

# 技术介绍文档：SketchFaceGS

基于 3D 高斯泼溅的实时草图驱动人脸编辑与生成

## 1. 作者信息

作者：李波<sup>1</sup>，康家浩<sup>1</sup>，马煜博<sup>1</sup>，刘锋林<sup>2</sup>，刘彬<sup>1</sup>，张方略<sup>3</sup>，高林<sup>2\*</sup>

单位：

<sup>1</sup> 南昌航空大学

<sup>2</sup> 中国科学院计算技术研究所

<sup>3</sup> 惠灵顿维多利亚大学

\* 通讯作者邮箱： [gaolin@ict.ac.cn](mailto:gaolin@ict.ac.cn)

## 2. 研究背景

3D 高斯泼溅（3D Gaussian Splatting, 3DGS）因其照片级逼真的渲染质量和实时的渲染速度，已经成为数字人头建模的强大范式。然而，如何直观且交互式地创建或编辑 3D 高斯人头模型仍然是一个巨大的挑战。

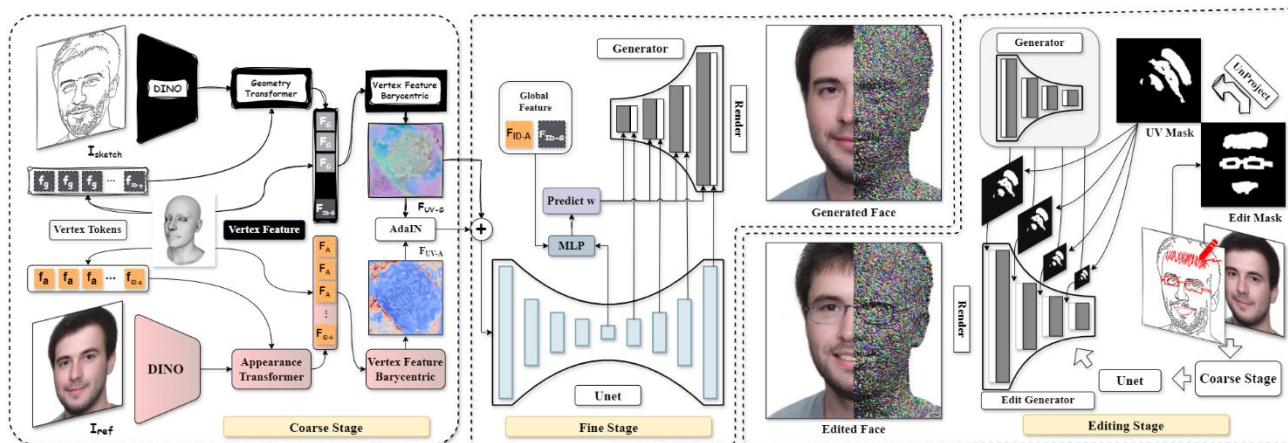
- 现有痛点：**虽然 2D 草图是进行快速、直观概念设计的理想交互方式，但草图本身具有稀疏性、深度模糊性且缺乏高频外观信息，这使得从笔画中推断出密集且几何一致的 3D 高斯结构非常困难，特别是在实时要求的约束下。
- 现有方法的局限：**之前的草图驱动 3D 生成方法（如 SketchFaceNeRF）依赖于耗时的逐场景优化（per-instance optimization），容易累积误差且无法实现实时交互；而基于 GAN 或文本驱动的方法则缺乏对局部细节的精细和便捷控制。

## 3. 创新点

- 首个统一的端到端框架：**提出了首个能够从单张 2D 草图实时生成和交互式编辑照片级 3D 高斯人头模型的端到端框架，提供了直观且对艺术家友好的工作流。
- 粗到细（Coarse-to-fine）生成管线：**设计了并行的 Transformer 分支，分别从草图提取几何特征和从参考图像提取外观特征，生成粗略的 UV 图；随后通过基于 3D-GAN 的调制模块为其注入高频、逼真的细节。

3. **UV 掩码融合技术 (UV Mask Fusion)**：提出了一种用于实时编辑的机制，在生成器的多尺度特征图上进行逐层特征融合。该方法避免了在 3D 空间中直接组合高斯点所产生的空间伪影，实现了自然、稳定、连续且支持自由视角的编辑。

## 4. 技术描述



SketchFaceGS 框架总览图

SketchFaceGS 采用前向传播 (feed-forward) 的粗到细架构，主要包含以下技术模块：

- **粗阶段 (基于 Transformer 的 UV 特征预测)：**  
给定一张输入草图和一张参考图像，系统通过并行的 Transformer 架构分离几何与外观信息。使用预训练的 DINOv2 提取特征后，利用可学习的顶点查询 (vertex queries) 生成逐顶点的几何特征和外观特征，并利用 AdaIN (自适应实例归一化) 对齐网络解决两者间的几何-外观冲突，最终投影出粗略的几何一致的 UV 特征图。
- **细阶段 (3D UV 特征增强)：**  
为了弥补粗阶段缺乏的照片级细节，引入了一个 U-Net 网络，将粗 UV 图映射为全局潜在变量 (Global latent) 和多尺度空间调制参数。这些参数被逐层应用到预训练的 GGHead (基于 StyleGAN2) 生成器中，将粗略的表示提升为高保真、可渲染的完整 3D 高斯属性。
- **草图驱动的 3D 头部编辑：**  
用户的 2D 编辑笔画 (绘制/擦除) 被转换为 2D 像素掩码，反向投影到 3D 空间找到对应的高斯点，并映射为规范化的 UV 掩码。通过\*\*逐层特征融合 (Layer-by-Layer Feature Fusion)\*\*策略，在 StyleGAN 生成器的特征空间内将编辑后与未编辑的特征进行融合，从而在保证边缘无缝衔接的同时维持非编辑区域的稳定。

## 5. 性能指标

SketchFaceGS 在图像质量和交互/渲染速度上均达到了 sota 水平（在合成数据集和真实手绘草图集上进行了全面测试）：

- **生成性能（Generation）：**
  - FID ↓：92.65（明显优于基线模型 S3D 的 96.03 和 SketchFaceNeRF 的 94.94）
  - KID ( $\times 100$ ) ↓：4.00  $\pm$  0.4（达到最优）

Method	FID ↓	KID ( $\times 100$ ) ↓
S3D [47]	96.03	4.50 $\pm$ 1.0
Nano-LAM [12, 20]	133.72	7.61 $\pm$ 0.9
SketchFaceNeRF [15]	94.94	4.53 $\pm$ 0.6
Ours	92.65	4.00 $\pm$ 0.4

生成性能对比表格

- **编辑与渲染速度（Editing & Rendering）：**
  - 端到端单次编辑延迟：约 0.3 秒（对比 SketchFaceNeRF 需要耗时约 10 秒）
  - 渲染帧率（FPS）：高达 243 FPS（远超同类基于优化的方案，完全满足实时交互需求）
  - 编辑图像质量：FID 低至 44.60，KID 为 0.69，编辑后的模型细节自然且无拼接伪影。

Method	FID ↓	KID ( $\times 100$ ) ↓	Time (s) ↓	FPS ↑
MagicQuill	46.48	0.78 $\pm$ 0.2	~6.0	—
Nano-LAM	74.26	3.01 $\pm$ 0.3	~15.0	281
SketchFaceNeRF	62.49	2.65 $\pm$ 0.3	~10.0	42
Ours	44.60	0.69 $\pm$ 0.2	~0.7	243

编辑性能对比表格

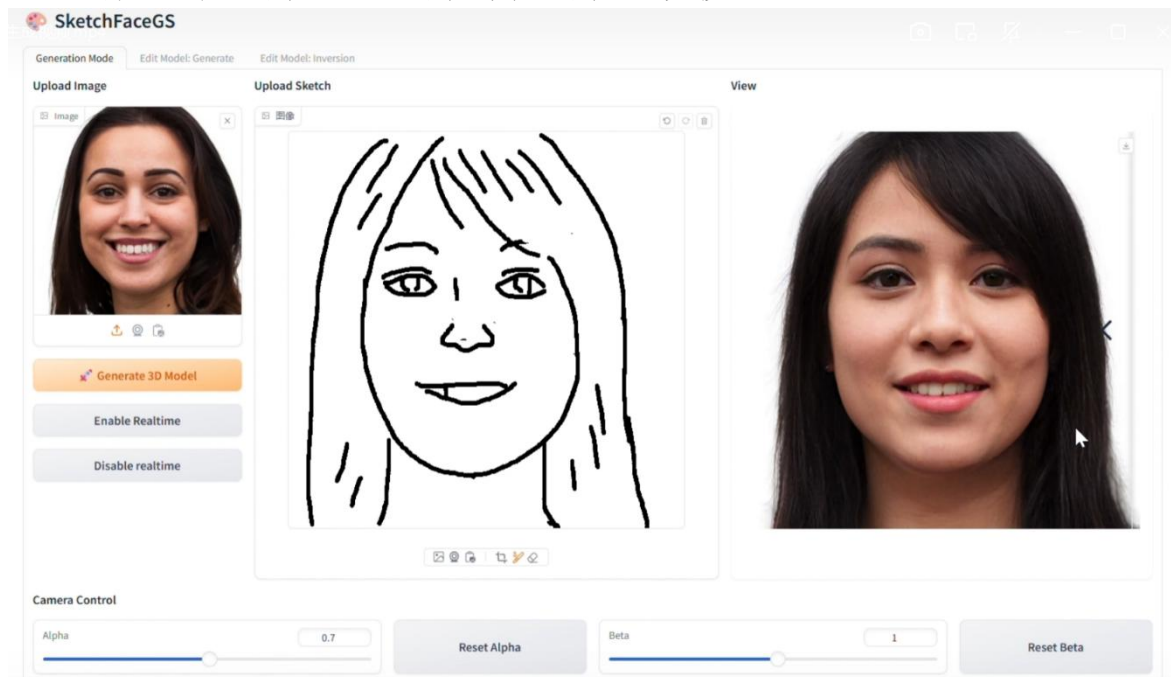
## 6. 视频演示界面及效果展示

演示系统根据功能分为“生成模式（Generation）”与“编辑模式（Editing）”两大核心模块，提供直观的交互体验：

**模式一：基于草图的 3D 人脸生成模式（Generation Mode）**

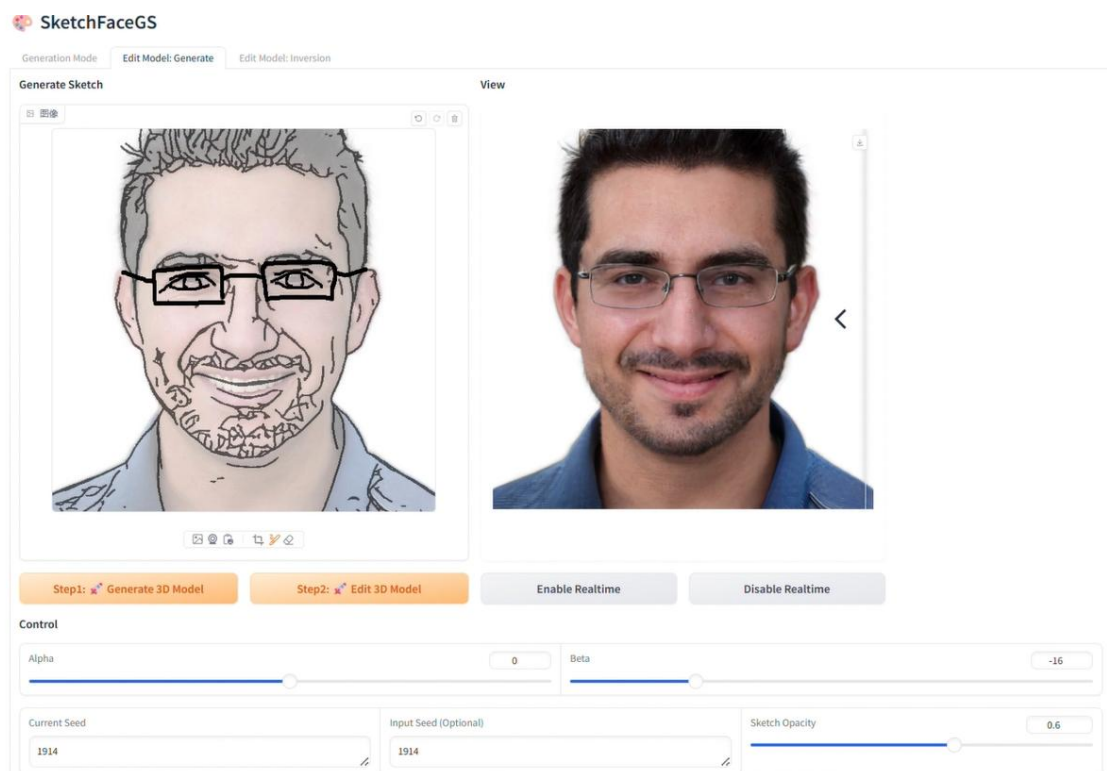
- **交互逻辑：**在此模式下，用户只需输入一张手绘的单视角面部草图（Sketch），并可选择性地提供一张参考图像（Reference Image，用于提取肤色、光照等外观特征）。

- **一键生成：**得益于系统前向传播（feed-forward）的粗到细架构，系统无需耗时的优化过程，即可在极短时间内将稀疏的 2D 线条直接转化为密集、且几何高度一致的高保真 3D 高斯人头模型。



上图展示了在生成模式下，系统通过单张手绘草图及参考图像，直接合成的照片级逼真、且支持任意视角渲染的 3D 人头模型，几何结构与草图高度吻合。

## 模式二：交互式 3D 人脸编辑模式（Editing Mode）



上图展示了用户生成原始 3D 人脸后，在不同视角下进行连续的局部笔画编辑，实时生成高逼真度编辑结果及多视角一致性展示。

- **界面布局与操作：**系统允许用户导入基础 3D 人脸模型（或通过真实照片反演重建的模型）。在界面中直接提供 2D 笔刷工具，用户可以自由切换视角，在任意视角的屏幕空间下绘制或擦除草图线条（例如修改发型轮廓、添加眼镜、修改脸型轮廓，甚至进行局部外观如纹身的添加）。
- **实时交互反馈：**得益于创新的 UV 掩码融合（UV Mask Fusion）技术，系统在不到 0.3s 内即可将草图的修改精准转化为 3D 高斯模型的更新。用户可立即拖拽改变视角（View Change），全方位查看修改效果，且未编辑区域依然保持完美的原有纹理和多视角一致性。

## 7. 相关论文

### 1. 本项目核心论文

- **SketchFaceGS: Real-Time Sketch-Driven Face Editing and Generation with Gaussian Splatting**  
Bo Li, Jiahao Kang, Yubo Ma, Feng-Lin Liu, Bin Liu, Fang-Lue Zhang, Lin Gao. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

### 2. 基础工作

- **3D Gaussian Splatting for Real-Time Radiance Field Rendering**  
Bernhard Kerbl et al. *ACM Transactions on Graphics (TOG)*, 2023.
- **GGHead: Fast and Generalizable 3D Gaussian Heads**  
Tobias Kirschstein et al. *SIGGRAPH Asia*, 2024.

### 3. 基线工作

- **SketchFaceNeRF: Sketch-based facial generation and editing in neural radiance fields**  
Lin Gao et al. *ACM Transactions on Graphics (TOG)*, 2023.
- **MagicQuill: An intelligent interactive image editing system**  
Zichen Liu et al. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.