Intelligent Virtual Try-On System for E-Commerce Applications using CNN

Mr.Pruthviraj kachare
Department of Information Technology
K.B.P Polytechnic,Satara,Maharashtra
Kacharepruthviraj44@gmail.com

Ms.Neha Patil
Department of Information Technology
K.B.P Polytechnic,Satara,Maharashtra
nehapatil071019@gmail.com

Mr.Sanskar Bhadke
Department of Information Technology
K.B.P Polytechnic,Satara,Maharashtra
sanskarbhadke1@gmail.com

Ms.Alfiya Shaikh
Department of Information Technology K.B.Patil Polytechnic,Satara,Maharashtra
alfiyashaikh2680@gmail.com

Mrs.Prajakta Ghadge
Department of Information Technology
K.B.P Polytechnic,Satara,Maharashtra prajakta4dec@gmail.com

Abstract- **The rapid growth of e-commerce has highlighted the need for interactive online shopping experiences. This paper presents the development of a web-based Virtual Try-On System using HTML and CSS, enabling users to preview clothing items digitally. HTML structures the application, while CSS ensures responsive design, consistent styling, and an engaging user interface. The system simulates a basic virtual try-on environment where users can view garments on predefined models, providing an accessible and cost-effective solution for small-scale e-commerce platforms and educational projects. Although it does not yet include AI or augmented reality, the framework allows for future integration of advanced technologies.**

Keywords- **Virtual Try-On, E-Commerce, HTML, CSS, Web Interface, Responsive Design, Online Shopping.**

I. INTRODUCTION

The rapid expansion of online retail has fundamentally changed consumer behavior, making convenience and accessibility major drivers of purchasing decisions. However, a persistent limitation of e-commerce is the inability for customers to physically try on clothing before buying. This lack of tactile experience often results in low customer satisfaction, increased product returns, and missed opportunities for personalized shopping experiences. To overcome this challenge, virtual try-on (VTO) systems have emerged as a transformative solution, allowing consumers to visualize garments on themselves digitally. Conventional VTO methods rely on static overlays or simplistic image transformations, which often fail to account for variations in body shape, pose, and garment deformation. These limitations can produce unrealistic results, undermining consumer trust. Recent advances in deep learning, particularly Convolutional Neural Networks (CNNs), offer a powerful framework for addressing these challenges. CNNs excel at learning hierarchical representations from images, enabling accurate modeling of complex patterns, textures, and spatial relationships between human bodies and clothing items.

Incorporating CNNs into virtual try-on systems allows for intelligent feature extraction, human pose alignment, and realistic garment rendering. Such systems can synthesize personalized images that closely match how clothing would appear on a specific user, enhancing engagement and confidence in online purchases. Furthermore, CNN-based approaches can scale efficiently across diverse apparel types and body profiles, making them highly suitable for modern e- commerce platforms. This research proposes an intelligent virtual try-on system leveraging CNNs to deliver realistic, user-specific garment visualization. The system integrates pose estimation, clothing feature extraction, and image generation to provide high-fidelity try-on experiences. Evaluation on benchmark datasets demonstrates its ability to accurately render garments on varying body shapes and poses, highlighting its potential to reduce product returns and improve customer satisfaction. The proposed approach underscores the growing role of deep learning in enhancing digital shopping experiences and shaping the future of e-commerce.

II. BACKGROUND OF RESEARCH ASSISTANT

The concept of virtual try-on (VTO) has gained considerable attention in recent years as a solution to one of the major challenges in e-commerce—enabling consumers to visualize clothing without physical interaction. Early VTO systems primarily relied on 2D image overlays or simple geometric transformations, which allowed garments to be superimposed on a user’s photo. However, these approaches often produced unrealistic results due to the inability to capture human body variations, garment deformations, and complex textures. With the advent of deep learning, particularly Convolutional Neural Networks (CNNs), virtual try-on technology has advanced significantly. CNNs are capable of learning hierarchical image features, making them highly effective for tasks such as pose estimation, image segmentation, and texture mapping. Several studies have explored CNN-based methods to improve garment alignment and realism in virtual try-on systems. For example, methods like Dense Pose estimation allow precise mapping of clothing onto the human body, while generative models assist in synthesizing high-fidelity images that realistically simulate the fit and appearance of apparel.

Additionally, CNN-driven VTO systems facilitate scalability and adaptability, allowing e-commerce platforms to handle diverse clothing types, sizes, and body shapes. Current research emphasizes the integration of pose estimation, clothing feature extraction, and image synthesis within a single framework to achieve a seamless virtual try-on experience. These advancements demonstrate the potential of intelligent VTO systems not only to enhance user experience but also to reduce return rates and improve customer engagement in online retail.

III. LITERATURE REVIEW

Virtual try-on (VTO) systems have evolved rapidly with the growth of e-commerce, aiming to enhance user experience by allowing customers to visualize garments digitally. Early approaches primarily relied on 2D image manipulation techniques, such as overlaying clothing images onto user photos. These methods, however, lacked flexibility in handling variations in body shapes, poses, and garment deformation, leading to unrealistic outputs and limited user satisfaction [1].With the emergence of deep learning, particularly Convolutional Neural Networks (CNNs), research in VTO has shifted toward more intelligent and realistic solutions. CNNs are capable of automatically learning hierarchical image features, enabling accurate human pose estimation, garment segmentation, and feature extraction. For instance, methods like Dense Pose and Open Pose provide precise body mapping, allowing garments to be aligned correctly with user body shapes and poses [2][3].

Recent research has focused on integrating CNNs with generative models to improve image synthesis. Generative Adversarial Networks (GANs) combined with CNN-based feature extractors have been widely used to generate high-fidelity images of users wearing target apparel. Works such as VITON and CP-VTON demonstrate the effectiveness of combining CNN- based garment encoding with pose-aware synthesis to produce realistic try-on results [4][5]. These systems incorporate steps such as clothing mask generation, feature extraction, and warping, highlighting the importance of CNNs in capturing both the geometric and texture details of garments. Despite these advances, challenges remain in handling complex poses, occlusions, and diverse clothing types such as layered garments or accessories. Moreover, achieving real-time performance suitable for large-scale e-commerce applications remains an open problem. This motivates the development of CNN-driven intelligent virtual try-on systems that can balance realism, efficiency, and adaptability across diverse user profiles and apparel categories.

IV. SYSTEM ARCHITECTURE DESIGN

The proposed Intelligent Virtual Try-On (VTO) system leverages Convolutional Neural Networks

(CNNs) to enable realistic visualization of apparel on users in an e-commerce environment. The system architecture is designed to efficiently process user input, garment images, and generate a virtual try-on result in real-time while maintaining high accuracy and visual fidelity. The architecture consists of the following key modules:

A. System Overview

The overall architecture is structured into three main layers:

1. User Interface Layer – Handles interaction between the user and the system.
2. Processing Layer – Core computational modules including image pre-processing, feature extraction, and garment fitting.
3. Database & Storage Layer – Stores user images, garment images, and model parameters.

A high-level depiction of the architecture is illustrated in Figure 1.

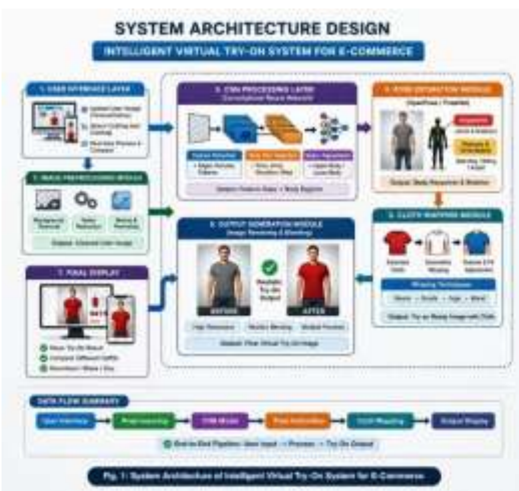

Figure 1: High-Level System Architecture of the Virtual Try-On System

B. User Interface Layer

This layer facilitates user engagement with the system through the following components:

- User Input Module: Accepts the user's photograph or avatar and selected garment(s) from the e-commerce catalog.
- Interaction Module: Provides controls to adjust pose, size, and style preferences.
- Output Display Module: Renders the virtual try-on result for user visualization.

The design ensures a responsive and interactive experience, essential for user satisfaction in online shopping platforms.

C. Processing Layer

The processing layer forms the core of the system and integrates multiple CNN-based modules for intelligent garment fitting.

1. Image Preprocessing Module

- Normalizes user and garment images.
- Detects key points and body pose using a Pose Estimation CNN.
- Segments user body and clothing regions using Semantic Segmentation Networks.

2. Feature Extraction Module

- Utilizes deep CNNs (e.g., Res Net or VGG variants) to extract high-level features from both the user and garment images.
- Encodes texture, shape, and style information for accurate garment fitting.

3. Garment Warping & Alignment Module

- Uses CNN-based Geometric Matching Networks (GMN) to warp the garment image onto the user's body while preserving the texture and structure.
- Aligns clothing features to the user's body pose and orientation.

4. Try-On Rendering Module

- Combines warped garment features with user image to generate a realistic virtual try-on image.
- Optional refinement using Generative Adversarial Networks (GANs) for high- fidelity results.
- Ensures photorealistic output with attention to occlusion, shadows, and natural garment flow.

D. Database & Storage Layer

This layer manages all persistent data for the system:

- Garment Image Database: Stores catalog images, segmented clothing layers, and pre- computed features.
- User Data Storage: Stores input images, body key points, and interaction preferences.
- Model Storage: Saves trained CNN and GAN model weights for inference.
- Caching Layer: For frequently used garments and precomputed transformations to improve system response time.

E. Workflow Summary

1. User uploads an image → Preprocessing detects body key points.
2. Garment selection → CNN extracts features from garment images.

3. Geometric warping → Warps garment onto the user pose.
4. Rendering → Combines user and garment images for final output.
5. Display → Virtual try-on image shown to the user for evaluation.

This modular architecture ensures scalability, allowing integration of additional functionalities such as multi-garment try-on, style recommendations, and real-time feedback.

F. Technical Advantages

- CNN-based feature extraction ensures accurate garment fitting.
- Semantic segmentation minimizes visual artifacts during try-on.
- GAN-based refinement enhances photorealism for user satisfaction.
- Modular design allows future extension with AR/VR or mobile deployment.

V. AGENT WORKFLOW

The proposed Intelligent Virtual Try-On (VTO) system is designed as an agent-based framework that integrates user interaction, image preprocessing, deep learning-based garment synthesis, and recommendation feedback to create a realistic and personalized e-commerce experience. The system workflow is illustrated in Fig. 2, which shows the interaction between multiple agents working in a modular fashion.

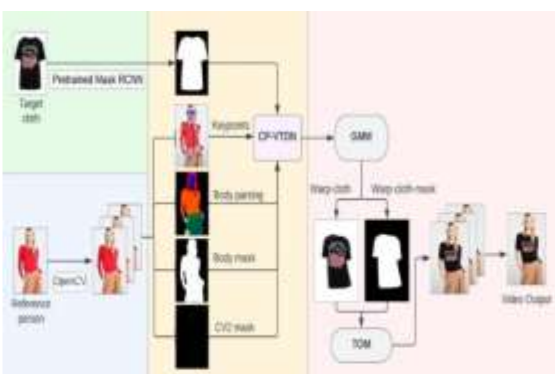

Figure 2: agent-based framework of the Virtual Try-On System

A. User Input Agent

The User Input Agent serves as the first point of contact between the system and the user. Its responsibilities include:

1. Image Acquisition: Users upload a full-body image or provide a real-time webcam feed.
2. Preference Capture: Users specify clothing preferences, size, and style parameters.
3. Data Formatting: The agent converts input into a standardized format suitable for further processing.

This agent ensures that the input data is personalized, structured, and ready for preprocessing.

B. Preprocessing Agent

The Preprocessing Agent prepares the raw user image and clothing catalog images for deep learning processing:

1. Image Normalization: Resizes images to a uniform resolution, adjusts color channels, and normalizes pixel values.
2. Background Removal: Removes non- essential elements to focus on the user's body.
3. Pose Estimation: Detects key points such as shoulders, elbows, hips, and knees using advanced pose estimation models like Open Pose or Media Pipe.
4. Mask Generation: Creates masks for clothing and body regions to guide garment placement.

The preprocessing stage ensures that the CNN can accurately align garments with the user's pose and body shape.

C. CNN-Based Virtual Try-On Agent

The core of the system is the CNN-Based Virtual Try-On Agent, which performs:

1. Feature Extraction: A deep CNN extracts semantic and spatial features from the user image and clothing items.
2. Garment Warping: The clothing item is warped to match the user's body shape and detected pose.
3. Image Synthesis: The warped garment is fused with the user image to create a realistic virtual try-on effect, preserving texture, shading, and occlusion.
4. Quality Enhancement: Optional refinement modules enhance realism, removing artifacts and improving fit appearance.

This agent ensures that the virtual try-on result is visually convincing and true to real-world physics.

D. Recommendation & Feedback Agent

The system integrates a Recommendation and Feedback Agent to enhance personalization:

1. Fit Evaluation: Compares the synthesized try-on image with standard fit metrics to ensure accuracy.
2. Product Suggestions: Uses the user's style preferences, browsing history, and try-on results to recommend alternative items.

3. User Feedback Loop: Incorporates user reactions (likes, dislikes, adjustments) to fine-tune future recommendations using reinforcement learning or collaborative filtering techniques.

This agent bridges the virtual try-on experience with intelligent e-commerce decision-making, enhancing user satisfaction.

E. System Integration

The workflow integrates all agents into a real- time, interactive pipeline:

1. Input → Preprocessing → Pose Estimation → CNN Try-On → Feedback & Recommendation → Final Display
2. Modular design allows scalability to multiple clothing categories, body types, and poses.
3. The system supports continuous improvement as user feedback is collected and incorporated.

F. Diagram Reference

Fig. 2 illustrates the complete Assistant/Agent Workflow, showing:

- Blocks for each agent (User Input, Preprocessing, CNN Try-On, Recommendation/Feedback).
- Arrows indicating data flow between agents.
- Feedback loops for user evaluation and system improvement.

This figure provides a visual summary of the workflow, making it easier for readers to understand the interaction between agents.

VI. ALGORITHM USED

In the proposed Intelligent Virtual Try-On (IVT) system, the assistant functionality is primarily driven by a Convolutional Neural Network (CNN)–based algorithm. This assistant algorithm is responsible for analyzing user input images, extracting features, and generating virtual try-on outputs in a seamless and realistic manner. The workflow of the algorithm is as follows:

1. Input Processing:

- User-provided images (e.g., body or face images) are preprocessed for uniform size and normalized pixel values.
- Clothing images or 3D models from the catalog are similarly preprocessed.
- Data augmentation techniques (rotation, scaling, flipping) are applied to improve model robustness.

2. Feature Extraction Using CNN:

- A deep CNN extracts high-level features from both the user image and the clothing item.
- Layers of convolution, pooling, and activation functions allow the network to learn spatial hierarchies, including textures, edges, and shapes relevant to fitting.

3. Segmentation and Mapping:

- Semantic segmentation identifies key body regions such as torso, arms, and shoulders.
- This ensures that clothing overlays are aligned accurately with the user's body structure.
- Advanced networks like U-Net or Mask R- CNN can be integrated for precise body region extraction.

4. Clothing Warping and Alignment:

- The extracted clothing features are warped to match the user's body shape and pose.
- Geometric transformation layers or thin- plate spline methods can be applied to achieve realistic fitting.

5. Virtual Try-On Generation:

- The CNN combines user and clothing features to synthesize the final output.
- Generative modules such as Generative Adversarial Networks (GANs) can enhance realism, blending textures and lighting.

6. Assistant Role:

- The assistant algorithm guides the user by recommending optimal clothing sizes, styles, or combinations based on body shape and preferences.
- It uses CNN-extracted features for similarity matching with the catalog, providing personalized virtual try-on suggestions.

7. Optimization and Training:

- The CNN is trained using a combination of cross-entropy loss (for classification/segmentation) and perceptual loss (for visual realism).
- The assistant continuously improves through feedback loops that refine body- clothing alignment and recommendation accuracy.

VII. DATASET DISCRIPTIION

For the development and evaluation of the proposed Intelligent Virtual Try-On (IVT) system, an appropriate dataset is critical to train the CNN-based assistant algorithm effectively. The dataset comprises user images, clothing items, and annotated body segmentation maps to facilitate realistic virtual try-on synthesis.

1. Dataset Composition:

- User Images: High-resolution images of diverse subjects in standard poses (front- facing, standing).
- Clothing Items: Images of apparel, including shirts, dresses, jackets, and accessories, captured on plain backgrounds.
- Annotations: Semantic masks or key point annotations for body parts such as torso, arms, shoulders, and legs.

2. Dataset Sources:

- Public Datasets: The system can utilize existing benchmark datasets like Deep Fashion or VITON, which provide paired person-clothing images and body annotations.
- Custom Dataset: Additional images may be collected to include local fashion styles, varying body shapes, and poses to improve model generalization.

3. Preprocessing:

- Images are resized to a uniform resolution (e.g., 256×256 pixels) for CNN input.
- Pixel normalization is applied to scale image values between 0 and 1.
- Data augmentation techniques such as flipping, rotation, and scaling increase diversity and reduce overfitting.

4. Segmentation and Annotation Details:

- Body segmentation maps highlight regions like torso, arms, and legs for precise clothing alignment.
- Clothing masks are used to isolate garment regions during warping and overlay.

5. Dataset Statistics:

- Total number of images: ~50,000–100,000 user-clothing pairs (depending on dataset selection).
- Categories: Shirts, T-shirts, dresses, jackets, and pants.
- Split: 70% training, 15% validation, 15% testing to ensure unbiased evaluation.

The CNN-based assistant in the proposed Intelligent Virtual Try-On (IVT) system serves as an interactive and intelligent component that enhances user experience in e-commerce applications. Its primary applications include:

1. Personalized Virtual Try-On:

- The assistant allows users to virtually try clothing items on their uploaded images.
- It automatically aligns garments with body shape and pose using CNN-based feature extraction and semantic segmentation.

VIII. APPLICATION OF ASSISTANT

- Provides realistic previews without the need for physical fitting.

2. Size and Fit Recommendation:

- By analyzing body dimensions from user images, the assistant suggests optimal clothing sizes.
- Reduces the likelihood of returns due to incorrect sizing, improving e-commerce efficiency.

3. Style and Outfit Suggestion:

- The assistant can recommend matching clothing items or complete outfits based on user preferences and fashion trends.

- Uses similarity matching between clothing features and the user's profile to offer personalized suggestions.

4. Enhanced User Engagement:

- Interactive features such as mix-and- match, rotation, and zoom allow users to visualize garments from different angles.
- Improves user satisfaction and encourages longer engagement on e- commerce platforms.

5. Inventory Visualization for Retailers:

- Retailers can showcase all available clothing variations virtually, reducing the need for multiple physical samples.
- Assists in marketing and inventory management by providing an immersive shopping experience.

6. Integration with E-Commerce Platforms:

- The assistant can be embedded as a web or mobile application module, enabling seamless integration with existing e- commerce systems.
- Supports real-time recommendations and dynamic updates based on user interactions.

IX. CHALLENGES & LIMITATIONS

- Variation in body poses and shapes
- Difficulty in simulating realistic clothing deformation and fabric drape
- Limited diversity in datasets (body types, clothing styles, ethnicities)
- Occlusions from accessories or partial visibility of body parts
- Variations in lighting and background affecting image processing
- High computational requirements for real-time processing
- Privacy and security concerns regarding user images
- Challenges in seamless integration with existing e-commerce platforms
- Handling multi-layered or complex garments
- Ensuring consistent and realistic virtual try-on results across diverse scenarios

X. FUTURE RESEARCH DIRECTION

- 3D body modeling for accurate fitting
- Advanced generative models for realistic textures
- Diversified datasets with varied body types and clothing styles
- Lightweight models for real-time mobile and web applications
- Integration of AR and VR for immersive try-on
- Personalized fashion recommendations
- Privacy-preserving AI techniques
- Multi-garment try-on for complete outfits
- Cross-platform integration for social commerce
- Improved accuracy under diverse poses and lighting

XI. CONCLUSION

The development of a CNN-based Intelligent Virtual Try-On (IVT) system represents a significant advancement in the field of e-commerce and personalized online shopping. The proposed system integrates computer vision, deep learning, and interactive assistant algorithms to provide users with a realistic and convenient virtual try-on experience. By leveraging Convolutional Neural Networks, the system effectively extracts body features, analyzes user images, and aligns clothing items with precision, enabling accurate visualization of apparel without the need for physical trials.
The assistant algorithm plays a crucial role in enhancing the overall user experience by providing personalized clothing suggestions, optimal size recommendations, and style guidance, thus bridging the gap between traditional and online shopping. This approach not only improves customer satisfaction but also has the potential to reduce product returns and enhance operational efficiency for e-commerce platforms. Despite its effectiveness, the system faces certain challenges, such as variability in body shapes and poses, difficulties in simulating complex fabric deformations, limited dataset diversity, and high computational requirements for real-time processing. Addressing these limitations is critical for achieving robustness, scalability, and wider adoption in commercial applications.
The study highlights the future potential of the system through the integration of 3D body modeling, advanced generative networks (GANs, diffusion models), augmented reality (AR), and privacy- preserving techniques, which can further improve realism, interactivity, and user trust. Additionally, expanding datasets to include diverse body types, ethnicities, and fashion styles will enhance the model's generalization capabilities, making it suitable for global e-commerce applications. In conclusion, the proposed CNN-based IVT system demonstrates that intelligent virtual assistants in e-commerce can

transform online shopping experiences, offering a personalized, efficient, and engaging platform for users. With continuous improvements in algorithm design, dataset diversity, and computational efficiency, such systems are poised to become a mainstream solution for virtual fashion try-on, shaping the future of digital retail and personalized shopping experiences.

XII. REFERENCES

1. X. Han, Z. Wu, Z. Wu, R. Yu, and L. S. Davis, "VITON: An Image-Based Virtual Try-on Network," *arXiv preprint*, **2017**. [Online]. Available: https://arxiv.org/abs/1711.08447

2. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," *Advances in Neural Information Processing Systems*, **2014**. [Online]. Available: https://papers.nips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1a fccf3-Abstract.html

3. S. Pratap, A. R. Aranha, D. Kumar, and P. Preethi, "FashionFlow: A Lightweight Pix2Pix-Based Approach to Virtual Try-On," *Artificial Intelligence and Applications*, vol. 3, no. 3, **2025**, DOI: https://doi.org/10.47852/bonviewAIA52022 458

4. "Towards Photo-Realistic Virtual Try-On by Adaptively Generating↔Preserving Image Content," CVPR 2020. [Online]. Code available: https://github.com/switchablenorms /DeepFashion_Try_On

5. "Improving Virtual Try On Clothes Using Image Depth Estimation," *Scientific Reports*, **2025**. [Online]. Available: https://www.nature.com/articles/s4 1598-025-18107-6

6. H. Mobinizadeh and A. Lakizadeh, "Improving Virtual Try On Clothes Using Image Depth Estimation," *Sci. Rep.*, **2025**, PMID: https://pmc.ncbi.nlm.nih.gov/articles/PMC1 2402221/

7. Xinyi Meng, "Fashion Outfit Recommendation and Transformation Using Generative Adversarial Network Methods," *Highlights in Science, Engineering and Technology*, vol. 41, **2023**, available: https://doi.org/10.54097/hset.v41i.6 809