USC Adobe

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Complete 3D Human Reconstruction From a Single Incomplete Image

Junying Wang [1] , Jae Shin Yoon [2], Tuanfeng Y. Wang [2], Krishna Kumar Singh [2], Ulrich Neumann [1]
[1] University of Southern California [2] Adobe Research

WED-AM-050

- Introduction
  - Research goal

Single image with occlusion

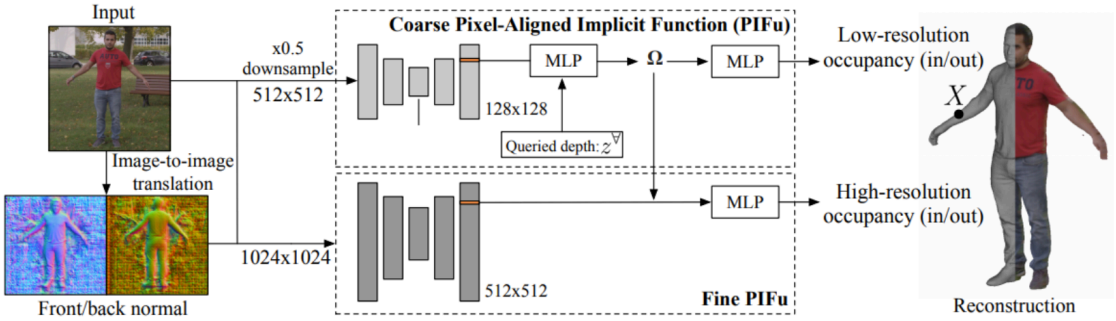
Complete 3D human model

Complete 3D human model with texture

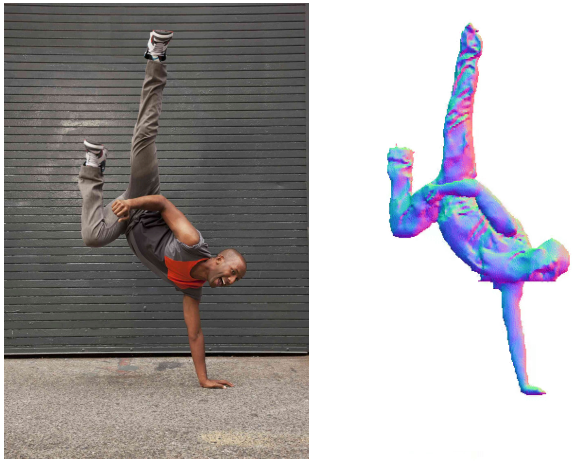
- Introduction
  - Contributions

- A new design of **generative volumetric feature** enable an implicit network to reconstruct a 3D human from an incomplete image.
- A novel **multi-view normal fusion** approach that upgrades the quality of local geometry details in a view-coherent way.
- An effective **texture inpainting** pipeline using the reconstructed 3D geometry.

- Related work
  - Single image human body reconstruction



PIFuHD [1] Pipeline



3D Reconstruction

[1] Saito, Shunsuke, et al. "Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization." *CVPR* 2020.

- Related work
  - Single image human body reconstruction

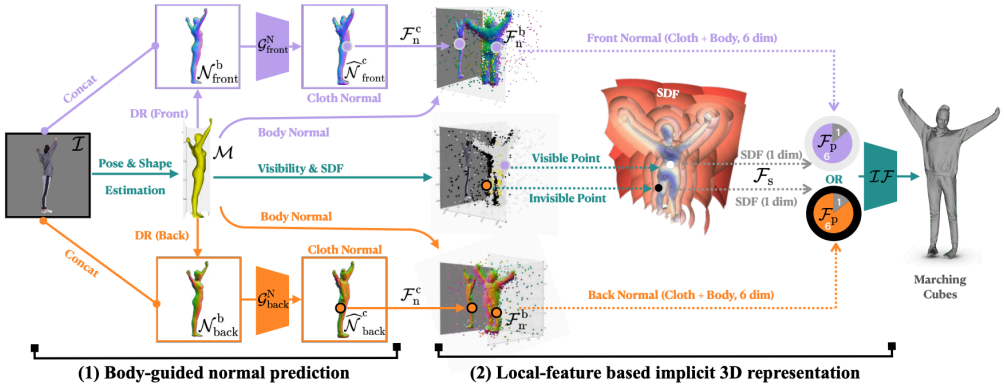
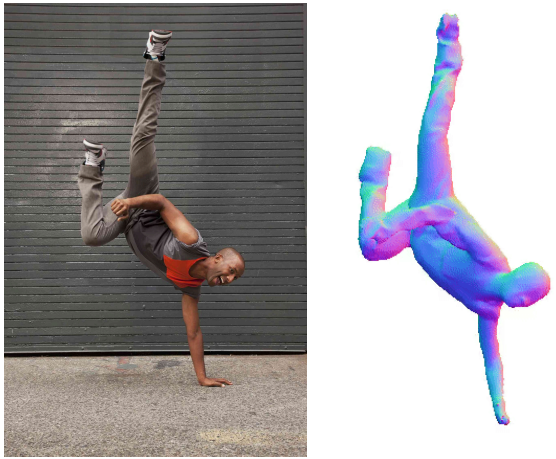


Figure 3. ICON's architecture contains two main modules for: (1) body-guided normal prediction, and (2) local-feature based implicit 3D reconstruction. The dotted line with an arrow is a 2D or 3D query function. The two $\mathcal{G}^N$ networks (purple/orange) have different parameters.

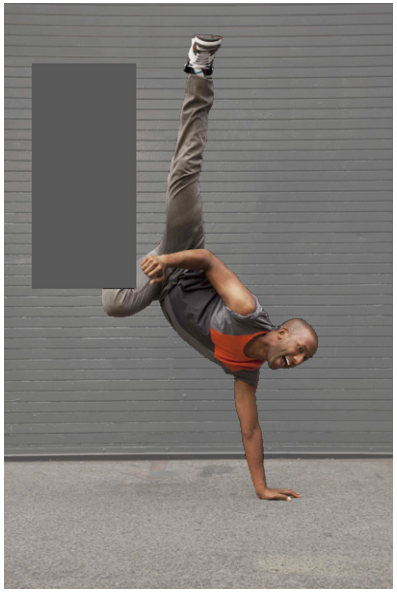
ICON [2] Pipeline



3D Reconstruction

[2] Xiu, Yuliang, et al. "ICON: implicit clothed humans obtained from normals." *CVPR* 2022.

- Related work
  - Single image human body reconstruction



Incomplete image

PIFuHD [1]

ICON [2]

[1] Saito, Shunsuke, et al. "Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization." *CVPR* 2020.
[2] Xiu, Yuliang, et al. "ICON: implicit clothed humans obtained from normals." *CVPR* 2022.

- Related work
  - 2D image inpainting then 3D reconstruction

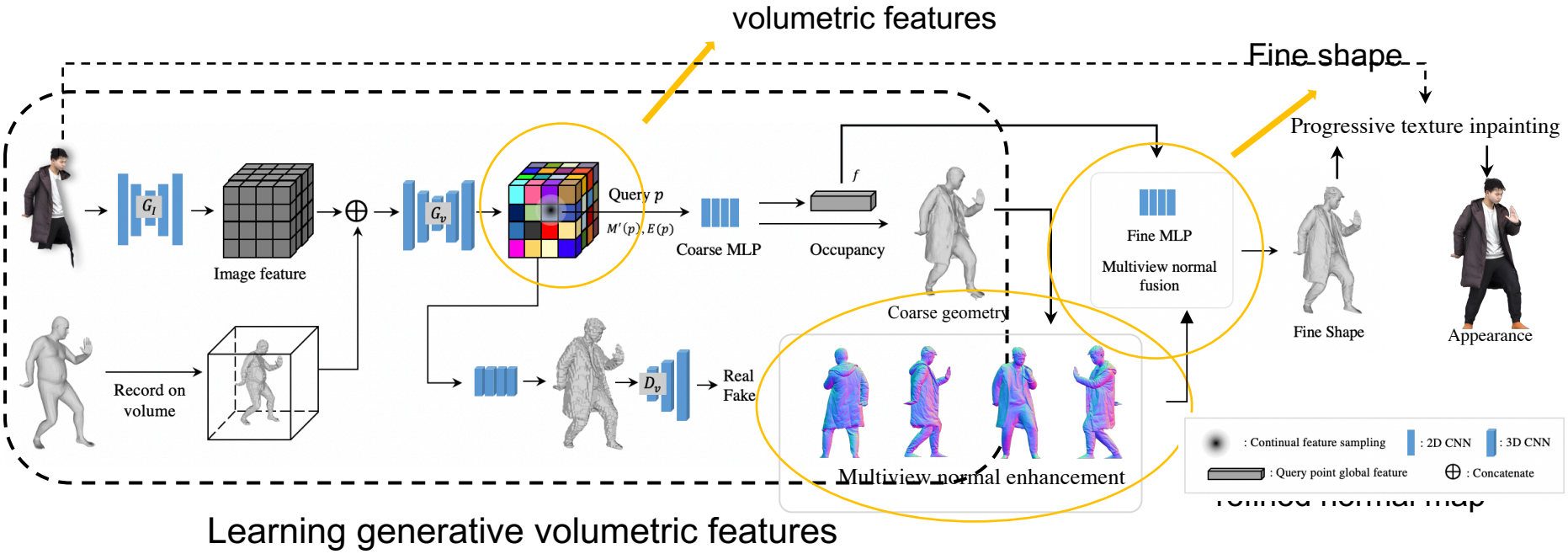
Inpainted image [3]

PIFuHD [1]

ICON [2]

[1] Saito, Shunsuke, et al. "Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization." *CVPR* 2020.
[2] Xiu, Yuliang, et al. "ICON: implicit clothed humans obtained from normals." *CVPR* 2022.
[3] Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models." *CVPR* 2022.

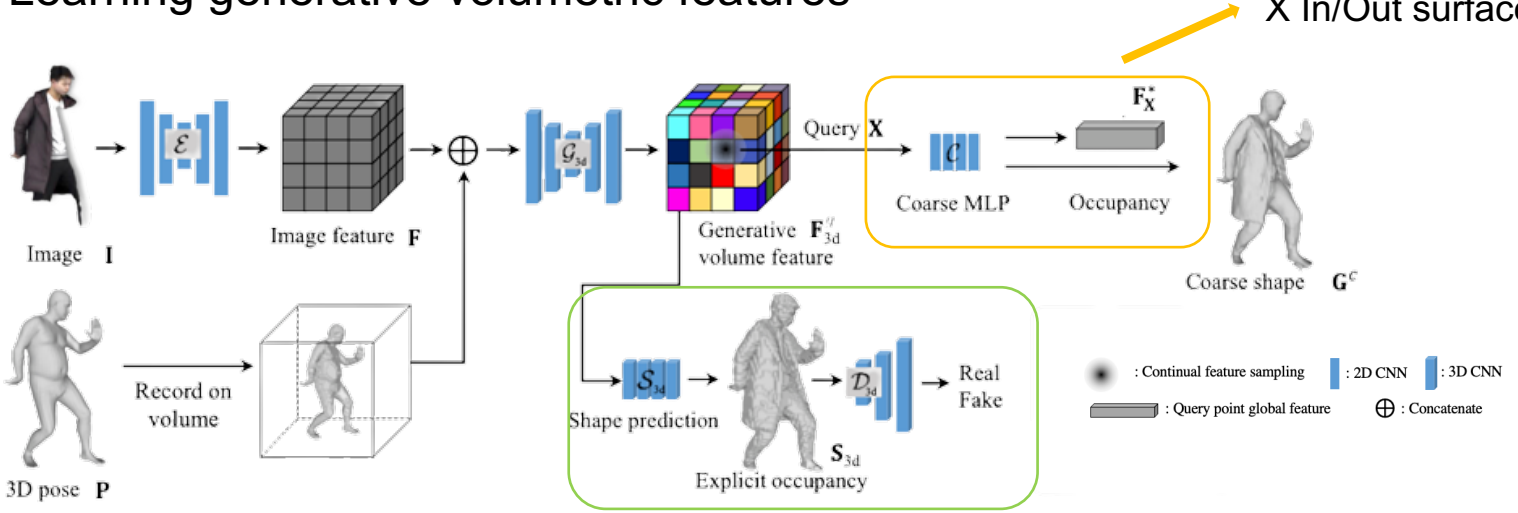
- Methods
  - Overview



Coarse-to-Fine 3D Generative Framework

❖ Methods

➢ Learning generative volumetric features



- Explicit volume feature
- Implicit function
- 3D discriminator

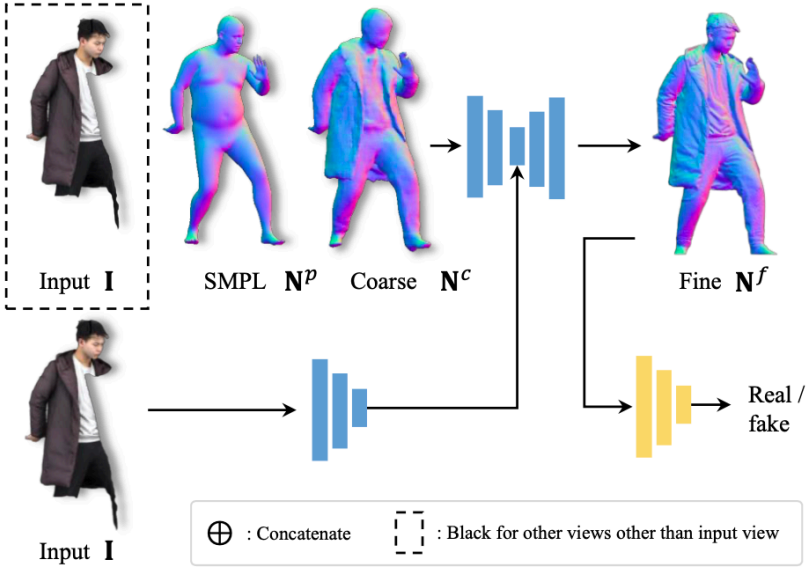
$$\mathcal{G}_{3d}(\mathbf{F};\mathbf{P}_{3d}) = \mathbf{F}^g_{3d}, \qquad \mathcal{C}(\mathbf{F}^g_{3d},\mathbf{X};\mathbf{X}) \rightarrow [0,1],$$
$$\mathcal{S}_{3d}(\mathbf{F}^g_{3d}) = \mathbf{S}_{3d}, \qquad \mathcal{D}_{3d}(\mathbf{S}_{3d};\mathbf{P}_{3d}) \rightarrow [0,1],$$

$$\mathcal{L}_c = \sum_i \|\mathcal{C}(\mathbf{F}^g_{3d,\mathbf{X}};\mathbf{X}) - \mathcal{C}_{gt}(\mathbf{X})\|^2,$$

$$L_{feat} = \mathcal{L}_c + \lambda_g \mathcal{L}_g + \lambda_{cGAN}\mathcal{L}_{cGAN},$$

❖ Methods

➢ Multiview normal enhancement



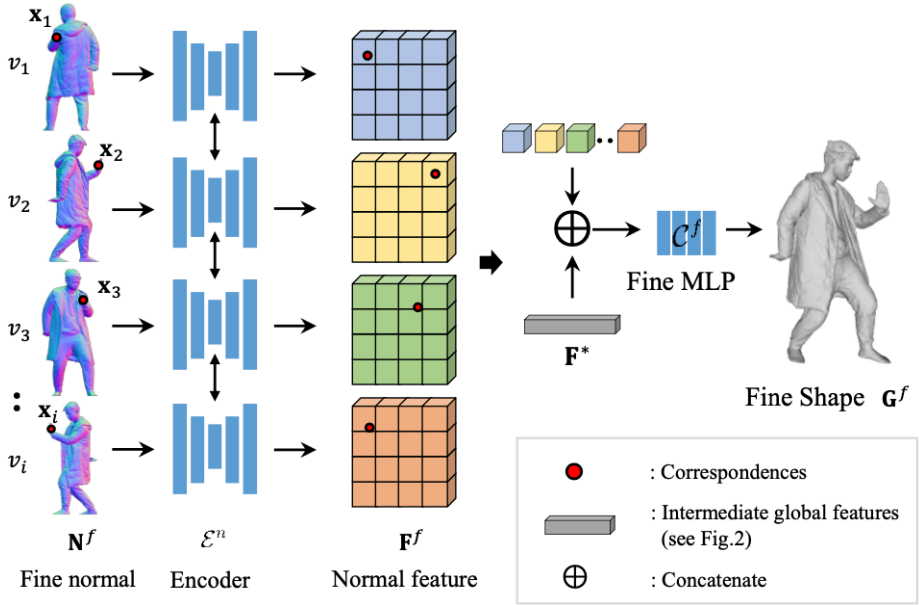
- Novel view coarse normal map rendering
- Multiview normal enhancement
- 2D discriminator

$$\mathcal{R}(\mathbf{G}^c; v_i) = \mathbf{N}^c_{v_i}, \quad \mathbf{N}^f_{v_i} = \mathcal{G}^n(\mathbf{N}^c_{v_i}; \mathbf{I}),$$

$$L_{\text{enhance}} = \mathcal{L}_1 + \lambda_{\text{vgg}}\mathcal{L}_{\text{vgg}} + \lambda_{\text{Adv}}\mathcal{L}_{\text{Adv}},$$

Methods

➢ Multiview normal fusion

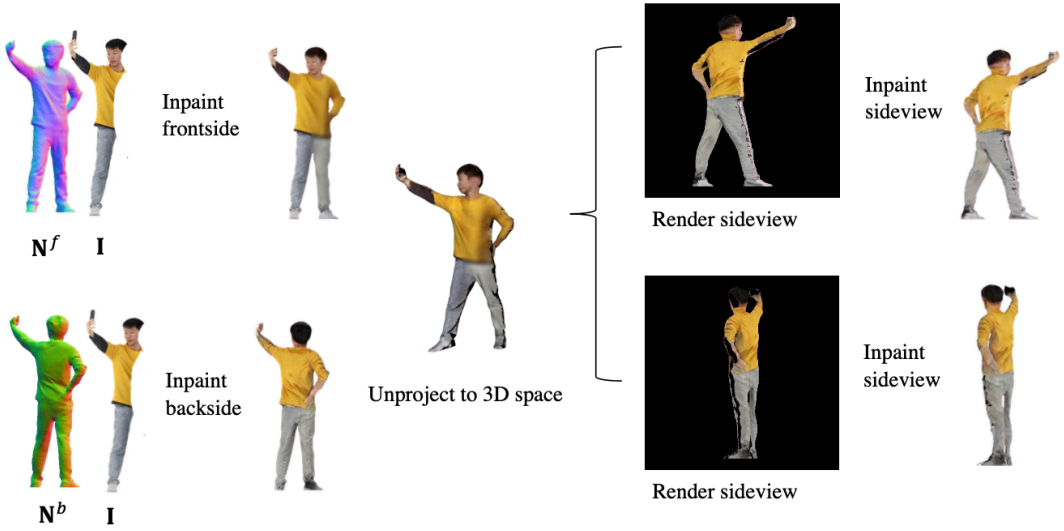


- Surface normal feature fusion
- Fine implicit function

$$\mathcal{L}_{\text{fusion}} = \sum_i \|\mathcal{C}^f(\{\mathbf{F}^n_{v_1,\mathbf{x}_1}, .., \mathbf{F}^n_{v_i,\mathbf{x}_i}\}; \mathbf{F}^*_{\mathbf{X}}) - \mathcal{C}_{\text{gt}}(\mathbf{X})\|.$$

❖ Methods

➢ Progressive texture inpainting [4]



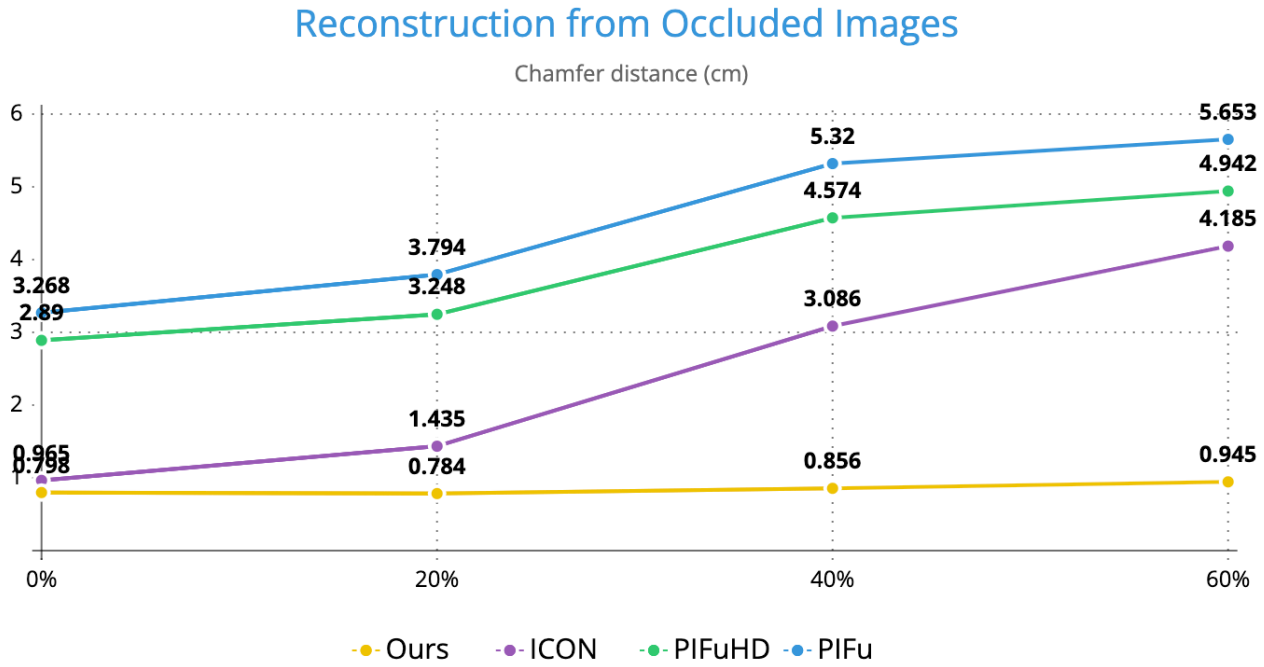
- Fine surface normal rendering
- Texture inpainting
- 3D warping to other views

[4] Yoon, Jae Shin, et al. "Pose-guided human animation from a single image in the wild." *CVPR* 2021.

- Results
  - Test on THuman2.0[5] unseen objects

[5] Yu, Tao, et al. "Function4d: Real-time human volumetric capture from very sparse consumer rgbd sensors." *CVPR* 2021.

- Results
  - Evaluation



Partial Body Image Reconstruction

Accumulative Occlusions

Input image

Ours

ICON

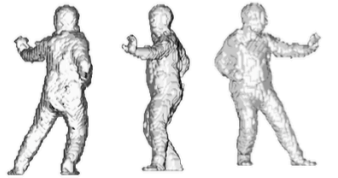
PIFuHD

PIFu

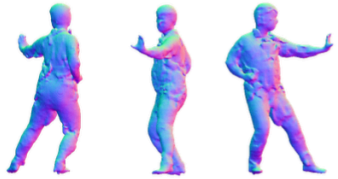
❖ Results

➢ Ablation study

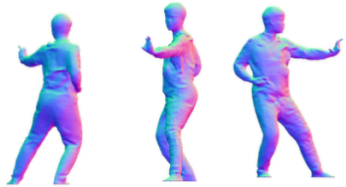
Input image



Explicit prediction



Coarse prediction



Fine prediction

- Explicit network
- Explicit + coarse implicit network
- Explicit + coarse implicit + fine implicit network

Method	Chamfer↓	P2S↓	Normal↑
Ours - coarse MLP - fine MLP	1.978	1.720	6.320
Ours - fine MLP	0.818	0.926	10.704
Ours w/o GT SMPL	1.224	1.062	12.106
Ours	**0.798**	**0.808**	**12.441**

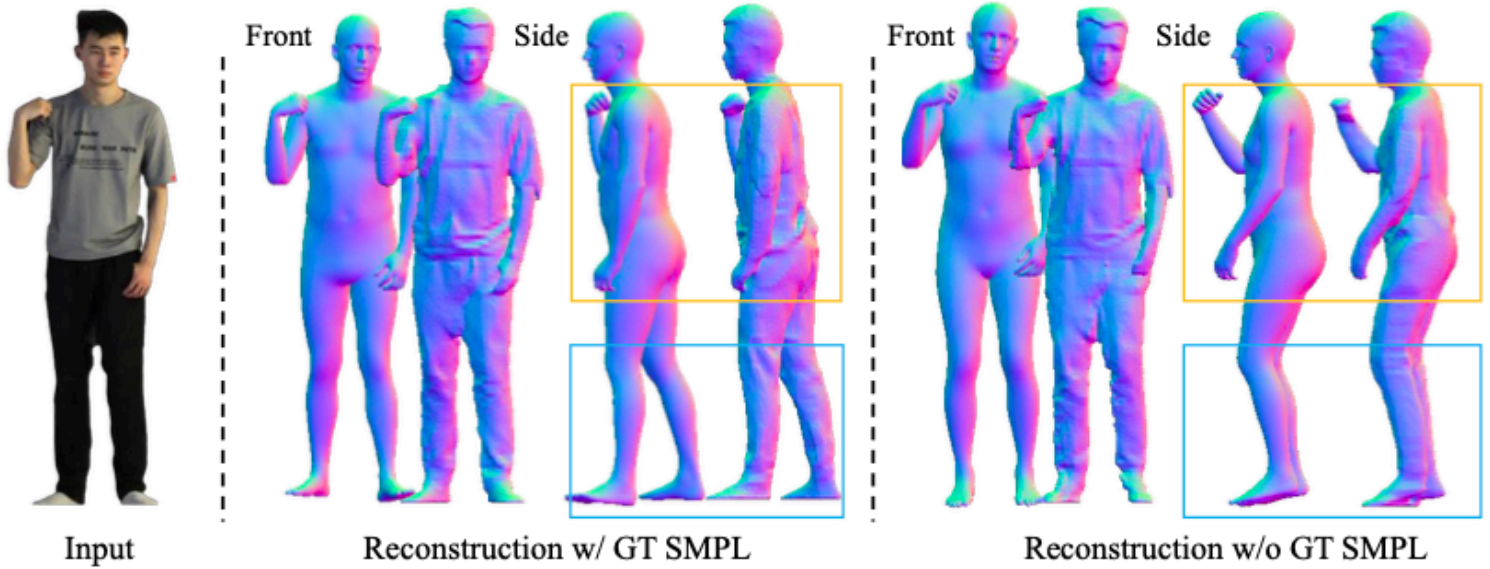
Group-shot image reconstruction

In-the-wild Group Shot Image

In-the-wild reconstruction

- Discussion
  - Limitation

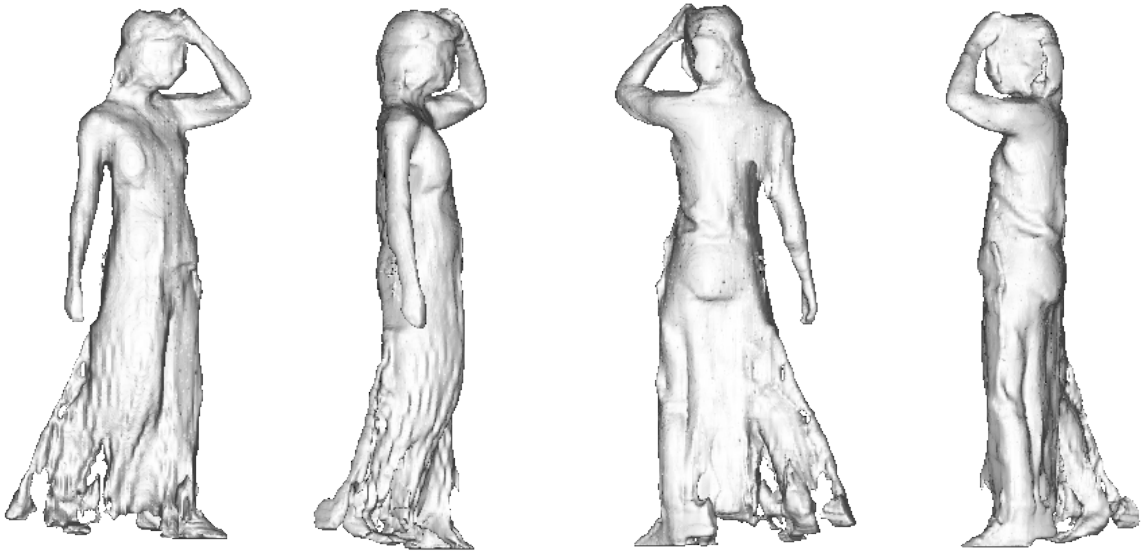
The requirement of an accurate 3D posed SMPL model



- Discussion
  - Limitation

Degradation of reconstruction quality with loose clothes

Input image



Our Reconstruction

❖ Conclusion

We present a method to reconstruct a complete human 3D model from a single image of a person with a partial body.

- We developed a new **coarse-to-fine framework** for human reconstruction using **generative volumetric features** learned by **3D GANs**, resulting in a complete 3D human geometry.
- The implicit fusion network **improves local geometry** quality by combining learned volumetric features and enhanced **multi-view surface normals** from coarse geometry.
- Our framework **performs well on scenes with occlusion**, as demonstrated by evaluations on diverse subjects with various testing setups. It shows significant improvement over existing methods.
- We also show that the complete and high-quality geometry from our method makes it possible to reconstruct **fully textured 3D human appearance** by applying an existing inpainting model in a view-progressive way.

Thank You !