

World and Human Action Models towards gameplay ideation (WHAM / Muse)

Nature volume 638, pages 656–663 (2025)

Anssi Kanervisto, Dave Bignell, Linda Yilin Wen, Martin Grayson, Raluca Georgescu,
Sergio Valcarcel Macua, Shan Zheng Tan, Tabish Rashid, Tim Pearce, Yuhan Cao,
Abdelhak Lemkhenter, Chentian Jiang, Gavin Costello Gunshi Gupta, Marko Tot,
Shu Ishida, Tarun Gupta, Udit Arora, Ryen W. White, Sam Devlin, Cecily Morrison & **Katja Hofmann**
Microsoft Research Game Intelligence team

分享者: 陈宇婷

领域: 生成式人工智能

日期: 2025.3.25

Muse (基于WHAM-1.6B) 生成的游戏片段示例



世界模型：通过学习环境的动态特性来构建一个简化的模型，使得智能体能够在此模型中进行规划和决策

生成式AI在创意产业中的潜力

- **应用领域：**文本、图像、音频、音乐、视频、游戏
- **当前挑战：**迭代调整与发散性思维支持不足

游戏开发作为研究切入点

- **产业规模：**全球超30亿用户
- **独特机遇：**3D游戏开发的复杂性（多模态数据、跨学科协作）

ChatGPT的启示

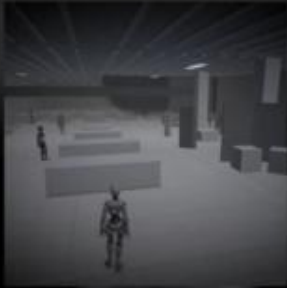
- **关键问题：**如何将生成模型应用于游戏动态建模？
- **数据优势：**《Bleeding Edge》10亿帧画面+玩家操作数据（=7年）

用户研究设计

- **面向用户：**游戏创意、开发人员
- **参与者：**8个工作室、27位创意工作者（4家独立工作室+1家3A工作室+3支游戏易用性开发团队）
- **设计探针系统：**主动式用户研究工具，将用户转化为“共同研究者”

基于Unity引擎构建的**交互式原型**，模拟生成式AI支持游戏创作的三种核心功能：

1. 自然语言修改：通过文本指令调整生成场景
2. 视觉引导生成：通过绘图编辑直接引导内容生成
3. 跨模态参考：使用图像/视频示例传递设计概念



A modern museum with a number of guards around protecting the displays.

Try This

Try Some Different Options

YOU
Create a prototype level for a new Hitman game where the player must steal from a well protected museum.

CYRUS AI
Awesome, here is one option.

What would you like to do?

Assets

Draw

Send

CYRUS AI
Ready, here are 2 options.

YOU
Add a balcony like this

CYRUS AI
Awesome, here is one option.

YOU
What strategies could help me get past the guard by the stairs?

CYRUS AI
Ready, here are 2 options.

YOU
Make the distraction longer

CYRUS AI
Ready, made the enemy distraction longer.

What would you like to do?

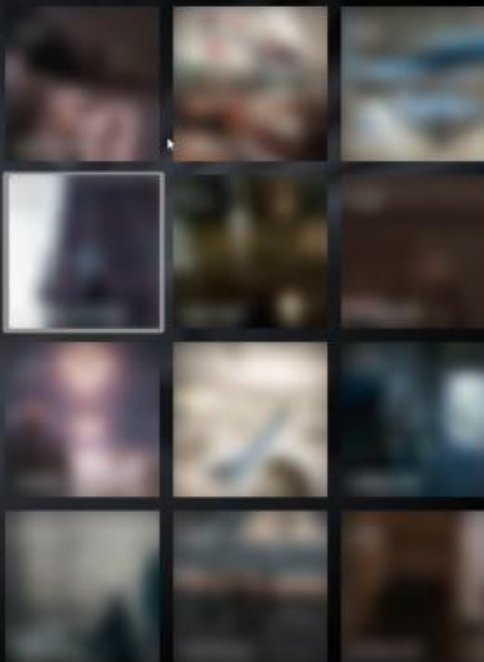
Assets

Draw

Send

Search assets

1 of 1



Click and drag on an object to change scale
Click and drag on an enemy to draw a path
Hold right mouse button to turn



CYRUS AI
Ready, here are 2 options.

YOU
Make the distraction longer

CYRUS AI
Ready, made the enemy distraction longer.

YOU
Add a grappling hook and some high platforms like in the video

CYRUS AI
Sounds good, here is one option.

YOU
Make these high platforms harder to move between.

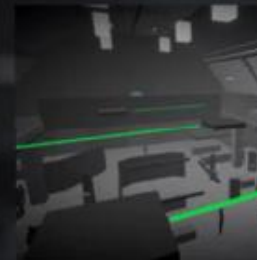
CYRUS AI
OK, here are 3 options.

What would you like to do?

Assets

Draw

Send



High platforms will move back and forth.

Try This



Rotating enemies will be placed on high platforms.

Try This



Add moving lasers to the high platforms.

Try This

Try Some Different Options

用户需求分析

用户研究设计

- **面向用户：**游戏创意、开发人员
- **参与者：**8个工作室、27位创意工作者（4家独立工作室+1家3A工作室+3支游戏易用性开发团队）
- **设计探针系统：**主动式用户研究工具，将用户转化为“共同研究者”

基于Unity引擎构建的**交互式原型**，模拟生成式AI支持游戏创作的三种核心功能：

1. 自然语言修改：通过文本指令调整生成场景
2. 视觉引导生成：通过绘图编辑直接引导内容生成
3. 跨模态参考：使用图像/视频示例传递设计概念

- **数据分析方法：**主题分析法处理转录文本，提炼出用户的两大需求方向，分别是增强工作流程（38个用例）和支持创作实践（64个用例）



用户需求分析：主题 – 示例

Theme	Example quote
Category 1: Augmenting workflows	
1 Assist with concept formation	Seek inspiration: “As game developers, sometimes we want to evoke a certain feeling or a certain atmosphere, but we’re not even sure where to start. And again, that’s just been a hole that’s been very nicely filled with AI.”
2 Assist with prototyping	Scope technical requirements: “Through very rapid prototyping, we’d be able to discover all of the things that would emerge throughout the course of development as things we’d have to go and change or address.”
3 Facilitate multidisciplinary collaboration	Encourage contribution from all disciplines: “Everybody can try out stuff that could bring something to the debate. Sometimes the issue is that people rely on someone else to (build prototypes and) try out stuff, that sometimes that costs time and sometimes it is not the exact same idea that the (person has in mind).”
Category 2: User requirements for supporting creative practice	
4 The need to help creators build a robust mental model to enable effective usage	“The one thing that’s really frustrating about generating images with AI is, if it doesn’t work, you have to go back to the starting point, redo your prompt, redo another prompt, try another prompt. It is not a fast process, and it doesn’t always have a clear path to the outcome that you’re looking for because it’s so unpredictable.”
5 The need to support iterative tweaking	“It’s hard to know what the right output is until we see it, and that, I think is one of the tough things. It takes just a lot of finessing it and playing with it.”
6 The need to support mixed modality of inputs to allow creators express their ideas in their preferred medium	“I think the mixture of being able to prompt with video and with text would be really beneficial, because sometimes, if you aren’t a programmer, it’s really difficult to actually work out exactly what you want.”
7 The need to provide multiple options to help creators explore and refine their ideas	“It could be nice to save several states of the project, so you can get back to something that you tried out at the very first and get back to this point, or maybe mixed it with the current state and do stuff like iteration mixing”

用户需求分析

用户研究设计

- **面向用户：**游戏创意、开发人员
- **参与者：**8个工作室、27位创意工作者（4家独立工作室+1家3A工作室+3支游戏易用性开发团队）
- **设计探针系统：**主动式用户研究工具，将用户转化为“共同研究者”
基于Unity引擎构建的**交互式原型**，模拟生成式AI支持游戏创作的三种核心功能：
 1. 自然语言修改：通过文本指令调整生成场景
 2. 视觉引导生成：通过绘图编辑直接引导内容生成
 3. 跨模态参考：使用图像/视频示例传递设计概念
- **数据分析方法：**主题分析法处理转录文本，提炼出用户的两大需求方向
- **研究启示：**
 - ① **现存问题：**现有模型无法支持跨模态迭代编辑，缺乏长周期内容生成的一致性保障
 - ② **关键需求：**“人类编辑-模型响应”的动态闭环；单输入生成多分支设计方案的能力
 - ③ **技术路径：**优先提升模型的一致性、多样性和持久性，开发专用评估体系

用户需求分析 -> 模型能力映射

a

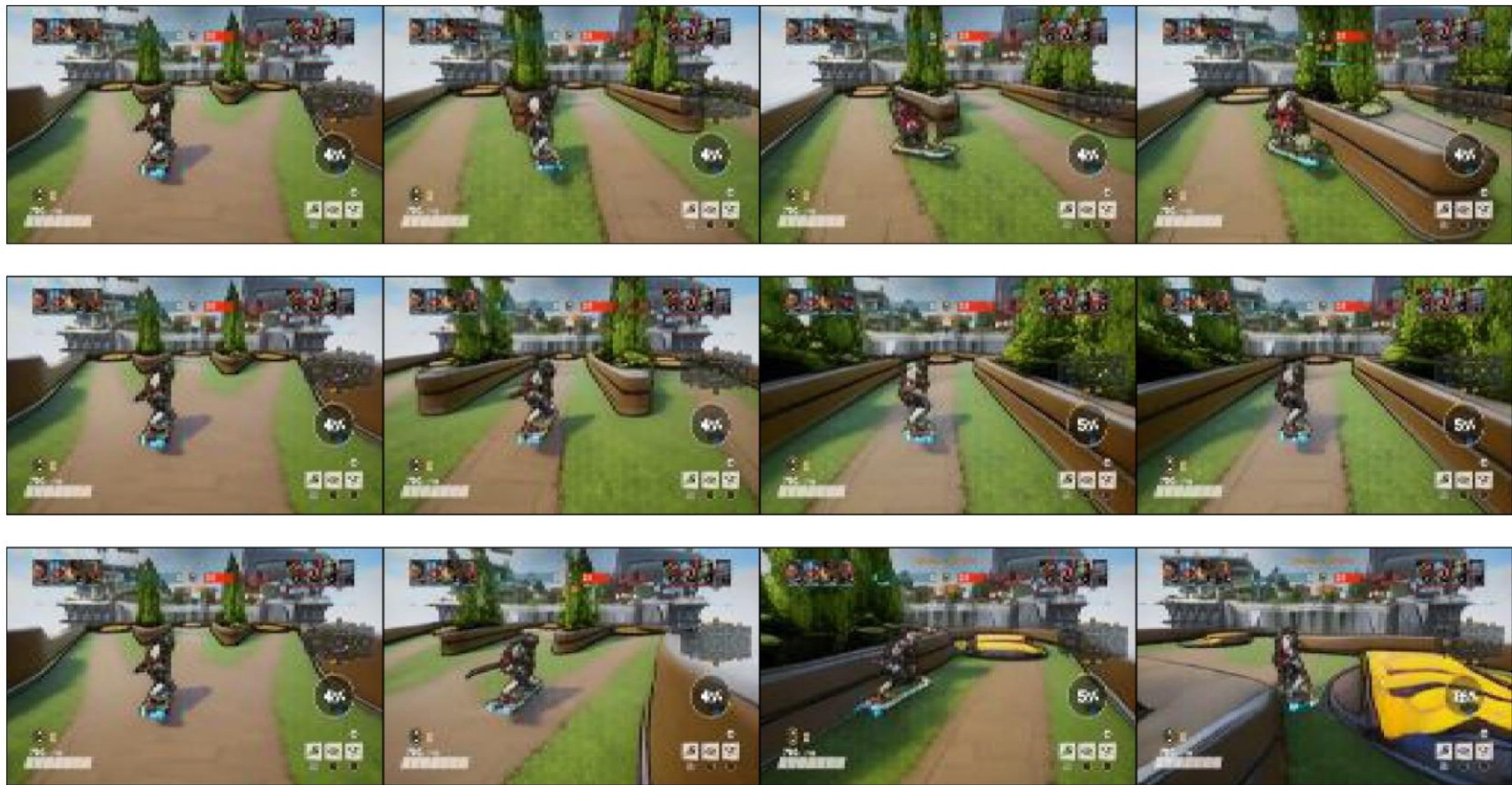


Consistency

一致性：生成序列需保持时间连贯性并符合游戏机制

用户需求分析 -> 模型能力映射

b



多样性：模型应产出反映不同潜在结果的多样化序列以支持发散性思维

用户需求分析 -> 模型能力映射

a



Consistency

b



Diversity

c



Persistency

d

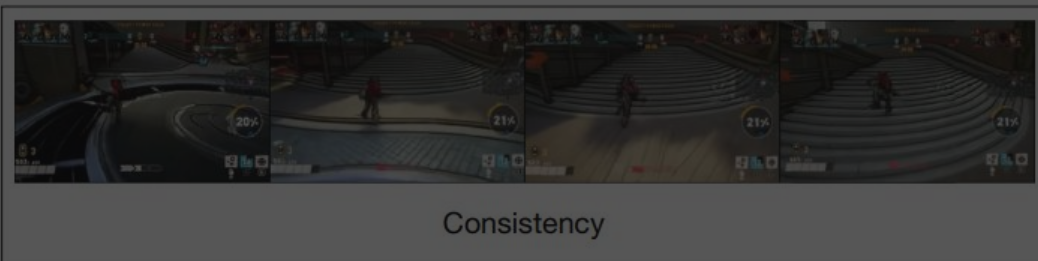


WHAM Demonstrator

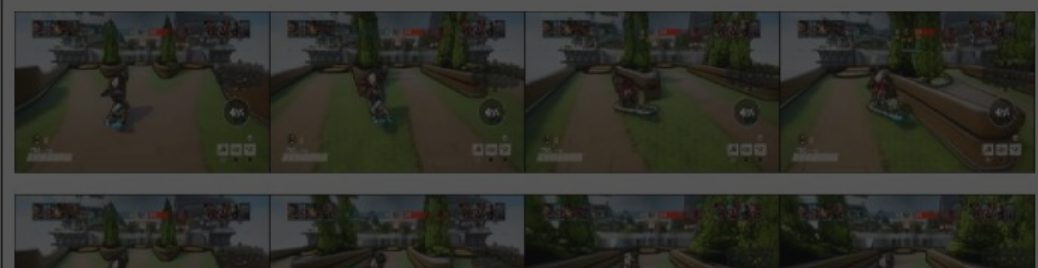
持久性：模型应持续保留用户对游戏视觉元素与操控指令的修改

用户需求分析 -> 模型能力映射

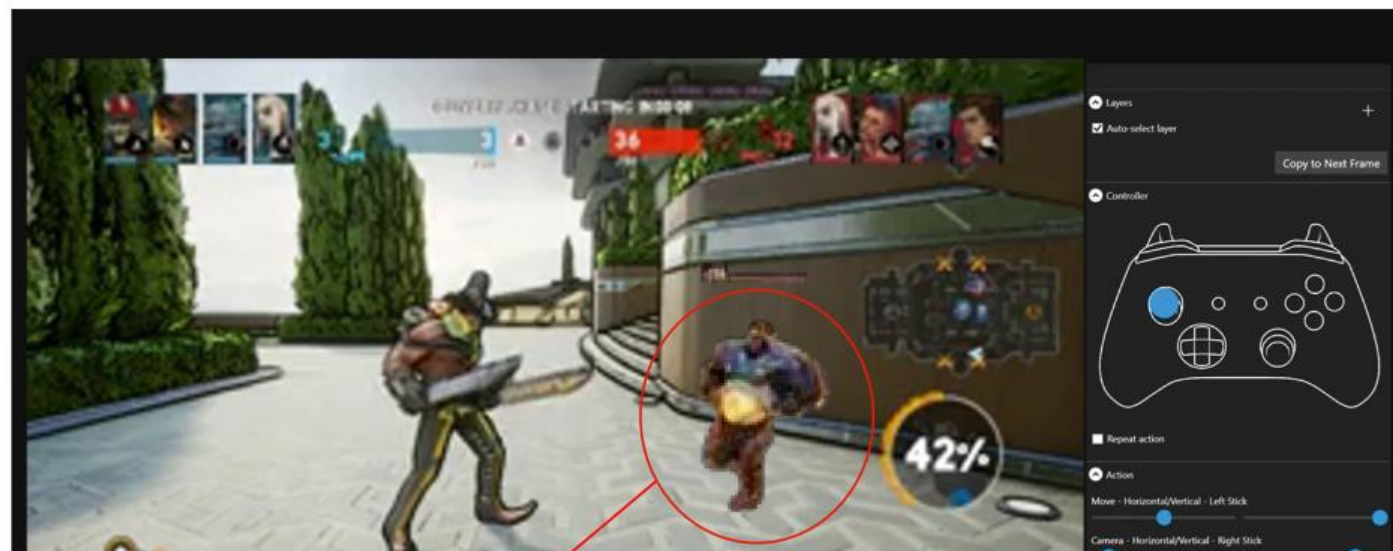
a



b



d



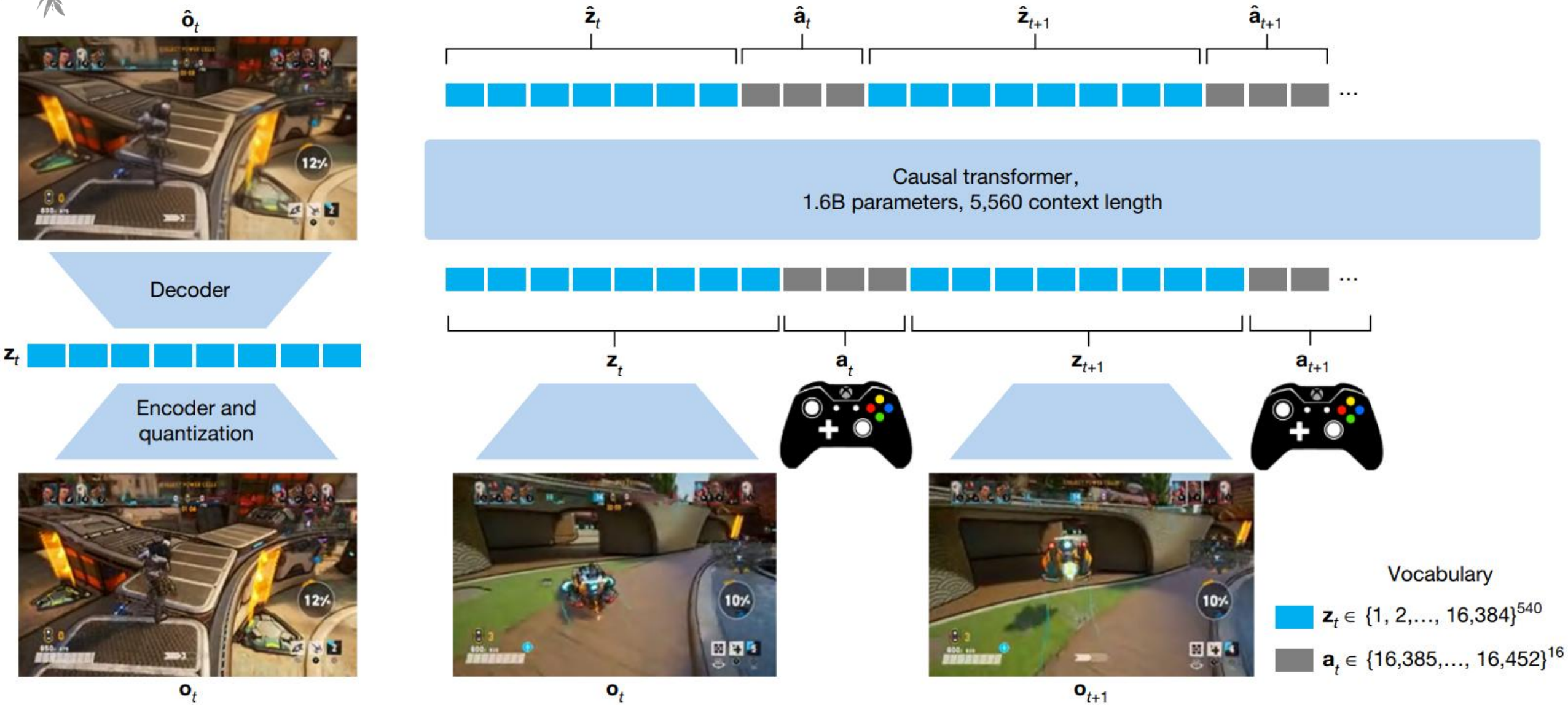
c



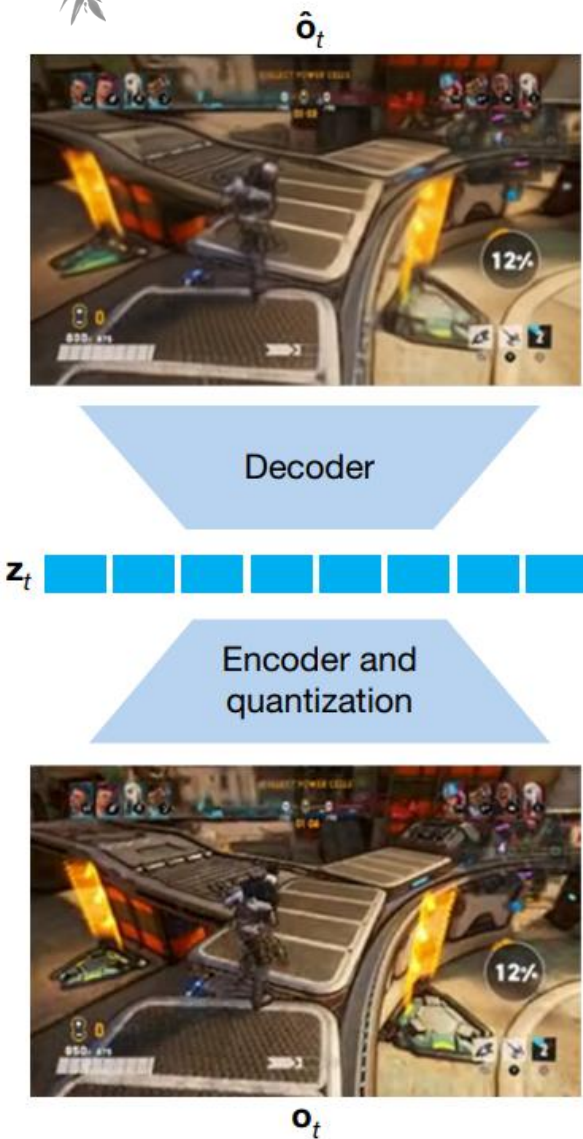
Persistence

持久性：模型应持续保留用户对游戏视觉元素与操控指令的修改

Muse模型架构



Muse模型架构



变量	定义	说明
o_t	t 时刻原始游戏帧 ($H \times W \times 3$)	300 像素 \times 180 像素 \times 3 (RGB)
\hat{o}_t	t 时刻解码重构的游戏帧	同上
z_t	t 时刻游戏帧的540个token	通过VQGAN编码器将 o_t 转换为离散token序列 z_t
a_t	t 时刻玩家操作的16个token	12个按键+2个摇杆X/Y轴分箱

压缩表示

常量	定义
16384	游戏帧码本 (codebook) 大小
16452	$-16384 = 68$, 玩家操作码本大小

按下/未按下
-> 二进制1/0

24

+

轴取值 $[-1, 1]$ 分箱成11个区间, 每一个区间对应一个码本中的值

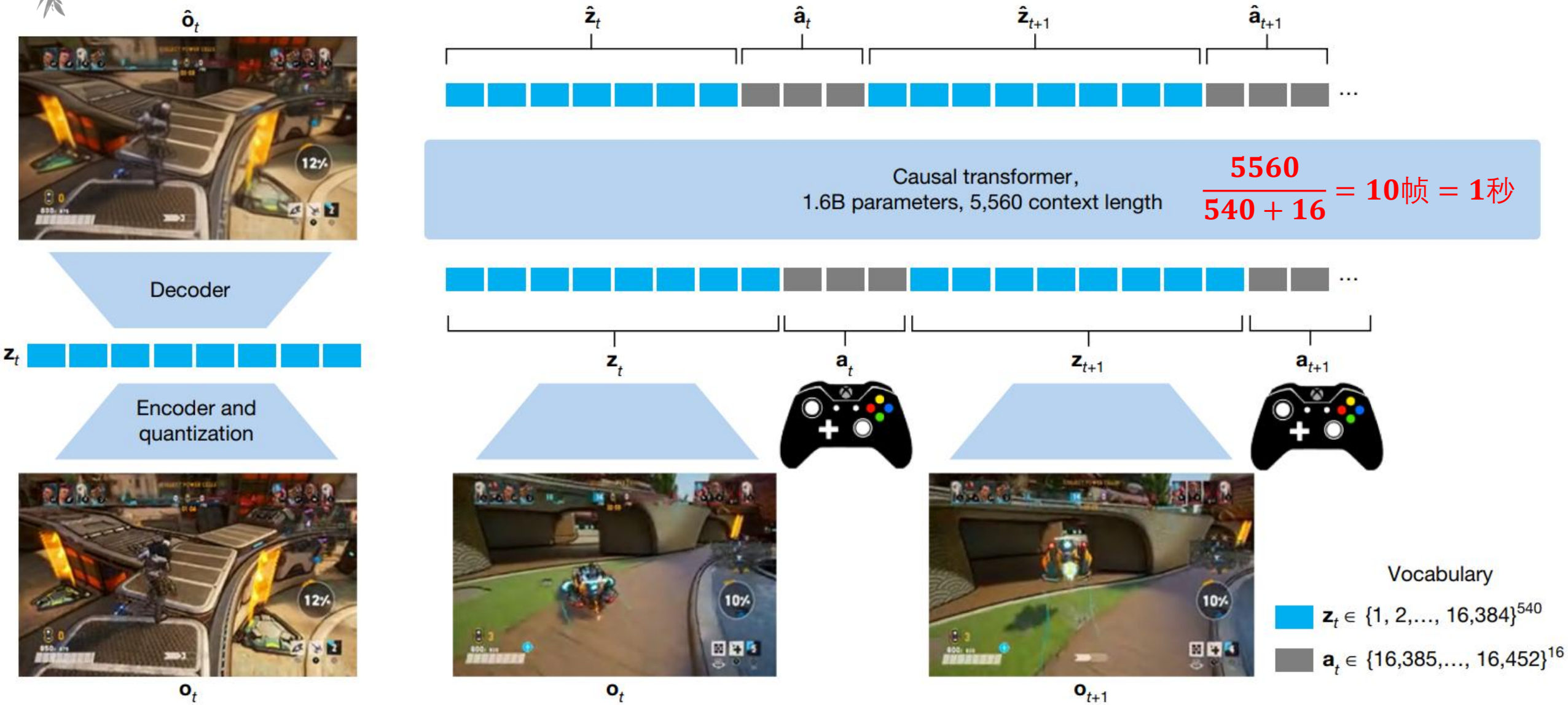
44

Vocabulary

$z_t \in \{1, 2, \dots, 16,384\}^{540}$

$a_t \in \{16,385, \dots, 16,452\}^{16}$

Muse模型架构



数据集构建

通过与Ninja Theory的合作，获取格斗游戏《嗜血边缘》的大规模玩家行动数据集

时间跨度：2020年9月—2022年10月

数据特征：视频 – H.264编码MP4格式（分辨率300×180）

操作 – 二进制时间序列（与视频帧严格同步）

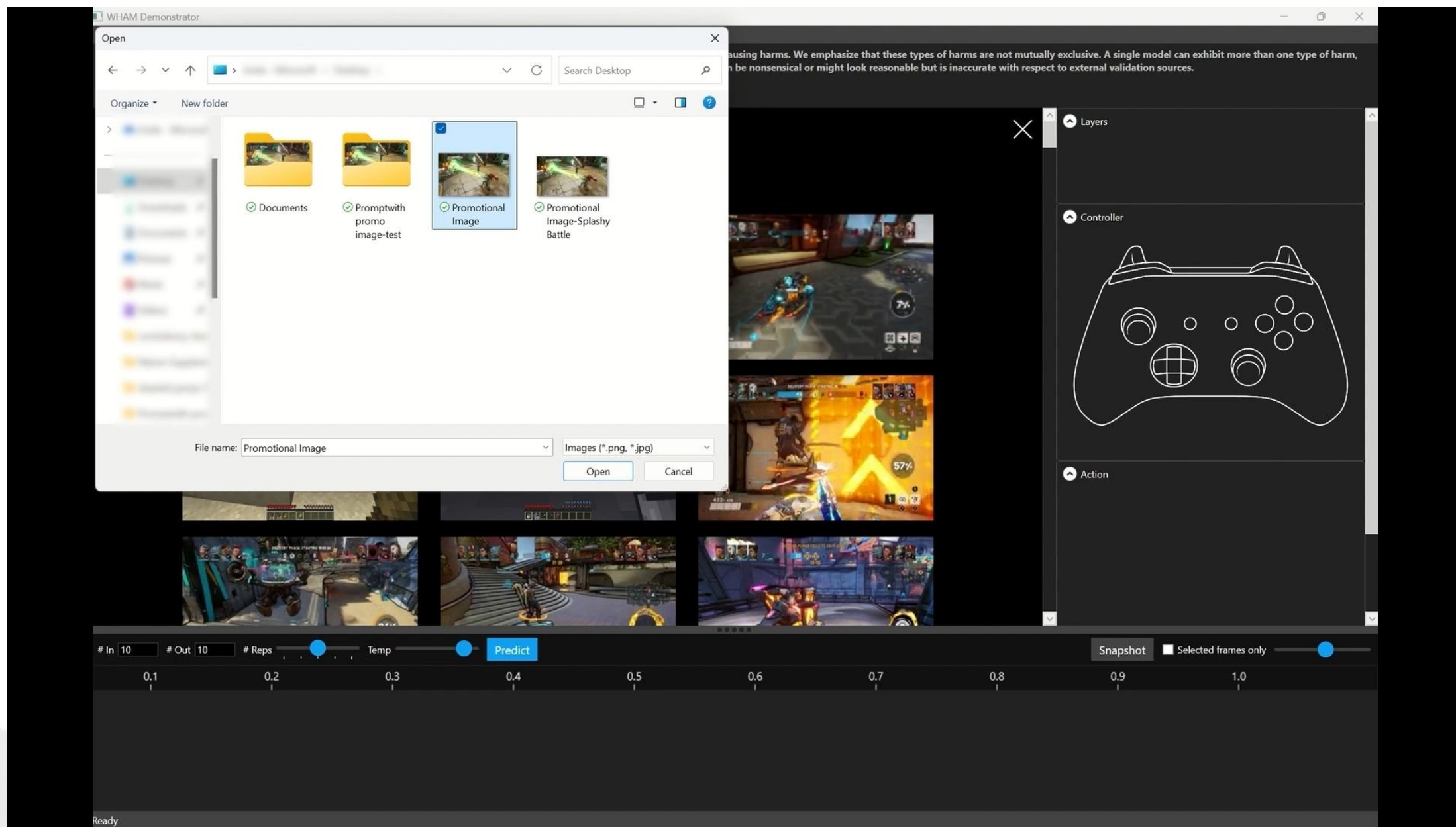
预处理：去除无效数据（非活跃玩家）、用户ID匿名化、降采样至10Hz

数据集	7 Maps基准集	Skygarden专项集
构成	7张地图的60,986场对战	单地图深度采样
帧率处理	60fps→10Hz降采样	同左
数据规模	14亿帧（27.89TiB）	3.1亿帧（6.2TiB）
时间等效	7年连续游戏时长	1年连续游戏时长
划分比例	训练80%/，验证10%/，测试10%	同左

模型训练参数

参数类别	15M - 894M WHAM	1.6B WHAM
模型规模	15M - 894M参数	1.6B参数
图像分辨率	128×128×3 (降采样)	300×180×3 (原始分辨率)
编码器类型	VQGAN卷积自编码器 (约60M参数)	ViT-VQGAN (约300M参数)
帧Token配置	<ul style="list-style-type: none">- 码本大小: 4096- 每帧token数: 256	<ul style="list-style-type: none">- 码本大小: 16384- 每帧token数: 540
玩家动作配置	12个按钮+摇杆分箱数: 68	同左
Transformer	<ul style="list-style-type: none">- 上下文长度: 2720 token (1秒)- 批次大小: 2M token- 训练步数: 170k	<ul style="list-style-type: none">- 上下文长度: 5560 token (1秒)- 批次大小: 2.5M token- 训练步数: 170k
优化器	<ul style="list-style-type: none">- AdamW ($\beta_1 = 0.9, \beta_2 = 0.999$)- 固定学习率0.00036 (线性预热)	<ul style="list-style-type: none">- AdamW ($\beta_1 = 0.9, \beta_2 = 0.95$)- 余弦退火学习率
计算资源	V100集群 (100 GPU)	H100集群 (更高计算效率)

WHAM演示器 – 与Muse模型交互的可视化接口




WHAM演示器 – 与Muse模型交互的可视化接口

WHAM Demonstrator

File Export

AI Generated Content: Models trained using game data may potentially behave in ways that are unfair, unreliable, or offensive, in turn causing harms. We emphasize that these types of harms are not mutually exclusive. A single model can exhibit more than one type of harm, potentially relating to multiple different groups of people. For example, the output of the model can be nonsensical or might look reasonable but is inaccurate with respect to external validation sources.

Service Uri: [REDACTED]



1200x720


Anti-aliasing

Layers

Auto-select layer

Copy to Next Frame

Controller



Action

In 10 # Out 10 # Reps Temp Predict

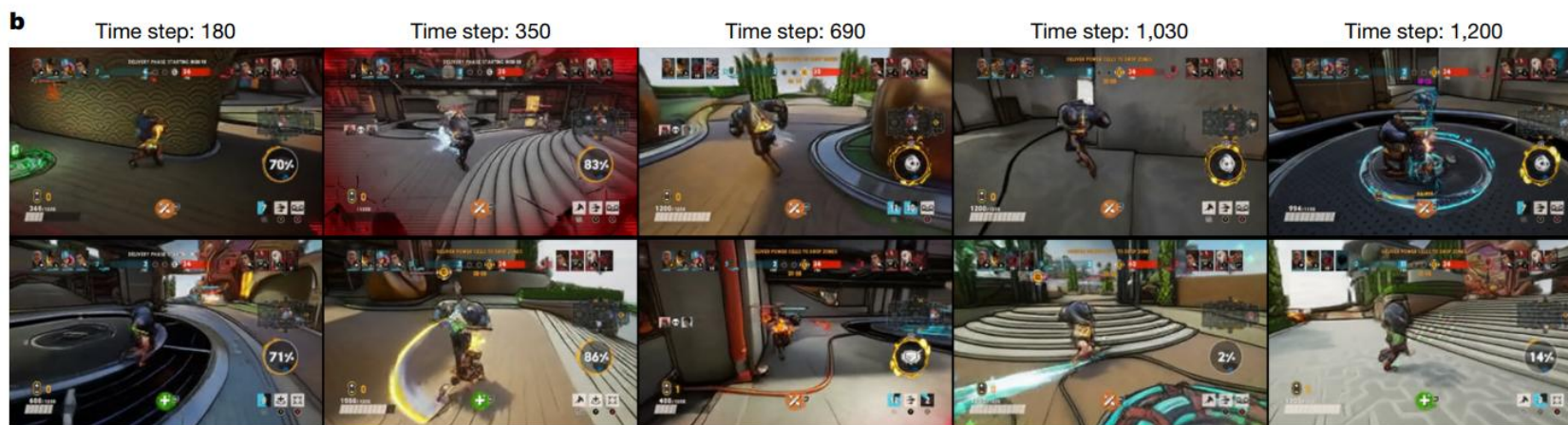
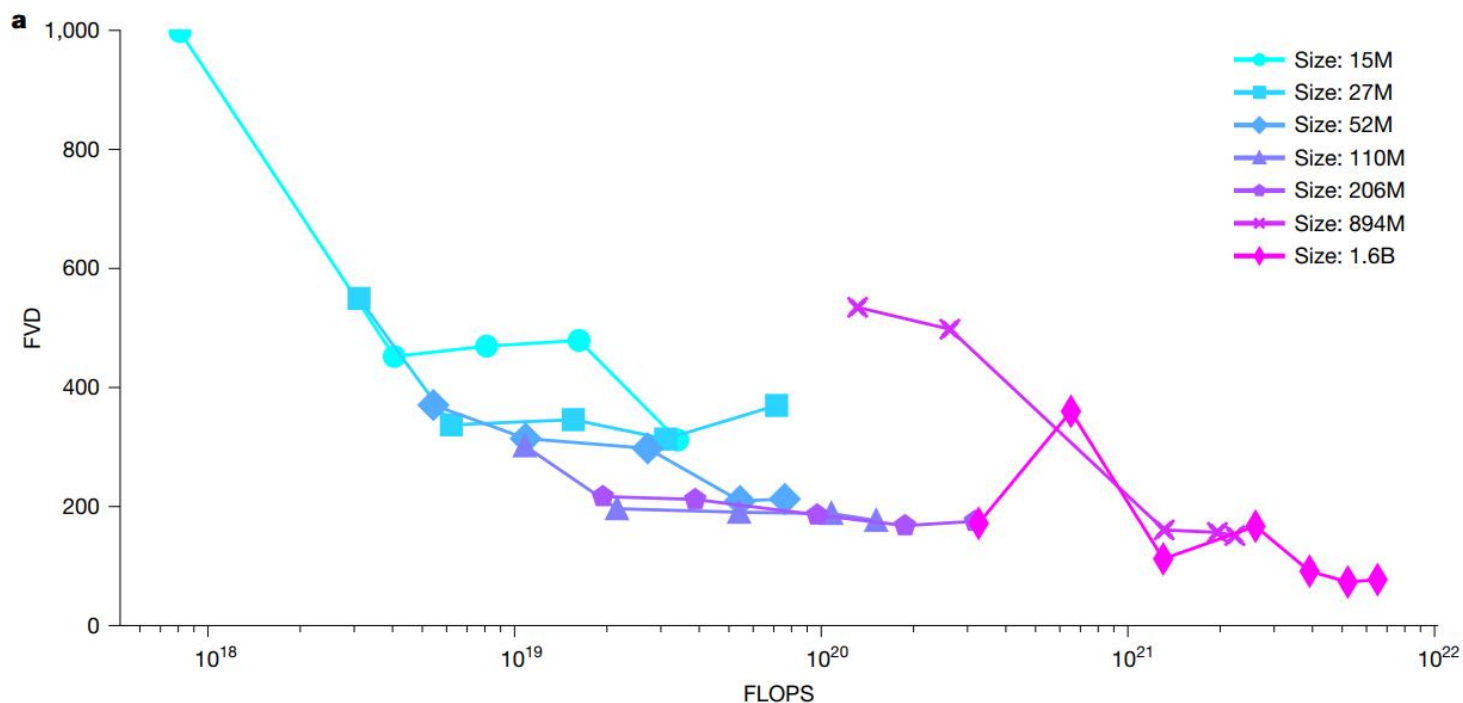
3.1 3.2 3.3 3.4 3.5 3.6 3.7 3.8 3.9 4.0 4.1 4.2 4.3 4.4 4.5 4.6 4.7 4.8 4.9 5.0 5.1

Snapshot Selected frames only

Ready

实验设计与评估指标 – 一致性评估

- 评估方法：生成序列与真实视频的Fréchet视频距离 (FVD)
- 评估结果：1.6B模型FVD最低
- 生成效果：1.6B的WHAM可以生成长达2分钟的连贯游戏序列



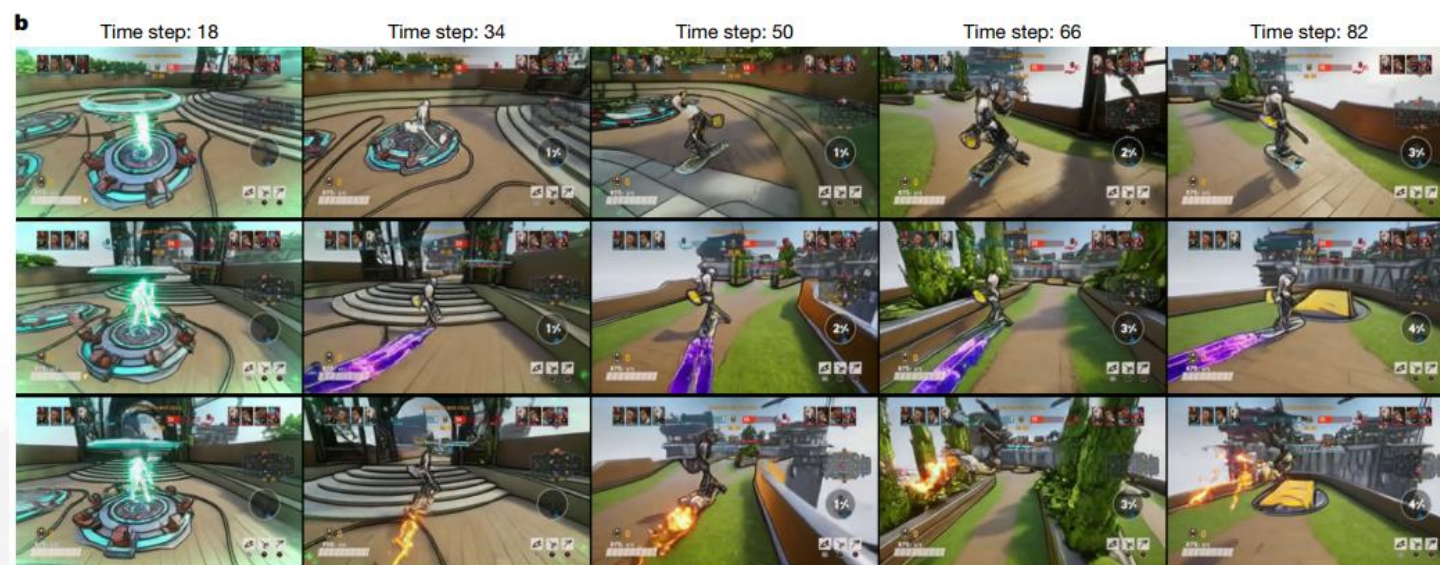
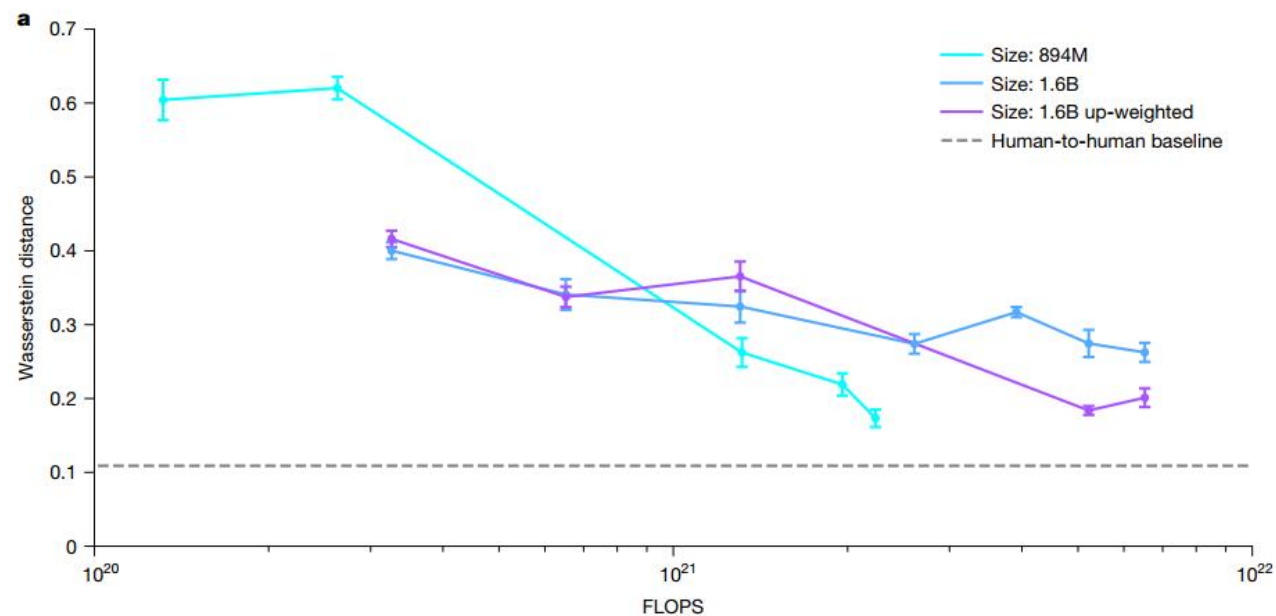
实验设计与评估指标 – 一致性评估

- 评估方法：生成序列与真实视频的Fréchet视频距离（FVD）
- 评估结果：1.6B模型FVD最低
- 生成效果：1.6B的WHAM可以生成长达2分钟的连贯游戏序列



实验设计与评估指标 – 多样性评估

- 评估方法: Wasserstein距离
生成动作分布 vs 人类玩家分布
- 评估结果: 动作损失加权后接近人类基线
- 生成效果: 行为多样性 + 视觉多样性

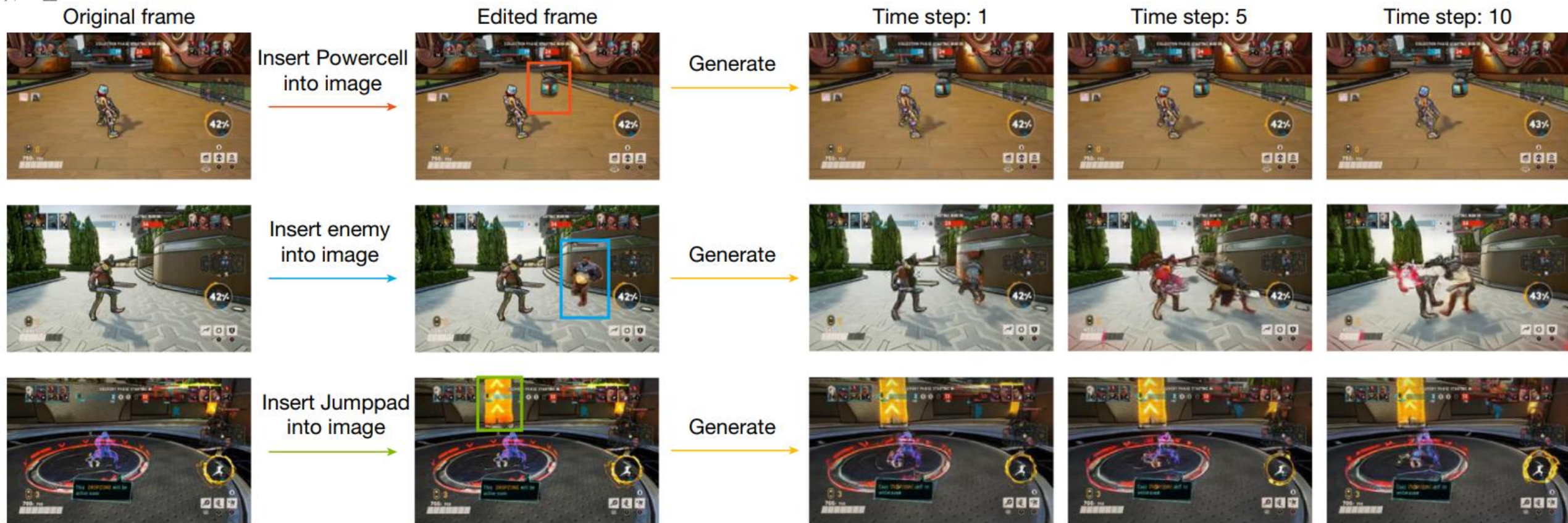


实验设计与评估指标 – 多样性评估

- 评估方法: Wasserstein距离 – 生成动作分布 vs 人类玩家分布
- 评估结果: 动作损失加权后接近人类基线
- 生成效果: 行为多样性 + 视觉多样性



实验设计与评估指标 – 持久性评估



- 评估方法：人工标注

插入三类新元素后生成10秒序列

游戏道具：能量核心

玩家角色：友方/敌方

地图组件：垂直弹射器

实验设计与评估指标 – 持久性评估

- 评估结果：5帧条件设置下持久性成功率超85%，
元素类型与初始位置影响持久效果

Table 1 | Quantitative persistency results

Conditioned frames	Powercell	Character	Vertical Jump pad
1	58%	45%	58%
5	86%	85%	98%





World and Human Action Models towards gameplay ideation

Supplementary Video

Microsoft Research Game Intelligence & Teachable AI Experiences¹,

Anssi Kanervisto¹, Dave Bignell¹, Linda Yilin Wen¹, Martin Grayson¹, Raluca Georgescu¹, Sergio Valcarcel Macua¹, Shan Zheng Tan¹,
Tabish Rashid¹, Tim Pearce¹, Yuhan Cao¹

Abdelhak Lemkhenter¹, Chentian Jiang³, Gavin Costello², Gunshi Gupta⁴, Marko Tot⁵, Shu Ishida⁴, Tarun Gupta¹, Udit Arora¹

Ryen W. White¹, Sam Devlin¹, Cecily Morrison¹, Katja Hofmann¹

¹Microsoft Research. ²Ninja Theory. ³University of Edinburgh, work completed while at Microsoft. ⁴University of Oxford, work completed while at Microsoft. ⁵Queen Mary University of London, work completed while at Microsoft.

学习过程中的疑惑

- 《Bleeding Edge》是一款3D游戏，加入的元素为什么可以在后续生成过程中保持3D状态？
- 通过隐式学习3D游戏动态生成的2D帧序列，在视觉上模拟3D效果，而非显式构建3D模型
与传统方法的对比：不同于静态贴图生成3D模型，WHAM更注重时序连贯性与交互逻辑的合理性
适用场景：快速生成游戏流程原型、自动化测试用例，或为设计者提供灵感，但无法替代3D建模

核心贡献

- **技术突破**：首个支持3D游戏动态的生成模型
- **开源价值**：
 - ① **模型权重**：包含完整训练参数，支持直接加载生成游戏序列，无需从头训练；提供不同规模模型（1.6B/200M）适应不同硬件条件。
 - ② **评估数据集**：完整7 Maps数据集因商业原因未公开，仅提供单地图（Skygarden）样本。
 - ③ **WHAM演示器**：支持通过图像或操作序列提示模型生成；允许用户直接编辑帧内容（如添加角色），并观察修改如何影响后续生成。

- **高分辨率：**300×180不足以支撑当前的游戏市场，需要优化结构或者投入更多硬件设备，生成更高质量的游戏画面。
- **更新现有游戏：**简化开发流程，提高效率，让创意者和开发者专注于真正需要人类干预的部分。
- **实时生成画面：**模型根据用户操作实时扩展，做出符合逻辑的反应并与环境正确交互。
- **让老旧游戏重焕生机：**原硬件过时，理论上Muse能通过游戏数据学习整个游戏，从而在无需原引擎的情况下，在新平台重新渲染。



欢迎探讨

分享者：陈宇婷

领域：生成式人工智能

日期：2025.3.25