



计算机工程
Computer Engineering
ISSN 1000-3428, CN 31-1289/TP

《计算机工程》网络首发论文

题目: 基于改进的 Faster R-CNN 小尺度行人检测
作者: 陈泽, 叶学义, 钱丁炜, 魏阳洋
DOI: 10.19678/j.issn.1000-3428.0055817
网络首发日期: 2019-09-29
引用格式: 陈泽, 叶学义, 钱丁炜, 魏阳洋. 基于改进的 Faster R-CNN 小尺度行人检测. 计算机工程. <https://doi.org/10.19678/j.issn.1000-3428.0055817>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

基于改进的 Faster R-CNN 小尺度行人检测

陈泽, 叶学义, 钱丁炜, 魏阳洋

(杭州电子科技大学 模式识别与信息安全实验室, 杭州 310018)

摘 要: 小尺度行人的检测一直是行人检测的难点问题, 本文以 Faster R-CNN 目标检测网络为基础, 采用一种基于双线性插值的对齐池化策略以避免感兴趣区域池化过程中两次量化操作带来的位置偏差; 然后提出一种基于级联的多层特征融合策略, 将具有丰富细节信息的浅层特征图和具有抽象语义信息的深层特征图进行通道叠加以补充小尺度行人在深层特征图中特征信息的不足。实验结果表明, 针对 INRIA 和 PASCAL VOC2012 中的小目标行人数据, 相比于 Faster R-CNN, 在检测效率相当的条件下, 平均准确率(MAP)分别提高 17.58%和 23.78%。

关键词: 小尺度行人检测; RPN; 感兴趣区域池化; Faster R-CNN; 特征融合

Small-scale Pedestrian Detection Using Improved Faster R-CNN Network

CHEN Ze YE Xueyi QIAN Dingwei WEI Yangyang

(Lab of Pattern and Information Security, Hangzhou University 310018)

【Abstract】 A key issue in pedestrian detection is the detection of small-scale pedestrians. This paper is based on Faster R-CNN, and using a new alignment of the pool of the pool area of interest in a manner to avoid the positional deviation caused by the quantization operation twice; Then propose a cascade-based multi-layer feature fusion strategy, concatenate shallow feature maps with rich detail information and deep feature maps with abstract semantic information to complement the insufficiency of small-scale pedestrians' feature information in deep feature maps. Experimental results show, For the small target pedestrian data in INRIA and PASCAL VOC2012, the average accuracy rate (MAP) increased by 17.58% and 23.78%, respectively, compared with the Faster R-CNN.

【Key words】 small-scale pedestrian detection; RPN; Regions of Interest Pooling; Faster R-CNN; feature fusion

DOI:10.19678/j.issn.1000-3428.0055817.

1 概述

近年来, 行人检测作为自动驾驶、智能监控等实际应用的核心部分, 引起了人们的广泛关注^[1]。行人检测的任务是对于输入的图像或者视频帧通过计算机视觉技术输出包含行人的矩形框, 这可以看作是目标检测中的一个实例。由于深度学习(主要是卷积神经网络)强大的特征表达能力, 基于深度学习的检测算法已经在目标

检测领域占据了主导地位^[2,3]。目前基于深度学习的目标检测算法根据预测流程可分为两大类: 一类是以 Faster R-CNN^[4]为代表的基于区域建议的目标检测算法, 如 SPP-Net^[5]、Fast R-CNN^[6]、R-FCN^[7]等, 这类算法都有两个流程, 首先通过区域建议算法生成可能包含目标的候选区域, 然后通过卷积神经网络对候选区域进行分类和位置回归得到最终的检测框。另一类是以 YOLOv3^[8]为代表的基于回归^[9]的目标检测算法, 如 SSD^[10]、RetinaNet^[11]等, 这类算法没有区域建议环节, 它们将预

基金类别: 国家自然科学基金项目(60802047);

作者简介: 陈泽(1995—), 男, 硕士研究生, 主研究方向为深度学习、计算机视觉; 叶学义, 副教授、博士; 钱丁炜, 硕士研究生; 魏阳洋, 硕士研究生。 E-mail: 1453137539@qq.com

先设置在原图上的窗口直接映射回卷积神经网络生成的特征图上,通过深度特征回归出窗口的类别和位置偏移量最终得到目标的检测框。区域建议环节以牺牲检测时间为代价,提供了质量更好的检测框也使得检测的结果更加精准。由于行人检测任务对于精度要求较高,所以目前研究人员大多使用基于区域建议的目标检测算法应用于行人检测。ZHANG 等^[12]发现 Faster R-CNN 目标检测网络中的区域建议网络 (Region Proposal Network, RPN) 作为独立的检测器表现良好,而分类回归网络却受小尺度行人影响较大,进而大幅降低了整体网络的检测性能,所以他们使用提升树算法(Boosted Forests, BF)替换了分类回归网络,这种方法一定程度上提高了网络对于小尺度行人的检测能力;Li 等^[13]针对行人尺度变化过大问题提出一种 SA-Fast RCNN 方法,分别设计两个子网络分开检测大尺度和小尺度行人;类似地,Cai 等^[14]提出了一种统一的多尺度卷积神经网络 (MS-CNN),它在卷积网络中的多个中间层同时进行检测以匹配不同尺度的目标;Mao 等^[15]试图引入其它特征 (如梯度、热力信息、光流等) 来改善 Faster R-CNN 的检测性能,这些特征在提升检测性能的同时也带来了更多的计算消耗;同样地,Brazil 等^[16]将语义分割信息注入检测网络的共享特征图来帮助检测行人。

在现实场景下的行人检测任务存在背景复杂多变,遮挡和尺度过小等诸多问题,而其中小尺度的行人对于检测性能影响最大且出现的情况更普遍^[17],比如在自动驾驶场景下存在很多距离汽车较远的小行人,这些目标很难用激光雷达探测到,这个时候就需要计算机视觉的帮助。本文专注于提高对于小尺度行人的检测能力,以 Faster R-CNN 为基础,通过重新设置包围框参数,取消感兴趣区域池化过程中的量化操作和多层特征融合的方法,在不增加大量检测时间的前提下,提高了小尺度行人的检测能力。

2 Faster R-CNN 目标检测网络

2.1 Faster R-CNN 简介

Faster R-CNN 是基于区域建议的目标检测算法,其检测流程如图 1 所示,首先通过卷积神经网络 (Convolutional Neural Network, CNN) 对输入图像进行多层卷积特征提取,其次区域建议网络 (Region Proposal Network, RPN) 根据提取到的卷积特征生成可能包含目

标的候选区域也叫做感兴趣区域 (Regions of Interest, RoI),同时将 RoI 映射回卷积神经网络生成的特征图上,接着分类回归子网 Fast R-CNN 通过感兴趣区域池化 (Regions of Interest Pooling, ROI pooling) 提取出长度固定的特征张量,最后再对提取的特征张量进行分类和位置回归得到最终的检测框。

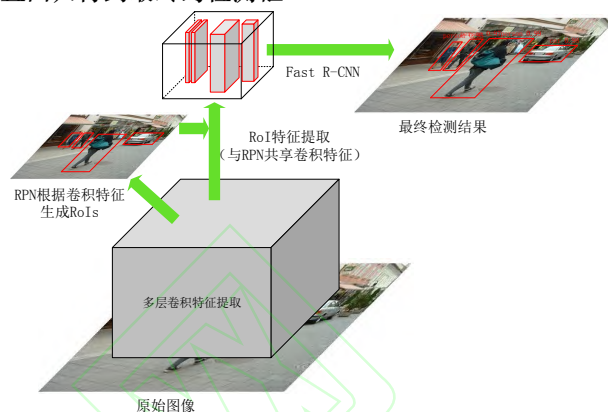


图 1 Faster R-CNN 的网络结构

2.2 RPN

RPN 是 Faster R-CNN 中非常重要的一部分,它根据 CNN 提取的特征,在原图上铺设不同比例的 Anchor (参考框) 来产生匹配各种尺度目标的候选框,这种方式大幅降低了区域建议环节带来的计算成本,例如,传统的选择性搜索 (Selective Search) 算法^[18]对一张图像进行区域建议的时间为 2s,而 RPN 只需要 10ms。

RPN 进行区域建议的流程如图 2 所示,首先在从 CNN 提取到的卷积特征图上以每个点为中心生成三种尺度 (128, 256, 512)、三种宽高比 (1:1, 1:2, 2:1) 共 9 种不同大小的 Anchor (参考框),值得注意的是,Anchor 虽然以特征图上的点为中心,但它是铺设在原图的尺度上的,如图 2 右边所展示的一样。这样的设计方案是为了更高效地应对多尺度目标的检测,相比于传统的图像金字塔^[19]方案 (把图像在多个尺度上进行缩放,并针对每个尺度进行特征提取),Anchor 的设置更有效且计算量要少的多。接着我们在卷积特征图上滑动一个 3×3 的窗口做卷积操作,每一个滑动窗口生成一个 512 维的特征向量 (特征向量的维度由 CNN 决定,使用 VGG-16^[20]做特征提取网络,则最后一层卷积得到的特征图有 512 个通道),之后这个特征向量被分别输入到两个 1×1 的全连接层,一个分类层输出参考框作为前景的置信度,一个回归层输出参考框相较于标注框的坐标偏移量。我们通过设置损失函数来指导训练 RPN 网络,测试时 RPN

会输出大约 2000 个只包含前景并经过位置修正的候选

区域, 以供 Fast R-CNN 进行进一步地分类回归。

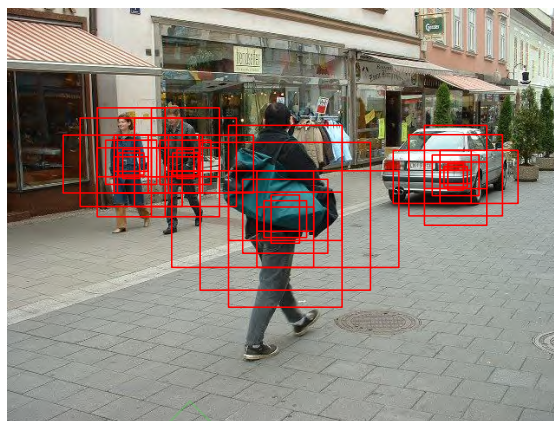
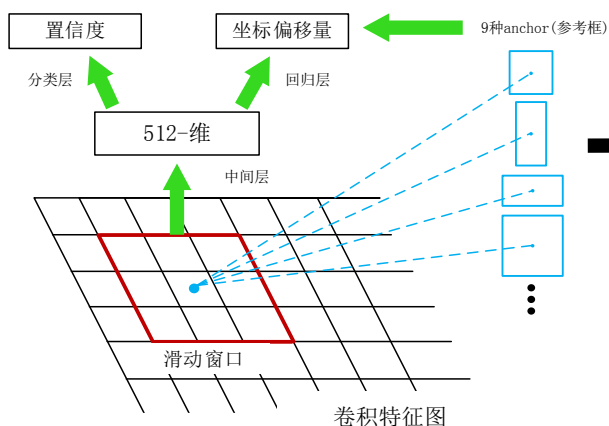


图 2 RPN 网络示意图

2.2.1 RPN 损失函数

为了训练 RPN, 我们为与标注框的重叠交并比 (Intersection over Union, IoU) 最大, 或者与标注框的 IoU 超过 0.7 的 Anchor 分配正标签, 为 IoU 低于 0.3 的 Anchor 分配负标签。IoU 的计算如公式 (1) 所示:

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \quad (1)$$

有了以上的定义, RPN 的损失函数可以定义为公式 2 所示:

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (2)$$

$$smooth_{L_1}(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & otherwise \end{cases} \quad (3)$$

其中, $L(\{p_i\}, \{t_i\})$ 是 RPN 的损失函数, i 是一个小批量数据中的 Anchor 的索引, p_i 是 Anchor i 作为前景的预测概率。如果 Anchor 为正样本, 则它的真实标签 p_i^* 为 1, 如果 Anchor 为负样本, 则为 0。 t_i 表示预测边界框 4 个参数化坐标的向量, 而 t_i^* 是与正 Anchor 相关的真实标注框的向量。分类损失 L_{cls} 是两个类别上 (前景或背景) 的对数损失。回归损失 $L_{reg}(t_i, t_i^*) = smooth_{L_1}(t_i - t_i^*)$, $p_i^* L_{reg}(t_i, t_i^*)$ 表示回归损失仅对于正样本 Anchor 激活, 否则被禁用 $p_i^* = 0$ 。

2.3 Fast R-CNN

在 RPN 之后, 我们已经得到了大约 2000 个可能包

含目标的候选区域, 文献[6]将这两千多个候选区域裁剪到相同的尺寸然后分别送入 CNN 提取特征最后进行分类, 显然这样的计算效率非常低。Fast R-CNN 通过 RoI Pooling 层重用现有的卷积特征, 极大地提升了计算效率。RoI pooling 的流程如图 3 所示, 首先将候选区域映射到 CNN 最后一层卷积得到的特征图上, 由于候选区域形状, 大小各异, 其对应的特征图形状也各不相同, 而全连接层要求输入维度相同的特征向量, 所以 Fast R-CNN 将不同的特征池化成 $7 \times 7 \times 512$ 固定维度的张量, 最后通过全连接层分别输出各个类别的置信度和位置修正参数。值得注意的是, RPN 输出的候选区域已经经过了一次位置的修正, Fast R-CNN 对于候选区域位置的第二次修正也使得最终生成的检测框的位置更加准确。

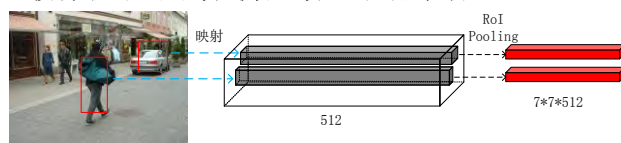


图 3 RoI pooling 的示意图

3 Faster R-CNN 网络分析及改进方法

Faster R-CNN 网络在通用的目标检测领域表现优异, 但是对于行人检测任务却表现不好, 尤其在检测小尺度行人方面出现了大量的误检和漏检情况^[15]。本小节在之前研究的基础上, 分析了 Faster R-CNN 网络在行人检测尤其是对于小尺度行人检测上的应用: 1) 在 RPN 网络中, Anchor 的设置是针对于通用目标的, 如图 2 所示, 为了同时检测到行人和车辆, Anchor 需要使用不同的宽高比。但是行人的包围框多数为瘦长型, 且具有固定的

宽高比: 0.41^[1], 所以这个重要的超参设置对于行人检测任务来说有着明显的短板。2) 原始网络在 RoI pooling

过程中存在两次量化 (即将浮点数取整) 操作。

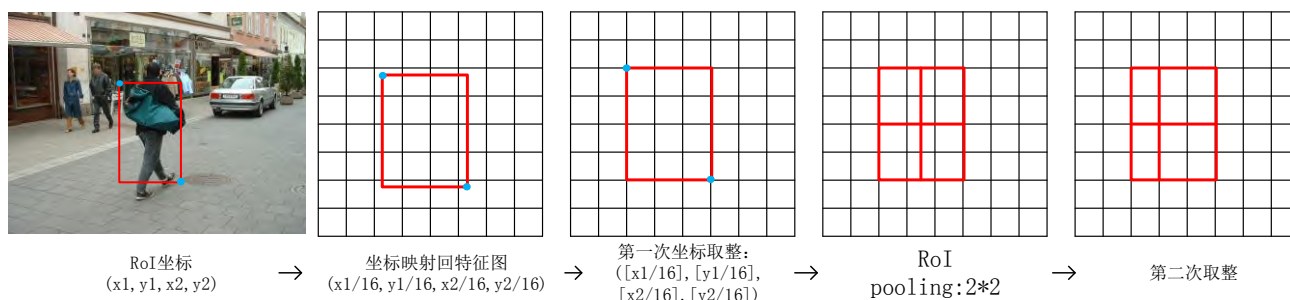


图 4 RoI pooling 中的两次量化

如图 4 所示, 原网络首先将对应于原图上的 RoI 坐标 (分别为 RoI 左上角和右下角两个点的四个坐标值) 映射到特征图上, 因为 CNN 的池化操作, 特征图的尺寸相比于原图缩小了 16 倍, 所以通过 $x/16$ 来计算对应坐标, 但是这样计算是有小数的, 所以要进行第一次取整操作, 即 $x/16$ 。接下来为方便讨论我们假设进行 2×2 的池化操作 (原网络使用的是 7×7 池化), 即对于每一个映射到特征图上的 RoI 都要分成 2×2 个区域, 再进行一个除法, 又出现小数, 所以要进行第二次取整, 经过两次取整之后就使得 RoI pooling 最终提取的特征和原图的 RoI 不再对齐, 这在特征图上可能是一个或两个点的偏差, 但是这几个点相对于原图则是 16 个甚至更多像素的偏差, 这对于一般尺度的目标不会有太大的影响, 但是对于自身包含像素就较少的小尺度行人来说却是严重的误差。3) RPN 和 Fast R-CNN 共享 CNN 提取的卷积特征图, 这个特征图的质量直接影响了前期区域建议的质量和后续的分类回归的准确度。在卷积神经网络中, 越深层的特征图分辨率越低且具有较大的感受野和丰富的语义信息, 相反越浅层的特征图分辨率越高且具有的较小的感受野和丰富的细节信息。Faster R-CNN 网络使用最后一层的卷积特征图, 以期望于通过深层特征更为丰富且抽象的语义信息来提高网络对于物体形变及遮挡的鲁棒性。这对于大尺度目标是有效的, 但是这些分辨率过小的深层特征图并不适用于小尺度行人, 小尺度行人自身所包含的像素点就较少更是难以提取到较为丰富的语义信息, 随着网络层数的加深, 其细节特征不断与周围提取的特征组合, 这使得深层特征图的每个点都包含了太多小尺度目标周围的信息, 进而降低了对于小尺度

行人的辨别能力。

3.1 改进 RPN

本文在 RPN 的基础上做出改进, 根据行人较为固定的宽高比的特性将 Anchor 设置为一种宽高比: 0.41 和 11 种尺度, 其具体数值可以根据数据集中行人的高度分布进行设置。图 5 展示了 RPN 和改进后提供的 Anchor 对比, 图中红色框为部分 Anchor, 可以看出改进 RPN 后提供的 Anchor 与行人更为对齐。并且由于 RPN 中的损失函数是由分类损失和用于位置修正的回归损失组成的, 更为对齐标注框的 Anchor 使得回归的损失函数较小, 在训练过程中, 网络将更侧重于学习分类, 进而使得最终 RPN 提供的候选区域更为准确。

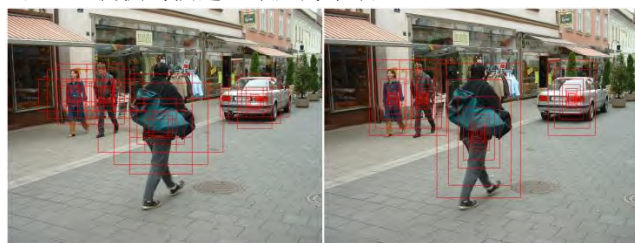


图 5 RPN 的 Anchor (左) 和改进后的 Anchor (右)

3.2 双线性插值的对齐 RoI Pooling

为避免 RoI Pooling 中两次量化操作所带来的像素偏差, 本文采取一种基于双线性插值的对齐 RoI Pooling^[21] 方式。具体方法如图 6 所示, 首先将 RoI 映射到特征图上, 计算过程中保留小数, 不进行取整操作。之后为方便讨论同样是将 RoI 对应的特征图均分为 2×2 个区域, 同时每个区域内设置 4 个采样点, 我们可以看到所有的采样点都不是整数坐标, 没有对应的具体的值, 所以我们通过双线性插值来分别对每个采样点进行估值。插值完成之后, 对每个区域内进行 Max Pooling, 即对每个

区域内的四个点取最大值,最终得到一个 2×2 大小的特征张量。这就是对齐 RoI Pooling 的整个过程,我们可以

发现整个过程中没有用到取整操作,这样就可以很好的保存每个 RoI 的空间位置,使得检测框更精确。

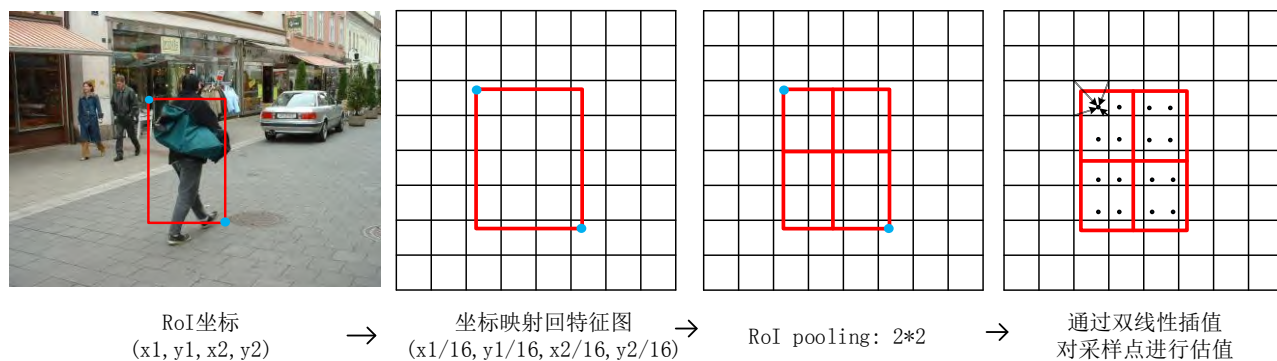


图 6 对齐 RoI Pooling

3.3 基于级联的特征融合

Faster R-CNN 只使用最后一层的卷积特征图,其较低的分辨率对于检测小尺度行人来说是不够的。直观上使用分辨率更高的第四层卷积特征图(VGG-16^[20]中有一共 5 个卷积模块)就可以提高对小尺度行人的检测能力,文献^[2]就是采用了这种方法。但是这种方法并没有充分利用到卷积神经网络的特征提取能力,并可能影响对于大尺度目标的检测。考虑到浅层特征分辨率较高且细节信息丰富,深层特征分辨率较低但语义信息丰富的特点,故本文采用特征融合的策略,实现特征复用,丰富小尺度行人的特征。特征融合一般分为特征通道叠加和特征图求和两种方式,对于两路特征融合来说,通道叠加和求和的计算分别由公式 (4) (5) 所示:

$$Z_{concat} = \sum_{i=1}^c X_i * K_i + \sum_{i=1}^c Y_i * K_{i+c} \quad (4)$$

$$Z_{add} = \sum_{i=1}^c (X_i + Y_i) * K_i = \sum_{i=1}^c X_i * K_i + \sum_{i=1}^c Y_i * K_i \quad (5)$$

式中, X_i 和 Y_i 分别表示两路待融合的特征, K_i 表示卷积核, $*$ 为卷积操作。特征求和和要求两路特征通道数必须相同,而通道叠加并不需要;并且特征求和是通过特征图逐元素相加的方式进行的,如果两路特征不具有同类的特征信息,则融合的结果很可能会带来负影响。通道叠加的方式则可以理解为,增加了图像的不同特征,而每一层的特征信息没有变化。故本文采用通道叠加的方式进行多层的通道融合,以丰富小尺度行人的特征信息。具体的融合策略如图 7 所示首先将 conv3_3 层后的特征图和 conv5_3 后的特征图调整到 conv4_3 后特征图

的尺寸,然后将二者和 conv4_4 进行通道叠加,再将通道叠加后的特征通过 1×1 卷积降维到原始的 512 通道数,最后得到融合以后的特征。

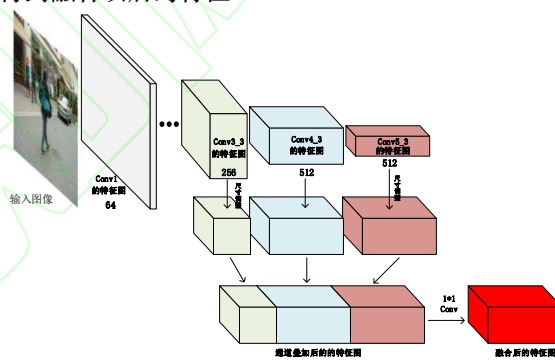


图 7 特征融合示意图

4 实验与分析

为了验证本文改进方法的有效性,在计算机上进行了仿真,对 Faster R-CNN 目标检测网络和改进后的 Faster R-CNN 进行了比较。本文使用 PyTorch 深度学习框架,在 16GB 内存,4GB 显存 NVIDIA GeForce GTX1050Ti GPU,并搭有 CUDA8.0 和 CUDNN5.1 运算平台的笔记本上进行训练。本文实验基于 PASCAL VOC2007、PASCAL VOC2012 和 INRIA 数据集。采用 VOC2007 数据集中的包含行人的部分作为训练集(共 6383 张),采用 INRIA 数据集中训练集里的正样本图片作为测试集(共 614 张),而测试集中的正样本作为验证集(共 576 张)。取 PASCAL VOC2012 数据集中小目标行人图片制作小尺度行人测试集(共 154 张),本文定义行人高度小于图片高度四分之一为小尺度行人。

4.1 实验细节

4.1.1 特征融合细节

本文将 conv3_3 后的特征图通过 Max Pooling 下采样到 conv4_4 的尺寸, 对于 conv5_3 后特征图的上采样则是使用双线性插值的方式而不是流行的反卷积 (Deconvolution) [22], 这是因为在实际训练中, 反卷积由于参数过多导致难以训练, 而双线性插值根据临近像素的位置进行估值无需参数, 其实实验效果相较于反卷积要好得多。在特征通道叠加前本文又分别通过 Batch Normalization 层[23]对调整过尺寸的不同层特征图进行归一化。最后加入了 1×1 的卷积层将融合特征的维度降到 512, 同时卷积层后增加了非线性激活函数 Relu (The Rectified Linear Unit), 这也提高了网络的表达能力。

4.1.2 Soft-NMS

原网络在测试时会通过 NMS 算法去除同一目标上的重叠框, NMS 算法是将检测框按置信度排序, 然后保留置信度最高的框, 同时直接删除与该框 IoU 大于一定阈值 (如 0.5) 的其它框, 然而这种过滤方式并不适用于重叠情况较为普遍的行人检测, 如图 8 所示红色框的置信度最高得到保留, 而同样检测到行人的绿色框会因为与红色框的 IoU 高于设定阈值而被过滤掉, 最终导致了红色行人的漏检。Bodla [24]针对这个问题做出改进, 提出了 Soft-NMS, 它并没有直接删除图中的绿色框, 而是通过线性加权或是高斯加权的方式降低其置信度, 与红色框重叠程度越高则置信度下降的越快, 这种方式很大程度上保留了因重叠而导致被误删的检测框, 故本文在测试时使用 Soft-NMS 来替换传统的 NMS, 在实验中获得了平均 2% 的平均准确度提升。



图 8 重叠行人示意图

4.2 结果分析

4.2.1 改进 RPN

为验证改进 RPN 的有效性, 本小节使用不同的区域建议网络 (分类回归子网络使用相同的原 Fast R-CNN) 在 INRIA 数据集进行比较, 结果如表 1 所示:

表 1 改进 RPN 前后对比

方法	候选区域数	MAP(%)	Time(ms)
RPN	2000	69.21	231
改进的 RPN	2000	75.43	235
RPN	300	67.74	224
改进的 RPN	300	73.22	224

从表 1 可以看出, 改进的 RPN 在和原 RPN 提供相同数量的候选区域的同时分别提升了 6.22% 和 5.48% 的精度。这说明改进后所提供的候选区域质量更高。

4.2.2 特征融合

本小节为了说明特征融合策略的有效性, 采用了不同卷积层的特征图进行融合, 在 INRIA 数据集做了对比实验, 结果如表 2 所示:

表 2 不同卷积层特征融合效果的对比

方法	MAP(%)	Time(ms)
改进 Faster R-CNN	83.15	347
Faster R-CNN	59.37	336
HOG+SVM	46.86	-

4.2.3 小尺度行人检测

本小节对比了不同算法在 PASCAL VOC2012 数据集中提取出的小尺度行人测试集上的表现, 结果如表 3 所示:

表 3 PASCAL VOC2012 数据集检测结果对比

Conv3_3	Conv4_3	Conv5_3	MAP(%)
		✓	69.21%
	✓	✓	77.43%
✓		✓	78.17%
✓	✓	✓	86.79%

表 3 显示本项目改进后的网络对于小尺度行人检测的精确度提高了 23.78%, 检测时间略有延长。这表明本项目针对于检测小尺度行人的改进是有效的, 虽然检测时间相比原始网络有少许增加, 但换来的是对于小尺度行人检测能力的大幅提升。图 9 (a), (b), (c) 为改进网络前后在行人遮挡及尺度变化较大的场景和自动驾驶场景下的行人检测结果, 它们更加直观地展示了改进网络检测小尺度行人的能力, 图中左侧为 Faster R-CNN 检测结果, 右侧为改进网络检测结果:



图 9 Faster R-CNN 与改进网络的检测结果对比

从图 9 (a) 中可以看出改进网络不仅检测出了 Faster R-CNN 漏检的远处小行人, 同时也检测出了被部分遮挡的行人, 这说明本文的改进方法对于行人的遮挡具有一定的鲁棒性。图 9 (b), (c) 是由车载摄像头采集到的图像, 可以看出改进网络均检测出了道路远端的小尺度行人, 这也说明在自动驾驶场景下本文的改进方法对于小尺度行人检测能力也有很大的提高。

5 结束语

本文为了提高对于小尺度行人的检测能力, 以 Faster R-CNN 为基础, 针对行人的特点重新设置了 RPN 中的候选框参数; 为了避免在感兴趣区域池化中两次量化操作带来的位置偏差采用了一种基于双线性插值的对齐池化方式; 同时针对小尺度行人在深层特征图中特征信息的不足, 提出一种基于级联的多层特征融合策略, 将具有丰富细节信息的浅层特征图和具有抽象语义信息的深层特征图进行融合, 最后使用融合的特征检测行人。实验结果表明, 本文的改进方法在和文献[1]检测效率相当

的条件下有效地提高了网络对于小尺度行人的检测能力。但本文的网络在面对密集人群互相遮挡的场景时依然存在漏检情况, 下一步将在改进网络的基础上, 对行人的自遮挡问题进行研究, 以进一步提高行人检测的准确度。

参考文献

- [1] Zhang SS, Benenson R, Schiele B, et al. CityPersons: A diverse dataset for pedestrian detection. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA, 2017, 4457-4465.
- [2] 薛丽霞, 钟欣, 汪荣贵, 杨娟, 胡敏. 基于深度特征融合的中低分辨率车型识别[J]. 计算机工程, 2019, 45(1): 233-238, 245.
- [3] 李红艳, 李春庚, 安居白, 任俊丽. 注意力机制改进卷积神经网络的遥感图像目标检测[J]. 中国图象图形学报, 2019, 24(08): 1400-1408.
- [4] Ren S, Girshick R, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.
- [5] He K, Zhang X, Ren S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9): 1904-16.
- [6] Girshick R. Fast R-CNN[J]. Computer Science, 2015.
- [7] DAI J F, LI Y, HE K M, et al. R-FCN: object detection via region-based fully convolutional networks [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2016: 379-387.
- [8] REDMON J, FARHADI A. YOLO V3: an incremental improvement [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2018: 1-6.
- [9] 焦天驰, 李强, 林茂松, 贺贤珍. 结合反残差块和 YOLOv3 的目标检测法[J]. 传感器与微系统, 2019, (09): 144-146, 156.
- [10] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands. 2016. 21-37.
- [11] Lin T Y, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017,

PP(99):2999-3007.

5562-5570.

- [12] ZHANG L, LIN L, LIANG X, et al. Is Faster R-CNN Doing Well for Pedestrian Detection?[C] // European Conference on Computer Vision. Springer International Publishing, 2016: 443-457.
- [13] Li J, Liang X, Shen S M, et al. Scale-aware Fast R-CNN for Pedestrian Detection[J]. IEEE Transactions on Multimedia, 2017: 1-1.
- [14] Cai Z W, Fan Q F, Feris R S, et al. A unified multi-scale deep convolutional neural network for fast object detection[C]//Proceedings of the 14th European Conference on Computer Vision, 2016: 354-370.
- [15] Mao J, Xiao T, Jiang Y, et al. What Can Help Pedestrian Detection?[C]// Computer Vision & Pattern Recognition. 2017.
- [16] Li C, Song D, Tong R, et al. Illumination-aware Faster R-CNN for Robust Multispectral Pedestrian Detection[J]. Pattern Recognition, 2018.
- [17] Liu S, Huang D, Wang Y. Adaptive NMS: Refining Pedestrian Detection in a Crowd[J]. 2019.
- [18] Kulkarni A, Callan J. Selective Search[J]. ACM Transactions on Information Systems, 2015, 33(4): 1-33.
- [19] Girshick R, Donahue J, Darrelland T, et al. Rich feature hierarchies for object detection and semantic segmentation[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2014.
- [20] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.
- [21] He K, Gkioxari G, Dollar P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99): 1-1.
- [22] Shi W, Caballero J, Theis L, et al. Is the deconvolution layer the same as a convolutional layer?[J]. 2016.
- [23] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]// International Conference on International Conference on Machine Learning. JMLR.org, 2015.
- [24] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMS -Improving object detection with one line of code[C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: