Report of ICPR MTWI Challenge 2

Team: nelslip(iflytek&ustc)
Yixing Zhu*, Jianshu Zhang*, Jun Du*, Lirong Dai*, Jiajia Wu†, Mingjun Chen† and Jinshui Hu†
*National Engineering Laboratory for Speech and Language Information Processing
University of Science and Technology of China, Hefei, Anhui, P. R. China
†iFLYTEK Research, Hefei, Anhui, P. R. China

I. Methodology

Our method is based on FPN [1], and we augmented the network as PANet [2] does. Because text line may be arbitrary quadrilateral, we use SLPR [3] to fit the outline of the text line, this method isometrically regress the points on the edge of text by using the vertically and horizontally sliding lines which can solve ambiguity of four vertices label, then we use its outline to generate a oriented rectangle for RoIRotate, so we adopt Cascade R-CNN as [4] with two steps, in the first step we propose a horizontal rectangle, in the second step we propose a oriented rectangle with SLPR, finally, we regress four quadrilateral vertices.

II. Experimental details

We use some tricks for better performance, including image rotation, multi-scale training & testing, when model assembling, we use ResNeXt-101 (32*8d) [5] pre-trained on ImageNet as backbone network and we only use a single model. All experiments were implemented in Caffe2 by using the NVIDIA GTX 1080Ti GPU.

III. Dateset

The official training dataset contains 10,000 images and we split them into 9,000 for training and the other 1,000 for validation. We did not use any extra data.

IV. Team introduction

The team leader is Jianshu Zhang, his email address: xysszjs@mail.ustc.edu.cn, telephone number: 15856910468, address: West campus of University of Science and Technology of China, No.8 apartment building, Room 640, No.443 HuangShan Road, Hefei, Anhui, 230027, China.

This team is a cooperation between National Engineering Laboratory for Speech and Language Information Processing in University of Science and Technology of China (USTC-NELSLIP) and iFLYTEK Research. We hope to have others' attention on our technology for addressing text line detection.

References

[1] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *CVPR*, vol. 1, no. 2, 2017, p. 4.

[2] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," *arXiv preprint arXiv:1803.01534*, 2018.

[3] Y. Zhu and J. Du, "Sliding line point regression for shape robust scene text detection," *arXiv preprint arXiv:1801.09969*, 2018.

[4] Z. Cai and N. Vasconcelos, "Cascade r-cnn: Delving into high quality object detection," *arXiv preprint arXiv:1712.00726*, 2017.

[5] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*. IEEE, 2017, pp. 5987–5995.