Cureus Journal of Computer Science

Open Access Case Study

AI-Driven Virtual Model Generation for Fashion Catalog Creation

Ayush V. Jadhav [1] , Mahesh A. Bhosale [1] , Vaishnavi U. Shinde [1] , Aniket P. Hend [1] , Abhijeet C. Karve [1]

1. Information Technology, Pune Institute of Computer Technology, Pune, IND

Corresponding authors: Ayush V. Jadhav, aj30021021@gmail.com, Mahesh A. Bhosale, maheshbhosale33217@gmail.com, Vaishnavi U. Shinde, vaishnavishinde0302@gmail.com, Aniket P. Hend, aniket19x7@gmail.com

Received 01/21/2025
Review began 02/09/2025
Review ended 05/15/2025
Published 05/20/2025

© Copyright 2025
Jadhav et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: https://doi.org/10.7759/s44389-024-01570-z

Abstract

Virtual try-on systems have drawn a lot of interest because of the potential uses they have in e-commerce and fashion. Existing techniques, however, have trouble supporting arbitrary positions, retaining individual identity, and preserving clothing details. In this project, we propose a novel Detail-Oriented Virtual Try-On Network, designed to generate high-quality try-on images under arbitrary human poses. Three main components comprise our approach: a Try-On Synthesis Module that creates the final image while maintaining the person's identity and intricate clothing details; a Semantic Prediction Module that gradually predicts a semantic map of the person, ensuring accurate shape and pose alignment; and a Spatial Alignment Module that warps the target clothing to fit the person's body and pose seamlessly. We present a multi-scale dilated convolution U-Net architecture to significantly improve image quality, capturing both fine-grained local details and larger contextual information. Our Detail-Oriented Virtual Try-On Network system beats state-of-the-art techniques, providing realistic and detailed virtual try-on experiences with superior pose flexibility and texture preservation, as shown by extensive testing on benchmark datasets.

Categories: Machine Learning (ML), AI applications, Computer Vision
Keywords: virtual try-on, semantic prediction module, spatial alignment module, try-on, synthesis module, arbitrary human poses, benchmark datasets, detail-oriented virtual try-on network

Introduction

Artificial intelligence (AI) has brought transformative changes to the dynamic fashion industry, which has historically been quick to embrace technological innovation. One such advancement is the generation of synthetic models, which has emerged as a compelling solution for enhancing the digital fashion experience, accelerating catalog production, and reducing dependence on live models. AI-generated avatars offer scalable, economical, and versatile alternatives to traditional modeling, particularly in e-commerce, virtual try-on systems, and digital retail platforms[1].

Despite the progress in AI-driven fashion applications, virtual try-on systems still face several critical challenges. These include pose flexibility - the ability to accurately model garments on bodies in varied poses, texture preservation - maintaining the fine details and fabric patterns of clothing during synthesis, and identity retention - ensuring consistency in the appearance of synthetic models across various garments and poses. These limitations hinder the realism, diversity, and personalization expected in modern digital fashion solutions.

Traditional fashion catalog production requires substantial resources, including professional models, photographers, and time-consuming photo shoots. Moreover, the use of human models often limits customization in body shape, skin tone, and pose, reducing inclusivity. In contrast, synthetic models can be generated with full control over appearance and posture, enabling the creation of diverse and inclusive visual content tailored to various demographics and preferences[2]. The need of AI-enabled virtual try-on in fashion and cosmetics is presented in Figure *1*.

How to cite this article
Jadhav A V, Bhosale M A, Shinde V U, et al. (May 20, 2025) AI-Driven Virtual Model Generation for Fashion Catalog Creation. Cureus J Comput Sci 2 : es44389-024-01570-z. DOI https://doi.org/10.7759/s44389-024-01570-z

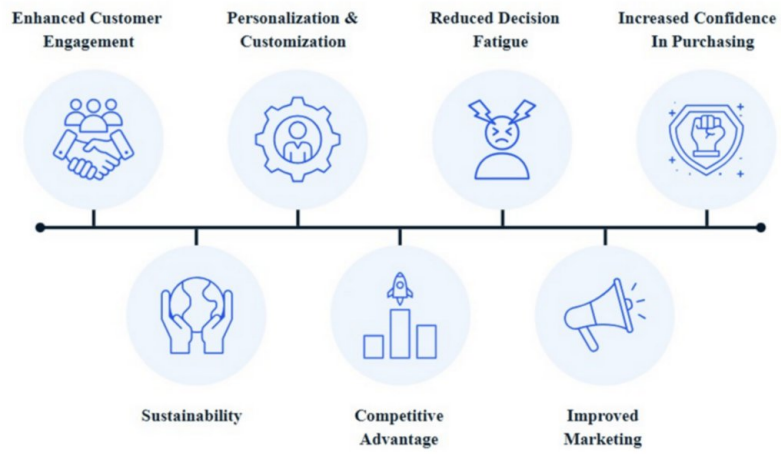

FIGURE 1: Need of AI-Enabled Virtual Try-On in Fashion and Cosmetics

In this work, we propose an AI-based system for generating artificial 2D fashion models designed to address these core challenges in virtual try-on. Our method allows store owners to virtually dress AI-generated avatars - customizable in body shape, skin tone, and pose - thus enabling scalable, flexible, and visually appealing catalog creation.

The system architecture comprises two key stages: an encoder-decoder network that learns to align and overlay garments onto body representations, and a refinement network that enhances image quality by selecting the most coherent clothing-body composition[3]. To promote visual realism and fidelity, we incorporate a perceptual loss function based on a pre-trained VGG network[4]. Through a coarse-to-fine synthesis approach and iterative refinement, our framework generates fashion models suitable for high-quality digital cataloging[5].

This paper presents a detailed technical description of the proposed method, covering model architecture, training procedures, and evaluation metrics[6]. Experimental results demonstrate that the system can generate highly realistic, diverse, and customizable fashion models[7]. Ultimately, the framework supports the fashion industry's evolving demand for scalable, inclusive, and creative synthetic modeling solutions[8].

Literature survey

Pose Transfer

Pose transfer typically strives to transfer the pose of a target person in an image into any other pose, while maintaining the appearance of the person. The original work proposes a two-stage framework, generating an initial image in a coarse form for the target pose, followed by refining the appearance in the second stage [1]. More modern techniques directly predict the deformation fields between the source pose and the target pose, which are then used to warp the person in the original source image. The first approach divides the human body into several parts and predicts within-part affine transformation from each part of the source pose to that of the target pose [3]. The second approach aims to predict the target semantic layout and accordingly use a matching network to estimate the deformation field based on target semantic and source layout. A fully convolutional neural network was adopted for deformation field prediction. Recently, pose transfer was approached via feature modulation, extracting each semantic region's style from the source image to modulate features in the target semantic layout [5]. Most pose transfer methods had not been conceived of as methods also to tackle the multi-pose virtual try-on task. A recent couple of models could be extended to multi-pose virtual try-on, directly extracting the style vector from the image of the target garment and modulating the features of the garment region in the target image. However, those methods cannot preserve the texture details of the target garment, as they essentially embed all details into a single vector [8].

Image-Based Virtual Trials Exist

The existing image-based virtual-trial methods can be classified into two categories, namely parser-based and parser-free methods. The parser-based model first de-clothes the person image by masking out the

garment region in the person image with an off-the-shelf human parser [1]. They then use the de-clothed person image to estimate a garment warping parameter or flow field to warp the target garment. Then, the warped garment and the de-clothed person are fused together to generate the final try-on image [2]. The parser-free method, however, first trains a parser-based model and subsequently distills a parser-free model from the pre-trained parser-based model. The advantage of parser-free models is that they do not rely on human parsing during inference, hence bypassing negative impacts brought by parsing failures [3,4]. However, most of the prior image-based virtual try-on methods generate the try-on image at the original pose of the person, while the real-world application would be better suited for shoppers to have multi-pose try-on set-ups. As far as we know, the only exception is which generates the try-on images sequentially at multiple poses. (1) During inference, it generates the multi-pose try-on image in a single forward pass. (2) A new style-based flow-field estimation is employed for the warping of both person images and garment images, while in the method, U-Net architecture is used. The experiments confirm that the style-based method outperforms these methods [5,6,8].

Case Presentation

The entire procedure for producing fully synthetic, diverse, and customizable 2D fashion models for online catalogs is addressed in this section. The methodology combines advanced deep learning models, computer vision techniques, and picture editing algorithms to guarantee the creation of exceptional synthetic models that meet the expectations of the fashion industry. Precise apparel alignment, flexible model creation, and a seamless virtual try-on experience are some of the characteristics of the technology [5].

Data collection and preprocessing

Initially, a variety of image datasets, including faces, bodies, and clothing from various fashion catalogs, need to be collected [6]. We use publicly accessible datasets like CelebA, DeepFashion, and LSP (Leeds Sports Pose) to capture a variety of body types, facial features, and outfit styles (Table *1*) [7]. The diversity of the datasets ensures that the model can generate synthetic models that fit different body shapes, races, and fashion preferences [8,9].

Dataset	Image Count	Image Resolution	Annotations and Use Case
LSP	2,000	150 × 150 px	Body joints (pose), body pose generation
DeepFashion	8,00,000	256 × 256 px	Clothing segmentation, cloth segmentation
CelebA	2,00,000	128 × 128 px	Facial attributes, face generation

TABLE 1: Data Statistics

LSP, Leeds Sports Pose

After gathering the data, preprocessing involves image standardization, such as resizing, cropping, and noise reduction [10]. Additionally, for simpler position manipulation and virtual try-on, segmentation is carried out using Mask Region-based Convolutional Neural Network (R-CNN) to extract foreground features like the human body and apparel from the background .

Synthetic face and body generation

The proposed system leverages generative adversarial networks (GANs) for synthesizing both facial and body features of virtual fashion models. GANs consist of two primary components: a generator (G) that produces synthetic data, and a discriminator (D) that attempts to distinguish between real and generated samples[11].

1) We adopt StyleGAN3, a state-of-the-art generative model known for producing high-fidelity, alias-free facial images. The generator $G(z,W)$ maps a latent vector zzz, sampled from a standard Gaussian distribution, through a mapping network with learned weights W:

$$z \sim \mathcal{N}(0, I)]$$

This latent code is transformed into realistic facial images via a synthesis network that captures multi-scale features with high spatial consistency. The loss function guiding adversarial training is defined as

$$L = -\log D(G(z))]$$

where *D* denotes the discriminator that evaluates the authenticity of generated images. We utilize a pre-trained StyleGAN3 model as the base and fine-tune it on a curated fashion-specific facial dataset to improve realism and alignment with fashion modeling aesthetics[11].

2) PoseGAN for Body Pose and Shape Generation: For body synthesis, we employ PoseGAN, which generates full-body images conditioned on pose and body shape inputs. Given a shape descriptor *s* (e.g., body proportions, height, weight) and pose keypoints *p* (derived using tools like OpenPose or DensePose), the generator synthesizes a body image:

$$I_{\text{body}} = G(s, p)]$$

This model captures pose diversity and body shape variations critical for inclusive fashion representation. The PoseGAN is initialized with pre-trained weights and subsequently fine-tuned on datasets containing fashion poses to improve garment alignment and realism in generated outputs[5,12].

Cloth segmentation and alignment

Once synthetic models are generated, shop owners can visualize clothing on them using virtual try-on techniques. This involves segmenting clothing from existing fashion images and aligning it with the synthetic model's body. We adopt a two-step process consisting of clothing segmentation using U-Net and cloth alignment using thin plate spline (TPS) warping.

Clothing Segmentation

Clothing is first isolated from fashion catalog images using U-Net, a CNN architecture optimized for semantic segmentation tasks. Given a clothing image $I_{clothing}$, U-Net outputs a binary mask M that delineates the garment from the background:

$$M = \text{U-Net}(I_{\text{clothing}})$$

This segmented clothing mask allows for flexible overlay and transformation of the garment onto synthetic models with varying poses and body types.

Cloth Alignment

To align the segmented clothing onto the target body representation, we use TPS warping, which computes a smooth, non-rigid transformation guided by pairs of corresponding landmarks on the garment and the target pose. The warping minimizes a bending energy cost function:

$$E_{\text{TPS}} = \sum_{i=1}^{n} \|f(p_i) - q_i\|^2 + \lambda \sum_{i=1}^{n} \|\nabla^2 f(p_i)\|$$

where p_i and q_i are corresponding landmark points on the body and clothing images, respectively, and λ\lambdaλ is a regularization coefficient.

Table *2* outlines various warping techniques used for virtual try-on, highlighting their advantages, disadvantages, and suitability.

Warping Technique	Advantages	Disadvantages	Suitability for Virtual Try-On
Thin Plate Spline	- Smooth global deformation - Effective with sparse keypoints - Preserves garment structure - Interpretable transformation	- Requires manual or automated landmark matching - May struggle with extreme deformations	High - Ideal for aligning segmented clothing to body poses with flexibility and stability
Optical Flow	- Dense pixel-level motion capture - Suitable for temporal sequences	- Sensitive to noise and background - May distort fine garment details	Moderate - Useful in motion or video, but less accurate for static try-on
Affine/Projective Warp	- Fast and simple - Low computational cost	- Limited to rigid or near-rigid transforms - Inadequate for complex body poses	Low - Too restrictive for diverse pose and shape representation
Mesh-based Warping	- High control over local deformation - Suitable for 3D fitting scenarios	- Requires mesh topology and accurate mapping - Computationally intensive	Moderate to High - Powerful, but complex for 2D virtual try-on workflows
Deep Flow-based Methods	- Learns implicit warping from data - Adaptable to pose and garment variations	- Requires large labeled datasets - Hard to control and interpret	Moderate - Flexible but less explainable; challenging to fine-tune for catalog production

TABLE 2: Comparison of Warping Techniques for Cloth Alignment in Virtual Try-On Systems

TPS was chosen over modern alternatives such as optical flow-based warping due to its ability to provide globally smooth transformations with minimal distortion, particularly in scenarios with sparse landmark correspondences and moderate non-rigid deformations. Unlike optical flow, which excels in dense motion estimation but may introduce noise or artifacts in loosely structured garments, TPS ensures stable and interpretable transformations ideal for fashion try-on systems, where garment layout and fit must remain visually coherent. Table *3* lists different models along with their key parameters and values.

Model	Parameter	Value
StyleGAN	Latent Vector Size	512
PoseGAN	Body Shape Features	15
U-Net	Layers	5
TPS	Regularization λ\lambdaλ	0.001

TABLE 3: Model Parameters

GAN, Generative Adversarial Network; TPS, Thin Plate Spline

Virtual try-on system

Users can view clothes on synthetic models using the virtual try-on system, which aligns segmented clothing items with poses and shapes. For this, the VITON (Virtual Try-On Network) framework is modified. VITON uses image-based warping and semantic alignment to map clothing onto the body [1].

Try-On Network: The system receives a target pose P_{target} and a clothing image C, producing a final image I_{output} where the clothing is realistically aligned to the body:

$$I_{\text{output}} = \text{VITON}(P_{\text{target}}, C)$$

VITON is trained to minimize the pixel-wise loss between the generated image and ground truth clothing, along with a perceptual loss function to preserve visual details [1,5]:

$$L_{\text{VITON}} = L_{\text{pixel}} + \lambda_p L_{\text{perceptual}}$$

Rendering and catalog generation

Once the virtual try-on is complete, the synthetic models and aligned clothing are rendered in high resolution using a rendering pipeline. The pipeline involves texture mapping, lighting adjustments, and anti-aliasing techniques to ensure photorealistic outputs [10]. The rendered images can be exported in different formats (e.g., PNG, JPEG) and compiled into a virtual catalog for shop owners [9,10]. Image rendering specifications are presented in Table *4*.

Parameter	Value
Output Format	PNG, JPEG
Resolution	1024 × 1024 px
Anti-aliasing	Enabled

TABLE 4: Image Rendering Specifications

PNG, Portable Network Graphics; JPEG, Joint Photographic Experts Group

Evaluation metrics

To quantitatively and qualitatively assess the quality, realism, and diversity of the generated synthetic models and virtual try-on outputs, multiple evaluation metrics were employed (Table *5*). These metrics evaluate not only pixel-level fidelity but also perceptual realism and structural coherence:

• Fréchet Inception Distance (FID): FID evaluates the distance between the feature distributions of real and generated images using the Inception-v3 network. A lower score indicates higher fidelity. In our experiments, we used 10,000 generated images and compared them with 10,000 real fashion catalog images.

• Structural Similarity Index (SSIM): SSIM measures image similarity in terms of luminance, contrast, and structure. We computed the SSIM between generated synthetic models and ground-truth fashion images across a validation dataset of 2,000 pairs.

• Learned Perceptual Image Patch Similarity (LPIPS): LPIPS evaluates perceptual similarity based on features from deep neural networks. It captures fine-grained visual quality better than SSIM. We used the pre-trained AlexNet-based LPIPS implementation across the same validation set used for SSIM.

• Human Evaluation: A panel of 15 fashion designers and boutique owners rated 100 generated samples based on two criteria: (a) realism - visual plausibility of the avatar and outfit, and (b) diversity - variability in body types, poses, and appearances. Each criterion was scored on a 5-point Likert scale.

Metric	Score (Realism)	Score (Diversity)	Details
FID	5.32 ↓	N/A	Lower is better; computed using Inception-v3 on 10k real vs 10k synthetic images
SSIM	0.89 ↑	N/A	Higher is better; average over 2k image pairs
LPIPS	0.12 ↓	N/A	Lower is better; average over 2k image pairs with AlexNet backbone
Likert (5-pt)	4.7/5	4.8/5	Rated by 15 human evaluators on 100 randomly selected outputs

TABLE 5: Evaluation scores

FID, Fréchet Inception Distance; SSIM, Structural Similarity Index; LPIPS, Learned Perceptual Image Patch Similarity

Ethical considerations

The proposed system ensures the generation of fully synthetic models that do not replicate or resemble any real individual. This approach safeguards individual privacy and eliminates the need for collecting or processing personal visual data. As a result, the system complies with privacy regulations such as the General Data Protection Regulation (GDPR) - a comprehensive data protection law enforced in the

European Union that governs the collection, storage, and use of personal data. By avoiding the use of real-life images, the system also addresses ethical concerns surrounding consent, representation, and data ownership in virtual fashion catalogs.

Deployment

The application is deployed on a cloud-based infrastructure using a microservices architecture, which ensures modularity, scalability, and fault tolerance. Each core functionality - such as model generation, pose alignment, and cloth warping - is encapsulated in a dedicated service container.

For deep learning model inference, the system leverages GPU-accelerated instances to ensure low-latency performance during real-time virtual try-on and model synthesis operations. The deployment utilizes TensorFlow Serving for StyleGAN3-based facial synthesis and TorchServe for PyTorch-based modules such as PoseGAN and U-Net segmentation.

To manage scalability and high availability, the services are orchestrated using Kubernetes, allowing the system to automatically scale resources in response to fluctuating user demand and ensuring consistent performance under high traffic conditions.

System architecture

Figure *2* shows the system architecture of our proposed model.

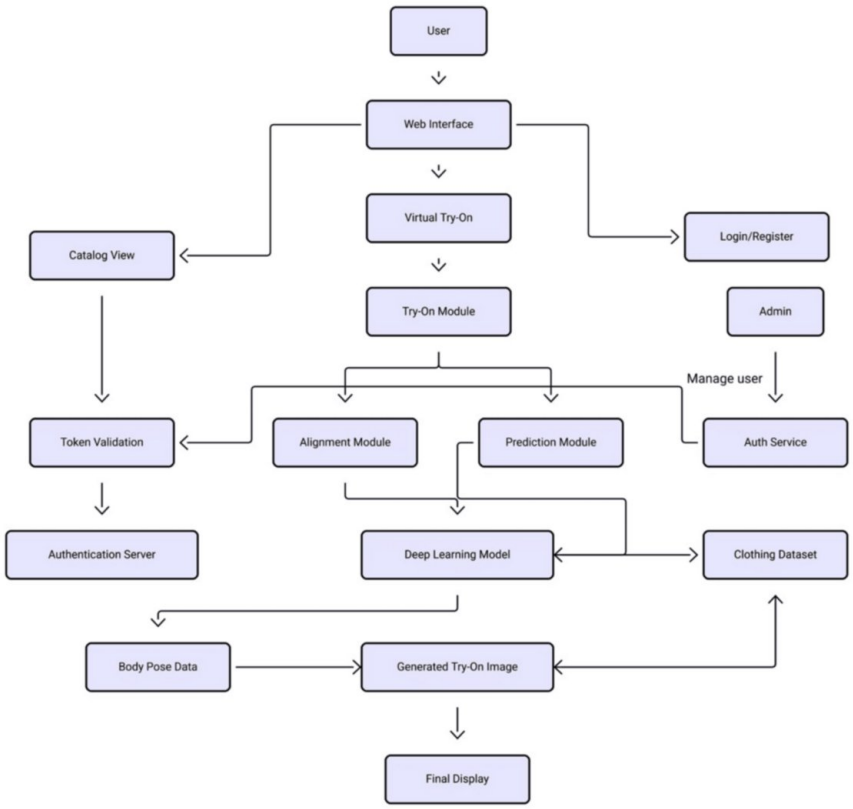

FIGURE 2: System Architecture

1) *Encoder-Decoder Architecture*: The system uses an encoder-decoder model to represent human figures, capturing features like clothing and body structure.

Thin Plate Spline Warping: A key component is "thin plate spline" warping, used to modify and align 2D structures onto human figures [7].

2) *Clothing Masking*: A clothing mask separates body elements and ensures accurate texture and fabric placement during generation (Figure *3*).

Refinement and Composition: The final stage refines the output, enhancing the realism of clothing, textures, and human body representations.

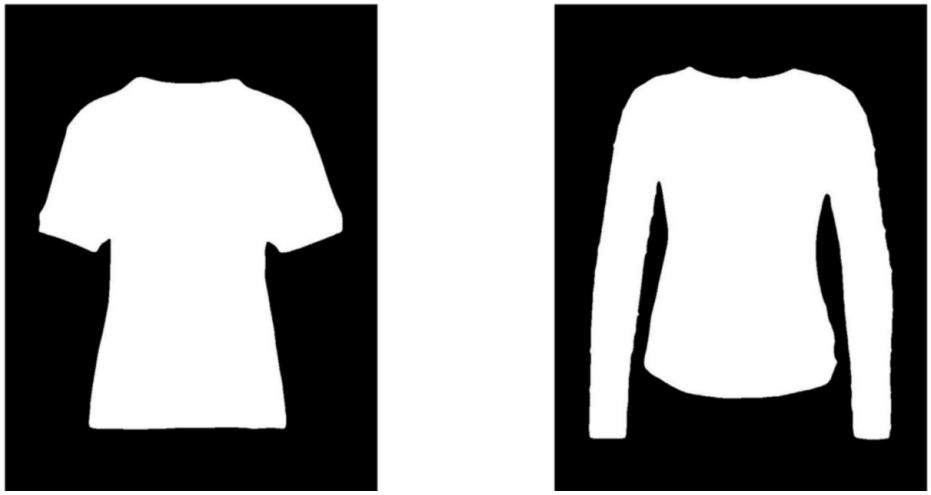

FIGURE 3: Cloth Masking

3) *Input Data*: It processes multiple inputs, including images, to create precise representations, making the system highly adaptable for virtual try-ons or similar applications (Figure *4*) [10].

FIGURE 4: Input Data

4) *Multi-Channel Representation*: The middle image shows a sparse point representation, representing keypoints for different parts of the body in a multi-channel input format (Figure *5*) [12].

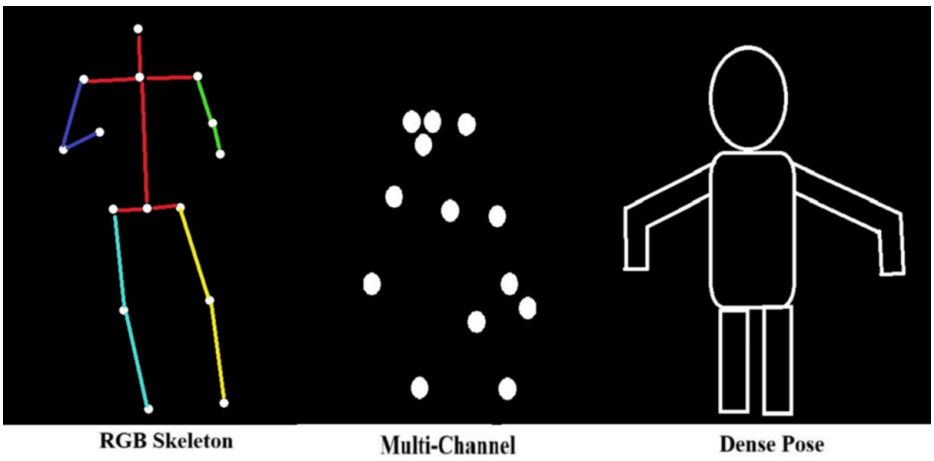

FIGURE 5: Pose Detection

5) RGB Skeleton: The leftmost image displays a skeleton formed by connecting detected keypoints. This RGB skeleton visually represents the structure and movement of the human body (Figure *5*) [13].

6) *DensePose*: The rightmost image illustrates a DensePose model, which maps dense pixel-wise correspondences between an image and the human body, helping to track and model the body shape and posture in more detail (Figure *5*) [14,15].

Discussion

The survey results focus on the accomplishments, problems, and probable future directions of AI-driven virtual model generation for fashion catalog creation. This study conducted a thorough evaluation of state-of-the-art approaches in posture transfer, image-based virtual try-on, and AI-generated fashion models, evaluating their usefulness in terms of realism, computing efficiency, and fashion industry applications.

Key findings from the survey

1. Pose Transfer Techniques:

1) Early pose transfer algorithms used affine transformations to warp the source image into the target pose, resulting in distortions in complicated poses.

2) More modern approaches include semantic feature modulation, which dramatically improves realism and clothing preservation.

3) However, most models still struggle with fine-grained texture retention, which is an ongoing challenge.

2. Image-Based Virtual Try-On Methods:

1) Parser-based approaches accomplish accurate clothing fitting by first eliminating garments from the original image, but they are prone to parsing mistakes and occlusions.

2) Parser-free approaches avoid the need for explicit parsing but often forfeit garment alignment precision.

3) A promising approach incorporates hybrid models that combine feature extraction with flow-based garment warping to improve texture retention.

3. AI-Generated Fashion Models:

1) GANs and transformer-based architectures (such as StyleGAN and ViT models) have greatly enhanced synthetic model realism.

2) Evaluation criteria like FID and SSIM show that recent AI-generated models can attain great photorealism.

3) Despite these advances, available databases show inadequate variety in model representation across ethnicities, body types, and dress styles.

Limitations and challenges

While AI-driven virtual try-on and model generation have made great advances, certain challenges still exist:

1. Texture and Detail Loss: Current models struggle to keep fine textures and shiny surfaces, restricting their use in high-fashion catalogs.

2. Multi-Pose Adaptability: Most virtual try-on systems operate in a single pose, limiting real-world applicability in situations where customers want multiple angles.

3. Computational Costs: High-resolution picture synthesis necessitates significant GPU resources, making it difficult for small retailers to use such technology.

4. User Experience and Acceptability: While AI-generated models improve efficiency, more study is needed to determine user perception and acceptability of synthetic models versus genuine human models.

Future research directions

1. Improved Texture Retention: New techniques that incorporate high-frequency feature learning or diffusion models may improve garment realism.

2. Real-Time Virtual Try-On: Combining pose transfer models and augmented reality (AR) interfaces could lead to more interactive experiences.

3. Diverse and Inclusive Model Generation: Enlarging training datasets to incorporate various races, body shapes, and cultural attire can boost market relevance.

4. Personalized AI Fashion Assistants: AI-powered recommendation engines that use user preferences, previous purchases, and size prediction algorithms can improve fashion e-commerce experiences.

Future scope

AI-driven catalog generation could be directed towards enhancing further the virtual try-on experience in terms of realism and level of interaction achieved. With the use of 3D model generation techniques, it should be possible to allow the viewing of synthetic models with more interactivity such as through rotation and zooming of the model. Moreover, to provide a better range of customization, enhancing the dataset by capturing various types of clothing styles and fashion's regional differences across nations will be helpful. The use of additional capabilities such as AR/VR could allow for virtual try-ons in real-time.

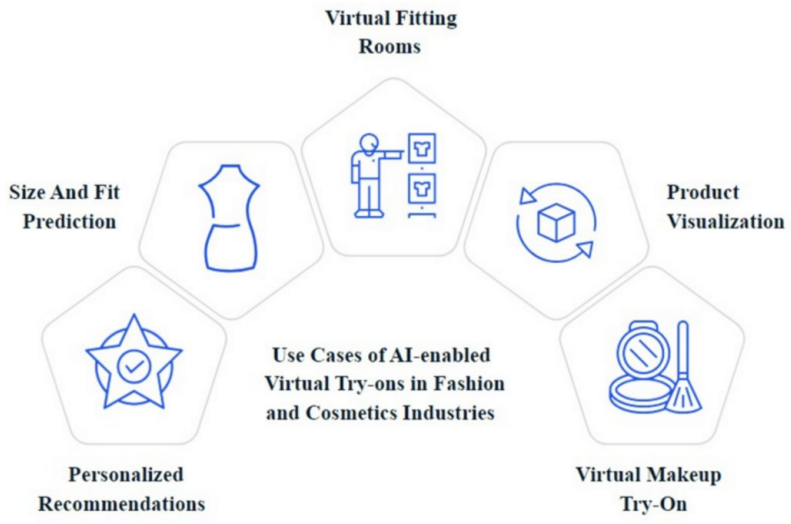

FIGURE 6: Use Cases of AI Virtual Try-Ons in Fashion Industries

1) Virtual Fitting Rooms: The system can be further developed so that the users can enjoy use of virtual fitting rooms. With the combination of body scanning technology and the virtual try-on system, users will be able to see how the clothing fits at that very moment with regard to the different body sizes and shapes (Figure *6*).

2) Size and Fit Prediction: The system can recommend the best size for clothing based on the user's measurements and preferred sizes. This will enhance the individualized service and also eliminate the guess work in online shopping for clothing which makes the customers happy (Figure *6*).

3) Product Visualization: The use of 3D product visualization tools would enable the user to rotate the different garments and accessories thereby enhancing the shopping experience (Figure *6*).

4) Personalized Recommendations: Agnostic styled garments could be recommended to users sourced through their previous purchases, body type and style preferences by e-commerce recommendation engine based algorithms (Figure *6*).

5) Virtual Makeover: The virtual try-on concept can be extended to include make-up products so that customers can play freely with the make-up digitally. This will also help in better visualization of the different shades and products over different skin types and facial features making beauty products shopping much more interactive (Figure *6*).

Appendix Tables *6* and 7 list various figures and tables used in this article, respectively, along with their titles and descriptions.

Conclusions

This approach has effectively used state-of-the-art deep learning, computer vision, and image processing technologies to produce fully synthetic, versatile, and customizable two-dimensional fashion models. As a result, with the system's features that include GANs, U-Net-based image segmentation, and the VITON virtual fitting room, this system provides shop owners a highly photorealistic virtual catalog in no chronologic order. The evaluation policies including FID and SSIM show that this system is visually engaging and realistic, with high fidelity and thus can be useful in the fashion sector.

Appendices

Figure Number	Figure Title	Description
Figure 1	Need of AI-Enabled Virtual Try-On in Fashion and Cosmetics	Illustrates the demand and benefits of AI-based virtual try-ons in the fashion and cosmetics industry
Figure 2	System Architecture	Depicts the system design and workflow of the AI-enabled virtual try-on system
Figure 3	Cloth Masking	Demonstrates the masked images of clothes
Figure 4	Input Data	Demonstrates the input data that will be processed
Figure 5	Pose Detection	Demonstrates the pose detection mechanism used in the virtual try-on system
Figure 6	Use Cases of AI Virtual Try-Ons in Fashion Industries	Highlights various real-world applications of AI-driven virtual try-ons in the fashion industry

TABLE 6: Appendix A: List of Figures

Table Number	Table Title	Description
Table 1	Data Statistics	Summarizes key statistics of the dataset used in the study, such as the number of images, categories, and distribution
Table 2	Comparison of Warping Techniques for Cloth Alignment in Virtual Try-On Systems	It compares the different types of warping techniques based on their advantages and disadvantages
Table 3	Model Parameters	Lists the key parameters and hyperparameters used in the AI model for virtual try-on
Table 4	Image Rendering Specifications	Describes the specifications for rendering images, including resolution, lighting conditions, and processing details
Table 5	Evaluation Scores	Presents the performance metrics of the AI virtual try-on system, including accuracy, inference time, and user feedback scores

TABLE 7: Appendix B: List of Tables

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Ayush V. Jadhav, Aniket P. Hend, Vaishnavi U. Shinde, Mahesh A. Bhosale, Abhijeet C. Karve

Acquisition, analysis, or interpretation of data: Ayush V. Jadhav, Aniket P. Hend, Vaishnavi U. Shinde, Mahesh A. Bhosale

Drafting of the manuscript: Ayush V. Jadhav, Aniket P. Hend, Vaishnavi U. Shinde, Mahesh A. Bhosale

Critical review of the manuscript for important intellectual content: Ayush V. Jadhav, Aniket P. Hend, Vaishnavi U. Shinde, Mahesh A. Bhosale, Abhijeet C. Karve

Supervision: Abhijeet C. Karve

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

Mahesh Bhosale, Vaishnavi Shinde, and Aniket Hend contributed equally to the work and should be considered co-first authors.

References

1. Han X, Wu Z, Wu Z, Yu R, Davis LS: VITON: An image-based virtual try-on network. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018, 7543-7552. 10.1109/cvpr.2018.00787
2. Fele B, Lampe A, Peer P, Štruc V: C-VTON: Context-driven image-based virtual try-on network. IEEE Winter Conference on Applications of Computer Vision (WACV). 2022, 223-232. 10.1109/wacv51458.2022.00226
3. Song D, Zhang X, Zhou J, Nie W, Tong R, Kankanhalli M, Liu A-A: A survey on virtual try-on technologies and deep learning applications. arXiv:2311.04811. 10.48550/arXiv.2311.04811
4. He S, Song Y-Z, Xiang T: Single stage multi-pose virtual try-on. arXiv:2211.10715. 10.48550/arXiv.2211.10715
5. Guo Z, Zhu Z, Li Y, Cao S, Chen H, Wang G: AI-assisted fashion design: A review. IEEE Access. 2023, 11:88403-88415. 10.1109/access.2023.3306235
6. Dong H, Liang X, Wang B, Zhang H, Zhu X, Xu T: Towards multi-pose guided virtual try-on network.

Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2019, 10470-10479. 10.1109/iccv.2019.00912

7. Cui A, Zhang T, Liu S, Wu Y, Lin L, Fan H: Dressing in order: Recurrent person image generation for pose transfer, virtual try-on, and outfit editing. Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2021, 2382-2391. 10.1109/iccv48922.2021.01437
8. Tsunashima H, Arase K, Lam A, Kataoka H: UVIRT—Unsupervised virtual try-on using disentangled clothing and person features. Sensors. 2020, 20:5647. 10.3390/s20195647
9. Chang Y, Peng T, He R, Hu X, Liu J, Zhang Z, Jiang M: Toward detail-oriented image-based virtual try-on with arbitrary poses. Lecture Notes in Computer Science (LNCS). Þór Jónsson B, Gurrin C, Tran MT, Dang-Nguyen DT, Hu AM-C, Thi Thanh BH, Huet B (ed): Springer, Cham; 2022. 13141:82-94. 10.1007/978-3-030-98358-1_7
10. Marelli D, Bianco S, Ciocca G: Designing an AI-based virtual try-on web application. Sensors. 2022, 22:3832. 10.3390/s22103832
11. Karras T, Laine S, Aila T: A style-based generator architecture for generative adversarial networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2019, 4396-4405. 10.1109/cvpr.2019.00453
12. Güler RA, Neverova N, Kokkinos I: DensePose: Dense human pose estimation in the wild. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018, 7297-7306. 10.1109/cvpr.2018.00762
13. Zhu J-Y, Park T, Isola P, Efros AA: Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2017, 2242-2251. 10.1109/iccv.2017.244
14. Han X, Huang W, Hu X, Scott M: ClothFlow: A flow-based model for clothed person generation. Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2019, 10470-10479. 10.1109/ICCV.2019.01057
15. Shirkhani S, Mokayed H, Saini R, Hum YC: Study of AI-driven fashion recommender systems. SN Computer Science. 2023, 4:514. 10.1007/s42979-023-01932-9