Virtual Try on System for Women's Clothing

Priyanka N. Sandhan[1], Pooja M. Borade[2], Pratiksha S. Badhan[3] and Hemangi J. Deokar [4]
[1,2,3,4] *Department of Artificial Intelligence& Data Science Engineering, SNJB's Late Sau. Kantabai Bhavarlalji Jain College of Engineering, Chandwad, Nashik - 423101, Maharashtra, India*

***Abstract*—Online fashion platforms have grown quickly, but many customers still feel unsure when buying clothes because they cannot try them on beforehand. This challenge is even more noticeable with women's ethnic wear, where proper fitting, draping, and overall look play a crucial role in the final appearance. This paper introduces a virtual try-on system that enables users to see how different garments would look on their own images. The system is built using an image processing pipeline and a modular backend developed with FastAPI. It also includes a category-aware approach, allowing it to handle different types of clothing, such as kurtis and lehengas, in a more effective and tailored manner. To make the results look more natural, a post-processing step is used to preserve the user's facial features and body structure. The system is designed to run efficiently without the need for high-end hardware. Experimental results indicate that it can generate visually realistic outputs with proper garment alignment. Overall, the system proves to be practical, scalable, and suitable for real-world use in online fashion platforms.**

***Index Terms*—Virtual Try-On, Diffusion Models, Computer Vision, FastAPI, Image Processing, Deep Learning, Ethnic Wear, Realistic Visualizations, Garment Fit, E-commerce**

I. INTRODUCTION

The growth of e-commerce has transformed the fashion industry by allowing users to browse and purchase clothing from anywhere. However, one major limitation of online shopping is the inability to try garments before purchasing, leading to uncertainty regarding fit, appearance, and overall satisfaction.

This issue becomes more challenging in the case of ethnic wear such as kurtis and lehengas. These garments have complex structures, loose fittings, and varied draping styles, making it difficult for users to imagine how they would look when worn. Traditional product images and size charts fail to provide sufficient visualization.

Virtual try-on systems have been introduced to address this problem by generating images of users wearing selected garments. Early approaches relied on basic image transformation techniques, but these often resulted in poor alignment and unrealistic outputs.

Recent advancements in image generation techniques have improved the quality of such systems. However, many existing solutions focus mainly on visual output and do not address system scalability, backend integration, or real-world deployment.

This paper presents a complete virtual try-on system that integrates an advanced image processing pipeline with a structured backend. The system introduces a category-aware approach to handle different garment types effectively, especially ethnic wear. It also includes a post-processing module to enhance realism and preserve identity, making the system suitable for practical applications.

II.MOTIVATION

The increasing demand for personalized and interactive online shopping experiences has highlighted the need for better visualization tools in fashion platforms. A major challenge faced by users is the inability to see how a garment will appear when worn. This issue is more critical for ethnic wear, where garment structure, draping, and length significantly affect the overall appearance. Existing systems are mostly designed for standard clothing and often fail to produce realistic results for such garments. Another limitation in existing approaches is the lack of proper system design for real-world deployment. Many solutions focus only on visual output without considering scalability, integration, and usability. This project is also motivated by a real-world application, where the system will be used by a fashion designer to visualize garments digitally. This reduces the need for physical trials, saves time, and improves design decisions. The aim of this work is to develop a system that is not only visually accurate but also practical, scalable, and suitable for real-world use.

III. LITERATURE REVIEW

Virtual try-on systems have gained significant attention in recent years as a solution to improve online fashion shopping experiences. These systems aim to generate images of users wearing selected garments, helping customers make better purchasing decisions.

A. Early Image-Based Methods

Initial approaches to virtual try-on relied on traditional image processing techniques such as segmentation, geometric transformation, and image warping. These methods attempted to align garment images with the human body using pose estimation and predefined mapping rules. While these approaches were computationally efficient, they often struggled with maintaining proper alignment and failed to preserve fine details such as fabric texture and folds. As a result, the generated outputs appeared unrealistic, especially for loose or layered garments.

B. Two-Stage Generation Approaches

Later methods introduced structured pipelines that divided the process into two main stages:

1. Garment alignment (warping)
2. Image refinement (generation)

These approaches improved visual quality compared to earlier techniques. However, they still suffered from several limitations, including:

- Blurred outputs
- Misalignment between garment and body
- Distortion of facial features
- Difficulty handling occlusions (e.g., arms overlapping clothing)

Additionally, these systems often required careful parameter tuning and were not always stable across different inputs.

C. Advanced Image Synthesis Techniques

Recent developments have focused on more advanced image generation strategies that produce higher-quality and more realistic outputs. These methods follow a progressive refinement process that improves image quality step by step, allowing better preservation of textures and structural details.

Such techniques have shown improved performance in:

- Maintaining garment patterns
- Handling different body poses
- Producing smoother and more natural outputs

Despite these improvements, most existing works primarily focus on enhancing visual quality and do not sufficiently address system-level challenges such as scalability, deployment, and integration.

D. Challenges in Ethnic Wear Visualization

A major gap in existing research is the limited focus on ethnic wear. Garments such as kurtis and lehengas differ significantly from standard clothing due to:

- Complex structures
- Loose and flowing designs
- Layered fabric arrangements
- Variable garment lengths

Most existing systems are optimized for Western clothing and fail to handle these characteristics effectively. This often leads to unrealistic outputs, poor fitting, and incorrect garment alignment. This highlights the need for specialized processing strategies that consider garment-specific properties

E. System Design and Deployment Limitations

Another key limitation observed in existing work is the lack of focus on backend architecture and real-world usability. Many systems are developed as research prototypes without considering:

- API-based integration
- Scalability
- Efficient request handling
- Storage and resource management

This makes it difficult to deploy such systems in real-world environments like fashion platforms or design studios

F. Identified Research Gaps

Based on the review of existing methods, the following gaps are identified:

- Limited support for ethnic wear visualization
- Inadequate preservation of user identity
- Poor handling of complex garment structures
- Lack of scalable and modular system design
- High dependency on powerful local hardware

G. Motivation for Proposed Work

To address these challenges, there is a need for a system that:

- Handles different garment categories effectively
- Maintains visual realism and identity
- Uses a structured and scalable backend
- Is suitable for real-world applications

The proposed system is designed to overcome these limitations by combining an advanced image processing pipeline with a modular architecture, specifically optimized for ethnic wear visualization.

IV. PROPOSED METHODOLOGY

The proposed virtual try-on system is designed to generate realistic images of users wearing selected garments, with a special focus on ethnic wear. The system follows a structured, multi-stage pipeline to ensure accuracy, consistency, and visual quality. Each stage of the methodology is described below.

The system's overall workflow has several steps, which are listed below.

A. Identified Research Gaps

The system accepts two primary inputs:

- User Image (I_u): A photograph of the person
- Garment Image (Ig): An image of the selected clothing item

These inputs are uploaded through an API interface and temporarily stored for processing. The system validates the input formats (such as JPEG or PNG) and ensures that the images meet the required resolution and quality standards.

B. Image Preprocessing

Before further processing, both input images undergo a preprocessing stage to ensure uniformity and compatibility. This includes:

- Resizing images to a fixed resolution
- Normalizing pixel values
- Removing unwanted background noise

This step improves the consistency of the data and enhances the overall quality of the final output.

C. Category-Aware Pipeline Selection

A key feature of the system is its category-aware design. Based on the type of garment provided, the system selects an appropriate processing pipeline.

Let:

$C \in$ {kurti, lehenga, shirt, pant}

Each category has specific configurations that consider:

- Garment length
- Structural properties
- Draping style

This ensures better alignment and more realistic representation of different clothing types, especially ethnic wear.

D. Garment Mapping and Image Generation

In this stage, the system maps the garment image onto the user's body to generate a realistic try-on output.

The process can be represented as:

$I\_gen = F(I_u, Ig, C)$

Where:

- I_gen = Generated try-on image
- F = Processing function of the system

this stage ensures:

- Accurate placement of the garment
- Preservation of texture and patterns
- Adaptation to body pose and structure

The processing is performed using a cloud-based system, which allows efficient execution without requiring high-end local hardware.

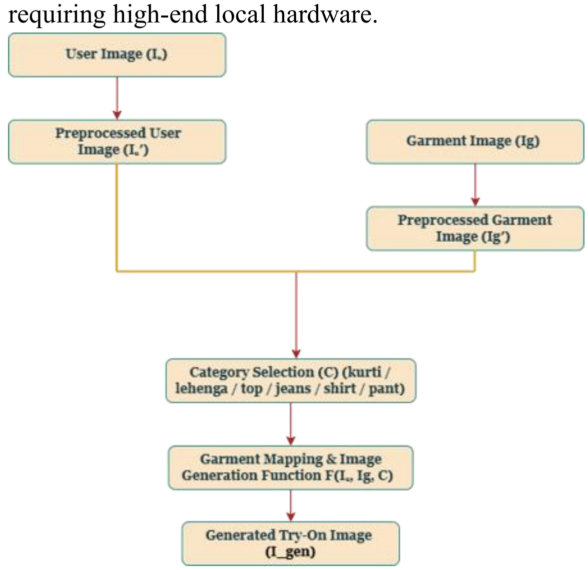

Figure 1: Garment Mapping and Image Generation Process for Virtual Try-On System

E. Post-Processing and Identity Preservation

After generating the try-on image, a post-processing step is applied to enhance visual quality and realism.

This stage focuses on:

- Preserving the user's facial identity
- Reducing distortions and visual artifacts
- Improving sharpness and color consistency

The refined output can be expressed as:

$$I\_final = P(I\_gen)$$

Where:

- P = post-processing function

This step plays a crucial role in maintaining realism and user trust.

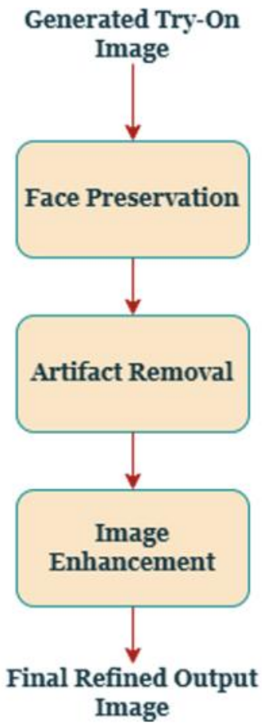

Figure 2: Post-processing pipeline for identity preservation and image quality enhancement.

F. Backend Integration and Workflow Management

The system is supported by a FastAPI-based backend that manages the complete workflow.

The backend is responsible for:

- Handling file uploads and validation
- Routing requests to appropriate pipelines
- Managing communication with processing modules
- Returning the final output to the user

Additionally, background tasks are implemented to:

- Remove temporary files
- Optimize storage usage
- Maintain system efficiency

G. Output Generation and Response

The final output of the system is a realistic image of the user wearing the selected garment.

Along with the generated image, the system returns:

- Generation Status
- Result Image Path
- Job ID (for tracking)

This output can be used by both end users and designers for visualization and decision-making purposes. The overall methodology of the proposed virtual try-on system integrates preprocessing, category-aware routing, diffusion-based generation, and backend processing to produce realistic results.

Table 1. Summary of the proposed virtual try-on system workflow.

Stage	Description
Input Acquisition	Upload user image and garment image
Preprocessing	Resize and normalize images
Category Selection	Identify garment type (kurti/lehenga)
Image Processing	Apply image synthesis pipeline for try-on
Post-Processing	Enhance output and preserve face identity
Backend Handling	FastAPI manages request and response
Output Generation	Return final try-on image

V. PROPOSED RESULTS

The proposed virtual try-on system was evaluated based on the quality, realism, and consistency of the generated outputs. The system successfully produces visually convincing images in which the selected garment is accurately mapped onto the user's body while maintaining natural appearance.

The results demonstrate that the system effectively handles different garment categories, particularly ethnic wear such as kurtis and lehengas, by preserving their structural properties and draping patterns. The inclusion of a category-aware pipeline significantly improves alignment and overall visual consistency.

Step 1: Input Processing and Preparation

The system begins by processing the input images to ensure consistency and compatibility.

Let:

- I_u = User image
- Ig = Garment image

Both images are resized and normalized to a standard format:

$$I_u' = N(I_u), \quad Ig' = N(Ig)$$

Where:

- N( ) = Normalization and preprocessing function

This step ensures that all inputs are uniform, which improves the stability and quality of the generated output.

Step 2: Category-Aware Routing

The system identifies the garment category and routes the input through a specialized processing pipeline.

Let:

$C \in \{kurti, lehenga, shirt, pant\}$

Each category uses a dedicated configuration to handle:

- Garment length
- Structural characteristics
- Draping variations

This improves the accuracy of garment placement and ensures better visual realism.

Step 3: Garment Mapping and Image Generation

The processed inputs are then used to generate the try-on image through the system pipeline:

$$I\_gen = F(I_u', Ig', C)$$

Where:

- I_gen = Generated try-on image
- F = Image synthesis function

This stage ensures:

- Proper garment alignment
- Preservation of texture and design patterns
- Adaptation to user body pose

The system produces outputs that closely resemble real-world appearance, even for complex garments.

Step 4: Post-Processing and Enhancement

To further improve the output quality, a post-processing step is applied:

$$I\_final = P(I\_gen)$$

Where:

- P = Enhancement function

Key improvements include:

- Preservation of facial identity
- Reduction of artifacts and distortions
- Improved sharpness and color consistency

This stage plays a crucial role in making the output appear natural and visually appealing.

Step 5: Output Packaging and Response

The final processed image is returned through the API along with metadata:

- Generation status
- Result image path
- Job ID

This allows users and designers to easily access and track the generated outputs.

Step 6: Performance Evaluation

The system was evaluated based on multiple performance parameters:

- Visual Quality: The generated images appear realistic and clear
- Alignment Accuracy: Garments are properly fitted to the user's body
- Identity Preservation: Facial features remain consistent
- Ethnic Wear Handling: Complex garments are handled effectively
- Processing Time: Approximately 10–20 seconds per request

Table 2: Observed Results of Proposed System

Parameter	Result
Visual Quality	High
Alignment Accuracy	Accurate
Identity Preservation	Good
Ethnic Wear Handling	Effective
Processing Time	10–20 seconds
System Stability	Stable

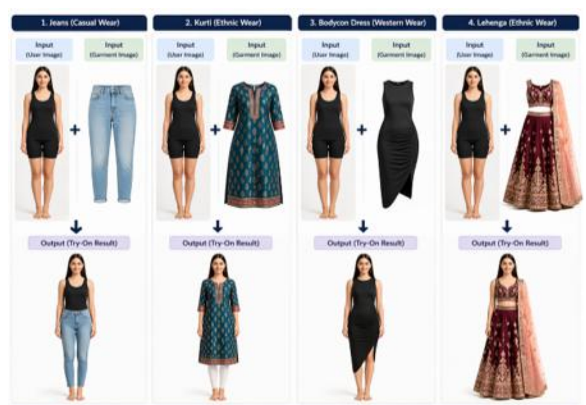

Figure 3: Sample outputs for different garment categories

V. DISCUSSION

The experimental results demonstrate that the proposed system is capable of generating realistic virtual try-on outputs through a well-structured processing pipeline. The integration of preprocessing, category-aware routing, and post-processing plays a significant role in improving the overall quality and consistency of the generated images.

A. Performance and Realism

The adopted image generation approach contributes significantly to enhancing visual quality by preserving garment texture and structural details. Compared to traditional techniques, the outputs show improved alignment, reduced distortions, and better adaptation to user body pose. The system performs particularly well for ethnic garments such as kurtis and lehengas, which typically involve complex draping patterns and layered structures. The category-aware pipeline ensures that garment-specific characteristics are handled effectively, resulting in more natural and visually accurate outputs

B. System Design and Scalability

The backend architecture, implemented using FastAPI, ensures efficient request handling and smooth workflow management. The modular design allows easy integration of different processing components and supports scalability for real-world deployment. Additionally, the use of cloud-based processing reduces dependency on high-end local hardware, making the system lightweight and accessible. This design choice also enables better resource management and flexibility in handling multiple requests.

C. Practical Usability

The system demonstrates strong potential for real-world applications in areas such as fashion design and online retail. Designers can use the system to visualize garments on different body types without requiring physical trials, while users can preview outfits before making purchasing decisions. This improves overall user experience, increases confidence in product selection, and reduces uncertainty commonly associated with online shopping.

D. Limitations

Despite its effectiveness, the system has certain limitations:

- Output quality depends on the clarity and pose of the input image
- Complex backgrounds and occlusions may affect accuracy
- Dependency on external services may introduce response delays
- Limited garment categories may restrict flexibility

E. Future Implications

Future improvements can further enhance the system by:

- Enabling real-time virtual try-on capabilities
- Expanding support for additional garment categories
- Integrating with mobile and web-based applications
- Improving identity preservation techniques
- Incorporating 3D garment visualization for enhanced realism

VI. CONCLUSION

In this work, a virtual try-on system was developed to generate realistic images of users wearing selected garments. The system integrates an advanced image processing pipeline with a scalable backend architecture to produce high-quality outputs. A category-based approach is used to handle different types of clothing effectively, particularly ethnic wear such as kurtis and lehengas. The inclusion of a post-processing stage ensures preservation of facial identity and enhances overall visual realism. The results indicate that the system is practical and suitable for real-world applications, especially in fashion design and online shopping. By allowing users to visualize garments before purchase, the system improves decision-making and reduces uncertainty. Furthermore, the proposed approach demonstrates how modern computational techniques can enhance digital fashion experiences. By providing realistic visual representations, the system increases user confidence and has the potential to reduce product return rates in online shopping. This approach benefits both consumers and designers and opens opportunities for future applications in virtual fashion platforms and interactive digital environments. Future work may focus on real-time processing, expanding garment categories, and improving system efficiency to deliver a more seamless user experience.

Table 3: Final Performance Summary of Proposed Virtual Try-On System

Parameter	Result
Visual Quality	High
Alignment	Accurate
Identity	Preserved
Ethnic Wear	Supported
Time	10–20 sec

REFERENCES

[1] Wang, B., H. Zheng, X. Liang, Y. Chen, and L. Lin, "Toward characteristic-preserving image-based virtual try-on network," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 589–604.

[2] Yang, H., R. Zhang, X. Guo, W. Liu, W. Zuo, and P. Luo, "Towards photo-realistic virtual try-on by adaptively generating-preserving image content," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020, pp. 7847–7856.

[3] Yu, R., X. Wang, and X. Xie, "VTNFP: An image-based virtual try-on network with body and clothing feature preservation," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2019, pp. 10511–10520.

[4] Ge, Y., et al., "Parser-free virtual try-on via distilling appearance flows," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2021, pp. 8485–8493.

[5] Choi, S., S. Park, M. Lee, and J. Choo, "VITON-HD: High-resolution virtual try-on via misalignment-aware normalization," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2021, pp. 14126–14135.

[6] He, S., Y.-Z. Song, and T. Xiang, "Style-based global appearance flow for virtual try-on," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2022, pp. 3470–3479.

[7] Lee, S., et al., "High-resolution virtual try-on with misalignment and occlusion handling," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2022, pp. 204–219.

[8] Bookstein, F., "Principal warps: Thin-plate splines and the decomposition of deformations," IEEE Trans. Pattern Anal. Mach. Intell., vol. 11, no. 6, pp. 567–585, 1989.

[9] Zhou, T., et al., "View synthesis by appearance flow," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2016, pp. 286–301.

[10] Karras, T., et al., "Analyzing and improving the image quality of StyleGAN," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020, pp. 8107–8116.

[11] Baraff, D., and A. Witkin, "Large steps in cloth simulation," in Proc. SIGGRAPH, 1998, pp. 43–54.

[12] Bridson, R., R. Fedkiw, and J. Anderson, "Robust treatment of collisions, contact and friction for cloth animation," ACM Trans. Graph., vol. 21, no. 3, pp. 594–603, 2002.

[13] Bridson, R., et al., "Simulation of clothing with folds and wrinkles," in Proc. ACM SIGGRAPH/Eurographics Symp. Comput. Animation, 2003, pp. 28–36.

[14] Selle, A., et al., "Robust high-resolution cloth simulation," IEEE Trans. Vis. Comput. Graph., vol. 15, no. 2, pp. 339–350, 2009.

[15] Guan, P., et al., "DRAPE: Dressing any person," ACM Trans. Graph., vol. 31, no. 4, Art. no. 35, 2012.

[16] Hahn, F., et al., "Subspace clothing simulation using adaptive bases," ACM Trans. Graph., vol. 33, no. 4, Art. no. 105, 2014.

[17] Pons-Moll, G., et al., "ClothCap: Seamless 4D clothing capture and retargeting," ACM Trans. Graph., vol. 36, no. 4, Art. no. 73, 2017.

[18] Stoll, C., et al., "Video-based reconstruction of animatable human characters," ACM Trans. Graph., vol. 29, no. 6, Art. no. 151, 2010.

[19] Zhao, F., et al., "M3D-VTON: A 3D virtual try-on network," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2021.

[20] Bhatnagar, B. L., et al., "Multi-garment net: Learning to dress 3D people from images," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2019.

[21] Patel, C., et al., "TailorNet: Predicting clothing in 3D," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020.

[22] Mir, A., et al., "Learning to transfer texture from clothing images to 3D humans," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020.

[23] Zhu, H., et al., "DeepFashion3D: Dataset for 3D garment reconstruction," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2020.

[24] Corona, E., et al., "SMPlicit: Topology-aware generative model," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2021.

[25] Loper, M., et al., "SMPL: A skinned multi-person linear model," ACM Trans. Graph., vol. 34, no. 6, Art. no. 248, 2015.

[26] Park, J. J., et al., "DeepSDF: Learning signed distance functions for shape representation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019.

[27] Sitzmann, V., et al., “Implicit neural representations with periodic activation functions,” arXiv preprint arXiv:2006.09661, 2020.
[28] Tewari, A., et al., “State of the art on neural rendering,” Computer Graphics Forum, vol. 39, no. 2, pp. 701–727, 2020.
[29] Sitzmann, V., et al., “Scene representation networks: Continuous 3D-structure-aware neural scene representations,” arXiv preprint arXiv:1906.01618, 2019.
[30] Saito, S., et al., “PIFu: Pixel-aligned implicit function for high-resolution clothed human digitization,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2019.