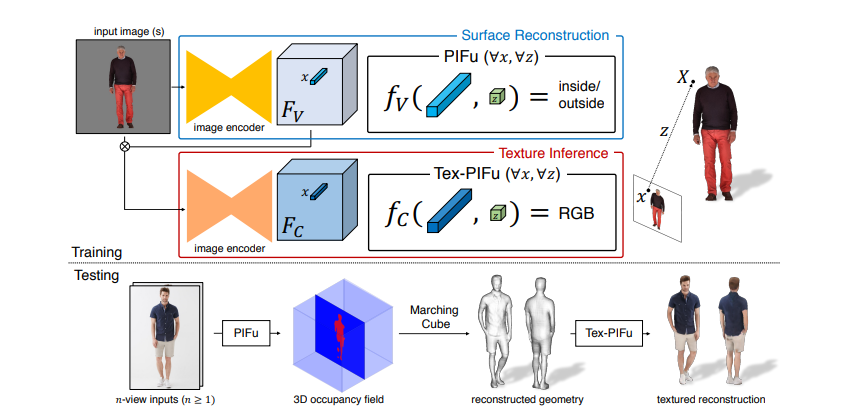
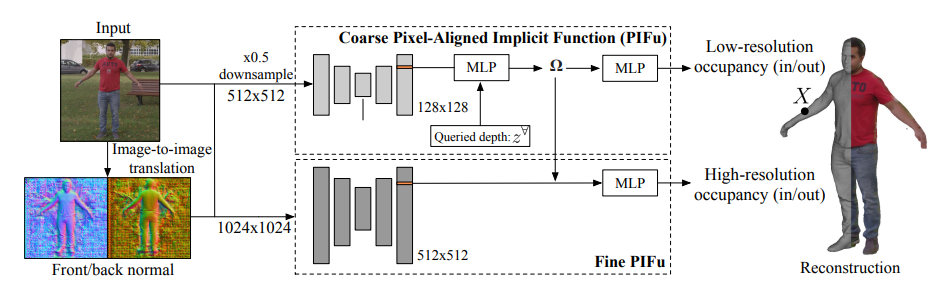
1. System pipeline





2. Dataset:

- Deepfashion Dataset (~800K diverse fashion images, 46 categories, 1,000 descriptive attributes, bounding boxes and landmark information)

- BUFF Dataset (5 videos, 360-degrees view video)

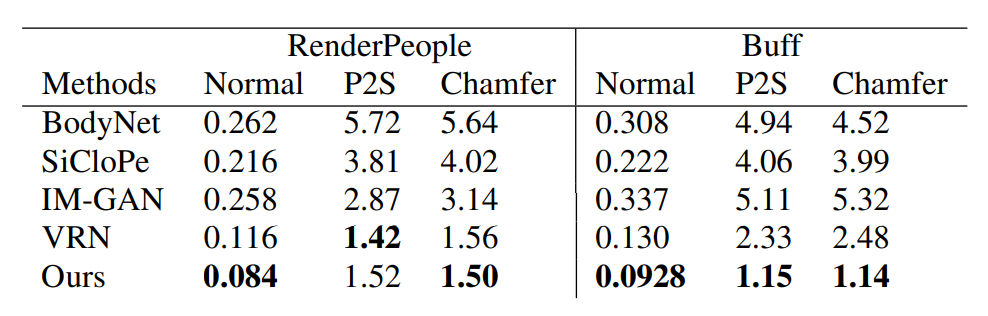
- RenderPeople Dataset (~719 Posed People)

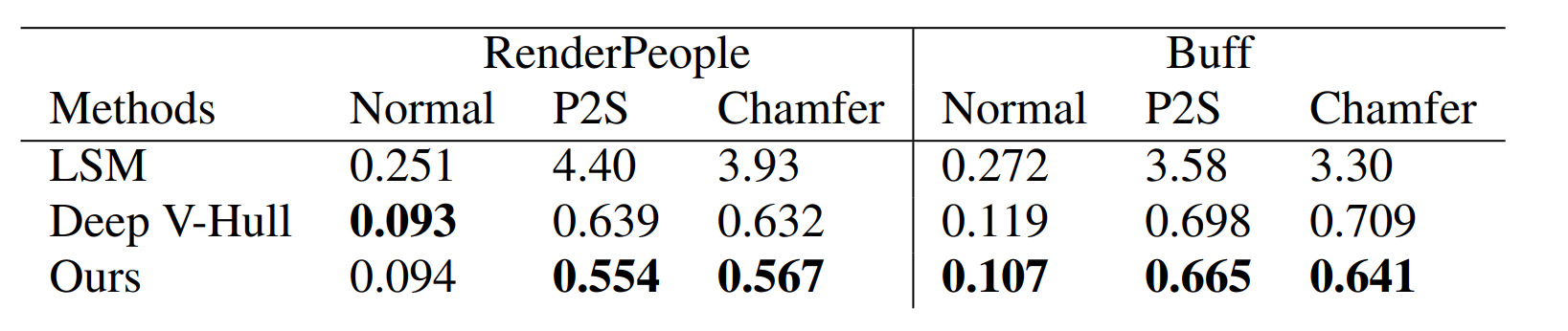
3. Measure:

Average Point-to-surface (P2S) in cm

Chamfer distance between recontructed & ground truth surfaces in cm

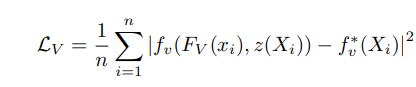
Normal = L2 error between normal maps of recontruced & ground truth surfaces





Quantitative evaluation on RenderPeople recontruction

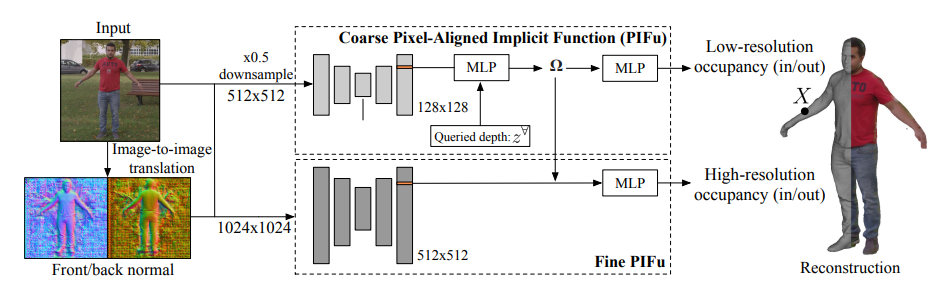
We train a pixel-aligned implicit function (PIFu) fv by minimizing the average of mean squared error:



where Xi ∈ R^3, Fv (x) = g(I(x)) is the image feature from the image encoder g at x = π(X) and n is the number of sampled points.

Our method builds on the recently introduced Pixelaligned Implicit Function (PIFu)

we first predict normal maps for the front and back sides in image space, and feed these to the network as additional input



\*pix2pixHD network used for img translation