HVTR: Hybrid Volumetric-Textural Rendering for Human Avatars

Tao Hu[1], Tao Yu[2], Zerong Zheng[2], He Zhang[2], Yebin Liu[2], Matthias Zwicker[1]

[1]University of Maryland, College Park [2]Tsinghua University

International Conference on 3D Vision
September 12-15, 2022
Hybrid: Prague, Czechia / Online
3DV 2022

Goal & Motivation

Goal: Learn 3D human avatars from videos, which can be rendered under arbitrary poses and viewpoints **efficiently** and at **high quality**.

Motivation: We propose a hybrid rendering method, which learns geometry reconstructions by optimizing a pose-conditioned downsampled NeRF, and uses a GAN for efficient rendering.

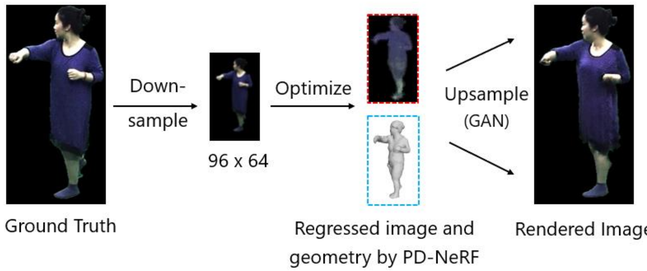
Representations	Renderer	Efficient Rendering	Geometry Recons
2D Plus (e.g., SMPL + DNR [1])	GAN	✓	✗
3D (e.g., Neural Body [2])	Volume Rendering	✗	✓
3D (ours)	**Hybrid**	✓	✓

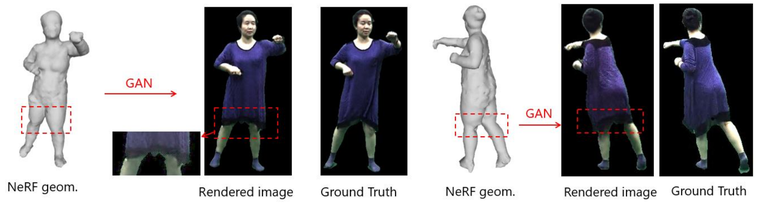
Experiments & Results

Methods	LPIPS ↓	FID ↓	Time (FPS) in inference
SMPL + DNR [1]	.113	85.752	5.4
Neural Body [2]	.210	149.924	0.05
Ours	**.100**	**72.142**	2.9

Render Loose Clothing

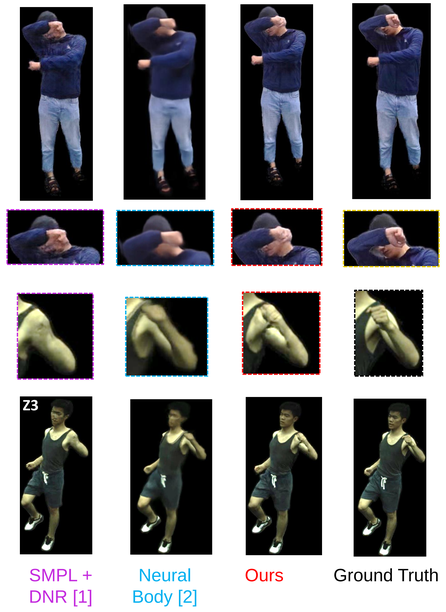
- We learn rough geometry (PD-NeRF) from low-resolution images.
- Yet our method can render loose clothing with the rough geometry.





Solve Self-Occlusions

- Our method can handle boundaries, solve occlusions, and render high-quality images.

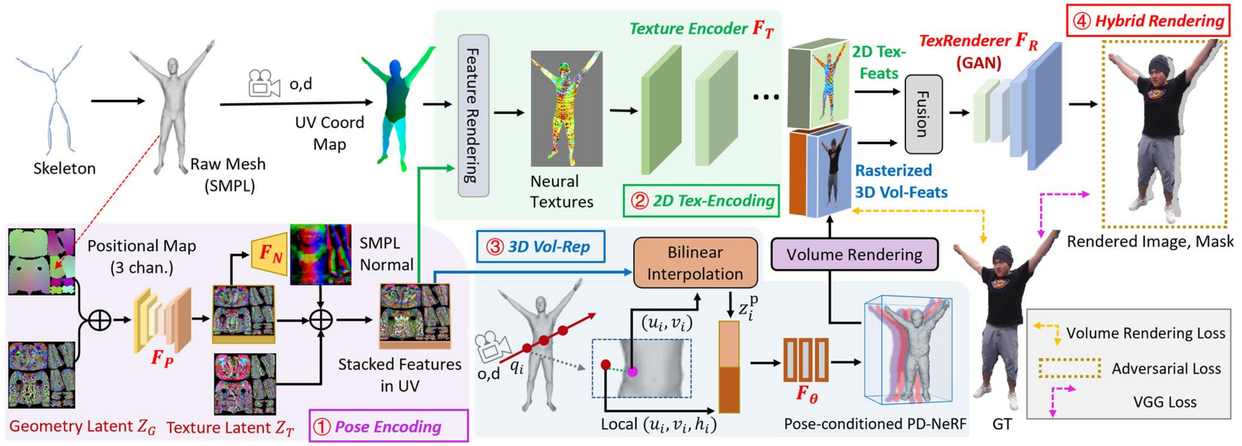


[1] Thies et al. Deferred Neural Rendering. TOG 2019.
[2] Peng et al. Neural Body. CVPR 2021.

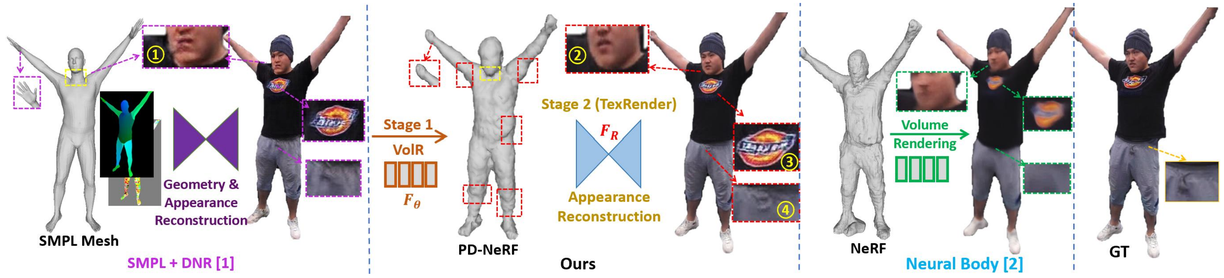
Our Approach

A novel neural rendering pipeline, Hybrid Volumetric-Textural Rendering (HVTR), which has four components:

- **Pose Encoding in UV space**: Parameterize poses on the UV manifold of human body surface.
- **2D Textural Encoding**: Transform the features from UV space to 2D image space for rendering.
- **3D Volumetric Representation**: To handle self-occlusions, we learn a pose-conditioned downsampled NeRF.
- **Hybrid Rendering**: Get 3D vol-features by volume rendering, which are then up-sampled to avatars by GAN.



Differences from Existing Methods



- DNR [1] uses a GAN for one-stage rendering, which suffers from artifacts (closeup ①) due to geometric misalignments.
- Our method works at two stages by first constructing a PD-NeRF to handle geometric alignments, and then utilizing a GAN for appearance synthesis. We can handle self-occlusions (②③④ vs. ①) better than DNR, and preserve more details (③④) than Neural Body [2].