School

of

Electronics and Communication Engineering

Minor Project Report

on

Virtual Try-On: A Virtual Fashion Store

By:

1. **Prajwal M Patang** USN:01FE18BEC106
2. **Pratiksha Naik** USN:01FE18BEC114
3. **Prajwal Banagar** USN:01FE18BEC241
4. **A Saniya** USN:01FE18BEC351

Semester: VI

Year: 2021

Under the Guidance of

Uma Mudenagudi

Ramesh Ashok Tabib

K.L.E SOCIETY'S
KLE Technological University,
HUBBALLI-580031
2021

SCHOOL OF ELECTRONICS AND COMMUNICATION ENGINEERING

CERTIFICATE

This is to certify that project entitled **"Virtual Try-On : A Virtual Fashion Store"** is a bonafide work carried out by the student team of **Prajwal M Patang (01FE18BEC106), Pratiksha Naik (01FE18BEC114), Prajwal Banagar (01FE18BEC241) and A Saniya (01FE18BEC351)**. The project report has been approved as it satisfies the requirements with respect to the minor project work prescribed by the university curriculum for BE (VI Semester) in School of Electronics and Communication Engineering of KLE Technological University for the academic year 2021.

Uma Mudenagudi **Ramesh Ashok Tabib** **Guide**	**Nalini C. Iyer** **Head of School**	**N. H. Ayachit** **Registrar**

External Viva:

Name of Examiners	**Signature with date**
1.	
2.	

ACKNOWLEDGMENT

Motivation, Inspiration, and presentation have played a key role in the success of this venture.

We would like to express our admiration to all those who provided us the possibility to complete this project. We would sincerely like to thank Uma Mudenagudi, Ramesh A. Tabib, Chaitra Desai and the CEVI-KLETECH for their active support and helpful insights while building this project. We offer our most sincere appreciation for the learning opportunuties provided to us. We owe our deepest gratitude to Dikshit Hegde and Sumitra Malagi for mentoring us throughout the project. We would also like to thank the School of Electronics and Communication Engineering, KLETECH for giving us this great learning opportunity.

-Project Team Members

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.

Contents

1 Introduction **9**

1.1 Motivation . . . 9

1.2 Objectives . . . 9

1.3 Literature survey . . . 9

1.4 Problem statement . . . 11

1.5 Application in Societal Context . . . 11

1.6 Organization of the report . . . 11

2 System design of Virtual Try-on **12**

2.1 Functional block diagram for try-on network . . . 12

2.2 Design alternatives . . . 13

2.2.1 CP-VTON . . . 13

2.2.2 CP-VTON+ . . . 14

2.2.3 ACGPN . . . 15

2.2.4 Parser-Free Virtual Try-on via Distilling Appearance Flows . . . 16

2.3 Final design: PF-AFN Approach . . . 16

3 Implementation details **17**

3.1 Dataset . . . 17

3.2 Training . . . 17

3.3 Algorithm . . . 18

3.4 Flowchart for Parser-Free Virtual Try-On . . . 19

4 Results and discussions **20**

4.1 Result Analysis . . . 20

5 Conclusions and future scope **23**

5.1 Conclusion . . . 23

5.2 Future scope . . . 23

References **23**

List of Figures

2.1 Block diagram of a virtual try-on network 12

2.2 CP-VTON Architecture taking input in-shop cloth, warped cloth and person representation in its two modules, GMM and TOM to generate output result. 13

2.3 Pipeline diagram of the CP-VTON+ architecture. The target and person image are given as input to the down sampling layers. The cloth is then warped according to the person representation. Warped cloth and the person representation are downsampled and upsampled to produce mask and rendered person. The mask and the rendered person are composed to give out the desired try-on result. 14

2.4 Pipeline diagram of the ACGPN architecture. Module 1 involves two GAN's to generate synthesized body image mask and cloth mask from target cloth, pose map and fused body model. The target clothing mask undergoes module 2 which involves STN to generate body part mask and image. Module 3 consists of GAN 3 which produces our required output image. 15

2.5 The training pipeline of PF-AFN. Training image - Cloth and person image. $p*$ is obtained from person image to randomly select different cloth images to generate fake image U_I as tutor. Tutor and clothes image are inputted to PF-AFN to generate student S_I to be directly supervised by real image. 16

3.1 Parser-Based Appearance Flow Network: Takes person representation and cloth image as input and gives the warped module and generative module as the outputs. 19

3.2 Parser-Free Appearance Flow Network: Takes the knowledge from the PB-AFN Tutor network and gives the warped module of the PF-AFN as outputs. 19

4.1 Cloth texture is retained in case of PF-AFN, and is not retained in CPVton+, while the cloth is not rendered correctly in ACGPN 20

4.2 Cloth texture and type is retained in case of PF-AFN, cloth type is not retained in CPVton+, while the cloth is rendered correctly but extra skin is generated in ACGPN. 21

4.3 Cloth shape is retained in case of PF-AFN and CPVton+, while the cloth is not rendered correctly in ACGPN. The quality of CPVton+ is inferior compared to PF-AFN 21

4.4 Occlusion has been handled perfectly in PF-AFN and ACGPN, but the cloth shape is not retained in ACGPN. 21

4.5 Cloth does not fit the body in CP-Vton+, whereas cloth shape is not perfectly rendered in ACGPN. 22

4.6 PFAFN and CP-Vton+ results for custom image in an uncontrolled environment 22
4.7 In a controlled environment the cloth is rendering correctly on a target custom image whereas in an uncontrolled environment we can observe distorted output. 22

Chapter 1

Introduction

In this chapter, we would be looking at why we need Virtual Try-On and how it would be beneficial in these days.

1.1 Motivation

- 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.

1.2 Objectives

- Achieve Virtual Try-On
- Perform realistic rendering of clothes.

1.3 Literature survey

1. **Toward Characteristic-Preserving Image-based Virtual Try-On Network (Wang et al. ECCV 2018)**[5]
   Authors : Bochao Wang, Huabin Zheng, Xiaodan Liang, Yimin Chen, Liang Lin, and Meng Yang.
   Published in : ECCV 2018
   Inference : CP-VTON outperforms the primitive virtual try-on methods by overcoming the spatial misalignments between the cloth and person's images due to 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 Geometric matching module and Try-on module. 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 composition mask to integrate 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

2. **CP-VTON+: Clothing Shape and Texture Preserving Image-Based Virtual Try-On : (Minar et al. CVPRW 2020**[3]
   Authors : Matiur Rahman Minar, Thai Thanh Tuan, Heejune Ahn, Paul L. Rosin, Yu-Kun Lai.
   Published in : CVPRW 2020
   Inference : 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.

3. **Towards Photo-Realistic Virtual Try-On by Adaptively Generating↔Preserving Image Content: (CVPR 2020)**[6]
   Authors : Han Yang, Ruimao Zhang, Xiaobao Guo, Wei Lu, Wangmeng Zuo, Ping Luo..
   Published in : CVPR 2020

   Inference : ACGPN performs well with various poses even with occlusions and cross-arms. It can preserve the fine scale details, unlike VITON, CP-VTON and VTNFP which are easily lost in them . Since the main aim of ACGPN is to adaptively generate distinct human parts/skin , it sometimes generates result of skin occluding over cloth. Followed by three modules i.e., Semantic Generation Module , Cloth Warping Module and Content Fusion Module it preserves clothes with complex texture . Since apart from TPS it also makes use of second order difference constraint which helps it to align the target cloth appropriately in line with target person image.

4. **Parser-Free Virtual Try-on via Distilling Appearance Flows : (Ge et al. CVPR 2021)**[1]
   Authors : Ge Yuying, Yibing Song, Ruimao Zhang, Ge Chongjian and Wei Lu,

Ping Luo.
Published in : CVPR, March 2021
Inference : 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 th 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 PFAFN 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.

1.4 Problem statement

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.

1.5 Application in Societal Context

- Virtual try-on technology helps customer to try products which they like using their own camera devices. This will help the customer to try the product in several ways for selecting the style and size fit before purchasing.
- According to a report by statista, in 2017 the worth of AR market was 3.5 billion dollars but now developing exponentially and may reach up to 198 billion dollars by 2025.
- In the view of pandemic, shopping and other main chores are being done online. Also in colleges graduation ceremonies are being held online, Virtual Try-on can be used for wearing academic dress/Graduation Gown.
- Virtual Try-On let customers to judge how certain products looks on them before they really buy and order the item. This give customers a complete liberty of making the decision, trying and selecting products at their own comfort, without actually feeling pressurized about buying something and later regretting about it.

1.6 Organization of the report

Chapter 2 describes the system design with functional block diagram and design alternatives.
Chapter 3 describes the implementation details of the project.
Chapter 4 describes the results obtained in the project.
Chapter 5 describes the conclusion drawn from the project implementation and the future scope is discussed.

Chapter 2

System design of Virtual Try-on

In this chapter, we explore different possibilities to design our framework. We implement various methods for our framework and select the best-suited one.

2.1 Functional block diagram for try-on network

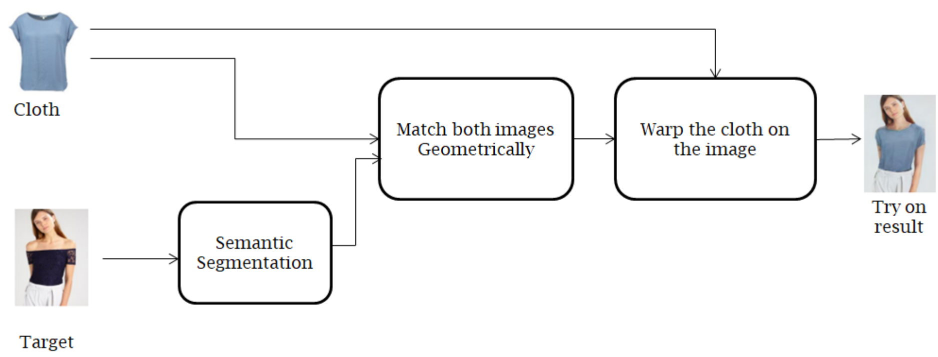

Figure 2.1: Block diagram of a virtual try-on network

As shown in Figure 2.1 the functional block diagram consists of semantic segmentation,matching of cloth and target images geometrically and warping the cloth on image. In semantic segmentation function we will obtain the semantic segmentation of target image and generate its parsing points. These points are used to align the target cloth in the pose of the given human body. And using this functionality, we can obtain our Try On result.

2.2 Design alternatives

2.2.1 CP-VTON

This architecture consists of two modules namely, Geometric Matching Module and Try On Module. Geometric module learns the thin plate spline transformation using convolutional neural network, in order to warp the cloth image on the person's image like a wrapper, rather than taking pixel-wise correspondence. First they extract pose, body shape and face from a given reference image and then cloth warp is generated. Try-On module fuses the warped cloth with the target person's image. Here U-Net renders the person's image and predict the composition mask. This composition mask is used to fuse target person's image with cloth and generator generates the final Try-On image. At last, the refinement network is utilized to learn how to integrate the warped clothing item to the reference image.

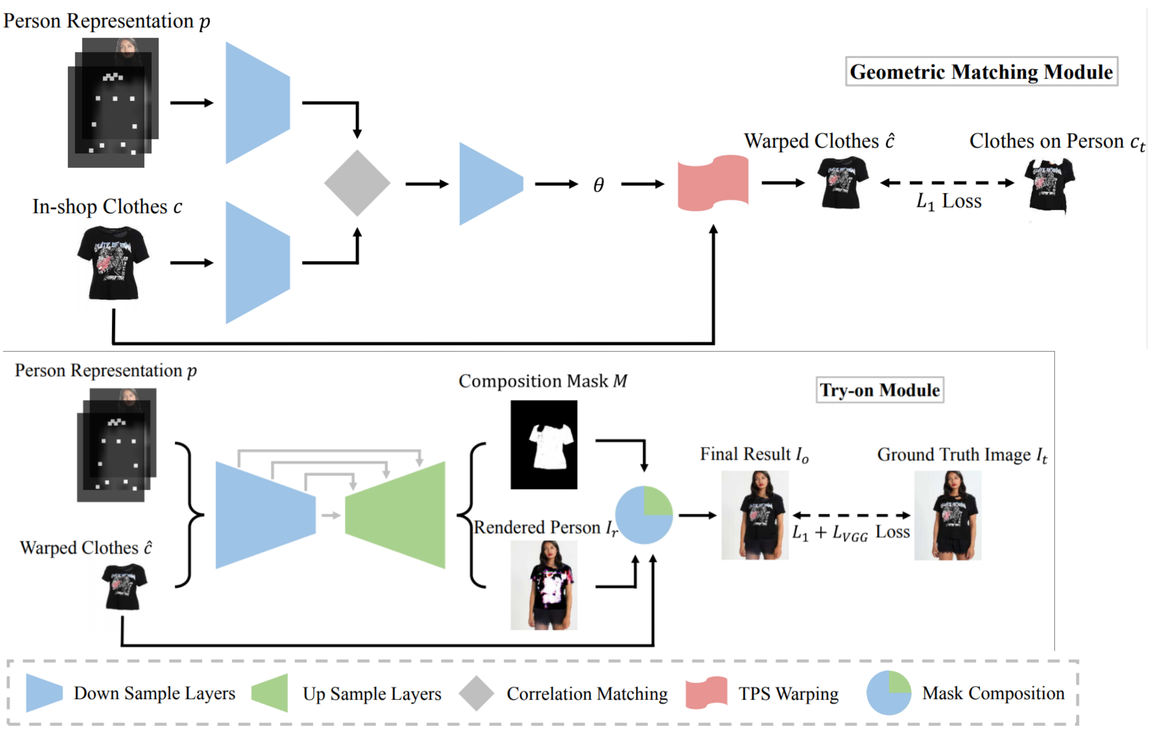

Figure 2.2: CP-VTON Architecture taking input in-shop cloth, warped cloth and person representation in its two modules, GMM and TOM to generate output result.

2.2.2 CP-VTON+

This design is based on the pipeline structure of CP-VTON[5] hence named CP-VTON+. The input to this model will be the person representation and the target cloth. The person representation and the target cloth is both fed to two down sampling layers. The two down sampled outputs are fed to correlation matching and again fed to the down sampling layers. After this process the cloth has to b warped. This is done using TPS warping technique. This will result in warped cloth and warped cloth mask. This concludes the GMM. For TOM, the person representation and the cloth mask and cloth is fed to the down sampling and up sampling layers to generate the composition mask and the rendered person. The composition mask and rendered person after mask composition will give the desired result of cloth being masked on the person, satisfying the Virtual Try-On. This architecture works pretty well for mono-coloured clothes but not for cloth with rich texture.

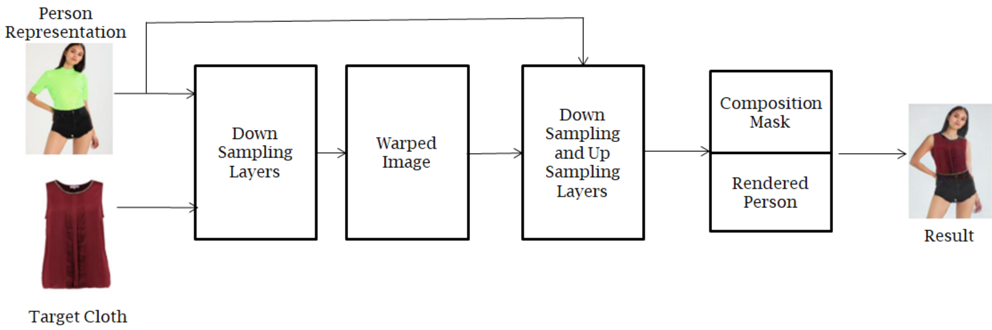

Figure 2.3: Pipeline diagram of the CP-VTON+ architecture. The target and person image are given as input to the down sampling layers. The cloth is then warped according to the person representation. Warped cloth and the person representation are downsampled and upsampled to produce mask and rendered person. The mask and the rendered person are composed to give out the desired try-on result.

The training losses and the person representation was re-employed giving comparatively better results. Parameters used were: $\lambda_{\text{VGG}} = 1, \lambda_1 = 1, \lambda_{\text{mask}} = 1, \lambda_{\text{reg}} = 0.5$.

2.2.3 ACGPN

ACGPN consists of 3 major modules:

1. *Layout Generation Module*: This module will help to separate the portion of the body where the cloth has to be masked retaining the other parts of the body with two GANs. GAN1 will generate the synthesised body part mask and GAN2 will generate the synthesised cloth mask.

2. *Cloth Warping Module*: According to the generated semantic layout, clothes are warped. In order to do that, Spatial Transformation Network can be used by applying thin plate spline transformation. In order to make the warping more appropriate and also to preserve complex textures of cloth, second order difference constraint is added in the warping loss.

3. *Content Fusion Module*: It adds up the features from the generated body mask, original body and the warped cloth to determine generating and preserving human body parts in the generated image.

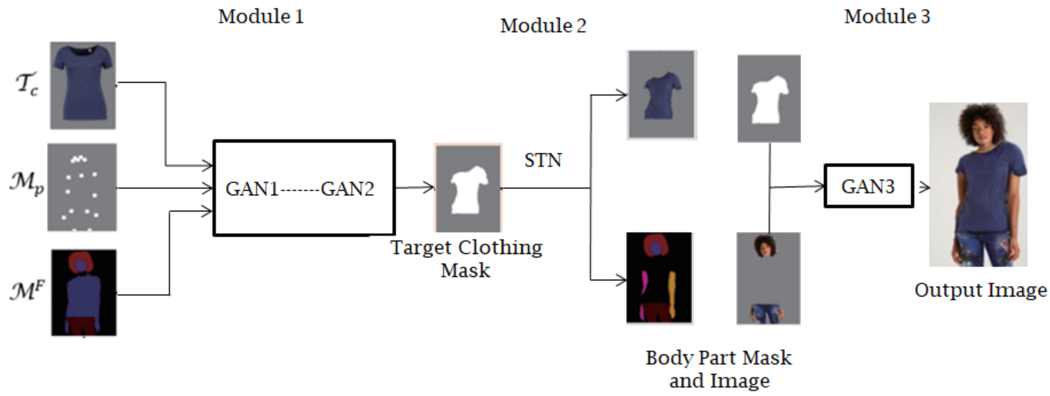

Figure 2.4: Pipeline diagram of the ACGPN architecture. Module 1 involves two GAN's to generate synthesized body image mask and cloth mask from target cloth, pose map and fused body model. The target clothing mask undergoes module 2 which involves STN to generate body part mask and image. Module 3 consists of GAN 3 which produces our required output image.

2.2.4 Parser-Free Virtual Try-on via Distilling Appearance Flows

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[2] and U-Net[4] architectures, which combined is called as the Res-U-Net architecture.

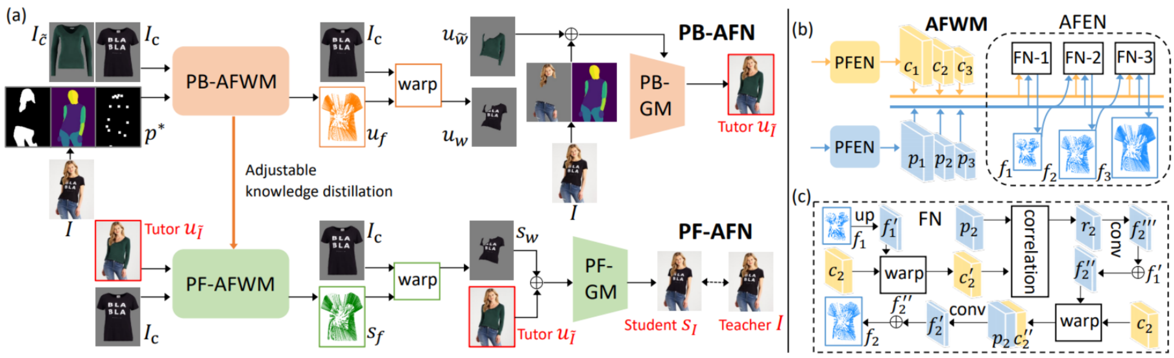

Figure 2.5: The training pipeline of PF-AFN. Training image - Cloth and person image. p^* is obtained from person image to randomly select different cloth images to generate fake image U_I as tutor. Tutor and clothes image are inputted to PF-AFN to generate student S_I to be directly supervised by real image.

2.3 Final design: PF-AFN Approach

The one optimal solution for effectiveness, ease of implementation as well as working was PF-AFN. Hence the architecture adopted in the PF-AFN paper was used for the implementation.

Chapter 3

Implementation details

In this chapter, we discuss the specifications and final system architecture with its algorithm and flowchart.

3.1 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.

Image **Keypoints** **Image parse** **Image mask** **Cloth** **Cloth Mask**

3.2 Training

Both PB-AFN and PF-AFN have been trained for 200 epochs. The initial learning rate was set to 3 x 10^{-5}. The hyperparameters set for PF-AFN were same as that of PB-AFN. The parameters set were: $\lambda_{\mathrm{l}} = 1.0$, $\lambda_{\mathrm{p}} = 0.2$, $\lambda_{\mathrm{sec}} = 6.0$. $\lambda_{\mathrm{hint}} = 0.04$, $\lambda_{\mathrm{pred}} = 1.0$.

3.3 Algorithm

Algorithm 1 PBAFN

Input: Person representation and cloth image
Output: Warp module and generative module

1. initialize warp and GM module randomly at first.
2. **for** 200 epochs **do**
3. get person representation with masked cloth image
4. predict appearance flow between person and cloth
5. warp cloth wrt appearance flows predicted
6. synthesise images with person and warped cloth
7. calculate losses
8. update modules
9. **end for**

Algorithm 2 PFAFN

Input: Parser based AFN - tutor network
Output: Warp Module and generative module of PF-AFN

1. randomly initialise the warp and generative module
2. **for** 200 epochs **do**
3. randomly select cloth image
4. obtain the person representation with cloth masked
5. predict appearance flow between person representation and cloth selected
6. warp selected cloth to appearance flow
7. synthesize fake image from person representation and warped cloth
8. calculate the losses
9. minimise the loss to update the warp and generative module
10. **end for**

3.4 Flowchart for Parser-Free Virtual Try-On

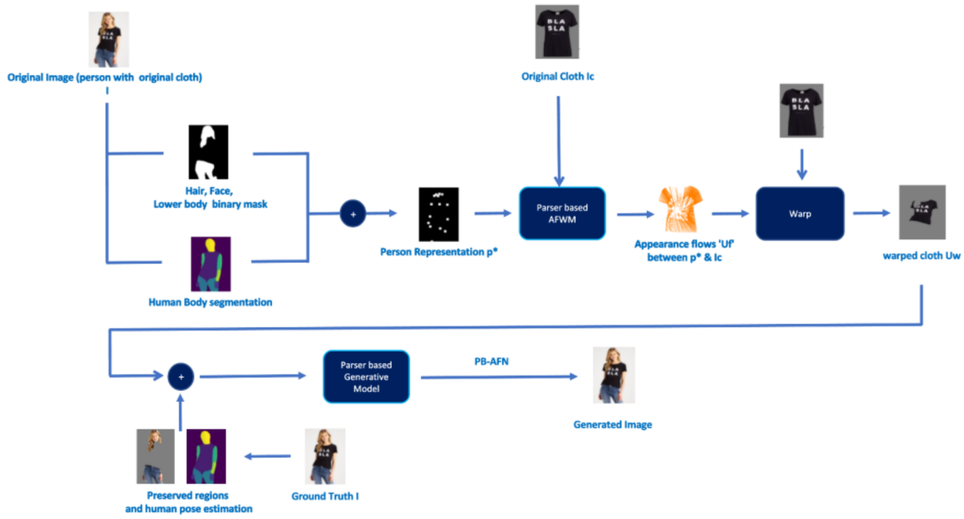

Figure 3.1: Parser-Based Appearance Flow Network: Takes person representation and cloth image as input and gives the warped module and generative module as the outputs.

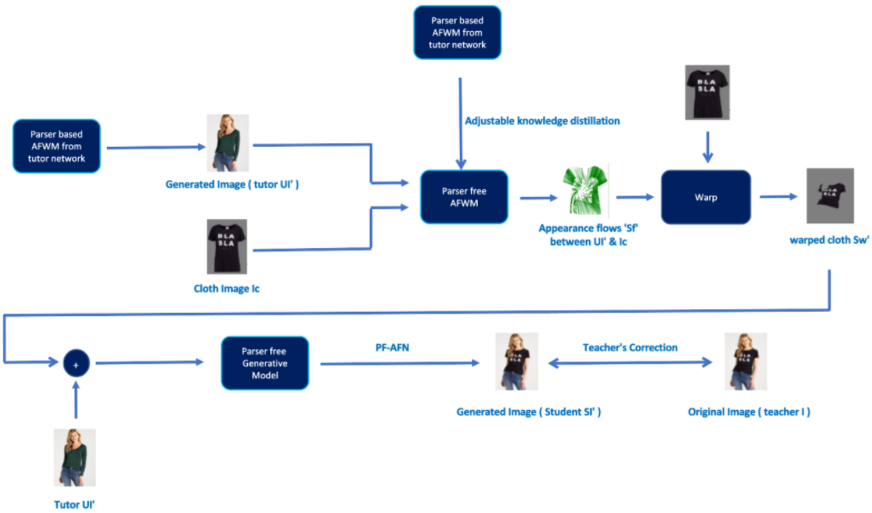

Figure 3.2: Parser-Free Appearance Flow Network: Takes the knowledge from the PB-AFN Tutor network and gives the warped module of the PF-AFN as outputs.

Chapter 4

Results and discussions

4.1 Result Analysis

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. General Results for a target cloth on target image can be seen in Figure 4.1.
2. Simple Pose-Simple Cloth combination result can be seen in Figure 4.2.
3. Simple Pose-Complex Cloth combination in Figure 4.3.
4. Complex Pose-Simple Cloth result in Figure 4.4.
5. Complex Pose-Complex Cloth results can be observed from Figure 4.5
6. 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.

Image **Cloth** **PFAFN** **CPVton+** **ACGPN**

Figure 4.1: Cloth texture is retained in case of PF-AFN, and is not retained in CPVton+, while the cloth is not rendered correctly in ACGPN

Image **Cloth** **PFAFN** **CPVton+** **ACGPN**

Figure 4.2: Cloth texture and type is retained in case of PF-AFN, cloth type is not retained in CPVton+, while the cloth is rendered correctly but extra skin is generated in ACGPN.

Image **Cloth** **PFAFN** **CPVton+** **ACGPN**

Figure 4.3: Cloth shape is retained in case of PF-AFN and CPVton+, while the cloth is not rendered correctly in ACGPN. The quality of CPVton+ is inferior compared to PF-AFN

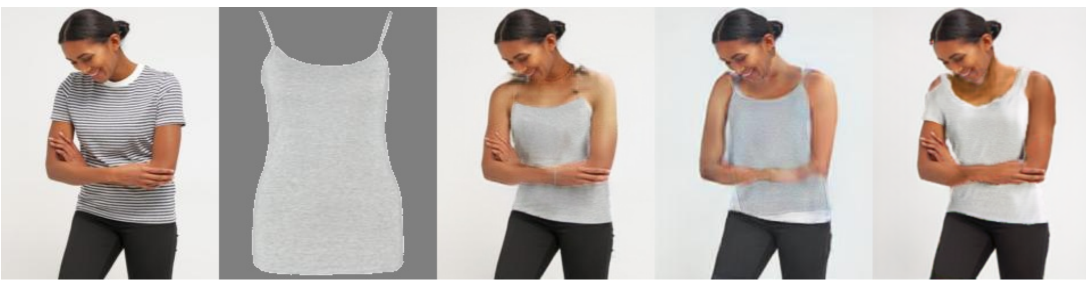

Image **Cloth** **PFAFN** **CPVton+** **ACGPN**

Figure 4.4: Occlusion has been handled perfectly in PF-AFN and ACGPN, but the cloth shape is not retained in ACGPN.

Image **Cloth** **PFAFN** **CPVton+** **ACGPN**

Figure 4.5: Cloth does not fit the body in CP-Vton+, whereas cloth shape is not perfectly rendered in ACGPN.

Custom Image Results

Image **Cloth** **PFAFN** **CPVton+**

Figure 4.6: PFAFN and CP-Vton+ results for custom image in an uncontrolled environment

Image **Cloth** **PFAFN**

Figure 4.7: In a controlled environment the cloth is rendering correctly on a target custom image whereas in an uncontrolled environment we can observe distorted output.

Chapter 5

Conclusions and future scope

In this chapter, we brief our conclusion and future scope for Virtual Try-On.

5.1 Conclusion

We were successful in implementing the above said model.

- The distortions like fine finger details can be overcome by improving the loss function.
- PF-AFN generates better results 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.
- Virtual Try-On can be used instead of tedious offline shopping or clueless selection of products.

5.2 Future scope

This concept of virtual try-on can be used by online stores and help their costumers know how they would look in that attire. Collecting the appropriate data of both target cloth and image is necessary towards which we are working. Also this project is useful for online shopping enthusiasts to reduce their time and number of returns. This Virtual Try-On can be extended to 3D and real-time also.

Bibliography

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

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

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

[4] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for 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 International Publishing.

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

[6] Han Yang, Ruimao Zhang, Xiaobao Guo, Wei Liu, Wangmeng Zuo, and Ping Luo. Towards photo-realistic virtual try-on by adaptively generating-preserving image content. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7850–7859, 2020.

Virtual Try-On

ORIGINALITY REPORT

12%	10%	6%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	arxiv.org Internet Source	6%
2	minar09.github.io Internet Source	1%
3	export.arxiv.org Internet Source	1%
4	Submitted to University of Michigan-Shanghai Jiao Tong University Joint Institute Student Paper	1%
5	"Computer Vision – ECCV 2020", Springer Science and Business Media LLC, 2020 Publication	1%
6	Submitted to City University of Hong Kong Student Paper	1%
7	Submitted to University of Warwick Student Paper	<1%
8	www.scilit.net Internet Source	<1%

9	Matiur Rahman Minar, Thai Thanh Tuan, Heejune Ahn. "CloTH-VTON+: Clothing Three-Dimensional Reconstruction for Hybrid Image-Based Virtual Try-ON", IEEE Access, 2021 Publication	<1 %
10	Lecture Notes in Computer Science, 2015. Publication	<1 %
11	"Computer Vision – ACCV 2020", Springer Science and Business Media LLC, 2021 Publication	<1 %
12	Matteo Fincato, Federico Landi, Marcella Cornia, Fabio Cesari, Rita Cucchiara. "VITON-GT: An Image-based Virtual Try-On Model with Geometric Transformations", 2020 25th International Conference on Pattern Recognition (ICPR), 2021 Publication	<1 %

Exclude quotes	Off	Exclude matches	< 5 words
Exclude bibliography	On		