

基于改进U-Net的无人机可见光影像植被精细分割研究

刘 洋, 王英伟, 费美玲, 廖亚迪, 马浩鸣

东北林业大学林学院, 黑龙江 哈尔滨

收稿日期: 2025年12月20日; 录用日期: 2026年1月17日; 发布日期: 2026年1月27日

摘 要

利用高分辨率无人机可见光影像实现植被类型的自动、精细分割, 对于生态环境监测与量化研究具有重要意义。然而, 可见光影像光谱信息有限, 加之植被类型间光谱特征相似(特别是乔木、灌木与草地), 导致精细分割精度面临挑战。本文提出一种改进的U-Net语义分割模型, 旨在提升植被精细分割的精度与模型泛化能力。主要改进包括: 1) 采用在ImageNet上预训练的ResNet34网络替换原始编码器, 以增强多层次特征提取能力; 2) 在跳跃连接中引入空间与通道挤压激励注意力模块, 实现特征的自适应融合; 3) 设计一种结合交叉熵损失、Dice损失与Focal损失的复合损失函数, 以优化训练过程并聚焦难分样本。本研究在自主构建的无人机影像数据集上进行实验验证。实验采用五折交叉验证与OneCycleLR学习率调度策略。结果表明, 本文模型在B测试集上的总体精度达到90.2%, 较原始U-Net提升6.3个百分点, 其中对灌木和草地的分割精度提升尤为显著。本研究为基于低成本可见光数据的植被精细识别提供了一种有效的深度学习解决方案。

关键词

无人机可见光影像, 植被分类, 深度学习, 注意力机制

Research on Vegetation Segmentation in UAV Imagery Based on Improved U-Net

Yang Liu, Yingwei Wang, Meiling Fei, Yadi Liao, Haoming Ma

College of Forestry, Northeast Forestry University, Harbin Heilongjiang

Received: December 20, 2025; accepted: January 17, 2026; published: January 27, 2026

Abstract

Realizing the automatic and fine-grained segmentation of vegetation types using high-resolution

文章引用: 刘洋, 王英伟, 费美玲, 廖亚迪, 马浩鸣. 基于改进 U-Net 的无人机可见光影像植被精细分割研究[J]. 计算机科学与应用, 2026, 16(1): 364-373. DOI: 10.12677/csa.2026.161030

UAV visible imagery is of great significance for ecological environment monitoring and quantitative research. However, the limited spectral information of visible imagery, coupled with the similar spectral characteristics among different vegetation types (especially arbor, shrub, and grassland), poses challenges to the accuracy of segmentation. This study proposes an improved U-Net semantic segmentation model aiming to enhance the segmentation accuracy and model generalization ability. The main improvements include: 1) replacing the original encoder with the ResNet34 network pre-trained on ImageNet to strengthen multi-level feature extraction capability; 2) introducing spatial and channel squeeze-and-excitation (SE) attention modules into skip connections to achieve adaptive feature fusion; 3) designing a hybrid loss function combining cross-entropy loss, Dice loss, and Focal loss to optimize the training process and focus on hard-to-segment samples. Experimental verification was conducted on a self-constructed UAV imagery dataset, adopting five-fold cross-validation and the OneCycleLR learning rate scheduling strategy. The results show that the overall accuracy (OA) of the proposed model on Test Set B reaches 90.2%, improving by 6.3 percentage points compared with the original U-Net. Particularly, the segmentation accuracy of shrubs and grassland is significantly enhanced. This study provides an effective deep learning solution for fine-grained vegetation identification based on low-cost visible light data.

Keywords

UAV Imagery, Vegetation Classification, Deep Learning, Attention Mechanism

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

植被是陆地生态系统进行能量转换与物质循环的核心载体,其类型、空间分布及生物量的精确信息,是生态学研究、生物多样性保护、林业碳汇评估及城市绿地规划不可或缺的数据基础。传统的植被调查方法主要依赖野外人工勘测,存在效率低下、成本高昂且难以获取大面积连续空间信息的局限性。近年来,无人机遥感技术凭借其灵活机动、高空间分辨率(可达厘米级)和低成本的数据获取优势,为快速、动态监测植被资源开辟了全新途径。特别是消费级无人机搭载的可见光相机所获取的 RGB 影像,数据来源最为广泛,但其光谱信息相对匮乏,为植被类型的精细辨识带来了巨大挑战[1]。

在遥感影像分析领域,植被信息提取方法主要经历了从像元光谱分析、面向对象分类到深度学习语义分割的发展历程。早期研究多依赖于目视解译或基于像元光谱特征的机器学习方法(如支持向量机、随机森林),这些方法需要大量人工干预进行特征设计与筛选,且在复杂场景下泛化能力有限[2]。随着深度学习技术的突破,全卷积神经网络因其端到端的像素级预测能力,在遥感影像语义分割任务中取得了革命性进展。其中,U-Net 网络凭借其独特的编码器-解码器架构与跳跃连接,能有效融合浅层空间细节与深层语义信息,在医学影像与遥感地物提取中均表现出色,成为图像分割领域的基准模型之一[3]。

然而,将标准 U-Net 直接应用于高分辨率无人机可见光影像的植被精细分割,仍面临若干瓶颈。首先,其编码器通常由简单的卷积堆叠构成,特征提取能力有限,难以充分学习植被复杂多样的形态与纹理特征。其次,标准的跳跃连接对编码器所有特征进行简单拼接,未能区分不同层次特征对最终分割贡献的差异性,可能导致细节模糊或噪声引入。最后,植被类别(尤其是“灌木”与“草地”)常存在样本数量不均衡与边界模糊的问题,单一的交叉熵损失函数难以引导模型专注于难分样本的学习。

针对上述问题,本文旨在构建一个鲁棒、高效的改进型 U-Net 模型,以提升对无人机可见光影像中

乔木、灌木、草地等植被类型的精细分割精度。本文的主要贡献如下：(1) 引入在大型自然图像数据集上预训练的 ResNet34 作为编码器骨干网络，利用其强大的深度残差结构提取更具判别力的多层次特征；(2) 在跳跃连接路径中集成空间与通道挤压激励注意力模块，使模型能够自适应地校准与融合编码器特征，增强对关键信息的捕捉能力；(3) 设计一种加权复合损失函数，协同优化像素级分类精度、区域重叠度以及对难例的关注度。本文通过在自建数据集上的系统性实验，验证了改进模型的有效性与泛化能力。

2. 研究区与数据获取

2.1. 研究区概况

本文所选的研究区域为东北林业大学校区，位于黑龙江省哈尔滨市香坊区，如下图 1 所示。该区域地物特征多样，并且数据采集较为方便，易于获取。其中部分道路和操场与植被颜色相似，这使得后续的分类实验难度提高，且植被类型丰富，林地与草地覆盖较多，为验证植被精细分割算法提供了理想的复杂场景。

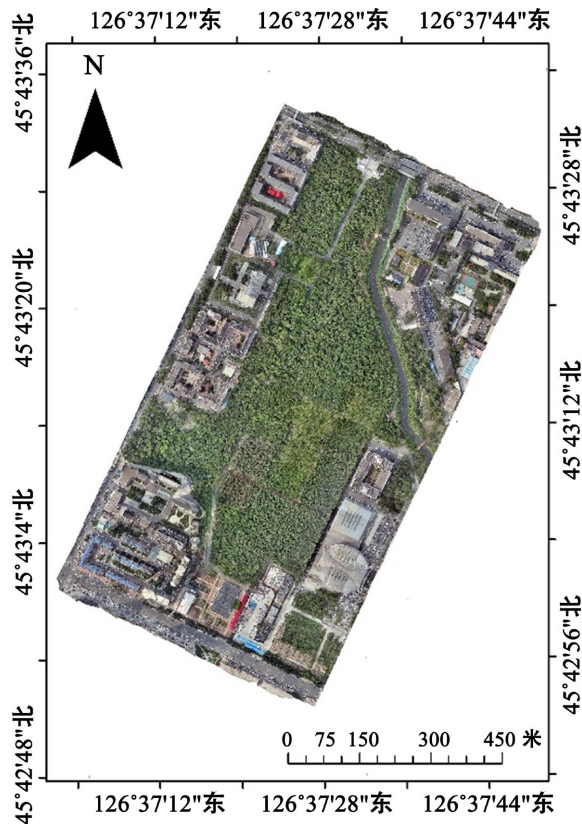


Figure 1. Schematic diagram of the geographical location of the study area
图 1. 研究区地理位置示意图

2.2. 无人机遥感影像的采集与处理

数据获取于 2025 年 6 月，利用大疆 DJI Matrice 4E 无人机获取研究区无人机影像，该款无人机搭载三轴机械云台，4800 万像素的变焦相机和 2000 万的广角相机。飞行高度设定为 100 米，航向与旁向重叠率均设置为 80%，以确保生成高精度正射影像。通过专业软件对采集的原始影像进行拼接、几何校正与辐射校正，最终生成空间分辨率为 0.02 m 的研究区正射影像。

本文选取研究区地物类型丰富且空间上完全独立的 A、B 两个区域来制作植被分类样本数据集，两个样本区在研究区的分布如图 2 所示。



Figure 2. UAV image data of the sample area
图 2. 样本区无人机影像数据

2.3. 植被多分类标签数据的制作

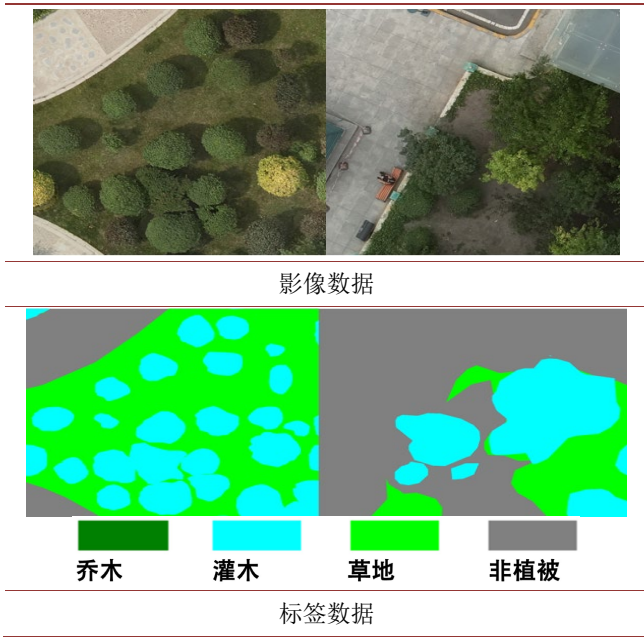
在进入模型训练之前，需要对实验所用数据集进行制作，达到模型训练的输入格式要求。数据集的质量也会间接影响最终的分类精度，所以，数据集制作也是本文实验关键的一部分。制作主要分为以下几个步骤：

(1) 使用 ArcGis 10.8 软件对样本区域进行高精度人工目视解译，分为乔木、灌木和草地和非植被 4 个类别，生成像素级分类标签图。本研究对样本区的植被划分为 3 种类型：乔木、灌木和草地。3 种类型植被的影像解译特征如表 1 所示。部分多分类数据的对应影像与对应标签如表 2 所示。

Table 1. Image interpretation features of three vegetation types in the sample area
表 1. 样本区 3 种植被类型对应影像解译特征

类别	解译特征	影像
乔木	颜色偏深绿色，分布广泛	
灌木	形状规整，天然的灌木纹理粗糙	
草地	色调偏浅绿色，边界有具体的形状	

Table 2. Display of corresponding images and labels for partial multi-class classification data
表 2. 部分多分类数据的对应影像与对应标签展示



(2) 将 A 和 B 影像及其对应的标签图采用滑动窗口裁剪法裁剪成统一尺寸的样本切片。

为科学评估模型性能，A 区域作为训练与验证集，用于模型参数的训练与调整；B 区域作为严格的独立测试集，仅用于最终评估模型的泛化性能，确保评价结果的客观性。在模型训练阶段，对 A 区域的影像样本施加了随机水平翻转、垂直翻转及小角度旋转等数据增强操作[4]。

3. 研究方法

U-Net 是一种经典的编码器 - 解码器架构语义分割网络，因其高效、精准的像素级预测能力而在众多密集预测任务中备受青睐。该模型主要由对称的编码器、解码器以及连接二者的跳跃连接三部分构成。编码路径能有效学习从光谱、纹理到空间上下文的多层次遥感特征，而跳跃连接确保了地块边缘、小型植被斑块等细节空间信息在解码过程中的高保真度恢复。该模型在遥感影像语义分割领域已拥有大量成功应用先例，证明了其处理复杂自然场景与人工地物的强大能力[5]。因此，选择 U-Net 作为基准模型，不仅基于其广泛的技术验证，更因其专精于像素级预测的特性与本研究目标高度契合。因此本文选择 U-Net 作为基准模型[6]。

3.1. 改进的 U-Net 模型架构

本文提出的模型以 U-Net 经典架构为蓝本，在编码器、特征融合及损失函数三个层面进行系统改进，改进后的 U-Net 模型结构如图 3 所示。

3.1.1. 强化编码器：ResNet34 骨干网络

摒弃原始 U-Net 的浅层编码器，采用在 ImageNet 大规模数据集上预训练的 ResNet34 网络作为编码器。ResNet34 通过引入残差连接，允许网络有效加深至 34 层，从而能够提取从低层边缘、纹理到高层语义信息的丰富特征层次[7]。预训练权重使模型继承了强大的通用视觉表征能力，通过迁移学习可大幅提升在特定任务上的收敛速度与性能上限。

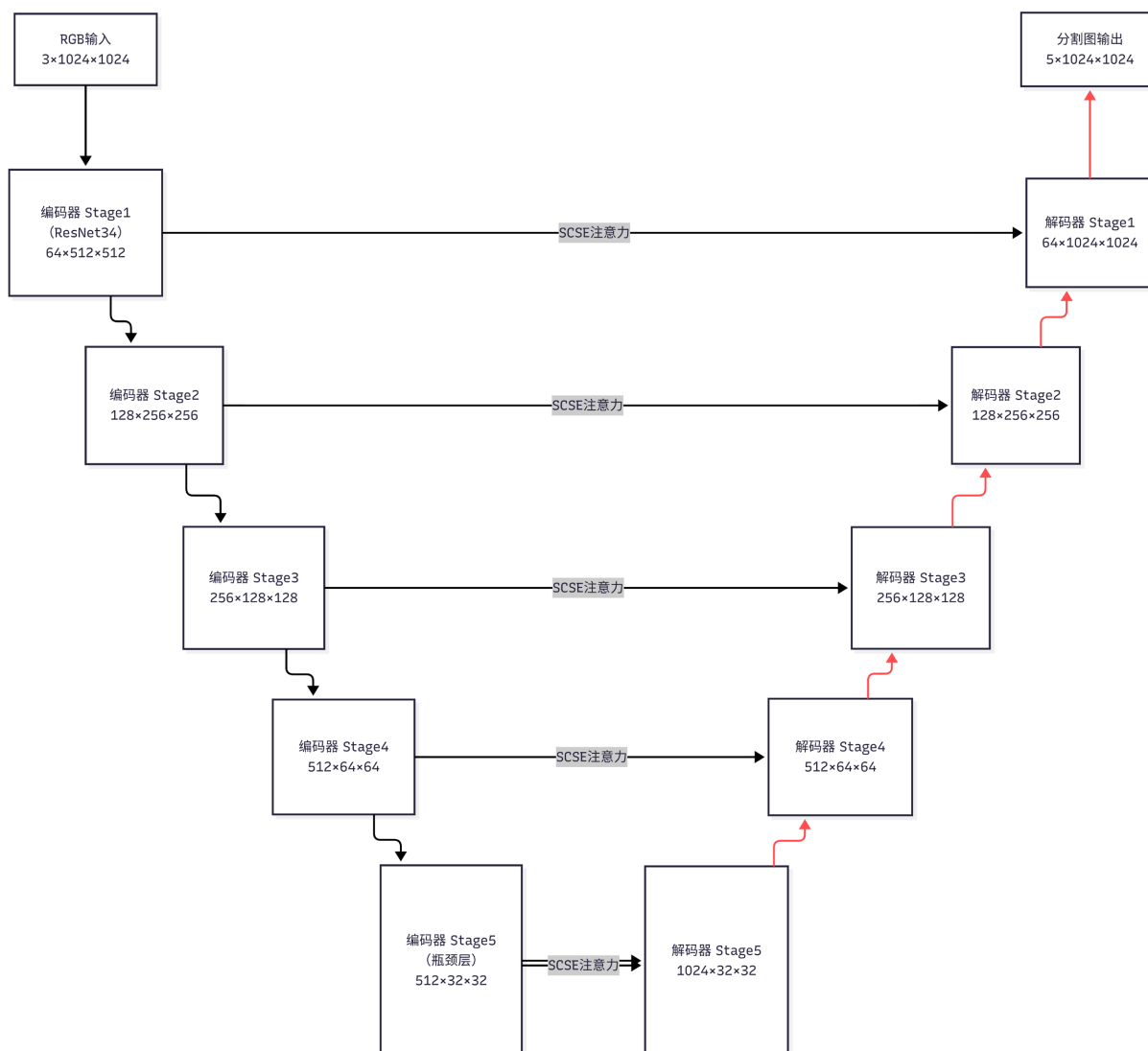


Figure 3. Structure of the improved U-Net model

图 3. 改进后 U-net 模型结构

3.1.2. 优化特征融合：SCSE 注意力跳跃连接

标准跳跃连接直接拼接编码器与解码器的特征，忽略了不同特征通道和空间位置对当前分割任务的重要性差异。为此，我们在每条跳跃连接路径上集成了空间与通道挤压激励注意力模块。

该模块并行工作：通道注意力分支通过分析全局上下文，学习各个特征通道的权重，增强信息量丰富的通道；空间注意力分支则聚焦于学习特征图在空间维度上的重要性分布，突出关键区域。两个分支的输出相乘后对原特征图进行加权，从而实现自适应特征校准。经 SCSE 模块处理后的特征再与解码器特征拼接，使得解码过程能够更精准地利用编码器传递的细节与语义信息[8]。

3.1.3. 输出与复合损失函数

解码器末端通过 1×1 卷积将通道数映射为类别数(5)，并经由 Softmax 函数输出每个像素的类别概率。

为解决植被分割中常见的类别不平衡(如“灌木”样本较少)和边界模糊问题，本文设计了复合损失函

数 L_{total} :

$$L_{total} = 0.5 * L_{CE} + 0.3 * L_{Dice} + 0.2 * L_{Focal}$$

其中, L_{CE} (交叉熵损失)确保像素级分类准确性; L_{Dice} (Dice 损失)直接优化预测区域与真实区域的重叠度, 对小目标类别友好; L_{Focal} (Focal 损失)通过动态缩放因子, 降低易分类像素的权重, 使训练聚焦于难以区分的像素(如灌 - 草边界)。三者加权协同, 共同引导模型优化方向。

3.2. 实验设置

所有实验在 PyTorch 框架下进行, 使用 NVIDIA GeForce RTX 2080 Ti GPU 加速。输入影像统一裁剪为 1024×1024 像素。优化器选用 AdamW, 初始学习率为 1×10^{-4} , 并配合 OneCycleLR 调度策略动态调整。批处理大小设为 4。

模型训练采用五折交叉验证法在 A 数据集上进行, 即每次将 A 数据分为 5 份, 轮流以其中 4 份训练、1 份验证, 最终取 5 次验证结果的平均值作为模型性能估计, 以此确保评估的稳健性。选择在交叉验证中平均交并比最高的模型权重, 在完全未参与训练的 B 独立测试集上进行最终性能评估。评价指标包括: 总体精度、平均交并比、召回率、F1 分数和准确率。

4. 实验结果与分析

4.1. 混合损失函数权重合理性量化分析

针对混合损失函数(交叉熵损失 + Dice 损失 + Focal 损失)权重设置的合理性, 采用控制变量法, 固定交叉熵(CE)损失权重为 1, 调整 Dice 损失权重(α : 0.5~2.0)与 Focal 损失权重(β : 0~1.0), 以 MIoU 为核心评价指标, 分析权重对模型性能的影响[9], 结果如表 3 所示。

Table 3. Settings and results of the weight experiment
表 3. 权重实验设置与结果

实验编号	混合损失权重(CE:Dice:Focal)	平均交并比(MIoU)	总体精度(OA)	性能评价
L0	1:0:0 (仅 CE 损失)	73.0	84.5	性能最差, 无法应对样本不平衡
L1	1:0.5:0 (CE + 低权重 Dice)	73.5	85.0	性能提升有限, Dice 权重不足
L2	1:1.0:0 (CE + 等权重 Dice)	73.9	86.0	性能显著提升, 达到单一组合最优
L3	1:1.0:0.5 (本研究设置)	74.2	86.3	性能最优, 权重配比合理
L4	1:1.0:1.0 (高权重 Focal)	73.8	85.8	性能下降, Focal 权重过高
L5	1:2.0:0.5 (高权重 Dice)	73.6	85.5	性能下降, Dice 权重过高

4.2. 模型改进组件消融实验

为验证改进 U-Net 模型中各核心组件(ResNet34 编码器、SCSE 注意力模块、混合损失函数)的独立贡献及组合有效性, 本研究设计了 5 组对比实验, 以原始 U-Net 为基准, 逐步添加改进组件。实验保持训练策略(学习率、迭代次数、优化器等)完全一致, 基于自主构建的无人机植被影像数据集, 采用五折交叉验证的平均指标作为评价结果[10]。

如表 4 所示, ResNet34 编码器是核心增益来源, 单一组件带来 OA 提升 2.9 个百分点, 远超 SCSE 与混合损失, 说明预训练 ResNet34 是精度提升的基础。混合损失函数是重要优化补充, 单一组件 OA 提升 2.1 个百分点, 且与 ResNet34 组合后能额外带来 1.7 个百分点提升。SCSE 模块是精度突破的关键, 单一组件增益虽最低, 但在 ResNet34 + 混合损失的基础上, 仍能带来 1.7 个百分点 OA 提升。

Table 4. Results of the ablation study
表 4. 消融实验结果

模型配置	总体精度(OA)	平均交并比(MIoU)	召回率(Recall)	F1 分数	准确率(Accuracy)
U-Net	83.9	72.4	80.2	84.9	84.1
U-Net + SCSE	85.5	73.5	81.5	85.8	84.8
U-Net + 混合损失	86.0	73.9	81.9	86.0	85.0
U-Net (ResNet34)	86.8	74.8	82.0	86.2	85.3
ResNet34-U-Net + 混合损失	88.5	76.6	83.5	87.6	86.6
改进模型	90.2	78.2	84.6	88.3	87.2

由表 5 可知，改进模型的参数量与计算量分别增加 36.5%和 34.9%，主要源于 ResNet34 编码器的引入和 SCSE 注意力模块的添加。推理帧率从 42.5 FPS 下降至 35.8FPS，降幅为 15.8%，但仍保持在 30 FPS 以上，满足无人机影像实时分割的基本需求。

改进模型在精度提升 7.5% (OA)的同时，计算负载增加约 35%，推理效率下降约 16%，整体处于可接受范围内。对于无人机遥感植被分割的实际应用场景，精度提升带来的生态监测准确性改善，远大于计算复杂度增加带来的存储与推理开销。改进模型的参数量为 42.3M，可轻松部署于便携式 GPU 设备，满足野外无人机作业的实时数据处理需求，同时，相比其他高精度模型，本模型在精度与效率之间实现了更优平衡，具备更强的实际应用价值。

Table 5. Comparison of model accuracy and efficiency metrics
表 5. 模型精度与效率指标对比

模型	总体精度(OA, %)	平均交并比(mIoU, %)	参数量(M)	计算量(GFLOPs)	推理帧率(FPS)
U-Net	83.9	76.2	31.0	15.2	42.5
改进模型	90.2	83.7	42.3	20.5	35.8
相对变化	+7.5%	+9.8%	+36.5%	+34.9%	-15.8%

4.3. 在自建数据集上的性能表现

为验证改进模型的有效性，在 B 独立测试集上对比了本文模型与两个基线模型：原始 U-Net、仅替换 ResNet34 编码器的 U-Net。定量结果如表 6 所示。

本文模型取得了全面最优的性能，总体精度达到 90.2%，相比原始 U-Net (83.9%)和 ResNet34-U-Net (86.3%)分别有显著提升，准确率为 87.2%，SCSE 注意力模块和复合损失函数对于改善光谱、纹理相似类别的区分能力起到了关键作用。

Table 6. Performance comparison of different models on Test Set B (%)
表 6. 不同模型在 B 测试集上的性能对比(%)

模型	总体精度(OA)	平均交并比(MIoU)	召回率(Recall)	F1 分数	准确率(Accuracy)
U-Net	83.9	72.4	80.2	84.9	84.1
U-Net (ResNet34)	86.3	74.1	82.1	86.2	85.2
改进模型	90.2	78.2	84.6	88.3	87.2

4.4. 可视化结果对比

图 4 展示了 B 测试集中一个典型区域的分割结果可视化对比。原始 U-Net 的结果在灌草交错区域存在大量混淆，且乔木冠层边界较为粗糙。而本文模型的分割结果与真实标签高度吻合：能够清晰地地区分出单棵乔木的完整冠层轮廓，准确地将低矮、丛生的灌木与均一的草地分离开来，并在植被与道路、建筑的边界处保持了更高的分割锐度。

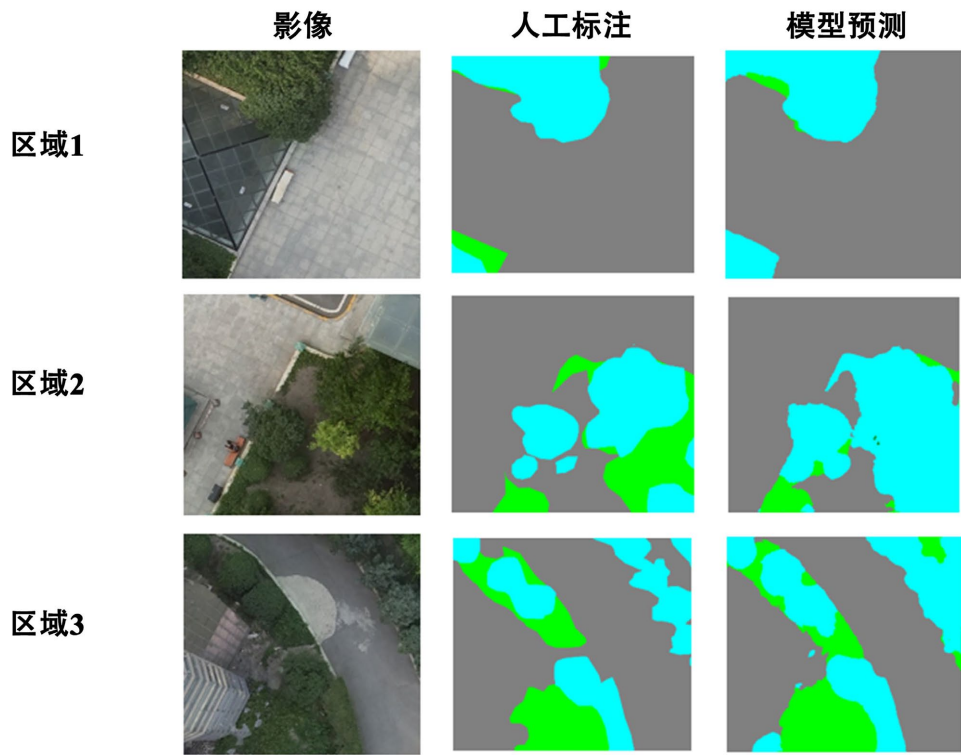


Figure 4. Comparison of segmentation results in typical areas
图 4. 典型区域的分割结果对比

5. 讨论

基于前述实验结果，本节将深入讨论本文所提出改进措施的有效性机制、模型的综合性能及其在植被遥感解译中的应用价值与潜力。

首先，引入预训练的 ResNet34 编码器带来了最显著的性能增益。这源于其深度残差结构能够从高分辨率无人机影像中提取更丰富、更具判别力的多层次特征。浅层网络捕获精细的纹理和边缘(对区分灌草边界至关重要)，而深层网络则理解复杂的空间上下文和语义信息(如识别成片乔木冠层)，这种强大的特征提取能力是后续所有改进的基石。

其次，集成于跳跃连接的 SCSE 注意力模块通过“软筛选”机制优化了特征融合。在复杂的植被场景中，编码器传递的特征包含大量信息，但也混杂着阴影、土壤背景等干扰。SCSE 模块通过通道注意力强调对植被敏感的特征通道，并通过空间注意力聚焦于疑似植被的区域，从而抑制了无关信息的干扰，实现了自适应、有重点的特征融合。这直接解释了为何在加入该模块后，模型对形态不规则、光谱易混淆的灌木类别的分割精度得到了特别明显的改善。

最后，所采用的复合损失函数从优化目标层面引导模型聚焦于难点。交叉熵损失确保全局分类正确；

Dice 损失直接优化预测区域与真实区域的重叠,提升整体分割完整性;而 Focal 损失则自动降低易分样本(如大片均质草地)的权重,迫使模型集中学习难分样本(如阴影中的植被、灌草交错边缘)。三者协同工作,确保了模型在提升全局指标的同时,也能精细化处理分类薄弱的环节,使各类别精度得到均衡提升。

实验结果表明,本文模型利用数据集 A 进行训练和验证,在自建的数据集 B 上测试达到了最优性能,这充分证明了改进模型所学习的特征表示具有普适性,而非对训练集特定分布的过拟合。这种泛化能力对于资源环境领域的实际应用至关重要,意味着模型在经过特定区域数据微调后,有望快速部署至其他地理区域进行植被监测,极大地提升了方法的实用价值。与原始 U-Net 及仅替换骨干网络的变体相比,本文模型在平均交并比(MIoU)和总体精度(OA)上均取得全面领先,说明其改进是系统且有效的,显著提升了模型在复杂自然场景下的综合解译稳健性。

参考文献

- [1] 郑远攀,李广阳,李晔.深度学习在图像识别中的应用研究综述[J].计算机工程与应用,2019,55(12):20-36.
- [2] 赵什陆,张强.深度学习多模态图像语义分割前沿进展[J].中国图象图形学报,2023,28(11):3320-3341.
- [3] 杨柳,陈延辉,岳德鹏,等.无人机遥感影像的城市绿地信息提取[J].测绘科学,2017,42(2):59-64.
- [4] 杨惠雯.基于深度特征学习的高分辨率遥感影像分类[D]:[硕士学位论文].北京:中国科学院大学(中国科学院遥感与数字地球研究所),2018.
- [5] 李倩楠,张杜娟,潘耀忠,等.MPSPNet 和 UNet 网络下山东省高分辨耕地遥感提取[J].遥感学报,2023,27(2):471-491.
- [6] 姚方方,骆剑承,沈占锋,等.高分辨率影像城市植被自动提取算法[J].地球信息科学学报,2016,18(2):248-254.
- [7] 徐知宇.基于深度学习的城市绿地遥感分类及应用研究[D]:[硕士学位论文].北京:中国科学院大学(中国科学院空天信息创新研究院),2021.
- [8] 高永刚,林悦欢,温小乐,等.基于无人机影像的可见光波段植被信息识别[J].农业工程学报,2020,36(3):178-189.
- [9] 赖丽琦.基于语义分割模型的无人机遥感影像植被分类研究[D]:[硕士学位论文].杭州:浙江农林大学,2020.
- [10] 汪小钦,王苗苗,王绍强,等.基于可见光波段无人机遥感的植被信息提取[J].农业工程学报,2015,31(5):152-157.