文章编号:1001-5078(2023)09-1449-08

·图像与信号处理·

基于注意力机制的语义增强损失函数与全景分割

郑立冬[1],滕书华[2],谭志国[2],元志安[3],马燕新[3]
(1. 河北省迁安市职业技术教育中心,河北 迁安 064400; 2. 湖南第一师范学院 电子信息学院,湖南 长沙 410205;
3. 国防科技大学 气象海洋学院,湖南 长沙 410073)

摘　要:全景分割是计算机视觉中重要的研究方向。考虑到不同应用场景对语义分割精度的要求不同,本文提出一种基于注意力机制的语义增强损失函数和全景分割方法。首先将语义类别按照重要程度分组,加入注意力机制来对不同语义信息进行区分,并通过对损失权重的设计有效抑制了分类失衡问题;其次设计一种全景分割网络,利用 MaskR-CNN 网络作为实例分割子分支并加入 FPN 结构作为语义分割基准,提高了所需物体种类的分割精度;最后通过设计重叠结果剔除规则避免了网络结构中的实例和语义分割分支输出的重叠问题。通过对 COCO 数据集的对比实验表明,本文提出的语义增强损失函数有效提高了优先级较高语义类别的分割效果,为不同应用场景的全景分割提供了更加高质量的语义信息。

关键词:损失函数;注意力机制;全景分割;实例分割;语义分割
中图分类号:TN249;TP391　　**文献标识码:**A　　**DOI:**10.3969/j.issn.1001-5078.2023.09.023

Attention-based semantic enhancement loss function and panoptic segmentation

ZHENG Li-dong[1], TENG Shu-hua[2], TAN Zhi-guo[2], YUAN Zhi-an[3], MA Yan-xin[3]
(1. Qian'an Vocational Education Center, Qian'an 064400, China;
2. College of Electronic Information, Hunan First Normal University, Changsha 410205, China;
3. College of Meteorology and Oceanography, National University of Defense Technology, Changsha 410073, China)

Abstract: Panoramic segmentation is an important research direction in computer vision. Considering that different application scenarios have different requirements for semantic segmentation accuracy, a semantic enhancement loss function and panoramic segmentation method based on attention mechanism is proposed in this paper. Firstly, the semantic categories are grouped according to their importance, and the attention mechanism is added to distinguish different semantic information, and the classification imbalance is effectively suppressed through the design of loss weight. Secondly, a panoramic segmentation network is designed using MaskR-CNN network as the instance segmentation subbranch and adding FPN structure as the semantic segmentation benchmark to improve the segmentation accuracy of the required object types. Finally, the overlapping problem of instance and semantic segmentation branch output in network structure is avoided by designing overlapping result elimination rules. The comparative experiments on COCO data sets show that the semantic enhancement loss function proposed in this paper effectively improves the segmentation effect of semantic categories with higher priority, and provides more high-quality semantic information for panoramic segmentation of different application scenarios.

Keywords: loss function; attention mechanism; panoptic segmentation; instance segmentation; semantic segmentation

基金项目:湖南省自然科学基金项目(No. 2023JJ0185);湖南省教育厅科学研究重点项目(No. 22A0640)资助。
作者简介:郑立冬(1972 -),硕士,高级讲师,硕士生导师,主要研究方向为智慧教育、职业教育管理与职业培训。
通讯作者:滕书华(1979 -),博士,正高级工程师,硕士生导师,主要研究方向为粗糙集理论、数据挖掘、智能信息处理等。
E-mail:27385918@qq.com
收稿日期:2023-04-23

1 引 言

图像分割技术是计算机视觉领域重要的研究方向。传统分割算法仅仅从人类的直观视觉特征出发,针对图像的颜色分布、纹理特征、点特征来进行分割和分类,其分割效果极大受限于不同场景。随着深度学习的推广和发展,图像分割进入全新的发展时期,Facebook AI 研究院[1]于 2018 年提出全景分割的概念,并给出相关基准。相比于传统的语义分割和实例分割算法,全景分割任务需要在对图像像素点分类的同时区分不同实例并给出识别号。全景分割将语义和实例分割的优点进行了有效结合,既可以得到图像所有物体的分类结果,又可以区分不同物体实例个体,即同时实现了图像背景语义信息和前景实例对象分割的同时处理。

全景分割任务主要分为三个部分:特征提取、语义与实例分割分支和信息融合。通过对输入图像进行特征提取操作,为后续分割过程提供特征信息。主干网络 ResNet[2]作为经典的特征提取网络模块,通过残差结构,在层数增加时依旧可以提升网络收敛性,为高级语义特征提取和分类提供了可行性。SENet[3]通过加入注意力机制,得到特征重要程度与特征之间的连接关系,使模型更加关注信息量大的特征。语义与实例分割分支分别进行语义分割和实例分割,提供语义类别和实例信息,常用网络结构为 PSPNet[4]和 Mask R-CNN[5]。信息融合部分将语义类别和实例信息进行融合,得到最终的全景分割结果。在融合过程中,主要有启发式算法和全景头部(Panoptic Head)两种方法,启发式算法可以在有限时间给出相对不错的结果,但全景头部结构可以得到与语义和实例分割结果一致性较高的融合结果,例如 UPSNet[6]与 OCFusion[7]算法。目前的全景分割算法大多聚焦于提高所有种类的平均分割精度,忽略了不同任务对于不同语义分割结果的需求和重视程度不同,进而导致很多全景分割的精度不理想,实用性不强。本文针对不同的语义类别,提出一种基于注意力机制的语义增强损失函数和全景分割方法,以提高对重要语义类别的分割精度,进而提高分割结果的实用性。

2 算法描述

2.1 语义增强损失函数

全景分割的损失函数通常由语义分割子分支损失函数和实例分割子分支损失函数结合构成:

$$L = \alpha L_{\text{semantic}} + \beta L_{\text{instance}} \tag{1}$$

式中,L_{semantic} 为语义分割子网络损失函数;L_{instance} 为实例分割子网络损失函数;α 和 β 为权重系数。该传统损失函数针对所有语义种类给出了相同的损失代价,导致不同任务中重要性较高语义信息准确性不高。为解决该问题,本文提出一种基于注意力机制的语义增强损失函数,来区分不同重要性的语义信息,并提高重要语义信息的分类精度。本文仍然采取分开计算损失函数的结构:

$$L = \lambda_1 L_{\text{semantic}} + \lambda_2 L_{\text{instance}} \tag{2}$$

其中,λ_1 为语义分割权重系数;λ_2 为实例分割权重系数。因为两个分支的损失设计具有不同规模和规范化策略,因而通过加权来进行不同损失校正。

在全景分割中,语义分割是决定语义分类的关键因素,所以主要对 L_{semantic} 进行重新设计。常用于语义分割的损失函数有 Cross-Entropy Loss[8]和 Focal Loss[9],Focal Loss 具有更高的准确率,Cross-Entropy Loss 则具有更高的召回率,本文选用 Cross-Entropy Loss 来作为损失函数基准。Cross-Entropy Loss 表示实际输出与期望输出之间的距离,用以刻画预测值与真值相似度,交叉熵越小,两个概率分布越接近,传统定义为:

$$L = -\sum_{i=1}^{H}\sum_{j=1}^{W} q_{i,j} \cdot \log(p_{i,j}) \tag{3}$$

其中,$q_{i,j}$ 和 $p_{i,j}$ 都是长度为 C(分类总数)的 one-hot 编码;$q_{i,j}$ 为(i,j) 处的真值向量,正确语义标签位置标注为 1,其他标注为 0;$p_{i,j}$ 为(i,j) 处的预测向量,每个数组元素对应相应分类预测概率;H 和 W 分别为图像的高和宽。

对语义分割来说,交叉熵损失并不理想。因为对一张图来说,交叉熵损失是每一个像素损失的和,它并不鼓励邻近像素保持一致。此外,交叉熵损失无法在像素间采用更高级的结构,所以交叉熵最小化的标签预测一般都是不完整或者是模糊的,它们都需要进行后续处理。

在不同任务需求中,不同物体的语义信息往往重要性程度不同,所以需要对分类的语义信息进行重要程度的划分。如图 1 所示,本文在分割过程中加入注意力机制,将语义种类划分为四个等级,重要程度从 R4 到 R1 依次降低。其划分依据以回环检

测任务为例:相比于动态物体(R2 和 R1),静态物体(R4 和 R3)能够提供更多的可靠鲁棒的参考信息;静态的实例物体(things,R4)又比静态背景(stuff,R3)更具有参考价值和路标功能。动态物体中,相较于车辆等物体(部分情况为静止,R2),人和动物(R1)属于高频移动物体,且出现频次较高,属于干扰信息。

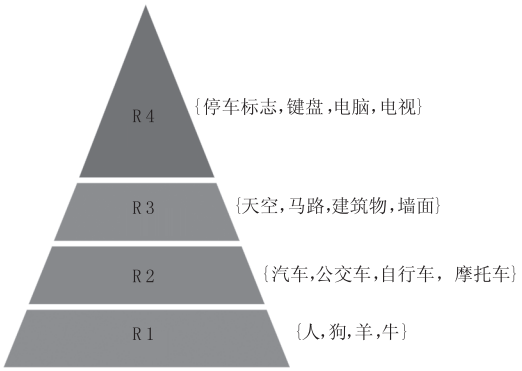


图 1　语义重要性等级划分示意图

Fig. 1 Semantic importance classification diagram

根据分组定义向量 V_1、V_2、V_3 和 V_4,分别储存 4 个分组中物体分类的交叉熵损失。例如第 i 组、第 j 个像素的损失值定义为:

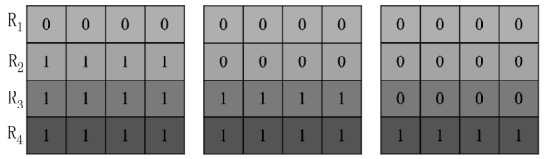
$$(L_{R_i})_j = -\sum_c q_c \log\left(\frac{\exp(O_{c,i,j})}{\sum_{k=1}^{C}\exp(O_{k,i,j})}\right) \tag{4}$$

式中,$O_{c,i,j}$ 为图像(i,j) 处输出 c 分类可能性的张量;q_c 是第 c 个元素为 1 的 one-hot 编码。同样,定义 W_1、W_2、W_3 和 W_4 四个向量来记录 4 个分组中物体分类对应的损失权重,利用损失权重来有效抑制分类失衡问题。下面给出 W_i 的定义为:

$$W_{i,c} = \frac{\alpha}{\log(\beta + s_{i,c})} \tag{5}$$

式中,$s_{i,c}$ 代表第 c 分类在训练集像素点中出现的总次数;α 和 β 代表控制参数。在这里利用 IF-IDF 原理来进行分类元素加权过程,出现频率较高语义类别需要通过降低损失权重来减少迭代过程的每次调整幅度,减少训练发散问题,对于频次较低类别通过提高权重来提高迭代过程的学习过程,对学习过程进行加速。结合上述两部分,定义强损失函数为 $W_i^T L_{Pi}$,其中 $i = 1,2,3,4$。

利用加权操作来调整学习速度之后,设计重要性矩阵 M_t 来完成对不同语义信息的损失函数设定,其定义方式如图 2 所示。



(a)M_1　(b)M_2　(c)M_3

图 2　重要性矩阵 M_t 示意图

Fig. 2 Importance matrix M_t diagram

重要性矩阵 M_t 主要有三部分:M_1、M_2 和 M_3,M_t 的大小为 $H \times W$。与图 1 一致,矩阵中第 1 行区域代表 R1、第 2 行区域代表 R2、第 3 行区域代表 R3、第 4 行区域代表 R4。矩阵中,1 代表重要性高,0 代表重要性低。例如 R4 所包含的 1 数量最多,代表 R4 包含的物体分类最重要。基于 M_t 定义重要性系数 $\theta(M_t)(t = 1,2,3)$ 定义为:

$$\theta(M_t) = \frac{1}{2} \parallel (M_t + \gamma E)^{0.5} \odot (G - M_t) \odot (M_t)^{0.5} \parallel_2^2 \tag{6}$$

式中,E 为全 1 矩阵(幺矩阵);$\gamma \in \mathbb{R}^+$ 为调参,本文实验取 0.5,G 为输出 $O_{c,i,j}$ 在图像(i,j) 处真值分类标签对应的预测概率;$\odot$ 运算表示矩阵对应元素相乘。γ 取值决定了 $\theta(M_t)$ 的大小。当 γ 放大,输出 G 与 M_t 之间的差距便会扩大,尤其当 $M_t = 1$ 时。

语义增强损失函数的设计原则是提高对重要物体损失偏差的敏感性,利用图 3 所示结构来计算损失值。

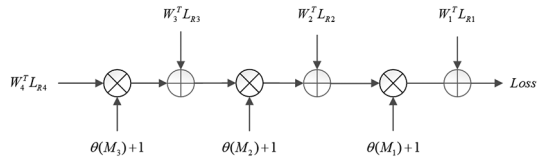


图 3　损失函数计算结构

Fig. 3 Loss function calculation structure

R1 所在分组的重要性最低,设置 R1 组的重要性系数为 1;R2 组的重要性系数为 $\theta(M_1)+1$;R3 组的重要性系数为 $(\theta(M_1)+1)(\theta(M_2)+1)$;R4 组的重要性系数为 $(\theta(M_1)+1)(\theta(M_2)+1)(\theta(M_3)+1)$。最后,语义分割损失函数计算公式为:

$$\begin{aligned} L_{\text{semantic}} &= W_1^T L_{R1} + (\theta(M_1)+1) W_2^T L_{R2} + \\ &(\theta(M_1)+1)(\theta(M_2)+1) W_3^T L_{R3} + (\theta(M_1)+1) \cdot \\ &(\theta(M_2)+1)(\theta(M_3)+1) W_4^T L_{R4} \end{aligned} \tag{7}$$

实例分割子分支的损失函数设计由三部分组成:

分类损失 L_{class}、边界损失 L_{box} 和掩码损失 L_{mask}。L_{class} 和 L_{box} 由采样的感兴趣区域(Region of Interest,RoIs)数量归一化[10],L_{mask} 通过前景 RoIs 数量归一化[11]。给出实例分割子网络的损失函数 $L_{instance}$ 定义为:

$$L_{instance} = L_{class} + L_{box} + L_{mask} \tag{8}$$

最终,语义增强损失函数定义为:

$$L = \lambda_1 L_{semantic} + \lambda_2 (L_{class} + L_{box} + L_{mask}) \tag{9}$$

通过调整权重系数 λ_1 和 λ_2,可实现对语义和实例子分支的不同侧重,同时也可以分别对两个独立任务模块进行单模型训练,且计算量减半。当设计 $\lambda_1 = 0$,即实例分割单模型训练;当 $\lambda_2 = 0$,即语义分割单模型训练。

2.2 全景分割网络结构

本文构建全景分割的思路是利用特征金字塔网络(Feature Pyramid Network,FPN)[12]来修改 Mask R-CNN,其结构思路如图4所示:完成全景分割任务的网络结构需满足如下条件:分辨率足够高以解析微小结构;语义编码足够多以准确预测物体分类;具备多尺度信息,以在不同分辨率上进行预测。FPN(初始用于物体检测)具备高分辨、丰富的多尺度特征提取的作用,因此可以通过附加语义分割网络来完成全景分割任务。如图4所示,FPN 由两部分组成:多种空间分辨率特征的标准网络(本文应用 ResNet[2])和一个带有横向连接的自上而下的轻型通道。自上而下的通道从最深层网络开始,逐步进行上采样,同时从自下而上的路径获取更高分辨率的特征转换。

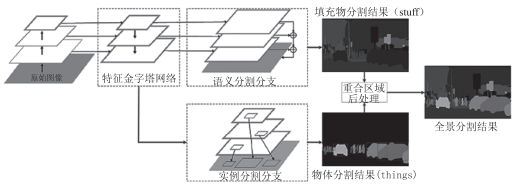


图4 全景分割网络示意图

Fig. 4 Panoramic segmentation network diagram

FPN 结构可以与基于区域的物体检测器直接相连(尤其是相同维度金字塔结构)。Faster R-CNN[10]在不同金字塔层级上执行 RoIs 汇集,并应用共享网络预测每个区域的精炼框和分类标签。如图5所示,使用 Mask R-CNN 来获得实例分割结果,通过添加 FCN(Full Convolutional Networks)分支来扩展 Faster R-CNN[10],以预测候选区域的二进制分割掩码。

为了利用 FPN 特征来获取语义分割结果,利用 Panoptic FPN[13]中提出的一种设计方式,将 FPN 的所有金字塔等级的特征信息合并。FPN 最顶层为1/32分辨率比例,利用三次上采样操作得到 1/4 分辨率比例的特征图,其中每个上采样操作由 3×3 卷积、群体规范、ReLU 和 2 倍双线性上采样组成。

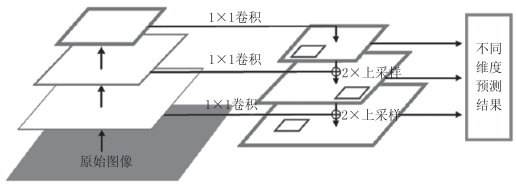


图5 实例分割分支示意图

Fig. 5 Instance segmentation branch diagram

在分辨率比例分别为 1/16,1/8 和 1/4 的 FPN 上重复此操作。每层的上采样结果是相同的 1/4 分辨率比例的特征图,之后按元素求和。最终加入 4 倍双线性上采样和 1×1 卷积来获取与原始图像相同分辨率的像素分类标签。除了填充物(stuff)类之外,在这个分支中增加一个特殊的"其他"类,以作为物体对象的额外像素点输出,可以避免强行预测像素点的填充物种类,造成误判。

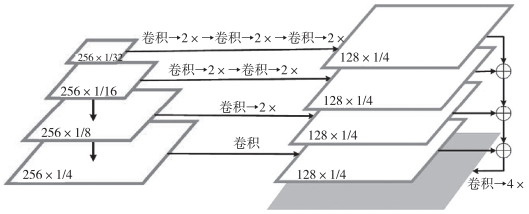


图6 语义分割分支示意图

Fig. 6 Semantic segmentation branch diagram

本文采用的 FPN 配置每个尺度有 256 个输出通道,语义分割分支减少通道数至 128。对于 FPN 之前的主干网络,使用批量标准(Batch Norm,BN)[14]在 ImageNet[15]上预训练 ResNet[2]模型。在微调时,用固定通道仿射变换来代替 BN。

全景输出格式[16]需要为每个图像像素分配类标签和实例 ID(stuff 类不具备实例 ID)。为避免网络结构中的实例和语义分割分支输出重叠问题,加入一种后处理方法,操作方式如下:

(1)不同实例重叠,据置信度得分进行取舍;

(2)语义和实例分割结果重叠,以实例结果优先;

(3)删除标注的"其他"类。

2.3 基于注意力机制的语义增强损失函数与全景分割方法步骤

下面给出本文基于注意力机制的语义增强损失函数与全景分割方法步骤如下：

(1)按重要程度将语义种类划分为若干语义等级，并利用预设初始语义损失函数对若干语义等级加权学习获得语义加强损失函数；

(2)利用预设重要性矩阵确定目标语义分割损失函数；

(3)基于预设权重系数对预设实例分割损失函数以及目标语义分割损失函数进行处理得到目标语义增强损失函数；

(4)利用目标语义增强损失函数对原始图像处理得到实例分割结果以及语义分割结果；

(5)根据预设重叠结果剔除规则对实例分割结果以及语义分割结果进行处理，输出最终目标分割结果。

3 实验及结果分析

3.1 实验数据与评价指标

为对算法进行系统的测算，采用 COCO 数据集作为全景分割训练和测试的数据集。COCO 数据集提供 80 类语义种类，基本覆盖生活中常见物体的学习和分类。同时该数据集也包含了不同分辨率、不同视角和光线下的数据。

全景分割实验中，通常使用 6 种评价指标：平均准确度(Average Precision，AP)、平均召回率(Average Recall，AR)、交并比(Intersection-over-Union，IoU)、分割质量(Segmentation quality，SQ)、识别质量(recognition qulity，RQ)、全景分割质量(Panoptic quality，PQ)。上述评价指标的相关定义如下：

3.1.1 平均准确度

平均准确度(*AP*)用来反映语义种类的分割准确程度，其定义为：

$$AP = \frac{TP}{TP + FP} \tag{10}$$

式中，*TP* 为真阳性数值(true positives)，表示预测正例样本正确的个数；*FP* 为假阳性数值(false positives)，表示预测正例样本错误的个数。*AP* 越高，说明模型的分割效果越好。

3.1.2 平均召回率

平均召回率(*AR*)用来反映语义种类的真正例召回比例，其定义为：

$$AR = \frac{TP}{TP + FN} \tag{11}$$

式中，*FN* 为假阴性数值(false negatives)，表示预测反例样本错误的个数。*AR* 越高，说明模型的分割效果越好。

3.1.3 交并比

交并比(*IoU*)是模型对某一类别预测结果和真实值的的交集与并集比值，其定义为：

$$IoU = \frac{TP}{TP + FP + FN} \tag{12}$$

在检测过程中，当 *IoU* 大于阈值，则判定检测结果为正，反之判错。一般约定，$IoU = 0.5$ 是阈值，*IoU* 越高，说明模型的分割效果越好。

平均交并比(*mIoU*)是模型对每一类预测的结果和真实值的交集与并集的比值，求和再平均的结果，其定义为：

$$mIoU = \frac{1}{2}\left(\frac{TP}{TP + FP + FN} + \frac{TN}{TN + FN + FP}\right)$$

频权交并比(*FWIoU*)是根据每一类出现的频率设置权重，权重乘以每一类的 IoU 并进行求和，其定义为：

$$FWIoU = \frac{TP + FN}{TP + FP + TN + FN} \times \frac{TP}{TP + FP + FN}$$

3.1.4 分割质量

分割质量(*SQ*)指标用来测评语义分割网络，是匹配实例中常用的平均 *IoU* 度量。其定义为：

$$SQ = \frac{\sum_{(i,j) \in TP} IoU(i,j)}{|TP|} \tag{13}$$

式中，$\sum_{(i,j) \in TP} IoU(i,j)$ 为匹配分割区域结果的 *IoU* 总和，而 *SQ* 代表匹配分割区域的平均 *IoU* 值。

3.1.5 识别质量

识别质量(*RQ*)用来测评实例分割子网络，即计算全景分割中每个实例物体识别的准确性。其定义为：

$$RQ = \frac{|TP|}{|TP| + 0.5|FP| + 0.5|FN|} \tag{14}$$

3.1.6 全景分割质量

全景分割质量(*PQ*)指标联合分割质量参数和识别质量参数来对整体全景分割网络框架进行评价，其定义为：

$$PQ = SQ \times RQ = \frac{\sum_{(i,j) \in TP} IoU(i,j)}{|TP| + 0.5|FP| + 0.5|FN|} \tag{15}$$

3.2 全景分割实验结果

首先讨论梯度平衡的损失系数(公式 9)对分割质量的影响。实验过程中,因为语义分割子网络是语义增强损失函数的主要作用点,所以将实例分割子网络系数 λ_2 设置为 1,通过测试不同语义分割损失函数系数 λ_1 来进行分割预测,测试结果如表所示。PQ^{th}表示物体(things)的分割质量,PQ^{st}表示填充物(stuff)的分割质量。从表 1 可以看出,λ_1 过大或过小,网络平衡状态都会被打破,导致两个子网络学习效率均降低。另外当 λ_1 过大,例如 $\lambda_1 = 0.6$ 时,填充网络传入基础网络的梯度幅值过大,此时 $RQ = 50.94$,大大降低了实例分割子网络的预测准确性。该实验表明,当 $\lambda_1 = 0.4$ 时,实例与语义分割网络处于最佳平衡位置,分割质量达到 43.02 %。

表 1 不同分割损失系数权重对应的分割实验结果

Tab. 1 Experiment results of segmentation loss coefficient weights

λ_1	PQ	PQ^{th}	PQ^{st}	SQ	SQ^{th}	SQ^{st}	RQ	RQ^{th}	RQ^{st}
0.25	41.47	48.26	31.23	79.08	82.21	74.36	50.52	57.85	39.44
0.35	42.59	49.86	32.43	79.56	82.93	74.25	51.94	59.67	40.59
0.40	**43.02**	49.70	32.94	79.99	82.88	75.63	**52.06**	59.21	41.26
0.45	42.87	49.45	32.56	79.21	82.42	75.56	52.04	31.27	40.35
0.60	38.12	46.31	32.43	**80.64**	83.06	76.89	50.94	30.04	40.35
0.75	37.56	40.32	31.26	76.57	80.34	75.26	46.35	29.64	40.68

找到最佳参数之后,采用 2.2 节所述的网络结构,进行语义增强损失函数和交叉熵损失函数的语义分割子分支对比实验。保持两组实验的超参配置不变,其实验结果如表所示。从表 2 可以看出,语义增强损失函数相比于交叉熵损失函数,IoU(即 mIoU 和 FWIoU)指标有小幅度提升,填充物(stuff 类)分割表现提高比较明显,全景分割质量 PQ 提高0.97 %,分割质量 SQ 提高了 4.30 %,这是因为在分割过程中,填充物代表的语义分类大多被分到了较高的优先级,所以填充物分割质量提升较大。

表 2 交叉熵损失函数与语义增强损失函数语义分割对比试验

Tab. 2 Comparison test of cross-entropy loss function and semantic enhancement loss function for semantic segmentation

损失函数	mIoU	FWIoU	PQ^{st}	SQ^{st}
交叉熵损失函数	42.56	69.26	30.26	70.06
语义增强损失函数	42.94	69.42	31.23	74.36

为更加直观地测试语义增强损失函数在全景分割过程的作用,下面按照图 1 中的语义种类等级对全景分割进行分组实验,表给出了不同语义分组的分割结果。从表 3 可以看出,相比于交叉熵损失函数,语义增强损失函数对 R4 中的电视语义分类准确度提高了 3.93 %,对 R4 组内均值提高了2.17 %,

表 3 语义增强损失函数与交叉熵损失函数全景分割 AP(%)结果对比

Tab. 3 Comparison of semantic enhancement loss function and cross entropy loss function for panoramic segmentation AP(%) results

分组	语义分类	交叉熵损失函数	语义增强损失函数
R4	停车标志	65.20	**65.97**
	键盘	51.28	**53.31**
	电脑	59.24	**61.18**
	电视	57.38	**61.31**
	组内均值	58.27	**60.44**
R3	天空	39.26	**42.81**
	马路	52.04	**53.83**
	建筑物	67.54	**67.96**
	墙面	53.67	**56.43**
	组内均值	53.13	**55.26**
R2	汽车	**41.85**	40.86
	公交车	64.54	**66.53**
	自行车	**19.75**	18.17
	摩托车	32.18	**34.70**
	组内均值	39.58	**40.07**
R1	人	48.37	**49.59**
	狗	**57.53**	56.24
	羊	**44.53**	42.30
	牛	46.60	**47.10**
	组内均值	**49.26**	48.81
平均值		50.06	**51.14**

对R3组内均值提高了2.13 %,语义增强损失函数有效提高了R3和R4分组中的语义分类准确度。在R1和R2组中,语义增强损失函数的分割准确度和交叉熵损失函数基本持平,语义增强损失函数对R2组内均值提高了0.49 %,对R1组内均值降低了0.45 %。综合上述结果可知,语义增强损失函数有效提高了优先级较高的语义分类的分割准确度,而其他非重要目标的语义分类准确性,少量类别会有轻微程度下降,达到了预期设计函数目标。

图7给出了语义增强损失函数与交叉熵损失函数在不同实测环境下的全景分割结果。由图7可知,交叉熵损失函数对图7(1)中的绿化带以及电线杆上的静态标识、(2)中红绿灯旁边的静态标识以及背景中的天空、(3)中路边的限速标识等分割有误,而语义增强损失函数则有效的提高了上述场景中静态物体的分割效果。

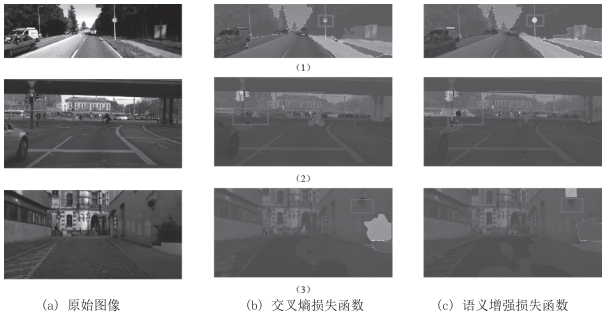


图7 语义增强损失函数与交叉熵损失函数全景分割结果对比图

Fig. 7 Comparison of panoramic segmentation results between semantic enhancement loss function and cross entropy loss function

为了进一步验证本文分割网络在全景分割中的性能,我们选用COCO全景分割挑战赛上出现的Artemis、LeChen、MPS-TU Eindhoven[17]、MMAP-seg以及Facebook AI工作室提出的Panoptic FPN[13]方法与本文网络进行对比,实验结果如表4所示。由表4可以看出,本文网络对填充物的分割质量PQ^{st}明显优于其他5种方法,比效果较好的Panoptic FPN方法还高1.5;物体分割质量PQ^{th}稍稍低于Panoptic FPN方法,但明显高于其他四种方法;本文分割网络的全景分割质量均优于其他5种方法。因此,本文设计的带有语义增强损失函数的分割网络在COCO数据集上取得了较好的分割效果。

表4 COCO数据库中全景分割实验对比

Tab. 4 Comparison of panoramic segmentation experiments in COCO database

	PQ	PQ^{th}	PQ^{st}
Artemis	16.9	16.8	17.0
LeChen	26.2	31.0	18.9
MPS-TU Eindhoven	27.2	29.6	23.4
MMAP-seg	32.1	38.9	22.0
Panoptic FPN	40.9	**48.3**	29.7
本文分割网络	**41.4**	48.2	**31.2**

4 结 论

为了提高不同应用场景下重要目标分割的准确度和可靠性,本文提出一种基于注意力机制的语义增强损失函数和全景分割方法。通过增加注意力机制,增强对任务关注语义信息的敏感度,提高对特定物体和背景的分类精度;同时设计相应的全景分割网络,提高对所需物体种类的分割精度。最后通过设计重叠结果剔除规则避免了网络结构中的实例和语义分割分支输出的重叠问题。对COCO数据集的对比实验表明,本文提出的语义增强损失函数有效提高了优先级较高语义类别的分割效果,为不同应用场景的全景分割提供了更加高质量的语义信息,进而增强了全景分割方法的实用性。

参考文献:

[1] Kirillov A, He K, Girshick R, et al. Panoptic segmentation [C]//In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16 - 18 June 2020: 9396 - 9405.

[2] Kim D, Woo S, Lee J Y, et al. Video panoptic segmentation US: 2021326638A[P]. 2023 - 08 - 22.

[3] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132 - 7141.

[4] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2881 - 2890.

[5] He K, Gkioxari G, Dollar P, et al. Mask R-CNN[C]// Proceedings of the IEEE International Conference on Computer Vision, Italy, 24 - 27 Oct, 2017: 2980 - 2988.

[6] Xiong Y, Liao R, Zhao H, et al. Upsnet: a unified panoptic segmentation network [C]//Proceedings of the IEEE/

CVF Conference on Computer Vision and Pattern Recognition,2019:8818 – 8826.

[7] Lazarow J, Lee K, Shi K, et al. Learning instance occlusion for panoptic segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2020:10720 – 10729.

[8] De Boer P T, Kroese D P, Mannor S, et al. A tutorial on the cross-entropy method[J]. Annals of Operations Research,2005,134(1):19 – 67.

[9] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//IEEE International Conference on Computer Vision,2017:2999 – 3007.

[10] 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 and Machine Intelligence, IEEE,2017,39(6):1137 – 1149.

[11] He K, Gkioxari G, Dollar P, et al. Mask R-CNN[C]// Proceedings of the IEEE International Conference on Computer Vision, Italy,24 – 27 Oct,2017:2980 – 2988.

[12] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//In the Proceedings-30th IEEE Conference on Computer Vision and Pattern Recognition,2017:936 – 944.

[13] Kirillov A, Girshick R, He K, et al. Panoptic feature pyramid networks[C]//In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA,16 – 18 June 2020:9396 – 9405.

[14] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]// 32nd International Conference on Machine Learning, 2015:448 – 456.

[15] Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision,2015,115(3):211 – 252.

[16] Elharrouss O, Al-Maadeed S, Subramanian N, et al. Panoptic segmentation: a review[J]. arXiv:2111.10250V1.

[17] De Geus D, Meletis P, Dubbelman G. Panoptic segmentation with a joint semantic and instance segmentation network[J]. arXiv:1809.02110.