ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-7605-5/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICASSP39728.2021.9413416

ATTENTION ENHANCED SPATIAL TEMPORAL NEURAL NETWORK FOR HRRP RECOGNITION

Yuchen Chu, Zunhua Guo

School of Mechanical, Electrical & Information Engineering, Shandong University, Weihai, China

ABSTRACT

The high resolution range profile (HRRP) is an important signal for radar automatic target recognition (RATR). Recent publications have shown that exploring spatial or temporal features via neural networks is essential for this task. However, it remains a challenging problem to effectively extract and combine discriminative spatial and temporal features for HRRP recognition. In this work, we propose a novel Attention Enhanced Convolutional Gated Recurrent Unit network (AC-GRU) for HRRP recognition which improves the representation of the spatial and temporal co-occurrence in the HRRP sequences. Furthermore, an attention mechanism is employed to select key information in spatial-temporal domains. The simulation results show that the AC-GRU network can achieve better recognition rates compared with several popular classifiers under the condition of limited training data. Finally, further experiments demonstrate that our model also gets robust results under low signal-to-noise ratio.

Index Terms— High resolution range profile, Gated recurrent attention mechanism, Radar automatic target recognition

1. INTRODUCTION

Radar has the advantages of long-distance detection and continuous working capability in modern warfare. Target recognition based on radar imaging can create favorable conditions for real-time reconnaissance of battlefield situations and it is a key factor in determining the outcome of the war. The high resolution range profile (HRRP) is a kind of one-dimensional radar imaging. When the high-resolution radar emits a burst of pulses to the target, the electromagnetic amplitude returned from the target in each range cell can be characterized with a series of backscatter-points, which reflect the shape, the scatterers distribution and the structure information of the target. Compared with synthetic aperture radar (SAR) images, HRRP has the advantages of easy acquisition, low dimensional and real-time computation. Thus, the HRRP has been extensively used in the radar automatic target recognition (RATR) community.

Due to the practical value of the HRRP based RATR, there are various attempts on how to effectively extract discriminative and robust features from the original data. These methods can be approximately classified into two categories. Among the first category, most approaches performed feature extraction by hand-crafted methods, such as the amplitude of the Fourier transform, the coefficients of the Gabor transform, the differential power spectrum, and the sequence from the hidden Markov model, etc. These extracted features are then fed to the classifier to achieve the target recognition task [1-4].

The second category mainly focuses on deep learning techniques. The existing neural network models for HRRP recognition mainly fall into two architectures. One is based on convolutional neural network (CNN) architecture. Li et al. and Wan et al. use CNN to identify spatial features in HRRP [5, 6]. The convolutional layers in CNN usually learn the local features, while the global structure of HRRP might be neglected. These CNN based classification methods intend to use the pooling layer to find the significant spatial features to predict the results correctly, but the down-sampling technique like max-pooling will lose some valuable information in the feature maps and neglect the correlation between the part and the whole. The other framework is recurrent neural network (RNN) with sequential architecture capturing the temporal information of the target. Li et al. and Wan et al. introduce the bidirectional RNN for processing HRRP feature maps, and Jithesh et al. use the long short-term memory (LSTM) based network to predict the category of the HRRP signals [7-9]. The GRU was proposed by Kyunghyun Cho et al, which can be regarded as a simplified version of the LSTM cell [10]. Compared with the traditional RNN, GRU uses the gate unit to solve long-term dependence and avoid gradient vanish or explode during training. However, it remains difficult to find the dependencies between features and fully utilize the spatial information included in HRRP.

Generally, for the HRRP data, there are two prominent characteristics: 1) the HRRP sequences are one-dimensional signals, and the whole sequences contain ample structural information of the radar targets. 2) The sequential coherence not only exists in a single range profile but also in the whole range profiles from different aspect angles. Previous methods have tried to design effective models to extract either temporal or spatial features from the HRRP

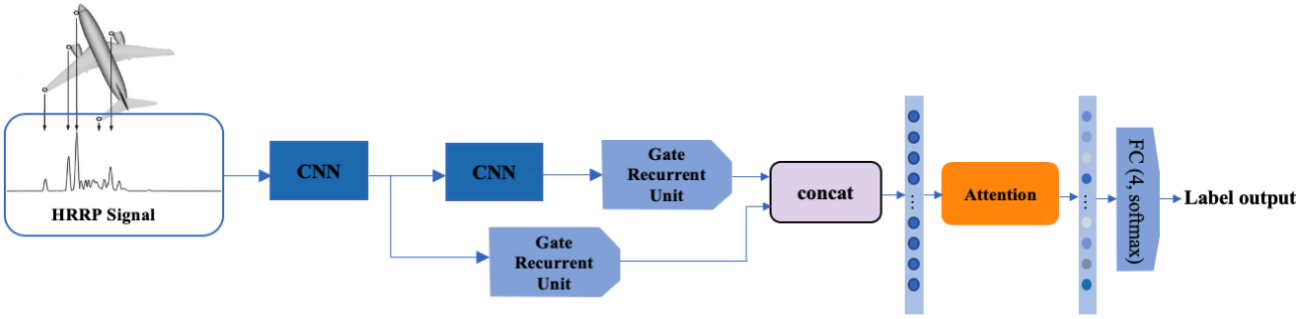

Fig. 1. The AC-GRU network. The AC-GRU estimate the label of the HRRP sequence using a spatiotemporal hierarchical architecture, which combine the spatial features from CNN and the temporal features from GRU. Followed by the attention mechanism to aggregate hidden states. Finally, a fully connected layer outputs the probability for each input sequence.

sequence to achieve good performance on RATR tasks. Nevertheless, how to model both spatial structure and temporal correlation to get discriminative features of the HRRP and give robust recognition results under a low signal-to-noise ratio are still two challenging problems.

In this work, we revisit the role of neural network for HRRP recognition and propose a novel framework called attention enhanced convolutional gated recurrent unit network (AC-GRU) for HRRP recognition. The AC-GRU network can improve the representation of the spatial and temporal co-occurrence in the HRRP sequences. As a brief summary, our contributions are as follows:

We present a novel deep learning approach for HRRP recognition, which is the first attempt to use convolutional GRU in this task;

In order to improve the spatiotemporal representation, a spatiotemporal hierarchical structure is proposed to learn multi-scale features. And an attention mechanism is applied to enhance the key features in a global aspect, so the AC-GRU network can capture discriminative features from the HRRP more effectively;

The simulations show that the AC-GRU network is more robust than the CNN-based and RNN-based models even in the condition of a low signal-to-noise ratio.

2. SPATIAL TEMPORAL NEURAL NETWORK

The architecture of the proposed AC-GRU network is shown in Fig. 1. Firstly, a one-dimensional CNN is used to transform the raw HRRP data into spatial features. Then the spatial features are fed forward to the gated recurrent unit to capture temporal dependency in the HRRP sequence. Inspired by the multi-scale feature maps combination methods in the two-dimensional CNN [11, 12], we introduce spatiotemporal hierarchical architecture with an attention mechanism. A lot of works in natural language processing and computer vision have shown that the attention mechanism is a powerful tool [13, 14]. The models with the attention mechanism outperform other models under the same network architecture. The attention mechanism aims to calculate the importance of the features predicted by GRU in a global aspect and make an alignment on elements weights according to their importance. Finally, the enhanced features are passed into a classification module to compute the final output. By using less than 0.15 million parameters, the prediction results can be obtained in real-time.

2.1. Spatiotemporal Feature Encoder

The convolutional neural network (CNN) is a general and effective model for learning the representation of the spatial structured data. For the AC-GRU network, a one-dimensional CNN is used to extract local features of the input. Followed by a hierarchical architecture, the temporal features are extracted by gated recurrent units and the deep spatial features with bigger receptive fields are extracted by another one-dimensional CNN in parallel. These features are concatenated to a linear layer, then a multi-head attention model is used to align the GRUs encoded features through their importance. Finally, the classifier, a fully connected network with one hidden layer, evaluates the score on each label.

Given an input HRRP signal with 256 sampling points, the HRRP can be denoted as $X = \{x_n\}_{n=1}^{256}$, where x_n is the n-th value of X. The temporal encoding model can be simply denoted as a series of gated recurrent units within time step. For each GRU cell, there are two gates called update gate z_t and reset gate r_t, which control the information from current input and previous hidden state h_{t-1}. The update gate, reset gate and the hidden state are defined as follows [10]:

$$r_t = \sigma_g(W_r x_t + U_r h_t - 1 + b_r) \quad (1)$$

$$z_t = \sigma_g(W_z x_t + U_z h_t - 1 + b_z) \quad (2)$$

$$h_t = (1 - z_t) h_t - 1 + z_t \sigma_h(W_h x_t + U_h(r_t h_t - 1) + b_h) \quad (3)$$

where W, U are the parameter metrics, σ_g is the sigmoid function, σ_h is the hyperbolic tangent function, and h_t is the hidden state from the previous time step. We define the GRU with 64 hidden states, which means there are 64 GRU cells in our decode model, the last GRU cell comprises all the hidden states into an output vector with 64 dimensions.

2.2. Attention Mechanism

Since the attention mechanism has the ability to assign different weights to each part of the input and extracting more critical information, it helps the model to make a more accurate judgement. We introduce the multi-heads attention from reference [15] because of the superior quality for translation tasks. The structure of our attention model is shown in Fig. 2.

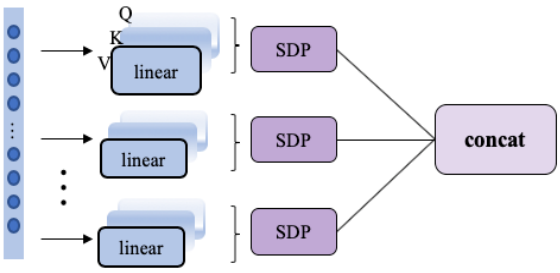

Fig. 2. Attention mechanism

Taking the output of the decode model, we can apply the attention mechanism to align the output value. The number of the multi-heads was set to 4, dividing the 64d output of the GRUs into 4 parts. Then for each part, the 16d vector is multiplied by three matrices: W_i^Q, W_i^K, W_i^V. Each of the three vectors Q, K, V is linearly projected 4 times respectively as follows [15]:

$$Q'_i = QW_i^Q, \qquad i = 1,2,...,h \tag{4}$$

$$K'_i = KW_i^K, \qquad i = 1,2,...,h \tag{5}$$

$$W'_i = VW_i^V, \qquad i = 1,2,...,h \tag{6}$$

These weight matrices have the same shape as the input vector, which keeps the dimension of the output vectors Q', K', V' also in 16d. We calculate the scaled dot product of vectors Q, K, V and apply the softmax function to obtain the weights on values. The four 16d vectors are finally concatenated together. The operation of the attention model can be summarized as [15]:

$$Attention(Q,K,V) = \text{softmax}(\frac{QK^T}{\sqrt{d_m}})V \tag{7}$$

where $\frac{1}{\sqrt{d_m}}$ is the scaling factor, controlling the fluctuation of the dot product.

Finally, to simplify the parameters of the network, the classifier is only composed of a fully connected layer for the four classes. For training the AC-GRU network, the loss function is defined as the following focal loss, where p_t is the probability of the target class:

$$L(p_t) = -(1-p_t)^2 \log(p_t) \tag{8}$$

3. EXPERIMENT

3.1. Dataset

In our experiment, four aircrafts HRRP data include B2, F117, J6 and YF22. These data are achieved by bursting a stepped frequency waveform to the symmetric aircrafts, which carries 256 pulses. The interval of the azimuth (ranging from 0° -180°) is set to 0.6°. Thus, we get 300 HRRP sequences for each aircraft and 1200 sequences in total. The dataset is randomly shuffled and divided into training, validation and testing set by ratio 4:2:4.

3.2. Training Parameters

For the spatiotemporal encoder, we apply a kernel size of 1x3 with stride=1 for the convolutional layers. The size of the hidden state of GRUs is set to 64, followed by a concat layer that outputs $f_i \in \mathbb{R}^{128}$. For multi-head attention, we use the method described in 2.2 to learn the attention weights. The classifier has one fully connected layer with 128 neurons as input, followed by a final output layer with four classes. The batch sizes for the HRRP dataset are 32. During the training phase, we use the Adam optimizer to optimize the network. The initial learning rate is set to 0.001 and reduced by 0.1 every 20 epochs. We train the models on one NVIDIA 1080 GPU.

3.3. Comparative evaluation

The recognition performance of the five models is shown in table 1. Compared with other approaches, it can be seen that the AC-GRU network provides more accurate recognition rates under the same training set, which is because the model takes the advantage of both spatial and temporal information contained in the range profiles.

Table 1. Recognition performance (mean average precision)

Models	CNN [6]	LSTM [9]	Bi-SRU [7]	CNN-BiRNN [8]	AC-GRU
mAP (%)	75.0	79.37	87.5	88.75	91.66

Next, we evaluate the effect of the attention mechanism by removing the multi-head attention. Thus, the GRUs are directly connected to the classifier. After removing the attention model, the recognition accuracy drops 4.16% on the test set. This suggests that the attention mechanism is necessary for our recognition test, which promotes the capability of feature representation.

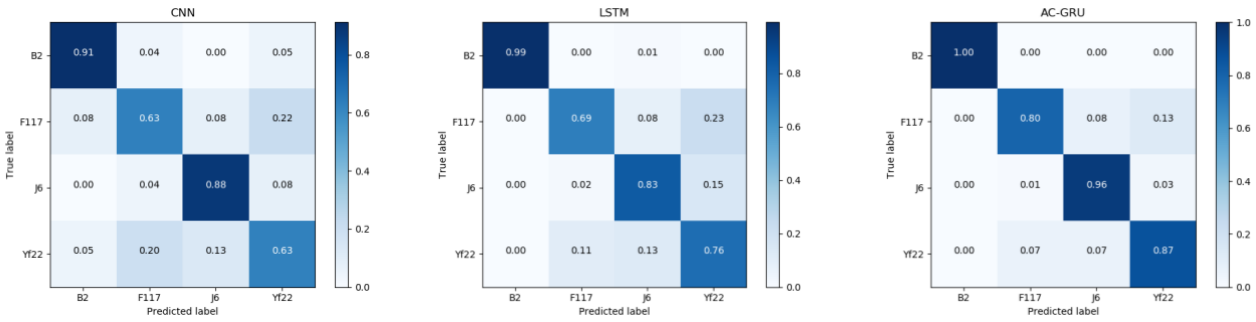

Fig. 3. Confusion matrix for each model

We show the experimental results via the confusion matrix on the test set. The scores in the principal diagonal line indicate the percentage of accuracy for each class, other scores show the wrong classification percentage. As shown in Fig. 3, it is very confusing for a CNN model and LSTM model to recognize F117 and YF22 correctly. This is because the original range profiles of the two aircrafts have some similarities. Nevertheless, the proposed AC-GRU can significantly improve the recognition accuracy.

Furthermore, in order to investigate the effect of the temporal encoder in our network, we compare the GRU with RNN and LSTM. By only replacing the GRU with RNN and LSTM, we keep the spatiotemporal hierarchical architecture unchanged. Thus, we get the classification results of the AC-RNN and the AC-LSTM. The mean average precision for AC-RNN is 85.62% and 88.13% for AC-LSTM. Fig. 4 shows the loss curve of three networks on the validation set during the training procedure. We can see that the AC-GRU network converges fastest among the three networks.

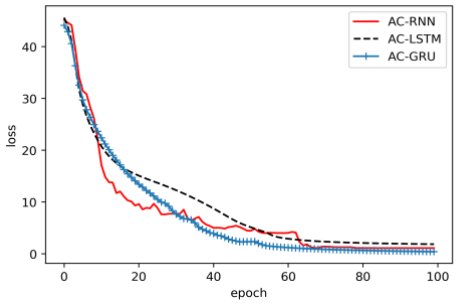

Fig. 4. Loss curves of three networks

Finally, the noise is added to the test set to investigate the noise influence on the classification performance and the robustness of the model. The HRRP in our dataset can be viewed as 256 discrete points and its power P_s is calculated by the following formula:

$$P_s = \sum S(t)^2 \tag{9}$$

where $S(t)$ is the amplitude of the HRRP at timestep t. We adopt Gaussian noise and adjust the standard deviation to get different signal-to-noise ratios. The SNR of the HRRP signal in the test set is 25dB, 20dB, 15dB and 10dB, respectively.

The recognition results of the five models are presented in table 2. As it can be seen, the recognition performance of all models decreases when the noise is increasing, but the AC-GRU still outperforms other models in different SNR conditions. The results show that the proposed AC-GRU is an effective method for HRRP recognition.

Table 2. Evaluation results of mAP under different SNR

Models mAP(%)	*25dB*	*20dB*	*15dB*	*10dB*
CNN [6]	72.91	70.83	68.75	62.5
LSTM [9]	79.16	75.0	70.83	61.25
Bi-SRU [7]	81.25	77.08	71.87	69.58
CNN-BiRNN [8]	83.3	80.83	78.95	73.95
AC-GRU	86.4	81.45	80.2	76.66

4. CONCLUSION

We have presented an attention enhanced convolutional gated recurrent unit (AC-GRU) network for HRRP recognition, which is the first attempt to use convolutional GRU for this application. The proposed AC-GRU can not only capture discriminative spatial and temporal features but also improve the representation of the spatial and temporal co-occurrence relationship. Furthermore, the attention mechanism is adopted to enhance the key features in a global aspect. The simulation results demonstrate that the AC-GRU network can obtain better recognition rates, and the spatiotemporal hierarchical architecture is helpful to improve the classification performance.

5. ACKNOWLEDGEMENT

This work was supported by National Natural Science Foundation of China (61401252).

7. REFERENCES

[1] J. P. Zwart, R. V. D. Heiden, S. Gelsema, and F. Groen, “Fast translation invariant classification of HRR range profiles in a zero phase representation,” IEEE Proceedings - Radar, Sonar and Navigation, vol. 150, no. 6, pp. 411–418, December 2003.

[2] F. Zhu, X. Zhang, and Y. Hu, “Gabor Filter Approach to Joint Feature Extraction and Target Recognition,” IEEE Transactions on Aerospace and Electronic Systems, vol. 45, no. 1, pp. 17–30, January 2009

[3] F. Zhu, X.-D. Zhang, Y.-F. Hu, and D. Xie, ‘‘Nonstationary hidden Markov models for multiaspect discriminative feature extraction from radar targets,’’ IEEE Trans. Signal Process., vol. 55, no. 5, pp. 2203–2214, May 2007.

[4] Z. H. Guo and S. H. Li, “One-Dimensional Frequency-Domain Features for Aircraft Recognition from Radar Range Profiles,” IEEE Transactions on Aerospace and Electronic Systems, vol. 46, no. 4, pp. 1880–1892, October 2010.

[5] Li, X., Li, C., Wang, P., Men, Z., & Xu, H. (2015). SAR ATR based on dividing CNN into CAE and SNN. 2015 IEEE 5th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR). IEEE.

[6] J.Wan, B.Chen, B.Xu, H.Liu, andL.Jin, “Convolutional neural network for radar HRRP target recognition and rejection”, EURASIP J. Adv. Signal Process., vol. 2019, no. 1, p. 5, 2019.

[7] X. Li and Z. Guo, "A Bi-SRU Neural Network Based on Soft Attention for HRRP Target Recognition," 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Suzhou, China, 2019, pp. 1-5.

[8] J. Wan, B. Chen, Y. Liu, Y. Yuan, H. Liu and L. Jin, "Recognizing the HRRP by Combining CNN and BiRNN With Attention Mechanism," in IEEE Access, vol. 8, pp. 20828-20837, 2020.

[9] Jithesh, V., Sagayaraj, M. J., & Srinivasa, K. G. (2017). LSTM recurrent neural networks for high resolution range profile based radar target classification. International Conference on Computational Intelligence & Communication Technology. IEEE.

[10] Cho K, Van Merrienboer B, Gulcehre C, et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation[J]. Computer Science, 2014.

[11] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., & Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. European Conference on Computer Vision. Springer International Publishing.

[12] Lin, T. Y., Dollár, Piotr, Girshick, R., He, K., Hariharan, B. , & Belongie, S, "Feature Pyramid Networks for Object Detection," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[13] J. Fu, H. Zheng and T. Mei, "Look Closer to See Better: Recurrent Attention Convolutional Neural Network for Fine-Grained Image Recognition," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 4476-4484.

[14] Luong, Minh-Thang, Pham, Hieu, and Manning, Christopher D. "Effective Approaches to Attention-based Neural Machine Translation." 2015 Association for Computational Linguistics (ACL), pp. 1412-1421.

[15] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., & Gomez, A. N. (2017). Attention is all you need.