结构约束先行的视觉生成：AI 绘画与视频生成的三层混合逻辑

YunFei M. Wang
（王·Micheal·雲飛）
Independent Researcher

摘要

当前 AI 图片与视频生成主要依赖位图生成逻辑，在局部细节填充上表现优异，但在宏观结构一致性、物理合理性与帧间连贯性上存在结构性缺陷。本文提出一套三层混合生成逻辑的理论框架：第一层构筑三维矢量结构，约束具有长期一致性但随剧情产生局部变化的全局元素；第二层以二维投影逻辑约束镜头视角、光照、明暗、阴影、物理碰撞等单帧空间关系与因果连贯性；第三层以位图逻辑填充单帧局部细节。本文阐述该框架的理论依据——矢量图与位图的知识模式区分，以及从计算模拟到 AI 生成的联想过程，并论证其与 3D 设计软件及 3D 动画工作流程的对应关系。本文属于方法论研究，不包含实证数据。本文为上篇，聚焦理论框架与技术架构；下篇《从三层混合逻辑到生成产线：AI 视觉生成的工程实现与产业迁移》将讨论生产流程、算力生态与产业应用。

关键词

AI 生成；矢量约束；位图细节；分层架构；三维结构；二维投影

1 引言

1.1 当前 AI 生成的技术现状与结构性缺陷

当前 AI 图片与视频生成技术已取得显著进展。扩散模型在图像生成上接近照片级真实感，基于 Transformer 的视频生成模型能够生成长达数十秒的连贯视频。然而，这些技术在结构性缺陷上仍未取得根本突破。

结构崩坏。人物肢体比例失调、物体边界模糊、空间关系矛盾——这些现象在复杂场景中反复出现。模型能够生成逼真的皮肤纹理，却无法保证手指数量正确；能够渲染精细的光影效果，却无法维持物体在三维空间中的几何一致性。

帧间不一致。视频生成中，物体在帧与帧之间发生形变、闪烁或消失。同一角色在前后帧中的面部特征可能不一致，同一场景在不同镜头中的空间布局可能矛盾。

物理违规。光影方向与光源位置不一致，物体运动不符合力学规律，碰撞响应不符合

动量守恒。模型生成的是视觉上"看起来像"的图像，而非物理上"说得通"的场景。

语义漂移。在长序列生成中，模型逐渐偏离初始意图。起始帧是一只猫，中间帧变成了狗，最终帧变成了完全无关的物体。

这些缺陷在 3D 动画从业者眼中，是不可理解、不可接受的错误。对 3D 动画从业者而言，建模完成后角色模型不会在帧间变形，场景设置完成后光照关系不会矛盾，物理模拟遵循力学规律——这些是工作流程的基本前提。当前 AI 视觉生成跳过这些前提直接逐帧渲染，在 3D 动画从业者看来，等于在没有骨架的情况下直接填充血肉。

这些缺陷的结构性根源在于：传统 3D 动画工作流程包含三个阶段——建模、场景设置、渲染。建模创建角色、道具、场景的三维几何结构；场景设置确定摄像机、灯光、材质、物理模拟与动画关键帧；渲染生成最终的二维像素图像。其中，建模与场景设置是逐帧渲染的前提条件——它们分别提供全局元素的三维结构约束与单帧空间关系约束。当前 AI 视频生成跳过了前两个阶段，直接进行第三阶段的逐帧渲染。这种对前置阶段的省略，恰恰是画面崩坏和一致性难以保证的结构性根源。

1.2 现有改进方案的局限

针对上述缺陷，现有研究提出了多种改进方案。后处理修正通过外部模块检测并修复生成结果中的错误，但属于事后补救，无法根除。训练数据扩充提升了统计覆盖范围，但未触及结构性根源。提示词工程优化了条件输入的精确度，但模型内部仍缺乏对宏观结构的显式建模。这些方案的共同缺失是：缺乏显式的宏观结构约束层。

当前模型在位图层面的局部生成能力已经很强，但缺乏一个独立的、先于位图生成的宏观结构约束层来保证全局一致性。更深一层的问题是，现有方案把本应由架构承担的状态管理负担转移给了用户。设定集与关键词模板试图用外部文本约束补偿内部状态管理的缺失，在短程与受控场景中有效，但状态一旦需要随时间演化、升格、在分支间独立维护，外部文本约束即失效——它没有时间维度、没有升格机制、没有分支标识。

1.3 本文任务

本文提出一套三层混合生成逻辑的理论框架，将宏观结构约束、单帧空间关系约束与局部细节生成分层处理。这一框架的理论依据来自矢量图与位图的区分——宏观结构以矢量逻辑约束，单帧空间关系以投影逻辑约束，局部细节以位图逻辑填充。

本文提出的三层混合生成逻辑，可类比为从"小作坊工匠"到"工业级流水线"的范式差异。当前 AI 视觉生成模型类似于传统小作坊的不熟练工匠——无前期设计、边做边改、依赖灵感、无法规模化、废品率高。三层混合生成逻辑构建的是严谨可靠的大规模工业级流水线——先设计后生产、分工协作、结构前置、过程控制、一次合格、可

规模化。

本文为上篇，聚焦理论框架与技术架构。下篇将讨论生产流程重构、算力调用逻辑、跨领域借鉴与产业迁移路径。

1.4 结构安排

本文按以下序列展开：

- 第 2 章：区分矢量图与位图的认知模式，为三层架构奠定理论依据。
- 第 3 章：从计算模拟到 AI 生成的联想过程，推导三层架构的基本形态。
- 第 4 章：三层混合生成逻辑的技术框架，逐层展开职责、内容与协作机制。
- 第 5 章：完全拟真逻辑与范式差异，论证结构约束如何使生成内容在结构层面与真实世界同等自洽。
- 第 6 章：结论，总结核心贡献、理论意义与开放问题。

2 理论依据：矢量图与位图的认知模式区分

2.1 矢量图与位图的基本区分

矢量图	**位图**
描述方式：数学方程（点、线、曲线）	描述方式：像素阵列（离散值）
缩放特性：无限缩放不失真	缩放特性：缩放失真
存储内容：描述几何形状的指令	存储内容：每个像素的具体值
本质：描述规则	本质：离散实现
框架对应：宏观物理→场	框架对应：微观物理→粒子

2.2 框架内定位

矢量图与位图作为描述方式是键层面的算法规则；它们所体现的"连续/离散"认知模式才是相层面的特征。

矢量图模式：物理实体的相以连续、微分、场论的方式呈现。这是宏观物理的相。

位图模式：物理实体的相以离散、统计、粒子的方式呈现。这是微观物理的相。

两者都是对同一投影结构的不同认知模型，不是"哪个更真实"的问题，而是物理实体的相在不同尺度上的适配模式。

2.3 位图作为动态快照

位图不是静态结构，而是动态快照。每一帧对应时间序列的一个标记点，记录了该时刻的物理实体分布。

要素	位图对应	框架内定位
静帧	某一时刻的位图	时间序列中的一个标记点
像素	基本物理实体	坍缩后的确定性实现
像素密度	区域物理实体分布密度	局部坍缩事件的时空分布密度
帧序列	位图序列	时间序列标记的连续记录

位图模式的粒度下限可以理解为一种"本体论意义上的分辨率下限"——不是看不清更细的结构，而是放不下更小的确定性单元。这一特征可以类比为简并态中的量子态占据：低能态被填满后，进一步压缩不会产生"更低的能态"，而是迫使后续电子沿新的维度向上层排布。类似地，位图模式在 2D 平面内达到粒度下限后，试图获得更精细的细节，不会产生"更小的 2D 像素"，而是被迫引入 Z 轴来承载新的确定性单元——微观上仍是像素占位，宏观上则呈现为从 2D 到 2.5D/3D 的维度切换。这一类比仅用于说明"粒度下限意味着占位机制被迫向新维度延伸"的结构特征。

每一帧都是某一时刻的"物理实体分布快照"。它描述的是该时刻该区域的实现状态，而非规则本身。规则是矢量图，实现是位图快照。

2.4 缩放极限与框架内禀属性

矢量图可以无限缩放，但物理场不能。物理场的缩放极限受制于框架载体的极限算力性能和基础规则预设。这些属性是潜存场涌现时携带的框架全局内禀属性，不是物理场自身的性质。物理场在接近缩放极限时，从矢量图模式切换到位图模式。在当前框架中，这一切换发生在普朗克尺度附近。

这一区分对 AI 生成的启示是：宏观结构约束应当采用矢量图模式，微观细节填充应当

采用位图模式。两者不是替代关系，而是协同关系。

3 从计算模拟到 AI 生成的联想过程

3.1 联想链条的起点

计算模拟中，宏观物理场用连续方程描述（矢量图模式），微观物理结构用离散状态描述（位图模式）。两者在模拟中协同运作：矢量约束宏观演化，位图处理局部细节。

AI 生成本质上是一种计算模拟：从噪声到图像的演化过程，是对物理世界结构形成过程的一种模拟。计算模拟需要矢量约束层来控制模拟物的宏观结构呈现——矢量约束宏观演化，位图处理局部细节。当前 AI 生成只模拟了位图层面的局部演化，缺失了计算模拟中天然存在的矢量约束层。因此，AI 生成需要恢复这一矢量约束层。

这一推进是类比性的：从"计算模拟需要矢量约束才能稳定"到"AI 生成同样需要矢量约束层"，依据的是两者在"从局部规则生成宏观结构"上的结构相似性。类比给出的是本文框架的动机，而非严格证明。框架的严格性不来自这一类比本身，而来自后续各章对三层分工、约束传播与偏差反馈的具体展开——第 4 章说明三层各自的功能与协作机制，第 5 章说明结构约束如何使生成内容在结构层面与真实世界同等自洽。

3.2 从二维到三维的推导

AI 图像生成是二维投影，但被生成的对象本身是三维的。因此，矢量约束层应当先构筑三维结构，再通过投影逻辑转换为二维画面。

这一推导产生了三层生成逻辑的基本架构：三维矢量结构约束全局框架与一致性；二维投影逻辑约束单帧空间关系与因果连贯性；位图逻辑填充单帧局部细节。

3.3 与 3D 设计软件及 3D 动画工作流程的对应

三层混合生成逻辑并非凭空构想——它借鉴了当下主流 3D 设计软件与 3D 动画工作流程的成熟原理。

三层混合生成逻辑	**3D 设计软件/动画工作流程对应功能**
第一层：三维矢量结构	建模（Modeling）：创建角色、道具、场景的三维几何结构
第二层：二维投影逻辑	场景设置（Scene Setup）：设置摄像机、灯

	光、材质、物理模拟、动画关键帧
第三层：位图细节填充	渲染（Rendering）：生成最终像素输出，填充纹理、色彩、光影细节

核心差异：传统 3D 工作流程中，三个层级由人类艺术家分别操作。三层混合生成逻辑保留相同的分层结构，但将各层级的执行由人类操作转为 AI 在约束下生成。

这一对应的意义：3D 设计软件与 3D 动画工作流程已经证明了分层约束的有效性。三层混合生成逻辑不是理论空想，而是将已验证的工作流程原理引入 AI 生成领域。现有 3D 工具链（建模软件、渲染引擎、物理引擎）可以作为三层混合生成逻辑的工程基础。

4 三层混合生成逻辑的技术框架

4.1 总体架构

生成过程 = 三维矢量结构→二维投影逻辑→位图细节填充

层级	职责	对应认知模式
第一层	三维矢量结构：约束具有长期一致性但随剧情产生局部变化的全局元素	矢量图模式
第二层	二维投影逻辑：约束镜头视角、光照、明暗、阴影、物理碰撞等单帧空间关系与因果连贯性	投影几何
第三层	位图逻辑：填充单帧局部细节	位图模式

4.2 第一层：三维矢量结构约束

功能：构筑三维矢量结构，约束具有长期一致性但随剧情产生局部变化的全局元素。

内容：全局元素（某一角色、重要道具、场景、某一线索性细节特征等，这些元素需要在长时间跨度内保持一致，同时允许随剧情产生局部变化）；几何约束（全局元素的轮廓、比例、对称性、拓扑关系）；空间约束（全局元素在三维空间中的位置、朝向、相互距离）；运动约束（全局元素在时间中的运动轨迹、速度、加速度）；一致性约束（同一全局元素在不同帧、不同镜头中的几何一致性）。

实现方式：以参数化结构显式建模全局元素的三维结构。这些结构是概念性的、可能性的描述——它们不是具体的像素值，而是像素值必须满足的约束条件。

作用：三维矢量结构提供全局元素的确定性框架。局部生成（位图）在这个框架内进行，不会偏离全局一致性。全局元素在保持长期一致性的同时，允许随剧情产生局部变化——如同本底键提供长期稳定的倾向基础，经验键记录阶段性变化，活跃键驱动当下表现（详见文献[1]）。

4.2A 第一层实现路径

第一层不是端到端生成模型，而是"意图解析 + 结构化资产检索与组合 + 硬约束求解"的协作系统。

意图解析器（LLM 驱动）把自然语言或剧情设定翻译成结构化参数：基本结构描述映射到参数区间，特异标记描述（伤疤、缺指、磨损、畸形、异色）解析为五元组（类型/位置/成因标签/程度/合法形态范围），模糊描述映射到软约束。

结构化资产检索与组合器在 3D 资产库中检索匹配的基础结构，调用参数化修改器做调整。

约束求解器检查组合结果是否满足硬约束（比例范围、拓扑合法性、关节自由度），不满足则回退重新组合。当资产库中没有可精确匹配的基础结构，或反复回退仍不满足硬约束时，系统启动概率比对近似取值机制：在资产库中寻找与目标结构最接近的候选，按匹配算法（轮廓、比例、拓扑关系、关节自由度等维度的加权比较）给出近似度评分，取评分最高的候选进入约束求解。近似取值需两个前提：一是相对严格的取值区间，明确"多接近才算可接受"；二是匹配算法，定义结构相似度的计算方式。这一机制与当前大模型生成中"概率比对、近似取值"的逻辑同构——不是精确检索，而是在候选空间中按相似度排序取最优。

当近似取值仍无法满足硬约束时，系统向两个方向反馈：其一，将缺失的基础结构类型与约束条件反馈给工程师，作为增加此类训练数据的依据；其二，向用户请求附加参考——例图、工程文件、结构草图——作为临时补充的匹配依据。反馈闭环使系统不是"失败即终止"，而是"失败即触发补全"。

训练层面：资产库预训练提供结构先验，意图到结构的条件映射做微调，硬约束作为损失惩罚或采样过滤器贯穿全程。

传统 3D 建模与动画中的骨骼—蒙皮系统为这一层提供了可直接借鉴的工程范式：骨骼层级树对应结构关系函数图，关节自由度与运动限位对应结构参数硬约束，蒙皮权重对应约束从结构向细节的传播机制，形态键对应瑕疵与表情的参数化表示，IK/FK 求解对应约束求解器。

锁定后的结构参数、瑕疵五元组与结构关系函数图，按时间尺度写入分层记忆的本底键（身份级）与经验键（场景级）；瑕疵的合法形态范围可用认知对象数据库的 L1 锚定数据，而非由设计者手写每一条规则。

4.3 第二层：二维投影逻辑约束

功能：以二维投影逻辑约束镜头视角、光照、明暗、阴影、物理碰撞等单帧空间关系，同时约束帧与帧之间的因果连贯性与全局环境对因果关系的泛化响应。

内容：镜头约束（焦距、视角、景深、透视关系）；光照约束（光源位置、方向、强度、色温）；阴影约束（投影方向、软硬程度、遮挡关系）；物理约束（碰撞响应、力学合理性）；因果连贯性约束（帧与帧之间的上下文因果关系；全局环境对因果事件的泛化响应；非局域关联；非直接接触式操控）。

因果连贯性的三个层次：

层次	内容	对应平行并发关系取值
经典力学连锁	作用与反作用对等，力沿空间传播，响应局域化	低（独立）
非局域关联	元素之间存在不依赖空间传播的因果联系，如心灵感应、命运绑定	中（纠缠）
非直接接触式操控	通过非局域关联接口，直接作用于目标的存在形式——能量、物质实体、思想	高（强制并发）

表格中"强制并发"指作用不受空间距离与因果链长度限制，直接改变目标状态。它是"非直接接触式操控"在强度维度上的极端形态：操控仍经由某个接口，强制并发则表现为接口本身被绕过或压缩至零——作用与结果之间不存在可观测的中间环节。

非局域关联与非直接接触式操控是同一作用机制在观察者认知界面上的两种呈现。两者在本体层面都是不依赖空间传播的作用：某一存在对另一存在产生直接影响，这一影响不经过空间中的因果链传递。区别在观察者一侧：当观察者的键系统无法解析其作用机制时，它呈现为"关联"——心灵感应、命运绑定，在叙事中表现为玄学或神秘学；当观察者的键系统能够解析其作用机制时，它呈现为"操控"——通过某个接口直接作用于目标的存在形式。所谓"玄学"，是尚未被理解的底层规律；量子纠缠在理解其机制的人眼中是物理作用，在不理解的人眼中是"心灵感应"。这一区分在叙事设计中可用于表达不同角色对同一作用的不同理解：理解机制的角色视其为"操控"，不理解的角色

视其为"命运"或"感应"。

三层次的划分可与因果律的平行并发关系、因果关联向性两个维度形成粗略对应，作为创作者设定因果规则时的参考标尺（详见文献[2]）；本文不展开该对应的形式化细节。

实现方式：将三维矢量结构通过投影几何转换为二维画面约束。这些约束不仅是单帧层面的空间关系描述，更是帧序列层面的因果连贯性描述——确保每一帧在物理上"说得通"，在时间序列上"接得上"。

作用：二维投影逻辑连接三维结构与二维画面，确保单帧空间关系的物理合理性与帧序列的因果连贯性。

4.4 第三层：位图细节填充

功能：在矢量约束与投影约束的框架内，进行单帧局部细节的位图生成。

内容：纹理细节（皮肤纹理、布料纤维、金属质感）；色彩细节（渐变、反射、折射、次表面散射）；微观细节（颗粒感、噪点、笔触）；局部变化（单帧内的微小形变、表情变化）。

实现方式：扩散模型、Transformer 等位图生成模型在约束框架内运作。局部预测必须满足结构约束，偏差被约束条件纠正或拒绝。

与逐帧渲染的对应：第三层的计算特征与 3D 动画工作流程中的逐帧渲染完全对应——逐帧独立、高密度、持续消耗、帧间不可复用。其算力调用逻辑可直接借鉴逐帧渲染的成熟工程实践：渲染农场、分布式渲染、分块渲染、优先级队列。

作用：位图细节填充在宏观结构框架内进行，保证细节丰富性的同时不破坏全局一致性。

4.4A 三层协作机制

三层协作是"约束向下传播、偏差向上反馈、冲突分级仲裁"的闭环。

约束传播：第一层输出结构约束包（含瑕疵参数及合法区间），同时传给第二层和第三层；第二层输出投影约束包（遮挡、光照、阴影、透视、因果连贯性标签），传给第三层；第三层接收两份约束包，在硬约束内生成细节。

偏差反馈：第三层生成后检查硬约束违反，标记偏差类型（结构越界/投影矛盾/因果断裂），反馈给对应层级；第二层判断是投影参数配置问题还是结构约束本身有问题；第一层判断是参数越界还是约束范围设定过窄。

分级仲裁：硬约束违反（手指数量错误、物体穿模、光影与光源矛盾）直接拒绝回退；软约束偏离（纹理不够丰富、风格不统一）记录放行；约束冲突按"结构优先于投影、投影优先于细节"的优先级仲裁。

失败惩罚与失败报告的设计原则：硬约束违反的拒绝回退需配套合法的失败报告通道，而非在第三层内部自行消化。失败报告须携带可验证的数据支撑：约束冲突项、近似取值记录、已尝试的降级路径。训练目标对三类结果作三档区分——对"有数据支撑的建设性失败"给予奖励（这类失败不是系统缺陷，而是系统在约束边界上的有效探测，它告诉工程师资产库缺什么，告诉用户需要补什么）；对"无数据支撑的敷衍失败"给予惩罚（这类结果不携带可行动的信息，只是把失败的成本转嫁给下游）；对"伪造合规"给予更重的惩罚（伪造合规破坏的是约束的可信度，比单次生成质量下降更危险）。失败报告反馈给对应层级：结构约束问题反馈第一层，投影参数问题反馈第二层。

4.5 三层协同机制

约束传播：三维矢量结构从粗到细传播到二维投影层，再传播到位图生成层。偏差反馈：位图生成的局部偏差反馈到投影层，投影层的偏差反馈到三维结构层。迭代优化：反复迭代直至满足所有约束。三维结构提供硬约束（必须满足），投影逻辑提供中约束（倾向性满足），位图生成提供软约束（在允许范围内自由生成）。

三层架构不仅适用于生成方向，也适用于理解方向：视频理解可视为"整体结构→关系约束→按需细节"的反向应用——先建立场景级结构与关系，再按实际需要触发局部细节分析。这一对称性表明，三层分工是结构与细节之间的一般性组织原则，不限于生成。

5 完全拟真逻辑与范式差异

5.1 四个维度的升维

实际生成效率包含单次生成有效性与整体有效生成速度。单次生成有效性指一次生成即满足结构约束与因果连贯性要求的比例。整体有效生成速度指从输入设定集到输出合格内容的完整周期。

维度	当前 AI 生成	三层混合生成逻辑
视觉美感	局部逼真，全局崩塌	结构约束保证全局美感
细节合理性	细节与结构脱节	细节在物理约束与因果连贯性

		框架内生成
元素一致性	角色、道具、场景在长序列中漂移、崩坏	全局元素在三维矢量结构中长期一致
实际生成效率	反复重试、后处理修正、人工干预	约束前置，单次生成有效性提升，整体有效生成速度加快

5.2 完全拟真逻辑的构建

三层混合生成逻辑的本质是以完全拟真的逻辑构建幻想世界。

第一层（三维矢量结构）：为幻想世界建立与真实世界同等严格的三维结构约束。角色的身体比例、道具的物理属性、场景的空间关系——这些在幻想世界中同样遵循严格的几何与力学规则。

第二层（二维投影逻辑）：为幻想世界建立与真实世界同等严格的投影与因果约束。镜头视角、光照阴影、物理碰撞、因果连贯性——这些在幻想世界中同样遵循严格的物理与逻辑规则。

第三层（位图细节填充）：在严格约束框架内填充细节，使幻想世界的视觉呈现与真实世界同等丰富、同等细腻。

关键区分：幻想世界的"幻想"不在于打破规则，而在于遵循不同的规则体系。魔法遵循魔法的规则，异能遵循异能的规则，但这些规则本身是自洽的、严格的、可推导的。三层混合生成逻辑要求创作者先构建这套规则体系，再让 AI 在规则体系内生成。

5.2A 投影逻辑的统一参数化

投影逻辑为同一算子，其参数组包括：Z 区间（深度维度是否存在与范围）、深度表达机制（深度以何种方式被编码为视觉线索）、投影类型（正交/透视/混合）、深度规则（遮挡与光照如何计算）。

Z 区间决定深度维度是否存在：2D 取[0,0]，2.5D 取有限区间，3D 取全区间。深度表达机制决定深度如何被编码：2D 依靠绘制惯例（近大远小、形变、边缘强调），2.5D 依靠表面法线与高度图的"纸片鼓起"，3D 依靠几何坐标与真实投影。两者正交组合，加上投影类型与深度规则，构成完整的投影参数组。约束：Z = [0,0]强制绘制惯例；Z =有限区间允许绘制惯例或表面法线；Z =全区间要求几何坐标。

Z = [0,0]在逻辑上强制投影退化为 2D 约束框架：深度维度不存在，遮挡、光照、透视

均由平面内的约束关系决定。但 2D 绘制惯例的具体规则——风格化变形、线条优先级、透视夸张——不是从 Z 区间推导出来的，而是在这一约束框架内通过经验积累形成的独立参数组。推导给出的是"必须如此"的逻辑边界，独立参数给出的是"实际如此"的经验内容。两者不矛盾：绘制惯例在逻辑上被 Z=[0,0]强制为 2D 约束，在经验上则表现为一组需要独立标定的风格化参数。

这一"逻辑强制 + 独立参数"的分层结构，为 2D、2.5D、3D 等几何风格提供了统一的解析框架。三者的区别不在投影逻辑本身，而在 Z 区间的取值与对应的独立参数组：Z=[0,0]时投影逻辑强制 2D 框架，独立参数组提供 2D 绘制惯例；Z 取有限区间时投影逻辑允许深度表达但不要求几何坐标，独立参数组提供 2.5D 的表面法线与高度图规则；Z 取全区间时投影逻辑要求几何坐标与真实投影，独立参数组提供 3D 的材质、光照与渲染参数。同一投影逻辑在不同 Z 区间下强制出不同的约束边界，独立参数组在不同约束边界内填充不同的经验内容。2D、2.5D、3D 因此不是三套独立的生成逻辑，而是同一解析框架在不同参数下的具体化。

这一统一化使本方案从"3D 生成架构"扩展为"全维度视觉生成架构"：2D 插画、2.5D 游戏美术、3D 渲染都是同一套三层逻辑在不同投影参数下的具体应用，不需要为每种媒介单独设计生成范式。

5.2B 正常瑕疵与逻辑错误的区分

三层协作需区分"正常瑕疵"与"逻辑错误"。核心判据：瑕疵是否有可追溯的成因标签、是否在第一层定义的合法区间内、视觉异常是否能追溯到合法瑕疵参数。

从底层机制看，不存在无成因的瑕疵。随机磨损、自然老化、表面崩坏，在系统中都有其具体成因——风蚀、霉变、衰变、虫蛀、腐化、机械碰撞，均与目标事物所处环境的变化和遭遇相关。所谓"随机"不是本体论意义上的无因，而是具体模型在算力约束下的分辨率模糊化妥协：混沌系统的大气或粒子环境在更高分辨率下仍有其确定性成因，只是在当前算力下被表达为统计分布。因此，"每个瑕疵参数绑定成因标签"在逻辑层是严格成立的。

工程实现中，可以根据实际算力资源水平对成因表达进行约等或劣化处理：成因标签可以是具体的（如"刀伤"），也可以是类别性的（如"长期磨损""自然老化"），还可以在算力受限时合并为粗粒度的成因类别。这种约等或劣化是工程层的妥协，不是逻辑层对机制的否定。第一层的结构约束包含瑕疵参数及其合法区间：生物畸形（类型/程度/成因标签）、物体磨损（类型/程度/成因标签）、器具不完美（制造公差/装配偏差/表面缺陷）。成因标签决定瑕疵的形态范围；约等或劣化后的成因表达仍然约束形态范围，只是约束的精度随算力资源调整。完全无成因、无法归类的随机崩坏，在逻辑层是逻辑错误；在工程层，如果系统因算力约束无法追溯成因，应标记为"成因待定"并触发回溯，而非直接判定为合法瑕疵。

第二层在计算投影约束时区分两种视觉异常：瑕疵导致的视觉异常（与第一层瑕疵参数一致，正常投影，不标记偏差）与逻辑错误导致的视觉异常（无法追溯到合法瑕疵参数，标记为硬约束违反）。第三层在约束内填充瑕疵细节，不允许自由发明瑕疵。

系统提供瑕疵溯源查询：给定画面中任意不完美视觉特征，系统能回答它对应第一层的哪个瑕疵参数、成因标签是什么、是否在合法区间内。能回答，是正常瑕疵；不能回答，是逻辑错误，应触发回退或警告。

5.3 观众难以区分真假的结构性原因

当前 AI 生成的内容之所以容易被识别为"假"，不是因为细节不够丰富，而是因为结构不真实：光影矛盾、物理违规、因果断裂。观众可能无法说出哪里不对，但能感受到"不自然"。

三层混合生成逻辑从结构层面解决了这个问题：

不自然的来源	三层混合逻辑的解决
光影矛盾	第二层光照约束保证光影一致性
物理违规	第二层物理约束保证力学合理性
因果断裂	第二层因果连贯性约束保证帧间逻辑
元素漂移	第一层三维矢量结构保证全局一致性

结果：生成内容在结构层面与真实世界同等自洽。观众无法通过结构矛盾识别"假"，因为结构层面没有矛盾。

5.4 与现有技术路径的路径差异

当前 AI 视觉生成模型的工作方式，类似于传统小作坊的不熟练工匠——无前期设计、边做边改、依赖灵感、无法规模化、废品率高。每次生成都是个案，质量波动大，一致性无法保证。三层混合生成逻辑构建的是严谨可靠的大规模工业级流水线——先设计后生产、分工协作、结构前置、过程控制、一次合格、可规模化。全局元素在长序列中保持一致，单帧空间关系与因果连贯性得到保障，局部细节在约束框架内自由生成。

核心差异：不是"更好的工匠"，而是"从工匠到流水线"的路径差异。当前路径在小作坊范式内不断优化工匠技艺；三层混合生成逻辑在工业级流水线范式内建立标准、规范

与流程。

5.5 与统一多模态世界模型及分层视频生成的对话

Puffin-World（2026）的统一框架表明，感知、仿真、生成、重建可以在一个模型中完成，其核心是显式建模物理、几何、外观三种原生 3D 状态，而不是仅生成 RGB 帧序列（详见文献[11]）。本文的三层生成与它在目标上一致：都拒绝跳过结构直接逐帧生成。差异在于，Puffin-World 追求模型内的端到端统一，本文追求架构上的显式分层与约束传播。统一世界模型在单帧/短序列上可以很强，但长程叙事中的分支一致性要求状态快照与坍缩链（详见文献[1]）来管理哪些结构在分叉时被复制、哪些被保留在封存池中。统一生成模型不显式提供坍缩链与分支状态管理机制，需要外部架构补充；它可以生成不同分支的内容，但不记录分支之间的坍缩关系与状态继承规则。

分层视频生成的产业实践（2026）已经出现拆解角色/场景/镜头约束、首尾帧锁定、素材资产库的分层产线思路，以及分层生成、Alpha 流与 RGB 流分离的工程方案。本文的贡献是把这种工程直觉提升为与 3D 动画工作流程同构的三层理论框架，而不是又一个工程 trick。

与现有工作的定位差异。当前主流策略（RLHF、3D-aware reward、World-R1 等）在输出层施加后验信号，模型学到的是避开被扣分的输出模式，而非错误在哪一层形成。本方案将结构一致性交给架构协议，使错误可定位到具体层级与约束缺失。代差不在"是否引入结构约束"，而在"约束放在哪里、约束什么、是否可审计"。

6 结论

6.1 核心贡献

本文提出了一套三层混合生成逻辑的理论框架：三维矢量结构约束全局元素；二维投影逻辑约束单帧空间关系与因果连贯性；位图逻辑填充单帧局部细节。

这一框架的理论依据来自矢量图与位图的认知模式区分——宏观结构以矢量逻辑约束，单帧空间关系以投影逻辑约束，局部细节以位图逻辑填充。

从计算模拟到 AI 生成的联想过程表明，AI 生成本质上是一种计算模拟，需要矢量约束层才能实现结构稳定性。与 3D 设计软件及 3D 动画工作流程的对应表明，三层混合生成逻辑借鉴了已被验证数十年的成熟工程原理。

6.2 理论意义

三层混合生成逻辑的核心价值不在于"更好的位图生成"，而在于"从位图层面升维到结构层面"——将生成过程从"局部拼凑"转变为"结构约束下的坍缩实现"。

方案的最大创新在于：将 3D 设计软件与 3D 动画工作流程的成熟原理，系统性地引入 AI 生成领域，并用 AI 恢复被当前 AI 视频生成跳过的前两个阶段。这一思路既不是对当前技术的简单改进，也不是对传统 3D 工作流程的简单复制，而是两者的有机融合。

6.3 开放问题

本文为上篇，聚焦理论框架与技术架构。下篇将讨论生产流程重构、算力调用逻辑、跨领域借鉴与产业迁移路径。

在技术框架层面，仍存在若干开放问题：非局域关联与非直接接触式操控的建模方法；三层架构与现有扩散模型、视频生成模型的工程整合；设定集构建方法论的标准化与工具化；投影参数组在不同媒介间的迁移校准；瑕疵参数五元组与认知对象数据库的接口规范。近似取值匹配算法与取值区间的校准——如何定义结构相似度、如何设定可接受阈值，需结合具体资产库与生成任务展开；失败报告质量判据与奖惩权重的校准——如何定义"数据支撑充分"，如何防止系统学会"堆砌数据但无实质建设性"，需后续研究结合训练框架展开。这些问题为后续研究提供了明确方向。

参考文献

[1] YunFei M. Wang 从长上下文到分层状态：面向长程智能体的设计框架[PP/OL]. AiraXiv, 2026. eprint: 2611.0018

[2] YunFei M. Wang 因果律的形式化分析与异质框架比较——四维分解、耦合运算与回归趋势判别[预印本]. PhilPapers, 2026. 链接：https://philpapers.org/rec/WANAFA-4

[3] YunFei M. Wang 论经典 AI 训练方式认知局限的产生原因与破局思路——迈向基于多尺度机器感知与认知对象数据库的下一代人工智能[PP/OL]. AiraXiv, 2026. eprint: 2611.0009

[4] YunFei M. Wang 数学有效性的哲学解释：基于计算模拟视角的"键-值-信息-相"四层认知模式[预印本]. PhilPapers, 2026. 链接: https://philpapers.org/rec/WANPOT-8

[5] YunFei M. Wang 时间、多重基轴与平行现实：计算模拟中的时间分支与界面选择[预印本]. PhilPapers, 2026. 链接: https://philpapers.org/rec/WANAPE

[6] YunFei M. Wang 关系实体：值层面纯关系结构的形而上学分析[预印本]. PhilPapers, 2026. 链接: https://philpapers.org/rec/WANRSA-2

[7] YunFei M. Wang 键-值-信息-相框架的回应能力：一项基于框架内部资源的批判性推演[预印本]. PhilPapers, 2026. 链接: https://philpapers.org/rec/WANTRO-23

[8] YunFei M. Wang AI 视觉功能的可行性构建方法——基于多尺度机器感知与三层映射模型的工程实现路线图[PP/OL]. AiraXiv, 2026. eprint: 2611.0010

[9] YunFei M. Wang AI 听觉功能的可行性构建方法——基于多尺度机器感知与三层映射模型的工程实现路线图[PP/OL]. AiraXiv, 2026. eprint: 2611.0011

[10] YunFei M. Wang AI 嗅觉功能的可行性构建方法——基于多尺度机器感知与三层映射模型的工程实现路线图[PP/OL]. AiraXiv, 2026. eprint: 2611.0012

[11] Puffin-World: Unified 3D World Model, 2026. .
[12] YunFei M. Wang AI 味觉功能的可行性构建方法——基于多尺度机器感知与三层映射模型的工程实现路线图[PP/OL]. AiraXiv, 2026. eprint: 2611.0013
[13] YunFei M. Wang AI 触觉功能的可行性构建方法——基于多尺度机器感知与三层映射模型的工程实现路线图[PP/OL]. AiraXiv, 2026. eprint: 2611.0014
[14] YunFei M. Wang 论 AI 通感（机器联觉）功能的可行性构建方法——基于五感感知体系与认知对象数据库的跨模态联想与验证机制[PP/OL]. AiraXiv, 2026. eprint: 2611.0015
[15] YunFei M. Wang 论 AI 认知对象数据库的设计与构建——基于五感感知映射的多模态认知对象数据体系与基础设施[PP/OL]. AiraXiv, 2026. eprint: 2611.0016
[16] Wolfram S. A new kind of science[M]. Champaign: Wolfram Media, 2002.

AIGC 使用声明

本文作者提出了论文的核心观点、理论框架、方案设计思路与全部逻辑原理。在文本撰写过程中，作者使用了 DeepSeek AI 模型辅助进行文献观点的归纳、部分文本的起草与修辞润色。作者对 AI 生成的所有内容进行了全面的审阅、批判性修改与最终定稿，并对论文的全部学术内容、观点、数据引用及结论负最终责任。