

一种针对路口监控图像的区域分割方法

李昕蔚^{1,2,3} 丁正彦⁴ 尚岩峰⁴ 祝永新¹ 汪辉¹ 钟雪霞⁴ 田犁¹ 黄尊恺¹ 封松林¹

¹(中国科学院上海高等研究院 上海 201210)

²(上海科技大学信息科学与技术学院 上海 201210)

³(中国科学院大学 北京 101408)

⁴(公安部第三研究所 上海 201204)

摘要 针对智能交通系统中对交通路口场景理解的需求,提出一种基于线特征先验和凸包损失函数的空间分割网络,目标是对斑马线以及斑马线所围路口区域进行精确检测和分割。利用公安交通管理系统平台采集并标注路口数据集;引入线特征先验,将 RGBL 图像作为网络输入,为深度学习实例分割提供显著的物体边缘特征以加强深度网络对图像特征学习的针对性;在分割网络中引入 SCNN 网络结构,构成空间分割网络以增强网络对空间结构的学习;引入凸包二值交叉熵动态损失函数来优化网络的输出精度。实验结果表明,该空间分割网络对斑马线及路口区域的检测正确率和分割完整度和精确度都有了显著的提升。

关键词 智能交通系统 斑马线检测 实例分割 Mask R-CNN

中图分类号 TP183

文献标志码 A

DOI:10.3969/j.issn.1000-386x.2020.03.040

A REGION SEGMENTATION METHOD FOR INTERSECTION MONITORING IMAGES

Li Xinwei^{1,2,3} Ding Zhengyan⁴ Shang Yanfeng⁴ Zhu Yongxin¹ Wang Hui¹

Zhong Xuexia⁴ Tian Li¹ Huang Zunkai¹ Feng Songlin¹

¹(Shanghai Advanced Research Institute, Chinese Academy of Science, Shanghai 201210, China)

²(School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China)

³(University of Chinese Academy of Sciences, Beijing 101408, China)

⁴(The Third Research Institute of Ministry of Public Security, Shanghai 201204, China)

Abstract Aiming at the need of understanding the traffic intersection scene in intelligent transportation system, this paper proposes a spatial segmentation network based on line feature prior and convex hull loss function. The task is to accurately detect and segment the zebra crossing and the intersection area surrounded by the zebra crossing. We used the public security traffic management system platform to collect and mark the intersection data set. Then, in order to provide significant object edge features for deep learning instance segmentation, we introduced a line feature prior and used RGBL image as network input, so as to enhance the pertinence of deep network on image features learning. In addition, SCNN network structure was introduced into the segmentation network to form a spatial segmentation network to enhance the learning of the network on the spatial structure. In the end, a convex hull binary cross-entropy dynamic loss function was introduced to optimize the output accuracy of the network. The experimental results show that the detection accuracy, segmentation integrity and accuracy of zebra crossing and intersection area are greatly improved by the spatial segmentation network.

Keywords Intelligent transportation system Zebra crossing detection Instance segmentation Mask R-CNN

收稿日期:2019-03-04。国家重点研发计划项目(2016YFC0801304);上海市科学技术委员会科研计划项目(18511111302)。

李昕蔚,硕士生,主研领域:深度学习语义分割。丁正彦,助理研究员。尚岩峰,副研究员。祝永新,研究员。汪辉,研究员。钟雪霞,助理研究员。田犁,副研究员。黄尊恺,助理研究员。封松林,研究员。

0 引言

近年来,中国的城市智能交通系统(ITS)得到快速发展,北京、上海、广州等一线城市已经建立了道路交通控制、公共交通指挥与调度、高速公路管理、紧急事件管理、静态交通管理等多个 ITS 协同系统。在智能交通系统中,城市道路交通控制是非常重要的组成部分。其中,对交通路口状态的监控尤为重要,因为交通路口是交通拥堵和交通事故的多发区,主要原因是交通路口的交通规则复杂、车辆行驶路线多样等,往往一些微小交通事故(如车辆逆行、撞车等)都会引发其关联路段(交汇于此路口的两条及以上交通线路)的一系列的交通拥堵。在城市智能交通系统中,智能化算法及时检测路口交通情况发现交通事故源发地点、迅速通知交警路况信息并协调车辆通行是亟需实现的功能。自动化检测交通情况、分析路口车辆行驶状态、分析人与马路的关系、判定事故责任方等任务都建立在对路口环境精确理解的基础上。为了使系统理解路口场景,首先需要场景的内容进行分割,并对车辆行人进行检测。在现有的交通监控系统中,为了获取路口车辆通行时间等信息,会人为对路口区域进行标注,然而支持任意角度旋转的球机摄像头的存在,一方面扩大了监控视角、提升了监控灵活程度,另一方面也增加了人工标注路口信息的难度。所以,本文旨在针对视角可动态改变的交通路口监控任务,设计一种智能的区域分割算法。

1 相关研究

1.1 道路标识检测

斑马线是路口场景道路标识的最重要的组成部分之一。以往的斑马线检测算法多应用于行人和车辆视角图像的盲人辅助和自动驾驶等系统中。这样场景下的斑马线具有物体尺度大、局部线特征明显的特点。目前自动驾驶的研究主要侧重于对道路、行人以及路标的识别检测。

在这种情况下,过去道路标识检测多应用传统的模式识别算法,依赖于人工特征。文献[11]通过提取多种感兴趣区域的最大稳定极值区域,利用 FAST 和 HOG 特征建立每种道路标识的模板。相似地,文献[16]使用 HOG 特征并训练支持向量机来生成类别标

签。另外有一些专门针对斑马线检测的方法,利用斑马线条纹状、黑白两极的特点进行边缘检测,如文献[1]。文献[5]运用霍夫直线检测改进双极系数的方法,对全景图像进行俯视投影,再运用形态学开闭运算处理来获得比较好的检测效果。文献[19]在双极系数的基础上进行改良,应用综合线性判别公式 LDA 和直线检测来进行导盲系统的斑马线识别。为了解决遮挡问题,文献[2]依赖于交通视频进行背景提取以减少在斑马线识别中的前景干扰,利用最大稳定极值区域(MSER)提取图像特征,然后利用 ERANSAC 算法对提取的特征进行挑选。

近年来,在计算机视觉领域,深度学习的方法获得巨大的成功。在路面标识检测方面,一些工作使用神经网络作为特征提取的工具。文献[14]提出了一种提取 MSERs 并融合同一类别区域最后使用 PCANet 和神经网络进行分类的方法。VPGNet^[15]构建了一个包含 20 000 幅标注了车道线、道路标识图像的数据集,实现了对路面标识以及斑马线的检测,他们将图像的像素级标注转换为网格标注,使分割问题转变为检测问题,并在多种天气环境下能够对距离较近的斑马线进行定位。然而,该算法对实时性要求较高,所以使用轻量级网络结构,存在近距离斑马线分割不准确、很难检测到远距离斑马线的问题。

现有的道路标识检测算法对于斑马线的研究主要针对斑马线的识别和检测,而很少针对斑马线的区域分割。同时现有的自动驾驶交通数据集中,通常不对斑马线等标识进行单独标注,并且与自动驾驶等低视角拍摄下的道路图片相对比,路口监控图像具有对比度和分辨率低,光照、遮挡情况复杂,目标区域尺度小或磨损严重等特点,这些目标图片特点导致了在本文任务中直接使用已有算法进行斑马线分割的难度以及不适用性。

1.2 实例分割方法

实例分割(Instance Segmentation)和超像素(Super Pixel)语义分割同为像素级的精确的物体轮廓识别任务,而它相比语义分割更加困难,因为它需要在识别出物体所属类别的同时将同一类别的不同个体区分开来。

Mask R-CNN^[4]是在物体检测深度学习框架 Faster R-CNN^[7]基础上的扩展,采用了同 Faster R-CNN 相似的两阶段处理结构,其中第一阶段 RPN 完全相同。在第二阶段,Mask R-CNN 在分类和检测的基础上加入了分割分支,为每个感兴趣区域(RoI)输出一个二值

掩膜,最后的损失函数是分类、检测框补偿和实例分割三个平行任务损失函数的和。同时,Mask R-CNN 还引入了 RoIAlign 方法,提高了对小目标的分割和检测准确率。Mask R-CNN 相较于其他的实例分割算法,如 MNC、FCIS 等,在分割精确度上有较大的提升。值得一提的是,何凯明等^[6]利用 Mask R-CNN 算法进行 3 000 个类别的检测框标注和 80 个类别的分割标注,实现了远超前于分割标注类别数量的物体分割。

2 算法描述

本文设计了一种基于深度学习的路口场景分割算法。首先我们采集并标注了交通监控路口图片,然后利用神经网络有监督地学习一个从路口图片到目标区域掩膜的映射。然而实验结果显示,路面状况、天气、距离等原因会导致分割目标模糊、尺寸过小,深度神经网络经常出现误检、漏检等现象。此外,路口监控摄像数据涉及公共信息安全,属于敏感文件,很难大量获取。通过与公安部第三研究所合作,采集了 400 个路口摄像头图像。然而,Mask R-CNN 应用的复杂骨干架构在小数据集上的训练存在精度不高的问题。为了提高精度,Mask R-CNN 在训练规模为 3 000 幅左右精确标注的 Cityscapes 数据集时采用了 COCO 的预训练模型初始化权重。然而,现有规模庞大的公开数据集中没有与本文任务场景相似的斑马线和路口区域类别,所以本文旨在不使用预训练模型初始化参数的情况下对交通路口场景的 Mask R-CNN 训练效果进行提升,提出了一种加入线先验的凸包 Mask R-CNN 方法,网络结构如图 1 所示。

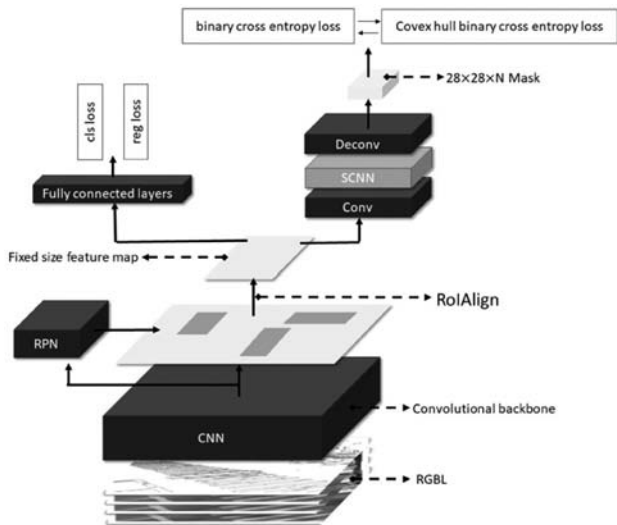


图 1 路口区域分割网络结构示意图

2.1 线先验

为了能让网络更容易且更有效地学习到图像中的重要特征,一种方法是将我们已有的先验知识整合到数据和网络中去。人类接受和处理外界光线信息的人眼,具有非常强的物体分割和区域检测的能力。有研究表明,在人眼视网膜中有一层名为双极细胞的结构,它不通过直接感知光照强度,而是通过感知光线分布的阶梯变化来处理接收到的光照信息,所以双极细胞对物体的边缘信息非常敏感。本文将人眼功能与神经网络相结合,对数据集图片的边缘信息进行提取,与原图一起,构建增加了线特征通道的图像作为深度神经网络的输入,以加强网络对图像特定特征的学习,进而提高检测和分割的精确度。

LSD(Line Segment Detector)^[3]是一个亚像素精度的、线性时间且非常鲁棒的线分割检测器。算法的主要思路可以分解为如下几步:

- 1) 通过 Gaussian 亚采样将输入图像横纵坐标缩短为 80%。
- 2) 计算水平和垂直方向梯度并对梯度进行伪排序。
- 3) 设置梯度临界值排除梯度较小的像素点,用区域生长方法生成线支持区域。
- 4) 将线支持区域矩形化,并计算它的错误警报数(NFA),将不符合要求的矩形区域放入步骤 5) 中继续处理。验证矩形区域的方法基于 Helmholtz principle 和 a-contrario approach。错误警报数的计算公式为:

$$NFA = N_{\text{test}} P_{H_0} [k(r, I) \geq k(r, i)] \quad (1)$$

式中: N_{test} 为当前图像中矩形框的数量; P_{H_0} 是对应 contrario model H_0 的概率; I 是 model H_0 下的随机图像; $k(r, I)$ 是 contrario model I 中 r 矩形里的 aligned point 数量; $k(r, i)$ 观测图像 i 中 r 矩形里的 aligned point 数量。

- 5) 针对被划分在一个矩形区域内的曲线进行处理;对 NFA 不符合要求的矩形区域进行进一步精度提高处理,使其符合要求。

整个 LSD 算法复杂度与图像的像素点数量成正比,可以在线性时间内完成。

由于本文待分割目标之一为斑马线,它的形态特征非常典型,即由多条相对平行但存在透视变化的白色长条矩形构成,线检测算法应用于斑马线可以比较完整地保留它的条纹状特征。然而由于数据集原图中斑马线存在对比度低边缘模糊的情况,我们加入一个图像增强步骤去优化对比度并且减小阴影和白噪声的干扰,使用的 Wallis filter^[17]是局部图像变换。通过估

算不同部分的灰度均值与方差对灰度对比度过大或过小的地方进行平均,公式如下:

$$f(x,y) = r_1 \times g(x,y) + r_0$$

$$r_1 = c \times \frac{s_f}{c \times s_g + (1 - c)s_f} \quad (2)$$

$$r_0 = b \times m_f + (1 - b - r_1)m_g$$

式中: m_f 为局部平均值; s_f 为 f 的方差值; m_g 为原始图像均值; s_g 为原始方差; b 和 c 为亮度系数和对比度提取常数。

同时,输入图片也通常具有非常多的噪声,使得直线检测对噪声过于敏感,故采用 bilateral filter 的改进版本 rolling guidance filter^[18]减少图像中过多的噪声并依然较好地保留边缘信息,方便轮廓的获取和图像分割。本文采用 LSD 线检测算法对预处理后图像进行线检测处理,将线检测处理后的图片作为输入图像的线先验通道。如图 2 所示,网络输入为 RGLB(RGB + Line)四通道信息。

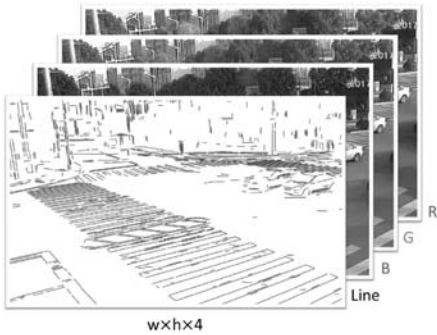


图 2 RGLB 四通道图像

2.2 空间分割网络

在监控图像中,斑马线与路面对比度低甚至部分磨损不清晰的情况较为多见。在一个面积较大的斑马线区域中,斑马线局部的颜色和边缘特征多变,分布不均匀,还有斑马线被车辆等障碍物遮挡中断,这些特征都会导致在进行斑马线区域分割时出现分割不完整的情况。所以,我们针对斑马线局部特征不均匀、特征间断这种空间特性,提出了一种空间分割网络。

本文采用 Mask R-CNN 作为斑马线和路口区域分割任务的基础深度学习网络结构。Mask R-CNN 在 Faster R-CNN 上的结构改进使它在检测分割小物体时具有更高的精确度,具体体现在 Mask R-CNN 网络中的 RoIAlign 在将 RoI 映射到特征图和将不同尺寸特征图采样到统一尺寸的过程中均放弃取整操作。采样过程将特征图划分为 7×7 个同等大小的区域,每个区域平分 n 份(例如, 2×2 份),每份取中心点像素值,采用双线性插值法进行计算。双线性插值公式如下:

$$f(x,y) \approx \frac{f(x_1,y_1)}{(x_2 - x_1)(y_2 - y_1)}(x_2 - x)(y_2 - y) + \frac{f(x_2,y_1)}{(x_2 - x_1)(y_2 - y_1)}(x - x_1)(y_2 - y) + \frac{f(x_1,y_2)}{(x_2 - x_1)(y_2 - y_1)}(x_2 - x)(y - y_1) + \frac{f(x_2,y_2)}{(x_2 - x_1)(y_2 - y_1)}(x - x_1)(y - y_1) \quad (3)$$

接着再将 n 个双线性插值结果进行最大值采样或平均值采样。改进的 RoIAlign 通过插值等方法显著地提升了对小目标的检测分割精确度。

Mask R-CNN 在进行 RoIAlign 操作后,得到所有目标检测框的固定尺寸特征图。在识别和检测框回归分支中,特征图将通过全连接网络,而分割分支会对特征图进行卷积和反卷积操作得到固定尺寸的掩膜。接着进行二值化,使用二值交叉熵损失函数计算损失值。

我们的空间分割网络是在 Mask R-CNN 第二阶段的分割分支基础上进行改进。卷积神经网络具有公认的强大特征提取能力,然而,对于图像中行和列之间空间关系的特征探索却尚为欠缺。文献[12]提出了一种 Spatial CNN 网络,这种网络对于具有强先验形状和图像像素连续性很弱的目标具有更强的空间关系提取能力。这一网络结构首先被应用于自动驾驶的道路线检测当中,对间断的没有外观连续性的道路线有非常好的推理检测能力,同样地,本文的分割目标斑马线和路口区域也具有相似的被车辆大面积遮挡、大面积间断的特点。Spatial CNN 突破了传统的基于马尔可夫随机场或条件随机场来进行空间关系建模的思路,使 CNN 能通过精心设计的网络结构融合空间信息,更有效地学习空间关系,达到识别有潜在连续性和强先验的目标的目的。SCNN 的网络结构如图 3 所示。

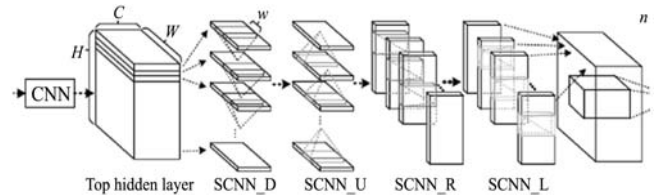


图 3 SCNN 结构示意图^[12]

以 SCNN_D 层为例,为了实现空间关系的学习,网络将 $C \times H \times W$ 的三维向量以 $C \times W$ 为截面进行切片处理,接着使用 $C \times w$ 的卷积核进行分片卷积,将每个切片的处理结果输入到下面一个切片。通过这种方式,信息可以在同一层的不同神经元之间传递。相比于条件随机场等方法,SCNN 可以用较少的计算量达到快速的信息传递效果,它的向前传播计算公式为:

$$X'_{i,j,k} = \begin{cases} X_{i,j,k} & j = 1 \\ X_{i,j,k} + f(\sum_m \sum_n X'_{m,j-1,k+n-1} \times X_{m,i,n}) & j = 2, 3, \dots, H \end{cases} \quad (4)$$

式中: X 为一个三维张量; $X_{i,j,k}$ 为最后一通道 i 元素和当前片通道 j 元素之间的权重, 偏移量为 k 列; $X_{i,j,k}$ 为张量 X 的元素, i, j, k 代表通道、行和列。

SCNN 网络结构的另一个优势是它可以应用于卷积层的任何部位而不仅仅在输出, 来实现对空间信息的学习。在我们的实验数据集中, 存在大量颜色特征不连续、外观线索不连贯的难以完整分割的斑马线。所以为了使 Mask R-CNN 能够更加充分学习空间特征, 在分割分支中, 我们在特征图的卷积操作后加入 SCNN 结构, 对卷积后的特征图的行和列进行向下、向上、向右、向左的切片卷积, 做行与列间的信息交流传递, 最后再进行反卷积的操作。通过引入局部信息的传播, 实现分割网络对斑马线空间关系以及连续性的学习, 使范围过大、中间被遮挡或局部磨损严重的斑马线可以更为精确和完整地分割出来。图 4 为我们的空间分割网络示意图。

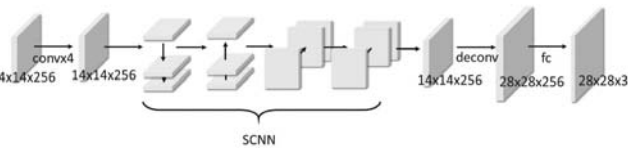


图 4 空间分割网络结构

2.3 凸包损失函数和基于几何形状的实例分割

由于 Mask R-CNN 使用自动编解码的网络结构, 其反卷积层很容易导致输出的斑马线分割结果边界信息模糊甚至丢失, 因而导致在后续的路口分割任务中引入噪声。所以我们希望能对分割任务进行几何上的约束。在对数据集进行分析时发现, 目标斑马线与路口区域为不规则的凸多边形。我们希望对网络的输出结果进行形态结构上的优化, 因此选择采用凸包算法对分割目标的几何形状进行约束。

Mask R-CNN 中的分割分支为每一个 RoI 生成二值掩膜, 然后用目标区域标签类别所对应的掩膜计算损失函数。采用的二值交叉熵损失函数表达式为:

$$L(mask) = -\frac{1}{m} \sum_{1 \leq i, j \leq m} [y_{ij} \log \hat{y}_{ij}^k + (1 - y_{ij}) \log (1 - \hat{y}_{ij}^k)] \quad (5)$$

式中: y 为目标区域真实掩膜; \hat{y} 为网络输出掩膜; m 为掩膜边长; k 代表掩膜对应的真实类别。

在实验中我们发现 Mask R-CNN 分割存在边缘不规则的问题。为了解决这个问题, 最直接的方案是对网络的输出结果进行后续凸包处理, 然而直接对结果

使用凸包算法会将网络的输出膨胀为凸多边形, 从而导致最终的输出结果比真实的目标区域范围大, 进而降低了分割的精确程度。

因此我们将凸包计算融合到网络训练中, 联合优化网络的权重而不是将凸包算法和网络训练分离。为了使深度学习网络能够输出规整的凸多边形形态, 本文改进了分割损失函数, 在原有损失函数的基础上引入凸包二值交叉熵损失项。我们将网络生成的掩膜进行凸包计算, 然后将网络输出的原始掩膜和真实掩膜、凸包掩膜进行二值交叉熵损失函数的计算。引入了凸包二值交叉熵后的损失函数表达式为:

$$L(mask) = -\frac{1}{m} \sum_{1 \leq i, j \leq m} \{ [y_{ij} \log \hat{y}_{ij}^k + (1 - y_{ij}) \log (1 - \hat{y}_{ij}^k)] + [\tilde{y}_{ij}^k \log \hat{y}_{ij}^k + (1 - \tilde{y}_{ij}^k) \log (1 - \hat{y}_{ij}^k)] \} \quad (6)$$

式中: \hat{y}^k 代表第 k 类掩膜的凸包结果。

改进的损失函数使训练网络在拟合目标分割区域的同时不断对生成的分割区域强化几何形态凸包约束, 使训练结果可以更好地适应斑马线的分割任务。在应用改进后的网络生成掩膜时也加入了凸包函数, 在不牺牲斑马线分割精确度的前提下得到形状规则、边缘清晰的斑马线及路口区域分割结果。

3 实验

3.1 交通路口图像区域分割数据集

2012 年, 著名的 GoogleBrain 项目触发了深度学习新一轮的高速发展, 深度学习的快速发展得益于 GPU 并行计算能力的提升和多年来互联网数据的积累。对于自动驾驶任务, 目前已经存在非常丰富的公开数据集资源, 如 KITTI^[8]、cityscapes^[9] 等。另外, SYNTHIA^[10] 数据集由图像模拟器生成, 数据集中存在一部分类似交通摄像头视角图片, 但是它的主要功能是辅助真实数据集的训练以提高分割精确度。这些数据集中图像采集自欧美, 道路特征与国内道路区别较大, 不满足我们对交通路口监控视角真实场景的分析需求, 并且这些数据集的标注类别集中在行人、车辆、周边环境等, 缺少对路面具体区域如斑马线、路口的标注。考虑到在现实生活中交通路口在整个交通系统中突出的重要性, 针对交通路口的算法研究和实际应用会有长远的发展, 专门针对交通路口的监控视角图像数据集是现在非常稀缺的资源。

如表 1 所示, 本文为进行实验构建了规模为 400 幅的交通路口数据集, 图像采集自中国某市区公安交通管理系统平台。数据集内标注物体类别两类, 分别

为斑马线区域和路口区域(十字路口、三岔路口等路口类型),每幅图片含有路口区域一个(大部分图片)或零个(少数图片),斑马线一条及以上,数据集中共标注斑马线条数 1 221 条,路口区域 380 个。数据集包含有日出、正午、日落等不同时刻的采集图片,图像的光照强度以及光照角度多样,也有大量摄像头模糊、斑马线不清晰的图像。此外,数据集内图像中的路口场景复杂,斑马线和十字路口被车辆遮挡、路口周围环境多样、阳光照射在十字路口中而形成大片阴影区域等,这些特征能很好地适应实际应用场景,但同时也极大地影响了斑马线和路口区域的检测。

表 1 实验数据集统计

数据类别	训练集	测试集	总计
图片/幅	360	40	400
斑马线/条	1 095	126	1 221
路口区域/个	343	37	380

数据集中图像大致可分为三类,如图 5(a)所示,分别为常规斑马线、不规则斑马线和远景,数量占比为 40%、40% 和 20%。图 5(b)为三种路口数据标注效果图。

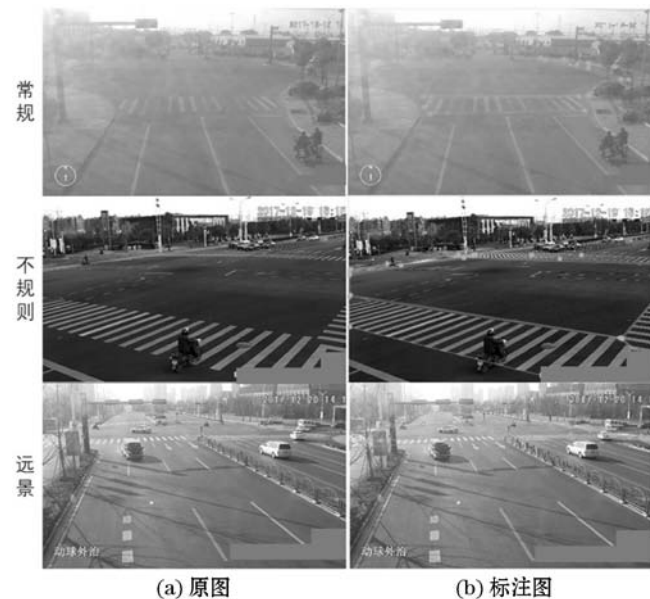


图 5 三种类型路口原图及标注示意图

数据集实用性强,与国内交通环境相适应。本数据集暂未公开,但本文涉及的线增强凸包 Mask R-CNN 算法,对于与本数据集类似场景的斑马线检测和分割任务同样适用,尤其适用于交通摄像图像中视角较远、画质不清,画面内容较丰富的情况。

数据集标注采用 Labelme 工具进行像素级标注。本数据集定义斑马线区域为含有同一方向白色条纹的一条完整的斑马线区域,每一条完整斑马线为一个实

例。定义路口区域为由斑马线和路边石或其他物体(如减速带)等所围成的路面区域。标注工具生成的标注文件为 VOC 数据集类型标注文件,本文将其转换为 coco 类型标注文件以方便数据训练。

3.2 实验结果

本文用 Mask R-CNN 深度学习网络和凸包 Mask R-CNN 网络分别对数据集中的原图、RGBL 类型图像进行训练。本文将数据集中图片按路口类型比例构建由 40 幅图片组成的测试集,其余图像为训练集,其中训练集中包含斑马线 1 095 条,路口区域 287 个,测试集中包含斑马线 126 条,路口区域 37 个。实验采用 ResNet-101-FPN_3x 为基本卷积架构,在两个 GPU 上训练,迭代 50 000 次。训练中使用初始学习速率为 0.013,在迭代 40 000 次和 45 000 次时分别削减十分之一,权重衰减为 0.000 1,动量为 0.9。另外,在训练模型时采用了数据增强处理,在输入数据上添加零均值、随机数 γ 和增益 ρ ,这种方法可以有效地增强网络在不同光照下的性能。公式为:

$$I = \rho \times \left(\frac{I - \tilde{I}}{127} \right) \gamma \quad (7)$$

3.2.1 线先验空间分割网络实验结果

通过使用测试图像验证,我们发现加入线特征先验空间分割网络提高了分割任务的精确度。如图 6 所示,在(b)、(c)、(d)、(e)中,从左到右三幅图片分别为原图、Mask R-CNN 模型和线先验空间分割模型测试结果细节放大图。如图 6(b)所示,在原图中,深色矩形所框出的斑马线在图片边缘只露出狭窄的一部分且特征不突出,Mask R-CNN 网络只将半条斑马线检测和分割出来,而线先验空间分割网络将整条斑马线检测出来并分割得相对准确,提升了分割的精确度。但是,通过观察细节图可以发现,它对分割边缘的处理非常粗糙,不能非常细致和规整地对斑马线和路口进行分割。如图 6(c)所示,在浅色矩形框出的图片右下角是原始图像中的文字部分,Mask R-CNN 将其误检为斑马线并进行错误分割,而加入线先验后纠正了这一错误。另外,加入线先验空间分割网络也提高了网络的检测正确率。如图 6(d)所示,Mask R-CNN 分割结果细节放大图中,深色矩形框出了网络对较远路面区域的误检,将斑马线外的路面区域检测为路口区域;如图 6(e)所示,网络漏检了远处不清晰的斑马线,使路面区域与斑马线连结成一片。而线先验空间分割网络纠正了这两个错误,提高了检测的正确率。实例分割结果见表 2。

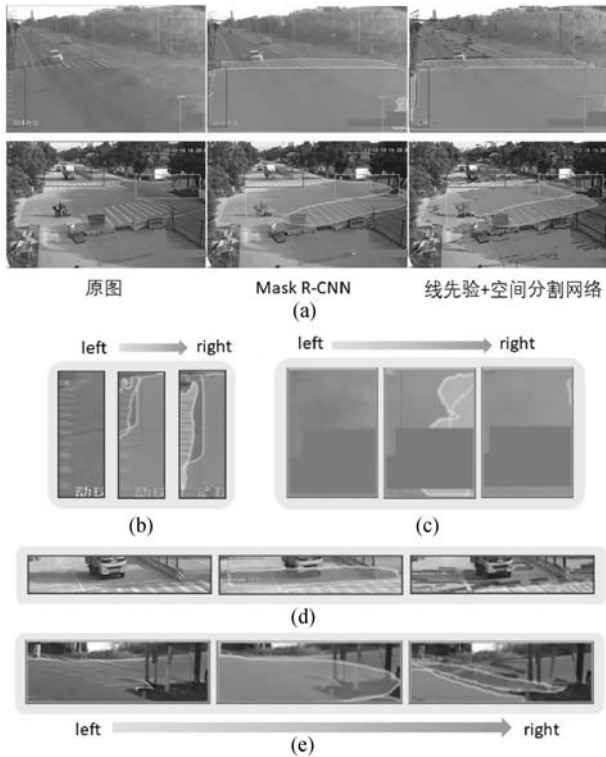


图6 线先验空间分割网络实验结果

表2 实例分割结果数据对比

	Mask R-CNN	线先验 + 空间分割网络
AP	0.491 8	0.516 6
AP ₅₀	0.815 0	0.769 0
AP ₇₅	0.533 3	0.572 0
AP _s	0.235 6	0.265 0
AP _M	0.281 0	0.318 0
AP _L	0.517 0	0.533 0
AP _L	0.517 0	0.533 0

图7列出了三组原图、Mask R-CNN 测试结果图、线先验空间分割测试结果图,分别为数据集中三种类型路口图像的测试效果,即常规、不规则和远景。同时我们与 VPGNet 中的道路线检测结果进行分割结果对比(VPGNet 没有公布预先训练好的模型和数据集,所以我们从它的结果视频里选取了包含斑马线的图帧进行对比)。VPGNet 使用的数据集中标注有 10 000 多条斑马线,可以看到自动驾驶数据集在分辨率、清晰度和对比度质量比较高,而本文所针对的交通路口画质更模糊、斑马线清晰度低,具有更大的分割难度。但是在斑马线分割这一任务对比下,本文算法能够在不规则和远景斑马线具有更好的分割完整度和精确度。而且对于这三种类型的路口图像,本文提出的线先验空间分割网络相比于 Mask R-CNN 网络分别都很大程度提升了分割和检测的精确度。

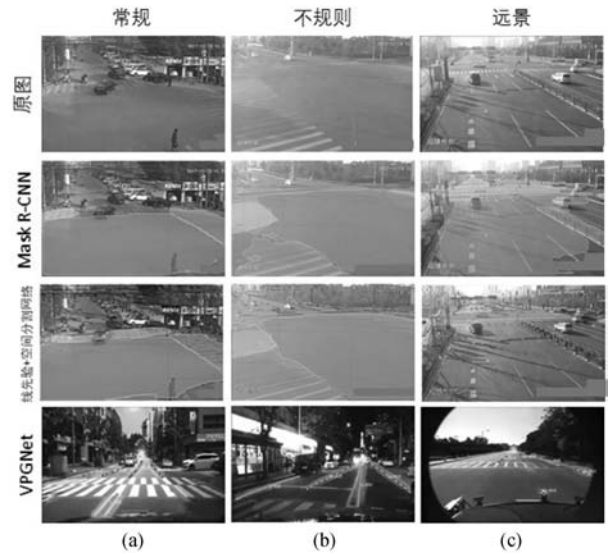


图7 不同类型图像分割结果

3.2.2 凸包损失函数实验结果

线先验空间分割网络虽然提升了分割和检测精确度,但是依然保留着 Mask R-CNN 的缺点。如在分析图6分割结果时所述,训练模型对分割边缘的处理不规整。另外,在图8中,如第一行图片所示,当斑马线分割边界的倾斜角度较大时,分割边界呈现非常不规则的锯齿状。经过分析得出原因,斑马线的外观特性是许多条平行白色线条,没有清晰的边界,在人眼识别或者标注数据集时,通常按照日常生活经验将斑马线的整个区域用一个凸多边形进行拟合。然而深度学习模型没有人类的生活常识,由于模型学习到的是斑马线的黑白相间条纹状特征,当对斑马线和路口区域边界进行划分时,模型产生了混淆,从而形成了不规则的锯齿状边缘。

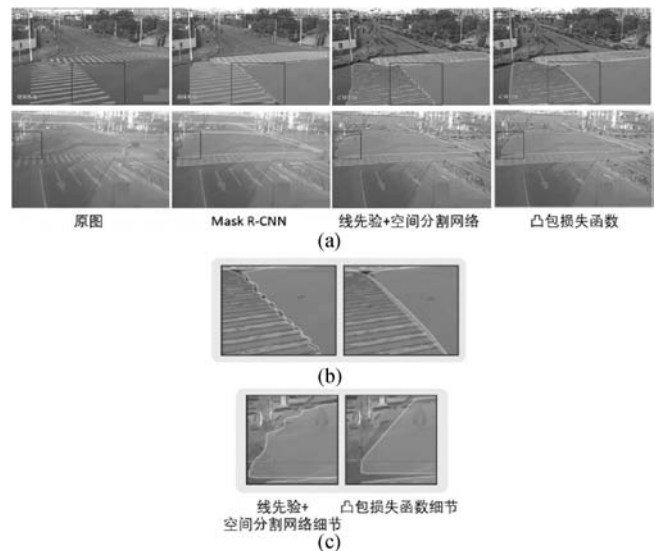


图8 凸包损失函数实验结果

为了解决边缘不规则这一问题,本文在 Mask R-CNN 框架中引入了凸包算法,在损失函数中加入了凸包二值交叉熵损失项。这一方法使模型在训练过程

中逐步学习用凸多边形进行分割的方法,改善了 Mask R-CNN 算法对于斑马线虚线类型的边界识别不清、边界划分不规则的问题。图 8 中,(b)、(c)分别为两组测试图片的细节放大对比图,展示了线先验空间分割网络和加入凸包损失函数后网络对分割边缘的处理效果。如图 8(b)所示,加入凸包损失函数后,网络在不牺牲分割精确度的情况下将锯齿状边缘规范化,形成易于表达和适用于后期路口车辆分析的分割边缘。如图 8(c)所示,模型将原来因模糊而分割不清的路口边界划分整齐,不仅达到了规整的分割效果也提升了分割精确度。

4 结 语

本文构建的包含 1 221 条斑马线和 138 块路口区域的交通路口监控图像区域分割数据集具有视角多样、识别困难等特点。针对斑马线和路口区域分割这一特定任务,线先验特征的加入为网络学习提供了明确的线特征。将 SCNN 网络结构与 Mask R-CNN 分割分支进行融合的空间分割网络让网络具有更好的空间学习能力,相较于其他的斑马线检测方法具有更高的准确度和鲁棒性。引入凸包二值交叉熵损失函数,使网络学习到使用凸多边形进行几何分割的方法。通过以上实验及分析可以发现,结合了线先验和凸包损失函数的空间分割网络与其他斑马线检测算法相比,在更困难的数据集中获得了更精确的斑马线分割结果,因而更适用于智能交通监管系统。然而监控视角的场景分割除了斑马线和路口区域外还有许多内容可以进行进一步细致研究,最终才能达到对于路口场景全面智能理解。

参 考 文 献

- [1] Mascetti S, Ahmetovic D, Gerino A, et al. ZebraRecognizer: Pedestrian crossing recognition for people with visual impairment or blindness [J]. Pattern Recognition, 2016, 60:405 - 419.
- [2] Zhai Y, Cui G, Gu Q, et al. Crosswalk detection based on MSER and ERANSAC [C]//IEEE International Conference on Intelligent Transportation Systems. IEEE, 2015.
- [3] Gioi R G V, Jakubowicz J, Morel J M, et al. LSD: A fast line segment detector with a false detection control [J]. IEEE Transactions on Software Engineering, 2010, 32(4): 722 - 732.
- [4] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN [C]//2017 IEEE International Conference on Computer Vision(ICCV), Venice, 2017:2980 - 2988.
- [5] 闫利,黄亮.街景影像中的斑马线检测与定位[J].测绘科学,2015,40(11):168 - 171.
- [6] Hu R H, Dollár P, He K M, et al. Learning to segment every thing[EB]. arXiv:1711.10370, 2017.
- [7] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 39(6):1137 - 1149.
- [8] Geiger A, Lenz P, Stiller C, et al. Vision meets robotics: The KITTI dataset[J]. International Journal of Robotics Research, 2013, 32(11):1231 - 1237.
- [9] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding[EB]. arXiv:1604.01685, 2016.
- [10] Ros G, Sellart L, Materzynska J, et al. The SYNTHIA dataset: A large collection of synthetic images for semantic segmentation of urban scenes [C]//Computer Vision & Pattern Recognition. IEEE, 2016.
- [11] Wu T, Ranganathan A. A practical system for road marking detection and recognition [C]//2012 IEEE Intelligent Vehicles Symposium. IEEE, 2012.
- [12] Pan X, Shi J, Luo P, et al. Spatial as deep: Spatial cnn for traffic scene understanding [C]//Thirty-Second AAAI Conference on Artificial Intelligence. 2018.
- [13] Greenhalgh J, Mirmehdi M. Automatic detection and recognition of symbols and text on the road surface [C]//International Conference on Pattern Recognition Applications and Methods. Springer, Cham, 2015: 124 - 140.
- [14] Bailo O, Lee S, Rameau F, et al. Robust road marking detection and recognition using density-based grouping and machine learning techniques [C]//2017 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2017: 760 - 768.
- [15] Lee S, Kim J, Yoon J S, et al. VPGNet: Vanishing point guided network for lane and road marking detection and recognition [C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 1947 - 1955.
- [16] Matas J, Chum O, Urban M, et al. Robust wide-baseline stereo from maximally stable extremal regions [J]. Image and vision computing, 2004, 22(10): 761 - 767.
- [17] Tan D. Image enhancement based on adaptive median filter and Wallis filter [C]//2015 4th National Conference on Electrical, Electronics and Computer Engineering. Atlantis Press, 2015.
- [18] Zhang Q, Shen X, Xu L, et al. Rolling guidance filter [C]//European conference on computer vision. Springer, Cham, 2014: 815 - 830.
- [19] 黄新,林倩.一种基于双极系数法的斑马线检测改良方法 [J]. 计算机应用与软件, 2017, 34(12): 202 - 205.