FIND: An Unsupervised Implicit 3D Model of Articulated Human Feet

Oliver Boyne, James Charles, Roberto Cipolla
University of Cambridge

Motivation

- Generative models of human bodies, hands and faces have been well developed
- Foot models are a relatively unexplored category - modelling feet is useful for shoe fitting and orthotics
- Challenging task due to limited available data

Contributions

- FIND (Foot Implicit Neural Deformation) model to generate explicit, textured feet with pose, shape and texture
- Unsupervised shape/pose disentanglement
- Unsupervised part-based learning
- Foot3D – Dataset of high resolution 3D foot scans

FIND Model

- Define template mesh
- Use latent codes $z_s$ (shape), $z_p$ (pose), $z_t$ (texture)
- Sample point $x$ on template surface
- Feed positional encoding $\gamma(x)$ through MLP $F$ to predict colour and displacement

$$F(\gamma(x), z_s, z_p, z_t) \rightarrow (\Delta x, c)$$

- Unsupervised pose representation learning
- Constraint: feet of same identity have same $z_s$
- Contrastive loss: similar poses have similar $z_p$; different poses have different $z_p$
- Resolution of template model chosen depending on task (eg low vertex count for mobile applications)

Learning 2D parts

- StyleGAN generates synthetic foot images
- Encode foot images to StyleGAN style codes
- k-means clustering on StyleGAN feature maps produces unsupervised 'part' segmentations
- Train classifier to predict these parts
- Fully differentiable image-to-parts pipeline (Figure 3)
- At train time, use pipeline to learn parts directly on template mesh of FIND
- For inference on 2D images, use cross entropy between image-to-parts pipeline and projected 3D FIND parts

Experimental Results

- Model evaluated by fitting to Foot3D validation scans, with 3D chamfer loss
- Quantitative results

<table>
<thead>
<tr>
<th>Model</th>
<th>Trained on</th>
<th>Chamfer $\mu m$</th>
<th>Keypoint, mm</th>
<th>IoU</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUPR</td>
<td>Foot3D</td>
<td>48.0</td>
<td>11.2</td>
<td>0.756</td>
</tr>
<tr>
<td>PCA</td>
<td>Foot3D</td>
<td>11.2</td>
<td>15.7</td>
<td>0.892</td>
</tr>
<tr>
<td>FIND</td>
<td>Foot3D</td>
<td>7.3</td>
<td>5.9</td>
<td>0.931</td>
</tr>
</tbody>
</table>

- Qualitative results

- Model renders
- Baselines
- FIND
- StyleGAN
- Unsupervised 'part' segmentations

Dataset • Code • Web demo

ollieboyne.github.io/FIND

Figure 1: 5 scans from Foot3D dataset with pose descriptions

To develop an accurate generative foot model, we contribute:

- FIND (Foot Implicit Neural Deformation) model to generate explicit, textured feet with pose, shape and texture
- Unsupervised shape/pose disentanglement
- Unsupervised part-based learning
- Foot3D – Dataset of high resolution 3D foot scans

Figure 2: FIND model overview

Figure 3: Pipeline for predicting per-pixel classes from an input image

Figure 4: Renders of our model, and baselines, optimised in 3D to fit to our ground truth scans