Virtual Try-On: A Virtual Fashion Store

Prajwal M Patang
Centre of Excellence for Visual Intelligence
KLE Technological University
01fe18bec106@kletech.ac.in

Pratiksha R Naik
Centre of Excellence for Visual Intelligence
KLE Technological University
01fe18bec114@kletech.ac.in

Prajwal A Banagar
Centre of Excellence for Visual Intelligence
KLE Technological University
01fe18bec241@kletech.ac.in

A Saniya
Centre of Excellence for Visual Intelligence
KLE Technological University
01fe18bec351@kletech.ac.in

***Abstract*—In this project, we try to design deep neural network towards Virtual Try On, a network that shows us how a person would look if they were to wear a given cloth. In the past five years we can observe rapid growth in online shopping and the way technology has evolved to provide better experience for the customers. All these evolved with Augmented Reality based mirrors and applications redefining the way one discover and try on products for past three years. Virtual Try On enhances the shopping experience and gives the customers confidence to make the purchase. Our over arch of methodology includes estimating the pose of the person, semantic segmentation, geometric matching of cloth and image and warping of cloth. Deep Neural Networks provide efficient way to design our methodology using its key processes. This can be applied to the virtual fashion stores for costumers to know how they would look in that outfit.**

***Index Terms*—Virtual Try-On, PF-AFN**

I. Introduction

For the past couple of years, due to increasing online shopping and rent cost, offline stores have been tremendously threatened – Pandemic made this problem worse. Now companies have started looking for more desirable way to attract people be that online or in-store. Number of returns have also been increased because of lack of try on as they would have in offline shopping. Therefore, customers find it confusing and difficult to select appropriate style and suitable colour of garment for them. The objective here is to achieve Virtual Try-On and perform realistic rendering of clothes. Aim was to design an end-to-end pipeline such that given a pair of target cloth and person image, generate an output image of the person wearing the target cloth.

II. Related work

A. Toward Characteristic-Preserving Image-based Virtual Try-On Network [1]

CP-VTON outperforms the primitive virtual try-on methods by overcoming the spatial misalignments between the cloth and person's images due to application of coarse-to-fine architectures in previous methods, thus CP-VTON meets some critical requirements of virtual try-on. The two modules used in the architecture are GMM and TOM. The former learns the thin plate spline transformation and fits the cloth image like a wrapper on the person's image, whereas the latter alleviates the boundary artifacts of wrapped cloth and learns the make up the mask to mask the warped cloth and generate image. Though this architecture preserves the clothing details better than the primitive methods, yet there were some occlusions and loss of finer textural details of complex clothes

B. CP-VTON+: Clothing Shape and Texture Preserving Image-Based Virtual Try-On [2]

Performs better than CP-VTON and other methods while preserving the fine details of the cloth and the facial appearance of the human image. Two main modules were used in the paper which was same as that of CP-VTON with some improvements in input representation as well as the losses obtained. Some of the chest labels being wrongly labelled, unbalanced GMM inputs and losses and mask composition were all outperformed in the CP-VTON+ paper. The hair occlusion proved to be a deformity to body posture. In addition to the other labels a new label namely 'skin' was put on. This model proved to work certainly well for easy poses, mono-colored cloth to mask and matching sleeves but not for those of cloth with rich texture or sleeves with discrepancy. Some similarity check results were posed to determine the ablation of the paper to compare with previously proposed methods.

C. Towards Photo-Realistic Virtual Try-On by Adaptively Generating and preserving image content [3]

ACGPN performs well with different occlusions and poses. It can preserve the fine scale and details, unlike some other models where the details are lost. Since the main aim of ACGPN is to adaptively generate distinct human parts/skin , it sometimes generates result of skin occluding over cloth. ACGPN consists of 3 major modules, namely Layout Generation Module, Cloth Warping Module, Content Fusion Module . Apart from using the TPS warping technique, it also makes use of difference constraint i.e, second

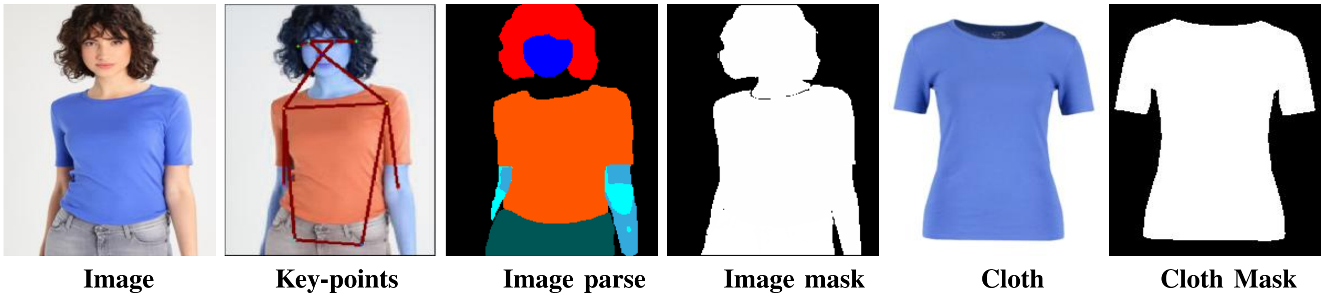

Fig. 1. Intermediate key-points and mask generation of the input images

order difference constraint which helps it to align the target cloth appropriately in line with target person image.

D. Parser-Free Virtual Try-on via Distilling Appearance Flows [4]

To address the problem of low image quality seen in the existing parser- based virtual try-on methods, this method uses the technique of knowledge distillation. PF-AFN treats the fake person image as input to the parser-free model that is supervised by the ground truth real person original image, to act as if the student is mimicking the teacher's knowledge. in this model, the PF-AFN network or the student network works on the generated image and puts its clothes onto the real person image. Due to the proposed tree block, the model could preserve more details and better fuse the spatially aligned cloth with the coarse rendered person image. The fine characteristics of clothes like stripes, logo and designs are preserved by the second order smoothening constraint used.

III. Proposed Approach

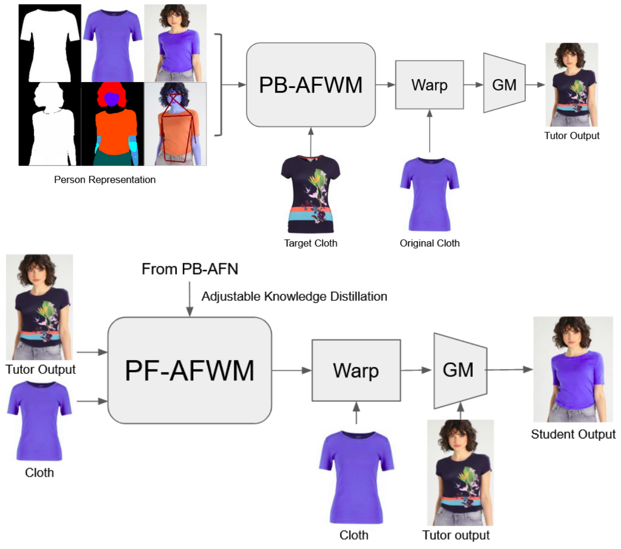

Fig. 2. Proposed Block Diagram. PB-AFN at the top and PF-AFN at the bottom. Person representation and the cloth image are input to the PB-AFN. The output from the PB-AFN are input to the PF-AFN

The model of Parser-Free Virtual Try-on via Distilling Appearance Flows has two main networks namely Parser Based Appearance Transfer Network (PB-AFN) and the Parser Free Appearance Transfer Network (PF-AFN). This model makes use of the technique knowledge distillation. Both the above mentioned networks consist of an Appearance Flow Warping module (AFWM). This module consists of two Pyramid Feature Extraction Network (PFEN) and Appearance Flow Estimation Network (AFEN). PFEN extracts pyramid level representation from the cloth image and person representation and at each level, the AFEN learns how to generate the coarse appearance flows.

Both main networks also have Generative modules. The parser based generative module concatenates the warped cloth, human pose keypoints and the preserved region onto the body. The parser free generative module uses the tutor knowledge as inputs. It concatenates the tutor image and the warped cloth. Both these generative modules are built upon a mixture of ResNet [5] and U-Net [6] architectures, which combined is called as the Res-U-Net architecture.

IV. Experiments

A. Dataset

We conducted experiment on VITON dataset. The VITON dataset has 19000 images of women and top clothes, all of which are of 256x192 pixels. Out of these, 16253 pairs are considered to be good and are selected. They are split into 14221 training pairs and 2032 testing pairs with the resolution of 256 x 192. Dataset along with its parsed and masked image for image and cloth is shown in Figure 1

B. Implementation Details

The pyramid feature extraction network used 5 FPN networks with 2 stride. AFEN had 5 flow network blocks having two ConvNets with 4 layers of convolution. The generative module made use of the comination of ResNet and UNet called the ResUNet architecture. Both the networks PB-AFN and the PF-AFN were trained for 200 epochs. The initial Learning Rate of the model was set to 3 x 10^{-5}. The hyperparameters set for PF-AFN were same as that of PB-AFN. The parameters set were: $\lambda_{l} = 1.0$, $\lambda_{p} = 0.2$, $\lambda_{sec} = 6.0$. $\lambda_{hint} = 0.04$, $\lambda_{pred} = 1.0$.

C. Results

1) Comparison of results for VITON DATASET:

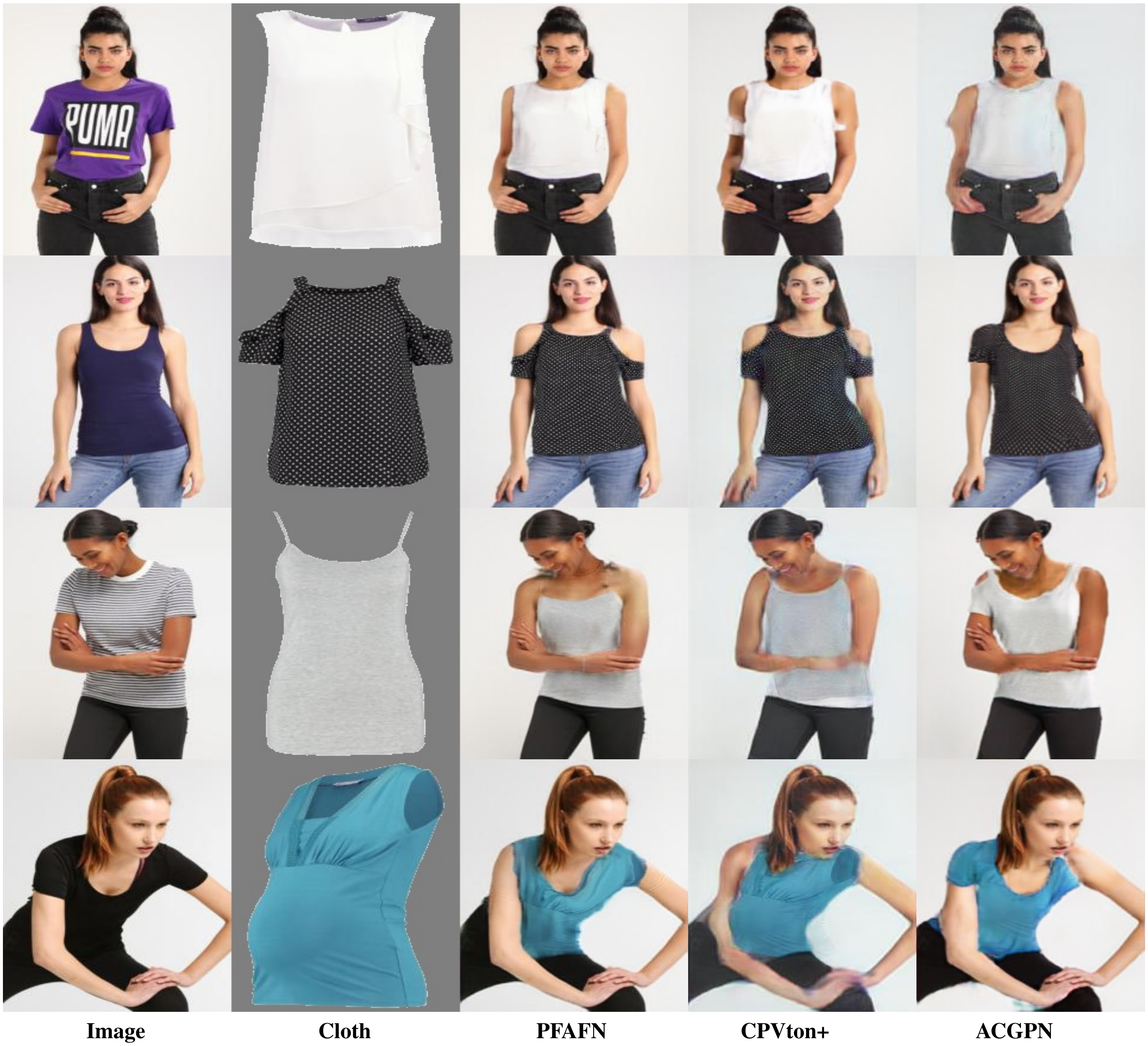

Fig. 3. The above figure shows the comparison of results of the proposed PF-AFN, CP-VTON+ and ACGPN. The first image is the person image, the second image is the target cloth. The following three images are try-on results of PF-AFN, CP-VTON+ and ACGPN respectively.

We performed the above said model along with two other models for comparison.
The results has been divided into five different categories as mentioned below.

1) Simple Pose-Simple Cloth combination result can be seen in first row in Fig 3.
2) Simple Pose-Complex Cloth combination in first row was second row in Fig 3.
3) Complex Pose-Simple Cloth result in third row in Fig 3.
4) Complex Pose-Complex Cloth results can be observed from fourth row in Fig 3.
5) The architectures were also tried on custom images

All the above mentioned categories have been implemented with PFAFN, CPVton+ and ACGPN for better analysis of results and architectures that can be seen from below figures.

Some Inferences from results were:

- Occlusion has been handled perfectly in PF-AFN and ACGPN, but the cloth shape is not retained in ACGPN.
- Cloth shape and texture is retained in case of PF-AFN and CPVton+

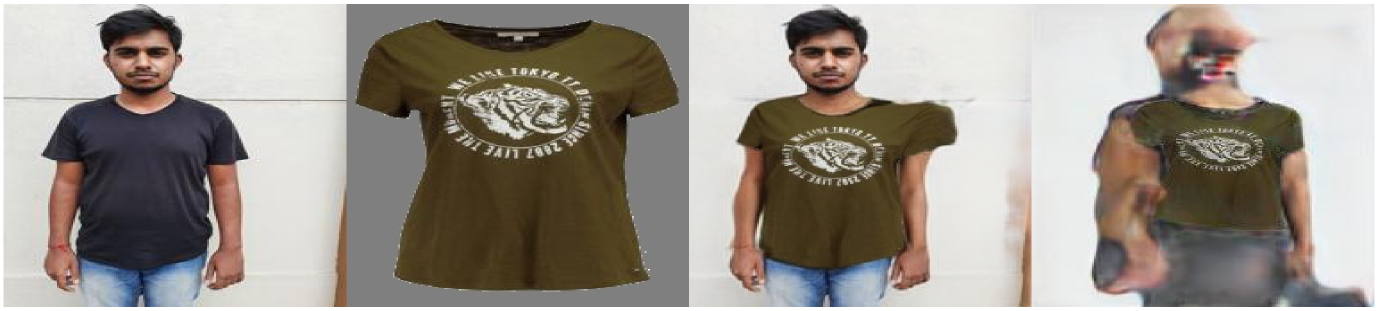

Fig. 4. Testing the Try-On for a custom image in uncontrolled environment led to distorted outputs.

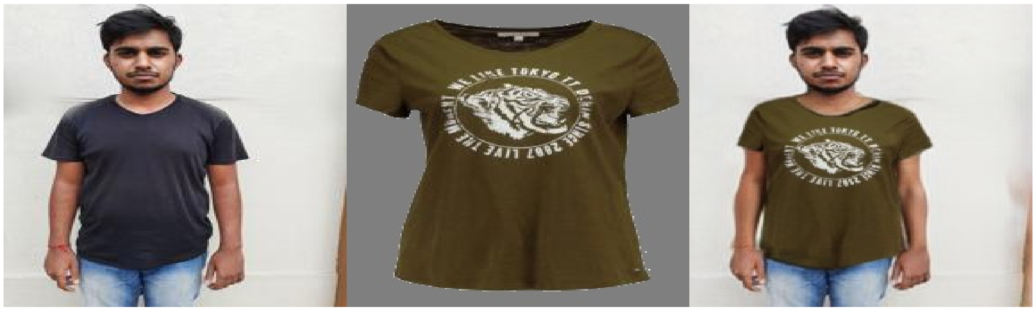

Fig. 5. The Try-On of the custom image by PF-AFN in a controlled environment produced better results compared to that in uncontrolled environment. No distortions were observed and all the desired features of the cloth and also the person were retained.

D. Conclusion

- PF-AFN generates better results than than CP-VTON, CP-VTON+ and ACGPN.
- CP-VTON+ also generates good results but has challenges for complex clothes and poses.
- Images should be taken in a proper controlled environment and must be taken in good lighting conditions.
- PF-AFN outperformed the rest other architectures.
- We were successful in implementing the above said model.
- Virtual Try-On can be used instead of tedious offline shopping or clueless selection of products.

References

[1] Bochao Wang, Huabin Zheng, Xiaodan Liang, Yimin Chen, Liang Lin, and MengYang. Toward characteristic-preserving image-based virtual try-on network. 2018.

[2] MR Minar, TT Tuan, H Ahn, P Rosin, and YK Lai. Cp-vton+: Clothing shapeand texture preserving image-based virtual try-on. InThe IEEE/CVF Conference onComputer Vision and Pattern Recognition (CVPR) Workshops, volume 2, page 11,2020

[3] Han Yang, Ruimao Zhang, Xiaobao Guo, Wei Liu, Wangmeng Zuo, and Ping Luo.Towards photo-realistic virtual try-on by adaptively generating-preserving image con-tent. InProceedings of the IEEE/CVF Conference on Computer Vision and PatternRecognition, pages 7850–7859, 2020.

[4] Yuying Ge, Yibing Song, Ruimao Zhang, Chongjian Ge, Wei Liu, and Ping Luo.Parser-free virtual try-on via distilling appearance flows. 2021.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun.Deep Residual Learning for Image Recognition. 2015

[6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networksfor biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M.Wells, and Alejandro F. Frangi, editors,Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, pages 234–241, Cham, 2015. Springer Interna-tional Publishing