

## Floor-SP: 平面图的逆 CAD 顺序房间最短路径

陈家诚<sup>1</sup>

<sup>1</sup>西蒙弗雷泽大学

{jca348,古川}@sfu.ca

陈柳<sup>2</sup>

吴家业<sup>2</sup>

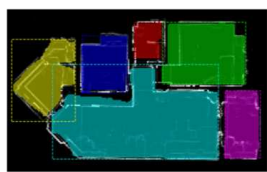
<sup>2</sup>圣路易斯华盛顿大学

{chenliu,jiaye.wu}@wustl.edu

古川康孝<sup>1</sup>



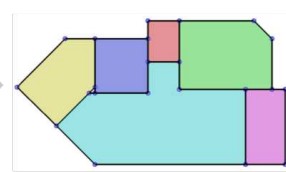
Aligned panorama RGBD scans



Raster room segments



Vectorized room polygonal loops



Vector-graphics floorplan

图 1. 建议的系统, 称为 *地板-SP*, 将对齐的全景 RGBD 扫描作为输入, 找到房间段, 解决优化问题以将平面图重建为多个多边形环 (每个房间一个), 并通过简单的后处理启发式将它们合并为 2D 图。优化是本文的技术贡献, 它采用逐室坐标下降策略, 依次求解最短路径问题来优化房间结构。

### 抽象的

本文提出了一种从 RGBD 扫描中自动重建平面图的新方法, 这是室内地图研究的一个重要里程碑。这种被称为 *Floor-SP* 的方法提出了一个新的优化问题, 其中按房间的坐标下降顺序解决最短路径问题以优化平面图结构。目标函数由深度神经网络引导的数据项、鼓励相邻房间共享角落和墙壁的一致性项以及模型复杂性项组成。与大多数其他方法不同, 该方法不需要角/边缘图元提取。我们对 527 套公寓或房屋的生产质量 RGBD 扫描评估了我们的系统, 其中包括许多具有非曼哈顿结构的单元。定性和定量评估表明, 与当前最先进的技术相比, 性能有了显著提升。代码和数据请参考我们的项目网站<http://jcchen.me/floor-sp/>。

### 一、简介

建筑平面图在设计、理解和改造室内空间方面发挥着至关重要的作用。从原始传感器数据自动重建平面图是室内地图研究的一个重要里程碑。核心技术挑战在于墙图结构的推断, 其拓扑结构未知且因示例而异。

计算机视觉在图推理任务中取得了显著进展, 例如, 人体姿态估计

化[3]和手部跟踪[30]。不幸的是, 成功仅限于固定已知拓扑的情况 (例如, 一个人有两条手臂)。具有未知变化拓扑的图结构推断仍然是一个悬而未决的问题。

图重建的一种流行方法是原始检测和选择[11,27,22], 例如, 检测角点, 选择角点的子集形成边缘, 以及选择边缘的子集形成区域。这种自下而上过程的主要问题是它无法从早期阶段的单个假阴性 (即缺失的原语) 中恢复。随着原始空间的自由度呈指数增长, 这项任务变得越来越困难, 特别是对于大多数现有方法无法处理的非曼哈顿场景[11、2、21、20]。

本文旨在通过三个关键思想在平面图重建领域取得突破。

- 首先, 我们通过实例语义分割技术从房间分割开始 (我们使用 Mask-RCNN [12])。房间分割将平面图推断减少为多个多边形环的重建, 每个房间一个。这种减少使我们能够将平面图重建制定为由房间建议引导的多个循环上的声音能量优化。

- 第二, 我们雇用 *房间坐标下降* 优化目标函数的策略。通过利用房间拓扑是一个简单的循环这一事实, 我们的公式通过依次为每个房间一个接一个地解决最短路径问题来找到 (接近) 最优图结构, 同时强制与其他房间保持一致性。

• 第三，我们利用深度神经网络评估优化问题的数据项，测量与输入传感器数据的差异。数据项与 ad-hoc 1) 一致性项相结合，鼓励相邻房间共享房间边界处的角落和墙壁，以及 2) 模型复杂性项，惩罚图中角落的数量。

我们已经评估了对 527 套公寓或房屋的生产质量 RGBD 扫描提出的方法，比当前最大的数据库 [20] 大几倍。我们的方法对当前最先进的技术 [20] 进行了重大改进。我们参考我们的项目网站 <http://jcchen.me/floor-sp/> 获取代码和数据。

## 2. 相关作品

我们讨论两个领域的相关工作：图重建和室内扫描数据集。

图重建：图结构推断一直是计算机视觉中的热门研究领域，例如推断人体姿势 [3] 或类别的语义关系 [14, 28]。在这些问题中，图拓扑是在标签空间上定义的，对所有实例都是通用的（例如，头总是连接到一个身体）。我们在这里关注重构上下文中的图推理问题，其中拓扑因实例而异。

房间布局估计从单个图像中推断出建筑特征线的图形，其中节点是房间角落，边缘是墙壁边界。大多数方法假设一个 3D 盒子房间来限制 2D 图像中可见的房间布局的拓扑变化 [13、25、18、5]。对于超出盒子形状的房间，应用动态规划 (DP) 来搜索最佳房间结构 [8, 9]。DP 类似地用于通过将拓扑结构限制为循环来解决平面布置图 [2]。

自下而上处理是一种流行的图重建方法，其中检测角等低级基元，然后选择这些基元以形成边缘或区域等高级基元。基于 DNN 的结点检测器被提出用于平面图图像矢量化 [21]，其中结点指示曼哈顿框架中的入射边缘方向。连接信息用于通过整数规划 (IP) 推断边缘。同样，黄等人 [16] 使用 DNN 检测由一组入射边缘方向表示的连接点，并通过启发式方法推断边缘，用于人造场景的单图像线框重建。

虽然许多以前的工作利用 RGBD 扫描/点云进行高质量的室内重建 [17、19、23、20]，但 FloorNet [20] 是目前在大规模室内测试的平面图重建任务的最新技术基准。FloorNet 在自下而上的过程中结合了 DNN 和 IP，但它具有三种主要的故障模式。首先，就像在任何

自下而上的过程中，检测阶段缺少角落会自动导致最终模型中缺少墙壁和房间。其次，错误的候选基元可能导致无关的墙壁和房间的重建。第三，为了能够使用强大的 IP，FloorNet 需要将解决方案空间限制在曼哈顿场景。

Ikehata 的结构化室内模型等人 [17] 是我们工作的灵感来源，它从房间分割开始，然后解决最短路径问题以重建房间形状，然后是房间合并和房间添加。虽然他们的系统是一系列用于室内建模的启发式算法，但我们的方法制定了一个合理的能量最小化问题来恢复平面图结构。

室内扫描数据集：经济实惠的深度传感硬件使研究人员能够构建许多室内扫描数据集。ETH3D 数据集包含 16 次室内多视图立体扫描 [24]。ScanNet 数据集 [6] 和 SceneNN 数据集 [15] 捕获了各种室内场景。然而，他们的大多数扫描只包含一两个房间，不适合平面图重建问题。Matterport3D [4] 为 90 座豪宅构建了高质量的全景 RGBD 图像集。2D-3D-S 数据集 [1] 使用相同的 Matterport 系统提供 6 次大规模室内办公空间扫描。最后，一个大型合成数据集 SUNCG [26] 提供了各种室内场景。

对于平面图重建任务，FloorNet [20] 为 155 个住宅单元提供了完整的平面图注释和来自智能手机的相应 RGBD 视频的基准。本文利用具有平面图注释的 527 栋房屋或公寓的生产质量全景 RGBD 扫描。

## 3. Floor-SP：系统概述

Floor-SP 分三个阶段将对齐的全景 RGBD 图像转换为平面图：房间分割、房间感知平面图重建和循环合并（见图 2）。本节提供了系统概述和最少的细节。对齐的全景 RGBD 扫描首先转换为 2D 点密度/法线图，这是 Floor-SP 的输入。与 FloorNet [20] 不同，我们专注于墙壁结构，在给定适当的墙壁结构的情况下，可以添加门/窗、图标和房间语义。

房间分割：输入全景扫描转换为 4 通道  $256 \times 256$  自上而下视图中的点密度/法线贴图（参见第 6 节）。我们利用实例语义分割技术（Mask R-CNN [12]）在给定 4 通道图像的情况下找到房间段。房间段通过提供大致形状的房间提议为平面图重建奠定了良好的基础，但它们距离良好的平面图还很远，因为 1) Mask R-CNN 段具有栅格表示（即未知数量和位置的角落）；和 2) 墙壁是

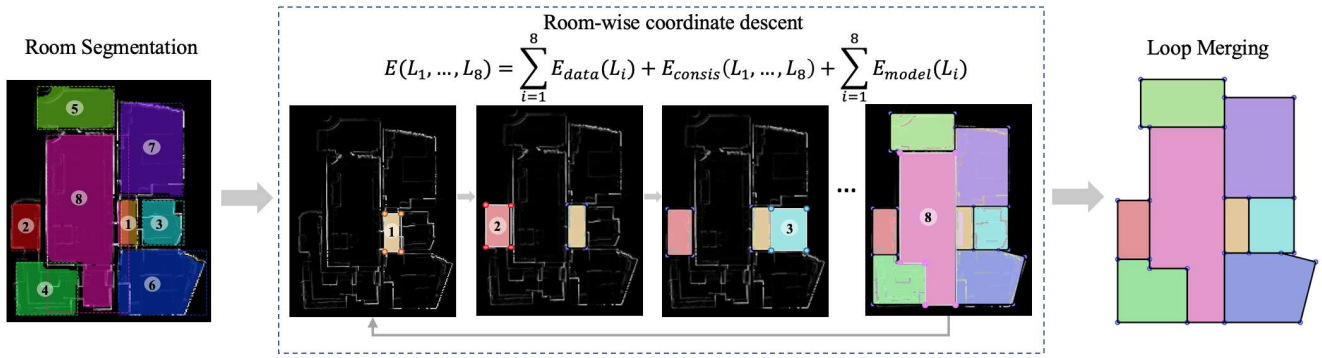


图 2. 系统概述：（左）Mask-RCNN 从由点密度和平均表面法线组成的自上而下的投影图像中找到房间段（光栅），使我们能够将平面图重建为多个房间循环。（中）逐室坐标下降通过最小化数据、一致性和模型复杂性项的总和来逐一优化矢量化房间结构。（右）简单的图合并操作将循环组合成一个平面图结构。

并非始终在房间之间共享。

房间感知平面图重建：给定一组房间段和输入点密度/法线图，我们制定了一个优化问题，将平面图重建为多个多边形环，每个房间一个。深度神经网络在目标中导出数据项。我们提出了一种新颖的按房间坐标下降算法，通过顺序解决最短路径问题直接优化角的数量和位置。

循环合并：简单的图形合并操作将多个多边形环组合成最终的平面图。

房间感知平面图重建是本文的技术核心，其中 Sect. 4 定义了问题的表述和 Sect. 图 5 给出了优化算法。房间分割和循环合并基于现有技术，其中 Sect. 图 6 提供了它们的算法细节和剩余的系统规范。

#### 4. 房间感知平面图重建

房间分割（ $R_{\text{seg}}$ ）来自 Mask R-CNN 允许我们将平面图推断减少为多个循环的重建（ $R_{\text{seg}}$ ），每个房间一个。 $R_{\text{seg}}$  定义为整数坐标处的像素序列，形成具有环形拓扑的多边形曲线。我们的问题是最小化关于多边形环集的以下目标大数：

$$\sum_{R_{\text{seg}} \in L} \mathcal{E}_{\text{数据}}(R_{\text{seg}}) + \mathcal{E}_{\text{一致}}(L) + \sum_{R_{\text{seg}} \in L} \mathcal{E}_{\text{模型}}(R_{\text{seg}}),$$

受制于  $R_{\text{seg}}$  是一个包含  $R_{\text{seg}}$  里面。请注意，房间有任何数量的角（即自由度），必须通过算法进行优化。

数据术语： $\mathcal{E}_{\text{数据}}$  是房间一元电位，测量与输入传感器数据在集合上的差异

沿每个循环的像素。

$$\mathcal{E}_{\text{数据}}(R_{\text{seg}}) = \sum_{p \in C(R_{\text{seg}})} \lambda_1 \mathcal{E}_{\text{数据}}(p) + \sum_{p \in E(R_{\text{seg}})} [\lambda_2 \mathcal{E}_{\text{数据}}(p) + \lambda_3 \mathcal{E}_{\text{数据}}(p)].$$

•  $\mathcal{E}_{\text{数据}}(p)$  是在像素处放置一个角的惩罚  $p$ （见图 3a），因此，对所有角像素求和  $C(R_{\text{seg}})$  在大号  $R_{\text{seg}}$ 。惩罚定义为 1 减去逐像素角的可能性。我们使用扩张残差网络（DRN）[29] 从输入点密度/法线图估计角似然图。

•  $\mathcal{E}_{\text{数据}}(p)$  是在像素上放置边缘的惩罚  $p$ 。该术语被定义为 1 减去逐像素边缘似然性（见图 3b），对所有边缘像素求和  $E(R_{\text{seg}})$  沿着大号  $R_{\text{seg}}$ 。我们使用 Bresenham 的线算法来获得给定角的边缘像素。相同的 DRN 从输入点密度/法线图估计边缘似然性。

•  $\mathcal{E}_{\text{数据}}(p)$  也是对边缘像素求和的惩罚，强制执行大号  $R_{\text{seg}}$  不通过房间段  $R_{\text{seg}}$ 。如果像素属于任何房间段，则该项是一个大常数，否则为 0。

一致性术语： $\mathcal{E}_{\text{一致}}$  是一个房间的高阶势，鼓励循环在房间边界处保持一致（即，共享角落和边缘）。我们将惩罚定义为像素数被使用所有循环的角（或边缘）在一起。例如，如果两个角彼此靠近，则该术语建议移动

它们到相同的像素，因此只施加一次惩罚：

$$\mathcal{E}_{\text{一致}}(L) = \sum_p [\lambda_4 1c(p, L)] + \sum_p [\lambda_5 1z(p, L)]$$

第一学期  $1c(p, L)$  是一个指示函数，如果一个像素  $(p)$  是至少一个循环的角，则它变为 1。同样，第二项是边的指示函数。有关玩具示例的说明，请参见图 3。

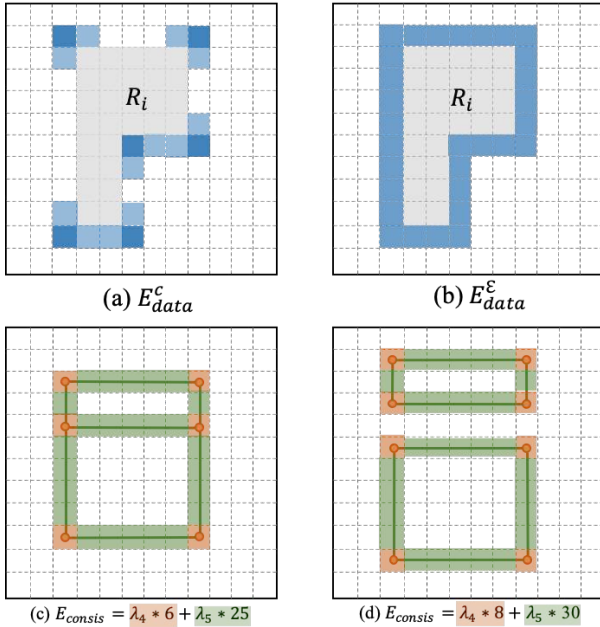


图 3. 数据和一致性术语的说明。  
乙数据是基于角和边缘似然图定义的。  
在这些玩具示例中，蓝色像素表示成本较低。乙一致  
计算房间角落和房间边缘使用的像素数。当相邻房间共享角落和  
边缘时，如 (c) 所示，乙一致下跌降落。

模型复杂度项：乙模型是模型复杂度损失，计算循环中的  
角数，更喜欢紧凑的形状。

$$\text{乙模型 (大号-世)} = \lambda_6 \{\# \text{ 中的角大号-世}\}.$$

$\lambda_i$  是定义惩罚项的相对权重的标量。我们发现我们的  
系统对这些参数具有鲁棒性，并在整个实验中使用以下设  
置： $\lambda_1=0.2, \lambda_2=0.2, \lambda_3=100.0, \lambda_4=0.2, \lambda_5=0.1, \lambda_6=1.0$ 。

## 5. 连续房间最短路径

我们优化策略的灵感来自先前的工作，该工作解决了  
最短路径问题并将平面图重建为循环 [2]。该公式将每个  
像素视为图的一个节点，将目标编码为边权重，并找到最  
短路径作为循环。

我们的问题解决了多个房间的多个循环。我们设计逐  
室坐标下降策略通过将逐室坐标下降步骤减少为最短路  
径问题，依次优化房间结构。虽然该算法对处理顺序具有  
鲁棒性，但我们以区域的递增顺序访问房间（即首先处理  
较小的房间），以便在相同输入的情况下获得固定结果。  
在我们的实验中，优化运行了两轮。

本节解释 1) 最短路径问题减少；  
2) 包含约束满足；3) 速度提升的两种近似方法。

最短路径问题减少：减少过程很简单，因为我们的成本函  
数是像素惩罚和角点数量的总和。不失一般性，假设我们  
正在优化大号<sub>1</sub>同时修复其他循环。我们的优化问题等价  
于解决最短路径问题 $R_1$ 每个边的权重定义如下 (e) (推导  
见补充文件)：

$$\begin{aligned} & \sum_{p \in C(e)} \frac{\lambda_1 Z_c}{2} \text{数据}(p) + \\ & \sum_{p \in E(e)} \left[ \lambda_2 Z_{\text{数据}}(p) + \lambda_3 Z_{\text{世数据}}(p) \right] + \\ & \sum_{p \in C(e)} \lambda_4 (1 - 1_c(p, \text{大号} \setminus \{\text{大号}_1\})) + \\ & \sum_{p \in E(e)} \lambda_5 (1 - 1_z(p, \text{大号} \setminus \{\text{大号}_1\})) + \lambda_6. \end{aligned}$$

滥用符号， $C(e)$  表示端点的两个像素 $e, E(e)$ 表示沿的像  
素集 $e$ 通过 Bresenham 的线算法获得，并且左 $\{\text{大号}_1\}$ 表  
示循环的集合，不包括大号<sub>1</sub>。

包含约束满足：最短路径是一种强大的公式，它搜索角  
的最佳数量和位置，但有一个警告：需要一个额外的约束来  
避免琐碎的解决方案（即空循环）。我们使用与先前工作  
[2] 在精神上相似的启发式方法来实现此约束：“大号-世  
包含（或四处走动） $R_{\text{世}}$ ”。我们将详细信息参考补充文  
件并在此总结过程。

首先，我们从用于数据项的相同角点似然图中找到候  
选角点（见图 4）。其次，我们查看边缘似然图来识别形  
成开始边缘的循环。第三，我们画一个起跑线从房间面具  
开始 ( $R_{\text{世}}$ )并在其中点垂直通过起始边。最后，我们删  
除所有与起始线相交的边，以确保路径必须绕行 $R_{\text{世}}$ 。

请注意，将起始边缘固定为循环的一部分会破坏我们  
的坐标下降步骤的局部最优性，但在实践中效果很好，因  
为不难以高置信度识别一个墙段。

边界框近似：我们对修剪节点和边进行了近似，以减少最  
短路径算法 (SPA) 的计算开销。我们限制了 SPA 的  
域，因为在整个图像域上运行它来重建一个房间是很浪费  
的。给定一个房间面具 $R_{\text{世}}$ ，我们应用二元膨胀 10 次来展  
开



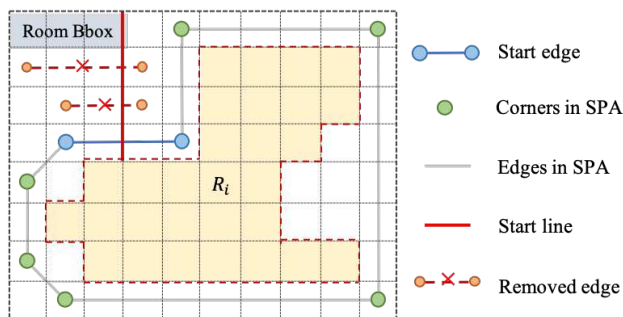


图 4. 我们解决了每个房间的最短路径问题，其中成本函数被编码为边缘权重。为了避免琐碎的解决方案（即空图）并强制路径绕过粗糙的房间段（ $R_i$ ），我们首先以高置信度确定作为房间形状一部分的起始边。接下来，我们垂直绘制一条（红色）起始线来分割域。我们禁止越过起跑线，通过处以非常高的处罚  $R_i$ ，然后求解在起始边的两个端点处开始和结束的最短路径。

掩码并找到它的轴对齐边界框，边距为 5 像素，我们在其中求解 SPA。

主导方向近似：Floor-SP 超越了传统的曼哈顿假设，允许每个房间有多个曼哈顿框架。我们训练相同的 DRN 架构以在每个像素处以 10 度的增量估计墙壁方向的可能性。我们执行一个简单的统计分析来全局提取四个曼哈顿框架（即八个方向），然后将其子集分配给每个房间。我们只允许沿选定的主要方向的边缘对离散化误差有一定的容忍度（有关详细信息，请参阅补充文档）。

## 6. 系统详情

输入处理：给定一组全景 RGBD 扫描，其中 Z 轴与重力方向对齐，我们计算水平面上点的紧密轴对齐边界框。我们将矩形扩展为 2.5% 在四个方向的每一个上，将非均匀缩放应用到  $256 \times 256$  像素网格，并计算每个像素中的点密度和法线。点密度是落在像素内的 3D 点的数量，我们将其线性重新缩放到  $[0.0, 1.0]$  以便最高密度变为 1.0。点法线是与像素关联的 3D 点的平均表面法线向量。

房间分割：我们使用具有默认超参数的公开可用的 Mask R-CNN 实现 [7]，只是我们将检测阈值从 0.7 降低到 0.2。给定来自 Mask R-CNN 的片段，我们应用二元腐蚀操作进行 2 次迭代，使用 8 个连通邻域来获得房间片段（ $R_i$ ）。

房间感知平面图重建：估计像素-

对于角、边缘和边缘方向的明智可能性，我们使用 Dilated Residual Networks [29] 的官方实现，它产生  $32 \times 32$  特征图。为了产生与输入具有相同分辨率的输出，我们添加了 3 个额外的残差块 [10] 层，具有步长为 2 的转置卷积，以达到  $256 \times 256$ 。对于角似然监督，我们将每个真实角点渲染为  $7 \times 7$  磁盘。对于边缘似然和墙方向监督，我们绘制边缘掩码和方向信息，宽度为 5 像素。损失是二元交叉熵，学习率为  $1e-4$ 。Dijkstra 算法解决了最短路径问题。

循环合并：我们使用简单的图合并操作将房间循环转换为最终的平面图结构。更具体地说，我们将一组连续的共线线段表示为段组。我们重复识别 5 个像素内的一对平行线段组，并在合并角点的同时将它们捕捉到中间点的新线段组。在对所有兼容对应用边缘合并后，我们合并 3 像素以内的角。

## 7. 实验

我们已经在 527 组对齐的全景 RGBD 扫描上评估了所提出的系统。1) 点密度/法线图像的输入 3D 点、2) 注释中的角点、3) 注释中的墙段和 4) 注释中的房间的平均数量为 432,552、28.87、35.88 和 7.73，分别。在 4072 间客房中，有 489 间客房不遵循该单元的主要曼哈顿结构。图 5 显示了我们数据集中的四个示例。

527 个单元被分成 433 个和 94 个，分别用于训练和测试。我们故意使测试集更具挑战性以进行评估：94 个测试单元中的 48 个包含具有挑战性的非曼哈顿结构，667 个测试室中的 199 个遵循非曼哈顿几何结构。

我们在 Python 中实现了建议的系统，同时使用 PyTorch 作为 DNN 库。我们使用了配备 NVIDIA 1080Ti 和 12GB GPU 内存的工作站。我们对 Mask-RCNN 训练了 70 个 epoch，batch size 为 1，DRNs 训练了 35 个 epoch，batch size 为 4。每个 DNN 模型的训练最多需要一天时间。在测试时，处理一套公寓/房屋大约需要 5 分钟。瓶颈是最短路径问题（CPU 密集型）的图构建。

### 7.1. 定性评价

图 6 将 Floor-SP 与当前最先进的 FloorNet [20] 和我们系统的变体进行了比较。Floor-Net 遵循自下而上的过程，它首先检测角点，然后使用整数规划找到它们的有效连接。FloorNet 存在三种故障模式：1) 由于第一个角落的角落缺失而导致房间缺失

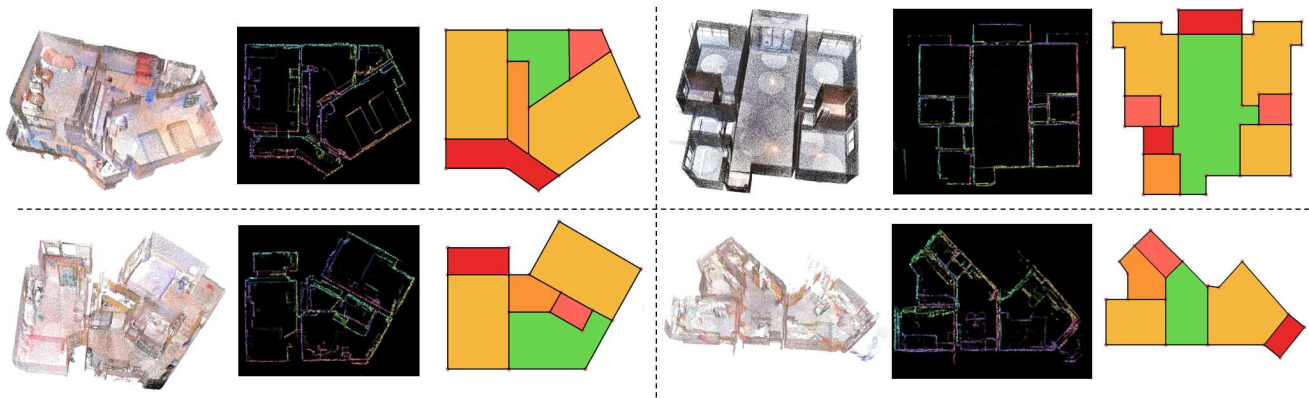


图 5. 我们的数据集为 527 间房屋/公寓提供生产级全景 RGBD 扫描。我们从自上而下的视图将每次扫描转换为点密度/法线图，这是我们系统的输入。我们将平面图结构注释为二维多边形图。请注意，为了可视化点密度/法线贴图（中间列），强度编码点密度，色相/饱和度编码平均表面法线的 2D 水平分量。

表 1 主要定量评价结果 颜色青色,橙,品红代表前三名。

方法	角落		边缘		房间		房间++
	前。召回	Prec.	召回	Prec.	召回	Prec.	记起
地板网 [20]	95.0	76.6	94.8	76.8	81.2	72.1	42.3 37.5
我们的（不含乙数据, E一致）	84.4	80.4	82.3	79.8	75.1	61.3	23.3 22.0
我们的（不含乙一致）	93.9	82.3	89.2	81.2	83.8	81.7	49.4 48.5
我们的（第一轮坐标下降）	94.6	82.8	89.4	81.7	83.9	81.8	49.5 48.7
我们的（第二轮坐标下降）	95.1	82.2	90.2	81.1	84.7	83.0	51.4 50.4

检测步骤；2）来自无关角落检测的无关房间；3）破碎的非曼哈顿结构，由于整数规划中的搜索空间过多，这使得具有挑战性。

右三列显示了提议的 Floor-SP 的变体。左边没有一致性项，并用先前工作中的 ad-hoc 成本函数替换了基于 DNN 的数据项 [2]。我们的整体配方保证了在每个检测到的房间部分进行房间重建，产生合理的结果。关于添加我们基于 DNN 的数据项乙数据(中)，每个房间的结构显著改善。但是，房间边界的不一致通常很明显。最后，加上一致性项（右），我们看到干净的平面图结构和一致的共享房间边界。

图 7 说明了多轮逐室坐标下降的效果。红色椭圆表示具有挑战性的结构，导致房间重叠或孔洞，在第二轮优化后得到解决。

## 7.2. 定量评价

我们遵循 FloorNet [20] 并为定量评估定义了以下四个指标：

角球精度/召回率：我们声明，如果在 10 个像素内存在真实房间角点，则角点重建成功。当在单个真实角点周围检测到多个角点时，我们只将最接近的一个角点视为正确，而将其他角点视为误报。

边缘精度/召回：我们声明如果图的两个端点通过了上述角点测试并且相应的边属于地面实况，则该图的边被成功重建。

房间精度/召回率：我们声明一个房间成功重建，如果 1) 它不与任何其他房间重叠，并且 2) 在 ground-truth 中存在一个房间，其交叉联合 (IOU) 分数超过 0.7。请注意，此度量不考虑角和边缘的定位和共享。

Room++ 精度/召回率：除了通过上述两个房间条件之外，如果房间连接（即共享边）到正确的一组成功重建的房间，我们声明在这个度量中成功地重建了房间。

表 1 显示了主要的定量评价。低级基元（即角和边）的精度指标

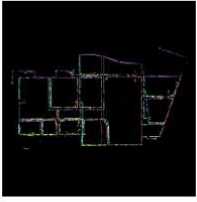
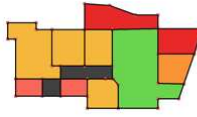
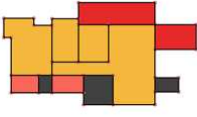
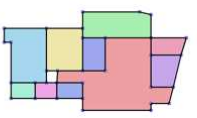
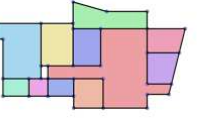
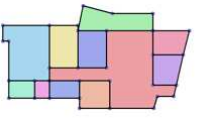

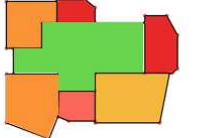
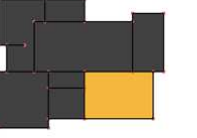
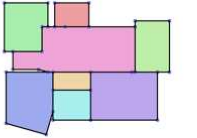
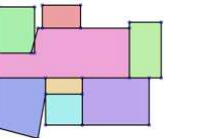
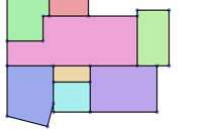


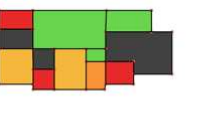
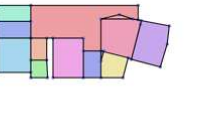
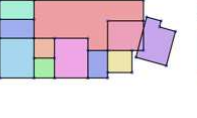
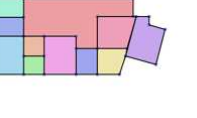

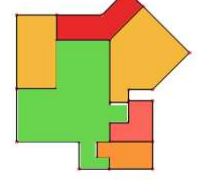
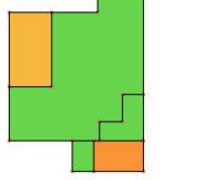
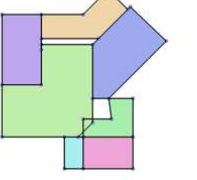
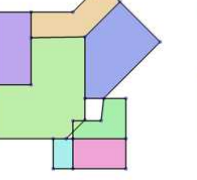
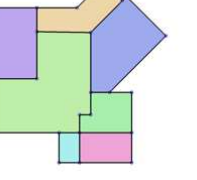

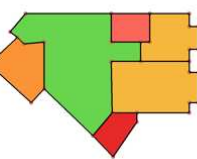
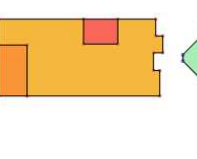
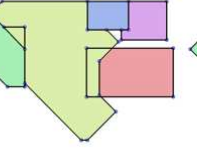
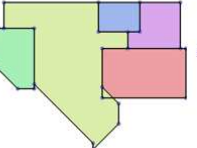
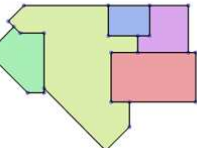

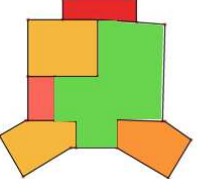
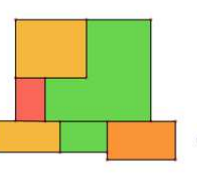
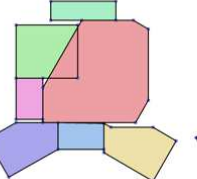
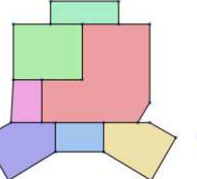
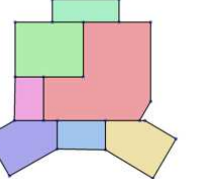
Processed Input	Ground Truth	FloorNet	Ours (w/o $E_{data}, E_{consis}$ )	Ours (w/o $E_{consis}$ )	Ours
					
					
					
					
					
					

图 6.与 FloorNet [20] 和我们方法的变体的定性比较。我们在这里选择硬非曼哈顿示例来说明我们数据集中的重建挑战。对于 Floor-SP 变体的重建，房间颜色由来自 Mask R-CNN 的相应房间段确定。对于 ground-truth 和 FloorNet，颜色取决于房间类型。

对 FloorNet 来说很高，因为这项任务不需要高级结构推理，而且大多数角点都很容易（例如，曼哈顿角点）。另一方面，即使对于低级基元，它们的召回指标也很低，因为某些房间角落由于遮挡而没有足够的 3D 点，其中基于 DNN 的角落检测

失败。Floor-SP 通过连续的房间优化过程恢复了这些具有挑战性的角落。

在房间级指标上，Floor-SP 始终优于 FloorNet。此外，数据和一致性术语的添加改进了房间级别的指标。最后，房间坐标下降进一步提升了



表现。所有 94 个测试示例的定量结果和可视化都在补充文档中

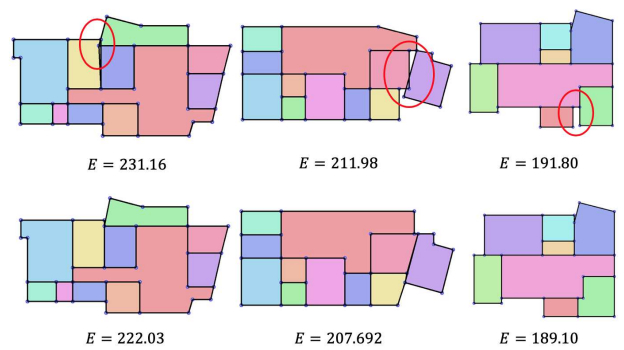


图 7. 多轮坐标下降修复了具有挑战性的平面图结构中的错误。上排显示第一轮坐标下降优化后的结果，下排显示第二轮后的结果。我们还显示了每一轮之后的总能量。相应的groundtruth注释如图6所示。

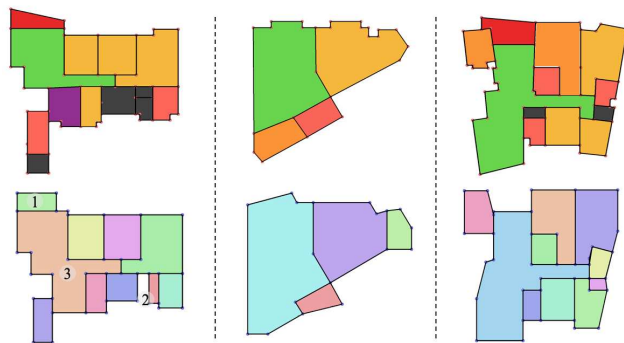


图 8. 典型故障模式。顶部是真实标注，底部是我们对每个示例的结果。对于复杂的场景和具有挑战性的非曼哈顿结构，我们的系统仍然会出错。

### 7.3. 讨论

Floor-SP 为曼哈顿结构产生了近乎完美的结果。大多数失败都集中在非曼哈顿案例上。从数量上看，我们的 Room++ 指标仅略高于 50。但是，我们想指出的是，即使在 Room++ 指标较差的极具挑战性的情况下，我们的重建也不是非常糟糕。

请看图 8 中的第一个示例。我们重建的 Room++ 精度和召回率均为 0，而重建看起来相当合理。如数字所示，原因有三重。1) 非曼哈顿小房间在前处理步骤中有错误的主导方向，导致 Floor-SP 无法恢复，IOU 测试失败；2) 凹形结构等小细节难以保留，房间未通过 IOU 测试；3) Mask R-CNN 的房间分割出错

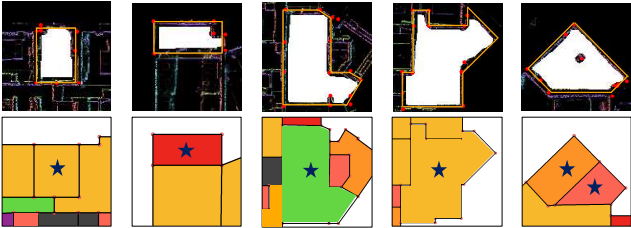


图 9. 标准角点检测容易出错（红色磁盘）。Mask R-CNN 会产生不精确的栅格房间片段（白色遮罩）或遗漏整个房间（最右边的示例）。Floor-SP 使用优化来稳健地解决拐角位置及其连接问题。在顶部，橙多边形显示了我们焦点房间的重建。底部显示了相应的基本事实。

关于复杂案件的房间数量，这又是不可能恢复的。一旦单个房间发生故障，所有相邻房间在 Room++ 指标中自动失败，导致此示例中的精度和召回率为零。

在图 9 中，我们进一步分析了我们的方法的稳健性。使用非最大抑制的角点检测总是会产生嘈杂的结果，并且实例分割网络生成的房间片段在细节上也不精确。Floor-SP 不是直接使用这些原始检测，而是制定了一个能量最小化问题来解决平面图角落的数量和位置，并且对这两种类型的错误具有鲁棒性。但是，当房间实例分割在房间数量上出错时（如图 9 中的最后一个示例所示），我们的系统无法恢复，而是产生了房间分隔错误的近似室内结构。在图 8 中的两个示例中也观察到了这个错误。未来的一项研究是从房间分割阶段的错误中恢复，以生成更准确的平面图。

我们还想指出，我们系统的输入是来自自上而下视图的单元密度/法线图像。我们通过将点投影到 2D 图像上来丢弃 3D 信息，如第 1 节所述。6. 我们没有使用高分辨率全景 RGB 图像，这些图像在数据集中可用，可以使系统像 FloorNet [20] 一样更加健壮。

我们相信这篇论文为室内测绘研究树立了一个重要的里程碑。所提出的系统在大量生产质量具有挑战性的场景中产生了令人信服的平面图重建结果。我们在项目网站上公开分享我们的代码和数据，以促进进一步的研究。

致谢：这项研究得到了美国国家科学基金会 IIS 1618685、NSERC Discovery Grants、NSERC Discovery Grants Accelerator Supplements 和 DND/NSERC Discovery Grant Supplement 的部分支持。我们感谢北客 (<https://www.ke.com>) 的 3D 房屋扫描和注释。



## 参考

- [1] Iro Armeni, Ozan Sener, Amir R Zamir, Helen Jiang, Ioannis Brilakis, Martin Fischer 和 Silvio Savarese。大型室内空间的 3D 语义解析。在 *IEEE 计算机视觉和模式识别会议 (CVPR)*, 2016 年。
- [2] Ricardo Cabral 和 Yasutaka Furukawa。从图像中重建分段平面和紧凑平面图。在 *IEEE 计算机视觉和模式识别会议 (CVPR)*. IEEE, 2014。
- [3] 曹哲、托马斯·西蒙、魏世恩和亚瑟·谢赫。使用部分亲场的实时多人 2d 姿势估计。在 *IEEE 计算机视觉和模式识别会议论文集 (CVPR)*, 2017 年。
- [4] Angel X. Chang, Angela Dai, Thomas A. Funkhouser, Maciej Halber, Matthias Nießner, Manolis Savva, Shuran Song, Andy Zeng 和 Yinda Zhang。Matterport3d: 从室内环境中的 rgb-d 数据中学习。在 *2017 年 3D 视觉国际会议 (3DV)*, 2017 年。
- [5] 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 *图像分析与处理国际会议 (ICIAP)*, 2013 年。
- [6] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser 和 Matthias Nießner。Scannet: 对室内场景进行了丰富注释的 3D 重建。在 *IEEE 计算机视觉和模式识别会议 (CVPR)*, 2017 年。
- [7] pytorch-mask-rcnn. <https://github.com/multimodallearning/pytorch-mask-rcnn>。
- [8] 亚历克斯·弗林特、克里斯托弗·梅、大卫·默里和伊恩·里德。一种重建建筑内部的动态规划方法。在 *欧洲计算机视觉会议 (ECCV)*, 2010 年。
- [9] 亚历克斯·弗林特、大卫·默里和伊恩·里德。使用单目、立体和 3d 特征理解曼哈顿场景。在 *IEEE 国际计算机视觉会议 (ICCV)*, 2011 年。
- [10] 傅焕, 龚明明, 王朝辉, 凯汉·巴特曼盖利奇, 陶大成。用于单目深度估计的深度序数回归网络。在 *IEEE 计算机视觉和模式识别会议 (CVPR)*, 2018 年。
- [11] Yasutaka Furukawa, Brian Curless, Steven M. Seitz 和 Richard Szeliski。曼哈顿世界立体声。在 *IEEE 计算机学会计算机视觉和模式识别会议 (CVPR)*, 2009 年。
- [12] Kaiming He, Georgia Gkioxari, Piotr Dollár 和 Ross B. Girshick。掩码 r-cnn。在 *IEEE 国际计算机视觉会议 (ICCV)*, 2017 年。
- [13] Varsha Hedau, Derek Hoiem 和 David A. Forsyth。恢复杂乱房间的空间布局。在 *IEEE 国际计算机视觉会议 (ICCV)*, 2009 年。
- [14] 胡贺祥、周广通、邓志伟、廖子成、Greg Mori。学习具有标签关系的结构化推理神经网络。在 *IEEE 计算机视觉和模式识别会议 (CVPR)*, 2016 年。
- [15] Binh-Son Hua, Quang-Hieu Pham, Duc Thanh Nguyen, Minh-Khoi Tran, Lap-Fai Yu 和 Sai-Kit Yeung。场景: 带有注释的场景网格数据集。在 *2016 第四届 3D 视觉国际会议 (3DV)*, 2016 年。
- [16] 黄坤, 王一凡, 周子涵, 丁天娇, 高盛华, 马毅。学习解析人造环境图像中的线框。在 *IEEE 计算机视觉和模式识别会议 (CVPR)*, 2018 年。
- [17] 池畑聪、杨航和古川康孝。结构化的室内建模。在 *IEEE 国际计算机视觉会议 (ICCV)*, 2015 年。
- [18] Chen-Yu Lee, Vijay Badrinarayanan, Tomasz Malisiewicz 和 Andrew Rabinovich。Roomnet: 端到端的房间布局估计。在 *IEEE 国际计算机视觉会议 (ICCV)*, 2017 年。
- [19] 李明磊、彼得·旺卡、南良良。从点云重建曼哈顿世界城市。在 *欧洲计算机视觉会议 (ECCV)*, 2016 年。
- [20] 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 (ECCV)*, 2018 年。
- [21] 刘晨、吴家军、Pushmeet Kohli 和 Yasutaka Furukawa。光栅到矢量: 重新审视平面图转换。在 *IEEE 国际计算机视觉会议 (ICCV)*, 2017 年。
- [22] Aron Monszpart, Nicolas Mellado, Gabriel J. Brostow 和 Niloy Jyoti Mitra。Rapter: 用有规律的平面布局重建人造场景。ACM 翻译。图形, 34:103:1–103:12, 2015。
- [23] 梁良楠和彼得·旺卡。Polyfit: 从点云重建多边形表面。在 *IEEE 国际计算机视觉会议 (ICCV)*, 2017 年。
- [24] Thomas Schöps, Johannes L. Schönberger, Silvano Galliani, Torsten Sattler, Konrad Schindler, Marc Pollefeys 和 Andreas Geiger。具有高分辨率图像和多摄像头视频的多视图立体基准。在 *IEEE 计算机视觉和模式识别会议 (CVPR)*, 2017 年。
- [25] Alexander G Schwing, Tamir Hazan, Marc Pollefeys 和 Raquel Urtasun。用于 3d 室内场景理解的高效结构化预测。在 *IEEE 计算机视觉和模式识别会议 (CVPR)*. IEEE, 2012。
- [26] Shuran Song, Fisher Yu, Andy Zeng, Angel X. Chang, Manolis Savva 和 Thomas A. Funkhouser。从单个深度图像完成语义场景。在 *2017 IEEE 计算机视觉与模式识别会议 (CVPR)*, 2016 年。
- [27] 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):243–258, 2014。
- [28] 徐丹飞、朱玉可、Christopher B Choy、李飞飞。通过迭代消息传递生成场景图。在 *IEEE 计算机视觉和模式识别会议论文集 (CVPR)*, 2017 年。
- [29] Fisher Yu, Vladlen Koltun 和 Thomas A. Funkhouser。扩张的残差网络。在 *IEEE 计算机视觉和模式识别会议 (CVPR)*, 2017 年。

- [30] Shanxin Yuan, Qi Ye, Bjorn Stenger, Siddhant Jain, and Tae-Kyun Kim. 大手2。2m 基准：手姿势数据集和最先进的分析。在 *IEEE 计算机视觉和模式识别会议 (CVPR)*, 2017 年。