End-to-End Learning of Geometric Deformations of Feature Maps for Virtual Try-On

Arxiv

2019.10.15

발표자 박성현

1

Introduction

Virtual Try-on

[Virtual Try-on]

1 Introduction

Background - Virtual Try-on Network (VITON)

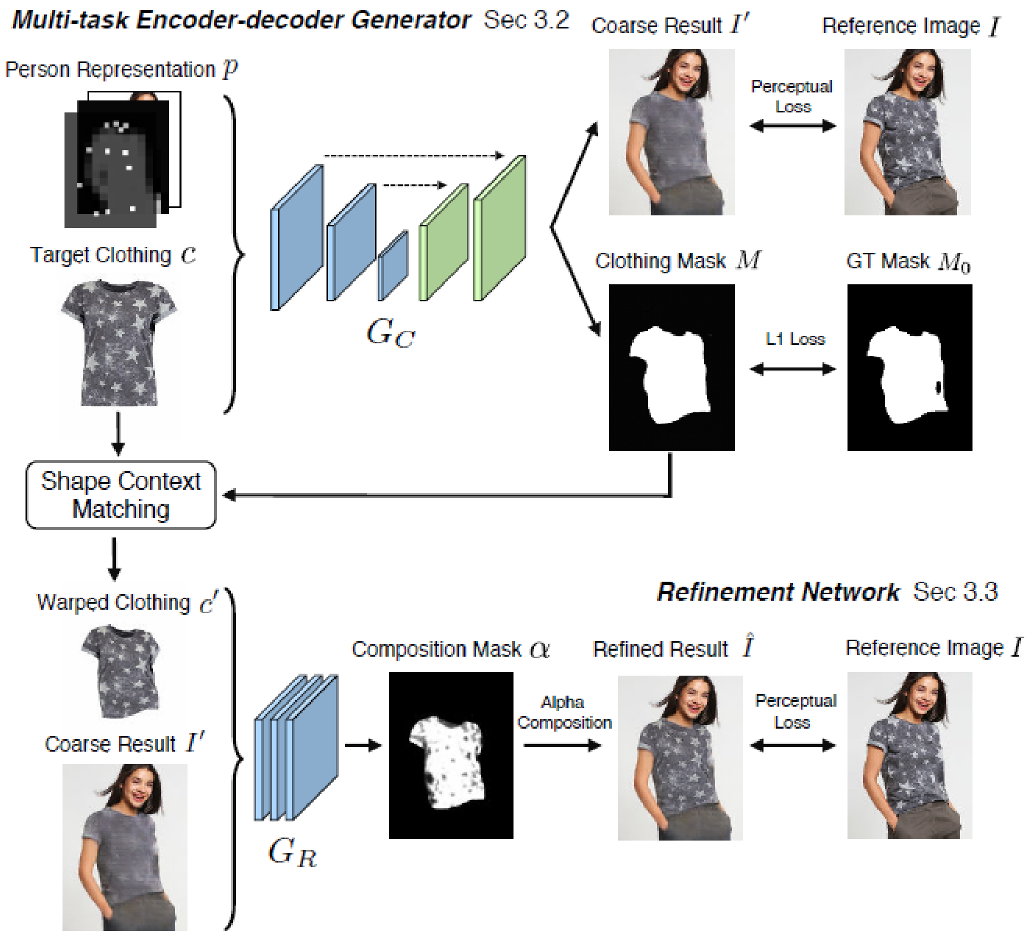


Figure 2: **An overview of VITON.** VITON consists of two stages: (a) an encoder-decoder generator stage (Sec 3.2), and (b) a refinement stage (Sec 3.3).

DAVIAN Data and Visual Analytics Lab

KOREA UNIVERSITY

1 Introduction

Background - CP-VITON

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

1 Introduction

Background - CP-VITON

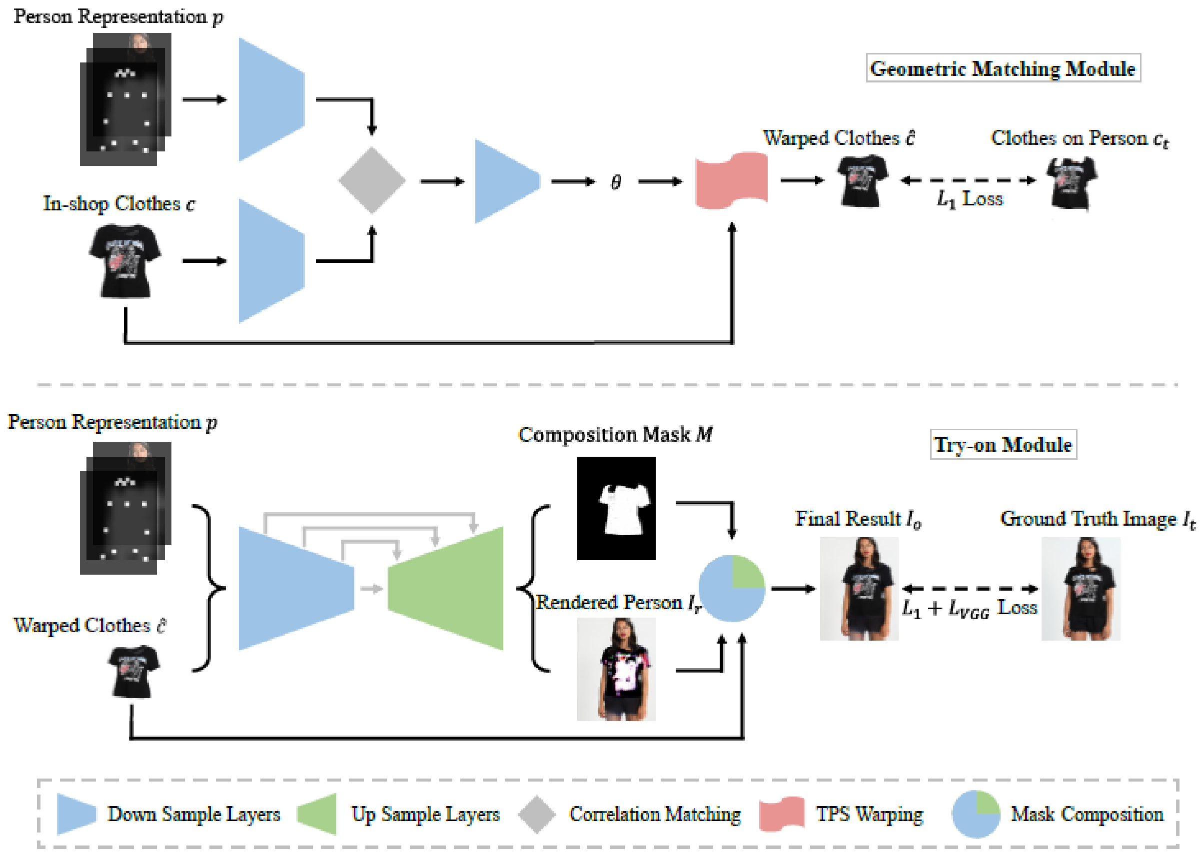
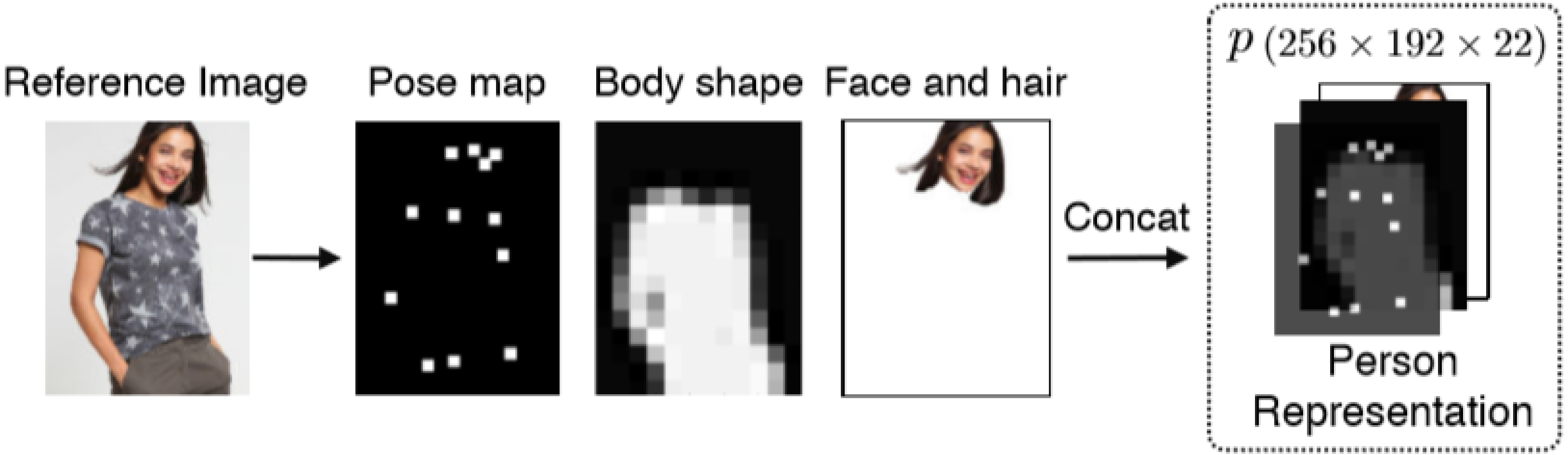


Fig. 2. An overview of our CP-VTON, containing two main modules. (a) Geometric Matching Module: the in-shop clothes c and input image representation p are aligned via a learnable matching module. (b) Try-On Module: it generates a composition mask M and a rendered person I_r. The final results I_o is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

1 Introduction

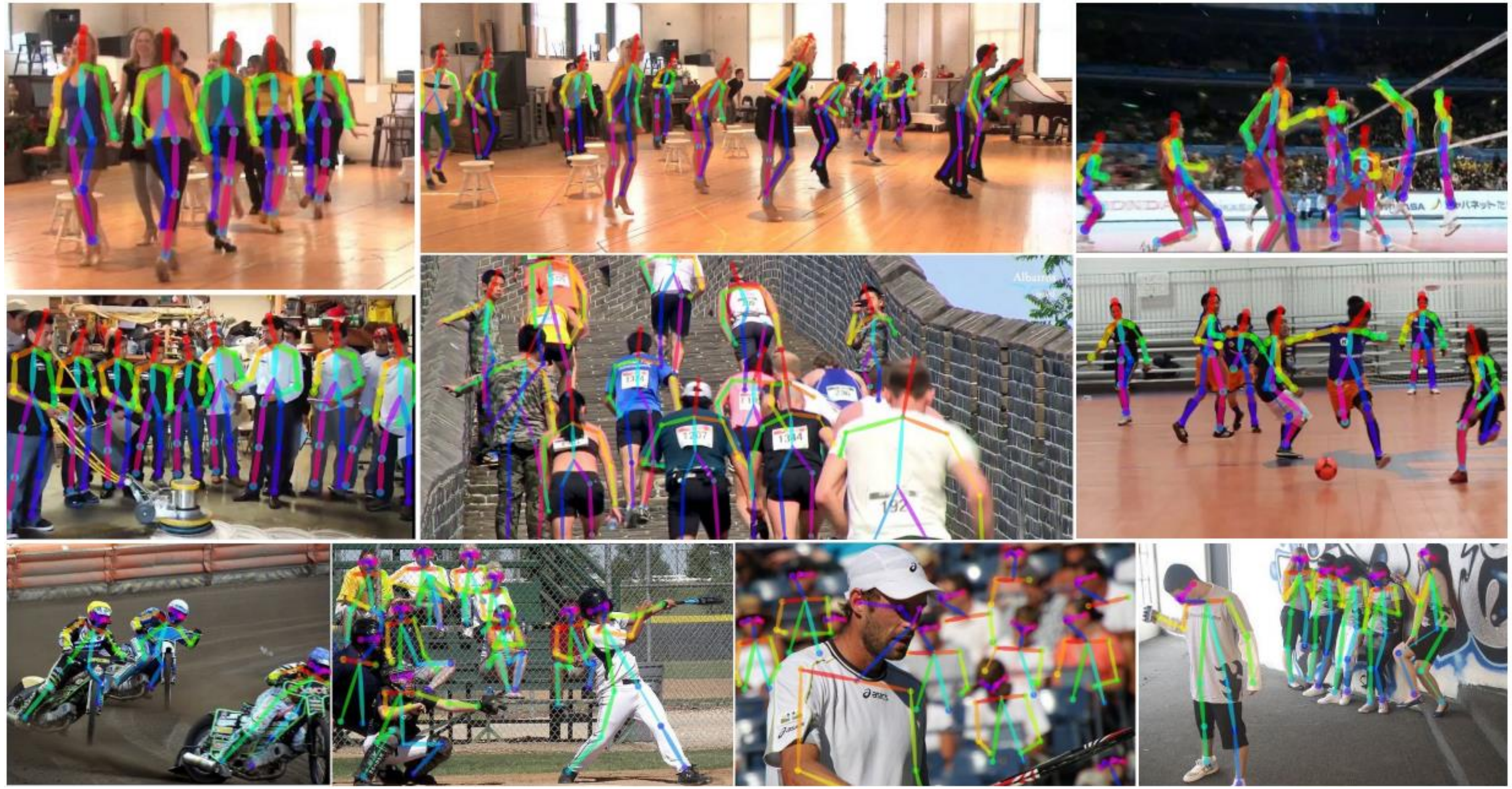
Person Representation

- **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



1 Introduction

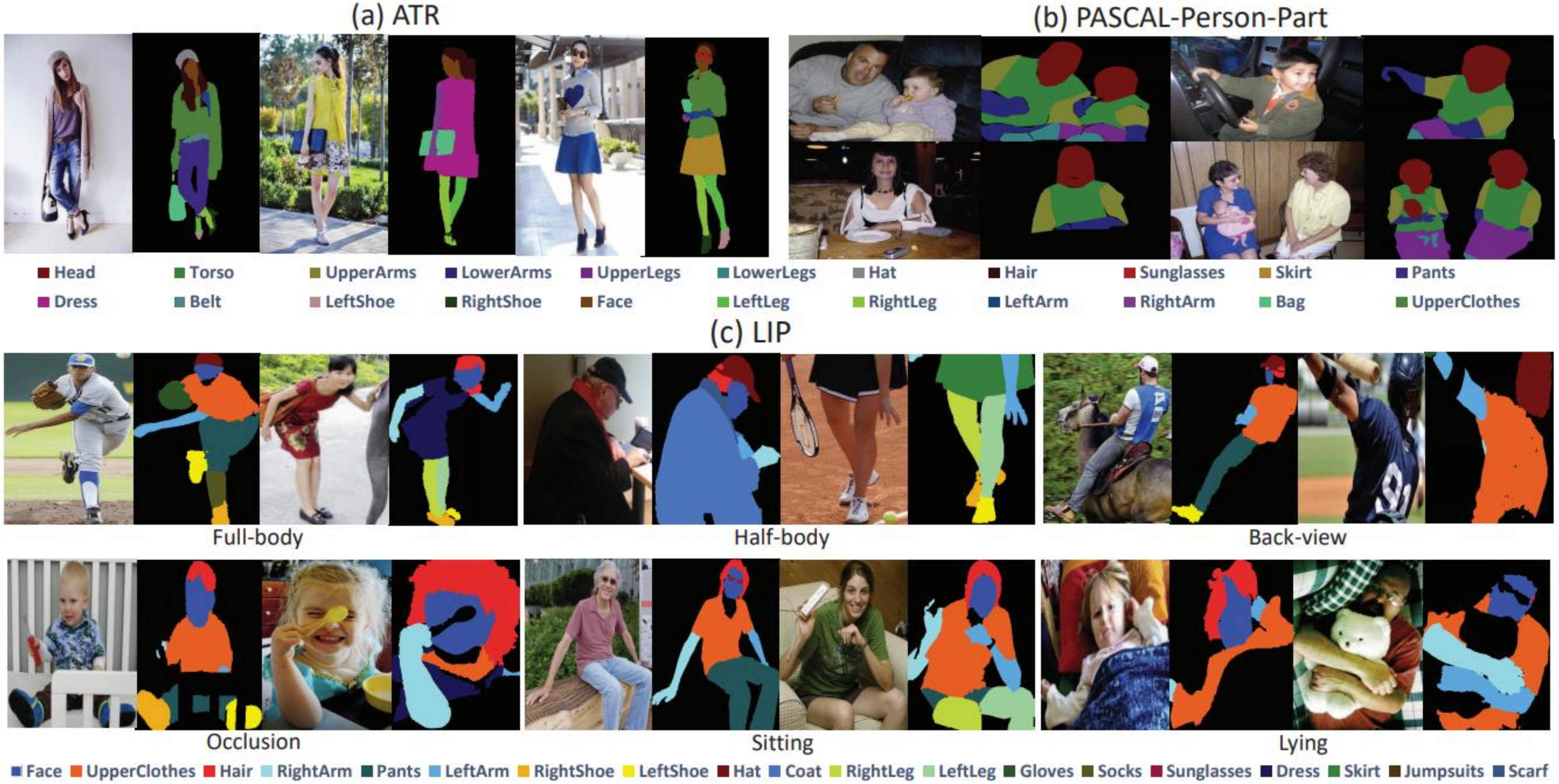
Person Representation



[Real-time Multi-Person Pose Estimation]

1 Introduction

Person Representation



[Self-supervised Structure-sensitive Learning]

2 Model

Overview of the proposed model

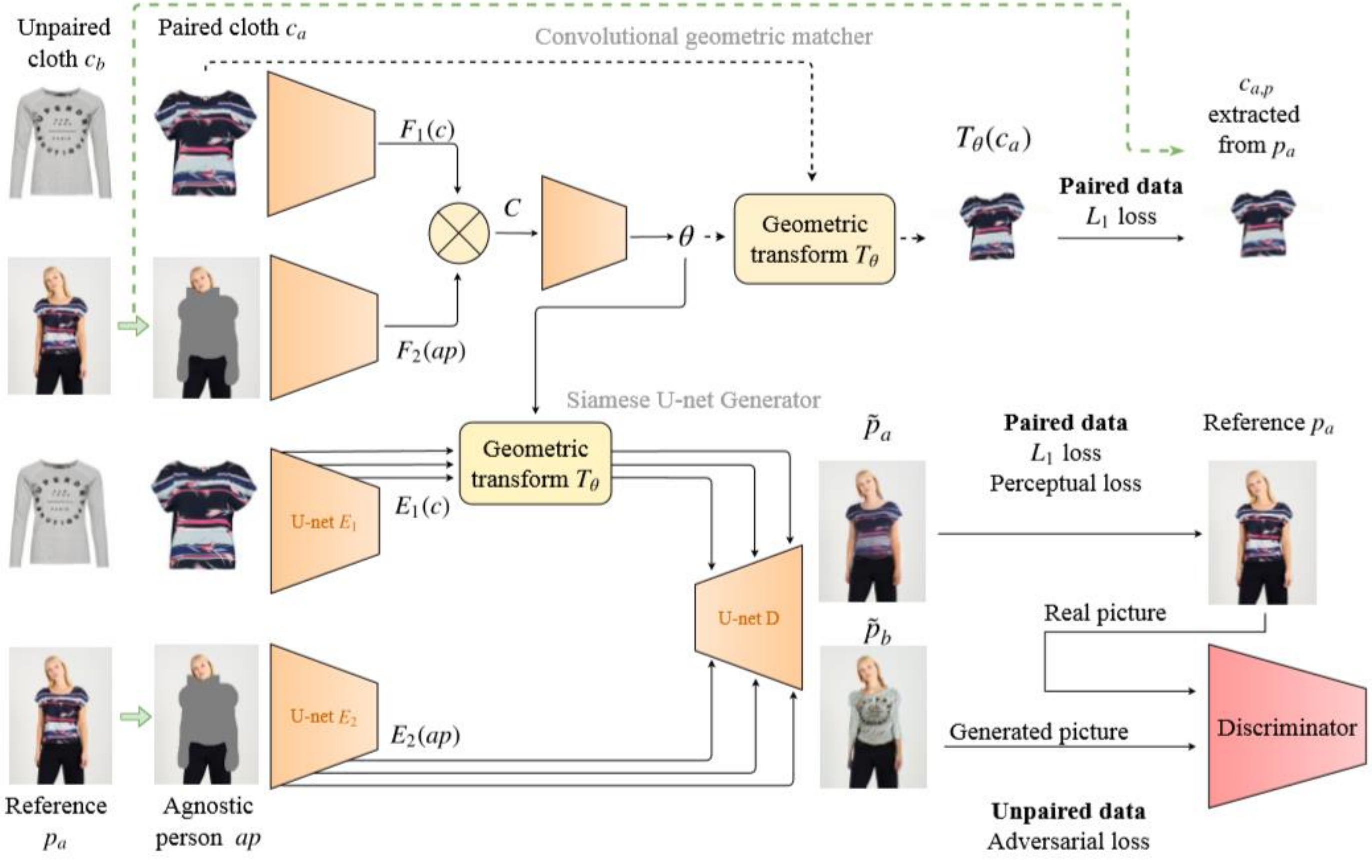


Figure 1: WUTON : our proposed end-to-end warping U-net architecture. Dotted arrows correspond to the forward pass only performed during training. Green arrows are the human parser. The geometric transforms share the same parameters but do not operate on the same spaces. The different training procedure for paired and unpaired pictures is explained in section 3.2.

2 Model

Agnostic Person Representation

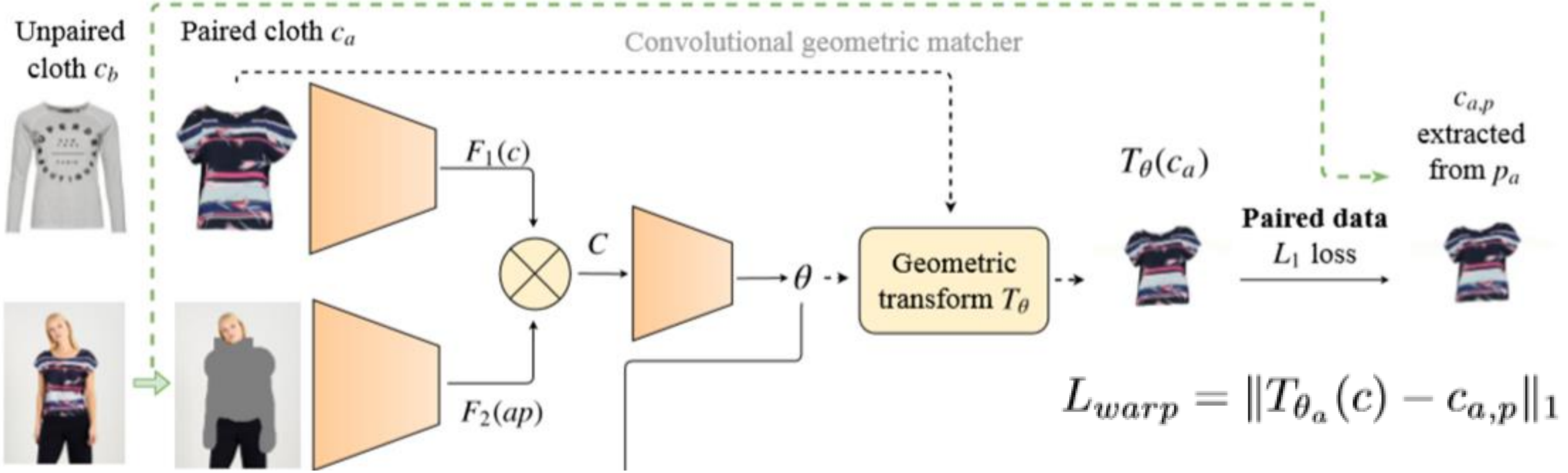
(1) Compute the upper-body mask from pose and body parsing information

(2) Mask the areas corresponding to the arms, the upper-body cloth and a fixed bounding box around the neck keypoint

Reference p_a Agnostic person ap

2 Model

Convolutional Geometric Matcher



(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

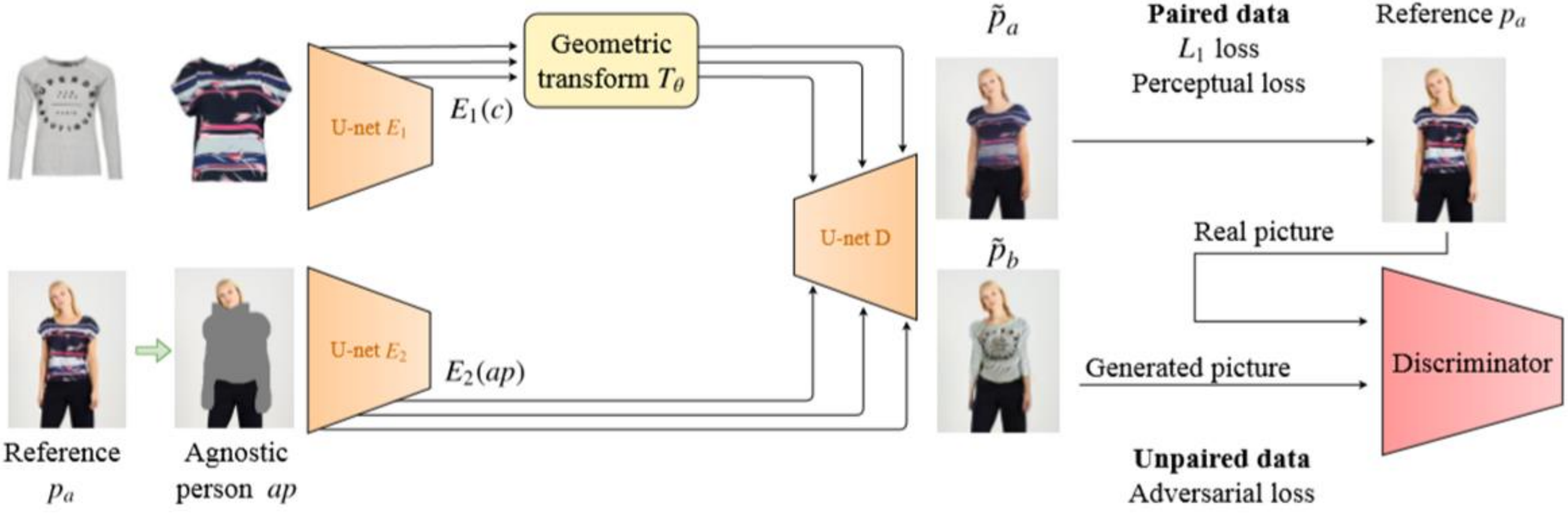
(3) The regression network for predicting the spatial transformation parameters θ

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

2 Model

Warping U-Net

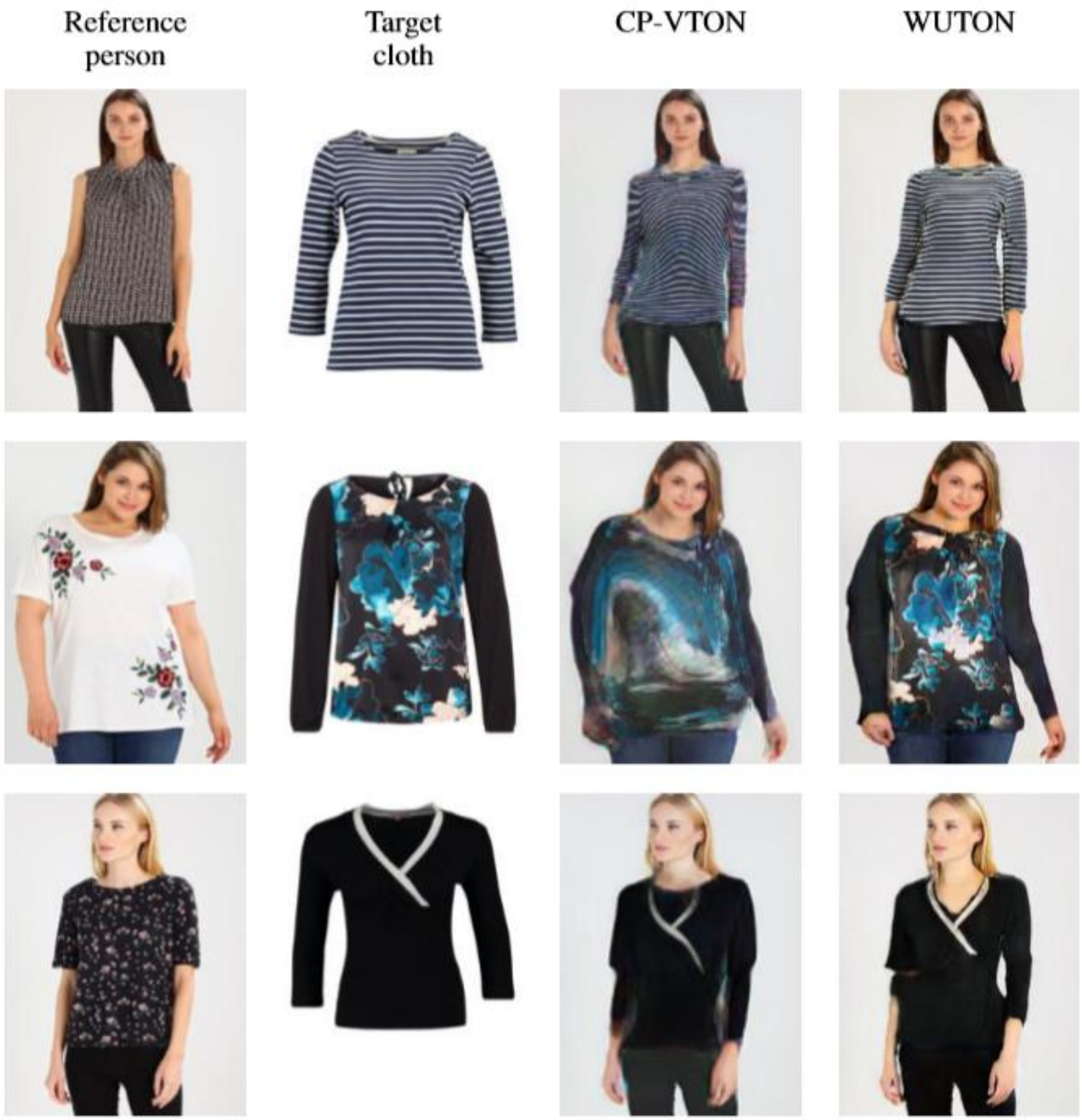
$$L_{perceptual} = \sum_{i=1}^{5} \|\phi_i(\tilde{p}_a) - \phi_i(p_a)\|_1$$



$$L = \lambda_w L_{warp} + \lambda_p L_{perceptual} + \lambda_{L_1} L_1 + \lambda_{adv} L_{adv}$$

3 Experiments

Visual Results



3 Experiments

Ablation Study

Figure 3: Our unpaired adversarial loss function improves the performance of our generator in the case of significant shape changes from the source cloth to the target cloth. Specifically, when going from short sleeves to long sleeves, it tends to gum the shape of the short sleeves. With the paired adversarial loss, we do not observe this phenomenon since the case never happens during training.

3 Experiments

Ablation Study

Figure 4: Left: Our method can handle low-quality masks at cost of generic arm pose. Right: Some common failure cases of our method. Detection of initial cloth can fail beyond the capacity of our U-net generator (first row), and uncommon poses are not properly rendered (second row).

3 Experiments

LPIPS metric

Table 1: LPIPS metric on paired setting. Lower is better, ± reports std. dev.

Method	LPIPS
CP-VTON on ap_{viton}	0.182 ± 0.049
CP-VTON on ap_{wuton}	0.131 ± 0.058
WUTON	0.101 ± 0.047
Impact of loss functions on WUTON:	
W/o adv. loss	0.107 ± 0.049
W. paired adv. loss	0.099 ± 0.046
Not end-to-end	0.112 ± 0.053

Method	LPIPS
Impact of composition on WUTON:	
W. composition	0.105 ± 0.047
Impact of mask quality box masked person:	
CP-VTON	0.185 ± 0.078
WUTON	0.151 ± 0.069