Learning Neural Parametric Head Models with 2D Adversarial Objectives

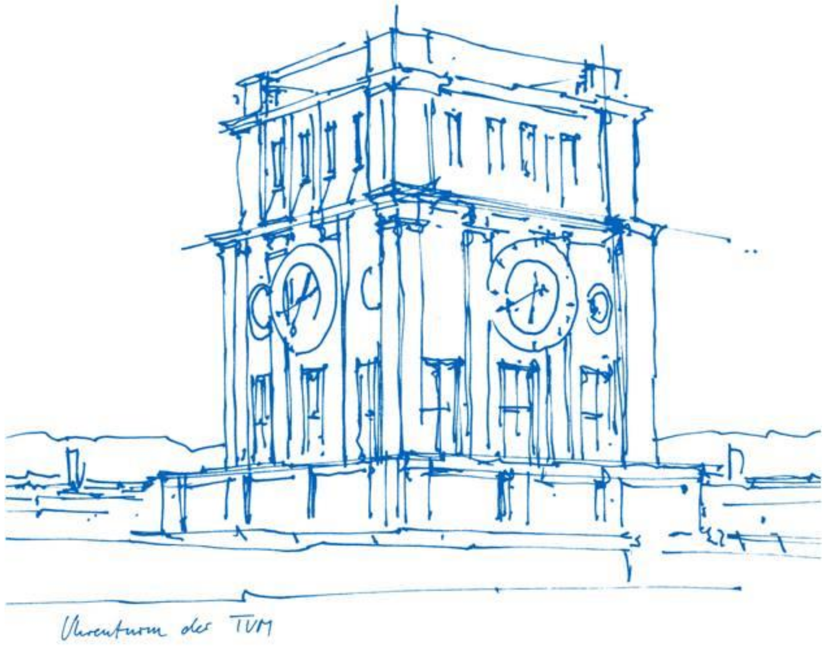
Tathagata Bandyopadhyay

Technische Universität München

Department of Informatics

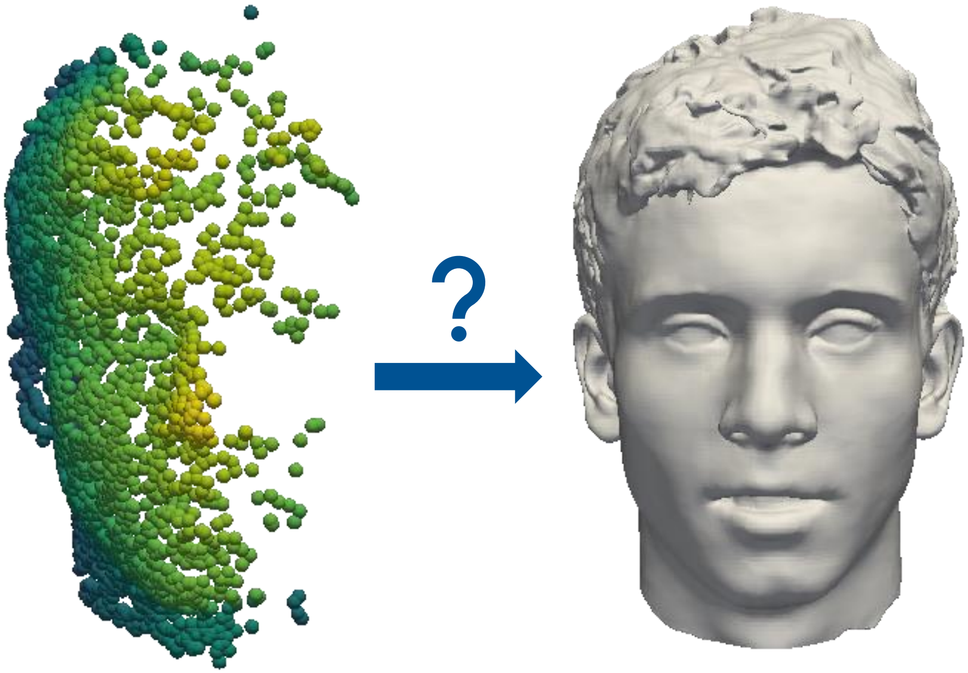
Visual Computing Lab

Munich, 22 March 2024



Introduction

The 3D Head Reconstruction Problem



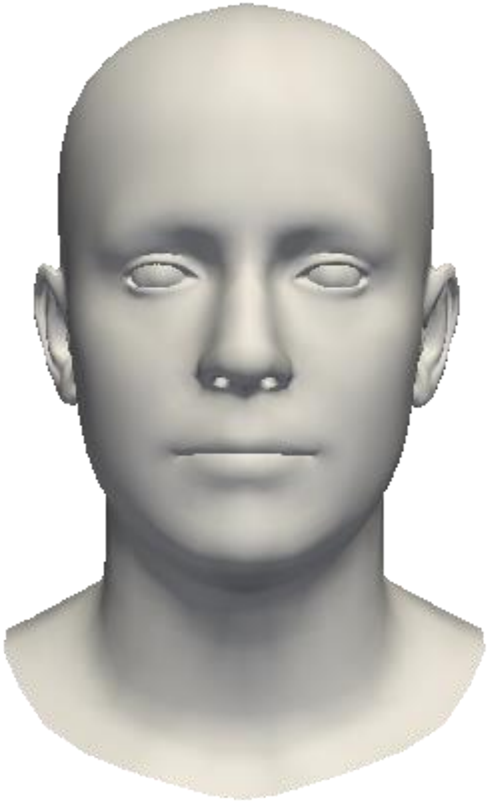
Aditional requirements:

- Generalizable identity
- Controllable expression

Free-form reconstruction:

- Highly under-constrained
- Difficult to parameterize
- Noisy point cloud

Leveraging Common Head Structure

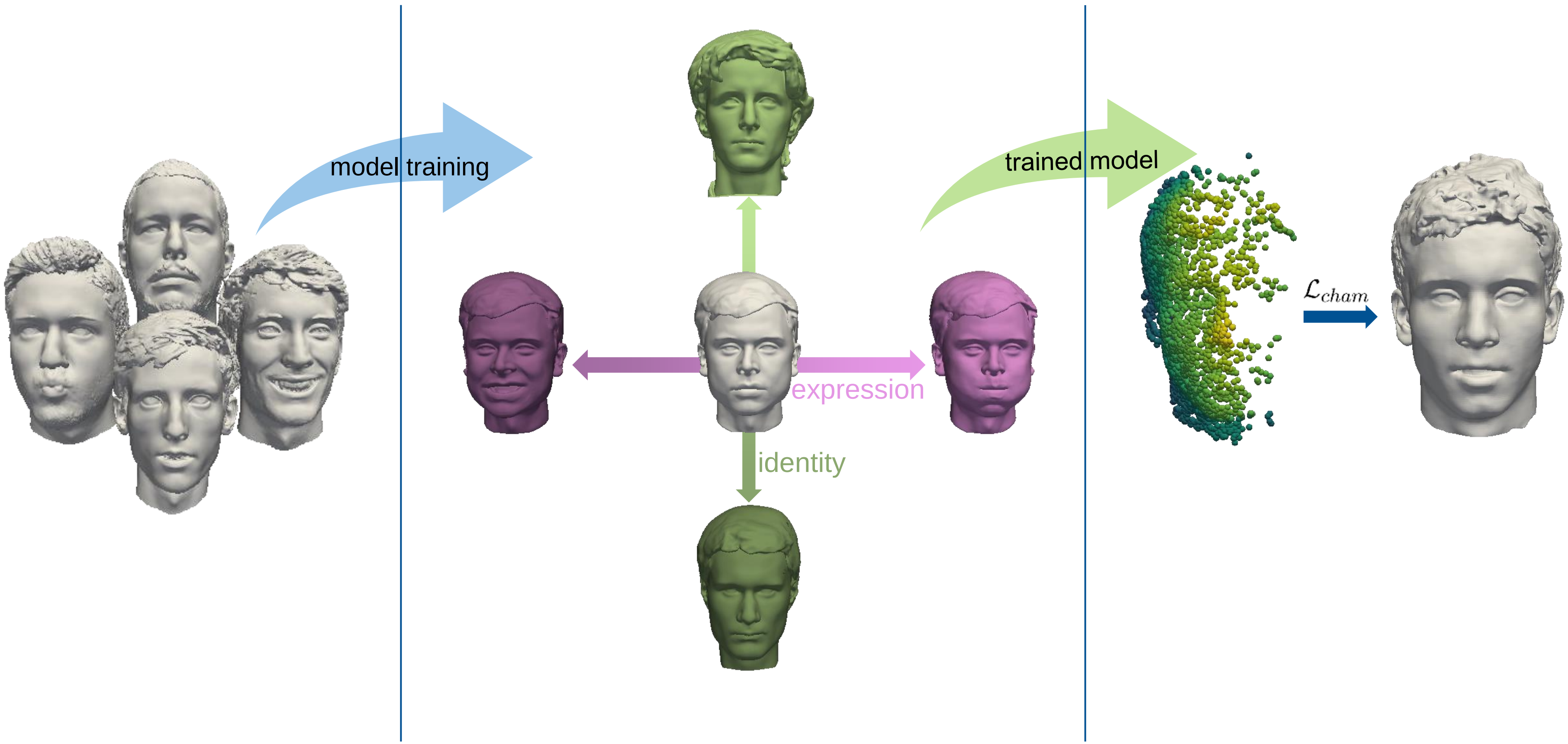


Human Head / Face:

- Common underlying geometry
- Identity / expression specific deformation
- Utilize this constraint as strong regularizer

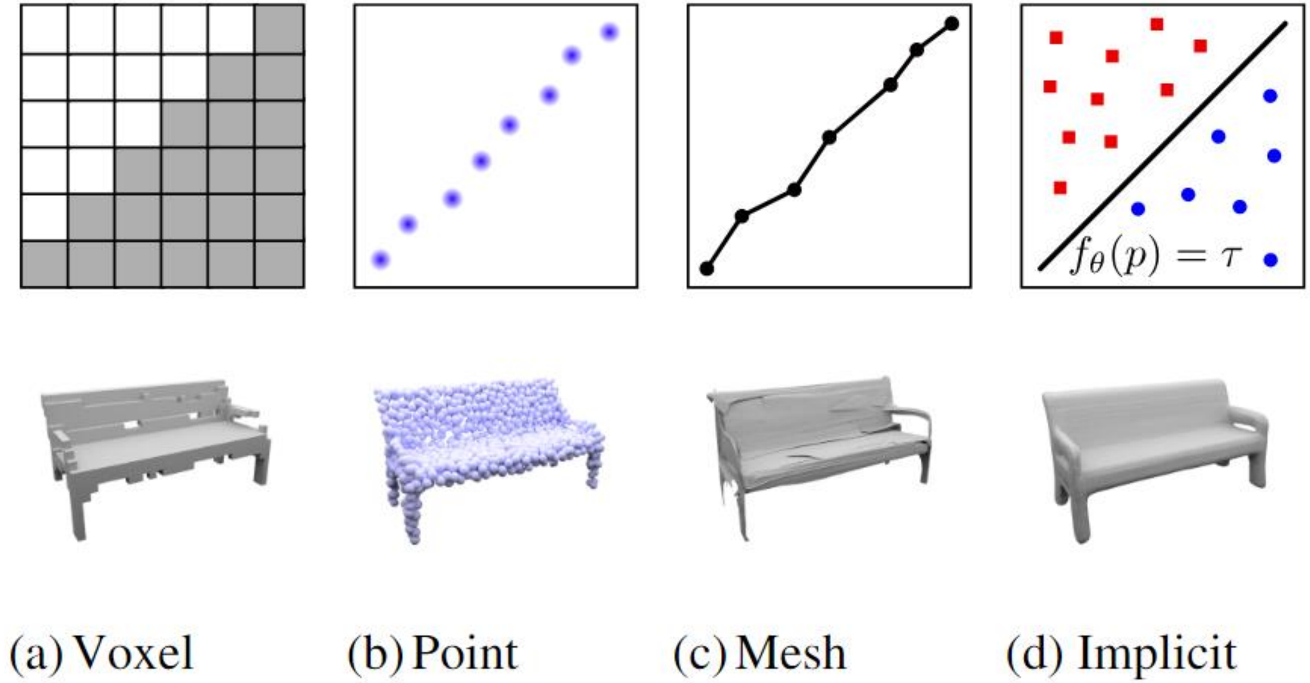
Parametric Head Model

3D Head Scans | Learnt Representation | Inference / Model-fitting



Related Work

Related Work: 3D Representation



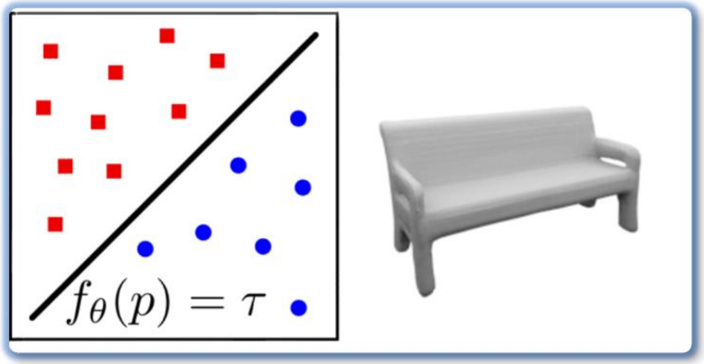
(a) Voxel
+ Simple
+ High volumetric details
- High memory usage

(b) Point
+ Direct from 3D scan
+ Memory efficient
- No surface / topology

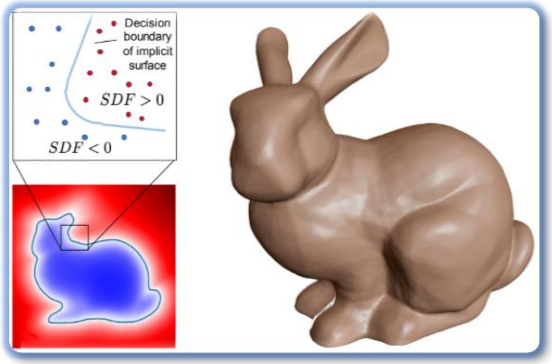
(c) Mesh
+ Graphics ready
+ Supports textures
- Hard to modify

(d) Implicit
+ Infinite resolution
+ Smooth surfaces
- Hard to define

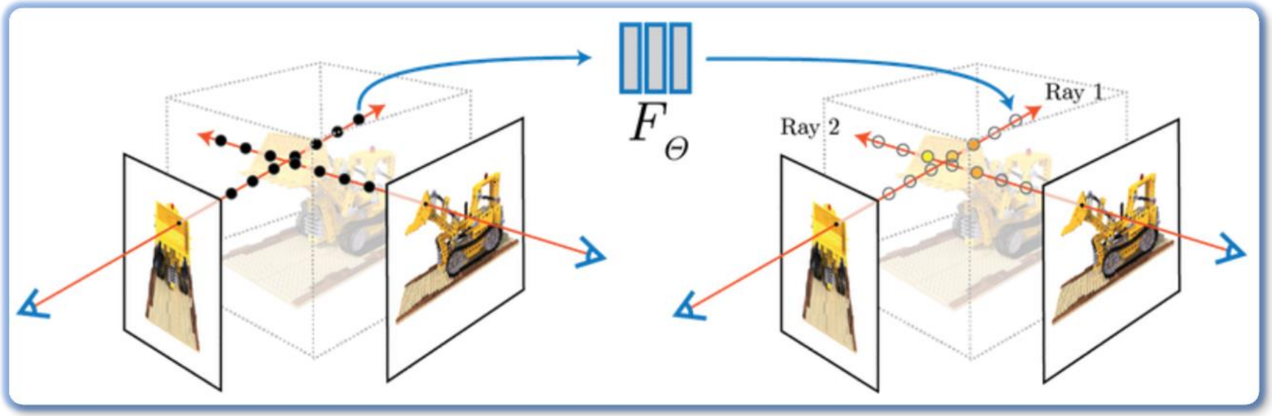
Related Work: Implicit 3D Representation



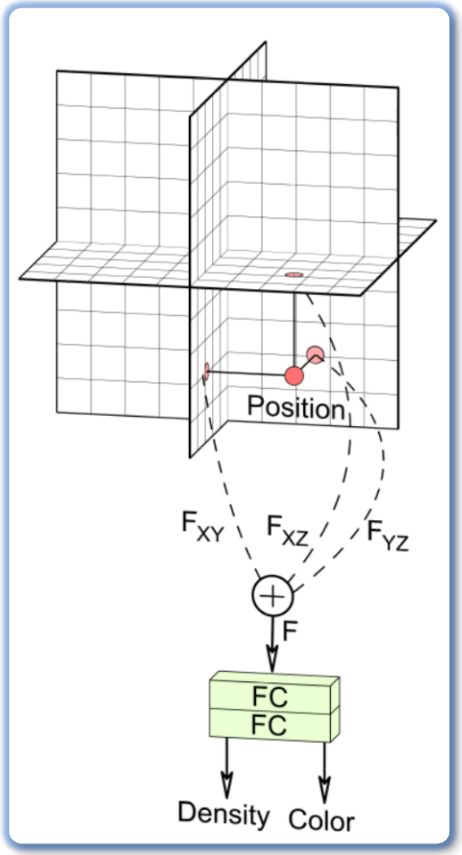
Occupancy Networks



DeepSDF



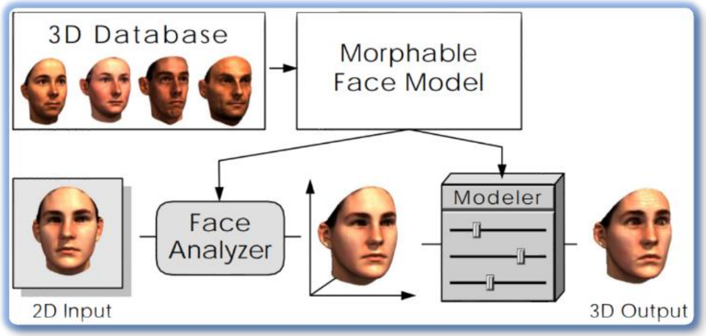
NeRF



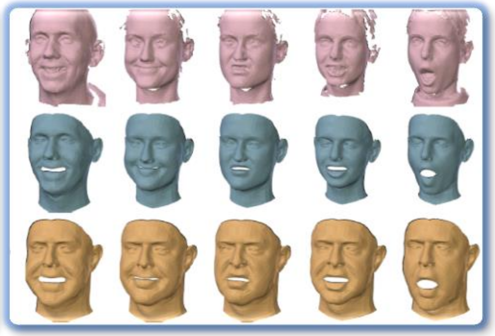
EG3D (tri-plane)

Related Work: Parametric Head / Face Models

PCA / Linear

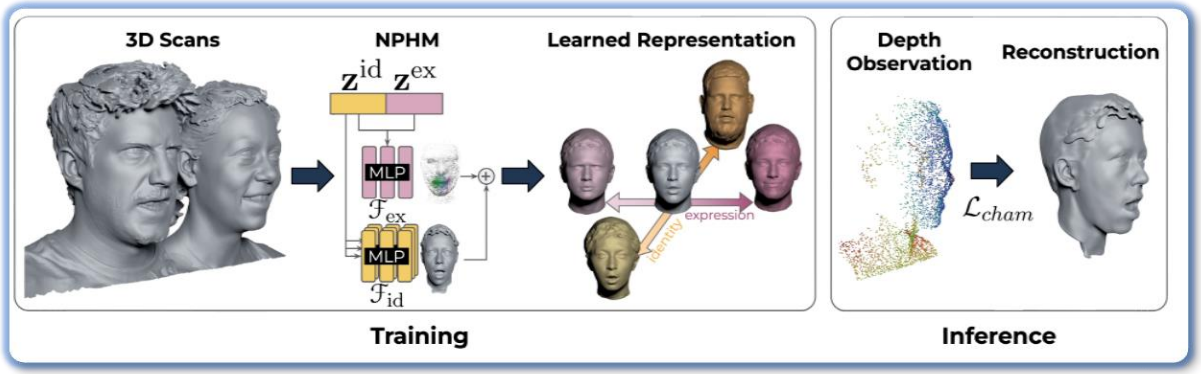


Blanz & Vetter 1999


FLAME

+ Simple
+ Highly regularized
- Fixed topology
- Less expressive

Neural / Non-linear

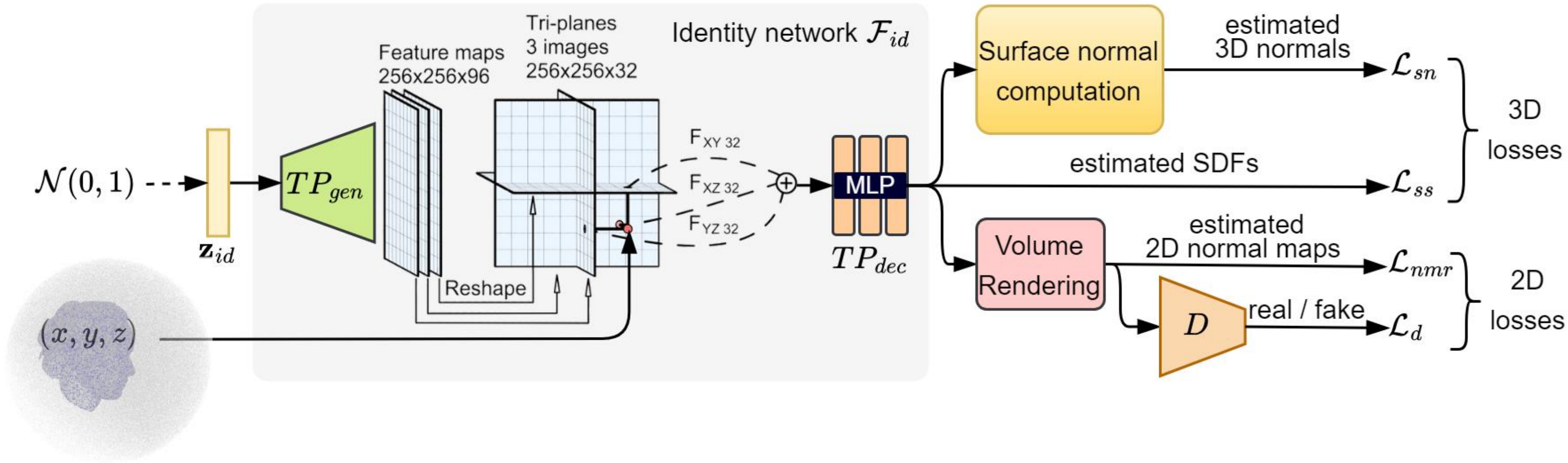


NPHM

+ Highly expressive
+ More surface detail
- Less regularized
- Can overfit noise

Method

Modeling Identity



$$\mathcal{L}_{ss} := \sum_{\mathbf{x}\in\delta\mathbf{X}} |\mathcal{F}_{\mathbf{id}}(\mathbf{x}, \mathbf{z}_{id})|$$

$$\mathcal{L}_{sn} := \sum_{\mathbf{x}\in\delta\mathbf{X}} (1 - \langle \nabla \mathcal{F}_{\mathbf{id}}(\mathbf{x}, \mathbf{z}_{id}), n_{id}(\mathbf{x}) \rangle)$$

$$\mathcal{L}_{nmr} := \sum_{v=1}^{M} ||\hat{N}_{id}^{(v)} - N_{id}^{(v)}||_1$$

$\delta\mathbf{X}$: set of 3D point samples on the surface

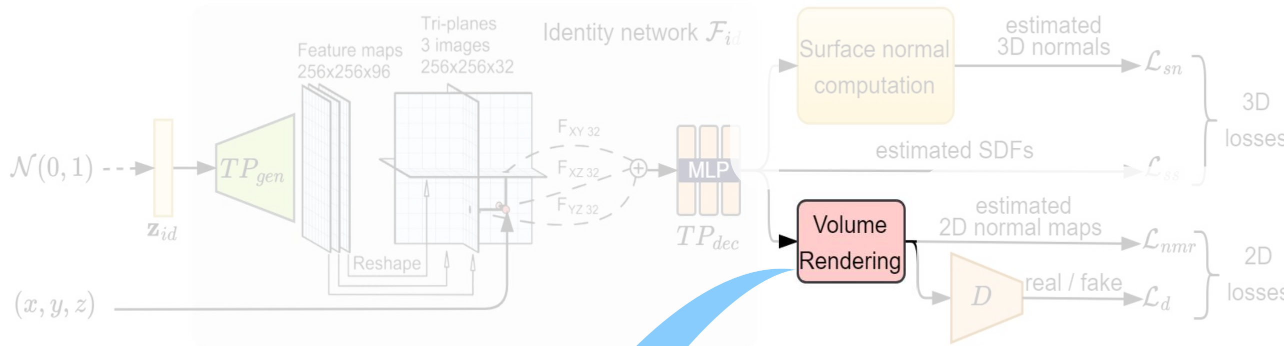
$n_{id}(\mathbf{x})$: ground truth 3D surface normal at **X**

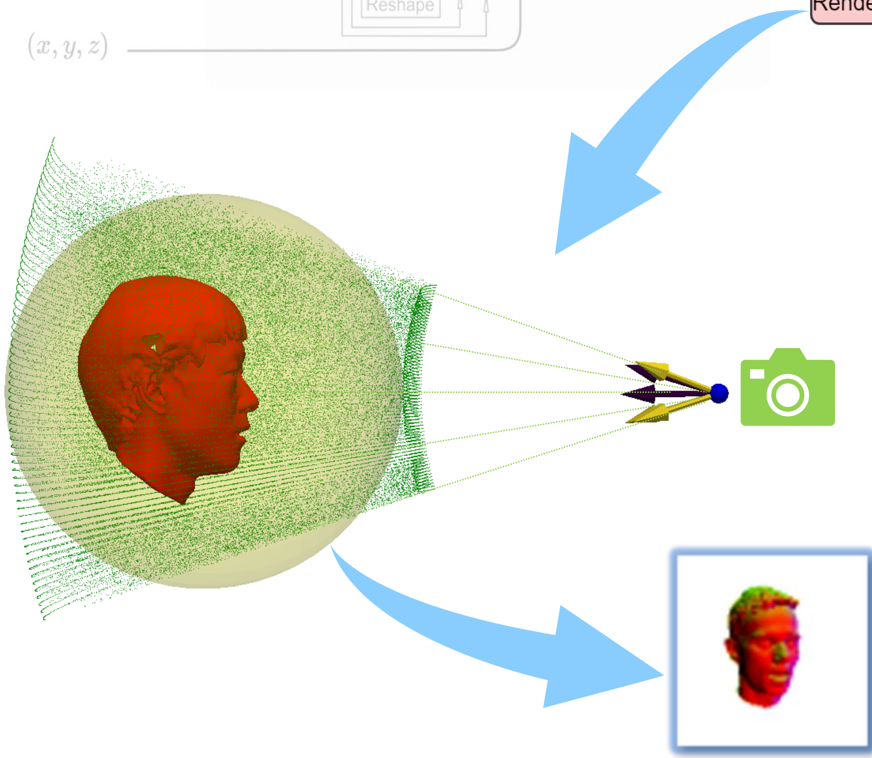
M : number of camera views

$N_{id}^{(v)}$: ground truth 2D normal map at view v

$\hat{N}_{id}^{(v)}$: diff. rendered 2D normal map at view v

Modeling Identity: Differentiable Volume Rendering





SDF(s) to density(σ):

$$\sigma_\beta(s) = \begin{cases} \frac{1}{2\beta}\exp\left(\frac{s}{\beta}\right) & \text{if } s \leq 0 \\ \frac{1}{\beta}\left(1 - \frac{1}{2}\exp\left(-\frac{s}{\beta}\right)\right) & \text{if } s > 0 \end{cases}$$

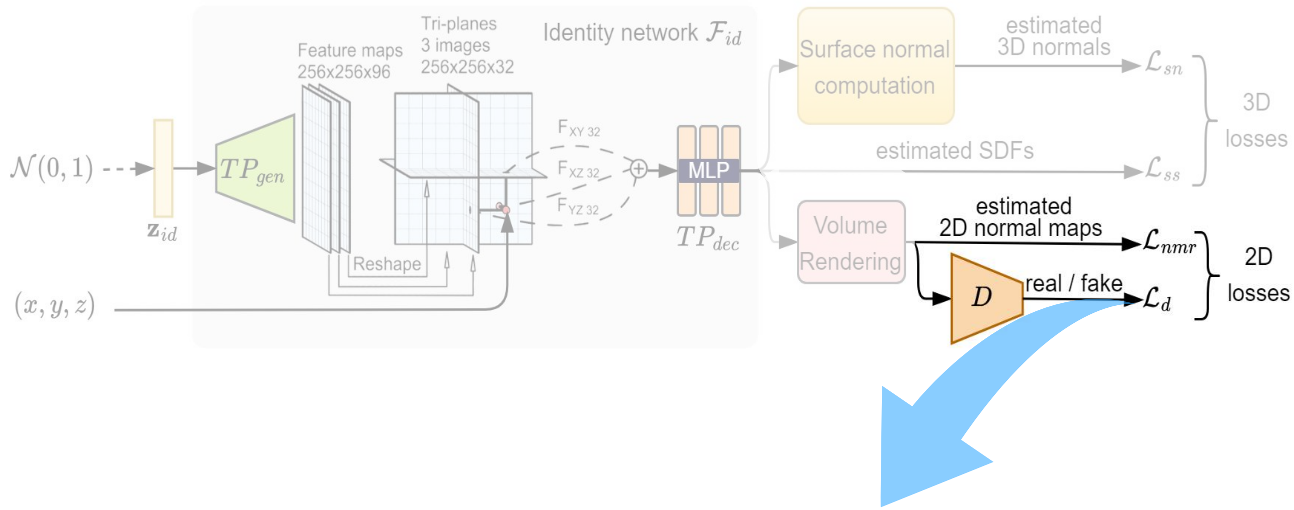
NeRF integration:

$$\hat{N}(\mathbf{r}) = \sum_{i=1}^{M} T_{\mathbf{r}}^i \alpha_{\mathbf{r}}^i \hat{\mathbf{n}}_{\mathbf{r}}^i$$

$$T_{\mathbf{r}}^i = \prod_{i=1}^{i-1}\left(1 - \alpha_{\mathbf{r}}^j\right)$$

$$\alpha_{\mathbf{r}}^i = 1 - \exp\left(-\sigma_{\mathbf{r}}^i \delta_{\mathbf{r}}^i\right)$$

Modeling Identity: 2D Adversarial Objectives



$$\mathcal{L}_{\text{d}}(D;G) := \mathbb{E}_{(\mathbf{z}_{id},v)}\left[f(D(G(\mathbf{z}_{id},v)))\right] + \mathbb{E}_{(N_{id}^{(v)})}\left[f(-D(N_{id}^{(v)}))\right] + \lambda\mathcal{L}_{gp}$$

$$\mathcal{L}_{\text{g}}(G;D) := \mathbb{E}_{(\mathbf{z}_{id},v)}\left[f(-D(G(\mathbf{z}_{id},v)))\right]$$

$$\mathcal{L}_{gp} := \mathbb{E}_{(N_{id}^{(v)})}\left[|\nabla D(N_{id}^{(v)})|^2\right]$$

$$f(u) := \log(1+\exp(u))$$

$$G(\mathbf{z}_{id},v) := \hat{N}_{id}^{(v)} = \mathcal{R}(\mathcal{F}_{\mathbf{id}}(\mathbf{z}_{id}), \mathbf{K}^{(v)}, \mathbf{Rt}^{(v)}, res)$$

$\mathcal{R}$: differentiable volume renderer

res : rendering resolution in pixels

v : viewing direction

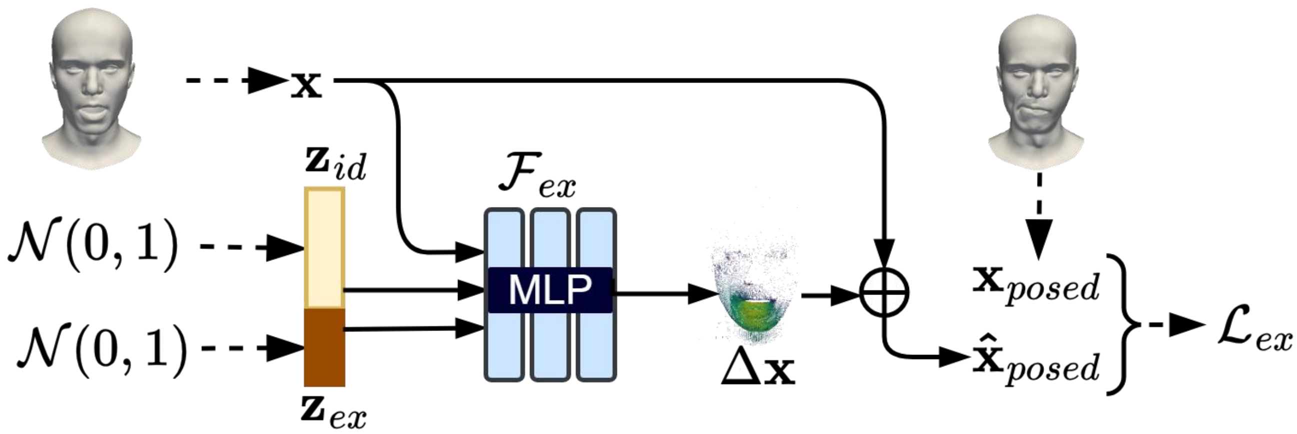
$\mathbf{K}$: camera intrinsics

$\mathbf{Rt}$: camera pose / extrinsics

$N_{id}^{(v)}$: ground truth 2D normal map

$\hat{N}_{id}^{(v)}$: diff. rendered 2D normal map

Modeling Expression as Forward Deformation



$$\mathcal{L}_c := \sum_{\mathbf{x}\in\mathbf{X}} \left((\mathbf{x}+\Delta\mathbf{x}) - \mathbf{x}_{posed}\right)^2$$

$$\Delta\mathbf{x} := \mathcal{F}_{\mathbf{ex}}(\mathbf{x}, \mathbf{z}_{ex}, \mathbf{z}_{id})$$

$$\mathcal{L}_{dp} := \sum_{\mathbf{x}\notin\mathbf{X}} ||\mathcal{F}_{\mathbf{ex}}(\mathbf{x}, \mathbf{z}_{ex}, \mathbf{z}_{id})||_2^2$$

$$\mathcal{L}_r := ||\mathbf{z}_{ex}||_2^2$$

$\mathbf{X}$: neutral pose face points in registered mesh

$\Delta\mathbf{x}$: local forward deformation / displacement

$\mathcal{L}_c$: correspondence loss to model deformation

$\mathcal{L}_{dp}$: deformation penalty for non-face regions

$\mathcal{L}_r$: latent regularizer for expression code

Experiments
&
Results

Dataset, Baselines and Evaluation Metrics

Dataset:

- NPHM[1] 3D face scans
- 255 subjects, ~21 expression / subject
- Roughly 5200 scans
- Registered in FLAME[2] template

Baselines:

- FLAME[2] (linear, PCA based)
- NPM[3] (neural, global MLP)
- NPHM[1] (neural, ensemble of local MLPs)

Evaluation metrics:

- **Chamfer-L_1**: point cloud similarity ↓
- **Normal Consistency (N.C.):** better surface reconstruction (orientation) ↑
- **F-Sore@5 mm:** accuracy and completeness ↑
- All metrics are computed with 2.5M point samples

[1]Giebenhain, Simon, et al. "Learning neural parametric head models." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.
[2]Li, Tianye, et al. "Learning a model of facial shape and expression from 4D scans." ACM Trans. Graph. 36.6 (2017): 194-1.
[3]Palafox, Pablo, et al. "Npms: Neural parametric models for 3d deformable shapes." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

Model Summary Comparison

	NPM	NPHM	Ours
Model size (Mio.)	7.349 / 7.351	3.014 / 1.362	7.337 / 7.351
3D representation	Global MLP	Local MLPs	Tri-Plane
Regularizer	eikonal	eikonal, symmetric anchors	eikonal, adversarial loss
Mesh extraction time at res 256 (Sec.)	18.319	33.119	**03.717**

Giebenhain, Simon, et al. "Learning neural parametric head models." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.
Palafox, Pablo, et al. "Npms: Neural parametric models for 3d deformable shapes." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

Identity Fitting: Qualitative Results

Identity Fitting: Quantitative Results

Method	L_1-Chamfer ↓		N.C. ↑		F-Score @ 1 mm ↑	
	face	head	face	head	face	head
FLAME	0.643	5.829	0.975	0.894	0.998	0.636
NPM	0.451	2.037	0.991	0.897	0.999	0.901
NPHM	**0.320**	1.360	**0.994**	0.924	**1.000**	0.957
Ours	0.359	**0.820**	0.993	**0.948**	**1.000**	**0.989**

Identity Interpolation

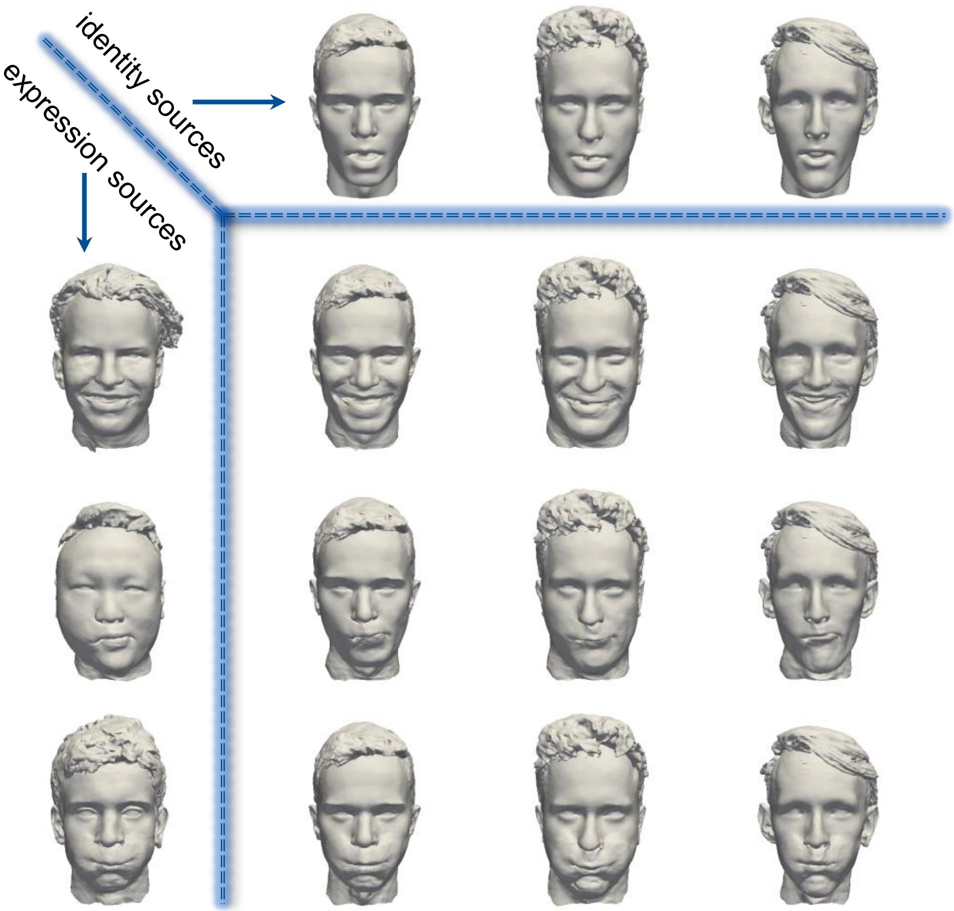
Expression Fitting: Qualitative Results

Expression Fitting: Quantitative Results

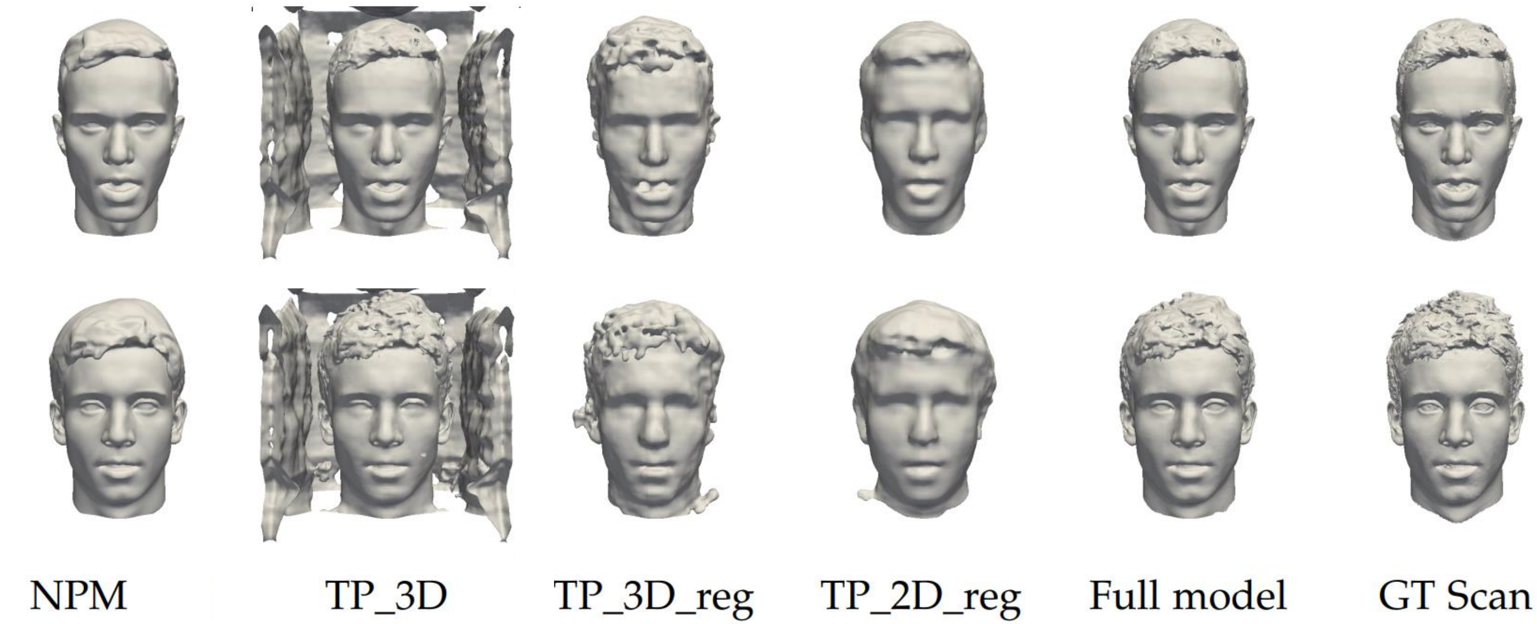
Method	L_1-Chamfer ↓		N.C. ↑		F-Score @ 1 mm ↑	
	face	head	face	head	face	head
FLAME	0.769	6.016	0.972	0.882	0.999	0.636
NPM	0.416	1.659	0.988	0.888	**1.000**	0.934
NPHM	**0.368**	1.313	**0.991**	0.909	**1.000**	0.965
Ours	0.650	**1.179**	0.981	**0.915**	0.998	**0.984**

Expression Interpolation

Expression Transfer



Ablation study: Qualitative



(MLP baseline)

(tri-plane
+ 3D loss)

(tri-plane
+ 3D loss
+ regularization)

(tri-plane
+ 2D loss
+ regularization)

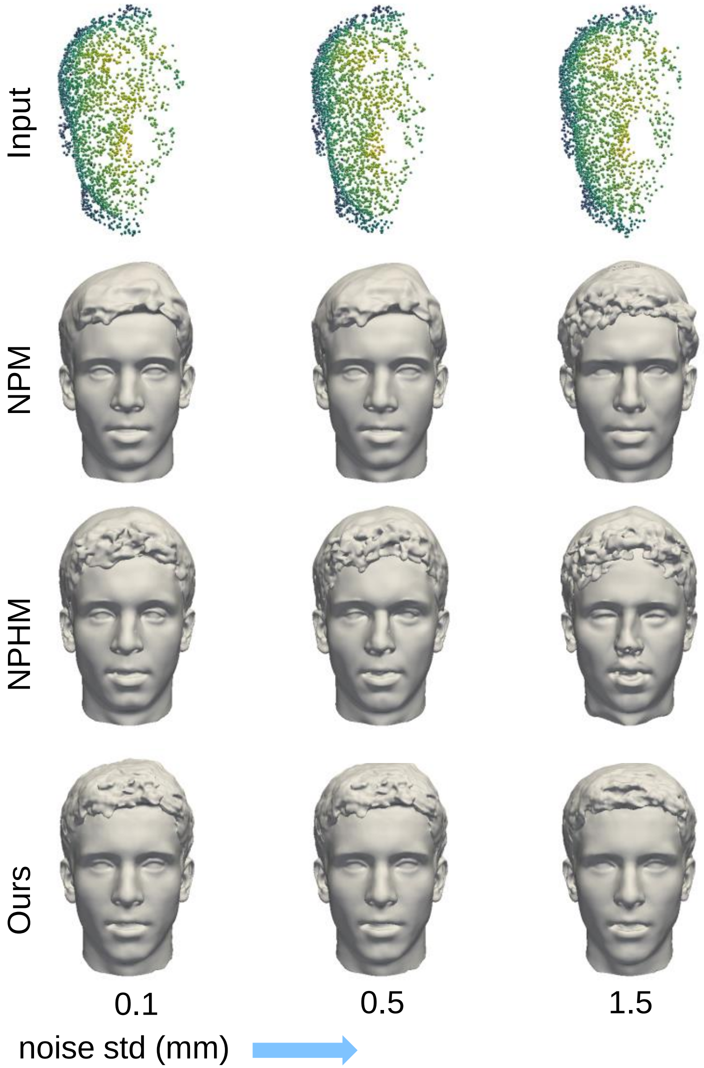
(tri-plane
+ 2D loss
+ 3D loss
+ regularization)

Ablation study: Quantitative

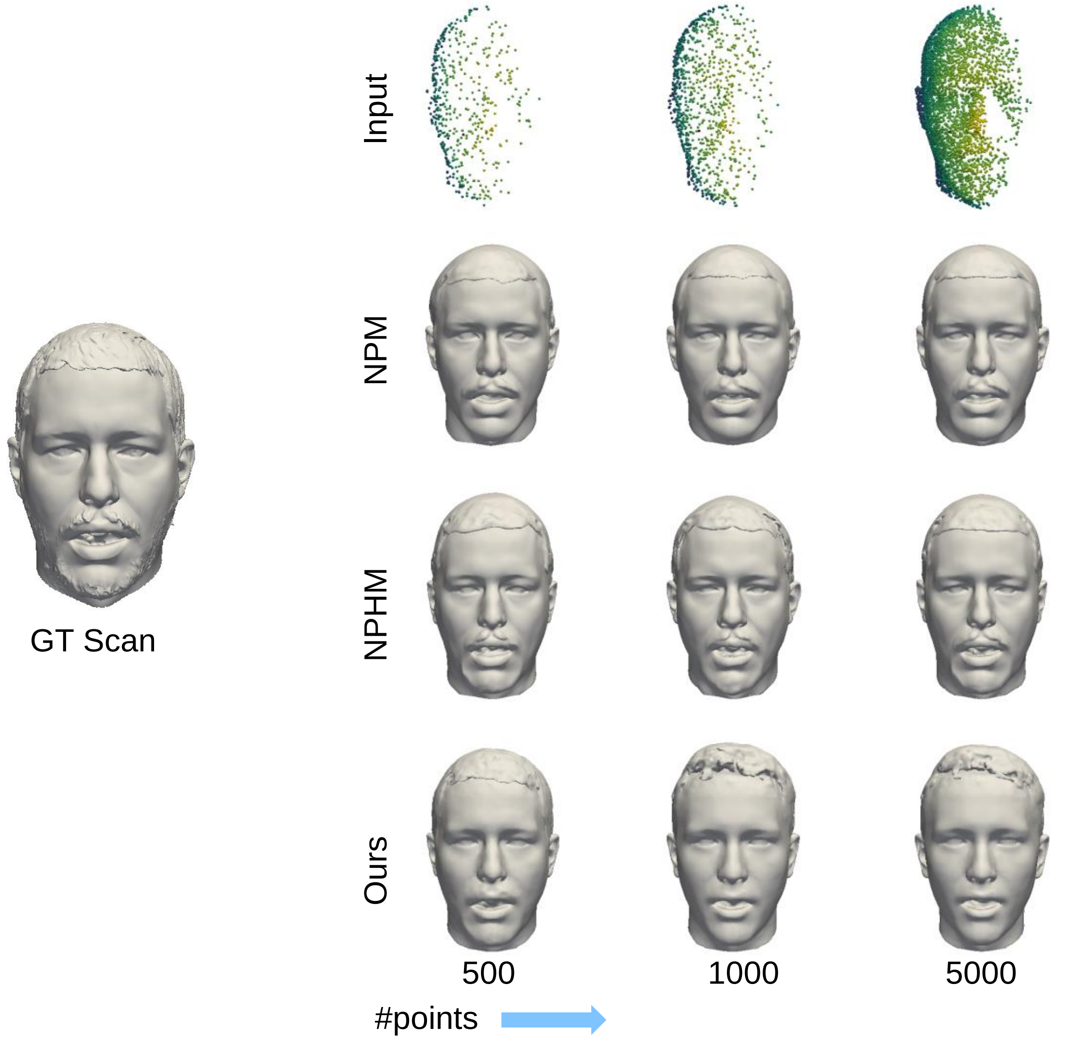
Method	L_1-Chamfer $\downarrow$	N.C. $\uparrow$	F-Score @ 1 mm $\uparrow$
NPM	1.361	0.924	0.957
TP_3D	17.226	0.869	0.642
TP_3D_reg	3.150	0.857	0.791
TP_2D_reg	3.785	0.866	0.747
Full model	**0.819**	**0.948**	**0.989**

Aditional ablation: Additive Gaussian Noise

GT Scan



Aditional ablation: Sparse Point Colud



Conclusion

Conclusion and Future Scope

In summary:

- A Parametric Head Model with adversarial loss
- Tri-plane representation capture more details and faster to infer
- Adversarial regularization helps to avoid unwanted artifacts

Limitation:

- Per-pixel ray shooting is inefficient
- Deformation model is not adversarially constrained

In future:

- Add GAN loss to deformation model as well
- Efficient sampling and ray shooting for vol rendering
- Use diffusion approach instead of GAN loss

Thank you!

Any questions?