Advances in Intelligent Systems Research, volume 147
International Conference on Network, Communication, Computer Engineering (NCCE 2018)

Research on the Recognition of Offline Handwritten New Tai Lue Characters Based on Bidirectional LSTM

Yongqiang Wang [1, a)], Pengfei Yu [2, b)], Hongsong Li [2)] and Haiyan Li [2)]

[1]*School of information, Yunnan University, Kunming 650000, China*
[2]*School of information, Yunnan University, Kunming 650000, China.*

[a)] yongqiang.w@139.com, [b)] pfyu@ynu.edu.cn

Abstract. Deep learning has made breakthrough progress in image recognition, target detection and tracking in recent years. It is proved too good at classification tasks. In this paper, we have compared use of Convolutional neural network(CNN), VGG16, Long Short-Term Memory (LSTM), and Bidirectional LSTM to perform offline handwriting New Tai Lue Characters recognition. These methods have been tested on a dataset build by our laboratory. For testing purpose 58795 samples including 9834 test samples of handwriting New Tai Lue Characters are used in these experiments. The experimental results show that the recognition rates are 91.23%, 89.33%, 92.78% for CNN, VGG16 and LSTM. Moreover, the best recognition result is obtained with the Bidirectional LSTM based method, whose recognition rate is 94.87% on the dataset.

Key words: Offline handwritten character recognition, Bidirectional LSTM, New Tai Lue characters.

INTRODUCTION

Thai alphabet, used by Thai people, emerges as alphabetic writing evolved from Brahmi of ancient India, including seven types, that is, Siam Thai(Thai), Lan Xang Thai(Lao), Lanna Thai (Tai Tham script), Vietnamese Thai (Daiduan writing), Southern Thai (Old shan writing), Northern Thai (North Thai written words) and Assam Thai (Old Assam written words). China covers Lanna Thai, Vietnamese Thai, Southern Thai and Northern Thai, four types of characters in total. During the 1950s, New Tai Lue is developed based on the traditional Tai Tham scripts in Xishuangbanna, China. From then on, it is become the official alphabet for Dai people lived in Xishuangbanna.

Optical character recognition(OCR) [1] has been quite hot in the area of pattern recognition, enjoying a significant application prospect, in fields like car license number recognition, mail sorting and bill identification. It is very mature in research aspects of OCR such as numbers, English letters and Chinese characters. While Chinese minorities' languages, New Tai Lue characters in particular, have in fact seen few studies about them. Research on the recognition of their offline handwritten characters will contribute to the development of their economy, culture and education, so it is more than essential to conduct recognition study about New Tai Lue characters.

OCR consists of online [2] and offline [3] in the light of the attaining methods of the target character. Online character recognition means the target character to be recognized is the real-time handwriting obtained through the tablets, mobile screens, etc. As for the offline one, it refers to the recognition about the intact character which is already written down or scanned. This process involves preprocessing, feature extraction and classification. The offline handwritten character recognition is more difficult than the online one. Methods available in character recognition can be generally sorted into three ones, that is, template matching, feature extraction and classification, and deep learning [4,5]. The template matching method was the most commonly used method in earlier research. To begin with, it needs to build a standard template set for the target characters. What follows is matching the image of the target with the templates in the very set one by one. The recognition result shows up as the character

corresponding to the template with the highest matching score. Feature extraction and classification is the method mostly seen in OCR. Containing two parts, its first step is to extract the characteristics of the character image with an algorithm such as Scale Invariant Feature Transform(SIFT), Speeded Up Robust Features(SURF), Histogram of Oriented Gradient(HOG), etc. Its second step is to classify the acquired features by a classifier (such as Bayesian classifier, K-nearest neighbor algorithm, artificial neural network algorithm, etc.); Deep learning methods are currently the most popular in many areas, while in the field of character recognition, one of these models is called CNN [6] which has achieved great progress in character recognition tasks with fantastic results since LeNet-5[7].

This paper applied many deep learning methods in offline handwritten New Tai Lue characters recognition, such as Convolutional neural network(CNN), VGG16[8], Long Short-Term Memory (LSTM) [9,10], and Bidirectional LSTM [11].

The rest of this paper is organized as follows: related works are introduced in Section II; in Section III we further describe the basic mechanism and theoretical derivation of Bi-LSTM; Section IV shows the experimental results of many deep learning methods; a conclusion and a brief outlook on future research is given in Section V.

RELATED WORK

Neural network commonly used in OCR includes CNN, classical CNN-VGGNet [12], Recurrent Neutral Network(RNN)-LSTM and Bidirectional LSTM.

CNN

The CNN, one of the deep learning models, mainly differs from those common models for two strong assumptions in the model. The first one is parameter sharing. Usually, a filter in CNN does not have many parameters, for instance, a filter at the size of 5×5×3 merely asks for 75 parameters, which is equal to the connection of the hidden layers and the partial input in comparison with the neural network. The second one is the correlation of partial pixels, based on which, Max Pooling [13], a kind of processing technology, is derived, that is, take the maximum of the partial pixel in certain image block and the image dimensionality will be reduced at a square-ratio speed, as a result, time and space consumed by CNN are decreased greatly. The structure of CNN is as shown in Fig. 1.

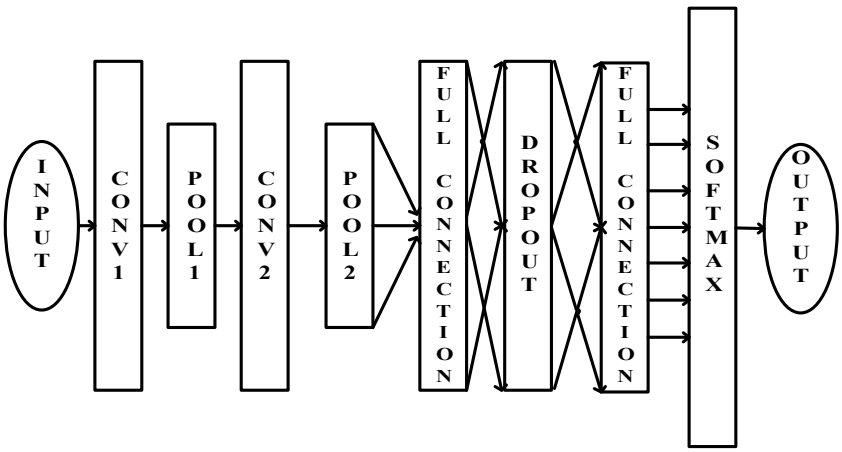

FIG. 1. The structure of CNN

As shown in the Fig. 1, a CNN is mainly made of input layers, convolution layers, pooling layers, fully connected layers and a SoftMax layer. (The detailed introduction about each layer omitted here.) The first convolution layer in the CNN would directly accept the input of the images at the pixel level. Only one piece of the image would be processed in each convolution operation and it would then be fed into the following network. The most effective feature in a block of an image would be extracted in each convolution operation. After that these fundamental features of an image would be re-combined and abstracted to make the advanced feature possible.

Theoretically speaking, these features mentioned above are stable and unaffected by the scaling, translation and rotation of the image.

Classical CNN-VGGNet.

VGGNet, the deep convolution neural network, was developed with joint efforts of the Computer Visual Group from Oxford University and the researchers from Google DeepMind Company. They studied the relationship between the depth of the CNN and its performances. Then, a convolution network of 16 to 19 layers is made by VGGNet via repeatedly piling the small 3×3 convolution kernel and the largest pooling layer of 2×2. The network structure of VGG16 is shown in Fig. 2 below.

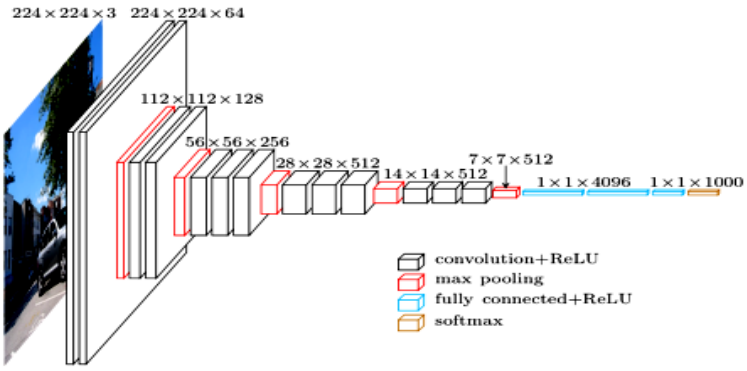

FIG. 2. The network structure of VGG16

VGGNet has many advantages, such as excellent abilities of extensible and generalization for image data. In addition, its structure is also quite concise. Image feature can usually be extracted with VGGNet for it could improve performance through consistently deepening the network structure.

Long Short-Term Memory.

Long Short-Term Memory(LSTM), a special RNN, was proposed by Professor Schmidhuber in 1997, designed to solve the problem of long-term dependencies. The interior structure of LSTM tends to be more intricate than the RNN and it makes information selectively cast impact on the RNN's state at each moment through some 'gate' structure. Its structure is presented in Fig. 3.

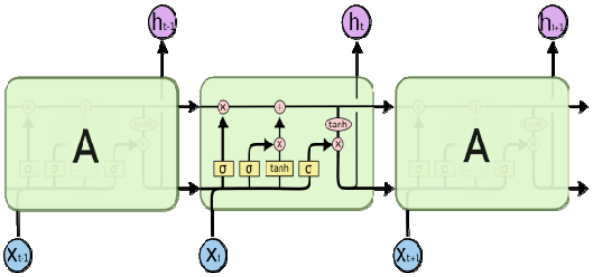

FIG. 3. The illusion of the LSTM structure

As shown in the above figure, LSTM comprises 4 layers of neural network, inside which, the small circle stands for the point-wise operation while the small rectangle refers to a neural network layer capable of leaning parameters. The straight line on LSTM reflecting the status of it runs through all the series LSTM units, flowing from the first unit to the last one merely with few linear interference and change. When the status is conveyed in this channel, the unit of LSTM could add or delete information on it, which is controlled by the 'gate'. The LSTM unit structure is presented in the Fig. 4.

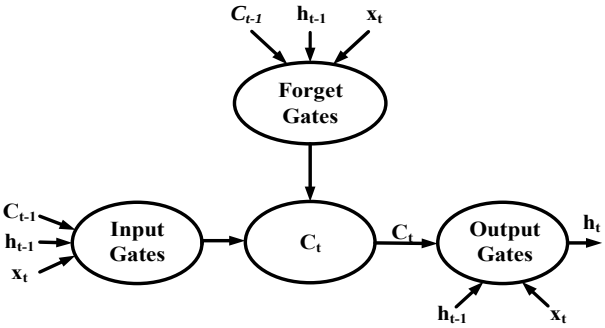

FIG. 4. The illusion of LSTM's unit structure

The composition of the 'gate' is an operation of a neural network with the sigmoid function and a bitwise multiplication. As a popular activation function of artificial neurons in a fully connected neural network, the return value of sigmoid function, which ranges from 0 to 1, describes how much information could get through this structure. If the return value is 1, all the information could pass it, but if the return value is 0, none of the information could pass it then.

METHODS INVOLVED

Based on the previous section, this paper conducted training and recognition about the offline handwritten New Tai Lue characters on the Bi-LSTM as the Bidirectional LSTM is superior to the unidirectional LSTM and could employ the information in the positive time direction (forward states), and another for negative time direction (backward states), making the final prediction more accurate.

The Basic Structure and Theory of Bi-LSTM

The major structure of the Bi-LSTM is the combination of two LSTM. At every moment t, the input would be provided to these two neural networks in opposite directions while the output would be decided by these two unidirectional LSTM. The illusion of its network is demonstrated in the Fig. 5 and Fig. 6.

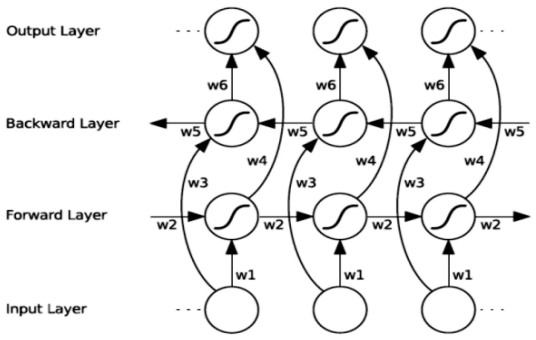

FIG. 5. The illusion of Bi-LSTM's structure

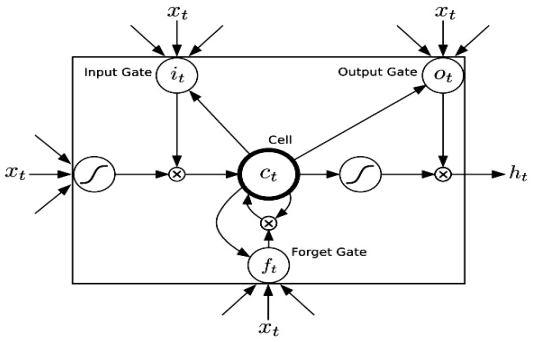

FIG. 6. The illusion of Bi-LSTM's memory module

The output sequence of the Bi-LSTM with a fully connected layers was followed by a SoftMax layer at length, which is similar to the classification through the convolution network's output.

Besides, the construction of its memory module(Cell) [14] is shown in the Fig. 6. Cell is similar to a processor judging whether the information is useful or not. There is an input gate, forget gate and output gate placed in one Cell, that is, the employment of one inlet and two outlets working principle. The complex and artificial long-time-lag tasks in the neural network could be settled via its repeated operation.

Mathematical Representation of Bi-LSTM

As for an input x whose length is T, in the network lies I input units, H hidden units and K output units. Let us define the x_i^t as the ith input at time t, the a_j^t and b_j^t as the input of the net unit j at time t and the output of the unit j's nonlinear differentiable activation function at time t, respectively.

Calculating forward:

Input Gate:

$$a_l^t = \sum_{i=1}^{I} w_{il} x_i^t + \sum_{h=1}^{H} w_{hl} b_h^{t-1} + \sum_{c=1}^{C} w_{cl} s_c^{t-1} \tag{1}$$

$$b_l^t = f(a_l^t) \tag{2}$$

Forget Gate:

$$a_\phi^t = \sum_{i=1}^{I} w_{i\phi} x_i^t + \sum_{h=1}^{H} w_{h\phi} b_h^{t-1} + \sum_{c=1}^{C} w_{c\phi} s_c^{t-1} \tag{3}$$

$$b_\phi^t = f(a_\phi^t) \tag{4}$$

Cell:

$$a_c^t = \sum_{i=1}^{I} w_{ic} x_i^t + \sum_{h=1}^{H} w_{hc} b_h^{t-1} \tag{5}$$

$$s_c^t = b_\phi^t s_c^{t-1} + b_l^t g(a_c^t) \tag{6}$$

Output Gate:

$$a_w^t = \sum_{i=1}^{I} w_{iw} x_i^t + \sum_{h=1}^{H} w_{hw} b_h^{t-1} + \sum_{c=1}^{C} w_{cw} s_c^t \tag{7}$$

$$b_w^t = f(a_w^t) \tag{8}$$

Cell Output:

$$b_c^t = b_w^t h(s_c^t) \tag{9}$$

Calculating backwards:

$$\varepsilon_c^t \overset{def}{=} \frac{\partial O}{\partial b_c^t} \quad \varepsilon_s^t \overset{def}{=} \frac{\partial O}{\partial s_c^t} \tag{10}$$

Cell Output:

$$\varepsilon_{\mathrm{c}}^{t}=\sum_{k=1}^{K} w_{ck}\delta_{k}^{t}+\sum_{h=1}^{H} w_{ch}\delta_{h}^{t+1} \tag{11}$$

Output Gate:

$$\delta_{\mathrm{w}}^{t}=f^{'}(a_{w}^{t})\sum_{c=1}^{C} h(s_{c}^{t})\varepsilon_{c}^{t} \tag{12}$$

Cell:

$$\delta_{\mathrm{c}}^{t}=b_{l}^{t}g^{'}(a_{c}^{t})\varepsilon_{s}^{t} \tag{13}$$

$$\varepsilon_{\mathrm{s}}^{t}=b_{w}^{t}h^{'}(s_{c}^{t})\varepsilon_{c}^{t}+b_{\phi}^{t+1}\varepsilon_{\mathrm{s}}^{t+1}+w_{cl}\delta_{l}^{t+1}+w_{c\phi}\delta_{\phi}^{t+1}+w_{cw}\delta_{w}^{t} \tag{14}$$

Forget Gate:

$$\delta_{\phi}^{\mathrm{t}}=f^{'}(a_{\phi}^{t})\sum_{c=1}^{C} s_{c}^{t-1}\varepsilon_{s}^{t} \tag{15}$$

Input Gate:

$$\delta_{l}^{\mathrm{t}}=f^{'}(a_{l}^{t})\sum_{c=1}^{C} g(a_{c}^{t})\varepsilon_{\mathrm{s}}^{t} \tag{16}$$

The equations above only illustrate the unidirectional long short memory network while the major structure of the bidirectional LSTM is actually the combination of two unidirectional ones.

EXPERIMENT RESULTS AND DISCUSSION

For comparison these deep learning methods used in handwriting New Tai Lue characters recognition, some experiments are carried out on a handwriting New Tai Lue characters database which is collected by our laboratory.

Experiment Platform

The experiments of the offline handwritten New Tai Lue characters recognition in this paper are implemented on a Lenovo desktop PC (Ubuntu16.04 64-bit OS, Core i7-6700 processor, NVIDIA GT730 graphics card) with Keras and TensorFlow deep learning framework, python language and PyCharm developing tool involved.

Experiment Data Set

Data used in this paper is covered by a lab-built database. An android APP and a web software were developed for the collection task of handwritten New Tai Lue characters. The processed sample is as shown in Fig. 7.

FIG. 7. Sample data of handwriting New Tai Lue characters.

Note that there are 83 scripts in New Tai Lue alphabet, and all New Tai Lue characters are normalized to the same size 32×32. Besides, these samples where was converted into the TF Records format which is the default format of TensorFlow. The total number of samples is 58795 while that for the test sample is 9834, As a result, the ratio of the training sample to the test sample is about 5:1.

Experiment Results and Discussion

The training results of CNN, VGG16, LSTM, Bi-LSTM and the recognition rate curves of these methods are shown in the figures below.

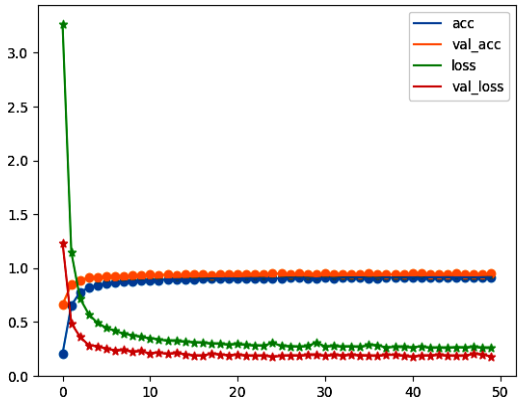

FIG. 8. CNN's recognition rate and the loss curve

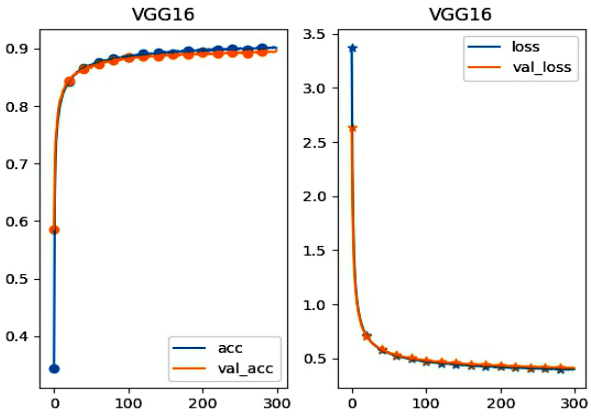

FIG. 9. VGG16's recognition rate and the Loss curve

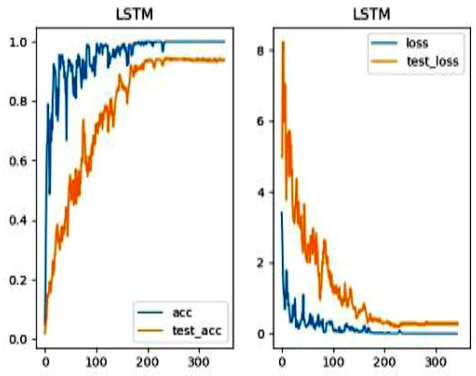

FIG. 10. LSTM's recognition rate and the loss curve

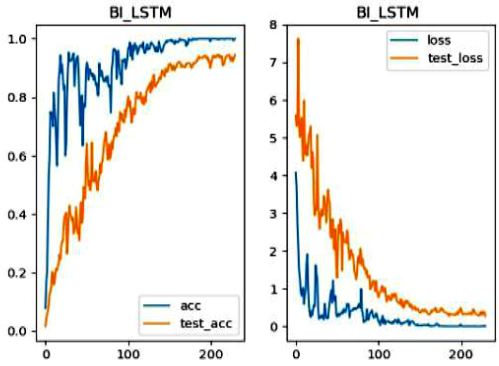

FIG. 11. Bi-LSTM's recognition rate and the Loss curve

TABLE 1. Comparison of experiment results

Net	accuracy	loss
CNN	91.23%	0.3087
VGG16	89.33%	0.4353
LSTM	92.78%	0.3223
Bi-LSTM	94.87%	0.2836

The recognition rate in this experiment for CNN, VGG16 and LSTM is 91.23%, 89.33% and 92.78% respectively while that for Bi-LSTM hits 94.87%. Compared with other three methods, it has achieved encouraging progress.

CONCLUSION

This paper applied four popular Deep Learning methods, including CNN, VGG16, LSTM and Bidirectional LSTM to recognition of offline handwriting New Tai Lue characters. The best recognition rate 94.87% is obtained with the Bidirectional LSTM. The experimental results show the effectiveness of the proposed method. In the future work, we will study to adjust the structure of the layers in Deep Learning methods, such as replacing the SoftMax layer with a support vector machine.

ACKNOWLEDGEMENTS

The authors would like to thank the financial supports by National Natural Science Foundation of China (Grant No.61462094, 61561050) and Natural Science Foundation of Yunnan Province of China (Grant No.2015FB116).

REFERENCES

1. Fujii Y, Driesen K, Baccash J, et al. Sequence-to-Label Script Identification for Multilingual OCR[J]. 2017.
2. Lin I J, Kung S Y. A novel learning method by structural reduction of DAGs for on-line OCR applications[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2007:1069-1072 vol.2.
3. Vamvakas G, Gatos B, Pratikakis I, et al. Hybrid off-line OCR for isolated handwritten Greek characters[M]. 2008.
4. Lecun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553):436.
5. Schmidhuber J. Deep Learning in neural networks: An overview. [J]. Neural Networks the Official Journal of the International Neural Network Society, 2015, 61:85-117.
6. Fedorovici L O, Precup R E, Dragan F, et al. Evolutionary optimization-based training of convolutional neural networks for OCR applications[M]. 2013.
7. Lecun Y. LeNet-5, convolutional neural networks[J].
8. Liu B, Zhang X, Gao Z, et al. Weld Defect Images Classification with VGG16-Based Neural Network[M]// Digital TV and Wireless Multimedia Communication. 2018.
9. Otte S, Liwicki M, Krechel D. Investigating Long Short-Term Memory Networks for Various Pattern Recognition Problems[C]// International Workshop on Machine Learning and Data Mining in Pattern Recognition. Springer, Cham, 2014:484-497.
10. Zhu W, Lan C, Xing J, et al. Co-occurrence feature learning for skeleton based action recognition using regularized deep LSTM networks[C]// Thirtieth AAAI Conference on Artificial Intelligence. AAAI Press, 2016:3697-3703.
11. Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures[J]. Neural Networks the Official Journal of the International Neural Network Society, 2005, 18(5):602-610.
12. Wang L, Guo S, Huang W, et al. Places205-VGGNet Models for Scene Recognition[J]. Computer Science, 2015.
13. Graham B. Fractional Max-Pooling[J]. Eprint Arxiv, 2014.
14. Rahman L, Mohammed N, Azad A K A. A new LSTM model by introducing biological cell state[C]// International Conference on Electrical Engineering and Information Communication Technology. IEEE, 2017:1-6.