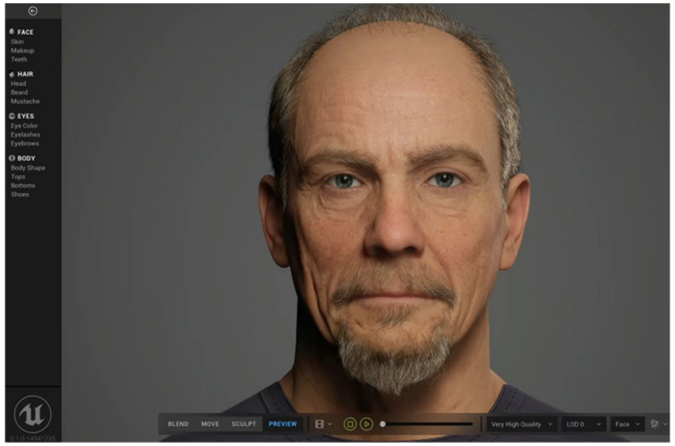
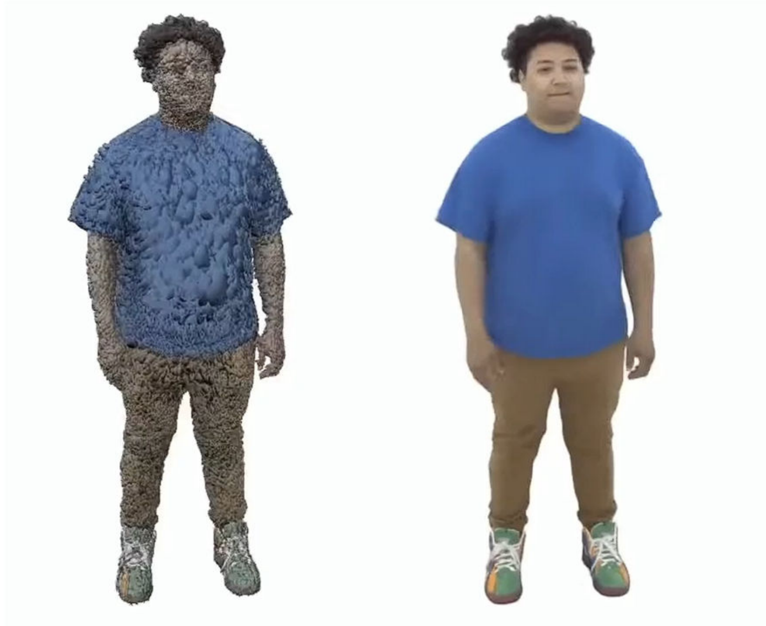
中国科学技术大学
University of Science and Technology of China

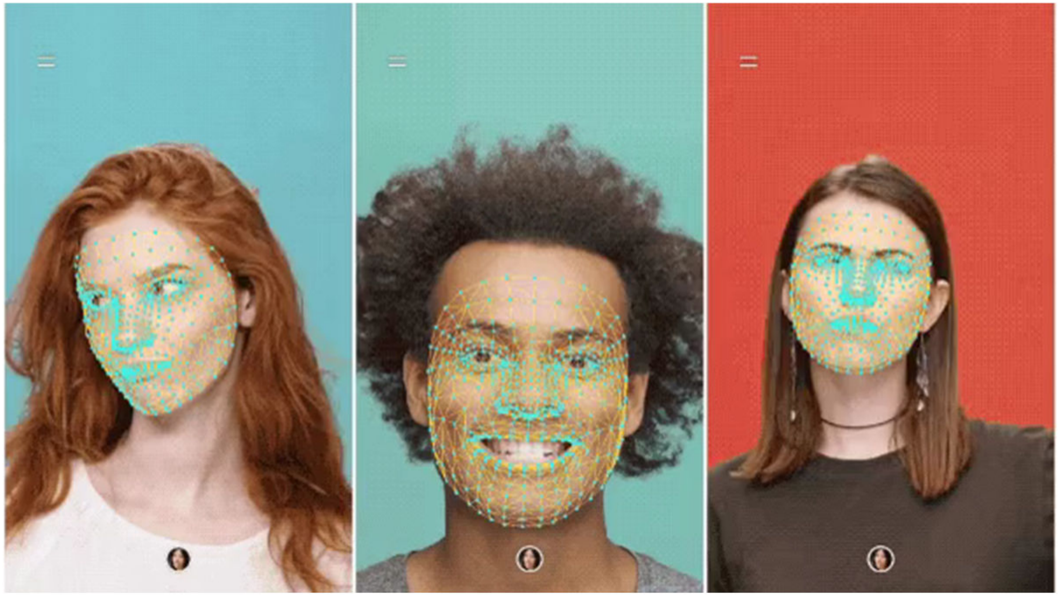
数字人课程：三维数字人表示





郭玉东
中国科学技术大学

应用驱动的数字人表示



视频人脸特效

CG电影

https://developers.google.com/ar/develop/augmented-faces
https://www.youtube.com/watch?v=IPQ5vTqqdgE

应用驱动的数字人表示

面部动态几何、精细几何模型

视频人脸特

电影

显式网格表示

应用驱动的数字人表示

超写实数字主持人

沉浸式视频广告

https://tv.cctv.com/2024/03/07/VIDEUrHr2Zkw6l0ZrmhN7KXP240307.shtml
https://www.youtube.com/watch?v=34KeBnSwvmc

应用驱动的数字人表示

外观形象精确、立体式还原

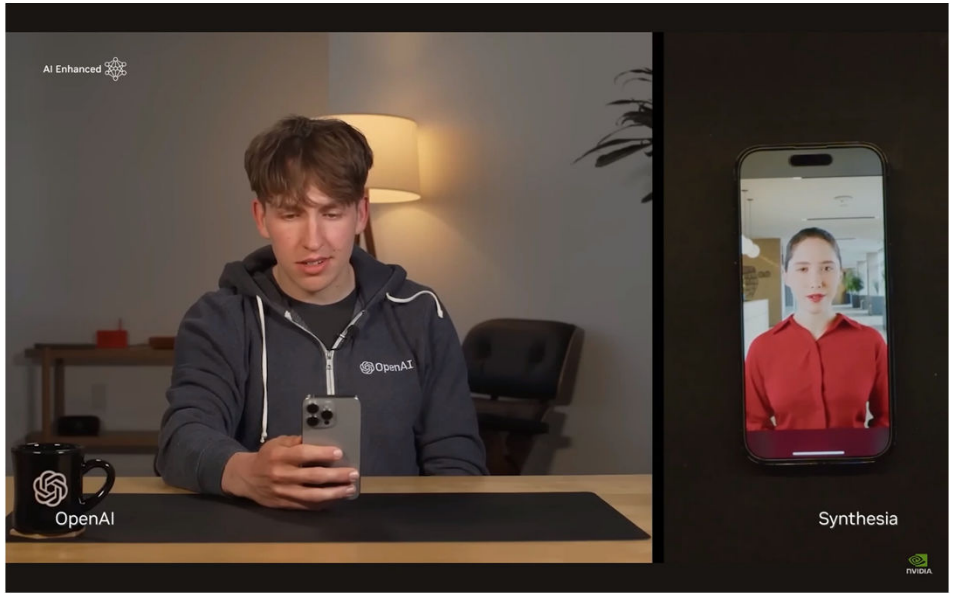
超写实数字主

视频广告

神经辐射场表示

应用驱动的数字人表示

全息沉浸式会议



超写实AI Agent

https://www.youtube.com/watch?v=xWwMrw7GeX8
https://www.youtube.com/watch?v=8xMeIwBnIpU

应用驱动的数字人表示

实时全息、高保真互动

全息沉浸式会

Agent

三维高斯泼溅表示

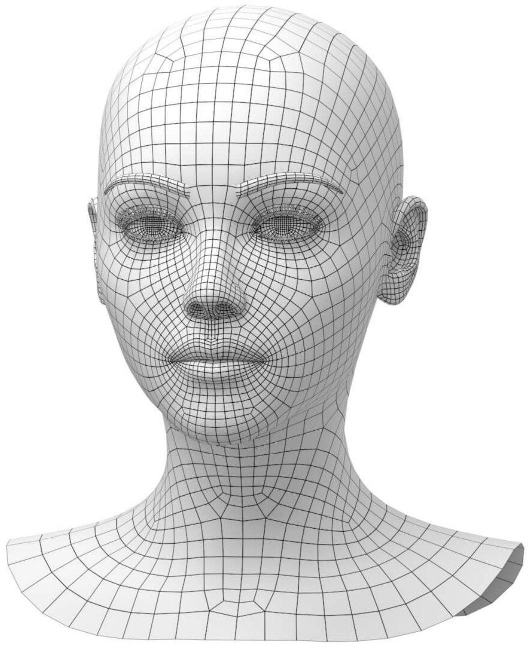
表　示

1 Mesh数字人

2 NeRF数字人

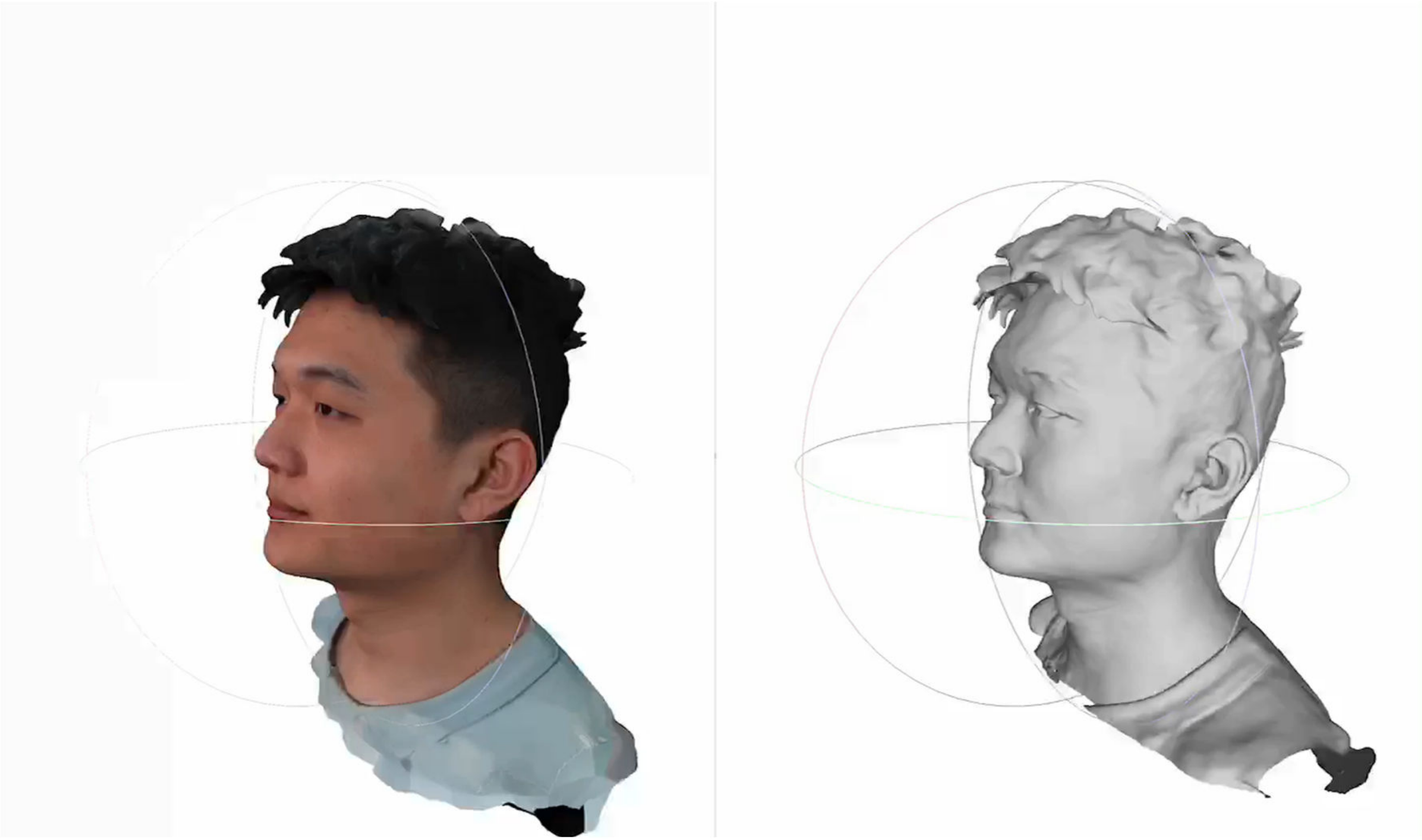
3 3DGS数字人

Mesh简介



(V, E, F)

Mesh简介



精细几何

渲染高效

驱动方便

特定身份建模Challenge

https://github.com/NVlabs/ffhq-dataset

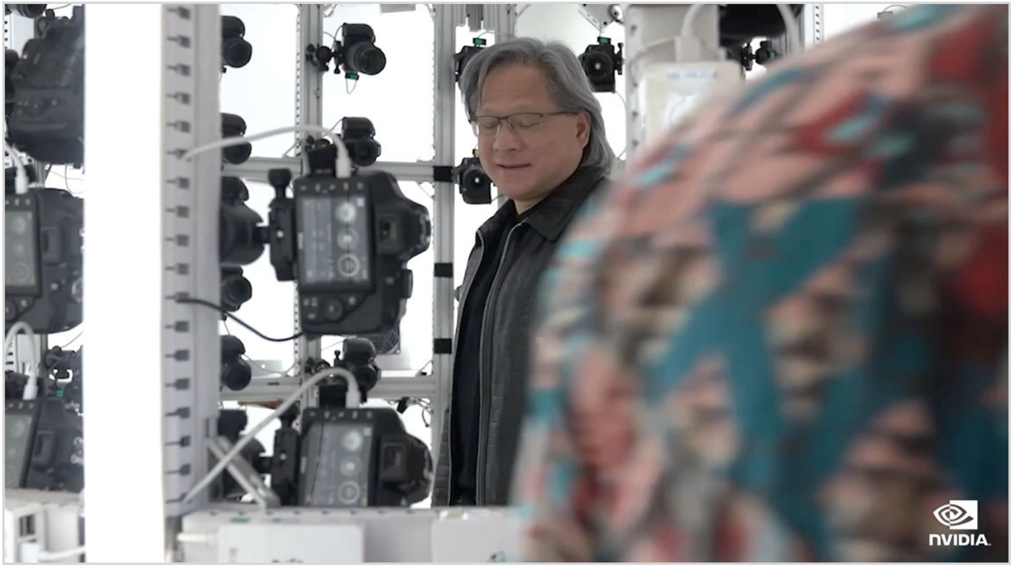
多样性高

几何复杂

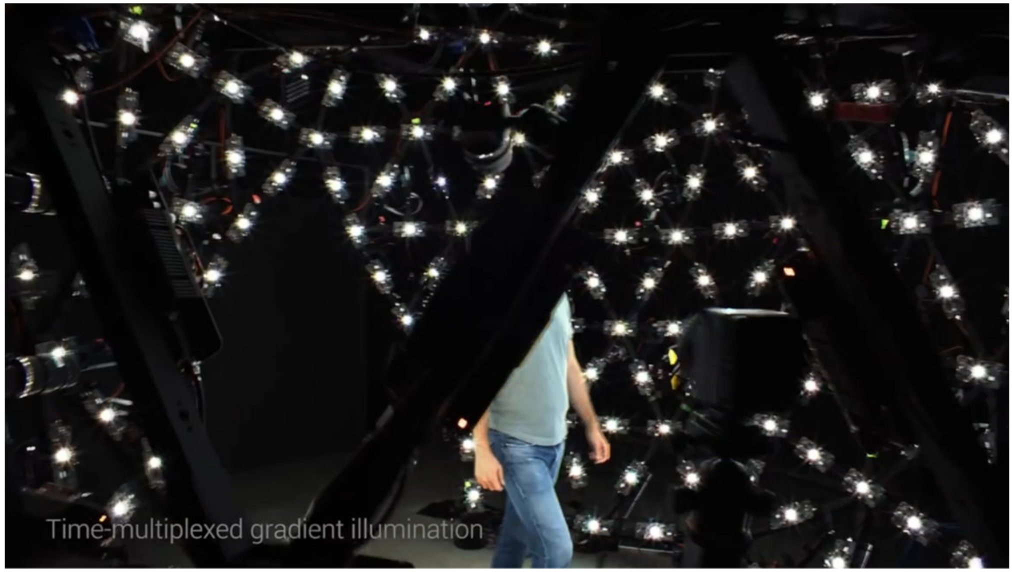
维度高

变量维度高，需要更多等式约束（观测）

精细Mesh数字人制作流程



GTC 2021, NVIDIA
Multi-view Stereo



The Relightables, Google
Photometric Stereo

精细建模

设备昂贵，成本高

只适用于采集者，无法scale到他人

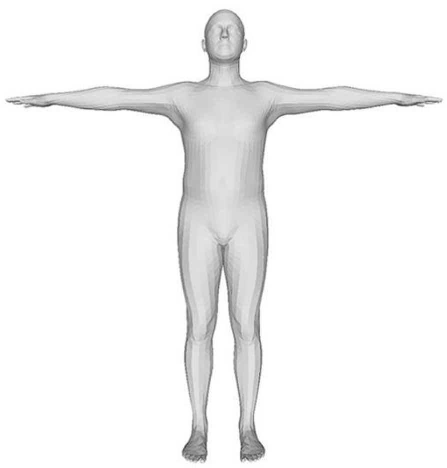
https://www.youtube.com/watch?v=f_V30ueEXE4
https://www.youtube.com/watch?v=anBRroZWfzI

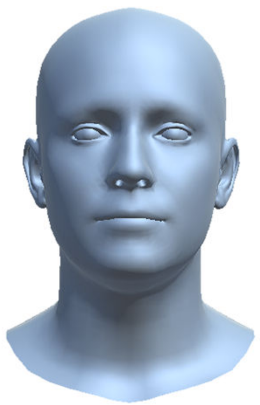
解决泛化问题Key Idea — 生成模型

- 利用低维Code生成高维真实数据的生成模型，将解高维方程转化为求解低维方程，从而可减少观测/输入

Latent Space
128 Dim Vector Z

参数化表示—降维





Domain先验

- Human Body
- Faces
- Hands

实用性

- 可泛化
- 参数量低
- 求解简便

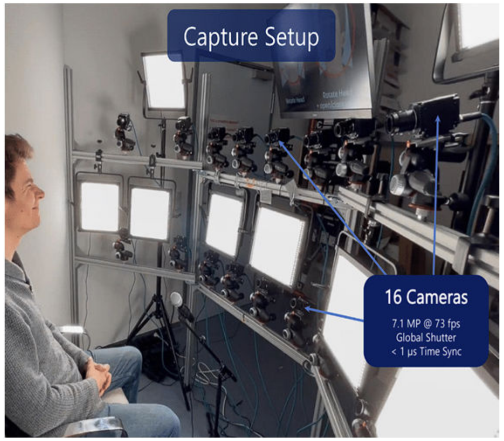
$$\boldsymbol{x} \in \mathbb{R}^D \xrightarrow{f(\boldsymbol{x},\theta)} \boldsymbol{z} \in \mathbb{R}^d$$

- 拓扑共享
- 顶点变形
- 高维向量

降维

- PCA
- Auto Encoder
- VAE

How to Build

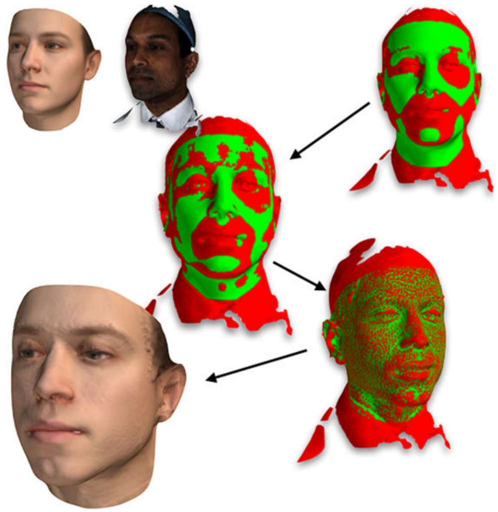


Data Capturing

Multi-view Images

Multi-view Stereo

Raw Scan Data



Non-Rigid ICP

拓扑一致的样本网格

$X \in R^{N \times D}$

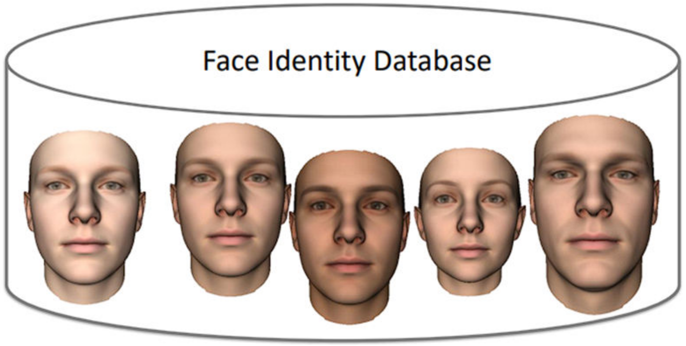
```
X_std = (X - mean(X)) / std(X)
C = cov(X_std)
vals, vecs = eig(C)
P_k = vecs[:, argsort(vals)[::-1]]
Components = P_k[:, :k]
Z = X_std @ Components
```

```
Z = Encoder(X)
X_rec = Decoder(Z)
loss = recon_loss(X, X_rec)
for _ in range(epochs):
    Z = Encoder(X)
    X_rec = Decoder(Z)
    loss = recon_loss(X, X_rec)
    optimizer.step(loss.backward())
```

Dim Reduction

$$\boldsymbol{x} \in \mathbb{R}^D \xrightarrow{f(\boldsymbol{x},\theta)} \boldsymbol{z} \in \mathbb{R}^d$$

人脸参数化表示—3DMM

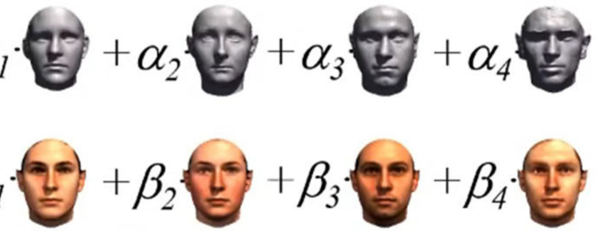


A vector space of 3D shapes and colors of a class of objects

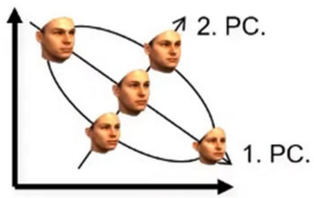
- linear combinations of shapes $\mathbf{S}$ and textures $\mathbf{T}$

$$\mathbf{S} = \sum_i \alpha_i \boldsymbol{S}_i = \alpha_1 \cdot + \alpha_2 \cdot + \alpha_3 \cdot + \alpha_4 \cdot + \ldots$$

$$\boldsymbol{T} = \sum_i \beta_i \boldsymbol{T}_i = \beta_1 \cdot + \beta_2 \cdot + \beta_3 \cdot + \beta_4 \cdot + \ldots$$



- Often: Principal Component Analysis (PCA)



A morphable model for the synthesis of 3D faces. SIGGRAPH 1999.

人脸参数化表示—3DMM

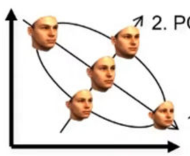
A vector space of 3D shapes and colors of a class of objects

- linear combinations of shapes $\mathbf{S}$ and textures $\mathbf{T}$

$$\mathbf{S} = \sum_i \alpha_i \boldsymbol{S}_i = \alpha_1 \cdot \quad + \alpha_2 \cdot \quad + \alpha_3 \cdot \quad + \alpha_4 \cdot \quad + \ldots$$

$$\boldsymbol{T} = \sum_i \beta_i \boldsymbol{T}_i = \beta_1 \cdot \quad + \beta_2 \cdot \quad + \beta_3 \cdot \quad + \beta_4 \cdot \quad + \ldots$$

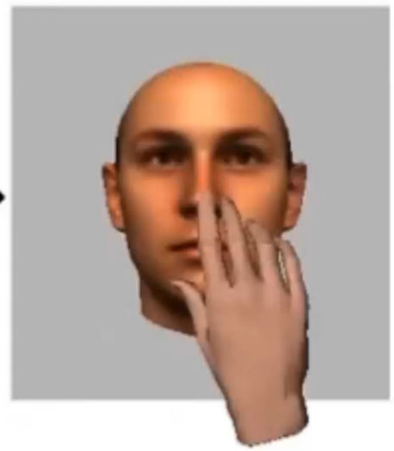
- Often: Principal Component Analysis (PCA)



Input Image

$\leftrightarrow$

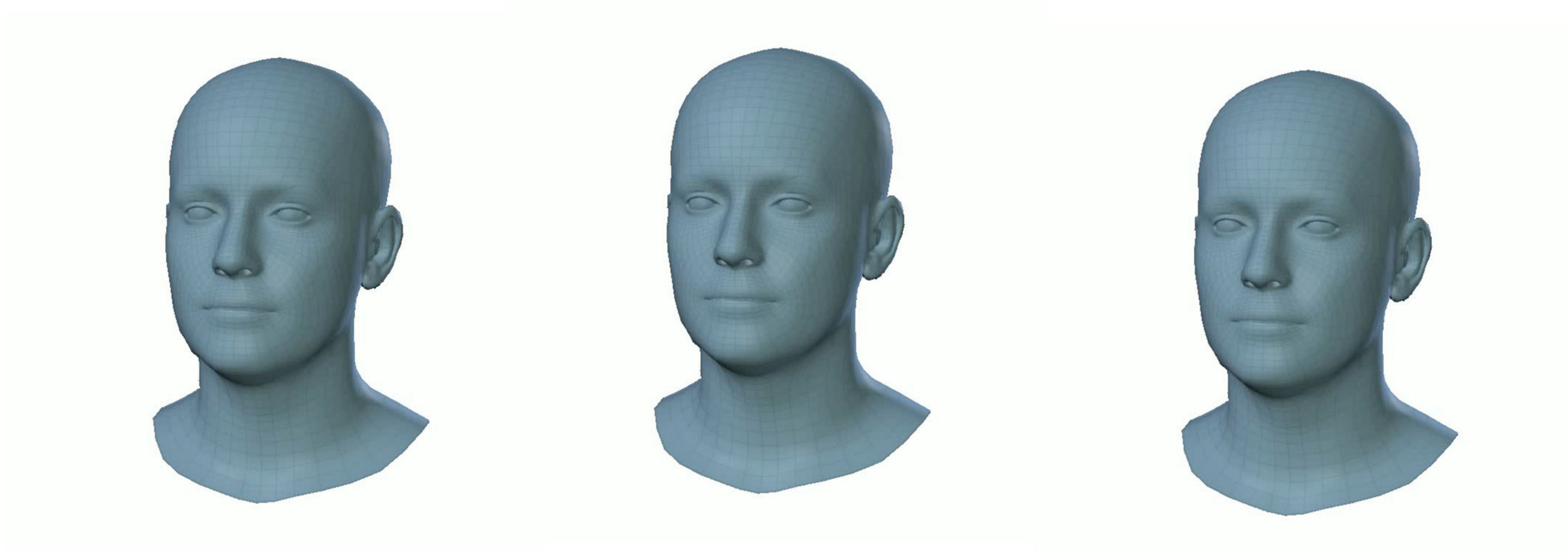
Synthetic Image



Fitting on in-the-wild images

A morphable model for the synthesis of 3D faces. SIGGRAPH 1999.

带表情参数化表示—FLAME



Identity Expression Pose

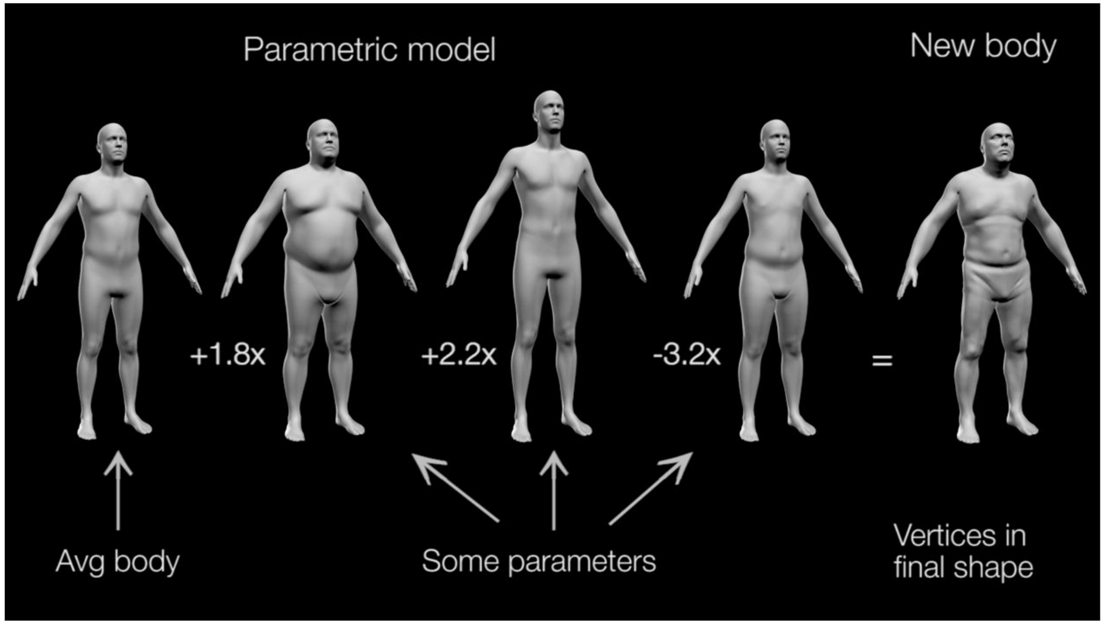
Learning a model of facial shape and expression from 4D scans. SIGGRAPH Asia 2017.

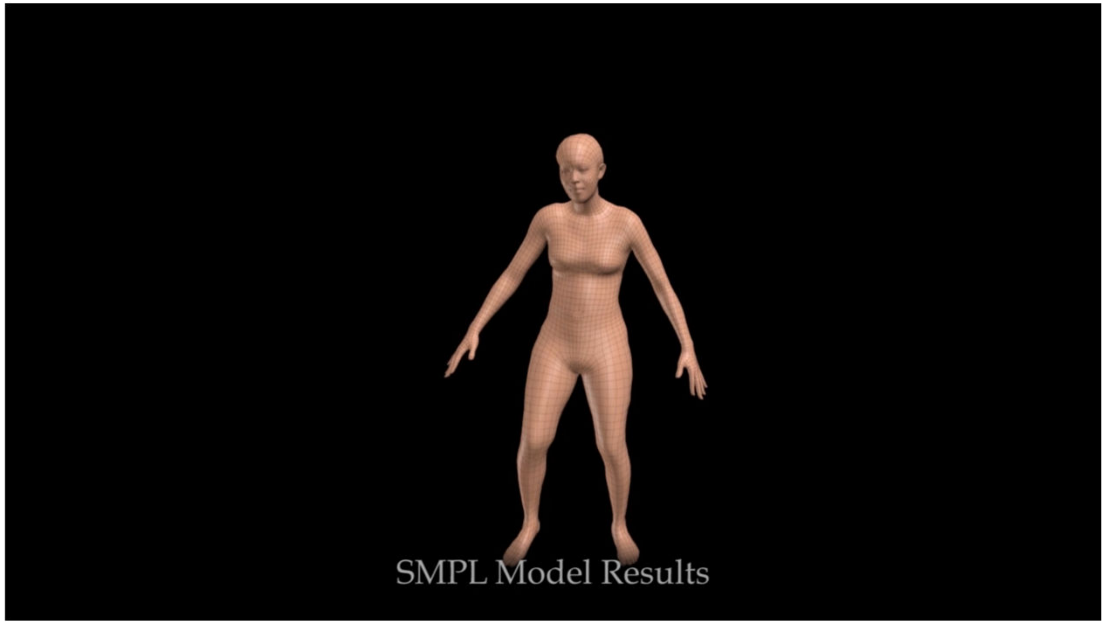
带表情参数化表示

- 得益于低维表示及表情编码，可高效建模与驱动

CNN-based Real-time Dense Face Reconstruction with Inverse-rendered Photo-realistic Face Images. TPAMI 2018.

Body参数化表示

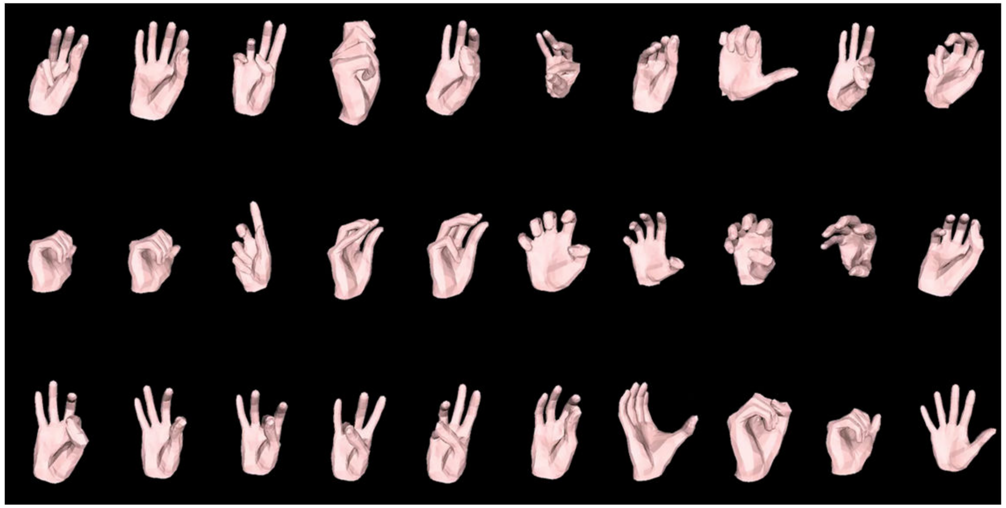




1860个人体各种姿势下的扫描数据

SMPL: A Skinned Multi-Person Linear Model. SIGGRAPH Asia 2015.

Hand参数化表示

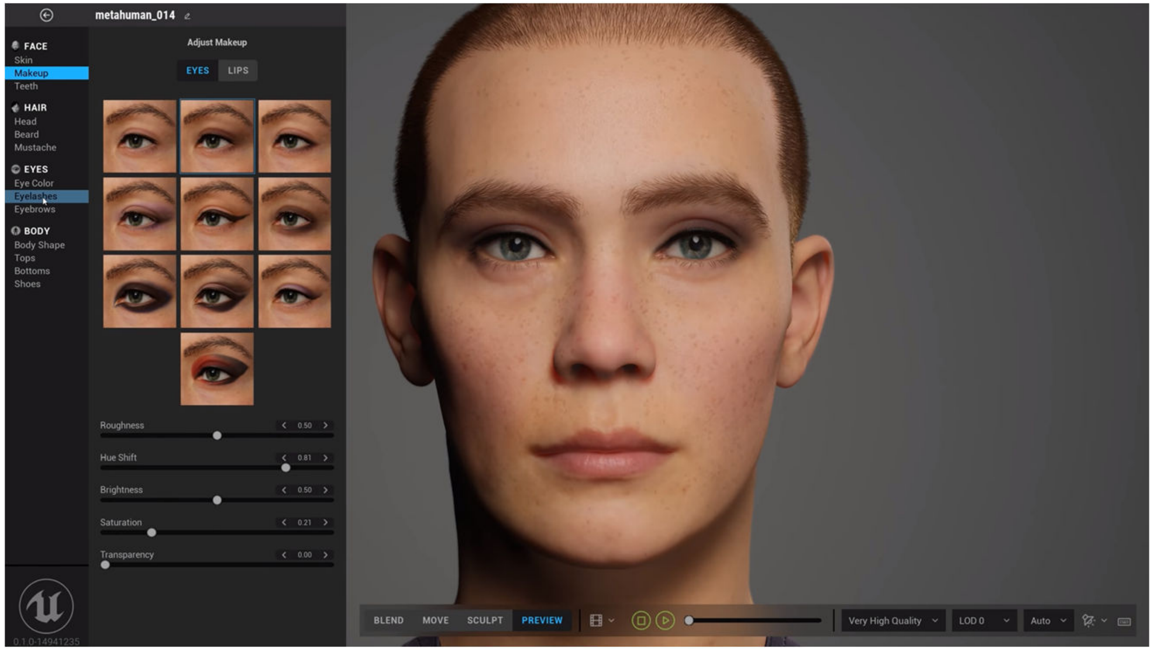


31个人的2018个手部扫描数据

Embodied Hands: Modeling and Capturing Hands and Bodies Together. SIGGRAPH Asia 2017.

产业界MetaHuman

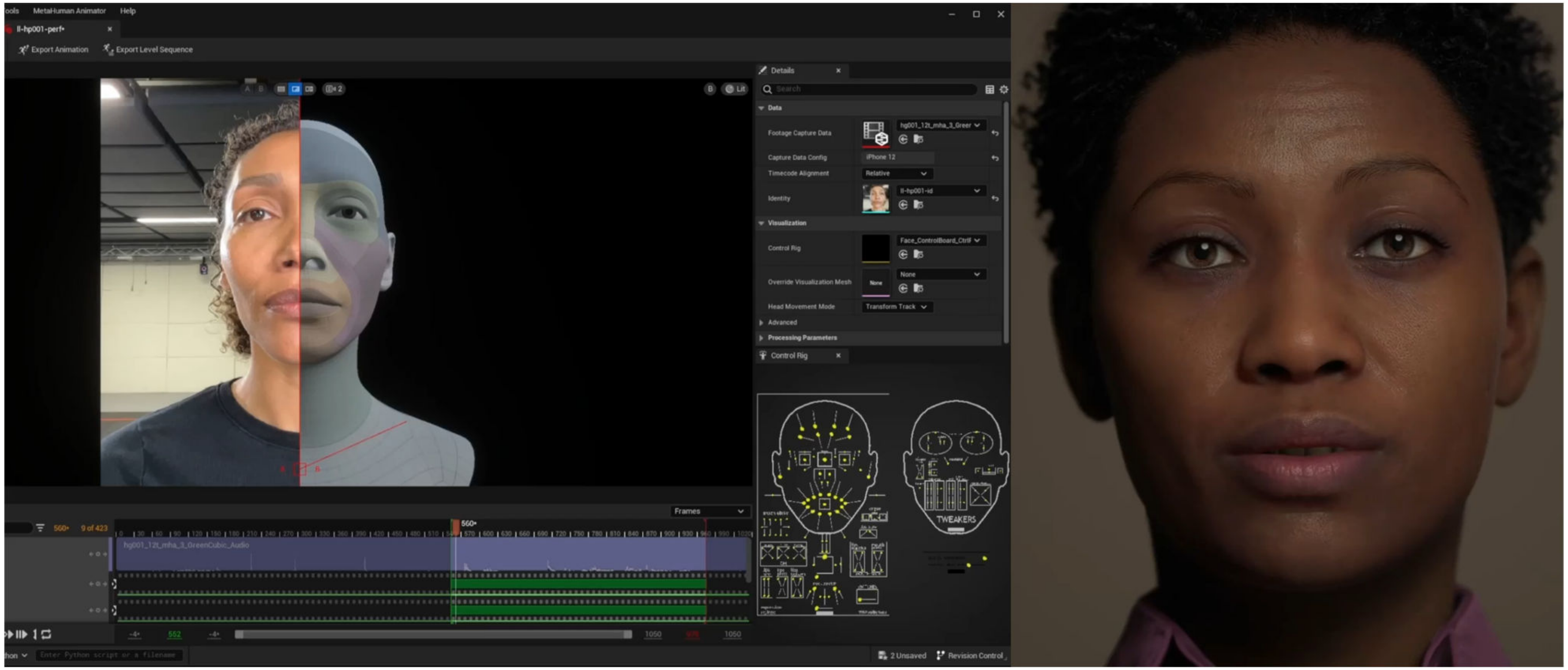
基于上万个人体的扫描数据构建



https://www.unrealengine.com/en-US/metahuman

- 更完整：头发、牙齿等手工建模的资产
- 更真实：相对学术界参数化模型
- 不够Photo-Realistic
- 难以方便地对特定用户进行复刻

Mesh驱动—MetaHuman Animator



https://www.unrealengine.com/en-US/metahuman

驱动Mesh人体—Siren数字人

https://www.youtube.com/watch?v=9owTAISsvwk

Mesh驱动人体—Siren数字人

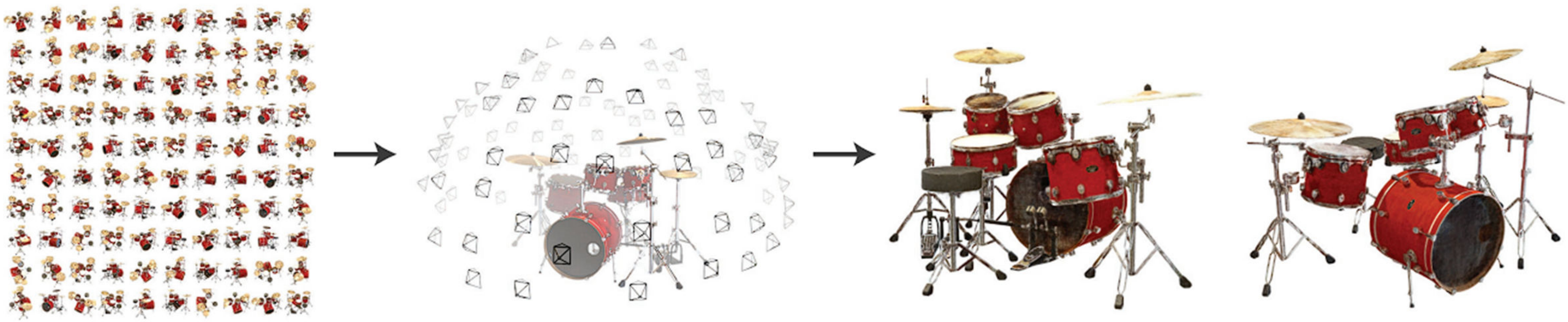
表　示

1 Mesh数字人

2 NeRF数字人

3 3DGS数字人

NeRF简介—任务



Input:
一个场景的相对
稀疏的图片

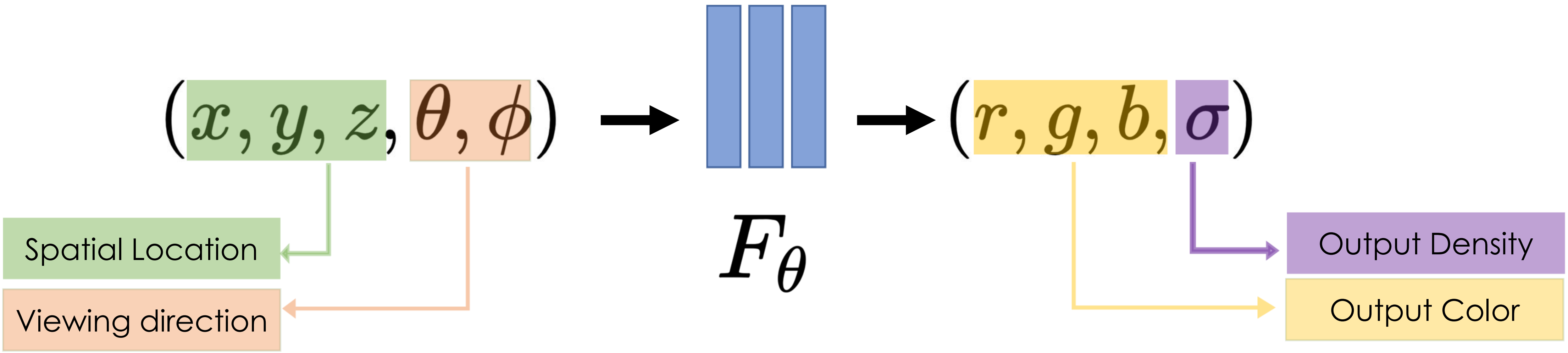
Learn scene
representation

场景的新视角合成

NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. ECCV 2020.

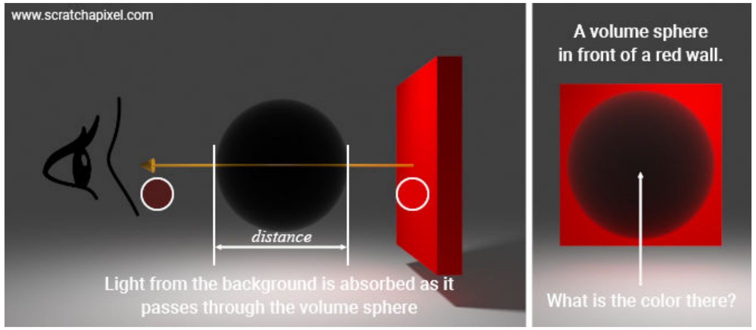
NeRF简介—KeyIdea

学习场景的辐射场表示



NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. ECCV 2020.

NeRF简介—体渲染



Given color and density (r, g, b, σ) , we calculate the color of every camera ray using:

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))\mathbf{c}(\mathbf{r}(t))dt$$

Camera ray: $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$

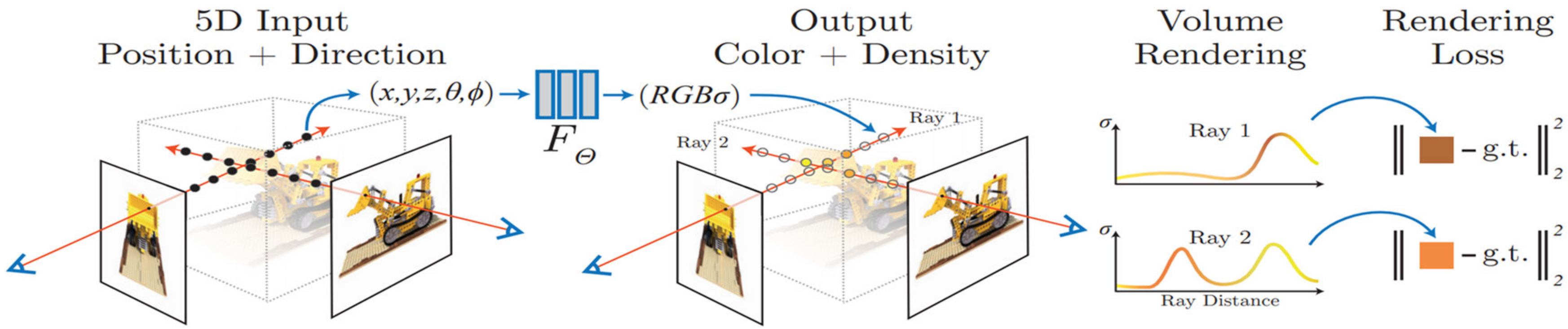
Near and far bounds

Color (r,g,b) at r(t)

Volume density: Probability of a ray terminating at an infinitesimal particle at location r(t)

Accumulated transmittance along ray :
$T(t) = \exp(-\int_{t_n}^{t_f} \sigma(\mathbf{r}(s))ds)$

NeRF简介



<u>Training NeRF</u>:

1) 逐像素沿着相机的射线在场景中生成一组**采样**的3D点
2) 采样点位置及视线方向输入**MLP**中，输出采样点的**颜色和密度**
3) 使用**体渲染**技术将这些颜色和密度累积成一个2D图像
4) 最小化渲染的颜色与GT颜色，Loss梯度回传

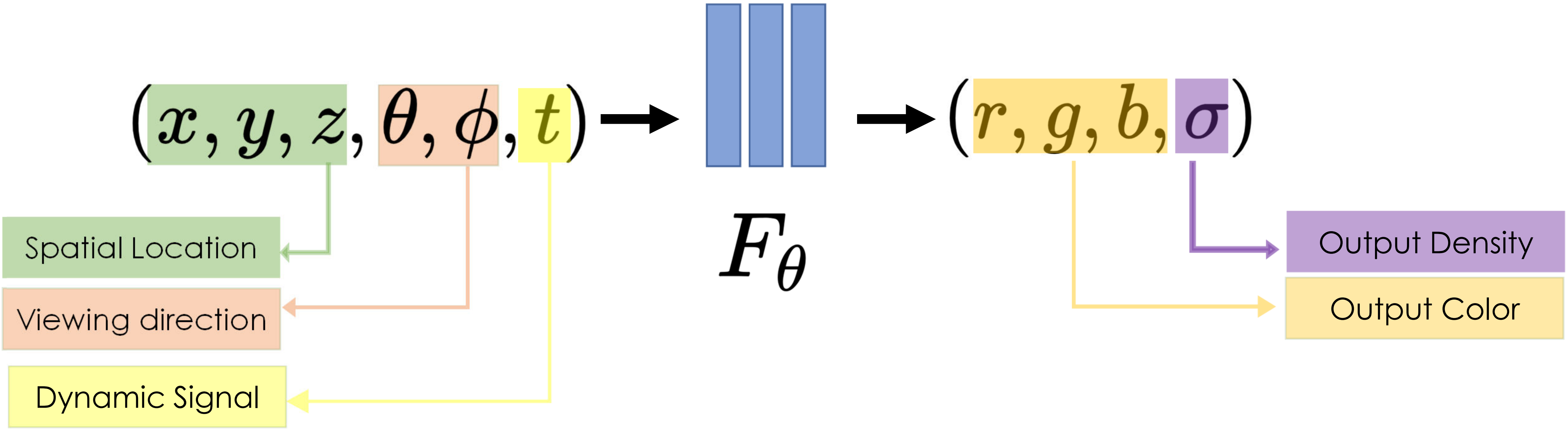
NeRF优势 — PhotoRealistic建模

- 得益于NeRF强大的表达能力和端到端的可微过程

Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. SIGGRAPH 2022.

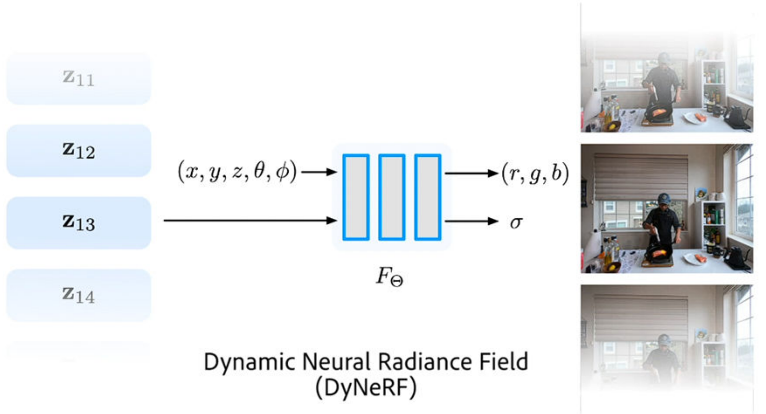
NeRF数字人表示—KeyIdea

- 额外加一个动态输入（时间/驱动信号等）预测动态的辐射场



NeRF数字人表示—DyNeRF

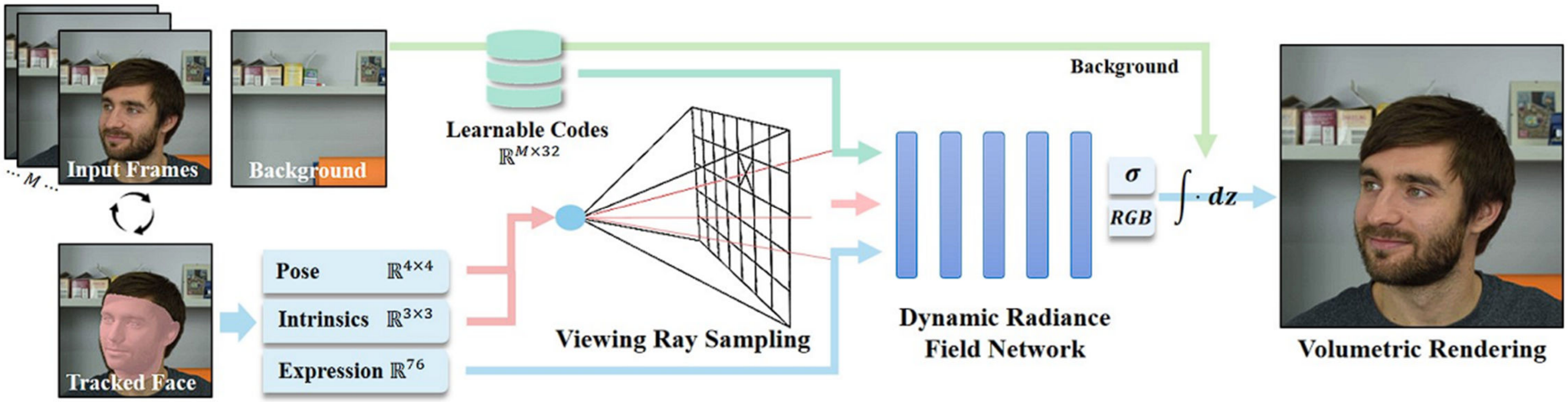
- 动态辐射场可渲染动态场景



Neural 3D Video Synthesis from Multi-view Video. CVPR 2022.

NeRF数字人表示—NeRFace

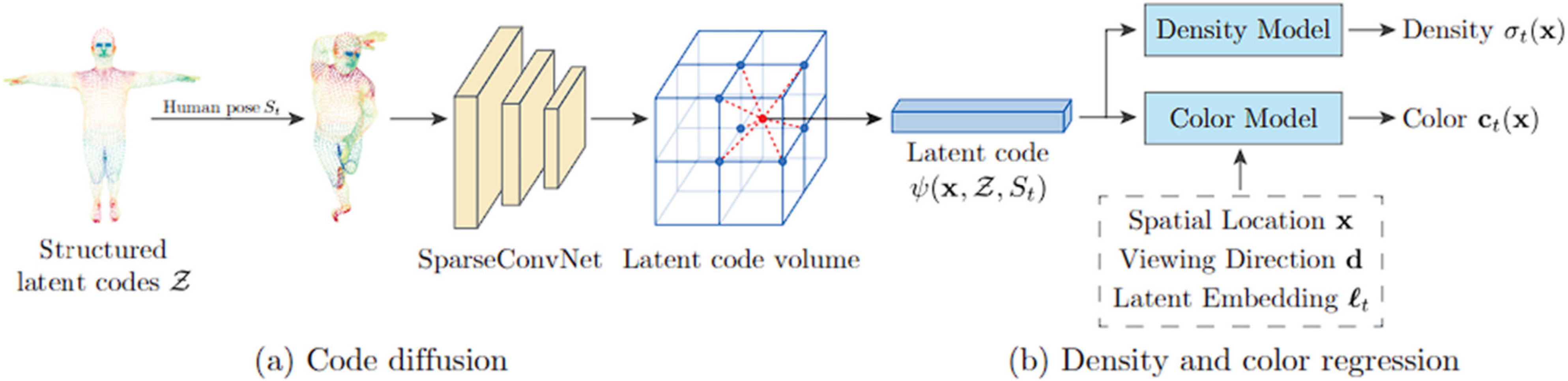
- 将表情信号作为动态辐射场的输入



Dynamic Neural Radiance Fields for Monocular 4D Facial Avatar Reconstruction. CVPR 2021.

NeRF数字人表示—NeuralBody

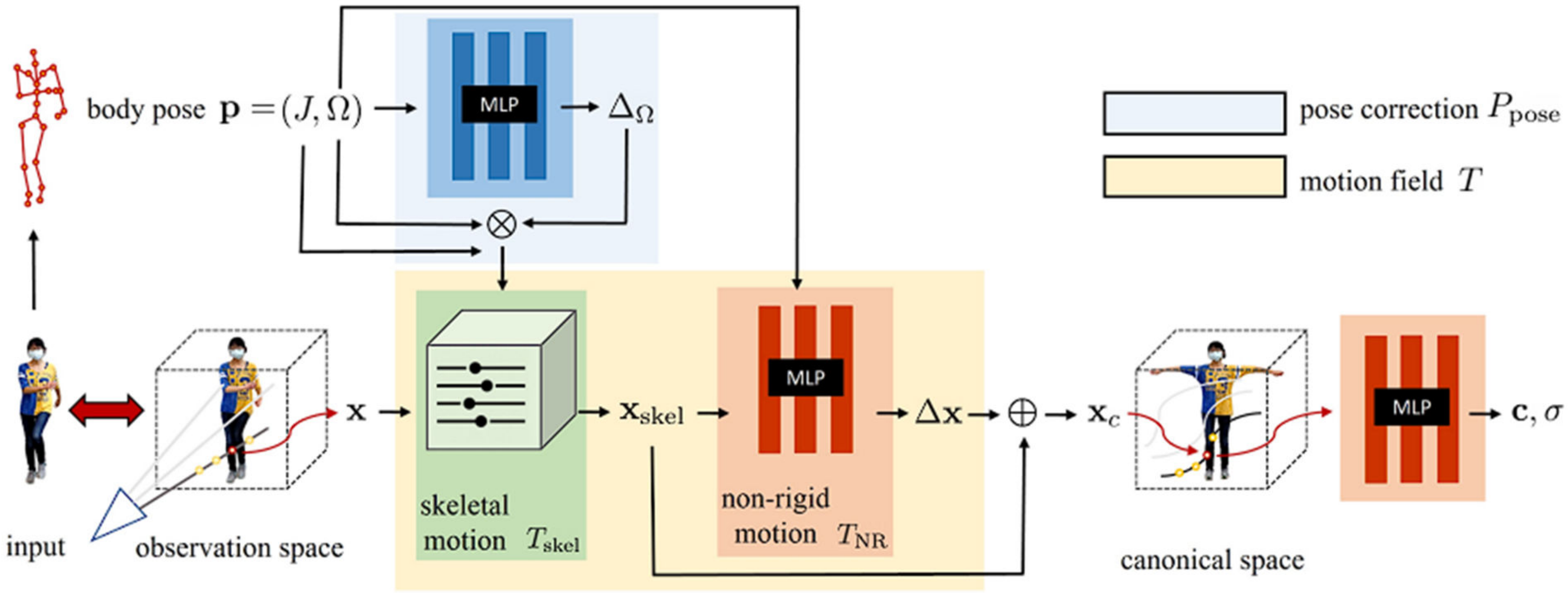
- 将辐射场绑定在动态Mesh人体上



Neural Body: Implicit Neural Representations with Structured Latent Codes for Novel View Synthesis of Dynamic Humans. CVPR 2021.

NeRF数字人表示—HumanNeRF

- 将辐射场绑定在骨架驱动的人体上



HumanNeRF: Free-viewpoint Rendering of Moving People from Monocular Video. CVPR 2022.

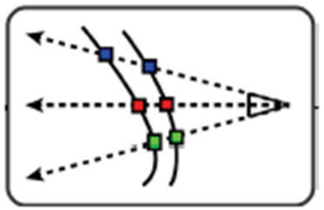
NeRF数字人表示—HumanNeRF

HumanNeRF: Free-viewpoint Rendering of Moving People from Monocular Video. CVPR 2022.

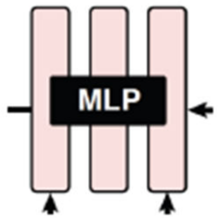
NeRF的一些问题

Rendering NeRF:

1) 逐像素沿着相机的射线在场景中生成一组**采样**的3D点
2) 采样点位置及视线方向输入**MLP**中，输出采样点的**颜色和密度**
3) 使用**体渲染**技术将这些颜色和密度累积成一个2D图像



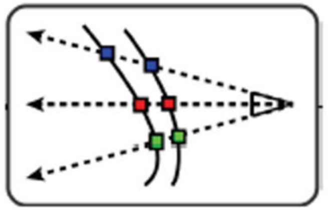
*



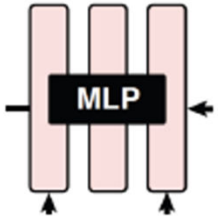
采样点数量 * 逐点计算成本 = 渲染成本

Training: 1~2 days
Rendering: ~10s / frame, 0.1 fps

NeRF加速表示—Key Idea

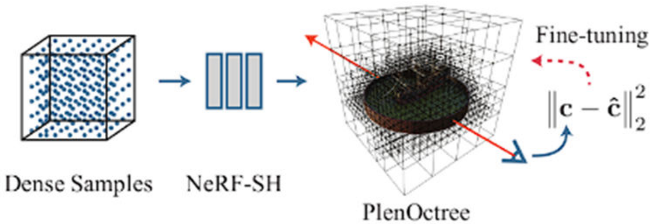


*

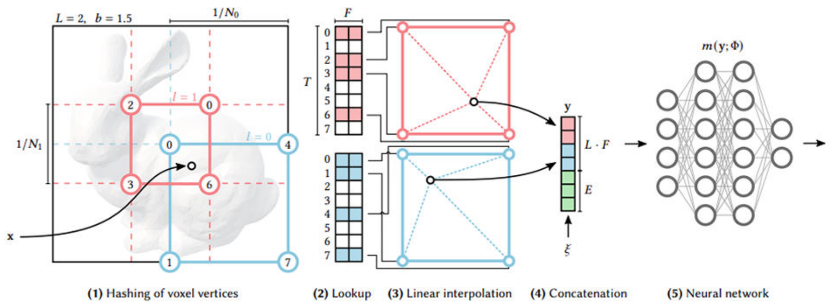


采样点数量 * 逐点计算成本 = 渲染成本

- 结构化 / 显式存储，消除/降低MLP计算成本，减少采样



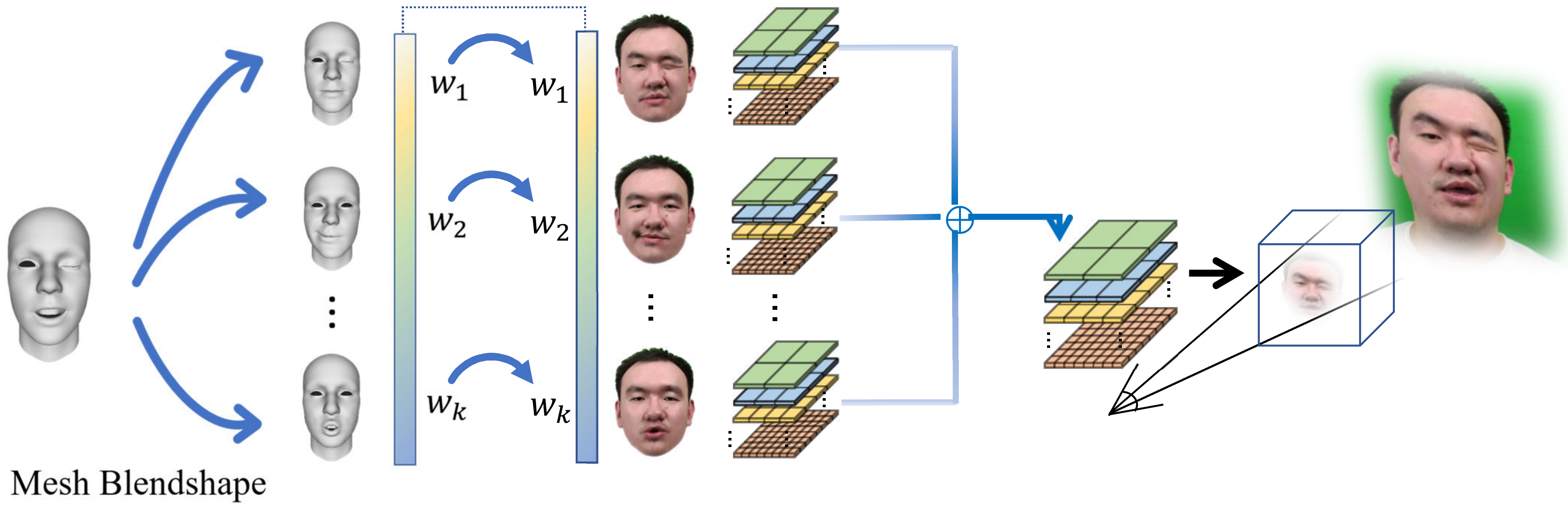
PlenOctrees, ICCV 2021



InstantNGP, SIGGRAPH 2022

NeRF加速数字人表示—NeRFBlendshape

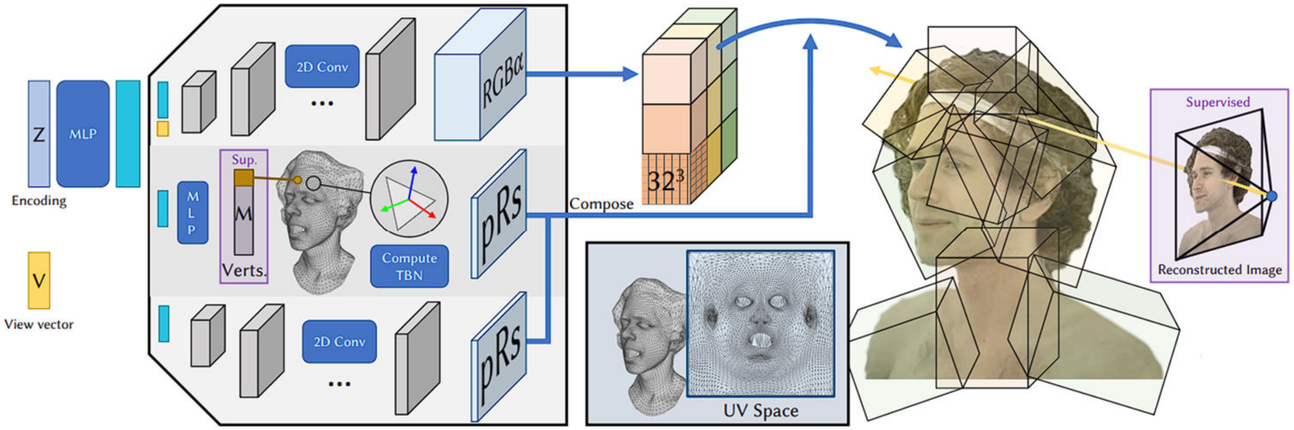
- 将每个表情Blendshae建模一个高效NeRF表示



Reconstructing Personalized Semantic Facial NeRF Models From Monocular Video. SIGGRAPH Asia 2022.

NeRF加速数字人表示—MVP

- 将三维辐射场规则化到体元 (Primitives) 表示:



Mixture of Volumetric Primitives. SIGGRAPH 2021.

Still not that Efficient

20 FPS: 512x512 on RTX 3090

Super fast than original NeRF

Not enough for real-time interaction

Poor performance on mobile devices

表 示

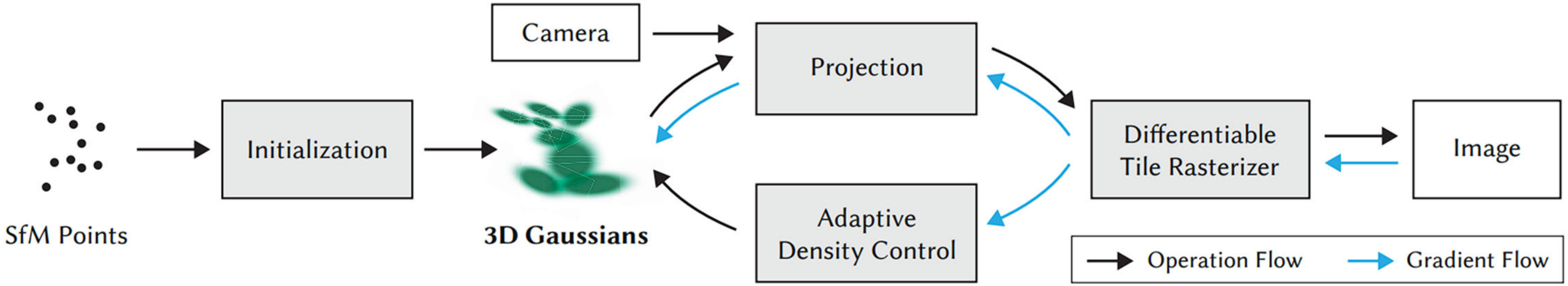
1. Mesh数字人
2. NeRF数字人
3. 3DGS数字人

3DGS简介

- 将辐射场表示为分布在物体表面的具有一定辐射范围（高斯分布）的球体，并赋予每个球体颜色和密度属性表示附近的辐射场

- 显式表达，无需MLP
- Splatting, 无需采样

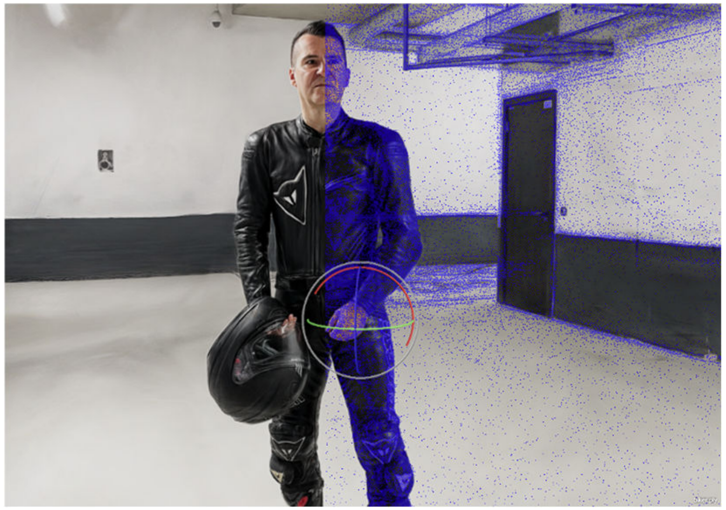
- 实时渲染，大于100FPS
- 快速训练, 几十分钟



3D Gaussian Splatting for Real-Time Radiance Field Rendering. SIGGRAPH 2023.

3DGS简介

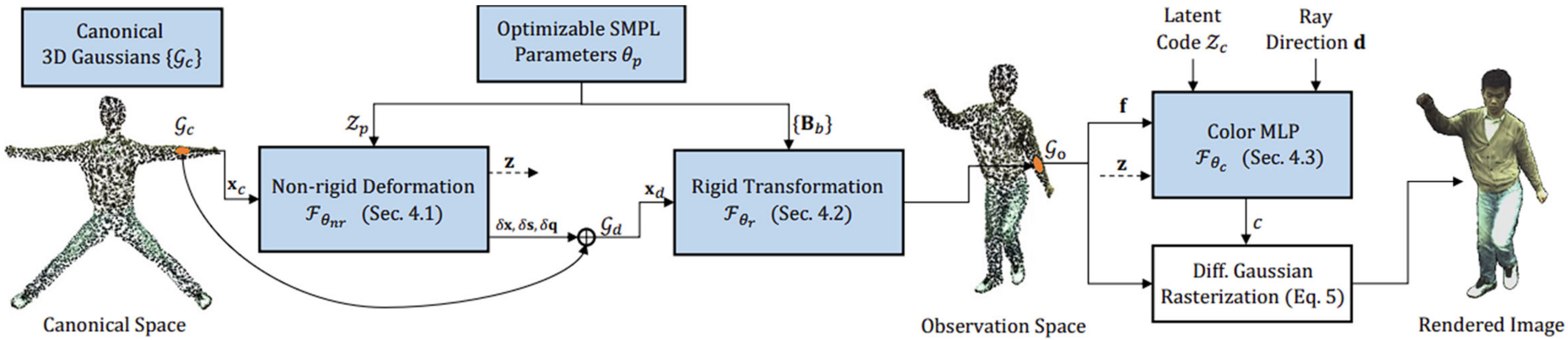
- 3DGS可实现高保真建模与Web端实时渲染



3D Gaussian Splatting for Real-Time Radiance Field Rendering. SIGGRAPH 2023.

3DGS数字人表示—Key Idea

- 将3DGS绑定在参数化Mesh上



3DGS-Avatar: Animatable Avatars via Deformable 3D Gaussian Splatting. CVPR 2024.

3DGS数字人表示

3. [CVPR '24] Animatable Gaussians: Learning Pose-dependent Gaussian Maps for High-fidelity Human Avatar Modeling

Authors: Zhe Li, Zerong Zheng, Lizhen Wang, Yebin Liu

▶ Abstract

Paper | Project Page | Code

4. [CVPR '24] GART: Gaussian Articulated Template Models

Authors: Jiahui Lei, Yufu Wang, Georgios Pavlakos, Lingjie Liu, Kostas Daniilidis

▶ Abstract

Paper | Project Page | Code | Short Presentation

5. [CVPR '24] Human Gaussian Splatting: Real-time Rendering of Animatable Avatars

Authors: Arthur Moreau, Jifei Song, Helisa Dhamo, Richard Shaw, Yiren Zhou, Eduardo Pérez-Pellitero

▶ Abstract

Paper | Project Page | Short Presentation

6. [CVPR '24] HUGS: Human Gaussian Splats

Authors: Muhammed Kocabas, Jen-Hao Rick Chang, James Gabriel, Oncel Tuzel, Anurag Ranjan

▶ Abstract

Paper | Project Page | Code (not yet)

7. [CVPR '24] Gaussian Shell Maps for Efficient 3D Human Generation

Authors: Rameen Abdal, Wang Yifan, Zifan Shi, Yinghao Xu, Ryan Po, Zhengfei Kuang, Qifeng Chen, Dit-Yan Yeung, Gordon Wetzstein

▶ Abstract

8. GaussianHead: High-fidelity Head Avatars with Learnable Gaussian Derivation

Authors: Jie Wang, Jiu-Cheng Xie, Xianyan Li, Chi-Man Pun, Feng Xu, Hao Gao

▶ Abstract

Paper | Project Page | Code

9. [CVPR '24] GaussianAvatars: Photorealistic Head Avatars with Rigged 3D Gaussians

Authors: Shenhan Qian, Tobias Kirschstein, Liam Schoneveld, Davide Davoli, Simon Giebenhain, Matthias Nießner

▶ Abstract

Paper | Project Page | Code | Short Presentation

10. [CVPR '24] GPS-Gaussian: Generalizable Pixel-wise 3D Gaussian Splatting for Real-time Human Novel View Synthesis

Authors: Shunyuan Zheng, Boyao Zhou, Ruizhi Shao, Boning Liu, Shengping Zhang, Liqiang Nie, Yebin Liu

▶ Abstract

Paper | Project Page | Code | Short Presentation

11. GauHuman: Articulated Gaussian Splatting from Monocular Human Videos

Authors: Shoukang Hu Ziwei Liu

▶ Abstract

Paper | Project Page | Code | Short Presentation

12. HeadGaS: Real-Time Animatable Head Avatars via 3D Gaussian Splatting

Authors: Helisa Dhamo, Yinyu Nie, Arthur Moreau, Jifei Song, Richard Shaw, Yiren Zhou, Eduardo Pérez-Pellitero

▶ Abstract

13. [CVPR '24] HiFi4G: High-Fidelity Human Performance Rendering via Compact Gaussian Splatting

Authors: Yuheng Jiang, Zhehao Shen, Penghao Wang, Zhuo Su, Yu Hong, Yingliang Zhang, Jingyi Yu, Lan Xu

▶ Abstract

Paper | Project Page | Short Presentation | Dataset

14. [CVPR '24] GaussianAvatar: Towards Realistic Human Avatar Modeling from a Single Video via Animatable 3D Gaussians

Authors: Liangxiao Hu, Hongwen Zhang, Yuxiang Zhang, Boyao Zhou, Boning Liu, Shengping Zhang, Liqiang Nie

▶ Abstract

Paper | Project Page | Code | Short Presentation

15. [CVPR '24] FlashAvatar: High-fidelity Head Avatar with Efficient Gaussian Embedding

Authors: Jun Xiang, Xuan Gao, Yudong Guo, Juyong Zhang

▶ Abstract

Paper | Project Page | Code

16. [CVPR '24] Relightable Gaussian Codec Avatars

Authors: Shunsuke Saito, Gabriel Schwartz, Tomas Simon, Junxuan Li, Giljoo Nam

▶ Abstract

Paper | Project Page

17. MonoGaussianAvatar: Monocular Gaussian Point-based Head Avatar

Authors: Yufan Chen, Lizhen Wang, Qijing Li, Hongjiang Xiao, Shengping Zhang, Hongxun Yao, Yebin Liu

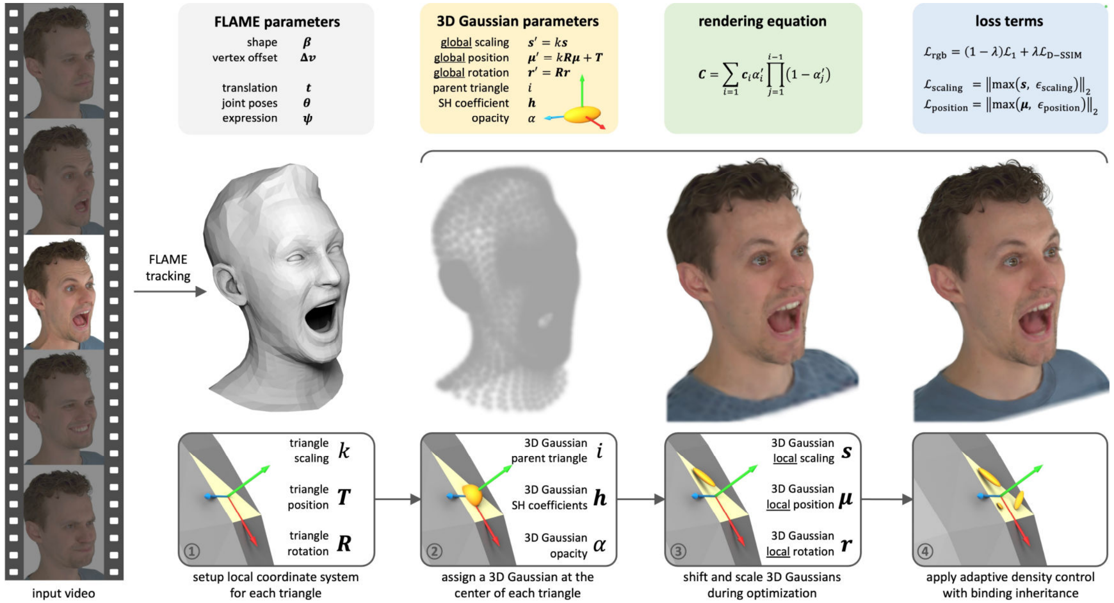
▶ Abstract

Paper | Project Page | Code (not yet) | Short Presentation

https://github.com/MrNeRF/awesome-3D-gaussian-splatting

3DGS数字人表示—GaussianAvatars

- 将3DGS绑定在FLAME Mesh



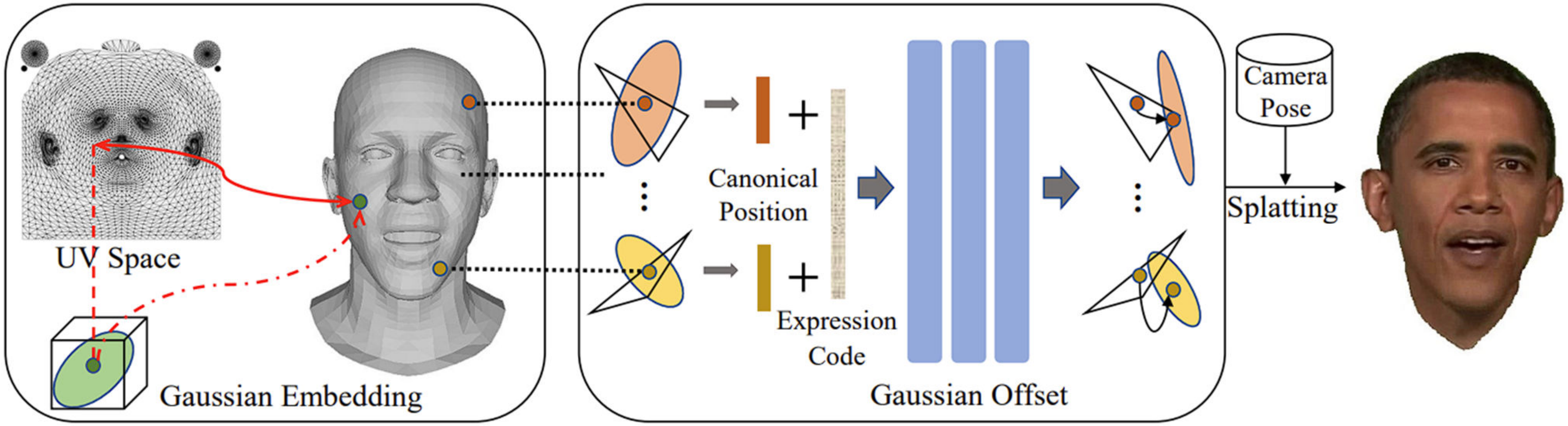
Deform Gaussians based on the underlying parametric head model

---

GaussianAvatars: Photorealistic Head Avatars with Rigged 3D Gaussians. CVPR 2024.

3DGS数字人表示—FlashAvatar

- 将3DGS绑定在FLAME Mesh UV Space

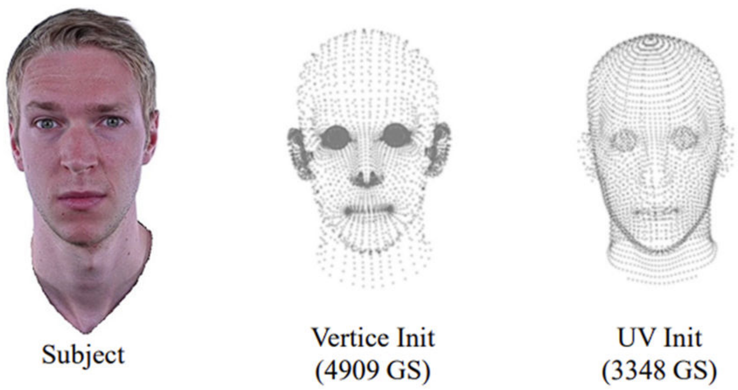


Deform Gaussians based on the underlying parametric head model

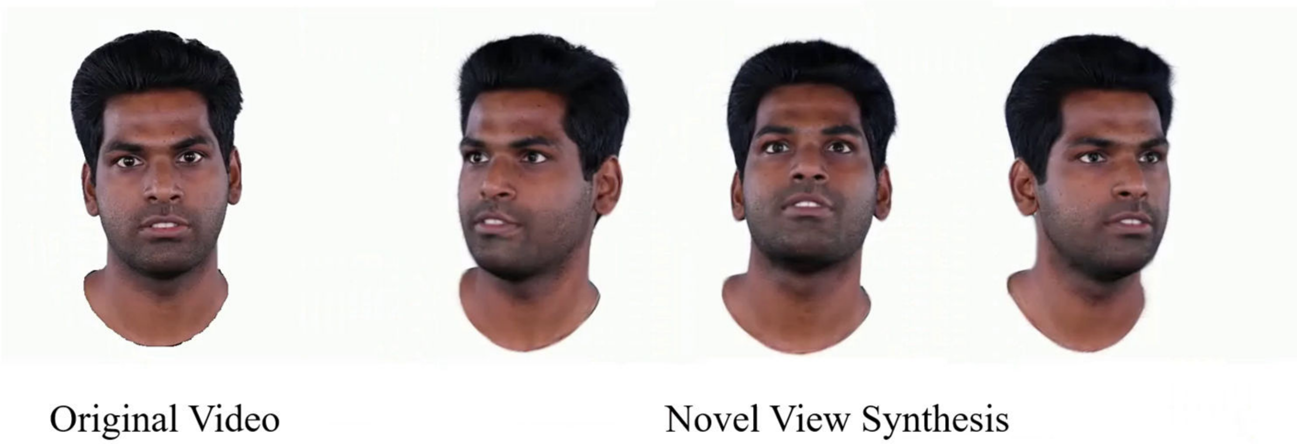
FlashAvatar: High-fidelity Head Avatar with Efficient Gaussian Embedding. CVPR 2024.

3DGS数字人表示—FlashAvatar

- 将3DGS绑定在FLAME Mesh UV Space



Uniformly resampling in the UV space
for compression

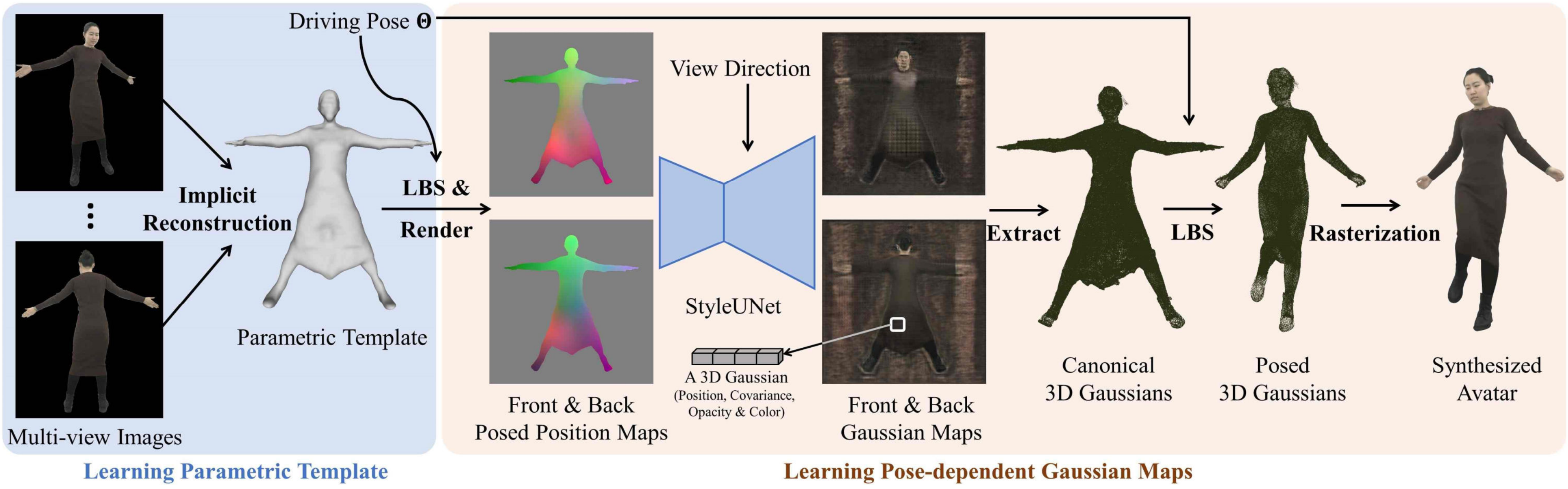


Achieve 300+ FPS rendering

FlashAvatar: High-fidelity Head Avatar with Efficient Gaussian Embedding. CVPR 2024.

3DGS数字人表示—Animatable Gaussians

• 将3DGS绑定在SMPL Mesh上

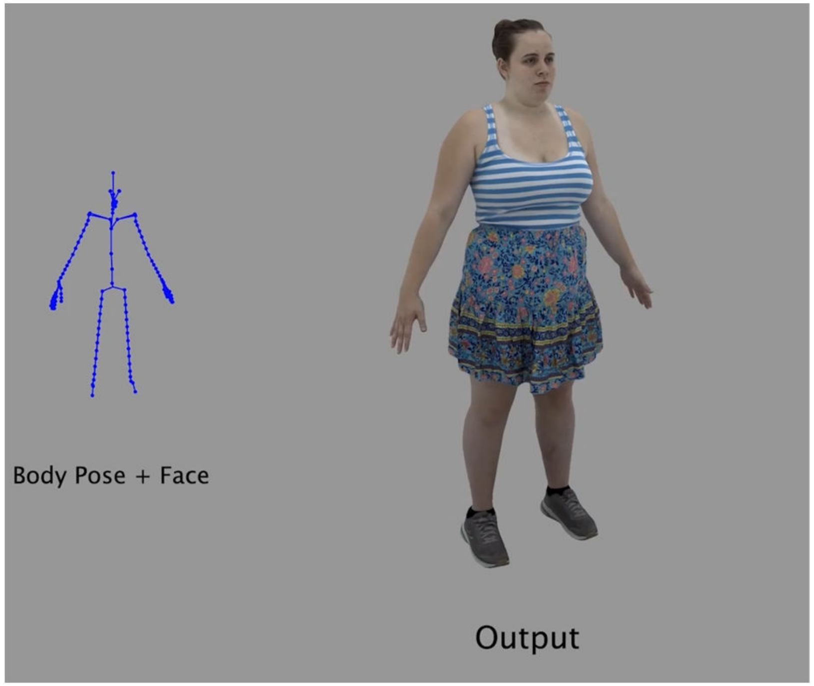


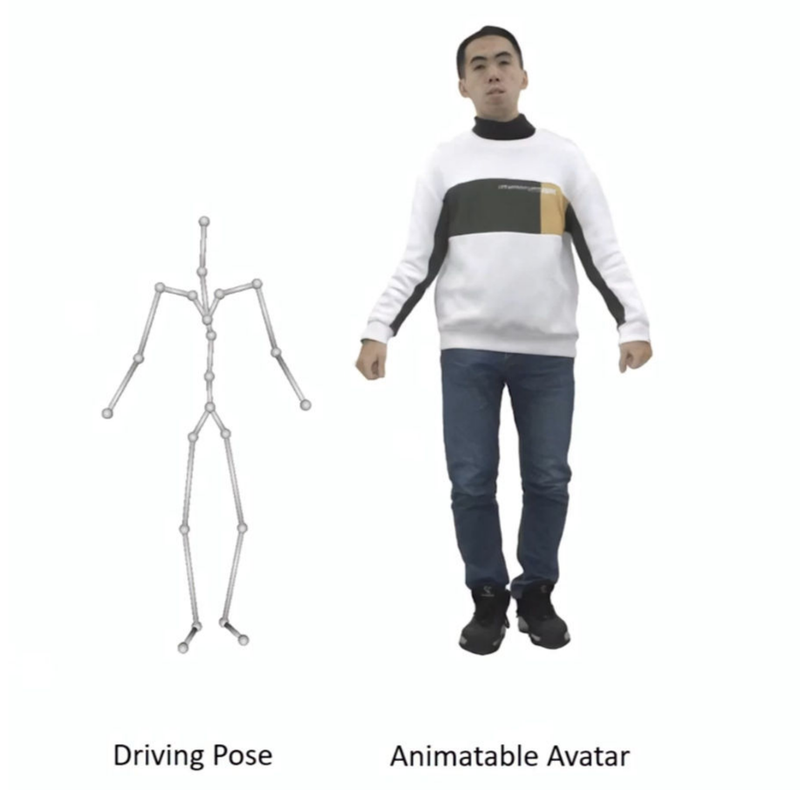
Animatable Gaussians: Learning Pose-dependent Gaussian Maps for High-fidelity Human Avatar Modeling. CVPR 2024.

3DGS数字人表示—Animatable Gaussians

Animatable Gaussians: Learning Pose-dependent Gaussian Maps for High-fidelity Human Avatar Modeling. CVPR 2024.

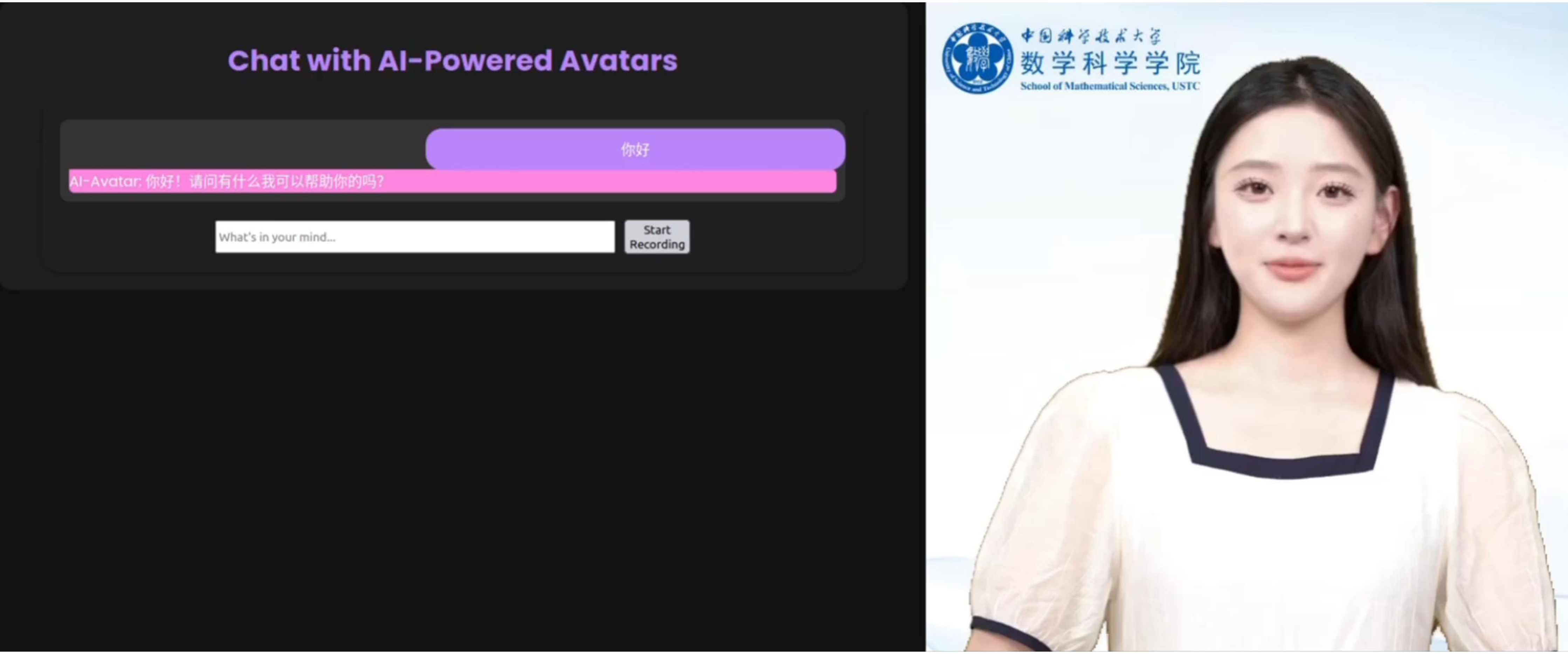
驱动人体NeRF/3DGS





Faithful Full-Body Telepresence with Dynamic Clothing Driven by Sparse RGB-D Input. SIGGRAPH Asia 2023.
Animatable Gaussians: Learning Pose-dependent Gaussian Maps for High-fidelity Human Avatar Modeling. CVPR 2024.

实时互动


Chat with AI-Powered Avatars
你好
AI-Avatar: 你好！请问有什么我可以帮助你的吗？
What's in your mind...
Start Recording
中国科学技术大学
数学科学学院
School of Mathematical Sciences, USTC

Summary

- Mesh数字人：
  - 高效渲染
  - 便于驱动

  - 不够Photo-Realistic
  - 1:1复刻真人难度高

- NeRF数字人：
  - 建模方便
  - 可实现高保真建模

  - 效率低
  - 无原生好的几何
  - 无法原生驱动

- 3DGS数字人：
  - 建模方便
  - 可实现高保真建模
  - 建模、渲染效率高

  - 无原生好的几何
  - 无法原生驱动