PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization

Peiyuan Zhu

March 31, 2020

1 Main contribution

1. Proposed an end-to-end deep learning framework for digitizing clothed human with high resolutions

2. This framework can be generally applied to inferring both 3-D surface and texture, and can take arbitrary number of perspective images as input

3. Achieved state-of-the-art performance on RenderPeople and Buff datasets

2 Strength point

1. Compared with voxel representation, using implicit function to represent the surface is much more memory efficient

2. This allows the use of fully-convolutional network that preserves spatial alignment between the image and the output

3 Weakness point

1. Need to to exhaustively sample points from the 3-D space and decide whether it is inside or outside the surface mesh

2. From the result, it seems that the extrapolation of the clothes on unseen area of the human were flawed with mixing them up with thje human skins and the textures weren't exactly preserved.

3. From the result of the video, the body shape of a human is far from realistic when the body is behind the human.

4 Questions

1. Is it possible to perform segmentation first to distinguish human bodies from clothes, then reconstruct them with separate implicit function networks?

2. Can we represent the surface as a differential equation and use neural network to learn the differential operator? When the views are limited, extrapolation by differential equation can be a better way.

3. For more than three views, should we incorporate some dependencies between views instead of average pooling? This might give better prediction for the unseen area.