文本检测–TextSnake

<Excerpt in index | 首页摘要>
TextSnake: A Flexible Representation for Detecting Text of Arbitrary Shapes
KeyWords Plus: **ECCV 2018** **Curved Text**

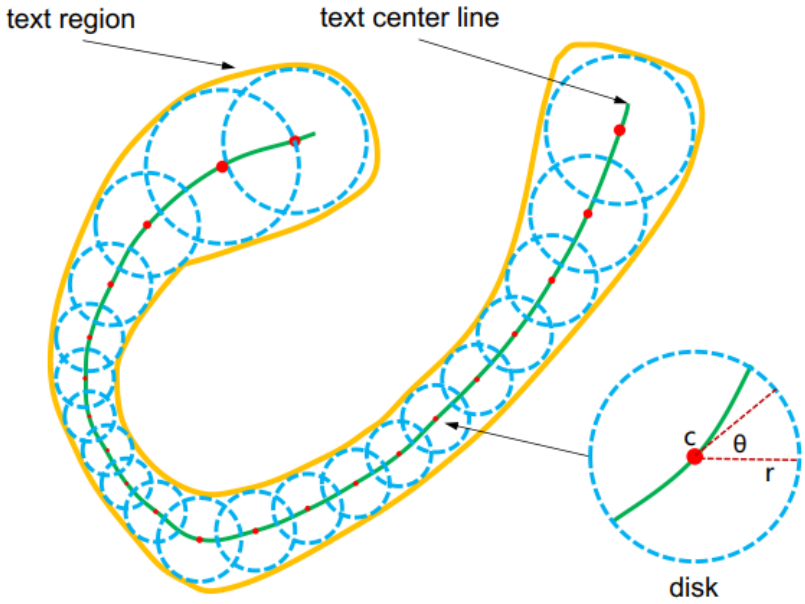
- **relevant blog**： 旷视科技提出TextSnake
- **paper**： TextSnake
- **Github**: TextSnake.pytorch

<The rest of contents | 余下全文>

Introduction

1、论文创新点

1、Propose a flexible and general representation for scene text of `arbitrary shapes`
2、Predict the Text Center Line (TCL) Radius r Orientation θ Text regions (TR)



如上图所示，该论文的创新点主要在于提出一个类似于文本蛇的检测方式对**不规则文本进行预测**，事实上自然场景中多数如下图所示的文本，所以按照正常矩形框的检测方式显然无法有效的解决这种情况。因为自然场景中的文本可以各种形状，但本质不变的就是他必然是一个不断层的文本

（所以我相信作者可能也是基于这个想用很多圆去拟合文本。）

该算法主要做的是五个任务：**1、预测文本 2、预测文本中心线 3、预测一个文本中15个圆的半径 4、预测中心线与圆心的sin 5、预测cos**

Fig. 1. Comparison of different representations for text instances. (a) Axis-aligned rectangle. (b) Rotated rectangle. (c) Quadrangle. (d) TextSnake. Obviously, the proposed TextSnake representation is able to effectively and precisely describe the geometric properties, such as location, scale, and bending of curved text with perspective distortion, while the other representations (axis-aligned rectangle, rotated rectangle or quadrangle) struggle with giving accurate predictions in such cases.

对于这种弯曲并且非平行视角的文本而言，传统的矩形框检测显然不够用了，而且**不规则文本也是之后的发展方向**，有心的读者如何观察18年下半年的论文趋势，可以发现基本找不到以往的矩形框文本检测方式的，大多顶会论文都是针对不规则文本或者是通用文本所提出的解决方法。

2、算法主体

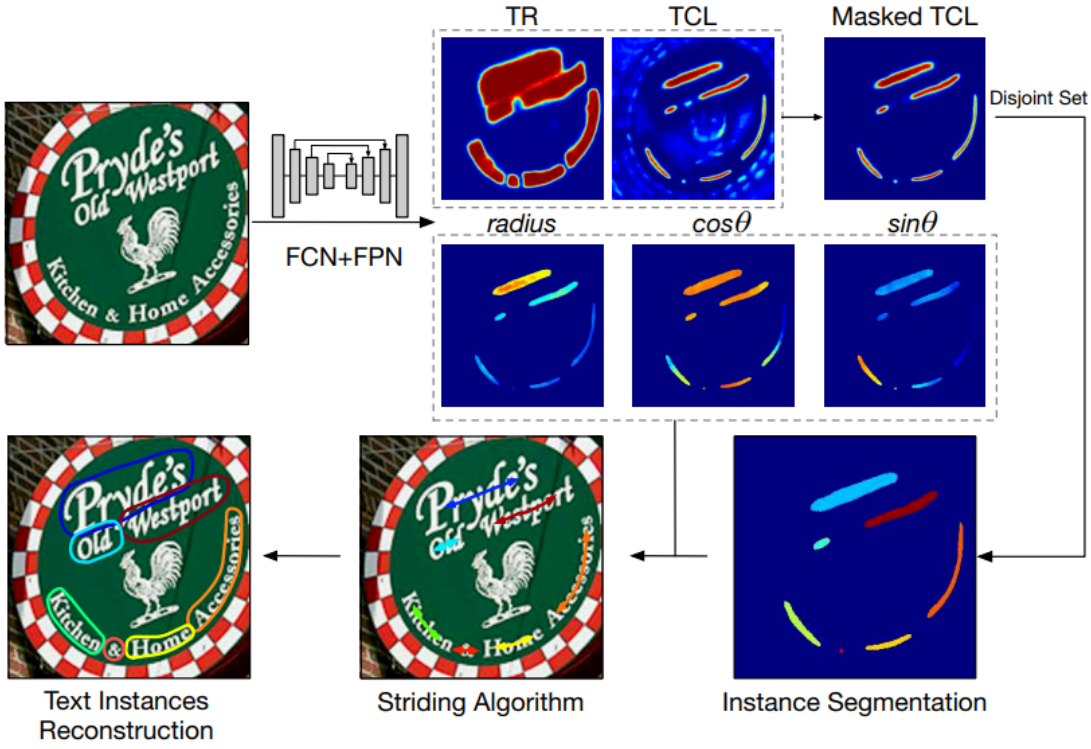


Fig. 3. Method framework: network output and post-processing

In order to detect text with **arbitrary shapes**, we employ an **FCN** model to predict the geometry attributes of text instances.The FCN based network predicts score maps of text center line **(TCL)** and text regions **(TR)**, together with geometry attributes, including **r, cosθ and sinθ.**

网络框架

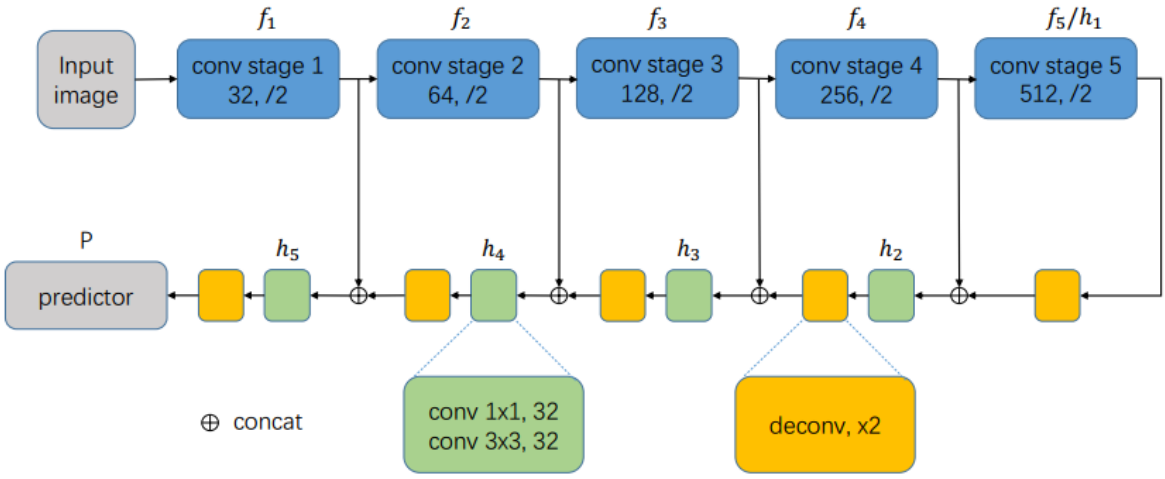


Fig. 4. Network Architecture. Blue blocks are convolution stages of VGG-16.

网络框架如上图所示，采用VGG16，抽取五层的feature map进行融合预测。

下面就是融合上采样的方法，采样五层特征，深层先上采样然后与浅层进行融合，再用一个（1*1）卷积和（3*3）的卷积核进行卷积运算。

上采样

$$h_1 = f_5$$

$$h_i = conv_{3\times3}(conv_{1\times1}[f_{i-1}; UpSampling_{\times2}(h_{i-1})]), \text{ for } i \geq 2$$

产生label

Extracting Text Center Line

For **triangles and quadrangles,** it's easy to directly calculate the TCL with algebraic methods, since in this case, TCL is a straight line.

It has two edges that are respectively the **head and the tail**. The two edges near the head or tail are running **parallel but in opposite direction.**

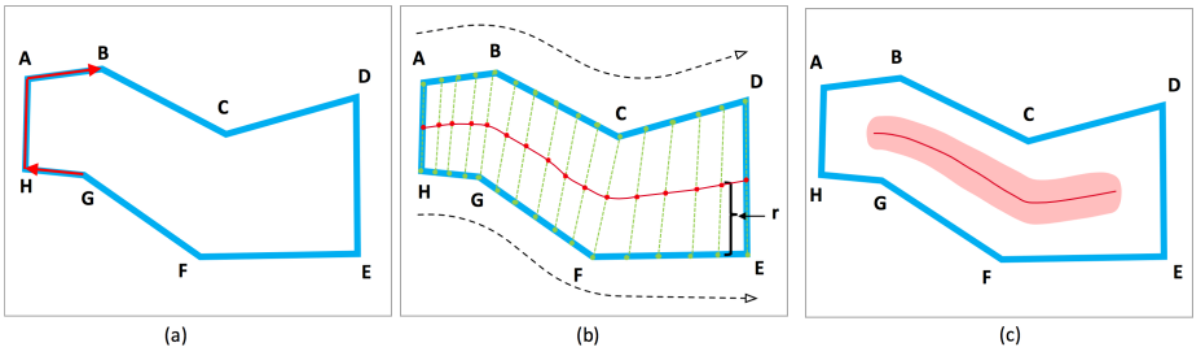


Fig. 7. Label Generation. (a) Determining text head and tail; (b) Extracting text center line and calculating geometries; (c) Expanded text center line.

主要思路是先找到文本的两个端点和边线，然后按照边线的1/2去寻找中心线，除去两边的端点就是**中心线的label。**

3、Loss

Loss 主要分为**分类loss和回归loss**，分类TR和TCL为分类loss，半径角度这些为回归loss，TR和TCL使用的是交叉熵，并加入了**Oinline hard negative mining** 去解决正负样本不平衡问题。

$$L = L_{cls} + L_{reg}$$
$$L_{cls} = \lambda_1 L_{tr} + \lambda_2 L_{tcl}$$
$$L_{reg} = \lambda_3 L_r + \lambda_4 L_{sin} + \lambda_5 L_{cos}$$

Regression loss使用Smoothed loss计算，并且这些只对tcl内的计算，对tcl外的像素没有任何意义。

$$\begin{pmatrix} L_r \\ L_{cos} \\ L_{sin} \end{pmatrix} = SmoothedL1 \begin{pmatrix} \frac{\widehat{r}-r}{r} \\ \widehat{cos\theta} - cos\theta \\ \widehat{sin\theta} - sin\theta \end{pmatrix}$$

4、Datasets

SynthText

Contains about **800K** synthetic images.

TotalText

Newly-released benchmark for text detection. Besides horizontal and multi-Oriented text instances.The dataset is split into **training and testing sets with 1255 and 300 images**, respectively.

CTW1500

another dataset **mainly consisting of curved text**. It consists of **1000 training images and 500 test images**. Text instances are annotated with polygons with **14 vertexes.**

ICDAR 2015

MSRA-TD500

A dataset with **multi-lingual, arbitrary-oriented and long text lines.** It includes **300 training images and 200 test images** with text line level annotations

5、Experiment Results

Method	Precision	Recall	F-measure
SegLink [2]	30.3	23.8	26.7
EAST [3]	50.0	36.2	42.0
Baseline (DeconvNet[42])	33.0	40.0	36.0
TextSnake	**82.7**	**74.5**	**78.4**

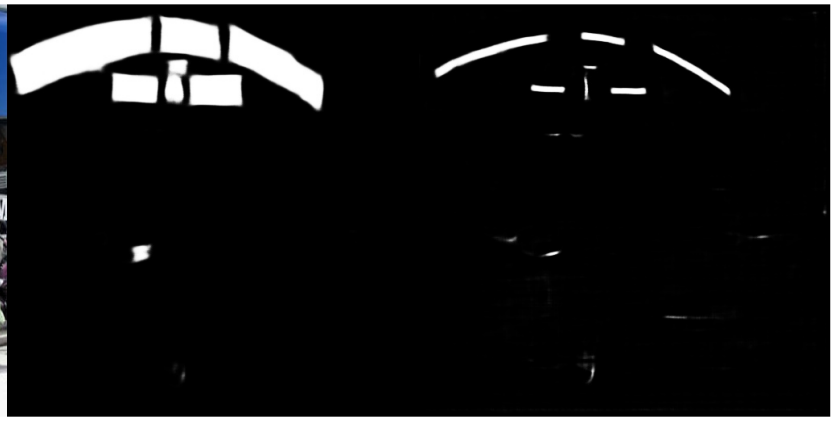
Total-Text

Method	Precision	Recall	F-measure
SegLink [2]	42.3	40.0	40.8
EAST [3]	78.7	49.1	60.4
DMPNet [43]	69.9	56.0	62.2
CTD[13]	74.3	65.2	69.5
CTD+TLOC[13]	**77.4**	69.8	73.4
TextSnake	67.9	**85.3**	**75.6**

CTW1500

Method	Precision	Recall	F-measure	FPS
Zhang *et al.* [27]	70.8	43.0	53.6	0.48
CTPN [44]	74.2	51.6	60.9	7.1
Yao *et al.* [1]	72.3	58.7	64.8	1.61
SegLink [2]	73.1	76.8	75.0	-
EAST [3]	80.5	72.8	76.4	6.52
SSTD [45]	80.0	73.0	77.0	**7.7**
WordSup * [8]	79.3	77.0	78.2	2
EAST * † [3]	83.3	78.3	80.7	-
He *et al.* * † [25]	82.0	80.0	81.0	1.1
PixelLink [46]	**85.5**	**82.0**	**83.7**	3.0
TextSnake	84.9	80.4	82.6	1.1

ICDAR 2015



6、Conclusion and Future work

这篇**paper**在不规则场景文本检测里面也算是先锋者了，不规则场景文本检测的**paper**大多数都是18年后半年迸发的，但是人个感觉得这个**paper**的方法不是很好，比较繁琐，有很多可以改进的地方。

反馈与建议

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