School

of

Electronics and Communication Engineering

Mini Project Report

on

Virtual Try-On: 2D clothing try-on

By:

1. **Adnan Ladji** 01FE19BEC252
2. **Niranjan Bhuti** 01FE19BEC248
3. **Shivkumar Ghamani** 01FE19BEC084

Semester: V, 2021-2022

Under the Guidance of

Uma Mudenagudi

Ramesh Ashok Tabib

K.L.E SOCIETY'S
KLE Technological University,
HUBBALLI-580031
2021-2022

SCHOOL OF ELECTRONICS AND COMMUNICATION ENGINEERING

CERTIFICATE

This is to certify that project entitled **" Virtual Try-On: 2D clothing try-on "** is a bonafide work carried out by the student team of **"Adnan Ladji (01FE19BEC252), Niranjan Bhuti (01FE19BEC248), Shivkumar Ghamani (01FE19BEC084) "**. The project report has been approved as it satisfies the requirements with respect to the mini project work prescribed by the university curriculum for BE (V 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 opportunities provided to us. We owe our deepest gratitude to Dikshit Hegde and Prajwal Banagar 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.

Adnan Ladji-01FE19BEC252

Niranjan Bhuti-01FE19BEC248

Shivkumar Ghamani-01FE19BEC084

ABSTRACT

In this project, we try to design a 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

List of Figures

Chapter 1

Introduction

When we talk about shopping for clothes, it is one of the most enjoyable experiences for a customer, but at the same time can become cumbersome and tiring. The usual hassles involved in traditional in-store shopping experiences like standing in long queues or having to dress-undress again and again can make the whole experience a little dull. With virtual try-on, all these hassles are done away with.

With virtual try-on, a customer can try different clothing items without having to go through any of the above-mentioned hassles, right from the comfort of their homes.

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.

When it comes to marketing and sales, a virtual try-on can make a big difference for firms. A virtual try-on provides customers with a realistic, personalised product, whether it's a smart mirror in a physical store or a customised mobile application.presentation in a matter of seconds Customers may try out dozens of things right now, and they can even buy them.Get product recommendations that are relevant to your current purchase, which will encourage you to buy more.Having the opportunity to try on a product makes purchasing selections much easier for customers, especially when shopping online. Knowing how the desired product fits them enhances the likelihood of purchasing it and being satisfied with it. This improves and personalises the entire user experience. Customers are more inclined to shop from the same brand again as a result.

1.2 Objectives

- Achieve Virtual Try-On.

  Given with a pair of human image and a target cloth the deep learning algorithm should be able to produce a human image wearing the target clothes

- Perform realistic rendering of clothes.

  The algorithm should be able to achieve realistic rendering of clothes so that the results are accurate and give good quality images.

- Compare virtual try-on algorithms available.

  comparing different algorithms such as PFAFN, CP-VTON, ACGPN for different types of images.

1.3 Literature survey

1. **CP-VTON+: Clothing Shape and Texture Preserving Image-Based Virtual Try-On : (Minar et al. CVPRW)**
   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.

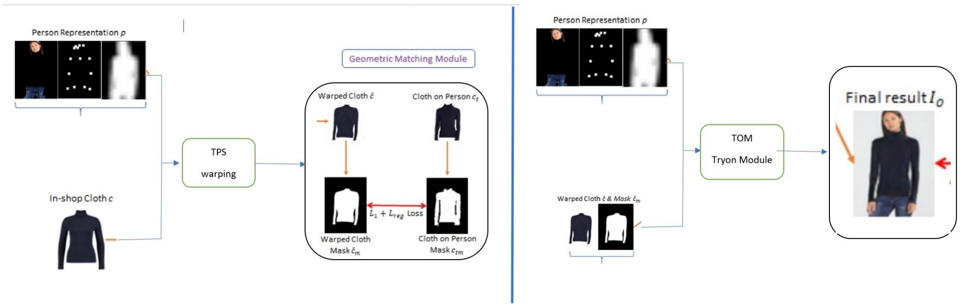

Figure 1.1: Block diagram of CP-VTON+ Model

2. **Towards Photo-Realistic Virtual Try-On by Adaptively GeneratingPreserving Image Content: (CVPR 2020)**[4]

   Inference : ACGPN is capable of handling a variety of positions, including occlusions and cross-arms. Without the secondary constraint, the warping module may produce deformed images. It is able to preserve the fine scale details which are easily lost in other methods like VITON, CP-VTON

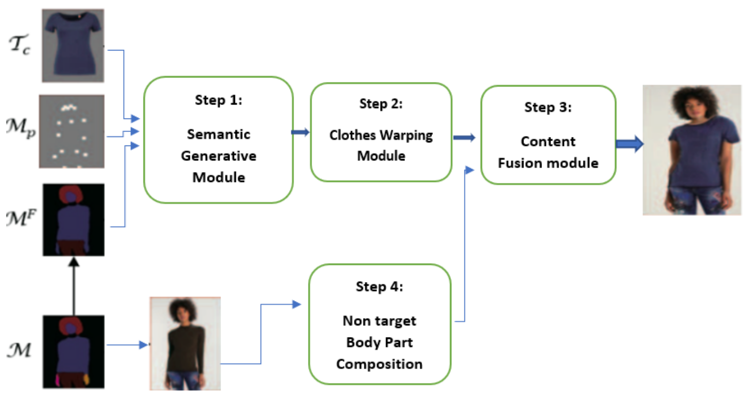

Figure 1.2: Block diagram of ACGPN Model

3. **Parser-free appearance flow network**

Inference : This model could keep more details and better merge the spatially aligned fabric with the coarse produced person image thanks to the proposed tree block.

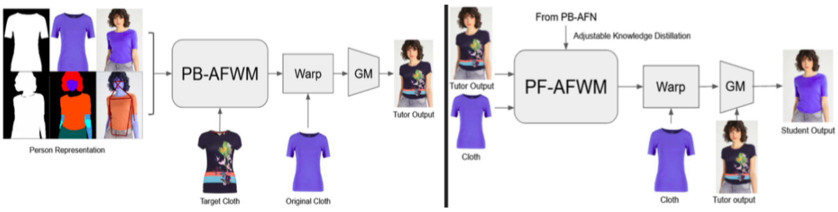

Figure 1.3: Block diagram of PFAFN Model

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.
- The applications prepared by fashion giants like nike, loreal, gucci etc are designed to scan the customer body shape and size using AR technology. The application allows customers to find the right item size and visualize how the item will look on the customer. This solution is becoming very trendy these days.

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 model

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 Virtual Try-On model

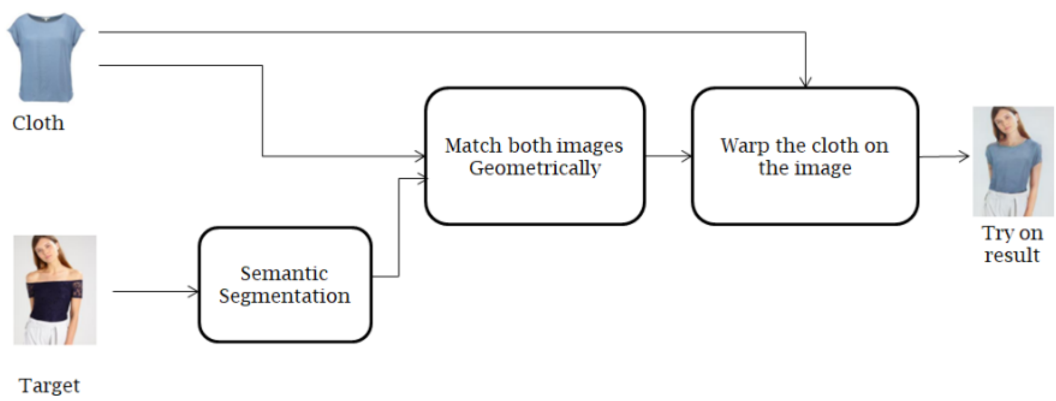

Figure 2.1: Block diagram of a virtual try-on network

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+

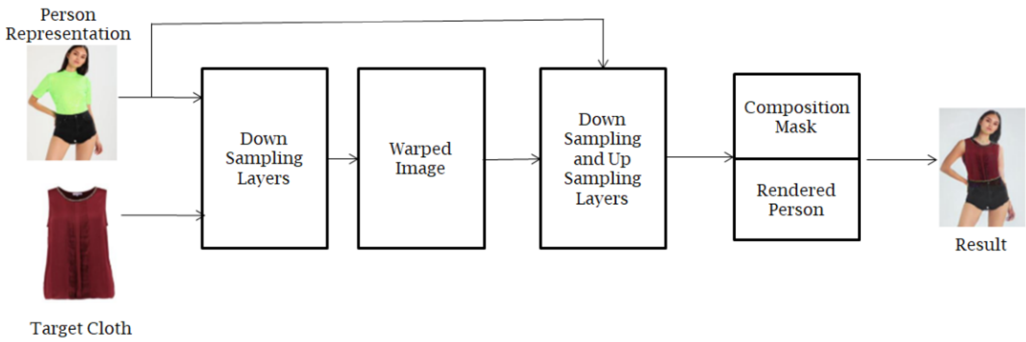

Figure 2.2: Pipeline diagram of the CP-VTON+ architecture. The target and person image are given as input to the down sampling layers. The cloth is then warped according to the person representation. Warped cloth and the person representation are downsampled and up sampled to produce mask and rendered person. The mask and the rendered person are composed to give out the desired try-on result.

This design is based on the pipeline structure of CP-VTON hence named CP-VTON+. It consists of two main stages cloth warping stage and Blending Stage. As shown in figure 2.4, cloth warping stage works in three main aspects. As output, we get a warped cloth. In the second stage, try on module network gets added with the binary mask of warped clothing.

This design suggested CPVTON+, a revised image-based VTON system that addresses difficulties with previous techniques such as human representation and dataset inaccuracies, network architecture, and a sloppy cost function. Despite the fact that CP-VTON+ enhances performance, a 2D image-based technique is discovered to have inherent limitations when dealing with diversely posed target human situations. As a result, the application would be restricted to basic attire and standardly positioned human targets. 3D reconstruction would be more appropriate for more diversified cases.

2.2.2 ACGPN

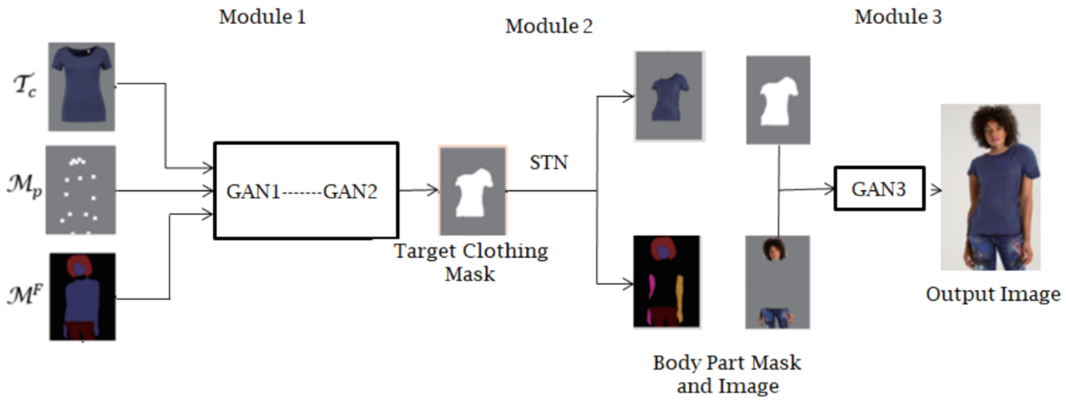

Figure 2.3: Pipeline diagram of the ACGPN architecture. Module 1 involves two GAN's to generate synthesized body image mask and cloth mask from target cloth,pose map and fused body model. The target clothing mask undergoes module 2 which involves STN to generate body part mask and image. Module 3 consists of GAN 3 which produces our required output image.

This design offers the ACGPN, a unique adaptive content generating and preservation network that attempts to generate photo-realistic try-on results while maintaining both the character of clothes and nuances of human identity (posture, body parts, bottom clothes). Mask Generation Module (GMM), Clothes Warping Module (CWM), and Content Fusion Module are three meticulously built modules (CFM). On the VI-TON [12] dataset, we test our ACGPN on three levels of try-on difficulty. In terms of quantitative measures, visual quality, and user studies, the results clearly reveal that ACGPN outperforms state-of-the-art approaches.

ACGPN consists of 3 major modules:

1. **Layout Generation Module**: GAN1 is used to train the body in order to generate synthesized body part mask by using information from fused map, pose map and target cloth and GAN2 is used to generate synthesize clothing mask.

2. **Cloth Warping Module**: According the generated semantic layout, clothes are warped. In order to do that, Spatial Transformation Network can be used applying thin plate spline. A second order difference constraint on warping network is introduced to avoid misalignments and blurry results.

3. **Content Fusion Module**: Combines data from the warped image, the original image, and the synthesised body part image to determine the preservation and production of distinct human parts in the synthetic image.

2.2.3 PF-AFN

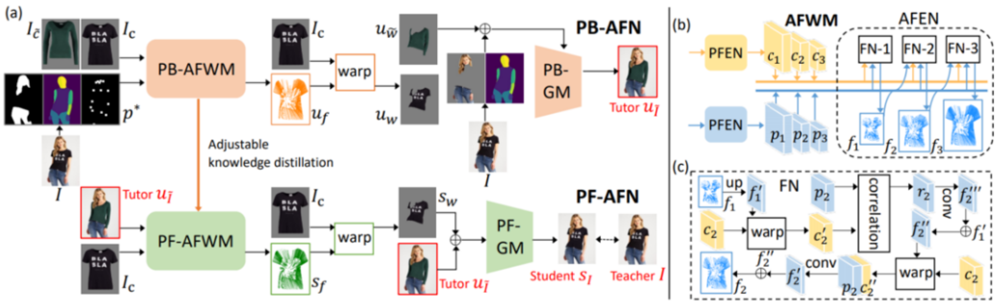

Figure 2.4: The training pipeline of PF-AFN. Training image - Cloth and person image. p* is obtained from person image to randomly select different cloth images to synthesize fake image UI as tutor. Tutor and clothes image is fed as inputs to train PF-AFN to generate student SI to be directly supervised by real image.

This design proposes the Parser Free Appearance Flow Network (PF-AFN), a novel "teacher-tutor-student" knowledge distillation system that uses a novel "teacher-tutor-student" knowledge distillation technique to produce highly photorealistic try-on images without human parsing. This further formulates the try-on problem's knowledge distillation as distilling the appearance flows between the human picture and the garments image. With the "teacher-tutor-student" knowledge distillation, we first clarify the entire training strategy. We employ a generative module (GM) to generate the try-on image and an appearance flow warping module (AFWM) to construct accurate dense correspondences between the person image and the garments image. Finally, we'll go over how we distil appearance flows to produce high-quality photos.

This work consists three main contributions:

1. A strategy of teacher-tutor-student has been presented for generating photo realistic results without the use of human segmentation.

2. They used the try-on problem to define knowledge distillation, which is necessary for finding accurate dense correspondences between pixels in order to make high-quality images.

3. Experiments on common datasets show that their strategy is qualitatively and quantitatively superior to recent state-of-the-art approaches.

2.3 Final design: PF-AFN Approach

Unlike previous work, this methodology uses the fake photos generated by the parser-based method as "tutor knowledge," with the artifacts being repaired by genuine "teacher knowledge" gathered in a self-supervised manner from real person images. This algorithm formulates knowledge distillation in the try-on problem as distilling the appearance flows between the human image and the clothing image, allowing us to uncover accurate dense correspondences between them and generate high-quality outcomes. Extensive testing has revealed that this strategy is vastly superior.

Hence 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

Detailed Explanation of VITON:Resolution: 256x192

- Total image pairs: 19000 (front view woman and top-clothing image pairs).
- Cleaned pairs : 16253.
- Training set : 14221.
- Validation set : 2032.

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

3.2 Training

On all three datasets, the training approach is the same. The clothes picture and the image of the person wearing the clothes are the first things we teach PB-AFN. In this phase, the parser findings and human pose predictions are also used. With an initial learning rate of 3 *105, PB-AFN is optimised for 200 epochs with l = 1.0, p = 0.2, and sec = 6.0. The training schedule for PF-AFN is the same as for PB-AFN, and the hyper-parameters are set to hint = 0.04 and pred = 1.0.

3.3 Algorithm

Algorithm 1 PB-AFN: This algorithm masks the human picture's clothing region and reconstructs the image with the relevant in-shop apparel and person representations.

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 w.r.t appearance flows predicted
6. Synthesise images with person and warped cloth
7. calculate losses
8. update modules
9. **end for**

Algorithm 2 PF-AFN: Generated fake images is used as the input to PF-AFN under the supervision of the real images then distill the appearance flows to get the accurate correspondences between the target and the person clothes PF-AFN

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 PF-FAN Approach

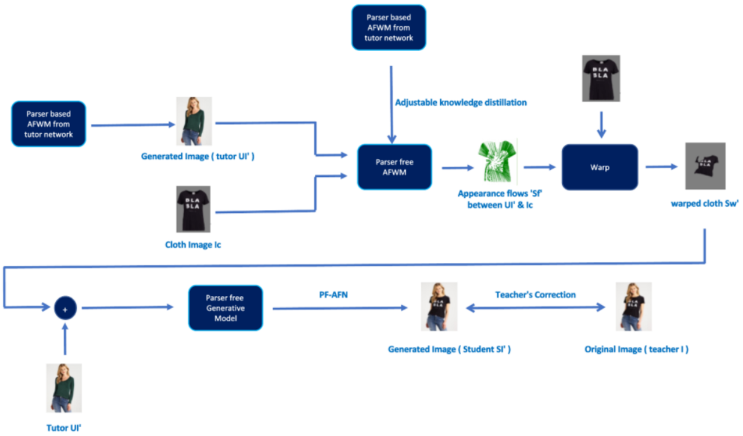

Figure 3.1: Parser-Based Appearance Flow Network: Takes person representation and cloth image as input and gives the warped module and generative module as the outputs.

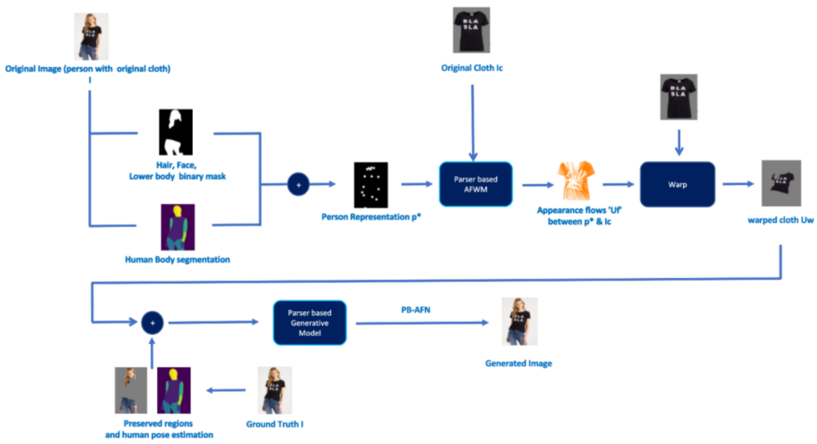

Figure 3.2: Parser-Free Appearance Flow Network: Takes the knowledge from the PBAFN Tutor network and gives the warped module of the PF-AFN as outputs.

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 fig:4.1
2. simple Pose-Simple Cloth result in Figure 4.2.
3. simple Pose-Simple Cloth result 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 tried on custom images and can be seen in fig:4.6 and fig:4.7. for controlled environment.
7. The architectures were also tried on custom images and can be seen in fig:4.8 for uncontrolled environment.

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.

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

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.

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

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

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

Figure 4.6: PFAFN and ACGPN results for custom image results for T-shirt 1 in controlled environment.

Figure 4.7: PFAFN and ACGPN results for custom image results for T-shirt 2 in controlled environment.

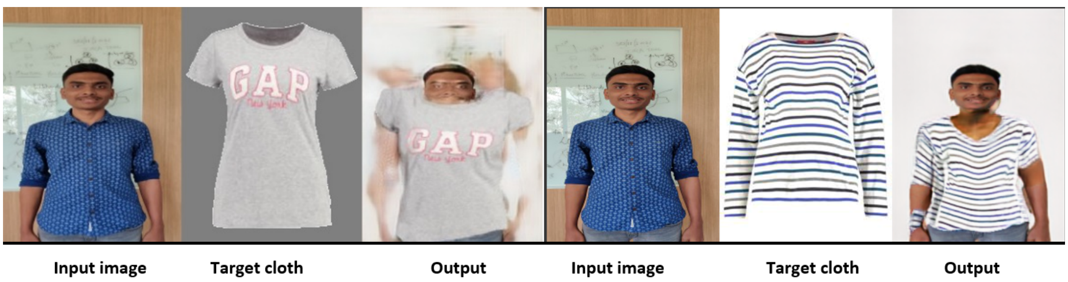

Figure 4.8: PFAFN and ACGPN results for custom image results in an uncontrolled environment.

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.

- Referring fig:4.5 the distortions like fine finger details can be overcome by improving the loss function.
- From all the figures seen above of the comparison between all the three algorithms PF-AFN generates better results than CPVTON, CP-VTON+ and ACGPN.
- CP-VTON+ also generates good results but has challenges for complex clothes and poses and it can seen in fig:4.4 and fig:4.5.
- From fig:4.8 we can conclude that images should be taken in a proper controlled environment with 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 customers 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. When we talk about shopping for clothes, it is one of the most enjoyable experiences for a customer, but at the same time can become cumbersome and tiring. The usual hassles involved in traditional in-store shopping experiences like standing in long queues or having to dress-undress again and again can make the whole experience a little dull. With virtual try-on, all these hassles are done away with.

With virtual try-on, a customer can try different clothing items without having to go through any of the above-mentioned hassles, right from the comfort of their homes.

[1] [2] [3]

Bibliography

[1] Yuying Ge, Yibing Song, Ruimao Zhang, Chongjian Ge, Wei Liu, and Ping Luo. Parser-free virtual try-on via distilling appearance flows. *arXiv preprint arXiv:2103.04559*, 2021.

[2] Matiur Rahman Minar, Thai Thanh Tuan, Heejune Ahn, Paul Rosin, and Yu-Kun Lai. Cp-vton+: Clothing shape and texture preserving image-based virtual try-on. In *CVPR Workshops*, 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 content. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7850–7859, 2020.

Virtual Try-on

ORIGINALITY REPORT

15%	6%	10%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Yuying Ge, Yibing Song, Ruimao Zhang, Chongjian Ge, Wei Liu, Ping Luo. "Parser-Free Virtual Try-on via Distilling Appearance Flows", 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021 Publication	7%
2	arxiv.org Internet Source	3%
3	minar09.github.io Internet Source	1%
4	divante.com Internet Source	1%
5	Submitted to University of Sydney Student Paper	1%
6	Han Yang, Ruimao Zhang, Xiaobao Guo, Wei Liu, Wangmeng Zuo, Ping Luo. "Towards Photo-Realistic Virtual Try-On by Adaptively Generating↔Preserving Image Content", 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020 Publication	1%

7	lup.lub.lu.se Internet Source		1 %
8	Submitted to Istituto Marangoni LTD Paris Campus Student Paper		1 %
9	Submitted to B.V. B College of Engineering and Technology, Hubli Student Paper		1 %
10	Submitted to University of Glasgow Student Paper		<1 %
11	Kathleen M Lewis, Srivatsan Varadharajan, Ira Kemelmacher-Shlizerman. "TryOnGAN", ACM Transactions on Graphics, 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	On	Exclude matches	Off
Exclude bibliography	On		