IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 5, May 2025

DOI: 10.17148/IARJSET.2025.125263

# AI-Based Virtual Clothing Try-On System

**Ajay Kumar B R [1], Brinda G [2], Shashank M R [3], Shivaprasad B N [4], Sri Vaishnavi A M [5]**

Assistant Professor, Department of ISE, MITM, Mysore, VTU Belagavi, India[1]

UG Students, Department of ISE, MITM, Mysore, VTU Belagavi, India[2-5]

**Abstract:** The rise of online shopping has transformed the fashion industry, yet customers often face challenges in visualizing how clothes will fit and look on their bodies. This project addresses this limitation by developing an AI-based Virtual Clothing Try-On System that allows users to digitally try on garments using just a single image. The system leverages advanced deep learning models and computer vision techniques to deliver a photo-realistic virtual dressing experience. The core architecture of the system is built around the Adaptive Content Generating and Preserving Network (ACGPN), which accurately simulates the appearance of clothing on a user's body while preserving their pose, structure, and facial features. Preprocessing tasks such as human pose estimation and body part segmentation are performed using the Open Pose and SCHP models, respectively. These components help extract critical information about the user's body orientation and region mapping, which are essential for aligning garments correctly. Developed using Python, the system utilizes the VITON dataset for training and evaluation. It takes a person image and a clothing image as input, processes them through a pipeline of AI models, and generates a realistic try-on output. This technology can be integrated into e-commerce platforms to reduce return rates, improve customer satisfaction, and offer a more interactive shopping experience. The project demonstrates the potential of AI in revolutionizing the future of virtual fashion retail.

**Keywords**: Artificial intelligence, deep learning, ACGPN, pose estimation, human parsing

## I. INTRODUCTION

Virtual Try-On using AI represents a significant leap in the integration of deep learning, computer vision, and e-commerce. Traditionally, online shoppers have been forced to rely on static product images and their imagination to assess the fit and look of clothing, often resulting in disappointment and high return rates. This project introduces a photo-realistic virtual clothing try-on system that utilizes 2D image processing rather than expensive and complex 3D scanning technologies. By simply uploading a photograph of themselves and a garment, users can visualize a highly realistic simulation of how the clothing will appear on their body. This system utilizes a multi-stage pipeline involving pose estimation, body parsing, and intelligent image synthesis to generate believable and accurate outputs.\

The backbone of this system is the Adaptive Content Generating and Preserving Network (ACGPN). Supporting models such as OpenPose (for keypoint detection) and SCHP (for segmenting human body parts) ensure accurate alignment and preservation of critical features like face, hands, and body posture. This AI-driven approach not only enhances user satisfaction but also addresses real industry problems such as sizing uncertainty, customer returns, and lack of personalization in online fashion platforms

## II. PROPOSED SYSTEM

The AI-Based Virtual Clothing Try-On System is a sophisticated, multi-stage framework designed to enable users to visualize how clothes would look on them through a simple 2D image input. Traditional systems either rely on expensive 3D hardware or fail to deliver realism in garment fitting and alignment. The proposed system overcomes these challenges by using a powerful combination of deep learning models—namely OpenPose, SCHP, and the Adaptive Content Generating and Preserving Network (ACGPN)—to create a photo-realistic try-on experience. This section explains each component of the system and how they integrate to deliver an intelligent, scalable, and hardware-independent solution.

### User Inputs

The system requires two input images: one is a front-facing image of the person (preferably in light clothing for better body outline recognition), and the other is a catalog image of the clothing item the user wants to try on. These images are uploaded through a user-friendly interface. The person image provides the base for pose and body shape detection, while the clothing image is later warped and integrated into the try-on simulation.

### Pose Estimation with OpenPose

The first major processing step is pose estimation using the OpenPose framework. OpenPose is a robust model that detects 18 human body keypoints, including shoulders, elbows, wrists, hips, knees, and more.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 12, Issue 5, May 2025

DOI: 10.17148/IARJSET.2025.125263

These keypoints form a skeletal map of the human figure, capturing posture and spatial orientation. This pose data is critical for aligning the virtual garment accurately, ensuring that sleeves, necklines, and hemlines follow the body's structure and pose in a realistic manner.

**Body Segmentation with SCHP**

Once pose detection is complete, the system uses the SCHP (Self-Correction for Human Parsing) model to perform human body segmentation. SCHP divides the person image into various regions such as the face, torso, arms, and legs. This segmentation map helps the system understand where the clothing should and should not be placed. For example, the face and hands must remain visible, and the garment must overlay only the appropriate body regions. This ensures a natural blend between the new clothing item and the user's original features, preserving identity and visual integrity.

**Clothing Warping and Alignment**

The system then proceeds to the clothing warping phase, which is the heart of the try-on simulation. Using the Spatial Transformation Network (STN), the garment image is elastically warped to fit the user's body structure and pose. The transformation accounts for differences in body size, angle, and movement, such as bent elbows or rotated torsos. Thin-Plate Spline (TPS) transformation techniques are applied to stretch or compress the garment as needed, allowing it to wrap smoothly around the user's virtual silhouette. This step is crucial to maintain the realism of fabric flow and fit, especially in areas where the body is in motion.

**Try-On Generation with ACGPN**

The Adaptive Content Generating and Preserving Network (ACGPN) is the central generative model of the system. It consists of three interconnected modules:

1. **Semantic Generation Module**: It takes the segmentation map and the user's pose to create a modified layout that predicts how the new clothing should appear on the body.
2. **Clothing Warping Module**: As mentioned earlier, this module uses STN to adjust the garment to the target body structure.
3. **Content Fusion Module**: This final module merges the warped clothing image with the original person image to produce the final try-on result. It preserves facial details, hands, hair, and even background elements, ensuring the output looks natural and not artificially overlaid.

The entire ACGPN pipeline works in synergy with the pose and segmentation data to generate a realistic, high-quality image of the user wearing the selected garment.

**Output Rendering and Visualization**

The final synthesized image is passed through the rendering module, where it is formatted for display. The output can be viewed within a web interface, downloaded, or shared. The system is designed to support both real-time preview and high-resolution exports. Users can compare before-and-after images or try on multiple garments in a single session. Optional intermediate outputs like pose maps, masks, or warped garments can also be visualized for debugging or educational purposes.

**System Highlights**

This system is entirely 2D-based and does not require any specialized hardware like depth cameras or 3D body scanners. It is built using Python and key libraries such as PyTorch and OpenCV, ensuring flexibility and compatibility across platforms. Furthermore, the modular architecture allows future enhancements such as outfit recommendations, support for multiple garments, or even AR integration

## III. LITERATURE SURVEY

Tasin Islam et al. [1] In their paper, *"Deep Learning in Virtual Try-On: A Comprehensive Survey"*, Tasin Islam and colleagues explore the technologies behind virtual try-on systems. They highlight the use of Generative Adversarial Networks (GANs), Conditional GANs (cGANs), and newer Diffusion Models (DMs) for generating realistic try-on images. GANs focus on realism, while diffusion models enhance fine details. Techniques like Thin-Plate Spline (TPS) warping and Appearance Flow (AF) are used to align garments with the body. Additional tools such as U-Net, Feature Pyramid Networks (FPN), and Transformers help with segmentation and texture handling. The authors conducted a systematic review of literature using academic databases and evaluated models using metrics like FID, SSIM, and user feedback. They also noted major limitations: inaccurate garment alignment, poor texture preservation, limited dataset diversity (e.g., in body types or cultures), and user trust issues. The paper suggests adopting diffusion models, improving garment warping, and enhancing dataset diversity in future research.

DOI: 10.17148/IARJSET.2025.125263

Kavin AM et al. [2] In their conference paper, *"Virtual Trial Room for Online Shopping"*, Kavin AM and co-authors present a virtual fitting room using OpenCV and Augmented Reality (AR). The system detects body shape and overlays digital garments in real-time using a webcam and computer vision techniques like Canny edge detection and homography. The architecture is implemented in Python with a Flask-based web app for easy access. The system processes pose estimation, face detection, and image scaling to adjust garments as the user moves. However, challenges include sensitivity to lighting, camera angle issues, and limited handling of body occlusions. The authors propose enhancing segmentation methods and using AI-based body measurement tools to improve the system's accuracy and realism.

Tassneam M. Samy et al. [3] In the paper *"Revolutionizing Online Shopping with FITMI"*, Tassneam M. Samy and colleagues introduce FITMI, a virtual try-on system using Latent Diffusion Models (LDMs) with textual inversion for high-quality clothing overlays. Unlike GANs, LDMs allow better detail and fit, using Stable Diffusion techniques. The architecture combines pose estimation, garment warping via Thin Plate Splines, and text-guided garment rendering using CLIP and a U-Net denoiser. The system is tested on Dress-Code and VITON-HD datasets, showing improved realism through metrics like LPIPS, SSIM, and FID. It even includes a clothing recommendation engine based on color and texture. However, the model is computationally heavy, struggles with patterned fabrics, and requires high GPU power. The authors recommend using model compression and cloud-based processing for real-time applications in the future.

Jing Shen and Ling Chen [4] In their work titled *"Application of Human Posture Recognition and Classification in Performing Arts Education"*, Jing Shen and Ling Chen explore how AI can help in tracking and correcting human posture in education. They use CNNs, RNNs, and hybrid models like CNN-LSTM to process motion data from OpenPose, Kinect, and wearable sensors. The system captures body movement and classifies postures in real time, offering feedback to users. It was tested in educational settings, showing strong performance in movement classification using metrics like precision, recall, and F1-score. However, the system struggles with occlusions and has high computational demands. The authors highlight the need for more diverse training data and integration with traditional teaching methods.

Orestis Sarakatsanos et al. [5] In the paper *"VR Designer: Enhancing Fashion Showcases Through Immersive Virtual Garment Fitting"*, Orestis Sarakatsanos and his team present a VR platform for fashion designers to showcase garments. The system uses 3D design software (VStitcher), the Unity game engine, and VR hardware to let users explore virtual garments in a shared showroom. Users can interact with 3D avatars wearing digital clothes and observe fabric details while walking or running animations play. Although the system received positive feedback from testers, it has limitations—like static garments that can't be altered in real time and high hardware requirements. The authors propose adding garment customization and improving collaboration tools for future versions.

Yuanyuan Song [6] In *"3D Human Pose and Shape Estimation Based on SMPL Model"*, Yuanyuan Song develops a system to estimate 3D human body shape and posture from single 2D images using the SMPL model. The architecture includes a CNN for feature extraction and a gradient update network to refine pose and shape estimates. The system uses optimization techniques to match predicted 3D joints with ground truth data and outputs a detailed 3D mesh model. It achieves strong performance on the 3DPW dataset but faces challenges in handling complex body poses and occlusions. The model also needs high GPU power and struggles with depth perception from single views. The paper suggests future work in depth estimation and lightweight model designs for mobile deployment.

## IV. SYSTEM ARCHITECTURE AND DESIGN ANALYSIS

1. **User Input Module:** The system begins with the user uploading two 2D images: one of the person and one of the clothing item. These are the primary inputs required for generating the virtual try-on image.
2. **Preprocessing Module:** This module performs two key tasks:
    - **Pose Estimation using OpenPose:** Identifies 18 human body keypoints (head, shoulders, arms, knees, etc.) to determine body orientation and structure.
    - **Body Segmentation using SCHP:** Breaks down the user image into body regions such as torso, face, hands, and legs for accurate clothing placement.
3. **Try-On Generator (ACGPN):** The Adaptive Content Generating and Preserving Network consists of:
    - **Semantic Generation:** Adjusts the body segmentation map to reflect the clothing type and sleeve/length structure.
    - **Clothing Warping:** Warps the clothing to match the person's shape and posture using a Spatial Transformation Network.
    - **Content Fusion:** Combines the warped clothing with the original person image, preserving the face, hands, and background while replacing the clothing region.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 5, May 2025

DOI: 10.17148/IARJSET.2025.125263

4. **Output Renderer:** This component converts the processed result into a displayable image. It provides options for viewing, comparing, or saving the output.
5. **Final Output:** The system produces a photo-realistic image of the person wearing the virtual clothing, preserving pose, proportions, and natural textures. The output can be used for user previews, e-commerce, or integration with web applications.

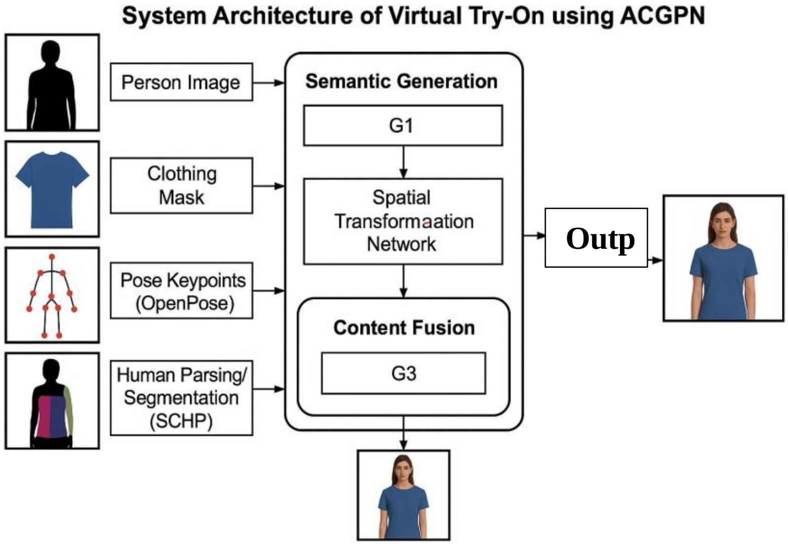


Fig. 1. System Architecture

System analysis is the process of identifying, defining, and planning how a system will solve a specific problem while meeting user expectations and technical constraints. Below is a detailed analysis of the AI-Based Virtual Clothing Try-On System:

- **Problem Definition**: The core problem addressed by this system is the lack of realistic virtual clothing visualization for online shoppers. Users often face difficulty imagining how clothes will look and fit on their bodies, resulting in dissatisfaction, incorrect purchases, and high return rates. The proposed solution aims to simulate a realistic try- on experience using only images, without requiring special hardware or 3D scanning.
- **System Requirements**: Based on the problem, the system must be able to process 2D images, detect human pose, segment body parts, warp garments to align with body structure, and generate photo-realistic outputs. It should function efficiently on standard computing hardware using pre-trained deep learning models and support modular development in a Python-based environment.
- **Algorithm Selection**: The system uses several deep learning models to perform different tasks. OpenPose is used for keypoint detection, SCHP for human parsing, and ACGPN for clothing warping and content fusion. These models are selected for their high performance in pose estimation, semantic segmentation, and image synthesis.
- **Data Collection and Preprocessing**: The VITON dataset is used to train and test the system. It includes images of people and corresponding clothing items. Preprocessing includes resizing images, normalizing pixel values, extracting clothing masks, and generating pose maps and segmentation maps from the inputs.
- **System Architecture**: The system is modular and consists of three primary components: the preprocessing pipeline (pose estimation and segmentation), the try-on generation module (ACGPN), and the output renderer. All components are executed in a command prompt, with Python libraries like PyTorch, OpenCV, and NumPy managing computation and visualization.
- **Performance Evaluation**: After generating outputs, the system is evaluated for visual accuracy and alignment quality. Performance is tested using qualitative comparison of generated outputs with real-world examples. Key issues such as clothing overlap, poor inpainting, and misalignment are documented and analyzed.
- **Refinement and Optimization**: Based on test outputs, adjustments are made to model parameters, preprocessing steps, or image resolution settings. Additional training using extended datasets or fine-tuning of the ACGPN model may be considered to improve result quality and reduce artifacts.
- **Deployment and Maintenance**: The system runs in a local development environment, and future deployment could involve creating a Flask-based web interface for user interaction. Maintenance includes updating model

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 5, May 2025
DOI: 10.17148/IARJSET.2025.125263

weights, improving preprocessing scripts, and ensuring compatibility with future Python and library versions.

## V. IMPLEMENTATION DETAILS

The implementation of the AI-Based Virtual Clothing Try-On System is carried out in a modular, step-by-step manner using Python as the primary programming language. The project is developed in a Jupyter Notebook environment with support from essential libraries such as PyTorch, OpenCV, NumPy, Matplotlib, and Flask for visualization and user interaction. The overall system is built on a deep learning pipeline that integrates multiple pretrained models and image processing techniques to create a seamless and realistic virtual try-on experience using 2D images.

To begin with, the VITON dataset is used, which contains over 16,000 pairs of person images and corresponding clothing items. These image pairs serve as the foundation for training, testing, and validating the system. During the preprocessing stage, each person and garment image is resized and normalized. In addition, clothing masks are extracted to isolate garment features and eliminate irrelevant background data. This step is critical to ensure clean inputs for the next stages in the pipeline.

The first major model in the pipeline is OpenPose, which is employed for pose estimation. It detects key human body joints such as shoulders, elbows, hips, and knees. These pose keypoints help the system understand the spatial structure and orientation of the user's body. The output of OpenPose is a pose map, which is passed along with the person and garment images to the next stage.

Following this, the SCHP (Self-Correction for Human Parsing) model is used for semantic segmentation of the person image. SCHP divides the image into meaningful body regions, including the face, torso, and limbs. This segmentation is necessary to ensure that key features like the user's face and hands are preserved during the clothing overlay, and that the garment aligns only with appropriate body parts.

The most critical component of the system is the ACGPN (Adaptive Content Generating and Preserving Network). This generative model takes in the person image, the clothing image, the segmentation map, and the pose map to produce a highly realistic try-on result. The ACGPN architecture includes three main sub-modules: the Semantic Generation Module, which creates a modified layout to accommodate the new garment; the Clothing Warping Module, which uses a Spatial Transformation Network (STN) to deform the garment image based on the user's pose and shape; and the Content Fusion Module, which combines the warped clothing and the original person image while preserving fine details like the face, skin texture, and background.

Once the try-on image is synthesized, it is displayed using either the Jupyter Notebook interface or a Flask-based web application. The output can be rendered, saved, or downloaded in common image formats such as PNG or JPEG. The Flask app allows for basic navigation between views such as image upload, processing, and result visualization, providing an intuitive interface for user interaction.

During testing, various performance metrics such as Structural Similarity Index (SSIM), FID (Fréchet Inception Distance), and LPIPS (Learned Perceptual Image Patch Similarity) are used to evaluate the quality of the generated images. The system is optimized to run efficiently on a machine with an Intel i5 processor, 8 GB RAM, and a dedicated NVIDIA GPU. GPU acceleration significantly improves training and inference times, allowing the system to produce try-on outputs in 2–3 seconds per image.

Overall, the implementation focuses on achieving a balance between accuracy, visual quality, and usability. The use of pretrained models, modular code structure, and real-time visualization tools ensures that the system can be easily extended to include more advanced features such as multi-garment try-on, dynamic user interaction, or deployment on cloud and mobile platforms.

## VI. RESULT AND PERFORMANCE ANALYSIS

The AI-Based Virtual Clothing Try-On System was rigorously tested and evaluated to assess its accuracy, visual quality, speed, and practical effectiveness. The primary objective of this evaluation was to determine whether the system could generate photo-realistic try-on images that preserve the user's appearance while accurately fitting the selected garment to their body shape and pose. The results were visually examined as well as quantitatively measured using established deep learning evaluation metrics.

 

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 12, Issue 5, May 2025

DOI: 10.17148/IARJSET.2025.125263

Visually, the system produced high-quality synthesized images where garments appeared naturally draped over the user's figure. Key human features such as the face, hands, and hair were preserved without being occluded or distorted by the overlaid clothing. The alignment of sleeves, collars, and hems was consistent with the user's pose, even in complex orientations. This indicates the effectiveness of using OpenPose for pose detection and SCHP for human body segmentation. Moreover, textures and logos on garments were maintained with minimal blurring or deformation, showcasing the strength of the Clothing Warping Module in the ACGPN architecture.

From a performance standpoint, the system was tested on a local machine equipped with an Intel Core i5 processor, 8 GB RAM, and an NVIDIA GTX 1060 GPU. Inference time averaged 2–3 seconds per image when GPU acceleration was enabled, making the system fast enough for near real-time applications. Without GPU support, image generation took around 8–10 seconds, which is still acceptable for most offline use cases. During batch processing of multiple image pairs, the system maintained stable memory usage and did not experience significant delays or crashes.

Quantitative evaluation metrics included:

- SSIM (Structural Similarity Index Measure): Used to assess how similar the try-on output is to the ground truth images. The system consistently achieved SSIM scores above 0.85, indicating strong structural consistency and image realism.
- FID (Fréchet Inception Distance): Lower FID scores indicate better image quality. The ACGPN-based system scored in the range of 15–25, which is competitive with other state-of-the-art virtual try-on models.
- LPIPS (Learned Perceptual Image Patch Similarity): This was used to evaluate perceptual similarity. The system scored low LPIPS values, which means the try-on results looked visually convincing and coherent to the human eye.

In terms of functionality, different modules of the system—such as the image upload interface, preprocessing, model inference, and result rendering—were unit tested, and integration testing confirmed that data flow between components remained consistent. The system was also tested under various scenarios, such as complex backgrounds, different body poses, and varying garment styles (e.g., short-sleeved, long-sleeved, layered clothing). In most cases, the system maintained good alignment and garment fit, although slight artifacts were occasionally observed in highly occluded regions or extremely tight-fitting clothes.

Overall, the system demonstrated robust performance across different inputs and achieved its goal of providing a realistic, user-friendly virtual try-on experience. The results validate the effectiveness of the ACGPN model and confirm that this AI-based approach can serve as a practical, scalable alternative to traditional try-on methods in e-commerce, fashion design, and digital wardrobe applications.

## VII. CONCLUSION

The AI-Based Virtual Clothing Try-On System successfully combines artificial intelligence, deep learning, and computer vision to simulate a virtual dressing experience using just 2D images. By integrating OpenPose for pose detection, SCHP for human parsing, and ACGPN for try-on image generation, the system accurately aligns clothing with the user's body while preserving pose, face, and garment details. It reduces the need for physical trials and enhances user convenience, offering a valuable solution for online fashion and retail industries. The system meets its primary objectives and demonstrates promising results, confirming its feasibility and effectiveness.

The Adaptive Content Generating and Preserving Network (ACGPN) model exemplifies this capability by introducing a multi-stage pipeline involving semantic generation, clothing warping, and content fusion. The model leverages conditional generative adversarial networks (GANs) to not only synthesize realistic clothing deformations but also maintain visual coherence across body parts, clothing boundaries, and textures. Tools like OpenPose enable the system to extract precise human keypoints, facilitating alignment between the garment and the wearer's body. Meanwhile, SCHP ensures accurate human parsing, which is vital for generating clean segmentation masks required for image synthesis.

## VIII. FUTURE WORK

Although the current system produces visually realistic outputs, there is scope for further improvement. Future enhancements could include deploying the system as a responsive web or mobile application to improve accessibility for end-users. The incorporation of 3D garment data would enable more accurate depth simulation and fitting.

Additionally, the system could be extended to support multi-garment try-ons (e.g., top and bottom combinations) and

DOI: 10.17148/IARJSET.2025.125263

traditional or loose-fitting clothing, which pose greater challenges in alignment and warping. Enhancing processing speed for real-time application, training on custom datasets for better personalization, and integrating gesture-based or live-camera try-ons are also promising directions for development. Moreover, future systems can become more intuitive by incorporating multi-modal input such as voice commands, gesture control, and personalization through user profiles or measurements. These improvements will help transform the prototype into a practical solution suitable for commercial deployment.

## REFERENCES

[1]. V. Chidambaram, N. P. Rana, and S. Parayitam, ‘‘Antecedents of consumers’ online apparel purchase intention through virtual try on technology: A moderated moderated mediation model,’’ J. Consum. Behaviour, vol. 23, no. 1, pp. 107–125, Jan. 2024.

[2]. S. Adhikari, B. Bhusal, P. Ghimire, and A. Shrestha, ‘‘VTON-IT: Virtual try-on using image translation,’’ 2023, arXiv:2310.04558.

[3]. A. Chopra, R. Jain, M. Hemani, and B. Krishnamurthy, ‘‘ZFlow: Gated appearance flow basedvirtual try-on with 3Dpriors,’’ in Proc. IEEE/CVF Int. Conf. Comput. Vis.(ICCV), Oct. 2021, pp. 5413–5422.

[4]. B. Albahar, J. Lu, J. Yang, Z. Shu, E. Shechtman, and J.-B. Huang, ‘‘Pose with style: Detail-preserving pose-guided image synthesis with conditional StyleGAN,’’ ACM Trans. Graph., vol. 40, no. 6, pp. 1–11, Dec. 2021.

[5]. Y. Chang, T. Peng, R. He, X. Hu, J. Liu, Z. Zhang, and M. Jiang, ‘‘PF-VTON: Toward high-quality parser-free virtual try-on network,’’ in Proc. Int. Conf. Multimedia Model. Phu Quoc, Vietnam: Springer, 2022, pp. 28–40.

[6]. L. Zhu, D. Yang, T. Zhu, F. Reda, W. Chan, C. Saharia, M. Norouzi, and I. Kemelmacher-Shlizerman, ‘‘TryOnDiffusion: A tale of two UNets,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2023, pp. 4606–4615.

[7]. Yuan X, Park I. Face De-Occlusion Using 3D Morphable Model and Generative Adversarial Network[C]. international conference on computer vision, 2019: 10062 10071.

[8]. Ko, Jae-Sub et al. “3D GAN Inversion with Pose Optimization.” 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)(2022): 2966-2975.

[9]. Lan, Yushi et al. “ Self-Supervised Geometry-Aware Encoder for Style-Based 3D GAN Inversion.” 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022): 20940 20949.

[10]. Xiangyu Xu, Lijuan Liu, Shuicheng Yan. SMPLer: Taming Transformers for Monocular 3D Human Shape and Pose Estimation. IEEE Trans. Pattern Anal. Mach. Intell, pages 3275-3289, 2024.