第 46 卷 第 9 期 计 算 机 工 程 2020 年 9 月
Vol. 46 No. 9 Computer Engineering September 2020

·图形图像处理· 文章编号：1000-3428(2020)09-0226-07 文献标志码：A 中图分类号：TP391

基于改进 Faster R-CNN 的小尺度行人检测

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： 为提高小尺度行人检测的准确性，提出一种基于改进 Faster R-CNN 的目标检测方法。通过引入基于双线性插值的对齐池化层，避免感兴趣区域池化过程中两次量化操作导致的位置偏差，同时设计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，从而解决小尺度行人在深层特征图中特征信息缺乏的问题。在 INRIA 和 PASCAL VOC2012 数据集上的实验结果表明，在小尺度行人检测效率相同的情况下，该方法相比基于 Faster R-CNN 的检测方法平均精确率均值分别提高了 17.58% 和 23.78% 。

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

开放科学（资源服务）标志码（OSID）：

中文引用格式： 陈泽，叶学义，钱丁炜，等. 基于改进 Faster R-CNN 的小尺度行人检测[J]. 计算机工程，2020，46(9)：226-232，241.

英文引用格式： CHEN Ze，YE Xueyi，QIAN Dingwei，et al. Small-scale pedestrian detection based on improved Faster R-CNN[J]. Computer Engineering，2020，46(9)：226-232，241.

Small-Scale Pedestrian Detection Based on Improved Faster R-CNN

CHEN Ze，YE Xueyi，QIAN Dingwei，WEI Yangyang

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

【Abstract】 To improve the accuracy of small-scale pedestrian detection，this paper proposes a target detection method based on improved Faster R-CNN. The network structure uses a new aligned pooling layer based on bilinear interpolation to avoid the positional deviation caused by two quantization operations in Region of Interest（ROI）pooling. Then a cascade-based multi-layer feature fusion strategy is designed，which concatenates shallow feature maps with rich detail information and deep feature maps with abstract semantic information to address the insufficiency of feature information of small-scale pedestrians in deep feature maps. Experimental results on INRIA and PASCAL VOC2012 datasets show that the proposed method increases the mean Average Precision（mAP）by 17.58% and 23.78% respectively compared with detection method based on Faster R-CNN with the same efficiency of small-scale pedestrian detection.

【Key words】 small-scale pedestrian detection；Region Proposal Network（RPN）；Region of Interest（ROI）pooling；Faster R-CNN network；feature fusion

DOI： 10.19678/j.issn.1000-3428.0055817

0 概述

近年来，行人检测作为智能交通应用的核心技术，引起了人们的广泛关注[1]。行人检测的任务是对输入的图像或视频帧通过计算机视觉技术输出包含行人的矩形框，可将其看作目标检测中的一个实例。由于深度学习强大的特征表达能力，因此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]。根据预测流程可将基于深度学习的目标检测算法分为两大类：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（Convolutional

基金项目： 国家自然科学基金（60802047）。

作者简介： 陈 泽（1995—），男，硕士研究生，主研方向为深度学习、计算机视觉；叶学义，副教授、博士；钱丁炜、魏阳洋，硕士研究生。

收稿日期： 2019-08-26 **修回日期：** 2019-10-19 **E-mail：** 1453137539@qq.com

Neural Network,CNN)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,而是将预先设置在原图上的窗口直接映射到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,可提供质量更好的检测框,也能使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,因此目前多数研究在行人检测中使用基于区域建议的目标检测算法。文献[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,大幅降低了整体网络的检测性能,因此,该文使用提升树(Boosted Forests,BF)算法替换分类回归网络,有效提高了网络对于小尺度行人的检测能力。文献[13]针对行人尺度变化过大的问题提出 SA-Fast RCNN 方法,分别设计两个子网络用于检测大尺度和小尺度行人。类似地,文献[14]提出多尺度卷积神经网络 MS-CNN,在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。文献[15]通过引入其他特征(如梯度、热力信息、光流等)来改善 Faster R-CNN 的检测性能,但同时带来了更多的计算消耗。文献[16]将语义分割信息引入检测网络的共享特征图用于协助行人检测。

现实场景下的行人检测任务存在背景复杂多变、目标遮挡和尺度过小等问题,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],如在自动驾驶场景下存在很多距离汽车较远的小尺度行人,这些目标很难用激光雷达探测到,因此需要引入计算机视觉技术。本文以 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 3 种方式,在不增加大量检测时间的前提下提高网络对小尺度行人的检测能力。

1　Faster R-CNN 目标检测网络

1.1　网络结构

Faster R-CNN 是基于区域建议的目标检测网络,其检测流程如图 1 所示。首先利用卷积神经网络对输入图像进行多层卷积特征提取,然后在区域建议网络中根据提取到的卷积特征生成可能包含目标的候选区域,也称为感兴趣区域(Region of Interest,RoI),并将其映射到卷积神经网络生成的特征图上,分类回归子网 Fast R-CNN 通过感兴趣区域池化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。

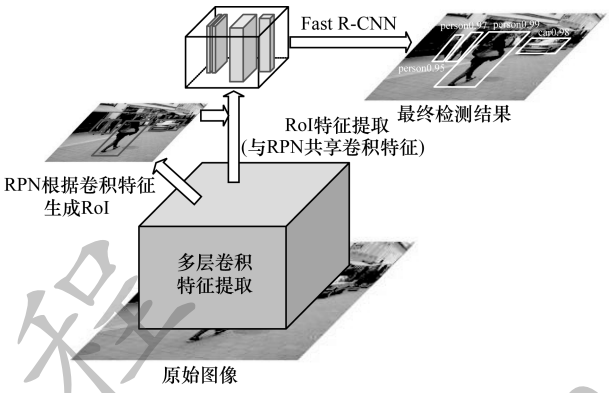


图 1　Faster R-CNN 网络结构

Fig. 1　Network structure of Faster R-CNN

1.2　区域建议网络

区域建议网络是 Faster R-CNN 中非常重要的一部分,其根据卷积神经网络(CNN)提取的特征,在原图上铺设不同比例的参考框(Anchor)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,传统的选择性搜索算法[18] 对一张图像进行区域建议的时间为 2 s,而 RPN 只需要 10 ms。

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 3 种尺度(128,256,512)、3 种宽高比(1∶1,1∶2,2∶1)共 9 种不同大小的 Anchor。值得注意的是,Anchor 虽然以特征图上的点为中心,但它是铺设在原图尺度上的,如图 2 右边部分所示。这样的设计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大幅减少。本文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,如使用 VGG-16[20] 构建特征提取网络,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,分类层输出参考框作为前景的置信度,回归层输出参考框相较于标注框的坐标偏移量。本文通过设置损失函数来指导训练 RPN 网络,测试时 RPN 会输出约 2 000 个只包含前景并经过位置修正的候选区域,以供 Faster R-CNN 做进一步分类回归。

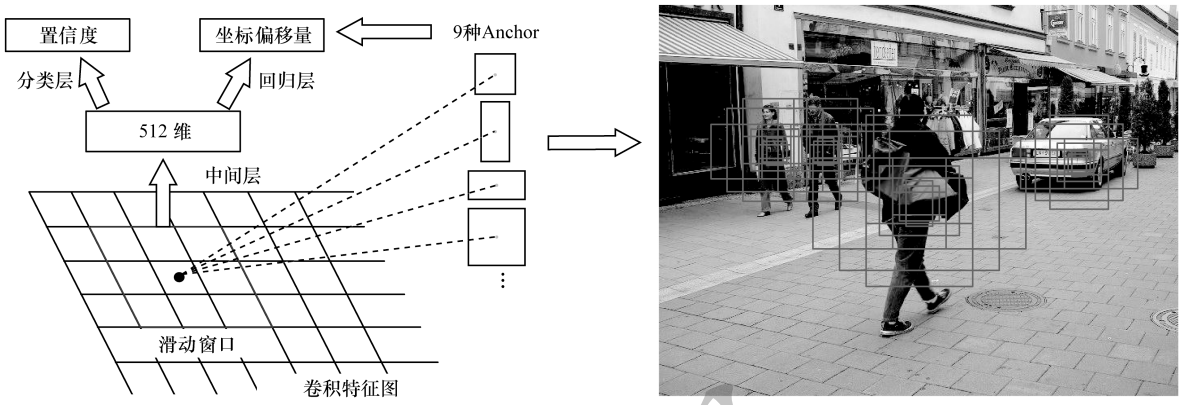


图 2 RPN 示意图

Fig. 2 Schematic diagram of RPN

为训练 RPN，对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，对 IoU 低于0.3 的 Anchor 分配负标签。IoU 的计算如式（1）所示：

$$\text{IoU}=\frac{\text{交集区域}}{\text{并集区域}} \tag{1}$$

基于上述计算公式，RPN 的损失函数可以定义为：

$$L(\{p_i\},\{\boldsymbol{t}_i\})=\frac{1}{N_{\text{cls}}}\sum_i L_{\text{cls}}(p_i,p_i^*)+\lambda\frac{1}{N_{\text{reg}}}\sum_i p_i^* L_{\text{reg}}(\boldsymbol{t}_i,\boldsymbol{t}_i^*) \tag{2}$$

$$\text{smooth}_{L_1}(x)=\begin{cases}0.5x^2, & |x|<1\\ |x|-0.5, & \text{其他}\end{cases} \tag{3}$$

其中，$L(\{p_i\},\{\boldsymbol{t}_i\})$表示 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。Anchor 为正样本时其真实标签 p_i^* 为 1，负样本时则为 0；$\boldsymbol{t}_i$表示预测边界框 4 个参数化坐标的向量，而$\boldsymbol{t}_i^*$ 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$L_{\text{reg}}(\boldsymbol{t}_i,\boldsymbol{t}_i^*)=\text{smooth}_{L_1}(\boldsymbol{t}_i-\boldsymbol{t}_i^*)$，$p_i^* L_{\text{reg}}(\boldsymbol{t}_i,\boldsymbol{t}_i^*)$表示回归损失仅对于正样本 Anchor 激活，否则被禁用，$p_i^*=0$。

1.3 RoI 池化

通过 RPN 能够得到约 2 000 个可能包含目标的候选区域，文献[6]将这些候选区域裁剪到相同的尺寸后分别送入 CNN 提取特征进行分类，显然这样的方式计算效率非常低。Faster R-CNN 通过 RoI 池化层重用现有的卷积特征，以提高计算效率。RoI 池化工作流程如图 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，因此 Fast R-CNN 将不同的特征池化成 7×7×512 固定维度的张量，然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一次位置的修正，Faster R-CNN 对于候选区域位置的第 2 次修正也使得最终生成的检测框的位置更加准确。

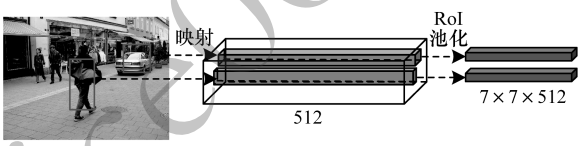


图 3 RoI 池化示意图

Fig. 3 Schematic diagram of RoI pooling

2 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佳，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]。Faster R-CNN 网络应用于行人检测尤其是小尺度行人检测时存在的不足具体如下：

1）在 RPN 网络中，Anchor 的设置是针对通用目标的，为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，该超参数设置对于行人检测任务而言具有明显缺陷。

2）原始网络在 RoI 池化过程中存在两次量化（即将浮点数取整）操作。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 2 个点的 4 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,所以应通过 $x/16$ 来计算对应坐标,但是这样计算是有小数的,因此,需要进行第一次取整操作,即 $x/16$。为方便讨论,本文假设进行 2×2 的池化操作(原网络使用的是 7×7),即对于每一个映射到特征图上的 RoI 都要分成 2×2 个区域,再进行一次除法操作又会出现小数,所以,需要进行第 2 次取整。经过两次取整之后就使得 RoI 池化层最终提取的特征和原图的 RoI 不再对齐,这在特征图上可能仅是 1 个或 2 个点的偏差,但是相对于原图则是 16 个点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目标而言却是严重的误差。

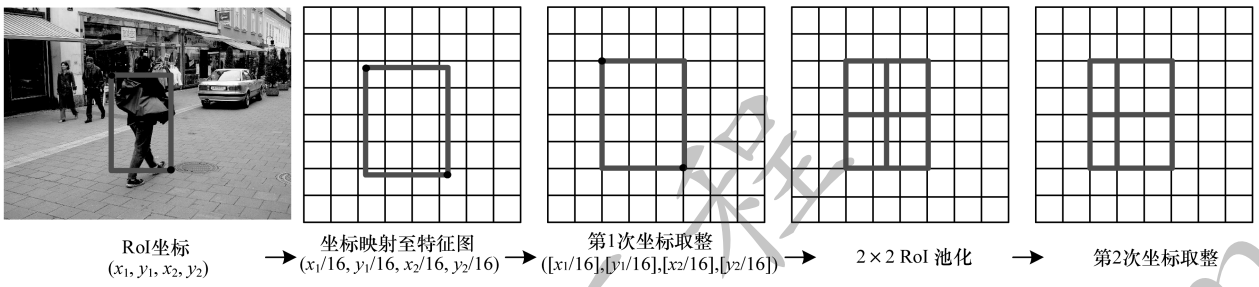


图 4　RoI 池化中的两次量化

Fig. 4　Two quantizations in RoI pooling

3)RPN 和 Faster 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,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,因为小尺度行人自身所包含的像素点较少,难以从中提取到丰富的语义信息,并且随着网络层数的增加,其细节特征不断与周围提取的特征组合,使得深层特征图中的每个点都包含很多小尺度目标周围的信息,降低了检测的准确性。

2.1　RPN 改进

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 1 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,图中方框部分为 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　改进前后的 Anchor 对比

Fig. 5　Comparison of Anchor before and after improvement

2.2　双线性插值的对齐 RoI 池化

为避免 RoI 池化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 池化[21]方式,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。为方便讨论,同样将 RoI 对应的特征图均分为 2×2 个区域,并在每个区域内设置 4 个采样点,可以看到,所有的采样点都不是整数坐标,没有对应具体值,因此,本文通过双线性插值分别对每个采样点进行估值。插值完成之后,对每个区域内进行最大池化操作,即对每个区域内的 4 个点取最大值,最终得到一个 2×2 大小的特征张量。在对齐 RoI 池化的整个过程中没有用到取整操作,从而能够很好地保存每个 RoI 的空间位置,进一步提高检测框精度。

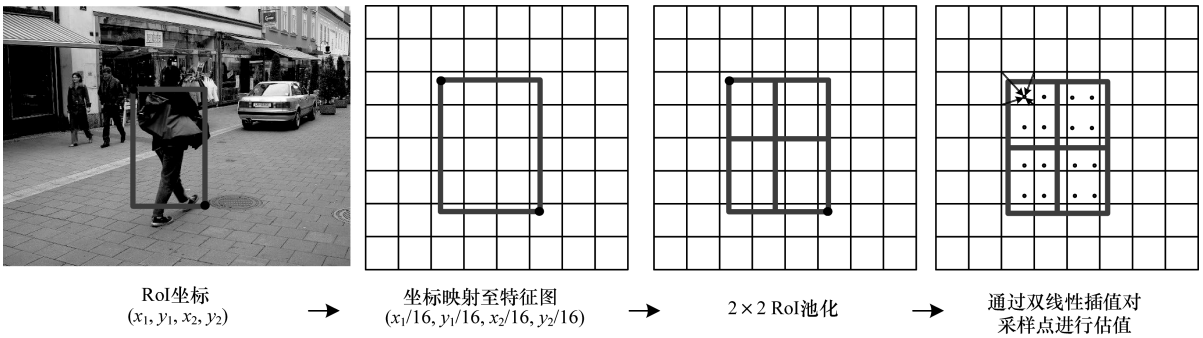


图 6 对齐 RoI 池化示意图

Fig. 6 Schematic diagram of aligned RoI pooling

2.3 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，其较低的分辨率不能满足小尺度行人检测的要求，而使用分辨率较高的第 4 层卷积特征图（VGG-16[20]中有共 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]即采用这种方法。但是该方法并没有充分利用到卷积神经网络强大的特征提取能力，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，本文采用特征融合的策略实现特征复用，以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，对于两路特征融合而言，通道叠加和与求和的计算公式分别由如式(4)和式(5)所示：

$$Z_{\text{concat}} = \sum_{i=1}^{c} X_i * K_i + \sum_{i=1}^{c} Y_i * K_{i+c} \tag{4}$$

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

其中，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，则很可能会给融合结果带来负面影响。通道叠加的方式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。因此，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 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，然后将两者与 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，最终得到融合后的特征。

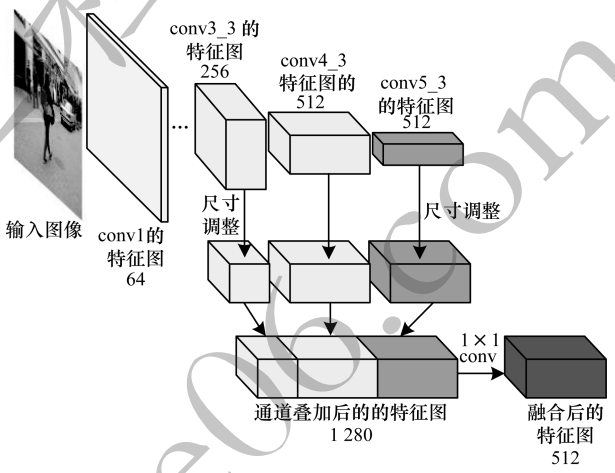


图 7 特征融合示意图

Fig. 7 Schematic diagram of feature fusion

3 实验与结果分析

为验证本文改进方法的有效性，在计算机上进行仿真，对原 Faster R-CNN 目标检测网络和改进后的 Faster R-CNN 进行比较。使用 PyTorch 深度学习框架，在 16 GB 内存、4 GB 显存 NVIDIA GeForce GTX1050Ti GPU 并搭有 CUDA8.0 和 CUDNN5.1 运算平台的笔记本电脑上进行训练。本文实验基于 PASCAL VOC2007、PASCAL VOC2012 和 INRIA 数据集。采用 VOC2007 数据集中包含行人的部分作为训练集（共 6 383 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），并将测试集中的正样本作为验证集（共 576 张）。采用 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 1/4 为小尺度行人。

3.1 实验过程

3.1.1 特征融合

本文将 conv3_3 后的特征图通过最大池化下采

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当前流行的反卷积[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,其实验效果相较于反卷积更好。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 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 Relu 非线性激活函数,进一步增强网络的表达能力。

3.1.2　Soft-NMS 算法

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)的其他框,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,深灰色框的置信度最高,得到保留,而同样检测到行人的浅灰色框会因为与实线框的 IoU 高于设定阈值而被过滤掉,最终导致实线框中行人的漏检。文献[24]针对此问题做出改进,提出了 Soft-NMS 算法。该算法没有直接删除图中的虚线框,而是通过线性加权或是高斯加权的方式降低其置信度,与实线框重叠程度越高则置信度下降得越快,这种方式能够很大程度地保留因重叠而导致被误删的检测框。因此,本文在测试时使用 Soft-NMS 算法替换传统的 NMS 算法,使平均精确率均值(mean Average Precision,mAP)约提升 2% 。

图 8　行人重叠示意图

Fig. 8　Schematic diagram of pedestrians overlapping

3.2　结果分析

3.2.1　RPN 实验

为验证改进 RPN 的有效性,本节使用不同的区域建议网络(分类回归子网络使用相同原 Faster R-CNN)在 INRIA 数据集中进行比较,结果如表 1 所示。

表 1　改进 RPN 前后实验结果对比

Table 1　Comparison of experimental results before and after RPN improvement

候选区域数	方法	mAP/%	时间/ms
2 000	RPN	69.21	231
	改进 RPN	75.43	235
300	RPN	67.74	224
	改进 RPN	73.22	224

从表 1 可以看出,在和原 RPN 提供相同数量的候选区域的情况下,改进 RPN 的 mAP 分别提升了 6.22% 和 5.48% ,这说明改进后所提供的候选区域质量更高。

3.2.2　特征融合实验

为验证特征融合策略的有效性,本节采用不同卷积层的特征图进行融合,使用 INRIA 数据集进行对比实验,结果如表 2 所示,其中,√表示叠加的卷积层。可以看出,3 层叠加的情况能够获得较高的 mAP。

表 2　不同卷积层特征融合效果对比

Table 2　Comparison of feature fusion effects of different convolutional layers

Conv3_3	Conv4_3	Conv5_3	mAP/%
		√	69.21
	√	√	77.43
√		√	78.17
√	√	√	86.79

3.2.3　小尺度行人检测实验

本节使用 PASCAL VOC2012 数据集中提取出的小尺度行人测试集进行实验,结果如表 3 所示。

表 3　PASCAL VOC2012 数据集实验结果对比

Table 3　Comparison of experimental results in PASCAL VOC2012 dataset

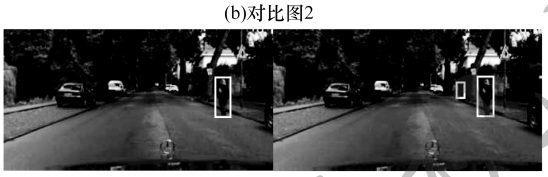
方法	mAP/%	时间/ms
改进 Faster R-CNN	83.15	347
Faster R-CNN	59.37	336
HOG + SVM	46.86	—

表 3 显示本文改进方法对于小尺度行人检测的 mAP 提高了 23.78% ,检测时间略有延长。这表明本文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方法有少许增加,但换来的是对于小尺度行人检测能力的大幅提升。

图 9(a)~图 9(c)为改进方法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,它们更直观地展示了改进方法检测小尺度行人的性能,图中左图为 Faster R-CNN 检测结果,右图为改进方法检测结果。

(a)对比图1

(b)对比图2



(c)对比图3

图 9 Faster R-CNN 与改进方法的检测结果对比

Fig. 9 Comparison of detection results between Faster R-CNN and improved method

从图 9(a)中可以看出,改进方法不仅检测出了 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,这说明改进方法对于行人的遮挡具有一定的鲁棒性。图 9(b)和图 9(c)是由车载摄像头采集到的图像,可以看出改进网络均能检测出道路远端的小尺度行人,这也说明在自动驾驶场景下改进方法对于小尺度行人检测能力也有很大的提高。

4 结束语

为提高小尺度行人的检测能力,本文以 Faster R-CNN 为基础,针对行人的特点重新设置 RPN 中的候选框参数,采用基于双线性插值的对齐池化方式以避免在感兴趣区域池化中两次量化操作带来的位置偏差,同时针对小尺度行人在深层特征图中特征信息的不足,设计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,使用最终融合的特征进行行人检测。实验结果表明,本文方法能够有效提高改进 Faster R-CNN 网络对于小尺度行人的检测性能。但该方法的改进 Faster R-CNN 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,因此,后续将在本文工作基础上对行人的自遮挡问题进行研究,进一步提高行人检测的准确度。

参考文献

[1] ZHANG S S, BENENSON R, SCHIELE B, et al. CityPersons: a diverse dataset for pedestrian detection[C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 4457-4465.

[2] XUE Lixia, ZHONG Xin, WANG Ronggui, et al. Mid-low resolution vehicle type recognition based on deep feature fusion[J]. Computer Engineering, 2019, 45(1): 233-238, 245. (in Chinese)
薛丽霞, 钟欣, 汪荣贵, 等. 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] LI Hongyan, LI Chungeng, AN Jubai, et al. Attention mechanism improves CNN remote sensing[J]. Journal of Image and Graphics, 2019, 24(8): 1400-1408. (in Chinese)
李红艳, 李春庚, 安居白, 等. 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(8): 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 Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9): 1904-1916.

[6] GIRSHICK R. Fast R-CNN[EB/OL]. [2019-05-10]. https://arxiv.org/pdf/1504.08083.pdf.

[7] DAI Jifeng, LI Yi, HE Kaiming, et al. R-FCN: object detection via region-based fully convolutional networks[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 379-387.

[8] REDMON J, FARHADI A. YOLOv3: an incremental improvement[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 1-6.

[9] JIAO Tianchi, LI Qiang, LIN Maosong, et al. Target detection method combining inverted residual block and YOLOv3[J]. Transducer and Microsystem Technologies, 2019, 38(9): 144-146, 156. (in Chinese)
焦天驰, 李强, 林茂松, 等. 结合反残差块和 YOLOv3 的目标检测法[J]. 传感器与微系统, 2019, 38(9): 144-146, 156.

[10] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands: [s. n.], 2016: 21-37.

[11] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 2999-3007.

(下转第 241 页)

(上接第 232 页)

[12] ZHANG Liliang, LIN Liang, LIANG Xiaodan, et al. Is Faster R-CNN doing well for pedestrian detection? [C]// Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 443-457.

[13] LI Jianan, LIANG Xiaodan, SHEN Shengmei, et al. Scale-aware Fast R-CNN for pedestrian detection[C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 1-6.

[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. Amsterdam, the Netherlands: [s. n.], 2016: 354-370.

[15] MAO Jiayuan, XIAO Tete, JIANG Yuning, et al. What can help pedestrian detection? [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 1-5.

[16] LI Chengyang, SONG Dan, TONG Ruofeng, et al. Illumination-aware Faster R-CNN for robust multispectral pedestrian detection [EB/OL]. [2019-05-10]. https://arxiv. org/pdf/1803. 05347. pdf.

[17] LIU Songtao, HUANG Di, WANG Yunhong. Adaptive NMS: refining pedestrian detection in a crowd[EB/OL]. [2019-05-10]. https://arxiv. org/pdf/1904. 03629v1. pdf.

[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]//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2014: 1-5.

[20] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2019-05-10]. https://arxiv. org/pdf/1409. 1556. pdf.

[21] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[C]// Proceedings of 2017 IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 1-12.

[22] SHI W, CABALLERO J, THEIS L, et al. Is the deconvolution layer the same as a convolutional layer? [EB/OL]. [2019-05-10]. https://arxiv. org/ftp/arxiv/papers/1609/1609. 07009. pdf.

[23] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C]// Proceedings of International Conference on International Conference on Machine Learning. Miami, USA: [s. n.], 2015: 1-11.

[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. Washington D. C., USA: IEEE Press, 2017: 5562-5570.

编辑　陆燕菲