

IDM-VITON: DIFFUSION-BASED VIRTUAL TRY-ON WITH GARMENT-GUIDED CROSS-ATTENTION

Project Reference No.: 49S_BE_4983

College : *Sambhram Institute of Technology, Bengaluru*
Branch : *Department of Artificial Intelligence and Data Science*
Guide : *Dr. Mamatha. M*
Student(s) : *Mr. Thejas C*
Ms. K Sakshi
Ms. Shruthi G
Mr. Dhanush C M

Keywords:

Virtual Try-On, Diffusion Model, Computer Vision, Image Processing, Cross-Attention, Deep Learning, Image Synthesis.

Introduction:

1. Online fashion shopping has grown rapidly, but customers often face difficulty in visualizing how a garment will look on them without physically trying it. This leads to uncertainty, poor user experience, and a high rate of product returns. Traditional virtual try-on systems based on image warping or GAN models often fail to preserve garment details, distort textures, or alter the user's identity. Recent advancements in diffusion-based generative models have significantly improved image synthesis by producing high-quality and realistic outputs. However, existing diffusion-based virtual try-on systems still face challenges such as improper garment alignment, loss of fine textures, and unintended modifications to the user's face and body.
2. This project proposes IDM-VTON, a diffusion-based virtual try-on system that integrates garment-guided cross-attention with multi-scale feature extraction. The system uses pose estimation, human parsing, and DensePose information to ensure proper garment alignment and realistic fitting. By combining these techniques, the proposed system generates high-resolution images where the garment is accurately fitted onto the user while preserving identity and maintaining fine texture details.
3. Additionally, the system is designed to be scalable and adaptable for real-world e-commerce platforms, enabling seamless integration with existing online retail systems. It can support personalized recommendations, virtual styling, and interactive shopping experiences by allowing users to experiment with multiple outfits in real time. This not only reduces return rates and operational costs for businesses but also increases customer satisfaction and confidence in online purchases. Furthermore, the approach opens opportunities for future innovations in digital fashion, including virtual wardrobes, AI-driven fashion assistants, and metaverse-based shopping experiences.

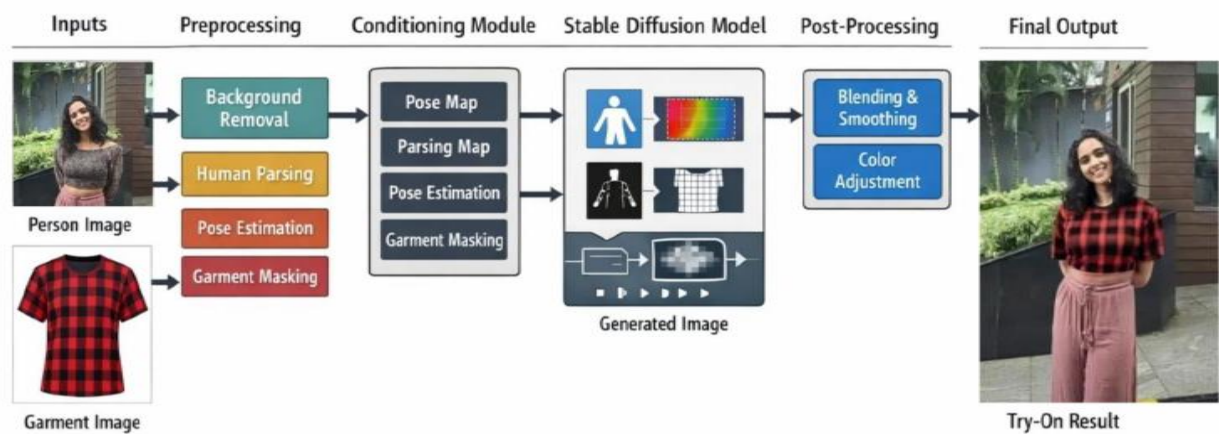


Figure 1: Architecture of the IDM-Based Virtual Try-On System

IDM-VTON Preprocessing Pipeline

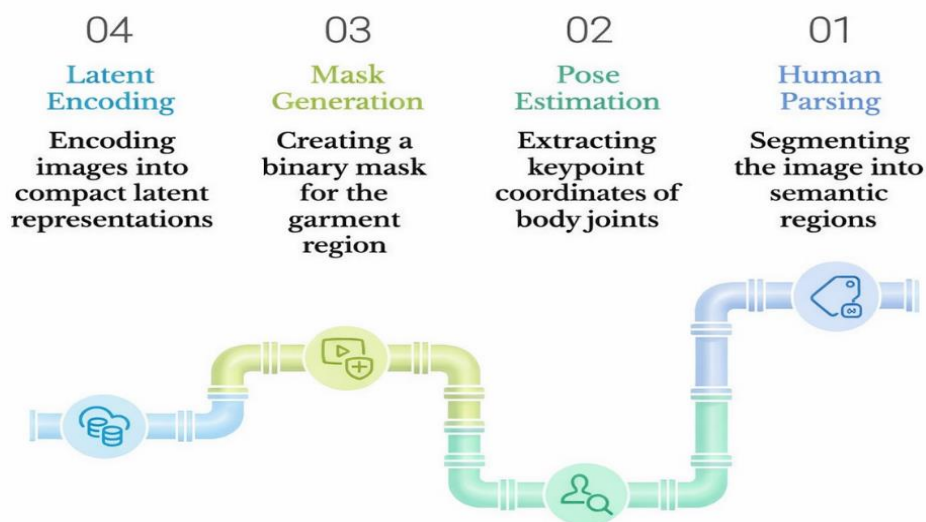


Figure 2: Preprocessing Pipeline

Objectives:

1. To develop a diffusion-based virtual try-on system for realistic garment visualization.
2. To accurately align garments with the user's body using pose estimation and DensePose.
3. To preserve identity features such as face, hair, and body structure during image generation.
4. To improve garment texture quality and overall visual realism using cross-attention mechanisms.

Methodology:

1. The proposed system takes a person image and a garment image as input. Background removal is applied to isolate the subject, and human parsing is used to segment body regions such as face, hair, and clothing areas.
2. Pose estimation and DensePose capture body structure and geometry for accurate garment alignment. A multi-scale garment encoder extracts texture and pattern features from the garment image.
3. A diffusion-based model with garment-guided cross-attention integrates these features, while latent inpainting ensures only clothing regions are modified, preserving identity.
4. Finally, the system generates a high-resolution output where the garment is realistically fitted with proper alignment and smooth blending.

Result and Conclusion:

The proposed IDM-VTON system successfully generates realistic virtual try-on images with improved garment alignment and texture quality. The use of diffusion models and cross-attention ensures high-fidelity image synthesis while preserving the user's identity.

The system demonstrates better performance compared to traditional GAN-based and warping-based approaches, reducing distortion and improving visual consistency. Overall, it provides an effective solution for online fashion applications, enhancing user experience and reducing product returns.

References:

- [1] Han et al., "VITON: An Image-based Virtual Try-On Network," CVPR, 2018.
- [2] Wang et al., "CP-VTON: Characteristic-Preserving Virtual Try-On Network," ECCV, 2018.
- [3] Choi et al., "VITON-HD: High-Resolution Virtual Try-On," CVPR, 2021.
- [4] Kim et al., "LaDI-VTON: Latent Diffusion for Virtual Try-On," AAAI, 2023.
- [5] Gupta et al., "FITMI: Multi-Input Diffusion for Try-On," 2024.

Future Scope:

The future scope of this project includes:

1. Extending the system for full-body virtual try-on applications.
2. Implementing real-time virtual try-on in mobile and web platforms.
3. Enhancing accuracy using advanced deep learning and 3D modelling techniques.