

# 基于改进实例分割网络的步行道违停车辆检测算法<sup>①</sup>

郑雅羽<sup>②\*</sup> 王济浩<sup>\*</sup> 冯 杰<sup>\*\*</sup>

( \* 浙江工业大学信息工程学院 杭州 310023)

( \*\* 浙江理工大学信息学院 杭州 310018)

**摘 要** 步行道上违法占道停车现象存在安全隐患,传统的监控摄像设备检测和执法人员人工巡检手段分别存在投入成本大、效率低的缺点。针对以上问题,本文基于移动设备采集图像提出了一种对实例分割网络改进后的违停车辆检测算法。通过改进网络的特征传递方式配合连通区域融合的方法改善实例的定位分割结果,利用违停判定方法以实现合规停放车辆与违停车辆的有效区分。实验结果表明,改进网络方法与连通区域融合方法均提升了实例分割网络的识别准确度。相较于直接使用实例分割网络进行违停判定,本文提出的算法对步行道上违停车辆的查全率与查准率达到了 90.5% 和 80.8%,分别提升了 21.4% 与 6.4%。

**关键词** 步行道;违法占道停车;实例分割;特征传递;连通区域融合;违停判定

## 0 引 言

在步行道上占道停车的行为增加了行人因绕行而被迫进入非机动车道乃至机动车道进而引起交通事故的可能性。城市管理部门常通过固定监控设备摄像<sup>[1]</sup>和执法人员巡检来整治违停现象。但由于监控摄像设备的视角固定,容易存在监控盲区、易被人为干扰与躲避,而执法人员巡检存在效率低的问题。随着类似谷歌眼镜等智能可穿戴移动设备的快速发展,可提高巡检时采集相应场景图片的效率。开发和推广应用移动设备采集的图像并自动进行违停识别的技术可加强对违停行为的监管。

移动设备采集到的图像往往由于步行道的宽窄不一、拍摄时因为行进方向的变化导致拍摄视角差异较大,使得不同图像中车辆尺度差别大、包含的背景信息复杂,只采集到车头或车尾等局部信息的特点。使用传统的图像检测方法<sup>[2]</sup>进行违停车辆的自动识别主要有以下 2 个难点:一是传统方法提取

的特征是基于车辆完整出现的前提,因此对图像中只含局部信息的车辆难以有效检出;二是不同于固定监控设备可预先设置步行道的可停区域,由于图片是移动采集的,步行道的可停区域会随时变化,因此需要对图片中的可停区域也能实现自动识别,但由于拍摄视角、车位线磨损、光照影响、可停区域被车辆大面积遮挡形变等原因造成检测困难。

针对以上 2 个难点,可以利用近年来兴起的深度学习领域中的目标检测网络来提高检出率<sup>[3]</sup>。目标检测网络主要分为一步法(YOLO<sup>[4]</sup>系列为代表)和两步法(R-CNN<sup>[5]</sup>系列),本文出于精度和后期多分类的需求选取两步法进行研究。Girshick 等人<sup>[5]</sup>提出的 R-CNN 方法,实现了以 CNN 进行目标检测。R-CNN 在 VOC 数据集上达到了 53.7% 的准确率,相较于传统方法平均 35% 的准确率有了较大的提升,但 R-CNN 的模型复杂度较高。SPP-net<sup>[6]</sup>通过空间金字塔池化加速了提取特征的速度,而后的 Fast R-CNN<sup>[7]</sup>、Faster R-CNN<sup>[8]</sup>在精度和检测速

① 国家自然科学基金(61401398,61501402)资助项目。  
② 男,1978 年生,博士;研究方向:计算机视觉,嵌入式系统应用;联系人,E-mail:yayuzheng@zjut.edu.cn  
(收稿日期:2019-10-10)

度有效提升且实现了端对端的 CNN 目标检测模型。

但在遭遇检测效果不佳的情况下,仅依靠检测网络输出的包围框坐标与类别难以分析出原因,其给出的矩形包围框在面对形状不规则物体时也难以直观体现出对目标物体的检出效果。Mask R-CNN<sup>[9]</sup>是在 Faster R-CNN 基础上结合 FCN<sup>[10]</sup>支路提出的实例分割模型,可以实现在输出物体包围框与类别置信度的同时增加框内目标物体的轮廓掩码信息,使得对物体检测的精准度有了额外的评价指标。此外,可停区域的特征不如车辆特征明显,其一般由车位线框定,但由于车位线内的区域与车位线外的区域存在高度相似的原因,当可停区域存在车位线磨损、光照影响、遮挡变形等情况时,会使包围框和轮廓掩码与真值相差过大,进而影响检出车辆的进一步违停判定。因此需要改进实例分割网络在保证定位准确度的前提下加强可停区域的特征信息。

本文以识别步行道上违停车辆为目标,通过改

进实例分割网络结合连通区域融合的方法提升车辆与可停区域的检测效果。在模型检测的基础上,本文设计违停判定算法以检测违停车辆。

### 1 算法概述

根据检测任务的需求,本文提出的检测步行道上违停车辆的算法框架如图 1 所示。首先将图片输入到对 Mask R-CNN 改进后的网络,通过迁移学习得到输入图片中车辆与可停区域的精准的实例分割结果。步行道上的可停区域往往是由多个紧密相连车位组成的整体区域,而由于实例分割网络具有分割相同类别不同个体的特性使得完整的可停区域被分割成重叠的单个车位的包围框与轮廓掩码,破坏了可停区域的完整性,因此本文设计了连通区域的融合方法改善实例分割网络对可停区域的分割。最后根据处理后的结果设计了一种违停判定方法来有效识别违停车辆。

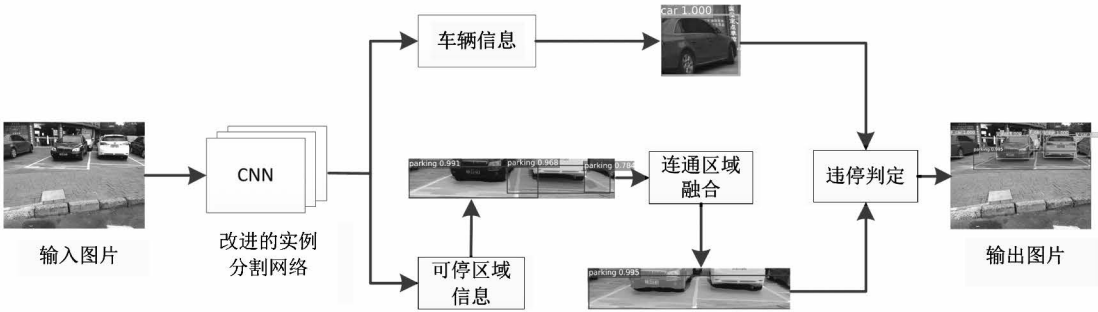


图 1 整体框架图

### 2 改进的实例分割网络

实例分割网络 Mask R-CNN 在实例检测与分割等任务中获得了较好的性能,网络结构如图 2 所示。对于输入图片,首先进入基础网络提取特征,为了更好地提取特征 Mask R-CNN 引入了特征金字塔结构 (feature pyramid network, FPN)<sup>[11]</sup>。将 FPN 处理后产生的多层特征图送入区域提议网络 (region proposal network, RPN) 得到可能包含物体的前景提议框,ROI (region of interest) Align 层根据 FPN 产生的特征图将前景提议框映射成固定尺寸的特征向量。

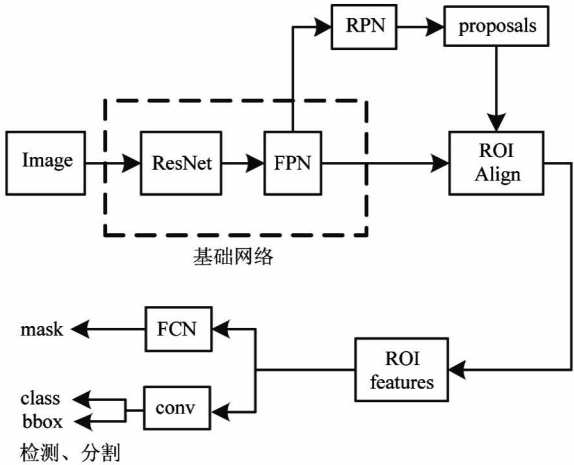


图 2 Mask R-CNN 结构

产生的提议框特征向量最后复用于经过 FCN 支路得到分割图与经过卷积全连接支路得到相应类别与包围框的坐标。

由于本文的连通区域融合与违停判定依赖实例分割网络结果的准确性,而网络对于车辆等物体的强特征性较为敏感,在遇到可停区域中车位线内外存在高度相似的特征时,可能出现检测失准的情况。用原始的 Mask R-CNN 网络训练后的检测效果如图 3 所示,通过包围框辅助以轮廓掩码的主观结果可以发现网络对于车辆的检测效果较为精准,但可停区域的包围框与掩码存在偏大的现象,这将直接导致在后续进行违停判定时易产生误判现象。由于 Mask R-CNN 网络会大量复用基础网络中的特征信息输出最终的检测与分割结果,因此特征信息的完整尤为重要,本文为了加强基础网络的特征信息提出了一种低代价的改进方法。



图 3 可停区域检测失准

2.1 网络架构改进

有别于 Faster R-CNN 中基础网络得到单一尺寸的特征图,Mask R-CNN 算法在基础网络中引入 FPN 得到不同尺寸的特征图,解决了多尺度检测的问题。基础网络具体结构如图 4 所示,输入图片经过左侧的 ResNet<sup>[12]</sup> 中不同残差块得到 C2 ~ C5 特征层,右侧特征金字塔 FPN 结构中主要包含 2 个结构:(1)横向连接,使用  $1 \times 1$  卷积提取 ResNet 中的特征层信息并将通道数统一成 256。(2)自上而下,将含有整体抽象信息的上层特征图做 2 倍上采样,然后与下一层经过横向连接的特征相加得到 M2 ~ M5,使得含有定位信息的下层特征图能额外拥有整

体抽象信息,并将相加过后的特征层经过  $3 \times 3$  卷积得到 P2 ~ P5,使得在不改变特征图尺寸的情况下提高感受野(receptive field)。

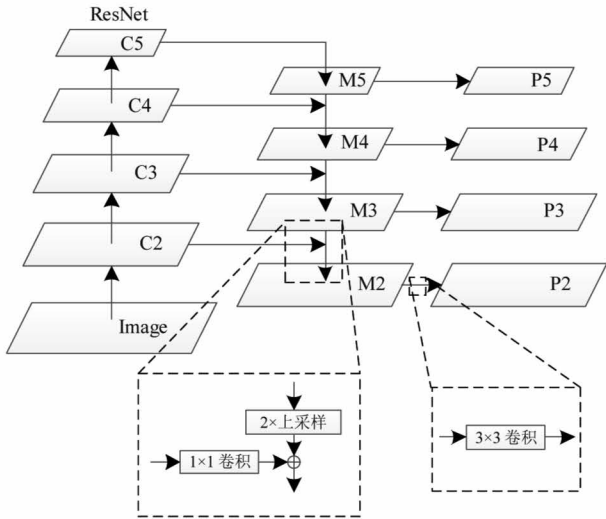


图 4 基础网络结构

Mask R-CNN 的基础网络中的 FPN 结构将上层信息通过上采样的方式传递给下层,虽然使得下层信息更加丰富,但上层含有的定位信息并没有得到加强,特别是输入图片经过 ResNet 中 C2 ~ C5 共计 101 层后会损失过多残差网络提取到的下层的定位信息。基于以上不足,本文提出了一种改进的基础网络结构以改善特征信息的传递,具体的连接方式如图 5 所示,在保持原有的横向连接与自上而下的结构的基础上,额外增加了一种自下而上的连接。

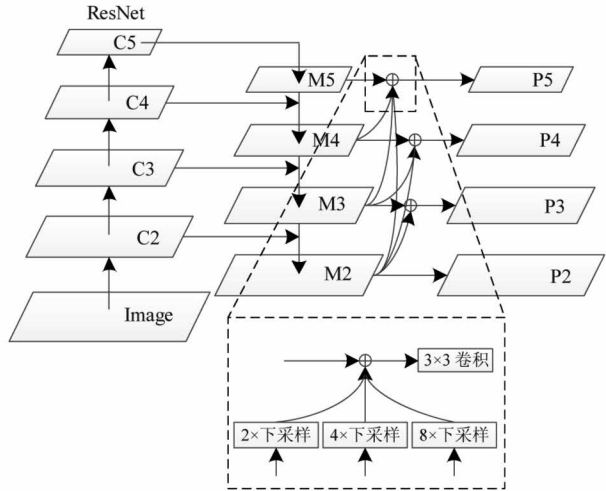


图 5 改进的基础网络结构

图片输入到改进后的网络中,首先经过 ResNet,其卷积核与步长的选取遵循 ResNet 网络的设计,C2 ~ C5 的特征层的选取也与基础网络结构中的 ResNet 相同。在得到 C2 ~ C5 特征层后,使用基础网络中相同的 FPN 网络中的横向连接与自下而上的连接得到 M2 ~ M5 层,对于 M5 层首先将位于该层之下的 M4 ~ M2 层做相应 2、4、8 倍的下采样,将各层与 M5 层进行尺寸对齐,之后通过稠密连接<sup>[13]</sup>与 M5 层相加后通过  $3 \times 3$  卷积得到 P5 层,M4 与 M3 的修改机制同上。通过稠密连接的方式使得上层快速获得下层的定位信息,这一过程中只包含不到 10 层,但能有效加强上层的定位信息,使得网络对大物体的定位更加精准。

2.2 损失函数

本文所使用的模型为多任务模型,其整体损失函数为

$$L = L_{cls} + L_{box} + L_p + L_{mask} \tag{1}$$

其中, $L_{cls}$ 为检测类别的损失, $L_{box}$ 为检测包围框坐标的回归损失, $L_{mask}$ 为包围框内分割损失, $L_p$ 为 RPN 网络的损失。其中  $L_{cls}$ 、 $L_{box}$ 、 $L_p$  三者均与 Faster R-CNN 的损失函数相同,这里不做详细介绍。对于  $L_{mask}$  损失,由于在训练时,网络会得到固定数目为  $n$  的提议框,在经过 FCN 分割后会得到  $n$  个  $28 \times 28$  特征向量,其中每一个元素均为  $0 \sim 1$  的概率值,对于单个元素,分类的交叉熵表示如式(2)所示:

$$H(y, \hat{y}) = -y \log \hat{y} \tag{2}$$

其中, $y$  为类别的真值,  $\hat{y}$  为该类别的预测值。由于特征向量为  $28 \times 28$ ,且总数为  $n$ ,则  $L_{mask}$  的损失函数如式(3)所示。

$$L_{mask} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{28 \times 28} H(y_j, \hat{y}_j) \tag{3}$$

$L_{mask}$  的损失函数为各类别交叉熵和的平均,考虑到本文模型的类别数较小,将  $n$  从原有的 200 减少为 100 以减少模型过多的参数。

2.3 迁移学习

深度学习模型与传统分类器的显著区别之一是前者需要海量含有真值标签的样本,而获得这些样本以及标注工作往往成本巨大。迁移学习(transfer learning)借助通用数据集上训练完成的模型得到图像特征的先验知识,例如,模型具有帮助识别边缘、

纹理、形状和物体组成的能力,因此迁移学习不受传统的数据同分布假设的限制,在开始新任务时只需在该模型上进一步训练而无需从零开始训练参数。

本文借助于迁移学习中的微调(fine tuning)方法修改模型。(1)获得源数据集上训练完成的源模型。(2)根据目标数据集得到目标模型,在训练开始前复制源模型上除输出层外的所有模型设计及参数。(3)为目标模型添加输出大小为目标任务所需分类个数的输出层,并随机初始化该输出层参数。(4)在目标数据集上训练目标模型,其中输出层从头开始训练,其余层的参数都是基于源模型的参数微调得到的。

3 连通区域的融合

图像的二维性、拍摄视角、车位线紧密相连等原因使得步行道上的停车区域在采集图像中存在形变、被遮挡的可能,造成可停区域的面积大小、区域轮廓不存在统一的实例特征,因此模型输出的可停区域包围框存在分割相连通区域的问题,这易造成同一块连通的可停区域具有多个包围框的现象。实例分割输出包围框前使用非极大值抑制(non-maximum suppression, NMS)<sup>[14]</sup>来移除同一实例重叠的包围框,框与框之间的重叠率使用 IOU(intersection over union)来计算。由于 NMS 方法是基于重叠包围框之间的面积、长宽比近似的假设,因此当 2 个包围框之间面积与长宽比差异较大时,NMS 并不能有效抑制重叠的包围框,例如大包围框中包含小包围框,这在对可停区域的检测中大概率出现。

基于以上的不足,本文在第 2 节得到的模型基础上提出了一种连通区域的融合方法来加强后处理。假设实例分割得到的可停区域包围框为  $a$  和  $b$ ,且存在相交的矩形区域记为  $I_{ab}$ ,该区域中属于框  $a$  和  $b$  的可停区域掩码分别记为  $m_a$  和  $m_b$ ,根据  $I_{ab}$  中存在的掩码来判定包围框之间是否存在相连通的区域,其判定如式(4)所示:

$$M = \frac{area(m_a) \cap area(m_b)}{area(m_a) \cup area(m_b)} \tag{4}$$

其中,  $area(m_a)$  和  $area(m_b)$  代表  $I_{ab}$  中属于包围框  $a$  和  $b$  的掩码面积,其掩码交集如图 6 所示,计算区

域内相应掩码的交并比  $M$ , 当  $M$  大于阈值  $\theta$  时, 则判定包围框  $a$  和  $b$  存在相连通的区域, 将实例分割结果中的分割掩码做相并操作进行融合, 其融合后的包围框为掩码区域的外接矩形框, 其相应的置信度选取融合前掩码面积最大的包围框置信度。当存在 2 个以上的可停区域包围框时, 按照上述连通检测方法构造图, 使用深度优先搜索策略得到相连通的包围框, 重复上述的融合方法。由于计算交并比限制在包围框的相交区域, 且基于掩码的叠加程度来衡量是否相连通, 相比传统的 NMS 检测方法, 上述方法一方面缩小了搜寻区域, 另一方面基于轮廓掩码的交并比对于物体重叠的判定相较于包围框会更加精确。

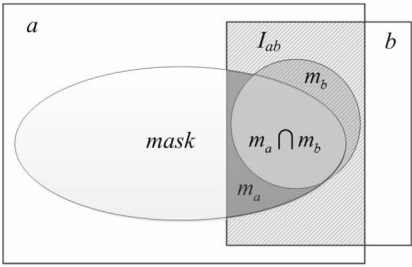


图6 可停区域掩码交集

## 4 违停判定

为了将合规停放车辆与违停车辆区分开来, 需要进一步将违停与非违停的车辆标记, 具体流程如图7所示, 首先借助改进后的网络与融合方法得到的车辆与可停区域的检测结果判定是否检测到车辆, 若车辆不存在, 则无需进行违停判定; 若车辆存在, 则进一步判定是否存在可停区域。若不存在可停区域, 则将检测到的车辆标记为违停车辆; 若存在可停区域, 需要进一步计算车辆包围框  $B_c$  与可停区域包围框  $B_p$  的重叠率  $a$  与相对位置关系, 对于重叠率  $a$  的计算如式(5)和(6)所示。

$$a = \max(a_p, a_c) \quad (5)$$

$$a_p = \frac{\text{area}(B_p) \cap \text{area}(B_c)}{\text{area}(B_p)} \quad (6)$$

$$a_c = \frac{\text{area}(B_p) \cap \text{area}(B_c)}{\text{area}(B_c)}$$

对于检测结果包含 2 种情况: (1) 当可停区域

大面积被遮挡时其包围框面积较小, 计算  $B_c$  与  $B_p$  的相交矩形区域面积与  $B_p$  面积的比率  $a_p$ ; (2) 当可停区域包围框面积较大时, 车辆包围框面积相对较小, 计算  $B_c$  与  $B_p$  的相交矩形区域面积与  $B_c$  面积的比率  $a_c$ 。重叠率  $a$  为  $a_p$  与  $a_c$  中的较大值。当  $a$  大于阈值  $t$  时认为车辆与可停区域可能存在空间上的重叠关系, 同时根据先验知识, 若为合规停车则该车必定在可停区域的上方, 空间位置关系可以用包围框中心点的纵坐标  $y$  来表示, 当满足以上 2 种情况时, 则判定该车为合规停放车辆, 若不满足则判定为违停车辆。

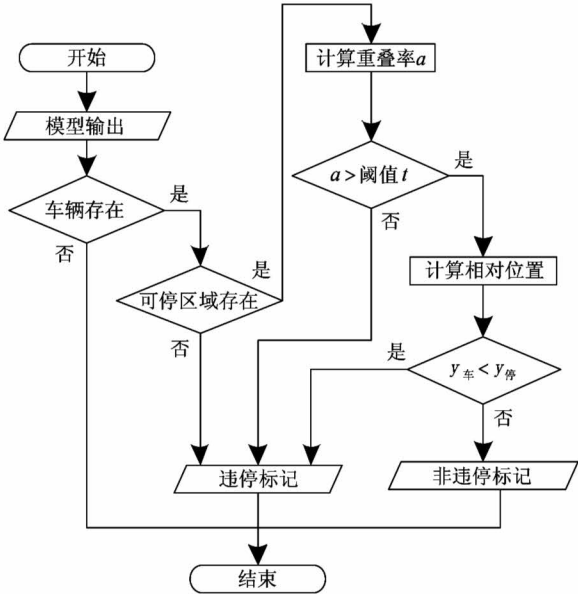


图7 违停判定步骤

## 5 实验

### 5.1 数据集扩充

由于步行道上的违停车辆数据集较为特殊, 网上没有相应公开的数据集, 因此需要自建数据集与其标注信息。本文的实验数据集采集于杭州市城区某步行道, 数据集包括 607 张车辆停放于步行道的图片, 尺寸统一为  $3\,264 \times 2\,448$  像素。每张图片均含有待分类的目标, 并且通过标注软件描绘关键点构建多边形对相应车辆与可停区域进行人工标注获得所需的标注信息。同时, 为了验证改进模型的效果, 本文使用 COCO2017 数据集<sup>[15]</sup>进行对比实验。

由于原始城区步行道数据集的数量并不能完全拟合模型, 数据量不足时易发生过拟合现象。因此,



本文在原数据集的基础上对图片进行随机改变,产生相似但又不同的图片以达到扩充数据集的目的。扩充后的数据集能有效防止过拟合现象,同时随机改变样本能减少模型对某些属性的依赖,从而提高模型的泛化能力。

考虑到变换后的图像应当尽可能与真实采集到的图像相类似,本文使用了随机水平翻转、小比例缩放与旋转、颜色空间抖动以减少模型对于颜色空间、目标物体所处位置的依赖。随机叠加以上几种变换后得到的数据集图片能使原始数据集规模提高 2 ~ 4 倍。在对图片做各种变换的同时,每张图片的标注信息也做相同的变换,从而避免大量的标注操作。

5.2 评价指标

在评价指标方面,相较于 VOC 数据集的评价指标<sup>[16]</sup>,COCO 官方提供的标准评价指标更为严格与全面。因此,后文基于 COCO 评价指标进行实例分割结果的分析。

COCO 评价指标是基于不同 IOU 阈值下的计算每个类别在不同查全率 (recall) 下查准率 (precision) 的平均值,称为平均准确率 (average precision, AP)。对于综合所有类别的 AP 结果,使用 mAP (mean average precision) 进行定量分析。AP 与 mAP 的计算如式(7)和(8)所示:

$$AP = \int_0^1 P(R) dR \tag{7}$$

$$mAP = \frac{\sum_{n=1}^N AP(n)}{N} \tag{8}$$

其中, $P$  为查准率, $R$  为查全率, $N$  为模型需要分类的类别总数, $P(R)$  为查全率为  $R$  时的查准率, $AP(n)$  为类  $n$  对应的 AP 值。 $P$  和  $R$  的计算如式(9)和(10)所示:

$$P = \frac{TP}{TP + FP} \tag{9}$$

$$R = \frac{TP}{TP + FN} \tag{10}$$

其中, $TP$  表示预测正例正确, $FP$  表示预测正例错误, $FN$  表示预测负例错误。COCO 评价指标中的默认 AP 为 mAP,为避免歧义,后续统一使用 mAP 来表示模型的性能。

5.3 实验结果与分析

实验所用服务器搭载 Ubuntu16.04 操作系统,16 G 内存,Intel(R) Core(TM) i7-7800X 处理器,主频为 3.5 GHz,Nvidia GeForce GTX 1080 Ti 显卡。

首先,为验证针对基础网络进行修改具有一定的通用性,本文以未修改的 Mask R-CNN 作为比较对象,在 COCO2017 数据集上进行实验,比较最后评价指标的差异。数据集所带标注含有 80 类不同物体的信息。2 个网络所含有的特征提取网络均为 ResNet101,加载相同的预训练参数,动量参数设置为 0.9,权值衰减设置为  $1e-4$ ,学习率设置为  $1e-3$ ,batch-size 设置为 1。通过迁移学习微调 ResNet、RPN、掩码与分类的输出层,训练周期个数为 25 个,保存最优模型。从验证集中随机抽取 500 张作为测试集,在 COCO 数据集上训练 2 个模型,最终得到的结果如表 1 所示。

表 1 基础网络修改在 COCO 数据集上的影响

| 模型         | mAP  | mAP <sub>50</sub> | mAP <sup>bb</sup> | mAP <sup>bb</sup> <sub>50</sub> |
|------------|------|-------------------|-------------------|---------------------------------|
| Mask R-CNN | 32.2 | 51.9              | 36.3              | 54.1                            |
| 修改基础网络     | 32.5 | 52.5              | 36.8              | 55.0                            |

表 1 中 mAP<sub>50</sub>指的是在 IOU = 0.5 时,计算掩码的 mAP 值,其中 mAP 是 mAP<sub>50</sub>,mAP<sub>55</sub>,mAP<sub>60</sub>,...,mAP<sub>95</sub>的平均值。表 1 中 mAP<sup>bb</sup>指的是在 IOU = 0.5 时,计算包围框的 mAP 值,计算均值方法同上。由表 1 可知,相较于 Mask R-CNN,COCO 数据集上基础网络修改后的模型在掩码以及包围框的 mAP 评价指标精度上分别提升了 0.3 与 0.5。由于 COCO 数据集中的物体类别大多具有强特征性,不存在像可停区域类别的物体,因此提升效果不显著。

为验证模型对数据集中车辆与可停区域的检测分割效果,本文从 607 张原始图片中随机挑选 70 张作为测试集,剩下的图片进行数据增强最终得到 1 396 张图片作为训练集。由于类别数为 2,因此将原始 Mask R-CNN 中掩码与分类的输出层中的类别数由 80 改为 2,考虑到类别数量少,将特征提取网络 ResNet101 改为 ResNet50 来降低模型复杂度防止过拟合,其中判定连通域融合的阈值  $\theta$  设为 0.2,

违停判定的阈值  $t$  设为 0.2。训练机制同上,在此基础上使用本文提出的方法进行实验,并将经过迁移学习后的 Mask R-CNN 模型参数作为比较对象,最终得到的结果如表 2 所示。

表 2 迁移学习后的 Mask R-CNN 与本文方法的对比

| 模型         | 修改基础网络 | 区域融合 | mAP <sup>bb</sup> | mAP  |
|------------|--------|------|-------------------|------|
| Mask R-CNN | ×      | ×    | 58.0              | 56.2 |
|            | ✓      | ×    | 59.8              | 58.3 |
| 本文方法       | ×      | ✓    | 59.5              | 57.0 |
|            | ✓      | ✓    | 62.2              | 59.7 |

由表 2 可知,单独使用基础网络的修改对于步行道数据集在掩码以及包围框的 mAP 评价指标上分别提升了 2.1 与 1.8;单独使用连通区域融合的后处理方法在指标上提升了 0.8 与 1.5;叠加使用 2 种方法最终提升了 3.5 与 4.2。以上的实验结果表明,本文提出的修改基础网络与连通区域融合方法使得 Mask R-CNN 的输出结果更为精准,且叠加使用能使精度进一步提升。

分析实例分割网络得到的车辆与可停区域的检测分割结果可以验证本文所提出的违停判定方法的有效性。经统计,在 70 张测试图片中步行道上停放的车辆数目为 123 辆,其中 42 辆车疑似违停。本文统计包围框在 IOU = 0.5 时,车辆与违停车辆的 2 项检测指标即查全率与查准率,使用迁移学习得到的 Mask-R-CNN 模型参数作为基准进行对比实验,最终结果如表 3、4 所示。

表 3 车辆的查全率与查准率

| 模型         | 查全率(%) | 查准率(%) |
|------------|--------|--------|
| Mask R-CNN | 96.8   | 94.5   |
| 本文方法       | 98.4   | 94.6   |

表 4 违停车辆的查全率与查准率

| 模型                | 查全率(%) | 查准率(%) |
|-------------------|--------|--------|
| Mask R-CNN + 违停判定 | 69.1   | 74.4   |
| 本文方法              | 90.5   | 80.8   |

根据表 3 的数据可知,迁移学习后的 Mask R-CNN 模型对于车辆的识别性能较强,在该模型的基础上使用本文方法使得车辆的查准率与查全率相较于原有模型有略微的提升。由表 4 可得,仅依靠迁移学习后训练得到的 Mask R-CNN 模型结合违停判定方法对于违停车辆的检测结果较差,而使用本文算法后违停车辆的查全率与查准率均有显著提升,分别达到了 90.5% 与 80.8%,其中违停车辆的查准率提高 6.4%,查全率提高 21.4%,提升效果明显。以上结果表明,对于基础网络的修改与连通区域融合方法有助于改进违停检测精度,改进后的精度基本满足违停检测任务。

图 8 为测试集中的部分图片的主观视觉效果,第 1 行为原始图片,第 2 行为仅使用迁移学习的 Mask R-CNN 模型的输出结果,car 代表车辆,parking 代表可停区域,每个包围框都包含有相应的置信度与框内的分割掩码。第 3 行为使用本文方法的输出结果,V\_car 代表合规停放的车辆,F\_car 代表违停车辆,parking 代表可停区域。观察对比图可以发现,使用本文算法有效地减少了实例分割网络的误识别概率。以从左往右第 2 列图片为例,未改进的 Mask R-CNN 网络出现了可停区域误识别,进而会影响最后的违停判定结果,而本文提出的检测方法对可停区域的识别与分割更为精准。第 1、3、4 列图片对重叠连通的可停区域进行了有效的融合,且正确区分了合规车辆与违停车辆。如第 5 列图所示为误识别情况,本文的模型假定原始图片均采集于步行道,针对停放于步行道上的车辆进行违停与否的自动化检测。但第 5 组图中的背景并不全为步行道,存在另外 2 辆疑似违停车辆,违反了本文的假定。因而,虽然模型检出了近处的合规车辆,但由于模型统一检测具有车辆特征的对象导致远处背景的违停车辆同样被识别,由此干扰了识别结果并产生了误判。

表 5 是基于服务器上的图形处理单元(GPU)对本文提出的方法进行实验对比,统计了测试集中单张图片的平均运行耗时。根据表 5 的数据可得,Mask R-CNN 处理单张图片平均耗时 1 050 ms,仅修改基础网络增加运行耗时 53 ms,仅使用连通区域融

合方法增加运行耗时 69 ms,使用违停判定增加运行耗时 31 ms。当修改基础网络叠加连通区域融合方法后增加运行耗时 110 ms。结合表 2 数据可以发现,在原有模型运行耗时增加 10.5% 的低代价下,可使掩码与包围框的 mAP 最高提升 3.5 与 4.2。结合表 4,在原有模型的基础上使用违停判定方法可

进一步检测出违停车辆,最后通过叠加基础网络的修改与连通区域融合的方法,使得违停车辆的查全率与查准率分别提升 21.4% 与 6.4%,总耗时为 1 185 ms,相比最初的 Mask R-CNN 检测仅增加了 12.8%。



图 8 主观检测结果

表 5 单张图片的运行耗时

| 模型         | 修改基础网络 | 区域融合 | 违停判定 | 运行耗时 (ms) |
|------------|--------|------|------|-----------|
| Mask R-CNN | ×      | ×    | ×    | 1 050     |
|            | ✓      | ×    | ×    | 1 103     |
| 本文方法       | ×      | ✓    | ×    | 1 119     |
|            | ×      | ×    | ✓    | 1 081     |
|            | ✓      | ✓    | ×    | 1 160     |
|            | ✓      | ✓    | ✓    | 1 185     |

6 结 论

本文针对步行道上违法占道停车现象提出了基于改进实例分割网络的步行道违停车辆检测算法。本文通过对原有的 Mask R-CNN 网络进行基础网络改进显著强化了特征信息的提取过程,通过针对检测任务进行迁移学习有效加速了模型训练,通过设计连通区域融合的方法明显改善了停车区域识别不佳的情况,减少了识别不全与重叠现象。为了合理

区分合规车辆与违停车辆,本文利用前述步骤得到的结果,设计了一种区分算法,其最终对于违停车辆的查全率与查准率高达 90.5% 与 80.8%,且检测的图片中包含包围框、置信度、类别信息与分割掩码等更为丰富的输出信息。

进一步分析实验结果可以发现,由于最终检测效果较依赖于实例分割模型的输出,因而当实例分割模型输出出现错误时,将会影响违停车辆的准确识别。因此,如何进一步改进实例分割网络、设计更多相关类别以帮助减少违停车辆的误识别、进一步减小算法的复杂度是今后的研究重点。

参考文献

[ 1 ] 姜明新,王洪玉,沐方顺. 基于视频监控的违章停车检测算法[J]. 计算机工程, 2012, 38(19): 151-153, 158

[ 2 ] 刘琼. 导引概率图与显著特征相结合的行人目标检测[J]. 高技术通讯, 2016, 26(5): 464-474

[ 3 ] 付若楠. 基于深度学习的目标检测研究[D]. 北京:北京交通大学计算机与信息技术学院,2017: 1-2



- [ 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, Las Vegas, USA, 2016: 779-788
- [ 5 ] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[ C ] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA, 2014: 580-587
- [ 6 ] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[ C ] // European Conference on Computer Vision, Zurich, Switzerland, 2014: 346-361
- [ 7 ] Girshick R. Fast R-CNN[ C ] // Proceedings of the IEEE International Conference on Computer Vision, Boston, USA, 2015: 1440-1448
- [ 8 ] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [ C ] // Advances in Neural Information Processing Systems, Montréal, Canada, 2015: 91-99
- [ 9 ] He K, Gkioxari G, Dollár P, et al. Mask R-CNN[ C ] // Proceedings of the IEEE International Conference on Computer Vision, Honolulu, USA, 2017: 2961-2969
- [ 10 ] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[ C ] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 3431-3440
- [ 11 ] 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, Honolulu, USA, 2017: 2117-2125
- [ 12 ] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[ C ] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770-778
- [ 13 ] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks [ C ] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 4700-4708
- [ 14 ] Neubeck A, Van Gool L. Efficient non-maximum suppression[ C ] // International Conference on Pattern Recognition (ICPR06), Hong Kong, China, 2006: 850-855
- [ 15 ] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context[ C ] // European Conference on Computer Vision, Zurich, Switzerland, 2014: 740-755
- [ 16 ] Everingham M, Van Gool L, Williams C K I, et al. The pascal visual object classes (voc) challenge[ J ]. *International Journal of Computer Vision*, 2010, 88 ( 2 ): 303-338

## An illegal sidewalk parking car detection algorithm based on improved instance segmentation network

Zheng Yayu<sup>\*</sup>, Wang Jihao<sup>\*</sup>, Feng Jie<sup>\*\*</sup>

( <sup>\*</sup> College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023 )

( <sup>\*\*</sup> School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018 )

### Abstract

Illegal sidewalk parking has traffic safety hazards. The traditional surveillance camera detection and manual inspection have the disadvantages of high cost and low efficiency. Aiming at the above problems, an illegal parking car detection algorithm is proposed based on the instance segmentation network, which improves the image captured by mobile devices. By optimizing feature transmission mode of the network and the fusion method of connected region, the position and segmentation results of the instance are improved. Also, illegal parking determination method is used to effectively distinguish between legal parking car and illegal parking car. The experimental results show that both the improved network method and the connected region fusion method improve the recognition accuracy of the instance segmentation network. Compared with using instance segmentation network directly to detect illegal parking car, the proposed algorithm has achieved a recall rate and a precision rate of 90.5% and 80.8% for the car on the sidewalk, which is increased by 21.4% and 6.4% respectively.

**Key words:** sidewalk, illegal sidewalk parking, instance segmentation, feature transmission, connected region fusion, illegal parking determination