

文章编号:1671-1637(2026)06-0167-19

基于多尺度特征融合及通道特征自适应的 道路坑洞分割网络

王司宇¹, 方虹苏¹, 杨 炜^{*1}, 周勇军²

(1. 长安大学 汽车学院, 陕西 西安 710064; 2. 长安大学 公路学院, 陕西 西安 710064)

摘要:为准确感知道路坑洞并提高模型泛化分割能力,提出了一种基于空洞空间金字塔融合及特征自适应的道路坑洞分割网络(potholes-FBConvNet)。将卷积网络模型(ConvNext)作为主干结构,进行不同网络深度的多尺度特征提取;将统一感知解析网络架构(UPerNet)作为颈部融合网络,充分利用主干网络所提取的多级特征表示,在颈部网络中引入空洞空间金字塔融合模块,进一步增强网络捕获多尺度上下文特征信息能力;在颈部及解码头部之间嵌入特征适应模块,在颈部融合特征解码之前对通道及空间特征进行自适应校准,有效提升网络分割性能;分别收集并精细化标注2 097张道路坑洞(普通破损坑洞1 065张,完全破损坑洞507张,重度破损坑洞525张)分割数据库,并依次进行对比及消融试验以验证网络性能。试验结果表明:提出的 potholes-FBConvNet 模型交并比、相似系数、精确率及召回率在普通破损坑洞上可达85.43%、92.14%、92.96%及91.33%,在完全破损坑洞上可达87.76%、93.48%、92.33%及94.67%,在重度破损坑洞上可达90.76%、95.15%、95.41%及94.90%;相比16种基于Transformer及卷积Conv的典型对比模型,提出的分割网络具备最优泛化分割能力及鲁棒性。

关键词:自动驾驶;坑洞分割;复杂环境;空洞空间金字塔融合;特征适应

中图分类号:U445.7 **文献标志码:**A **DOI:**10.19818/j.cnki.1671-1637.2026.031

Road pothole segmentation network based on multi-scale feature fusion and channel feature adaptation

WANG Si-yu¹, FANG Hong-su¹, YANG Wei^{*1}, ZHOU Yong-jun²

(1. School of Automobile, Chang'an University, Xi'an 710064, Shaanxi, China;

2. School of Highway, Chang'an University, Xi'an 710064, Shaanxi, China)

Abstract: To accurately identify road potholes and the improve model generalization segmentation ability, a road pothole segmentation network (potholes-FBConvNet) based on astrous spatial pyramid fusion (ASPF) and feature adaptation was proposed. The convolutional neural network model (ConvNext) was used as the main structure for multi-scale feature extraction at different network depths. The unified perception and resolution network architecture (UPerNet) was used as the neck fusion network. The multi-level feature representations extracted through the main

出版历程:2025-01-12 收稿,2025-03-27 修回,2025-06-05 录用

基金项目:国家重点研发计划(2021YFB2601000);陕西省重点研发计划(2024CY2-GJHX-31)

作者简介:王司宇(1999-),女,山东菏泽人,博士研究生,E-mail:2023022005@chd.edu.cn。

***通信作者:**杨 炜(1985-),男,陕西蒲城人,讲师,工学博士,E-mail:yw@chd.edu.cn。

引用格式:王司宇,方虹苏,杨 炜,等. 基于多尺度特征融合及通道特征自适应的道路坑洞分割网络[J]. 交通运输工程学报,2026,26(6): 167-185.

Citation: WANG Si-yu, FANG Hong-su, YANG Wei, et al. Road pothole segmentation network based on multi-scale feature fusion and channel feature adaptation[J]. Journal of Traffic and Transportation Engineering, 2026, 26(6): 167-185.

network were fully utilized, and the ASPF module was introduced into the neck network to further enhance its ability to capture multi-scale contextual feature information. A feature adaptation module was embedded between the neck and decoding head to perform adaptive calibration on the channel and spatial features before neck fusion feature decoding. In this way, the network segmentation performance was effectively improved. A total of 2 097 road pothole segmentation databases were separately collected and meticulously annotated (1 065 generally damaged potholes, 507 completely damaged potholes, 525 severely damaged potholes), and comparative and ablation experiments were conducted in sequence to verify network performance. The experimental results show that, the proposed potholes-FBConvNet model achieves an intersection over union (IoU), similarity coefficient, precision, and recall of 85.43%, 92.14%, 92.96%, and 91.33% on generally damaged potholes. On completely damaged potholes, it reaches 87.76%, 93.48%, 92.33%, and 94.67%. On severely damaged potholes, it reaches 90.76%, 95.15%, 95.41%, and 94.90%. Compared with 16 typical comparison models based on Transformer and Conv, the proposed segmentation network has the optimal generalization segmentation ability and robustness.

Keywords: autonomous driving; pothole segmentation; complex environment; astrous spatial pyramid fusion; feature adaptation

Publication history: Received 2025-01-12; Received in revised form 2025-03-27; Accepted 2025-06-05

Funding: National Key R&D Program of China (2021YFB2601000); Key R&D Program of Shaanxi Province (2024CY2-GJHX-31)

* **Corresponding author:** YANG Wei, lecturer, PhD, E-mail: yw@chd.edu.cn.

0 引 言

路面坑洞感知是自动驾驶环境感知的重要环节,其核心任务是通过图像分析识别坑洞并定位其区域,可为速度及路径规划提供决策依据。道路是城市基础设施的重要部分^[1],将直接影响车辆行驶安全。中国车辆超载严重,导致路面损坏加剧^[2],甚至是坑洞的形成,因此及时发现道路坑洞成为提升交通安全及道路使用的关键问题,但传统人工巡查方法成本高且效率低^[3],而重型机械设备昂贵且操作复杂,不适合大规模检测^[4]。自动驾驶中路面积洞检测则可通过图像分析以准确识别,对于车辆速度控制及路径规划具有重要实际意义^[5]。

近年来,深度学习图像分割技术的突破,为道路坑洞检测提供了新的解决方案。国外学者 Gao 等^[6]提出了一种整合灰度及纹理特征处理的方法,通过结合工业相机以实现对坑洞的快速准确检测;Abbas 等^[7]通过提取裂缝及坑洞,从图像与视频中对其进行分类,并比较手动与自动方法以评估路面损坏;Kim 等^[8]提出了一种数值边缘算法来提高微小物体分割准确性,可在沥青路面上精确检测坑洞;Ye 等^[9]提出了卷积神经网络,利用数字图像可以进

行坑洞检测方面应用;Wang 等^[10]策划了路面坑洞、剥落遇险检测及分割数据集,提出了一种两步路面坑洞及裂缝检测分割方法,以增强网络对各种形状大小坑洞与裂缝的分割能力;Lakmal 等^[11]提出了一种坑洞检测应用程序,可用于检测充满水的坑洞,并使用随机森林分类器训练分割网络。尽管上述坑洞检测方法在一定程度上提高了识别效率,但面临复杂环境时仍精度较低,难以提供稳定且准确检测结果。此外,袁朝春等^[12]为实时、有效获取路面坑洞损毁情况并提高智能汽车行驶安全性,提出了一种新的无接触式坑洞检测及主动避障控制算法;宣以国等^[13]针对道路实时检测中存在漏检及错检等问题,提出了一种基于改进 YOLOv7 的道路裂缝及坑洞检测算法,具有出色的检测效果;王鑫等^[14]使用 YOLOv5x 结合透视变换及图像分割设计了路面缺陷检测系统;胡晓伟等^[15]针对现有路面病害检测方法存在精确度不足问题,提出了一种融合状态空间模型的 YOLOM 检测方法,以增强模型对局部细节及全局语义的表达能力;罗向龙等^[16]针对多种病害并发的复杂场景下道路病害识别精度不高、泛化能力不足的问题,提出了一种复杂场景的病害检测模型 RGT-YOLOv7,可在复杂场景下有效检测道

路病害;郭道俊等^[17]为避免依托人工进行隧道定期检测方式中由于工作人员专业水平差异及个人主观性对检测结果的影响,提出了一种基于 DeepLabV3+ 的改进病害分割方法;任泳洁等^[18]针对道路病害快速准确实时识别需求,提出了一种基于 MobileNetV3 的轻量化道路病害识别网络模型。上述方法虽可以对单类坑洞进行较好检测,但面对多样化路面坑洞时,泛化性及适应性仍较弱^[19]。

自动驾驶系统通常依赖图像传感器识别感知坑洞^[20],但实际场景中,多样化的路面背景及坑洞形态对网络提出了较高要求。鉴于此,本文提出了一种基于空洞空间金字塔融合及特征适应的 potholes-FBConvNet 坑洞分割模型,通过结合空洞空间金字塔融合(Astrous Spatial Pyramid Fusion, ASPF)模块及特征适应模块(Bottleneck Attention Module, BAM),可更好地捕获多尺度信息并自适应调整解码特征,实现对不同道路类型坑洞的准确分割,并具备更强的泛化分割感知能力,以增强自动驾驶系统在多种环境下对坑洞的感知能力。

1 整体流程及数据库构建

1.1 道路坑洞分割流程

所提出的道路坑洞分割流程如图1所示。坑洞根据损坏程度可以划分为普通破损(小型坑洞,车辆经过时仅造成轻微颠簸)、完全破损(存在多种类型、大小不一且深浅不一坑洞)及重度破损(坑洞直径及深度较大)三类坑洞。此外,对所构建坑洞数据库进行裁剪及归一化,并应用数据增强技术对其进行扩充整理,完成道路坑洞数据库搭建;将所构建的坑洞数据库传入所提出的 potholes-FBConvNet 分割模型,并设定适宜的模型训练超参数,完成迭代训练,实现对不同道路坑洞的特征提取及准确分割;其中主干网络采取 ConvNext 进行不同深度的特征提取,颈部网络则结合空洞空间金字塔融合模块,进一步丰富所提取特征,并在解码头之前嵌入特征适应模块,对解码特征进行再次调整优化,从而可输出道路坑洞分割结果;最后对各典型网络分割结果进行可视化对比,以评估所提出网络的分割泛化性。

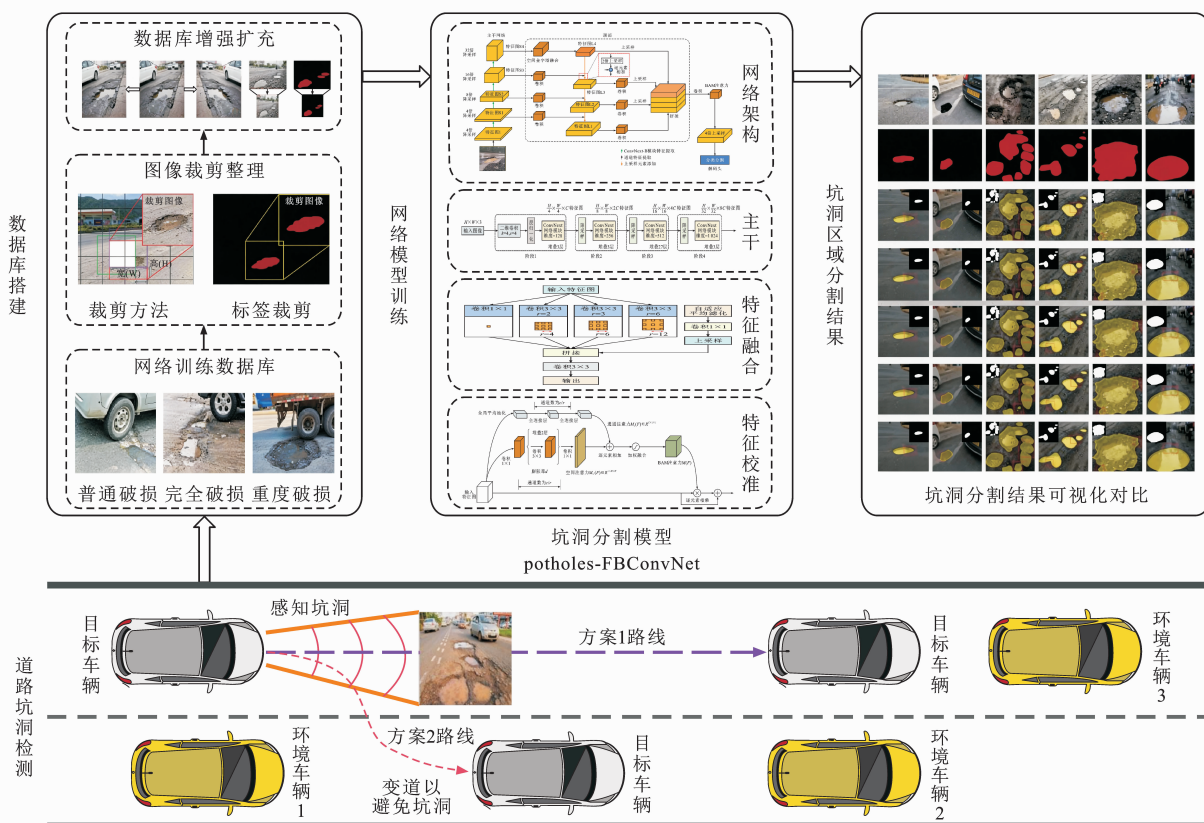


图1 道路坑洞分割流程

Fig. 1 Road pothole segmentation process

1.2 数据处理方法

1.2.1 数据库构建及预处理

为建立高质量道路坑洞分割数据库,其通过相

机在实际环境中拍摄及网络搜集方式进行构建。然而,鉴于实际采集图像数量较为有限,且网络搜集图像分辨率大小不一,为对数据集进行统一整理,并考

考虑到实际训练中对计算资源的合理使用,故进行了以下预处理。

所有原始图像处理为 480 像素 $\times$ 480 像素,以减少无关背景信息对计算资源的浪费,并适应小批量训练,具体裁剪过程如图 2 所示,该方法可高效去除冗余背景,以保留道路坑洞关键特征。

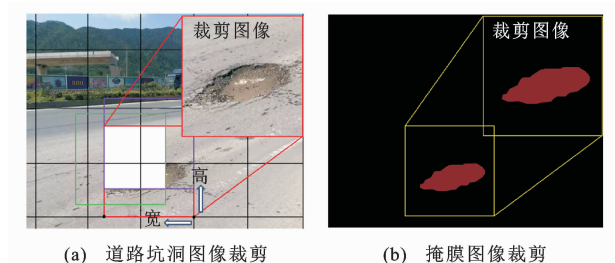


图 2 数据库图像裁剪示例

Fig. 2 Example of database image cropping

为确保数据集多样性及有效性,对较大分辨率图像进行了区域分块,并通过随机偏移裁剪,以获取多角度、多视角坑洞图像样本,并优先保留包含足够坑洞像素图像,以保证训练数据有效性。其中所有图像尺寸均被统一为 480 像素 $\times$ 480 像素,色深均为 24 位,颜色空间为标准 RGB 格式。图 2(a)展示了从原始图像中裁剪图像块的方法,核心过程是沿不同路径对图像进行 2 次遍历(白色箭头),并保存符合标准的道路坑洞图像(白色区域为 3 个示例框重叠区域),其中不同颜色框则代表不同大小图像块;图 2(b)则展示了在掩膜图像上应用裁剪的方法示例。

此外,所有道路坑洞图像均通过 LabelMe 标注软件进行精细化标注,以支持后续网络算法训练。标注过程中,对道路坑洞轮廓进行了较为精确勾画,以确保其尽量贴合真实图像坑洞形状;其后,将标注生成的 JSON 文件转码为掩膜图,以二进制格式存储,其中坑洞区域像素标记为 1,背景像素标记为 0。同时,为便于后续分割网络效果对比,将标注的坑洞区域设置为调色板模式(红色)。

为确保模型泛化能力,数据集被随机严格划分为训练集、验证集及测试集,比例大致为 70 : 15 : 15,且三者之间无重叠。其中训练集用于模型参数训练,验证集用于调整及优化,测试集则用于评估模型的分割性能。

此外,部分分辨率接近 480 像素 $\times$ 480 像素的原始图像均被调整为 480 像素 $\times$ 480 像素,如图 3 所示。

由于训练图像数量有限,故进行了数据增强操作,包括随机翻转及噪声添加,以增加数据集的多样

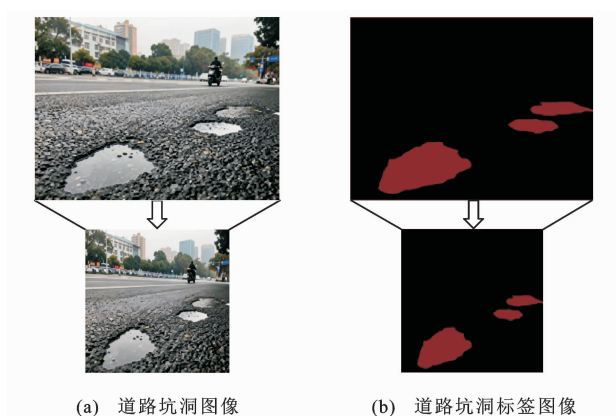


图 3 图像 resize 调整

Fig. 3 Image resize adjustment

性。通过以上数据增强方法不仅可扩充数据量,还可保留原始数据结构特征,使模型在不同场景道路坑洞分割中具备更强适应性,具体增强示例如图 4 所示。



图 4 图像翻转及噪声增强

Fig. 4 Image flipping and noise enhancement

通过以上数据标注及处理步骤,分别建立了 3 个高质量道路坑洞分割数据集,可为路面坑洞分割提供多尺度及多视角的可靠训练数据,由于完全破损坑洞与重度破损坑洞之间界限存在一定重叠,因此在所构建的 3 类数据库中,2 类坑洞会存在部分重叠。

1.2.2 普通破损坑洞

普通破损坑洞在路面上较为常见,坑洞破坏程度较小。经数据预处理后,普通破损坑洞数据集共包含 1 065 张 RGB 彩色图像,部分代表性图像如图 5 所示。该数据集涵盖了多种道路场景,包括轻微破损坑洞、橡胶路面坑洞、近距离破损场景及复杂背景下的普通破损坑洞,以确保算法可在不同环境中测试其泛化能力。

为对普通破损坑洞数据集进一步分析其质量,绘制了如图 6 所示坑洞分割区域比例分布,其中横轴表示坑洞区域比例(坑洞区域像素占整个图像的比例),纵轴表示该比例区间内频数。大部分数据目标区域比例集中在较低范围(0~0.2),随着比例增加,频数逐渐减少。表明大部分普通坑洞目标区域较小。此外,坑洞区域比例均值为 0.101,表明多数图像的坑洞区域仅占图像较小部分,且集中在较低比例附



图5 普通破损坑洞示例

Fig. 5 Examples of ordinary damaged potholes

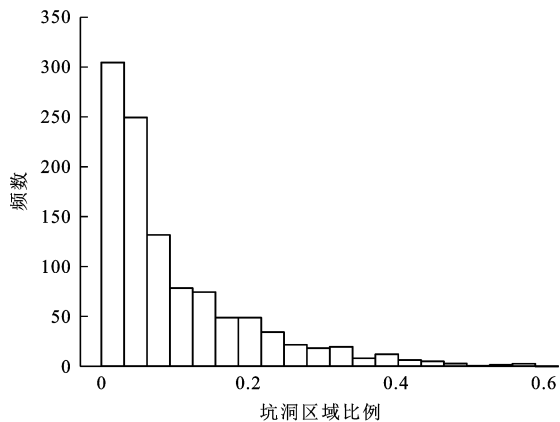


图6 普通破损坑洞分割区域比例分布

Fig. 6 Proportion distribution of ordinary damaged potholes segmentation area

近,方差为 0.010 则代表目标区域比例变化相对较小,数据分布较为集中,变化幅度不大。

1.2.3 完全破损坑洞

完全破损坑洞通常具有多种不同类型坑洞,且坑洞中往往包含积水,由于完全破损坑洞较为少见,因此该数据集收集较为困难,经数据处理后共包含 507 张 RGB 彩色图像,部分完全破损坑洞的代表性图像如图 7 所示。所构建的完全破损数据集涵盖了多种道路场景,如泥泞路面、沙砾路面、道路施工场景及乡村公路完全破损坑洞,可模拟车辆在实际行驶中可能遇到的极端道路条件,以全面测试算法在复杂环境的鲁棒性及适应性。

为进一步分析完全破损坑洞数据集质量,绘



图7 完全破损坑洞示例

Fig. 7 Examples of completely damaged potholes

制了如图 8 所示的完全破损坑洞分割区域比例分布图。大部分数据的坑洞区域比例集中在较低的范围(0~0.3)。随着目标区域比例增加,出现的频数也逐渐下降。数据分布相对平滑,且存在一定集中性。此外,完全破损坑洞数据库中目标区域比例均值为 0.299,数据波动程度方差为 0.019,表明数据相对集中且波动不大,目标区域比例变化较为稳定。

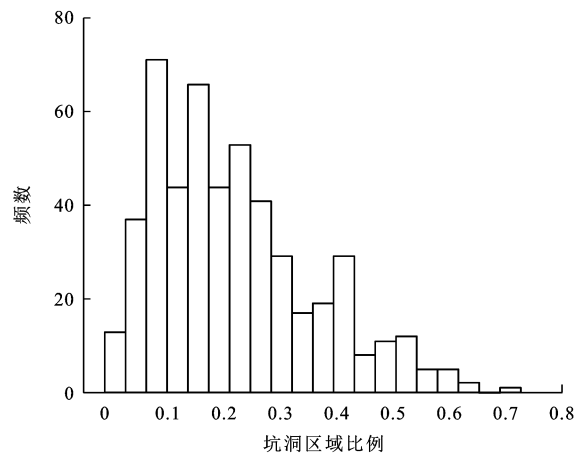


图8 完全破损坑洞分割区域比例分布

Fig. 8 Proportion distribution of completely damaged potholes segmentation area

1.2.4 重度破损坑洞

重度破损坑洞通常指直径及深度均较大坑洞,对车辆行驶安全有显著影响。与完全破损坑洞数据集类似,该类坑洞同样难以在实地及网络上收集。经

过数据处理后,重度破损坑洞数据集共包含 525 张 RGB 彩色图像,部分代表性图像如图 9 所示。该数据集涵盖了多种典型道路场景,包括水泥、橡胶、积水及沙砾路面重度坑洞,可测试算法在极端路况下表现,以评估其对道路大型缺陷的感知能力。



图 9 重度破损坑洞示例

Fig. 9 Examples of severely damaged potholes

为对重度破损坑洞数据集进一步分析其质量,绘制如图 10 所示的重度破损坑洞分割区域比例分布图。大部分数据集中在较小比例区间(0~0.3),

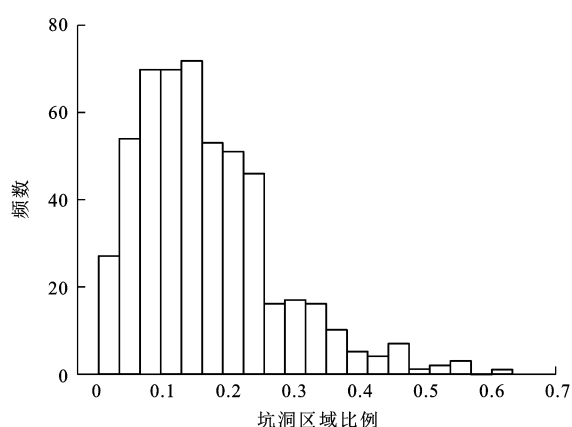


图 10 重度破损坑洞分割区域比例分布

Fig. 10 Proportion distribution of severely damaged potholes segmentation area

随着比例增大,出现频数逐渐减少,表明坑洞区域比例有较为集中分布,其中坑洞分布平均值 0.170 代表了数据的中心位置,而数据波动方差 0.011 则表明重度破损坑洞区域数据较为集中。

2 试验原理分析

2.1 道路坑洞分割网络

为对道路坑洞进行高效精确分割,提出一种高分割精度及泛化性的 potholes-FBConvNet 坑洞分割网络,主要由 3 个核心组件组成:主干网络(特征提取模块)、颈部(特征融合模块)及解码头(预测模块),具体架构如图 11 所示(后续小节将对此网络展开详细论述)。

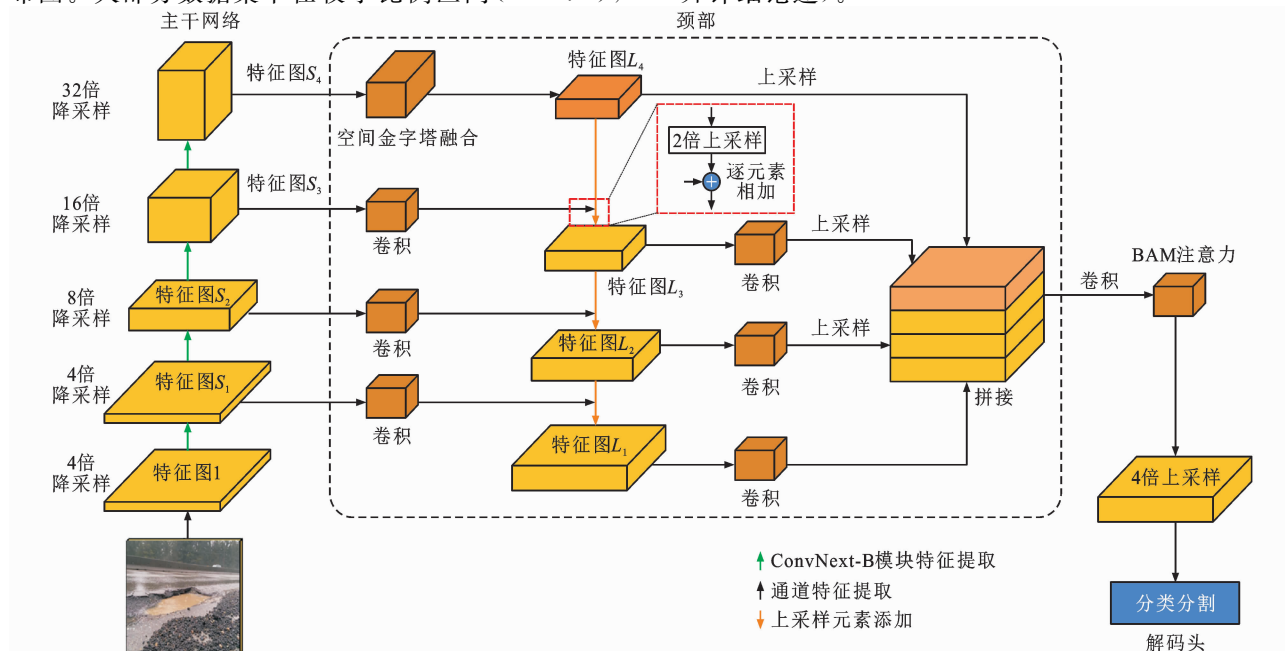


图 11 potholes-FBConvNet 坑洞分割网络

Fig. 11 Potholes segmentation network of potholes-FBConvNet

2.2 主干网络

本文以基于卷积的 ConvNext^[21] 网络作为基础算法,该网络在设计上借鉴了 Transformer^[22] 模型的成功策略,通过引入深度卷积模块,并搭配层归一化(LayerNorm)与 GELU 非线性激活函数等,改进了 ResNet,但其核心模块仍类似于 ResNet 残差块^[23],但在通道分离及归一化处理能力上更为强大,具备更深网络结构及更大感受野,适用于道路坑洞分割。此外,ConvNext 可输出不同层次特征图,

可提取多尺度特征信息,并捕捉图像中细节及全局信息,网络主干结构如图 12 所示(H 、 W 及 C 分别为输入特征图高、宽及通道数;卷积模块 k 、 s 及 p 分别为卷积核尺寸、移动步长及填充参数)。

鉴于 ConvNext 兼具卷积网络在局部建模与 Transformer 中长期依赖建模能力的优势,尤其适于处理具有不规则形态与复杂背景的坑洞区域。其多层次特征输出能力可为后续多尺度融合模块提供丰富的语义与细节信息基础。

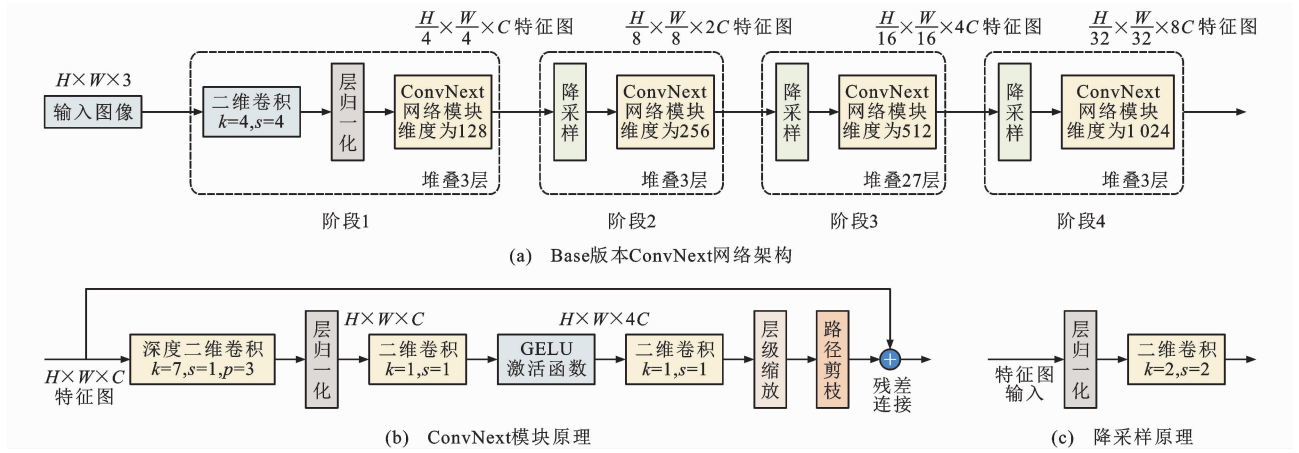


图 12 分割网络主干结构原理

Fig. 12 Principles of segmented network backbone structure

由图 12 可知:每个 ConvNext 网络模块均包含以下关键步骤。深度二维卷积:对每个通道分别进行卷积操作,并保持通道数不变。层归一化:在通道维度进行网络层标准化。二维卷积:使用 1×1 卷积以改变通道数。GELU 激活函数:经过 1×1 二维卷积后的特征图通过该激活函数。层级缩放:残差连接之前,需对激活后特征图进行层级缩放操作,通过一个可学习的缩放因子 λ 以调整每个通道幅度。路径剪枝:在应用残差连接前引入该操作,通过随机丢弃某些路径以避免过拟合。残差连接:为最后步骤,将原始输入与经过层级缩放及路径剪枝处理后的特征图相加,以得到最终结果。此外,ConvNext 主干在每个阶段末尾均对特征图进行下采样以增加通道数。下采样通过步幅为 2 的 2×2 深度卷积实现, 1×1 二维卷积则用于扩展通道数。

最终,通过以上的分层表示,ConvNext 主干在每个阶段可增加通道数、减少空间分辨率,最终可提取到丰富的多尺度特征,其各阶段所输出特征图可表示为 $\{S_1, S_2, S_3, S_4\}$, 对应下采样率分别为 4、8、16 和 32 倍,这些特征图被用于后续解码操作,即道路坑洞分割下游任务。

2.3 颈部网络

颈部模块设计基于统一感知解析网络架构 UPerNet^[24],主要目的是充分利用多级特征表示,以增强模型整体表现力。该模块在语义分割任务中以其高效的多尺度特征融合机制被广泛应用,可有效整合来自主干网络不同阶段的语义与空间信息。考虑到道路坑洞的复杂背景及多样形态,单一尺度特征往往难以准确定位与分割,在引入 UPerNet 作为颈部结构后,不但可自顶向下对特征进行融合,增强模型对高层语义的理解能力,还可充分保留底层细节信息,以提升模型对坑洞区域的感知精度与边界区分能力。此外,该模块选取了从金字塔骨干网络中每个阶段提取的特征图 $\{S_1, S_2, S_3, S_4\}$ 作为颈部输入,其中 S_1 、 S_2 、 S_3 及 S_4 分别来自骨干金字塔网络不同阶段。为将高层语义信息融入中低层特征中,颈部模块构建了一个自上而下结构,并通过横向连接以实现特征融合,其中每个横向连接包含一个 1×1 卷积用于调整通道维度。具体试验中,将 ConvNext 的 Base 版本作为骨干网络,各阶段的通道维度可看为 $\{C, 2C, 4C, 8C\}$, 其中 $C=128$ 。因此每级输出特征图维度为 $\{128, 256, 512, 1024\}$, 通过自上而下架构进行特征融合,可将输出特征图集

$\{L_1, L_2, L_3, L_4\}$ 通道数统一为 512 大小(输入分辨率为 480 像素 $\times$ 480 像素的场景下,统一通道数为 512 有助于充分保留多尺度特征的关键信息;同时,该通道数作为语义分割中常用且验证有效的通道配置,可更好承载多层次语义表达)。

在颈部模块顶部则应用了 ASPF,该模块借鉴了空洞金字塔池化的设计^[25]。ASPF 通过多个不同膨胀率 d 并行空洞卷积(依次设定为 1、2、3 及 6)对输入图像进行多尺度特征提取,并构建不同感受野以捕捉道路坑洞多尺度信息,以减少采样过程中信息丢失。同时,ASPF 通过跳跃连接以融合来自骨干网络不同层级的多尺度特征图,特征图包含丰富空间分辨率及语义深度信息,对精确分割复杂轮廓及不规则形状道路坑洞至关重要。

ASPF 模块具体架构如图 13 所示,其通过不同膨胀率空洞卷积以实现输入路面坑洞图像的特征提取,多尺度信息融合则有助于坑洞分割结果精细化,以进一步提升模型表达及分割精度。

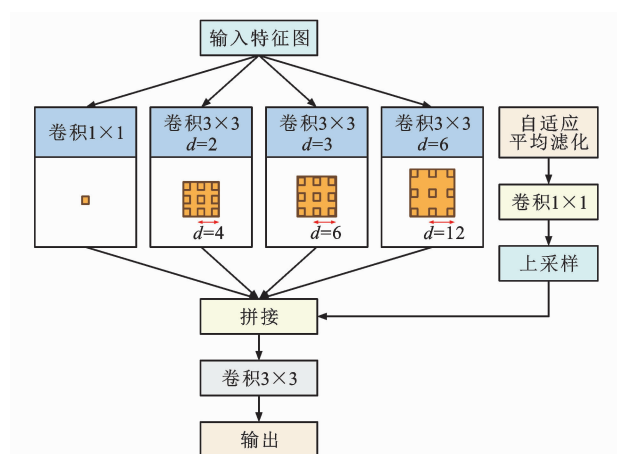


图 13 ASPF 模块原理

Fig. 13 Principle of ASPF module

ASPF 模块共包含 5 个并行分支,分别为 1 个 1×1 卷积、3 个膨胀率为 2、3、6 的 3×3 空洞卷积及残差连接。由于道路坑洞分割中仅有 1 个坑洞目标,因此通过 3 个空洞卷积分支已足够提取多尺度语义信息,残差连接则用于保留位置信息;此外,ASPF 模块中特征图膨胀率则根据阶段 4 输出特征图尺寸 S 设置。对于本研究中输入尺寸为 480 像素 $\times$ 480 像素图像,经总步长为 32 的下采样后,输出特征图尺寸 S 为

$$S = \frac{480}{32} \times \frac{480}{32} = 15 \times 15 = 225 \quad (1)$$

此时,其膨胀率最大设置为 6,对应感受野 f 为

$$f = (k-1)d + 1 = (3-1)d + 1 = 13 \quad (2)$$

式中: d 为膨胀率。

此感受野允许网络捕获广泛的上下文信息,同时使用膨胀率为 2 与 3 的空洞卷积以捕获局部信息,其中 ASPF 模块输出的特征图则被记为 L_4 。为将不同尺寸特征图 $\{L_1, L_2, L_3, L_4\}$ 进行融合,首先需对 L_1, L_2 及 L_3 应用 3×3 卷积(用于调整通道数并减少冗余);然后通过双线性插值将其调整到与 L_4 的相同尺寸 15×15 ;最后,将所有特征图拼接为 L 。其中拼接后的通道数目为 2 048,随后再通过包含 512 个滤波器的卷积对来自不同级别的特征进行融合,并将通道维度从 2 048 减少为 512。

通过以上操作可确保多尺度特征在空间维度对齐的同时,压缩其通道数,以为路面坑洞分割提供更具代表性的语义信息。综合而言,颈部模块通过多尺度特征融合,可显著提升模型捕捉坑洞不同尺度特征的能力。其自上而下的架构设计使高层语义信息逐层传递至中低层特征,可确保模型在保留丰富上下文信息的同时,能有效处理坑洞分割中的细节及边缘问题,并保持较高分割精度,而将 ASPF 模块集成于颈部网络,不仅可增强网络捕捉多尺度上下文能力,还可有效应对坑洞尺寸变化及复杂背景带来的干扰,显著提升整体分割有效性,以确保精确定位坑洞。

2.4 解码头部

解码头部主要作用是对颈部融合的特征图进行逐步上采样,以完成分割预测。此外,坑洞分割网络 potholes-FBConvNet 在颈部及解码头部之间嵌入 BAM 特征适应模块^[26],用于重新校准特征图通道及空间响应。该模块通过自适应调整通道及空间维度注意力权重,可突出关键特征并抑制无关信息,以提升解码头解码效果。

此外,解码头部通过经 BAM 所优化的特征图进行逐步上采样及细节恢复,同时通过多级上采样及特征融合,可将高层语义信息及低层空间细节结合,以生成高分辨率分割结果,该设计不仅可有效应对道路坑洞分割中不规则及大小多样的坑洞,还可准确聚焦坑洞区域,并显著提高分割精度及鲁棒性。

BAM 的具体结构如图 14 所示,图中: R 为特征域; r 为放缩率; $M_o(F)$ 为空间注意力图; $M_c(F)$ 为通道注意力图; $M(F)$ 为对原始特征图 F 经通道注意力与空间注意力融合后得到的三维注意力图。该模块通过捕捉局部及全局的空间于通道信息,可实现多尺度特征融合,以提升网络特征提取能力。

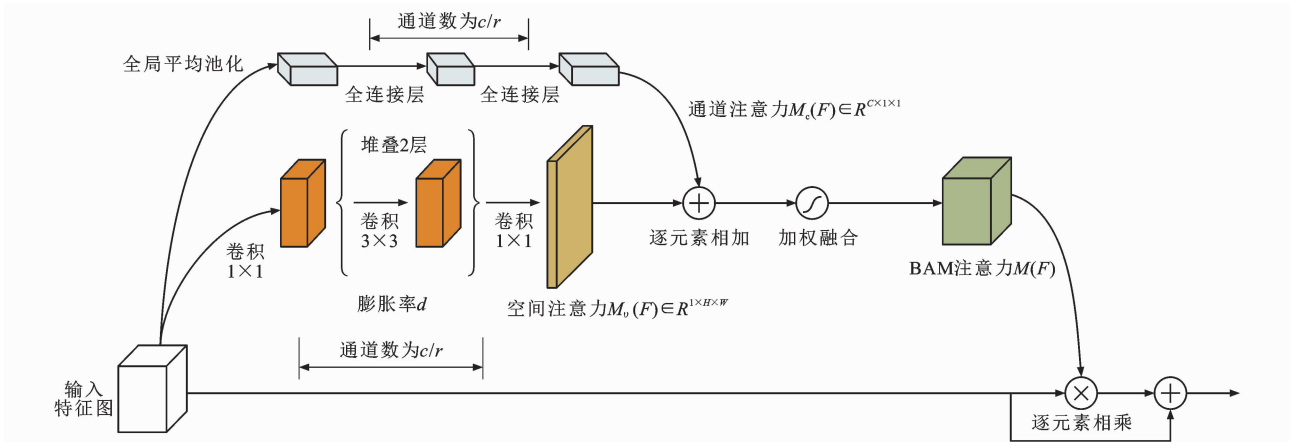


图 14 BAM 原理

Fig. 14 Principle of BAM

由图 14 可知, BAM 模块主要分为分割、融合及注意力加权三大步骤。分割操作: BAM 模块将输入特征图分割为空间注意力及通道注意力双分支, 其中空间注意力分支将输入特征图通过 1×1 卷积, 再经过 3×3 空洞卷积(膨胀率 d) 进行特征提取, 生成空间注意力图 $M_o(F) \in R^{H \times W \times 1}$ 。此外, 通道注意力分支则是对输入特征图进行全局平均池化, 生成通道维度的全局信息向量, 同时通道注意力经 2 个全连接层及非线性激活, 以生成通道注意力图 $M_c(F) \in R^{C \times 1 \times 1}$ 。融合操作: 通道注意力图 $M_c(F)$ 与空间注意力图 $M_o(F)$ 通过逐元素相加融合以生成最终注意力特征图 $M(F)$ 。该融合方式可同时结合通道及空间维度信息, 并增强模型对输入特征图表达; 将最终注意力图 $M(F)$ 应用到原始输入特征图, 可生成经加权的特征图, 该过程包括空间及通道维度的加权调整, 其中空间加权则是将空间注意力图 $M_o(F)$ 作用在输入特征图上, 此外, 通道加权则是把通道注意力图 $M_c(F)$ 对输入特征图进行通道维度逐元素加权。最终, 空间加权及通道加权结果相加, 可得到最终加权输出特征图。

此外, 在图 11 的分类分割解码头中, 还包含 1 个路径剪枝层(剪枝比例为 0.1) 及一个具有 2 个滤波器的卷积层, 其卷积核大小为 1×1 , 以得到最终特征图的分类输出结果。

综合而言, BAM 作为头部模块中的关键组件, 通过联合通道及空间注意力, 可提升模型对复杂特征的表达能力, 其通过自适应调整输入特征图中每个通道及空间位置的权重, 还可使模型更敏锐捕捉重要的语义信息, 并抑制噪声及无关特征, 模型可在路面坑洞分割中更为准确地处理细节及边界信息。

3 试验及分析

3.1 评价指标

对于本试验, 采用分割试验处理中的交并比 (Intersection over Union, IoU)、相似系数、准确率及召回率 4 种评价指标以评估网络在道路坑洞分割中的优劣。此外, 为对网络分割性能进行合理表示, 还引入混淆矩阵以表示每类别识别正确或错误的数量, 具体如表 1 所示。

表 1 混淆矩阵

Table 1 Confusion matrix

混淆矩阵		真实值	
		坑洞	背景
预测值	坑洞	N_{TP}	N_{FN}
	背景	N_{FP}	N_{TN}

表 1 中: N_{TP} 为正确预测为坑洞的像素点数量; N_{TN} 为正确预测为背景的像素点数量; N_{FP} 为背景被错误预测为坑洞像素点数量; N_{FN} 为坑洞被错误预测为背景的像素点数量。其中交并比 I_{oU} 、相似系数 D 、准确率 P 及召回率 R 计算公式分别为

$$I_{oU} = \frac{N_{TP}}{N_{TP} + N_{FP} + N_{FN}} \quad (3)$$

$$D = \frac{2N_{TP}}{N_{FP} + 2N_{TP} + N_{FN}} \quad (4)$$

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (5)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (6)$$

在道路坑洞图像分割问题中, 交并比用于衡量预测分割区域与真实坑洞区域在像素级别的重叠情况; 相似系数用于评估分割结果与真实像素的重叠度; 准确率表示预测为坑洞的像素点中, 实际属于坑

洞像素所占比例;召回率则衡量真实坑洞像素点中,正确预测为坑洞的比例。此外,还选取参数量 Param、浮点运算数 FLOPs 及检测速度作为额外评估指标。

3.2 试验平台及参数

本文试验基于 PyTorch 及 MMDetection 库实现。在网络中,每个卷积层后均采用批归一化及 ReLU 激活函数,以加速训练过程的收敛速度。对于各类主干网络编码器,其卷积层权重均由 ImageNet 预训练模型初始化,颈部卷积层权重则使用 Xavier 方法进行初始化,偏置及增益分别设置为 0 与 1。解码器卷积层权重则通过 Normal 方法初始化,均值、标准差及偏置分别设为 0、0.01 及 1,此外,解码器中还采用双线性插值进行上采样操作。训练过程中,图片分辨率均设定为 480 像素 $\times$ 480 像素,考虑到不同对比分割网络对图像预处理的兼容性,故图像仅采取了归一化预处理,学习率策略采用 Poly 方法,初始全局学习率设置为 0.000 1,最小学习率为 0.000 01,指数设为 0.9,权重衰减系数为 0.01。网络参数通过自适应动量估计(AdamW)优化算法进行更新,其中贝塔系数设定为 0.900 0~0.999 0,且每次迭代的批次大小为 4。

在普通破损、完全破损及重度破损坑洞数据集上进行试验时,训练迭代被设定为 100 回合。其中,试验环境为 Windows 10 操作系统,采用 13 代 Intel(R) Core(TM) i9-13900H 处理器及 NVIDIA GeForce RTX 4080 GPU 显卡,由于网络深度较大,试验中网络还使用了辅助损失以帮助优化学习过程,平衡辅助损失权重被设定为 0.4。

3.3 对比试验

为验证 potholes-FBConvNet 网络对坑洞的分割效果,选取了基于 Transformer 及卷积 Conv 网络的多种对比典型模型以进行对比试验,最后将各模型在 100 轮迭代中所得到的最佳权重在划分测试集及验证集上的分割效果进行汇总,后续的所有试验中,均在 3 个不同类型的坑洞数据集上分别进行独立训练与测试,以确保每个模型在对应场景下的适应性和泛化能力。

各模型在普通破损坑洞数据集上的分割表现如表 2 所示。重点对比了各网络模型在 I_{ou} 、相似系数 D 、精确率 P 及召回率 R 等多项指标上的性能表现。由表 2 可知:potholes-FBConvNet 坑洞网络在普通破损坑上的 I_{ou} 为 85.43%,为最佳模型,相比 ConvNext 的 84.94%,可提升 0.49%;与基于 Transformer 的 Segformer-MiT-B0 及 Swin-T 相

比,分别可提升 2.66%及 3.49%,展现了该网络在普通破损场景下对坑洞分割的优势;在精确率 P 上,potholes-FBConvNet 网络可达 92.96%,相较 ConvNext 的 92.14%,可提高 0.82%;与 Twins-S 及 Swin-T 相比,分别提升 0.18%及 1.6%,表明所提出的 potholes-FBConvNet 网络在分割坑洞特征时可具备更好精确性;此外,potholes-FBConvNet 召回率 R 可达 91.33%,相比 Swin-T 及 Segformer-MiT-B0,分别可提升 2.5%及 1.33%,表明该网络可更全面分割坑洞特征;尽管 potholes-FBConvNet 参数量为 134.65×10^6 ,浮点运算数 FLOPs 为 259.84×10^9 ,但相较于 ConvNext(参数量为 121.99×10^6 ,FLOPs 为 255.92×10^9)仅有小幅增加,且改进模型性能可得到显著提升;在检测速度上,potholes-FBConvNet 略低于 ConvNext,但鉴于其精度的显著提升,故仍具备最优的综合分割性能(对于检测速度,由于网络模型结构相同,故仅在普通破损坑洞类型上进行测试)。

综合而言,potholes-FBConvNet 在普通破损场景下,相比其余模型具备更优分割精度,体现了其强大的特征捕捉能力。

各模型在完全破损坑洞数据集上的分割表现则如表 3 所示,用以评估模型在大面积破损坑洞下的分割能力(表中提升效果指改进网络相对 ConvNext-B 的各参数变化幅度)。

所提出的 potholes-FBConvNet 坑洞网络 I_{ou} 可达 87.76%,为最优模型;相比 ConvNext-B 的 86.95%,可提升 0.81%;与 Swin-T 的 86.22%相比,可提升 1.54%;相较 Segformer-MiT-B0 的 85.87%,提升 1.89%;对于精确率 P ,potholes-FBConvNet 可达 92.33%,与 ConvNext-B 相比提升了 0.15%,表明该模型在检测大面积坑洞时仍可保持较高的精确性;在召回率 R 上,potholes-FBConvNet 可达 94.67%,比 ConvNext-B 提升 0.80%,并高于 Swin-T 模型 1.64%,表明改进模型可更好地识别完全破损区域,并完整捕捉坑洞的形态。在相似系数 D 及精确率 P_t 上,所提出的 potholes-FBConvNet 分别可达 93.48%及 92.79%,表现优于其他模型,具备更强的鲁棒性及泛化能力。

综合而言,potholes-FBConvNet 具备最优的精确率及召回率^[43],在完全破损坑洞场景分割任务中为最佳选择,具备优异的分割能力。

最后,表 4 为各模型在重度破损坑洞数据集上的分割表现(表中提升效果同样指改进网络相对

表 2 普通破损坑洞对比试验

Table 2 Comparison experiments of ordinary damaged potholes

模型		$I_{ou}/\%$	$D/\%$	$P/\%$	$R/\%$	$I_{ou_t}/\%$	$D_t/\%$	$P_t/\%$	$R_t/\%$	Param/ 10^6	FLOPs/ 10^9	检测速度/(帧 $\cdot s^{-1}$)
基于 Transformer	Segformer-MiT-B0 ^[27]	82.77	90.57	91.14	90.00	81.29	89.68	91.55	87.88	3.72	5.59	113.30
	Twins-S ^[28]	83.77	91.17	92.78	89.61	83.46	90.98	93.43	88.66	53.97	197.97	32.73
	Swin-T ^[29]	81.94	90.08	91.36	88.83	81.22	89.64	92.52	86.93	59.83	207.11	34.31
	DPT ^[30]	83.03	90.73	93.06	88.51	81.98	90.10	92.88	87.48	109.67	148.19	28.95
	ViT ^[31]	82.61	90.48	92.26	88.76	81.71	89.93	92.54	87.46	144.06	346.15	18.35
基于 Conv 网络	MobileNetV3 ^[32]	76.36	86.59	90.67	82.87	76.23	86.51	88.93	84.22	3.28	7.39	118.79
	Fast-SCNN ^[33]	78.40	87.89	90.33	85.58	76.74	86.84	90.93	83.10	1.45	0.80	197.32
	STDC ^[34]	76.71	86.82	89.66	84.15	73.87	84.97	85.41	84.54	8.57	7.41	166.47
	PoolFormer-S12 ^[35]	64.89	78.71	80.54	76.96	59.62	74.70	70.80	79.06	15.64	26.96	100.61
	U-Net ^[36]	69.95	82.32	88.88	76.66	63.38	77.58	89.85	68.27	29.06	178.03	31.71
	BISNetV2 ^[37]	77.00	87.01	90.33	83.92	78.97	88.25	90.03	86.53	14.76	10.81	158.21
	DeepLabV3 ^[38]	81.54	89.83	92.28	87.50	80.29	89.07	91.79	86.51	68.10	236.99	33.69
	PoinrRend ^[39]	81.80	89.99	93.23	86.97	79.79	88.76	91.44	86.23	28.70	43.10	61.66
	K-Net ^[40]	82.48	90.40	93.28	87.69	80.79	89.37	91.81	87.06	81.31	239.85	31.80
	ISANet ^[41]	81.34	89.71	91.82	87.10	80.36	89.11	91.38	86.95	37.69	132.21	44.32
	ANN ^[42]	80.46	89.17	91.55	86.92	80.14	88.98	91.97	86.17	46.22	162.45	36.76
	ConvNext-B	84.94	91.86	92.14	91.58	82.86	90.63	90.42	90.83	121.99	255.92	27.02
	改进网络	85.43	92.14	92.96	91.33	83.19	90.82	91.51	90.14	134.65	259.84	26.77
	提升效果	0.49	0.28	0.82	-0.25	0.33	0.19	1.10	-0.69	12.66	3.92	-0.25

注:带 t 下标变量代表测试集上的评估结果,后续表格均以此规则表示;提升效果指改进网络相对 ConvNext-B 的各参数变化幅度。

ConvNext-B 的各参数变化幅度),该分割任务对于模型识别精度、边缘处理能力及复杂场景适应性提出了更高要求。

由表 4 可知:potholes-FBConvNet 网络交并比为 90.76%,相比 ConvNext-B 的 90.22%提升了 0.54%;与 Swin-T 的 89.98%相比,可提升 0.78%,更适于复杂坑洞分割问题。对精确率 P ,potholes-FBConvNet 网络可达 95.41%,略高于 ConvNext-B,可提升 0.10%,表明该模型在检测重度破损坑洞时具备更高的准确性。

在召回率 R 上,potholes-FBConvNet 为 94.9%,相比 ConvNext-B 提升 0.49%,表明该模型可检测到更多的重度破损区域,尤其在坑洞形态复杂且边缘不规则的情况下,仍能保持较好的分割效果;对于相似系数 D_t 及精确率 P_t ,potholes-FBConvNet 分别可达 94.90%与 95.85%,相比 ConvNext-B,其 D_t 可提升 0.34%。

综上所述,potholes-FBConvNet 在重度破损坑洞分割研究中仍为最优模型,其在交并比、召回率及精确率等方面均有显著提升,对于坑洞边缘信息处理及分割复杂形状方面具备更为强大能力及较高的

分割精度于鲁棒性。

综合表 2~4 可知:potholes-FBConvNet 在 3 个数据集上均具备卓越的分割能力,特别在完全破损及重度破损场景中表现突出。虽然该模型浮点运算数 FLOPs 略高于基准模型,但其在精度、召回率及整体分割性能方面显著优于其他模型,特别适于要求高精度及高可靠性的坑洞分割任务,在保证精度的同时,可有效处理不同类型的道路坑洞数据,并在实际应用中具备显著优势。

3.4 消融试验

为评估改进模块对算法的提升效果,在普通破损、完全破损及重度破损道路坑洞数据集上进行了消融试验,以分析其在基准模型上分别添加 BAM 及 ASPF 模块对模型的性能影响,进而验证所提出的 potholes-FBConvNet 在 3 个不同破损场景下的有效性,其中普通破损坑洞消融试验结果如表 5 所示。

由表 5 可知:基准模型 B 的交并比及精确率分别为 84.94%及 92.14%,当加入 BAM 模块后,交并比可提升至 85.03%,但召回率 R 略微下降至 90.66%;相比基准模型,BAM 模块在精确率 P_t 上可带来 1.39%的提升,表明了BAM模块对于精确

表 3 完全破损坑洞对比试验
Table 3 Comparison experiments of completely damaged potholes %

模型		I_{oU}	D	P	R	I_{oUt}	D_t	P_t	R_t
基于 Transformer	Segformer-MiT-B0	85.87	92.40	91.75	93.06	83.28	90.88	92.27	89.53
	Twins-S	86.94	93.01	92.58	93.45	85.54	92.20	92.72	91.49
	Swin-T	86.22	92.60	92.18	93.03	82.90	90.65	91.12	90.18
	DPT	87.00	93.05	92.81	93.28	85.57	92.22	93.84	90.67
	ViT	86.56	92.79	92.32	93.27	85.37	92.11	92.94	91.29
基于 Conv 网络	MobileNetV3	76.61	86.76	86.45	87.07	76.68	83.90	89.91	83.90
	Fast-SCNN	80.84	89.40	91.35	87.54	77.44	87.28	88.71	85.90
	STDC	81.33	89.70	89.53	89.88	76.82	86.89	87.88	85.93
	PoolFormer-S12	65.47	79.13	79.97	78.30	62.19	76.69	76.65	76.72
	U-Net	71.28	83.23	88.58	78.48	67.12	80.32	85.80	75.50
	BISNetV2	77.66	87.43	89.68	85.28	75.38	85.96	90.96	81.70
	DeepLabV3	83.25	90.86	90.60	91.11	81.63	89.89	92.87	87.09
	PoinrRend	82.11	90.17	91.20	89.17	81.13	89.58	91.97	87.31
	K-Net	84.04	91.33	91.76	90.91	81.42	89.76	91.74	87.86
	ISANet	82.88	90.64	91.44	89.85	80.15	88.98	90.96	87.09
	ANN	83.81	91.19	91.71	90.68	80.26	89.05	90.89	87.27
	ConvNext-B	86.95	93.02	92.18	93.87	84.62	91.67	92.51	90.85
	改进网络	87.76	93.48	92.33	94.67	85.06	91.92	92.79	91.08
	提升效果	0.81	0.46	0.15	0.80	0.44	0.25	0.28	0.23

表 4 重度破损坑洞对比试验
Table 4 Comparison experiments of severely damage potholes %

模型		I_{oU}	D	P	R	I_{oUt}	D_t	P_t	R_t
基于 Transformer	Segformer-MiT-B0	88.82	94.08	94.09	94.07	89.54	94.48	96.53	92.52
	Twins-S	89.96	94.72	95.12	94.31	90.34	94.78	95.60	94.36
	Swin-T	89.98	94.72	95.27	94.19	89.98	94.67	96.31	93.09
	DPT	89.10	94.24	94.95	93.54	89.89	94.68	95.60	93.77
	ViT	88.73	94.03	95.03	93.05	89.43	94.42	96.25	92.65
基于 Conv 网络	MobileNetV3	83.87	91.23	91.19	91.27	84.58	91.64	92.23	91.07
	Fast-SCNN	83.88	91.23	93.75	88.84	83.86	91.22	94.70	88.00
	STDC	85.45	92.15	94.09	90.29	86.11	92.54	94.43	90.72
	PoolFormer-S12	70.48	82.68	85.45	80.09	70.53	82.72	85.80	79.85
	U-Net	76.03	86.38	88.35	84.50	72.81	84.27	93.06	76.99
	BISNetV2	84.42	91.55	93.38	89.79	85.62	92.25	92.58	91.93
	DeepLabV3	88.62	93.97	95.35	92.63	89.36	94.38	95.81	92.99
	PoinrRend	86.64	92.84	94.96	90.82	87.29	93.22	95.34	91.18
	K-Net	88.10	93.67	95.15	92.24	88.36	93.82	96.32	91.45
	ISANet	87.55	93.36	94.67	92.09	87.60	93.39	94.22	92.57
	ANN	88.07	93.65	95.32	92.05	87.91	93.57	95.27	91.92
	ConvNext-B	90.22	94.86	95.31	94.41	89.68	94.56	96.26	92.92
	改进网络	90.76	95.15	95.41	94.90	90.29	94.90	95.85	93.96
	提升效果	0.54	0.29	0.10	0.49	0.61	0.34	-0.41	1.04

表 5 普通破损坑洞消融试验

Table 5 Ablation experiments of ordinary damaged potholes

模型	$I_{oU}/\%$	$D/\%$	$P/\%$	$R/\%$	$I_{oUt}/\%$	$D_t/\%$	$P_t/\%$	$R_t/\%$	Param/ 10^6	FLOPs/ 10^9	检测速度/(帧 $\cdot s^{-1}$)
基准模型 B	84.94	91.86	92.14	91.58	82.86	90.63	90.42	90.83	121.99	255.92	27.02
B+BAM	85.03	91.91	93.19	90.66	83.13	90.79	91.81	89.78	122.07	256.57	26.71
B+ASPF	85.18	92.00	92.90	91.12	83.86	91.22	91.19	90.28	134.57	259.19	27.14
改进网络	85.43	92.14	92.96	91.33	83.19	90.82	91.51	90.14	134.65	259.84	26.77

识别坑洞的形态可起到积极作用;当 ASPF 模块加入后,模型交并比提升至 85.18%,并在精确率及召回率上取得了较为平衡提升,分别为 92.90%及 91.12%,表明 ASPF 可有效增强模型对多尺度特征的捕捉能力,进一步提升其在复杂场景下对道路普通破损坑洞的分割性能;最终,所提出的 potholes-FBConvNet 取得了 85.43% 的交并比,相比基准模型可提升 0.49%,且模型在精确率及召回率上可保持较好平衡,表明多尺度特征融合及特征适应模块的结合有助于提高模型分割的准确性及稳健性。

此外,完全破损坑洞的消融试验结果如表 6 所示,基准模型可取得 86.95% 的交并比及 93.02% 的相似系数 D ,当加入 BAM 模块后,交并比提升至 87.52%,但召回率略微下降至 93.57%,表明 BAM 可帮助模型更精确地识别完全破损区域的边界;当加入 ASPF 模块后,其交并比可达 87.31%,同时精确率及召回率也保持了较好的提升,分别为 92.75% 和 93.70%,即 ASPF 模块的多尺度上下文信息捕捉进一步加强了模型对大面积完全破损区域的检测能力;此外,potholes-FBConvNet 模型在完全破损场景下取得了 87.76% 的交并比及 94.67% 的召回率,相比基准模型,交并比可提升 0.81%,召回率可提升 0.8%;以上结果表明 BAM 与 ASPF 模块结合可显著增强模型对完全破损区域的整体检测效果,并提高模型对边缘及复杂结构的分割性能。

表 6 完全破损坑洞消融试验

Table 6 Ablation experiments of completely damage potholes %

模型	I_{oU}	D	P	R	I_{oUt}	D_t	P_t	R_t
基准模型B	86.95	93.02	92.18	93.87	84.62	91.67	92.51	90.85
B+BAM	87.52	93.34	93.11	93.57	84.44	91.57	92.85	90.32
B+ASPF	87.31	93.22	92.75	93.70	84.96	91.87	92.19	91.54
改进网络	87.76	93.48	92.33	94.67	85.06	91.92	92.79	91.08

重度破损坑洞消融试验结果如表 7 所示。基准模型可取得 90.22% 的交并比及 95.31% 的精确率,为一个较高基准;当加入 BAM 模块后,交并比可提升至 90.43%,精确率可达 95.41%,进一步增强了

表 7 重度破损坑洞消融试验

Table 7 Ablation experiments of severely damage potholes %

模型	I_{oU}	D	P	R	I_{oUt}	D_t	P_t	R_t
基准模型B	90.22	94.86	95.31	94.41	89.68	94.56	96.26	92.92
B+BAM	90.43	94.97	95.41	94.54	89.37	94.38	95.32	93.47
B+ASPF	90.56	95.04	95.23	94.86	90.00	94.74	95.60	93.90
改进网络	90.76	95.15	95.41	94.90	90.29	94.90	95.85	93.96

模型对重度破损边缘的识别能力;ASPF 模块加入后,可使其交并比提升至 90.56%,召回率达 94.86%,表明 ASPF 模块可帮助模型捕捉到更多的全局特征,对重度破损区域的边缘细节处理具有积极作用;而所提出的 potholes-FBConvNet 模型取得了 90.76% 的交并比与 95.41% 的精确率,相比基准模型,其交并比提升 0.54%,在保持较高精确率的同时,召回率也可达 94.90%,表明模型在处理形状复杂及边缘模糊重度破损坑洞场景时具备更为优异性能。

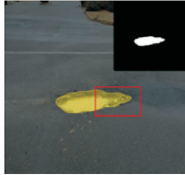
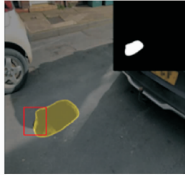
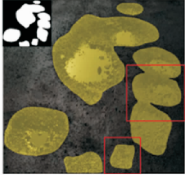
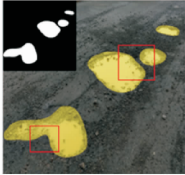
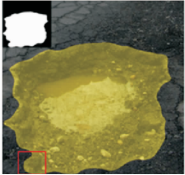
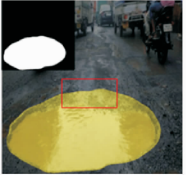
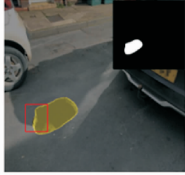
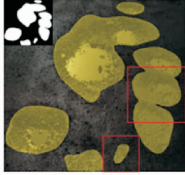
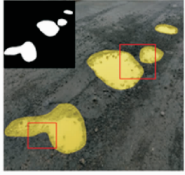
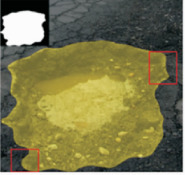
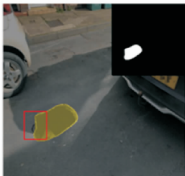
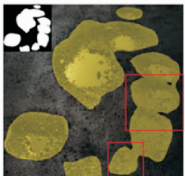
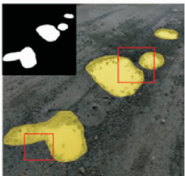
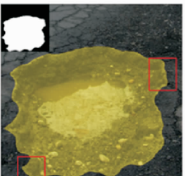
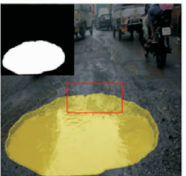
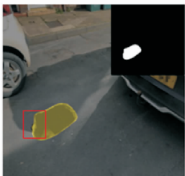
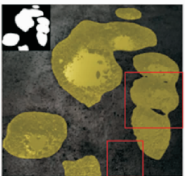
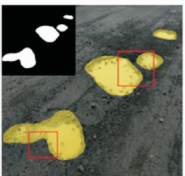
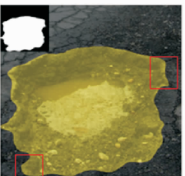
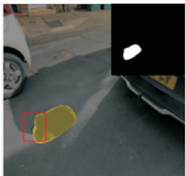
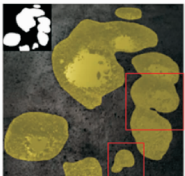
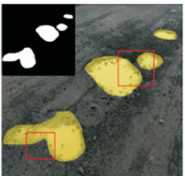
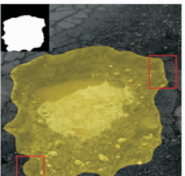
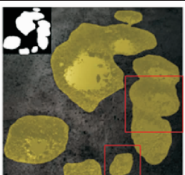
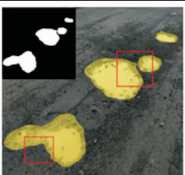
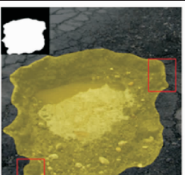
综合而言,上述消融试验结果表明在普通破损、完全破损及重度破损三类坑洞分割任务中,BAM 及 ASPF 模块均可显著提升模型分割性能。特征适应 BAM 模块可帮助模型在精确识别坑洞形态及边界上取得更优的表现;ASPF 模块通过多尺度特征融合,可有效增强模型在复杂场景中的分割能力;最终结合以上 2 个模块的坑洞分割网络可取得最佳结果,展示了其在处理各类道路坑洞分割中的优越性,尤其在精确率及召回率之间可达到较好平衡,适用于实际复杂环境应用下的高精度道路坑洞分割。

3.5 道路坑洞分割可视化

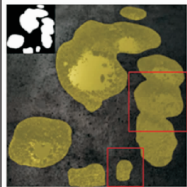
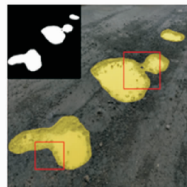
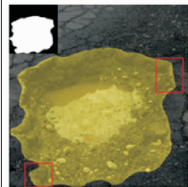
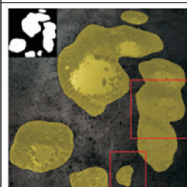
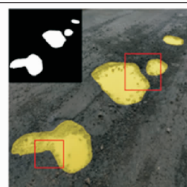
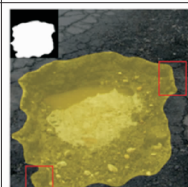
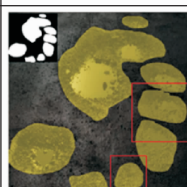
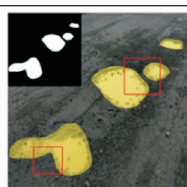
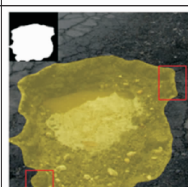
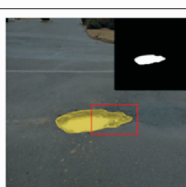
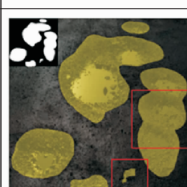
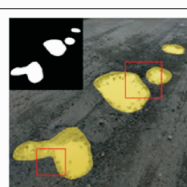
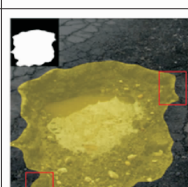
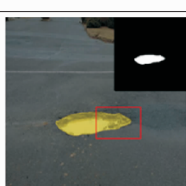
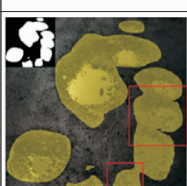
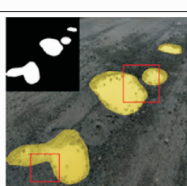
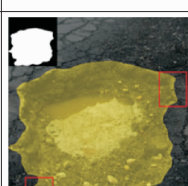
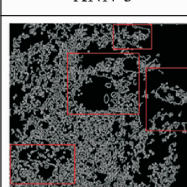
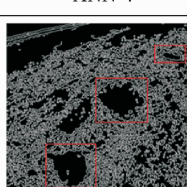
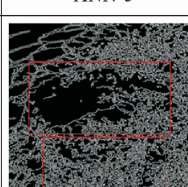
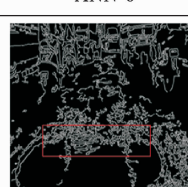
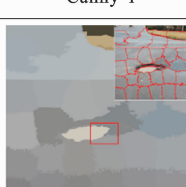
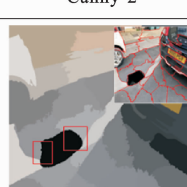
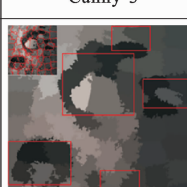
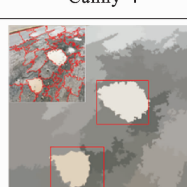
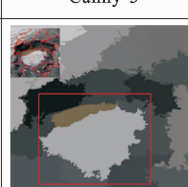
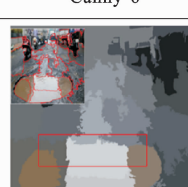
为验证改进网络 potholes-FBConvNet 在道路坑洞分割中的优化效果,从各类型坑洞数据库测试集中随机选取 3 张典型坑洞图片进行分割可视化对比(同时选取传统分割算法进行可视化对比),各网络模型均可分割到坑洞具体位置,但分割区域、真实标签匹配程度及边缘连续性上存在部分差异;传统算法存在可视化结果见表 8(仅对部分典型模型的分割效果进行对比)。

表 8 各典型算法坑洞分割效果可视化对比

Table 8 Visual comparison of potholes segmentation effects of various typical algorithms

坑洞	普通破损坑洞		完全破损坑洞		重度破损坑洞	
初始图像						
	普通坑洞1	普通坑洞2	完全坑洞3	完全坑洞4	重度坑洞5	重度坑洞6
真值标签图						
	普通坑洞标签1	普通坑洞标签2	完全坑洞标签3	完全坑洞标签4	重度坑洞标签5	重度坑洞标签6
potholes-FBConvNet						
	坑洞网络-1	坑洞网络-2	坑洞网络-3	坑洞网络-4	坑洞网络-5	坑洞网络-6
ConvNext-B						
	ConvNext-B-1	ConvNext-B-2	ConvNext-B-3	ConvNext-B-4	ConvNext-B-5	ConvNext-B-6
U-Net						
	U-Net-1	U-Net-2	U-Net-3	U-Net-4	U-Net-5	U-Net-6
DPT						
	DPT-1	DPT-2	DPT-3	DPT-4	DPT-5	DPT-6
ViT						
	ViT-1	ViT-2	ViT-3	ViT-4	ViT-5	ViT-6
Segformer-MiT-B0						
	Segformer-1	Segformer-2	Segformer-3	Segformer-4	Segformer-5	Segformer-6

续表 8

坑洞	普通破损坑洞		完全破损坑洞		重度破损坑洞	
Swin-T						
	Swin-T-1	Swin-T-2	Swin-T-3	Swin-T-4	Swin-T-5	Swin-T-6
Twins-S						
	Twins-S-1	Twins-S-2	Twins-S-3	Twins-S-4	Twins-S-5	Twins-S-6
DeepLabV3						
	DeepLabV3-1	DeepLabV3-2	DeepLabV3-3	DeepLabV3-4	DeepLabV3-5	DeepLabV3-6
K-Net						
	K-Net-1	K-Net-2	K-Net-3	K-Net-4	K-Net-5	K-Net-6
ANN						
	ANN-1	ANN-2	ANN-3	ANN-4	ANN-5	ANN-6
Canny算子						
	Canny-1	Canny-2	Canny-3	Canny-4	Canny-5	Canny-6
超像素分割						
	$n=40, c=10$ 超像素分割-1	$n=40, c=10$ 超像素分割-2	$n=65, c=15$ 超像素分割-3	$n=40, c=8$ 超像素分割-4	$n=30, c=10$ 超像素分割-5	$n=30, c=12$ 超像素分割-6

由表 8 可知:传统图像分割方法 Canny 算子及超像素分割检测方法,在多场景中存在较大局限性,分割效果较差。Canny 算子虽然能够有效检测边

缘,但对噪声敏感,且提取的边缘信息不连续,难以形成完整的坑洞轮廓;而超像素分割虽能提高分割精度,但其最佳参数选择困难,无法同时准确分割不

同尺寸坑洞,且在坑洞边缘精确分割与处理相邻坑洞模糊边界时表现不足(表8超像素分割中 n 及 c 分别为超级像素个数与像素紧凑程度)。此外,基于深度学习的各对比模型可能在某些坑洞类型上的分割表现较好,但在处理复杂坑洞场景时依然存在泛化能力不足,难以同时实现精准及全面的分割,尤其多坑洞及背景复杂情况,部分模型甚至出现较大区域漏分割。

在普通破损坑洞场景中,potholes-FBConvNet可精准分割出坑洞区域,分割轮廓与真实标签高度吻合。相比之下,ConvNext-B则在细节处理上则稍显不足;而U-Net由于受到背景噪声干扰,导致了在第2张测试图像上的分割结果出现了偏差;由于普通道路坑洞形状较为单一,因此其余模型均可近乎标签的将其分割,且差异不大。

在完全破损坑洞场景中,potholes-FBConvNet凭借其多尺度特征融合机制,可准确捕捉坑洞整体轮廓,并有效避免背景误分。其中ConvNext-B在第3张图片测试中分割不太全面,且其余较多模型均难以达到较好分割效果,未能将坑洞区域边缘进行较好分割,存在误分割现象,其中K-Net及ANN等网络均无法对所有坑洞达到全部准确分割,在处

理多个相邻坑洞时易出漏分割及过度分割现象,即面对复杂背景及多坑洞时,无法对坑洞边缘进行准确分割。而potholes-FBConvNet则在坑洞分割边缘连续性、切合度方面表现更为细致,展现了优异的泛化性及鲁棒性。

对于重度破损坑洞,potholes-FBConvNet分割网络仍表现出较优的细节捕捉及边缘处理能力,其分割结果与标签图近乎完全一致,并可有效抑制噪声,精准区分坑洞与背景。相比之下,U-Net边缘处理不太连续;DPT及ViT分割效果较为一般,其中ViT甚至对抗洞边缘出现了误分割。

通过表8的分割效果可视化对比,3种典型场景中的potholes-FBConvNet分割效果均优于各对比模型,通过引入多尺度特征融合及特征适应模块,其可更好理解全局及局部特征;此外,该模型在细节处理及背景噪声抑制方面具备优异表现,尤其在复杂背景完全破损坑洞下,具备更强分割泛化能力及鲁棒性。

为进一步探索网络的可解释性,在图15、16中对改进的坑洞分割模型potholes-FBConvNet与基础模型的几个选定阶段进行可视化展示。由于特征通道数量较多,故仅计算每个阶段均值进行展示输出。

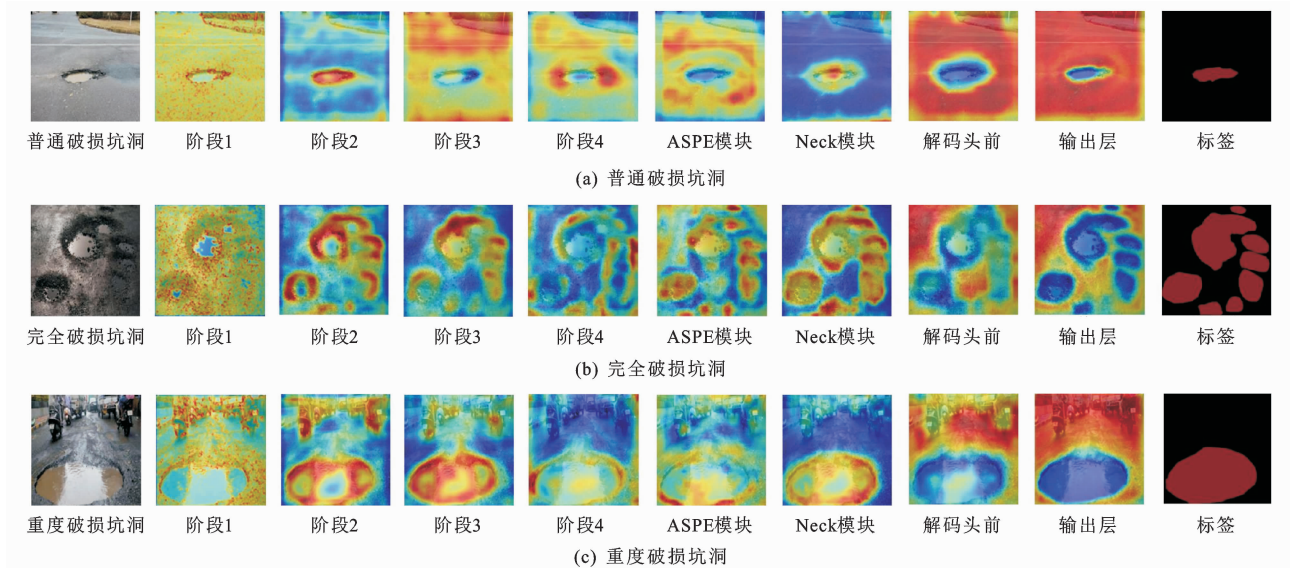


图15 potholes-FBConvNet在不同类型坑洞的特征图可视化结果

Fig. 15 Visualization results of feature maps of different types of potholes using potholes-FBConvNet

由图15、16可知:基于卷积神经网络的两模型在寻找物体边缘方面较为有效,而注意力机制的引入则更擅长寻找标记之间的关系,但两者均在更深阶段探索到了更高层次的语义信息。其中图15、16的第6列显示了ASPF模块的输出与UPerNet结构中金字塔池化(Pyramid Pooling Module, PPM)

模块的输出,如该列所示,所提出的ASPF模块促使网络可捕捉更多的上下文多尺度信息,有助于区分坑洞和背景信息。相比之下,UPerNet中的PPM模块采用池化操作,导致收集到较多的全局信息,并混合了细节。由图15、16第8列可知,在BAM模块之后,通道注意力使得特征图通道得到了重新分

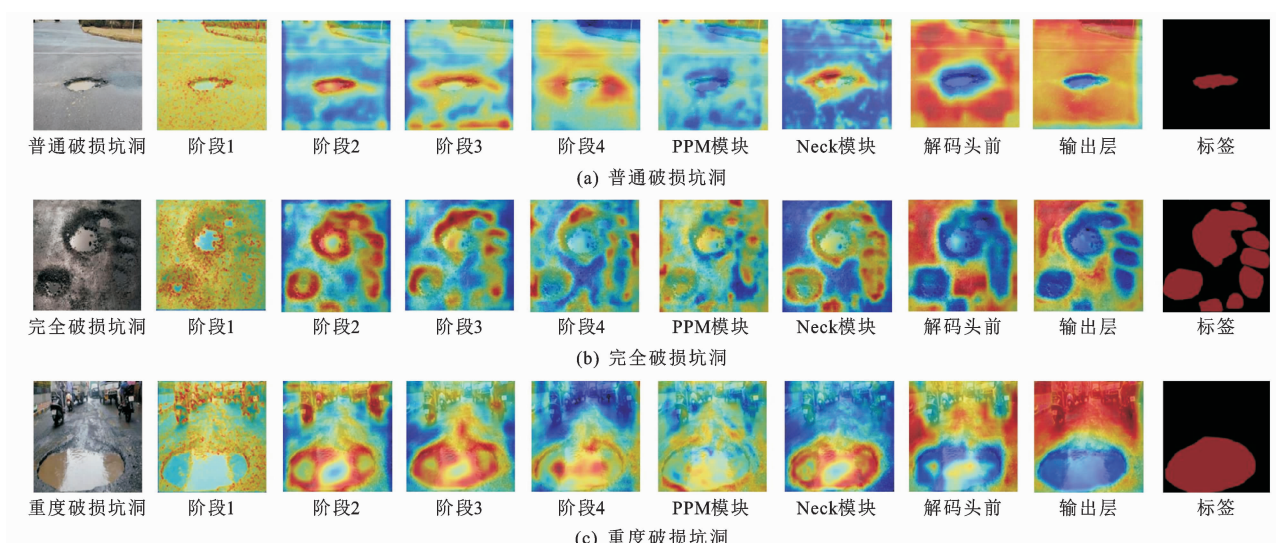


图16 基础模型在不同类型坑洞的特征图可视化结果

Fig. 16 Visualization results of feature maps of different types of potholes using basic model

配,因此每个通道作用更为明显,使得解码头在坑洞信息的分类方面具有更好效能。

4 结 语

(1)将路面坑洞分为普通破损、完全破损及重度破损3类场景,并构建对应的道路坑洞分割数据库。基于该特征库,提出一种道路坑洞分割网络 potholes-FBConvNet,通过设定超参数及网络迭代以完成模型训练,实现了对道路不同类型坑洞的准确分割。

(2)所提出的模型通过主干 ConvNet 网络以提取特征,结合空洞空间金字塔融合模块用于捕捉多尺度上下文信息,并引入特征适应模块,用于自适应调整特征图通道及空间维度,能有效应对道路坑洞分割的复杂性及多样性。

(3)在3类坑洞分割数据库上进行的训练测试结果表明,potholes-FBConvNet 相比16种典型分割模型,在精度、召回率及边缘处理能力方面均具备显著优势,尤其在复杂场景及多坑洞分布情况下,具备最优分割效果,且该网络架构具备较强灵活性,可迁移应用到类似的分割场景。

(4)对道路坑洞应用深度学习分割技术具有广泛应用前景,后续研究还可对坑洞分割数据库进行不断扩充,尤其是不同光照条件及道路类型的坑洞样本,以持续提升模型泛化分割能力。此外,结合分割结果,可为车速控制及路径规划提供数据支持,提升行车安全性及用户体验,实现智能决策规划。同时,还可将模型逐步部署至实际系统,以评估其在边缘设备上的效果,使其可更好适应

低功耗设备的需求。

参 考 文 献 :

References :

- [1] 姚楚溪,蔡皓楠,张远波,等. 基于轻量化车载设备的道路病害检测方法[J]. 测绘通报, 2024(5): 147-150.
YAO Chu-xian, CAI Hao-nan, ZHANG Yuan-bo, et al. Road damage detection method based on lightweight vehicle equipment[J]. Bulletin of Surveying and Mapping, 2024(5): 147-150.
- [2] 崔新壮,黄丹,刘磊,等. 沥青路面病害力学研究进展[J]. 山东大学学报(工学版), 2016, 46(5): 68-87.
CUI Xin-zhuang, HUANG Dan, LIU Lei, et al. A review of mechanics of asphalt pavement disease[J]. Journal of Shandong University (Engineering Science), 2016, 46(5): 68-87.
- [3] 管进超,丁玲,杨旭,等. 多维图像融合驱动的复杂场景路表破损识别[J]. 交通运输工程学报, 2024, 24(3): 154-170.
GUAN Jin-chao, DING Ling, YANG Xu, et al. Pavement surface distress detection in complex scenarios driven by multi-dimensional image fusion[J]. Journal of Traffic and Transportation Engineering, 2024, 24(3): 154-170.
- [4] 《中国公路学报》编辑部. 中国路面工程学术研究综述·2024[J]. 中国公路学报, 2024, 37(3): 1-81.
Editorial Department of China Journal of Highway and Transport. Review on China's pavement engineering research: 2024[J]. China Journal of Highway and Transport, 2024, 37(3): 1-81.
- [5] 《中国公路学报》编辑部. 中国路面工程学术研究综述·2020[J]. 中国公路学报, 2020, 33(10): 1-66.
Editorial Department of China Journal of Highway and Transport. Review on China's pavement engineering research · 2020[J]. China Journal of Highway and Transport, 2020, 33(10): 1-66.
- [6] GAO M X, WANG X, ZHU S L, et al. Detection and

- segmentation of cement concrete pavement pothole based on image processing technology[J]. *Mathematical Problems in Engineering*, 2020(1): 1360832.
- [7] ABBAS I H, ISMAEL M Q. Automated pavement distress detection using image processing techniques[J]. *Engineering, Technology and Applied Science Research*, 2021, 11(5): 7702-7708.
- [8] KIM S, SEO D, JEON S. Improvement of tiny object segmentation accuracy in aerial images for asphalt pavement pothole detection[J]. *Sensors*, 2023, 23(13): 5851.
- [9] YE W L, JIANG W, TONG Z, et al. Convolutional neural network for pothole detection in asphalt pavement[J]. *Road Materials and Pavement Design*, 2021, 22(1): 42-58.
- [10] WANG A D, LANG H, CHEN Z, et al. The two-step method of pavement pothole and raveling detection and segmentation based on deep learning[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2024, 25(6): 5402-5417.
- [11] LAKMAL H K I S, DISSANAYAKE M B. Pothole detection with image segmentation for advanced driver assisted systems[C]//IEEE. 2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE). New York: IEEE, 2021: 308-311.
- [12] 袁朝春,王俊娴,何友国,等.基于路面坑洞检测的智能汽车主动避障控制[J]. *江苏大学学报(自然科学版)*, 2022, 43(5): 504-511, 518.
- YUAN Chao-chun, WANG Jun-xian, HE You-guo, et al. Active obstacle avoidance control of intelligent vehicle based on pothole detection[J]. *Journal of Jiangsu University (Natural Science Edition)*, 2022, 43(5): 504-511, 518.
- [13] 宣以国,余成波,蒋启超,等.基于改进 YOLOv7 的道路裂缝和坑洞检测算法[J]. *科学技术与工程*, 2024, 24(17): 7205-7213.
- XUAN Yi-guo, YU Cheng-bo, JIANG Qi-chao, et al. Improved YOLOv7 road crack and pothole detection algorithm[J]. *Science Technology and Engineering*, 2024, 24(17): 7205-7213.
- [14] 王 鑫,李 琦.基于深度学习的路面缺陷自动检测系统[J]. *光电子·激光*, 2022, 33(11): 1165-1172.
- WANG Xin, LI Qi. Automatic detection of pavement defects based on deep learning[J]. *Journal of Optoelectronics · Laser*, 2022, 33(11): 1165-1172.
- [15] 胡晓伟,闫奕昕,王大为,等.基于 YOLOM 算法的路面病害轻量化检测方法[J]. *中国公路学报*, 2024, 37(12): 381-391.
- HU Xiao-wei, YAN Yi-xin, WANG Da-wei, et al. Lightweight pavement disease detection based on YOLOM algorithm[J]. *China Journal of Highway and Transport*, 2024, 37(12): 381-391.
- [16] 罗向龙,王彦博,蒲亚亚,等.多病害并发复杂场景下的道路病害检测 RGT-YOLOv7 模型[J]. *湖南大学学报(自然科学版)*, 2024, 51(12): 107-118.
- LUO Xiang-long, WANG Yan-bo, PU Ya-ya, et al. Road disease detection RGT-YOLOv7 model under multiple diseases complicated scenarios[J]. *Journal of Hunan University (Natural Sciences)*, 2024, 51(12): 107-118.
- [17] 郭道俊,周幸宇,许崇帮,等.基于 DeepLabV3+网络改进的快速隧道病害分割方法[J]. *公路交通科技*, 2023, 40(8): 127-135.
- GUO Dao-jun, ZHOU Xing-yu, XU Chong-bang, et al. Improved fast tunnel disease segmentation method based on DeepLabV3+ network[J]. *Journal of Highway and Transportation Research and Development*, 2023, 40(8): 127-135.
- [18] 任泳洁,吴立朋.融合注意力机制和新型卷积神经网络的市政道路病害识别[J]. *科学技术与工程*, 2024, 24(20): 8663-8672.
- REN Yong-jie, WU Li-peng. Municipal road disease recognition based on attention mechanism and novel convolutional neural network[J]. *Science Technology and Engineering*, 2024, 24(20): 8663-8672.
- [19] LI D R, DUAN Z D, HU X Y, et al. Automated classification and detection of multiple pavement distress images based on deep learning[J]. *Journal of Traffic and Transportation Engineering (English Edition)*, 2023, 10(2): 276-290.
- [20] GUAN T, CAI J Y, WANG Y, et al. Pavement pothole detection system based on deep learning and binocular vision[J]. *Journal of Traffic and Transportation Engineering (English Edition)*, 2025, 12(4): 1100-1123.
- [21] LIU Z, MAO H Z, WU C Y, et al. A ConvNet for the 2020s[C]//IEEE. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2022: 11966-11976.
- [22] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[PP/OL]. V7. ArXiv (2023-08-02). <https://arxiv.org/abs/1706.03762>.
- [23] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//IEEE. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2016: 770-778.
- [24] XIAO T T, LIU Y C, ZHOU B L, et al. Unified perceptual parsing for scene understanding[C]//Springer. Computer Vision-ECCV 2018. Munich: Springer, 2018: 432-448.
- [25] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[PP/OL]. V3. ArXiv (2017-12-05). <https://arxiv.org/abs/1706.05587>.
- [26] PARK J, WOO S, LEE J Y, et al. BAM: Bottleneck attention module[PP/OL]. V2. ArXiv (2018-07-18). <https://arxiv.org/abs/1807.06514>.
- [27] XIE E Z, WANG W H, YU Z D, et al. SegFormer: Simple and efficient design for semantic segmentation with Transformers[PP/OL]. V3. ArXiv (2021-10-28). <https://arxiv.org/abs/2105.15203>.
- [28] CHU X X, TIAN Z, WANG Y Q, et al. Twins: Revisiting the design of spatial attention in vision transformers[PP/OL]. V4. ArXiv (2021-09-30). <https://arxiv.org/abs/2104.13840>.
- [29] LIU Z, LIN Y T, CAO Y, et al. Swin Transformer: Hierarchical vision transformer using shifted windows[C]//IEEE. 2021 IEEE/CVF International Conference on Computer

- Vision (ICCV). New York: IEEE, 2022: 9992-10002.
- [30] RANFTL R, BOCHKOVSKIY A, KOLTUN V. Vision Transformers for dense prediction[C]//IEEE. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). New York: IEEE, 2022: 12159-12168.
- [31] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale[PP/OL]. V2. ArXiv (2021-06-03). <https://arxiv.org/abs/2010.11929>.
- [32] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3[C]//IEEE. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). New York: IEEE, 2019: 1314-1324.
- [33] POUDEL R P K, LIWICKI S, CIPOLLA R. Fast-SCNN: Fast semantic segmentation network[PP/OL]. V1. ArXiv (2019-02-12). <https://arxiv.org/abs/1902.04502>.
- [34] FAN M Y, LAI S Q, HUANG J S, et al. Rethinking BiSeNet for real-time semantic segmentation[C]//IEEE. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2021: 9711-9720.
- [35] YU W H, LUO M, ZHOU P, et al. MetaFormer is actually what you need for vision[C]//IEEE. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2022: 10809-10819.
- [36] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]//Springer. Medical Image Computing and Computer-assisted Intervention-MICCAI 2015. Munich: Springer, 2015: 234-241.
- [37] YU C Q, GAO C X, WANG J B, et al. BiSeNet V2: Bilateral network with guided aggregation for real-time semantic segmentation[J]. International Journal of Computer Vision, 2021, 129(11): 3051-3068.
- [38] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Springer. Computer Vision-ECCV 2018. Munich: Springer, 2018: 833-851.
- [39] KIRILLOV A, WU Y X, HE K M, et al. PointRend: Image segmentation as rendering[C]//IEEE. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2020: 9796-9805.
- [40] ZHANG W W, PANG J M, CHEN K, et al. K-net: Towards unified image segmentation[PP/OL]. V2. ArXiv (2021-11-01). <https://arxiv.org/abs/2106.14855>.
- [41] JIN Z C, LIU B, CHU Q, et al. ISNet: Integrate image-level and semantic-level context for semantic segmentation[C]//IEEE. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). New York: IEEE, 2022: 7169-7178.
- [42] ZHU Z, XU M D, BAI S, et al. Asymmetric non-local neural networks for semantic segmentation[C]//IEEE. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). New York: IEEE, 2019: 593-602.
- [43] 朱彦洁,王瑜晨,熊文,等. 基于桥检领域-任务迁移的检测报告信息提取少样本模型[J]. 交通运输工程学报, 2025, 25(1): 248-262.
- ZHU Yan-jie, WANG Yu-chen, XIONG Wen, et al. Few-shot model for extracting inspection report information based on bridge inspection domain-task transfer[J]. Journal of Traffic and Transportation Engineering, 2025, 25(1): 248-262.