Journal of Data Acquisition and Processing

ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.754975

ADVANCING VIRTUAL TOP TRY-ON: INTEGRATING ACGPN, U-2 NET ARCHITECTURE, AND HUMAN PARSING

Komala K V [a*], Deepa V P [b], Veena Yadav S [c]

[a.] Assistant Professor, Department of Computer Science and Engineering, Government Engineering College,
Ramanagara – 562 159, Karnataka, India.

[b] Assistant Professor, Department of Electronics and Communications Engineering, Government Engineering College,Ramanagara – 562 159, Karnataka, India.

[c] Assistant Professor, Department of Electronics and Communications Engineering, Government Engineering College, Chamarajanagara –571 313 Karnataka, India.

* Corresponding author: komala K V

Abstract—Virtual top try-on systems have revolutionized online shopping, allowing consumers to visualize clothing items before purchasing. This paper presents a novel approach to virtual top try-on, integrating attribute-controlled and Geometry-Preserving GAN (ACGPN), U-2 Net architecture, and human parsing. ACGPN facilitates the generation of realistic top simulations with customizable attributes, while the U-2 Net architecture enhances segmentation accuracy for precise garment placement. Human parsing ensures accurate detection and segmentation of body parts, optimizing virtual garment fitting. Experimental results demonstrate the effectiveness of the proposed approach in achieving lifelike virtual try-on experiences. The model is verified on the VITON dataset and the Custom dataset.

Keywords: Virtual top try-on, Attribute-Controlled and Geometry-Preserving GAN (ACGPN), U-2 Net architecture, Human parsing.

1. INTRODUCTION

Virtual try-on technology is becoming increasingly important in the age of Internet shopping, which has led to much research being done in this area. Virtual try-on solutions based on photos that function without specific hardware or imaging equipment using standard-intensity images are very appealing. This makes it easier to create virtual fitting rooms, which let customers try on clothing from a distance and eliminate the need for in-person store visits. Moreover, it benefits retailers by lowering product returns and delivery expenses. Most current picture-based virtual try-on techniques involve two stages: (i) Geometric matching, which involves lining up the target garment with the subject's pose in the input image and projecting the garment's approximate location and shape in the final try-on result. (ii) Image synthesis creates the final try-on image using generative models such as Generative Adversarial Networks (GANs) and combining several refining techniques and auxiliary data sources, including clothing annotations, pose vital points, and parsed body parts. A few improvements have been suggested to improve this framework, such as: (i) Better data sent to the geometric matching stage. (ii) Segments for clothing should be included in the synthesis process. (iii) Using knowledge distillation techniques to reduce errors caused by parsers. Even while these developments have greatly improved the quality of the generated try-on results, there are still issues, most notably the loss of garment details due to problems with the geometric matching step. Furthermore, poor human/clothing parsing frequently produces convincing try-on photographs when body portions are mistakenly covered in garment texturing. We suggest a Context-Driven Virtual Tryon Network (C-VTON) to address these problems. To minimize pre-processing stages and improve the quality of created garment textures and logos, we provide a unique geometric matching module that conditions the pose-matching process only on body segmentations. In addition, we design training objectives for the module that penalize the appearance of aligned clothing exclusively inside the body region, guaranteeing strong performance even in the face of difficult pose configurations and self-occlusions.

Minaee, Shervin et al. [1] proposed a human pose, hand, and mesh estimation survey. The difficult field of human position estimation in computer vision involves locating body key points in space from pictures or movies. The paper emphasizes the localization of anatomical key points and discusses the advancements in significant models for 2D and 3D human posture estimation. Prospective avenues for research encompass tackling obstacles such as blockage from garments or other persons, inter-individual interactions, limitations imposed by human anatomy, and the identification of hardly perceptible joints. Pang presents a virtual try-on method based on 3D matching, Shanchen et al. [2]. It is comparable to beauty effects, in which virtual images simulate face features without learning their relative placements. The method converts virtual outfit changes based on photos into accurate fitting results. Munea, Tewodros Legesse et al. [3] employed image segmentation, a key difficulty in computer vision. It entails classifying pixels into instance segmentation (identifying specific objects), semantic labels (semantic segmentation), or both (panoptic segmentation). Meanwhile, instance segmentation further distinguishes each unique object inside the image, such as individual people. Semantic segmentation labels pixels based on predetermined categories like human, car, tree, or sky. Han, Xintong et al. [4] presented V¨ ITON employing a cascaded appearance flow estimation network. Cloth Flow can efficiently handle geometric deformations and occlusions while synthesizing target pictures from source images and poses. Even though handling non-rigid objects. An Image-based Virtual Try-on Network in 2018 addresses the present difficulties of cloth Flows. encouraging outcomes of the growing demand for online fashion shopping and the need for consumers to visualize how clothing items would look on them before purchasing. Their work employs the Inception Score (IS) to evaluate image synthesis quality, focusing on visually diverse and semantically meaningful results while emphasizing the importance of preserving clothing details and body pose in synthesized images. Despite advancements in 3D modeling techniques for realistic clothing simulations, challenges remain in hardware installation costs and data collection for large-scale deployment, highlighting avenues for future research in refinement networks and dataset enhancements. Dabolina, I et al. [5] evaluated clothing fit in their 2018 publication, emphasizing the development of textile materials with smart applications and the importance of functional clothing design for various environments. Their study utilized virtual prototyping to assess ease, fit, and appearance, highlighting differences in fit among test subjects and the significance of garment functionality. While gaps between jackets and the body indicate spatial ease for movement, the research aims to streamline garment size selection for specific target groups in the apparel industry, supporting Sustainable Development Goals.Yang, Han et al. [6] presented that the ACGPN approach attempts to provide realistic try-on results while keeping clothing traits and aspects of human identity, thereby addressing the difficulties in developing a photorealistic virtual try-on system. Although the suggested network improves try-on quality regarding semantic alignment, character retention, and layout adaptability, it might not be able to recreate fine details, which opens new research opportunities through user surveys. Ren, Bin et al. [7] introduced Virtual try-on (VTON) derived from fashion editing. It aims to seamlessly overlay clients' chosen in-store apparel onto their body photos, providing a quicker and more efficient purchasing experience. Because of its convenience, this technology has been useful in e-commerce platforms and large-brand clothing retailers. Raffiee, Amir Hossein [8] tackles the problem of providing online clothing customers with virtual try-on experiences. Their Garment GAN model efficiently transfers clothing properties and synthesizes high-quality photos, overcoming complicated situations, including occlusions, body posture, and garment shape. The approach, which uses adversarial training techniques and innovative network architecture, presents a viable way to improve the ease of online buying. Han, Xintong et al. [9] designed Cloth Flow: A Flow-Based Model for Clothed Person Generation to address pose-guided person generation, an important aspect of fashion in which it is desirable to transfer clothing between individuals. By encouraging more research into computer vision for better geometric capture. Hsieh, Chia-Wei et al. [10] presented F¨ IT-ME: Image-Based Virtual Try-On with Arbitrary Poses in response to the problem of online fashion buying, where customers cannot sample products before buying. Fit Me is their technique that converts user poses to target positions and warps in-store clothing to create virtual try-on images with arbitrary poses. Creating virtual try-on photos with random poses is still difficult, even though this challenge needs conditional GANs for realistic image generation. This suggests potential future research directions, such as using human segmentation information for better outcomes.

2. PROPOSED METHODOLOGY

As seen in the Fig. 1, the ACGPN comprises three modules. First, the Semantic Generation Module (SGM) generates body parts and distorted clothing masks by semantic segmentation, which guarantees spatial alignment. Secondly, a second-order difference constraint on thin-plate spline (TPS) is incorporated for geometric matching and character retention when the Clothes Warping Module (CWM) warps target clothing images based on the clothing mask. Lastly, to determine the creation or preservation of discrete human parts in the synthesized output image, the Content Fusion Module (CFM) combines data from the preceding modules. This enables flexible handling of non-target body parts and effective layout adaptation for pictures with different levels of difficulty. The model has four primary modules: i) Semantic Generation, ii) Warping of Clothes, iii) Composition of non-target bodily parts, iv) The Fusion Unit.

Semantic Generation

Computer vision tasks such as semantic segmentation require classifying each pixel in an image into an object or class. In this case, the goal of the Semantic Generation Module (SGM) is to precisely identify the target clothing region while maintaining the body components (e.g., arms) and avoiding modifying the stance or other human body characteristics. The SGM incorporates a mask generation mechanism to accurately segment body parts and the target clothing region based on input data, including the target clothing image, mask, pose information, segmentation map of body parts, and identified clothing items. This contrasts with previous approaches focusing only on target clothes and ignoring human body generation.

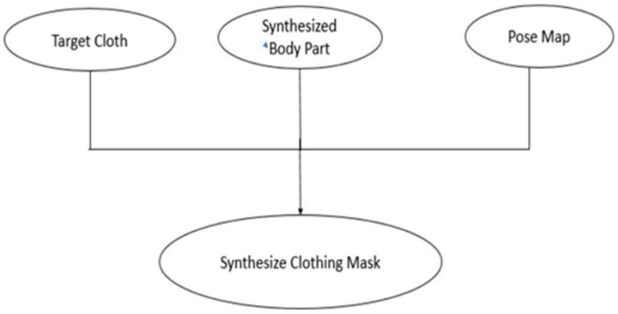

Fig. 1. Flow Chart of Semantic Generation Module.

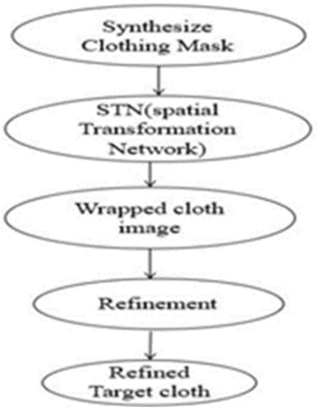

Fig. 2. Flow Chart of Clothes Warping Module.

Clothes Warping Module

Clothes must be warped to fit clothing to the target clothing region while maintaining their beauty and guaranteeing natural distortion by human stance. However, training a Spatial Transformation Network (STN) by itself could provide inaccurate transformations, especially for intricate textures and vibrant colors, which could cause misalignment and hazy results. To address these issues and enhance the quality of the synthesized images, we accomplish accurate geometric matching and character retention by adding a second-order difference constraint to the garments warping network.

Non-target body part Composition

The final result is produced in the third generative module by integrating the person’s image, the changed segmentation map, and the warped clothing image. Non-target body portions are precisely retained by combining masks MSω and Mω, guaranteeing the coherence and detail retention directed by MGα. Furthermore, the system exhibits versatility in delivering coherent findings across various contexts.

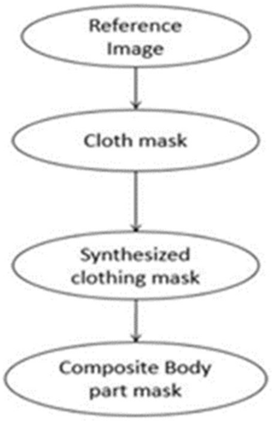

Fig. 3. Flow Chart of Nontarget body part Composition.

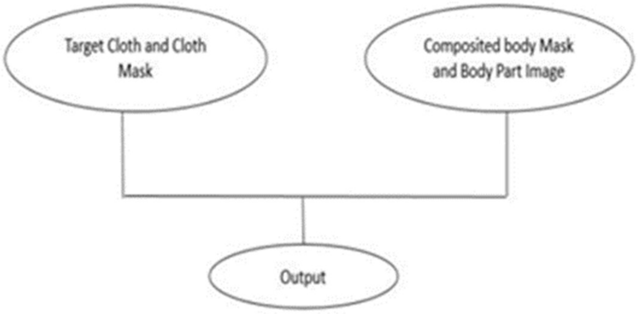
Target Cloth and Cloth Mask
Composited body Mask and Body Part Image
Output

Fig. 4. Flow Chart of Content Fusion Module.

Content Fusion Module

Beyond semantic alignment and character retention, a significant obstacle in visual try-on jobs is still accomplishing layout adaptation. To do this, the target garment region must be rendered clearly, and fine-scale body component characteristics, such as finger gaps, must be preserved adaptively. By entirely keeping untargeted body parts and adaptively conserving variable body components, such as arms, the proposed content fusion module (CFM) overcomes this issue, in contrast to previous approaches that frequently fail to recreate fine details and rely on coarse body outlines. Step 3 guarantees the preservation of non-targeted body parts and adaptive retention of changing body parts. In contrast, Step 4 fills in the changeable body part using masks and pictures from earlier steps in an inpainting-based fusion GAN.

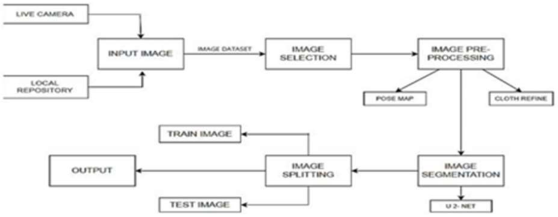

Fig. 5. The High-Level System Design.

SYSTEM DESIGN

Fig. 5 shows the high-level system design. Choosing an image from local storage, a live camera, or an image dataset with frontal ladies and top photos is the first step in the procedure. The dataset is downloaded from a dataset source and is usually in the.jpg or.png format. After choosing a picture, preprocessing is done to fine-tune the fabric and pose map to obtain the appropriate cloth image. The you 2-net architecture, a neural network renowned for its accuracy in image segmentation, is then used to carry out picture segmentation. The preprocessed data is subsequently split into train and test sets for evaluating and predicting the model. Lastly, the output recommends or fits the relevant top to a woman's matching input frontal photograph.

Data Flow Diagram

The data flow and procedures involved in the virtual try-on system are usually depicted in a data flow diagram (DFD) for Virtual Top Try-On in Fig. 6. The following is a synopsis of its constituent parts:

Sources of Input: These are the places where the system gets its input data. Images from local storage, live camera feeds,

Image Selection Process: Local storage, a live camera, or an image dataset are used to choose the input images. This process includes selecting the top and the frontal woman image's input image.

Preprocessing: The input photos go through preprocessing after they are chosen. This phase entails fine-tuning the fabric and position map to achieve the ideal cloth image.

Image Segmentation: Next, a segmentation technique, like the you 2-net architecture, is applied to the pre-processed images. By giving each pixel a label, this technique creates different sections inside the image while providing comprehensive information about the image.

Train and Test Data Splitting: The pre-processed data is divided into train and test sets following segmentation. The test set is used to forecast the model’s performance, and the training set is used to evaluate the model.

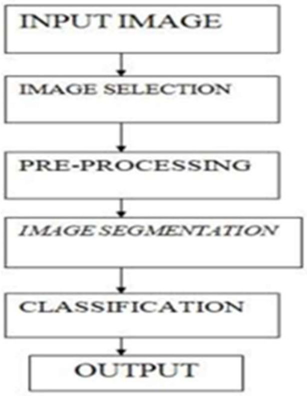

Fig. 6. Data Flow Diagram of the Proposed System.

Model Evaluation and Prediction: Using the train set, the system assesses the model and uses the test set to forecast its performance. Usually, this calls for the use of segmentation methods like U-Net.

Output: The system produces output by recommending or fitting the proper top to the matching input.

Use Case Diagram

The system's functionality and the actors' roles are highlighted in a use case diagram for Virtual Top Try-On, which shows the different interactions between users, or actors, and the system.

In general, the use case diagram in Fig. 7 helps to identify critical system requirements and features by offering a high-level overview of the interactions between actors and the system’s functionality in the context of Virtual Top Try-On.

3. IMPLEMENTATION

DATASETS

The VITON dataset serves as a benchmark for research in virtual try-on systems. It consists of paired images: one depicting a person in casual clothing and another with the person wearing a virtual top. The dataset encompasses various poses, body shapes, and clothing styles to facilitate diverse experiments. Each image pair is meticulously annotated with pixel-level clothing masks and poses vital points, enabling accurate evaluation of virtual try-on algorithms. Researchers often utilize the VITON dataset to train and test their models, aiming to improve the realism and accuracy of virtual clothing try-on applications.

The dataset contains 19,000 image pairs, including a front-view woman image and a top clothing image. Removing the invalid image pairs yields 16,253 pairs, further splitting into a training set of 14,221 pairs and a testing set of 2,032 pairs.

Virtual Top Try On MODEL

Virtual top try-on, powered by innovative technologies such as Attribute-Controlled and Geometry-Preserving GAN (ACGPN), U-2 Net architecture, and human parsing, revolutionizes online shopping. ACGPN ensures lifelike top simulations with customizable attributes, while the U-2 Net architecture enhances segmentation accuracy, optimizing garment placement. Human parsing accurately detects and segments body parts, facilitating precise virtual garment fitting and collectively improving the virtual try-on experience with realism and accuracy.

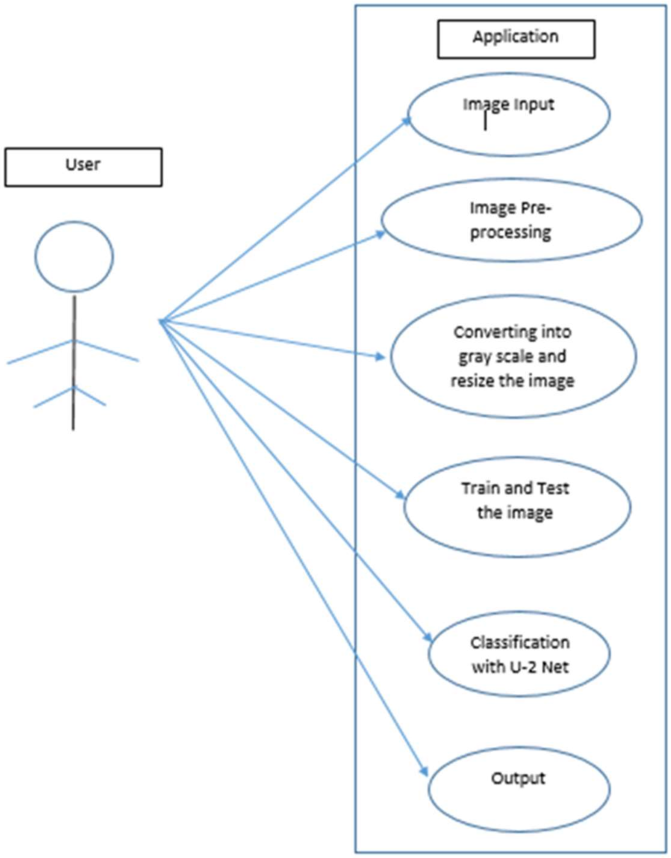

Fig. 7. Use a Case Diagram of the Proposed System.

ACGPN Model: A deep learning model designed for image-to-image translation, the Attribute-Controlled and Geometry-Preserving GAN (ACGPN) emphasizes the production of realistic human faces with adjustable properties. ACGPN, first presented in a research paper in 2019, uses the GAN architecture and consists of a discriminator network for realism assessment and a generator network for image production. Notably, the discriminator guarantees the retention of geometric aspects like facial structure, while the generator permits attribute control, allowing the synthesis of images with traits like gender and age. To maintain geometric fidelity and improve attribute-controlled generation, ACGPN goes through a multi-stage training process that gradually improves image quality until realism is consistent with natural imagery. ACGPN is a critical development in deep learning, exhibiting excellent performance on datasets like as CelebA and excelling in both quantitative benchmarks and qualitative evaluation.

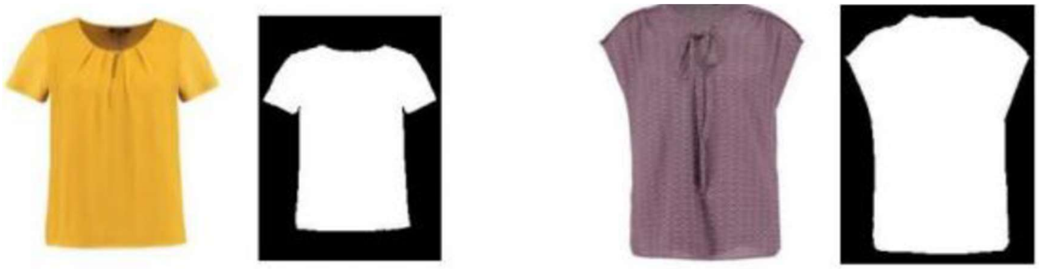

Fig. 8. A. Cloth and its Cloth Mask after using U-2 Net Architecture

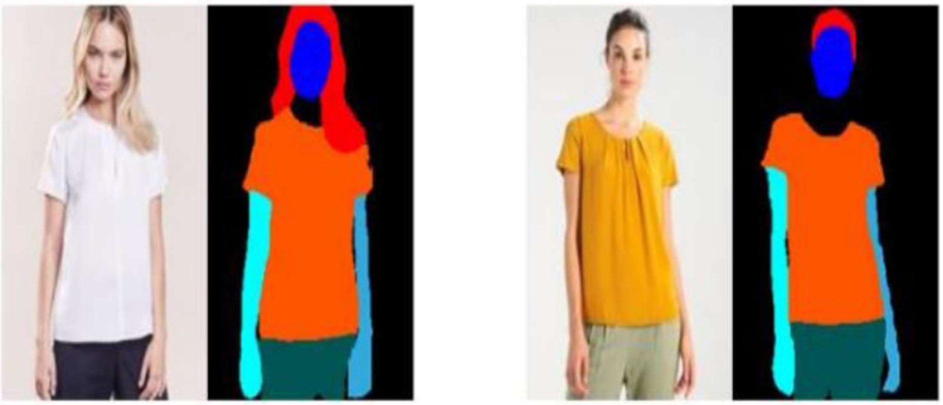

Fig. 9. Reference Image and Synthesized Human Image

The proposed model, the Virtual Try-On System, implements the ACGPN MODEL and U-2 Net architecture and is combined with a Human Parser to create a generative model that can replace a person's garment in any pose with the appropriate item.

U-2 Net architecture: Fig. 8 depicts cloth masking using U-2 Net architecture. Because the U-2 Net architecture uses a hybrid loss function, dilated convolutions, and residual connections, it is a breakthrough in deep learning for image segmentation. This combination has significantly contributed to the discipline by improving segmentation accuracy.

Human Parsing: Human parsing in virtual top try-on involves precisely detecting and segmenting human body parts from images. By identifying key anatomical landmarks and distinguishing between different body regions, human parsing enables accurate overlaying of virtual tops onto the wearer's body. This process ensures that the virtual garment fits naturally and seamlessly, enhancing the realism of the virtual try-on experience. Human parsing plays a crucial role in aligning the virtual clothing with the wearer's pose and anatomy, ultimately improving virtual top try-on systems' overall accuracy and effectiveness. Fig. 9 shows the Reference Image and Synthesized Human Image after Human Parsing.

4. RESULTS

The results of the proposed model are depicted in Fig.10. A custom dataset was created, and the model was applied to it. The results are shown in Fig. 11. The results obtained from the virtual top try-on system utilizing ACGPN, U-2Net, and human parsing showcase significant advancements in realism and accuracy. Quantitative evaluations reveal improved garment fitting precision, with ACGPN ensuring lifelike top simulations through attribute control. U-2Net enhances segmentation accuracy, enabling precise placement of virtual garments. Human parsing accurately detects and segments body parts, facilitating seamless integration of virtual clothing onto the wearer's body. Qualitative assessments demonstrate visually appealing try-on experiences, indicating the effectiveness of the integrated system in providing realistic virtual representations of clothing items. Overall, the results underscore the system's capability to enhance the virtual try-on experience, setting a promising foundation for further advancements in the field.

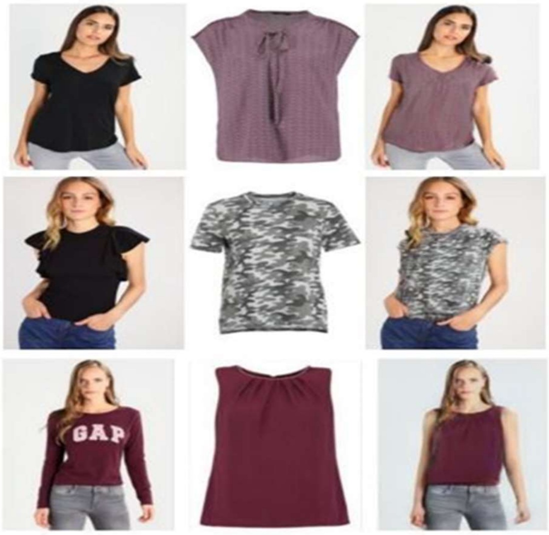

Fig. 10. Proposed System Results on VITON Dataset.

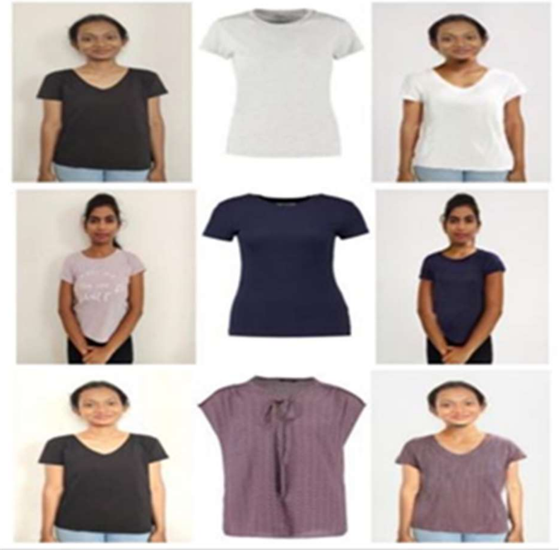

Fig. 11. Proposed System Results on Custom Dataset.

5, CONCLUSION AND FUTURE ENHANCEMENT

In conclusion, virtual top try-on technology has significantly transformed the online shopping experience, allowing users to preview clothing items in a realistic virtual environment. Virtual try-on systems have achieved remarkable accuracy and realism by leveraging advanced techniques such as attribute-controlled GANs, U-2 Net architecture, and human parsing. However, there are still areas for future enhancement. These include improving the fidelity of virtual garment simulation, enhancing user interaction and customization options, and integrating augmented reality for even more immersive try-on experiences. As technology advances, virtual top try-on systems will become even more integral to online retail, providing consumers with an unparalleled shopping experience from the comfort of their homes. Future research directions include further refinement of garment simulation and exploration of augmented reality integration for enhanced user interaction.

REFERENCES

[1]Minaee, Shervin and Boykov, Yuri and Porikli, Fatih and Plaza, Antonio and Kehtarnavaz, Nasser and Terzopoulos, Demetri, ”Image segmentation using deep learning: A survey,” IEEE, Vol. 44, No. 7, pp.3523–3542, 2021.

[2]Pang, Shanchen and Tao, Xixi and Xiong, Neal N and Dong, Yukun, ”An efficient style virtual try on the network for clothing business industry,” arXiv preprint arXiv:2105.13183, 2021.

[3]Munea, Tewodros Legesse and Jembre, Yalew Zelalem and Weldege- briel, Halefom Tekle and Chen, Longbiao and Huang, Chenxi and Yang, Chenhui, ”The progress of human pose estimation: A survey and taxonomy of models applied in 2D human pose estimation,” IEEE, vol. 8, pp. 133330–133348, 2020.

[4]Han, Xintong and Wu, Zuxuan and Wu, Zhe and Yu, Ruichi and Davis, Larry S, “Viton: An image-based virtual try-on network,” Proceedings of the IEEE Conference on Computer vision and pattern recognition, pp. 7543–7552, 2018.

[5]Dabolina, I. and Silina, L. and Apse-Apsitis, P.” Evaluation of clothing fit,” IOP Conference Series: Materials Science and Engineering, no. 459, no. 1, pp. 012077, 2018.

[6]Yang, Han and Zhang, Ruimao and Guo, Xiaobao and Liu, Wei and Zuo, Wangmeng and Luo, Ping,” Towards Photo-Realistic Virtual Try-On by Adaptively Generating Preserving Image,” arXiv preprint arXiv:2003.05863, 2020.

[7]Ren, Bin and Tang, Hao and Meng, Fanyang and Runwei, Ding and Torr, Philip HS and Sebe, Nicu,” Cloth interactive transformer for virtual try-on,” ACM Transactions on Multimedia Computing, Communications and Applications, vol. 20, no. 4, pp. 1-20, 2023.

[8]Raffiee, Amir Hossein and Sollami, Michael,” Garmentgan: Photorealistic adversarial fashion transfer,” IEEE, pp. 3923–3930, 2021.

[9]Han, Xintong and Hu, Xiaojun and Huang, Weilin and Scott, Matthew R, ”Clothflow: A flow-based model for clothed person generation,” Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10471–10480, 2019.

[10]Hsieh, Chia-Wei and Chen, Chieh-Yun and Chou, Chien-Lung and Shuai, Hong-Han and Cheng, Wen-Huang, ”Fit-me: Image-based virtual try-on with arbitrary poses,” IEEE International conference on image processing (ICIP), pp.4694–4698, 2019.