Detail-Preserving Image-based Virtual Try-On Network with Patch GAN

Chia-Ming Wang[1], Shangquan Sun[2], Rongcong Xu[1], Hanwen Miao[1]

[1]EECS, University of Michigan, [2]LSA, University of Michigan

[1]{jaywang, rongcong, hwmiao}@umich.edu, [2]sunsean@umich.edu

MICHIGAN ENGINEERING
UNIVERSITY OF MICHIGAN

Introduction

With the rapid growth of online shopping, image-based virtual try-on systems have attracted many research interests.Many virtual try-on systems have been proposed.Characteristic-preserving Image-based Virtual Try-on Network (CP-VTON) is one of state-of-art systems. Based on CP-VTON, we propose a Detail-preserving Vritual Try-On Network (DP-VTON). Our system references the geometric matching module in CP-VTON and adds a patch GAN in the try-on module. The objective of our system is preserving more details of clothes after tried on a person.

Figure 1: From left to right: person image, clothes image,try-on image

Data Preporcessing

- ❖ Data Collection
- ❖ Data processing
  - ➢ Pose shape
  - ➢ Binary Mask
  - ➢ Human Parsing

Figure 2:Data set from [1]

Figure 3: From left to right are person image, human parsing, cloth image, cloth mask

Figure 4: From left to right: reserved regions, pose heatmap, body shape

Human Parsing Tools

- ❖ Look Into Person (LIP) [3]:
  - ➢ Fail in images with pattern
  - ➢ Fail when colors are close
  - ➢ Ignore human neck
- ❖ Graphonomy [4]:
  - ➢ Parse human parts perfectly
  - ➢ Render more accurate edge

Figure 5: From left to right: original human image, human parsing result from LIP [3], and human parsing result from Graphonomy [4]

Method & Approach

- ❖ Geometric Matching Module (GMM) [2]

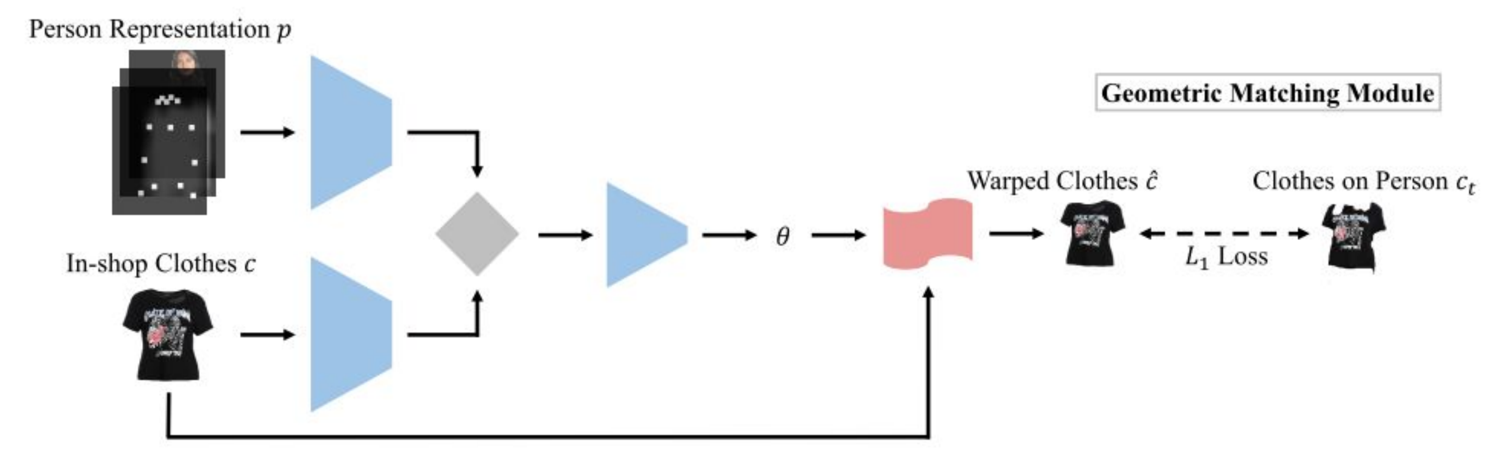

Figure 6: Structure of Geometric Matching Module [2]

- ❖ Try-on Module (TOM) with Patch GAN

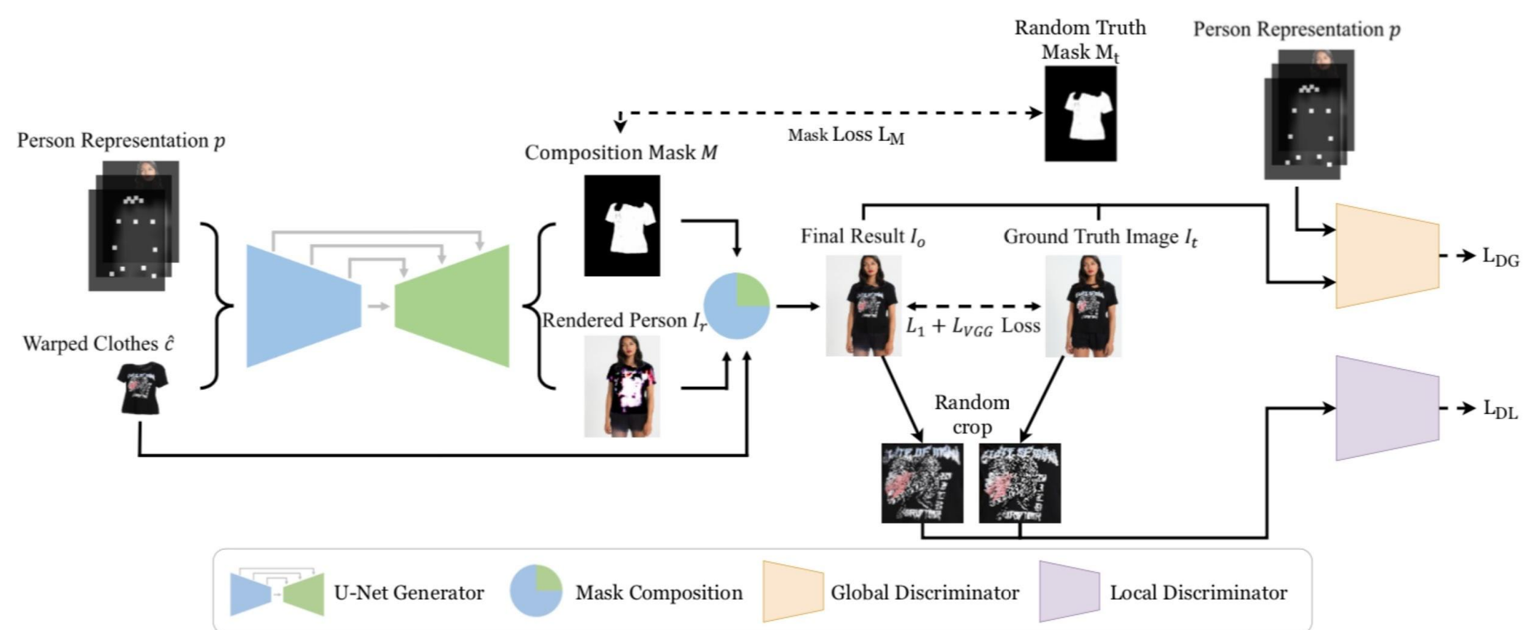

Figure 7: Structure of Try-on Module with Patch GAN

- ❖ Network Structure of Patch GAN

Layer	Filter Size	Stride	Output Size
Input			$D \times H \times W$
Conv2D Leaky-ReLU	4×4	2×2	$64 \times H/2 \times W/2$ $64 \times H/2 \times W/2$
Conv2D InstanceNorm2d Leaky-ReLU	4×4	2×2	$128 \times H/4 \times W/4$ $128 \times H/4 \times W/4$ $128 \times H/4 \times W/4$
Conv2D InstanceNorm2d Leaky-ReLU	4×4	2×2	$256 \times H/8 \times W/8$ $256 \times H/8 \times W/8$ $256 \times H/8 \times W/8$
Conv2D InstanceNorm2d Leaky-ReLU	4×4	1×1	$512 \times H/8 \times W/8$ $512 \times H/8 \times W/8$ $512 \times H/8 \times W/8$
Conv2D AvgPool2d	4×4 $H/8 \times W/8$	1×1	$1 \times H/8 \times W/8$ 1
Output			1

Table 1: Architecture of $\mathcal{D}_G$(D=28) & $\mathcal{D}_L$(D=3).

Experiment Result

Figure 8: From left to right: person images, warped clothes results using human parsing from LIP[3], and warped clothes results using human parsing from Graphonomy [4]

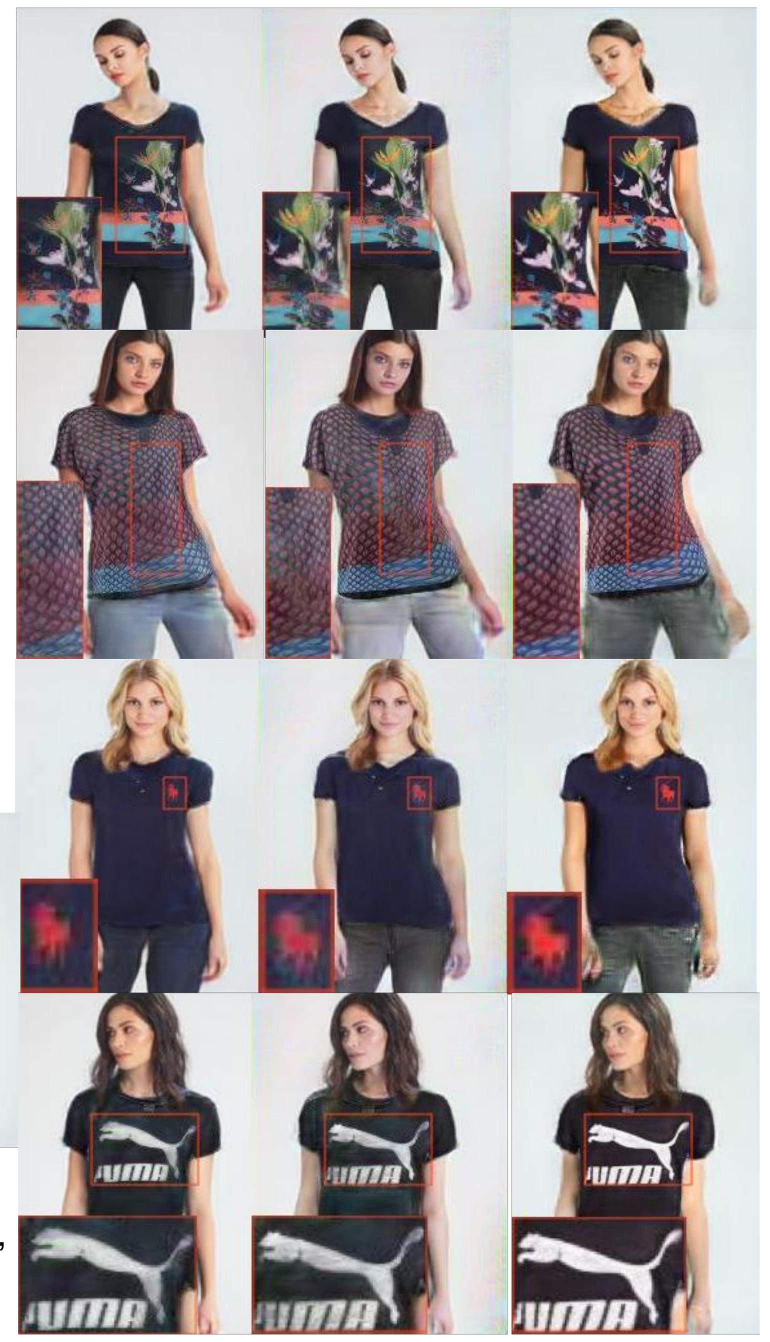

Figure 9: From left to right: results among CP-VTON, our DP-VTON without patch GAN, DP-VTON with patch GAN

Conclusions

Based on CP-VTON, we improved the details of generated images by making two updates on CP-VTON. One is that we substituted human parsing method in CP-VTON with a newest method, Graphonomy [4]. With the better parsed human images as part of inputs, the warping result of clothes generated by GMM model could be much more reasonable. The other contribution is that we introduced GAN in TOM, not only a global GAN, but also a local patch-wise GAN.

From the comparison of try-on results from CP-VTON, DP-VTON without patch GAN and our DP-VTON with patch-wise local GAN, we can see that our results preserve more details than other two methods. When there are logos, patterns, shades on the clothes, our results preserve the realistic details of shape and color.

References

[1] Xintong Han, Zuxuan Wu, Zhe Wu, Ruichi Yu, and Larry S. Davis. VITON: An Image-based Virtual Try-on Network. *arXiv e-prints*, page arXiv:1711.08447, 2017.

[2] Bochao Wang, Huabin Zhang, Xiaodan Liang, Yimin Chen, Liang Lin, and Meng Yang. Toward characteristic-preserving image-based virtual try-on network. *CoRR*, abs/1807.07688, 2018

[3] Ke Gong, Yiming Gao, Xiaodan Liang, Xiaohui Shen, Meng Wang, and Liang Lin. Graphonomy: Universal human parsing via graph transfer learning. *In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2019.

[4] Ke Gong, Xiaodan Liang, Xiaohui Shen, and Liang Lin. Look into person: Self-supervised structure-sensitive learning and A new benchmark for human parsing. *CoRR*, abs/1703.05446, 2017.

[5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv e-prints,* page arXiv:1505.04597, 2015

[6] Augustus Odena, Christopher Olah, and Jonathon Shlens. Conditional Image Synthesis With Auxiliary Classifier GANs. *arXiv e-prints*, page arXiv:1610.09585, 2016.