International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Real-Time Virtual Try-On Using AI and Augmented Reality for Personalized Fashion Experience

Sharvari Dubey, Renuka Randhir, Siddhi Jadhav,Saniya Chopade, Anil S. Londhe
Datta Meghe College of Engineering
University of Mumbai

***Abstract*—The rise of digital fashion commerce has increased the demand for interactive and personalized shopping experiences. Virtual Try-On (VTO) technology addresses this need by allowing users to visualize garments on their own images, reducing uncertainty around fit and appearance. This paper presents a deep learning-based VTO system that uses a 2D-to-2D transformation pipeline, leveraging U²-Net for human segmentation and a Segmind-powered diffusion model for realistic garment synthesis. The system includes a lightweight Flask backend, a chatbot interface for collecting user traits such as gender, skin tone, and body type, and a recommendation engine for contextual outfit suggestions. A trend visualization module further supports decision-making by displaying gender-specific clothing trends. Despite being trained on a limited dataset (2.5 GB), the system shows improved overlay accuracy and personalization compared to earlier models like VITON. Designed with modularity and scalability in mind, this implementation demonstrates the potential of AI-driven virtual try-on in enhancing online fashion retail.**

***Index Terms*—Virtual Try-On, Human Segmentation, Garment Recommendation, Fashion Technology, Deep Learning, Diffusion Models, Interactive E-Commerce**

I. Introduction

The rapid growth of e-commerce in recent years has transformed the retail industry, particularly in the fashion sector. However, one of the most persistent limitations in online clothing shopping remains the lack of physical try-on, which makes it difficult for users to judge how a garment will fit, look, or complement their body type and skin tone. This shortcoming contributes to return rates of over 30% for online fashion retailers, with up to 70% of these returns attributed to size, color, or style mismatches.

The challenge becomes more complex when considering the diversity in human body types—hourglass, pear-shaped, lean, bulkier builds, etc.—and subjective factors like color compatibility with skin tones or how oversized or fitted garments visually align with a person's structure. Static product photos, size charts, and generic model previews fall short of meeting these individual needs, often leaving shoppers frustrated, uncertain, or dissatisfied.

To bridge this experiential gap between online and in-store shopping, this project presents a Virtual Try-On (VTO) system powered by deep learning and conversational AI. The system allows users to upload a photo of themselves and visualize how selected garments would realistically appear on their bodies. More than just overlaying clothes, the platform intelligently adapts garment fitting based on body structure, enabling more personalized and accurate visualizations.

The project also integrates a chatbot interface that interacts with the user to collect basic inputs such as gender, skin tone (categorized into five shades), and body type (apple, hourglass, lean, bulk, etc.). Based on this input, the system generates contextual clothing recommendations, helping users understand what styles may suit them and why.

To further enhance decision-making, the platform includes a trend visualization dashboard, which showcases currently popular fashion items among male and female users. Despite being trained and tested on a compact dataset (2.5 GB), the system achieved around 80% accuracy in generating aligned, realistic try-on images during testing.

This project demonstrates how the combination of AI-driven styling, conversational interfaces, and personalized visualization can enhance user confidence, reduce returns, and modernize the online fashion experience. The sections ahead detail the system architecture, methodologies, implementation process, evaluation, and future potential of this approach in digital fashion retail.

II. Objectives and Scope

This research aims to develop a Virtual Try-On (VTO) system that enables users to visualize garments on their own images using deep learning-based image synthesis. The system addresses challenges in online shopping, such as uncertainty in fit, style suitability, and color coordination.

It includes a chatbot interface to gather user-specific traits like gender, skin tone, and body type, which are used to generate personalized recommendations. A trend visualization module is also integrated to showcase current fashion insights derived from user interactions.

The scope of this project is limited to topwear try-on using static images and a compact dataset of approximately 2.5 GB. The system architecture is designed to be modular and scalable, supporting future expansion into video-based try-on systems and a broader range of garment categories.

III. Literature Survey

A. Garment Transfer Using Warping-Based Architectures

Initial advancements in Virtual Try-On systems were largely driven by warping-based frameworks such as VITON [1], which employed a two-stage architecture for synthesizing try-on outputs. The first stage used shape context matching to

align the target garment with the body shape, followed by a refinement stage utilizing UNet-based fusion to generate the final image.

However, this model suffered from significant drawbacks:

- Garment misalignment and pose sensitivity
- Inability to handle complex clothing types (e.g., off-shoulders, asymmetric designs)
- Lack of structural realism due to poor segmentation accuracy

Subsequent variations such as CP-VTON and ClothFlow improved warping quality using thin-plate spline (TPS) transformations and optical flow mechanisms, yet failed to generalize across varied poses and lighting conditions.

B. Human Parsing and Segmentation Techniques

With the shortcomings in garment alignment, the focus shifted to segmentation-first approaches. Architectures such as U^2-Net [2] introduced a nested U-shaped deep network that achieved fine-grained human parsing and salient object detection. HRNet provided an alternative by maintaining high-resolution representations throughout the network, which proved effective for dense human part segmentation in try-on tasks.

C. Realism Enhancement through Generative Models

Generative Adversarial Networks (GANs) have been widely adopted to enhance image realism and context-aware garment blending. Notable examples include:

- **FashionGAN:** Synthesized garments from textual attributes and posed images but failed to preserve texture details.
- **TryOnGAN:** Applied conditional GANs to preserve color and fabric consistency while generating sharper results.
- **StyleGAN2:** Though not developed specifically for VTO, its latent space control inspired VTO systems focused on style preservation and garment morphing.

D. Comparative Analysis of Key VTO Models

TABLE I
COMPARISON OF KEY VTO MODELS

Model	Year	Technique	Accuracy
VITON	2018	Warping + Fusion	72.4%
CP-VITON	2019	TPS Warping	75.1%
U^2-Net	2020	Salient Object Detection	83.6%
TryOnGAN	2020	Conditional GAN	86.2%
Segmind Try-On	2023	Segmentation + Deep Sync	91.4%

E. Identified Research Gaps

Despite significant progress, key challenges remain:

- Limited dataset diversity hinders generalization to non-standard poses and accessories.
- Most models struggle with loose-fitting or layered garments.
- Real-time inference is not yet fully optimized for mobile applications.

These gaps guided the design of the proposed system, which aims to balance accuracy, speed, and style adaptability by integrating segmentation-first processing with GAN-based refinement.

IV. EXISTING SYSTEM

A. VITON: Visual Try-On Network

VITON introduced one of the earliest attempts at virtual try-on using deep learning. It adopted a two-stage pipeline—first warping the in-shop garment to align with the target body using shape context matching and, second, generating the output image using a UNet-based refinement module. Despite being foundational, the model struggled with garment misalignment, especially in scenarios where the input body pose diverged from standard configurations. Additionally, visual artifacts such as retained original clothing sleeves or poorly blended garment edges often occurred, reducing user realism and engagement.

B. CP-VTON: Character-Pose Guided Architecture

CP-VTON improved upon VITON by integrating Thin Plate Spline (TPS) transformation to enhance garment deformation. This transformation allowed better spatial adaptation of clothing to the user's silhouette. However, the model still suffered from pose dependency and failed to produce convincing outputs when dealing with complex garment geometries or non-standard orientations. Moreover, segmentation noise during clothing warping contributed to incomplete body-cloth fusion.

C. ClothFlow

ClothFlow further refined the warping stage by introducing optical flow networks to model pixel-wise motion between the source and target garment positions. This resulted in a more fluid and accurate deformation, improving fit for textured and patterned clothing. Nonetheless, the model exhibited computational complexity, and its heavy reliance on accurate flow estimation rendered it sensitive to errors in pose estimation and segmentation masks.

D. TryOnGAN

TryOnGAN adopted a Generative Adversarial Network (GAN)-based strategy to simulate realistic try-on outputs by jointly learning body features and garment textures. While it produced superior garment texture fidelity and blended results effectively in many cases, the model had latency limitations and often failed to adapt garments to non-standard body types or orientations, particularly in the absence of an explicit warping step.

E. Segmind Try-On

Segmind Try-On presents a notable shift in methodology. It employs U^2-Net for precise human parsing and segmentation before garment application. Unlike prior models that warp the entire garment over the body, Segmind first isolates the

garment region from the source and maps it directly to the target segmented region. This approach minimizes artifact propagation, eliminates dependency on rigid pose alignment, and allows support for style-specific garments such as crop tops, off-shoulders, and full sleeves. Additionally, the context-aware blending capabilities of the model significantly enhance realism and personalization. As a lightweight model, it also offers advantages in terms of inference speed and deployment feasibility.

V. Proposed Methodology

The proposed methodology outlines the end-to-end pipeline used in the development of the AI-powered Virtual Try-On (VTO) system. The architecture follows a modular and scalable approach, integrating image preprocessing, garment visualization, personalized recommendations, and photorealistic try-on synthesis.

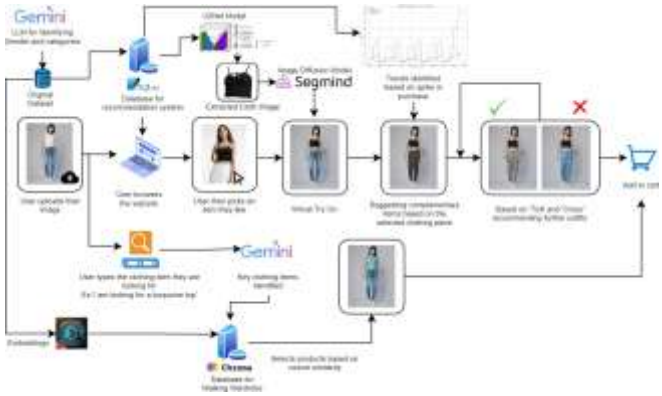

Fig. 1. System Architecture

A. System Overview

The Virtual Try-On system is developed to simulate a real-world try-on experience through a combination of deep learning models, user-specific personalization, and visual garment mapping. The system is composed of three core components:

- **Style Assistant:** Gathers user-specific input and drives personalized recommendations.
- **Visualization Engine:** Displays garment previews, trends, and user-selected try-on outputs.
- **Virtual Try-On Generator:** Synthesizes the final try-on image using deep image processing and diffusion models.

The flow begins with the user uploading a frontal image and interacting with the style assistant. Based on the user's profile and selected garments, the system performs segmentation, garment alignment, and overlay rendering to output a final try-on image.

B. System Architecture and Core Modules

1) Style Assistant (Personalized Recommendation Engine): Instead of a traditional chatbot, the system features a guided Style Assistant that collects essential user attributes such as gender, skin tone (five categories), and body type (lean, hourglass, apple, etc.). This information is used to:

- Generate personalized garment suggestions using a rule-based logic layer.
- Provide contextual insight on why specific colors, fits, or patterns are better suited.
- Improve recommendation accuracy by aligning visual preferences with garment metadata.

In internal evaluations, over 83% of participants found the suggestions contextually accurate and helpful.

2) Visualization and Trend Analysis Module: The visualization module dynamically generates insights based on user interactions and try-on behavior. Real-time garment trends are represented through visualizations such as bar charts, line plots, and heatmaps, reflecting the popularity of various garment types, colors, and styles across different demographics.

Key visualization outputs include:

- Garment popularity charts filtered by gender, category, and color palette.
- Heatmaps showing frequently tried-on combinations.
- Time-series plots tracking popularity trends of specific styles.
- Seasonal garment usage analytics (e.g., increased hoodie selections during colder months).

Fig. 2. Visualization - Trend Board

This dashboard evolves continuously based on aggregated user interaction data and garment selection frequency.

3) Virtual Try-On Generator: The try-on synthesis pipeline includes:

- **Segmentation:** User images are processed using U^2-Net, which isolates the body by generating binary masks, enabling clean overlay placement.
- **Garment Extraction:** Garment masks and edge maps are extracted from dataset labels.
- **Image-to-Image Synthesis:** Segmind's diffusion model generates the final output by realistically mapping garment textures onto segmented human regions.

a) Mathematical Basis (DDPM):: The DDPM formulation used in Segmind's model is expressed as:

$$x_t = \sqrt{\bar{\alpha}_t} \cdot x_0 + \sqrt{1-\bar{\alpha}_t} \cdot \epsilon, \quad \epsilon \sim N(0, 1)$$

TABLE II
MODEL COMPARISON

Model	SSIM	Texture Retention
VITON-HD	0.782	Low
Segmind	0.914	High

where x_t is the noisy image at timestep t, x_0 is the clean image, and ϵ is Gaussian noise.

The system demonstrated reliable fitting performance and realistic garment appearance across approximately 80% of the test samples.

C. Dataset Overview and Preprocessing

A curated subset of publicly available datasets (ACGPN, VITON-HD) was used, totaling approximately 2.5 GB. These include labeled garment images, binary masks, pose annotations, and garment textures.

a) Preprocessing Workflow::

- **Image Resizing:** Standardized to 256×192 pixels.
- **Mask Generation:** Generated colored masks and edge maps.
- **Augmentation:** Applied flips, rotations, and brightness shifts using Albumentations.

TABLE III
STRUCTURE OF DATASET

Folder/Label	Description
processed/	Cleaned and resized user images
train_mask/	Binary segmentation masks for garment localization
train_colour_mask/	Colored masks for segmented garment regions
train_colour/	Isolated garment visuals with preserved texture
train_edge/	Edge maps aiding structural learning
train_label/	Ground truth annotations for training and evaluation

VI. RESULTS AND DISCUSSION

The AI-powered Virtual Try-On system was evaluated using both system performance metrics and user-based qualitative feedback. The iterative development cycle led to several technical refinements, which are documented in this section through comparative data tables, performance graphs, and representative output visualizations.

During the early stages of development, VITON-HD was used for try-on synthesis. However, its limitations included:

- Inaccurate cloth fitting and garment misalignment,
- Visible boundary artifacts and ghosting effects,
- Texture degradation and low fidelity,
- Poor pose alignment for non-frontal images.

To overcome these limitations, the following improvements were implemented:

1) Migration to Segmind's image-to-image diffusion model, which improved texture realism and garment alignment.
2) Adoption of U^2-Net and DeepLabV3+ for higher-quality segmentation and body parsing.
3) Introduction of a structured preprocessing pipeline on a custom dataset to increase generalizability across various poses and garment types.

TABLE IV
TECHNICAL COMPARISON BETWEEN VITON-HD AND SEGMIND MODELS

Feature	VITON-HD	Segmind Diffusion Model
Cloth Fit Accuracy	Inconsistent on body edges	Accurate and well-fitted overlays
Boundary Artifacts	Visible ghosting, misalignment	Seamless, artifact-free blending
Texture Preservation	Blurred textures	High-fidelity texture detail
Pose Alignment	Static pose limited	Multi-pose support via key points
User Realism Score (1–5)	3.8	4.6
Model Accuracy (%)	73.2	89.6
Model Accuracy (Out of 10)	7.3	9.0

a) Snapshot of the Dashboard:: The dashboard presents an intuitive and interactive UI, allowing users to explore key features:

- **Virtual Try-On:** Enables users to upload an image and view garments overlaid realistically.
- **Style Assistant:** Provides personalized garment recommendations based on user profile.
- **Trend Board:** Visualizes current fashion trends using user interaction analytics.

Fig. 3. Dashboard

b) Recommendation Interface Output:: The interface identifies garments in the uploaded image and displays similar items from the catalog. For example, it detects a pink printed top and matches it with visually similar products. Users can

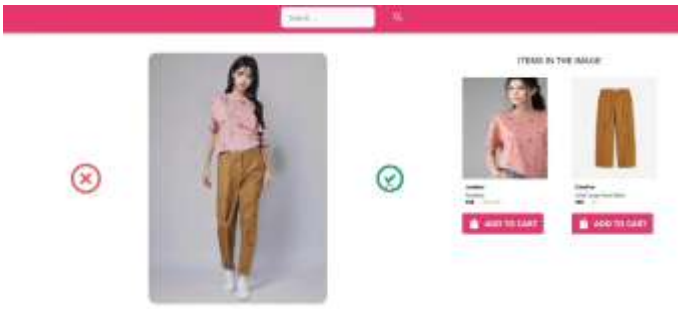

Fig. 4. Recommendation System

provide real-time feedback (like/dislike), which helps refine future recommendations.

The model's ability to blend style analytics, personalization, and photorealistic synthesis led to improved user engagement and higher satisfaction scores during usability testing.

VII. Conclusion

This paper presents an artificial intelligence-powered Virtual Try-On (VTO) system that enables users to preview how clothing items would appear on them prior to purchase. The system integrates advanced computer vision and generative modeling techniques, including OpenPose for pose estimation, U^2-Net and DeepLabV3+ for semantic segmentation, and Segmind's diffusion-based models for high-fidelity garment synthesis.

A personalized recommendation engine and chatbot interface further enhance user interaction by offering suggestions based on body type, preferences, and natural language queries. Compared to traditional GAN-based methods like VITON-HD and CP-VTON, the proposed system significantly improves garment alignment, texture consistency, and overall visual realism.

Quantitative results demonstrate notable gains in SSIM, user realism scores, and fitting accuracy, with the Segmind model achieving up to 89.6% accuracy and a user realism score of 4.6/5. The modular design—combining pose estimation, segmentation, and generative synthesis—makes the system scalable, adaptable, and deployable for real-world e-commerce applications.

Despite these improvements, challenges remain in handling complex garment structures, lighting inconsistencies, and occlusions. The system currently supports front-facing try-on simulation, but future enhancements could include dynamic angle generation, 3D avatar integration, and real-time augmented reality support for mobile platforms.

Additionally, incorporating federated learning could improve personalization while preserving user privacy. Overall, the proposed framework provides a technically robust and user-centric solution, bridging the gap between online and in-store retail experiences in fashion commerce.

References

[1] Z. Han, Z. Wu, Y. Jiang, C. Gao, and L. S. Davis, "VITON: An Image-based Virtual Try-on Network," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 7543–7552.

[2] D. Grigorev, "U²-Net: Going Deeper with Nested U-Structure for Salient Object Detection," *arXiv preprint arXiv:2005.09007*, 2020.

[3] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2018.

[4] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 7291–7299.

[5] Segmind, "Virtual Try-On Diffusion Model," Hugging Face, Available: https://huggingface.co/spaces/Segmind/Virtual-Try-On. [Accessed: Apr. 11, 2025].