TCATD: Text Contour Attention for Scene Text Detection

ZiLing Hu, XingJiao Wu, Jing Yang
East China Normal University

Contribution

Contribution 1

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.

01

Contribution 2

We design a Text Contour Attention Module to separate text instances accurately by employing the contour information

02

Contribution 3

The TCATD achieves the state-of-the-art results in both quadrangle and curved text detection benchmarks.

03

Structure

01  Problem

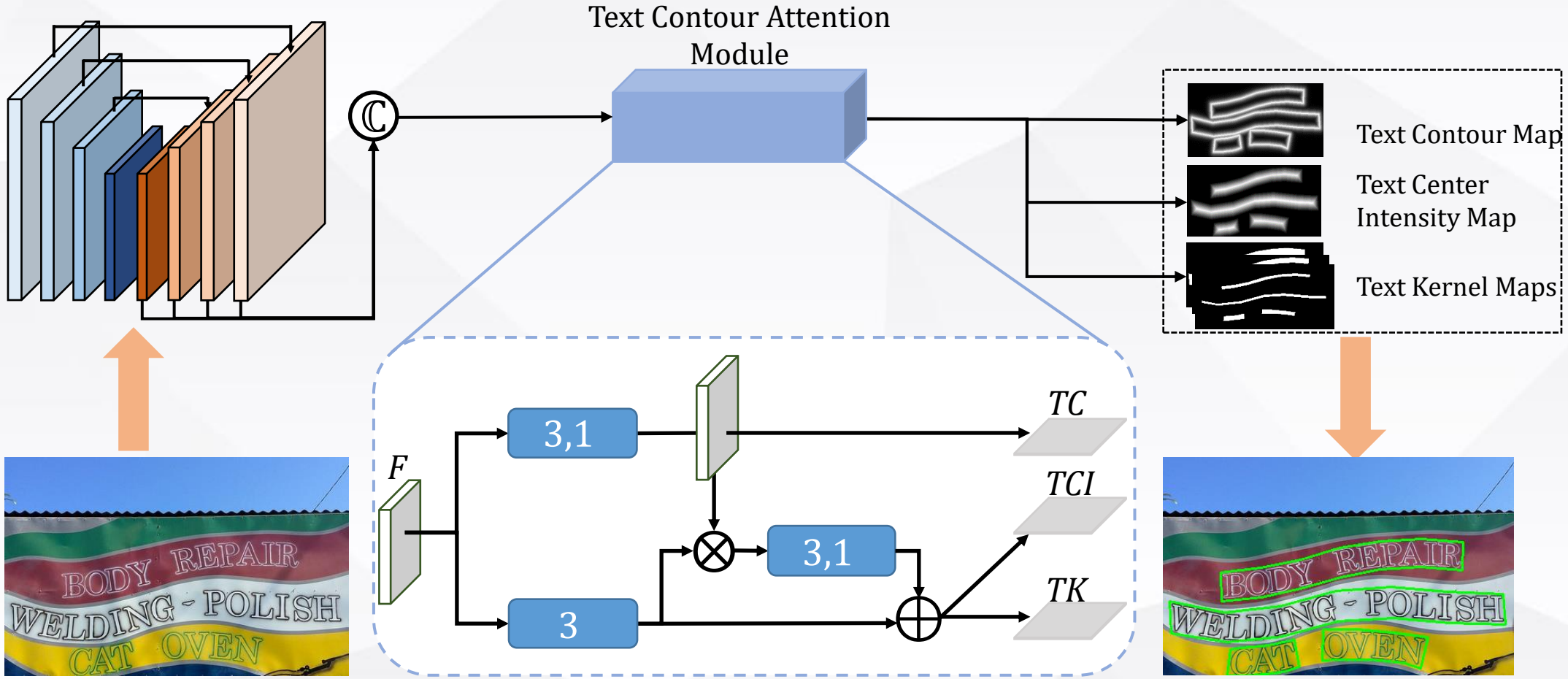
02  Model

03  Result

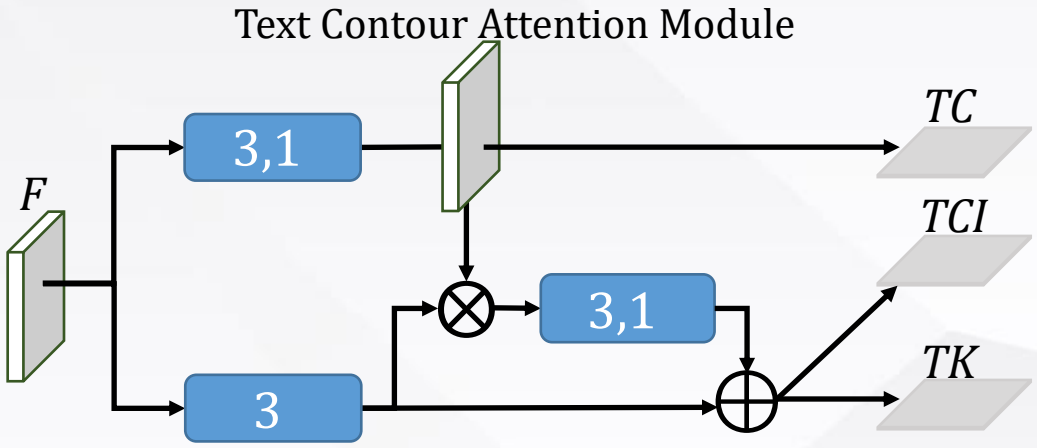
Problem

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.

Model


Text Contour Attention Module
Text Contour Map
Text Center Intensity Map
Text Kernel Maps
F
3,1
3
3,1
TC
TCI
TK
BODY REPAIR
WELDING - POLISH
CAT OVEN

Model



The first branch includes a 3×3 convolutional layer and a 1×1 convolutional layer.

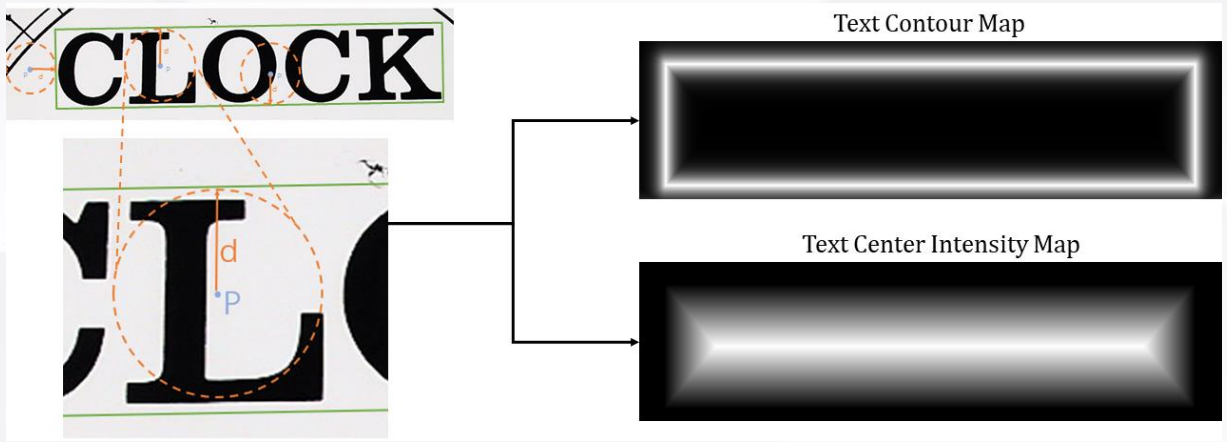
In the second branch, we employ the skip connect attention module based on the shortcut in ResNet and generate TCI and TK.

$$F_s = Conv_3(F)$$
$$F_s' = TC' \otimes F_s$$

After that, F_s' will through a 3×3 convolution layer and a 1×1 convolution layer.

$$\widehat{F_s} = F_s \oplus F_s'$$

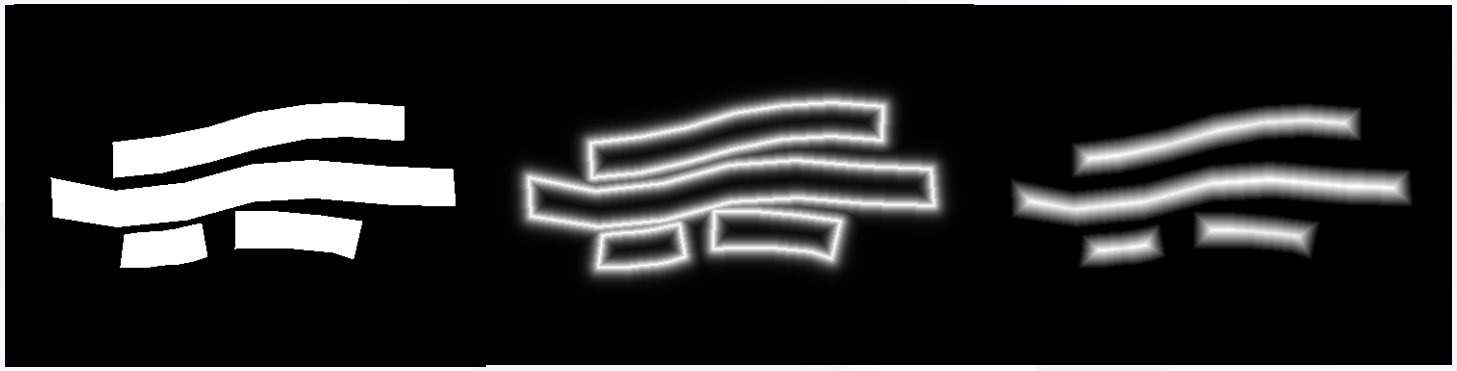
Method



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

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

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



Result

Text detection results

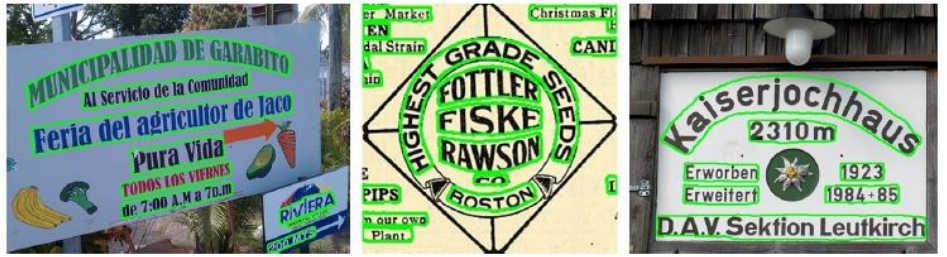
ICDAR-15 Ablation study result

Method	Precison	Recall	F-measure
Baseline	81.4%	79.6%	80.5%
Baseline+TCI	82.7%	80.7%	81.7%
Baseline+TCI+TC	83.3%	82.9%	82.9%
Baseline+TCI+TC+TCAM	86.6%	82.4%	84.5%

ICDAR-15 results "Ext" indicates external data.

Method	Ext	Precison	Recall	F-measure
SegLink	√	73.1%	76.8%	75.0%
RRPN	-	82.0%	73.4%	77.0%
EAST	-	83.5%	73.4%	78.2%
Lyu et al.	√	94.1%	70.7%	80.7%
TextSnake	√	84.9%	80.4%	82.6%
LOMO	√	91.3%	83.5%	87.2%
PSENet	-	81.4%	79.6%	80.5%
Ours	-	86.6%	82.4%	84.5%
Ours	√	88.9%	85.2%	87%

(a) ICDAR 2015



(b) CTW1500

(c) Total-Text

基准数据集上的检测结果样例

Result

Text detection results

CTW-1500 results "Ext" indicates external data.

Method	Ext	Precison	Recall	F-measure
SegLink	-	42.3%	40.0%	40.8%
EAST	-	78.7%	49.1%	60.4%
CTD+TLOC	-	77.4%	69.8%	60.4%
TextSnake	√	67.9%	85.3%	75.6%
He et al.	√	80.0%	73.3%	77.0%
LOMO	√	89.2%	69.6%	78.4%
MSR	√	83.8%	77.8%	80.7%
PSENet	-	80.6%	75.5%	78.0%
Ours	-	85.6%	78.9%	82.1%
Ours	√	86.3%	80.7%	83.4%

Total-Text results "Ext" indicates external data.

Method	Ext	Precison	Recall	F-measure
SegLink	-	30.3%	23.8%	26.7%
EAST	-	50.0%	36.2%	42.0%
TextSnake	√	82.7%	74.5%	78.4%
LOMO	√	88.6%	75.7%	81.6%
MSR	√	85.2%	73.0%	78.6%
PSENet	-	81.7%	75.1%	78.3%
Ours	-	85.0%	77.3%	81.0%
Our	√	86.5%	78.4%	82.3%