



This ECCV 2018 paper, provided here by the Computer Vision Foundation, is the author-created version.

The content of this paper is identical to the content of the officially published ECCV 2018
LNCS version of the paper as available on SpringerLink: <https://link.springer.com/conference/eccv>

FloorNet: Floorplan 的统一框架 从 3D 扫描重建

陈柳¹, 吴家业¹和古川康 孝²

¹美国圣路易斯华盛顿大学
¹{chenliu,jiaye.wu}@wustl.edu
²加拿大西蒙弗雷泽大学
furukawa@sfu.ca

抽象的。本文提出了一种新颖的深度神经架构，该架构通过使用智能手机穿过房屋自动重建平面图，这是室内地图研究的最终目标。挑战在于处理跨大型 3D 空间的 RGBD 流。所提出的神经架构，称为 FloorNet，通过三个神经网络分支有效地处理数据：1) 具有 3D 点的 PointNet，利用 3D 信息；2) 具有俯视图的二维点密度图像的 CNN，增强了局部空间推理；3) 带有 RGB 图像的 CNN，利用完整的图像信息。FloorNet 在分支之间交换中间特征以利用所有架构。我们通过使用 Google Tango 手机获取 155 个住宅或公寓的 RGBD 视频流并注释完整的平面图信息，为平面图重建创建了基准。我们的定性和定量评估表明，三个分支的融合有效地提高了重建质量。我们希望这篇论文和基准测试将成为解决具有挑战性的矢量图形平面图重建问题的重要一步。

关键词：平面图重建；3D 计算机视觉；3D 卷积神经网络

1 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。他们的绘图有效地传达了场景的几何和语义信息。例如，我们可以快速识别房间范围、门的位置或对象排列。我们还可以通过文本或图标样式轻松识别房间、门或对象的类型。不幸的是，北美 90% 以上的房屋没有平面图。室内地图研究的最终目标是通过使用智能手机穿过房屋来自动重建平面图。

¹ 前两位作者对这项工作做出了同等贡献。

消费级深度传感器通过成功的产品彻底改变了室内 3D 扫描。Matterport [1] 通过使用专用硬件获取一组全景 RGBD 图像来生成室内空间的详细纹理映射模型。Google Project Tango 手机 [14] 将 RGBD 图像流转换为 3D 或 2D 模型。这些系统产生详细的几何图形，但不能作为平面图或建筑蓝图，其几何图形必须简洁并尊重底层场景分割和语义。

由于其较大的 3D 范围，重建整个房屋或具有多个房间的公寓的平面图对现有技术提出了根本性挑战。标准方法将 3D 信息投影到 2D 横向域 [10]，从而丢失高度信息。PointNet [26, 28] 直接使用 3D 信息，但缺乏局部邻域结构。多视图表示 [27, 34] 避免了显式的 3D 空间建模，但主要用于对象，而不是大型场景和复杂的相机运动。3D 卷积神经网络 (CNN) [29, 36] 也显示出有希望的结果，但迄今为止仅限于物体或小规模场景。

本文提出了一种新颖的深度神经网络 (DNN) 架构 Floor-Net，它将覆盖大型 3D 空间的 RGBD 视频转换为对平面图几何和语义的逐像素预测，然后使用现有的整数规划公式 [17] 来恢复向量-图形平面图。FloorNet 由三个 DNN 分支组成。第一个分支使用具有 3D 点的 PointNet，利用 3D 信息。第二个分支在自上而下的平面图中使用具有 2D 点密度图像的 CNN，增强了局部空间推理。第三个分支使用带有 RGB 图像的 CNN，利用完整的图像信息。PointNet 分支和点密度分支在自顶向下视图中的 3D 点及其对应单元之间交换特征。图像分支将深度图像特征贡献到自上而下视图中的相应单元格中。这种混合 DNN 设计利用了所有架构中的精华，并有效地处理了覆盖具有复杂摄像机运动的大型 3D 场景的完整 RGBD 视频。

我们通过使用 Google Tango 手机获取 155 套住宅或公寓的 RGBD 视频流，为平面图重建创建了基准，并注释了它们的完整平面图信息，包括建筑结构、图标和房间类型。广泛的定性和定量评估证明了我们的方法优于竞争方法的有效性。

综上所述，本文的主要贡献有两个：1) 用于 RGBD 视频的新型混合 DNN 架构，直接处理 3D 坐标，在 2D 域中对局部空间结构进行建模，并结合完整的图像信息；2) 使用 RGBD 视频的新平面图重建基准，其中存在许多室内场景数据库 [4, 33, 1]，但没有一个解决矢量图形重建问题，这对数字地图、房地产或土木工程应用有直接影响。

2 相关工作

我们讨论了三个领域的相关工作：室内场景重建、3D 深度学习和室内扫描数据集。

室内场景重建：消费级深度传感器的进步为室内 3D 扫描带来了革命性的变化。KinectFusion [23] 可以对物体和小规模场景进行高保真 3D 扫描。惠兰等人。[40] 将工作扩展到建筑规模的扫描。虽然细节准确，但这些密集的重建不如 CAD 模型，它必须具有 1) 用于高效数据传输的简洁几何和 2) 用于架构分析或有效可视化的适当分割/语义。

对于 CAD 质量的重建，研究人员通过使用几何图元表示场景来应用基于模型的方法。利用室内建筑结构的 2.5D 属性，可以通过将线拟合到自上而下视图中的点来分隔房间 [25,37,35]。原始类型已扩展到平面 [5、6、31、43、22] 或长方体 [42]。虽然它们为选定的扫描产生了有希望的结果，但它们基于启发式的原始检测面临着嘈杂和不完整的 3D 数据的挑战。我们的方法通过 DNN 对整个输入进行全局分析，以更加稳健地检测原始结构。

另一条研究方向研究使用来自单个图像 [47] 或一组全景 RGBD 图像 [10, 21] 的形状语法的自上而下的场景重建。图像和 WiFi 指纹等人群感知数据也被用于构建场景图 [8, 7, 19, 12]。虽然语义分割 [4, 26, 28] 和场景理解 [45] 在室内场景中很流行，但还没有基于学习的强大的矢量图形平面图重建方法。本文提供了这样一种方法及其与 ground-truth 的基准。

恢复上述矢量图形平面图模型的一种方法是从光栅化平面图图像 [17]。我们共享相同的重建目标，并在最后一步使用他们的整数规划公式来恢复最终的平面图。然而，我们的输入不是单个图像作为输入，而是覆盖大 3D 空间的 RGBD 视频，这需要一种根本不同的方法来有效地处理输入数据。

3D 深度学习：CNN 在 2D 图像上的成功激发了通过 DNN 进行 3D 特征学习的研究。Volumetric CNNs [41, 20, 27] 是 CNN 到 3D 域的直接扩展，但存在两个主要挑战：1) 数据稀疏性和 2) 3D 卷积的计算成本。FPNN [15] 和 Vote3D [38] 试图解决第一个挑战，而 OctNet [29] 和 O-CNN [39] 通过八叉树表示来解决计算成本。

具有多视图渲染的 2D CNN 已成功用于对象识别 [27、34] 和部分分割 [16]。它们有效地利用了所有图像信息，但迄今为止仅限于常规（或固定）相机布置。扩展到具有复杂相机运动的最大场景并非易事。

PointNet [26] 直接使用 3D 点坐标来利用稀疏性并避免量化误差，但它没有提供明确的局部空间推理。PointNet++ [28] 对点进行分层分组并添加空间结构

最远点采样。Kd-Networks [13] 类似地通过 KD-tree 对点进行分组。由于分组, 这些技术会产生额外的计算费用, 并且在对象规模上受到限制。对于场景, 他们需要将空间分割成更小的区域 (例如, $1m \times 1m$ 块) 并独立处理每个区域 [26, 28], 损害全局推理 (例如, 识别走廊的长墙)。

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 数据集仅包含 16 个室内扫描 [30], 其目的是用于多视图立体而不是 3D 点云处理。ScanNet 数据集 [4] 和 SceneNN 数据集 [9] 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 [3] 为 90 座豪宅构建高质量的全景 RGBD 图像集。2D-3D-S 数据集 [2] 使用相同的 Matterport 相机提供 6 次大规模室内办公空间扫描。然而, 它们专注于 2D 和 3D 语义注释, 并没有解决矢量图形重建问题。同时, 它们需要昂贵的专用硬件 (即, Matterport 相机) 来进行高保真 3D 扫描, 而我们的目标是应对数据质量低的消费级智能手机的挑战。

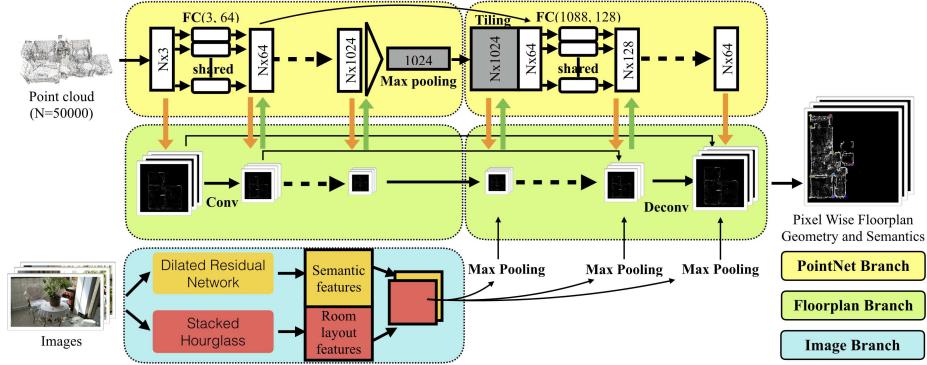
最后, 一个大型合成数据集 SUNCG [32] 提供了各种具有 CAD 质量几何和注释的室内场景。但是, 它们是合成的, 无法模拟真实场景的复杂性或替代真实照片。我们为 155 个住宅单元的智能手机提供完整的平面图注释和相应的 RGBD 视频的基准。

3 地板网

所提出的 FloorNet 将带有相机姿势的 RGBD 视频转换为像素级平面图几何和语义信息, 这是 Liu 等人引入的中间平面图表示。[17]。我们首先解释自包含的中间表示, 然后提供细节。

3.1 预赛

中间表示由几何和语义信息组成。几何部分包含房间角、对象图标角和门/窗端点, 其中每个角/点类型的位置由 $256 \times 2D$ 平面图图像域中的 256 个热图, 然后是标准的非最大抑制。例如, 房间角落是 I 形、L 形、T 形或 X 形的, 具体取决于入射墙壁的数量, 考虑到它们的旋转变体, 特征图的总数为 13。语义部分被建模为 1) 12 个特征图作为概率分布函数 (PDF) 超过 12 种房间类型, 以及 2) 8 个特征图作为 PDF 超过 8 种图标类型。我们遵循他们的方法, 并在最后使用他们的整数规划公式从这个表示中重建平面图。



图。1。FloorNet 由三个 DNN 分支组成。第一个分支使用 PointNet [26] 直接消费 3D 信息。第二个分支采用完全卷积网络 [18] 的平面图域中的自上而下的点密度图像，并产生逐像素的几何和语义信息。第三个分支通过在语义分割任务 [44] 上训练的扩张残差网络以及在房间布局估计 [24] 上训练的堆叠沙漏 CNN 产生深度图像特征。PointNet 分支和平面图分支在每一层交换中间特征，而图像分支将深度图像特征贡献到平面图分支的解码部分。这种混合 DNN 架构有效地处理带有相机姿势的输入 RGBD 视频，覆盖较大的 3D 空间。

3.2 三分支混合设计

Floornet 由三个 DNN 分支组成。我们在每个分支中使用现有的 DNN 架构，无需修改。我们的贡献在于它的混合设计：如何组合它们并共享中间特征（见图 1）。

点网分支：第一个分支是具有原始架构的 PointNet [26]，除了每个 3D 点由没有归一化位置的 XYZ 和 RGB 值表示。我们为每个数据随机抽取 50,000 个点。我们手动校正旋转并将重力方向与 Z 轴对齐。我们添加平移以将质心移动到原点。

平面图分公司：第二个分支是一个全卷积网络 (FCN) [18]，在编码器和解码器之间有跳跃连接，它在自上而下的视图中获取具有 RGB 值的点密度图像。每个单元格中的 RGB 值计算为 3D 点的平均值。我们计算曼哈顿校正的 3D 点的 2D 轴对齐边界框以定义矩形平面图域，同时忽略 2.5% 的异常点并将矩形在四个方向的每个方向上扩展 5%。矩形放在 256 的中间 \times 生成几何和语义特征图的 256 个正方形图像。分支的输入是同一域中的点密度图像。

图像分支：第三个分支通过两个 CNN 架构计算深度图像特征：1) 在语义上训练的扩张残差网络 (DRN) [44]

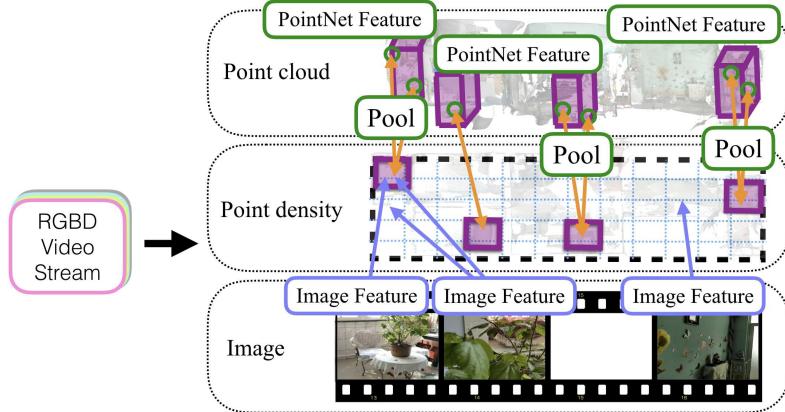


图 2。FloorNet 跨分支共享功能，以利用所有架构中的精华。3D 点的 PointNet 特征被汇集到平面图分支中相应的 2D 单元中。2D 单元的平面图特征被分解为 PointNet 分支中相应的 3D 点。基于深度图和相机位姿信息，深度图像特征被汇集到平面图分支中相应的 2D 单元中。

使用 ScanNet 数据集 [4] 进行 tic 分割；2) 堆叠沙漏 CNN (SH) [24] 使用 LSUN 数据集 [46] 对房间布局估计进行训练。

3.3 分支间特征共享

不同的分支学习不同领域（3D 点、平面图和图像）中的特征。图 2 显示了基于相机位姿和 3D 信息通过池化和反池化操作共享三个分支间特征。

PointNet 到平面图池：该池化模块从 PointNet 分支的每一层获取无序点的特征，并在平面图分支的相应层中生成 2D 自上而下的特征图。该模块简单地将 3D 点特征投影到平面图特征图中的单元格中，然后通过求和或最大操作聚合每个单元格中的特征。我们在前三个卷积层中使用求和操作来保留更多信息，同时在其余层中取最大值来引入竞争。构造的特征图与平面图分支中的特征图具有相同的维度。我们已经探索了点网分支的特征图和平面图分支的特征图之间的几种不同的聚合方案。我们发现 sum-pooling (即元素加法) 效果最好。

平面图到 PointNet 去池化：该模块反转了上述池化操作。它只是将平面图单元的特征复制并添加到单元内投影的每个相应的 3D 点中。时间复杂度再次与点数呈线性关系。

表格1。数据集统计。从左到右：房间数、图标数、开口数（即门或窗）、房间角数和总面积。报告每个条目的平均值和标准偏差。

	# room	#icon	#opening	#corner	area
平均	5.2	9.1	9.9		18.1 63.8[米 ²]
标准	1.8	4.5	2.9	4.2	13.0[米 ²]

图像到平面图池：图像分支为每个视频帧从 DRN 和 SH 生成尺寸为 512x32x32 和 256x64x64 的两个深度图像特征。我们首先通过深度图和相机姿势将图像特征解投影到 3D 点，然后将相同的 3D 应用于上面的平面图池化。一种修改是我们在所有层使用最大池化，以便将 3D 点投影到平面图域上，而不是混合使用 sum 和 max pooling。原因是我们在使用预训练模型进行图像特征编码，更复杂的混合池化效果会更小。我们对视频序列中的每 10 帧进行图像分支池化。

3.4 损失函数

我们的网络以相同的分辨率输出对平面图几何和语义信息的逐像素预测 256×256 。对于几何热图（即房间角、对象图标角和门/窗端点），使用 sigmoid 交叉熵损失。地面实况热图是通过将值 1.0 放入每个地面实况像素周围半径为 11 像素的圆盘内来准备的。对于语义分类特征图（即房间类型和对象图标类型），使用像素级 softmax 交叉熵损失。

4 平面图重构基准

本文为具有相机姿势的 RGBD 视频的矢量图形平面图重建问题创建了一个基准。我们使用 Google Tango 手机（Lenovo Phab 2 Pro 和 Asus ZenFone AR）对美国和中国的住宅单元进行了大约两百个 3D 扫描（见图 3）。在手动删除质量较差的扫描后，我们为剩余的 155 次扫描注释了完整的平面图信息：1) 房间角落作为点，2) 墙壁作为房间角落对，3) 对象图标和类型作为轴对齐的矩形和分类标签，4) 门窗（即开口）作为墙壁上的线段，以及 5) 房间类型作为由墙壁包围的多边形区域的分类标签。对象类型列表是{柜台、浴缸、马桶、水槽、沙发、橱柜、床、桌子、冰箱}。房间类型列表是{客厅、厨房、卧室、浴室、衣柜、阳台、走廊、餐厅}。表 1 提供了我们数据收集的统计数据。

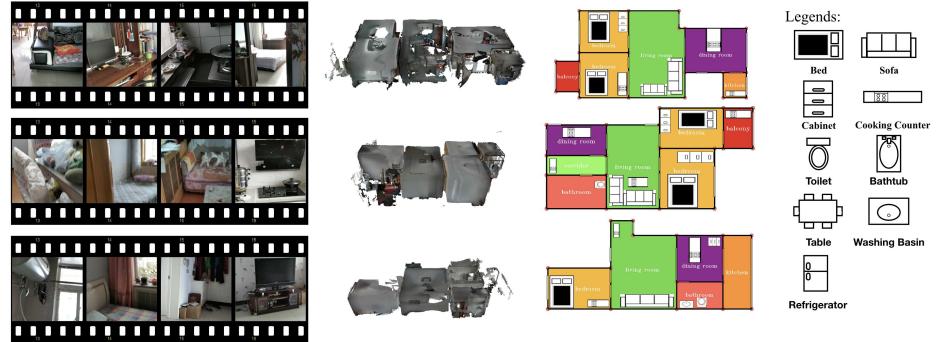


图 3。平面图重建基准。从左到右：二次采样视频帧、彩色 3D 点云和真实平面图数据。平面图数据存储在矢量图形表示中，它使用简单的渲染引擎进行可视化（例如，房间根据其类型分配不同的颜色，对象显示为规范图标）。

重建的平面图在三个不同级别的几何和语义一致性与地面实况进行评估。我们遵循 Liu 等人的工作。[17]并定义低级和中级指标如下。

- 低级指标是房间角落检测的精度和召回率。如果角点检测到地面实况的距离小于 10 像素并且是所有其他房间角点中最近的，则角点检测被宣布为成功。
- 中级指标是检测到的开口（即门窗）、对象图标和房间的精度和召回率。如果相应端点的最大距离小于 10 个像素，则表明开口检测成功。如果与 ground-truth 的交集超过联合 (IOU) 高于 0.5 (resp. 0.7)，则物体（或房间）的检测被宣布成功。
- 建筑构件的关系在评估室内空间中起着至关重要的作用。例如，人们可能会寻找卧室不与厨房相连的公寓。建筑规范可能会强制每间卧室在发生火灾时通过门窗快速疏散到室外。我们引入高级度量作为与相邻房间具有正确关系的房间的比率。更准确地说，我们声明一个房间有正确的关系，如果 1) 它通过门连接到正确的房间集合，如果两个房间的公共墙壁包含至少一个门，则两个房间是连接的，2) 房间有一个 IOU 分数大于 0.5 与相应的真实值，并且 3) 房间有正确的房间类型。

5 实施细节

5.1 DNN 训练

在我们收集的 155 次扫描中，我们随机抽取 135 次进行训练，留下 20 次进行测试。我们通过随机缩放和

每次我们输入训练样本时都会旋转。首先，我们使用从 $[0.5, 1.5]$ 范围内均匀采样的随机因子对点云和注释应用重新缩放。其次，我们随机应用围绕 z 轴的旋转 $0^\circ, 90^\circ, 180^\circ$, 或 270° .

我们使用了两个图像编码器 DRN [44] 和 SH [24] 的官方代码。我们使用 ScanNet 数据库 [4] 对 DRN 进行语义分割任务的预训练，并使用 LSUN [46] 在房间布局估计任务上对 SH 进行预训练。在 FloorNet 训练期间，DRN 和 SH 是固定的。我们在 TensorFlow 中使用现代 API 自行实现了剩余的 DNN 模块，即 PointNet [26] 用于 Pointnet 分支，FCN [18] 用于 Floorplan 分支。

使用 TitanX GPU 训练 FloorNet 大约需要 2 小时。我们将批量大小设置为 6。FloorNet 具有三种类型的损失函数。为了避免图标损失的过度拟合，我们根据测试损失分别训练了最多 600 个时期的图标损失，并提前停止。其他损失联合训练 600 个 epoch。³训练消耗 $81,000 = 135(\text{samples}) \times 600(\text{epochs})$ 个增强训练样本。最初令我们惊讶的是，FloorNet 甚至可以从少量 3D 扫描中进行泛化。然而，FloorNet 进行逐像素的低级预测。每个 3D 扫描包含大约 10 个对象图标、10 个开口和几十个房间角落，这可能导致良好的泛化性能以及数据增强，其中 Liu 等人观察到了类似的现象。[17]

5.2 增强启发式

我们使用以下两种增强启发式方法来增强整数规划公式 [17]，以处理更具挑战性的输入数据（即大规模原始传感器数据），从而在网络预测中产生更多噪声。

原始候选生成：标准的非最大抑制通常会检测到单个地面实况周围的多个房间角落。在将房间角落热图的阈值设置为 0.5 后，我们只需从每个面积超过 5 个像素的连接组件中提取最高峰。为了处理定位错误，我们连接两个房间角落，并在它们对应的连接组件沿 X 或 Y 方向重叠时生成候选墙。我们不会增加路口以保持候选人的数量易于处理。

目标函数：墙壁和开口候选者最初在目标函数[17]中分配了统一的权重。我们通过沿宽度为 7 像素（分别为 5 像素）的线取“墙”类型的语义热图得分的平均值来计算候选墙（分别为开口）的置信度。我们通过置信度分数减去 0.5 来设置每个基元的权重，以便仅当置信度至少为 0.5 时才鼓励选择基元。

³我们考虑了合成数据集 SUNCG [32] 和真实数据集 Matterport3D [3]

使用图标损失进行训练，同时使用它们的语义分割信息来生成图标注释。然而，联合训练仍然会出现过度拟合，而这种简单的早期停止启发式在我们的实验中效果很好。

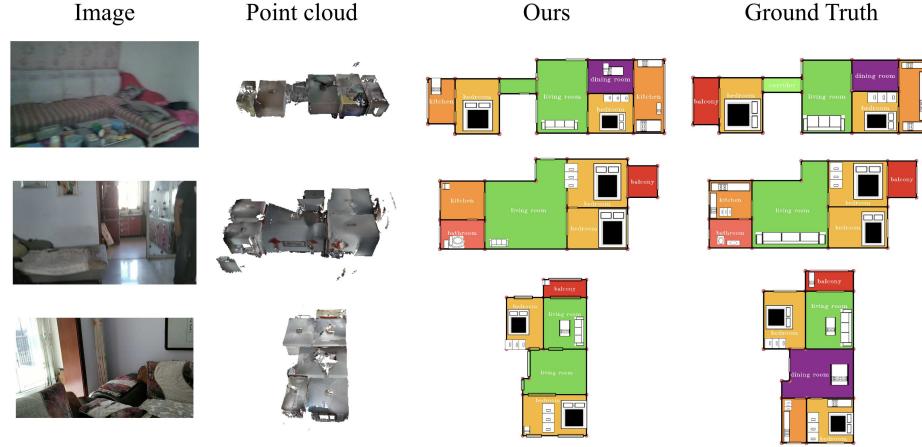


图 4。平面图重建结果。

表 2。针对竞争方法和我们的变体对低、中和高级指标进行定量评估。橙色和青色表示每个条目的最佳和次佳结果。

	墙	门	图标	房间	关系
点网 [26]	25.8/42.5	11.5/38.7	22.5/27.9	27.0/40.2	5.0
平面图-分支	90.2/88.7	70.5/78.0	43.4/42.8	76.3/75.3	50.0
图像分支	40.0/83.3	15.4/47.1	21.4/17.4	25.0/57.1	0.0
OctNet [29]	75.4/89.2	36.6/ 82.3	32.8/48.8	62.1/72.0	13.5
我们的没有 PointNet-Unpooling	92.6 /92.1	75.8/76.8	55.1 /51.9	80.9/77.4	52.3
我们的没有 PointNet-Pooling	88.4/ 93.0	73.0/ 87.2	50.0/42.2	75.0/80.6	52.8
我们的 w/o Image-Pooling	92.6 /89.7	77.1 /74.4	50.5/ 57.8	84.2 / 83.1	56.8
我们的	92.1/ 92.8	76.7 /80.2	56.1 / 57.8	83.6 / 85.2	56.8

6个实验

图 4 显示了我们对一些代表性示例的重建结果。我们的方法成功地恢复了复杂的矢量图形平面图数据，包括房间几何形状及其通过门的连接性。主要的故障模式之一是图标检测，因为对象检测通常需要比低级几何检测更多的训练数据[17]。我们相信更多的训练数据将克服这个问题。另一个典型的故障是由于杂乱或扫描不完整而导致房间角落丢失。房间的成功重建需要成功检测房间的每个角落。这是一个具有挑战性的问题，引入更高级别的约束可能会找到解决方案。

图 6 和表 2 定性和定量地比较了我们的方法与竞争技术，即 OctNet [29]、PointNet [26] 和我们的 FloorNet 的一些变体。OctNet 和 PointNet 代表最先进的 3D

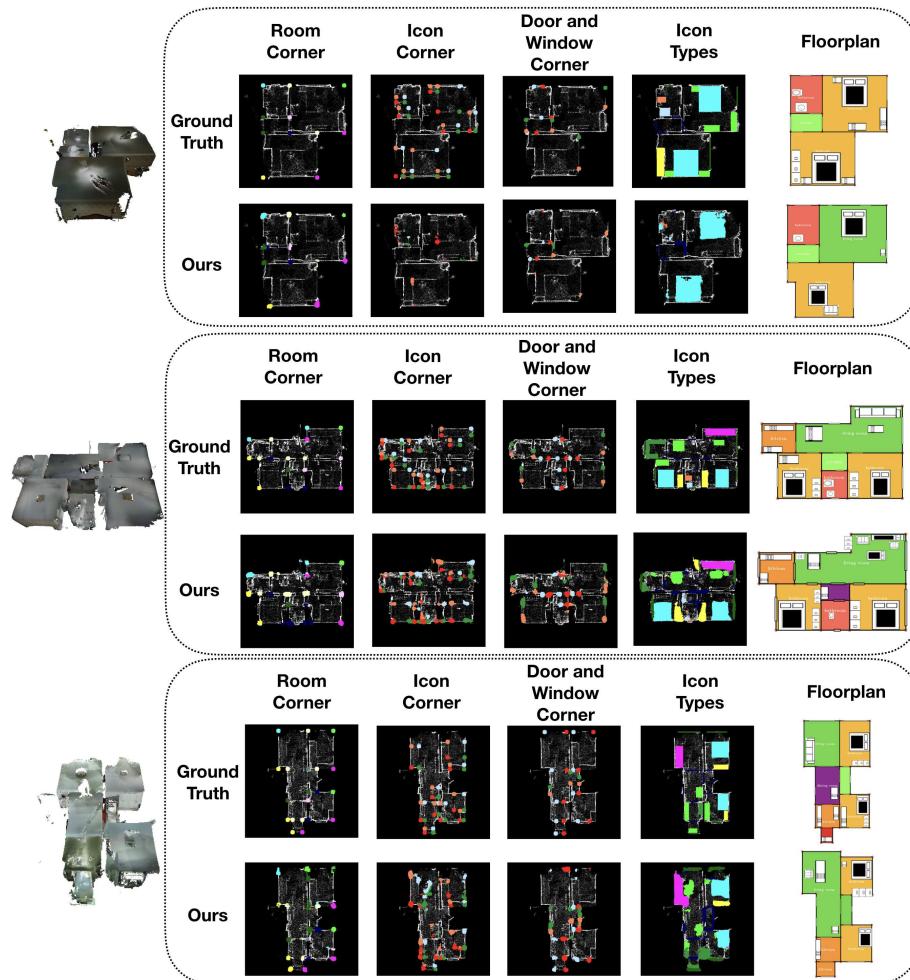


图 5。中间结果。对于每个示例，我们将网络的原始输出（房间角、图标角、开口角和图标类型）与真实情况进行比较。在第二个示例中，由于 3D 点质量差，我们在顶部生成了一个假房间（蓝色）。在第三个例子中，由于嘈杂的 3D 点，重建的房间在左下角附近再次出现不准确的形状，说明了我们问题的挑战。



图 6。与竞争方法的定性比较。最上面的是 OctNet [29]，一种最先进的 3D CNN 架构。接下来的三行显示了我们 FloorNet 的变体，其中仅启用了一个分支。FloorNet 与所有分支机构整体生成更完整和准确的平面图。

表3。在 PointNet to floorplan 分支间池化中，我们混合使用 sum 和 max pooling 将 3D 点投影到 2D 平面图域。为了验证这种混合方案，我们还评估了在所有层仅使用最大池或和池时的性能。

汇集法	墙	门	图标	房间	关系
最大限度	88.0/89.9	70.9/86.6	59.1/47.8	76.3/77.3	47.0
和	88.3/97.0	69.6/85.9	55.6/53.4	76.3/82.9	52.3
总和/最大值（默认）	92.1/92.8	76.7/80.2	56.1/57.8	83.6/85.2	56.8

DNN。更准确地说，我们基于官方 OctNet 库实现了体素语义分割网络，⁴它将 256x256x256 体素作为输入并输出相同分辨率的 3D 体素。然后我们添加三个单独的 $5 \times 3 \times 3$ 个卷积层，步幅为 $4 \times 1 \times 1$ 用相同的损失函数集预测相同的逐像素几何和语义特征图。PointNet 只是我们的 FloorNet，没有点密度或图像输入。同样，我们通过仅启用 3D 点（对于 PointNet 分支）或点密度图像（对于平面图分支）作为输入来构建 FloorNet 变体。

该表显示，平面图分支是最能提供信息的，因为它是平面图重建任务最自然的表示，而单独的 PointNet 分支或图像分支不能很好地工作。我们还将整个点云拆分为 1 米 \times 1 米块，训练仅 PointNet 模型，该模型分别对每个块进行预测，然后进行简单的合并。然而，这表现得更糟。OctNet 在中低级指标上表现相当不错，但在高级指标上表现不佳，所有房间和相关门必须以高精度重建以报告良好的数字。

为了进一步评估所提出的 FloorNet 架构的有效性，我们通过禁用每个分支间池化/非池化操作来进行消融研究。表 2 的底部显示，特征共享总体上会带来更好的结果，尤其是对于中高级指标。

表 3 比较了 PointNet 与平面图池的不同分支间池/解池方案。该表显示，早期层中的最大操作丢失了太多信息并导致更差的性能。

最后，图 7 与内置的 Tango Navigator 应用程序 [11] 进行了比较，后者在手机上实时生成平面图图像。请注意，他们的系统不会 1) 生成房间分割，2) 识别房间类型，3) 检测对象，4) 识别对象类型，或 5) 生成 CAD 质量的几何图形。因此，我们通过测量真实墙和预测墙之间的线距离来定量评估几何信息。更准确地说，我们 1) 从每个墙线段中采样 100 个点，2) 对于每个采样点，在另一个线段中找到最接近的点，以及 3) 计算所有采样点和线段的平均距离。Tango Navigator App 和我们的 FloorNet 的平均线距离分别为 2.72 [像素] 和 1.66 [像素]。这是一个令人惊讶的结果，

⁴OctNet 库：<https://github.com/griegler/octnet>



图 7。与商业平面图生成器 Tango Navigator App 的比较。上图：来自 Tango 的平面图。底部：我们的结果。

在整数规划期间，当相应的房间没有被重建时，确定的线段。另一方面，这是一个预期的结果，因为我们的方法利用了所有的几何和图像信息。

7 结论

本文提出了一种新颖的 DNN 架构 FloorNet，它从带有相机姿势的 RGBD 视频中重建矢量图平面图。FloorNet 采用混合方法并利用三种 DNN 架构中最好的一种来有效地处理覆盖具有复杂摄像机运动的大型 3D 空间的 RGBD 视频。该论文还为新的矢量图形重建问题提供了一个新的基准，该问题在最近的计算机视觉室内场景数据库中是缺失的。两个主要的未来工作摆在我们面前。第一个是学习在 DNN 内部强制执行更高级别的约束，而不是在单独的后处理（例如，整数规划）内部。学习高级约束可能需要更多的训练数据，第二个未来的工作是获得更多的扫描。

北美 90% 以上的房屋没有平面图。我们希望这篇论文和基准测试将成为解决这一具有挑战性的矢量图形重建问题的重要一步，并能够通过智能手机穿过房屋来重建平面图。我们公开分享我们的代码和数据以促进进一步的研究。

8 确认

这项研究得到了美国国家科学基金会 IIS 1540012 和 IIS 1618685、谷歌学院研究奖、Adobe 礼品基金和 Zillow 礼品基金的部分支持。我们感谢 Nvidia 慷慨的 GPU 捐赠。

参考

1. 物质港。<https://matterport.com/>
2. Armeni, I., Sener, O., Zamir, AR, Jiang, H., Brilakis, I., Fischer, M., Savarese, S.: 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论文集。第 1534-1543 页 (2016 年)
3. Chang, A., Dai, A., Funkhouser, T., Halber, M., Nießner, M., Savva, M., Song, S., Zeng, A., Zhang, Y.: Matterport3d：学习来自室内环境中的 rgb-d 数据。arXiv 预印本 arXiv:1709.06158 (2017)
4. Dai, A., Chang, AX, Savva, M., Halber, M., Funkhouser, T., Nießner, M.: Scannet：带丰富注释的室内场景 3D 重建。在：过程。IEEE 会议。关于计算机视觉和模式识别 (CVPR)。卷。1 (2017)
5. Furukawa, Y., Curless, B., Seitz, SM, Szeliski, R.: 曼哈顿世界立体声。在：计算机视觉和模式识别，2009 年。CVPR 2009 年。IEEE 会议。第 1422-1429 页。IEEE (2009)
6. Furukawa, Y., Curless, B., Seitz, SM, Szeliski, R.: 从图像重建建筑内部。在：计算机视觉，2009 年 IEEE 第 12 届国际会议上。第 80-87 页。IEEE (2009)
7. Gao, R., Zhao, M., Ye, T., Ye, F., Luo, G., Wang, Y., Bian, K., Wang, T., Li, X.: 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), 1427-1442 (2016)
8. Gao, R., Zhao, M., Ye, T., Ye, F., Wang, Y., Bian, K., Wang, T., Li, X.: 拼图：通过移动设备重建室内平面图人群感应。在：第 20 届移动计算和网络年度国际会议论文集。第 249-260 页。ACM (2014)
9. Hua, BS, Pham, QH, Nguyen, DT, Tran, MK, Yu, LF, Yeung, SK: Scenenn：带有注释的场景网格数据集。见：3D Vision (3DV)，2016 年第四届国际会议。第 92-101 页。IEEE (2016)
10. Ikehata, S., Yang, H., Furukawa, Y.: 结构化室内建模。在：IEEE 计算机视觉国际会议论文集。第 1323–1331 页 (2015 年)
11. Inc., G.: 探戈计划。<https://developers.google.com/tango/>
12. Jiang, Y., Xiang, Y., Pan, X., Li, K., Lv, Q., Dick, RP, Shang, L., Hannigan, M.: 使用房间指纹的基于走廊的自动室内平面图构建。见：2013 年 ACM 普适计算国际联合会议论文集。第 315-324 页。ACM (2013)
13. Klokov, R., Lempitsky, V.: 逃离细胞：用于识别 3d 点云模型的深度 kd 网络。在：2017 年 IEEE 计算机视觉国际会议 (ICCV)。第 863-872 页。IEEE (2017)
14. Lee, J., Dugan, R. 等人：谷歌探戈项目
15. Li, Y., Pirk, S., Su, H., Qi, CR, Guibas, LJ: Fpnn：3d 数据的现场探测神经网络。在：神经信息处理系统的进展。第 307–315 页 (2016 年)
16. Limberger, FA, Wilson, RC, Aono, M., Audebert, N., Boulch, A., Bustos, B., Giachetti, A., Godil, A., Le Sa ux, B., Li, B., et al.: Shrec'17 track: 非刚性玩具的点云形状检索。在：关于 3D 对象检索的第 10 届 Eurographics 研讨会。第 1-11 页 (2017 年)
17. Liu, C., Wu, J., Kohli, P., Furukawa, Y.: 光栅到矢量：重新审视平面图转换。在：IEEE 计算机视觉和模式识别会议论文集。第 2195-2203 页 (2017 年)

18. Long, J., Shelhamer, E., Darrell, T.: 用于语义分割的全卷积网络。在：IEEE 计算机视觉和模式识别会议论文集。第 3431–3440 页 (2015 年)
19. Luo, H., Zhao, F., Jiang, M., Ma, H., Zhang, Y.: 使用基于磁指纹的众包构建室内平面图。传感器17(11), 2678 (2017)
20. Maturana, D., Scherer, S.: Voxnet：用于实时对象识别的 3d 卷积神经网络。在：智能机器人和系统 (IROS)，2015 年 IEEE/RSJ 国际会议。第 922–928 页。IEEE (2015)
21. Mura, C., Mattausch, O., Pajarola, R.: 具有任意墙壁布置的多房间内部的分段平面重建。在：计算机图形学论坛。卷。35，第 179–188 页。威利在线图书馆 (2016)
22. Mura, C., Mattausch, O., Villanueva, AJ., Gobbetti, E., Pajarola, R.: 房间布局复杂的杂乱室内环境中的自动房间检测和重建。计算机和图形44,20–32 (2014)
23. Newcombe, RA, Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, AJ, Kohi, P., Shotton, J., Hodges, S., Fitzgibbon, A.: Kinectfusion：实时密集表面映射和跟踪。见：混合和增强现实 (ISMAR)，2011 年第 10 届 IEEE 国际研讨会。第 127–136 页。IEEE (2011)
24. Newell, A., Yang, K., Deng, J.: 用于人体姿势估计的堆叠沙漏网络。在：欧洲计算机视觉会议。第 483–499 页。施普林格 (2016)
25. Okorn, B., Xiong, X., Akinci, B., Huber, D.: 走向平面图的自动化建模。在：关于 3D 数据处理、可视化和传输的研讨会论文集。卷。2 (2010)
26. Qi, CR, Su, H., Mo, K., Guibas, LJ: Pointnet：用于 3d 分类和分割的点集的深度学习。arXiv 预印本 arXiv:1612.00593 (2016)
27. Qi, CR, Su, H., Nießner, M., Dai, A., Yan, M., Guibas, LJ: 用于 3d 数据对象分类的体积和多视图 cnns。在：IEEE 计算机视觉和模式识别会议论文集。第 5648–5656 页 (2016 年)
28. Qi, CR, Yi, L., Su, H., Guibas, LJ: Pointnet++：度量空间中点集的深度分层特征学习。在：神经信息处理系统的进展。第 5105–5114 页 (2017 年)
29. Riegler, G., Ulussoys, AO, Geiger, A.: Octnet：以高分辨率学习深度 3d 表示。arXiv 预印本 arXiv:1611.05009 (2016)
30. Schöps, T., Schönberger, JL, Galliani, S., Sattler, T., Schindler, K., Pollefeys, M., Geiger, A.: 具有高分辨率图像和多摄像头视频的多视图立体基准。在：过程。CVPR。卷。3 (2017)
31. Sinha, S., Steedly, D., Szeliski, R.: 基于图像渲染的分段平面立体 (2009)
32. Song, S., Yu, F., Zeng, A., Chang, AX, Savva, M., Funkhouser, T.: 来自单个深度图像的语义场景补全。arXiv 预印本 arXiv:1611.08974 (2016)
33. Song, S., Yu, F., Zeng, A., Chang, AX, Savva, M., Funkhouser, T.: 来自单个深度图像的语义场景补全。IEEE 计算机视觉和模式识别会议 (2017)
34. Su, H., Maji, S., Kalogerakis, E., Learned-Miller, E.: 用于 3d 形状识别的多视图卷积神经网络。在：IEEE 计算机视觉国际会议论文集。第 945–953 页 (2015 年)
35. Sui, W., Wang, L., Fan, B., Xiao, H., Wu, H., Pan, C.: 城市建筑自动重建的分层平面图提取。IEEE 可视化和计算机图形学交易22(3), 1261–1277 (2016)

36. Tatarchenko, M., Dosovitskiy, A., Brox, T.: 八叉树生成网络：用于高分辨率 3d 输出的高效卷积架构。arXiv 预印本 arXiv:1703.09438 (2017)
37. Turner, E., Cheng, P., Zakhori, A.: 室内环境纹理 3d 模型的快速、自动化、可扩展生成。IEEE 信号处理选题杂志9(3), 409–421 (2015)
38. Wang, DZ, Posner, I.: 在线点云目标检测中的投票投票。在：机器人：科学与系统 (2015 年)
39. Wang, PS, Liu, Y., Guo, YX, Sun, CY, Tong, X.: O-cnn：用于 3d 形状分析的基于八叉树的卷积神经网络。ACM 图形事务 (TOG)36(4), 72 (2017)
40. Whelan, T., Kaess, M., Fallon, M., Johannsson, H., Leonard, J., McDonald, J.: Kintinuous：空间扩展 kinectfusion (2012)
41. Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J.: 3d shapenets：体积形状的深度表示。在：IEEE 计算机视觉和模式识别会议论文集。第 1912-1920 页 (2015 年)
42. Xiao, J., Furukawa, Y.: 重建世界博物馆。国际计算机视觉杂志110(3), 243–258 (2014)
43. Xiong, X., Adan, A., Akinci, B., Huber, D.: 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31, 325–337 (2013)
44. Yu, F., Koltun, V., Funkhouser, T.: 扩张残差网络。在：计算机视觉和模式识别。卷。1 (2017)
45. Zhang, Y., Bai, M., Kohli, P., Izadi, S., Xiao, J.: Deepcontext：用于 3D 整体场景理解的 Contextencoding 神经通路。arXiv 预印本 arXiv:1603.04922 (2016)
46. Zhang, Y., Yu, F., Song, S., Xu, P., Seff, A., Xiao, J.: 大规模场景理解挑战：房间布局估计。九月访问15 (2015)
47. Zhao, Y., Zhu, SC: 使用随机场景语法进行图像解析。在：神经信息处理系统的进展。第 73–81 页 (2011 年)