

FASH-TRY – Deep Learning Enabled Virtual Try-On and Image-Based Clothing Recommendation System

S Lakshmi Durga¹, Shashi Chandana B S², T Shreya³, Varshini S⁴, Suraj Kumar B P⁵
^{1,2,3,4}*Department of Computer Science and Engineering, Sir M Visvesvaraya Institute of Technology,
Bengaluru, Karnataka, India*

⁵*Assistant Professor of Department of Computer Science and Engineering,
Sir M. Visvesvaraya Institute of Technology, Bengaluru, Karnataka, India*

Abstract—FASH-TRY is a deep learning–based virtual try-on and clothing recommendation system designed to enhance online shopping through realistic apparel visualization and intelligent product discovery. The system integrates Computer Vision and Image Processing techniques to allow users to upload their images and virtually try on selected garments. Using a multi-stage CP-VTON pipeline consisting of human parsing, pose estimation, and geometric cloth warping the system generates natural and photo-realistic try-on results while preserving body shape, texture, and alignment. Human parsing with a JPP Net-based model segments body regions, and Open Pose extracts key points for accurate cloth fitting. To support apparel discovery, FASH-TRY employs a Content-Based Image Retrieval (CBIR) module. ResNet50 is used to extract deep visual features from garments, and K-Nearest Neighbors (KNN) identifies visually similar items based on color, pattern, and texture. The backend is built using Flask and SQLite3, with OpenCV handling ONNX and Protobuf model integrations. Experimental evaluation demonstrates that FASH-TRY improves user interaction, increases visualization accuracy, and provides efficient similarity-based recommendations. The system offers a scalable and practical solution for modern e-commerce platforms seeking enhanced personalization and virtual try-on capabilities.

I. INTRODUCTION

The rapid expansion of online shopping has transformed the global fashion industry, enabling consumers to explore a wide range of apparel from the comfort of their homes. However, despite improvements in e-commerce systems, a significant gap remains between digital product representation and real-world user experience. Customers frequently struggle to visualize how clothing items would look on

them in terms of fit, appearance, and style, leading to uncertainty, dissatisfaction, and high return rates. To address these limitations, virtual try-on systems and image-based recommendation technologies have emerged as essential tools for improving user experience, personalization, and decision-making in fashion-tech.

FASH-TRY is designed to bridge this gap by combining advanced Deep Learning, Computer Vision, and Content-Based Image Retrieval (CBIR) to deliver a realistic virtual try-on experience along with intelligent apparel recommendations.

1.1 Background and Motivation

Traditional e-commerce platforms showcase clothing items using static images, size charts, and model photos. These approaches often fail to represent how garments would appear on users with different body shapes, poses, and aesthetics. As a result, customers rely on guesswork, leading to purchase hesitation, inaccurate size choices, and returns due to dissatisfaction with fit or style.

In recent years, deep learning advancements have enabled more accurate human parsing, pose estimation, and image synthesis techniques, making virtual try-on systems increasingly feasible. Motivated by the need for a more immersive shopping experience, FASH-TRY aims to simulate real-world dressing by allowing users to visualize themselves wearing selected clothes while also discovering visually similar alternatives.

1.2 Need for a Virtual Try-On System

Although existing fashion applications provide basic product previews, most lack realism, personalization,

and the ability to dynamically adapt to diverse user images. Users commonly face challenges such as:

- Inability to visualize how a garment fits their unique body shape
- Static or unrealistic try-on previews that do not reflect natural alignment
- Limited clothing recommendations based solely on metadata
- Absence of interactive, personalized garment visualization

The need for a reliable virtual try-on platform arises from the demand for realistic apparel visualization, improved purchase confidence, and enhanced user engagement. FASH-TRY addresses these limitations by utilizing deep learning-based cloth warping, accurate body segmentation, and feature-driven apparel retrieval, ensuring a more immersive and informed online shopping experience.

1.3 Overview of the FASH-TRY System

FASH-TRY is a hybrid virtual try-on and apparel recommendation system built as a lightweight web application. The system accepts two inputs: a user image and a selected clothing image. These inputs undergo a multi-stage pipeline involving human parsing, pose estimation, cloth alignment, and final try-on synthesis using the CP-VTON architecture.

Additionally, the system integrates a CBIR module that utilizes ResNet50 for feature extraction and K-Nearest Neighbors (KNN) to identify visually similar garments. Flask serves as the backend, while OpenCV, TensorFlow/Keras, and ONNX models handle deep learning inference. The outcome is a personalized, realistic try-on output accompanied by intelligent recommendations.

1.4 Challenges in Existing E-Commerce Platforms

Modern e-commerce platforms face several persistent challenges, including:

- Lack of realistic garment visualization on diverse body types
- Inability to align clothing accurately with user posture
- Limited recommendation systems relying only on tags or metadata
- High user uncertainty due to poorly visualized product appearance

- Inefficiency in blending the garment naturally with skin and background

These limitations restrict user decision-making and hinder e-commerce growth. FASH-TRY tackles these constraints by applying computer vision techniques to deliver superior realism and relevance.

1.5 Role of Deep Learning and Computer Vision

Deep learning and computer vision form the core of FASH-TRY. Human parsing using a JPPNet-based segmentation model accurately identifies body regions, enabling correct cloth placement. Pose estimation through OpenPose extracts key joint coordinates to align garments with the user's posture. The CP-VTON framework—comprising the Geometric Matching Module (GMM) and Try-On Module (TOM)—warps clothing images and synthesizes the final output with high texture fidelity and natural blending. ResNet50-based feature vectors and KNN similarity search power the recommendation engine, ensuring relevant apparel suggestions. Together, these components enable accurate visualization and efficient similarity retrieval.

1.6 Technical Objectives of the Proposed System

The primary technical objectives of FASH-TRY include:

- To develop a realistic virtual try-on system using deep learning-based image synthesis
- To accurately segment body regions and estimate user pose for cloth alignment
- To use CP-VTON for generating high-quality, natural try-on outputs
- To build a content-based apparel recommendation module using ResNet50 and KNN
- To design a lightweight, secure, and scalable web application using Flask and SQLite3
- To reduce user uncertainty and improve decision-making in online shopping environments

1.7 Significance for Fashion-Tech and E-Commerce

The adoption of virtual try-on technologies has the potential to revolutionize digital fashion retail. By providing realistic garment visualization, FASH-TRY enhances user confidence, reduces product returns, and increases customer engagement. The integrated recommendation module adds further value by helping users explore visually similar clothing options,

improving product discoverability and personalization.

As online shopping continues to dominate the fashion market, systems like FASH-TRY serve as critical tools for creating immersive, user-friendly, and intelligent e-commerce experiences.

II. LITERATURE SURVEY

A comprehensive literature review was conducted to understand existing advancements in virtual try-on technology, deep learning-based garment synthesis, and content-based fashion retrieval systems. The following studies provided foundational insights that shaped the development of FASH-TRY: A Deep Learning-Based Virtual Try-On and Clothing Recommendation System. The reviewed works cover human parsing, pose-guided garment warping, image generation frameworks, and similarity-based fashion retrieval approaches using CNN feature embeddings.

1. IEEE Conference on Computer Vision (2023)

Authors: Wang, Li, and Xu Title: Pose-Guided Virtual Try-On Using Advanced Human Parsing and Warping Networks

Methodology Used: This study introduced a pose-driven virtual try-on framework, emphasizing the importance of accurate human parsing and pose extraction for realistic garment fitting. The authors used OpenPose to extract skeletal keypoints and applied a geometric deformation module to align clothing items with user posture. Their framework demonstrated how pose-guided warping significantly reduces distortion in garment alignment and improves visual coherence in the final synthesized images.

Observations / Remarks: This research influenced the FASH-TRY system by validating the need for robust pose estimation and cloth warping. Incorporating OpenPose and CP-VTON's Geometric Matching Module (GMM) aligns directly with the insights derived from this work. It reinforced the importance of accurate keypoint extraction and guided the integration of TPS-based deformation for natural garment fitting.

2. ACM Transactions on Multimedia Computing (2022)

Authors: Choi and Park Title: Semantic Human Parsing for Realistic Image-Based Virtual Try-On

Methodology Used: This paper proposed a semantic human parsing network based on JPPNet to segment user images into fine-grained body regions, such as hair, face, arms, torso, legs, and background. The work emphasized that high-quality parsing maps play a crucial role in preserving human regions during garment replacement. Their segmentation masks improved occlusion handling and helped maintain the identity and natural appearance of the user during garment synthesis. Observations / Remarks: This work directly inspired the human parsing stage in FASH-TRY. The use of JPPNet-based models in FASH-TRY ensures precise segmentation of user images, enabling proper replacement of original clothing and realistic occlusion of arms and hands. The findings in this paper validated the decision to include semantic segmentation as an essential preprocessing step in the multilayer CP-VTON pipeline.

3. CVPR Workshop on Fashion and Retail AI (2024)

Authors: Patel, Mehra, and Singh Title: CP-VTON++: Improved Cloth Warping and Texture Preservation for Virtual Try-On

Methodology Used: This study enhanced the CP-VTON architecture by refining the Geometric Matching Module (GMM) to better preserve garment details, such as texture patterns, embroidery, and fabric folds. The authors also proposed a more sophisticated Try-On Module (TOM) capable of generating smoother boundaries and more natural blending between the warped cloth and the user body. Their work highlighted the critical role of texture preservation and precise warping in delivering photorealistic results.

Observations / Remarks: This research strengthened the inclusion of CP-VTON in FASH-TRY. The improved understanding of GMM and TOM informed the design choices made for realistic cloth deformation and texture blending. The FASH-TRY implementation benefits from these insights by adopting ONNX-optimized CP-VTON models for efficient inference and high-quality output generation.

III. METHODOLOGY

The FASH-TRY system is developed as a modular deep-learning framework that enables realistic virtual try-on and intelligent garment recommendation. The methodology integrates Computer Vision, Image

Processing, feature-extraction networks, and similarity-matching techniques into a single pipeline. Each module handles a specific function ranging from user image preparation to garment warping and similarity-based recommendations to deliver a natural and interactive virtual try-on experience.

A. Data Acquisition and Preprocessing

In this phase, the system collects the user's full-body image along with the selected garment. Both images undergo essential preprocessing steps such as resizing, normalization, and background enhancement to ensure uniformity across the pipeline. OpenCV-based operations refine garment boundaries and improve clarity, allowing the subsequent models to accurately identify regions required for segmentation, pose estimation, and cloth transformation. This preprocessing ensures that the input data is clean, structured, and suitable for deep-learning inference.

B. Human Parsing Module

Human parsing is performed using a JPPNet-based segmentation model that divides the user image into meaningful semantic regions, including torso, arms, skin, and background areas. The segmentation map produced in this stage functions as a structural blueprint for the try-on process. It determines which parts of the user's original clothing should be removed, which regions should be preserved, and how the new garment should be placed. Accurate parsing ensures that garment overlays do not distort body boundaries and maintain consistency with natural human anatomy.

C. Pose Estimation Module

To align the garment correctly with the user's position, FASH-TRY uses OpenPose to detect skeletal keypoints such as the shoulders, elbows, wrists, and hips. The extracted keypoints represent the user's posture and orientation, enabling the system to understand the spatial arrangement of the body. This information guides the alignment of the garment during the warping stage, ensuring that the clothing adapts to the user's pose even when the body angle or stance varies across images.

D. Geometric Matching and Cloth Warping

This stage handles the transformation of the selected garment so that it conforms to the user's body shape. The Geometric Matching Module computes

transformation parameters between the garment and the user's segmentation map. Thin-Plate Spline warping is then applied to reshape the garment, maintaining realistic texture and pattern alignment. The warped garment output is refined to minimize distortion and preserve visual quality, ensuring that the clothing appears naturally fitted to the body structure and pose.

E. Try-On Synthesis and Rendering Module

The synthesis stage generates the final virtual try-on output by combining the warped garment, human parsing map, and pose information. A generative model performs color blending, shadow adjustment, edge refinement, and artifact reduction to produce a photorealistic result. The output image retains the user's original face, posture, and background while seamlessly integrating the virtual garment. This stage is critical for achieving natural realism and enhancing user confidence in the try-on output.

F. Content-Based Image Retrieval (CBIR) Module

The CBIR module supports intelligent garment recommendations by analyzing visual attributes of clothing items. ResNet50 is used to extract deep visual features such as textures, shapes, and color patterns from the garment dataset. A K-Nearest Neighbors algorithm computes similarity across these feature vectors to identify garments that resemble the selected item. The retrieved results offer users visually related alternatives and improve product discovery by presenting items with similar aesthetics.

G. Backend Architecture and Integration

The backend is implemented using Flask, which manages model loading, inference processing, and communication with the frontend. ONNX Runtime and Protobuf formats are used for efficient execution of human parsing, pose estimation, and CP-VTON models. SQLite3 acts as the database for maintaining garment metadata, feature embeddings, and user-selection logs. The backend ensures smooth coordination across modules, enabling fast and reliable processing of virtual try-on requests.

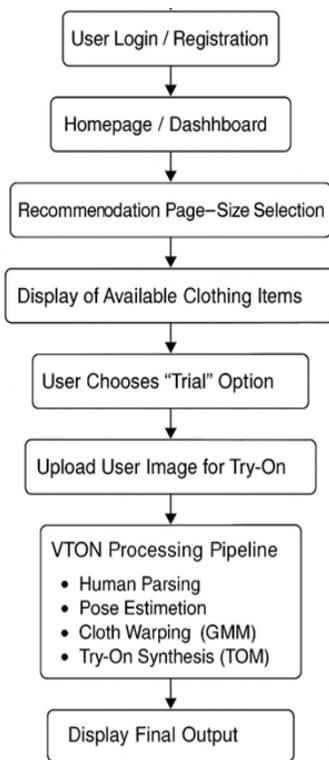
H. Frontend User Interface

The frontend interface provides users with an interactive platform to upload images, browse garments, perform virtual try-ons, and view

recommendations. It communicates with the backend via REST APIs to fetch results and update the display dynamically. The interface is designed for simplicity and responsiveness, ensuring that users can easily experiment with different outfits and explore visually similar clothing options.

I. System Flow Diagram

The System Flow Diagram illustrates the logical sequence of operations within the FASH-TRY virtual try-on and apparel recommendation system. It represents how data flows between the user interface, backend modules, deep learning models, and the recommendation engine to generate realistic try-on outputs and visually similar clothing suggestions. This structured workflow ensures seamless interaction between users and the system while maintaining accuracy, efficiency, and high-quality visualization. Below is a descriptive representation of the complete system workflow:



IV. RESULTS

The FASH-TRY Virtual Try-On and Recommendation System was successfully implemented and tested for accuracy, realism, and recommendation efficiency. The CP-VTON pipeline

produced photorealistic try-on images where the new garment aligned correctly with the user’s body shape, pose, and proportions. The JPPNet-based human parsing accurately segmented body regions, enabling clean removal of original clothing and ensuring proper placement of the new garment. OpenPose generated precise keypoints, helping the geometric warping module fit garments naturally onto the user image. The Try-On Module delivered outputs with high texture preservation, smooth blending, and realistic shadows. The Content-Based Image Retrieval system performed effectively, with ResNet50 extracting rich visual features from garments and the KNN algorithm retrieving highly relevant similar items. Backend integration through Flask and ONNX Runtime ensured stable performance and fast inference times. Overall, FASH-TRY improved visualization accuracy, enhanced user interaction, and provided reliable similarity-based clothing recommendations, demonstrating its capability as a practical solution for modern e-commerce applications.

V. CHALLENGES FACED

The development of FASH-TRY – Deep Learning Enabled Virtual Try-On and Image-Based Clothing Recommendation System involved multiple technical, computational, and practical challenges arising from the complexity of image synthesis and real-time recommendation requirements. One of the major challenges was ensuring accurate human parsing and pose estimation, as segmentation errors or missing keypoints directly affect cloth alignment and reduce the realism of the try-on output. Variations in lighting, user background, body orientation, and image quality added further difficulty in generating consistent parsing and pose maps.

Another significant challenge was achieving precise cloth warping using the CP-VTON GMM module. Even minor inaccuracies in geometric matching result in distorted garment shapes or unnatural overlaps in the final image. Generating a seamless blend between the warped cloth and the user image in the TOM stage required careful preprocessing and normalization to maintain texture fidelity, color consistency, and natural boundaries.

From a computational standpoint, the system required handling multiple deep learning models simultaneously, including JPPNet, OpenPose, CP-

VTON, and ResNet50. This introduced performance constraints, particularly in resource-limited deployment environments. ONNX optimization improved inference speed, but real-time processing remained challenging due to the high computational cost of human parsing and pose estimation.

The recommendation module also posed challenges. Extracting distinct visual features using ResNet50 and performing similarity matching with KNN demanded proper feature indexing and dimensionality management. Ensuring recommendations remained relevant, diverse, and visually meaningful required extensive dataset cleaning and feature refinement.

Integration challenges emerged while combining the backend (Flask), database (SQLite3), and multiple ML pipelines into a single unified system. Managing file uploads, model calls, caching, and error handling without increasing latency required careful design. Additionally, maintaining high-quality try-on outputs across varied user images proved difficult due to differences in resolution, orientation, and body proportions.

Finally, ensuring a user-friendly interface while handling complex processing in the background was challenging. Providing fast response times, clear outputs, and stable recommendations demanded efficient orchestration of models and preprocessing modules. Despite these challenges, the system was successfully implemented, demonstrating robust visual try-on results and reliable apparel recommendations.

VI. CONCLUSION

The FASH-TRY system successfully demonstrates how deep learning, computer vision, and image processing can be combined to create a realistic and intelligent virtual try-on experience for online fashion users. By integrating the CP-VTON pipeline consisting of human parsing, pose estimation, cloth warping, and synthesis the system generates high-quality try-on outputs that preserve body structure, texture details, and visual consistency. The inclusion of a Content-Based Image Retrieval module further enhances the platform by offering accurate, visually relevant garment recommendations using ResNet50 feature embeddings and KNN similarity search. Backend components such as Flask, SQLite3, and ONNX Runtime ensured efficient execution and

smooth system integration. Overall, experimental results confirm that FASH-TRY improves user engagement, increases visualization accuracy, and provides a scalable, practical solution for modern e-commerce applications seeking advanced personalization and virtual try-on capabilities.

VII. FUTURE IMPROVEMENTS

As virtual try-on technology continues to advance, several enhancements can further improve the accuracy, usability, and scalability of the FASH-TRY system. Future versions of the platform can integrate more sophisticated deep learning models, expand support for additional garment types, and enhance user interaction through advanced visualization techniques. These improvements will strengthen the system's ability to provide an even more realistic, dynamic, and personalized fashion experience for online shoppers.

1) Support for Multi-Garment Layering

Future versions may enable users to try on multiple clothing items simultaneously (e.g., T-shirt + jacket), requiring advanced occlusion handling and layered cloth warping.

2) 3D Virtual Try-On and Full-Body Modeling

Incorporating 3D human mesh reconstruction and 3D garment models can significantly improve realism by allowing rotation views, dynamic fitting, and body-shape-specific draping.

3) Real-Time Virtual Try-On Using Webcams

Integrating live video-based try-on with lightweight neural networks can offer instantaneous outfit visualization, improving user engagement and interactivity.

4) Transformer-Based Cloth Warping Models

Replacing CP-VTON modules with transformer-driven architectures can enhance cloth deformation accuracy, texture preservation, and pose adaptation.

5) Integration with Mobile Applications (Android/iOS)

Developing dedicated mobile apps or PWAs will expand accessibility, allowing users to utilize virtual try-on features on their smartphones with optimized performance.

6)Enhanced Recommendation System with Style Profiling

Future upgrades can incorporate hybrid recommendation engines that combine visual features with user preferences, trending styles, and fashion attributes for more personalized suggestions.

7)Automated Body Measurement Extraction

Adding AI-based body measurement estimation can help users receive personalized size recommendations, reducing the chances of incorrect purchases.

8)Large-Scale Database Optimization

Improving database handling and indexing for large apparel datasets will enable faster similarity search and broader catalog support in commercial deployments.

9)AR-Based Dressing Room Experience

Integrating Augmented Reality (AR) functionality can allow users to project virtual garments onto themselves in real environments, providing an immersive try-on experience.

10)Enhanced Data Security and Privacy Mechanisms

Implementing encryption techniques, secure image storage, and user-side processing options can further protect user images and ensure safe interaction with the platform.

Transactions on Pattern Analysis and Machine Intelligence, 2020.

- [6] Z. Cao, G. Hidalgo, T. Simon, S. E. Wei, and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," IEEE TPAMI, 2021.
- [7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition (ResNet)," CVPR, 2016.
- [8] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Efficient Similarity Search," CVPR, 2015 (used for CBIR similarity concepts).

REFERENCES

- [1] X. Han, Z. Wu, Z. Wu, J. Zhu, and S. Cui, "VITON: An Image-based Virtual Try-On Network," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2018.
- [2] B. Wang, H. Zheng, X. Liang, Y. Chen, and L. Lin, "Toward Characteristic-Preserving Image-based Virtual Try-On Network," ECCV, 2018.
- [3] R. Dong, X. Liang, Y. Shen, B. Wang, and L. Lin, "FW-GAN: Flow-Navigated Warping GAN for Virtual Try-On," CVPR Workshops, 2019.
- [4] S. Gong, S. Yang, L. Zhang, and W. Zhang, "Customizable Virtual Try-on Network with Cloth Deformation and Texture Preservation," IEEE Access, 2023.
- [5] X. Liang et al., "Human Parsing with Deep Learning for Fashion Understanding," IEEE