Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 16, No. 2, pp. 247–259, 2020

RADAR BEHAVIOR RECOGNITION BASED ON CENTER LOSS LSTM*

Yongjian Shen, Chunyu Xie, Zhendong Liu, Jing Dai, Baochang Zhang†, Lei Zhu and Jingwen Li

Abstract: Radar behavior recognition is a challenging problem in the field of pattern recognition. In this paper, we incorporate center loss into Bidirectional Long Short-Term Memory (BiLSTM), termed C-BiLSTM, to efficiently solve the problem. With the center loss, the C-BiLSTM model can effectively classify the radar behavior based on the frequency information. Moreover, we collect a large Radar Behavior Dataset including five modes, which have 4091 samples. We achieve 99.51% using the proposed deep learning model.

Key words: *radar behavior, deep learning, center loss*

Mathematics Subject Classification: *62M45, 62H30, 68T10*

1 Introduction

Radar behavior recognition, as a key technology of electronic interception system, has the characteristics of identifying and detecting radar behavior mode in various application [5, 15, 17, 32]. Since the increasingly new complex system radars result in denser signal stream, radar behavior identification becomes a crucial problem in electronic countermeasure field.

Radar behavior recognition is challenging in the field of radar measurement and radar interception. Once the radar behavior is recognized, in-depth evaluation of the radar's capability and electronic countermeasure will be performed. Traditional interception systems usually recognize radars with stable working mode. Nevertheless, such technologies cannot adapt to the development of modern programmable digital radar systems. More researchers pay attention to this problem, which shows that the machine learning methods have been a promising way toward high performance, including Hidden Markov Model (HMM) [18], Artificial Neural Network (ANN) [17], Support Vector Machine (SVM) [2]. All these methods are based on conventional handcrafted features, however, which fail to deal with the fuzzy signals for the complicated and changeable radar systems so that the performance of them is not considered sufficiently.

Recently, deep learning is investigated in radar signal recognition. The supervised deep learning network can directly use the labeled training data to classify radar signals. More importantly, instead of handcrafted features in use, the raw data is directly fed into the

*The work is supported in part by National Natural Science Foundation of China under Grants 61672079. This work is supported by Shenzhen Science and Technology Program KQTD2016112515134654.

†Corresponding author.

network to train a model, which can learn meaningful and discriminative features in a complex situation. In fact, deep learning has been a new state-of-the-art in the machine learning field, which are designed with multi-layer neural networks. With the development of parallel computation techniques, it is possible to fulfill the deep multi-layer neural network. In particular, deep learning techniques have been successfully applied to the recognition tasks. Through increasing the number of layers and the number of neurons in every layer, the deep learning neural network can express complex functions. Based on its super modeling ability, the deep learning neural network can be used to map input data vectors to output results. The Long Short-Term Memory (LSTM) [8] and its variant Gated Recurrent Unit have been successfully applied to many fields, including language modeling [13,14,20], image captioning [26,30], video analysis [1,6,10,23] and 3D action recognition [4,12,24,27]. Such methods can also be used to solve our problems since the radar signal has a temporal relationship [11,22].

The publicly available datasets are very important to researchers, as far as we know, but most of the previous research in the field of radar signal recognition were conducted with their own data. And few datasets related to radar signal recognition are publicly available. In this paper, we collect a large radar behavior dataset including five modes of radar signals which have 4091 samples, helping more researchers to evaluate and compare their algorithms by using the same radar dataset. In addition, we first introduce LSTMs to recognize the radar signal, including the searching, tracking, ISAR imaging and their combine modes. As is known to us, the Center Loss [28] helps to minimize the intra-class variations. As illustrated in Figure 1, we introduce the center loss into Bidirectional Long Short-Term Memory (BiLSTM) to cope with internal variations. The results show the effectiveness of our proposed algorithm. In summary, we have following contributions:

1. We collect a large radar signal dataset, helping more researchers to evaluate and compare their algorithms.
2. We develop a radar signal recognition system via BiLSTM and achieve a significant performance.
3. We introduce the center loss into the BiLSTM network to improve the traditional softmax loss function.

The rest of the paper is organized as follows. Section 2 introduces the related works, and Section 3 describes the details of the proposed method. Experiments and results are presented in Section 4. Finally, Section 5 concludes the paper.

2 Related Work

Recently, a lot of improved machine learning based radar emitter behavior identification technique emerged. Petrov has proposed an Artificial Neural Network (ANN) based method [17]. By this means, it deals with 29094 intercepted radar signal samples, which is consisted of 12 radar parameters (such as frequency, modulation, pulse repeated intervals). For conventional radar signal, the correct probability of identification is more than 80%. This algorithm can extract key features from abundant radar emitter data and parameters, taking full advantage of statistical analysis and characteristic modeling ability to acquire prior-knowledge of radar emitter. It can realize automatic identification of radar emitter type. Another improvement of ANN-based radar emitter behavior identification is to use the evolutionary algorithm to

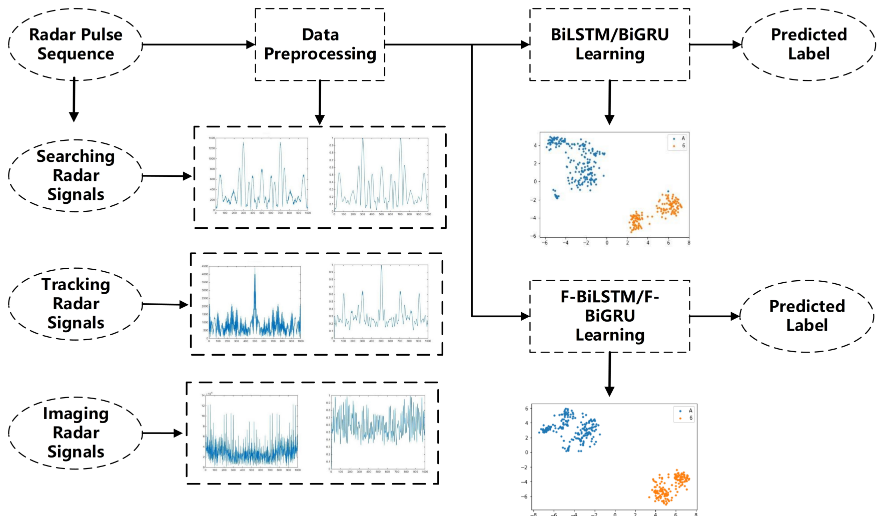

Figure 1: The architecture of radar behavior recognition system. Data preprocessing is applied to radar sequence, yielding a variety of radar signals, such as searching, tracking and imaging, etc. Bidirectional Long Short-Term Memory (BiLSTM) is for the first time applied to this field, and the center loss is introduced into BiLSTM to minimize the intra-class variations. The t-SNE method is used to visualize the data distribution, which shows that C-BiLSTM indeed deals well with internal variations efficiently.

overcome local extreme point problem of traditional artificial neural network. By establishing a general radar knowledge dataset, radar signal behavior can be influenced using the evolutionary algorithm based radial basis function network. But this method is based upon accurate radar parameter estimation and correct radar signal sorting, pre-treatment of radar signal will have a great impact on this method.

Except for ANN-based methods, SVM based methods are also being proposed in the field of radar behavior identification. SVM, based on statistical learning theory, has a good ability of recognition and generalization capability and can be used in radar signal identification. Normal SVM based algorithm can be split into two individual parts, training process, and recognition process. But the traditional SVM is based on batch-wise training, causing that when a new training sample is adopted, the whole training process has to restart with all the training samples. In order to solve this problem, M. Zhu has proposed an online support vector machine-based radar behavior identification method [32]. Firstly, signal features are extracted from labeled radar signal. Then, the signal features and the labels are used to train a series of recognition models. In the application stage, signal feature extraction is processed when a new radar signal is entered. The signal feature transfer to OISVM classifier to complete the recognition process. On the one hand, the affirmed recognition results are used to support the decision-making process. On the other hand, they can also be used to retrain the OISVM models to improve the model performance.

Other radar behavior identification methods include the graphic representation of Pulse Description Word (PWD) method [5]. This method is based on the assumption that graphic representation of PWDs contains an important message to distinguish different radar. This algorithm adopts an off-line processing technique to plot the PWDs figures, which need experts' participation to achieve radar behavior identification while lacking the ability of automatic parameter extraction and on-line processing. According to practical engineering need, J. Matuszewski has proposed a Karhunen-Loeve transformation-based radar signal identification algorithm [15]. This algorithm can provide the probability of radar behavior. However, this algorithm still has poor extensibility. Similar to above methods, its performance has a strong dependence on prior-knowledge and it cannot process online. To cope with these issues, we develop hereafter our radar signal recognition method based on BiLSTM that learns to automatically extract relevant features from input data.

In the past, various models have been used to process sequential signal, such as hidden semi-Markov models [18], conditional random fields [2], and finite-state machines [9]. Recently, RNN became well-known with the development of deep learning. As a special RNN, LSTM has been widely used in the field of voice and video because of its ability to handle gradient disappearance in traditional RNNs. It has the less conditional independence hypothesis compared with the previous models and facilitates integration with other deep learning networks. Researchers have recently combined spatial/optical flow CNN features with vanilla LSTM models for global temporal modeling of videos [3, 7, 21, 25, 29]. These studies have demonstrated that deep learning models have a significant effect on action recognition [3,16,21] and video description [25,31]. But to our best of knowledge, the fusion of CNN and LSTM is never investigated to solve the Automatic Modulation Classification (AMC) problem.

3 The Center Loss Based LSTM

Leveraging the insights from recent works, we design a center loss LSTM network for radar behavior recognition. Figure 1 shows the architecture of the proposed network, which has a bidirectional LSTM layer and a softmax layer give the predictions. What's more, the successful center loss is introduced into the loss function, which helps the BiLSTM learn more discriminative features. In the following, we first review LSTM briefly to make the paper self-contained. Then we introduce our method of center loss LSTM.

3.1 Overview of LSTM

The recurrent neural network (RNN) is a successful model for sequence data modeling and feature extraction. Figure 2 shows an RNN neuron, for the recurrent neurons at some layers, the output responses h_t are calculated based on the inputs x_t to this layer and the responses h_{t-1} from the previous time step

$$h_t = \theta(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \tag{3.1}$$

where $\theta(\cdot)$ denotes the activation function, b_h denotes the bias vector, W_{xh} is the matrix of weights between the input and hidden layer and W_{hh} is the matrix of recurrent weights from the hidden layer to itself at adjacent time steps.

LSTM is an advanced RNN architecture which mitigates the vanishing gradient effect of RNN. As illustrated in Figure 3, a LSTM neuron contains a memory cell c_t, an input gate i_t, a forget gate f_t, an output gate o_t, and an output response h_t. At each time step t, the neuron can choose to write, reset and read the memory cell governed by the input gate i_t forget gate f_t and output gate o_t. The memory cell has a self-connected recurrent edge, ensuring that the gradient can pass through many steps without vanishing or exploding. Therefore, it copes with the problem of vanishing gradient. The updates in a layer of LSTM are summarized as follows:

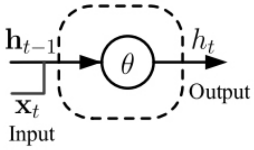

Figure 2: The structure of RNN.

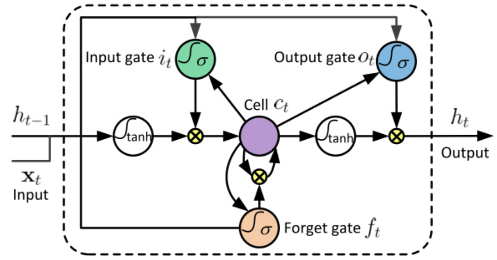

Figure 3: The structure of LSTM block.

$$\begin{cases} i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i), \\ f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f), \\ c_t = f_t c_{t-1} + i_t tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c), \\ o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o), \\ h_t = o_t tanh(c_t) \end{cases} \tag{3.2}$$

To utilize the information from both the past and the future efficiently, Schuster and Paliwal [19] proposed the bidirectional recurrent neural network (BRNN), which presents the sequential forwards and backwards for two separate recurrent hidden layers. These two recurrent hidden layers share a single output layer. Figure 4 shows a bidirectional recurrent network. We can replace the nonlinear units in Figure 4 with LSTM blocks to obtain bidirectional LSTM. Considering the sequential information of radar signal, bidirectional LSTM can be applied to radar signal recognition.

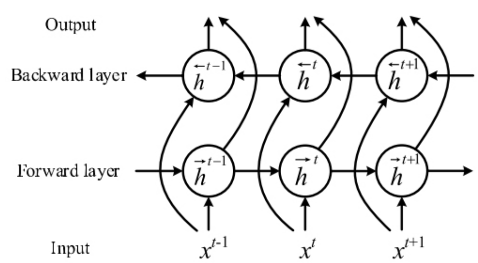

Figure 4: Bidirectional Recurrent Neural Network.

$\boxed{3.2}$ Center Loss BiLSTM

The radar behavior recognition system in Figure 1 has very powerful learning capability. We introduce center loss into Bidirectional Long-Short Term Memory (BiLSTM), which obtains a better performance than the state-of-the-art. The intuition behind this model is that the auxiliary loss, which helps to differentiate the deeply learned features in BiLSTM model. As far as we know, the center loss helps to minimize the intra-class variations, which is incorporated into softmax loss of BiLSTM, and the resulting networks achieve a better recognition ability. The softmax loss function is presented as follows:

$$L_s = -\sum_{i=1}^{m} log \frac{e^{W_{y_i}^T O_i + b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T O_i + b_j}} \tag{3.3}$$

Where O_i denotes the output vector. Based on this, we also introduce the center loss as follows:

$$L_f = \sum_{i=1}^{m} \|O_i - \mu_{y_i}\|_2^2 \tag{3.4}$$

Where μ_{y_i} is the y_ith class mean of output vectors. We update the mean vectors of each class in every iteration and combine the softmax loss with the center loss as follows:

$$L = L_s + \theta L_f \tag{3.5}$$

Where θ is used for balancing two parts of loss functions, restricted in $[0, 1]$. Output vector is set to O_i, with e as iteration number. In each iteration, we compute the loss of BiLSTM and compute the back propagation error by

$$\frac{\partial L^e}{\partial O_i^e} = \frac{\partial L_s^e}{\partial O_i^e} + \theta \frac{\partial L_f^e}{\partial O_i^e} \tag{3.6}$$

In addition, we set scalar parameter α, which is restricted in $[0, 1]$. We update the mean vector μ_j in the $e + 1$ iteration by the following formulas. Meanwhile, all the parameters are updated in the process of training.

$$\mu_j^{e+1} = \mu_j^e - \alpha \cdot \Delta\mu_j^e \tag{3.7}$$

With the additional loss, we can enhance the discriminative power of BiLSTM radar behavior recognition. Mean pooling is used to obtain the final BiLSTM features. By using the way of classical online backpropagation, the network is learned. For classifying a new radar signal, we use a majority voting rule over the outputs, which keeps only the most probable class ($\text{argmax}_{i\in[1,n]} O_i$) to determine the final radar class. The algorithm of the Center Loss BiLSTM (C-BiLSTM) can be seen in Table 1.

Table 1: Algorithm of Center Loss BiLSTM

Algorithm: Center Loss BiLSTM (C-BiLSTM)
1: set $e = 0$
2: Initialize w_{y_i}, b_{y_i} and μ_{y_i}, $i = 1, 2, ..., m$
3: Initialize balancing factor θ and update rate α
4: **repeat**
5: $e = e + 1$;
6: Update w_{y_i} and b_{y_i} according to $\partial L^e / \partial O_i^e$
7: Update $\mu_{y_i}^{e+1} = \mu_{y_i}^e - \alpha \cdot \Delta\mu_{y_i}^e$
8: **until** convergence

4 Experiments

4.1 Radar Signal Dataset

4.1.1 Input Radar Signals

Radar Behavior dataset including five modes: Scanning radar mode, Tracking radar mode, TWS (Track While Scan) radar mode, Imaging radar mode and ISAR Imaging radar mode. These modes have the corresponding samples 990, 990, 1485, 349, 205 respectively. Those different modes can accomplish different tasks,In this paper, we use Center-loss LSTM to recognize them.

Scanning radar mode Generally, scanning radar mode is a radar mode used to accurately determine range, angle, and speed of a target. For scanning mode, we set the parameters as follows: pulse width: 30us, bandwidth: 5MHz, pulse repetition interval: 5ms, inter-pulse modulation: linear frequency modulation. The Scanning radar mode signal sketch is shown in Figure 5.

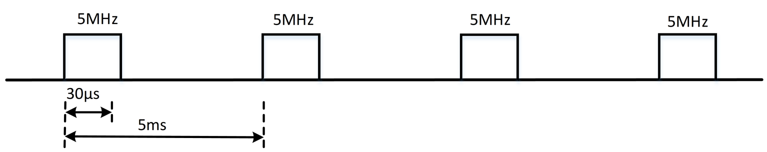

Figure 5: Scanning radar mode signal sketch.

Tracking radar mode In tracking radar mode, Radar detects target and determines location as well as predict its trajectory, and generally use a narrow and symmetrical beam. For tracking mode, we set the parameters as follows: Pulse width: 50us, bandwidth: 10MHz, pulse repetition interval: 5ms, inter-pulse modulation: linear frequency modulation. The Tracking mode signal sketch is shown in Figure 6.

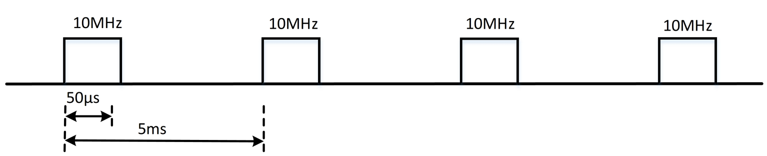

Figure 6: Tracking radar mode signal sketch.

TWS radar mode The TWS (Track while scan) is a mode of radar operation in which the radar allocates part of its power to tracking the targets while part of its power to scanning the airspace. In the TWS mode, the radar has a possibility to acquire additional targets as well as providing an overall view of the airspace and helping maintain better situational awareness. The TWS radar model signal sketch is shown in Figure 7.

Imaging radar mode Imaging radar mode is a radar mode which is used to create two-dimensional images, typically landscapes. Imaging radar usually provides its power to

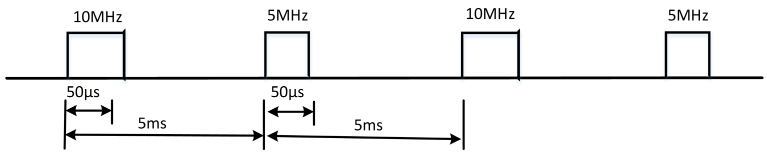

Figure 7: TWS radar mode signal sketch.

illuminate an area on the ground and take a picture from radio wavelength. This mode includes three kinds of signal pulse, every five pulses are divided into a group. The Imaging radar mode signal sketch is shown in Figure 8.

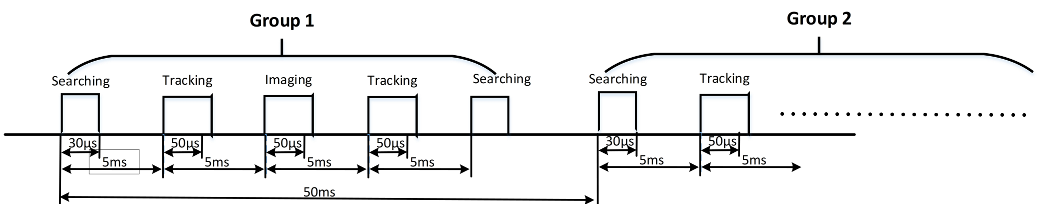

Figure 8: Imaging radar mode signal sketch.

ISAR Imaging radar mode ISAR (Inverse Synthetic Aperture Rada) Imaging radar mode is a radar mode which is widely employed to obtain the high-resolution image of moving targets such as missiles and aircrafts. This mode includes three kinds of signal pulse, every seven pulses are divided into a group. The ISAR Imaging radar signal sketch is shown in Figure 9.

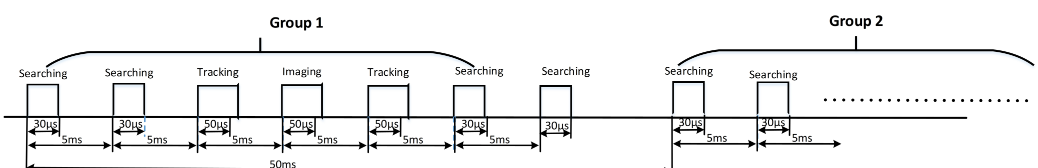

Figure 9: ISAR Imaging radar mode signal sketch.

4.1.2 Radar Signals Preprocessing

In order to simplify the operation, the down-sampling is applied in this part. The radar signals after down-sampling record the whole information of five modes. Then, we split the entire radar signal to obtain the datasets by utilizing FFT and autocorrelation. Obviously, the radar signal characterized by multi-pulse combination can be used to detect the cycle through autocorrelation analysis. For different signals, the complete cycle of radar signals can be obtained by autocorrelation and other operations. In practical applications, under the known prior cases of the maximum period of radar modes, the real-time autocorrelation analysis can be achieved by intercepting the maximum period of the radar mode. The complete cycle of radar signals can be obtained by setting the autocorrelation threshold.

In view of the characteristics of the multi-pulse combination of radar signals, we use the method of Fast Fourier Transform (FFT) for the whole radar period to obtain the behavior features. Through the comprehensive application of the data processing method, the cutting-edge time of multiple pulses can be obtained. The cutting-edge time of each pulse is taken as the starting point, and the FFT operation is performed for the whole period. Figure 10 shows the spectrum map of three modes of radar behavior signals after the operation of FFT.

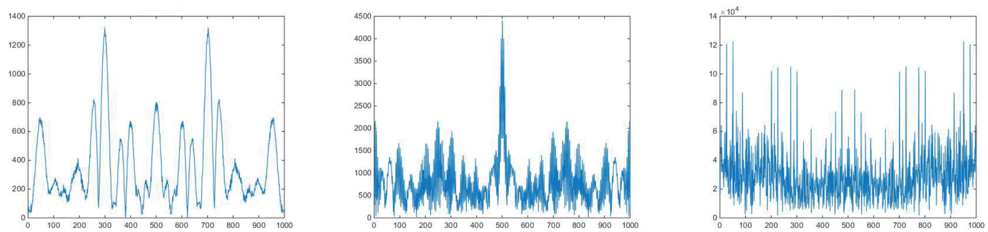

Figure 10: The spectrum map of three modes of three radar behavior signals after FFT. Signal sections for three radar pulses modes are processed by FFT. the results show different FFT features.

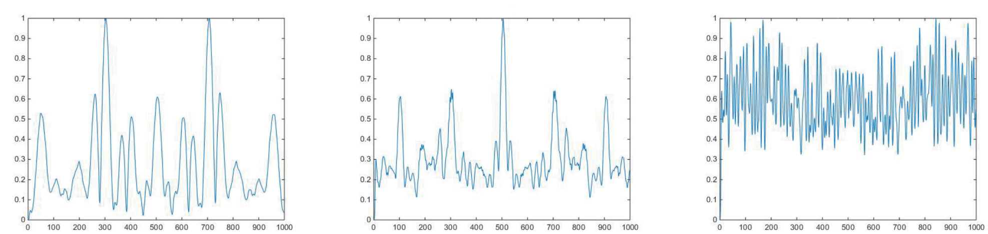

Figure 11: Three radar signal samples from Radar Behavior Dataset. Signal sections for three radar pulses modes are processed by FFT. Then, smoother FFT features are derived by Butterworth filter and normalization processing.

This method can detect and judge each pulse in real time, and it can meet the requirement of real-time and fast. We can also use the filtering and normalization processing to regularize the data. The Butterworth filter is performed. Figure 11 shows the final radar data, the same three modes with Figure 10. The radar data after above processing sets up the radar dataset of 4091 samples and five modes, and we conduct our experiments on this dataset.

4.2 Experiments on C-BiLSTM and SVM

We validate the proposed approach on the Radar Behavior Dataset. To investigate the effectiveness of our network, we conduct extensive experiments with the architecture of center loss LSTM. Our experiments are performed based on the TensorFlow framework and an NVIDIA GTX 1070 GPU. Stochastic gradient descent (SGD) algorithm is used to train our end-to-end network. We set the learning rate, decay rate and momentum to 0.001, 0.95, 0.9, respectively. The applied dropout probability in our network is 0.5. The dimensions of the cell state of LSTM are 128. We use a mini-batch of 64 and the scalar parameter α and θ are set to 0.5 and 0.01. We select 3000 samples for training and remaining samples for testing. The training is stopped after 1000 iterations, obtaining the accuracy of 99.51%. For comparison, we use traditional SVM under the same split of dataset, which obtains an accuracy of 94.60%, and the traditional LSTM acquires an accuracy of 99%. The comparison is illustrated in Figure 12. The performance proves the effectiveness of our proposed C-BiLSTM method.

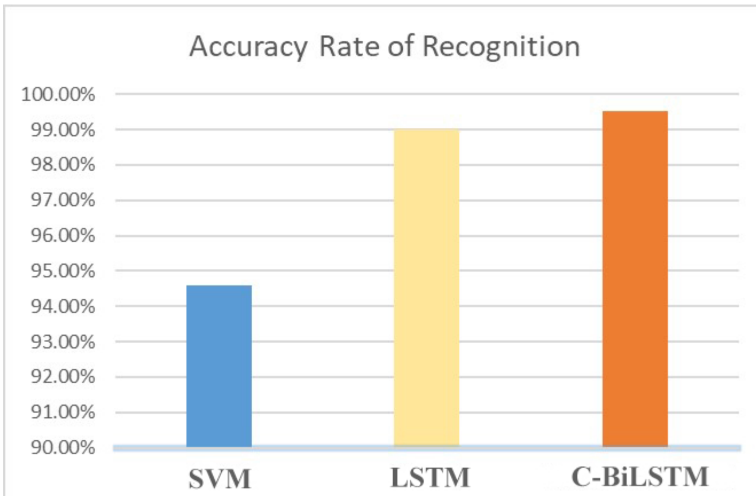

Figure 12: Comparing with the state-of-the-arts in terms of accuracy rate.

5 Conclusion

In this paper, we extend the BiLSTM network to achieve a center loss BiLSTM network for radar behavior recognition, which has a strong discriminative capability with the assistance of additional loss. The experimental results obtained from our collected Radar Behavior Dataset validate the contributions by achieving state-of-the-art performance. Based on this, future work will focus on deeper models and the analysis of multi data type, further improve model robustness and radar behavior identification ability.

References

[1] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei and S. Savarese, Social lstm: Human trajectory prediction in crowded spaces, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 961–971.

[2] C. Cortes and V. Vapnik, Support-vector networks, *Machine learning* 20 (1995) 273–297.

[3] J. Donahue, L.A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko and T. Darrell, Long-term recurrent convolutional networks for visual recognition and description, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 2625–2634.

[4] Y. Du, W. Wang and L. Wang, Hierarchical recurrent neural network for skeleton based action recognition, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1110–1118.

[5] J. Dudczyk and A. Kawalec, Specific emitter identification based on graphical representation of the distribution of radar signal parameters, *Bulletin of the Polish Academy of Sciences Technical Sciences* 63 (2015) 391–396.

[6] Z. Deng, A. Vahdat, H. Hu and G. Mori, Structure inference machines: Recurrent neural networks for analyzing relations in group activity recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4772–4781.

[7] C. Feichtenhofer, A. Pinz and A. Zisserman, Convolutional two-stream network fusion for video action recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1933–1941.

[8] S. Hochreiter and J. Schmidhuber, Long short-term memory, *Neural Comput.* 9 (1997) 1735–1780.

[9] V. Iglesias, J. Grajal, O. Yeste-Ojeda, M. Garrido, M.A. Sanchez and M. Lopez-Vallejo, Real-time radar pulse parameter extractor, in: *Radar Conference*, IEEE, 2014, pp. 37-1-375.

[10] M.S. Ibrahim, S. Muralidharan, Z. Deng, A. Vahdat and G. Mori, A hierarchical deep temporal model for group activity recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1971–1980.

[11] G. Lefebvre, S. Berlemont, F. Mamalet and C. Garcia, BLSTM-RNN based 3D gesture classification, in: *Proceedings of the International Conference on Artificial Neural Networks*, Springer, 2013, pp. 381–388.

[12] J. Liu, A. Shahroudy, D. Xu and G. Wang, Spatio-temporal lstm with trust gates for 3d human action recognition, in: *Proceedings of the European Conference on Computer Vision*, Springer, 2016, pp. 816–833.

[13] T. Mikolov, S. Kombrink, L. Burget, J. Černockỳ and S. Khudanpur, Extensions of recurrent neural network language model, in: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2011, pp. 5528–5531.

[14] G. Mesnil, X. He, L. Deng and Y. Bengio, Investigation of recurrent-neural-network architectures and learning methods for spoken language understanding, in *Proceedings of the Interspeech*, 2013, pp. 3771–3775.

[15] J. Matuszewski, Knowledge-based signal processing for radar identification, *International Journal of Computing* 7 (2014) 80–87.

[16] B. Mahasseni and S. Todorovic, Regularizing long short term memory with 3D human-skeleton sequences for action recognition, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3054–3062.

[17] N. Petrov, I. Jordanov and J. Roe, Radar emitter signals recognition and classification with feedforward networks, *Procedia Computer Science* 22 (2013) 1192–1220

[18] L.R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE* 77 (1989) 257–286.

[19] M. Schuster and K.K. Paliwal, Bidirectional recurrent neural networks, *IEEE Trans. Signal Process.* 45 (1997) 2673–2681.

[20] M. Sundermeyer, R. Schlüter and H. Ney, LSTM neural networks for language modeling, in *Proceedings of the Thirteenth annual conference of the international speech communication association*, 2012.

[21] L. Sun, K. Jia, D.Y. Yeung and B.E. Shi, Human action recognition using factorized spatio-temporal convolutional networks, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 4597–4605.

[22] S. Shin and W. Sung, Dynamic hand gesture recognition for wearable devices with low complexity recurrent neural networks, in *IEEE International Symposium on Circuits and Systems*, IEEE, 2016, pp. 2274–2277.

[23] S. Vijayanarasimhan, O. Vinyals, R. Monga and G. Toderici, Beyond short snippets: Deep networks for video classification, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp.4694–4702.

[24] V. Veeriah, N. Zhuang and G.J. Qi, Differential recurrent neural networks for action recognition, in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 4041–4049.

[25] S. Venugopalan, M. Rohrbach, J. Donahue, R. Mooney, T. Darrell and K. Saenko, Sequence to sequence-video to text, in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 4534–4542.

[26] O. Vinyals, A. Toshev, S. Bengio and D. Erhan, Show and tell: A neural image caption generator, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3156–3164.

[27] J. Wang, Z. Liu, Y. Wu and J. Yuan, Learning actionlet ensemble for 3D human action recognition, *IEEE transactions on pattern analysis and machine intelligence* 36 (2014) 914–927.

[28] Y. Wen, K. Zhang, Z. Li and Y. Qiao, A discriminative feature learning approach for deep face recognition, in *European Conference on Computer Vision*, Springer, 2016, pp. 499–515.

[29] X. Wang, L. Gao, J. Song and H. Shen, Beyond frame-level CNN: saliency-aware 3-D CNN with LSTM for video action recognition, *IEEE Signal Processing Letters* 24 (2017) 510–514.

[30] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel and Y. Bengio, Show, attend and tell: Neural image caption generation with visual attention, in *International conference on machine learning*, 2015, pp. 2048–2057.

[31] J. Xu, T. Mei, T. Yao and Y. Rui, Msr-vtt: A large video description dataset for bridging video and language, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 5288–5296.

[32] M. Zhu, K. Fu, X. Huang, S. Wu and W. Jin, A novel recognition approach for radar emitter signals based on on-line independent support vector machines, *Advances in Computer Science and Its Applications* 2 (2013) 390–396.

Manuscript received 5 March 2018
revised 10 September 2018
accepted for publication 20 November 2018

Yongjian Shen
School of Electronics and Information Engineering
Beihang University, Beijing, China
E-mail address: shenyongshen@163.com

Chunyu Xie
School of Automation Science and Electrical Engineering
Beihang University, Beijing, China
E-mail address: yuxie@buaa.edu.cn

Zhendong Liu
School of Automation Science and Electrical Engineering
Beihang University, Beijing, China
E-mail address: liuzhendong@buaa.edu.cn

Jing Dai
China Academy of Launch Vehicle Technology
Beijing, China
E-mail address: buaadaij@126.com

Baochang Zhang
School of Automation Science and Electrical Engineering
Beihang University, Beijing, China
Shenzhen Academy of Aerospace Technology, Shenzhen, China
E-mail address: bczhang@buaa.edu.cn

Lei Zhu
School of Life Information and Instrument Engineering
Hangzhou Dianzi University, Hangzhou, China
E-mail address: zhulei@hdu.edu.cn

Jingwen Li
School of Electronics and Information Engineering
Beihang University, Beijing, China
E-mail address: lijingwen@buaa.edu.cn