Radar HRRP Target Recognition Method Based on Multi-Input Convolutional Gated Recurrent Unit With Cascaded Feature Fusion

Zhiqiang Zeng, Jinping Sun, *Member, IEEE*, Zhu Han, *Student Member, IEEE*, and Wen Hong, *Senior Member, IEEE*

***Abstract*— Over the past decades, radar high-resolution range profile (HRRP) has been one of the research highlights in the field of radar automatic target recognition (RATR) due to its advantages of easy acquisition, small amount of data, and rich target structure information. However, most of the existing methods only consider its amplitude (time domain) characteristics, thereby neglecting the temporal dependence and multi-domain features inside the HRRP sequence. To this end, we propose an end-to-end multi-input convolutional gated recurrent unit neural network, called MIConvGRU, for RATR by both exploiting the multi-domain and temporal information to improve the recognition performance of HRRP target. Initially, the data-preprocessing module is employed to extract the multi-domain features of the target, including time domain, frequency domain, and time–frequency domain features, in order to further enhance the target representation. In addition, a cascaded multi-input GRU structure is designed to acquire the multi-domain temporal dependence feature of HRRP sequence from low to high level. Finally, these temporal features are adaptively fused by a parameter learnable strategy. The experimental results show that the proposed MIConvGRU can effectively learn the multi-domain temporal dependence correlation features in HRRP sequences, improving the target recognition performance.**

***Index Terms*— Gated recurrent unit (GRU), high-resolution range profile (HRRP), multi-domain feature extraction, radar automatic target recognition (RATR), temporal dependence.**

I. Introduction

THE radar high-resolution range profile (HRRP) contains the information of the target scattering centers distribution along the line of sight (LOS). Compared with traditional synthetic aperture radar (SAR) and inverse SAR (ISAR), HRRP has the advantages of easy acquisition, small data volume, and fast processing speed [1]. The HRRP-based target recognition technology is also a hotspot in the research field of radar automatic target recognition (RATR) [2], [3]. HRRP contains rich target structure feature information; however, it is easily affected by issues such as orientation, translation, and amplitude sensitivity, resulting in poor target recognition performance.

In recent years, with the development of deep learning [4]–[6], an increasing number of researchers focus on this data-driven learning method expecting to improve the recognition performance of radar HRRP targets [7], [8]. To solve the problem of feature extraction for HRRP targets, Karabayir *et al.* [9] utilize a convolutional neural network (CNN) to identify the HRRP sequences of ship targets, including six warships and four civilian ships, and finally an overall recognition accuracy of 93.90% is achieved on their own well-prepared dataset. CNN has the characteristics of orientation and translation invariance, which can improve the expression of HRRP features. However, CNN cannot effectively utilize the temporal characteristics that between the different nodes in HRRP sequences. To overcome this limitation, Liu *et al.* [10] propose a target aware two-dimensional recurrent neural network (TATDRNN) to solve the cross-temporal dependencies between nodes in a single HRRP, effectively improving the recognition performance of HRRP targets. Moreover, for the purpose of solving the long-term dependency problem that is prone to occur in traditional recurrent neural networks (RNNs) and avoiding gradient vanishing, Pan *et al.* [11] investigate an HRRP automatic target recognition method based on stacked CNN-BiLSTM. By combining CNN and long short-term memory (LSTM), the learning ability of the model for HRRP sequences is effectively enhanced. Nevertheless, these aforementioned methods only consider the amplitude information (time domain) of HRRP, and lack the exploitation of the multi-domain features inside the HRRP sequences, such as frequency and time–frequency domain features.

For the sake of fully taking advantage of the target information in HRRP sequences and improving the recognition performance of radar HRRP targets, this letter proposes a multi-input convolutional gated recurrent unit (MIConvGRU) network to effectively learn and extract the HRRP sequence features by both considering multi-domain and temporal dependence information. The experimental results demonstrate that MIConvGRU can effectively learn the multi-domain temporal features in HRRP sequences and improve the recognition performance of radar HRRP targets.

Manuscript received 6 May 2022; revised 21 June 2022; accepted 14 July 2022. Date of publication 19 July 2022; date of current version 28 July 2022. This work was supported by the National Natural Science Foundation of China under Grant 62131001 and Grant 62073334. *(Corresponding author: Jinping Sun.)*

Zhiqiang Zeng and Jinping Sun are with the School of Electronic and Information Engineering, Beihang University, Beijing 100191, China (e-mail: zengzq@buaa.edu.cn; sunjinping@buaa.edu.cn).

Zhu Han is with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: hanzhu19@mails.ucas.ac.cn).

Wen Hong is with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: whong@mail.ie.ac.cn).

Digital Object Identifier 10.1109/LGRS.2022.3192289

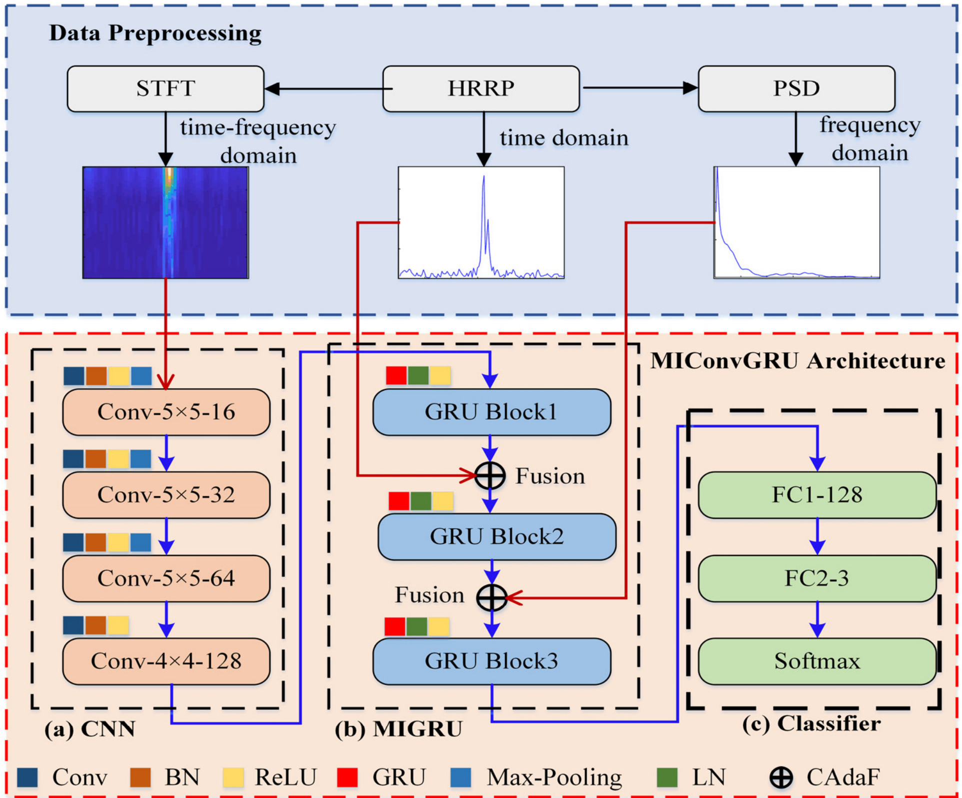

Fig. 1. Overall framework of radar HRRP recognition (a) CNN, (b) MIGRU, and (c) classifier.

To sum up, the main contributions of this letter can be summarized as follows.

1) An end-to-end radar HRRP recognition model is proposed to achieve accurate recognition of target by exploiting the spatial–temporal dependence correlation and multi-domain features in HRRP sequences. In addition, multi-domain radar HRRP sequences (time, frequency, and time–frequency domains) are extracted by data-preprocessing module to improve the representation ability of target.

2) In order to fully learn and fuse the multi-domain HRRP sequences, a cascaded multi-input GRU structure is well designed to extract the multi-domain temporal dependence feature between these sequences.

II. Sequence Operation Based Architecture for HRRP Recognition

Fig. 1 illustrates the overall framework of the proposed MIConvGRU model. To begin with, the target original HRRP sequence is processed by the data-preprocessing module to generate multi-domain features data (time domain, frequency domain, and time–frequency domain). Next, they are input to MIConvGRU for fusion and extraction of multi-domain temporal features of radar HRRP sequence, and finally the classifier completes the accurate recognition of the target.

A. Multi-Domain Feature Generation and Extraction

Radar HRRP sequences contain abundant information about the distribution of scattering intensity from the LOS of the target, which is usually closely related to the shape and contour of the target [2], [3], [8], [11]. According to the distribution characteristics of scattered echoes, long-distance accurate identification and threat assessment of targets can be achieved. In this letter, to make full use of the target information in HRRP sequences, we execute the data-preprocessing method in Fig. 1 to extract and generate target features in multiple data domains for the original HRRP sequences. Specifically, the time-domain and frequency-domain features adopt the amplitude information and power spectral density (PSD), respectively, and the short-time Fourier transform (STFT) is used as the time–frequency feature extractor.

The PSD calculation is as follows:

$$\hat{P}(\omega) = \frac{1}{N}\left|X_N\left(e^{j\omega}\right)\right|^2 \quad (1)$$

where N represents sequence length, and $X_N(e^{j\omega})$ is the Fourier transform (FT) of the raw HRRP sequence. The STFT calculation formula is as follows:

$$X(n,\omega) = \sum_{m=-\infty}^{+\infty} x(m)\cdot w(n-m)\cdot e^{-j\omega m} \quad (2)$$

where $x(m)$ is the input signal, and $w(m)$ indicates the window function that reverses in time and has an offset of n samples. $X(n,\omega)$ is a two-dimensional function of time, n and frequency, ω that relates the time and frequency domains of the signal. Through STFT, we can perform time–frequency analysis on the target HRRP sequence and extract rich time–frequency features.

TABLE I

CONFIGURATION INFORMATION OF CNN FEATURE EXTRACTION NETWORK

Conv-5×5-32	
Components	**Parameter setting**
Conv2d	channel=32, kernel size=5, stride=1, padding=0
BN	channel=32
ReLU	–
Max-Pooling	kernel size=2, stride=1

TABLE II

CONFIGURATION INFORMATION OF MIGRU

GRU Block1		GRU Block2&3	
Components	**Parameters**	**Components**	**Parameters**
GRU	input size=8, hidden size=20, num layers=2	GRU	input size=1, hidden size=4, num layers=2
LN	[128, 20]	LN	[128, 4]
ReLU	–	ReLU	–

B. Multi-Input Convolutional Gated Recurrent Unit

This section presents an end-to-end radar HRRP recognition model, namely MIConvGRU, shown in Fig. 1. MIConvGRU is mainly divided into three parts, CNN time–frequency feature extraction, multi-input GRU adaptive sequence feature fusion, and fully connected feature classifier. First of all, the two-dimensional time–frequency features processed by STFT utilize CNN for feature dimension reduction and extraction. Then, a multi-input GRU hierarchical fusion method is adopted to effectively fuse and represent the extracted multi-domain temporal dependence HRRP features. Finally, these target features are classified using the fully connected layer with Softmax.

1) CNN Time–Frequency Feature Extraction and Dimensionality Reduction: The original HRRP sequence is preprocessed to generate multi-domain feature data, namely time domain, frequency domain, and time–frequency domain. Among them, the time-domain and frequency-domain features are 1-dimensional sequences, and the time–frequency features are 2-dimensional time–frequency distribution maps. To effectively fuse these features, it is necessary for CNN to perform dimensionality reduction and feature extraction on the time–frequency distribution to ensure the consistency of the multi-domain data. The CNN network structure is displayed in Fig. 1(a), which is mainly composed of four convolutional layers. Each convolutional layer performs target feature abstraction and representation in the order of Conv → BN → ReLU → Max-Pooling. The configuration of a convolutional layer Conv-5 × 5 −32 is listed in Table I, and the other convolutional layers are similar to it. After CNN processing, the 2-dimensional time–frequency feature is converted into a 128 × 8 feature matrix and then it will be input into MIGRU for multi-domain temporal dependence feature fusion and extraction.

2) Multi-Input GRU Adaptive Feature Fusion: The network architecture of the multi-input GRU (MIGRU) is shown in Fig. 1(b). It can be seen that MIGRU is mainly composed of three GRU Blocks and two adaptive fusion layers. Each GRU Block consists of GRU, layer normalization (LN), and ReLU activation. The configuration information about the GRU Block is depicted in Table II. It is worth noting that due to the inconsistent size of multi-domain features, the input size of GRU Block 1 is 8, while that of GRU Block 2 and 3 is both 1.

a) Gated recurrent unit: GRU retains and forgets the previous memory unit by introducing the mechanism of the gated control unit, which can effectively solve the problems of long-term dependence and gradient disappearance that are prone to occur in traditional RNNs [12]. Therefore, GRU is very suitable for handling data with temporal characteristics, and can make full use of temporal dependence correlation information in sequence signals to improve the ability to express the target. A standard GRU structure is exhibited in Fig. 2.

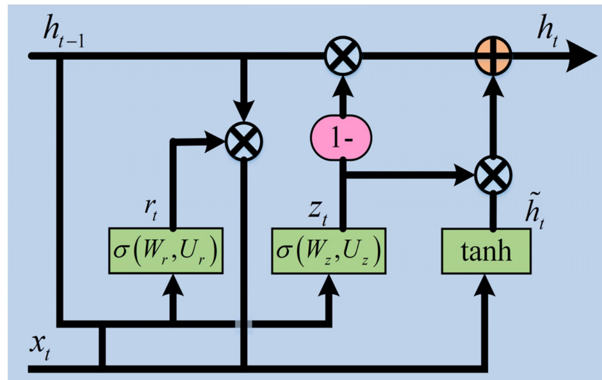

Fig. 2. Principle of GRU.

Assuming that the output state of a node at t time is h_t, the previous output state, h_{t-1} and the candidate state, $\tilde{h}_t$, we have

$$h_t = (1 - z_t)h_{t-1} + z_t\tilde{h}_t \tag{3}$$

where z_t is update gate, which controls the influence of the h_{t-1} on the h_t. And the formula for calculating z_t is

$$z_t = \sigma(W_z x_t + U_z h_{t-1}) \tag{4}$$

where x_t is the input data, σ represents the activation function, and W_z and U_z denote the weights of update gate z_t, respectively. The candidate activation $\tilde{h}_t$ is computed by

$$\tilde{h}_t = \tanh(W x_t + U(r_t \odot h_{t-1})) \tag{5}$$

where W and U represent the GRU weights, r_t is reset gate, and $\odot$ means an element-wise multiplication. When the update gate tends to 0, the update value of the current state is only related to the input, x_t, and has nothing to do with the state at the previous moment. The reset gate r_t can be expressed as

$$r_t = \sigma(W_r x_t + U_r h_{t-1}). \tag{6}$$

Similarly, W_r and U_r represent the weight of reset gate corresponding to x_t and h_{t-1}.

b) Multi-domain temporal feature fusion: In terms of multi-domain target features, how to effectively fuse and characterize the rich temporal features extracted by GRU is crucial for subsequent target recognition. For this purpose,

TABLE III
CONFIGURATION INFORMATION OF HRRP SEQUENCE DATASET

Target	Training set			Testing set		
	Number	Size	Aperture (s=25)	Number	Size	Aperture (s=10)
2S1	1196	1×128	0.075°	1196	1×128	0.03°
D7	1196	1×128	0.075°	1196	1×128	0.03°
ZSU234	1196	1×128	0.075°	1196	1×128	0.03°
Total	3588	-	-	3588	-	-

this letter proposes a cascaded structure-based adaptive feature fusion method (CAdaF) to learn the multi-domain temporal dependence features. The process of CAdaF is shown in Fig. 1(b). First, the time–frequency features extracted by CNN are fed into GRU Block1 for temporal feature extraction. Next, the extracted feature is sequentially fused with time-domain and frequency-domain feature to obtain the multi-domain temporal dependence correlation feature.

Suppose the time–frequency feature is x_{tfd}, time-domain feature is x_{td}, the CAdaF at the primary level can be expressed as

$$\hat{x} = \alpha \cdot x_{tfd} + \beta \cdot x_{td} \tag{7}$$

where α and β are learnable parameters, which are designed to control the exposure of each data domain features to achieve the purpose of adaptive feature fusion. Similarly, high-level feature fusion of these three data domains can be realized by following the above steps.

III. EXPERIMENTAL RESULTS ANALYSIS AND DISCUSSION

In this experimental section, we adopt the Moving and Stationary Target Acquisition and Recognition (MSTAR) dataset [5], [13] to evaluate the performance of the proposed method. Since the targets in MSTAR dataset are 2-dimensional SAR image, this letter inverts the HRRP sequence following the work of [14]. Meanwhile, according to the experimental requirements, three types of targets, 2S1, D7, and ZSU234, are selected for training and testing. The configuration information of the experimental dataset is listed in Table III. In the training set, each sample is the average of the original 25 HRRPs, and it contains a 0.075° aperture. In the testing set, ten HRRPs are randomly selected from 25 HRRP sequences for averaging calculation, and each sample retains the amount of information with an aperture of 0.03°.

A. Three-Type Target Test

This section mainly conducts three types of target tests. In order to more intuitively reflect the target recognition performance of our proposed MIConvGRU, five HRRP recognition models, Traditional CNN [9], TATDRNN [10], LSTM [6], 2CNN+BiLSTM [6], and Deep Nested NN [11], are introduced for comparison purposes. Table IV shows the recognition results of HRRP sequences.

As can be seen from the Table IV, the recognition results of the traditional CNN are: cross entropy (CE) loss (9.7425), accuracy (75.95%), and the comprehensive recognition performance is poor. This may be because the traditional CNN can only memorize the current state information of the target and cannot effectively utilize the temporal information in the HRRP sequence. Compared with the other state-of-the-art HRRP recognition models, MIConvGRU not only considers spatial–temporal information but also introduces multi-domain HRRP features, which effectively improves the model's ability to represent the target and achieves the highest target recognition performance, loss (0.7729), and accuracy (86.68%). From the perspective of computational parameters, the proposed MIConvGRU can achieve excellent HRRP target recognition performance with the least amount of parameters. Moreover, Fig. 3 illustrates the confusion matrix of it, and it manifests that MIConvGRU maintains a high recall rate and target precision and can accurately recognize HRRP targets. This fully indicates the effectiveness and efficiency of our proposed MIConvGRU.

TABLE IV
THREE-TYPE TARGET RECOGNITION RESULTS

Models	Paras (M)	CE Loss	Accuracy
Traditional CNN	0.81	9.7425	75.95%
TATDRNN	0.78	5.4739	79.68%
LSTM	1.09	3.2586	81.24%
2 CNN+BiLSTM	0.96	1.0743	85.37%
Deep Nested NN	1.12	0.8536	86.12%
MIConvGRU (ours)	**0.67**	**0.7729**	**86.68%**

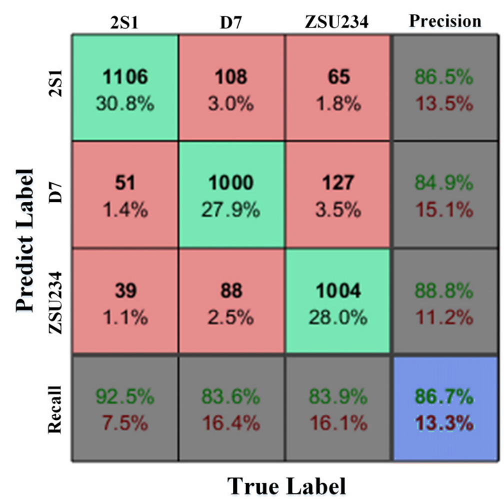

Fig. 3. Confusion matrix of 2S1, D7, and ZSU234.

TABLE V

ABLATION EXPERIMENTS ON DIFFERENT NETWORK TRAINING TRICKS

model	GRU layer	GRU direction	CAdaF	dropout	LN	GRU ReLU	Acc
MIConvGRU	1	1	×	×	×	×	78.12%
	1	1	✓	×	×	×	82.39%
	2	1	✓	×	×	×	84.98%
	2	2	✓	×	×	×	84.25%
	2	1	✓	✓	×	×	85.20%
	3	1	✓	✓	×	×	83.92%
	2	1	✓	✓	✓	×	85.34%
	2	1	✓	✓	✓	✓	**86.68%**

B. Ablation Experiments and Limitation Analysis

MIConvGRU can learn the multi-domain temporal dependence information in HRRP sequences, which improves the recognition performance of targets. For the sake of enhancing the performance of MIConvGRU, we have leveraged many training tricks, such as CAdaF, bidirectional GRU, dropout, LN, and GRU activation. To explore the impact of each module on the recognition performance, this section conducts detailed ablation experiments, and the experimental results are listed in Table V.

According to Table V, the influence of the number of GRU layers on the target recognition accuracy of radar HRRP is to increase first and then decrease, and the best performance is obtained when the number of GRU layers is 2. In the direction of GRU, single direction GRU can obtain higher target recognition accuracy than that of bidirectional GRU, and the proposed CAdaF can significantly improve model's ability to recognize targets. In addition, dropout, LN, and GRU ReLU have a positive impact on the recognition results, which also reflects the effectiveness of these training tricks. Furthermore, as seen from the experimental results, adding an activation layer behind the GRU Block can greatly improve the performance of MIConvGRU. The reason for this may be that the ReLU layer can increase the nonlinear mapping capability of the GRU Block and improve the representation ability of multi-domain temporal HRRP features.

Since MIConvGRU adopts the time–frequency feature map as the main input, the quality of feature map will directly affect the recognition performance of targets. Therefore, appropriate time–frequency resolution or reliable time–frequency analysis methods are necessary. Besides, the model complexity of MIConvGRU is still higher than that of traditional expert systems. This usually means that we need to find a compromise between target recognition performance and time delay according to the task requirements.

IV. CONCLUSION

This letter proposes an end-to-end MIConvGRU network based on multi-domain temporal dependence information to improve the recognition performance of radar HRRP targets. Through the extraction of HRRP multiple data domain features, the characteristic description of the target is increased effectively. Meanwhile, the introduction of MIGRU also reinforces the model's learning ability of target temporal features. To effectively fuse the multi-domain features extracted by GRU, this letter presents a parameter-learnable adaptive feature fusion strategy with cascaded structure to improve the representational capacity of the model. The experimental results demonstrate that MIConvGRU achieves promising performance compared with the current state-of-the-art HRRP target recognition model, which verifies the effectiveness and advancement of the proposed method.

REFERENCES

[1] Y. Huang, Y. Wen, L. Shi, and X. Ding, "Recognition-aware HRRP generation with generative adversarial network," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[2] 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.

[3] Y. Wen, L. Shi, X. Yu, Y. Huang, and X. Ding, "HRRP target recognition with deep transfer learning," *IEEE Access*, vol. 8, pp. 57859–57867, 2020.

[4] C. Li, L. Du, Y. Li, and J. Song, "A novel SAR target recognition method combining electromagnetic scattering information and GCN," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[5] Z. Zeng, J. Sun, C. Xu, and H. Wang, "Unknown SAR target identification method based on feature extraction network and KLD–RPA joint discrimination," *Remote Sens.*, vol. 13, no. 15, p. 2901, Jul. 2021.

[6] C.-L. Lin, T.-P. Chen, K.-C. Fan, H.-Y. Cheng, and C.-H. Chuang, "Radar high-resolution range profile ship recognition using two-channel convolutional neural networks concatenated with bidirectional long short-term memory," *Remote Sens.*, vol. 13, no. 7, p. 1259, Mar. 2021.

[7] L. Shi, Z. Liang, Y. Wen, Y. Zhuang, Y. Huang, and X. Ding, "One-shot HRRP generation for radar target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[8] D. Guo, B. Chen, W. Chen, C. Wang, H. Liu, and M. Zhou, "Variational temporal deep generative model for radar HRRP target recognition," *IEEE Trans. Signal Process.*, vol. 68, pp. 5795–5809, 2020.

[9] 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)*, Jun. 2017, pp. 1–9.

[10] J. Liu, B. Chen, W. Chen, and Y. Yang, "Radar HRRP target recognition with target aware two-dimensional recurrent neural network," in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput. (ICSPCC)*, Sep. 2019, pp. 1–6.

[11] M. Pan *et al.*, "Radar HRRP target recognition model based on a stacked CNN–Bi-RNN with attention mechanism," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022.

[12] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*.

[13] O. Kechagias-Stamatis and N. Aouf, "Automatic target recognition on synthetic aperture radar imagery: A survey," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 36, no. 3, pp. 56–81, Mar. 2021.

[14] D. C. Gross, M. W. Oppenheimer, B. Kahler, B. L. Keaffaber, and R. L. Williams, "Preliminary comparison of high-range resolution signatures of moving and stationary ground vehicles," *Proc. SPIE*, vol. 4727, pp. 205–212, Aug. 2002.