



中国科学院自动化研究所
模式识别实验室
New Laboratory of Pattern Recognition



中国科学院自动化研究所
INSTITUTE OF AUTOMATION
CHINESE ACADEMY OF SCIENCES



中国科学院大学
University of Chinese Academy of Sciences

Exploring Robust and Expressive Face Representations for Multi-Modal Manipulation



Songlin Yang (杨嵩林)



- **M.S. Degree** at Institute of Automation, Chinese Academy of Sciences (2021.9-2024.6)
- **Intern Researcher** at MicroSoft Research Asia (2022.3-2022.8), and SenseTime Research (2021.6-2021.9, 2022.9-2023.8)
- **M.S. Thesis:** Dynamic Facial Editing based on Neural Radiance Fields
- **Interested Topics:** Face Analysis by Synthesis, 3D Human Motion Generation, Human-Object-Scene Interaction, Neural Rendering, Generative Models, and Embodied Intelligence

**Do not hesitate to drop me an email for
any possible collaboration! :D**

1 / 研究背景与意义

2 / 研究内容

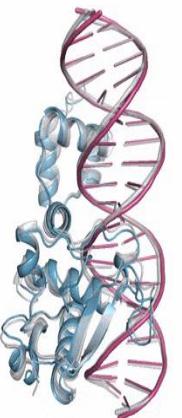
- 内容一 基于神经辐射场反演的人脸属性动态编辑
- 内容二 基于神经辐射场多模态融合的动态人脸编辑
- 内容三 基于神经辐射场稠密对应的动态人脸编辑

3 / 总结和展望

生成式人工智能 → 以人为中心的数字内容创作

生成式人工智能 (Generative AI)

7R6R

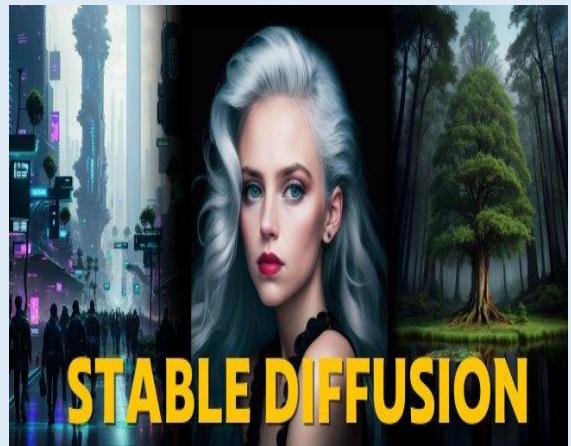


Ground truth shown in gray

分子结构预测



元宇宙



文生图模型



大语言模型

以人为中心 (Human-Centric) 的数字内容创作

Results identity 1

Original Dynamic Nerf [1]



"Photo of a bronze bust"



"Photo of a marble statue"

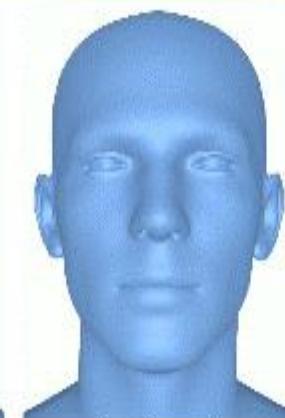


Teotla, Kartik, et al. "HQ3DAvatar: High Quality Controllable 3D Head Avatar." arXiv preprint arXiv:2303.14471 (2023).

No Personalization

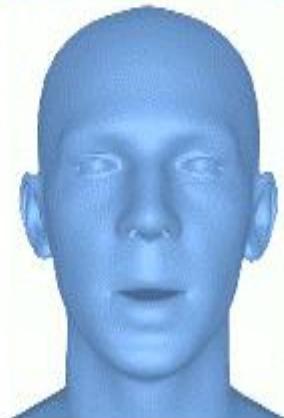


VOCA*
[Cuderio et al., CVPR 2019]



FaceFormer*
[Pan et al., CVPR 2022]

With Personalization



Ours



GT

应用背景：从“静态”人脸编辑迈向“动态”人脸编辑

静态
人脸编辑



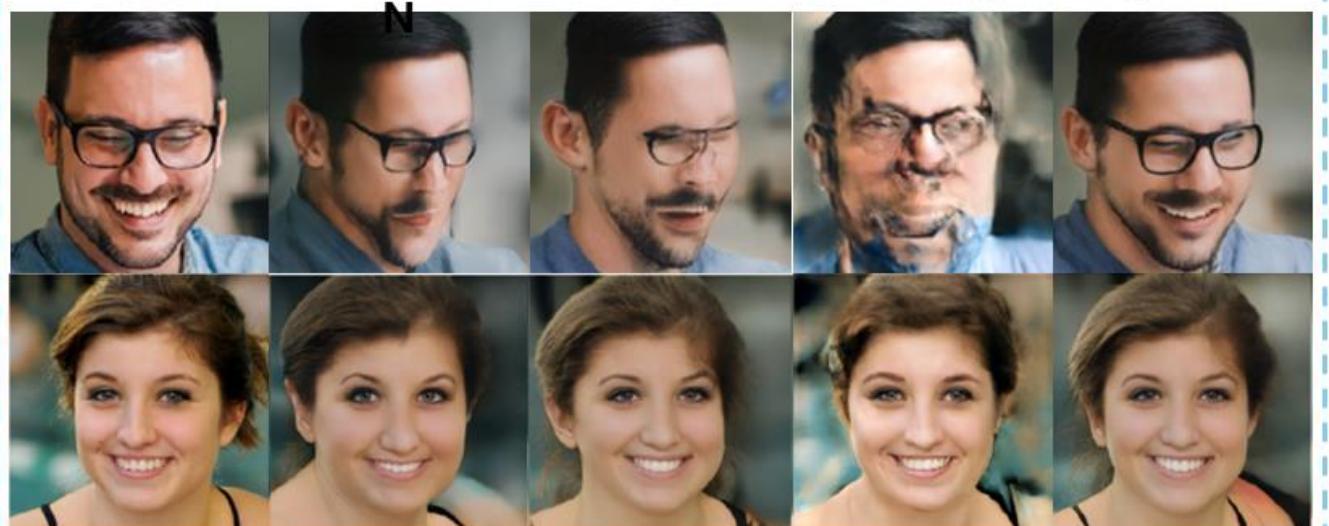
动态
人脸编辑



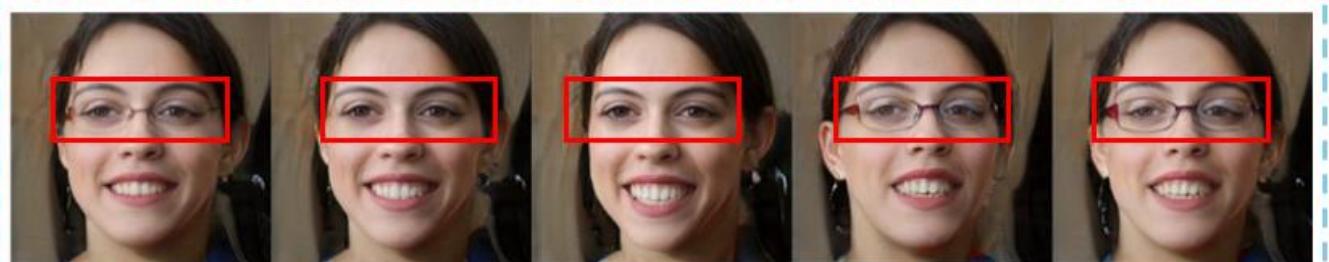
性能问题：缺乏三维感知导致较差的三维（空间）、时序（时间）一致性

1) 空间视角变化下编辑结果的三维一致性较差

原始图像 InterfaceGA GANSpace StyleRig StyleFlow



(a) 缺乏人脸三维感知的生成模型导致新视角生成下的伪影



(b) 视角变化下编辑结果出现不一致

2) 人脸视频编辑结果时序一致性难以保持



(a) 属性编辑帧间跳跃

原始帧

有上下文感知的新帧插入

原始帧



无上下文感知的新帧插入

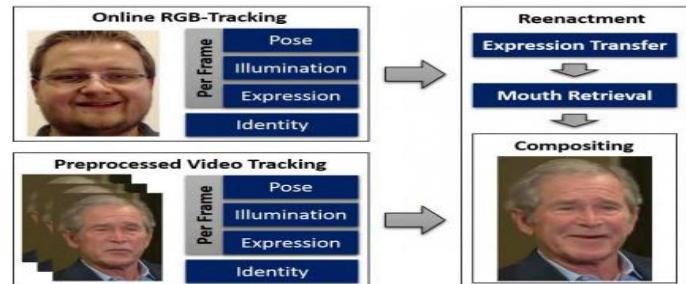


(b) 视频编辑帧间抖动

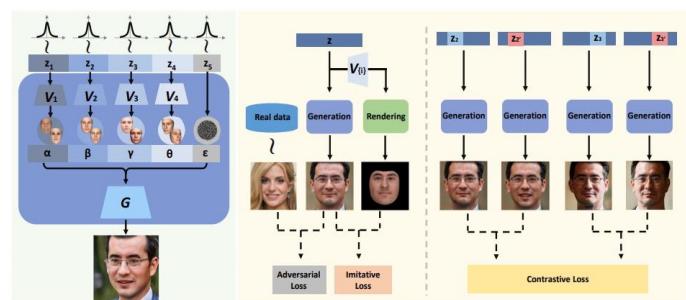
技术背景：基于三维人脸参数化模型的优劣与简单引入神经辐射场的问题

三维人脸参数化模型

利用模型表征作为条件信息

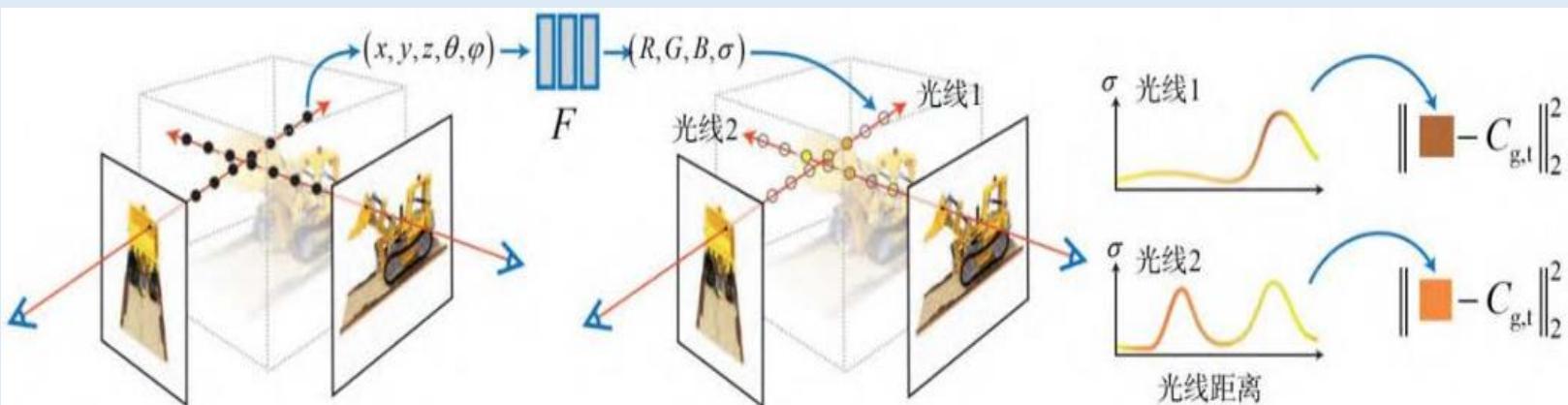


利用模型生成结果作为约束



- 有限3D数据导致多样性受限
- 对齐2D、3D表征的高昂训练成本

神经辐射场 (Neural Radiance Field, NeRF)



- 通过隐式函数来学习场景表示，利用体渲染 (Volume Rendering) 生成图像
- 学习从空间位置和视角方向的5D坐标到颜色和密度的映射

优势

- 新视角图像生成
- 高保真渲染效果

问题

- 过拟合场景信息，编辑操作破坏3D一致性
- 作为基础表征用于建模通用人脸难度大
- 训练成本高 (数据需求大、训练时间长)

1. 引入三维人脸参数化模型先验到二维生成模型的做法是存在局限性的

2. 替换为基于三维感知的神经辐射场表征具有潜力，但是如何应用于动态人脸编辑是需要进一步探索的

1 研究背景与意义

2 研究内容

- 内容一 基于神经辐射场反演的人脸属性动态编辑
- 内容二 基于神经辐射场多模态融合的动态人脸编辑
- 内容三 基于神经辐射场稠密对应的动态人脸编辑

3 总结和展望

研究内容：基于神经辐射场的动态人脸编辑

基于神经辐射场的动态人脸编辑

编辑内容

编辑方式

编辑方法

编辑效果

渲染视角编辑及时序拓展

属性标签驱动

神经辐射场反演

三维一致的人脸属性编辑

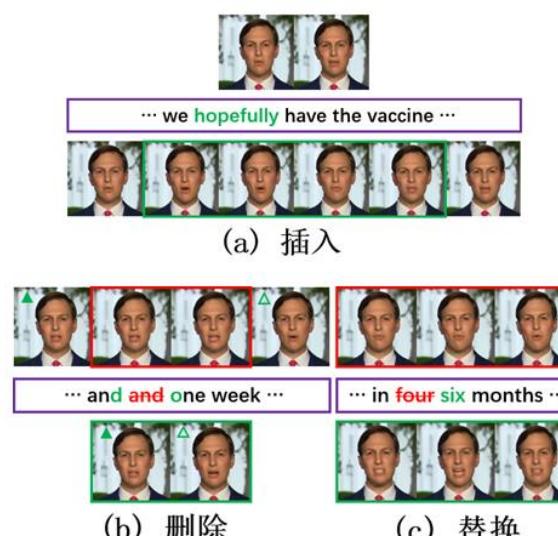


人脸运动编辑及时序平滑

文本驱动

神经辐射场多模态融合

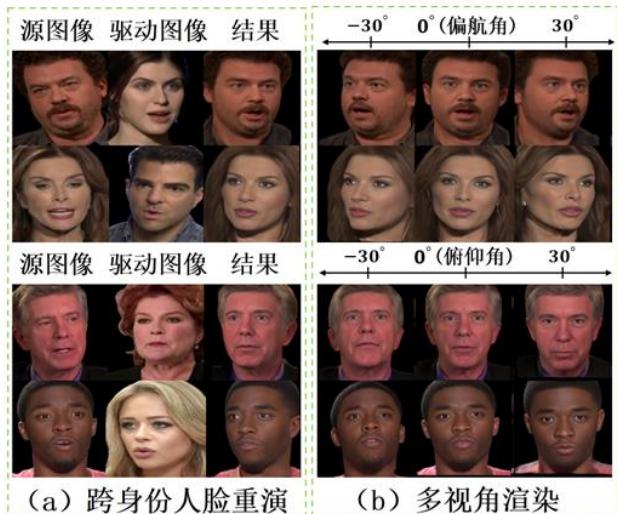
时序平滑的人脸唇部编辑



参考图像驱动

神经辐射场稠密对应

精细化的人脸表情、姿态编辑



核心问题

空间三维一致性 + 视频时序一致性

对应章节

第三章

第四章

第五章

1 研究背景与意义

2 研究内容

➤ 内容一 基于神经辐射场反演的
人脸属性动态编辑

➤ 内容二 基于神经辐射场多模态
融合的动态人脸编辑

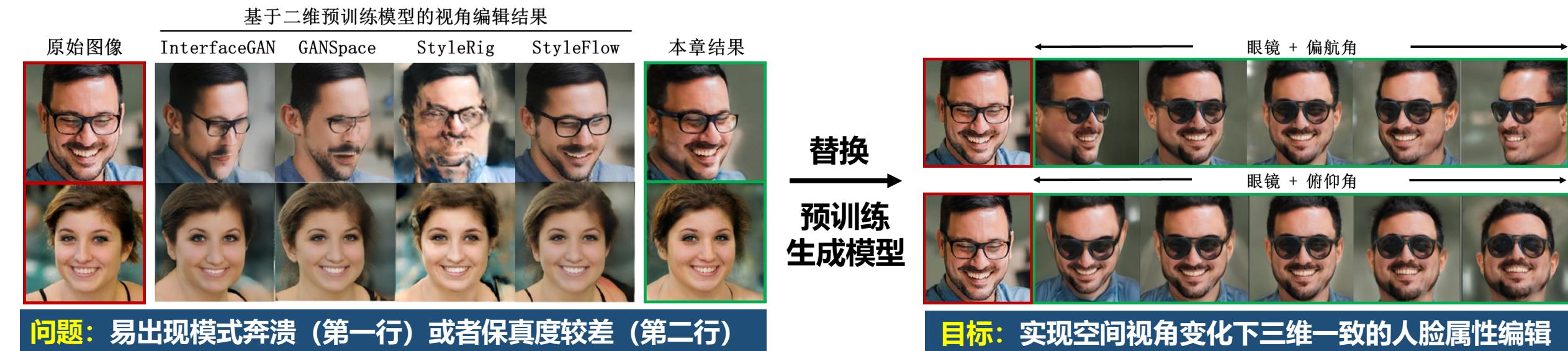
➤ 内容三 基于神经辐射场稠密对
应的动态人脸编辑

3 总结和展望

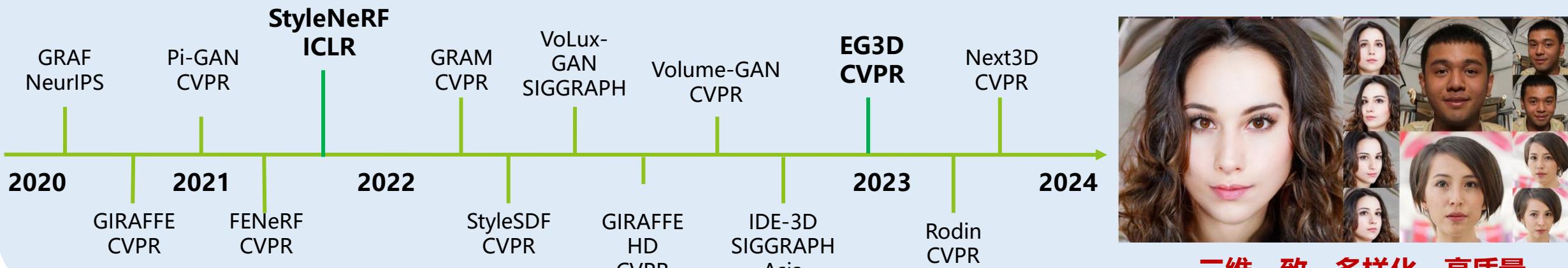


研究内容一：基于神经辐射场反演的人脸属性动态编辑

➤ 研究动机：探索适配基于三维感知的预训练生成模型的身份编码器与属性编辑器

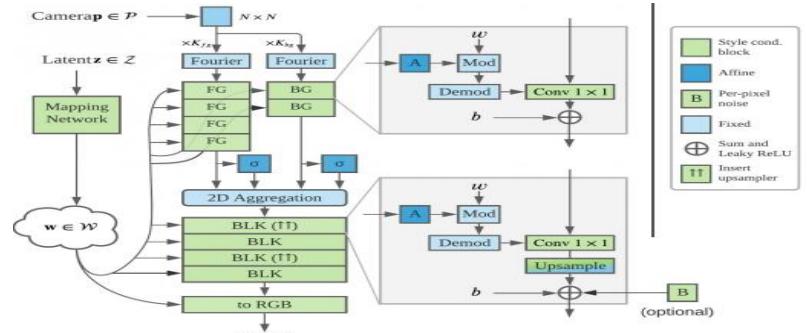
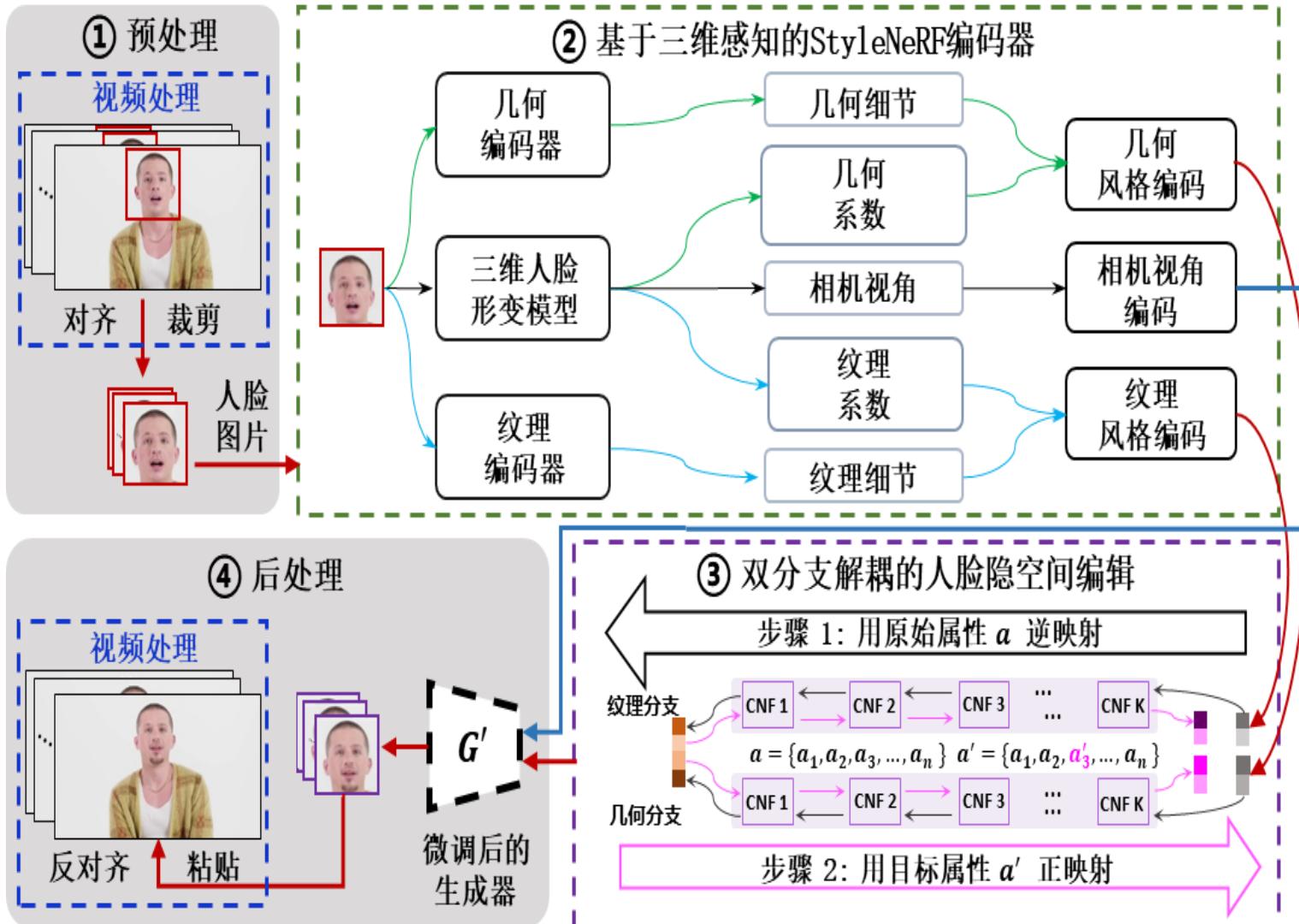


基于三维感知的预训练生成模型



研究内容一：基于神经辐射场反演的人脸属性动态编辑

➤ 研究方案：基于三维感知的人脸编码与双分支解耦的人脸属性编辑方法



预训练生成模型特点

- 浅层、深层编码分别影响几何、纹理
- $W +$ 隐空间破坏3D一致性

编码器 (Encoding)

- 参数化模型表征+可学习的细节编码器
- 生成几何、纹理、视角三个编码信息

编辑器 (Editing)

- 采取StyleFlow基础模型架构
- 采用双分支解耦几何、纹理编辑

视频序列帧处理和微调流程

研究内容一：基于神经辐射场反演的人脸属性动态编辑

➤ 实验结果：人脸图像反演与属性编辑



- ✓ 缓解基于二维生成模型的人脸属性编辑的身份偏移问题
- ✓ 利用三维信息感知的优势，实现多样化的人脸属性编辑，且更好地保留源身份

研究内容一：基于神经辐射场反演的人脸属性动态编辑

➤ 实验结果：人脸图像反演与属性编辑

身份保持定量测试

方法	年龄 ↑	眼镜 ↑	胡须 ↑	头发 ↑
StyleGAN + e4e	0.637	0.653	0.651	0.694
StyleNeRF + 3Da	0.794	0.791	0.803	0.811

✓ 能够在属性编辑中实现较好的身份保持

生成质量定量测试

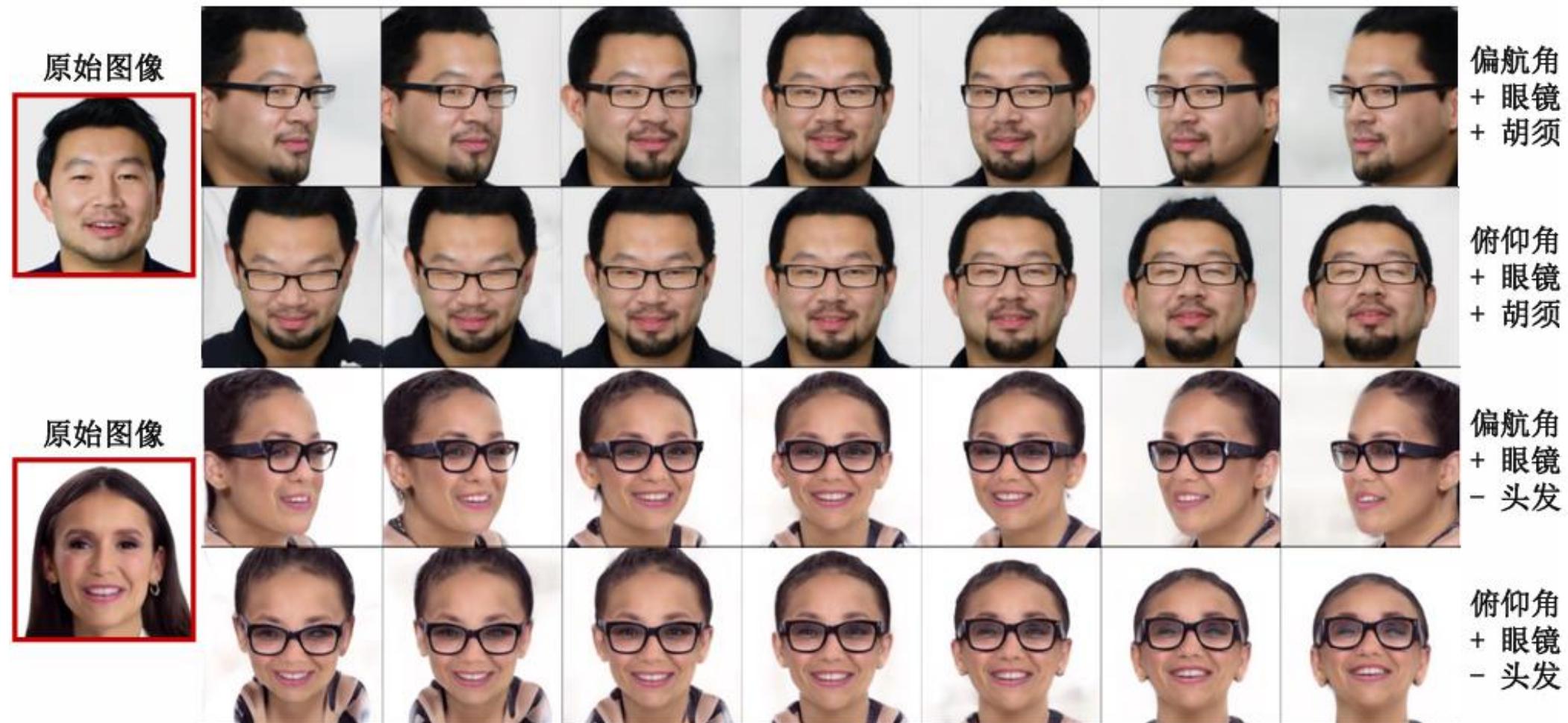
方法	PSNR ↑	SSIM ↑	VIF ↑	FVD ↓
StyleGAN + e4e	40.3	0.996	0.75	64.2
StyleNeRF + e4e	32.1	0.995	0.71	168.1
StyleGAN + 3Da	38.4	0.995	0.78	121.7
StyleNeRF + 3Da	41.2	0.997	0.79	51.3

✓ 能够在属性编辑中实现较好的生成质量



研究内容一：基于神经辐射场反演的人脸属性动态编辑

➤ 实验结果：人脸图像反演与属性编辑



- ✓ 实现多属性编辑与多视角生成
- ✓ 不会因为视角编辑而出现模式崩溃或者保真度下降的问题

研究内容一：基于神经辐射场反演的人脸属性动态编辑

➤ 实验结果：人脸视频反演与属性编辑

原始
视频序列



StyleGAN
+ e4e



StyleNeRF
+ e4e



StyleGAN
+ 3Da



StyleNeRF
+ 3Da



✓ 能够实现时序一致的人脸视频编码

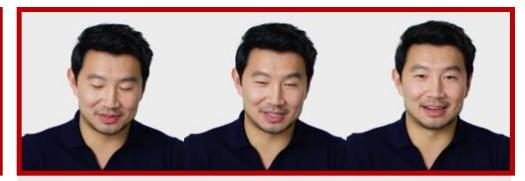
原始
视频序列



StyleGAN
+ e4e



StyleNeRF
+ 3Da



属性时序保持定量测试

方法	年龄 ↓	眼镜 ↓	胡须 ↓	头发 ↓
StyleGAN + e4e	0.498	0.533	0.502	0.497
StyleNeRF + 3Da	0.454	0.223	0.385	0.441

✓ 能够保持属性编辑的时序一致性

1

研究背景与意义

研究内容

- 内容一 基于神经辐射场反演的人脸属性动态编辑
- 内容二 基于神经辐射场多模态融合的动态人脸编辑
- 内容三 基于神经辐射场稠密对应的动态人脸编辑

总结和展望



研究内容二：基于神经辐射场多模态融合的动态人脸编辑

➤ 研究动机：基于神经辐射场的训练高效、渲染高质量、时序平滑的说话人脸视频编辑

ID #1 :插入 (使用本章方法编辑说话人脸视频样例)

	视频	文本
原始视频		... goodbye, Trump. Hello, vaccines. We're glad to see you ...
编辑结果		... goodbye, Trump. Hello, vaccines. We're so glad to see you ...

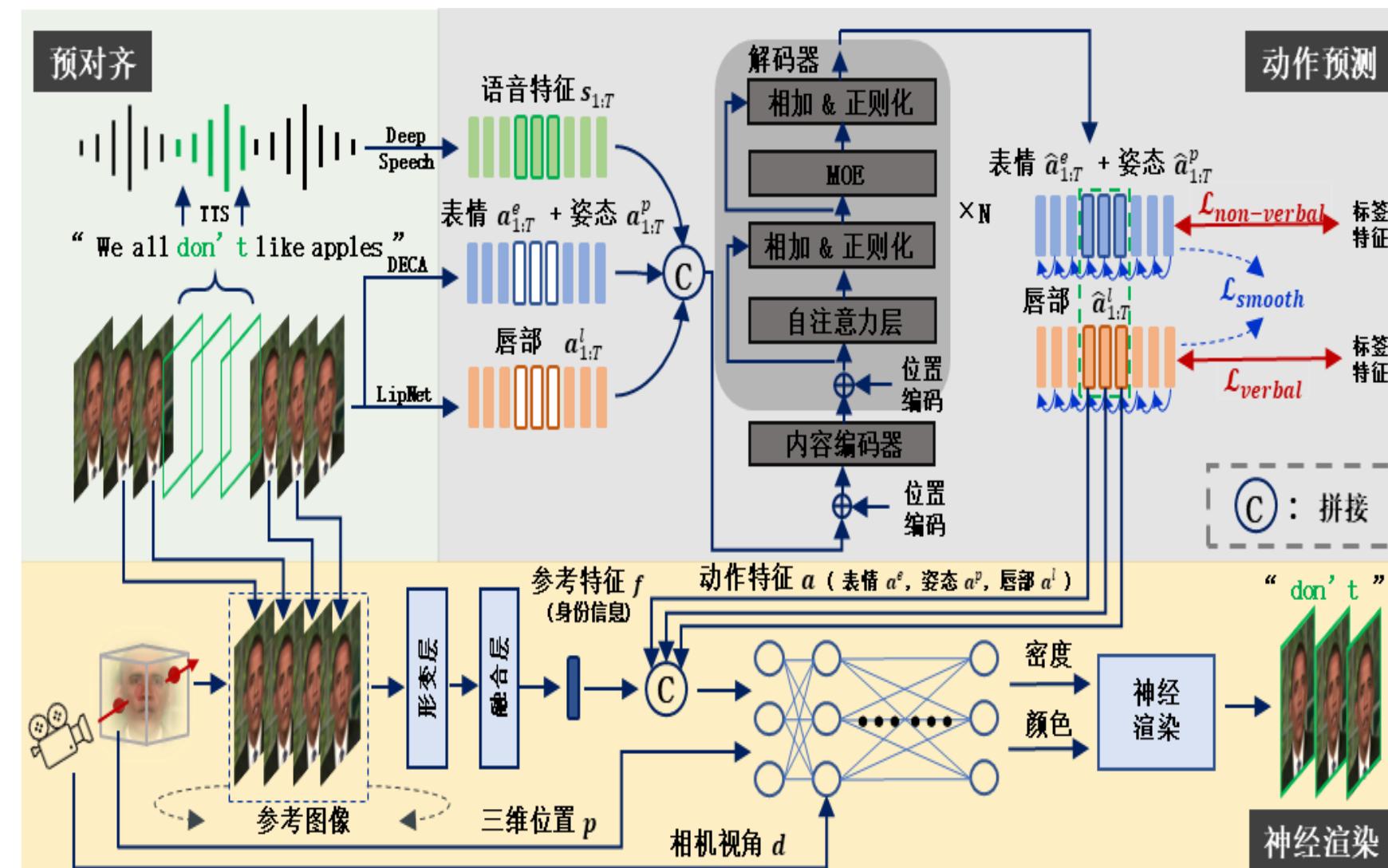
任务：对预录制说话人脸视频的单词级别增加、删除以及替换

对比方法	数据需求	时间需求	动作平滑	全头建模
ATVG [206]	-	-	✗	人脸
MakeitTalk [156]	-	-	✗	人脸
PC-AVS [17]	-	-	✗	部分人头
Wav2Lip [207]	-	-	✗	唇部
AD-NeRF [161]	5 min	36 h	✗	✓
SSP-NeRF [162]	5 min	36 h	✗	✓
DFRF [208]	15 s	2 h	✗	✓
Fried 等人 [196]	1 h	42 h	✗	人脸下部
Yao 等人 [197]	2.5 min	17 h	✗	人脸下部
本章方法	15 s	25 min	✓	✓

要求：训练数据少、训练时间短、时序平滑以及全头建模

研究内容二：基于神经辐射场多模态融合的动态人脸编辑

➤ 研究方案：基于多模态信息融合与上下文感知的人脸视频编辑方法



文本-语音-视觉预对齐

- 解耦语言（语音、唇部相关）与非语言（头部姿态和表情）相关特征

动作预测

- 采用基于上下文感知的输入格式，保留“说话风格”，用上下文先验提升时序预测平滑性
- 采用非自回归序列建模，提升训练和推理效率
- 在大规模说话人脸数据集上训练，学习更具泛化性和准确性的唇部预测

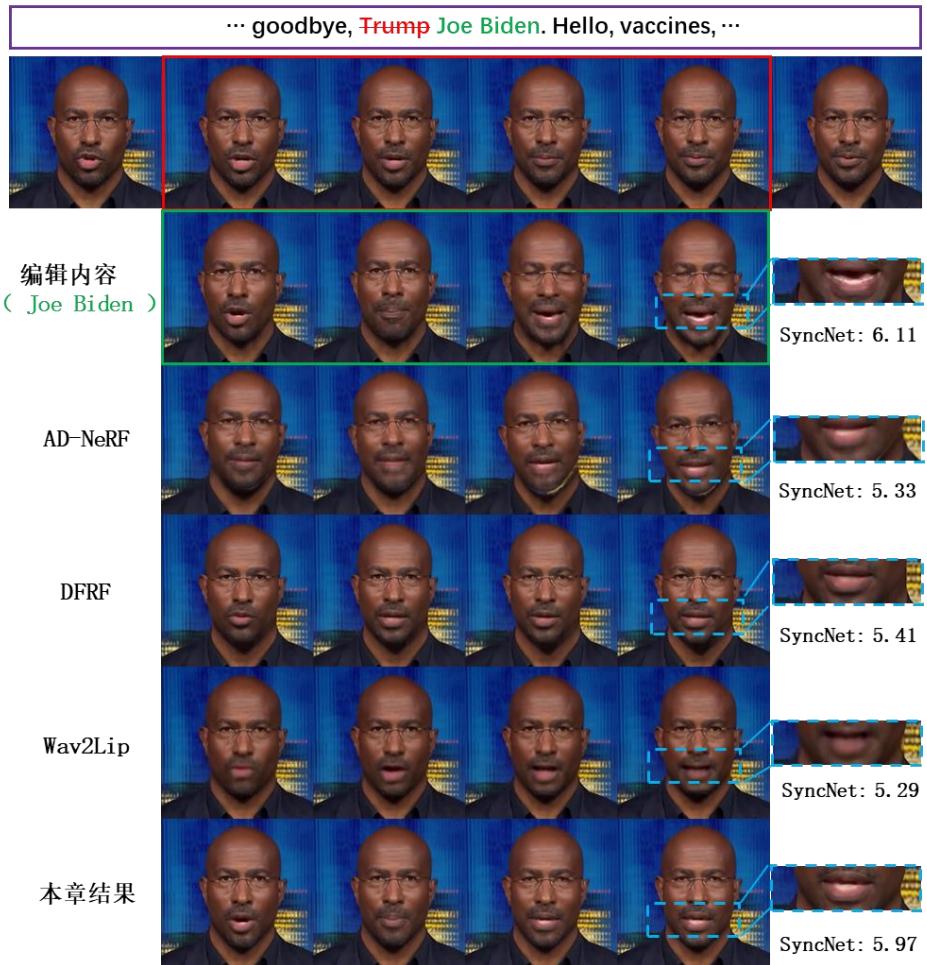
神经渲染

- 采用显式动作条件控制画面生成，缓解“语音-动作”一对多映射
- 采用“预训练+微调”策略，高效获取特定身份说话人渲染模型

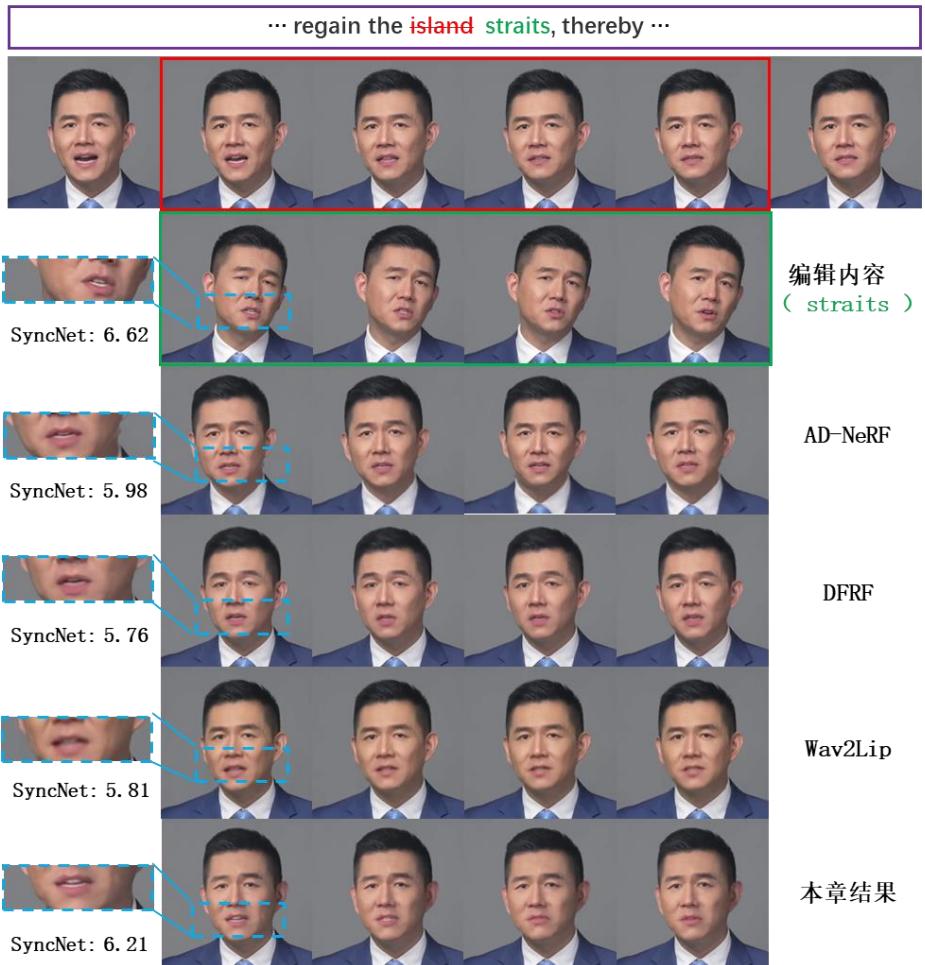
研究内容二：基于神经辐射场多模态融合的动态人脸编辑

➤ 实验结果：质量、准确性以及平滑性

测试语音对应的**唇部运动与原始帧非常不同的替换编辑**



测试视频的**头部姿态与原始帧非常不同的替换编辑**



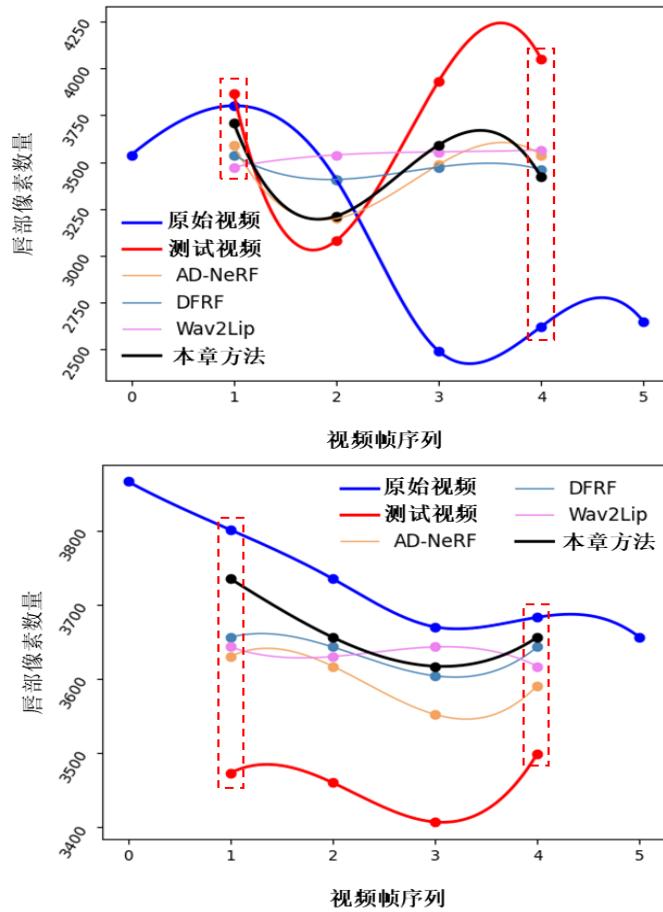
✓ 保持和衔接帧较为**平滑的过渡** (平滑性-准确性的平衡)

✓ 表明对语言和非语言表征**显式、解耦控制**的重要性

研究内容二：基于神经辐射场多模态融合的动态人脸编辑

➤ 实验结果：质量、准确性以及平滑性

平滑性测试可视化（唇部像素数量）



✓ 首尾衔接处更为平滑

说话人脸视频编辑定量评估

	LMD [209]↓	SyncNet [237]↑	PSNR↑	SSIM↑	LPIPS↓	LF [238]↓	LDS↓
原始视频	0.0	6.24	N/A	1.0	0.0	0.12	2.23
AD-NeRF [161]	5.63	5.24	29.75	0.842	0.045	0.89	2.77
DFRF (5s/10k) [208]	7.16	5.33	28.21	0.814	0.092	0.91	2.84
DFRF (15s/40k) [208]	5.03	6.02	30.98	0.871	0.041	0.80	2.45
Wav2Lip [207]	5.97	6.16	26.12	0.791	0.087	0.85	2.66
本章方法 (5s/10k)	6.46	5.52	30.23	0.838	0.089	0.88	2.73
本章方法 (15s/40k)	4.97	6.11	33.79	0.879	0.039	0.79	2.36

说话人脸视频编辑用户实验

	5	4	3	2	1	平均↑	真实↑
原始视频	45	36	11	6	2	4.160	81.0
AD-NeRF	7	12.6	18.5	22.1	39.8	2.249	19.6
DFRF (5s/10k)	6	8.4	23.9	30.1	31.6	2.271	14.4
DFRF (15s/40k)	10.9	11.2	20.3	22.8	34.8	2.406	22.1
Wav2Lip	9.8	10.3	20.1	23	36.8	2.333	20.1
本章方法 (5s/10k)	8	9.2	25.4	31.9	25.5	2.423	17.2
本章方法 (15s/40k)	12	14.7	28.1	36	9.2	2.843	26.7

✓ 生成质量好
✓ 音画一致性好
✓ 编辑平滑性好

研究内容二：基于神经辐射场多模态融合的动态人脸编辑

➤ 实验结果：效率-质量的平衡



训练数据及时间的定量测试

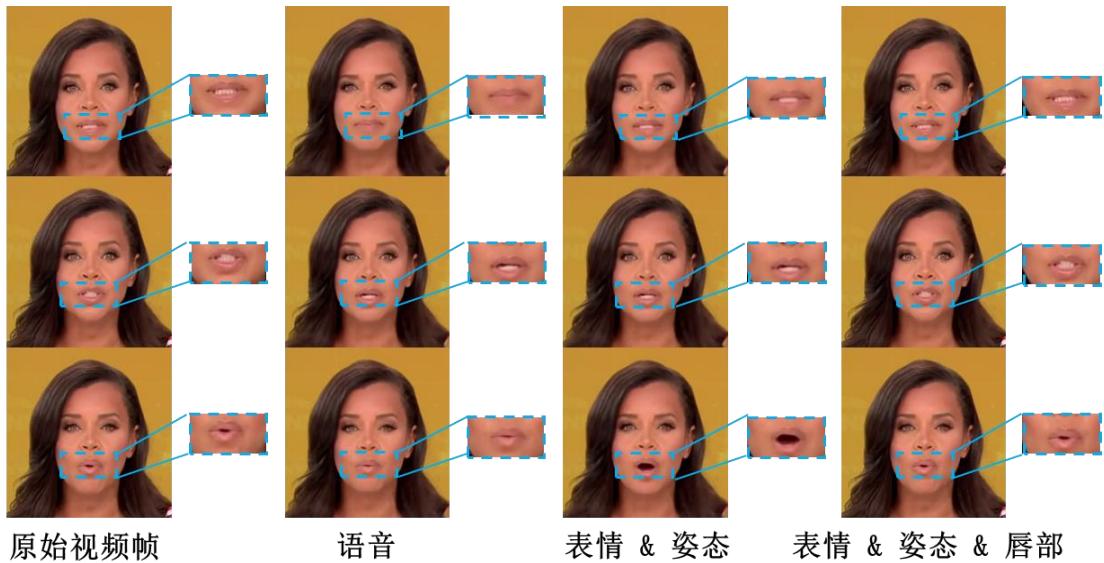
	DFRF	本章	LMD [209]↓		SyncNet [237]↑		PSNR↑		SSIM↑		LPIPS↓	
			DFRF	本章	DFRF	本章	DFRF	本章	DFRF	本章	DFRF	本章
5 s	0.1 k	8.34	6.99	4.88	5.40	27.33	28.01	0.725	0.802	0.168	0.160	
	10 k	7.02	6.40	5.39	5.65	28.21	30.23	0.811	0.837	0.092	0.089	
	40 k	6.44	5.72	5.64	5.68	28.92	31.00	0.829	0.848	0.071	0.066	
10 s	0.1 k	8.01	6.81	4.97	5.45	27.89	28.62	0.734	0.811	0.152	0.149	
	10 k	6.85	6.08	5.48	5.67	28.67	30.68	0.845	0.849	0.089	0.070	
	40 k	5.42	5.08	6.11	6.12	30.43	32.85	0.852	0.870	0.045	0.041	
15 s	0.1 k	7.87	6.65	5.11	5.61	28.02	29.75	0.750	0.825	0.141	0.138	
	10 k	6.04	5.39	5.88	6.06	29.75	31.11	0.849	0.891	0.065	0.059	
	40 k	5.03	4.94	6.19	6.17	30.98	33.72	0.881	0.897	0.041	0.039	

- ✓ 采用显式、解耦的条件信息，大大减少将语音与非语言视觉内容（如头部姿态和表情）对齐的数据要求
- ✓ 以更少的训练数据和迭代次数，快速拟合说话人脸视频中画面占比小但是变化多的部分（如眼睛和嘴巴）

研究内容二：基于神经辐射场多模态融合的动态人脸编辑

➤ 实验结果：不同条件信息以及上下文感知的消融实验

采用不同的条件信息的消融实验



是否引入动作上下文以及相邻帧参考图像

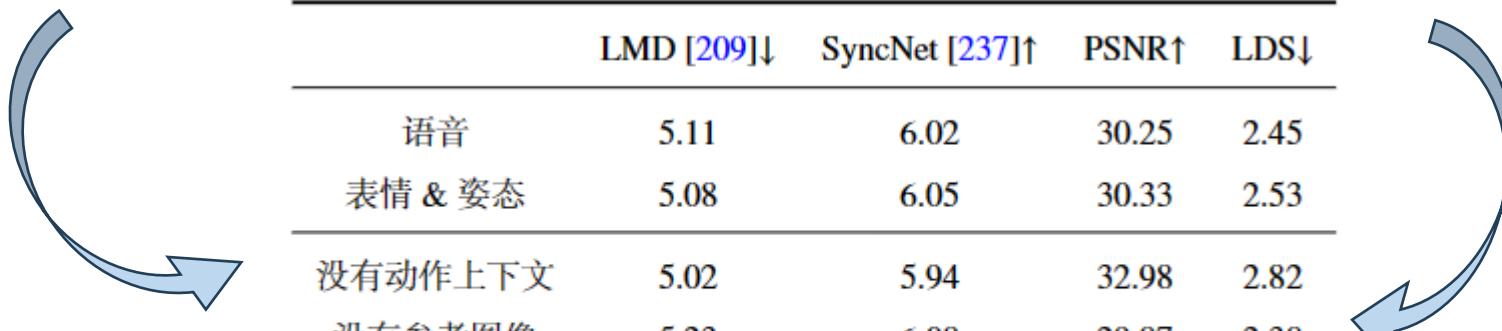
没有
动作上下文

没有
相邻帧
参考图像

本章方法



- ✓ 解耦得更细
- ✓ 控制得更好



- ✓ 缺失上下文（仅用语音驱动）会引起头部姿态帧间抖动
- ✓ 使用任意帧而非前/后帧，会产生光影抖动（尤其是嘴巴部分）

1 研究背景与意义

2 研究内容

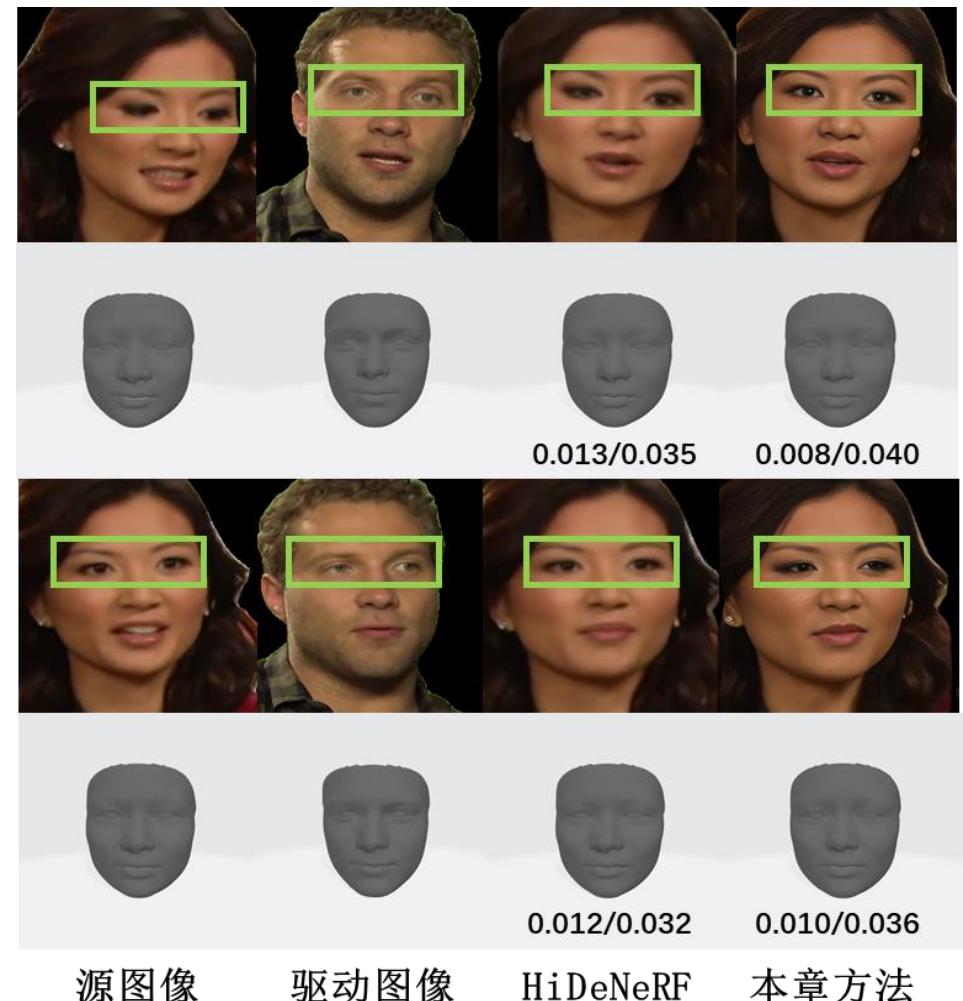
- 内容一 基于神经辐射场反演的人脸属性动态编辑
- 内容二 基于神经辐射场多模态融合的动态人脸编辑
- 内容三 基于神经辐射场稠密对应的动态人脸编辑

3 总结和展望



研究内容三：基于神经辐射场稠密对应的动态人脸编辑

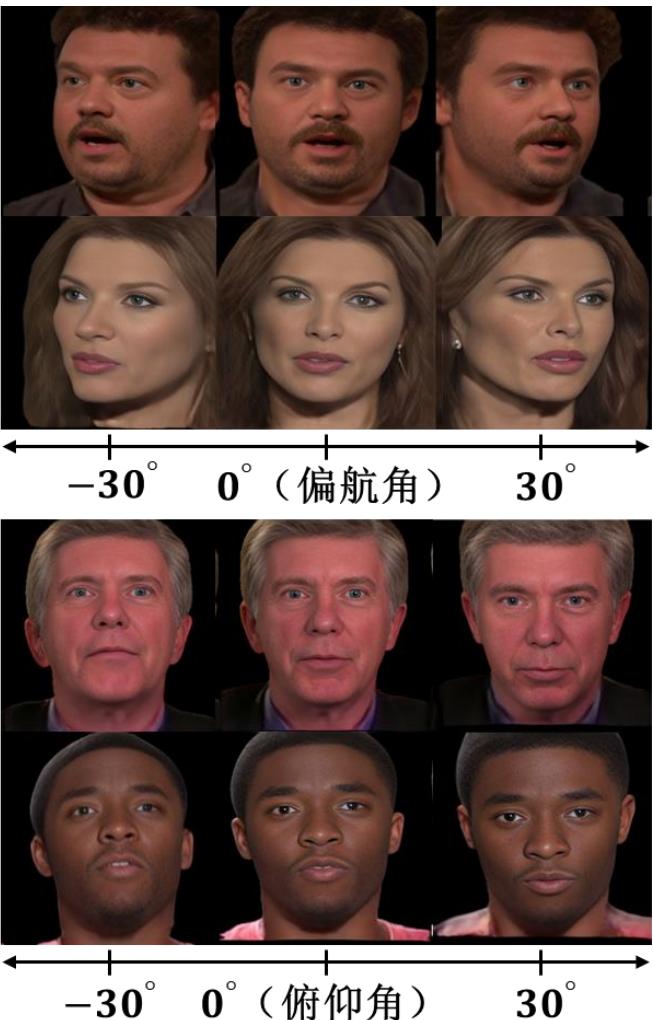
➤ 研究动机：实现无需三维先验的神经辐射场稠密对应与精细化动作迁移



问题：精细化程度不高（比如眼睛部分）

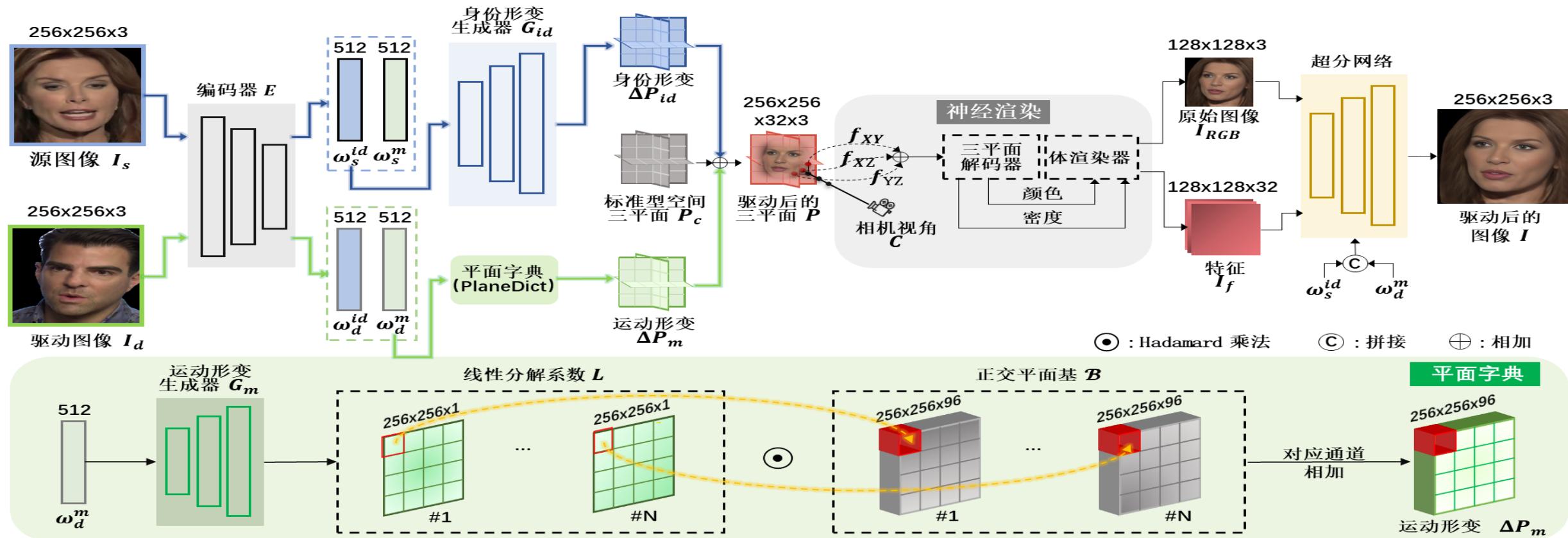


任务：单图人脸重演与多视角渲染



研究内容三：基于神经辐射场稠密对应的动态人脸编辑

➤ 研究方案：基于神经辐射场稠密对应的人脸运动迁移方法



基础表征：三平面 (Tri-Plane)

- 结合StyleGAN生成器，构建多身份、多姿态与多表情且三维一致的隐空间
- 取得信息密度与照片真实度之间的平衡

人脸身份与运动形变解耦

- 将一个人脸表示三部分：标准型空间三平面、身份形变三平面与运动形变三平面

神经辐射场稠密对应

- 将运动条件映射为一组正交平面基的线性加权和
- 平面基正交性保持：将具有相同索引的向量进行QR分解

研究内容三：基于神经辐射场稠密对应的动态人脸编辑

➤ 实验结果：同身份人脸运动编辑



✓ 能够实现同身份人脸重演

同身份重演生成质量测试

	SSIM↑	PSNR↑	LPIPS↓
FOMM	0.690	19.2	0.112
DaGAN	0.807	23.2	0.088
FNeVR	0.901	21.1	0.092
ROME	0.833	21.6	0.085
HiDeNeRF	0.862	21.9	0.084
本章方法	0.870	22.1	0.079

✓ 取得更好的重演生成质量

身份保真、头部姿态、表情、多视角身份保真以及眼神测试

	CSIM↑	AUCON↑	PRMSE↓	AVD↓	ET↓
FOMM	0.837	0.872	2.88	0.021	1.98
DaGAN	0.875	0.921	1.79	0.016	4.08
FNeVR	0.880	0.929	2.22	0.016	2.01
ROME	0.906	0.918	1.68	0.013	5.28
HiDeNeRF	0.931	0.956	1.66	0.010	5.44
本章方法	0.946	0.961	1.60	0.009	1.72

✓ 取得更好的身份保真度和动作准确度

研究内容三：基于神经辐射场稠密对应的动态人脸编辑

➤ 实验结果：跨身份人脸运动编辑



跨身份重演的定量测试

VoxCeleb1						
	CSIM↑	AUCON↑	PRMSE↓	FID↓	AVD↓	ET↓
FOMM	0.748	0.752	3.66	86	0.044	6.08
DaGAN	0.790	0.880	3.06	87	0.036	6.16
FNeVR	0.812	0.884	3.32	82	0.041	6.10
ROME	0.833	0.871	2.64	76	0.016	7.08
HiDeNeRF	0.876	0.917	2.62	57	0.012	7.02
本章方法	0.911	0.928	2.50	49	0.011	5.18

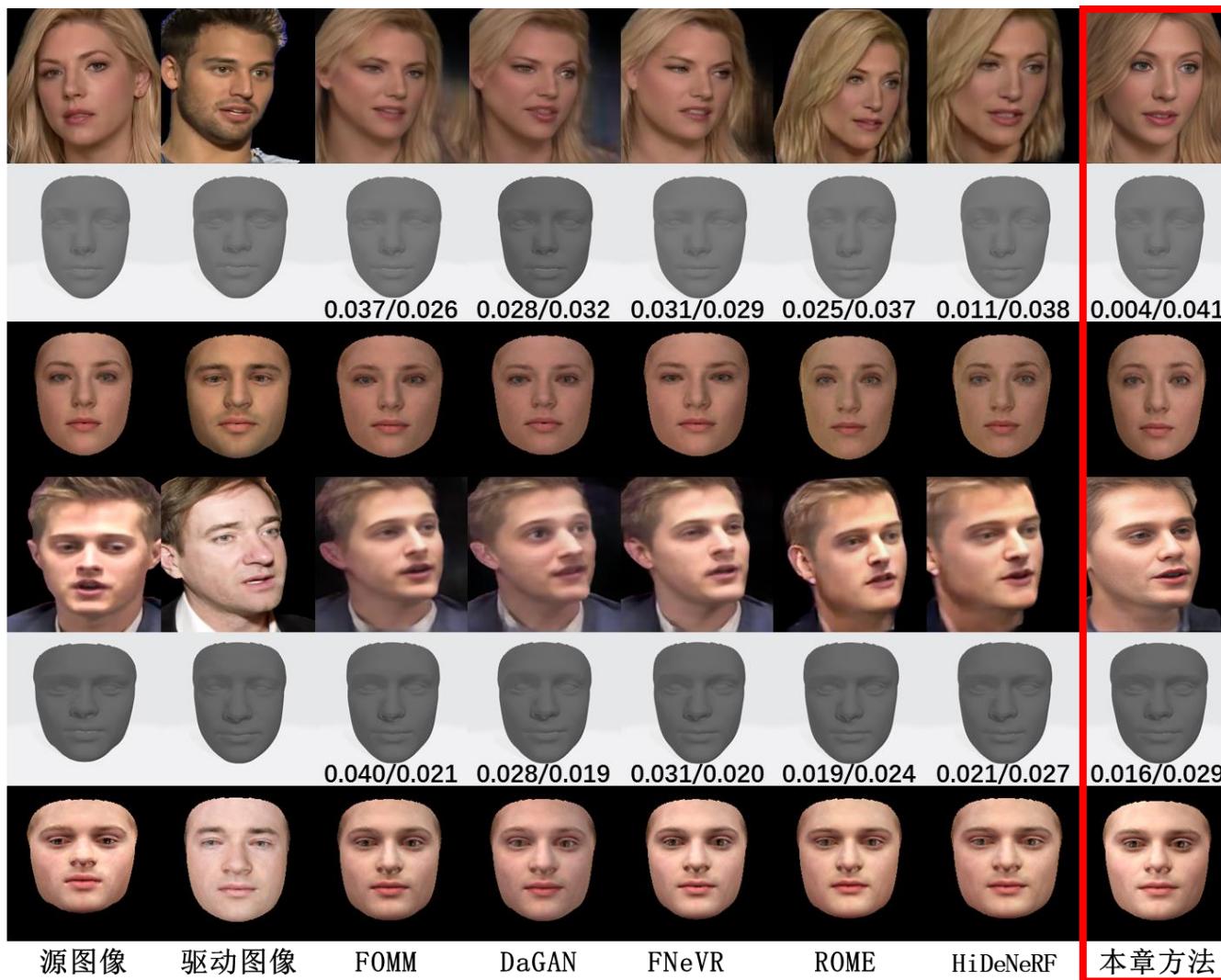
VoxCeleb2						TalkingHead-1KH							
	CSIM↑	AUCON↑	PRMSE↓	FID↓	AVD↓	ET↓		CSIM↑	AUCON↑	PRMSE↓	FID↓	AVD↓	ET↓
FOMM	0.680	0.707	4.16	85	0.047	6.23	FOMM	0.723	0.741	3.71	76	0.039	6.17
DaGAN	0.693	0.815	3.93	86	0.040	6.62	DaGAN	0.766	0.872	2.98	73	0.035	6.59
FNeVR	0.699	0.829	3.90	84	0.047	5.99	FNeVR	0.775	0.879	3.39	73	0.037	6.03
ROME	0.710	0.821	3.08	76	0.019	7.29	ROME	0.781	0.864	2.66	68	0.017	6.97
HiDeNeRF	0.787	0.889	2.91	61	0.014	7.30	HiDeNeRF	0.828	0.901	2.60	52	0.011	7.09
本章方法	0.790	0.894	2.83	58	0.012	5.33	本章方法	0.831	0.912	2.55	49	0.010	5.42

- ✓ 能够实现跨身份人脸重演，保留源身份的同时，实现不同人脸三平面之间的精细运动迁移
- ✓ 直接从驱动图像中提取运动特征，不仅可以实现细粒度运动控制，还可以处理眼睛等特殊区域

研究内容三：基于神经辐射场稠密对应的动态人脸编辑

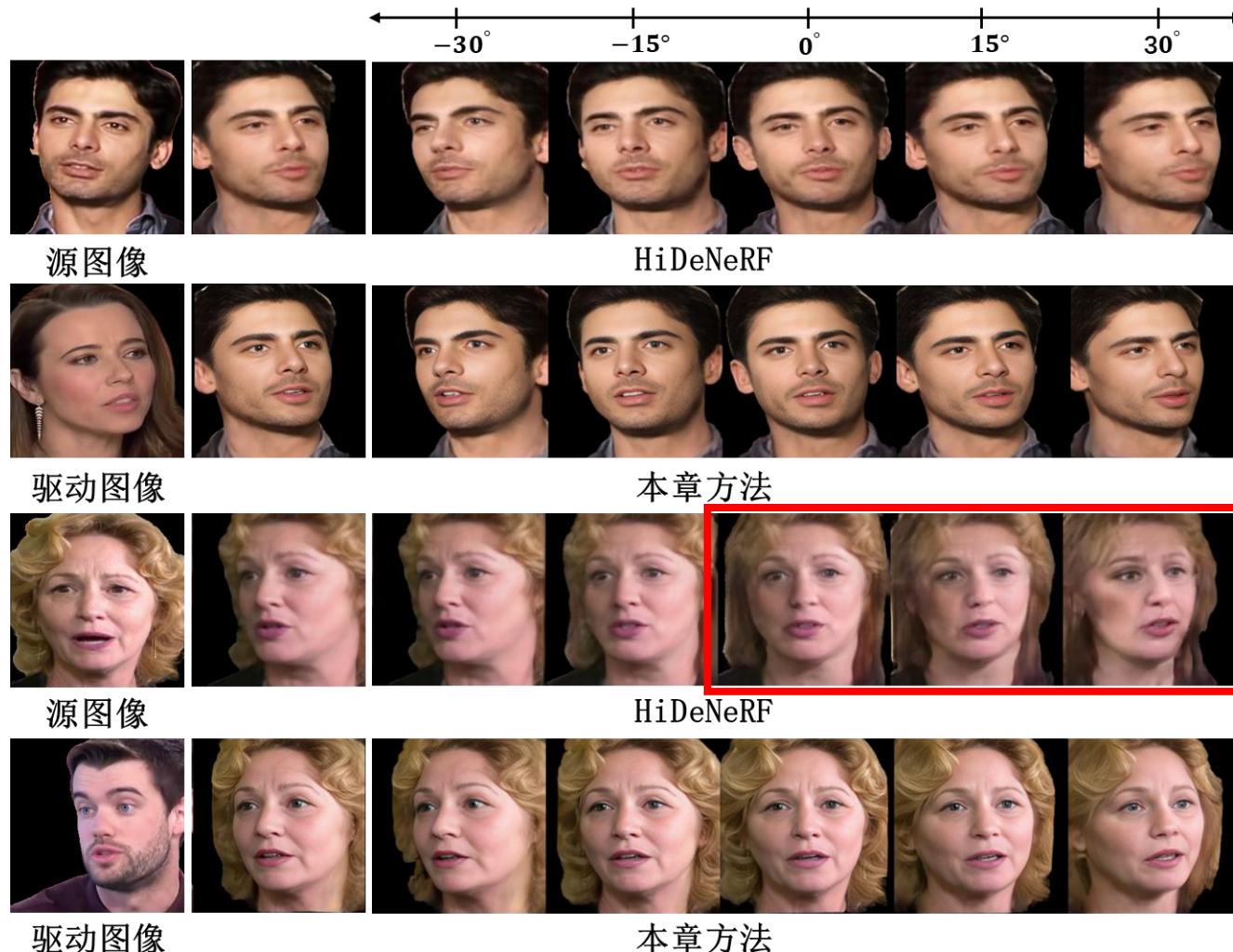
➤ 实验结果：跨身份人脸运动编辑

跨身份人脸重演的三维重建衡量（身份保真度）



研究内容三：基于神经辐射场稠密对应的动态人脸编辑

➤ 实验结果：多视角渲染与消融实验



✓ 能够实现跨身份人脸重演与多视角渲染

多视角渲染定量测试

	CSIM↑	AUCON↑	PRMSE↓	AVD↓
HiDeNeRF	0.829	0.864	3.78	0.014
本章方法	0.840	0.881	3.53	0.008

消融实验

	CSIM↑	AUCON↑	PRMSE↓	AVD↓	ET↓
不用 PlaneDict	0.763	0.809	3.10	0.035	7.58
使用 PlaneDict (5)	0.679	0.718	3.93	0.058	9.92
使用 PlaneDict (10)	0.802	0.824	3.18	0.038	7.36
使用 PlaneDict (15)	0.899	0.871	2.92	0.019	6.69
使用 PlaneDict (20)	0.911	0.928	2.50	0.011	5.18

✓ 找到质量和优化难度的平衡点

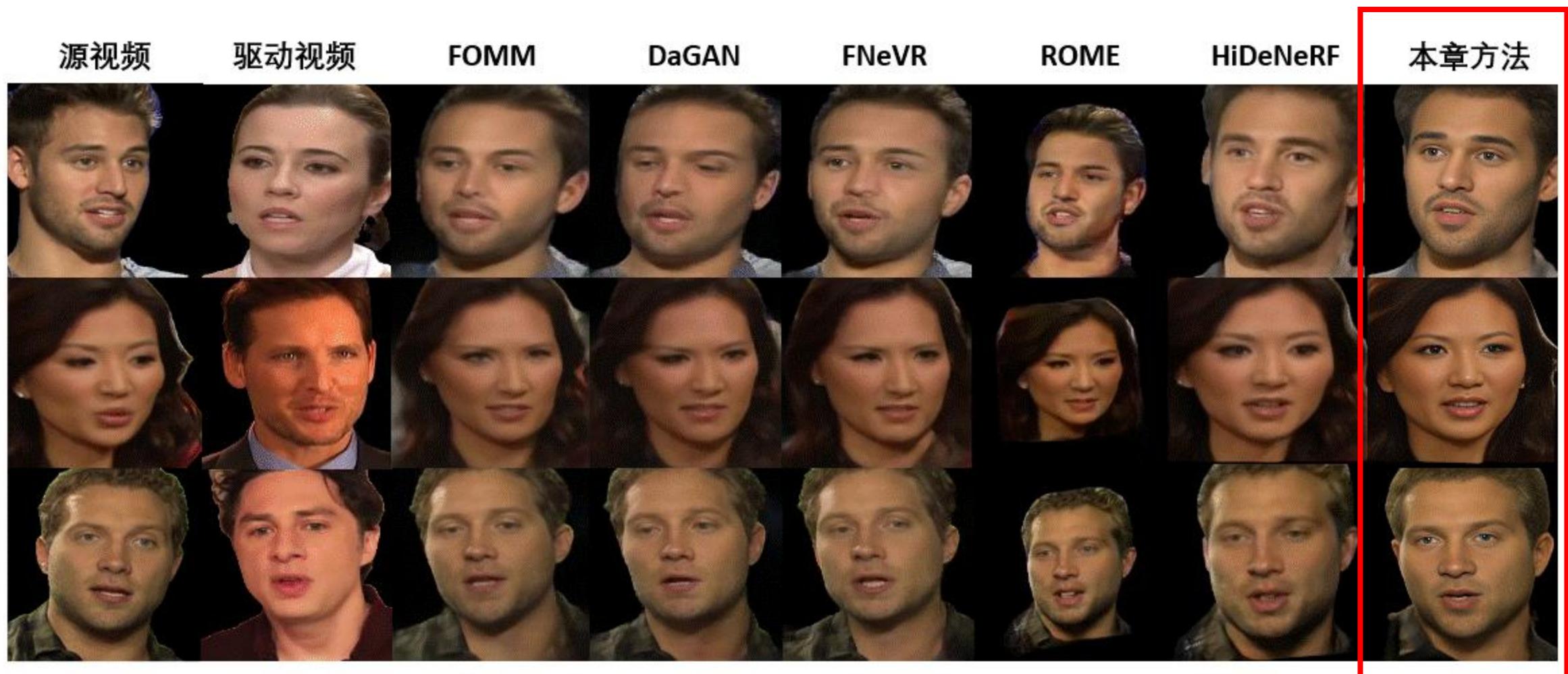
研究内容三：基于神经辐射场稠密对应的动态人脸编辑

➤ 实验结果：跨身份重演视频



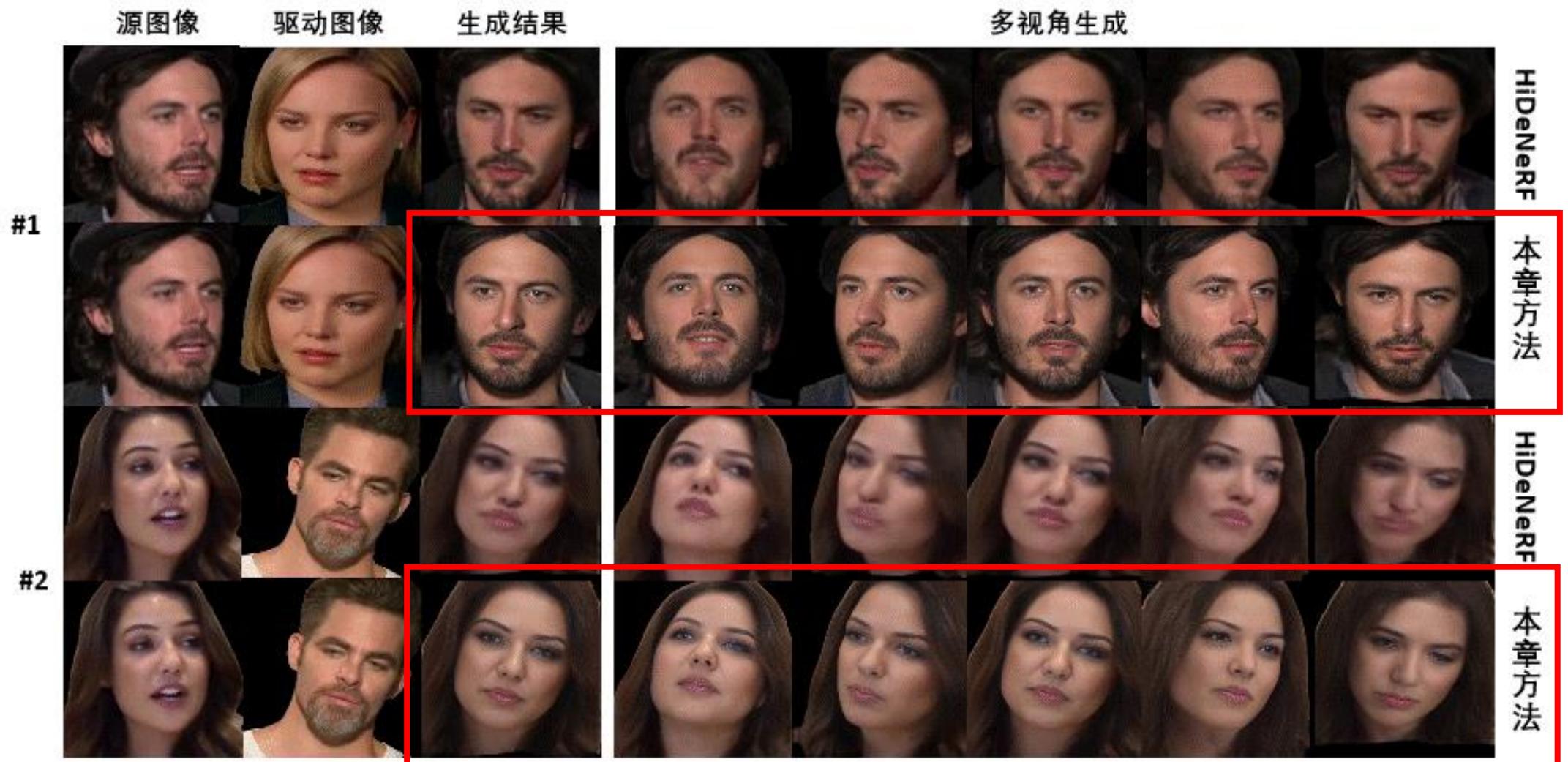
研究内容三：基于神经辐射场稠密对应的动态人脸编辑

➤ 实验结果：跨身份重演视频



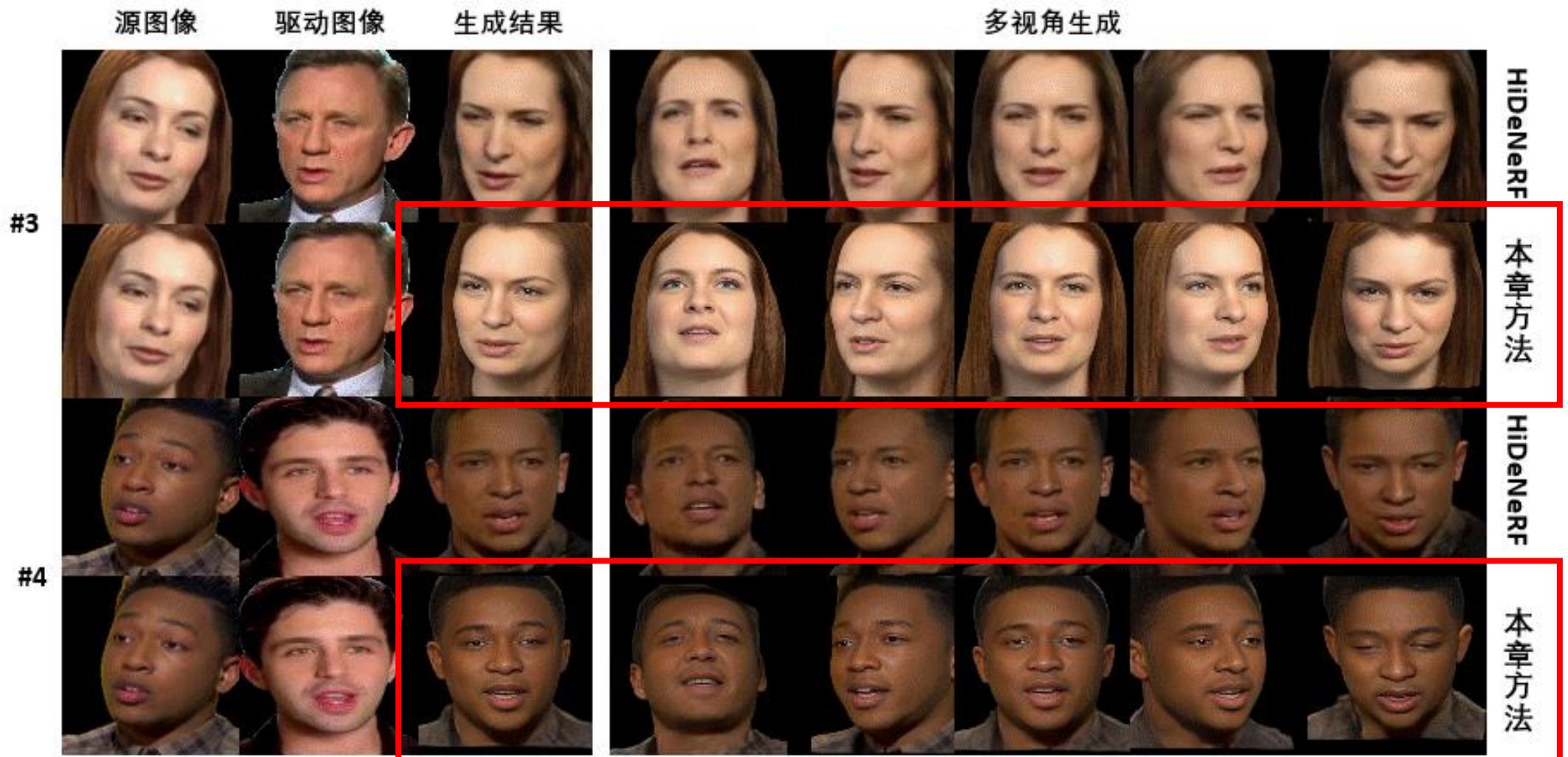
研究内容三：基于神经辐射场稠密对应的动态人脸编辑

➤ 实验结果：跨身份重演与多视角渲染



研究内容三：基于神经辐射场稠密对应的动态人脸编辑

➤ 实验结果：跨身份重演与多视角渲染

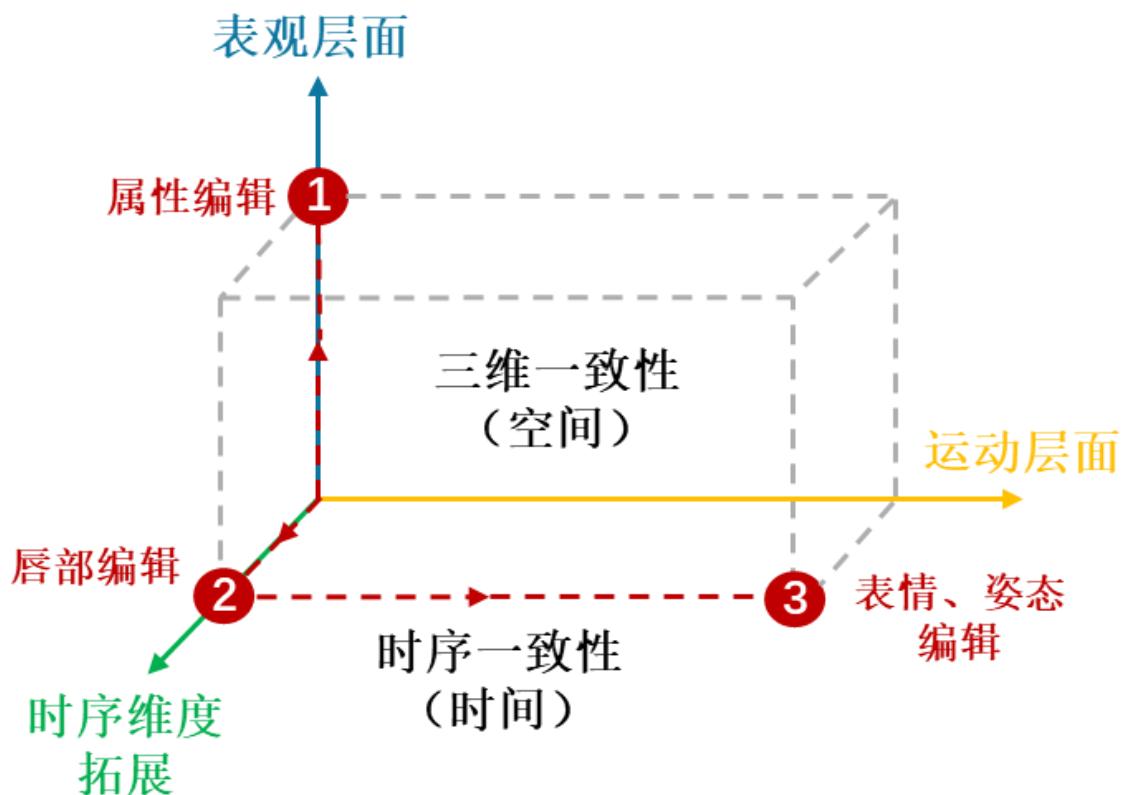
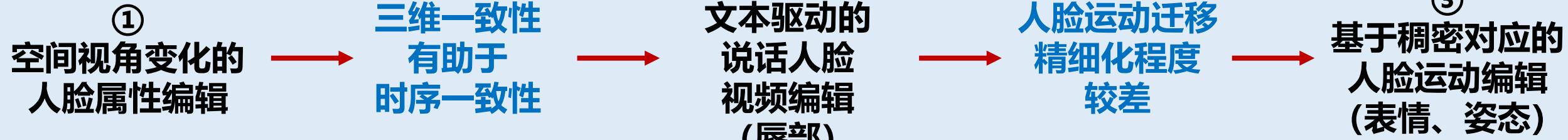


1 研究背景与意义

2 研究内容

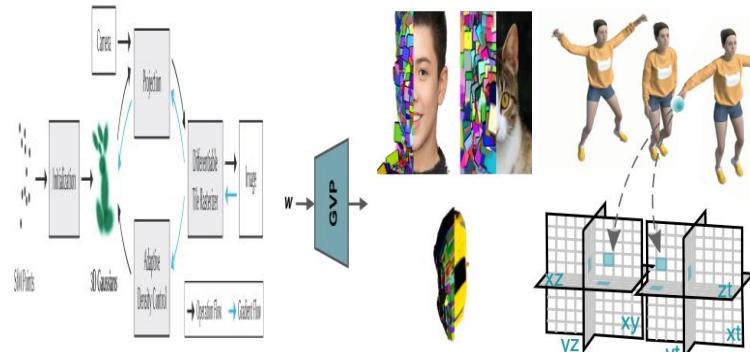
- 内容一 基于神经辐射场反演的人脸属性动态编辑
- 内容二 基于神经辐射场多模态融合的动态人脸编辑
- 内容三 基于神经辐射场稠密对应的动态人脸编辑

3 总结和展望



- ① **神经辐射场反演问题**
提出一种基于三维感知的编码器适配神经辐射场生成模型以提高编辑的三维一致性
- ② **神经辐射场多模态融合问题**
提出一种基于文本的说话人脸视频编辑框架以高效、灵活编辑视频内容
- ③ **神经辐射场稠密对应问题**
提出首个无需三维人脸先验的稠密对应方法以实现单图输入的多视角人脸重演

探索新一代三维隐式表征

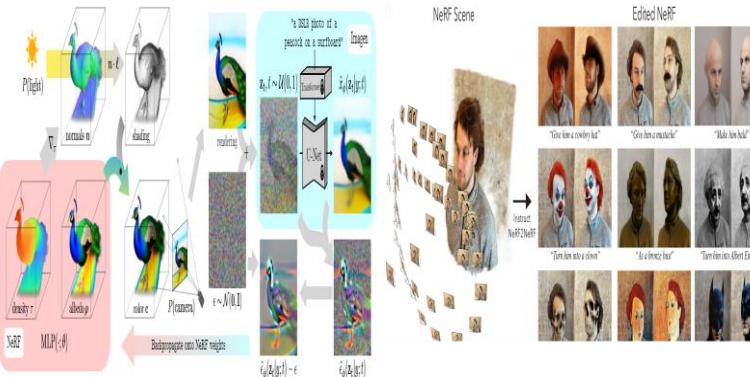


高斯泼溅

GVP

K-Planes

探索文生图扩散模型先验



迭代数据集更新
Instruct-NeRF2NeRF

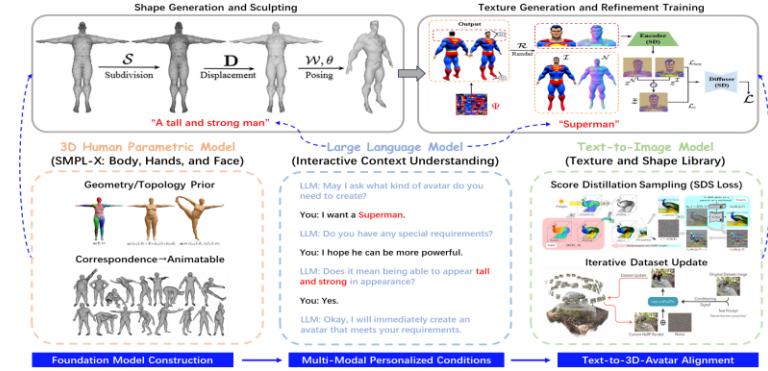
局限性

- 多视角渲染质量的优势仍无法取代传统三维建模的便捷性与可扩展性

未来研究

- 减少数据依赖
- 减少拟合所需训练时间
- 可动画能力
- 适配现有工业渲染管线

探索多模态、上下文学习



支持人机交互、多轮迭代的编辑框架

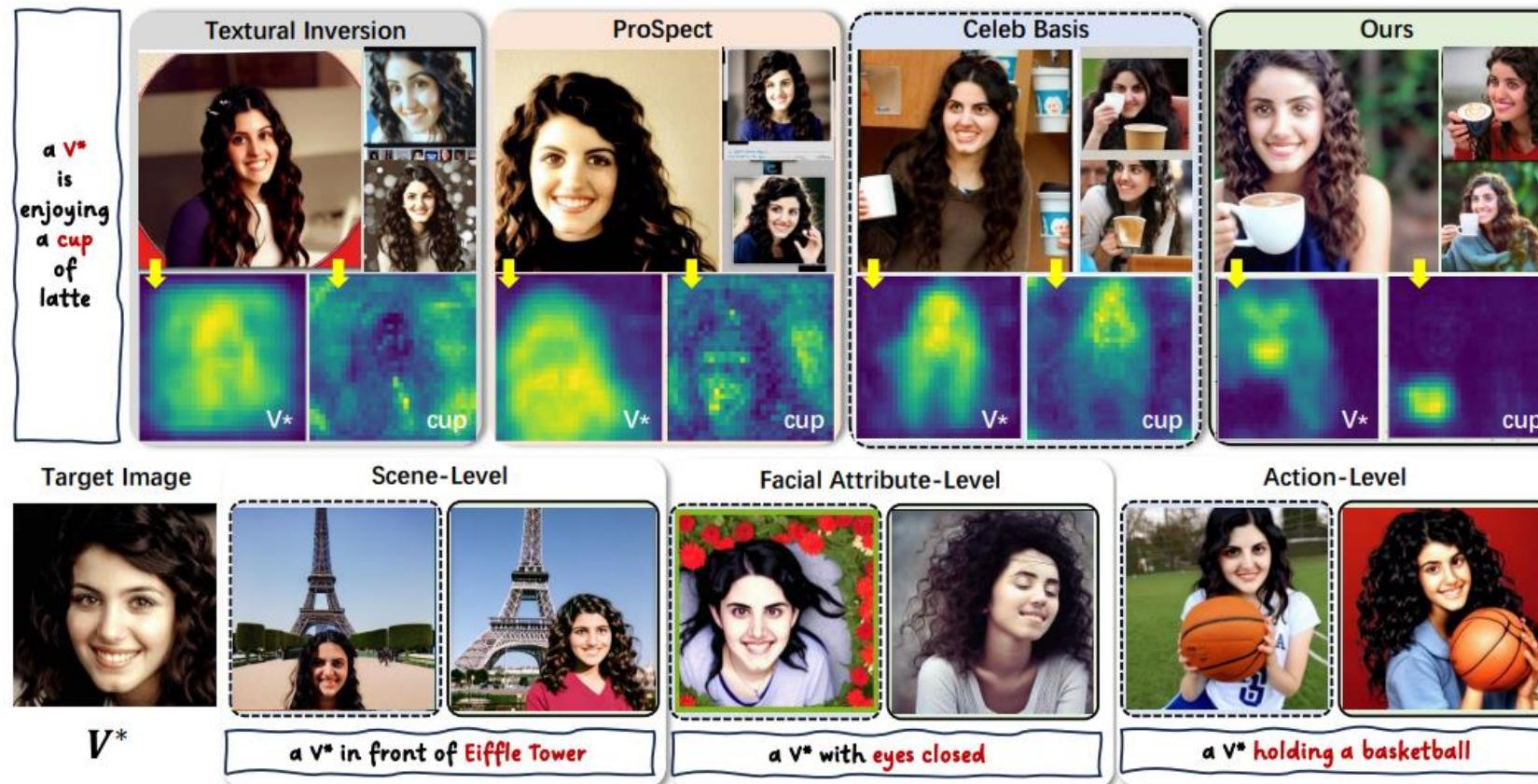
局限性

- 编辑终端不够用户友好
- 不支持用户微调指令（多轮迭代）

未来研究

- 参考文生图大模型研究思路，基于提示词（Prompt）的人脸编辑亟待探索
- 参考上下文学习研究背景，人脸编辑系统需要处理多轮指令与反馈

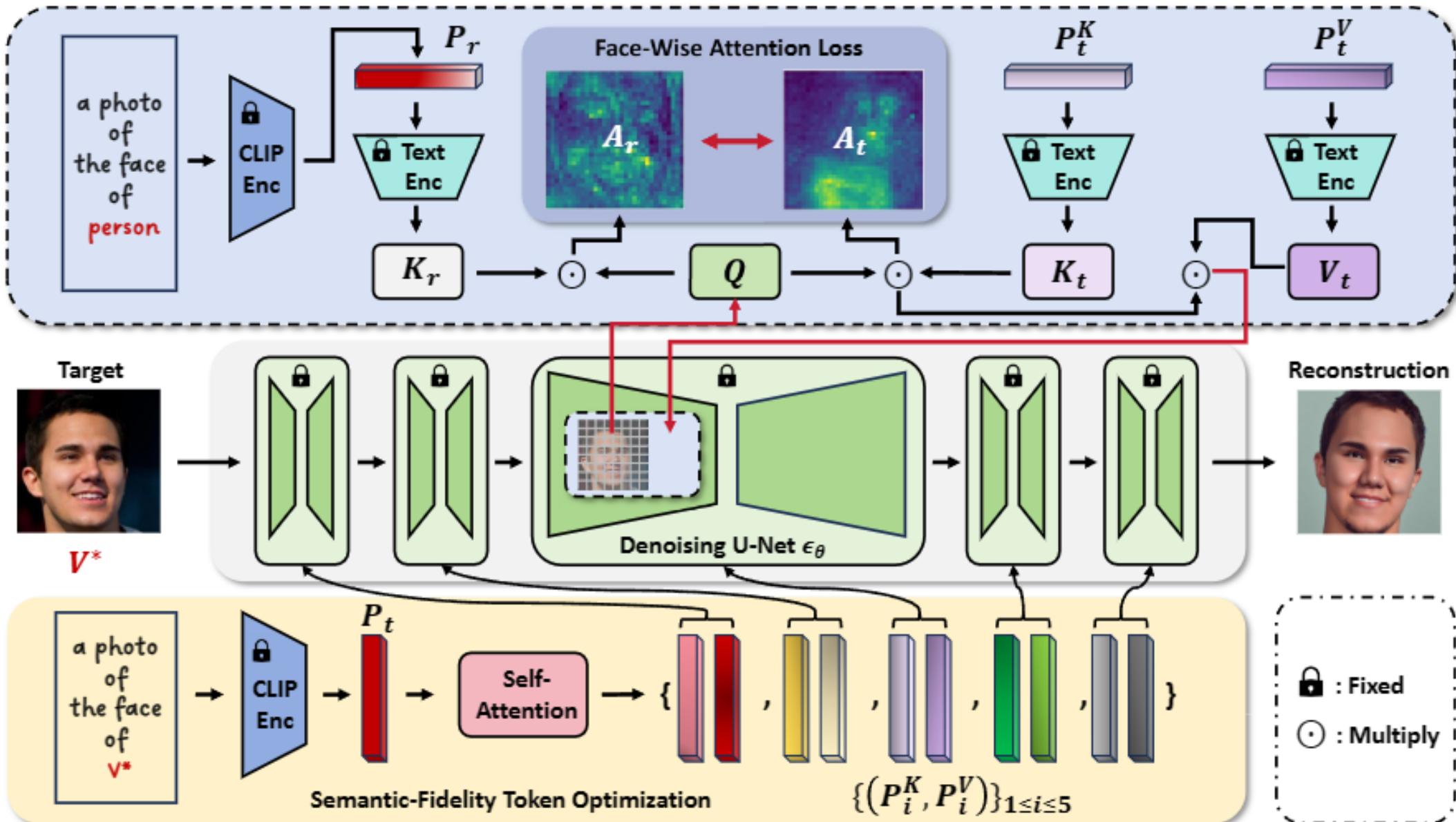
Interactive Identity Embedding for Semantic-Fidelity Diffusion Generation



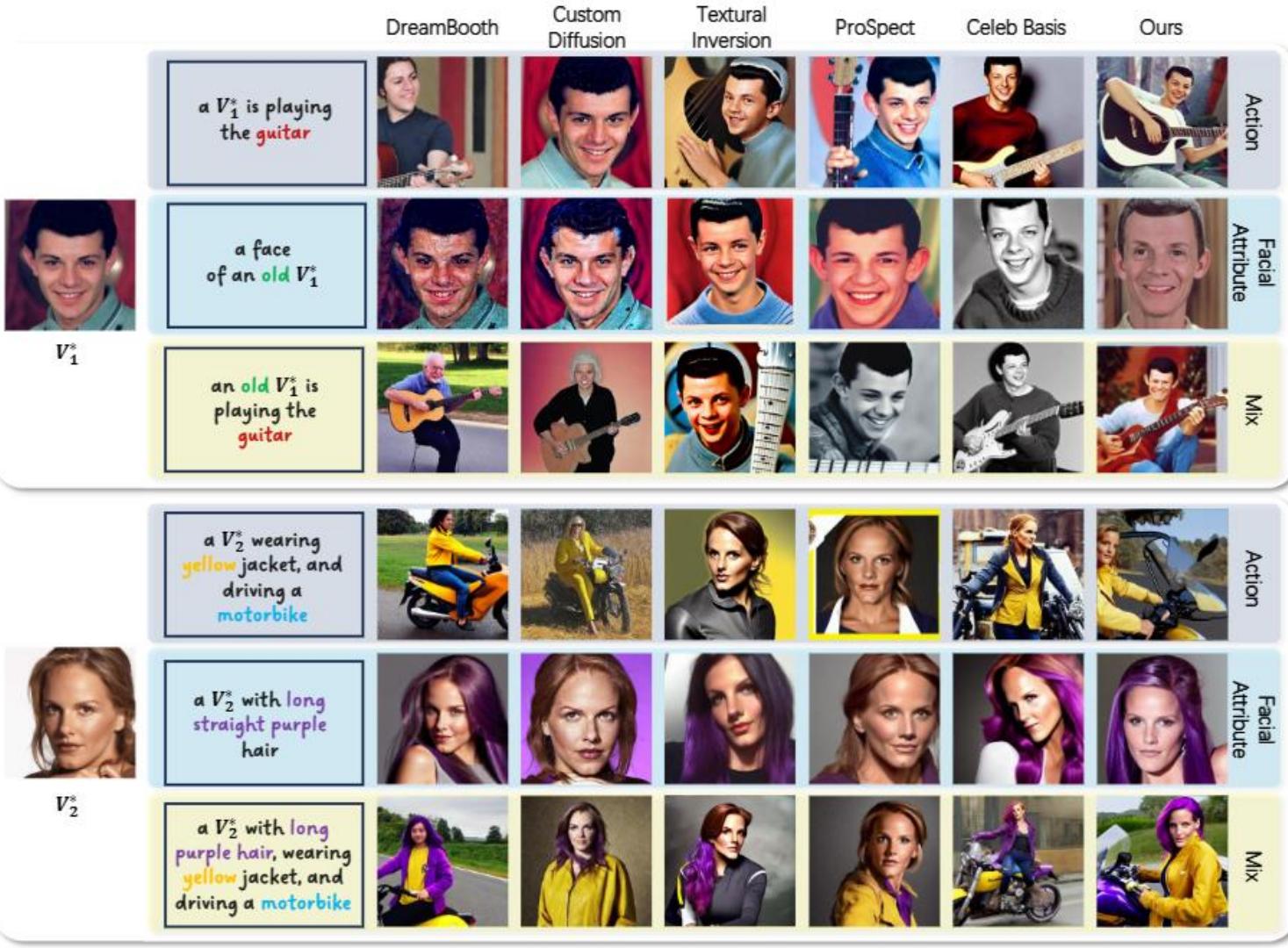
Motivation

- (1) **Attention Overfit:** As shown in the activation maps of Textural Inversion and ProSpect, their “ V^* ” attention nearly takes over the whole images.
- (2) **Limited Semantic-Fidelity:** Despite alleviating overfit, Celeb Basis introduces excessive face prior, limiting the semantic-fidelity of the learned ID embedding and hindering the control of facial attributes such as “eyes closed”.

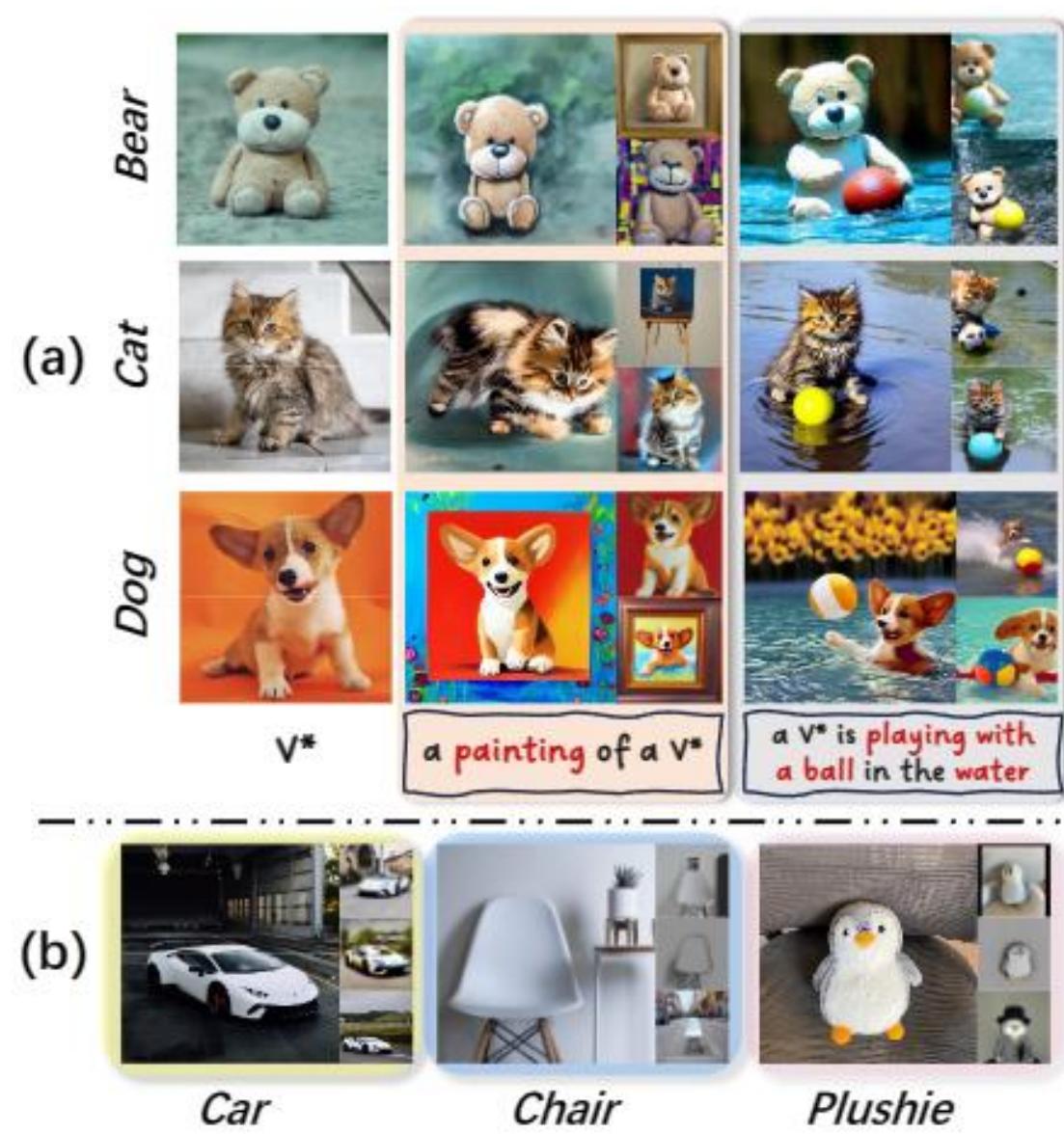
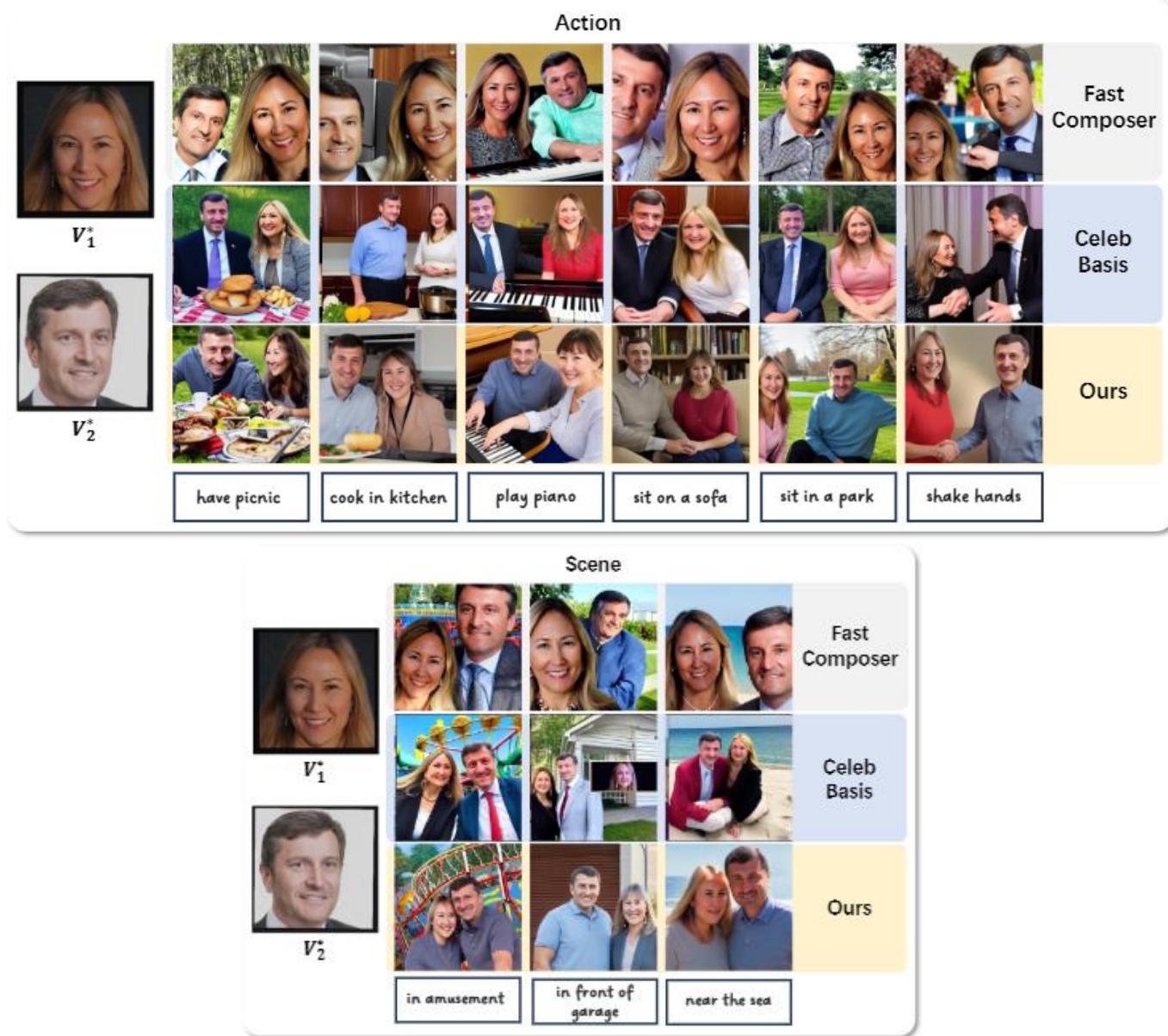
Interactive Identity Embedding for Semantic-Fidelity Diffusion Generation



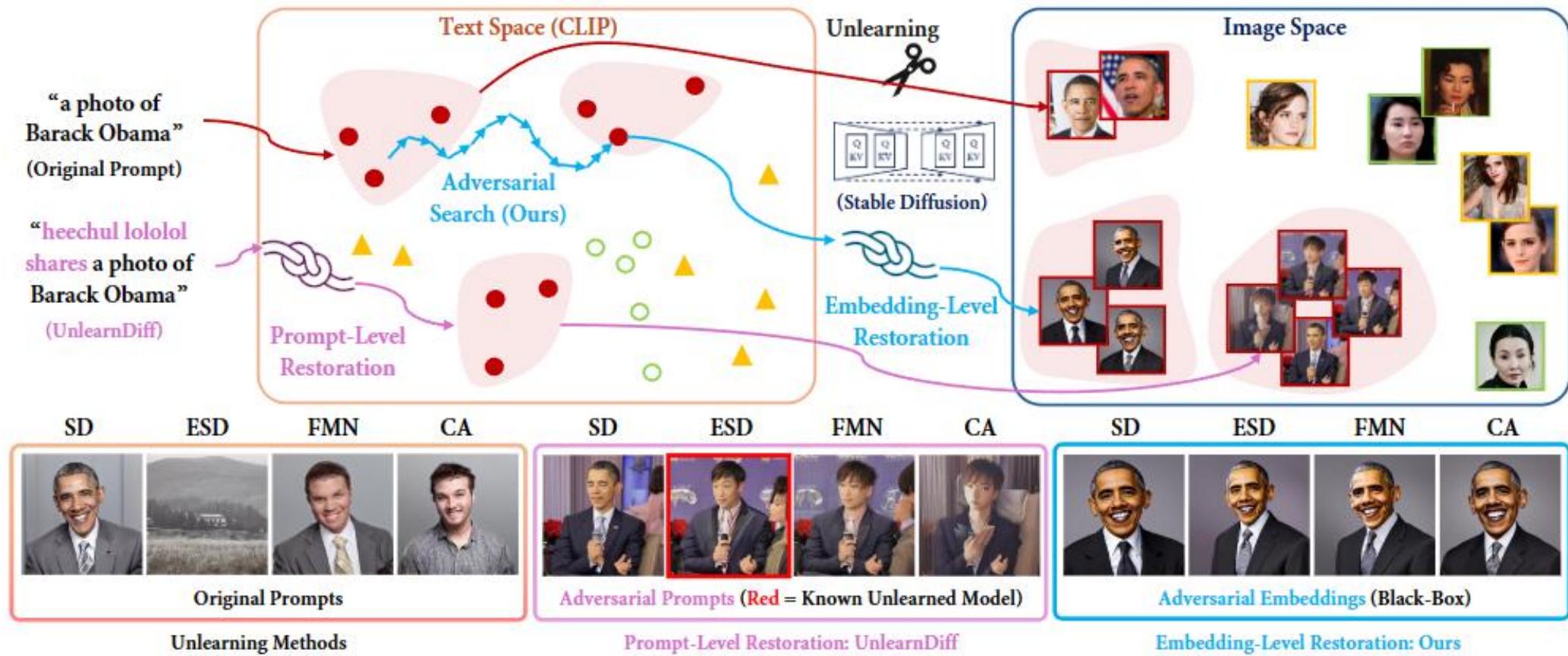
Interactive Identity Embedding for Semantic-Fidelity Diffusion Generation



Interactive Identity Embedding for Semantic-Fidelity Diffusion Generation

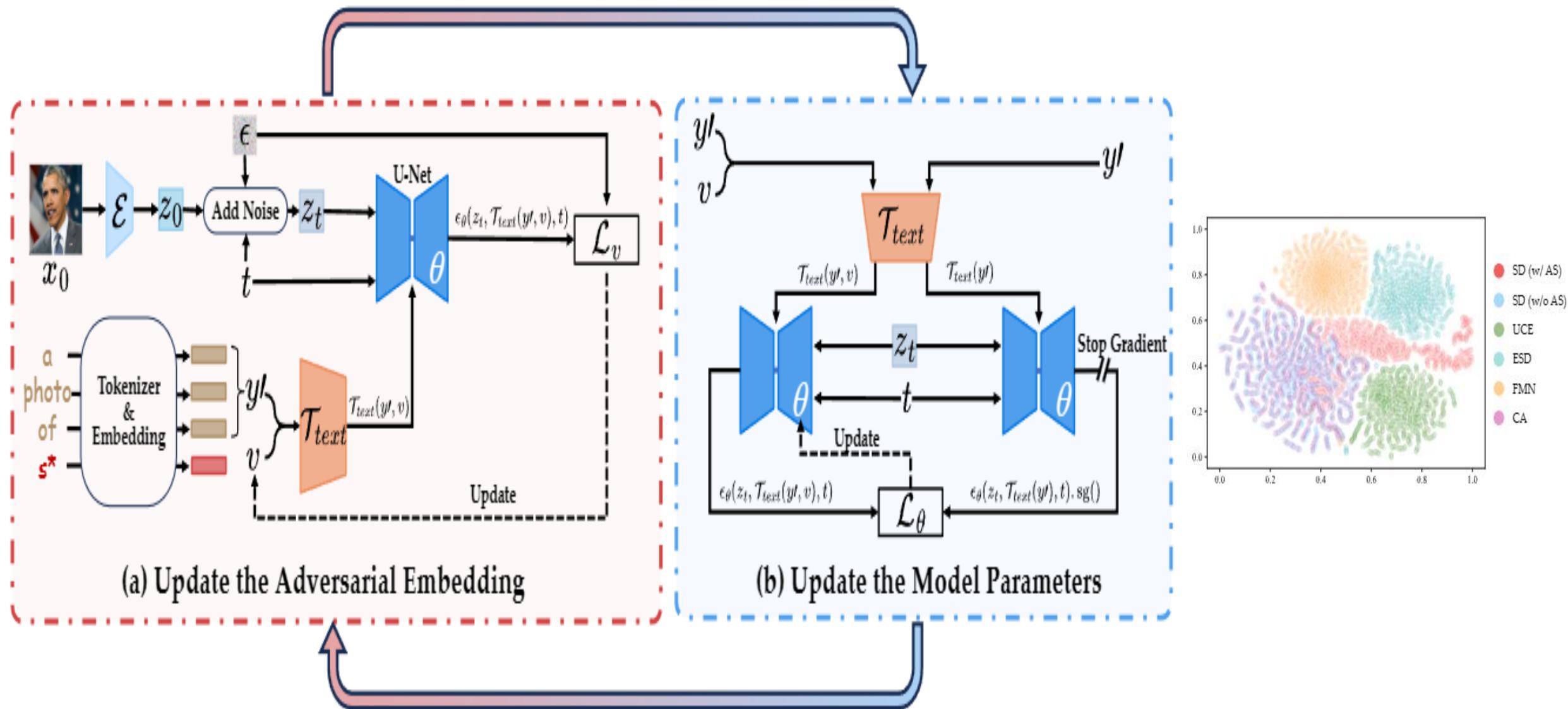


Evaluating Diffusion Model Unlearning from the Perspective of Transferable Adversarial Attacks

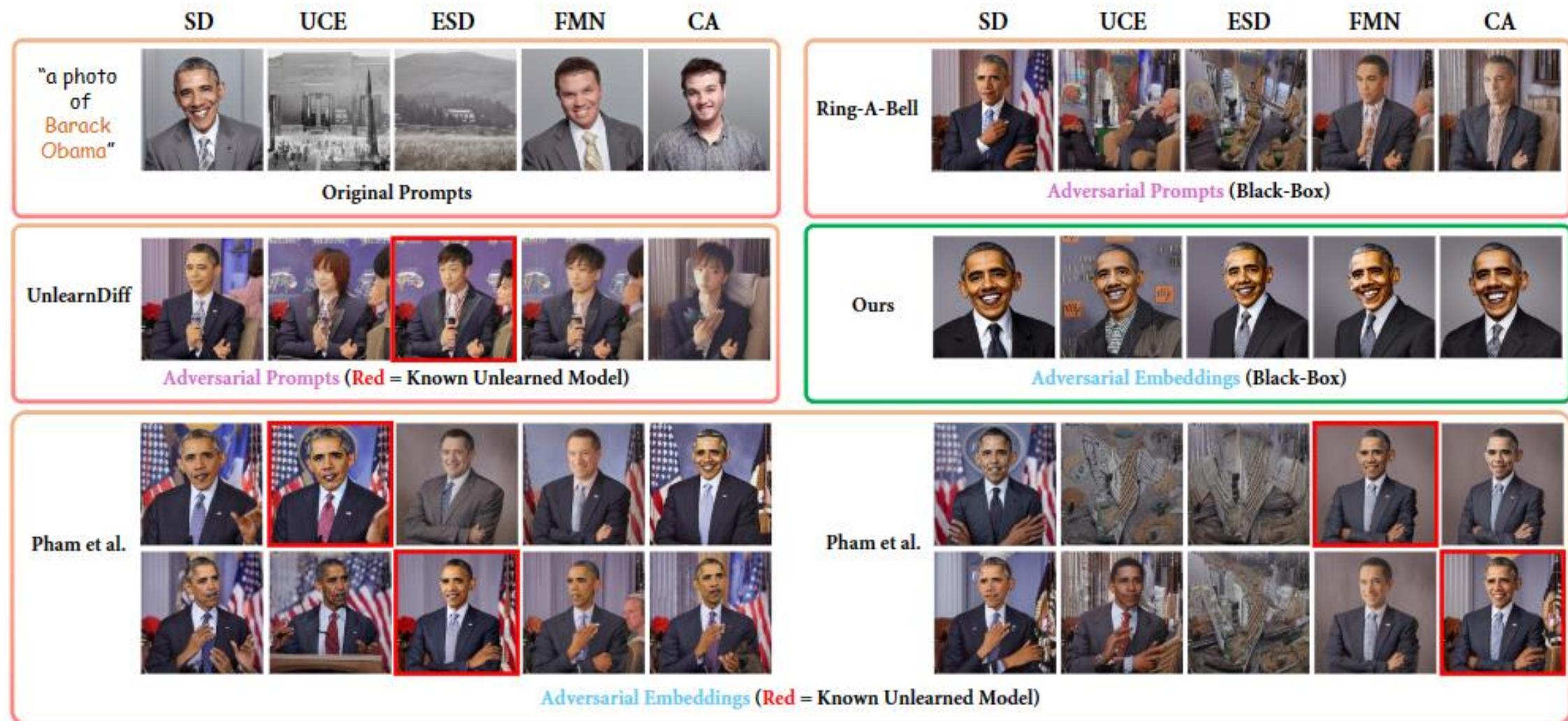


Motivation

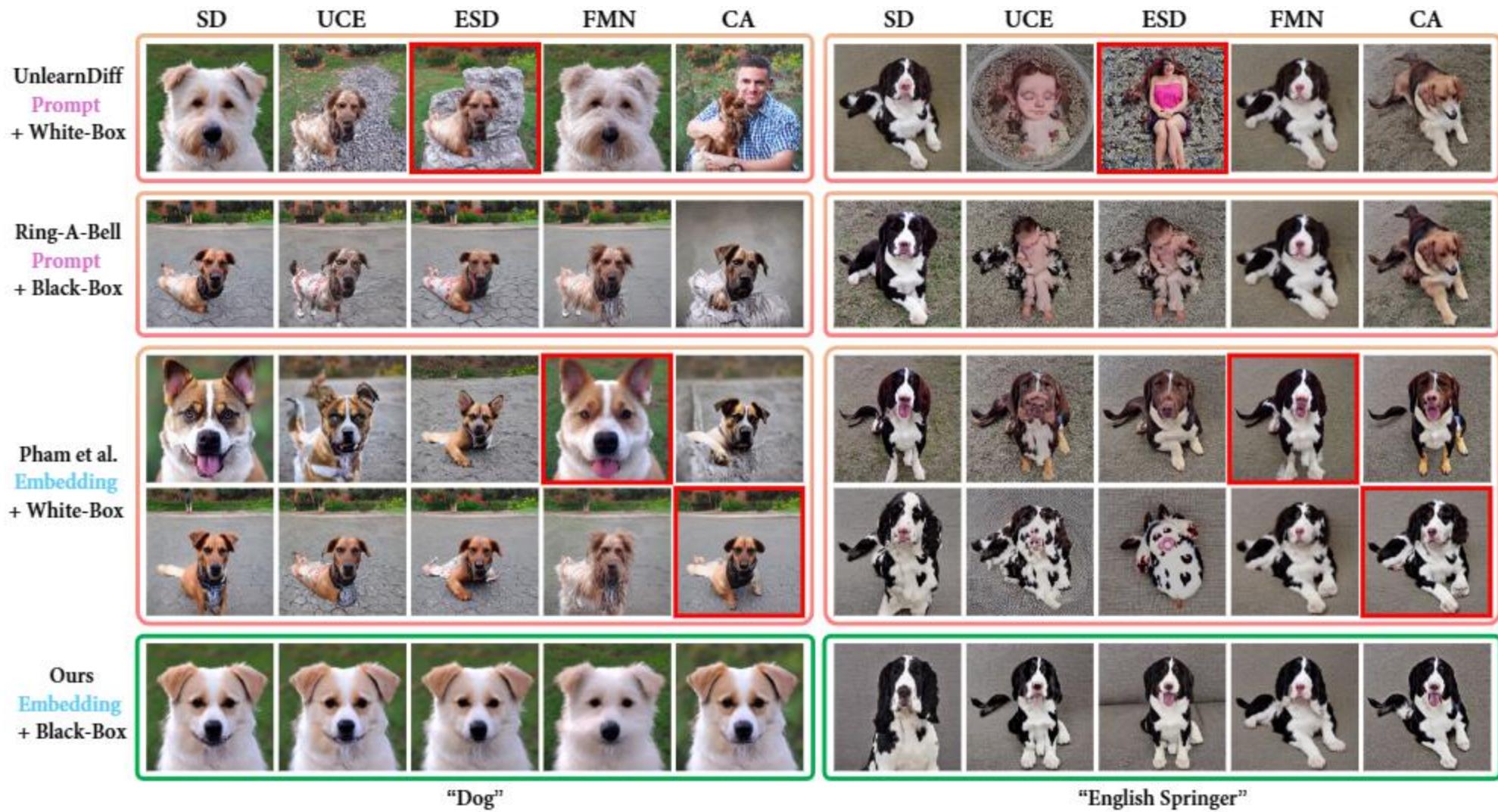
- (1) **Lack of transferability:** Some methods operate within a white-box setting, requiring access to the unlearned model. And the learned adversarial input often fails to transfer to other unlearned models for concept restoration;
- (2) **Limited attack:** The prompt-level methods struggle to restore narrow concepts from unlearned models, such as celebrity identity.



Evaluating Diffusion Model Unlearning from the Perspective of Transferable Adversarial Attacks



Evaluating Diffusion Model Unlearning from the Perspective of Transferable Adversarial Attacks



Evaluating Diffusion Model Unlearning from the Perspective of Transferable Adversarial Attacks

Target Concepts	Erasure Methods	Attack Methods				
		w/o Attack	UD	RAB	CI (FMN)	CI (CA)
Dog (Object)	UCE	0.0	37.0	29.0	<u>86.0</u>	82.0
	ESD	16.0	57.0*	62.0	<u>93.0</u>	78.0
	FMN	65.0	64.0	66.0	<u>82.0*</u>	74.0
	CA	12.0	62.0	59.0	<u>90.0</u>	<u>98.0*</u>
	Average	23.3	55.0	54.0	<u>87.8</u>	<u>83.0</u>
English Springer (Object)	UCE	1.0	0.0	1.0	<u>12.0</u>	24.0
	ESD	0.0	0.0*	0.0	<u>5.0</u>	2.0
	FMN	69.0	70.0	16.0	<u>92.0*</u>	46.0
	CA	1.0	2.0	0.0	2.0	<u>58.0*</u>
	Average	17.8	18.0	4.3	27.8	<u>32.5</u>
Van Gogh (Artist Style)	UCE	0.0	0.0	1.0	<u>1.0</u>	<u>26.0</u>
	ESD	0.0	0.0*	0.0	<u>9.0</u>	3.0
	FMN	0.0	0.0	1.0	<u>51.0*</u>	1.0
	CA	0.0	0.0	3.0	<u>60.0</u>	44.0*
	Average	0.0	0.0	1.3	<u>30.3</u>	<u>18.5</u>
Nudity (NSFW)	UCE	0.0	0.7	0.0	<u>1.5</u>	<u>5.2</u>
	ESD	10.4	13.5*	31.9	34.3	<u>67.2</u>
	FMN	56.0	75.9	61.7	<u>70.9*</u>	<u>76.1</u>
	CA	2.2	3.5	51.1	19.4	<u>64.9*</u>
	Average	17.2	23.4	36.2	<u>31.5</u>	<u>53.4</u>
Barack Obama (ID)	UCE	0.0	0.0	0.0	<u>0.0</u>	<u>10.0</u>
	ESD	0.0	0.0*	0.0	<u>5.0</u>	0.0
	FMN	0.0	0.0	0.0	<u>56.0*</u>	0.0
	CA	0.0	0.0	0.0	<u>47.0</u>	41.0*
	Average	0.0	0.0	0.0	<u>27.0</u>	12.8



中国科学院自动化研究所
模式识别实验室
New Laboratory of Pattern Recognition



中国科学院自动化研究所
INSTITUTE OF AUTOMATION
CHINESE ACADEMY OF SCIENCES



中国科学院大学
University of Chinese Academy of Sciences

Exploring Robust and Expressive Face Representations for Multi-Modal Manipulation



Songlin Yang (杨嵩林)



- **M.S. Degree** at Institute of Automation, Chinese Academy of Sciences (2021.9-2024.6)
- **Intern Researcher** at MicroSoft Research Asia (2022.3-2022.8), and SenseTime Research (2021.6-2021.9, 2022.9-2023.8)
- **M.S. Thesis:** Dynamic Facial Editing based on Neural Radiance Fields
- **Interested Topics:** Face Analysis by Synthesis, 3D Human Motion Generation, Human-Object-Scene Interaction, Neural Rendering, Generative Models, and Embodied Intelligence

**Do not hesitate to drop me an email for
any possible collaboration! :D**