

弱监督学习在机载激光雷达点云中的树实例分割

Swann Emilien Céleste Destouches¹, Jesse Lahaye¹, Laurent V. Jospin¹, Jan Skaloud¹

¹ Environmental Sensing & Observation Laboratory (ESO), Ecole Polytechnique Fédérale de Lausanne (EPFL), Swann.Destouches@alumni.epfl.ch, (jesse.lahaye, laurent.jospin, jan.skaloud)@epfl.ch

Keywords: ALS, Lidar, 树实例分割, 迁移学习, 远程 sensing

Abstract

机载激光扫描 (ALS) 数据中的树木实例分割对于森林监测极为重要,但由于传感器分辨率、采集时的植被状态、地形特征等因素导致的数据变化,这一过程仍然具有挑战性。此外,获得足够数量的精确标注数据来训练完全监督的实例分割方法成本高昂。为了解决这些挑战,我们提出了一种弱监督方法,其中由未经微调的模型或封闭形式算法得出的初始分割结果的标签将由人工操作员提供质量评级。在质量评估过程中生成的标签随后用于训练一个评分模型,该模型的任务是按照人类操作员规定的类别对分割输出进行分类。最后,通过评分模型提供的反馈来调整分割模型。这反过来使原始的分割模型正确识别树木实例的能力提高了 34%,同时显著减少了预测出的非树木实例的数量。在稀疏森林区域或复杂环境中(含有灌木、巨石等)仍然存在挑战,在这些情况下,低矮树木(高度不足两米)或可能被误认为树木的对象会导致所提出方法的表现降低。

1. 介绍

量化森林中的树木分布是三维视觉的重要应用,可以用来估算其经济潜力(例如,可利用木材体积)(Ma et al., 2020),测量气候变化的影响(Fouqueray et al., 2020),监测对斜坡稳定性的冲击(Jiang et al., 2023),或评估碳捕获的潜力(Dalponte and Coomes, 2016)。

个体树冠分割 (ITCs) 在通用 3D 森林数据中仍是一项具有挑战性的任务。虽然深度学习模型在某种程度上已经改善了这一点,但它们仍然难以应对数据分辨率、物种、季节变化(有叶/无叶)和地形中存在的伪影等变异性问题,特别是在山地地区,悬崖和巨石可能会干扰分割。尽管最近在泛化方面有所改进(Wielgosz et al., 2024),微调仍然是必要之举,这给现有方法的普遍适用性带来了挑战。为分割任务生成手动标签费时且容易出错,无论是语义(Shimoda and Yanai, 2019)分割还是实例(Cheng et al., 2023)分割都是如此。另一方面,深度学习方

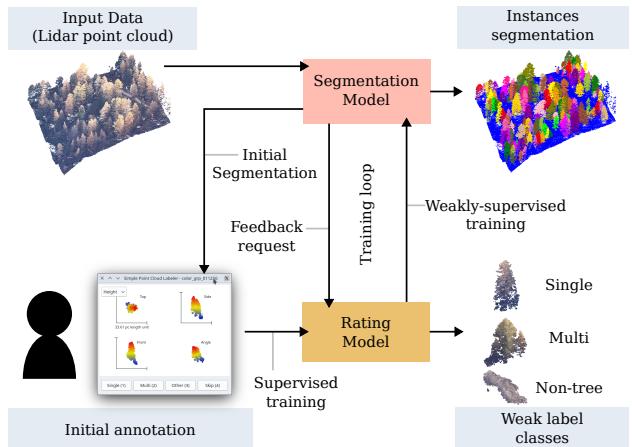


Figure 1. 提出的弱监督训练树分割方法。

法需要访问高质量的标注数据来进行微调,以便应用于新的目标区域(例如山地生态过渡带例如),这仍然是一个问题(Fan et al., 2024)。

为了解决这些挑战,我们提出了一种受人类反馈强化学习(RLHF(Chaudhari et al., 2025))启发的方法。我们不对数据进行分割掩码标记,而是对分割结果中的样本提供反馈,这由人工操作员生成的速

度要快得多。然后我们训练一个分类模型（评分模型）来模仿人工操作员的评分，并反过来为分割模型提供反馈；参见图 1。虽然这种方法在技术上可以推广到任何分割任务，但对于那些现有算法能够产生足够质量的初始结果以启动训练过程的任务特别有用，这种情况在树实例分割中尤为常见。

总之我们的贡献是：

- 我们建议在训练分割模型时使用评分模型作为“弱监督”。我们的分类器可以获得约 90% 的准确率，该准确率可以通过人工操作员通过数据检查合理快速建立的分割质量评分标签获得。
- 我们通过在一个具有挑战性的高山数据集中，将分割的单棵树实例数量增加 34% 来验证所提出的方法，并且不对应实际树木的实例比例显著下降。
- 我们评估了联合分割和评分模型在数据集的手动标记子集上的性能。我们确认我们的方法成功地将研究区域内密集森林中 80% 到 90% 的树木进行了分割，与现有技术水平相比有了显著的性能提升。

2. 相关工作

2.1 点云分类

点分类模型可以大致分为两类。第一类将数据转换为规则结构，如 2D 图像网格 (Rizaldy et al., 2018)、3D 体素表示 (Maturana and Scherer, 2015) 或切片 (Huang et al., 2018)，在进一步处理之前。规则结构可以通过经典架构进行处理，例如 2D 卷积神经网络 (Rizaldy et al., 2018)、3D 卷积神经网络 (Maturana and Scherer, 2015)、RNN (Huang et al., 2018) 等。虽然 CNN 已被证明在图像分类任务中易于使用且可靠，但将其应用于 3D 点云并非没有额外的不便之处。实际上，在 3D 数据中，CNN 要么强制使用体素网格，这在高分辨率时可能内存密集或在低分辨率时导致信息丢失；或者将 3D 数据投影到 2D 网格上，从而进一步从 3D 结构 (Qi et al., 2017a) 中移除信息。然而，这些限制可以通过使用稀疏体素网格来在一定程度上解决，后者具有线性处理和内存复杂度 (Chen et al., 2023)。

另一种方法是将激光雷达点云作为不规则数据结构进行处理。这一想法的先驱是 PointNet 架构，它单独处理每个点，并使用一种排序不变函数（例如最大池化或平均池化）来提取特征 (Qi et al., 2017a)。这种方法的主要缺点在于难以捕捉局部细节，同时关注重要的结构点 (Qi et al., 2017a)。为了部分缓解这种不良效果，PointNet++ 对其输入采用了嵌套的局部划分 (Qi et al., 2017b)。自我注意力提供了一种处理对象无结构集合的有效方法，在诸如 Point Transformer (Zhao et al., 2021) 等架构中实现了这一点。

在为提议的学习系统选择分类架构时，不同标准的重要性包括但不限于准确性、训练的简便性和计算速度。我们将评估三种架构：(i) 3DmFV (Ben-Shabat et al., 2018)，一种密集的体素网络，通过使用 3D 修改后的费舍尔向量编码丰富的局部特征来缓解较低的体素网格分辨率；(ii) Point Transformer (Zhao et al., 2021)，作为一种最先进的无结构模型；以及 (iii) 我们从 VoxNet (Maturana and Scherer, 2015) 衍生出的一种更高分辨率的 3D CNN。

2.2 单木分割

闭式点云分割解决方案，如分水岭 (Yang et al., 2020)、图割 (Lee et al., 2016) 或区域生长 (Ma et al., 2020)，似乎正逐渐被更高性能的深度学习模型所取代。然而，它们仍然提供了合理的准确性水平，并因此作为基线模型而保持流行，部分原因也在于其对个体树检测 (ITD) 和单个树冠分割 (ITC) 任务的一般适用性。

不同的深度学习架构已被适配到 ITD 和/或 ITC 任务中，包括 PointNet (Chen et al., 2021)、PointNet++ (Liu et al., 2023) 以及 2D CNN 版本如 YOLOv5 (Straker et al., 2023) 或 Mask R-CNN (Fan et al., 2024)（对于后者，点云数据被投影到栅格深度图）或 3D UNET (Xiang et al., 2023)。

尽管计算速度有所提升，处理 ITD 和 ITCs 的非常大的户外数据集仍然具有挑战性。为了解决这一问题，Xiang 等人。 (Xiang et al., 2023) 提出了一个使用自下而上实例分组的架构。与自顶向下实例检测不同，在自顶向下实例检测中需要先检测到对象然后再进行分割，并且平铺可能会将对象分开，自

下而上的实例分组为云中的每个点分配一个特征向量，这些特征向量随后使用无监督方法进行分割。自下而上实例分组的一个主要缺点是，无监督的分割步骤可能非常不可靠 (Jiang et al., 2020)。为了缓解这一问题，模型基于 PointGroup，在其中使用了多种聚类变体以生成多个候选聚类。这些候选聚类随后通过 ScoreNet 进行筛选，整个过程提高了分割步骤的可靠性。

相等人。的分割方法后来与专门针对森林的几何过滤和数据增强相结合，创造了 ForAINet (Xiang et al., 2024)。这种全景分割模型（即输出单个树分割和点云语义分割的模型）专用于林业。该模型随后通过添加多分辨率数据来提高其泛化能力，并成为了任意分割树模型 (Wielgosz et al., 2024)。

ForestFormer3D 模型由 (Xiang et al., 2025) 提出，作为 ForAINet (Xiang et al., 2024) 和 SegmentAnyTree (Wielgosz et al., 2024) 的替代品。尽管所有这些发展仍然依赖于 3D UNet 进行编码，ForestFormer3D 将无监督分割步骤替换为基于变压器的层，使模型能够端到端完全训练。然而，在撰写本文时，其代码库尚未公开发布且该方法未经过同行评审；因此在本工作中未对其进行评估。

尽管第 3 节中详细描述的方法可以适应使用其他（或任何）骨干分割架构，我们选择了任意分割树 (Wielgosz et al., 2024) 作为当前最先进的方法。这是因为它是最新的同行评审方法，并且通过数据增强进行训练，这应该使其能够泛化到具有不同分辨率的数据集上。这两个方面对于第 4 节及后续结论（第 5 节）中的比较来说都很有趣。

3. 方法

评估分割算法的输出比手动标记（树）集群的确切轮廓快一个数量级。但是，如何利用评分中编码的信息最大限度地改进分割任务？我们建议通过图 2 所示的迭代过程来实现这一目标，具体描述如下：

- 首先，我们构建数据的初始分割，使用预训练的分割模型版本（只要其性能足够准确）或闭式解，如，例如，分水岭 (Yang et al., 2020)。
- 人类操作员将对步骤 1 中获得的初始聚类样本进行评分。对于 ITCs 任务，提出的聚类被评定

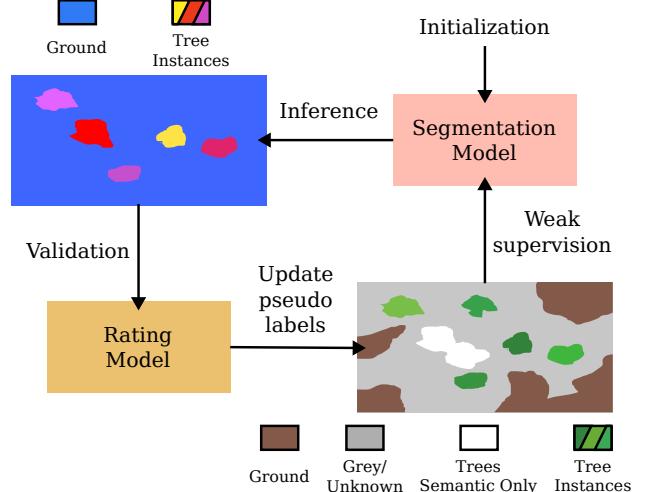


Figure 2. 训练循环：知识通过一个交互过程从评分模型转移到分割模型，在这个过程中，分割模型的输出被评分并用于构建伪标签。

为属于三种伪标签树类别之一（单个、多或非树），第 3.1 节。

- 然后，训练了一个评分模型，即，基于人类操作员标注的“真实值”评分来预测一个聚类的评分分类模型（第 3.1 节中的评分模型之一）。
- 随后将评级模型用于对初始分割输出中预测的所有聚类进行评级。在此基础上，构建伪标签图。
- 之前获得的伪标签用于训练分割模型（第 3.2 节）。
- 更新的分割预测被输入到评分模型中，其输出用于更新伪标签（第 3.3 节）。从这里开始，我们迭代回到步骤 5，直到新识别出的树实例数量稳定。

评估工具在补充材料中有更详细的描述。本节的其余部分将重点介绍评分和分段模型。

3.1 监督评级模型

为了简化评分过程，我们没有赋予一个分数（这会有点主观，并且需要操作员花费更多时间来思考），而是将评分问题转化为分类问题。每个聚类被分配到一个类别，当它们是单一树木时为单个，当同一聚类中包含多棵树木时为多，或者当其他非树木元素被聚类时，或至少在这些聚类中的树木不是主导元素时为非树结构。负责初步评分的人类操作员只需要几秒钟的时间就可以对大多数聚类做出决定。

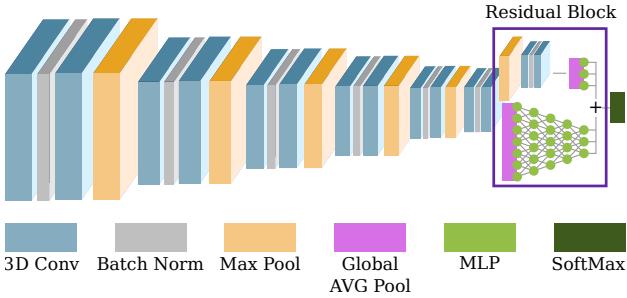


Figure 3. 提议的评分模型架构

对于自动评分模型, 如前所述, 评估了三种不同的架构。使用了“原样”的 3DmFV(Ben-Shabat et al., 2018) 和 Point Transformer(Zhao et al., 2021) 架构, 仅将最后一层调整为匹配三个评分类别。

我们还提出了一种 VoxNet (Maturana and Scherer, 2015) 架构, 如图 3 所示, 该架构用较高的分辨率代替了 3DmFV (Ben-Shabat et al., 2018) 中使用的 Fisher Vectors 的丰富特征。我们使用 Kernel Density Estimation (KDE) 来构建体素网格, 而不是简单的占用函数 (Maturana and Scherer, 2015), 这在牺牲略微稀疏的体素网格的情况下, 提供了对局部点密度更精确的估计。这一点本身就是一个重要的特征, 可以被分类器 (Hu et al., 2022)(例如, 考虑到点与传感器之间的距离, 这将影响点密度) 使用。

给定一个点云 P , KDE 计算体素网格 V 中体素 i 的值为:

$$V[i] = \sum_{p \in P} k(p - i), \quad (1)$$

使用 k 作为核函数。我们使用了:

$$k(x) = \frac{1}{\sqrt{(2\pi)^3}} \exp\left(-\frac{1}{2} \langle x, x \rangle\right), \quad (2)$$

是一个方差为单位的高斯核。

体素网格随后通过一系列 3D 卷积和最大池化层进行处理, 分辨率减半但特征数量增加。在 1/32 分辨率下的最后一层使用残差块聚合所有特征。残差块的第一个连接使用两个卷积层将体素网格减少到 1/64 分辨率下的 3 个特征, 并通过全局平均池化进行聚合。残差块的第二个连接直接在 1/32 分辨率下

通过全局平均池化聚合 1024 个特征, 然后使用多层感知机处理这些特征。两个分支的特征相加, 最终激活值通过 Softmax 层获得。

训练使用 ADAM 优化器进行。我们使用批归一化和权重衰减来进行规范化。对于所有三个模型, 我们也使用了数据增强以通过沿 z 轴添加随机旋转来限制过拟合的风险。分类损失是交叉熵损失, 加权使得每个类别的总权重等同 (因为某些类别, 特别是非树, 在初始数据中可能被过度表示), 即。, 权重 w 的计算方式如下:

$$w_i c_i = w_j c_j, \quad (3)$$

对于任何类 i 和 j , 其中 c_i 是类 i 中样本的数量。

3.2 自监督分割模型

虽然我们所有的实验都将专注于任意分割树 (Wielgosz et al., 2024) 方法的最新状态, 但这里提出的方法论应该能够适应任何执行树全景分割的方法, 只要它接受语义分割的二值掩码、实例分割的实例掩码, 并且不要求训练数据完全标注 (对于大多数损失函数, 可以通过将没有信息的所有区域权重设置为 0 来实现这一点)。

伪标签如下构建自树聚类:

1. 首先, 点云中的所有点都被标记为地面。
2. 然后, 所有被识别为树木的簇或点都被标记为灰度/未知。
3. 最后, 在数据中标记单个树的实例。

灰色/未知区域被分配了权重 0, 在损失中, 而地面和树木则按常规由网络处理 (任意树段默认支持灰色/未知标签, 尽管任何方法都可以通过设置损失中的权重来适应, 而这通常的支持所有常见的损失函数)。

伪标签用于训练模型若干个周期 n , 之后再进行更新。我们执行了一个简单的网格搜索来调整 n , 并且在我们的实验中使用了 $n = 3$ 。尽管最优值会根据数据集的大小及其他外部因素有所不同, 并且需要在不同的设置下重新估计。

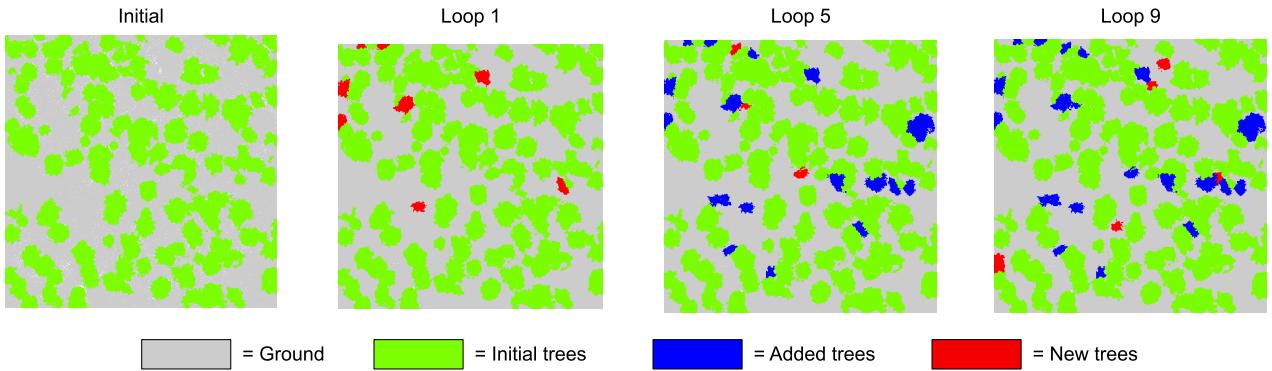


Figure 4. 新树的迭代检测

3.3 更新伪标签

经过一个训练循环（步骤5）后，点云使用当前的分割模型参数进行处理，并使用评分模型对聚类进行分类。被分类为单棵树的聚类随后将被测试是否可以添加到伪标签中。

如果候选单株树不与其他树木相交（即，它仅包含被分类为地面或灰/未知的点），则将其作为新识别的单株树添加到伪标签地图中。

当候选点与现有树点重叠时，会应用一系列测试来确定该聚类是否对应于一棵新树或一个已知实例。特别地，为了被视为一棵新树，这些聚类需要符合以下标准：

1. 候选聚类中最高点的坐标，假设为树尖，应与所有相交聚类中的最高点不同，因为假定两棵树不能有相同的树尖。
2. 树不能在超过固定阈值的距离上与其他树相交。我们将阈值设定为2米，这是训练集中估计的树宽（8米）的25%。
3. 每个簇的簇内交集（IoC）小于0.7，

$$\text{IoC} = \frac{\# \text{ points of intersection}}{\# \text{ points of cluster}}. \quad (4)$$

使用簇内交集而不是交并比，以避免小树与大树相交时被过滤掉。

一旦簇通过了所有测试，它就会被添加到伪标签中。位于簇交集中的点会根据它们与每个簇中心的距离在旧簇和新簇之间分配。随着时间的推移，将检测并添加越来越多的树到伪标签中，如图4所示，直到不再有树木被检测到。

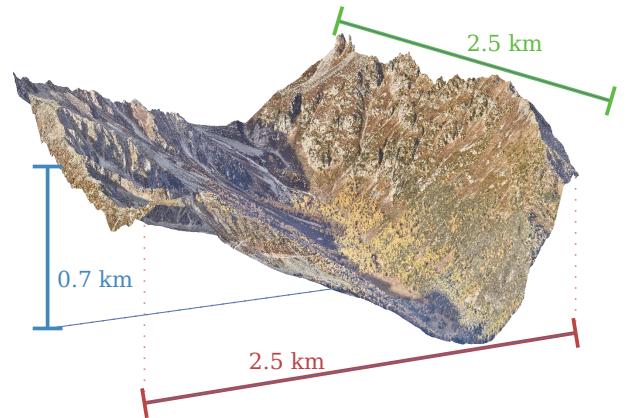


Figure 5. 瑞士瓦尔达佩特地区研究地点的着色激光雷达点云。

4. 实验

为了验证提出的程序，我们在一个新的且有些非典型的遥感数据集上进行了测试，就像在实践中应用的一样。本节介绍了所使用的数据集以及对评分模型和分割模型的实验评估。

4.1 数据集

我们将用于本研究的数据集是由一架飞机在瑞士西部一个小山地区域获取的激光雷达点云数据；参见图5，使用最先进的双通道ALS，在尽可能低的高度飞越山脉以最大化点密度。该数据与中格式航拍相机一同获取，后者用于为点云上色。

该区域大约覆盖了 6.5 km^2 ，垂直差异高达 $\approx 700 \text{ m}$ ，包括2.72亿个点，并且平均密度为 38 pt/m^2 。这个小山谷中的树木总数估计在30'000到40'000之间。为了比较，用于训练任意树段的FOR实例航空数据集的点密度范围从大约 500 pt/m^2 到 $10'000 \text{ pt/m}^2$ ，仅包含 0.002 km^2 的标注数据，并且

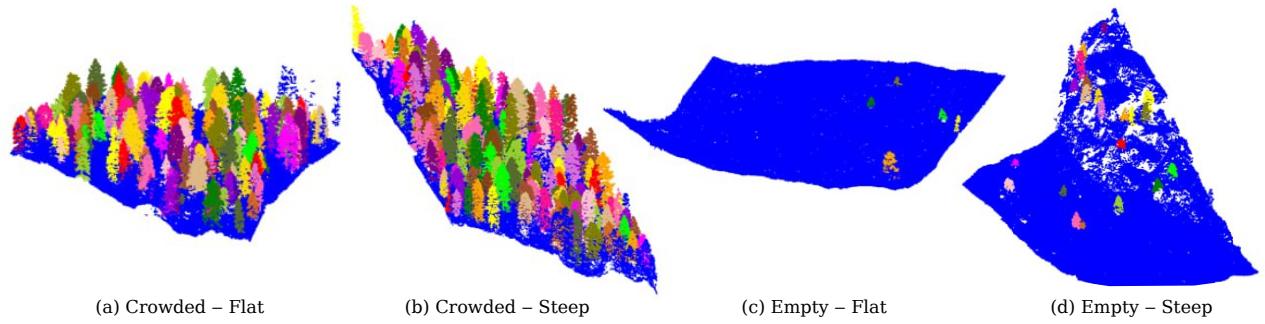


Figure 6. 标记的地面真实瓦片。

有 1'130 个标注树（还使用了一个地面数据集用于任意分割树 (Wielgosz et al., 2023) 的初始训练，最多达到 $20'000 \text{ pt}/\text{m}^2$ ）。

我们的点云对树分割方法（无论是封闭形式还是深度学习）提出了多方面的挑战，原因如下：

- 极其崎岖的地形和陡峭的斜坡可能会对分割模型产生负面影响。在某些情况下，岩石可能被错误地分类为树木，或者斜坡的部分可能被包含在分割的树实例集群中。
- 白天的时间和阴影使得基于颜色的分割更具挑战性。
- 均值点密度显著低于在同一地形上使用 ALS 可能达到的密度（几乎低了一个数量级），甚至比通过无人机或地面测量获得的点密度还要低。它也处于用于训练分割任意树的合成密度较低的一端，最低密度为 $10 \text{ pt}/\text{m}^2$ ，第二低的密度为 $100 \text{ pt}/\text{m}^2$ 。
- 点云数据是从一架在陡峭地形上方恒定高度飞行的飞机上收集的，该地形旁边有高山脊。因此，激光雷达传感器与地面之间的距离有所变化，导致数据中的点密度不一，并增加了模型训练过程的复杂性。

使用 (Yang et al., 2020) 的商业实现对整个数据集进行了初步分割，得到了 38'609 个聚类。在这些聚类中，13'316 个被手动评为代表第 3.1 节描述的三类之一。表 1 报告了手分类数据的不同类别比例。

由于训练分割模型所需的计算资源，数据集被分块处理。块的大小设置为 $100 \text{ m} \times 100 \text{ m}$ ，这是在确保每个块包含足够树木的数量和用于训练的 GPU 处

单个	3'790	28.88%
多	1'448	10.31%
非树结构	7'985	60.81%

Table 1. 初始手动分类的聚类于 Val d'Arpette 测试。

理能力之间的一种折衷。选定的块被分成一个训练集，在该集合上执行我们的方法，以及一个测试集，保留用于评估目的。

我们手动标注了测试集中四个标题的所有树实例。为了研究所提出的管道在不同条件下的行为，我们选择了一个平坦且拥挤的区域、一个平坦且陡峭的区域、一个空旷且平坦的区域和一个空旷且陡峭的区域；参见图 6。地面实况标签是手动选取的树实例与人类操作员遗漏但自动检测并后来通过视觉验证的实例的混合，显示了手工标注树群的局限性。事实上，手动分割树木不仅耗时，而且容易出错。我们在标注过程中观察到的主要错误来源与非常靠近在一起且被人类操作员标记为单棵树的情况有关。

4.2 分类准确性

对于考虑的三个模型中的每一个，都进行了一次网格搜索以调整模型最重要的超参数。准确率和加权准确率（即，即假设每个类权重相同的情况下模型的准确率）报告在表 2 中。我们开发的 KDE 基础模型达到了最高的准确率 ($\sim 90\%$)，然而，这一点紧随其后的则是 Point Transformer 架构。其中一个主要原因是，在使用 lidar 数据检测树木时，局部纹理非常重要。事实上，机载激光扫描仪经常在允许其激光穿透植被上层并探测到更深层次的树枝和叶子，有时甚至可以探测到植被下的地面的频率下工作。3DmFV(Ben-Shabat et al., 2018) 模型由于分辨率较低难以捕捉这一点，并且优化用于处理不规则数据结构的 Point Transformer 架构也稍微有些困难。

	准确性	加权准确率
3DmFV (Ben-Shabat et al., 2018)	83.4%	73.0%
Point Transformer (Zhao et al., 2021)	87.3%	85.5%
KDE (ours)	91.6%	88.6%

Table 2. 所研究的 3 个分类器模型的准确性

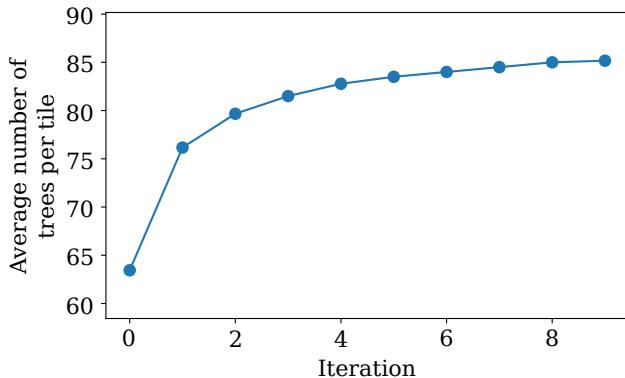


Figure 7. 每次管道迭代中的每块平均树木数量

请注意，当每个类别被重新加权时，所有模型的性能都会下降。尽管在交叉熵损失中使用了权重，模型还是稍微过拟合了**非树结构**类别，因为这个类别在训练数据中占比过高。然而，这种影响微乎其微，所以我们选择继续使用 KDE VoxNet 基础分类器作为我们的评分模型。

4.3 单木分割

为了评估所提出的迭代分割方法，我们将使用分类评分模型来估计分割的质量。这是为了检查每次迭代后单个树木的数量是否增加，以及多树和非树聚类的数量是否减少。我们以由评分模型建立的伪标签，在学习率为 $5 \cdot 10^{-5}$ 的情况下对任意分割树模型进行了重新训练，这被发现可以在训练的速度和稳定性之间提供最佳平衡。

如图 7 所示，每块地（一公顷的面积）识别到的树木平均数量从第 0 次迭代时的 63.4 增加到了第 9 次迭代时的 85.2，增加了每公顷 34.3 棵树。检测到的树木数量在开始时急剧上升，并随着时间逐渐稳定下来。图 8 还表明，随着训练的进行，**非树结构**样本的比例下降了，而**多重集群**的比例基本保持不变。这表明提出的训练方案增加了检索到的**单个树**实例的数量，并减少了被归类为**非树结构**的集群数量。

4.4 相对于手动标注的真实值的分割

虽然分割模型检测到的树木数量增加令人鼓舞，但将其与地面实际树木数量进行比较也很重要。为此我们采用了图 6 中所示的手动标记的真实数据。

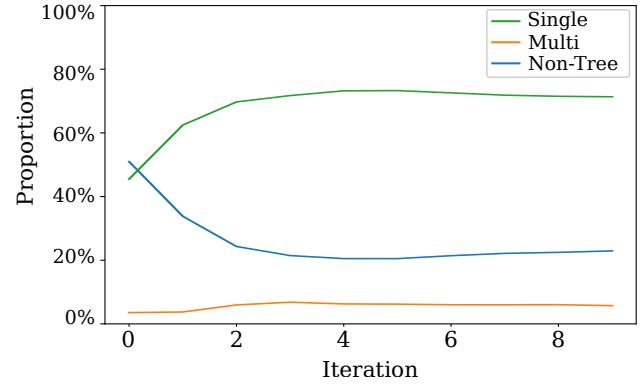


Figure 8. 随着训练的进行，评级模型预测的不同评级类别 的比例

	# 实例
(a) 拥挤的 - 平坦的	
SegmentAnyTree Original	136
SegmentAnyTree + Ours	162
Ground truth	182
(b) 拥挤——陡峭	
SegmentAnyTree Original	162
SegmentAnyTree + Ours	189
Ground truth	200
(c) 空 - 平坦	
SegmentAnyTree Original	2
SegmentAnyTree + Ours	2
Ground truth	5
(d) 空的 - 陡峭的	
SegmentAnyTree Original	10
SegmentAnyTree + Ours	15
Ground truth	22

Table 3. 相对于标记的真实值检测到的实例数目

比较如下。我们采用原始的任意分割树模型或其通过我们的方法演变出的版本，并使用描述的评分模型对其进行处理，从而过滤掉**非单个树**聚类。这显示了我们方法的另一个优势：评分模型可以在推理时与分割模型结合，以精化其输出。最后，我们将聚类中预测的树的数量与真实值进行比较。

结果如表 3 所示，根据地形陡峭程度和树木种群密度显示，虽然我们的方法未达到 100% 的覆盖率，但在至少 9 次迭代后，它增加了检测到的实例数量。对于包含少量孤立树木的稀疏地块，性能并不强劲。这表明该方法在面对与数据集中典型树木相比稍微偏离领域的较小灌木时遇到了挑战。

图 9 中所示结果的详细分析与密集区域的实际地面真值非常吻合，这与斜率陡峭程度无关。只有实际存在的树木中的一部分未被检测到，导致 IOU 非常小。对于拥挤的瓦片，超过 80% 的实际树木聚类与其

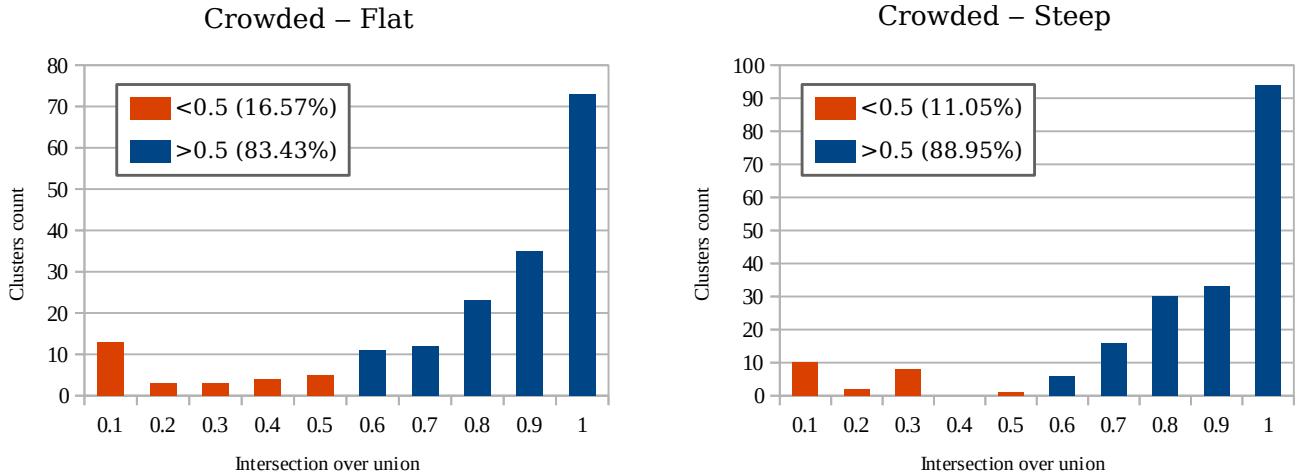


Figure 9. 地面真实集群与微调模型在地面真实拥挤瓷砖上相应预测之间的 IOU

对应的预测聚类具有大于 0.5 的 IOU。对于“空”瓦片（在补充材料中有讨论）的结果更令人担忧，在这些区域不仅大部分实际存在的小树未被检测到，而且部分分割出来的树木也不匹配地面真值。

5. 结论

我们提出了一种方法，该方法能够利用高效获取的人工评分标签来促进在具有挑战性的环境中从激光雷达点云数据中进行树木实例分割。所提出的方法显著提高了最先进的深度学习模型在接近树线的高山地形中检测单个树木的比例，并且当与我们的自动评分模型结合使用时，使在手动标注的密集验证图块中标记了超过 80% 的树木得以恢复。然而，在崎岖地形中检测小而孤立的树木仍然存在挑战。这部分原因是这些区域中的树木相对于森林中的主要树木群体来说非常小且分布不同。

我们的结果展示了在树木个体实例分割中使用简单评分模型进行弱监督的潜在益处。该评分模型也可以与分割模型在推理时结合使用，以过滤掉无效簇。这为分割模型提供了额外的价值。

6. 致谢

斯万·埃米利安·塞莱斯特·德图什在杰西·拉哈伊和劳伦特·约斯平的监督下进行了本文所述的实验。实验设计由劳伦特·约斯平完成，斯万·埃米利安·塞莱斯特·德图什做出了额外贡献。评估点簇的工具是由劳伦特·约斯平开发的。用于在点云数据中标记簇的手动标签工具由斯万·埃米利安·

塞莱斯特·德图什开发。总体研究由扬·斯卡鲁德指导。感谢瑞士飞行服务公司和 SixSense Helimap SA 为我们的数据集获取和预处理提供了便利。

References

- Ben-Shabat, Y., Lindenbaum, M., Fischer, A., 2018. 3DmFV: Three-Dimensional Point Cloud Classification in Real-Time Using Convolutional Neural Networks. *IEEE Robotics and Automation Letters*, 3(4), 3145-3152. [2](#), [4](#), [6](#), [7](#)
- Chaudhari, S., Aggarwal, P., Murahari, V., Rajpurohit, T., Kalyan, A., Narasimhan, K., Deshpande, A., Castro da Silva, B., 2025. RLHF Deciphered: A Critical Analysis of Reinforcement Learning from Human Feedback for LLMs. *ACM Comput. Surv.* [1](#)
- Chen, X., Jiang, K., Zhu, Y., Wang, X., Yun, T., 2021. Individual Tree Crown Segmentation Directly from UAV-Borne LiDAR Data Using the PointNet of Deep Learning. *Forests*, 12(2). <https://www.mdpi.com/1999-4907/12/2/131>. [2](#)
- Chen, Y., Liu, J., Zhang, X., Qi, X., Jia, J., 2023. Voxelnext: Fully sparse voxelnet for 3d object detection and tracking. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 21674-21683. [2](#)
- Cheng, T., Wang, X., Chen, S., Zhang, Q., Liu, W., 2023. Boxteacher: Exploring high-quality pseudo labels for weakly supervised instance segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3145-3154. [1](#)
- Dalponte, M., Coomes, D. A., 2016. Tree-centric mapping of forest carbon density from airborne laser scanning

- and hyperspectral data. *Methods in Ecology and Evolution*, 7(10), 1236-1245. [1](#)
- Fan, W., Tian, J., Troles, J., Döllerer, M., Kindu, M., Knoke, T., 2024. Comparing Deep Learning and MCWST Approaches for Individual Tree Crown Segmentation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-1-2024, 67-73. <https://isprs-annals.copernicus.org/articles/X-1-2024/67/2024/>. [1](#), [2](#)
- Fouqueray, T., Charpentier, A., Trommetter, M., Frascaria-Lacoste, N., 2020. The calm before the storm: How climate change drives forestry evolutions. *Forest Ecology and Management*, 460, 117880. [1](#)
- Hu, J. S. K., Kuai, T., Waslander, S. L., 2022. Point density-aware voxels for lidar 3d object detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 8469-8478. [4](#)
- Huang, Q., Wang, W., Neumann, U., 2018. Recurrent slice networks for 3d segmentation of point clouds. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [2](#)
- Jiang, H., Zou, Q., Zhou, B., Jiang, Y., Cui, J., Yao, H., Zhou, W., 2023. Estimation of Shallow Landslide Susceptibility Incorporating the Impacts of Vegetation on Slope Stability. *International Journal of Disaster Risk Science*, 14(4), 618-635. [1](#)
- Jiang, L., Zhao, H., Shi, S., Liu, S., Fu, C.-W., Jia, J., 2020. Pointgroup: Dual-set point grouping for 3d instance segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. [3](#)
- Lee, J., Cai, X., Lellmann, J., Dalponte, M., Malhi, Y., Butt, N., Morecroft, M., Schönlieb, C.-B., Coomes, D. A., 2016. Individual Tree Species Classification From Airborne Multisensor Imagery Using Robust PCA. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(6), 2554-2567. [2](#)
- Liu, Y., You, H., Tang, X., You, Q., Huang, Y., Chen, J., 2023. Study on Individual Tree Segmentation of Different Tree Species Using Different Segmentation Algorithms Based on 3D UAV Data. *Forests*, 14(7). [2](#)
- Ma, Z., Pang, Y., Wang, D., Liang, X., Chen, B., Lu, H., Weinacker, H., Koch, B., 2020. Individual Tree Crown Segmentation of a Larch Plantation Using Airborne Laser Scanning Data Based on Region Growing and Canopy Morphology Features. *Remote Sensing*, 12(7). [1](#), [2](#)
- Maturana, D., Scherer, S., 2015. Voxnet: A 3d convolutional neural network for real-time object recognition. *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 922-928. [2](#), [4](#)
- Qi, C. R., Su, H., Mo, K., Guibas, L. J., 2017a. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [2](#)
- Qi, C. R., Yi, L., Su, H., Guibas, L. J., 2017b. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (eds), *Advances in Neural Information Processing Systems*, 30, Curran Associates, Inc. [2](#)
- Rizaldy, A., Persello, C., Gevaert, C. M., Oude Elberink, S. J., 2018. FULLY CONVOLUTIONAL NETWORKS FOR GROUND CLASSIFICATION FROM LIDAR POINT CLOUDS. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2, 231-238. [2](#)
- Shimoda, W., Yanai, K., 2019. Self-supervised difference detection for weakly-supervised semantic segmentation. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. [1](#)
- Straker, A., Puliti, S., Breidenbach, J., Kleinn, C., Pearse, G., Astrup, R., Magdon, P., 2023. Instance segmentation of individual tree crowns with YOLOv5: A comparison of approaches using the ForInstance benchmark LiDAR dataset. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 9, 100045. [2](#)
- Wielgosz, M., Puliti, S., Wilkes, P., Astrup, R., 2023. Point2Tree(P2T)—Framework for Parameter Tuning of Semantic and Instance Segmentation Used with Mobile Laser Scanning Data in Coniferous Forest. *Remote Sensing*, 15(15). [6](#)
- Wielgosz, M., Puliti, S., Xiang, B., Schindler, K., Astrup, R., 2024. SegmentAnyTree: A sensor and platform agnostic deep learning model for tree segmentation using laser scanning data. *Remote Sensing of Environment*, 313, 114367. [1](#), [3](#), [4](#)
- Xiang, B., Peters, T., Kontogianni, T., Vetterli, F., Puliti, S., Astrup, R., Schindler, K., 2023. TOWARDS ACCURATE INSTANCE SEGMENTATION IN LARGE-SCALE LIDAR POINT CLOUDS. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-1/W1-2023, 605-612. [2](#)
- Xiang, B., Wielgosz, M., Kontogianni, T., Peters, T., Puliti, S., Astrup, R., Schindler, K., 2024. Automated forest inventory: analysis of high-density airborne LiDAR point

clouds with 3D deep learning. *Remote Sensing of Environment*, 305, 114078. [3](#)

Xiang, B., Wielgosz, M., Puliti, S., Král, K., Krůček, M., Missarov, A., Astrup, R., 2025. Forestformer3d: A unified framework for end-to-end segmentation of forest lidar 3d point clouds. [3](#)

Yang, J., Kang, Z., Cheng, S., Yang, Z., Akwensi, P. H., 2020. An Individual Tree Segmentation Method Based on Watershed Algorithm and Three-Dimensional Spatial Distribution Analysis From Airborne LiDAR Point Clouds. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 1055-1067. [2](#), [3](#), [6](#)

Zhao, H., Jiang, L., Jia, J., Torr, P. H., Koltun, V., 2021. Point transformer. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 16259–16268. [2](#), [4](#), [7](#)