Marble

——从语言到世界：生成式AI的下一前沿

西南财经大学计算机与人工智能学院 CCCI团队

2023级 谢天翊 42303054

PART 1 LLM的困境

PART 2 AI的下一前沿：空间智能

PART 3 Marble世界模型

PART 4 Marble核心能力

LLM的困境："黑暗中的文匠"

尽管以GPT-4为代表的LLM在语言处理上表现出色，但它们本质上缺乏对现实物理世界的理解，如同"黑暗中的文匠"，能精妙运用语言，却无法真正感知和交互三维世界。

缺乏空间想象力

在"心智旋转"（从不同角度想象物体）等空间推理任务中表现有限，难以理解物体在三维空间中的变换和关系。

估算能力差

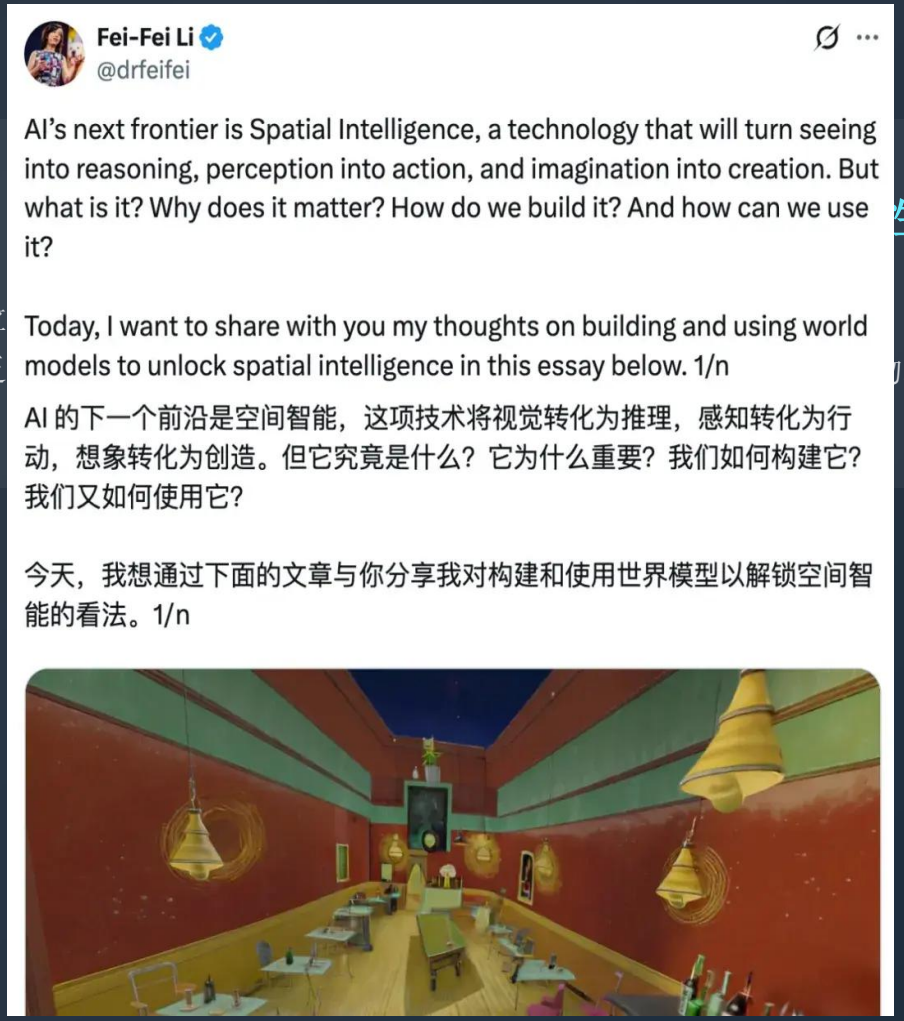
在涉及距离、方向和大小的估算佳，对物理世界的量化属性缺乏

生

因为AI

理规律

李飞飞教授指出：AI的下一步必须从"文字世界"走向"物理世界"



From Words to Worlds: Spatial Intelligence is AI’s Next Frontier

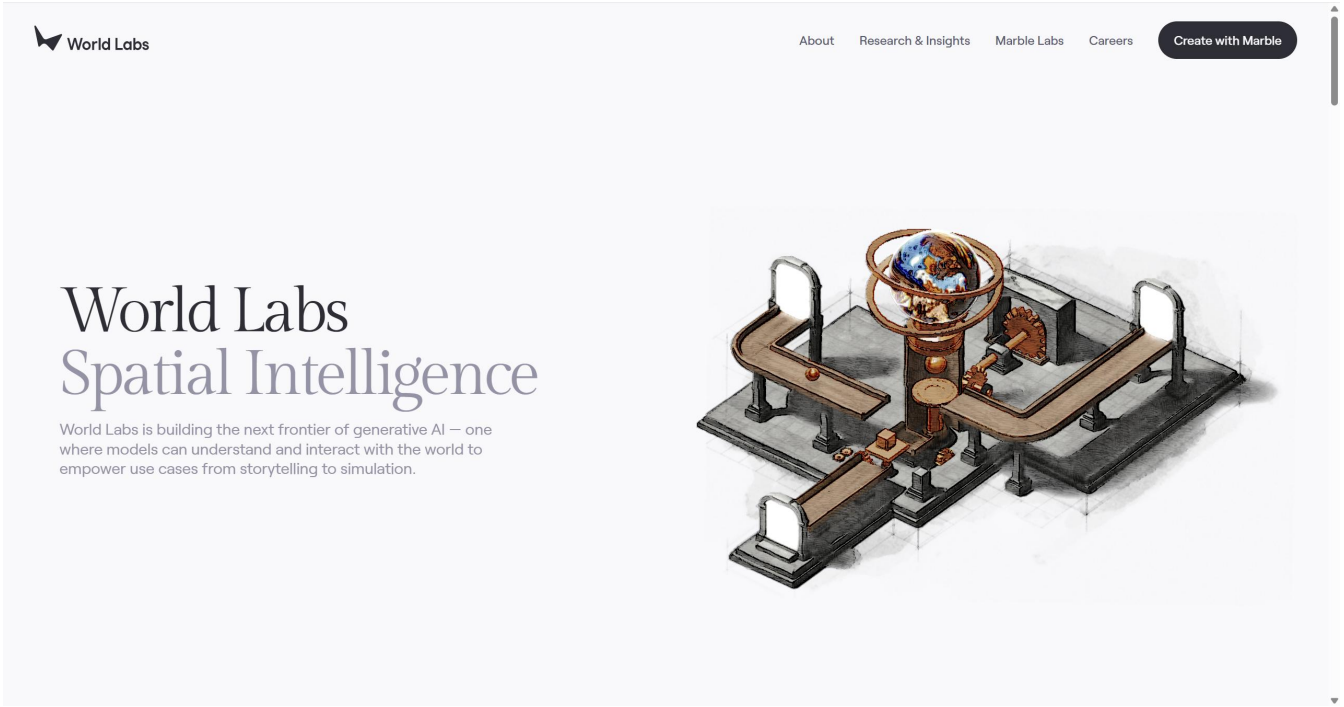
https://drfeifei.substack.com/p/from-words-to-worlds-spatial-intelligence

1950 年，当计算机还只是自动化的算术和简单逻辑时，艾伦·图灵提出了一个至今仍余音不绝的问题：机器能思考吗？他拥有非凡的想象力，看到了一个超越时代的可能 —— 智能或许可以被「构建」，而非「诞生」。这一洞见开启了一个持久而伟大的科学征程 —— 人工智能（AI）。在我投身人工智能研究二十五年后的今天，图灵的愿景仍让我心怀敬意与灵感。但我们究竟走到了哪一步？答案并不简单。

如今，以大语言模型（LLM）为代表的前沿 AI 技术，已经开始改变我们获取和运用抽象知识的方式。然而，它们依然像是「黑暗中的文匠」：能言善辩却缺乏经验，知识丰富却脱离现实。**空间智能将彻底改变我们创造和交互现实与虚拟世界的方式 —— 它将重塑叙事、创意、机器人学、科学发现等领域。这正是 AI 的下一个前沿。**

自我踏入这一领域以来，对视觉与空间智能的追求一直是我心中的北极星。这也是我花费多年时间创建 ImageNet 的原因 —— 这是首个大规模视觉学习与评测数据集，与神经网络算法和现代计算（如图形处理器 GPU）一道，构成了现代人工智能诞生的三大基石。这也是为什么我的斯坦福实验室在过去十年中，持续探索将计算机视觉与机器人学习相结合。

而这一追求，也促使我与合伙人 Justin Johnson、Christoph Lassner、Ben Mildenhall 共同创立了 World Labs—— 在一年多前，我们立志首次将这一愿景彻底实现。在这篇文章中，我将阐述什么是空间智能、它为何重要，以及我们如何构建能够释放空间智能潜力的世界模型 —— 这种能力将深刻影响创造力、具身智能*与人类的未来进步。



节选自李飞飞最新长文：AI的下一个十年——构建真正具备空间智能的机器

AI的下一前沿：空间智能

定义

空间智能将彻底改变我们创造和交互现实与虚拟世界的方式，赋予AI对物理世界的深层次理解能力。

空间智能的重要性

降低3D创作门槛

为游戏设计师、电影制作人、建筑师等提供前所未有的创造力，通过更直观、高效的方式生成和编辑3D内容。

实现"具身智能"

让机器人能够理解并安全地与物理世界互动，更好地感知环境、规划路径、执行操作。

解决训练数据难题

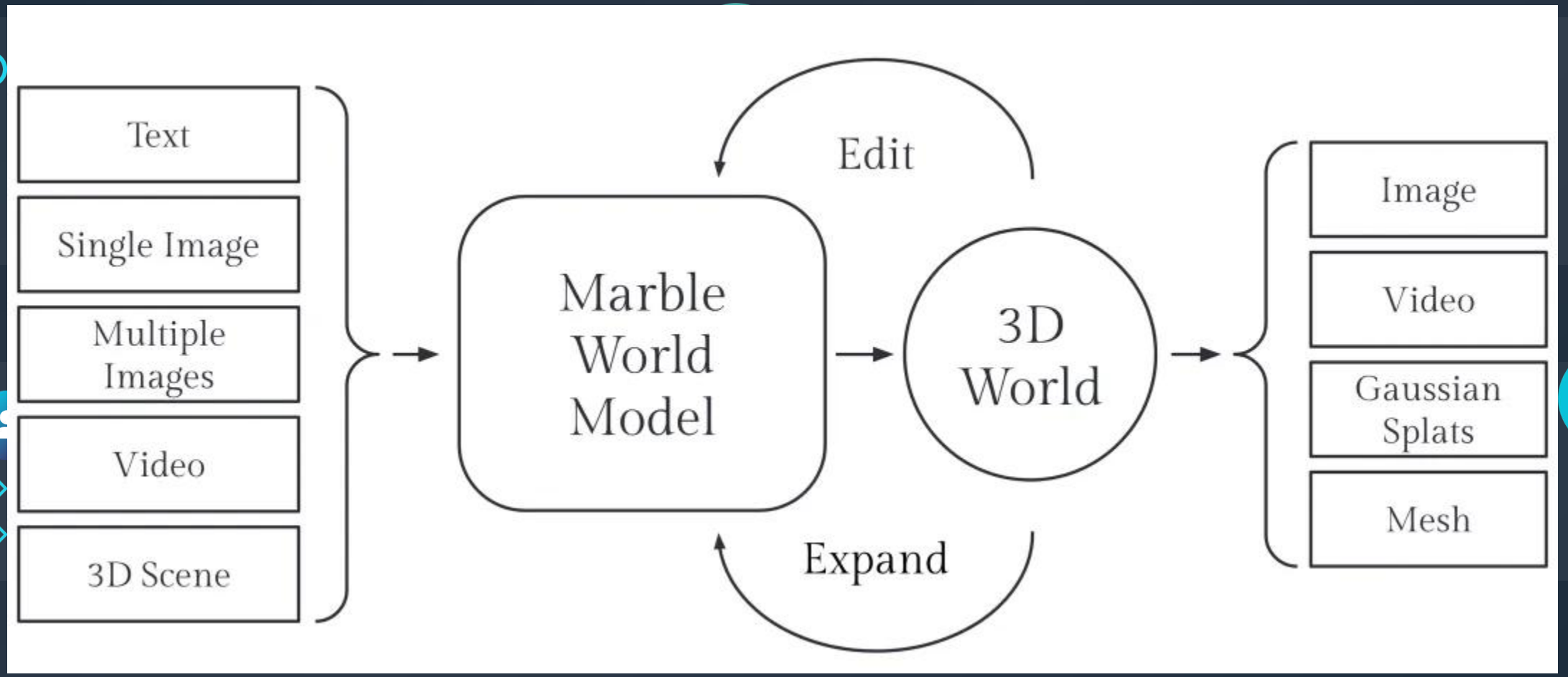
通过模拟帮助机器人在无数仿真环境中进行训练，有效解决训练数据不足的问题。

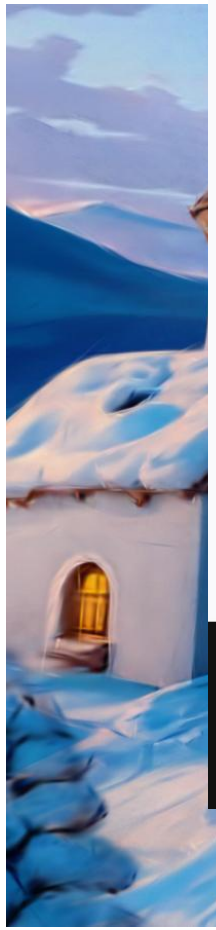
推动科学研究

模拟人类难以到达的环境（如深海、外太空）或复杂问题（如气候科学），为科学研究提供强大的虚拟实验平台。

解决方案：Marble世界模型

World Labs（由李飞飞教授创立的AI公司）推出的首款商业化产品——Marble，是一款前沿的生成式多模态世界模型，旨在将AI的能力从语言理解拓展到物理世界的感知与交互。





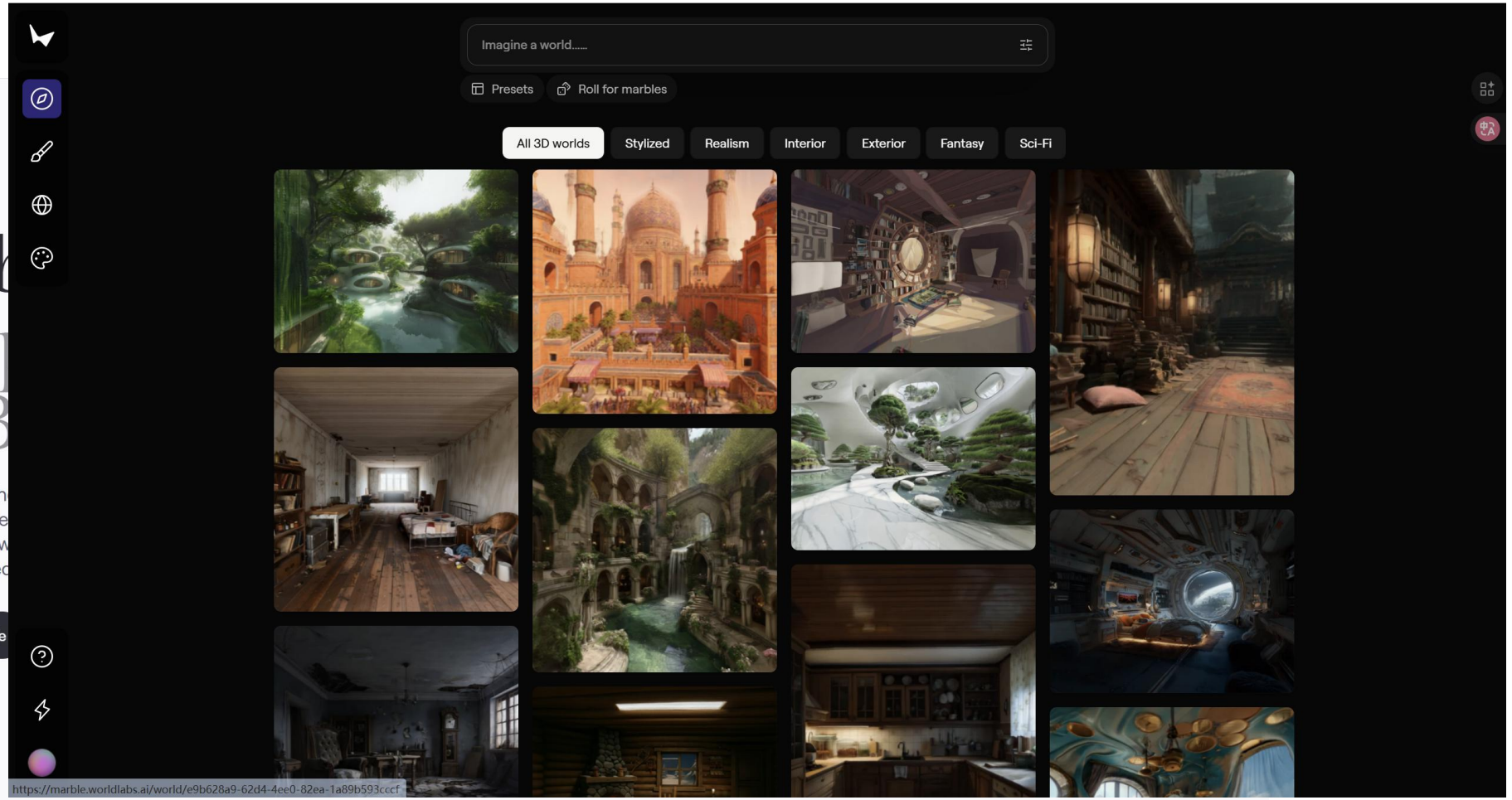
World Labs

IMAGINE
EXPLORE
SIMULATE
LEARN

Mark
Blue
For B

Marble Labs is wh
engineers, and de
case studies show
how to get started

Create with Marble



核心能力一：多模态3D世界生成

Marble通过"重度多模态"输入功能，允许用户通过多种形式的输入直接生成一个完整的、可探索的3D世界，极大地降低了3D内容创作的门槛。

文本 (Text)

通过简单的文字描述生成复杂的3D场景。例如："一个细致的、居住的霍比特人厨房"，Marble能生成包含方格地板、不锈钢配件和柔和灯光等细节的可漫游3D空间。

单张图片 (Single Image)

即使只提供一张照片，Marble也能"脑补"出图片中未显示的部分，如背面和侧面细节，从而生成一个完整的3D世界。

多张图片 (Multiple Images)

用户可上传不同角度的照片，Marble能将这些信息无缝拼接融合，构建出统一且精确的3D空间，如通过正面和侧面照片还原儿童房或会议室。

视频 (Video)

支持输入短视频片段，Marble能从中提取空间信息并生成对应的3D世界。

多模态输入

Multi-Image Prompt

Front

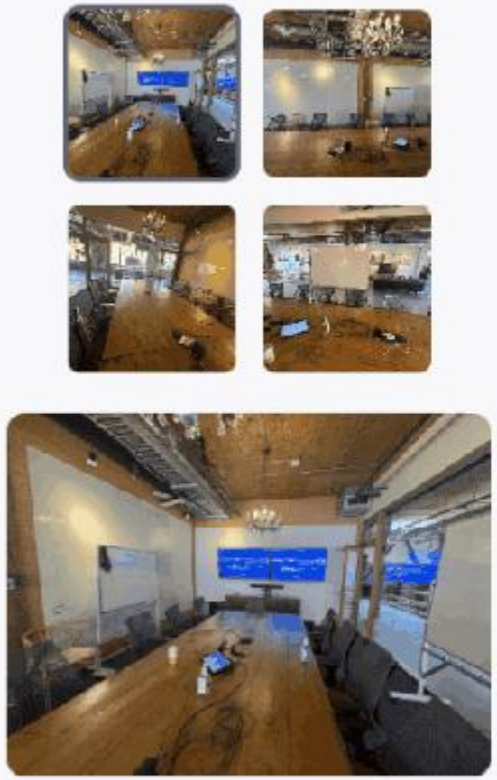
Back

Generated World

Multiple Images to World

多模态输入

Multi-Image Prompt



Generated World

核心能力二：AI原生迭代式编辑

Marble提供了AI原生的世界编辑工具，使3D世界的生成不再是"一次性产物"，而是进入了"可迭代模式"，极大提升了创作的灵活性和效率。

局部微调

用户可以对生成的3D世界进行细致的局部修改：

- ✓ 移除不想要的物体
- ✓ 调整特定区域的细节
- ✓ 更换厨房台面为黑色花岗岩等

全局调整

Marble支持对整个场景进行大规模的全局调整：

- ✓ 改变艺术风格：将现代风转换为复古风
- ✓ 结构重构：通过文字指令改变场景结构

> *"把后墙变成一个舞台，桌子换成面向舞台的长凳"*

AI生成世界 → 迭代编辑 → 再次生成 → 完善作品

核心能力三：Chisel工具——结构与风格解耦

搭建"骨架"

用简单的3D几何形状或导入的3D资产，搭建世界的粗略"骨架"。

填充"风格"

通过文字描述为"骨架"填充材质、细节和艺术风格，如"一座现代艺术博物馆，铺着木地板，摆满色彩缤纷的绘画和雕塑"。

精确控制

增强创作者控制维度，精确实现从结构到风格的完全掌控。

结构

风格

核心能力四：世界扩展与专业导出

构建更大的世界

扩展 (Expand)

选择世界某个区域，让Marble自动向外生成更多内容，或修补边缘细节，使世界变得更大、更完整。例如，当房间的边角在初始生成中不够细致时，可以通过扩展功能进行补充。

组合 (Combine)

在"组合模式"下，用户可以将多个独立生成的3D世界像搭积木一样拼接起来，从而构建出规模庞大、层次丰富的虚拟环境。

多种格式导出

高保真度输出

Gaussian Splats (高斯溅射)：通过大量半透明粒子构建3D场景，呈现逼真光影与复杂材质。

最高保真度　实时渲染

网格与视频导出

Mesh (三角网格)

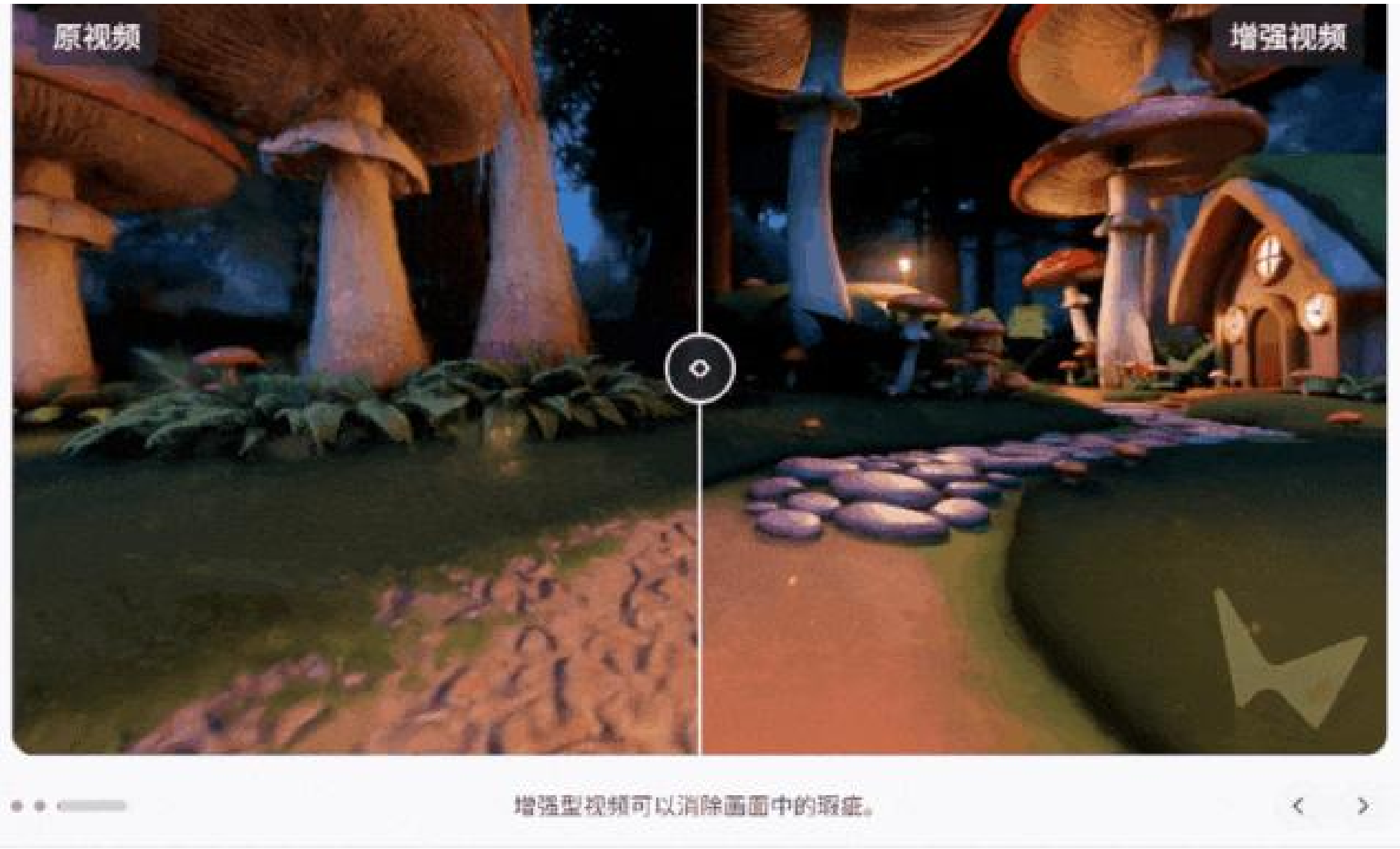
适用于Unity、Unreal Engine等平台

碰撞网格　高质量网格

视频渲染

支持精确镜头控制与动态增强

视频增强&专业导出



结语

在人类历史上，我们首次有能力构建出与物理世界深度契合的机器，让它们成为我们在应对重大挑战时值得信赖的伙伴。无论是加速我们在实验室中对疾病的理解，革新我们讲述故事的方式，还是在疾病、伤痛或衰老带来的脆弱时刻给予支持，我们正站在一项能够提升人类最珍视生活要素的技术门槛上。这是一个让生命更加深刻、更加丰盈、更加有力量的愿景。

距自然在远古动物身上首次点燃空间智能的火花，已过去近五亿年。

而我们有幸身处这样一个时代，或许很快，我们将让机器也拥有同样的能力；更幸运的是，我们能够将这种能力用于造福全人类。

如果没有空间智能，我们对「真正智能机器」的梦想就永远无法完整。


西南财经大学
SOUTHWESTERN UNIVERSITY OF FINANCE AND ECONOMICS
YI DE BUILDING

THANKS !