INDUSTRIAL TECHNOLOGY ADVANCES

A deep learning-based method for vehicle license plate recognition in natural scene

JIANZONG WANG,[1] XINHUI LIU,[2] AOZHI LIU[2] AND JING XIAO[3]

Vehicle license plate recognition in natural scene is an important research topic in computer vision. The license plate recognition approach in the specific scene has become a relatively mature technology . However, license plate recognition in the natural scene is still a challenge since the image parameters are highly affected by the complicated environment. For the purpose of improving the performance of license plate recognition in natural scene, we proposed a solution to recognize real-world Chinese license plate photographs using the DCNN-RNN model. With the implementation of DCNN, the license plate is located and the features of the license plate are extracted after the correction process. Finally, an RNN model is performed to decode the deep features to characters without character segmentation. Our state-of-the-art system results in the accuracy and recall of 92.32 and 91.89% on the car accident scene dataset collected in the natural scene, and 92.88 and 92.09% on Caltech Cars 1999 dataset.

I. INTRODUCTION

With the development of deep learning, image data become more and more accessible to researchers and engineers. Many companies of different industrial areas such as finance technology company or automobile manufacture may accumulate a large amount of license plate images. These license plate images are probably obtained in different scenes, which has different influences on the image recognition task. For this reason, automatic identification of license plate is a significant research topic since the environments to identify a license plate may be unknown. Our research to recognize vehicle license plate (VLP) in natural scene is to identify the characters on images regardless of the environment where the image is obtained.

Traditional license plate recognition systems are mostly based on a known environment. For example, a set of images may be derived from the automobile on the highway. While an image is obtained from a car in the parking lot, the traditional system may have relatively poor performance on the image recognition task. Concretely, fixed cameras, fixed shooting angles, similar illumination, and even the same moving directions should be known with the traditional systems.

In reality, license plate images taken in natural scenes have different shooting angles and illumination parameters. As a result, recognition license plate in natural scene has two main challenges. On one hand, auto-detect license plate from giving images is necessary; on the other, it is hard to segment characters accurately.

In the license plate detection module, we first use SSD model [1] to detect license plate. Then we perform image processing techniques to correct license plate blocks detected in the previous step. Since character segmentation is a challenge, we propose a framework for the recognition process that could avoid license plate character segmentation.

The main contributions of this paper are summarized as three aspects. (A) We deploy a complete system to recognize VLP in natural scene; (B) an automatic encoder–decoder model (DCNN-RNN) is presented to recognize VLP; (C) we apply a gate mechanism to the encoder.

The rest of the paper is organized as follow: in Section II, we describe the present model in detail. We summarize the experimental setup and report experimental results in Section III, and related work and conclusions are in Sections IV and V, respectively.

[1]Federated Learning Technology Department of Ping An Technology Co., Ltd., Shenzhen, China
[2]Ping An Technology Co., Ltd., Shenzhen, China
[3]Ping An Group , AI Center of Ping An Technology Co., Ltd., Shenzhen, China

Corresponding author:
Aozhi Liu,
Email: liuaozhi1989@gmail.com

II. RELATED WORK

A) License plate detection

License plate detection inputs a real-life image and outputs potential license plate bounding boxes. Although there

are lots of existing algorithms to detect license plate under certain natural environment, it is still a challenging problem to detect the license plate well under arbitrary natural environment. There are four types of algorithms for this problem [2, 3]: edge-based, color-based, texture-based, and character-based.

Edge-based approaches try to find large edge-density regions as the license plates. In [4], an improved Prewitt arithmetic operator is a method used to locate license plate position. Rasheed *et al.* [5] applied canny operator to detect regions and used Hough transform to find the best vertical and horizontal boundaries of the license plate. Edge-based approaches benefit with fast speed, while they are sensitive to useless background edges. Color-based algorithms are based on the fact that the color of VLPs is different from others in most cases. HSI color model in [6] is used to detect the candidate VLP region. Then the algorithm uses rectangularity, aspect ratio, and edge density to distinguish license plates. The disadvantage of the color-based method is that they can not drop out other objects with almost the same color.

Texture-based models try to detect by unconventional pixel density distribution in plate regions. In [7], sliding concentric window algorithm is applied to identify license plate by its local irregularity property.

Character-based algorithms detect license plate via examining the presence of characters in the image. Lin [8] published an algorithm using image saliency to locate license plate. These approaches segment out characters with a high recall to the image saliency map. After that, it applies a rolling window on the selected characters to compute some detailed properties in image saliency and finally detect the license plate. Recent work in license plate recognition applies deep neural network to the recognition task [9]. This approach implemented the character segmentation method, which is different to the pipeline of our solution.

B) Optional character recognition

Character sequence recognition is directly related to VLP recognition. There are two methods which are well established by other researchers, the input segmentation (INSEG) method and the output segmentation (OUTSEG) method. The INSEG algorithm aims at determining the possible segmentation points of the characters in a certain sequence. Only when points containing a possible character are chosen to go through the recognizer, it generates a sequence that contains all the most likely characters, which are selected by another algorithm like Viterbi [10]. Some approaches to determine the points for a possible character are connected component analysis [10]. In the OUTSEG approach, unlike the INSEG, many characters are generated by sweeping a rolling window over the entire image with small steps. At each step, the contents of the windows will be treated as a tentative character. Then, all the tentative characters come through the recognition machine. The final sequence is selected if the corresponding actual characters

overlap significantly over the image [11]. INSEG and OUTSEG methods have problems: the recognizer is not able to gather enough useful information, when the partition of the image is inputted into the recognition machine.

III. APPROACH

A) System description

Our VLP recognize framework is shown in Fig. 1. It contains three steps: (1) VLP detection, (2) license plate adjustment, and (3) license plate recognition. We first apply SSD model or VLP detection [12]. Then, an image processing algorithm is applied for license plate adjustment. For license plate recognition, we propose DCNN-RNN neural network to improve the accuracy of vehicle character recognition.

1) VEHICLE LICENSE PLATE DETECTION

VLP detection is to detect the rectangle of the license plate in the image obtained in natural scene. We apply SSD model to detect the structure which is based on VGG16 [12] network.

We normalize the scale of the image in the input layer. In the training phase, according to the default box and the ground truth overlap rate, we find the corresponding positive sample. We also enlarge the training dataset by cropping, mirroring, and adding noise. The VGG-16 convolution layer is adopted to sample the multi-scale feature on the image. According to the granularity of the feature map extracted from different layers, we select the target bounding box with various scales. The prediction results for each target bounding box contain both the category and the confidence, and then the final prediction result is selected on the basis of the maximum inhibition criterion.

The superiority of this license plate detection model is that it could avoid multiple extraction of the characters of the overlapping images in the candidate frame, like Faster-RCNN [13]. This model just extracts feature once, so it achieves efficient license plate detection. The training set has 9597 images, containing five color classes: blue plate, yellow plate, white plate, black plate, and green plate.

2) VEHICLE LICENSE PLATE ADJUSTMENT

The aim of VLP adjusting is to correct the incline license plate. We adopt traditional image processing algorithms to rectify the license plate. The pipeline of the algorithm is shown in Fig. 3(a). First, we take a series of operations to find the circumscribed rectangle of the license plate. Then, the rectangle must be corrected if the angle of the license plate is too large. After that, all the license plates are adjusted to the front visual angle. Finally, we normalize the size of the license plate, which is acted as the input of the following recognition part. Figure 2 shows our correct results.

3) VEHICLE LICENSE PLATE RECOGNITION

VLP recognition is the most vital building block of our system. We choose the deep learning models to recognize the
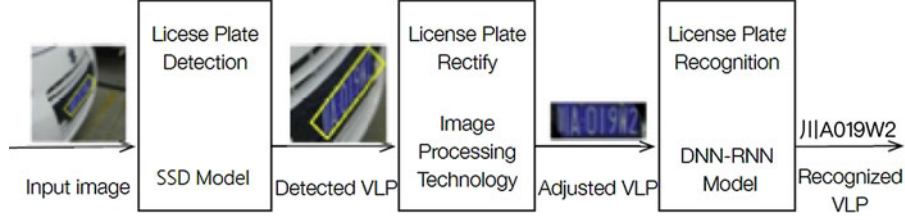
Fig. 1. The framework of the proposed vehicle license plate (VLP) detection and recognition system which contains three essential parts: detection, adjusting, recognition.

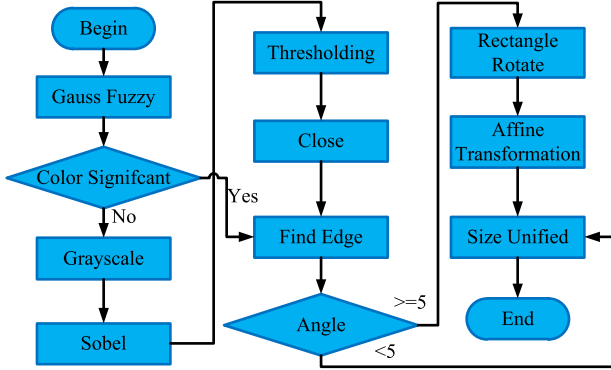**Fig. 2.** The correct results after license plate adjusting.

Fig. 3. The process of vehicle license plate adjusting adopts several image processing algorithms.

characters on the license plate, which consists of a deep convolution network for feature extracting and a stacked RNN for character recognition as shown in Fig. 4. First, DCNN layer is selected for extracting the features from the image that have been sliced. Then, the obtained feature sequence is input into the RNN Layers for recognizing characters in the row from left to right in order, in which multiple features identify one character. Finally, we process the result by the Translation Layer and define the recognition result in accordance with the char dictionary of the license plate. The specific theory will be described in detail in the next section.

B) DCNN-RNN

It is natural to use convolutional operation to scan the images which contain a row of characters and use the recurrent neural network to predict the characters. Following, we build a network that includes deep convolutional neural networks and recurrent neural networks, and we name it DCNN-RNN.

1) ENCODER
Giving an image $X \in R^{w \times h}$ with a standard size, the network automatically extracts a sequence of feature vectors $H = \langle H_1, \ldots H_T \rangle$:

$$H_1 = DCNN(X), \tag{1}$$

where T is the sequence length, and we set it 16 in our system.

$$H_2 = MLP(X), \tag{2}$$

where MLP is a multilayer perceptron layer.

$$H = g \odot H_1 + (1 - g) \odot H_2, \tag{3}$$

where g is a gating mechanism, which transports the raw information from input X to feature vectors H. $g = \sigma(W_T X + b)$, W_T is the square matrices.

By adding a gating mechanism to the neural network, it is possible to selectively preserve and output information through the gate, which solves the problem of information disappearing in the process of transmission. The convolution neural network with gating mechanism through the global shared weight connection enhanced the relationship between the front and back, so that we can extract more abstract features in larger context within the context.

2) DECODER
In the decoder layer, we use a recurrent neural network to predict vehicle characters. Long short-term memory (LSTM) [14] is used as basic cell:

$$S_t = LSTM(H_{t-1}, S_{t-1}) \tag{4}$$

and the probability distribution over the label space is estimated by:

$$\widehat{y}_t = softmax(WS_t), \tag{5}$$

where $W \in R^{o \times c}$, o is LSTM cell size and c is the number of vehicle characters.

We choose the argmax of the k-th output being the k-th character of the VLP and the CTC [15] loss is calculated as
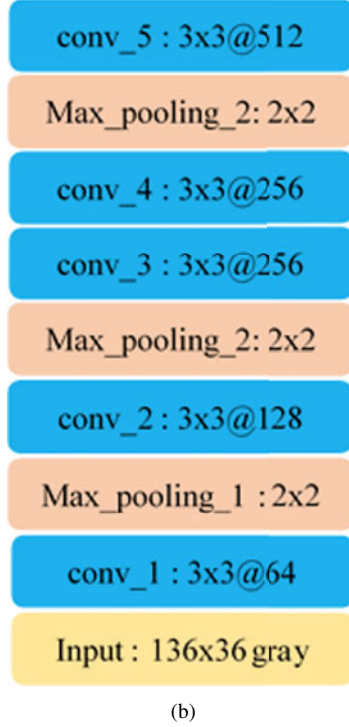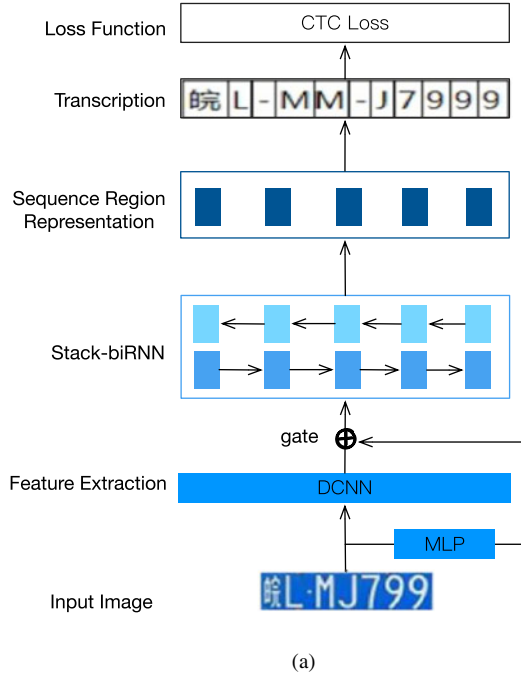
(a)

(b)

Fig. 4. (a) Overall structure of our sequence labeling-based model for Chinese plate recognition chooses DCNN to encode input image and stack RNN to decode deep features to characters. (b) The details of DCNN part.

the objection:

$$loss = \frac{1}{m} \sum_{i}^{m} CTC(DCNN\text{-}RNN(X_i), y_i), \qquad (6)$$

where m is the number of samples and y_i is the ground truth vehicle character sequence.

Table 1. Statistics of our dataset PingAnCarPL(a). Image is the number of VLP images, Class means the class of license plate.

PingAnCarPL(a)	Image	Class
Train	81 000	5
Test	19 000	5

IV. EXPERIMENTS

To evaluate the effectiveness of our proposed model, we conduct experiments on two datasets: Caltech Cars 1999 which is a standard dataset and PingAnCarPL which is collected by our own. The more details about dataset and training settings can be found in next parts.

A) Datasets

A sufficient training dataset is essential to the success of deep neural networks. To train our model, we collected a large VLP dataset which primarily contains Chinese and USA license plate. All stages of training apply to our own dataset, while the test carries on both datasets.

- **Caltech Cars 1999** is a standard dataset which consists 126 images of cars rear with 896×896 jpg format. The images are taken in the Caltech parking lots, which contain a US license plate with the cluttered background, such as trees, grass, wall, etc.
- **PingAnCarPL** is a dataset collected by our own from natural scenes. It contains 100 000 images of cars (with 700×700 jpg format) including 31 words of provinces in China and two special kinds of VLP. The data are processed into two parts for the different stages, which are used to train the license plate detection model and the license plate character recognition model, respectively.

B) Training

For the license plate detection model, we implement the network based on SSD using Caffe framework. For the license plate recognition model, we perform DCNN-RNN under Torch framework. The training of the two models is independent of each other.

In training the license plate detection phases, we use the original images in PingAnCarPL(a) dataset, shown in Table 1, whose size is normalized to 512×512. We extract seven candidate rectangular boxes in each convolution layer, and the aspect ratio of these candidate boxes is {1, 2, 3, 4, 1/2, 1/3, 1/4}. There are five classes: blue plate (general plate), yellow plate (coach car plate), white plate (police car plate), black plate (car plate with foreign elements), green plate (clean energy license plate). The color information of license plate will help to adjust the plate blocks that are detected by license plate detection model. In training the license plate recognition phases, we use the plate blocks that have been corrected by image processing before, shown in Table 2, whose size is normalized to 136 * 36. Batch size is

Table 2. Statistics of our dataset PingAnCarPL(b). Image is the number of VLP blocks corrected before. Kind means the class of VLP which from 31 Chinese provinces and two special kinds. Chars is the kind of characters in the license plate.

PingAnCarPL(b)	Image	Kind	Chars
Train	40000	33	67
Validation	10000	33	67
Test	10000	33	67

Table 3. Our license plate detection model result on PingAnCarPL(a).

Classes	mAP (%)
Blue plate	99.35
Yellow plate	99.10
White plate	98.87
Black plate	99.84
Green plate	89.57
Average	97.24

tuned over 128, 256. We tune on two optimization schemes: stochastic gradient descent and Adam [16] with start learning rate lr over {0.1, 0.05}. We employ a decreasing learning rate $lr = lr/(10 \times epoch)$ for all layers. Our experiments are carried out on NVIDIA M60 GPU.

C) Results

The result of the license plate detection model is shown in Table 3. The average mAP rate achieves 97.24%. The mAP of the green plate is lower to other classes, because of a lot less training data. In the field of object detection, mAP represents mean average precision. Mean average precision for a set of queries is the mean of the average precision scores

Table 4. Comparison of plate detection results by different methods on Caltech Cars dataset.

Method	Precision (%)	Recall (%)
Le& Li [17]	71.40	61.60
Bai & Liu [3]	74.10	68.70
Lim & Tay [18]	77.73	75.47
Zhou [19]	82.50	79.80
Shi [20]	88.02	87.87
Ours	**92.88**	**92.09**

Table 5. Comparison of plate detection results by different methods on PingAnCarPL evaluate dataset.

Method	Precision (%)	Recall (%)
Jaderberg et al. [21]	88.2	84.81
Shi [20]	89.80	87.00
Ours	**92.32**	**91.89**

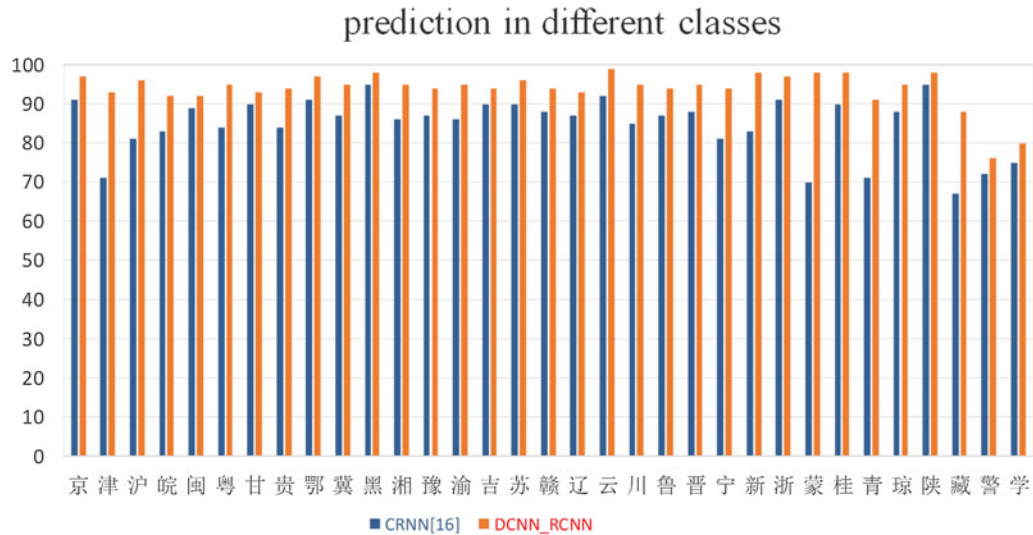**Fig. 5.** Samples from the PingAnCarPL.

Fig. 6. Our license plate recognition model's precision in different class of PingAnCarPL(b).

for each query. The results of our license plate recognition model(DCNN-RNN) are visualized in Fig. 5.

For the whole model, we compare with other license plate recognition algorithms on two datasets. The results on Caltech Cars 1999 and the results on PingAnCarPL are shown in Tables 4 and 5. There is an incredible improvement of performance with our model on Caltech Cars 1999 and PingAnCarPL dataset. As shown in Tables 4 and 5, our method outperforms the two state-of-the-art models by a large margin.

V. CONCLUSION

In this paper, a unified framework is presented for VLP recognition. The DCNN-RNN model is used to encode input images and stacked RNN model is applied to decode deep features to characters. Our method uses DCNN to locate the license plate, after correction using DCNN with gating mechanism it extracts the characteristics of the license plate, which could extract more abstract features in larger context within the context. Finally, leverage RNN to decode the deep features to characters, without character segmentation. Experimental results on PingAnCarPL and Caltech Cars 1999 dataset demonstrate that our proposed framework achieves state-of-the-art result compared to the traditional machine learning methods and former CNN models as shown in Fig. 6.

REFERENCES

[1] Liu, W *et al.*: Single shot multibox detector, in *European Conf. on Computer Vision*. 2016, 21—37.

[2] Du, S. *et al.*: Automatic license plate recognition (ALPR): a state-of-the-art review. *IEEE Trans. Circuits Syst. Video Technol.*, **23** (2) (2013), 311–325.

[3] Hongliang, B; Changping, L: A hybrid license plate extraction method based on edge statistics; morphology, in *Pattern Recognition, 2004. ICPR 2004. Proc. of the 17th Int. Conf. on*. vol. 2, 831–834. IEEE 2004.

[4] Chen, R; Luo, Y: An improved license plate location method based on edge detection. *Phys. Procedia.*, **24** (Part B) (2012), 1350–1356.

[5] Rasheed, S; Naeem, A; Ishaq, O: Automated Number Plate Recognition Using Hough Lines; Template Matching, in *Proc. of the World Congress on Engineering; Computer Science*. 2012, **1**, 24–26.

[6] Deb, K; Jo, K.H.: HSI color based vehicle license plate detection, in *Int. Conf. on Control, Automation; Systems*. IEEE, 2008, 687–691.

[7] Giannoukos, I. *et al.*: Operator context scanning to support high segmentation rates for real time license plate recognition. *Pattern. Recognit.*, **43** (11) (2010), 3866–3878.

[8] Lin, K.H.; Tang, H; Huang, T.S.: Robust license plate detection using image saliency, in *Image Processing (ICIP), 2010 17th IEEE Int. Conf. on*. 2010, 3945—3948.

[9] Li, H; Shen, C: Reading car license plates using deep convolutional neural networks and LSTMs." arXiv preprint arXiv:1601.05610, 2016.

[10] Weissman, H. *et al.*: Recognition-based segmentation of on-line run-on handprinted words: input vs. output segmentation. *Pattern. Recognit.*, **27** (3) (1994), 405–420.

[11] Zhang, Y; Zhang, C.: A new algorithm for character segmentation of license plate, in *Intelligent Vehicles Symp., 2003. Proc. IEEE*. 2003, 106–109.

[12] Simonyan, K; Zisserman, A: Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint arXiv*, 1409.1556, 2014.

[13] Girshick, R: Fast R-CNN, in *Proc. of the IEEE Int. Conf. on Computer Vision*, 2015, 1440–1448.

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

[15] Graves, A; Gomez, F: Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks, in *Int. Conf. on Machine Learning*. ACM, 2006, 369–376.

[16] Kingma, D. P.; Ba, J. L.; Adam, B.J.: A method for stochastic optimization. *Comput. Sci.* arXiv preprint arXiv:1412.6980 (2014).

[17] Le, W; Li, S: A hybrid license plate extraction method for complex scenes, in *Pattern Recognition, 2006. ICPR 2006. 18th Int. Conf. on.*, **2**, 2006, 324–327.

[18] Hao, W.L.; Tay, Y.H.: Detection of license plate characters in natural scene with MSER and SIFT unigram classifier, in *Sustainable*

Utilization and Development in Engineering and Technology. IEEE, 2010, 95–98.

[19] Zhou, W. *et al.*: Principal visual word discovery for automatic license plate detection. *IEEE Trans. Image. Process.*, **21** (9) (2012), 4269–4279.

[20] 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. Machine Intell.*, **39** (11) (2017), 2298.

[21] Jaderberg, M. *et al.*: Reading text in the wild with convolutional neural networks. *Int. J. Comput. Vis.*, **116** (1) (2016), 1–20.

Dr. Jianzong Wang is the Deputy Chief Engineer and Senior AI Director at PingAn Technology. He's responsible for worldwide PingAn Technology innovations with respect to Deep Learning, Cloud Computing, Big Data and committed use AI in Financial and Insurance Service like Smart Auto Damage Detection, Smart Outbound Robots, Virtual Reality, Voiceprint Recognition, Animal Face Identification, AI Music Composition, AI Translation, Text to Speech, Automatic Speech Translation and Crowd Sourcing. Dr. Wang is the true innovator who leads three core platforms like Deep Learning Platform, Heterogeneous Computing Platform and Smart Voice Platform in the largest financial service company within China and he is the first man who adopts Nvidia RAPIDS in China. Dr. Wang has been actively involved in numerous social activities. He is a representative of Shenzhen Futian District People's Congress. He is a senior member of the Big Data Expert Committee in China Computer Federation and vice chairman in CCF YOC-SEF Shenzhen. He is also a producer of many conferences both in international and domestic.

Ms. Xinhui Liu is a senior algorithm engineer in the Intelligent Vision team of PingAn Technology, focusing on the research of computer vision and medical imaging assisted diagnosis. Graduated from Harbin Institute of Technology, majoring in computer science and technology.

Mr. Aozhi Liu is the leader of the AI Music group of PingAn Technology Deep Learning Team. He's responsible for the research of the fields of AI composition as well as Singing Audio Synthesis. He is an interdisciplinary researcher of computer science and music composition. As a senior data scientist, he is the algorithm designer of the Applet Composer for Everyone. As an artist, Mr. Aozhi Liu was a member of the Nanjing University Orchestra, who was invited to performance in different countries such as the United States, the United Kingdom, Russia, Canada, Germany and South Korea.

Dr. Xiao Jing, China "Thousand Talents Program" Distinguished Expert, is the Chief Scientist of Ping An Insurance (Group) Company of China, LTD. He received his PhD degree from Carnegie Mellon University, and has published over 50 academic papers and 90 granted US patents. Before joining Ping An, he worked as Principal Applied Scientist Lead in Microsoft Corp. and Manager of Algorithm Group in Epson Research and Development, Inc. Dr. Xiao started R&D in artificial intelligence and related fields since 1995, covering a broad range of application areas. He is now leading research and development in AI-related technologies and their applications on finance and healthcare in Ping An Group.