DOI: 10.16452/j.cnki.sdkjzk.2025.05.005 文章编号:1672-3767(2025)05-0052-12

HR-DeepLabV3+:融合多尺度上下文与注意力机制的道路提取模型

马召恒[1],张 雷[1],刘如飞[2],冯 飞[1],王友雷[1]
(1. 河北省地理信息集团有限公司,河北 石家庄 050031;
2. 山东科技大学 测绘与空间信息学院,山东 青岛 266590)

摘 要:城市路网是城市发展的重要基础设施,也是社会经济活动的重要载体。但现有遥感影像路网提取方法在应对高楼阴影、植被遮挡等复杂环境时往往效果不佳。为解决该问题,提出一种 HR-DeepLabV3+深度学习模型,主干网络设计采用 HRNetV2-W18(high-resolution network W-18),用于保留高分辨率影像中的细节信息;解码器中设计了多层级融合的空洞空间金字塔结构,并添加多头注意力结构,提升对图像上下文信息的学习能力;损失函数中通过降低背景环境特征信息的权重,更加关注细小的道路目标,从而降低背景复杂环境因素的干扰。在公开数据集 CHN6-CUG、DeepGlobe 和 SJZ Road 上进行实验分析表明,该网络的 I_{OU} 精度分别达 0.711、0.685 和 0.751,相比其他主流网络效果更稳健。

关键词:城市路网;遥感识别;深度学习;空间金字塔;路网提取
中图分类号:P236 **文献标志码**:A

HR-DeepLabV3+: Amulti-scale context fusion network model with attention mechanism for road extraction

MA Zhaoheng[1], ZHANG Lei[1], LIU Rufei[2], FENG Fei[1], WANG Youlei[1]
(1. Hebei Geographic Information Group Co., Ltd., Shijiazhuang 050031, China;
2. College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao 266590, China)

Abstract: Urban road networks are critical infrastructure for city development and serve as essential carriers for socio-economic activities. However, the existing road extraction methods based on remote sensing imagery often perform poorly under complex conditions with tall building shadows and vegetation occlusion. To address this challenge, we proposed an advanced deep learning model named HR-DeepLabV3+. The backbone was designed using HRNetV2-W18 (High-Resolution Network W-18) to preserve the fine-grained details in high-resolution imagery. The decoder incorporates a multi-level fused atrous spatial pyramid structure along with a multi-head attention mechanism to enhance the model's ability to capture contextual information. The loss function was optimized by reducing the weight of background features to guide the model to focus more on subtle road structures and mitigate the interference from complex background elements. Extensive experiments conducted on publicly available datasets of CHN6-CUG, DeepGlobe, and SJZ Road demonstrated that the proposed method achieved Intersection over Union (I_{OU}) scores of 0.711, 0.685, and 0.751, respectively. These results indicate that HR-

收稿日期:2024-06-07
基金项目:国家自然科学基金项目(42171439);山东省科技型中小企业创新能力提升工程项目(2022TSGC1135);山东科技大学"菁英计划"科研支持项目(0104060541613)
作者简介:马召恒(1985—),男,河北石家庄人,高级工程师,主要从事遥感图像处理与激光点云数据处理.
刘如飞(1986—),男,江苏南京人,副教授,博士后,主要从事多平台激光点云数据处理、点云与图像目标特征识别、三维自动建模与 GIS 应用方面的研究,本文通信作者. E-mail:liurufei_2007@126.com

DeepLabV3+ exhibits more robust and reliable performance compared to other state-of-the-art networks.

Key words: urban road network; remote sensing recognition; deep learning; spatial pyramid; road network extraction

随着城市经济及规模的快速发展,国土空间监测成为至关重要的任务之一,运用卫星遥感图像进行城市路网监测成为研究热点。传统的道路路网提取方法通常先采用人工标注提取样本特征,再通过监督学习方法来识别道路区域[1]。深度学习技术[2]在遥感影像处理中的广泛应用,为路网提取提供了新的技术路线。

目前,针对路网提取中存在的高大建筑物阴影或树木遮挡干扰问题,国内外众多学者进行了相关研究,主要通过优化网络结构或改进损失函数等方法,提升模型在复杂场景下的特征判别能力和遮挡区域的恢复精度。Diakogiannis 等[3]通过扩张卷积和加深模型架构增强了编码器的特征提取能力,并通过改进损失函数缓解了道路和背景类别之间的不平衡。Hetang 等[4]以分割一切模型(segment anything model,SAM)为基础,通过添加轻量级的基于 Transformer 的图神经网络来预测大区域的图顶点和边以推算出道路,避免了复杂的后处理算法,但无法准确处理立交桥道路。Xu 等[5]增强了 RNGDetby 算法,在其基础上添加实例分割头,并利用骨干网络的多尺度特征监督训练,但提取效率较低且无法处理复杂的道路交叉口或立交桥图像。Qiu 等[6]提出具有两个分支主干的 SGNet 框架进行道路提取,并集成特征细化模块增强道路细节和纹理特征,但在部分道路的连通性上丢失像素。上述方法通过优化网络结构及损失函数使模型在精度和全局建模方面具有较高优势,但在应对建筑物阴影或树木遮挡问题上仍有不足。

此外,一些学者在模型中引入注意力机制优化其性能,提升在阴影和遮挡环境下的识别能力。如王谦等[7]提出一种 CE-DeepLabv3+网络,加入 ECANet 注意力机制和改进损失函数来解决道路与背景不平衡问题,但当道路和背景颜色、纹理相似时,提取道路存在一定难度。Chen 等[8]提出一种基于多尺度移动的注意力检测方法。赫晓慧等[9]为解决道路提取中结构完整性问题,提出使用 DCNN 与短距条件随机场相结合的方法来增强上下文信息提取能力,但在乡村道路中仍存在部分断连情况。Dai 等[10]提出一种道路增强可变形注意力网络来学习特定道路像素的长距离依赖关系,并添加道路增强模块提取道路特征,实现多尺度道路语义信息的提取。Chen 等[11]基于 ResNet 主干网络构建空间注意力模型,通过在通道和空间两个维度引入注意力机制,增强模型在遮挡条件下的特征提取能力,但在处理大面积阴影覆盖或密集植被遮挡等复杂场景时,道路边缘的提取精度仍有待提高。此外,当道路与周边建筑物纹理相近时,该模型仍会出现误检和漏检现象,特别是在城市高密度建成区的道路提取任务中表现不够稳定。

受空洞空间金字塔池化(atrous spatial pyramid pooling, ASPP)与多头注意力机制原理的启发,本研究提出一种基于 HRNet 的路网提取模型——HR-DeeplabV3+:通过引入多层次融合的空洞空间金字塔池化结构,提升模型对道路特征的学习能力;结合多头注意力机制(multi-head attention,MHA),有效增强上下文特征的提取效果;同时设计了一种自适应调整的损失函数,降低背景环境特征信息的权重。

1 HR-DeepLabV3+模型整体架构

现有分割网络在路网提取中常受固定感受野限制,难以兼顾道路连通性与边缘细节:大感受野虽有利于捕捉长距离结构,但易忽略边缘细节;小感受野则无法有效感知道路拓扑结构。为此,为重点解决高楼阴影和遮挡环境下的城市路网提取问题,本研究以 HRNetV2-W18 为主干网络,利用其高分辨率保持与多尺度特征交互能力,有效提取丰富道路信息。并对 HRNetV2-W18 进行三方面的改进:①引入多层级空洞空间金字塔池化模块,通过不同膨胀率融合多尺度上下文,提升遮挡区域下的道路识别连通性;②融合多头注意力机制,增强模型对遮挡环境下路网的感知能力;③设计自适应加权损失函数,动态调整前背景权重,对遮挡区域路段赋予更大权重进行计算,上述模块协同构成 HR-DeepLabV3+整体架构,如图 1 所示。

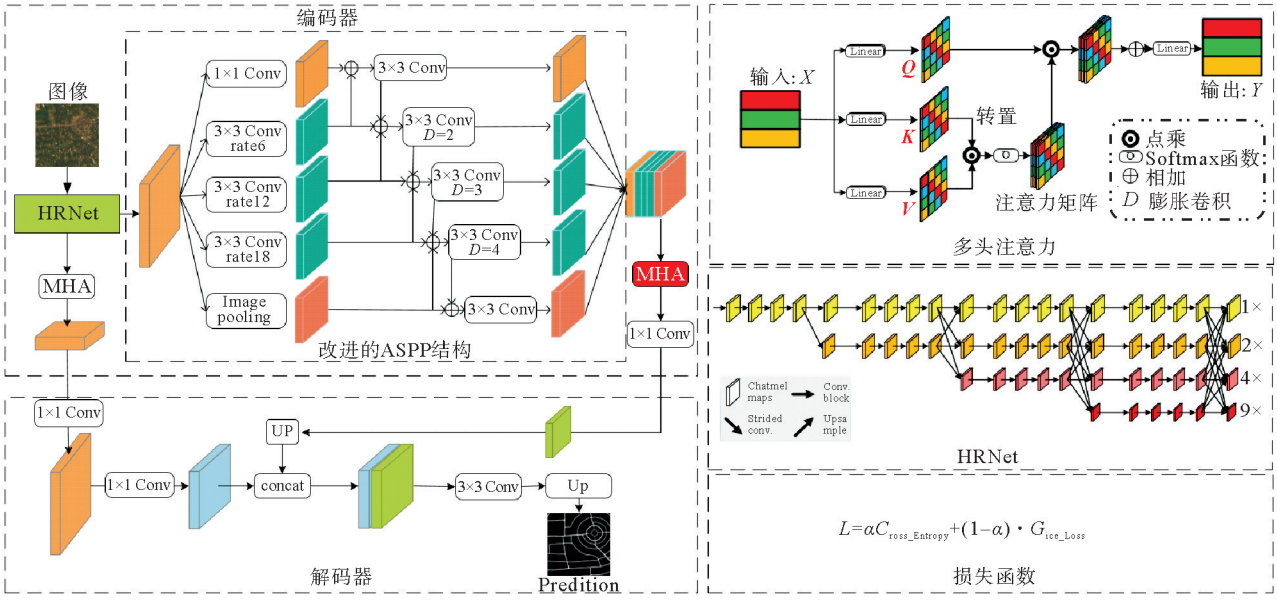


图 1 HR-DeepLabV3＋网络结构

Fig. 1 HR-DeepLabV3＋ network structure

2 HR-DeepLabV3＋模型核心模块设计

2.1 多层级融合的空洞空间金字塔模块

在道路提取任务中,模型不仅需要具备对道路全局结构的建模能力,还必须对局部细节信息保持高度敏感,才能较好地适应城市高楼阴影、植被遮挡等场景。原始 DeepLabV3＋ 网络中的空洞空间金字塔池化(ASPP)模块采用固定空洞率的设置,在下采样过程中容易导致高维特征图的细节信息严重丢失,如图 2 所示。

为进一步提升特征表达能力并缓解该问题,提出一种多层级融合的空洞空间金字塔池化模块,通过引入多尺度空洞率和特征融合机制,增强对多尺度上下文信息的建模能力,并保留关键细节特征。将空洞空间金字塔池化模块结构中四个卷积层和池化层进行交互学习(如图 3 所示),接收 HRNet 主干网络中的高维信息,分别进行一个 1×1 卷积、三个 3×3 卷积与一个全局池化,其中在 3×3 卷积中分别以不同的空洞卷积率进行卷积,目的是扩大模型的感受野。经过卷积后,将五个并列分支进行交互学习,步骤为:①将分支 f_2 特征与 f_1 特征相加,经过 3×3 卷积进一步提取特征,将卷积后的特征图与 f_2 分支进行交互;②在 f_2 分支中,将 f_3 分支和 f_1 合并后特征相加,进一步增强 f_2 的特征;③在 f_3 分支中,将 f_2 分支融合原始特征的低空洞率特征相加用来补充 f_3 分支大空洞率的特征图;④对于在 f_4 分支中,先采用相同方法融合低空洞卷积下的特征图 f_5 和 f_3 分支的特征;⑤将 f_4 分支特征图信息添加到 f_5 中,保留原始特征,防止信息损失。

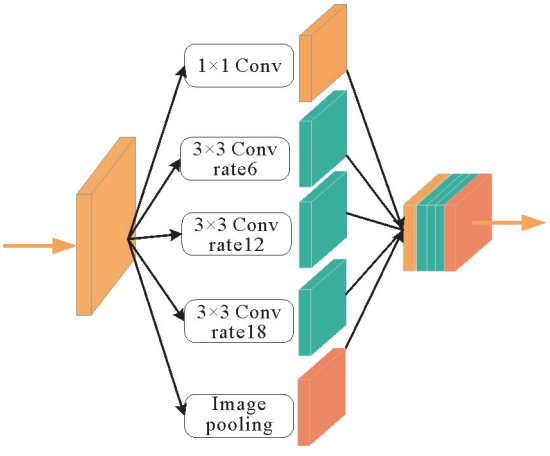


图 2 改进前结构

Fig. 2 Structure before improvement

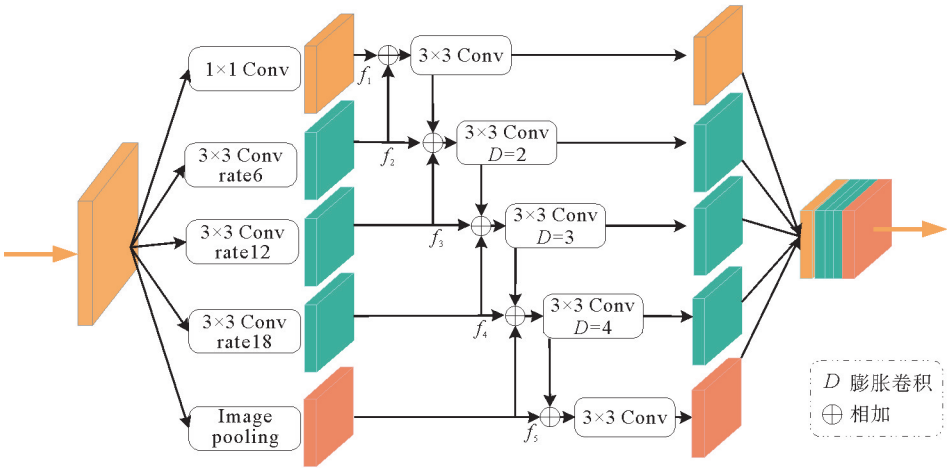


图 3　改进后的空洞空间金字塔池化结构

Fig. 3　Improved ASPP structure

2.2　多头注意力机制

在道路受遮挡的情况下,阴影或遮挡物两侧的道路像素之间存在较大的空间分离和特征差异,导致在特征提取过程中难以有效建立远距离的语义关联。本研究将多头注意力机制引入网络当中,增强网络在复杂遮挡场景下道路的连续性推理能力。多头注意力机制是 Transformer 架构的核心思想,其核心为多个自注意力机制拼接成一个多头注意力机制。通过该机制动态捕捉影像中道路的跨区域长程拓扑关联,并利用多头子空间的特征解耦特性,在频域和空域上聚焦关键道路响应区域,提高对道路联通性结构的理解和感知能力,从而提高提取路网的连通性。自注意力机制的计算式为:

$$\boldsymbol{F}_{\mathrm{a}}=f_{\mathrm{s}}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_{\mathrm{k}}}}\right)\boldsymbol{V}\ 。\tag{1}$$

式中:$\boldsymbol{F}_{\mathrm{a}}$ 为输出特征矩阵;$\boldsymbol{Q}$ 为查询矩阵;$\boldsymbol{K}$ 为索引矩阵;$\boldsymbol{V}$ 为内容矩阵;d_{k} 为维度大小;f_{s} 为 Softmax 函数。

多头注意力模型结构如图 4 所示。先将输入特征经过三个线性变换,分别得到 $\boldsymbol{Q}$、$\boldsymbol{K}$ 和 $\boldsymbol{V}$;再对 $\boldsymbol{V}$ 和 $\boldsymbol{K}^{\mathrm{T}}$ 矩阵点乘,利用维度为 d_{k} 的 $\boldsymbol{K}$ 矩阵实现加权,生成注意力矩阵;最后,通过 Softmax 函数 f_{s} 对注意力矩阵进行归一化。归一化后的注意力矩阵与 $\boldsymbol{Q}$ 矩阵相乘,得到输出特征矩阵 $\boldsymbol{F}_{\mathrm{a}}$。

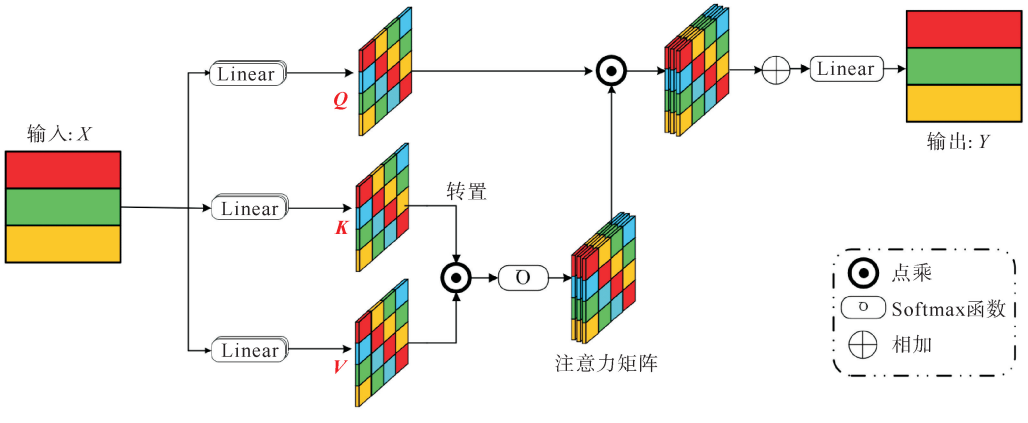


图 4　注意力机制

Fig. 4　Attention mechanism

2.3 自适应损失函数

在高分辨率遥感影像中,道路背景像素所占比例过小,造成前景与背景不平衡问题。为解决该问题,需要对前景数据赋予不同的权重系数,以平衡分割前后背景不同时的场景。故基于二分类的交叉熵损失函数($C_{\text{ross-Entropy}}$)和损失函数($G_{\text{ice\_Loss}}$),提出一种自适应的损失函数。其中,$C_{\text{ross-Entropy}}$ 可以平等计算每个像素点的损失,$G_{\text{ice\_Loss}}$ 可以缓解前景背景不平衡问题,更关注对前景信息的挖掘,即更关注细小道路区域,但会存在损失过饱和问题。计算式为:

$$C_{\text{ross\_Entropy}} = -\frac{1}{N}\sum_{1}^{N}\sum_{j=1}^{C} y_{ij}\ln p_{ij} \ , \tag{2}$$

式中:$C_{\text{ross\_Entropy}}$ 为交叉熵损失函数;N 为样本数量;C 是类别数量;y_{ij} 是样本 i 属于类别 j 的真实标签(one-hot 编码),属于样本 j,则为 1,否则为 0;p_{ij} 是模型预测样本 i 属于类别 j 的概率。

$$G_{\text{ice\_Loss}} = 1 - \frac{2\times|X\cap Y|}{|X|+|Y|+\varepsilon}。 \tag{3}$$

式中:X 是模型预测分割结果;Y 是实际分割结果;$|X\cap Y|$ 是预测结果和实际结果交集的元素个数;$|X|$、$|Y|$ 分别是预测结果和实际结果的元素个数;ε 为一个很小的常数,避免分母为零。本研究统计同一批次中前景与背景真值像素,计算所有前景像素数量与背景值数量的平均值作为阈值,当前景值较小时,赋予 $G_{\text{ice\_Loss}}$ 更大的权重系数。自适应损失函数式为:

$$L = \alpha C_{\text{ross\_Entropy}} + (1-\alpha)\cdot G_{\text{ice\_Loss}} \ , \tag{4}$$

$$\alpha = \begin{cases} \lambda\cdot\dfrac{N_{\text{fg}}}{N_{\text{fg}}+N_{\text{bg}}}, \\ 0.5。 \end{cases} \tag{5}$$

式中:N_{fg} 为同一批次中前景像素总数;N_{bg} 为批次中像素总数;λ 为放大因子,用于增强小目标场景下 $G_{\text{ice\_loss}}$ 的权重。

3 实验与分析

3.1 实验数据集

为了验证实验的准确性,本研究以河北省石家庄市遥感影像为基础,建立一套河北省石家庄市的遥感影像数据集(SJZ Road),图像数据来自国产高分辨率卫星,图像空间分辨率 1 m/像素,数据集经人工精细标注,涵盖城市、农村、郊区多种类别,共 9 190 张,大小为 512×512 像素。

CHN6-CUG 是中国城市道路遥感影像数据集,图像数据来源于北京、上海、武汉、深圳、香港和澳门 6 个城市,图像的空间分辨率为 50 cm/像素。CHN6-CUG 包含 4 511 张大小为 512×512 像素的标记图像,其中 3 608 张用于模型训练,903 张用于测试。

DeepGlobe 道路提取数据集包含 6 226 张 1 024×1 024 像素的卫星遥感图像和标签,每幅图像的空间分辨率为 50 cm/像素,图像包含城市、郊区和乡村道路,来源于泰国、印度和印度尼西亚。将数据集按照 8∶2 进行训练集和验证集划分,其中 2 384 张为训练集,596 张为验证集,3 246 张为测试集。数据集汇总如图 5 所示,其中第 1、3 行为图像,2、4 行为对应标签。

3.2 实验环境

所有实验均基于 Python 3.8、opencv 4.5 以及 windows10 进行构建,内存 32 G、cuda11.3、cudnn8.2.0,采用 24 GB 显存的英伟达 3090 GPU 进行单卡加速训练。实验数据集为 SJZ Road、CHN6-CUG 以及 DeepGlobe,图像入网大小为 512×512 像素。网络超参数,批次为 4、Adam 优化器、初始学习率 0.000 01,采用余弦退火优化策略,每个网络均训练 60 轮。

3.3 实验评价指标

为全面评估算法的准确性,采用 Precision(P)、Recall(R)、F_1 值和 I_{OU}(intersection over union)对模型分割性能进行全面评估,计算式分别为式(5)~(8)。此外,还扩展了评估范围,Param 指神经网络模型中的

参数(权重和偏差),FLOPs 表示模型在推理或训练期间执行的浮点运算的数量,以提供更全面的分析。

$$P=\frac{T_P}{T_P+F_P}, \tag{6}$$

$$R=\frac{T_P}{T_P+F_N}, \tag{7}$$

$$F_1=2\times\frac{P\times R}{P+R}, \tag{8}$$

$$I_{OU}=\frac{T_P}{T_P+F_P+F_N}。 \tag{9}$$

式中:T_P 为被模型预测为正类的正样本(预测正确的正样本),T_N 为被模型预测为负类的负样本(预测正确的负样本),F_P 为被模型预测为正类的负样本(预测错误的正样本),F_N 为被模型预测为负类的正样本(预测错误的负样本)。

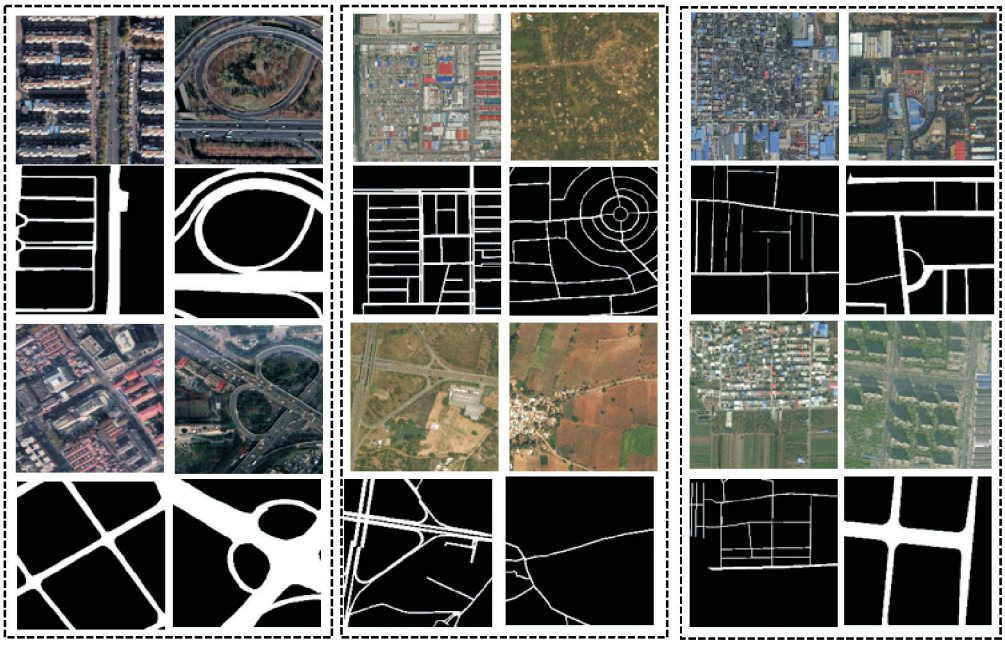


图 5 数据集

Fig. 5 Dataset

3.4 对比实验

为验证本研究方法的有效性,选取当前主流的路网提取网络 SDUNet[12]、DDU-Net[13]、TRNet[14]、SC-RDNet[15]、BDTNet[16]、RCFSNet[17]、DPENet[18]和 RADANet[19]作为对比对象进行实验。通过与上述模型对比表明,本方法以 HRNet-W18 作为主干网络,有效保持了高分辨率特征的表达能力,并通过引入动态损失函数与多头注意力机制实现协同优化,在 DeepGlobe 数据集的性能显著提升。特别是在城市场景中复杂植被干扰场景下,本研究方法的性能明显优于大多数对比网络,其中 I_{OU} 达 0.711,F_1 值达 0.945。

HR-DeepLabV3+与其他主流网络在 DeepGlobe 数据集的精度指标值如表 1 所示,RADANet、DPENet、SC-RDNet 网络的 I_{OU} 精度分别达 0.679、0.659、0.656。本研究网络拥有较低的参数量,得益于轻量化的主干网络 HRNet-W18,其轻量化特性体现在相对较浅的网络层次结构并通过高效的多尺度特征融合,减少了深层网络的重复计算,这与经典的加深网络架构形成了对比,深层网络能够提取更高级别的特征,但带来了更大的计算开销和参数量。SDUNet 和 DDU-Net 均采用基于 U-Net 的结构框架,主要依赖对称的编码-解码路径提取特征,虽然对整体语义信息有较强的感知能力,但在高分辨率细节恢复上存在一定局

限,容易在复杂遮挡下导致道路边界模糊,I_{OU} 分别为 0.668 和 0.665。SC-RDNet 利用多尺度上下文融合策略提升了复杂场景的适应能力,但其融合方式相对静态,难以根据不同遮挡强度自适应调整注意力权重,I_{OU} 为 0.656,F_1 值偏低。BDTNet 通过双分支结构并行建模语义与边缘信息,虽提升了边界精度,但整体特征提取能力偏弱,在深度复杂场景中难以保证连续道路识别的完整性,I_{OU} 仅为 0.649。RCFSNet 尽管引入递归特征融合模块用于增强上下文感知,但其结构计算量大,浮点运算高达 506.760 GFLOPs,未能在复杂环境下体现出对应性能的提升,I_{OU} 仅为 0.583,说明其对高干扰区域适应能力有限。DPENet 和 RADANet 分别通过深度特征增强和多尺度残差机制增强特征表达能力,表现相对较好,但由于缺乏细粒度的空间保持机制,在道路与干扰物边界模糊时仍存在判别不准的问题,I_{OU} 分别为 0.659 和 0.679。HR-DeepLabV3+与其他主流网络在 DeepGlobe 上的识别效果如表 2 所示。其中,TRNet 引入 Transformer 模块以增强全局建模能力,但其下采样路径过深,缺乏高分辨率保留机制,在局部细节处理上效果一般,尤其在细碎道路与植被交叠区域容易出现漏检。

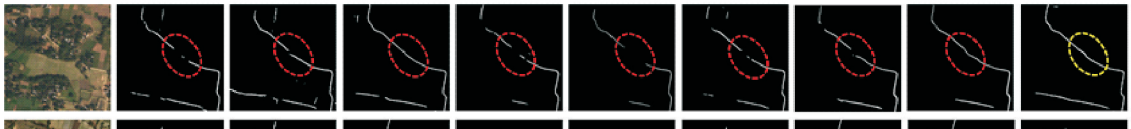
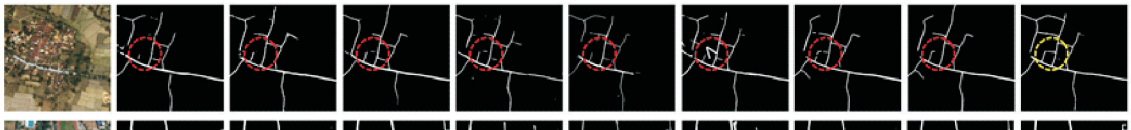
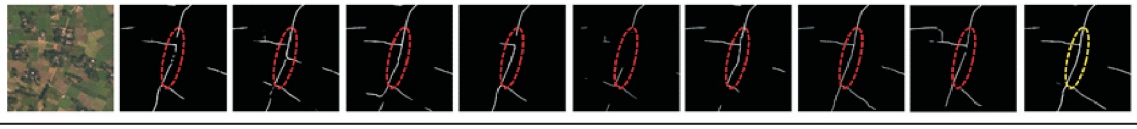
表 1 HR-DeepLabV3+与其他主流网络在 DeepGlobe 测试集的结果对比

Table 1 Comparison of results between improved DeepLabV3+ and other mainstream networks on the DeepGlobe test set

网络结构	F_1	I_{OU}	FLOPs/GFLOPs	Param/MB
SDUNet[12]	0.891	0.668	132.872	67.243
DDU-Net[13]	0.887	0.665	86.541	24.516
TRNet[14]	0.889	0.651	152.845	27.863
SC-RDNet[15]	0.883	0.656	81.184	32.741
BDTNet[16]	0.876	0.649	38.572	57.325
RCFSNet[17]	0.889	0.583	506.760	58.228
DPENet[18]	0.891	0.659	167.516	27.846
RADANet[19]	0.912	0.679	212.452	73.852
本研究方法	0.945	0.711	71.409	9.663

表 2 DeepGlobe 数据集识别对比

Table 2 DeepGlobe dataset recognition comparison

场景	图像	SDUNet[12]	DDU-Net[13]	TRNet[14]	SC-RDNet[15]	BDTNet[16]	RCFSNet[17]	DPENet[18]	RADANet[19]	本研究方法
较多树木遮挡										
少量低矮建筑遮盖										
较多楼宇阴影										
较多低矮建筑遮挡										
较多树木遮挡										

综上所述,其他对比路网提取网络在结构设计上多注重语义增强或上下文建模,但普遍忽视了对高分辨率细节的保持和动态特征的调整能力,导致在高楼阴影、树冠遮挡等复杂场景下容易产生误分或漏分。而本研究方法在结构设计上兼顾高分辨率特征保持与多尺度动态感知,显著提升了复杂遮挡环境下的道路提取精度,验证了其结构优势和场景适应性。

本研究方法在 CHN6-CUG 测试集进行对比实验分析,HR-DeepLabV3+与其他网络在 CHN6-CUG 测试集的精度指标如表 3 所示。该数据集涵盖了树木遮挡、建筑物遮挡及楼宇阴影道路环境,具有较强的代表性。对比结果显示,DPENet 与 RCFSNet 在该数据集的表现相对较弱,其 I_{OU} 分别为 0.643 和 0.659,F_1 分别为 0.856 和 0.861,说明在处理复杂空间结构和遮挡情况下的道路提取能力有限。TRNet、SC-RDNet、BDTNet 等主流模型在该测试集的性能有所提升,其 I_{OU} 分别为 0.652、0.644 和 0.661,F_1 分别为 0.867、0.859 和 0.872,表现出一定的特征表达能力和鲁棒性。

表 3 HR-DeepLabV3+与其他网络在 CHN6-CUG 测试集的结果对比

Table 3 Comparison of results between improved HR-DeepLabV3+and other networks on the CHN6-CUG test set

网络结构	F_1	I_{OU}	FLOPs/GFLOPs	Param/MB
SDUNet[12]	0.865	0.648	132.872	67.243
DDU-Net[13]	0.869	0.657	86.541	24.516
TRNet[14]	0.867	0.652	152.845	27.863
SC-RDNet[15]	0.859	0.644	81.184	32.741
BDTNet[16]	0.872	0.661	38.572	57.325
RCFSNet[17]	0.861	0.659	506.760	58.228
DPENet[18]	0.856	0.643	167.516	27.846
RADANet[19]	0.875	0.662	212.452	73.852
本研究方法	0.897	0.685	71.409	9.663

本研究方法在融合多头注意力机制后,提升了对复杂场景中道路目标的建模能力,有效强化了特征表达的精度与上下文感知能力。在遮挡、阴影及道路形态复杂条件下,仍能保持较强的连贯性和完整性,实现了更加精确的道路提取。实验结果表明,本研究方法在多个主流模型中达到领先水平,其 I_{OU} 提升至 0.685,F_1 达 0.897,充分验证了本研究方法的有效性与优越性。

HR-DeepLabV3+与其他网络在 CHN6-CUG 测试集的识别效果如表 4 所示,本研究进一步可视化对比了本研究方法与其他网络在 CHN6-CUG 测试集的结果。CHN6-CUG 数据集中道路环境更加复杂,空间结构立体性更强,道路形态多变,颜色与纹理差异较小,且普遍存在高大建筑密集、遮挡严重、光照变化明显等挑战性因素。为全面评估该方法的适应性,本研究选取数据集中典型的较多树木遮挡地区,并与纹理差异较小的建筑混淆区域进行实验验证。

由表 4 可以看出,在楼房遮挡区域,由于楼体具备显著的光谱纹理特征,而路面信息往往因遮挡而丢失,导致传统方法提取困难。尽管 SDUNet、BDTNet 等模型在处理单一、无遮挡的道路提取任务中表现良好,但在复杂环境中的表现仍不理想,尤其是在路面纹理与周围建筑相似的场景中,容易出现误识或漏检。

相比之下,本研究方法依靠多头注意力机制能够有效识别道路的空间分布、纹理差异以及道路线型结构,特别是在高层建筑遮挡区域,在全局信息建模与特征聚合方面具有明显优势,显著提升了识别准确性。在树木遮挡区域,本研究方法不仅能准确识别连续的道路,还能精确分割城市道路中的复杂结构如环岛,说明 HRNet 采用的多分支并行卷积结构在保持高分辨率信息的同时,具备优秀的细节保持和纹理判别能力。

为验证网络的先进性,本研究进一步在 SJZ Road 测试集进行对比分析,该测试集涵盖了 CHN6-CUG、DeepGlobe 数据集的特征,既包括高楼阴影也存在大量植被遮挡,能够全面评估网络在不同环境下的表现。HR-DeepLabV3+与其他主流网络在 SJZ Road 测试集的精度指标如表 5 所示,在该数据集中,本研究方法

精度最高，I_{OU} 达 0.751，比其他网络提高了 3.5%～8.7%；F_1 达 0.937，相比其他网络精度提高了 2.5%～5.1%。然而，TRNet 和 RCFSNet 网络在 SJZ Road 数据集的精度较低，I_{OU} 分别为 0.677 和 0.664，F_1 分别为 0.887 和 0.886。

表 4 与 CHN6-CUG 数据集的各网络对比

Table 4 Comparison of various networks on the CHN6 CUG dataset

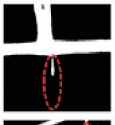
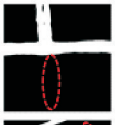
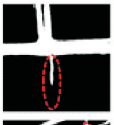
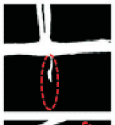
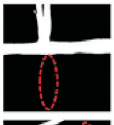
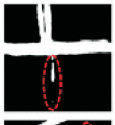
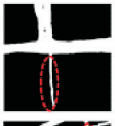
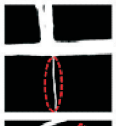
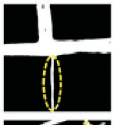
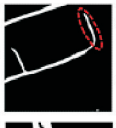
场景	图像	SDUNet[12]	DDU-Net[13]	TRNet[14]	SC-RDNet[15]	BDTNet[16]	RCFSNet[17]	DPENet[18]	RADANet[19]	本研究方法
低矮灌木遮挡										
较多树木遮挡										
较多楼宇阴影										
较多树木遮挡										
纹理差异较小										

表 5 HR-DeepLabV3＋与其他主流网络在 SJZ Road 测试集的对比结果

Table 5 Comparison of results between improved DeepLabV3＋and other mainstream networks on the SJZ Road test set

网络结构	F_1	I_{OU}	FLOPs/GFLOPs	Param/MB
SDUNet[12]	0.892	0.687	132.872	67.243
DDU-Net[13]	0.895	0.691	86.541	24.516
TRNet[14]	0.887	0.677	152.845	27.863
SC-RDNet[15]	0.897	0.712	81.184	32.741
BDTNet[16]	0.893	0.701	38.572	57.325
RCFSNet[17]	0.886	0.664	506.760	58.228
DPENet[18]	0.905	0.709	167.516	27.846
RADANet[19]	0.912	0.716	212.452	73.852
本研究方法	0.937	0.751	71.409	9.663

HR-DeepLabV3＋与其他主流网络在 SJZ Road 测试集的识别效果如表 6 所示，本研究对 SJZ Road 数据集的路网提取结果进行可视化分析。在道路与周围建筑纹理差异小的区域和树木遮挡地区，尤其是城市厂区与居民小区等场景中，道路多为硬化水泥铺设，边界不明显，给道路提取带来较大困难。在此类区域中，RADANet 及 SC-RDNet 均存在不同程度的漏检现象，未准确捕捉完整的道路结构。

相比之下，本研究方法在上述复杂场景的表现更加稳定，不仅能够准确提取边界模糊的水泥道路，而且在高层建筑阴影区域中，依然能够有效识别被部分遮挡的道路信息。这得益于所采用的 HRNet-W18 主干网络与金字塔池化结构的协同作用。HRNet-W18 能够持续保持高分辨率特征表达，而金字塔池化模块则

增强了上下文信息的融合能力,两者结合有效提升了模型对纹理模糊和遮挡严重区域的判别能力。此外,由表 6 对比可知,尽管 RCFSNet 在阴影遮挡区域的道路连通性方面表现尚可,但其对细粒度道路的提取仍存在明显不足,导致整体 I_{OU} 精度偏低。这表明,其对局部细节信息的建模能力有限,难以适应复杂多变的实际道路场景。

表 6 SJZ Road 数据集的网络对比

Table 6 Network comparison of SJZ Road dataset

场景	图像	SDUNet[12]	DDU-Net[13]	TRNet[14]	SC-RDNet[15]	BDTNet[16]	RCFSNet[17]	DPENet[18]	RADANet[19]	本研究方法
纹理差异小										
纹理差异小										
较多树木遮挡										
纹理差异小										
较多楼宇遮挡										

综上所述,本研究提出的网络结构在 SJZ Road 数据集中对楼宇阴影遮挡、树木遮挡等复杂环境时具有较好的性能。尤其在阴影覆盖和植被干扰显著的区域,其道路识别能力明显优于现有主流方法,适应性与精度均较好。

3.5 消融实验

为验证本研究各模块在应对复杂道路提取场景中的有效性,在引入动态损失函数的基础上,设计了一系列消融实验,依次评估多头注意力机制(MHA)、空洞空间金字塔池化(ASPP)模块及 HRNetV2-W18 主干网络对整体性能的贡献。

实验结果如表 7 所示,基线模型在 DeepGlobe、CHN6-CUG 及 SJZ Road 三个数据集的 I_{OU} 分别为 0.618、0.638、0.715,F_1 分别为 0.871、0.871 和 0.906。基线网络结构相对简单,缺乏对空间上下文与高分辨率细节的有效建模能力,导致在高楼阴影、植被遮挡等复杂环境下,道路边界模糊、连接性不足,提取效果不稳定。当在基线模型中加入多头注意力模块后,DeepGlobe 数据集的 I_{OU} 和 F_1 分别为 0.639 和 0.882,提升了 2.1%和 1.1%。该模块可增强网络对显著区域的关注力,有效抑制背景干扰,对易被遮挡或低对比度道路区域的识别能力有所提升。尽管在其他两个数据集的提升幅度相对较小,但仍体现出一定的稳健性。进一步,在模型中加入 ASPP 模块后,DeepGlobe 数据集 I_{OU} 提升了 1.7%、F_1 提升了 1.2%。ASPP 能够通过多尺度空洞卷积扩大感受野,且增强了对不同尺度与形态道路的建模能力,尤其对郊区及非规则道路结构具备更强的表达效果,有助于缓解遮挡或断裂造成的识别难题。

进一步,将原有主干替换为 HRNetV2-W18 后,DeepGlobe 数据集的 I_{OU} 提升了 3.4%、F_1 提升了 1.4%,提升幅度最大。HRNet 通过多分支并行架构保持特征图的分辨率,具备良好的边缘保持能力,特别适用于城市密集区因高楼阴影和道路变窄而引发的细节丢失问题,有效提升了模型的结构识别与边界定位

能力。多模块组合 MHA+HRNet 的 I_{OU} 提升至 0.666，F_1 提升至 0.893，验证了在高分辨率基础上加入注意力机制可进一步增强目标区域的响应。多模块组合 ASPP+HRNet 的 I_{OU} 进一步提升至 0.699，F_1 提升至 0.896，说明在结构表达能力提升的同时，结合多尺度感受野对复杂道路结构更具适应性。多模块组合 MHA+ASPP+HRNet(本研究方法)的 I_{OU} 提升至 0.711，F_1 达 0.945，较基线分别提升了 9.3%与 7.4%。CHN6-CUG 与 SJZ Road 数据集也取得了稳定提升，验证了三模块协同作用的有效性。

表 7 消融实验

Table 7 Ablation experiment

研究方法	I_{OU}			F_1		
	DeepGlobe	CHN6-CUG	SJZ Road	DeepGlobe	CHN6-CUG	SJZ Road
Basebline	0.618	0.638	0.715	0.871	0.871	0.906
Basebline+MHA	0.639	0.642	0.721	0.882	0.873	0.909
Basebline+ASPP	0.635	0.643	0.723	0.883	0.875	0.910
Basebline+HRNet	0.652	0.649	0.726	0.885	0.877	0.913
Basebline+MHA+HRNet	0.666	0.658	0.734	0.893	0.885	0.918
Basebline+ASPP+HRNet	0.699	0.661	0.736	0.896	0.886	0.920
本研究方法	**0.711**	**0.685**	**0.751**	**0.945**	**0.897**	**0.937**

消融实验结果表明，HRNet 主干网络在保持高分辨率特征方面表现优异，是提升道路提取精度的关键因素；ASPP 模块增强了模型的上下文建模与尺度适应能力，MHA 机制则提升了对关键目标区域的响应能力。三者联合使用，有效缓解了原始方法在高楼阴影、植被遮挡等复杂环境下道路识别效果不佳的问题。

4 结论

针对遥感影像城市路网自动提取的问题，本研究提出一种基于 HR-DeepLabV3+路网的提取方法，分别采用 DeepGlobe、CHN6-CUG 以及 SJZ Road 数据集与多种主流算法进行对比实验，I_{OU} 精度分别达 0.711、0.685、0.751，在 DeepGlobe 数据集上，改进网络与基线网络 DeeplabV3+网络相比，I_{OU} 最大提升了 9.3%，与最新的 DPENet 网络相比精度提升了 5.2%。本研究方法在三个数据集中均有效解决了高楼阴影与植被遮挡问题，证明了本研究方法具有良好的识别稳健性。

下一步，针对现有模型在高密度建筑区、多尺度道路交叉口中的细粒度分割能力不足的问题，拟引入 SAR 影像多源数据融合特征，进一步优化 HR-DeepLabV3+的上下文建模能力，构建融合多光谱的跨模态训练框架，通过特征互补增强阴影与遮挡区域的识别精度。

参考文献：

[1] DAI J G，ZHU T T，ZHANG Y L，et al. Lane-level road extraction from high-resolution optical satellite images[J/OL]. Remote Sensing，2019，11(22). DOI：10.3390/rs11222672.

[2] CHEN J Q，LU J C，ZHU X T，et al. Generative semantic segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver：IEEE，2023：7111-7120.

[3] DIAKOGIANNIS F I，WALDNER F，CACCETTA P，et al. ResUNet-a：A deep learning framework for semantic segmentation of remotely sensed data[J]. ISPRS Journal of Photogrammetry and Remote Sensing，2020，162：94-114.

[4] HETANG C R，XUE H R，LE C，et al. Segment anything model for road network graph extraction[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle：IEEE，2024：2556-2566.

[5] XU Z H，LIU Y X，SUN Y X，et al. RNGDet++：Road network graph detection by transformer with instance segmentation and multi-scale features enhancement[J]. IEEE Robotics and Automation Letters，2023，8(5)：2991-2998.

[6] QIU L Y，YU D Y，ZHANG C X，et al. A semantics-geometry framework for road extraction from remote sensing images

[J/OL]. IEEE Geoscience and Remote Sensing Letters,2023,20. DOI:10.1109/LGRS.2023.3268647.

[7] 王谦,何朗,王展青,等.基于改进 DeepLabv3+的遥感影像道路提取算法[J].计算机科学,2024,51(8):168-175.
WANG Qian,HE Lang,WANG Zhanqing,et al. Road extraction algorithm for remote sensing images based on improved DeepLabv3+[J]. Computer Science,2024,51(8):168-175.

[8] CHEN J D,WEN Y X,NANEHKARAN Y A,et al. Multiscale attention networks for pavement defect detection[J/OL]. IEEE Transactions on Instrumentation and Measurement,2023,72. DOI:10.1109/TIM.2023.3298391.

[9] 赫晓慧,陈明扬,李盼乐,等.结合 DCNN 与短距条件随机场的遥感影像道路提取[J].武汉大学学报(信息科学版),2024,49(3):333-342.
HE Xiaohui,CHEN Mingyang,LI Panle,et al. Road extraction from remote sensing images by integrating DCNN with short range conditional random fields[J]. Geomatics and Information Science of Wuhan University,2024,49(3):333-342.

[10] DAI L,ZHANG G Y,ZHANG R T. RADANet:Road augmented deformable attention network for road extraction from complex high-resolution remote-sensing images[J/OL]. IEEE Transactions on Geoscience and Remote Sensing,2023,61. DOI:10.1109/TGRS.2023.3237561.

[11] CHEN R N,HU Y,WU T,et al. Spatial attention network for road extraction[C]// IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium. Athens:IEEE GRSS,2020,24(1):1841-1844.

[12] YANG M X,YUAN Y,LIU G C,et al. SDUNet:Road extraction via spatial enhanced and densely connected UNet[J/OL]. Pattern Recognition,2022,126. DOI:10.1016/j.patcog.2022.108549.

[13] CHENG J L,TIAN S W,YU L,et al. DDU-Net:A dual dense U-structure network for medical image segmentation[J/OL]. Applied Soft Computing,2022,126. DOI:10.1016/j.asoc.2022.109297.

[14] YANG Z G,ZHOU D X,YANG Y,et al. TransRoadNet:A novel road extraction method for remote sensing images via combining high-level semantic feature and context[J]. IEEE Geoscience and Remote Sensing Letters,2022,19:1-8.

[15] ABDOLLAHI A,PRADHAN B,ALAMRI A. SC-RoadDeepNet:A new shape and connectivity-preserving road extraction deep learning-based network from remote sensing data[J]. IEEE Transactions on Geoscience and Remote Sensing,2022,60:13-19.

[16] LUO L,WANG J X,CHEN S B,et al. BDTNet:Road extraction by bi-direction transformer from remote sensing images [J]. IEEE Geoscience and Remote Sensing Letters,2022,19:1-5

[17] YANG Z G,ZHOU D X,YANG Y,et al. Road extraction from satellite imagery by road context and full-stage feature[J/OL]. IEEE Geoscience and Remote Sensing Letters,2022,20. DOI:10.1109/LGRS.2022.3228967.

[18] CHEN Z Y,LUO Y H,WANG J,et al. DPENet:Dual-path extraction network based on CNN and transformer for accurate building and road extraction[J/OL]. International Journal of Applied Earth Observation and Geoinformation 2023,124. DOI:10.1016/j.jag.2023.103510.

[19] DAI L,ZHANG G Y,ZHANG R T. RADANet:Road augmented deformable attention network for road extraction from complex high-resolution remote-sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing,2023,61:1-13.

(责任编辑:高丽华)