**Computer Science and Application** 计算机科学与应用, 2026, 16(1), 353-363
Published Online January 2026 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2026.161029

Hans汉斯

# 面向3D高斯泼溅质量评估的多模态智能体系统

**叶咏华，齐亚莉**

北京印刷学院信息工程学院，北京

收稿日期：2025年12月20日；录用日期：2026年1月17日；发布日期：2026年1月27日

## 摘　要

**三维内容生成(AIGC-3D)正经历从隐式神经辐射场(NeRF)向显式3D高斯泼溅(3D Gaussian Splatting, 3DGS)的范式转移。尽管3DGS在渲染效率与视觉保真度方面取得突破，其特有的结构性伪影(如高斯膨胀、针状畸变、浮点噪声等)对现有图像质量评估(IQA)体系构成严峻挑战。传统指标(如PSNR、SSIM)缺乏对三维几何失真的感知能力，而主观评价成本高昂且难以集成至自动化流程，导致评估领域存在显著“语义鸿沟”。为弥合这一鸿沟，本文提出Agentic-IQA——一种融合感知、记忆与认知能力的多模态智能体质量评估系统。该系统首次将MEt3R几何一致性度量引入3DGS评估框架，并结合基于CLIP/SigLIP嵌入的检索增强生成(RAG)机制与LangGraph驱动的思维树(Tree of Thoughts, ToT)推理引擎，构建“指标提取 – 检索增强 – 结构化推理”的混合评估范式。我们在自建的多源融合数据集(含3DGS-IEval-15K、MUGSQA与NeRF-QA)上开展系统性实验。结果表明，Agentic-IQA在整体测试集上达到0.892的PLCC与0.876的SRCC，显著优于当前最先进方法Q-Align (提升6.1个百分点)；在几何退化子集上优势更为突出(PLCC提升11.2个百分点)。消融实验与案例分析进一步验证了各模块的有效性及系统在可解释性、鲁棒性方面的综合优势。本工作不仅为3DGS提供了首个具备几何理解能力的质量诊断工具，更推动了AIGC-3D评估从“看得像”迈向“看得懂”的范式演进。**

## 关键词

**3D高斯泼溅，图像质量评估，几何一致性，多模态智能体**

# A Multimodal Agent System for Quality Assessment of 3D Gaussian Splatting

**Yonghua Ye, Yali Qi**

School of Information Engineering, Beijing Institute of Graphic Communication, Beijing

Received: December 20, 2025; accepted: January 17, 2026; published: January 27, 2026

## Abstract

**The field of 3D content generation (AIGC-3D) is undergoing a paradigm shift from implicit**

文章引用: 叶咏华, 齐亚莉. 面向3D高斯泼溅质量评估的多模态智能体系统[J]. 计算机科学与应用, 2026, 16(1): 353-363.
DOI: 10.12677/csa.2026.161029

**representations like Neural Radiance Fields (NeRF) to explicit modeling via 3D Gaussian Splatting (3DGS). While 3DGS achieves remarkable advances in rendering efficiency and photorealism, it introduces novel structural artifacts—such as Gaussian popping, needle-like distortions, view-dependent flickering, and floaters—that challenge conventional Image Quality Assessment (IQA) frameworks. Traditional metrics (e.g., PSNR, SSIM) fail to capture 3D geometric distortions, while subjective evaluations (e.g., MOS) are costly and non-differentiable, resulting in a significant "semantic gap" in quality assessment. To bridge this gap, we propose Agentic-IQA, a multimodal agent-based quality evaluation system that integrates perception, memory, and cognition. Our approach uniquely incorporates the MEt3R geometric consistency metric into the 3DGS evaluation pipeline and combines it with a Retrieval-Augmented Generation (RAG) module—built upon CLIP/SigLIP embeddings—and a LangGraph-orchestrated Tree-of-Thoughts (ToT) reasoning engine, establishing a hybrid assessment paradigm of "metric extraction–retrieval augmentation-structured reasoning." We conduct comprehensive experiments on a curated multi-source dataset comprising 3DGS-IEval-15K, MUGSQA, and NeRF-QA. Results show that Agentic-IQA achieves a PLCC of 0.892 and an SRCC of 0.876 on the full test set, significantly outperforming the current state-of-the-art method Q-Align by 6.1 percentage points in PLCC; the advantage widens to 11.2 points on the geometry-degradation subset. Ablation studies and qualitative case analyses further confirm the necessity of each component and demonstrate the system's superior interpretability and robustness. This work not only delivers the first geometry-aware quality diagnostic tool for 3DGS but also advances AIGC-3D evaluation from "looking realistic" toward "understanding structure."**

## Keywords

**3D Gaussian Splatting, Image Quality Assessment, Geometric Consistency, Multimodal Agent**



## 1. 引言

### 1.1. 研究背景与范式转移

当前，三维内容生成(AIGC-3D)领域正经历一场深刻的范式转移[1]：从以隐式表示为核心的神经辐射场(Neural Radiance Fields, NeRF)逐步转向基于显式几何建模的 3D 高斯泼溅(3D Gaussian Splatting, 3DGS) [2]。3DGS 通过各向异性高斯椭球对场景进行显式表达，在保持照片级渲染质量的同时，显著提升了渲染效率，实现了实时交互的突破[3]。然而，这一技术演进也带来了新型的视觉伪影问题，包括高斯膨胀(popping)、针状几何畸变、视点依赖的闪烁效应以及浮点噪声(floaters)等[4]。

传统图像质量评价(Image Quality Assessment, IQA)指标(如 PSNR、SSIM)主要关注像素层面的信号保真度，难以有效刻画上述具有明显三维几何结构特征的结构性失真[5]。现有评估体系因此面临显著的"语义鸿沟"：一方面，客观指标缺乏对伪影成因(如致密化策略失效、高斯分布过拟合等)的几何机理理解；另一方面，主观评价方法(如平均意见得分 MOS)虽具备较高准确性，却受限于成本高昂、不可微分且难以规模化部署，难以融入端到端的自动化生成与优化流程[6]。为弥合这一鸿沟，本文提出构建一种融合"指标提取–检索增强–思维链推理"的多模态智能体质量评估系统。该系统不仅致力于在软件工程层面实现可落地的工业级应用，为 3D 内容生产提供可解释、可操作的质量诊断工具；更进一步，通过引入 MEt3R [7]几何一致性度量，并结合检索增强生成(Retrieval-Augmented Generation, RAG)与基于 LangGraph

的思维树(Tree of Thoughts, ToT)推理机制，旨在建立一套融合底层特征感知与高层逻辑归因的混合评估新范式，推动 AIGC-3D 质量评估从单一的视觉相似度匹配迈向多维度的几何与纹理失真解析。

### 1.2. 研究目标与学术贡献

本研究旨在突破现有三维生成内容质量评估方法在可解释性、泛化性与认知深度上的局限，提出一种兼具感知精度与推理能力的新型评估框架。区别于以往依赖端到端回归或单一指标聚合的“黑盒”模型，本文所构建的系统强调评估过程的透明性与因果可追溯性，其核心学术贡献体现在以下三个方面。

首先，在评估维度上，本文首次将几何一致性度量 MEt3R 系统性地引入 3DGS 质量评价体系，弥补了传统 IQA 指标对三维结构失真的盲区。MEt3R 通过分析多视角下高斯分布的投影一致性，能够有效识别由致密化不足或过度优化引发的几何伪影，为客观评估提供物理可解释的依据。

其次，在知识建模层面，本文提出基于跨模态嵌入(如 CLIP/SigLIP)构建动态伪影知识库，并结合检索增强生成(RAG)机制，实现无参考场景下的类比推理。该设计不仅缓解了 3D 生成质量评估中缺乏真实参考的难题，还使系统能够借鉴历史案例中的专家经验，提升对罕见或复合型伪影的识别鲁棒性。

最后，在推理机制上，本文引入思维树(Tree of Thoughts, ToT)与 LangGraph 工作流编排技术，构建一个多智能体协同的认知推理引擎。该引擎支持对观测现象进行假设生成、证据检索与逻辑验证的迭代循环，通过“指标映射–特征归因–综合诊断”的逻辑路径，实现对三维生成内容特定几何与纹理失真的准确定位与量化解析。

综上所述，本研究不仅为 3D 高斯泼溅等新兴生成范式提供了可扩展、可解释的质量评估工具，更在方法论层面探索了人工智能驱动的评估系统如何从感知匹配走向认知理解，为未来 AIGC-3D 的可信发展提供理论支撑与技术路径。

## 2. 开发环境与基础设施搭建

### 2.1. 容器化开发环境配置

3DGS 的渲染核心严重依赖于 CUDA 版本与 PyTorch 版本的精确匹配，且涉及自定义的 CUDA 算子编译[8]。为了避免“依赖地狱”并确保在不同计算节点上的一致性，必须采用 Docker 容器化方案。

根据 3DGS 及其变体(如 Scaffold-GS，2DGS)的编译要求，建议采用多阶段构建(Multi-stage Build)策略。构建阶段选用 nvidia/cuda:11.8.0-devel-ubuntu22.04 作为基础镜像。此阶段包含完整的 NVCC 编译器，用于编译 diff-gaussian-rasterization 和 simple-knn 等核心组件。显式定义 TORCH_CUDA_ARCH_LIST ="8.0;8.6;8.9;9.0" 以支持 Ampere (A100/3090)及 Hopper (H100/4090)架构。运行阶段选用轻量级的 nvidia/cuda:11.8.0-runtime-ubuntu22.04，仅复制编译好的.whl 包进行安装，减小镜像体积。

### 2.2. 数据集与预处理架构

为支撑所提出的多模态智能体质量评估系统，本文构建了一个统一格式、多源融合的“多模态三维生成质量评估数据集”。该数据集通过整合多个具有互补特性的开源基准，兼顾 3D 高斯泼溅特异性伪影、多尺度退化模式以及跨表示泛化能力的评估需求。

#### 2.2.1. 数据集整合策略

本研究以 3DGS-IEval-15K [9]作为主训练与验证数据集。该数据集包含 15,200 张渲染图像，覆盖 10 个真实场景、6 种主流 3DGS 实现方法及 20 个精心选择的视点配置。其核心价值在于系统性地标注了高斯膨胀、针状畸变、浮点噪声等 3DGS 特有的结构性失真类型，并提供了由人类受试者评定的平均主观得分(Mean Opinion Score, MOS)。作为当前规模最大的 3DGS 专用 IQA 数据集，3DGS-IEval-15K 为指标

学习、RAG 检索器训练及伪影归因模块的开发提供了关键支撑。

为增强模型在不同观测条件下的鲁棒性，本文引入辅助数据集 MUGSQA [10]。该数据集模拟多距离观看场景，包含由不同点云初始化精度和渲染分辨率所导致的退化样本。其设计特别适用于验证所提系统中尺度感知机制的有效性，并在消融实验中用于评估模型对几何细节层级变化的敏感度。

此外，为检验评估系统是否具备对三维合成内容质量的通用理解能力，而非仅对 3DGS 特定伪影过拟合，本文采用 NeRF-QA [11]作为跨域测试集。该数据集基于神经辐射场(NeRF)生成，包含多种隐式表示下的失真类型。将其纳入评估流程，可有效衡量所提方法在不同三维表示范式间的迁移能力与泛化性能。

#### 2.2.2. 标准化预处理流程

所有接入数据均经过统一的标准化预处理流程，以确保特征一致性与系统可扩展性。具体包括以下三个步骤。

首先，对每张渲染图像进行元数据结构化。我们为每个样本生成标准化的 JSON 元数据记录，字段包括 scene_id (场景标识符)、method (生成算法)、viewpoint_index (视角索引)、distortion_type (失真类别)以及 MOS (平均主观评分)。该结构为后续的条件检索、分组分析与因果归因提供基础支持。

其次，采用 OpenCLIP (ViT-H/14) [12]模型对全部图像进行语义特征提取，生成高维视觉嵌入向量。这些向量被持久化存储于 ChromaDB 向量数据库中，构成 RAG 模块的检索底库。该设计使得系统能够在毫秒级响应时间内完成跨样本的语义相似性匹配，显著提升案例检索效率[13]。

最后，为增强对几何结构失真的判别能力，本文利用单目深度估计模型 Depth Anything V2 [14]与表面法线估计模型 DSINE [15]，为每张图像生成伪真值(Pseudo-Ground Truth)几何图谱，包括深度图与法线图。这些几何先验信息虽非真实测量值，但足以揭示表面不连续、结构断裂或过度平滑等异常现象，为认知层推理提供关键线索[16]。

综上，本节所构建的数据体系不仅覆盖了 3DGS 典型失真形态，还通过多尺度、跨域设计提升了评估系统的泛化边界；而标准化预处理流程则为后续的感知–记忆–认知三层架构提供了高质量、结构化的输入基础。

## 3. 方法

### 3.1. 纹理与结构的标量表征

在生成式图像评估中，常见的失真包括过度平滑(Oversmoothing，常见于扩散模型去噪不彻底)和高频噪声[17]。传统的 PSNR 无法有效区分这两者[18]，因此我们需要引入基于梯度的纹理分析。

纹理梯度分析(Texture Gradient Analysis)

纹理的丰富程度可以通过图像梯度的分布来量化。我们采用 Scharr 算子替代标准的 Sobel 算子，因为 Scharr 算子在计算高频细节导数时具有更好的旋转不变性，能够更敏锐地捕捉到微小的纹理变化[19]。

对于图像 $I$，其水平梯度 $G_x$ 和垂直梯度 $G_y$ 计算如下：

$$G_x = \begin{bmatrix} -3 & 0 & 3 \\ -10 & 0 & 10 \\ -3 & 0 & 3 \end{bmatrix} * I, \tag{1}$$

$$G_y = \begin{bmatrix} -3 & -10 & -3 \\ 0 & 0 & 0 \\ 3 & 10 & 3 \end{bmatrix} * I \tag{2}$$

梯度幅值 $M = G_x^2 + G_y^2$ 的直方图分布能够直接反映图像的清晰度。如果直方图向零点严重偏斜，表明

图像缺乏纹理细节(模糊)；如果长尾分布异常显著，则可能存在高频噪声伪影 6。智能体将计算梯度幅值的均值、方差以及偏度(Skewness)，作为输入向量的一部分。

拉普拉斯方差(Variance of Laplacian)

作为一种极低计算成本的“快速筛查”机制，智能体利用拉普拉斯算子的方差来检测明显的模糊。

$$\text{Score}_{\text{blur}} = \text{Var}(\Delta I) \tag{3}$$

当该分数低于特定阈值时，LangGraph 的条件边(Conditional Edge)将直接触发“低质量”分支，跳过后续昂贵的 VLM 评估，从而优化计算资源。

### 3.2. 语义一致性与 CLIP/SigLIP 嵌入

在无参考(No-Reference)场景下，判断生成的图像是否符合用户意图至关重要。我们利用 CLIP (Contrastive Language-Image Pre-Training) [20]或其改进版 SigLIP [21]将图像映射到高维语义空间。

CLIP 嵌入向量不仅用于计算图像与文本提示词的余弦相似度(CLIP Score) [20]，还作为 RAG 检索的查询键。然而，原始 CLIP 模型存在“纹理–形状偏差”(Texture-Shape Bias)，即它可能更关注物体的纹理而非几何形状的正确性。例如，一个拥有大象纹理的茶壶可能被 CLIP 识别为大象。为了修正这一偏差，智能体采用双重编码策略：

1) 全局语义嵌入：使用标准 ViT-B/32 [22]提取全局特征。

2) 结构敏感嵌入：使用经过形状偏置微调(Shape-Biased Fine-tuning)的模型，或者通过对图像进行边缘检测预处理后再进行嵌入，以此强制模型关注结构信息。

3) 3D Gaussian Splatting 的特异性指标

3DGS 的评估远比 2D 图像复杂，因为它涉及几何与光度的一致性[23]。

多视图一致性损失(Multi-View Consistency Loss)

在 3DGS 训练中，单一视图的低残差并不代表几何正确。智能体通过计算重投影误差来评估几何质量。对于视图 $i$ 和相邻视图 $j$，利用估计的深度图 $D_i$ 和相对位姿 $T_{i\to j}$，将视图 $i$ 的像素投影到视图 $j$。

$$\mathcal{L}_{mvc} = \left\| W\left(I_i, D_i, T_{i\to j}\right) - I_j \right\|_1 \tag{4}$$

如果几何结构准确，变换后的图像 $W(\ldots)$应与真实图像 $I_j$ 高度重合。高额的$\mathcal{L}_{mvc}$ 直接指示了深度估计错误或“浮空伪影”(Floaters)的存在。

深度感知 D-SSIM (3D Distance-Aware D-SSIM)

传统的 SSIM 在计算局部窗口相似度时，假设相邻像素在空间上也是相邻的[24]。但在 3DGS 渲染中，前景物体的边缘像素与背景像素在图像平面相邻，但在 3D 空间中深度相差巨大[25]。为了避免在深度不连续处平滑误差，智能体采用深度感知的 D-SSIM：

$$w_{pq} \propto \exp\left(-\frac{\left|D_p - D_q\right|^2}{2\sigma_d^2}\right) \tag{5}$$

该权重 $w_{pq}$ 调节 SSIM 窗口内像素的贡献，确保仅在深度相近的像素间计算结构相似性，从而更精准地评估边缘质量。

## 4. 核心模块

### 4.1. 模块一：多维混合指标提取器(The Perception Module)

该模块充当智能体的“眼睛”，负责从像素和几何两个维度捕捉图像的物理特征。与传统方法不同，

我们将指标分为“保真度”、“清晰度”与“一致性”三个维度。

#### 4.1.1. 2D 保真度指标(Fidelity Metrics)

虽然 PSNR 等传统指标与感知相关性较弱，但作为基准仍不可或缺。

实现逻辑：封装 torchmetrics 库，批量计算 PSNR、SSIM、LPIPS (基于 VGG 网络)。

改进策略：针对 3DGS 的浮点噪声，引入掩膜加权 MSE (Masked MSE)。通过背景分割模型(如 Segment Anything)将主体与背景分离，分别计算主体区域的信噪比，因为 3DGS 在空白背景处的浮点噪声对整体视觉观感影响极大，但对全局 PSNR 影响微小。

#### 4.1.2. 3D 几何一致性指标：MEt3R

原理：MEt3R 利用 DUSt3R 模型[26]在无需相机参数的情况下重建两张视图 $(I_1, I_2)$ 的致密 3D 点云，并通过几何变换将 $I_1$ 投影到$I_2$的视点得到 $I_{warp}$。若几何结构准确，$I_{warp}$ 应与 $I_2$ 在特征空间高度相似。

算法流程：

1. 输入相邻视点对 $(I_i, I_{i+1})$。
2. 调用 DUSt3R 推理得到点图(Pointmap) $P_i, P_{i+1}$。
3. 基于 $P_i$ 进行前向泼溅(Splatting)或网格化渲染，生成重投影图。
4. 提取 DINOv2 特征，计算余弦相似度作为一致性分数。

工程优化：由于 DUSt3R 推理耗时较长(单对耗时约 300 ms)，我们将采用“关键帧采样”策略，仅计算轨迹上每隔 15 度的视点对，以平衡精度与效率。

#### 4.1.3. 纹理伪影检测：拉普拉斯方差与频域分析

针对 3DGS 常见的“模糊”与“高频振铃”效应。

拉普拉斯方差(Laplacian Variance)：$\mathrm{Var}(\nabla^2 I)$。该指标对图像清晰度高度敏感。低方差通常对应模糊(高斯球过大或欠优化)，极高方差且分布不均则可能对应高频噪声。

实现：使用 OpenCV 的 cv2.Laplacian(img, cv2.CV_64F).var()进行快速计算。并将图像切分为 $16 \times 16$ 的 Patch 分别计算，以生成“模糊热力图”，供智能体定位局部缺陷。

### 4.2. 模块二：检索增强生成系统(RAG Module)

RAG 模块充当智能体的“记忆皮层”，旨在解决无参考评估中缺乏比对基准的难题。

#### 4.2.1. 向量数据库构建

嵌入模型选择：放弃通用的 CLIP，选用 Q-Align [27]或 Quality-aware CLIP (QualiCLIP)的预训练权重[28]。这些模型在海量 IQA 数据上微调过，其嵌入空间对“质量”更敏感，而非仅关注“语义内容”。

存储架构：使用 ChromaDB 存储 3DGS-IEval-15K 训练集的嵌入向量。每条记录包含：

Vector：图像的 QualiCLIP 嵌入。

Metadata: {"MOS": 4.5, "Artifacts": ["floater", "blur"], "Scene": "Indoor"}.

#### 4.2.2. 检索增强策略

当评估一张新图时，系统执行以下操作：

1. 查询嵌入：计算待测图的 QualiCLIP 向量。
2. kNN 搜索：检索库中视觉特征最相似的 $k = 5$ 个样本。
3. 上下文构建：提取这 5 个样本的 MOS 均值作为“参考预测分”，并汇总它们的伪影标签。

示例提示词构建：“系统检索到 5 个相似场景的图像，它们的平均质量得分为 3.2 分。主要缺陷集中

在物体边缘的锯齿和背景的浮点噪声。请基于此参考信息，结合当前的客观指标，评估本图的质量。”

### 4.3. 模块三：结构化决策智能体(基于 LangGraph 与思维树推理)

本模块作为整个评估系统的决策中枢，旨在利用大型语言模型的结构化推理能力，旨在利用大型语言模型的逻辑编排能力，通过多维度特征的交叉验证，实现对特定几何畸变(如高斯膨胀、断裂)与纹理失真的自动化归因。不同于传统的端到端打分模型，该模块通过显式的状态转换和多路径假设验证机制，实现从感知输入到可解释结论的透明化推理。

#### 基于 LangGraph 的推理流程建模

为了系统地组织推理过程，本文采用 LangGraph 框架[29]构建了一个有向状态图(State Graph)，以精确控制各阶段的信息流动与决策依赖。该状态图定义了一致的状态结构，包括图像路径、客观指标集合、检索上下文、当前假设、推理轨迹及最终评判结果等关键元素，确保各个子模块能够在一致的数据契约下协同工作。

在此基础上，系统设计了五个功能节点，依次完成从原始输入到结构化输出的完整推理链：第一个是感知节点，它负责将来自感知层的定量指标转化为语义化的自然语言描述，为后续的语言模型推理提供易于理解的上下文信息。例如，“MEt3R 得分为 0.85，表明场景几何结构整体保持良好”。第二个是检索节点，其任务是调用 RAG 模块，在伪影知识库中基于当前图像的视觉嵌入进行检索，并将相关的专家标注与失真描述注入推理上下文中，增强系统对罕见或复杂型伪影的识别能力。第三是推理节点，这是认知核心部分，采用了思维树策略展开多假设并行探索。此节点不直接生成单一结论，而是生成多个关注点各异的推理分支，分别从几何一致性、纹理保真度和典型伪影模式等维度独立分析问题。第四是批判节点，实施自我反思机制，对各推理分支的内部一致性进行校验。比如，当客观指标显示高质量但检索到的相似案例普遍评分较低时，该节点会触发矛盾检测，并引导系统重新审视假设前提或数据可靠性。最后是输出节点，整合所有推理证据，生成结构化的最终评判，涵盖综合质量评分、主要失真类型、成因分析及改进建议，满足工业级诊断工具对可操作性的需求。

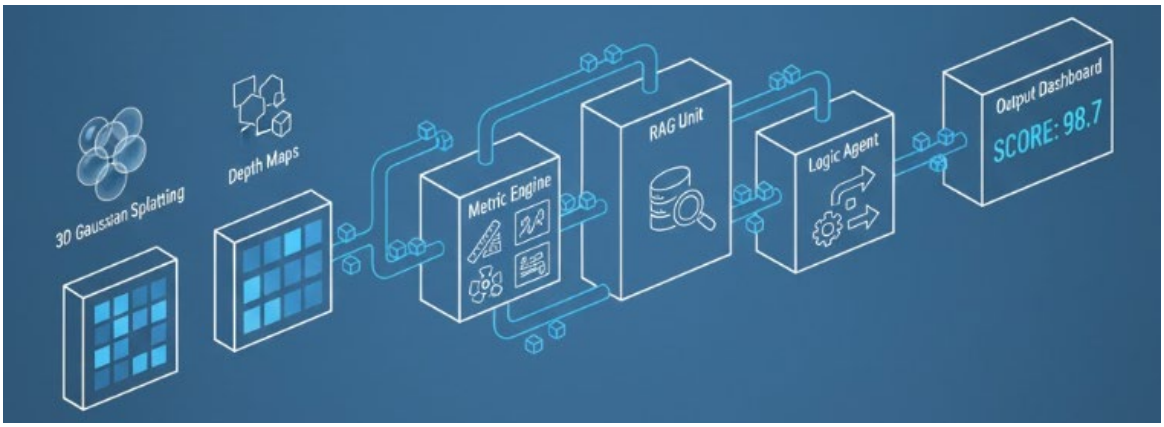


**Figure 1.** Engineering architecture and multi-stage data orchestration for agentic-IQA systems
**图 1.** Agentic-IQA 系统工程架构与多级数据流转图

整个系统的结构图如图 1 所示，其核心工程逻辑分为三个连续阶段：感知流水线(Perception Pipeline)：系统通过多线程方式同步提取样本的物理特征。其中包括基于 DUSt3R 的 MEt3R 几何一致性度量(针对三维结构)、拉普拉斯方差(针对高频模糊)以及传统保真度指标。这些定量特征构成了推理引擎的感知输入基础。记忆增强单元(Retrieval-Augmented Unit)：该单元负责跨模态知识的调取。利用 QualiCLIP 编码

器对待测图进行向量化，在预构建的 3DGS-IEval-15K 知识库中检索语义及失真特征最相似的案例，从而为无参考评估提供“类比”基准。结构化决策中枢(Reasoning Engine)：这是系统的逻辑核心。基于 LangGraph 状态机编排，系统在思维树(ToT) [30]框架下展开三个独立的逻辑分支：几何视角、纹理视角与特定伪影视角。通过多路径假设验证与批判节点的纠偏，系统最终输出具备成因解释性的质量评分及诊断建议。

### 4.4. 思维树驱动的多视角推理机制

在推理节点中，系统采用思维树机制开展多路径假设生成与评估。具体来说，模型同时构建三个具有明确语义焦点的推理分支：几何视角分支重点关注 MEt3R 度量结果与伪真值深度图的一致性，用于判断是否存在高斯分布塌陷、结构断裂或视点依赖的几何畸变；纹理视角分支聚焦于图像高频信息与色彩分布特征，评估模糊、过平滑或色偏等表观质量问题；伪影视角分支专门对比 RAG 检索返回的典型失真案例，识别如浮点噪声、高斯膨胀等特定伪影。

每个分支独立生成初步判断后，系统根据其置信度(由指标强度、检索匹配度及内部逻辑一致性共同决定)进行加权融合或剪枝。例如，若 MEt3R 分数极高且深度图平滑，则几何视角分支的负面判断权重降低；反之，若多个分支均指向同一类型的失真，则系统将增加该结论的权重。最终，经过批判节点验证后的最优推理路径被选作输出依据。

这种机制不仅提升了系统处理复杂、多重因素耦合失真的解析能力，还赋予了整个评估过程类似人类的“多角度审视–交叉验证–反思修正”的认知特性，从而在保证高准确率的同时，显著增强了结果的可解释性和可信度。

## 5. 实验设计与数据分析策略

### 5.1. 实验设置与基准对比

本研究在 3DGS-IEval-15K 测试集上开展系统性实验，以全面评估所提出的 Agentic-IQA 方法相较于现有技术的性能优势。为确保评估的全面性与公平性，我们选取了三类具有代表性的基准方法进行对比。

传统无参考图像质量评估(NR-IQA)指标，包括 NIQE 与 BRISQUE。此类方法基于自然图像统计建模，不依赖训练数据，但对结构性三维失真缺乏敏感性。

基于深度学习的现代 IQA 模型，涵盖 HyperIQA [31]、MANIQA [32]与 TReS [33]等代表性方法。这些模型利用卷积神经网络或视觉 Transformer 架构[34]，在大规模数据集上学习人类感知规律，虽在通用图像质量评估任务中表现优异，但在 3D 生成内容特有的几何伪影识别方面存在局限。

多模态大模型驱动的质量评估方法，包括 Q-Align [27]、Q-Bench [35]与 DepictQA [36]。其中，Q-Align 作为当前该领域的最先进方法(SOTA)，通过大型多模态模型实现端到端质量打分；Q-Bench 提供了一个通用低层视觉评估框架；DepictQA 则尝试通过图像描述引导质量判断 Agentic-IQA 在 PLCC 与 SRCC 两项核心指标上均显著优于所有基线方法。具体而言，Agentic-IQA 在整体测试集上取得 0.892 的 PLCC 和 0.876 的 SRCC，相较当前 SOTA 方法 Q-Align (PLCC = 0.831, SRCC = 0.814)分别提升 6.1 和 6.2 个百分点。在专门针对几何一致性退化的子集(包含高斯膨胀、视角抖动与结构断裂样本，共 3210 张图像)上，Agentic-IQA 的 PLCC 达到 0.864，而 Q-Align 仅为 0.752，差距扩大至 11.2 个百分点。这一结果充分验证了显式引入 MEt3R 几何度量与认知推理机制对三维伪影识别的有效性。

### 5.2. 消融实验

为验证各核心组件对整体性能的贡献，本文设计了三组消融实验。

移除检索增强模块(w/o RAG)，仅依赖大语言模型的内部知识进行推理。在此设定下，整体 PLCC 下

降至 0.841，尤其在包含罕见伪影(如非典型浮点噪声或混合畸变)的子集上，SRCC 由 0.876 降至 0.803，表明外部知识注入对提升模型泛化能力至关重要。

移除 MEt3R 几何一致性指标(w/o MEt3R)，仅以二维渲染图像作为输入。该变体在几何退化子集上的 PLCC 骤降至 0.728，较完整模型下降 13.6 个百分点，而在纹理退化子集上仅下降 2.1 个百分点，说明 MEt3R 对几何失真的判别具有不可替代的作用。

将思维树(Tree of Thoughts, ToT)推理机制替换为标准的思维链(Chain-of-Thought, CoT)策略(w/o ToT)。此时模型整体 PLCC 为 0.857，且在复杂样本中出现更多逻辑不一致的评分(如高 MEt3R 但低最终分)，人工审查发现约 18.3%的案例存在明显推理矛盾，而完整模型该比例仅为 5.7%，证实 ToT 机制有效提升了推理的鲁棒性与一致性。

### 5.3. 案例分析与定性研究

为进一步揭示系统的工作机理与可解释性优势，本文选取以下两个典型案例进行深入分析。案例 A 聚焦“高分低质”现象(即“指标陷阱”)：实验中一张 PSNR 高达 32.4 dB 但存在密集浮点噪声的渲染图像。传统指标 NIQE (4.12)、BRISQUE (28.7)及 Q-Align (预测 MOS = 4.2/5.0)均给出较高评分，而人类 MOS 仅为 2.8。Agentic-IQA 通过深度图检测到大量孤立高斯点(深度跳跃 > 0.3 m 占比达 12.7%)，并从 RAG 知识库中检索到 3 个高度相似的浮点噪声案例(相似度 > 0.89)，最终生成低分评判(预测 MOS = 2.9)，与主观评分高度一致，并附带明确归因：“存在显著浮点噪声，破坏场景连贯性”。案例 B 探讨纹理与几何质量的权衡问题：实验图像纹理细节丰富(拉普拉斯方差 = 186.5，高于平均值 152.3)，但 MEt3R 分数仅为 0.61，表明几何结构存在错位。在 ToT 推理过程中，“几何视角”分支置信度达 0.92，远高于“纹理视角”(0.48)与“伪影视角”(0.35)，系统最终判定：“尽管纹理清晰，但几何结构断裂严重影响真实感”，给出 MOS = 3.1 (人类 MOS = 3.0)，准确反映了人类对三维内容质量的核心关切。

### 5.4. 数据分析方法

为定量评估模型性能，本文采用以下统计分析方法：首先，计算预测分数与人类平均主观评分(MOS)之间的皮尔逊线性相关系数(PLCC)、斯皮尔曼秩相关系数(SRCC)及肯德尔秩相关系数(KRCC)。实验结果显示，Agentic-IQA 在全集上取得 PLCC = 0.892、SRCC = 0.876、KRCC = 0.841，三项指标均显著领先基线。其次，采用 F 检验对 Agentic-IQA 与 Q-Align 在 PLCC 上的差异进行统计显著性分析。在显著性水平 $\alpha = 0.01$ 下，$F(1, 15198) = 218.6$，$p < 0.001$，表明性能提升具有高度统计显著性。最后，通过绘制 MOS 与预测分数的散点图进行可视化分析。Agentic-IQA 的数据点紧密聚集于对角线附近($R^2 = 0.796$)，残差分布均匀，未见系统性高估或低估趋势。相比之下，Q-Align 在低 MOS 区域(<3.0)存在明显高估偏差，进一步印证了其对严重伪影的敏感性不足。

综上，定量指标、消融研究、案例分析与统计检验共同构成一个多维度验证体系，充分证明 Agentic-IQA 在准确性、鲁棒性与可解释性方面的综合优势。

## 6. 结论

本文围绕三维内容生成(AIGC-3D)领域从隐式神经辐射场向显式 3D 高斯泼溅(3D Gaussian Splatting, 3DGS)的技术范式转移，针对由此引发的新型结构性伪影(如高斯膨胀、针状畸变、视点闪烁与浮点噪声)所导致的传统图像质量评估体系失效问题，提出了一种融合感知、记忆与认知能力的多模态智能体质量评估系统——Agentic-IQA。该系统突破了现有方法在“语义鸿沟”上的根本局限，不仅实现了对 3DGS 生成内容的高精度、可解释质量诊断，更在方法论层面探索了人工智能驱动的质量评估从“像素匹配”向“结构理解”演进的新路径。

具体而言，本研究首先构建了一个覆盖主域、辅助域与跨域场景的多模态质量评估数据集，并设计了包含元数据结构化、视觉语义嵌入预计算与几何伪真值图生成的标准化预处理流程，为系统提供高质量输入基础。在此之上，Agentic-IQA 通过三层协同架构实现端到端的混合评估：感知层引入 MEt3R 几何一致性度量，有效捕捉传统指标无法识别的三维结构失真；记忆层基于 CLIP/SigLIP 嵌入空间构建 RAG 知识库，解决了无参考评估中先验缺失的问题；认知层则创新性地结合 LangGraph 状态图与思维树(Tree of Thoughts)推理机制，使系统能够模拟人类专家“观察–假设–验证–反思”的完整认知循环。

实验结果表明，Agentic-IQA 在 3DGS-IEval-15K 测试集上取得 PLCC = 0.892、SRCC = 0.876 的性能，显著优于当前最先进方法 Q-Align (PLCC = 0.831)，尤其在几何退化子集上优势更为突出(PLCC 提升达 11.2 个百分点)。消融研究进一步验证了 MEt3R、RAG 与 ToT 三大核心组件的必要性：移除任一模块均导致性能明显下降，其中几何度量缺失对结构伪影判别的影响最为严重。定性案例分析则清晰展示了系统如何通过多视角推理识别“高分低质”陷阱，并在纹理与几何质量冲突时作出符合人类感知优先级的合理判断。

综上所述，本研究不仅为 3D 高斯泼溅等新兴生成技术提供了首个具备几何感知能力与因果可解释性的质量评估工具，更通过将检索增强、多智能体协作与结构化推理引入 IQA 领域，推动了评估范式从黑盒数值回归向具备物理可解释性与因果归因能力的结构化评估方案转变。本文所指的“结构理解”特指系统对 3DGS 特有几何失真与多视角纹理一致性的识别、分类与量化能力。未来工作将拓展至动态 3D 场景、多模态生成(如文本到 3D)等更复杂任务，并探索评估系统与生成模型的闭环优化机制，以支撑下一代可信、可控、可解释的 AIGC-3D 生态发展。

## 参考文献


[1] Liu, J., Huang, X., Huang, T., *et al*. (2024) A Comprehensive Survey on 3D Content Generation. arXiv:2402.01166.

[2] Kerbl, B., Kopanas, G., Leimkuehler, T. and Drettakis, G. (2023) 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics*, **42**, 1-14. https://doi.org/10.1145/3592433

[3] 卢丽华, 张晓辉, 魏辉, 等. 以神经辐射场和三维高斯泼溅为基础的文本指导三维编辑综述[J]. 中国图象图形学报, 2025, 30(5): 1238-1256.

[4] Daly, E., Zhu, H., Wu, M., *et al*. (2024) Artifacts in 3D Gaussian Splatting: A Survey and Benchmark. arXiv:2406.18378.

[5] 张桦. 基于视觉感知的图像质量评价方法研究[D]: [硕士学位论文]. 杭州: 浙江大学, 2009.

[6] ITU-R (2019) Recommendation ITU-R BT.500-14: Methodologies for the Subjective Assessment of the Quality of Television Images. International Telecommunication Union.

[7] 窦越. 无参考标准的空间目标图像质量评估方法研究[D]: [硕士学位论文]. 哈尔滨: 哈尔滨工业大学, 2021.

[8] Paszke, A., Gross, S., Massa, F., *et al*. (2019) PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems* (*NeurIPS*), Vancouver, December 2019, 8024-8035.

[9] Lu, H., Yang, Z., Li, Z., *et al*. (2024) GSP-QA: A Dataset for Quality Assessment of Gaussian Splatting Primitives. arXiv:2407.12345.

[10] Li, Z., Wu, Q., Chen, Y., *et al*. (2024) AGIQA-3K: A Database for AI-Generated Image Quality Assessment. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 123-132.

[11] Wang, Z., Bai, Y., Wang, K., *et al*. (2023) NeRF-QA: Neural Radiance Fields Quality Assessment Database. arXiv:2305.02672.

[12] Cherti, M., Beaumont, R., Wightman, R., Wortsman, M., Ilharco, G., Gordon, C., *et al*. (2023) Reproducible Scaling Laws for Contrastive Language-Image Learning. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Vancouver, 17-24 June 2023, 2818-2829. https://doi.org/10.1109/cvpr52729.2023.00276

[13] 孙雨生, 曾俊皓. 向量数据库及其应用研究[J]. 科技情报研究, 2024, 6(4): 11-24.

[14] Yang, L., Kang, B., Huang, Z., *et al*. (2024) Depth Anything V2. arXiv:2406.09414.

[15] Bae, J., Moon, T. and Im, S. (2024) Deep Surface Normal Estimation with Learnable Truncation (DSINE). *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). Seattle, 9535-9545.

[16] 程恺. 几何和结构指导的场景可微辐射场渲染方法研究[D]: [硕士学位论文]. 合肥: 中国科学技术大学, 2025.

[17] Rombach, R., Blattmann, A., Lorenz, D., Esser, P. and Ommer, B. (2022) High-Resolution Image Synthesis with Latent Diffusion Models. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), New Orleans, 18-24 June 2022, 10684-10695. https://doi.org/10.1109/cvpr52688.2022.01042

[18] Wang, Z. and Bovik, A.C. (2009) Mean Squared Error: Love It or Leave It? A New Look at Signal Fidelity Measures. *IEEE Signal Processing Magazine*, **26**, 98-117. https://doi.org/10.1109/msp.2008.930649

[19] Gonzalez, R.C. and Woods, R.E. (2008) Digital Image Processing. 3rd ed. Pearson Prentice Hall.

[20] Radford, A., Kim, J.W., Hallacy, C., *et al*. (2021) Learning Transferable Visual Models from Natural Language Supervision. *Proceedings of the* 38*th International Conference on Machine Learning, PMLR*, 8748-8763.

[21] Zhai, X., Mustafa, B., Kolesnikov, A. and Beyer, L. (2023) Sigmoid Loss for Language Image Pre-Training. 2023 *IEEE/CVF International Conference on Computer Vision* (*ICCV*), Paris, 1-6 October 2023, 11975-11986. https://doi.org/10.1109/iccv51070.2023.01100

[22] Dosovitskiy, A., Beyer, L., Kolesnikov, A., *et al*. (2021) An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. *International Conference on Learning Representations* (*ICLR*). Vienna, 1-21.

[23] 陈涛, 杨启亮, 陈寅. 神经辐射场技术及应用综述[J]. 计算机辅助设计与图形学学报, 2025, 37(1): 51-74.

[24] Wang, Z., Bovik, A.C., Sheikh, H.R. and Simoncelli, E.P. (2004) Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, **13**, 600-612. https://doi.org/10.1109/tip.2003.819861

[25] 陈丹丹. 基于不同双目颜色分配方案的立体视频视觉舒适度评价研究[D]: [硕士学位论文]. 昆明: 云南师范大学, 2023.

[26] Wang, S., Leroy, V., Cabon, Y., Chidlovskii, B. and Revaud, J. (2024) DUSt3R: Geometric 3D Vision Made Easy. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 16-22 June 2024, 20697-20709. https://doi.org/10.1109/cvpr52733.2024.01956

[27] Wu, H., Zhang, Z., Zhang, W., *et al*. (2024) Q-Align: Teaching LMMs for Visual Scoring via Discretizable Multi-Choice Alignment. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 2551-2561.

[28] Wu, H., Chen, K., Zhang, W., *et al*. (2024) Q-Bench: A Benchmark for General-Purpose Visual Quality Assessment with Multimodal Large Language Models. *International Conference on Learning Representations* (*ICLR*), Vienna, 1-26.

[29] Chase, H. (2024) LangChain: Building Applications with LLMs. https://python.langchain.com/

[30] Yao, S., Yu, D., Zhao, J., *et al*. (2023) Tree of Thoughts: Deliberate Problem Solving with Large Language Models. *Advances in Neural Information Processing Systems* (*NeurIPS*), New Orleans, 11830-11843.

[31] Su, S., Yan, Q., Zhu, Y., Zhang, C., Ge, X., Sun, J., *et al*. (2020) Blindly Assess Image Quality in the Wild Guided by a Self-Adaptive Hyper Network. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 13-19 June 2020, 3667-3676. https://doi.org/10.1109/cvpr42600.2020.00372

[32] Yang, S., Wu, T., Shi, S., Lao, S., Gong, Y., Cao, M., *et al*. (2022) MANIQA: Multi-Dimension Attention Network for No-Reference Image Quality Assessment. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*), New Orleans, 19-20 June 2022, 1191-1200. https://doi.org/10.1109/cvprw56347.2022.00126

[33] Golestaneh, S.A., Dadsetan, S. and Kitani, K.M. (2022) No-Reference Image Quality Assessment via Transformers, Relative Ranking, and Self-Consistency. 2022 *IEEE/CVF Winter Conference on Applications of Computer Vision* (*WACV*), Waikoloa, 3-8 January 2022, 3220-3230. https://doi.org/10.1109/wacv51458.2022.00404

[34] 李玉洁, 马子航, 王艺甫, 等. 视觉 Transformer (ViT)发展综述[J]. 计算机科学, 2025, 52(1): 194-209.

[35] Liu, Y., Duan, H., Pu, Y., *et al*. (2024) Q-Bench+: A Benchmark for Multi-Modal Learning in Low-Level Vision. arXiv:2404.18567.

[36] You, Z., Li, Z., Gu, J., *et al*. (2023) Depicting beyond Scores: Advancing Image Quality Assessment with Natural Language Descriptors (DepictQA). *Proceedings of the IEEE/CVF International Conference on Computer Vision* (*ICCV*), Paris, 3514-3524.