S^3: Neural Shape, Skeleton, and Skinning Fields for 3D Human Modeling

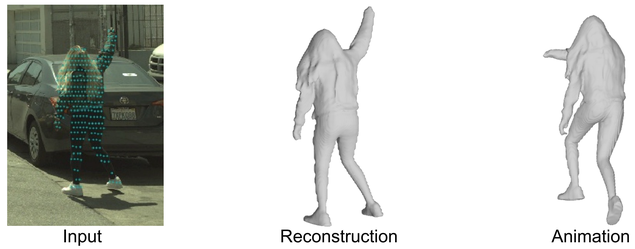
Ze Yang, Shenlong Wang, Sivabalan Manivasagam, Zeng Huang,
Wei-Chiu Ma, Xinchen Yan, Ersin Yumer, Raquel Urtasun

Uber ATG · University of Toronto · USC · MIT

CVPR VIRTUAL JUNE 19-25

Introduction

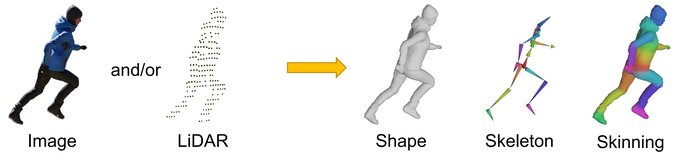
- Constructing and animating humans is an important task and has a wide variety of applications



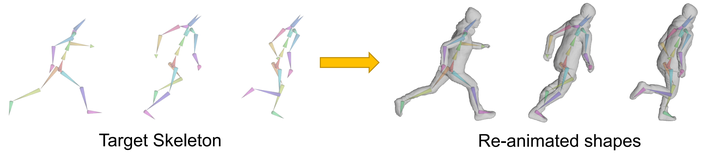
- Automating the human reconstruction and animation is very challenging due to large variations in shape, pose, clothing
- Traditional methods are time-consuming and do not scale
- Existing methods rely on controlled environment and are prone to sensor noise

Overview

- We use a unified neural field to jointly learn the shape, skeleton and skinning for 3D human in an end-to-end fashion
- We explore multi sensor setting captured at a single viewpoint



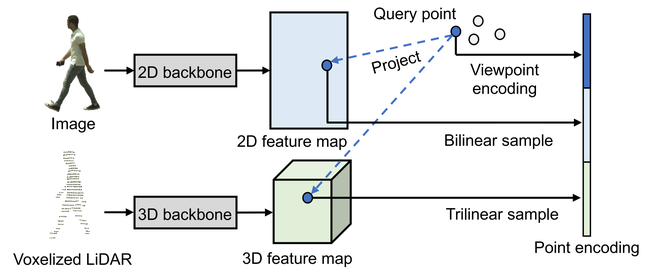
- Our model is not constrained by a parametric model with a fixed (mesh) topology
- The resulting model can be directly deformed to novel poses using either motion capture data or artist-created animations



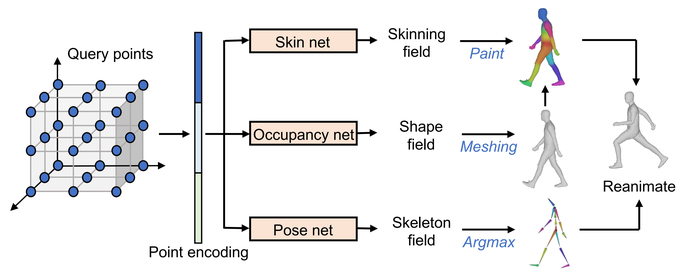
Method

We extract **Neural S^3 fields** to generate human animation model.

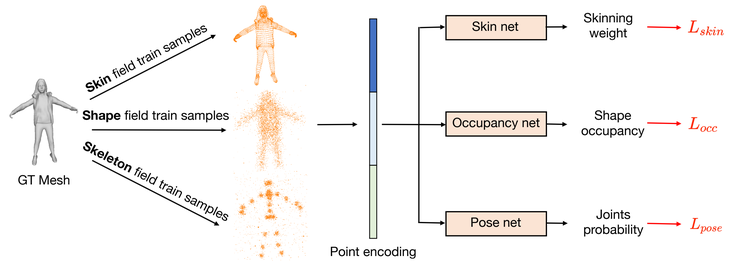
- **Multi-sensor feature:** We process a single camera image and a sparse point cloud to obtain the sensor feature maps
- **Extract point encoding:** Given an arbitrary query point, we interpolate its local features and fuse them with viewpoint encoding



- **Neural S^3 field:** We use MLPs to predict skinning weight, shape occupancy and skeleton joints probability
- **Animatable model extraction:** We post-process the neural S^3 field to obtain the mesh, skeleton and skinning weights

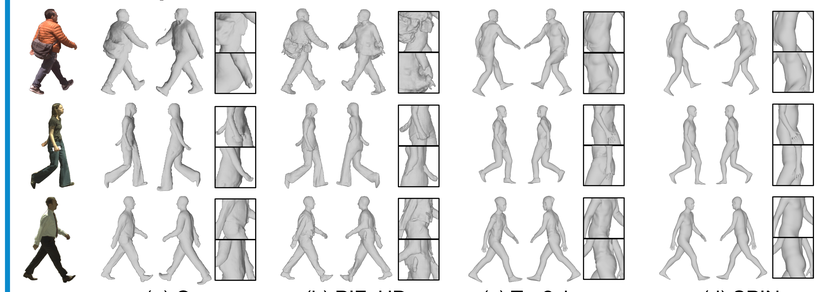


To learn the **neural S^3 field** efficiently, we adaptively sample query points during training.



Experiments

Human Shape Reconstruction: Real-world data from urban scenes.



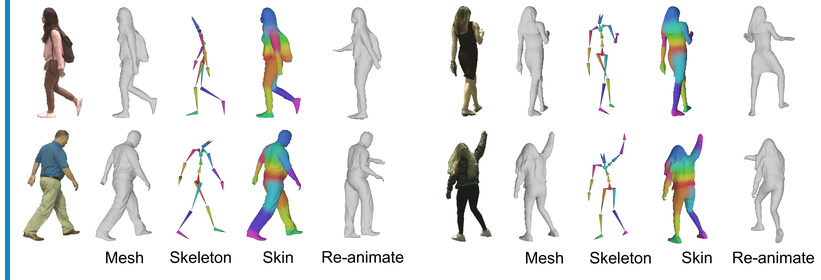
Quantitative comparisons on RenderPeople dataset.

	Seen Poses (CD) ↓	Unseen Poses (CD) ↓
SPIN*	3.25	3.26
PIFu	1.03	1.61
Ours (Image)	0.92	1.58
Ours (Image+LiDAR)	**0.66**	**0.76**

*Denotes model from official repo: https://github.com/nkolot/SPIN

Project Page

Human Animation: The reconstructed mesh, skeleton and skinning weights from real data, which can be used for re-animation.



Motion Retarget: We compute Chamfer distance between our retarget shape and GT shape to measure the overall performance of our model.