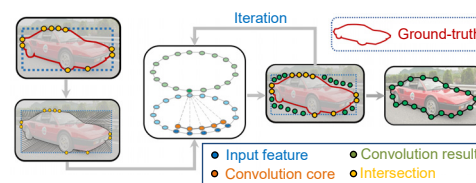


DOI: 10.12086/oe.2021.210245

实时实例分割的深度轮廓段落匹配算法

曹春林¹, 陶重霖^{1,2*}, 李华一¹, 高涵文¹¹苏州科技大学电子与信息工程学院, 江苏 苏州 215009;²清华大学苏州汽车研究院, 江苏 苏州 215134

摘要: 针对实例分割算法在进行轮廓收敛时, 普遍存在目标遮挡增加轮廓处理的时间以及降低检测框的准确性的问题。本文提出一种实时实例分割的算法, 在处理轮廓中增加段落匹配、目标聚合损失函数和边界系数模块。首先对初始轮廓进行分段处理, 在每一个段落内进行分配局部地面真值点, 实现更自然、快捷和平滑的变形路径。其次利用目标聚合损失函数和边界系数模块对存在目标遮挡的物体进行预测, 给出准确的检测框。最后利用循环卷积与 Snake 模型对匹配过的轮廓进行收敛, 对顶点进行迭代计算得到分割结果。本文算法在 COCO、Cityscapes、Kins 等多个数据集上进行评估, 其中 COCO 数据集上取得 32.6% mAP 和 36.3 f/s 的结果, 在精度与速度上取得最佳平衡。

关键词: 实例分割; 目标检测; Snake 模型; 目标遮挡; 初始轮廓

中图分类号: TP391.4

文献标志码: A

曹春林, 陶重霖, 李华一, 等. 实时实例分割的深度轮廓段落匹配算法[J]. 光电工程, 2021, 48(11): 210245

Cao C L, Tao C B, Li H Y, et al. Deep contour fragment matching algorithm for real-time instance segmentation[J].

Opto-Electron Eng, 2021, 48(11): 210245

Deep contour fragment matching algorithm for real-time instance segmentation

Cao Chunlin¹, Tao Chongben^{1,2*}, Li Huayi¹, Gao Hanwen¹¹School of Electronics and Information Engineering, Suzhou University of Science and Technology, Suzhou, Jiangsu 215009, China;²Tsinghua University Suzhou Automotive Research Institute, Suzhou, Jiangsu 215134, China

Abstract: During the instance segmentation for contour convergence, it is a general problem that target occlusion increases the time for contour processing and reduces the accuracy of the detection box. This paper proposes an algorithm for real-time instance segmentation, adding fragment matching, target aggregation loss function and boundary coefficient modules to the processing contour. Firstly, fragment matching is performed on the initial contour formed by evenly spaced points, and local ground truth points are allocated in each fragment to achieve a more natural, faster, and smoother deformation path. Secondly, the target aggregation loss function and the boundary coefficient modules are used to predict the objects in the presence of object occlusion and give an accurate

收稿日期: 2021-07-20; 收到修改稿日期: 2021-11-10

基金项目: 国家自然科学基金资助项目(61801323, 61972454); 中国博士后科学基金资助项目(2021M691848); 苏州市科技项目基金资助项目(SS2019029); 江苏省高校自然科学基金资助项目(19KJB110021, 20KJB520018)

作者简介: 曹春林(1997-), 男, 硕士研究生, 主要从事图像处理和自动驾驶的研究。E-mail: ccl021916@163.com

通信作者: 陶重霖(1985-), 男, 博士后, 讲师, 主要从事自动驾驶的研究。E-mail: tom1tao@163.com

版权所有©2021 中国科学院光电技术研究所

detection box. Finally, circular convolution and Snake model are used to converge the matched contours, and then the vertices are iteratively calculated to obtain segmentation results. The proposed method is evaluated on multiple data sets such as Cityscapes, Kins, COCO, et al, among which 30.7 mAP and 33.1 f/s results are obtained on the COCO dataset, achieving a compromise between accuracy and speed.

Keywords: instance segmentation; object detection; snake model; object occlusion; initial contour

1 引言

实例分割是一项新兴的计算机视觉任务,需要在具备语义分割和目标检测基础上,对同一类别下物体进行逐个区分预测。预测处理对象是物体实例,如人、交通工具和建筑物等,而像天空、地面等背景则不需要处理。借助实例分割可以进一步的理解场景上下文信息,然后设计出可以操作复杂任务的机器人系统^[1],可用于车辆的定位^[2]和识别^[3],并将这些用于改善自动驾驶的感知系统^[4]。但实例分割是一项具有挑战性的任务,目标遮挡和物体模糊会使得分割更加困难。

为了解决目标遮挡和运动模糊的问题,常用的是两阶段^[5-6]过程的实例分割算法,首先,建立对象的候选区域,然后在候选区域执行分类与校准,实现前景与背景分割。大多数两阶段的算法利用目标检测器给出的边界框进行逐像素预测,而边界框的准确度直接决定后续的分割精度。并且,这种将目标表示为稠密像素的分割算法需要极大的算力和昂贵的后处理。

另一种处理目标的方法是通过轮廓表示,该轮廓由多个顶点组成。与基于稠密像素的表示方法不同,物体变形范围不再局限于边界框内,而且所需的参数点远小于稠密像素,可以进一步提升分割的速度。在 Witkin 等^[7]提出主动轮廓模型(active contour models, ACM)后,基于轮廓的表示方法被广泛应用于实例分割,一般将这种表示方式称为 Snake 模型。Snake 模型算法分为两部分,首先,需要给物体一个初始轮廓,其次,通过能量函数进行变形,最终得到一个贴合目标的轮廓。但是传统的 Snake 模型算法,需要手工作用于轮廓变形的优化函数,这样的优化函数是非凸的,存在局部最优解,不适合用于实例分割中。

Peng 等^[8]提出结合循环卷积的方式解决 ACM 的局部最优问题,循环卷积还可以充分利用轮廓的拓扑特性,通常将这种方法称为深度蛇(Deep Snake, DS)。对于基于 ACM^[7-8]的算法来说,初始轮廓的获取与处理决定分割精度和速度。例如,DS 获取初始轮廓需要更多的步骤,首先,给目标一个检测框,其次把四条边界框中点连接成一个菱形,最后利用 DS 进行变形

获得一个八边形,将这个八边形作为实例分割的初始轮廓,这种获取初始框的方法需要较长的处理时间,实际上会降低实例分割的实时性。因此,本文提出一种新颖的处理初始框的匹配方法,不同于常用的统一匹配,我们提出的匹配方法将检测框四条边界分为许多小段,降低初始轮廓与目标轮廓之间的总误差,使得处理速度更快。在一些分割算法中,使用 Mask RCNN 的 ROI-Align^[6]解决目标遮挡的问题。该方法依赖于检测框的准确性以及需要复杂的后处理,所花费的时间会加长。为此,本文利用目标聚合损失函数进行目标遮挡的优化,获取被遮挡物体准确的检测框,可以减少处理时间,以更快的获得处理结果。

本文的主要贡献:

- 1) 提出新颖的初始轮廓处理方法段落匹配,根据交点将轮廓划分片段,得到更自然的学习路径,避免了变形时发生重叠与交叉,能获得更快的处理速度,还能保证分割的精度。
- 2) 使用目标聚合损失函数,通过对目标预测框进行吸引和抑制周围的检测框,这样提高预测框的准确性,可在密集和遮挡的场景中更好地检测目标。
- 3) 提出一种边界系数用于处理目标与相邻目标的边界关系,将边界系数生成的特征与循环卷积提取的特征进行对齐,将相邻物体的信息保留下来,对相邻物体实现更精准的分割。
- 4) 基于 Snake 模型与循环卷积,利用段落匹配处理初始轮廓,减少处理轮廓时间;以及利用目标聚合损失函数和边界系数优化目标遮挡情况,增加分割准确性,在速度与精度上达到最佳平衡。

2 理论推导

本文的轮廓变形过程为,首先,CenterNet 的检测器^[9]给出一个检测框,将该检测框输入到段落匹配中,得到更为平滑的初始轮廓,最后再根据循环卷积和 Snake 模型进行轮廓变形。本文算法深度轮廓段落匹配(deep contour fragment matching, DCFM)的流程图如图 1 所示。

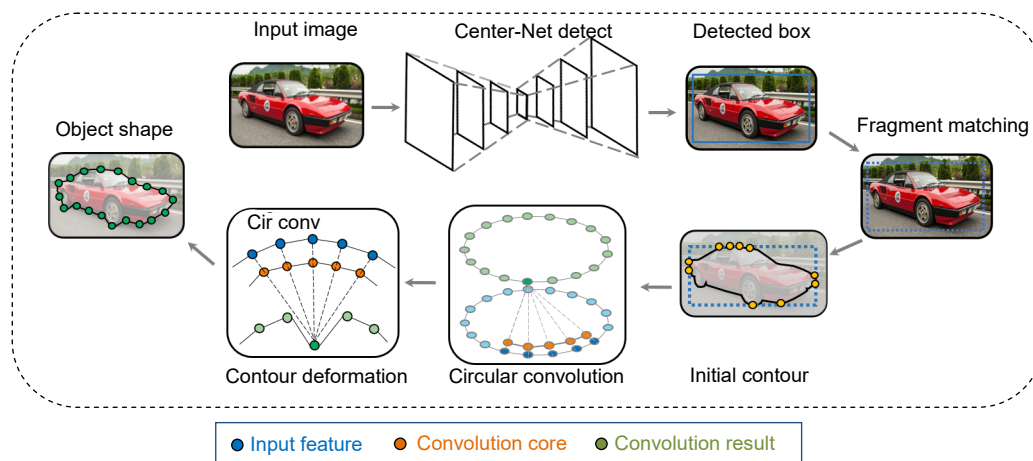


图 1 DCFM 变形流程

Fig. 1 DCFM deformation pipeline

2.1 Center Net 检测器

检测器给对象输出一个边界框(bounding box, BBox), 这个边界框决定了实例分割的精度与速度。本文选用 CenterNet 作为检测器, 该网络有两个特点: 检测过程只定义一个“锚”, 无过多冗余部分; 区分背景和前景的步骤简单, 可直接检测的是目标的中心点和大小。基于这两个优点, 满足本文算法对实时性的要求。

CenterNet 的检测过程中, 每一个类别都有相应的一张热图, 当热图上出现目标的中心点, 则在该点生成一个关键点的高斯分布区域。在关键点(热图峰值点)周围有八个邻点, 经过最大池化后, 只剩下热值最大的中心点, 具体检测过程如图 2 所示。再通过直接回归的目标框尺寸, 然后基于目标框宽高及目标框中心和后处理得到边界框, 本文将基于这个边界框进行后续实验。

2.2 段落匹配

检测器输出的边界框是初始轮廓准确性的关键参考, 而初始框的获取与处理是基于轮廓实例分割方法的重要组成部分。从最早的主动轮廓模型^[7]处理方式来看, 能量函数是轮廓变形的工具, 这也是较为简单直接的方法。Snake 模型使用的能量函数由内部公式和外部公式组成, 内部公式控制轮廓的收敛程度, 外部公式约束着轮廓形状, 使轮廓收敛到真实的形状。但这个模型容易受到噪声干扰以及在凹陷形状收敛时效果差, 所以一般在使用 Snake 模型时采用深度学习的方式避免出现这些问题。DS^[8]的初始轮廓由点拓展与连接组成, 首先取检测框的四条边界的中心构成一个菱形, 其次, 通过 DS 模型将菱形变为八边形, 八边形即初始轮廓。而且这种获取初始轮廓的方法依赖学习过程, 需要花费多一些时间去处理极点变形, 所以本文提出简化该初始化过程的方法, 通过修改初

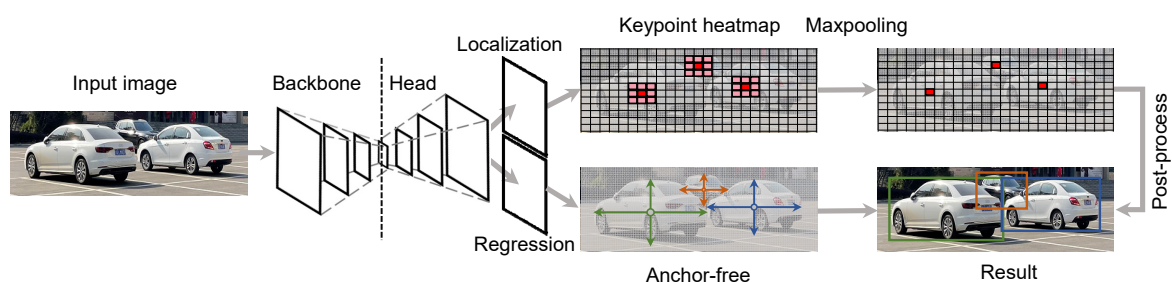


图 2 CenterNet 的检测过程

Fig. 2 Detection process of CenterNet

始轮廓的形状,使用段落匹配处理优势提高处理速度。

段落匹配的过程分为 3 步,首先,取真实轮廓和边界框的交点作为划分片段的端点,设交点数量为 n ,每两个相近的交点为一个片段,得到 $n-1$ 个片段。其次,对 $n-1$ 个片段按照统一距离取 $98-n$ 个点,加上 n 个交点共 98 个,符合后续的采样点数。最后,进行段落匹配处理,对所有点进行二进制赋值,交点为 1,非交点为 0。把 98 个点所在区域像素进行最大池化来提取特征,将池化后的值与点进行相乘。交点赋值为 1 可以保留完整的特征,实现特征与点匹配,匹配后的交点直接作为固定顶点不参与后续变形;而剩余的 $98-n$ 个点,由于赋值为 0,当点与特征匹配时,不能保留相应的特征信息,而这些空白点则作为非固定顶点参与变形。

如图 3 所示,图 3(a)为段落匹配的结果图,图 3(b)为段落匹配与 DS 对比图,图中深蓝色与绿色颜色线条皆为学习偏移量的模拟路径。从图 3(a)左可以看出 DS 生成的初始轮廓不能将物体完整的包裹,导致图 3(b)左的学习路径有一部分处于红色轮廓之外,深蓝色路径是正常的学习偏移量路径,绿色路径表示未被包裹部分的学习偏移量路径,出现学习路径交叉或重叠等情况,导致最终的分割效果不佳。而本文使用的段落匹配,从图 3(a)右可以看出,初始轮廓不能将物体完整的包裹,再看图 3(b)右,学习偏移量的路径全部均在轮廓中,这样可以保证轮廓收敛的准确。但在大部分场景中,会出现许多遮挡和模糊等情况,物体碎片化后可能被检测器忽略而导致分割失败,本文利用目标聚合损失函数和边界系数模块优化这些情况。

2.3 目标遮挡的优化

在实例分割中,目标不会单一出现在背景中,通

常会出现两种遮挡情况,第一种是出现同类物体相互遮挡,第二种是目标被干扰物遮挡。当目标物体被遮挡后,可能被分成几个部分,也可能物体只剩一小部分。由于目标信息的缺少和不完整,会出现误检和漏检,使得目标定位不准,严重限制分割的准确度。可以使用 ROI Align^[6]去解决目标遮挡,该方式准确度较高,但是所需时间较长。受 He^[6]、Peng^[8]和 Wang^[10]对目标遮挡处理的启发,本文使用目标聚合损失函数(target aggregation loss, Target-Agg Loss)优化遮挡和模糊的问题。

使用的损失函数有两个部分组成,目标的牵引与周围物体的抑制。

$$L_{\text{Tow/Ex}} = L_{\text{Tow}} + \alpha \cdot L_{\text{ExGT}} + \beta \cdot L_{\text{ExBox}}, \quad (1)$$

式中: L_{Tow} 是用于目标的牵引,将预测框拉向该目标框。 L_{ExGT} 和 L_{ExBox} 是用于抑制周围物体的,前者将预测框远离其他目标框,后者将预测框远离周围其他目标框的不同预测框。系数 α 与 β 是权重因子用于平衡辅助损失。

L_{Tow} 的作用是为了减少矩阵测量的预测框与真实框的距离。首先针对给出的候选框 $B \in B_+$, 然后将最大 IoU 的预测框作为目标框 G_{Tow}^B , D^B 是回归的预测框。

$$L_{\text{Tow}} = \frac{\sum_{B \in B_+} \text{Smooth}_{L1}(D^B, G_{\text{Tow}}^B)}{|B_+|}, \quad (2)$$

$$G_{\text{Tow}}^B = \operatorname{argmax}_{G \in G} \text{IoU}(G, B), \quad (3)$$

L_{ExGT} 的作用是该目标预测框与其他目标的真实框分开。使用 IoG 进行 Smooth_{L1} 损失,可以防止预测框向周围的真值偏移。 G_{Ex}^B 是除了目标外,具有最大 IoU 的真实框。

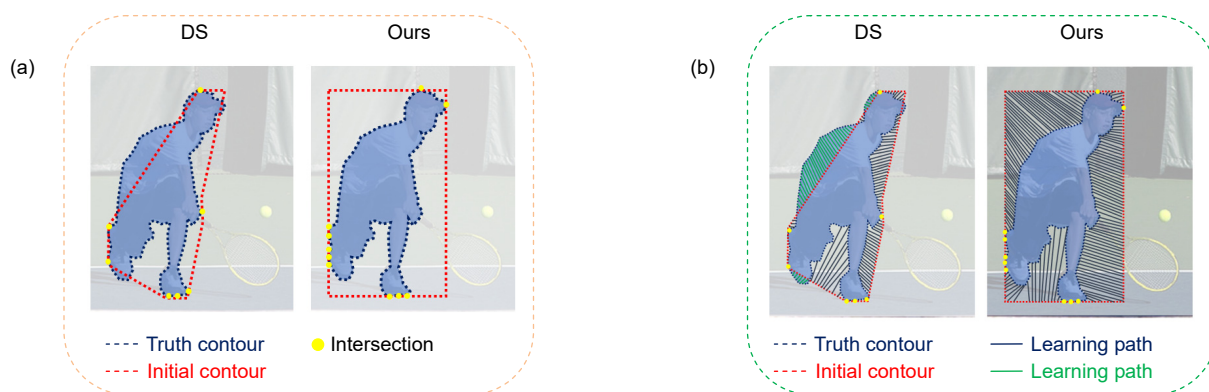


图 3 段落匹配的优势

Fig. 3 Advantages of fragment matching

$$L_{\text{ExGT}} = \frac{\sum_{B \in B_+} \text{Smooth}_{\text{in}}(\text{IoG}(D^B, G_{\text{Ex}}^B))}{|B_+|}, \quad (4)$$

$$G_{\text{Ex}}^B = \underset{G \in g\{G_{\text{ToW}}^B, T\}}{\text{argmax}} \text{IoU}(G, B), \quad (5)$$

$$\text{Smooth}_{\text{in}} = \begin{cases} -\ln(1-x) & x \leq \varphi \\ \frac{x-\varphi}{1-\varphi} - \ln(1-x) & x > \varphi \end{cases}, \quad (6)$$

式中: $\text{Smooth}_{\text{in}}$ 函数, 不但可在(0, 1)范围连续微分, 还可在(0, 1)中调节平滑参数 φ 的值, 当 φ 的值越小, 则表示异常值的抑制效果越好。

L_{ExBox} 的作用将该目标预测框与其他目标的不同预测框分开, 这使得检测器达到接近 NMS 的效果。本文将预测框的集合 B_+ 根据目标的不同分成互不相交的子集 $B_+ = B_1 \cap B_2 \cap \dots \cap B_n$, 然后从中选出两个子集作为预测框, 这两个子集的重叠越小越符合要求。

$$L_{\text{ExBox}} = \sum_{i \neq j} \text{Smooth}_{\text{in}}(\text{IoU}(D^{B_i}, D^{B_j})), \quad (7)$$

式中: $L_{\text{ToW/Ex}}$ 很好地解决遮挡问题, 即使在密集场景下, 也可以提高检测的精度, 给物体较准确的检测框。

目标聚合损失函数提高边界框的准确性, 使得遮挡的物体也能被完整检测, 为后续的轮廓变形提供帮助。给遮挡后的目标输出准确的边界框, 但在边界框内包含多个物体的信息, 相邻物体由于碎片化有时被忽略, 所以本文通过下节的边界系数模块保留更多边界信息, 提高相邻物体分割精度。

2.4 边界系数模块

检测器对物体给出准确的边界框, 每一个边界框中包含物体特征信息, 当出现遮挡情况时, 相邻物体的部分特征信息也留在这一个边界框, 所以在进行轮廓收敛时, 可能会出现将相邻物体信息剔除, 造成相邻物体的变形轮廓不够准确。本文提出一种可以保留

边界框中的边界信息的系数—边界系数(boundary coefficient), 该方法可以保留同一个边界框中多个物体的边界信息, 边界信息包括物体的特征信息和空间信息, 从而更好地描述相邻的物体, 这样可以对遮挡物体进行精准的实例分割。受到 Yolact^[14]的启发, 为了生成边界系数, 在预测头中增加一个分支—全局平均池化(global average pooling, GAP)层, 生成边界系数过程详见 2.5 节的图 4。

边界系数是将检测框划分为 $k \times k$ 的区域, 考虑到精度和速度的平衡, $k=2$ 更符合本文的要求, 一共有四个区域(左上角、左下角、右上角和右下角); 每一个区域都生成一组单独的边界系数 m_{ij} , 这些系数可以将特征信息保留在边界框中去优化轮廓收敛; 边界系数生成的特征与卷积层获得的特征进行对齐, 增加分割的精度。

$$m_{ij} \in R^{W \times H \times 3}, \quad (8)$$

式中: i 表示为第 i 块子区域, 本文将检测框分成四块子区域; j 表示第 j 个检测框, 在本文图像中, 一共有 d 个检测框, 则 j 为第 1, 2, ..., d 个。而 m_{ij} 是一个 $c \times n$ 矩阵, c 表示通道数, n 则表示预测框的数量。

在轮廓的收敛变形中, 容易出现系数生成的特征与循环卷积提取的特征不一致, 本文利用特征对齐把特征进行改进。

$$F(S_0) = \sum_i \sigma[F_{\text{in}}(S_0 + \Delta S_0) \times m_{ij}], i \in \{1, 4\}, \quad (9)$$

式中: $F(S_0)$ 是在 S_0 处产生的对齐特征, F_{in} 为输入特征, S_0 是某一个轮廓点, ΔS_0 是 S_0 回归偏移量(用于减小特征对齐时的偏差), σ 是 sigmoid 函数, 将边界系数与 F_{in} 进行线性组合, 得到系数生成的特征。边界系数生成的特征包含相邻物体, 通过与循环卷积的特征对齐, 这样避免将相邻物体误判, 造成相邻物体的分

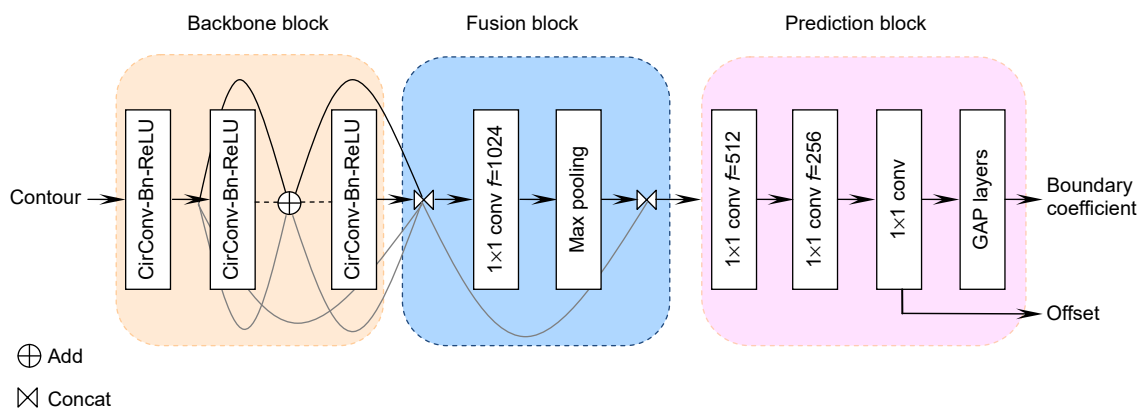


图 4 轮廓变形过程

Fig. 4 Iterative deformation process

割不均匀。该系数的特征对齐,与算法中的循环卷积相辅相成,让遮挡后的呈碎片化物体也能被精准分割。对齐后特征直接用于后续的迭代变形,减小计算量。

2.5 循环卷积与变形

特征的提取和传输是图像处理的重要步骤,而卷积是最常用的方式。通过控制步长与卷积核的大小,可以获得同一图像的多层信息,体现深度学习对特征的深度提取。由于轮廓节点的邻点数量与方向不固定,普通卷积在提取特征时会破坏轮廓的结构,降低分割的准确度。在 DS 中将轮廓看成一个有序点组成的特殊环形,每个节点的邻点数量与方向都可以确定,由此可以定义卷积核。本文使用的初始轮廓由均匀间隔的点构成 $\{x_i | i=1, \dots, N\}$, 循环卷积可以在不破坏结构前提下,充分利用轮廓的拓扑特性提取图像特征。

利用循环卷积提取轮廓上顶点的特征,将初始轮廓上的点设为有序点集。为了不破坏轮廓的结构,使用周期函数 $(R_K)_i$ 代替普通函数。

$$(R_K)_i = \sum_{j=-\infty}^{\infty} R_{i-jK}, \quad (10)$$

$$(R_K * U)_i = \sum_{j=-\infty}^{\infty} (R_K)_{i+j} U_j, \quad (11)$$

式中: K 是轮廓上的顶点数, U 是可学习的核函数, $*$ 是标准卷积。

使用循环卷积也能像普通卷积一样搭建网络层,并将其用于轮廓的特征学习。本文的网络结构与 DS 一致,由主干网络、融合模块和预测头三个模块构成。首先,将初始轮廓输入到主干网络中,主干网络中有 8 个循环卷积层用于多尺度的特征信息提取,将每个顶点的特征输出到融合模块。融合模块将每层的各尺度上的轮廓顶点特征进行连接,然后使用 1×1 的卷积核进行转发,再通过最大池化压缩特征信息,融合多

尺度的轮廓特征。最后,在预测头模块,利用三个卷积层对融合后的特征进行处理,生成偏移量用于每个顶点的轮廓变形;在此基础上增加 GAP 层生成边界系数,该系数则用于描述物体边界。

首先,从经过分配处理的轮廓交点 x 开始初始轮廓进行均匀 K 点采样,其次,对目标的形状也进行 K 点采样,再定义第一个离交点 x 最近的点为第一顶点(一共有 K 个顶点),然后分配各顶点所对应的偏移量,最终定义学习的目标。每个顶点的偏移量由下式计算:

$$(\Delta x_i, \Delta y_i)_{i=1}^K = S\{F_{\text{input}}(x_i, y_i)_{i=1}^K\}, i \in \{1, K\}, \quad (12)$$

式中: (x_i, y_i) 表示第 i 个顶点的坐标, S 是 Snake 模型, F_{input} 是图像的输入特征, $F_{\text{input}}(x_i, y_i)$ 是可学习特征和相对顶点坐标 (x'_i, y'_i) 的级联。 (x'_i, y'_i) 是顶点的相对坐标,由每一个顶点坐标减去顶点中最小值得到的,可以保证变形不影响轮廓的平移。

$$x'_i = \frac{(x_i - x_{\min})}{(x_{\max} - x_{\min})}, y'_i = \frac{(y_i - y_{\min})}{(y_{\max} - y_{\min})}. \quad (13)$$

将每个顶点经过偏移量的迭代计算,结合之前的坐标,更新顶点坐标。

$$\begin{bmatrix} x_i^n \\ y_i^n \end{bmatrix} = \begin{bmatrix} x_i^o \\ y_i^o \end{bmatrix} + \begin{bmatrix} \Delta x_i \\ \Delta y_i \end{bmatrix}, i \in \{1, K\}, \quad (14)$$

式中: (x_i^n, y_i^n) 是更新后的顶点坐标, (x_i^o, y_i^o) 是原始的顶点, $(\Delta x_i, \Delta y_i)$ 是顶点的偏移量。

因为顶点与目标的距离不固定,特别是离目标较远的顶点,进行回归偏移时难度较大,所以本文采用迭代优化的方法解决这一问题。将上一次的偏移量用于下一次的顶点更新,经过更新的顶点离目标更近,起到逐渐变形的效果,其中共有 K 个顶点(K 为 98, 可以覆盖大多数物体)。变形结果如图 5 所示,图中黄色点是物体与初始轮廓的交点,交点为偏移量计算提

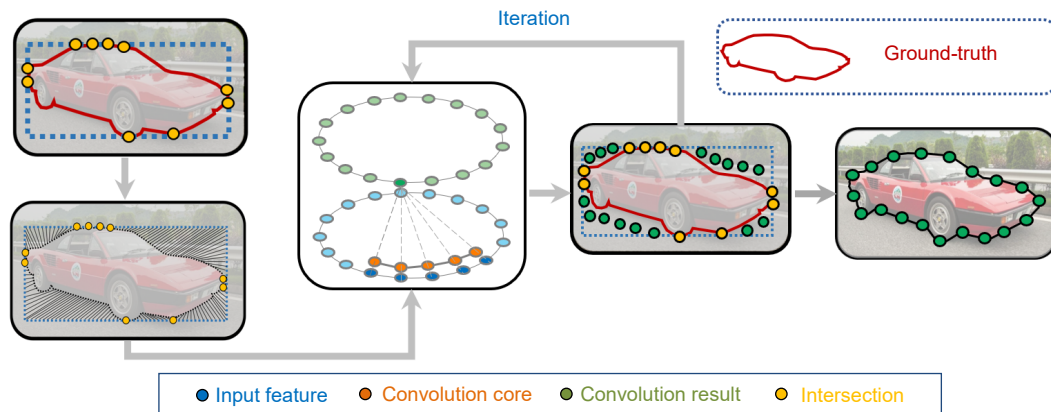


图 5 轮廓变形过程

Fig. 5 Iterative deformation process

供参考, 这些点划分的区域是需要变形的区域; 绿色点是经过更新后的顶点, 慢慢地向物体贴合, 最终在第三次迭代结束收敛, 得到变形结果。

3 实验结果与分析

本文算法在 Cityscapes、Kins、COCO 和 SBD 数据集上与先进算法进行评估比较, 以及在 COCO 与 SBD 上进行消融实验。实验在 Ubuntu18.04 环境下进行, 处理器型号是 Intel i9-9900K, 64 GB 内存, RTX2080 Ti 显卡进行实验。训练时, 先对 CenterNet 检测器训练 60 个 epochs, 学习率为 $5e^{-3}$, 在 20, 40 个 epochs 时学习率下降一半。检测器训练后, 对 Snake 变形模块训练 80 个 epochs, 在 20, 40, 60 个 epochs 时学习率下降至 e^{-3} 。

3.1 与其他先进方法对比

Cityscapes 的场景很丰富, 其中包含数十个城市的不同背景、街景; 同类与不同类的物体相互遮挡且呈碎片化分布, 被遮挡的物体被分成几个部分, 给实例

分割增加挑战性。本文加入目标聚合损失函数, 使得被遮挡的物体也能准确地被边界框包围, 不会因为遮挡过多而造成边界框无法准确定位。表 1 展示与其他先进的方法在 Cityscapes 的测试和验证集的结果。在验证集上本文方法的 AP 分别比 DS 和 PANet 提高 1.7% 和 3.4%, 在 AP_{50} 上分别比 DS 和 PANet 增加 2.2% 和 3.5%, 本文算法在“人”、“车”和“货车”类别取得最佳成绩。本文算法在 Kins 和 Cityscapes 数据集的分割效果如图 6 所示, 两个数据集中有大量的密集和物体重叠场景, 通过本文提出的目标聚合损失函数对重叠物体给出准确的检测框, 以及边界系数模块对重叠物体的特征改进, 避免相邻物体特征的丢失, 保证精确分割物体的真实形状。

Kins 是最大的模态实例分割数据集, 该数据集增加许多附加注释, 主要针对模态实例分割掩码、语义标签和像素级实例标注。该数据集采用单独的遮挡分类模块和多级编码, 去优化目标遮挡的情况, 为解决遮挡问题的算法提供数据集验证。表 2 展示本文方法与多个先进方法在 Kins 数据集的对比结果, “Amdoel”

表 1 与其他先进算法在 Cityscapes 的验证集与测试集对比结果

Table 1 Comparison results of the validation set and test set with state-of-the-art in Cityscapes

Method	Speed/(f/s)	$AP_{val}/\%$	$AP/\%$	$AP_{50}/\%$	Person	Car	Rider	Truck	Bicycle
Mask RCNN ^[6]	2.2	31.5	26.2	49.9	30.5	46.9	23.7	22.8	16.0
SECB ^[11]	11.0	-	27.6	50.9	34.5	52.4	26.1	21.7	18.9
PANet ^[12]	<1.0	36.5	31.8	57.1	36.8	54.8	30.4	27.0	20.8
Deep Snake ^[8]	4.6	37.4	31.7	58.4	37.2	56.0	27.0	29.5	16.4
Ours	5.9	39.1	33.8	60.6	38.5	56.8	27.7	30.7	17.3



图 6 Cityscapes 与 Kins 的效果图

Fig. 6 The renderings of Cityscapes and Kins

表 2 与其他实例分割方法在 Kins 的 AP 对比

Table 2 Comparison of the AP of other instance segmentation methods in Kins

Method	Detection	Amodal seg	Inmodal seg
ORCNN ^[13]	30.9	29.0	26.4
Mask RCNN ^[6]	31.3	29.3	26.6
PANet ^[12]	32.3	30.4	27.6
Deep Snake ^[9]	32.8	31.3	-
Ours	34.5	33.6	29.7

与“Inmodel”是被遮挡和未知物体的两种注释,前者产生的数据较为简单,相比之下后者更复杂,也进一步增加分割难度。本文在“Amdoel”注释 AP 比 DS 提升了 2.3%;在复杂的“Inmodel”注释 AP 依旧稳定保持在 29.7%,经过对比展示本文算法良好的分割性能。

COCO 数据集标注的物体超过 90 类,数据集中的个体数目超过 150 万个,场景选自复杂的日常生活。本文方法与其他先进方法在 COCO 数据集评估分割的性能。直接对原始像素的图像进行测试,不需要进行额外的处理。表 3 展示本文算法与其他实时算法的性能对比,在测试集 test-dev,本文的 AP 比 DS 提高

1.2%;在处理时间取得 34.5 ms,由于不一样的初始轮廓获取方式,比 DS 减少 2.3 ms 的时间,说明本文算法在实时性和精度上达到了先进水平。图 7(a)展示与其他实时算法的折线对比,本文算法如梅花折线所示,左上方梅花块表明取得 36.3 f/s 和 32.6% mAP 的成绩,右下方表示取得 38.4 f/s 和 28.3% mAP 的评估结果,为了在精度与速度取得最佳平衡,本文算法在 COCO 数据集上最佳成绩为 36.3 f/s 和 32.6% mAP;图 7(b)展示与其他算法的 PR 曲线对比,在召回率增加到 0.5 附近时,其他算法精度下降更快,而本文算法下降趋势较缓,基本保持 0.7~0.8 的精度,在精度与召回率上取得最佳平衡。

表 3 与其他的实时方法在 COCO 上的处理时间对比

Table 3 Comparison of processing time on COCO with other real-time methods

Method	Yolact ^[14]	ESE ^[15]	Deep Snake ^[8]	Ours
Val(seg AP/%)	29.9	21.6	30.5	31.8
Test-dev(seg AP/%)	29.8	-	30.3	31.5
Time/ms	34.2	26.0	36.8	34.5

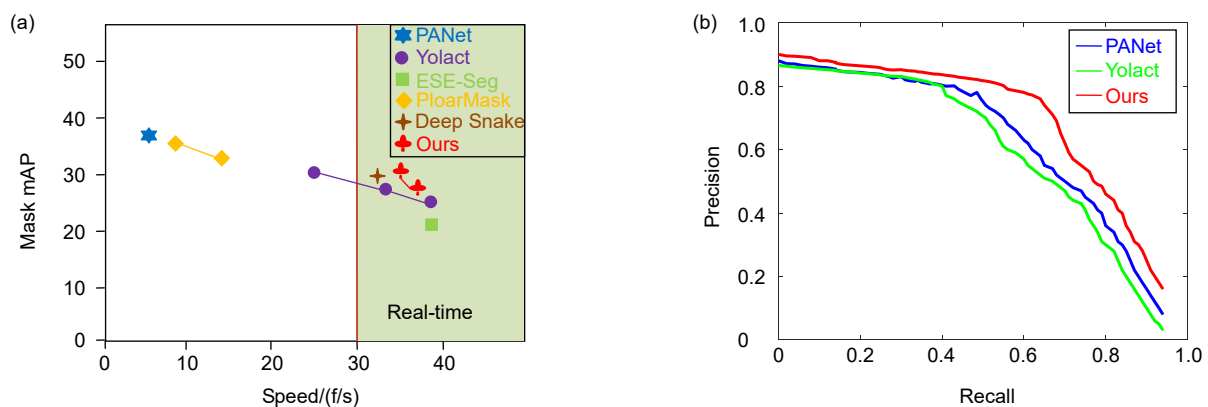


图 7 (a) 各先进方法在 COCO 的速度和实时性能; (b) PR 曲线对比

Fig. 7 (a) The speed and real-time performance of each state-of-the-art in COCO; (b) PR curve comparison

在 SBD 上的对象很多都是单一的, 存在的遮挡情况比 Cityscapes 少。针对这些遮挡, 采用与 Cityscapes 一致的解决方法, 给被遮挡物体一个准确的边界框, 然后用于后续的预测变形。表 4 展示了各算法的实时性, 本文的方法以 35.7 f/s 运行, 比 DS 提升 3.4 f/s, 达到先进的实时水平; 取得 55.8% 的 AP_{vol} , 63.5% 的 AP_{50} 和 50.7% 的 AP_{70} 的结果, 即使 IoU 阈值发生变化, AP 仍然可以保持较好水平, 在速度和精度取得平衡。

图 8 展示本文算法在 SBD 和 COCO 数据集的分割结果。从分割结果看出, 单个物体的分割效果表现良好, 可以准确地将人的身体完整分割。在遮挡过多或者物体非常远时, 也能精准预测物体的形状, 得到一个准确的轮廓。图 9 显示我们整理生活和学习中场景分割效果, 它展示本文算法的泛化性和鲁棒性。

3.2 消融实验

本文对 SBD 与 COCO 数据集进行消融实验, SBD 数据集更适合基于轮廓的分割算法使用; COCO 数据集在实例分割上有 80 种语义类别, 可以全面评估本文

算法处理不同类别目标时的能力。在 CenterNet 给出的定位框基础上, 评估基线与初始轮廓、分段匹配和目标聚合损失函数对实例分割的效果。基线由 CenterNet 与循环卷积组成。循环卷积比普通卷积更加适合轮廓方案, 不会破坏轮廓的拓扑结构。具体过程在目标的定位框基础上, 给出包围目标的轮廓, 再通过循环卷积使轮廓慢慢变形直至贴合目标真实形状。该基线的处理方式是将轮廓表示为环形, 然后使用循环卷积网络进行轮廓的变形。

表 5 和表 6 展示消融实验的结果。表 5 使用的基线(Baseline)是 CenterNet 与循环卷积组合而成。在基线上增加初始轮廓, 该初始轮廓是段落匹配后的, 使得 AP_{vol} 增加 1.9%; 在基线和初始轮廓基础上增加边界系数模块, 使得 AP_{vol} 增加 0.7%。表 6 是在基线、初始轮廓和边界系数组合基础上进行评估, 迭代次数均为 3 次。该消融实验中, 第一种组合使用段落匹配方法和 $Smooth_{L1}$ 损失函数; 第二种组合使用目标聚合损失函数(表 5、表 6 和图 10 用 Target-Agg 表示)和均

表 4 与先进方法的实例分割在 SBD 上的实时比较

Table 4 Real-time comparison of instance segmentation with state-of-the-art on SBD

Method	$AP_{vol}/\%$	$AP_{50}/\%$	$AP_{70}/\%$	Speed/(f/s)
STS ^[16]	29.0	30.0	6.5	37.0
ESE-50 ^[15]	32.6	39.1	10.5	-
ESE-20 ^[15]	35.3	40.7	12.1	38.5
Deep Snake ^[8]	54.4	62.1	48.3	32.3
Ours	55.8	63.5	50.7	35.7



图 8 SBD 与 COCO 的效果

Fig. 8 The segmentation effect of SBD and COCO

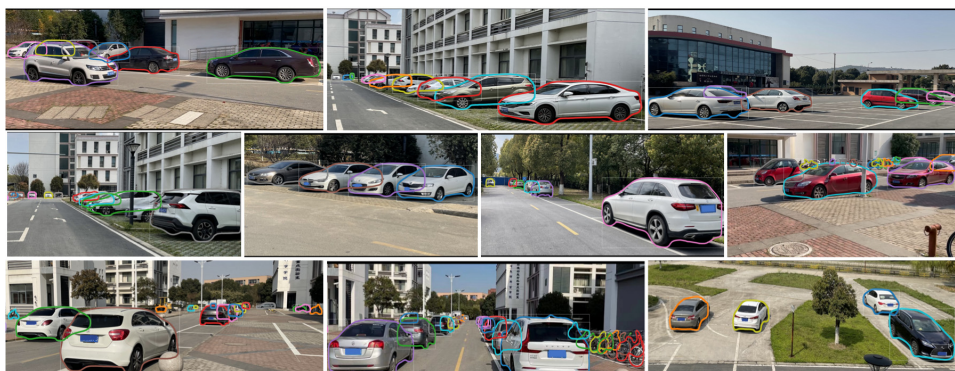


图 9 现实生活中的常见场景分割效果

Fig. 9 Common scene splitting effect in real life

表 5 消融实验在 SBD 的评估结果

Table 5 Evaluation results of ablation experiments in SBD

Models	AP _{vol} /%	AP ₅₀ /%	AP ₇₀ /%
Baseline	53.9	61.5	48.1
+Initial contour	55.8	62.4	49.0
+Boundary coefficient	56.5	63.4	50.3

表 6 在 COCO 数据集中进行消融实验

Table 6 Ablation experiments performed in the COCO

Initial contour	Matching	Loss	AP/%	AP ₅₀ /%	AP ₇₅ /%	AP _S /%	AP _M /%	AP _L /%
BBox	Fragment	Smooth _{L1}	30.3	50.5	32.4	16.4	34.0	46.2
BBox	Uniformity	Target-Agg	30.7	50.8	32.0	16.9	34.5	46.3
BBox	Fragment	Target-Agg	32.6	53.5	34.7	18.9	36.1	48.7

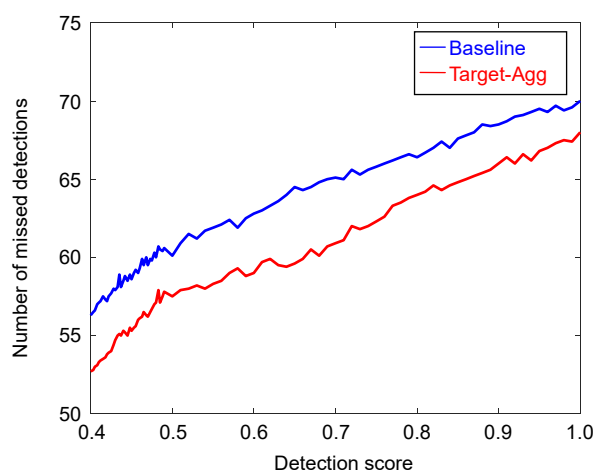


图 10 漏检次数的消融实验

Fig. 10 Ablation experiment of the number of missed detections

匀匹配(uniformity matching), 其中均匀匹配是 DS 所使用的匹配方式; 第三种组合使用段落匹配和目标聚合损失函数, 其 AP 达到 32.6%, 比第一种组合的 AP 提高 2.3%, 比第二种组合 AP 增加 1.9%。图 10 展示在检测中, 因为目标遮挡和目标模糊被漏检的情况, 特别是随着检测阈值的增加, 漏检的次数会增加, 图中蓝线是基线的漏检次数, 在增加目标聚合损失函数后, 漏检次数明显降低 9%~12%, 证明目标聚合损失函数的有效性和鲁棒性。

表 5 表示消融实验结果, 段落匹配后的轮廓使得轮廓变形更加自然, 边界系数对多物体的特征对齐增加分割精度, 实现 AP_{val} 2.6% 的提升。最后为了验证段落匹配与目标聚合损失函数的可靠性, 对不同的匹配方式和损失函数进行比较。表 6 的定性与定量评估都显示, 段落匹配比均匀匹配具有更强的轮廓收敛能力。表 5 和表 6 共同展示了段落匹配的优越性以及本文算法的先进性。

4 结 论

本文提出的处理轮廓变形方法, 将段落匹配后的轮廓作为初始轮廓, 通过目标聚合损失函数对存在遮挡的物体进行准确预测, 最后利用循环卷积和 Snake 模型对轮廓进行迭代变形。在消融实验中, 段落匹配与目标聚合损失的组合, 在 COCO 数据集上取得 32.6% mAP 和 36.3 f/s 的优秀成绩, 证明本文算法可以更快捷和更准确进行轮廓变形。

参考文献

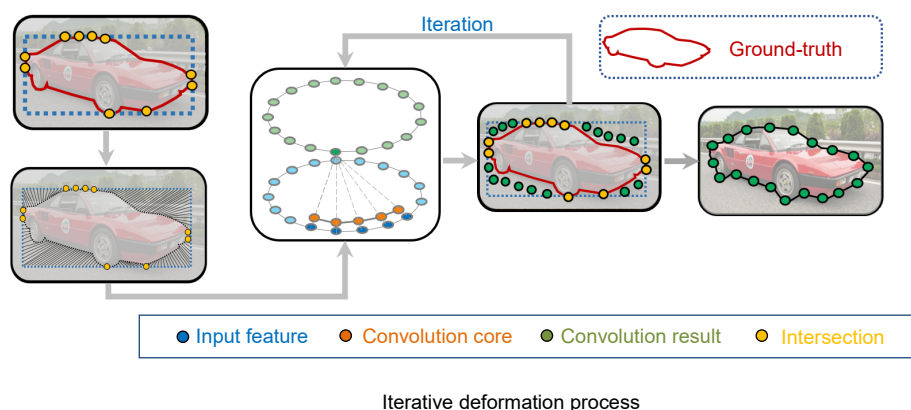
- [1] Fazeli N, Oller M, Wu J, et al. See, feel, act: Hierarchical learning for complex manipulation skills with multisensory fusion[J]. *Sci Robot*, 2019, 4(26): eaav3123.
- [2] Zhang Y, Yang J, Yang Y F. The research on fog's positioning performance of vehicles based on visible light[J]. *Opto-Electron Eng*, 2020, 47(4): 85–90.
张颖, 杨晶, 杨玉峰. 雾对基于可见光的车辆定位性能的研究[J]. *光电工程*, 2020, 47(4): 85–90.
- [3] Meng F J, Yin D. Vehicle identification number recognition based on neural network[J]. *Opto-Electron Eng*, 2021, 48(1): 51–60.
孟凡俊, 尹东. 基于神经网络的车辆识别代号识别方法[J]. *光电工程*, 2021, 48(1): 51–60.
- [4] Ma W C, Wang S L, Hu R, et al. Deep rigid instance scene flow[C]//*Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, 2019: 3609–3617.
- [5] Cao J L, Cholakal H, Anwer R M, et al. D2det: Towards high quality object detection and instance segmentation[C]//*Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 2020: 11482–11491.
- [6] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN[C]//*Proceedings of the 2017 IEEE International Conference on Computer Vision*, Venice, Italy, 2017: 2980–2988.
- [7] Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models[J]. *Int J Computer Vis*, 1988, 1(4): 321–331.
- [8] Peng S D, Jiang W, Pi H J, et al. Deep Snake for real-time instance segmentation[C]//*Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 2020: 8530–8539.
- [9] Zhou X Y, Wang D Q, Krähenbühl P. Objects as points[Z]. arXiv preprint arXiv: 1904.07850, 2019.
- [10] Wang X L, Xiao T T, Jiang Y N, et al. Repulsion loss: detecting pedestrians in a crowd[C]//*Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018: 7774–7783.
- [11] Neven D, Brabandere B D, Proesmans M, et al. Instance segmentation by jointly optimizing spatial embeddings and clustering bandwidth[C]//*Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, 2019: 8829–8837.
- [12] 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*, Salt Lake City, UT, USA, 2018: 8759–8768.
- [13] 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, 2019: 1328–1336.
- [14] Bolya D, Zhou C, Xiao F Y, et al. YOLACT: Real-time instance segmentation[C]//*Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision*, Seoul, Korea (South), 2019: 9156–9165.
- [15] Xu W Q, Wang H Y, Qi F B, et al. Explicit shape encoding for real-time instance segmentation[C]//*Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision*, Seoul, Korea (South), 2019: 5167–5176.
- [16] Jetley S, Sapienza M, Golodetz S, et al. Straight to shapes: real-time detection of encoded shapes[C]//*Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 2017: 4207–4216.

Deep contour fragment matching algorithm for real-time instance segmentation

Cao Chunlin¹, Tao Chongben^{1,2*}, Li Huayi¹, Gao Hanwen¹

¹School of Electronics and Information Engineering, Suzhou University of Science and Technology, Suzhou, Jiangsu 215009, China;

²Tsinghua University Suzhou Automotive Research Institute, Tsinghua University, Suzhou, Jiangsu 215134, China



Overview: With the help of instance segmentation, the scene information can be better understood, and the perception system of autonomous driving can be effectively improved. However, due to the problems such as object occlusion and object blur during detection, the accuracy of instance segmentation is greatly reduced. Deep neural network is a common method to solve object occlusion and blur. Based on computing resources and real-time considerations, contour-based algorithms are other solutions. Active Contour Model (ACM) is a classic contour algorithm, which is called Snake model. Its parameters are less than those based on dense pixels, which speeds up the segmentation. A novel segmentation algorithm based on ACM combined with cyclic convolution is proposed. The algorithm uses center net as the target detector to update the vertices using the iterative calculation of cyclic convolution and vertex offset calculation, and finally fits the real shape of the body. The algorithm has three main contributions. Firstly, for object occlusion and blurring, a loss function (target aggregation loss) is introduced, which increases the positioning accuracy of the detection box by pulling and repelling surrounding objects to the target. Secondly, the initial contour processing is an important step based on the contour algorithm, which affects the accuracy and speed of subsequent instance segmentation. This paper proposes a method of processing the initial contour, which is fragment matching. The initial contour to be processed is caused by evenly spaced points. The detection box is adaptively divided into multiple segments. The segments correspond to the initial contour. Each segment is matched point by point and assigned vertices. These vertices are the key to subsequent deformation. Finally, in dense scenes, it is easy to lose the information of adjacent objects in the same detection box. This paper proposes a boundary coefficient module to correct the misjudged boundary information by dividing the area and aligning the features to ensure the accuracy of boundary segmentation. The algorithm in this paper is compared with multiple advanced algorithms in multiple data sets. In the Cityscapes dataset, an AP_{vol} of 37.7% and an AP result of 31.8% are obtained, which is an improvement of 1.2% AP_{vol} compared to PANet. In SBD dataset, the results of 62.1% AP_{50} and 48.5% AP_{70} were obtained, indicating that even if the IoU threshold changes, the AP does not change much, which proves its stability. Compared with other real-time algorithms in the COCO dataset, a trade-off between accuracy and speed was achieved, reaching 33.1 f/s, while the COCO test-dev has 30.7% mAP. After the above data analysis, it is proved that the algorithm in this paper has reached a good level in accuracy and speed.

Cao C L, Tao C B, Li H Y, *et al.* Deep contour fragment matching algorithm for real-time instance segmentation[J]. *Opto-Electron Eng*, 2021, 48(11): 210245; DOI: 10.12086/oe.2021.210245

Foundation item: National Natural Science Foundation of China (61801323, 61972454), China Postdoctoral Science Foundation (2021M691848), Science and Technology Projects Fund of Suzhou (SS2019029), and Natural Science Foundation of Jiangsu Province (BK20201405, 19KJB110021, 20KJB520018)

* E-mail: ccl021916@163.com