TCATD: Text Contour Attention for Scene Text Detection

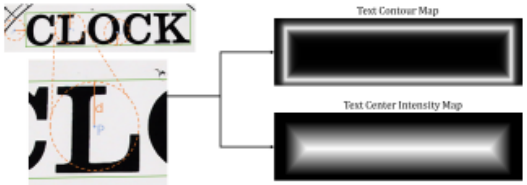
ZiLing Hu, Xingjiao Wu, Jing Yang
East China Normal University, Shanghai, China

The Problem

Segmentation-based approaches have enabled state of-the-art performance in long or curved text detection tasks. However, false detection still is a challenge when two text instances are close to each other.

We introduce the Text Contour Map and Text Center Intensity Map which can help model learn more robust features and improve the effect of text detection.

Approach – Details



The define the distance of pixel p:

$$d_p = \begin{cases} 0, & \text{if } p \in S_{border} \\ \min(euler(p,q)), & \text{if } q \notin S_{border} \text{ and } \exists q \in S_{border} \end{cases}$$

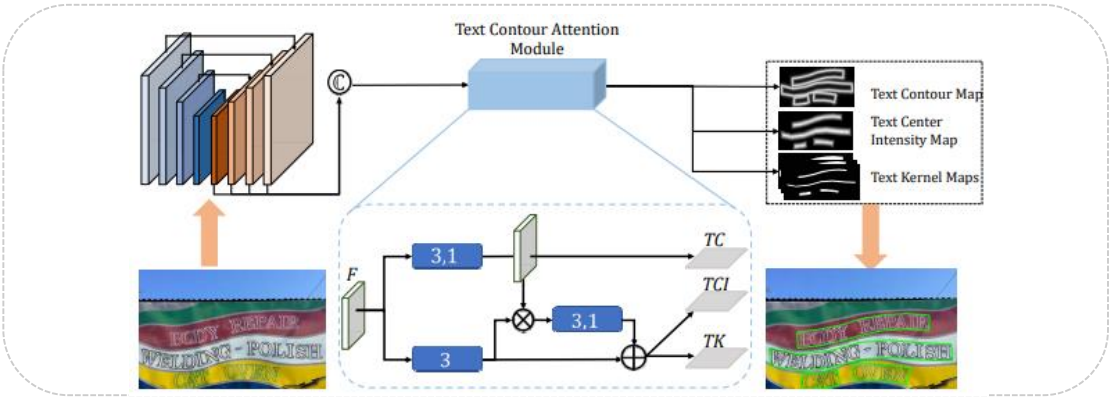
TC of pixel p is given by

$$TC_p = \frac{2 \times e^{-\lambda \times d_p}}{1 + e^{-\lambda \times d_p}}$$

TCI is given by

$$TCI_p = \frac{d_p}{\max(d_q)}, \forall q \in T$$

The Proposed Framework



1.The backbone is the Resnet50/FPN model pre-trained from the ImageNet dataset and fine-tuned with the scene text detection dataset.
2. The Text Contour Attention Module generate Text Contour Map and then use Text Contour Map as attention help generate Text Center Map and Text Kernel Maps.
3. Progressive scale expansion algorithm is applied to Text Kernel Maps to capture text instances .

Evaluations

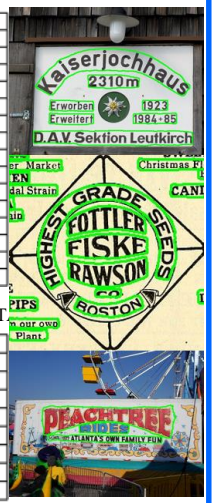
Comparison with the state-of-the-arts on the bench marks.

ICDAR15 EXT" INDICATES EXTERNAL DATA.

Method	Ext	Precision	Recall	F-measure
SegLink [4]	✓	73.1	76.8	75.0
RRPN [15]	-	82.0	73.0	77.0
EAST [3]	-	83.5	73.4	78.2
Lyu *et al.* [14]	✓	94.1	70.7	80.7
TextSnake [6]	✓	84.9	80.4	82.6
LOMO [9]	✓	91.3	83.5	**87.2**
PixelLink [29]	-	82.9	81.7	82.3
RRPN [15]	-	82.0	73.0	77.0
RRD [30]	✓	85.6	79.0	82.2
Textmountain [31]	✓	88.5	84.1	86.2
Baek *et al.* [10]	✓	89.8	84.3	86.9
Tian *et al.* [32]	✓	85.0	88.3	86.6
baseline	-	81.4	79.6	80.5
ours	-	86.6	82.4	84.5
ours	✓	88.9	85.2	87.0

Total-Text EXT" INDICATES EXTERNAL DAT

Method	Ext	Precision	Recall	F-measure
SegLink [4]	-	30.3	23.8	26.7
EAST [3]	-	50.0	36.2	42.0
TextSnake [6]	✓	82.7	74.5	78.4
MSR [17]	✓	85.2	73.0	78.6
SAST [33]	✓	83.7	76.8	80.1
baseline	-	81.7	75.1	78.3
ours	-	85.0	77.3	81.0
ours	✓	86.5	78.4	**82.3**



Key Related Works

[1] W. Wang et al, "Shape robust text detection with progressive scale expansion network," in CVPR, 2019.
[2] Y. Wu et al, "Self-organized text detection with minimalpost-processing via border learning," in ICCV, 2017.
[3] J. Yang et al, "Object contourdetection with a fully convolutional encoder-decoder network," in CVPR,2016.
[4] J. Long et al, "Fully convolutional networks for semantic segmentation," in CVPR, 2015.

Conclusions

1. Our model can separate adjacent text instances by introducing text contour map and text center intensity map.
2. . Employ contour information not only maintains the recall but also improve precision significantly;
3. Our proposed framework achieves state-of-the-art performance.