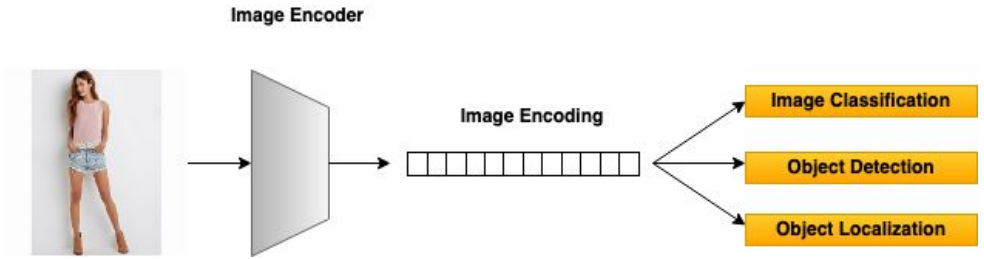
Hierarchical Part based Pose and Shape Disentanglement

Farnoosh Javadi

Summer 2020

Motivation

- Learning Image Representation is a crucial task in Computer Vision.

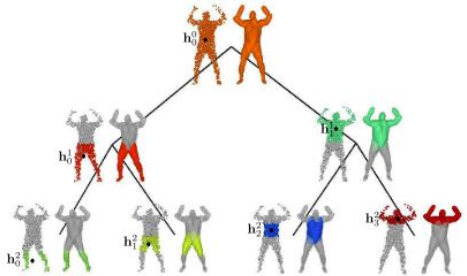


- The more interpretable representation, the better.
  - Easier to understand, evaluate.
  - Enables Novel Image Synthesis.

Our Goal

- Learn interpretable representations
  - Learn parts that form an object

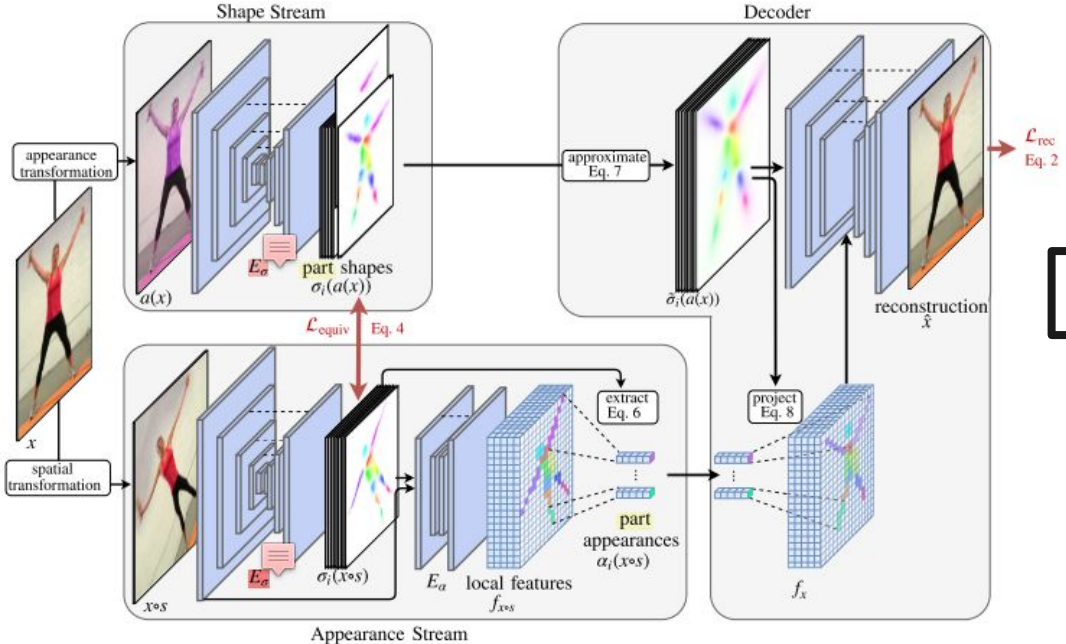
  - Learn the disentangled appearance and pose for each part
  - Learn the hierarchy of parts
- Learned representations could be used in
  - Landmark Discovery
  - Video to Video Translation
  - Pose and Appearance Transfer



Related Work

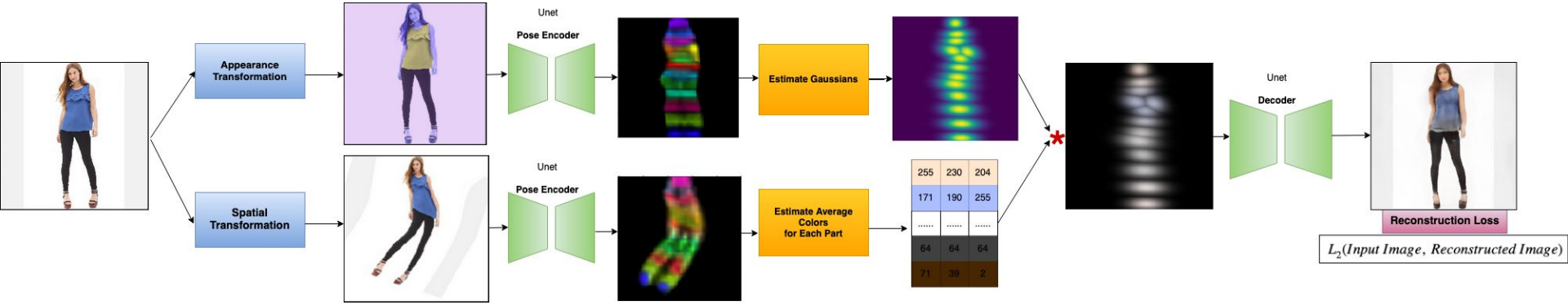
- Shape-supervised approaches
  - Conditioning generative models on shape information.
- Unsupervised disentangling approaches
  - Using holistic models.
- Supervised structure-aware representations
  - Needing supervision in terms of segmentation of objects into their primitive parts or the hierarchies.
- Unsupervised part-based representations
  - Learned in discriminative tasks.

Unsupervised Part-Based Disentangling of Object Shape and Appearance



$$\mathcal{L} = \mathcal{L}_{\text{rec}} + \mathcal{L}_{\text{equiv}}.$$

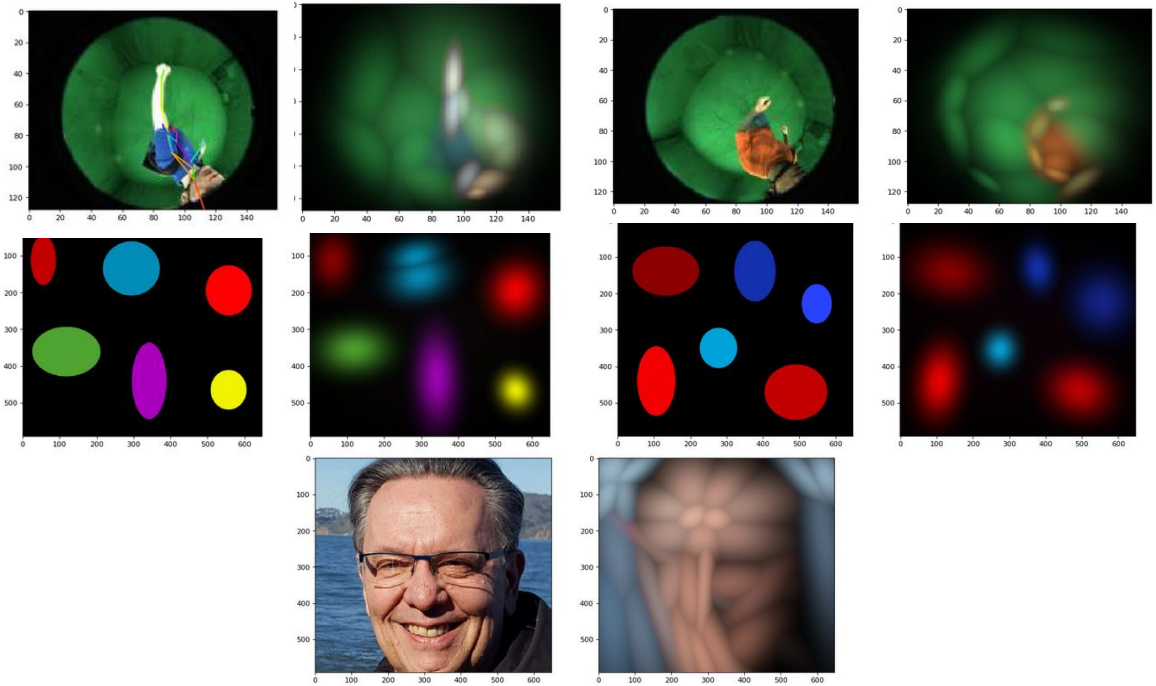
Our Model


Appearance Transformation
Spatial Transformation
Unet
Pose Encoder
Unet
Pose Encoder
Estimate Gaussians
Estimate Average Colors for Each Part
255 230 204
171 190 255
64 64 64
71 39 2
Unet
Decoder
Reconstruction Loss
$L_2(Input\ Image, Reconstructed\ Image)$

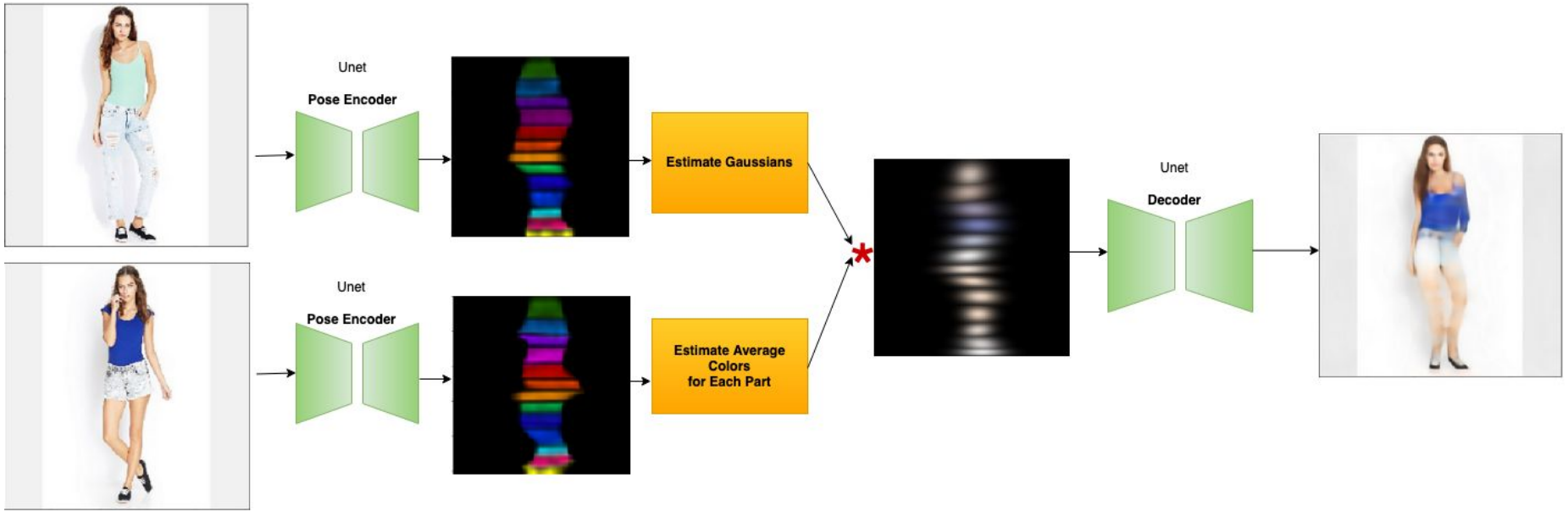
Transformations

- Appearance Transformation
  - Shift H in HSV space
  - Mix pixels' color with a random color in RGB space
    - New_color = a * old_color + (1-a) * random_color
- Spatial Transformation
  - Linearly combine 7 predefined TPS transformations
  - Rotate images

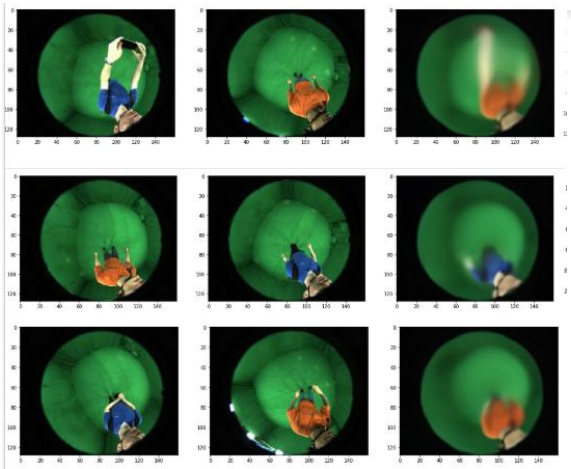
Results



Pose and Appearance Transfer



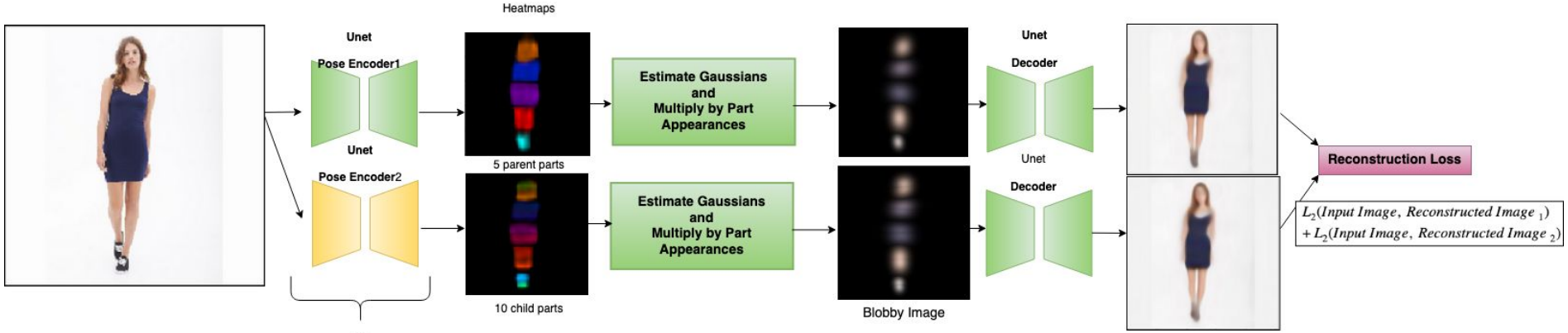
Results



Results

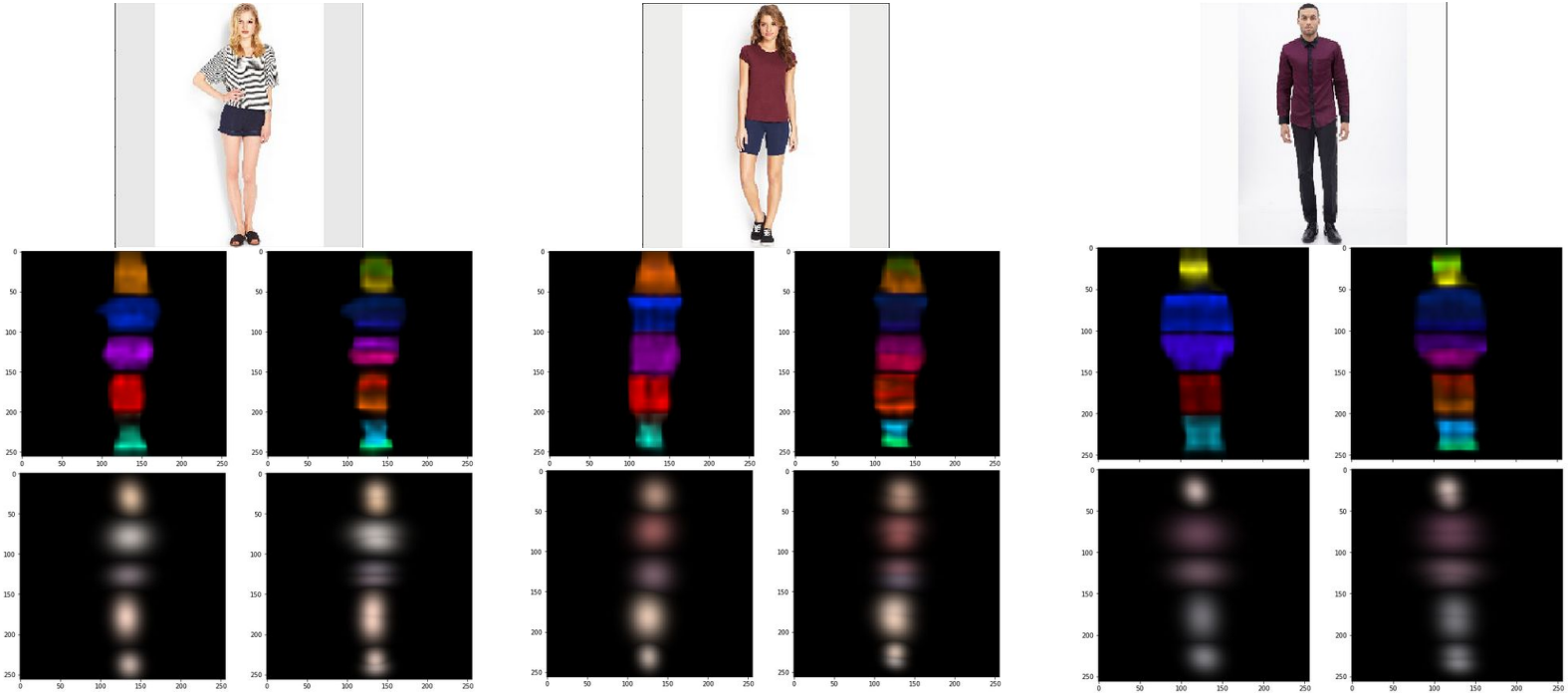
Results

Hierarchical Model



Similarity Loss = $-\sum_{i=1}^{5} H1_i \cdot H2_{2i} + H1_i \cdot H2_{2i+1}$

Results



Future Works

- Increase levels of the hierarchy
- Try different loss functions for enforcing the structure
  - $-\sum_{i=1}^{5} G1_i \cdot G2_{2i} + G1_i \cdot G2_{2i+1}$
  - $\frac{1}{5}\sum_{i=1}^{5} \frac{1}{2}\left((C1_i - C2_{2i})^2 + (C1_i - C2_{2i+1})^2\right)$
  - $\frac{1}{5}\sum_{i=1}^{5} \left(C1_i - \left(\frac{C2_{2i}+C2_{2i+1}}{2}\right)\right)^2$
- Make the networks deeper to prevent blurry images
- Try different batch-sizes

Thanks for your attention!