arXiv:2602.13022v1 [cs.CV] 13 Feb 2026

利用增强的基于激光雷达的伪标签学习基于图像的树冠分割

Julius Pesonen[a, b, c, *], Stefan Rua[a], Josef Taher[a], Niko Koivumäki[a], Xiaowei Yu[a], Eija Honkavaara[a]

[a]芬兰地理空间研究所遥感与摄影测量系，埃斯波，*02150*，芬兰[b]阿尔托大学计算机科学系，埃斯波，*02150*，芬兰[c]*DOTA*，*ONERA*，图卢兹大学，图卢兹，*31000*，法国

摘要

单木树冠制图对于维护城市树木清查和监测森林健康等任务至关重要，这些工作有助于我们了解并呵护周围环境。然而，由于纹理以及树冠部分重叠等因素，要在航空影像中自动分离彼此相邻的树冠颇具挑战性。本研究提出了一种训练深度学习模型的方法，利用航空激光扫描（ALS）数据生成的伪标签，从 RGB 和多光谱图像中分割并分离单木。研究表明，可使用零样本实例分割模型 Segment Anything Model 2（SAM 2）增强基于 ALS 的伪标签。我们的方法无需任何人工标注成本，即可为基于光学图像的模型获取特定领域的训练标注，进而训练出在同一任务上优于所有现有通用领域模型的分割模型。

关键词： 遥感、深度学习、传感器融合、弱监督学习、实例分割、单木树冠勾绘

1. 引言

单木树冠制图有助于我们保护环境，并支持城市树木清查和森林健康监测等关键工作。例如，城市树木的高分辨率地图有助于我们了解树木对城市居民健康的影响 [1] ，并量化其碳储量 [2, 3]。绘制森林地图对林业至关重要 [4], ，此外还有其他重要用途，例如帮助我们保护森林免受日益增多的害虫侵害 [5, 6] ，并评估森林火灾风险 [7]。

树木清查可以利用多种数据源开展和维护，从地面专家测量 [8] 到自动化卫星图像分析 [9, 10]。航空数据采集介于这两种方法之间，与野外测量相比，它能更快地绘制更大范围的地图；与卫星成像相比，则具有更高的空间分辨率。然而，要在此类数据中自动将单棵树木彼此分离，并从背景中区分出来，从而创建高分辨率

*通讯作者电子邮箱： julius.pesonen@nls.fi (Julius Pesonen)

树木地图，并非易事。近年来，最佳结果大多通过机器学习方法 [11], 获得，这类方法依赖大量通常由人工标注的训练数据。用于检测树木的影像也会因地理位置、所用传感器、天气以及一年中的时节或一天中的时段而有很大差异。因此，为了获得能在这些不同情境下可靠运行的方法，需要高效的训练数据采集方法。

分离单棵树木的过程称为树冠实例分割；在本研究中，也将其称为单木树冠勾绘，两者可以互换使用。关于该任务自动化方法的研究，大致可明确分为基于激光雷达的方法和基于相机的方法。近年来，这两类方法的研究都着重于利用基于学习的方法。然而，除了少数直接使用融合传感器数据来生成树木形状估计的研究 [12, 13]外，基于这两种不同传感器类型的方法交集仍在很大程度上尚未得到探索。获取足够的标注训练数据是基于学习的方法面临的一项重大挑战，因为不可能为所有可能的地点类型和树种都创建人工标注的数据集。

本研究表明，可以利用基于冠层高度模型（CHM）的分割区域训练 RGB 和多光谱图像模型；这些分割区域源自低分辨率航空激光扫描（ALS）测量数据。我们将这些模型与其他基于图像的方法进行了比较，结果显示，即使使用粗略或带有噪声的标签训练基于图像的模型，其结果也比当前的零样本模型选项更稳健。这为获取已开展 ALS 测量和航空影像采集区域的树木分割信息提供了新方案，也为获取树冠实例分割模型的训练数据提供了新途径；这些模型可部署于仅有图像数据的区域。此外，我们的结果表明，在处理流程中使用零样本实例分割模型 SAM 2 [14], ，将粗略的基于冠层高度模型的分割区域转化为树冠图像标签，具有显著优势。针对这一具体任务使用 SAM 2 进行预处理，最早由 Stefan Rua 在其硕士论文 [15], 中提出，本研究在此基础上进行了扩展。

2. 相关工作

近年来，基于单木的方法日益受到关注，使树冠分割成为森林监测与分析中的重要课题。树冠通常通过激光雷达、光学传感器或两者结合来检测。由于传感器相对昂贵且并不常见，激光雷达数据不像光学相机图像那样容易获取；而使用光学传感器数据且性能最佳的方法通常基于监督学习，这意味着它们通常需要手工制作的标签进行训练。针对这一任务，人们已在不同数据集上训练了多种监督模型。本研究撰写时已公开发布、因而可用于我们的数据进行测试的模型，已在第 3.6 节中介绍。使用这类模型得到的指标结果会因数据而有很大差异。这种差异由多种因素造成，例如数据分辨率、测试数据与训练数据的相似程度，以及训练样本数量。

一项综述汇总的相关指标（也考虑了激光雷达数据和语义分割）包括：F1分数为 0.82 至 0.98，精确率为 0.85 至 0.91，召回率为 0.88 至 0.91，平均交并比（mIoU）为 0.49 至 0.94 [16]。介绍基线方法之一 Detectree2 [17], 的论文报告的测试指标包括：精确率为 0.60 至 0.71，召回率为 0.54 至 0.66，F1 分数为 0.57 至 0.69，具体数值取决于测试区域。类似地，一项研究使用监督式U-Net处理航空影像 [18], ，其结果也被本研究用作

一种基线方法，在航空图像上的精确率为 0.71，召回率为 0.66。在卫星影像上，该模型的表现略差，精确率为 0.63，召回率为 0.64。

利用激光雷达衍生树木分割片段训练基于图像的模型，是兼顾两种模态优势的一种方法。已有研究将其应用于树木密度较低的针叶林—苔原交错带数据 [19]，以及德国的高密度激光雷达数据 [20]。然而，与分辨率相近的光学影像相比，获取高密度激光雷达数据的成本较高。使用较低分辨率的激光雷达数据生成树冠分割片段，则会产生噪声标签。为利用光学图像中的信息增强标签，可以使用任意分割模型，例如 Segment Anything Model（SAM） [21]，及其较新的变体 SAM 2 或 3 [22] 。已有研究使用基于学习的提示 [23]或数字表面模型衍生提示 [24]，探索了 SAM 在树木分割中的应用。研究人员也将 SAM 2 与通用树木检测模型 DeepForest [25, 26]结合，用于这项任务。最后，Grounded SAM [27]，（Grounding DINO [28] 与 Segment Anything Model [21], 的组合）已用于灾害管理背景下的树木制图 [29]。此外，研究人员还评估了 SAM 2 在不同提示方式下（使用激光雷达、人工提示和基于模型的提示）的单木树冠勾绘直接表现 [30]。

以往研究尚未探索使用经 SAM 增强的激光雷达分割区域来训练基于光学图像的模型，以分割单木树冠的可能性。这为获取质量更高、针对特定领域的标注数据来训练模型提供了独特机会。

3. 材料与方法

3.1. 数据

我们使用多光谱影像和点云研究了这些方法。所用图像数据包括一幅空间分辨率为 5 cm、于 2023年7月采集的正射影像。完整正射影像的尺寸为 31 254 × 36 871 像素，覆盖面积约为 2 km^2。多光谱影像由直升机搭载 MicaSense Altum-PT 在 120 m 高度采集。Micasense Altum-PT 可采集五个分辨率为 3.2 MP 的多光谱波段、一个 12.4 MP 的全色波段以及一个热红外波段。影像的摄影测量处理采用 Agisoft Metashape Professional [31] 完成。在光束法区域网平差步骤中，使用了从激光雷达数据中提取的 66 个地面控制点，以改善影像的对齐效果。本研究仅使用五个多光谱波段，分别为 475 nm（蓝色）、560 nm（绿色）、668 nm（红色）、717 nm（红边，RE）和 842 nm（近红外，NIR）。相应的带宽分别为 32 nm、27 nm、16 nm、12 nm 和 57 nm。

多光谱机载激光扫描（ALS）点云于2016年6月14日在展叶期条件下，使用 Optech Titan 传感器（Teledyne Optech，加拿大安大略省）采集。TerraTec Oy（芬兰赫尔辛基）在研究区域上空以700米的高度执行了一次固定翼数据采集飞行，生成了三个不同的点云，分别对应1550 nm（红外）、1064 nm（近红外）和532 nm（绿色）激光波长通道。

原始点云经过了初始辐射校准和几何校准，其中包括强度范围归一化（Matikainen等的研究 [32]对此进行了进一步描述），以及移除冗余的航线重叠，以确保空间均匀性。所得点云的平均点密度分别为通道1、2和3的11、13和11点/m^2 。所有数据均已地理配准至ETRS-TM35FIN坐标系，垂直值则参照N2000正高系统。

我们预留了一部分数据用于测试。该部分选取为正射影像中的一个单独区域，大小为4471×4893像素。根据人类对同一正射影像的感知，对该区域内的树木进行了人工标注。主要结果包括模型生成的树木分割区域与人工绘制标注之间的典型分割指标。该区域的正射影像及分离出的测试区域如图1所示。测试区域包含362棵人工分割的树木，其余用于训练和验证的正射影像包含约24,000棵树木。

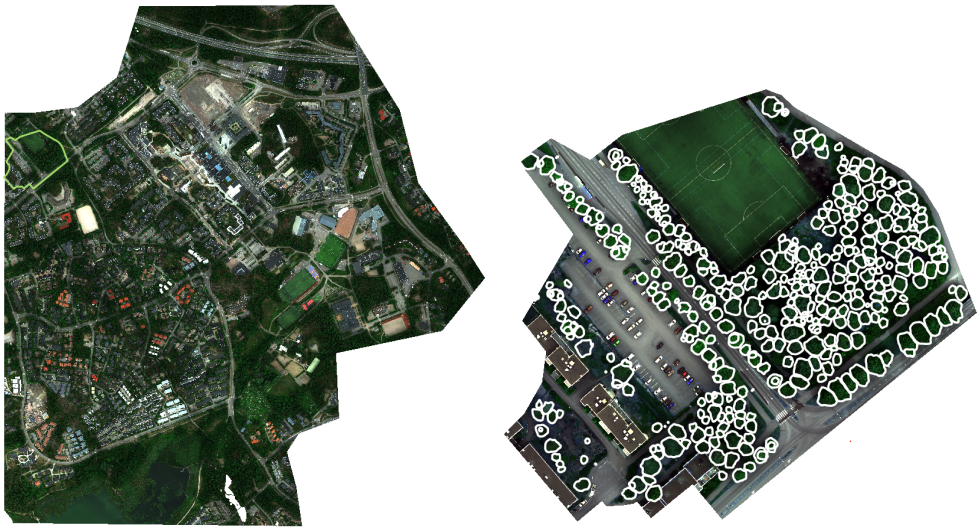

图1：左侧为覆盖整个数据集区域的正射影像 RGB 可视化，手动标注的测试区域以浅绿色勾勒并单独显示在右侧。手动测试标注以白色勾勒。

该数据集采集自芬兰埃斯波的埃斯波恩拉赫蒂区（中心约位于 60.1462°N 和 24.6587°E），该地区森林类型多样，既有典型的芬诺斯堪的亚沿海景观，也有经人工经营的芬兰北方森林。根据 Taher 等人 [33],在该地区的研究统计，树种分布约为：37% 的松树（欧洲赤松）、25% 的桦树（桦属植物）、11% 的槭树（Acerplatanoides）、11% 的云杉（云杉属植物）、8% 的欧洲山杨（欧洲山杨）、3% 的花楸（Sorbussp.）、3% 的椴树（椴属植物）、2% 的桤木（桤木属植物），以及 1% 的橡树（欧洲栎）。

3.2. 粗分割

初步处理完成后，从 Optech Titan 点云中提取了单木树冠标签。首先使用 LAStools 软件套件 [34] 将回波分类为地面、植被、建筑物和噪声类别。随后根据地面分类点插值得到数字地形模型（DTM），从而将植被回波归一化为离地高度。

Kaartinen 等人 [35]采用可变窗口分水岭分割方法进行了单木树冠划分。利用通道 1 和通道 2（1550 nm 和 1064 nm）的合并首回波生成了分辨率为 0.5 m 的冠层高度模型，以便详细捕捉冠层结构。对冠层高度模型进行高斯滤波器平滑后，使用局部

最大值滤波器识别树梢。这些树梢位置作为标记，用于标记控制分水岭算法，以划分单株树冠边界。随后，所得单株树冠分割片段被用作后续处理步骤中的伪标签。Karila 等人 [36]和 Taher 等人 [33]的研究对激光雷达数据的采集和处理流程进行了更详细的说明。

由于点云的采集时间比多光谱影像早七年，因此被建成环境取代的区域中出现了相当多的假阳性分割片段。我们使用归一化差异植被指数阈值过滤这些假阳性分割片段

$$\mathrm{NDVI} = (\mathrm{RE} - \mathrm{Red})/(\mathrm{RE} + \mathrm{Red}). \tag{1}$$

该指数根据多光谱正射影像计算得出。阈值设为 0.2，平均 NDVI 低于该阈值的分割区域将被舍弃。阈值是根据对数据的人工观察设定的。附录中的图 A.1 展示了分割区域 NDVI 分布的直方图，表明大多数分割区域的 NDVI 明显高于或低于设定阈值。

3.3. SAM 2 增强

由于可用激光雷达点云的分辨率较低，基于冠层高度模型的分割片段呈现明显的块状。为了生成更精确的标注，我们采用了 SAM 2。SAM 2 是一种模型，它以图像和查询（例如点或边界框）作为输入，并输出与这些查询相对应的分割区域。为了生成树冠掩膜，我们将粗略标签的边界框（包含所有粗略标签像素的最小矩形）作为 SAM 2 的查询输入。

为避免只分割部分可见的树木，我们将正射影像裁切为大小为 1024 × 1024 像素的图像块，步长为 512 像素。这样，只需对每个图像块中央 512 × 512 像素区域内质心所在的树木进行伪标注，即可捕获所有树木。之所以采用这一特定配置，是因为根据树冠大小的分布，只有极少数树木的树冠足够大，可能从中央区域延伸至图像外。图2展示了 SAM 2 生成的伪标签在五个测试集样本上的可视化结果，并同时呈现了人工分割区域和粗分割区域。

3.4. 伪监督模型

我们使用冠层高度模型和 SAM 2 衍生的标签训练下游模型，使其能够直接从 RGB 或多光谱图像中预测单棵树冠的掩膜。图3概述了模型训练和评估方案。我们选择了改进版 Mask R-CNN [37, 38]，其骨干网络为经 ImageNet [39] 预训练的 ResNet-50-FPN [40, 41]。选择这一特定配置，是因为它在通用实例分割和树冠实例分割文献中均有重要地位。对于使用 RGB 及红边或近红外波段的多光谱模型变体，我们通过重复图像首层的参数对模型进行了调整。

模型基于简单的求和损失函数进行优化 $L = L_{cls} + L_{box} + L_{mask}$，其中 L_{cls} 是分类损失，L_{box} 是边界框预测损失，L_{mask} 是掩码损失。原始介绍 Mask R-CNN 架构的研究 [37] 对该损失进行了详细说明。模型使用 Adam 优化器 [42]和解耦权重衰减 [43, 44]进行训练。训练损失和验证损失均基于自动生成的粗略

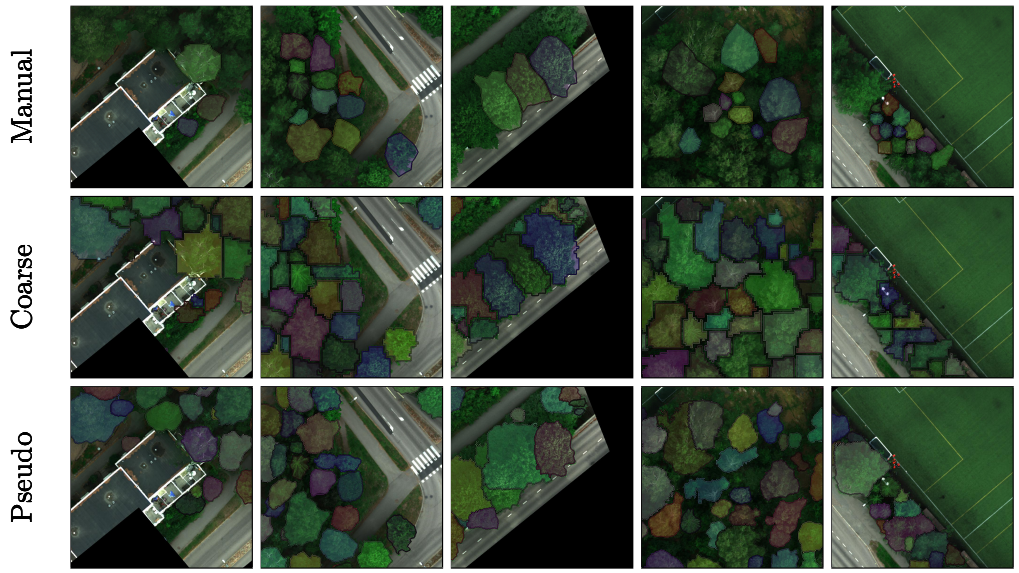

图2：测试集上不同标签的可视化。

标签与经 SAM 2 增强的伪标签计算得出。在训练过程中，训练和验证均使用预先抽样的一组图像块，大小为 1024 × 1024 像素。验证集采用伪随机算法抽样，所有模型变体均使用相同的验证集划分。验证集包含 209 张（19%）图像，训练集包含 886 张（81%）图像。按照 Liu 等人 [45] 的研究方法，对训练图像进行水平翻转和垂直翻转、随机旋转 [-180°, 180°], 以及随机 IoU 裁剪等数据增强。随机 IoU 裁剪的尺度范围为 0.3 至 1.0，宽高比范围为 0.5 至 2.0。每种数据增强均独立执行，并在相应范围内均匀抽样。

验证集用于在 400 个训练轮次的训练过程中选择所获得的最佳模型，并通过试错优化训练与推理超参数。在最终测试集上评估的模型以 1e-6 的学习率、8 的批次大小和 1e-2 的权重衰减进行训练。推理时，我们使用 0.5 的掩膜阈值、0.3 的分数阈值，以及 0.3 的非极大值抑制（NMS）IoU。NMS 根据设定阈值移除重叠的输出掩膜，并保留分数较高的掩膜。

3.5. 评估

我们在地理上分离且经人工标注的测试区域上评估了所有方法。在大多数实验中，该区域被划分为 70 个大小为 1024×1024 像素的图像块。由于其中一个基线模型使用较小的图像块进行训练，我们采用了 512×512 大小的图像块，因为它们产生了更好的结果。这些图像块用于计算每个模型预测结果的精确率、召回率、F1 分数和杰卡德指数（平均交并比）。在计算精确率、召回率和 F1 时，预测结果与真实分割区域之间采用了 50% 的重叠阈值。

由于测试集规模很小，我们通过测试集自助法估算结果的 95% 置信区间。自助抽样以单棵树级别进行，

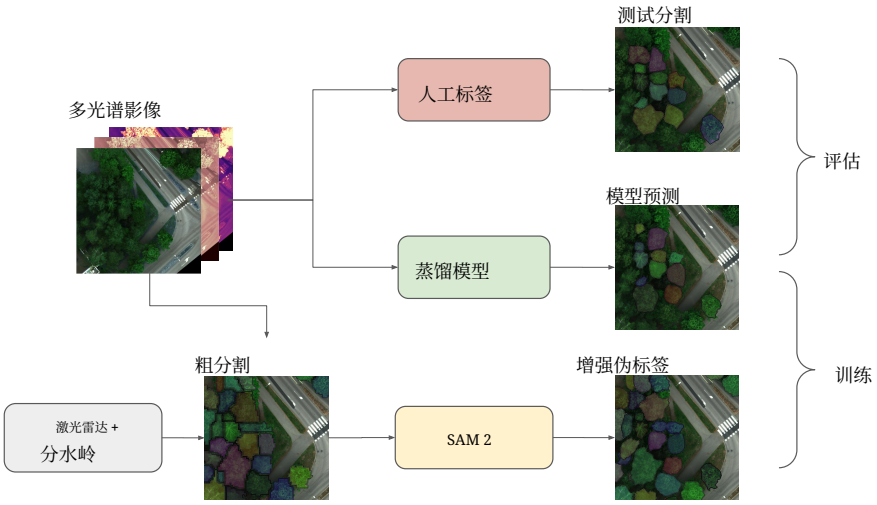

图 3：所提出的伪监督模型训练数据融合方法概览。

共生成 1000 组重采样数据。置信区间能更具代表性地估算该方法在其他数据上的表现，同时排除显著的域偏移。我们对所有方法采用相同的评估指标，包括分别使用冠层高度模型和 SAM 2 增强生成的粗略标签和伪标签。

3.6. 基线方法

我们将伪监督 Mask R-CNN 模型与多种不同的基线方法进行了比较。由于我们的方法无需使用本地采集的人工标注训练数据即可部署，因此排除了需要此类数据的方案。因此，这些基线方法代表了在不利用本地人工标注数据进行训练或微调的情况下，获取树冠分割预测的替代方案。下文介绍了所测试的基线方法。此外，我们还将粗略掩膜和 SAM 2 增强掩膜视为替代的树冠分割方法，因为它们可直接部署在具有 ALS 数据的地点，例如本研究的测试地点。我们仅考虑了源代码公开可用的方法。

Detectree2。 Detectree2 是一系列使用其他可用树冠实例分割数据训练的全监督 Mask R-CNN 模型，是该任务中最常用的模型之一 [17]。截至本研究开展时，共有七个预训练 Detectree2 模型可用，我们对每个模型的性能都进行了评估。这些模型均采用相同的 Mask R-CNN 架构和 ResNet 101 FPN 骨干网络，但训练数据组成和训练策略各不相同。

预训练模型检查点可在 Zenodo [46] 上获取。按照作者的命名惯例，这些模型分别称为 *withParacouUAV*，即使用热带地区无人机和飞机拍摄的影像训练的模型；*randresize_full*，即使用相同数据训练、并采用随机缩放增强的模型；*urban_trees_Cambridge*，即使用剑桥市区数据训练的模型；

数据来自英国剑桥；*base*，即使用飞机拍摄的热带地区数据训练的模型，在 Ball 等人较早的一项研究中被引入并称为*base*模型。 [47], *05dates*，即以*base* 模型为基础，进一步使用来自五个不同日期、由无人机拍摄的 ~5cm 分辨率 RGB 影像进行微调的副本；在 Ball 等人的同一项早期研究中，该模型称为*5 date* 模型。*flexi*，即使用来自城市和森林区域的数据训练的模型；以及*tropical_closed_canopy*，即专注于热带森林区域的模型的后续版本。所有模型均使用约 1000×1000 像素的图像块进行训练，空间分辨率约为 5 至 10 cm，因此我们将这些模型直接应用于与所提方法相同的 1024×1024 像素图像块。

监督式***U-Net***. 一种公开可用的完全监督树冠实例分割模型也已基于 U-net 架构开发 [18]。该模型已在~5厘米分辨率的航空影像上测试，我们将预训练模型按原样部署到我们的数据上。除其他基线模型所使用的 RGB 图像通道外，U-Net 还使用红边通道。预处理还包含一个步骤，即根据红色和红边波段生成 NDVI 图像通道，如公式 1 所示。该模型是为在 224×224 像素的图像块上进行推理而开发的。因此，我们考察了几种不同图像分块与缩放配置的精度。在结果中，我们报告了所获得的最佳指标，其配置是将正射影像切分为 512×512 像素的图像，并在输入神经网络之前将其缩放为 224×224 像素。

Grounded SAM (2). 这是 Grounded DINO [28] 与任意分割模型 [21], 的融合，以及利用 SAM 2 [14] 的后续版本。该模型利用来自 Grounded DINO 模型的零样本边界框来提示 SAM (2) [27]。该方法已应用于树冠分割场景，以辅助构建 FMARS 数据集 [29]。FMARS 数据集的作者指出，提示词“bushes”在航空图像中检测树木的效果最佳。我们将其与其他几个提示词一同测试，并得出了相同的结论。

DeepForest + SAM2。 DeepForest 是一种用于树冠边界框检测的模型。它使用各种树冠数据进行训练，有望至少在不同数据集之间具备一定的泛化能力 [25]。为了利用该模型进行分割，我们将其与 SAM 2 结合使用，并通过提示输入 DeepForest 生成的边界框来使用 SAM 2。

SAM 3。 Segment Anything Model 的最新版本 3 加入了原生文本提示功能，这意味着无需微调或借助外部目标检测模型，即可直接将其应用于图像 [22]。我们使用几种不同的提示在自己的数据上测试了 SAM 3，包括“灌木”和“树冠”，并发现简单的提示“树”效果最好。

4. 实验与结果

结果表明，在埃斯波-拉赫蒂数据集上，我们的方法优于所有开箱即用的替代方法。直观而言，这一结果并不令人意外，因为我们的模型利用了来自同一地点的训练数据。然而，需要注意的是，这些数据均未经过完整标注；相反，在有机载激光扫描数据可用的区域，用于训练模型的伪标签生成过程可以实现完全自动化。我们表现最佳的模型（使用 RGB 和近红外输入（RGB 近红外））、仅使用 RGB 输入的模型（RGB）以及所有基线方法的结果汇总于表1中。

方法	F1	精确率	召回率	mIoU	平均交并比 95%置信区间
Detectree2*	0.556	**0.911**	0.400	0.324	[0.281, 0.364]
U-Net	0.306	0.214	0.536	0.497	[0.472, 0.520]
Grounded SAM	0.166	0.430	0.103	0.094	[0.070, 0.121]
DeepForest	0.501	0.544	0.464	0.396	[0.363, 0.428]
SAM 3	0.558	0.822	0.422	0.352	[0.311, 0.397]
粗略掩膜	0.571	0.550	0.594	0.477	[0.447, 0.505]
伪掩膜	0.597	0.584	0.611	0.513	[0.478, 0.545]
RGB	0.758	0.783	0.733	0.599	[0.566, 0.632]
红绿蓝近红外	**0.778**	0.783	**0.772**	**0.621**	[0.587, 0.652]

表1：我们提出的方法（下方）与适用基线方法（上方）的比较。*Detectree2 的结果对应于使用*Flexi* 检查点所获得的最佳结果。七个检查点的完整结果见附录表A.1。

在纯基于图像的基线模型中，表现最佳的监督式U-Net 的平均交并比超过了粗略的基于冠层高度模型的分割区域，但仍不及所提出的方法。尽管如此，对于一个其他模型都无法产生稳健结果的数据集而言，该模型的准确率仍然令人印象深刻。这可能是因为它是唯一一个同时利用了可用红边波段的模型，而其他模型仅针对 RGB 图像进行训练。基线方法中的另一个显著成功案例是 SAM 3 和 Detectree2。虽然它们的平均交并比和 F1 分数明显低于我们最终模型所取得的结果，但由于假阳性很少，这些模型的精确率表现令人印象深刻。然而，这在一定程度上是因为模型过于谨慎，只识别出了图像中少数可见的树冠；图4中的低召回率和定性结果也体现了这一点。

尽管即使是我们表现最佳的模型，其指标仍有明显的提升空间，但结果表明，自动化训练标签收集方法与零样本分割模型增强相结合，使我们更接近于获得足够的数据，以训练具有广泛泛化能力的树冠实例分割模型。所有指标均处于以全监督数据训练的模型在先前文献中所观察到的范围内。对于一种仅使用自动收集的弱标签进行训练的方法而言，这是一个重大成就。

在图4中，我们展示了根据定量结果选出的、我们认为最值得重点呈现的模型输出。未展示的基线方法见附录中的图 A.2。可视化测试样本的选择旨在突出数据中存在的各种情形，完全依据输入和真实标注掩膜进行。为节省空间，可视化帧裁剪至原始 1024 像素宽高的 3/4；每个图块仅评估了中间 512×512 像素区域内的掩膜，因此可视化也仅显示该区域。

图4所示结果支持了定量分析结果。图中突显了 Detec-tree2 和 SAM 3 模型的预测似乎过于保守，而监督式U-Net模型则产生了更多假阳性和过大的分割区域。这与表1中给出的精确率和召回率数值相吻合。所提出的方法生成的结果更加均衡。从同一图中还可以观察到，仅使用 RGB 输入的所提模型与使用 RGB+近红外输入的模型相比，预测结果略有改善。

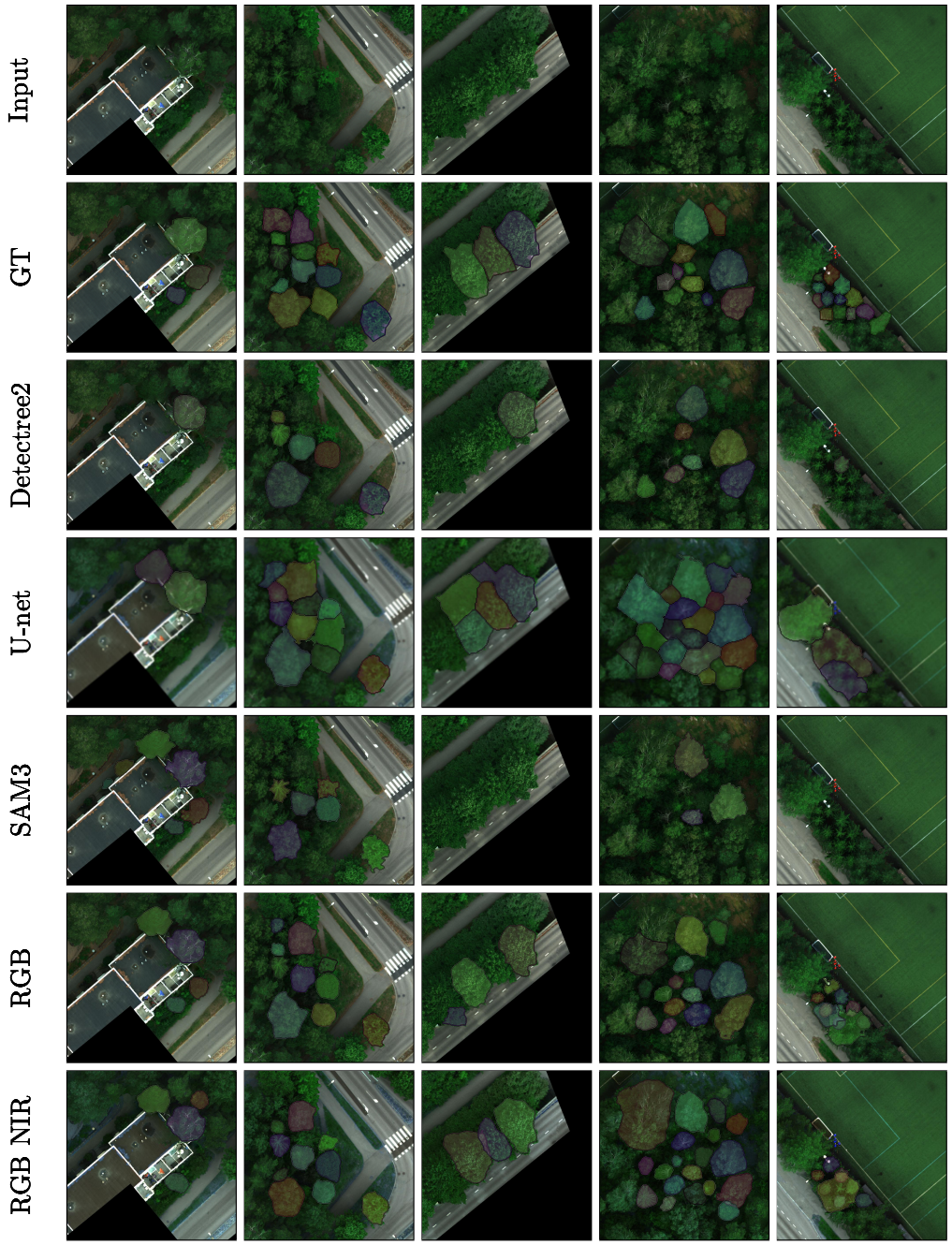

图4：所选模型输出结果的可视化。Detectree2 的结果来自性能最佳的 *Flexi* 检查点。

4.1. 消融

我们对 SAM 2 的标签平滑以及多光谱图像中不同通道的使用进行了消融实验，结果均列于表 2。如表 2 和图 5 所示，SAM 2 的标签增强全面提升了下游模型的性能。不同输入通道配置的效果则不太明确。值得注意的是，不同指标都凸显了近红外通道的作用，但遗憾的是，将其与其他输入通道结合使用的模型未能充分发挥其价值。对此应进一步研究改进方法，但这个问题也可能因数据而异，并与通道归一化等因素有关。

监督	输入	F1	精确率	召回率	mIoU	95%置信区间
粗略	RGB	0.743	0.784	0.706	0.548	[0.515, 0.579]
粗略	RE	0.775	**0.805**	0.747	0.576	[0.541, 0.607]
粗略	NIR	0.739	0.777	0.706	0.539	[0.505, 0.568]
粗略	RGB+红边	0.740	0.766	0.717	0.562	[0.528, 0.591]
粗略	RGB+近红外	0.746	0.762	0.731	0.565	[0.532，0.595]
粗略	RE+近红外	0.760	0.792	0.731	0.559	[0.525，0.590]
粗略	RGB+RE+近红外	0.748	0.786	0.714	0.549	[0.515, 0.579]
伪	RGB	0.758	0.783	0.733	0.599	[0.566, 0.632]
伪	RE	0.767	0.776	0.761	0.597	[0.565, 0.631]
伪	NIR	0.767	0.785	0.750	0.586	[0.554, 0.617]
伪	RGB+红边	0.774	0.793	0.756	0.610	[0.577, 0.643]
伪	RGB+近红外	**0.778**	0.783	**0.772**	**0.621**	[0.587, 0.652]
伪	红边+近红外	0.772	0.787	0.758	0.610	[0.576, 0.640]
伪	RGB+红边+近红外	0.761	0.771	0.750	0.602	[0.567, 0.635]

表 2：不同训练所得 Mask-RCNN 模型的消融结果。**95%置信区间** 是平均交并比的置信区间。

消融实验为 SAM 2 标签增强的有效性提供了确凿证据，并清楚表明，将 RGB 图像与近红外或红外通道融合很可能会带来更好的结果。这些结果在图 5 的箱线图中最为明显。可视化的置信区间表明，这种改进在不同输入模态下都很稳定。为了可靠地确定此任务的最佳输入通道配置，还应使用更多样的数据开展更多实验。不同输入模态和训练标签配置下的定性结果见附录中的图 6 和图 A.3。

4.2. 失效点

虽然我们的方法在测试数据集上的表现大幅优于基线方法，但我们也观察到了一些明显的失效点。在实际场景中直接部署这些方法，或基于本研究开发新模型时，都应考虑这些问题。模型的弱点包括对新数据集的可迁移性、图像边缘区域附近的预测，以及容易生成未经非极大值抑制筛除的重叠掩膜。

重叠预测。 所提出方法的主要弱点（在所使用的测试样本中可以观察到）是，模型倾向于生成这样的输出：某些掩膜完全包含在一个更大的预测掩膜中。对于大小相近的掩膜，这并不是问题，因为它们会在非极大值抑制后处理步骤中被移除；但在其中一个掩膜明显大于另一个的情况下，由于过滤是基于简单的交并比阈值进行的，非极大值抑制无法识别这种重叠。不过，在实际推理中，还可以加入另一种

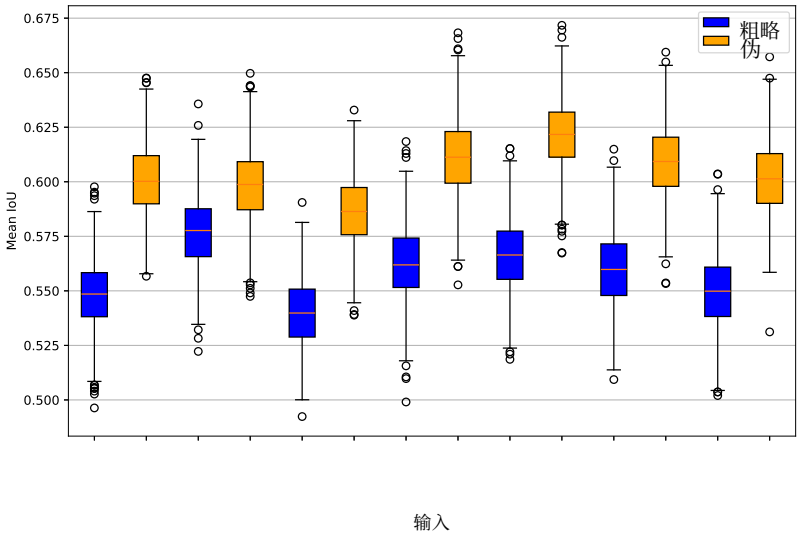

图 5：标准四分位数箱线图，展示不同输入模态以及粗略监督与伪监督之间的预测质量（平均交并比）。

过滤步骤，该步骤基于交集以及单个掩膜的面积进行。这个问题也凸显了数据集中树木尺寸差异悬殊所带来的难点。

可迁移性。 我们最终的基于 RGB 图像的方法应用于一些公开可用的树冠图像时会失效。这很可能是由于与其他可用数据相比，我们的数据集具有独特的空间分辨率和颜色。更强的数据增强或许可以缓解这一问题，但我们尚未对此进行进一步探索。

图像边缘附近的预测。 由于模型训练所用的标签仅出现在图像中间部分，因此图像边缘处的预测效果会变差。这意味着，在较大区域进行实际推理时，应将栅格切分为图像块，并使各维度上图像块之间的重叠部分占图像尺寸的 1/4。采用完整图像步幅进行评估所得的指标见表3。

输入	mIoU	变化	相对(%)
RGB	0.580	-0.020	-3.33
RE	**0.598**	+0.001	1.70
NIR	0.553	-0.035	-5.97
RGB + 红边	0.587	-0.024	-3.93
RGB + 近红外	0.594	-0.027	-4.35
红边 +近红外	0.584	-0.016	-2.62
RGB + 红边 + 近红外	0.579	-0.023	-3.82

表3：基于整幅图像预测评估模型时的结果（包括更靠近图像边缘的部分树冠）。所有结果均来自伪监督模型。

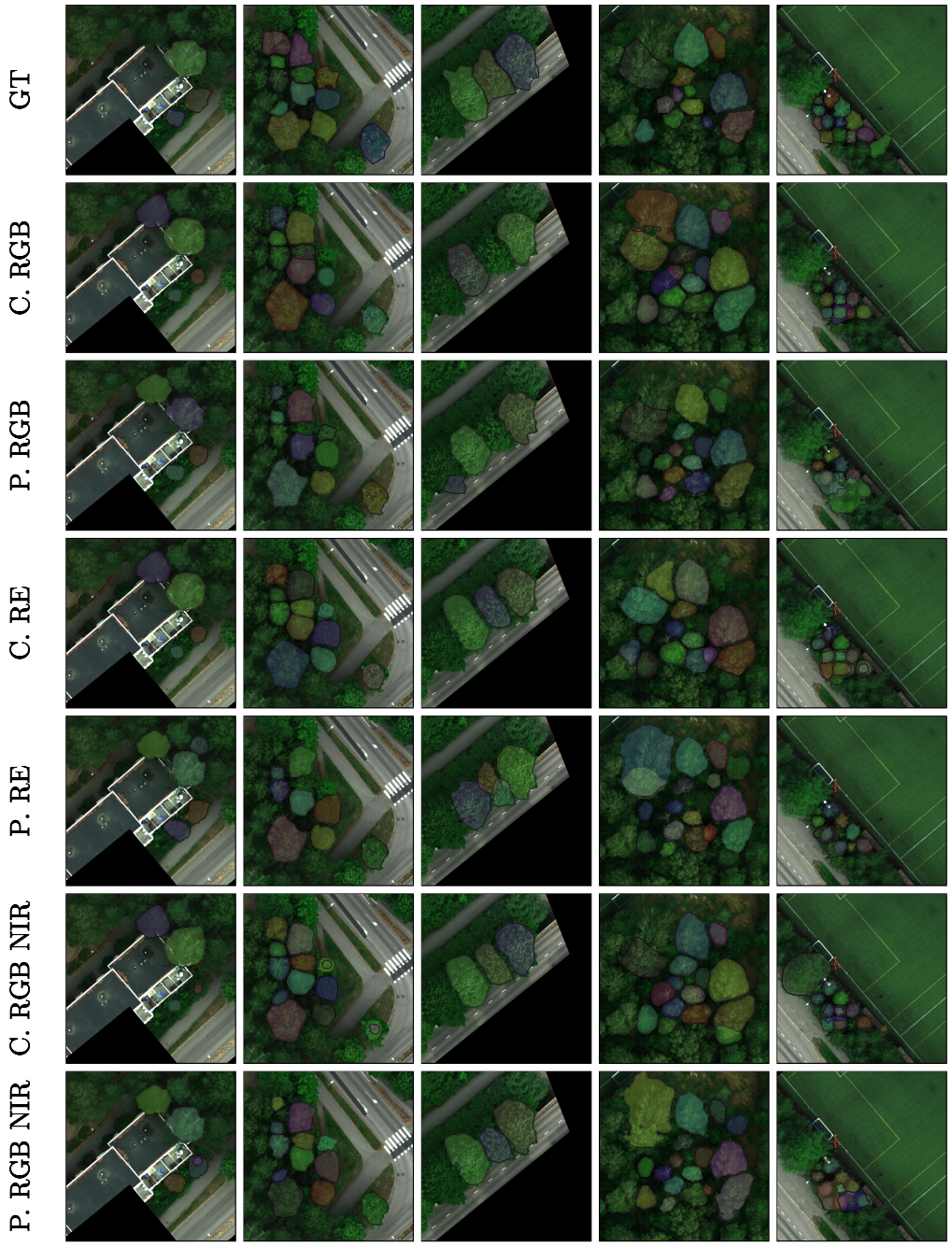

图 6：使用不同输入模态和训练标签的模型输出可视化结果。C. 是粗略（coarse）的缩写，P. 是伪标签（pseudo-labels）的缩写。

5. 结论

本研究表明，由冠层高度模型生成的标注可用于训练基于图像的树冠分割模型，并且将 SAM 2 用作预处理步骤能明显提升这些标签的质量。我们的方法提供了一种无需任何人工标注成本即可为基于光学图像的模型获取相对高质量训练标注的途径。这使得模型能够以较低成本针对特定地点进行训练，并优于在不同数据上针对同一任务训练的任何最先进模型。

遗憾的是，大多数基于相机的树冠预测研究更侧重于边界框检测，这一点也反映在可用数据的数量上。虽然我们提供了一种生成分割数据的替代方法，但我们缺乏来自更多样化地点的大量数据，而这些数据本可用于训练或测试更具泛化能力的模型。此外，尤其是在不同输入模态之间，所观察到的差异很小，并且相较于人工标注略优的分割指标并不一定与真实世界性能 [48]完全相关。

为了进一步利用这一方法，生成能够在任何特定场景中表现更好的树冠实例分割模型，我们认为应将这些数据与来自更多不同地理区域和时间段的类似数据结合起来。此外，还可以改进模型，例如将 Mask R-CNN 换成经验证更适用于树冠的模型，如 StarDist [49], ，仅采用更新的骨干架构，或利用经过更大规模预训练的模型。

致谢

我们感谢 Teemu Hakala 为本研究采集多光谱图像数据准备传感器系统。我们还要感谢芬兰 CSC – IT科学中心提供计算资源

经费来源

本研究由芬兰研究理事会通过“基于自主无人机的森林植被高光谱分析学习技术”项目（决议编号 357380）和芬兰研究理事会研究基础设施“森林生态系统时空变化测量”（346382）资助；此外，还通过芬兰研究理事会获得欧盟“下一代欧盟”工具的资助（多重风险项目资助编号 353263）。本研究隶属于芬兰研究理事会旗舰项目“森林—人类—机器交互——构建韧性、重塑价值网络并创造有意义的体验”（UNITE）（决议编号 357908）。

参考文献

[1] Y. Lai、C. E. Kontokosta，《城市行道树树种对空气质量和呼吸系统疾病的影响：基于大规模高分辨率城市数据的空间分析》，《健康与场所》56（2019）80–87。doi:https://doi.org/10.1016/j.healthplace.2019.01.016。[2] C. Zhao、H. A. Sander，《量化并绘制城市树木碳储量与碳封存服务的供需分布》，《公共科学图书馆·综合》10（8）（2015）e0136392。doi:https://doi.org/10.1371/journal.pone.0136392。

[3]T. Yang、Y. Ryu、R. Kwon、C. Choi、Z. Zhong、Y. Nam、S. Jo，《基于激光雷达与相机融合的移动测绘系统：绘制行道树单木碳储量及其增长情况》，Remote Sensing of Environment 328（2025）114895。doi:https://doi.org/10.1016/j.rse.2025.114895。[4] K. D. Brosofske、R. E. Froese、M. J. Falkowski、A. Banskota，《森林经营管理中清查属性制图与预测方法综述》，森林科学 60（4）（2013）733–756。doi:https://doi.org/10.5849/forsci.12-134。[5] T. Hlásny、L. König、P. Krokene、M. Lindner、C. Montagné-Huck、J. Müller、H. Qin、K. F. Raffa、M.-J. Schelhaas、M. Svoboda 等，《欧洲树皮甲虫暴发：现状与管理对策》，Current Forestry Reports 7（3）（2021）138–165。doi:https://doi.org/10.1007/s40725-021-00142-x。[6] M. Patacca、M. Lindner、M. E. Lucas-Borja、T. Cordonnier、G. Fidej、B. Gardiner、Y. Hauf、G. Jasinevičius、S. Labonne、E. Linkevičius 等，《自1950年以来欧洲森林自然干扰的影响显著增加》，全球变化生物学 29（5）（2023）1359–1376。doi:https://doi.org/10.1111/gcb.16531。[7] W. You、L. Lin、L. Wu、Z. Ji、J. Yu、J. Zhu、Y. Fan、D. He，《基于地理信息系统、整合国家森林清查数据的森林火灾风险评估及其时空变异分析》，生态指标 77（2017）176–184。doi:https://doi.org/10.1016/j.ecolind.2017.01.042。[8] 美国农业部林务局，《城市森林清查与分析计划》，网址：https://research.fs.usda.gov/programs/urbanfia，（访问日期：2026年1月12日）（2025）。[9] J. Zhou、C. Proisy、X. Descombes、G. Le Maire、Y. Nouvellon、J.-L. Stape、G. Viennois、J. Zerubia、P. Couteron，《利用高空间分辨率卫星影像进行单木检测，绘制幼龄桉树人工林的局部密度》，Forest Ecology and Management 301（2013）129–141。doi:https://doi.org/10.1016/j.foreco.2012.10.007。[10] S. Zhang、Y. Cui、Y. Zhou、J. Dong、W. Li、B. Liu、J. Dong，《利用高分辨率卫星影像绘制中国柳州桉树人工林树冠及单木分布》，IEEE地球科学与遥感汇刊 61（2023）1–13。doi:https://doi.org/10.1109/TGRS.2023.3327128。[11]J. Zheng、S. Yuan、W. Li、H. Fu、L. Yu、J. Huang，《基于光学遥感影像的单木树冠检测与 delineation 综述：当前进展与未来方向》，IEEE地球科学与遥感杂志（2024）。doi:https://doi.org/10.1109/MGRS.2024.3479871。[12] M. Favorskaya、A. Tkacheva、I. M. Danilin、E. M. Medvedev，《融合机载激光雷达与数码摄影数据进行树冠分割和测量》，载于：智能交互式多媒体系统与服务，Springer，2015，第191–201页。doi:https://doi.org/10.1007/978-3-319-19830-9_18。[13] M. Aubry-Kientz、A. Laybros、B. Weinstein、J. G. Ball、T. Jackson、D. Coomes、G. Vincent，《多传感器数据融合提升稠密热带森林中的单木树冠分割效果》，IEEE应用地球观测与遥感专题期刊 14（2021）3927–3936。doi:https://doi.org/10.1109/JSTARS.2021.3069159。

[14] N. Ravi、V. Gabeur、Y.-T. Hu、R. Hu、C. Ryali、T. Ma、H. Khedr、R. Rädle、C. Rolland、L. Gustafson、E. Mintun、J. Pan、K. V. Alwala、N. Carion、C.-Y. Wu、R. Girshick、P. Dollar、C. Feichtenhofer，SAM 2：图像和视频中的任意分割，载于：《第十三届国际学习表征会议》，2025年。网址：https://openreview.net/forum?id=Ha6RTeWMd0

[15] S. Rua，利用SAM2伪标签对航空影像中的树冠进行实例分割 -，硕士论文，阿尔托大学（2025年）。网址：https://urn.fi/URN:NBN:fi:aalto-202512179400

[16] K. Wołk、M. S. Tatara，利用激光雷达和影像数据进行林业语义分割与实例分割技术综述，《Electronics》13（20）（2024）4139。doi：https://doi.org/10.3390/electronics13204139。

[17] J. G. Ball、S. H. Hickman、T. D. Jackson、X. J. Koay、J. Hirst、W. Jay、M. Archer、M. Aubry- Kientz、G. Vincent、D. A. Coomes，《利用 Mask R-CNN 从航空 RGB 影像中准确 delineation 热带森林中的单木树冠》，《生态与保护遥感》9 (5) (2023) 641–655。doi:https://doi.org/10.1002/rse2.332.

[18] M. Freudenberg、P. Magdon、N. Nölke，基于 U-Net 的高分辨率遥感图像单木树冠勾绘，NCAA（2022）。doi:https://doi.org/10.1007/s00521- 022-07640-4。[19] Y. Lin、H. Li、L. Jing、H. Ding、S. Tian，利用机载激光雷达数据和航空影像在针叶林—苔原交错带进行单木树冠勾绘，Remote Sensing 16（21）（2024）3920。doi:https://doi.org/10.3390/rs16213920。[20]S. Dersch、A. Schöttl、P. Krzystek、M. Heurich，利用航空多光谱影像和激光雷达数据进行半监督多类别树冠勾绘，ISPRS 摄影测量与遥感杂志 216（2024）154–167。doi:https://doi.org/10.1016/j.isprsjprs.2024.07.032。[21]A. Kirillov、E. Mintun、N. Ravi、H. Mao、C. Rolland、L. Gustafson、T. Xiao、S. Whitehead、A. C. Berg、W.-Y. Lo、P. Dollár、R. Girshick，任意分割，arXiv:2304。02643[预印本]（2023）。doi:https://doi.org/10.48550/arXiv.2304.02643。[22] N. Carion、L. Gustafson、Y.-T. Hu、S. Debnath、R. Hu、D. Suris、C. Ryali、K. V. Alwala、H. Khedr、A. Huang、J. Lei、T. Ma、B. Guo、A. Kalla、M. Marks、J. Greer、M. Wang、P. Sun、R. Rädle、T. Afouras、E. Mavroudi、K. Xu、T.-H. Wu、Y. Zhou、L. Momeni、R. Hazra、S. Ding、S. Vaze、F. Porcher、F. Li、S. Li、A. Kamath、H. K. Cheng、P. Dollár、N. Ravi、K. Saenko、P. Zhang、C. Feichtenhofer，SAM 3：利用概念进行任意分割，[预印本]（2025）。arXiv:2511.16719，doi:https://doi.org/10.48550/arXiv.2511.16719。[23] K. Chen、C. Liu、H. Chen、H. Zhang、W. Li、Z. Zou、Z. Shi，RSPrompter：基于视觉基础模型学习用于遥感实例分割的提示方法，IEEE地球科学与遥感汇刊 62（2024）1–17。doi:https://doi.org/10.1109/TGRS.2024.3356074。[24] M. Teng、A. Ouaknine、E. Laliberté、Y. Bengio、D. Rolnick、H. Larochelle，评估 SAM 在无人机影像树冠实例分割中的表现，arXiv:2503。20199[预印本]（2025）。doi:https://doi.org/10.48550/arXiv.2503.20199。

[25] B. G. Weinstein、S. Marconi、M. Aubry-Kientz、G. Vincent、H. Senyondo、E. P. White，DeepForest：用于 RGB 深度学习树冠勾绘的 Python 软件包，Methods in Ecology and Evolution 11（12）（2020）1743–1751。doi:https://doi.org/10.1111/2041- 210X.13472。[26]M. Chen、D. Russell、A. Pallavoor、D. Young、J. Wu，基于森林航空影像的零样本单木检测与分割，arXiv:2506。03114[预印本]（2025）。doi:https://doi.org/10.48550/arXiv.2506.03114。[27] T. Ren、S. Liu、A. Zeng、J. Lin、K. Li、H. Cao、J. Chen、X. Huang、Y. Chen、F. Yan、Z. Zeng、H. Zhang、F. Li、J. Yang、H. Li、Q. Jiang、L. Zhang，Grounded SAM：组装开放世界模型以完成多样化视觉任务， [预印本]（2024）。arXiv:2401.14159, doi:https://doi.org/10.48550/arXiv.2401.14159。[28] S. Liu、Z. Zeng、T. Ren、F. Li、H. Zhang、J. Yang、C. Li、J. Yang、H. Su、J. Zhu等，Grounding DINO：将 DINO 与 grounded 预训练相结合以进行开放集目标检测，arXiv:2303。05499[预印本]（2023）。doi:https://doi.org/10.48550/arXiv.2303.05499。[29] E. Arnaudo、J. L. Vaschetti、L. Innocenti、L. Barco、D. Lisi、V. Fissore、C. Rossi，Fmars：利用基础模型为灾害管理标注遥感图像，载于：IGARSS 2024-2024 IEEE 国际地球科学与遥感研讨会，IEEE，2024，第3920–3924页。doi:https://doi.org/10.1109/IGARSS53475.2024.10641130。[30] Y. Zhu、W. Locke、J. Yuan、Y. Zhang、Q. Ma、L. Liang，利用 SAM 2 和激光雷达实现自动单木树冠勾绘：提示方法的比较评估，Information Geography（2025）100025doi:https://doi.org/10.1016/j.infgeo.2025.100025。[31] Agisoft，Agisoft Metashape Professional，（v.2.0.2）。 [软件], https://www.agisoft.com/。[32] L. Matikainen、K. Karila、J. Hyyppä、P. Litkey、E. Puttonen、E. Ahokas，面向土地覆盖分类与地图更新的多光谱机载激光扫描仪数据面向对象分析，ISPRS 摄影测量与遥感杂志 128（2017）298–313。doi:https://doi.org/10.1016/j.isprsjprs.2017.04.005。[33]J. Taher、E. Hyyppä、M. Hyyppä、K. Salolahti、X. Yu、L. Matikainen、A. Kukko、M. Lehtomäki、H. Kaartinen、S. Thurachen等，多光谱机载激光扫描用于树种分类：机器学习与深度学习算法的基准测试，ISPRS 摄影测量与遥感杂志 233（2026）278–309。doi:https://doi.org/10.1016/j.isprsjprs.2026.01.031。[34] rapidlasso GmbH，LAStools 软件套件， [软件],https://rapidlasso.de/。[35] H. Kaartinen、J. Hyyppä、X. Yu、M. Vastaranta、H. Hyyppä、A. Kukko、M. Holopainen、C. Heipke、M. Hirschmugl、F. Morsdorf等，利用机载激光扫描进行单木检测与提取的国际比较，Remote Sensing 4（4）（2012）950–974。doi:https://doi.org/10.3390/rs4040950。[36]K. Karila、L. Matikainen、P. Litkey、J. Hyyppä、E. Puttonen，季节变化对基于多光谱机载激光扫描数据的自动土地覆盖制图的影响，International Journal of Remote Sensing 40（9）（2019）3289–3307。doi:https://doi.org/10.1080/01431161.2018.1528023。

[37]K. He, G. Gkioxari, P. Dollár, R. Girshick，Mask R-CNN，载于：IEEE国际计算机视觉会议论文集，2017年，第2961–2969页。doi:https://doi.org/10.1109/ICCV.2017.322。[38]Y. Li, S. Xie, X. Chen, P. Dollar, K. He, R. Girshick，使用视觉变换器对检测迁移学习进行基准测试，arXiv:2111。11429[预印本]（2021）。doi:https://doi.org/10.48550/arXiv.2111.11429。[39] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei，ImageNet：大规模分层图像数据库，载于：2009年IEEE计算机视觉与模式识别会议，Ieee，2009年，第248–255页。doi:https://doi.org/10.1109/CVPR.2009.5206848。[40] K. He, X. Zhang, S. Ren, J. Sun，用于图像识别的深度残差学习，载于：IEEE计算机视觉与模式识别会议论文集，2016年，第770–778页。doi:https://doi.ieeecomputersociety.org/10.1109/CVPR.2016.90。[41] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie，用于目标检测的特征金字塔网络，载于：IEEE计算机视觉与模式识别会议论文集，2017年，第2117–2125页。doi:https://doi.org/10.1109/CVPR.2017.106。[42] D. P. Kingma，Adam：一种随机优化方法，arXiv:1412。6980[预印本]（2014）。doi:https://doi.org/10.48550/arXiv.1412.6980。[43] S. Hanson, L. Pratt，比较用于最小网络构建的反向传播偏差，载于：D. Touretzky（编），《神经信息处理系统进展》，第1卷，Morgan-Kaufmann，1988年。[44] I. Loshchilov, F. Hutter，解耦权重衰减正则化，载于：国际学习表征会议，2019年。URL https://openreview.net/forum?id=Bkg6RiCqY7[45] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg，SSD：单次多框检测器，载于：欧洲计算机视觉会议，Springer，2016年，第21–37页。doi:https://doi.org/10.1007/978-3-319-46448-0_2.[46] J. Ball, C. Kotthoff, Detectree2训练模型（2025）。doi:10.5281/zenodo.15863800。URL https://doi.org/10.5281/zenodo.15863800[47] J. G. C. Ball, S. Jaffer, A. Laybros, C. Prieur, T. Jackson, A. Madhavapeddy, N. Barbier, G. Vincent, D. A. Coomes，通过利用时间和光谱维度，实现多样化热带森林中全面的单株树种制图，bioRxiv[预印本]（2025）。doi:https://doi.org/10.1101/2024.06.24.600405。[48] T. Khan, J. Krebs, S. K. Gupta, J. Renkel, C. Arnold, N. Nölke，利用深度学习从遥感影像中大规模分割树冠的验证挑战：德国案例研究，载于：数字图书馆理论与实践国际会议，Springer，2025年，第311–323页。doi:https://doi.org/10.1007/978-3-032-06136-2_30.[49] F. Tong, Y. Zhang，使用基于StarDist的模型对高分辨率航空RGB影像进行单木树冠勾绘，Remote Sensing of Environment 319（2025）114618。doi:https://doi.org/10.1016/j.rse.2025.114618。

附录

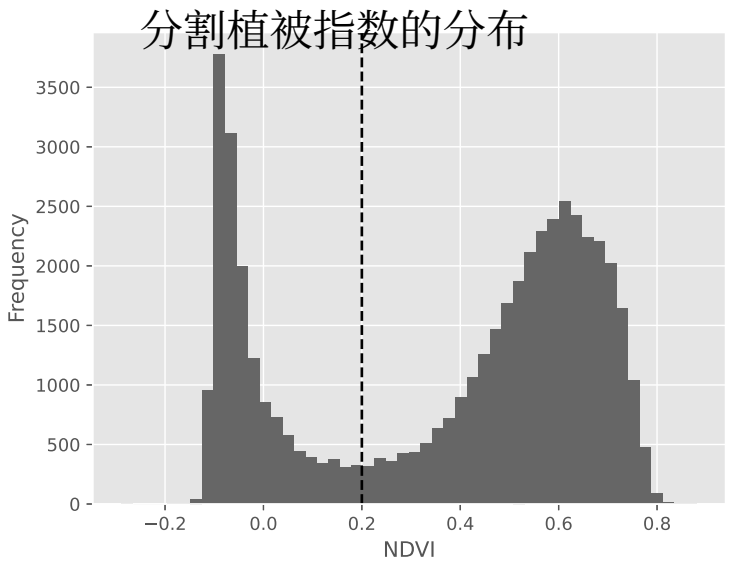

图 A.1：初始树冠分割片段的归一化植被指数分布，黑色竖向虚线表示截断阈值。

Detectree2 检查点	F1	精确率	召回率	mIoU	平均交并比 95%置信区间
base	0.298	0.540	0.206	0.182	[0.150, 0.215]
*05*日期	0.423	0.606	0.325	0.261	[0.227, 0.298]
flexi	**0.556**	**0.911**	**0.400**	**0.324**	[0.281, 0.364]
随机缩放_完整	0.447	0.718	0.325	0.269	[0.228, 0.308]
热带_郁闭_冠层	0.064	0.177	0.039	0.039	[0.025, 0.056]
城市_树木_剑桥	0.393	0.600	0.292	0.243	[0.211, 0.276]
withParacouUAV	0.415	0.543	0.336	0.280	[0.242, 0.319]

表 A.1：所有不同 Detectree2 模型检查点的测试集指标。

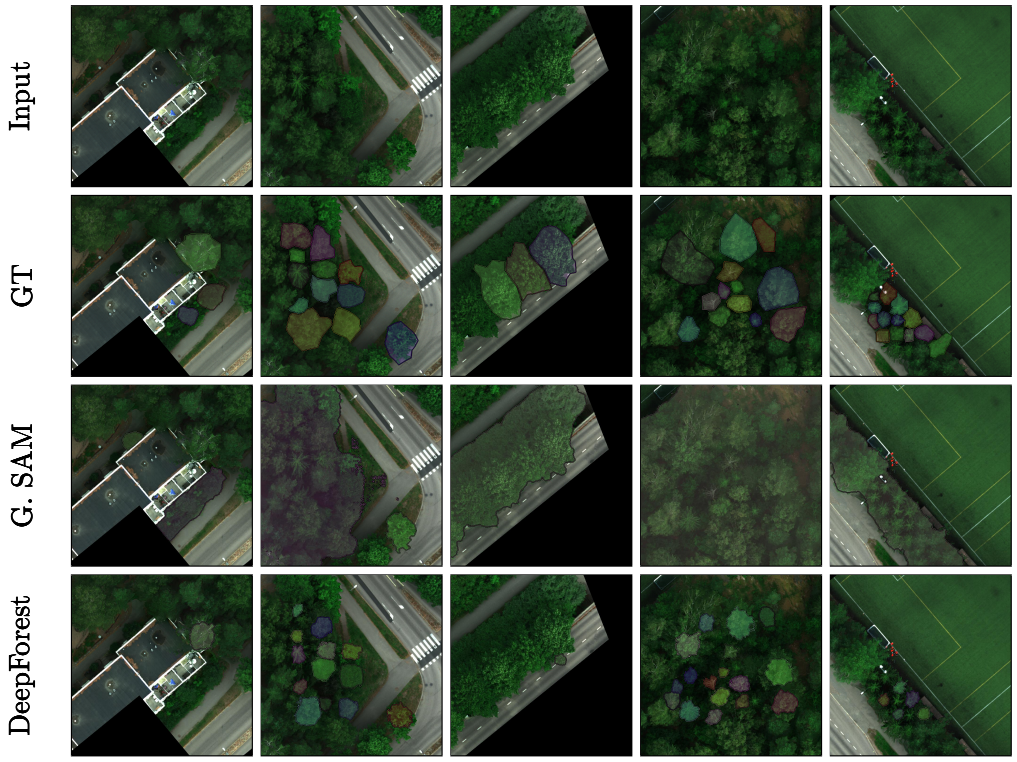

图 A.2：Grounded SAM（记作 G. SAM）和 DeepForest 模型的输出可视化。

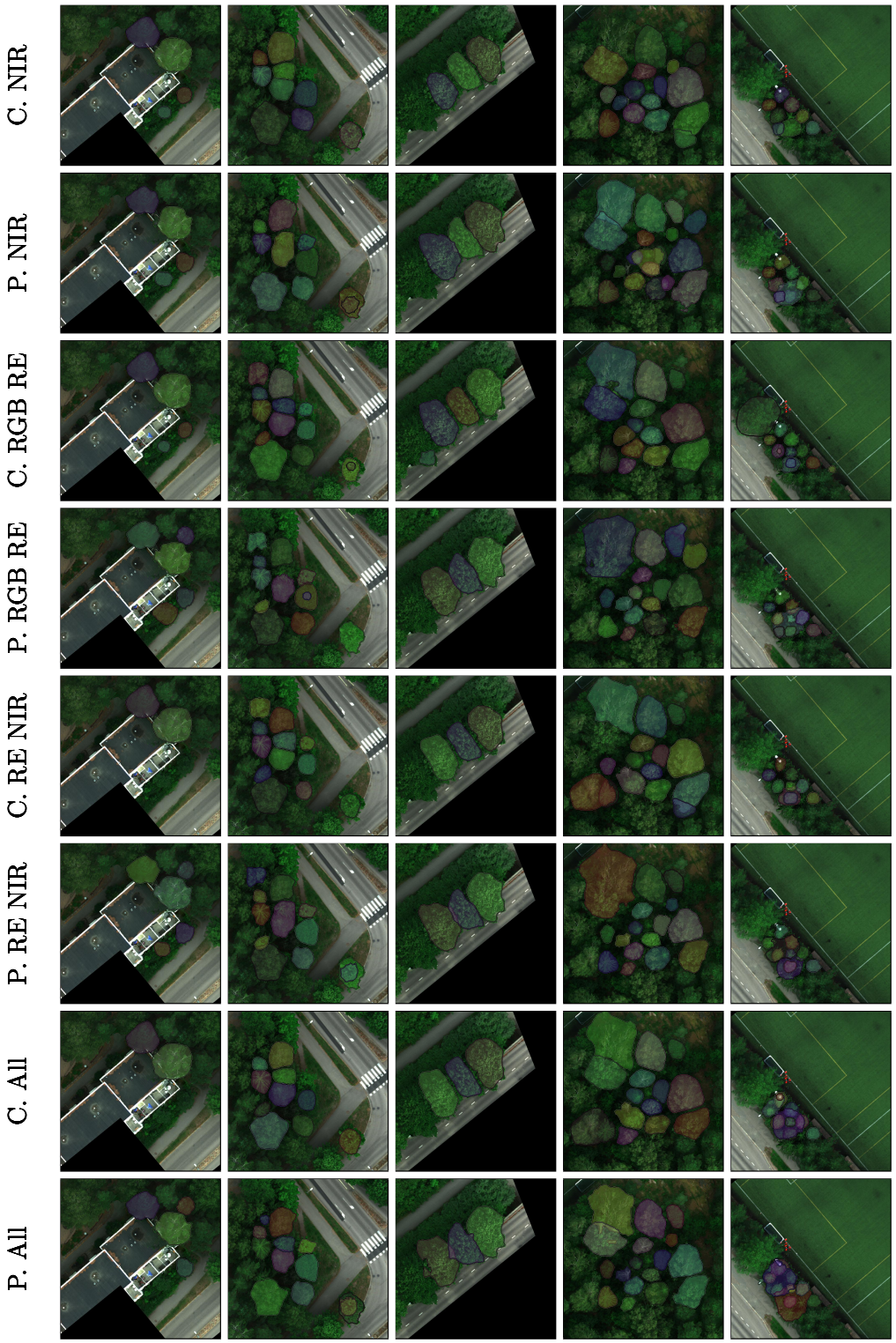

图 A.3：使用不同输入模态和训练标签的模型输出可视化。C. 分别是粗略的缩写，P. 是伪标签的缩写。**所有**指使用全部三种输入模态（RGB、红边和近红外）的模型。