A Bi-SRU Neural Network Based on Soft Attention for HRRP Target Recognition

Xin LI, Zunhua GUO
School of Mechanical, Electrical & Information Engineering
Shandong University
Weihai, China

Abstract—**Radar automatic target recognition (RATR) plays a significant role in military applications. The high resolution range profiles (HRRP) contain a large amount of information, such as the distribution of the scatterers, target size and structure, so the HRRP has great prospects in the field of RATR. The traditional HRRP based RATR systems need to extract features manually, which is very difficult and complex to obtain excellent features. To solve this problem, this paper proposed a bidirectional simple recurrent unit network based on soft attention mechanism (SAMBi-SRU) to achieve HRRP target recognition. The simulation results demonstrate that the proposed model can extract robust features from the HRRP data effectively and obtain good performance and noise immunity in HRRP based target recognition.**

Keywords- recurrent neural networks; high resolution range profiles; radar automatic target recognition; deep learning

I. INTRODUCTION

Radar automatic target recognition (RATR) systems can provide a great deal of important information, such as target position, category and shape [1], which is the key factor to the victory of modern war. The high resolution range profiles (HRRP) are the coherent summation of echo amplitude, and obtained by wideband radars in the radar observation direction [2]. Therefore, the range profiles contain plenty of information like the distribution of radar cross section (RCS) and the target scatterers. In addition, the range profiles are easy to acquire and store compared with Synthetic aperture radar (SAR) and inverse synthetic aperture radar (ISAR) images [3]. As a result, it is widely recognized that the HRRP can reduce the need for computation and inter-radars communication bandwidth. The HRRP-based RATR system has been successfully applied into some military target recognition fields such as the recognition of fighters, tanks and missiles, and become an important research topic.

Due to its high research and practical values, the HRRP-based RATR system has drawn the attention of many scholars. As shown in Fig. 1, a traditional HRRP target recognition system usually includes three stages: data preprocessing, feature extraction, and identification. Data preprocessing is generally aimed to eliminate external noise, translation and amplitude sensitivity of the original HRRP data [4-6]. For sake of extracting robust features, researchers have carried out extensive experiments and proposed many feature extraction methods such as the amplitude of the Fourier transform [5], the differential power spectrum [7], and the coefficients of the Gabor transform [8], etc. After extracting the HRRP features, the classifiers are designed to achieve radar target recognition. The traditional classifiers can be divided into three classes: correlation filters, classifiers based on HRRP statistical features and machine learning. It is worth mentioning that with the developments of machine learning in recent years, the support vector machine and relevance vector machine have been gradually applied to the area of HRRP recognition and achieved great performance.

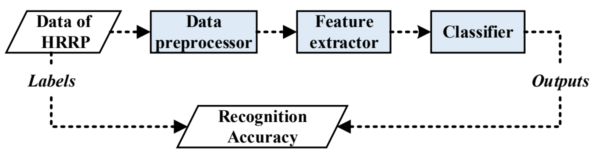

Figure 1. The HRRP-RATR system

Feature extraction has always been a hot research issue in the community of signal processing, because features with low dimension, high-robustness and strong distinguishability can improve the accuracy of target recognition. The traditional HRRP-based RATR systems usually need to extract features manually, but it is very complicated to acquire excellent features. In order to address this issue, many researchers have introduced the end-to-end deep neural networks (DNNs) into the HRRP target recognition. For example, M. Pan processed data with t-distributed stochastic neighbor embedding and achieved target recognition with the discriminant deep belief network [9]. L. Du proposed a truncated stick-breaking hidden Markov model to extract features and classify the targets [2]. Besides, there are also many HRRP target recognition algorithms composed of convolutional neural networks and other DNNs [10-13].

Elman believed that how to represent and learn timing information in time series is a very important issue. Therefore, he proposed a simple recurrent network to solve this problem [14]. In recent years, recurrent neural networks (RNNs) have made great progress in the field of speech recognition [15], machine translation [16], and text summarization [17], etc.

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

978-1-7281-4852-6/19/$31.00 ©2019 IEEE

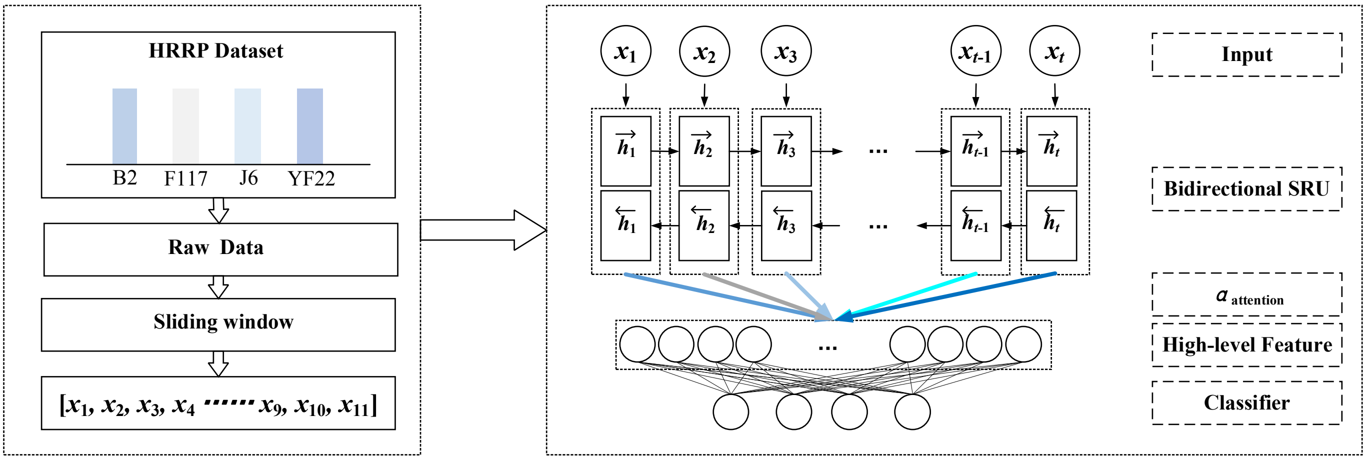

Figure 2. The proposed model for HRRP target recognition

Similar to the speech and text, the HRRP data are also one-dimensional signals and there is temporal and spatial correlation between HRRP points [7]. To make full use of the temporal and spatial correlation of the HRRP sequences, this paper proposed a bidirectional simple recurrent unit network based on soft attention mechanism (SAMBi-SRU) to achieve HRRP target recognition. The bidirectional simple recurrent unit (Bi-SRU) network can extract robust features while the soft attention mechanism can highlight important segments in the HRRP points. The simulation results show that the SAMBi-SRU network can extract effective features from the HRRP data and obtain higher recognition accuracy.

II. The Soft Attention Based SRU Network

A. The Basic Network Architecture

The basic model architecture is shown in Fig. 2, which mainly comprises four components: the data segmentation part, the Bi-SRU network part, the soft attention part and the softmax part. The Bi-SRU network can capture the time dependency both forward and backward in the HRRP sequences at the same time. The soft attention algorithm helps the network to extract robust features while the softmax performs the target identification.

B. Data segmentation

When using traditional systems to perform HRRP target recognition tasks, some preprocessing operations will be conducted, such as L1-regularization and principal component analysis. To validate the recognition performance of the proposed model with the raw data, there are no de-noising operations in the proposed model. However, to reduce the time steps of the RNN and its complexity of the computation, the one-dimensional HRRP data are divided into several short units with sliding window.

As shown in Fig. 3, 11 units are intercepted from the one-dimensional HRRP sequence which is composed of 256 sampling points. To guarantee the correlation of the input data, the size of the sliding window is set to 40 points while the size of the sliding step is set to 20 points. After that, the original HRRP data are segmented into 11 units, which can improve the training efficiency practically.

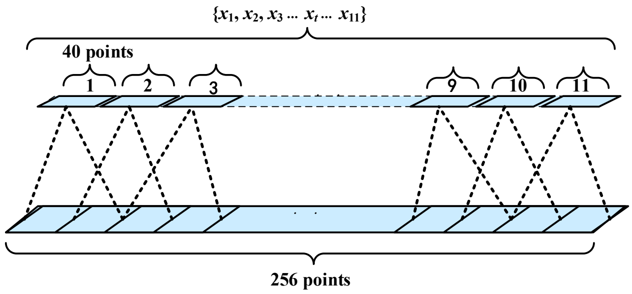

Figure 3. Segmentation of the HRRP data

C. The description of RNNs

The RNN is capable of processing sequential data with long periods. Fig. 4 shows the architecture of a simple RNN. It can be drawn from the figure that the RNN has not only the connection from the input layer to the hidden layer but also the connection from the previous hidden layer to the current one.

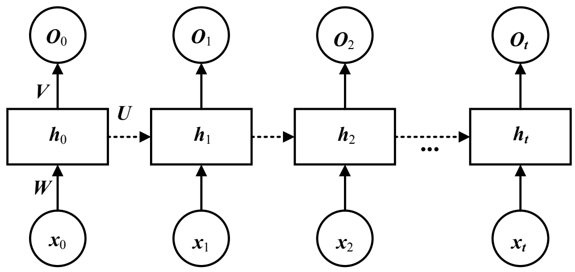

Figure 4. The recurrent neural networks

With the input $X=\{x_1,x_2\cdots x_t\cdots x_{11}\}$, the update process of the hidden state h_t and the output O_t can be calculated as follows:

$$\boldsymbol{h}_t = f(\boldsymbol{U}\boldsymbol{h}_{t-1} + \boldsymbol{W}\boldsymbol{x}_t + \boldsymbol{b}_h), \tag{1}$$

$$\boldsymbol{O}_t = g(\boldsymbol{V}\boldsymbol{h}_t + \boldsymbol{b}_y), \tag{2}$$

where $\boldsymbol{U}$, $\boldsymbol{W}$, and $\boldsymbol{V}$ denote the learnable weight parameters, the $f(\bullet)$ and $g(\bullet)$ represent the activation functions, such as the sigmoid function $\sigma(x)$ and the $\tanh(x)$ function, the $\boldsymbol{b}_h$ and $\boldsymbol{b}_y$ are bias matrices.

Although the traditional RNNs can capture the time dependency of the sequential data, Y. Bengio pointed out that the gradients are prone to vanish or explode during the training process with long sequence [18]. The gradients of the loss function $\boldsymbol{C}_t$ can be given by:

$$\frac{\partial \boldsymbol{C}_t}{\partial \boldsymbol{W}} = \sum_{\tau \le t} \frac{\partial \boldsymbol{C}_t}{\partial \boldsymbol{h}_\tau} \frac{\partial \boldsymbol{h}_\tau}{\partial \boldsymbol{W}} = \sum_{\tau \le t} \frac{\partial \boldsymbol{C}_t}{\partial \boldsymbol{h}_t} \frac{\partial \boldsymbol{h}_t}{\partial \boldsymbol{h}_\tau} \frac{\partial \boldsymbol{h}_\tau}{\partial \boldsymbol{W}}, \tag{3}$$

in which

$$\frac{\partial \boldsymbol{h}_t}{\partial \boldsymbol{h}_\tau} = \frac{\partial \boldsymbol{h}_t}{\partial \boldsymbol{h}_{t-1}} \frac{\partial \boldsymbol{h}_{t-1}}{\partial \boldsymbol{h}_{t-2}} \cdots \frac{\partial \boldsymbol{h}_{\tau+1}}{\partial \boldsymbol{h}_\tau}. \tag{4}$$

With the increase of the sequences length and the accumulation of the time steps, the gradient value tends to be zero or infinity.

D. *Simple recurrent units*

The long short-term memory (LSTM) network proposed by Hochreiter in 1997 and the gated recurrent unit (GRU) network proposed by Kyunghyun in 2014 solved this problem by using the gating unit [19-20]. Furthermore, a great quantity of experiments has been conducted to prove that the LSTM and GRU networks can solve the long-term dependence problem effectively. The SRU is also a network with gating units [21]. However, compared with the LSTM and GRU, the SRU network eliminates the dependence of $\boldsymbol{h}_t$ on that of the previous time step while training, and therefore it provides the possibility of parallelization, speeding up the training process without reducing the accuracy of the network. The diagrammatic sketch of the SRU network is shown in Fig. 5.

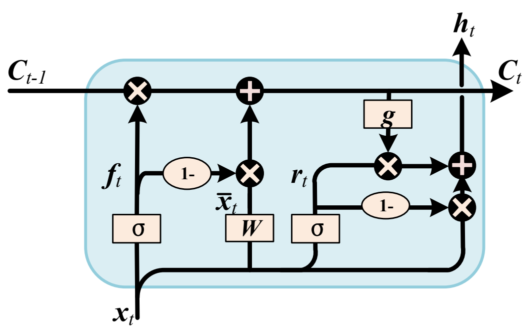

Figure 5. The SRU cell

Moreover, at the current time step, the memory $\boldsymbol{C}_t$ and the hidden state $\boldsymbol{h}_t$ can be inferred by (5) - (9):

$$\tilde{\boldsymbol{x}}_t = \boldsymbol{W}\boldsymbol{x}_t, \tag{5}$$

$$\boldsymbol{f}_t = \sigma(\boldsymbol{W}_f \boldsymbol{x}_t + \boldsymbol{b}_f), \tag{6}$$

$$\boldsymbol{r}_t = \sigma(\boldsymbol{W}_r \boldsymbol{x}_t + \boldsymbol{b}_r), \tag{7}$$

$$\boldsymbol{C}_t = \boldsymbol{f}_t \odot \boldsymbol{C}_{t-1} + (1 - \boldsymbol{f}_t) \odot \tilde{\boldsymbol{x}}_t, \tag{8}$$

$$\boldsymbol{h}_t = \boldsymbol{r}_t \odot g(\boldsymbol{C}_t) + (1 - \boldsymbol{r}_t) \odot \boldsymbol{x}_t, \tag{9}$$

where $\boldsymbol{f}_t$ and $\boldsymbol{r}_t$ stand for the forget gate and the reset gate at time respectively.

E. *Soft attention mechanism*

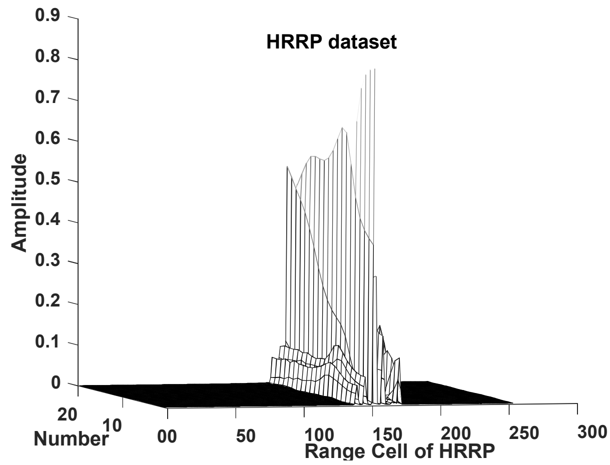

Figure 6. The HRRP samples

From Fig. 6, we can see that the importance of the target information is distinct at different range cells, which has a great impact on target recognition. The attention mechanism lets the model pay more attention to significant information by giving different weights to features according to their contribution to the ultimate result, thereby improving the accuracy of the model. As a consequence, it has shown excellent performance in quantities of applications such as object recognition, image capture, natural language processing, and so forth. The soft attention mechanism proposed in [22] is introduced into our proposed model to make the model focus on more significant features. Multiplying the features produced by the Bi-SRU network with the attention weights, we can obtain a more robust feature distribution $\boldsymbol{M}$. $\boldsymbol{M}$ is defined as:

$$\boldsymbol{M} = \sum_{t=1}^{T} \left(\frac{\exp(\boldsymbol{v}_t)}{\sum_{i=1}^{T} \exp(\boldsymbol{v}_i)} \right) \boldsymbol{h}_t, \tag{10}$$

where

$$\boldsymbol{v}_i = \boldsymbol{U}_{att} \tanh(\boldsymbol{W}_{att} \boldsymbol{h}_i + \boldsymbol{b}). \tag{11}$$

III. SIMULATIONS

A. Experimental data

The data set used in simulation experiments is HRRP data of four aircrafts, including B2, F117, J6 and YF22. The azimuth ranges from 0º-180º and its interval is set to 0.6°. Thus for an aircraft, we can get 300 HRRP sequences. Each sequence is a 256-dimensional vector and the number of HRRP sequences is 1200. The data set is randomly shuffled and divided into train and test set by ratio 2:1.

B. Principal model parameters

The proposed SAMBi-SRU network is implemented using Tensorflow. In the model, the number of hidden states is defined as 20, which means there are 40 SRU cells in our model. Furthermore, the dimension of $\boldsymbol{h}_t$ is 40. The tanh function is chosen as the activation function, the batch size is set to 64 and the training epoch is 2000, the initial learning rate is set to 0.01. To prevent over-fitting, a regularization operation is added to the model during training, and the regularization coefficient is initialized to 0.001.

C. Recognition performance

For the purpose of making comparison with the proposed model, in our experiments, we built one shallow model: the k-nearest neighbor (KNN) classifier, and five deep networks, including fully connected neural network (FNN), one-dimensional convolutional neural network (CNN1D), two-dimensional convolutional neural network (CNN2D), deep belief network (DBN) and Bi-SRU network to perform the HRRP target recognition task. The primary structure and parameter description of these models are shown in Table I. The recognition rates of each model are shown in Table II.

As the results showed, the SAMBi-SRU network obtain the best HRRP target recognition rates, since RNNs can not only extract the amplitude-based features but also can capture the time dependency information of the HRRP sequences. Furthermore the recognition rates of the SAMBi-SRU network are higher than that of the Bi-SRU network, which is due to the fact that soft attention mechanism can spotlight important data segments in HRRP sequences. As a result, this algorithm is more suitable for modeling and predicting HRRP sequences. It is necessary to note that the recognition rates of KNN are the lowest among the seven algorithms, which proves that the HRRP target recognition systems using DNNs are better choices. The recognition rates of the CNN1D, CNN2D and DBN models are all over 95%. When it comes to DBN model, it assigns initial values of parameters with the weights of restricted boltzmann machine which is based on non-supervision training. Therefore, the recognition rates of the DBN are higher than that of FNN. It is also worth noting that the recognition rates of CNN2D model are worse than that of CNN1D model, because when transforming one-dimensional HRRP sequences into two-dimensional matrices, the temporal and spatial correlation between HRRP points are destroyed.

TABLE I. KEY STRUCTURE PARAMENTERS

Algorithms	**Key structure parameters**
KNN	The k is set to 3; Distance measurement is Manhattan distance.
FNN	The number of the hidden layer is 1; The number of hidden units is 30.
CNN1D	The number of convolutional layers is 3; The size of conv kernels is 5x16, 3x32, 3x64.
CNN2D	The number of convolutional layers is 2; The size of conv kernels is 5x5x16, 3x3x32.
DBN	The number of Restricted Boltzmann Machines is 3; The k in the contrast divergence method is set to 1.
Bi-SRU	The number of hidden units in the SRU network is 40.
SAMBi-SRU	The i in soft attention mechanism is set to 20; The number of hidden units in the SRU network is 40.

TABLE II. RECOGNITION RATES (%)

Algorithms	**Targets**			
	B2	***F117***	***J6***	***YF22***
KNN	84	84	89	81
FNN	100	90	100	95
CNN1D	100	96	100	98
CNN2D	100	96	95	99
DBN	100	96	97	99
Bi-SRU	100	99	98	98
SAMBi-SRU	100	100	98	100

D. Noise immunity

To assess the anti-noise performance of the proposed model, we performed experiments on the HRRP target recognition task when the signal-to-noise ratio (SNR) is set to 20DB, 15DB, 10DB and 5DB respectively. The recognition results are listed in Table III.

TABLE III. AVERAGE RECOGNITION RATES (%) WITH DIFFERENT SNR

Algorithms	**SNR**			
	20DB	***15DB***	***10DB***	***5DB***
KNN	80	78	78	75
FNN	92	85	82	77
CNN1D	94	92	82	72
CNN2D	92	85	82	76
DBN	79	76	72	71
Bi-SRU	94	89	82	75
SAMBi-SRU	95	89	83	77

With the increase of the noise, the recognition rates of the FNN, CNN1D, CNN2D, Bi-SRU and SAMBi-SRU network models are still acceptable, but the recognition rates of the

DBN model and KNN classifier are no higher than 80%. From the experiment results, we can see that the anti-noise performance of the SAMBi-SRU network is superior to the other six models.

IV. Conclusion

In this paper, a SAMBi-SRU network is introduced into modeling and predicting HRRP sequences. In the preprocessing phase, the one-dimensional HRRP points are intercepted into short units through the sliding window to reduce the complexity of computation. Then the SAMBi-SRU network is trained. The simulation results show that the SAMBi-SRU network can extract the effective features of HRRP sequences and then obtain good recognition performance. We also investigate the noise immunity of the proposed model, and the results demonstrate that the overall anti-noise performance of the proposed algorithm is superior to other algorithms.

References

[1] L. Du, H. W. Liu, Z. Bao, and J. Y. Zhang, "A two-distribution compounded statistical model for Radar HRRP target recognition," IEEE Transactions on Signal Processing, vol. 54, no. 6, pp. 2226–2238, June 2006.

[2] L. Du, H. W. Liu, Z. Bao, and M. D. Xing, "Radar HRRP target recognition based on higher order spectra," IEEE Transactions on Signal Processing, vol. 53, no. 7, pp. 2359–2368, July 2005.

[3] L. Du, P. H. Wang, H. W. Liu, M. Pan, F. Chen, and Z. Bao, "Bayesian Spatiotemporal Multitask Learning for Radar HRRP Target Recognition," IEEE Transactions on Signal Processing, vol. 59, no. 7, pp. 3182–3195, July 2011.

[4] K. T. Kim, D. K. Seo, and H. T. Kim, "Efficient radar target recognition using the MUSIC algorithm and invariant features," IEEE Transactions on Antennas and Propagation, vol. 50, no. 3, pp. 325–337, March 2002.

[5] 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," IEE Proceedings - Radar, Sonar and Navigation, vol. 150, no. 6, pp. 411–418, December 2003.

[6] H. J. Li, Y. D. Wang, and L. H. Wang, "Matching score properties between range profiles of high-resolution radar targets," IEEE Transactions on Antennas and Propagation, vol. 44, no. 4, pp. 444–452, April 1996.

[7] 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.

[8] 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.

[9] 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, no. 9, pp. 1609–1613, September 2017.

[10] C. Guo, Y. He, H. Wang, T. Jian, and S. Sun, "Radar HRRP Target Recognition Based on Deep One-Dimensional Residual-Inception Network," IEEE Access, vol. 7, pp. 9191–9204, January 2019.

[11] Q. Zhang, J. Lu, T. Liu, P. Zhang, and Q. Liu, "Ship HRRP Target Recognition Based on CNN and ELM," 2019 4th International Conference on Electromechanical Control Technology and Transportation, Guilin, China, pp. 124-128, 2019.

[12] S. Yu and Y. Xie, "Application of a Convolutional Autoencoder to Half Space Radar Hrrp Recognition," 2018 International Conference on Wavelet Analysis and Pattern Recognition, Chengdu, China, pp. 48-53, 2018.

[13] V. Jithesh, M. J. Sagayaraj, and K. G. Srinivasa, "LSTM recurrent neural networks for high resolution range profile based radar target classification," 2017 3rd International Conference on Computational Intelligence Communication Technology, Ghaziabad, India, pp. 1-6, 2017.

[14] J. L. Elman, "Finding structure in time," Cognitive Science, vol. 14, no. 2, pp. 179–211, March 1990.

[15] A. Graves, N. Jaitly and A. Mohamed, "Hybrid speech recognition with Deep Bidirectional LSTM," 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, Olomouc, pp. 273-278, 2013.

[16] K. Cho, B. V. Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, and H. Schwenk, et al. "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," https://arxiv.org/abs/1406.1078.

[17] N. Raphal, H. Duwarah and P. Daniel, "Survey on Abstractive Text Summarization," 2018 International Conference on Communication and Signal Processing, Chennai, pp. 0513-0517, 2018.

[18] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," IEEE Transactions on Neural Networks, vol. 5, no.2, pp. 157–166, March 1994.

[19] S. Hochreiter and J. Schmidhuber, "Long Short-term Memory," Neural Computation, vol. 9, no.8, pp. 1735–1780, December 1997.

[20] J. Chung, C. Gulcehre, K. Cho, Y. Begio "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," https://arxiv.org/abs/1412.3555.

[21] T. Lei and Y. Zhang, "Training RNNs as Fast as CNNs," https://arxiv.org/abs/1709.02755v1.

[22] D. Bahdanau, K. Cho, and Y. Bengio "Neural Machine Translation by Jointly Learning to Align and Translate," https://arxiv.org/abs/1412.3555.