Virtual Humans – Winter 23/24

Lecture 9_1 – Neural Implicits and Point Based Clothing Models

Prof. Dr.-Ing. Gerard Pons-Moll
University of Tübingen / MPI-Informatics
PART1: Neural Implicits for 3D Shapes

PART2: Neural Implicits for Humans

PART3: Neural Implicits – Generative Models

PART4: Point-based Clothing Models
PART1: Neural Implicits for 3D Shapes

PART2: Neural Implicits for Humans

PART3: Neural Implicits – Generative Models

PART4: Point-based Clothing Models
What is a good representation for 3D data?
What is a good representation for 3D data?

- Compatible with neural networks.
- Flexible
- High fidelity
Voxels

- Discretization of 3D space into grid.
- Easy to process with neural networks.
- Cubic memory $\mathcal{O}(n^3) \rightarrow$ limited resolution.

Image credit: Mescheder et al. CVPR'19

[Liao et al. CVPR'18]
[Chov et al. ECCV'16]
Pointclouds

- Discretization of 3D space into 3D points.
- Does not model connectivity/topology.
- Limited number of points.

Image credit: Mescheder et al. CVPR'19

[Liao et al. CVPR'18]
[Chov et al. ECCV'16]
Meshes

- Discretization into vertices and faces.
- Limited number of vertices/granularity.
- Requires class specific template.
- Leads to self-intersections.

Image credit: Mescheder et al. CVPR'19

[Wang et al. ECCV'18]
Meshes

- Discretization into vertices and faces.
- Limited number of vertices/granularity.
- Requires class specific template.
- Leads to self-intersections.

We have covered mesh-based human/clothing models.

[Wang et al. ECCV’18]

Image credit: Mescheder et al. CVPR’19
Implicit representation

- Implicit representation $\rightarrow$ No discretization.
- Arbitrary topology and resolution.
- Low memory footprint.
- Not restricted to specific class.

Mescheder et al. CVPR’19
Chen et al. CVPR’19
Park et al. CVPR’19

Image credit: Mescheder et al. CVPR’19
Surfaces as an Implicit Function

A function tells us whether a point is inside or outside an object

\[ f(p) = \begin{cases} 
0, & \text{if } p \in \text{outside} \\
1, & \text{if } p \in \text{inside} 
\end{cases} \quad p = (x, y, z) \in \mathbb{R}^3 \]

If the function is continuous, a levelset of it defines a surface

\[ S = \{ p, \quad f(p) = \tau \} \]
Surfaces as an Implicit Function

\[ p = (x, y, z) \in \mathbb{R}^3 \]

\[ f(p) = \begin{cases} 
0, & \text{if } p \in \text{outside} \quad \bullet \\
1, & \text{if } p \in \text{inside} \quad \blacksquare
\end{cases} \]

\[ S = \{ p, \ f(p) = \tau \} \]

✔️ With implicit functions, Topology changes only require changing \( f(p) \)

✗ Mesh based representations would struggle
Neural Implicits for common objects
Neural Implicits for common objects

Work well for rigid objects:

- ✔ Continuous
- ✔ Multiple topologies

[Park et al. CVPR’19]

[Chen et al. CVPR’19]

[Mescheder et al. CVPR’19]
Previous Implicit Function Learning Architecture

\( X \) is a shape observation:
- Sparse point-cloud
- Partial point-cloud
- Low-res voxels

\( z \in \mathbb{R}^N \)

\[ f(z, p) \]

\[ [0, 1] \]

\[ p = (x, y, z) \]

[Mescheder et al. CVPR’19
Chen et al. CVPR’19
Park et al. CVPR’19]
Previous Implicit Function Learning
Architecture

\( \mathbf{x} \) is a shape observation:
- Sparse point-cloud
- Partial point-cloud
- Low-res voxels

\[ \mathbf{z} \in \mathbb{R}^N \]

\( f(\mathbf{z}, \mathbf{p}) \)

\[ [0, 1] \]

\[ \mathbf{p} = (x, y, z) \]

[Mescheder et al. CVPR’19
Chen et al. CVPR’19
Park et al. CVPR’19]
Problem with Previous Work

X Reconstruct Articulations

[Chen et al. CVPR'19]
[Mescheder et al. CVPR'19]

X Retain Details

[Chen et al. CVPR'19]
[Mescheder et al. CVPR'19]
Implicit Functions in Feature Space for 3D Shape Reconstruction and Completion

Julian Chibane\textsuperscript{1,2}, Thiemo Alldieck\textsuperscript{1,3}, Gerard Pons-Moll\textsuperscript{1}

CVPR 2020
Problems with previous work

1) Loss of 3D structure

2) Point coordinates carry no information about local shape

[Mescheder et al. CVPR’19
Chen et al. CVPR’19]
Implicit Feature Networks (IF-Nets)

3D Grid
$K \times K \times K$

Chibane et al. IF-Nets CVPR’20
Implicit Feature Networks (IF-Nets)

3D Grid
$K \times K \times K$

Chibane et al. IF-Nets CVPR’20
Representation of IF-Nets

Previous:
\[ f(z, p) \mapsto [0, 1] \]

\[ F_1, \ldots, F_n, \quad F \in \mathcal{F}^{K \times K \times K} \]

Ours:
\[ f(F_1(p), \ldots, F_n(p)) \mapsto [0, 1] \]

Chibane et al. IF-Nets CVPR’20
IF-Nets for 3D Shape Reconstruction and Completion

- Reconstruct Articulations
- Retain Details
- Complete Shape
IF-Nets for Texture completion
ECCV SHARP CHALLENGE

[Chibane and Pons-Moll, IF-Nets for texture. SHARP 2020
Chibane et al. IF-Nets CVPR’20]
• Surfaces that do not divide the space in two regions can not be represented.

• We need a different **output representation**.
Neural Unsigned Distance Fields for Implicit Function Learning

Julian Chibane, Aymen Mir, Gerard Pons-Moll
NeurIPS 2020

Chibane et al. *NDF*, NeurIPS 2020
Our Solution

\[ f(F_1(p), \ldots, F_n(p)) \mapsto [0, 1] \]

Change the output representation

\[ f(F_1(p), \ldots, F_n(p)) \mapsto \mathbb{R}^+ \]

**Unsigned distance:**

\[ f(p) = \min_{q \in S} \|p - q\| \]

---

Only water-tight surfaces

Open surfaces and manifolds

Functions

Complex shapes

Chibane et al. **NDF, NeurIPS 2020**
Neural Distance Fields

\[ f(p) = \min_{q \in S} \|p - q\| \]

\[ S = \{ p \in \mathbb{R}^d \mid f(p) = 0 \} \]

\[ q = p - f(p) \nabla_p f(p) \]
Neural processing of arbitrary surfaces

Next, we illustrate the capabilities of NDF to neurally process arbitrary surfaces, not representable by prior learned implicit work:

- **Mathematical Functions and Manifolds** – We train a single NDF on a dataset consisting of 1000 functions per type: linear function, parabola, sinusoids and spirals.

- **Garments** – Open Surfaces, without thickness. Training on ~300 garments of five types from [Bhatnagar et al. ICCV’19].

- **Scenes** – Open surfaces with holes and no thickness. Training on 34 real world scenes captured by RGBD Sensors from [Xia et al. CVPR’18].
NDF results

1) Comparison with IF-Nets
   - IF-Net (cut)
   - Ours (cut)
   - IF-Net (Transparent)
   - Ours (Transparent)

2) Garment reconstruction

   GT Mesh
   Input
   Output 1Mio. PC
   Direct Rendering (Front View)
   Direct Rendering (Side View)
   Open surface

Chibane et al. NDF, NeurIPS 2020
Direct Rendering of NDF

Chibane et al. NDF, NeurIPS 2020
Representation and Regression of Functions

\[ y = h(x_1, \ldots, x_n) \quad \rightarrow \quad f(x_1, \ldots, x_n, y) = 0 \]

\[ p(\lambda) = (x_1, \ldots, x_n, 0) + \lambda(0, \ldots, 0, 1) \quad \rightarrow \quad f(p(\lambda)) = 0 \]

Classical regression using NDFs and an adapted sphere tracing (ray tracing method)

Chibane et al. NDF, NeurIPS 2020
Representation and Completion of Scenes

Input  
Output  
Ground Truth

Chibane et al. *NDF*, NeurIPS 2020
Meshes vs Implicits

1) Parametric Meshes

[Alldieck et al. CVPR’18
Bhatnagar et al. ICCV’19, ECCV’20
Tiwari et al. ECCV’20]

2) Implicit Functions

[Chibane et al. CVPR’20
Chibane et al. NeurIPS’20]

General objects and humans

<table>
<thead>
<tr>
<th></th>
<th>Control /Meaning</th>
<th>Topology</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>1)</td>
<td>✓</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>2)</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

2) Compatible with learning
PART1: Neural Implicits for 3D Shapes

PART2: Neural Implicits for Humans

PART3: Neural Implicits – Generative Models

PART4: Point-based Clothing Models
Human and Clothing Models

Prior works → mesh based

Guan et al., 2012  Danerek et al., 2016  Lähner et al., 2018

Gundogdu et al., 2019  Santesteban et al., 2019  Ma et al., 2020

Wang et al., 2018  Patel et al., 2020  Tiwari et al., 2020
Human and Clothing Models

Prior works $\rightarrow$ mesh based

- Guan et al., 2012
- Danerek et al., 2016
- Lähner et al., 2018
- Gundogdu et al., 2019
- Santesteban et al., 2019
- Ma et al., 2020
- Patel et al., 2020
- Tiwari et al., 2020

Fixed topology
Topology has to be manually predefined
Limited resolution
Meshes vs Implicits

1) Parametric Meshes
   - [Alldieck et al. CVPR’18]
   - [Bhatnagar et al. ICCV’19, ECCV’20]
   - [Tiwari et al. ECCV’20]

2) Implicit Functions
   - General objects and humans
     - [Chibane et al. CVPR’20]
     - [Chibane et al. NeurIPS’20]

<table>
<thead>
<tr>
<th></th>
<th>Control /Meaning</th>
<th>Topology</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>1)</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>2)</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

2) Compatible with learning
Human and clothing model using Neural Implicits

- High fidelity
- Flexible topology
- Pose/Shape/Style controllable
- Learned directly from scans

Corona et al. CVPR’21
Tiwari et al. ICCV’21
Deng et al. ECCV’20
Saito et al. CVPR’21
Human and clothing model using Neural Implicits

- **High fidelity**
- **Flexible topology**
- Pose/Shape/Style controllable
- Learned directly from scans

Corona et al. CVPR’21

Tiwari et al. ICCV’21

Deng et al. ECCV’20

Saito et al. CVPR’21
Human and clothing model using Neural Implicits

- High fidelity
- Flexible topology
- Pose/Shape/Style controllable
- Learned directly from scans

Corona et al. CVPR’21
Tiwari et al. ICCV’21
Deng et al. ECCV’20
Saito et al. CVPR’21
Controllable Neural Implicits for Human

Vertex based human model: SMPL

\[ M(\theta, \beta) : \theta \times \beta \rightarrow \mathbf{V} \in \mathbb{R}^3 \]

Neural Implicit for common objects:

\[ f(\mathbf{p}, z) : \mathbb{R}^3 \times \mathbb{R}^d \rightarrow d \in \mathbb{R} \]

\[ S = \{ \mathbf{p}, \ f(\mathbf{p}) = \tau \} \]
Controllable Neural Implicit humans:

\[ f(p, \theta, \beta) : \mathbb{R}^3 \times \theta \times \beta \rightarrow d \in \mathbb{R} \]

\[ S = \{ p, f(p) = \tau \} \]
Learning pose-conditioned occupancy

- Naïve solution (Unstructured)

\[ \mathcal{O}(p|\theta) = f_w(p, \{B_b^{-1}t_0\}) \]

- \( B_b \): Bone transformations
- \( t_0 \): Root translation
- \( p \): Query point

Deng et al. NASA, ECCV 2020
Learning pose-conditioned occupancy

- Naïve solution (Unstructured)

\[ \mathcal{O}(p|\theta) = f_w(p, \{B_b^{-1}t_0\}) \]
Incorporating prior knowledge about human models

Vertex based human model: SMPL

\[ M(\theta, \beta) : \theta \times \beta \rightarrow \mathbf{V} \in \mathbb{R}^3 \]

\[ T(\theta, \beta) = \mathbf{T} + B_s(\beta) + B_p(\theta) \]
Learning pose/shape conditioned neural implicits using part composition
• Naïve solution (Unstructured)

\[ \mathcal{O}(p|\theta) = f_w(p, \{B^{-1}_b t_0\}) \]

Unstructured (U)
**NASA**

- Piecewise-rigid model

\[ O(p|\theta) = \max \{ f_w(B_b^{-1}p) \} \]

A point is occupied if it is occupied by *any* of the parts. Done with max operator

Deng et al. *NASA*, ECCV 2020
• Piecewise-rigid model

\[ O(p|\theta) = \max \{ f_w(B_b^{-1}p) \} \]
NASA

- Piecewise-deformable model

\[ \mathcal{O}(p|\theta) = \max \{ f_w(B_b^{-1}p|\theta) \} \]
NASA results

- Piecewise-deformable model

\[ O(p|\theta) = \max\{f_w(B^{-1}_b p|\theta)\} \]

Deng et al. NASA, ECCV 2020
Limitations of NASA:
- Part-based artefacts
  - No information about neighbouring body parts
- Limited pose generalization
  - Low-dimensional pose encoding does not fully remove long-range spurious correlations

Deng et al. NASA, ECCV 2020
COAP: Compositional Articulated Occupancy of People

Marko Mihajlovic1  Shunsuke Saito2  Aayush Bansal2  Michael Zollhoefer2  Siyu Tang1
1ETH Zurich  2Reality Labs Research at Meta

CVPR 2022
COAP: Compositional Articulated Occupancy of People
How is COAP different from NASA?

COAP is not subject-specific model
How is COAP different from NASA?

Per-part features = body part + few points from neighbouring parts
How is COAP different from NASA?
Part-based model

– Skinning weights are needed for part-decomposition
– Part-artefacts are prominent for out-of-distribution poses.
– Cannot model clothing/ loose clothing.
Part-based model limitations

– Skinning weights are needed for part-decomposition
– Part-artefacts are prominent for out-of-distribution poses.
– Cannot model clothing/loose clothing.
Part-based model

- Skinning weights are needed for part-decomposition
- Part-artefacts are prominent for out-of-distribution poses.
- Cannot model clothing/loose clothing.
Learning pose/shape conditioned neural implicits using learned LBS and canonical shape
Learning pose/shape conditioned neural implicits using learned LBS and canonical shape

Given an input pose/shape and 3D query point

\[ p' \in \mathbb{R}^3 \]

Predict pose/shape dependent deformation field in canonical space

\[ \Delta p = f_d(p, \theta, \beta) \]

Get the final posed mesh
Learning pose/shape conditioned neural implicits using learned LBS and canonical shape

Given an input pose/shape and 3D query point

Map query point to canonical space? Using learned LBS?

Predict pose/shape dependent deformation field in canonical space

Get the final posed mesh
Neural-GIF: Neural Generalized Implicit Functions for Animating People in Clothing

Garvita Tiwari  Nikolaos Sarafianos  Tony Tung  Gerard Pons-Moll
University of Tuebingen  MPI for Informatics, Saarland Informatics Campus, Germany  Facebook Reality Labs, Sausalito, USA

ICCV 2021
NeuralGIF

- A generalized framework to animate people in clothing (or clothing), which learns directly from scans

NeuralGIF is trained on set of raw scans for a given subject

Given a query pose (left), NeuralGIF animates the subject (right)

Tiwari et al. ICCV’21
Generalized Implicit Function

\[ p = (x, y, z) \in \mathbb{R}^3 \quad \Rightarrow \quad p' = Rp + t \]

\[ S = \{ p, \; f(p) = \tau \} \quad S' = \{ p', \; f(R^{-1}(p' - t)) = \tau \} \]
Generalized Implicit Function

\[ p = (x, y, z) \in \mathbb{R}^3 \quad \Rightarrow \quad p' = p + d(x, y, z) \]

\[ S = \{ p, \, f(p) = \tau \} \quad \Rightarrow \quad S' = \{ p', \, f(R^{-1}(p' - t)) = \tau \} \]
SMPL model

$$p' = \left( \sum_{i=1}^{K} w_i B_i \right) p$$

$$T(\theta, \beta) = T_\mu + B_s(\beta) + B_p(\theta)$$

Vertices in a 0-pose
Neural-GIF

How to predict the signed distance for a point in the posed space?

\[ p' \in \mathbb{R}^3 \]

\[ p = \left( \sum_{i=1}^{K} w_i B_i \right)^{-1} p' \]

\[ w = f_c(p', \theta) \quad w \in \mathbb{R}^{24} \]

\[ \mathbf{B}_i \rightarrow \text{Joints transformation matrix} \]
Neural-GIF: Pose driven Animation

CAPE
DFAUST
TailorNet-Shirt
TailorNet-Skirt

Tiwari et al. ICCV'21
NeuralGIF as Multi-shape model

Tiwari et al. ICCV'21
Neural-GIF vs Scanimate

\[ S = \{ \mathbf{p}, f_{\theta}(\mathbf{p}; \theta) \} = \tau \]  \[ S = \{ \mathbf{p}', f(\mathbf{p} + \Delta \mathbf{p}(\theta)) \} = \tau \]

Advantages of Neural-GIF
- A single Canonical f(\cdot) is learned.
- More flexibility in topologies
- Better detail
- Simpler model

Figure 7. Comparison with SCANimate: We compare the results of our method on the CAPE dataset with SCANimate [51]. Our model preserves more details and does not have posing artifacts.

Tiwari et al. ICCV'21
Comparison with State-of-the-art methods

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Model</th>
<th>NASA [18]</th>
<th>SCANimate [51]</th>
<th>Ours (Neural-GIF)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Point2Surface ↓</td>
<td>IoU ↑</td>
<td>F-Score ↑</td>
<td>Point2Surface ↓</td>
</tr>
<tr>
<td>CAPE [31]</td>
<td>10.67</td>
<td>0.918</td>
<td>94.32</td>
<td>5.82</td>
</tr>
<tr>
<td>ClothSeq</td>
<td>23.26</td>
<td>0.780</td>
<td>57.29</td>
<td>7.32</td>
</tr>
<tr>
<td>DFAUST [13]</td>
<td>10.52</td>
<td>0.939</td>
<td>95.48</td>
<td>3.79</td>
</tr>
</tbody>
</table>

Comparison with NASA [1] and SCANimate [2]. We report point to surface distance (in mm) and IoU and F-Scores(%) for comparison

1. NASA: Neural Articulated Shape Approximation, Deng et al., ECCV2020
2. SCANimate: Weakly Supervised Learning of Skinned Clothed Avatar Networks, Saito et al., CVPR2021

Most of the improvement is in modelling fine geometric details.

Tiwari et al. ICCV’21
NeuralGIF as Multi-shape model

<table>
<thead>
<tr>
<th>Dataset</th>
<th>LEAP [36]</th>
<th>Ours (Neural-GIF)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Point2Surface ▼</td>
<td>IoU ▲</td>
</tr>
<tr>
<td>DFAUST [13]</td>
<td>3.42</td>
<td>0.958</td>
</tr>
<tr>
<td>MoVi [19]</td>
<td>3.19</td>
<td><strong>0.969</strong></td>
</tr>
<tr>
<td>SMPL</td>
<td>3.26</td>
<td>0.968</td>
</tr>
</tbody>
</table>

We quantitatively compare the results of our method with LEAP[1] on various datasets. We report point to surface distance (in mm) and IoU for comparison.
Neural-GIF

Pose space to unpose space using skinning weights

\[ p' \in \mathbb{R}^3 \]

\[ p = \left( \sum_{i=1}^{K} w_i B_i \right)^{-1} p' \]

\[ \Delta p = f_d(p, \theta) \]

\[ f_{sdf}(p + \Delta p) \]

\[ w = f_c(p', \theta) \quad w \in \mathbb{R}^{24} \]

\[ B_i \rightarrow \text{Joints transformation matrix} \]

Tiwari et al. ICCV'21
Backward and Forward skinning

Backward skinning

Posed space skinning field predicted

Source: https://autonomousvision.github.io/snarf/
Backward and Forward skinning

**Backward skinning**

In backward skinning, \( w \) is predicted from the deformed point \( x' \) and the pose \( \theta \)

Source: https://autonomousvision.github.io/snarf/
Backward and Forward skinning

**Backward skinning**

In backward skinning, \( w \) is predicted from the deformed point \( x' \) and the pose \( \theta \)

**Forward skinning**

In forward skinning, \( w \) is predicted from the canonic point \( x^* \)

Source: https://autonomousvision.github.io/snarf/
SNARF: Differentiable Forward Skinning for Animating Non-Rigid Neural Implicit Shapes

Source: https://autonomousvision.github.io/snarf/
SNARF: Differentiable Forward Skinning for Animating Non-Rigid Neural Implicit Shapes

Source: https://autonomousvision.github.io/snarf/
SNARF: Differentiable Forward Skinning for Animating Non-Rigid Neural Implicit Shapes

SNARF is a forward skinning method:
Forward skinning explicitly defines

\[ \mathbf{x} \rightarrow \mathbf{x}' \ \forall \mathbf{x} \]

Learning canonical shape from posed scans requires,

\[ \mathbf{x}' \rightarrow \mathbf{x} \ \forall \mathbf{x}' \]

Given \( \mathbf{x} \rightarrow \mathbf{x}' \), determine \( \mathbf{x}' \rightarrow \mathbf{x} \)
- Implicit relation, no closed form solution
- Non-bijective mapping, multiple solution may exist

Source: Chen et al., ICCV 2021
SNARF: Differentiable Forward Skinning for Animating Non-Rigid Neural Implicit Shapes

**Key Idea:** Differentiable Forward Skinning

Requires differentiating through the solution of a non-linear system

Source: Chen et al., ICCV 2021
SNARF: Understand the training objective

\[ \text{find } \mathbf{x}^* \text{ such that } \mathbf{x}' - \sum_{i=0}^{n_b} B_i \sigma_{w,i}(\mathbf{x}^*) = 0 \]

Canonical Point (multiple solutions) \quad Posed point

Neural network which predicts skinning weights from canonic point
\[ \sigma_{w,i}(\mathbf{x}^*) = \mathbf{w}(\mathbf{x}^*) \in \mathbb{R}^{n_b} \]

\[ \mathcal{L}_w = \mathcal{L}(\max_b \{ f(\mathbf{x}_b^*(\mathbf{w})) \}_{b=1}^{n_b}, \varrho(\mathbf{x}')) \]

Cross-entropy loss \quad Ground truth occupancy

Challenge: compute \[ \frac{\partial \mathcal{L}}{\partial \mathbf{w}} \] Possible to backprop iterative root finding
SNARF results

Source: Chen et al., ICCV 2021
Backward and Forward skinning

Forward skinning models have better generalization w.r.t. unseen pose.

Backward-LBS

Forward-LBS

Within distribution poses

Out of distribution poses
Backward and Forward skinning

Forward skinning models have better generalization w.r.t. unseen pose.

Reconstructions of Novel Poses outside Training Distribution (PosePrior)

Backward Skinning  NASA  Ours  Ground Truth
Backward and Forward skinning

- **Forward** skinning models have **better generalization** w.r.t. unseen pose.
- **Backward** skinning models have **higher fidelity** (for distribution poses) and **more flexible** to model loose clothing.
Works like NASA, COAP, Neural-GIF, SNARF are not generative models
PART1: Neural Implicits for 3D Shapes

PART2: Neural Implicits for Humans

PART3: Neural Implicits – Generative Models

PART4: Point-based Clothing Models
Works like NASA, COAP, Neural-GIF, SNARF are not generative models

What’s next??

Neural Implicit based generative model of people in clothing.
gDNA: Towards Generative Detailed Neural Avatars

Generative model of human in clothing using Neural Implicits
gDNA: Towards Generative Detailed Neural Avatars

Chen et al., CVPR 2022
gDNA: Towards Generative Detailed Neural Avatars

Chen et al., CVPR 2022
gDNA: Towards Generative Detailed Neural Avatars

Training is based on auto-decoders for the 3D shape and GANs for stochastic detail

Chen et al., CVPR 2022
Canonical Implicit model

\[ S(z_{\text{shape}}) = \{ x \mid O(x, z_{\text{shape}}) = \tau \} \]

\[ O : \mathbb{R}^3 \times \mathbb{R}^{L_{\text{shape}}} \to [0, 1] \times \mathbb{R}^{L_f} \]

\[ (x, z_{\text{shape}}) \mapsto (o, f) \]

\[ z_{\text{shape}} \in \mathbb{R}^{L_{\text{shape}}} \]

Chen et al., CVPR 2022
Canonical Implicit model

\[ S(z_{\text{shape}}) = \{ x \mid \mathcal{O}(x, z_{\text{shape}}) = \tau \} \]

\[ \mathcal{O} : \mathbb{R}^3 \times \mathbb{R}^{L_{\text{shape}}} \rightarrow [0, 1] \times \mathbb{R}^{L_f} \]

\[ (x, z_{\text{shape}}) \mapsto (o, f) \]

\[ z_{\text{shape}} \in \mathbb{R}^{L_{\text{shape}}} \]

\[ \mathcal{N} : \mathbb{R}^3 \times \mathbb{R}^{L_{\text{detail}}} \times \mathbb{R}^{L_f} \rightarrow \mathbb{R}^3 \]

\[ (x, z_{\text{detail}}, f) \mapsto n \]

\[ z_{\text{detail}} \in \mathbb{R}^{L_{\text{detail}}} \]

Chen et al., CVPR 2022
Multi-subject forward-skinning model

Based on SNARF

\[ \mathcal{W} : \mathbb{R}^3 \times \mathbb{R}^{L_{\text{shape}}} \to \mathbb{R}^{n_b} \]

\[ (x, z_{\text{shape}}) \mapsto w \]

Skinning field in a body-shape-independent space

Chen et al., CVPR 2022
Multi-subject forward-skinning model

Based on SNARF

$\mathcal{M} : \mathbb{R}^3 \times \mathbb{R}^{L_\beta} \rightarrow \mathbb{R}^3$

$\mathcal{W} : \mathbb{R}^3 \times \mathbb{R}^{L_{\text{shape}}} \rightarrow \mathbb{R}^{n_b}$

$(x, z_{\text{shape}}) \mapsto w$

Skinning field in a body-shape-independent space

Body shape dependent warping field

Chen et al., CVPR 2022
Multi-subject forward-skinning model

Based on SNARF

\[ \text{Find } \hat{x}^*, \text{ s.t. } d(\hat{x}^*, \beta, \theta, z_{\text{shape}}) = x' \]

\[ d(\hat{x}, \beta, \theta, z_{\text{shape}}) - x' = 0, \]

Chen et al., CVPR 2022
Multi-subject forward-skinning model

Based on SNARF

\[
d(\hat{x}, \beta, \theta, z_{shape}) - x' = 0, \\
x^* = M(\hat{x}^*, \beta)
\]

Chen et al., CVPR 2022
gDNA: Towards Generative Detailed Neural Avatars

Pose conditioned
gDNA: Towards Generative Detailed Neural Avatars

Body shape conditioned
gDNA: Towards Generative Detailed Neural Avatars

Clothing style/shape conditioned
SMPLicit: Topology-aware Generative Model for Clothed People

Enric Corona Albert Pumarola Guillem Alenyà  Gerard Pons-Moll Francesc Moreno-Noguer

Institut de Robotica i Informatica Industrial, CSIC-UPC, Barcelona, Spain  Max Planck Institute for Informatics

CVPR 2021

Cloth interpolation (Using a single model)
Vertex-based SMPL to Implicit SMPL (SMPLicit)

Vertex based Clothing model
E.g, TailorNet predicts vertex displacement \( D \) as a function of pose, shape and clothing style (requires multiple cloth templates)

\[
M(\theta, \beta, D) : \theta \times \beta \times D \rightarrow V \in \mathbb{R}^3
\]

\[
T(\theta, \beta, D) = T + B_s(\beta) + B_p(\theta) + D
\]

Neural implicit clothing model
Predicts the unsigned distance of the surface as a function of pose, shape, clothing cut, and style

\[
C(p, \theta, z_{\text{cut}}, z_{\text{style}}) \rightarrow \mathbb{R}^+
\]

Clothing cut controls how much clothing overlaps with the body (sleeve length, pant length)

Clothing style controls the size, and fit

Corona et al., SMPLicit, CVPR’21
Moving to new topologies: Implicit representations

Unsigned distance field

\[ D(p) \leftarrow C(P_\beta, z) \]
Moving to new topologies: Implicit representations

Input point $\mathbf{p}$

Unsigned distance field

$D(p) \leftarrow C(\mathbf{P}_\beta, \mathbf{z})$

Corona et al., *SMPLicit*, CVPR’21
Moving to new topologies: Implicit representations

Cloth style

Unsigned distance field

\[ D(p) \leftarrow C(P_\beta, z) \]

Corona et al., SMPLicit, CVPR’21
Dressing humans

Attributes

MLP

Corona et al., SMPLicit, CVPR’21
Dressing humans

Attributes

MLP

Corona et al., *SMPLicit*, CVPR’21
Interpolation in latent space

Attributes

MLP

Corona et al., SMPLicit, CVPR’21
Key advantages of SMPLicit

Interpolation

Interpolating clothing of different topology using single model with SMPLicit

TailorNet (vertex-based model) uses one model/garment type

Corona et al., SMPLicit, CVPR'21
Key advantages of SMPLicit

Fitting to scans

- Represents multiple topologies in one network
- No need to pre-define clothes and train independently per template

Reference person

[TailorNet, CVPR 2020]

SMPLicit

Corona et al., SMPLicit, CVPR’21
Model fitting with SMPLicit

Input image

Cloth Segmentation
[R-P-R-CNN, ECCV 2020]

Body Estimation
[FrankMocap, ICCVW 2021]

3D Reconstruction

Semantic Labels

Fitting SMPLicit by minimizing projection error

Corona et al., SMPLicit, CVPR'21
Model fitting with SMPLICIT

\[
\mathcal{L}_I(z) = \begin{cases} 
|C(P_\beta, z) - d_{\text{max}}|, & \text{if } s_p = 0 \\
\min_i |C(P^i_\beta, z)|, & \text{if } s_p = 1
\end{cases}
\]

Min over points along the ray

- $P_\beta$: Body relative representation of a sampled point in canonic space
- $s_p = 0$: Point projects outside segmentation mask → force to predict maximum distance or off-surface
- $s_p = 1$: Point projects inside segmentation mask → force to predict 0 distance (on-surface)
Combining the flexibility of implicit representations with the control of explicit parametric models

3D Reconstruction
- Represents multiple topologies
- Automatic training from a general dataset
- High-resolution details
Meshes vs Implicits

1) Parametric Meshes

- [Alldieck et al. CVPR’18]
- [Bhatnagar et al. ICCV’19, ECCV’20]
- [Tiwari et al. ECCV’20]

2) Implicit Functions

- [Chibane et al. CVPR’20]
- [Chibane et al. NeurIPS’20]

<table>
<thead>
<tr>
<th></th>
<th>Control /Meaning</th>
<th>Topology</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>1)</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>2)</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

2) Compatible with learning
More works on Human modeling using Neural Implicits

- SCANimate, Saito et al. CVPR’21
- LEAP, Mihajlovic et al. CVPR’21
- imGHUM, Alldieck et al. ICCV’21
- MetaAvatar, Weng et al. NeurIPS’21
- ICON, Xiu et al. CVPR’22
- PINA, Dong et al. CVPR’22
- AutoAvatar, Bai et al. ECCV’22

And Many more.....

NeuralFields
https://neuralfields.cs.brown.edu/
PART1: Neural Implicits for 3D Shapes

PART2: Neural Implicits for Humans

PART3: Neural Implicits – Generative Models

PART4: Point-based Clothing Models
# Meshes vs Implicits vs PointClouds

<table>
<thead>
<tr>
<th></th>
<th>Control /Meaning</th>
<th>Topology</th>
<th>Details</th>
<th>Speed</th>
<th>Continuous</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Meshes</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>2) Implicits</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>3) PointClouds</td>
<td>?</td>
<td>✓</td>
<td>?</td>
<td>✓</td>
<td>?</td>
</tr>
</tbody>
</table>
Human and clothing model using PointClouds

- High fidelity
- Flexible topology
- Pose/Shape/Style controllable
- Learned directly from scans
- Fast rendering
Human and clothing model using PointClouds

- High fidelity
- Flexible topology
- Pose/Shape/Style controllable
- Learned directly from scans
- Fast rendering

Ma et al. ICCV'21

Zakharkin et al. ICCV'21

Ma et al. 3DV'22
The Power of Points for Modeling Humans in Clothing

Using pointcloud for humans/clothing

Single Global Feature (Groueix et al., 2018)

Per-Patch Features (Ma et al., 2021)

Fine-grained Point Features (Ours)

Lack Details

“Patchy” Artifacts

Clean overall shape, sharp local details

Ma et al. ICCV’21
How is this different from prior Point-based works?

Using pointcloud for humans/clothing

- Single Global Feature (Groueix et al., 2018)
- Per-Patch Features (Ma et al., 2021)
- Fine-grained Point Features (Ours)

Lack Details
“Patchy” Artifacts
Clean overall shape, sharp local details

Ma et al. ICCV’21
The Power of Points for Modeling Humans in Clothing

Ma et al. ICCV'21
The Power of Points for Modeling Humans in Clothing

Ma et al. ICCV'21
The Power of Points for Modeling Humans in Clothing

Ma et al. ICCV’21

\[ \mathbf{p}_i, \quad \mathbf{I}_{u_i} = \mathbf{p}_i \in \mathbb{R}^3 \]

\[ \mathbf{z}_i^P, \quad \mathbf{z}_i^G, \quad \mathbf{u}_i \]

\[ \mathbf{T}_i \]

\[ \mathbf{r}_i \in \mathbb{R}^3 \]

Neutral Body Shape

Trainable Networks
Optimizable Geometric Features
Known Rigid Transform

Clothing on Per-Subject Body Shape
The Power of Points for Modeling Humans in Clothing

Ma et al. ICCV’21
The Power of Points for Modeling Humans in Clothing

\[ r_i = f_w(u_i; z_i) : \mathbb{R}^2 \times \mathbb{R}^Z \rightarrow \mathbb{R}^3, \]

Ma et al. ICCV'21
The Power of Points for Modeling Humans in Clothing

Ma et al. ICCV'21
The Power of Points for Modeling Humans in Clothing

Ma et al. ICCV'21
The Power of Points for Modeling Humans in Clothing

Results: POP produces high-quality and fine-detailed results than LBS
Limitations:
Discontinuity due to SMPL UV maps
- This results in visible “seams” between certain body parts.
- More significant for skirts.
Discontinuity in POP

Ma et al., 3DV 2022
# Meshes vs Implicits vs PointClouds

<table>
<thead>
<tr>
<th></th>
<th>Control /Meaning</th>
<th>Topology</th>
<th>Details</th>
<th>Speed</th>
<th>Continuous</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Meshes</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>2) Implicits</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>3) PointClouds</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
</tr>
</tbody>
</table>
Slides credit and resources

Thanks to
Julian Chibane, Enric Corona and Qianli Ma for providing materials.

TUM AI Lecture Series - Neural Implicit Representations for 3D Vision
(talk by Prof. Pons-Moll)