CHROME: Clothed Human Reconstruction with Occlusion-Resilience and Multiview-Consistency from a Single Image

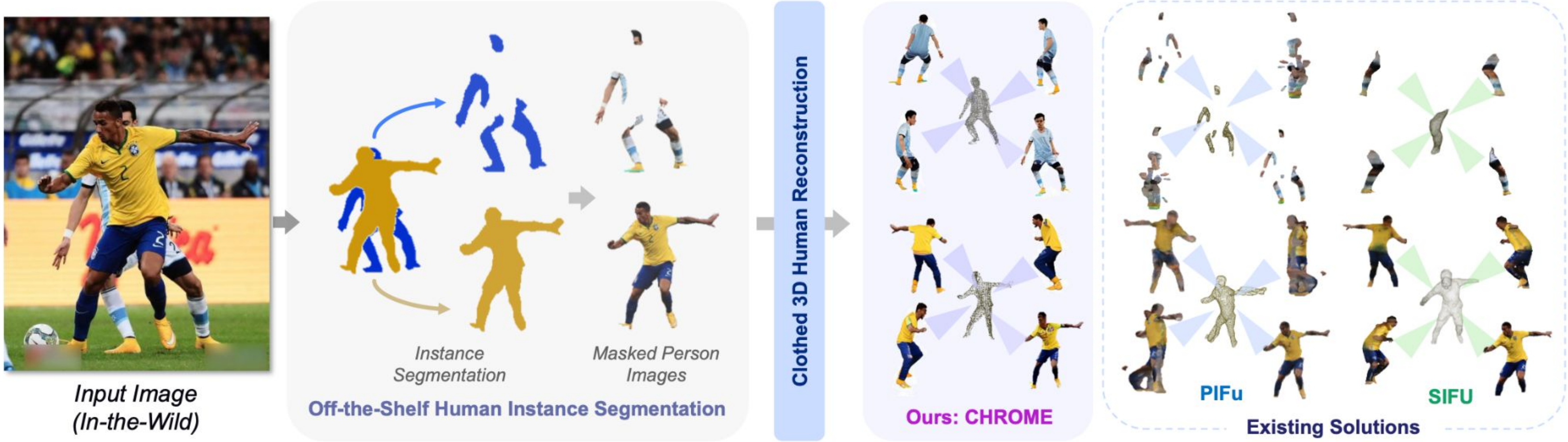
Arindam Dutta, Meng Zheng, Zhongpai Gao, Benjamin Planche, Anwesa Chaudhuri, Terrence Chen, Amit K. Roy-Chowdhury, Ziyan Wu

UCR VLG

UNITED IMAGING

ICCV OCT 19-23, 2025 HONOLULU HAWAII

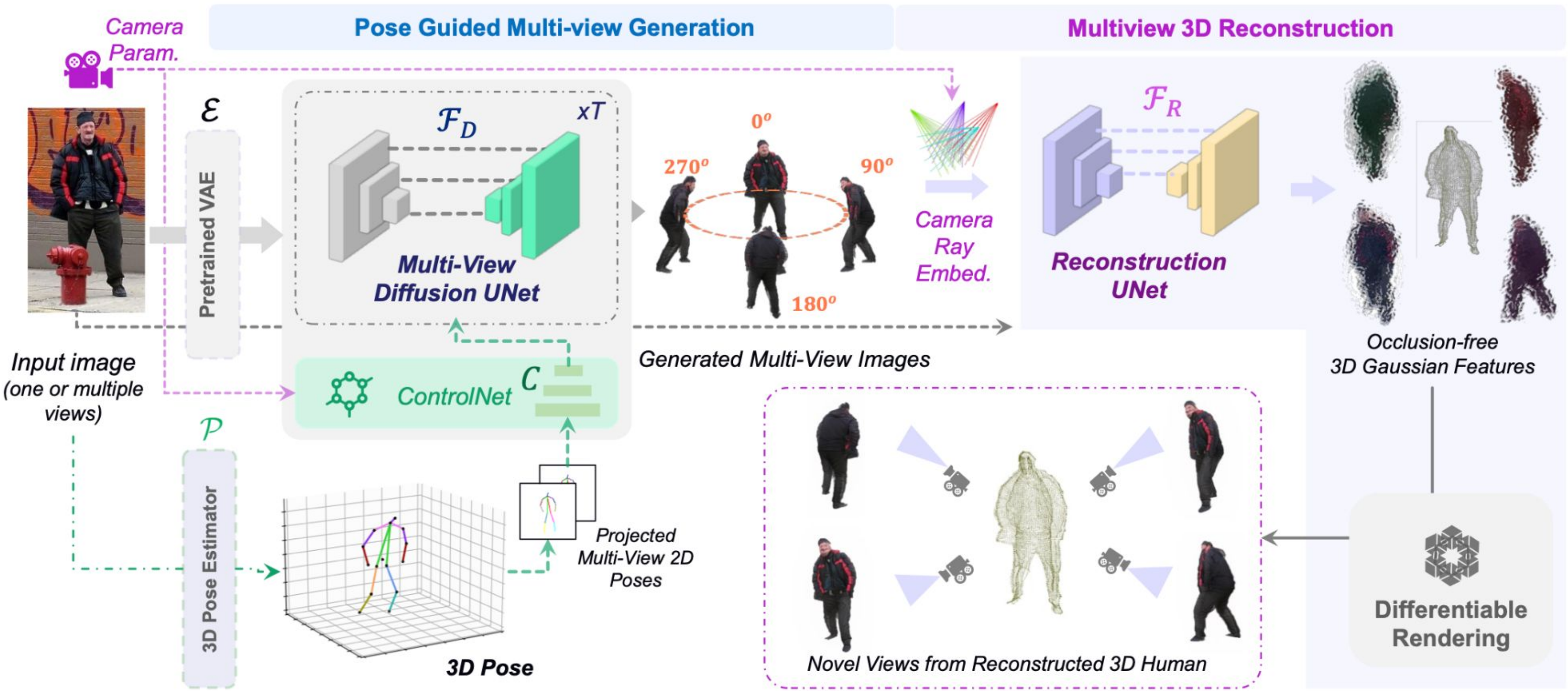
Introduction



Introduction

- **Problem & gap:** Single-view 3D clothed human reconstruction breaks under occlusions—current SMPL-, NeRF-/3DGS-based, and LRM methods yield fragmented, multi-view-inconsistent results and rely on costly supervision; multiview capture is impractical.
- **Method (CHROME):** A two-stage pipeline: (i) multiview diffusion with off-the-shelf pose control synthesizes consistent, de-occluded views from a single occluded image; (ii) a 3D Gaussian reconstructor, conditioned on the occluded input plus synthesized views and trained with 2D photometric loss, builds a coherent 3D model—no SMPL/3D GT required; stereo extension supported.
- **Outcomes:** Occlusion-resilient, multiview-consistent geometry and texture enabling robust novel-view synthesis, with strong results in both in-domain and zero-shot settings.

Methodology


Pose Guided Multi-view Generation
Multiview 3D Reconstruction
Camera Param.
$\mathcal{E}$
Pretrained VAE
$\mathcal{F}_D$
xT
Multi-View Diffusion UNet
ControlNet
$\mathcal{C}$
Input image (one or multiple views)
$\mathcal{P}$
3D Pose Estimator
3D Pose
Projected Multi-View 2D Poses
0°
90°
180°
270°
Generated Multi-View Images
Camera Ray Embed.
$\mathcal{F}_R$
Reconstruction UNet
Occlusion-free 3D Gaussian Features
Differentiable Rendering
Novel Views from Reconstructed 3D Human

Methodology

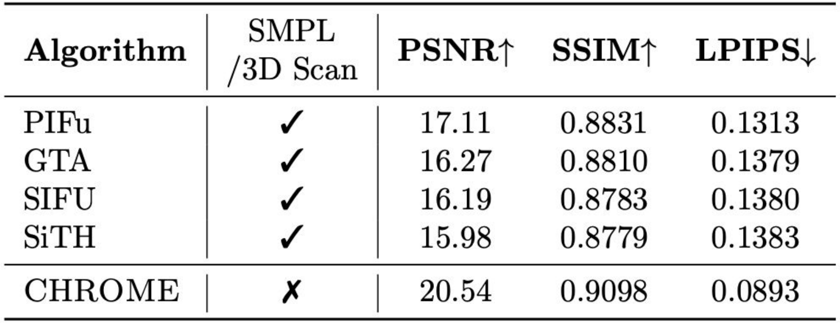
Stage-1 (Diffusion): From a single occluded image, a pose-controlled latent diffusion model—conditioned on VAE features of the visible regions and ControlNet with 2D poses (from a 3D pose estimator)—synthesizes four de-occluded, cross-view-consistent images.

Stage-2 (Reconstruction): A UNet-based reconstructor takes [occluded input + synthesized views] and predicts a 3D Gaussian field, differentiably rendered to enforce cross-view geometric/texture consistency and enable novel-view synthesis.

Training & robustness: End-to-end fine-tuning (Zero123++ init) on rendered human scans with synthetic occlusions, supervised only by 2D photometric/perceptual/silhouette losses—no SMPL/3D GT—yields occlusion-resilient reconstructions and supports multiview inputs.

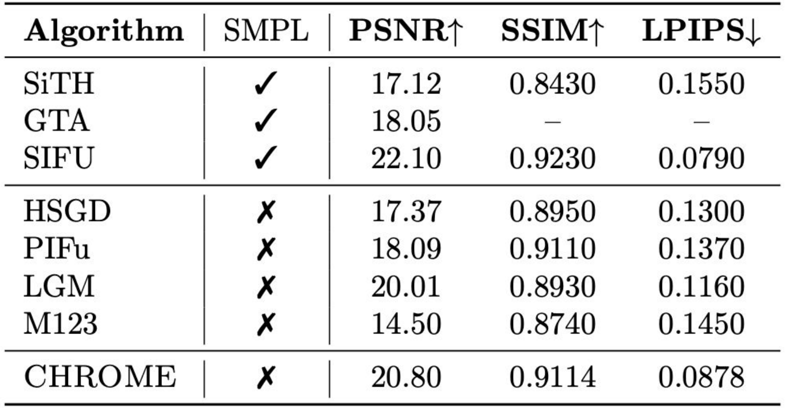
Quantitative Results

Quantitative comparison for Novel View Texture Reconstruction on Occluded THuman2.0. “SMPL” denotes requiring ground-truth SMPL annotation for training. “3D Scan” denotes requiring scan-level supervision.



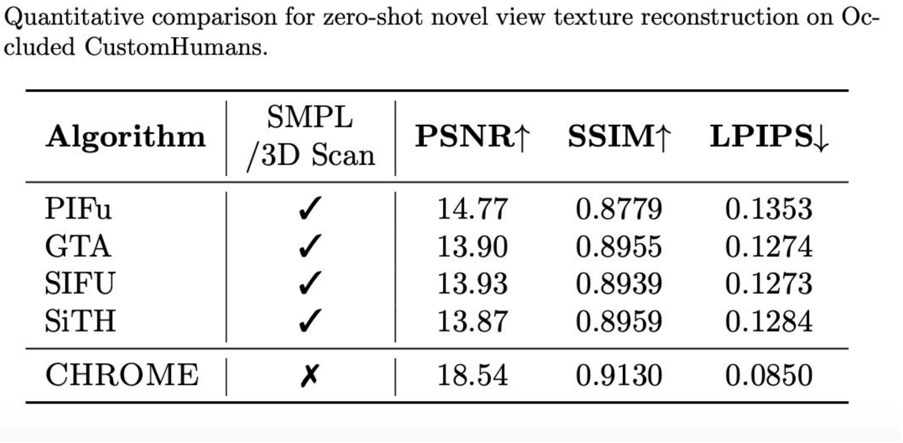
Algorithm	SMPL /3D Scan	**PSNR↑**	**SSIM↑**	**LPIPS↓**
PIFu	✓	17.11	0.8831	0.1313
GTA	✓	16.27	0.8810	0.1379
SIFU	✓	16.19	0.8783	0.1380
SiTH	✓	15.98	0.8779	0.1383
CHROME	✗	20.54	0.9098	0.0893

Quantitative comparison for novel view texture reconstruction on **Clean** THuman2.0.



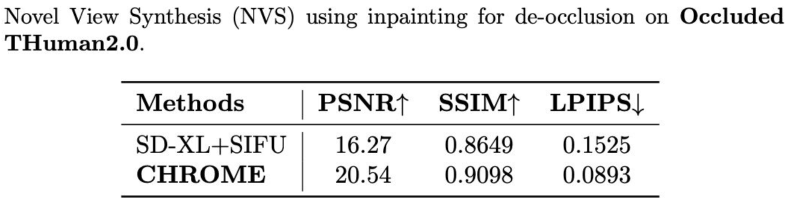
Algorithm	SMPL	**PSNR↑**	**SSIM↑**	**LPIPS↓**
SiTH	✓	17.12	0.8430	0.1550
GTA	✓	18.05	–	–
SIFU	✓	22.10	0.9230	0.0790
HSGD	✗	17.37	0.8950	0.1300
PIFu	✗	18.09	0.9110	0.1370
LGM	✗	20.01	0.8930	0.1160
M123	✗	14.50	0.8740	0.1450
CHROME	✗	20.80	0.9114	0.0878

Quantitative Results



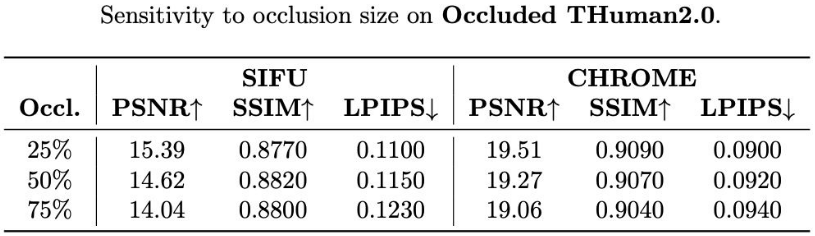
Quantitative comparison for zero-shot novel view texture reconstruction on Occluded CustomHumans.

Algorithm	SMPL /3D Scan	**PSNR↑**	**SSIM↑**	**LPIPS↓**
PIFu	✓	14.77	0.8779	0.1353
GTA	✓	13.90	0.8955	0.1274
SIFU	✓	13.93	0.8939	0.1273
SiTH	✓	13.87	0.8959	0.1284
CHROME	✗	18.54	0.9130	0.0850



Novel View Synthesis (NVS) using inpainting for de-occlusion on **Occluded THuman2.0**.

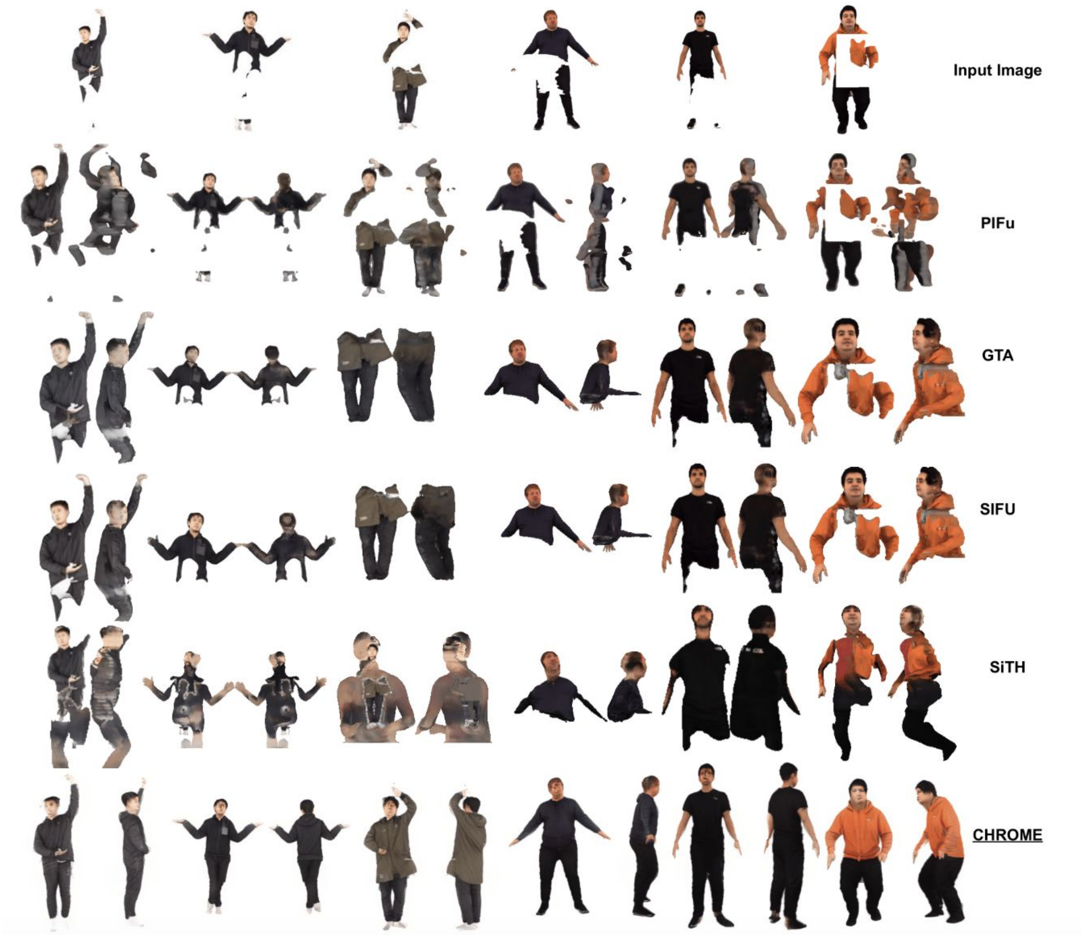
Methods	**PSNR↑**	**SSIM↑**	**LPIPS↓**
SD-XL+SIFU	16.27	0.8649	0.1525
CHROME	20.54	0.9098	0.0893

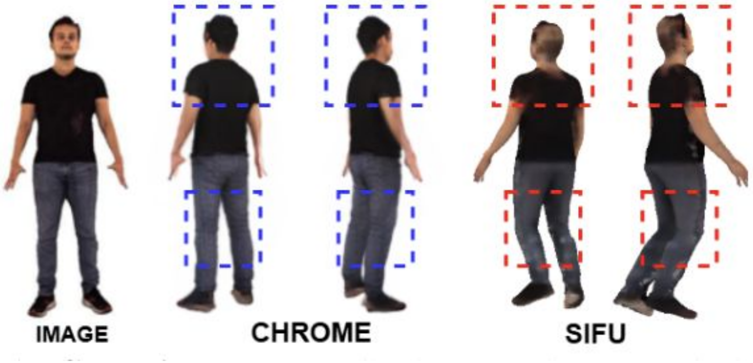


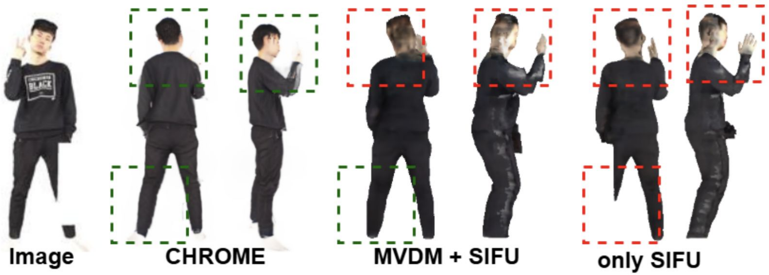
Sensitivity to occlusion size on **Occluded THuman2.0**.

	SIFU			**CHROME**		
Occl.	**PSNR↑**	**SSIM↑**	**LPIPS↓**	**PSNR↑**	**SSIM↑**	**LPIPS↓**
25%	15.39	0.8770	0.1100	19.51	0.9090	0.0900
50%	14.62	0.8820	0.1150	19.27	0.9070	0.0920
75%	14.04	0.8800	0.1230	19.06	0.9040	0.0940

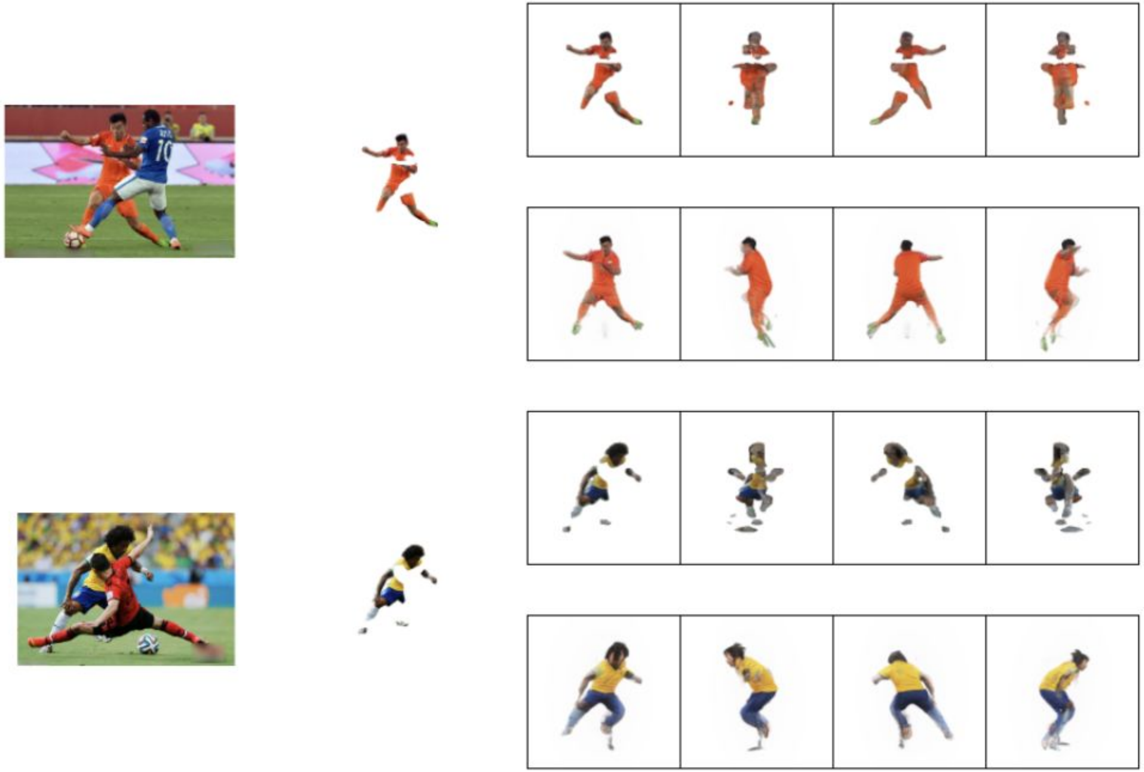
Qualitative Results







Qualitative Results (in-the-wild)



PIFu

CHROME (ours)

PIFu

CHROME (ours)

Extension to Stereo Reconstruction

Novel view texture reconstruction with **stereo inputs** on Occluded THuman2.0. Angle is the separation between the two views relative to the front-facing frame.

Stereo Angle	PSNR↑	SSIM↑	LPIPS↓
45°	24.32	0.9280	0.0542
90°	24.70	0.9310	0.0521
135°	24.78	0.9313	0.0511