JOINT LICENSE PLATE SUPER-RESOLUTION AND RECOGNITION IN ONE MULTI-TASK GAN FRAMEWORK

Minghui Zhang, Wu Liu, Huadong Ma

Beijing Key Lab of Intelligent Telecommunication Software and Multimedia,
Beijing University of Posts and Telecommunications, Beijing 100876, China

ABSTRACT

License plate recognition (LPR) plays an important role in intelligent transport systems. The existed LPR systems are mostly based on hand-crafted methods for detection, segmentation, and recognition, which cannot accurately recognize the license plate in unconstrained surveillance environments. In this paper, we propose a Multi-Task Generative Adversarial Network (MTGAN) based LPR system, which combines the license plate super-resolution and recognition in one end-to-end framework. In the proposed MTGAN, we design a Fully Connected Network (FCN) as generative network (GN), which can combine knowledge from data distribution and domain prior knowledge of license plate to generate the spatial corresponding and high-resolution plate images in the synthesis pipeline. More important, a multi-task discriminative network is designed in MTGAN to combine the super-resolution and recognition in an adversarial manner to enhance each other. The experiments on the built real-world license plate dataset show that the proposed LPR system can generate high-resolution license plates as well as recognize them with higher accuracy than state-of-the-art LPR systems.

Index Terms— Generative Adversarial Network, Super-Resolution, License Plate Recognition, Multi-Task

1. INTRODUCTION

Automatic license plate recognition (LPR) system plays an important role in daily life, such as road traffic monitoring, ETC, and parking management system. The main task of LPR system is to recognize the numbers and characters of the license plate images captured by surveillance cameras. As mentioned in [1], traditional LPR systems mainly follow three steps: 1) detect license plates in the images based on hand-crafted features, such as the color and shape of the license plates, or the local texture feature of the characters [2]; 2) segment the whole plate license into individual blocks of characters; 3) recognize the segmented characters with hand-crafted features and classifiers like SVM [3], artificial neural networks [4] or Convolutional Neural Network [5].

However, the existing LPR systems mainly focus on constrained scenarios, such as the entrances of parks or the toll gates of highways. With the auxiliary equipments (e.g., high definition camera, flashlights), and strictly limited scenarios, the captured license plate images are usually well-aligned with high resolution and bright illumination. However, in unconstrained urban surveillance conditions, the cameras are deployed in unrestricted areas and locations with uncertain illumination [6] and viewpoints. Therefore, due to the complex surveillance background, lighting, camera viewpoints, resolution, motion blur, and occlusion, the captured license plate images are blurred, distorted, and in low-resolution[7, 8], which makes LPR a very challenging task. Therefore, before recognition, many existed works try to exploit the super-resolution technology to recover the high-resolution plate images [9, 10], which can improve the recognition accuracy.

To automatically recognize the license plate in unconstrained surveillance conditions, we propose a multi-task [11] generative adversarial network (MTGAN) for joint license plate super-resolution and recognition in one end-to-end framework, which is named as MTGAN. Recently, GAN [12] achieves excellent performance on super-resolution [13] and style transfer [14]. The original GAN contains two subnets: Generator Network (GN) and Discriminator Network (DN). The GN tries to generate high-resolution images according to the data distribution of original images, while the DN tries to discriminate the input images belong to the original images or belong to the generated images. GN and DN are trained in an adversarial manner to enhance each other. Inspired by GAN, the proposed MTGAN is also consist of a GN and a DN. Differently, as the goal of the super-resolution is to improve the recognition rate, we change the function of DN to judge whether the license plate is high-definition enough to be correctly recognized. Thus, the DN is redesigned as a multi-task convolutional neural network, which has two task in the training phase: judgment as original GAN and character classification for plate recognition.

Moreover, in the MTGAN, GN is a Fully Connected Network (FCN) to learn a generative model, which can generate
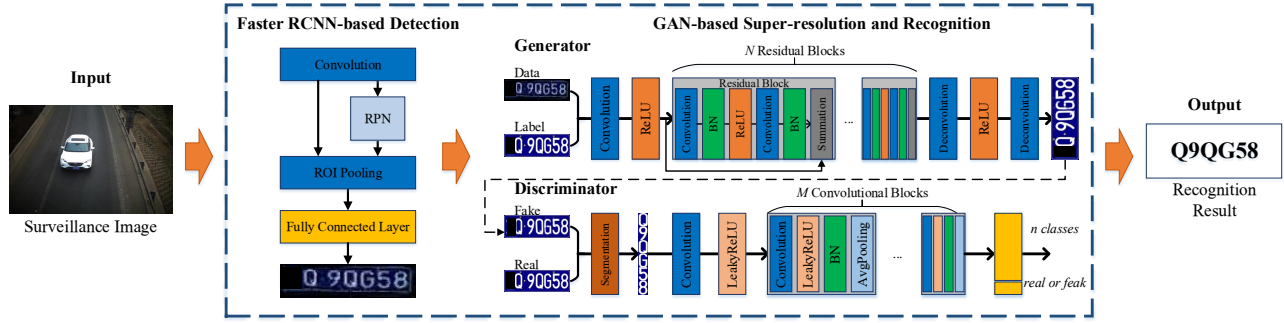
Fig. 1. The overview of the proposed joint license plate super-resolution and recognition system based on MTGAN.

high-resolution license plates from low-resolution ones. In particular, we generate a large-scale dataset, which contains many low-resolution license plates (i.e., inputs) and standard license plates (i.e., labels) to train GN. The standard plates are make according to the strict license plate manufactory standard. In this way, the GN tries to generate standard license plates with high-resolution, unified viewpoint, consistent color, and same illumination. Furthermore, we design a Direct Segmentation Layer (DSL) in DN, which directly acquires the pixels of one character according to the production standard of license plates. Then the segmented characters in a license plates are fed into the convolution layer of DN. The DN produces the probability of each class as well as the probability whether the input is from standard license plates or from generated ones by GN. Adversarial training of GN and DN can make super-resolution and recognition models enhance each other. In the experiment section, we describe the dataset proposed in this paper and show the effectiveness of the proposed LPR system based on the dataset.

In summary, this paper makes the following contributions:

- We propose an end-to-end deep learning-based license plate super-resolution and recognition system for unconstrained urban surveillance scenes, which is different from the traditional process of LPR.

- We propose a multi-task discriminative network in MT-GAN to combine the super-resolution and recognition in an adversarial manner to enhance each other.

- We propose a generative network in MT-GAN to combine prior knowledge from data distribution and domain knowledge of license plate to generate the spatial corresponding and high-resolution plate images.

2. PROPOSED APPROACH

The architecture of the proposed LPR system is shown in Fig.1. The input of the system is images captured by traffic surveillance cameras. They vary in viewpoints, resolution, light condition, background and so on. First, we utilize Faster RCNN to detect the license plates in the images. The output

of the Faster RCNN is the bounding boxes and probabilities of the detected license plates. Then, the detected license plates are fed into GAN. The GN can generate an unified and high-resolution license plate accord with the manufacture standard. The DN segments the generated standard license plates into character blocks and recognizes them. At last, the output of the end-to-end LPR framework is the identification result.

2.1. Faster RCNN based Detection

The first step of the LPR system is to detect the license plates in the images. Recently, there are many methods for the object detection, such as Fast RCNN [15], Faster RCNN [16], YOLO [17], etc. Because the Faster RCNN is an end-to-end detection neural network and have higher speed and accuracy than the state-of-the-art methods, we choose to utilize it to detect the license plates. Faster RCNN introduced a Region Proposal Network (RPN) which predicts the bounding boxes and probabilities of the detection object at each position. In this paper, we modify the Faster RCNN to detect the license plates in images captured in urban surveillance.

2.2. GN for Super-Resolution

Considering that the license plates detected by Faster RCNN may have low resolution and different viewpoints, it is difficult to segment and recognize the characters directly. Fortunately, GAN can learn the data distribution through the adversarial training of GN and DN. As the method proposed in [9], we adopt GAN to generate standard license plates for each detected license plate. As shown in Fig.1, we adopt a hierarchy of K residual blocks proposed in [18] as the backbone of GN. For each residual block, it contains two convolutional layers with 64 filters of 3×3, two batch-normalization layers, and one ReLU activation function in a specific order. A shortcut connects the input and output of the block by pixel-wise summation as the identity mapping to guarantee the gradient flow through the deep network. After the K residual blocks, a deconvolutional layer with stride of two is adopted to enlarge the resolution of the feature map. At the tail of the network, we use a deconvolutional layer with three filters of 1×1 to

generate the three-channel high-resolution images.

In the training process of the GN, the detected images by Faster RCNN are as the inputs of GN. As the [9] proposed, because the license plates have strict manufacture standards, it is reasonable and easier to synthesize the groundtruth plate for each original low-resolution one with these standards. With the rendered groundtruth, the GN can learn the data distribution of the groundtruth, and try to generate high-resolution and standard image plates.

2.3. DN for Segmentation and Recognition

As the generated license plates by GN are in accord with the manufacture standards, we utilize this property to segment the characters directly according to the pixel position. Concretely, for one license plate, the image size is $W \times H$ and has N characters, where W is the width and H is the height of the image. Furthermore, according to the manufacture standard, the i_{th} character occupies a_i percent of the width W. Then, we segment the characters as follows:

$$W_i = W * a_i, \qquad (1)$$

$$H_i = H. \qquad (2)$$

Moreover, the start of each character block is from the end of the frontier character block. Accordingly, the Direct Segmentation Layer (DSL) segments the each license plate into 6 character blocks. Then the backbone of the discriminator employs M convolution layers and a fully connected (FC) layer ($M = 8$ in our implementation). The LeakyReLU activation function and batch-normalization are employed as in [13]. At last, a sigmoid function is connected to the tail of the network as a classifier.

The DN is trained to classify the input image into 35 categories: 10 digits, 24 English characters (except the "I" and "O"), and fake label (1 for fake and 0 for real). The characters segmented from generated images are fed into the DN as the fake images. The characters segmented from the groundtruth plates following the above method are fed into the DN as the real images. The last bits for the labels represent whether the images fed into the DN are from GN or groundtruth.

2.4. Adversarial Loss Function

As mentioned above, the aim of GAN is to train a generator network GN which outputs high-resolution images (I_h) ensemble with the groundtruth (I_g). Therefore, it is critical to define a loss function for the GN to measure the differences between the I_h and I_g. In this paper, we utilize the Mean Squared Error (MSE), and the concrete formulation is:

$$L_{MSE} = \frac{1}{WH} \sum_{x=1}^{W} \sum_{y=1}^{H} (I_{g(x,y)} - G(I_l)_{(x,y)})^2, \qquad (3)$$

where W and H are width and height of I_g. I_l presents the input low-resolution images of GN. $G(\cdot)$ represents the output of GN. Furthermore, just as the existing GANs, an adversarial

loss from DN is returned to GN. In this paper, we return the loss of the fake images of DN, i.e., $-L_F$, to GN. It tells the GN whether DN can recognize the generated images belong to fake or real. Therefore, the total loss function of GN is :

$$L_G = L_{MSE} - L_F. \qquad (4)$$

For the DN, the classification loss function adopts the Binary Cross Entropy (BCE) between the label and the output of DN for both real images and fake images. The concrete formulation is:

$$L_D = L_{BCE} = L_R + L_F$$

$$= -\frac{1}{B * N} \sum_i (t_i * log(o_i) + (1 - t_i) * log(1 - o_i)), \qquad (5)$$

where L_F represents the loss of license plate images generated from GN, L_R represents the loss of groundtruth images from standard license plates, o_i is the output of the sigmoid function in the DN, t_i is the one-hot vector label of the license plate number, N is the character block number in one license plate, and B is the batch size of one iteration.

3. EXPERIMENTAL RESULTS

3.1. Experimental Setting

To evaluate the proposed method, we build a new dataset with 10,000 images, which are captured by real-world cameras in the unconstrained city surveillance environments. There is at least one vehicle in each image. We label the bounding boxes of license plates in each image. There are 12,170 license plates in total. For each license plates, we first record the license plate numbers, then synthesize a standard license plate accord with the manufacture standard.

In the evaluation, we use 10,000 license plates as train set and 2,170 ones as test set. When training the GN, the standard license plates are exploited as the labels. When training the DN, the real label is the license plate number which is converted to 34-dimension one-hot vector concatenated with zero in the end and the fake label is a 35-dimension vector with 34 zeros and a one in the end. To compared with the state-of-the-art super-resolution methods, we choose peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) as the evaluation criteria. Moreover, we use EasyPR [19], an open LPR project, to recognize the generated licnese plates by different methods to evaluate their super-resolution effect.

3.2. Evaluation on Super-resolution

In this section, we compare the following five super-resolution methods on the proposed dataset to demonstrate the effectiveness of the proposed method:

(1) **SCN [20].** SCN is a sparse coding model particularly designed for super-resolution.

(2) **SRCNN [21].** SRCNN is a deep learning method for single image super-resolution, which directly learns an end-to-end mapping between the low and high-resolution images.

Table 1. The best PSNR/SSIM comparison among different methods.

	SCN[20]	SRCNN[21]	SRGAN[13]	DPGAN[9]	**MTGAN**
PSNR	7.31	8.13	7.59	10.1	**22.1**
SSIM	0.105	0.124	0.100	0.439	**0.893**

Fig. 2. The examples recovered by different super-resolution approaches.

(3) **SRGAN [13].** SRGAN is a GAN-based method for image super-resolution. It proposes a perceptual loss function which consists of an adversarial loss and a content loss.

(4) **Domain Priori GAN (DPGAN) [9].** DPGAN designs a particular GAN based super-resolution framework for license plate recognition.

(5) **MTGAN.** The proposed method which combines super-resolution and recognition.

Table 1 shows the PSNR and SSIM comparisons for different methods. We can see that the results of DPGAN and MTGAN are greatly higher than the other methods. The reason is that DPGAN and MTGAN can both keep the domain priori knowledge of the license plate for image super-resolution. Although the DP-GAN can generate standard license plates with domain knowledge, the proposed MTGAN also achieves better performance than DPGAN. It is because that we return the BCE loss of fake characters, i.e., $-L_F$, to GN, which can assist the GN to generate the more clear and accurate characters than DPGAN. From the examples of the super-resolution results shown in Fig.2, we can also find the same conclusions. The charactors generated by the MT-GAN are more correspond with the manufacture standards of license plate.

3.3. Evaluation on Recognition

Besides the super-resolution, the main goal of the proposed method is to improve the license plate recognition performance. Therefore, we compare the proposed method with the state-of-the-art recognition methods. We use EasyPR as the comparative automatic license plate recognition method as it is a famous open source project for Chinese plate recognition in unconstrained situation [19]. The comparative results are shown in Table 2. The inputs of the methods shown in this table are license plates detected by Faster RCNN on testset. The accuracy of EasyPR for the detected origin images is the baseline. The next column shows the accuracy of super-resolution images by SRCNN. It achieves 7% higher accuracy than the baseline, which shows that the super-resolution can improve the recognition. Furthermore, the "GN + EasyPR"

Table 2. The recognition accuracy of different methods.

Method	Origin + EasyPR	SRCNN + EasyPR	GN + EasyPR	**MTGAN (Ours)**
Accuracy	36%	43%	54%	**75%**

Fig. 3. Some results of the proposed LPR system: (a) the input images captured by cameras; (b) the license plates detected by Faster RCNN; (c) the outputs of GN; (d) the recognition results of DN.

exploits the GN in MTGAN to recover the images and uses EasyPR to recognize the characters. It achieves 11% high accuracy, which demonstrate the super-resolution power of the proposed MTGAN. Finally, MTGAN achieves the best performance, which significantly improve the recognition performance. This result demonstrates that MTGAN which combines super-resolution and recognition in one end-to-end framework can simultaneously improve the two task through adversarial learning.

For the performance of DN, we also compare it with the classical method LeNet for character recognition. We modify the LeNet into a 34-category classification network and train it using the characters segmented by the groundtruth and the generated images. We test the DN of our method and the modified LeNet on the characters segmented by GN. The *single character recognition accuracy* of DN achieves 92.5%, while the accuracy of the modified LeNet is 76.5%. Owing to lack of adversarial training, the new trained LeNet cannot recognize characters generated by GN well. At last, Fig.3 shows the performance of the proposed LPR system at each step. There are some misidentified characters. The reason is that some character pairs are too similar such as "0" and "Q", "Z" and "2", "Z" and "7", "8" and "B", etc.

4. CONCLUSION

In this paper, we propose a deep learning-based and end-to-end LPR system for unconstrained urban surveillance scenes, which jointly sharpens the license plates and recognizes the characters in one multi-task based GAN framework. First, the Faster RCNN is exploited to detect the license plates. Then GN is responsible to generate high-resolution license plates. Next the multi-task DN, which combines the discrimination and recognition in one N-category classification network, is proposed to segment and recognize the characters. The experiments on the built real-world license plate dataset demonstrate that the proposed LPR system can achieve better super-resolution and recognition performance than the state-of-the-art LPR systems.

5. REFERENCES

[1] Shan Du, Mahmoud Ibrahim, Mohamed Shehata, and Wael Badawy, "Automatic license plate recognition (alpr): A state-of-the-art review," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 2, pp. 311–325, 2013.

[2] Huadong Ma and Wu Liu, "Progressive search paradigm for internet of things," *IEEE MultiMedia*, 2017.

[3] Kap Kee Kim, KI Kim, JB Kim, and Hang Joon Kim, "Learning-based approach for license plate recognition," in *IEEE Signal Processing Society Workshop on Neural Networks for Signal Processing*, 2000, vol. 2, pp. 614–623.

[4] Jianbin Jiao, Qixiang Ye, and Qingming Huang, "A configurable method for multi-style license plate recognition," *Pattern Recognition*, vol. 42, no. 3, pp. 358–369, 2009.

[5] Xinhao Liu, Takahito Kawanishi, Xiaomeng Wu, and Kunio Kashino, "Scene text recognition with high performance cnn classifier and efficient word inference," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2016, pp. 1322–1326.

[6] P R. Hill, Harish Bhaskar, Mohammed Al-Mualla, and D R. Bull, "Improved illumination invariant homomorphic filtering using the dual tree complex wavelet transform," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2016, pp. 1214–1218.

[7] Xinchen Liu, Wu Liu, Tao Mei, and Huadong Ma, "A deep learning-based approach to progressive vehicle re-identification for urban surveillance," in *European Conference on Computer Vision*. Springer, 2016, pp. 869–884.

[8] Xinchen Liu, Wu Liu, Tao Mei, and Huadong Ma, "Provid: Progressive and multi-modal vehicle re-identification for large-scale urban surveillance," *IEEE Transactions on Multimedia*, 2017.

[9] Wu Liu, Xinchen Liu, Huadong Ma, and Peng Cheng, "Beyond human-level license plate super-resolution with progressive vehicle search and domain priori gan," in *ACM Conference on Multimedia*, 2017, pp. 1618–1626.

[10] Kaggere V Suresh, G Mahesh Kumar, and AN Rajagopalan, "Superresolution of license plates in real traffic videos," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 2, pp. 321–331, 2007.

[11] Wu Liu, Tao Mei, Yongdong Zhang, Cherry Che, and Jiebo Luo, "Multi-task deep visual-semantic embedding for video thumbnail selection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3707–3715.

[12] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, 2014, pp. 2672–2680.

[13] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[14] Justin Johnson, Alexandre Alahi, and Li Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision*, 2016.

[15] Ross Girshick, "Fast r-cnn," in *IEEE International Conference on Computer Vision*, 2015, pp. 1440–1448.

[16] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.

[17] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi, "You only look once: Unified, real-time object detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779–788.

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *IEEE conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[19] "Easypr," *https://github.com/liuruoze/EasyPR*.

[20] Zhaowen Wang, Ding Liu, Jianchao Yang, Wei Han, and Thomas Huang, "Deep networks for image super-resolution with sparse prior," in *IEEE International Conference on Computer Vision*, 2015, pp. 370–378.

[21] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016.