豪泓科技
HAOHONG TECHNOLOGY

H2 AIDM · L1 世界模型 · 白皮书 03

RayWorld
相机，成了参数。

R1 离线世界模型 × Live 实时直播模型。把相机位姿编成视线参数，作为条件叠进注意力——相机成了请求里的参数。

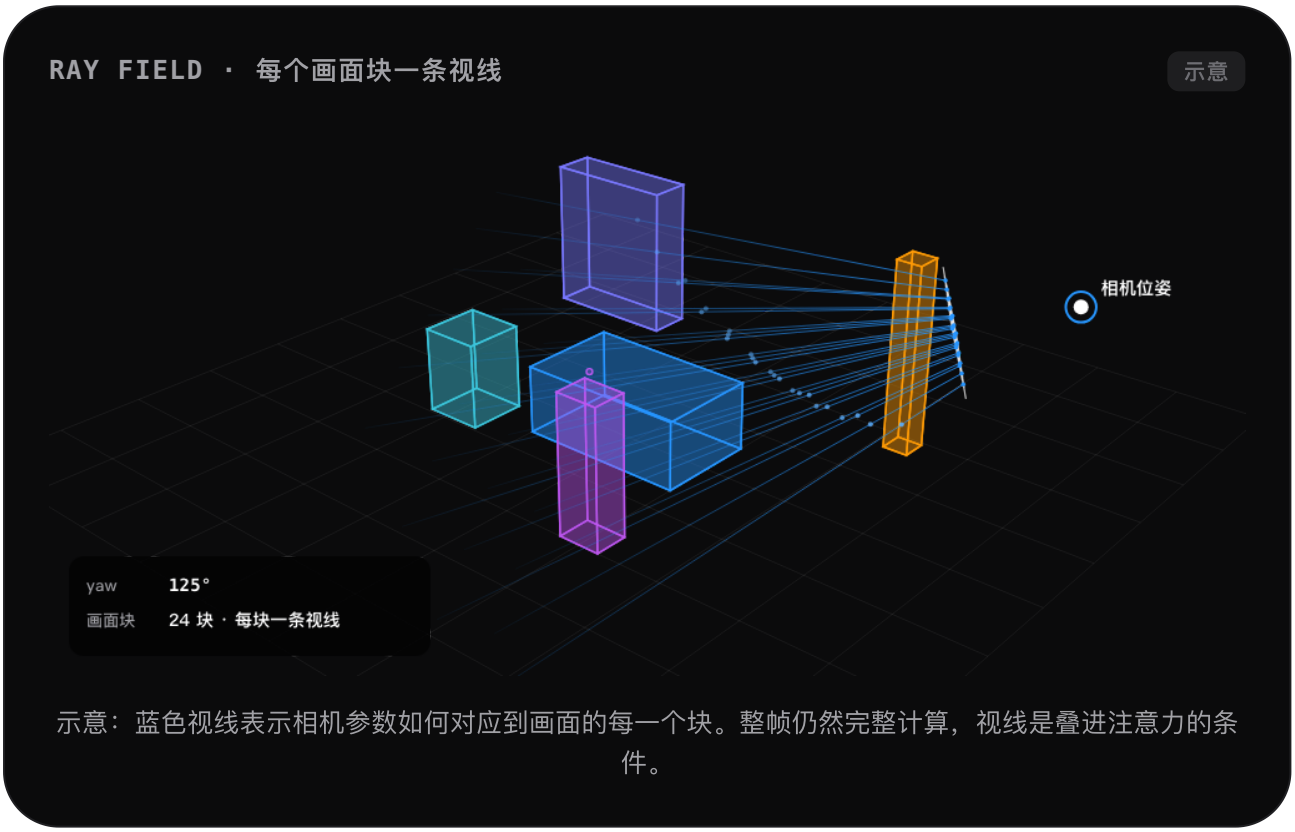
上海豪泓信息技术有限公司 | v5.1 | 2026 年 11 月

数字口径：实测 / 生产实证 / 已上线 / 规划测算，见各处标注。规划值以实际运行数据为准。

RayWorld R1 离线 × Live 实时

相机，成了参数。

常规视频模型逐帧猜像素，大幅转镜时容易不一致。RayWorld 把相机位姿作为条件叠进注意力，镜头成了可传、可复现的参数。



示意：蓝色视线表示相机参数如何对应到画面的每一个块。整帧仍然完整计算，视线是叠进注意力的条件。

0.07 元/秒

RayWorld R1 768p 单秒价格；在小影马上即每秒 1 积分。

已上线

1/14

约为主流同类（约 1 元/秒）的 1/14。

已上线

2 路 · 25fps

RayWorld Live 单张 RTX PRO 4500 并发直播，0 掉帧。

生产实证

2K

R1 单段 4–15 秒，最高 2K，可直接带声成片。

已上线

第三条路

不赌像素，
也不仿真整个世界。

常规视频模型没有显式的几何条件；全量 3D 仿真算力高，也出不了成片。RayWorld 走第三条路。

「视线场（Ray Field Conditioning）」：把相机位姿编码成与画面块对齐的视线参数，作为条件叠进 DiT 的注意力——模型每一帧都知道镜头从哪里、朝哪里看。

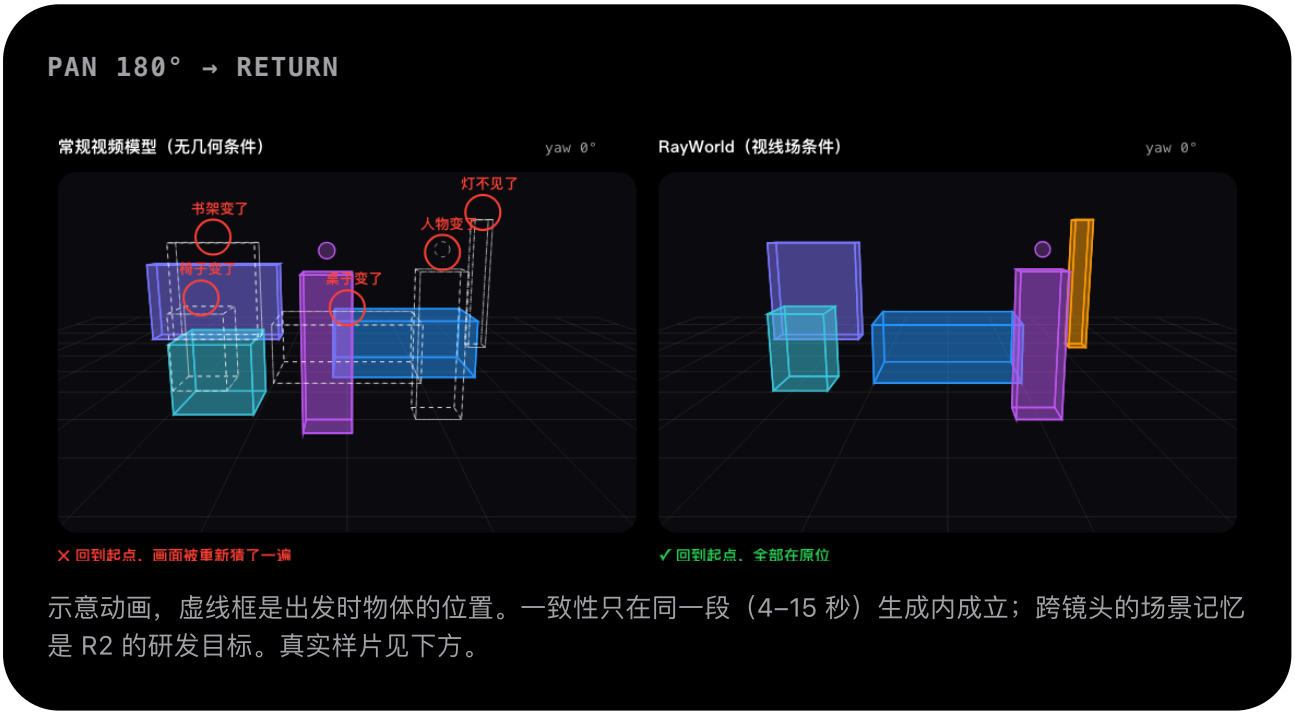
维度	常规 DiT 视频模型	全量 3D 仿真 / 重建	RayWorld
空间一致性	没有几何条件，大幅转镜时容易不一致	强（显式网格）	**强（视线场几何条件，同一段生成内）**
相机控制	文字粗描述，不可复现	完全可控	**相机位姿与路径作为请求参数，可被智能体调用**
推理算力 行业观察	1×	10–100×（机房级集群）	**与常规 DiT 同量级（约 1×），单张 RTX PRO 4500 可运行**
输出形态	视频	可漫游空间，不是成片	**直接交付带声成片，最高 2K**

维度	常规 DiT 视频模型	全量 3D 仿真 / 重建	RayWorld
价格（768p）	约 1 元/秒	订阅制，并发难	约 0.07 元/秒 已上线

回望测试

同一段生成里，
镜头转开再转回，东西还在原位。

相机转身 180° 再回到起点。没有几何条件，模型只能重新猜一遍画面；RayWorld 每一帧都带着同一套相机视线。



示意动画，虚线框是出发时物体的位置。一致性只在同一段（4–15 秒）生成内成立；跨镜头的场景记忆是 R2 的研发目标。真实样片见下方。

真实样片 · RayWorld R1

不是动画，是 R1 直接生成的画面。

每段都是一次生成里的极限运镜，看镜头剧烈运动时空间是否还成立。





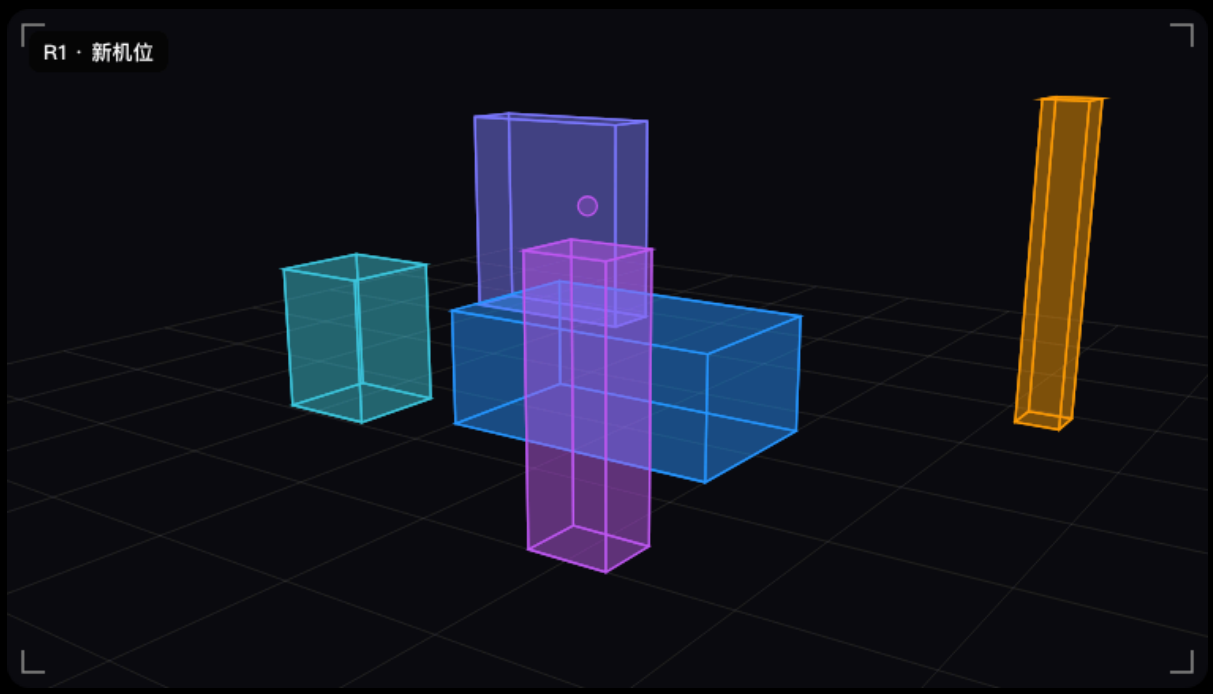


样片来自 rayworld.ai，每段约 5 秒，已压缩为网页预览。

RayWorld R1 · 离线世界模型

智能体写不好「镜头往右一点」，但能写准一行 JSON。

相机位姿（位置 xyz + 朝向 yaw / pitch / roll）成为请求参数：yaw ±180°、pitch ±40°，支持 roll；四种原生模式，可传参、可复现。H2 Harness 与小影马直接调用这个接口编排镜头。



reshoot
锁场景，换机位重拍

gimbal
云台平移与俯仰

path
环绕、推进、升降

multicam
同一时刻多机位

```
{
  "mode": "reshoot",
  "camera": {
    "position": [4.7, 2.7, 6.9],
    "rotation": { "yaw": -35, "pitch": -15, "roll": 0 },
    "fov": 50
  }
}
```

示意交互：线框按参数实时投影，右侧是智能体发送的同一结构 JSON。R1 同时支持全模态参考——文字、首尾帧、最多 9 张图、3 段视频、3 段音频，都可以和相机参数叠加。

怎么做到的

在开源基座上做结构改造，分三步。

方法由豪泓在 RayWorld R1 上首先落地，再沉淀为 H2MC 的后训练服务，同样用来把企业的业务约束写进模型。

01 · 拆解

找到最小改动面

逐层探查梯度与表征敏感度，只改必要的那一层。改大了，原有能力会崩掉。

02 · 接线

把相机位姿接进注意力

视线参数与画面块对齐，作为条件叠进注意力。接错层，条件就不生效。

03 · 保能力

逐项回归，压回单卡

计数、相对位置、文字稳定、极限旋转逐项回归，再用步数蒸馏与量化把成本压回单卡。

RayWorld Live · 实时直播模型

一张照片，两种开法。

两路交替占用同一张卡，每路每 40 毫秒出一帧（25fps）。相机作为每帧的条件输入，再叠加时空注意力稀疏化、步数蒸馏与帧间潜变量复用，把单帧计算压进预算。

模式一

数字人实时互动直播

上传一张照片，装入人设与企业知识库。冷场主动暖场、送礼即时感谢、弹幕提问实时解答，口型与微表情同步。三步推流到抖音、视频号、小红书。

适合：品牌带货、本地生活、企业客服与科普直播间。

模式二

互动剧情直播

观众的弹幕、点赞与投票实时变成相机指令和剧情分支，边看边改镜头与走向，生成连续不穿帮的剧情流。

适合：互动影游直播、虚拟偶像、文旅实时导览、沉浸式发布会。

演示直播间仍在持续调试与升级，画面和功能可能随时变化；如暂时打不开，请稍后再试。

25 fps

实时出流，每路每 40 毫秒出一帧。

生产实证

2 路

同一张 RTX PRO 4500（32GB · 200W）并发，两路交替占用 GPU。

生产实证

0 掉帧

两路长时间运行。

生产实证

流式

每一帧依据当前相机参数生成，帧间复用潜变量。

已上线

推理成本 已上线

成本，
是工程做出来的。

价格由成本结构决定：H2GC 把模型放进单卡，H2MC 用蒸馏与量化压下步数和精度。

清晰度	单价	15 秒一段	适合
480p	0.03 元/秒	0.45 元	分镜预演、批量试拍、智能体自动草稿
768p	0.07 元/秒	1.05 元	短剧 / 漫剧量产、广告电商、短视频成片（约为同类 1/14）
2K	0.20 元/秒	3.00 元	影视预演、高端广告、合成数据

RayWorld 在小影马上按积分售卖：1 积分 = 0.07 元，768p 每秒 1 积分。在 vingma.tv 使用 ↗



技术路线

已上线两个，在做两个。

R1 · 离线世界模型

已上线

视线场几何条件，单卡商用，四种相机模式。

Live · 实时直播

生产实证

单卡两路 25fps，一张照片两种开法。

R2 · 场景记忆

研发中

分钟级连续镜头，跨镜头记住场景与角色状态。

R3 · 空间智能底座

远期规划

焦距、景深、物理动力学可控；多智能体共享持久化世界。

在 H2 AIDM 中的位置

下面两层把成本压到单卡，上面两层把它用起来。

↓ L0 算力 · 托起

H2 GPU Combo

算子融合、FP4 量化与 PCIe 5.0 预取：同一张 RTX PRO 4500 可并发两路 Live 直播，也可单独运行 R1。

了解 H2GC ›

↔ L1 模型 · 同层协同

H2 Model Combo

后训练方法由豪泓在 R1 上首先落地，再沉淀为 H2MC 的后训练服务；步数蒸馏、量化与统一网关计费由 H2MC 提供。

了解 H2MC ›

↑ L2 智能体 · 调用

H2 Harness

向 R1 下发相机参数 JSON，向 Live 推送话术与互动分支——世界模型成了智能体可控的镜头。

了解 H2 Harness ›

↑ L3 应用 · 落地

小影马与小影马直播

短剧镜头不穿帮、互动影游多分支渲染、小影马直播 25fps 出流，核心都是 RayWorld。

了解小影马 ›