CNN Based on Multiscale Window Self-Attention Mechanism for Radar HRRP Target Recognition

Yujia Diao
College of Electronic Science and Technology
National University of Defence Technology
Changsha, China
dyjnudt@nudt.edu.cn

Shuowei Liu
College of Electronic Science and Technology
National University of Defence Technology
Changsha, China
liushuowei@nudt.edu.cn

Xunzhang Gao*
College of Electronic Science and Technology
National University of Defence Technology
Changsha, China
gaoxunzhang@nudt.edu.cn

Afei Liu
College of Electronic Science and Technology
National University of Defence Technology
Changsha, China
2737606186@qq.com

Zhiwei Zhang
College of Electronic Science and Technology
National University of Defence Technology
Changsha, China
514131141@qq.com

2022 7th International Conference on Signal and Image Processing (ICSIP) | 978-1-6654-9563-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICSIP55141.2022.9887058

***Abstract*—In this paper, we propose a convolutional neural network (CNN) with multiscale window self-attention mechanism for radar high-resolution range profile (HRRP) target recognition task. Specifically, we take a one-dimensional CNN (1-D CNN) to extract shallow feature and fully excavate the rich local structural features of HRRP data. Then, we utilize a multiscale window convolutional self-attention mechanism to capture the regional difference of HRRP data. The proposed self-attention module divides the features obtained by CNN into equal width bands through sliding windows. Multi-level features of different regions can be obtained by the continuous expansion of the windows. It can improve the model's attention to target regions and suppress the influence of irrelevant noise. Experimental results show the superiority of the proposed method in HRRP recognition task.**

Keywords—HRRP target recognition, convolutional neural network, self-attention mechanism

I. Introduction

High resolution range profile(HRRP) is the projection of complex returned echo from target scattering center in the radar ray direction. It is formed by the amplitude of the coherent complex returned echoes of the target scatterers in each range cell [1]. Hence, HRRP contains rich information of targets, such as structure size and scattering distribution. Furthermore, one-dimensional HRRP data can be easily obtained, saved and processed. Due to the above advantages, it has gained extensive attention and application in the field of Radar Automatic Target Recognition(RATR) [2].

Radar HRRP target recognition aims to extract features of heterogeneous separability for classification. Traditional hand-crafted methods mainly extract some inherent characteristics of HRRP data, such as the statistical characteristics [3] or the spectral characteristics [4] . Du et al. [5] used Gaussian distribution to model HRRP data and introduced a label cofactor analysis (LA-FA) model to improve the recognition effect. In [6], the feasibility of HRRP classification based on high-order spectral features is studied, where the features are calculated in the form of Euclidean distance in higher-order spectra space. However, these hand-crafted methods rely on prior knowledge, which limits their application.

In order to learn data features autonomously, Pan et al. [7] used t-SNE to preprocess data and then classify with discriminant depth belief network (DDBN), which improved the recognition performance in imbalanced HRRP data. Wan et al. [8] adopt one-dimensional CNN and two-dimensional CNN to capture time-domain features and spectral features respectively. However, the above networks ignore the differences in the contribution of different regions to the recognition tasks, eventually leading to mediocre performance. To solve this problem, attention mechanism [9] has been introduced into the HRRP recognition task to capture the discriminable target region. Zhang et al. [10] introduced the self-attention mechanism into the model to focus on discriminative range cells, which helps to improve the learning ability of Conv-LSTM. Xu et al. [11] proposed a target-aware recurrent attentional network(TARAN), which can make use of temporal dependence within HRRP sequence and assign reasonable weight to each HRRP timestep to find the target area. Although RNN is effective for sequence modeling, the shared parameters in each timestep are inconsistent with the characteristics of HRRP data [2]. Because there is usually noise redundancy on both sides of the HRRP data, it is unreasonable to model the noise and the target area with shared parameters. Chen et al.[12] used bidirectional GRU (Bi-GRU) unit to weight the features of convolution neural network output to obtain the importance of different local features regions. However, it is difficult for GRU units to acquire the global feature awareness.

To addressing these issues, we propose CNN based on multiscale window self-attention mechanism to obtain multi-level features, which enable the model can generate different attention proposals. Our contributions include the following aspects:

1) Instead of original linear operation, we propose a convolutional self-attention mechanism module to obtain translation invariance and inductive biases of CNN.

2) We propose multiscale window self-attention mechanism to capture local features while expanding the receptive field of the window to realize the perception of multi-level features.

3) The superiority of the proposed method is proved by relevant experiments, and the effectiveness of the proposed method is illustrated by feature visualization.

II. METHOD

A. Overall Framework

In this paper, the structure of our model is shown in Fig. 1. The overall framework mainly consists of three components: the feature extraction module, the attention module, and the recognition module. In the recognition module, the extracted features are flattened along the channel dimension after the full connection layer, and then the prediction label is output by softmax normalization. Here, we introduce the feature extraction module and the attention module in detail as follows.

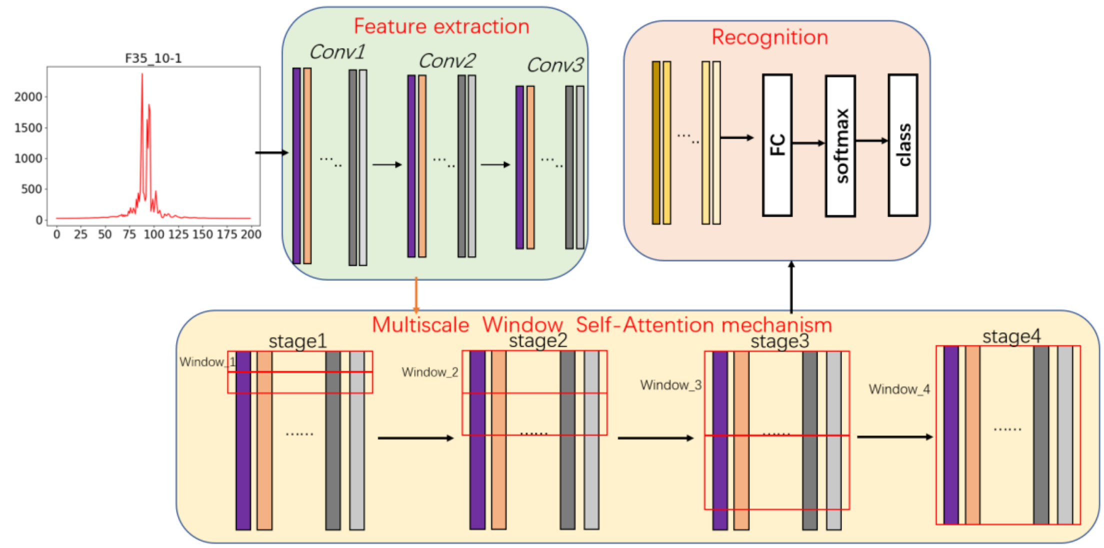

Fig. 1. Network Structure

B. Feature Extraction Based on 1-D CNN module

We build a simple one-dimensional CNN to mine the feature structure in HRRP data and obtain the shallow data feature information. The obtained local features by CNN are independent, which facilitates the extraction of the multi-level deep features using attention mechanism.

We construct a 1-D CNN with L convolution layers, in which has S convolution kernels, each layer has Q channels, and the convolution kernels in L th layer are recorded as K_L. If the input of the L th convolution layer is F_{L-1}, the output of the convolution layer is F_K^{L-1}, the calculation formula is as follows:

$$F_K^{L-1} = f(F_{L-1} * K_L + b_L) \tag{1}$$

where $*$ represents convolution operation, b_L denotes bias, and $f(\bullet)$ is nonlinear function (ReLu).

We use the max-pooling operation to remove all elements except the maximum element in each convolution layer, which reduce the redundancy and increase the convolution receptive field. In addition, we adopt the 1-D wide convolution and the edge of input data can be filled with zeros, so that the edge information of the data can not be lost.

C. Convolutional Self-Attention Module

Since CNN has translation invariance and inductive biases [13], we introduce it into self-attention, so it is no longer necessary to add position embedding. In this paper, query vector Q, key vector K and value vector V are obtained through the convolution operation:

$$\begin{aligned} q &= Conv1d(F) \\ k &= Conv1d(F) \\ v &= Conv1d(F) \end{aligned} \tag{2}$$

where $Conv1d(\bullet)$ denotes 1-D convolution, q,k,v represents the features obtained by 1-D convolution.

Similar to [14], the self-attention function is calculated as:

$$Attention(q,k,v) = soft\max(qk^T / \sqrt{d_k})v \tag{3}$$

where d_k are the dimensions of the key vector q.

Here, Multi-Head Self-Attention mechanism is adopted to stack multiple single-layer attention mechanisms and the function is described as:

$$\begin{aligned} MultiHead(q,k,v) &= (head_1, head_2, ..., head_h)W^o \\ head_i &= Attention(qW_i^q, kW_i^k, vW_i^v) \end{aligned} \tag{4}$$

where h denotes the number of heads, W_i^q, W_i^k, W_i^v represents projection matrices, W^o denotes the transformation matrix.

D. Multiscale Window Self-Attention Mechanism

Generally, the range cell echo in the HRRP target area reflects the structural information of the target, which is useful for the identification task. The receptive field of the attention mechanism methods in the previous literature is usually small, and only the attention weight generated in the single step-time of HRRP sequence is considered [11,12]. To make the attention area larger and fully consider the contribution of different regional features, we propose a multiscale window convolution self-attention module by continuously enlarging the size of the window to expand the receptive field.

We divide the features extracted by 1-D CNN into several local features by horizontal sliding window. We set the length of the window to M . The HRRP feature F can be divided into several segments $[F_1, F_2, ...F_n]$, where $length(F_i) = M$. Then the self-attention mechanism is calculated as follows:

$$\begin{aligned} \tilde{F}_k^i &= Attention(F_i W_k^Q, F_i W_k^K, F_i W_k^V) \\ Attention_{all}(F) &= [\tilde{F}_k^1, \tilde{F}_k^2, ..., \tilde{F}_k^n] \end{aligned} \tag{5}$$

The calculation of $Attention_{all}(F)$ is the same as section C above.

III. EXPERIMENTS

A. Dataset

In this paper, we employ datasets that simulated three different types of aircraft HRRP to evaluate our method. The models of these three types of aircraft are F-35, F-117 and P-51 respectively. The specific parameters of the aircrafts are shown in Table I.

TABLE I. THREE AIRCRAFT PARAMETERS

Plane	Length(m)	Width(m)	Height(m)	Wing area(m^2)
F-35	15.67	10.70	4.33	42.7
F-117	20.08	13.20	3.78	84.8
P-51	9.82	11.30	4.17	21.83

The working parameters of radar are as follows: the simulation band is X-band, the frequency range is 9.5GHz~10.5GHz , and its step length is 5MHz, the polarization mode is HH polarization. The pitch angle and azimuth angle range of the radar are 0 °~ 10 ° and 0 °~ 90 ° respectively, and the step length are all 0.1 °. Therefore, the dataset contains 901×101×201=91001 samples, where 201 represents the dimension of one HRRP, 101 and 901 mean the number of pitching angles and azimuth angles respectively. We take each data as an integer for ease of calculation and remove the last cell so that each HRRP is 200 dimensions. The training data should include all azimuth angles to ensure the completeness. Secondly, the pitch angles of the test data should be different from the training data [15]. In all experiments, we choose the training data with pitch angle $\theta \in \{0^\circ, 2^\circ, 4^\circ, 6^\circ, 8^\circ, 10^\circ\}$, and the rest as the test data.

B. Experimental Setup

In this experiment, we adopt normalization to process each HRRP echo and the preprocessed HRRP echo data to train the neural network. For the network details, the 1-D CNN in the shallow feature extraction module contains three convolution layers and Max-pooling layer with pool rate of 2×1. The size of convolution kernel is 9×1. The numbers of convolution kernels of the three convolution layers are set to 32, 32, and 64, respectively. In the multiscale window convolution self-attention module, the obtained shallow features are segmented through the sliding window on each feature channel, and the size of window are set to 4, 8, 16 and 64 in proper order. In this way, we use convolution self-attention mechanism. During training, we choose Cross-Entropy as the loss function and exploit Adam [16] optimizer with learning rate of 0.0001 until convergence.

C. Model analysis

We give two-dimensional projections from all original test HRRP data through UMAP [17] method, as shown in Fig. 2. It can be seen that HRRP samples of all aircraft are gathered together, and the HRRP samples of different targets are overlapping, which is difficult to identify. Therefore, it is also a great challenge for the recognition task.

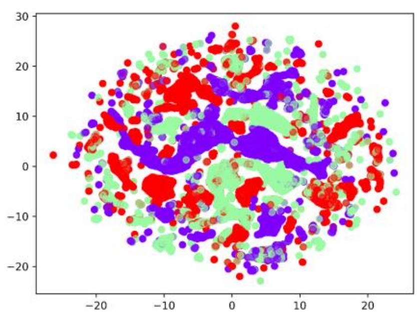

Fig. 2. 2-D UMAP visualizations of original HRRP data.

Fig. 3 shows 2-D UMAP visualizations of test HRRPs' features extracted by our proposed model. Compared with the original HRRP distribution (Fig. 2), we can see that the features extracted by our model are separable, demonstrating the effectiveness of the proposed method.

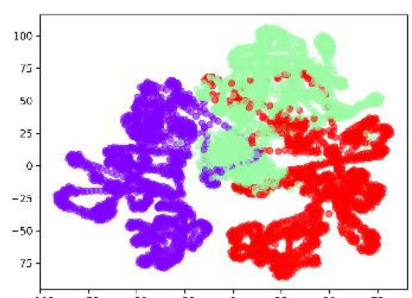

Fig. 3. 2-D UMAP visualizations of test HRRPs' features.

In order to highlight the effectiveness of multiscale window convolution self-attention mechanism, we trained an ordinary 1-D CNN without attention mechanism and randomly selected a test HRRP sample to extracted the features, as shown in Fig.4 (a). Correspondingly, Fig.4 (b) shows the features by our model with multiscale window convolution self-attention mechanism. For 1-D CNN, the dimension of extracted features is 25 and the number of feature channels is 64. The above features can be obtained by learning in the 8 range cells of the HRRP samples. For each feature vector, the target region elements are mostly located in the part marked with the red dotted line in Fig. 4. We can see that the features learned by ordinary 1-D CNN still have large values in the non-target areas of HRRP sample, which has a negative effect on classification. In contrast, the feature extracted by our proposed model is almost 0 in the non-target region of HRRP sample, which is consistent with the conclusion of our analysis. It indicated that our proposed model can suppress the noise and irrelevant information of non-target areas, so as to better mine the characteristics of target areas.

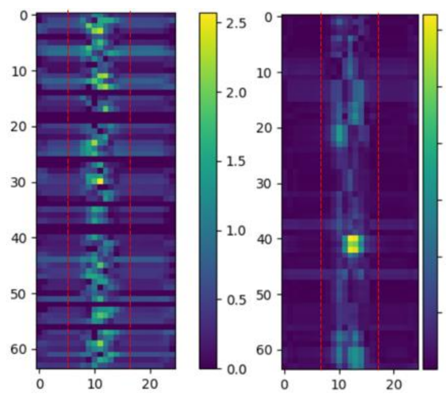

Fig. 4. Visualizations of test HRRPs' features.

D. Recognition Performance

As shown in Table II, our method outperform others three methods. As a traditional method, the PCA is difficult to learn deep features and hence its performance is the worst among the four methods. Ordinary 1-D CNN without attention mechanism treats all features equally, which is not enough to completely distinguish the features of target region and non-target region, resulting in general recognition results. TACNN [12] introduces bidirectional GRU units to generate attention weights through sliding windows to find the target area, but it lacks long-distance dependence, so that attention fails to cover the global receptive field. The proposed method is composed of multiscale window attention, which can capture multi-level features of different regions and achieve obvious improvement of the recognition performance.

TABLE II. RECOGNITION RATE

Methods	Accuracy（%）
PCA	88.9
1-D CNN	90.6
TACNN	92.8
Ours	**94.3**

E. Ablation Experiments

We conducted ablation experiments to demonstrate the effectiveness of multiscale window convolution self-attention mechanism. The abbreviations are described in Table III: 1) O-1-DCNN: denotes that ordinary 1-D CNN without attention mechanism; 2) SE-1-D CNN: denotes that 1-D CNN with channel attention mechanism (SENet [18]); 3) GSA-1-D CNN: denotes that 1-D CNN with global self-attention mechanism; 4) Ours: denotes that our proposed model with multiscale window self-attention mechanism.

In Table III, we show the comparative experimental results of the above methods. For SE and GSA-1-D CNN, although the importance of the target region is depicted, the impact of non-target regions is also enlarged. We use the multiscale window convolution self-attention mechanism to emphasize the target region and suppresses other areas at the same time. Therefore, compared with the other three methods, our recognition performance is further improved.

TABLE III. RESULTS OF ABLATION EXPERIMENTS

Methods	Accuracy（%）
O-1-D CNN	90.6
SE-1-D CNN	91.0
GSA-1-D CNN	94.1
Ours	94.3

IV. CONCLUSIONS

In this paper, we propose an CNN model based on multiscale window self-attention mechanism for Radar HRRP target recognition. Not only the shallow structure features are fully extracted by CNN, but also the deep multi-level features are obtained by multiscale window convolution self-attention mechanism. The proposed network suppresses the noise region of HRRP and reduce the impact of non-target region. Experiments show that the proposed multiscale window self-attention mechanism can improve the recognition performance of ordinary CNN.

REFERENCES

[1] J. J. Wang, Z. Liu, T. Li, L. Ran, and R. Xie, "Radar HRRP target recognition via statistics‐based scattering centre set registration," IET Radar, Sonar & Navigation, vol. 13, pp. 1264-1271, 2019.

[2] W. Chen, B. Chen, X. Peng, J. Liu, Y. Yang, H. Zhang, and H. Liu, "Tensor RNN With Bayesian Nonparametric Mixture for Radar HRRP Modeling and Target Recognition," IEEE Transactions on Signal Processing, pp. 1995-2009, 2021.

[3] M. Pan, L. Du, P. Wang, H. Liu, and Z. Bao, "Noise-Robust Modification Method for Gaussian-Based Models With Application to Radar HRRP Recognition," IEEE Geoscience and Remote Sensing Letters, vol. 10, pp. 558-562, 2013.

[4] 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 transactions on signal processing, vol. 60, pp. 3546-3559, 2012-01-01 2012.

[5] L. Du, J. Chen, J. Hu, Y. Li, and H. He, "Statistical Modeling With Label Constraint for Radar Target Recognition," IEEE transactions on aerospace and electronic systems, vol. 56, pp. 1026-1044, 2020-01-01 2020.

[6] L. Du, H. Liu, Z. Bao, and M. Xing, "Radar HRRP target recognition based on higher order spectra," IEEE transactions on signal processing, vol. 53, pp. 2359-2368, 2005-01-01 2005.

[7] 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 Geoscience and Remote Sensing Letters, vol. 14, pp. 1609-1613, 2017.

[8] J. Wan, B. Chen, B. Xu, H. Liu, and L. Jin, "Convolutional neural networks for radar HRRP target recognition and rejection," EURASIP Journal on Advances in Signal Processing, vol. 2019, 2019.

[9] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based LSTM for aspect-level sentiment classification," in Proceedings of the 2016 conference on empirical methods in natural language processing, 2016, pp. 606-615.

[10] L. Zhang, Y. Li, Y. Wang, J. Wang, and T. Long, "Polarimetric HRRP Recognition Based on ConvLSTM With Self-Attention," IEEE Sensors Journal, vol. 21, pp. 7884-7898, 2021.

[11] B. Xu, B. Chen, J. Wan, H. Liu, and L. Jin, "Target-Aware Recurrent Attentional Network for Radar HRRP Target Recognition," Signal Processing, vol. 155, pp. 268-280, 2019.

[12] J. Chen, L. Du, G. Guo, L. Yin, and D. Wei, "Target-attentional CNN for Radar Automatic Target Recognition with HRRP," Signal processing, vol. 196, 2022-01-01 2022.

[13] S. D Ascoli, H. Touvron, M. L. Leavitt, A. S. Morcos, G. Biroli, and L. Sagun, "Convit: Improving vision transformers with soft convolutional inductive biases," in International Conference on Machine Learning, 2021, pp. 2286-2296.

[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention Is All You Need," in 2017 31st Conference on Neural Information Processing Systems (NIPS), 2017, pp. 5998-6008.

[15] C. Du, B. Chen, B. Xu, D. Guo, and H. Liu, "Factorized discriminative conditional variational auto-encoder for radar HRRP target recognition," Signal Processing, vol. 158, pp. 176-189, 2019.

[16] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Proceedings of the International Conference on Machine Learning (ICML), China ., 2014-01-01 2014.

[17] L. McInnes and J. Healy, "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction," Statistics, 2018.

[18] L. S. S. A. Jie Hu, "Squeeze-and-Excitation Networks.," IEEE transactions on pattern analysis and machine intelligence, pp. 2011-2023, 2020.