QueryDet: Cascaded Sparse Query for Accelerating High-Resolution Small Object Detection

级联稀疏query加速高分辨率下的小目标检测

CVPR2022

汇 报 人： 叶伟欣

汇报时间： 2022年11月29日

内容

1. 引言

➔ 目标检测快速发展，但小目标检测依然是个大问题。

例如 **RetinaNet** 在 **COCO** 上的中目标、大目标检测上实现了**44.1 mAP**，**51.2 mAP**，但对小目标的检测精度只有**24.1 mAP**。造成此情况的原因作者分析有如下**3**条：

- 一方面小目标的特征随着网络的多次下采样而逐渐消失或者被背景特征淹没。
- 低分辨特征的感受野可能与小物体的大小不匹配（参考 **Scale-Aware Trident Networks for Object Detection**）。
- 定位小物体比大物体更困难，因为边界框的小扰动可能会导致对联合交集（IOU）度量的显着干扰。

1. 引言

➔ 补充

其他原因：

（1）小目标在原图中的数量较少，检测器提取的特征较少，导致小目标的检测效果差。

（2）神经网络在学习中被大目标主导，小目标在整个学习过程被忽视，导致导致小目标的检测效果差。

Tricks:

（1） **data-augmentation.**简单粗暴，比如将图像放大，利用 **image pyramid**多尺度检测，最后将检测结果融合.缺点是操作复杂，计算量大，实际情况中不实用;

（2） 特征融合方法：**FPN**这些，多尺度**feature map**预测，**feature stride**可以从更小的开始;

1. 引言

一般来说，小目标检测的精度可以通过放大图像或者减少网络的降采样率来实现。

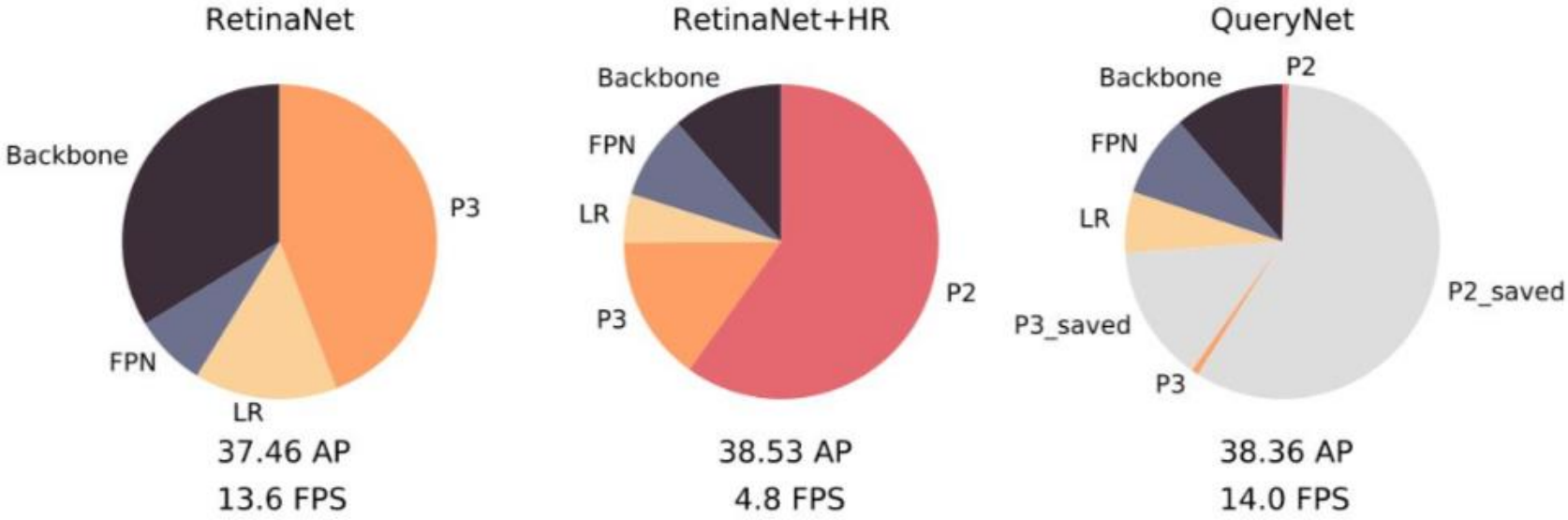
不过放大图像可能会带来后续巨大的计算量，而单纯提高特征图的分辨率也会导致计算量激增。

不同尺度的目标在不同的层次上被处理：大目标倾向于在高层次特征上被检测到，而小目标通常在低层次上被检测到。特征金字塔范式节省了在主干中从浅到深维护高分辨率特征图的计算成本。尽管如此，检测头对低级特征的计算复杂度仍然是巨大的。例如，在**RetinaNet**中添加一个额外的金字塔级别 $\mathbf{P_2}$将在检测头中带来大约 **300%** 的计算量（**FLOPs**）和内存成本。因此在 **NVIDIA 2080Ti GPU**上将推理速度从**13.6 FPS**严重降低到**4.85 FPS**。

1. 引言

➔ 研究背景

- **RetinaNet**有两部分：一个带有**FPN**的输出多尺度特征图的主干网络和两个用于分类和回归的检测头。
- 图中计算量分析可以看到，在原生**RetinaNet**中，分辨率最高的$\mathbf{P_3}$（相对原图**8**倍下采样）占据了最大的计算量（应该包含检测头的计算量），如果将**RetinaNet**的**FPN**进一步拓展到$\mathbf{P_2}$（相对原图**4**倍下采样），那$\mathbf{P_2}$带来的额外计算量更是占据了绝对的大头。



2. 方法论

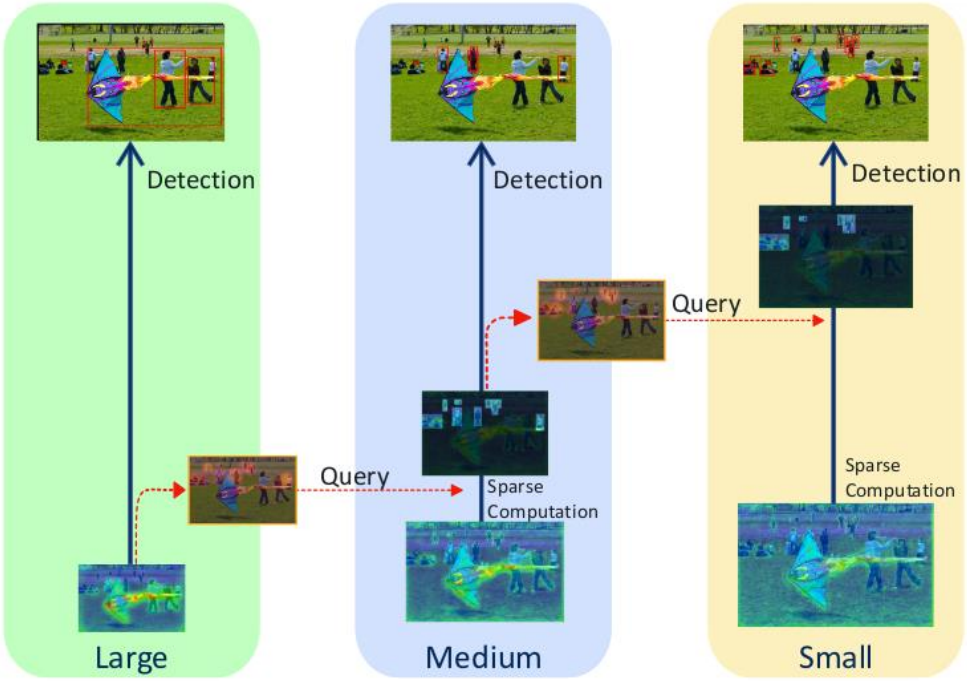
➔ 总体介绍

- 在本文中，提出了一种简单有效的方法 **QueryDet**，以节省检测头的计算量，同时提高小目标的性能。
- 动机来自两个关键观察：**1）**对低级特征的计算是高度冗余的。 在大多数情况下，小目标的空间分布非常稀疏：它们只占据高分辨率特征图的一小部分； 因此浪费了大量的计算。 **2）**特征金字塔（**FPN**）是高度结构化的。这意味着我们无法准确检测低分辨率特征图中的小目标，但我们仍然可以高度自信地推断出它们的存在和大致位置。

3. 方法论

➔ 总体介绍

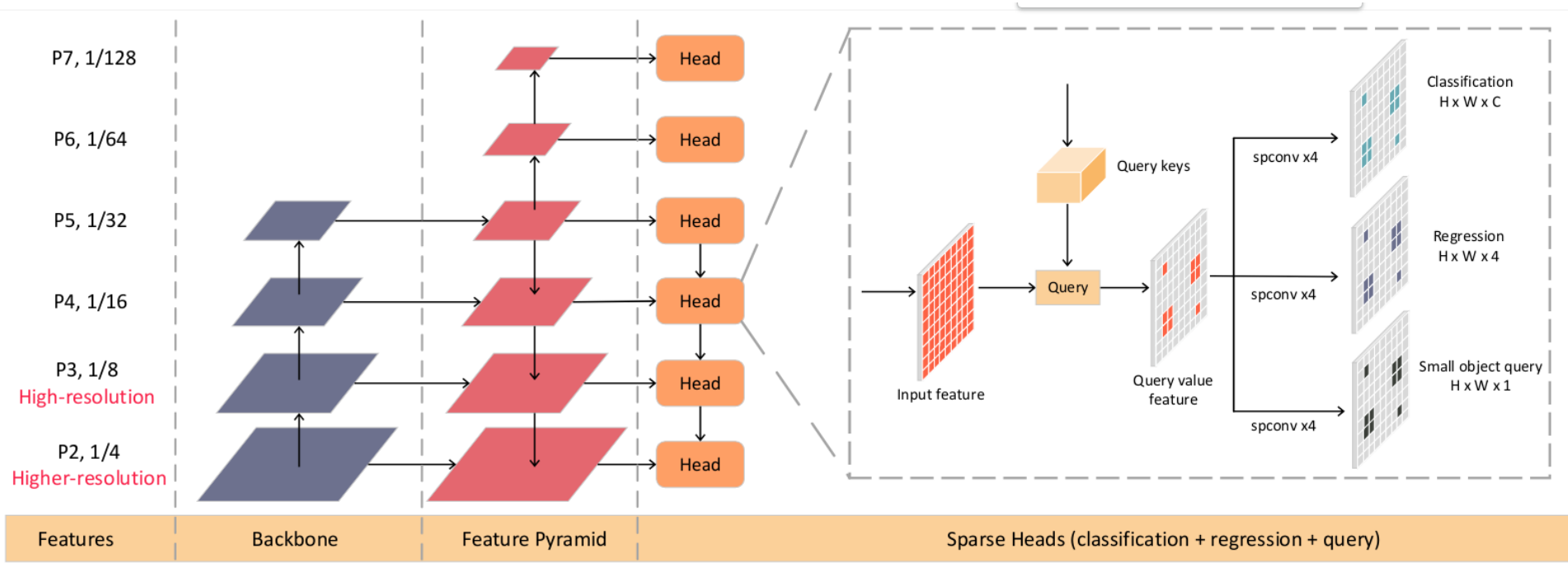
- 根据低层特征的计算高度冗余和特征金字塔高度结构化，**QueryDet**首先预测低分辨率特征上小物体的粗略位置，然后使用由那些粗略位置稀疏引导的高分辨率特征来计算准确的检测结果。这样，不仅可以收获高分辨率特征图的好处，而且还可以避免对背景区域进行无用的计算。 此**pipeline**以级联方式应用，可实现快速，准确的小物体检测。



3. 方法论

➔ 网络结构

- 图像被送到主干和特征金字塔网络（FPN）中，生成一系列具有不同分辨率的特征图。 从查询起始层（此图像中的P_5）开始，每个层从上一层接收一组关键位置，并应用查询操作来生成稀疏值特征图。 然后，稀疏检测头和稀疏查询头会预测所检测到的下一层相应的比例和关键位置的框。



3. 实验

➔ 实验设计

- 用两个数据集比较了本文方法和**RetinaNet**，揭示了检测小目标时使用高分辨率的重要性，然而高分辨率特征图会显著降低推理速度，当采用级联稀疏查询(**CSQ**)时可以提高推理速度(该方法将检测**mAP**提高了**1.0**，将小**mAP**提高了**2.0**，高分辨率推理速度平均提高到**3.0**倍。相较于其他方案亦有明显提升。)

Comparison of accuracy (AP) and speed (FPS) of our QueryDet and the baseline RetinaNet on COCO mini-val set

Method	CSQ	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	FPS
RetinaNet	-	37.46	56.90	39.94	22.64	41.48	48.04	13.60
RetinaNet (3x)	-	38.76	58.27	41.24	22.89	42.53	50.02	13.83
QueryDet	×	38.53	59.11	41.12	24.64	41.97	49.53	4.85
QueryDet	✓	38.36	58.78	40.99	24.33	41.97	49.53	14.88
QueryDet (3x)	×	39.47	59.93	42.11	25.24	42.37	51.12	4.89
QueryDet (3x)	✓	39.34	59.69	41.98	24.91	42.38	51.12	15.94

3. 实验

➔ 实验设计

- **Comparison of detection accuracy (AP) and speed (FPS) of our QueryDet and the baseline RetinaNet on VisDrone validation set**

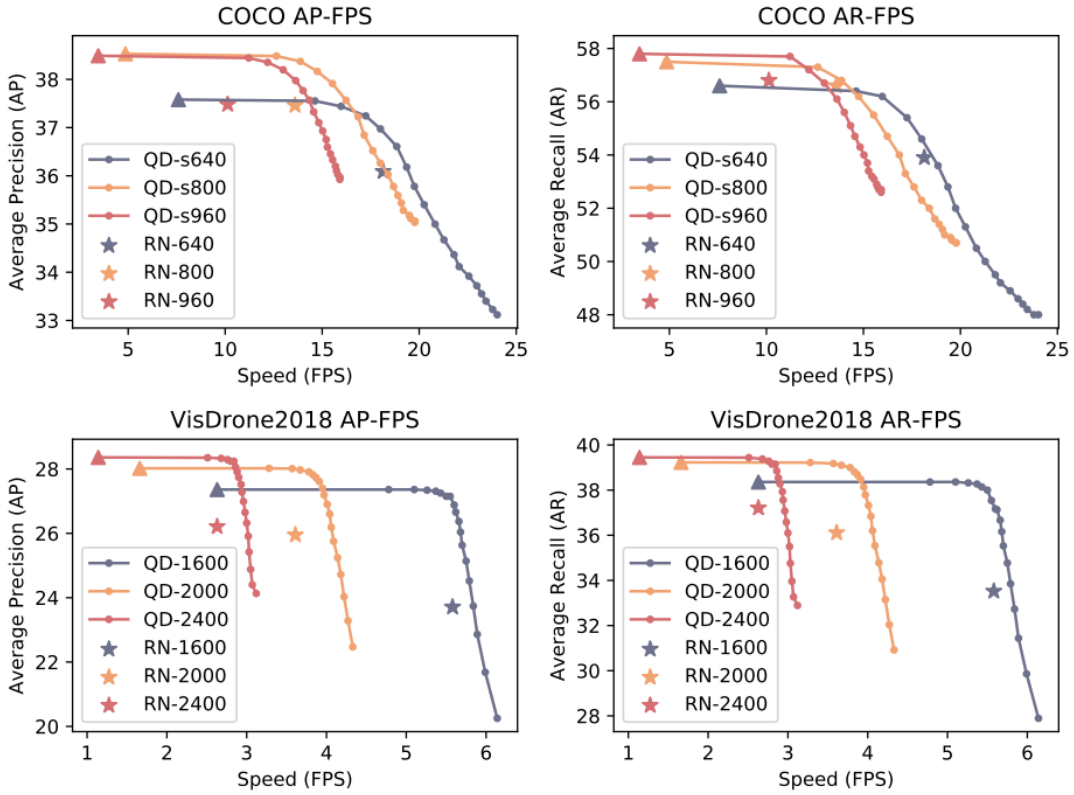
Method	CSQ	AP	AP_{50}	AP_{75}	AR_1	AR_{10}	AR_{100}	AR_{500}	FPS
RetinaNet	-	26.21	44.90	27.10	0.52	5.35	34.63	37.21	2.63
QueryDet	×	28.35	48.21	28.78	0.51	5.96	36.48	39.42	1.16
QueryDet	✓	28.32	48.14	28.75	0.51	5.96	36.45	39.35	2.75

- **Comparison of detection accuracy (AP) and speed (FPS) of our QueryDet and the baseline RetinaNet on VisDrone validation set**

Method	CSQ	AP	AP_{50}	AP_{75}	AR_1	AR_{10}	AR_{100}	AR_{500}	FPS
RetinaNet	-	26.21	44.90	27.10	0.52	5.35	34.63	37.21	2.63
QueryDet	×	28.35	48.21	28.78	0.51	5.96	36.48	39.42	1.16
QueryDet	✓	28.32	48.14	28.75	0.51	5.96	36.45	39.35	2.75

3. 实验

- **The speed and accuracy (AP and AR) trade-off with input images with different sizes on COCO and VisDrone. The trade-off is controlled by the the query threshold σ. The leftmost marker (the ▲ marker) of each curve stands for the result when Cascade Sparse Query is not applied. QD stands for QueryDet and RN stands for RetineNet**



4. 实验

- 可视化了 **COCO** 和 **VisDrone** 上小目标的检测结果和查询热力图。从热力图中可以看出，我们的查询头可以成功找到小目标的粗略位置，使我们的 **CSQ** 能够有效地检测到它们。 此外，通过结合高分辨率特征，我们的方法可以非常准确地检测小目标。

(a) COCO detection (b) COCO query heatmap (c) VisDrone detection (d) VisDrone query heatmap

4. 结果与讨论

➔ 结果

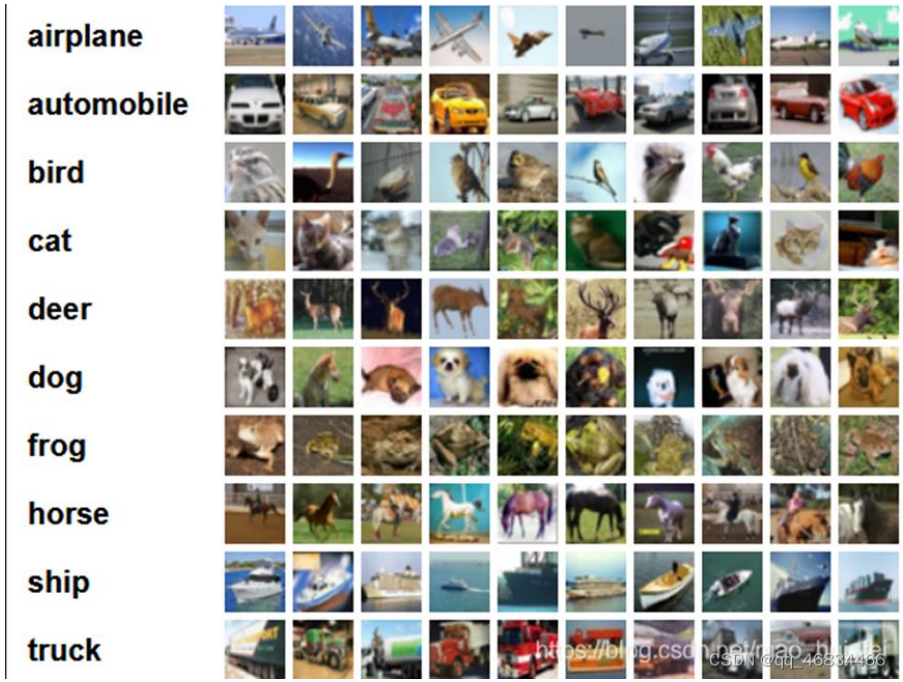
- 提出 **QueryDet**，它使用一种新颖的查询机制级联稀疏查询 **(CSQ)** 来加速基于特征金字塔的密集目标检测器的推理。 **QueryDet** 使目标检测器能够以低成本检测小目标并易于部署，使其能够在自动驾驶等实时应用程序中部署。 对于未来的工作，我们计划将 **QueryDet** 扩展到以 **LiDAR** 点云作为输入的更具挑战性的 **3D** 目标检测任务，其中 **3D** 空间通常比 **2D** 图像更稀疏，并且计算资源对于昂贵的 **3D** 卷积操作来说更加密集。

➔ 启发

- 图像、特征图空间稀疏特性的新用法。

Cifar10

CIFAR-10 是一个用于识别普适物体的小型数据集。一共包含 **10** 个类别的 **RGB** 彩色图 片：飞机（**airplane**）、汽车（**automobile**）、鸟类（**bird**）、猫（**cat**）、鹿（**deer**）、狗（**dog**）、蛙类（**frog**）、马（**horse**）、船（**ship**）和卡车（**truck**）。图片的尺寸为 **32×32**，数据集中一共有 **50000** 张训练圆片和 **10000** 张测试图片。 **CIFAR-10** 的图片样例如图所示。



Cifar10

Table 1. model performance based on Cafir10

Model	EPOCH	LR	Train	Test
VGG16	2000	0.001	0.955	0.900
VGG19	2000	0.001	0.964	0.903
ResNet18	2000	0.001	0.960	0.904
ResNet34	2000	0.001	0.964	0.903
GoogleNet	2000	0.001	0.9999	0.9497

本周主要工作和下周计划

1. 近期工作

1）.目标检测的基础知识学习（YOLO和RNN系列）；

2）.学习overleaf模板使用；

3）.使用VGG,GoogleNet,ResNet训练Cifari10;

4）.毕设论文（第一、二、三章的撰写），模型的建立（分类）；

2. 下周计划

1）.申博资料的准备

2）.论文审稿意见回复；

3）.目标检测论文阅读；

4）.毕设数据处理与撰写（第三章完善）；

谢　谢!

欢迎批评指正