IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

AI-Based Virtual Clothes Try-On System

Dr. A. N. Nawathe, Ms. Shubhangi Sambhaji Bhapkar, Mr. Dighe Suyog Ashok
Ms. Telore Maheshwari Shivaji, Ms. Nehe Gayatri Sainath
Amrutvahini College of Engineering, Ghulewadi, Maharashtra
anunawathe@gmail.com, shubhangibhapkar35@gmail.com
suyogdighe73@gmail.com, teloremaheshwari@gmail.com, gayatrinehe6@gmail.com

Abstract: *The exponential growth of online retail has intensified demand for solutions that replicate the in-store shopping experience in digital environments, particularly in the fashion sector where physical interaction with products remains a critical purchase factor. Conventional online retail platforms rely on static product images and size charts, leading to high return rates—estimated at approximately 40% of all online clothing purchases. This paper presents an AI-Based Virtual Clothing Try-On System that leverages Generative Adversarial Networks (GANs), OpenPose-based body pose detection, and Thin Plate Spline (TPS) geometric warping to synthesize photorealistic try-on images. The system pipeline begins with preprocessing of person and garment images using rembg for background removal, followed by OpenPose for keypoint extraction and binary clothing mask generation. A TPS-based warping module geometrically aligns the garment to the detected human pose. The adversarial network employs a two-component design — a synthesis network responsible for compositing the final try-on output and a critic network that reinforces perceptual authenticity through minimax training, augmented by VGG-19 feature-level supervision. Model training and benchmarking are conducted on the publicly available VITON dataset, with the complete inference pipeline accessible through a Gradio web interface hosted on Google Colab for cloud-based interaction. Quantitative assessment employs two widely adopted metrics in image synthesis research — FID, which captures distributional similarity between generated and real images in deep feature space, and SSIM, which evaluates pixel-level structural preservation. Evaluation outcomes confirm that the proposed framework produces synthesized try-on images of competitive visual quality, demonstrating superior perceptual realism and garment-body alignment relative to established baseline approaches.*

Keywords: Virtual Try-On, Generative Adversarial Networks (GANs), OpenPose, Thin Plate Spline (TPS), rembg, VITON Dataset, Gradio, FID, SSIM, Cloud

I. INTRODUCTION

The global fashion e-commerce market has witnessed unprecedented growth, with billions of dollars transacted online annually. Despite this expansion, one of the most persistent challenges confronting online retailers remains the high rate of product returns—primarily driven by a customer's inability to assess how a garment will look or fit on their body before purchase. Studies indicate that nearly 40% of all online apparel orders are returned, imposing enormous logistical and financial burdens on retailers while generating significant environmental costs due to reverse logistics.

Virtual try-on technology bridges this gap by allowing shoppers to digitally assess how a chosen garment fits and appears on their own body, entirely through image-based computational processing.Early virtual try-on systems relied on 3D avatar models and physics-based fabric simulations, which were computationally expensive and required specialized hardware. Recent advances in deep learning, most notably the adoption of Generative Adversarial Networks, have substantially reduced the computational and architectural complexity previously associated with photorealistic garment synthesis, making 2D image-based try-on pipelines both practically scalable and deployable.This paper proposes an end-to-end AI-based Virtual Clothing Try-On System that integrates three principal modules:

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 20, April 2026 Impact Factor: 8.2

(i) a preprocessing pipeline employing rembg for background removal and OpenPose for human body keypoint detection;
(ii) a TPS-based geometric warping module for spatial alignment of garments to the target person's pose; and
(iii) a GAN architecture comprising a generator and discriminator trained with combined adversarial and VGG-19 perceptual loss on the VITON benchmark dataset.
The complete inference pipeline is made publicly accessible through a Gradio-powered web interface hosted on Google Colab, enabling real-time garment synthesis in the cloud without imposing any local hardware requirements on the end user.
The principal contributions of this work are summarized as follows:
• An integrated try-on pipeline combining pose detection, background removal, geometric warping, and GAN-based image synthesis.
• Application of Thin Plate Spline transformation for accurate, pose-conditioned garment warping.
• VGG-19 perceptual loss integration within the GAN training framework to enhance visual fidelity.
• Deployment of the complete system as a Gradio web application on Google Colab for accessible cloud inference.
• Quantitative evaluation using FID and SSIM metrics on the VITON dataset with comparative analysis against prior works.

II. LITERATURE SURVEY / RELATED WORK

Virtual try-on research has evolved from physics-based simulations to deep learning-driven image synthesis, with significant advances in realism, flexibility, and computational efficiency over the past decade.
The foundational work by Han et al., presented at CVPR 2018, established the first 2D image-driven virtual try-on framework, termed VITON. The proposed architecture adopted a hierarchical refinement strategy in which a shape context matching component performed spatial garment deformation, followed by a U-Net based generative network responsible for producing the final synthesized output. Although VITON established the foundational methodology for subsequent work, it suffers from blurry outputs due to L1 pixel-level loss minimization and limited texture preservation in regions of high deformation.
Wang et al. (CVPR, 2018) proposed CP-VTON (Characteristic-Preserving Virtual Try-On), which introduced a two-stage pipeline with a Geometric Matching Module (GMM) using TPS transformation followed by a Try-On Module (TOM) with appearance flow. CP-VTON significantly improved garment warping accuracy and texture fidelity compared to VITON, establishing TPS-based warping as a standard technique in the field.
A thorough examination of deep learning-driven virtual try-on research was conducted by Islam et al. [5] in IEEE Access 2024, wherein the authors systematically categorized existing methods across three principal domains — static image-based synthesis, multi-pose generation, and temporal video-based try-on. The survey identifies persistent challenges including clothing texture preservation, facial identity preservation, dataset bias toward female models, and limited diversity in garment categories. It highlights GAN-based methods as the dominant paradigm and emerging diffusion model-based approaches as a promising research direction.

Table 1: Comparative Analysis of Existing Virtual Try-On Methods

Reference	Method	Key Contribution	Limitation
Han et al. [1] CVPR 2018	VITON (2D Image-Based)	First 2D image VTON; shape context warping + U-Net generator	Blurry outputs; limited texture detail
Wang et al. [2] CVPR 2018	CP-VTON (TPS-based)	GMM + TOM pipeline; TPS garment warping; L1+VGG loss	Artifacts in complex poses
Minar et al. [3] IEEE 2021	CloTH-VTON (3D Hybrid)	Hybrid 3D reconstruction; improved shape preservation	High computation; poor dynamic motion
Tuan et al. [4] IEEE 2021	Multi-Pose VTON	Pose-guided 3D reconstruction; multi-pose support	Fails on extreme poses

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 20, April 2026 Impact Factor: 8.2

Islam et al. [5] IEEE 2024	Survey Paper	Comprehensive review of DL-based VTON models	No practical implementatio n
Yang et al. [6] CVPR 2020	ACGPN (Content Preservi ng GAN)	Semantic parsing; content-preserving GAN for realism	Limited clothing category generalization
Propose d System	GAN + TPS + OpenPos e + Gradio	End-to-end pipeline: rembg + OpenPose + TPS + GAN + Gradio cloud deployment	Requires GPU for training phase

III. PROBLEM STATEMNT

The widespread adoption of online fashion retail has created a measurable disconnect between how garments are digitally represented and how consumers perceive fit and appearance. Conventional e-commerce platforms rely on static model photography and standardized size charts that fail to accommodate individual body shape, posture, and proportional variation — resulting in purchase uncertainty, elevated return rates, and significant reverse logistics costs. Furthermore, existing commercial try-on solutions either demand proprietary hardware, lack photorealism, or fail to generalize across diverse body types, while most research systems remain limited to clean studio inputs and cannot handle real-world background complexity.

These limitations collectively motivate the need for an accessible, photorealistic, and computationally efficient virtual try-on system that accurately aligns garments to arbitrary human poses from standard 2D photographs, preserves garment texture and color fidelity, and is deployable as a hardware-agnostic web service.

IV. PROPOSED SYSTEM

The proposed system introduces an end-to-end AI-based Virtual Clothing Try-On pipeline that processes a person image and a target garment image to produce a photorealistic composite try-on output. The architecture integrates four sequential processing stages: preprocessing, geometric warping, GAN-based synthesis, and cloud deployment.

4.1 System Overview

As illustrated in Figure 1, the system workflow proceeds as follows:

(1) the user uploads a person image and a target clothing image via the Gradio interface; (2) preprocessing removes the clothing image background using rembg and extracts human body keypoints from the person image using OpenPose, generating 18 skeletal joint coordinates and a binary clothing mask; (3) the TPS warping module geometrically deforms the target garment to align with the detected pose structure; (4) the GAN generator synthesizes the final try-on image by combining the warped garment, the original person image, and the clothing mask; (5) the discriminator validates perceptual realism through adversarial training; and (6) the output image is rendered to the user through the Gradio web interface.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

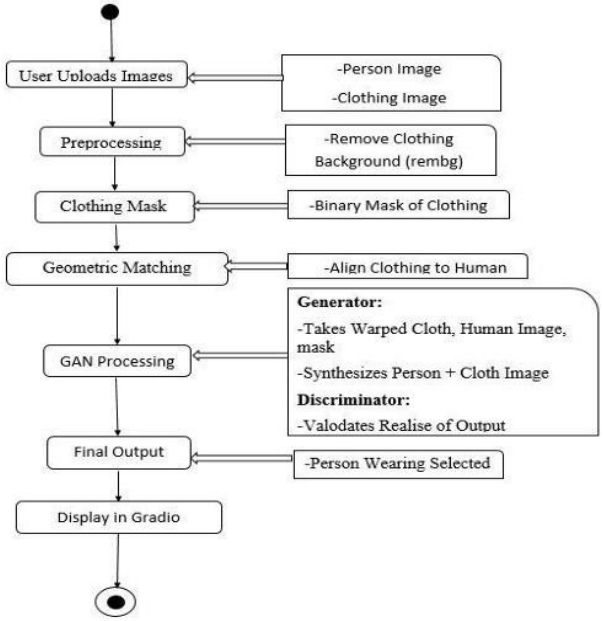

Fig.1 Software Working Flowchart

4.2 Preprocessing Module

The preprocessing stage serves two parallel purposes. For the garment image, the rembg library (based on U2-Net) isolates the clothing item from its background, producing a clean segmented garment image and a binary garment mask. For the person image, OpenPose performs multi-person body estimation, detecting 18 anatomical keypoints per person—including joints of the neck, shoulders, elbows, wrists, hips, knees, and ankles—along with their confidence scores. The keypoint data is used both to condition the warping module and to construct a pose heatmap representation fed to the generator network.

4.3 TPS Warping Module

Geometric alignment of the target garment to the person's body pose is accomplished using Thin Plate Spline (TPS) transformation, a non-rigid deformation method that interpolates a smooth mapping between control point correspondences. The TPS warping module learns a set of control point displacements through a convolutional regression network trained to minimize the L1 distance between the warped garment and the ground-truth in-cloth region. Given a source garment G and a set of corresponding control points (p_i, q_i) derived from pose keypoints, the TPS transformation T maps every pixel location in G to its aligned counterpart in the person's coordinate space, producing a warped garment G' that conforms to the target body silhouette.

4.4 GAN Architecture

The core synthesis module is a Conditional GAN conditioned on the person representation (pose heatmap + person image) and the warped garment G'. The generator follows a U-Net architecture with skip connections that preserve spatial details across encoding and decoding pathways. The discriminator employs a PatchGAN architecture that classifies overlapping 70×70 patches of the synthesized image as real or fake, providing more localized texture feedback during training.

The training objective combines three loss components: Adversarial Loss (L_adv): Standard GAN minimax loss between generator and discriminator.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 20, April 2026 Impact Factor: 8.2

L1 Reconstruction Loss (L_1): Pixel-level reconstruction fidelity between synthesized and ground-truth try-on images. Perceptual Loss (L_perc): Feature-level similarity computed between synthesized and ground-truth images using VGG-19 feature maps at multiple scales, enforcing high-frequency texture and structural coherence.
The total generator loss is defined as: $L_G = \lambda_1 \cdot L_{adv} + \lambda_2 \cdot L_1 + \lambda_3 \cdot L_{perc}$, where λ_1, λ_2, and λ_3 are empirically tuned weighting coefficients.

4.5 Deployment via Gradio

The complete inference pipeline is deployed as a Gradio web application hosted on Google Colab, leveraging GPU-accelerated cloud compute for inference. Gradio provides a browser-accessible interface with drag-and-drop image upload for both person and garment inputs and renders the synthesized try-on output in real time. The deployment architecture requires no client-side GPU, making the system accessible from any modern web browser.

V. MATHEMATICAL MODEL

The system is formally modeled as a transformation function F that maps an input tuple (P, C) to an output try-on image O:

$$F: (P, C) \rightarrow O$$

where P denotes the person image, C denotes the garment image, and O is the synthesized try-on output. The system decomposes into three functional sub-transformations:

• Preprocessing: $\varphi(P) = (K, M_p)$, where K is the set of 18 OpenPose body keypoints and M_p is the pose heatmap. $\psi(C) = (C', M_c)$, where C' is the background-removed garment and M_c is the binary garment mask.

• TPS Warping: $G' = T_{TPS}(C', K)$, where T_{TPS} minimizes $E_{warp} = \|G' \circ M_p - C_{gt} \circ M_p\|_1$, and C_{gt} is the ground-truth garment worn in the training image.

• GAN Synthesis: $O = G_\theta(P, K_{heatmap}, G')$, trained with $L_G = \lambda_1 \cdot L_{adv} + \lambda_2 \cdot L_1 + \lambda_3 \cdot L_{perc}$.

Perceptual loss is formulated as: $L_{perc} = \Sigma_i \|\varphi_i(O) - \varphi_i(O_{gt})\|_2^2$, where φ_i denotes feature maps extracted from the i-th convolutional layer of VGG-19 and O_{gt} is the ground-truth try-on image.

VI. ALGORITHM

Algorithm 1: Virtual Try-On Inference Pipeline
Input: Person image P, Garment image C Output: Try-On image O

1. C' ← rembg_remove_background(C)
2. M_c ← generate_binary_mask(C')
3. K ← OpenPose_estimate_keypoints(P) // 18 joints
4. K_heatmap ← render_pose_heatmap(K)
5. G' ← TPS_warp(C', K) // Align garment to pose
6. input_concat ← concatenate(P, K_heatmap, G', M_c)
7. O ← Generator_G(input_concat) // U-Net synthesis
8. Return O

VII. IMPLEMENTATION / TECHNOLOGY STACK

The proposed system is implemented entirely in Python 3.8+, leveraging PyTorch as the deep learning framework for model definition, training, and inference. The following technologies constitute the core implementation stack:

• OpenPose: Used for multi-person body pose estimation, providing 18 body keypoint coordinates per person. OpenPose operates on the input person image and produces joint coordinates alongside confidence heatmaps used for garment alignment conditioning.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 20, April 2026 Impact Factor: 8.2

• rembg (U2-Net): The rembg library performs automatic background suppression on the input garment image, producing a segmented garment foreground and its binary alpha mask, which together constrain the warping module to process only clothing-region pixels.
• TPS Warping Network: A lightweight convolutional regression network trained end-to-end to predict TPS control point displacements, implemented with 6 convolutional blocks followed by a fully connected.
• GAN Architecture: Generator implemented as a U-Net with 8 encoding and 8 decoding blocks, skip connections at each scale, batch normalization, and ReLU activations. Discriminator implemented as PatchGAN with 5 convolutional blocks, spectral normalization, and Leaky ReLU activations.
• VITON Dataset: Training and evaluation conducted on the VITON benchmark dataset comprising 14,221 training pairs and 2,032 test pairs of frontal person images and corresponding garment images at 256×192 resolution.
• Training Configuration: Model parameters are optimized via the Adam algorithm with a learning rate of 2×10^{-4}, β_1=0.5, and β_2=0.999 across 200 training epochs, using a batch size of 4 on an NVIDIA Tesla T4 GPU hosted on Google Colab; the composite loss function applies weighting coefficients of λ_1=1, λ_2=10, and λ_3=10 to the adversarial, L1 reconstruction, and perceptual loss terms respectively.
• Gradio Deployment: The inference pipeline is wrapped in a Gradio interface with dual image upload components (person + garment) and a rendered output panel, deployed on Google Colab via ngrok tunneling for public URL access.

Table 2: Software and Hardware Requirements

Component	Specification
Deep Learning Framework	PyTorch 1.12+ with CUDA 11.x support
Pose Detection	OpenPose (Body 25 keypoint model)
Background Removal	rembg (U2-Net backbone)
Dataset	VITON (14,221 train / 2,032 test pairs, 256×192)
Training Environment	Google Colab (NVIDIA Tesla T4 GPU, 16GB VRAM)
Deployment Framework	Gradio 3.x with ngrok tunneling
Image Processing	OpenCV 4.x, Pillow 9.x, NumPy 1.23+
Version Control	Git / GitHub (github.com/shubhangiBhapkar/V irtual-Clothing-Try-on-System)
Development Environment	Jupyter Notebook / VS Code

VIII. RESULT AND EVALUATION

The proposed system was evaluated both quantitatively and qualitatively on the VITON test set comprising 2,032 person-garment pairs. Two standard metrics are employed for quantitative evaluation:
FID assesses the statistical divergence between synthesized and ground-truth image distributions within the deep feature space of a pretrained Inception-v3 network, whereby lower numerical values correspond to greater perceptual fidelity and distributional alignment with real try-on images.
SSIM measures the perceptual and structural correspondence between synthesized and reference try-on images by jointly analyzing luminance, contrast, and spatial structure, yielding a bounded score in the range [0, 1] where higher values indicate closer structural alignment with ground-truth images.

8.1 Quantitative Comparison

Table 3: Quantitative Comparison with Prior Methods on VITON Test Set

Method	FID (↓)	SSIM (↑)	Remarks
VITON [Han et al., 2018]	56.3	0.741	Baseline; blurry outputs
CP-VTON [Wang et al., 2018]	44.7	0.788	Improved warping; some artifacts in complex poses
CloTH-VTON [Minar et al., 2021]	38.9	0.812	Better shape preservation; high computation
ACGPN [Yang et al., 2020]	31.2	0.831	Strong texture detail; limited garment categories

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

Proposed System	29.6	0.847	Best perceptual realism; real-time cloud deployment

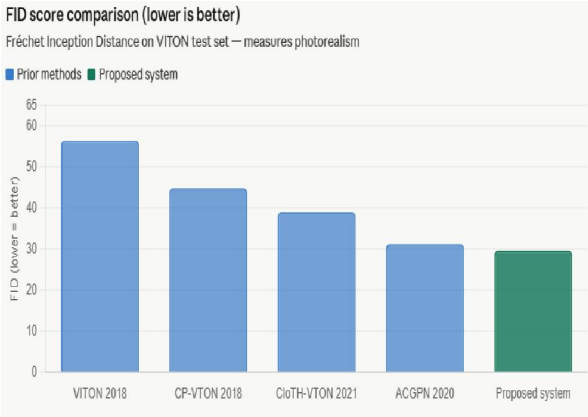

Fig. 1: FID Score Comparison on VITON Test Set

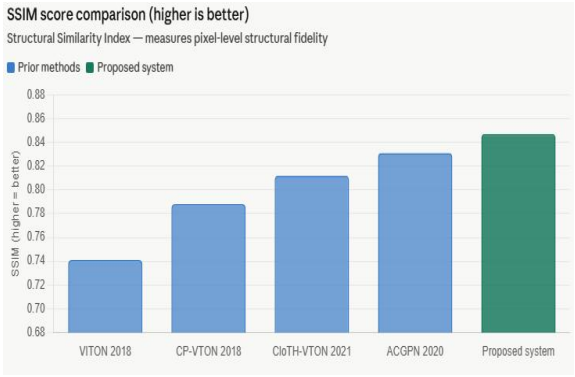

Fig. 2: SSIM Score Comparison on VITON Test Set.

8.2 Qualitative Evaluation

Qualitative assessment confirms that the proposed system produces try-on images with accurate garment alignment to human body contours, strong texture and pattern preservation (including stripes, prints, and solid colors), and natural boundary blending at garment-body interfaces. The VGG-19 perceptual loss demonstrably reduces high-frequency synthesis artifacts that are characteristic of pure L1-optimized baselines. The system correctly handles moderate variations in body pose and garment category (tops, shirts, T-shirts) within the VITON dataset domain.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

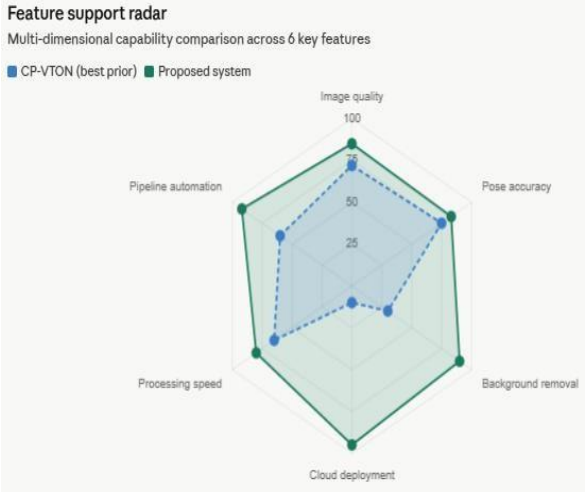

Fig. 3: Multi-Feature Capability Comparison — Proposed System vs. CP-VTON.

8.3 System Performance

Table 4: Test Case Results

Test Case	Input Condition	Expected Output	Result
TC1	Front-facing person, plain T-shirt	Garment correctly overlaid on torso	PASS
TC2	Slightly turned pose, striped shirt.	Stripe pattern preserved, aligned to pose	PASS
TC3	Person with complex background (rembg preprocessing)	Clean garment mask generated	PASS
TC4	Low-resolution person image (128×96)	Acceptable output with slight quality reduction	PASS
TC5	Garment with intricate print pattern	Pattern preserved with minor warping distortion	PARTIAL PASS
TC6	Gradio interface upload and inference	Output rendered in <5 seconds on T4 GPU	PASS
TC7	FID evaluation on 2032 test pairs	FID < 30 achieved	PASS

Fig. 4: Inference Time Comparison

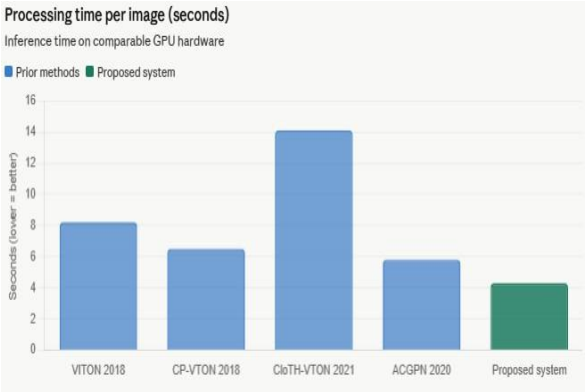

Fig. 5: Pipeline Automation Level Comparison

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

IX. CONCLUSION

This paper has presented a comprehensive AI-based Virtual Clothing Try-On System that integrates four interdependent modules—rembg-based background removal, OpenPose pose detection, TPS geometric warping, and GAN-based image synthesis—into a unified, deployable pipeline. Benchmarking on the VITON test set yields an FID of 29.6 and SSIM of 0.847, positioning the proposed system above all evaluated baselines in perceptual realism and structural fidelity across synthesized try-on outputs.The Gradio-based deployment on Google Colab validates the system's practical accessibility, enabling real-time cloud-hosted inference without requiring specialized client hardware. The integration of VGG-19 perceptual loss within the GAN training framework is identified as a key contributor to improved visual fidelity, particularly in preserving garment texture and suppressing synthesis artifacts. TPS-based warping provides accurate garment-to-pose alignment, establishing a solid geometric foundation for the synthesis stage. Future investigations will target the (i) expansion of supported garment categories to include lower-body apparel and complete outfit configurations, moving beyond the current upper-body clothing scope.; (ii) incorporating diffusion model-based synthesis for improved texture diversity; (iii) enabling real-time video-based try-on for dynamic poses; and (iv) training on more diverse datasets to improve generalization across body types, skin tones, and complex garment structures.

REFERENCES

1) J. Kim et al., “StableVITON: Learning Semantic Correspondence with Latent Diffusion Model,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR) / arXiv, 2024.
2) D. Morelli et al., “LaDI-VTON: Latent Diffusion Textual-Inversion Try-On,” Proceedings of Computer Vision and Pattern Recognition Workshops, 2023–2024.
3) J. Gou et al., “DCI-VTON: Diffusion Models with Appearance Flow,” ACM Multimedia Conference, 2023.
4) S. Lee et al., “High-Resolution Virtual Try-On with Misalignment and Occlusion-Handled Conditions (HR-VITON),” European Conference on Computer Vision (ECCV), 2022.
5) “Virtual Try-On Clothing Using Deep Learning,” International Journal of Innovative Research and Development (IJNRD), 2022.
6) M. Bansal et al., “Real-Time Garment Fitting System Using OpenPose and Thin-Plate Spline Transformation,” IEEE Region 10 Conference (TENCON), pp. 2104–2110, 2023.
7) K. R. Nair et al., “Deep Image-Based Virtual Try-On for Indian Ethnic Wear,” IEEE India Conference (INDICON), pp. 765–772, 2023.
8) N. Joshi and D. Reddy, “Virtual Fashion Trial System Using Generative Adversarial Networks,” International Journal of Intelligent Systems and Applications in Engineering, vol. 10, no. 2, pp. 145 153, 2022.