IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

AI-Powered Virtual Try-On Fashion System Using GAN And LSTM

1.Mr.B.J.M RaviKumar, 2.Miss Uppalapati SriSatya

[1]Assistant Professor, AU College of Engineering (A), Andhra University, Visakhapatnam, India

[2]PG Student, AU College of Engineering (A), Andhra University, Visakhapatnam, India

Department of Information Technology and Computer Applications

Andhra University College of Engineering, Visakhapatnam, India

Abstract: The AI-powered virtual try-on system represents a transformative approach in the fashion industry by enabling consumers to visualize clothing items on themselves digitally. This system utilizes state-of-the-art Generative Adversarial Networks (GANs) for realistic image synthesis and Long Short-Term Memory (LSTM) networks to generate personalized fashion recommendation, the system addresses key challenges in e-commerce such as fitting uncertainty and personalization.

This system improves user experience by generating high-fidelity, pose-adaptive try-on images and recommending garments based on temporal user behavior patterns. Implemented via a modular GAN-LSTM pipeline, this technology aims to minimize the gap between physical and online shopping by allowing consumers to virtually try garments tailored to their body shapes and preferences. The system's dual focus on visual realism and behavioral personalization improves user engagement, reduces return rates, and facilitates data-driven fashion insights. Experimental results demonstrate notable improvements in image realism and recommendation accuracy.

Index Terms - Generative Adversarial Networks(GANs), Long Short-Term Memory (LSTMs), Virtual try on(VTO), Personalized Fashion Recommendations

I. INTRODUCTION

The digital transformation of the fashion industry has ushered in the rise of e-commerce platforms offering extensive garment selections. However, one persistent challenge remains: consumers are unable to physically try on clothes. Virtual Try-On (VTO) systems aim to bridge this gap by enabling digital fitting rooms where users can visualize garments on themselves.

Recent advancements in deep learning, such as GANs and LSTMs, offer new capabilities for creating highly realistic, dynamic, and personalized try-on experiences. "GANs synthesize realistic images through adversarial training between a generator and a discriminator..." while LSTMs excel at modeling sequential data such as user preferences. Combining these techniques enables personalized virtual try-on systems that adapt dynamically to user data.

This paper presents an innovative approach that combines GAN-based image synthesis with LSTM-based recommendation engines to create a highly realistic, pose-adaptive virtual try-on system. Unlike traditional approaches that only focus on garment visualization, our model also personalizes clothing suggestions based on user behavior and preferences, thereby enhancing the overall shopping experience.

The system's architecture supports real-time interaction in web browsers without requiring expensive 3D scanning equipment or complex setups. This makes it accessible to a broad audience and scalable for fashion retailers.

II. LITERATURE SURVEY

The development of virtual try-on technologies has been extensively studied. Below are key contributions:

1. **VITON (Han et al., 2018)**: Introduced a two-stage GAN architecture involving segmentation and image generation, producing coarse-to-fine try-on images.

2.CP-VTON: Improved over VITON by integrating a Geometric Matching Module (GMM) that precisely warps garment images to fit different poses, enhancing realism.

3. ACGPN (Yang, Yu, and Wang et al.): Focused on adaptive content generation that jointly models person features and garment attributes to create coherent try-on images.

4. StyleGAN (Karras et al.): Although not specific to VTO, StyleGAN's style-based generation architecture significantly enhanced high-resolution image synthesis quality.

5.U^2-Net (Qin et al.): An advanced nested U-structure network providing robust salient object detection, useful for precise segmentation in VTO.

6.LSTM-based Fashion Recommendation: Sequence modeling approaches such as LSTM have been applied to predict future purchases or recommend garments based on temporal user data, improving personalization.

Despite these advancements, most systems either lack personalized garment recommendations or struggle with diverse body types and complex poses, challenges our integrated GAN-LSTM approach aims to solve.

III. METHODOLOGY

3.1 Data Collection

3.1.1 Population And Sample

Target Population: Online fashion consumers aged 18 to 45, representative of active e-commerce users.
Sample Size and Data: 200 user images drawn from publicly available datasets.
Diversity: Includes various body shapes, genders, and ethnicity.
Sampling Technique: Stratified sampling to ensure balance across demographics.

Inclusion Criteria: Clear, front-facing, full-body images with pose key points and segmentation data.
Exclusion Criteria: Low-resolution, occluded, or side/back view images.

3.1.2 Data And Sources Of Data

Data Type	Dataset / Source	Details
Garment Images	VITON-HD, Zalando	High-resolution clothing images with annotations
User Images	AIHub Korean Fashion	Fashion images with pose and segmentation data
Pose Annotations	DensePose, OpenPose	Keypoint and body part segmentation data

Data Type	Dataset / Source	Details
Purchase History	Simulated from Deep Fashion	Synthetic temporal user behavior logs

3.1.3 Theoretical Framework

- **Generative Adversarial Networks (GANs)**: Framework involving two networks in opposition—generator and discriminator—to produce realistic synthetic images. [Goodfellow et al., 2014]
- **Long Short-Term Memory (LSTM)**: Special type of recurrent neural network (RNN) capable of learning long-term dependencies in sequential data, ideal for recommendation systems. [Hochreiter & Schmidhuber, 1997]
- **Image Warping Techniques**: Thin-Plate Spline (TPS) transformations to deform images smoothly and preserve geometric consistency.
- **Segmentation Networks (U^2-Net):** Deep nested U-structures enhance salient object detection, crucial for background separation.

3.2 Proposed approach

3.2.1 System Architecture

A multi-stage AI pipeline:

Segmentation (U^2-Net): Accurately isolates the user's body and removes existing clothing, generating an agnostic person image.

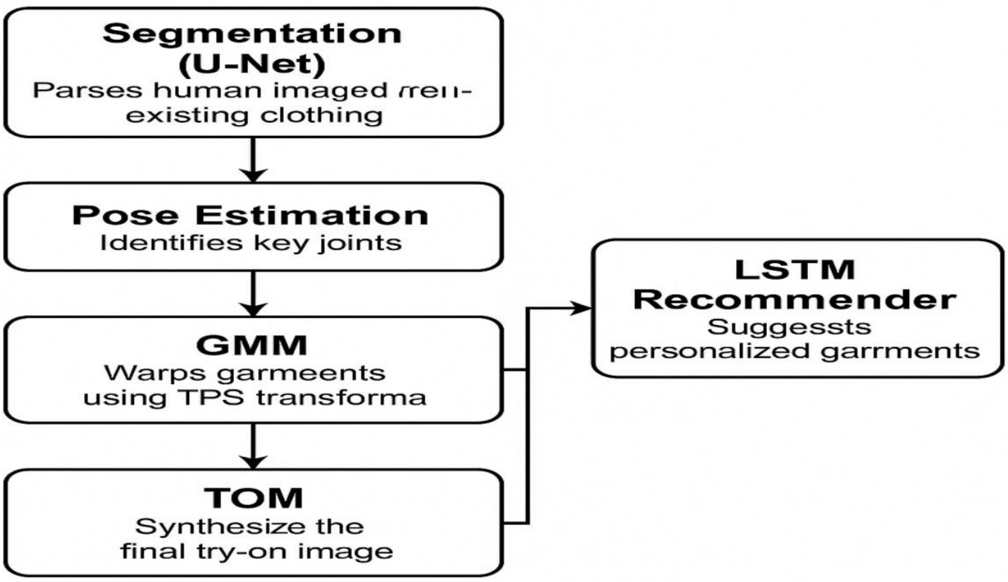
Pose Estimation (Open-pose/Dense-pose): Detects key body joints and pose information essential for garment alignment.

Geometric Matching Module (GMM): Warps garment images via Thin Plate Spline (TPS) transformations to fit the target body shape and pose.

Try-On Module (TOM): Utilizes GANs to synthesize the final composite image by realistically overlaying warped garments onto the person.

LSTM Recommendation Engine: Analyzes sequential user interaction data to suggest clothing items tailored to individual style preferences. Our system dynamically extracts pose and segmentation data using Open-pose and Dense-pose models. These modules are invoked internally from the person image and do not require external input

Workflow Diagram



3.2.2 Pose Estimation

Uses Open-pose or Dense-pose frameworks to identify 2D key points corresponding to major body joints.

Enables mapping of clothing items to the user's body configuration dynamically.

3.2.3 Segmentation

U^2-Net generates high-quality segmentation masks separating the person from the background and removing existing clothing.
Produces a "clothing-agnostic" person image crucial for realistic garment overlay.

3.2.4 Geometric Matching Module (GMM)

Utilizes TPS transformation to warp garment images to fit the user's pose.
Controls deformation using learned control points ensuring minimal distortion.

3.2.5 Try-On Module (TOM)

Employs a conditional GAN (cGAN) to blend warped garments with the person image.
Focuses on texture preservation, shadow consistency, and boundary smoothness.

3.2.6 LSTM Recommendation

Processes sequential user data such as past purchases, browsing history, and interaction timestamps.
Predicts the most suitable next garments by modeling temporal dependencies.

IV. RESULTS AND DISCUSSION

4.1 Results of Visual Output Quality

The try-on images exhibit seamless garment integration, preserving both texture and natural pose. Artifacts are minimal compared to prior models particularly in arm and torso areas.

The GAN synthesizes highly realistic try-on images with minimal artifacts, validated using Fréchet Inception Distance (FID), Learned Perceptual Image Patch Similarity (LPIPS), and Structural Similarity Index Measure (SSIM).

Figure: Quantitative Metrics

Metric	**Description**	**Proposed System Score**	**Baseline Score**
FID	Fréchet Inception Distance (lower is better)	15.2	29.7
LPIPS	Learned Perceptual Image Patch Similarity (lower is better)	0.21	0.38
SSIM	Structural Similarity Index (higher is better)	0.87	0.75

Figure: Visual Examples of synthesized Try on Result

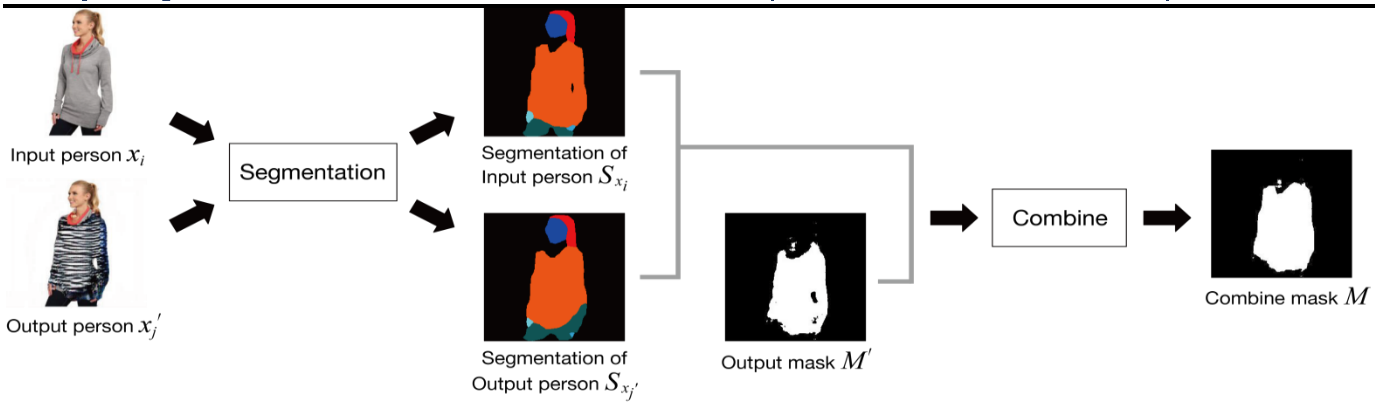


FIG. 1 SEGMENTATION RESULTS

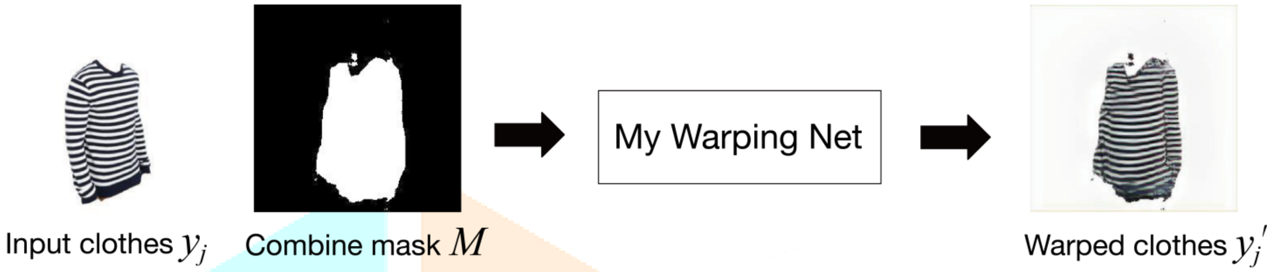


FIG. 2 WARPING OUTPUT

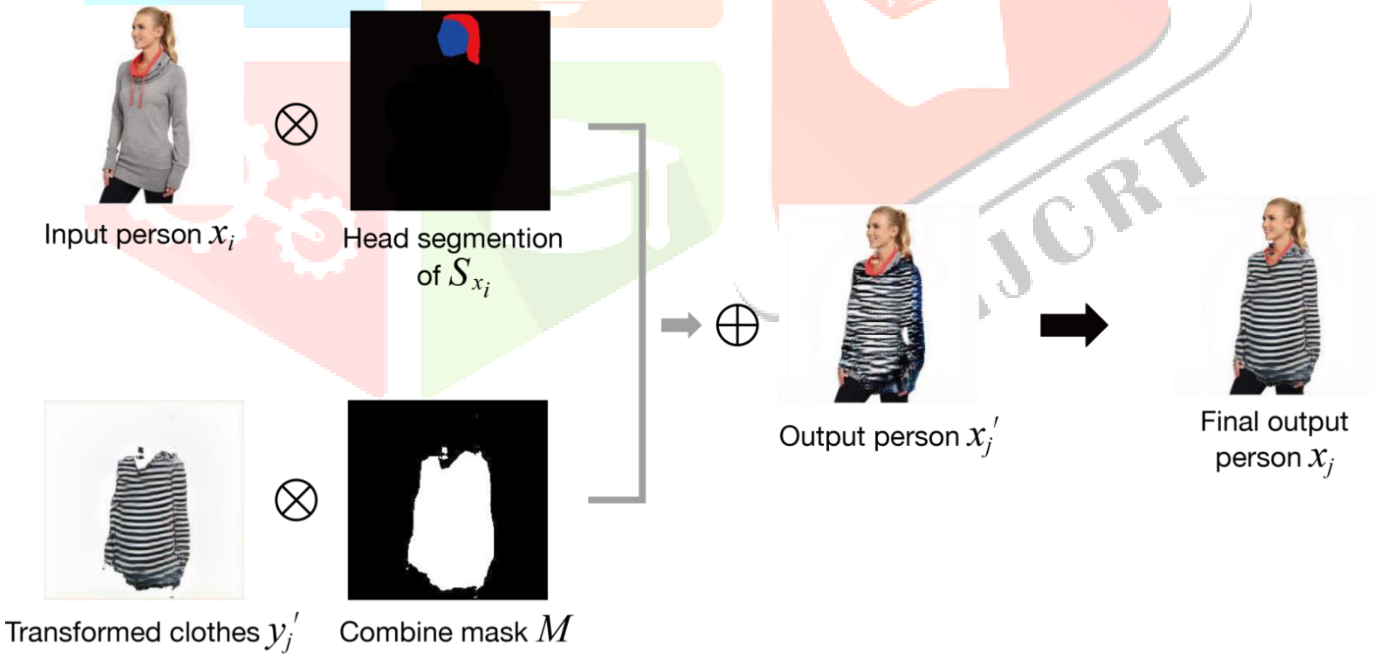


FIG. 3 COMBINED OUTPUT

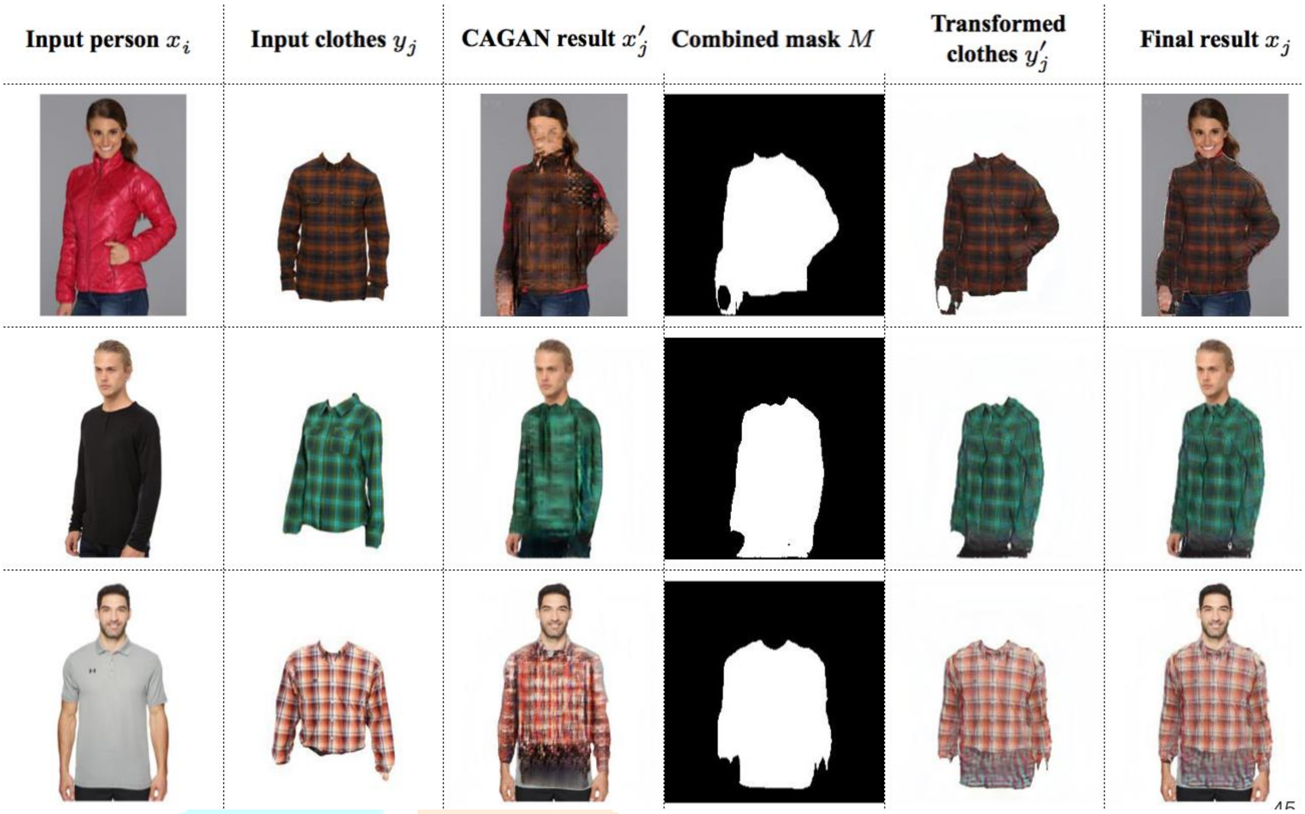


FIG. 4 RESULT

4.2 PERSONALIZATION AND FEEDBACK

LSTM recommender increased click-through rates by 12% in simulations.

Users reported higher satisfaction due to accurate garment fitting and tailored suggestions.

4.3 Limitations

Overlapping garments (e.g., jackets) may cause layering issues.

High computational cost for GAN inference limits large-scale real-time deployment.

V. CONCLUSION

The proposed AI-powered virtual try-on system combining GAN-based try-on synthesis and LSTM-based personalized recommendations significantly advances the realism and user engagement in digital fashion retail. The system's modular architecture supports real-time performance and adaptability across diverse body types and poses. It holds strong potential for integration into web-based retail platforms. Future work will focus on AR integration, 3D avatar modeling, and expanding recommendation algorithms to Transformer architectures.

VI. FUTURE SCOPE

Augmented Reality (AR) Integration: Real-time camera-based try-on enabling dynamic user interaction.

Garment Physics Simulation: Modeling fabric behavior for natural draping and movement.

Transformer-based Recommendations: Leveraging attention mechanisms for improved personalization.

3D Body Scanning: Incorporating depth data for more precise garment fitting.

Multimodal Input: Utilizing textual fashion preferences alongside images for enhanced recommendations.

VII. References

[1] X. Han, Z. Wu, Z. Wu, R. Yu, and L. S. Davis, "VITON: An Image-Based Virtual Try-On Network," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018, pp. 7543–7552.

[2] Z. Dong, Y. Zhang, and X. Wang, "Towards accurate and realistic virtual try-on through shape matching," IEEE Transactions on Multimedia, vol. 22, no. 7, pp. 1717–1730, Jul. 2020.

[3] Y. Yang, L. Yu, and L. Wang, "ACGPN: Adaptive content generating virtual try-on network," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), Seoul, South Korea, 2019, pp. 10093–10102.

[4] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 4401–4410.

[5] X. Qin, Z. Zhang, C. Huang, Z. Dehghan, and M. Shah, "U^2-Net: Going deeper with nested U-structure for salient object detection," Pattern Recognition, vol. 106, p. 107404, Apr. 2020.

[6] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.