arXiv:2304.00793v1 [cs.CV] 3 Apr 2023

# FinnWoodlands 数据集

Juan Lagos[1],Urho Lempiö[1],和 Esa Rahtu[1][0000−0001−8767−0864]

芬兰坦佩雷，坦佩雷大学 {juanpablo.lagosbenitez, urho.lempio, esa.rahtu}@tuni.fi

**摘要**。尽管大型、多样化数据集的普及推动了自动驾驶和室内应用领域的重大突破，但林业应用仍相对滞后，而新的森林数据集无疑将有助于推动面向森林类场景的数据驱动方法的开发取得显著进展。本文介绍了一个名为 $FinnWoodlands$ 的森林数据集，其中包含 RGB 双目图像、点云和稀疏深度图，以及用于语义分割、实例分割和全景分割的真实值人工标注。$FinnWoodlands$ 共包含 4226 个人工标注的对象，其中 2562 个对象（60.6%）为树干，分为三个不同的实例类别，即“云杉树”“桦树”和“松树”。除树干外，我们还将“障碍物”对象标注为实例，并标注了“湖泊”“地面”和“轨道”等语义 stuff 类别。我们的数据集可用于需要环境整体表征的林业应用。我们使用三个模型，分别针对实例分割、全景分割和深度补全提供了初步基准测试，并展示了此类非结构化场景带来的挑战。$Finn$ $Woodlands$ 数据集可在 https://github.com/juanb09111/FinnForest. git 获取。

**关键词：** 机器学习 · 深度学习 · 林业 · 数据集

## 1 引言

公共数据集有助于激发人们对计算机视觉领域研究的兴趣。数据可用性加速了新范式、技术和模型的发展，尤其是旨在解决图像分割、目标检测、深度估计、光流估计和目标跟踪等不同计算机视觉任务的数据驱动方法与深度学习模型 [32]。一些公共数据集甚至已成为评测基准，使不同方法能够宣称达到最先进水平 [4, 5, 7–9, 11, 13, 15,16, 22, 25, 26, 28, 36]。尽管大多数热门数据集都采集自城市环境，这对自动驾驶应用的发展显然大有裨益，但在其他场景中，数据可用性却相对滞后，尤其是森林等越野景观。与城市场景相比，使用合适的数据采集装置在森林中行进要困难得多，这一点尤其如此

，这在一定程度上解释了森林数据集的可用性为何不及室内或城市数据集。

尽管如此，增加森林数据集的可用性仍能让众多应用受益。例如，在农业、农耕和探索应用中 [2, 3, 23, 30], ，与城市和室内场景相比，物体的结构和定义不那么明确，物体之间的边界也不那么清晰，移动、交互和外观都更加模糊。因此，在森林类场景中执行自主导航等任务极具挑战性，因为这类环境的可控性较低，且没有车道线、导航标志或清晰的路径。

此外，在此类非结构化场景中的导航标准化程度较低。场景多种多样，既包括森林深处的土路和碎石路，大型车辆和收割机、拖拉机等用于农业应用的重型机械 [27] 可以在这些路上行驶，也包括仅供徒步者和探险者通行的较窄徒步小径。虽然森林数据集与特定应用群体，尤其是林业，密切相关，但此类数据集也有一些共同特征，例如缺乏结构性，而其他应用或许也能加以利用。

从森林中采集数据并非易事。森林在不同季节会发生显著变化，植被类型也因地理位置而异。因此，打造一个优质的森林数据集无疑需要集体努力。基于这一动机，我们推出了***FinnWoodlands***，这是一个采集自芬兰森林徒步小径的数据集。它包含 5170 个立体 RGB 帧、对应的激光雷达点云，以及每帧对应的稀疏深度图。此外，我们还为 300 帧提供语义分割、实例分割和全景分割标注，其中包含 4226 个由人工使用 CVAT [29], 这一开源图像和视频标注工具标注的对象。我们为其他科学家提供了扩展 FinnWoodlands、添加更多帧的指南；这也将通过纳入来自世界各地的森林图像，提升数据集的多样性。由此，FinnWoodlands 为林业应用奠定了一个可靠的种子数据集，我们希望借此吸引数据科学家群体的关注。

我们使用三种不同的模型评估了我们的数据集，分别用于实例分割、全景分割和深度补全，即 Mask R-CNN [18],EfficientPS [24], 和 FuseNet [10],。由此，我们为该数据集建立了初步基准测试。我们工作的主要贡献如下：

– 我们提供了一个名为 *FinnWoodlands* 的森林数据集，其中包含 RGB 双目影像帧、点云和稀疏深度图，以及用于语义分割、实例分割和全景分割的真实标注（GT）。据我们所知，在林业应用领域，还没有其他数据集提供全景分割真值标注。因此，我们旨在促进森林环境中整体场景理解方面的研究。– 我们展示了如何在移动和导航具有挑战性的场景中采集数据，所用的简易数据采集装置由安装在背包上的激光雷达传感器和双目相机组成。我们的装置可以

与其他更昂贵的解决方案相比，该装置易于在其他地方复制，从而采集可靠的数据。

## 2 相关工作

在众多公共数据集中，只有少数是在森林场景中采集的，其中一些面向导航任务，其余则专注于工业林业应用中的树木检测与分割。本文概述了公开可用的森林数据集，并总结了表 1 所示各数据集提供的标注。图 1 和图 2 展示了所列数据集的代表性图像及其标注。

***CANATREE100*** 数据集 [17] 包含 100 张 RGB 图像和 100 张深度图像，并标注了约 920 棵树木。该数据集采集于加拿大的森林中，并提供了树干的分割掩码标注。

***ForTrunkDet*** [31] 是一个用于树干检测的数据集，采集自三个不同的森林地点。它包含人工标注的可见光图像和热成像图像，涉及桉树和松属两个树种。该数据集包含 2029 张可见光谱图像和 866 张热成像图像。

***RELLIS-3D*** 数据集 [20] 由从越野环境中采集的多模态同步传感器数据帧组成。它包含五个序列，涵盖 RGB 图像、IMU 数据、GPS 数据、激光雷达点云和立体图像。

机器人非结构化地面行驶数据集（***RUGD***） [35]是由地面机器人在半城市化地点以及森林等非结构化场景中采集的数据集。它由视频序列组成，包含几何形态不规则且不一致的物体。机器人在不同类型的地形中移动，这些地形标记为 “溪流”、“公园”、“小径”和“村庄”。

***SYNTHTREE43K*** [17] 是一个用于树木检测的合成数据集，采集自模拟森林环境，包含多达 17 种不同类型的树木。它由使用 Unity 游戏引擎生成的 $43k$ 合成图像组成 [21]。

***TartanAir*** [34] 是一个用于导航任务的大型照片级真实感数据集。它提供室外和室内场景以及各种环境，其中包括两个涵盖不同季节（秋季和冬季）森林场景的子集。由于具有合成特性，TartanAir 数据集还提供多模态真实标注，例如语义分割标签、深度、相机位姿、光流、立体视差、合成激光雷达点和合成 IMU 读数。

***TimberSeg 1.0*** 数据集 [14] 由 220 张在不同季节和环境光照条件下采集的 RGB 图像组成。它为单个树木原木提供边界框和分割掩码标注。*TimberSeg 1.0* 数据集中共有 2500 根经过分割的原木。该数据集尤其面向集材和原木抓取应用。

**表 1.** 数据集与真值标注

| 数据集 | | | | | | | |
|---|---|---|---|---|---|---|---|
| CANATREE100 [17] | ✔ | ✔ | ✗ | ✗ | ✗ | ✗ | ✔ |
| ForTrunkDet [31] | ✔ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |
| RELLIS-3D [20] | ✗ | ✗ | ✔ | ✗ | ✔ | ✗ | ✗ |
| RUGD [35] | ✗ | ✗ | ✔ | ✗ | ✗ | ✗ | ✗ |
| SYNTHTREE43K [17] | ✔ | ✔ | ✗ | ✗ | ✗ | ✗ | ✔ |
| TartanAir [34] | ✗ | ✗ | ✔ | ✗ | ✔ | ✔ | ✔ |
| TimberSeg 1.0 [14] | ✔ | ✔ | ✗ | ✗ | ✗ | ✗ | ✗ |
| **FinnWoodlands（我们）** | ✔ | ✔ | ✔ | ✔ | ✗ | ✗ | ✗ |

## 3 数据集特征

我们的数据集 *FinnWoodlands*包含 5170 组同步立体 RGB 图像，每帧均配有激光雷达点云和稀疏深度图。图 4 展示了一个样本帧的可视化结果及其对应的点云（图 4a），以及激光雷达点在图像上的投影（图 4b）。数据采集自芬兰坦佩雷附近的三个不同地点。标注均由人工完成，包括语义分割图像、边界框、类别标签、实例分割和全景分割真实标注图像。我们还提供了用于实例分割和全景分割的 COCO 格式标注文件。表 2 汇总了数据集所包含的数据。

我们的数据集使用“背景物质”和“可数物体”作为类别。在计算机视觉领域，“背景物质”类别指不可数的物体，例如“天空”或“地面”；而“可数物体”类别指可数的物体，例如“汽车”或“人” [1]。“可数物体”类别下的物体可以使用边界框和实例分割掩码进行标注；“背景物质”类别下的物体通常使用逐像素分割掩码进行标注，且无法被限制在单个边界框内。

*FinnWoodlands* 包含三类“背景物质”和五类“可数物体”。“背景物质”类别为“湖泊”“地面”和“轨道”，其中“轨道”

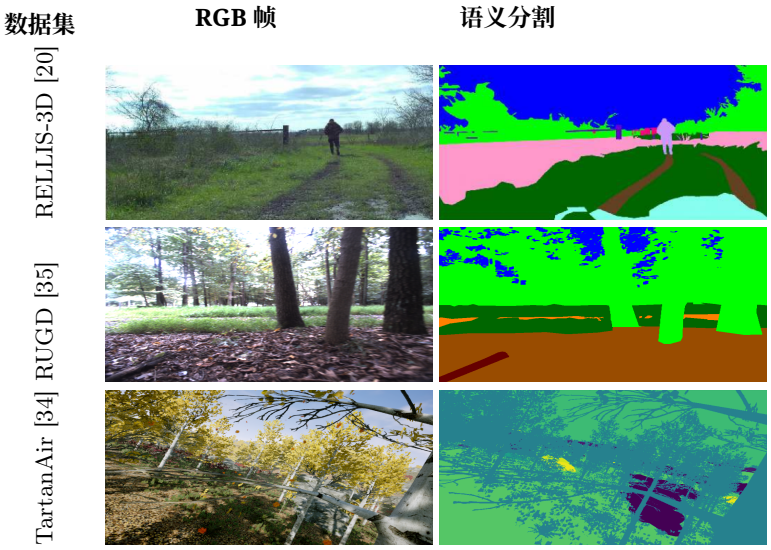



**图 1.** 语义分割标注。来自三个森林数据集的 RGB 图像样本及其对应的语义分割真实标注：RELLIS- 3D [20], RUGD [35], 和TartanAir [34]。

**表 2.** FinnWoodlands 数据集概览

| Data | 样本数量 | 格式 |
|---|---|---|
| 立体 RGB 帧 | 5170 | .jpg |
| 激光雷达点云 | 5170 | .pcd |
| 语义真值 | 300 | 标签和 RGB PNG图像 |
| 实例真值 | 300 | RGB PNG图像 |
| 全景真值 | 300 | RGB PNG图像 |
| 稀疏深度图 | 5170 | 深度PNG图像 |

指步行小径或道路。“可数物体”类别包括“障碍物”“云杉”“桦树”“松树”和“树木”。“障碍物”类别指步行小径上或附近的障碍物。“云杉”“桦树”和“松树”指芬兰森林中常见的树种。其他不属于这些类别的树木均归入通用类别“树木”。通过在数据集中纳入“背景物质”和“可数物体”类别，我们旨在利用场景的全景分割标注和稀疏深度图，为森林场景提供更全面的三维表示。

绝大多数经过标注和分割的对象都是树干，约占对象总数的 60.6%。表 3 展示了各类别的总体表示情况。在树种中，“云杉”是我们的数据集中最常见的树种，约占已标注对象总数的

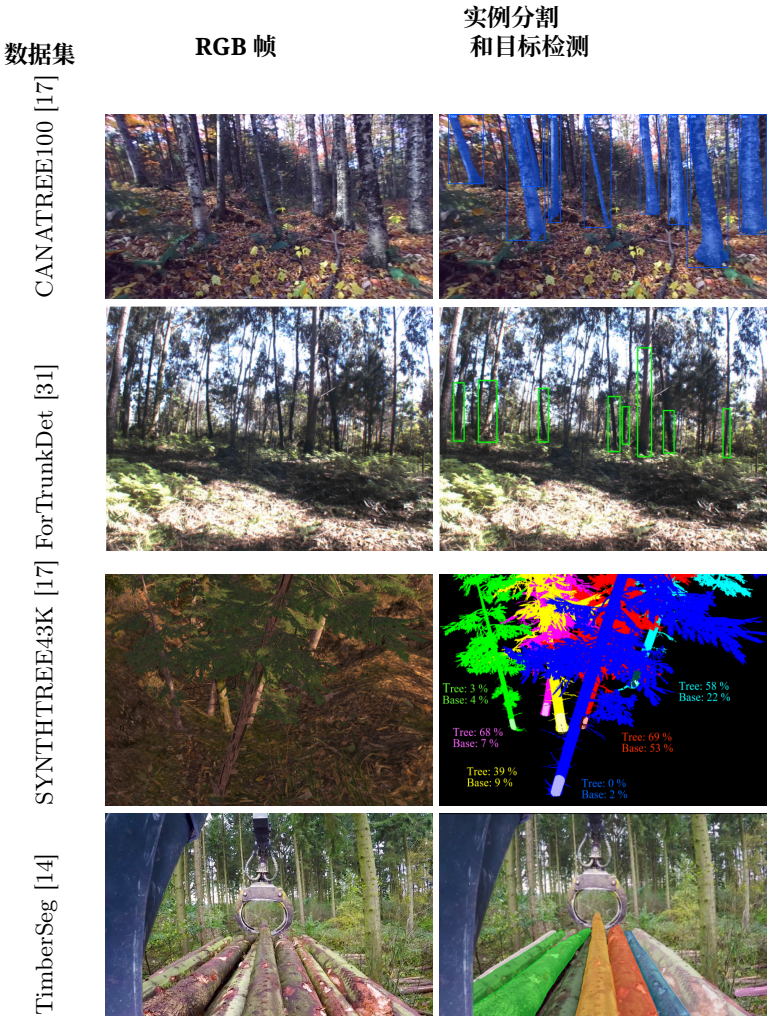



**图 2.** 实例分割标注。来自四个不同森林数据集的样本 RGB 图像及其对应的实例分割真值标注：CANATREE100 [17], ForTrunkDet [31], SYNTHTREE43K [17], TimberSeg [14]。

32.5%，如表 4 所示。

图 3 展示了标注的可视化效果，其中选取了 *FinnWoodlands* 数据集的代表性样本来呈现真实图像。每列对应一个不同的场景；各行展示 RGB 帧及其对应的标注。第一行展示了三个不同的场景；最左侧的画面呈现了一片典型的芬兰森林，其中有一条步行 trail。中间列的画面呈现了一片森林，其中有一条步行 trail、一些障碍物，右侧还有一片湖泊。

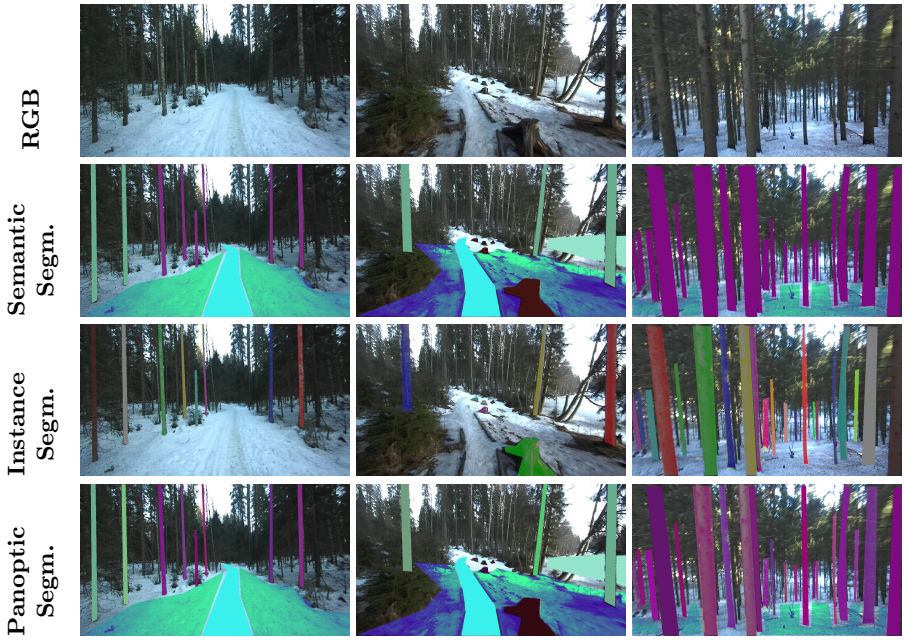



**图 3.** FinnWoodlands 真实标注。FinnWoodlands 数据集的代表性样本。*FinnWoodlands*。第一行展示了三张不同的 RGB 图像。第二至第四行分别展示了语义分割、实例分割和全景分割的真实标注。

**表 3.** FinnWoodlands 物体类别

| 对象类别 | 总数 | 总面积（像素） | 表示 |
|---|---|---|---|
| Lake | 378 | 8.728M | 8.9% |
| 障碍物 | 525 | 0.907M | 12.4% |
| 地面 | 554 | 73.873M | 13.1% |
| 轨道 | 207 | 7.778M | 4.8% |
| Tree | 2562 | 27.906M | 60.6% |

**表 4.** FinnWoodlands 树木对象类别

| 树种 | 总数 | 总面积（像素） | 表示 |
|---|---|---|---|
| 云杉 | 1374 | 16.439M | 32.5% |
| 桦树 | 683 | 6.884M | 16.1% |
| Pine | 430 | 4.262M | 10.1% |
| 其他类型的树 | 75 | 0.319M | 1.7% |

最后，最右侧一列展示了一片没有步行 trail、云杉树密度很高的森林。从第二行到第四行，

图中展示了对应的语义分割真实标注、实例分割真实标注和全景分割真实标注。

## 4 数据采集

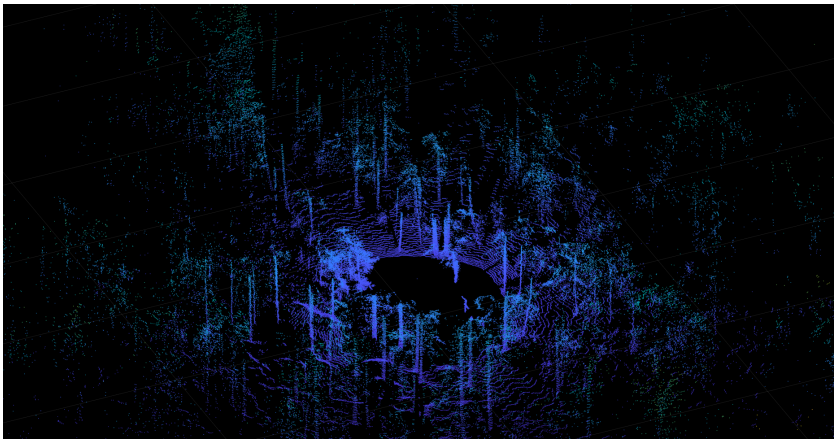
对于$FinnWoodlands$ 数据集，使用一台 64 线 Ouster OS1 激光雷达传感器和一台 ZED2 立体相机，以每秒 10 帧的速度记录点云和分辨率为 720p 的立体图像，因此每帧都包含立体 RGB 图像和一个点云。激光雷达和相机牢固地安装在背包的金属底盘上，位于佩戴者头顶，OS1 在上方，ZED2 紧贴其下方，如图 5 所示。使用一台联想 ThinkPad T440P 笔记本电脑采集并同步来自相机和传感器的数据。计算机运行 Ubuntu Linux（64 位，版本：18.04）、机器人操作系统（ROS，版本：Melodic）、Nvidia CUDA 工具包（版本：10.2）和 Linuxptp 软件。使用 Linuxptp 提供的 PTP 工具作为主时钟来同步传感器。随后使用 ROS 工具记录数据。

每次录制时长约为 50-100 秒。录制时，研究人员以悠闲的步伐沿着徒步小径或在森林中行走。地点选自芬兰坦佩雷附近不同类型的森林，包括 Kyötikkälä、Hervantajärvi 和 Suolijärvi。由于录制是在早春进行，地面有积雪，湖泊也已结冰。在这三个地点中，Hervantajärvi 和 Suolijärvi 周围环绕着湖泊，桦树较多；而 Kyötikkälä 则以云杉为主。

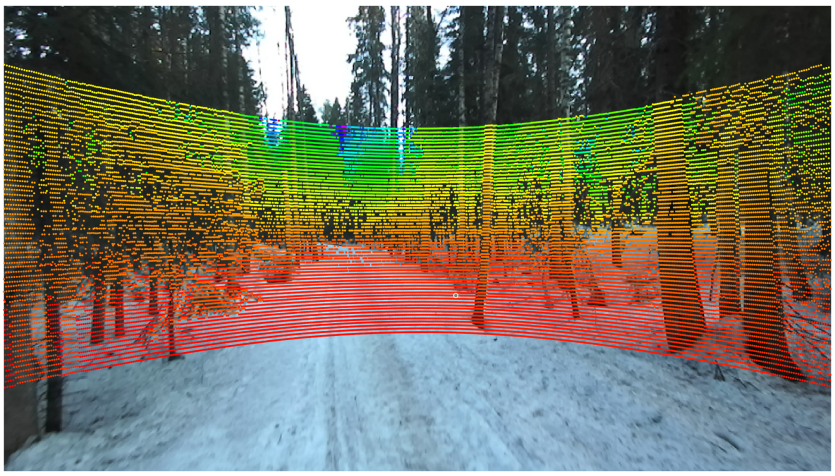
与城市环境相比，森林是一个难以开展工作的环境。地面凹凸不平，而且有许多障碍物，例如岩石、树木、坑洼和山丘。由于地面积雪及其他障碍物，有些地方步行采集数据颇具挑战。不过，我们的数据采集设置虽然简单，却足够灵活，能够在此类条件下采集可靠的数据。

## 5 实验

在林业应用中，与其他应用一样，理解给定图像的语义含义至关重要。通常还需要与场景中的物体进行交互。因此，仅靠语义分割并不足够，因为它只提供逐像素分类，而实例分割则支持目标检测和目标分割。此外，还可以通过全景分割结合实例分割与语义分割，从而更全面地表征场景。另一方面，深度补全以不完整的稀疏深度图为输入，为输入帧中的每个像素提供深度值。这对于需要在三维空间中与物体交互的应用尤其重要。

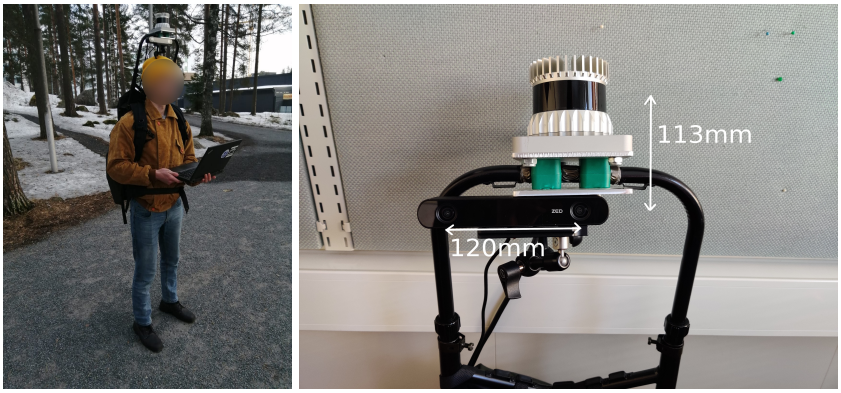


(a) 点云



(b) 点云投影

**图4.** *FinnWoodlands* 样本点云和稀疏深度图可视化。稀疏深度图对应于三维点云投影到二维图像上的结果。




(a) 背包 (b) 装置

**图 5.** 用户背着的背包及装置的特写。OS1 激光雷达和 ZED2 双目相机安装在背包顶部。ZED2 相机位于激光雷达正下方，与激光雷达垂直中心的距离为 $113mm$ 。激光雷达和 ZED2 相机均通过可调节金属底盘安装在用户头部上方。ZED2 相机的基线为 $120mm$。

因此，我们针对三种不同模型进行了基线实验，分别是用于实例分割的 Mask R-CNN [18]、用于全景分割的 EfficientPS [24]，以及基于稀疏深度图和 RGB 图像进行深度补全的 FuseNet [10] 。训练集包含从两个不同地点采集的 150 帧图像，评估集包含从一个未用于训练集的不同地点采集的 50 帧图像。

为了训练深度补全模型 $FuseNet$ [10], ，我们对稀疏深度图进行下采样，并将得到的稀疏深度图用作训练数据，同时保留原始稀疏深度图作为真实标注。下采样后的稀疏深度图包含 3500 个深度值，约占原始深度图中点数的 25%。

***Mask R-CNN*** [18] 于 2017 年达到了端到端实例分割的最先进水平，此后一直被广泛用作基准模型。它以 RGB 图像为输入，并返回三种不同类型的输出。具体而言，它会返回用于目标定位的边界框、每个检测到对象的类别标签，以及分割掩码。它由用于特征提取的骨干网络、区域建议网络（RPN）以及三个并行的输出头组成，用于预测给定输入图像中相应的类别标签、边界框和分割掩码。我们使用了以 EfficientNet-B5 为骨干网络的 Mask R-CNN，该模型在 ImageNet 数据集上进行了预训练 [33], [12]。我们将批归一化层 [19] 替换为同步原位激活批归一化 [6] ，以进行 GPU 优化。

***EfficientPS*** [24] 是一种用于端到端全景分割的模型。与 Mask R-CNN 类似，它使用骨干网络提取多个特征图，并将这些特征图馈送到两个输出分支，最后还包含一个全景融合模块。第一个分支基于 Mask R-CNN 执行实例分割，使用区域建议网络和三个输出头来预测边界框、类别标签和分割掩码。第二个分支执行语义分割，最后，全景融合模块将语义逻辑值与实例分割掩码逻辑值融合，生成全景分割图。模型名称指的是其骨干网络，该网络基于可扩展的 *Efficient- Net* [33] 架构，并结合双向特征金字塔网络（FPN），以便在多个尺度上进行特征提取。

***FuseNet*** [10] 使用 RGB 图像和稀疏深度图进行端到端深度补全。它输出完全稠密的深度图。FuseNet 通过一个构建模块联合学习二维和三维特征：该模块使用二维卷积层处理二维张量，并使用连续卷积处理三维点。随后，构建模块提取的特征会在二维空间中融合。

### 5.1 评估

我们采用了标准的 COCO 评估指标 [22]。具体而言，我们计算了用于评估实例分割的平均精度均值（mAP）、用于评估语义分割的平均交并比（mIoU），以及用于评估全景分割的全景质量（PQ）、分割质量（SQ）和识别质量（RQ）。我们使用均方根误差（RMSE）评估深度补全任务。

### 5.2 结果

表 5 展示了各个模型的定量性能结果。根据 mAP@50，Mask R-CNN [18] 的整体性能为 28%。从图 6 可以看出，尽管它能够检测树干，但分割掩码的准确度并不高，尤其是在树干彼此非常接近时。这表明，在树木密集的森林中仍有改进空间。同样，如图 7 所示，EfficientPS [24] 能够有效检测树干。然而，在茂密的森林场景中，分割掩码与对应树木的边界贴合得不够紧密。总体 PQ 达到 27.8%；但分别针对“可数物体”和“背景物质”评估 PQ 时，结果存在显著差异。EfficientPS 在“可数物体”实例上的表现较差，SQ 和 RQ 指标也呈现相同情况。这些结果再次凸显了茂密森林场景给机器学习方法和数据驱动算法带来的挑战，尤其是在分割单个物体方面。我们还计算了 EfficientPS 的 mAP@50，结果显示它在这一指标上优于 Mask R-CNN。另一方面，即使在图像中未提供稀疏深度信息的区域，FuseNet[10] 的泛化能力也相当不错。然而，精细结构和边界

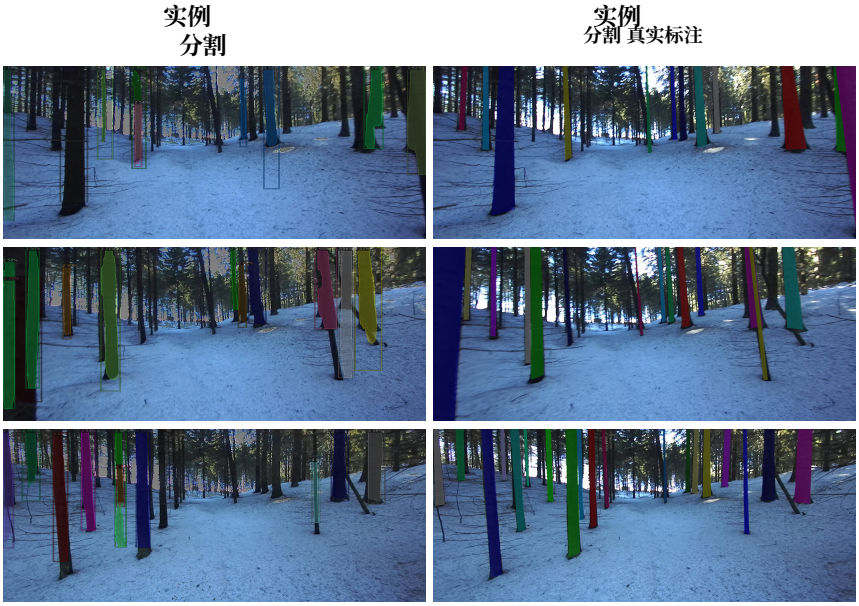



**图6.** Mask R-CNN [18] 定性结果。第一列展示了三个不同的输入帧。第一列和第二列分别展示了 Mask R-CNN [18] 的实例分割输出和真实标注分割。

如图 8 所示，树木等物体的信息有所丢失，因此深度补全也有改进空间。

**表 5.** 定量结果

| 模型 | 指标 | 数值 |
|---|---|---|
| Mask R-CNN | mAP@50 | 28% |
| EfficientPS | PQ | 27.8 |
| | SQ | 32.5% |
| | RQ | 34.1% |
| | PQ Stuff | 41.5% |
| | SQ Stuff | 41.5% |
| | RQ Stuff | 50.0% |
| | PQ Things | 18.7% |
| | SQ Things | 26.5% |
| | RQ Things | 23.5% |
| | mAP@50 | 50.0% |
| FuseNet | 均方根误差(mm) | 489.96 |

## 6 结论

本文介绍了*FinnWoodlands*，这是一个用于森林环境场景理解的独特数据集。*FinnWoodlands* 提供实例分割、语义分割和全景分割的人工真实标注，以及整体场景表示所必需的稀疏深度图。我们的数据集包含森林场景中常见的非结构化物体，并重点检测和分割三种不同树种的树干，即“云杉”“桦树”和“松树”。我们使用相对简单的数据采集设置收集数据，该设置易于复现，可用于生成类似数据并扩展*FinnWoodlands* 数据集。我们还通过使用三种深度神经网络架构测试数据，提供了初步基准测试：用于实例分割的 Mask R-CNN [18] 、用于全景分割的 EfficientPS [24]，以及用于深度补全的 FuseNet [10] 。结果揭示了在森林等高度非结构化场景中部署模型时计算机视觉面临的一些挑战，也凸显了在类似场景中改进的机会。我们的数据集有望推动林业应用的发展，并促进计算机视觉领域在森林类场景中的研究。

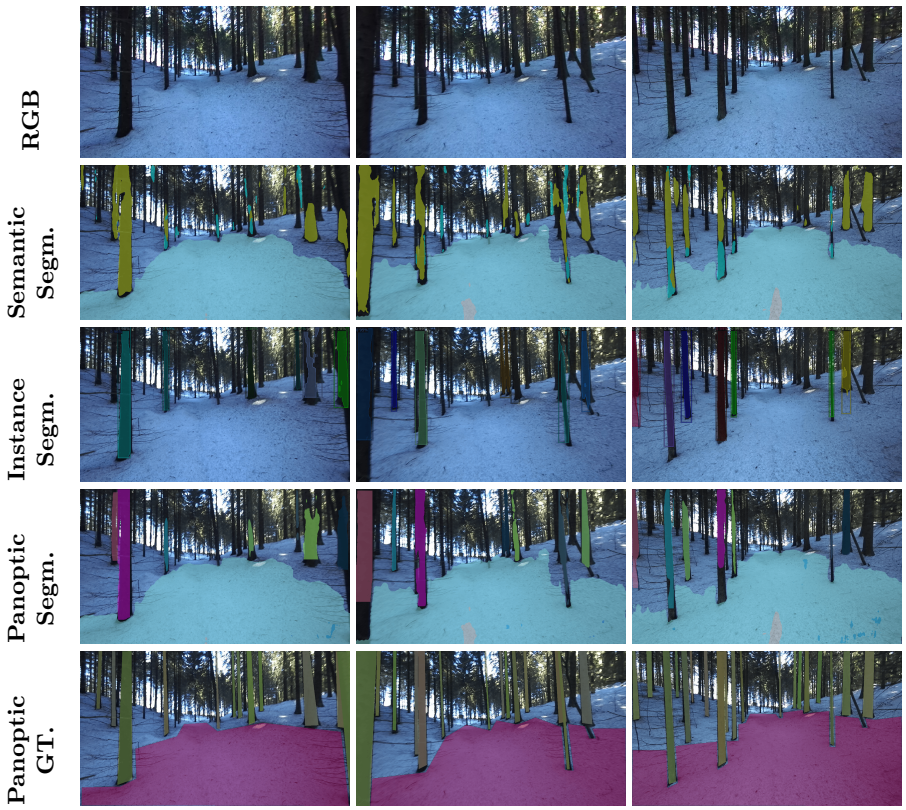



**图7。** EfficientPS [24] 定性结果。第一行展示了三张不同的 RGB 输入图像。从第二行到第四行，依次展示 EfficientPS 在各项任务上的分割结果，具体来说，分别是语义分割、实例分割和全景分割。最后一行展示了真实标注分割。

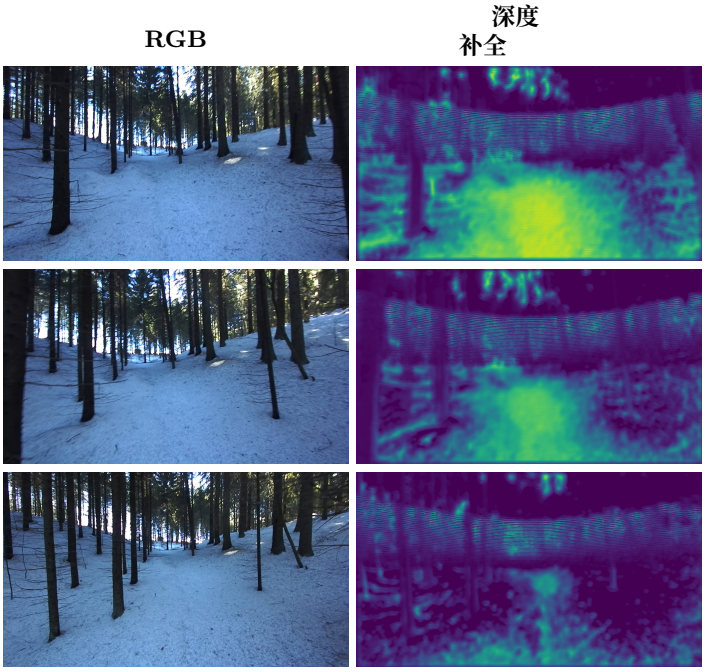



**图 8.** FuseNet [10] 定性结果。第一列显示了三张不同的输入图像，第二列则展示了每个输入场景对应的深度补全结果。FuseNet [10] 具有较好的泛化能力，包括稀疏深度图中未提供深度信息的区域；然而，森林中物体的一些精细边缘和结构未能恢复。

# 参考文献

[1] Adelson, E.H.：论如何看见背景物质：人类与机器的材料感知。载于：IS&T/SPIE Electronic Imaging（2001）[2] Bac, C.W., Van Henten, E., Hemming, J., Edan, Y.：高价值作物采收机器人：最先进水平综述与未来挑战。《现场机器人学》**31**（2014年7月）。https://doi.org/10.1002/rob.21525[3] Bechar, A., Vigneault, C.：用于田间作业的农业机器人：概念与组件。《生物系统工程》**149**，94–111页（2016 年）。https://doi.org/https://doi.org/10.1016/j.biosystemseng.2016. 06.014, https://www.sciencedirect.com/science/article/pii/ S1537511015301914 [4] Behley, J., Garbade, M., Milioto, A., Quenzel, J., Behnke, S., Stachniss, C., Gall, J.：用于点云序列语义分割的数据集。CoRR **abs/1904.01416**（2019年），http://arxiv.org/abs/1904.01416[5] Brostow, G.J., Fauqueur, J., Cipolla, R.：视频中的语义对象类别：高分辨率真值数据库。《模式识别快报》**xx**（x），xx–xx（2008年）[6] Bulò, S.R., Porzi, L., Kontschieder, P.：用于深度神经网络内存优化训练的原位激活批归一化。CoRR **abs/1712.02616**（2017 年），http: //arxiv.org/abs/1712.02616[7]Cabon, Y., Murray, N., Humenberger, M.：Virtual KITTI 2。CoRR**abs/2001.10773**（2020 年），https://arxiv.org/abs/2001.10773[8] Caesar, H., Bankiti, V., Lang, A.H., Vora,S.,Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O.：nuScenes：用于自动驾驶的多模态数据集。CoRR **abs/1903.11027**（2019年），http://arxiv.org/ abs/1903.11027[9] Che, Z., Li, M.G., Li, T., Jiang, B., Shi, X., Zhang, X., Lu, Y., Wu, G., Liu, Y., Ye, J.：D2-City：多样化交通场景的大规模行车记录仪视频数据集。ArXiv **abs/1904.01975**（2019年）[10] Chen, Y., Yang, B., Liang, M., Urtasun, R.：学习用于深度补全的联合二维-三维表征。CoRR **abs/2012.12402**（2020 年），https: //arxiv.org/abs/2012.12402[11] Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benen- son, R., Franke, U., Roth, S., Schiele, B.：用于语义城市场景理解的 Cityscapes数据集。CoRR **abs/1604.01685**（2016年），http: // arxiv.org/abs/1604.01685[12]Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.：ImageNet：大规模分层图像数据库。载于：2009 IEEE计算机视觉与模式识别会议。第248–255页（2009年）。https: // doi.org/10.1109/CVPR.2009.5206848[13]Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.：PASCAL视觉目标类别

2012 年挑战赛（VOC2012）结果。http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html（2012 年）[14] Fortin, J.M., Gamache, O., Grondin, V., Pomerleau, F., Giguère, P.：林业作业中用于自主木材抓取的实例分割（2022 年）。https://doi.org/10.48550/ARXIV.2203.01902, https://arxiv. org/abs/2203.01902[15] Geiger, A., Lenz, P., Urtasun, R.：我们准备好自动驾驶了吗？KITTI 视觉基准套件。载于：计算机视觉与模式识别会议（CVPR）（2012 年）[16] Geyer, J., Kassahun, Y., Mahmudi, M., Ricou, X., Durgesh, R., Chung, A.S., Hauswald, L., Pham, V.H., Mühlegg, M., Dorn, S., Fernandez, T., J änicke, M., Mirashi, S., Savani, C., Sturm, M., Vorobiov, O., Oelker, M., Garreis, S., Schuberth, P.：A2D2：奥迪自动驾驶数据集。CoRR**abs/2004.06320**（2020 年），https://arxiv.org/abs/2004.06320[17] Grondin, V., Fortin, J.M., Pomerleau, F., Giguère, P.：基于深度学习的树木检测与直径估算。《林业：国际森林研究期刊》（2022年10月）[18]He, K., Gkioxari, G., Dollár, P., Girshick, R.B.：Mask R-CNN。CoRR**abs/1703.06870**（2017 年），http://arxiv.org/abs/1703.06870[19] Ioffe, S., Szegedy, C.：批归一化：通过减少内部协变量偏移加速深度网络训练。CoRR **abs/1502.03167**（2015 年），http://arxiv.org/abs/1502.03167[20] Jiang, P., Osteen, P.R., Wigness, M.B., Saripalli, S.：RELLIS-3D 数据集：数据、基准测试与分析。CoRR **abs/2011.12954**（2020 年），https: //arxiv.org/abs/2011.12954[21] Juliani, A., Berges, V., Vckay, E., Gao, Y., Henry, H., Mattar, M., Lange, D.：Unity：智能体通用平台。CoRR **abs/1809.02627**（2018 年），http://arxiv.org/abs/1809.02627[22] Lin, T., Maire, M., Belongie, S.J., Bourdev, L.D., Girshick, R.B., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.：Microsoft COCO：上下文中的常见物体。CoRR **abs/1405.0312**（2014 年），http://arxiv. org/abs/1405.0312[23] Mitra, A., Vangipuram, S.L.T., Bapatla, A.K., Bathalapalli, V.K.V.V., Mo- hanty, S.P., Kougianos, E., Ray, C.：关于智慧农业，你想了解的一切。CoRR **abs/2201.04754**（2022 年），https://arxiv.org/ abs/2201.04754[24] Mohan, R., Valada, A.：EfficientPS：高效全景分割。CoRR**abs/2004.02307**（2020 年），https://arxiv.org/abs/2004.02307[25] Nathan Silberman, Derek Hoiem, P.K., Fergus, R.：基于RGBD图像的室内场景分割与支撑关系推断。载于：ECCV（2012 年）[26] Neuhold, G., Ollmann, T., Bulò, S.R., Kontschieder, P.：用于街景语义理解的Mapillary Vistas数据集。载于：2017 IEEE国际计算机视觉会议（ICCV）。第5000–5009页（2017 年）。https://doi.org/10.1109/ICCV.2017.534 [27]Ringdahl, O.：林业自动化：无人集材机的研发。博士论文，于默奥大学（2011年5月）

[28] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M.S., Berg, A.C., Fei-Fei, L.: ImageNet 大规模视觉识别挑战。CoRR **abs/1409.0575**（2014 年），http://arxiv.org/abs/1409.0575[29]Sekachev, B., Manovich, N., Zhiltsov, M., Zhavoronkov, A., Kalinin, D., Hoff, B., TOsmanov, Kruchinin, D., Zankevich, A., DmitriySidnev, Markelov, M., Johannes222, Chenuet, M.,a andre, telenachos, Melnikov, A., Kim, J., Ilouz, L., Glazov, N., Priya4607, Tehrani, R., Jeong, S., Skubriev, V., Yonekura, S., vugia truong, zliang7, lizhming, Truong, T.: opencv/cvat：v1.1.0（2020 年 8 月）。https://doi.org/10.5281/zenodo.4009388, https: // doi.org/10.5281/zenodo.4009388[30] Shamshiri, R., Weltzien, C., Hameed, I., Yule, I., Grift, T., Balasundram, S., Pitonakova, L., Ahmad, D., Chowdhary, G.: 农业机器人技术的研究与发展：数字农业的视角。《国际农业与生物工程杂志》**11**，1–14（2018 年 7 月）。https:// doi.org/10.25165/j.ijabe.20181104.4278 [31] da Silva, D.Q., dos Santos, F.N.: ForTrunkDet——用于目标检测的可见光与热成像标注图像森林数据集。《成像杂志》（2021 年 8 月）。https://doi.org/10.5281/zenodo.5213825, https://doi. org/10.5281/zenodo.5213825[32] Sinha, R.K., Pandey, R., Pattnaik, R.: 计算机视觉任务中的深度学习：综述。CoRR **abs/1804.03928**（2018 年），http://arxiv.org/abs/ 1804.03928[33] Tan, M., Le, Q.V.: EfficientNet：重新思考卷积神经网络的模型缩放。CoRR **abs/1905.11946**（2019年），http://arxiv.org/abs/ 1905.11946[34] Wang, W., Zhu, D., Wang, X., Hu, Y., Qiu, Y., Wang, C., Hu, Y., Kapoor, A., Scherer, S.A.: TartanAir：一个旨在突破视觉同步定位与地图构建极限的数据集。CoRR **abs/2003.14338**（2020 年），https://arxiv.org/abs/2003.14338[35] Wigness, M., Eum, S., Rogers, J.G., Han, D., Kwon, H.: 一个用于非结构化户外环境中自主导航和视觉感知的 RUGD 数据集。见：国际智能机器人与系统会议（IROS）（2019年）[36] Yu, F., Xian, W., Chen, Y., Liu, F., Liao, M., Madhavan, V., Darrell, T.: BDD100K：一个具有可扩展标注工具的多样化驾驶视频数据库。CoRR **abs/1805.04687**（2018年），http://arxiv.org/abs/1805.04687