranch structure and the connection of different branches are beneficial to extract global and local features.

In ResXNet, S refers to the number of branches of each convolution module, larger S can learn the features of larger receptive field size, and the calculation and memory overhead introduced is negligible. ResXNet further improves the multiscale capability of convolutional neural networks and can be integrated with existing state-of-the-art methods to improve recognition accuracy and generalization performance. The proposed ResXNet model can be regarded as an improvement of Res2Net [27]. Changing the grouping convolution in the Res2Net model to several 1×1 channel-wise convolutions makes the parameters and structural design of the model more flexible.

3.3. Convolution Block Attention Module. The convolution block attention module (CBAM) is an attention module for convolutional neural networks that is simple and efficient [18]. The diagram of CBAM is depicted in Figure 4. CBAM sequentially calculates the channel and spatial attention of the feature map and then multiplies the two attention maps with the input feature map to refine the adaptive feature. CBAM is a lightweight general module that can be seamlessly inserted into any convolutional neural network architecture, with negligible computing and memory overhead, and can be trained end-to-end together with the base convolutional neural network.

The attention module allows the model to concentrate on informative features and suppress irrelevant features. CBAM applies the channel attention module and the spatial attention module sequentially to enable the model to

FIGURE 4: The convolution block attention module (CBAM).

reinforce effective features. Given the feature map $F \in \mathbb{R}^{C \times H \times W}$ as input, CBAM calculates one-dimensional channel attention map $M_c \in \mathbb{R}^{C \times 1 \times 1}$ and two-dimensional spatial attention map $M_s \in \mathbb{R}^{1 \times H \times W}$. The overall attention module is expressed as

$$\begin{aligned} F' &= M_c(F) \otimes F, \\ F'' &= M_s(F) \otimes F', \end{aligned} \quad (7)$$

where $\otimes$ is the element-wise multiplication and F'' denotes the feature map of the final CBAM attention module output. In the multiplication process, the attention feature is broadcasted accordingly: the channel attention value is broadcast along the spatial dimension and vice versa [18]. The specific details of channel attention and spatial attention are introduced as follows.

3.3.1. Channel Attention Module. The channel attention module (CAM) exploits the interchannel relationship of the feature maps to generate the channel attention map. Figure 5 depicts the computation process of the channel attention map. Average pooling and maximum pooling are used to aggregate spatial information, then the fully connected layer is used to compress the channel dimensions of the feature map, and the multilayer perceptron is used to generate the final attention map.

Specifically, first, we use global average pooling and global maximum pooling to aggregate the spatial context information of the feature mapping to generate two different spatial descriptors, which denote the global average pooling feature F_{avg}^c and the global maximum pooling feature F_{max}^c . Then, the two feature descriptors are forwarded into a shared network to generate channel attention map $M_c \in \mathbb{R}^{C \times 1 \times 1}$. The shared network is a multilayer perceptron that contains one hidden layer. To reduce the computational overhead, the activation unit size of the hidden layer multilayer perceptron is set to $\mathbb{R}^{c/r \times 1 \times 1}$, where r is the reduction rate. The shared network is applied to the different spatial descriptors, and then, the two feature descriptors are merged using an element-wise summation to obtain the final channel feature vector. Therefore, the channel attention can be written as

$$\begin{aligned} M_c(F) &= \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \\ &= \sigma\left(W_1\left(W_0\left(F_{\text{avg}}^c\right)\right) + W_1\left(W_0\left(F_{\text{max}}^c\right)\right)\right), \end{aligned} \quad (8)$$

where σ represents the sigmoid activation function, W_0

$\in \mathbb{R}^{c/r \times c}$ and $W_1 \in \mathbb{R}^{c \times c/r}$ denote the weights of MLP, and the ReLU activation function is used in each layer.

3.3.2. Spatial Attention Module. The spatial attention module (SAM) utilizes the interspatial relationship of feature maps to generate spatial attention maps. As illustrated in Figure 6, the average pooling and maximum pooling are applied to the feature map along the channel dimension, and then, the two feature maps are concatenated to create the spatial feature descriptor. The spatial attention map highlights informative regions [18]. Finally, the spatial attention map is generated by convolution operation on the spatial feature descriptor to enhance or suppress the feature region.

Specifically, using spatial global average and max pooling to calculate the spatial information of the feature map, two 2-dimensional spatial feature maps $F_{\text{avg}}^s \in \mathbb{R}^{1 \times H \times W}$ and $F_{\text{max}}^s \in \mathbb{R}^{1 \times H \times W}$ are generated. Then, the 2D spatial attention map is calculated by a convolution operation. The spatial attention is computed as

$$\begin{aligned} M_s(F) &= \sigma\left(f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])\right) \\ &= \sigma\left(f^{7 \times 7}\left(\begin{bmatrix} F_{\text{avg}}^s \\ F_{\text{max}}^s \end{bmatrix}\right)\right), \end{aligned} \quad (9)$$

where σ represents sigmoid activation function and $f^{7 \times 7}$ denotes a convolution operation with the kernel size of 7×7 .

The convolution block attention module (CBAM) divides attention features into the channel and spatial attention modules and achieves a significant performance improvement while keeping a small overhead. CBAM can be seamlessly integrated into any convolutional neural network architecture and trained end-to-end with the CNN model. The CBAM can prompt the network to learn and aggregate the feature information in the target area, effectively strengthen or suppress the features of a specific space or a specific channel, and guide the convolutional neural network to make good use of the feature maps [18].

3.4. Class Activation Map. The class activation map (CAM) is portable and applied to a variety of computer vision tasks for weakly supervised object localization. The CAM is trained end-to-end based on image-level annotation and localizes objects simply in a single forward pass. CAM avoids the flattening of the feature map by replacing the fully connected layer with the global average pooling, which completely preserves the spatial information of the objects in the embedded features. More importantly, CAM can make the existing state-of-the-art deep models interpretable and transparent and help researchers understand the logic of predictions hidden inside the deep learning models.

For each image I , $A = F(I)$ denotes the feature map of the last convolutional layer. Suppose that there are K feature maps in the last convolution layer. The $A_k(x, y)$ indicates the activation of the k th channel at spatial coordinate (i, j) , where $k = 1, 2, \dots, K$. Each node of the GAP (global average pooling) layer F_k is spatial average of the activation and can be computed by

FIGURE 5: Channel attention module.

FIGURE 6: Spatial attention module.

$$F_k = \sum_{i,j} A_k(x, y). \quad (10)$$

The function definition of CAM is

$$M_c(x, y) = \sum_k w_k^c A_k(x, y), \quad (11)$$

where M_c is the class activation map for class c and w_k^c represents the weight corresponding to class c for unit k . $M_c(x, y)$ signifies the importance of the activation at spatial grid (x, y) for class c .

Each class activation map consists of the weighted linear sum of these visual patterns at different spatial locations, which contain a series of part saliency maps. Upsampling is applied to resize the class activation map to the size of the input time-frequency image to localize the image saliency regions most relevant to a particular class.

3.5. ResXNet Architecture. The group dimension indicates the number of groups within a convolutional layer. This dimension converts single-branch convolutional layers to multibranch, improving the capacity of multiscale representation [31]. The CBAM block adaptively recalibrates channel and spatial feature responses by explicitly modeling interdependencies among channel and spatial [18].

Table 2 shows the specification of ResXNet, including depth and width. ResXNet makes extensive use of 3×3 convolution and uses average pooling layers to reduce the feature size. The proposed models are constructed in 5 stages, and the first stage is a 7×7 conventional convolution block with stride = 2. The following 4 stages, respectively, contain [1, 2, 4] ResXNet modules to construct the proposed ResXNet model. The number of channels in each stage of

TABLE 2: Architectural specification of ResXNet.

Stage	Output size	Channels	Modules
Input	600×600	3	—
0	300×300	32	—
1	150×150	32	2
2	75×75	64	2
3	37×37	128	4
4	18×18	256	1

TABLE 3: Model information.

S	Parameters	Model size
1	5.8 M	22.5 M
2	3.2 M	13.0 M
4	1.9 M	8.5 M
8	1.3 M	6.7 M

the module is [32, 64, 128, 256], respectively. The global average pooling is followed by a fully connected layer as the head for the classification task. A dropout layer is inserted before the fully connected layer to prevent the model from overfitting. A CBAM attention mechanism is implemented after the output of each ResXNet module to improve the capability of global feature extraction.

Each convolutional module is composed of the convolutional layer without bias, batch normalization layer, and ReLU activation function. The input size of the proposed model is 600×600 . The last convolutional layer connects the fully connected layer through global average pooling (GAP) instead of flattening the features, such that changes

TABLE 4: Radar signal parameter variation.

	Parameter	Data scale
General parameters	Sampling frequency f_s	1
	Carrier frequency f_c	$U(0, 1/4)$
Linear frequency modulation	Bandwidth Δf	$U(1/12, 1/4)$
	Modulation period τ	(600, 800)
Binary phase codes	Carrier frequency f_c	$U(1/12, 1/4)$
	Subpulse frequency	$f_c/4$
	Modulation period	$U(350, 650)$
Costas code	Number of subpulse N	4
	Lowest frequency f_0	1/24
	Frequency hop Δf	$U(1/24, 1/12)$
	Modulation period τ	$U(400, 600)$
Polyphase codes	Carrier frequency f_c	$U(1/12, 1/4)$
	Subpulse frequency	$f_c/4$

in the input size do not affect the number of parameters. Table 3 shows the model information with different groups, including model size and the number of parameters. When $S = 1$, the ResXNet model degenerates to a normal conventional convolutional neural network.

4. Simulation and Analysis

The proposed ResNet model was evaluated and compared with the Res2Net and GroupNet on the simulation dataset to verify the effectiveness and robustness. In this section, we evaluate the performance of the proposed ResXNet model to recognize radar signals with different modulations and the ability of the class activation map (CAM) to localize radar signals in time-frequency images. We implement the proposed models using the TensorFlow 2.0 framework. Cross-entropy is used as a loss function. We trained the networks using SGD and a minibatch of 16 on NVIDIA 3090 GPU. The learning rate was initially set to 0.01 and divided by 0.1 every 10 epochs. All models, including the proposed and compared models, are trained for 30 epochs with the same training strategy.

The dataset consists of 8 types of radar signals, including Barker, Costas, Frank, LFM, P1, P2, P3, and P4. In the training dataset, the SNR of each signal category is from -18 dB to 10 dB with the step of 2 dB, and each SNR contains 500 signals, totaling 60000 signal samples in the training dataset. In the test dataset, each SNR contains 100 samples for a total of 12000 signal samples. The parameters variations of each kind of radar signal are listed in Table 4.

4.1. Recognition Accuracy of the Different Models. The first experiment is to compare the effectiveness of group dimensions on model accuracy. As shown in Table 5, the accuracy of the ResXNet improves with the increase of the number of groups. The number of parameters of the ResXNet model decreases as the increase of group; meanwhile, the training

TABLE 5: Recognition accuracy for different numbers of groups.

Input size	Group dimensions	Accuracy	Train time/epoch
224	$S = 2$	88.1%	207 s
	$S = 4$	88.4%	242 s
	$S = 8$	88.6%	325 s
600	$S = 2$	88.4%	555 s
	$S = 4$	88.6%	595 s
	$S = 8$	88.9%	690 s

TABLE 6: Overall recognition accuracy ($S = 8$).

Model	Accuracy
ResXNet	88.9%
Res2Net	87.5%
GroupNet	86.9%

time will gradually increase. Because the convolution output of the proposed model is connected to the classification layer through the global average pooling (GAP) layer, changing the input size of the model does not change the size and the number of parameters of the model. The proposed model has the combined effect of the hierarchical residual structure, which can generate abundant multiscale features and effectively improve the recognition accuracy. Increasing the input data size of the model can significantly reduce the information loss of the input data, which is beneficial to the convolutional neural network to extract the informative features from the time-frequency images.

4.2. Recognition Accuracy of the Different Models. Table 6 lists the overall recognition accuracy of the proposed and compared models for radar signal modulations with the group dimension is 8. Compared with Res2Net and

FIGURE 7: The performance of the proposed ResXNet models under different SNRs.

FIGURE 8: The average recognition accuracy of different models under different SNRs.

GroupNet, ResXNet has higher recognition accuracy. The recognition accuracy of the proposed ResXNet model can achieve 88.9%, which outperforms Res2Net by 1.4%.

Figure 7 depicts the recognition accuracy of the proposed ResXNet model for different radar signals at various

SNRs. The accuracy of Barker code, Frank code, and LFM is still 100% even at -18 dB. The average recognition accuracies of all radar modulations under different SNRs for Res2Net, GroupNet, and proposed ResXNet models are presented in Figure 8. The average accuracy of all radar

FIGURE 9: The confusion matrixes of different models.

modulations is still above 50% at -18 dB. The average recognition accuracy of the model is above 90% at -8 dB. The radar signals are disrupted by Gaussian white noise at low SNR, resulting in misclassifications. As the SNR increases, the features extraction becomes more distinguishable from each other.

Figure 9 illustrates the confusion matrix of different models to further analyse the recognition capability for radar modulation types. Frank code and P3 code, as well as P1 code and P4 code, are similar and easy to confuse. The recognition accuracy of the other four radar signals is 100%.

4.3. Comparison of Object Visualization. The visualization of model features can accurately explain the features learned by the model and provide reasonable explanations for the pre-

diction results. In this paper, the predictions are visualized by the class activation mapping (CAM) [19], which is commonly used for localizing the discriminative regions in image classification. CAM generates the object saliency map by calculating the weighted sum of the feature maps of the last convolutional layer. By simply upsampling the class activation map to the size of the input time-frequency image, we can identify the image regions most relevant to the particular category [19]. CAM can reveal the decision-making process of the convolutional neural network model. CAM makes the model based on the convolutional neural network more transparent by generating visual explanations. And CAM can localize objects without additional bounding box annotations.

FIGURE 10: CAM comparison of different models.

FIGURE 11: CAM results of different input sizes.

Figure 10 depicts the CAM visualization results of the ResXNet model and the grouped convolution model for different radar signals, with the discriminative CAM regions highlighted. Figure 10 shows that the ResXNet model outperforms GroupNet in terms of visual object localization. Compared with GroupNet, the CAM results of ResXNet have more concentrated activation maps. The ResXNet localizes the radar signal more precisely and tends to encompass the entire object in the time-frequency image due to its better multiscale capability. The ResXNet model proposed in this paper can help researchers develop confidence in artificial intelligence systems by improving the interpretability and trustworthiness of the model. This ability to precisely localize discriminative regions makes ResXNet potentially valuable for object region mining in weakly supervised localization tasks of radar signals.

4.4. Comparison of Object Visualization with Different Input Sizes. In addition, the paper also explored the impact of the input size of the proposed models on CAM visualization.

Figure 11 shows the CAM visualization results with various input sizes. As demonstrated in Figure 11, the model can generate more precise CAM with a higher resolution as the input image size increases, improving the visual localization capacity of the model. Especially for the Costas signals, ResXNet can accurately localize each signal component.

Therefore, the multiscale feature extraction capability of ResXNet helps to better localize the most discriminative regions in radar time-frequency images compared to GroupNet. For the same ResXNet model, increasing the input image size of the model can obtain more precise and complete discriminative regions. As a result, the proposed ResXNet model contributes to improving the target localization capability of the model and the interpretability of the prediction.

5. Conclusions

In this paper, a radar signal recognition approach based on the ResXNet lightweight model is proposed. The proposed model substantially improves the multiscale representation

ability and recognition accuracy by grouping and cascading convolutional layers. The CBAM attention mechanism is employed to effectively aggregate the channel features and spatial features, so that the convolutional neural network model can make better use of the given feature maps.

Experiments demonstrate that the proposed model has capable multiscale representations and achieves higher recognition performance in low SNR. And the model is tiny, the number of parameters is low, and it is suitable for the embedded platform application. ResXNet adds the grouped dimension by channel-wise convolution, allowing to adjust the size and number of parameters of the model and improve the multiscale representation capability of the model. The proposed ResXNet model provides a superior capacity for object localization, and the radar signal can be localized more precisely through CAM.

For future research, more lightweight models for radar signal recognition, as well as the use of CAM in radar signal recognition and localization, will be explored.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was financially supported by the National Natural Science Foundation of China (Grant Nos. 61971155 and 61801143), Natural Science Foundation of Heilongjiang Province of China (Grant No. JJ2019LH1760), Fundamental Research Funds for the Central Universities (Grant No. 3072020CF0814), and Aeronautical Science Foundation of China (Grant No. 2019010P6001).

References

- [1] X. Ni, H. L. Wang, F. Meng, J. Hu, and C. K. Tong, "LPI radar waveform recognition based on multi-resolution deep feature fusion," *IEEE Access*, vol. 9, pp. 26138–26146, 2021.
- [2] X. Ni, H. Wang, Y. Zhu, and F. Meng, "Multi-resolution fusion convolutional neural networks for intrapulse modulation LPI radar waveforms recognition," *IEICE Transactions on Communications*, vol. E103B, no. 12, pp. 1470–1476, 2020.
- [3] S. H. Kong, M. Kim, M. Hoang, and E. Kim, "Automatic LPI radar waveform recognition using CNN," *IEEE Access*, vol. 6, pp. 4207–4219, 2018.
- [4] L. T. Wan, R. Liu, L. Sun, H. S. Nie, and X. P. Wang, "UAV swarm based radar signal sorting via multi-source data fusion: a deep transfer learning framework," *Information Fusion*, vol. 78, pp. 90–101, 2022.
- [5] J. Lunden and V. Koivunen, "Automatic radar waveform recognition," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 1, pp. 124–136, 2007.
- [6] C. C. Xu, J. Y. Zhang, Q. S. Zhou, and S. W. Chen, "Modulation classification for radar pulses in low SNR levels with graph features of ambiguity function," in *2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN)*, pp. 779–783, Guangzhou, China, 2017.
- [7] T. R. Kishore and K. D. Rao, "Automatic intrapulse modulation classification of advanced LPI radar waveforms," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 2, pp. 901–914, 2017.
- [8] T. Wan, K. L. Jiang, Y. L. Tang, Y. Xiong, and B. Tang, "Automatic LPI radar signal sensing method using visibility graphs," *IEEE Access, Article*, vol. 8, pp. 159650–159660, 2020.
- [9] G. Vanhoy, T. Schucker, and T. Bose, "Classification of LPI radar signals using spectral correlation and support vector machines," *Analog Integrated Circuits and Signal Processing*, vol. 91, no. 2, pp. 305–313, 2017.
- [10] C. M. Jeong, Y. G. Jung, and S. J. Lee, "Deep belief networks based radar signal classification system," *Journal of Ambient Intelligence and Humanized Computing*, 2018.
- [11] J. Li, G. Zhang, Y. Sun, E. Yang, L. Qiu, and W. Ma, "Automatic intra-pulse modulation recognition using support vector machines and genetic algorithm," in *2017 IEEE 3rd Information Technology and Mechatronics Engineering Conference (ITOEC)*, pp. 309–312, Chongqing, China, 2017.
- [12] Q. Wang, P. F. Du, J. Y. Yang, G. H. Wang, J. J. Lei, and C. P. Hou, "Transferred deep learning based waveform recognition for cognitive passive radar," *Signal Processing*, vol. 155, pp. 259–267, 2019.
- [13] L. M. Hoang, M. Kim, and S. H. Kong, "Automatic recognition of general LPI radar waveform using SSD and supplementary classifier," *IEEE Transactions on Signal Processing*, vol. 67, no. 13, pp. 3516–3530, 2019.
- [14] C. Wang, J. Wang, and X. Zhang, "Automatic radar waveform recognition based on time-frequency analysis and convolutional neural network," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2437–2441, New Orleans, LA, USA, 2017.
- [15] A. Khan, A. Sohail, U. Zahoor, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5455–5516, 2020.
- [16] X. Li, W. H. Wang, X. L. Hu, J. Yang, and I. C. Soc, "Selective kernel networks," in *2019 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 510–519, Long Beach, CA, 2019.
- [17] J. Hu, L. Shen, S. Albanie, G. Sun, and E. H. Wu, "Squeeze-and-excitation networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 2011–2023, 2020.
- [18] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: convolutional block attention module," <http://arxiv.org/abs/1807.06521>.
- [19] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2921–2929, Seattle, WA, 2016.
- [20] K. Huang, F. Meng, H. Li, S. Chen, Q. Wu, and K. N. Ngan, "Class activation map generation by multiple level class grouping and orthogonal constraint," in *2019 Digital Image Computing: Techniques and Applications*, pp. 182–187, Perth, Australia, 2019.

- [21] K. Konopko, Y. P. Grishin, and D. Janczak, "Radar signal recognition based on time-frequency representations and multi-dimensional probability density function estimator," in *2015 Signal Processing Symposium*, Debe, Poland, 2015.
- [22] M. Zhu, W.-D. Jin, J.-W. Pu, and L.-Z. Hu, "Classification of radar emitter signals based on the feature of time-frequency atoms," in *2007 International Conference on Wavelet Analysis and Pattern Recognition*, pp. 1232–1237, Beijing, China, 2007.
- [23] A. Pavy and B. Rigling, "SV-means: a fast SVM-based level set estimator for phase-modulated radar waveform classification," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 191–201, 2018.
- [24] C. Wang, H. Gao, and X. D. Zhang, "Radar signal classification based on auto-correlation function and directed graphical model," in *2016 IEEE International Conference on Signal Processing, Communications and Computing*, Hong Kong, 2016.
- [25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 60, no. 6, pp. 84–90, 2017.
- [26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, <http://arxiv.org/abs/1409.1556>.
- [27] S. H. Gao, M. M. Cheng, K. Zhao, X. Y. Zhang, M. H. Yang, and P. Torr, "Res2Net: a new multi-scale backbone architecture," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, pp. 652–662, 2021.
- [28] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Seattle, WA, 2016.
- [29] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2261–2269, Honolulu, HI, 2017.
- [30] A. G. Howard, M. Zhu, B. Chen et al., "MobileNets: efficient convolutional neural networks for mobile vision applications," <http://arxiv.org/abs/1704.04861>.
- [31] S. N. Xie, R. Girshick, P. Dollar, Z. W. Tu, and K. M. He, "Aggregated residual transformations for deep neural networks," in *30th IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5987–5995, Honolulu, HI, 2017.