AIIInOneAvatar:

An All-Inclusive, End-to-End Framework for Rapid Avatar Creation from Monocular Videos

Boyu Zhang

Senyang Jiang

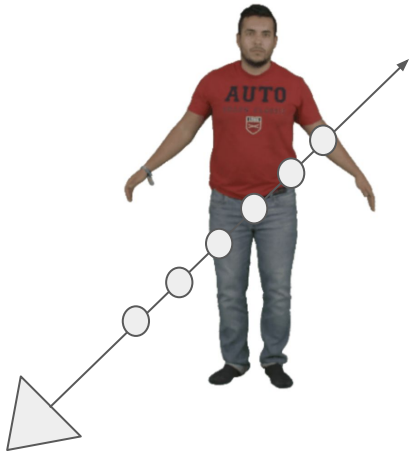
Yining Mao

System Input and Output

Monocular Video

Input

NeRF Representation



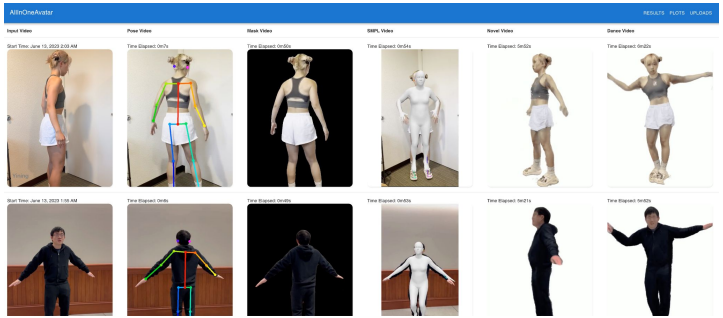
Output

Novel View

Novel Animation

Goal and Constraints

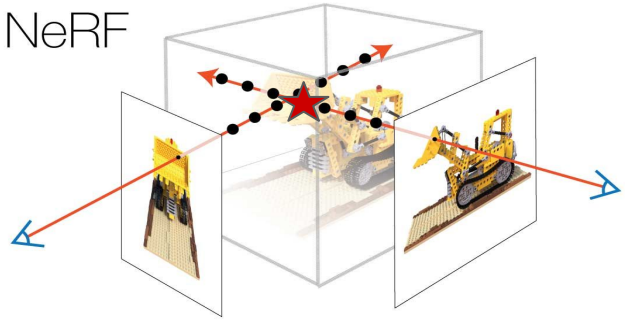
1. End to end web application generates novel view and novel animation videos in **5 mins**!



2. Input is only a **monocular** video

Demo!

Nerf to HumanNeRF



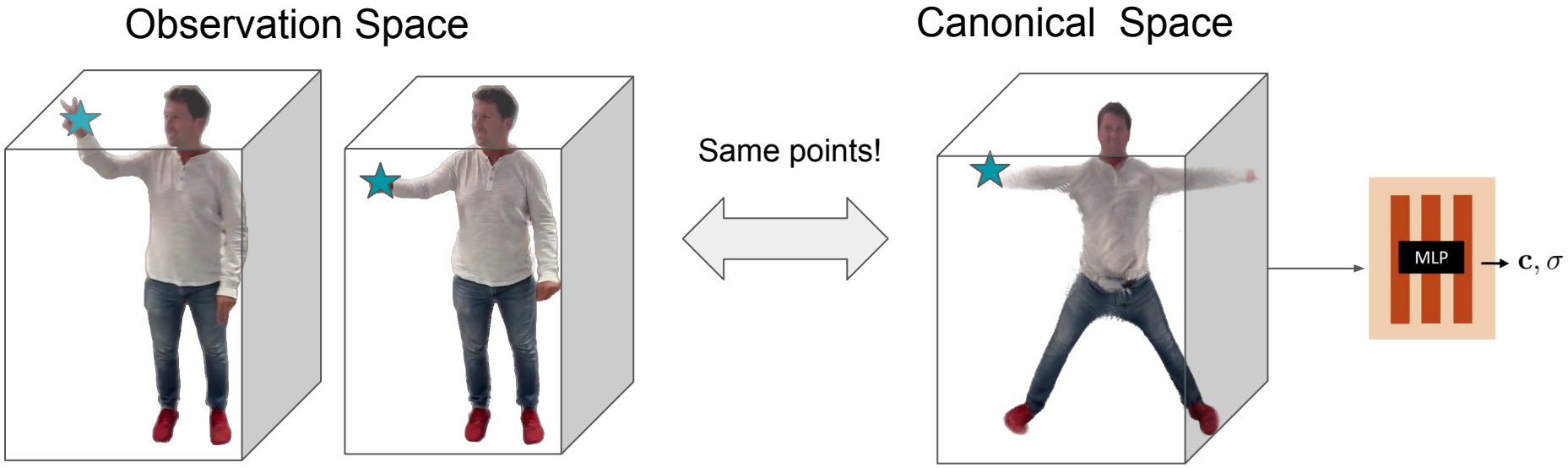
Color and density at the same sample point is **consistent**

HumanNeRF

People **move** in the input video!

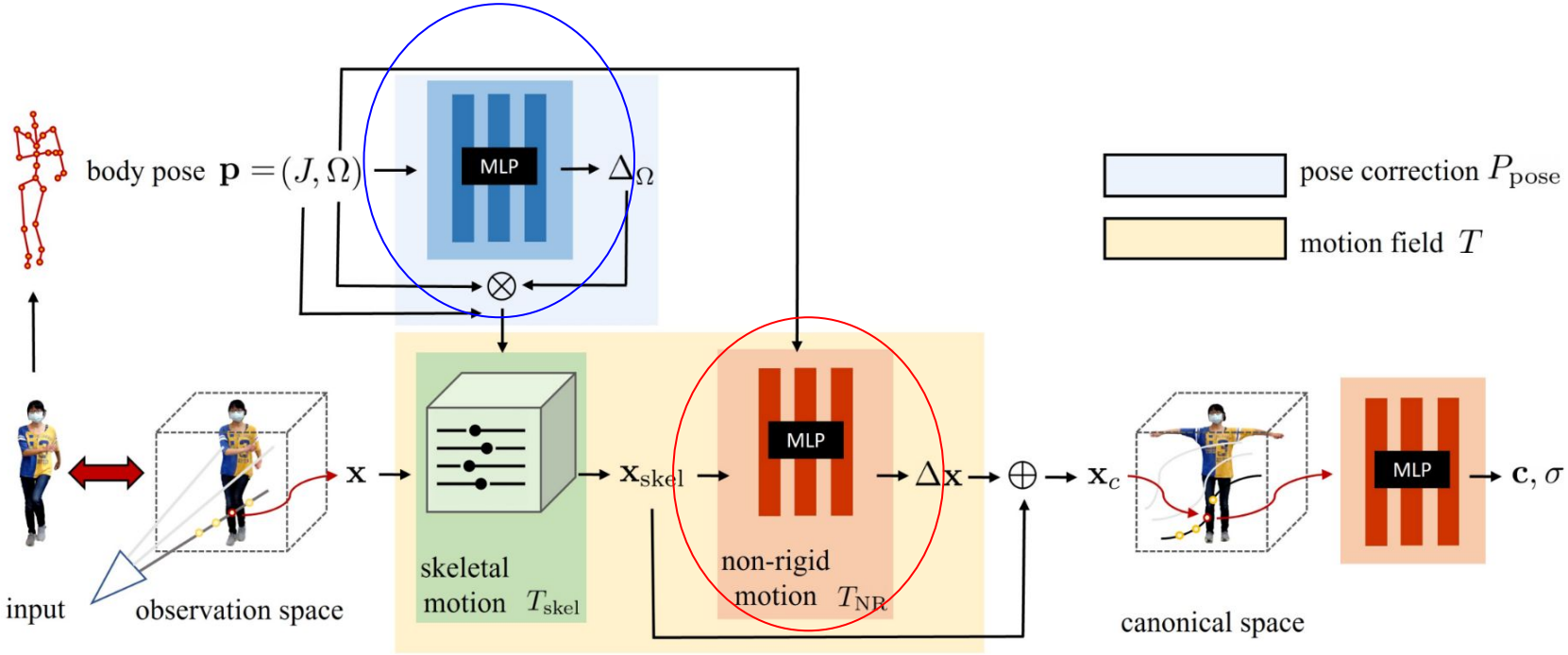
Color and density at the same sample point is **inconsistent**

Sample points mapping is important

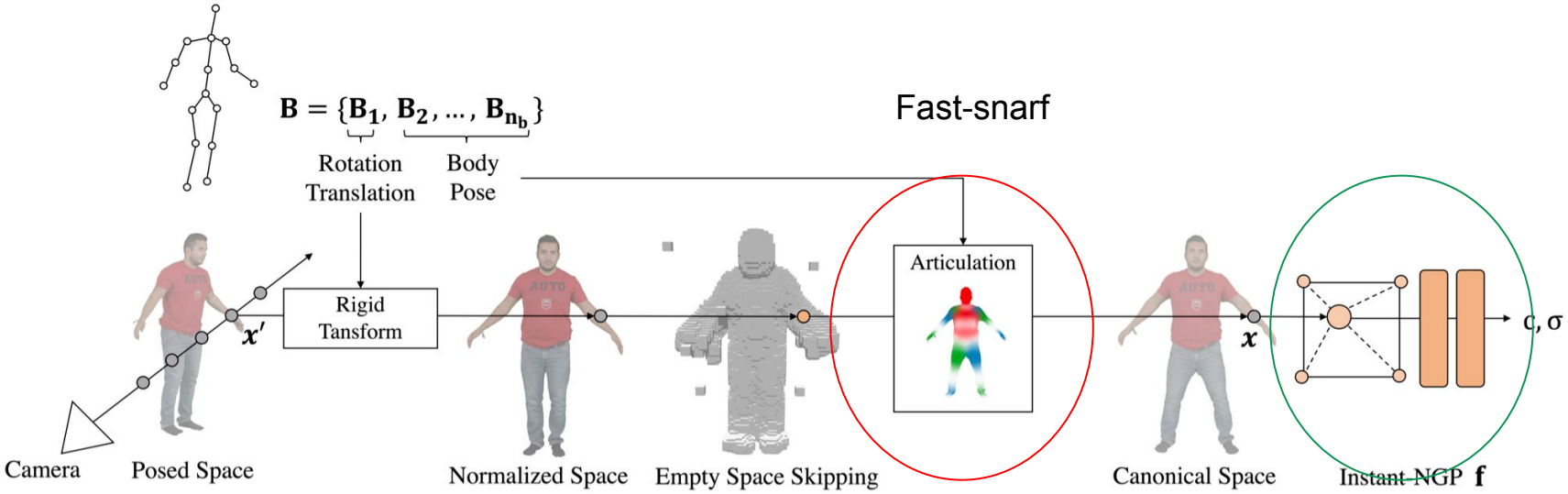


Original HumanNeRF

Three **MLPs**?



InstantAvatar Pipeline



End-to-End System

Monocular Video

fit (SMPL)

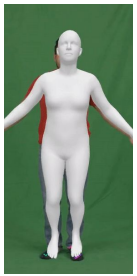
train (SNARF)

Input

Preprocessing

Two-Stage Training

Inference



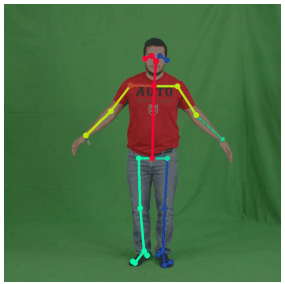
Preprocessing

Pose estimation + Body segmentation + SMPL fitting

Selected frames

LIVE

Video



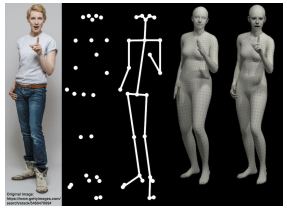
Pose estimation

OpenPose

Parallel tasks

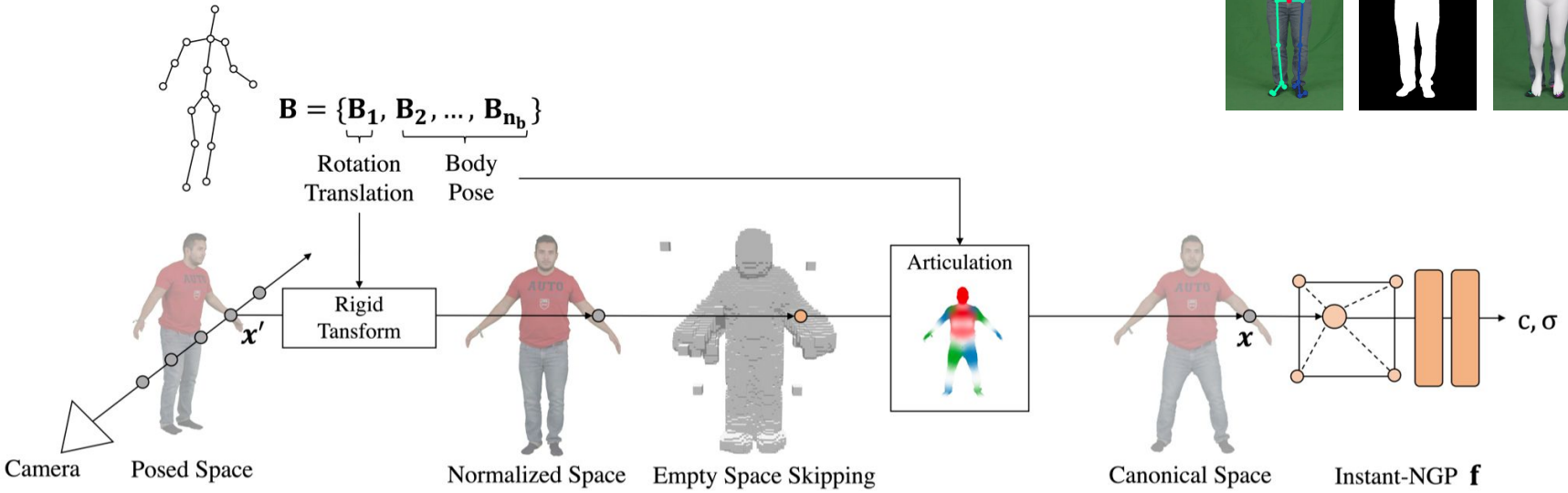
Body segmentation

ultralytics YOLOv5

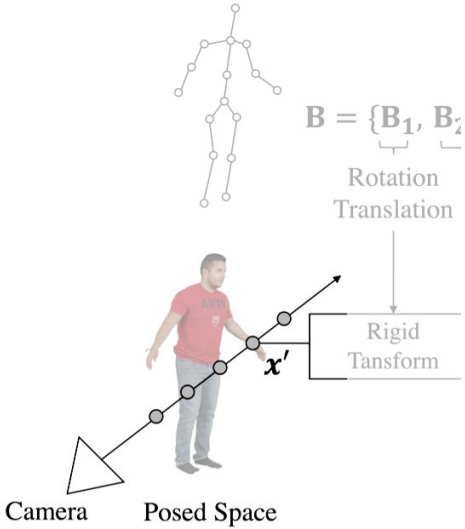


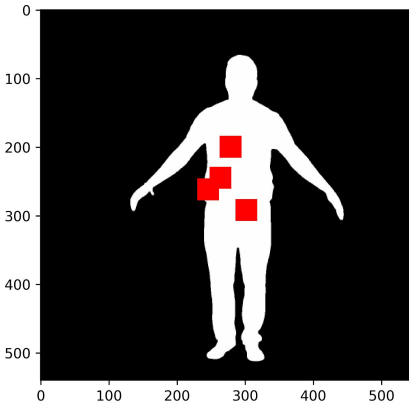
SMPL fitting (multi-thread)

Human NeRF pipeline

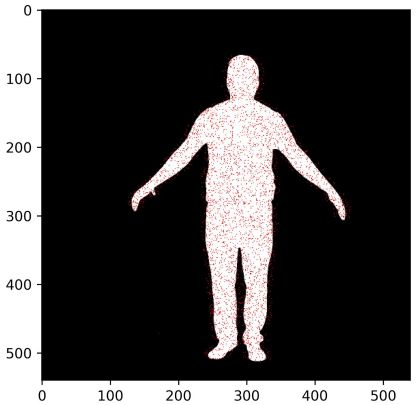


Ray Marching and Sampling

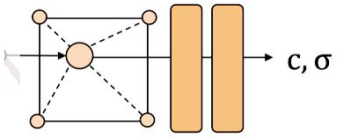




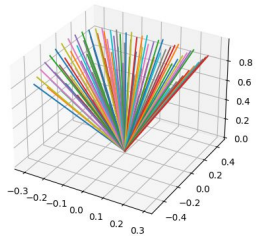
Patch Sampler



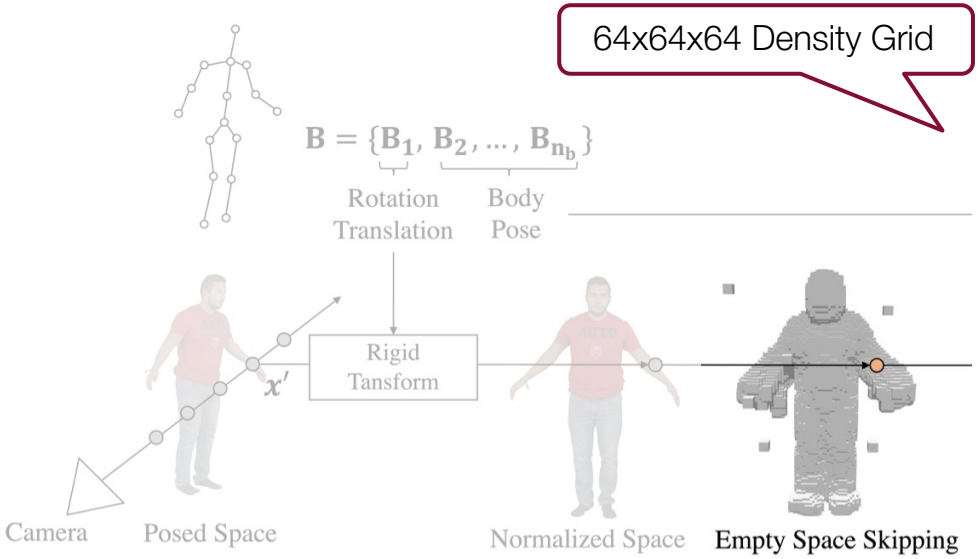
Edge Sampler

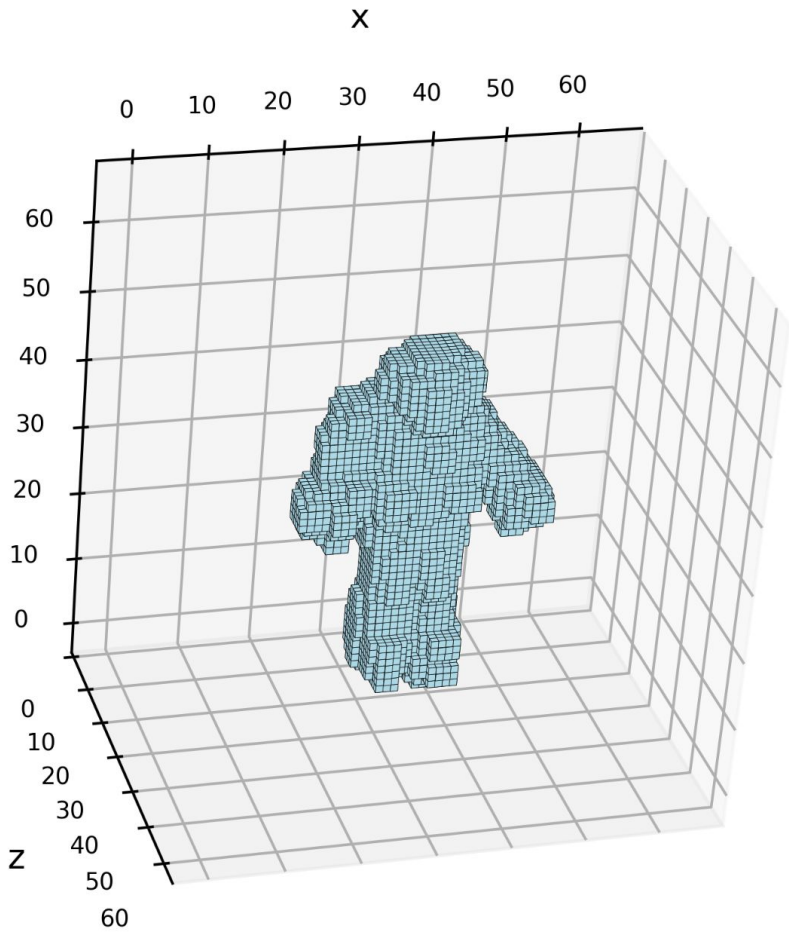


Instant-NGP **f**



Empty Space Skipping

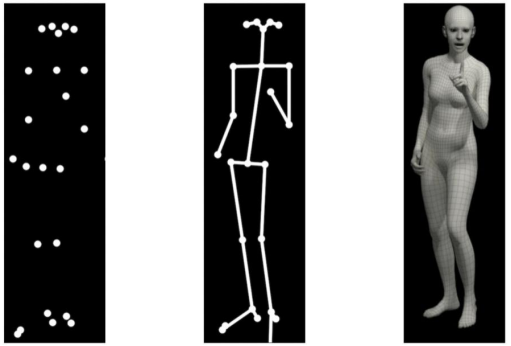


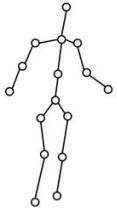


Mapping to Canonical Space

SMPL

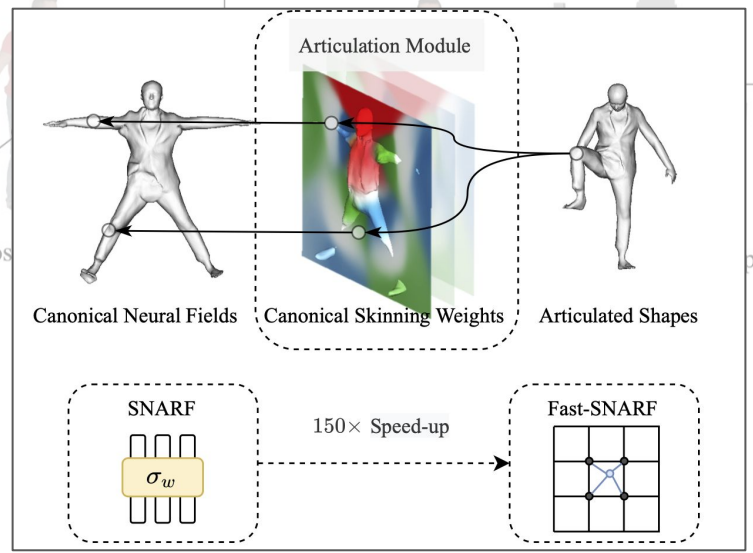
Human mesh model

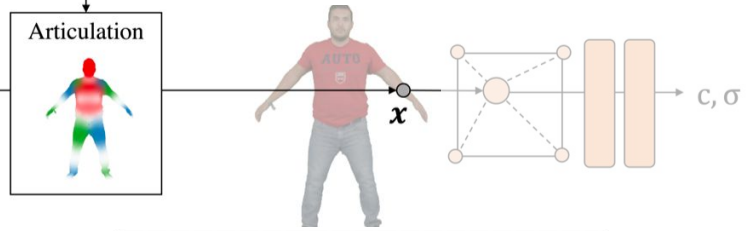


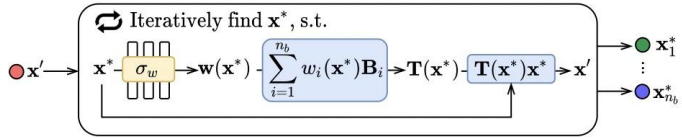


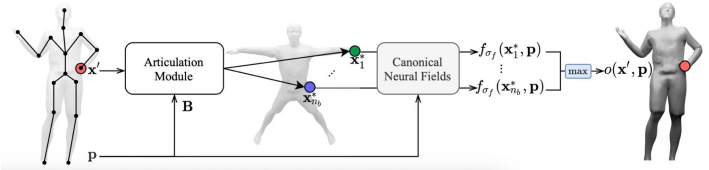
$$\mathbf{B} = \{\underbrace{\mathbf{B_1}}_{\text{Rotation Translation}}, \underbrace{\mathbf{B_2}, \ldots, \mathbf{B_{n_b}}}_{\text{Body Pose}}\}$$

Fast SNARF



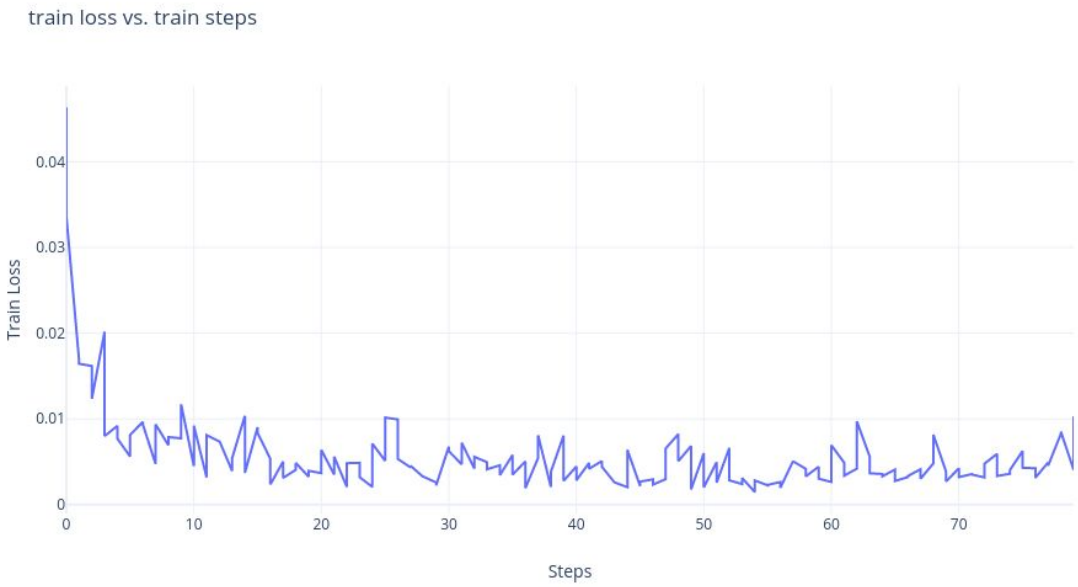


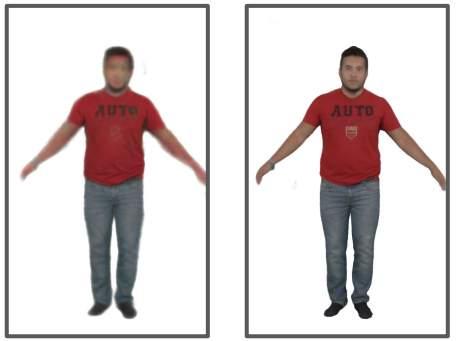




Experiments & Improvements

Training Strategy: Early Stopping



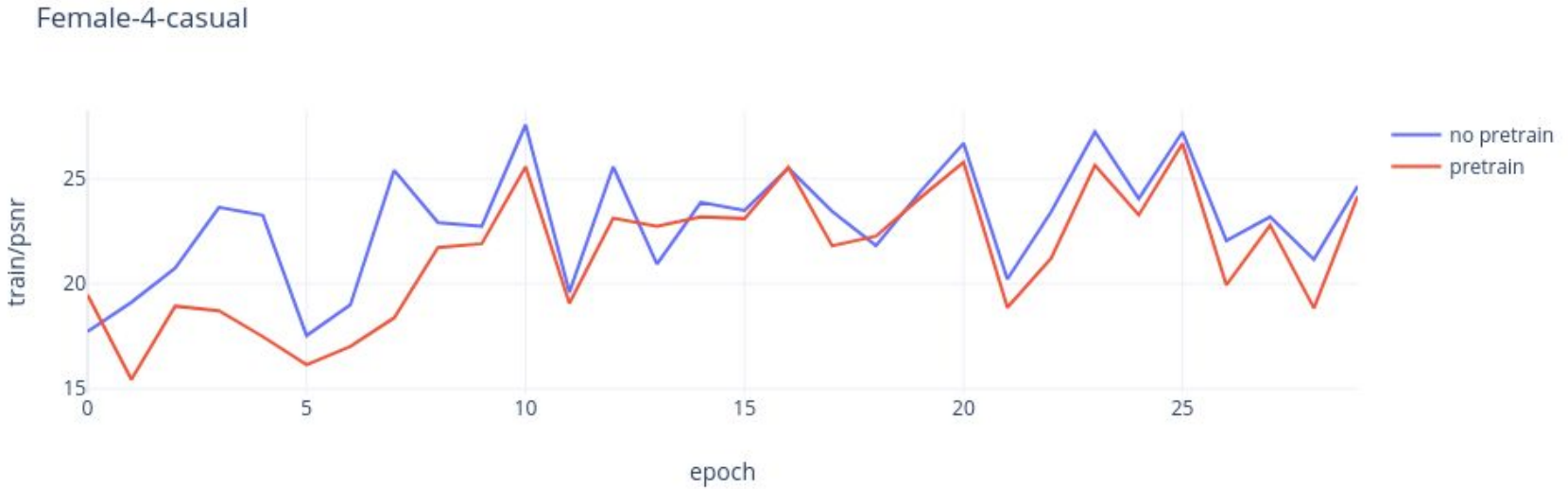


fit (SMPL) train (SNARF)

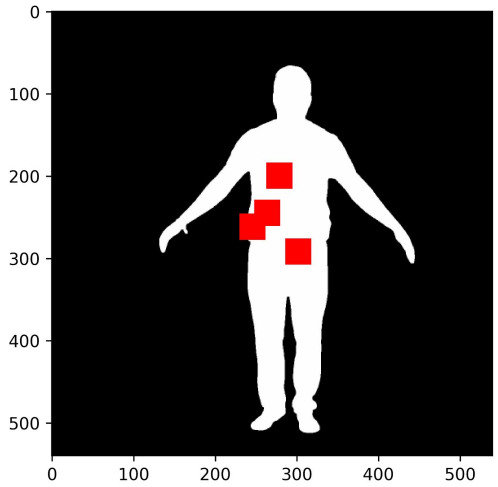
Early Stopping works pretty well

Stopping Criteria: No improvement in train/psnr in 5 train epochs										
	female-4-casual					male-4-casual				
	Train Epochs	Train Time	PSNR	SSIM	LPIPS	Train Epochs	Train Time	PSNR	SSIM	LPIPS
InstantAvatar	80	2m55.5s	23.57	0.9397	0.03632	80	2m10.5s	22.88	0.9357	0.05309
InstantAvatar + Early Stopping	12	0m27.3s	23.59	0.9379	0.04600	16	0m34.5s	23.12	0.9355	0.06092

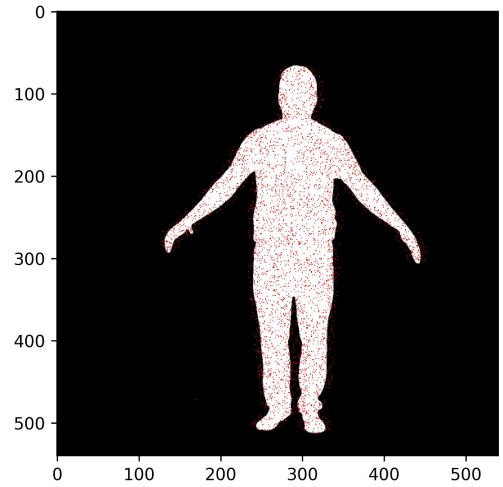
Pretraining doesn’t help



Patch Sampler is better than Edge Sampler



Patch Sampler



Edge Sampler

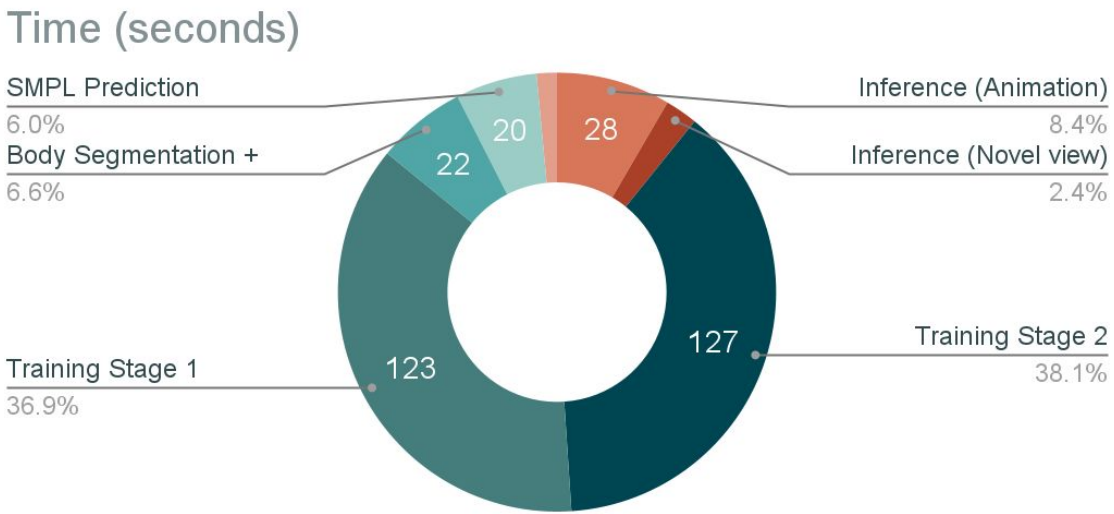
SMPL performs similar to Fast-SNARF

Table: SMPL vs. Fast-SNARF										
	female-4-casual					male-4-casual				
	Train Epochs	Train Time	PSNR	SSIM	LPIPS	Train Epochs	Train Time	PSNR	SSIM	LPIPS
SMPL	80	2m6.74s	23.54	0.9392	0.03543	80	2m27.1s	22.88	0.9351	0.04815
Fast-SNARF	80	2m5.9s	23.56	0.9396	0.03630	80	2m28.4s	22.97	0.9357	0.05102

Time Analysis

Average training time: ~120s for each stage
Average training epoch number (early stopping): ~55 epochs for each stage
Total time: ~5mins

Parallel



Thank you

Boyu Zhang

Senyang Jiang

Yining Mao