

基于分割的行人检测

一、 背景

行人检测作为一种重要的目标检测应用，在自动驾驶、视频监控、刑事侦查等领域得到了广泛的关注。按照实现原理，可将行人检测算法分为三类：基于运动检测算法，基于机器学习的算法，基于深度学习的算法。早期的行人检测方法，如 HOG 检测器，ICF 检测器，在特征表示，分类器的设计，检测加速度方面，为一般的目标检测奠定了坚实的基础。近年来，一些通用的目标检测算法，如 Faster RCNN，已经被引用到行人检测中，极大地推动了该领域的研究进展。不过现存的深度学习目标检测算法普遍需要 Region proposal 机制，尽管通常在正常环境下能够达到很好的效果，但带极端环境下召回率很低，这主要因为 bounding box 的标注方式把很多环境噪声也包含了进来，除此之外还需要用 NMS 来筛选这些 bounding box。因此，用语义分割或实例分割来进行行人检测具有以下优势：(i) 这是一种 anchor free, NMS free 的检测模型，避免了所有与 bounding box 以及 NMS 相关的超参数。(ii)语义分割和实例分割是像素级别的分割，能够达到更精准的描述行人，避免了上述的环境噪声问题。

二、 实验设计

(一) 模型选择

本文采用 Unet 语义分割模型，Unet 模型最初是为解决生物医学图像分割的问题而提出的模型，因为在生物医学中只能获得很少的训练数据，这个模型在已有数据上应用弹性变形来对数据增强。换句话说，可以用相对较少的数据来训练。

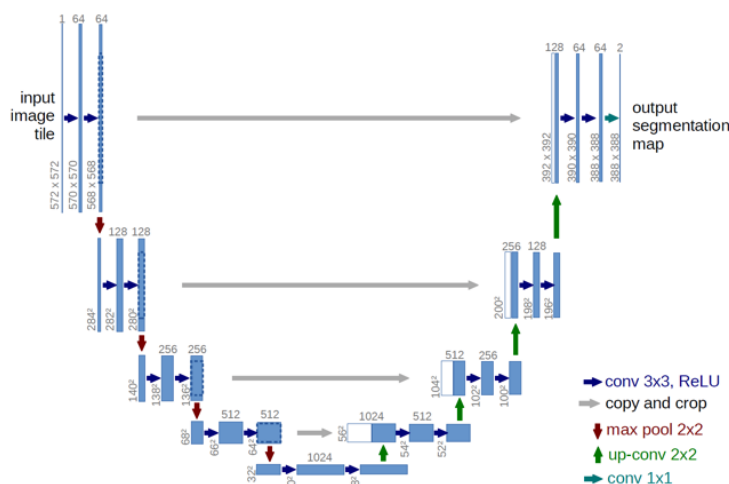


图 1 Unet 模型结构

(二) 数据集选择

选择 Cityscape 数据集，Cityscape 划分的训练集中分为了 18 个城市，读取 18 个城市的所有训练图像，共 2975 张图像作为训练集。用划分好的测试集中图像进行测试。

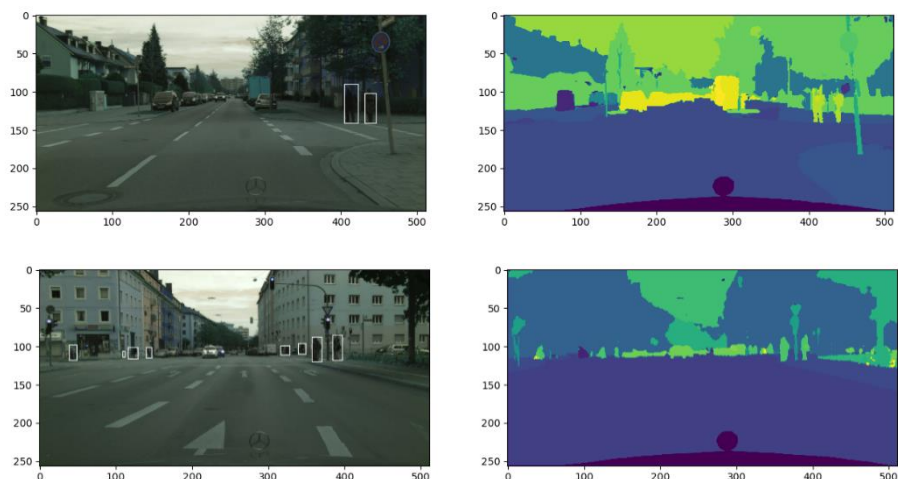


图 2 训练数据示例

(三) 标注 bounding box

现得到了分割出的 label 图，将行人的部分提取出来则可用框标注出行人的 bounding box。在这一部分显示出了语义分割用于行人检测的一些缺点：当行人比较密集是 label 图中行人会连在一起，没法区分开每个单独的人。如果用实例分割可避免这个问题。

三、 结果展示



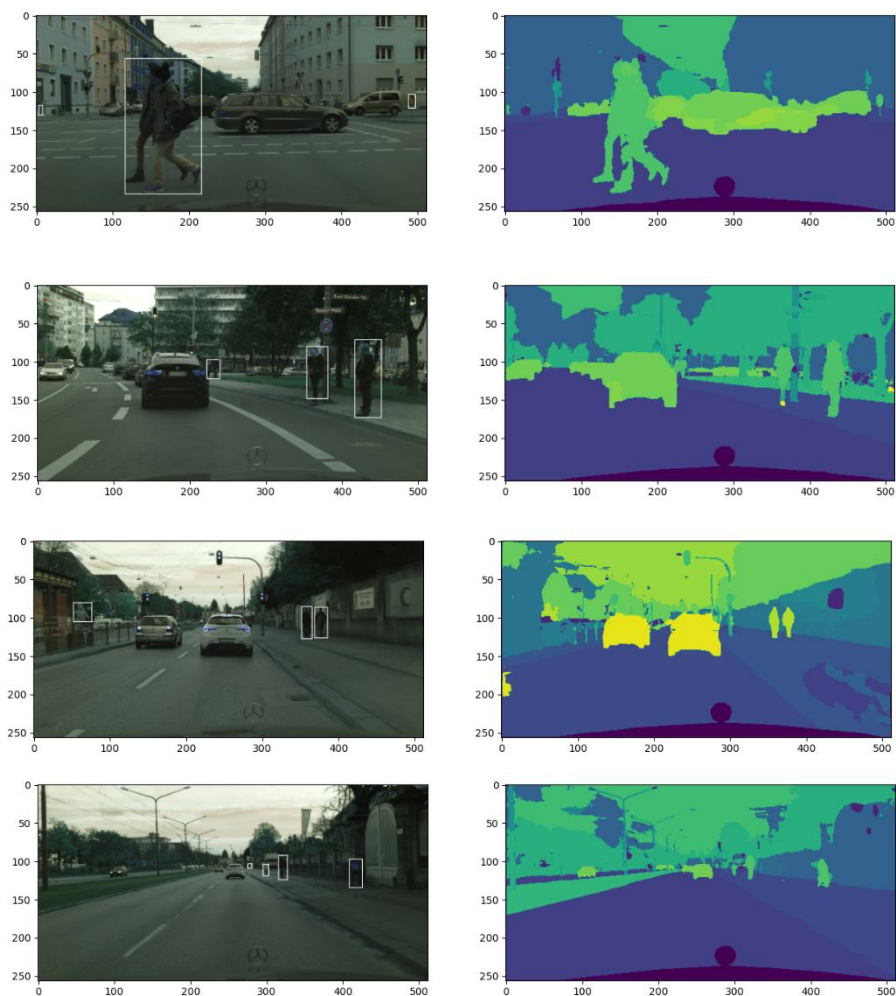


图 3 左侧根据分割结果在原图中标注出 bounding box，右侧是测试集上分割出的 label 图。语义分割能够检测出很远处

处的行人，但是行人挤在一起时检测框会合并

从以上展示的结果中可以很清楚的发现，由于分割是以像素为单位，即便是非常远处的行人，在图像中只占非常小部分，也能够分割出来。另一方面，这种方法不适用于人群密集的场景，因为行人挤在一起时只能标记出一个 bounding box。

四、 总结

尝试了一种用分割来进行行人检测的方法，它是 anchor free 和 NMS free 的。NMS 是现有深度学习方法在图像中目标检测的瓶颈之一。需要调优 NMS 上的超参数，这严重阻碍了高性能检测框架的延展性。因此，在 Cityscape 数据集上的实验结果表明，运用 Unet 语义分割结果进行行人检测能够取得很好的效果，不足在于没法区分相互遮挡的行人，不过这一问题可通过实例分割解决，并不影响分割方法用于行人检测的可行性。