SEOUL NATIONAL UNIVERSITY Google BMVC 2022

Learning to Wear: Details-Preserved Virtual Try-on via Disentangling Clothes and Wearer

Sangho Lee[1], Seoyoung Lee[1], Joonseok Lee[1,2]

[1]Seoul National University [2]Google Research

Introduction

Virtual Try-on

- Synthesizing a realistic image of a person wearing the given clothing

- Need to disentangle human and clothes & different types of clothes data
- Should be generalizable to various human poses and body shapes

Motivation

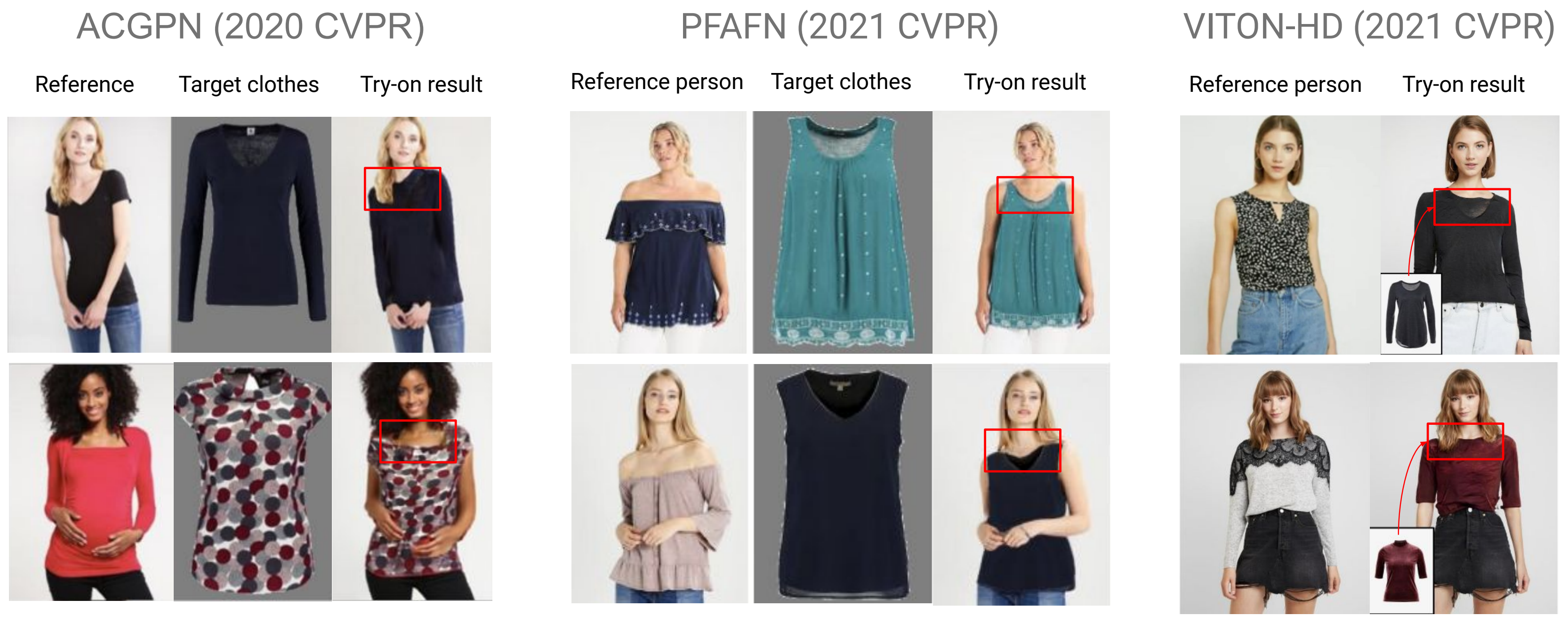
Inability to reflect details of the target clothing

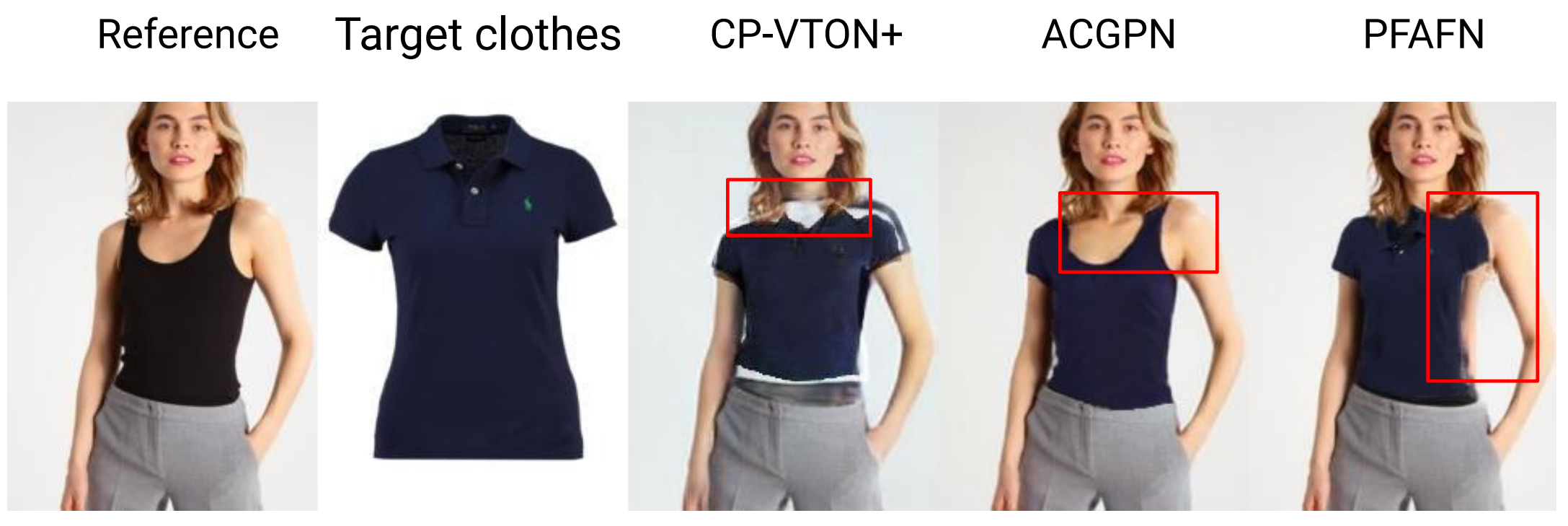
- Detailed characteristics of of the target clothing (*e.g.*, shape of neckline and sleeves) are not retained.
- Output often reveals characteristics of clothes in the reference image.

Limited understanding in 3D semantics of wearing clothes

- Parts that should not be seen when worn (*e.g.*, inner side of shirt neckline) are still visible when clothing is worn.
- Overall struggle in synthesizing well-fitted images indicate a weak generalizability of the models.

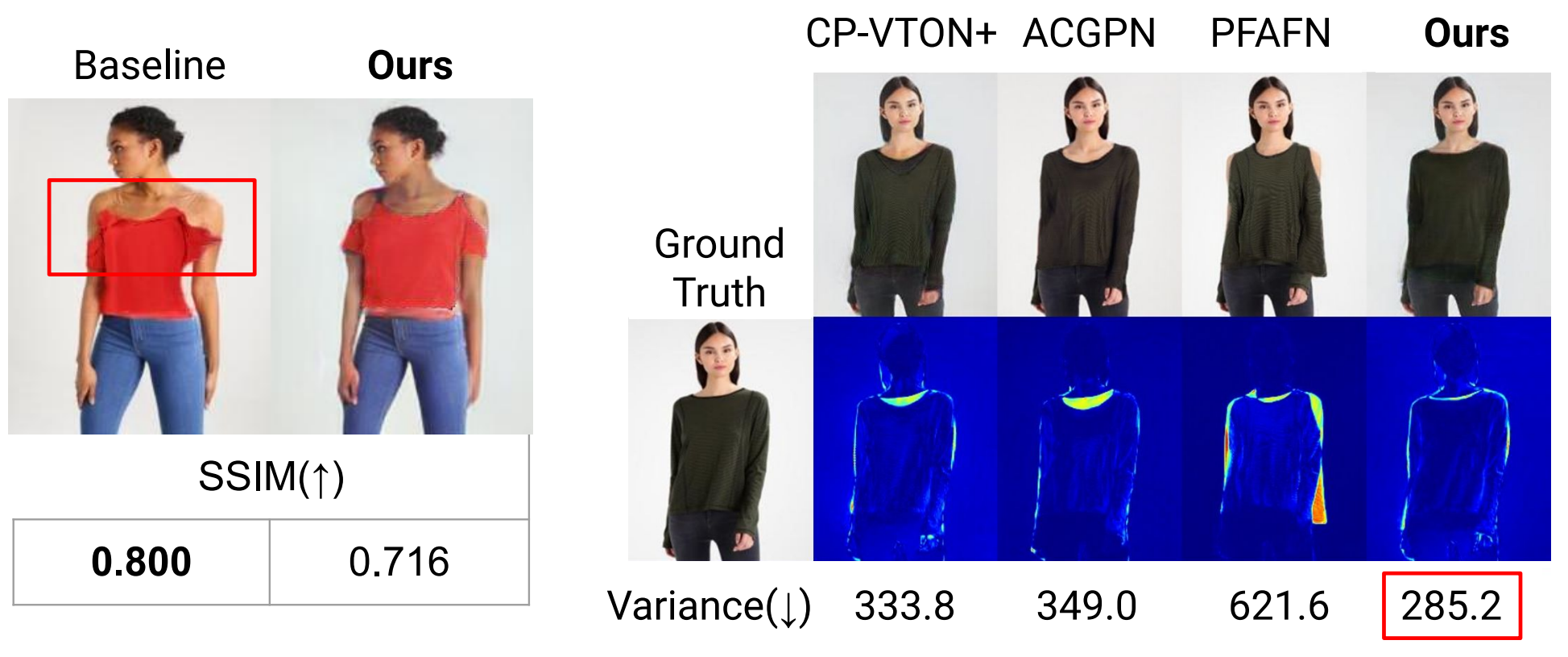
→ ***Disentangle <u>clothes & wearer</u> and <u>target & source clothes</u>***





Evaluation

Flaw of existing metrics based on entire images

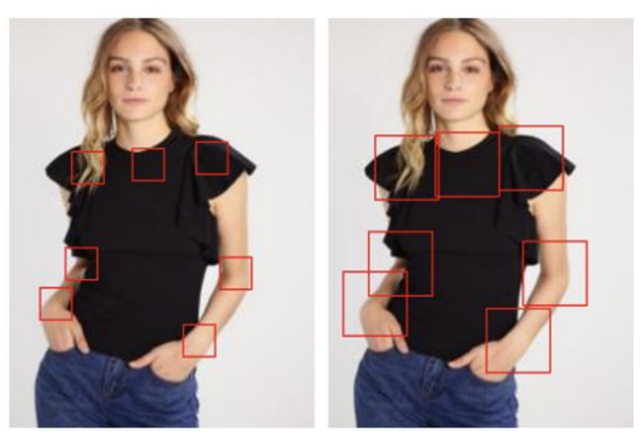


- Misalignment between qualitative and quantitative evaluation
- Considers even parts irrelevant for virtual try-on (*e.g.*, background).

Novel approach to evaluation metrics

- Focus on areas relevant to the task (*i.e.*, major body keypoints) for better alignment with human perception.

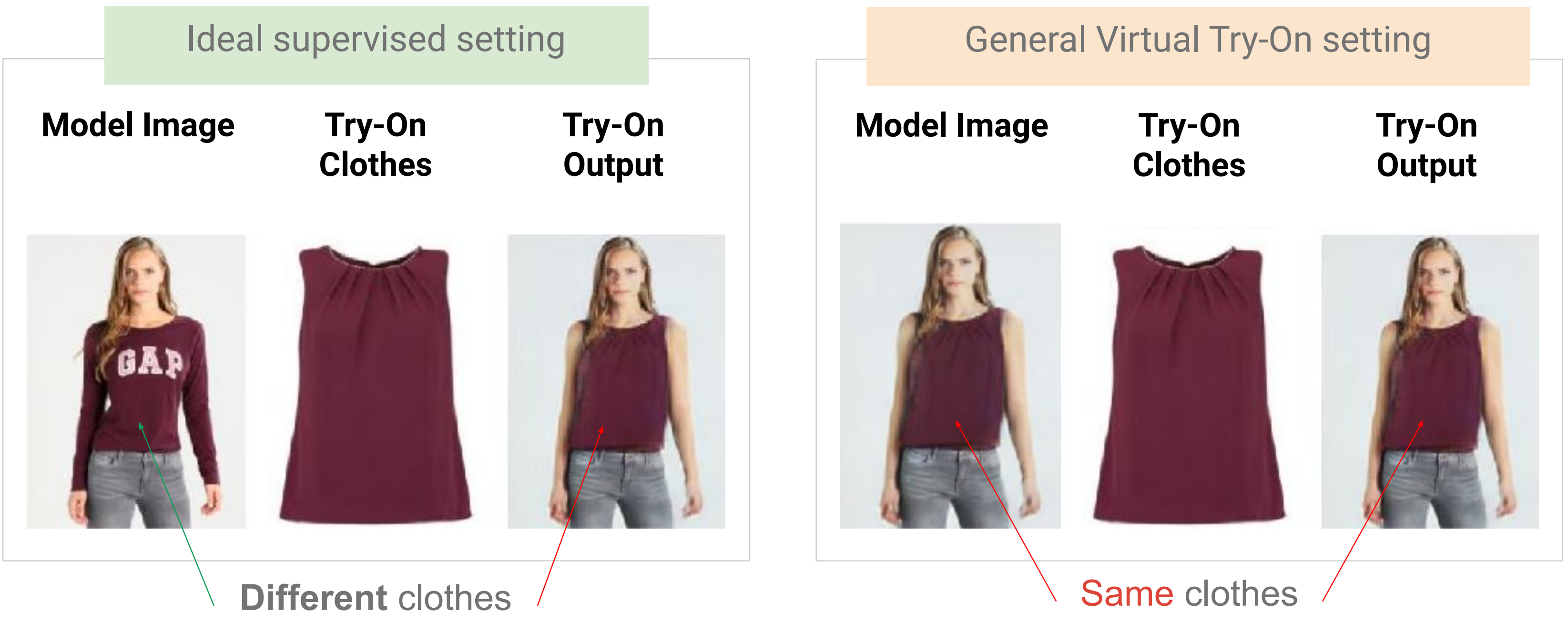
$$\texttt{Metric}^p_\epsilon(I) = \frac{1}{k}\sum_{i=1}^{k}\texttt{Metric}^{\text{all}}\left(I\left[x_i-\frac{\epsilon}{2}:x_i+\frac{\epsilon}{2}, y_i-\frac{\epsilon}{2}:y_i+\frac{\epsilon}{2}\right]\right)$$



Core Issue

Lack of paired datasets

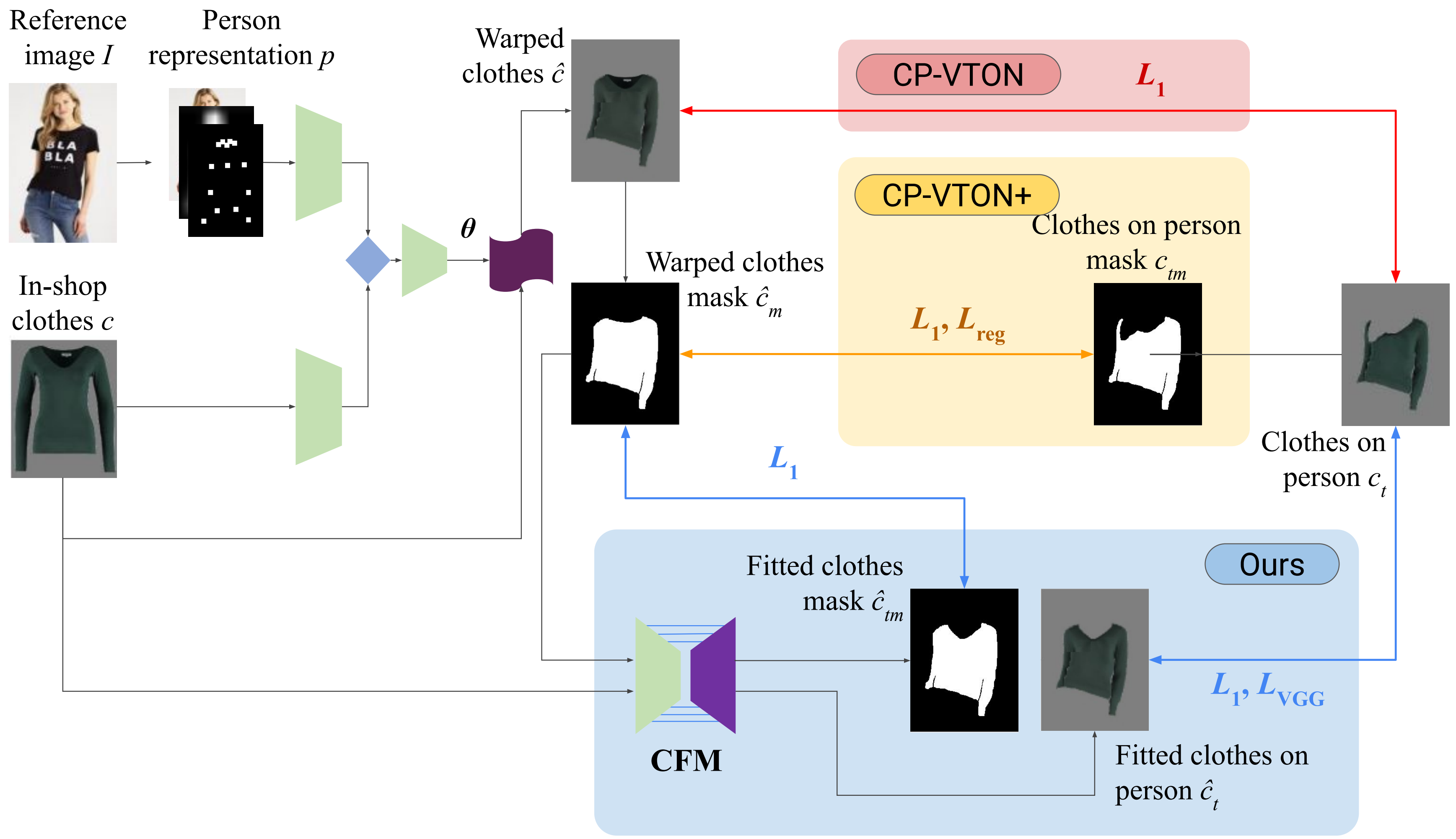
- Impossible to have two photos of a model with just different clothes on.
- Usually, models are trained to wear the same clothes that is already worn.



Method

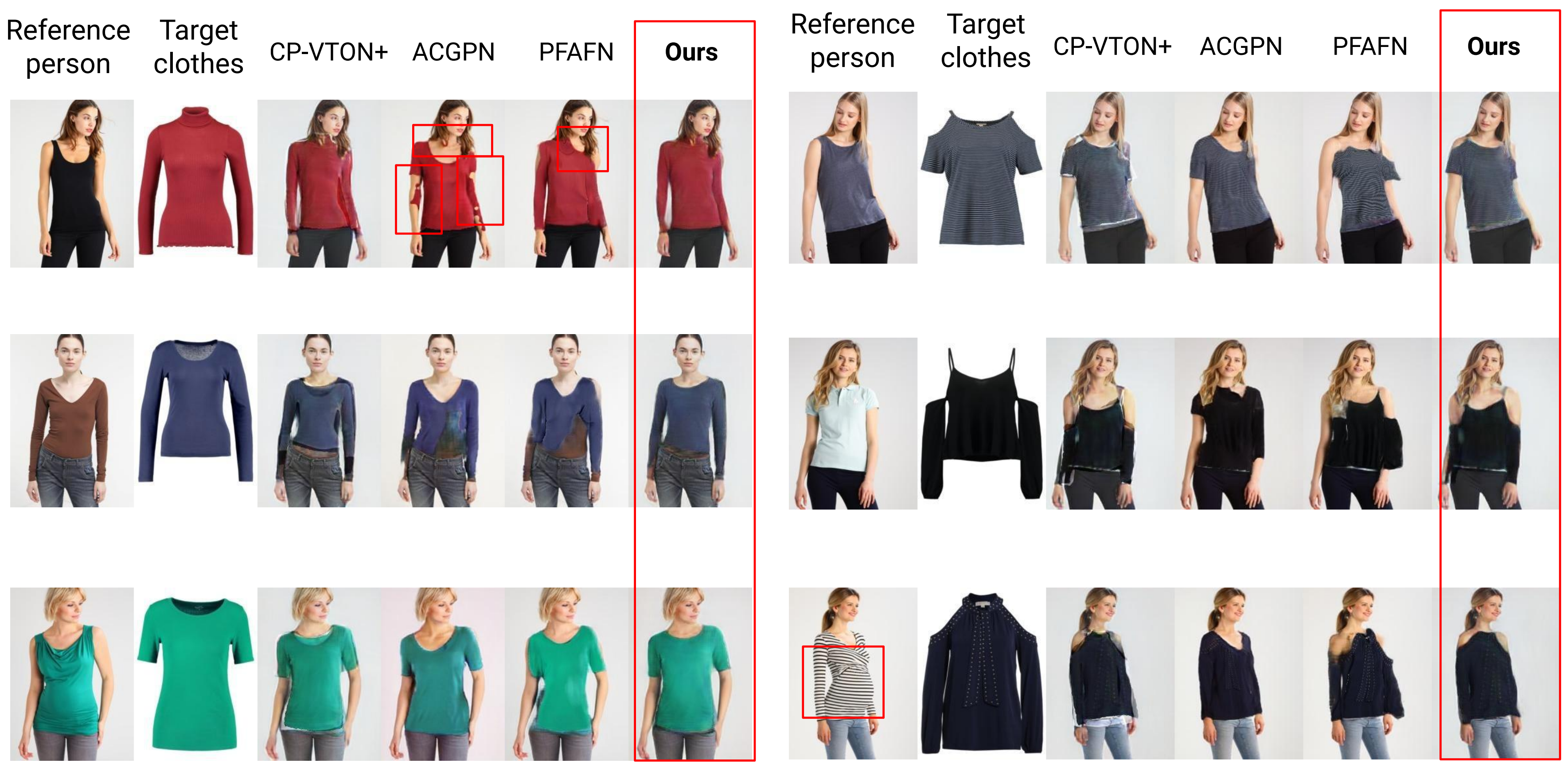
Clothes Fitting Module (CFM): Learning to Wear

- Inserted between Geometric Warping Module (GWM) and Try-on Module.
- Allows GWM to perceive clothes as the source clothes in the reference.
- Allows CFM to perceive clothes as the target ground truth.
- ***Successfully disentangles <u>source & target clothes</u> and <u>clothes & wearer</u>***


Reference image I
Person representation p
In-shop clothes c
Warped clothes $\hat{c}$
θ
Warped clothes mask $\hat{c}_m$
CP-VTON
L_1
CP-VTON+
Clothes on person mask c_{tm}
L_1, L_{reg}
Clothes on person c_t
L_1
Ours
Fitted clothes mask $\hat{c}_{tm}$
L_1, L_{VGG}
Fitted clothes on person $\hat{c}_t$
CFM

Results

- Retain properties of target clothes, disentangled from the reference image
- Generalizable to various designs, as well as body shapes and poses.



Take-home Messages

- Previous virtual try-on models learned entangled representations that lack generalizability due to the lack of paired datasets.
- With CFM, we disentangle important factors of virtual try-on and detour the inherent limitation in data.
- Our patch-based evaluation metrics better correspond to qualitative results.