

文本检测-CRATD

<Excerpt in index | 首页摘要>

Character Region Awareness for Text Detection

KeyWords Plus: CVPR2019 Curved Text

- paper : [new version paper](#)
- NAVER: line的母公司，韩国的最大的互联网公司，字符级别的文字检测，采用了**CAM热力图**的操作去检测每一个字符

<The rest of contents | 余下全文>

Introduction

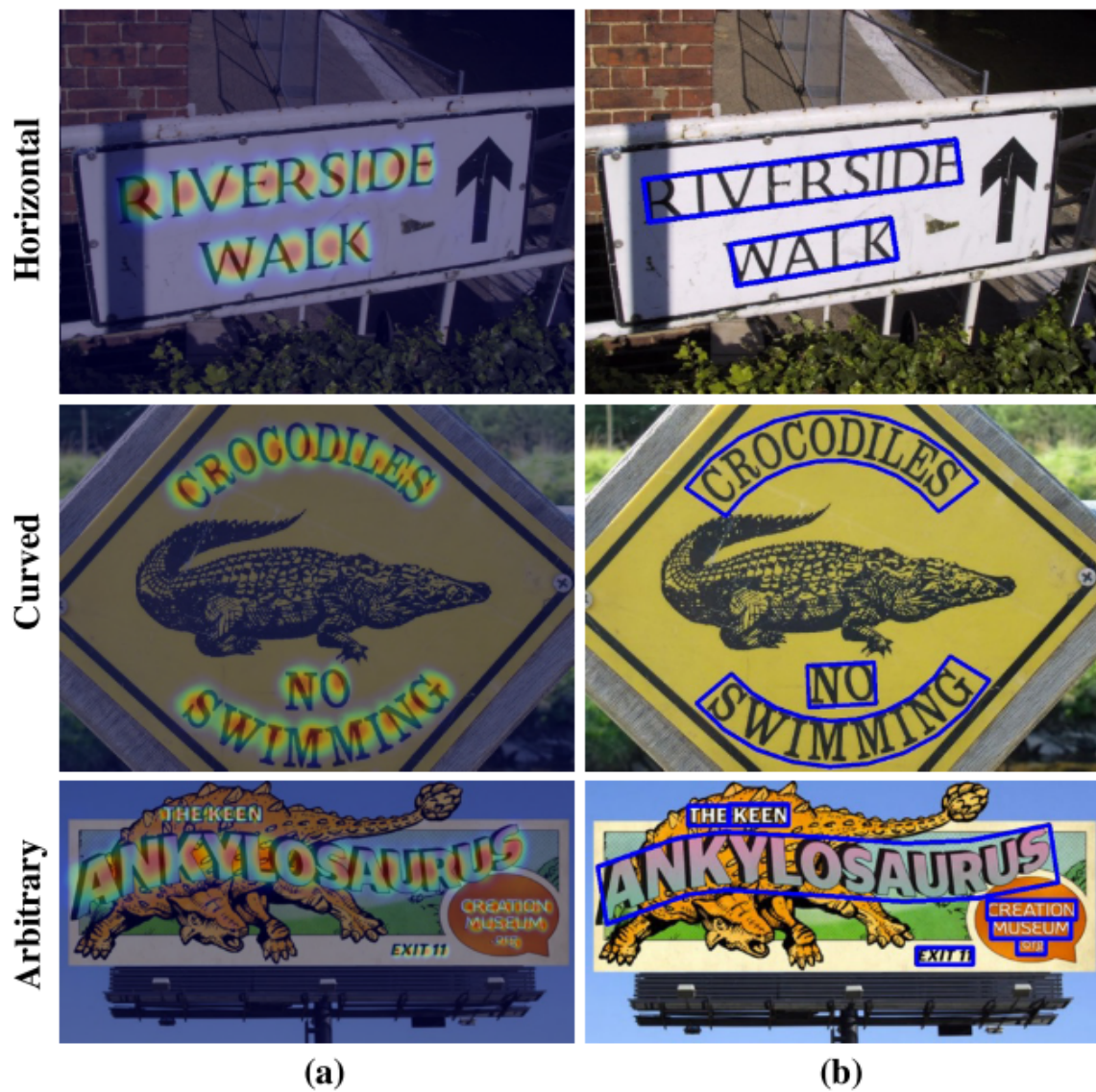
字符级别的文本检测网络，用的是分水岭算法生成label，采用heatmaps去得到激活值最大的目标区域，有点attention的感觉。

1、论文创新点

1.提出了一篇字符级别的检测算法

2.预测得到 :**1.The character region score** **2. Affinity score**. The region score is used to **localize individual characters** in the image, and the affinity score is used to **group each character into a single instance**.

3.Propose a **weakly- supervised learning framework** that estimates character- level ground truths in existing real word-level datasets.



2、算法主体

改论文主要预测每个字符区域和字符之间的一个紧密程度预测，因为没有字符级别的label，所以模型训练在一个弱监督的方式下。网络的backbone采用VGG16，之后接上采样最终输出两个通道：the region score and the affinity score

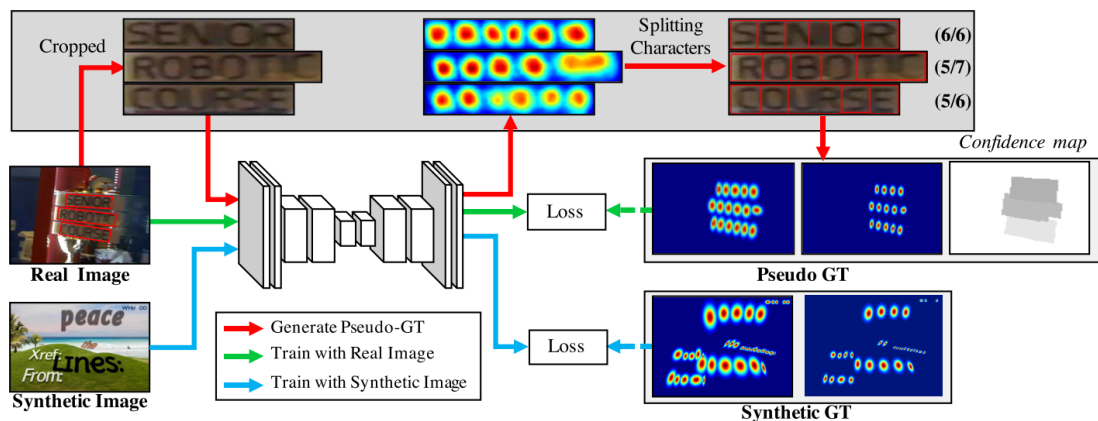


Figure 4. Illustration of the overall training stream for the proposed method. Training is carried out using both real and synthetic images in a weakly-supervised fashion.

训练在一个**弱监督学习的方式下**，有人合成成的数据集具有字符级别的label，real image没有字符级别的标注时，自己检测合成产生label再进行训练。如上图所示，对真实场景中的数据集和人造合成的数据集有不同的训练方式。

3、label generation

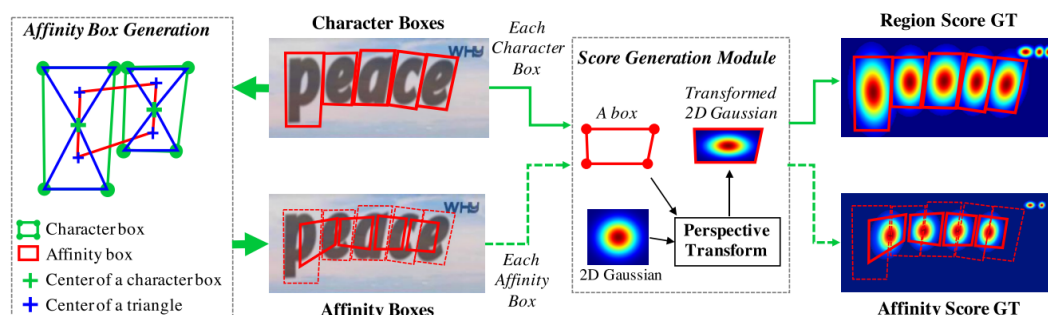


Figure 3. Illustration of ground truth generation procedure in our framework. We generate ground truth labels from a synthetic image that has character level annotations.

分别产生**Region Score GT** 和**Affinity Score GT**

the following steps to approximate and generate the ground truth for both the region score and the affinity score:

- 1) prepare a **2-dimensional isotropic Gaussian map**;
- 2) **compute perspective transform** between the Gaussian map region and each character box;
- 3) **warp Gaussian map** to the box area.

使用**小感受野也能预测大文本和长文本**，只需要关注字符级别的内容而不需要关注整个文本实例。

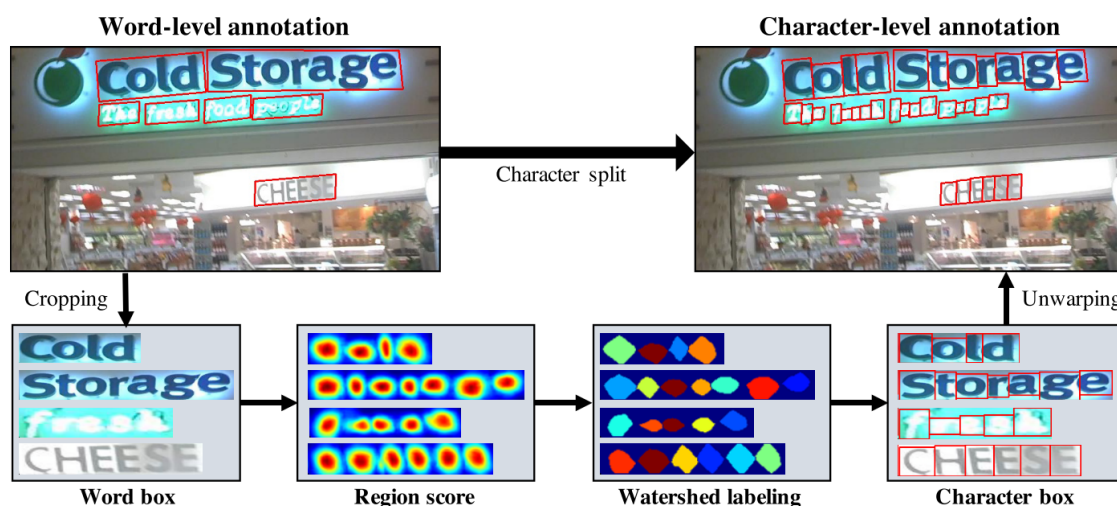


Figure 6. Character split procedure for achieving character-level annotation from word-level annotation: 1) crop the word-level image; 2) predict the region score; 3) apply the watershed algorithm; 4) get the character bounding boxes; 5) unwarp the character bounding boxes.

三步产生字符级别的label:

- 1、抠出文本级别的内容
- 2、预测region score区域
- 3、运用分水岭算法
- 4、得到字符基本的文字框
- 5、贴上文字框

为了防止在弱监督方式下产生的错误label带偏网络，该论文提出了一个评价方式（虽然我没有咋看明白），大概就是计算一个

$$s_{conf}(w) = \frac{l(w) - \min(l(w), |l(w) - l^c(w)|)}{l(w)},$$

4、Post-processing

正常文本后处理分为以下几步：

- 1、首先对0-1之间的概率图进行取阈值计算
- 2、使用Connected Component Labeling(CCL) 进行区域连接
- 3、最后使用opencv的MinAreaRect去框出最小的四边形区域

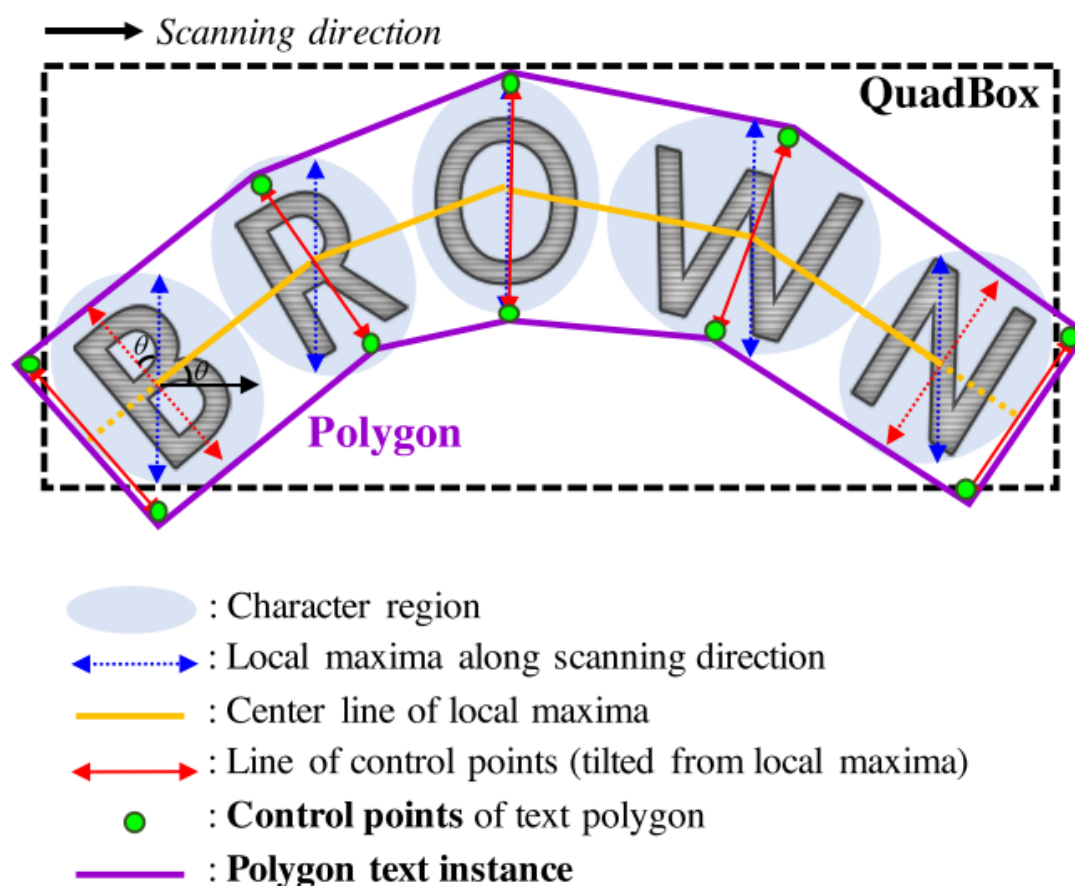


Figure 7. Polygon generation for arbitrarily-shaped texts.

不规则文本检测后处理分为以下几步（如上图所示）：

- 1、先找到扫描方向的局部最大值 (blue line)
- 2、连接所有the local maxima上的中心点叫做中心线
- 3、然后将the local maxima lines旋转至于中心线垂直

4、the local maxima lines上的端点是文本控制点的候选点，为了能更好的覆盖文本，将文本最外端的两个控制点分别向外移动the local maxima lines的半径长度最为最终的控制点。

5、Experiment Results

Method	IC13(DetEval)			IC15			IC17			MSRA-TD500			FPS
	R	P	H	R	P	H	R	P	H	R	P	H	
Zhang et al. [39]	78	88	83	43	71	54	-	-	-	67	83	74	0.48
Yao et al. [37]	80.2	88.8	84.3	58.7	72.3	64.8	-	-	-	75.3	76.5	75.9	1.61
SegLink [32]	83.0	87.7	85.3	76.8	73.1	75.0	-	-	-	70	86	77	20.6
SSTD [8]	86	89	88	73	80	77	-	-	-	-	-	-	7.7
Wordsup [12]	87.5	93.3	90.3	77.0	79.3	78.2	-	-	-	-	-	-	1.9
EAST* [40]	-	-	-	78.3	83.3	80.7	-	-	-	67.4	87.3	76.1	13.2
He et al. [11]	81	92	86	80	82	81	-	-	-	70	77	74	1.1
R2CNN [13]	82.6	93.6	87.7	79.7	85.6	82.5	-	-	-	-	-	-	0.4
TextSnake [24]	-	-	-	80.4	84.9	82.6	-	-	-	73.9	83.2	78.3	1.1
TextBoxes++* [17]	86	92	89	78.5	87.8	82.9	-	-	-	-	-	-	2.3
EAA [10]	87	88	88	83	84	83	-	-	-	-	-	-	-
Mask TextSpotter [25]	88.1	94.1	91.0	81.2	85.8	83.4	-	-	-	-	-	-	4.8
PixelLink* [4]	87.5	88.6	88.1	82.0	85.5	83.7	-	-	-	73.2	83.0	77.8	3.0
RRD* [19]	86	92	89	80.0	88.0	83.8	-	-	-	73	87	79	10
Lyu et al.* [26]	84.4	92.0	88.0	79.7	89.5	84.3	70.6	74.3	72.4	76.2	87.6	81.5	5.7
FOTS [21]	-	-	87.3	82.0	88.8	85.3	57.5	79.5	66.7	-	-	-	23.9
CRAFT(ours)	93.1	97.4	95.2	84.3	89.8	86.9	68.2	80.6	73.9	78.2	88.2	82.9	8.6

ICD13-ICD15-ICD17

Method	TotalText			CTW-1500		
	R	P	H	R	P	H
CTD+TLOC [38]	-	-	-	69.8	77.4	73.4
MaskSpotter [25]	55.0	69.0	61.3	-	-	-
TextSnake [24]	74.5	82.7	78.4	85.3	67.9	75.6
CRAFT(ours)	79.9	87.6	83.6	81.1	86.0	83.5

Table 2. Results on polygon-type datasets, such as TotalText and CTW-1500. R, P and H refer to recall, precision and H-mean, respectively. The best score is highlighted in **bold**.

Total-CTW

6、Conclusion and Future work

个人观点：不太受感受野的限制，只关注单个文字，对于长文本和不规则文本不必特意去设置相应大小的卷积提升感受野。

反馈与建议

- 邮箱：<weijia_wu@yeah.net>

