

MonteFloor: 扩展 MCTS 以重建准确的大型平面图

西尼萨·斯特科维奇¹, 马赫迪拉德¹, 弗里德里希·弗朗多弗¹, 文森特·莱佩特^{2,1}
¹计算机图形与视觉研究所, 格拉茨科技大学, 格拉茨, 奥地利

²Université Paris-Est, École des Ponts ParisTech, 巴黎, 法国

{sinisa.stekovic, rad, 欺诈者}@icg.tugraz.at, vincent.lepetit@enpc.fr

项目页面: <https://www.tugraz.at/index.php?id=52770>

抽象的

我们提出了一种从嘈杂的 3D 点云重建平面图的新方法。我们的主要贡献是一种有原则的方法, 它依赖于蒙特卡洛树搜索 (MCTS) 算法来有效地最大化合适的目标函数, 尽管问题很复杂。与之前的工作一样, 我们首先将输入点云投影到顶视图以创建密度图并从中提取房间建议。我们的方法选择并优化了这些房间提案的多边形形状共同即使对于大型复杂场景, 也可以拟合密度图并输出准确的矢量化楼层图。为此, 我们采用最初设计用于学习玩游戏的算法 MCTS, 通过最大化目标函数来选择房间提案, 该目标函数将适应度与深度网络预测的密度图相结合, 并对房间形状的术语进行正则化。我们还向 MCTS 引入了一个细化步骤, 用于调整房间提案的形状。对于这一步, 我们提出了一种新颖的可微分方法来渲染这些提议的多边形形状。我们在最近且具有挑战性的 Structured3D 和 Floor-SP 数据集上评估了我们的方法, 并显示了对现有技术有着显著的改进, 而没有对平面图配置施加任何硬性约束或假设。

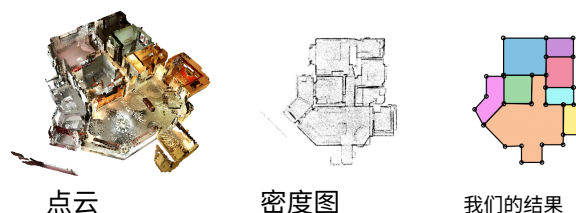


图 1: 给定密度图 I 在一个楼层的 3D 点云的顶视图中, 我们检索了一个准确的楼层地图, 该地图成功地恢复了各种房间形状。

地图 [36,39]。在这里, 我们专注于非结构化 3D 点云, 如 [2,7,23,26], 因为它们现在可以使用 RGB-D 相机轻松生成, 并且可以覆盖整个楼层。

要从给定的点云估计平面图, [7,23] 建议首先将点云投影到虚拟顶视图中以创建“密度图”, 因为墙壁 (用于创建楼层图的主要特征) 在密度图中显示相对清晰。如图 1, 密度图可能很嘈杂, 并且将房间表示为具有最少边数的矢量图仍然具有挑战性, 因为人类设计师特别是对于非曼哈顿楼层地图会这样做。为了处理这种复杂性, [7] 提出了一种基于图的解决方案, 该解决方案具有声能项, 但仍假设场景中存在一些主要的墙方向。

一、简介

从图像中理解场景是计算机视觉的主要主题之一, 因为它既旨在复制人类的一项关键能力, 又为机器人或增强现实等许多应用提供解决方案。我们在这里专注于创建结构化的平面图, 其中室内环境的每个房间都表示为一个多边形, 每面墙有一个边。已经考虑了许多类型的输入: 单眼透视彩色视图 [18,19,22,31], 全景 [32,38,40], 深度

在本文中, 我们还旨在从密度图估计平面图。我们的贡献是一种我们称之为 MonteFloor 的方法, 它在概念上简单而强大, 并返回高质量的平面图。数字 1 展示了 Floor-SP 测试集中的一个示例, 该示例演示了我们可以重建复杂的楼层地图, 包括具有复杂房间形状的非常大的地图, 而无需调整超参数。

像 [7], 我们的方法从 Mask-RCNN [17] 来自密度图。然而,

我们处理这些房间提案的方式与 [7]。在哪里 [7] 以贪婪的方式调整房间的墙壁和角落，我们选择正确的房间建议共同调整它们的位置和形状，由学习的评分函数指导。

这要归功于两个主要贡献。我们的第一个贡献是基于蒙特卡洛树搜索 (MCTS) 算法 [3,8]。MCTS 是一种随机算法，可以有效地探索搜索树，并已在例如 AlphaGo 和 AlphaZero 中用于在玩围棋或其他具有高组合数的游戏时选择移动 [30]。我们使用它在房间建议中搜索实际属于正确平面图的建议。在我们的例子中，移动对应于房间建议的选择。与其他树搜索算法相比，MCTS 仅基于叶子的评估。这意味着我们可以根据它们对密度图的解释程度来选择一组提案一起。在评估叶子之后，MCTS 更新访问节点中的分数，该分数将用于指导下一次树的探索。

为了评估一组建议对输入密度图的解释程度，我们引入了一个目标函数，该函数结合了由深度网络预测的“度量分数”和正则化项。该网络将密度图和所选提案的图像作为输入，以预测所选提案与地面实况之间的交集。正则化项鼓励选定的房间提案相互接触而不重叠，并且角度接近 90°。正好是 90°——请注意，这与强制执行曼哈顿世界条件不同，因为也接受其他角度。

此外，为了适应 MCTS 并获得准确的计划估计，我们通过在评估目标函数之前添加一个细化步骤来扩展它。该步骤执行目标函数的优化并调整所选房间建议的形状以更好地拟合密度图。我们的第二个贡献使这成为可能，这是一种优化 2D 多边形形状的新型可微分方法。请注意，最近，[16] 还使用 MCTS 来解决场景理解问题。但是，它提出了 MCTS 的简单应用。相比之下，我们依赖于适合我们问题的学习目标函数，并且我们引入了优化步骤来获得准确的估计。

虽然我们在这项工作中专注于平面图估计，但我们相信我们的方法是通用的，并且可以应用于其他场景理解问题，因为它的组件是通用的：我们从目标对象（在本例中为房间）的建议开始。由于我们基于 MCTS 的算法可以处理许多误报，因此此步骤不必执行良好即可获得良好的最终结果。该算法通过最大化数据驱动分数来寻找最终解决方案，因此可以轻松替换该分数以适应另一个问题。我们的

改进建议的解决方案更具体到 2D 多边形，但可以启发其他作者开发适合其目标对象的自己的方法。

评估我们的方法并将其与 Floor-SP [7]，这是我们问题的最新技术，我们首先在 Structured3D 数据集上进行实验 [37]，其中包含各种复杂的布局配置。与 Floor-SP 相比，我们在准确性和时间复杂度方面表现出显着改进（在 Structured3D 上重新训练他们的方法之后）。由于 Floor-SP 的作者无法为他们的的方法提供训练集（如他们的项目页面所述¹），我们无法重新训练我们的网络来预测专门针对该数据集的指标，我们必须使用在 Structured3D 上训练的那个。尽管存在这一领域差距，但我们在 Floor-SP 测试集上实现了更好的性能，而没有对平面图配置施加任何硬约束或假设。

二、相关工作

从 3D 数据创建平面图的早期方法依赖于基本的图像处理方法，例如直方图或平面拟合 [1,4,27,29,34,35]。例如，[27] 通过构建所有测量点的垂直位置的直方图来检测 3D 点云中的垂直平面来创建平面图。以类似的方式，[4] 通过应用扫描技术来识别曼哈顿世界方向，提取平面结构，从 3D 点云中的定位墙壁创建平面图。然而，这些技术严重依赖启发式算法，并且容易在噪声数据上失败。

后来通过使用图模型取得了重大进展，如 [5,10,11,12,20]。[10] 在体积 MRF 公式中使用图割优化。然而，所提出的方法容易受到噪声数据的影响，因为 MRF 中的正则化仅基于成对交互项。[20] 将 MRF 与稳健主成分分析相结合，以获得更紧凑的 3D 模型。图形模型也用于 [12] 从众包图像和位置数据中恢复布局和平面图。

基于图的方法定义了由表示计划元素的一元项和一次只涉及两个元素的二元项组成的目标函数（这里，元素主要是墙）。在我们的例子中，我们使用 MCTS 作为优化算法。MCTS 不对目标函数的形式施加限制，我们使用捕获复杂约束的目标函数。特别是，我们的目标函数的主要术语是同时考虑所有元素的深度网络。此外，我们通过添加细化步骤来补充 MCTS，以基于相同的目标函数调整元素的位置。

¹<https://github.com/woodfrog/floor-sp>

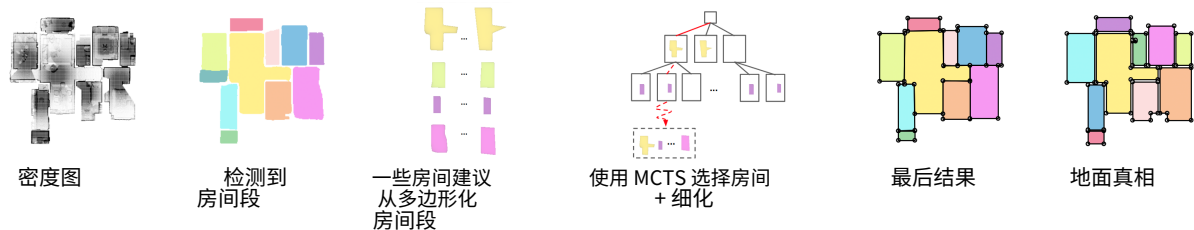


图 2: 我们的 MonteFloor 方法概述。给定一个 3D 点云, 我们首先创建一个楼层的密度图。然后我们使用 Mask-RCNN 检测房间段, 如 Floor-SP [7]。请注意左侧绿色部分底部的误报。我们以不同的方式对每个部分进行多边形化, 并从每个房间部分获得多个房间建议。我们依靠 MCTS 和我们的目标函数来选择正确的房间建议, 以及我们的细化步骤来联合调整房间建议的形状以适应输入密度图。

最近的工作依赖于其他优化技术[6,7,23]。然而, 这些技术面临的挑战是成本函数的定义和优化过程。其中一种方法称为 Floor-Net [23] 提出了一个深度网络, 用于从场景的给定密度图中检测可能的角落位置, 然后是整数规划公式。但是, 不正确的角落检测和错误检测会导致墙壁和房间丢失或增加。此外, 解决方案空间仅限于曼哈顿场景, 推广到非曼哈顿场景将导致更大的解决方案空间。相比之下, 我们的方法是可扩展的, 因为它依靠 MCTS 的效率来减少搜索空间, 并且可以考虑具有相同复杂性的曼哈顿和非曼哈顿场景。它通过全局优化选择最能解释输入的房间检测, 因此对误报不敏感。

我们方法的起点受到 Floor-SP 的启发 [7], 它建议首先分割房间实例, 然后通过顺序解决最短路径问题来重建房间的多边形表示。在他们的例子中, 离散密度图中的每个像素位置都是图中的一个节点, 可能属于房间的多边形曲线。错误的分割可能仍会导致不准确的平面图结构, 而我们会在早期处理不正确的房间分割。此外, Floor-SP 将房间的边缘方向离散化, 并为每个房间建模多个曼哈顿框架, 而我们的方法可以考虑任何角度。它仍然鼓励接近 90 的角度。正好是 90°, 当房间实际上遵循曼哈顿结构同时允许其他形状时, 这会产生更好的形状。正如我们将在实验中展示的那样, 我们的方法优于 Floor-SP 的准确性。

可微分渲染。3D 计算机视觉领域的一些作品已经表现出对可微渲染的兴趣 [13,14,15,25,28,33]。然而, 这些方法专注于 3D 表示的渲染, 例如点云、体素、网格和隐式 3D 表示。在

相比之下, 在这项工作中, 我们专注于 2D 表示的快速可微渲染, 即多边形, 并引入用于光栅化目的的可微渲染算法。

3.方法

数字 2 概述了我们的 MonteFloor 方法: 给定场景的 3D 点云, 我们首先创建该点云的顶视图密度图, 如后面部分所述 3.6。我们使用 Mask R-CNN [17] 训练以检测此类密度图中的房间, 然后我们将检测多边形化以获得一组房间建议。一些建议将至少粗略地对应于实际房间, 但其他建议只是误报。我们使用 MCTS 来找出哪些房间提案最适合输入密度图。MCTS 搜索由一个“度量网络”引导, 该网络经过训练可以预测所选房间建议与楼层地图地面实况之间的交集。由于来自 Mask R-CNN 的正确房间提议的形状仅与真实房间粗略对应, 因此我们在 MCTS 中执行搜索时优化了它们的形状。这是通过引入用于渲染多边形形状的可微分方法来完成的。

下面, 我们详细介绍:

- 我们究竟是如何获得房间建议的;
- 我们如何使用 MCTS 来选择房间建议;
- 我们的目标函数, 涉及我们的度量网络和正则化项;
- 我们如何在 MCTS 中改进房间建议的位置和形状;
- 我们如何精确计算给定 3D 点云的密度图。

3.1. 生成房间建议

我们训练了 Mask R-CNN [17] 在从 Structured3D 数据集的训练集创建的密度图上 [37] 从给定的密度图中提取单个房间段。虽然生成的细分质量很高, 但它们可以

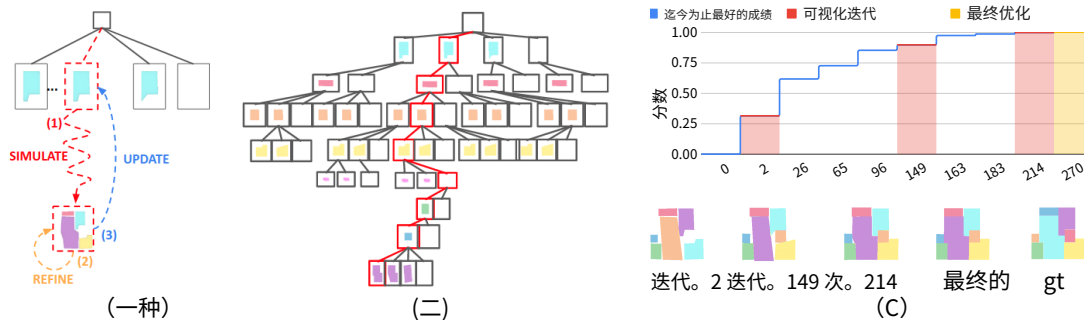


图 3：使用 MCTS 构建平面图树。(a) 在我们的例子中，一个节点对应于一个房间提案的选择，或跳过从一个房间段生成的所有房间提案。当一个节点第一次被访问时，MCTS 运行一个“模拟”步骤。这一步随机探索树的其余部分，直到到达叶子，在我们的例子中，当没有空间建议可以考虑时。当到达叶子时，我们执行我们的“细化步骤”，优化从根节点到叶子的路径中的房间提议的目标函数。目标函数的值用于更新路径中所有节点的预期分数。(b) MCTS 仅构建和探索树的一部分。与其他树搜索算法相比，MCTS 的剪枝仅基于对叶子的评估，这意味着我们可以根据它们对密度图的解释程度来选择一组建议。(c) 经过几次迭代后，我们的 MonteFloor 方法专注于并同时优化具有有希望的预期分数的解决方案。这使我们能够快速重建场景的准确平面图，大约在60此插图中使用的场景的秒数。

仍然包含误报，但是，它们将被 MCTS 过滤。数字2显示了房间段的示例以及我们从中生成的房间建议。我们将在下面详细介绍此过程。

有时，一个房间被检测为两个部分重叠的部分。因此，我们合并两个显著重叠的段（超过5%在实践中）进入一个额外的房间段，同时保留两个原始段。

在实践中，Mask-RCNN 提供的真阳性片段的形状与房间的确切形状并不对应，因为它们通常过于平滑。因此，我们将房间段多边形化以生成房间建议。为此，我们应用 Douglas-Peucker 多边形化算法 [9] 到房间段的轮廓。该算法取决于一个参数 ϵ 控制轮廓的简化。更准确地说，这个参数是原始曲线与其近似值之间的最大距离。

由于他的阶段的确切复杂
房间形状不详a tt性，我们生成 mul-
通过使用不同的值从每个细分中提取提案
为了 ϵ 。在实践中，我们取 $\epsilon = d \cdot \text{大号}$ ，在哪里 d 在预定义的
的集合中取不同的值 D 和大号是周长
段，与 $D = \{0.04, 0.02, 0.01\}$ 。有时，2个不同-
热切的 ϵ 导致相同数量的顶点，并且只有两个多 我们保持
边形之一。

即使在多边形化之后，真实位置的形状
可能的房间建议可能与实际房间形状不对应。调整
他们的形状，我们将优化它
通过我们的目标函数选择 们。我们将在下一小节描述提
M CTS 在 案，以及目标
功能函数 往后。

3.2. 使用 MCTS 进行房间建议选择

MCTS 是一种有效探索大树的算法，其中最大化的分数只能针对树的叶子进行评估。因此，我们对其进行了调整，以根据用作分数的目标函数来选择房间建议。我们将在下一小节中描述这个目标函数。

如图3，在我们的例子中，移动包括选择通过多边形化一个房间段生成的一个房间建议。对于每个房间段，还有一个额外的移动，包括不从该段中选择任何房间建议。根节点最多有 $|D|+1$ 孩子，对应于选择其中之一 $|D|$ 从第一个房间段发出的房间建议，并且没有从这个房间段中选择。

全树的节点数最多为 $(|D|+1)^k$
在哪里 k 是房间段的数量，并且，如 k 增加，遍历树中的所有路径很快变得不可行。幸运的是，MCTS 只会在需要时种植树，并避免进行详尽的评估。我们依靠标准的 Select-Expand-Simulate 算法。

更新 strategy，我们描述 be 低于 br 为松司飞米-
完整性。有关更详细的 de e, s 抄本 MCTS l-
算法， w 请参阅感兴趣的 r e 向 s 调查 [3] .

MCTS 一个 算法。MCTS 存储 es 在 n 颂歌 X-
预期的 SC (他们相信的补丁 和 使用它们 f 或者
指导 朝着新节点 i 的最佳选 C 化。前任 p 躺在贝洛 w,
期望 哪分数，并且可以更新 s 初始化 使用模拟 你-
升华 p 哪进一步 X 探索。在
每次迭代化，从 根节点，t 他的树是 tr 一种-

精通使用标准的置信上限 (UCB) 标准来选择每个节点。该标准取决于存储在节点中的预期分数，并平衡开发和探索。

当到达一个新节点时，MCTS 会执行一个模拟步骤来初始化这个节点的预期分数。这个模拟步骤随机探索树的其余部分，直到到达叶子，在我们的例子中，当不再需要考虑空间建议时。然后我们可以评估包含在从根节点到叶子的路径中选择的建议的解决方案的分数。我们将在下一段中解释我们如何计算这个分数。分数用于更新存储在路径节点中的预期分数。我们在补充材料中提供了有关我们实施 MCTS 的更多详细信息。

评分和细化步骤。 计算一个分数
与到达叶子时的路径相对应的解决方案，我们依赖于我们的目标函数，这将在下一小节中详细介绍。为了获得更准确的结果，除了标准的 MCTS 步骤外，我们还引入了优化目标函数的细化步骤，然后将其值作为解决方案的分数：房间提案的位置和形状可能与实际房间，如果没有这种细化，目标函数的值可能相对较低，不能很好地反映所选提案的实际质量。此细化步骤调整房间建议的位置和形状以获得更准确的解决方案。

目标函数。我们的目标函数可以写成：

$$\text{大号}(\text{磷}) = -\lambda_F F(D, F(\text{磷})) + \text{大号注册}(\text{磷}), \quad (1)$$

在哪里磷是一组用于评估解决方案的房间建议。 $F(D, F(\text{磷}))$ 是我们的度量网络，应用于输入密度图和房间建议的平面图磷，加权 λ_F 大号注册(磷)是正则化损失。我们将在接下来的两个小节中详细介绍这两个术语。我们用-大号(磷)作为 MCTS 最大化的分数。

最终解决方案推断。在 500 次 MCTS 迭代后，我们在具有最高预期分数的节点之后对树进行最终遍历，并通过最小化目标函数来优化所选提案。对于一些罕见的多边形，顶点小于5像素彼此分开。我们合并相应的顶点以获得最终的解决方案。

3.3. 公制网络

我们的度量网络 $F(D, F(\text{磷}))$ 评估一个集合的好坏磷选定的房间建议符合输入密度图。

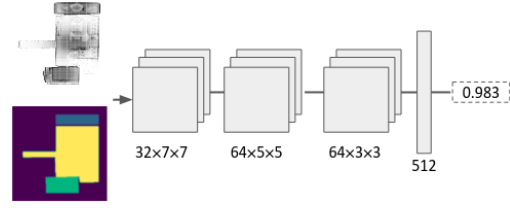


图 4：我们的度量网络。该网络采用密度图和地板表示，为可视化目的着色，作为输入，并输出一个分数，用于衡量地板图与输入密度图的匹配程度。我们训练它来预测估计的平面图和地面实况之间的交集。

如图4，这个网络结构简单，有两个输入：第一个输入 D 是密度图。第二个输入 $F(\text{磷})$ 是房间提案的图像，我们使用它们的索引作为像素值进行渲染：

$$F(\text{磷}) = \sum_{\text{一世}} \text{红外线}(\text{磷}-\text{一世}), \quad (2)$$

和 $R(\text{磷}-\text{一世})$ 是一个二值图像磷-一世, 里面的像素磷-一世设置为 1，其他设置为 0。

F 只输出一个值，它应该反映房间提议和密度图之间的适应度。我们训练它使用来自 Structured3D 数据集的训练数据以监督的方式预测所选房间提议和地面实况房间之间的 Intersection-over-Union (IOU)，用于密度图。37]。有关培训程序的更多详细信息，请参阅补充材料。

3.4. 正则化损失

正则化损失大号注册分解为：

$$\text{大号注册}(\text{磷}) = \lambda_{\text{昂}} \text{大号昂}(\text{磷}) + \lambda_{\text{球体}} \text{大号球体}(\text{磷}) + \lambda_0 \text{大号}_0(\text{磷}), \quad (3)$$

在哪里 $\lambda_{\text{昂}}$, $\lambda_{\text{球体}}$, 和 λ_0 加权三个术语。我们对所有场景使用相同的权重，并在补充材料中提供实际值。

大号昂(磷)规范房间建议的角度磷：

$$\text{大号昂}(\text{磷}) = - \frac{1}{|P|} \sum_{\text{磷}-\text{一世} \in \text{磷}} \frac{1}{|\text{磷}-\text{一世}|} \sum_{(\text{你}, \text{v}, \text{w}) \in \text{磷}-\text{一世}} \text{日志} p((\hat{u}, \text{v}, \text{w})), \quad (4)$$

在哪里 $|\text{磷}-\text{一世}|$ 表示多边形中的顶点数磷-一世, $(\text{你}, \text{v}, \text{w})$ 表示多边形的任意三个连续顶点

磷-一世, 和 $(\hat{u}, \text{v}, \text{w})$ 他们在顶点的角度 v . $p(\alpha)$ 是我们在房间角度上假设的先验分布。如图5，我们在余弦和均匀分布上使用高斯分布的混合。它不鼓励

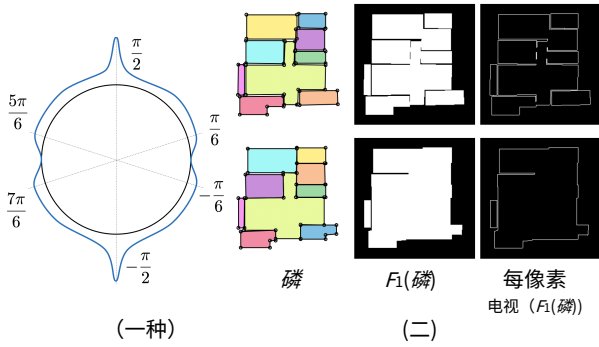


图 5: 正则化损失的可视化。(a) 事先分配 $p(\alpha)$ on angle 不鼓励平角并鼓励直角, 但其他角度仍然可以接受。(二) 大号球体是基于总变异。最佳: 当房间建议在磷没有接触或重叠, 总变化电视 ($F_1(\text{磷})$) 他们的形象 $F_1(\text{磷})$ 很大。底部: 当房间建议组合在一起时, Total Variation $TV(F_1(\text{磷}))$ 低得多。

平角 (0° 和 180°), 鼓励直角 (90° 和 270°), 以及之间的角度 $\pi/6$ 和 $5\pi/6$ 和之间 $7\pi/6$ 和 $-\pi/6$ 服从均匀分布。更正式地说, 我们取 $p(\alpha) =$

$$\begin{aligned} & -G \text{ (因为 } \alpha/\pi, \sigma_0) & \text{如果 } \alpha \in [-\pi/6, \pi/6], \\ & 1-\eta+G \text{ (因为 } \alpha/\pi, \sigma_2) & \text{如果 } \alpha \in [\pi/6, 5\pi/6], \\ & Z-G \text{ (因为 } \alpha/\pi, \sigma_1) & \text{如果 } \alpha \in [5\pi/6, 7\pi/6], \text{ 和} \\ & -\eta+G \text{ (因为 } \alpha/\pi, \sigma_2) & \text{如果 } \alpha \in [7\pi/6, \pi], \end{aligned} \quad (5)$$

在哪里 G 表示高斯分布, η 是常数 G (因为 $\pi/6$ 和 $\pi/6$), 和 Z 是归一化因子。在练习, 我们使用 $\sigma_1=0.1$ 和 $\sigma_2=0.08$ 。

大号球体(磷)鼓励房间提案相互联系而不重叠。如图所示5建议图像的总变化 (梯度绝对值的总和) 是一个很好的标准:

$$\text{大号球体} = \text{电视} (F_1(\text{磷})), \quad (6)$$

其中 TV 表示总变化, 并且 $F_1(\text{磷})$ 是计算为提案的图像

$$F_1(\text{磷}) = \sum_{\text{米} \in \text{磷}} R(\text{磷}-\text{米}). \quad (7)$$

数字5表明这种损失会惩罚重叠并将提案推向彼此, 并通过这样做, 在相邻房间的墙壁之间强制执行相似的方向。

大号0用于防止提案在优化期间偏离其初始位置。我们采取:

$$\text{大号0(磷)} = \frac{1}{|P|} \sum_{\text{磷}-\text{米} \in \text{磷}} \text{MSE}(R(\text{磷})-\text{米}-\text{米}) \quad (8)$$

在哪里米-世是生成提案的片段的二值图像磷-世, 和 $\text{MSE}(\cdot)$ 将此图像与二值图像进行比较 $R(\text{磷}-\text{世})$ 的提案。

3.5. 细化步骤和可微分多边形渲染

如前所述, 当 MCTS 到达叶节点时, 我们对等式中的目标函数执行几个优化步骤。(1) 在计算其值并将其用作 MCTS 的分数之前。在实践中, 我们使用 Adam 优化器 [21] 这个任务。

优化大号(磷), 我们需要使其可微。唯一不可微的部分是二进制图像创建 $R(\text{磷}-\text{世})$ 一个房间的建议磷-世, 在哪里磷-世表示为一个多边形。可微分渲染已经开发 [24], 然而, 可用的实现是为渲染 3D 三角形的网格而设计的。我们没有调整这些实现以使其在 2D 多边形上工作, 而是通过使绕组数算法可微来开发了一种更简单的方法。原来的绕组数算法检查一个像素位置是否米在多边形内磷-世通过计算:

$$W(\text{米}, \text{磷}-\text{世}) = \frac{1}{2\pi} \sum_{(\text{紫外线}) \in \text{磷}-\text{世}} \text{符号}(\text{确定}(\text{嗯}, \text{vm}))(\hat{u}mv), \quad (9)$$

在哪里 (紫外线) 是任意 2 个连续的顶点磷-世和确定 (·) 是向量的行列式嗯和虚拟机。标志 (·) 学期等于 1 如果角度 ($\hat{u}mv$) 在 $[0, \pi]$ 之间, 和 0 除此以外。因此, 对于一个有效的非相交、闭合和逆时针方向的多边形, $W(\text{米}, \text{磷}-\text{世}) \in [0, 1]$, 是具有值的阶跃函数 1 如果米在里面磷-世和 0 除此以外。

为了使其可微, 我们使用以下表达式代替:

$$W(\text{米}, \text{磷}-\text{世}) = \frac{1}{2\pi} \sum_{(\text{紫外线}) \in \text{磷}-\text{世}} \frac{C \cdot \text{确定}(\text{嗯}, \text{vm})}{1 + |C \cdot \text{确定}(\text{嗯}, \text{vm})|} (\hat{u}mv). \quad (10)$$

分数项实现了符号函数的软形式, 用于测量三角形的方向 (微博) 和 $C=1000$ 以可微分的方式逼近符号函数的阶跃形式。为了提高渲染效率, 我们只计算像素位置的缠绕值米位于 Mask R-CNN 为相应房间段检测到的边界框内。

3.6. 计算密度图

获取密度图 D 在场景中, 我们采用与 [7]。给定一组已注册的 RGB-D 全景图像, 我们生成一个点云

的场景。从场景点云的顶视图，我们将这些点投影到一个 256×256 图像空间，使得投影后的顶视图保持不变，完整的场景仍然可见。给定像素位置的密度值是投影到同一像素位置的点数。密度图的值最终归一化为范围 $[0,1]$ 。

4. 实验

在本节中，我们通过将其与 Floor-SP 进行比较来评估我们的方法 [7]，当前最先进的平面图重建，在两个数据集上。我们还提供了一项消融研究，以显示细化步骤对我们的方法的重要性。

4.1. 指标

为了评估恢复的平面图的质量，我们首先将恢复的房间与地面真实房间进行匹配。更准确地说，从最大的 ground truth 房间开始，我们找到了具有最大 Intersection-Over-Union (IOU) 值的匹配恢复房间。正如我们认为在 [7] 对于真正评估比较方法的质量过于宽松，我们对定量评估更加严格：

- 1.房间公制。该指标与 [7]。如果房间多边形不与其他房间重叠并且与真实房间匹配，则认为房间多边形已成功恢复。我们允许房间之间有一个像素重叠，因此我们不会惩罚与该指标接触的房间多边形。
- 2.角度量。如果一个角对应的房间多边形被成功恢复，并且它是与匹配的地面真实房间多边形中的任何一个角最近的角，在 10 个像素的距离内，则认为该角被成功恢复。该指标的灵感来自 [7] 没有考虑角是否真的属于正确的多边形。
- 3.角角公制。如果房间多边形的对应角被成功恢复并且与对应的ground truth角度的绝对差小于 5° 。

4.2. 与 Floor-SP 的评估和比较

结构化3D。我们对 Structured3D 数据集进行第一次评估 [37]，其中包含平面图注释 3500场景：3000训练场景，250验证场景，以及250测试场景。为了模仿标准的场景重建管道，我们投影了注册的 RGB-D

	房间		角落		角度		嘛	
	Prec	Rec	Prec	Rec	Prec	Rec	预录	
结构化3D								
DP ($\epsilon=0.01$)	0.93	0.94	0.74	0.79	0.49	0.52	0.89	0.88
地板-SP [7]	0.81	0.73	0.80	0.72	0.96	0.94	0.89	0.77
MonteFloor (我们的)	0.86	0.75						
[7] 测试集								
地板-SP [7]	0.85	0.83	0.72	0.58	0.65	0.52	0.74	0.64
MonteFloor (我们的)	0.88	0.85	0.78	0.63	0.68	0.54	0.78	0.67

表 1: Structured3D 的定量结果 [37]和来自[的测试集7]。MA 是三个指标（Room、Corner 和 Angle）的平均值。我们将我们的方法与 Mask-RCNN (DP) 和 Floor-SP 获得的房间段的简单 Douglas-Peucker 多边形化进行比较。7]。我们的方法稍微优于其他方法，即使我们无法在 [7]。

全景图像获取场景的点云。我们处理重建以获得 Mask R-CNN 和度量网络的训练数据。为了公平比较，我们重新训练了 Floor-SP 使用的网络，用于预测从 Structured3D 数据集生成的训练集上的角和边缘似然图，我们用同样在 Structured3D 上训练的 Mask R-CNN 网络替换了他们的 Mask R-CNN 网络。

Floor-SP 测试集。不幸的是，Floor-SP [7] 无法发布他们的 Floor-SP 数据集的训练场景，但我们可以在100 公开可用的测试场景，其中包括各种平面图配置。我们使用在作者提供的训练集上预训练的 Mask R-CNN 网络。然而，由于我们无法在 Floor-SP 训练集上训练我们的度量网络，我们不得不使用仅在 Structured3D 上训练的那个。因此，Floor-SP 方法在该数据集上具有优势。

桌子1显示了两个数据集的定量结果。为了更好地展示我们方法的好处，我们还展示了简单房间检测的结果，该结果是通过使用我们用来初始化房间提议的 Douglas-Peucker (DP) 方法对 Mask R-CNN 检测到的房间掩码进行多边形化。DP 在房间度量上获得了非常高的性能，这表明 Mask R-CNN 在大多数情况下都输出了高质量的掩码。然而，角度度量清楚地表明，这些多边形通常看起来不像实际的房间形状。

对于 Floor-SP，与 Douglas-Peucker 方法相比，房间指标有所下降。这与 Floor-SP 方法中的包含约束满足有关，该方法强制检索的多边形完全包含分割掩码。如果无法强制执行此约束，则重建也将失败。然而，更重要的是，角度度量清楚地表明他们的结果仍然优于 DP 获得的结果。

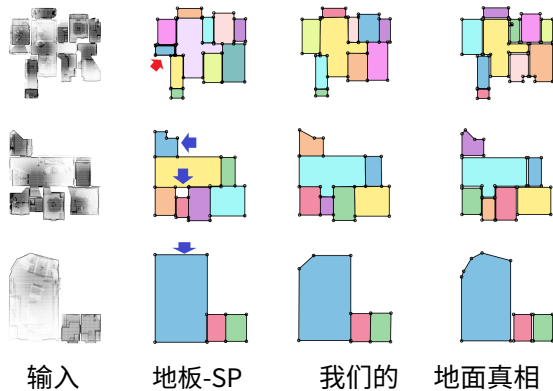


图 6: Structured3D 数据集的定性结果 [37], 最好在颜色中看到。红色箭头: 与 Floor-SP 相比, 我们的方法可以很好地处理误报检测。蓝色箭头: 与 Floor-SP 相比, 我们能够模拟更多种类的房间形状。

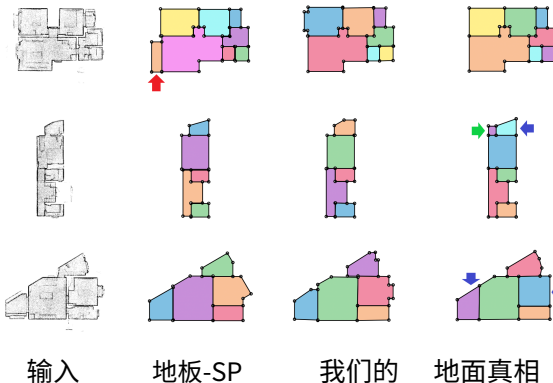


图 7: 测试集上的定性结果来自 [7]。即使我们的度量网络没有在 Floor-SP 训练集上进行训练, 我们的方法在 Floor-SP 测试集上的表现仍然略好于 Floor-SP。红色箭头: 我们删除了误报房间检测。绿色箭头: 地面实况中的紫色房间似乎是注释错误。蓝色箭头: 我们的重建有时比手动注释的房间更符合输入。

我们的方法在很大程度上优于两种基线方法, 因为我们在所有指标上都保持了非常高的性能。即使对于 Floor-SP 测试集也是如此, 即使我们无法在相应的训练集上重新训练我们的度量网络。

与 Douglas-Peucker 方法相比, 我们提高了房间指标的性能, 因为我们的细化步骤调整了可能最初重叠的房间建议的形状, 并且 MCTS 的选择消除了误报。与 Floor-SP 相比, 我们的方法受益于直接优化多边形形状, 这使我们能够避免

	房间		角落		角度		嘛	
	Prec	Rec	Prec	Rec	Prec	Rec	Prec	Rec
没有精炼。步长	0.95	0.93	0.86	0.76	0.65	0.57	0.82	0.75
无大号	0.96	0.94	0.86	0.75	0.73	0.68	0.85	0.85
无大号	0.84	0.78	0.69	0.74	0.66	0.79	0.73	0.92
无大号	0.84	0.72	0.88	0.80	0.94	0.22	0.89	0.10
无大号	0.10	0.96	0.94	0.89	0.77	0.94	0.89	0.77
完全的	0.86	0.75	0.90	0.82				

表 2: 消融研究。移除大对角度度量有很大影响; 移除大对角落的位置有很大的影响; 移除大可能会导致漂移。我们的度量网络 $f(\cdot)$ 对于 MCTS 的选择步骤至关重要, 因为其他术语不是平面图生成任务的合适评分函数。

掩码包含和角度离散化约束。

另外, 我们比较了两种方法在同一台机器上的执行时间。在 Structured3D 数据集上, Floor-SP 的平均计算时间为 785 ± 549 秒。相反, 我们的 MonteFloor 方法的平均计算时间是 71 ± 40 秒, 和 12 ± 8 秒跳过细化步骤时的秒数。我们对 Floor-SP 数据集进行了类似的观察。

定性结果。数据 6 和 7 展示了一些定性结果并证明我们的方法能够消除误报检测并检索平面图的高精度多边形重建。

消融研究。我们进行了消融研究, 以评估我们细化过程中每个单独术语的有效性。如表所示 2, 我们所有的正则化项都有助于检索位置和形状更好的房间多边形。主要消融表明度量网络在我们的方法中也起着至关重要的作用。如果没有度量网络, 目标函数不会强制与输入场景保持一致。然后在大场景中选择一个正确的房间可以最大限度地提高精度, 因为确实没有误报, 但可以最大限度地减少召回率。

5. 结论

我们提出了一种从 3D 点云估计平面图的方法。我们展示了如何将 MCTS 算法应用于此问题, 以及如何添加细化步骤以稳健的方式获得准确的计划。

除了平面图估计, 我们相信我们的方法是通用的。使其适应其他场景理解问题所需的只是 (1) 生成建议的方法和 (2) 用于评估解决方案质量的可微函数。我们希望我们的工作能够激发研究人员考虑对象之间复杂交互的问题, 并获得稳健和准确的结果。

确认。这项工作得到了 Christian Doppler 语义 3D 计算机视觉实验室的支持, 部分由高通公司资助。

参考

- [1] 安东尼奥·阿丹和丹尼尔·胡贝尔。遮挡和杂波下内墙表面的 3D 重建。在 *3D 成像、建模、处理、可视化和传输国际会议*, 2011 年。2
- [2] Armen Avetisyan、Tatiana Khanova、Christopher Choy、Denver Dash、Angela Dai 和 Matthias Nießner。SceneCAD: 预测 RGB-D 扫描中的对象对齐和布局。 *欧洲计算机视觉会议*, 2020 年。1
- [3] 卡梅隆·布朗、爱德华·鲍利、丹尼尔·怀特豪斯、西蒙·卢卡斯、彼得·考林、菲利普·罗尔夫沙根、斯蒂芬·塔文纳、迭戈·佩雷斯·利埃巴纳、斯皮里登·萨莫特拉基斯和西蒙·科尔顿。蒙特卡洛树搜索方法的调查。 *IEEE Transactions on Computational Intelligence and AI in Games*, 4:1, 2012。2,4
- [4] 安吉拉·布德罗尼和扬·博姆。从点云自动 3D 重建内部。 *国际建筑计算杂志*, 2010 年。2
- [5] Ricardo Cabral 和 Yasutaka Furukawa。从图像中重建分段平面和紧凑平面图。在 *计算机视觉与模式识别会议*, 2014 年。2
- [6] Yu-Wei Chao、Wongun Choi、Caroline Pantofaru 和 Silvio Savarese。使用几何和语义线索对高度杂乱的室内场景进行布局估计。在 *图像分析与处理*, 2013 年。3
- [7] 陈嘉诚、刘陈、吴嘉业、古川康孝。Floor-SP: 按顺序房间最短路径对平面图进行逆 CAD。在 *计算机视觉与模式识别会议*, 2019 年。1,2,3,6,7,8
- [8] 雷米库洛姆。蒙特卡洛树搜索中的高效选择性和备份算子。在 *计算机与游戏国际会议*, 2006 年。2
- [9] David H. Douglas 和 Thomas K. Peucker。减少表示数字化线或其漫画所需的点数的算法。 *制图: 国际地理信息和地理可视化杂志*, 1973 年。4
- [10] Yasutaka Furukawa、Brian Curless、Steven M. Seitz 和 Richard Szeliski。从图像重建建筑内部。在 *计算机视觉国际会议*, 2009 年。2
- [11] 高瑞鹏、赵明民、叶涛、叶凡、罗国杰、王益洲、边凯贵、王涛、李晓明。通过移动人群感应重建多层室内平面图。 *IEEE 移动计算汇刊*, 15(6), 2016。2
- [12] 高瑞鹏、赵明敏、叶涛、叶凡、王益洲、边凯桂、王涛、李晓明。Jigsaw: 通过移动人群感应重建室内平面图。在 *移动计算与网络国际会议*, 2014 年。2
- [13] 凯尔·热那亚、福雷斯特·科尔、Avneesh Sud、Aaron Sarna 和 Thomas Funkhouser。3D 形状的局部深层隐式函数。在 *计算机视觉与模式识别会议*, 2020 年。3
- [14] 格鲁吉亚 Gkioxari、Jitendra Malik 和贾斯汀约翰逊。网格 R-CNN。在 *计算机视觉国际会议*, 2019 年。3
- [15] Thibault Groueix、Matthew Fisher、Vladimir G Kim、Bryan C Russell 和 Mathieu Aubry。学习 3D 表面生成的纸浆方法。在 *计算机视觉与模式识别会议*, 2018 年。3
- [16] Shreyas Hampali、Sinisa Stekovic、Sayan Deb Sarkar、Chetan S. Kumar、Friedrich Fraundorfer 和 Vincent Lepetit。蒙特卡洛场景搜索以了解 3D 场景。在 *arXiv 预印本*, 2021 年。2
- [17] Kaiming He、Georgia Gkioxari、Piotr Dollár 和 Ross Girshick。掩码 R-CNN。在 *计算机视觉国际会议*, 2017 年。1,3
- [18] Varsha Hedau、Derek Hoiem 和 David Forsyth。恢复杂乱房间的空间布局。在 *计算机视觉国际会议*, 2009 年。1
- [19] H. Howard-Jenkins、S. Li 和 V. Prisacaru。跳出框框思考: 生成不受约束的 3D 房间布局。在 *亚洲计算机视觉会议*, 2019 年。1
- [20] 池畑聪、杨航和古川康孝。结构化室内建模。在 *计算机视觉国际会议*, 2015 年。2
- [21] Diederik P Kingma 和 Jimmy Ba。Adam: 一种随机优化方法。 *arXiv 预印本 arXiv:1412.6980*, 2014 年。6
- [22] Chen-Yu Lee、Vijay Badrinarayanan、Tomasz Malisiewicz 和 Andrew Rabinovich。Roomnet: 端到端房间布局估计。在 *计算机视觉国际会议*, 2017 年。1
- [23] 刘晨、吴家业、古川康孝。Floornet: 从 3D 扫描重建平面图的统一框架。在 *欧洲计算机视觉会议*, 2018 年。1,3
- [24] 马修·M·洛珀和迈克尔·J·布莱克。OpenDR: 近似可微渲染器。在 *欧洲计算机视觉会议*, 2014 年。6
- [25] Ben Mildenhall、Pratul P Srinivasan、Matthew Tancik、Jonathan T Barron、Ravi Ramamoorthi 和 Ren Ng。NeRF: 将场景表示为用于视图合成的神经辐射场。在 *欧洲计算机视觉会议*, 2020 年。3
- [26] Srivathsan Murali、Pablo Speciale、Martin R Oswald 和 Marc Pollefeys。Indoor scan2bim: 房屋内部的建筑信息模型。在 *智能机器人与系统国际会议*, 2017 年。1
- [27] Brian Okorn、Xuehan Xiong、Burcu Akinci 和 Daniel Huber。走向平面图的自动建模。在 *3DPVT*, 2009 年。2
- [28] Nikhila Ravi、Jeremy Reizenstein、David Novotny、Taylor Gordon、Wan-Yen Lo、Justin Johnson 和 Georgia Gkioxari。使用 PyTorch3D 加速 3D 深度学习。 *arXiv:2007.08501*, 2020 年。3
- [29] 维克多·桑切斯和阿维德·扎科尔。从点云数据对建筑内部进行平面 3D 建模。在 *图像处理国际会议*, 2012 年。2

- [30] David Silver、Thomas Hubert、Julian Schrittwieser、Ioannis Antonoglou、Matthew Lai、Arthur Guez、Marc Lanctot、Laurent Sifre、Dharshan Kumaran、Thore Graepel、Timothy Lillicrap、Karen Simonyan 和 Demis Hassabis。一种掌握国际象棋、将棋和自我对弈的通用强化学习算法。《科学》, 362(6419), 2018. [2](#)
- [31] Sinisa Stekovic、Shreyas Hampali、Mahdi Rad、Sayan Deb Sarkar、Friedrich Fraundorfer 和 Vincent Lepetit。通过渲染和比较从单一视图中获得一般 3D 房间布局。在《欧洲计算机视觉会议》, 2020 年。 [1](#)
- [32] Cheng Sun, Chi-Wei Hsiao, Min Sun, and Hwann-Tzong Chen. Horizo nNet: 具有一维表示和全景拉伸数据增强的学习室布局。在《计算机视觉与模式识别会议》, 2019 年。 [1](#)
- [33] 王南洋, 张银达, 李竹文, 付艳伟, 刘伟, 姜玉刚。Pixel2Mesh: 从单个 RGB 图像生成 3D 网格模型。在《欧洲计算机视觉会议》, 2018 年。 [3](#)
- [34] 肖建雄、古川康孝。重建世界博物馆。《国际计算机视觉杂志》, 110(3), 2014。 [2](#)
- [35] Xuehan Xiong、Antonio Adan、Burcu Akinci 和 Daniel Huber。从激光扫描仪数据自动创建语义丰富的 3D 建筑模型。《建筑自动化》, 2013 年 31 月。 [2](#)
- [36] 张健、陈侃、Alexander G Schwing 和 Raquel Urtasun。从深度传感器估计室内场景的 3D 布局及其杂波。在《计算机视觉国际会议》, 2013 年。 [1](#)
- [37] 贾政, 张俊飞, 李静, 唐睿, 高盛华, 周子涵。Structured3D: 用于结构化 3D 建模的大型逼真数据集。在《欧洲计算机视觉会议》, 2020 年。 [2,3,5,7,8](#)
- [38] Chuhan Zou、Alex Colburn、Qi Shan 和 Derek Hoiem。LayoutNet: 从单个 RGB 图像重建 3D 房间布局。在《计算机视觉与模式识别会议》, 2018 年。 [1](#)
- [39] 邹楚航、郭瑞琪、李志忠、德瑞克·霍伊姆。从 RGBD 图像完成 3D 场景解析。《国际计算机视觉杂志》, 2019 年。 [1](#)
- [40] Chuhan Zou, Jheng-Wei Su, Chi-Han Peng, Alex Colburn, Qi Shan, Peter Wonka, Hung-Kuo Chu, and Derek Hoiem。从单个 360 度图像重建 3D 曼哈顿房间布局。《国际计算机视觉杂志》, 2019 年。 [1](#)