

基于掩码重建的域自适应夜间语义分割方法

丁美荣, 郎济聪, 刘庆龙, 卓金鑫

(华南师范大学软件学院, 广东 佛山 528225)

摘要: 语义分割在自动驾驶中扮演着至关重要的角色, 然而日间场景源域的数据训练模型难以适应夜间场景目标域的数据。为此, 提出一种基于掩码建模的域自适应夜间语义分割网络, 主要包含自训练模块与掩码重建模块。在自训练模块中, 预训练模型生成的伪标签能辅助模型预测夜间目标域, 通过缩小日间源域与夜间目标域间的分布差异, 以提升模型在夜间场景的语义分割性能。在掩码重建模块中, 针对不同类别目标使用特定的掩码策略, 以在重建掩码图像过程中充分提取不同类别间的上下文语义信息, 增强模型在夜间场景上的鲁棒性。在多个城市街道场景数据集上的实验表明, 所提模型的平均交并比为64.3%, 有效提升了分割精度。

关键词: 夜间语义分割; 域自适应; 掩码重建; 自训练; 上下文信息

DOI: 10.11907/rjdk.241404

中图分类号: TP391.14

文献标识码: A

开放科学(资源服务)标识码(OSID):

文章编号: 1672-7800(2025)007-0200-06



Domain Adaptive Nighttime Semantic Segmentation Method Based on Mask Reconstruction

DING Meirong, LANG Jicong, LIU Qinglong, ZHUO Jinxin

(School of Software, South China Normal University, Foshan 528225, China)

Abstract: Semantic segmentation plays a crucial role in autonomous driving, however, training models on daytime scene source domain data is difficult to adapt to nighttime scene target domain data. To this end, a domain adaptive nighttime semantic segmentation network based on mask modeling is proposed, which mainly includes a self training module and a mask reconstruction module. In the self training module, the pseudo labels generated by the pre trained model can assist the model in predicting the nighttime target domain. By reducing the inter domain distribution difference between the daytime source domain and the nighttime target domain, the semantic segmentation performance of the model in nighttime scenes can be improved. In the mask reconstruction module, specific mask strategies are used for different categories of targets to fully extract contextual semantic information between different categories during the reconstruction of mask images, in order to enhance the robustness of the model in nighttime scenes. Experiments on multiple urban street scene datasets have shown that the proposed model has an average intersection to union ratio of 64.3%, effectively improving segmentation accuracy.

Key Words: nighttime semantic segmentation; domain adaptation; mask reconstruction; self-training; context information

0 引言

语义分割在自动驾驶、医学图像等实际场景发挥着重要作用, 目的是将每一个像素分配到对应类别^[1,2]。近年来, 深度学习技术的发展和大规模数据集的扩充降低了语义分割像素级分类任务的难度, 但在实际应用场景中, 精

细标注的标签需要耗费大量人力时间, 单张图像的标注工作需消耗1.5 h^[3]。同时, 现有语义分割模型主要针对正常天气条件下的日间场景(源域), 夜间场景(目标域)受光线过曝或过暗、图像运动模糊的影响, 导致语义分割这种像素级任务的注释工作十分困难。并且, 如果模型的训练数据仅为日间图像, 将会发生域偏移现象, 导致模型在夜间场景中的分割效果很难达到预期。

收稿日期: 2024-04-25

扫描二维码阅读全文:



作者简介: 丁美荣(1972-), 女, CCF高级会员, 硕士, 华南师范大学软件学院副教授, 研究方向为自然语言处理和智能软件开发、人工智能教育; 郎济聪(1998-), 男, CCF会员, 华南师范大学软件学院硕士研究生, 研究方向为域自适应、语义分割; 刘庆龙(1999-), 男, 华南师范大学软件学院硕士研究生, 研究方向为迁移学习、图像分类; 卓金鑫(1999-), 男, 华南师范大学软件学院硕士研究生, 研究方向为迁移学习、图像分类。本文通讯作者: 郎济聪。

为了解决域偏移问题,域自适应方法将模型在源域上学到的知识迁移到目标域上,使模型适应目标域的数据分布,以提升其在无标签目标域上的语义分割准确性。域自适应根据其训练策略的不同,可分为基于对抗训练和基于自训练的方法^[4-7]。在域自适应方法中,对抗训练通过特征生成器与域鉴别器之间互相博弈提取域不变特征。在整个过程中,特征生成器的任务是生成在源域和目标域中具有良好泛化能力的特征;判别器的任务是区分这些特征来自源域还是目标域,迫使生成器尽可能学习到域不变特征,以实现从源域到目标域的迁移目的。自训练在预训练的基础上,生成无标签目标域数据的伪标签来训练语义分割模型。由于生成伪标签的原始图像来自目标域,所以自训练在一定程度上提升了模型在目标域上的性能。尽管,基于对抗训练和自训练的方法在一定程度上缓解了域偏移带来的性能下降问题,但在实际应用场景中,特别是夜间环境下,对自动驾驶这类涉及生命财产安全的领域,现有方法的准确度远远不够。

为了解决域自适应夜间语义分割存在的问题,不少人提出许多基于深度学习的方法。例如,Wu等^[8]提出一个单阶段域自适应网络DANet,通过带标签的日间场景源域数据集与包含粗略对齐的日一夜图像对的无标签夜间场景目标域数据集,结合对抗性训练实现了夜间场景图像的语义分割。同时,该模型还使用重加权策略处理日一夜图像对的对齐误差和日间图像预测错误,以提升模型在小物体类别上的预测准确性。Yang等^[9]提出的Bi-Mix采用了一种新的数据增强技术,通过混合样本进行双向学习以增强模型鲁棒性。Lee等^[10]提出GPS-GLASS,首先利用日间与夜间图像的GPS对齐进行跨领域对应匹配,以在目标域上获得像素级的伪标签监督,其次在日间视频帧之间进行光流估计,最后应用基于GPS的缩放操作获取另一个像素级的标签来微调伪标签,提升伪标签精度。然而,上述域自适应方法主要依托对抗训练来缩小不同域之间的差异,无法避免对抗训练过程中的一个固有问题——无法充分把握每个类别的全局语义信息,以致于影响了训练过程中的稳定性。

目前,大多数域自适应语义分割算法采用自训练方法避免对抗学习过程中训练不稳定的问题。Hoyer等^[11]提出DAFormer对基于卷积的网络架构和基于Transformer的网络架构在域自适应上的性能进行基准测试和比较,揭示了Transformer的潜力,具体为通过Transformer编码器和一个多级上下文感知特征融合解码器,利用稀有类采样、ImageNet类特征距离、学习率预热3个训练策略避免模型对源域发生过拟合^[12]。Hoyer等^[13]在DAFormer基础上提出HRDA,结合高分辨率与低分辨率图像的多尺度融合策略捕获长距离上下文依赖性和细节分割信息,显著提升了域自适应在语义分割任务上的性能。虽然,以上方法在域自适应语义分割上的成效明显,但夜间场景中过暗的光线通

常会使语义分割模型难以准确、清晰地划分大类别(路面、建筑、天空等)边界。同时,模型在低光照区域也难以捕获小物体类别(交通标志、行人、自行车等)信息。由图1(彩图扫描OSID码可见,下同)绿色区域可见,树木/天空、建筑/天空、机动车道/人行道的类别边界非常容易发生预测错误。由图1黄色区域可见,由于低光照和低像素占比的影响,小物体类别(栅栏、栏杆、交通标志、灯光等)在模型预测的过程中容易被忽略。图1中,左侧为原图,右侧为分割结果。

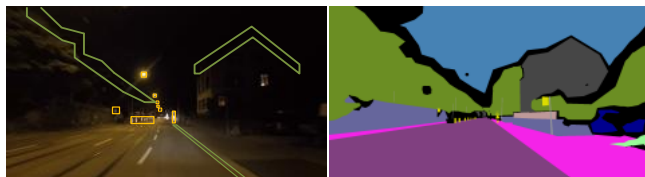


Fig. 1 Classification results of large category boundaries and small object categories

图1 大类别边界与小物体类别划分结果

为了提升模型在夜间场景下对大类别边界与小物体类别识别的准确性,本文提出一种基于掩码重建的域自适应语义分割网络。首先,对大类别边界与小物体类别按照特定掩码策略进行掩码;其次,语义分割模型对被掩码的图像进行重建恢复,促使模型在重建过程中充分获取目标域的上下文信息,以增强模型在目标域上的鲁棒性,进而提升了模型在夜间场景中的语义分割能力。

本文主要贡献为:①提出一种基于掩码重建的域自适应语义分割网络,旨在夜间场景下能有效进行语义分割;②针对不同类别对象,从大类别边界局部与小物体类别全局两方面执行特定的掩码策略,强化模型分割能力;③在DarkZurich-test、DarkZurich-val、BDD100k-night和Night-time Driving数据集上评估域自适应语义分割网络的有效性^[14,15]。

1 本文方法

图2为基于掩码重建的域自适应语义分割模型,由源域监督训练模块、目标域自训练模块、掩码重建(Mask Reconstruction, MR)模块组成。源域监督训练模块利用源域中可用的标注数据来训练一个基础模型,以理解和学习源域数据特征和模式,为后续域自适应步骤奠定基础。目标域自训练模块在源域监督训练模块的基础上,进一步提升模型在目标域上的性能,通过目标域未标注的数据来调整和优化模型,使其更好地适应目标域分布。掩码重建模块则进一步针对夜间场景下的大类别边界与小物体类别进行掩码,使模型能以上下文信息为线索重建掩码区域,增强模型在复杂场景下的鲁棒性。

1.1 源域监督训练模块

无监督的域自适应语义分割任务定义为:给定一个有

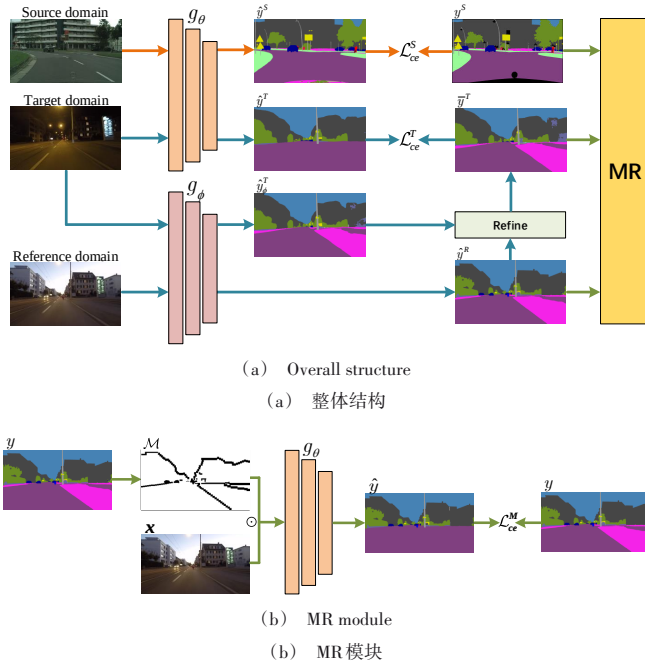


Fig. 2 Nighttime semantic segmentation network architecture

图2 夜间场景语义分割网络结构

标签源域 $S \in \{X^S, Y^S\}$ (例如 Cityscapes^[16]) 和一个目标域 $T \in \{X^T, R^T\}$ (例如 DarkZurich^[15]), $X = \{x_k\}_{k=1}^N, Y = \{y_k\}_{k=1}^N$ 分别表示图像及其对应的标签, $R = \{x_k^R\}_{k=1}^N$ 表示在正常场景下与目标域对齐的参考域图像。由于源域数据存在精细标注的标签 $Y^S = \{y_k^S\}_{k=1}^N$, 因此可通过交叉熵损失 (cross-entropy loss) 对语义分割模型 g_θ 进行训练。具体计算过程如式(1)所示。

$$\mathcal{L}_{ce}^S = - \sum_{j=1}^{H \times W} \sum_{c=1}^C y_{i,j,c}^S \log g_\theta(x_i^S)^{j,c} \quad (1)$$

式中: H, W, C 分别表示图像长度、宽度和维度。

1.2 目标域自训练模块

源域与目标域导致的域间差异难以通过源域上的监督训练进行弥补。为了进一步缩小域间差异, 采用自训练方法, 通过教师网络 g_ϕ 为无标签目标域数据生成伪标签 $\hat{y}$, 以更好地将模型从源域转移到目标域。具体计算公式如式(2)所示。

$$\hat{y}_{\phi,i,j,c}^T = [c = \operatorname{argmax}_c g_\phi(x_i^T)^{j,c'}] \quad (2)$$

式中: $[\cdot]$ 表示 Iverson bracket。

由于伪标签中包含的噪声会误导和扰乱模型学习过程, 而自训练过程中通过伪标签生成的置信度能减少噪声影响, 确保模型从带噪声的伪标签中有效地进行学习。

$$q_i^T = \frac{\sum_{j=1}^{H \times W} [\max_{c'} g_\phi(x_i^T)^{j,c'} > \tau]}{H \times W} \quad (3)$$

式中: τ 为控制置信度的阈值。

进一步, 使用参考域的图像修正伪标签确保伪标签的精确性。

$$\hat{y}^T = \operatorname{Refine}(\hat{y}_\phi^T, g_\phi(x^R)) \quad (4)$$

式中: $\operatorname{Refine}(\cdot)$ 表示对伪标签进行过滤和修正。

在伪标签 $\hat{y}$ 、置信度 q_i^T 的约束下, 目标域自训练模块可表示为:

$$\mathcal{L}_{ce}^T = - \sum_{j=1}^{H \times W} \sum_{c=1}^C q_i^T \cdot \hat{y}^T \log g_\theta(x_i^T)^{j,c} \quad (5)$$

1.3 掩码重建模块

目标域自训练模块在一定程度上弥补了源域与目标域之间的差异, 但在目标域夜间场景中面临的一个难点是: 低光照和光线过曝造成的噪声容易使大类别边界和小类别难以呈现出完整清晰的原型, 使语义分割模型难以理解到夜间场景中类别间的边界语义信息与小物体类别语义信息。由于各类别的语义信息受其相邻类别和位置关系的影响, 本文尝试引入掩码重建模块强化相邻类别的边界关系和小物体类别的局部上下文关系, 以解决该问题。

掩码重建模块对大类别边界与小物体类别分别执行两种不同的掩码策略, 以提升类别边界和小物体类别在夜间场景中的语义分割准确性。针对大类别边界, 掩码重建模块对特定类别之间的边界进行掩码, 语义分割模型在后续的重建过程中提取不同类别之间的边界信息以增强类别之间的联系, 确保在夜间复杂场景中也能获得清晰的边界分割结果。具体计算公式如式(6)所示。

$$\mathcal{M}^b = \begin{cases} 0, & \text{if } j = N(i), \\ 1, & \text{otherwise,} \end{cases} \quad (i, j) \in \mathcal{B} \quad (6)$$

式中: i, j 表示 i, j 类别的像素; $N(\cdot)$ 表示该像素周围像素的类别; $\mathcal{B}$ 表示两两相邻的类别对集合。

针对夜间场景中的小物体类别, 掩码重建模块对选取的小物体类别整体进行掩码, 以增强模型对小物体类别的识别能力, 确保语义分割网络能在夜间环境中捕捉到小物体类别的语义信息, 从而提升分割准确性。具体计算公式如式(7)所示。

$$\mathcal{M}^s = \begin{cases} 0, & \text{if } c \in \mathcal{S}, \\ 1, & \text{otherwise,} \end{cases} \quad (7)$$

式中: c 表示 c 类别的像素; $\mathcal{S}$ 表示小物体类别像素的集合。

大类别边界掩码 $\mathcal{M}^b$ 与小物体类别掩码 $\mathcal{M}^s$ 组成的整体掩码 $\mathcal{M}$ 为:

$$\mathcal{M} = \mathcal{M}^b + \mathcal{M}^s \quad (8)$$

为了充分利用图像的局部信息和上下文信息, 对掩码区域周围的像素通过池化操作进行扩充:

$$\mathcal{M} = \operatorname{Pooling}(\mathcal{M}) \quad (9)$$

式中: $\operatorname{Pooling}(\cdot)$ 表示池化操作。

由此, 掩码图像 x^M 可由掩码 $\mathcal{M}$ 与输入图像 x 通过对应的位置元素相乘得到, 如式(10)所示。

$$x^M = \mathcal{M} \odot x \quad (10)$$

式中: $\odot$ 表示哈达玛乘积。

在重建过程中, 语义分割模型根据掩码图像 x^M 对每个

位置的像素进行还原, 未掩码区域的语义信息作为被掩码区域的提示信息进行重建, 旨在还原掩码区域的过程中增强模型理解大类别边界与小物体类别的能力, 提升其在夜间场景中的分割能力。重建过程的损失函数为:

$$\mathcal{L}_{ce}^M = - \sum_{j=1}^H \sum_{c=1}^C q_i^{S,T,R} \cdot y_{i,j,c}^{S,T,R} \log g_{\theta}(\mathbf{x}_i^M)^{j,c} \quad (11)$$

总损失函数为:

$$\mathcal{L} = \mathcal{L}_{ce}^S + \mathcal{L}_{ce}^T + \lambda^M \mathcal{L}_{ce}^M \quad (12)$$

式中: λ^M 为超参数。

2 实验结果与分析

2.1 实验环境与参数设置

本文实验编程语言为 Python 3.8.5, 深度学习框架为 Pytorch 1.7.1, 在 NVIDIA A40 GPU 上采用 AdamW 优化器进行训练, 权重衰减为 0.01, 动量为 0.9, 初始学习率为 0.000 6, batchsize 大小为 2, 进行 40 k 次迭代训练, 具体参数如表 1 所示^[17]。

Table 1 Experimental setup and parameter settings

表 1 实验环境及参数设置

配置	参数
编程语言	Python 3.8.5
深度学习框架	Pytorch 1.7.1
GPU	NVIDIA A40
优化器	AdamW
Batchsize	2
迭代轮次/k	40

2.2 数据集

本文选用 Cityscapes 数据集作为源域, DarkZurich 数据集作为目标域, 域自适应的过程可表示为 Cityscapes $\rightarrow$ DarkZurich^[15,16]。Cityscapes 数据集汇集了 5 000 张城市街道景观的图像, 每幅图像都精细标注了 19 个类别, 其中 2 975 张图像用于训练, 500 张图像用于验证, 1 525 张图像用于测试。DarkZurich 数据集包含了 2 416 张夜间场景、2 920 张黄昏场景和 3 041 张日间场景的无标签图像, 并通过 GPS 将 3 个域的图像进行粗略对齐。此外, 还包含 201 张标注好的夜间场景图像, 50 张属于 DarkZurich-val 的图像用于验证, 以防止模型过拟合, 151 张属于 DarkZurich-test 的图像用于测试, 以评估模型的有效性。

2.3 评价指标

本文采用平均交互比 (Mean Intersection over Union, mIoU, mIoU) 评估模型的有效性。

$$mIoU = \frac{1}{N} \sum_{c=1}^N IoU_c \quad (13)$$

式中: N 为类别数; IoU_c 为第 c 个类别的交并比 (Intersection Over Union, IoU)。

IoU_c 计算公式为:

$$IoU_c = \frac{TP_c}{TP_c + FP_c + FN_c} \quad (14)$$

式中: TP_c 为正确预测为类别 c 的像素数; FP_c 为错误预测为类别 c 的像素数; FN_c 为真实属于类别 c 但未被预测的像素数。

2.4 比较实验

为了评估本文方法的有效性, 在 DarkZurich-test 数据集上与 DANNet、GPS-GLASS、DAFormer、SePiCo、HRDA、MIC 模型进行比较, 结果如表 2 所示^[8,10,11,13,18,19]。由此可知, 所提模型在 DarkZurich-test 数据集上对 19 个类别的整体 mIoU 为 63.4%, 8 个子类别为最优结果, 8 个子类别为次优结果; 针对大类别 (building, wall, fence, vegetation, terrain, sky) 模糊边界, 所提模型在 building、fence、sky 上的 mIoU 分别为 86.2%、31.6%、89.3%, 相较于最优的比较模型提升 2.2%、3.2%、3.0%; 在类别 wall、vegetation、terrain 上也取得了次优结果; 针对小物体类别 (pole, light, sign, person, rider, motorcycle, bicycle), 所提模型在类别 light、motorcycle 上的 mIoU 分别为 59.9%、56.7%, 相较于最优的比较模型提升 7.2%、4.1%, 在其余 5 个类别上也均取得了次优结果。

Table 2 Experimental results of different models on DarkZurich-test

表 2 不同模型在 DarkZurich-test 上的实验结果 (%)

Semantic Class	DAN-Net ^[8]	GPS-GLASS ^[10]	DAFormer ^[11]	SePiCo ^[18]	HRDA ^[13]	MIC ^[19]	Ours
road	90.0	91.6	93.5	93.2	90.4	94.8	91.4
sidewalk	54.0	63.1	65.5	68.1	56.3	75.0	63.3
building	74.8	71.2	73.3	73.7	72.0	84.0	86.2
wall	41.0	34.7	39.4	32.8	39.5	55.1	51.4
fence	21.1	26.7	19.2	16.3	19.5	28.4	31.6
pole	25.0	41.4	53.3	54.6	57.8	62.0	57.9
light	26.8	39.7	44.1	49.5	52.7	35.5	59.9
sign	30.2	38.4	44.0	48.1	43.1	52.6	66.8
vegetation	72.0	68.6	59.5	74.2	59.3	59.2	77.7
terrain	26.2	34.8	34.5	31.0	29.1	46.8	45.8
sky	84.0	83.7	66.6	86.3	70.5	70.0	89.3
person	47.0	41.3	53.4	57.9	60.0	65.2	64.4
rider	33.9	40.8	52.7	50.9	58.6	61.7	55.8
car	68.2	69.6	82.1	82.4	84.0	82.1	83.5
truck	19.0	21.5	52.7	52.2	75.5	64.2	84.3
bus	0.3	0.0	9.5	1.3	11.2	18.5	19.8
train	66.4	63.5	89.3	83.8	90.5	91.3	92.0
motorcycle	38.3	32.1	50.5	43.9	51.6	52.6	56.7
bicycle	23.6	19.4	38.5	29.8	40.9	44.0	43.7
mIoU	44.3	46.4	53.8	54.2	55.9	60.2	64.3

为了评估本文方法的泛化性, 在 DarkZurich-val、BDD100k-night 和 Nighttime Driving 等 3 个数据集上进行比较, mIoU 分别为 48.8%、40.0%、55.6%。其中, 在 DarkZurich-val 数据集上, 所提模型相较于最优的比较模型 mIoU 提升了 1.3%; 在 BDD100k-night 数据集上实现了次优结果, 相较于 Refign-HRDA 模型下降了 0.6%^[20]; 在 Nighttime Driving 数据集上, 相较于 MIC-HRDA 下降了 3.1%, 具体结果如表 3 所示。

Table 3 Experimental results of different models on DarkZurich-val, BDD100k-night, and Nighttime Driving**表 3** 不同模型在 DarkZurich-val、BDD100k-night 和 Nighttime Driving 上的实验结果 (%)

Method	DarkZurich-val	BDD100k-night	Nighttime Driving
DANNet-PSPNet	36.8	—	47.7
SePiCo-DAFormer	37.1	36.9	57.1
HRDA	42.1	39.1	55.6
MIC-HRDA	43.6	38.0	58.7
Refign-HRDA	<u>47.5</u>	<u>40.6</u>	<u>57.8</u>
Ours	<u>50.3</u>	<u>40.0</u>	55.6

2.5 消融试验

为了验证所提模型各模块对夜间场景语义分割的有效性,保持相同的实验环境,在 DarkZurich-test 数据集上对各模块进行消融实验,结果如表 4 所示。具体为,本文选用文献[13]的方法作为基准模型 Baseline,在 Baseline 上融合伪标签修正 Refine、MIC 中的掩码方法及本文掩码策略 MR 后进行性能比较^[19,20]。

由此可知,Baseline 在 DarkZurich-test 数据集上的 mIoU 为 55.90%,加入伪标签修正 Refine 后 mIoU 为

61.95%。为了验证 MR 掩码重建模块的有效性,在 Baseline+Refine 的基础上使用 MR 掩码重建模块后 mIoU 为 64.29%,相较于使用 MIC 中的掩码策略提升了 0.64%,证明了 MR 掩码重建模块的优越性。

Table 4 Results of ablation experiments on DarkZurich-test**表 4** 在 DarkZurich-test 上的消融实验结果 (%)

Method	mIoU	increase value
Baseline	55.90	—
Baseline+Refine	61.95	6.05
Baseline+Refine+MIC	63.65	7.75
Baseline+Refine+MR(Ours)	64.29	8.39

2.6 可视化结果

图 3 展示了在 Dark Zurich-val、NighttimeDriving 和 BDD100k-night 等 3 个数据集上选取 9 个样本的分割结果,前 3 行为 Dark Zurich-val 数据集的 3 个样本,中间 3 行为 NighttimeDriving 数据集的 3 个样本,最后 3 行为 BDD100k-night 数据集的 3 个样本。图 3 从左至右依次为原始图像、真实标签、DANNet、HRDA、MIC 及本文方法的可视化预测结果。

**Fig. 3** Visualization results of different models on DarkZurich-val, NighttimeDriving, and BDD100k-night datasets**图 3** 不同模型在 DarkZurich-val、NighttimeDriving 和 BDD100k-night 数据集上的可视化结果

由此可见,本文方法在大类别的边界上表现更规整,建筑类别、天空类别的边界混淆情况有所缓解,模型错误

地将天空预测为植被类别的情况有所减少。同时,模型对小类别的识别更敏感,忽略小物体类别的情况有所减少,

例如在预测分割中能准确识别电线杆、交通标志、骑手、摩托车等小物体类别。

3 结语

本文针对夜间场景中大规模边界不清晰和小物体类别易丢失的问题,提出基于掩码重建的域自适应夜间语义分割模型。首先,引入大规模局部边界掩码与小物体类别全局掩码策略;其次,语义分割模型在重建过程中以不同类别间的上下文信息为线索,促使模型在夜间场景中提取更具鲁棒性的信息,有效提升了所提模型在夜间场景中的语义分割能力。

在4个数据集的基准测试发现,所提模型在mIoU指标和可视化方面表现优秀,在夜间场景下的语义分割效果更优,仅在夜间场景复杂的天气情况下难以满足实际需求,例如雨、雪、雾等恶劣天气。未来,将考虑引入风格迁移的方法,使模型能适用于更复杂条件下的多目标域场景。

参考文献:

- [1] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2012: 3354–3361.
- [2] CHEN X, WILLIAMS B M, VALLABHANENI S R, et al. Learning active contour models for medical image segmentation [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2019: 11624–11632.
- [3] SAKARIDIS C, DAI D, GOOL L V. Guided curriculum model adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation [C]// IEEE International Conference on Computer Vision, 2019: 7373–7382.
- [4] GONG R, LI W, CHEN Y, et al. DLOW: domain flow for adaptation and generalization [J]. International Journal of Computer Vision, 2021, 128: 2865–2888.
- [5] TSAI Y, HUNG W, SCHULTER S, et al. Learning to adapt structured output space for semantic segmentation [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7472–7481.
- [6] TRANHEDEN W, OLSSON V, PINTO J, et al. DACS: domain adaptation via cross-domain mixed sampling [C]// IEEE Winter Conference on Applications of Computer Vision, 2021: 1379–1389.
- [7] WANG Q, DAI D, HOYER L, et al. Domain adaptive semantic segmentation with self-supervised depth estimation [C]// IEEE International Conference on Computer Vision, 2021: 8495–8505.
- [8] WU X, WU Z, GUO H, et al. DANNet: a one-stage domain adaptation network for unsupervised nighttime semantic segmentation [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2021: 15764–15773.
- [9] YANG G, ZHONG Z, TANG H, et al. Bi-Mix: bidirectional mixing for domain adaptive nighttime semantic segmentation [DB/OL]. <https://arxiv.org/pdf/2111.10339.pdf>.
- [10] LEE H, HAN C, YOO J, et al. GPS-GLASS: learning nighttime semantic segmentation using daytime video and gps data [C]// IEEE International Conference on Computer Vision Workshops, 2023: 4003–4012.
- [11] HOYER L, DAI D, GOOL L V. DAFormer: improving network architectures and training strategies for domain-adaptive semantic segmentation [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2022: 9914–9925.
- [12] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: transformers for image recognition at scale [DB/OL]. <https://arxiv.org/abs/2010.11929>.
- [13] HOYER L, DAI D, GOOL L V. HRDA: context-aware high-resolution domain-adaptive semantic segmentation [C]// European Conference on Computer Vision, 2022: 372–391.
- [14] YU F, CHEN H, WANG X, et al. BDD100K: a diverse driving dataset for heterogeneous multitask learning [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2020: 2636–2645.
- [15] DAI D, GOOL L V. Dark model adaptation: semantic image segmentation from daytime to nighttime [C]// International Conference on Intelligent Transportation Systems, 2018: 3819–3824.
- [16] CORDTS M, OMRAN M, RAMOS S, et al. The cityscapes dataset for semantic urban scene understanding [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2016: 3213–3223.
- [17] LOSHCHILOV I, HUTTER F. Decoupled weight decay regularization [DB/OL]. <https://arxiv.org/abs/2010.11929>.
- [18] XIE B, LI S, LI M, et al. SePiCo: semantic-guided pixel contrast for domain adaptive semantic segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(7): 9004–9021.
- [19] HOYER L, DAI D, WANG H, et al. MIC: masked image consistency for context-enhanced domain adaptation [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2023: 11721–11732.
- [20] BRUGGEMANN D, SAKARIDIS C, TRUONG P, et al. Refign: align and refine for adaptation of semantic segmentation to adverse conditions [C]// IEEE Winter Conference on Applications of Computer Vision, 2023: 3173–3183.

(责任编辑:刘嘉文)