

基于改进 U-Net 的遥感图像道路检测

张桐瑞¹, 杨延宁^{1,2}, 张佳豪¹

(1. 延安大学 物理与电子信息学院, 陕西 延安 716000;

2. 陕西省能源大数据智能处理省市共建重点实验室, 陕西 延安 716000)

摘要:针对复杂遥感影像中因道路材质多样、背景复杂和阴影遮挡等因素所导致的识别精度不高、现有模型参数量大等问题,设计了一种基于 U-Net 的改进模型,用于遥感图像道路检测;在编码器中采用 MobileNetV2 替代原主干网络以降低计算复杂度;在解码器中引入深度可分离卷积以高效恢复高分辨率细节,并结合空间组增强注意力机制提升空间特征建模能力;实验结果表明,相较于对照模型,改进模型在马萨诸塞州道路数据集上的 $mIOU$ 和 F_1 分别达到了最高的 79.50% 和 87.46%;实现了检测性能与模型复杂度之间的有效平衡,为遥感影像中的道路提取提供了有力支持。

关键词: U-Net; MobileNetV2; 深度可分离卷积; 注意力机制; 道路检测

Road Detection in Remote Sensing Images Based on Improved U-Net

ZHANG Tongrui¹, YANG Yanning^{1,2}, ZHANG Jiahao¹

(1. School of Physics and Electronic Information, Yan'an University, Yan'an 716000, China;

2. Shaanxi Provincial Key Laboratory of Energy Big Data Intelligent Processing, Yan'an 716000, China)

Abstract: To address the challenges of low recognition accuracy caused by diverse road materials, complex backgrounds, and shadow occlusions in complex remote sensing images, as well as the large parameter size of existing models, this paper proposes an improved U-Net-based model for detecting roads in remote sensing images. In the encoder, the MobileNetV2 is used to replace the original backbone network, significantly reducing computational complexity while maintaining detection accuracy. In the decoder, a depthwise separable convolution is employed to efficiently recover high-resolution details, and a spatial group attention mechanism is integrated to enhance the modeling of spatial hierarchical features. Experimental results on the Massachusetts road dataset demonstrate that, compared with the baseline model, the proposed model achieves a maximum mean intersection over union ($mIOU$) of 79.50% and an F_1 -score of 87.46%, respectively, which shows that the proposed approach effectively make a balance between detection performance and model complexity, providing a practical solution for extracting roads from remote sensing images.

Keywords: U-Net; MobileNetV2; depthwise separable convolution; attention mechanism; road detection

0 引言

近年来,我国交通基础设施建设取得显著成效,公路网络规模持续扩大,对高效、精准的道路信息获取提出了更高要求。而遥感图像作为一种获取地表信息的重要手段,已在道路信息提取中展现出巨大优势^[1-2]。准确的遥感图像道路提取在城市道路规划、导航与定位服务、灾害应急响应等领域具有重要的意义^[3-4]。

基于传统机器学习的遥感图像道路提取方法通常依赖于图像的几何特征、光谱特征、纹理特征等信息,通

过规则设计或模型学习进行道路识别和提取。一种基于多特征融合的高分辨率遥感影像道路中心线提取方法被提出^[5]。该方法从光谱和空间两方面提取道路特征,通过多特征融合构建初始道路网络,结合形状特征进行网络优化,并利用张量投票算法提取精细的道路中心线。

作为一种经典的机器学习方法,随机森林被用于构建面向遥感图像检测的算法^[6]。该算法先利用熵率算法进行超像素分割,提取 Gabor 特征和光谱特征作为输入,训练随机森林模型,并通过多决策树投票生成最终检测结果。这些传统方法通常依赖人工特征设计和选

收稿日期:2025-06-30; 修回日期:2025-08-08。

基金项目:国家自然科学基金(52365069)。

作者简介:张桐瑞(2002-),男,硕士研究生。

通讯作者:杨延宁(1969-),男,博士,教授。

引用格式:张桐瑞,杨延宁,张佳豪. 基于改进 U-Net 的遥感图像道路检测[J]. 计算机测量与控制,2026,34(6):42-48.

择, 不仅增加了特征工程的计算成本, 还需要多次迭代优化模型。传统方法包含特征提取、降维、分类等多个环节, 难以满足实时检测需求。

随着深度学习与人工智能技术的持续进步, 遥感图像中的道路提取正朝着自动化与智能化方向加速发展, 已成为近年来广泛关注的研究热点之一^[7]。遥感图像能够获取地表物体的高分辨率信息, 而深度学习技术凭借其强大的特征提取能力, 近年来在遥感图像分割任务中得到了广泛应用^[8]。现有研究在遥感图像分割领域提出了多种结构改进与融合策略。其中, 多尺度跳跃连接结合注意力机制可整合不同层级特征, 显著提升边缘区域分割精度^[9]; 轻量化方向的工作将 U-Net 与 ShuffleNetV2 融合, 并引入逐点卷积及卷积块注意力机制 (CBAM), 以增强深层特征提取能力^[10]。面向多方向目标特征的任务中, 旋转多尺度交互网络通过尺度内交互与自适应旋转卷积捕捉细粒度特征, 从而提升性能^[11]。道路提取方法则利用改进的深度残差网络结合空洞感知模块, 并在后处理阶段引入数学形态学与张量投票以优化结果^[12]。针对特征不一致与大目标碎片化问题, 双分支语义分割网络融合动态权重、语义筛选注意力及空间-语义信息耦合模块, 有效改善精度^[13]。将高光谱图像与深度学习结合, 并在改进 U-Net 结构中引入注意力机制与失活层的设计, 可同时提升分割精度与检测速度^[14]。这些方法在提升算法性能方面具有一定优势, 但都伴随着较大的计算开销, 难以满足对实时性和高效性的需求。此外, 遥感影像中背景信息复杂, 易导致分割精度下降, 同时边缘特征易发生丢失, 影响提取结果的完整性与准确性。因此, 在保证结果精度的同时, 如何有效降低计算复杂度仍是需要进一步探讨的问题。

为了解决上述问题, 本文提出了一种改进模型, 首先引入 MobileNetV2 轻量级网络结构作为主干提取网络, 显著降低了模型计算参数量, 提升了模型的推理速度, 增强了模型的特征提取能力; 其次, 将模型中解码器的普通卷积替换为深度可分离卷积 (DWConv, depthwise separable convolution), 结合 MobileNetV2 形成整体轻量级网络, 减少过拟合风险; 最后, 在解码器上采样后加入空间组增强注意力机制 (SGE, spatial group-wise enhance), 增强了恢复空间细节和边界信息时的特征表达能力, 并提升解码质量。通过消融实验, 并与现有主流模型进行对比实验, 以全面验证所提出模型在分割精度与模型效率方面的优势。

1 模型结构设计

1.1 MobileNetV2 网络

MobileNetV2 由 Sandler 等人^[15]提出, 作为一种轻

量级网络结构, 可灵活集成至多种语义分割框架中。凭借其出色的边缘特征提取能力和高效的结构设计, MobileNetV2 尤其适用于高分辨率遥感图像的分析任务。相比 MobileNetV1, MobileNetV2 引入了倒残差结构与线性瓶颈设计, 并在激活函数选择方面进行了关键改进。在 MobileNetV2 中, 采用了 ReLU6 激活函数替代传统的 ReLU, 以在保留非线性能力的同时, 通过对激活值施加上限约束, 有效缓解特征信息在低维空间中的损失。传统 ReLU 函数在负值输入时输出为零, 可能导致信息截断与梯度传播障碍, 尤其在低维特征表示中影响更为显著。而 ReLU6 将激活值限制在 $[0, 6]$ 内, 有助于降低过大激活值对网络训练稳定性的影响, 特别适用于轻量级网络设计。

在结构设计方面, MobileNetV2 首先通过一个 1×1 卷积层对输入通道进行升维, 将特征映射扩展至高维空间; 随后在高维空间中应用深度可分离卷积以高效提取空间特征; 最终通过另一个 1×1 卷积层进行线性降维。该降维操作采用线性激活, 以避免在低维空间中丢失关键信息。该结构能够在保证模型非线性表达能力的同时, 最大限度保留特征信息, 并支持残差连接的引入, 从而促进梯度的有效传播, 提升模型训练的稳定性。而且线性瓶颈的设计有助于保留道路的细长形态与边缘细节信息, 多尺度特征提取机制与 U-Net 的跳跃连接高度契合, 从而在复杂背景、材质变化及阴影干扰下, 依然能够有效提升道路检测的精度与鲁棒性。

在将 MobileNetV2 作为 U-Net 编码器主干网络时, 需对其结构进行阶段性划分, 以实现与解码器部分的有效对接。MobileNetV2 主干由多个倒残差模块构成, 不同于原始 U-Net 所采用的逐层输出特征, 用于跳跃连接的编码结构。因此, 为实现跳跃连接与解码器各层空间维度对齐, 以网络中 $\text{stride}=2$ 的倒残差模块作为阶段边界, 用这些特征图作为跳跃连接中不同尺度的输入来源, 如图 1 (a) 所示。与此同时, 各阶段内部包含的 $\text{stride}=1$ 的倒残差模块, 虽不改变特征图的空间尺寸, 但通过深度堆叠, 有效提升网络在当前尺度下的特征建模能力, 增强语义表达, 同时借助残差连接促进梯度传播与信息保留, 如图 1 (b) 所示。

1.2 深度可分离卷积

深度可分离卷积是一种高效的卷积运算方式, 其核心思想在于将标准卷积操作解耦为两个独立的阶段, 从而显著降低计算复杂度并提升模型性能。第一阶段为逐通道卷积, 该操作在不进行跨通道信息交互的前提下, 对每个输入通道分别执行空间卷积, 用以提取局部空间特征; 第二阶段为逐点卷积, 通过使用 1×1 卷积核在每个空间位置上对所有输入通道进行线性组合, 实现通道间特征的融合与重构。深度可分离卷积因其将空间特

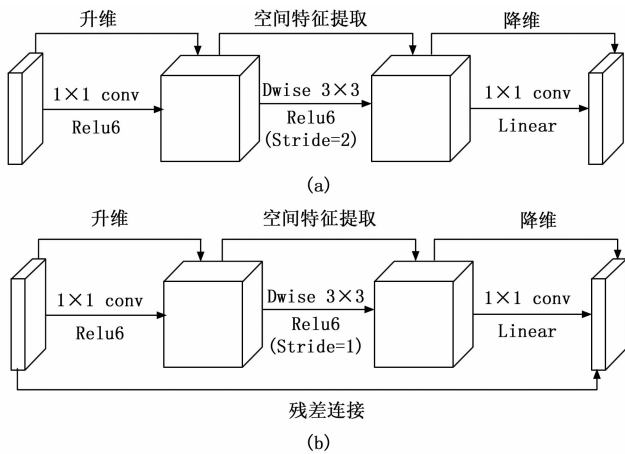


图 1 MobileNetV2 结构图

征提取与通道特征融合解耦的结构设计,使模型在保持较低计算复杂度的同时,具备更强的细粒度特征提取能力,特别适用于遥感图像中道路等细长结构的提取任务。这是因为其逐通道卷积 (Depthwise Convolution) 部分可对每个通道进行独立的空间特征建模,有助于捕捉细小的边缘、纹理和具有方向性的结构,避免了通道之间过早混合可能导致的细节丢失。而逐点卷积 (Pointwise Convolution) 则在保持局部结构完整的基础上完成跨通道的信息整合,从而提升整体语义表达能力。

相比之下,传统的普通卷积在同时进行空间和通道融合的过程中,容易模糊目标边界,尤其在遥感图像道路检测任务中处理狭窄且形态连续的结构时,可能出现边缘不清、结构断裂等问题。如图 2 所示,深度可分离卷积中的卷积核为单通道卷积核,即每个卷积核仅作用于输入特征图的一个通道,分别对各通道独立进行卷积操作。与标准卷积相比,这种分解方式显著减少了参数量和计算量。

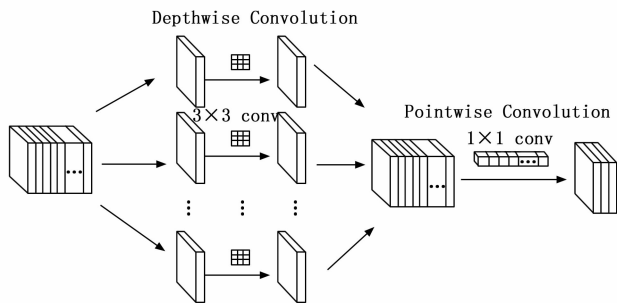


图 2 深度可分离卷积

1.3 SGE 注意力机制

SGE^[16] 模块在通道维度上将特征图划分为 8 个子组,并在各子组内独立实施注意力建模。分组机制通过将具有相似语义的特征聚合到同一子组中,有效保持了子组内部的语义一致性。每组内部进行独立的空间增强,随后重组以还原原始结构,从而在保留局部细节的

同时提升模型的轻量化特性。在遥感图像道路检测任务中,该结构显著提升了特征表达的精度与区分能力。SGE 模块采用轻量化的全局设计策略,在几乎不增加额外参数和计算开销的前提下,有效增强了模型的表达能力与鲁棒性,体现出较强的实际应用潜力。其模型结构如图 3 所示。

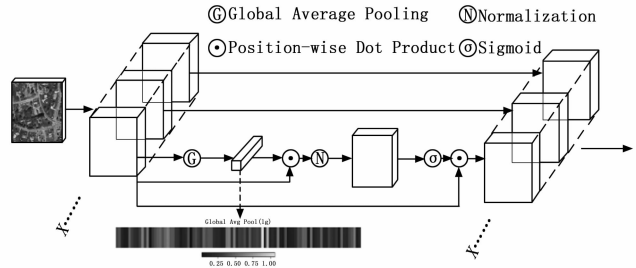


图 3 SGE 注意力机制结构图

给定输入图像,图像经过处理后生成多个特征图,每张特征图表示卷积神经网络中不同的视觉信息(如边缘、纹理、形状等)。然后,应用全局平均池化 (GAP, global average pooling) 将每张特征图的信息压缩为一个单一的向量,表示每个通道的全局特征。对于每个通道 c ,通过对特征图的空间维度进行平均,得到全局特征向量 z_c :

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_{ijc} \quad (1)$$

其中: X_{ijc} 表示位置 i, j 和通道 c 上的特征值。

全局平均池化的输出与输入特征图的每个位置特征进行点积操作,计算每个像素的权重:

$$a_{ijc} = z_c \cdot X_{ijc} \quad (2)$$

计算得到的注意力权重 a_{ijc} 使用 Sigmoid 函数进行归一化处理,确保权重在 $[0, 1]$ 内。最后,通过将归一化后的注意力权重 $\hat{a}_{ijc}$ 与每个通道的特征图 X_c 进行逐像素相乘,得到加权后的特征图 $\hat{X}_c$:

$$\hat{X}_c = \hat{a}_{ijc} \cdot X_{ijc} \quad (3)$$

其中: $\hat{X}_c$ 是最终的加权特征图,包含了增强的空间信息。

2 模型整体框架

鉴于传统 U-Net^[17] 模型在面临大规模数据和高计算复杂度任务时易产生过拟合现象,本文提出了一种改进的轻量级 U-Net 结构,以在减少参数量的同时提升模型的预测精度。整体网络结构如图 4 所示,所提出的模型采用 MobileNetV2 作为编码器主干网络,保留了 U-Net 的对称结构特点。图中标注的 IRB-1 (stride=1) 与 IRB-2 (stride=2) 分别对应于 MobileNetV2 中的倒残差模块 (IRB, inverted residual block),通过其高效的深度可分离卷积与线性瓶颈结构,实现特征提取与逐

层下采样。在每个尺度下采样后,编码器提取的特征将通过跳跃连接(Skip Connection)传递至解码器相应层,保留空间细节信息,从而实现多尺度语义特征的有效融合。

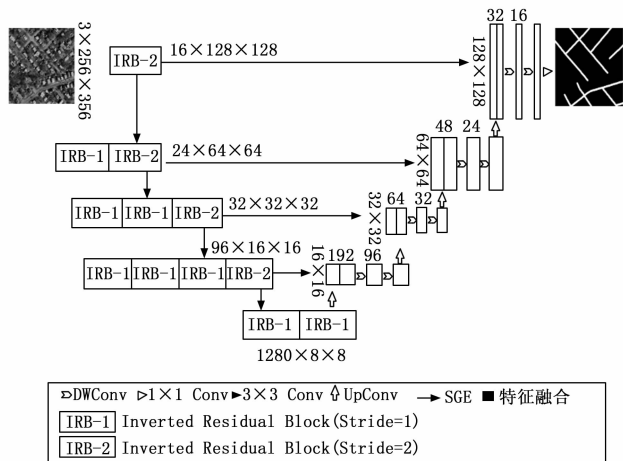


图4 改进后的 U-Net 网络框架

解码器部分采用逐层上采样的方式逐步恢复特征图的空间分辨率,并在每一阶段引入空间组增强注意力机制,用以增强显著区域的特征表达能力。在每一级解码模块中,将上采样后的特征图与编码器对应层通过跳跃连接传递过来的高分辨率特征图沿通道维度进行拼接,以整合编码器提供的高分辨率细节特征与解码器逐步还原的全局语义特征。拼接后特征图依次经过两次深度可分离卷积操作:第一层卷积用于融合拼接后的特征并进行通道压缩,减少冗余信息;第二层卷积进一步提取关键语义特征,增强特征表达的判别性,从而提升最终分割结果的精细度和准确性。对于遥感图像中具有细长形态且与背景颜色差异较小的道路目标,改进模型显著提升了对细小目标与边缘轮廓的感知能力。该模型在保持空间细节信息完整性的同时,能够有效增强对低对比度目标的检测,突出道路区域的特征表达,从而提升对道路结构的完整性建模能力与检测精度。

3 实验与分析

3.1 数据集介绍

本文采用马萨诸塞州道路数据集(Massachusetts Roads Dataset),该数据集由1 171幅航拍图像构成,每幅图像尺寸为1 500×1 500像素,单幅图像覆盖区域约为2.25平方千米。为适应深度学习模型的输入需求,并增强训练样本的多样性与泛化能力,对原始数据进行了预处理操作。具体而言,采用滑动窗口策略以256×256像素为单位对图像及其对应的标签图进行裁剪,裁剪过程采用固定步长实现图像块的系统性提取。在裁剪完成后,依据统一的随机划分策略,将生成的样本划分

为训练集、验证集与测试集,分别包含9 620张训练图像、202张验证图像和577张测试图像。该预处理过程有效提升了数据利用率,为后续模型的训练与评估提供了充分的数据支持。

3.2 评价指标

本实验采用了语义分割任务中常见的多种评价指标,以全面评估模型的性能表现,包括 F_1 、平均交并比(mIOU, mean intersection over union)、总体精度(OA, overall accuracy)、召回率(R, recall)以及精确率(P, precision)。其中,mIOU是最具代表性的评价指标,用于衡量模型预测区域与真实标注区域之间的重叠程度。该指标综合考虑了假正例(FP)与假负例(FN),对类别不平衡具有较强的鲁棒性,广泛应用于语义分割领域作为主指标。

F_1 是衡量分类平衡性的重要指标,尤其适用于存在类别不均衡的数据情形,能够反映模型在各类别上的综合判别能力。每个类别均单独计算 F_1 值,并最终取平均以衡量整体性能。其计算方式为精确率(Precision)与召回率(Recall)的调和平均,如公式(4)所示:

$$F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

mIOU通过计算每个类别预测结果与真实标签之间的重叠程度,衡量预测的准确性,然后对所有类别的IOU值取平均,如公式(5)所示:

$$mIOU = \frac{1}{C} \sum_{c=1}^C IOU_c = \frac{1}{C} \sum_{c=1}^C \frac{TP}{TP + FP + FN} \quad (5)$$

OA则用于衡量模型预测正确的像素数量占总像素数量的比例,是一种较为直观的整体性能评估方式,如公式(6)所示:

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Recall是评价模型识别正类样本能力的指标,反映了模型在所有实际正类样本中正确识别出的比例,如公式(7)所示:

$$R = \frac{TP}{TP + FN} \quad (7)$$

Precision用于分析模型预测偏差,表示在所有被预测为正类的像素中,实际为正类的比例,其定义如公式(8)所示:

$$P = \frac{TP}{TP + FP} \quad (8)$$

上述多个指标的联合使用,有助于从不同维度对模型进行综合性评价,既反映了整体准确率,又揭示了模型在各类别上的判别能力与偏差特征。

3.3 实验平台具体参数设计

所有实验均在Windows11专业版操作系统环境下开展,硬件平台为Intel Core i5-12600KF处理器与

NVIDIA GeForce RTX 4060 图形处理器, 配备 8GB 显存。软件环境方面, 选用 Python 3.8 作为开发语言, 深度学习框架为 PyTorch 2.0.1, 并通过 CUDA 11.8 加速实现 GPU 计算。

模型设置为二分类任务, 采用 Adam 优化器, 初始学习率 1×10^{-4} , 批量大小 16, 训练 100 轮, 数据加载线程数 4。为提升泛化性能与抑制过拟合, 在网络中引入丢弃率为 0.1 的 Dropout 及正则化项以约束参数幅度并增强训练稳定性。

表 1 超参数设置对比

学习率	Dropout	$F_1 / \%$	$mIOU / \%$
1×10^{-3}	0.1	85.24	76.67
1×10^{-3}	0.3	83.93	75.08
1×10^{-3}	0.5	85.83	77.39
1×10^{-4}	0.1	87.46	79.50
1×10^{-4}	0.3	85.26	76.69
1×10^{-4}	0.5	82.78	73.73

由表 1 可见, 学习率设为 1×10^{-4} 时的 F_1 分数与 $mIoU$ 均整体优于 1×10^{-3} , 表明较小学习率更适合当前模型与数据, 有助于提升收敛质量与整体性能。在较小学习率条件下, 模型已具备较强泛化能力, 过高的 Dropout 会削弱有效特征的学习能力, 导致性能下降。当学习率为 1×10^{-4} , 且 Dropout 取 0.1 时, 模型性能达到最优。

在损失函数设计上, 采用了联合损失函数策略, 将交叉熵损失 (Cross-Entropy Loss) 与 Dice 损失 (Dice Loss) 进行加权融合, 以兼顾类别分布不均衡与再细小。具体地, 损失函数加权组合比例设置为: 交叉熵损失占比 0.8, Dice 损失占比 0.2。其公式为:

$$L_{ce} = - \sum_{i=1}^N y_i \cdot \log(\hat{y}_i) \quad (9)$$

$$L_{Dice} = 1 - \frac{2N_{TP}}{2N_{TP} + N_{FP} + N_{FN}} \quad (10)$$

$$L_{loss} = 0.8L_{ce} + 0.2L_{Dice} \quad (11)$$

式中, N 表示类别总数, y_i 表示真实的标签, $\hat{y}_i$ 模型的预测概率。交叉熵损失函数衡量的是模型预测结果 $\hat{y}_i$ 与真实标签 y_i 之间的距离, 当预测概率 $\hat{y}_i$ 趋近于真实标签为 1 的位置时, 说明预测准确。

3.4 消融实验

以 U-Net 作为基准架构, 本文通过定量评价指标结果, 对比分析了 3 种模块 (MobileNetV2、Depthwise Convolution 以及 SGE) 在遥感图像道路检测任务中的性能表现, 以验证其在特征提取与模型优化方面的有效性与适用性。消融实验结果如表 2 所示。

不同模块组合在多个评估指标上均显著提升。以 U-Net 为基础框架, 引入 MobileNetV2 轻量化主干后,

表 2 消融实验对比结果

U-Net	MobileNetV2	DWConv	SGE	$F_1 / \%$	$mIOU / \%$	OA / %	R / %	P / %
✓				85.61	77.05	97.00	85.54	85.68
✓	✓			86.88	78.72	97.36	85.61	88.26
✓		✓		86.06	77.65	97.19	84.91	87.30
✓			✓	86.52	78.26	97.30	85.10	88.08
✓	✓	✓		87.20	79.16	97.45	85.57	89.02
✓	✓		✓	87.25	79.22	97.47	89.24	85.48
✓	✓	✓	✓	87.46	79.50	97.46	86.35	88.65

相较于原始模型, F_1 和 $mIOU$ 分别提升了 1.27% 和 1.67%。进一步引入深度可分离卷积后, 模型性能得到进一步增强, F_1 分数达到 87.25%, $mIOU$ 提升至 79.22%。这一结果验证了深度可分离卷积在提升模型特征表达能力方面的有效性。最后, 引入 SGE 注意力机制后, 组合模型的 F_1 达到 87.46%, $mIOU$ 提升至 79.50%, 表明 SGE 模块进一步增强了模型对关键信息的感知能力, 显著提升了分割精度。随着 MobileNetV2、DWConv 和 SGE 注意力机制的引入, OA 精度不断提升, 这表明模型在目标区域的检测精度上得到了显著改善, 同时有效减少了背景区域的错误分类。尽管 P 和 R 在某些情况下可能高于最终模型的值, 但这并不意味着模型性能更优。从实验结果可以看出, P 和 R 在极值时存在较大差异, 这导致 F_1 分数出现下降, 表明模型可能存在对某一方面的偏向, 且在某些场景下表现不够均衡。 F_1 分数的最佳优化通常发生在精度和召回率达到较好平衡时。引入 MobileNetV2、DWConv 和 SGE 注意力机制后, 模型在精度与召回率的平衡上得到了进一步优化, 从而使 F_1 分数达到了最佳表现。通过消融实验可知, 主干提取网络, 上采样更换卷积结构以及注意力机制每个模块的引入都能够在保证模型轻量化的前提下进一步提升分割性能, 验证了 3 个模块在遥感图像道路分割的有效性。

3.5 对比实验

为进一步验证本文所提出改进方法有效性, 除基于改进的 U-Net 网络进行训练外, 还引入了多种主流的语义分割网络模型, 包括 DeepLabv3+^[18]、MANet^[19]、LEDNet^[20] 以及 DFANet^[21] 进行对比实验, 各种对比图结果如表 3 所示。本文提出的模型在精度指标 F_1 、 $mIOU$ 、OA、 R 、 P 和计算效率指标参数量、 $FLOPs$ 两个方面均优于其他的语义分割模型, 体现出良好的综合性能。相比于经典的网络分割模型 DeepLabv3+, 其 F_1 和 $mIOU$ 指标分别提升了 1.79% 和 2.35%。与基础网络模型 U-Net 对比 F_1 和 $mIOU$ 分别提升了 1.85% 和 2.45%。对比模型中, MANet 虽利用注意力机制提升

了局部特征表达能力,但其网络结构复杂,参数量高达 35.86 M,且 $FLOPs$ 高达 21.51 G,导致部署成本极高。与 LEDNet 相比, F_1 和 $mIOU$ 指标也分别提升 1.55%和 2.04%。DFANet 采用了轻量化的网络结构设计,在推理效率上具备一定优势,但其在分割精度方面表现相对较弱。其分割精度 F_1 和 $mIOU$ 相比于本文方法下降 3.8%和 4.82%,说明其存在较高的误检率,并且在目标轮廓边缘的判定上不够精确,预测区域与真实标注之间存在较大偏差。同时在整个精度 OA 和精确率 P 下降进一步印证问题,说明模型在像素级分类中容易产生将非目标区域误判为目标,尤其在对小目标和边界区域的识别上准确性不足。与 SAG-LoANet 模型相比,所提方法在精度、召回率等性能指标上均表现出更优的检测性能,并在保证精度提升的同时保持了参数规模不增加。通过对比实验验证,本文所提出的改进方法在语义分割任务中展现出显著的性能优势。与现有主流道路提取网络模型相比,本文模型在有效降低计算复杂度与模型参数量的同时,进一步提升了分割精度与边界识别能力。该方法在保证模型轻量化的基础上,实现了更优的预测效果。

表 3 对比实验结果

模型	F_1 / %	$mIOU$ / %	OA / %	P / %	R / %	$Params$ / M	$FLOPs$ / G
U-Net	85.61	77.05	97.00	85.68	85.54	5.69	7.42
DeepLabv3+	85.67	77.15	97.10	86.78	84.63	5.81	6.60
MANet	85.19	76.56	97.08	87.41	83.25	35.86	21.51
LEDNet	85.91	77.46	97.14	86.89	84.99	2.27	1.52
DFANet	83.66	74.68	96.81	86.16	81.53	2.01	1.75
SAG-LoANet	—	66.6	—	84.50	—	1.4	5.5
Ours	87.46	79.50	97.46	88.65	86.35	3.03	0.91

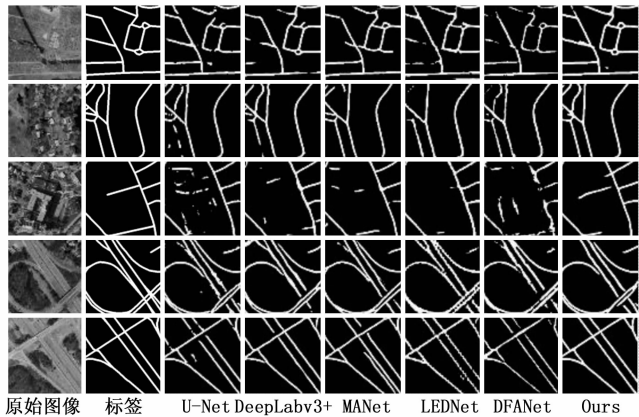


图 5 对比试验可视化结果

为进一步直观地展示本文改进模型在道路检测任务的有效性,在数据集上选取具有代表性的样本进行可视化分析,如图 5 所示,样本涵盖了城市,乡镇等多种典

型的应用场景,部分图像中存在阴影遮挡,对比度复杂等因素,显著增加了道路识别难度。综合分析各组可视化结果可以看出,本文所提出的改进模型在细节还原、边缘保持以及抗噪能力方面均明显优于现有对比模型。U-Net 原始模型对道路提取存在很多噪声和伪边界,细节结构不够清晰。DeepLabv3+对主干道路提取较为完整,但存在明显断裂现象,尤其在道路交叉口或遮挡区域无法形成连续结构。MANet 具有一定的细节恢复能力,边界更加平滑,但部分区域存在过度预测的问题,对背景干扰仍不够鲁棒。LEDNet 轻量化设计在推理速度上有优势,但从视觉效果来看,存在明显信息缺失,细节保留能力较差,边缘较为粗糙。DFANet 在复杂的场景下表现不稳定,更容易产生错检和漏检。本文改进模型相比其他模型,本文提出的方法在多个场景中均能准确还原道路结构,边缘清晰且连贯,表现出更强的语义理解能力与空间一致性。尤其在遮挡、交叉复杂区域依然保持较高的预测质量,体现出良好的鲁棒性和泛化能力。

4 结束语

本研究针对当前遥感影像道路检测任务中常见的提取不完整、边界模糊以及模型计算开销较大的问题,设计并提出了一种基于 U-Net 框架的高效改进模型。该模型在保持分割精度的同时,降低了计算资源的消耗。首先,采用轻量化的 MobileNetV2 作为编码器主干网络,通过其高效的瓶颈结构捕捉多尺度特征,增强对细长、非规则结构的感知能力;其次,引入深度可分离卷积,在减少参数量的同时,有效地提升了模型的表达能力,为后续提供了更精细的特征表示;最后,融合 SGE 注意力机制,通过分组方式对特征图进行空间加权,突出关键区域的特征提取,抑制背景噪声,从而在保证高效计算的同时实现精细空间注意力建模,显著提升了模型在复杂背景下的鲁棒性与分割精度。综合上述设计,本文所提出的模型兼具结构轻量化和特征提取精确性,在遥感图像道路检测任务中更能表现出更优的分割效果与更强的泛化能力。后续实验结果从精度指标与可视化方面都对本文方法的有效性进行了验证。未来,将进一步拓展和优化遥感影像道路数据集,以覆盖更多样化的地理场景。同时还将探索更高效的网络架构,不断优化模型的表达能力与计算效率,推动其在实际应用场景中的部署。

参考文献:

[1] MO S, SHI Y, YUAN Q, et al. A survey of deep learning road extraction algorithms using high-resolution remote sensing images [J]. Sensors, 2024, 24 (5): 1708.
[2] YUAN X, SHI J, GU L. A review of deep learning

- methods for semantic segmentation of remote sensing imagery [J]. *Expert Systems with Applications*, 2021, 169: 114417.
- [3] 闵 蕾, 高 昆, 李 维, 等. 光学遥感图像分割技术综述 [J]. *航天返回与遥感*, 2020, 41 (6): 1-13.
- [4] 姜凯悦, 赵竹新, 马 璐, 等. 基于深度学习的遥感影像变化检测方法进展综述 [J]. *无线电工程*, 2025, 55 (2): 343-356.
- [5] 蒋星详, 肖 莉. 一种多特征融合的高分辨率遥感影像道路中心线提取算法 [J]. *测绘地理信息*, 2019, 44 (4): 98-101.
- [6] 刘 霞, 郭亚男. 基于随机森林的遥感影像变化检测 [J]. *测绘通报*, 2020 (5): 16-20.
- [7] LI A, JIAO L, ZHU H. Multitask semantic boundary awareness network for remote sensing image segmentation [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 60: 1-14.
- [8] LI J, CAI Y, LI Q, et al. A review of remote sensing image segmentation by deep learning methods [J]. *International Journal of Digital Earth*, 2024, 17 (1): 2328827.
- [9] 胡功明, 杨春成, 徐 立, 等. 改进 U-Net 的遥感图像语义分割方法 [J]. *测绘学报*, 2023, 52 (6): 980-989.
- [10] LIU D, DU J, LI C, et al. Multi-unit stacked architecture: an urban scene segmentation network based on UNet and ShuffleNetv2 [J]. *Applied Soft Computing*, 2024, 165: 112065.
- [11] LIU S, MA Y, ZHANG X, et al. Rotated multi-scale interaction network for referring remote sensing image segmentation [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024: 1-8.
- [12] GAO L, SONG W, DAI J, et al. Road extraction from high-resolution remote sensing imagery using refined deep residual convolutional neural network [J]. *Remote Sensing*, 2019, 11 (5): 552.
- [13] 房 梁, 谢 刚, 续欣莹, 等. 融合动态权重与语义筛
- (上接第 33 页)
- [18] SULICA L. Laryngoscopy, stroboscopy and other tools for the evaluation of voice disorders [J]. *Otolaryngologic Clinics of North America*, 2013, 46 (1): 21-30.
- [19] BEYNON R A, LANG S, SCHIMANSKY S, et al. Tobacco smoking and alcohol drinking at diagnosis of head and neck cancer and all-cause mortality: results from head and neck 5000, a prospective observational cohort of people with head and neck cancer [J]. *International Journal of Cancer*, 2018: 1-14.
- [20] IGISSINOV N, ZATOSKIKH V, MOORE M A, et al. 选的双分支遥感图像语义分割 [J]. *计算机工程与应用*, 2025, 61 (20): 295-305.
- [14] 王克明, 公维佳, 王海明, 等. 改进 Unet 网络的汽油管道泄漏高光谱图像检测 [J]. *光谱学与光谱分析*, 2025, 45 (5): 1476-1484.
- [15] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: Inverted residuals and linear bottlenecks [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 4510-4520.
- [16] ZHU B, HU S, SONG H, et al. Group-wise semantic-enhanced interaction network for remote sensing spatio-temporal fusion [C] // *2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2025: 1-5.
- [17] RONNEBERGER O, FISCHER R P, BROX T. U-net: Convolutional networks for biomedical image segmentation [C] // *Proceedings of the Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, Springer*, 2015: 234-241.
- [18] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C] // *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018: 801-818.
- [19] LI R, ZHENG S, ZHANG C, et al. Multiattention network for semantic segmentation of fine-resolution remote sensing images [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 60: 1-13.
- [20] WANG Y, ZHOU Q, LIU J, et al. Lednet: a lightweight encoder-decoder network for real-time semantic segmentation [C] // *Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2019: 1860-1864.
- [21] LI H, XIONG P, FAN H, et al. Dfanet: deep feature aggregation for real-time semantic segmentation [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 9522-9531.
- Laryngeal cancer in Kazakhstan-ethnic and gender variation with age over [J]. *Asian Pac. J. Cancer Prev.*, 2013, 14 (11): 7033-8.
- [21] MAJSZYK D, BRUZGIELEWICZ A, OSUCHWOJCIEWICZ E, et al. Gender-related incidence, risk factors exposure and survival rates of laryngeal cancers-the 10-years analysis of trends from one institution [J]. *Otolaryngol Pol.*, 2019, 73 (3): 6-10.
- [22] KRAFT M, FOSTIROPOULOS K, GÜRTLER N, et al. Value of narrow band imaging in the early diagnosis of laryngeal cancer [J]. *Head & neck*, 2016, 38 (1): 15-20.