

# GeoTexBuild：从地图轮廓生成3D建筑模型

王瑞哲

杨俊彦

王乔



2025年3月

东南大学

## 摘要

我们介绍了GeoTexBuild，这是一个用于从地图轮廓生成3D建筑模型的模块化生成框架。该框架采用三阶段过程，包括高度图生成、几何重建和外观风格化，最终生成具有复杂几何形状和外观属性的建筑模型。通过整合定制的ControlNet和Text2Mesh模型，我们探索了在生成过程中控制几何和视觉属性的有效方法。通过这一方法，我们消除了现有3D生成技术中单一立面照片背后的结构变化问题。各阶段的实验结果验证了GeoTexBuild从场地规划或地图设计中提取的轮廓生成详细且准确的建筑模型的能力。我们的框架显著减少了建筑建模中的人工劳动，并能为设计师提供灵感。

**关键词：**3D建筑生成、模块化框架、几何控制、外观风格化。

## 1 引言

建筑模型在城市规划 [68]、旅游、视频游戏设计、电影制作 [61] 和虚拟现实 [79] 应用中发挥着关键作用。然而，建筑结构的内在复杂性意味着，针对特定设计要求的自动化和可控建模仍然是一个需要进一步探索的领域 [81]。一方面，使用 CAD 软件 [5] 手动构建许多建筑模型既耗时又费力。另一方面，尽管 3D 重建技术 [26, 40, 20, 71, 39, 32, 50, 2, 48] 在近年来取得了显著进展，但这些方法在多个方面受到限制：它们无法生成现实中不存在的新建筑，通常需要大量且昂贵的数据收集 [73, 92, 85]，因此不太适合在早期规划和设计阶段应用。

由于建筑几何结构的固有复杂性 [44]，准确的 3D 建筑建模需要的不仅仅是单一的立面图像。这种复杂性不仅包括立面，还包括结构设计、高度和组件的比例，而这些在街景图像或相同立面设计中无法充分捕捉到 [57]。因此，这一限制要求探索能够更有效地概括和调节建筑整体几何特征的特征。

不幸的是，现有的 3D 生成技术 [60, 63, 41, 89, 4, 76, 9, 67, 72, 82, 29, 96, 53] 依赖于单一照片或文本提示，往往缺乏详细结构所需的几何精度。这些方法严重依赖于预训练模型，未能考虑几何中关键的设计特定参数，使其在细致的建筑应用中显得不足。

地图上识别出的自然特征可以作为有效的几何控制参数：

建筑轮廓（见图 1）。在城市规划和游戏设计中，专业系统和人员

<sup>1</sup>王瑞 (R. Wang) 曾在中国江苏省南京市东南大学信息科学与工程学院工作  
(电子邮件 : rz\_wang@seu.edu.cn)

<sup>2</sup>杨俊 (J. Yang) 曾在中国江苏省南京市东南大学建筑学院工作 (电子邮件 :  
yangjy\_seu@163.com)

<sup>3</sup>Q. Wang曾在中国江苏省南京市东南大学的信息科学与工程学院和经济与管理学院任职 (通讯作者, 邮箱:qiaowang@seu.edu.cn)

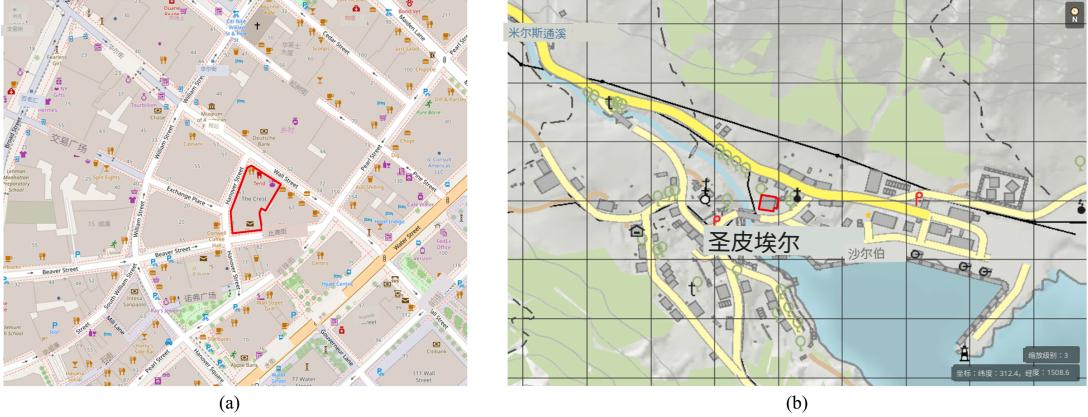


图1：在(a) OpenStreetMap上的真实世界地图和(b) 游戏设计中的虚构地图中的建筑轮廓。一个轮廓用红色突出显示，以展示其轮廓。

通常提前建立场地规划[43]或地图设计[49]，这些计划决定了关键参数，如建筑位置、楼面面积和轮廓形状。我们的目标是开发一种生成性方法，生成符合这些设计师指定控制条件的纹理建筑模型，有效地将预定义属性与简单的手绘草图和文本提示相结合，从而显著减少人工劳动并为设计师提供灵感。

表1：建筑建模方法的目标和功能比较。

类别	方法	特征	前向生成	轮廓控制屋顶形状控制	外观	监督	条件（一般）	表示
	航空摄影Yu 等人[13]		×	○	×	×	单一2D	-
	语法	Milde 等人[47] Storcz 等人[70]	×	○ ○	○ ○	×	人类知识 人类知识	网格 线框(图)
	Dick 等人[15] Schönberger 等人[65]	×	○	○ ○	○ ○	多个2D	-	网格
	Mildenhall 等人[48]	×	○	○ ○	○ ○	多个2D	-	Pcd NeRF
	Kerbl 等人[32] Li 等人[39] Debevec 等人[14]	×	○	○ ○	○ ○	多个2D	-	3DGs
重建	Nan 等人[52]	×	○	○ ○	○ ○	3D	-	网格
	Paden 等人[56]	×	○	○ ○	○ ○	3D	-	网格
	Chen 等人[10]	×	○	○ ○	○ ○	3D	-	网格
	Wu 等人[83]	×	○	○ ○	○ ○	3D	-	网格
	程序化	Nishida 等人[54] Parish 等人[58]	✓ ✓	✗ ✓	○ ○	单一2D 人类知识	图像 图表	网格 网格
	激光雷达	Elaikshler 等人[17] 黄 等人[27] Kada 等人[30]	×	○ ○ ○	○ ○ ○	3D 3D 3D	-	线框(图) 网格 网格
	卫星图像	Partovi 等人[59]	×	○ ○	○ ○	多光谱2D	-	网格
	单幅图像	Pang 等人[57]	✓	✗ ○	✗ ○	单一2D	图像	网格
	图像到3D	魏 等人[81] 龙 等人[41] 吴 等人[82]	✓ ✓ ✓	✗ ✗ ✗	✗ ✓ ✓	3D 3D 3D	图像 图像 图像	Pcd 网格 网格
	生成	普尔 等人[60] 拉杰 等人[63] 陈 等人[9] 易 等人[89]	✓ ✓ ✓ ✓	✗ ✗ ✗ ✗	✓ ✓ ✓ ✓	多个2D 多个2D 多个2D 多个2D	文本 文本 文本 文本	NeRF NeRF NeRF 3DGs
生成	文本到3D	Bensadoun 等人[4] Ukarapol 等人[76] Shi 等人[67] Nichol 等人[53]	✓ ✓ ✓ ✓	✗ ✗ ✗ ✗	✓ ✓ ✓ ✓	多个2D 多个2D 多个2D 多个2D	文本 文本 文本 文本	NeRF 3DGs, 网格 Mv 图像, NeRF Pcd
	文本/图像到3D	Tang 等人[72] Jun 等人[29] Zhao 等人[96]	✓ ✓ ✓	✗ ✗ ✗	✓ ✓ ✓	多个2D 多个2D 多个2D	文本/图像 文本/图像 文本/图像	3DGs 网格 网格
	文本, 图像到3D	我们的	✓	✓	✓	单一2D, 文本	图像, 文本	网格

✗：无法， ○：通过额外数据间接控制， ✓：能够。

为了实现这一目标，我们提出了GeoTexBuild（图2），这是一个三阶段生成框架策略：高度图生成、几何重建和外观风格化。在初始步骤中，我们结合ControlNet [94]，将建筑轮廓和手绘高度图的信息结合起来，从而实现对屋顶结构的受控生成。随后，建筑的整体几何形状从生成的屋顶结构和特定输入参数中重建，最终得到一个无纹理模型。最后，我们利用Text2Mesh [46]的风格化功能，为模型赋予详细的几何形状和颜色，形成一个完全纹理化的建筑模型。我们在表1中比较了所提方法的目标和特征与其他方法的异同。

在我们的实验中，我们训练了几个定制的ControlNet，以研究不同图像条件对屋顶结构生成控制的影响。此外，我们修改了Text2Mesh模块，以提高建筑生成的风格化质量。从各个角度进行的广泛实验证实了我们框架在建筑模型生成中的有效性。

GeoTexBuild有潜力弥合创意愿景与技术执行之间的差距，

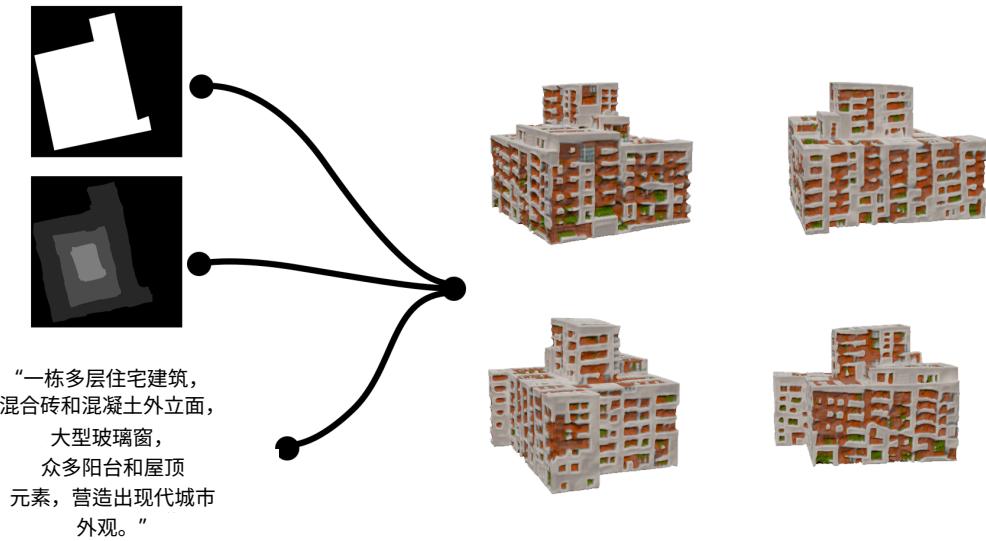


图2：GeoTexBuild从输入的建筑轮廓、高度草图和

文本提示生成3D纹理建筑模型。

使高度定制的建筑设计得以轻松实现。此外，这种方法还可以激发相邻领域的进步，例如室内和外部的生成设计结合，或用于设计和规划的自动化城市美学分析。我们模块化框架中的模块可以被更新的预训练模型替换，从而使我们能够利用快速发展的视觉和语言研究中的最新和未来的进展，因此它可以成为未来创新的基础。

总之，我们的贡献如下：

- 我们提出了一个模块化框架GeoTexBuild，用于从建筑轮廓生成建筑模型，具有详细可控的几何形状和外观属性，从而显著减少所需的人工劳动，并为设计师提供灵感。
- 我们研究如何通过将建筑生成过程分为三个阶段并利用正交视图中的2D空间条件来控制几何属性。
- 我们训练了一个定制的ControlNet和一个Text2Mesh模块，并将它们集成到我们的框架中用于建筑生成，使用了变体数据集Building3D和精心挑选的建筑图像。
- 每个阶段的实验证明了我们框架和微调模型的有效性。

## 2 相关工作

### 2.1 从重建建模建筑

除了手动建模，建筑自动建模的最早方法包括基于从现实中获得的数据或专业人士的专业知识重建结构的各种方法[42, 37, 1]。这些方法通常涉及从卫星图像[33, 59]、航拍照片[8, 13]、多张照片[15, 14, 16, 57, 65, 48, 32, 39]、点云[52, 56, 84, 10, 23, 83]或激光雷达数据[30, 17, 27]重建建筑。基于专业经验的自动建模通常采用程序建模技术[54, 58]和从收集的数据生成建筑语法的方法[47, 70]。

近年来，传统方法与深度学习或优化技术相结合 [8, 13, 57, 65, 52, 10, 83]，要么更新这些方法中的计算逻辑，要么改变数据量要求。此外，针对一般3D重建的新方法也相继出现，超越了仅限于建筑物的范围 [7, 18, 19]。这些创新方法 [26, 20, 39, 32, 50, 2, 48] 重新定义了3D数据表示，或引入了新的重建算法，提升了重建的效率和保真度。它们能够同时将几何形状和颜色恢复到照片级真实感。

这些重建方法提高了我们对建模和生成过程的理解，并为生成建立了方法论框架和数据基础。

## 2.2 3D生成

3D生成技术主要分为两大类：一类强调单个对象的生成，另一类则专注于场景（背景）生成 [36, 28, 55, 93, 90]。也存在用于建模3D城市环境的生成方法 [98, 35, 87, 66]，但它们的重点通常不包括单个建筑的详细几何形状。因此，本节主要集中在单对象生成的技术上。

起源于重建任务的3D生成技术最近发展迅速。与其说是生成特定对象，这些技术可以生成由其数据集决定的各种对象。一些3D生成模型只能生成形状 [86]，而其他模型则可以同时生成形状和外观 [38]。

根据输入，有两种主流的3D生成模型：文本到3D [34] 和图像到3D [38]。根据其3D表示，这些模型可以进一步分类为显式表示（如基于点云的 [53, 81, 97]、基于网格的 [69, 95]、基于体素的 [64] 和基于3DGS的 [89, 76]）、隐式表示（如基于NeRF的 [60, 63] 和基于神经隐式表面的 [29]），以及混合表示（如基于潜在的 [96, 45]、基于三平面的 [25, 74] 和基于多层表示的 [90]）。

关于生成策略，模型可以采用基于多视图优化的 [72, 67, 76, 89, 63, 60]、前馈 [41, 4, 82, 25, 74, 95, 88] 或程序化方法 [22]。根据训练数据，它们可以分为在3D数据上训练的模型 [4, 82, 25, 74, 95, 88]、在多视图2D图像上训练的模型 [72, 76, 89, 63, 60] 和在单视图图像上训练的模型 [81]。这些模型在复杂性、训练成本、所需数据量、生成质量、可控性和应用便利性等方面存在差异。

## 2.3 3D模型编辑与风格化

在几何生成阶段之后，需要将外观融入3D模型。在其他情况下，需要对现有几何模型的外观进行修改。这些修改通过编辑和外观风格化技术实现。某些方法专注于仅改变3D模型的视觉属性，例如，通过生成纹理和调整颜色属性 [12, 91, 3]，而其他方法则强调局部几何特征的细化 [80]。鉴于几何形状和外观之间的相互依赖性，一些方法旨在同时修改局部几何形状和颜色属性 [46, 21, 77]。

# 3 方法

## 3.1 公式化

我们的目标是生成具有详细外观的3D建筑模型，这些模型依赖于建筑物的轮廓、屋顶结构的提示和建筑风格的描述。考虑到可行性、控制精度和便利性之间的权衡，我们将建筑轮廓和屋顶结构作为图像输入，即掩模和高度草图，将风格描述作为文本输入。为此，我们提出了GeoTexBuild，这是一个模块化框架，允许用户轻松而准确地控制并生成多种建筑模式的风格化3D模型。

为了更好地理解我们的框架，我们首先在3.2节介绍两个重要的外部模块，然后在3.3节概述我们的处理流程，并在3.4节详细介绍每个组件。

## 3.2 基础知识

### 3.2.1 控制网络

控制网络 [94] 是一种强大的神经网络架构，用于指导文本到图像的扩散模型 [24]，并具有空间局部化的、特定任务的图像条件。它将一个可训练的扩散模型副本附加到主干上，以注入所需条件的空间信息。在这里，我们简要回顾其原理。

首先，空间条件图像  $c_i$  被传递到一个小型网络  $E(\cdot)$ ，以编码到特征空间，作为

$$c_f = E(c_i). \quad (1)$$

该向量随后作为一个重要输入。然后，控制网络对原始扩散模型实施了几项巧妙的修改。一个预训练的神经块  $F(\cdot; \theta)$  具有参数  $\theta$ ，将输入特征图  $x$  转换为输出  $y$ ，可以写成

$$y = F(x; \theta). \quad (2)$$

控制网络冻结原始块的参数  $\theta$ ，同时添加一个具有新参数  $\theta_c$  的可训练克隆。添加的块将外部条件向量  $c_f$  作为输入，并与锁定模型通过零卷积层连接，表示为  $Z(\cdot; \cdot)$ 。对于单个神经块

$F$ ，完整的ControlNet然后输出

$$y_c = F(x; \theta) + Z(F(x + Z(c_f; \theta_z); \theta_c); \theta_{z2}), \quad (3)$$

下标中的数字表示参数属于哪个层和类型。

对于整个稳定扩散模型，插入多个ControlNet模块以实现对“全局”和“详细”上下文不同尺度的精细控制。可训练的副本应用于扩散U-net的每个编码器层，输出被添加到跳跃连接和U-net的中间块中。ControlNet是一种计算效率高的架构，因为只有可训练的副本在微调时需要梯度反向传播，从而实现高速、低内存消耗的单GPU训练。

ControlNet允许多种控制方法，例如各种边缘图、分割图、深度图、掩模，甚至骨架姿势。它还支持对图像不同方面（如形状和颜色）的串联或同时控制。得益于ControlNet带来的极大便利，我们能够实现对底部形状的精细控制，不仅来自于轮廓，还能从任意草图中控制屋顶的高度和详细结构。这也帮助我们在仅使用2D条件的情况下实现低成本训练，相比于其他使用3D训练数据的3D生成模型。

### 3.2.2 Text2Mesh

Text2Mesh [46] 是一种通过自然语言或图像条件对给定粗糙网格进行风格化的方法。该模块将全局内容视为由给定3D网格规定的大结构，定义整体形状和拓扑，而风格则是由其颜色和表面细致几何细节决定的特定外观。它通过三个主要组件实现网格风格编辑：神经风格场 (NSF)、可微渲染器和相似性比较器。

输入网格  $\mathcal{M}$  具有顶点  $\mathcal{V} \in \mathbb{R}^{n \times 3}$  和面  $\mathcal{F} \in \{1, \dots, n\}^{m \times 3}$ ，首先被归一化为位于单位边界框内，并在整个操作过程中固定为参考全局内容。

神经风格场表示为三个MLP： $N_s$ 、 $N_c$  和  $N_d$ ，将单位空间中的点  $p$  映射为一个向量和一个标量，如下所示

$$(c_p, d_p) = NSF(\gamma(p)) \in (\mathbb{R}^3, \mathbb{R}), \quad (4)$$

其中  $\gamma(\cdot)$  是位置编码算子。当点位于网格的顶点集合中时，NSF输出的值是有效的，并作为顶点颜色  $c_p$  和位移  $d_p \cdot \vec{n}_p$  沿表面法线传递给网格。在实践中，场中的查询点被设置为  $\mathcal{V}$  中的顶点，然后通过给定的三角剖分由可微渲染器进行可视化。Text2Mesh在每次迭代中渲染  $n\theta = 5$  个修改网格的视图，并对渲染图像实施多种2D增强，以增强风格修改的细节。

然后，这些增强图像被传递到相似性比较器中以计算语义损失。

比较器的核心组件是一个预训练的CLIP模型 [62]，它在多模态中起作用

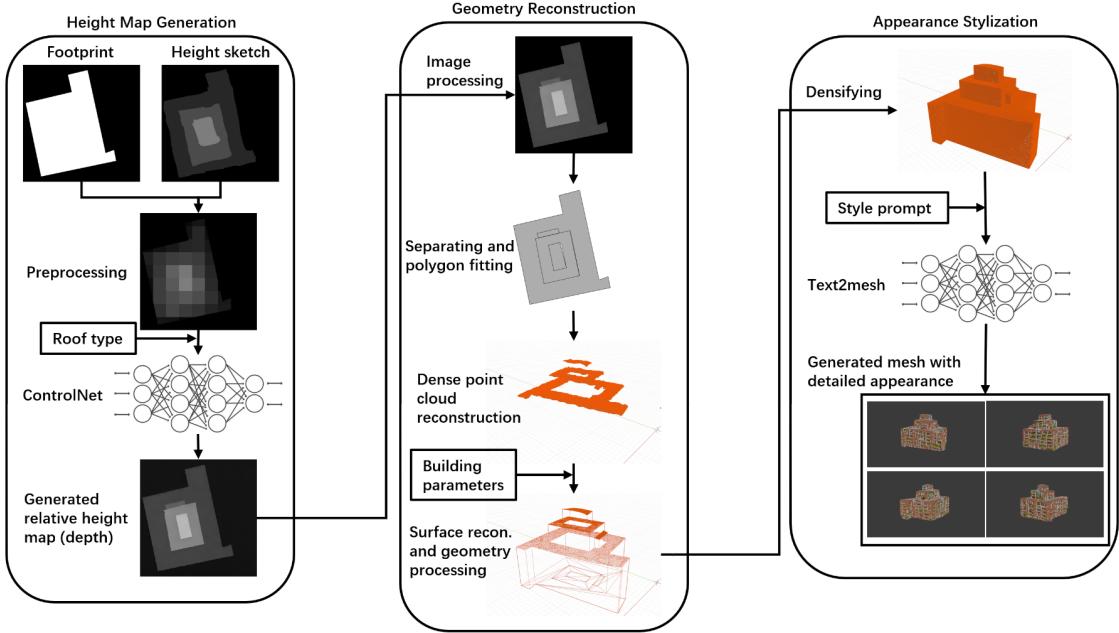


图3：我们管道的示意图。我们的框架有三个主要功能模块：高度图生成、几何重建和外观风格化。

嵌入空间。所需风格的文本描述  $t$  首先传入文本编码器，如  $\phi_{target} = E_t(t) \in \mathbb{R}^{512}$ 。然后，图像跟随其后，作为  $S^{type} = 1/n \sum_{\theta} (E_i(\psi_{type}(I_{\theta}))) \in \mathbb{R}^{512}$ ，其中  $\psi_{type}$  是相应类型的2D增强操作符。损失函数为：

$$L_{sim} = - \sum_S sim(S, \phi_{target}), \quad (5)$$

其中  $sim(a, b) = (a \cdot b) / (|a| \cdot |b|)$  是  $a$  和  $b$  之间的余弦相似度。损失通过可微渲染器反向传播，然后更新神经风格场中的MLP。当最小化损失时，渲染图像与提示之间的相似度增加，这意味着网格的外观逐渐接近文本描述。

Text2Mesh模块允许我们修改网格表面上的详细几何和颜色属性，因此，在我们的建筑模型生成框架中，它用于细化粗糙网格，以提供与提示词匹配的精细外观。

### 3.3 管道概述

我们在图3中展示了我们框架的示意图。我们的框架有三个主要功能模块：高度图生成、几何重建和外观风格化。我们从给定的建筑平面图和手绘的高度草图开始。高度图生成器中的ControlNet将输出一个规则和平滑的高度图，准备进行几何重建。下一个模块将首先处理生成的图像，以去除扩散模型带来的噪声，然后分离高度图中的每个不连续部分，并将其形状拟合为多边形。之后，我们根据高度图上的像素值重建屋顶结构的稠密点云，并进一步基于点云构建表面。根据输入的建筑参数构建更多的表面，例如立面。在几何重建的最后，我们处理粗糙网格，以去除小孔和裂缝，使表面光滑干净。最后，我们使用Text2Mesh模块和输入文本提示对稠密建筑网格进行风格化。输出的网格以标准3D数据格式保存，以便于后续应用或进一步手动修改。

## 3.4 GeoTexBuild的组件

### 3.4.1 高度图生成

我们的高度图生成模块基于ControlNet的一个变体，它将一个矩形的颜色调色板转换为常规的彩色图像。高度草图是决定生成高度图的主要控制因素，可以在几秒钟内轻松手绘。为了适合建筑生成，我们有以下要求：1) 图像上的像素值仅表示屋顶上相应点的高度，因此输入和输出应转换为灰度，即深度图。2) 为了确保生成建筑的底部与建筑平面完全匹配，后者应以某种形式作为输入。3) 为了对屋顶类型和形状进行细粒度控制，我们应能够使用不同的调色板网格大小和指示屋顶类型的多个固定模式的提示。

为此，我们将高度草图定义为一幅灰度图像，其值表示屋顶上结构的相对高度，值越大（越亮）表示越高。相对意味着它们不是绝对值，可以在后续操作中进行平移或缩放。与大多数使用线条边缘控制生成形状的ControlNet不同，我们决定将建筑平面作为二进制掩码  $I_n$  输入，以实现更强的控制并减少混淆。

预处理。 我们首先将草图处理为一个网格大小为  $n_g$  的方形亮度调色板，

$$I_p^{i,j} = \frac{1}{w_p^2} \sum_{\substack{(i-1)w_p < x \leq iw_p \\ (j-1)w_p < y \leq jw_p}} I_s(x, y), \quad (6)$$

其中  $w_p = w_s/n_g$  是调色板网格的宽度，  $w_s$  是草图图像的宽度（等于高度），  $I_p^{i,j}$  是第  $i$  行第  $j$  列方块的像素值。然后，亮度调色板被建筑平面掩盖，以创建清晰的边界，如

$$I_{mp} = I_p \odot I_m, \quad (7)$$

其中  $\odot$  是矩阵的哈达玛积。然后，  $I_{mp}$  是我们微调的亮度调色板ControlNet用于高度图生成的输入。

应用ControlNet。为了帮助ControlNet理解来自单一亮度调色板的不同可能类型的屋顶，我们构建了几种固定模式的提示文本  $t$ ，描述屋顶是否具有多个部分及其一般形状。提示在表2中列出。然后，生成过程可以表示为

$$I_h = CN(I_{mp}, t). \quad (8)$$

为了学习掩盖的亮度调色板图像与屋顶的相对高度图之间的映射，我们使用精心挑选的现实建筑屋顶重建数据集训练了ControlNet模型。数据集和训练细节将在4.1节中解释。

### 3.4.2 几何重建

图像处理。一旦我们获得相对高度图，就可以开始重建建筑的几何形状。

首先，由于相对高度图是由一个零卷积的ControlNet模型生成的，在控制边缘附近不可避免地会出现一些突发噪声。由于扩散模型的特性，其他噪声也会存在于平坦区域。我们使用双边滤波器[75]对图像进行滤波，以平滑高度值，同时保留不同部分的边缘，正如

$$I_{bf}(x) = \frac{1}{W(x)} \sum_{x' \in \Omega} I_h(x') \cdot g_r(\|x - x'\|) \cdot g_s(\|x - x'\|), \quad (9)$$

其中  $W(x)$  是归一化因子，  $\Omega$  是  $x$  的邻域，  $g_r$  和  $g_s$  是高斯核。为了确保足够的平滑，滤波器大小设置为7，  $\sigma_{gr} = 9$ ，  $\sigma_{gs} = 55$ 。我们使用形态学算子  $E(\cdot)$  将掩膜腐蚀一个像素，以进一步去除边缘附近的亮点，

$$I_{bfe} = I_{bf} \odot E(I_m). \quad (10)$$

分离和多边形拟合。在处理后的图像中，我们使用Vtracer（最先进的光栅到矢量图形转换器之一）将每个颜色区域拟合为一个多边形。在此模块中，像素

通过Kopelman算法对其进行聚类，然后从聚类中自下而上构建二叉树，以便在画布上进行分层布局。聚类是基于亮度差异进行的，我们将其颜色精度设置为6，这意味着一个聚类必须至少具有  $256/6 \approx 42$  的颜色差异，才能被视为一个独立的聚类。较低或较高的精度会导致颜色区域缺失或多边形中出现不必要的孔。这些聚类将经过路径行走、路径简化、平滑和曲线拟合，最终被矢量化为多边形。我们从Vtracer的输出矢量图中提取每个多边形作为独立的屋顶部分。通过这些矢量化的多边形，我们获得重建屋顶  $\{S_r\}$  外部  $\{\partial S_r\}$  顶点的位置。

将图像提升为点云。我们从每个矢量化多边形包围的高度图像素重建一个密集的点云。内部像素的位置和高度值直接设置为  $(x, y)$  和点的z坐标。通过去除统计异常值和z值低于输入高度限制  $h_{min}$  的噪声点来清理点云。

重建表面。点云的每个部分都被发送到法线估计器和Poisson表面重建[31]，八叉树深度设置为6，之后根据屋顶类型应用平滑。由于我们在具有相似方向法线的开放边界点云上进行重建，重建的屋顶表面  $\{\mathcal{S}_r\}$  将不会被封闭。这给我们带来了优势和劣势。

优势在于我们可以通过乘以输入建筑参数中的因子来调整绝对建筑高度与屋顶高度之间的比例。劣势在于我们发现重建后的直接修剪会破坏我们试图保留的平滑边界，导致条纹外立面。为了解决这个问题，我们利用布尔运算。我们将表面提升到所需的建筑高度，并从上到下挤出到地面，以构建每个屋顶部分的网格  $\{\mathcal{M}_r\}$ 。

与此同时，从  $\{\partial S_r\}$  自下而上挤出的棱柱  $\{M^{\partial S_r}\}$  被构建。然后，重建的最终网格是

$$\mathcal{M}_b = \cup(\mathcal{D}(\{\mathcal{M}_r\} \cap \{\mathcal{M}_{\partial S_r}\})), \quad (11)$$

其中  $\cap$  指的是两个集合中对应网格的交集，形态学算子  $\mathcal{D}(\cdot)$  指的是每个对象在  $x$  和  $y$  方向上的1单位膨胀。膨胀补偿了之前施加的腐蚀，并消除了多边形拟合造成的孔或裂缝。

### 3.4.3 外观风格化

通过几何重建，我们获得了一个表面光滑、没有细微结构或颜色的网格。

我们在这个功能模块中丰富了它的外观，使其更接近现实。网格  $\mathcal{M}_b$  首先被重新网格化，以在表面上加密三角形，以满足Text2Mesh的分辨率要求。我们将八叉树深度设置为8，以确保顶点能够准确地表现颜色和位移。重新网格化的模型和输入提示被传递给经过微调的CLIP的Text2Mesh模块。

我们在Text2Mesh的渲染器中进行了一些微调。在原始设置中， $n_\theta$  视图是围绕前视图以正态分布渲染的。然而，尽管在原始论文中对于小物体效果良好，这些视图并没有均匀覆盖建筑网格的表面。我们将方位角的分布修改为  $[0, 2\pi)$  的均匀分布，将仰角的分布修改为方差  $3\sigma = \pi/4$  的正态分布。

## 4 实验

为了完成整个生成过程，我们在Building3D数据集的特殊渲染上训练了一个定制的ControlNet模型用于高度图生成，并在高质量手工挑选的图像上微调了CLIP模型以进行外观风格化。详细信息见4.1和4.3节。几何重建的实验在4.2节中，完整流程的结果和分析在4.4节中。

### 4.1 定制ControlNet

#### 4.1.1 数据集

在这个模块中，我们尝试学习从具有亮度条件的建筑平面图到高度图的映射，因此，我们需要创建相应的配对数据。开源地图和城市数据，如OpenStreetMap，通常不包含足够准确的建筑高度数据，以显示房屋屋顶内部的高度变化。我们转向城市规模的建筑重建数据，并从它们的屋顶提取高度信息。幸运的是，Building3D [78] (图4) 提供了超过32,000个建筑重建。

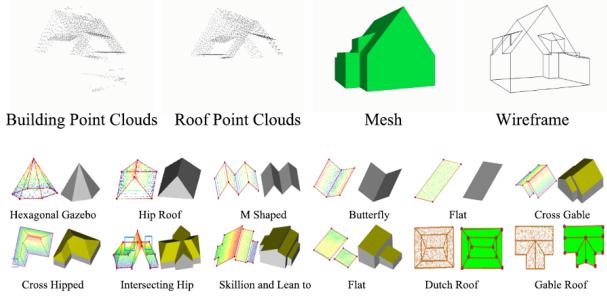


图4：来自Building3D数据集的数据示例。该数据集提供了超过32,000个以多种数据形式呈现的建筑重建。我们从填充表面的线框中渲染了深度图。

以点云、线框和网格的多种数据形式。我们选择了线框，因为它足够准确，而且没有像简单的点云那样嘈杂。我们在线框中插入了表面，并在Blender中渲染了正交顶视图；深度图和掩膜被用作ControlNet训练中的配对数据。我们通过旋转、翻转和随机裁剪对数据集进行了增强 $8\times$ 。

#### 4.1.2 训练

在训练过程中，渲染的深度图既作为输入也作为真实值。作为输入，它们经过第3.4.1节中描述的预处理，输出一个带掩码的亮度调色板图像。我们设置 $n_g$ 为一个从5到9的随机整数，以便在不同精度下进行条件控制。作为真实值，深度图用于计算扩散过程中的噪声预测器的损失。

为什么输入的部分可以与输出目标相同？这只是重建图像吗？这发生在训练中。回顾一下，我们的目标是从手绘草图生成，但来自现实世界的重建建筑没有配对的高度草图。多亏了填补渲染深度与手绘草图之间差距的预处理，我们可以以重建的方式训练模型，但以生成的方式进行推理。

为了准备不同屋顶类型的提示，我们根据顶结构和线框的几何形状对屋顶进行了分类。线框被视为图形，并根据其连通性分为单件或多件，根据节点的度数分为简单或复杂，以及根据节点的z坐标范围分为斜面或平面。我们为每个类别构建了简单的提示模式，如表2所示。

表2：不同屋顶类型的提示

连通性	节点度数	Z坐标范围	提示
多件	> 4(复杂)	倾斜的	灰度深度图，多个复杂的坡度和层，呈多边形形状，黑色背景
多件	/	平坦的	灰度深度图，多个平坦的层，呈多边形形状，黑色背景
多件	< 3(简单)	倾斜的	灰度深度图，多个简单的棚屋层，呈多边形形状，黑色背景
多件	= 3, 4(中等)	倾斜的	灰度深度图，多个坡度和层，呈多边形形状，黑色背景
单件	> 4(复杂)	倾斜的	灰度深度图，复杂斜坡组合的一个联合部分，呈多边形形状，黑色背景
单件	/	平坦的	灰度深度图，平面层，多边形形状，黑色背景
单件	< 3(简单)	倾斜的	灰度深度图，简单的棚屋层，多边形形状，黑色背景
单件	= 3, 4(中等)	倾斜的	灰度深度图，简单斜坡的连接部分，多边形形状，黑色背景

我们在260k对图像上，从Stable Diffusion 1.5基础模型训练了一个自定义的ControlNet，直到32618个全局步骤。批量大小设置为4，梯度累积步骤设置为4。在最后1000个步骤中，SD解码器被解锁，学习率从 $1e-5$ 变更为 $2e-6$ ，遵循ControlNet作者的指示。整个训练过程在62小时内使用单个Nvidia RTX 3090 GPU完成。

#### 4.1.3 比较不同的控制组合

我们在这里比较从足迹到高度图的不同控制组合，以展示掩模和亮度调色板组合带来的优越性。

这里是图5中使用相同超参数训练的四个模型，但输入了不同的控制：

(1)仅线条边缘，(2)线条边缘和亮度调色板，(3)仅掩模，(4)掩模和亮度调色板。  
结果显示，首先，在(1)(3)中，未使用亮度调色板时高度是随机生成的；其次，在(1)中，线条边缘导致内部和外部的区分模糊。至于(2)，

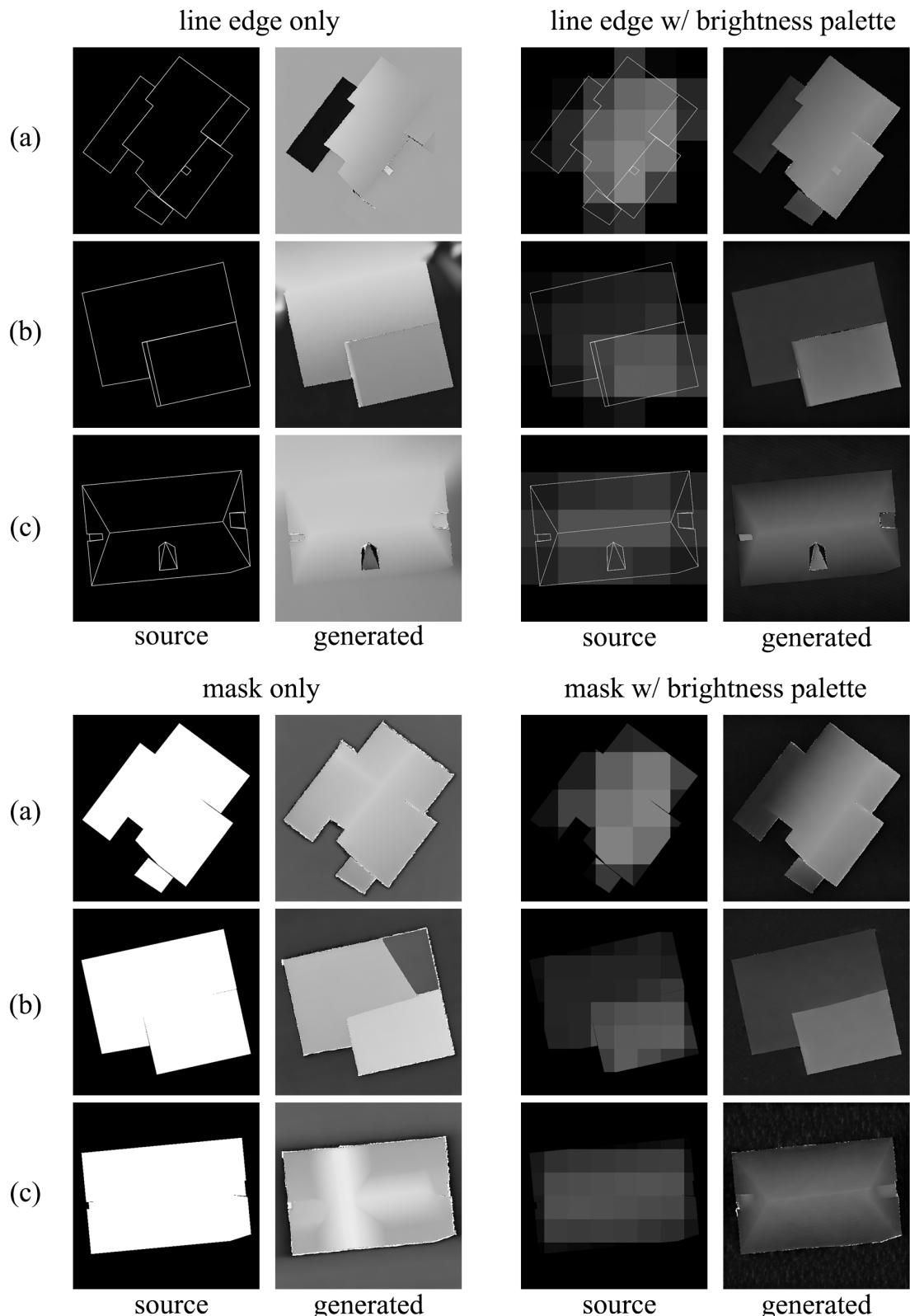


图5：四个定制的ControlNet模型输入不同的几何控制，使用相同的超参数进行训练。推理时使用的相应提示为“灰度深度图，简单坡度的一个连接部分，呈多边形形状，黑色背景”用于（a）（b），而“灰度深度图，复杂坡度组合的一个连接部分，呈多边形形状，黑色背景”用于（c）。

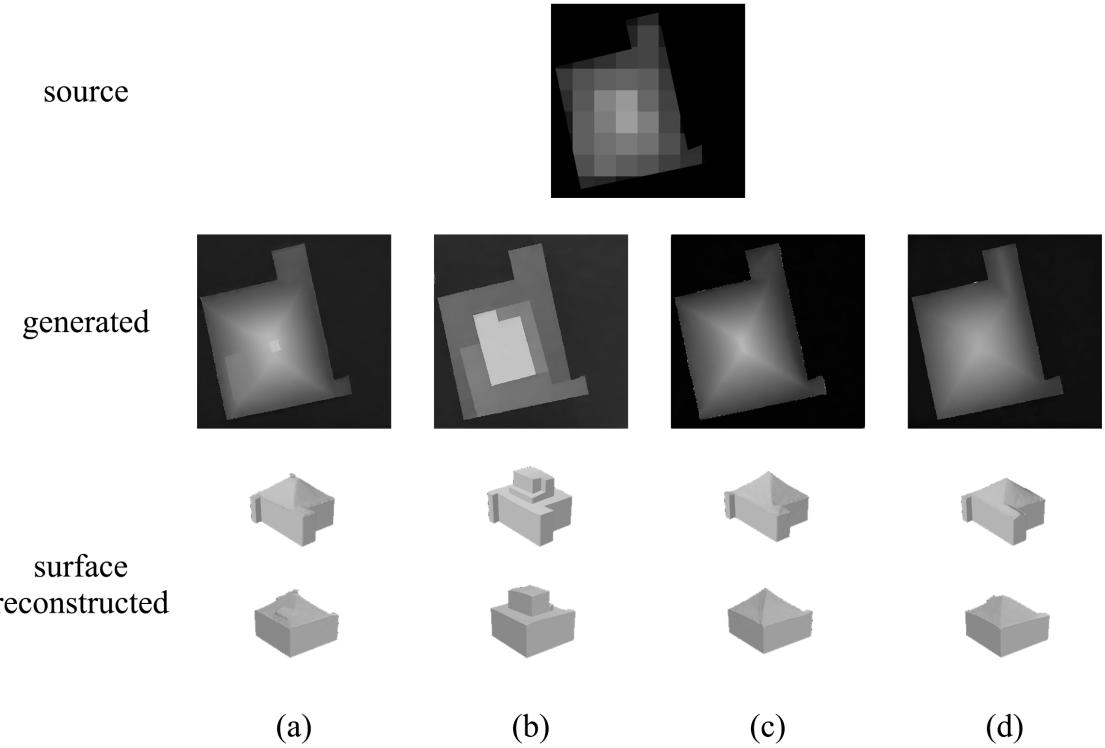


图6：通过在输入相同控制时更改不同的提示，我们展示了模型可以从单一亮度调色板理解不同可能类型的屋顶。相应的提示为“灰度深度图，复杂坡度和层的多个部分，呈多边形形状，黑色背景”用于（a），“灰度深度图，多个平坦层，呈多边形形状，黑色背景”用于（b），“灰度深度图，复杂坡度组合的一个连接部分，呈多边形形状，黑色背景”用于（c），“灰度深度图，简单坡度的一个连接部分，呈多边形形状，黑色背景”用于（d）。重建的表面也显示出来。提示（a）表示更多的层，例如烟囱、天窗和檐口的结构。由于提示的相似性，（c）和（d）的结果也相似。

同时绘制线条边缘和高度草图需要额外考虑它们的兼容性，但与（4）相比显示出相似的质量。尽管（4）丢失了一些非常小和精致的细节，但它在内部部分的边缘产生的噪声较少，这很难过滤。

此外，我们展示了通过改变屋顶类型的提示，模型可以生成与图6中相应的高度图。

## 4.2 几何重建

在这个组件中，主要是图像和几何操作，没有任何可训练的神经块。

我们使用Easy3D [51]和Blender构建了它们。在这里，我们展示了一些消融研究的比较。

无图像处理。重建的表面显示在图7的第二部分。图像处理阶段有助于去除噪声并保持重建屋顶部分的清晰边界。

不进行分离和多边形拟合。重建的表面显示在第七图的第三部分。在没有分离和多边形拟合阶段的情况下，各种屋顶部分合并，导致在泊松重建中出现过于平滑的表面或面包状边界伪影。虽然对单层屋顶的影响相对较小，但在第二列中应用于多层屋顶时则会变得灾难性。

不使用布尔操作。重建的表面显示在图7的第四部分。  
用简单的修剪替代布尔操作将破坏我们试图保持的平滑边界，导致条纹外立面。

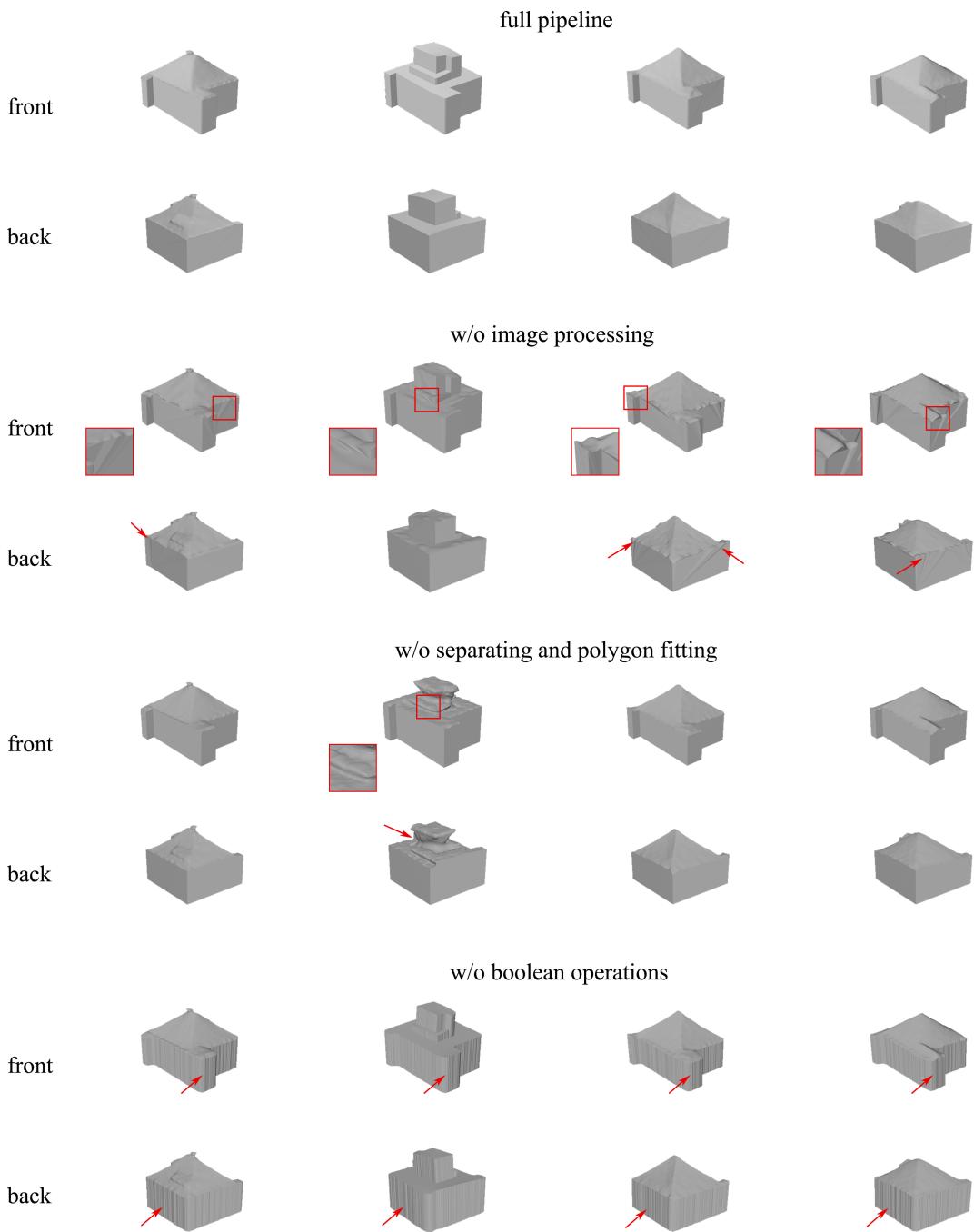


图7：几何重建的消融研究结果。请注意红框中或箭头指示的非规则形状和不均匀边缘，这些都是省略关键步骤的直接后果。

### 4.3 自定义Text2Mesh

当我们尝试使用默认的Text2Mesh与ViT-B/32 CLIP模型来美化外观时，发现该模型经常无法正确区分图像的语义部分。在严重的情况下，它甚至会将背景（如庭院、风景或树木）与前景（建筑物）混淆，从而在目标上绘制背景。由于生成过程的主要指导是相似性比较器，即CLIP模型，我们决定使用手动选择的高质量建筑图像对其进行微调。

#### 4.3.1 数据

为了解决背景混淆的问题，我们消除了用于微调的图像的背景，仅保留前景建筑。在Text2Mesh的渲染器中，背景设置为白色并使用遮罩，因此我们相应地在训练数据中的图像中填充了白色。我们从互联网上选择了800多张图像，涵盖了不同视角的各种类型建筑。

样式的相应文本描述由GPT-4生成，要求为“给我上传图像中建筑的提示，描述其形状、设计、材料、颜色和其他特征，不超过30个单词。”

#### 4.3.2 训练

我们使用OpenCLIP[11]代码库对来自'laion2b\_s34b\_b79k'检查点的ViT-B/32 CLIP模型进行了微调。该模型训练了16个周期，批量大小为64。使用一台RTX 3090在20分钟内完成。

#### 4.3.3 比较

我们展示了微调模型与我们的渲染视角调整和默认的Text2Mesh之间的定性比较，见图8的前两行，以及定量比较，见表3。默认模型倾向于给整个建筑赋予重复的纹理，或将其他背景（如绿化）视为建筑的一部分，而微调模型能够生成规则的建筑表面和细节。

### 4.4 结果

定性结果如图9所示。我们展示了纹理网格以及它们的俯视投影与底图之间的差异，见图10。样式提示与渲染视图之间的CLIP相似性得分的定量结果如表3所示。表中的数字是图8中显示的四个固定相机视角的平均得分。CLIP相似性计算为

$$sim_{CLIP} = \langle E_i(I), E_t(t) \rangle, \quad (12)$$

这是来自CLIP编码器的编码图像向量和编码文本向量的内积。

表3：生成模型与提示之间的CLIP相似性得分

提示	1	2	3	4	5	总平均值	GPU小时数
我们的	0.3164	0.2934	0.2945	0.2845	0.3113	0.3000	63(RTX3090)
w/o finetune	0.2898	0.1669	0.2912	0.2612	0.2606	0.2540	/
GradeADreamer	0.2759	0.1726	0.2351	0.1020	0.2321	0.2035	/
TripoSR 1.4	0.3295	0.2641	0.3115	0.2709	0.3039	0.2960	21100(A100)
Hunyuan3D 1.0	0.2745	0.1725	0.2589	0.1882	0.2530	0.2294	/

一座小房子，带有山墙屋顶，由石头和砖块建成，配有方形窗户和门，设计传统。

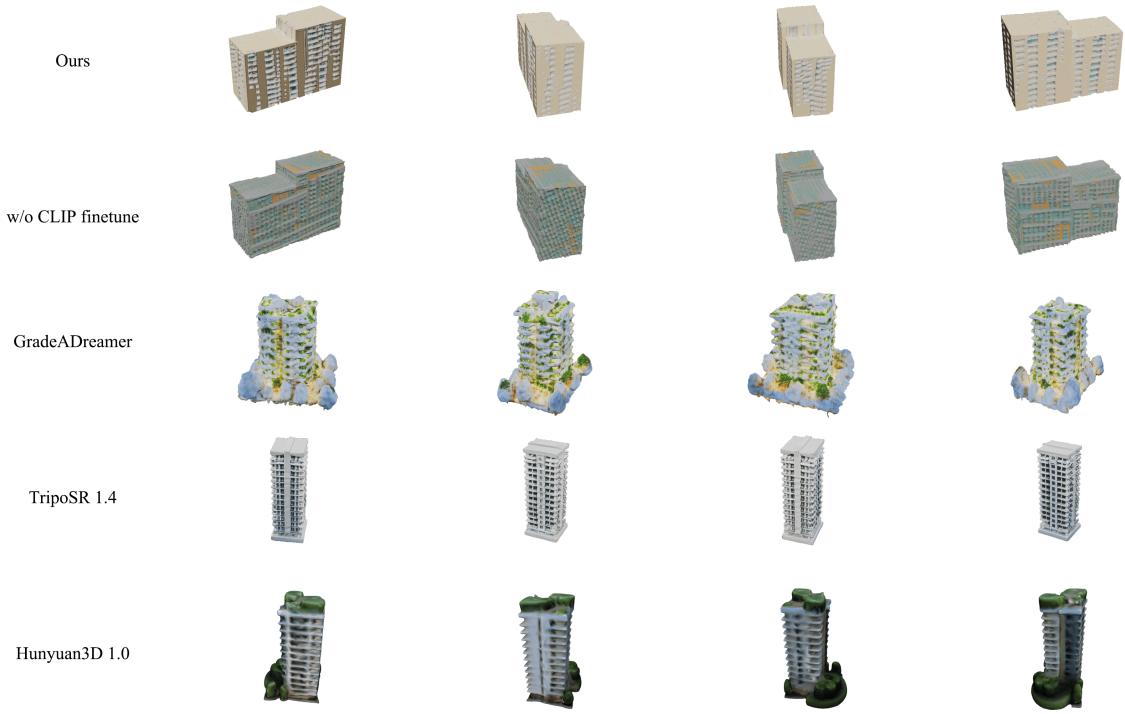
3: "一座矩形的多层建筑，外立面为米色，墙面呈网格图案，配有大型蓝色窗户，红色瓦屋顶，以及一扇棕色的单门，具有古典设计元素。"

4: "现代摩天大楼综合体，具有光滑的玻璃外立面、几何图案，以及蓝色和灰色的混合色调。"

5: "一座高大的现代高层建筑，拥有众多阳台、干净的线条、上部使用浅色材料、平屋顶，以及纹理基础。"

我们还将生成的网格与基于3D高斯分割的多视图生成方法[76]、基于在3D数据上训练的NeRF的直接3D生成方法[74]以及3D-DiT模型进行比较

Prompt: A tall, modern high-rise building with numerous balconies, clean lines, with light color materials, flat roof, and textured base.



Prompt: A multi-story residential building with a mix of brick and concrete facades, large glass windows, numerous balconies, and rooftop elements, creating a modern, urban appearance.

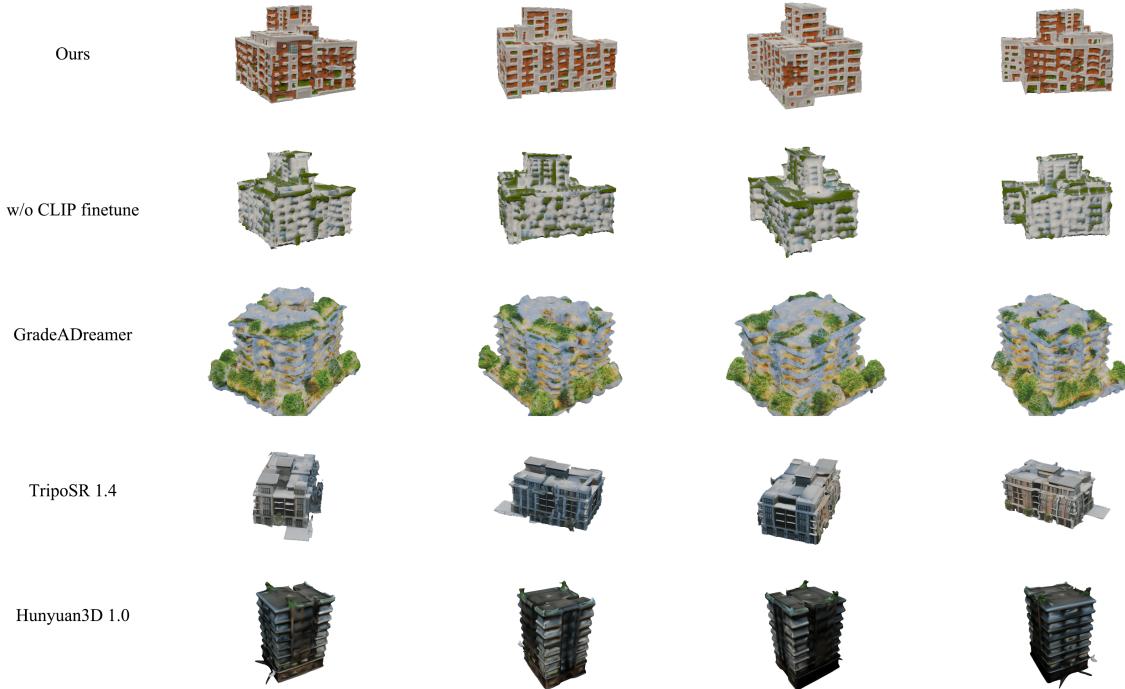


图8：不同生成方法的比较。请注意，在其他方法中，底部形状没有对轮廓的限制，生成模型的底部形状可能会有所不同。我们生成模型的轮廓比较见图10。

训练于3D数据[88]。由于缺乏对其足迹的控制方法，我们仅展示了从文本提示中进行外观建模的比较。除此之外，从3DGS提取的网格由于3D高斯的椭球形状，表面存在噪声。尽管直接的3D生成方法似乎能够生成详细的几何形状和纹理，但请注意，它们是通过更昂贵的3D数据训练的，并且使用了多达176个A100 GPU，持续超过21,100个GPU小时。

## 4.5 限制与进一步研究

如第4.4节所述，端到端直接3D生成模型[95, 74, 25, 82, 4, 41]，在全面的3D数据集上训练，并拥有广泛的GPU资源，可能会产生更优越的细致几何和纹理分辨率。然而，与数据格式和计算能力相关的限制阻碍了在这些前馈模型中实现有效的几何控制，以及它们在我们管道中的集成。因此，我们采用了一种替代方法，将生成过程分解为离散步骤，并提出了一个模块化框架，其中每个阶段在特定条件下生成目标数据。研究在渲染或文本之外集成几何控制的方法，同时降低直接3D生成模型的训练成本，以促进领域特定的微调，代表了未来研究的一个有前景的方向。

## 5 结论

在这项工作中，我们介绍了GeoTexBuild，这是一个用于从足迹、高度草图和风格提示生成纹理化3D建筑模型的模块化框架。我们研究了通过分离生成过程和利用高度图来控制建筑几何属性的方法。三个组件实现了高度图生成、几何重建、外观风格化，并构建了整个系统。

该框架通过集成定制的ControlNet和Text2Mesh模块，实现了对屋顶结构和详细外观的精细控制。生成的建筑网格可以轻松用于下游任务或进一步操作，从而显著减少人工劳动，并为设计师提供灵感。从各个角度的实验证实了我们框架在建筑模型生成中的有效性。

GeoTexBuild架起了创意愿景与技术执行之间的桥梁，使高度定制的建筑设计成为可能。它还为生成设计集成和自动化城市美学分析等创新奠定了基础。

## 参考文献

- [1] Al., M.B.E.: 深度学习在3D建筑重建中的应用：综述。在：第24届ISPRS大会，今日成像，明日预见（2022）
- [2] Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Mip-nerf 360：无界反锯齿神经辐射场。在：IEEE/CVF计算机视觉与模式识别会议论文集中。第5470–5479页（2022）
- [3] 本萨杜恩, R., 克莱曼, Y., 阿祖里, I., 哈罗什, O., 维达尔迪, A., 内维罗娃, N., 加夫尼, O.: Meta 3d 纹理生成: 快速且一致的3D物体纹理生成 (2024),<https://arxiv.org/abs/2407.02430>
- [4] 本萨杜恩, R., 莫尼耶, T., 克莱曼, Y., 科基诺斯, F., 西迪基, Y., 卡里亚, M., 哈罗什, O., 沙波瓦洛夫, R., 格雷厄姆, B., 加雷乌, E., 卡尔内瓦尔, A., 曹, A., 阿祖里, I., 马卡罗夫, I., 勒, E.T., 托伊索尔, A., 诺沃特尼, D., 加夫尼, O., 内维罗娃, N., 维达尔迪, A.: Meta 3d 生成 (2024),<https://arxiv.org/abs/2407.02599>
- [5] 布朗, F., 库珀, G., 福特, S., 阿乌阿德, G., 布兰登, P., 查尔德, T., 柯克汉姆, J., 奥克斯曼, R., 杨,B.: 一种集成的CAD方法: 建筑设计与施工中的建模概念。设计研究 16(3), 327–347 (1995). [https://doi.org/https://doi.org/10.1016/0142-694X\(94\)00002-U](https://doi.org/https://doi.org/10.1016/0142-694X(94)00002-U), <https://www.sciencedirect.com/science/article/pii/0142694X9400002U>
- [6] 布朗, T.B., 曼, B., 莱德, N., 苏比亚, M., 卡普兰, J., 达哈里瓦尔, P., 尼拉坎坦, A., 夏姆, P., 萨斯特里, G., 阿斯凯尔, A., 阿加瓦尔, S., 赫伯特-沃斯, A., 克鲁格, G., 亨尼汉, T., 查尔德, R.,

Prompt:

Modern multi-story building with flat roof, large windows, light and dark facade materials, small rooftop structure, possible commercial or institutional use.



Modern, multi-level house with flat roof, large wooden deck, spacious terrace, sleek lines, minimalist design, large windows, and outdoor furniture.



A modern architectural building with stacked cubic glass structures, reflective panels, sleek entrance, and geometric contemporary design.



A modern, multi-level building with a mix of white and gray materials, featuring large glass windows, vertical lines, and a unique, layered design.



Tall, modern high-rise, rectangular shape, beige and gray facade, numerous windows and balconies, clean lines, rooftop structures for ventilation. residential or commercial high-rise.



A multi-story residential building with a mix of brick and concrete facades, large glass windows, numerous balconies, and rooftop elements, creating a modern, urban appearance.



Modern skyscraper complex with sleek, glass facade, geometric patterns, and a mix of blue and gray tones.



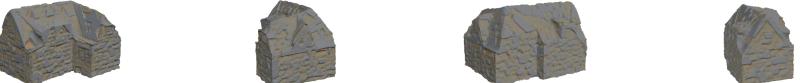
Modern two-story house with white walls, pitched dark gray roof with red trim, multiple large windows, chimneys.



Modern single-story house with slanted roof, stone and wood elements, large windows, wooden door with glass panels, stone columns, potted plants, warm exterior lighting.



Elegant stone house with steeply pitched roof, multiple gables, arched windows, grand entrance, prominent chimney, mix of traditional and modern elements.



A charming small wooden house with a gable roof, the house features a front porch with a railing and steps leading up to the entrance.



Small orange wooden house with steeply pitched roof, white central door, two flanking windows, small side window, simple rustic design, countryside setting.



Small wooden house with steeply pitched roof, white central door, two flanking windows, small side window, simple rustic design, countryside setting.



A small house with a gable roof built of stone and brick with square windows and doors, a traditional design.



图9：具有其风格提示的定性结果。

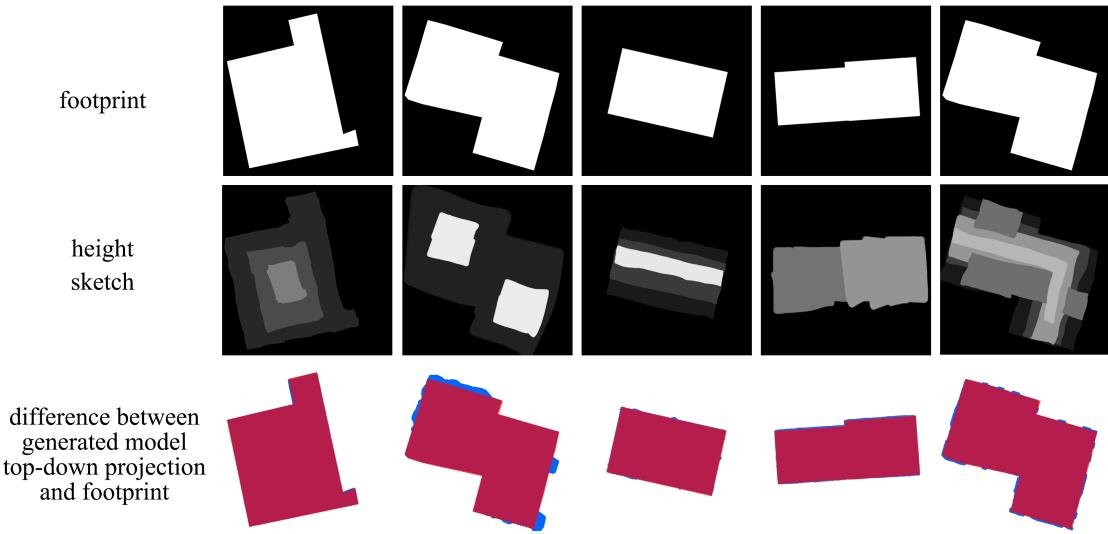


图10：我们展示了生成模型的足迹与自上而下投影之间的差异。投影为蓝色，足迹为品红色。因此，投影超过目标的区域用蓝色表示，而不足的区域用红色表示。

- 拉梅什, A., 齐格勒, D.M., 吴, J., 温特, C., 赫斯, C., 陈, M., 西格勒, E., 利特温, M., 格雷, S., 切斯, B., 克拉克, J., 伯纳, C., 麦肯德利什, S., 拉德福德, A., 萨茨基弗, I., 阿莫代, D. :语言模型是少量学习者。在：第34届国际神经信息处理系统会议论文集。NIPS '20, Currant Associates Inc., 纽约州红钩 (2020) [7] 陈, G., 王, W.:关于3D高斯喷溅的调查 (2025) , <https://arxiv.org/abs/2401.03890>
- [8] 陈, Q., 王, L., Waslander, S.L., 刘, X.: 一个端到端的形状建模框架, 用于从航空图像生成矢量化建筑轮廓。ISPRS摄影测量与遥感杂志 170, 114–126 (2020). <https://doi.org/10.1016/j.isprsjprs.2020.10.008>, <https://www.sciencedirect.com/science/article/pii/S092427162030280X>
- [9] 陈, R., 陈, Y., 焦, N., 贾, K.: Fantasia3d: 解耦几何和外观以创建高质量的文本到3D内容。发表于：IEEE/CVF国际计算机视觉会议 (ICCV) 论文集。第22246–22256页 (2023年10月)
- [10] 陈, Z., Ledoux, H., Khademi, S., 南, L.: 使用深度隐式场从点云重建紧凑建筑模型。ISPRS摄影测量与遥感杂志 194, 58–73 (2022). <https://doi.org/10.1016/j.isprsjprs.2022.09.017>, <https://www.sciencedirect.com/science/article/pii/S0924271622002611>
- [11] Cherti, M., Beaumont, R., Wightman, R., Wortsman, M., Ilharco, G., Gordon, C., Schuhmann, C., Schmidt, L., Jitsev, J.: 可重复的对比语言-图像学习的缩放法则。2023年IEEE/CVF计算机视觉与模式识别会议 (CVPR) 第2818–2829页 (2022), <https://api.semanticscholar.org/CorpusID:254636568>
- [12] Cheskidova, E., Arganaidi, A., Rancea, D.I., Haag, O.: 几何感知纹理。发表于：SIGGRAPH亚洲2023海报。SA '23, 计算机协会, 纽约, NY, 美国 (2023). <https://doi.org/10.1145/3610542.3626152>
- [13] Dawen, Y., Wei, S., Liu, J., Ji, S.: 从航空图像自动重建3D建筑的先进方法ISPRS - 国际摄影测量、遥感与空间信息科学档案 XLIII-B2-2020, 541–546 (08 2020). <https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-541-2020>

- [14] Debevec, P.E., Taylor, C.J., Malik, J.: 从照片建模和渲染建筑：一种混合几何和基于图像的方法。在：第23届计算机图形学与交互技术年会论文集。第11-20页。SIGGRAPH '96, 计算机协会, 纽约, NY, 美国 (1996). <https://doi.org/10.1145/237170.237191>
- [15] Dick, A., Torr, P., Cipolla, R.: 自动3D建筑建模。 (01 2000)
- [16] Dick, A., Torr, P., Cipolla, R.: 建模和解释建筑A.,Torr,P.,Cipolla,R.: 从多张图像建模和解释建筑。国际计算机视觉杂志 **60** (11 2004). <https://doi.org/10.1023/B:VISI.0000029665.07652.61>
- [17] Elaksher, A., Bethel, J.: 从激光雷达数据重建3D建筑。国际摄影测量遥感与空间信息科学档案 **34** (05 2002)
- [18] Fei, B., Xu, J., Zhang, R., Zhou, Q., Yang, W., He, Y.: 3D高斯点云作为新时代：一项调查。IEEE可视化与计算机图形学交易 PP (2024), <https://api.semanticscholar.org/CorpusID:267626952>
- [19] 高凯, 高阳, 何辉, 陆东, 许立, 李杰: Nerf: 3D视觉中的神经辐射场, 综合评述 (2023), <https://arxiv.org/abs/2210.00379>
- [20] Guédon, A., Lepetit, V.: Sugar : 用于高效3D网格重建和高质量网格渲染的表面对齐高斯喷溅。CVPR (2024)
- [21] 哈克, A., 坦西克, M., 埃夫罗斯, A.A., 霍林斯基, A., 叶川, A.: Instruct-nerf2nerf: 使用指令编辑3D场景. 2023 IEEE/CVF 国际计算机视觉会议 (ICCV) 第19683–19693页 (2023),<https://api.semanticscholar.org/CorpusID:257663414>
- [22] Hart, J.C.: 线性分形建模的对象实例化范式。在：1992年图形界面会议论文集。第224–231页。摩根·考夫曼出版社, 旧金山, 加利福尼亚州, 美国 (1992)
- [23] Henricsson, O., Gruen, A.: 自动化的建筑三维重建和城市模型可视化 (10 1996)
- [24] Ho, J., Jain, A., Abbeel, P.: 去噪扩散概率模型。在：第34届国际神经信息处理系统会议论文集。NIPS '20, Curran Associates Inc., 美国纽约州红钩 (2020)
- [25] 洪, Y., 张, K., 顾, J., 毕, S., 周, Y., 刘, D., 刘, F., Sunkavalli, K., Bui, T., 谭, H.: LRM: 单幅图像到3D的大型重建模型。发表于：第十二届国际学习表征会议 (2024) , <https://openreview.net/forum?id=s11U8vvsFF>
- [26] 黄, B., 余, Z., 陈, A., 盖格, A., 高, S.: 用于几何准确的2D高斯点云辐射场。在：SIGGRAPH 2024 会议论文。计算机协会 (2024). <https://doi.org/10.1145/3641519.3657428>
- [27] 黄, J., 斯托特, J., 彼得斯, R., 南, L.: City3d: 基于空中激光雷达点云的大规模建筑重建。遥感 **14** (9) (2022). <https://doi.org/10.3390/rs14092254>, <https://www.mdpi.com/2072-4292/14/9/2254>
- [28] 黄, I., 金, H., 金, Y.M.: Text2scene: 基于文本的室内场景风格化与部件感知细节. 在: IEEE/CVF计算机视觉与模式识别会议论文集 (CVPR). 第1890–1899页 (2023年6月)
- [29] Jun, H., Nichol, A.: Shap-e: 生成条件3D隐式函数 (2023) , <https://arxiv.org/abs/2305.02463>
- [30] Kada, M., McKinley, L.: 基于细胞分解方法的激光雷达三维建筑重建。国际摄影测量、遥感与空间信息科学档案 **38** (09 2009)

- [31] Kazhdan, M., Bolitho, M., Hoppe, H.: 泊松表面重建。在：第四届欧洲图形学几何处理研讨会论文集。第 61-70 页。SGP '06，欧洲图形学协会，戈斯拉尔，德国（2006）
- [32] Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 实时辐射场渲染的3D高斯点云场景。ACM图形学会会刊42(4) (2023年7月),<https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/>
- [33] Lafarge, F., Descombes, X., Zerubia, J., Deseilligny, M.: 一种自动建筑重建方法：使用高分辨率卫星图像的结构方法。在：国际图像处理会议。第1205 - 1208页（2006年11月）。<https://doi.org/10.1109/ICIP.2006.312541>
- [34] 李, H.H., Savva, M., Chang, A.X.: 文本到3D形状生成 (2024)
- [35] 李, J., 李, S., 乔, C., 林, W., 宣, J., 尹, S.E.: SemCity: 语义场景  
三平面扩散生成。在：2024年IEEE/CVF计算机视觉与模式识别会议 (CVPR)。第28337-28347页。  
IEEE计算机学会，美国加利福尼亚州洛杉矶（2024年6月）。<https://doi.org/10.1109/CVPR52733.2024.02677>  
[ieeecomputersociety.org/10.1109/CVPR52733.2024.02677](https://ieeecomputersociety.org/10.1109/CVPR52733.2024.02677)
- [36] 李华, 施浩, 张伟, 吴伟, 廖阳, 王磊, 李立华, 周鹏宇：梦场：  
基于3D高斯的文本到3D场景生成通过形成模式采样。在：Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., Varol, G. (编) 计算机视觉 - ECCV 2024。  
第214-230页。施普林格自然瑞士, 查姆 (2025) [37]
- 李强, 牟丽, 孙阳, 华阳, 施阳, 朱小星：遥感影像中建筑提取的综述：几何结构和语义属性。IEEE地  
球科学与遥感学报 62, 1-15 (2024)。<https://doi.org/10.1109/TGRS.2024.3369723>
- [38] 李晓, 张强, 康东, 程伟, 高阳, 张杰, 梁志, 廖杰, 曹阳平, 单宇：3D生成的进展：一项调查  
(2024) , <https://arxiv.org/abs/2401.17807>
- [39] 李志, 穆勒, 艾文斯, 泰勒, R.H., Unberath, M., 刘明阳, 林承浩：Neuralangelo：  
高保真神经表面重建。在：IEEE计算机视觉与模式识别会议 (CVPR) (2023)
- [40] 林杰, 李志, 唐晓, 刘杰, 刘思, 刘杰, 卢阳, 吴晓, 徐晟, 严阳, 杨伟：  
Vastgaussian: 大场景重建的广泛3D高斯。在：CVPR (2024)
- [41] 龙, X., 郭, Y., 林, C., 刘, Y., 斗, Z., 刘, L., 马, Y., 张, S.H., 哈贝尔曼, M.,  
西奥巴特, C., 王, W.: Wonder3d: 使用跨域扩散将单幅图像转换为3D。2024  
IEEE/CVF计算机视觉与模式识别会议 (CVPR) 第9970-9980页 (2023) ,  
<https://api.semanticscholar.org/CorpusID:264436465>
- [42] 罗, H., 张, J., 刘, X., 张, L., 刘, J.: 从多视角图像进行大规模3D重建：综合评述。遥感 16(5) (2024)。<https://doi.org/10.3390/rs16050773>,  
<https://www.mdpi.com/2072-4292/16/5/773>
- [43] 林奇, K., 哈克, G.: 场地规划。麻省理工学院出版  
社 (1984) [44] 梅里特, F.S.: 建筑设计与施工手册。麦格劳-希尔 (2001) [45]
- 梅策尔, G., 理查德森, E., 帕塔什尼克, O., 吉尔耶斯, R., 科恩-奥尔, D.: 潜在神经辐射场用于形状引  
导的3D形状和纹理生成。2023 IEEE/CVF计算机视觉与模式识别会议 (CVPR) 第12663-12673页  
(2022) , <https://api.semanticscholar.org/CorpusID:253510536>
- [46] 米歇尔, O., 巴尔-昂, R., 刘, R., 贝奈姆, S., 哈诺卡, R.: Text2mesh: 文本驱动的  
Text2mesh: 文本驱动的  
网格神经风格化。在：2022 IEEE/CVF计算机视觉会议  
和模式识别 (CVPR)。第13482-13492页。IEEE计算机学会，洛斯阿拉米托斯，加利福尼亚州，  
美国 (2022年6月)。<https://doi.org/10.1109/CVPR52688.2022.01313>, <https://doi.ieeecomputersociety.org/10.1109/CVPR52688.2022.01313>

- [47] Milde, J., Zhang, Y., Brenner, C., Plümer, L., Sester, M.: 基于形式语法的结构描述进行建筑重建。国际摄影测量、遥感与空间信息科学档案 37 (2008年01月)
- [48] Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: 将场景表示为神经辐射场以进行视图合成。在 : ECCV (2020)
- [49] Moore, M.: 游戏设计基础。CRC出版社 (2016)
- [50] Müller, T., Evans, A., Schied, C., Keller, A.: 具有 多分辨率哈希编码的即时神经图形原语。ACM图形学期刊 41(4), 102:1–102:15 (2022年7月)。  
<https://doi.org/10.1145/3528223.3530127>, <https://doi.org/10.1145/3528223.3530127>
- [51] Nan, L.: Easy3d: 一个轻量级、易于使用且高效的C++库，用于处理和渲染 3D数据。开放源代码软件期刊6(64), 3255 (2021). <https://doi.org/10.21105/joss.03255>, <https://doi.org/10.21105/joss.03255>
- [52] Nan, L., Wonka, P.: Polyfit: 从点云进行多边形表面重建。在 : 2017年IEEE国际计算机视觉会议 (ICCV) 。第2372–2380页 (2017年)。  
<https://doi.org/10.1109/ICCV.2017.258>
- [53] Nichol, A., Jun, H., Dhariwal, P., Mishkin, P., Chen, M. : Point-e : 一个用于从复杂提示生成3D点云的系统 (2022) , <https://arxiv.org/abs/2212.08751>
- [54] Nishita, S., Bousseau, A., Aliaga, D.G. : 程序化 建模 的 一个 建筑- 来自单一图像的计算机图形论坛 37(2), 415–429 (2018).  
<https://doi.org/https://doi.org/10.1111/cgf.13372>
- [55] 欧阳, H., 希尔, K., 隆巴尔迪, S., 孙, T.: Text2immersion: 生成沉浸式场景与3D 高斯分布 (2023), <https://arxiv.org/abs/2312.09242>
- [56] Páden, I., Peters, R., García-Sánchez, C., Ledoux, H.: 自动高细节建筑- 自动高细节建筑- 城市微观尺度模拟的重建工作流程。建筑与环境 265, 111978 (2024). <https://doi.org/https://doi.org/10.1016/j.buildenv.2024.111978>, <https://www.sciencedirect.com/science/article/pii/S0360132324008205>
- [57] 庞, H.E., 比尔耶基, F.: 使用深度学习从单一街景图像进行3D建筑重建。应用地球观测与地理信息国际期刊 112, 102859 (2022). <https://doi.org/https://doi.org/10.1016/j.jag.2022.102859>, <https://www.sciencedirect.com/science/article/pii/S1569843222000619>
- [58] 帕里什, Y.I.H., 米勒, P.: 城市的程序化建模。在 : 第28届计算机图形学与交互技术年会论文集 。第301–308页。SIGGRAPH '01, 计算机协会, 纽约, NY, 美国 (2001). <https://doi.org/10.1145/383259.383292>
- [59] 帕尔托维, T., 弗朗多费尔, F., 巴赫曼亚尔, R., 黄, H., 莱因哈茨, P.: 从超高分辨率立体卫星影像自动重建3D建筑模型。遥感 11(14)
- [60] 普雷萨·雷耶斯, M.E., 陈, S.C.: 一个用于风暴潮洪水动画的3D虚拟环境。在 : 2017 IEEE第三届国际多媒体大数据会议 (BigMM)。第244–245页 (2017).  
<https://doi.org/10.1109/BigMM.2017.754>
- [61] Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: 从自然语言监督中学习可转移的视觉模型。在 : 国际机器学习会议 (2021) , <https://api.semanticscholar.org/CorpusID:231591445>
- [62] Raj, A., Kaza, S., Poole, B., Niemeyer, M., Mildenhall, B., Ruiz, N., Zada, S., Aberman, K., Rubenstein, M., Barron, J., Li, Y., Jampani, V.: Dreambooth3d: 基于主题的文本到3D生成。ICCV (2023)

- [64] Ren, X., Huang, J., Zeng, X., Museth, K., Fidler, S., Williams, F.: Xcube: 使用稀疏体素层次的大规模3D生成建模。在：IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集中。第4209-4219页（2024年6月）
- [65] Schönberger, J.L., Frahm, J.M.: 运动重建结构再探。2016年IEEE计算机视觉与模式识别会议（CVPR）第4104-4113页（2016），<https://api.semanticscholar.org/CorpusID:1728538>
- [66] Shang, Y., Lin, Y., Zheng, Y., Fan, H., Ding, J., Feng, J., Chen, J., Tian, L., Li, Y.: Urbanworld:一个用于3D城市生成的城市世界模型（2024），<https://arxiv.org/abs/2407.11965>
- [67] Shi, Y., Wang, P., Ye, J., Mai, L., Li, K., Yang, X.: Mvdream: 多视角扩散用于3D生成。arXiv:2308.16512 (2023)
- [68] Shiode, N.: 3D城市模型：城市环境数字建模的最新进展在三维中。GeoJournal 52, 263-269 (2012年04月)。<https://doi.org/10.1023/A:1014276309416>

<https://doi.org/10.3390/en16052254>, <https://www.mdpi.com/1996-1073/16/5/2254>

- [71] Tancik, M., Casser, V., Yan, X., Pradhan, S., Mildenhall, B., Srinivasan, P.P., Barron, J.T., Kretzschmar, H.: Block-nerf: 可扩展的大场景神经视图合成。CVPR (2022)
- [72] 唐杰, 任杰, 周辉, 刘志, 曾国辉 : Dreamgaussian : 用于高效3D内容创作的生成高斯喷溅。ICLR (2024)
- [73] 陶, Y., 安赫尔·穆尼奥斯-巴尼翁, M., 张, L., 王, J., 傅, L.F.T., 法伦, M.: 牛津尖塔数据集 : 大规模激光雷达-视觉定位、重建和辐射场方法的基准测试 (2024), <https://arxiv.org/abs/2411.10546>
- [74] Tochilkin, D., Pankratz, D., Liu, Z., Huang, Z., Letts, A., Li, Y., Liang, D., Laforte, C., Jampani, V., Cao, Y.P.: Triposr: 从单张图像快速重建3D物体 (2024), <https://arxiv.org/abs/2403.02151>
- [75] Tomasi, C., Manduchi, R.: 双边滤波用于灰度和彩色图像。在：第六届国际计算机视觉会议论文集。第839页。ICCV '98, IEEE计算机学会, 美国 (1998)
- [76] 乌卡拉波尔, T., 普鲁沃斯特, K.: Gradeadreamer: 使用高斯喷涂和多视角扩散增强文本到3D生成 (2024)
- [77] 王杰, 方杰, 张晓, 谢丽, 田强 : GaussianEditor : 编辑3D GaussianEditor: 编辑3D 通过文本指令精细处理高斯分布。发表于 : 2024年IEEE/CVF计算机视觉与模式识别会议（CVPR）。IEEE计算机学会, 加利福尼亚州洛杉矶 (2024年6月)。<https://doi.org/10.1109/CVPR52733.2024.01975> <https://doi.ieeecomputersociety.org/10.1109/CVPR52733.2024.01975>
- [78] 王, R., 黄, S., 杨, H.: Building3d: 一个用于学习点云屋顶结构的城市规模数据集和基准。2023 IEEE/CVF国际计算机视觉大会(ICCV) 第20019-20029页 (2023), <https://api.semanticscholar.org/CorpusID:260125967>

<https://doi.org/10.1109/GEOINFORMATICS.2010.5567608>

- [80] 王, X., 朱, J., 叶, Q., 霍, Y., 冉, Y., 钟, Z., 陈, J.: Seal-3D: 交互式像素- Seal-3D: 互动像素- 神经辐射场的级别编辑。在 : 2023 IEEE/CVF国际计算机视觉会议 (ICCV)。第17637-17647页。IEEE计算机学会, 加利福尼亚州洛杉矶, 美国 (2023年10月)。<https://doi.org/10.1109/ICCV51070.2023.01621> <https://doi.ieeecomputersociety.org/10.1109/ICCV51070.2023.01621>

- [81] 魏, Y., 沃斯曼, G., 杨, M.Y.: Buildiff: 使用单幅图像的3D建筑形状生成条件点云扩散模型。在：IEEE/CVF国际计算机视觉会议（ICCV）研讨会论文集。第2910–2919页（2023年10月）
- [82] 吴, S., 林, Y., 张, F., 曾, Y., 许, J., 托尔, P., 曹, X., 姚, Y.: Direct3d: 可扩展的图像到3D生成通过3D潜在扩散变换器。神经信息处理系统进展 37, 121859–121881 (2024)
- [83] 吴, Y., 薛, F., 李, M., 陈, S.H.: 一种新颖的建筑截面骨架用于从点云进行紧凑的3D重建: 高密度城市场景的研究。ISPRS摄影测量与遥感杂志 209, 85–100 (2024). <https://doi.org/https://doi.org/10.1016/j.isprsjprs.2024.01.020> <https://www.sciencedirect.com/science/article/pii/S0924271624000297>
- [84] 维索基, O., 霍格纳, L., 斯蒂拉, U.: MLS2Iod3: 使用MLS点云细化低LOD建筑模型以重建语义LOD3建筑模型。在：3D地理信息科学的最新进展. pp. 367–380. 施普林格自然瑞士, 查姆 (2024)
- [85] 熊, B., 郑, N., 刘, J., 李, Z.: Gauu-scene v2: 使用3DGS和NeRF评估基于图像的度量的可靠性，基于扩展的激光雷达图像数据集。ArXiv abs/2404.04880 (2024), <https://api.semanticscholar.org/CorpusID:269005139>
- [86] 许, Q.C., 穆, T.J., 杨, Y.: 基于深度学习的3D形状生成综述。计算视觉媒体 9(3), 407–442 (2023年9月). <https://doi.org/10.1007/s41095-022-0321-5>
- [87] 许, Y., 吴, Y., 王, Y., 萨, I., 段, Y., 李, Y., 吉, P., 李, H.: Sketch2scene: 从用户的随意草图自动生成互动3D游戏场景。ArXiv abs/2408.04567 (2024), <https://api.semanticscholar.org/CorpusID:271769138>
- [88] 杨, X., 施, H., 张, B., 杨, F., 王, J., 赵, H., 刘, X., 王, X., 林, Q., 余, J., 王, L., 许, J., 何, Z., 陈, Z., 刘, S., 吴, J., 连, Y., 杨, S., 刘, Y., 杨, Y., 王, D., 江, J., 郭, C.: Hunyuan3d 1.0: 一个统一的文本到3D和图像到3D生成框架 (2025), <https://arxiv.org/abs/2411.02293>
- [89] 易, T., 方, J., 王, J., 吴, G., 谢, L., 张, X., 刘, W., 田, Q., 王, X.: 高斯-dreamer: 通过连接2D和3D扩散模型从文本快速生成3D高斯分布。在：CVPR (2024)
- [90] 于浩轩, 段辉, 赫尔曼, 威尔士·弗里曼, 吴俊: Wonderworld: 从单幅图像生成交互式3D场景。CoRR abs/2406.09394 (2024), <https://doi.org/10.48550/arXiv.2406.09394>
- [91] 于晓阳, 于佳欣, 周立波, 魏阳, 欧丽丽: Instantstylegaussian: 使用3D高斯点云的高效艺术风格迁移 (2024), <https://arxiv.org/abs/2408.04249>
- [92] 于阳, 哈迪, 李凯, 崔杰, 具博: Archshapesnet: 用于基准测试建筑信息建模元素分类算法的新数据集。计算设计与工程杂志 9(4), 1449–1466 (2022年7月). <https://doi.org/10.1093/jcde/qwac064>
- [93] 张杰, 李晓, 万志, 王超, 廖杰: Text2NeRF: 基于文本驱动的3D场景生成 Text2NeRF: 基于文本驱动的3D场景生成与神经辐射场。IEEE可视化与计算机图形学汇刊 30(12), 7749–7762 (2024年12月). <https://doi.org/10.1109/TVCG.2024.3361502>
- [94] 张, L., 饶, A., 阿格拉瓦, M.: 向文本到图像扩散模型添加条件控制。在：IEEE国际计算机视觉会议 (ICCV) (2023)
- [95] 赵, Z., 赖, Z., 林, Q., 赵, Y., 刘, H., 杨, S., 冯, Y., 杨, M., 张, S., 杨, X., 石, H., 刘, S., 吴, J., 连, Y., 杨, F., 唐, R., 何, Z., 王, X., 刘, J., 左, X., 陈, Z., 雷, B., 翁, H., 许, J., 朱, Y., 刘, X., 许, L., 胡, C., 杨, S., 张, S., 刘, Y., 黄, T., 王, L., 张, J., 陈, M., 董, L., 贾, Y., 蔡, Y., 余, J., 唐, Y., 张, H., 叶, Z., 何, P., 吴, R., 张, C., 谭, Y., 肖, J., 陶, Y., 朱, J., 薛, J., 刘, K., 赵, C., 吴, X., 胡, Z., 秦, L., 彭, J., 李, Z., 陈, M., 张, X., 牛, L., 王, P., 王, Y., 匡, H., 范, Z.,

- 郑, X., 庄, W., 何, Y., 刘, T., 杨, Y., 王, D., 刘, Y., 江, J., 黄, J., 郭, C.: Hunyuan3d 2.0: 扩展扩散模型以生成高分辨率纹理3D资产 (2025),  
<https://arxiv.org/abs/2501.12202>
- [96] 赵, Z., 刘, W., 陈, X., 曾, X., 王, R., 程, P., 傅, B., 陈, T., 余, G., 高, S.: 米开朗基罗: 基于形状-图像-文本对齐潜在表示的条件3D形状生成. 在: 第三十七届神经信息处理系统会议 (2023),  
<https://openreview.net/forum?id=xmxgMij3LY>
- [97] 周, L., 杜, Y., 吴, J.: 通过点-体素扩散进行3D形状生成和补全. 在:  
IEEE/CVF国际计算机视觉会议 (ICCV) 论文集. 第5826-5835页  
(2021年10月)
- [98] 庄, J., 李, G., 许, H., 许, J., 田, R.: 文本到城市: 可控的3D城市- 禁止使用潜在扩散模型生成区块。在：  
文本到城市： 可控的3D ur- CAADRIA 2024 (04 2024).  
<https://doi.org/10.52842/conf.caadria.2024.2.169>