Tea-UNet 茶园无人机影像语义分割算法研究

熊盛佳[1] 叶智国[1] 吴章云[1] 吴传宇[1*]
XIONG Shengjia YE Zhiguo WU Zhangyun WU Chuanyu

摘 要 针对传统方法在复杂地形和相似作物背景下难以精确识别茶园区域的问题，文章提出了一种基于增强编码器与多级特征融合的方法，旨在开发一种名为 Tea-UNet 的新型语义分割算法，专门用于从无人机获取的高分辨率影像中高效、准确地提取茶园区域。通过增强编码器，采用 RepVit 架构，结合深度可分离卷积和结构重参数化技术，在保持计算效率的同时增强了全局特征建模能力。此外，引入了基于 Transformer 的多级多尺度特征融合模块（MFT），通过多头交叉注意力机制实现复杂的特征融合，有效缩小了浅层与深层特征间的语义差距。解码器部分则设计了多级注意力引导的特征融合模块（MAF），通过连续的注意力机制更新，精炼特征表示，减少不相关信息干扰。实验结果表明，Tea-UNet 模型在茶园区域提取任务上取得了显著的效果，不仅提高了分割精度，还增强了对细长或形状不规则目标的识别能力，该研究为茶园管理和可持续发展提供了强有力的技术支持。

关键词 语义分割；无人机影像；增强编码器；多级特征融合；茶园提取

doi：10.3969/j.issn.1672-9528.2025.05.004

0 引言

茶树作为一种重要的常绿木本经济作物，其叶子加工而成的饮品与咖啡、可可并列为全球三大非酒精饮料之一，对许多发展中国家的经济繁荣贡献显著[1]。中国作为最大的茶叶生产国，占全球产量约 38%，全国有超过 20 个省份从事茶叶种植[2]。随着茶园面积扩大促进了地方经济，但也带来生态破坏等挑战。因此，亟须开发高效获取茶园空间分布的方法，维持经济效益的同时减轻环境影响，支持可持续发展。

当前无人机搭载普通相机的遥感图像因其独特的视角和高分辨率，为地物特征的提取提供了丰富信息，但同时也带来了分割上的挑战[3]。与自然图像不同，无人机影像是从空中垂直拍摄的，覆盖了广阔的区域且物体大小差异很大，这使得在这些图像中分割较小的物体变得困难。为了解决这些问题，许多研究者采用特征金字塔网络（FPN）[4]来整合来自不同层次的图像特征，从而有效识别和分割遥感图像中的小物体。为了进一步增强分割能力，大多数学者利用空间和通道注意力机制来调整模型对各种特征的关注点[5]。Li 等人[6]提出了 SCAttNet，连接两种机制，从两个维度顺序调整模型对特征图像的关注点，从而增强了分割性能。然而，这些优化方法并不适用于构建作物提取模型。在地形破碎的丘陵地区，物体之间存在大小差异，使得传统方法难以准确识别小物体。此外，在茶种植区，周边其他类别的作物会影响卷积神经网络的特征提取，导致语义混淆，影响茶园区域提取的准确性。因此，提高模型区分相似特征的能力至关重要。

考虑到茶园区域提取的特点，提出一种基于增强编码器与多级特征融合的语义分割算法，专门用于识别和提取中国南方丘陵地区的茶园区域。Tea-UNet 模型基于 U-Net 架构[7]，分别针对编码器、跳跃链接、解码器进行针对性优化改进，从而提升分割的精度，精确地提取茶园区域。

1 基于增强编码器与多级特征融合的语义分割算法

1.1 Tea-UNet 网络结构

本研究提出的模型结构如图 1 所示，Tea-UNet 是一个 U 型网络，主要由编码器、特征融合模块、解码器构成。在编码器中，包含 4 个 Stage，每个 Stage 由一个 RepVitSEBlock 和一个 RepVitBlock 组成。基于 RepVitBlock 的编码器设计，可以学习到更多的特征，获得更高的分割精度。特征融合模块利用多尺度特征嵌入、多头交叉注意力和 MLP，增强了全局上下文理解与长距离依赖捕捉，同时保留了空间细节。实现了复杂的特征融合，提高了对数据变化的鲁棒性，并在处理细长物体或形状不规则区域等复杂结构目标时，提供了更精准的分割结果和更丰富的语义特征表

1. 福建农林大学机电工程学院 福建福州 350002
[基金项目] 福建农林大学科技创新项目“咖啡烘焙温度智能控制研究”（KFB23162A）

示。在解码器中，MAF 模块通过多级特征融合与连续的注意力机制更新，逐步精炼特征表示，减少不相关信息干扰，整合跨尺度信息，生成聚焦于关键信息的高质量特征图，从而提高模型效率、泛化能力和训练稳定性，特别有助于精准提取茶园区域。

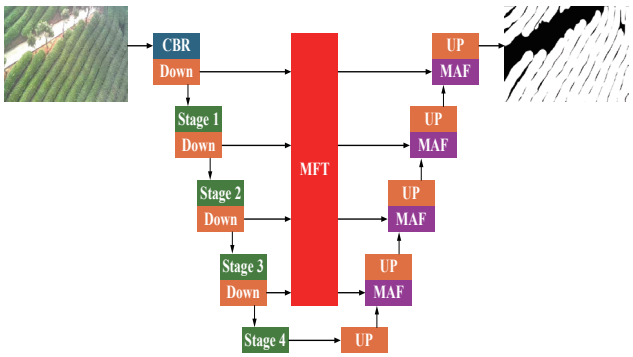

图 1 Tea-Unet 网络结构图

1.2 基于 RepVit 的轻量化增强编码器

经典的 U-Net 常使用 VGG16 作为编码器，虽然 VGG16 擅长提取局部特征，但它在捕捉长距离依赖和全局上下文方面存在局限，并且由于其参数量大、计算复杂度高，在资源受限环境下部署时面临挑战。为了解决这些问题，本文提出了一种基于 RepVit 的轻量化增强编码器来替代传统的 U-Net 编码器。如图 2 所示，该编码器包括 CBR 模块（3×3 卷积 +BN 层 +ReLU）、Down 模块（3×3 深度卷积 +1×1 逐点卷积 +FFN）以及 4 个 Stage，每个 Stage 包含一个带有 SE 注意力机制的 RepVitSEBlock 和一个 RepVitBlock[8]。这些组件利用深度可分离卷积、SE 模块和 FFN，以高效捕获全局信息，同时保持较低的计算成本。特别是 RepVitBlock 通过结构重参数化技术，在训练中引入多分支拓扑以提升性能，而在推理时简化为单一分支以降低计算和内存需求。相较于 VGG16，基于 RepVit 的轻量化编码器不仅增强了全局特征建模能力，还降低了参数量和计算复杂度，使得模型在资源受限环境下的部署更加高效。

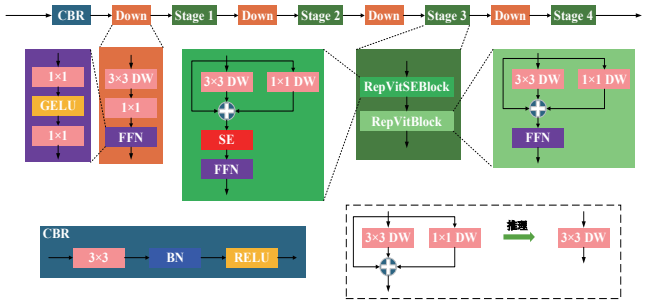

图 2 基于 RepVit 的轻量化增强编码器

1.3 基于 Transformer 的多级多尺度特征融合模块

U-Net 的跳跃连接通过直接关联编码器与解码器相应层的特征图，有效地保留和传递了空间信息，提升了分割效果。然而，这种方法也增加了挑战：浅层与深层特征间的语义差异可能导致性能下降；简单地引入编码器特征可能带来噪声或冗余，并因未加权的特征叠加而引发梯度问题，忽视了各层特征的重要性。为应对上述问题，提出了一种基于 Transformer 的多级多尺度特征融合模块（MFT），部署于 U-Net 的编码器与解码器之间，旨在优化特征融合、减小语义差距，从而提高模型的学习效率和分割精度。MFT 结构如图 3 所示，它由多尺度特征嵌入、多头交叉注意力机制和多层感知机（MLP）组成。

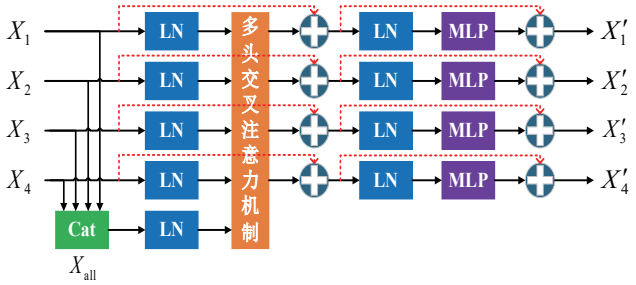

图 3 基于 Transformer 的多级多尺度特征融合模块

多尺度特征嵌入：给定 4 个跳跃连接层的输出 $\boldsymbol{E}_i \in \mathbb{R}^{H_i W_i \times C_i} (i = 1,2,3,4)$，首先将这些特征重塑为一系列扁平化的二维（2D）补丁序列，以进行标记化处理。各个补丁的尺寸分别为 P、P^2、P^4、P^8，确保其能映射到编码器特征图的相同区域，同时保持原有的通道维度不变。将来自 4 个不同层次的标记 $\boldsymbol{X}_i (i = 1,2,3,4)$，每个 $\boldsymbol{X}_i \in \mathbb{R}^{H_i W_i \times C_i}$，拼接起来作为键和值，形成 $\boldsymbol{X}_{\text{all}} = \text{Concat}(X_1, X_2, X_3, X_4)$

多头交叉注意力机制与 MLP：多头注意力机制有五个输入，包括 4 个作为查询 $\boldsymbol{Q}$ 的 $\boldsymbol{X}_i$ 和一个作为键 $\boldsymbol{K}$ 和值 $\boldsymbol{V}$ 的 $\boldsymbol{X}_{\text{all}}$：

$$\boldsymbol{Q}_i = \boldsymbol{X}_i \boldsymbol{W}_{Q_i},\quad \boldsymbol{K} = \boldsymbol{X}_{\text{all}} \boldsymbol{W}_K,\quad \boldsymbol{V} = \boldsymbol{X}_{\text{all}} \boldsymbol{W}_V \tag{1}$$

式中：$\boldsymbol{W}_{Q_i} \in \mathbb{R}^{C_i \times d}$、$\boldsymbol{W}_K = \mathbb{R}^{C_\Sigma \times d}$、$\boldsymbol{W}_V = \mathbb{R}^{C_\Sigma \times d}$分别为不同输入的权重矩阵；$d$ 是序列长度（块数）；C_i（$i = 1, 2, 3, 4$）是 4 个跳跃连接层的通道维度，在本研究中，C_1=64、C_2=128、C_3=256、C_4=512。

通过$\boldsymbol{Q}_i \in \mathbb{R}^{C_i \times d}$、$\boldsymbol{K} = \mathbb{R}^{C_\Sigma \times d}$、$\boldsymbol{V} = \mathbb{R}^{C_\Sigma \times d}$，生成相似性矩阵 $\boldsymbol{M}_i$，并通过交叉注意力机制 CA 用 $\boldsymbol{M}_i$ 对 $\boldsymbol{V}$ 进行加权：

$$\text{CA}_i = \boldsymbol{M}_i \boldsymbol{V}^{\mathsf{T}} = \sigma\left[\psi\left(\frac{\boldsymbol{Q}_i^{\mathsf{T}} \boldsymbol{K}}{\sqrt{\boldsymbol{C}_\Sigma}}\right)\right] \boldsymbol{V}^{\mathsf{T}} = \sigma\left[\psi\left(\frac{\boldsymbol{W}_{Q_i}^{\mathsf{T}} \boldsymbol{T}_i^{\mathsf{T}} \boldsymbol{T}_\Sigma \boldsymbol{W}_K}{\sqrt{\boldsymbol{C}_\Sigma}}\right)\right] \boldsymbol{W}_V^{\mathsf{T}} \tag{2}$$

式中：$\psi(\cdot)$ 和 $\sigma(\cdot)$ 分别表示实例归一化和 softmax 函数。

与 Transformer 原始自注意力的主要区别在于，本文的注意力操作是在通道轴上而不是在块轴上执行，且使用了实例归一化，这可在相似度映射上对每个实例的相似度矩阵进行归一化，从而使得梯度可以平滑传播。在 N 头注意力的情

况下，多头交叉注意力后的输出计算为：

$$\mathrm{MCA}_i = \left(\mathrm{CA}_i^1 + \mathrm{CA}_i^2 + , ..., + \mathrm{CA}_i^N\right)/N \quad (3)$$

式中：N 是头的数量。

然后，应用一个 MLP 和残差运算符，得到输出：

$$O_i = \mathrm{MCA}_i + \mathrm{MLP}(Q_i + \mathrm{MCA}_i) \quad (4)$$

为简化起见，上述方程中省略了层归一化（LN）。经过上述操作后，得到 4 个输出 $X_1^{'}$、$X_2^{'}$、$X_3^{'}$、$X_4^{'}$，再通过上采样操作后跟一个卷积层重建，即可分别与解码器相应层拼接。

相较于 U-Net 的跳跃连接，本文提出的 MFT 模块不仅增强了全局上下文理解和长距离依赖捕捉，同时保留了关键的空间分辨率和细节信息。通过引入多头交叉注意力机制，MFT 实现了更为复杂的特征融合，使模型能够学习到更抽象、语义丰富的表示，从而提升复杂结构目标的识别精度。此外，MFT 展示了更高的鲁棒性，抗干扰性更强。特别是在处理具有复杂内部结构的对象时，如细长物体或形态不规则的茶园区域，MFT 能更准确地建模这些特征，提供更加精细的分割结果。

1.4 多级注意力引导的特征融合模块 MAF

U-Net 的跳跃连接设计虽然有效地融合高低层次的特征信息，有助于保留图像细节并改善梯度流，但直接将尺寸一致的跳跃链接过来的特征图与上采样后的特征图进行拼接仍然会存在一些问题：特征冗余，引入不必要的低级特征；造成特征冲突，干扰高级语义特征的学习；以及缺乏选择性，无法有效区分和增强对任务有益的重要特征。为了解决这些问题，更好地融合语义不一致的特征，因此提出了一种多级注意力引导的特征融合模块（MAF），公式表示为：

$$Z_1 = A(X + Y) \otimes X + \left(1 - A(X + Y)\right) \otimes Y \quad (5)$$

$$Z_2 = AZ_1 \otimes X + (1 - AZ_1) \otimes Y \quad (6)$$

式中：X 为跳跃链接过来的特征图；Y 为上采样后的特征图；A 为注意力模块；Z_1 为一级特征融合后的特征；Z_2 为二级特征融合后的特征。

多级注意力引导特征融合模块，如图 4 所示，通过两次迭代优化特征表示。每级中将输入特征 X 和 Y 相加后，经由结合局部与全局注意力机制的模块处理，再与原始特征相乘并相加生成融合特征。二级融合基于一级生成的特征 Z_1 重复这一过程。每个级别的注意力模块结构相同，包括两个分支：局部注意力分支采用 Point-wise 卷积捕捉细节，全局注意力分支则通过平均池化和卷积层获取长距离依赖。此设计确保捕捉局部结构并强化整体上下文，经 Sigmoid 激活生成权重矩阵，实现平滑融合与对比度增强。多级融合不断应用注意力模块，依据前一级结果更新特征图，聚焦关键信息，减少干扰。这不仅促进了高质量特征图的生成，提升了模型效率和泛化能力，还改善了深层网络中的梯度传播，使训练更加稳定。

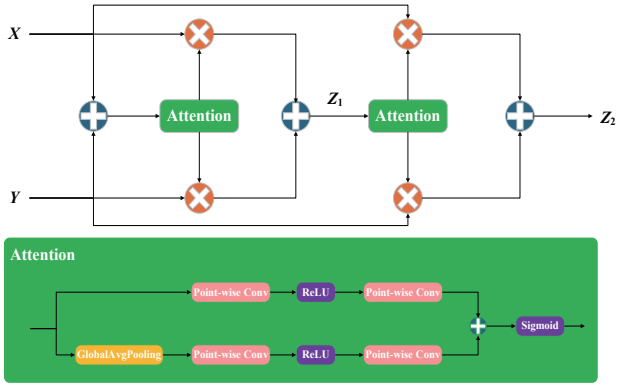

图 4 多级注意力引导的特征融合模块

2 实验

2.1 数据集采集及预处理

本研究中使用的数据集是在福建省南平市建阳区名芳茶园进行采集的。在本研究中，使用了搭载有高分辨率摄像头的兽 SG906 型号无人机作为数据采集设备，该摄像头的分辨率为 1 920 px×1 080 px。首先，在不同的天气条件和时间段内对茶园的不同区域进行了数据采集，总共收集 10 段、每段时长为 60 min 的视频；其次，对采集到的视频进行了抽帧处理，并剔除了模糊和无效的视频帧；接着，将筛选后的视频帧利用 Labelme 工具进行数据标注，得到每一张图片对应的标签图，原图及其标签示意图如图 5 所示；最后，获得 2 000 张标注图像，构成了原始数据集。在实验过程中，按照 7:1:2 的比例将原始数据集划分为训练集、验证集和测试集。

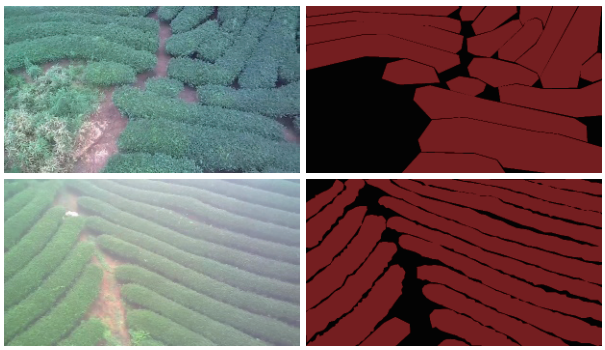

图 5 数据及其标签示意图

2.2 实验设置

本研究中的模型训练是在 Nvidia RTX 4080 GPU 上使用深度学习框架 PyTorch 进行的，并采用了 CUDA 和 cuDNN 加速环境来加快训练速度。安装的硬件和环境配置如表 1 所示。

表 1 实验配置

硬件 & 配置	参数
CPU	Intel i7-10700K CPU
GPU	NVIDIA GeForce RTX 3090
操作系统	Ubuntu 18.04
加速环境	CUDA 10.2+ cuDNN 7.6.5
深度学习框架	PyTorch 1.9.1

在训练过程中，首先通过标准差对图像数据进行了归一化处理，并将其分割成若干小块。随后，这些图像小块被输入到模型中进行特征提取和特征映射，从而生成初步预测结果。接下来，采用损失函数来量化预测结果与实际值之间的差异，并使用自适应矩估计（Adam）优化器执行反向传播迭代过程，以动态调整模型参数及学习率。此外，为了防止模型在训练数据集上过拟合，本研究设置了早停策略：如果连续 10 个周期内验证数据集上的损失没有改善，则提前结束训练。最终，在验证数据集中表现最佳（即损失值最低）的模型被选为最优模型。具体的超参数配置如表 2 所示。

表 2 超参数配置

图像输入尺寸	初始学习率	批次大小	迭代次数
640 px×640 px×3 px	1e-4	4	100

2.3 评价指标

本研究中的茶园区域提取任务是一个语义分割二分类任务，因此本研究引入了在语义分割中常用的评估指标精确率（Precision）、召回率（Recall）、交并比（IoU）来评价测试数据集上的预测结果。精确率衡量的是模型预测为正例（Positive）的样本中有多少是真正正确的。召回率衡量的是所有实际正例中被模型正确识别出来的比例。IoU（交并比）度量了预测区域与真实标注区域之间的重叠程度，是一个从 0 到 1 的值，值越接近 1 表示预测越准确。这些指标共同提供了对模型性能的不同方面进行评价的方法，其中精确率关注于模型预测出的正例中实际有多少是正确的；召回率关注于模型能够找到多少实际存在的正例；而 IoU 则综合考虑了预测与真实标签之间的重叠程度。

上述评估指标的公式分别为：

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} = \frac{S_{\text{pred}} \cap S_{\text{truth}}}{S_{\text{pred}} \cup S_{\text{truth}}} \tag{9}$$

式中：TP（true positive）是指正确预测为正类别的像素数量；FP（false positive）是指错误地将负类别预测为正类别的像素数量；FN（false negative）是指实际为正类别但被模型错误地预测为负类别的像素数量；$S_{\text{pred}} \cap S_{\text{truth}}$ 是指预测和真实的茶园区域的交集；$S_{\text{pred}} \cup S_{\text{truth}}$ 是指预测和真实的茶园区域的并集。

3 实验结果与分析

3.1 消融实验

为评估 Tea-UNet 模型中各个组件对最终分割效果的具体贡献，本研究进行详尽的消融实验。表 3 展示不同配置下模型在测试集上的表现。

表 3 消融实验评价指标

模型	Precision	Recall	MIoU
U-Net	0.888 0	0.884 2	0.803 5
U-Net+RepVit	0.912 2	0.884 0	0.820 6
U-Net+MFT	0.905 3	0.887 8	0.819 0
U-Net+MAF	0.893 1	0.884 9	0.807 8
U-Net+RepVit+MAF	0.916 7	0.894 4	0.832 6
U-Net+RepVit+MFT+MAF	0.918 1	0.904 3	0.841 9

通过引入基于 RepVit 的轻量化增强编码器，模型的 Precision 从 0.888 0 提升至 0.912 2，MIoU 亦由 0.803 5 升至 0.820 6。此改进彰显了增强编码器在捕捉茶园特征信息、提升全局理解与局部细节保留上的优越性。MFT 模块的应用则使 Recall 从 0.884 2 微增至 0.887 8，MIoU 维持在 0.819 0 的高水平，表明其在减少语义差距及优化复杂结构目标分割方面的贡献。MAF 模块单独使用时，虽对 Precision 和 MIoU 的增益不如 RepVit 显著，但 MIoU 仍见小幅增长至 0.807 8。结合 RepVit（U-Net+RepVit+MAF）后，Precision 达 0.916 7，Recall 为 0.894 4，MIoU 升至 0.832 6，显示连续注意力机制更新对精炼特征表示、排除无关信息以及强化关键特征学习的重要性。最终配置实现了 Precision 0.918 1、MIoU 0.841 9 和 Recall 0.904 3 的最佳成绩，证明各组件协同工作的高效性，尤其是增强编码器、MFT 和 MAF 的有效集成，为应对实际应用挑战提供了强有力的技术解决方案。

通过图 6 中的视觉对比可以更直观地看到不同配置下模型的表现差异。从原图到标签，再到各模型的预测结果，可以清晰地观察到每个组件对茶园区域提取效果的影响。相较于传统的 U-Net 模型，引入基于 RepVit 的轻量化增强编码器显著提升了茶园边界的检测准确性，减少了误报现象；添加 MFT 模块后，细长或形状不规则的目标如茶树行间的识别精度明显提高；而 MAF 模块则进一步精炼了特征表示，降低了无关信息的干扰，使得茶园内部的连贯性和边缘清晰度都有所改善。完整模型（U-Net+RepVit+MFT+MAF）不仅保持了对复杂结构目标的良好识别能力，而且在整个茶园区域提取任务上展现了最优异的整体性能，包括更好的边缘清晰度、

内部连贯性以及更高的细节保留，证明了各个组件协同工作的优越性。

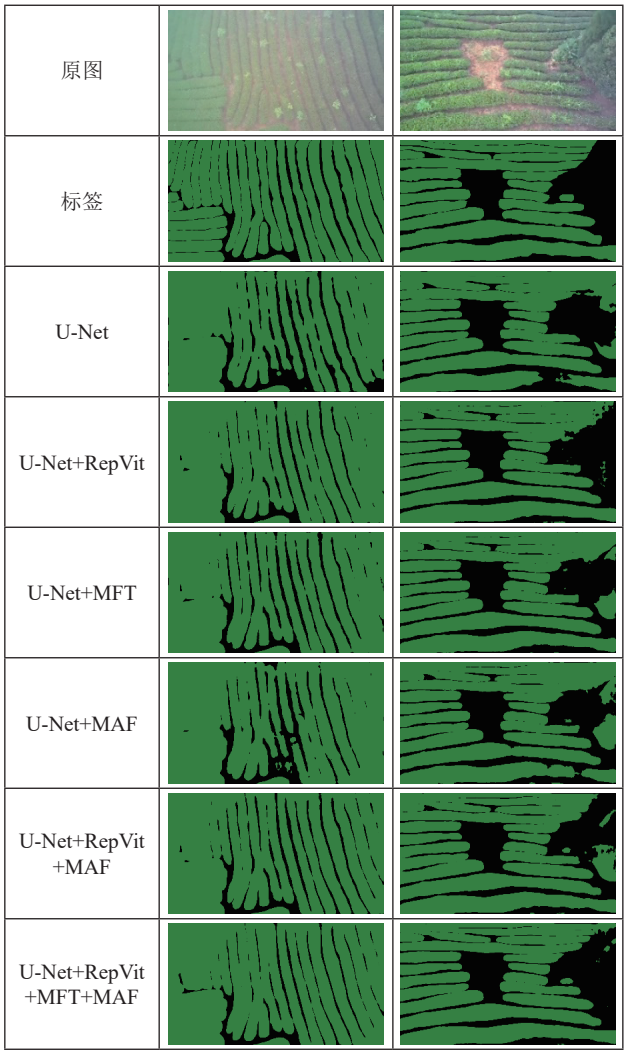

图 6 消融实验 - 茶园区域提取效果对比

3.2 横向对比实验

为评估 Tea-UNet 模型在茶园无人机影像语义分割任务上的表现，本文进行了横向对比实验，验证了其有效性与优越性，如表 4 所示。

表 4 横向对比实验评价指标

模型	Precision	Recall	MIoU
PSPNet[9]	0.854 3	0.781 6	0.702 6
HRNetv2[10]	0.894 4	0.865 0	0.793 0
TransUNet[11]	0.903 9	0.893 3	0.822 3
SegFormer[12]	0.860 7	0.822 0	0.738 4
UNeXt[13]	0.863 5	0.845 0	0.757 4
UCTransNet[14]	0.907 7	0.890 3	0.822 7
SA2Net	0.919 8	0.874 4	0.817 5
FusionU-Net	0.898 9	0.893 2	0.818 5
Tea-UNet	0.918 1	0.904 3	0.841 9

实验中，Tea-UNet 与其他先进模型如 PSPNet、HRNetv2、SegFormer 等在同一条件下训练和测试。结果显示，Tea-UNet 在 Precision、Recall 和 MIoU 上分别达到了 0.918 1、0.904 3 和 0.841 9 的最优成绩，尤其在捕捉全局信息和处理复杂背景方面表现突出，优于其他模型。与 UCTransNet 和 SA2Net 相比，Tea-UNet 提供了更高的分割准确性和更平滑的边缘检测，彰显其在茶园无人机影像分析中的优势。

通过图 7 中的视觉对比可以更直观地看到不同模型在茶园区域提取任务上的表现差异。

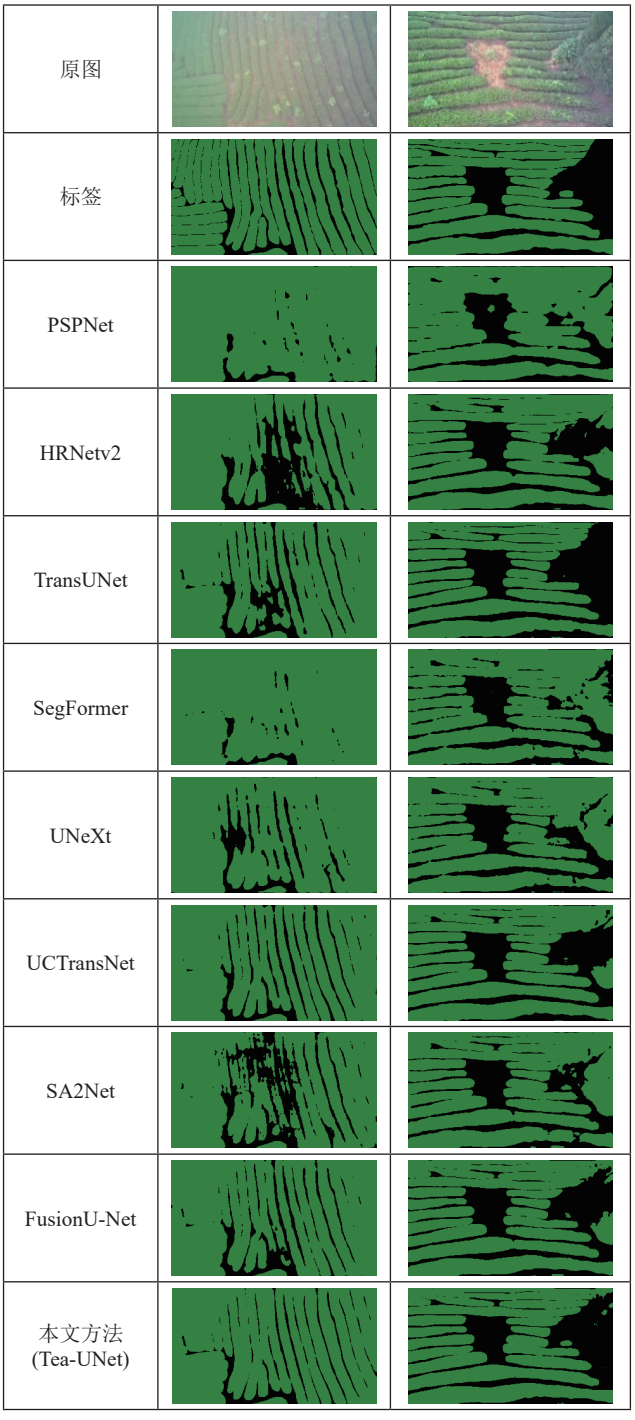

图 7 横向对比实验 - 茶园区域提取效果对比

从原图到标签，再到各模型的预测结果，可以清晰地观察到 Tea-UNet 相较于其他模型的优势。相较于其他模型，Tea-UNet 不仅能够更好地处理复杂的背景和相似作物的干扰，还能准确识别细长或形状不规则的目标。此外，Tea-UNet 生成的分割结果边缘更加清晰，内部连贯性更好，显著减少了误报现象，体现了其在细节保留和噪声抑制方面的卓越能力。最终，这些特点使得 Tea-UNet 在整个茶园区域提取任务上展现了最优异的整体性能，进一步验证了该模型在实际应用中的优越性和可靠性。

4 结论

本文开发了名为 Tea-UNet 的新型语义分割算法，专门针对无人机高分辨率影像高效准确提取茶园区域，通过引入轻量化的 RepVit 增强编码器、基于 Transformer 的多级多尺度特征融合模块和多级注意力引导的特征融合模块，不仅提升了细长或形状不规则目标的识别精度，还增强了全局上下文理解和局部细节保留能力。实验表明，Tea-UNet 在 MIoU 等关键指标上优于现有模型，展示了较高的鲁棒性和泛化能力，为茶园管理和监测提供了技术支持，促进了茶叶生产和生态环境的可持续发展。未来研究将致力于优化模型结构与训练策略以提高计算效率、降低参数量，并探索 Tea-UNet 在更多农业场景及其他领域的应用。

参考文献：

[1] 叶志立 .CsMYB184 调控茶树生物碱合成机制研究 [D]. 合肥 : 安徽农业大学 , 2023.

[2] 邱海蓉 . 基于需求视角的中国茶叶出口贸易研究 [D]. 武汉 : 华中农业大学 , 2010.

[3] 何少林 . 基于无人机遥感影像的土地信息提取及专题图制作研究 [D]. 成都 : 西南交通大学 ,2013.

[4] LIN T Y, DOLLAR P, GIRSHICK R, et al.Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 936-944.

[5] BRAUWERS G, FRASINCAR F. A general survey on attention mechanisms in deep learning[J]. IEEE transactions on knowledge and data engineering,2021,99:1.

[6] LI H F, QIU K J, CHEN L, et al. SCAttNet: semantic aegmentation network with spatial and channel attention mechanism for high-resolution remote sensing images[J].IEEE geoscience and remote sensing letters, 2021,18(5):905-909.

[7] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Berlin: Springer, 2015: 234-241.

[8] WANG A, CHEN H, LIN Z J, et al. RepViT: revisitingmobile CNN from ViT perspective[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2024:15909-15920.

[9] ZHAO H S, SHI J P, QI X J, et al. Pyramidscene parsing network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).Piscataway: IEEE, 2017: 6230-6239.

[10] SUN K, ZHAO Y, JIANG B R, et al. High-resolution representations for labeling pixels and regions[EB/OL]. (2019-04-09)[2024-09-13].https://doi.org/10.48550/arXiv.1904.04514.

[11] CHEN J N, LU Y Y, YU Q H, et al.TransUNet: transformers make strong encoders for medical image segmentation[EB/OL]. (2021-02-08)[2024-03-10].https://doi.org/10.48550/arXiv.2102.04306.

[12] XIE E Z, WANG W H, YU Z D, et al. SegFormer: simple and efficient design for semantic segmentation with transformers[EB/OL]. (2021-10-28)[2024-01-11].https://doi.org/10.48550/arXiv.2105.15203.

[13] VALANARASU J M J, PATEL V M. UNeXt: mlp-based rapid medical image segmentation network[C]//Medical Image Computing and Computer Assisted Intervention–MICCAI 2022. Berlin:Springer,2022:23-33.

[14] WANG H N, CAO P, WANG J Q, et al. UCTransNet: rethinking the skip connections in U-Net from a channel-wise perspective with transformer[C/OL]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto:AAAI,2021[2024-03-03].https://aaai.org/papers/02441-uctransnet-rethinking-the-skip-connections-in-u-net-from-a-channel-wise-perspective-with-transformer. DOI:10.1609/aaai.v36i3.20144.

【作者简介】

熊盛佳（2000—），女，湖南益阳人，硕士研究生，研究方向：人工智能，email:1483551312@qq.com。

叶智国（1999—），男，江西上饶人，硕士研究生，研究方向：人工智能，email:71972144@qq.com。

吴章云（1999—），男，湖南郴州人，硕士研究生，研究方向：人工智能，email:741124643@qq.com。

吴传宇（1976—），通信作者（email:360626964@qq.com），男，福建建瓯人，博士，高级工程师，研究方向：农业机械化。

（收稿日期：2025-01-02 修回日期：2025-05-12）