

DOI:10.11918/202505066

ChatHouseDiffusion: 提示词引导的建筑平面图生成与编辑方法

覃思中¹, 陈思齐¹, 何承昱¹, 陈巧云¹, 杨森¹, 廖文杰², 顾焱¹, 陆新征¹

(1. 清华大学 土木水利学院, 北京 100084; 2. 西南交通大学 土木工程学院, 成都 610031)

摘要: 建筑平面图的生成与编辑是建筑智能设计中的关键环节, 需要兼具灵活性与设计效率。针对现有方法过度依赖信息输入、缺乏交互性以及难以实现局部精确编辑等问题, 文中提出 ChatHouseDiffusion 模型。该方法首先通过大语言模型将用户自然语言输入解析为结构化 JSON 提示信息, 继而利用 Graphormer 捕捉房间之间的拓扑关系, 最后通过扩散模型在房间轮廓约束下生成满足条件的建筑平面图。研究表明: 在编辑阶段, 借助交叉注意力图的替换机制可以实现对局部区域的精确修改, 避免整体重构。基于 RPLAN 数据集的测试表明, ChatHouseDiffusion 在微平均交并比 (Micro-IoU) 和宏平均交并比 (Macro-IoU) 两项评价指标上均优于现有模型, 尤其在使用准确输入条件时, 其生成结果与真实标注高度一致, 展现出良好的实用性与泛化能力。该模型不仅严格遵循用户要求, 更能通过交互功能实现更直观的设计流程, 为智能化建筑平面图设计提供了新的技术路径。同时基于该方法开发了一个支持绘图、文本输入与交互式编辑的可视化设计平台, 提升了建筑平面图设计的实用性与可操作性。

关键词: 建筑平面图生成; 建筑平面图编辑; 交互设计; 扩散模型; 大语言模型; Graphormer

中图分类号: TU24 **文献标志码:** A **文章编号:** 0367-6234(2025)12-0245-09

ChatHouseDiffusion: Prompt-guided generation and editing of floor plans

QIN Sizhong¹, CHEN Siqi¹, HE Chengyu¹, CHEN Qiaoyun¹, YANG Sen¹, LIAO Wenjie², GU Yi¹, LU Xinzheng¹

(1. School of Civil Engineering, Tsinghua University, Beijing 100084, China;

2. School of Civil Engineering, Southwest Jiaotong University, Chengdu 610031, China)

Abstract: The generation and editing of floor plans are critical in intelligent architectural planning, requiring a high degree of flexibility and efficiency. In response to the limitations of existing methods—such as excessive reliance on input information, lack of interactivity, and insufficient support for precise local modifications—this paper presents ChatHouseDiffusion. The proposed approach first utilizes a large language model to parse natural language instructions into structured JSON prompts, then applies Graphormer to capture the topological relationships between rooms, and finally adopts a diffusion model to generate floor plans under room boundary constraints. The results indicate that during the editing phase, the cross-attention maps replacement mechanism enables localized modifications without reconstructing the entire layout. Experiments conducted on the RPLAN dataset demonstrate that ChatHouseDiffusion outperforms existing models in both Micro-IoU and Macro-IoU, especially when accurate input prompts are used, achieving results highly consistent with ground truth, which exhibits strong practical utility and generalization performance. The model not only strictly follows user requirements but also enables intuitive and iterative design through interactive operations, providing a novel pathway for intelligent floor plan design. Based on this method, we further developed a visual platform enabling the drawing of outlines, inputting text prompts, and generating and editing floor plans, enhancing the practicality and usability of floor plan design.

Keywords: floor plan generation; floor plan editing; interaction design; diffusion model; large language model; Graphormer

自动化建筑平面图生成技术对于提升设计效率、降低设计成本具有重要意义。在实际设计过程中, 建筑师常常需要依据变化频繁的设计意图, 快速

生成方案以供评估与修改。传统设计方法依赖于 CAD 工具, 需要人工绘制, 难以满足高频迭代和高效协作的现代设计需求。为此, 如何借助人工智能

收稿日期: 2025-05-29; 录用日期: 2025-07-21; 网络首发日期: 2025-09-23

网络首发地址: <https://link.cnki.net/urlid/23.1235.T.20250923.0932.002>

基金项目: 北京市科学技术委员会 (Z231100005923043)

作者简介: 覃思中 (2001—), 男, 博士研究生; 陆新征 (1978—), 男, 教授, 博士生导师

通信作者: 陆新征, luxz@tsinghua.edu.cn

(artificial intelligence, AI) 技术实现“以人为本、可交互、高灵活性”的建筑平面图生成与编辑,已成为建筑智能设计的重要研究方向。

在深度学习算法出现之前,Merrell 等^[1]通过优化算法实现了房间布局的自动生成;Regateiro 等^[2]则将块代数与约束满足方法相结合,用于自动化建筑平面图设计。近年来,随着生成式 AI 技术的发展,建筑平面图生成领域涌现出大量具有突破性的研究。House-GAN^[3]和 House-GAN++^[4]利用生成对抗网络(generative adversarial networks, GAN)实现房间布局和门的位置生成。Rahbar 等^[5]提出两阶段生成方法,先构造气泡图,再使用条件生成对抗网络生成建筑平面图。HouseDiffusion^[6]结合离散与连续去噪过程的扩散模型,可以在连续坐标系中精准控制房间与门的布局生成。

上述方法普遍依赖房间拓扑或气泡图作为输入^[7],在此基础上,iPLAN^[8]和 MaskPLAN^[9]进一步采用了交互式设计方法,通过添加、编辑、删除、冻结等操作实现房间布局的修改。但它们的交互主要依赖于图形界面操作,设计师需要有明确清晰的思路,手绘绘制或选择房间关系、设定属性掩码,才能传递设计意图。图形界面操作难以处理普通用户高层次的语义需求,如“客厅采光好”、“书房尽量安静”等,而自然语言则可以更好地传递复杂的语义信息。随着大语言模型等技术的发展,基于文本的交互式设计展现出广泛的应用前景。Tell2Design^[10]利用深度学习序列生成模型,根据自然语言指令直接生成房间布局坐标。ChatDesign^[11]进一步引入预训练大语言模型,并支持迭代优化过程。Wei 等^[12]利用大语言模型将文本转换为图谱,结合图谱和文本采用序列生成方法得到房间坐标信息。HouseLLM^[13]利用大语言模型基于链式思维提示生成初步布局草图,再通过条件扩散模型对坐标进行细化设计。Zeng 等^[14-15]则通过大语言模型以及多智能体等方式实现平面布局的生成与编辑。但这些方法大多受限于大语言模型的图像理解能力,且难以保留前期生成的建筑平面图信息,从而影响房间的精准生成和局部编辑。

受限于编辑任务专用数据集的稀缺性,大量研究开始探索基于扩散模型的免训练编辑方法。Huang 等^[16]在其综述中总结了 5 类编辑策略,其中修改注意力机制因其无需训练、实现简单,成为目前最常见、最直接的编辑方式。Prompt2Prompt^[17]是该方向的代表性工作之一,通过直接修改、替换扩散过程中的注意力图,引导模型进行目标图像的编辑。TF-ICON^[18]与 Object-Shape Variation^[19]则通过微调

局部注意力图,在文本生成图像的流程中实现更精细的图像合成与形状变化控制。

针对上述方法交互能力不足且难以实现局部精确编辑的问题,本文提出 ChatHouseDiffusion 模型,融合大语言模型^[20]的交互能力、图结构表示的拓扑关系提取能力,以及扩散模型的生成与编辑功能。利用大语言模型对用户输入进行语义解析,采用 Graphormer^[21]来编码房间的拓扑关系,并引入扩散模型进行建筑平面图预测。此外,通过替换注意力图,实现建筑平面图设计中的局部精准编辑。最终,本文构建了一个更符合实际设计流程的自动化建筑平面图设计系统:用户可在模糊构想阶段通过输入部分信息快速获得初步方案,并在后续多轮交互中不断调整优化,最终生成满意的设计成果(图 1)。

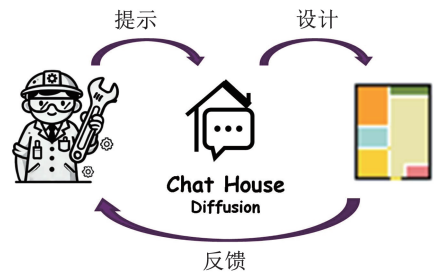


图 1 建筑平面图的迭代生成和编辑

Fig. 1 Iterative generation and editing of floor plans

1 模型构建

在建筑平面图设计任务中,去噪扩散概率模型(denoising diffusion probabilistic models, DDPM)的训练过程可以表示为如下式的优化问题:

$$\min_{\theta} \sum_{i=1}^N E_{T^{(i)}, I^{(i)}} [-\log P(P^{(i)} | T^{(i)}, I^{(i)}; \theta)] \Leftrightarrow \min_{\theta} E_{T^{(i)}, I^{(i)}, \epsilon_0^{(i)}} E_{\epsilon, t \sim [1, T]} [\|\epsilon - \epsilon_{\theta}(x_t, t, T^{(i)}, I^{(i)})\|^2] \quad (1)$$

式中: θ 为 DDPM 模型的参数, N 为数据集 D 中的样本数量, $P^{(i)}$ 为第 i 个样本对应的真实建筑平面图, $P(P^{(i)} | T^{(i)}, I^{(i)}; \theta)$ 为在给定文本提示 $T^{(i)}$ 、条件图像 $I^{(i)}$ 以及模型参数 θ 的条件下,生成真实建筑平面图 $P^{(i)}$ 的概率。该优化目标旨在最小化整个数据集上基于提供的文本提示与图像条件生成真实建筑平面的负对数似然函数,即提升模型在给定条件下生成真实样本的概率。该任务可等效为扩散过程中的噪声 ϵ 预测问题,其中: t 为当前扩散步骤, T 为扩散总步数, x_t 为第 t 步的带噪图像, $\epsilon_{\theta}(x_t, t, T^{(i)}, I^{(i)})$ 为模型对噪声的预测结果。

通过上述建模方式,建筑平面图生成问题被重新表述为一个噪声预测网络的训练任务。其核心挑战在于如何将文本描述与条件图像转化为可比较的表示形式,从而实现完整的语义理解与精确的跨模

态对齐。为此,本文引入了 Graphormer 模块与交叉注意力机制,以提升最终建筑平面图预测的准确率。

2 建筑平面图自动生成与编辑方法

本文旨在通过文本提示与房间轮廓信息,实现建筑平面图的自动生成与编辑,从而辅助设计人员进行房间布局规划。由于扩散模型具备生成准确、多样方案的能力,且能有效避免模式坍塌,保证生成结果的稳定性与精确性,因此将其作为主要框架。通过将文本提示与结构化的多模态输入相结合,可

以使模型在理解用户意图与满足具体设计需求方面表现出更强的能力,进一步提高在生成过程中灵活调控与交互控制的潜力。

2.1 整体框架

本文提出的模型框架如图 2 所示,可分为两个主要部分:1)基于大语言模型对自然语言房间设计条件的理解与解析;2)基于扩散模型进行建筑平面图的条件生成与编辑。此外,在模型训练阶段,为确保训练数据的准确性,以规则驱动的手动解析方式构建了高质量的训练样本。

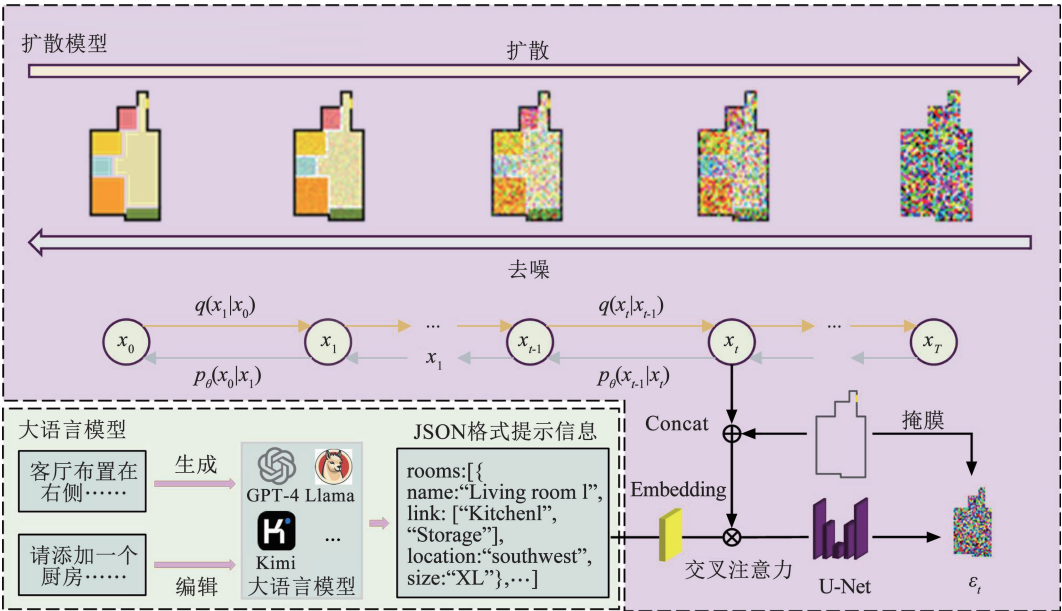


图 2 基于扩散模型的文本提示驱动的房间布局生成框架

Fig. 2 Framework of text prompt floor plan generation based on diffusion models

在大语言模型的解析流程中,当用户输入自然语言提示后,模型首先对文本进行语义分析,输出 JSON 格式的结构化提示信息,作为扩散模型的生成条件。

为了实现扩散与去噪过程,本文基于文本提示与房间轮廓构建了一个用于噪声预测的神经网络。首先,将房间轮廓作为掩膜,对当前扩散步骤 t 的图像 x_t 进行处理;其次,将房间轮廓与 x_t 进行拼接,并结合文本嵌入信息,作为交叉注意力的输入特征。最后,使用 U-Net 进行噪声预测,并在输出阶段再次应用房间轮廓掩码。该方法能够在建筑平面图的生

成与编辑过程中充分利用房间轮廓信息,确保扩散模型专注于轮廓区域内的图像生成,从而有效避免不相关区域的干扰。同时,为提高生成和编辑的效率,本文引入了去噪扩散隐式模型(denoising diffusion implicit models, DDIM)^[22]进行加速采样,在采样步数设置为 50 步时能取得较好的结果,单次设计时间可控制在 20 s 以内。

2.2 设计意图解析

图 3 展示了将用户自然语言描述转换为结构化 JSON 格式的完整流程。该过程开始于用户输入,其内容通常为关于房间位置及房间关系的文本描述。

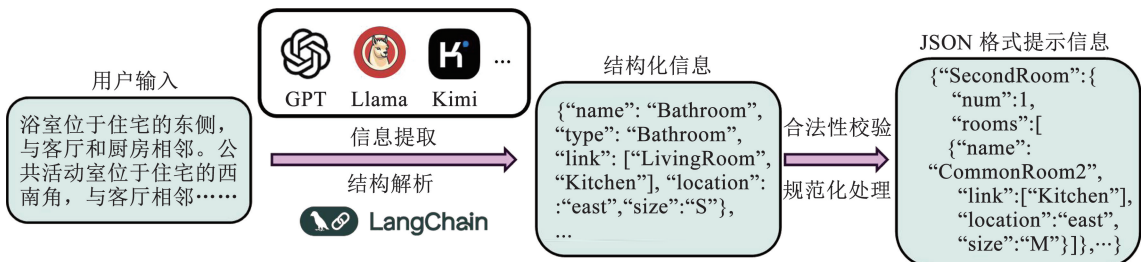


图 3 从自然语言到 JSON 的转换流程

Fig. 3 Workflow of prompt to JSON

首先通过大语言模型提取关键信息。本文基于 LangChain 框架,设计了一套结构化少样本提示模板,指导大语言模型从非结构化的自然语言中精准提取房间名称、类型、尺寸、位置及连通关系等属性,转化为可解析的 JSON 结构化数据。

在信息提取完成后,系统基于 JSON 模板对大语言模型的输出结果进行解析与格式化,最终组织为结构化的字典列表。之后进行枚举类型校验,用于验证 JSON 属性值的合法性,确保所有结构化数据均符合预设的数据类型与约束规则。同时,还引入基于莱文斯坦距离的模糊字符串匹配技术,对 JSON 属性值进行自动纠错与标准化处理。经过上述步骤处理与验证后,最终生成合规的 JSON 格式提示信息,作为后续建筑平面图生成的输入条件。

2.3 建筑平面图生成

基于扩散模型的建筑平面图生成方法主要包括两个核心模块。第 1 个模块负责对文本输入进行编码,提取并保留诸如房间属性、拓扑关系等关键信息;第 2 个模块则利用文本嵌入与房间轮廓作为输入,训练神经网络以进行噪声预测。

2.3.1 文本嵌入

在获得 JSON 格式的文本提示信息后,需要对房间的属性与拓扑关系进行更合理的特征表示。以往的文本生成图像方法^[23]通常采用如 T5 等^[24]模型对文本直接编码,虽然适用于一般自然语言,但难以准确表达不同房间之间的拓扑结构。

房间拓扑结构通常采用图神经网络来表征和计算,然而常见的图神经网络往往只关注局部拓扑信息,难以理解房间的整体布局。Graphormer^[21]通过 Transformer 的自注意力机制计算图中各节点间的语义相似性,建立个体单元与整体结构之间的联系,同时采用 3 种简单而高效的空间编码方法来利用图结构信息,不仅可以理解图的整体布局,还能准确把握节点之间的关系。

因此,本文提出了一种结合 T5^[24]与 Graphormer^[21]的特征表示方法,见图 4。首先,从结构化房间数据中提取每个房间的类型、位置与尺寸信息;其次,使用预训练的 T5 编码器对这些属性进行编码,并将其拼接形成表示房间的初始特征向量;最后,利用 Graphormer 方法进一步嵌入房间之间的拓扑关系。

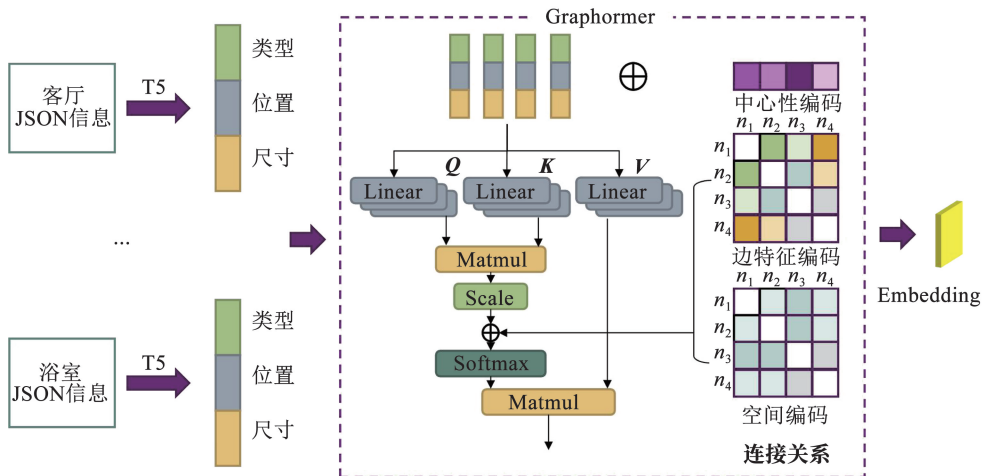


图 4 融合房间属性与拓扑关系的建筑平面图信息编码方法

Fig. 4 Floor plan information encoding method combining attribute features and topological relationships

图 4 中,Graphormer 模型在注意力机制中引入房间之间的连接关系,帮助模型学习拓扑结构信息。Graphormer 通过构建房间的无向图来显式引入拓扑结构信息,节点表示各个房间,边表示空间或功能上的相邻关系,边是双向的。具体而言,采用了以下 3 种编码方式将图特征融入注意力偏置中进行学习:1)中心性编码。基于房间的度数衡量其在布局中的重要性,并将该信息加入输入特征中。由于房间布局为无向图,入度与出度相同。2)空间编码。利用无权最短路径长度衡量节点之间的相对关系强度。3)边编码。捕捉房间连接之间的边信息,初始时每条边赋值为 1,表示基础的拓扑连接关系。最

终输出的特征向量将作为扩散模型的文本嵌入表示。

2.3.2 噪声预测

借鉴 Imagen^[23]的设计思想,构建增强型 U-Net 噪声预测网络,并引入多种条件输入,其模型结构见图 5。

该模型在初始卷积模块中采用多种卷积核,以对输入特征进行初始化处理,从而获得多尺度的空间表示能力。在下采样路径中,包含多个子模块,每个子模块由 1 个交叉注意力残差块、3 个标准残差块、1 个自注意力层,以及 1 个空间下采样卷积层组成。需要注意的是,首个下采样模块中不包含自注

注意力机制。每个下采样阶段不仅将特征映射至更高维度的表示空间,同时还保留中间特征,作为上采样阶段的跳跃连接。在网络的中间层,交叉注意力残差块被分别置于起始与末端,自注意力层位于其中,以进一步增强模型的特征建模能力。上采样路径则

整合了下采样过程中保留的跳跃连接特征,其结构与下采样路径对称,逐步恢复图像分辨率。最后,模型在上采样阶段通过融合模块整合多尺度信息,并通过最终的残差层与卷积层输出单通道的噪声预测结果。

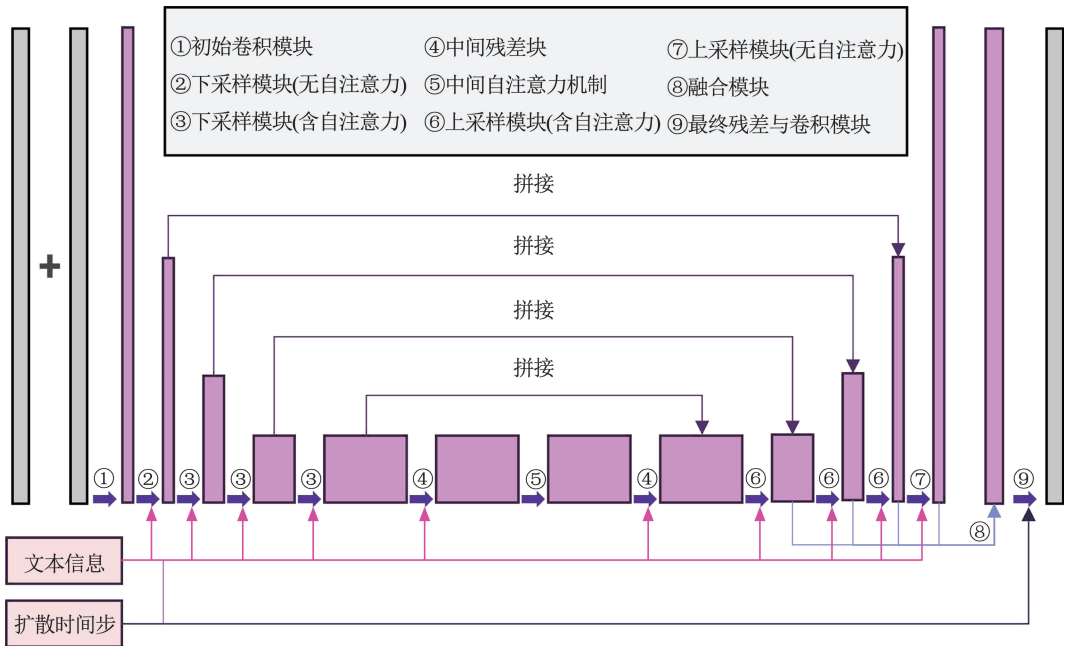


图 5 增强型 U-Net 网络结构

Fig. 5 Architecture of the enhanced U-Net network

这一增强型 U-Net 架构通过引入多层次的交叉注意力与自注意力机制,显著提升了图像生成与重建任务的性能,在处理复杂的跨模态信息输入时表现出良好的鲁棒性与准确率。

2.4 建筑平面图编辑

在初始建筑平面图生成完成后,用户可能会对某些局部区域不满意,但又不希望重新生成整个布局。为此提出一种建筑平面图的局部编辑方法,见图 6。

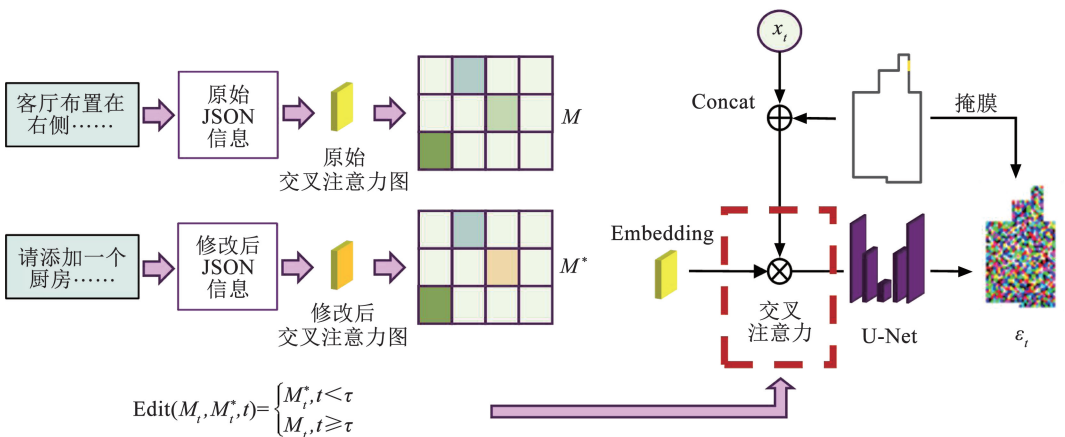


图 6 基于交叉注意力图的建筑平面图局部编辑方法

Fig. 6 Floor plan editing method based on cross-attention map

该方法受到 Prompt-to-Prompt^[17] 方法的启发,通过在去噪过程中替换交叉注意力图,实现仅对局部区域进行修改,而保留整体布局结构不变。

具体而言,在编辑过程中,首先固定随机种子,并保存初始生成过程中各个扩散步骤的交叉注意力图。随后,使用相同的随机种子,通过大语言模型与

Graphormer 获取新的文本嵌入表示。在初始去噪阶段,使用原始的交叉注意力图替换修改后的交叉注意力图,从而实现生成过程的局部控制。此外,编辑操作的强度可以通过调整阈值 τ 来控制,以实现更为精准可控的建筑平面图编辑。图 7 展示了一个具体的编辑案例。

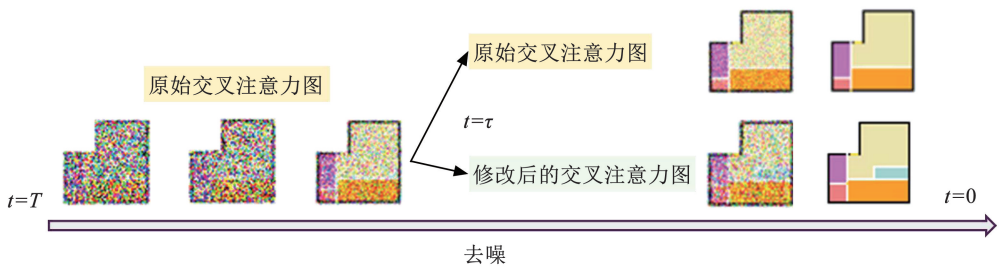


图 7 替换注意力图实现建筑平面图局部编辑的案例

Fig. 7 Substitution of attention maps to enable localized editing of floor plans

3 实验与讨论

本文的实验在如下计算平台上进行:操作系统为 Ubuntu 22.04 LTS;处理器为 Intel Xeon E5-2682 v4@ 64 × 3 GHz;内存为 32 GB;显卡为 NVIDIA GeForce RTX 3090,显存为 24 GB。

3.1 数据集与评价指标

在训练阶段,本文采用了 RPLAN 数据集^[25],该数据集是从真实住宅建筑中人工收集的大规模建筑平面图数据集。由于 Tell2Design^[10]构建了文本驱动建筑平面图生成数据集,本文直接采用该数据集和实验结果作为基准对照,包括训练集数据 78 480 个,测试集数据 2 308 个。

房屋的相关信息,包括房间之间的连接关系、每个房间的尺寸以及位置,均通过向量表示方式进行解析。最终,生成描述房屋信息的 JSON 文件作为模型的另一个输入。

为评估生成的建筑平面图与真实标注之间的一致性,本文采用了 Leng 等^[10]提出的两项指标:微平均交并比(Micro IoU)与宏平均交并比(Macro IoU)。Micro IoU 通过汇总所有房间类型的交集与并集,计算整体层面的 IoU。Macro IoU 则对各类房间分别计算 IoU,再取平均,用于反映模型对不同房间类型的预测准确率。其计算公式如下:

$$\text{Micro-IoU} = \frac{\sum_{r=1}^R I_r}{\sum_{r=1}^R U_r} \quad (2)$$

$$\text{Macro-IoU} = \frac{1}{R} \sum_{r=1}^R \frac{I_r}{U_r} \quad (3)$$

式中: I_r 、 U_r 分别为第 r 类房间在真实标注与预测图像中的交集面积与并集面积, R 为房间类型的总数。

3.2 建筑平面图生成结果讨论

在本文中,文本提示条件最初通过人工标注方式获得,即直接从建筑平面图中提取数据并转换为

JSON 格式,通常被视为真实标签(ground truth)。相比于依赖大语言模型从文本中提取数据的方式,该方法在准确性上显著更高,因此被用作后续实验的基准参照。

将本方法与 Tell2Design 等^[10-11]现有方法进行了比较,同时评估不同大语言模型对设计结果的影响。IoU 指标结果见表 1。图 8 展示了多种方法生成的建筑平面图示例,包括人工标注版本和基于 DeepSeek-V3 模型生成的版本。

表 1 不同方法的 IoU 指标对比

Tab. 1 IoU scores using different methods

方法	Micro-IoU/%	Macro-IoU/%
Obj-GAN	10.68	8.44
CogView	13.30	11.43
Imagen	12.17	14.96
T2D (w/o boundary)	35.95	29.95
T2D	54.34	53.30
ChatDesign	58.31	55.43
ChatDesign-iterative	54.57	57.24
ChatHouseDiffusion (Llama3)	54.64	51.22
ChatHouseDiffusion (GPT-4-turbo)	60.51	56.27
ChatHouseDiffusion (moonshot-v1-8k)	60.57	57.34
ChatHouseDiffusion (DeepSeek-V3)	63.60	59.32
ChatHouseDiffusion (ma)	85.04	82.32

针对多种方法的生成结果,使用如下的统一文本提示进行比较:住宅的北侧若无阳台将显得不完整。该阳台面积约为 1.49 m²,可通过客厅或其旁的公共活动室进入。浴室 1 位于住宅东部,靠近客厅,面积约为 1.39 m²。浴室 2 面积更大,约为 2.79 m²,位于主卧与第 2 公共活动区之间,沿住宅西侧布局。公共活动室 1 位于住宅的东北角,面积约为 7.43 m²,位置便利,紧邻阳台。公共活动室 2 位于西北角,面积接近 9.29 m²,可通过旁边的厨房或客厅的公共通道进入,通行便利。厨房位于住宅北侧,位于客厅与第 2 公共活动区之间,面积约为 4.65 m²。客厅位

于住宅的东南角,面积约为 23.23 m²,几乎可通往住宅中的所有房间,布局合理、连接性强。主卧室位

于住宅的西南角,面积约为 11.15 m²,紧邻客厅。相关建筑平面图见图 9。

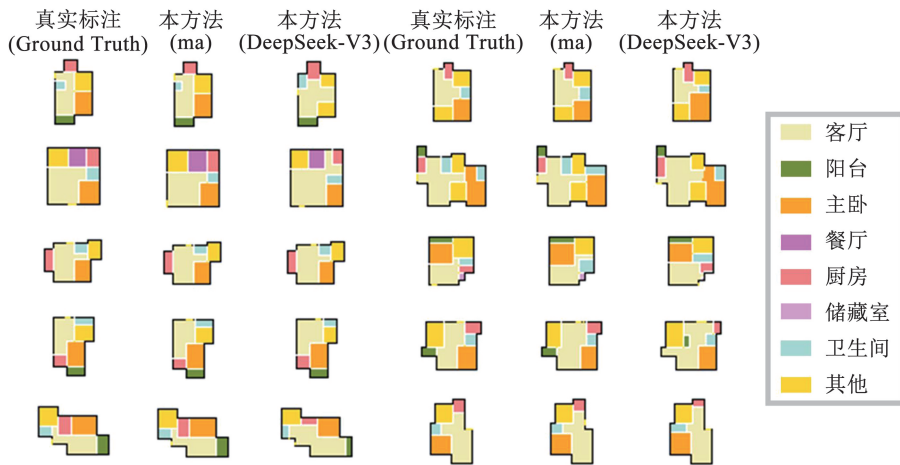


图 8 ChatHouseDiffusion 生成的房间布局示例

Fig. 8 Cases of floor plans generated by ChatHouseDiffusion

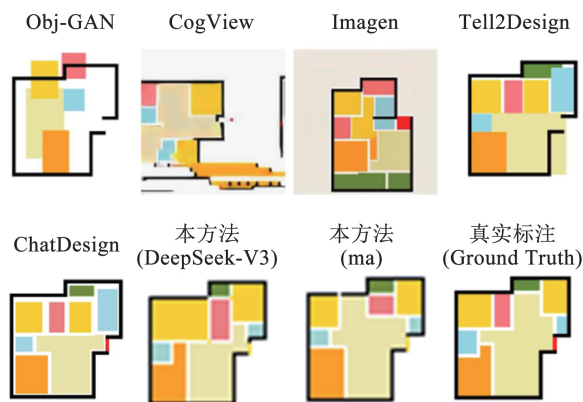


图 9 不同方法生成的房间布局对比

Fig. 9 Comparison of generated floor plans from different methods

与现有方法相比,本文方法在 IoU 指标上取得了显著提升。特别是 ChatHouseDiffusion (人工标注)所达到的得分远高于其他方法,充分展现了该方法的潜力。结果表明,只要大语言模型能够准确解析文本,就可以获得更优的生成结果。在所有参与实验的大语言模型中,ChatHouseDiffusion (DeepSeek-V3) 的表现最佳,超过了其他现有方法,进一步验证了本方法的有效性。从图 8 可以看出,本方法生成的建筑平面图整体与真实标注一致,虽然在局部存在些许差异,但可通过后续编辑进一步调整。图 9 中的对比也表明,本方法生成的建筑平面图在满足外部轮廓要求的同时,其内部房间布局更接近真实结构,也更为合理。

在生成阶段,本方法充分利用了文本描述与房间轮廓信息。ChatHouseDiffusion (人工标注) 基于完全准确地描述信息, Micro-IoU 与 Macro-IoU 均超过 80%, 在当前输入信息有限的条件下,进一步提升较为困难。具体案例分析显示,生成结果与真实标注

的差异主要体现在房间形状和相对位置,这表明若能在模型中引入这些特征,有望实现更加精确的建筑平面图生成。

在多个大语言模型的对比中,DeepSeek-V3 表现最好,moonshot-v1-8k 与 GPT-4-turbo 的表现非常接近,而 Llama3 的效果明显较差。总体来看,DeepSeek-V3 基本达到了当前大语言模型在结构化解析方面的性能上限,但与真实标注之间仍存在差距。造成这一现象的原因主要有两点:1) Tell2Design 数据集中存在部分不精确的文本描述,未能准确反映实际建筑平面图结构;2) 文本描述对 IoU 影响较大,例如在处理面积较小的房间时,即使位置非常接近,IoU 也可能为零,这种情况会严重影响最终得分。因此,在实际应用中,使用更准确地描述或通过后期编辑优化不清晰的文本输入,对于提升生成结果的准确率具有重要意义。

3.3 建筑平面图编辑结果讨论

为了验证建筑平面图编辑的能力,本文通过案例分析展示了房间新增、修改与删除等操作的效果。如图 10 所示,本方法能够准确实现局部房间的编辑,且在编辑过程中不会破坏整体的建筑平面图结构。

在编辑阶段,用户可以有效地执行房间的添加、删除和修改操作。在某些情况下,仍可能出现其他房间被误修改的问题,但可通过调整阈值 τ 进行控制与优化。然而,由于语言表达的模糊性,当前方法尚难以实现诸如精确设定房间尺寸等细粒度调整。未来的改进方向可考虑引入图形化交互界面,结合拖拽操作,实现更灵活的布局调整,从而提升建筑平面图编辑的交互性与实用性。

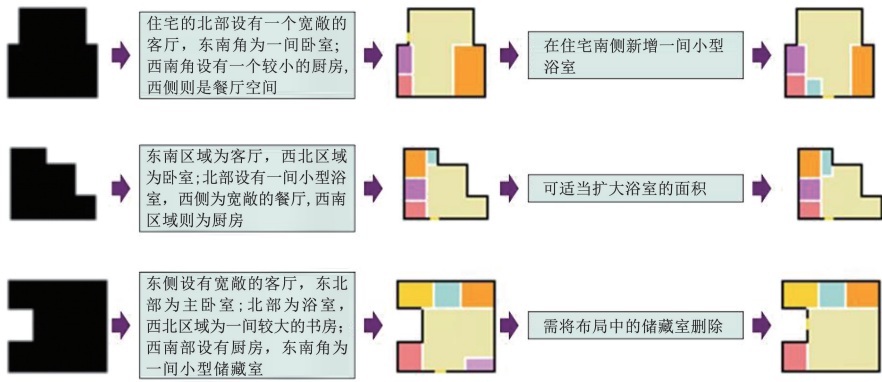


图 10 建筑平面图生成与编辑案例

Fig. 10 Cases of floor plan generation and editing

4 方法与应用

本文基于 Python 的 Tkinter 库实现了用户界面设计,支持绘制房间轮廓、输入文本提示以及生成与编辑建筑平面图等操作,整体流程见图 11。用户先

在“Line”模式下手绘房屋外轮廓,在文本框中输入建筑平面布置需求,点击“Generate”自动生成房间布局后,如需微调可进一步输入编辑需求,点击“Edit”进行布局编辑,在得到合适结果后点击“Save”导出最终平面图。

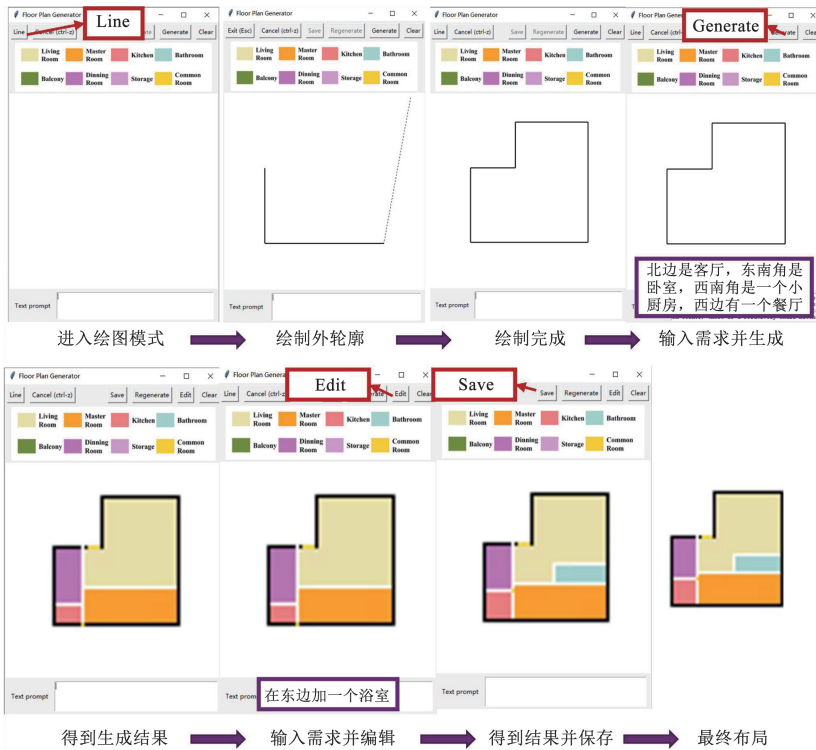


图 11 ChatHouseDiffusion 的用户界面设计

Fig. 11 UI design of ChatHouseDiffusion

为提升用户操作体验,系统中集成了多项实用功能。在绘图阶段,系统支持绘制虚线辅助线、正交捕捉、端点捕捉,以及撤销上一步操作等辅助功能;在生成与编辑阶段,系统支持记忆上一次输入文本与通过修改随机种子进行图像重生成等操作。

该方法可实现高效的房间概念设计,显著增强了建筑平面图设计的交互性。用户可参考 GitHub 项目主页提供的说明 (<https://github.com/>

ChatHouseDiffusion/chathousediffusion),配置 API 密钥并实现建筑平面图的生成与编辑。

5 结论

本文提出了 ChatHouseDiffusion,一种面向建筑平面图自动生成与编辑的新型方法和软件系统。该方法基于文本提示驱动,融合大语言模型、Graphormer 与扩散模型完成建筑平面图生成,利用

交叉注意力机制实现平面图精准编辑,显著提升了建筑设计过程的智能化水平。本文的主要结论如下:

1) 本文构建了一套建筑平面图概念设计框架,支持基于自然语言的交互式生成与编辑,具备良好的交互性。

2) 在建筑平面图生成方面,有效结合大语言模型的语义理解、Graphormer 的拓扑关系建模以及扩散模型的空间位置理解能力,所生成结果能较好地满足用户设定的设计条件,在 IoU 指标上明显优于现有方法。

3) 编辑模块基于交叉注意力图进行局部调整,能够在不重新设计整体布局的前提下,灵活实现指定区域的精准编辑,具备较强的实用性与工程适应能力。

4) ChatHouseDiffusion 拓展了自动化建筑平面图设计的界限,为更直观、高效的建筑设计流程提供了可能。然而,该方法在文本理解能力上仍存在一定局限,可能导致生成结果与用户意图存在偏差。未来的研究可进一步增强模型的多模态理解能力,构建更具通用性和可扩展性的交互式设计平台。

参考文献

[1] MERRELL P, SCHKUFZA E, KOLTUN V. Computer-generated residential building layouts [J]. *ACM Transactions on Graphics*, 2010, 29(6): 1. DOI:10.1145/1882261.1866203

[2] REGATEIRO F, BENTO J, DIAS J. Floor plan design using block algebra and constraint satisfaction [J]. *Advanced Engineering Informatics*, 2012, 26(2): 361. DOI:10.1016/j.aei.2012.01.002

[3] NAUATA N, CHANG K H, CHENG C Y, et al. House-GAN: relational generative adversarial networks for graph-constrained house layout generation [C]//*Computer Vision-ECCV 2020*. Cham: Springer, 2020: 162. DOI:10.1007/978-3-030-58452-8_10

[4] NAUATA N, HOSSEINI S, CHANG K H, et al. House-GAN++: generative adversarial layout refinement network towards intelligent computational agent for professional architects [C]//*2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville: IEEE, 2021: 13627. DOI:10.1109/CVPR46437.2021.01342

[5] RAHBAR M, MAHDAVINEJAD M, MARKAZI A H D, et al. Architectural layout design through deep learning and agent-based modeling: a hybrid approach [J]. *Journal of Building Engineering*, 2022, 47: 103822. DOI:10.1016/j.jobee.2021.103822

[6] SHABANI M A, HOSSEINI S, FURUKAWA Y. HouseDiffusion: vector floorplan generation via a diffusion model with discrete and continuous denoising [C]//*2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver: IEEE, 2023: 5466. DOI:10.1109/CVPR52729.2023.00529

[7] WEBER R E, MUELLER C, REINHART C. Automated floorplan generation in architectural design: a review of methods and applications [J]. *Automation in Construction*, 2022, 140: 104385. DOI:10.1016/j.autcon.2022.104385

[8] HE Feixiang, HUANG Yanlong, WANG He. iPLAN: interactive and procedural layout planning [C]//*2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans: IEEE, 2022: 7783. DOI:10.1109/CVPR52688.2022.00764

[9] ZHANG Hang, SAVOV A, DILLENBURGER B. MaskPLAN: masked generative layout planning from partial input [C]//*2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*

(CVPR). Seattle: IEEE, 2024: 8964. DOI:10.1109/CVPR52733.2024.00856

[10] LENG Sicong, ZHOU Yang, DUPTY M H, et al. Tell2Design: a dataset for language-guided floor plan generation [C]//*Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Stroudsburg: ACL, 2023: 14680. DOI:10.18653/v1/2023.acl-long.820

[11] LI Jinmin, LUO Yilu, LU Shuai, et al. ChatDesign: bootstrapping generative floor plan design with pre-trained large language models [J]. *Proceedings of the 29th Conference on Computer Aided Architectural Design Research in Asia (CAADRIA) [Volume 1]*, 2024, 1: 99. DOI:10.52842/conf.caadria.2024.1.099

[12] WEI Yinyi, LI Xiaoh. Graph-augmented text-based floorplan generation [C]//*2024 International Conference on Automation in Manufacturing, Transportation and Logistics (ICaMaL)*. Hong Kong: IEEE, 2025: 1. DOI:10.1109/ICaMaL62577.2024.10919609

[13] ZONG Ziyang, ZHAN Zhaohuan, TAN Guang. HouseLLM: LLM-assisted two-phase text-to-floorplan generation [EB/OL]. 2024: arXiv: 2411.12279. <https://doi.org/10.48550/arXiv.2411.12279>

[14] ZENG Pengyu, JIANG Maowei, WANG Zihang, et al. CARD: cross-modal agent framework for generative and editable residential design [C]//*38th Conference on Neural Information Processing Systems*. Amherst: *NeurIPS 2024 Workshop on Open-World Agents*, 2024

[15] ZENG Pengyu, GAO Wen, LI Jizhizi, et al. Automated residential layout generation and editing using natural language and images [J]. *Automation in Construction*, 2025, 174: 106133. DOI:10.1016/j.autcon.2025.106133

[16] HUANG Yi, HUANG Jiancheng, LIU Yifan, et al. Diffusion model-based image editing: a survey [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025, 47(6): 4409. DOI:10.1109/TPAMI.2025.3541625

[17] HERTZ A, MOKADY R, TENENBAUM J, et al. Prompt-to-prompt image editing with cross-attention control [EB/OL]. 2022: arXiv: 2208.01626. <https://arxiv.org/abs/2208.01626>

[18] LU Shilin, LIU Yanzhu, KONG A W K. TF-ICON: diffusion-based training-free cross-domain image composition [C]//*2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. Paris: IEEE, 2024: 2294. DOI:10.1109/ICCV51070.2023.00218

[19] PATASHNIK O, GARIBI D, AZURI I, et al. Localizing object-level shape variations with text-to-image diffusion models [C]//*2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. Paris: IEEE, 2024: 22994. DOI:10.1109/ICCV51070.2023.02107

[20] COLABIANCHI S, COSTANTINO F, SABETTA N. Assessment of a large language model based digital intelligent assistant in assembly manufacturing [J]. *Computers in Industry*, 2024, 162: 104129. DOI:10.1016/j.compind.2024.104129

[21] YING Chengxuan, CAI Tianle, LUO Shengjie, et al. Do transformers really perform badly for graph representation? [C]//*Thirty-Fifth Annual Conference on Neural Information Processing Systems*. San Diego: *NeurIPS*, 2021, 34: 28877

[22] SONG Jiaming, MENG Chenlin, ERMON S. Denoising diffusion implicit models [EB/OL]. 2020: arXiv: 2010.02502. <https://arxiv.org/abs/2010.02502>

[23] SAHARIA C, CHAN W, SAXENA S, et al. Photorealistic text-to-image diffusion models with deep language understanding [EB/OL]. 2022: arXiv: 2205.11487. <https://arxiv.org/abs/2205.11487>

[24] RAFFEL C, SHAZEER N, ROBERTS A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer [J]. *Journal of Machine Learning Research*, 2020, 21(1): 5485. DOI:10.5555/3455716.3455856

[25] WU Wenming, FU Xiaoming, TANG Rui, et al. Data-driven interior plan generation for residential buildings [J]. *ACM Transactions on Graphics*, 2019, 38(6): 1. DOI:10.1145/3355089.3356556