

R3Det: Refined Single-Stage Detector with Feature Refinement for Rotating Object

笔记本：待解决记事本

创建时间：2019/10/25 10:37

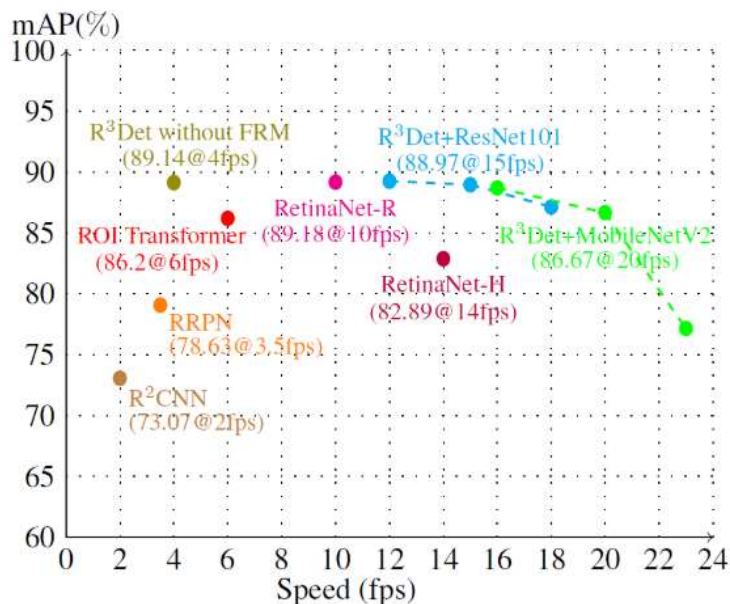
更新时间：2019/10/25 10:38

作者：ming71

2019.8.5

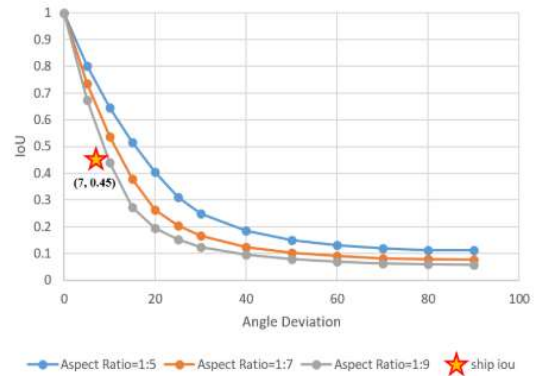
1. Introduction

旋转目标检测的三个问题：悬殊的宽高比、密集的分佈、类别极度不平衡。本文提出了端到端的单阶段旋转目标检测器，能够较好地权衡精度和速度。为了解决当前检测器的特征对齐不佳问题，设计特征精炼模块，通过特征插值重构bbox的位置信息，能够在长尾分布的数据集上取得明显的效果。在DOTA, HRSC2016, ICDAR2015上都取得SOA的成绩。

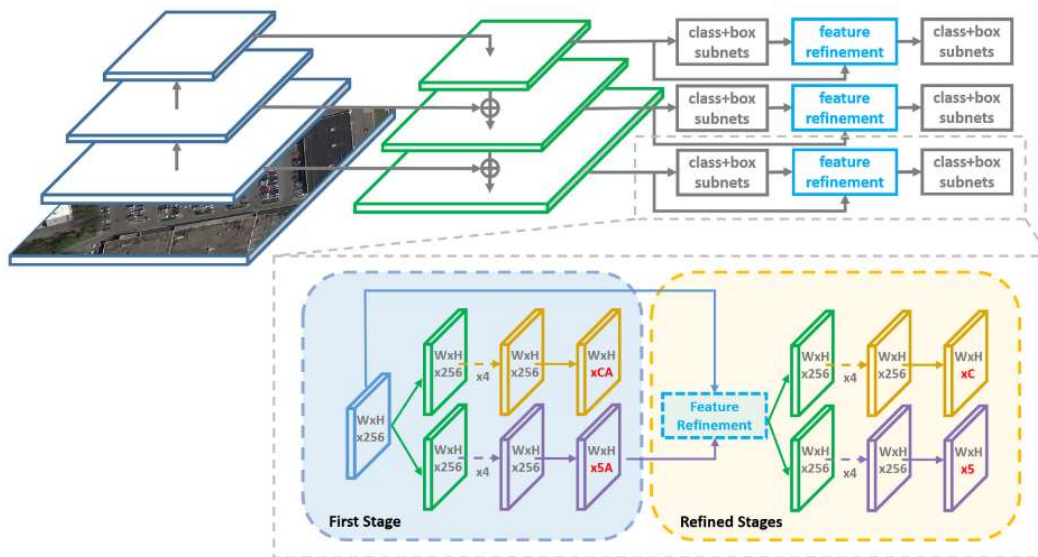


- Large aspect ratio

斜IoU(SkewIoU)在悬殊的宽高比下，随着角度预测的偏差增大，结果会更低。如下图，所有曲线都是随着预测角度的偏差增大而使IoU下降，但是ratio越大的gt下降更厉害。



2. The Proposed Method



大体结构是，以RetinaNet为基础，在三个FPN尺度上预测。每个尺度使用不共享权重的分类回归网络，经过特征对齐后级联refine。

- **Rotation RetinaNet**

分类采用focal loss。

- **Refined Rotation RetinaNet**

- Refined Detection

采用级联优化的方式。第一阶段的前景背景IoU阈值设置为0.5和0.4，后面的都是0.6和0.5。

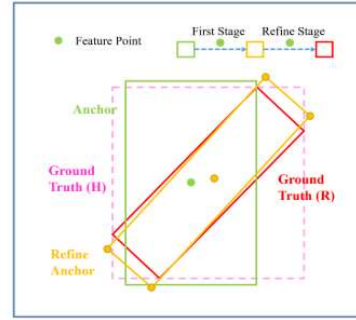
- Feature Refinement Module

解决旋转后特征的对齐问题。通过一定的方式将特征点对齐。如下图所示，采取的插值计算方式为：

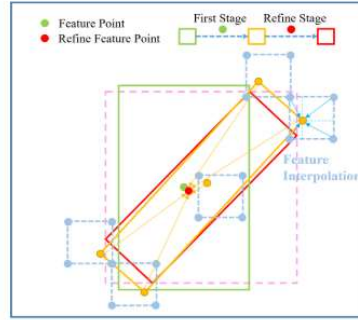
$$val = val_{lt} * area_{rb} + val_{rt} * area_{lb} + val_{rb} * area_{lt} + val_{lb} * area_{rt}$$



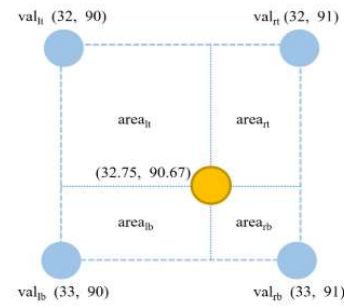
(a)



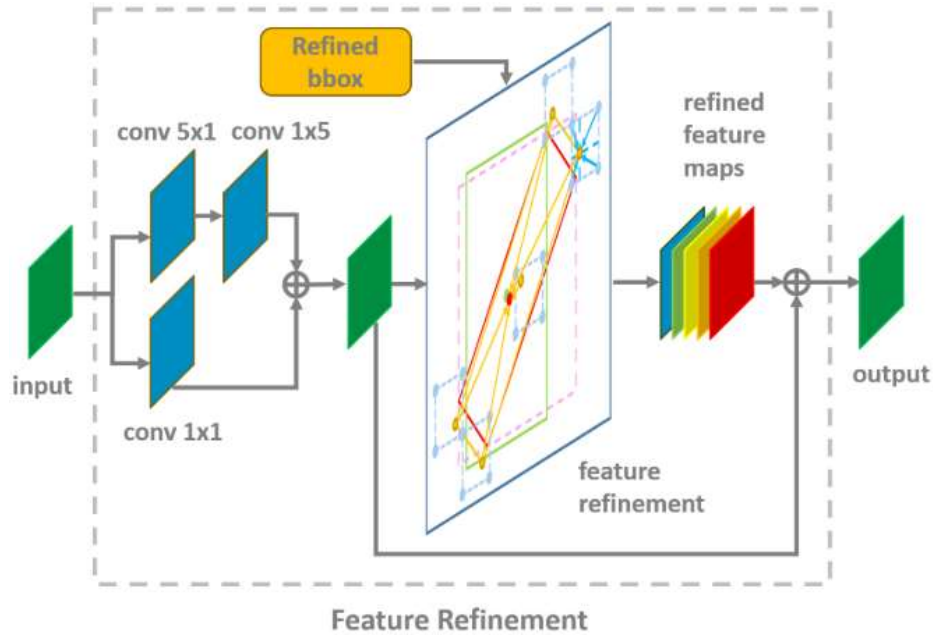
(b)



(c)



(d)



3. Experiments

• Protocol

anchor的设置: P3-P7特征, 分别是area 32*32到512*512; 每个特征图的ratio {1, 1/2, 2, 1/3, 3, 5, 1/5}, scale {20, 21/3, 22/3}, 角度为{-90°, -75°, -60°, -45°, -30°, -15°}。实际采用combine-anchor的形式。

观察：H-R anchor具有更高的召回，可以用很少的anchor获得很好的覆盖率；但是Rotate anchor具有更好的精度。所以首先H-Ranchor获得较好的proposal，后面再使用rotate anchor进行回归。

• Result

Method	FRM	Backbone	Image Size	Data Aug.	mAP	Speed
R ² CNN [15]	-	ResNet101	800*800	×	73.07	2fps
RC1 & RC2 [26]	-	VGG16	-	-	75.7	slow
RRPN [27]	-	ResNet101	800*800	×	79.08	3.5fps
R ² PN [39]	-	VGG16	-	✓	79.6	slow
RetinaNet-H	-	ResNet101	800*800	✓	82.89	14fps
RRD [21]	-	VGG16	384*384	-	84.3	slow
RetinaNet-R	-	ResNet101	800*800	✓	89.18	10fps
RoI-Transformer [8]	-	ResNet101	512*800	×	86.20	6fps
R ³ Det (proposed)	×	ResNet101	800*800	✓	89.14	4fps
	✓	ResNet152	800*800	✓	89.33	10fps
	✓	ResNet101	300*300	✓	87.14	18fps
	✓	ResNet101	600*600	✓	88.97	15fps
	✓	ResNet101	800*800	✓	89.26	12fps
	✓	MobileNetV2	300*300	✓	77.16	23fps
	✓	MobileNetV2	600*600	✓	86.67	20fps
	✓	MobileNetV2	800*800	✓	88.71	16fps

Table 4: Comparison of the accuracy and speed of different methods on the HRSC2016 dataset.