

DOI:10.19344/j. cnki. issn1671-5276.2024.03.039

基于改进高分辨率网络的多语义图像分割方法

张少杰, 彭富明, 方斌, 张子祥, 相福磊, 何浩天

(南京理工大学, 江苏 南京 210094)

摘要:针对室外复杂场景下图像分割难度较大的问题,提出一种基于 HRNet 的多语义图像分割模型(HR_DfeNet)。该模型通过引入通道注意力和空间注意力模块优化特征提取,通过改进金字塔池化模块设计 ASPP_M 模块形成高分辨率特征提取分支,并与多种注意力机制融合。在 Cityscape 数据集上,HR_DfeNet 相较于传统分割模型表现出不同程度的分割优化效果。

关键词:室外复杂场景;图像分割;注意力模块;金字塔池化模块

中图分类号:TP391.41 **文献标志码:**A **文章编号:**1671-5276(2024)03-0181-04

A Multi Semantic Image Segmentation Method Based on Improved High Resolution Networks

ZHANG Shaojie, PENG Fuming, FANG Bin, ZHANG Zixiang, XIANG Fulei, HE Haotian

(Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract:To address the difficulty of image segmentation in complex outdoor scenes, this paper proposes a multi semantic image segmentation model based on HRNet (HR_DfeNet), which optimizes feature extraction by introducing channel attention and spatial attention modules, designs a high-resolution feature extraction branch by improving the pyramid pooling module and ASPP_M module, and integrates with multiple attention mechanisms. On the Cityscape dataset, HR_DfeNet exhibits varying degrees of segmentation optimization performance compared to traditional segmentation models.

Keywords:outdoor complex scenes; image segmentation; attention module; pyramid pooling module

0 引言

与室内环境不同的是,室外结构化或非结构化环境下的多语义图像中的信息区分度差,类别边缘更模糊,从而识别分割难度更大。所以室外环境下如何实现对图像信息的精确快速区分成为当前研究的重点内容^[1]。

基于深度学习的语义分割方法是现在的主流研究方向,这类研究主要基于卷积神经网络(CNN)。包子涵等^[2]提出改进余弦型高斯核函数的非局部均值滤波算法,减小了图片预处理阶段噪点的影响。LONG 等^[3]提出第一个深度学习语义分割模型 FCN,通过将全连接层替换为反卷积层并进行上采样以生成空间特征映射,从而产生密集的像素级特征。这一创新证明了深层网络可用于输入尺寸可变的语义分割模型中,进行端到端训练来预测像素级别的标签。此外,HRNet

在分割过程中能够始终保持高分辨率特征,通过高低分辨率特征的并行连接和信息交换来不断优化特征表示^[4]。CHEN 等^[5]提出的 DeepLab 系列模型使用深度卷积的方式进行特征提取,在卷积层中引入空洞卷积和空洞空间金字塔池化模块(atrous spatial pyramid pooling, ASPP)扩大了感受野并捕获了多尺度空间信息。张艺杰^[6]设计的双路径网络可分别提取空间和语义信息,并通过一个高效的特征融合模块来融合这两种信息。同时,近年来注意力机制也被广泛应用于语义分割模型,注意力机制通过注意力分布计算特征加权融合,能够有效地处理多个特征向量^[7-8]。

目前深度学习的语义分割方法准确率较高,但仍存在上采样阶段特征提取不充分、分辨率限制等问题。本文针对室外复杂场景图像分割改进,优化特征提取并设计高分辨率特征提取分支,最后通过实验验证改进方法的有效性。

基金项目:国家重点研发计划项目(2021YFE0194600);江苏省科技计划项目(BZ2023023)

第一作者简介:张少杰(1999—),男,河南周口人,硕士研究生,研究方向为智能驾驶,335002867@qq.com。

通信作者:彭富明(1965—),男,江苏宜兴人,研究员级高级工程师,硕导,硕士,研究方向为车辆检测与控制,新能源汽车,机电控制工程,pfm@njust.edu.cn。

1 模型改进

本文以高分辨率识别分割网络(HRNet)模型为基础,通过引进多种注意力机制和优化ASPP模块等方法对模型进行优化设计,构建一种新型高效多语义图像分割网络模型。

1.1 基于ECA/SA下的模块优化

图像分割中,层级模块之间传递的特征信息包含空间特征和通道特征信息。本节结合高分辨率网络结构的特点,通过引入通道注意力机制(ECA)和空间注意力机制(SA),提高HRNet网络的特征传递效率。

1) ECA 的优化设计

通道注意力机制通过赋予不同通道的特征不同的重要性,使网络更高效地提取与任务相关的特征。

针对HRNet的结构特点,本文设计将ECA通道模块引入主干网络的Concat之后,对合并的高分辨率信息进行进一步优化,具体如图1所示。

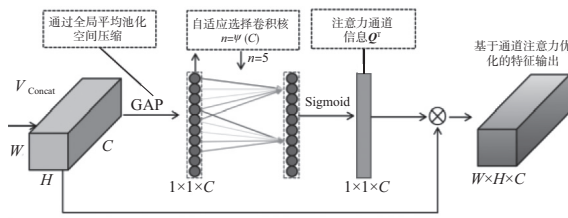


图1 通道注意力机制

在对多个串行分支上采样Concat后,获得尺寸大小为 $H \times W \times C$ 的特征图 V_{Concat} ,其中 C 是 V_{Concat} 的通道数。通过全局平均池化对其在空间尺度上进行压缩,使其变为 $1 \times 1 \times C$,在归一化空间尺度的同时,通道数保持不变。然后通过使用一个自适应大小为 5×5 的卷积操作,捕获不同通道之间的信息,利用激活函数输出特征图的通道注意力权重信息 Q^T 。具体计算公式如式(1)所示。

$$\begin{cases} Q^T = \text{Sigmoid} \{ C_{ov}^{n \times n} (A_{Pool}(V_{Concat})) \} \\ n = \psi(C) = \left\lfloor \frac{\log_2(C)}{2} + \frac{1}{2} \right\rfloor \end{cases} \quad (1)$$

式中: $C_{ov}^{n \times n}$ 为自适应卷积操作; n 为卷积核大小; A_{Pool} 为平均池化函数。

最后,将 Q 与输入特征图数据 V_i 进行乘积,即完成对Concat输出特征图的通道优化。

2) SA 的优化设计

在HRNet中,浅层分支可能包含干扰信息,

直接上采样会导致伪影和影响语义分割结果,尤其在需要精确边界的任务中。

因此本文对HRNet的每个高分辨率提取分支引入空间注意力更新模块以提高图像特征传递效率,具体如图2所示。在每个分支进行上采样Concat前,对输出尺寸为 $C \times H \times W$ 的数据特征 F 进行通道上的压缩处理,使 $C=1$ 。其中每个像素权值相同,但在整个平面中权值不同; H 和 W 是 F 的高度和宽度。对压缩后的数据进行最大池化和平均池化操作,然后将两种池化数据进行级联拼接,最后通过卷积层和激活函数确定通道平面的注意力权值分布 Q 。具体公式如式(2)所示。

$$Q_i = \text{Sigmoid} \{ C_{ov}^{n \times n} [(A_{Pool}(V_i); M_{Pool}(V_i))] \}, \quad (i=1,2,3,4) \quad (2)$$

式中 $A_{Pool}(V_i)$ 和 $M_{Pool}(V_i)$ 分别表示第 i 个分支的平均池化与最大池化结果; $C^{n \times n}$ 表示 $n \times n$ 大小的卷积运算。

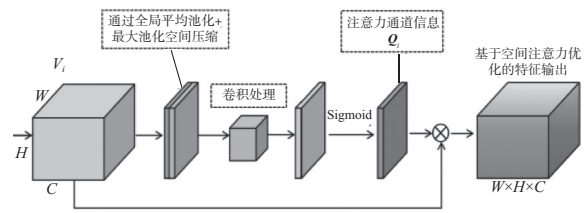


图2 空间注意力机制

最后,将第 i 个分支的权值分布 Q_i 与输出特征图 V_i 乘积,即完成对特征图的空间权值优化过程,得到了具有更明显空间语义信息的新特征图 V_i' 。

1.2 基于优化ASPP下的分支优化

HRNet的高分辨率并行结构使其可以进行多次特征卷积融合,但常规融合方式对精度提升有限,且增加训练复杂度。因此,为优化并行分支融合效果,本文进一步引入空洞空间卷积池化金字塔(ASPP)模块。

ASPP模块通过多个并行分支使用不同尺度的空洞卷积和池化操作,增大感受野,提取多尺度特征。其结构如图3所示,模块通过膨胀卷积层、全局池化层和Concat融合层构建金字塔池化,以不同膨胀率捕获多尺度信息。适当的膨胀率选择对感受野和信息提取至关重要,能避免网格问题并提升精度。针对膨胀率选取问题,本文设计了混合扩张卷积框架(HDC)来构建金字塔的膨胀率,以减轻网格问题。首先对 k 个不同尺寸下的膨胀卷积模块,定义其膨胀率分别为 $[p_1, p_2, \dots, p_k]$,则有

$$D_i = \max(D_{i+1} - 2p_i, -D_{i+1} + 2p_i, p_i),$$

$$(i=1, 2, \dots, k-1) \quad (3)$$

式中 $D_i = p_i$ 定义为两个非零点之间的最大距离。为最小化“网格问题”的影响,选择膨胀率时,要确保一个组内卷积的变换因子不固定,即不包含大于1的公约数。因此本文选择的空洞率分别为(2,3,7,13)这4种尺度,加上池化层和原特征图层,共6层结构,对ASPP模块改进设计为ASPP_M,具体如图4所示。

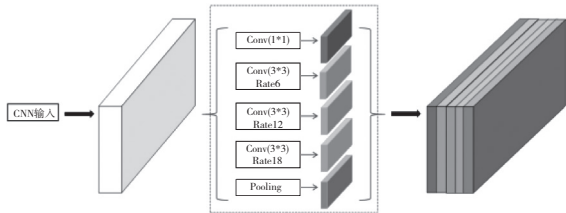


图3 空洞空间金字塔池化模型图

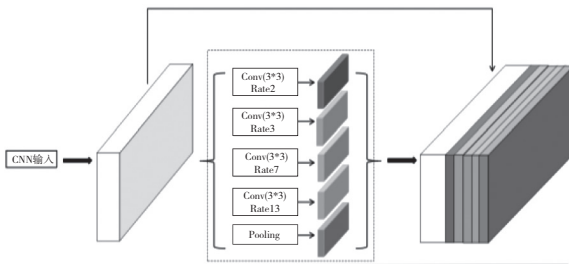


图4 基于HRNet模块改进的ASPP_M设计

在改进ASPP模块中,首先对原特征数据不做处理直接进行下级传递,然后利用上述设计的4种不同膨胀率下的空洞卷积对原特征图进行不同尺度下的特征提取;最后利用平均池化完成全局下的语义提取,同时增加了批归一化层提高模型的训练效率。最后进行Concat数据融合,完成ASPP_M模块的特征增强任务。

1.3 整体模型搭建

通过上述改进,本文基于HRNet构建了双特征提取分支下的优化模型(HR_DfeNet)。该模型一方面通过多种注意力优化模块针对性提取模型的特征辨识度;另一方面通过优化金字塔模型加强对特征图多尺度信息的语义提取能力,整体架构如图5所示。

2 算法验证与对比分析

2.1 实验环境和参数配置

本文的语义分割方法主要针对室外一般结构

化或非结构化复杂场景进行研究,并基于仿真平台进行实验验证,因此可通过网上查找和自行拍摄等建立数据集。该数据集包含了各类室外场景的1000张图片。实验采用传统MIoU作为模型性能的核心评估指标,使用交叉熵损失函数和Adam优化器来训练模型。训练过程包括320个Epoch轮次,初始学习率设定为0.0001,BatchSize设置为12,选取数个目前主流多语义分割算法进行综合对比分析,包括HRNet、DeepLabV3+、U-Net、PSPNet。

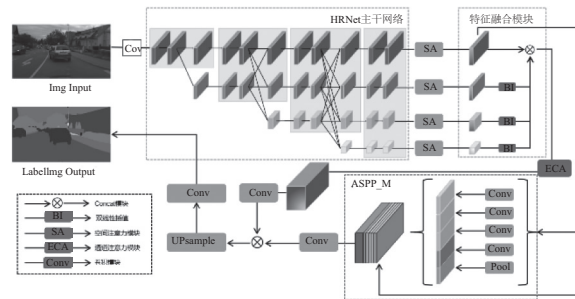


图5 HR_DfeNet模型框架

2.2 消融实验

本文针对上述多个优化模块进行了消融实验,以验证每个模块对分割效果的影响。取数据集数量15%比例的图像进行模型训练,为确保实验评估的准确性,每项消融实验均重复3次,以平均值作为数据基准,并记录最值以评估结果的波动情况。

本文设计5种方案对模型的消融部分进行对比分析,分别为:1)原HRNet模型;2)SA优化的HRNet模型;3)SA+ECA改进(即单特征提取分支)下的HRNet模型;4)ASPP_M(单特征提取分支)下的HRNet模型;5)HR_DfeNet(双特征特区分支融合)模型。

具体的实验结果如表1所示,其中√代表在原HRNet模型中添加该(改进的)模块,×代表不添加。可以观察到,相较于原HRNet网络,方案2—方案5所建的其余模块均在不同程度上优化了分割效果,验证了本文设计的多种优化方案的有效性。其中方案1—方案3的优化效果逐级明显,说明基于注意力机制下的高分辨率特征提取分支有效地保留了特征图的空间细节和全局信息。同时,方案4—方案5表明基于ASPP_M的高分辨率特征提取分支通过多种尺度的卷积和池化操作,显著优化了图像在多尺度上的语义信息处理效果。

表 1 基于 HR_DfeNet 的消融实验

消融方案	SA 模块	ECA 模块	ASPP_M 模块	MIoU 值/%
1	×	×	×	70.6±0.5
2	✓	×	×	71.1±0.2
3	✓	✓	×	71.9±0.2
4	×	×	✓	71.4±0.4
5	✓	✓	✓	72.3±0.2

2.3 HR_DfeNet 分割效果验证

为评估 HR_DfeNet 模型的分割精度,对 5 种算法进行了对比分析。

图 6 展示了 HRNet 与 HR_DfeNet 模型在测试集上的部分分割对比效果。由于本文的主干网络是 HRNet,图中仅展示了 HRNet 与 HR_DfeNet 的分割效果。HRNet 在远视距和类别边缘区分不明显的物体分割上效果较差,而 HR_DfeNet 模型能够更好地提取类别之间的语义差别。尽管未能完全筛选出所有类别,但相比 HRNet 有明显优化,并且 HR_DfeNet 算法对细小枝叶轮廓及墙体等大面积内细小物体的识别能力显著改善。

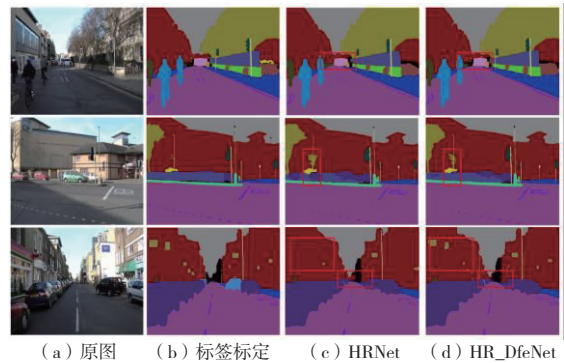


图 6 HRNet、HR_DfeNet 对比效果图

将训练集图片分割为 11 种类标签,各个算法模型在测试集下的分割效果如表 2 所示。较 HRNet 模型,HR_DfeNet 模型在多个细小类别如指示牌、杆、栅栏等的分割精度对比原 HRNet 模型优化效果明显;HR_DfeNet 的平均交并比较 HRNet 模型提高了 2.6%,较 DeepLabV3+ 模型提升了约 3.7%,比其他模型也有不同程度的优越性。

表 2 各模型分割 MIoU 对比结果 单位: %

标签类别	PSPNet	DeepLabV3+	U-Net	HRNet	HR_DfeNet
Sky	88.9	92.5	89.5	90.3	92.1
Building	81.6	86.1	79.3	81.2	83.4
Pole	39.1	46.7	25.2	48.6	49.2
Road	90.4	91.5	82.7	89.5	90.3

续表 2

单位: %

标签类别	PSPNet	DeepLabV3+	U-Net	HRNet	HR_DfeNet
Sidewalk	79.5	80.0	65.9	81.0	83.7
Tree	71.0	79.1	70.4	78.1	81.2
Sign	48.2	57.0	42.6	61.3	63.4
Fence	45.7	36.9	40.7	34.6	42.1
Car	81.4	78.1	80.3	81.5	84.9
Pedestrian	49.6	45.6	40.7	51.3	52.8
Bicyclist	45.1	49.7	39.4	52.6	50.9
MIoU	65.5	67.5	59.7	68.6	71.2

3 结语

本文通过研究室外多语义图像分割算法,提出了一种基于 HRNet 的多语义图像分割模型 (HR_DfeNet)。该模型通过引入空间/通道注意力机制和 ASPP_M 模块,显著提升了模型对高分辨率特征图的表征能力。通过在自建数据集上的实验验证,本文所建模型在分割精度上较多种传统算法皆有明显的优化和提升效果。

参考文献:

[1] 余京蕾.浅谈计算机视觉技术进展及其新兴应用[J].北京联合大学学报,2020,34(1):63-69.

[2] 包子涵,李龙海,刘丽丽,等.基于机器视觉的救援机器人自动避障技术研究[J].机械制造与自动化,2024,53(1):202-208.

[3] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,39(4):640-651.

[4] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE,2019:5686-5696.

[5] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs[EB/OL]. (2016-06-07)[2024-05-01]http://arxiv.org/abs/1412.7062.

[6] 张艺杰.基于深度学习的高分辨率遥感图像语义分割方法研究[D].成都:电子科技大学,2022.

[7] LIU G P, KE J. End-to-end full-waveform echo decomposition based on self-attention classification and U-net decomposition[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,2022,15:7978-7987.

[8] 田壮壮,张恒伟,王坤,等.改进 CenterNet 在遥感图像目标检测中的应用[J].遥感学报,2023,27(12):2706-2715.

收稿日期:2024-05-21