arXiv:2603.02662v1 [cs.HC] 3 Mar 2026

面向人类使用的三维布局行为感知人体测量场景生成

Semin Jin*设计信息学实验室
汉阳大学 韩国大韩民国首尔 以人为中心的人工智能设计研究所 汉阳大学 韩国大韩民国首尔
tpals97@gmail.com

Donghyuk Kim*
设计信息学实验室
汉阳大学
韩国大韩民国首尔
以人为中心的人工智能设计研究所
汉阳大学
首尔，大韩民国
oververitas@gmail.com

Jeongmin Ryu
设计信息学实验室
汉阳大学
首尔，大韩民国
2002rjm@gmail.com

Kyung Hoon Hyun†
设计信息学实验室
汉阳大学
韩国首尔
以人为中心的人工智能设计研究所
汉阳大学
韩国首尔
hoonhello@gmail.com

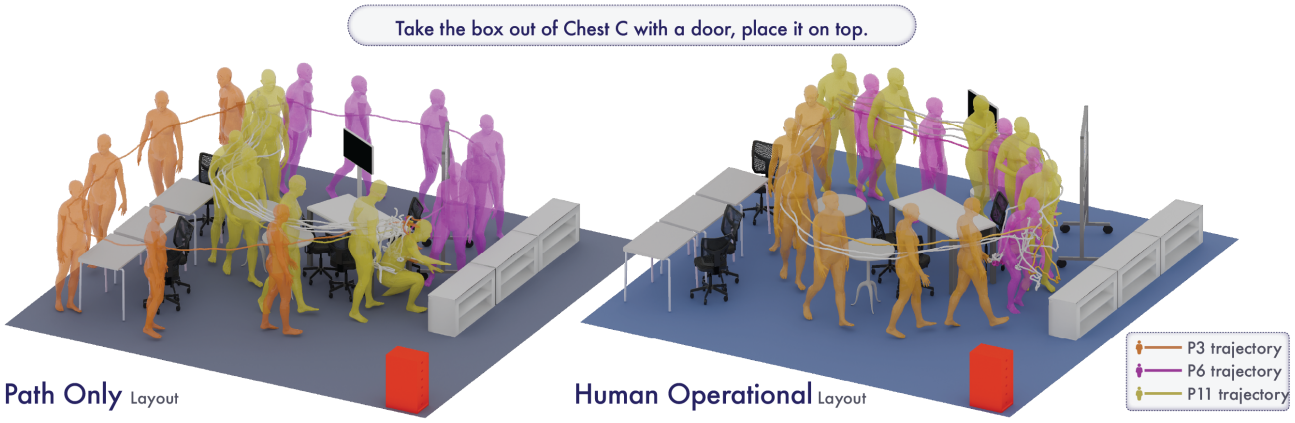


图 1：可视化　生成的仅通道布局与人工操作布局中的运动轨迹可视化。

摘要

设计良好的室内场景应优先考虑人们如何在空间中活动，而不只是摆放哪些物体。然而，现有的3D场景生成方法侧重于视觉和语义合理性，却未能充分考虑人们能否舒适地行走、坐下或操作物体。为弥合这一差距

针对这一空白，我们提出了一个行为感知的人体测量场景生成框架。我们的方法利用视觉语言模型（VLM）分析物体—行为关系，并将空间需求转化为适配用户个体人体测量数据的参数化布局约束。我们使用几何度量指标与一项用户感知研究（**N**=16），对该方法与最先进模型进行了比较研究。我们还开展了深入的真人尺度研究（个体，**N**=20；群体，**N**=18）。结果表明，该方法在任务完成时间、轨迹效率和人-物操作空间方面均有所改善。本研究提出了一个框架，将基于 VLM 的交互推理与人体测量约束相结合，并通过技术指标和真实尺度的人体可用性研究进行了验证。

*共同第一作者。
†通讯作者。

本作品依据知识共享署名-非商业性使用-禁止演绎 4.0 国际许可协议授权。
CHI'26，西班牙巴塞罗那© 2026年版权归所有者/作者所有。ACM ISBN 979-8-4007-2278-3/2026/04 https://doi.org/10.1145/3772318.3790341

CCS概念
• **以人为中心的计算** → **人机交互设计与评估方法**；• **计算方法** → **计算机视觉**。

关键词
适合人类使用的三维布局、室内场景生成、人体测量数据、视觉语言模型。

ACM引用格式：
Semin Jin、Donghyuk Kim、Jeongmin Ryu和Kyung Hoon Hyun。2026年。面向人类可用三维布局的行为感知人体测量场景生成。载于*2026*年*CHI*计算系统中的人为因素会议论文集（*CHI '26*），*2026*年*4*月*13*日至*17*日，西班牙巴塞罗那。ACM，美国纽约州纽约市，20 页。https://doi.org/10.1145/3772318.3790341

1 引言
物理环境从根本上塑造人体运动与行为 [25]。在现实世界中，布局很少是静态的；使用者会自然地调整周围环境——将椅子向后挪动以腾出伸腿空间，或将桌子拉近以便够到物体——从而适应其特定的身体尺寸与运动。因此，可用布局即是已针对其使用者的人体测量数据与行为完成优化的布局。三维场景生成的最新进展已能够自动创建合理的布局 [4, 20, 29, 35, 42]。然而，一个关键局限在于，这些模型依据数据集统计信息 [20, 29, 42] 或大语言模型（大语言模型）先验 [4, 35],生成布局，缺乏对人体尺寸或行为模式的任何显式推理。

考虑一个办公室场景：生成模型可能会根据训练数据集中发现的视觉对称性，将书桌背靠背地摆放。然而，由于模型忽略了实际将椅子向后拉并站起所需的人体测量间隙，生成的布局会立即形成一个冲突区域——这是一种功能性失效，而人类用户通常会本能地通过调整家具之间的距离来避免这种情况。在需要高保真具身交互的扩展现实（XR）应用中，这一问题变得更加关键，例如训练模拟、远程呈现，或用于人体工学评估的数字孪生。例如，用户可能会反复调整自身位置，以触及超出其运动范围的目标；或者出现视觉穿模，即虚拟手部穿透表面——这些问题的根源在于布局未针对用户特定的身体尺寸进行优化。尽管现有设计标准涵盖了抽屉间隙和通行宽度等功能性约束，但将其应用于布局生成仍颇具挑战：相关指南规定的是建议范围，无法脱离具体情境简化为固定值；而生成可用布局则需要综合考虑物体类型、尺寸、房间配置以及用户身体特征。我们的框架通过利用视觉语言模型（VLM）推断特定物体的空间约束，并将其与个性化人体测量数据相结合，从而应对这些挑战。

为解决这一局限，我们提出了**行为感知人体测量场景生成**，这是一种基于视觉语言模型（VLM）的框架，通过行为推理和人体测量依据增强基于文本的布局生成。我们利用 VLM 从视觉线索中推断物体的功能，并根据场景类型和布局标准推理潜在的人类交互。我们的框架以布局标准和资产为输入，分两个阶段进行。首先，构建行为感知关系表示，整合物体语义、人—物交互模式和群体层级空间关系。其次，通过从这些关系中推断基于人体测量依据的空间约束，并将其编码为可微惩罚项，以进行基于梯度的优化，从而实现基于约束的布局生成。通过使用个性化的人体测量特征对这些约束进行参数化，我们的方法可确保生成的场景具备功能连通性、充足的操作净空以及高效的通行路径，并针对不同用户进行定制。

我们通过技术验证和针对人体操作可用性的用户研究评估了我们的框架。首先，我们将其与最先进的基于大语言模型的场景生成方法 LayoutVLM [35],进行性能比较，以衡量物体碰撞、平面图违规情况以及专业人士对生成布局的看法（N = 16）。其次，我们在实体办公室和休息室环境中按**1:1**比例实现生成的布局，并比较三种条件：**LayoutVLM（基线）**、**仅通行（PO）**，它依据静态身体尺寸确保通道达到最低可通行宽度；以及**人体操作（HO）**，它依据与运动相关的人体测量包络，确保有足够的人体操作空间。对于 PO 和 HO 条件，我们根据每位参与者的 Skinned Multi-Person Linear（SMPL） [22] 模型实例化了参与者特定的人体测量档案，并据此为每位参与者的实际身体尺寸设置约束参数（例如通道宽度、伸手包络和视野要求）。我们针对人体操作可用性的用户研究分为两部分：第一部分是采用结构化物体–动作–目标任务 [1, 44] 的个体研究（N = 20），用于验证基于人体测量的布局支持能否提升任务绩效；第二部分是采用自然情境协作场景的群体研究（N = 18，六个由三人组成的团队），用于考察布局能否在共同执行任务时最大限度减少通行冲突 [23, 24]。我们的研究主要贡献如下：

- 我们提出了一种行为感知的人体测量场景生成框架，利用视觉语言模型推理，从人与物体的交互中推断空间约束，并以个体人体测量数据为依据。
- 我们将这些约束融入可微分的布局优化流程，通过对通道宽度、操作净空和交互空间占用进行操作化建模，生成适合人体操作的布局。
- 我们开展了一项评估，结合技术验证、专业人士感知研究，以及针对实尺度环境中人体操作可用性的用户研究。

2 相关工作
近期的三维场景生成方法能够合成合理的布局，但依赖通用的空间约束，忽略了个体的身体尺寸和行为模式。生成便于人类使用的布局需要遵循人体测量学设计原则，但现有评估很少考察生成的布局是否支持实际的人类操作。为弥补这些不足，我们回顾了三个领域的既有研究：基于大语言模型的场景生成、人体测量学设计

虚拟与物理空间中的人体测量学设计，以及三维布局的评估方法。

2.1 基于大语言模型的场景生成

近年来，三维室内场景生成领域的进展利用 Transformer架构 [29, 37, 42], 扩散模型 [20, 38, 48], 和大语言模型 [4, 35, 47] 生成语义上合理的布局。在这些方法中，LayoutVLM [35] 通过将布局创建建模为约束满足问题，代表了当前基于大语言模型的场景生成领域的最先进水平。它利用大语言模型将自然语言指令转化为空间约束（例如距离、朝向和对齐），然后根据视觉语言模型反馈，通过基于梯度的方法优化物体位置。这种方法在生成语义连贯且满足用户指定关系的布局方面表现出色。然而，LayoutVLM依赖大语言模型的常识来设定距离参数，因此生成的方案千篇一律，无法考虑个体的身体尺寸或行为需求。

另一条研究路线探索利用大语言模型在现有场景中进行人类行为仿真 [11, 30, 34, 45]。这些方法擅长在预定义的空间布局中生成人体动作并规划动作序列，但与我们的目标不同：它们使人类行为适应现有空间，而我们则生成空间以适应人体操作需求。我们的工作基于LayoutVLM的约束优化框架，并通过将人体测量数据直接整合到约束量化过程中对其进行扩展。我们并非依赖大语言模型常识中的通用距离，而是根据个体的身体测量数据和预期交互计算个体特定的操作需求，从而将场景生成从语义上合理推进到人体操作层面。

2.2 虚拟与实体场景中的人体测量驱动设计

建筑学与人因学的基础理论确立了这样一种观点：空间设计不仅关乎视觉，更从根本上以交互为导向。在建筑理论中，人体与空间之间的关系居于核心地位；操作空间由人类活动所需的动态范围界定 [6, 8, 28]。同样，人体工程学文献区分了结构人体测量学——即静态身体尺寸——与功能人体测量学；后者描述任务所需的动态活动范围和净空，并强调真正的可用性取决于能否满足后者的要求 [41]。与这些理论框架相呼应，人体工程学和XR领域的研究表明，人体测量因素对功能空间至关重要。个性化家具设计研究显示，纳入个人身体测量数据（如可触及范围和关节活动范围）能够直接提升任务绩效 [31]。XR环境也进一步采用了这一方法，让虚拟物体根据用户身体尺寸动态调整，以实现最佳交互 [5, 18, 19]。这类系统会根据手臂触及范围调整架子高度，按照坐姿眼高缩放工作空间，并将控制装置布置在便于操作的区域。然而，这些方法主要评估的是设计过程本身，即衡量生成的家具尺寸是否符合身体测量数据，而非评估人在由此形成的空间中的实际行为。因此，人体测量规范与现实使用之间的差距仍在很大程度上未知，且其对人-场景和人-物体交互造成的连锁影响，尚未在场景生成领域得到系统研究。

人体测量设计方法不仅涵盖单件家具，还扩展到操作空间以及人类有效使用物体所需的体积。既定指南 [10, 28] 定义了抽屉操作的净空、书桌周围的通行路径以及观看距离。然而，当前的场景生成方法将这些视为固定常数。我们的人体操作方法通过将人体测量数据转化为空间约束来弥合这一差距，从而确保每个人的身体尺寸都拥有充足的操作空间。从通用净空转向个性化操作体积，标志着生成场景从仅仅合理迈向真正可用的进步。考虑到人类多样性，这种个性化的重要性便显而易见；第5百分位女性和第95百分位男性需要截然不同的操作空间 [28]；然而，当前方法采用统一约束，对体型较大者可能不足，对体型较小者则可能过于宽敞而造成浪费。通过将人体测量数据直接整合到生成过程中，我们确保布局能够容纳将居住于这些空间的特定个体。

2.3 三维场景评估方法

现有三维场景生成研究主要依赖弗雷歇特征距离和核初始距离等视觉合理性指标 [20, 29, 38] ，衡量生成结果与训练数据的分布相似度；或通过用户评分 [4, 34, 35] 评估语义连贯性。尽管这些指标能够有效验证视觉质量和学习性能，却无法捕捉决定实际可用性的行为与操作层面特征。相比之下，建筑学和室内设计领域采用全面的使用后评估，考察认知舒适度、寻路效率、空间满意度和长期行为适应 [15, 16, 27]；然而，这类评估需要较长时间的观察，因此不适合大规模评估生成场景。近期的基于仿真的方法尝试通过具身人工智能（AI）智能体评估家具移动性和步行性，以弥合这一差距 [46]。但这些方法无法捕捉人类的行为灵活性，因为人们会动态调整行为，例如跨越障碍物、弯腰通过狭窄空间，或开辟意想不到的最优路径。运动生成研究 [43, 44]能够更逼真地建模人-场景交互，但需要大量训练数据，且难以应对行为上的即兴变化。

为弥合视觉指标与真实世界可用性之间的评估差距，我们开发了专门针对生成的三维布局的基于行为的指标。这些指标衡量人类如何在空间中导航并利用空间。我们的方法捕捉可观察的行为模式：轨迹变化性揭示布局的直观性，动作序列暴露补偿性动作，

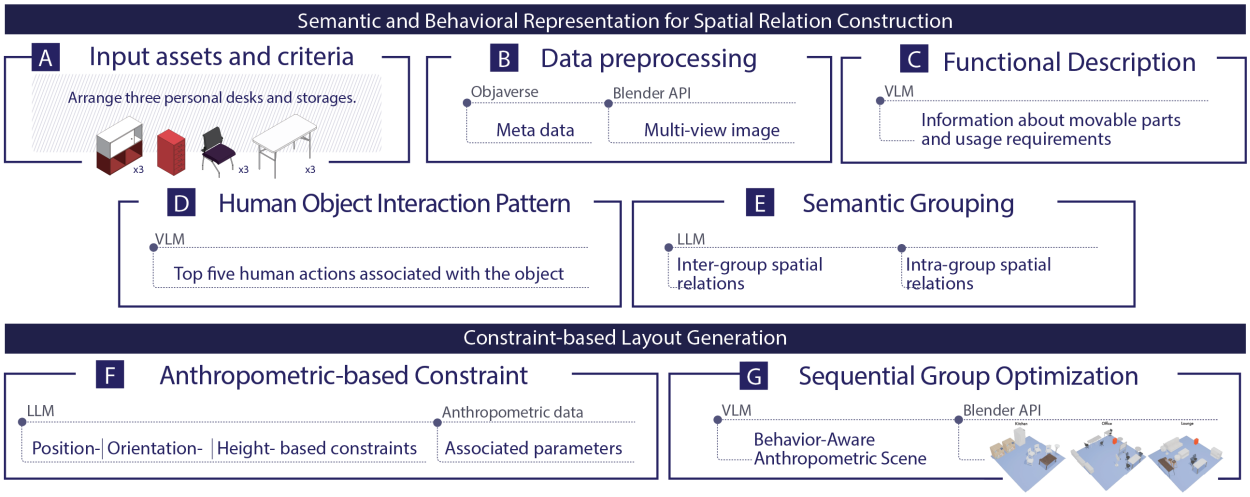


图 2：行为感知人体测量场景生成概览。该框架分为两个阶段：语义与行为表征（阶段 A–E）构建空间关系；基于人体测量约束的布局生成（阶段 F–G）则利用人体测量约束优化最终布局。

而占用率则验证操作空间充足性。该评估框架提供了实证证据，表明人体测量感知生成能够提升功能表现，从而建立了一种基于人体操作标准评估生成场景的新范式。

3 行为感知人体测量场景生成

我们提出**行为感知人体测量场景生成**，这是一种利用行为推理和人体测量依据增强语言驱动布局生成的方法。其目标是生成兼顾家具功能与人-物交互的空间布局。通过明确参照行为背景和人体测量数据，我们的方法能够使系统推断空间约束。如图 2 所示，我们的框架以场景指令和物件为输入，分为两个主要阶段：

(1) **用于空间关系构建的语义与行为表征**：构建行为感知的关系表征，整合物体语义、人体与物体的交互模式以及群体层级空间关系（图 2A–E）。(2) **基于约束的布局生成**：推断符合人体测量学依据的约束表征，以用于可微空间优化（图 2F–G）。

为确保可复现性，附录 A 提供了完整的提示模板。

3.1 用于空间关系构建的语义与行为表征

这些**A-E**阶段 会解读原始三维资产和布局标准，生成行为感知关系表示，将物体几何、功能与人机交互联系起来。这些表示通过建模场景中物体的使用、接触和协调方式，为约束推断奠定必要的上下文基础。以下章节将介绍我们如何组织资产数据、提取行为特征，并对物体进行语义分组，以便进行约束推断。

3.1.1 数据与输入。根据用户指定的布局标准（例如房间类型、所需家具），我们从开放宇宙数据集中检索三维资产，并将其放置到初始场景布局中：**[A] 输入资产与标准**（图 2 A）。对于每个输入资产，我们利用 OpenShape [21], ——一种通过对齐点云与文本和图像描述来检索 3D 资产的大规模多模态模型——从开放宇宙数据集 Objaverse 中提取候选物体 [9]。确定资产后，我们提取每个物体的 3D 包围盒和元数据（类别、宽度、深度、高度），并使用脚本化的 Blender 流程生成 3D 模型的多视角渲染图：**[B] 数据预处理**（图 3B）。局部坐标经过标准化处理（+X：右， +Y：前， +Z：上）。我们使用视觉语言模型（GPT-4o），通过描述物体几何形状、部件和可能交互的提示，联合解读视觉与文本线索。为向视觉语言模型提供视觉依据，我们使用 Blender 实现了自动化渲染流程。每个物体均从四个正交视角渲染（0°、 90°、 180° 270° 和），以全面呈现几何细节。多视角方法能够捕捉决定物体功能、但常常未出现在文本描述中的细粒度特征（例如脚轮、铰链或门把手）。我们将这些多视角图像与结构化元数据配对，使视觉语言模型能够交叉参照视觉证据与文本描述。组合输入使系统能够推断物体的功能属性，例如运动轴、关节连接点和运动学约束；这些属性无法从静态类别标签中获知。

3.1.2 基于人-物交互的特征提取。仅凭物体几何无法确定家具应如何接近或操作

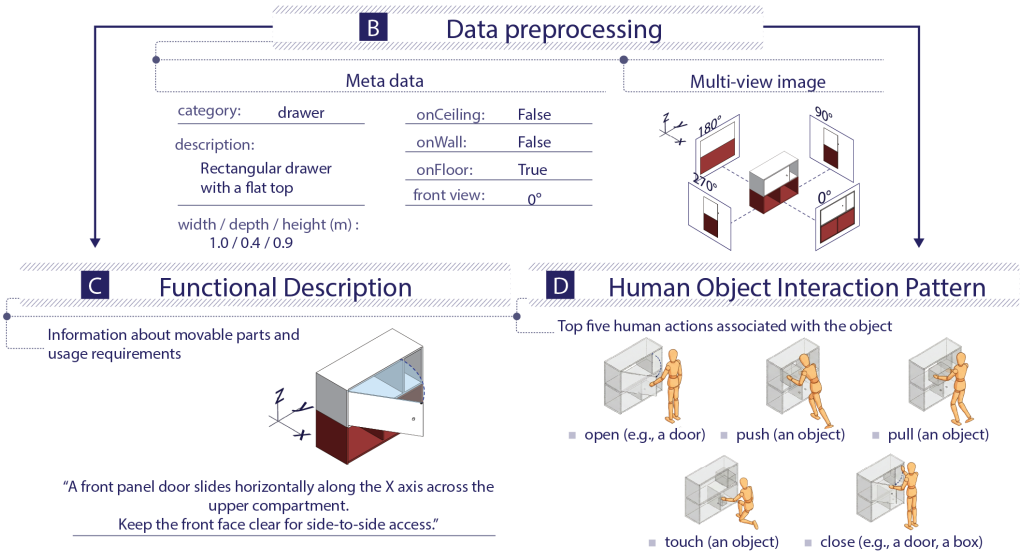


图 3：语义与行为表征（阶段 B–D）。对于场景中使用的每个三维资产，[B] 会预处理元数据和多视角渲染，[C] 会推断功能描述，[D] 会提取人-物交互模式。

——柜子可能向外打开或横向滑动，椅子可能旋转，也可能固定不动。本阶段推断每个物体的功能和行为属性：它如何运作，以及人如何与其交互。视觉语言模型以配对的多视角图像集和元数据为输入，并生成两个互补的输出：**[C] 功能描述**（图 3C），记录有关可动部件和使用要求的定性信息；以及 **[D] 人-物交互模式**（图 3D），基于原子视觉动作（如坐下、打开、拉动）识别与该物体相关的前五种人体动作 [13]。由此得到的视觉到文本处理流程会生成结构化的 JSON 表示，将物体与推断出的功能和交互语义关联起来。这些表示为约束推断提供了行为基础（详细提示见图 A1；输出示例见图 A2）。

3.1.3 语义分组。The **[E]语义分组**阶段（图 4E）会将物体组织成反映人们在场景中如何与其交互的功能组。虽然可以单独解读各个物体，但将它们视为行为配置的一部分时，才能进行有意义的空间推理；例如，围绕书桌摆放椅子组成工作区，或围绕咖啡桌摆放沙发形成休息区。语义分组有两个目的。首先，它会在组的情境中重新解读阶段 [D] 推断出的原子级人-物交互，以提取空间定义。具体而言，系统会识别**组内空间关系**（功能单元内部的布局）和**组间空间关系**（不同组之间的连通性），随后在阶段 [F]将其转换为结构化的符号程序。例如，在多柜存储组中理解时，原子级的“打开”或“拉动（某个物体）”动作会转化为更高层次的“整理”关系。

其次，它通过将每个组视为独立单元，并依据功能重要性和物体尺度确定其行为优先级，从而建立结构抽象，降低阶段 [G] 的优化复杂度（例如，床或书桌等场景定义元素会优先摆放）。我们使用附录图 A3 所示的系统进行语义资产分组，将每个组定义为一组由功能关系、几何邻近性和共享人类动作联系起来的物体（输出示例见图 A4）。

3.2 基于约束的布局生成

[F-G] 阶段会根据自然语言空间关系，推断基于人体测量依据的约束表征。它们将这些源自物体几何、功能和人-物交互的关系，转化为执行空间优化所需的可执行、可微约束。与依赖大语言模型常识的既有布局方法不同，[12, 35, 47], 我们的框架通过明确参考行为语义和人体测量依据来推断约束规范。

3.2.1 空间约束定义。为定义物体之间的空间关系，我们构建了一套扩展分类体系，改编自先前布局和场景生成方法中使用的几何关系表述 [29, 35, 47]。随后，我们以行为和人体测量学为基础，重新诠释约束语义并进行参数化。以往工作从几何角度定义约束，关注物件之间的距离和角度，而不考虑人们如何与这些物件交互。相比之下，我们的表述明确编码了人们如何触及、使用家具以及在家具周围活动。该分类体系将自然语言关系归纳为可学习的约束类型——位置、朝向和基于高度的约束，例如椅子靠墙、桌子与沙发对齐，或灯放在书桌上。表1列出了每种约束类型及其对应的约束名称。

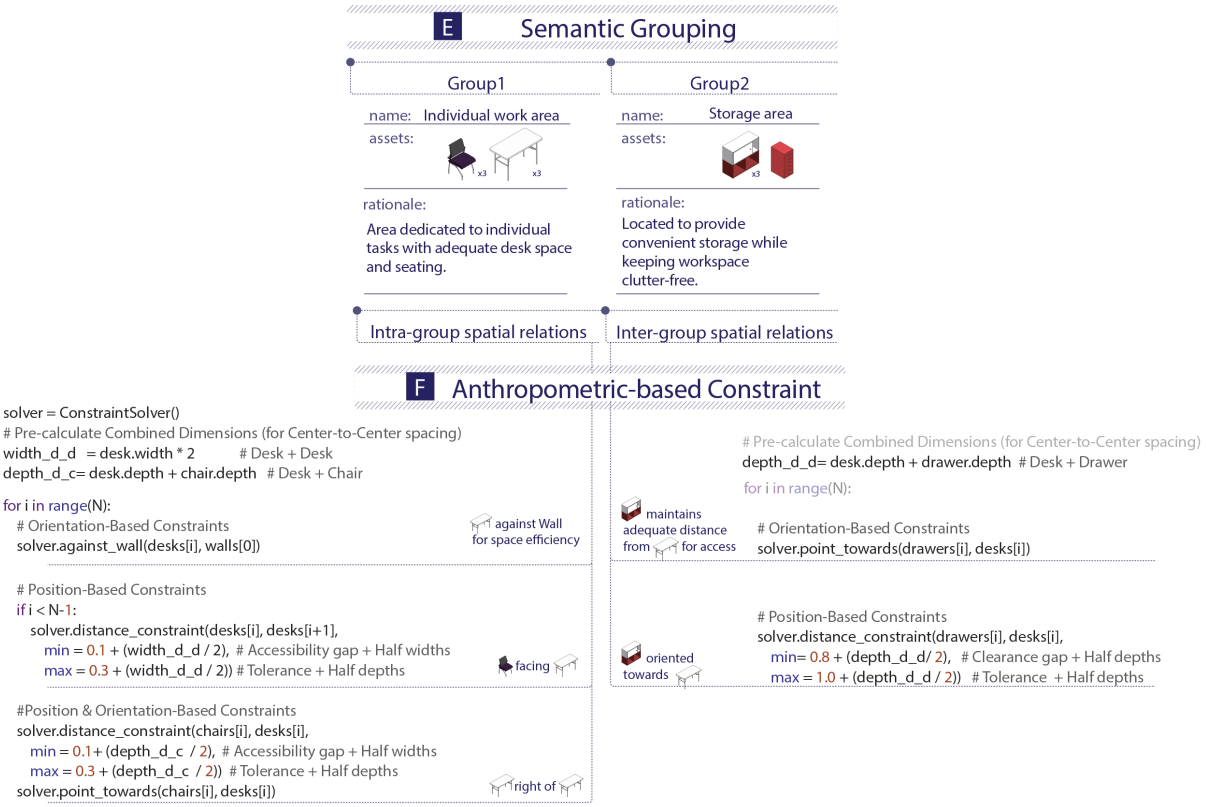


图 4： [E] 语义分组和 [F]基于人体测量的约束推断。

其中，距离约束需要进一步说明，因为它直接决定人体可达性和功能间隙。我们将每个距离约束表示为中心到中心的距离范围 $[d_{\min}, d_{\max}]$，并根据人们触及物体、站立以及在物体周围移动的方式明确推导得出。这里，$d_{\text{accessibility}}$ 表示触及或使用物体所需的最小距离，而 $d_{\text{clearance}}$ 表示进行操作所需的空间（例如，拉出椅子）。我们根据交互语义推断这些边界：对于以可达性为中心的关系， $d_{\min} = d_{\text{accessibility}}$ 和 $d_{\max} = d_{\min} + \tau$；对于以净空为中心的关系， $d_{\min} = d_{\text{clearance}} - \tau$和 $d_{\max} = d_{\text{clearance}}$，其中 τ是视觉语言模型根据物体功能和交互情境推断出的容差缓冲。

3.2.2 基于人体测量的约束推断。在 **[F]基于人体测量的约束推断**阶段（图 4F）中，我们根据自然语言空间关系推断完整的约束规范，包括约束类型、参数及相应的人体测量依据。我们首先解析每条关系（例如，组内关系办公椅面向书桌，组间关系双人箱柜与书桌保持足够距离），并将其归入预定义的约束类型之一。每种类型都包含一个参数模板，用于指定需要推断的空间量，例如距离、角度或垂直偏移。为推断这些参数，我们参考标准化的人体测量数据集《人体尺度与室内空间》[28], , 该数据集汇集了多个来源的人体测量数据 [8, 26, 40]。这些数据集可作为人群层面的参考。我们采用关键水平尺寸的第5至第95百分位范围，包括前向伸及距离、侧向伸及距离、体宽和体厚，以体现人群中的自然差异。这使我们推断出的约束能够适用于不同体型和运动能力的人群，而不必假设每个人都是同一个平均个体。

随后，我们将每种约束类型映射到特定的人体测量依据：与伸及相关的关系使用臂展测量，间隙与通行关系使用体宽或体深，邻接关系使用协同运作物体所需的最小偏移，视觉/朝向约束则使用眼位和偏好观看方向，使物体正面朝向主要交互或观看区域。通过整合关系类型、相关人体测量范围以及交互，我们推断约束类型及其相关参数，例如距离约束 $[d_{min}, d_{max}]$、align_0°和align_90°。最后，我们将该规范编码为可执行优化模式，记录约束类型、物体对、参数值以及人体测量依据。该结构化表示被直接传递给阶段 [G] 中的可微优化模块（系统提示见附录图A5；输出示例见图A6）。

表1：行为感知人体测量场景生成的空间约束分类体系，对比了我们基于行为感知和人体测量学的约束与几何关系 [35]。

约束类型	约束名称	方法	描述
基于位置	$L_{\text{distance}}(p_i, p_j, d_{\min}, d_{\max})$	LayoutVLM [35]	两个资产之间的距离应在范围 $[d_{\min}, d_{\max}]$ 内。
		Ours	两个物体之间的距离为 $[d_{\min}, d_{\max}]$，其中边界推断自基于人体测量数据的伸手范围和净空要求。
	$L_{\text{against 墙}}(p_i, w_j, b_i)$	LayoutVLM [35]	将物件放置在墙 w_j 旁。
		Ours	将物体靠着指定的墙放置，同时考虑可及性和邻近交互的净空要求。
基于方向	L_{align} 与 (p_i, p_j, Θ)	LayoutVLM [35]	以指定角度 Θ对齐两个资产。
		Ours	对齐两个物体的旋转；角度参数 Θ 反映面向任务对齐（例如，用于共同使用的平行或垂直配置）。
	朝向 (p_i, p_j, Θ)	LayoutVLM [35]	调整一个物件，使其以偏移角 Θ 朝向另一个物件。
		Ours	调整朝向，使物体正面朝向目标，其中 Θ 编码偏好的观看或交互方向（例如面向书桌或座位区）。
基于高度	L_{on} 顶部 (p_i, p_j, h)	LayoutVLM [35]	将一个物件放置在另一个物件上方。
		Ours	定义垂直堆叠关系，以便将较小物体放置在表面上同时确保有足够的交互区域。

符号说明： p_i, p_j 物体位姿；w_j 墙体索引；b_i 物体包围盒；$d_{\min}, d_{\max}$ 最小/最大中心距离；Θ 相对角度；h 高度偏移。

3.2.3 基于*VLM*的序列组优化。 在前几个阶段之后， **[G]序列组优化**阶段会编译一个约束程序，将布局标准以及组内和组间空间关系编码为可微的、以人体测量为依据的表达式。每条约束都纳入了先前阶段定义的功能依据和人类行为条件，使优化过程能够同时保持语义合理性和操作有效性。我们将所有约束表示为可微函数，从而能够使用基于梯度的优化。遵循 LayoutVLM [35], 的一般优化结构，我们将每次违反表述为针对物体位姿的连续惩罚。然而，与LayoutVLM不同——后者的损失主要来自视觉语言模型推断的空间关系和基于物理的碰撞项——我们的违反项以人体测量可达性、功能间隙以及基于行为的空间语义为依据。具体而言，我们将每条推断出的约束转换为一个可微惩罚函数：

- **距离约束**会在中心到中心的距离偏离推断出的$[d_{\text{最小}}, d_{\text{最大}}]$ 范围时产生违规。
- **朝向约束**会在物体朝向不同于根据交互推断出的对齐角度时产生违规。

优化过程会最小化汇总后的违反程度，$L = \sum_i w_i \cdot \text{violation}_i(\theta)$，其中 i 枚举所有行为感知且以人体测量学为依据的约束， w_i 控制优化过程中各项约束的相对影响，$\text{violation}_i(\theta)$则根据物体位姿 θ 与约束的目标距离、朝向或组级配置之间的偏差计算连续惩罚。我们采用自适应加权，其中 w_i 会在物体发生显著重叠时（包围盒重叠为 > 50%）优先避免碰撞，同时允许灵活调整行为约束。

优化按组依次进行。对于每个组，我们首先最小化组内约束违反程度。优化后续组时，会应用组间约束，以维持先前已放置组与当前组之间的行为关系。优化器运行 400 次基于梯度的迭代，最终布局在 Blender 中渲染，使用优化后的物体位姿和平面图坐标，单位为米。

4 技术验证

以下小节将介绍实验流程、基于几何的评估、用户感知研究，以及用于评估我们的方法相较于以往基于大语言模型的方法的布局质量的结果。

4.1 实验流程

在技术验证中，我们的目标是评估所提出的场景生成框架如何影响布局质量，并将其与以往基于大语言模型的方法进行比较。该框架以人-物交互推理和人体测量数据为基础。我们采用了三维场景生成研究中常用的验证框架，由两个部分组成：**基于几何的指标**（物体到物体碰撞分数和物体到地面边界内分数）[4, 12, 14, 35, 36]以及**用户感知研究** [35, 36]。

基线。 我们将方法与最先进的系统 LayoutVLM [35]进行了比较。我们选择基线，是因为它是具有代表性的基于大语言模型的布局方法，将场景合成建模为一个可微约束优化问题。这种优化结构使我们能够进行受控比较，两者唯一的主要差异在于约束的获取方式。基线依赖大语言模型的常识，而我们的方法则推断约束

来自人-物交互和人体测量参考数据。我们没有纳入其他基于大语言模型的布局方法 [4, 47]，因为用固定物件或确定性放置来约束它们，会损害其核心生成过程，无法在需要保持处理一致性的受控人类被试实验中进行公平比较。

设置。 验证涵盖五种房间类型（卧室、休息室、办公室、厨房和餐厅），每种类型包含两个场景，共计十个测试场景，每个房间平均有 9.9 个物件。所有场景均在标准化的 5.5 × 5.5 × 2.5 米平面图内生成。对于家具物件，我们按照先前的方法 [4, 35]，从经过验证和预处理的 Objaverse [9] 中为每种房间类型选取三维资产。针对每种方法和场景，我们使用不同的随机种子生成五个候选布局，并选取无碰撞得分最高的布局进行分析。由此得到一组匹配的 20 个布局（10 个场景 × 2 方法），用于本节基于几何的评估和专家感知评估。

人体测量数据。 由于我们的方法将人体尺寸纳入约束推断过程，我们利用合成人体测量特征，在多样化的用户体型上验证这一能力。首先，我们确定了六项关键身体尺寸（例如前向伸及距离、侧向伸及距离、体宽和体厚）的 5^{th}–95^{th} 百分位范围。随后，我们向大语言模型提供了在这些范围内采样的若干示例记录，并提示其生成仍处于相同百分位界限内的合理身体尺寸组合。在推断约束时，为每个场景生成一份合成档案，从而使技术验证能够覆盖多样化用户。

4.2 基于几何的评估

为确保对物理有效性进行标准化的定量比较，我们采用了广泛使用的无碰撞得分和边界内分数 [4, 12, 14, 35, 36]。选择这些指标是为了验证生成场景的基本物理有效性，因为物理可行性（即无重叠、位于墙内）是布局能够被有意义地评估可用性之前的必要条件。对于这两项指标，数值越高表示物理合理性越强。**无碰撞得分** 衡量互不重叠的物体对所占的比例：

$$\mathrm{CF} = \frac{1}{N_c} \sum_{(i,j)} \mathbb{I}(d_{ij} > r_i + r_j), \tag{1}$$

其中，d_{ij} 表示物体中心之间的距离，r_i 和 r_j 分别表示物体 i 和 j 的包围圆半径，$\mathbb{I}(\cdot)$ 在条件成立时返回 1，否则返回 0；N_c 表示评估的物体对数量。**边界内分数**用于衡量所有物体是否都位于房间内：

$$\mathrm{IB} = \frac{1}{N_o} \sum_{i} \mathbb{I}(0 \le x_{\min}(i), x_{\max}(i) \le W,\ 0 \le y_{\min}(i), y_{\max}(i) \le D), \tag{2}$$

其中，$x_{\min}(i)$ $x_{\max}(i)$ $y_{\min}(i)$、和 $y_{\max}(i)$ 表示物体的平面图包围盒范围，W 和 D 分别表示房间的宽度和深度，N_o 表示物体的数量。

4.3 对生成场景的用户感知研究

参与者。我们招募了 16 名参与者（六名男性和十名女性；年龄为 $M = 26.88$，$SD = 1.78$），他们是室内设计和建筑学领域的专业人士。所有参与者均已取得建筑学或室内设计专业的大学学位，因此熟悉空间布局推理和家具布置。

实验流程。本研究在线上进行。对于每个布局，参与者会看到两张场景渲染图（一张俯视图和一张透视图），并使用7点李克特量表（1 = 非常不同意，7 = 非常同意）对布局的五项评估标准进行评分。为帮助参与者准确判断，界面中提供了朝向箭头和带度量标注的平面图（图 5）。最终共获得 1,600 条评分（16 名参与者 × 20 个布局 × 5 项评估标准）。

测量指标。我们根据基线研究 [35] 对评估标准进行了细化，以体现我们方法的具体贡献。基线研究最初使用位置、朝向以及综合的、基于物理的语义对齐分数（综合了物理合理性和语义一致性）来评估布局。然而，使用综合分数会限制我们区分两种布局的能力：一种只是逻辑上正确，另一种则在实际中适合人类使用。由于我们的方法着重于人体操作可用性，因此将基线的综合指标拆分为三个更细致的评估标准：语义合理性、物理合理性和功能可用性。这样便能单独考察并衡量基于人体测量学的约束如何提升场景的具体可操作性。参与者对以下陈述表示认同程度：

- **位置恰当性：**每个物体是否被放置在房间内适合其功能与用途的位置（例如，不会不合理地远离相关家具）。
- **方向适切性：**每个物体的朝向是否与其功能相符（例如，椅子面向书桌，或电视面向座位区）。
- **语义合理性：**家具的布置是否在逻辑上与空间的预期用途一致（例如，床相对于床头柜的摆放是否合理）。
- **物理合理性：**布局在基本空间规则方面是否看起来在物理上可行（例如，没有明显的碰撞或边界违规）。
- **功能可用性：**人们是否看起来能够以现实的方式接近、触及并操作每个物体（例如，有足够空间完全打开抽屉、拉出椅子，或站立并在台面上工作）。

4.4 技术验证结果

基于几何的评估结果。 关于物理有效性的比较，我们的方法取得了90.4的无碰撞得分和73.4的边界内分数，而基线的得分分别为92.2和63.2。尽管我们的框架无碰撞得分略低（−1.8%），但边界遵循度显著提高（+10.2%）。无碰撞得分和边界内分数指标提供了客观的基于几何的物理

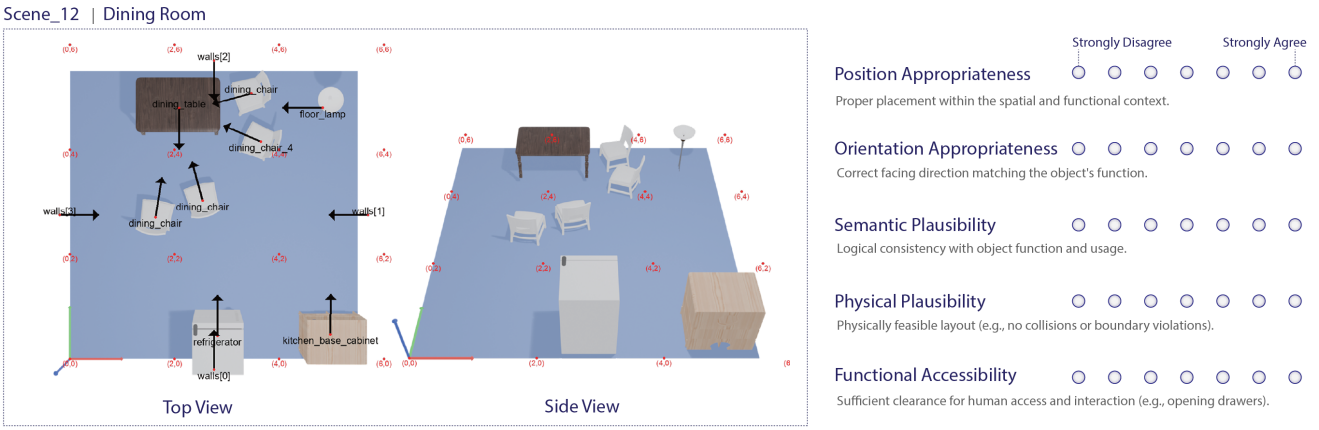


图 5：用户感知研究界面。红色标记表示以米为单位的平面图坐标，箭头表示家具朝向。参与者使用 7点李克特量表对各项标准进行评分。

表2：用户感知研究结果。数值表示中位数，方括号内为四分位距 [1st 四分位数、3rd 四分位数]。统计显著性采用带Holm-Bonferroni校正的Wilcoxon符号秩检验确定（*p < 0.01）。

测量指标	基线	Ours	p_{adj}	r
位置恰当性	3.00 [2.25, 4.00]	**5.00** [4.38, 6.00]	<0.01	0.85
方向适切性	3.00 [2.00, 3.25]	**5.00** [4.50, 6.13]	<0.01	0.88
语义合理性	3.00 [2.75, 3.63]	**5.00** [4.38, 6.00]	<0.01	0.86
物理合理性	3.00 [2.00, 4.13]	**5.00** [5.00, 6.00]	<0.01	0.78
功能可用性	3.00 [2.00, 4.00]	**5.00** [4.50, 6.00]	<0.01	0.77

通过测量物体是否相互重叠以及资产是否位于房间边界内，评估其合理性。

用户感知研究结果。 在考察用户感知研究的结果时，我们的方法在全部五项评估标准上均持续优于基线（表2）。我们的方法在所有指标上的中位数均为5.0，而基线的中位数仍为3.0。从分布来看，我们的方法的 1st 四分位数值介于4.375至5.0之间， 3rd 四分位数值介于6.0至6.125之间。相比之下，基线的表现显著更低，其 1st 四分位数值介于2.0至2.75之间， 3rd 四分位数值介于3.25至4.125之间。值得注意的是，尽管基线取得了更高的无碰撞得分，这一差异并未转化为更好的用户感知。采用Wilcoxon符号秩检验并结合Holm-Bonferroni校正的统计分析证实，这五项标准上的差异均具有统计学显著性（校正后 $p < 0.01$）。此外，效应量较大（$r > 0.77$）。这些结果表明专家认为，与基线相比，我们的布局明显更适合人类操作。

可用性研究的局限性与动机。 基于几何的评价指标在评估人类可用性时存在局限。首先，这些指标在固定物体假设（例如，不可活动的抽屉）下评估场景，因此无法考虑动态物体功能或实际的人类行为模式。其次，尽管用户感知研究通过纳入专家判断对这些指标进行了补充，但它完全依赖于对静态渲染的视觉估计。对布局进行视觉评估，并不能保证其支持实际的物理操作，例如舒适地伸手取物或无阻碍移动；对于身体尺寸各异的用户而言，这一点尤其如此。这一局限促使我们开展人体操作可用性研究，这是一项基于任务的用户研究，旨在真实使用场景中评估生成场景。

5 人体操作可用性用户研究

为探究我们提出的行为感知与人体测量方法能否转化为实际操作效益，我们基于每位参与者测得的人体测量数据参数化布局，并据此开展了以人为中心的可用性评估。我们设计了两项互补的用户研究，从不同角度考察参与者在这些场景中的实际移动与行为。两场实验的流程如图 6 所示。

- **个体环节** (N = 20)：通过可控的定量测量，评估基于人体测量学且考虑行为的布局对任务绩效的支持效果。借助结构化的单用户任务，我们记录了运动行为，量化了任务效率和交互空间使用情况，并收集了可用于统计评估本方法的数据。
- **小组会议** (N = 18，六个三人团队)：评估相同布局在真实多用户场景中的表现，在这些场景中，多个使用者会自然而然地共享空间。通过分析累计运动轨迹并评估任务后访谈

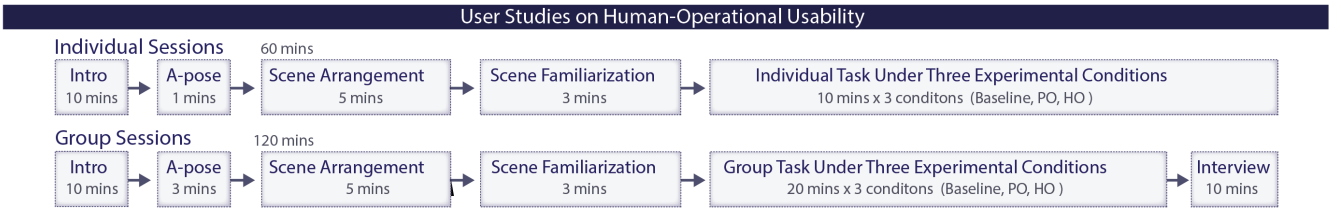


图 6：个人和小组会议的实验流程图。

反馈，我们记录了布局如何支持空间协商与协作活动，这些条件反映了真实世界中的共享环境。

5.1 个人和小组研究的设置

以下实验条件、环境、设备和数据收集流程均一致地应用于个人和小组研究。

5.1.1 实验条件。我们评估了三种条件：一种基于几何的**基线[35]**，以及我们框架的两个版本，**仅通行（PO）**和**人体操作（HO）**。这两个版本的区别仅在于注入空间约束中的人体测量参数类型（图 7）。

- **基线：** 未使用任何个性化人体测量信息生成的布局；空间关系仅由通用几何约束和碰撞检查确定。
- **PO：** 我们的框架采用静态人体测量尺寸（例如，身体宽度、躯干深度），以确保最小通行宽度和基本可导航性。
- **HO：** 我们的框架采用基于运动的人体测量尺寸（例如，伸展臂展、前向伸及距离、坐姿臀部至脚趾长度），以确保人-物体交互有足够的操作空间。

PO 和 HO 采用相同的空间关系和约束类型；二者仅在距离的最小值和最大值上有所不同：PO 使用静态身体尺寸，HO 则使用运动包络。这种设计支持两项关键比较：比较 PO 与 HO，可以单独考察动态任务感知人体测量的影响，因为可用性方面的任何差异都直接体现了对基于运动的操作空间进行建模所带来的价值。比较基线与 HO，则可以评估我们框架的整体收益；该框架融合了语义场景理解、空间关系推断和个性化人体测量参数化。这项比较凸显了行为感知布局如何通过使空间约束与实际人体运动相匹配，优于基于几何的方法。

5.1.2 评估环境。我们选取办公室和休息室环境作为具有代表性的室内空间，这些空间中会自然发生多样的物体布局和人类活动，支持工作、协作、储存、导航和休息任务，同时保持足够的实验控制。三种条件（基线、仅通道、人体操作）均采用相同的 5.5 × 5.5 × 2.5m 平面图，并按照技术验证中使用的相同实验流程生成。在用户研究中，仅通道和人体操作条件均根据参与者分别实例化，将每位参与者测得的人体测量档案代入距离约束，而基线则完全基于几何信息。在所有条件下，办公室和休息室场景共享相同的家具组和整体场景拓扑结构（图 7）。

5.1.3 动作捕捉与设备。 我们采用基于EasyMocap [33] 的无标记动作捕捉设置，以获取绝对人体坐标和身体测量数据，用于场景生成与人-场景交互评估。我们在 $9 \times 9 \times 2.5$米的室内空间中安装了八台RGB摄像机（Logitech网络摄像头），并使用 4×7 棋盘格 [33]进行内参与外参标定。标定后的摄像机在 $5.5 \times 5.5 \times 2.5$米的捕捉空间内以4K/60fps录制同步多视角视频。为重建三维人体运动，录制数据逐帧经过（1）二维关键点提取、（2）三维三角测量和（3）SMPL拟合处理。实验设置、摄像机布置以及采集的数据类型如图8所示。

5.1.4 人体测量数据采集。 我们从拟合后的SMPL模型中提取参与者特定的人体测量数据，用于参数化仅通道与HO条件中使用的距离约束。据此，我们采用了Bojanić等人 [3], 提出的与姿态无关的三维人体测量方法，该方法可根据稀疏的三维人体标志点预测标准化身体测量数据。我们选择该方法，是因为它能够准确重建标准A姿态尺寸，而不受采集过程中参与者实际姿态的影响。这消除了对严格A姿态扫描的需求，同时保持了与稠密几何方法相当的精度。对于每位参与者重建的SMPL人体，我们首先提取70个人体标志点的三维坐标，并将其输入预训练预测模型，以获得一组标准化人体测量尺寸。对于Bojanić模型未直接提供的额外任务特定操作尺寸，例如与我们的交互场景相关的特定伸及距离或关节间间距，我们计算了所选SMPL标志点对之间的欧几里得距离。由此得到的综合轮廓作为实例化个性化物体间距离约束的定量依据。

6 用户研究1：个体环节

个体环节考察了三种场景条件——基线、仅通道和人体操作——如何影响个性化布局对单个用户的任务层面可用性。本研究基于人体测量个性化和动作捕捉框架，重点对人体操作表现进行受控的定量评估。具体而言，我们比较了三种条件在以下方面的表现：

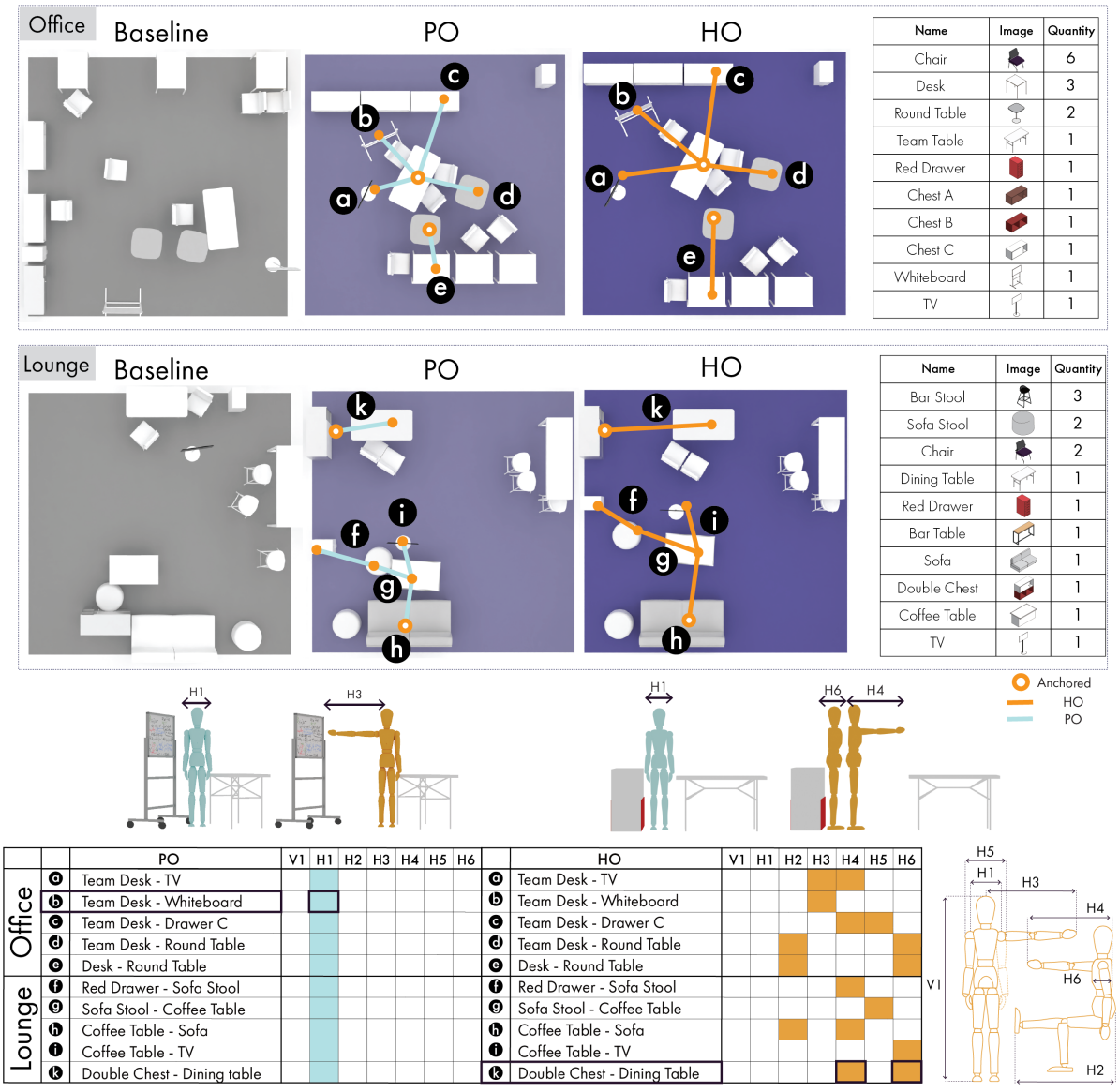


图 7：基线，　办公室和休息室环境中的 PO 和 HO 布局

任务完成时间、导航过程中的轨迹效率，以及关键物体周围的交互空间利用率。

6.1 实验设置

参与者。个体环节共有 20 名参与者：办公室空间中有 10 人（4 名男性和 6 名女性；年龄为 $M = 24$ $SD =$ ，），休息室空间中有 10 人（5 名男性和 5 名女性；年龄为 $M = 24.3$，$SD = 2.63$）。

实验流程与任务。每个环节开始时，我们会说明研究目的和实验流程，然后使用动作捕捉设备记录每位参与者在空旷空间中的 A姿态。我们将捕捉到的数据拟合到 SMPL模型，并计算人体测量数据，并使用这些测量数据为每位参与者生成 PO 和 HO 布局。每种环境的基线布局均为固定布局。一位研究人员负责说明任务并组织简短的练习环节，其他研究人员则按照指定布局布置家具，准备实际环境。三种布局条件（基线、PO、HO）的顺序通过拉丁方进行平衡。正式试验开始前，参与者接受了安全说明，并了解了家具命名规则。针对每种指定环境（办公室或休息室），我们定义了十项任务，以再现日常情境，例如在家具之间穿行、坐下和打开抽屉，确保每件家具至少使用一次（表3）。任务说明

表3：办公室和休息区环境的顺序任务列表。

办公室任务序列	休息室任务序列
O1. 从C号箱柜中取出箱子，并将其放在顶部。	L1. 移动到1号沙发凳旁，坐下。
O2. 将箱子搬到C号书桌上，坐在它前面。	L2. 从咖啡桌上拿起杯子，移动到2号沙发凳旁，坐下。
O3. 站起来，从红色抽屉的第三个隔间里拿出笔。	L3. 拿着杯子移动到吧台桌旁，坐下。
O4. 移动到2号圆桌旁，用遥控器打开电视。	L4. 手持杯子，走到双人箱柜，将杯子放入隔间内。
O5. 在1号圆桌上，将电视上的两个内容名称写到纸上。	L5. 在红色抽屉中找到遥控器，走到沙发，坐下。
O6. 手持纸，走到B号箱柜，取出白板笔。	L6. 从沙发处打开电视，找到内容，大声读出标题。
O7. 手持纸和白板笔，走到团队桌，坐下。	L7. 将遥控器放在咖啡桌上。
O8. 走到白板前，抄下你写的两个词。	L8. 走到双人箱柜前，取出笔和纸。
O9. 坐在团队桌旁，从右向左倒着读白板上的词。	L9. 走到餐桌旁，坐下，在纸上写下内容标题。
O10. 走到A号书桌旁，坐下，把记住的词写在纸上。	L10. 站起，将纸和笔放入双人箱柜的下层隔间中。

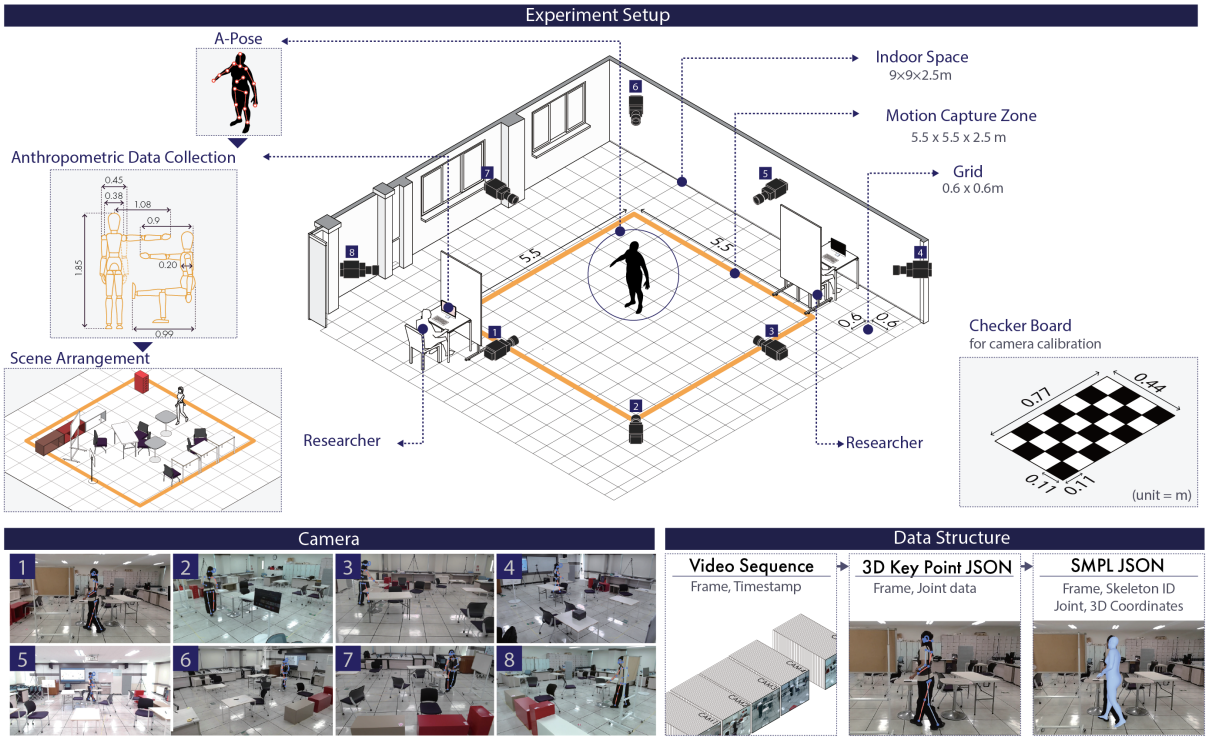


图8：实验环境、动作捕捉设置和数据处理流水线。该示意图展示了八相机阵列与捕捉区域的物理配置，以及从原始视频到三维SMPL模型重建的顺序数据流。所有空间测量均以米为单位。

遵循 物体–动作–目标 格式 [1, 44]。为保留初始布局，参与者被要求不得移动家具，椅子除外。各项任务按顺序衔接；一项任务的终点即为下一项任务的起点。每位参与者都在指定空间的三种条件下各完成十项任务，共记录了 60 段交互序列（20 名参与者 × 3 条件），每次试验约持续 60 分钟。

6.2 测量指标与数据分析

我们的测量指标集旨在评估所提出的人体测量约束是否能带来可测量的可用性提升。每项指标都将施加的人体测量约束与使用的某个特定方面关联起来：任务绩效、轨迹效率或交互空间利用率。我们重点关注四项指标：

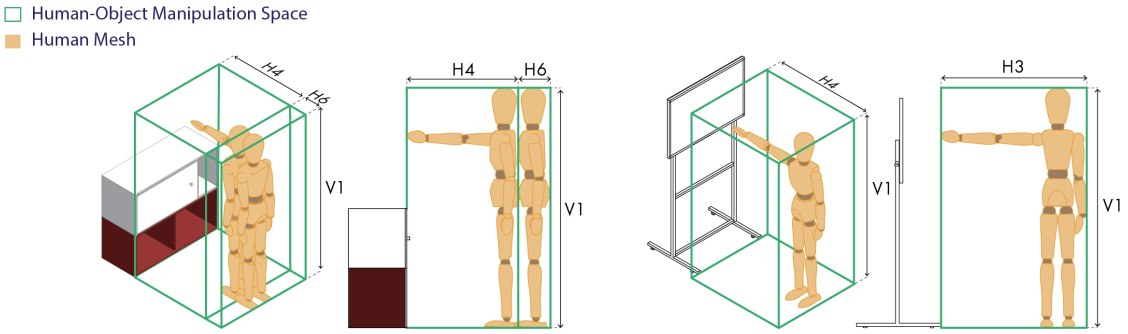


图9：人体—物体操控空间的定义。绿色线框定义了人体—物体操控空间，这是一个由参与者特定的人体测量尺寸（例如垂直高度 V1、水平触及距离 H3、H4）参数化的包围体。

任务完成时间。 我们测量了任务完成时间，起点为参与者收到完整的任务说明后开始行动的时刻，终点为他们口头表示完成的时刻。三位研究人员分别逐帧审阅所有录像，标注精确的起止点，并交叉核对标注结果以确保一致性。

轨迹计数。 为分析布局如何影响移动路径，我们将 SMPL 骨盆关节投影到二维平面，生成时间序列轨迹，然后统计不同路径的数量。当轨迹绕过家具或障碍物时采取了不同的迂回路线，便将其视为不同轨迹，由此量化参与者完成相同任务时所采用的不同路径数量。

序列动作标注。 为捕捉移动行为在动作层面的差异，我们使用 MMAction2 [7], 时空动作检测模型，自动标注参与者的动作序列。我们将每段视频划分为 6 帧的短片段，并检测每个片段中的动作。对于三个摄像机视角中的每个 6 帧片段，我们提取置信度高于 0.6 的前两个动作标签。随后，我们合并三个视角的标签，并按置信度加权，得到每项任务的最终动作序列。

体积占用率。 该指标衡量参与者对物体周围预期交互空间的使用程度。我们首先根据从每位参与者的人体测量档案中推导出的、针对各目标物体的人体操作距离，将人-物操作空间定义为每个目标物体周围的 3D 包围盒。参与者的身高被设为垂直范围，而人体测量距离约束则确定其宽度和深度（图 9）。我们关注带有可动部件且需要足够净空的物体：双人箱柜、单门柜、办公室中的白板，以及休息室中的红色抽屉。借助 Blender 和 Python 脚本，我们计算了：（1）参与者在任务执行期间身体所占据的累积 3D 体积（所有帧中人体网格体积的时间并集）；以及（2）该体积与人-物操作空间的交集。

M 表示累积人体网格体积（跨帧的时间并集），B 表示所定义的人-物操作空间（包围盒）。该比率定义为：

$$\text{Volumetric Occupancy Ratio} = \frac{\text{Vol}(M \cap B)}{\text{Vol}(B)}$$

其中，Vol(·)表示体积运算符，$\cap$ 表示交集。我们针对每项任务计算该比率，并对结果取平均，得到每位参与者的最终指标。

我们使用混合设计方差分析（ANOVA）分析实验结果，其中布局（基线、PO、HO）为被试内因素，环境（办公室、休息室）为被试间因素。对于每项指标，我们首先检验布局 × 环境交互作用，以确认各环境中的效应是否一致，然后考察布局主效应。考虑到样本量较小，后续两两比较采用 Wilcoxon符号秩检验，并使用 Bonferroni校正（$\alpha = 0.017$）这一保守方法。

6.3 用户研究1结果：个体环节

混合设计方差分析显示，布局对任务完成时间（$F(2,36) = 4.38$, $p = 0.020$）、轨迹计数（$F(2,36) = 7.45$, $p = 0.002$）、序列动作标注（$F_{=}$ (2,36) 32 49 $p < 0\ 001$ $F\ 2,36 = 66\ 12\ .\ ,\ .$）和体积占用率（( ) . , $p < 0\ 001$ × .）均具有显著主效应。未观察到显著的布局与环境交互作用（所有 ps > 0.05），表明两种环境中的效应一致。下文将报告各项指标的详细比较结果。

6.3.1 任务完成时间。 较短的任务完成时间表示布局更高效。当参与者执行相同的任务序列时，更快完成意味着布局能够提供更短或更少受阻碍的路径，以及更便捷的目标物体访问。因此，我们在一项个体研究中比较了完成十项连续任务所需的总时间，其中每项任务的终点都作为下一项任务的起点。在办公室环境中，人体操作的平均完成时间为14.87秒，而基线为17.11秒，仅通道为17.22秒（图10a）。基线与人体操作之间的差异为

统计显著（$p < 0.01$）。人体操作达到了最短的完成时间，比基线少用13.1%的时间。在休息室环境中也观察到了类似的模式：人体操作需要9.14秒，仅通道需要10.66秒，基线需要11.59秒（图10b）。休息室中的任务完成时间差异显示，基线与人体操作之间存在显著差异（$p < 0.01$），与基线相比减少了21.1%。

6.3.2 轨迹效率。 我们使用轨迹计数和序列动作标注来测量布局的轨迹效率，以评估这些布局是否促成了直观的路径选择，并通过所使用的轨迹数量分析其经济效率。

轨迹计数。 较低的轨迹计数表明布局更加直观且高效（图11）。若计数为1，则体验该布局的所有参与者都遵循了同一路径，表明该场景提供了清晰且自然的路线。所有序列的平均轨迹计数在办公室和休息室布局中均呈现出一致的模式：人体操作 < 基线 < 仅通道，其中人体操作的轨迹计数最低。在办公室中，人体操作在所有参与者中平均每个任务产生2.5条不同路径，基线产生2.6条路径，仅通道产生3.1条路径，尽管这些差异在统计上并不显著（图10c）。在休息室中，人体操作为1.6条路径，基线为1.9条路径，仅通道产生3.0条路径（图10d）。与仅通道相比，人体操作所需路径更少（$p < 0.01$），减少了46.7%，而仅通道的路径多于基线（$p < 0.05$）。

序列动作标注。我们统计了逐帧标注动作，以确定完成任务所需的动作序列，并计算了每项任务的平均动作数。动作数越少，表示轨迹越高效；额外动作通常源于不必要的行为，例如移动障碍物或进行纠正性调整。例如，在任务 L4 中，参与者需要手持杯子从吧台桌旁站起来，走到双人箱柜前，并将杯子放入隔间。在人体操作休息室布局中，P12 用四个动作完成了任务 {sit-hold-walk-stand}。在基线布局中，为了移开挡路的凳子，还需要额外执行一个“弯腰”动作，因此动作序列变为 {sit-hold-walk-stand-bend-stand}。因此，动作数更少意味着完成任务的轨迹更直接、更高效。在办公室布局中，人体操作平均需要 4.79 个动作，而基线需要 6.99 个动作，仅通道需要 8.95 个动作（图 10e）。与基线相比，人体操作所需动作显著更少（$p < 0.05$），减少了 31.5%；与仅通道相比也显著更少（$p < 0.01$），减少了 46.5%。在休息室布局中，人体操作平均需要 2.95 个动作，基线需要 4.18 个动作，仅通道需要 5.26 个动作（图 10f）。与基线相比，人体操作的动作数显著更少（$p < 0.001$），减少了 29.4%；与仅通道相比也显著更少（$p < 0.001$），减少了 43.9%。

6.3.3 交互空间利用率。对于体积占用率，数值越高，表示交互空间利用率越好；参与者占用了更多预期的人体—物体操作空间，也能更顺畅地操作物体。定性示例展示了这一规律（图12）。在 PO 办公室布局中，由于团队桌摆放得太近，P4 只能使用白板的一部分；因此，累积身体体积在白板操作框内仍然较低。在基线休息区布局中，由于沙发凳挡住了前方的通行区域，P17 不得不侧身在双人箱柜旁操作，因此操作空间内的占用也较为稀疏。相比之下，在两种场景中，参与者执行相关任务时，HO 布局的操作框内占用都更加密集且分布均匀。在办公室布局中，HO 的平均占用率为 0.25，PO 为 0.18，基线为 0.19（图10g）。休息室布局也呈现出类似结果（图10h）。在办公室布局中，HO 的占用率显著高于基线（$p < 0.001$），增幅为 31.6%；也显著高于 PO（$p < 0.001$），增幅为 38.9%。在休息室布局中，HO 的表现显著优于基线（$p < 0.001$），增幅为 52.9%；PO 的占用率也高于基线（$p < 0.01$），增幅为 11.8%。

6.3.4 总结。 在全部四项指标中，HO 条件始终比基线和 PO 布局具有更好的任务层面可用性。

- **任务完成时间：**在办公室和休息室环境中，HO 的任务完成时间均最短，而 PO 相较于基线没有任何改善。
- **轨迹计数：**平均轨迹数呈现一致的模式：HO < 基线 < PO，表明 HO 布局能引导更直观的路线，而 PO 往往会导致额外绕行。
- **序列动作标注：** 在两种环境中，HO 的每项任务所需动作数均最少；与基线和 PO 相比，它减少了移动障碍物或进行纠正性调整等额外行为。
- **体积占用率：** HO 实现了最高的交互空间利用率，而 PO 和基线则有更多预定的操控空间未被使用。

7 用户研究2：小组会议

尽管个体环节能够在受控条件下测量单人任务层面表现，但许多现实环境中都有多人长期共同活动，且活动内容相互重叠。为考察这类多用户动态——在此情境下，布局的可用性不仅取决于个人路径效率，还取决于使用者如何共享空间、经历相互干扰，以及如何据此调整行动——我们组织了小组会议，让小型团队在个体研究中使用的相同基线、PO 和 HO 布局中持续开展协作工作。除了动作捕捉数据外，我们还进行了事后访谈，以定性考察各布局条件下的感知拥挤、协调和工作流程支持情况。

7.1 实验设置

参与者 小组会议涉及18名参与者，组织成六个团队。在办公室空间中，三个团队（九名参与者：三名男性和六名女性；年龄为 $M = 23.56$、$SD = 2.92$）执行任务；其余三个团队（九名参与者：四名男性和五名女性；年龄为 $M = 23.78$、$SD = 2.49$）在休息室空间中执行任务。

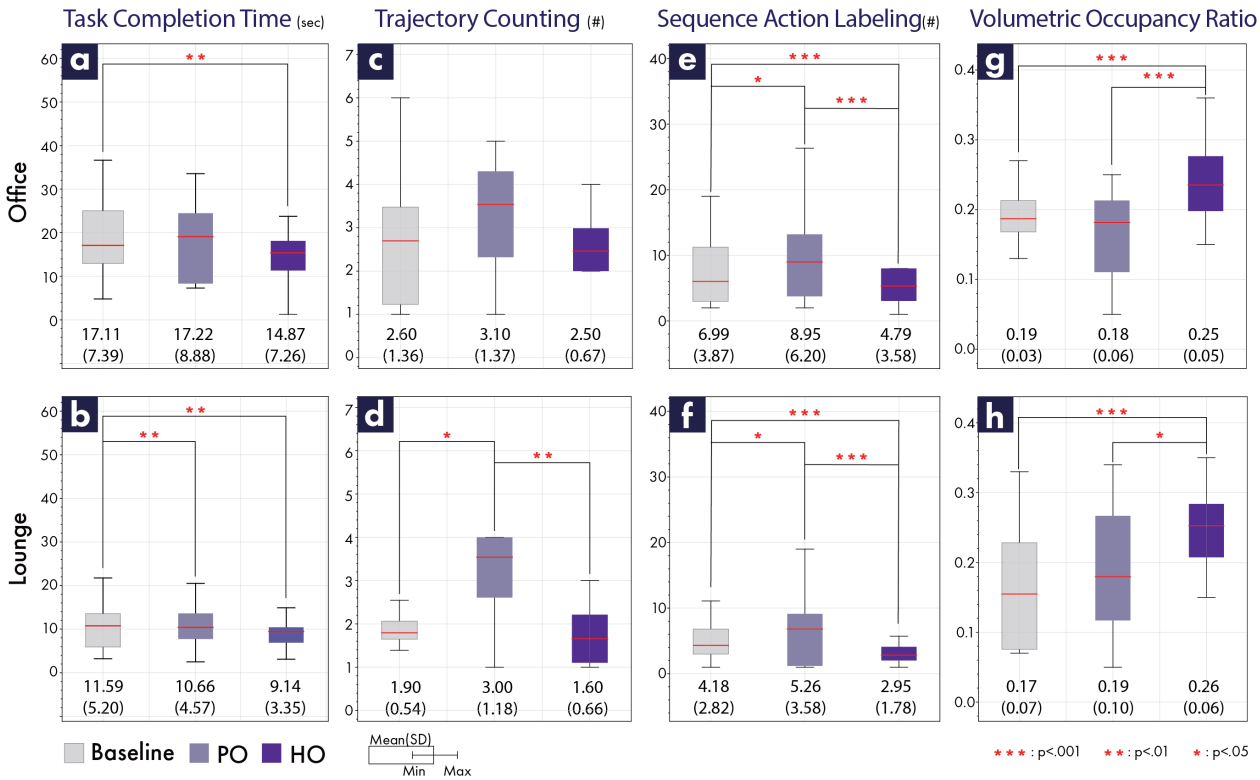


图 10：测量指标结果。(a, b)：任务完成时间。(c, d)：轨迹计数。(e, f)：序列动作标注。(g, h)：体积占用率。

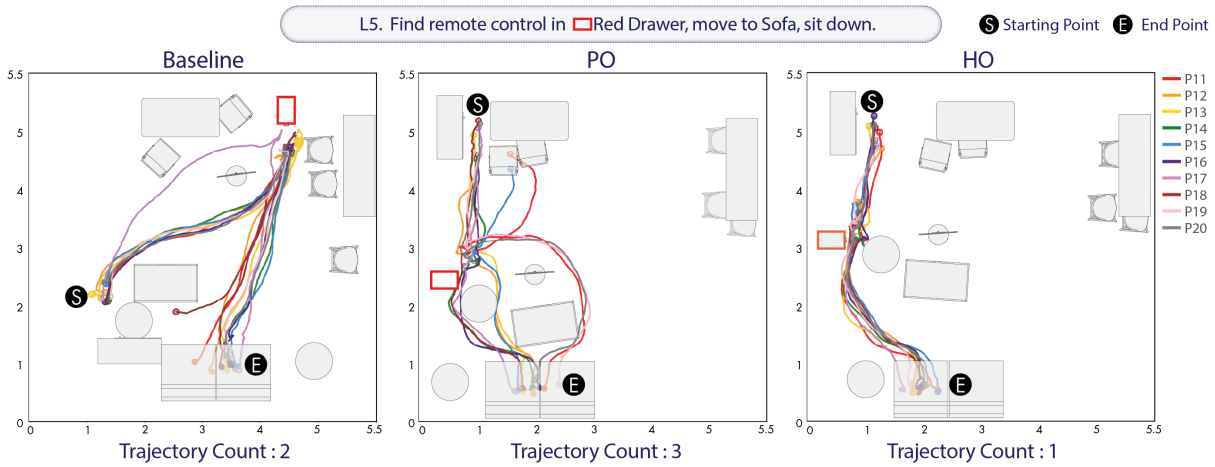


图 11：L5 的轨迹计数和参与者轨迹可视化示例。

实验流程和任务。每次会议开始时，研究人员都会向各团队介绍研究内容，并使用相同的动作捕捉和人体测量框架对其进行记录。为了进行场景生成，我们通过取团队三名成员在每个维度上的最大值，构建了团队级人体测量档案，并使用此配置文件来实例化 PO 和 HO 约束。完成安全说明和简短的练习环节后，我们根据当前条件布置家具，并要求团队按照空间交互协议，在三种布局（基线、PO、HO）中分别完成一系列协作涂色和分类任务。

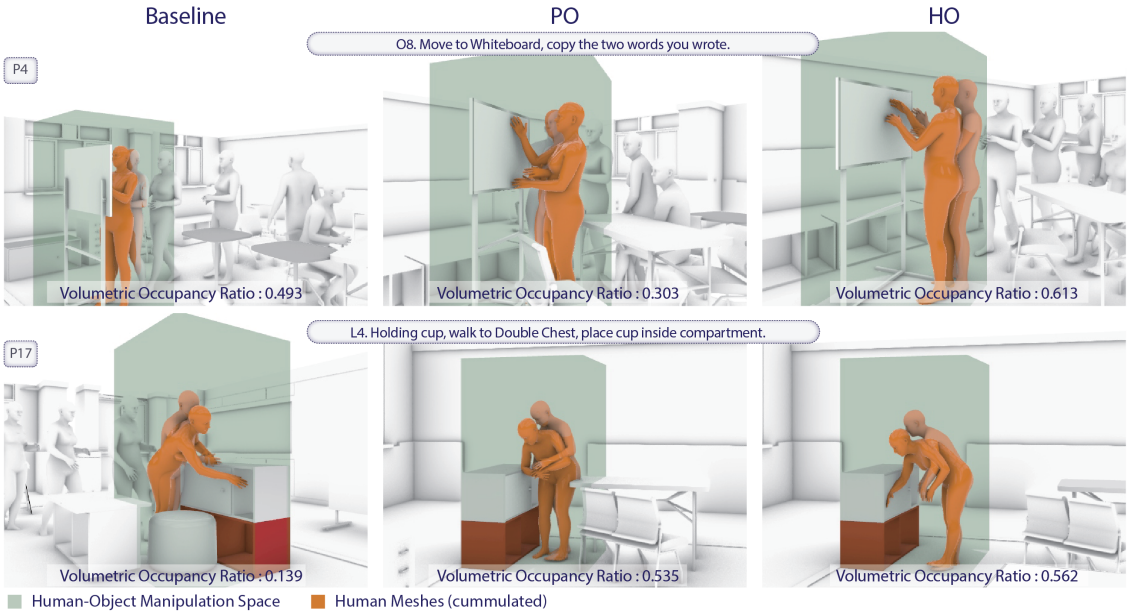


图12：P4（O8）和 P17（L4）的人-物操作空间（绿色箱子）及 SMPL 网格的图示示例。

改编自 [23, 24]。在每种布局中，参与者先查看电视上的模板，从储物单元中取出材料，在个人书桌旁借助共享空间中的提示开展工作；随后重新集合，讨论电视上显示的分类标准，并将最终模板贴到白板上。任务材料分散在空间各处，以鼓励资源交换和移动，自然而然地促成了流动、路径交叉和聚集。完成全部三种条件后，我们在单独的房间里对参与者进行了个别事后访谈。每次团队活动约持续 120 分钟，共获得 18 段群体交互序列（六个团队 × 三种条件）。

7.2 数据分析与可视化

空间使用可视化。为分析群体移动模式，我们开发了平均速度的热力图可视化。在空间的 1024×1024 个离散网格中，对于每个网格单元 $(i,\ j)$，平均移动速度按如下方式计算：

$$\bar{v}_{i,j} = \frac{\sum_{p=1}^{P}\sum_{t=1}^{T} v_{p,t} \cdot \delta_{i,j}(x_{p,t}) \cdot \Delta t}{\sum_{p=1}^{P}\sum_{t=1}^{T} \delta_{i,j}(x_{p,t}) \cdot \Delta t}$$

其中，$v_{p,\ t} = \|x_{p,\ t+1} - x_{p,\ t}\|\,\text{fps}$ 是参与者 p 在时间 t 的速度，而当位置 $x_{p,\ t}$ 落在网格单元 $(i,\ j)$ 内时，$\delta_{i,\ j}$ 等于 1。应用了高斯平滑（$\sigma = 0.01\text{m}$）以生成连续表示。各条件采用一致的颜色缩放，从而能够直接比较空间使用模式。

7.3 用户研究2结果：小组会议

对于每个团队，我们汇总了他们在整个任务序列中的移动情况，并将其可视化为热力图。我们观察到空间使用模式存在显著差异。下文将详细报告研究结果，并结合参与者访谈中的见解，分析观察到的这些运动行为（图13）。

7.3.1 功能可见性与物理可及性。在基线条件下，几何布局往往无法满足功能可见性和可及性的需求，导致不必要的移动。在办公室中，电视的朝向迫使参与者反复来回移动才能看清屏幕，而在 PO 和 HO 布局中则无需额外移动（图13）。访谈中，9 位参与者（P23–P29）中的 7 位认为电视是最不方便的物件；P28 提到：“我想把电视转向大家坐着的桌子，”P29 则表示：“为了看清电视屏幕而走的路太长了。”

休息室也出现了类似问题，参与者指出电视和被阻挡的双人箱柜是主要痛点。三位参与者（P30、P31、P36）表示，电视距离太远，只能从斜侧角度看到。此外，1号沙发凳挡住了双人箱柜的正面，阻碍了取用，只能从侧面交互。四位参与者（P30–P32、P38）认为这一阻碍是造成不便的主要原因，并明确指出凳子既妨碍材料取用，也影响流动。

7.3.2 轨迹连续性与操作净空。聚焦于人体操作（HO）与仅通道（PO）的差异，轨迹可视化显示，由于布局过于拥挤，PO 布局常导致轨迹中断和操作瓶颈（图13）。

在办公室中，HO 布局呈现出连续的移动轨迹，而 PO 布局在B号箱柜与白板之间，以及1号圆桌与书桌1之间出现了中断。P28谈及PO时表示：“由于圆桌紧挨着团队桌，坐着很不舒服；而在站立会议期间，我不得不绕很远的路。”

在休息室中也观察到了类似的模式。对于双人箱柜，人体操作布局提供了充足的人-物操作空间，使用者可以顺畅地沿正面轨迹移动。在仅通道布局中，餐

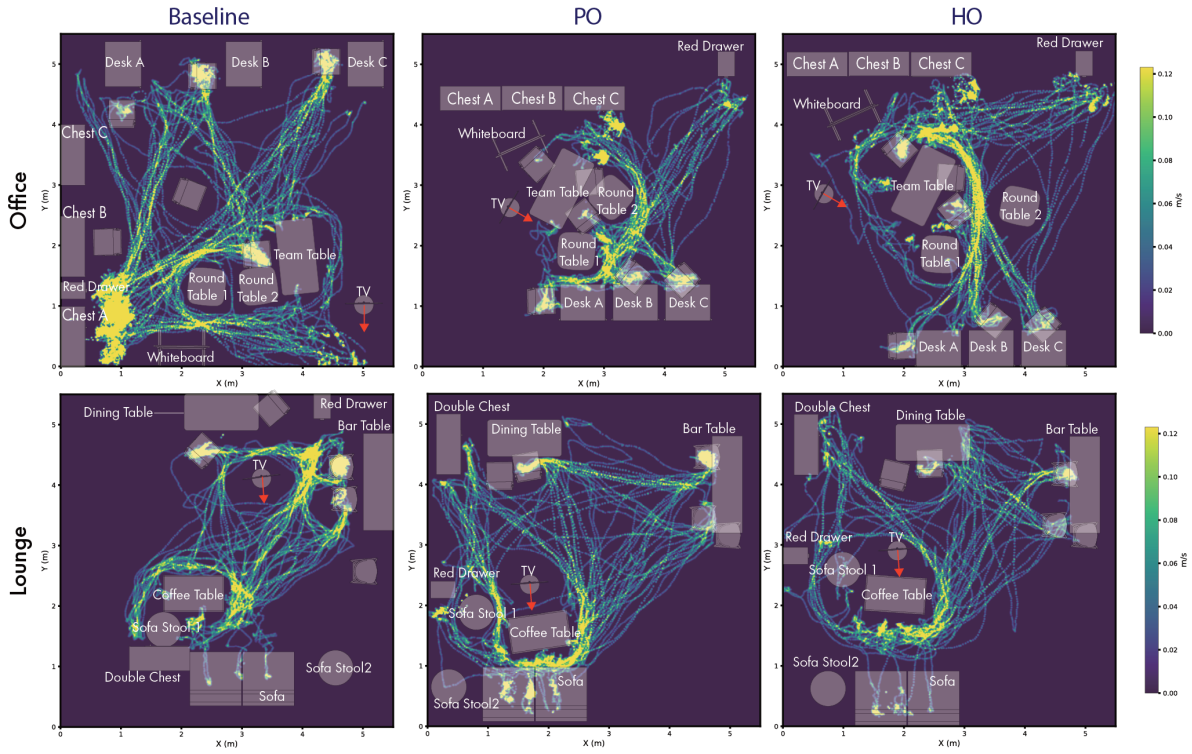


图13：团队1（P21-23；办公室；上方）和团队6（P36-38；休息室；下方）的平均速度热力图可视化

桌侵占了箱柜的操控空间，导致移动通道变窄；P22表示："双开储物柜门无法完全打开，门前形成了一个瓶颈。"四位参与者（P30、P36–P38）认为，高柜与餐桌之间的距离太窄，难以使用高柜，也难以坐在附近的椅子上。相比之下，参与者认为人体操作布局更宽敞，也更能有效利用可用空间，并强调在高柜与其他工作区域之间移动时毫无阻碍。P30特别指出，高柜的开启空间有所改善，并将人体操作布局评为最理想的布局。

7.3.3 总结。 在小组会议中，HO 条件始终比基线布局和 PO 布局都具有更好的群体层面的可用性，这与轨迹可视化和参与者报告一致。

- **功能可见性与可及性：** HO 确保电视、沙发和储物家具等共享资源在合适的距离和方向上便于使用。相比之下，基线布局经常影响可见性或造成物理阻挡，妨碍人们使用这些资源，导致不必要的移动。

- **轨迹连续性与操作净空：** 人体操作支持在关键工作区域周围保持连续的移动模式，减少了仅通道布局中观察到的瓶颈。尽管仅通道布局提供了基本的通行宽度，却未考虑操作空间侵占（例如，家具阻挡储物空间可达性），导致轨迹中断和用户投诉。

这些定性观察补充了我们的定量发现，表明基于人体测量学的布局不仅能提升个人任务绩效，还能促进共享空间中的群体互动更加自然、高效。

8 讨论

我们的研究结果表明，相较于语义合理性，行为感知的人体测量约束能显著提升布局的可用性。我们从四个关键方面探讨其影响：生成的布局中的功能关系、操作净空要求、在 XR 环境中的应用，以及局限性与未来研究方向。

8.1 行为感知布局中的功能关系

语义驱动基线与人体操作（HO）的比较表明，语义合理性并不能保证可用性。尽管语义驱动基线能够正确地将相关物体归为一组（例如将咖啡桌摆放在沙发附近），但它往往未能提供交互所需的物理净空。

我们关于人体操作可用性的两项用户研究清楚地揭示了这些局限性，表明基线布局中存在大量功能性脱节。例如，基线经常将电视放置在别扭的观看角度，或将桌子摆放在阻挡白板可达性的位置——这些布置在语义上是合理的，但实际上无法使用。相比之下，我们的行为感知布局明确推断了人体动作所需的操作空间范围。

人体操作保留了这些功能关系，使参与者能够维持自然的工作流程，而无需反复走到电视前，或以别扭的斜角观看电视。总体而言，这些结果表明，仅靠语义驱动方法不足以捕捉人类行为模式或物体之间的功能关系，而行为感知约束能切实提升生成空间的可用性。

8.2 操控操作净空

PO 与 HO 的比较凸显了一个关键区别：通行净空并不等同于操作净空。尽管这两种条件都采用人体测量数据，但它们对净空的定义不同。PO 仅根据身体深度和宽度确保通行净空，而 HO 还要求为操控和交互提供操作净空。

这一区别会带来显著的行为差异。与 PO 相比，HO 大幅减少了补偿性动作和轨迹中断，消除了用户重新摆放家具或绕开障碍物的需要。HO 下更高的空间占用量从数量上证实，为通行预留空间并不自动意味着操作空间也得到保障。尤其是在群体研究中，PO 缺乏操作约束，造成了实际瓶颈：物体摆放得过于靠近，妨碍人们在圆桌旁就座，或因家具占据空间而导致柜门无法完全打开。这些问题表现为原本无碰撞布局中的轨迹不连续。上述比较证实，即使采用相同的人体测量数据，为单纯满足通行净空而设计，与为操作净空而设计，仍然有所不同。这些发现为约束选择提供了实用指导：当布局涉及频繁的物体操控（例如打开抽屉、调整家具或取用储物空间）时，操作净空（HO）能带来可衡量的益处；而对于交互需求极少、以通行为主的空间，仅通行约束（PO）或许就已足够。通过阐明人体测量感知操作约束何时能提升可用性，本研究帮助从业者判断哪些地方值得投入额外的设计精力。

8.3 在 3D 环境中的应用 ts

我们的框架将人体测量因素编码为可微约束，从而为静态物理布局生成以外的应用创造了机会。我们将讨论行为感知场景生成可能发挥价值的三个方向，同时指出这些领域的实证验证仍有待未来开展。

*适配设备增强型交互。*在 XR 环境中，用户通常通过控制器进行交互，或执行幅度较大的手势指令，从而使其运动学空间范围实际上超出自身的身体尺寸 [2, 32]。我们的框架可以扩展到此类设备特定操作范围，从而有望生成能够考虑基于控制器或手势交互所需额外净空的布局。这代表了对我们人体测量参数化的扩展，不过具体要求仍需通过 XR 环境中的用户研究加以验证。

*提升 XR 中的具身交互保真度。*维持 XR 中的沉浸感取决于尽可能缩小用户意图与虚拟约束之间的差距 [39]。值得注意的是，虚拟环境无需付出体力即可即时修改空间布局，这使实时人体测量适配成为可能——而这一能力在物理环境中并不现实。将我们的框架应用于 XR 意味着，生成的环境可以帮助避免交互不匹配，例如用户必须过度伸展才能触及某个控制装置，或看到自己的虚拟手穿透桌子表面。通过根据用户（或虚拟化身）特定的人体测量特征优化布局，我们的框架有望支持自然且无碰撞的虚拟空间工作流程，从而在无需手动后处理的情况下，提升具身交互保真度。

*面向 AI 智能体和数字孪生的可扩展内容创作。*最后，对于需要大规模生成合成场景的情境——例如 AI 智能体训练、机器人导航或数字孪生仿真——我们的方法或可为提升行为真实性提供一条途径。用户无需进行静态资产放置，而是可以指定高层活动（例如，坐下并取回物体）。随后，系统会生成能够为这些动作提供所需运动学空间的布局，从而有望减少虚拟智能体或机器人在功能上彼此隔绝的空间中难以导航时，模拟中常见的别扭动作或不自然的逆运动学表现。

8.4 局限性与未来工作

我们的研究着重探讨人体测量数据和行为数据能否提升计算场景生成的可用性。基于这一研究范围，我们将在下文讨论若干局限性和未来方向。

首先，关于可推广性和验证范围，我们的技术验证涵盖了 20 个场景，用户研究则聚焦于 6 个场景。此外，尽管我们的参与者群体具有足够的统计效力，能够检测这些情境中的可用性改进，但样本量限制了我们对不同人群作出广泛结论的能力。未来研究应开展大规模评估，并将验证扩展到厨房、浴室或零售空间等复杂且受到严格监管的环境，以制定规范性设计指南。其次，由于资产库缺乏参数化可变性，我们当前的框架侧重于水平布局。因此，架子高度或站姿与坐姿转换等垂直交互尚未得到优化。未来工作应纳入参数化对象生成，以实现完整的三维优化，使系统能够根据个人的人体测量档案调整家具高度。第三，关于多用户优化，我们当前的方法优先考虑可及性，采用群体中最大的身体尺寸来确保所有人都有足够的净空。然而，这也揭示了一种固有的权衡：确保最大体型用户拥有几何净空，往往会导致较矮参与者的交互距离不舒适（P22）。更广泛地说，这一发现表明，适当的约束优先级可能因物体关系和预期交互而异——有些关系适合将物体放得更近，以便伸手触及；另一些则需要留出更大的间距，以便操控。由于我们的主要重点是考察

行为和人体测量信息如何影响布局质量，我们没有探讨如何自适应地平衡这类相互竞争的约束。未来研究应探索自适应约束优先级机制，根据每对物体的具体交互情境进行考量，并遵循面向多元用户群体的通用设计原则。第四，我们使用预处理数据集中的代表性家具资产，对基于视觉语言模型的约束推断进行了验证。系统能够可靠地推断受测家具资产的功能描述和交互模式。然而，由于视觉语言模型依赖视觉线索和常识，对于视觉上不明显的可供性（例如隐藏机构或非标准开启方向），若无补充文本信息则无法推断。对于非传统家具设计（例如形式多样的家具或功能模糊的多功能家具）的泛化能力仍未经检验。最后，我们当前的框架是在物理环境中实现并评估的；不过，同样的人体测量约束和行为感知约束也可以扩展到扩展现实环境。随着扩展现实技术在工作场所和住宅环境中日益普及，[17, 23, 24], 空间布局将越来越需要兼顾物理和虚拟交互区域。

9 结论

我们介绍了行为感知人体测量场景生成，这是一种通过行为推理和人体测量依据增强基于语言的布局生成的方法。该方法通过考察行为信息和人体测量信息如何影响布局质量，将关注重点从视觉与语义合理性转向人体操作可用性。我们的贡献主要有三点：首先，我们提出了一个两阶段场景生成框架，将行为推理转化为可微空间约束，从而能够通过基于梯度的优化提升布局的可用性。其次，我们定义了操作净空和交互区域的明确空间约束，确保布局能够适应动态人体动作（例如伸手或打开抽屉），而不只是避免静态碰撞。最后，我们通过技术验证和用户研究证明，行为中心指标能够捕捉人们实际使用空间的方式，超越传统的碰撞和边界评分。我们的行为感知人体测量视角，是迈向扩展现实、具身人工智能和数字孪生系统中功能性操作环境的重要一步；在这些环境中，人体移动、伸手和交互将成为首要设计约束。

致谢

本研究得到由贸易、工业和能源部（MOTIE，韩国）资助的技术创新计划（RS-2025-02317326，基于设计师意图的 AI 驱动设计生成技术开发）的支持。

参考文献

[1] Panos Achlioptas、Ahmed Abdelreheem、Fei Xia、Mohamed Elhoseiny 和 Leonidas Guibas。2020。Referit3d：用于真实场景中细粒度三维物体识别的神经听众。载于欧洲计算机视觉会议。Springer，422–440。doi:10.1007/978-3-030-58452-8_25[2] Karan Ahuja、Vivian Shen、Cathy Mengying Fang、Nathan Riopelle、Andy Kong 和 Chris Harrison。2022。Controllerpose：结合虚拟现实的由内向外的身体捕捉控制器摄像头。载于《2022年计算机系统中的人为因素CHI会议论文集》。1 13 10 1145 3491102 3502105 – 。doi：。/。David Bojanić、[3] Stefanie Wuhrer、Tomislav Petković和Tomislav Pribanić。2024。基于稀疏数据的姿态无关的三维人体测量。载于《欧洲计算机视觉会议论文集》。Springer，237–256。doi:10.1007/978-3-031-91575-8_15[4] Ata Çelen、Guo Han、Konrad Schindler、Luc Van Gool、Iro Armeni、Anton Obukhov和Xi Wang。2024。I-design：个性化大语言模型室内设计师。载于《欧洲计算机视觉会议论文集》。Springer，217–234。doi:10.1007/978-3-031-92387-6 17_[5] Pei Chen、Kexing Wang、Lianyan Liu、Xuanhui Liu、Hongbo Zhang、Zhuyu Teng和Lingyun Sun。2025。探讨混合现实在设计表征中的作用，以增强用户参与式协同设计沟通。《ACM人机交互汇刊》9, 2 2025 1 29 10 1145 3710979 ( ), –。doi：。/。[6] Francis DK Ching。2023。。《建筑：形式、空间与秩序》。约翰·威立父子出版公司。[7] MMAction2贡献者。2020。OpenMMLab的新一代视频理解工具箱与基准。https://github.com/open-mmlab/mmaction2。[8] Albert Damon、Howard W Stoudt和Ross A McFarland。1966。《设备设计中的人体》。哈佛大学出版社。doi:10.4159/harvard.9780674491892[9] Matt Deitke、Dustin Schwenk、Jordi Salvador、Luca Weihs、Oscar Michel、Eli VanderBilt、Ludwig Schmidt、Kiana Ehsani、Aniruddha Kembhavi和Ali Farhadi。2023。Objaverse：一个带标注的三维物体宇宙。载于《IEEE/CVF计算机视觉与模式识别会议论文集》。13142–13153。doi:10.1109/cvpr 52729 2023 01263 。。Iman Dianat、Johan Molenbroek和[10] Héctor Ignacio Castellucci。2018。人体测量学在人体工程学与产品设计中的方法与应用综述。《人体工程学》61, 12 2018 1696 1720 10 1080 00140139 2018 1502817 ( ), –。doi：。/。。[11] Yan Ding、Xiaohan Zhang、Chris Paxton和Shiqi Zhang。2023。。使用大语言模型进行物体重排的任务与运动规划。载于2023年IEEE/RSJ智能机器人与系统国际会议（IROS）。IEEE，2086–2092。doi:10.1109/iros55552.2023.10342169[12] Weixi Feng、Wanrong Zhu、Tsu-Jui Fu、Varun Jampani、Arjun Reddy Akula、Xuehai He、S Basu、Xin Eric Wang和William Yang Wang。2023。LayoutGPT：使用大语言模型进行组合式视觉规划与生成。载于第三十七届神经信息处理系统大会。https://openreview.net/forum?id=Xu8aG5Q8M3[13] Chunhui Gu、Chen Sun、David A Ross、Carl Vondrick、Caroline Pantofaru、Yeqing Li、Sudheendra Vijayanarasimhan、George Toderici、Susanna Ricco、Rahul Sukthankar等。2018。AVA：一个包含时空定位的原子视觉动作的视频数据集。载于《IEEE计算机视觉与模式识别会议论文集》。6047–6056。doi:10.1109/cvpr.2018.00633[14] Siyi Hu、Diego Martin Arroyo、Stephanie Debats、Fabian Manhardt、Luca Carlone和Federico Tombari。2024。用于三维室内场景合成的混合扩散。arXiv: 2405 21066 [ ] 10 48550 2405 21066 .cs.CV doi：。/arXiv.。Begoña Juliá Nehme、Eugenio Rodríguez和So-Yeon Yoon。2020。空间用户体[15] 验：评估物理环境的多学科方法。《室内设计杂志》45, 3 2020 7 25 10 1111 12177 ( ), –。doi：。/joid。Begoña Juliá Nehme、David Torres Irribarra、Patricio Cumsille和So-Yeon Yoon。2021。候诊[16] 室物理环境与门诊患者体验：作为分析工具的空间用户体验模型。《室内设计杂志》46，4（2021）， 27 48 10 1111 12205 –。doi：。/joid。Dooyoung Kim、Seonji Kim、Selin Choi和Woontack Woo。2024。面向混合现实远程呈现的空间可供性知觉交互子空间分配。arXiv: 2408 04297 [ ] 10 1109 62088 2024 00142 .[17] cs.ET doi：。/ismar。。Bokyung Lee、Minjoo Cho、Joonhee Min和Daniel Saakes。2016。[18] 将摆姿与动作作为个性化家具的输入。载于《第九届北欧人机交互会议论文集》。1 10 10 1145 2971485 2971487 –。doi：。/。Bokyung Lee、Joongi Shin、Hyoshin Bae和Daniel Saakes。2018。[19] 。交互式情境化指南，帮助用户设计适合自身身体的个人书桌。载于《2018年交互系统设计会议论文集》。637–650。doi:10.1145/3196709.3196725[20] Chenguo Lin和Yadong Mu。2024。Instructscene：基于语义图先验的指令驱动三维室内场景合成。arXiv预印本arXiv:2402.04717（2024）。doi:10.48550/arXiv.2402.04717[21] Minghua Liu、Ruoxi Shi、Kaiming Kuang、Yinhao Zhu、Xuanlin Li、Shizhong Han、Hong Cai、Fatih Porikli和Hao Su。2023。OpenShape：扩展三维形状表征以实现开放世界理解。《神经信息处理系统进展》36（2023），44860–44879。doi:10.48550/arXiv.2305.10764[22] Matthew Loper、Naureen Mahmood、Javier Romero、Gerard Pons-Moll和Michael J. Black。2015。SMPL：蒙皮多人线性模型。《ACM图形学汇刊》（SIGGRAPH亚洲计算机图形学国际会议论文集）（2015 10 1145 3596711 3596800 ）。doi：。/。[23] Weizhou Luo、Mats Ole Ellenberg、Marc Satkowski和Raimund Dachselt。2025。《掌中的文档：探索增强现实文档空间排列的交互技术》。载于《2025年计算机系统中的人为因素CHI会议论文集》。1–22。doi:10.1145/3706598.3713518

[24] Weizhou Luo、Anke Lehmann、Hjalmar Widengren和Raimund Dachselt。2022. 我们该把它放在哪里？用于协作式意义建构的增强现实文档布局与放置策略。载于《2022年计算机系统中的人为因素CHI会议论文集》。1–16。doi:10.1145/3491102.3501946[25] Weizhou Luo、Zhongyuan Yu、Rufat Rzayev、Marc Satkowski、Stefan Gumhold、Matthew McGinity 和 Raimund Dachselt。2023。Pearl：用于现场人体运动分析的基于物理环境的增强现实镜头。载于《2023年计算机系统中的人为因素CHI会议论文集》。1–15。doi:10.1145/3544548.3580715[26] 美国国家航空航天局。1978年。《人体测量资料手册》。第2卷：人体测量数据手册。Webb Associates，俄亥俄州黄泉市。[27] Binh Vinh Duc Nguyen 和 Andrew Vande Moere。2024。自适应建筑布局：如何通过共享半自主移动机器人隔断的控制权，协调开放式办公室的环境需求与资源。载于《2024年计算机系统中的人为因素CHI会议论文集》。1 20 10 1145 3613904 3642465 –。doi:。/。Julius Panero 和 Martin Zelnik。《人体尺度与室内空间：设计参考标准资料手册》。1979 Watson-Guptill，纽约。[29] Despoina Paschalidou、Amlan Kar、Maria Shugrina、Karsten Kreis、Andreas Geiger 和 Sanja Fidler。2021。Atiss：用于室内场景合成的自回归变换器。《神经信息处理系统进展》34（2021），12013–12026。doi:10.48550/arXiv.2110.03675[30] Haoxuan Qu、Ziyan Guo 和 Jun Liu。2024。GPT-Connect：以无需训练的方式实现文本驱动的人体运动生成器与三维场景之间的交互。arXiv预印本 arXiv:2403.14947（2024）。doi:10.48550/arXiv.2403.14947[31] Raf Ramakers、Danny Leen、Jeeeun Kim、Kris Luyten、Steven Houben 和 Tom Veuskens。2023。测量模式：在制作流程中处理测量值与尺寸的以用户为导向的策略。载于《2023年计算机系统中的人为因素CHI会议论文集》。1–17。doi:10.1145/3544548.3581157[32] Anthony Scavarelli 和 Robert J Teather。2017。VR Collide!：比较共处一地的虚拟现实用户之间的碰撞规避方法。载于《2017年CHI计算系统中的人为因素会议扩展摘要集》。2915–2921。doi:10.1145/3027063.3053180[33] Qing Shuai、Chen Geng、Qi Fang、Sida Peng、Wenhao Shen、Xiaowei Zhou 和 Hujun Bao。2022。基于稀疏多视角视频的新视角合成：人类交互。载于SIGGRAPH会议论文集。doi:10.1145/3528233.3530704[34] Zejia Su、Qingnan Fan、Xuelin Chen、Oliver Van Kaick、Hui Huang 和 Ruizhen Hu。2023。基于语言引导的场景感知活动程序生成：补充材料。《ACM图形学汇刊》42, 6 2023 10 1145 3618338 ( )。doi:。/。Fan-Yun Sun、Weiyu Liu、Siyi Gu、Dylan Lim、Goutam Bhat、Federico Tombari、Manling Li、Nick Haber 和 Jiajun Wu。2025。LayoutVLM：通过视觉语言模型对三维布局进行可微优化。载于《计算机视觉与模式识别会议论文集》。29469–29478。doi:10.1109/cvpr52734. 2025.02744[36] Jia-Mu Sun、Jie Yang、Kaichun Mo、Yu-Kun Lai、Leonidas Guibas 和 Lin Gao。2024。HAISOR：通过深度强化学习实现人体感知室内场景优化。《ACM图形学汇刊》43, 2 2024 1 17 10 1145 3632947 ( )，–。doi:。/ Qi Sun、Hang Zhou、Wengang Zhou、Li Li 和 Houqiang Li。2025 2。《Forest seq：通过序列式室内场景合成重振顺序先验》。载于《欧洲计算机视觉会议》。Springer，251–268。doi:10.1007/978-3-031-72698-9_15[38] Jiapeng Tang、Yinyu Nie、Lev Markhasin、Angela Dai、Justus Thies 和 Matthias Nießner。2024。DiffuScene：用于生成式室内场景合成的去噪扩散模型。载于《IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集》。20507 20518 10 1109 52733 2024 01938 –。doi:。/cvpr . . Yujie Tao、Cheng Yao Wang、Andrew D Wilson、Eyal Ofek 和 Mar Gonzalez-Franco。2023。在虚拟现实中构建具备物理感知能力的虚拟化身。载于《2023年计算机系统中的人为因素CHI会议论文集》。1–15。doi:10.1145/3544548.3580979[40] Harold P Van Cott 和 Robert G Kinkade。1972。《设备设计人因工程指南》。国防部。[41] Carlos Viviani、PM Arezes、Sara Braganca、Johan Molenbroek、Iman Dianat 和 HI Castellucci。2018。面向人体工程学的成年劳动人口人体测量调查的准确性、精确度与可靠性：文献综述。《国际工业人体工程学杂志》65（2018），1–16。doi:10. 1016 2018 01 012 /j.ergon. . . Xinpeng Wang、Chandan Yeshwanth 和 Matthias Nießner。2021。SceneFormer：使用变换器生成室内场景。载于《2021年国际三维视觉会议（3DV）》。IEEE，106–115。doi:10.1109/3dv53792.2021.00021[43] Zan Wang、Yixin Chen、Baoxiong Jia、Puhao Li、Jinlu Zhang、Jingze Zhang、Tengyu Liu、Yixin Zhu、Wei Liang 和 Siyuan Huang。2024。随言而动，因景而交互：语言引导的、具备场景可供性的人体运动生成。载于《IEEE/CVF计算机视觉与模式识别会议论文集》。433 444 10 1109 52733 2024 00049 –。doi:。/cvpr . . Zan Wang、Yixin Chen、Tengyu Liu、Yixin Zhu、Wei Liang 和 Siyuan Huang。2022。Humanise：三维场景中的语言条件控制人体运动生成。《神经信息处理系统进展》35（2022），14959–14971。doi:10. 48550/arXiv.2210.09729[45] Zhen Wu、Jiaman Li、Pei Xu 和 C Karen Liu。2025。根据人类层级指令进行人-物交互。载于*IEEE/CVF*国际计算机视觉会议论文集。11176 11186 10 48550 2406 17840 –。doi:。/arXiv.。Yandan Yang、Baoxiong Jia、Peiyuan Zhi 和 Siyuan Huang。2024。Physcene：面向具身人工智能的可物理交互3D场景合成。载于*IEEE/CVF*计算机视觉与模式识别会议论文集。16262–16272。doi: 10 1109 52733 2024 01539 。/cvpr。。Yue Yang、Fan-Yun Sun、Luca Weihs、Eli VanderBilt、Alvaro Herrasti、Winson Han、Jiajun Wu、Nick Haber、Ranjay Krishna、Lingjie Liu、Chris Callison-Burch、Mark Yatskar、Aniruddha Kembhavi 和 Christopher Clark。2024。Holodeck：语言引导的3D具身人工智能环境生成。arXiv:2312.09067 [cs.CV] doi: 10 1109 52733 2024 01536 。/cvpr。。Guangyao Zhai、Evin Pınar Örnek、Shun-Cheng Wu、Yan Di、Federico Tombari、Nassir Navab 和 Benjamin Busam。2023。Commonscenes：利用场景图扩散生成符合常识的3D室内场景。神经信息处理系统进展36（2023），30026–30038。doi:10.48550/arXiv.2305.16283