

文章编号:1007-2780(2022)04-0519-11

基于目标轮廓的实例分割方法

孙劲光, 王 雪*

(辽宁工程技术大学 电子与信息工程学院, 辽宁 葫芦岛 125105)

摘要:实例分割的目的是使用像素级实例掩码对单个对象进行分类和定位,是计算机视觉中一项具有挑战性的任务,目前存在分割速度慢、小物体分割精度低、分割边缘不平滑等问题。针对以上问题,提出了基于目标轮廓的实例分割方法,一方面,对物体轮廓上的结点进行特征提取,在不受检测框影响的同时避免了对目标物体内部像素点进行处理,以达到加快分割速度的效果;另一方面,对图像采取渐进式分割处理,多层次地提取物体轮廓上的特征,并配以多尺度融合的特征处理方法,更好地提取上下文语义信息,减少特征丢失。实验结果表明,本文算法在 Cityscapes 和 KINS 数据集中平均分割精度分别达到了 32.4%AP 和 32.0%AP,较目前许多优秀工作的分割精度都有所提升,在物体分割边缘的平滑程度以及贴合程度上均有更佳的处理效果。实验证明,基于目标轮廓的实例分割网络在实例分割任务中有较好的分割能力。

关键词:目标检测;实例分割;轮廓变形;圆形卷积;多尺度融合;渐进式分割

中图分类号:TP391.41 **文献标识码:**A **doi:**10.37188/CJLCD.2021-0236

Instance segmentation method based on target contour

SUN Jin-guang, WANG Xue*

(School of Electronic and Information Engineering, Liaoning Technical University,
Huludao 125105, China)

Abstract: The purpose of instance segmentation is to use pixel-level instance mask to classify and locate a single object, which is a challenging task in computer vision. At present, there are some problems such as slow segmentation speed, low segmentation accuracy for small objects, and uneven segmentation edge. Aiming at the above problems, a instance segmentation method based on target contour is proposed. On the one hand, the feature extraction is carried out for the nodes on the object contour, which is not affected by the detection box and avoids processing the internal pixels of the object, so as to accelerate the segmentation speed. On the other hand, the gradual segmentation of the image is adopted to extract the features on the contour of the object multi-level, and the multi-scale fusion feature processing method is used to better extract the context semantic information and reduce the feature loss. The average segmentation accuracy of Cityscapes and KINS is 32.4%AP and 32.0%AP, respectively, which is better than the segmentation accuracy of many excellent works. The smoothness and the degree of fit of the object segmentation edge have better processing effect. The instance segmentation network based on target contour has better

收稿日期:2021-09-09;修订日期:2021-11-02.

基金项目:国家重点研发计划(No. 2018YFB1403303)

Supported by National Key Research and Development Program of China(No. 2018YFB1403303)

*通信联系人, E-mail:1301629037@qq.com

segmentation ability in instance segmentation task.

Key words: object detection; instance segmentation; deformable contour; circle convolution; mulit-scale fusion; progressive segmentation

1 引言

实例分割是计算机视觉领域中的一个经典任务,是目标检测和语义分割两项任务的组合任务,通过目标检测定位图像中每个单独实例的位置,并对每个单独实例标记类别信息,区分同一类别的实例。随着深度学习的不断发展,许多实例分割框架被提出,并推陈出新,目前广泛应用于病灶检测、无人驾驶、虚拟现实、安防监控等领域,因此对实例分割精度和速度的要求也越来越高。

实例分割最早是 Hariharan^[1]等人受到 R-CNN 目标检测的启发,于 2014 年提出了 SDS 模型,可以同时完成检测与分割任务。2017 年, Li 等人提出 FCIS^[2],是首个实现端到端训练的图像实例分割算法。随着深度学习的不断发展,2017 年, He 等人提出了 Mask R-CNN^[3]模型,在 Faster R-CNN^[4]的基础上添加了掩码分支来预测分割掩膜,并设计 RoIAlign 层,消除 RoI Pooling 的量化误差,采用双线性插值的方法计算每个感兴趣区域采样点输入特征值,保证提取特征与输入对齐,得到了惊人的分割效果,为后续的图像分割工作奠定了好的基础。2019 年, Huang 等人针对 Mask R-CNN 模型中以分类分支置信度作为掩膜质量评价标准的问题,提出了 Mask Scoring R-CNN^[5]模型,采用预测掩膜与标注掩膜的交并比 MaskIoU 来描述掩膜的分割质量,并引入新的分支 MaskIoU Head,将 MakiIoU 与分类的分数相乘计算掩膜分数,这样的计算方式校准了掩码质量和掩码得分之间的偏差,提升了分割性能。对于 Mask R-CNN 的改进, Liu 等人还提出 PANnet^[6],引入了一种自适应特征池化,并设计了一种可以将不同层级特征融合的信息融合路径。以上的工作在基于检测框的实例分割领域取得了十分优秀的效果,但同时也存在两个问题:一是对于检测框检测不准确的物体,很难在后续得到精准的分割结果;二是基于检测框进行分割时,要对物体进行逐像素计算,将导致计算

量大、速度慢的问题。

不同于基于检测框的实例分割方法,无锚框的实例分割方法是沿着物体的外围轮廓进行划分,解决了物体检测框局限性的问题。2018 年, Liu 等人提出 GMIS^[7]模型,这是一种自底向上像素聚类的方法,用独立的两个子网络分别产生语义分割结合和像素亲和性信息,并利用语义分割的先验信息,得到每一类别所有个体的像素区域。2020 年, Wang 等人提出 CenterMask^[8]模型,不依赖预先设定的感兴趣区域(ROI)进行掩码预测,将分割分成两个子任务:局部形状预测,从每个对象的中心点表示预测的粗略的形状,以约束每个对象的局部区域;全局显著性映射,预测整个图像的显著性图,以实现精确分割,将两分支的输出相乘计算构造出每个实例的掩码。对于构建物体轮廓的实例分割方法, Liu 等人于 2017 年提出 SGN^[9]模型,构建了序列组合网络,采用由点到线再到区域的聚合方式,把像素聚合为线段,再对相邻的线段进行判断,进而得到一个连通区域,连通区域的组合结果即为实例分割的掩膜。2020 年, Xie 等人提出 PolarMask^[10]模型,一种全卷积、无锚框的实例分割方法,以 FCOS^[11]目标检测网络为基础,从 4 根射线扩成 36 根射线,以极坐标的方式对物体轮廓进行建模,不需要检测框,把实例分割问题转化为实例中心点分类问题和密集距离回归问题,虽然在精度和速度上优势并不大,但对基于无锚框实例分割方法的研究有着很大的意义。2020 年, Peng 等人提出 Deep Snake^[12]模型,以主动轮廓模型 Snake^[13-15]为基础,采用深度学习的方式替代人为能量函数,获得更加逼近目标物体的轮廓,从而达到高精度的实例分割,且加快了计算速度。但是现有方法仍然存在小物体分割精度低,分割边缘不平滑的问题。

针对现有方法的不足,本文提出了基于目标轮廓的实例分割网络,采取渐进式的分割模型,内部设计多尺度融合模块进行特征融合,以迭代的方式多次将分割模块输出的偏移量进行计算,得到新的目标物体轮廓并作为输入,以求得到最

优的分割结果,并配以一定的标准化方法减少目标物体对轮廓变形的影响。本文方法相较于 Deep Snake 方法以及其他优秀工作,在 Cityscapes 数据集和 KINS 数据集上,分割精度都有所提升。

2 相关工作

2.1 传统 Snake 算法

1987 年 Michael Kass 等人提出了 Snake 算法,可以用于图像实例分割。Snake 是一种主动轮廓模型,即预先给定一个初始轮廓,以初始轮廓为基础进行逐步迭代,将得到的结果多次作为输入,得到更加贴合图像边缘的轮廓,从而完成实例分割任务。Snake 算法将分割问题转化为能量函数最小化的数学问题,利用首尾相连的闭合曲线的函数规律,设计能量函数控制曲线的形变,最终目标是得到最小化能量函数使闭合曲线充分贴合分割目标的边缘,以此达到精确的物体分割结果。其中定义的能量函数包括轮廓能量和图像能量,具体公式如公式(1)所示:

$$E_{\text{snake}}^* = \int_0^1 E_{\text{snake}}(v(s)) ds = \int_0^1 E_{\text{int}}(v(s)) + E_{\text{image}}(v(s)) + E_{\text{con}}(v(s)) ds, \quad (1)$$

其中: $E_{\text{int}}(v(s))$ 是轮廓能量,即轮廓本身的能量也称为内部能量; $E_{\text{image}}(v(s))$ 为图像能量,即图像上轮廓对应点的能量,也称为外部能量; $E_{\text{con}}(v(s))$ 是方差相关项。

Snake 算法已经被应用在很多模型设计中,但这种人为设计的函数很容易得到局部最优解,对于边缘不平整的目标,想要轮廓和目标物体达到理想的贴合度是不可能的,从而很难得到理想的分割效果。

2.2 目标检测

目标检测的主要任务是定位出图像中实例个体的位置,是实例分割必不可少的一部分。CenterNet^[16]是在 CornerNet^[17]的基础上进行改进,提出的一种无锚框的目标检测方法。该方法只需将图像输入全卷积网络,得到一个热力图,热力图峰值即目标物体的中心点,且通过峰值点预测目标物体的宽和高。CenterNet 提出了 3 种 backbone 的网络结构,分别是 Resnet-50、DLA-34

和 Hourglass-104。

本文采用基于 DLA-34 的 Centernet 目标检测网络,该网络无锚框的检测方式更适用于针对物体轮廓的实例分割方法。其中 DLA-34 网络是通过多级的跳跃连接,迭代地将网络结构的特征信息融合起来,让模型有更高的精度和更少的参数,可以加快网络速度。

2.3 圆形卷积

圆形卷积是一维卷积的一种特殊卷积形式,多数应用于离散信号。本文方法针对目标物体轮廓上的特征点进行特征提取,由于组成物体轮廓的点之间的关系是离散的,相较于图像处理中常用的二维卷积,圆形卷积更加适合这种离散关系的特征计算。

圆形卷积的计算方式如图 1 所示,其中在下面圆形轮廓上的结点为轮廓上的输入特征,内部相连的结点为内核函数,在上面圆形轮廓上的结点是通过卷积后输出的特征。图中表示了输出特征由输入特征和内核函数计算所得的过程,计算与标准卷积类似,圆形卷积输出特征与输入特征长度相同。

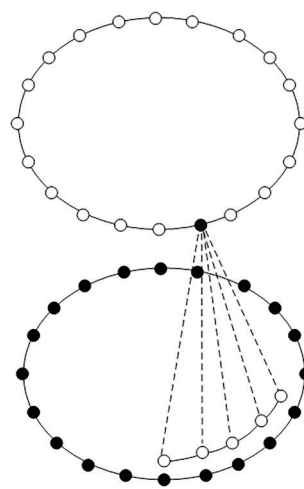


图1 圆形卷积

Fig. 1 Circle convolution

将轮廓顶点视为一个离散的一维信号, $f: Z \rightarrow R^D$, 将它拓展为一组周期信号,轮廓特征的定义如公式(2)所示:

$$(f_N)_i \triangleq \sum_{j=-\infty}^{\infty} f_{i-j} N. \quad (2)$$

采用圆形卷积计算的周期特征,具体计算如

公式(3)所示:

$$(f_N * k)_i = \sum_{j=-r}^r (f_N)_{i+j} k_j. \tag{3}$$

3 本文方法

针对小物体分割不准确,边缘不清晰问题,本文提出了基于目标轮廓的实例分割网络(Instance segmentation network based on target con-

tour points, TCPN),具体流程如图 2 所示,主要由目标检测、构建初始轮廓、计算偏移量 3 个任务完成整个分割工作,通过 3 次迭代的轮廓变形得到最终分割结果。在 Deep Snake 的基础上构建渐进式分割网络,对目标物体的边缘分割得到更加准确的效果。构建多尺度特征融合模块,对渐进式分割网络中经过粗割模块得到的特征经过不同的空洞卷积,融合多个尺度的特征,减少浅层特征的丢失。

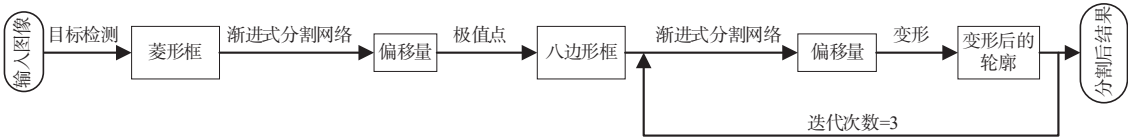


图 2 基于目标轮廓的实例分割网络流程图

Fig. 2 Flow chart of instance segmentation network based on target contour

3.1 渐进式分割网络

Deep Snake 方法是采用深度学习的方式实现传统 Snake 算法,以构建网络模型的方式代替能量函数。通过对 Deep Snake 实验结果的观察,发现对于自行车、行人这类边缘不平滑的物体,分割结果很难逼近物体边缘,从而降低了分割精度。针对这个问题,本文方法对 Deep Snake 算法

做了优化改进,构建了渐进式分割网络,如图 3 所示。通过两阶段的分割模块,由浅入深,以此使分割结果更好地逼近物体轮廓。通过渐进式分割网络对轮廓上结点的特征进行学习,得到指向目标物体轮廓的偏移量,通过得到的偏移量变形输入的轮廓,经过多次迭代,得到最终的分割结果。

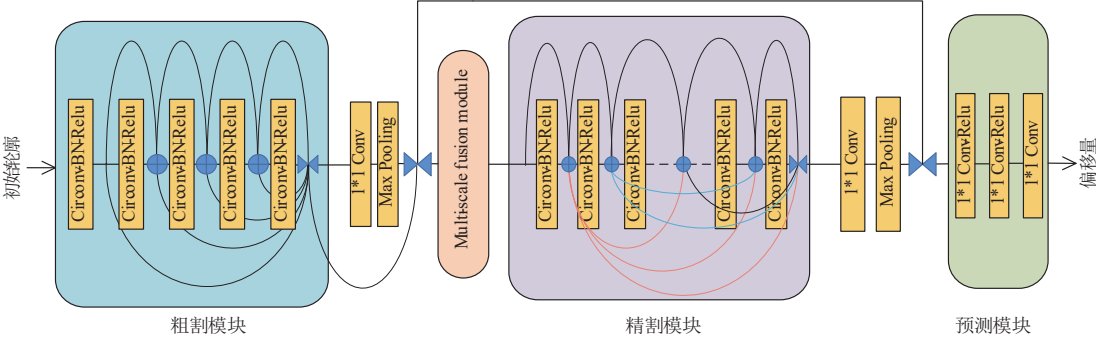


图 3 渐进式分割网络

Fig. 3 Progressive segmentation network

渐进式分割网络包括粗割模块、多尺度融合模块(Multi-scale-fusion module)、精割模块和预测模块。粗割模块和精割模块内部都是由相同的 Circonv-BN-Relu 层组成,包括圆形卷积、批量标准化和激活函数,但数量和连接方式不一样,由此达到一种渐进的效果。因为是对物体轮廓上的特征进行计算,而轮廓上的结点首尾相连,

度为 2,更类似于离散的信号,所以采用圆形卷积进行计算。粗割模块内部采用残差^[18-20]的连接方式,连接 5 个 Circonv-BN-Relu 层。精割模块内部采用类似于 DenseNet^[21-22]的连接方式,连接 8 个 Circonv-BN-Relu 层,通过这种连接方式可以更好地保留浅层特征,相较于粗割模型可以更好地提取轮廓上的特征值。

通过粗割模型对轮廓特征进行特征提取并将所有层的特征连接,通过 1×1 卷积层和最大池化层将融合的特征和每个顶点的特征连接,以此得到一个初步轮廓特征,并保留与下一步的精割结果连接。将结果通过多尺度融合模块后,输入精割模块,计算过程与粗割模型相同,加深了网络并更大程度地保留了浅层信息。最后将粗割结果和精割结果连接并输入预测模块,对轮廓特征应用了3个 1×1 卷积层,得到最终的偏移。通过对偏移量和初始轮廓的坐标位置进行计算,得到目标物体轮廓,以此作为初始轮廓输入分割模块,完成迭代过程。

3.2 多尺度融合模块

多尺度融合模块如图4所示,由多个圆形卷积组成,为了得到多尺度的上下文信息,本文方法中采用的是卷积率为1、3、5的空洞卷积分别提取特征,扩大感受野,防止经过粗割模块后大物体的特征丢失,将得到的不同感受野的特征进行拼接,通过 1×1 卷积进行融合,再通过 1×1 卷积输出作为精割模块的输入。将多尺度融合模块的设计加入渐进式分割网络,作为粗割模块后的特

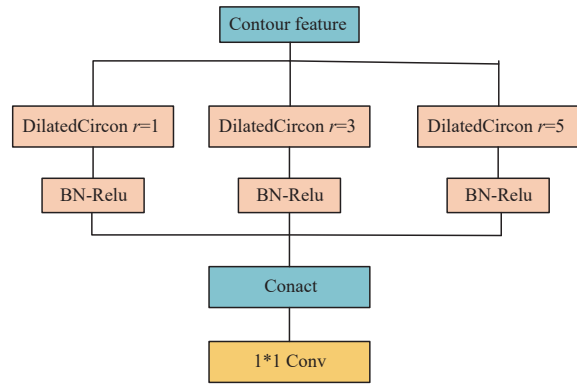


图4 多尺度融合模块

Fig. 4 Multi-scale fusion module

征融合和精割模块前的特征处理,是出于实验数据考虑,将在实验部分具体介绍。

3.3 初始轮廓

如图5所示,基于目标轮廓的实例分割过程通过CenterNet目标检测网络,得到目标物体的中心点以及宽和高,从而得到检测框。取矩形检测框4个边的中心点,记为 $\{P_i^p | i=1, 2, 3, 4\}$,将这4个点连接成菱形,作为初始轮廓,输入渐进式分割网络,得到4个偏移量,使 P_i^p 逼近 P_i^o 。

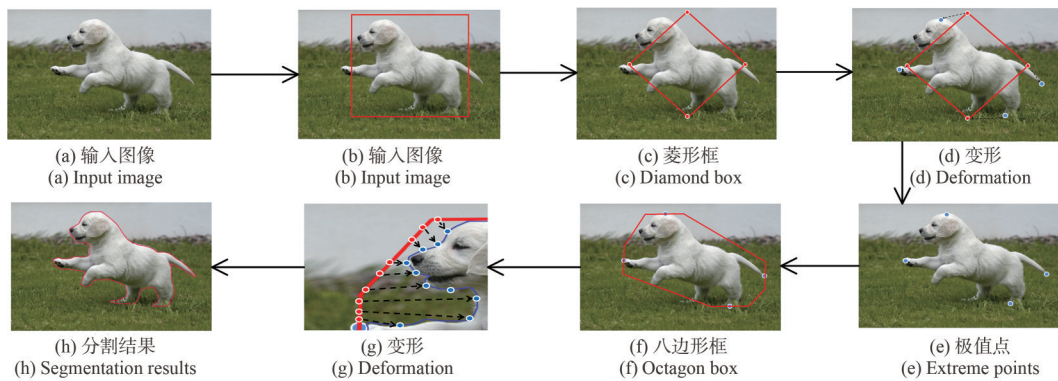


图5 分割过程

Fig. 5 Segmentation process

通过上一步得到的 P_i^p 定义一个新的边界框,以 P_i^p 点沿着边界框延伸方向向两侧画一条线段,长度为一条边的 $1/4$,连接4个线段的端点形成八边形。以这个八边形为输入,从 P_i^p 开始沿着八边形轮廓均匀采样 N 个点,记为 $\{P_i | i=1, 2, \dots, N, N=128\}$,且沿着目标物体轮廓均匀采样 N 个点,通过渐进式分割网络的计算,输出 N 个偏移值,使初始轮廓向目标物体轮廓偏移。采

用三次迭代的方式,使得图像中目标物体无论远近,都可以得到很好的分割结果。

在Deep Snake算法中,为了使轮廓变形不受图像中轮廓平移的影响,取轮廓上所有顶点的值减去最小的顶点值。本文为了更大程度地减少这种影响,在迭代过程中采用极差标准化,具体计算如下。

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad (4)$$

$$y' = \frac{y - y_{\min}}{y_{\max} - y_{\min}}. \quad (5)$$

使输入渐进式分割网络的初始轮廓在 $[0, 1]$ 的有界范围内,相应地可以使用激活函数 $\tanh$ 使偏移量也缩小到固定范围,输出再对应宽高比例放大,这样可以最大程度地减少对轮廓变形的影响。

3.4 损失函数

针对渐进式分割方法采用 Smooth L_1 方法设置两个损失函数约束轮廓偏移,一个是取极值点 P_i^{ep} 时的损失函数,记为

$$L_{\text{ep}} = \frac{1}{4} \sum_{i=1}^4 \text{smooth}_{L_1}(\tilde{p}_i^{\text{ep}} - p_i^{\text{ep}}), \quad (6)$$

$$\tilde{p}_i^{\text{ep}} = \begin{cases} \tilde{x}_i^{\text{ep}}/w \\ \tilde{y}_i^{\text{ep}}/h \end{cases}, \quad (7)$$

$$p_i^{\text{ep}} = \begin{cases} x_i^{\text{ep}}/w \\ y_i^{\text{ep}}/h \end{cases}, \quad (8)$$

这里 $\tilde{p}_i^{\text{ep}}$ 代表网络中所预测的极值点,另一个损

失函数是迭代变形轮廓时的损失函数,记为

$$L_{\text{iter}} = \frac{1}{N} \sum_{i=1}^N \text{smooth}_{L_1}(p_i, p_i^{\text{gt}}), \quad (9)$$

$$p_i = \begin{cases} x_i/w \\ y_i/h \end{cases}, \quad (10)$$

$$p_i^{\text{gt}} = \begin{cases} x_i^{\text{gt}}/w \\ y_i^{\text{gt}}/h \end{cases}, \quad (11)$$

损失函数对应轮廓变形时的极差标准化操作,对物体坐标做等比例变换处理,这里 p_i^{gt} 代表轮廓上标注的真实值。

3.5 二次检测

针对 Cityscapes 数据集中被遮挡而分为多部分的物体,由于数据集标注时是分别对多部分物体进行标注,若采用本文方法直接对整个物体进行实例分割,则会降低精度,因此采用二次检测的方式对检测框内被分为多部分的物体进行再次检测,这也是 Deep Snake 中所采取的处理方式,具体方式如图 6 所示。

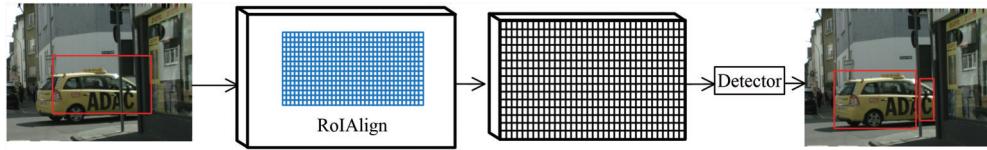


图 6 二次检测网络

Fig. 6 Secondary detection network

针对检测框,使用 RoIAlign^[2]方法来提取特征图,并在特征图上添加一个检测器分支对被遮挡的目标物体生成各部件的检测框;对于检测到的各部件,我们使用渐进式分割方法分别对它们进行分割,通过合并的方式得到最终的分割结果。

4 实验结果及分析

4.1 实验数据集

实例分割旨在解决不同实例个体像素分割的问题,为了提高网络对复杂场景的理解能力,需要高清晰度、数量规模庞大的数据库作为支撑。本文采取了以下两个公开的大型数据集。

Cityscapes 数据集^[23]: Cityscapes 数据集是关于城市街道场景的语义理解图片数据集。它主要包含来自 50 个不同城市的街道场景,拥有 5 000 张在城市环境中驾驶场景的高质量像素级

注释图像;此外,它有 20 000 张粗糙标注的图像。该数据集按照与城市场景的相关性,将 30 个目标类别分为 8 类数据集,最终的评价标准是根据数据集中 8 个语义类的平均精度来评估。

KITTI 数据集是目前国际上最大的自动驾驶场景下的计算机视觉算法评测数据集,包含市区、乡村和高速公路等场景采集的真实图像数据集,每张图像中最多达 15 辆车和 30 个行人,还有各种程度的遮挡和截断。KINS 数据集^[24]是对 KITTI 数据集进行实例级语义标注得到的一组数据。该数据集目的是在遮挡下恢复完整的实例形状,包括 7 474 张训练图像和 7 517 张测试图像。根据它的设置,最终的评价标准是根据数据集中 7 个对象类别的平均精度来评估的。

4.2 评价指标

为了验证本文所提出的基于目标轮廓的实例分割网络在实例分割任务中的有效性,采用平

均精度(Average Precision, AP)作为评价指标。平均精度是在图像分割情况下,将每个结果与正确标注结果计算IoU,且IoU从0.5~0.95之间以0.05为标准进行十等分,计算这10个不同IoU阈值下交并比的平均值,AP₅₀是指IoU阈值取0.5时的计算结果。

4.3 实验设置

本文的实验环境是Python3.7.0,操作系统为Ubuntu 18.04,主要基于开源的机器学习框架Pytorch1.3.1-gpu、CUDA和CuDNN来实现GPU加速。本文使用Cityscapes数据集和KINS数据集进行训练,在Cityscapes数据集上设计消融实验和对比实验,消融实验通过测试多种扩张率的扩张卷积,验证多尺度融合模块的有效性,并得到效果最佳的一组扩张率;对比实验通过和目前优秀的实例分割工作进行对比,验证本文方法的有效性;且通过KINS数据集测试迭代次数

对实例分割结果的影响,并验证本文方法对存在遮挡部分的图像进行分割的有效性。由于两种实验数据集的标注情况不同,所要实现的实验结果不同,所以采取不同的模型文件进行训练、验证和测试。

4.4 实验效果

在实验过程中,对于八边形作为初始轮廓时的采样点个数,本文设置 $N=128$ 。实验中也尝试选取过 $N=256$,不仅会降低分割精度,且大幅增加了运行时间。

4.4.1 迭代次数测试

在KINS数据集下进行多次测试,确定最终TCPN网络模型中渐进式分割网络的迭代次数,图7显示不同迭代次数下的实验效果,经过3次迭代时实验效果最佳。通过表1的测试数据也不难看出,迭代次数为3时实例分割精度最高。

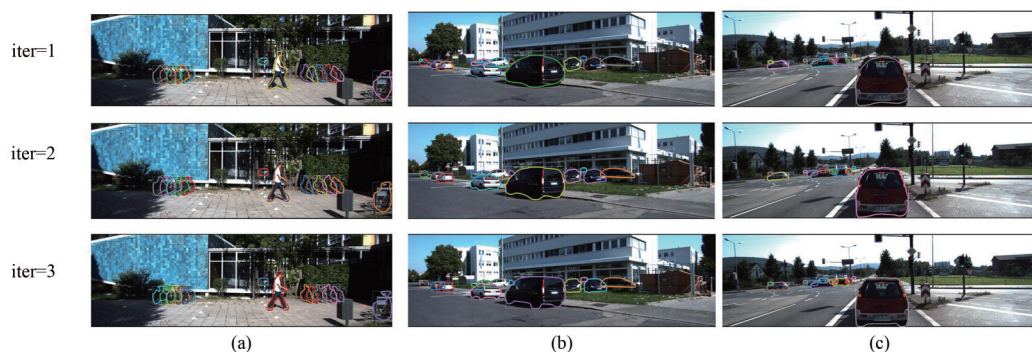


图7 不同迭代次数的实验效果对比。(a)自行车和行人;(b)轿车;(c)轿车和较远物体。

Fig. 7 Comparison of experimental effects with different iterations. (a) Bicycle and person; (b) Cars; (c) Cars and distance objects.

表1 不同迭代次数下的实验效果

Tab. 1 Experimental results of different iterations

迭代次数	AP
1	27.5
2	31.7
3	32.0
4	31.2
5	30.6

4.4.2 二次检测模块的表现

由于本文方法是通过闭合曲线进行实例分割,在Cityscapes中进行实验时,对于被遮挡而一分为多的数据情况,只能针对其中一部分。通过

对物体进行二次检测,识别其中的多部分,从而对每部分物体分别进行实例分割,得到最终结果,效果如图8所示。

4.4.3 多尺度融合模块的对比实验

多尺度融合模块的设计可以在粗割模块和精割模块中起到衔接作用,增强上下文语义信

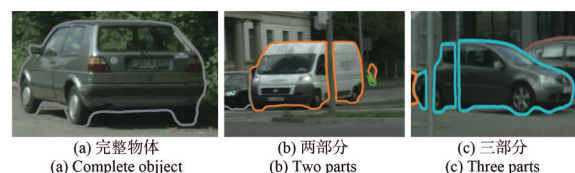


图8 二次检测效果图

Fig. 8 Experimental rendering of secondary detection

息,扩大感受野,从而得到更加精确的分割结果。对于多尺度融合模块的有效性,本文做了对比实验,表2的实验结果分别是不加多尺度融合模块,和加不同扩张率的多尺度融合模块的最终平均分割精度,表中−Multi-scale fusion module 代表不加多尺度融合模块的实验结果, +Multi-scale fusion module_1 代表加扩张

率为 2、4、8 的多尺度融合模块的实验结果, +Multi-scale fusion module_2 代表加扩张率为 1、3、5 的多尺度融合模块的实验结果。通过表2中的数据可以发现,添加多尺度融合模块有助于提升分割精度,尤其是对图像中大物体的分割精度,减少了精度损失,且扩张率为 1、3、5 时分割精度最高。

表 2 多尺度融合模块实验结果
Tab. 2 Experimental results of multi-scale fusion module

Module	AP	person	rider	car	truck	bus	train	mcycle	bicycle
−Multi-scale fusion module	31.4	39.0	27.8	57.5	28.8	40.3	20.4	19.6	17.8
+Multi-scale fusion module_1	31.8	39.7	28.9	57.9	28.5	38.7	22.9	19.5	18.1
+Multi-scale fusion module_2	32.4	39.6	29.0	57.8	28.5	40.4	24.8	20.8	18.5

4. 4. 4 在 Cityscapes 数据集上的表现

在 Cityscapes 数据集上采用分段式训练方式,先对检测网络 CenterNet 进行训练,再将训练好的结果与渐进式分割网络共同训练。实验中设置学习率 $lr=9e-5$, $batch-size=4$, 经过训练,最终的测试结果如图9所示。从图9中的实验效果可以看出,基于目标轮廓的实例分割网络可以很好地对图片中的物体进行实例分割,如图中的小汽车分割出的物体轮廓很好地贴合了物

体边缘,对于较远、较小的物体也都很好地给出了物体轮廓。通过表3和表4中的实验数据也不难发现,本文方法相较于 Mask R-CNN 经典的基于检测框的实例分割方法,无论在整体实例分割精度还是部分物体实例分割精度都有明显优势;相较于 PANet 自底向上“双塔战术”的方法,在平均精度以及部分实例物体的分割精度都有所提升;对比 Deep Snake 方法,平均精度也提升了 0.7%AP 和 1.5%AP₅₀。

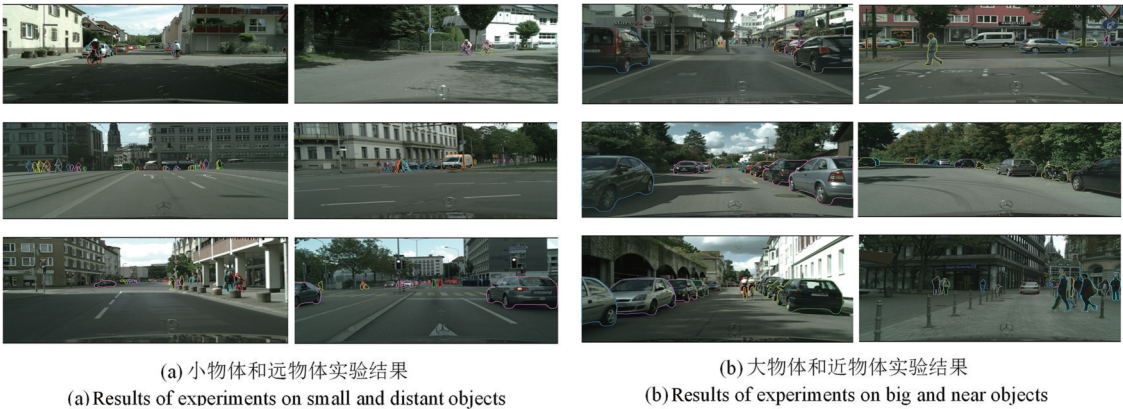


图 9 Cityscapes 数据集实验效果
Fig. 9 Experimental effect of Cityscapes dataset

4. 4. 5 在 KINS 数据集上的表现

将检测网络 CenterNet 与渐进式分割网络共同训练。实验中设置学习率 $lr=1e-4$, $batch-size=4$, 经过训练后,得到最终的测试结果如图10所示。从图10中的实验效果可以看出,图像中即使部分物体被遮挡,也可以很好地满足数据

集要求得到正确的分割结果,且对于图中小汽车、自行车和行人的分割轮廓贴合目标物体,光滑无突出棱角。通过表5中的实验结果可以看出,在 KINS 数据集中,本文方法较目前其他优秀工作在平均分割精度上有所提高,对比 Deep Snake 方法提升了 0.7%AP。

表 3 Cityscapes数据集实验结果 (AP)

Tab. 3 Results from the Cityscapes dataset(AP)

Algorithm	AP	person	rider	car	truck	bus	train	mcycle	bicycle
SGN ^[9]	25.0	21.8	20.1	39.4	24.8	33.2	30.8	17.7	12.4
PolygonRNN++ ^[25]	25.5	29.4	21.8	48.3	21.1	32.3	23.7	13.6	13.6
Mask R-CNN ^[3]	26.2	30.5	23.7	46.9	22.8	32.3	18.6	19.1	16.0
GMIS ^[7]	27.6	29.3	24.1	42.7	25.4	37.2	32.9	17.6	11.9
Spatial ^[26]	27.6	34.5	26.1	52.4	21.7	31.2	16.4	20.1	18.9
PANET ^[5]	31.8	36.8	30.4	54.8	27.0	36.3	25.5	22.6	20.8
Deep Snake ^[12]	31.7	37.2	27.0	56.0	29.5	40.5	28.2	19.0	16.4
TCPN	32.4	39.6	29.0	57.8	28.5	40.4	24.8	20.8	18.5

表 4 Cityscapes数据集实验结果 (AP₅₀)

Tab. 4 Results from the Cityscapes dataset(AP₅₀)

Algorithm	AP ₅₀	person	rider	car	truck	bus	train	mcycle	bicycle
Deep Snake ^[12]	58.3	72.2	64.3	81.5	41.2	58.8	51.3	50.4	46.6
TCPN	59.8	74.8	67.3	84.0	40.5	60.1	46.3	52.7	52.4



图 10 KINS数据集实验效果

Fig. 10 Experimental effect of KINS dataset

表 5 KINS数据集实验结果

Tab. 5 Results from the KINS dataset

Algorithm	amodal seg
MNS ^[27]	18.5
FCIS ^[2]	23.5
ORCNN ^[28]	29.0
Mask R-CNN ^[3]	29.2
PANET ^[5]	30.4
Deep snake ^[12]	31.3
TCPN	32.0

5 结 论

本文针对实例分割中存在小物体分割慢、分割精度低、分割边缘不平滑等,提出了基本目标轮廓的实例分割方法。通过提取物体轮廓上的结点,可加快分割速度;通过渐进式分割网络,多层次提取轮廓特征,再进行多尺度融合,利用扩张的圆形卷积增大感觉野,丰富了特征细节;通过极差标准化降低轮廓对轮廓回归变形的影响。

算法在实验数据集 Cityscapes 和 KINS 中的分割精度达到了 32.4% 和 32.0%, 分割边缘的平滑程度与贴合程度更佳, 实验结果验证了算法的有效性和正确性。

参 考 文 献:

- [1] HARIHARAN B, ARBELÁEZ P, GIRSHICK R, *et al.* Simultaneous detection and segmentation [C]//*13th European Conference on Computer Vision*. Zurich, Switzerland: Springer, 2014: 297-312.
- [2] LI Y, QI H Z, DAI J F, *et al.* Fully convolutional instance-aware semantic segmentation [C]//*Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA: IEEE, 2017: 4438-4446.
- [3] HE K M, GKIOXARI G, DOLLÁR P, *et al.* Mask R-CNN [C]//*2017 IEEE International Conference on Computer Vision (ICCV)*. Venice, Italy: IEEE, 2017: 2980-2988.
- [4] REN S Q, HE K M, GIRSHICK R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [5] HUANG Z J, HUANG L C, GONG Y C, *et al.* Mask scoring R-CNN [C]//*Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA: IEEE, 2019: 6402-6411.
- [6] LIU S, QI L, QIN H F, *et al.* Path aggregation network for instance segmentation [C]//*Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Salt Lake City, UT, USA: IEEE, 2018: 8759-8768.
- [7] LIU Y D, YANG S Y, LI B, *et al.* Affinity derivation and graph merge for instance segmentation [C]//*Proceedings of the 15th European Conference on Computer Vision (ECCV)*. Munich, Germany: Springer, 2018: 708-724.
- [8] LEE Y, PARK J. CenterMask: Real-time anchor-free instance segmentation [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE, 2020: 13903-13912.
- [9] LIU S, JIA J Y, FIDLER S, *et al.* SGN: Sequential grouping networks for instance segmentation [C]//*Proceedings of the 2017 IEEE International Conference on Computer Vision*. Venice, Italy: IEEE, 2017: 3516-3524.
- [10] XIE E, SUN P Z, SONG X G, *et al.* Polarmask: Single shot instance segmentation with polar representation [C]//*Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE, 2020: 12190-12199.
- [11] TIAN Z, SHEN C H, CHEN H, *et al.* FCOS: Fully convolutional one-stage object detection [C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South): IEEE, 2019: 9626-9635.
- [12] PENG S D, JIANG W, PI H J, *et al.* Deep snake for real-time instance segmentation [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE, 2020: 8530-8539.
- [13] KASS M, WITKIN A, TERZOPOULOS D. Snakes: Active contour models [J]. *International Journal of Computer Vision*, 1988, 1(4): 321-331.
- [14] 黄鑫. 基于四元数的彩色图像边缘检测研究[J]. *液晶与显示*, 2014, 29(3): 466-472.
HUANG X. Color image edge detection based on quaternion [J]. *Chinese Journal of Liquid Crystals and Displays*, 2014, 29(3): 466-472. (in Chinese)
- [15] 符喜优, 张风丽, 王国军, 等. 结合张量投票和 Snakes 模型的 SAR 图像道路提取[J]. *中国图象图形学报*, 2015, 20(10): 1403-1411.
FU X Y, ZHANG F L, WANG G J, *et al.* Road extraction from SAR images using tensor voting and Snakes model [J]. *Journal of Image and Graphics*, 2015, 20(10): 1403-1411. (in Chinese)
- [16] DUAN K W, BAI S, XIE L X, *et al.* Centernet: Keypoint triplets for object detection [C]//*Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South): IEEE, 2019: 6568-6577.
- [17] LAW H, DENG J. CornerNet: Detecting objects as paired keypoints [C]//*Proceedings of the 15th European Con-*

- ference on Computer Vision (ECCV). Munich, Germany: Springer, 2018: 734-750.
- [18] He K M, ZHANG X Y, REN S Q, *et al.* Deep residual learning for image recognition [C]//*Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, 2016: 770-778.
- [19] 徐胜军, 欧阳朴衍, 郭学源, 等. 多尺度特征融合空洞卷积 ResNet 遥感图像建筑物分割[J]. 光学精密工程, 2020, 28(7): 1588-1599.
- XU S J, OUYANG P Y, GUO X Y, *et al.* Building segmentation in remote sensing image based on multiscale-feature fusion dilated convolution ResNet [J]. *Optics and Precision Engineering*, 2020, 28(7): 1588-1599. (in Chinese)
- [20] 廖明哲, 吴谨, 朱磊. 基于 ResNet 和 RF-Net 的遥感影像匹配[J]. 液晶与显示, 2020, 35(9): 972-980.
- LIAO M Z, WU J, ZHU L. Remote sensing images matching based on ResNet and RF-Net [J]. *Chinese Journal of Liquid Crystals and Displays*, 2020, 35(9): 972-980. (in Chinese)
- [21] HUANG G, LIU Z, MAATEN L VAN DER, *et al.* Densely connected convolutional networks [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA: IEEE, 2017: 2261-2269.
- [22] 周涛, 霍兵强, 陆惠玲, 等. 融合多尺度图像的密集神经网络肺部肿瘤识别算法[J]. 光学精密工程, 2021, 29(7): 1695-1708.
- ZHOU T, HUO B Q, LU H L, *et al.* Lung tumor image recognition algorithm with densenet fusion multi-scale images [J]. *Optics and Precision Engineering*, 2021, 29(7): 1695-1708.
- [23] CORDTS M, OMRAN M, RAMOS S, *et al.* The cityscapes dataset for semantic urban scene understanding [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, 2016: 3213-3223.
- [24] QI L, JIANG L, LIU S, *et al.* Amodal instance segmentation with KINS dataset [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA: IEEE, 2019: 3009-3018.
- [25] ACUNA D, LING H, KAR A, *et al.* Efficient interactive annotation of segmentation datasets with polygon-RNN++ [C]//*Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Salt Lake City, UT, USA: IEEE, 2018: 859-868.
- [26] NEVEN D, BRABANDERE B D, PROESMANS M, *et al.* Instance segmentation by jointly optimizing spatial embeddings and clustering bandwidth [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA: IEEE, 2019: 8829-8837.
- [27] DAI J F, HE K M, SUN J. Instance-aware semantic segmentation via multi-task network cascades [C]//*Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, 2016: 3150-3158.
- [28] FOLLMANN P, KÖNIG R, HÄRTINGER P, *et al.* Learning to see the invisible: End-to-end trainable amodal instance segmentation [C]//*2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. Waikoloa, HI, USA: IEEE, 2019: 1328-1336.

作者简介:



孙劲光(1962—),女,辽宁阜新人,博士,教授,2006年于辽宁工程技术大学获得博士学位,主要从事计算机图像处理、计算机图形学、知识工程方面的研究。E-mail: sunjinguang@lntu.edu.cn



王雪(1996—),女,辽宁阜新人,硕士研究生,2019年于辽宁工程技术大学获得学士学位,主要从事计算机图像处理中图像分割、目标检测方面的研究。E-mail: 1301629037@qq.com