Towards Animatable Human Avatars

Andreas Geiger

Autonomous Vision Group

University of Tübingen and MPI for Intelligent Systems

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

MAX-PLANCK-GESELLSCHAFT

ellis

European Laboratory for Learning and Intelligent Systems

Covered Papers

SNARF: Differentiable Forward Skinning for Animating Non-Rigid Neural Implicit Shapes

X. Chen, Y. Zheng, M. Black, O. Hilliges and A. Geiger

ICCV 2021

MetaAvatar: Learning Animatable Clothed Human Models from Few Depth Images

S. Wang, M. Mihajlovic, Q. Ma, A. Geiger and S. Tang

NeurIPS 2021

Collaborators

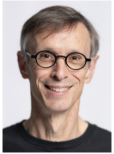

Xu Chen

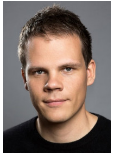

Shaofei Wang

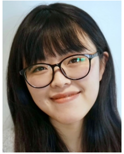

Yufeng Zheng


Marko Mihajlovic


Qianli Ma

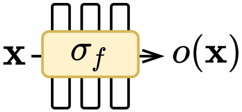

Michael Black


Otmar Hilliges


Siyu Tang

SNARF: Differentiable Forward Skinning for Animating Non-Rigid Neural Implicit Shapes

Neural Implicit Shapes



Neural implicit representations are useful for 3D human modeling:

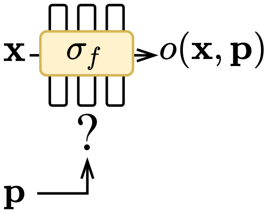
- Topological flexibility
- Resolution independent

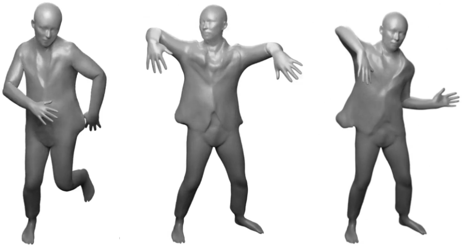
However:

- Animating such representations is not straightforward

Chen, Zheng, Black, Hilliges and Geiger: SNARF: Differentiable Forward Skinning for Animating Non-Rigid Neural Implicit Shapes. ICCV, 2021.

Animating Neural Implicit Shapes





Goal:

- Generate implicit shapes in given poses
- Learn shape representation from deformed observations

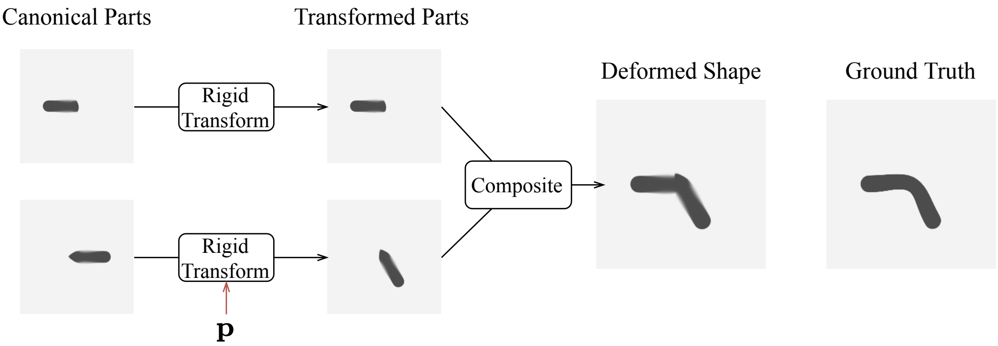
Key question:

- How to model and learn skeletal deformation of implicit shapes?

Chen, Zheng, Black, Hilliges and Geiger: SNARF: Differentiable Forward Skinning for Animating Non-Rigid Neural Implicit Shapes. ICCV, 2021.

Existing Solutions

Animating Implicit Shapes

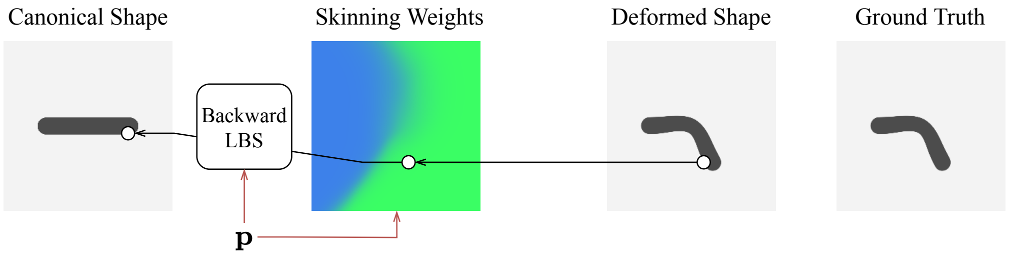


Piecewise rigid model: [Deng et al., ECCV 2020]

- Model shape as **parts,** and each part can be **rigidly transformed**
- Discontinuous artifacts at joints

Chen, Zheng, Black, Hilliges and Geiger: SNARF: Differentiable Forward Skinning for Animating Non-Rigid Neural Implicit Shapes. ICCV, 2021.

Animating Implicit Shapes

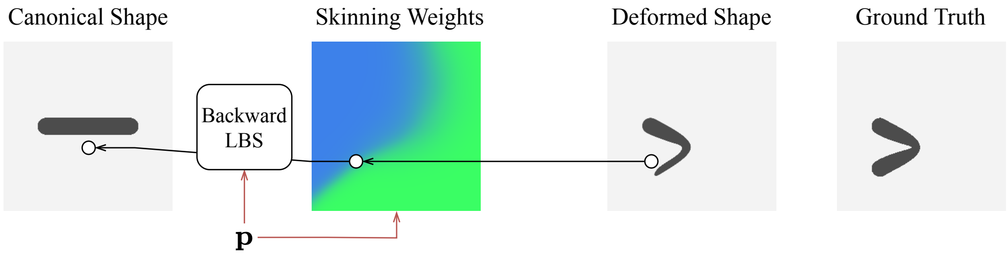


Backward skinning: [Jeruzalski et al., ArXiv 2020] [Mihajlovic et al., CVPR 2021]

- **Backward LBS** with **pose-dependent** skinning weights in deformed space

Chen, Zheng, Black, Hilliges and Geiger: SNARF: Differentiable Forward Skinning for Animating Non-Rigid Neural Implicit Shapes. ICCV, 2021.

Animating Implicit Shapes

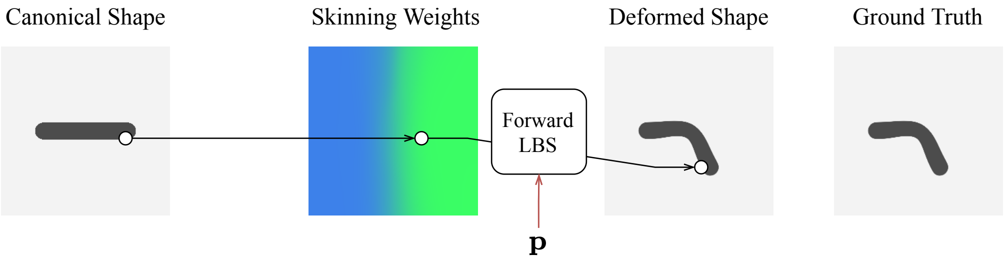


Backward skinning: [Jeruzalski et al., ArXiv 2020] [Mihajlovic et al., CVPR 2021]

- **Backward LBS** with **pose-dependent** skinning weights in deformed space
- Does not generalize to unseen poses
- Cannot handle one-to-many mapping

Chen, Zheng, Black, Hilliges and Geiger: SNARF: Differentiable Forward Skinning for Animating Non-Rigid Neural Implicit Shapes. ICCV, 2021.

Animating Implicit Shapes

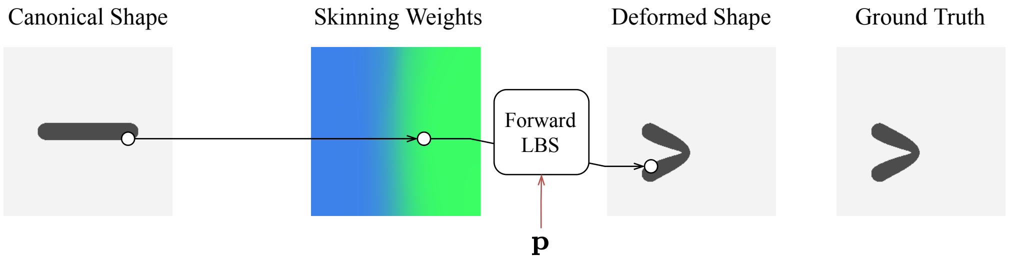


This work - forward skinning:

- Forward LBS with **pose-independent** skinning weights in canonical space

Chen, Zheng, Black, Hilliges and Geiger: SNARF: Differentiable Forward Skinning for Animating Non-Rigid Neural Implicit Shapes. ICCV, 2021.

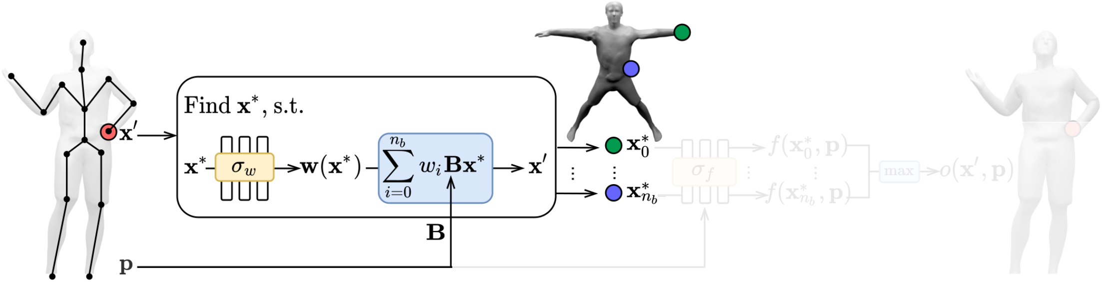
Animating Implicit Shapes



This work - forward skinning:

- Forward LBS with **pose-independent** skinning weights in canonical space
- Generalization to unseen poses
- Can handle one-to-many mapping

Differentiable Forward Skinning

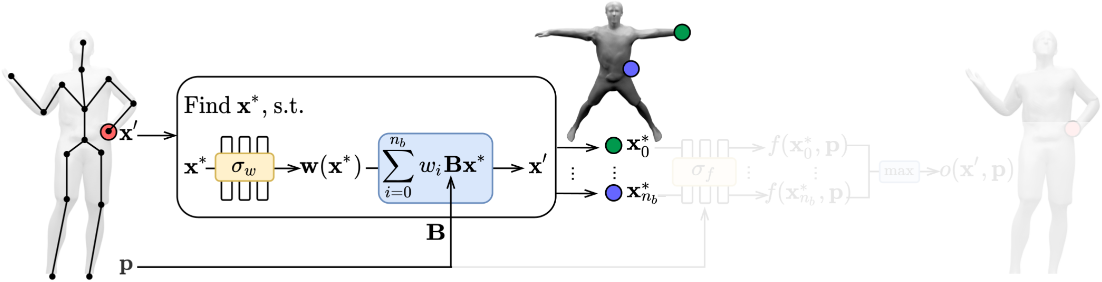


Correspondence search:

- Given query $\mathbf{x}'$, its canonical correspondences $\mathbf{x}^*$ satisfy $\mathbf{d}_{\sigma_w}(\mathbf{x}^*, \boldsymbol{B}) - \mathbf{x}' = \mathbf{0}$
- $\mathbf{x}^*$ can be numerically determined via iterative root finding

Chen, Zheng, Black, Hilliges and Geiger: SNARF: Differentiable Forward Skinning for Animating Non-Rigid Neural Implicit Shapes. ICCV, 2021.

Differentiable Forward Skinning

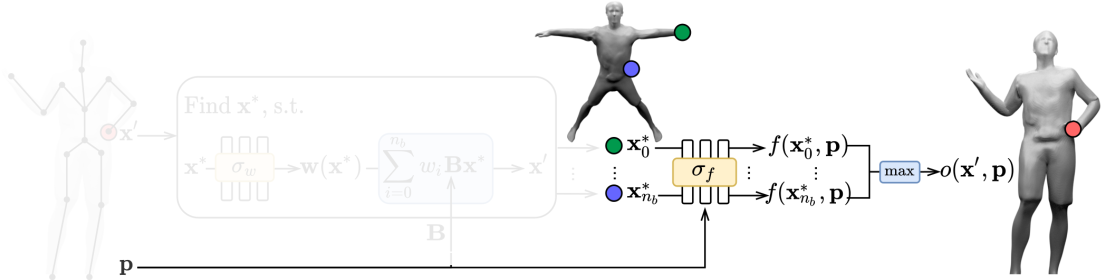


Multiple correspondences:

- Multiple solutions might exist $\rightarrow$ apply root finding with multiple initializations
- Rigidly transform the query point with each bone as initialization $\mathbf{x}_i^0 = \boldsymbol{B}_i^{-1} \cdot \mathbf{x}'$
- Collect valid solutions by convergence $\mathcal{X}^* = \{\mathbf{x}_i^* \mid \|\mathbf{d}_{\sigma_w}(\mathbf{x}_i^*, \boldsymbol{B}) - \mathbf{x}'\|_2 < \epsilon\}$

Chen, Zheng, Black, Hilliges and Geiger: SNARF: Differentiable Forward Skinning for Animating Non-Rigid Neural Implicit Shapes. ICCV, 2021.

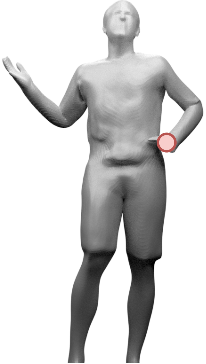
Differentiable Forward Skinning



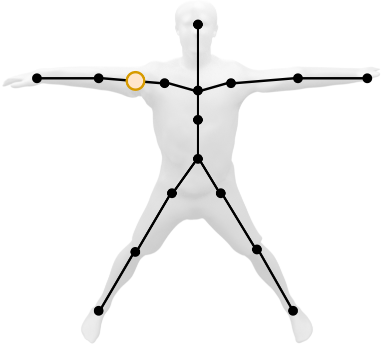
Occupancy query:

- Model the canonical shape as a single occupancy network
- Condition the canonical shape on pose to model pose-dependent deformations
- Aggregate multiple correspondences $o(\mathbf{x}', \mathbf{p}) = \max_{\mathbf{x}^* \in \mathcal{X}^*} \{f_{\sigma_f}(\mathbf{x}^*, \mathbf{p})\}$

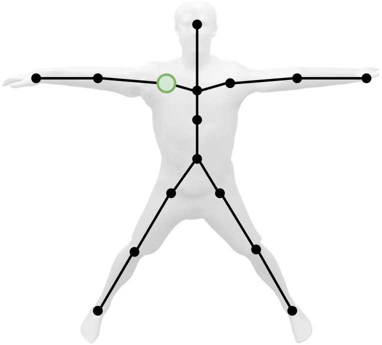
Training Objective



$$\mathcal{L}_{BCE}(o(\mathbf{x}', \mathbf{p}), o_{gt}(\mathbf{x}'))$$



$$\mathcal{L}_{bone} = \mathcal{L}_{BCE}(f_{\sigma_f}(\mathbf{x}_{bone}, \mathbf{p}), 1)$$

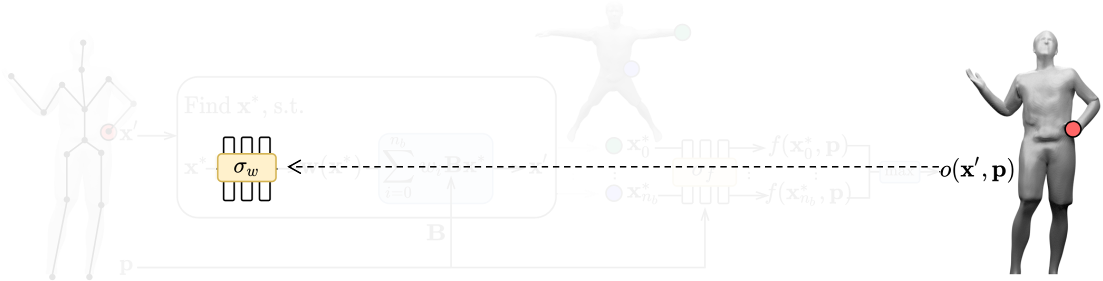


$$\mathcal{L}_{joint} = \|\mathbf{w}_{\sigma_w}(\mathbf{x}_{joint}) - \mathbf{w}_{target}\|_2^2$$

$$\mathcal{L} = \mathcal{L}_{BCE} + \underbrace{\mathcal{L}_{bone} + \mathcal{L}_{joint}}_{\text{only first epoch}}$$

Chen, Zheng, Black, Hilliges and Geiger: SNARF: Differentiable Forward Skinning for Animating Non-Rigid Neural Implicit Shapes. ICCV, 2021.

Gradients



- Analytical gradients via **implicit differentiation:**

$$\frac{\partial \mathcal{L}}{\partial \sigma_w} = \frac{\partial \mathcal{L}}{\partial o} \cdot \frac{\partial o}{\partial f_{\sigma_f}} \cdot \frac{\partial f_{\sigma_f}(\mathbf{x}^*)}{\partial \mathbf{x}^*} \cdot \frac{\partial \mathbf{x}^*}{\partial \sigma_w},$$

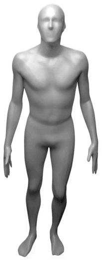
$$\frac{\partial \mathbf{x}^*}{\partial \sigma_w} = -\left(\frac{\partial \mathbf{d}_{\sigma_w}(\mathbf{x}^*, \boldsymbol{B})}{\partial \mathbf{x}^*}\right)^{-1} \cdot \frac{\partial \mathbf{d}_{\sigma_w}(\mathbf{x}^*, \boldsymbol{B})}{\partial \sigma_w}$$

Chen, Zheng, Black, Hilliges and Geiger: SNARF: Differentiable Forward Skinning for Animating Non-Rigid Neural Implicit Shapes. ICCV, 2021.

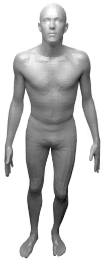
Results

Backward Skinning

NASA



Ours



Ground Truth

Chen, Zheng, Black, Hilliges and Geiger: SNARF: Differentiable Forward Skinning for Animating Non-Rigid Neural Implicit Shapes. ICCV, 2021.

Summary

Differentiable forward skinning:

- Learn forward skinning and shape fields in pose-independent space
- Learn from deformed shapes without direct supervision or prior (e.g., SMPL)
- Generalize to challenging unseen poses at test time

However:

- Root finding is time-consuming (10x slower than occupancy query)
  - Each iteration requires a skinning network query
- Requires 3D data $\Rightarrow$ combine with differentiable renderer to learn from images
- Requires accurate poses for training $\Rightarrow$ jointly optimize pose, shape and skinning
- So far only a single subject $\Rightarrow$ generative model of animatable avatars

Chen, Zheng, Black, Hilliges and Geiger: SNARF: Differentiable Forward Skinning for Animating Non-Rigid Neural Implicit Shapes. ICCV, 2021.

MetaAvatar: Learning Animatable Clothed Human Models from Few Depth Images

MetaAvatar

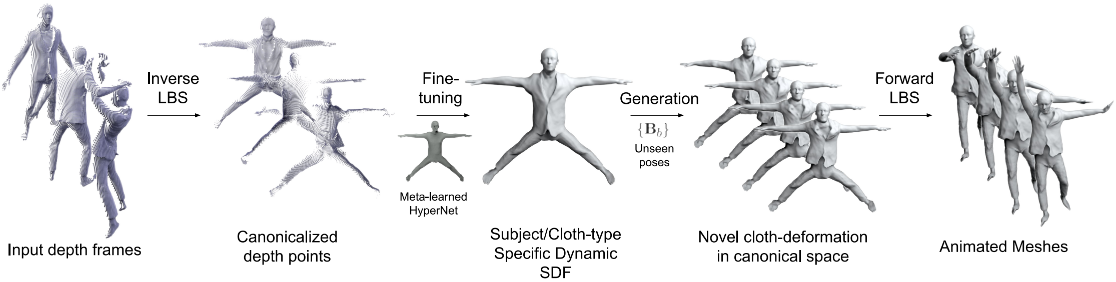
Goal:

- **Controllable avatars** learned from few **monocular depth** observations
- No fully-body scans or per-subject/cloth-type optimization required
- **Fast optimization** (2 minutes with 8 depth maps as input)

Idea: Meta-learn pose conditioned hypernetwork to predict parameters of neural SDF

Wang, Mihajlovic, Ma, Geiger and Tang: MetaAvatar: Learning Animatable Clothed Human Models from Few Depth Images. NeurIPS, 2021.

MetaAvatar

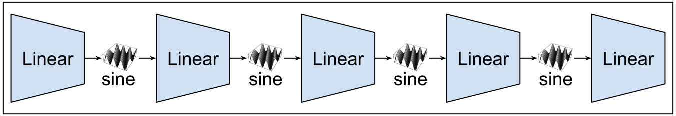


Approach:

- Using learned Inverse LBS, transform input depth frames into canonical pose
- Fine-tune a **meta-learned HyperNet** to predict parameters of neural SDF
- Given novel poses, our approach generates pose-dependent animated meshes

Wang, Mihajlovic, Ma, Geiger and Tang: MetaAvatar: Learning Animatable Clothed Human Models from Few Depth Images. NeurIPS, 2021.

Meta-learning a SDF

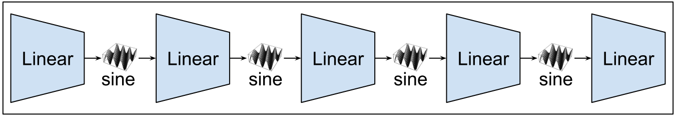


Meta-learning a SDF:

- 5-layer **SIREN** network with 256 neurons in each layer: $f_{\phi^*}(\mathbf{x})$
- **Point-based** on-surface and off-surface **loss** function [Gropp et al. 2020]
- **Meta-learn** network parameters on all subjects using Reptile [Nichol et al. 2018]
- Allows fast fine-tuning on new subject, but **no pose-dependent deformations**

Wang, Mihajlovic, Ma, Geiger and Tang: MetaAvatar: Learning Animatable Clothed Human Models from Few Depth Images. NeurIPS, 2021.

Meta-learning a pose-conditioned SDF



Meta-learning a pose-conditioned SDF:

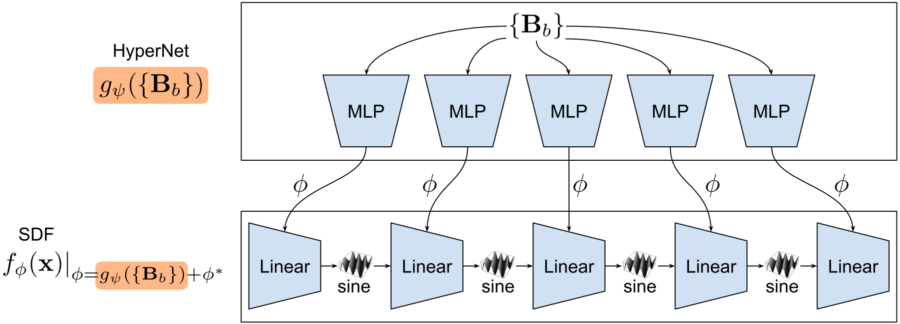
- 5-layer SIREN network with 256 neurons in each layer: $f_\phi(\mathbf{x}, \{\mathbf{B}_b\})$
- Condition network on **bone transformations** $\{\mathbf{B}_b\}$
- Does not work very well, leads to overly **smooth results**

Wang, Mihajlovic, Ma, Geiger and Tang: MetaAvatar: Learning Animatable Clothed Human Models from Few Depth Images. NeurIPS, 2021.

Meta-learning a pose-conditioned SDF

Wang, Mihajlovic, Ma, Geiger and Tang: MetaAvatar: Learning Animatable Clothed Human Models from Few Depth Images. NeurIPS, 2021.

MetaAvatar

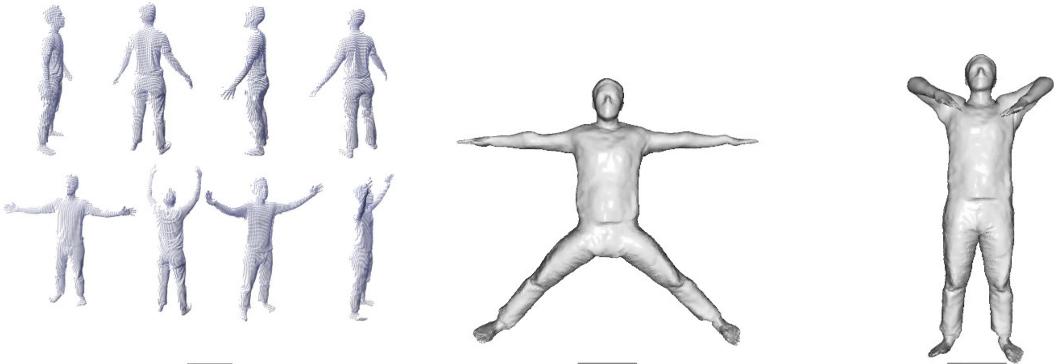


Meta-learning a pose-conditioned HyperNetwork:

- Learn **HyperNetwork** $g_\psi(\{\mathbf{B}_b\})$ on parameters of neural SDF
- HyperNetwork predicts **residuals** to meta-learned SDF parameters ϕ^*
- At test time, **fine-tune** parameters ψ of HyperNetwork

Wang, Mihajlovic, Ma, Geiger and Tang: MetaAvatar: Learning Animatable Clothed Human Models from Few Depth Images. NeurIPS, 2021.

Learning with Raw Sensor Inputs



Input
8 rendered monocular depth frames with estimated SMPL from PTF

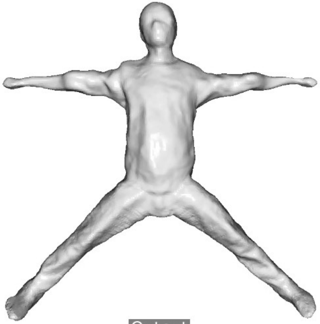
Output
Canonical body driven by novel poses

Output
Posed body

Wang, Mihajlovic, Ma, Geiger and Tang: MetaAvatar: Learning Animatable Clothed Human Models from Few Depth Images. NeurIPS, 2021.

Fine-tuning on Kinect Data

Input
8 filtered monocular depth frames with
estimated SMPL from POSEFusion

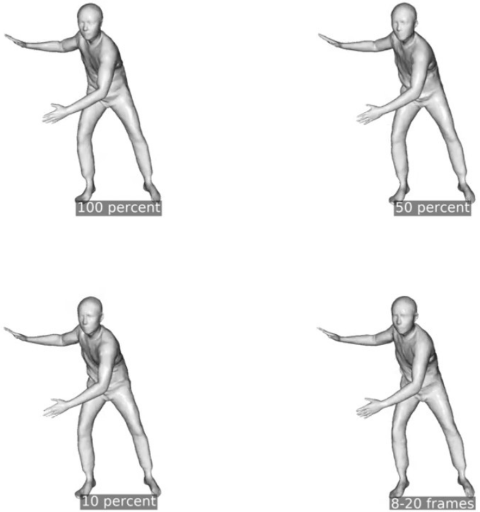


Output
Canonical body driven by novel poses

Output
Posed body

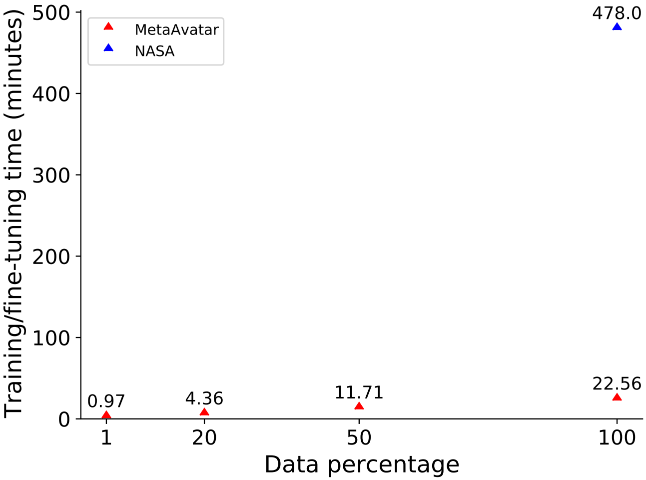
Wang, Mihajlovic, Ma, Geiger and Tang: MetaAvatar: Learning Animatable Clothed Human Models from Few Depth Images. NeurIPS, 2021.

Fine-tuning on Reduced Data



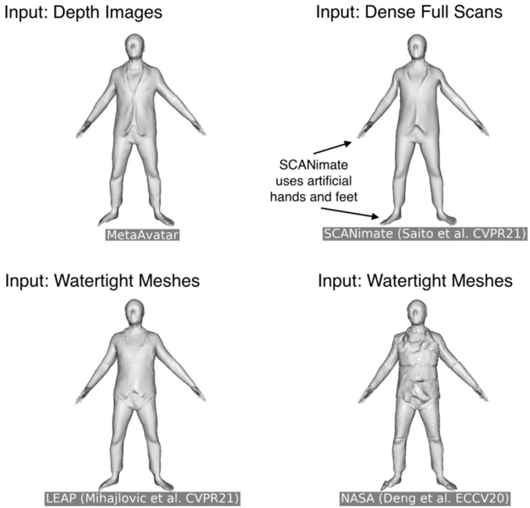
Wang, Mihajlovic, Ma, Geiger and Tang: MetaAvatar: Learning Animatable Clothed Human Models from Few Depth Images. NeurIPS, 2021.

Fine-tuning on Reduced Data



Wang, Mihajlovic, Ma, Geiger and Tang: MetaAvatar: Learning Animatable Clothed Human Models from Few Depth Images. NeurIPS, 2021.

Comparison to Baselines



Wang, Mihajlovic, Ma, Geiger and Tang: MetaAvatar: Learning Animatable Clothed Human Models from Few Depth Images. NeurIPS, 2021.

Summary

MetaAvatar:

- MetaAvatar enables generation of controllable clothed human avatars
- Meta-learning allows for fast subject-specific fine-tuning from few depth images
- MetaAvatar enables realistic clothed avatars in 2 minutes from 8 depth maps
- HyperNetworks are required to capture detailed pose dependent deformations
- Learned Inverse/Forward LBS models and bone transformations required as input

Wang, Mihajlovic, Ma, Geiger and Tang: MetaAvatar: Learning Animatable Clothed Human Models from Few Depth Images. NeurIPS, 2021.

Thank you!

http://autonomousvision.github.io

Federal Ministry of Education and Research

Federal Ministry for Economic Affairs and Energy