

如何评价商汤提出的目标检测新框架Grid R-CNN?

非利益相关方，仅在arxiv看到这篇论文（arxiv的论文感觉标题越短越可看？）

传统基于R-CNN的架构通过回归边界框实现定位，RoI的特征需要flatten+FCs预测边界框，属于曲线救国。ECCV2018的cornernet通过直接预测边界框的左上角和右下角，预测目标的边界框。而本文是商汤最新提出的Grid R-CNN（没有实习生），直接预测边界框包含四个边界在内的9宫格，在Faster R-CNN/Mask R-CNN基础上实现state-of-art.

具体细节，欢迎探讨。

arxiv.org/pdf/1811.1203...

关注问题

写回答

邀请回答

添加评论

分享

举报

收起

查看全部 11 个回答



TeddyZhang

机器学习/模式识别/深度学习

占坑，读完了再来回答

发布于 2018-12-05

赞同

添加评论

分享

收藏

感谢

更多回答



赫拉迪克方块

你身上的赫拉迪克方块是一个不折不扣的宝物

45 人赞同了该回答

首先motivation是合理的，det最后用fc去做regression是会损害定位能力的，在今年COCO比赛的时候Face++就用了Location Sensitive Header，在做bbox那一支的时候已经改用conv出了，附上slides。

文章的思路也是基于点去做，和CornerNet相比，Grid是基于proposal的，相当于已经确定了instance，所以不需要考虑bottom-up方法中将点组合的过程，难度上要小一些。其他的话用了ensemble的思想，出多个点考虑相关性，本质上是在做一些加权来得到更鲁棒的表达。

实现部分的话，grid branch这一个分支首先就是RoI Align到14x14，然后接了8 dilate conv再deconv到56x56，这个复杂度有点爆炸的吧（所以最后每张卡上只放的下一张图）。然后用了Sync-BN在Res50-FPN上涨了2.2(不知道BN在这里面有多少gain)。

发布于 2018-12-03


赞同 45

添加评论

分享

收藏

感谢



ChenJoya

做一个有趣又有理想的人

148 人赞同了该回答

下午那个匿名用户就是作者之一吧，你别跑，为什么删答案 我还想问问 Grid Points Feature Fusion 那块到底是怎么做的呢，真是感觉写的不清楚

读了读论文，我们先把文章的脉理理一下哈：

文章面向的问题主要是在讲如何干掉 R-CNN detector 在提取出 RoI feature 后的 regression 分支（虽然貌似传统的 regression 分支缺点在哪并没有特别讲清楚，只是反复强调了全卷积网络的



关于作者



TeddyZhang

机器学习/模式识别/深度学习

回答

4

文章

25

关注者

317

关注他

发私信

相关问题

目标检测领域还有什么可以做的？ 15 个回答

目标检测中region proposal的作用？ 8 个回答

目标检测的首选深度框架？ 4 个回答

目标检测算法有哪些？ 5 个回答

卷积神经网络可以用于小目标检测吗？ 17 个回答

相关推荐



深度学习之卷积神经网络 CNN

★★★★★ 66 人参与



这些年，从 ACCA 汲取框架

5 人参与

position sensitive property 是全连接不具有的), 如下图, 传统 regression 分支一般由全连接层来预测 4 个边界框相关的调整量, Grid R-CNN 中采用几个卷积层去预测几个 grid point 的 heat map, 而后进行边界框的输出。

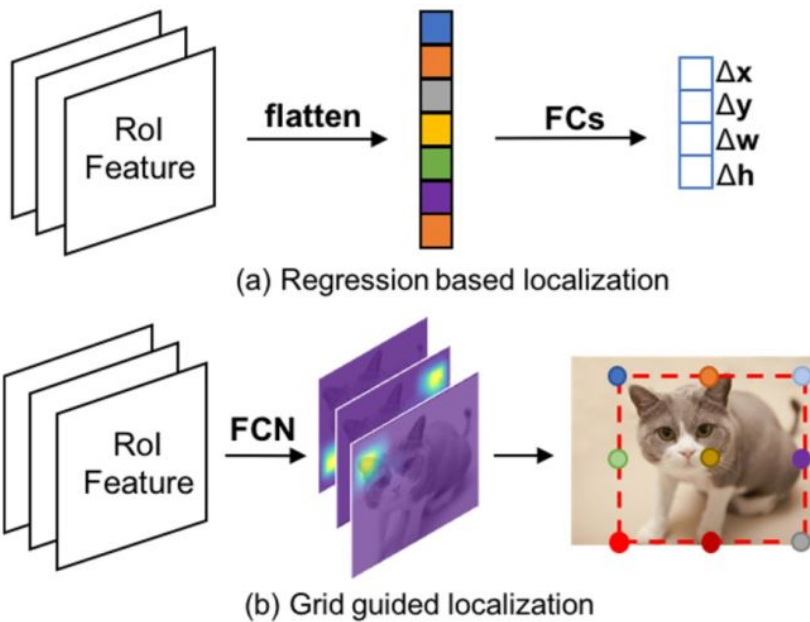


Figure 1. (a) Traditional offset regression based bounding box localization. (b) Our proposed grid guided localization in Grid R-CNN. The bounding box is located by a fully convolutional network.

像 CornerNet 这样基于角点的检测器 (没瞄过的直接看看这里 [CornerNet: 目标检测算法新思路](#)), 算是一个将关键点检测到引入到目标检测的先行工作。而 Grid R-CNN 与 CornerNet 有哪些不同呢?

首先 CornerNet 只预测边框的左上和右下角两个点, 而 Grid R-CNN 中采用了 multi-point 的形式, 作者认为两个点的预测较难, 因为某点落在的区域可能根本旁边都是背景, 这个点与周围的局部特征相似而难以辨别, 就像上面那只猫的右上角那个点。此外, 对于预测不准的点可以由其他点做一个矫正, 因而可以减少它们所带来的偏移。

其次, 它们对点预测的所述类别不同。在关键点检测中分为两种思路: Bottom-Up 和 Top-Down, Bottom-Up 这种方法先图片中所有类别的所有关键点全部检测出来, 然后对这些关键点进行聚类处理, 将不同人的不同关键点连接在一块, 从而聚类产生不同的个体; 而 Top-Down 这种方法先行人检测, 而后再对每个人体子图再使用关键点检测。CornerNet 是基于 Bottom-Up 的, 而 Grid R-CNN 是基于 Top-Down 的。Bottom-Up 的缺点很显而易见, 分组容易搞错, 弄到别的目标上去。

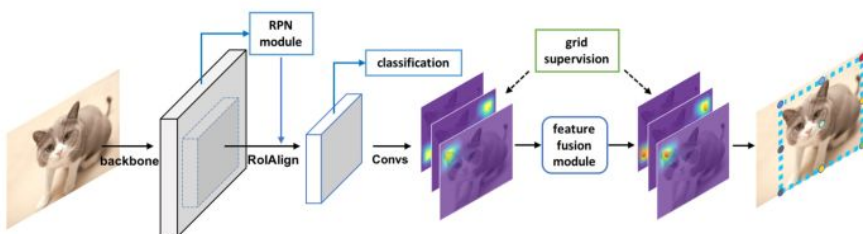


Figure 2. Overview of the pipeline of Grid R-CNN. Region proposals are obtained from RPN and used for RoI feature extraction from the output feature maps of a CNN backbone. The RoI features are then used to perform classification and localization. In contrast to previous works with a box offset regression branch, we adopt a grid guided mechanism for high quality localization. The grid prediction branch adopts a FCN to output a probability heatmap from which we can locate the grid points in the bounding box aligned with the object. With the grid points, we finally determine the accurate object bounding box by a feature map level information fusion approach.

整个 Grid R-CNN 的框架如上, pipeline 的前半部分相比于 R-CNN detector 没有变动, 我们关注怎么做 multi-point 预测的训练和前向推理。重点分为三个部分:

Grid Guided Localization.

看 pipeline 的图。RPN 得到的 proposal 在 RoI Align resampling 一下后, 通过一些空洞卷积扩大感受野 (eight 3×3 dilated convolutional layers), 然后通过两个反卷积把尺寸扩大, 再通过一个卷积生成与 multi-point 相关的 heat maps (9 个点就是 9 张图)。我们为每一个点提供一个交叉十字形状的 ground truth (5 pixels), 如下所示 (恩, 感受一下我的灵魂画功):



分布式实时计算框架原理及
实践案例

138 人读过

阅读



腾讯课堂 NEXT 学位
Next Degree

官方小程序课程
内部福利大放送

广告

刘看山 · 知乎指南 · 知乎协议 · 知乎隐私保护指引

应用 · 工作 · 申请开通知乎机构号

侵权举报 · 网上有害信息举报专区

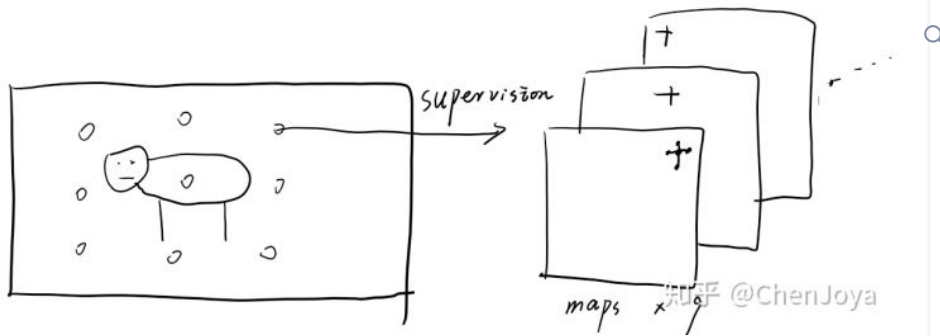
违法和不良信息举报: 010-82716601

儿童色情信息举报专区

电信与服务业务经营许可证

网络文化经营许可证

联系我们 © 2018 知乎



特征图上每个像素都是用 logistic regression 来训练的。在测试的时候，每个 heat map 上选最高概率的点，按照比例关系映射回原图，然后每条边上有三个点，做一个这个样子的概率平均来得到四条边（话说这里真的是 $1/N$ 吗 ... 预测概率是有可能特别低的噢）：

$$\begin{aligned} x_l &= \frac{1}{N} \sum_{j \in E_1} x_j p_j, & y_u &= \frac{1}{N} \sum_{j \in E_2} y_j p_j \\ x_r &= \frac{1}{N} \sum_{j \in E_3} x_j p_j, & y_b &= \frac{1}{N} \sum_{j \in E_4} y_j p_j \end{aligned} \quad (2)$$

Grid Points Feature Fusion.

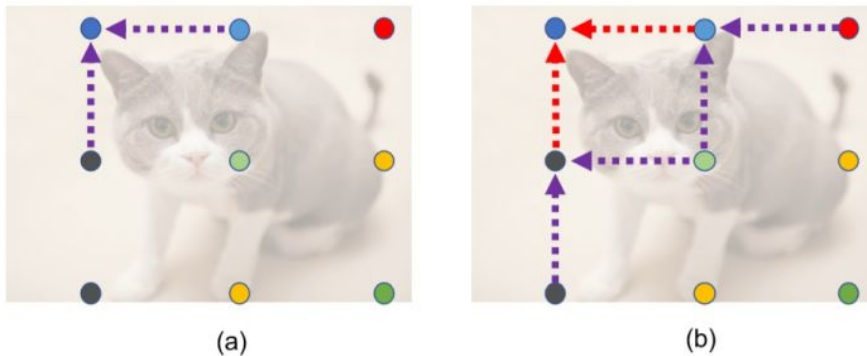


Figure 3. An illustration of the 3×3 case of grid points feature fusion mechanism acting on the top left grid point. The arrows represent the spatial information transfer direction. (a) First order feature fusion, feature of the point can be enhanced by fusing features from its adjacent points. (b) The second order feature fusion design in Grid R-CNN.

Grid points 之间存在内在关联，考虑充分利用不同 map 间的关系。看上图的 (a)，对于左上角点，可以考虑距离为 1 的 point 的 map 去完成 feature fusion，做法就是把这个 map 经过 5×5 的卷积（理解为提取需要的信息变换）后加到目标点的 map 上去。(b) 中展示了距离为 2 的 fusion 方法，紫色的箭头就理解成距离为 1 的 fusion，红色的箭头表示在这之后的 fusion。

这里我感觉文中也没太讲清楚吧，不知道理解的对不对，欢迎讨论哈。

Extended Region Mapping.

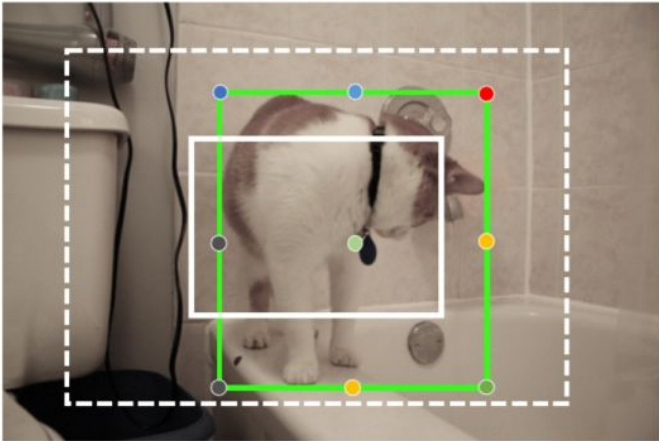


Figure 4. Illustration of the extended region mapping strategy. The small white box is the original region of the RoI and we extend the representation region of the feature map to the dashed white box for higher coverage rate of the grid points in the the ground truth box which is in green.

上图中白色的实线框表示 RPN 给出的 proposal，它不一定能完整的包含物体，这样会影响 multi-point 的预测。如果简单的扩大 proposal 的大小，那么不会带来提升，并且对小物体的检测还有害（在下面的 **Extended Region Mapping** 实验部分可以看到）。作者在这里并没有将其扩大或者缩小，而是在 heat map 上将这个 proposal 覆盖的区域看成是两倍大，这么做的依据，我觉得应该是在 heat map 上的感受野也足够这么表示了。那么这么做之后，只需修改 **Grid Guided Localization** 中点的映射公式即可。

我们选几个实验结果关注一下。

首先看点数目的影响：

| method | AP | AP _{.5} | AP _{.75} |
|--------------|------|------------------|-------------------|
| regression | 37.4 | 59.3 | 40.3 |
| 2 points | 38.3 | 57.3 | 40.5 |
| 4-point grid | 38.5 | 57.5 | 40.8 |
| 9-point grid | 38.9 | 58.2 | 41.2 |

Table 1. Comparison of different grid points strategies in Grid R-CNN. Experiments show that more grid points bring performance gains.

然后可以发现 AP 上的增益主要来源于 strict IoU 段。

Grid Points Feature Fusion:

| method | AP | AP _{.5} | AP _{.75} |
|-----------------------------|------|------------------|-------------------|
| w/o fusion | 38.9 | 58.2 | 41.2 |
| bi-directional fusion [26] | 39.2 | 58.2 | 41.8 |
| first order feature fusion | 39.2 | 58.1 | 41.9 |
| second order feature fusion | 39.6 | 58.3 | 42.4 |

Table 2. Comparison of different feature fusion methods. Bi-directional feature fusion, first order feature fusion and second order fusion all demonstrate improvements. Second order fusion achieves the best performance with an improvement of 0.7% on AP.

Extended Region Mapping:

| method | AP | AP _{small} | AP _{large} |
|-------------------------|------|---------------------|---------------------|
| baseline | 37.7 | 22.1 | 48.0 |
| enlarge proposal area | 37.7 | 20.8 | 50.9 |
| extended region mapping | 38.9 | 22.1 | 51.4 |

Table 3. Comparison of enlarging the proposal directly and extended region mapping strategy.

COCO 测试集:

| method | backbone | AP | AP _s | AP _m | AP _L |
|-------------------------|--------------------------|------|-----------------|-----------------|-----------------|
| YOLOv2 [14] | DarkNet-19 | 21.6 | 44.0 | 19.2 | 5.0 |
| SSD-513 [15] | ResNet-101 | 31.2 | 50.4 | 33.3 | 10.2 |
| DSSD-513 [16] | ResNet-101 | 33.2 | 53.3 | 35.2 | 13.0 |
| RefineDet512 [17] | ResNet101 | 36.4 | 57.5 | 39.5 | 16.6 |
| RetinaNet800 [18] | ResNet-101 | 39.1 | 59.1 | 42.3 | 21.8 |
| CornerNet | Hourglass-104 | 40.5 | 56.5 | 43.1 | 19.4 |
| Faster R-CNN+++ [8] | ResNet-101 | 34.9 | 55.7 | 37.4 | 15.6 |
| Faster R-CNN w FPN [4] | ResNet-101 | 36.2 | 59.1 | 39.0 | 18.2 |
| Faster R-CNN w TDM [19] | Inception-ResNet-v2 [22] | 36.8 | 57.7 | 39.2 | 16.2 |
| D-FCN [20] | Aligned-Inception-ResNet | 37.5 | 58.0 | - | 19.4 |
| Regionlets [21] | ResNet-101 | 39.3 | 59.8 | - | 21.7 |
| Mask R-CNN [5] | ResNeXt-101 | 39.8 | 62.3 | 43.4 | 22.1 |
| Grid R-CNN w FPN (ours) | ResNet-101 | 41.5 | 60.9 | 44.5 | 23.3 |
| Grid R-CNN w FPN (ours) | ResNeXt-101 | 43.2 | 63.0 | 46.5 | 25.1 |

Table 6. Comparison with state-of-the-art detectors on COCO test-dev.

不同 IoU 段上的提升: (增益主要来自于高 IoU段)

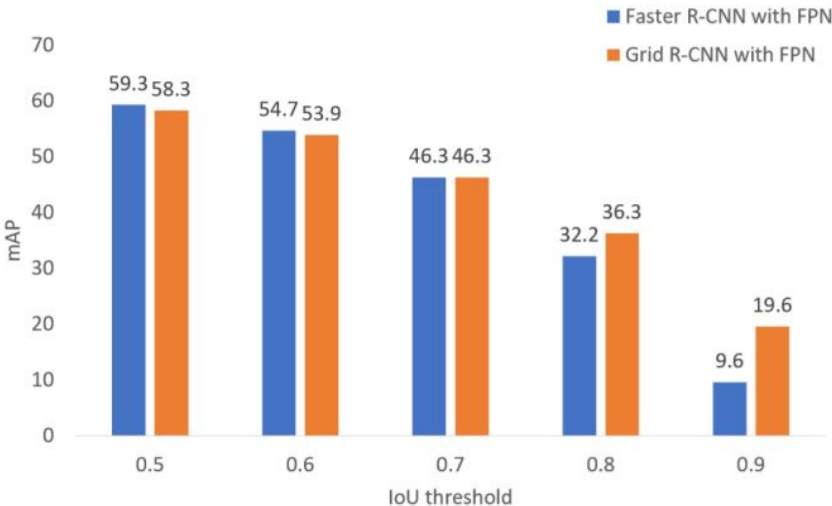


Figure 5. AP results across IoU thresholds from 0.5 to 0.9 with an interval of 0.1.

作者也对 COCO 中的不同类别物体在 AP 上得到的提升进行了统计, 发现规则物体 (keyboard, laptop, fork, train, and refrigerator) 得到的提升较大, 而对于球形物体 (sports ball, frisbee, bowl, clock and cup) 反而还有了一些下降, more details 去看论文啦。Grid R-CNN 也在 PASCAL VOC 2007 上做了实验, 但是采用了 COCO 的 [0.5:0.95] 标准, 主要还是为了说明在 higher IoU thresholds 上的提升吧。

有两个疑问:

- (1) Grid R-CNN 中引入了大量的额外的参数, 这需要一个说明, 把它们加到 baseline 中表现不会更好。
- (2) 代码会公开吗? 通篇没有看到 Code will be released 字样耶。

最后总结一下吧。目前 Detection 这一块的研究我个人把它们分为两块:

- (1) 特征, 特征, 深度学习也是要做 “特征工程” 的, 各种托马斯回旋的卷积操作;
- (2) 原 pipeline 中某个固定 mechanism 的深入研究后革新, 比如 anchor 这个东西啊对吧 (cornernet) , iou 这个东西啊对吧 (cascade r-cnn) , nms 啊对吧 (iouNet, soft softer

softererer....) , 在边框回归这个方面可能会出现越来越多借鉴关键点检测的思路来做, anchor based detector 会被取代吗? (In defense of anchor-based detector? 喵喵喵?)

Grid R-CNN 属于第二类的工作, 我也喜欢更多的有第二类工作来出现, 它能够更好地帮助窝们开拓思路, 去看到 detection 这个东西的 pipeline 到底怎么设计比较好。

错误之处, 还请批评指正 ~

有人看我就取匿 ٥_٥

编辑于 2018-12-02

▲ 赞同 148 ▼

● 25 条评论

➦ 分享

★ 收藏

♥ 感谢

...

收起 ^

查看全部 11 个回答

<https://www.zhihu.com/question/304322570/answer/545977596>

6/6