VITON-GAN: Virtual Try-on Image Generator Trained with Adversarial Loss

Shion Honda

The University of Tokyo / The International Research Center for Neurointelligence

Summary

- We develop a Virtual Try-on Generative Adversarial Network (VITON-GAN), that generates virtual try-on images from images of in-shop clothing and a model person.
- VITON-GAN enhances the quality of generated image when occlusion is present in a model person's image (e.g., arms crossed in front of the clothes).
- The main idea is to add an adversarial mechanism in the pipeline of the previous work called CP-VTON [3].

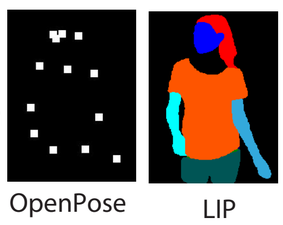
The models in the left column virtually wear the clothes from the top row.

Backgrounds and Methods

Despite the recent growth of online apparel shopping, there is a tremendous demand by consumers for buying clothes after trying them on in real shops. If e-commerce sites can offer virtual try-on images from a snapshot of the customer, they can improve their user experience.

CP-VTON [3]

The most successful approach in this field has been CP-VTON, that used **OpenPose** [1] and **Look Into Person (LIP)** [2] in its pipeline. The pose estimator and the human parser were used to extract clothes-agnostic person representations from a model's snapshot. This enabled training a mapping from a pair of an in-shop clothing image and a person representation to a model's snapshot.



OpenPose LIP

Clothes-agnostic person representation

CP-VTON consists of two modules: **geometry matching module (GMM)** and **try-on module (TOM)**. GMM is a neural network mapping an in-shop clothing image and a person representation into parameters for the thin plate spline (TPS) transform that warps the clothes to fit on the body. TOM is a U-net generator mapping a warped clothing image and a person representation into a rough try-on image and a mask. This mask was used to combine the rough try-on image and the warped clothing image.

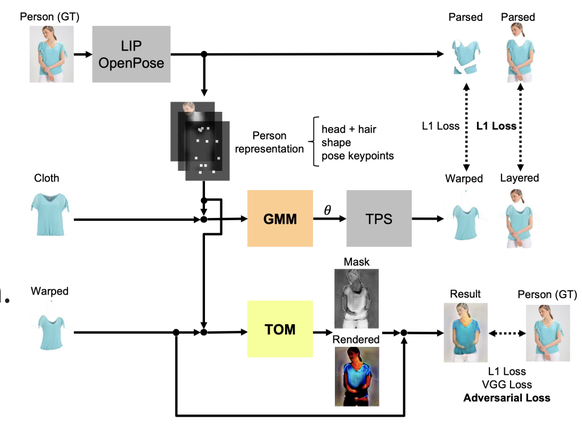
Occlusion problem

Although CP-VTON is able to preserve the patterns and letters of the clothes accurate, it fails when arms are crossed in front of the clothes (occlusion), generating blurred arms because it is trained only with reconstruction loss.

VITON-GAN (Proposed)

In VITON-GAN, we propose three new features.

- Added adversarial loss in TOM to tackle the occlusion problem. The discriminator tries to judge whether the given image is the ground truth or generated.
- Added new L1 loss in GMM to improve warping.
- Used random horizontal flipping both in GMM and TOM for data augmentation.



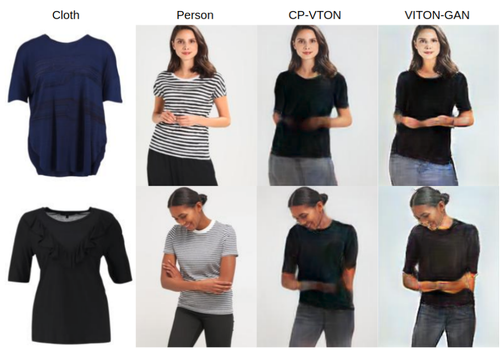
Training pipleline of VITON-GAN. New features are in the bold style.

Availability

The source codes and the trained model are available at: https://github.com/shionhonda/viton-gan/

Experiments and Results

- The dataset contained 16,253 female model's snapshots and top clothing image pairs.
- VITON-GAN generated hands and arms more clearly than CP-VTON in occlusion cases.
- However, arm generation failed when the model's original clothing was half-sleeve and the tried-on clothing was long-sleeve due to its topological change.



Successful cases.

Failed cases.

Conclusions

Here, we propose a virtual try-on image generator from 2D images of a person and top clothing that alleviates the occlusion problem. Future work will include improving the quality of generated parts of the human body and addressing topological changes in the clothes.

Acknowledgements

This work was supported by the Chair for Frontier AI Education, the University of Tokyo. Also, we would like to thank I. Sato, T. Harada, T. Mano, and C. Yokoyama for correcting this paper.

References

[1] Cao et al., CVPR, 2017. [2] Gong et al., CVPR, 2017. [3] Wang et al., ECCV, 2018.