

DOI:10.19651/j.cnki.emt.2518290

基于特征分支增强的室外场景图像语义分割^{*}

孙豪辰¹ 安毅^{1,2} 张永康¹ 熊攀¹

(1.新疆大学电气工程学院 乌鲁木齐 830017; 2.大连理工大学控制科学与工程学院 大连 116024)

摘要:语义分割是自动驾驶的关键技术。室外场景图像语义分割受环境复杂、样本不均衡等问题影响,致使分割效果不佳。针对上述问题,本文提出了一种基于特征分支增强的室外场景语义分割网络FBE-Net。FBE-Net采用编码器-解码器结构,专门设计了特征增强分支,使用多尺度膨胀注意力使模型聚焦关键特征,提高整体精度,使用记忆力模块改善样本不均衡的问题,同时兼顾轻量化设计。本文使用高清相机采集校园场景数据,并加注真实语义标签,制作校园场景语义分割数据集。实验在Cityscapes数据集和自建数据集上进行,实验结果表明,FBE-Net在Cityscapes数据集上mIoU达到了79.64%,在自制数据集上mIoU达到了78.01%,优于主流语义分割方法。

关键词:语义分割;分支增强;样本不均衡;记忆力模块

中图分类号:TP391.4; TN911.73 文献标识码:A 国家标准学科分类代码:510.4050

Outdoor scenes image semantic segmentation based on feature branch enhancement

Sun Haochen¹ An Yi^{1,2} Zhang Yongkang¹ Xiong Pan¹

(1. School of Electrical Engineering, Xinjiang University, Urumqi 830017, China;

2. School of Control Science and Engineering, Dalian University of Technology, Dalian 116024, China)

Abstract: Semantic segmentation is a key technology in autonomous driving. Outdoor scene image semantic segmentation faces challenges like environmental complexity and sample imbalance, leading to suboptimal performance. To address these issues, this paper proposes a semantic segmentation network for outdoor scenes based on feature branch enhancement, FBE-Net. FBE-Net adopts an encoder-decoder architecture and designs a feature enhancement branch. It utilizes multi-scale dilated attention to capture key features and enhance overall accuracy, and employs a memory module to address sample imbalance. Simultaneously considering lightweight design. We collected campus scene data using an HD camera, annotated it with semantic labels, and created a campus scene semantic segmentation dataset. Experiments were conducted on the Cityscapes dataset and the self-built dataset. The experimental results showed that FBE-Net achieved a mIoU of 79.64% on the Cityscapes dataset and 78.01% on the self-made dataset, outperforming mainstream semantic segmentation methods.

Keywords: semantic segmentation; branch enhancement; sample imbalance; memory model

0 引言

语义分割作为计算机视觉的重要任务,旨在对图像中每个像素分配特定的类别标签,实现精细化的场景理解。在室外场景中,由于背景包含建筑物、车辆、树木和行人等多种元素,这些元素可能与目标物体(如交通标志)在颜色或形状上相似,且存在遮挡现象^[1],使得分割难度较大。同时,数据集中存在样本不均衡的问题^[2]:尺度不均衡,比如

自行车和标志牌,其尺寸比较小,相应的,其像素在数据集中占比也比较低,容易被背景淹没,其关键特征在学习训练过程中容易被遗忘,且缺乏足够的外观信息,导致其与背景或相似目标难以分割;另一方面是数量不均衡,比如某些物体在数据集中很少出现,导致模型学习时忽视该少数类样本,或者在训练过程中出现特征遗忘,从而影响分割精度。

现研究者对于复杂场景的语义分割多使用多尺度特征提取^[3],构建特征金字塔^[4],增强上下文信息感知^[5],加强

收稿日期:2025-03-09

*基金项目:辽宁省自然科学基金计划项目(2023-MS-093)、国家自然科学基金(62173055)、山西省科技重大专项揭榜项目(20191101014)、新疆维吾尔自治区重大科技专项(2023A01005-1)资助

边缘信息提取^[6-7],优化损失函数^[8],数据增强^[9]以及使用注意力机制^[10]等方法加强对小尺度物体的特征提取,从而提升整体分割精度。

PSPNet 使用金字塔池化模块实现全局与局部信息结合^[11],提高了模型在复杂场景中分割的鲁棒性。DeepLabv3+ 使用空洞空间金字塔池化模块有效扩大感受野^[12],同时捕捉多尺度上下文信息,提高分割精度。DANet 使用双重注意力机制捕获全局和局部信息^[13],进行多尺度特征融合,提高整体精度。这些经典方法都有着不错的分割效果,但是忽略了样本不均衡的问题,通过多尺度特征提取信息和增强上下文信息感知,整体提升精度,对于小尺度物体和样本类别少的现象并没有进行相应的处理。此外,这些模型参数量和计算量都较大,不适用于网络模型轻量化的发展趋势。

Guo 等^[14]设计了 SegNeXt 网络,使用多尺度卷积注意力网络加强特征提取,但是对于物体边界的分割效果仍有不足。Wang 等^[15]提出了 BC-SegNeXt 网络,该网络增加了边界模块和多损失功能函数,加强了对边界的分割效果,但在更复杂的场景下如遮挡等,其边界分割精度仍有误差。Cui 等^[16]提出了 D-SegNeXt 网络,并设计了一种注意力单元,再通过混合扩张卷积扩大感受野,增强了多尺度特征的提取能力,提高了对狭窄边界的分割效果,但混合扩张卷积会导致局部细节丢失,影响分割精度。Cheng 等^[17]提出了 SF-SegNeXt 网络,网络中引入了额外的前景解码器和损失函数,并加入通道注意力机制增强特征提取能力,提高了边界分割效果,但引入的前景解码器是针对特定场景,其泛化能力有限。此外,这些网络都忽视了现实场景中样本数量不均衡和尺度不均衡的问题,而样本数量少、尺度小的物体,其关键特征在学习训练过程中容易被遗忘,致使语义分割精度不高。

徐晓龙等^[18]提出了 UA-Net,网络使用联合注意力模块进行多尺度特征增强,提升小尺度物体的权重,提高整体分割精度,而小尺度物体权重增强可能与背景或噪声抑制产生冲突,反而会引入噪声或导致边缘模糊。艾青林等^[19]提出了 MFF-STDC 网络,网络通过引入多种注意力模块增强上下文信息融合,并通过分层权重注意力优化,改善对小尺度物体的分割效果。邵玉文等^[20]提出了 DLA-Net,引入交叉注意力机制增强对全局上下文的感知,提高整体精度。这些方法本质上仍是对小尺度物体的特征权重动态调整,在逐层传播中,浅层细节因反复加权反而逐渐丢失。吴炳剑等^[21]结合生成对抗网络与注意力机制,提出了 GANA 网络,提高了小尺度物体的分割精度,但面对少样本物体时易导致生成器特征遗忘。朱磊等^[22]设计了融合多维特征的 MDFNet,并引入特征金字塔,使用双解码头进行细节恢复,提高了分割精度,但是参数量过大,易出现小尺度特征对齐误差,造成信息失真。对小尺度物体和少样本物体的分割,不仅在于特征选择,更在于对其特征的维护,通过记

忆其原本特征,实现更精准的分割。

针对背景复杂和样本不均衡导致语义分割精度低的问题,本文提出了一种基于特征分支增强的室外场景语义分割网络(semantic segmentation network for outdoor scenes based on feature branch enhancement, FBE-Net),网络使用编码器解码器结构,以 SegNeXt-Small 为基础网络,设计了多尺度膨胀注意力模块和记忆力模块,并采用了轻量化设计。本文贡献如下:

- 1) 针对室外场景数据集缺失的问题,构建了 Cityscapes 格式的校园场景数据集(campus scenery dataset, CSD),有效验证了模型的性能;
- 2) 针对室外场景中背景复杂导致分割困难、边界分割效果差的问题,设计了一条特征增强分支,设计了多尺度膨胀注意力(multi-scale dilated attention, MSDA)模块进行多尺度特征提取,增加局部上下文的感知,使用空间注意力机制和通道注意力机制突出重要特征,改善了边界分割效果;
- 3) 针对室外场景中存在样本尺度不均衡和样本数量不均衡导致语义分割精度低的问题,设计了记忆力模块储存特征信息以增强小尺度和少样本数据特征的表达能力,从而解决特征遗忘导致分割精度低的问题。

1 基于特征分支增强的室外场景图像语义分割网络

1.1 FBE-Net 网络框架

FBE-Net 的网络框架如图 1 所示,网络使用的是编码器解码器结构。在图 1 中,DWSConv 是深度可分离卷积(depthwise separable convolution, DWSConv),rate 表示膨胀率,CAM 是通道注意力机制(channel attention module, CAM)。FBE-Net 的输入是 $2048 \times 1024 \times 3$ 的图片,其中 3 表示像素信息(R, G, B)。

首先对输入图片进行处理作为编码器输入,编码器由 4 个下采样阶段和分支构成,4 个下采样阶段使用多尺度卷积注意力模块(multi-scale convolutional attention module, MSCA),图 1 中一系列模块堆叠起来构成多尺度卷积注意力网络(multi-scale convolutional attention network, MSCAN),多个 MSCAN 组成一个阶段,用于提取多尺度特征,MSCAN 和 MSCA 结构如图 2 所示。分支结构对 Stage1 的输出进行处理,使用多尺度膨胀注意力模块提取浅层的多尺度特征;将 Stage2、3、4 和分支输出的特征分别进行上采样后沿通道维度拼接,再使用记忆力模块增强特征表示,作为解码器头的输入,最终输出 19 个类别的预测图。

图 2(a)中全连接网络(feed-forward network, FFN)用于处理局部非线性变换,非线性激活函数(gaussian error linear unit, GELU)用于提升特征提取能力,批量归一化(batch normalization, BN)用于加速神经网络的训练和提

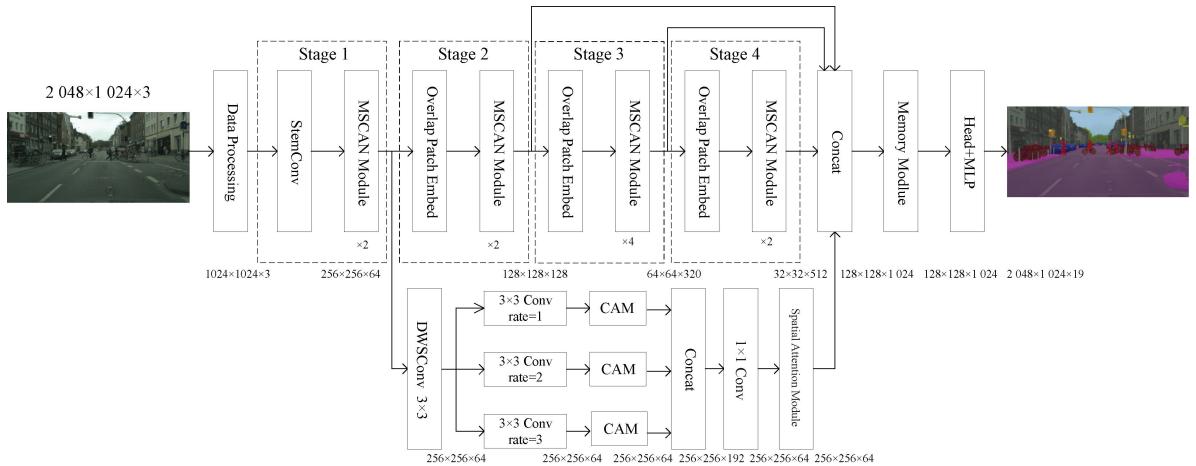


图 1 FBE-Net 网络模型架构

Fig. 1 Network model architecture of FBE-Net

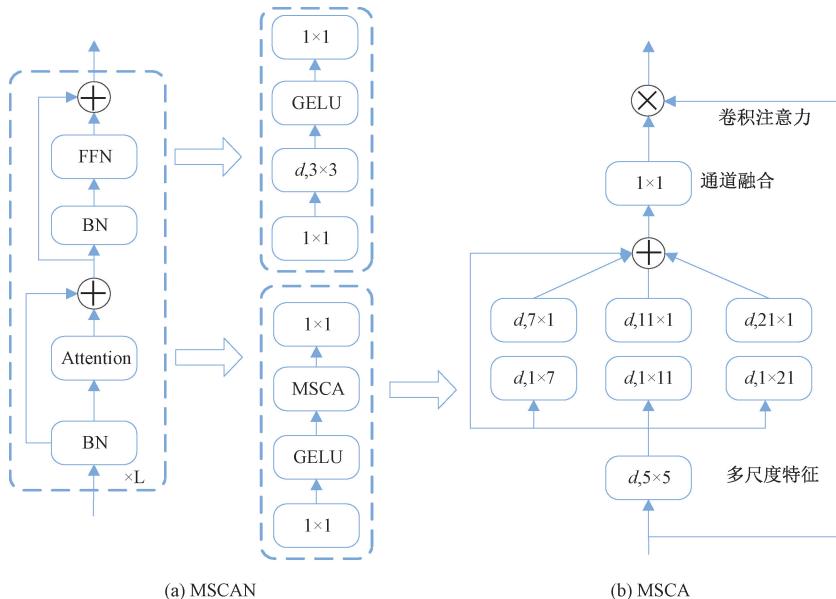


图 2 MSCAN 结构

Fig. 2 MSCAN structure

高模型的稳定性, d 表示深度卷积, \oplus 表示逐元素相加; 在图 2(b)中, MSCA 使用了条状卷积 $k \times 1$ 和 $1 \times k$ 来代替标准的深度卷积 $k \times k$, \otimes 表示逐元素相乘。

从图 2(b)可以看到, MSCA 首先对输出特征使用深度卷积提取局部特征, 经过多分支深度卷积进行多尺度上下文特征提取, 再对提取到的特征进行融合, 之后使用 1×1 卷积对不同通道间的相关性进行建模。 1×1 卷积输出的结果被用作注意力权重参数, 给 MSCA 的输入分配不同权重, 从而输出结果。MSCA 可以使用公式进行表示, 如式(1)所示。

$$\text{Att} = \text{Conv}_{1 \times 1} \left(\sum_{i=0}^3 \text{Scale}_i (\text{DW} - \text{Conv}(\mathbf{F})) \right) \quad (1)$$

$$\text{Out} = \text{Att} \otimes \mathbf{F}$$

式中: \mathbf{F} 代表输入的特征, Att 表示为注意力的权重参数, Out 是输出的特征, DW-Conv 表示深度卷积操作, Conv 是卷积操作, \otimes 为逐元素相乘; 在分支中, 使用两个不同的条状卷积来模拟标准的深度卷积, 使用条状卷积一方面可以用来代替标准的深度卷积, 降低参数量; 另一方面, 在室外场景中会出现很多条状物体, 如路灯、道路边缘等, 条状卷积有助于线状物体的特征提取。

解码器用于处理编码器所输出的特征, 获取高层次语义信息。FBE-Net 的解码器头对后三个阶段和分支输出的特征和进行特征拼接, 再使用轻量级解码器 Hamburger 模型进行全局信息建模。这样一个强效编码器配合轻量级的解码器, 可以保证计算效率的前提下, 提高模型的分割精度。

1.2 特征增强分支结构

在第一阶段提取的浅层特征具有高分辨率,与深层特征结合可提升整体分割精度。通过多尺度融合,这种结合既保留了全局语义信息,又捕捉了局部细节,从而改善了小目标分割效果。同时,使用注意力强化浅层特征中的有效信息并抑制噪声与冗余部分,不仅提高了融合质量,还增强了对复杂边界和精细结构的捕捉能力。

由此提出特征增强分支结构:从第一阶段引出分支,先用 3×3 深度可分离卷积进行初步特征提取,然后使用MSDA模块进行多尺度特征提取,输出的特征沿通道维度进行拼接然后再用 1×1 卷积进行通道调整,经空间注意力处理后输出至解码器。在每层卷积后使用激活函数GELU,以增加模型的非线性表达能力,使用双线性插值法对输出的多尺度特征进行上采样,然后与第二、三、四阶段的特征进行拼接,再使用记忆力模块增强特征表示,之后

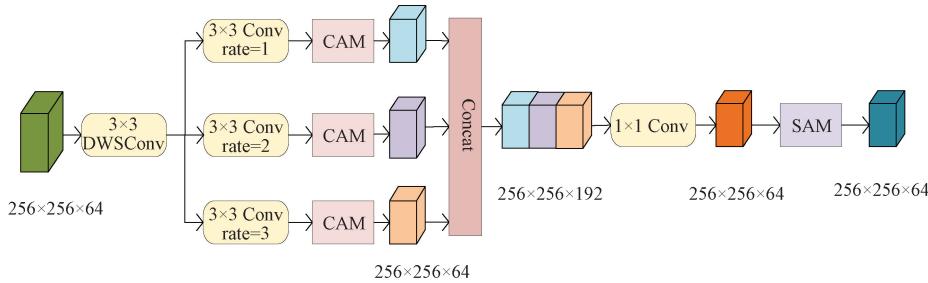


图 3 MSDA 模块

Fig. 3 MSDA module

多尺度膨胀卷积产生的特征可能包含大量冗余信息,在每个膨胀卷积分支输出后加入通道注意力模块,可以为不同感受野的特征分配合适的权重,避免不同分支的相互影响,导致某些重要特征被抑制,同时抑制冗余特征,为后续模块减少冗余信息的影响。空间注意力模块使网络聚焦关键位置,优化模型对局部和全局特征的空间分布,放置于多尺度特征融合之后,可以更有效地筛选和强化有用信息,避免对单一特征过度强化,保持特征表达的整体性。

CAM 旨在为每个通道分配一个权重,调整不同通道特征的重要性,该模块如图 4 所示,⊕表示逐元素相加,⊗表示逐通道相乘。它对输入特征 F 进行全局平均池化和全局最大池化,之后生成两个描述向量,将它们送入多层次感知机(multi-layer perceptron, MLP)中学习,然后再对生成的特征叠加并通过 Sigmoid 激活函数得到权重系数,将该权重与输入特征逐通道相乘,得到最终的通道注意力特征图。通道注意力权重 M_c 如式(2)所示。

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \quad (2)$$

式中:MaxPool 是最大池化操作,AvgPool 是平均池化操作,MLP 是全连接层, σ 是 Sigmoid 激活函数。

而 SAM 旨在为每个空间位置分配权重,调整特征图中不同位置的关注度,该模块如图 5 所示,⊗表示逐元素

将处理过的特征送入解码器处理。

1.3 多尺度膨胀注意力

多尺度膨胀注意力模块MSDA如图3所示,该模块用于提取多尺度特征,增强上下文感知,从而提高整体分割精度。它由3个基本单元组成:多尺度膨胀卷积、通道注意力模块(channel attention module, CAM)和空间注意力(spatial attention module, SAM)模块。多尺度膨胀卷积通过设置不同的膨胀率可以有效扩大卷积核的感受野,捕获从局部到全局的多尺度特征,膨胀率小的卷积关注局部细节,可以提高对物体边缘的分割效果,而膨胀率大的卷积关注全局语义信息。不同膨胀率可以从局部到全局构建出多尺度特征金字塔,适应不同尺度下的目标特征,有效应对目标尺度差异大、复杂背景干扰强的场景。相比直接增大卷积核尺寸,多尺度膨胀卷积无需额外增加参数,即可提升感受野,更适用于语义分割任务。

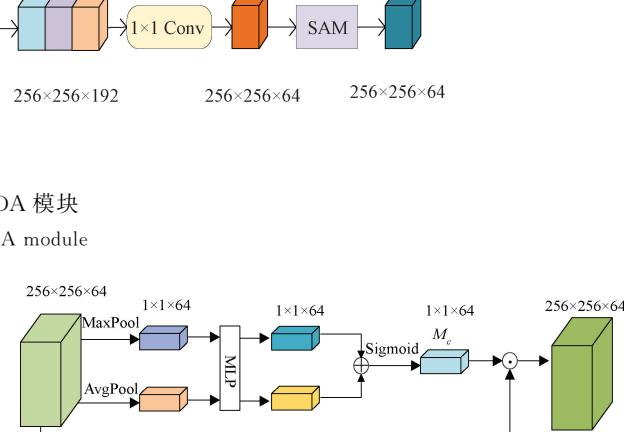


图 4 通道注意力模块

Fig. 4 Channel attention module

相乘。它对特征分别进行全局平均池化和全局最大池化,在通道维度进行压缩,得到两个单通道的特征,然后再将这两个特征沿通道维度进行拼接,对拼接后的特征图通过一个 7×7 的卷积操作生成单通道空间权重图,最后将它与输入特征逐位置相乘得到增强后的特征图。空间注意力权重 M_s 如式(3)所示。

$$M_s(F) = \sigma(f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])) \quad (3)$$

式中: $f^{7 \times 7}$ 表示 7×7 的卷积操作。

1.4 记忆力模块

在室外场景中,物体要素种类繁多且分布复杂,某些要素可能在特定场景中频繁出现,而在其他场景中则较为罕见,在数据采集过程中难以避免样本不均衡问题的出现。样本不均衡会显著影响模型的学习能力与特征提取能力,从而削弱分割效果。此外,室外场景中,如垃圾桶、

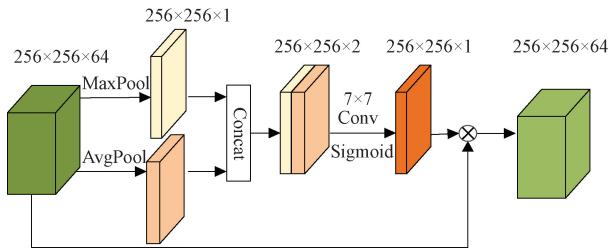


图 5 空间注意力模块
Fig. 5 Spatial attention module

标志牌等小尺寸物体,所占像素很少,分割困难。

因此,本文设计了记忆力模块,通过存储历史信息,维护其关键特征,能够有效增强模型对场景语义的理解。尤

其是对于小尺度物体分布零散且容易被忽视的情况,记忆力模块保留关键区域的历史特征,有助于提升小尺度物体和遮挡区域的分割准确度,改善对少样本物体的分割效果。复用上下文信息,减少了模型在训练过程中的误差传播,使模型能够更快达到最优状态,还可以减少重复计算,提升推理效率和训练速度。

记忆力模块包含记忆存储单元、记忆读写、记忆检索单元 3 个部分,该模块如图 6 所示。记忆存储单元的核心是一个动态存储结构 $M \in R^{N_a \times D}$,其中 N_a 是存储单元的数量,也就是需要分类的类别数量,D 是每个存储单元的特征维度,也就是通道数。在训练中,记忆存储单元会不断记录和更新分割目标图像的关键语义特征。

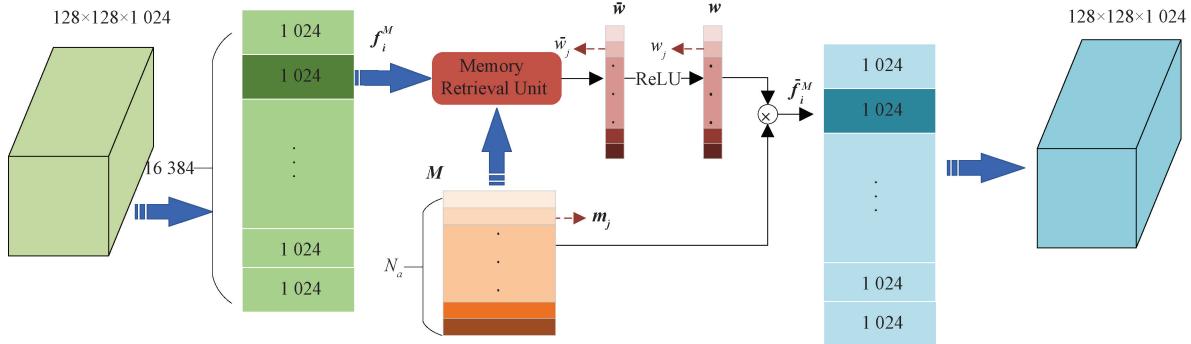


图 6 记忆力模块
Fig. 6 Memory module

首先获取编码特征, f_i^M 是编码器特征的一部分,为方便计算,将编码特征(H, W, C)转换为(N, C), N 表示像素点数。在记忆存储单元存储全局上下文信息,每个记忆单元 m_j 是一个特征向量。记忆检索是通过计算输入特征与记忆存储单元特征的相似度进行,可以用式(4)表示。

$$a(f_i^M, m_j) = \frac{f_i^M m_j^\top}{\|f_i^M\| \|m_j\|} \quad (4)$$

然后获取检索权重 \bar{w}_j , 表示每个记忆单元的重要性,如式(5)所示。

$$\bar{w}_j = \frac{\exp(a(f_i^M, m_j))}{\sum_{j=1}^N \exp(a(f_i^M, m_j))} \quad (5)$$

式中: $a(f_i^M, m_j)$ 是用于计算编码特征 f_i^M 和记忆单元 m_j 相似度的函数。

为了增强稀疏性和降低无关因素的影响,进一步对权重进行稀疏化操作:

$$\tilde{w}_j = \frac{\max(\bar{w}_j - \beta_a, 0) \bar{w}_j}{|\bar{w}_j - \beta_a| + \epsilon_a} \quad (6)$$

式中: $\max(\bar{w}_j - \beta_a, 0)$ 使用的是 ReLU 函数作为硬收缩操作来提高 \bar{w}_j 的稀疏性, β_a 为收缩阈值, ϵ_a 为非常小的正标量。

\tilde{w}_j 经过重新归一化处理后得到记忆检索函数:

$$w_j = \frac{\tilde{w}_j}{\sum_{j=1}^N |\tilde{w}_j|} \quad (7)$$

经过上述操作,从记忆力模块中检索到的对应特征 \bar{f}_i^M 为:

$$\bar{f}_i^M = \sum_{j=1}^N w_j m_j \quad (8)$$

再将该特征重新变回原本(H, W, C)的形状。

记忆力模块可以保证网络在训练过程中有效记忆和存储样本量少的物体特征,在训练中,通过记忆存储单元中存储的特征信息,避免因样本量少模型遗忘物体特征的问题。使用记忆力模块,增强了对小尺度物体和遮挡物体的特征建模能力,同时保留局部特征,提高了对少样本和小尺度物体分割的准确性,从而提高整体语义分割的精确度。

2 实验结果与分析

2.1 数据集

实验使用了两个数据集来评估网络性能,一个是公开的 Cityscapes 数据集,一个是使用高清相机在校园室外场景采集彩色图像得到,然后经人工标注生成的 Cityscapes 格式的校园场景数据集 CSD,两者均涵盖了复杂的城市场景和样本不均衡的物体,具有很高的挑战性和代表性。

Cityscapes 是专注于城市环境下的标准语义分割数据集,包含 5 000 张高质量像素级标注的图像(其中训练集 2 975 张,验证集 500 张,测试集 1 525 张),图像分辨率为 2048×1024 。涵盖 19 个语义类别,包括建筑、道路、行人等,数据来源于 50 个不同城市的街景拍摄,具有多样性和复杂性。

CSD 数据集是针对室外场景的复杂性和样本不均衡的需求构建的,该数据集覆盖了校园道路、停车场等多种场



图 7 CSD 数据集

Fig. 7 CSD dataset

2.2 实验环境及参数

本文实验在 UBUNTU18.04、CUDA10.1、PyTorch1.5.1、Python3.8 环境下,使用双卡 RTX3090(24 GB)-GPU 进行训练与测试,所有网络均使用基于迭代的训练模式,最大迭代次数是 160 000 次,使用初始学习率为 0.001 的 Adam 优化器,采用多项式衰减策略,使用交叉熵损失函数。

2.3 评价指标

为了全面评价网络模型在语义分割任务中的性能,本文采用了语义分割中 3 种常用的指标:平均交并比(mean intersection over union, mIoU),它是语义分割的核心指标,综合衡量每个类别的分割质量;全像素准确率(overall accuracy, aAcc),反映所有像素的总体分类准确率,平均像素准确率(mean accuracy, mAcc),评估每个类别的分类准确性,尤其适用于样本不平衡的数据集。各评价指标如式(9)~(11)所示。

$$\text{mIoU} = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}} \quad (9)$$

景。初始数据共采集了 328 张校园场景,如图 7(a)所示,图片分辨率为 1920×1080 ,经过标注后,如图 7(b)所示,使用数据增强扩充到了 2 400 张,其中训练集 1 400 张,验证集 400 张,测试集 600 张。数据集一共有 13 个类别,类别分布与 Cityscapes 数据集相似,但增加了垃圾桶、围栏和导航指示牌等小尺度和少样本类别。相较于 Cityscapes,自建数据集具有更多遮挡场景和小尺度物体以及样本不均衡的比例,能有效评估模型在实际应用场景中的性能。

$$\text{aAcc} = \frac{\sum_{i=0}^{k-1} p_{ii}}{\sum_{i=0}^{k-1} \sum_{j=0}^{k-1} p_{ij}} \quad (10)$$

$$\text{mAcc} = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij}} \quad (11)$$

式中: k 表示有效类别数, i 表示真实类别, j 表示预测类别, p_{ii} 表示类别预测正确的像素数, p_{ij} 表示假负的数量, p_{ji} 表示假真的数量,各指标值均为越高表示分割效果越佳。

2.4 对比实验

为了充分评估 FBE-Net 的性能,本文使用对比的方法进行验证,为了保证实验的有效性,分别在 Cityscapes 数据集和 CSD 数据集上进行训练与测试,主干网络均为 ResNet50。

为保证实验效果,所有算法相关参数与其他硬件条件均相同,并选取了小尺度物体和少样本的标签结果进行对比分析,具体训练结果如表 1 所示。

表 1 Cityscapes 数据集语义分割结果

Table 1 Semantic segmentation results for Cityscapes dataset

%

网络模型	mIoU	aAcc	mAcc	IoU					
				Traffic light	Traffic sign	Rider	Bus	Train	Motorcycle
PSPNet	71.27	84.81	81.83	56.09	59.68	50.42	65.52	63.11	60.45
DeepLabv3	74.25	84.02	82.22	59.28	64.66	56.11	66.78	70.72	62.11
DANet	75.48	89.71	83.13	60.32	65.30	60.92	70.68	73.94	62.22
DeepLabv3+	75.72	85.63	83.01	60.53	62.41	58.63	68.03	77.43	62.18
SegNeXt	77.79	95.98	85.41	69.74	78.51	62.74	88.79	79.99	65.94
FBE-Net (Ours)	79.64	96.14	86.57	71.44	80.15	66.89	90.5	83.78	68.84

由表 1 的实验结果表明, FBE-Net 的 mIoU 达到 79.64%, 与 SegNeXt 网络相比提高了 1.85%, 在全像素准确率和平均像素准确率中均有提升, 并在 3 项评价指标中均取得最高成绩。FBE-Net 结合记忆力模块与 MSDA 模块, 有效保留了细节特征和全局特征, 同时增强了对小尺度物体、少样本物体和边界区域的关注, 例如“Traffic light”、“Traffic sign”等标签, FBE-Net 因使用记忆力模块, 增强了对小尺度物体的特征表示, 在进行分割时更加精准。

为更加直观的展示分割效果, 将上述网络的分割结果进行可视化, 从而更好的观察分割效果, 具体如图 8 所示。

图 8 中标签及对应颜色如下: 人是红色, 汽车是蓝色, 自行车是深红色, 标志牌是黄色, 草地是绿色, 树木是深绿色, 建筑是深灰色, 路灯是浅灰色, 导航指示牌是天蓝色, 道路是粉色。图中未出现的标签及相应的颜色不再列出。此外, 两种数据集中, 相同标签物体使用的是同一种颜色。

从图 8 可以看出, 本文算法对于物体的边缘分割效果很好, 包括人、车的轮廓还有道路边缘分割十分精准。对于存在遮挡现象的标签, 也可以精准识别, 比如第 1 组中被挡住的人和第 4 组中被树和交通灯挡住的汽车, 在有遮挡时, 也可以提取其关键特征, 再与记忆力模块进行匹配, 从而精准分割。对于小尺度物体, 比如第 1 组、第 3 组和第 4 组中的标志牌, 还有第 4 组远方的行人, 也都可以精准识别并分割。相比于图中其他算法, 本文网络分割的完整性、连续性和效果均是最好最精准的。

表 2 CSD 数据集语义分割结果

Table 2 CSD dataset semantic segmentation results

%

网络模型	mIoU	aAcc	mAcc	IoU			
				Tree	Human	Road	Navigation
DANet	61.87	83.50	70.60	44.00	48.69	52.68	52.65
DeepLabv3	63.77	87.21	73.19	51.50	47.50	59.68	63.59
DeepLabv3+	63.81	88.06	73.68	57.33	46.23	60.81	66.22
PSPNet	69.17	90.31	76.61	57.39	64.78	76.44	71.95
SegNeXt	73.15	90.29	82.82	47.60	69.36	82.43	87.25
FBE-Net (Ours)	78.01	91.37	85.75	64.88	74.73	84.19	89.83

由表 2 的实验结果和图 9 可得, 与公开数据集相比, 各模型在 CSD 数据集上的性能均有所下降, 这可能是由于 CSD 数据集在图像质量、标注精度和类别分布等方面差异。而 FBE-Net 在 CSD 数据集上依然保持了最佳分割效果, 与 SegNeXt 网络相比, FBE-Net 的 mIoU 提升了 4.86%, 展现了性能优势。从类别标签看, FBE-Net 在多个标签上有明显提升。例如, 对于“人”和“树木”等在复杂背景的目标, 模型能够更精准地捕捉目标特征; 对于“道路”和“标志牌”等标签, 则体现出更强的多尺度特征提取能力。同时, 对于小尺度物体的识别更加精准, 在细节处

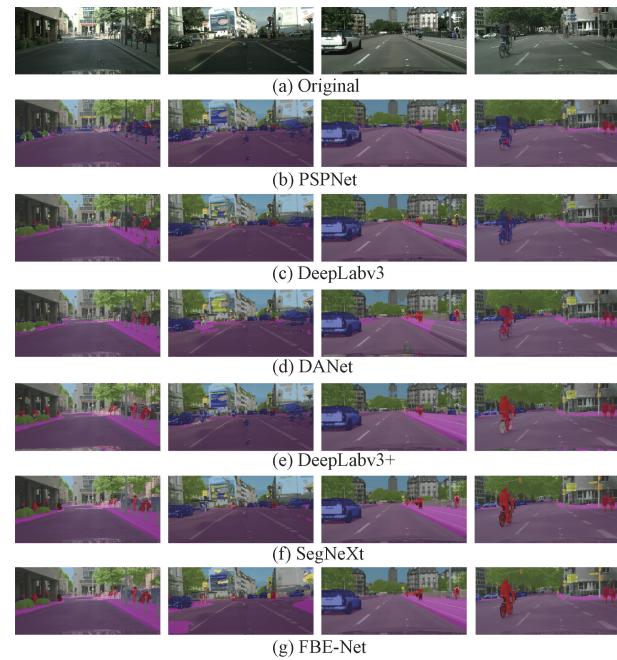


图 8 Cityscapes 数据集分割结果

Fig. 8 Cityscapes dataset segmentation results

为了充分评估 FBE-Net 的性能, 本文在 CSD 数据集上进行了相同的对比实验。在该数据集中存在更多样本数量不均衡和样本尺度不均衡的现象, 可以充分评估网络性能, 具体如表 2 所示, 分割可视化效果如图 9 所示。

理上更加高效, 对于边缘的分割更加精准。这些结果充分验证了 FBE-Net 在复杂场景下的鲁棒性和优秀的分割能力。

此外, 为了充分验证 FBE-Net 的性能, 本文又选取了 CSD 数据集中的 8 个场景进行预测, 预测结果如图 10 所示。

图 10 中, 对于路灯这种线状物体, FBE-Net 分割的效果很好, 识别准确, 对于其物体的边界分割也比较精准。而且, 对于道路边界和草地边界的分割也十分准确。此外, 每张图像的整体分割效果较好, 对于遮挡的物体也能

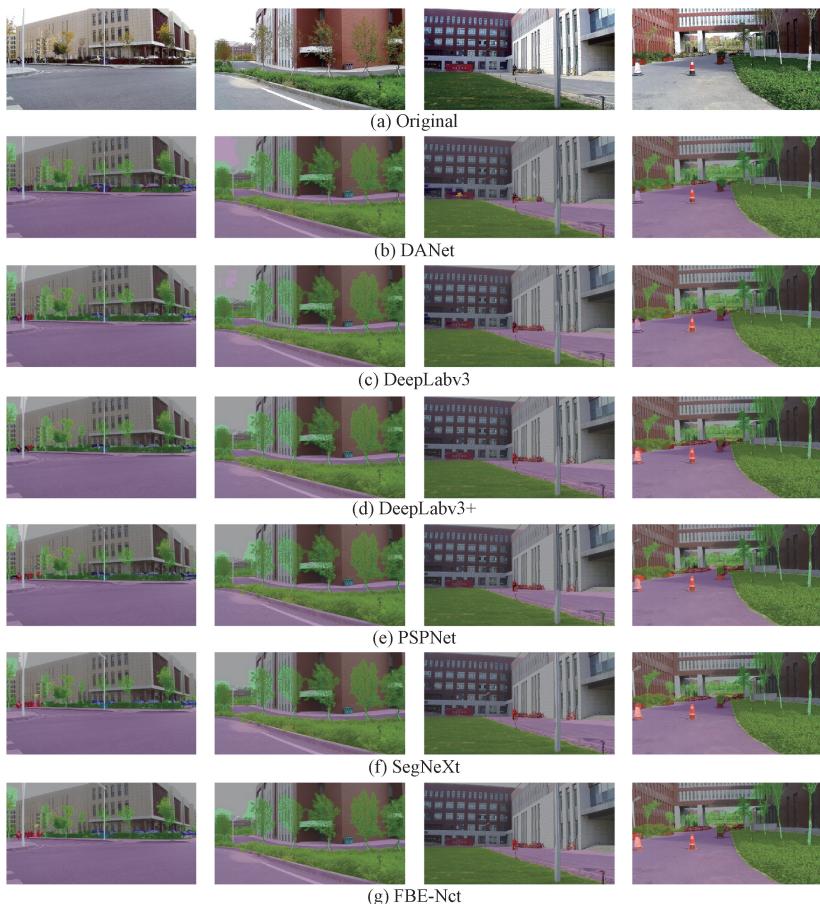


图 9 CSD 数据集分割结果

Fig. 9 CSD dataset segmentation results

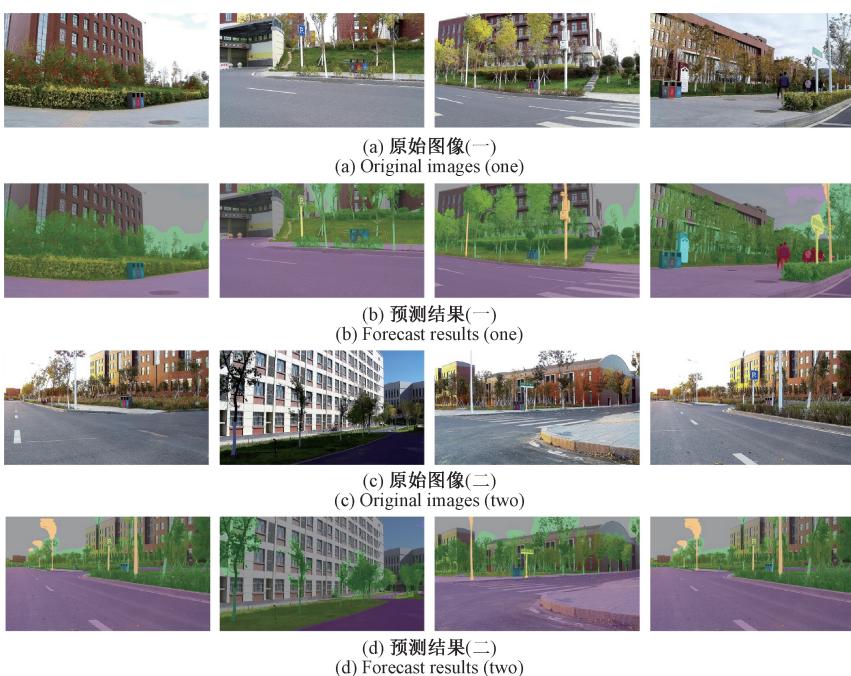


图 10 FBE-Net 分割结果

Fig. 10 FBE-Net segmentation results

精准识别。这些可视化的预测结果充分展现了 FBE-Net 优秀的分割性能。

2.5 网络效率分析

在实际应用中,语义分割模型需要兼顾精度和效率,特别是在资源受限的环境(如移动设备或实时场景)中。模型的参数量和计算量是衡量其资源消耗的重要指标,直接影响了模型的部署可行性和推理速度。而模型的参数量和计算量是衡量资源消耗的重要指标。为了验证模型的效率和性能,本文通过 FLOPs(计算量)和参数量(Params)对模型复杂度进行了详细的评估,如表 3 所示。

从表 3 可以看出,FBE-Net 因引入了一系列优化和模块,导致网络的参数量和计算量都有所提升,但是网络的整体复杂度低于 Base 模型,既提升了模型性能,又避免了计算资源的大幅消耗,适合实际场景中对模型规模和精度均有要求的任务。

2.6 与先进算法对比实验

此外,本文还与近年适用于室外场景语义分割的先进算法进行了精度对比和参数量以及计算效率的对比,充分

表 3 网络模型的计算效率和参数量

Table 3 Computational efficiency and number of parameters for network modeling

网络模型	FLOPs/G	Params/M
DeepLabv3	2159.70	68.11
DANet	1593.01	49.84
PSPNet	1427.4	48.97
DeepLabv3+	1410.60	43.59
SegNeXt-Base	256.48	27.57
SegNeXt-Small	122.72	13.90
FBE-Net (Ours)	189.10	18.92

验证本文算法的有效性和先进性。

这些算法针对尺度不均衡导致分割精度低的问题,对网络进行了各种优化,因此除了总体的分割效果,本文还选取了一部分小尺度物体标签进行对比,展示各个算法对于小尺度物体的分割优化效果,所用数据集均为 Cityscapes 数据集,具体如表 4 所示。

表 4 不同语义分割模型在 Cityscapes 对比

Table 4 Comparison results of different semantic segmentation models on Cityscapes

网络模型	mIoU/%	IoU/%				FLOPs/G	Params/M
		Traffic light	Traffic sign	Person	Bicycle		
UA-Net	70.10	63.00	68.50	80.10	67.80	—	—
GANA	73.40	—	—	—	—	—	24.45
DLA-Net	74.95	—	—	—	—	—	47.64
MFF-STDC	75.86	65.33	73.51	82.88	76.00	14.48	5.43
MDFNet	77.51	—	—	—	76.08	—	—
FBE-Net(Ours)	79.64	71.44	80.15	83.52	78.33	189.10	18.92

表 4 中,FBE-Net 参数量高于 MFF-STDC,但是低于其他算法,这是因为 MFF-STDC 在设计算法时使用了轻量化设计,而 FBE-Net 为了保证分割精度,使用了多种模块,因此引入了额外的参数。虽然 FBE-Net 参数量较高,但是整体分割精度是最高的,实现了对分割精度和网络效率的有效平衡。此外,相比于使用多种注意力机制加强对少样本物体和小尺度物体的特征提取,本文设计的记忆力模块更关注其特征本身,更证明了本文所提模块的有效行和本文算法的优越性。

通过与其他先进室外场景语义分割方法对比,可以看出本文算法整体精度均高于其他算法,且在小尺度物体上的分割效果,如行人、交通灯、交通标志和自行车等,其交并比均有比较明显的优势。更体现出,对于小尺度物体的识别和分割,除了加强对其特征的提取,还要关注其本身的特征,为后续小尺度物体分割提供了一个良好的解决思路。

2.7 消融实验

消融实验,通过逐步去除关键模块分析其对模型性能的影响。消融模块分别是记忆力模块(memory module)和MSDA 模块,通过逐步移除量化它们的贡献,并明确整体模型设计的优势。消融实验均在 Cityscapes 数据集上完成,为方便表述,使用 BaseNet 来代替基础网络,具体实验数据如表 5 所示。

表 5 消融实验

Table 5 Ablation experiment

%

网络模型	mIoU	aAcc	mAcc
BaseNet	77.79	95.98	85.41
BaseNet+Memory	78.09	96.08	85.05
BaseNet+MSDA	78.67	96.10	86.55
FBE-Net	79.64	96.14	86.57

如表 5 消融实验所示,记忆力模块和 MSDA 模块都对模型有积极作用。记忆力模块增强细节特征的记忆能力,有效增强了对尺度小物体和样本少物体的特征表示,而

MSDA 模块提取多尺度特征,并通过注意力模块使得模型聚焦于关键特征。两个模块结合更加强了对物体边缘信息和全局特征的提取,有效提高了边缘分割精度,进而提高了整体分割精度。

2.8 应用与分析

为了充分验证 FBE-Net 的实际性能,本文将 FBE-Net 部署到了智能机器人,设计了一个智能机器人室外场景语义分割系统。智能机器人如图 11 所示。



图 11 智能机器人

Fig. 11 Intelligent robot

该系统利用智能机器人的单目相机进行数据采集,然后对采集到室外场景图像数通过部署在智能机器人上的 FBE-Net 模型进行语义分割,并输出预测结果,预测结果如图 12 所示。



(a) 智能机器人拍摄图像
(a) Image captured by intelligent robot



图 12 系统预测结果

Fig. 12 Result predicted by the system

图 12 中,预测结果的整体分割效果较好,对于标志牌

的轮廓分割十分精细,边界分割效果很好,对于草地和道路的边界分割效果十分精准,对于远处的人也能准确识别,因此,模型性能整体满足预期,通过部署在智能机器人中,可以满足机器人对语义分割精度的需要。

在实际应用中,精准的语义分割结果对智能机器人的自动驾驶系统至关重要。后续将进一步优化网络架构以减少参数量,从而更好的满足实际工程中的任务需求。

3 结 论

在现实场景中,由于存在遮挡、目标尺度差异大以及小尺度物体等样本不均衡的问题,语义分割的效果难以达到预期,这对模型的性能提出了更高要求。为应对这些挑战,本文提出了一种基于特征分支增强的室外场景语义分割网络 FBE-Net,创新性地设计了多尺度膨胀注意力模块和记忆力模块,提升了网络分割性能并有效解决上述问题。同时,网络兼顾轻量化设计,拥有较低的参数量,在保持较低运算复杂度的同时,依然能够实现高精度分割。

多尺度膨胀注意力模块通过膨胀卷积提取多尺度特征,使模型更好地适应不同尺度的目标。记忆力模块通过保留目标特征,使模型在复杂背景或目标特征模糊的场景下能够更准确地识别并分割目标,同时改善了样本不均衡致使网络分割精度低的问题。

为了验证 FBE-Net 的有效性,本文分别在公开数据集 Cityscapes 和自建的 CSD 数据集上进行实验。实验结果表明, FBE-Net 显著提升了分割精度,相比其他方法在 mIoU 等评价指标和对小尺度物体以及样本不均衡方面均表现优越。此外,还通过实际工程应用,检测了 FBE-Net 的性能,为后续优化网络模型指明了方向。

参考文献

- [1] 林哲,潘慧琳,陈丹.融合改进 YOLO 和语义分割的遮挡目标抓取方法[J].电子测量与仪器学报,2024,38(12): 190-201.
LIN ZH, PAN H L, CHEN D. Grasp method for occlusion method by fusing improved YOLO with semantic segmentation[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38 (12): 190-201.
- [2] RAD M J, ARMENAKIS C. Active reinforcement learning for the semantic segmentation of urban images[J]. Canadian Journal of Remote Sensing, 2024, 50(1): 2374788.
- [3] KAR M K, NATH M K, NEOG D R. A review on progress in semantic image segmentation and its application to medical images [J]. SN Computer Science, 2021, 2(5): 397.
- [4] 周勇,刘泓滨,侯亚东.复杂城市交通场景下的自动驾驶语义分割方法[J].电子测量与仪器学报,2024,38(4): 241-247.
ZHOU Y, LIU H B, HOU Y D. Automatic driving semantic segmentation method for complex urban traffic

- scene [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(4): 241-247.
- [5] SUN C, AI Y B, WANG SH, et al. Mask-guided SSD for small-object detection [J]. Applied Intelligence, 2020, 51(6): 3311-3322.
- [6] 任莎莎, 刘琼. 小目标特征增强图像分割算法[J]. 电子学报, 2022, 50(8): 1894-1904.
- REN SH SH, LIU Q. Small target feature enhanced image segmentation algorithm [J]. Journal of Electronics, 2020, 50(8): 1894-1904.
- [7] LI X H, XIE L L, WANG C F, et al. Boundary-enhanced dual-stream network for semantic segmentation of high-resolution remote sensing images[J]. GIScience & Remote Sensing, 2024, 61(1), 2356355.
- [8] MA A L, WANG J J, ZHONG Y F, et al. FactSeg: Foreground activation-driven small object semantic segmentation in large-scale remote sensing imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60(1): 1-16.
- [9] 吴止缓, 高永明, 李磊, 等. 类别非均衡遥感图像语义分割的全卷积网络方法[J]. 光学学报, 2019, 39(4): 401-412.
- WU ZH H, GAO Y M, LI L, et al. Fully convolutional network method of semantic segmentation of class imbalance remote sensing images [J]. Acta Optica Sinica, 2019, 39(4): 401-412.
- [10] TSAI T H, TSENG Y W. BiSeNet V3: Bilateral segmentation network with coordinate attention for real-time semantic segmentation [J]. Neurocomputing, 2023, 532(1): 33-42.
- [11] ZHAO H S, SHI J P, QI X J, et al. Pyramid scene parsing network[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6230-6239.
- [12] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]. European Conference on Computer Vision, 2018: 801-818.
- [13] FU J, LIU J, TIAN H J, et al. Dual attention network for scene segmentation [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 3146-3154.
- [14] GUO M H, LU CH Z, HOU Q B, et al. SegNeXt: Rethinking convolutional attention design for semantic segmentation [J]. Advances in Neural Information Processing Systems, 2022, 35(1): 1140-1156.
- [15] WANG Z. Disaster remote sensing image semantic segmentation model with boundary constraints based on SegNeXt[C]. International Conference on Neural Networks, Information and Communication Engineering, IEEE, 2023: 734-737.
- [16] CUI P, MENG X, ZHANG W. Road extraction from high-resolution remote sensing images of open-pit mine using D-SegNeXt [J]. IEEE Geoscience and Remote Sensing Letters, 2024, 21(1): 1-5.
- [17] CHENG ZH X, YANG P L, ZUO Z CH, et al. SF-
- SegNeXt: A foreground-aware network based on SegNeXt for sea fog detection [C]. International Conference on Aerospace System Science and Engineering, 2024: 1-10.
- [18] 徐晓龙, 俞晓春, 何晓佳, 等. 基于改进 U-Net 的街景图像语义分割方法[J]. 电子测量技术, 2023, 46(9): 117-123.
- XU X L, YU X CH, HE X J, et al. Semantic segmentation method of street view image based on improved U-Net [J]. Electronic Measurement Technology, 2023, 46(9): 117-123.
- [19] 艾青林, 刘元宵, 杨佳豪. 基于 MFF-STDC 网络的室外复杂环境小目标语义分割方法[J/OL]. 吉林大学学报(工学版), 1-12[2025-04-26]. <http://kns.cnki.net/kcms/detail/22.1341.T.20240402.1737.003.html>.
- AI Q L, LIU Y X, YANG J H. Semantic segmentation method for small targets in outdoor complex environment based on MFF-STDC network[J/OL]. Journal of Jilin University(Engineering Edition), 1-12[2025-04-26]. <http://kns.cnki.net/kcms/detail/22.1341.T.20240402.1737.003.html>.
- [20] 邵玉文, 裴东. 融合交叉注意力机制的轻量化街景语义分割算法[J]. 西北师范大学学报(自然科学版), 2025, 61(1): 61-71.
- SHAO Y W, PEI D. A lightweight street scene segmentation algorithm incorporating criss-cross attention[J]. Journal of Northwest Normal University (Natural Science), 2025, 61(1): 61-71.
- [21] 吴炳剑, 高琳, 李衍志, 等. 结合生成对抗网络与混合注意力机制的街景图像语义分割[J]. 软件导刊, 2024, 23(11): 187-192.
- WU B J, GAO L, LI Y ZH, et al. Semantic segmentation of street view images based on generative adversarial networks and mixed attention mechanisms[J]. Software Guide, 2024, 23(11): 187-192.
- [22] 朱磊, 车晨洁, 姚同钰, 等. 融合多维特征的街景图像语义分割方法[J]. 液晶与显示, 2024, 39(7): 980-989.
- ZHU L, CHE CH J, YAO T Y, et al. Semantic segmentation method for street images with multi-dimensional features [J]. Chinese Journal of Liquid Crystals and Displays, 2024, 39(7): 980-989.

作者简介

孙豪辰, 硕士研究生, 主要研究方向为计算机视觉、语义分割。

E-mail: 275482164@qq.com

安毅(通信作者), 博士, 教授, 主要研究方向为三维环境感知与认知、模式识别与人工智能等。

E-mail: anyi@dlut.edu.cn

张永康, 硕士研究生, 主要研究方向为计算机视觉、三维重建。

E-mail: yorkenzhang@qq.com

熊攀, 硕士研究生, 主要研究方向为深度学习、点云配准。E-mail: 107552201553@stu.xju.edu.cn