

文章编号 1004-924X(2026)02-0267-12

跨层级交互与方位感知的航拍图像语义分割

刘 杰¹, 吴紫钰¹, 田 明², 韩 轲^{3*}

(1. 哈尔滨理工大学 模式识别与信息感知黑龙江省重点实验室, 黑龙江 哈尔滨 150080;

2. 中国电信股份有限公司黑龙江分公司, 黑龙江 哈尔滨 150040;

3. 哈尔滨商业大学 计算机与信息工程学院, 黑龙江 哈尔滨 150028)

摘要:针对航拍图像语义分割中提取特征尺度单一、细节丢失、边界模糊的问题,本文提出一种跨层级交互与方位感知的航拍图像语义分割网络。通过方向解耦注意力策略构建方位感知模块,增强模型对空间方向信息的处理能力;设计跨层级交互模块进行跨通道特征交互融合,提升空间感知能力,同时利用通道-空间注意力机制增强特征提取能力,在复杂场景中缓解细节模糊问题;最后对分割头进行轻量化设计,去除冗余操作,在降低计算量的同时保证分割性能。实验结果表明,本文网络在 UAVid 数据集与 Aeroscapes 数据集的平均交并比相较于基线模型 SegFormer 分别提升了 1.7% 和 1.3%,证明了该网络在航拍图像等复杂情况下语义分割的有效性。Human 类别相较于基线模型分割精度提升了 1.8%,证明了本文网络在小目标分割方面表现出色。与多个主流网络相比,本文方法在两个数据集上均实现了最高的分割精度,表现出更优的泛化能力。

关键词:跨层级交互;语义分割;航拍图像;轻量化;方向解耦

中图分类号:TP394.1;TH691.9 **文献标识码:**A

doi:10.37188/OPE.20263402.0267

CSTR:32169.14.OPE.20263402.0267

Semantic segmentation of aerial images based on multi-scale feature interaction and fusion

LIU Jie¹, WU Ziyu¹, TIAN Ming², HAN Ke^{3*}

(1. Harbin University of Science and Technology, Key Laboratory of Pattern Recognition and Information Perception in Heilongjiang Province, Harbin 150080, China;

2. Heilongjiang Branch of China Telecom Corporation Limited, Harbin 150040, China;

3. Harbin University of Commerce, School of Computer and Information Engineering, Harbin 150080, China)

* Corresponding author, E-mail: hanke@hrbcu.edu.cn

Abstract: To address the issues of single-scale feature extraction, detail loss, and blurred boundaries in aerial image semantic segmentation, this paper proposed an aerial image semantic segmentation network with cross-level interaction and orientation awareness. A position awareness module was constructed through a direction-decoupled attention strategy to enhance the model's ability to process spatial directional information; a cross-level interaction module was designed for inter-channel feature interaction and fusion

收稿日期:2025-08-20;修订日期:2025-09-17.

基金项目:黑龙江省自然科学基金(No. LH2023E086);黑龙江省交通运输厅科技项目(No. HJK2024B002)

to improve spatial perception, while a channel-spatial attention mechanism was used to enhance feature extraction capabilities and alleviate detail blurring issues in complex scenes; finally, a lightweight design was implemented for the segmentation head, removing redundant operations to reduce computational load while ensuring segmentation performance. Experimental results indicate that the proposed network achieves a 1.7% and 1.3% improvement in mean intersection over union on the UAVid and Aerialscapes datasets, respectively, compared to the baseline model SegFormer, demonstrating the network's effectiveness in semantic segmentation under complex conditions such as aerial images. The segmentation accuracy of the Human category improved by 1.8% compared to the baseline model, demonstrating that the network proposed in this paper performs excellently in small object segmentation. Compared with several mainstream networks, the method proposed in this paper achieves the highest segmentation accuracy on both datasets, showing superior generalization capability.

Key words: inter-level interaction; semantic segmentation; aerial imagery; lightweighting; direction decoupling

1 引言

语义分割技术是一种图像处理技术,其目标是将图像中的每个像素分配到预定义的类别中,实现像素级的分类^[1]。语义分割技术已经在多个领域得到广泛应用,包括航空航天领域的遥感影像地物分类和目标跟踪^[2],医学影像分析中的红细胞计数和白细胞核分割等精细任务^[3],以及智能驾驶中的道路、车辆和行人实时语义理解^[4]。特别是在无人机技术的推动下,语义分割技术的应用范围进一步扩大,为城市规划、土地资源管理和城市环境监测等提供了数据支持^[5]。

近年来,计算机视觉与深度学习不断发展进步。计算机视觉,经历了从传统几何与信号处理方法向深度学习模型的深刻变革,主要分为二维信息获取与三维信息获取两大方向。在二维视觉方面,传统方法如尺度不变特征变换(Scale-Invariant Feature Transform, SIFT)^[6]、方向梯度直方图(Histogram of Oriented Gradients, HOG)^[7]被广泛用于图像识别、图像分割与目标检测等任务。在三维视觉方面,研究者发展了多种高精度三维信息获取技术。傅里叶变换轮廓术(Fourier Transform Profilometry, FTP)^[8]、相移测量轮廓术(Phase Shift Profilometry, PSP)^[9]主要用于非接触式高精度三维形貌重建。调制度测量轮廓术(Modulation Measurement Profilometry, MMP)^[10]利用条纹调制度实

现高度映射,适用于高速三维形貌测量。基于计算莫尔轮廓术(Computational Moiré Profilometry, CGMP)^[11]的改进方法不断涌现,在实时三维测量领域展现出广阔的应用前景。激光雷达测距(LiDAR)^[12]通过主动测距,广泛应用于自动驾驶与机器人环境感知。本文主要聚焦于二维视觉中的语义分割任务。深度学习算法的出现,为计算机视觉带来了新的突破。其中卷积神经网络(Convolutional Neural Network, CNN)的广泛应用,显著推动了图像语义分割任务的发展。以 U-Net^[13]和 DeepLab^[14]为代表的经典架构,通过编码器-解码器结构与空洞卷积等技术,在遥感影像分析、医学图像分割等领域实现了高性能表现。然而,此类方法普遍存在对长距离空间依赖建模能力不足、计算复杂度较高等问题,具有一定的局限性。SETR^[15]将图像分割视为序列到序列的任务,采用多层 Transformer,但计算开销大且小目标分割性能受限;Swin Transformer^[16]通过移位窗口机制提升上下文融合能力,但计算复杂度高、训练数据需求大。MCSwinUnet^[17]以 SwinUnet 为基础,提出了一种多尺度上下文感知网络,系统提升了复杂场景分割精度与鲁棒性,但在边缘分割方面有待提升。闫河^[18]等人提出融合 SimAM 与 EA-Net 注意力机制,并以 Haar 小波下采样替代全局池化,提升语义理解能力,但其网络复杂度与计算成本较高。贾轩^[19]等人提出一种融合深度度量学习的语义分割框架,通过关注像素语义关系

提升精度,但模型的泛化能力具有一定局限性。MDFNet^[20]通过融合目标区域增强、双重注意力与特征金字塔网格,协同强化多维特征提取与多尺度融合,但其实时性受到限制。Xie提出的SegFormer^[21]将层次化Transformer编码器与轻量级全多层感知机(Multilayer Perceptron, MLP)解码器结合,在保持高精度的同时显著降低计算成本,但其对局部细节建模较弱,容易出现漏检或边界模糊的问题。

注意力机制源于自然语言领域,自Bahdanau等人首次提出并显著提升模型性能后,已广泛应用于事件抽取、3D测量、多模态情感分析、医学图像分析等领域。动态注意力匹配和图形注意力网络(Dynamic Attention Matching and Graph Attention Networks)^[22]有效识别模糊触发词并构建事件子图,在噪声环境下提升事件抽取鲁棒性;多注意力边缘恢复网络的条纹投影轮廓术(Ring Projection Profilometry by Multi-Attention Fringe Restoration Network)^[23]则通过重分配条纹特征提高3D测量精度;混合最优多尺度残差注意力网络(Hybrid Optimal Multi-Scale Residual Attention Network)^[24]有效捕捉并融合多模态信息中的关键情感线索,实现了对多模态情感分析性能的显著提升;多尺度交叉轴注意力(Multi-scale Cross-axis Attention, MCA)^[25]通过条带卷积捕捉医学图像中的微小结构。尽管应用场景不同,但核心思想一致,通过上下文自适应分配权重,突出关键特征、抑制噪声,从而在复杂、模糊或小目标场景中提升模型性能。然而,注意力机制仍面临计算开销大、对数据质量敏感及通用性有限等挑战。因此,设计高效注意力机制对航拍图像的语义分割具有重要意义。

本文以SegFormer作为基线模型,提出一种跨层级交互与方位感知的航拍图像语义分割方法。通过方位感知模块(Position Awareness Module, PAM)增强物体位置与方向特征,提升分割效果。接着设计跨层级交互模块(Cross-level Interaction Module, CIB)对不同层级特征进行交互融合,实现信息的互补。同时提出通道-空间注意力机制(Channel-Spatial Attention,

CSA),强调目标区域的同时有效抑制干扰。最后,提出轻量化分割头(Lightweight Splitting Head, LSH)优化网络结构,在保证分割效率的情况下简化网络。最后在UAVID与Aeroscapes数据集上进行实验,在参数量和复杂度小幅增长的情况下,网络的分割精度取得较好效果。

2 本文算法介绍

2.1 SegFormer算法介绍

SegFormer是一种用于语义分割任务的高效Transformer架构,整体的网络架构如图1(a)所示。 H, W, C 分别代表特征图的高度、宽度以及通道数。SegFormer编码器采用层次化设计,包含4个阶段。在每个阶段,使用一个独立的Transformer块处理输入特征。编码器输出4个不同尺度的特征图,分别为原图像分辨率的1/4, 1/8, 1/16和1/32。轻量级多层感知机(Multi-Layer Perceptron, MLP)解码器接收编码器输出的4个不同尺度的特征图,进行特征融合并生成最终的分割预测,结合损失计算和优化过程,实现高效准确的语义分割。

2.2 本文网络总体设计

虽然SegFormer具备多尺度特征提取能力,但其跨层级交互机制薄弱,容易导致小目标漏检和边界模糊等问题。因此,本文提出改进网络,结构如图1(b)所示。编码器四个阶段输出不同尺度特征图,在阶段4引入方位感知模块PAM,通过方向解耦注意力增强空间位置信息感知能力。随后,在四个阶段生成的不同尺度的特征图后加入跨层级交互模块CIB,构建特征金字塔双向信息流,融合各层级特征信息。模块内部由四个上下文增强模块(Context Enhancement Module, CEM)组成,分别作用于不同层级的特征以及经过方位感知模块PAM处理后的深层特征,对特征进行针对性的增强操作。接着,多层特征通过MLP层统一通道维度。解码器部分提出轻量化分割头(Lightweight Splitting Head, LSH),在原分割头基础上进行优化,有效降低了计算复杂度,实现了对图像的分割。

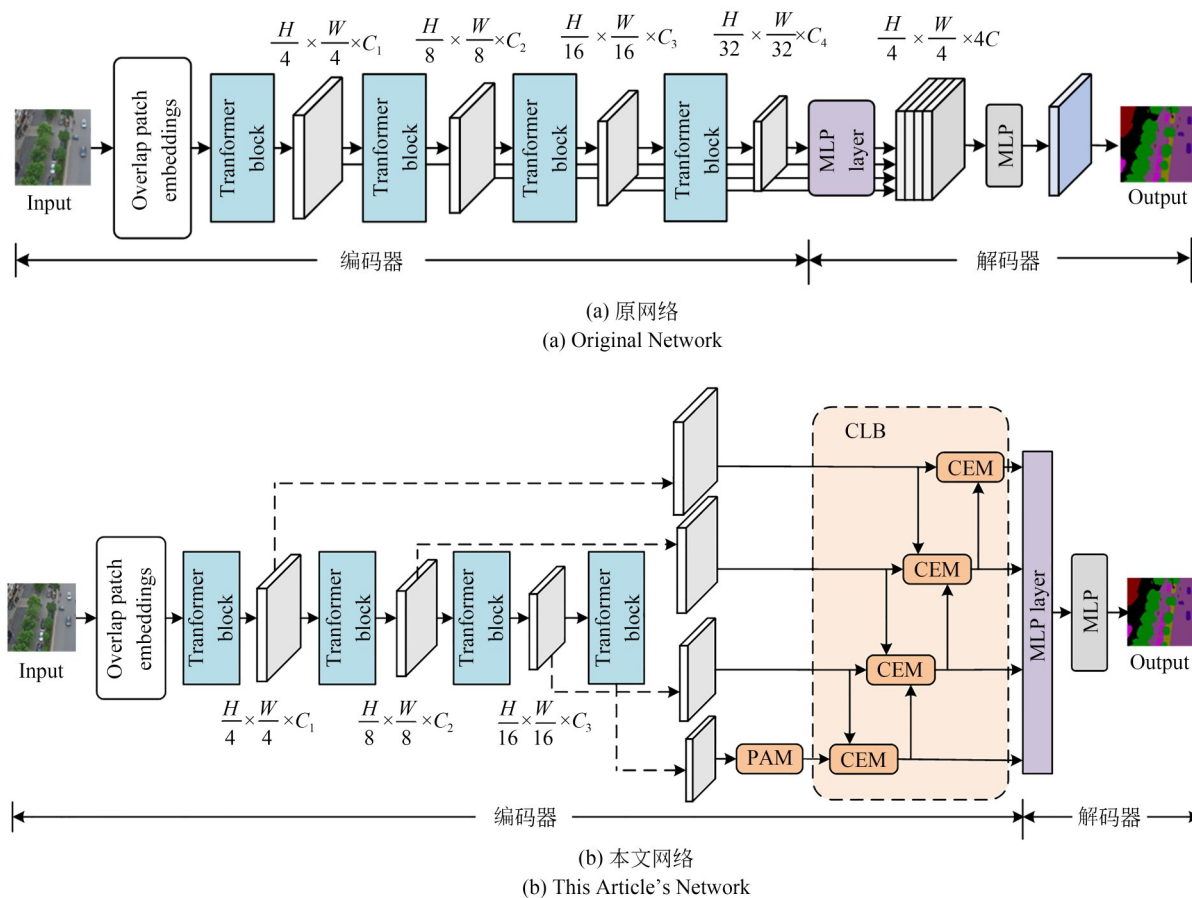


图 1 网络结构改进对比

Fig. 1 Structure of the SegFormer algorithm

3 算法改进设计

3.1 方位感知模块设计

传统特征融合方法常忽略目标在不同空间方向上的语义与几何差异,导致对方向敏感场景

的分割性能受限。因此,本文提出方位感知模块 PAM,增强模型对结构化方向信息的建模能力。模块结构如图 2 所示。

如图 2 所示,首先对输入张量沿高度与宽度方向分别进行全局平均池化,生成方向特征向

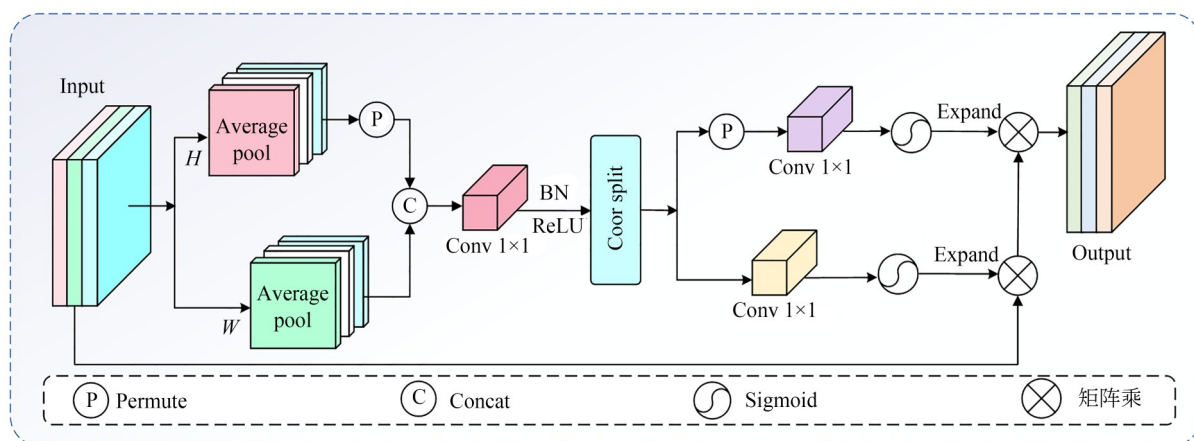


图 2 方位感知模块 PAM

Fig. 2 Position awareness module PAM

量。随后,将两者沿空间维度拼接,显式建模高度与宽度之间的空间依赖关系。接着,通过 1×1 卷积降维减少通道数,提取关键特征,并进行批归一化与激活。为了增强模型对特定方向特征的捕捉能力,避免不同维度的干扰,将得到的联合特征拆分为高度、宽度双分支,各自生成方向注意力权重。最终,将权重扩展至原始空间尺寸,并与输入特征逐元素相乘,实现了对特征图的加权,突出了重要区域的特征信息。

3.2 跨层级交互模块设计

尽管方位感知模块增强了对方向信息的感知能力,但面对航拍图像目标尺度差异大、空间布局复杂的问题,单一尺度的方向建模难以满足精度需求。因此,本文进一步引入跨层级交互模块CIB。通过双向融合机制强化多尺度上下文感知,有效缓解小目标漏检与边界模糊

问题。

设输入特征层级为 $\{F_1, F_2, F_3, F_4\}$ (分辨率从高到低),如式(1)所示,模块通过上、下采样与拼接实现跨层交互:

$$F^{(i)} = \begin{cases} \text{CEM}_1(\uparrow(F_2) \oplus F_1) & i=1 \\ \text{CEM}_i(\downarrow(F_{(i-1)}) \oplus F_{(i)} \oplus \uparrow(F_{(i+1)})) & 1 < i < 4, \\ \text{CEM}_4(\downarrow(F_3) \oplus F_4) & i=4 \end{cases} \quad (1)$$

其中: $\oplus$ 表示通道维度拼接。 $F^{(i)}$ 表示不同特征层级经过CEM模块拼接后的特征。 $i=1$,表示低层级融合, $1 < i < 4$,表示中间层级融合, $i=4$,表示高层级融合。

跨层级交互模块的核心为上下文增强模块CEM,其结构图如图3所示。

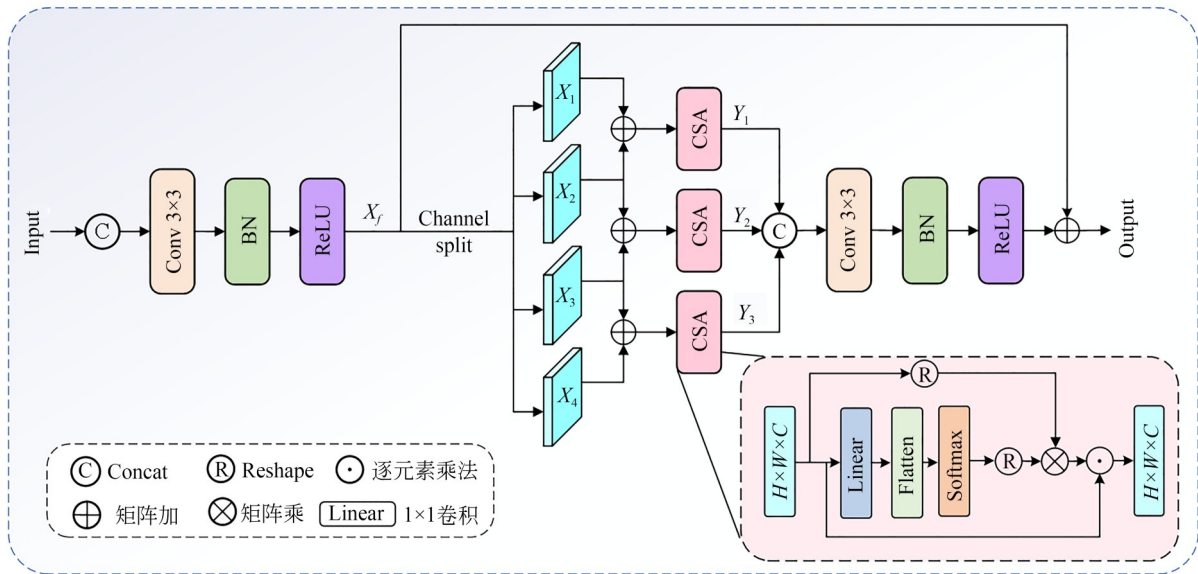


图3 上下文增强模块(CEM)

Fig. 3 Context Enhancement Module (CEM)

将Transformer块输出的相邻特征图通过 3×3 卷积进行特征融合生成 X_i ,整合多层级上下文信息。随后,对输入特征进行通道分解为4组,执行相邻相加促进跨组交互,并输入至通道-空间注意力(Channel-spatial Attention, CSA)模块。该模块首先通过 1×1 卷积压缩通道维度,生成空间注意力权重。权重展平后沿空间维度执行Softmax归一化。然后,分别将输入特征与注意力权重进行重塑,完成矩阵乘法的操

作,以得到通道调制系数。最后,将调制系数与原始输入特征逐通道相乘,实现通道维度上的注意力调制。该过程输出三组交互特征 Y_1, Y_2, Y_3 ,分别对应相邻两个分组经过交互和注意力处理后得到的结果。最终,将三组特征 Y_1, Y_2, Y_3 进行融合,并采用残差连接,以缓解梯度消失,确保模型的稳定性。具体计算过程如式(2)所示:

$$Y = \text{Conv}_{3 \times 3}(\text{Concat}[Y_1, Y_2, Y_3]) + X_f. \quad (2)$$

3.3 轻量化分割头设计

方位感知模块与跨层级交互模块的协同作用提升了模型对航拍场景中复杂目标的空间感知能力,但引入的多层级特征交互机制增加了计算负担。因此,本文通过对 SegFormer 原分割头的特征融合方式进行改进,提出了一种轻量化分割头 LSH,结构如图 4 所示,在降低计算成本的同时保持了分割精度。

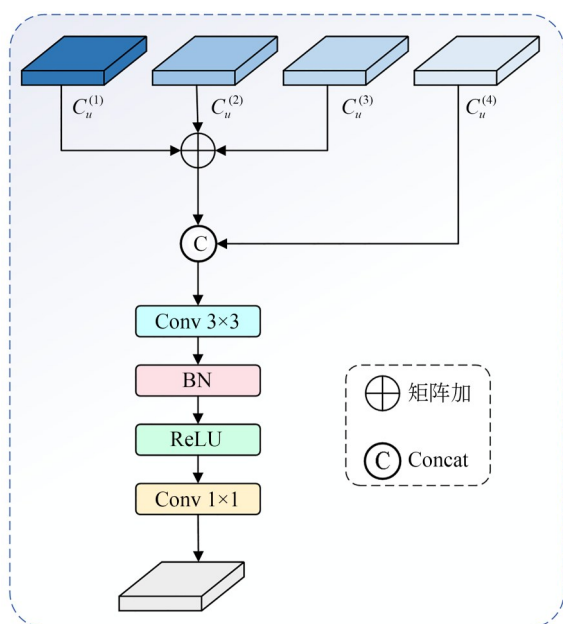


图 4 轻量化分割头 LSH

Fig. 4 Lightweight splitting head LSH

在 SegFormer 中,编码器输出四个不同尺度的特征图。这些特征图经双线性插值上采样至统一分辨率后,分别通过 MLP 进行非线性映射,得到变换后的特征图 $C_u^{(1)}$ 、 $C_u^{(2)}$ 、 $C_u^{(3)}$ 、 $C_u^{(4)}$ 。接着将其沿通道维度拼接为高维张量。最终经 1×1 卷积与 CBR($\text{Conv}3 \times 3 + \text{BN} + \text{ReLU}$) 模块输出最终分割结果 Y , 计算过程如式(3)所示:

$$Y = \text{Conv}_{1 \times 1} \left(\text{CBR} \left(\text{Concat} \left(C_u^{(1)}, C_u^{(2)}, C_u^{(3)}, C_u^{(4)} \right) \right) \right). \quad (3)$$

尽管 SegFormer 已经非常高效,但特征融合仍存在冗余。因此,本文提出了轻量化分割头 LSH。如图 4 所示,在所有特征通过上采样达到统一尺度后,浅层特征被用作信息整合特征进行处理。与直接拼接特征数量(通道)不同,轻量化分割头采用元素相加操作来减少计算量。随后,

将经过元素相加操作后的浅层特征与最后一层级特征进行拼接,进而完成预测。计算过程如式(4),式(5)所示:

$$C_{\text{agg}} = C_u^{(3)} + C_u^{(2)} + C_u^{(1)}, \quad (4)$$

$$Y = \text{Conv}_{1 \times 1} \left(\text{CBR} \left(\text{Concat} \left(C_{\text{agg}}, C_u^{(4)} \right) \right) \right), \quad (5)$$

其中: CBR 表示 $\text{Conv}3 \times 3 + \text{BN} + \text{ReLU}$, C_{agg} 表示浅层特征 $C_u^{(1)}$ 、 $C_u^{(2)}$ 、 $C_u^{(3)}$ 的元素相加融合, Concat 表示将融合后的浅层特征与最高层特征拼接。

4 实验结果与分析

4.1 实验数据集

本文实验采用 UAVid 与 AerialScapes 两个无人机语义分割数据集进行实验验证。其中, UAVid 数据集聚焦城市街景场景, 包含 3840×2160 和 4096×2160 两种分辨率的 420 张图像。所有图片以长度中心为基准, 左右对称填充, 以宽度中心为基准, 上下对称裁剪, 使图片尺寸变为 4096×2048 , 将每张图片平均分割成八张 1024×1024 的子图进行实验。数据集涵盖 8 个语义类别, 划分为 1600 张训练集、560 张验证集和 1200 张测试集, 并在训练过程中采用随机垂直、水平翻转, 翻转比率为 0.5, 采用光度失真对图像的颜色、亮度、对比度、饱和度等进行随机调整, 增强图像的多样性。同时选用 AerialScapes 数据集验证模型的泛化能力, 该数据集包含 3269 张 1280×720 分辨率的图像及其标注, 覆盖 11 个类别, 数据划分采用 2621 张训练集、324 张验证集和 324 张测试集的配置。

4.2 实验环境参数设计及评价指标

本实验采用 Windows11 系统, CPU 为 Intel Core i9-13900K, 显卡为 NVIDIA GeForce RTX 3090, 显存为 24 GB。Python 版本为 3.7, 使用 pytorch1.10.0 深度学习框架。实验过程中采用随机梯度下降 (Stochastic Gradient Descent, SGD) 优化器训练模型, 基础学习率设置为 0.01, 采用余弦退火策略调整学习率, 权重衰减系数设置为 1×10^{-4} , 训练轮数设置为 100, 批处理大小设置为 8。为评估模型稳定性, 将本文网络在相同设置下进行重复实验 5 次, mIoU 的标准差为 0.3%, 最终以平均值作为实验结果。

本实验所用指标主要包括平均交并比 (mIoU)、每秒浮点运算量 (FLOPs)、参数量 (Params)。其中,mIoU 用于衡量模型在语义分割任务中的整体精度。mIoU 值越高,模型的分割结果越接近真实标注,分割越准确。FLOPs 用于衡量模型的计算复杂度。FLOPs 越低,模型的计算效率越高。Params 用于衡量模型的复杂度。参数量越少,模型越轻量化。

4.3 实验

4.3.1 消融实验分析

为了对本文所设计的各模块的性能进行评估,本文在 UAVid 数据集上对方位感知模块 PAM、跨层级交互模块 CIB、轻量化分割头 LSH 及组合进行一系列消融实验,各个模块的实验结果如表 1 所示。

由表 1 可以看出,原网络在加入方位感知模块 PAM 后,通过方向解耦注意力机制,有效增强了模型对结构化目标(如道路边缘、建筑轮廓)的

空间方向建模能力。mIoU 提升了 0.62%,在计算复杂度和参数量基本不变的前提下,提高了分割精度。原网络在加入跨层级交互模块 CIB 后,强化高层语义对低层细节的指导能力,同时低层空间信息反向修正高层定位偏差,可有效抑制误检概率,显著提升分割精度。mIoU 提升了 1.41%,证明了该模块的有效性。但由于多尺度融合机制的加入,使得计算成本增加。参数量增加了 2.8 M,计算复杂度增加了 12.25 G FLOPs。原网络加入轻量化分割头 LSH 后,其浅层相加与高层拼接的特征融合策略使得计算成本大幅下降。网络的参数量下降了 0.13 M,计算复杂度减少了 8.59 G FLOPs,证明了其轻量化。在方位感知模块 PAM 基础上加入跨层级交互模块 CIB,从空间方向与尺度层级两个不同维度协同优化模型的表征能力。mIoU 达到了 67.12%,显著优于任一模块单独作用,证明了两个模块共同作用的有效性。在所有模块叠加后,计算复杂度只增加了 3.65 G FLOPs,mIoU 较原网络相比提升了约 1.7%,达到了 67.21%,实现了精度与效率的平衡。

4.3.2 对比实验分析

为了更全面地评估网络性能,分别在模型复杂度、参数量、八个类别的交并比以及平均交并比四个方面进行对比,在 UAVid 数据集上与当前主流网络进行对比,结果如表 2 所示。

(1)通过表 2 可以看出,本文网络的 mIoU 达到 67.2%,优于其他对比方法,较次优模型 PID-

表 1 各模块消融实验
Tab. 1 Ablation experiments for each module

实验	PAM	CIB	LSH	FLOPs /G	Params /M	mIoU/%
原网络				53.29	13.68	65.53
实验一	✓			53.30	13.88	66.15
实验二		✓		65.54	16.48	66.94
实验三			✓	44.70	13.55	65.29
实验四	✓	✓		65.55	16.59	67.12
实验五	✓	✓	✓	56.94	16.41	67.21

表 2 各个类别与主流网络在 UAVid 数据集上的比较结果
Tab. 2 Comparison results of various categories with mainstream networks on the UAVid dataset.

方法	FLOPs	Params	Clutter	Building	Road	Tree	LowVeg	Movin Car	Static Car	Human	mIoU
ABCNet ^[26]	62.9	14.0	<u>67.4</u>	86.4	81.2	79.9	63.1	69.8	48.4	13.9	63.8
SegFormer ^[21]	53.3	13.7	65.6	<u>87.5</u>	<u>80.7</u>	79.4	62.3	70.4	50.5	28.0	65.5
MANet ^[27]	51.7	12.0	64.5	85.4	77.8	77.0	60.3	67.2	53.6	14.9	62.6
AFFormer ^[28]	31.1	2.3	56.2	81.2	73.5	55.5	61.3	61.3	32.3	17.3	56.4
PIDNet ^[29]	<u>23.8</u>	7.6	66.7	85.9	80.5	79.4	63.0	71.5	<u>58.2</u>	28.1	<u>66.7</u>
MambaUNet ^[30]	160.3	35.9	56.6	78.5	76.1	71.5	53.0	66.4	40.5	15.9	57.3
DecoupleNet ^[31]	32.1	6.8	65.1	85.4	80.6	78.8	62.1	74.1	49.7	30.8	65.8
RSMamba ^[32]	193.5	28.6	69.5	84.2	68.8	75.2	66.5	57.7	47.7	27.9	62.7
UrbanSSF ^[33]	19.4	<u>3.6</u>	65.0	85.7	79.1	<u>79.6</u>	<u>63.3</u>	69.0	54.7	29.3	65.7
UNetMamba ^[34]	100.6	14.8	65.5	80.8	74.4	72.8	55.6	66.9	51.0	23.8	61.4
本文网络	56.9	16.4	64.3	88.3	80.5	79.9	62.6	<u>73.4</u>	58.9	<u>29.8</u>	67.2

Net提升0.5%。与参数量相近的ABCNet相比,本文方法计算复杂度降低6 G FLOPs且mIoU提升3.4%。本文网络在参数量上与UNetMamba相近,但其计算复杂度仅为UNetMamba的56.6%且mIoU高出5.8%。这表明本文网络在整体分割性能上具有明显的优势。

(2)相较于ABCNet,本文网络在大目标分割上的表现相当。由于跨层级交互模块的多尺度特征融合,使得本文网络在小目标分割上展现出显著优势,尤其在Human类别上,分割精度高出15.9%。在与其他网络的对比中,本文网络对MovingCar,StaticCar,Human这三种小目标的分割精度均位于前列,证明了该网络在小目标分割方面的优势。

(3)本文网络与轻量化分割网络AFFormer相比,分割精度高出10.8%。与MambaUNet相比,在参数量和计算复杂度均有显著优势的情况下,分割精度也高出9.9%。相较于原网络,在计算复杂度和参数量略有增加的情况下,分割精度提升1.7%,体现了本文网络在性能方面的优秀表现。

为了更直观地对比本文所提出网络与其他主流语义分割模型在UAVid数据集上的性能表现,本文选取了多个代表性网络的分割结果进行可视化展示。对比图如图5所示。

如图5所示,(a)列为输入图像,(b)列为标签,(c)列到(f)列分别为AFFormer,DecoupleNet,SegFormer以及本文网络的预测结果,在各

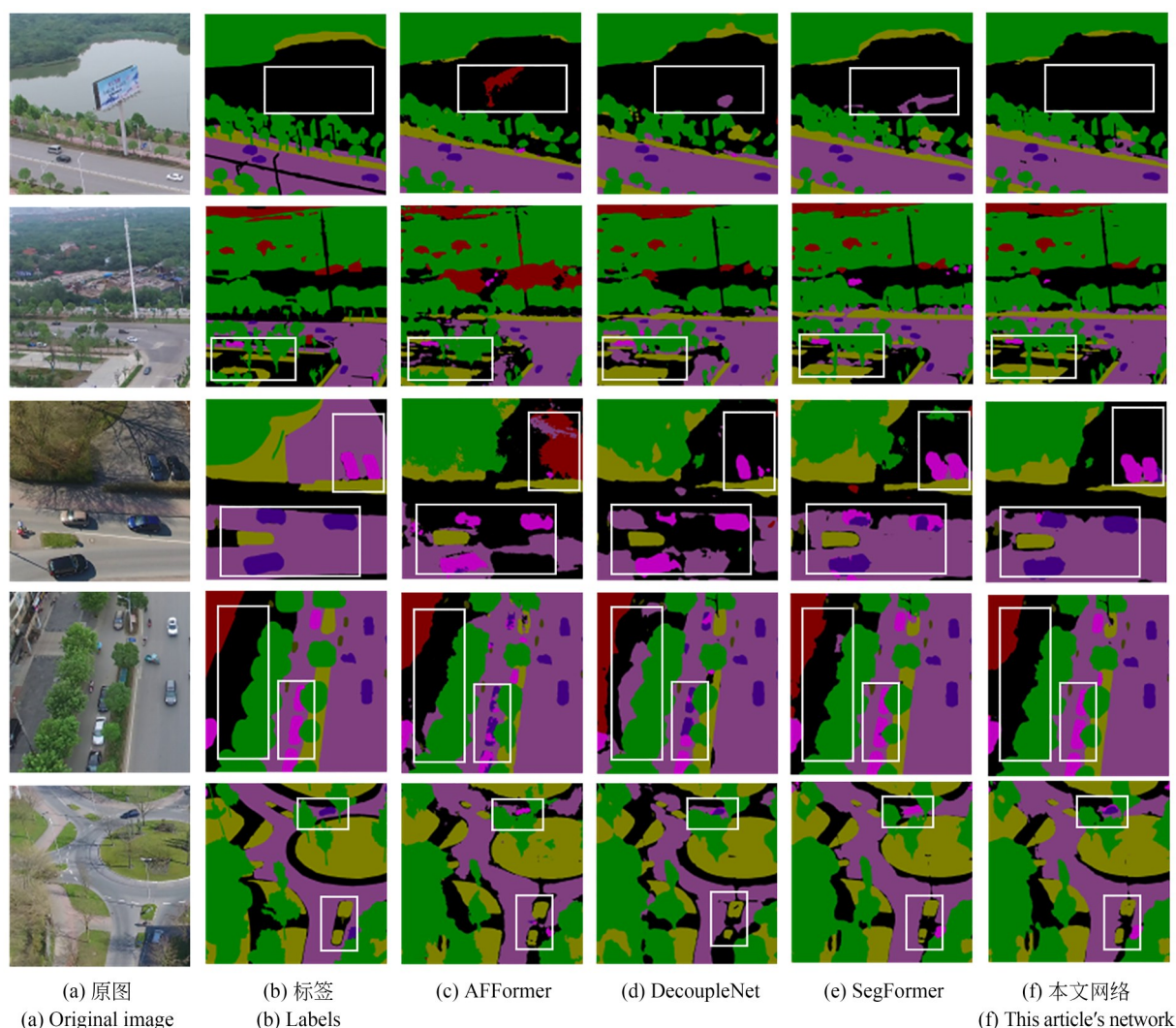


图5 部分网络在UAVid数据集上的可视化对比结果

Fig. 5 Visual comparison results of some networks on the UAVid dataset.

图像中,使用白色方框标注了分割效果较为突出的区域,以便于观察和分析。

在第一、二行可以看出,在标注区域,AFFormer,DecoupleNet,SegFormer在识别黑色背景时存在明显的误检现象,而本文算法能够正确识别黑色背景,降低误检概率,展现出更好的分割效果;在第三、四行中,AFFormer,DecoupleNet,SegFormer在识别移动的汽车与静止的汽车时表现不佳,而本文算法能够正确识别这两个类别,并生成更加流畅准确的边界;在第五行图像对比中可以看出,面对复杂环境下的语义分割,本文网络在识别移动的车这一小目标时表现出更好的效果,且能够完整识别物体。综上所述,本文网络在航拍图像语义分割中表现效果良好且具有较高的准确率。

为了验证本文网络的泛化性,将本文网络与其他网络在Aeroscapes数据集上的分割结果进行对比实验,结果如表3所示。从表3中可以看出,本文网络在Aeroscapes数据集上mIoU达到67.4%,相较于基线模型SegFormer提升了1.3%,表明本文网络在保持原有优势的基础上,进一步优化了模型的性能。与计算复杂度和参数量相当的ABCNet相比,mIoU超出6.2%。与其他网络相比,在计算复杂度与参数量位于中间

表3 与主流网络在Aeroscapes数据集上的比较结果

Tab. 3 Comparison results with mainstream networks on the Aeroscapes dataset

方法	FLOPs/G	Params/M	mIoU/%
ABCNet ^[26]	62.9	14.0	61.2
SegFormer ^[21]	53.3	13.7	<u>66.1</u>
MANet ^[27]	51.7	12.0	62.7
AFFormer ^[28]	<u>31.1</u>	2.3	52.3
PIDNet ^[29]	23.8	7.6	62.5
MambaUNet ^[30]	160.3	35.9	41.2
DecoupleNet ^[31]	32.1	<u>6.8</u>	57.8
RSMamba ^[32]	193.5	28.6	32.5
UAVSeg ^[35]	63.4	14.8	40.3
UNetMamba ^[34]	100.6	14.8	50.6
本文网络	56.9	16.4	67.4

的情况下,本文网络的mIoU值达到最高,充分证明了其卓越的泛化能力。通过在不同数据集上的实验验证,本文网络在复杂的实际场景中展现了更强的适应性和更高的分割精度,验证了本文模型的泛化能力。

为了更直观地对比本文网络与其他网络在Aeroscapes数据集上的性能表现,本文选取了多个代表性网络分割结果图进行可视化展示,结果图如图6所示。

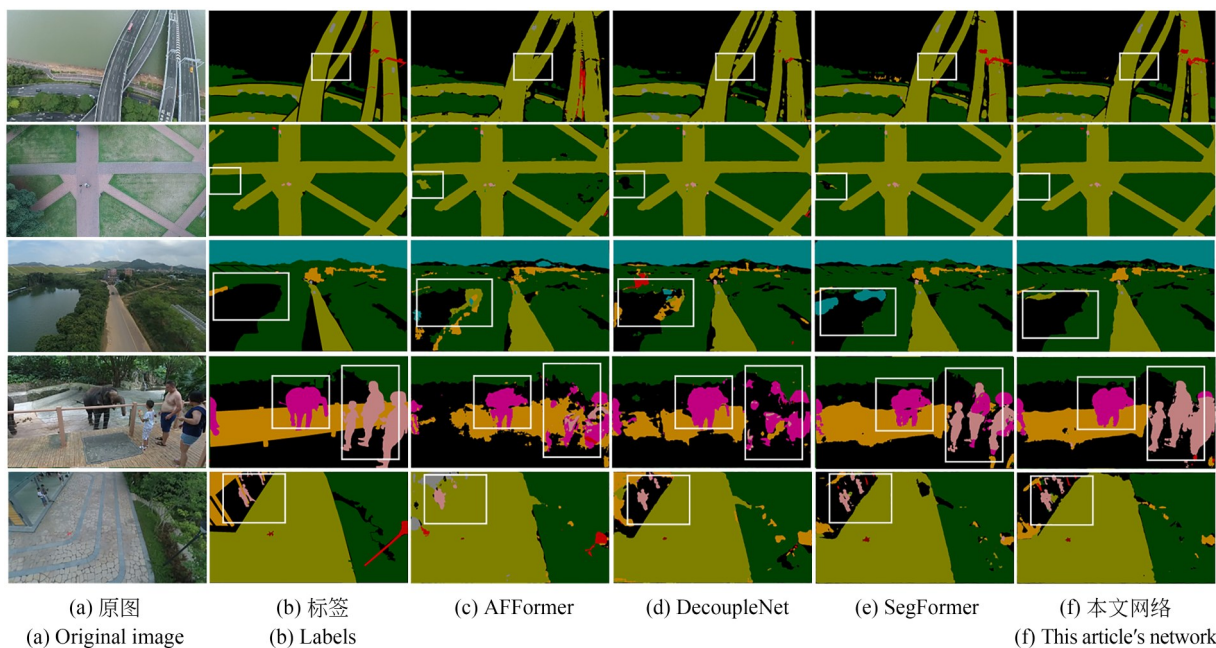


图6 部分网络在Aeroscapes数据集上的可视化对比结果

Fig. 6 Visual comparison results of some networks on the Aeroscapes dataset

从第一行可以看出,在标注区域,AFFormer, DecoupleNet, SegFormer在分割目标时存在粘连、边界不清晰的问题。由于方位感知模块的方向解耦注意力机制能增强模型对边缘、轮廓等方向敏感区域的感知能力,因此本文网络能正确识别边界,缓解了边界模糊的问题;从第二、三、四行可以看到,AFFormer, DecoupleNet, SegFormer存在着明显的误检现象,而本文网络则明显降低了误检概率且能正确识别轮廓;从第五行中可以看出本文网络在识别小目标方面的优越性,AFFormer, DecoupleNet, SegFormer并未完全识别到所有人,且识别到的人类并不完整。仅有本文网络识别到的人类数量更接近于标签真值,证明了本文网络在识别人类这一类别的有效性,也缓解了小目标漏检的现象。可视化结果对比表明,通过方位感知模块、跨层级交互模块和轻量化分割头的协同作用,本文网络在目标边界、小尺度目标以及复杂场景下的分割效果方面均表现出明显优势。

5 结 论

本文针对 SegFormer 在航拍图像语义分割任务中容易造成小目标漏检和边界模糊的问题,

提出了一种跨层级交互与方位感知的航拍图像的语义分割网络。通过设计方位感知模块提升网络对空间方向信息的感知利用能力,有效地提升了航拍图像语义分割精度。通过设计跨层级交互融合模块实现多尺度跨层级交互,同时采用动态分组的机制设计通道-空间注意力机制,促进不同组特征之间的信息交流,提升模型在复杂场景的分割能力。设计轻量化分割头降低计算开销,实现了轻量化。本文网络在 UA-Vid 数据集上 mIoU 达到了 67.2%,相较于原网络分割精度提升了 1.7%,在 Aeroscapes 数据集上 mIoU 达到了 67.4%。本文网络在高分辨率、目标密集、背景干扰强的航拍场景中表现出优异的分割性能,为无人机实时系统、城市遥感监测、灾害评估等实际应用场景提供了切实可行的技术方案。

作者贡献声明:

刘杰:实验方法与算法改进的指导,论文的审核与数据的分析;

吴紫钰:算法改进,论文的撰写与修改,数据的分析与整理;

田明:论文审核与编辑写作;

韩轲:测量实验数据分析。

参考文献:

- [1] 叶兆元,郑凯扬,张亮智,等. 图像语义分割的深度学习方法综述[J]. 中国设备工程, 2025(10): 23-25.
YE Z Y, ZHENG K Y, ZHANG L Z, *et al.* Overview of deep learning methods for image semantic segmentation [J]. *China Plant Engineering*, 2025 (10): 23-25. (in Chinese)
- [2] 柳长源,郭鹏岗,兰朝凤. 注意力置换与通道重建的无人机城市街景实时语义分割[J]. 控制与决策, 2025, 40(4): 1198-1206.
LIU C Y, GUO P G, LAN C F. Real-time semantic segmentation of UAV urban street scenes with attention permutation and channel reconstruction [J]. *Control and Decision*, 2025, 40 (4): 1198-1206. (in Chinese)
- [3] 张伟,刘潇霜,马宇张,等. 基于图卷积网络的疟疾感染红细胞图像识别[J]. 中国医学物理学杂志, 2025, 42(5): 606-612.
ZHANG W, LIU X S, MA Y Z, *et al.* Image recognition of malaria-infected erythrocytes based on graph convolutional network [J]. *Chinese Journal of Medical Physics*, 2025, 42(5): 606-612. (in Chinese)
- [4] SILVA M, SEOANE A, MURES O A, *et al.* Exploring the effects of synthetic data generation: a case study on autonomous driving for semantic segmentation [J]. *The Visual Computer*, 2025, 41 (10): 7379-7397.
- [5] ZHANG S J, ZHANG B, WU Y T, *et al.* Seg-CLIP: multimodal visual-language and prompt learning for high-resolution remote sensing semantic segmentation [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5646316.
- [6] ZHENG P J, FANG X J, ZHOU Z P, *et al.* Accurate and efficient privacy-preserving feature extraction on encrypted images [J]. *IEEE Transactions on Dependable and Secure Computing*, 2025, 22(4):

- 3192-3207.
- [7] XU J, CHENG M, LI B, *et al.* Oil slick identification in marine radar image using HOG, random forest, and PSO [J]. *IEEE Geoscience and Remote Sensing Letters*, 2024, 21: 1504305.
- [8] HAN Y H, FU Y J, TIAN S Y, *et al.* An improved Fourier transform profilometry and phase coding for 3D shape measurement [J]. *Optical Review*, 2025, 32(3): 486-496.
- [9] WANG J, ZHANG G Y, GUO W, *et al.* “3-2-1” PMP: adding an extra pattern to dual-band phase shift profilometry for higher precision 3-D imaging [J]. *IEEE Transactions on Instrumentation and Measurement*, 2024, 73: 5025814.
- [10] AI J, ZHANG Q C, SU X Y. Fast modulation measurement profilometry based on phase-shifting and modulation ratio [J]. *Optics Express*, 2025, 33(10): 21707-21718.
- [11] 曹益平, 张鹤晨. 计算莫尔轮廓术及其发展动态 [J]. *光学与光电技术*, 2023, 21(5): 1-23.
- CAO Y P, ZHANG H C. Computer-generated moiré profilometry and its development trends [J]. *Optics & Optoelectronic Technology*, 2023, 21(5): 1-23. (in Chinese)
- [12] BEHLEY J, GARBADE M, MILIOTO A, *et al.* SemanticKITTI: a dataset for semantic scene understanding of LiDAR sequences [C]. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*. October 27-November 2, 2019. Seoul, Korea. IEEE, 2019: 9296-9306.
- [13] SUGIRTHA T, SRIDEVI M. Semantic segmentation using modified U-Net for autonomous driving [C]. 2022 *IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*. June 1-4, 2022, Toronto, ON, Canada. IEEE, 2022: 1-7.
- [14] CHEN L C, PAPANDREOU G, KOKKINOS I, *et al.* DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 834-848.
- [15] ZHENG S X, LU J C, ZHAO H S, *et al.* Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers [C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 20-25, 2021, Nashville, TN, USA. IEEE, 2021: 6877-6886.
- [16] LIU Z, LIN Y T, CAO Y, *et al.* Swin transformer: hierarchical vision transformer using shifted windows [C]. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*. October 10-17, 2021, Montreal, QC, Canada. IEEE, 2022: 9992-10002.
- [17] 李智杰, 惠爱婷, 李昌华, 等. 面向遥感图像道路提取的多尺度上下文感知网络 [J]. *光学精密工程*, 2025, 33(4): 610-623.
- LI Z J, HUI A T, LI C H, *et al.* Multi-scale context-aware network for road extraction in remote sensing images [J]. *Opt. Precision Eng.*, 2025, 33(4): 610-623. (in Chinese)
- [18] 闫河, 雷秋霞, 王旭. 融合注意力机制的改进型 DeepLabv3+ 语义分割 [J]. *光学精密工程*, 2025, 33(1): 123-134.
- YAN H, LEI Q X, WANG X. Improved DeepLabv3+ semantic segmentation incorporating attention mechanisms [J]. *Opt. Precision Eng.*, 2025, 33(1): 123-134. (in Chinese)
- [19] 贾轩, 张叶, 常旭岭, 等. 基于深度度量学习和语义分割的场景识别 [J]. *液晶与显示*, 2025, 40(5): 740-750.
- JIA X, ZHANG Y, CHANG X L, *et al.* Scene recognition based on deep metric learning and semantic segmentation [J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(5): 740-750. (in Chinese)
- [20] 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 C 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. (in Chinese)
- [21] XIE E Z, WANG W H, YU Z D, *et al.* SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers [EB/OL]. 2021: arXiv: 2105.15203. <https://arxiv.org/abs/2105.15203>
- [22] CHENG J J, LIU W J, WANG Z F, *et al.* Joint event extraction model based on dynamic attention matching and graph attention networks [J]. *Scientific Reports*, 2025, 15: 6900.
- [23] QIN J Y, JIANG Y S, CAO Y P, *et al.* Single-

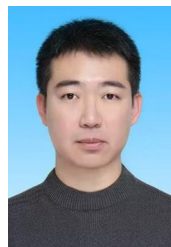
- shot phase-shifting composition fringe projection profilometry by multi-attention fringe restoration network [J]. *Neurocomputing*, 2025, 634: 129908.
- [24] SUBBAIAH B, MURUGESAN K, SARAVANAN P, *et al.* An efficient multimodal sentiment analysis in social media using hybrid optimal multi-scale residual attention network[J]. *Artificial Intelligence Review*, 2024, 57(2): 34.
- [25] BIAN J L, ZHANG X, ZHANG X Y, *et al.* MCANet: shared-weight-based MultiheadCrossAttention network for drug-target interaction prediction [J]. *Briefings in Bioinformatics*, 2023, 24(2): bbad082.
- [26] LI R, ZHENG S Y, ZHANG C, *et al.* ABCNet: Attentive bilateral contextual network for efficient semantic segmentation of Fine-Resolution remotely sensed imagery[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2021, 181: 84-98.
- [27] LI R, ZHENG S Y, ZHANG C, *et al.* Multiattention network for semantic segmentation of fine-resolution remote sensing images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5607713.
- [28] DONG B, WANG P C, WANG F. Head-free lightweight semantic segmentation with linear transformer [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023, 37(1): 516-524.
- [29] XU J C, XIONG Z X, BHATTACHARYYA S P. PIDNet: a real-time semantic segmentation network inspired by PID controllers[C]. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 17-24, 2023, Vancouver, BC, Canada. IEEE, 2023: 19529-19539.
- [30] WANG Z Y, ZHENG J Q, ZHANG Y C, *et al.* Mamba-UNet: UNet-like Pure Visual Mamba for Medical Image Segmentation [EB/OL]. 2024: arXiv: 2402.05079. <https://arxiv.org/abs/2402.05079>
- [31] LU W, CHEN S B, SHU Q L, *et al.* DecoupleNet: a lightweight backbone network with efficient feature decoupling for remote sensing visual tasks [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 4414613.
- [32] ZHAO S J, CHEN H, ZHANG X L, *et al.* RS-mamba for large remote sensing image dense prediction[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5633314.
- [33] WANG Z J, YI J Z, CHEN A B, *et al.* Accurate semantic segmentation of very high-resolution remote sensing images considering feature state sequences: From benchmark datasets to urban applications [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2025, 220: 824-840.
- [34] ZHU E Z, CHEN Z, WANG D K, *et al.* UNet-Mamba: an efficient UNet-like mamba for semantic segmentation of high-resolution remote sensing images[J]. *IEEE Geoscience and Remote Sensing Letters*, 2025, 22: 6001205.
- [35] WANG Z, YOU Z H, XU N, *et al.* UAVSeg: dual-encoder cross-scale attention network for UAV images' semantic segmentation [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 5600117.

作者简介:



刘 杰(1980—),女,黑龙江哈尔滨人,博士,副教授,研究生导师,2005年于哈尔滨工业大学获得硕士学位,2013年于哈尔滨理工大学获得博士学位,主要研究方向为人工智能与图像处理大数据模型及预测。E-mail: liujie@hrbust.edu.cn

通讯作者:



韩 轲(1980—),男,博士,教授,主要研究方向为数据挖掘与人工智能算法及其应用。E-mail: hanke@hrbcu.edu.cn