

View-Consistent 3D Editing with Gaussian Splatting

Yuxuan Wang¹, Xuanyu Yi¹, Zike Wu¹, Na Zhao², Long Chen³, and Hanwang Zhang^{1,4}

¹ Nanyang Technological University

² Singapore University of Technology and Design

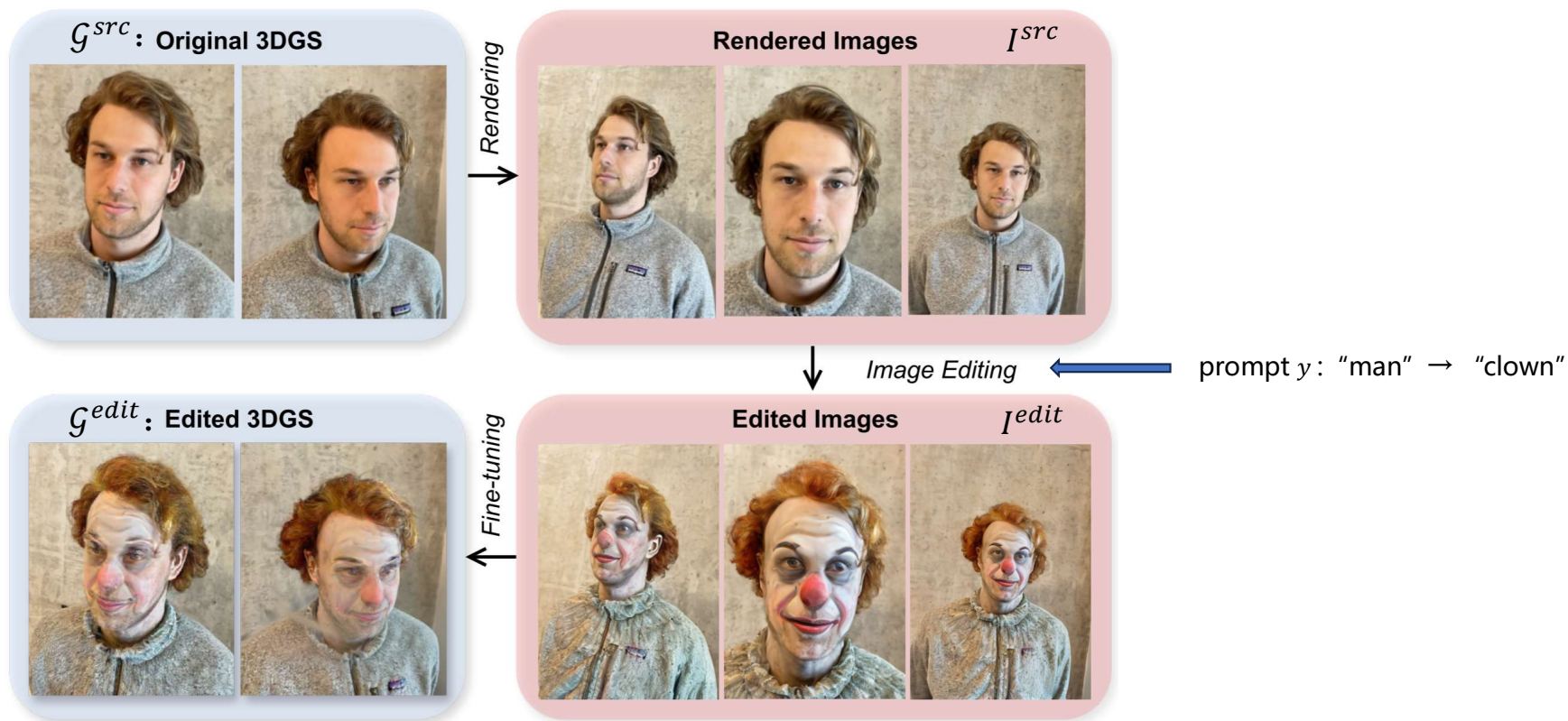
³ Hong Kong University of Science and Technology

⁴ Skywork AI

Accepted by ECCV 2024

一、介绍

One Iteration



目的: 给定 \mathcal{G}^{src} 和 修改提示词 y , 将 \mathcal{G}^{src} 转换为与 y 对齐的编辑版本 \mathcal{G}^{edit}

- 从多个视图 $\mathcal{V}=\{v\}$ 对3DGS进行渲染, 生成源图像集合 I^{src}
- 利用2D编辑模型将 I^{src} 根据文本提示词 y 转换为 I^{edit}
- 最后将 I^{edit} 用作训练指导, 将 \mathcal{G}^{src} 微调为 \mathcal{G}^{edit}

$$\mathcal{G}^{edit} = \operatorname{argmin}_{\mathcal{G}} \sum_{v \in \mathcal{V}} \mathcal{L}_{\text{Edit}}(\mathcal{R}(\mathcal{G}, v), \mathcal{I}^{edit}),$$

where \mathcal{R} represents the rendering function that projects 3DGS to image given a specific view v .

一、介绍



"man" -> "clown"



"man" -> add "red glasses"



"stone bear" -> "brown bear"



"stone bear" -> "pig"



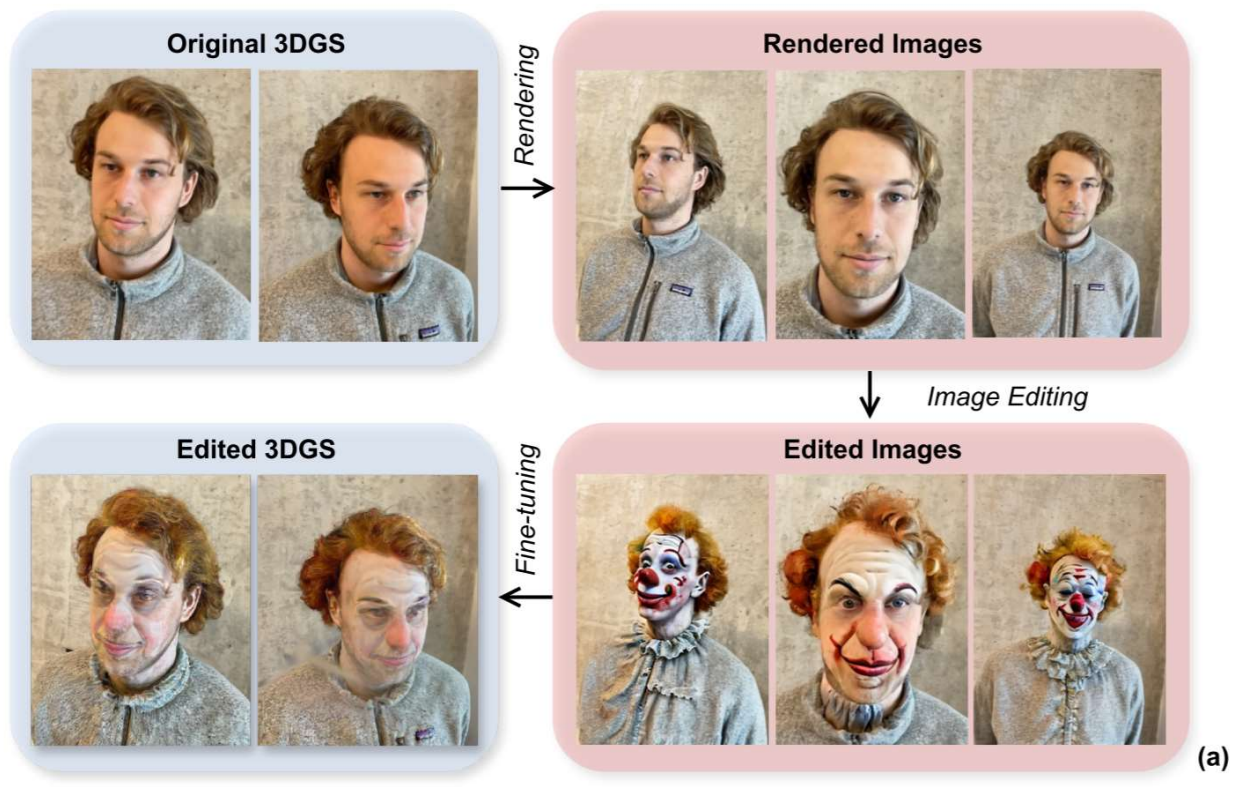
"whole scene" -> "autumn"



"bike" -> "blue bike"

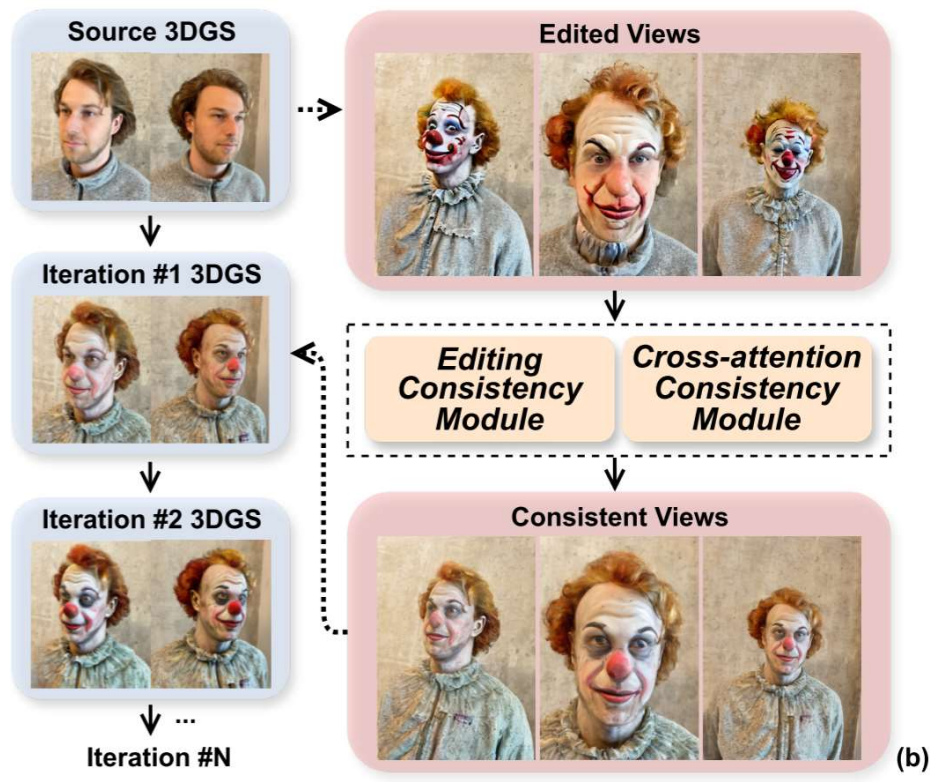
- 通过确保多视图一致的图像引导，VcEdit减轻了伪影，并在高质量的编辑中表现出色

二、存在的问题



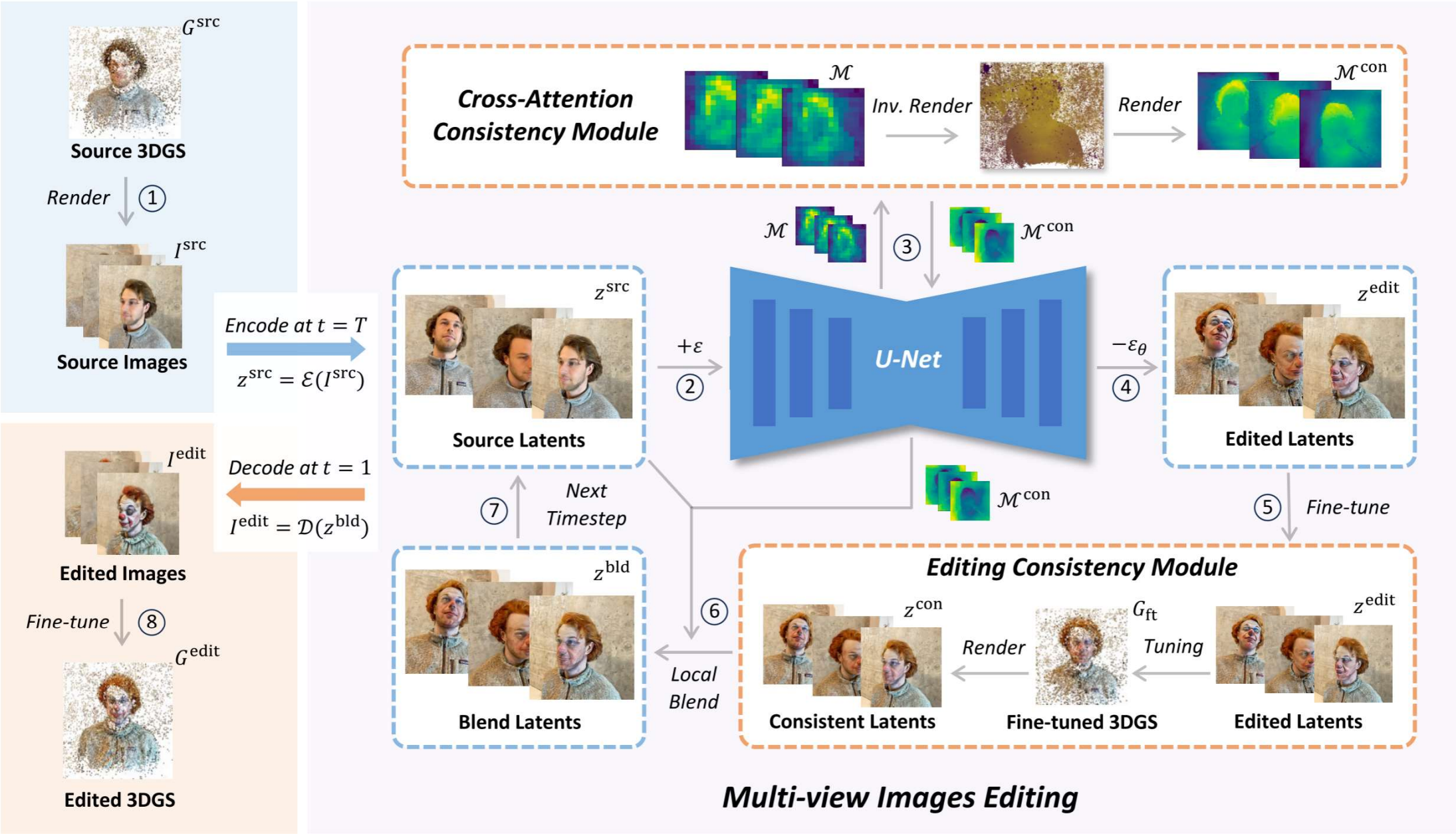
跨视图编辑不一致性:

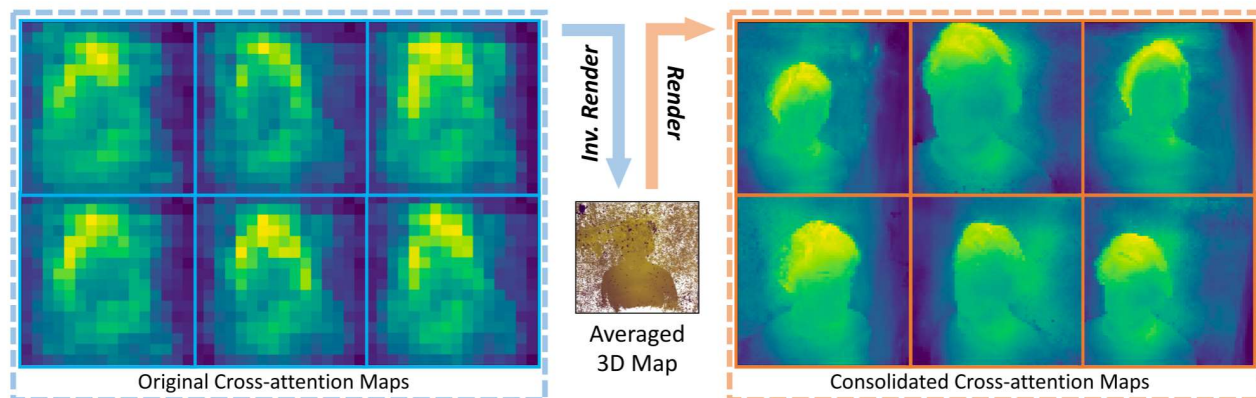
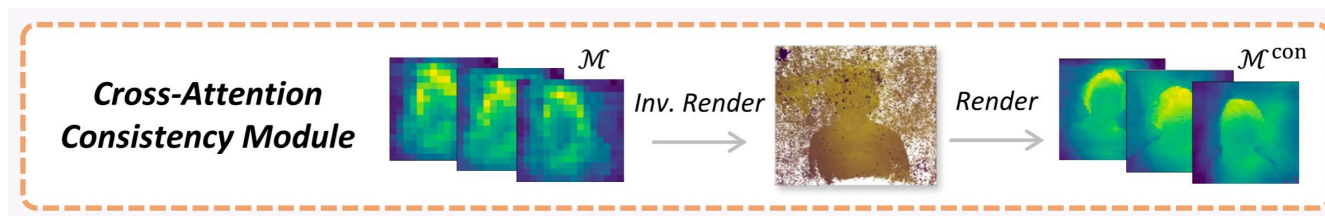
- 一个人的渲染视图被单独编辑为不同风格的小丑，导致学习3DGS的模式崩溃问题



论文解决方案:

- 采取迭代模式，对每一轮迭代采用两个一致性模块来增强编辑后的一致性



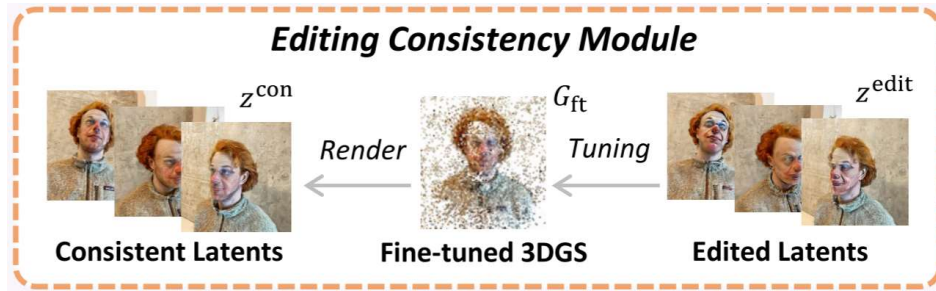


- 对于单个 Gaussian j , 涉及到的所有视图 v , 建立所有视图包含的像素点 p 和 Gaussian j 的联系

$$\mathbb{M}_j = \frac{1}{C_j} \sum_{v \in V} \sum_{\mathbf{p} \in \mathbb{M}_v} o_j(\mathbf{p}) \cdot T_{j,v}(\mathbf{p}) \cdot \mathbb{M}_{v,\mathbf{p}}$$

- o : opacity, T : transmittance matrix, \mathbb{M} : cross-attention weight
- C_j : Gaussian j 涉及到的像素点的个数

- In the created 3D map \mathbb{M} , each Gaussian receives an attention weight towards each word in the prompt

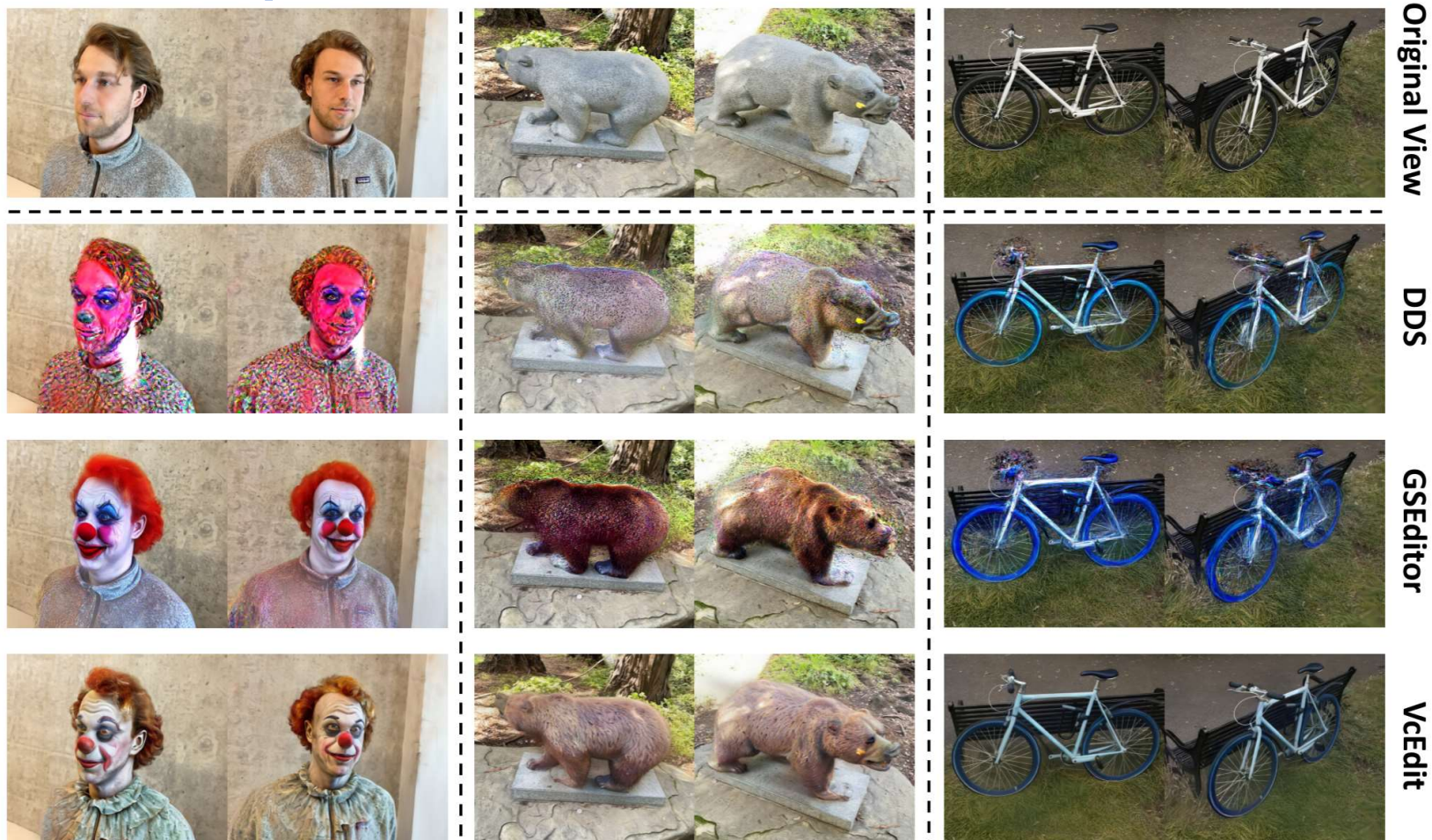


\mathcal{G}^{ft} : a copy of the original 3DGS model \mathcal{G}^{src}

We illustrate the proposed module in the *bottom-right* of Fig. 3. Initially, we decode the multi-view edited latents z^{edit} into images, denoted as $\mathcal{D}(z^{\text{edit}})$, which are subsequently used as guidance to fine-tune a copy of the original 3DGS model \mathcal{G}^{src} , which is denoted as \mathcal{G}_{ft} :

$$\text{fine-tuned: } \mathcal{G}_{\text{ft}} = \operatorname{argmin}_{\mathcal{G}} \sum_{v \in \mathcal{V}} \mathcal{L}_{\text{edit}}(\mathcal{R}(\mathcal{G}, v), \mathcal{D}(z_v^{\text{edit}})) \quad (7)$$

三、Qualitative comparison



"man" -> "clown"

"stone bear" -> "brown bear"

"bike" -> "blue bike"

VcEdit擅长于有效地解决多视图不一致，从而产生卓越的编辑质量
相比之下，其他方法遇到了模式崩溃的挑战，并表现出闪烁的伪影

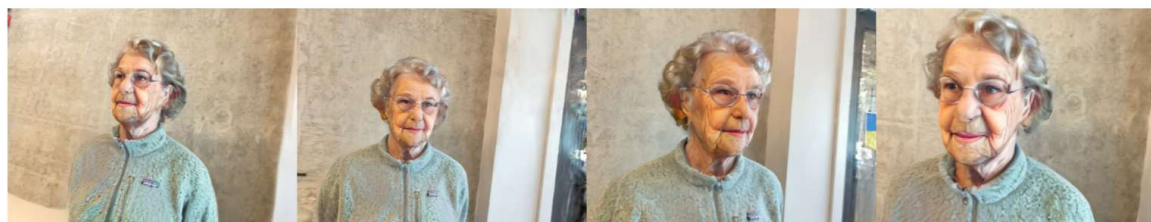
四、Quantitative Comparisons

Table 1: Quantitative Comparison: Our VcEDIT performs in both user study evaluations and CLIP T2I Directional Similarity [12] metrics.

Metrics\Methods	DDS [16]	GSEditor [6]	VcEDIT
User Study	1.57%	34.49%	63.93%
CLIP similarity [13]	0.1470	0.1917	0.2108

- VcEdit不仅在用户研究方面取得了优异的成绩
- 而且在CLIP文本-图像方向相似度方面也取得了优异的成绩
- 这一性能表明VcEdit提供了与预期提示密切相关的高质量编辑，在编辑质量和有效性方面明显超过了基线方法。

Editing with Diverse Source



"man" -> "grandma"



"doll" -> "robot"



"whole scene" -> "on fire"

original view

rendered view of edited 3DGS

➤ 呈现了目标“奶奶”的面部细节

➤ 保留了娃娃的关键特征

➤ 展示了其大规模场景修改的能力，
改变了林地场景，将其描绘成
“着火了”

➤ VcEdit的广泛结果：该方法能够完成各种编辑任务，包括人脸、物体和大规模场景编辑。

五、Ablation Study--Consistency Modules



- Top: 2D edited guidance images using three variant versions of VcEdit
- Bottom: 2D edited guidance images using GSEditor

Thank You