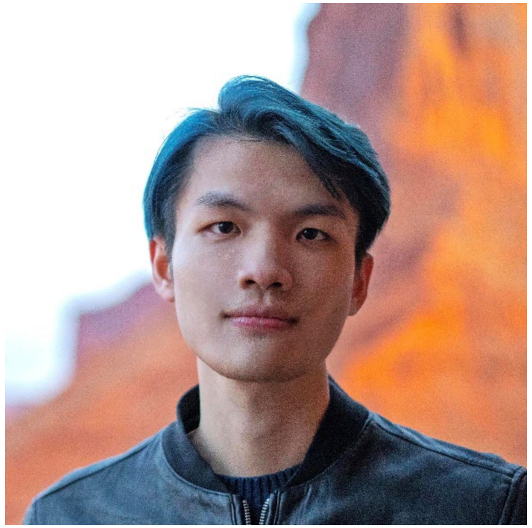
Invited Speakers

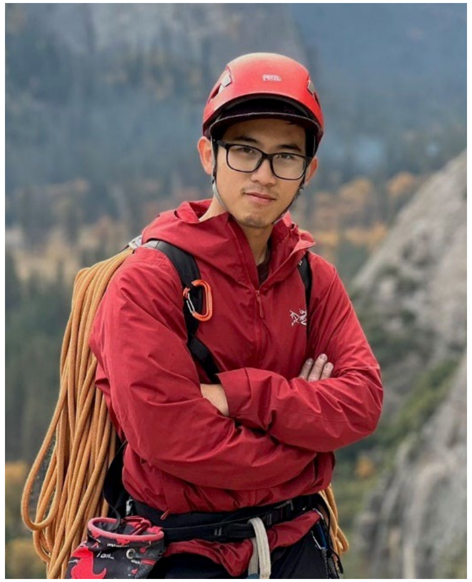
Jiatao Gu
Research Scientist at Apple
Incoming Assistant Professor at Penn

Oct 14 (Mon)



Ruoshi Liu
PhD at Columbia University

Oct 16 (Wed)



Guandao Yang
Postdoc at Stanford University

Oct 23 (Wed)

Lingjie Liu

3D GANs

Lingjie Liu

Oct 9, 2024

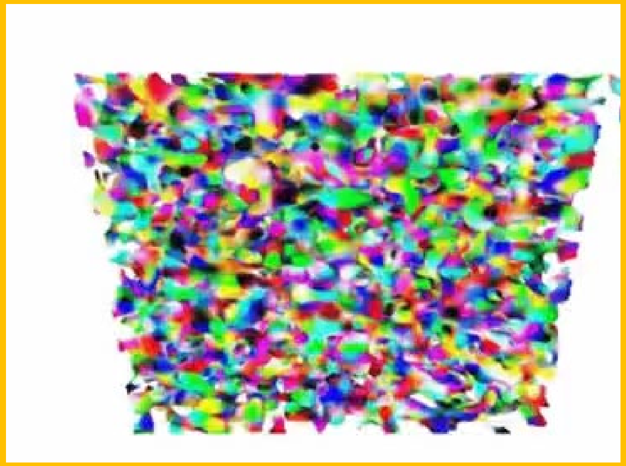
3D GANs

GRAF, pi-GAN, StyleNeRF, EG3D,
GIRAFFE, Next3D, …

Lifting 2D Diffusion Models for 3D Generation

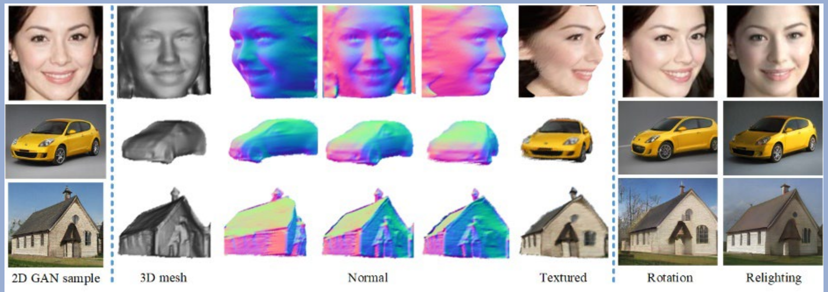
NerfDiff, DreamFusion, …

3D Diffusion Models



GAUDI, DiffRF, SSDNeRF, …

Lifting 2D GANs for 3D Generation



GAN2Shape, GAN2X, …

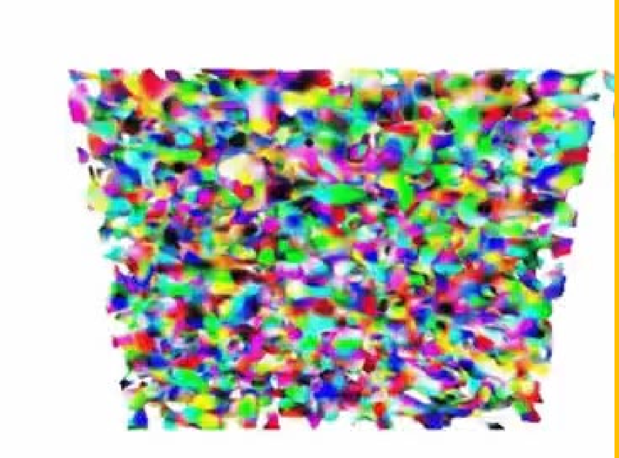
3D GANs

GRAF, pi-GAN, StyleNeRF, EG3D, GIRAFFE, Next3D, …

Lifting 2D Diffusion Models for 3D Generation

NerfDiff, DreamFusion, …

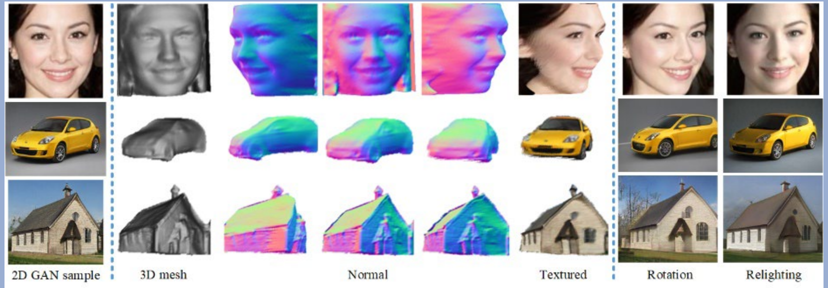
3D Diffusion Models



GAUDI, DiffRF, SSDNeRF, …

Lingjie Liu

Lifting 2D GANs for 3D Generation



GAN2Shape, GAN2X, …

3D GANs

GRAF, pi-GAN, StyleNeRF, EG3D,
GIRAFFE, Next3D, …

Lifting 2D Diffusion Models for 3D Generation

NerfDiff, DreamFusion, …

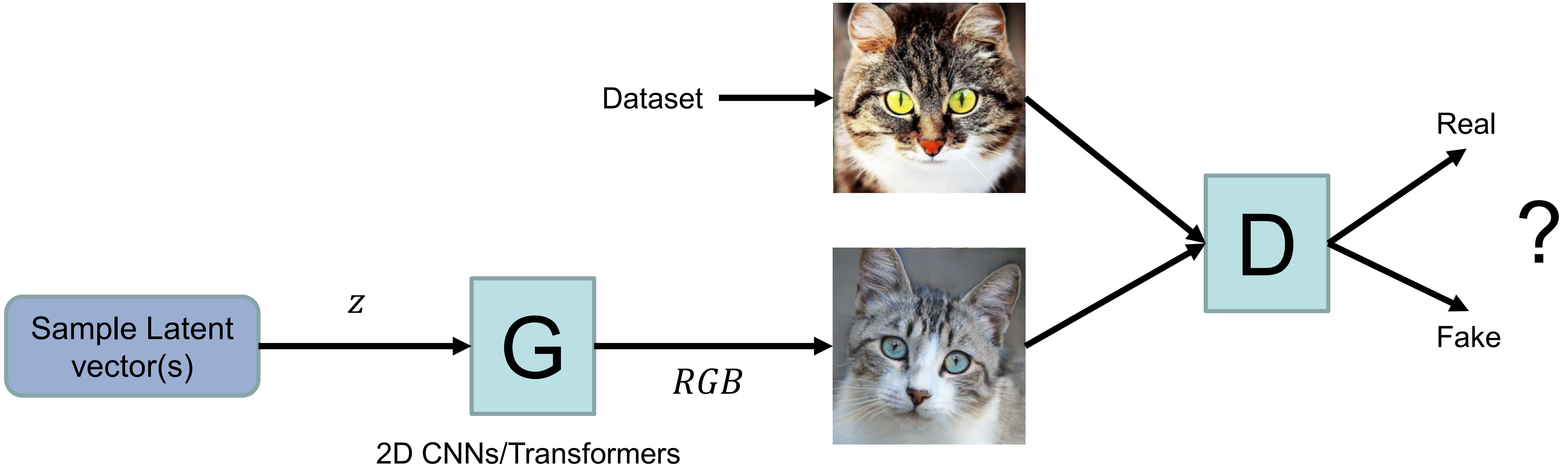
3D Diffusion Models

GAUDI, DiffRF, SSDNeRF, …

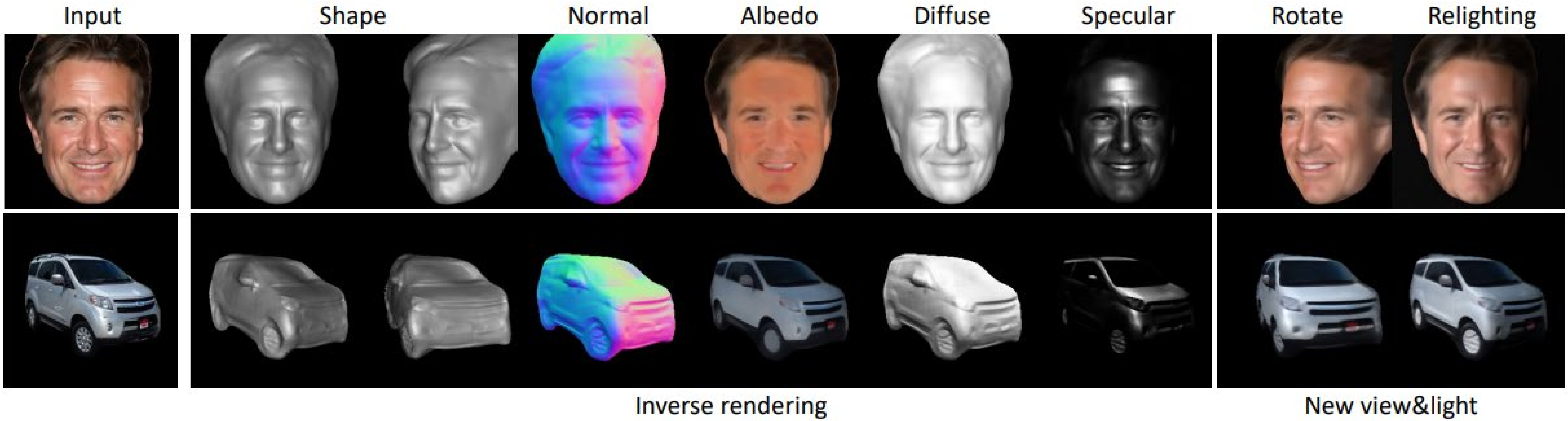
Lingjie Liu

Background

- Generative Adversarial Networks (GANs)

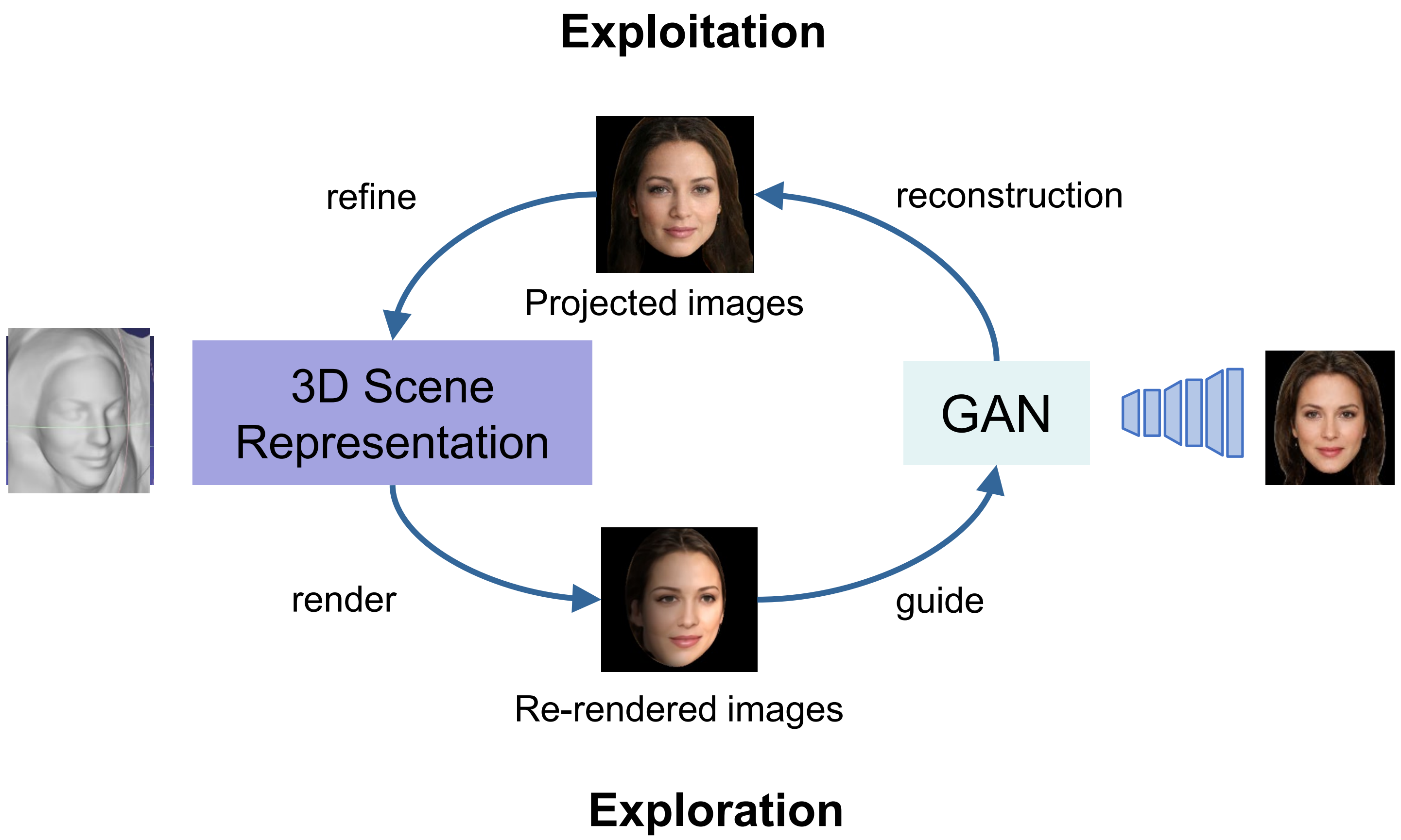


GAN2X: Non-Lambertian Inverse Rendering of Image GANs



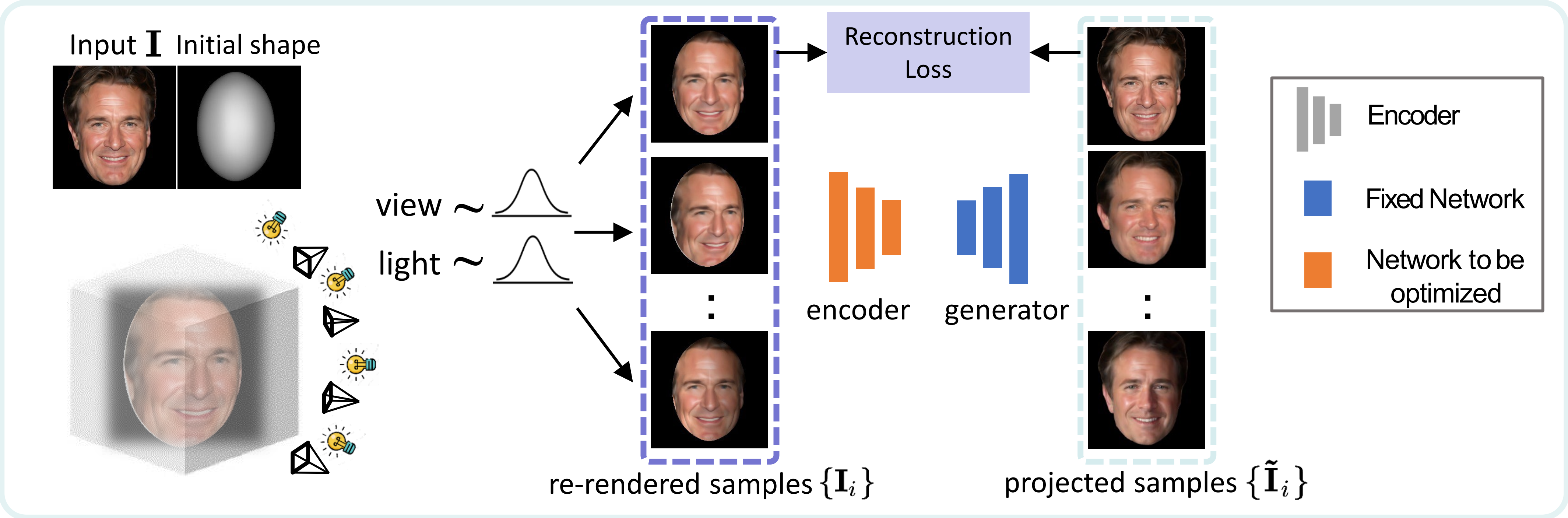
Pan et al. 2022. GAN2X: Non-Lambertian Inverse Rendering of Image GANs, 3DV 2022

Method



Lingjie Liu

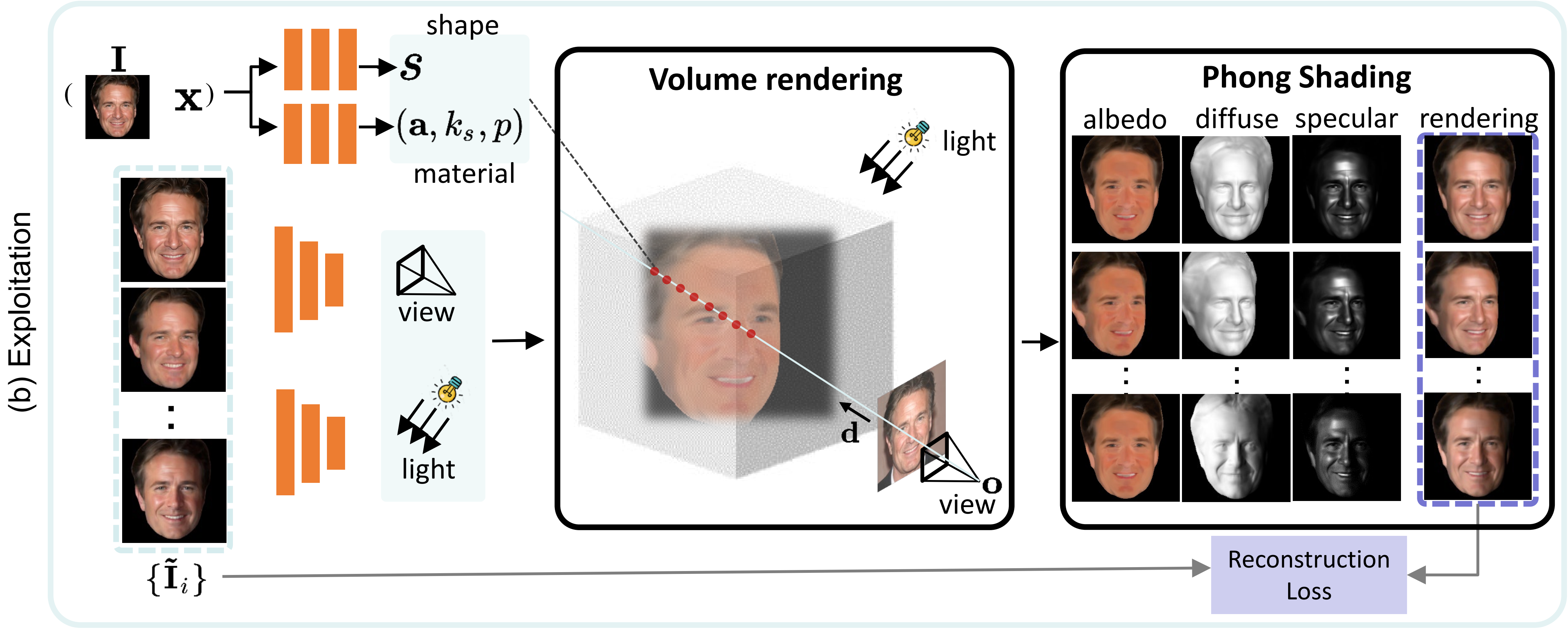
Method: Exploration



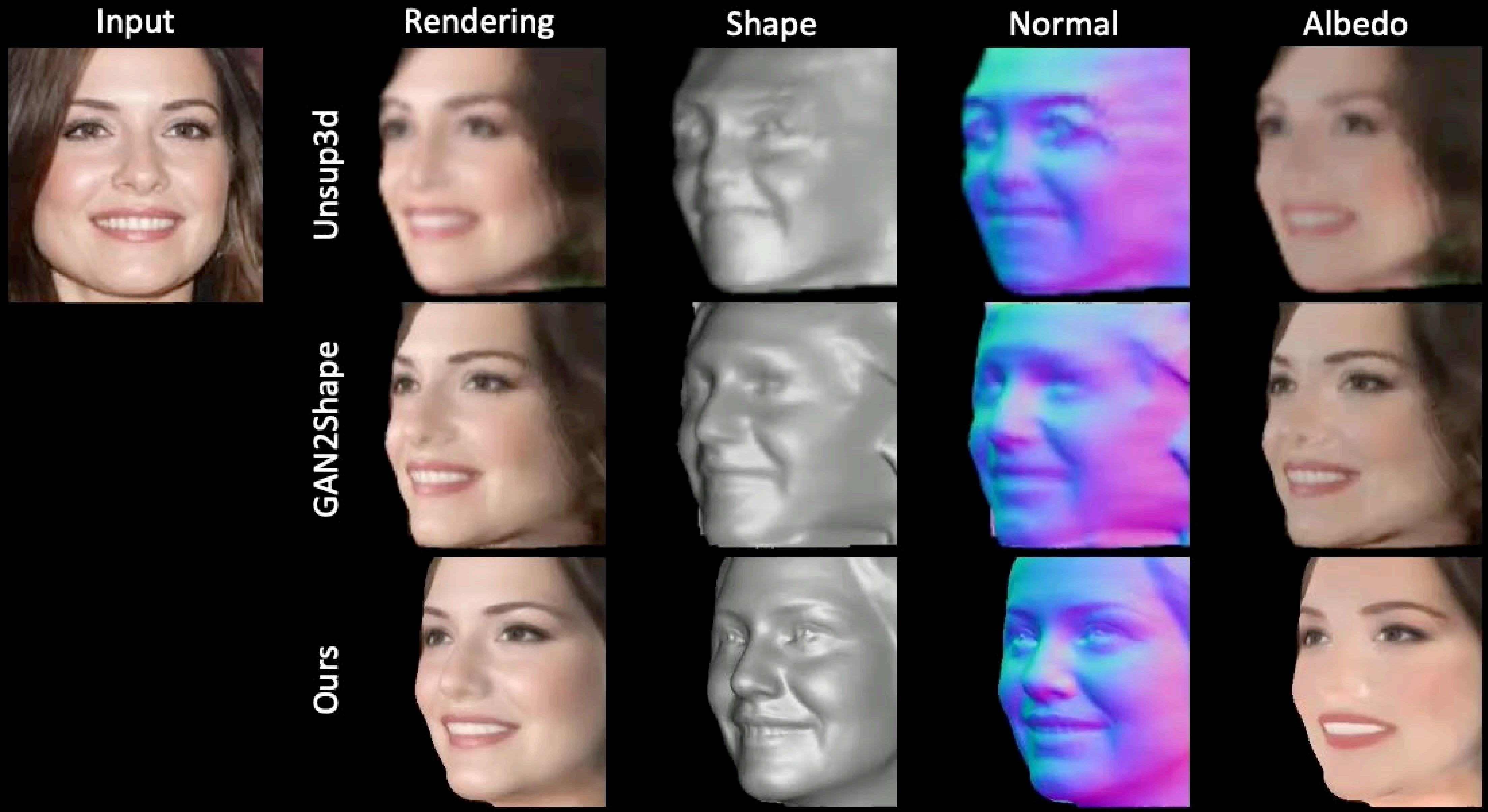
$$\boldsymbol{\theta}_w = \arg\min_{\boldsymbol{\theta}_w} \frac{1}{m} \sum_{i=0}^{m} \mathcal{L}\Big(\mathbf{I}_i, G\big(E_w(\mathbf{I}_i) + \boldsymbol{w}\big)\Big) + \lambda_1 \|E_w(\mathbf{I}_i)\|_2 \qquad (4)$$

Lingjie Liu

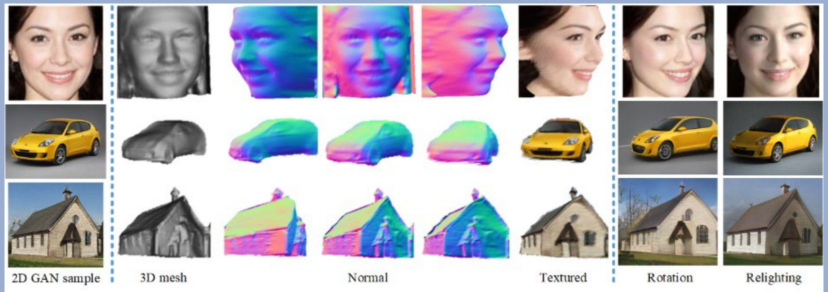
Method: Exploitation



Qualitative Comparison on CelebA: Rotation



Lifting 2D GANs for 3D Generation



GAN2Shape, GAN2X, …

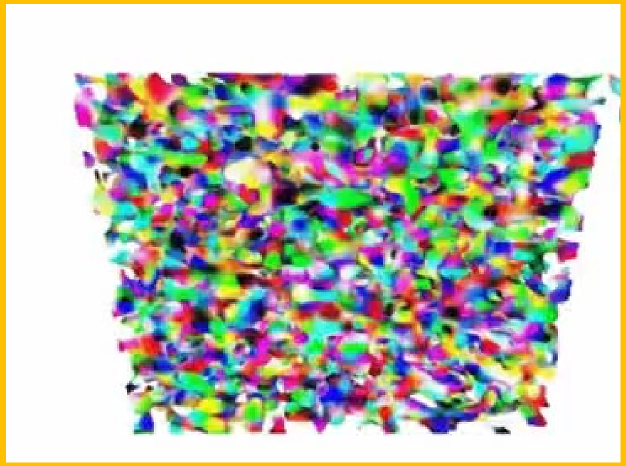
3D GANs

GRAF, pi-GAN, StyleNeRF, EG3D,
GIRAFFE, Next3D, …

Lifting 2D Diffusion Models for 3D Generation

NerfDiff, DreamFusion, …

3D Diffusion Models

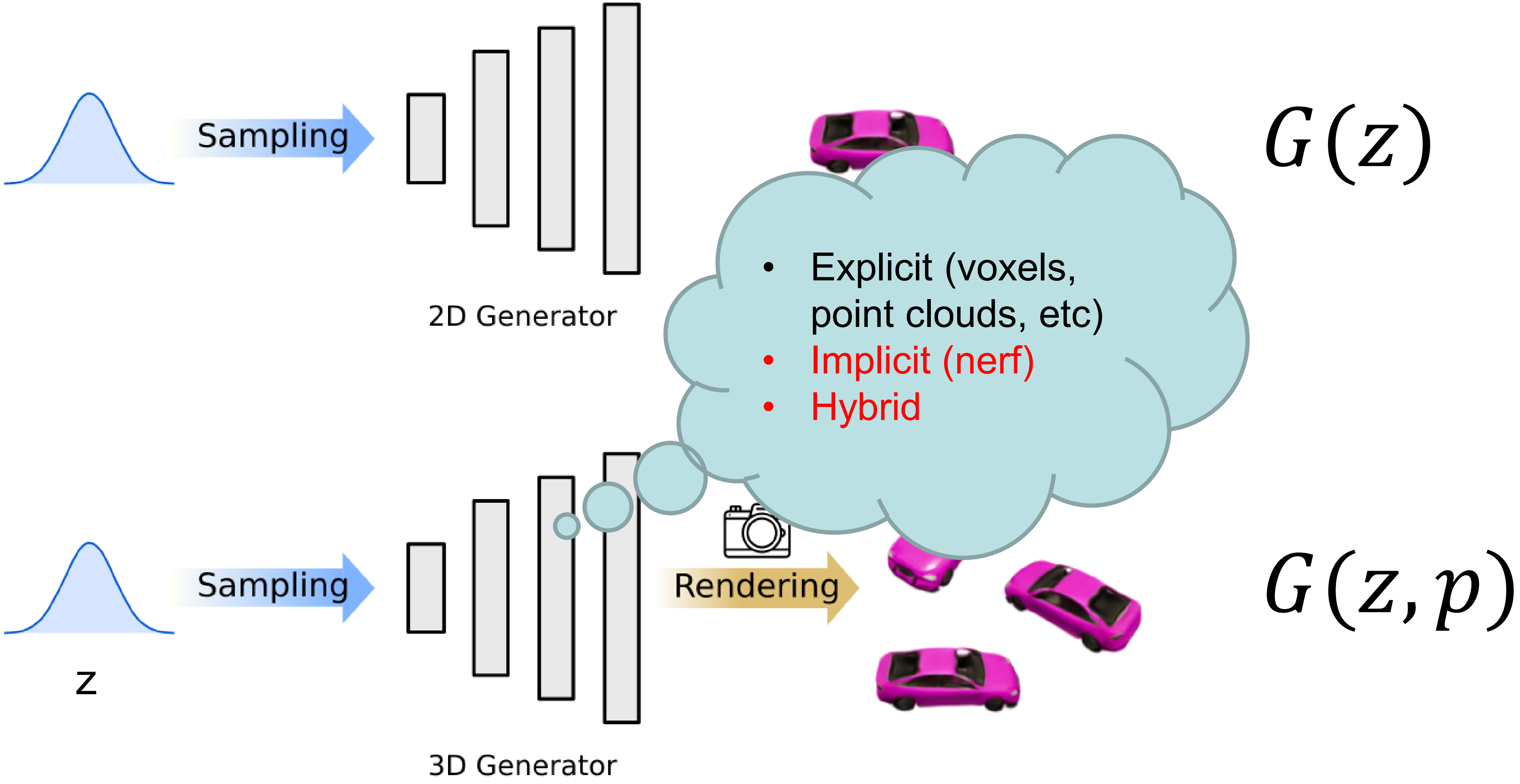


GAUDI, DiffRF, SSDNeRF, …

Lingjie Liu

2D GANs -> 3D GANs

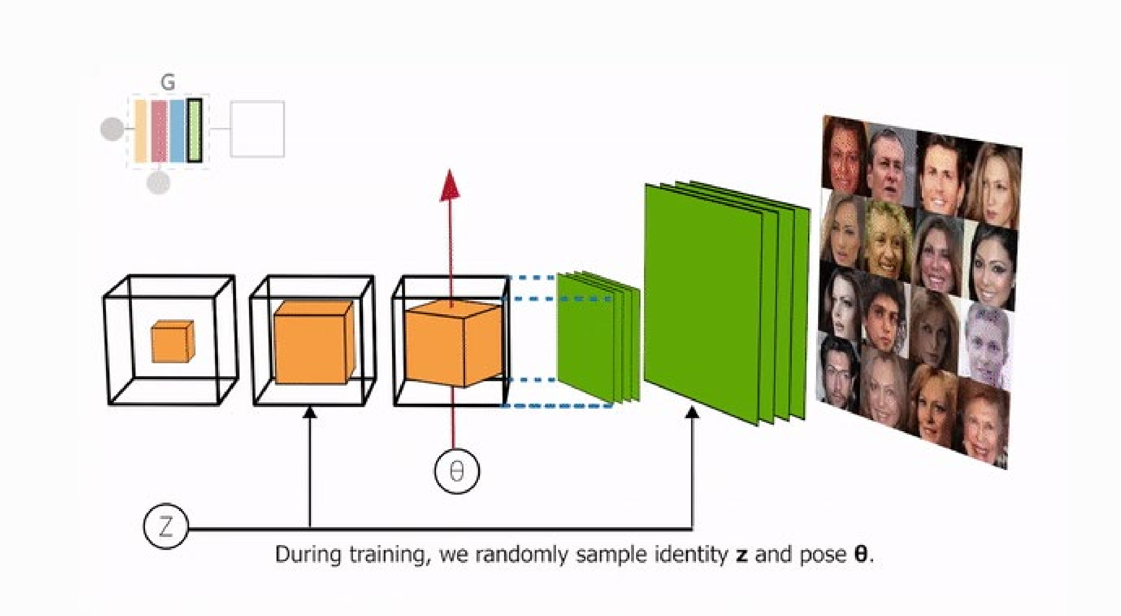
- Making GANs 3D aware/consistent



Lingjie Liu

Using Explicit Representation for 3D GANs

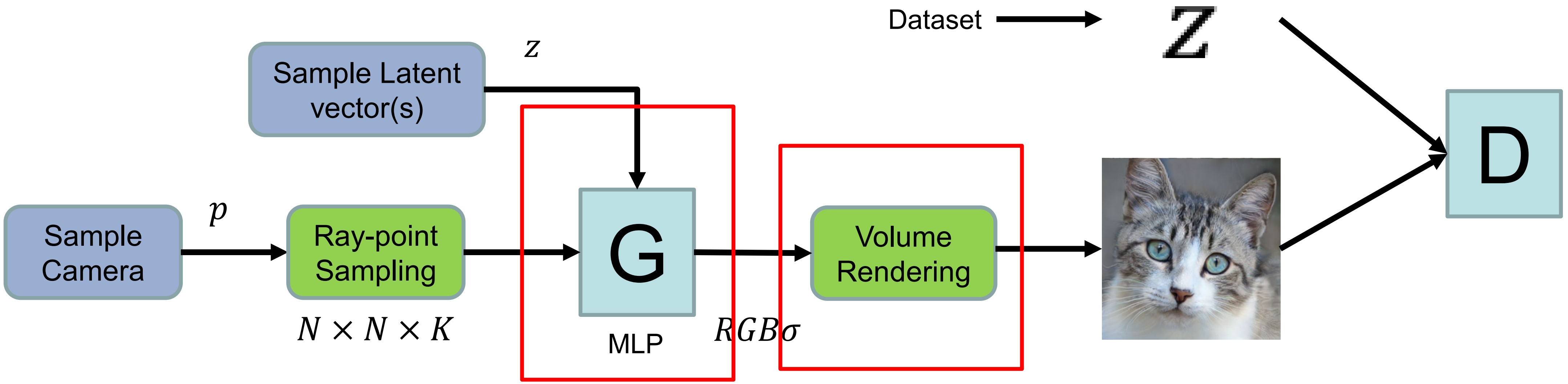
- Explicit representation of 3D GANs
  - HoloGAN
  - BlockGAN



Lingjie Liu

NeRF-GANs

Naïve implementation of putting NeRF into GANs



How to make the generator as expressive as possible

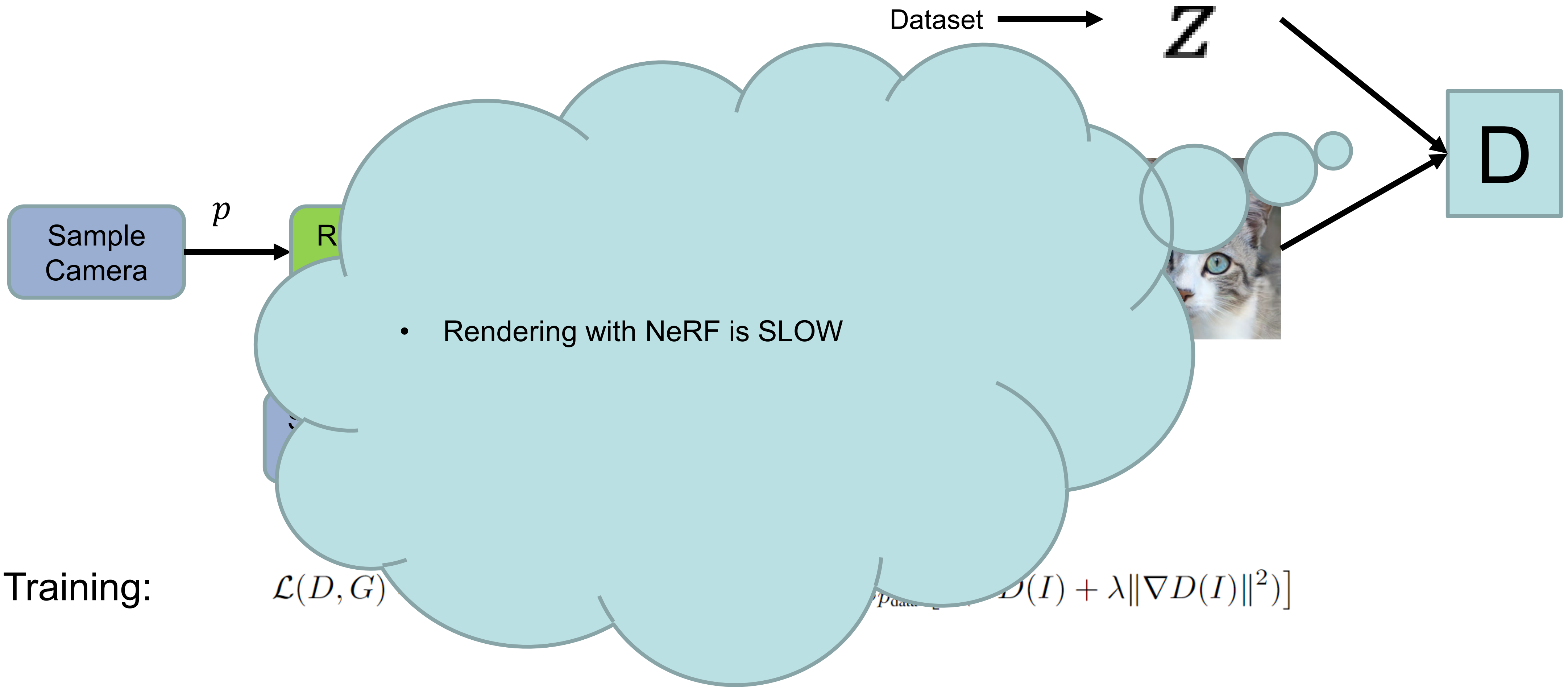
How to make rendering as efficient as possible (during training)

Training: $$\mathcal{L}(D,G) = \mathbb{E}_{\boldsymbol{z}\sim\mathcal{Z},\boldsymbol{p}\sim\mathcal{P}}\left[f(D(G(\boldsymbol{z},\boldsymbol{p})))\right] + \mathbb{E}_{I\sim p_{\text{data}}}\left[f(-D(I) + \lambda\|\nabla D(I)\|^2)\right]$$

Lingjie Liu

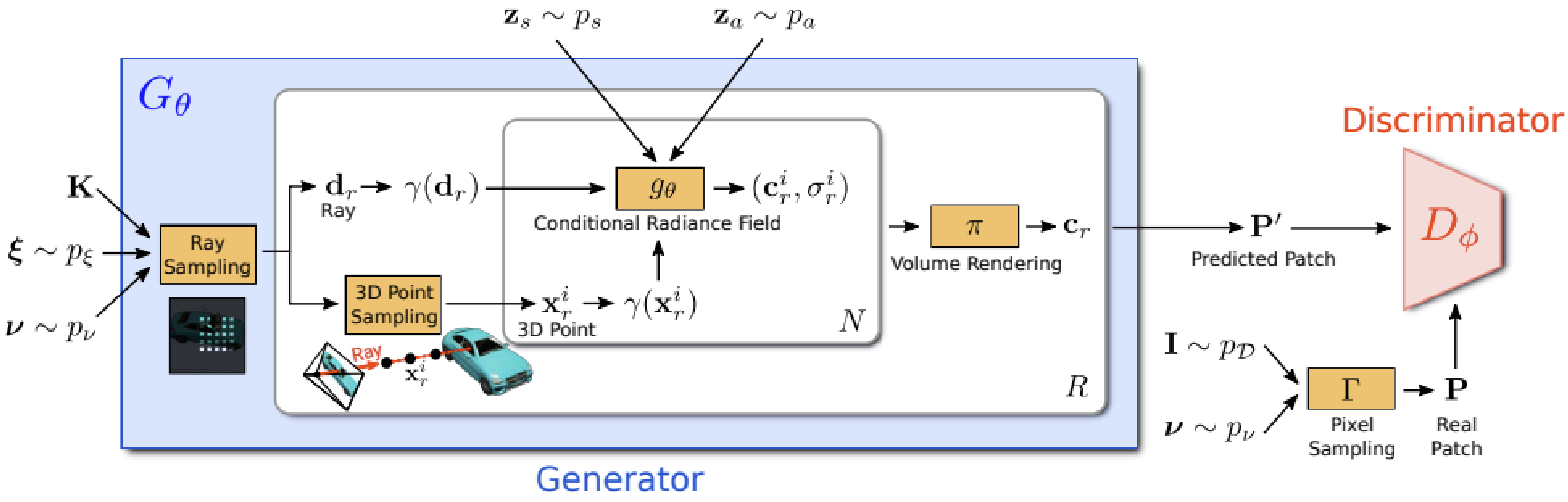
NeRF-GANs

Naïve implementation of putting NeRF into GANs



GRAF

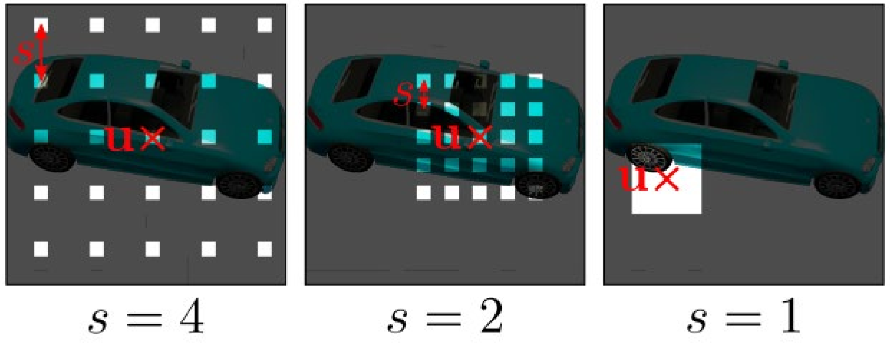
- GRAF is the first work combining NeRF in GAN framework
  - Simple MLP architecture, global Z concat with input position:
    - Not expressive enough to handle complex scenes
  - Sampling patches for discriminator to speed-up training
    - Worse performance on high-resolution images
    - Not really solve the rendering speed problem

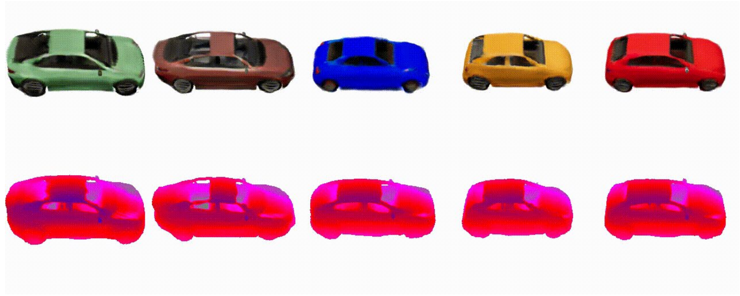


 Schwarz, K., Liao, Y., Niemeyer, M., & Geiger, A. (2020). Graf: Generative radiance fields for 3d-aware image synthesis. arXiv preprint arXiv:2007.02442.

GRAF

- GRAF is the first work combining NeRF in GAN framework
  - Simple MLP architecture, global Z concat with input position:
    - Not expressive enough to handle complex scenes
  - Sampling patches for discriminator to speed-up training
    - Worse performance on high-resolution images
    - Not really solve the rendering speed problem

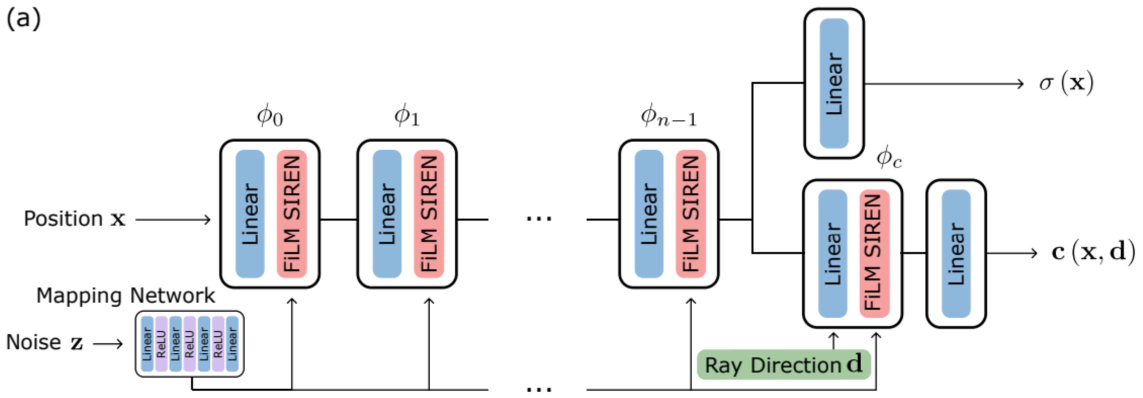


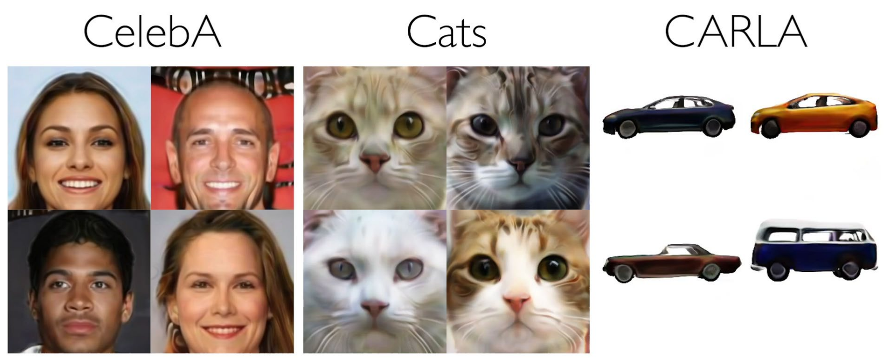


Schwarz, K., Liao, Y., Niemeyer, M., & Geiger, A. (2020). Graf: Generative radiance fields for 3d-aware image synthesis. arXiv preprint arXiv:2007.02442.

Pi-GAN

- Pi-GAN achieves much better visual quality than GRAF
  - StyleGAN like mapping and synthesis network
  - SIREN activations
    - The model is trained at full resolution, so it is slow and can only work in low-resolution
    - The quality is still not good enough and far from 2D models
    - Inference is also slow

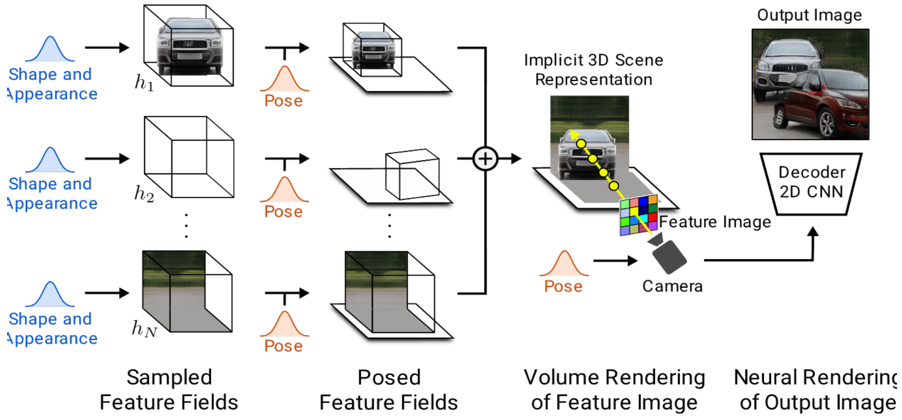


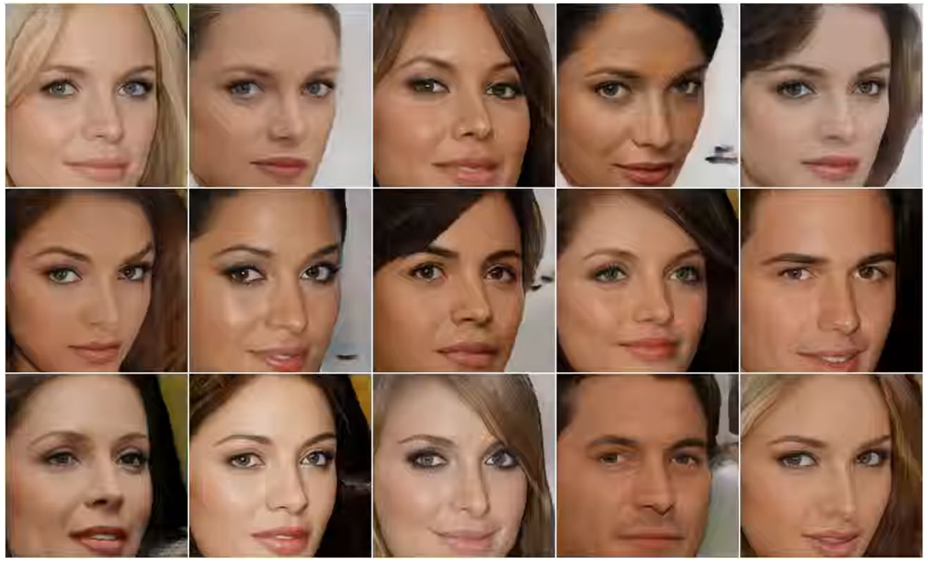


Chan, E. R., Monteiro, M., Kellnhofer, P., Wu, J., & Wetzstein, G. (2021). pi-gan: Periodic implicit generative adversarial networks for 3d-aware image synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 5799-5809).

GiRAFFE

- GiRAFFE proposes a compositional NeRF with a 2D CNN decoder
  - Same as GRAF, z concat with position
  - NeRF works in 16x16 resolution
    - Light-weight, rendering is pretty fast
    - Architecture is too simple to handle complex scenes
    - 2D CNN causes serious inconsistency in the rendered outputs





Niemeyer, M., & Geiger, A. (2021). Giraffe: Representing scenes as compositional generative neural feature fields. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 11453-11464).

StyleNeRF

Our Results: FFHQ Dataset (1024x1024)

This is the first time that a generative model can synthesize high-resolution images from novel views while preserving high 3D consistency

Gu et al. StyleNeRF: A Style-based 3D-Aware Generator for High-resolution Image Synthesis. ICLR 2022.

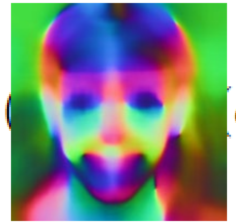
StyleNeRF - Method

- Approximated Volume Rendering

$$I_{\boldsymbol{w}}^{\text{NeRF}}(\boldsymbol{r}) = \int_0^\infty p_{\boldsymbol{w}}(t)\boldsymbol{c}_{\boldsymbol{w}}(\boldsymbol{r}(t), \boldsymbol{d})dt, \quad \text{where } \; p_{\boldsymbol{w}}(t) = \text{ex}$$

$$(\boldsymbol{r}(t))$$

$$I_{\boldsymbol{w}}^{\text{Approx}}(\boldsymbol{r}) = \int_0^\infty p_{\boldsymbol{w}}(t) \cdot h_c \circ [\phi_{\boldsymbol{w}}^{n_c}(\boldsymbol{r}(t)), \zeta\left(\boldsymbol{d}\right)]\, dt \approx h_c \circ [\phi_{\boldsymbol{w}}^{n_c, n_\sigma}$$



$$\boldsymbol{d})]$$

Early aggregation

$$\phi_{\boldsymbol{w}}^{n,n_\sigma}(\mathcal{A}(R_H)) \approx \texttt{Upsample}\left(\phi_{\boldsymbol{w}}^{n,n_\sigma}(\mathcal{A}(R_L))\right)$$

2D upsampling

StyleNeRF - Method

- Remove view direction input
  - We found that view direction will break the consistency and did not contribute to much quality (our dataset is single image)

- Up-sampler design

$$\texttt{Upsample}(X) = \texttt{Conv2d}(\texttt{Pixelshuffle}(\texttt{Repeat}(X, 4) + \psi_\theta(X), 2), K)$$

- NeRF-path regularization

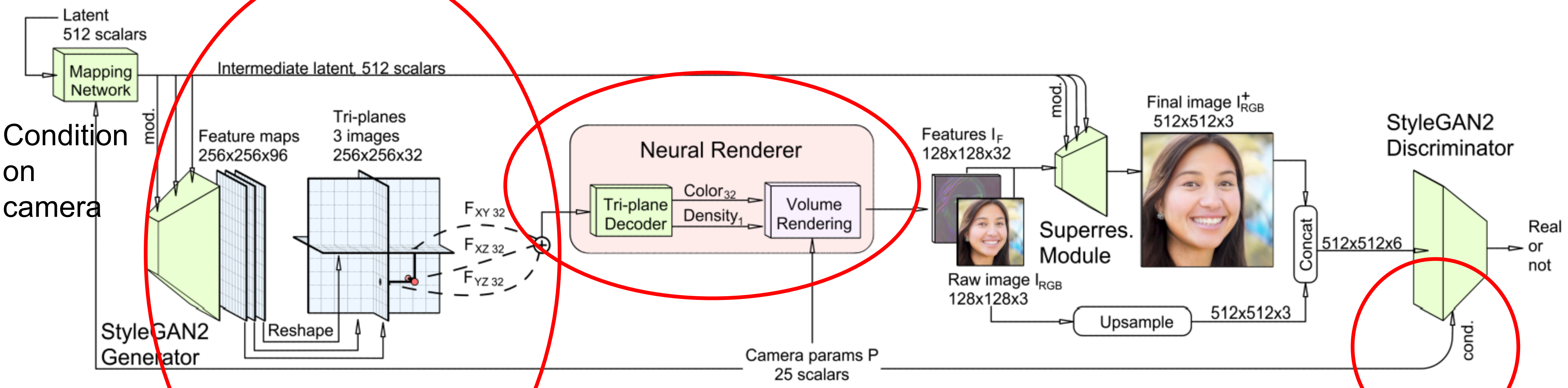
$$\mathcal{L}_{\text{NeRF-path}} = \frac{1}{|S|} \sum\nolimits_{(i,j)\in S} \left(I_{\boldsymbol{w}}^{\text{Approx}}(R_{\text{in}})[i,j] - I_{\boldsymbol{w}}^{\text{NeRF}}(R_{\text{out}}[i,j])\right)^2$$

Lingjie Liu

EG3D

- https://matthew-a-chan.github.io/EG3D/

Condition on camera



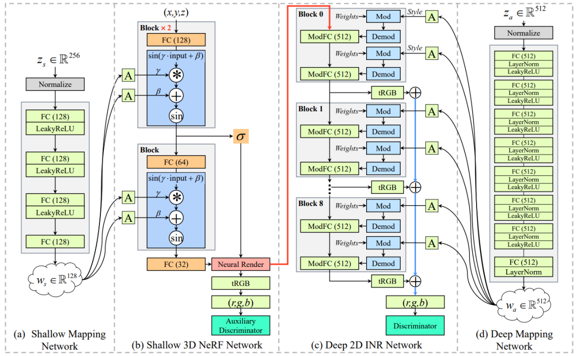
Tiny MLP

Condition on camera

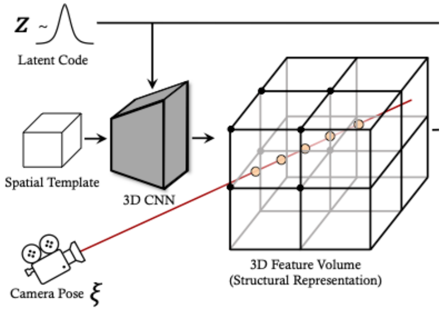
Tri-plane representations

Chan et al. EG3D: Efficient Geometry-aware 3D Generative Adversarial Networks. CVPR 2022.

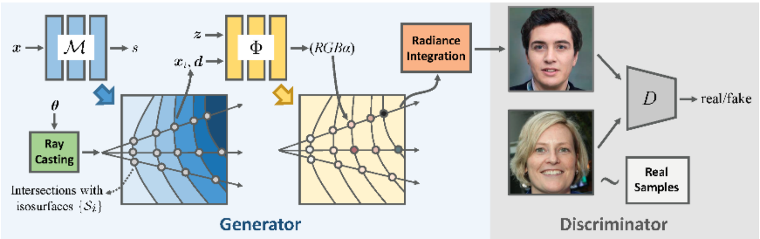
Explosion of 3D GANs



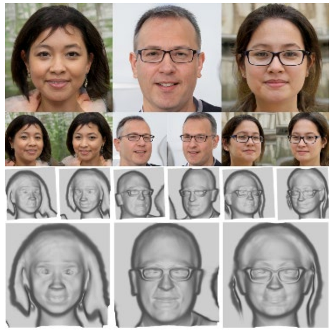
CIPS-3D [Zhou et al. 2021]



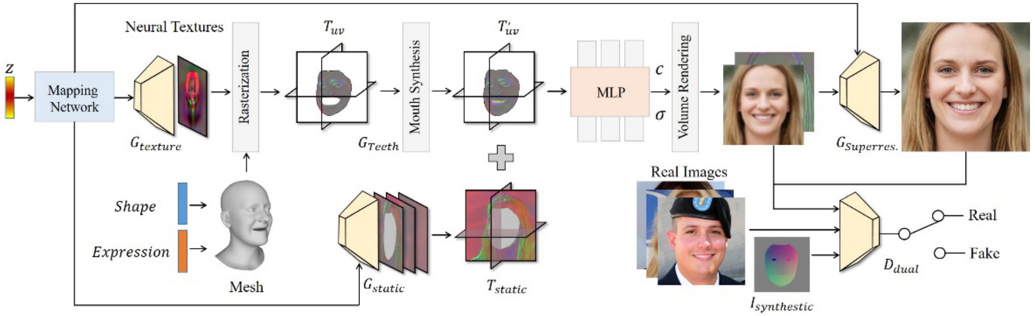
VolumeGAN [Xu et al. 2022]



GRAM [Deng et al. 2022]



StyleSDF [Or-El et al. 2022]



Next3D [Sun et al. 2023]

...