Toward Characteristic-Preserving Image-based Virtual Try-On Network

ECCV2018

2019.06.04

발표자 박성현

1 Introduction

Virtual Try-on

[Virtual Try-on]

1 Introduction

Background - Virtual Try-on Network (VITON)

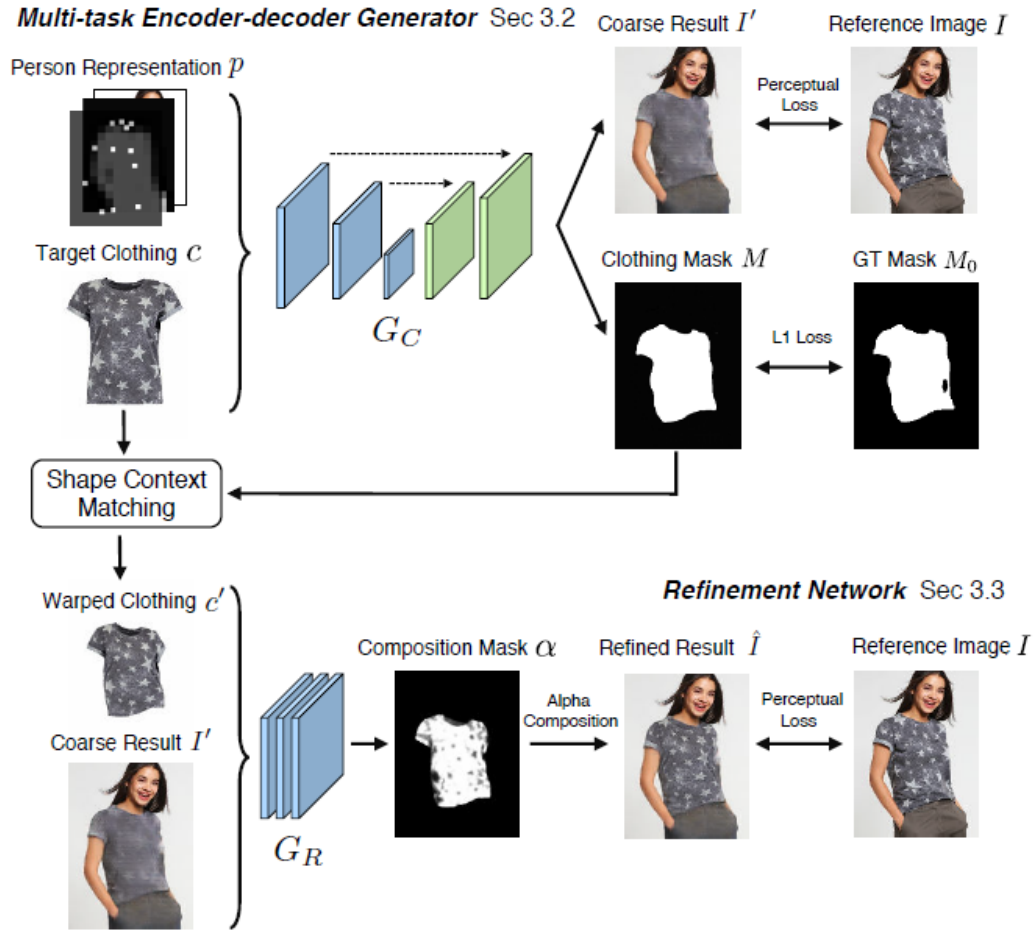


Figure 2: **An overview of VITON**. VITON consists of two stages: (a) an encoder-decoder generator stage (Sec 3.2), and (b) a refinement stage (Sec 3.3).

1 Introduction

Motivation

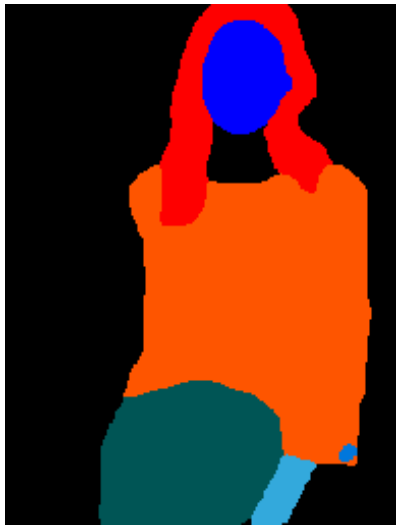
→ More realistic virtual try-on results
that preserve well key characteristics of the clothes

2 Model

Overview of the proposed model

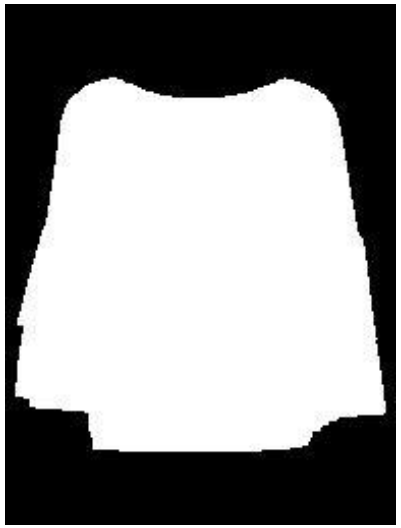
1. Person Representation
2. Geometric Matching Module
3. Try-On Module

[Image]



[Image-parse]

[Cloth]



[Cloth-mask]

2 Model

Overview of the proposed model

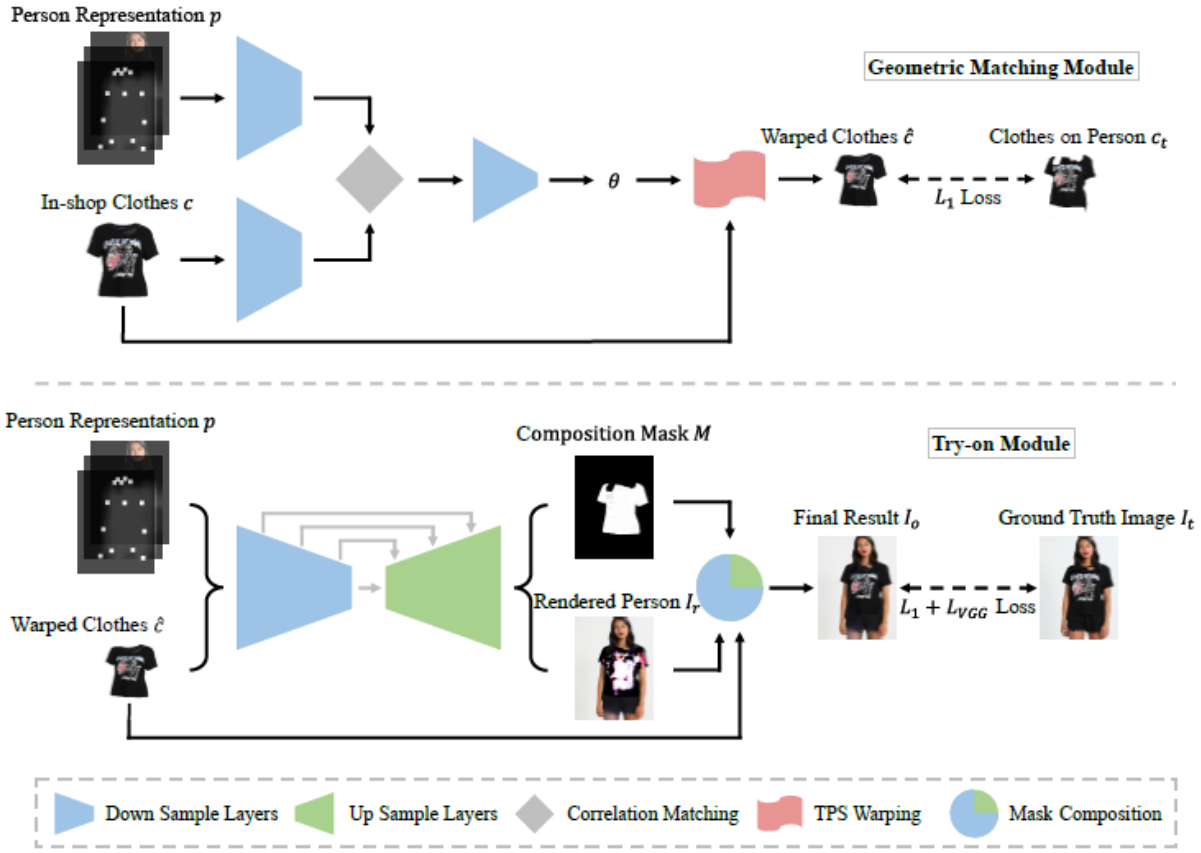
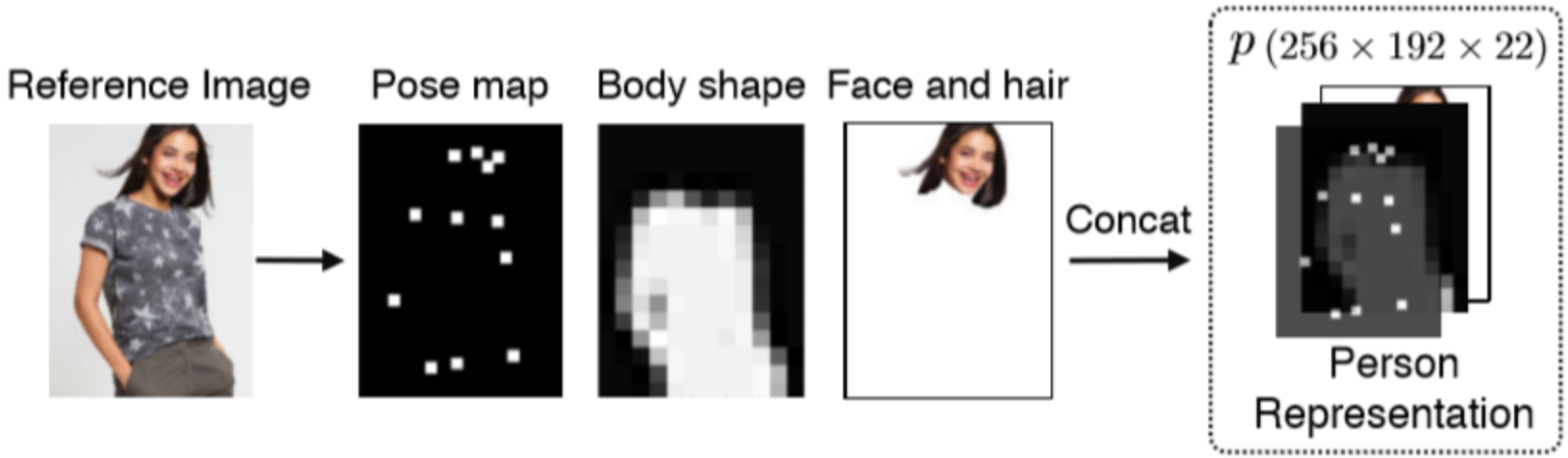


Fig. 2. An overview of our CP-VTON, containing two main modules. (a) Geometric Matching Module: the in-shop clothes c and input image representation p are aligned via a learnable matching module. (b) Try-On Module: it generates a composition mask M and a rendered person I_r. The final results I_o is composed by warped clothes $\hat{c}$ and the rendered person I_r with the composition mask M.

2 Model

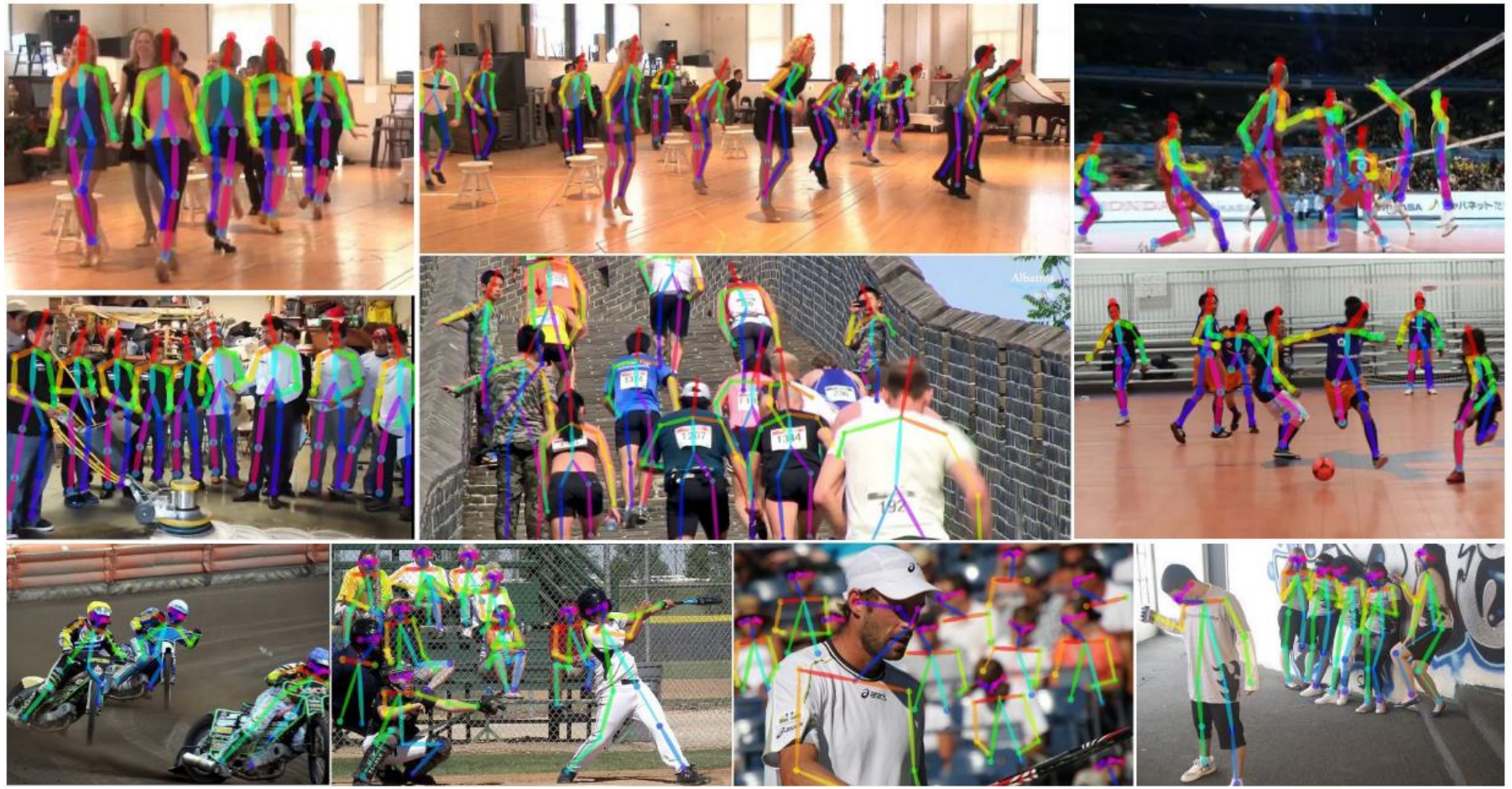
Person Representation

- **Pose heatmap** : a 18 channel feature map with each channel corresponding to one human pose keypoint, drawn as an 11x11 white rectangle
- **Body shape** : a 1-channel feature map of a blurred binary mask that roughly covering different parts of human body
- **Reserved regions** : a RGB image that contains the reserved regions to maintain the identity of a person, including face and hair



2 Model

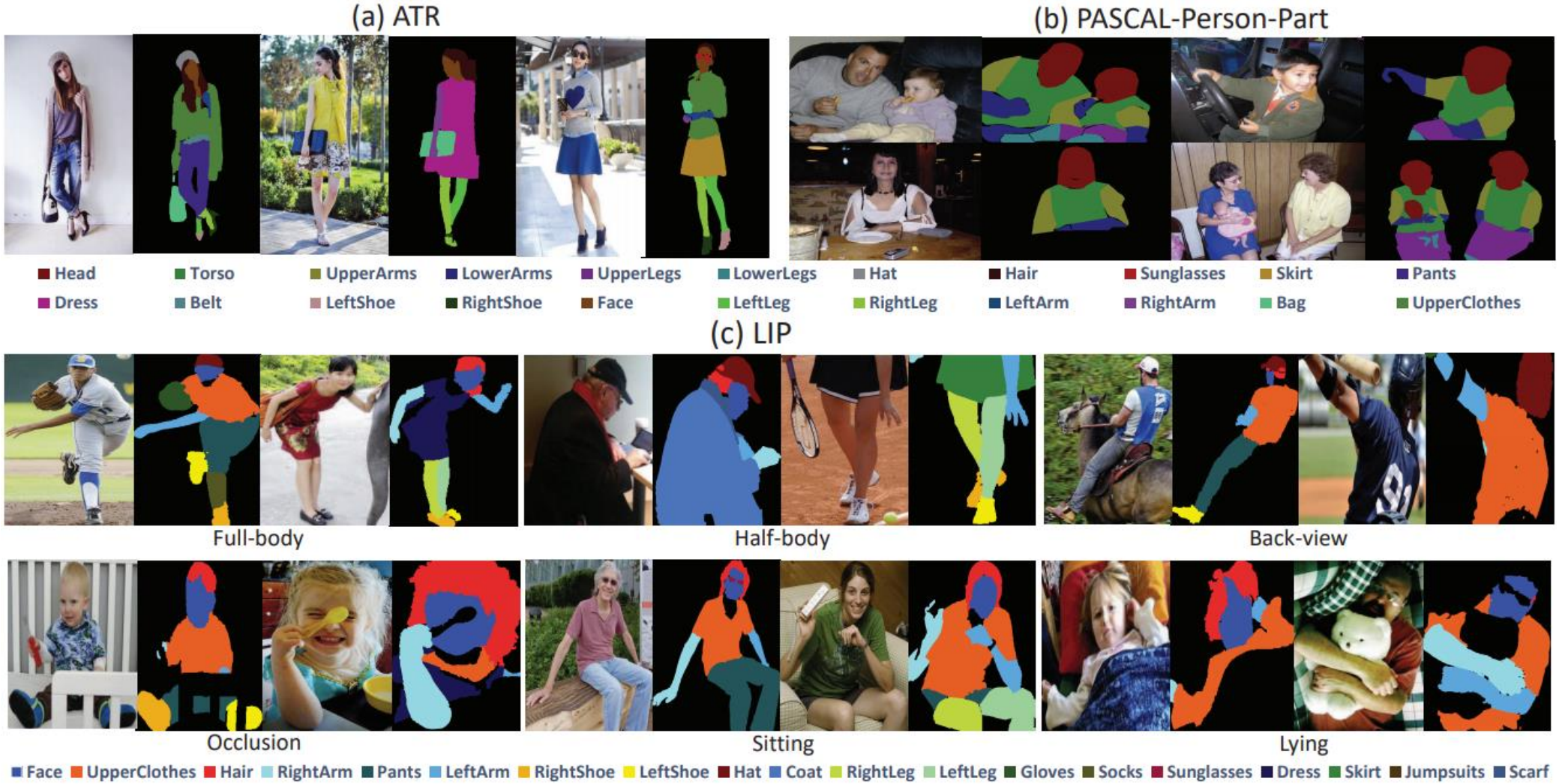
Person Representation



[Real-time Multi-Person Pose Estimation]

2 Model

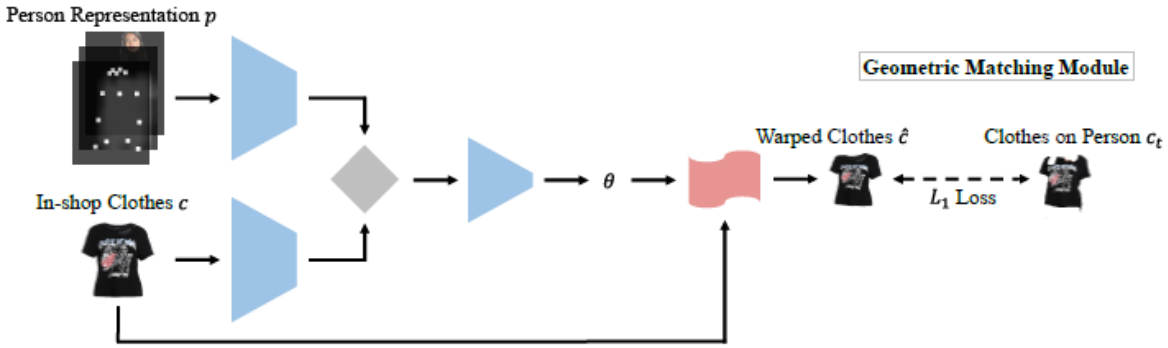
Person Representation



[Self-supervised Structure-sensitive Learning]

2 Model

Geometric Module



(1) Two networks for extracting high-level features of p and c respectively

(2) Correlation layer to combine two features into a single tensor as input to the regression network

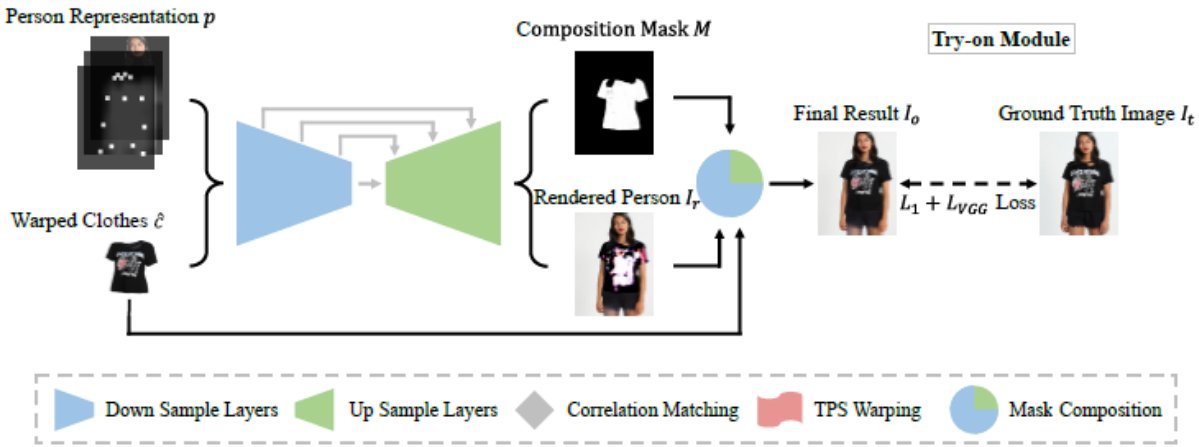
(3) The regression network for predicting the spatial transformation parameters θ

(4) Thin-Plate Spline (TPS) transformation module T for warping an image into the output $\hat{c} = T_\theta(c)$

$$\mathcal{L}_{GMM}(\theta) = ||\hat{c} - c_t||_1 = ||T_\theta(c) - c_t||_1$$

2 Model

Try-On Module



Given a concatenated input of person representation p and the warped clothes $\hat{c}$, UNet simultaneously renders a person image I_r and predicts a composition mask M.

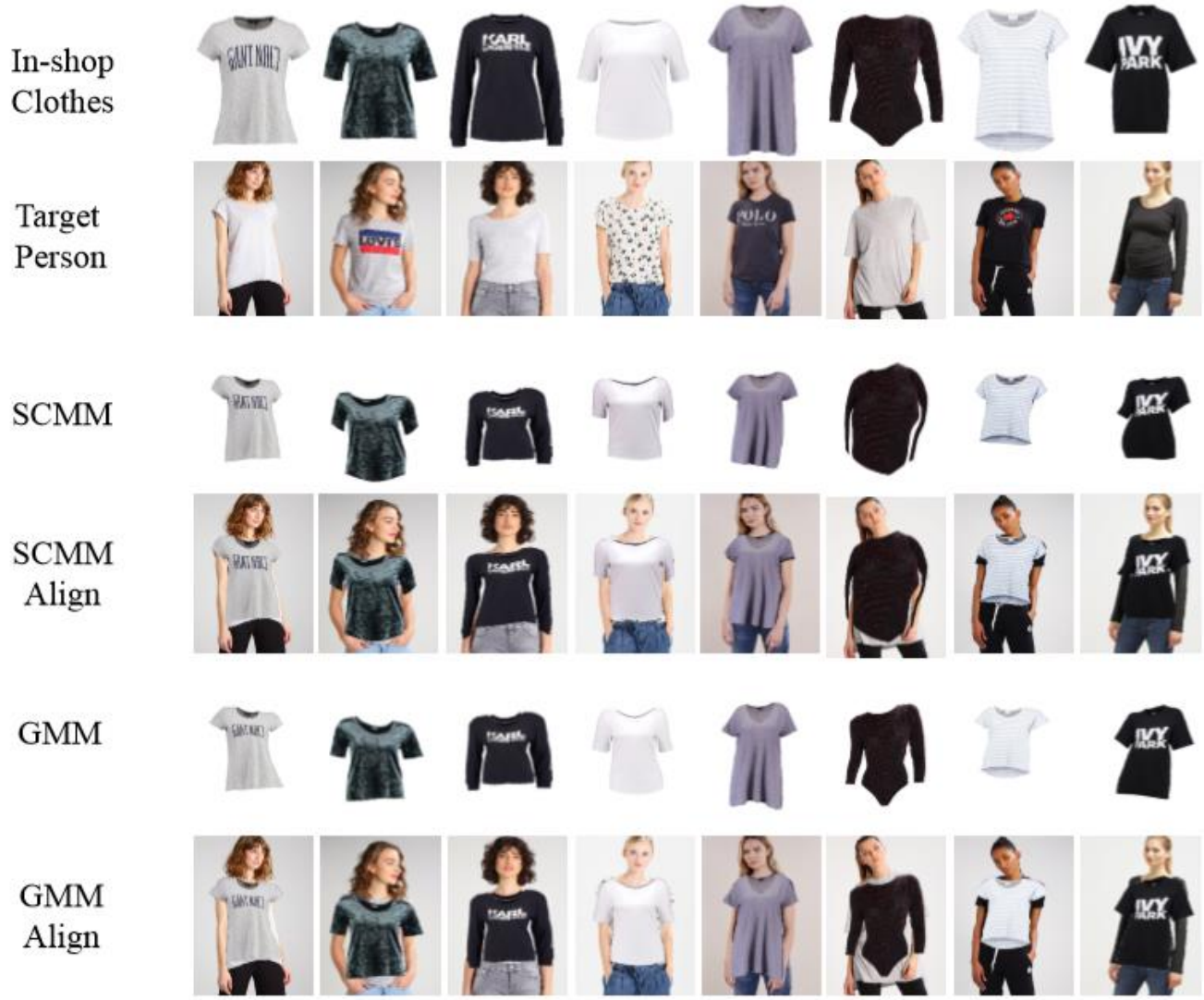
$$I_o = M \odot \hat{c} + (1 - M) \odot I_r$$

$$\mathcal{L}_{\text{VGG}}(I_o, I_t) = \sum_{i=1}^{5} \lambda_i \left\| \phi_i(I_o) - \phi_i(I_t) \right\|_1$$

$$\mathcal{L}_{\text{TOM}} = \lambda_{L1} ||I_o - I_t||_1 + \lambda_{vgg} \mathcal{L}_{\text{VGG}}(\hat{I}, I) + \lambda_{mask} ||1 - M||_1$$

3 Experiments

Comparison of Try-on Results



3 Experiments

Comparison of Try-on Results

3 Experiments

Ablation Study

3 Experiments

Failure cases

Fig. 9. Some failure cases of our CP-VTON.