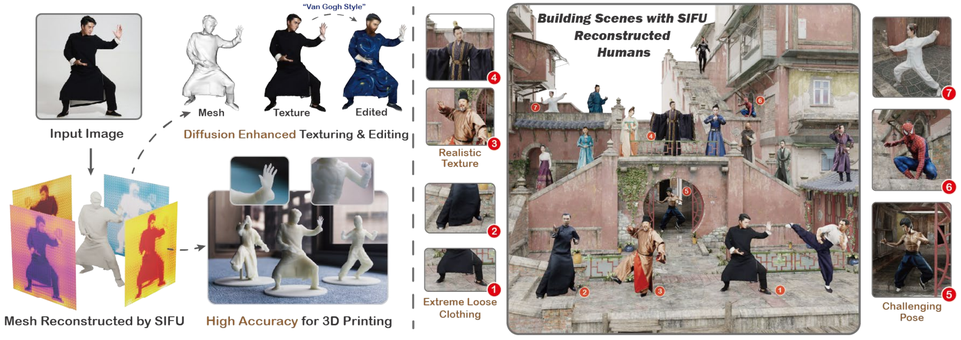
SIFU: Side-view Conditioned Implicit Function for Real-world Usable Clothed Human Reconstruction (Highlight)

Zechuan Zhang, Zongxin Yang,Yi Yang
ID: 3648

CVPR JUNE 17-21, 2024 SEATTLE, WA



Abstract

We introduce SIFU, a new method for creating high-quality 3D models of clothed humans from single images, tackling the challenge of reconstructing complex poses and predicting textures for unseen areas. SIFU combines a Side-view Decoupling Transformer and a 3D Consistent Texture Refinement pipeline to improve accuracy and texture realism. Utilizing cross-attention mechanisms and text-to-image diffusion techniques, it outperforms existing methods in both geometry and texture quality, proving useful in applications like 3D printing and scene building.

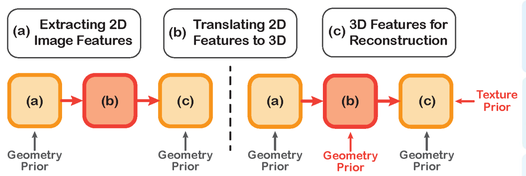
3D Printing and Texture Editing.

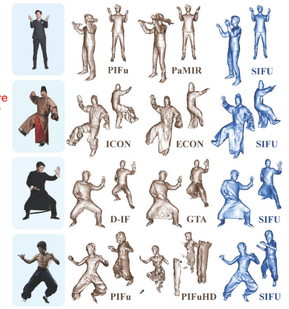
Scene Building.

Animation.

Applications of SIFU

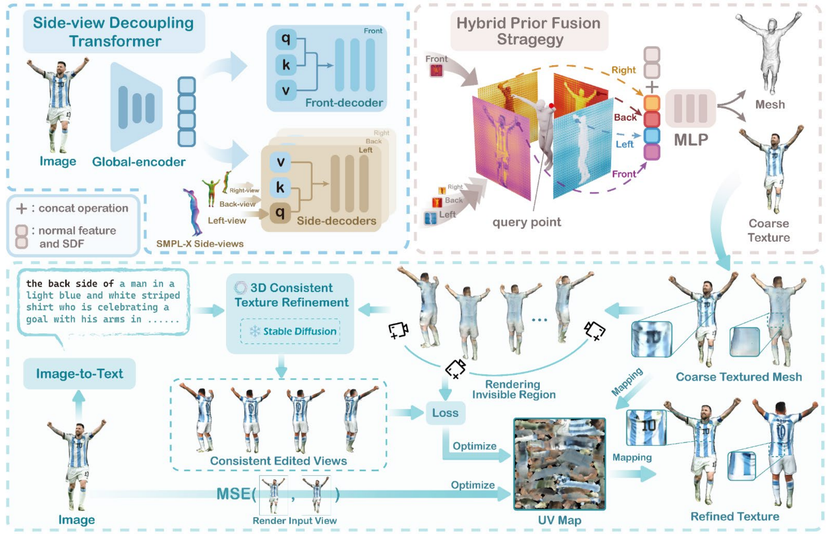
Motivation





Through analyzing existing methods, we pinpointed two key challenges in this field: **(i) Insufficient Prior Guid052 ance in Translating 2D Features to 3D (ii) Lack of Texture Prior.** Our approach improves the reconstruction process by incorporating additional guidance on geometry and texture priors.

Given a single image, SIFU constructs a 3D clothed human mesh with coarse textures using a Side-view Conditioned Implicit Function. This is followed by a step of 3D Consistent Texture Refinement to generate detailed textures. Specifically, SIFU employs a side-view decoupling transformer to decouple features from the input image and the side-view normals of the SMPL-X model. Then, these features are combined at a query point through a hybrid prior fusion strategy, aiding in the reconstruction of both the mesh and its texture. Finally, the mesh with its basic textures undergoes a diffusion-based 3D consistent texture refinement, ensuring feature consistency in the latent space and resulting in high-quality textures. Please see the paper for more details.



Project page: https://river-zhang.github.io/SIFU-projectpage/ Code: https://github.com/River-Zhang/SIFU

Experiment

Method	Publication	CAPE-NFP Chamfer ↓	CAPE-NFP P2S ↓	CAPE-NFP Normal ↓	CAPE-FP Chamfer ↓	CAPE-FP P2S ↓	CAPE-FP Normal ↓	THuman2.0 Chamfer ↓	THuman2.0 P2S ↓	THuman2.0 Normal ↓
w/o SMPL-X body prior										
PIFu * [71]	ICCV 2019	2.5609	1.9971	0.1023	1.8139	1.5108	0.0798	1.5991	1.4333	0.0843
PIFuHD [72]	CVPR 2020	3.7670	3.5910	0.1230	2.3020	2.3350	0.0900	-	-	-
w/ SMPL-X body prior										
PaMIR * [92]	TPAMI 2021	1.6313	1.2666	0.0730	1.481	1.1631	0.0727	1.2152	1.0582	0.0730
ICON [82]	CVPR 2022	0.8846	0.8569	0.0434	0.7247	0.6979	0.0371	0.9491	0.9846	0.0621
ECON [83]	CVPR 2023	0.9462	0.9334	0.0382	0.9039	0.8938	0.0373	1.2585	1.4184	0.0612
D-IF [85]	ICCV 2023	0.8237	0.8353	0.0575	0.7625	0.769	0.0503	1.1696	1.2900	0.0936
GTA [91]	NeurIPS 2023	0.8508	0.7920	0.0424	0.6525	0.6084	0.0349	0.7329	0.7297	0.0492
Ours	-	**0.7725**	**0.7354**	**0.0378**	**0.6297**	**0.5980**	**0.0327**	**0.5961**	**0.6058**	**0.0407**

Table 1. **Quantitative evaluation against SOTA (§4.1).** All models use a resolution of 256 for marching cubes and ground-truth SMPL-X models are used during testing. *Methods are re-implemented in [82] for a fair comparison. Top two results are colored as first second .

Method	Backbone	Chamfer ↓	P2S ↓	Normal ↓
PaMIR [92]	CNN	1.3224	1.1349	0.0767
ICON [82]	CNN	1.2935	1.3949	0.0781
D-IF [85]	CNN	1.5262	1.7296	0.1191
ECON [83]	-	2.1195	1.8074	0.1029
GTA [91]	Transformer	1.0473	1.0780	0.0649
Ours	Transformer	**0.9937**	**1.0645**	**0.0599**

Table 2. **Assessing model robustness to SMPL-X (§4.1).** To evaluate the models' robustness in reconstruction, we used the THuman2.0 dataset [86] and introduced random noise to the ground-truth SMPL-X models. This approach simulates inaccuracies in poses and shapes for robustness testing.

Method	Diffusion-based	PSNR ↑	SSIM ↑	LPIPS ↓
PIFu [71]	✗	18.0934	0.9117	0.1372
Impersonator++ [52]	✗	16.4791	0.9012	0.1468
TEXTure [69]	✓	16.7869	0.8740	0.1435
Magic123 [67]	✓	14.5013	0.8768	0.1880
S3F [11]	✗	14.1212	0.8840	0.1868
HumanSGD [1]	✓	17.3651	0.8946	0.1300
SIFU w/o refinement	✗	22.0256	0.9212	0.0849
SIFU	✓	**22.1024**	**0.9236**	**0.0794**

Table 6. **Texture comparison on THuman2.0 [86] (§6.4).** During testing, PIXIE [16] is used for SMPL-X estimation.