PlugNet: Degradation Aware Scene Text Recognition Supervised by a Pluggable Super-Resolution Unit

Yongqiang Mou[1]✉, Lei Tan[2]*, Hui Yang[1], Jingying Chen[2], Leyuan Liu[2], Rui Yan[1], and Yaohong Huang[1]

[1] AI-Labs, GuangZhou Image Data Technology Co., Ltd., China
yongqiang.mou@gmail.com, huiyang865@hotmail.com, reeyree@163.com, hyh362@me.com
[2] Nercel, Central China Normal University, China
lei.tan@mails.ccnu.edu.cn,{chenjy,lyliu}@mail.ccnu.edu.cn

1 Appendix

1.1 Failure Cases Analysis

Fig. 1. Some typical failure cases produced by PlugNet.

Although PlugNet reaches a satisfying performance in the text recognition issue, it still faces some nuisances. Figure. 1 shows some typical failure cases produced by the PlugNet. One of the remaining challenges is those occlusion texts like the image 'pioneer'. Of course, these types of error recognition will be significantly mitigated when using a lexicon. Vertical texts are also a tough case in the text recognition area. Especially, most of the methods (including PlugNet) using a horizontal template to resize the raw image before input which makes it much harder to recognize those vertical texts. Perhaps, this type of error

* Equal Contribution

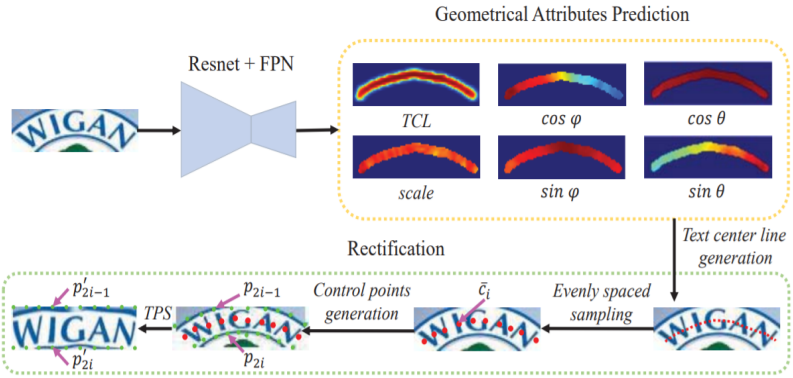

Fig. 2. Text rectification network of Yang et al.[3].

Table 1. Structure of Rectification Network in our paper. Herein, the 's' means the stride of each CNN layer and 'n' is the number of control points. In this paper, we set the n=20.

Layers	Output size	Configurations
Resize	32×64	–
CNN 1	16×32	$[2 \times 2conv, 32], s = 2$
CNN 2	8×16	$[2 \times 2conv, 64], s = 2$
CNN 3	4×8	$[2 \times 2conv, 128], s = 2$
CNN 4	2×4	$[2 \times 2conv, 256], s = 2$
CNN 5	1×2	$[2 \times 2conv, 256], s = 2$
FC1	1×512	–
FC1	$1 \times 2n$	–

could be solved by the bottom-up text recognition framework based on instance segmentation with the fully convolutional network. Some of the rare fonts may also make the PlugNet be confused like the image 'Angelo' for example. Since those extreme blur cases lose too much information, although having the PSU, PlugNet seems also inadequate when facing those texts. We think those extreme blur images may also challenging for humans. Some error labels also exist in those testing datasets like image 'Hosptal' which will decrease the final accuracy result.

In our work, the text rectification network (TRN) is attached to cope with those curved texts, but the TRN is training under weak supervision of the recognition loss. It makes PlugNet may not so efficient to solve extremely curved texts. Yang et al. [3] designed a Resnet-50 [1] based text rectification network as shown in Figure 2 and training under the dataset with character-level annotations, it helps them get better performance when compared to our work in the CUTE80 [2] dataset that contains many extremely curved texts with high-resolution. Ta-

ble. 1 shows the structure of our rectification network in detail. It is clear that our method seems more lite and reaches comparable results without character-level annotations.

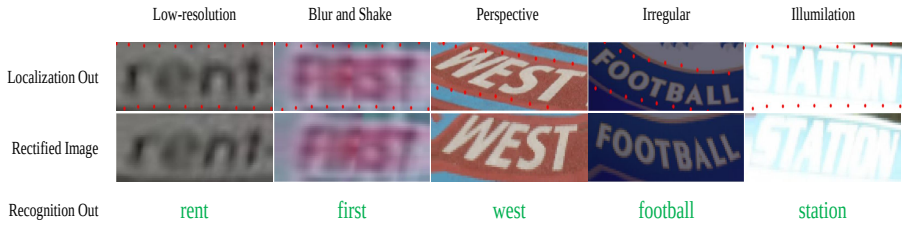

Fig. 3. PlugNet's recognition results of several typical challenges of text recognition issue.

Fig. 4. Some examples of text recognition results of PlugNet. From the left to right is the input image with control points, rectified image, ground truth, and recognition result.

1.2 Recognition Results of PlugNet

In this part, we choose several typical challenges among those testing dataset to validate the performance of PlugNet. As shown in Figure. 3, our proposed

PlugNet is able to recognize text from images while being robust to challenges such as low-resolution, blur shake, perspective, irregular, illumination, and etc. Additionally, we select some examples of text recognition results of PlugNet in Figure. 4 to demonstrate the superior performance of PlugNet.

References

1. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the CVPR. pp. 770–778 (2016)
2. Risnumawan, A., Shivakumara, P., Chan, C.S., Tan, C.L.: A robust arbitrary text detection system for natural scene images. Expert Systems with Applications **41**(18), 8027–8048 (2014)
3. Yang, M., Guan, Y., Liao, M., He, X., Bian, K., Bai, S., Yao, C., Bai, X.: Symmetry-constrained rectification network for scene text recognition. In: Proceedings of the ICCV (2019)