Learning Implicit Templates for Point-Based Clothed Human Modeling

Siyou Lin, Hongwen Zhang, Zerong Zheng, Ruizhi Shao, Yebin Liu
Department of Automation, Tsinghua University

INTRODUCTION & MOTIVATION

Our Goal:
Learning to model animatable humans in diverse clothing with high-fidelity pose-dependent details from unregistered scans.

Major challenges of this task:
- Learning the clothing topology;
- Learning the pose-dependent deformations with fine details.

Existing clothed human representations:

Type	SOTA	Pros	Cons
Mesh	CAPE [1]	• Efficient • Compatible with the rendering pipeline	• Fixed topology • Supports only tight clothing or requires clothing-specific templates
Point cloud	POP [2]	• Efficient • Topologically Flexible	• Requires a base-template (usually SMPL [5]), leading to nonuniform distribution for loose clothing
Implicit field	SNARF [3] SCANimate [4]	• Topologically Flexible	• Computationally heavy • Less details

Our motivation:
- Incorporate the merits of *implicit* and *explicit* representations using a ***First-Implicit-Then-Explicit (FITE)*** two-stage pipeline.

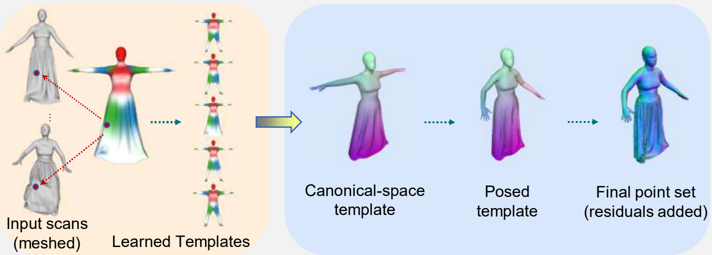
KEY IDEAS

Method overview:
- A two-stage pipeline that involves both implicit and explicit modeling.
- *Stage 1:* Learn implicit templates for different types of clothing.
- *Stage 2:* Learn pose-dependent deformations based on the learned implicit templates using a point-based representation.

Benefits:
Stage 1:
- captures the coarse topology of different types of clothing;
- avoids artifacts brought by a minimal body model for loose clothing.

Stage 2:
- models details with a point-based representation;
- Achieves higher geometric quality than purely implicit methods.

Our pipeline:



Stage one: Template training for obtaining coarse topology

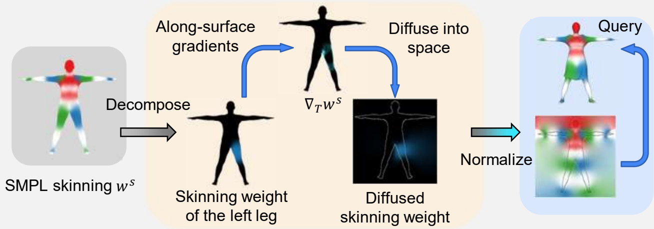
Stage two: Modeling pose-dependent deformations with a point-based representation

METHOD DETAILS: STAGE 1

Task:
Learning canonical-space implicit templates for different types of clothing.

Challenges:
- The canonical-to-posed (C2P) mapping is a many-to-one mapping.
- There is no well-defined skinning weights far from the SMPL surface.
- kNN-based/learned skinning often leads to discontinuity/local minima.

Key contribution (diffused skinning):
A smooth skinning field diffused from the SMPL [5] surface.

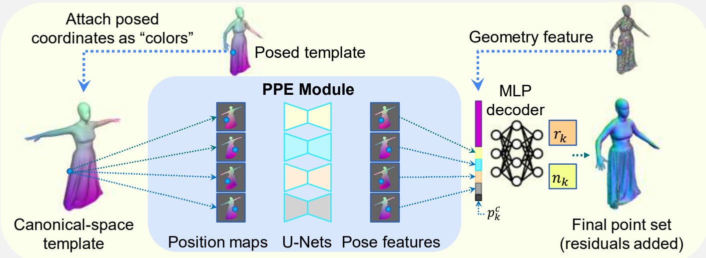


METHOD DETAILS: STAGE 2

Task:
Learning pose-dependent deformations for previously learned templates.

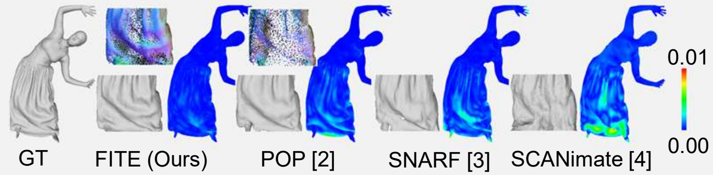
Challenges:
- Encode pose information into the learned templates.
- The learned templates have no predefined UV map. UV encoding as done in POP [2] is not applicable.

Key contributions (projection-based pose encoding, PPE):
- Attach the posed vertex coordinates as features to the canonical template;
- Encode these features with 2D U-Nets by rendering the canonical template above with multiview projections.
- The features encoded by the 2D U-Nets are decoded to residuals and normals that produce the final pose-dependent deformations.

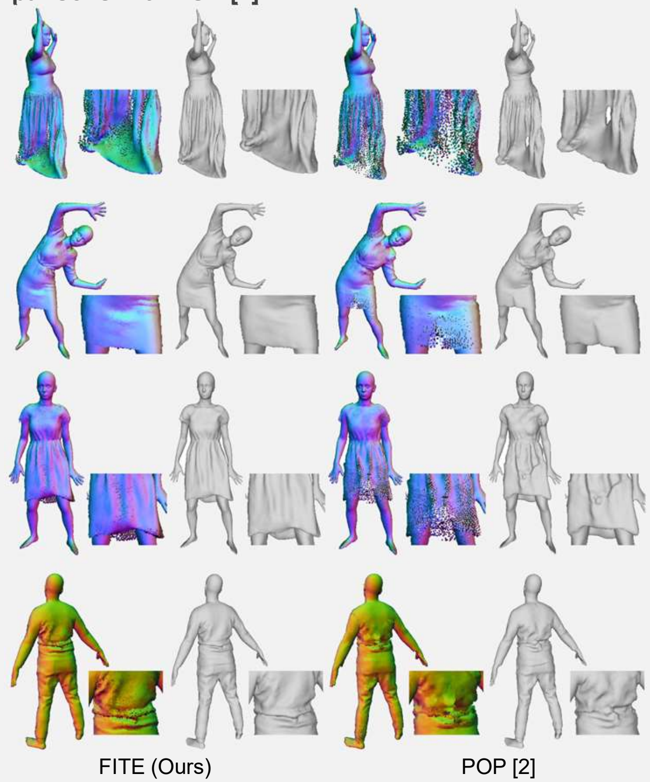


RESULTS ON THE RESYNTH DATASET

Comparison with POP [2], SNARF [3] and SCANimate [4]:



Comparisons with POP [2]:



Advantages of FITE:
- Better details compared with implicit methods [3,4].
- Better topology for dresses and skirts compared with POP [2].
- More continuous and uniform outputs compared with POP [2].

REFERENCE

[1] Ma et al. Learning to dress 3d people in generative clothing. CVPR 2020
[2] Ma et al. The power of points for modeling humans in clothing. ICCV 2021
[3] Chen et al. SNARF: Differentiable forward skinning for animating non-rigid neural implicit shapes. ICCV 2021
[4] Saito et al. SCANimate: Weakly supervised learning of skinned clothed avatar networks. CVPR 2021
[5] Loper et al. SMPL: A skinned multi-person linear model. ACMTOG 2015