Robust Chinese License Plate Generation via Foreground Text and Background Separation

Yi-Fan Sun, Qi Liu, Song-Lu Chen, Fang Zhou(✉), and Xu-Cheng Yin

University of Science and Technology Beijing, Beijing, China
{yifansun,qiliu7}@xs.ustb.edu.cn, zhoufang@ies.ustb.edu.cn,
xuchengyin@ustb.edu.cn

Abstract. To solve data scarcity, generating Chinese license plates with Generative Adversarial Network becomes an efficient solution. However, many previous methods are proposed to directly generate the whole license plate image, which causes the mutual interference of the foreground text and background. This way, it may cause unclear character strokes and an unreal sense of the overall image. To solve these problems, we propose a robust Chinese license plate generation method by separating the foreground text and background of the license plate to eliminate mutual interference. The proposed method can generate any Chinese license plate image while maintaining the precise character stroke and background of the real license plate. Specifically, we substitute the foreground text of the real license plate with the target text. To provide supervision data for text substitution, we propose to synthesize them via foreground text and background separation. Firstly, we erase the text of the real license plate to obtain the corresponding background image. Secondly, we extract the foreground text of another real license plate and merge it with the background obtained above. Qualitative and quantitative experiments verify that the license plates generated by our method are more homogeneous with the real license plates. Besides, we enhance license plate recognition performance with the generated license plates, which validates the effectiveness of our proposed method. Moreover, we release a generated dataset (https://github.com/ICIG2021-187) with 1,000 license plates for each province, including all 31 provinces of the Chinese mainland.

Keywords: License plate generation · Generative adversarial network · License plate recognition

1 Introduction

As the license plate is the unique identification of a car, license plate recognition (LPR) is widely used in many applications, such as traffic control, parking charge, and criminal investigation. However, recent LPR approaches need massive data for training, and it is labor-intensive and costly to collect and annotate

Y. Peng et al. (Eds.): ICIG 2021, LNCS 12890, pp. 290–302, 2021.
https://doi.org/10.1007/978-3-030-87361-5_24

license plate images. Hence, generating license plates with Generative Adversarial Network (GAN) becomes an efficient solution for data augmentation.

Many GAN-based license plate generation methods [14–16,18] are proposed based on pix2pix [8] and CycleGAN [21], and these methods generate the whole license plate image directly. As shown in Fig. 1 (b)–(e), the character strokes of the generated license plates are unclear, especially the Chinese characters. Moreover, the foreground text and background are less distinct than the real license plates.

Fig. 1. (a) Some examples of real license plates. Some examples generated by (b) CycleWGAN [15], (c) CycleWGAN-GP [16], (d) AsymCycleGAN [18], (e) Sun *et al.* [14], (f) our method.

These problems are mainly caused by the mutual interference of the foreground text and background, and the above GAN-based networks cannot generate both of them effectively. Specifically, background generation mainly focuses on texture features, such as background color, plate frame, space mark, rivets, and lighting factors. Meanwhile, the foreground text generation mainly focuses on the shape and stroke of the character. However, it is challenging for neural networks to generate them simultaneously. Hence, inspired by the modular approach [17], we propose to separate the foreground text and background of the license plate to eliminate mutual interference.

In this work, we propose a robust Chinese license plate generation method that can generate any Chinese license plate image while maintaining the precise character stroke and background of the real license plate. Specifically, we generate the license plate by substituting the foreground text of the real license plate with the target text, and the background texture is preserved appropriately. To provide supervision data for text substitution, we propose to synthesize them via foreground text and background separation. Firstly, we erase the text of the real license plate and fill this text region with appropriate background texture to obtain the corresponding background image. Secondly, we randomly select another license plate and extract its foreground text, then merge it with the background obtained above to obtain the merged license plate. The merged license plate and the real license plate with the same background are used as the input and output of the text substitution to supervise training. Compared with the existing license plate generation methods, our method can generate more clear and precise text stroke and a more real sense of the overall image. Qualitative and quantitative experiments prove that the license plates generated by our method are more homogeneous with the real license plates. Besides, we enhance license plate recognition performance with the generated license plates, especially for the Chinese characters, which validates the effectiveness of our proposed method. Moreover, we release a generated dataset with 1,000 license plates for each province, including all the 31 provinces of the Chinese mainland.

2 Related Work

2.1 Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) [3] generally consist of a generator and a discriminator, trained alternatively by an adversarial loss function. DCGAN [12] is the first work to introduce deep convolutional neural networks into GANs. Wasserstein GAN (WGAN) [1] presents Wasserstein distance loss to solve the vanishing gradient and mode collapse problems. WGAN-GP [4] adds gradient penalty to stabilize the training process. In terms of the image-to-image translation task, pix2pix [8] applies a conditional-GAN [11] to learn a mapping relationship from the input domain to the output domain but requires paired data. CycleGAN [21] designs a cycle consistency loss with unpaired training data to achieve cross-domain style transfer. However, these methods perform deficiently on text image generation due to the large variety of text shapes.

EnsNet [19] is an end-to-end GAN-based network to erase the scene text on the whole image, but it cannot deal with text image generation. Wu *et al.* [17] propose a style retention network to edit the text in natural images. However, the foreground text of the generated license plate is intertwined with background. To solve these problems, we separate the real license plate into the foreground text and background and finally fuse them.

2.2 License Plate Generation

With the development of GAN, using GAN for license plate generation is widely studied by researchers. Wang *et al.* [15] apply Wasserstein distance loss to CycleGAN [21] as CycleWGAN, which is the first work of using GAN-generated images for the license plate recognition task. Similarly, Wu *et al.* [16] present CycleWGAN-GP to improve license plate recognition performance. Huang *et al.* [7] conduct comparative experiments of GAN, WGAN, DCGAN, and CycleGAN on the license plate generation task and conclude that CycleGAN performs best. Sun *et al.* [14] employ a P-module for paired images and a U-module for unpaired images to solve the data imbalance problem. Zhang *et al.* [18] propose a robust framework for license plate recognition, in which a tailored CycleGAN model (called AsymCycleGAN) provides balanced data for plate recognition.

These methods are mainly based on CycleGAN [21], taking an entire image as the input. However, as shown in Fig. 1, these methods cannot generate precise character strokes, especially the Chinese characters, and the sense of the overall image is unreal. It is because foreground text and the background have different features, and they interfere with each other when trained together. Hence, we try to solve this problem by separating the foreground text and background of the license plate.

3 Methodology

We propose a Chinese license plate generation method to generate robust license plate images by separating the foreground text and background. As shown in Fig. 2, the overall architecture consists of background generation network, supervision synthesizing algorithm, and text substitution network.

3.1 Background Generation Network

The background generation network is a GAN-based network composed of a generator G and a discriminator D, which can obtain the corresponding background image of the real license plate. The generator G aims to generate an image, and the discriminator D judges whether the image is real. The input of the background generation network is a pair of images, i.e., the real license plate image x and its corresponding ground-truth background image z. The output is a generated license plate background image y. The above process is defined as a mathematical min-max problem, which can be regarded as optimization of the following formula. G and D are trained alternatively with the cross-entropy loss.

$$\begin{aligned}\min_G \max_D E_{x\sim p_{data(x)},z}&[\log(1-D(x,G(x,z)))]\\ &+E_{x\sim p_{data(x,y)},z}[\log D(x,y)]\end{aligned} \tag{1}$$

The framework is an encoder-decoder based on Fully-Convolution-Network (FCN). The encoder uses a lightweight ResNet18 [6] network for feature extraction. The decoder is composed of five deconvolutional layers which expand the size of the feature map layer by layer to ensure that the generated license plate can be the same size as the input.

3.2 Supervision Synthesizing Algorithm

In this subsection, the supervision synthesizing algorithm can synthesize a set of supervision images for the text substitution network by the real license plate

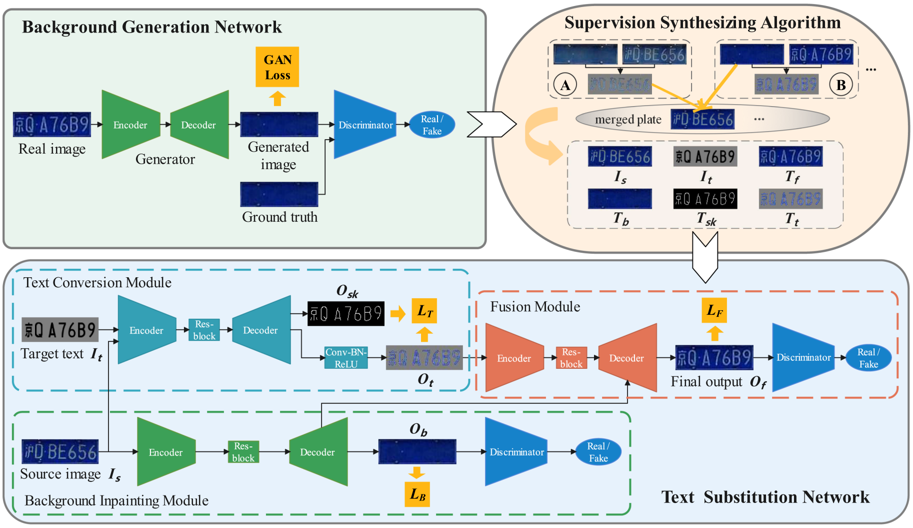

Fig. 2. Overall architecture. Firstly, the background generation network erases the text of the real license plate to obtain the background image. Secondly, the supervision synthesizing algorithm provides supervision data for the text substitution network. For example, it extracts the foreground text of the real license plate A and merges it with the background of another license plate B. The merged license plate I_s is the source image of the text substitution network, and the real license plate B is set as T_f to supervise the final output O_f. The target text I_t has the same text content with T_f. T_{sk} is the text skeleton of the real plate B. Finally, the text substitution network substitutes the foreground text of I_s with the target text I_t. T_{sk}, T_t, T_b, and T_f are the ground truth of O_{sk}, O_t, O_b, and O_f, respectively.

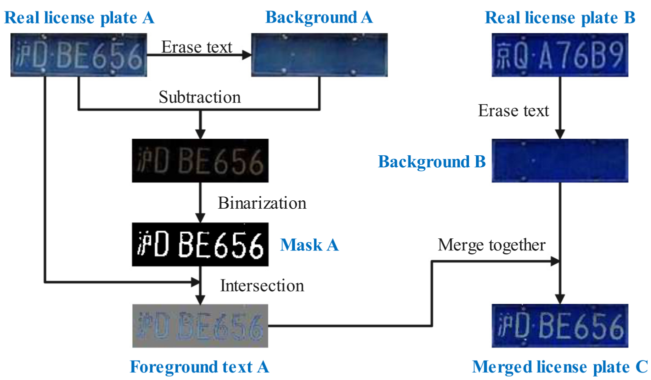

Fig. 3. The synthesizing process of the merged license plate. The merged license plate C is a synthetic image with the foreground text of real license plate A and the background of real license plate B.

image and its corresponding background image obtained before. The supervision images include the real license plate, background, foreground text, text skeleton, the standard format text, and the merged license plate.

The synthesizing process of the merged license plate is illustrated in Fig. 3. The background generation network outputs the background image of the input real license plate A. After subtraction and binarization, we get the mask image for extracting the foreground text. The foreground text A is obtained by the intersection of the real license plate A and its mask. Finally, the merged license plate C, a new image with the text of plate A and the background of plate B, is synthesized through conditional pixel-by-pixel traversal.

Correctly, we synthesize the supervision data via Algorithm 1. All the mathematical symbols are explained as follows, where I denotes input and T indicates target.

- I_s: the source image of the license plate;
- I_t: the target image with the target foreground text;
- T_f: the final output image;
- T_{sk}: the text skeleton of T_f;
- T_t: the foreground text of T_f with the grey background;
- T_b: the background of I_s and T_f;
- T_{mask}: the binary image of T_t.

T_f, T_{sk}, T_t, and T_b are the ground truth of O_f, O_{sk}, O_t, and O_b in Fig. 2, respectively.

Algorithm 1: Supervision Synthesizing Algorithm

Input: T_f, T_b;
Output: I_s, I_t, T_{sk}, T_t, T_b, T_f;
$T_{mask} = binarize(T_f - T_b)$;
$T_t = T_f \wedge T_{mask}$;
$T_{sk} = skeleton(T_{mask})$;
$I_t = drawIt(the\ text\ of\ T_f)$;
for $n = 1; n \leq 2000; n++$ **do**
    for $m = 1; m \leq 15; m++$ **do**
        $temp = randomly\ choose\ a\ T_t$;
        $I_s^{nm} = T_b^n \vee temp$; // T_b^n is reused 15 times.
        // Except for I_s^{nm}, others are copied 15 times.
        $I_t^{nm} = I_t^n$;
        $T_t^{nm} = T_t^n$;
        $T_f^{nm} = T_f^n$;
        $T_b^{nm} = T_b^n$;
        $T_{sk}^{nm} = T_{sk}^n$;
return I_s^{nm}, I_t^{nm}, T_{sk}^{nm}, T_t^{nm}, T_b^{nm}, T_f^{nm};

From a real license plate T_f and its background T_b, the mask T_{mask} and the foreground text T_t are synthesized by the process illustrated in Fig. 3. The text skeleton T_{sk} is extracted from T_{mask} by image corrosion and expansion. The target text I_t is drawn by a standard format with the same text content as the real license plate T_f. After that, each real license plate T_f has its corresponding T_b, T_t, T_{sk}, and I_t. We randomly choose 15 different foreground texts T_t and respectively merge them with the same background T_b to synthesize 15 different I_s for each T_f. Except for I_s, others are copied 15 times. Hence, we can expand 2,000 pairs of real license plates and their background images to 30,000 sets of supervision data that the text substitution network requires.

3.3 Text Substitution Network

The text substitution network consists of three modules: text conversion module, background inpainting module, and fusion module.

Text Conversion Module. The text conversion module aims to generate the foreground text image by substituting the text content of the source license plate with the target text while keeping the original text style. This module is an encoder-decoder FCN, including three down-sampling convolutional layers, four residual blocks [6], three deconvolutional layers, and a Conv-BN-ReLU block. The text conversion loss is the combination of L1 loss and skeleton loss [10]. The target text skeleton T_{sk} is used to obtain more precise skeleton and stroke of the foreground text.

$$L_T = \|T_t - G_T(I_t, I_s)\|_1 + \alpha(1 - \frac{2\sum_i^N (T_{sk})_i (O_{sk})_i}{\sum_i^N (T_{sk})_i + \sum_i^N (O_{sk})_i}), \tag{2}$$

where G_T denotes the text conversion module. O_{sk} is the output text skeleton image. α is the regularization parameter, which is set to 1.0 by default.

Background Inpainting Module. The background inpainting module erases the text and fills the text region with appropriate texture. This module is similar to the text conversion module, but the decoder is composed of three up-sampling transposed convolutional layers. A generator G_B and a discriminator D_B are trained alternatively with adversarial loss and L1 loss, where β is set to 10.

$$\begin{aligned} \mathrm{L}_B = {} & \mathrm{E}_{(T_b, I_s)}[\log D_B(T_b, I_s)] + \mathrm{E}_{I_s} \log[1 - D_B(G_B(I_s), I_s)] \\ & + \beta \|T_b - G_B(I_s)\|_1 \end{aligned} \tag{3}$$

Fusion Module. The fusion module fuses the substituted foreground text with the background and generates a new license plate. It includes a generator G_F, a discriminator D_F, and an encoder-decoder FCN. The fusion loss function is the sum of adversarial loss and L1 loss.

$$L_F = \mathrm{E}_{(T_f, I_t)}[\log D_F(T_f, I_t)] + \mathrm{E}_{I_t} \log[1 - D_F(O_f, I_t)] + \theta \|T_f - O_f\|_1, \tag{4}$$

$$O_f = G_F(G_T(I_t, I_s), G_B(I_s)), \tag{5}$$

where O_f is the output final image, $\theta = 10$.

4 Experiments

4.1 Datasets

Synthetic Dataset is synthesized by SynthText [5], containing 50,000 synthetic license plate images. The main idea is to select fonts, color, parameters of deformation randomly to generate styled text, then render it on the background image.

Yizhi2000 is a private license plate dataset containing 2,000 license plate images of good quality cropped from driving recorder images. They are resized to 150×48. The distribution of license plate data is unbalanced(the number of license plates in each province is extremely uneven). Besides, there are no data for some provinces, such as Qinghai, Tibet.

SUSY [20] is a public license plate dataset that contains all the 31 provinces in the Chinese mainland but without annotating the license plates. We extract the license plates with a YOLOv3 [13]. There are about 70 license plate images for each province.

GLPD is a generated license plate dataset we have released publicly, containing 1,000 license plates for each province, including all the 31 provinces of the Chinese mainland. To generate license plate images more effectively, we add SUSY to train the text substitution network to solve the data scarcity of some provinces.

4.2 Implementation Details

In our experiments, Adam [9] optimizer is adopted to train the model. We use the Synthetic Dataset to pretrain the background generation network, text substitution network, and license plate recognition network. We select 100 license plates from Yizhi2000 and erase the foreground text manually to finetune the background generation network. The initial learning rate is 5×10^{-4}, and the maximum iteration is 1,000. Afterward, we test the background generation network with Yizhi2000 and synthesize supervision data for the text substitution network. The input images of the text substitution network are resized to 300×96, and the batch size is set to 8. The learning rate is initially set to 1×10^{-4}, and the maximum iteration is 50,000. For the license plate recognition experiments in Sect. 4.5, we use SUSY to test the accuracy of different recognition models.

4.3 Evaluation Metrics

In this work, we use recognition accuracy (RA) and character recognition accuracy (CRA) to evaluate the effectiveness of generated data on the performance of the text recognition model [2]. Same with CycleWGAN [15], RA and CRA can be calculated as follows.

$$RA = \frac{Number\ of\ correctly\ recognized\ license\ plates}{Number\ of\ all\ license\ plates} \tag{6}$$

$$CRA = \frac{Number\ of\ correctly\ recognized\ characters}{Number\ of\ all\ characters} \tag{7}$$

Similar to RA and CRA, CRA-C is the recognition accuracy of the first character, namely the Chinese character, and CRA-NC is the character recognition accuracy of the last six characters, i.e., the letters and numbers. Besides, we also adopt other commonly used metrics to evaluate the image quality of the generated license plates.

- Fréchet Inception Distance (FID), which measures the distance between the distributions of synthesized images and real images;
- MSE, also known as l2 error;
- PSNR, which computes the ratio of peak signal to noise;
- SSIM, which computes the mean structural similarity index between two images.

A lower FID and MSE or higher SSIM and PSNR mean the generated images are more similar to ground truth.

4.4 Qualitative Experiments

As shown in Fig. 1, we compare some generated examples of the existing methods designed for license plate generation with our method visually. CycleWGAN [15], CylceWGAN-GP [16], and AsymCylceGAN [18] generate mottled background, ghosted, and blurred strokes. The images of Sun *et al.* [14] are better than the first three methods, but the texts are still unclear and hazy. Moreover, there is a big gap between the images generated by the four methods and the real license plate images on the real sense of overall image. The results indicate that the license plate images generated by our method have the best visual effect, and it is hard to distinguish them from the real license plate images.

Moreover, we reproduce pix2pix [8], CycleGAN [21], and SRNet [17] using the same dataset and training settings as our method. Figure 4 shows some examples generated by these methods with the same target text. Pix2pix can only generate some twisty numbers, and the foreground text and background are intertwined. CycleGAN performs better on letters and numbers but fails on the Chinese characters. Both pix2pix and CycleGAN have the mode collapse problem. These phenomena suggest that generating the whole license plate images can cause the mutual interference of the foreground text and background. The strokes of texts generated by SRNet are also blurred, and it cannot generate the Chinese characters with complex stroke skeleton. Finally, our methods can achieve better visual performance on the precise text strokes, background texture, and overall coordination and realism. Even for complex Chinese characters, such as the abbreviations of Jiangxi and Hubei, our method can generate precise strokes and skeletons. This experiment verifies that the license plates generated by our method are more homogeneous with the real license plates.

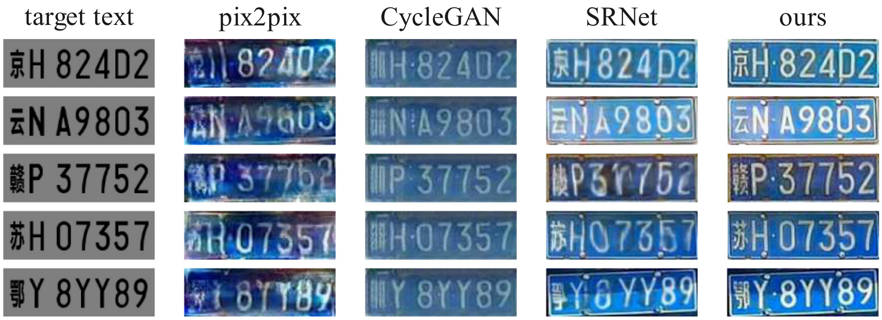

Fig. 4. Visual comparison of our method with pix2pix [8], CycleGAN [21], and SRNet [17]. The provinces of the five license plates are Beijing, Yunnan, Jiangxi, Jiangsu, Hubei (up to down).

4.5 Quantitative Experiments

Recognition Accuracy. We apply a text recognition model [2] to compare the effectiveness of different data generation methods for the LPR task. The total number of training sets is 30000 for all methods. Pix2pix, CycleGAN, SRNet, and ours are based on the same target texts. From Table 1, we can find that pretraining only makes the recognition model learn some letters and numbers. Pix2pix and CycleGAN are lower than traditional data augmentation methods on all four metrics because the generated images are much different from the real license plate images. Besides, CycleGAN cannot learn Chinese characters. SRNet also performs inadequately due to the gap between synthetic data and real data. Finally, even with a slight drop of CRA-NC, our method achieves

Table 1. The accuracy (%) of the text recognition model trained by different dataset. “pretrain” means training by the same synthetic dataset as GANs. “augmentation” means augmenting images with random crop, Gauss noise, etc.

Methods	RA	CAR	CRA-C	CRA-NC
Pretrain	6.4	73.1	22.8	79.3
Real data + augmentation	76.2	96.5	77.0	**99.8**
pix2pix [8]	55.9	93.0	60.7	98.1
CycleGAN [21]	6.4	74.0	17.3	81.1
SRNet [17]	58.5	93.4	66.2	97.8
Ours	**84.4**	**97.6**	**89.2**	98.9

the best RA, CRA, and CRA-C, especially for the CRA-C, which proves our method can significantly improve Chinese character recognition. In conclusion, this experiment validates the effectiveness of our proposed method to enhance license plate recognition performance. The four types of accuracy are generally low due to data scarcity of some provinces.

Other Metrics. For FID, we use the same 30,000 generated data of pix2pix [8], CycleGAN [21], SRNet [17], and our method as before. For MSE, PSNR, and SSIM, we set the target texts the same as the license plate texts in the real data, and then calculate these three metrics of the output images and the corresponding real license plate images. The results in Table 2 prove that the image quality of our generated license plates is much better than other methods. Notably, FID even drops to 10.64, and SSIM increases to 0.9404. This experiment proves that the license plates generated by our method are more homogeneous with the real license plates.

Table 2. Quantitative evaluation results on FID, MSE, PSNR, and SSIM.

Methods	FID	MSE	PSNR	SSIM
pix2pix [8]	192.43	0.5055	12.94	0.2238
CycleGAN [21]	110.43	0.3834	14.40	0.2566
SRNet [17]	48.13	0.1109	20.17	0.6221
Ours	**10.64**	**0.0222**	**27.77**	**0.9404**

5 Conclusion

In this work, we propose a robust Chinese license plate generation method by separating the foreground text and background of license plate images, improving the image quality of generated license plates. Our method can generate license plate images with the precise stroke and a real sense of the overall image. Besides, experiments on license plate recognition prove the effectiveness of our method for increasing LPR accuracy. Moreover, we release a generated license plate dataset with 1,000 images for each province, including all 31 provinces of the Chinese mainland.

In the future, we will try to solve inferior performance in some provinces due to data scarcity and explore to generate more diverse license plate data.

References

1. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: Proceedings of the 34th International Conference on Machine Learning (ICML), vol. 70, pp. 214–223. PMLR, Sydney, NSW, Australia (2017)

2. Baek, J., et al.: What is wrong with scene text recognition model comparisons? dataset and model analysis. In: Proceedings of the International Conference on Computer Vision (ICCV), pp. 4715–4723. IEEE, Seoul, Korea (South) (2019)
3. Goodfellow, I.J., et al.: Generative adversarial networks. CoRR abs/1406.2661 (2014)
4. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.: Improved training of wasserstein gans. In: Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems, pp. 5767–5777. Long Beach, CA, USA (2017)
5. Gupta, A., Vedaldi, A., Zisserman, A.: Synthetic data for text localisation in natural images. In: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2315–2324. IEEE, Las Vegas, NV, USA (2016)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. IEEE, Las Vegas, NV, USA (2016)
7. Huang, J., et al.: Research on vehicle license plate data generation in complex environment based on generative adversarial network. In: 11th International Conference on Digital Image Processing (ICDIP), vol. 11179, p. 1117932. International Society for Optics and Photonics (2019)
8. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1125–1134. IEEE, Honolulu, HI, USA (2017)
9. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA (2015)
10. Milletari, F., Navab, N., Ahmadi, S., V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 4th International Conference on 3D Vision (3DV), pp. 565–571. IEEE, Stanford, CA, USA (2016)
11. Mirza, M., Osindero, S.: Conditional generative adversarial nets. CoRR abs/1411.1784 (2014)
12. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. In: 4th International Conference on Learning Representations (ICLR), San Juan, Puerto Rico (2016)
13. Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. CoRR abs/1804.02767 (2018)
14. Sun, M., Zhou, F., Yang, C., Yin, X.: Image generation framework for unbalanced license plate data set. In: Neural Information Processing - 26th International Conference (ICONIP), vol. 1143, pp. 127–134. Springer, Sydney, NSW, Australia (2019)
15. Wang, X., You, M., Shen, C.: Adversarial generation of training examples for vehicle license plate recognition. CoRR abs/1707.03124 (2017)
16. Wu, C., Xu, S., Song, G., Zhang, S.: How many labeled license plates are needed? In: Lai, J.-H., Liu, C.-L., Chen, X., Zhou, J., Tan, T., Zheng, N., Zha, H. (eds.) PRCV 2018. LNCS, vol. 11259, pp. 334–346. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-03341-5_28
17. Wu, L., Zhang, C., Liu, J., Han, J., Liu, J., Ding, E., Bai, X.: Editing text in the wild. In: Proceedings of the 27th ACM International Conference on Multimedia, pp. 1500–1508. ACM, Nice, France (2019)
18. Zhang, L., Wang, P., Li, H., Li, Z., Shen, C., Zhang, Y.: A robust attentional framework for license plate recognition in the wild. CoRR abs/2006.03919 (2020)

19. Zhang, S., Liu, Y., Jin, L., Huang, Y., Lai, S.: Ensnet: ensconce text in the wild. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 801–808. AAAI, Honolulu, Hawaii, USA (2019)
20. Zhao, Y., Yu, Z., Li, X., Cai, M.: Chinese license plate image database building methodology for license plate recognition. J. Electron. Imaging **28**(1), 013001 (2019)
21. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: International Conference on Computer Vision (ICCV), pp. 2242–2251. IEEE, Venice, Italy (2017)