Available online @ https://jjem.jnnce.ac.in
https:www.doi.org/10.37314/JJEM.2023.070105
Indexed in International Scientific Indiexing (ISI)
Impact factor: 1.395 for 2021-22
Published on: 31 July 2023

Preserving Network and Adaptive Content Generating Method for Image Based Virtual Try-On Clothing System

Girish Mantha[1], Praveen G Shet[2], Rahul Athreya K M[3], Sathwik P Bhat[4], Shamanth G P[5]

[1]Asst professor, ISE Dept, JNNCE Shivamogga, [2-5]Student, ISE Dept, JNNCE Shivamogga

girish.mantha@jnnce.ac.in, praveengbharadvaj@gmail.com, rahulathreya312@gmail.com, sathwikbhatp@gmail.com, shamanthgp6@gmail.com

Abstract

The Virtual cloth Try-on is one of the biggest inventions took place in fashion industry which contributes to enhance user experience by allowing them to try out garments virtually without wearing it. The Virtual Try on Cloth is image-based technology to enhance the user experience on fashion-oriented e commerce websites, it will help customer's satisfaction. To get perfect body fit cloth or exact cloth fitting on body through Imaged based -virtual try on. This project suggests a better solution using Artificial Intelligence (AI) and Augmented Reality (AR) for non-tech-savvy customers who aims at transferring a target clothing image onto himself. Creating realistic try-on images is still a significant obstacle, especially when dealing with extensive occlusions and complex human poses in the reference person. The algorithm determines whether to generate or preserve the image content based on the anticipated semantic layout. This approach results in highly realistic try-on images and captures intricate clothing details. This project idea is applying some steps below to get exact results. Beginning, based on the initial pose of the given person our model adjusts the target clothing form to compatible with the given pose. After this next task is to come up with the body segmentation model of the person wearing the required clothing, to better understand the body parts and clothing regions. Finally, the body segmentation map is fused together with warped clothing and a given person image for fine-scale image synthesis. The body segmentation map prediction using CNN, helps to guide image synthesis where body part and clothing intersects and it's useful to preserving clothing and body part details.

Keywords: *Augmented Reality (AR), Artificial Reality (AI), Virtual Try-On (VTON), Photo-Realistic Image*

1. Introduction

Image based Virtual Try-On (VTON) clothing system has emerged as a popular research area, focusing on the task of transferring a desired clothing image onto a reference person. While previous studies have primarily concentrated on preserving the essential characteristics of the clothing (such as texture, logo, and embroidery) during the warping process to accommodate various human poses, generating photo-realistic try-on images becomes a significant hurdle when dealing with substantial occlusions and complex human poses in the reference person. This research comprises three primary modules. The first module involves the generation of a semantic layout, which progressively predicts the desired layout after clothing try-on by utilizing semantic segmentation of the reference image. The second module focuses on cloth warping, where clothing images are warped based on the generated semantic layout. To enhance stability during training, a second-order difference constraint is introduced. The third module is an inpainting module that integrates all available information, including the reference image, semantic layout, and warped clothes, to dynamically produce each semantic part of the human body. Compared to existing methods, this approach demonstrates superior capability in generating highly realistic images with enhanced percep-

tual quality and more intricate fine details. Inspired by the advancements in image synthesis [1]-[3], [9], image-based visual try-on [10][11][12][13], which focuses on transferring clothing items onto a reference person, has gained significant interest in recent years. Despite notable advancements [3], constructing a photo-realistic virtual try-on system for real-world scenarios continues to pose challenges. This can be attributed, in part, to the semantic and geometric disparities between the target clothes and reference images, as well as the occlusions that occur due to interactions between the torso and limbs. Fig 1

Figure 1: Virtual Try-on Sample Images

The proposed system will facilitate to Customers can now virtually try on various types of products before purchasing them order multiple variations of a single product. What's important here is that AR helps customers avoid disappointment and choose the best products for them. As a result, both online and brick-and-mortar store return rates tend to fall. Fig 2

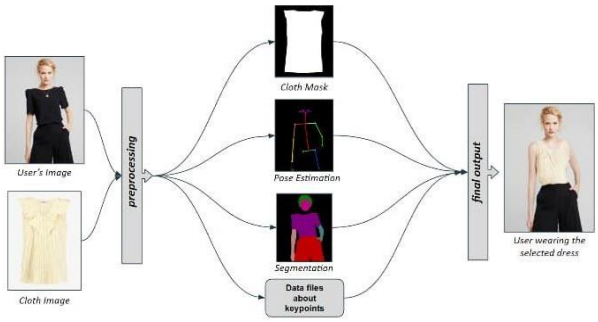

Figure 2. Overview of the method

1.1 Problem statement

Since there are problems in the field of clothing industry because of returning of the products due to mismatch of specific looks and fitting in clothing. This project is targeted to solve the designing and construction of 2D cloth model of the target cloth by matching silhouettes to the standard body model. The proposed method will transfer the cloth model to the estimated target human model, to produce the warped cloth. Generate target human skin parts and blend it to the rendered warped cloth along with human representations. This model helps in reducing return orders and for delivering products according to the customer poses and specification.

1.2 Objectives

Objectives of the study include;

a.To collect database of virtual fitted images of users and clothes. VTON dataset will be used from which 80% as training, 10% for testing and 10% for validation.

b.To collect and match the desired clothes to the user by importing the images to cGAN, re-sizing and pose mapping.

c.To determine whether the cloth is right fit in different poses by enhancing the pose map image using Pix2pixHD methods.

d.To try clothing without getting delivered to the customer by displaying augmented image grid.

2. Related Work

The following are the different works carried out in the given area:

Amit Raj et al. [4] discussed a system neural network architecture that tackles these sub-problems with two task-specific sub-networks. Since acquiring pairs of images showing the same clothing on different bodies is difficult, Authors proposed a novel weakly supervised approach that generates training pairs from a single image via data augmentation. They pre-

sented the first fully automatic method for garment transfer in unrestricted images without solving the difficult 3D reconstruction problem. They demonstrated a variety of transfer results and highlight their methods advantages over traditional image-to-image and analogy pipelines.

Xintong Han et al. [5] introduced a Virtual Try-On Network (VITON) that achieves seamless transfer of desired clothing items onto corresponding regions of a person. Unlike previous approaches, this network does not rely on any form of 2D information. Instead, it adopts a coarse-to-fine strategy. The framework utilizes a novel clothing-agnostic yet descriptive representation of the person. Initially, a coarse synthesized image is generated, where the target clothing item is overlaid on the person in the same pose. Subsequently, a refinement network enhances the initial blurry clothing area. The network is trained to determine the appropriate level of detail from the target clothing item and where to apply it on the person, resulting in a photo-realistic image where the target item naturally deforms with clear visual patterns. Experimental results on the newly collected Zalando dataset highlight the effectiveness of this approach in the image-based virtual try-on task, surpassing state-of-the-art generative models.

Haoye Dong et al. [6] presented a multi-pose guided virtual try-on system, which facilitates the transfer of clothes onto a person image with various poses. Their proposed approach, the Multi-pose Guided Virtual Try-on Network (MG-VTON) [11], takes as input a person image, a desired clothes image, and a desired pose. Utilizing these inputs, the network can generate a new person image by seamlessly fitting the desired clothes into the input image while manipulating human poses.

Yuying Ge Chen [7] introduced a technique called StarGAN, which presents a unique and scalable approach for conducting image-to-image translations across multiple domains using a single model. The unified architecture of StarGAN enables the concurrent training of diverse datasets with distinct domains within a single network. As a result, StarGAN exhibits remarkable quality in generating translated images, surpassing existing models. Additionally, it introduces a novel capability of flexibly translating an input image to any desired target domain. The effectiveness of this approach is empirically demonstrated through experiments involving facial attribute transfer and facial expression synthesis tasks.

In [8], Nilesh Pandey presented an approach called Poly-Gann Multi-Conditioned GAN for Fashion Synthesis. This technique, Poly-GAN, allows for conditioning on multiple inputs and is suitable for various tasks such as image alignment, image stitching, and inpainting. The architecture of Poly-GAN ensures that the conditions are incorporated at all layers of the encoder, and it utilizes skip connections between the coarse layers of the encoder and corresponding layers of the decoder. With Poly-GAN, it becomes possible to perform spatial transformations of garments based on the RGB skeleton of the model, even in arbitrary poses. Furthermore, Poly-GAN can handle image stitching irrespective of the garment orientation and perform inpainting on garment masks that contain irregular holes.

2.1 Review of papers

Following is the summary of papers.

Table 2.1: summary of review paper title and Methodology used.

Title of Paper	Methodology
“SwapNet: Image Based Garment Transfer”	SwapNet, A Neural Network Architecture
VITON: An Image base Virtual Try-on Network”	VITON, CRN, CAGAN using coarse-to-fine strategy

"Towards Multi-pose Guided Virtual Try-on Network"	GAN, MG-VTON by manipulaing human poses.
"StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation"	Conditional GAN by concurrent training of diverse datasets
"Poly-GAN: Multi-Conditioned GAN	Poly-GAN by image alignment, image stitching, and inpainting

3. System Design and Implementation

A multistage method to synthesize the image of person conditioned on both clothes and pose is proposed. Given an image of a person, a desired cloth, and a desired pose, Method will generate a realistic image that preserves the appearance of both desired clothes and perso as shown in fig 3.

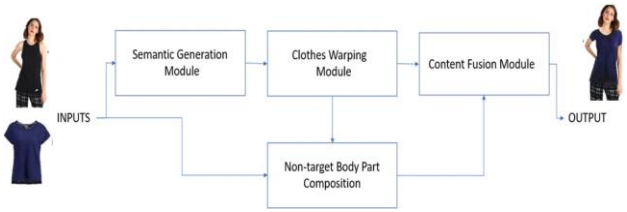

Figure 3: Block diagram of proposed methodology

3.1 Architecture

The proposed method comprises four modules, which are described below:

- '*Semantic Generation Module (SGM)':*The SGM is designed to separate the target clothing region while preserving the body parts of the person.

- '*Clothes Warping Module (CWM)':* The CWM focuses on fitting the clothes into the shape of the target clothing region, ensuring visually natural deformation according to the human pose. It also aims to retain the original characteristics of the clothes.

- Non-target Body Part Composition: This module effectively preserves the non-target body parts by combining the masks generated by the previous two modules.

- '*Content fusion module (CFM)':* The CFM consists of two main steps. Step 1 aims to preserve the untargeted body parts while adaptively maintaining the changeable body part, such as the arms. Step 2 utilizes the masks and images generated from the previous steps to fill in the changeable body part, employing an inpainting-based fusion GAN.

3.2 Methods and Technologies

Experiments were performed on the VITON dataset, as depicted in Figure 4 and 5. This dataset consists of approximately 19,000 image pairs, comprising front-view woman images and corresponding top clothing images. After filtering out any invalid image pairs, a total of 16,253 pairs remained. These pairs were then split into a training set of 14,221 pairs and a testing set of 2,032 pairs. For comparison, our approach was evaluated against VITON, CP-VTON, and VTNFP. Although the official code for VTNFP was not available, we recreated the visual results described in their paper to facilitate qualitative comparison.

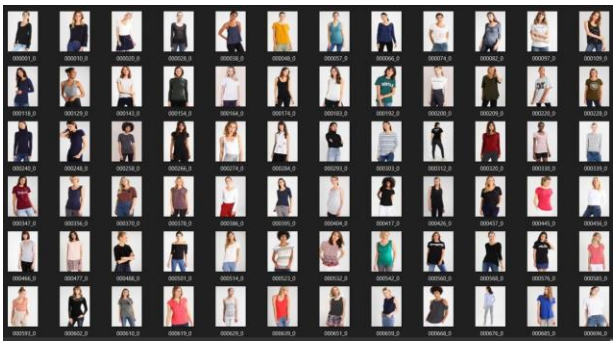

Figure.4: Cloth Image

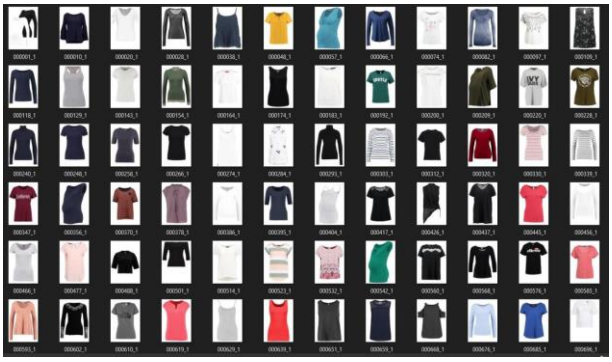

Figure 5: Sample of user images

3.3 Algorithm

The proposed method is implemented using the following pseudocode: Image based VTON

Step 1. Importing data-set which has user images and cloth images
Step 2. Importing models U-Net, human parsing which contains cGAN etc. Step 3. Input Images by using files.upload()
- Cloth image
- User image

Step 4. Segmentation of cloth image using U-Net
Step 5. Self-correction human parsing
- User image resizing
- Parsing the image using cGAN
- Generating pose map

Step 6. Enhance the output image using pix2pixHD method.
Step 7. Display Augmented image in grid.
- User image, cloth image, final output

4. Results

Results are recorded based on the poses of input images.

4.1 Easy

Test cases in easy level are mentioned in Fig 6 to 8. These figures have normal poses that are facing frontward and it contains mostly plain colored clothes this method works well on both half sleeve and full sleeve dresses.

4.2 Medium

Test cases in medium level are mentioned from Fig 9 to11. These figures have slightly different poses that are facing in some other direction, and it contains shirts having plain coloured clothes as well as printed shirts. This method works well on both half sleeve and full sleeve dresses.

Figure 6: Snapshot of easy level test case(1)

Figure 7: Snapshot of easy level test case(2)

Figure 8: Snapshot of easy level test case(3)

Figure 9: Snapshot of Medium level test case(1)

Figure 10: Snapshot of Medium level test case(2)

Figure 11: Snapshot of Medium level test case(3)

4.3 Hard

Test cases in Hard level are mentioned from Fig 12 to Fig 14. These figures have unorthodox poses that may have crossed hands or folded hands etc. They may be facing in some other direction as well and it contains shirts having plain colored clothes, printed cloths and complex patterns. The method works well on both half sleeve and full sleeve dress.

Figure 12: Snapshot of Hard level test case(1)

Figure 13: Snapshot of Hard level test case(2)

Figure 14: Snapshot of Hard level test case(3)

4.4 Failure test cases

- Large transformation of the semantic layout is hard to handle, partly ascribing to the agnostic input of fused segmentation.
- The shape of the original clues is not completely removed.
- Very difficult pose is hard to handle. Better solution could be proposed.

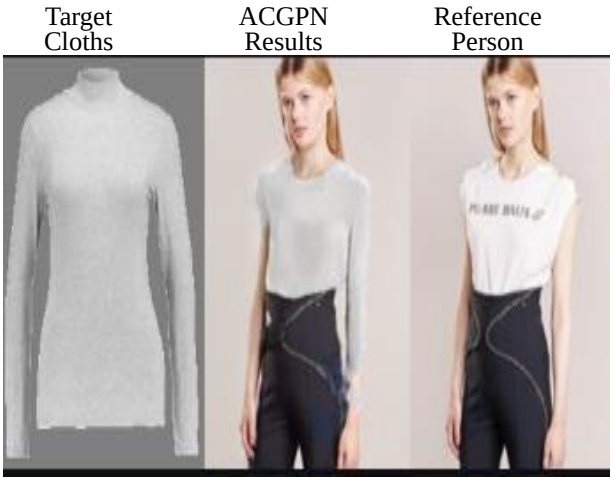

Figure 15: Failure case showing large semantic layout

In fig 15 Large transformation of the semantic layout is hard to handle, partly ascribing to the agnostic input of fused segmentation.

Figure 16: Failure case showing original clues not completely removed

In the fig 16, the shape of the original clues is not completely removed so this results in failure.

Figure 17: Failure case due to difficult pose

In fig 17, a very difficult pose is hard to handle. Better solution could be proposed.

5. Conclusion

There exist many challenges in the field of clothing industry such as mismatch of specific looks and fitting in clothing which leads to the lack of growth in E-commerce. The aim of this research is to propose a way to construct a software that will collect the images of users and targeted cloths and process it with the help of artificial intelligence and augmented reality by using semantic generation to separate the target clothing region as well as to preserve the body parts of the person without changing the pose and cloths wrapping to fit the clothes into the shape of target clothing region then fusing the above-mentioned methods. This application will prove to be portable and easy to use.

It will be much helpful for the people who isn't tech-savvy.

References

1. Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In CVPR, pages 5967–5976. IEEE Computer Society, 2017.
2. Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In CVPR, pages 2337–2346. Computer Vision Foundation / IEEE, 2019.
3. Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. In ICLR. OpenReview.net, 2018.
4. Amit Raj, Patsorn Sangkloy, Huiwen Chang, James Hays, Duygu Ceylan, and Jingwan Lu. "Swapnet: Image based garment transfer". In ECCV (12), volume 11216 of Lecture Notes in Computer Science, pages 679–695. Springer, 2018.
5. Xintong Han, Zuxuan Wu, Weilin Huang, Matthew R Scott and Larry S Davis. "Finet: Compatible and diverse fashion image inpainting". In Proceedings of the IEEE International Conference on Computer Vision, pages 4481–4491, 2019.
6. Haoye Dong, Xiaodan Liang, Bochao Wang, Hanjiang Lai, Jia Zhu, and Jian Yin. "Towards multi-pose guided virtual try-on network". CoRR, abs/1902.11026, 2019.
7. Yunjey Choi, Min-Je Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. "Stargan: Unified genernative adversarial networks for multi-domain image-to-image translation". In CVPR, pages 8789–8797. IEEE Computer Society, 2018.
8. Pandey, N., & Savakis, A. (2020). "Poly-GAN: Multi-Conditioned GAN for Fashion Synthesis". Neurocomputing Elsevier,volume 414, 13 November 2020, Pages 356-364.
9. Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In CVPR, pages 4401–4410. Computer Vision Foundation / IEEE, 2019.
10. Han Yang1,2 Ruimao Zhang2 Xiaobao Guo2 Wei Liu3 Wangmeng Zuo1 Ping Luo4 1Harbin Institute of Technology 2SenseTime Research 3Tencent AI Lab 4The University of Hong Kong, "Towards Photo-Realistic Virtual Try-On by Adaptively Generating↔Preserving Image Content, In CPVR 2020, pages 7850-7859. IEEE Explore, 2020.
11. X. Han, Z. Wu, Z. Wu, R. Yu, and L. S. Davis. "Viton: An image-based virtual try-on network". In CVPR, 2018.
12. G. Balakrishnan, A. Zhao, A. V. Dalca, F. Durand, and J. Guttag. Synthesizing images of humans in unseen poses. In CVPR, 2018.
13. Ruiyun Yu, Xiaoqi Wang, and Xiaohui Xie. Vtnfp: "An image-based virtual try-on network with body and clothing feature preservation". In The IEEE International Conference on Computer Vision (ICCV), October 2019.