

# Geometry-Guided Progressive NeRF for Generalizable and Efficient Neural Human Rendering

Mingfei Chen<sup>1,2</sup>, Jianfeng Zhang<sup>3</sup>, Xiangyu Xu<sup>1</sup>, Lijuan Liu<sup>1</sup>, Yujun Cai<sup>1</sup>, Jiashi Feng<sup>1</sup>, and Shuicheng Yan<sup>1</sup>

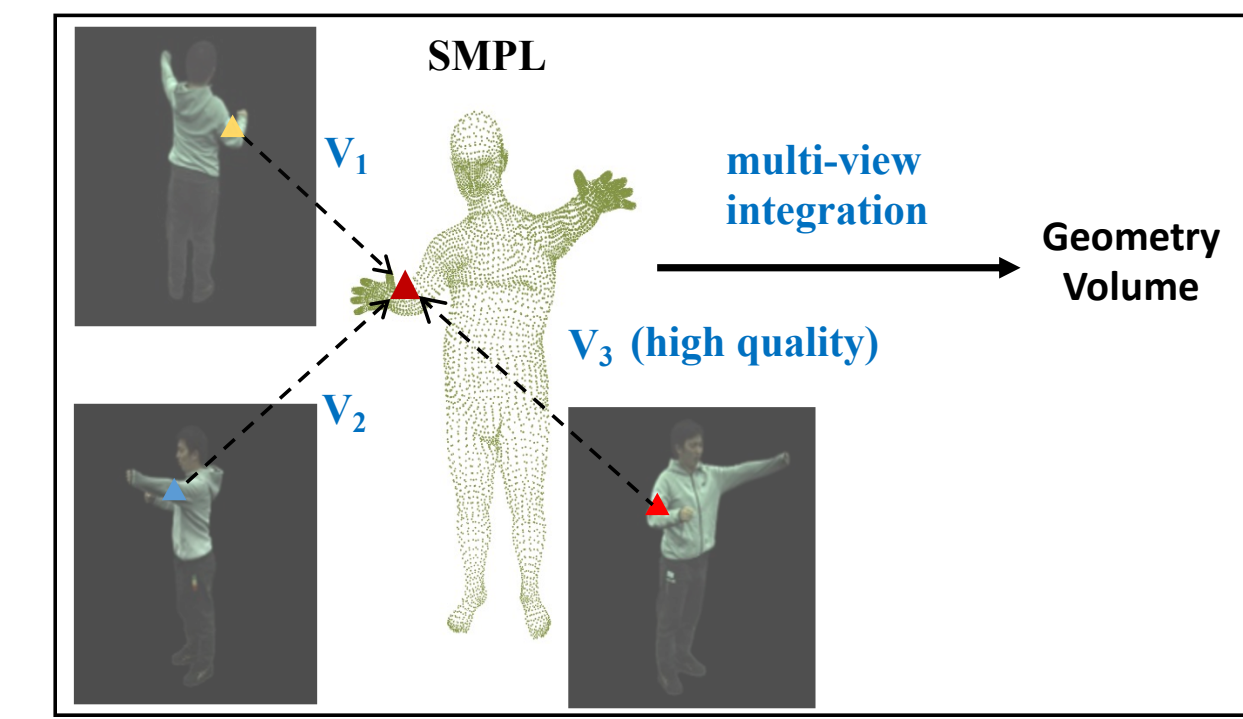
<sup>1</sup> Sea AI Lab

<sup>2</sup> University of Washington

<sup>3</sup> National University of Singapore

## Challenges

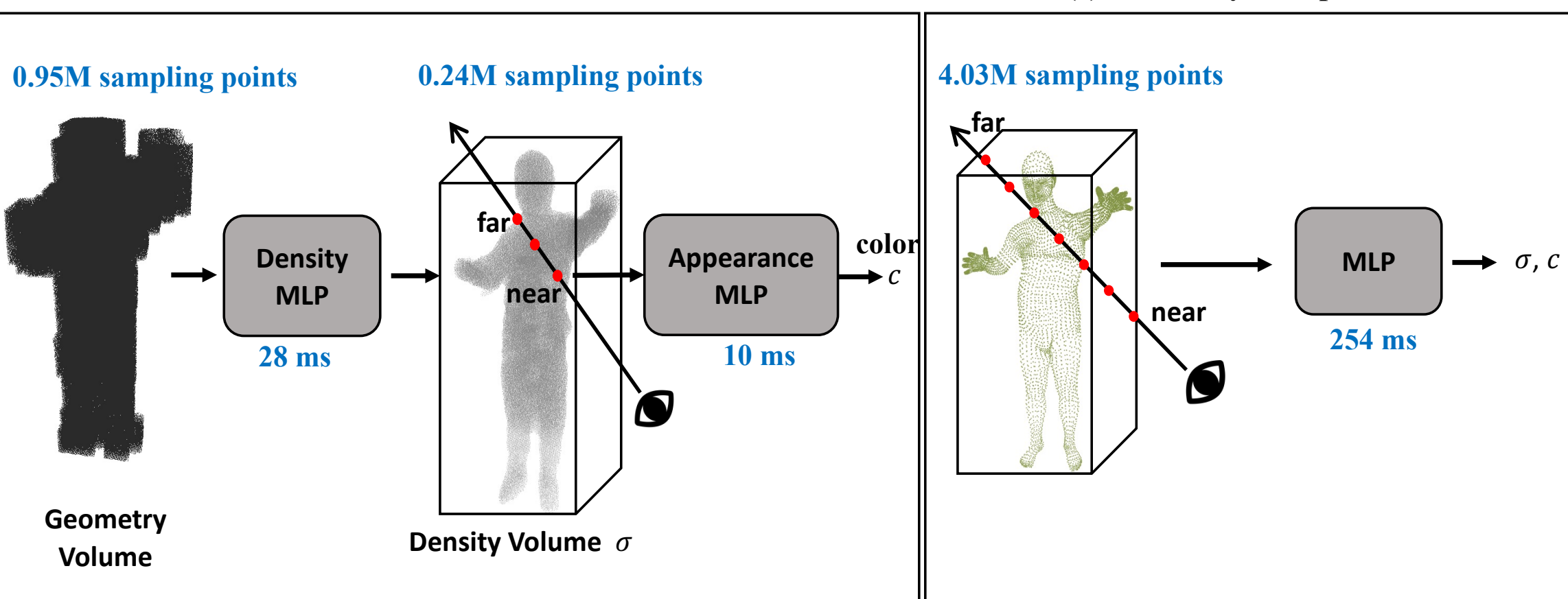
Free-viewpoint human body synthesis with sparse camera views



(a) Geometry-guided image feature integration: V for view.

	Previous	Ours
# Density Points (↓)	4.03M	0.95M (-76%)
Density MLP T (↓)	109ms	28ms (-74%)
# Color Points (↓)	4.03M	0.24M (-94%)
Color MLP T (↓)	145ms	10ms (-93%)
Memory (↓)	20.7GB	9.9GB (-52%)

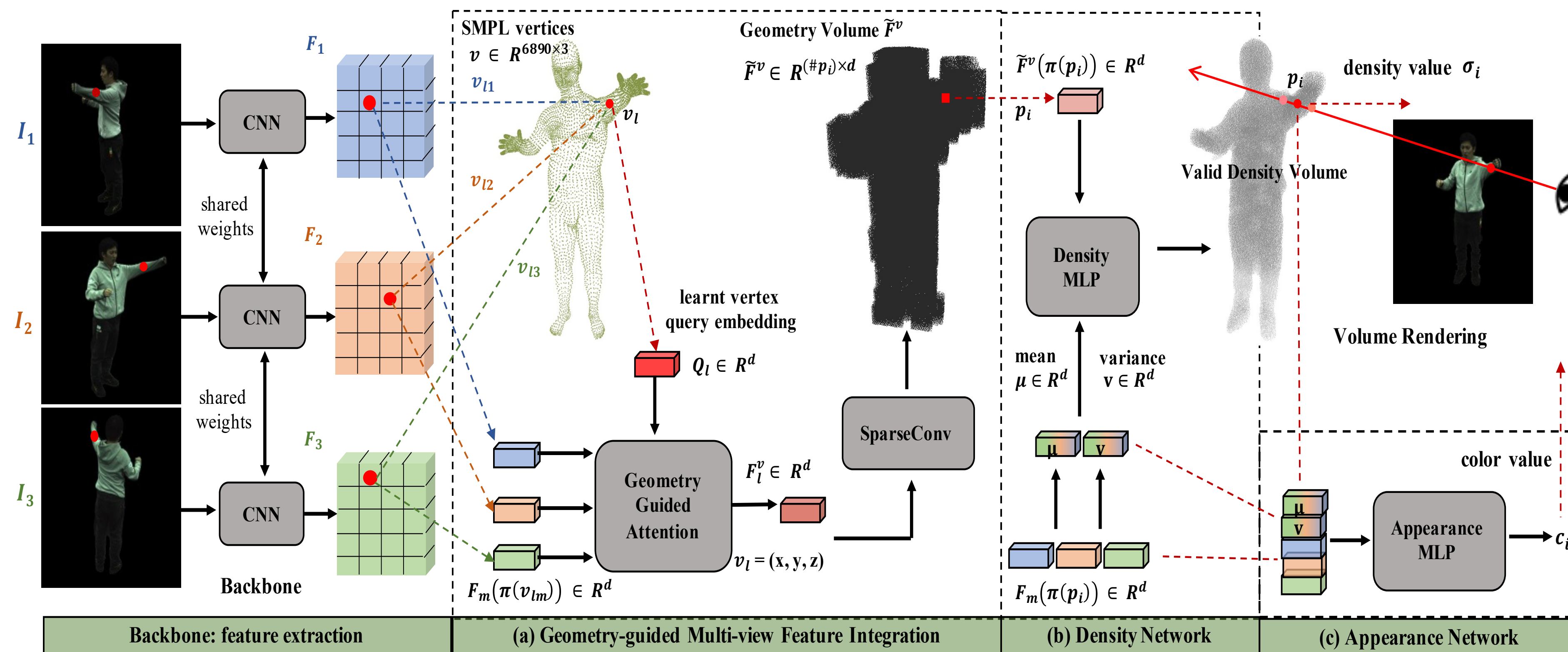
(c) Efficiency Comparison



(b) Rendering pipeline: our efficient geometry-guided progressive pipeline (left) vs. previous (right). The amount of sampling points and forward time in blue are measured on the same data and model parameters.

- The human body is highly non-rigid and commonly has self-occlusions over body parts, which may lead to ambiguous results
- High computational and memory cost of NeRF-based methods severely hinder human synthesis with accurate details in high-resolution.

## Solutions



- Propose a novel geometry-guided progressive NeRF (GP-NeRF) for generalizable and efficient human body rendering, which reduces the computational cost of rendering significantly and also gains higher generalization capacity simply based on the single-frame sparse views.
- Propose an effective geometry-guided multi-view feature integration approach, where we let each view compensate the low-quality occluded information for other views with the guidance of the geometry prior.

## Quantitative Results

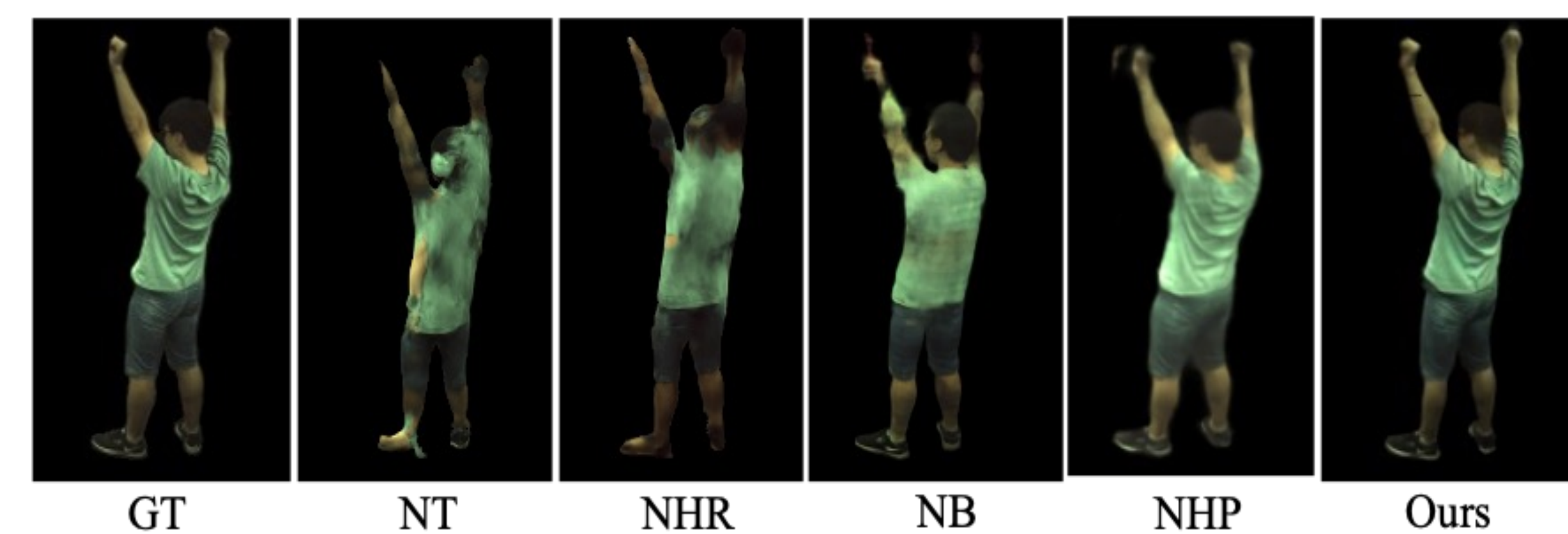
Method	Dataset Train	Test	Per-scene training	Unseen Pose	Unseen Body	Results PSNR (↑)	SSIM (↑)
Performance on training frames							
NT [37]	ZJU-7	ZJU-7	✓	✗	✗	23.86	0.896
NHR [39]	ZJU-7	ZJU-7	✓	✗	✗	23.95	0.897
NB [28]	ZJU-7	ZJU-7	✓	✗	✗	28.51	<b>0.947</b>
NHP [12]	ZJU-7	ZJU-7	✗	✗	✗	28.73	0.936
GP-NeRF (Ours)	ZJU-7	ZJU-7	✗	✗	✗	<b>28.91</b>	0.944
Performance on unseen frames from training data							
NV [19]	ZJU-7	ZJU-7	✓	✓	✗	22.00	0.818
NT [37]	ZJU-7	ZJU-7	✓	✓	✗	22.28	0.872
NHR [39]	ZJU-7	ZJU-7	✓	✓	✗	22.31	0.871
NB [28]	ZJU-7	ZJU-7	✓	✓	✗	23.79	0.887
NHP [12]	ZJU-7	ZJU-7	✗	✓	✗	26.94	0.929
GP-NeRF (Ours)	ZJU-7	ZJU-7	✗	✓	✗	<b>27.92</b>	<b>0.934</b>
Performance on test frames from test data							
NV [19]	ZJU-3	ZJU-3	✓	✓	✗	20.84	0.827
NT [37]	ZJU-3	ZJU-3	✓	✓	✗	21.92	0.873
NHR [39]	ZJU-3	ZJU-3	✓	✓	✗	22.03	0.875
NB [28]	ZJU-3	ZJU-3	✓	✓	✗	22.88	0.880
PVA [30]	ZJU-7	ZJU-3	✗	✓	✓	23.15	0.866
Pixel-NeRF [41]	ZJU-7	ZJU-3	✗	✓	✓	23.17	0.869
NHP [12]	ZJU-7	ZJU-3	✗	✓	✓	24.75	0.906
GP-NeRF (Ours)	ZJU-7	ZJU-3	✗	✓	✓	<b>25.96</b>	<b>0.921</b>
Generalization performance across datasets							
NHP [12]	AIST	ZJU-3	✗	✓	✓	17.05	0.771
GP-NeRF (Ours)	THuman-7	ZJU-3	✗	✓	✓	24.74	0.907
GP-NeRF (Ours)	THuman-all	ZJU-3	✗	✓	✓	<b>25.60</b>	<b>0.917</b>

Method	#r (M) (↓)	#p <sup>d</sup> (M) (↓)	#p <sup>c</sup> (M) (↓)	Time (ms) (↓)	Mem (GB) (↓)
NHP [10]	0.063	4.03	4.03	1160	14.20
NHR [34]	0.063	4.03	4.03	636	10.20
NB [24]	0.063	4.03	4.03	611	21.80
GP-NeRF <sup>+</sup> 3×	0.063 (-0.0%)	4.03 (-0.0%)	4.03 (-0.0%)	589 (-3.6%)	14.53 (-33.3%)
GP-NeRF <sup>+</sup> 2×	0.063 (-0.0%)	4.03 (-0.0%)	4.03 (-0.0%)	567 (-7.2%)	20.74 (-4.9%)
GP-NeRF 2×	<b>0.039 (-38.1%)</b>	<b>0.95 (-76.4%)</b>	<b>0.24 (-94.0%)</b>	243 (-60.2%)	<b>9.88 (-54.7%)</b>
GP-NeRF 1×	<b>0.039 (-38.1%)</b>	<b>0.95 (-76.4%)</b>	<b>0.24 (-94.0%)</b>	<b>175 (-71.4%)</b>	14.25 (-34.6%)

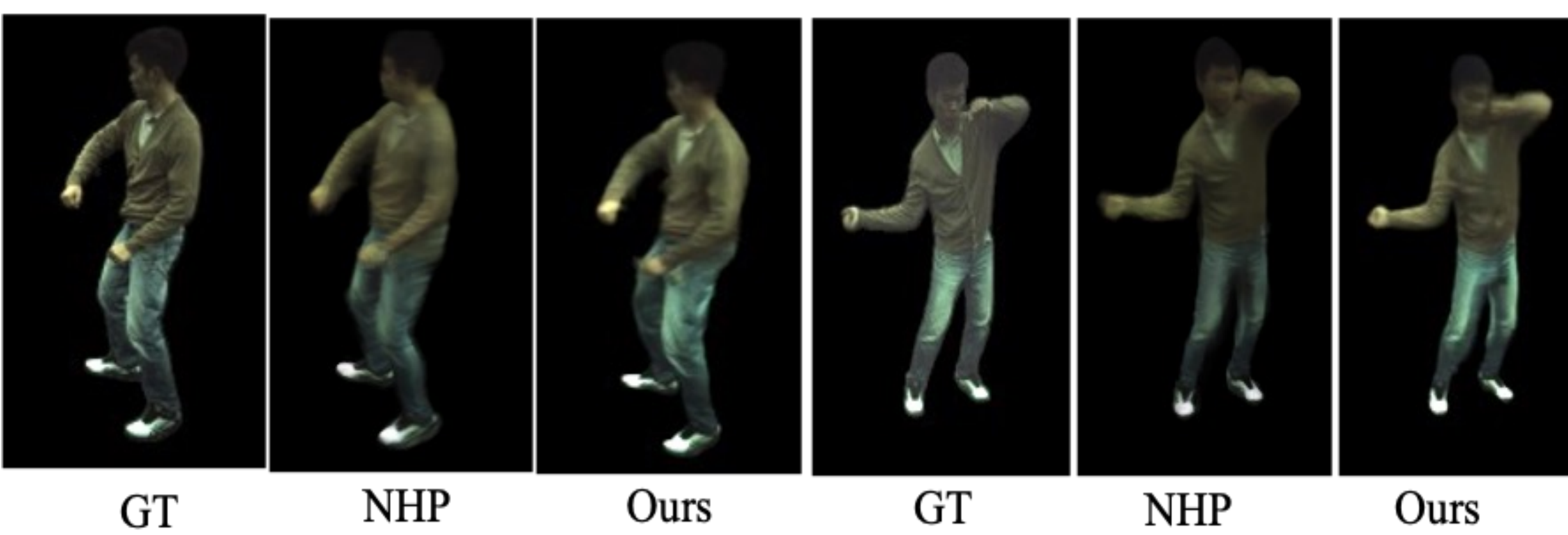
Method	T <sup>d</sup> -MLP (ms) (↓)	T <sup>d</sup> -total (ms) (↓)	T <sup>c</sup> -MLP (ms) (↓)	T <sup>c</sup> -total (ms) (↓)	PSNR (↑)
GP-NeRF <sup>+</sup> 2×	108.58	226.56	145.38	146.39	26.56
GP-NeRF 2×	28.08 (-74.1%)	83.65 (-63.1%)	10.02 (-93.1%)	11.4 (-92.2%)	<b>26.67 (+0.4%)</b>
GP-NeRF 1×	<b>23.55 (-78.3%)</b>	<b>74.07 (-67.3%)</b>	<b>9.50 (-93.5%)</b>	<b>10.27 (-93.0%)</b>	<b>26.67 (+0.4%)</b>

- Our GP-NeRF has achieved state-of-the-art performance on the ZJU-MoCap dataset, taking only 175ms on RTX 3090 and reducing time for rendering per image by over 70.

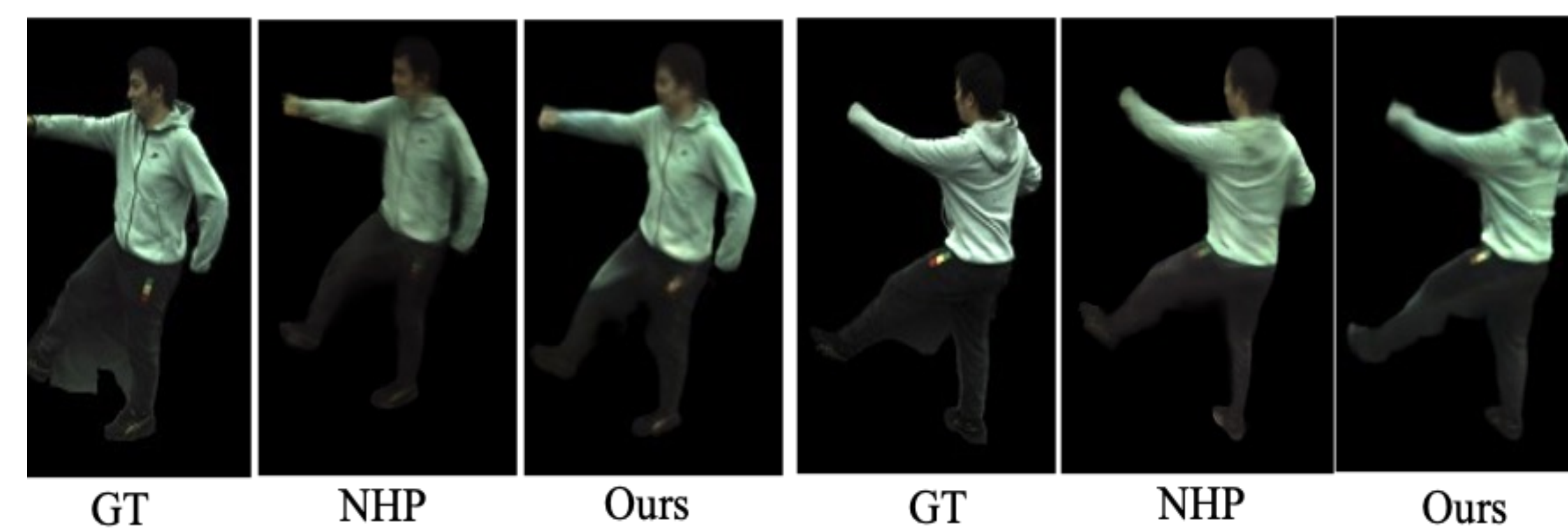
## Qualitative Results



(a) Seen dataset, seen body, unseen pose



(b) Seen dataset, unseen body (human #1)

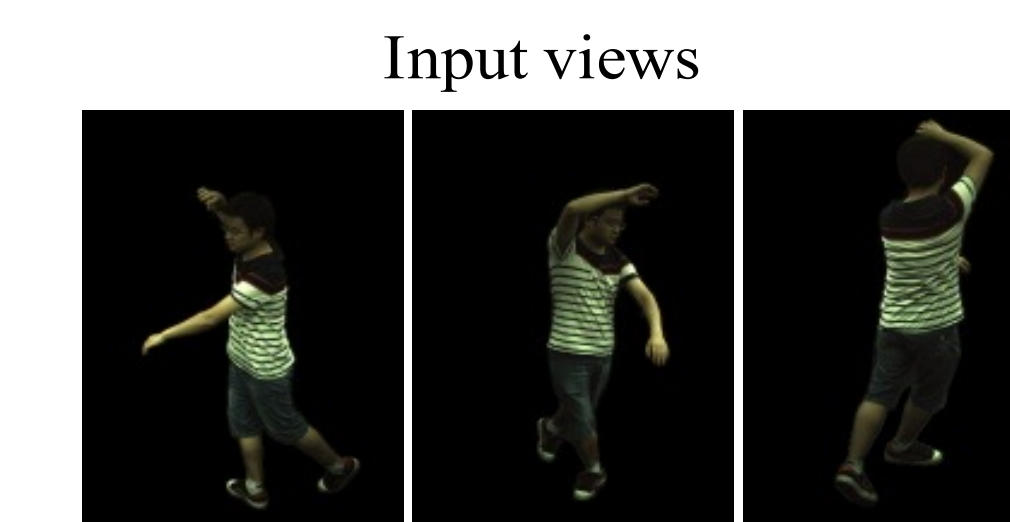


(c) Seen dataset, unseen body (human #2)

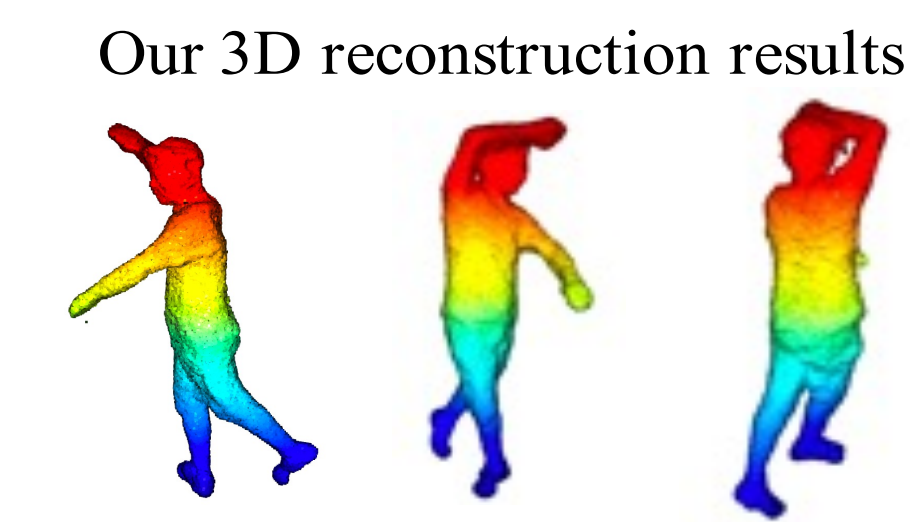


(d) Seen dataset, unseen body on THuman dataset (for each image pair, GT in the left, our results in the right)

## Reconstructed 3D Results

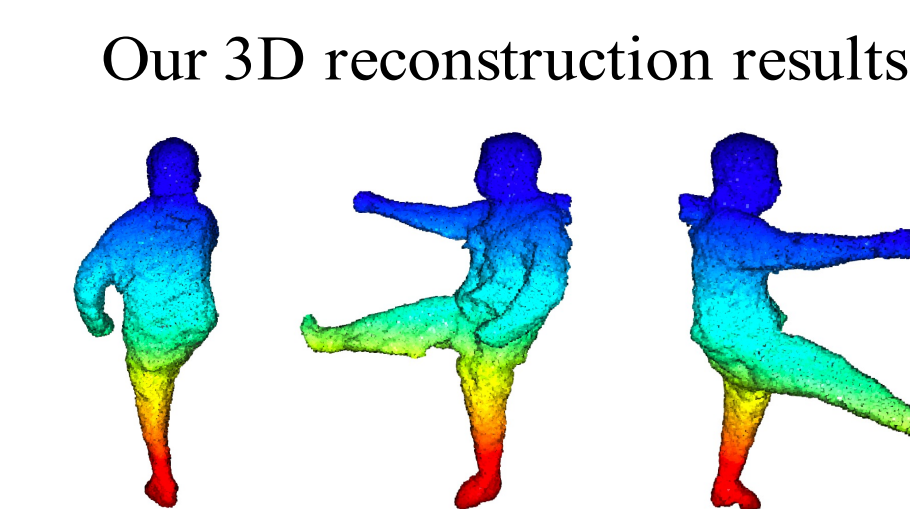
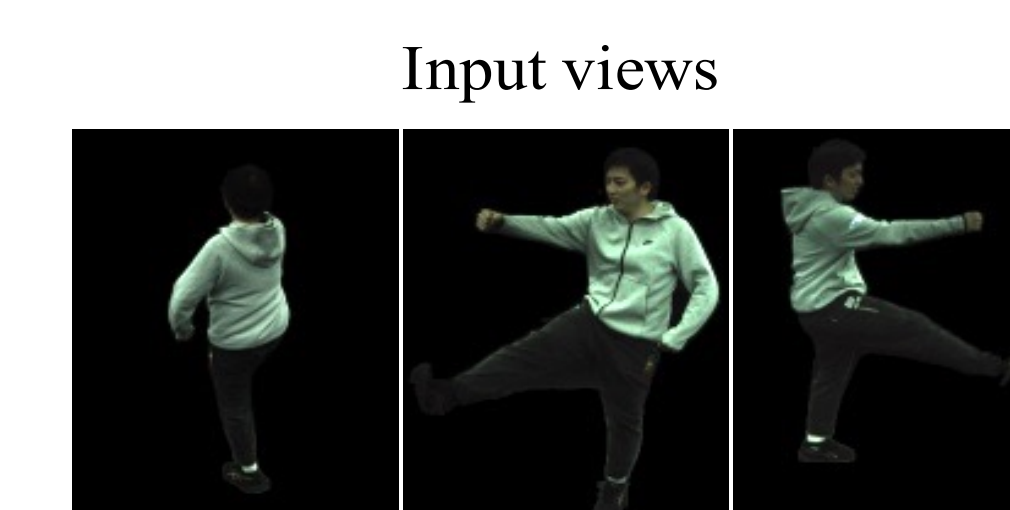


(a) Seen human body



(b) Unseen human body

- Ours can stick to the normal human body geometry better than methods without geometry priors and can reconstruct more accurate lighting conditions.



- Our synthesis can reconstruct very close human body shape and clothes details like hoods and folds on unseen human bodies.

PIFuHD reconstruction results