文本检测TIoU–metric

<Excerpt in index | 首页摘要>

Tightness-aware Evaluation Protocol for Scene Text Detection

KeyWords Plus: **CVPR2019** **Curved Text** **metric** 一种新的评价指标改进了以为评价指标的一些缺陷

- **paper**： Paper
- **Github**: Github

<The rest of contents | 余下全文>

Introduction

文本检测近年来发展迅速，不规则文本检测常用分割做检测也发展很快，之前大多数是直接套用了传统的评价指标**precision，recall，h-mean**,但是这里有一些问题，今年cvpr上这篇文章就是提出了一种新的评价指标解决这些问题。

Existing metrics exhibit some **obvious drawbacks**:

1) .They are **not goal-oriented**;
2).They cannot recognize the **tightness of detection methods;**
3).Existing **one-to-many** and **many-to-one** solutions involve inherent loopholes and deficiencies

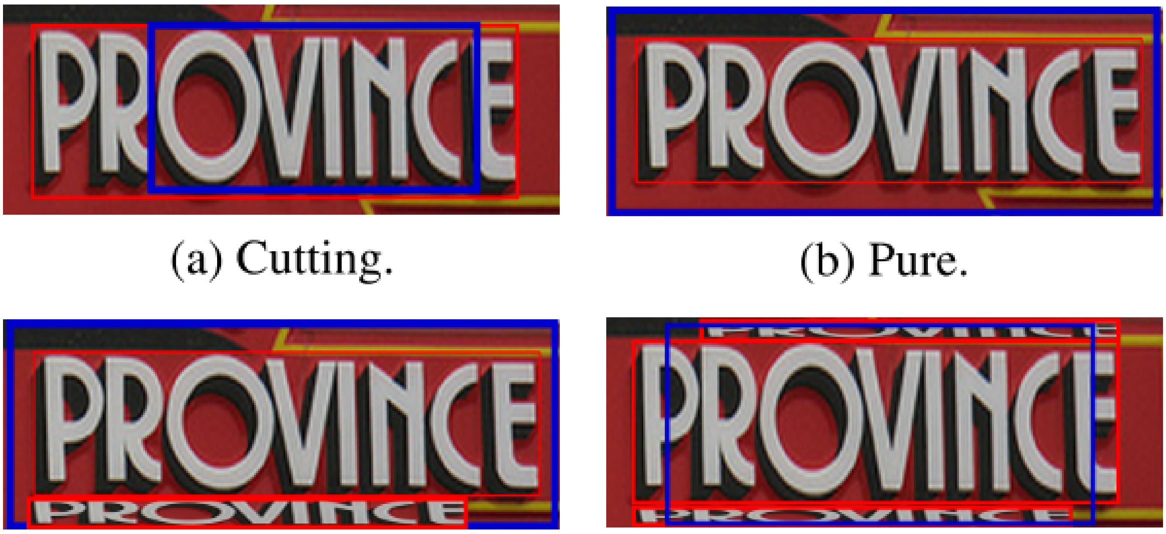
1、论文创新点

现有评价指标存在的问题:

1、As shown in Fig. 1 (a), detection over a fixed IoU threshold with the ground truth (GT) may not com- pletely recall the text (**some characters are missed**); however, previous metrics consider that the GT has been entirely recalled.在检测不完全的情况下，交并比达到一定阈值也为认为检测到了，这**在文本检测中会丢失信息。** `不合理`

2、As shown in Figs. 1 (b), (c), and (d), detection over a fixed IoU threshold with the GT may **still contain background noise**; however, previous metrics consider such detection to have 100% precision.**含有背景噪声**，但是认为precision已经是100%， `不是很合理。`

3、Previous metrics severely rely on an **IoU threshold**. However, if a relatively high IoU threshold is set, some satisfactory bounding boxes may be discarded (e.g., if 0.7 is set as the threshold, the detection in Fig. 1 (b) will be misjudged); if a low IoU threshold is set, sev- eral inexact bounding boxes would be included. `单纯靠IOU阈值来判断文本检测结果,造成了单个文本指标不是1就是0的局面，不合理的。`



(a) Cutting. (b) Pure.

(c) Outlier-GTs. (d) Cutting & Outlier-GTs.

Figure 1. Unreasonable cases obtained using recent evaluation metrics. (a), (b), (c), and (d) all have the same IoU of 0.66 against the GT. Red: GT. Blue: detection.

这个文章主要做的创新点分为以下三点：

1、**Completeness.**Using the TIoU metric would force methods to pay more attention to recalling every part of the GT, i.e., ensuring the completeness of GT,**完整性，要求指标更关注GT的每一个部分，确保文本的完整性**

2、**Compactness..**Because the detections of outlier-GT will be punished by TIoU, the compactness of the de- tection would receive more attention.**将其他文本的GT包含进来将会被惩罚，更关注检测结果的简洁。**

1、**Tightness-aware.**TIoU can distinguish the tightness among different detection methods, i.e., a 0.9 IoU de- tection would be much better than a 0.5 IoU detection in our metric.**有区分不单单是一个阈值，0.9的iou比0.5iou指标更高。**

2、以往各个主流的评价方法

1、ICDAR 2003 (IC03)
2、ICDAR 2013 (IC13)
3、ICDAR 2015 (IC15)
4、AP-based methods
这边就就不一个个介绍了，大家有兴趣就去看看相关论文和代码。

ICDAR 2013 (IC13)

$$\frac{A(G_i \cap D_j)}{A(D_j)} > tp,$$

$$\frac{A(G_i \cap D_j)}{A(G_i)} > tr,$$

tp和tr是precision和recall的两个阈值，分为三部分**OO,OM,MO**分别是一对一，一对多和多对一

OM一对多表示一个GT对多个detection，满足两个条件即可：a、足够多的检测覆盖GT b、每个检测结果要被GT充足的覆盖，如果满足这个两个要求，**precision和recall都为0.8**

MO一对多表示一个detection对多个GT，满足两个条件即可：a、检测必须包含充足的GT b、每一个检测对应足够的面积，如果满足这个两个要求，**precision和recall都为1**

这个评价指标主要有**两个问题**:

1、多对一，在很复杂的情况下，文本很多，**一个大框可以检测到达很高的指标**，但是检测结果是没有意义的，不能被识别所利用。

2、一对多，a method that separates a perfect detection into numerous small over-segmented OM detections (e.g., 20) can make the pre- cision close to 0.8.如下图公式所示，**将一个文本分割成多个会损失信息，但是却能拉高指标接近0.8左右。**

$$origin\_precision = \frac{0+1+0+0}{4} = 0.25 \qquad (7)$$

$$fake\_precision = \frac{0+\overbrace{0.8+\ldots+0.8}^{20}+0+0}{23} = 0.7 \qquad (8)$$

ICDAR 2015 IoU Metric

To be con- sidered a correct detection, **the value of Intersection-over- Union must exceed 0.5.**

$$\frac{A(G_j \cap D_i)}{A(G_j \cup D_i)} > 0.5.$$

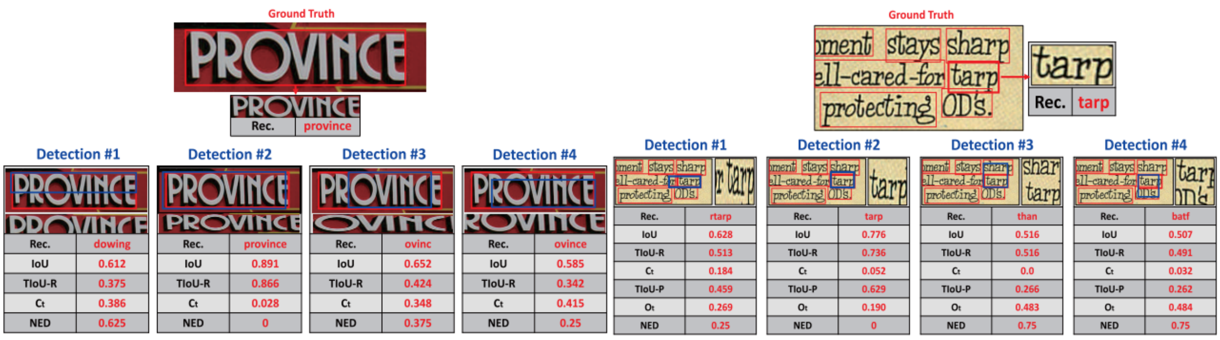
3、Methodology

检测的目的是为了识别，之前版本的检测并没有关注文本内容等信息，为此提出三个概念去加强文本内容信息：

1、**text instance不能被分割成多个文本区域**

2、**annotation应该尽可能包含更少的背景噪声，特别是别的文本实例内容**

3、**annotation应该尽可能的被检测得到的text instance完美匹配**



(a) Cutting effect. (b) Outlier effect.

Figure 3. Qualitative visualization of TIoU metric. Blue: Detection. Bold red: Target GT region. Light red: Other GT regions. Rec.: Recognition results by CRNN [24]. NED: Normalized edit distance. Previous metrics evaluate all detection results and target GTs as 100% precision and recall, respectively, while in TIoU metric, all matching pairs are penalized by different degrees. C_t is defined in Eq. 10. O_t is defined in Eq. 13.

TIoU-Recall

关于TIoU的计算，**引入了一个惩罚机制**，避免一个阈值定结果，出现对后面识别的干扰，如上图a随意，都是一样大小的Iou识别结果却差的很大。公式如下：

Firstly, we define the not-recalled area of G_i as C_t:

$$C_t = A(G_i) - A(D_j \cap G_i), C_t \in [0, A(G_i)], \tag{10}$$

where $A(*)$ means the area of the region. Then, the proportion of intersection in G_i is given by:

$$f(C_t) = 1 - x, x = \frac{C_t}{A(G_i)}. \tag{11}$$

Therefore, the final TIoU-Recall is defined as follows:

$$TIoU_{Recall} = \frac{A(G_i \cap D_j) * f(C_t)}{A(G_i \cup D_j)}. \tag{12}$$

主要是引入了**交集与GT的一个比例惩罚限制最终指标**。不再是单纯的一个阈值定高低。

TIoU-Precision

如果**一个检测结果覆盖了好几个GT**，这样的情况也会有个惩罚，毕竟框进来别的文本会对识别造成干扰而导致识别出错。

$$\begin{aligned} O_{t_{ij}} =& A((G_1 \cap D_j - G_1 \cap D_j \cap G_i) \cup \\ & \ldots \cup (G_{i-1} \cap D_j - G_{i-1} \cap D_j \cap G_i) \cup \\ & (G_{i+1} \cap D_j - G_{i+1} \cap D_j \cap G_i) \cup \ldots \cup \\ & (G_n \cap D_j - G_n \cap D_j \cap G_i)), \\ & O_{t_{ij}} \in [0, A(D_j - D_j \cap G_i)]. \end{aligned} \tag{13}$$

Note that for each $G_n (n \neq i)$ that does not intersect with D_j, it can be simply ignored, which can improve computing efficiency. Then, the proportion of intersection in D_j is given by:

$$f(O_t) = 1 - x, x = \frac{O_t}{A(D_j)}. \tag{14}$$

Using equation 14, we can define the TIoU-Precision in the same way as TIoU-Recall, as shown in equation 15:

$$TIoU_{Precision} = \frac{A(D_j \cap G_i) * f(O_t)}{A(D_j \cup G_i)}. \tag{15}$$

Tightness-aware Metric

以往计算**recall和precision**的方式是：

$$Recall_{ori} = \frac{\sum Match_{gt_i}}{Num_{gt}},$$

$$Precision_{ori} = \frac{\sum Match_{dt_j}}{Num_{dt}}.$$

计算match时不是1就是0，如果阈值是0.5，**导致了IOU0.51和1的结果是相同的**，这是不对的，在该评价方式中采用了联系的0-1的index

$$Recall_{TIoU} = \frac{\sum TIoU_{recall}}{Num_{gt}}$$

$$Precision_{TIoU} = \frac{\sum TIoU_{precision}}{Num_{dt}}$$

The Solution of One-to-many and Many-to-one Metrics

在该评价方式中，**解决一对多，多对一的方式是**：

$$TIoU^{*}_{Recall} = \frac{A(G_j \cap D_i) * f(C_t)}{A(G_j)}.$$

简单粗暴。

(a) East. (b) PixelLink. (c) RRD.

4、Experiments

Table 1. Comparison of evaluation methods on ICDAR 2013 for general detection frameworks and previous state-of-the-art methods. det: DetEval. i: IoU. $e1$: End-to-end recognition results by using CRNN [24]. $e2$: End-to-end recognition results by using RARE [25]. t: TIoU.

Methods	R_{det}	P_{det}	F_{det}	R_i	P_i	F_i	R_{e1}	P_{e1}	F_{e1}	R_{e2}	P_{e2}	F_{e2}	R_t	P_t	F_t
Faster R-CNN (VGG16) [22]	0.410	0.549	0.469	0.615	0.752	0.676	0.396	0.432	0.413	0.406	0.442	0.423	0.377	0.554	0.448
SSD (300x300) [14]	0.476	0.88	0.618	0.484	0.886	0.626	0.398	0.639	0.491	0.391	0.629	0.483	0.377	0.727	0.496
YOLO-v2 (320x320) [20]	0.431	0.772	0.553	0.481	0.877	0.621	0.372	0.548	0.443	0.526	0.571	0.547	0.339	0.682	0.453
YOLO-v3 (320x320) [21]	0.648	0.823	0.725	0.68	0.874	0.765	0.519	0.611	0.561	0.523	0.516	0.566	0.502	0.696	0.583
YOLO-v3 (512x512) [21]	0.694	0.867	0.771	0.721	0.895	0.799	0.566	0.65	0.605	0.585	0.672	0.625	0.549	0.73	0.627
Mask R-CNN [5]	0.767	0.793	0.780	0.718	0.715	0.716	0.544	0.494	0.518	0.58	0.525	0.551	0.527	0.545	0.536
R-FCN (resNet-50) [1]	0.603	0.796	0.686	0.656	0.869	0.748	0.527	0.627	0.573	0.543	0.647	0.59	0.488	0.712	0.579
Faster R-CNN-FPN [13]	0.674	0.882	0.764	0.686	0.875	0.769	0.578	0.678	0.624	0.597	0.699	0.644	0.551	0.737	0.631
RetinaNet (resNet-50-FPN) [13]	0.452	0.901	0.602	0.46	0.906	0.611	0.409	0.744	0.528	0.385	0.7	0.497	0.375	0.77	0.504
East [32]	0.707	0.816	0.758	0.731	0.835	0.779	0.588	0.595	0.591	0.6	0.607	0.603	0.567	0.684	0.620
SegLink [23]	0.6	0.739	0.662	0.572	0.666	0.615	0.485	0.497	0.491	0.495	0.507	0.501	0.387	0.471	0.425
PixelLink [2]	0.633	0.679	0.655	0.621	0.618	0.619	0.539	0.481	0.508	0.549	0.489	0.517	0.432	0.442	0.437
TextBox [11]	0.731	0.896	0.805	0.741	0.892	0.809	0.594	0.643	0.618	0.614	0.664	0.638	0.564	0.712	0.629
SWT-MSER [3, 19]	0.371	0.258	0.305	0.17	0.181	0.175	0.083	0.075	0.079	0.317	0.243	0.275	0.122	0.136	0.129
FEN [30]	0.899	0.947	0.923	0.885	0.934	0.909	0.719	0.716	0.717	0.759	0.757	0.758	0.721	0.783	0.751
R2CNN [7]	0.905	0.943	0.923	0.875	0.908	0.891	0.745	0.732	0.738	0.762	0.749	0.756	0.687	0.721	0.704
MaskTextSpotter [17]	0.886	0.95	0.917	0.873	0.935	0.903	0.751	0.752	0.752	0.766	0.766	0.766	0.733	0.809	0.769
WordSup [6]	0.871	0.928	0.899	0.702	0.821	0.757	0.611	0.648	0.629	0.624	0.662	0.642	0.533	0.626	0.575
AF-RPN [31]	0.896	0.945	0.92	0.854	0.902	0.877	0.731	0.72	0.725	0.756	0.744	0.75	0.665	0.711	0.687

Table 2. Comparison of metrics on the ICDAR 2015 challenge 4. Word&Text-Line Annotations use our new solution to address OM and MO issues. i: IoU. s: SIoU. t: TIoU.

Methods	Original Word-level-Only Annotations									Word&Text-Line Annotations					
	R_i	P_i	F_i	R_s	P_s	F_s	R_t	P_t	F_t	R_i	P_i	F_i	R_t	P_t	F_t
SegLink [23]	0.728	0.802	0.764	0.54	0.594	0.566	0.467	0.581	0.517	0.747	0.836	0.789	0.505	0.598	0.548
East [32]	0.772	0.846	0.808	0.593	0.65	0.62	0.528	0.635	0.576	0.785	0.864	0.823	0.567	0.64	0.601
RRD [12]	0.778	0.868	0.821	0.594	0.663	0.627	0.515	0.652	0.575	0.783	0.879	0.829	0.53	0.653	0.585
PixelLink [2]	0.817	0.829	0.823	0.616	0.626	0.621	0.552	0.618	0.583	0.829	0.851	0.84	0.585	0.627	0.605
TextBox++ [10]	0.808	0.891	0.847	0.619	0.683	0.649	0.537	0.672	0.597	0.812	0.9	0.854	0.549	0.67	0.603
DMPNet [15]	0.765	0.757	0.761	0.564	0.558	0.561	0.479	0.546	0.51	0.781	0.779	0.78	0.512	0.554	0.532
WordSup [6]	0.773	0.805	0.789	0.568	0.591	0.579	0.49	0.577	0.53	0.785	0.831	0.807	0.522	0.588	0.553
R2CNN [7]	0.828	0.887	0.855	0.641	0.687	0.663	0.559	0.676	0.612	0.831	0.901	0.865	0.577	0.676	0.622
AF-RPN [31]	0.832	0.891	0.861	0.645	0.69	0.667	0.577	0.677	0.623	0.844	0.912	0.877	0.607	0.681	0.642
MaskTextSpotter [17]	0.795	0.89	0.84	0.6	0.671	0.633	0.527	0.658	0.585	0.803	0.906	0.851	0.549	0.662	0.6

在icd13和icd15的对比实验如上图所示**可以看出，大多数算法框架普遍都直接掉了20多个百分点**，这简直是巅峰了文本检测行业，不过确实存在合理之处。

6、Conclusion and Future work

个人观点：人个对这个评价指标还是给予很高的期望，毕竟是按照文本检测的具体情况提出改善的，文本检测也是为了识别服务的，最终发展趋势肯定是端到端，分成两个单独网络实在是太冗余了，但是现有技术达不到这个程度还（虽然有几篇半监督提出了），但是还是蛮难的，这个指标也算是增强了这个趋势。

In future, we will try to use TIoU metric to guide train-ing because its characteristics may be benefited to provide a strong supervision. In addition, it can also be used to help incremental or semi-supervised learning because **TIoU can judge whether a detection is suitable to serve as a new GT annotation.**

文本检测还需要很长的路要走，希望各位大佬一起努力呀。

反馈与建议

- 微博：@柏林designer
- 邮箱：weijia_wu@yeah.net