https://doi.org/10.2352/ISSN.2470-1173.2020.6.IRIACV-052

Perceptual License Plate Super-Resolution with CTC Loss

Zuzana Bílková, Michal Hradiš; Charles University in Prague, Faculty of Mathematics and Physics, Ke Karlovu 3, Praha 2, 121 16, Czech republic; The Czech Academy of Sciences, Institute of Information Theory and Automation, Pod Vodárenskou věží 4, Praha 8, 182 08, Czech Republic; Brno University of Technology, Faculty of Information Technology, Božetěchova 2/1, Brno, 612 00, Czech republic

Abstract

We present a novel method for super-resolution (SR) of license plate images based on an end-to-end convolutional neural networks (CNN) combining generative adversial networks (GANs) and optical character recognition (OCR). License plate SR systems play an important role in number of security applications such as improvement of road safety, traffic monitoring or surveillance. The specific task requires not only realistic-looking reconstructed images but it also needs to preserve the text information. Standard CNN SR and GANs fail to accomplish this requirment. The incorporation of the OCR pipeline into the method also allows training of the network without the need of ground truth high resolution data which enables easy training on real data with all the real image degradations including compression.

Context

This paper presents an automatic license plate super-resolution (SR) method based on generative adversial networks [1] (GANs) and optical character recognition (OCR). License plate SR systems play an important role in number of security applications such as improvement of road safety, traffic monitoring or surveillance. It is one of the key components of modern intelligent transportation systems.

With the advance of deep-learning techniques, especially GANs, there has been a rapid development in SR methods. However, standard SR and GAN techniques in the literature focus mainly on the reconstruction of natural images which typically do not work well for numbers and letters. The specific task of license plate super-resolution requires not only realistic-looking reconstructed images but it also needs to preserve the text information.

Related work

There are several methods of super-resolution using deep learning. Neural networks are trained with pairs of low-resolution and corresponding ground truth high-resolution images minimizing a defined distance between them, such as in [2] or [3].

The first deep learning method for super-resolution to use GANs is a SRGAN [4]. The network incorporates perceptual loss into its loss function. Perceptual loss uses a pre-trained VGG network [5] to compute a distance between the VGG feature representations of a reconstructed image and reference high-resolution image which leads to promising results on natural images. However, the reconstructed images lack fine details for the applliccation of license plates SR.

The first SR method using GANs focusing on super-resolution of text is TextSR [6]. Like SRGAN, TextSR uses a perceptual loss; it is computed with a recognition network ASTER [7] for feature representations. Similar to all of the other mentioned SR methods, TextSR needs a ground truth high-resolution image.

License plate SR methods use traditional and deep learning approaches. A multi-frame SR using geometric K-NN SR is solved in [8]. Semantic information of the characters of license plates for SR in extremely low-resolution images is studied in [9]. Adversarial SR and one-stage character segmentation and recognition is used in [10]. MTGAN [11] is a method for LP SR and recognition using GANs with a special discriminator trained to judge whether the license plate is high-definition enough to be correctly recognized.

Objective

The purpose of our project is to develop an automatic method for super-resolution of images of license plates. We explore the state-of-the-art generative adversial networks which are used in many areas of image processing, including SR. However, the high resolution images generated directly from pure GANs lack the fine details desired for the preservation of the individual numbers an letters on license plates. We thus propose a new pipeline consisting of OCR neural network which guides the training of the generator network to produce high resolution images with the true text and it also allows training of the method without the need of the ground truth high resolution data. This advantage enables easy training on real data with all the real image degradations including compression.

Method

Our method is an end-to-end neural convolutional network combining the generative adversial networks and optical character recognition to output high resolution images preserving the license plate text information. The generator network takes as input up-sampled low-resolution images and produces high-resolution candidates. Discriminator is used to distinguish the generated images from the true data distribution. Furthermore, we introduce a third neural network that performs OCR using connectionist temporal classification (CTC). Incorporation of the CTC loss in the generator network forces the generator to produce content-aware images and enables training without the need of ground truth high-resolution data.

Generator network

The architecture of our generator network is inspired by the convolutional neural network U-net [12]. As shown in Figure 1., an input image, low-resolution image up-sampled by bilinear interpolation, goes through five down-sampling blocks. Each block consists of two convolutional layers followed by a max-pooling layer. Another convolutional layer is followed by five up-sampling blocks performing two convolutions, resizing and concatenating with the output of corresponding down-sampling block. Two other convolutional layers then yeild the resulting high-resolution image. All the convolutional layers use leaky ReLu activation function.

To train our model we use Wasserstein loss, first introduced for WGANs in [13]. This loss function ensures smoother gradients and thus improves stability of learning. The loss function for the generator is defined as follows:

$$l_G = \frac{1}{m}\sum_{i=1}^{m}[f(G(z_i))] + \lambda l_{CTC}, \qquad (1)$$

where $G(z)$ is the generator's output for a given input z, $f(G(z))$ is the discriminator's output for a fake instance, l_{CTC} is the CTC loss described in (3) and λ is a parameter.

Discriminator network

Discriminator with Wasserstein loss is usually called a ”critic” because it does not actually classify instances, instead, it outputs a scalar score rather than a probability. This score can be interpreted as how real the input images are. The formulas derive from the Wasserstein distance between the real and generated distributions. The discriminator function is demanded to be Lipschitz continuous. In the original article [13] authors propose weight clipping to enforce a Lipschitz constraint which still suffers from unstable training. We will use an improvement suggested in [14], precisely replacing weight clipping with gradient penalty.

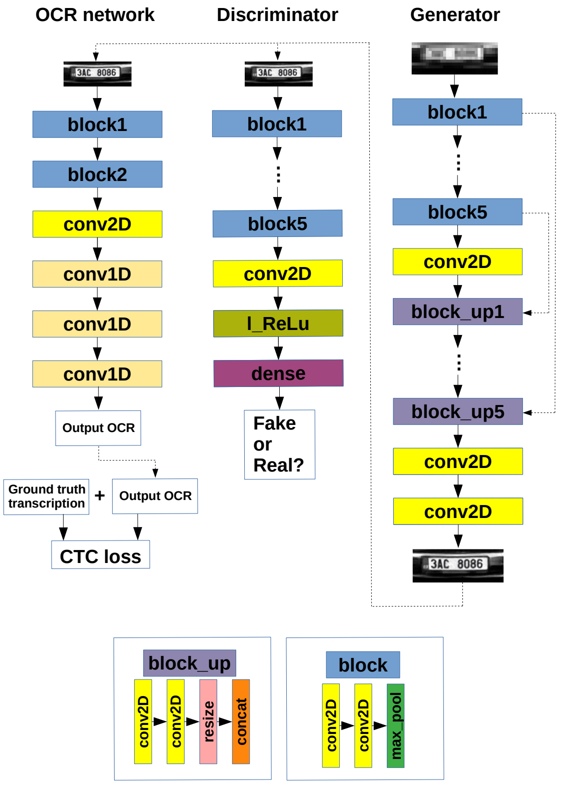


Figure 1. *The structure of the proposed method based on GAN with a recognition network.*

The discriminator's loss function l_D is calculated as follows:

$$l_D = \frac{1}{m}\sum_{i=1}^{m}[f(x_i) - f(G(z_i))] + GP, \qquad (2)$$

where $f(x)$ is the discriminator's output for a real instance, $f(G(z))$ is the discriminator's output for a generated instance and GP is the gradient penalty enforcing a Lipschitz constraint.

The architecture of our discriminator network is shown in Figure 1. The input is either real or generated image which then goes through five down-sampling blocks consisting of two convolutional layers and a max-pooling layer. Another convolutional layer is followed by the final dense layer.

OCR network

To improve the results of the generator network we propose to incorporate into its loss function a CTC loss controlling the generated text of a license plate. We use an optical character recognition network for a recognition of the text from the generated image. The output of the OCR pipeline is then used with the ground truth transcription to compute the CTC loss.

Our OCR network is a convolutional neural network with two blocks of two convolutional layers with a max-pooling layer folowed by other four convolutional layers with the final layer outputting the transcription, see Figure 1. All convolutional layers use leaky ReLU activation function.

The CTC loss function l_{CTC} used for trainig of the OCR pipeline, and for computing a part of the generator's loss, is defined as follows:

$$l_{CTC} = \sum_{(z,y)\in T} -\log p\left(y|G(z)\right), \qquad (3)$$

where $G(z)$ is the generator's output for a given input z, y is its corresponding transcription, T is the training data and p is a probability for a single training pair. The CTC objective for a single pair (g,y) is computed as:

$$p\left(y|g\right) = \sum_{A\in\mathscr{A}_{g,y}} \prod_{s=1}^{S} p_s\left(a_s|g\right), \qquad (4)$$

where A is a possible alignment of the text, $\mathscr{A}_{g,y}$ is a set of valid alignments and p_s computes the probability for a single alignment step-by-step.

Results

The proposed neural networks are trained on a dataset of more than 15 000 train images. The performance of the method is evaluated on more than 4000 test images. We add noise to the original images of size 240x96 pixels which are subsequently down-sampled to the size of 30x12 pixels. Figures 2. and 3. show a sample of the resulting pictures of our method and pure GAN architecture without OCR, see the Figures' description for details. Both methods were trained on the same dataset with the same

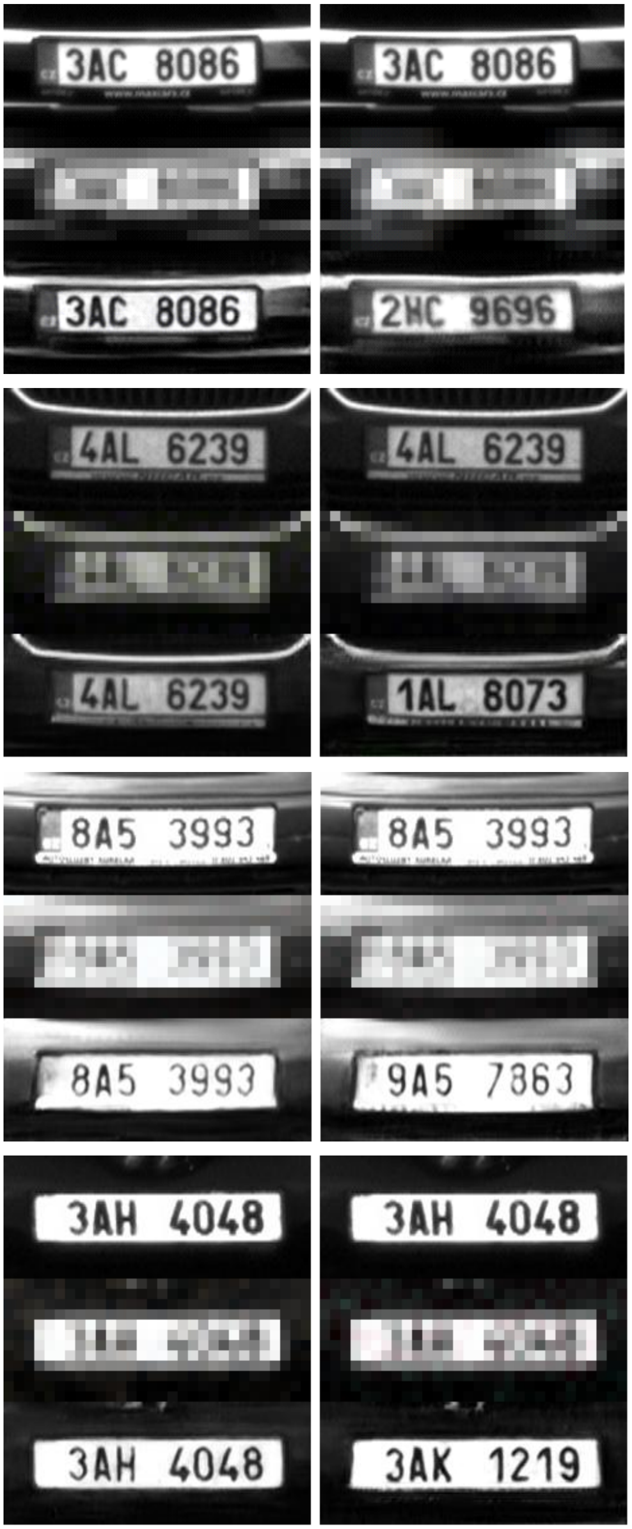


Figure 2. *Four examples of the outputs of our network on the left side of the pairs of the images and the outputs of pure GAN on the right side. First rows are original images, second rows show up-sampled noisy low-resolution images and the last rows present the outputs.*

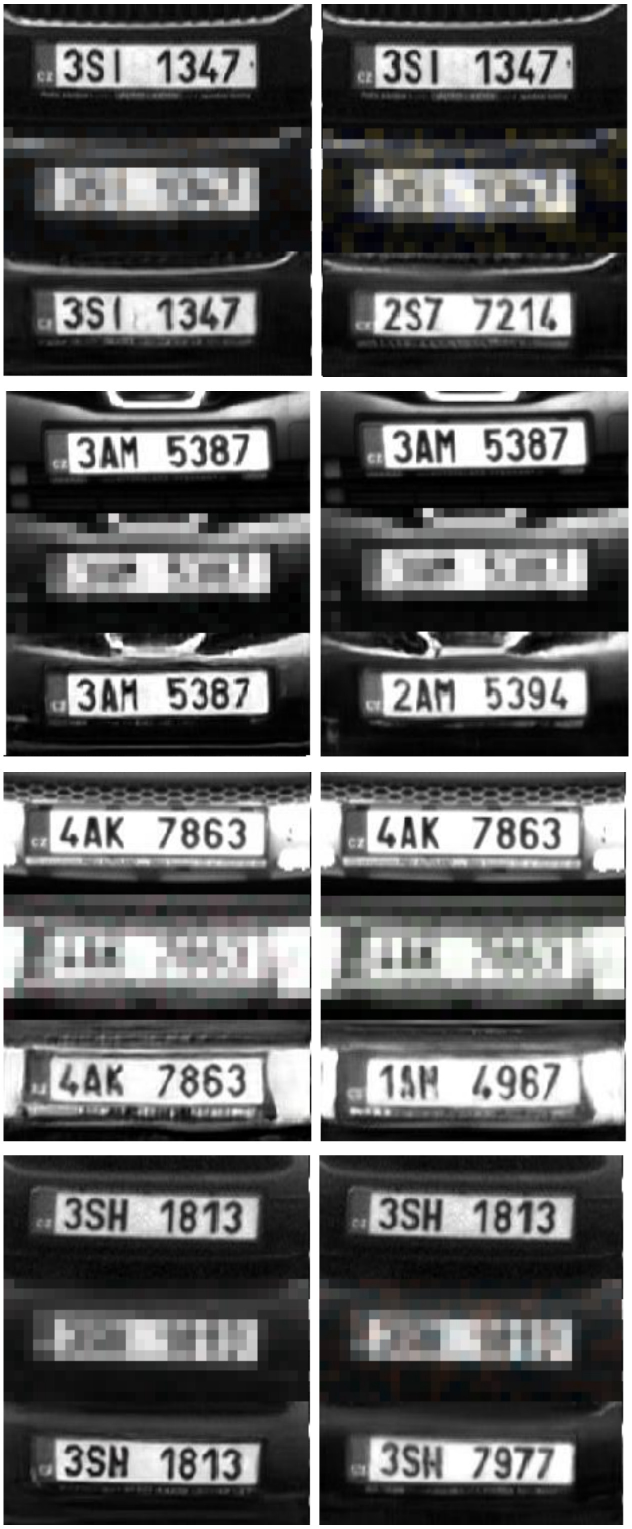


Figure 3. *Four examples of the outputs of our network on the left side of the pairs of the images and the outputs of pure GAN on the right side. First rows are original images, second rows show up-sampled noisy low-resolution images and the last rows present the outputs.*

number of epochs, i.e. 150 000. We can see that both methods achieve good quality high resolution images for inputs with different illumination conditions but the outputs of the pure GAN does not preserve the text information which is crucial in our task of SR of license plates.

Table 1: Mean Character Error Rate (CER) and Mean CTC loss for our method (GAN with CTC) and pure GAN

Method	CER	CTC loss
GAN with CTC	0.07	2.5
Pure GAN	0.7	59.5

Table 1 shows mean values of character error rate (CER) and CTC loss over the test dataset. We can see that the large differences in values of both metrics correspond with the fact that pure GANs are not able to reconstruct the true text information.

Novelty and Future Work

The presented method for super-resolution of license plates images combines an end-to-end architecture of GANs and OCR with the CTC loss. Unlike the classic SR our task faces a challenge of keeping the true numbers of license plates. The GANs in our method generate quality high resolution images and the incorporation of CTC loss ensures preservation of the text information of license plates and enables training of the method without the need of the ground truth high resolution data. This is in contrast to the standard GANs which produce realistic-looking high resolution images from the noisy down-sampled images but the text information does not correspond with the original image, thus showing that the popular pure GANs are inapplicable for our goal.

Future work will include training of the generator on a dataset with real degradations. We will also compare the results of our method with other CNN SR methods which use high resolution data for training (i.e. pixel loss and content loss) and we will incorporate mutliple inputs from a video sequence to output the high-resolution license plate image.

Acknowledgments

This work has been supported by the GAUK, grant No.1583117, the *Praemium Academiae*, and the grant SVV–2017–260452.

References

[1] 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, pg. 2672-2680. (2014).

[2] Chao Dong, Loy, C. C., He, K., Tang, X. Image super-resolution using deep convolutional networks, IEEE transactions on pattern analysis and machine intelligence, 38(2), pg. 295-307.(2015).

[3] Chao Dong, Chen Change Loy, Xiaoou Tang, Accelerating the super-resolution convolutional neural network, European conference on computer vision, Springer, Cham, pg. 391-407. (2016).

[4] Christian Ledig, Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... , Shi, W., Photo-realistic single image super-resolution using a generative adversarial network, In Proceedings of the IEEE conference on computer vision and pattern recognition, pg. 4681-4690.(2017).

[5] Karen Simonyan, Andrew Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556. (2014).

[6] Wenjia Wang, Xie, E., Sun, P., Wang, W., Tian, L., Shen, C., Luo, P., TextSR: Content-Aware Text Super-Resolution Guided by Recognition, arXiv preprint arXiv:1909.07113.(2019).

[7] Baoguang Shi, Yang, M., Wang, X., Lyu, P., Yao, C., Bai, X, Aster: An attentional scene text recognizer with flexible rectification, IEEE transactions on pattern analysis and machine intelligence, 41(9), pg. 2035-2048. (2018).

[8] Hilário Seibel, Siome Goldenstein, Anderson Rocha, Eyes on the target: Super-resolution and license-plate recognition in low-quality surveillance videos, IEEE access 5, pg. 20020-20035. (2017).

[9] Yuexian Zou, Wang, Y., Guan, W., Wang, W., Semantic Super-resolution for Extremely Low-resolution Vehicle License Plate, In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pg. 3772-3776. IEEE.(2019).

[10] Younkwan Lee, Jun, J., Hong, Y., Jeon, M., Practical License Plate Recognition in Unconstrained Surveillance Systems with Adversarial Super-Resolution, arXiv preprint arXiv:1910.04324.(2019).

[11] Minghui Zhang, Wu Liu, Huadong Ma, Joint License Plate Super-Resolution and Recognition in One Multi-Task Gan Framework, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pg. 1443-1447. (2018).

[12] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, U-net: Convolutional networks for biomedical image segmentation, International Conference on Medical image computing and computer-assisted intervention, Springer, Cham, pg. 234-241. (2015).

[13] Martin Arjovsky, Soumith Chintala, and Léon Bottou, Wasserstein gan, arXiv preprint arXiv:1701.07875. (2017).

[14] Ishaan Gulrajani, Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A. C., Improved training of wasserstein gans. In Advances in neural information processing systems, pg. 5767-5777. (2017).

Author Biography

Zuzana Bílková is a PhD. student at Charles University in Prague and she works in the Czech Academy of Sciences. She is oriented on applications of deep neural networks in image processing, especially in the area of medical imaging. She is a holder of two grants, one from Charles University on convolutional neural networks and the second from the Technology Agency of the Czech Republic.

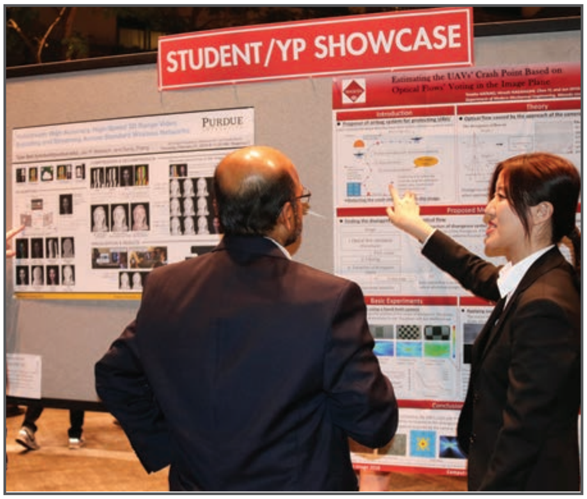
JOIN US AT THE NEXT EI!

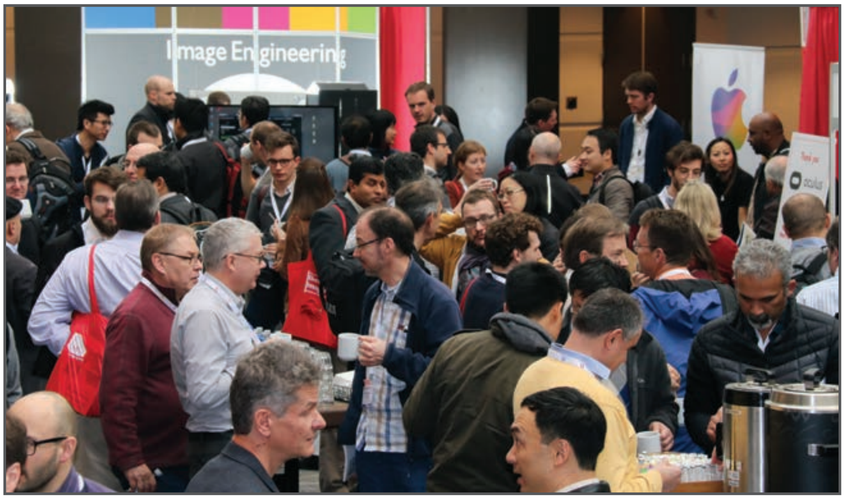
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





AR is a fundamental and transformative interface and platform.

20+ years from my initial working demo to today's AR displays

Optimistic that it won't take another 20 to reach ubiquitous consumer AR!

• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org