Research Center for Technology and Art

Seminar Presentation

基於生成對抗式網路之虛擬服裝穿戴合成技術

Virtual Try-on Image Synthesis Based on Generative Adversarial Networks

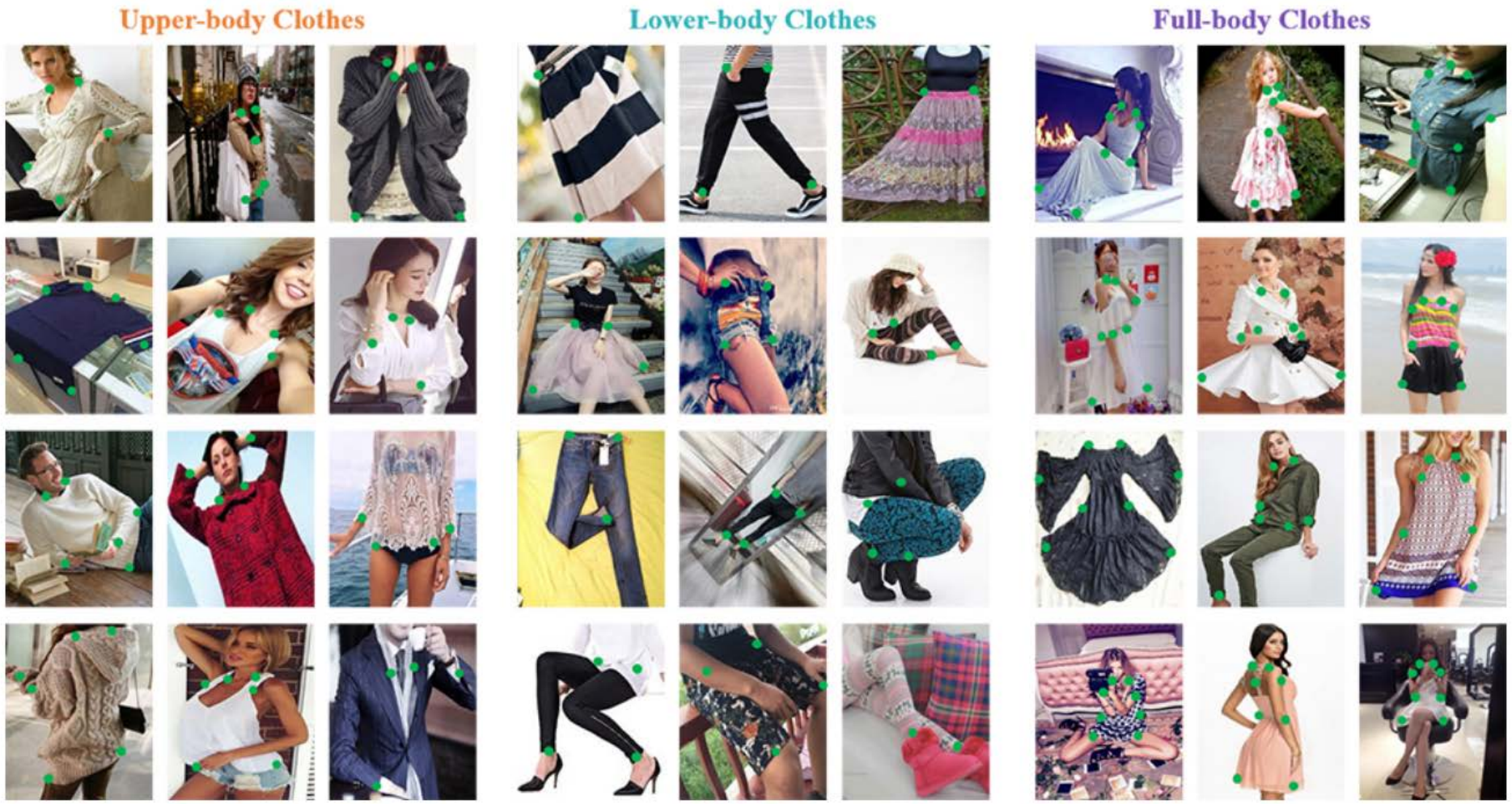
2019-01-31, NTHU, Department of Electrical Engineering

Reporter: NTHU IPHD 楊元福

Project Introduction

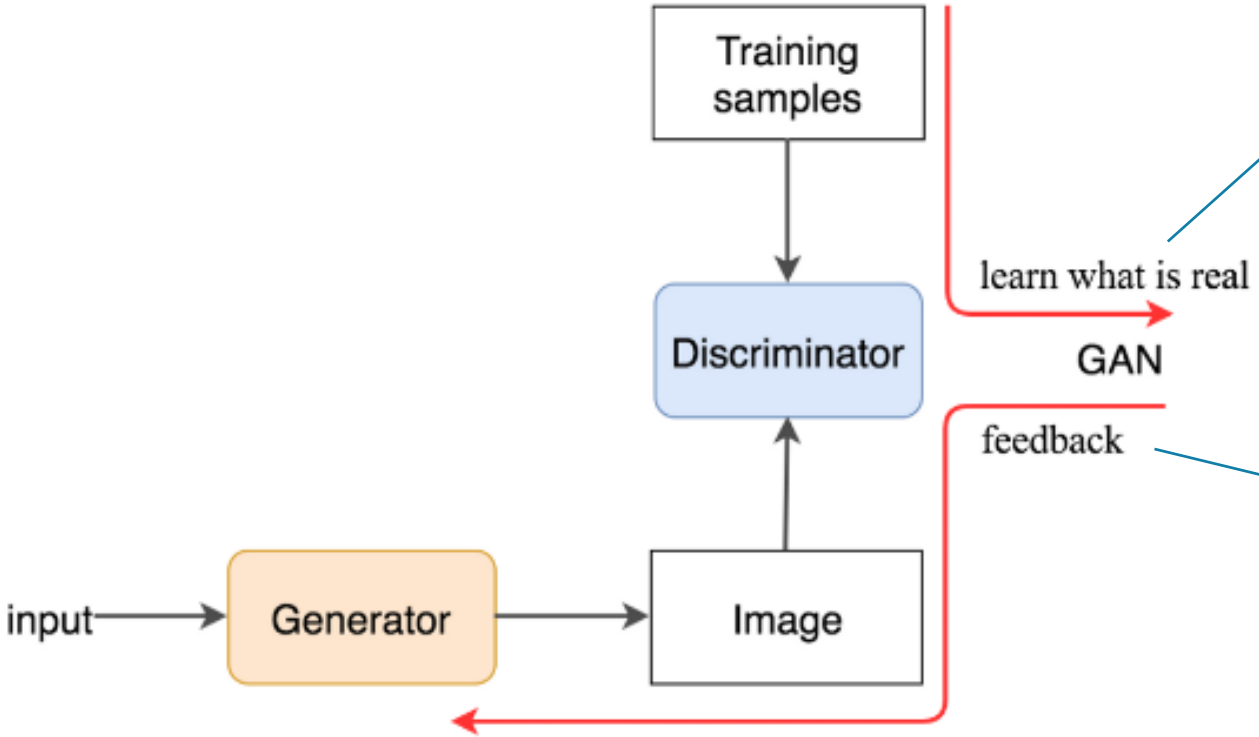
This study proposed a network which can redress the person in input image with the target clothes. Through the virtual try-on model, the clothes should perfectly fit human body in output image, the model wearing the assigned clothes with clothing detail well preserved. Also, the image would still keep the images photo-realistic.

Background – Deep Fashion Datasets



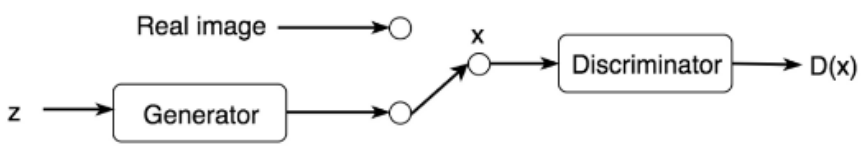
http://mmlab.ie.cuhk.edu.hk/projects/DeepFashion.html

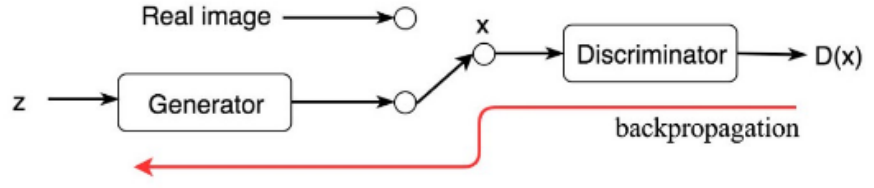
Methodology: Generative Adversarial Networks (GAN)



$$\max_D V(D) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]$$

recognize real images better | recognize generated images better





$$\min_G V(G) = \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]$$

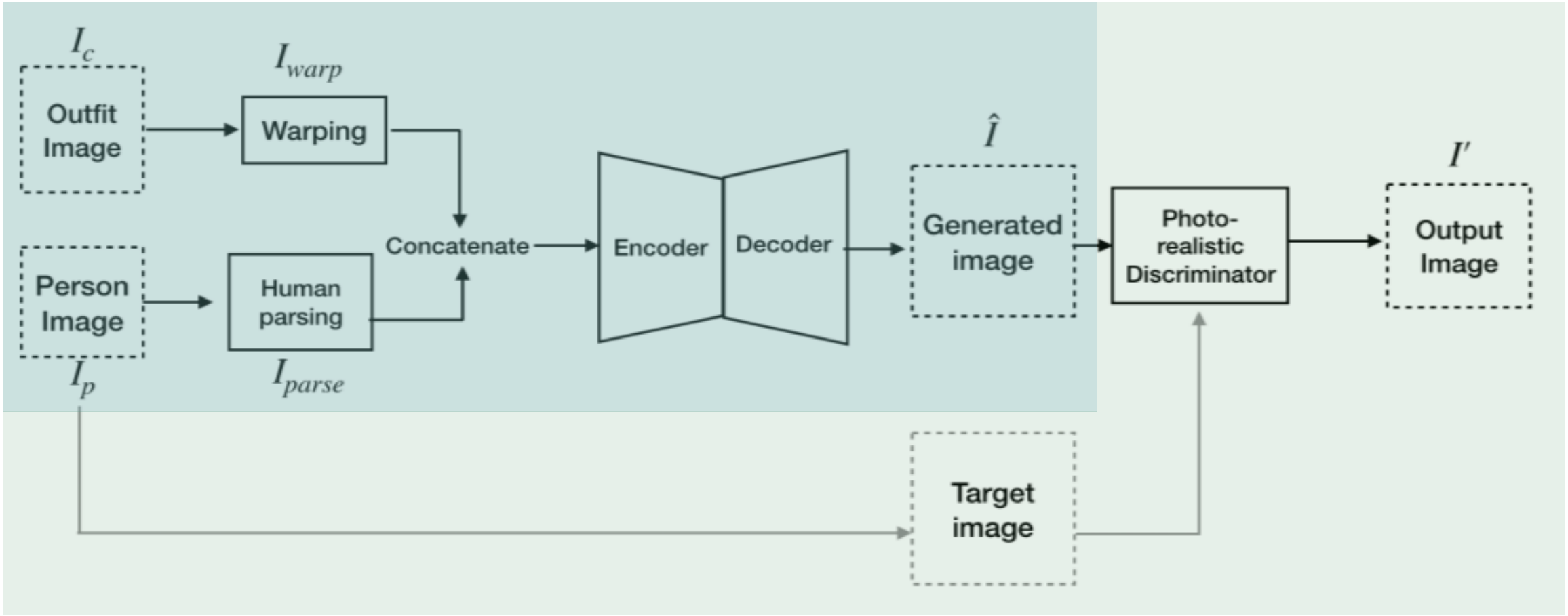
Optimize G that can fool the discriminator the most.

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))].$$

Proposed Model

Multi-task Encoder-Decoder Generator

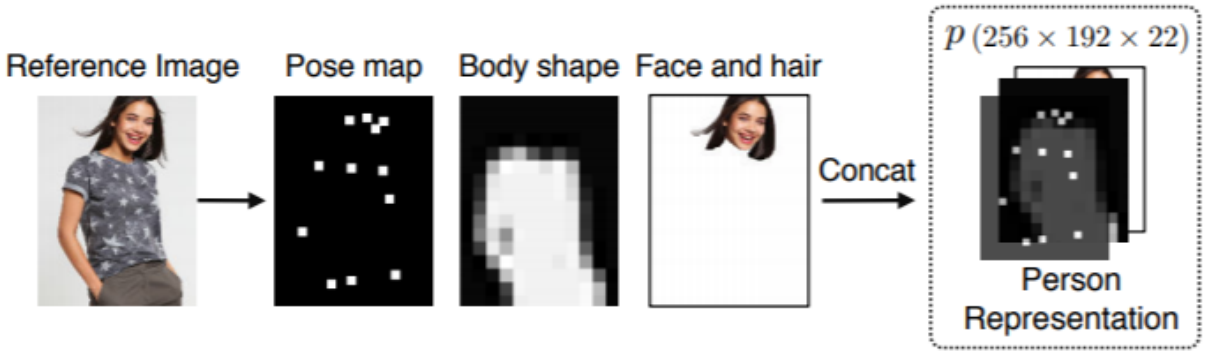
Refinement Network

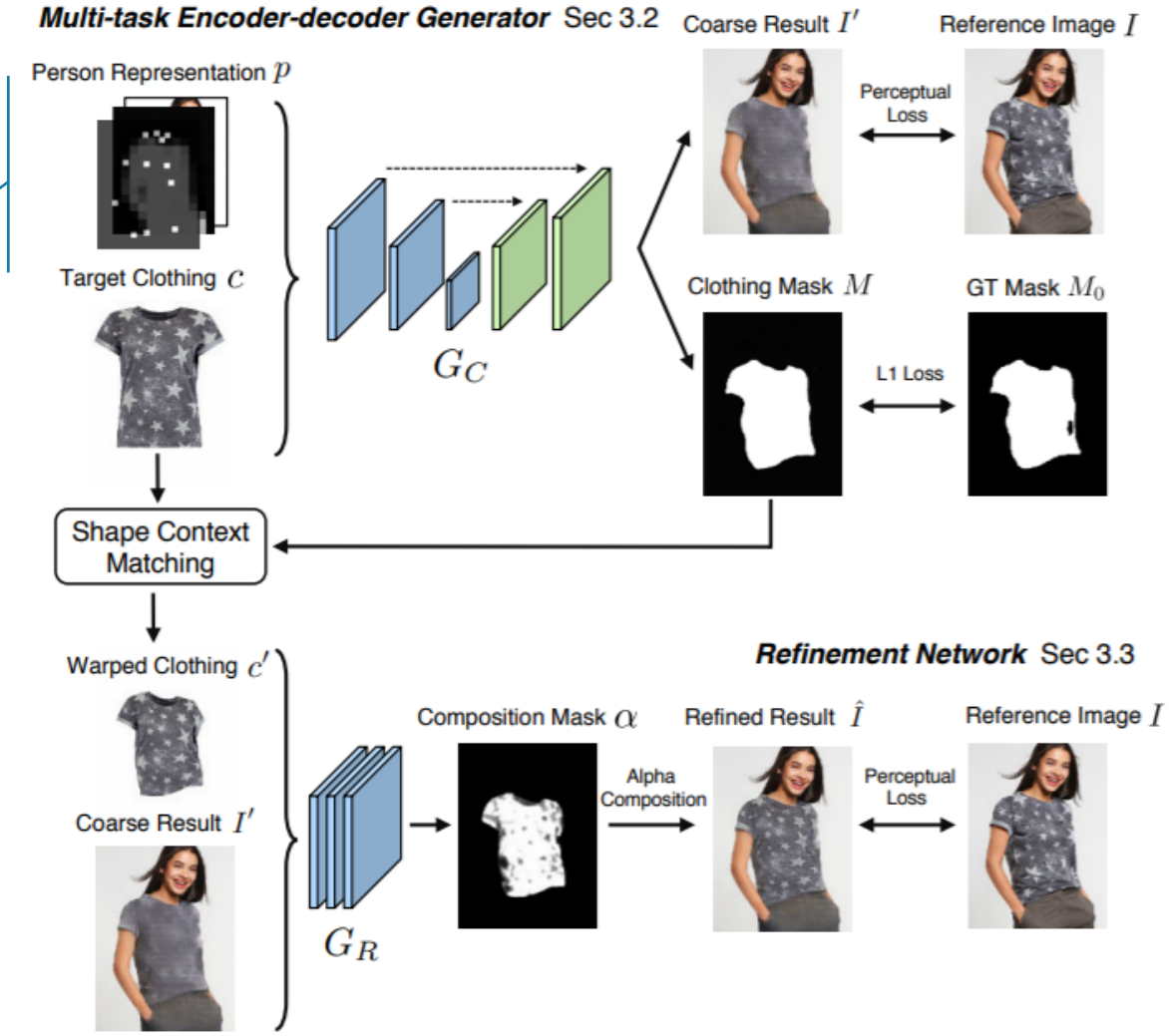


Multi-task Encoder-Decoder Generator

A main technical challenge of a virtual try-on synthesis is to deform the target clothing image to fit the pose of a person. To this end, we introduce a clothing-agnostic person representation, which contains a set of features, including **pose**, **body parts, face** and **hair,** as a prior to constrain the synthesis process.

$$L_{G_C} = \sum_{i=0}^{5} \lambda_i ||\phi_i(I') - \phi_i(I)||_1 + ||M - M_0||_1,$$





Refinement Network

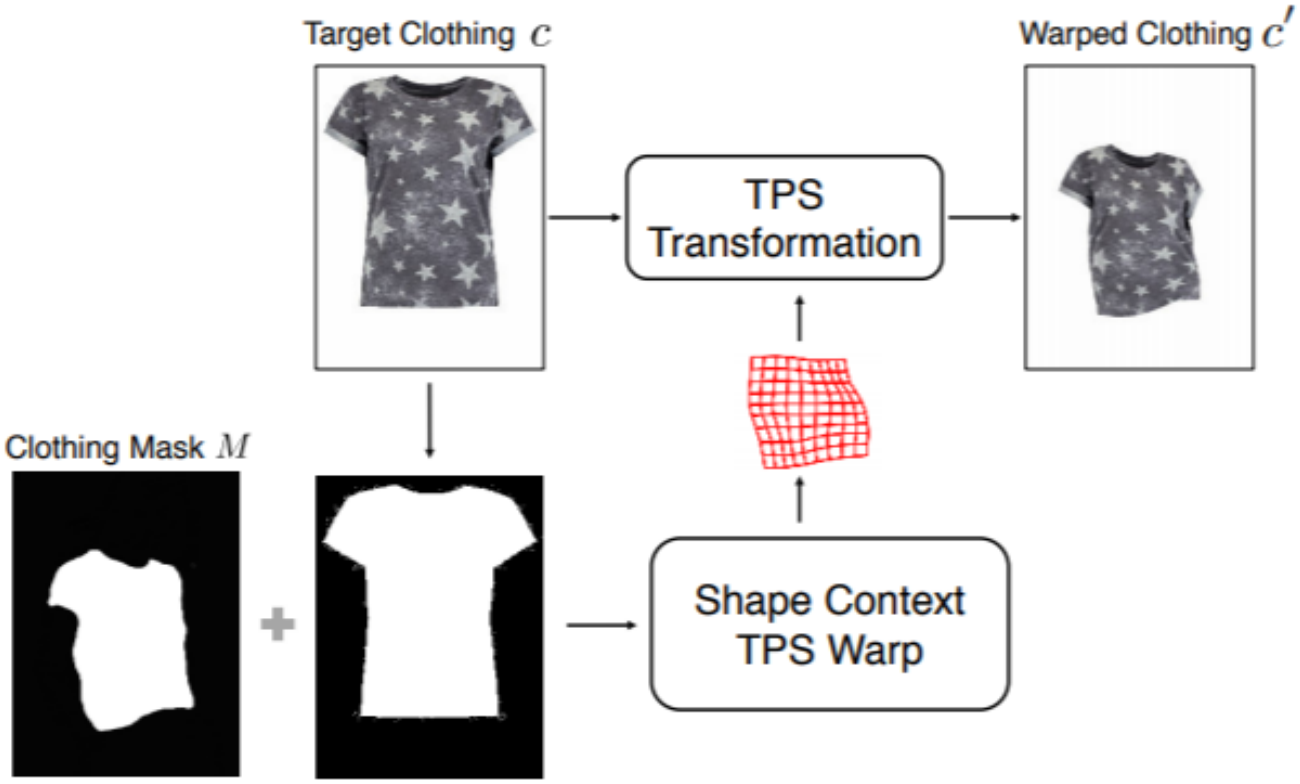
Warping a clothing image. Given the target clothing image and a clothing mask predicted in the first stage, we use shape context matching to estimate the TPS (Thin-Plate Spline) transformation and generate a warped clothing image.

synthesized image:

$$\hat{I} = \alpha \odot c' + (1 - \alpha) \odot I',$$

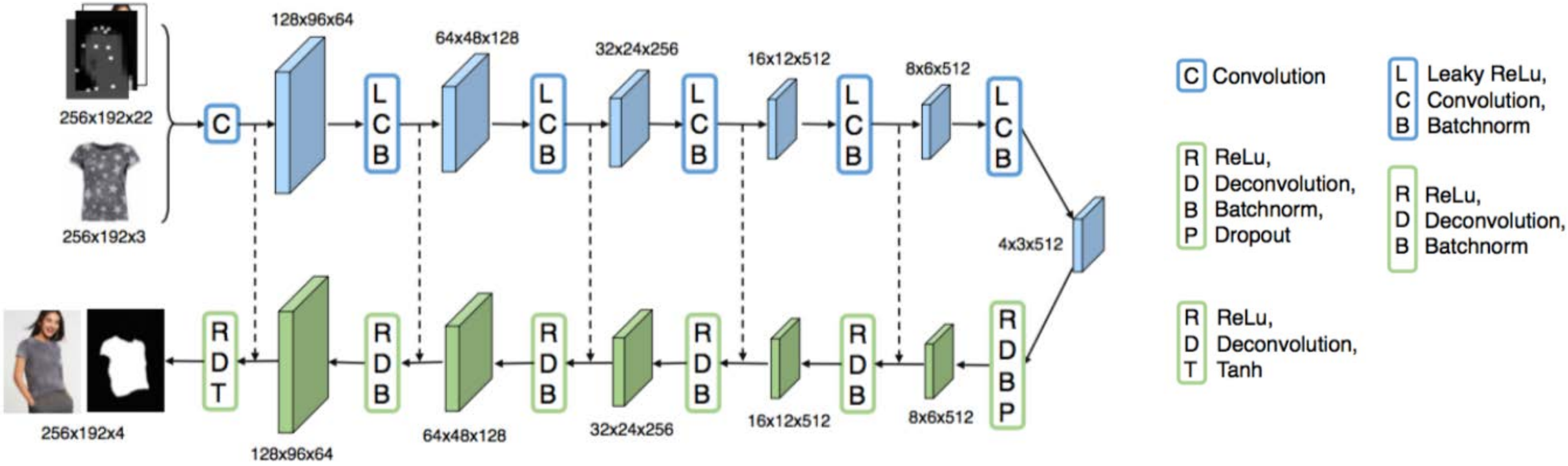
Refinement Network loss function：

$$L_{G_R} = L_{perc}(\hat{I}, I) - \lambda_{warp}||\alpha||_1 + \lambda_{TV}||\nabla\alpha||_1,$$



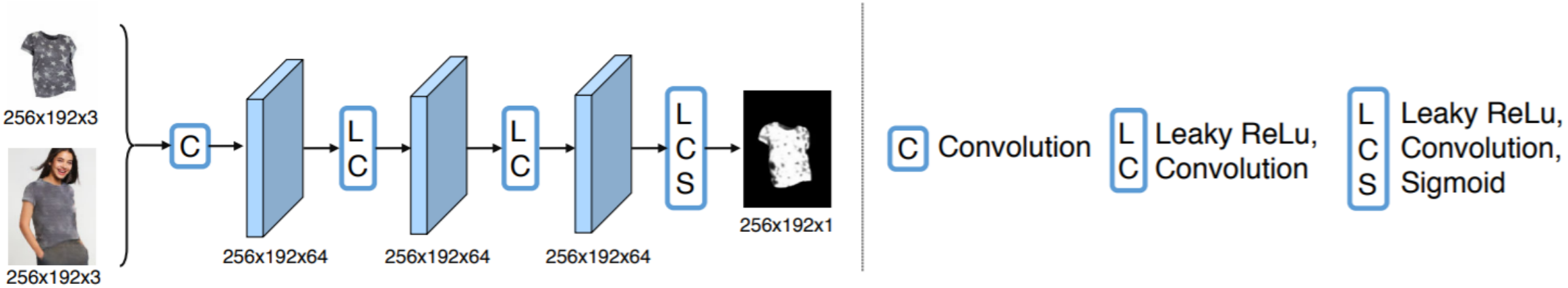
Implementation Details

Network structure of our encoder-decoder generator. Blue rectangles indicate the encoding layers and green ones are the decoding layers. Convolution denotes 4×4 convolution with a stride of 2. The negative slope of Leaky ReLu is 0.2. Deconvolution denotes 4×4 convolution with a stride of 1/2. The dropout probability is set to 0.5



Implementation Details

Network structure of our refinement network. Convolution denotes 3×3 convolution with a stride of 1. The negative slope of Leaky ReLu is 0.2.



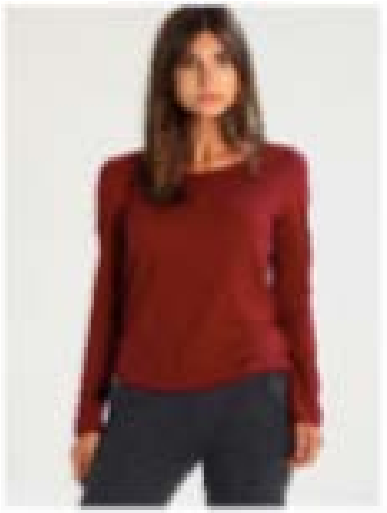
Comparisons

Compared Approaches

- GANs with Person Representation (PRGAN). Compared with VITON, there is no additional Refinement Network, the specific principle is similar to Multi-task Encoder-Decoder Generator, but the information of Face and hair segment is not considered when obtaining the person representation.
- Conditional Analogy GAN (CAGAN). Constructing CGAN directly for portraits and clothing pictures, but because the person representation information is not extracted, the effect of CAGAN in actual tests will not be ideal.
- Cascaded Refinement Network (CRN). Instead of using adversarial networks, the results can be obtained directly by constructing multiple layers of CNNs.
- Encoder-decoder generator. The output of VITON in the first stage, due to the limitations of ordinary CGAN, the obtained image will lose the logo and other details.
- Non-parametric warped synthesis. Without using the Refinement Network, the stretched clothing image is directly attached to the original portrait, skipping the learning process and leaving a lot of gaps in the generated image.

Connection

脖圍標簽陰影問題

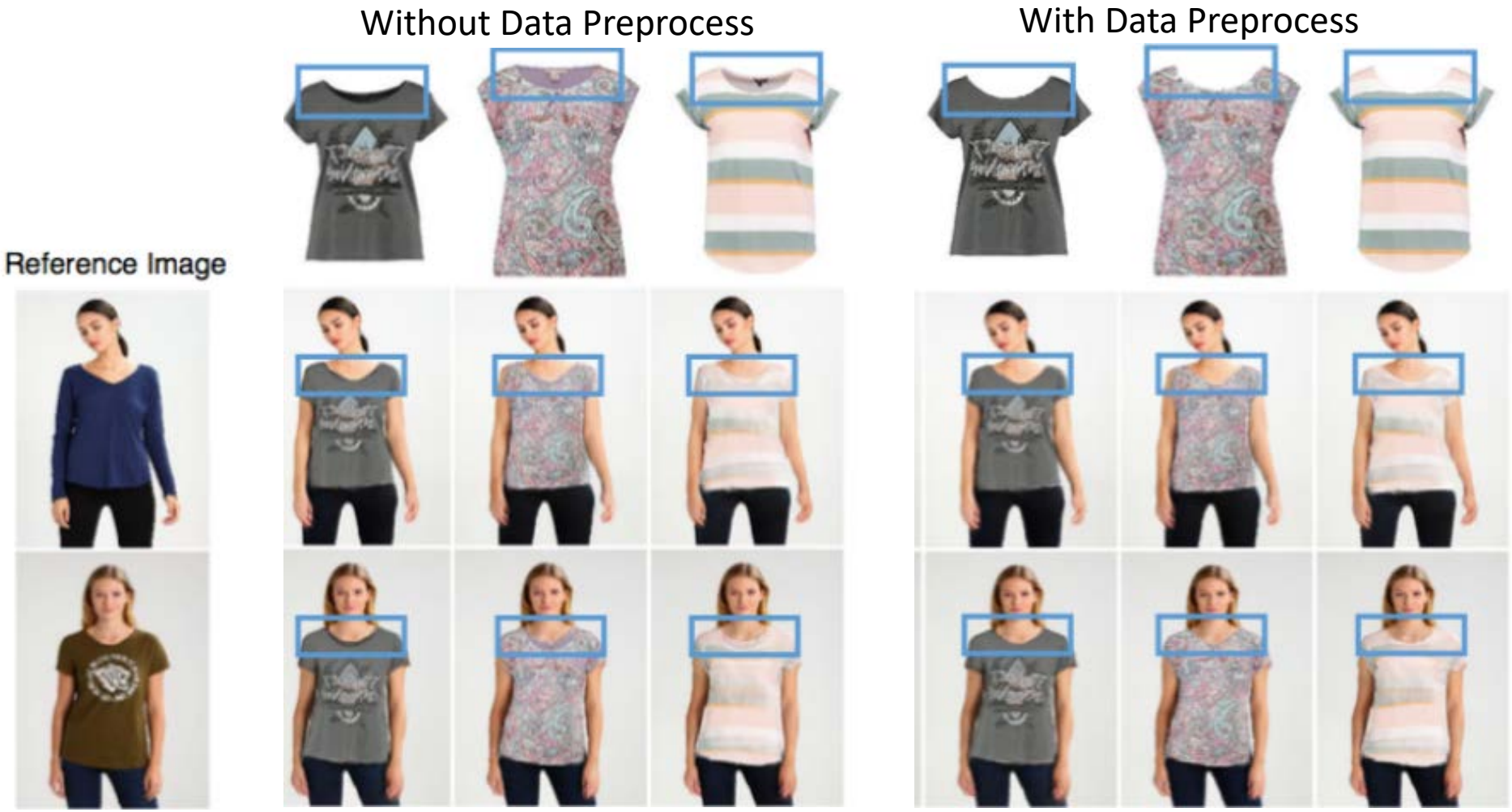


脖圍衣領形狀問題

Connection: 脖圍標簽陰影問題

Solution: Data Preprocess

Reference: " *An Image-based Virtual Try-on Network* ", Xintong Han, *CVPR 2018*



Connection: 脖圍衣領形狀問題

Solution: Data Preprocess

Reference: " *An Image-based Virtual Try-on Network* ", Xintong Han, *CVPR 2018*

Connection

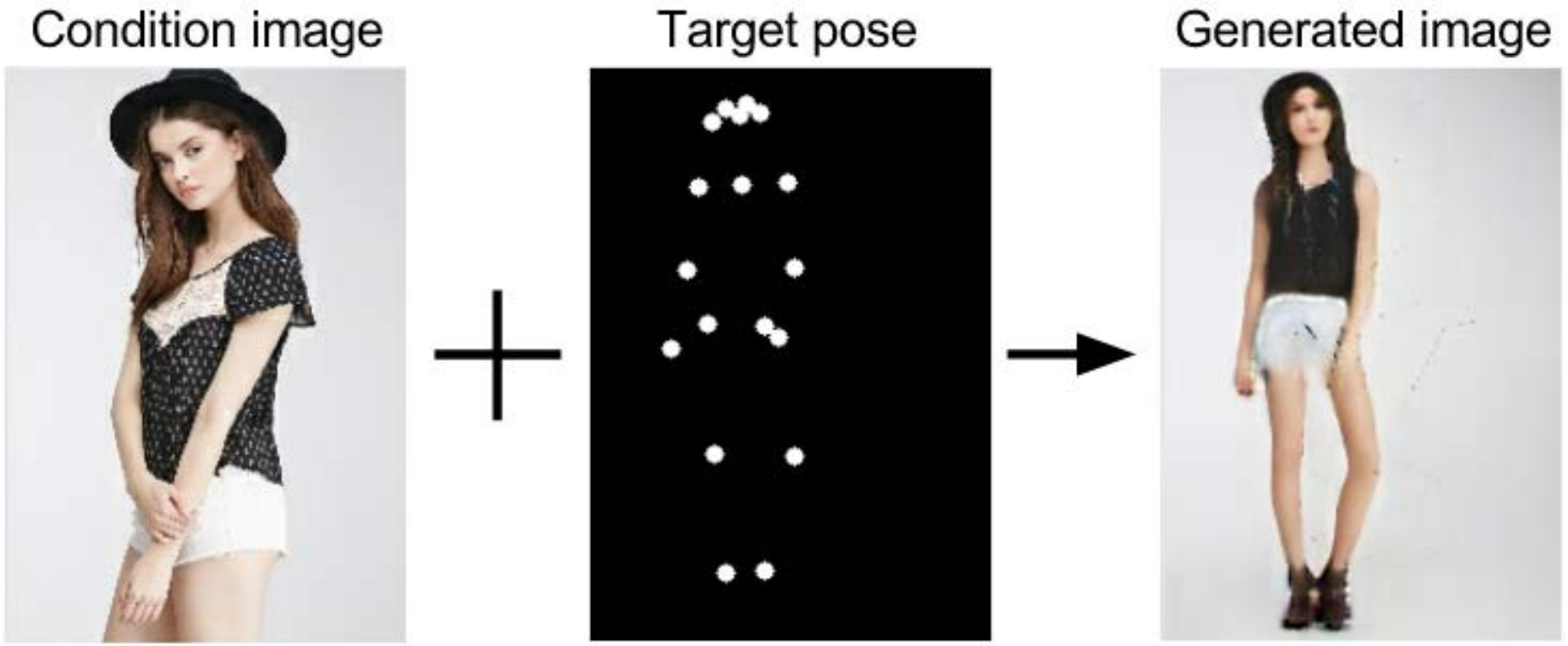
特殊紋理/透明度問題: 目前無解






Connection

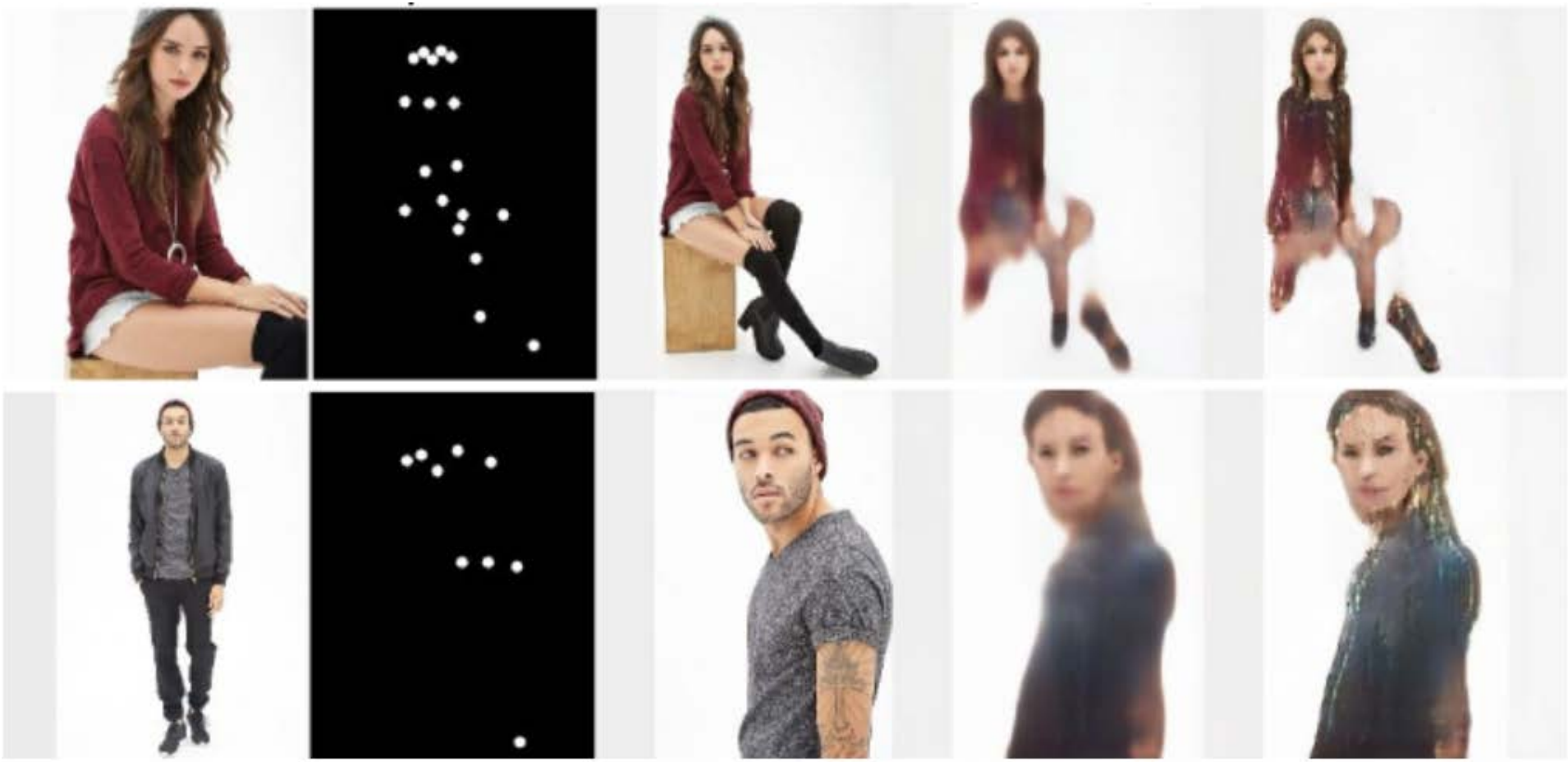
Pose Learning



Pose Guided Person Image Generation, NIPS 2017

Connection

Pose Learning



Pose Guided Person Image Generation, NIPS 2017