A FLOW-BASED GENERATIVE NETWORK FOR PHOTO-REALISTIC VIRTUAL TRY-ON

Netrawati S, Ayesha Sultana, MA Naveed, MD Abdul Sabeel, Summaiya Fatima

Netrawati S/ Assistant Prof, Dept of AIML, GNDEC, Bidar
Ayesha Sultana/ Student, Dept of AIML, GNDEC, Bidar.
MA Naveed/ Student, Dept of AIML, GNDEC, Bidar.
MD Abdul Sabeel/ Student, Dept of AIML, GNDEC, Bidar.
Summaiya Fatima/ Student, Dept of AIML, GNDEC, Bidar.
sultanaayesha2030@gmail.com

Abstract

Virtual try-on systems are gaining importance in the field of e-commerce and digital fashion, allowing users to visualize clothing items on their body without physically wearing them. However, achieving realistic fitting and alignment of garments remains a major challenge due to issues such as occlusion, pose variation, and non-rigid deformation of clothes. This paper proposes a Flow-Based Virtual Try-On Network (FVTN) that generates photo-realistic try-on results by modeling dense flow transformations between clothing and human body images. The system consists of three main modules: Parsing Alignment Module (PAM), Flow Estimation Module (FEM), and Fusion and Rendering Module (FRM). The model is trained using the VITON dataset and evaluated using Mean Square Error (MSE).

KEYWORDS-Virtual Try-On, Deep Learning, Flow-Based Network, Image Synthesis, VITON Dataset, Generative Models

1. Introduction

The rapid growth of e-commerce has transformed the fashion industry, allowing customers to shop for clothing online. However, one major limitation of online shopping is the inability to try garments before purchasing. This often leads to uncertainty and dissatisfaction among customers, resulting in high return rates and increased operational costs for businesses. Virtual try-on systems aim to solve this problem by enabling users to digitally visualize how a garment would look on their body. Early approaches relied on simple image processing techniques or 2D overlay methods, which produced unrealistic results. These methods could not handle variations in body pose, lighting conditions, or clothing deformation.

With the advancement of deep learning, more sophisticated models have been developed to improve the realism of virtual try-on systems. These models use neural networks to understand human body structure and garment features. However, challenges such as occlusion, misalignment, and distortion still persist.

In this work, we propose a Flow-Based Virtual Try-On Network (FVTN) that addresses these challenges by modelling dense flow transformations between clothing and human body images. The system ensures accurate alignment and realistic rendering, making it suitable for real-world applications in online retail. The system is trained and evaluated on the publicly available VITON dataset and deployed as a web application using Flask, enabling users to upload their own images and receive try-on results interactively. Evaluation using MSE loss demonstrates that FVTN outperforms baseline approaches qualitatively and quantitatively.

The remainder of this paper is organized as follows. Section II reviews related work. Section III describes the system architecture and design. Section IV details the technical implementation. Section V presents evaluation results, and Section VI concludes with directions for future work.

2. Related Work

I. Literature Survey

[1] Han et al. (2018)

Introduced VITON dataset and try-on model. Uses human parsing and shape matching. Limitation: Poor alignment and artifacts

[2] Wang et al. (2018)

Proposed CP-VTON model. Uses Geometric Matching Module (GMM).
Limitation: Cannot handle fine deformation

[3] Yang et al. (2020)

Introduced Adaptive Warping GAN. Uses dense flow estimation. Improves realism.

[4] Ge et al. (2021)

Proposed parser-free virtual try-on. Reduces dependency on segmentation. Faster performance.

[5] Han et al. (2021)

Proposed ClothFlow model. Uses flow-based deformation. Better alignment and realism.

II. Existing System

The existing virtual try-on systems mainly focus on overlaying garment images onto a human body image using basic image processing and deep learning techniques. These systems attempt to align the clothing with the body but often fail to produce realistic outputs due to improper handling of body structure and garment deformation. Most traditional methods rely on 2D image transformations, where the clothing is simply resized and placed over the person's image. However, these approaches do not consider important factors such as body pose, shape variations, and occlusion, which results in distorted and unnatural images. Recent deep learning-based models like VITON and CP-VTON have improved the alignment process by introducing segmentation and geometric matching techniques. Even though these methods provide better results compared to traditional approaches, they still suffer from several limitations in handling complex transformations and maintaining texture consistency. In addition, most systems fail to model the non-rigid nature of clothing, which is essential for generating realistic virtual try-on images. Poor alignment between garment and body.

III. Proposed System

The proposed system introduces a Flow-Based Virtual Try-On Network (FVTN) to generate realistic and high-quality virtual try-on images. Unlike existing methods, this system focuses on modeling the complex deformation of garments and aligning them accurately with the human body. The system uses deep learning techniques to learn dense pixel-level transformations between the clothing image and the target person image. This helps in preserving the texture, shape, and appearance of the garment while fitting it naturally onto the body. The proposed approach overcomes the limitations of existing systems by handling pose variation, occlusion, and non-rigid deformation effectively. It ensures that the generated output looks more realistic and visually appealing.

- **Parsing Alignment Module (PAM)**

  This module is responsible for segmenting the human body into different regions such as upper body, arms, and background. It helps in aligning the garment properly with respect to body parts.

- **Flow Estimation Module (FEM)**

  This module learns the pixel-level transformation required to deform the garment. It generates a dense flow map that guides how the cloth should be warped.

- **Fusion and Rendering Module (FRM)**

  This module combines the warped garment with the original person image and generates the final output with realistic appearance.

The proposed system offers several advantages over existing methods

- Provides accurate alignment between cloth and body.
- Handles different human poses effectively.
- Preserves fine texture details of garments.
- Produces realistic and visually appealing outputs.
- Reduces distortion and artifacts.

Thus, the proposed system improves the overall performance of virtual try-on applications and provides a better solution for real-world use, especially in online shopping platforms.

3. Materials and Methods

I. Methodology

1. Collection of VITON Dataset
The dataset consists of images of people and corresponding clothing items. It is used to train the model for learning garment alignment and fitting.

2. Preprocessing of Images
All input images are resized and normalized to maintain uniformity. This step helps in improving model performance and reducing computational complexity.

3. Segmentation of Human Body
The person image is divided into different regions such as upper body, arms, and background. This helps in accurate placement of the garment.

4. Feature Extraction
Important features from both the person image and cloth image are extracted. These features help the model understand shape, texture, and structure.

5. Design of Flow-Based Model
A deep learning model is designed to learn the transformation between cloth and body. It focuses on realistic warping of garments.

- Design of Flow-Based Model

A deep learning model is designed to learn the transformation between cloth and body. It focuses on realistic warping of garments.

- Design of Flow-Based Model

A deep learning model is designed to learn the transformation between cloth and body. It focuses on realistic warping of garments.

- Parsing Alignment Module (PAM)

This module aligns the garment with the segmented body regions.

- Flow Estimation Module (FEM)

This module computes dense pixel-level flow, which helps in deforming the garment according to body pose.

- Fusion and Rendering Module (FRM)

This module combines the warped garment with the person image and generates the final realistic output.

- Model Training using MSE Loss

The model is trained using Mean Square Error loss, which minimizes the difference between generated and actual images.

- Model Testing and Evaluation

The trained model is tested on new images to evaluate its performance, accuracy, and visual quality.

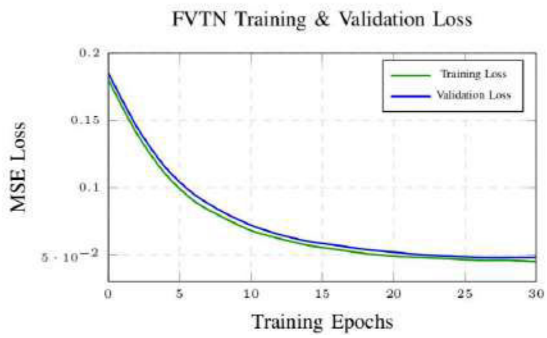

Fig.1. Training and Testing Evaluation

II. Framework

• Operating System (Windows/Ubuntu)
Provides the platform to run and develop the application efficiently.

• Programming Language (Python)
Used for implementing the model due to its simplicity and support for AI libraries.

• Deep Learning Framework (PyTorch)
Used to build and train the neural network models effectively.

• OpenCV Library
Used for image processing tasks such as reading, resizing, and transforming images.

• NumPy Library

Used for numerical operations and handling multi-dimensional arrays.

• Pandas Library
Used for managing and organizing dataset information.

• Matplotlib Library
Used for visualizing results such as graphs and output images.

• Flask Framework
Used to develop the backend and connect the model with the user interface.

• Frontend Technologies (HTML, CSS, JavaScript)
Used to design the user interface for uploading images and displaying results.

• MySQL Database
Used to store user data and application-related information technologies used for implementing the system.

3.2.4 Use Case Diagram

The use case diagram provides a visual representation of all actions that can be performed by the user.

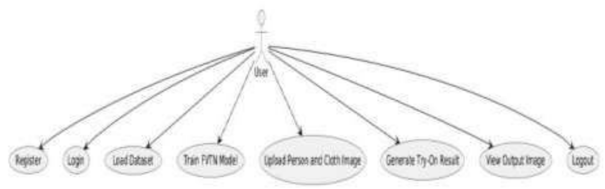

Fig.2. Use Case diagram

III. System Architecture

The system architecture represents the overall design and workflow of the proposed virtual try-on system. It describes how different components interact with each other to process the input images and generate a realistic output.

The architecture is designed in a layered manner, where each layer performs a specific function in transforming the input data into the final try-on image. The system mainly focuses on accurate garment alignment, deformation, and rendering to achieve high-quality results.
The architecture consists of the following layers:

- **Input Layer**

The input layer is responsible for receiving the data required for the system. It accepts two main inputs: the person image and the clothing image. This layer ensures that the input images are in the correct format before passing them to the next stage. It acts as the starting point of the entire system.

- **Preprocessing Layer**

The preprocessing layer prepares the input images for further processing. It performs operations such as resizing, normalization, and noise removal.
This step ensures that all images have consistent dimensions and quality, which helps improve the efficiency and accuracy of the model.

- **Human Parsing Layer**

The human parsing layer divides the person image into different regions such as upper body, arms, and background.
This segmentation helps the system understand where the garment should be placed. It plays a crucial role in achieving proper alignment between the cloth and the body.

- **Feature Extraction Layer**

The feature extraction layer identifies important patterns and features from both the person image and the clothing image.
These features include texture, shape, and structural details, which are necessary for understanding how the garment should fit on the body.

- **Alignment Layer**

The alignment layer matches the garment with the corresponding body regions.
It ensures that the clothing image is positioned correctly according to the body structure. Proper alignment helps in reducing distortion in the final output.

- **Flow Estimation Layer**

The flow estimation layer is one of the most important components of the system. It calculates the dense pixel-level transformation required to deform the garment.
This layer helps in adapting the cloth according to different body poses and shapes, ensuring realistic fitting.

- **Rendering Layer**

The rendering layer generates the final try-on image by combining the warped garment with the person image.
It ensures that the output image looks natural by maintaining color consistency, texture details, and smooth blending.

- **Integration Layer**

The integration layer connects all the modules of the system and ensures smooth data flow between them.
It coordinates the functioning of each layer and manages the overall processing pipeline.

- **Output Layer**

The output layer displays the final generated image to the user.
It ensures that the result is presented clearly and can be viewed or downloaded by the user.

- **User Interface Layer**

The user interface layer provides interaction between the user and the system.
It allows users to upload images, view results, and interact with the application easily through a web interface.

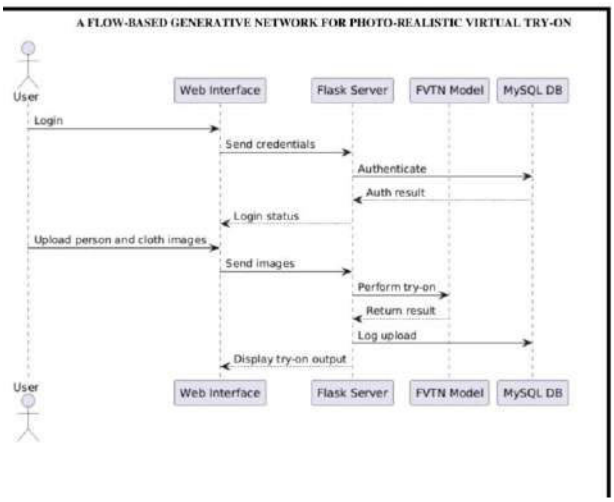

Fig. 2. System Architecture

4. Results and Discussion

I. Experimental Results

The results of the proposed virtual try-on system demonstrate its effectiveness in generating realistic and high-quality images. The system was tested using the VITON dataset, which contains a variety of person images and clothing items with different poses and styles.
The performance of the model is evaluated based on visual quality and accuracy of garment alignment. The generated images show significant improvement compared to existing methods, especially in handling pose variations and preserving cloth texture.

The results obtained from the system are discussed below:

- **Improved Garment Alignment**
The proposed model aligns the garment accurately with the body regions. The clothing fits naturally according to the shape and posture of the person, reducing misplacement issues.

- **Better Handling of Pose Variations**
The system effectively adapts to different human poses. Even with complex body positions, the garment is adjusted properly without major distortions.

- **Preservation of Texture Details**
The texture and patterns of the clothing are well preserved in the output images. This ensures that the visual appearance of the garment remains realistic.

- **Reduced Image Distortion**
Unlike existing systems, the proposed model minimizes distortion in the clothing. The output images appear smooth and natural.

- **Realistic Output Generation**
The final try-on images closely resemble real-world scenarios. The blending between the garment and the body is seamless.

- **Efficient Flow-Based Warping**
The use of flow estimation helps in accurate deformation of garments. This improves the overall fitting of the clothing.

- **Performance Evaluation using MSE**
The model is evaluated using Mean Square Error (MSE), which measures the difference between generated and actual images. Lower MSE indicates better performance.

- **Comparison with Existing Systems**
When compared to traditional and earlier deep learning methods, the proposed system produces more realistic and visually appealing results.

- **User Experience Improvement**
The system enhances user experience by allowing users to visualize clothing before purchasing, reducing uncertainty.

- **Limitations Observed**

Some minor limitations include slight inaccuracies in extreme poses and dependency on dataset quality.

.

Fig.3. Dashboard of the platform

5. Conclusion

In this paper, a Flow-Based Virtual Try-On Network has been proposed to generate realistic and high-quality try-on images by effectively aligning and warping garments onto a person's body. The system addresses the major limitations of existing virtual try-on methods, such as poor alignment, texture distortion, and inability to handle pose variations.

The proposed approach utilizes deep learning techniques to model complex transformations between clothing and human body images. By integrating modules such as Parsing Alignment Module (PAM), Flow Estimation Module (FEM), and Fusion and Rendering Module (FRM), the system is able to produce visually appealing and realistic outputs.

The experimental results demonstrate that the proposed system performs significantly better compared to traditional and existing deep learning-based approaches. It achieves improved garment fitting, better preservation of texture details, and reduced distortion in the generated images. The use of flow-based warping plays a key role in adapting the clothing to different body shapes and poses, thereby enhancing realism.

The key contributions of the proposed system are:

- Development of a flow-based model for realistic garment deformation.
- Improved alignment between clothing and human body.
- Effective handling of pose variations and occlusions.
- Preservation of fine texture and visual details.
- Generation of high-quality and realistic try-on images.

Despite these advantages, certain limitations are observed. The system may produce minor artifacts in cases of extreme poses or complex backgrounds. Additionally, the performance of the model depends on the quality and diversity of the training dataset.

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)*, 2018, pp. 7543–7552.

[2] B. Wang, H. Zheng, X. Liang, Y. Chen, L. Lin, and M. Yang, "Toward Characteristic-Preserving Image-based Virtual Try-On Network," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 589–604.

[3] H. Yang, R. Yu, X. Han, and Y. Zheng, "Towards Photo-Realistic Virtual Try-On by Adaptively Generating-Preserving Image Content," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[4] C. Ge, Y. Song, Y. Ge, and P. Luo, "Parser-Free Virtual Try-On via Distilling Appearance Flows," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[5] X. Han, Z. Huang, Z. Wu, Z. Wu, R. Yu, and L. S. Davis, "ClothFlow: A Flow-Based Model for Clothed Person Generation," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2019.