ICLR

THE UNIVERSITY OF BRITISH COLUMBIA

VECTOR INSTITUTE | INSTITUT VECTEUR

LINKÖPING UNIVERSITY

UNIVERSITÄT BIELEFELD

Locality Sensitive Avatars From Video

Chunjin Song[1]*, Zhijie Wu[1]*, Shih-Yang Su[1], Bastian Wandt[3], Leonid Sigal[1,2], Helge Rhodin[1,4]

(* Equal Contribution)

[1]University of British Columbia [2]Vector Institute for AI [3]Linköping University [4]Bielefeld University

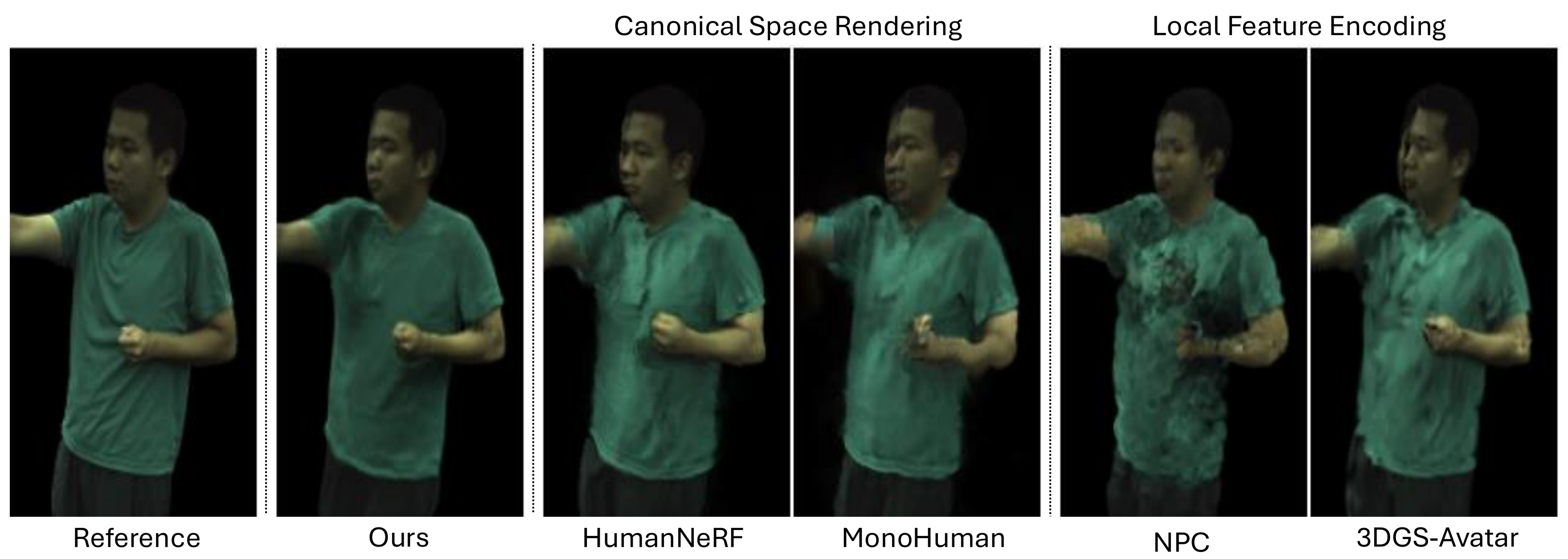
Github

Goal

We aim to reconstruct human avatars with realistic body shapes and vivid fine-grained details from monocular video.

Our goal is to improve performance across novel view synthesis, unseen pose rendering, and accurate shape reconstruction simultaneously.

Motivation

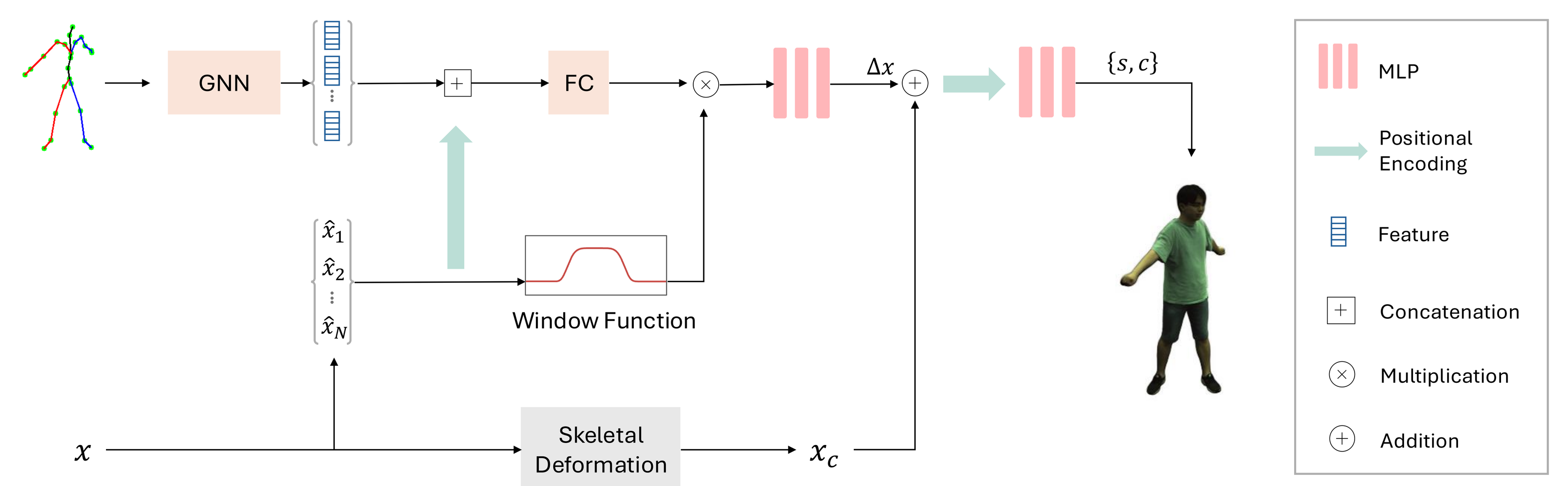


Stable without Details: Canonical space rendering provides stable avatars but struggles with fine-grained details, leading to distortions.

Detailed but Fragile: Local feature encoding captures fine details but is sensitive to sparse input, causing artifacts.

Our Approach: We learn a canonical representation via a non-linear mapping, decomposed into skeletal motion and locality-aware non-rigid counterpart to retain details.

Method



We apply inverse Linear Blend Skinning (LBS) to map a query point x from observation space to canonical space as x_c for rigid deformation modeling.

We use a Graph Neural Network (GNN) to capture body part correlations from the input pose skeleton.

We generate point-dependent offsets Δx in canonical space as the non-rigid deformation component, using features that encode local pose context.

Summary

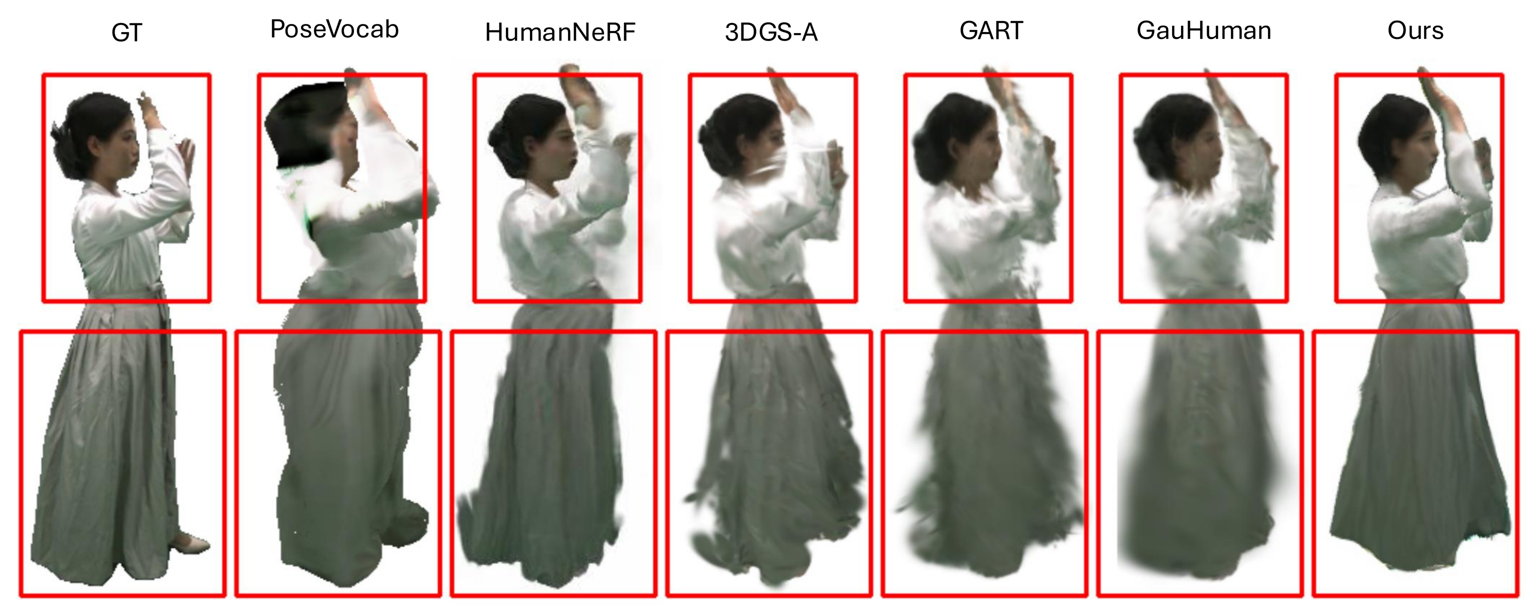
We propose a locality-sensitive avatar representation by integrating part-based locality encoding with canonical space rendering.

We use GNNs to model local context and estimate non-rigid offsets for adaptive detail generation.

Our method outperforms state-of-the-art in novel view synthesis, unseen pose rendering, and shape reconstruction simultaneously.

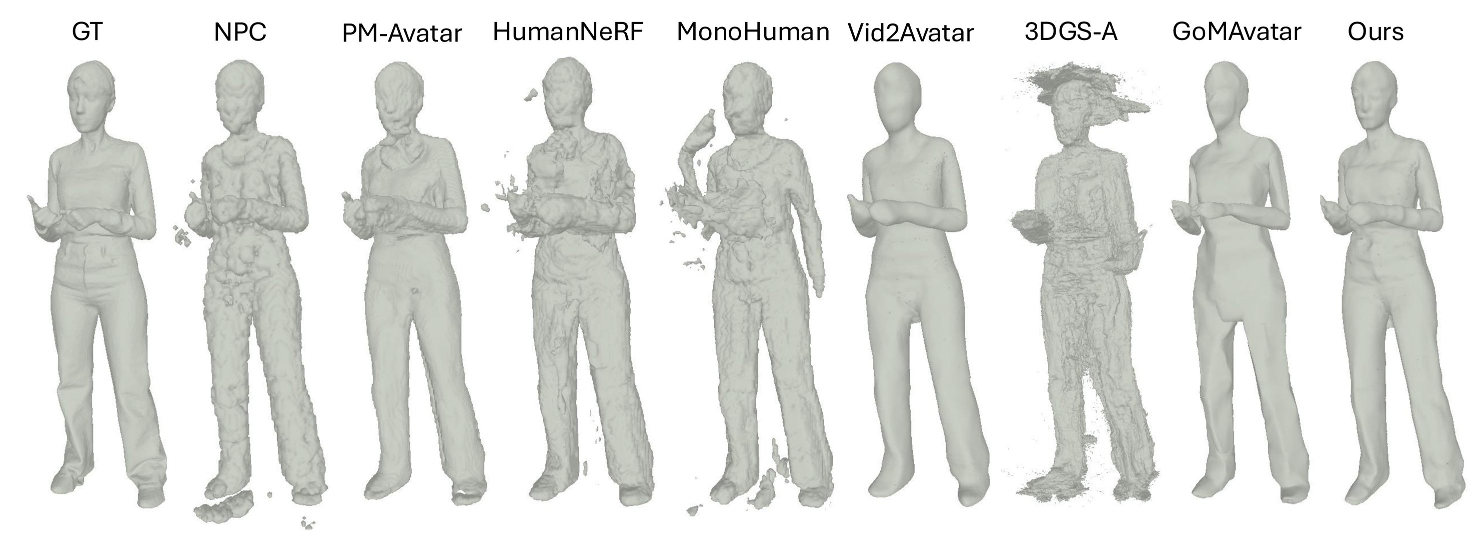
Results

Novel View Synthesis



Novel Pose Rendering

Geometry Results



Quantitative Comparisons

	ZJU-Mocap (Novel view)				ZJU-Mocap (Novel pose)				ZJU-Mocap (Shape)		SynWild (Shape)	
	PSNR↑	LPIPS↓	FID↓	KID↓	PSNR↑	LPIPS↓	FID↓	KID↓	CD ↓	NC ↑	CD ↓	NC ↑
HumanNeRF	29.66	36.78	28.35	14.23	29.57	34.17	42.84	12.32	0.051	0.765	0.507	0.635
MonoHuman	30.18	31.45	27.88	13.18	29.90	32.21	43.65	12.61	0.065	0.737	N/A	N/A
NPC	30.01	37.18	60.39	53.24	29.61	36.52	73.98	49.79	0.061	0.762	0.503	0.615
Vid2Avatar	29.76	35.61	36.83	27.65	29.53	35.69	54.16	31.51	0.042	0.852	0.499	0.687
PM-Avatar	30.24	38.38	49.58	39.64	29.87	39.26	65.71	40.16	0.051	0.766	0.500	0.632
3DGS-Avatar	30.09	31.30	29.61	15.33	29.77	30.69	43.23	13.24	0.079	0.695	0.553	0.560
GoMAvatar	30.29	32.40	26.90	12.80	30.20	32.03	43.09	13.81	0.044	0.820	0.567	0.661
Ours	30.25	28.36	22.13	8.26	30.18	27.84	35.22	7.16	0.041	0.845	0.485	0.690