We propose a novel solution for 2D single-pose virtual try-on which uses multiple geometric transformations to generate high-quality and photo-realistic images.

- We model can generate well defined images thanks to a two-stage geometric transformation of the input garment and a generative network.
- We conduct experiments on the VITON dataset [1] and on a collected set of upper-body clothes, and we demonstrate the effectiveness of our solution both in terms of visual similarity with ground-truth images and realism of the generated try-on results.

**Overview**

We generate an output image \( I \) representing the reference person wearing \( c \) by employing a U-Net architecture [3] consisting in two main components.

- **Transformation-Guided Encoder:** We apply the previous learned spatial transformations in the clothes branch, separated from the person branch:
  \[
  T(E(c), \theta_1, \theta_2) = T_{\theta_1}(T_{\theta_2}(E(c), \theta_1), \theta_2)
  \]

- **Try-On Decoder:** The final result \( I \) is guided by a pixel-level \( L_i \), a perceptual loss [2] and an adversarial loss:
  \[
  L_{\text{Try-On}} = p_1 L_{\text{con}} + p_2 L_{\text{perc}} + p_3 L_{\text{adv}}
  \]
  where \( p_1, p_2 \) and \( p_3 \) are weighting coefficients.

**Two-Stage Geometric Transformation Module**

We employ two different geometric transformations, namely affine and thin-plate spline, to warp the in-shop image \( c \) of a particular garment.

- Given an image \( c \) and a pose heatmap \( p_h \), we compute the parameters \( \theta_1 = (A, b) \) for the affine transformation \( T_{\theta_1} \):
  \[
  \begin{bmatrix}
  y_1 \\
  1
  \end{bmatrix} = \begin{bmatrix}
  A & b \\
  0 & 1
  \end{bmatrix}
  \begin{bmatrix}
  x_1 \\
  1
  \end{bmatrix}
  \]

- Given the input \( \xi_1 = T_{\theta_1}(c, \theta_1) \) and a 22-channel structure person representation \( P \), we predict the parameters \( \theta_2 \) to compute the thin-plate spline transformation. We generate the final output \( \xi_2 = T_{\theta_2}(\xi_1, \theta_2) \).
- The loss used to train this module is \( L_{\text{2-stage}} = \lambda_1 L_{\text{con}} + \lambda_2 L_{\text{perc}} \), where \( L_{\text{con}} \) and \( L_{\text{perc}} \) are \( L_1 \) distances between the results of the two learned transformations and the corresponding ground-truths.

**Transformation-Guided Try-On Module**

The affine transformation helps the TPS generating better warped clothes that are closer to the target body pose while reducing artifacts and distortions.

**Try-On Results**

Our VITON-GT better preserves textures and details of the original clothes, thus increasing the realism of generated images.

**References**


