

Arbitrary-oriented scene text detection via rotation proposals

笔记本: Application

创建时间: 2019/10/28 19:39

更新时间: 2019/10/28 21:26

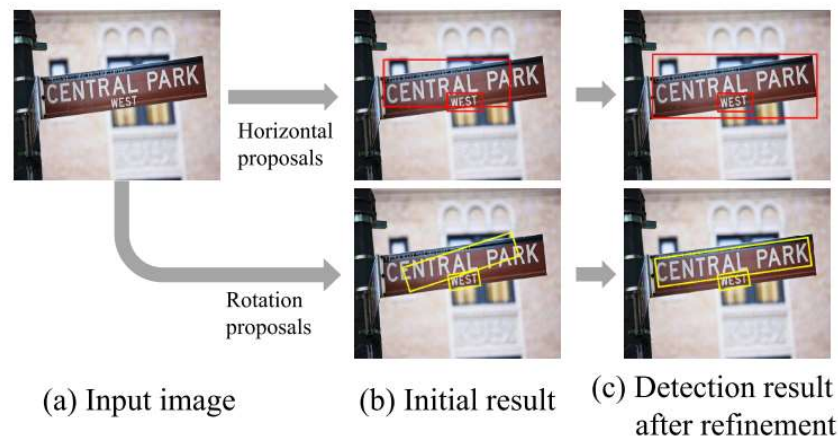
作者: ming71

IEEE Transactions on Multimedia, 2018

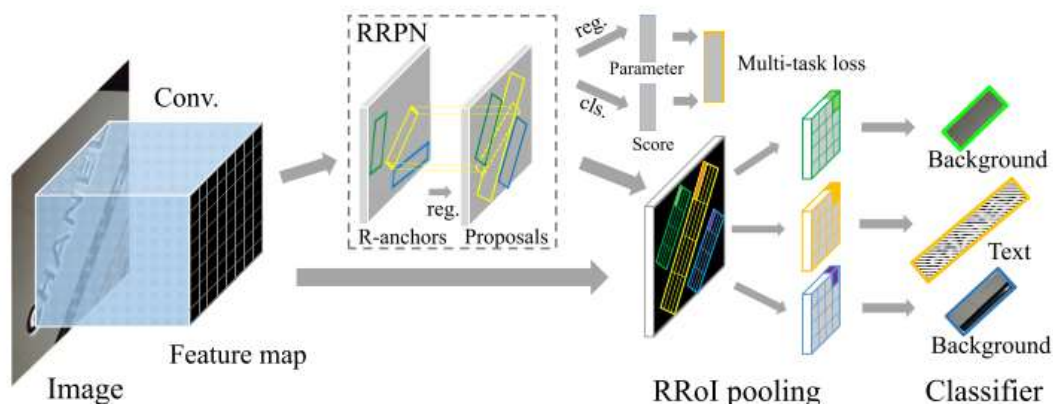
1. Introduction

这篇论文是最早使用RRPN和RRoI Pooling的，并且计算了斜框IoU，虽然本身的思路比较直白，但是在这块的实现上综合了很多当时没有的思路并且实现效果也较好，可以说是比较有意义。

其他斜框的如R2CNN，虽然使用斜框预测，但是在proposal生成阶段仍是正框。本部分使用的是完全的旋转proposal。



2. Approach



从pipeline看出和普通两阶段最大的不同就是加入角度回归，一阶段生成带角度的proposals，二阶段在RoI Pooling上依然使用带角度的特征进行回

归。

- **详细配置**

其中角度设定是 $[-0.25\pi, 0.75\pi]$ ，未约束角度的范围，而是以 $k\pi$ 为周期限定到给定的角度范围内（可能不好收敛，约束比较好）。旋转方式就是个仿射变换阵。旋转anchor上有三尺度，三ratio，六角度，共计54个。

- **损失计算**

loss上增加了额外的角度SmoothL1 loss，同时考虑到角度的周期性，如下：

$$v_{\theta} = \theta \ominus \theta_a$$

$$a \ominus b = a - b + k\pi \quad (v_{\theta} \text{约束在角度限定范围内})$$

- **正负样本划分**

由于新加了角度，在考虑正负anchor属性时还额外考虑角度影响。设定正Anchor满足：（1）和gt有最高iou （2）或者iou大于0.7，且和gt的角度偏差小于 $\pi/12$ 。负样本满足下列之一：（1）iou低于0.3 （2）iou大于0.7但角度偏差大于 $\pi/12$ 。

上述anchor正负样本的设置很大程度将角度的影响给分离出来了，单独为角度考虑了正负样本学习的容易性，从而能够帮助角度更好的回归。

- **skewIoU**

计算其实不难了，不过那个时候作为第一个用斜IoU的，他的复杂推导也是没办法的。

但是注意NMS，也是skew-NMS。传统只考虑iou，现在还必须加入角度（很好理解，如两个一模一样的box，如果角度不对iou就可能大不相同，所以正的iou不能直接作为nms的依据，需要角度修正）。Skew-NMS有两个条件：（1）保留大于0.7的最大iou的proposal （2）如果没有满足（1）的proposal，选择 $[0.3, 0.7]$ 的proposal，且其与gt的角度差小于 $\pi/12$ （总在提 $\pi/12$ ，因为6anchor下，每个anchor 30° ，左右 15° 是其domain）

- **RRoI**

3. Experiment

3.1 Ablation study

- Augmentation

Data Augmentation	Precision	Recall	F-measure
Without rotation	44.5%	38.9%	41.5%
With rotation	68.4%	58.9%	63.3%

针对这个任务，旋转可变性大大复杂，因此加旋转增强一定很有用，毫无悬念

- Context of the Text Region

Factor	Precision	Recall	F-measure
1.0	57.4%	54.5%	55.9%
1.2	59.3%	57.0%	58.1%
1.4	65.6%	58.4%	61.8%
1.6	63.8%	56.8%	60.1%

融合上下文信息。可以想到，斜框的目标检测使得目标特征将会更好地在proposal内响应，因而很可能随着训练的进行失去对复杂环境变化的泛化性，上下文信息应该是有用的。

具体实施是训练时将gt放大一定的系数（factor），加入一定的背景信息。在test时，预测出的box除以这个系数即可得到不包含背景的目标。

- Scale Jittering

就是多尺度训练。

- Border Padding

anchor分配时，一开始的anchor都是正的，这时候超出边界的会被剔除掉。但是实际上这部分的anchor在后面经过旋转后可能会回到图像内，为了尽可能保留这部分的anchor，在计算是将图像额外padding 0.25倍的边界，以尽可能保留可能的正样本。

3.2 Result

MSRA-TD500					ICDAR2015				ICDAR2013			
Approach	P	R	F	Time	Approach	P	R	F	Approach	P	R	F
Yin et al. [53]	71	61	65	0.8 s	CTPN [42]	74	52	61	Faster-RCNN [20]	75	71	73
Kang et al. [54]	71	62	66	-	Yao et al. [18]	72	59	65	Gupta et al. [55]	92	76	83
Yin et al. [56]	81	63	71	1.4 s	SCUT_DMPNet [57]	68	73	71	Yao et al. [18]	89	80	84
Zhang et al. [17]	83	67	74	2.1 s	UCSC_TextSpotter [58]	65	79	71	DeepText [43]	85	81	85
Yao et al. [18]	77	75	76	0.6 s	hust_orientedText [59]	77	75	76	CTPN [42]	93	83	88
RRPN	82	68	74	0.3 s	RRPN	<u>82</u>	73	<u>77</u>	RRPN	90	72	80
RRPN*	<u>82</u>	69	<u>75</u>	0.3 s	RRPN*	84	<u>77</u>	80	RRPN*	95	88	91

