Joint Image Deblurring and Binarization for License Plate Images using Deep Generative Adversarial Networks

Van-Giang Nguyen
Faculty of Information Technology,
Le Quy Don Technical University,
Hanoi, Vietnam
giangnv@mta.edu.vn

Duy Long Nguyen
Faculty of Information Technology,
Le Quy Don Technical University,
Hanoi, Vietnam
dlongmta@gmail.com

Abstract—**Image deblurring is a highly ill-posed inverse problem where it aims to estimate the sharp image from blurred image with or without the knowledge about the blurring process. Despite the success of model-based image deblurring methods where the deconvolution is a major step to recover the sharp image, its usage in practice is still limited, especially when many factors such as object motion, camera motion, non-uniform sensitivity of the imaging device contribute to imaging process. In automatic license plate recognition (ALPR) of moving vehicle, the blurred image severely reduces the accuracy of recognition. Meanwhile, though the binarized image of license plate has an important role in ALPR systems, its accuracy is largely affected by the blurred image. In this paper, we use a deep architecture based on Generative Adversarial Networks to jointly perform image deblurring and image binarization for license plate images. Our model directly maps from blurred image to binary image without going through the deblurring as in conventional method. The proposed method is benefited from the fact that the ground-truth, sharp license plates are difficult to acquire for moving object, while the accurate binary images can be manually derived from blurred ones.**

Keywords—Image Deblurring, Inverse Problems, License Plate Deblurring, Generative Adversarial Network (GAN).

I. Introduction

Motion blur is a fundamental problem in imaging acquisition. It shows up when the object and/or the imaging device move. The blur can also be resulted from the defocus of camera device, haze, fog, and the low-light environment [1, 2]. In vision-based traffic surveillance, the blurred image result in in-accurate results in automatic license plate recognition (ALPR). In the cases of weak light (from late afternoon to early morning of the next day), vehicle moving in high speed and/or degraded license plate, blurring might blinds ALPR systems.

There have been attempts to solve the motion deblur for a single image [2, 3] (video deblurring or deblurring for an image sequence is not considered in this paper). The existing methods can be grouped into three categories: (i) methods that based on the assumption of static scene and uniform (space-invariant) blur kernel; (ii) methods that based on the assumption of static scene and non-uniform (space-variant) blur kernel (which can be considered to equivalent to static camera and a moving object); (iii) methods that do not rely on kernel estimation.

In model-based image deblurring methods (categories (i) and (ii)), the blurring process is modeled as the combination of convolution ($\otimes$) and noise addition as follows:

$$\mathbf{b} = \mathbf{s} \otimes \mathbf{k} + \mathbf{n} \qquad (1)$$

where $\mathbf{b}, \mathbf{s}, \mathbf{k}, \mathbf{n}$ are blurred image, underlying sharp image, blur kernel and noise, respectively.

However, due to the difficulties in finding the blur kernel $\mathbf{k}$, the mode-based methods do not work well in practice when the image gets contribution from many blurring factors. In fact, the blurred image is not only resulted from the convolution of image with blur kernel, but also of other complex mechanics that are not easy to model with the blur kernel only.

To overcome the difficulties and limitations of kernel-based method, especially in case of non-uniform deblurring, attempts have been made to learning-based methods [4-10]. In [4], the image is divided into overlapping patches. Then the deblurring (including blur kernel estimation and latent image estimation) is performed for each patch. The deblurred patches are then added back later to form the final deblurred image. In [7], convolutional neural network (CNN) is trained on a large dataset consists of image pairs, each has a sharp image and an artificially generated, blurred image. In [8], the spatially-varying motion blur kernel is estimated using a CNN. In particular, the probabilities of different motions kernels are first estimated for each image patch. Then, the motion blur kernel for the whole image is estimated from the sub-region kernels. That global motion kernel is used in deconvolution to get the deblurred image. Recently, in [6], a multi-scale CNN is trained in a direct manner with the addition of adversarial loss. Then it is later used to directly estimate the sharp image from blurred image.

In this work, inspired by the works in [5, 9, 11], we treat image deblurring as a special case of image-to-image translation problem. In particular, a generative adversarial network based deep architecture is used to directly estimate the deblurred image from blurred image without the need of estimating the blur kernel.

978-1-5386-7983-8/18/$31.00 ©2018 IEEE

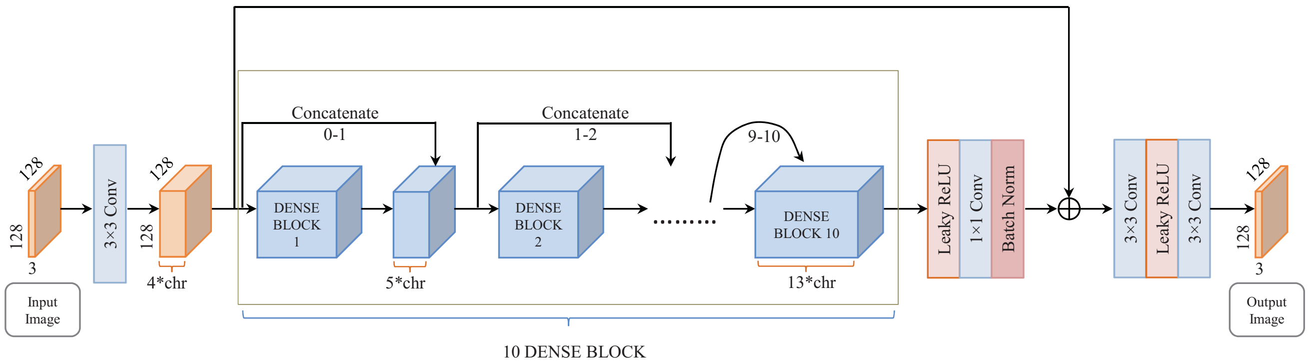


Fig. 1. Network architecture of generator.

An ALPR system usually has four stages [12]: (i) image acquisition; (ii) license plate extraction; (iii) license plate segmentation; (iv) character recognition. Among those stages, binarization has an important role in processing image. It could be used either in indicating the corner of plate (before image rectification/tilting onto a common image plane), in license plate segmentation or in extracting characters from image. In particular, the extracted license plate region from stage (ii) might have some problems such as tilt and non-uniform brightness, irregular texture in the plate region. When the vehicle is moving and/or the camera is a moderate one, the license plate region is highly blurred and it makes the segmentation more difficult.

To alleviate the above-mentioned issue, we propose to jointly perform image deblurring and binarization for license plate images. On the one hand, we improve the accuracy of license plate segmentation with the guide of deblurring. On the other hand, we take advantage of an image-to-image translation architecture, which has been successfully applied to other imaging problems such as super-resolution, style-transfer and image synthesis [5], to capture the complex nature of deblurring and binarization.

The existence of highly accurate binary license image (though manually generated) for each blurred image ensures a perfect pair for training the system. While the previous work relies on image synthesis to generate the blurred image from ground-truth one [9, 11], here we tackle in a opposite way where we already have the blurred image and we generate the deblurred one in favor of human visual system. Furthermore, we consider a particular application where the outcome is a deblurred and binarized image rather than a deblurred image only.

The remainder of this paper is organized as follows. Section II presents the model architecture of the adversarial network for deblurring and binarization. Section III presents method to generate blur/sharp pairs, followed by our experimental studies in Section IV. Section V discusses and concludes.

II. Methods

In [13], Goodfellow *et al* introduced a ground-breaking framework to estimate generative models via an adversarial process. It consists of two models that are trained together: The generative model learns to map from random noise to generate sample, while the discriminative model tries to discriminate whether the sample came from the real training data or from the generative model. Since both models are multilayer perceptrons, the adversarial modeling framework is denoted as Generative Adversarial Networks (GANs).

In imaging tasks, GANs is the game of two competitive networks: discriminator and generator. The *discriminator* tries to distinguish if an image is real (from training data) of fake (from generator) while the *generator* tries to generate the image that as close to the real image as possible. In most imaging cases, GANs are extended in conditional settings where it learn a conditional generative model which is a mapping from both observed image and noise to generate output image. [5]

In image deblurring, the main target is to recover the sharp image I^S with the only given information from blurred image I^B and without any information about blur kernel. The GANs for image deblurring train a convolutional neural network G, or generator, whose aim is to generate the (sharp) image I^S for each given input (blurred) image I^B. It also include another convolutional neural network *D*, named discriminator, which judge the performance of the estimated image I^S. The task of generator is to generate an image that leads the discriminator to believe it is a real image while the discriminator evaluates the performance of estimated image with respect to some given ground-truth images.

In the following sub-sections, we will give details about network architectures of generator and discriminator of a popular deblurring GANs [6, 9, 11]. The cost functions for these networks are also presented.

A. *Generator Network*

The architecture of generator is summarized in Fig. 1 and it works as follows:

(i) The input (blurred) image is first convolved with a 3×3 convolutional layer.

(ii) The result from step (i) is going through a dense field which consists of 10 dense blocks. These blocks are placed

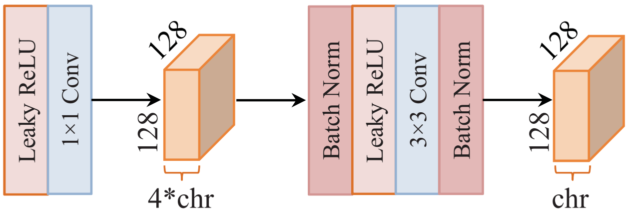


Fig. 2. A dense block used in Generator.

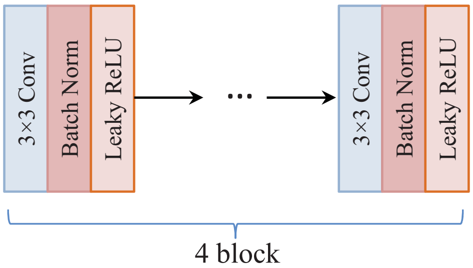


Fig. 3. Network architecture of discriminator.

sequentially after each other. The result of each dense block is concatenated with the output of its previous layer. The structure of a dense block is given in Fig. 2 and includes (in chronological order): a leaky rectified linear unit (LeakyReLU), a 1×1 convolution, a batch normalization, a LeakyReLU, a 3×3 convolution, and finally, a batch normalization.

(iii) The result from step (ii) is then going through the Tail part which consists of a LeakyReLU, a 1×1 convolutional layer followed by a batch normalization.

(iv) Finally, the global skip connection includes a concatenation of the output from (i) with the output from (iii) to improve the performance of generation.

B. Discriminator Network

For discriminator, we use the well-known convolutional PatchGANs (Markovian patch discriminator) classifier that penalizes structure at the scale of patches [14, 5]. In particular, it aims to classify if patches in image are real or fake. The averaging of response from each patch is used to calculate the output of discriminator. The network architecture for processing each patch in discriminator includes 4 convolutional layers and is illustrated in Fig. 3. (The number 4 is chosen empirically since it gives reasonable result in our tests)

C. Cost Functions

The cost function for discriminator in GANs is the standard cross-entropy when training a standard binary classifier with sigmoid output [13]. The classifier is trained with two batches of data, one from the real dataset (ground-truth image) with label 1 and the other from the generator with label 0. Its objective function, a.k.a. adversarial loss, is given by

Fig. 4. Sample license plate images from the same camera.

$$\begin{aligned} L_{adv} &= \mathrm{E}_{I^S \sim p_{sharp}(I^S)}[\log D_{\theta_D}(I^S)] \\ &+ \mathrm{E}_{I^B \sim p_{blurry}(I^B)}[\log(1 - D_{\theta_D}(G_{\theta_G}(I^B))] \end{aligned} \tag{2}$$

where D and G, are generator and discriminator, respectively, with its corresponding parameters denoted by θ_D, θ_G .

For generator in image deblurring, in addition to adversarial loss above, its cost function also includes: (i) content loss and (ii) perceptual loss. For the content loss, ℓ_1 loss between the blurred image and the ground-truth sharp image is used. Here the content loss is denoted by L_{ℓ_1} .

The perceptual loss, which is defined on high-level features extracted from pre-trained networks, helps to generate better image [15, 6] and is given by:

$$L_{percept} = \frac{1}{W_{i,j} H_{i,j}} \times \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left(\phi_{i,j}(I^S)_{x,y} - \phi_{i,j}(G_{\theta_G}(I^B))_{x,y} \right)^2 \tag{3}$$

where $W_{i,j}, H_{i,j}$ are the width and height of the $(i,j)^{th}$ ReLU layer of VGG16 network and $\phi_{i,j}$ is the feature map obtained by the j^{th} convolution (after activation) before the i^{th} max-pooling layer within the VGG16 network [16]. In this work, we choose *i=3* and *j=3*.

The overall cost function to be minimized is

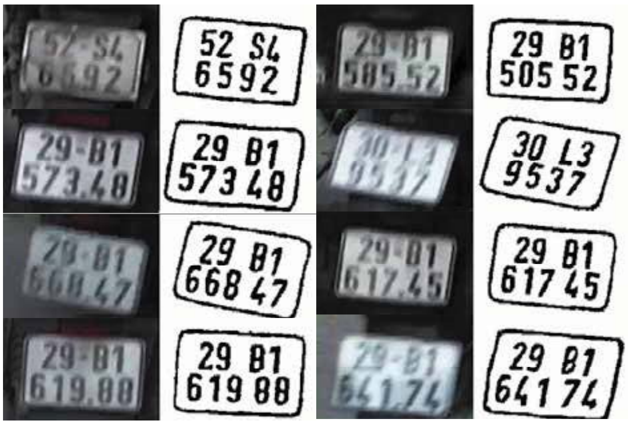
$$L_{total} = L_{adv} + K_1 L_{percept} + K_2 L_{\ell_1} \tag{4}$$

where K_1, K_2 are hyper-parameters. The Adam solver [17] is used to find minimization of the above optimization problem (in an alternating way for generator and discriminator).

III. GENERATING DATA

One of the key issues in image deblurring is the available of datasets for training. One can consider to acquire the ground-truth, sharp images using high speed camera [6] in which a high frame-per-second video was acquired. Then, consecutive frames were average to simulate photos taken at longer shutter-time. Though this method provides realistic blurred images, it is not easy to do in the absence of a high-end camera.

(a)



(b)

Fig. 5. Some blurred/sharp image pairs for training (first and third columns are for blurred images; second and fourth columns are for sharp images): (a) Dataset #1; (b) Dataset #2.

One can also consider to acquire ground-truth, sharp image (with short shutter-time) by camera and then convolves it with well-known blur kernels in the field to generate blurred images [2, 7, 11].

In ALPR system, especially in the tropical and developing countries, the license plate images are contributed by many factors such as the irregular movement of vehicle, motion, changing light, non-uniform license plate plane (See Fig. 4 for some samples). These difficulties prevents the acquisition of blurred/sharp image pair for training ALPR system. Furthermore, in our opinion, each ALPR system should be trained for a particular circumstance (and working condition) so that it can minimize the errors and reduce the computation time.

In this paper, we proposed to acquire the blurred license plate images from an existing ALPR system (with a moderate camera). Those images are then manually binarized by human. The resulting binarized images are free from blurring and then used as ground-truth images. (Fig. 5).

Note that, in predicting stage, given a blurred image, the generator predicts the intermediate deblurred image. This intermediate deblurred image is then binarized using the well-known Otsu's method [18] to generate the final binarized image.

With the ground truth (binarized and blur-free) image pair feeding to the deblurring GAN networks described in Section II, we expect the network will treat it as image-to-image translation where the expert knowledge has been added in making ground truth image. The network will do two things at a time (deblurring and binarization) in a jointly manner which is unique for ALPR systems.

IV. Results

A. Experimental Settings

We implemented our model with Keras library and performed the studies on a PC with Intel i7 processor with 16GB of RAM. The batch size for ADAM solver was set to 4. For hyper-parameters, K_1 and K_2 were chosen experimentally to achieve visually reasonable results and in this work, $K_1 = 145$ and $K_2 = 170$ (Optimizing these parameters are beyond the scope of this paper).

We trained the system with two datasets for two different applications. Dataset #1 consists of 500 images with the size of 128×128 (See Fig. 5(a)). All images in this dataset had been tilted onto a common plane in advanced. The bounder of license plate was also removed. This dataset was used to validate the performance of the proposed with respect to the image region having characters only.

Meanwhile, dataset #2 consists of 80 images with the size of 128×192 (See Fig. 5(b)). The small number of sample images used here was on purpose to see the power of GANs on such difficult task with limited data. These images were extracted from license plate extraction step. The binarized, sharp images for this dataset were designed to include the border of license plate, that is important information for tilting the image to standard plane.

To evaluate the performance of the proposed method, comparison is made with Otsu's method made on the blurred image.

B. Experimental Results

Fig. 6 shows some deblurred, binarized results for dataset #1 where the proposed method accurately binarizes the image in different conditions of blurring and lighting.

For realistic case of dataset #2, the test results on training data in Fig. 7 reveal the capability of the proposed method where it can sense and recover not only the character(s), but also the border(s) of the license plate region. These information are helpful for later processing of ALPR systems.

When testing the method on more severe images which were not used in training, the result is still promising as shown in Fig. 8. For moderately blurred images, the generator predicts quite good. However, for 'bad' ones (such as two images in the 2nd row from bottom of Fig. 8.), the generator could not predict accurately. When the number of training samples is increased (currently, for dataset #2, 80 samples were used), the

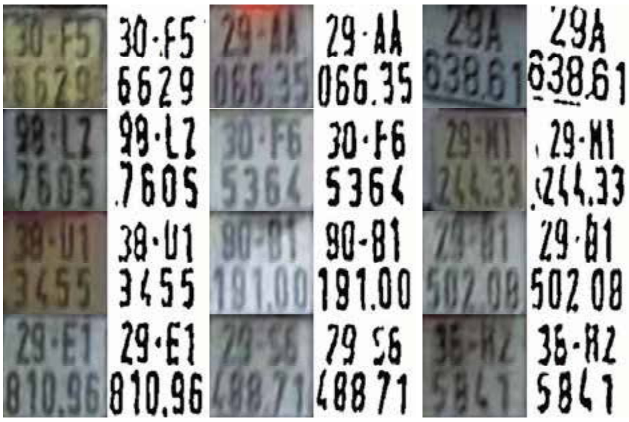


Fig. 6. Test results on dataset #1 (first, third and fifth columns are for blurred images; second, fourth and sixth columns are for deblurred, binarized images).

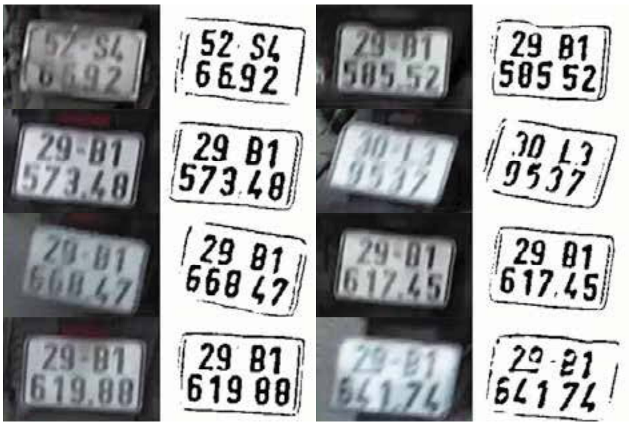


Fig. 7. Test results on (training part of) dataset #2 (first and third columns are for blurred images; second and fourth columns are for deblurred, binarized images).

performance of generator is expected to improve significantly. However, in that case, the training time will be an issue.

The visual comparison and numerical comparison (via root-mean-square error) of the proposed method with the result from binarizing directly the blurred images by using Otsu's method (not included here) also show that the proposed method is superior to the ordinary binarization of the blurred image. Finally, note that while the proposed method achieves more accurate result (than the conventional binarization method), it requires substantially more time to complete the computation. In our test, the runtime for the proposed method was about 0.25 second per image.

V. Discussions and Conclusions

We have presented a method to jointly deblur and generate binary images for blurred license plate images. The method is benefited from a deep generative adversarial network which efficiently learns a map from blurred image to binary image. The proposed method has proven to work well under different blurring conditions, lighting conditions and different irregular textures on license plate regions. It can be trained to recover not only the characters but also the borders of license plate

Fig. 8. Test results on (testing part of) dataset #2 (first and third columns are for blurred images; second and fourth columns are for deblurred, binarized images).

image, which consequently can improve the performance of ALPR systems.

Future work includes the evaluation of the proposed method via qualitative test on license plate recognition. More works are also needed to do with optimizing the network architecture, the number of dense blocks in generator, the number of blocks in discriminator and pruning the generator network for each particular application.

References

[1] P.C. Hansen, J.G. Nagy and D.O. O'Leary, *Deblurring images - Matrices, Spectra and Filtering*, Society for Industrial and Applied Mathematics, 2006.

[2] A.N. Rajagopalan and R. Chellappa, *Motion Debluring Algorithms and Systems*, Cambridge University Press, 2014.

[3] Ruxin Wang, Dacheng Tao, "Recent Progress in Image Deblurring", *arXiv:1409.6838*, 2014.

[4] S. Harmeling, H. Michael, and B. Scholkopf, "Space-variant single-image blind deconvolution for removing camera shake," *Advances in Neural Information Processing Systems*, pp. 829-837, 2010.

[5] P.Isola, J.-Y.Zhu, T.Zhou, and A.A. Efros, "Image-to-image translation with conditional adversarial networks," *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[6] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," *In IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[7] P. Svoboda, M. Hradis, L. Marsik and P. Zemcik, "CNN for License Plate Motion Deblurring," *IEEE ICIP* 2015.

[8] J. Sun, W. Cao, Z. Xu, and J. Ponce, "Learning a Convolutional Neural Network for Non-uniform Motion Blur Removal," *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[9] S. Ramakrishnan, S. Pachori, A. Gangopadhyay, S. Raman, "Deep Generative Filter for Motion Deblurring," *ICCV Workshop on Color and Photometry in Computer Vision*, 2017.

[10] M. Noroozi, P. Chandramouli, P. Favaro, "Motion Deblurring in the Wild," *German Conference on Pattern Recognition(GCPR)*, 2017.

[11] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin and J. Matas, "DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks," *arXiv:1711.07064*, 2018.

[12] S. Du, M. Ibrahim, M. Shehata and W. Badawy, "Automatic License Plate Recognition (ALPR): A State of the Art Review," *IEEE Trans. on Circ. and Syst. for Video Technology*, pp. 311-325, vol. 23, no. 2, 2013.

[13] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D.Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks", *Advances in Neural Information Processing Systems*, pp. 2672–2680, 2014.

[14] C. Li and M. Wand, "Precomputed Real-Time Texture Synthesis with Markovian Generative Adversarial Networks," *European Conference on Computer Vision*, 2016.

[15] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," *European Conference on Computer Vision*, 2016

[16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *International Conference on Learning Representations*, 2015.

[17] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *International Conference on Learning Representations*, 2015.

[18] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Sys. Man. Cyber.*, vol. 9, no. 1, pp. 62-66, 1979.