



# Attention Enhanced ConvNet-RNN for Chinese Vehicle License Plate Recognition

Shiming Duan, Wei Hu, Ruirui Li<sup>(✉)</sup>, Wei Li, and Shihao Sun

Beijing University of Chemical Technology, Beijing 100029, China  
ilydouble@gmail.com

**Abstract.** As an important part of intelligent transportation system, vehicle license plate recognition requires high accuracy in an open environment. While a lot of approaches have been proposed, and achieved good performance to some extent, these approaches still have problems, for example, in the condition of characters' distortion or partial occlusion. Segmentation-free VLPR systems compute the label in one pass using Long Short-Term Memory Network (LSTM), without individual segmentation step, their results tend to be not influenced by the segmentation accuracy. Based on the idea of Segmentation-free VLPR, this paper proposed an attention enhanced ConvNet-RNN (AC-RNN) for accurate Chinese Vehicle License Plate Recognition. The attention mechanism helps to locate the important instances in the step of recognition. While the ConvNet is used to extract features, the recurrent neural networks (RNN) with connectionist temporal classification (CTC) are applied for sequence labeling. The proposed AC-RNN was trained on a large generated dataset which contains various types of license plates in China. The AC-RNN could figure out the vehicle license even in cases of light changing, spatial distortion and partial blurry. Experiments showed that the AC-RNN performs better on the testing real images, increasing about 5% on accuracy, compared with classic ConvNet-RNN [8].

**Keywords:** Vehicle license plate recognition  
Recurrent neural networks · Long Short-Term Memory Network  
Attention

## 1 Introduction

As an important part of intelligent transportation system, vehicle license plate recognition (VLPR) has attracted considerable research interests. It is a useful technology for government agencies to track or detect stolen vehicles or collect data for traffic management and improvement. Due to its close relationship to public security, VLPR requires generalization and high accuracy in real applications.

While a lot of works have been proposed on the topic of VLPR, the VLPR task is still challenging, not only because of the environmental factors such as lighting, shadows, and occlusions, but also because of the image acquisition factors such as motion and focus blurs. For Chinese license plate recognition, the situation is more complicated. They are composed of Chinese characters and numbers. Their colors and sizes may be different and their lengths are not necessarily fixed, even placed in two lines.

Traditional image processing method needs a series of processing steps, including localization, segmentation and recognition. Many of them depend on handcrafted features and could only work well under controlled conditions. These handcrafted features are usually sensitive to image noises, and may result in many false positives under complex backgrounds.

CNNs have achieved great success in various tasks including image classification, object detection and semantic segmentation [11]. CNNs containing deep neural layers can learn efficient representations from a large amount of training data. For the VLPR tasks, extended CNNs transform the one-to-many problem into a single-label classification problem by classifying one character at a time. This requires the task to firstly segment characters and then recognize them one by one. More recent work performs segmentation before classification and use CNN-BRNN [14] and CTC to achieve the state of the art results. The segmentation-free VLPR [8] focus on images from real-world traffic cameras and applies the ConvNet RNN to the VLPR tasks. Unfortunately, their method takes no consideration of Chinese license plate recognition and requires specific optimization for higher accuracy.

In this paper, we proposed an attention enhanced ConvNet-RNN for Chinese vehicle license plate recognition. It is one-pass, end-to-end neural network. The proposed AC-RNN has two improvements. The original ConvNet-RNN is actually not fit for the vehicle license plate recognition with weak semantic connections. Thus, a novel semantic enhance strategy is introduced which inserts some trivial null characters into labeled strings. Secondly, an attention mechanism is added to learn the weights map, helping the neural network to perform classification better. The two techniques are not individual with each other. They work together to get higher accuracy. Furthermore, to avoid overfitting caused by lack of data. We also proposed a data generation method and generated a dataset containing one million labeled images. In summary, this paper makes the follows contributions to the community:

- A novel semantic enhance strategy for Chinese VLPR.
- An attention enhanced ConvNet-RNN.
- A data generation method and a new dataset of Chinese VLP.

This paper is organized as thus, Sect. 2 presents related works about this paper and Sect. 3 provides the proposed neural network, In Sect. 4, a series of experiments will be presented and the conclusion will be shown in Sect. 5.

## 2 Related Works

### 2.1 Vehicle License Plate Recognition

Approaches for VLPR problems contain two stages, localization and recognition. The main work of this paper focus on the latter stage—recognition.

Plate localization aims to detect plates from images. Lots of work has been done for detection problems, for example, Faster RCNN [28] is known for its high precision of detection. YOLO [27] is famous for its speed of detection. Like a combination of Faster RCNN and YOLO, SSD [20] performs very well on both accuracy and speed. In [30], CPTN is designed for texts detection in natural images. Plate localization aims to detect plates rather than texts, therefore, a SSD model is applied in our project to detect plates.

Previous works on plate recognition include traditional image processing methods [1, 2, 12, 13, 15, 25, 26, 31] and new deep learning approaches [3, 16, 18, 19, 21, 23, 33]. Most of them need to detect and segment the characters out from a license plate image before recognition.

The recognition of plates typically contains two-stages as well. Segmentation extracts characters from license plate image; and classification distinguishes the segmented characters one by one. For example, in [12], Gou *et al.* propose Extremal Regions (ER) to segment characters from coarsely detected license plates. Restricted Boltzmann machines were applied to recognize the characters. In [33], the license plate is segmented into seven blocks using a projection method, after which, two classifiers are designed to recognize Chinese characters, numbers, and alphabet letters. In [18, 23], two CNN models are used to recognize characters from plate image. Firstly, a binary deep network classifier is trained to confirm if a character exists. Another deep CNN is adopted for the task of character recognition. In [21], Liu *et al.* implement a CNN model which has shared hidden layers and two distinct softmax layers for the Chinese and the alphanumeric characters respectively.

In fact, character segmentation by itself is a challenging task since it is prone to be influenced by uneven lighting, shadows and noises in the images [18]. The plate cannot be recognized correctly if the segmentation is improper, even if we have a strong classifier for characters. Therefore, in this paper, VLPR is regarded as a sequence labeling problem, and our proposed method aims to recognize plates without segmentation.

### 2.2 ConvNet-RNN

One of the most popular methods for sequence-to-sequence problems is ConvNet-RNN (CRNN) [14]. ConvNet-RNN is proposed for image-based recognition in [29], Shi *et al.* integrate feature extraction, sequence modeling and transformation into a unified framework. CRNN is end-to-end trainable and can deal with sequence in various lengths, involving no character segmentation or horizontal scale normalization. CRNN has been popular since it is proposed, for example, in

[10], CRNN is adopted for offline handwriting recognition, in [7, 32], CRNN provides a useful method for Optical Music Recognition (OMR), and in [6], CRNN is used for script identification in natural scene image and video frame.

CRNN is also applied to VLPR. In [8], CRNN is adopted for plate recognition with CTC. However, due to the weak correlation between characters in plate, classic CRNN does not work well on VLPR, therefore, this paper introduces the method of generating interval characters to strengthen the correlation between characters combined with a fixed length CTC.

### 2.3 Attention Model

Attention mechanism becomes more and more popular since being used in image classification by Mnih *et al.* [24]. In [4], Dzmitry *et al.* first introduce attention mechanism to neural machine translation (NMT). By learning different weights from the source parts to different target words, the trained Model can automatically search for parts of a source sentence that are relevant to a target word. In [22], Luong *et al.* show that how to extend RNN with attention mechanism. Global attention and local attention are introduced in natural language processing (NLP). After [4, 22], attention mechanism is widely used in NLP tasks, including not only sequence-to-sequence models, but also various classification tasks. In [5], attention is used to allow a Recurrent Neural Network (RNN) to learn alignments between sequences of input frames and output label on the topic of Large Vocabulary Continuous Speech Recognition (LVCSR) system. In [9], Serdyuk adopted the attention mechanism for speech recognition. Attention is also used for OCR in [17].

## 3 The Proposed AC-RNN Framework

### 3.1 Network Architecture

The network architecture of AC-RNN is shown in Fig. 1. It contains three main parts: the ConvNet, the attention based RNN and the CTC. The AC-RNN takes plate images as input. Through deep convolutional neural layers, the AC-RNN learns a group of feature maps. It then departs the feature map into sequence feature blocks and sends them to Bi-LSTM RNN. After encoding and decoding training processes, pre-frame predictions are gotten. Then a length-fixed CTC is performed to classify the characters. The AC-RNN works consecutively in an end-to-end fashion which inputs a plate image and outputs the predicted labels.

### 3.2 Interval for Semantic Enhancement

LSTM is well known for its ability to capture long-range dependencies. Therefore, LSTM is widely used in voice recognition, Optical Character Recognition (OCR), text categories and so on. As often observed, the characters of vehicle license plate have semantic connections in the context. Since the LSTM is

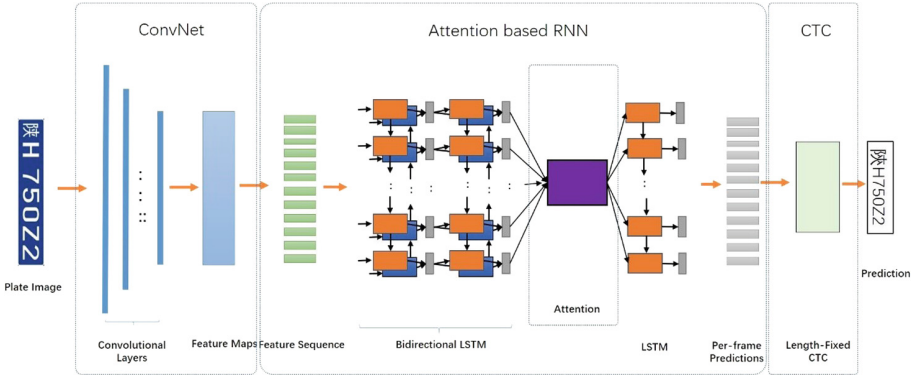


Fig. 1. Attention enhanced ConvNet-RNN.

directional, to make the best use of context information, the AC-RNN adopts a bidirectional LSTM. On the other hand, differed with language translation or OCR tasks, which the LSTM is intuitively fit for, the characters on a plate are weakly relevant. Following the rules, some characters are fixed according to the car properties while other characters may be generated randomly.

In order to strengthen the correlations between characters on a plate, characters named empty of sequence (EOS) are inserted at the intervals to the sequential labels when training the model.

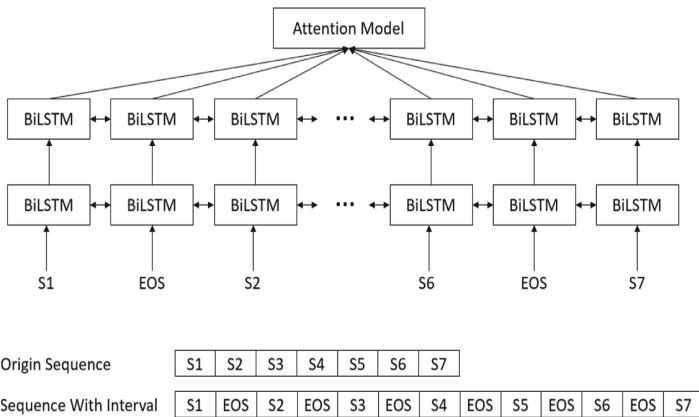


Fig. 2. Sample of interval character in sequence.

As shown in Fig. 2, the EOS participates in the training of the attention-based recurrent neural network. In the case of Chinese vehicle license plate, the AC-RNN adopts the rule that inserting the EOS at each interval between neighbouring pairs of characters showed with solid EOS labeled rectangles. The

interval characters could help distinguishing the gaps of characters in the plate. It also strengthens the correlation of sequence that the LSTM need. The EOS in the prediction sequence needs removing before output.

### 3.3 Attention Based RNN Decoder

Differed with classic CRNN, attention based CRNN has an additional attention model to learn the weights map, helping the neural network to perform classification better. As shown in Fig. 3, the attention model tries to learn a weights map  $a_{ij}$ , which could tell how relevant is the part of the source hidden  $h_j$  to a target frame. Then the RNN input  $c_i$  could be calculated by the following equation.

$$c_i = \sum_{j=1}^{T_X} a_{ij}h_j \tag{1}$$

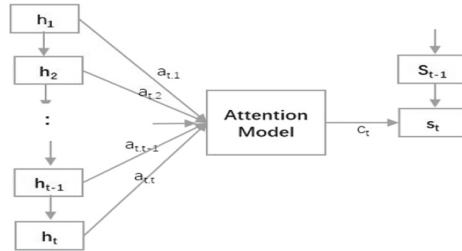


Fig. 3. Attention mechanism in our AC-RNN.

### 3.4 Length-Fixed CTC Decoder for Vehicle License Plate

The RNN will output a sequence that contains per-frame predictions. To get the final label sequence, two main steps are taken by the CTC. They are Merge Repeated and Remove Blank. As shown in the Fig. 4, to get the correct output ‘HELLO’, there must have a blank token between ‘LL’, and with this blank token, we obtain ‘HELLO’ rather than ‘HELO’.

The VLPs usually have a certain length, for example, in China, the length of VLP is seven. According to the proposed method, the length of the final output result is checked in the CTC step. If the CTC generates a sequence of uncommon length, the AC-RNN will find the longest continuous sub strings without blank. This step is illustrated by label (2) in Fig. 4. In this situation, the AC-RNN will put a blank into the longest substring in force, which guarantees the right output. This step is illustrated by label (1) in Fig. 4.

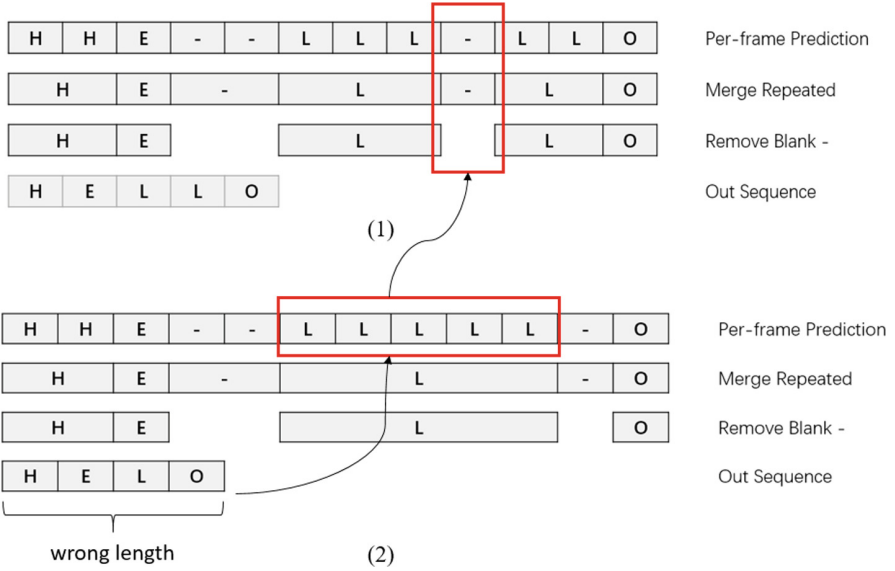


Fig. 4. Illustration of CTC progress.

## 4 Experiments

### 4.1 Experimental Environments

Our experiments are carried out on the 8-way GPU cluster, whose configuration is shown in Table 1. We design the experiments with caffe and try to compare general ConvNet-RNN and our AC-RNN with length-fixed CTC.

Table 1. Experimental environments.

Operating system	Red Hat 4.8.3-9
CPU	Intel(R) Xeon(R) CPU E5-2678 v3 @ 2.50 GHz
GPU	GeForce GTX TITAN X
Hard disk	1TB
cuda	4.0.7
CUDA	7.0.27

### 4.2 Generated Datasets

On account of various reasons, it is always difficult to obtain VLP datasets, not to mention balanced datasets over the country. To avoid overfitting caused by lack of data and to enhance the robustness of our model, a data generation

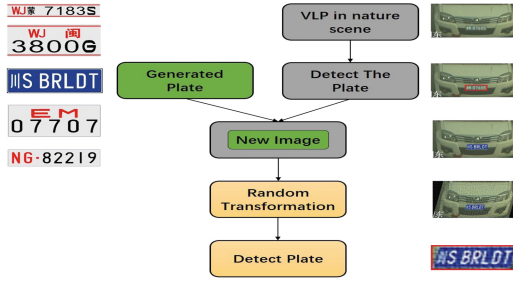


Fig. 5. The method proposed to generate VLP data.



Fig. 6. Examples of images generated by our method.

method is proposed. By this method, plenty of plate images are generated, which can be used both in VLP detection and VLP recognition.

As shown in Fig. 5, our method to generate data set is shown as following.

Step 1: A dataset from monitor cameras is necessary, with which some of lean VLPs will be detected by a detector. The remaining part without the plate will be used as background.

Step 2: Lots of VLPs with random character distribution will be plotted according to template from transportation department.

Step 3: The plates detected in step 1 will be replaced by plates from step 2, thus new images will be generated with nature backgrounds and manual plates.

Step 4: A series of transformations will be applied to these generated data, such as random scale, random blur, random rotate, random sharpen etc.

Step 5: After detected by a detector, our new VLP dataset will be ready.

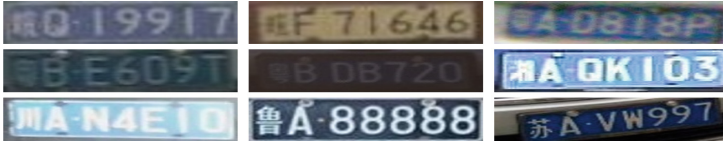
By our method, datasets of VLP containing millions of images can be generated for detection and recognition. Some examples of our generated images are shown in Fig. 6.

### 4.3 Experiments and Results

In our experiments, dataset for training and valuation contains 400 thousand images collected from natural scene and 800 thousand images generated by our data generation method. A dataset from EasyPR containing 260 images is set



for test, all these images are collected from natural scene. Some instances of the test dataset are shown in Fig. 7.



**Fig. 7.** Some instances in test dataset

The main work of this paper focus on the stage of recognition. As described in Sect. 3, contributions to network in this paper are semantic enhancement of plates, length-fixed CTC decoder and attention mechanism, therefore our contrast experiments are carried out on two different stages: The classic ConvNet-RNN with nothing enhanced and our AC-RNN. We test these models and summarize the results as Table 2. As shown in Table 2, compared with the classic ConvNet-RNN, our work in this paper makes an excellent improvement on accuracy.

**Table 2.** Evaluation results.


Comparison	Accuracy overall
Classic ConvNet-RNN	85.17%
AC-RNN	90.11%

A classic method to evaluate the performance of sequence labeling is to measure the percentage of perfectly predicted images in the test dataset, as shown in Table 2. Considering vehicle license plates in China contain a Chinese character and the repeated characters can only appeared in alphabets and numbers, the accuracy of the last six bits is calculated in Table 3. The accuracy of the Chinese characters is also listed in results. Meanwhile, some instances are given in Fig. 8.

**Table 3.** Evaluation results on our VLP dataset.

Comparison	Accuracy of last six bits	Accuracy of Chinese characters
Classic ConvNet-RNN	91.25%	88.97%
AC-RNN	95.44%	93.54%

As shown in Table 3 and Fig. 8, the classic ConvNet-RNN performs bad when text on a plate contains several same connected characters. As comparison, the AC-RNN proposed in this paper does not have this problem, and with attention mechanism, the AC-RNN performs better on recognition.

Test image	Ground truth	ConvNet-RNN	AC-RNN
	川A9J333	川A9J33	川A9J333
	渝BE7773	渝BE773	渝BE7773
	苏ANC818	苏AN0818	苏ANC818
	鲁A88888	鲁A8888	鲁A88888

**Fig. 8.** Some examples in experiments

## 5 Conclusion

In this paper, an attention enhanced ConvNet-RNN for Chinese Vehicle License Recognition, AC-RNN, is proposed. Compared with classic ConvNet-RNN, there are two improvements in AC-RNN. Firstly, intervals for semantic enhancement and length-fixed CTC decoder are firstly introduced to VLPR problems, in which intervals can strengthen the correlations between characters on a plate and a length-fix CTC decoder can perform better when there are several connected same characters on a plate. Secondly, attention mechanism is applied to learn a weights map, helping the neural network to perform classification better. Besides, a new method to generate datasets of VLP is proposed, thus, the overfitting caused by lack of data can be avoided.

## References

1. Aboura, K., Al-Hmouz, R.: An overview of image analysis algorithms for license plate recognition. *Organizacija* **50**(3), 285–295 (2017)
2. Abtahi, F., Zhu, Z., Burry, A.M.: A deep reinforcement learning approach to character segmentation of license plate images. In: 2015 14th IAPR International Conference on Machine Vision Applications (MVA), pp. 539–542, May 2015. <https://doi.org/10.1109/MVA.2015.7153249>
3. Angara, N.S.S.: Automatic license plate recognition using deep learning techniques (2015)
4. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. arXiv preprint [arXiv:1409.0473](https://arxiv.org/abs/1409.0473) (2014)
5. Bahdanau, D., Chorowski, J., Serdyuk, D., Brakel, P., Bengio, Y.: End-to-end attention-based large vocabulary speech recognition. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4945–4949. IEEE (2016)
6. Bhunia, A.K., Konwer, A., Bhowmick, A., Bhunia, A.K., Roy, P.P., Pal, U.: Script identification in natural scene image and video frame using attention based convolutional-LSTM network (2018)
7. Calvo-Zaragoza, J., Valero-Mas, J.J., Pertusa, A.: End-to-end optical music recognition using neural networks. In: Proceedings of the 18th International Society for Music Information Retrieval Conference, Suzhou, China, pp. 23–27 (2017)

8. Cheang, T.K., Chong, Y.S., Yong, H.T.: Segmentation-free vehicle license plate recognition using ConvNet-RNN (2017)
9. Chorowski, J.K., Bahdanau, D., Serdyuk, D., Cho, K., Bengio, Y.: Attention-based models for speech recognition. In: *Advances in Neural Information Processing Systems*, pp. 577–585 (2015)
10. Ding, H., et al.: A compact CNN-DBLSTM based character model for offline handwriting recognition with Tucker decomposition. In: *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 01, pp. 507–512, November 2017. <https://doi.org/10.1109/ICDAR.2017.89>
11. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587 (2014)
12. Gou, C., Wang, K., Yao, Y., Li, Z.: Vehicle license plate recognition based on extremal regions and restricted Boltzmann machines. *IEEE Trans. Intell. Transp. Syst.* **17**(4), 1096–1107 (2016)
13. He, S., Yang, C., Pan, J.S.: The research of chinese license plates recognition based on CNN and length\_feature. In: Fujita, H., Ali, M., Selamat, A., Sasaki, J., Kurematsu, M. (eds.) *Trends in Applied Knowledge-Based Systems and Data Science*, pp. 389–397. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-42007-3\\_33](https://doi.org/10.1007/978-3-319-42007-3_33)
14. Hui Li, C.S.: Reading car license plates using deep convolutional neural networks and LSTMS (2016)
15. Hurtik, P., Vajgl, M.: Automatic license plate recognition in difficult conditions - technical report. In: *Fuzzy Systems Association and International Conference on Soft Computing and Intelligent Systems* (2017)
16. Laroca, R., et al.: A robust real-time automatic license plate recognition based on the YOLO detector. In: *International Joint Conference on Neural Networks* (2018)
17. Lee, C.Y., Osindero, S.: Recursive recurrent nets with attention modeling for OCR in the wild. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2231–2239 (2016)
18. Li, H., Wang, P., Shen, C.: Towards end-to-end car license plates detection and recognition with deep neural networks (2017)
19. Li, H., Wang, P., You, M., Shen, C.: Reading car license plates using deep neural networks. *Image Vis. Comput.* **72**, 14–23 (2018). <https://doi.org/10.1016/j.imavis.2018.02.002>, <http://www.sciencedirect.com/science/article/pii/S0262885618300155>
20. Liu, W., et al.: SSD: single shot multibox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) *ECCV 2016. LNCS*, vol. 9905, pp. 21–37. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2)
21. Liu, Y., Huang, H.: Car plate character recognition using a convolutional neural network with shared hidden layers. In: *2015 Chinese Automation Congress (CAC)*, pp. 638–643, November 2015. <https://doi.org/10.1109/CAC.2015.7382577>
22. Luong, M.T., Pham, H., Manning, C.D.: Effective approaches to attention-based neural machine translation. arXiv preprint [arXiv:1508.04025](https://arxiv.org/abs/1508.04025) (2015)
23. Masood, S.Z., Shu, G., Dehghan, A., Ortiz, E.G.: License plate detection and recognition using deeply learned convolutional neural networks (2017)
24. Mnih, V., Heess, N., Graves, A., et al.: Recurrent models of visual attention. In: *Advances in Neural Information Processing Systems*, pp. 2204–2212 (2014)
25. Mubarak, H., Ibrahim, A.O., Elwasila, A., Bushra, S., Ahmed, A.: A framework for automatic license number plate recognition in sudanese vehicles (2017)

26. Pant, A.K., Gyawali, P.K., Acharya, S.: Automatic nepali number plate recognition with support vector machines. In: International Conference on Software, Knowledge, Information Management and Applications (2015)
27. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)
28. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, pp. 91–99 (2015)
29. Shi, B., Bai, X., Yao, C.: An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(11), 2298–2304 (2017). <https://doi.org/10.1109/TPAMI.2016.2646371>
30. Tian, Z., Huang, W., He, T., He, P., Qiao, Y.: Detecting text in natural image with connectionist text proposal network. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) *ECCV 2016*. LNCS, vol. 9912, pp. 56–72. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46484-8\\_4](https://doi.org/10.1007/978-3-319-46484-8_4)
31. Wang, J., Bacic, B., Yan, W.Q.: An effective method for plate number recognition. *Multimed. Tools Appl.* **77**(2), 1679–1692 (2018). <https://doi.org/10.1007/s11042-017-4356-z>
32. Wel, E.V.D., Ullrich, K.: Optical music recognition with convolutional sequence-to-sequence models (2017)
33. Zang, D.: Vehicle license plate recognition using visual attention model and deep learning. *J. Electron. Imaging* **24**(3), 033001 (2015)