

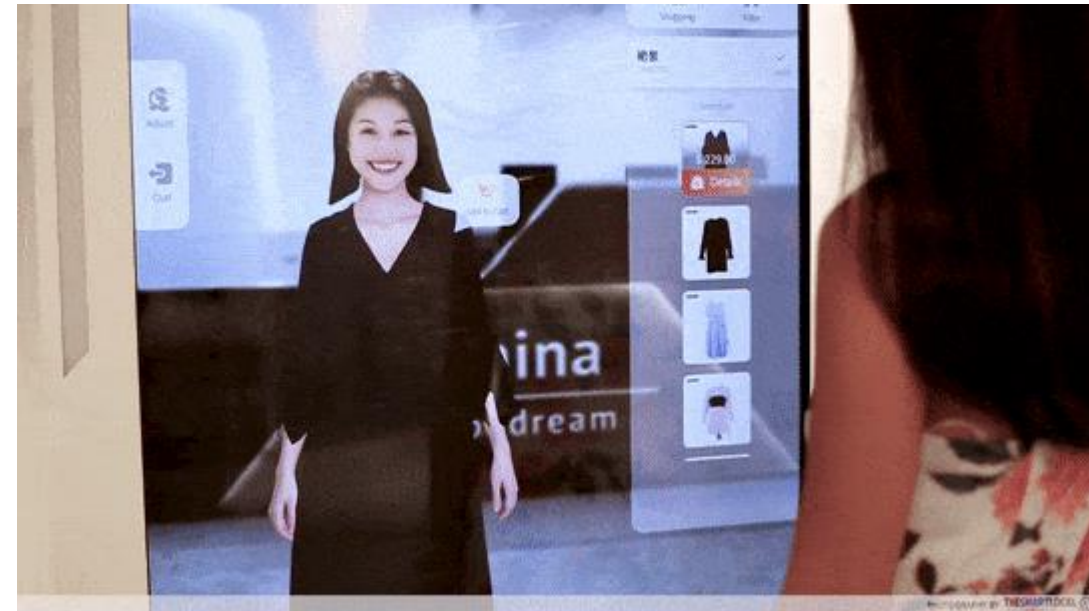
CP-VTON+: Clothing Shape and Texture Preserving Image-Based Virtual Try-On

Matiur Rahman Minar¹, Thai Thanh Tuan¹, Heejune Ahn¹, Paul Rosin², Yu-Kun Lai²

¹SeoulTech, ROK, ²Cardiff University, UK

Virtual try on

- Retain body shape and pose
- Reserve characteristics of target clothes
- Eliminate old clothes and replace with target clothes
- Retain non-relevant clothes



VITON

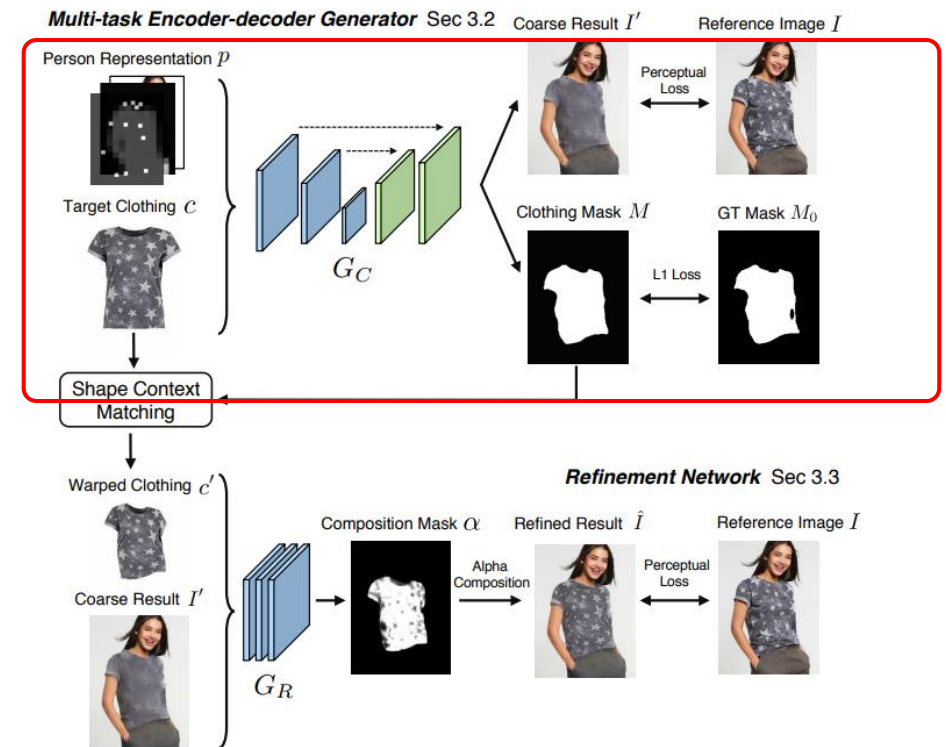
- Coarse-to-fine approach, using two-stage network

- Generator Stage

- encoder-decoder generator
- coarse synthesized image result I'

- Refinement Stage

- generate warped image c' using TPS
- refine using c' and I'



VITON: An Image-based Virtual Try-on Network, Xintong Han et al. CVPR 2018

VITON

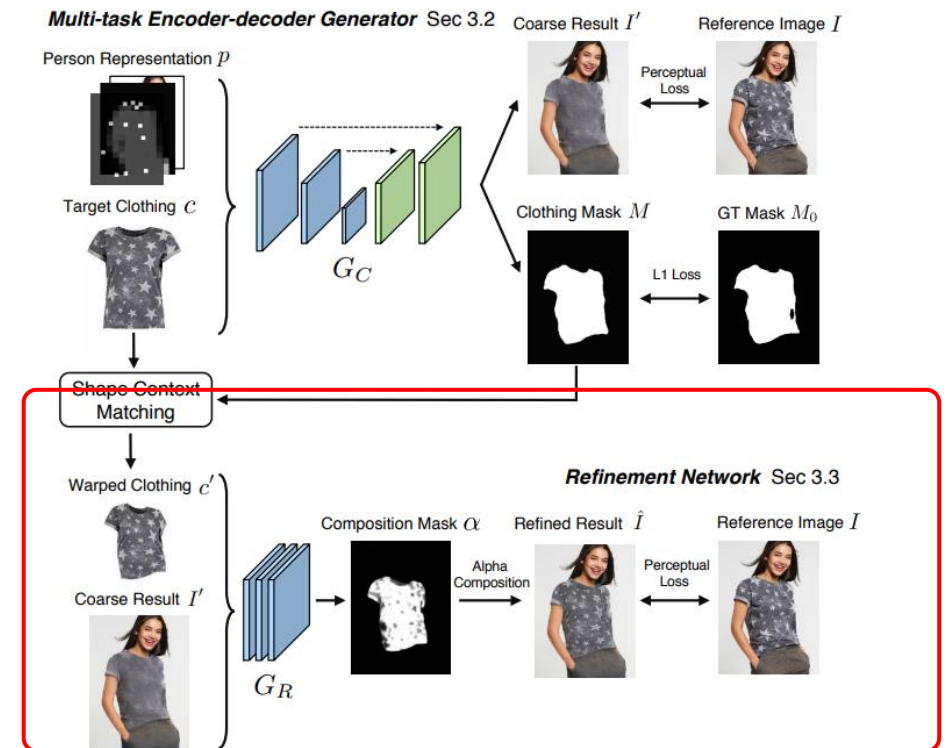
- Coarse-to-fine approach, using two-stage network

- Generator Stage

- encoder-decoder generator
- coarse synthesized image result I'

- Refinement Stage

- generate warped image c' using TPS
- refine using c' and I'



VITON: An Image-based Virtual Try-on Network, Xintong Han et al. CVPR 2018

VITON

- Coarse-to-fine approach, using two-stage network
- Problem
 - Warping is vulnerable to mask, blurry in rich details

Details on
body



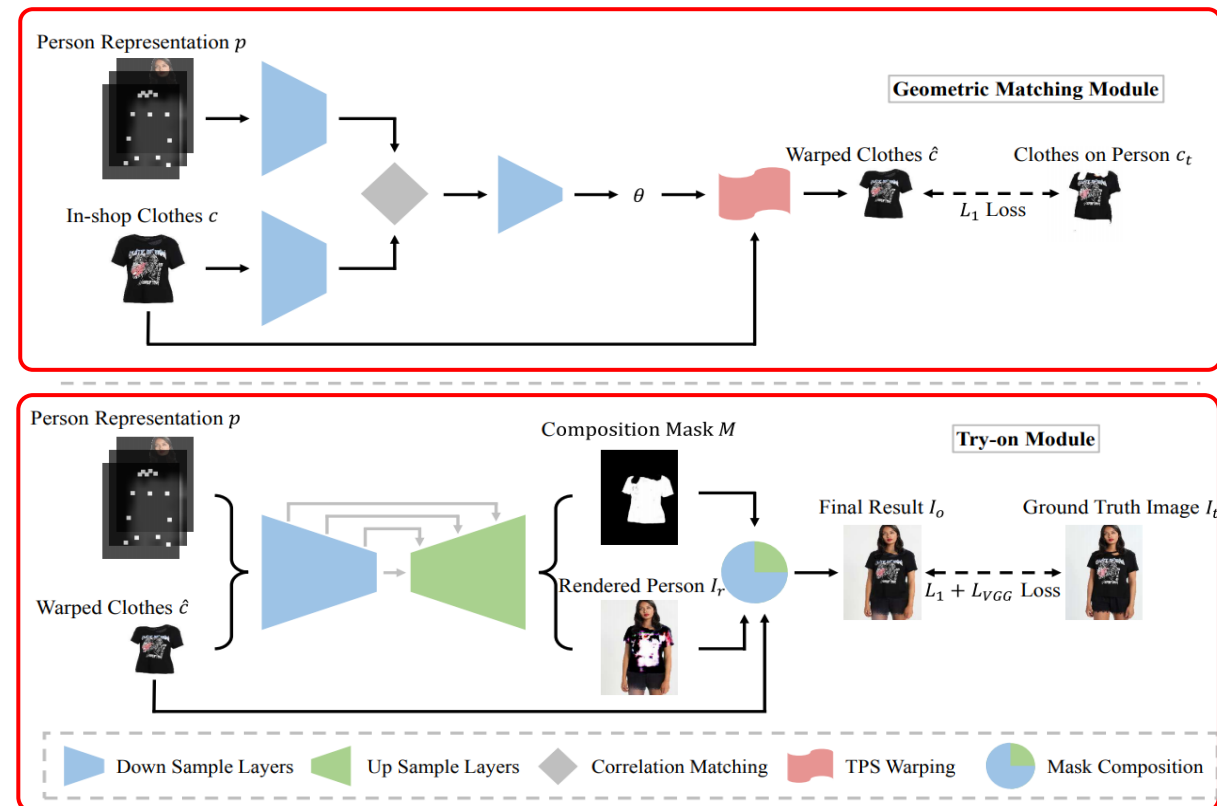
rich textures



blurred details

CP-VTON

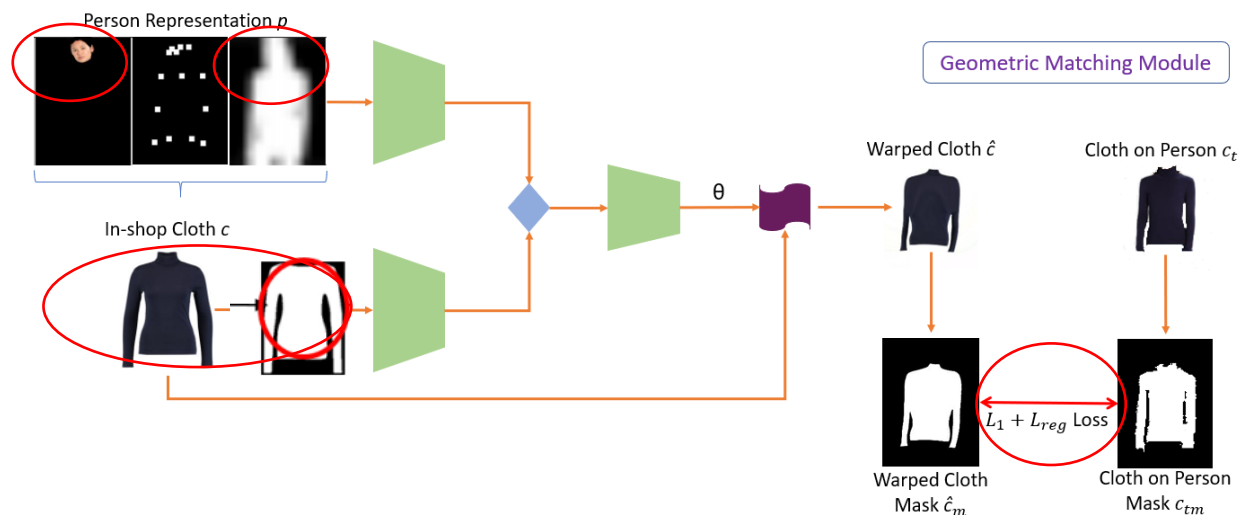
- Preserving the characteristics of clothes
- Geometric Matching Module (GMM)
 - estimating transformation parameters (TPS)
 - generate warped image \hat{c}
- Try-On Module (TOM)
 - A network to estimate M and coarse person image
 - generate final try on image I_r
 - fuse M , I_r and \hat{c}



CPVTON+

Add skin label to
VITON dataset

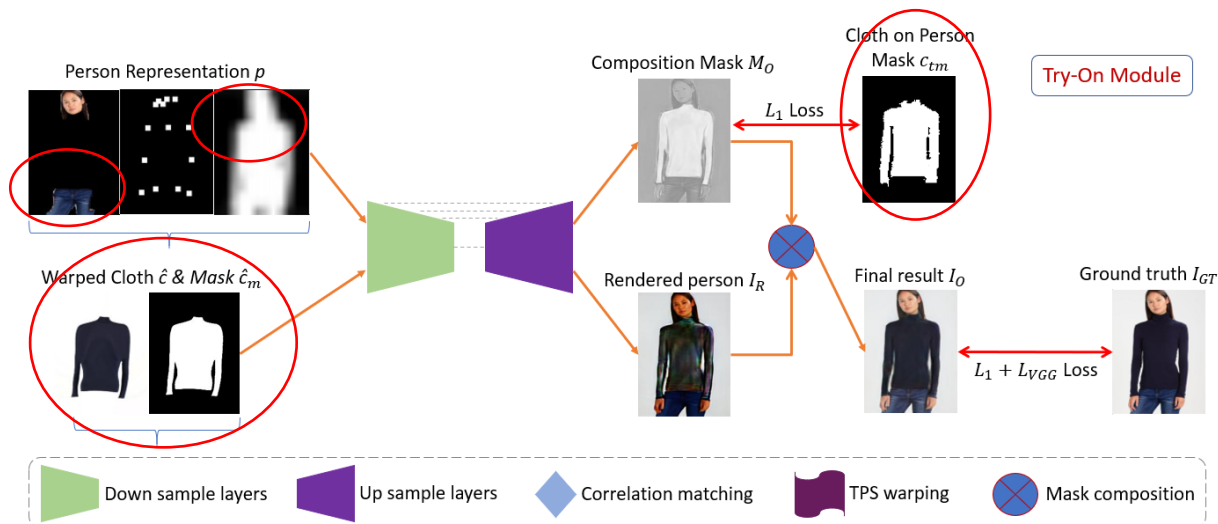
Using target
cloth mask



Add regularization
loss

Add un-upper
cloth

Add warped cloth
mask



Supervised ground
truth



Clothing Warping Stage: Adding skin label

- Neck and bare chest area → wrongly labeled as background
- Improvement:
 - Add new label 'skin'



Reference image

Human parsing
From VITON dataset

Body shape
In CPVTON

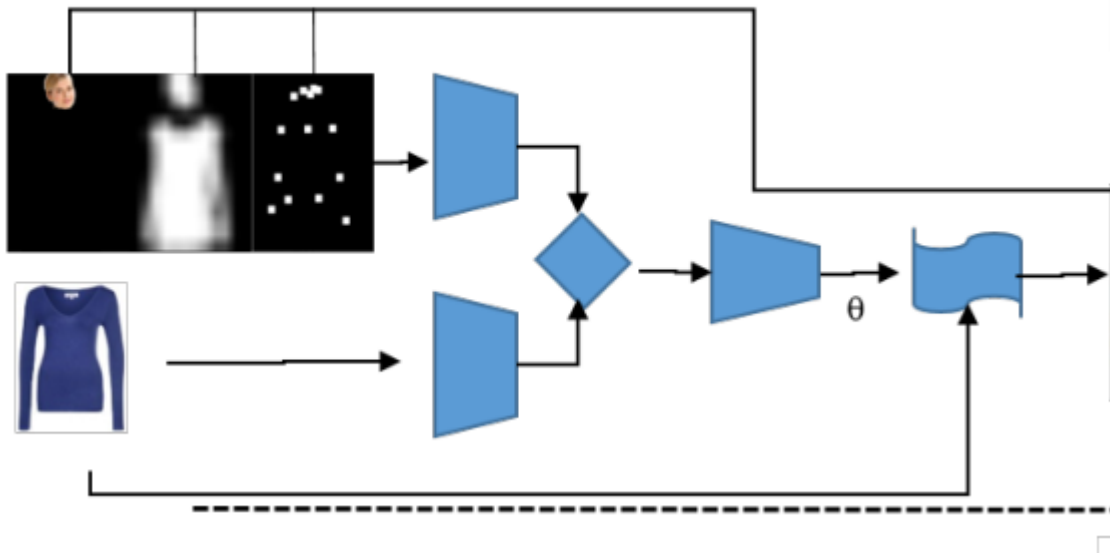
Update human
parsing
With skin label

Body shape
In CPVTON+

Clothing Warping Stage: using of cloth mask

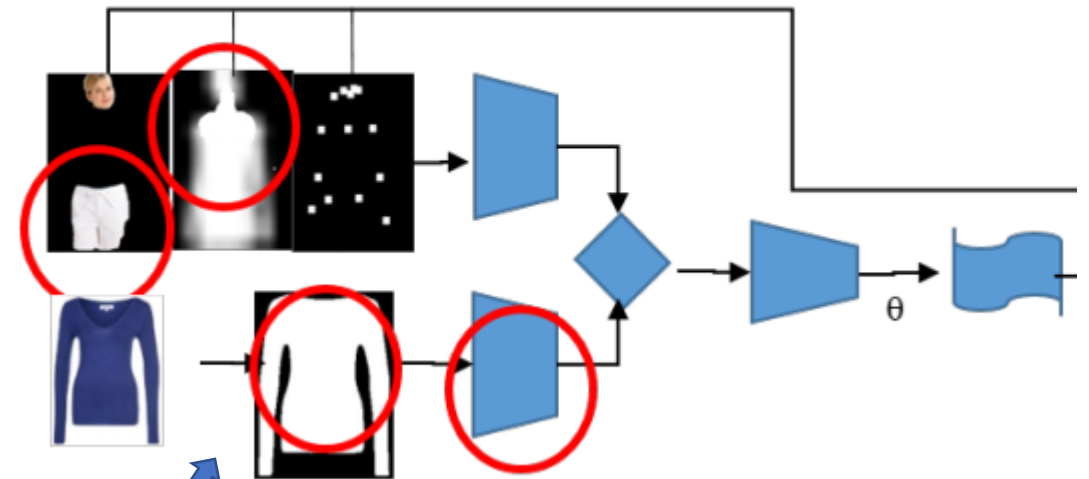
- Colored cloth → Cloth mask

CP-VTON



$$\theta = f_{\theta}(f_H(H_t), f_C(C_i))$$

CP-VTON+



$$\theta = f_{\theta}(f_H(H_t), f_C(M_{Ci}))$$

Clothing Warping Stage: TPS parameters regularization

- Reveal that warped clothing is often severely distorted. → Add regularization on the TPS parameters.

$$L_{GMM}^{CP_VTON+} = \lambda_1 \cdot L1(C_{warped}, I_{Ct}) + \lambda_{reg} \cdot L_{reg}$$

$$L_{reg}(G_x, G_y) = \sum_{i=-1,1} \sum_x \sum_y |G_x(x+i, y) - G_x(x, y)| + \sum_{j=-1,1} \sum_x \sum_y |G_x(x, y+j) - G_x(x, y)|$$

Blending Stage: Retain un-upper clothes area



Reference image

CPVTON
head only input

CPVTON result

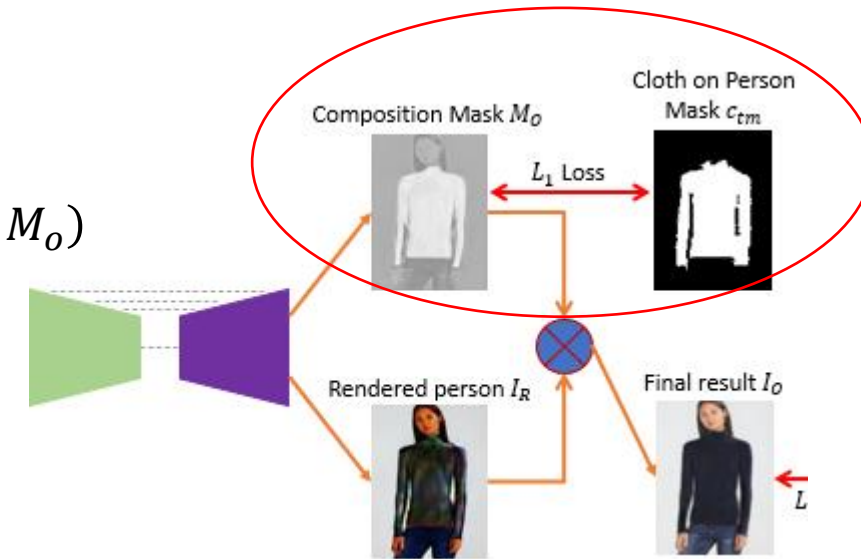
CPVTON+ up-upper
clothes area input

CPVTON+

Blending Stage: Supervised ground truth mask

$$L_{TOM}^{CP_VTON} = \lambda_1 \cdot L1(I_0 - I_{GT}) + \lambda_{VGG} \cdot LVGG(I_0, I_{GT}) + \lambda_{mask} \cdot L1(\mathbf{1}, M_o)$$

$$L_{TOM}^{CP_VTON+} = \lambda_1 \cdot L1(I_0 - I_{GT}) + \lambda_{VGG} \cdot LVGG(I_0, I_{GT}) + \lambda_{mask} \cdot L1(\mathbf{M}_{GT}, M_o)$$



Inshop cloth



CP-VTON warped cloth
and
composition mask



CP-VTON+ warped cloth
and
composition mask



Blending Stage: improve background color inshop clothes

- TOM could not recognize the white cloth area.



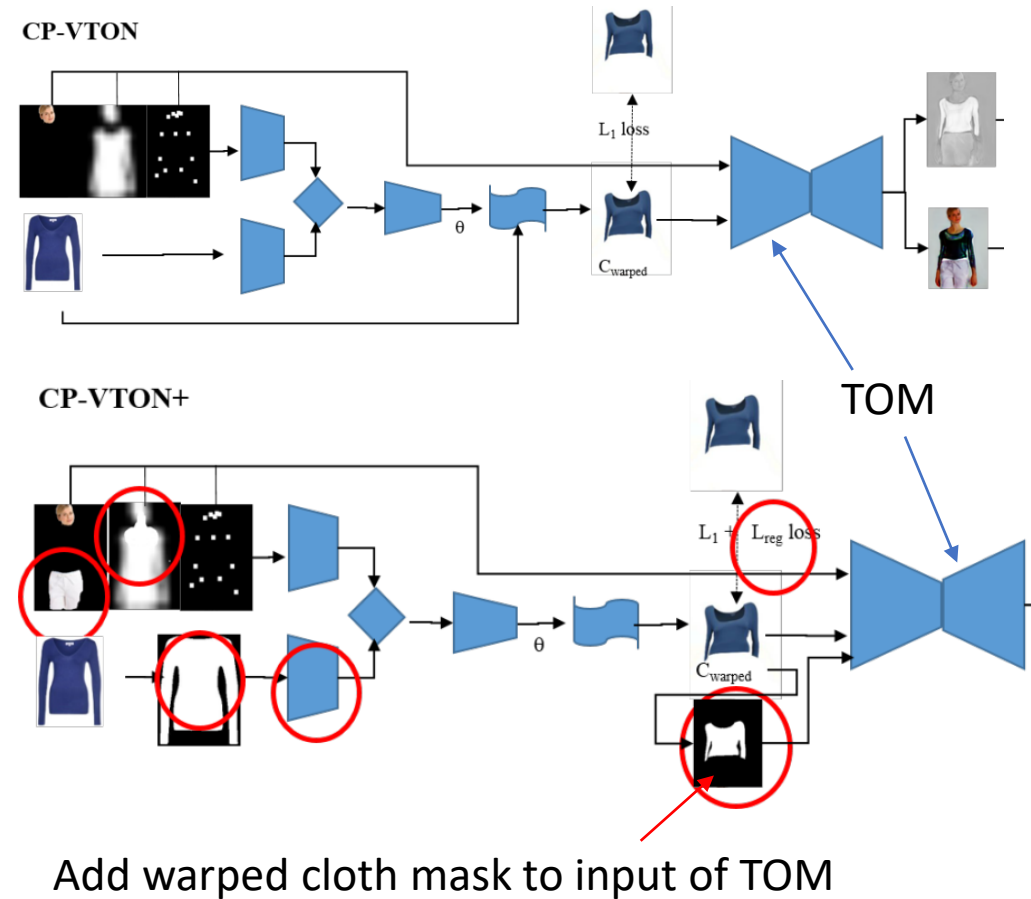
Inshop cloth



CP-VTON warped cloth and composition mask



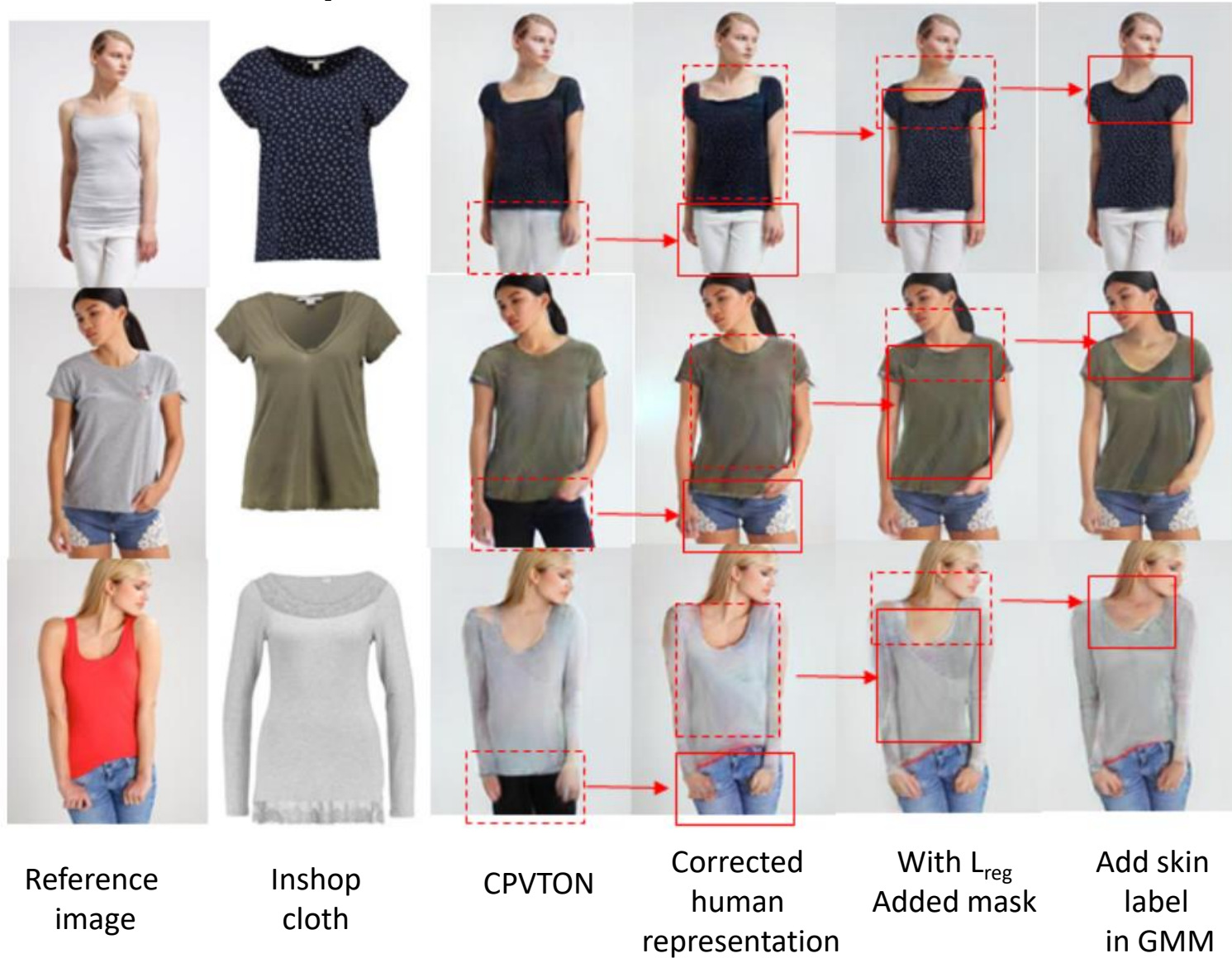
CP-VTON+ warped cloth and composition mask



Experiments and Results

Method	Warped (IoU)	Blended		
		SSIM	LPIPS	IS (mean \pm std.)
CP-VTON[4]	0.7898	0.7798	0.1397	2.7809 \pm 0.0594
CP-VTON+ (w/o GMM regularization & mask loss)	0.7602	0.8076	0.1263	3.0735 \pm 0.0531
CP-VTON+ (w/o GMM mask loss)	0.7920	0.8077	0.1231	3.1312 \pm 0.0837
CP-VTON+ (Ours)	0.8425	0.8163	0.1144	3.1048 \pm 0.1068

Ablation Study



Discussions

- 2D transformation can not handle strong 3D deformations.
- Better human parsing is crucial for better try on results



Figure 4. Failures of our CP-VTON+

Conclusion

- Proposed a refined image based VTON system, CPVTON+
- Solving issues in previous approaches:
 - Errors in human representation and dataset
 - Network design
 - Loose cost function
- Future work:
 - 3D reconstruction would be use for handle strongly clothing deformations

Project site

- <https://minar09.github.io/cpvtonplus/>
- <https://github.com/minar09/cp-vton-plus>

References

- [1] Xintong Han, Zuxuan Wu, Zhe Wu, Ruichi Yu, and Larry S. Davis. Viton: An image-based virtual try-on network. *CVPR*, pages 7543–7552, 2018. 1, 2, 3, 4
- [2] Ignacio Rocco, Relja Arandjelovic, and Josef Sivic. Convolutional neural network architecture for geometric matching. In *CVPR*, pages 6148–6157, 2017. 2
- [3] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. In *NeurIPS*, pages 2234–2242, 2016. 3, 4
- [4] Bochao Wang, Hongwei Zhang, Xiaodan Liang, Yimin Chen, Liang Lin, and Meng Yang. Toward characteristic-preserving image-based virtual try-on network. In *ECCV*, 2018. 1, 2, 3, 4
- [5] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 13(4):600–612, 2004. 3, 4
- [6] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, pages 586–595, 2018. 3, 4

