EfficientVITON: An Efficient Virtual Try-On Model using Optimized Diffusion Process

Abstract

Delivered by

Mostafa Kotb

Mariam Ayman

Ahmed Rashed

Ashrakat Saeed

Abdelrahman Said

Supervisor

Prof. Ahmed Fares

Abstract

Wouldn’t it be much more convenient for everybody to try on clothes by only looking into a mirror? The answer to that problem is “virtual try-on,” enabling users to digitally experiment with outfits. The core challenge lies in realistic image-to-image translation, where clothing must fit diverse human forms, poses, and figures. Early methods, which used 2D transformations, offered speed, but image quality was often disappointing and lacked the nuance of deep learning. Though GAN-based techniques enhanced realism, their dependence on paired data proved limiting. More adaptable methods offered great visuals but demanded significant computing power and time. Recent advances in diffusion models have shown promise for high-fidelity translation, yet the current crop of virtual try-on tools still struggle with detail loss and warping issues. To tackle these challenges, this paper proposes EfficientVITON: a new virtual try-on system leveraging the impressive pre-trained Stable Diffusion model for better images and deployment feasibility. The system includes a spatial encoder to maintain clothing’s finer details and zero cross-attention blocks to capture the subtleties of how clothes fit a human body. Input images are carefully prepared, and the diffusion process has been tweaked to significantly cut generation time without image quality loss. The training process involves two distinct stages of fine-tuning, carefully incorporating a balance of loss functions to ensure both accurate try-on results and high-quality visuals. Rigorous testing on the VITON-HD dataset, supplemented with real-world examples, has demonstrated that EfficientVITON achieves state-of-the-art results.