

Practical License Plate Recognition in Unconstrained Surveillance Systems with Adversarial Super-Resolution

Younkwan Lee, Jiwon Jun, Yoojin Hong and Moongu Jeon

Machine Learning and Vision Laboratory, Gwangju Institute of Science and Technology, Gwangju, South Korea

Keywords: Intelligent Transportation Systems, Visual Surveillance, License Plate Recognition, Super-Resolution, Generative Adversarial Networks.

Abstract: Although most current license plate (LP) recognition applications have been significantly advanced, they are still limited to ideal environments where training data are carefully annotated with constrained scenes. In this paper, we propose a novel license plate recognition method to handle unconstrained real world traffic scenes. To overcome these difficulties, we use adversarial super-resolution (SR), and one-stage character segmentation and recognition. Combined with a deep convolutional network based on VGG-net, our method provides simple but reasonable training procedure. Moreover, we introduce GIST-LP, a challenging LP dataset where image samples are effectively collected from unconstrained surveillance scenes. Experimental results on AOLP and GIST-LP dataset illustrate that our method, without any scene-specific adaptation, outperforms current LP recognition approaches in accuracy and provides visual enhancement in our SR results that are easier to understand than original data.

1 INTRODUCTION

License plate recognition (LPR) is a fundamental and essential process of identifying vehicles and can be extended to a variety of real-world applications. LPR methods have been widely studied over the last decade, and are especially of big interest in intelligent transport systems (ITS) applications such as access control (Chinomi et al., 2008), road traffic monitoring (Noh et al., 2016; Pu et al., 2013; Song and Jeon, 2016; Lee et al., 2017; Yoon et al., 2018) and traffic law enforcement (Zhang et al., 2011). Since all license plate recognition methods always deal with the letters and numbers in images, they are closely related to image classification (Simonyan and Zisserman, 2014; Russakovsky et al., 2015) and text localization (Anagnostopoulos et al., 2006).

Conventional LPR methods typically include two stages: character localization and character recognition. Those methods are widely designed for unrealistically most constrained scenarios: a high-quality resolution and an unrotated frontal or rear image. However, unlike the ideal situation, many traffic surveillance cameras scattered around the world are operating in a number of unconstrained scenarios: they produce poor-resolution images and tilted license plates as shown in Figure 1. Although considerable progress

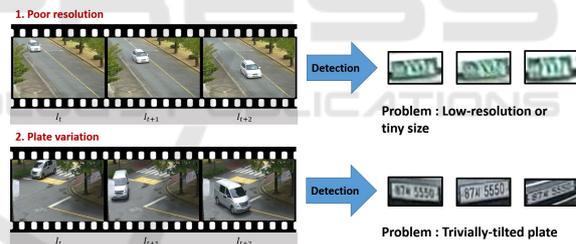


Figure 1: Example in GIST-LP dataset. Poor-resolution and plate variation are common challenging issues on license plate recognition problem.

of computer vision technology has been made, existing methods may fail to recognize license plates in such an environment without considering any unconstrained conditions. As a consequence, we find its limitations in three aspects: first, many license plate samples only constitute incomplete text search space; second, the projection angle of the sample is tilted with respect to the image plane at an angle of up to 30 degrees, interfering character exploitation; third, bad text localization often results in erroneous outputs.

Based on this finding, we propose a novel deep convolutional neural network based method for better LPR.

Adversarial Super-resolution. We suggest an adversarial super-resolution (SR) method including a ge-

erator and a discriminator networks over an image area. Modern SR method (Dong et al., 2014) commonly targets the pixel-wise average as optimization goal, minimizing the mean squared error (MSE) between the super-resolved image and the ground truth, which leads to the smoothing effect, especially across text. Instead, we follow (Ledig et al., 2017)’s generator network, which solves minimax game as optimization goal, avoiding a smoothing effect, which provide a sharpening effect. Combined with SR in generator, we introduce a new loss function that encourages the discriminator to count characters and distinguish whether SR or high-resolution(HR) sample concurrently. Character counting results from the discriminator help improve character recognition performance in one-stage recognition module as a conditional term.

Reconstruction Auto-encoder. We always reconstruct the samples to straighten when the horizontally or vertically tilted license plate is projected onto the image plane. To address this issue, we utilize the convolutional auto-encoder network with the objective function as the difference between the tilted image and the straightened image. By doing so, it serves as a preprocessing for correct character exploitation.

One-Stage Recognition. We do use the commonly used character segmentation and localization process. Instead, we propose a unified character localization and recognition approach as one-stage. One-stage recognition is not only more intuitive, but also more accurate than segmentation that requires precise an estimate of each pixel’s class. Our One-stage method divides the input image into a $1*S$ grid, and detects LP at three different scales which includes a conditional term. The result of our character localization using each grid cell is naturally unified with character classification.

In summary, our key contributions are:

- We show that adversarial SR module and AE based reconstruction module in the real world for unconstrained surveillance cameras can improve the recognition performance greatly by (2.57% (AOLP) and 8.06% (GIST-LP)) compared with the state-of-the-art methods.
- The One-stage method combined with the conditional term, instead of the two-stage method (character detection and classification), reduced the localization and classification error.
- We collected a dataset of challenging license plate samples from unconstrained conditions accompanied by the text annotations (1,800 samples, 50 different license plates).

2 RELATED WORK

2.1 License Plate Recognition

Traditionally, numerous LPR methods proposed consists of the two stages: semantic segmentation of the exact character region and recognition of the characters. The related methods generally utilize discriminate features, such as edge, color, shape and texture but does not show good results. Edge-based methods (Kim et al., 2000; Zhang et al., 2006; Wang and Lee, 2003; Kim et al., 2000; Zhang et al., 2006) and geometrical features (Wang and Lee, 2003) assume the presence of characters in the license plate. Many color-based methods (Shi et al., 2005; Chen et al., 2009) usually use the combination of the license plate and the characters.

However, since the two-stage methods are not only slow to run, but also take more time to converge for optimized training due to the double networks, one-stage pipeline based methods, segmentation-free approach (Zherzdev and Gruzdev, 2018; Cheang et al., 2017; Li and Shen, 2016; Wang et al.,), including segmentation and recognition at once, are proposed. Most segmentation-free models take advantage of deeply learned features which outperforms traditional methods on the task of classification by deep convolutional neural networks (DCNN) (Simonyan and Zisserman, 2014; He et al., 2016) and data-driven approaches (Russakovsky et al., 2015). The core underlying assumption of these methods extracts features directly without sliding window for LPR. As examples of these models, Sergey *et al.* (Zherzdev and Gruzdev, 2018) adopted a lightweight convolutional neural network to learn end-to-end way. In another work that use RNN module, Teik Koon *et al.* (Cheang et al., 2017) proposed CNN-RNN unification model that feed the entire image as input. It is assumed that the context of the entire image is further evaluated for exact classification than the sliding window approaches being. Also, Hui *et al.* (Li and Shen, 2016) utilized a cascade framework using DCNN and LSTM and Xinlong *et al.* (Wang et al.,) proposed DCNN and a bidirectional LSTM to use sequence labeling.

2.2 Adversarial Learning

The generative adversarial network (GAN) (Goodfellow et al., 2014; Radford et al., 2015; Radford et al., 2015) is an amazing solution for training deep neural network of generative models, which aim to learn the probability distributions of the input data. Originally, GAN is suggested to yield the more realistic-fake images (Frid-Adar et al., 2018), but recent researches

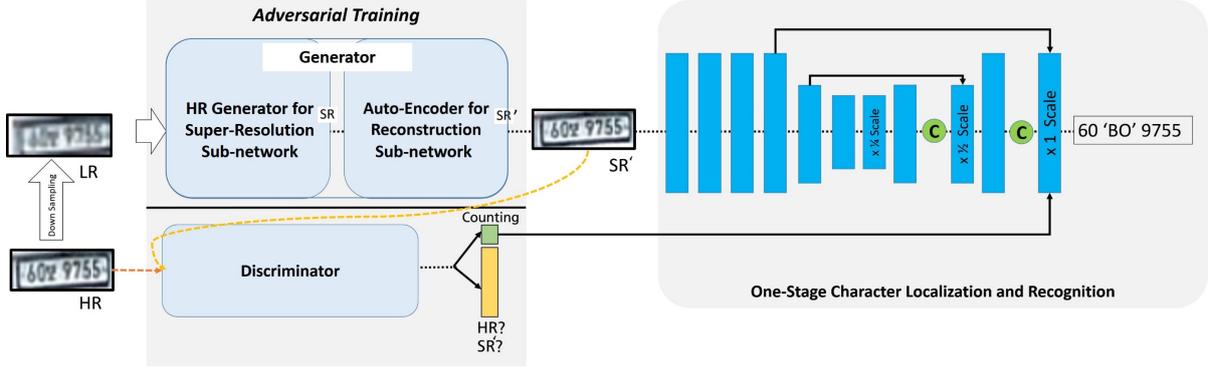


Figure 2: The proposed license plate recognition pipeline.

show that this adversarial technique can be utilized to produce the specific training algorithms. e.g., generative focused tasks; super-resolution (Nguyen et al., ; Ledig et al., 2017; Lee et al., 2018), style transfer (Zhu et al., 2017; Li et al., 2017), natural-language processing (Rajeswar et al., 2017) and discriminative focused tasks; human pose estimation (Chou et al., 2017; Peng et al., 2018).

3 PROPOSED METHOD

In this section, we describe the details of the proposed end-to-end pipeline for LPR. The schematics of the method is illustrated in Figure 2. We first introduce the adversarial network to super-resolve the input image, and reconstruct its output. Then, the details of the proposed one-stage character recognition network are presented for recognizing characters on the license plate and locating individual text regions without character segmentation. Finally, we describe a training process to find optimal parameters of our model.

3.1 Adversarial Network Architecture

Adversarial learning techniques have been widely used in many tasks (Frid-Adar et al., 2018; Zhu et al., 2017; Rajeswar et al., 2017; Chou et al., 2017), providing boosted performance through adversarial data or features. In vanilla GAN (Goodfellow et al., 2014), a minimax game is trained by alternately updating a generator sub-network G and a discriminator sub-network D simultaneously. The value function of the generator G and the discriminator D is defined as:

$$\min_{\theta_G} \max_{\theta_D} V(D, G) = \mathbb{E}_{x \sim p_{real}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_{fake}(z)} [\log(1 - D(G(z)))] \quad (1)$$

where p_{real} is the real data distribution observation from x and p_{fake} is the fake data distribution observation from a random distribution z . These sub-networks have conflicting goals to minimize their own cost and maximize the opposite's cost. Therefore, the conclusion to play the minimax game can be that the probability distribution (p_{fake}) generated by the generator G exactly matches the data distribution (p_{real}). After all, the discriminator D will not be able to distinguish between sampling distribution from the generator G and real data distribution. At this time, for the fixed generator, the optimal discriminator function is as follows:

$$D_G^*(x) = \frac{p_{real}(x)}{p_{real}(x) + p_{fake}(x)}. \quad (2)$$

In a similar way, we modified the minimax value function in the vanilla GAN for solving SR so that the generator G consisting of a HR generator G_{SR} and a reconstruction network G_{recon} creates an HR image from LP image, while the discriminator D trains to distinguish the HR fake image obtained by the generator from the actual LR image. This adversarial SR process can be defined as follows:

$$\min_{\theta_G} \max_{\theta_D} V(D, G) = \mathbb{E}_{I^{HR} \sim p_{train}(I^{HR})} [\log D_{\theta_D}(I^{HR})] + \mathbb{E}_{I^{LR} \sim p_G(I^{LR})} [\log(1 - D_{\theta_D}(G_{\theta_G} I^{LR}))], \quad (3)$$

where I^{HR} is the high-resolution image, I^{LR} is the low-resolution image, θ_G and θ_D denote the parameters trained by a feed-forward CNN G_{θ_G} and D_{θ_D} respectively.

Generator Network. Different from (Goodfellow et al., 2014), our generator network is composed of two sub-networks: (1) HR Generator G_{SR} and (2) Convolutional Auto-encoder for reconstruction G_{recon} as shown in Figure 2. The former is a series of convolutional layers and fractionally-strided convolution layers (i.e. *upsample layer*) inspired by (Ledig et al.,

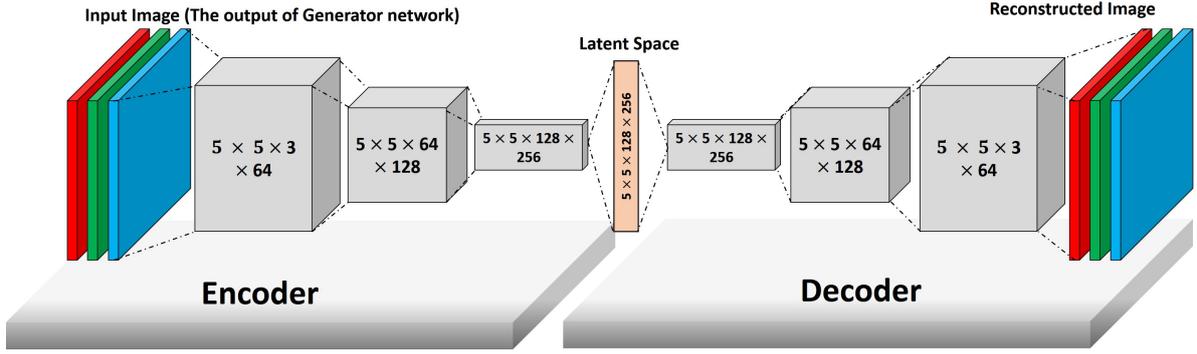


Figure 3: The proposed Auto-Encoder based reconstruction sub-network structure.

2017). We use two upsample layers (2 times upsampling) as proposed by Radford et al. (Radford et al., 2015), and acquire a 4 times enhanced image from them.

In addition to its network, we include a reconstruction sub-network for the refinement task of image with enhanced resolution. Given the output of 4 times super-resolved image, our proposed network aims at discovering that it corrects slightly distorted image through denoising learning manner. Basically, we employ a convolutional neural network (CNN) as encoder and decoder, as shown in Figure 3. Although both encoder and decoder consist of the same number of convolutional layers, the former adds MaxPooling2D layers for spatial down-sampling, while the latter adds UpSampling2D layers, with the BatchNormalization (Ioffe and Szegedy, 2015).

Discriminator Network. Figure 2 shows the architecture of the discriminator network and its output components. Inspired by VGG19 (Simonyan and Zisserman, 2014), we follow the same network structure. To discriminate exact object regions, we design all the fully-connected layers to split into two parallel branches to obtain two outputs: (1) how many characters are in the image as counting result f_{count} and (2) the HR vs. SR f_{GAN} .

3.2 Character Recognition Network Architecture

In this section, we describe the details of the proposed character recognition approach where localization and recognition are integrated into one-stage. We employ YOLO v3 (Redmon and Farhadi, 2018) as our detection network. To achieve scale-invariance, it detects characters at three scales, which are given by diminished dimensions of the image by 32, 16 and 8 each other, without the MaxPooling2D layer. Unlike previous model (Redmon and Farhadi, 2017), this allows better detection performance of small size character, which is optimized for character on a license

plate that is mostly expressed in small size localization and recognition with residual skip connections.

The shape of detection kernel denoted as $1 \times 1 \times (B \times (5 + C))$, where B is the number of bounding boxes, 5 is the sum of the four attributes of bounding boxes (coordinates (x, y) , width and height) and one object confidence score and C is the number of classes. In our method, we define the detection kernel size as $B = 3$ and C is 66 (10 numbers (0-9), 26 English letters and 30 Korean letters), result in $1 \times 1 \times 213$.

Furthermore, we add the counting information output f_{count} from the discriminator as a conditional term in our character recognition model. The last layer of recognition model has the previous layer’s output and f_{count} as inputs. We demonstrate that our recognition model can be extended to the sophisticated model where it can accurately count and localize any character in any input. These are further discussed later in Section 4.4.

3.3 Training

In this section, we discuss the objective to optimize our adversarial network and one-stage recognition network. Let I_i^{LR} , I_i^{HR} and I_i^{SR} denote a low-resolution image, high-resolution image and SR image, respectively. Given a training dataset $\{I_i^{LR}, I_i^{HR}, text_{j=1}, \dots, text_{j=count_i}, count_i\}_{i=1}^N$, our goal is to learn the adversarial model that predicts SR image from low-resolution image and recognition model that predicts character’s class and location from SR image.

Pixel-wise Loss. To force the generated plate image to high-resolution ground truth, our generator network is optimized for the MSE loss in each pixel values between the generated image sets and the small and blurry plate image sets calculated as follows:

$$L_{MSE}(w) = \frac{1}{N} \sum_{i=1}^N (\|G_{S1w1}(I_i^{LR}) - I_i^{HR}\|^2) + \|G_{S2w2}(G_{S1w1}(I_i^{LR}) - I_i^{HR})\|^2, \quad (4)$$

where G_{S1} means HR generator, G_{S2} denotes the reconstruction network, and w are the parameters of generator network.

Adversarial Loss. In order to provide a sharpening effect to the generated image different from the MSE loss that gives the smoothing effect, we define adversarial loss as:

$$L_{adv} = \frac{1}{N} \sum_{i=1}^N \log(1 - D_{\theta}(G_w(I_i^{LR}))). \quad (5)$$

Adversarial loss amplifies the photo-realistic effect and is trained in the direction of deception of the discriminator.

Reconstruction Loss. In order to let the quality of generated images by the $S1$ to be more photo-realistic, we propose the reconstruction loss that corrects changes in the generated image topology that interfere with the detection and is defined as follows:

$$L_{const} = \frac{1}{N} \sum_{i=1}^N \|G_{S1}(I_i^{LR}) - G_{S2}(G_{S1}(I_i^{LR}))\|. \quad (6)$$

The reconstruction loss is calculated as L1 loss, the difference between the output of G_{S1} and G_{S2} .

Classification Loss. The classification loss is playing both the roles of an character counting task as well as the discrimination task. To be more specific, the discriminator takes an image as input and classified it into two outputs: the HR real natural image or the SR fake image and the numbers of characters respectively. The loss of this multi-task is calculated as follows:

$$L_{clc} = \frac{1}{N} \sum_{i=1}^N (\log((y_i \wedge count_i) - D_{\theta}(G_w(I_i^{LR}))) + \log((y_i \wedge count_i) - D_{\theta}(I_i^{HR}))), \quad (7)$$

where y_n represents prediction value of the number of characters and the operations \wedge with y_n and $count_i$ output 1 if it predicts correctly or 0 respectively.

4 EXPERIMENTAL RESULTS

4.1 Setup

All the reported implementations are based on TensorFlow as learning framework, and our method has done on the NVIDIA TITAN X GPU. First of all, we use the YOLO-v3 for the pre-trained model on COCO (Lin et al., 2014) as our one-stage recognition model so that we trained license plate images by fine-tuning their network parameters.

Also, to avoid the premature convergence of the discriminator network, the generator network is updated more frequently than original one. In addition,

higher learning rate is applied to the training of the generator. For stable training, we use a technique called gradient clipping trick (Pascanu et al., 2013) and the Adam optimizer (Kingma and Ba, 2014) with a high momentum term. For the discriminator network, we use the VGG-19 (Simonyan and Zisserman, 2014) model pre-trained on ImageNet as our backbone network and we divide all the fully connected layers into two parallel f_{count} and f_{GAN} layers. The weights in all parallel fully connected layers are initialized from the standard Gaussian distribution with zero-mean, a standard deviation of 0.01 and the constant 0 as the bias in all layers. All models are trained on loss function for first 10 epochs with initial learning rate of 10^{-4} . After that, we set the learning rate to a further reduced 10^{-5} for the remaining epochs. Finally, batch normalization (Ioffe and Szegedy, 2015) is used in all layers of generator and discriminator, except the last layer of the G and the first layer of the D .

4.2 Dataset

AOLP: This dataset(Hsu et al., 2013) includes 2,049 images of Taiwan license plates, which are collected from the unconstrained surveillance scenes. AOLP dataset is divided into three subsets: access control (AC) with 681 samples, traffic law enforcement (LE) with 757 samples, and road patrol (RP) with 611 samples, based on diverse application parameters. 100 samples per subset are used for the training, and the rest of the 581(AC)/657(LE)/511(RP) samples are used for testing. More specifically, AC has a narrow range of variation conditions, while LE/RP have a wider range of variation conditions. Therefore, compared to the AC subset, LE/RP are more challenging subsets because they require a wider range of search conditions on the experiments. Besides, the RP samples collected via mobile have more challenging conditions because of the larger pan and orientation changes compared to the LE samples collected at road cameras with fixed viewing angles.

GIST-LP: We collected and annotated a new dataset GIST-LP for LPR. Our dataset is targeted on images captured from surveillance cameras under unconstrained scenes. We do not limit the license plate always to be large and front. We used traffic surveillance cameras which has 1920 x 1080 pixels of spatial resolution. We annotated the characters, including Korean (30 categories) and numbers (0-9, 10 categories) for all of the license plate images. In total, there are 1,800 license plates that appear in 1,569 frames. For license plate images, the characters are usually small-sized, blurred or tilted without occlusion. The dataset include information about bounding box for each cha-



Figure 4: Samples from the unconstrained surveillance cameras in GIST-LP dataset.

		08-‘DO’-7824
		68-‘RO’-0365
		60-‘BO’-9755
		64-‘SU’-6668
		74-‘JOO’-3203
		28-‘SO’-7323

Figure 5: Example in GIST-LP dataset (Laroca et al., 2018). Qualitative sample images of recognition results. The first column shows the original plates, the second column shows the character localization results and the third indicates the recognition results.

acter and text class (Koreans and numbers).

4.3 Comparison with other Methods

In the experiment with AOLP, we compared our method with the state-of-the-art license plate recognition approaches (Anagnostopoulos et al., 2006; Jiao et al., 2009; Smith, 2007; Hsu et al., 2013). The results are listed in Table 1, which are provided with accuracy of recognition to evaluate both text localization and classification are all performed well at the same time. We see that our method obtained the highest performance (*i.e.* 96.74%) on the all subsets, and outper-

formed the state-of-the-art LPR approaches by more than 2.5%. Also, it is important to note that, under the fairly tilted conditions, our method operated consistently robust and successfully detects the characters, while the baseline fail to detect. Furthermore, one interesting finding of these results is that, based on Figure 6 (b,c), the addition of adversarial loss lead to the highlighting of the positive features, while decimating of other irrelevant features. By doing so, it was further improved when detecting under night or confusing conditions. Based on these observations, our proposed method operated at least as well as others, which outperformed all other methods in most cases.

To show the results of experiment of LPR with GIST-LP, we compared our method with (Girshick et al., 2014; Ren et al., 2015) and followed the standard metrics (*i.e.* accuracy of recognition) of the GIST-LP. There were many tiny license plates in GIST-LP, making character detection not be accurate. Hence, we found that the state-of-the-art method (Redmon and Farhadi, 2018) that performed without considering the tiny size and blurred condition recorded on the inferior performance. However, our method mitigated the influence of these conditions and indicated these license plates successfully. Under such a challenging condition, our LPR performance still achieved a comparable performance (93.83%) over all other state-of-the-art LPR approaches, as shown Table 2.

4.4 Ablation Study

In the proposed method, the loss functions of adversarial networks locate different regions, each with their unique roles. In order to inspect its influence on character recognition performance, we removed one loss

Table 1: Comparison of our method with other state-of-the-art method on the AOLP dataset.

Method	Performance			
	AC	LE	RP	Avg
(Anagnostopoulos et al., 2006)	92.00%	88.00%	91.00%	86.34%
(Jiao et al., 2009)	90.00%	86.00%	90.00%	88.51%
(Smith, 2007)	96.00%	83.00%	83.00%	87.31%
(Hsu et al., 2013)	95.00%	93.00%	94.00%	94.17%
Baseline (YOLO v3) (Redmon and Farhadi, 2018)	94.66%	89.04%	89.04%	90.90%
without pixel-wise MSE loss	97.24%	94.67%	94.91%	95.60%
without reconstruction loss	96.21%	88.89%	94.32%	92.91%
without adversarial loss	95.18%	87.67%	93.93%	92.00%
without classification loss	96.39%	94.98%	96.48%	95.88%
Ours	97.59%	95.89%	96.87%	96.74%

Table 2: Comparison of our method with other state-of-the-art method on the GIST-LP dataset.

Method	Performance
RCNN based on VGG-16 (Girshick et al., 2014)	74.44%
RCNN based on ZFNET (Girshick et al., 2014)	72.11%
Faster-RCNN <i>et al.</i> (Ren et al., 2015)	86.77%
Baseline (YOLO v3) (Redmon and Farhadi, 2018)	84.16%
Ours without pixel-wise MSE loss	91.78%
Ours without reconstruction loss	89.00%
Ours without adversarial loss	87.72%
Ours without classification loss	90.78%
Ours	93.83%

function from the objective function at a time and performed an ablation study with it to compare the complete objective function. Most extremely, we perform experiments that compare the baseline and overall objective function, which obtain the superior performance by a considerable gap (5.84% / 9.67%) from Table 1 and 2.

Also, when removing one loss function from the overall objective function our method shows a considerable performance drop. First of all, even if the MSE loss is not suitable for tiny objects due to the smoothing effect, if there is no MSE loss, the performance degradation is up to 1.14% (in AOLP) / 2.05% (in GIST-LP), affecting the image up-scaling super-resolution. Then the reconstruction loss affects the correct converting of the tilted plate, because the SR performance of the generator is somewhat dependent on the degree of tilted angle of the license plate, and it leads to about 3.83% (in AOLP), 4.83% (in GIST-LP) improvement in performance. In another step, we observe that adversarial loss leads to the sharpened super-resolved result of minimax game. Thus it has a great influence on the detection performance as shown in Figure 6. The GIST-LP dataset which has relatively more tiny plates than AOLP dataset has found a performance improvement of almost 4.74% as shown Table 2, and the AOLP dataset also achieves performance improvement of nearly 6.11% as shown

Table 1. Finally removing classification loss in the objective function shows a significant impact on the character recognition performance, which observes an impressive improvement of 0.86% (in AOLP) and 3.15% (in GIST-LP). This proves that our two parallel fully-connected layers for classification affect the classification performance for our text localization of the detector as well as the SR performance of the generator. Also, we demonstrate that the counting term as conditional data benefits to better explore the space of the character localization as much as possible.

4.5 Qualitative Results

As shown in Figure 6., we give additional examples of the clear LP generated by the proposed generator network from the tiny ones. Upon thorough investigation of the generated images, we find that our method learn strong priors using the proposed new loss functions of GAN by focusing on images of plate contour, certain letters and numbers as shown in Figure 6 (a). It implies that the proposed loss significantly allows visually clearer LP and can be used to solve the ill-posed problem. Thus, SR module can capture the tiny LP without hallucination and it implies the proposed architecture has an impact on reducing the false negatives.



Figure 6: Example in AOLP dataset (Hsu et al., 2013). Poor-resolution and background clutter are common challenging issues on character recognition problem.

5 CONCLUSIONS

In this paper, we propose a new method based on GAN to recognize characters in unconstrained license plates. We design a novel network to directly generate a clear SR image from a blurry small one, and our up-sampling sub-network and reconstruction sub-network are trained in an end-to-end way. Moreover, we introduce an extra classification branch to the discriminator network, which can distinguish the HR/SR and the character counting probability simultaneously. Furthermore, the adversarial loss brings to generator network to restore a clearer SR image. Our experiment on AOLP and GIST-LP datasets demonstrate the substantial improvements, when compared to previous state-of-the-art methods.

REFERENCES

- Anagnostopoulos, C. N. E., Anagnostopoulos, I. E., Loumos, V., and Kayafas, E. (2006). A license plate-recognition algorithm for intelligent transportation system applications. *IEEE Transactions on Intelligent transportation systems*, 7(3):377–392.
- Cheang, T. K., Chong, Y. S., and Tay, Y. H. (2017). Segmentation-free vehicle license plate recognition using convnet-rnn. *arXiv preprint arXiv:1701.06439*.
- Chen, Z.-X., Liu, C.-Y., Chang, F.-L., and Wang, G.-Y. (2009). Automatic license-plate location and recognition based on feature saliency. *IEEE transactions on vehicular technology*, 58(7):3781.
- Chinomi, K., Nitta, N., Ito, Y., and Babaguchi, N. (2008). Prsurv: privacy protected video surveillance system using adaptive visual abstraction. In *International Conference on Multimedia Modeling*, pages 144–154. Springer.
- Chou, C.-J., Chien, J.-T., and Chen, H.-T. (2017). Self adversarial training for human pose estimation. *arXiv preprint arXiv:1707.02439*.
- Dong, C., Loy, C. C., He, K., and Tang, X. (2014). Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*, pages 184–199. Springer.
- Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., and Greenspan, H. (2018). Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *arXiv preprint arXiv:1803.01229*.
- Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- Hsu, G.-S., Chen, J.-C., and Chung, Y.-Z. (2013). Application-oriented license plate recognition. *IEEE transactions on vehicular technology*, 62(2):552–561.
- Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.
- Jiao, J., Ye, Q., and Huang, Q. (2009). A configurable method for multi-style license plate recognition. *Pattern Recognition*, 42(3):358–369.
- Kim, K. K., Kim, K., Kim, J., and Kim, H. J. (2000). Learning-based approach for license plate recognition. In *Neural Networks for Signal Processing X, 2000. Proceedings of the 2000 IEEE Signal Processing Society Workshop*, volume 2, pages 614–623. IEEE.
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Laroca, R., Severo, E., Zanlorensi, L. A., Oliveira, L. S., Gonçalves, G. R., Schwartz, W. R., and Menotti, D. (2018). A robust real-time automatic license plate recognition based on the YOLO detector. *CoRR*, abs/1802.09567.
- Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4681–4690.
- Lee, Y., Yu, J., and Jeon, M. (2017). Automatic part localization using 3d cuboid box for vehicle subcategory recognition. In *Control, Automation and Inform-*

- mation Sciences (ICCAIS), 2017 International Conference on*, pages 175–180. IEEE.
- Lee, Y., Yun, J., Hong, Y., Lee, J., and Jeon, M. (2018). Accurate license plate recognition and super-resolution using a generative adversarial networks on traffic surveillance video. In *2018 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, pages 1–4. IEEE.
- Li, H. and Shen, C. (2016). Reading car license plates using deep convolutional neural networks and lstms. *arXiv preprint arXiv:1601.05610*.
- Li, Y., Wang, N., Liu, J., and Hou, X. (2017). Demystifying neural style transfer. *arXiv preprint arXiv:1701.01036*.
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer.
- Nguyen, A., Bengio, Y., and Dosovitskiy, A. Plug & play generative networks: Conditional iterative generation of images in latent space.
- Noh, S., Shim, D., and Jeon, M. (2016). Adaptive sliding-window strategy for vehicle detection in highway environments. *IEEE Transactions on Intelligent Transportation Systems*, 17(2):323–335.
- Pascanu, R., Mikolov, T., and Bengio, Y. (2013). On the difficulty of training recurrent neural networks. In *International Conference on Machine Learning*, pages 1310–1318.
- Peng, X., Tang, Z., Yang, F., Feris, R., and Metaxas, D. (2018). Jointly optimize data augmentation and network training: Adversarial data augmentation in human pose estimation. *arXiv preprint arXiv:1805.09707*.
- Pu, J., Liu, S., Ding, Y., Qu, H., and Ni, L. (2013). T-watcher: A new visual analytic system for effective traffic surveillance. In *Mobile Data Management (MDM), 2013 IEEE 14th International Conference on*, volume 1, pages 127–136. IEEE.
- Radford, A., Metz, L., and Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.
- Rajeswar, S., Subramanian, S., Dutil, F., Pal, C., and Courville, A. (2017). Adversarial generation of natural language. *arXiv preprint arXiv:1705.10929*.
- Redmon, J. and Farhadi, A. (2017). Yolo9000: better, faster, stronger. *arXiv preprint*.
- Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.
- Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252.
- Shi, X., Zhao, W., and Shen, Y. (2005). Automatic license plate recognition system based on color image processing. In *International Conference on Computational Science and Its Applications*, pages 1159–1168. Springer.
- Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Smith, R. (2007). An overview of the tesseract ocr engine. In *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*, volume 2, pages 629–633. IEEE.
- Song, Y.-m. and Jeon, M. (2016). Online multiple object tracking with the hierarchically adopted gm-phd filter using motion and appearance. In *Consumer Electronics-Asia (ICCE-Asia), IEEE International Conference on*, pages 1–4. IEEE.
- Wang, S.-Z. and Lee, H.-J. (2003). Detection and recognition of license plate characters with different appearances. In *Intelligent Transportation Systems, 2003. Proceedings. 2003 IEEE*, volume 2, pages 979–984. IEEE.
- Wang, X., Man, Z., You, M., and Shen, C. Adversarial generation of training examples: Applications to moving vehicle license plate recognition.
- Yoon, Y.-c., Boragule, A., Yoon, K., and Jeon, M. (2018). Online multi-object tracking with historical appearance matching and scene adaptive detection filtering. *arXiv preprint arXiv:1805.10916*.
- Zhang, H., Jia, W., He, X., and Wu, Q. (2006). Learning-based license plate detection using global and local features. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 2, pages 1102–1105. IEEE.
- Zhang, J., Wang, F.-Y., Wang, K., Lin, W.-H., Xu, X., Chen, C., et al. (2011). Data-driven intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 12(4):1624–1639.
- Zherzdev, S. and Gruzdev, A. (2018). Lprnet: License plate recognition via deep neural networks. *arXiv preprint arXiv:1806.10447*.
- Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint arXiv:1703.10593*.