

基于 Faster R-CNN 的城市道路目标检测

ASHIKUR RAHMAN MOHAMMAD 李军

(重庆交通大学 机电与车辆工程学院 重庆市 400047)

摘要: 本文针对当前城市道路复杂场景下目标检测精度低下问题, 提出了一种多尺度 Faster R-CNN 模型应用于城市道路目标检测, 为提高了特征图卷积后的特征精度, 引入了特征金字塔结构, 以融合不同尺度特征, 以提高检测精度。通过区域建议网络 (RPN) 解决了区域对物体或背景的判断, 同时利用回归模型对目标框进行修正。在 BDD 100K 数据集进行实验验证, 结果表明本文方法与常见的目标检测方法检测的查准均都高。

关键词: Faster R-CNN; 目标检测; 特征金字塔; RPN

近几年来电动汽车的逐渐普及, 电动汽车与传统汽车相比, 除了绿色环保外, 其更多应用了大数据、机器学习等技术, 智能驾驶即通过机器学习或深度模型, 分析大数据样本来实现的, 其与一般的机器学习场景相比, 其时效性、运算性能要求更高, 5G 技术的出现, 为汽车智能驾驶提供了数据感知与数据通信基础。当前目标检测成为计算机视觉领域的重要研究方向, 研究目标就是在图像中的位置确定目标所属类型。道路目标即车辆在行驶过程中具有较大意义的目标, 如车辆、行人、障碍物等, 对目标进行定位、识别是智能驾驶的重要环节。随着城市化进程的发展, 城市道路交通流量越来越多, 需要识别的道路目标越来越多。

目标检测已成为智能驾驶的重要技术, 通过目标检测能够自动、高效地识别出道路前方的目标, 为智能汽车的变速、变道等操作提供支持。卷积神经网络对图像进行卷积、池化、输出等多层组合操作, 完成提取图像语义, 处理过程具有特征提取高效、泛化能力强特点。目标检测算法中, 传统的算法分为一阶段目标检测和两阶段目标检测两种, 其中一阶段目标检测是预测图片中圈定范围内的类型、偏移量等, 如 YOLO^[1]、SSD 算法^[2]; 两阶段目标检测算法是首先做初步检测, 然后生成候选区域并对其进行详细分类和微调, 如 Fast R-CNN。本文在 Fast R-CNN 模型基础上, 提出了基于 Faster R-CNN 模型的城市道路目标检测方法。Faster R-CNN 是在 Fast R-CNN 基础上提出了基于 RPN (Region Proposal Network) 候选区域选择, 同时把候选区域、选取特征图、检测目标和位置放在同一个网络模型中, 提高了检测效

率和检测的准确性, 提高了检测模型的性能。

1 Faster R-CNN的基本原理

Faster R-CNN 网络模型包括 4 部分, 分别为卷积层部分、RPN 层、池化层和全连接 FC 层, Faster R-CNN 检测流程图如图 1 所示。首先输入图像, 经过 VGG16 的特征提取, 来获取区域建议网络 (RPN) 需要的全部特征。RPN 利用特征图生成若干个候选区域以及每个区域的概率值, 按照非极大值抑制方法对候选区域的概率值进行计算, 经过 RPN 网络产生的候选区域进行映射处理, 投影到特征图上, 通过投影获得特征矩阵, 把得分最高的区域放入感兴趣区域的池化层 ROI pooling (Region of interest)。候选框的特征图经 ROI pooling 处理后, 缩放到 7×7 大小, 分别送入目标分类和回归两个分支, 计算出分类结果, 同时通过回归预测, 获取边界框区域, 即通过全连接层输出预测类别和位置。

Faster R-CNN 结构中, 特征图经过 RPN 层后, ROI 其产生的 proposal 对应的 box 框。ROI pooling 具体步骤如下: 首先, 根据输入的图像, 把 ROI 映射到 feature map 对应的位置上, 然后把映射后的区域划分为大小统一的 sections, 其数量与输出维度是一致的, 最后对每个 sections 完成 max pooling 操作。在 R-CNN 模型中, 在获取最终的特征后, 先采用二分类模型对类型进行判断, 然后通过 bounding-box 进行回归分析, 获取其准确位置, 这两步骤属于流程操作, 所以影响了网络模型的性能。在 Faster R-CNN 中, 把分类和回归两个任务合二为一, 即通过 multi-task 模型, 提高了网络检测性能。

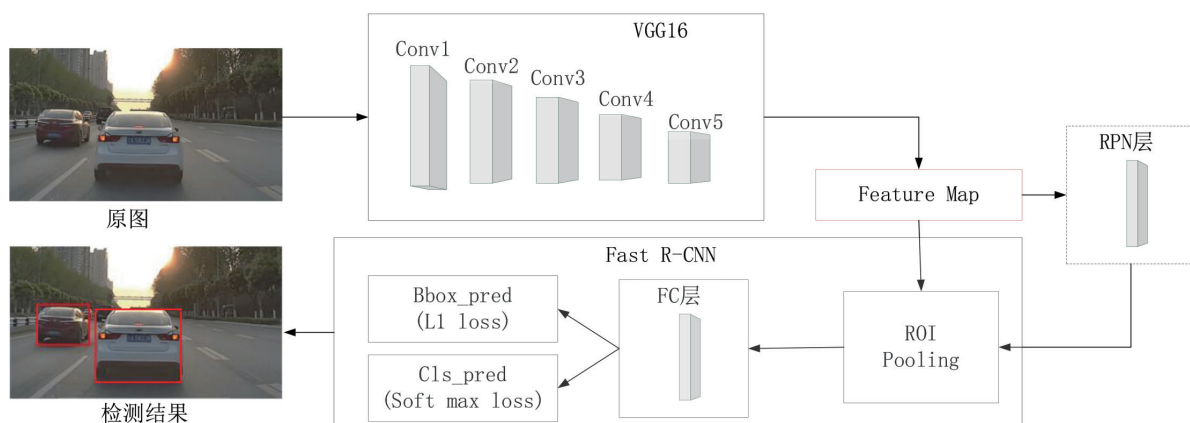


图 1: 基于 Faster R-CNN 网络结构的目标检测流程图

RoI Pooling 对原始图片的尺寸选择可为任意，为了减少因图片缩放产生的特征损失问题，池化层可把不同尺寸的图片产生的特征图统一转为尺寸固定的特征图，以便更好地使全连层和分类层更好的计算。

2 检测区域特征融合

Faster R-CNN 的核心网络采用了多层卷积分步提取了图像的语义，按照预置的语义分类生成多个特征图，在网络结构中，采用了 VGG16 作为卷积和分类处理的网络，其包括 5 个卷积模块和 1 个分类模块。由于 VGG16 网络层数较少，致使提取的图像特征信息不全，因此需要在原始 Faster R-CNN 网络结构基础上，对其进行替换，可用 ResNet50 作为特征提取的网络结构。

在 Faster R-CNN 网络结构中，候选区域的特征仅仅通过卷积层来获取，经过多层的卷积处理后，深层特征中基本不存在一些弱小目标的语义信息。为解决此类问题，提高卷积后图像特征精度，引入了特征金字塔结构（Feature Pyramid Networks, FPN）思想，采用了多尺度特征融合方法^[3]，对网络结构中多种尺度特征图分别完成上采样和下采样工作，特征图融合后，根据不同尺度的检测目标在相应层的融合特征图中生成候选区域。

图 2 为基于 FPN 结构的卷积特征融合示意图。

FPN 网络结构为一种全卷积的网络，每层的卷积的特征图的尺度与原始图的尺度比例是固定的，在进行特征融合时，主要步骤包括三个过程：自下而上、自上而下以及同层连接，在特征融合过程中，最后一层来构建层级结构，针对 ResNet 网络，每个过程阶段的最后一

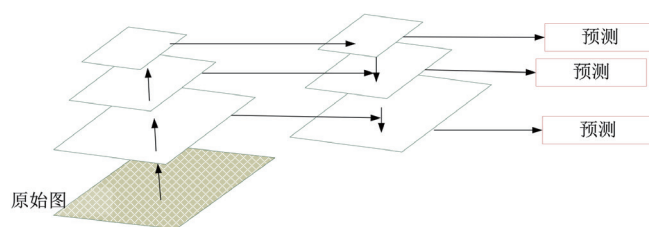


图 2: 基于 FPN 结构的卷积特征融合示意图

个残差块即为金字塔的特征图，因此要求每层的特征图采用相同的通道数，具体操作时可拉加一次 1×1 的特征卷积操作，使之成为相同的通道。

FPN 网络是把图像特征图作为模型的输入，模型中每一层的输出可作为部分信息内容作为下一层的输入项，所以 FPN 具有记忆上下文的特点，对输入数据中的时序特征和涵盖的图片特征语义较为敏感。FPN 可接受事先未规定的不规则的图像特征数据，可接收图像特征任何初始状态，输出结果受因前学习内容的影响，因此某一时刻的输出结果可能融入了以前学习获取特征。从输入输出结构上来看，可通过 1 个或多个输入序列生成 1 个或多个输出图像特征，且即使在相同输入条件下，输出结果也可能存在差异。

3 RPN 网络设计

RPN 替代了 R-CNN 与 Fast R-CNN 的选择性搜索方法，RPN 用于较快地生成候选区域和区域概率值（得分）。图 3 为 RPN 网络结构图。在图中，滑动窗口扫描特征图，生成低维特征向量，滑动过程时计算 bm 窗口与原图像中心点的对应，以中心点为锚生成不同比例尺的锚框，然后把低维特征向量分别向二分类层和回归层进行

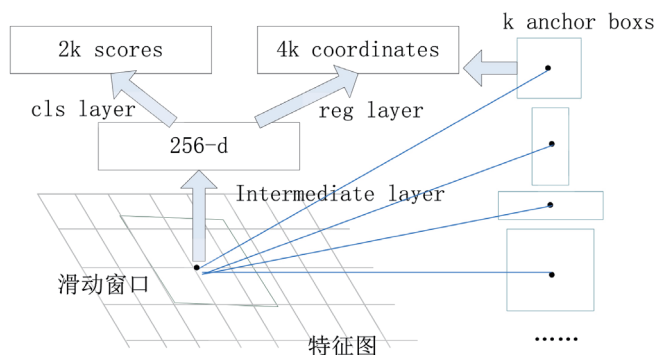


图 3: RPN 网络结构图

输入,二分类层对检测目标是否为背景或物体进行概率判断;回归层用于预测边界框的位置。RPN 节省了计算时间成本,提高了特征卷积速度。

对于卷积特征图中的所有像素点,都配置 k 种 anchors 作为初始化的检测框,其目的就是通过判断 anchor 是否覆盖检测目标来判断该框属于物体还是归属于背景,物体和背景问题属于二分类问题,即 cls layer 生成 $2k$ 个 scores,对应的坐标修改为 4 个值,所以 reg layer 生成 $4k$ 个 coordinates。

RPN 网络通过特征图上的滑动窗口,为每种状态生成多个目标框,这些目标框通过 128×125 、 256×256 和 512×512 这三种 anchor,按两两组合形成的共计 9 种 anchor,这些 anchor 的尺寸是不同。目标检测 Faster R-CNN 网络中,主要解决两方面的问题,第一个问题是通过二分类法对特征进行类型判断,识别目标是物体还是背景,给出每种类型的概率,另一个问题就是利用回归算法,对目标框进行修正,使目标框位置坐标更准确。

(1) 判断物体与背景时采用非极大值抑制的方法,并交比 IoU ((Intersection over Union) 的阈值值为 0.75,即预测的边框与实际边框的交集面积与并集面积的比值为 0.75, IoU 大于等于 0.75 的物体,相反如果 IoU 小于 0.25 时可视为背景。算法设计时把介于 0.25 至 0.75 的 anchor 舍弃掉,只保留小于 0.25 与大于等于 0.75 的 anchor。保留约为 2000 个的 anchor,然后取前 N 个 box 传输给 Fast-R-CNN。在进行二分类判别时,利用激活函数 softmax 来判断类型,在进行 softmax 处理的前后,都通过 Reshape 函数对每张特征图展平为向量,使

softmax 进行二分类处理时,降低了计算复杂度,提高了分类处理的性能。

(2) 目标框修正,通过缩放、平移等方式找到原 anchor 与真值框存在的映射关系,当 anchor 接近真值框时,可作为线性变换,采用线性回归模型对候选 anchor 进行修改,完成 anchor 调整后,可以计算出精确尺寸的 anchor。然后按照物体区域的概率值进行排序,剔除一些宽度、高度很小的无效 anchor,采用非极大值抑制算法,得到前 Top- N 的 anchors,把这 N 个 anchors 作为 RoI Pooling 层中。

4 图像多变化处理设计

由于输入的图像的尺寸是变化的,有时出现检测目标的角度、位置也是变动的,致使训练过程中模型缺少不同场景下的泛化能力。针对此问题,论文提出基于多变化图像处理模块,用于提高复杂场景下的城市道路目标检测效果。模块主要功能实现以下图像处理:

(1) 图像尺寸放缩处理:通过对尺寸放缩处理,以尽可能减少每个检测目标尺寸的差异,保证卷积、池化过程中的标准统一,以提高检测准确性。

(2) 图像翻转平移处理:对采集的检测目标样本进行翻转或平移处理,以增加训练样本的视角和位置的多样化。

(3) 调整样本图像的对比度、亮度等指标,以降低检测效果受光照环境影响的程度,以验证不同亮度、对比度下的检测效果。

(4) 增加图像的噪声,在图像中增加一些干扰像素,以提高训练网络模型的鲁棒性。

这些对图像进行处理的方式,可随机进行处理后得到新的补充数据,随后对新图像得到的检测结果与当原结果进行叠加分析,以提高目标检测的精度。

5 实验验证与分析

5.1 数据集

本文采用伯克利大学的 BDD 100K 数据集进行数据训练和验证模型,BDD 100K 是 2018 年伯克利大学对外公开的道路驾驶数据集,涉及 10 万张道路目标标注

图片。图片中目标标签类别包括行人、非机动车、大巴车、卡车、小型轿车等 10 类目标。本文采用的训练集为 7 万张、验证集为 1 万张，对数据做清洗处理后，训练样本数为 69719，测试样本数量为 1 万。

5.2 实验环境

实验硬件平台为 E5-2620 CPU，两块 GTX 1080ti GPU 的服务，软件为 Ubuntu16.04 操作系统，Python2.7 环境，Tensorflow13.0 框架。训练 Faster R-CNN 网络模型时，采用随机梯度下降 (Stochastic Gradient Descent, SGD)^[4] 算法进行优化。

5.3 实验结果

对于 Faster R-CNN 网络模型来说，模型评价验证指标包括正确率 (accuracy)、查准率 (precision) 和查全率 (recall)。在介绍这些指标前，描述一下混淆矩阵，混淆矩阵即表示真实值与预测值之间的差别。表 1 为混淆矩阵的定义形式。

表 1：混淆矩阵中各元素的定义形式

	预测为正类	预测为负类
真实为正类	TP	FN
真实为负类	FP	TN

根据表 1 可了解 TP、FP、FN、TN 的含义。本文通过 Faster R-CNN 对城市道路目标和目标类型 (行人、车辆) 进行识别，其识别结果均视为正类，即为 TP 或 FP，在系统中不是排除性预测，不存在预测为负类。本系统采用查准率为 Faster R-CNN 识别城市道路目标的评价指标。查准率 AP 的计算公式为如下所示。

$$\text{正类的查准率 (AP)} = \frac{TP}{TP + FP} = \frac{\text{预测正确的正类样本数}}{\text{预测为正类的样本数}} \quad (1)$$

本文实验采用 TensorFlow 深度学习框架，来完成网络模型的设计与训练。为了验证本文提出模型的可行性与检测效果，将本文算法的检测实验结果与当前常见的目标检测算法进行对比分析，对比的算法包括 SSD、YOLO V3 和 GANet (Guided Attention NetWork)，在相同数据集进行对比，检测准确率结果如下：SSD 模型的平均查准率为 68.2%，YOLO V3 网络模型的平均查准率为 73.8%，GANet 网络模型的平均查准率为 76.8%，

而本文采用的 Faster R-CNN 平均查准率为 86.3%。由此可知，Faster R-CNN 网络模型在城市道路目标检测准确率最高。

6 小结

本文提出了一种改进的 Faster R-CNN 算法，用于城市道路目标检测。首先分析了传统 Faster R-CNN 网络模型的工作原理，图像特征通过 VGG16 网络实现提取，为使提取特征信息更加全面，采用更深层次的 ResNet50 网络替换 VGG16，以提高网络模型的整体性能。由于经过多层的卷积操作后，一些弱小目标的语义信息会在深层特征图中消失，为提高卷积后图像特征精度，引入了 FPN 来实现多尺度特征融合，提高了图像特征精度。本文针对图像形状、角度、亮度等不同，设计了图像的多种处理模型，以提高在道路复杂场景下的目标检测效果。在 BDD 100K 数据集上进行实验验证，实验结果表明 Faster R-CNN 网络模型比 SSD、YOLO V3 和 GANet 等常见目标检测模型在查准率方面均高。

参考文献

- [1] 欧阳继红，王梓明，刘思光. 改进多尺度特征的 YOLO v4 目标检测方法 [J]. 吉林大学学报 (理学版), 2022, 60 (06): 1349-1355.
- [2] 钱雯倩，王军. 基于轻量化 SSD 算法的行人目标检测 [J]. 计算机仿真, 2022, 39 (09): 487-491.
- [3] 谢誉，包梓群，张娜，等. 基于特征优化与深层次融合的目标检测算法 [J]. 浙江大学学报 (工学版), 2022, 56 (12): 2403-2415.
- [4] 王昕. 梯度下降及优化算法研究综述 [J]. 电脑知识与技术, 2022, 18 (08): 71-73.

作者简介

ASHIKUR RAHMAN MOHAMMAD (1997-), 男，孟加拉国人。硕士研究生。研究方向为新能源与智能网联汽车技术。
李军 (1964-), 男，博士。研究方向为新能源与智能网联汽车技术。