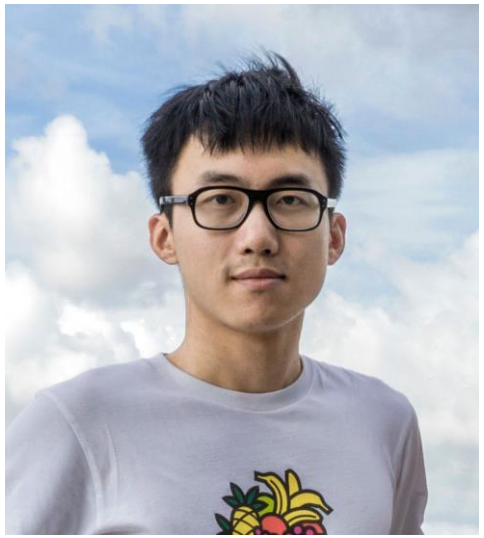
Learning Shapes from Unposed 2D Images using Generative Adversarial Networks

Dr. Dai Bo

Research Assistant Professor, S-Lab for Advanced Intelligence

Team Members



Pan Xingang

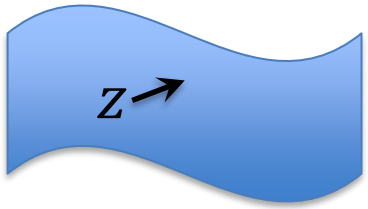
Xu Xudong

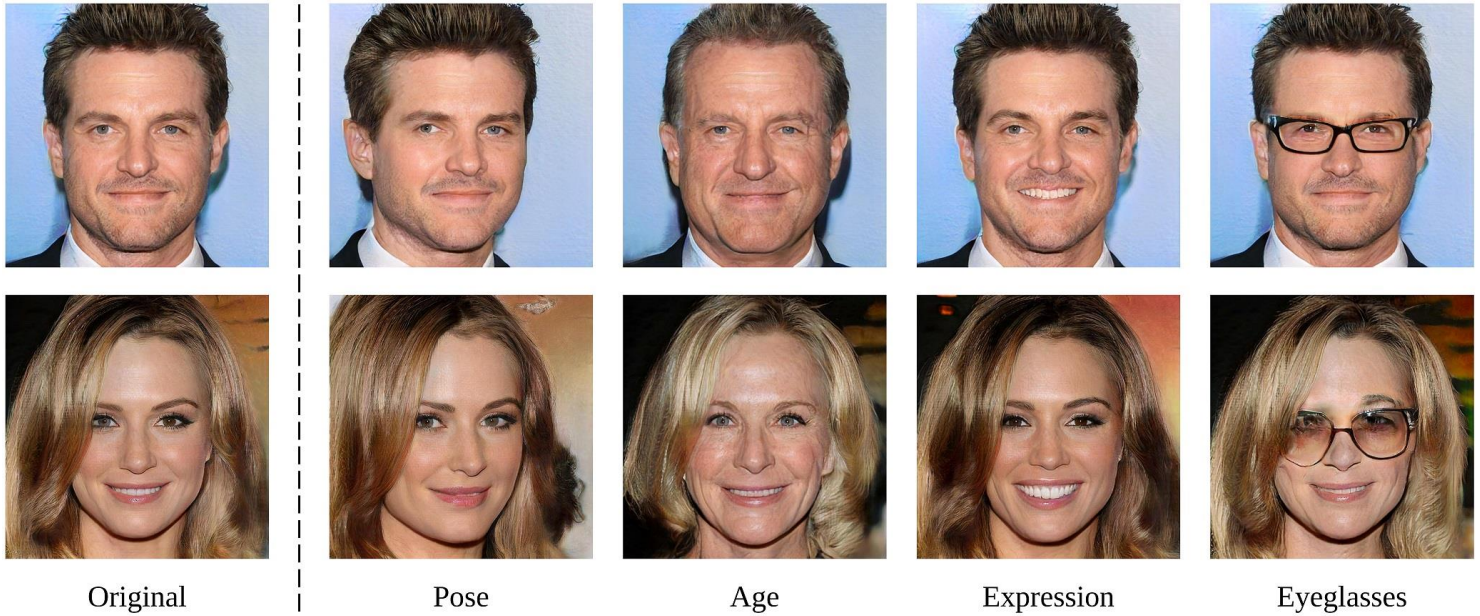
Generative Adversarial Networks (GAN)

StyleGAN

Img src: https://github.com/NVlabs/stylegan

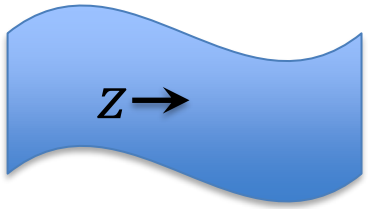
Generative Adversarial Networks (GAN)

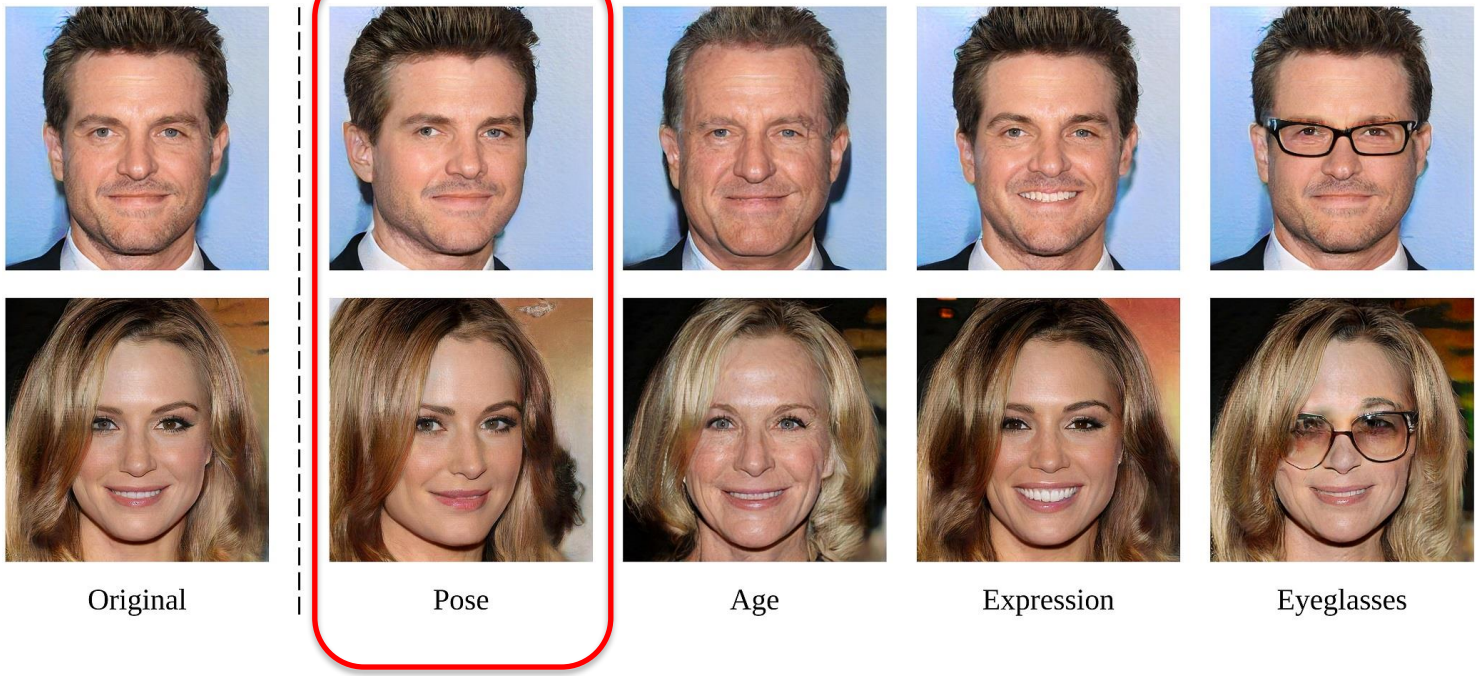




Img src: https://github.com/NVlabs/stylegan, https://genforce.github.io/interfacegan/

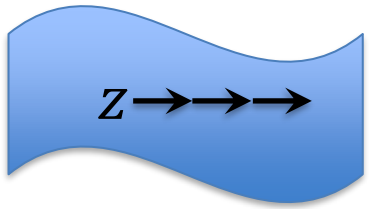
Generative Adversarial Networks (GAN)





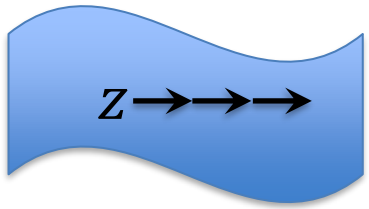
Img src: https://github.com/NVlabs/stylegan, https://genforce.github.io/interfacegan/

Generative Adversarial Networks (GAN)



Img src: https://github.com/NVlabs/stylegan, https://genforce.github.io/interfacegan/

Generative Adversarial Networks (GAN)



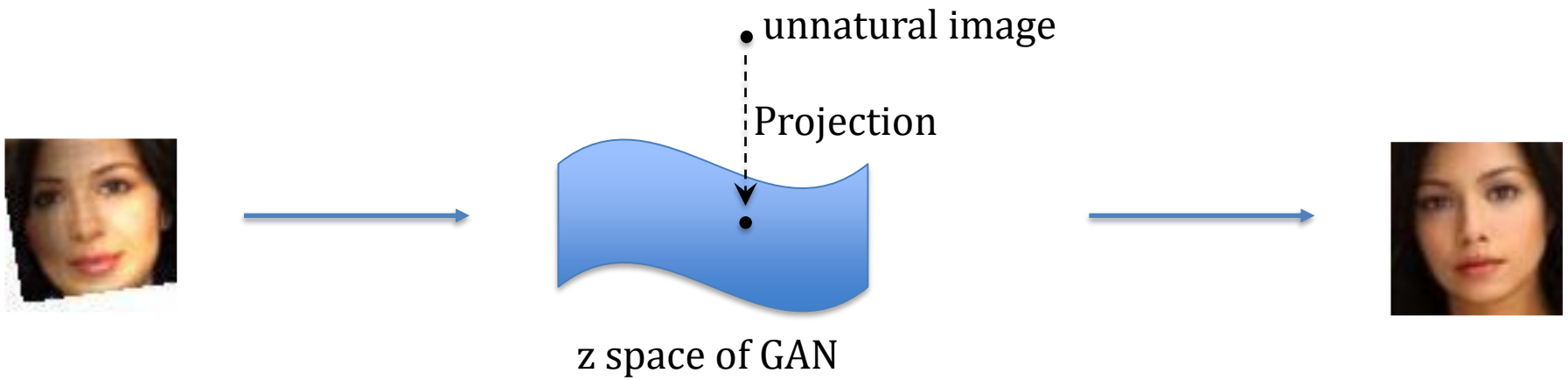
GANs are encouraged to learn 3D semantics from 2D images

Img src: https://github.com/NVlabs/stylegan, https://genforce.github.io/interfacegan/

Do 2D GANs Know 3D Shape?
Unsupervised 3D shape reconstruction from 2D Image GANs

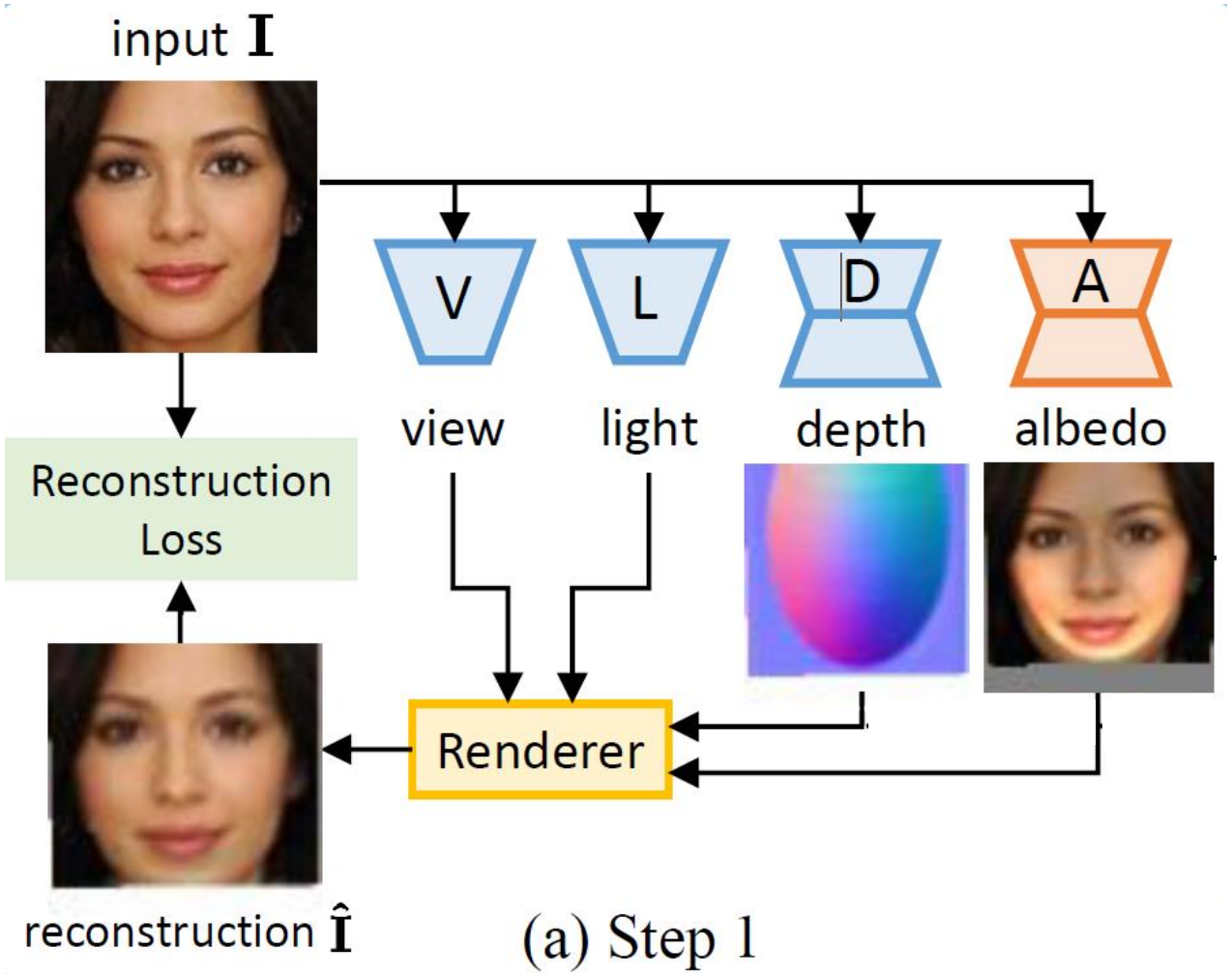
GAN2Shape

Key idea: if reconstructed shapes are incorrect, rendering new images from them will lead to unnatural results. We can use GANs to **correct** these unnatural results, which can be used as supervision.



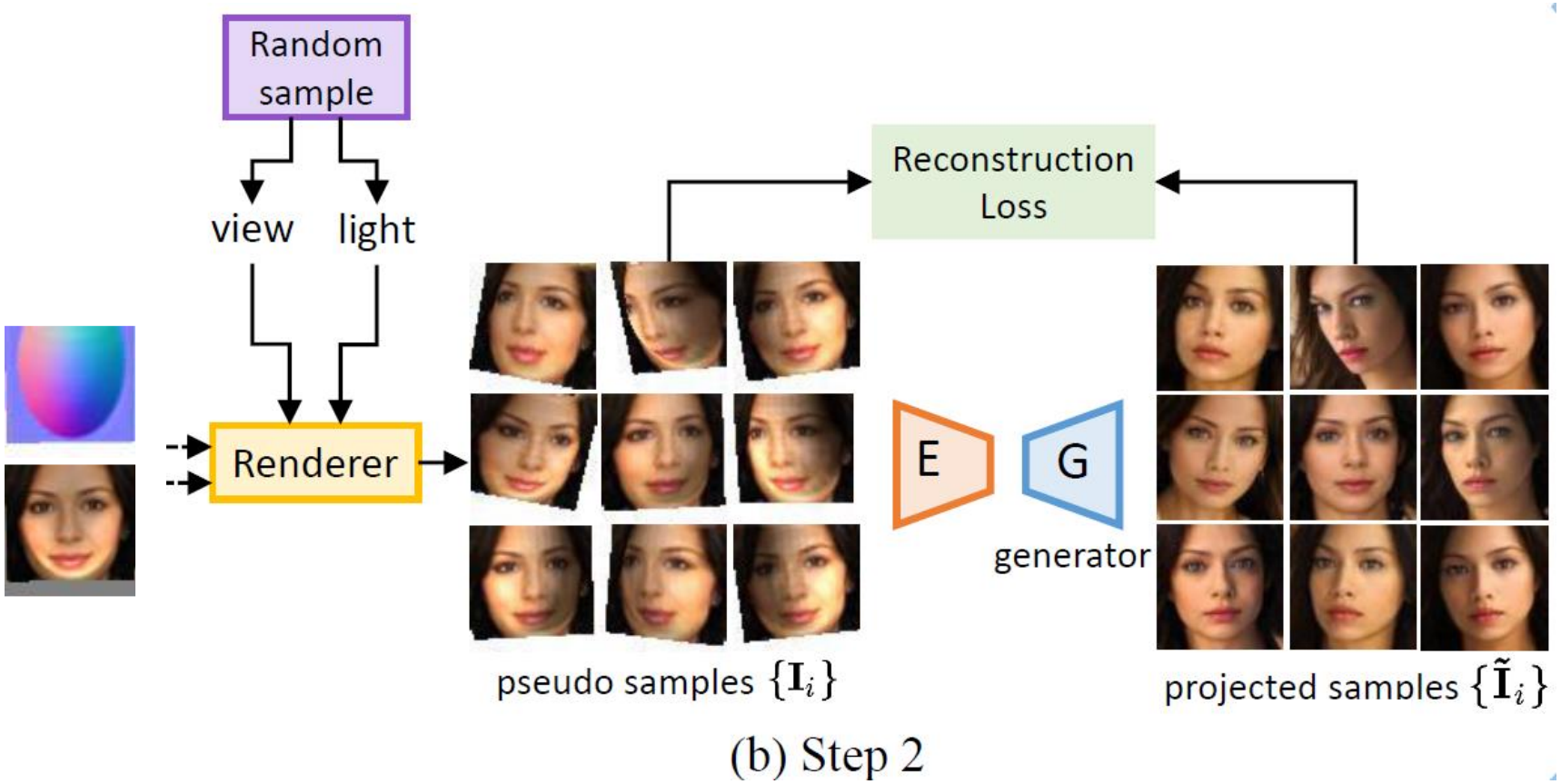
Img src: Pan Xingang et.al. ICLR2021

GAN2Shape Pipeline



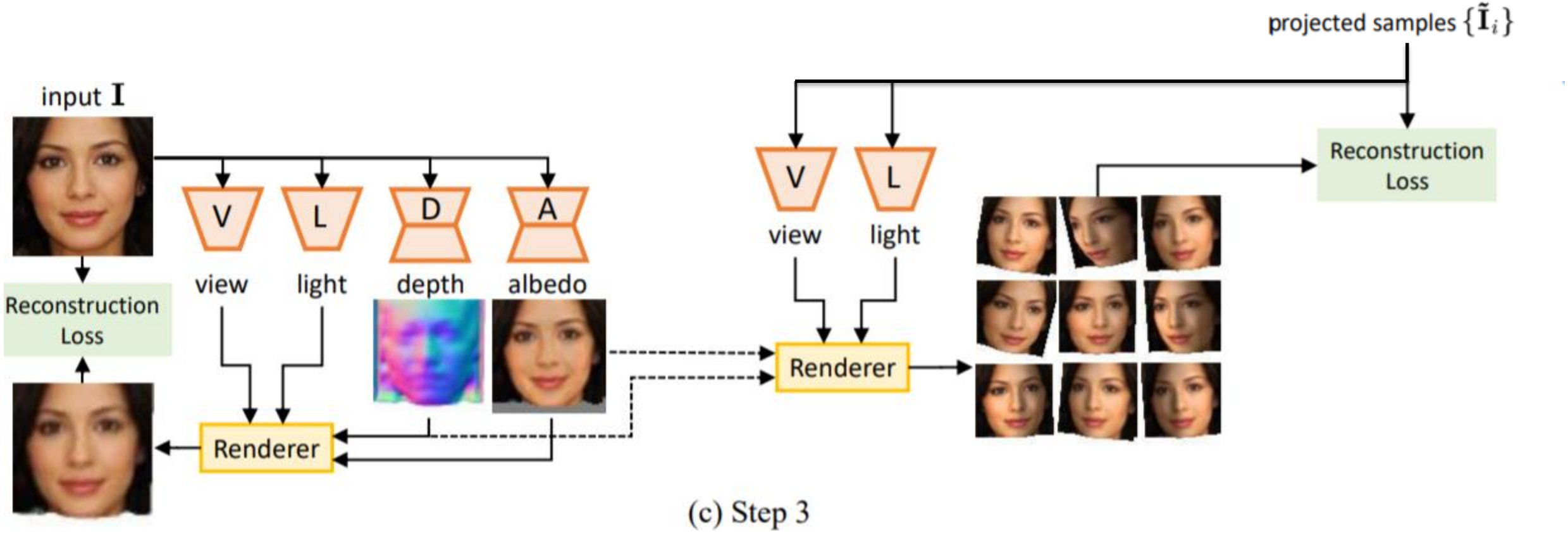
Img src: Pan Xingang et.al. ICLR2021

GAN2Shape Pipeline



Img src: Pan Xingang et.al. ICLR2021

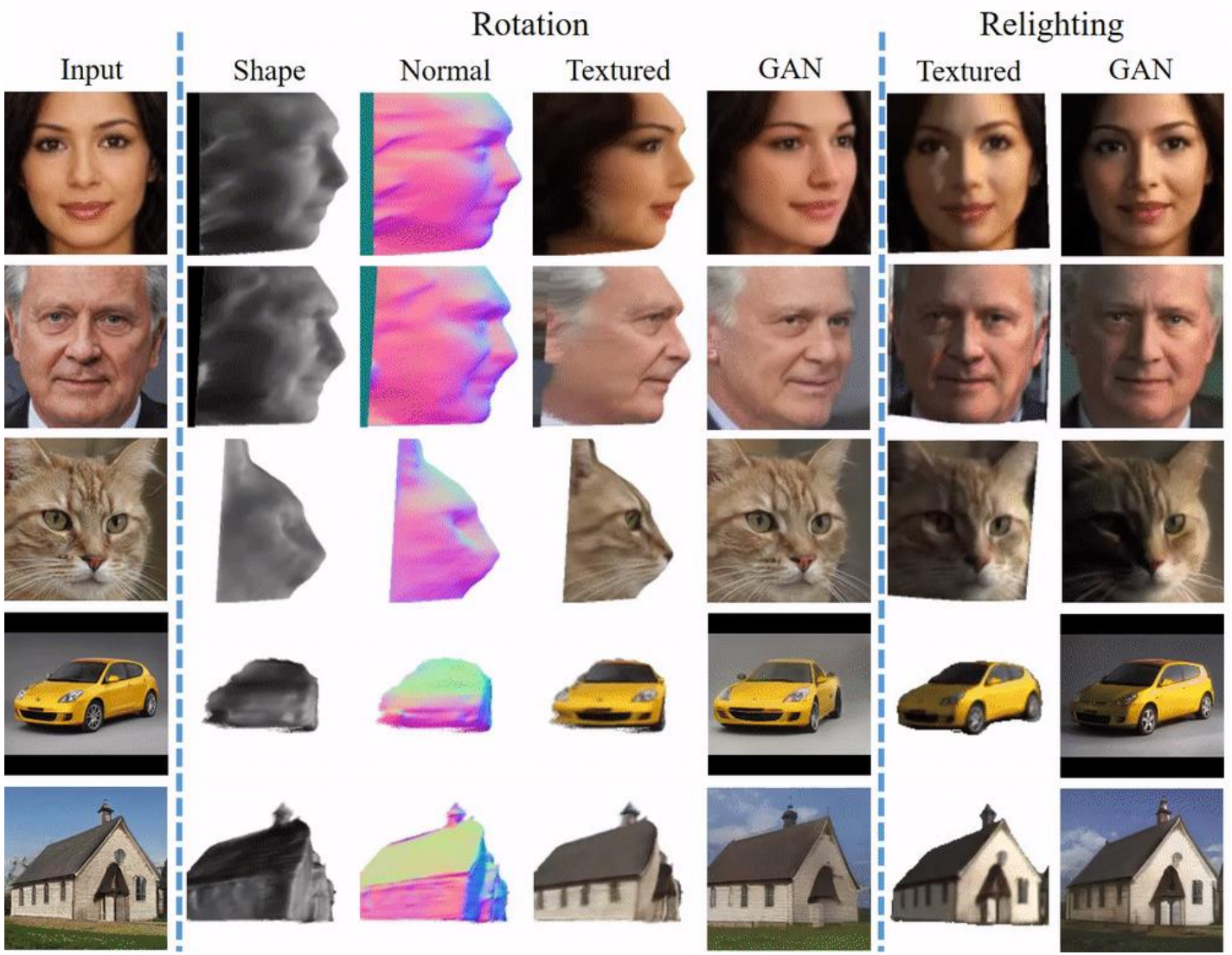
GAN2Shape Pipeline



Step 1 → Step 2 → Step 3 → Step 1 → …

Img src: Pan Xingang et.al. ICLR2021

GAN2Shape Results



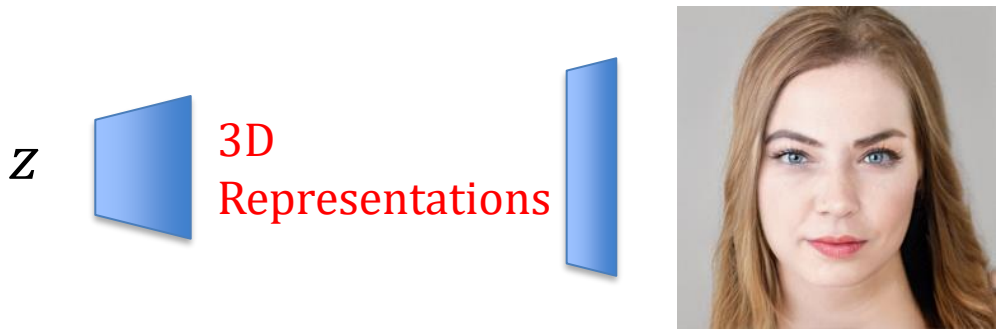
Img src: Pan Xingang et.al. ICLR2021

2D GANs indeed know 3D space to some extent.
Why stop at 2D GANs?

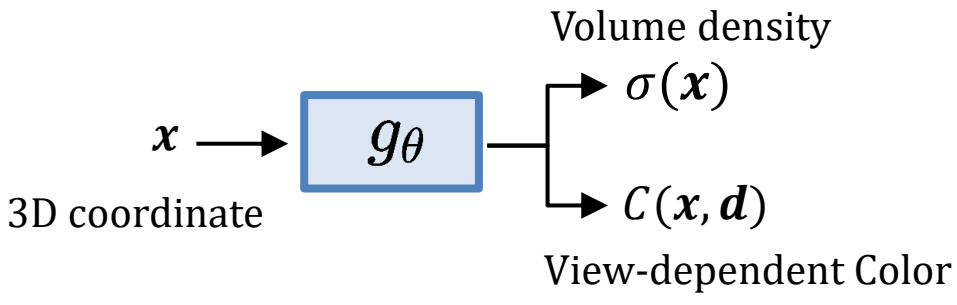
3D-aware Image GAN

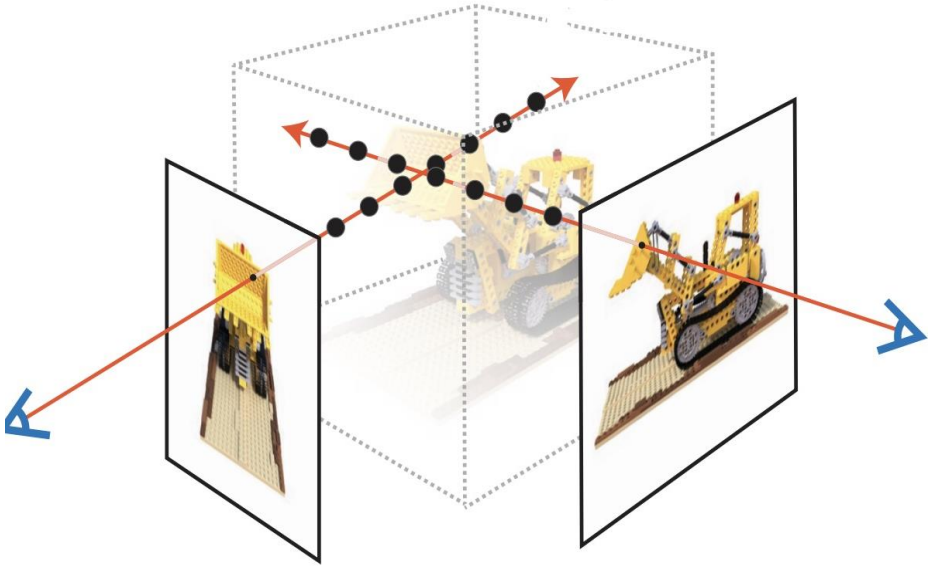
2D GAN:

3D-aware GAN:



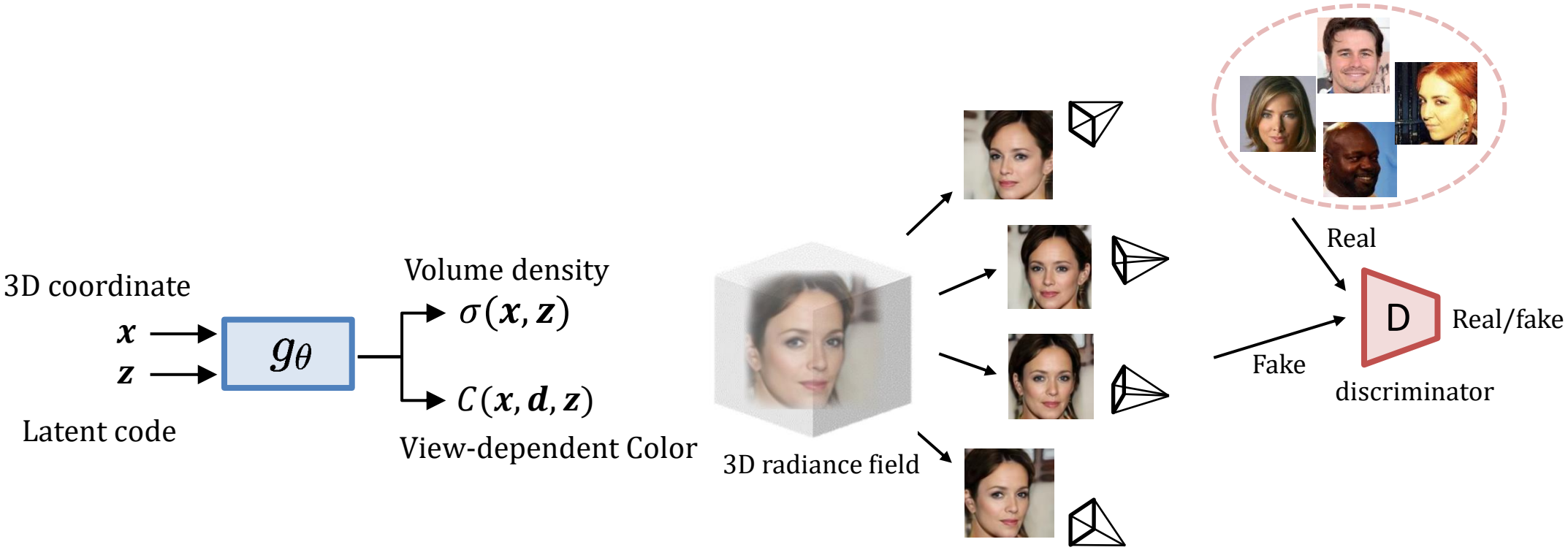
3D Representations: NeRF





Img src: Ben Mildenhall et.al. ECCV2020

NeRF-GAN: NeRF as intermediate representation



- Pi-GAN: Periodic Implicit Generative Adversarial Networks for 3D-Aware Image Synthesis (CVPR 2021)
- GRAF: Generative Radiance Fields for 3D-Aware Image Synthesis (NeurIPS 2020)

NeRF-GAN: Issues

learned shapes are not smooth and accurate!

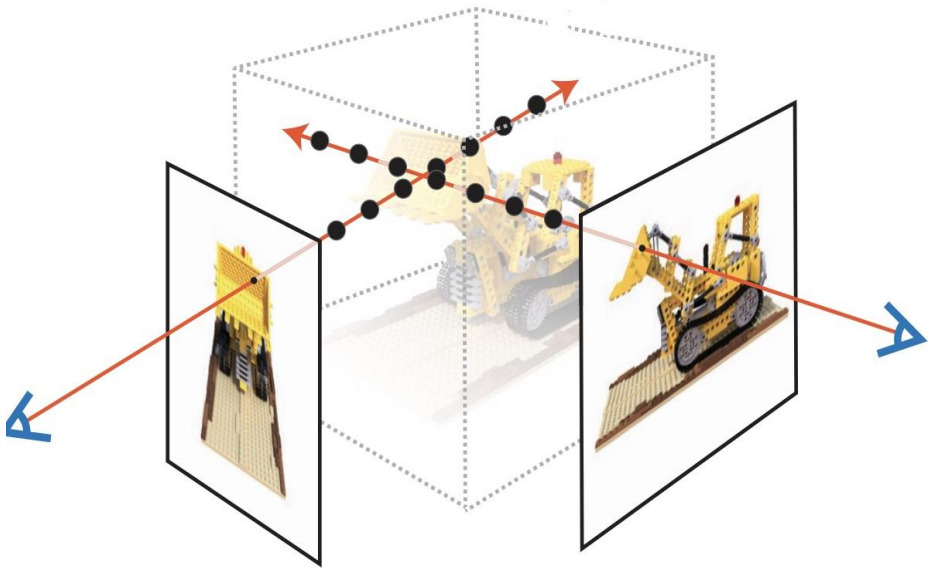
NeRF-GAN: Issues

learned shapes are not smooth and accurate!

- NeRF has no concept of surface, volume densities are diffused in the space.

Generative Occupancy Fields for 3D Surface-Aware Image Synthesis

GOF



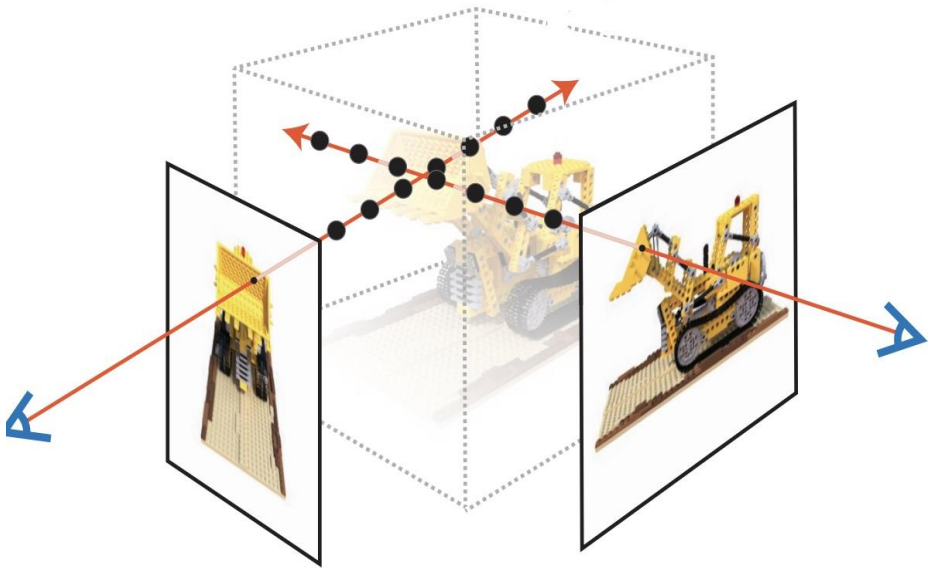
Volume Rendering Equation in NeRF-GAN:

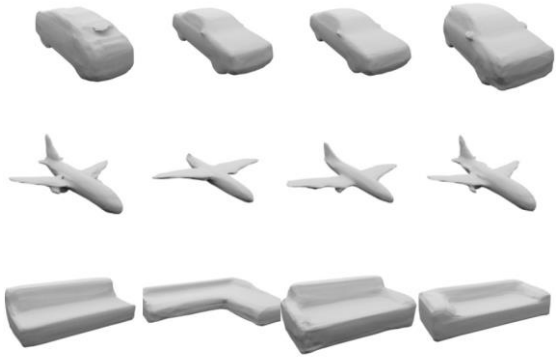
$$\hat{\mathbf{C}}(\mathbf{r},\mathbf{z}) = \sum_{i=1}^{N} T_i \Big(1 - \exp\left(-\sigma_\theta(\mathbf{x}_i,\mathbf{z})\delta_i\right)\Big)\mathbf{c}_\theta(\mathbf{x}_i,\mathbf{d},\mathbf{z})$$

Intermediate alpha values

$$\alpha_\theta(\mathbf{x}_i,\mathbf{z}) = 1 - \exp(-\sigma_\theta(\mathbf{x}_i,\mathbf{z})\delta_i) \in [0,1]$$

GOF





$$f_\theta : \mathbb{R}^3 \times \mathcal{Z} \rightarrow [0, 1]$$

Volume Rendering Equation in NeRF-GAN:

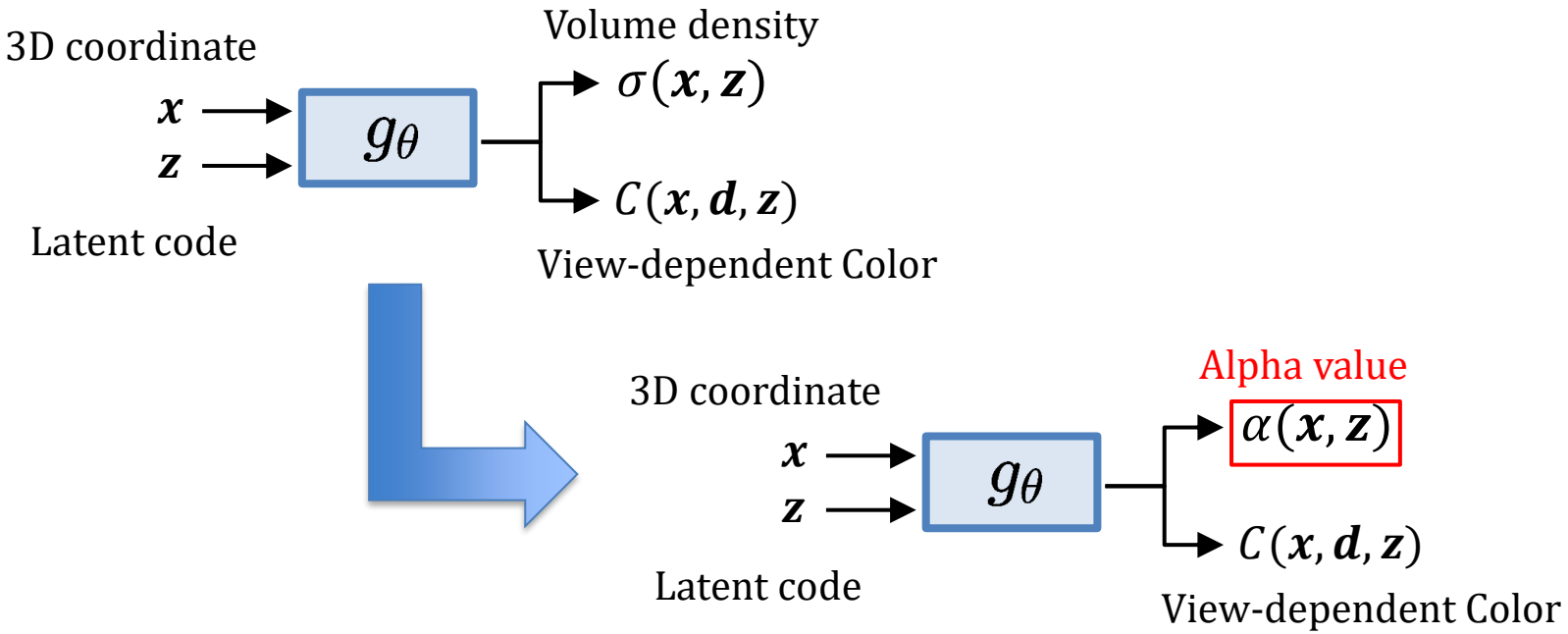
$$\hat{\mathbf{C}}(\mathbf{r}, \mathbf{z}) = \sum_{i=1}^{N} T_i \Big(1 - \exp\left(-\sigma_\theta(\mathbf{x}_i, \mathbf{z})\delta_i\right)\Big)\mathbf{c}_\theta(\mathbf{x}_i, \mathbf{d}, \mathbf{z})$$

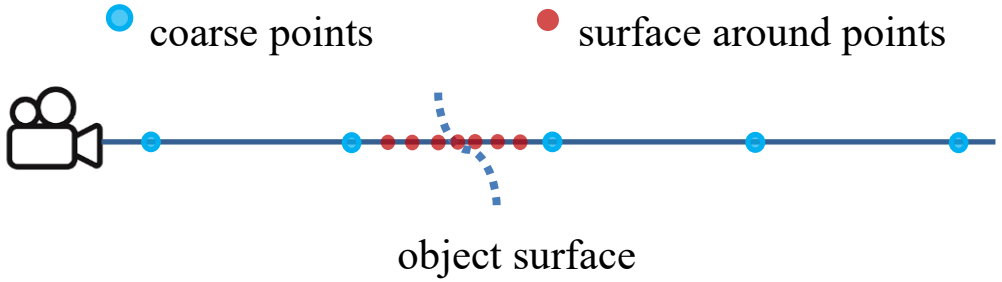
Intermediate alpha values

$$\alpha_\theta(\mathbf{x}_i, \mathbf{z}) = 1 - \exp(-\sigma_\theta(\mathbf{x}_i, \mathbf{z})\delta_i) \in [0, 1]$$

Resemble the **occupancy representation** that ensures good object surfaces inherently.

GOF

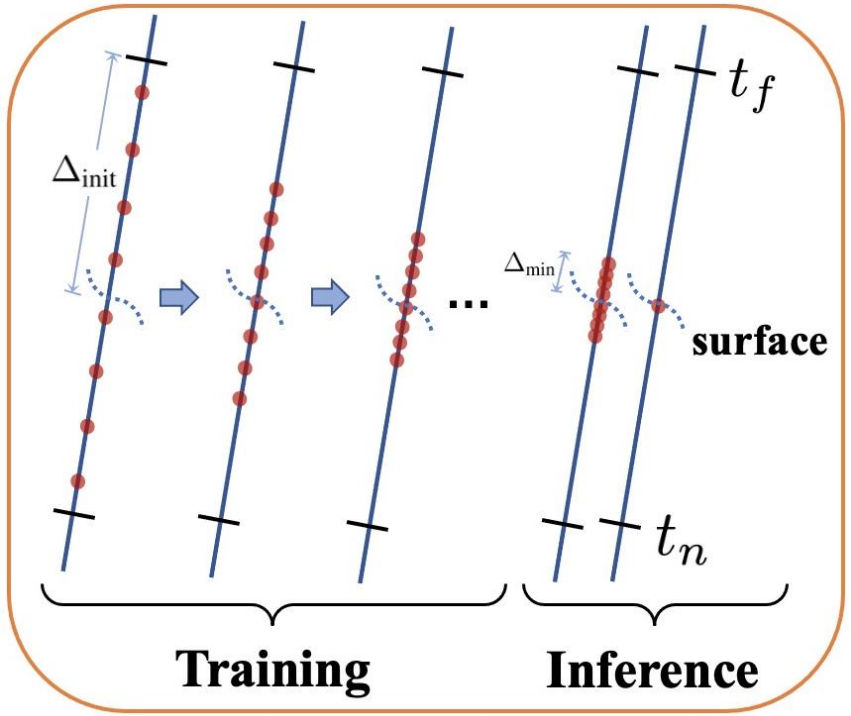




Root-finding:

$$k^{\mathcal{S}} = \underset{k}{\operatorname{argmin}} \left(\alpha_\theta(\mathbf{x}_k, \mathbf{z}) < \tau \leq \alpha_\theta(\mathbf{x}_{k+1}, \mathbf{z})\right)$$

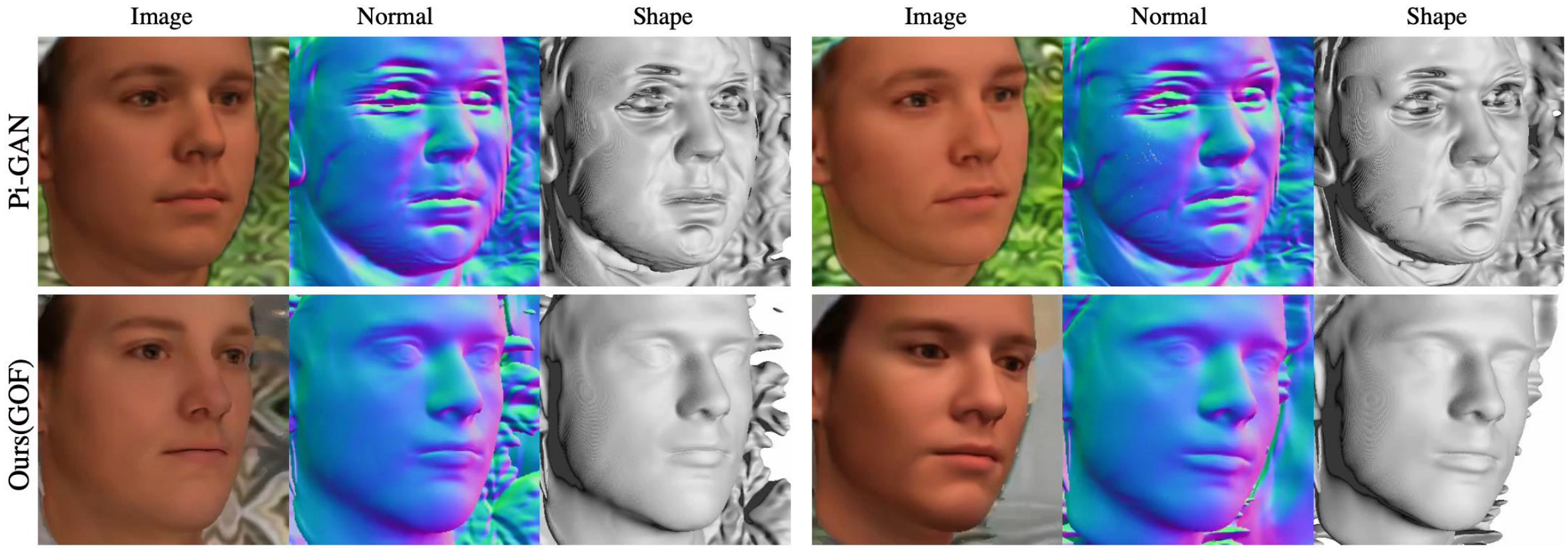
GOF: shrink sampling region



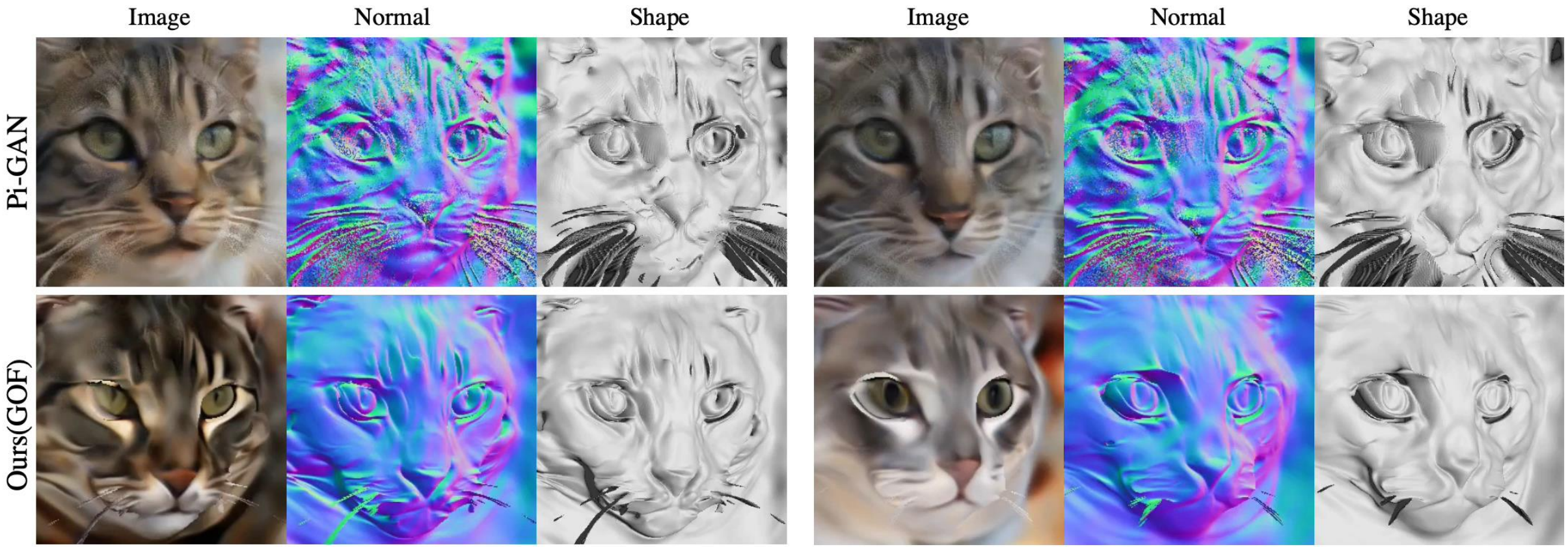
Shrinking sampling region during training:

$$\Delta_n = \max(\Delta_{\text{init}} \exp(-\gamma n), \Delta_{\text{min}})$$

GOF Results



GOF Results



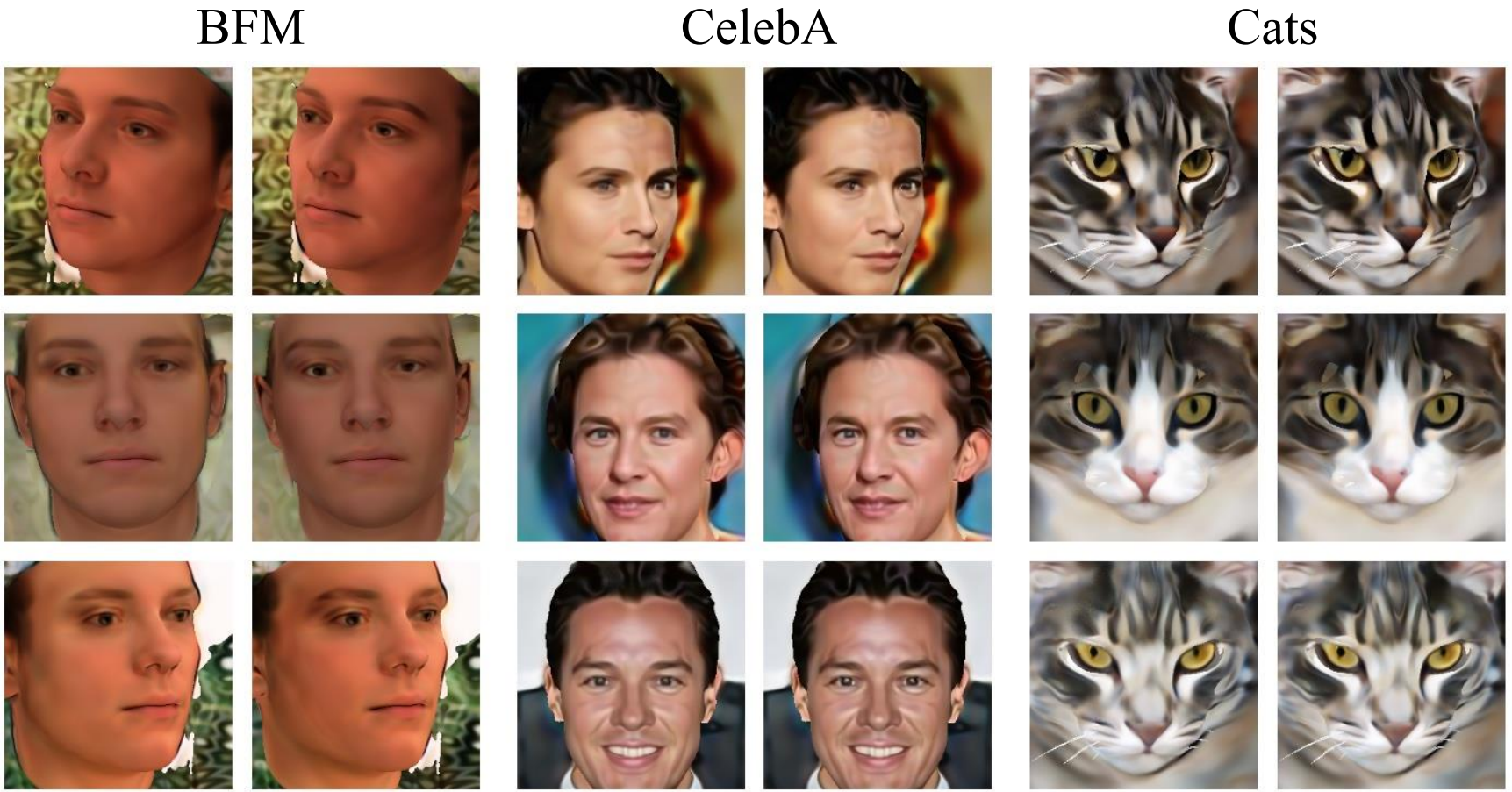
GOF Results

BFM

CelebA

Cats

GOF Results: Surface Rendering



Left side: volume rendering; Right side: surface rendering

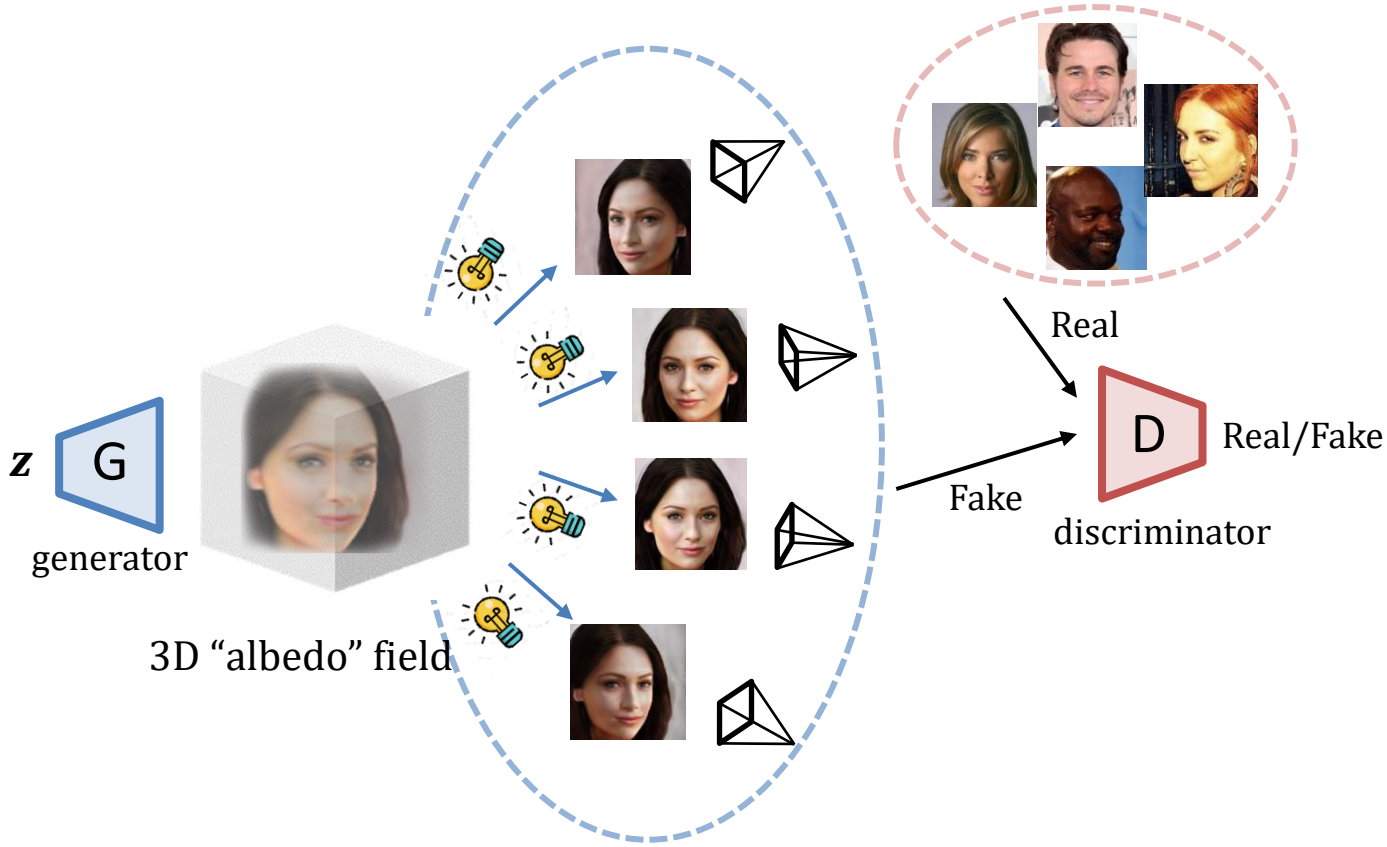
NeRF-GAN: Issues

learned shapes are not smooth and accurate!

- NeRF has no concept of surface, volume densities are diffused in the space.
- NeRF can render good images without learning accurate shapes, i.e, the shape-color ambiguity.

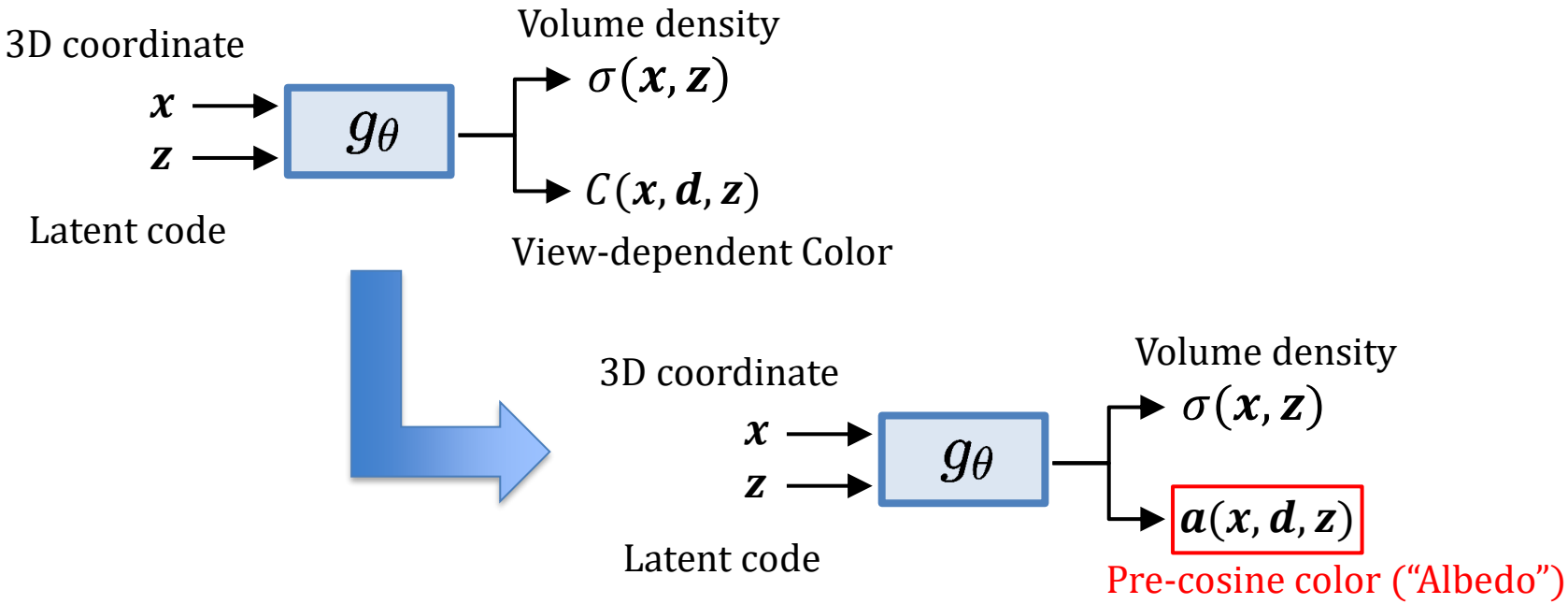
A Shading-Guided Generative Implicit Model for Shape-Accurate 3D-Aware Image Synthesis

ShadeGAN

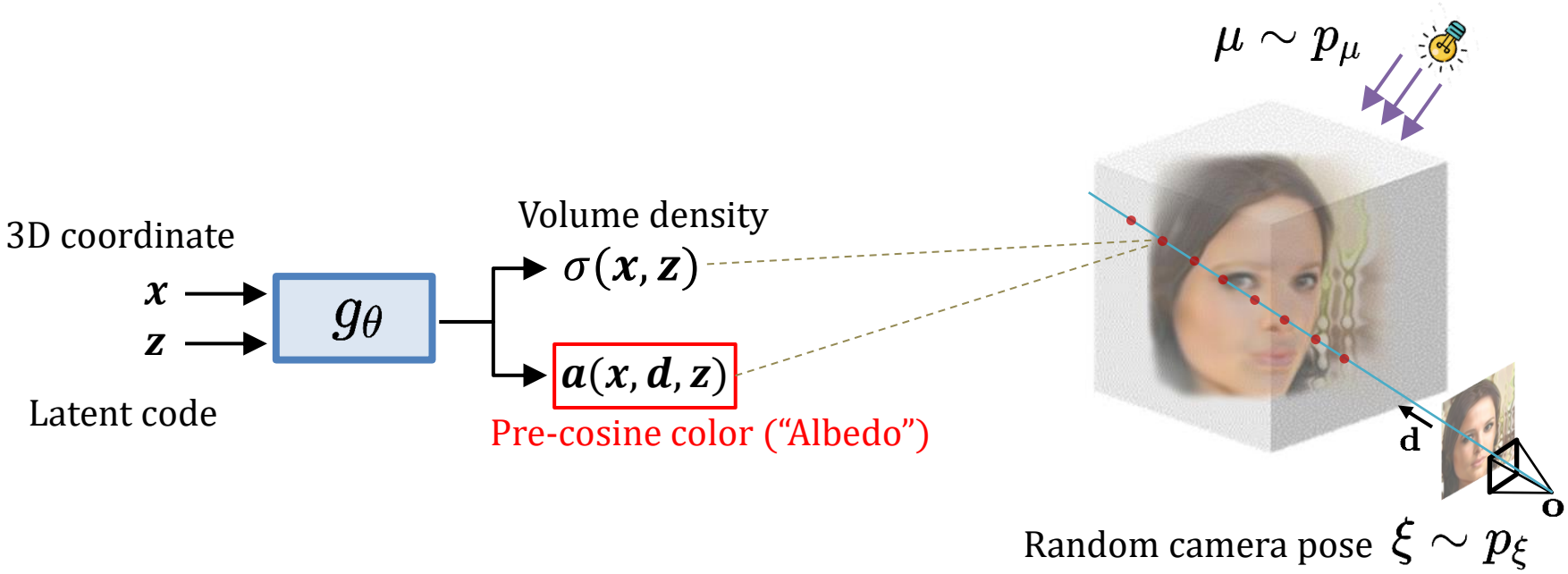


The learned shapes should look realistic not only from different **views** (**multi-view constraint**), but also from different lightings (multi-lighting constraint). Inaccurate shapes would be more clearly revealed, as shading is sensitive to surface normals.

ShadeGAN


3D coordinate
x
z
g_θ
Latent code
Volume density
$\sigma(x, z)$
$C(x, d, z)$
View-dependent Color
3D coordinate
x
z
g_θ
Latent code
Volume density
$\sigma(x, z)$
$a(x, d, z)$
Pre-cosine color ("Albedo")

ShadeGAN



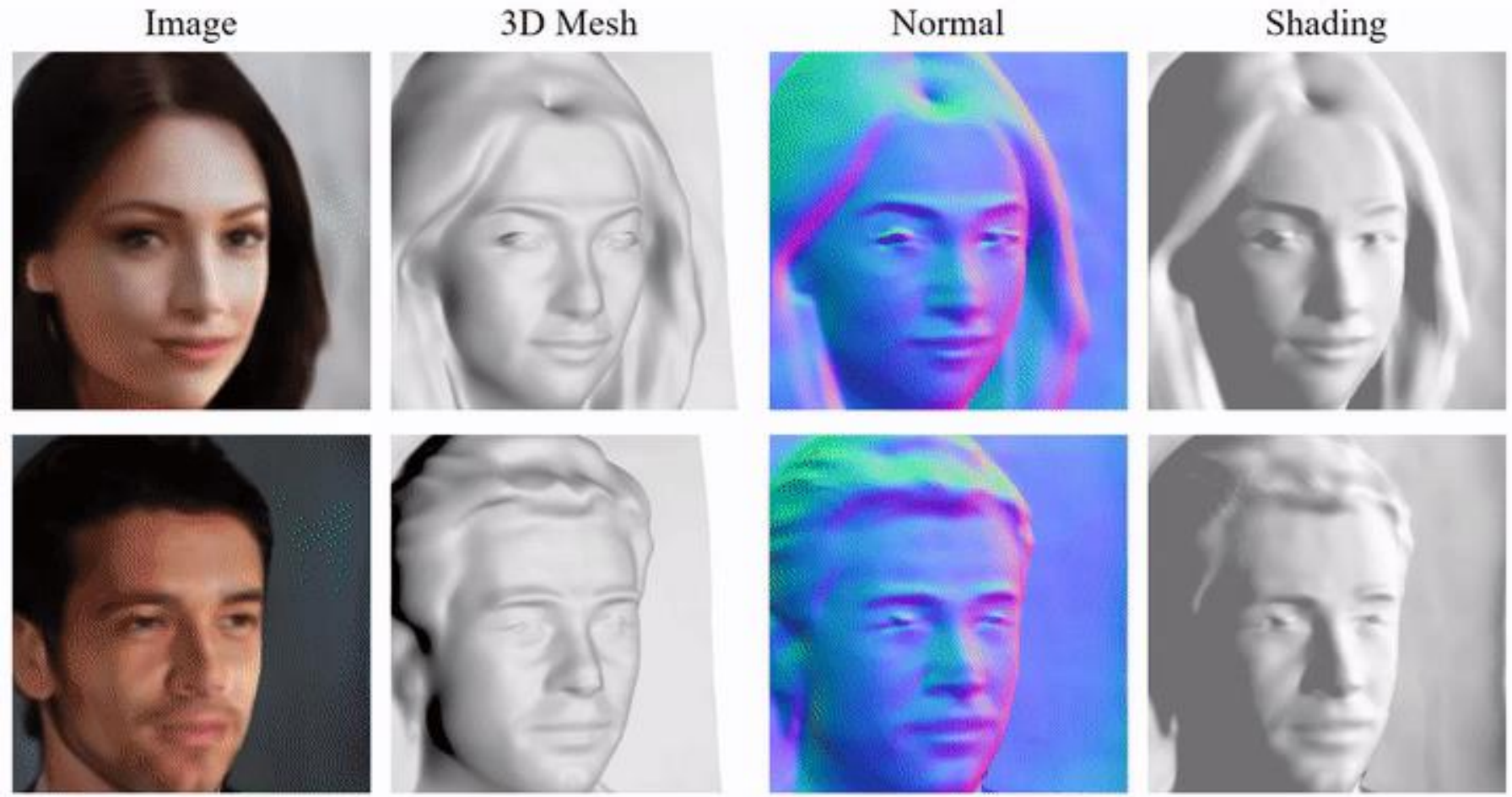
Calculate albedo: $$\boldsymbol{A}(\boldsymbol{r}, \boldsymbol{z}) = \int_{t_n}^{t_f} T(t, \boldsymbol{z})\sigma(\boldsymbol{r}(t), \boldsymbol{z})\boldsymbol{a}(\boldsymbol{r}(t), \boldsymbol{d}, \boldsymbol{\mu}, \boldsymbol{z})dt, \text{ where } T(t, \boldsymbol{z}) = \exp(-\int_{t_n}^{t} \sigma(\boldsymbol{r}(s), \boldsymbol{z})ds)$$

Surface normal: $$\boldsymbol{n}(\boldsymbol{r}, \boldsymbol{z}) = \hat{\boldsymbol{n}}(\boldsymbol{r}, \boldsymbol{z})/\|\hat{\boldsymbol{n}}(\boldsymbol{r}, \boldsymbol{z})\|_2, \text{ where } \hat{\boldsymbol{n}}(\boldsymbol{r}, \boldsymbol{z}) = -\int_{t_n}^{t_f} T(t, \boldsymbol{z})\sigma(\boldsymbol{r}(t), \boldsymbol{z})\nabla_{\boldsymbol{r}(t)}\sigma(\boldsymbol{r}(t), \boldsymbol{z})dt$$

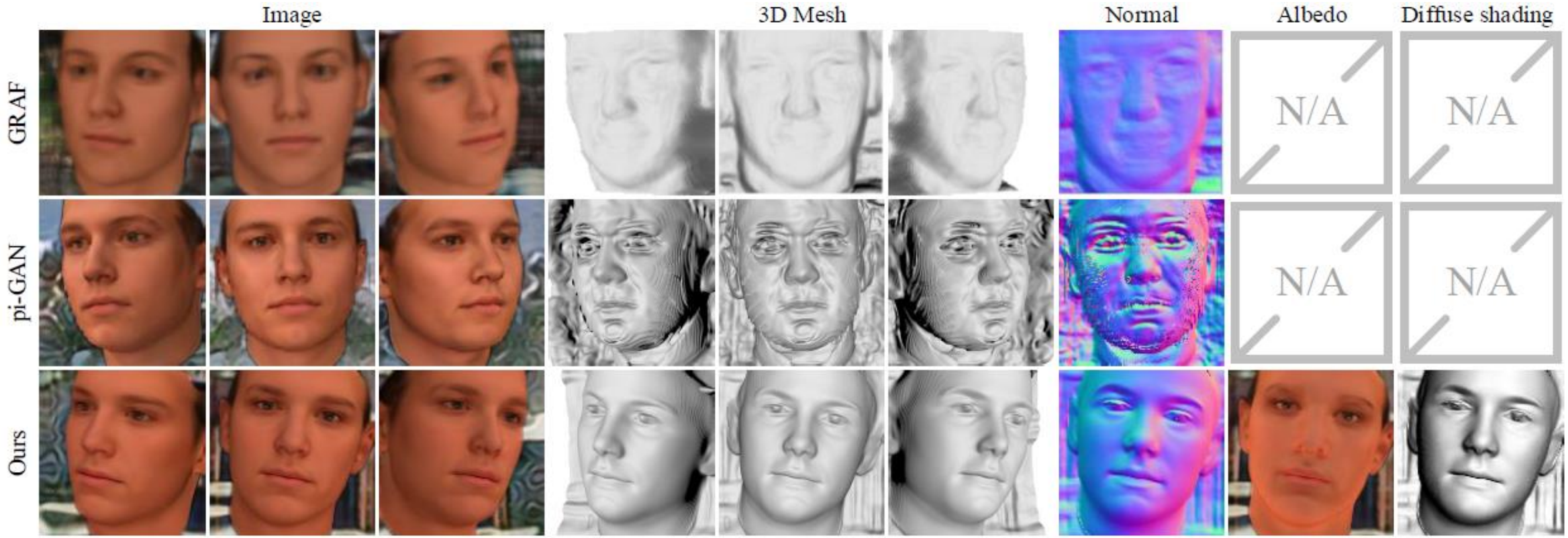
Lambertian shading: $$\boldsymbol{C}(\boldsymbol{r}, \boldsymbol{z}) = \boldsymbol{A}(\boldsymbol{r}, \boldsymbol{z})(k_a + k_d \max(0, \boldsymbol{l} \cdot \boldsymbol{n}(\boldsymbol{r}, \boldsymbol{z}))) \qquad \boldsymbol{\mu} = (\boldsymbol{l}, k_a, k_d) \sim p_\mu$$

Ambient diffuse light direction

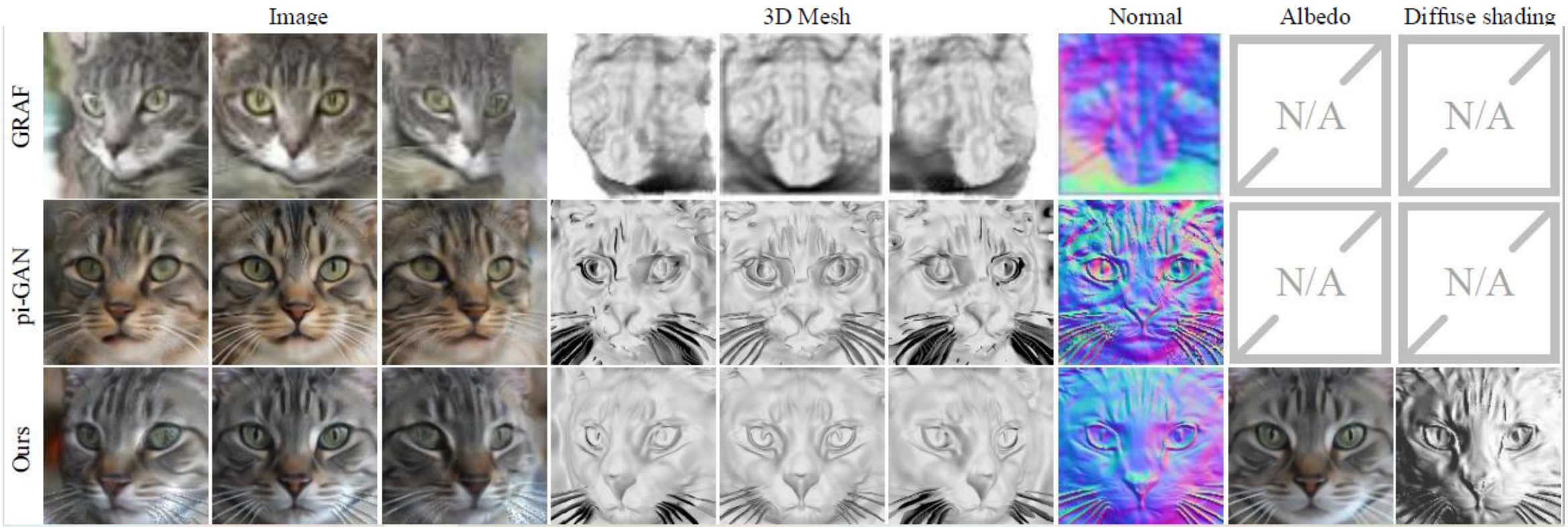
ShadeGAN Results



ShadeGAN Results



ShadeGAN Results



ShadeGAN Results: Real Image Editing

Real Image

Reconstruction

3D Mesh

Normal

Albedo

Diffuse Shading

View Synthesis

Relighting

What's next?

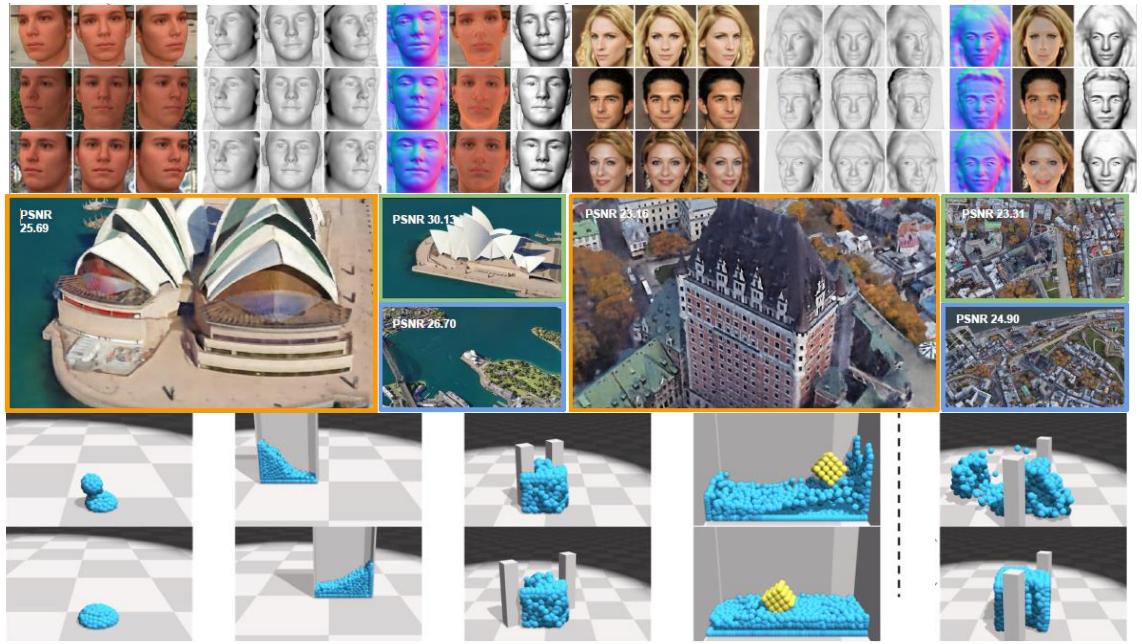
Future Work

- Shapes with more fine-grained details.
- Shapes of more complex categories.
- Improved learning efficiency and robustness.
- Application of these GANs in downstream tasks.
- ...

Advertisement

上海人工智能实验室

Shanghai Artificial Intelligence Laboratory



智能数字人研究中心

I'm looking for full-time researchers and part-time interns in the related fields of Digital Human and Metaverse!

Thank you!

GAN2Shape

GOF

ShadeGAN