

北京理工大学

本科生毕业设计（论文）开题报告

学 院： 计算机学院

专 业： 软件工程

班 级： 08111606

姓 名： 陶润洲

指导教师： 赵三元

校外指导教师：

二〇二〇年一月九日

毕业设计（论文）开题报告评审表

姓名	陶润洲	学号	1120162055	班级	08111606	专业	软件工程
导师	赵三元	校外导师（职称）				校外导师单位	
论文选题	题目名称	无锚点框的全卷积单阶段目标检测方法研究					
	题目性质	软件开发（ ） 理论研究（√）					
	题目来源	结合科研 （ ） 结合生产实际（ ） 结合实验室建设（ ） 自拟题目 （√）					
评审组成员	姓 名	职 称	工作单位及职务			签 字	
评审意见	(含：选题意义；选题是否满足毕业要求；技术方案是否可行；进度安排是否合理等)						
成 绩							
评审组长签字： <div style="text-align: right;">年 月 日</div>							

注：成绩以“合格”“不合格”记；评审组长为高级职称人员。

无锚点框的全卷积单阶段目标检测方法研究

1. 毕业设计（论文）选题的内容

目标检测作为计算机视觉中的一个重要研究方向，广泛应用于自动驾驶和智能监控等领域。本课题主要研究基于 anchor-free 的全卷积单阶段目标检测，通过 anchor-free 高效快速的解决目标检测问题。该方法不依赖预先定义的锚点框或者提议区域，避免了关于锚点框的复杂运算和与锚点框有关且对最终检测结果非常敏感的所有超参数，实现了高效快速的目标检测，比以往基于锚点框的一阶检测器更加简单。

2. 研究方案

2.1 本选题的主要任务

了解深度学习、目标检测和无锚点框思想相关应用领域背景知识，研究基于无锚点框的全卷积单阶段目标检测，通过无锚点框高效快速的解决目标检测问题。

采用特征图金字塔网络 FPN 结构来消除因为出现较多重叠框而产生的模糊现象，并设计出新分支“center-ness”打压距离目标中心较远位置的一定数量的低质量边界框，预测出一个像素对应边框中心的偏差，将所得的分数用于降低低质量的检测框，最后通过非极大值抑制 NMS 将检测结果进行融合。

2.2 技术方案的分析、选择

锚点框 anchor box 是在 Faster R-CNN 中提出的一个概念，目的在于取代原来的区域提案建议，用于解决一个窗口无法检测多个目标以及无法解决多尺度的问题。目标检测不同于目标识别，需要通过锚点框在训练和预测过程中来对目标进行位置的判定。而无锚点框思想 anchor free 就是不使用锚点框，通过将图片分割成 $N \times N$ 个网格，然后基于中心点或者关键点来进行目标的检测。

无锚点框思想也刚好符合单阶段目标检测的思想，单阶段目标检测，特点是一步到位，速度相对较快；相对于多阶段目标检测，需要选举候选框再进行分类，计算量相对较大，而单阶段检测方法，仅仅需送入一次网络就可以预测出所有的边界框。

而对于全卷积网络 FCN，对比于 CNN 而言，去除了全连接层，将原先所有的全连

接层都换成卷积层，最大程度上保留了图片的特征，对图像进行像素级的分类。FCN可以接受任意尺寸的输入图像，采用反卷积层对最后一个卷积层的特征图进行上采样，使它恢复到输入图像相同的尺寸，从而可以对每一个像素都产生预测，同时保留了原始输入图像中的空间信息。全卷积是从抽象的特征中恢复出每个像素所属的类别，即从图像级别的分类进一步延伸到像素级别的分类。

无锚点框的全卷积单阶段目标检测网络，类似于语义分割，针对每个像素进行预测。无锚点框不依赖于预先定义的锚点框或者区域提案，避免了与锚点框的复杂运算，锚点框会引入很多需要优化的超参数，并且为了检测的效果，需要大量的锚点框，这样一来就会造成正负样本类别不均衡的问题。在训练过程中，还需要计算锚点框和真实边界框的交并比，计算量极大。该网络则可以避免这些复杂计算，使得检测速度加快。

对于该目标检测算法而言，首先将输入的图片送入骨干网络之后，获得最终的特征图，基于锚点框的目标检测算法会在特征图的每个位置上使用预先定义好的锚点框来进行目标检测，但是该目标检测算法，通过直接对特征图上的每一点进行回归操作，来进行目标检测。

首先将特征图中的每一个点 (x, y) 映射回原始的输入图像中，然后如果这个映射回原始输入的点在相应的真实边界框范围之内，而且类别标签相对应，我们将其作为训练的正样本块，否为将其作为负样本块。再接着计算出该点到真实边界框的左边框，上边框，右边框，下边框的距离，即算法回归的目标，记为 (l, t, r, b) 。如果一个点在多个真实边界框的内部时，该样本为模糊样本，则选取区域最小的真实边界框作为目标；但是由于网络中使用了特征金字塔网络，这将导致模糊样本的数量大大减少。

当位置 (x, y) 与一个真实边界框相关联时，该点位置处的训练回归目标则为：

$$\begin{aligned} l^* &= x - x_0^{(i)}, t^* = y - y_0^{(i)} \\ r^* &= x_1^{(i)} - x, b^* = y_1^{(i)} - y \end{aligned}$$

其中 (x_0, y_0) ， (x_1, y_1) 分别表示真实边框的左上角和右下角坐标。通过这样的回归方式可以获得大量的正样本块，然后使用这样正样本块进行回归操作，获得比较好的性能提升。

对于 loss 函数，其公式为：

$$L(\{p_{x,y}\}, \{t_{x,y}\}) = \frac{1}{N_{pos}} \sum_{x,y} L_{cls}(p_{x,y}, c_{x,y}^*) + \frac{\lambda}{N_{pos}} \sum_{x,y} I_{\{c_{x,y}^* > 0\}} L_{reg}(t_{x,y}, t_{x,y}^*)$$

L_{cls} 为分类损失，采用的 Focal loss； L_{reg} 回归损失，采用的 IOU loss。

该目标检测方式会因为距离目标中心较远的位置而预测出大量低质量的边框。所以算法在不引起新参数的情况下来抑制这些低质量的边框。平行于分类分支增加了一个新的分支称为 center-ness 中心度，用于预测该位置到目标物中心位置的距离。其回归定义为：

$$centerness^* = \sqrt{\frac{\min(l^*, r^*)}{\max(l^*, r^*)} \times \frac{\min(t^*, b^*)}{\max(t^*, b^*)}}$$

center-ness 的取值范围为 [0, 1]，使用交叉熵损失进行训练。并把损失加入前面提到的损失函数中。测试时，将预测的中心度与相应的分类分数相乘，计算最终得分（用于对检测到的边界框进行排序）。因此，中心度可以降低远离对象中心的边界框的权重。因此，这些低质量边界框很可能被最终的非最大抑制（NMS）过程滤除，从而显着提高了检测性能。

在训练阶段，采用 ResNet-50 作为骨干网络，使用随机梯度下降优化器，初始学习率为 0.01，批次大小 batch_size 为 16，在迭代到 60k 和 80k 的时候权重衰减为 0.0001 和 0.9，使用 ImageNet 预训练权重进行初始化，将输入图片裁剪为短边不小于 800 像素，长边不小于 1333 像素大小。而整个网络则是在 COCO 数据集上面训练得到的。

2.3 实施技术方案所需的条件

- 操作系统：Ubuntu18.04 及以上操作系统（32 位或 64 位）
- 硬件环境：内存 8GB-16GB
- 处理器：Intel Core i5-8300H, 3900 MHz 及以上
- 显卡：Nvidia GeForce GTX 1050 ti 及以上
- 开发语言：Python3.6 以上
- 使用框架：Pytorch1.0 以上

2.4 存在的主要问题和关键技术

- 在以往基于锚点框的算法中，采用较大的步伐会导致召回率偏低，需要通过调整交并比阈值来补偿，该算法通过采用全卷积在步伐较大的情况下也能获得较好的召回率。
- 在训练时，多个真实边界框的重叠会造成难以理解的模糊情况。
- 对于 Pytorch 框架不够熟悉

目前存在的主要问题是：显卡算力不足

2.5 预期能够达到的研究目标

实现出一个基于无锚点框的全卷积单阶段目标检测系统，采用特征图金字塔网络 FPN 结构来消除因为出现较多重叠框而产生的模糊现象，并设计出新分支“center-ness”打压距离目标中心较远位置的一定数量的低质量边界框。通过该系统能达到高效目标检测的目的。

3. 课题计划进度表

- 阅读深度学习、目标检测相关 paper。（第 1 周-第 2 周）
- 学习相关框架 Pytorch 知识。（第 3 周-第 4 周）
- 理解相关无锚点框、全卷积算法思想。（第 5 周-第 6 周）
- 搭建系统架构并完成模块编码。（第 6 周-第 12 周）
- 构建相应的训练集，测试集和验证集，进行训练测试及验证。优化算法，完成实验。（第 12 周-第 14 周）
- 完成毕业论文，提交软件及相关文档。（第 14 周-第 15 周）
- 完成本科生毕业设计（论文）外文翻译；（第 1 周-第 15 周）
- 完成本科生毕业设计（论文）答辩；（第 1 周-第 15 周）

4. 参考文献

- [1] Tian Z, Shen C, Chen H, et al. FCOS: Fully Convolutional One-Stage Object Detection[J]. arXiv preprint arXiv:1904.01355, 2019.
- [2] Law H, Deng J. Cornernet: Detecting objects as paired

- keypoints[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 734-750.
- [3] Zhou X, Zhuo J, Krahenbuhl P. Bottom-up object detection by grouping extreme and center points[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 850-859.
- [4] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.
- [5] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.