RESEARCH ARTICLE OPEN ACCESS

Virtual Try on

Joe Joseph Raj*, Jeevan Jijo**, Jayasankar P A***,Vidhula Thomas****
*(Integrated MSc Student, Department of Data Science, MG University, Nirmala College Muvattupuzha, India
Email: joejr3387@gmail.com)
**(Integrated MSc Student, Department of Data Science, MG University, Nirmala College Muvattupuzha, India
Email: jeevanjijo10@gmail.com)
***Integrated MSc Student, Department of Data Science, MG University, Nirmala College Muvattupuzha, India
Email: c1.jayasankarp.a@gmail.com)
****(Assistant Professor, Department of Data Science, MG University, Nirmala College Muvattupuzha, India
Email: vidhulathomas90@gmail.com)

--************************----------------------------------

Abstract:

The rapid expansion of e-commerce in the fashion industry has driven the development of Virtual Try-On (VTON) technologies, which enable consumers to visualize garments on their own bodies or representative avatars, enhancing the online shopping experience. This paper presents a comprehensive review and analysis of state-of-the-art VTON methods, including 2D image-based approaches such as CP-VTON and VITON, hybrid systems like CloTH-VTON, and fully 3D models exemplified by 3D-MPVTON. These techniques leverage advanced deep learning architectures—Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), Capsule Networks, and transformers—to achieve photorealistic garment fitting and synthesis.Our study highlights significant technical achievements, such as high segmentation accuracy and improved garment shape preservation, alongside commercial impacts including increased sales and reduced return rates. Despite these advancements, challenges remain in handling complex clothing textures, diverse body shapes, occlusion, and real-time performance constraints. Additionally, ethical considerations related to privacy and data quality demand attention.Future directions emphasize the integration of multimodal inputs, 3D pose estimation, and personalized styling, aiming to develop more inclusive, efficient, and scalable VTON systems. This research underscores the transformative potential of Virtual Try-On technologies in reshaping digital fashion retail by bridging the gap between physical and virtual shopping experiences.

Keywords **—Virtual Try-On, Deep Learning, Generative Adversarial Networks (GANs), Computer Vision, 2D Image-Based Methods, 3D Reconstruction, CP-VTON, CloTH-VTON, Pose Estimation, Neural Body Fit, Fashion E-commerce, Image Synthesis, Multimodal Systems, Inclusivity, Body Shape Diversity, Real-Time Optimization, Hybrid Models, User Personalization, Ethical Considerations, Digital Fashion Retail, Hierarchical Cross-Attention Network (HCANet), Latent Diffusion Model (LDM).**

--************************----------------------------------

I. INTRODUCTION

The rapid growth of e-commerce, especially in the fashion industry, has created a demand for innovative solutions that bridge the gap between online shopping and physical store experiences.. The core objective of this field is to develop systems that allow users to visualize garments on their own body or a representative avatar from any angle, thereby increasing consumer confidence, boosting sales, and reducing returns.

From a technical perspective, Recent research has explored diverse approaches, from 2D image-based methods such as CP-VTON and VITON to hybrid and fully 3D systems like CloTH-VTON and 3D-MPVTON. These developments rely on deep learning models such as CNNs, ResNet, Inception, and newer approaches involving capsule networks and transformers. The actual try-on is frequently accomplished using Generative Adversarial Networks (GANs)—including CP-VTON+, TryOnGAN, and ClothFlow—which are trained to warp the target clothing item to fit the user's pose and then synthesize a photorealistic final image.

Research findings increasingly indicate that not only technical progress but also commercial potential. Studies demonstrate that virtual try-on systems can drive higher sales, improve customer satisfaction, and significantly reduce return rates, making them highly beneficial for retailers. Furthermore, a critical challenge is moving beyond standard garments to achieve high accuracy for custom outfits and ensuring inclusive representation of diverse body shapes and sizes. With e-commerce projected to dominate retail fashion markets, advancing Virtual Try-On technologies represents a critical step toward reshaping digital shopping experiences while balancing innovation, inclusivity, and efficiency.

II. LITERATURE REVIEW

[1]”Survey on Virtual Try-On and Fashion AI“,Aakash V B, Abdulla Akmal, Amrutha Chandrababu, Ashwin Jose, Alka Vijay (title in International Journal of Advances in Engineering and Management, Nov 2024). Authors review 10 studies, referencing datasets like Kaggle (40,000 photos) and DeepFashion2. Methods include CNN, ResNet50, YOLOv5, Inception-v3, and AR-based try-ons. Reported accuracies range 89.7% garment recognition to 96% YOLOv5 item detection. Challenges: lighting variation, real-time optimization. Future scope: hybrid systems integrating social media and inclusive body representation.

[2]”Survey of AI in Fashion with Capsule Networks”(Journal of Artificial Intelligence and Capsule Networks, Nov 2021). Reviews 110 articles, covering GANs, CNNs, and evaluation metrics like FID, IS. Reported accuracies vary, e.g., 76.62% in body-fitting. Challenges: outperforming humans, computational and energy efficiency, and lack of a standard evaluation metric. Concludes that the field is highly active for fashion e-commerce applications.

[3]“FashionFit: Analysis of Mapping 3D Pose and Neural Body Fit for Custom Virtual Try-On” by Hashmi, Kumar, Keskar, Bokde & Geem, published in IEEE Access (May 2020). Combines pose estimation, segmentation, Neural Body Fit (NBF), and GANs. Uses datasets DeepFashion2, MVC, MPII Pose, LIP. Reports 99.78% training patch accuracy and 97.73% testing accuracy in segmentation. Challenge: low accuracy for custom outfits.

[4]“Effects of 3D Virtual Try-On on Online Sales and Customers’ Purchasing Experiences” by Hwangbo, Kim, Lee & Jang, published in IEEE Access (Sept 2020). Uses actual customer data from L brand to evaluate sales impact. Method: 3D body models of different shapes/sizes. Findings: +14,000 won ($13) average sales/customer and 27% fewer returns. Challenges not stated, but study highlights real-world commercial benefits.

[5]”Multimodal Fashion Try-On System”(International Journal of Recent Research and Review, 2025). Integrates ML, computer vision, NLP, psychographics, and virtual try-on. Uses BERT for NLP, ResNet-50 for vision, Random Forest classifier, OpenPose/BlazePose for pose, MediaPipe for face mesh, and VITON, TryOnGAN, ClothFlow for try-on. Dataset: user inputs & images. Achieved >90% precision with clear images. Challenges: image quality, privacy, ethics, and high computational cost.

[6]“CP-VTON+: Clothing Shape and Texture Preserving Image-Based Virtual Try-On” by Minar, Tuan, Ahn, Rosin & Lai, uses the VITON dataset. Proposes an improved two-stage pipeline with Geometric Matching Module (GMM) and Try-On Module (TOM). Outperforms CP-VTON on metrics like IoU, SSIM, LPIPS, IS. Challenges: handling diverse poses, occlusions, and complex clothing. Suggests 3D approaches for better results.

[7]“CloTH-VTON: Clothing 3D Reconstruction for Hybrid Virtual Try-On” by Minar & Ahn, published in ACCV 2020. Proposes a hybrid pipeline combining 2D + 3D methods. Uses 3D SMPL body model, U-Net, Pix2PixHD to reconstruct 3D clothing from a single 2D image. Dataset not explicitly named. Accuracy not provided, but results are high-resolution and detailed. Challenges: 2D methods struggle with deformation, occlusion, and complex textures.

[8]"Multiple Pose Virtual Try-On Based on 3D Clothing Reconstruction (3D-MPVTON)" by Tuan, Minar, Ahn & Wainwright, published in IEEE Access (Aug 2021). Extends CloTH-VTON+ for multi-pose try-on. Uses 3D reconstruction pipeline with a new equalized entropy loss for better segmentation and faster training. Dataset not specified. Outperforms 2D-based methods, especially for occlusion and viewpoint changes. Challenges: clothing folding, self-occlusion, and complex pose variation.

[9]"Hierarchical Cross-Attention Network for Virtual Try-On" a two-stage framework combining geometric matching and try-on synthesis with a novel Hierarchical Cross-Attention (HCA) block. This block models long-range dependencies between clothing and person features for realistic virtual try-on results. It integrates a trainable thin-plate spline (TPS) transformation and a cross-modal attention mechanism, ensuring high texture preservation and pose alignment. Quantitative results outperform CP-VTON, CP-VTON+, and ACGPN, achieving IoU = 0.8662, SSIM = 0.859, LPIPS = 0.102, and FID = 8.25, establishing new state-of-the-art benchmarks.

[10]"Shining Yourself: High-Fidelity Ornaments Virtual Try-On with Diffusion Model" (CVPR 2024, Xi'an Jiaotong University & SenseTime Research) Yingmao Miao et al. introduce the first diffusion-based framework for ornament try-on, targeting accessories such as bracelets, rings, earrings, and necklaces. The system uses an iterative pose-aware mask prediction scheme to ensure precise alignment with body parts (like the wrist or neck) from only a coarse bounding box, and employs a mask-guided attention module to preserve high-fidelity identity and geometric details.

Title / Source	**Dataset Used**	**Method / Approach**	**Accuracy / Results**	**Challenges**
Survey on Virtual Try-On and Fashion AI (IJ of Advances in Engineering and Management, Nov 2024)	Kaggle (40,000 photos), DeepFashion2, others	CNN, ResNet50, YOLOv5, Inception-v3, AR-based systems	94% outfit rec., 89.7% garment recog., 96% item detection (YOLOv5)	Lighting variation, real-time optimization
Survey of AI in Fashion with Capsule Networks (J. of AI & Capsule Networks, Nov 2021)	Multiple datasets from 110 papers	GANs, CNNs, Capsule Nets	76.62% (body-fitting), others report FID & IS	Outperforming humans, efficiency, lack of standard metrics
FashionFit (IEEE Access, May 2020)	DeepFashion2, MVC, MPII Pose, LIP	Pose estimation + segmentation + Neural Body Fit + GANs	99.78% training, 97.73% testing patch accuracy	Low accuracy with custom outfits
Effects of 3D Virtual Try-On on Sales (IEEE Access, Sept 2020)	Real customer data from L brand	3D body modeling (different shapes/sizes)	+14,000 won ($13) avg. sales per customer; 27% fewer returns	Challenges not explicitly mentioned
Multimodal Fashion Try-On System (IJRRR, 2025)	User-provided inputs & images	BERT (NLP), ResNet-50 (vision), Random Forest, OpenPose/BlazePose, MediaPipe, VITON/TryOnGAN/ClothFlow	>90% precision with clear images	Image/text quality, privacy & ethics, high computation cost
CP-VTON+ (Authors: Minar et al.)	VITON	Improved 2-stage pipeline (GMM + TOM)	Outperforms CP-VTON on IoU, SSIM, LPIPS, IS	Struggles with diverse poses, occlusion, complex clothing
CloTH-VTON (ACCV 2020, Minar & Ahn)	Not specified (VITON-like)	Hybrid pipeline: 2D + 3D methods, SMPL, U-Net, Pix2PixHD	Claimed high-quality outputs with full detail	2D methods struggle with deformation, occlusion, textures
3D-MPVTON (IEEE	Not specified	Hybrid 3D reconstruction +	Outperforms 2D methods,	Self-occlusion, folding,

Access, Aug 2021, Tuan et al.)		equalized entropy loss	better segmentation, faster training	viewpoint changes
Hierarchical Cross-Attention Network for Virtual Try-On (IEEE Trans. Multimedia, 2024)	Han et al. VITON dataset (19K image pairs)	2-stage framework: Geometric Matching + Try-On; Hierarchical Cross-Attention Block (HCA)	IoU = 0.8662, SSIM = 0.859, LPIPS = 0.102, FID = 8.25	Hand-pose distortion, lighting inconsistencies
Shining Yourself: High-Fidelity Ornaments Virtual Try-On with Diffusion Model (CVPR 2024)	64K image triplets (bracelets, rings, necklaces, earrings)	Latent Diffusion + ReferenceNet; Iterative Pose-Aware Mask Refinement + Mask-Guided Attention	FID = 19.00, LPIPS = 0.0593, CLIP = 88.7	Specular reflection bias, orientation alignment issues

III. METHODOLOGY

1. Custom Long Short-Term Memory Model

LSTMs embody a particular architecture configured towards a class of recurrent neural networks (RNNs). An RNN employs a distinct connective architecture whereby the output of a neuron is fed back into the neural network at a later stage, thus maintaining a memory of the previous output. Order dependence is a major paradigm in RNNs, and LSTMs demonstrate the capability of learning such characteristics akin to RNNs, thus useful in credit card transaction anomaly detection. Another distinctive aspect of LSTMs and patterns of neural networks is the former's ability to retain patterns and remember ideas, which is essential in case of long sequences. Regardless, LSTMs can be costly and inefficient in domains requiring less computation and effort aimed toward trimming memory burden associated with model construction and pattern retaining. And, like other models, LSTMs do not work with streams of data and can only operate with static data.

2. Hidden Markov Model (HMM)

Markov processes are assumed in a statistical model that describes systems where one or more parameters change over time in a continuous manner, describing the Markov process in an BY an unobserved (hidden) states. The model derives the probabilistic distributions of the latent states with the resultant observations proffered these states. The context of the documents states that shocked card payments fraud detection systems with streams and real-time analytics can incorporate HMMs.

3. Haar Cascade Classifier

Unlike other frameworks, this one described an algorithm as a system that works within the probabilistic boundary of machine learning. Rather, it is used mainly for the real time detection of faces, but can also detect other parts of the body, such as eyes, noses, and other body parts . An object is a complex set of features known as Haar features, and for feature-based objects, it uses a cascade of classifiers to eliminate non-objects and concentrate its efforts on truly potential objects; this feature makes the process very efficient .

4. Naive Bayes

This is an algorithm which classifies data based on Bayes' theorem and considers every feature of a dataset as independent of each other. Even with such a naive assumption, it is still one of the fastest and most effective algorithms for an entire host of problems, which include text classification and recommendation systems. It estimates the class a particular data point belongs to, based on the estimates of its attributes.

5. Convolutional Neural Networks (CNN)

CNN is a type of deep learning architecture which is most commonly applied for studying spatial data. It has a deep structure of several layers which processes some data for example images by applying filter like functions to learn and extract the relevant attributes. That is why it is used for face recognition, image classification and object recognition. The network has the ability to learn and extract complex patterns and attributes from the raw data.

6. BlazePose

BlazePose clearly identifies limbs and joints and builds a skeletal model according to each detected position. The model was exposed to motion capture videos and learned to follow a human model in real-time. Easily, coordinates of all major body structures can be accessed allowing one to constantly know, standing still in a zero motion exercise, what body parts and areas they are in. Application fields, from physical exercises and sports to virtual fitting rooms and human-computer interaction, are virtually unlimited.

7. VibeMe

VibeMe also generates recommendations. However, a user does not just have to provide a picture of themselves in order to receive stylized sets. The AI analyzes various physical and psychological traits of the user and interprets the information to match the current mood and the occasion, contextually generating sets of clothes with proper reasoning. VibeMe leverages sophisticated machine learning and computer vision methodologies as well as psychography to understand user behavior in order to generate optimal sets.

8. CP-VTON+

Many of the issues in older models, in which the user ends up with clothes in a virtual fitting room with badly shaped or structurally malfunctioning outfits, stem from the user's image and the garment not integrating. The garment mismatches because the clothes lose detail and shape during the model's process of filtering and extraction. Setting the clothes onto the base image happens after filtering, where detail and shape are retained, in the two-step process the user model matches with the garment in the first stage. The user image and garment are blended in the second.

9. CloTH-VTON

This hybrid framework for virtual try-on technology leverages both 2D-image and 3D-model approaches. It uses cloth reconstruction techniques to transform 2D images into virtual 3D garments and then drapes the reconstructed clothing on a target person in a specific pose. This enables high-resolution try-on with complex, naturally deformed clothing while preserving fin detail.

10. 3D-MPVTON

This system pertains to 3D clothing deformation for multiple-pose virtual try-on scenarios where pre-existing 2D clothing shape transformation techniques are incapable. It employs a 3D clothing model reconstruction approach, which works best in multi-pose scenarios. The framework builds upon the principles of CloTH-VTON+.

11. Hierarchical Cross-Attention Network (HCANet)

The HCANet is a novel solution for garment VTON that follows the established two-stage architecture: Geometric Matching and Try-On. Its innovation lies in the Hierarchical Cross-Attention (HCA) block, which is incorporated into both stages. This block is specifically designed to capture long-range global interactive correlations between the person's various representations (pose heatmaps, body shape mask, reserved regions) and the in-shop clothing modality. This deep, synergistic fusion of information addresses the problem of inconsistent appearance and floating clothing seen in earlier 2D methods.

12. Latent Diffusion Model (LDM)

Shining Yourself (Ornament VTON) system is the first to address virtual try-on for ornaments, built upon the Latent Diffusion Model (LDM) and ReferenceNet. It operates as a zero-shot method, taking a reference ornament, a target person image, and a coarse bounding box as input. Iteratively predicts accurate wearing masks from coarse bounding boxes to align pose and scale between the ornament and body. Enforces mapping between reference and wearing masks to preserve geometry of tiny objects (rings, beads).

IV. RESULT AND DISCUSSION

The results demonstrate that virtual try-on technology has delivered compelling outcomes, both in terms of technical accuracy and commercial viability, with studies indicating a significant positive impact on the retail sector. Multiple models have achieved impressive performance metrics in specific tasks. For example, a "Survey on Virtual Try-On and Fashion AI" reported a notable 89.7% garment recognition accuracy and a 96% item detection accuracy with the use of YOLO. The "FashionFit" paper, which integrates pose estimation and Neural Body Fit, shows a remarkably high training patch accuracy of 99.78% and a testing accuracy of 97.73% in segmentation, showcasing the effectiveness of its combined approach. Furthermore, a Multimodal Fashion Try-On System achieved over 90% precision, particularly with clear images. From a commercial standpoint, a study titled "Effects of 3D Virtual Try-On on Online Sales and Customers' Purchasing Experiences" used real-world data from a brand to reveal substantial benefits, including an average sales increase of +14,000 won (approximately $13) per customer and a 27% reduction in product returns, highlighting the technology's ability to boost consumer confidence and satisfaction. The evolution of these systems is also evident in frameworks like, CP-VTON+, which is explicitly noted for outperforming its predecessor, CP-VTON, on key metrics such as IoU, SSIM, LPIPS, and IS. Similarly, the 3D-MPVTON system, which builds upon the principles of CloTH-VTON+, demonstrates its superiority by outperforming 2D-based methods, particularly when handling complex scenarios involving occlusion and viewpoint changes. The paper also mentions that the multimodal system leverages sophisticated machine learning, computer vision, and psychography to generate optimal sets, indicating a move toward more personalized and sophisticated user experiences. The results collectively underscore the technical maturity and commercial potential of this field, positioning it as a critical step toward reshaping the digital shopping experience. HCANet (Hierarchical Cross-Attention Network) has set a new state-of-the-art benchmark by introducing a two-stage generative process with geometric matching and try-on phases. Through its novel Hierarchical Cross-Attention block, HCANet effectively captures long-range global correlations between the person and clothing modalities, achieving significant improvements across multiple metrics—IoU (0.8662), SSIM (0.859), LPIPS (0.102), and FID (8.25)—thus ensuring highly realistic, pose-coherent, and texture-consistent virtual try-on results. Additionally, Shining Yourself, the first diffusion-based framework for virtual try-on of ornaments such as bracelets, rings, earrings, and necklaces, introduced pose-aware mask refinement and mask-guided attention mechanisms, yielding the best FID (19.00) and CLIP-similarity scores among competing methods. Its ability to maintain high-fidelity geometric details and identity preservation demonstrates the expanding versatility of virtual try-on beyond garments into accessories and other wearable domains. The results collectively underscore the technical maturity and commercial potential of this field, positioning it as a critical step toward reshaping the digital shopping experience.

Despite these significant advancements, the literature review also reveals a number of critical challenges and areas for improvement that the field must address to fully realize its potential. A primary technical hurdle is ensuring high accuracy for custom outfits, a challenge specifically noted in the "FashionFit" study. Models frequently struggle with various real-world factors, including lighting variations and the need for real-time optimization. Additionally, handling the intricate details of complex clothing, such as folding and textures, remains a significant challenge for many systems, especially those using 2D methods. The paper also points out that the field must move beyond standard garments to ensure inclusive representation of diverse body shapes and sizes, a critical challenge highlighted in the introduction. On a broader scale, high computational costs and energy inefficiency are notable barriers to widespread adoption. The lack of a standard evaluation metric is a key methodological limitation, making it difficult for researchers to consistently compare the performance of different models. Furthermore, ethical and data-related concerns, such as ensuring image and text quality, protecting user privacy, and navigating the ethical implications of the technology, are also critical issues that need to be carefully considered. The discussion also brings up the limitation that some models, like LSTMs, can only operate with static data and not data streams, which limits their application in real-time scenarios. These challenges underscore the fact that while virtual try-on technology has made impressive strides.

The future scope of virtual try-on technology is centered on addressing current challenges to create more advanced, inclusive, and efficient systems. One key area for future development is the creation of hybrid systems that integrate social media and ensure inclusive body representation. This is particularly important as a critical challenge is moving beyond standard garments to achieve high accuracy for custom outfits and representing diverse body shapes and sizes. The field is highly active, with significant potential for fashion e-commerce applications. Future advancements will also involve resolving existing technical hurdles, such as lighting variation and the need for real-time optimization. The document also suggests that moving toward 3D approaches will yield better results, especially in handling issues like deformation, occlusion, and complex textures. The document also suggests that moving toward 3D approaches will yield better results, especially in handling issues like deformation, occlusion, and complex textures. Moreover, integrating cross-modal attention mechanisms (as in HCANet) and iterative mask refinement (as in Shining Yourself) represents a promising direction for achieving seamless, high-fidelity results in both garments and accessories. Ultimately, advancing virtual try-on technologies is viewed as a critical step toward reshaping digital shopping experiences by balancing innovation, inclusivity, and efficiency.

V. CONCLUSION

The collective review of these research papers on Virtual Try-On (VTON) highlights the remarkable progress the field has made in bridging the gap between physical and digital fashion experiences, while also emphasizing persistent challenges that need to be addressed. Across the surveyed works, it is evident that both 2D and 3D approaches, powered by advanced deep learning models such as CNNs, GANs, ResNet, Capsule Networks, and hybrid pipelines, have significantly improved the realism, accuracy, and adaptability of try-on systems. Models like CP-VTON+ and CloTH-VTON demonstrated how geometric matching, hybrid 2D–3D pipelines, and enhanced reconstruction techniques outperform traditional methods in handling shape, texture, and pose variation. Likewise, studies incorporating pose estimation (FashionFit) and multimodal inputs (NLP, vision, and psychographics in the Multimodal Fashion Try-On System) extend the boundaries of what VTON systems can achieve, pointing toward more personalized and context-aware solutions.

Commercially, the research underscores the tangible business value of VTON systems, with real-world evidence such as improved sales and reduced return rates, confirming their relevance beyond academic experimentation. However, several consistent challenges remain: handling complex clothing textures and poses, self-occlusion, ensuring inclusivity across diverse body types, addressing computational costs, and developing standardized evaluation metrics. Additionally, ethical considerations such as privacy and responsible data use will become increasingly important as these systems move toward mass adoption.

In conclusion, Virtual Try-On technology has evolved from simple 2D warping methods to sophisticated hybrid 3D pipelines capable of producing high-resolution, photorealistic results. While accuracy and user personalization continue to improve, future research must focus on scalability, inclusivity, standardization, and ethical integration to realize the full potential of VTON in revolutionizing the fashion retail industry. The convergence of machine learning, computer vision, natural language processing, and human-centered design indicates that the next generation of VTON will not only enhance online shopping experiences but also redefine consumer engagement and trust in digital fashion ecosystems.

REFERENCES

[1] *Aakash V B, Abdulla Akmal, Amrutha Chandrababu, Ashwin Jose, Alka Vijay (2024) CP-VTON+: Clothing shape and texture preserving image-based virtual try-on.*

[2] Seyed Omid Mohammadi , Ahmad Kalhor (2021). Smart fashion: A review of AI applications in virtual try-on & fashion synthesis. Journal of Artificial Intelligence and Capsule Networks, 3(4), 284–304.

[3] Mohammad Farukh Hashmi , (Member, Ieee), B. Kiran Kumar Ashish, Avinash G. Keskar,Neeraj Dhanraj Bokde , Zong Woo Geem , (Member, Ieee) (2020). FashionFit: Analysis of mapping 3D pose and neural body fit for custom virtual try-on. IEEE Access, 8, 91603–91615.

[4] Hyunwoo Hwangbo , Eun Hie Kim , So-Hyun Lee, Young Jae Jang , (Member, Ieee) (2020). Effects of 3D virtual "try-on" on online sales and customers' purchasing experiences. IEEE Access, 8, 189479–189489.

[5] Vimal Daga, Preeti Daga, Purwanshi Mishra (2025). Vibeme: an ai stylist, wardrobe curator, and style expert for every body and mood with virtual try-on.

[6] Matiur Rahman Minar , Thai Thanh Tuan , Heejune Ahn , Paul L. Rosin , and Yu-Kun Lai (2025). CP-VTON: Clothing shape and texture preserving image-based virtual try-on.

[7] Heejune Ahn , Matiur Rahman Minar (2025). CloTH-VTON: Clothing Three-dimensional reconstruction for Hybrid image-based Virtual Try-ON. Seoul National University of Science and Technology, Seoul, South Korea.

[8] Thai Thanh Tuan , Matiur Rahman Minar , Heejune Ahn , John Wainwright (2021). Multiple pose virtual try-on based on 3D clothing reconstruction. IEEE Access, 9, 114367–114380.

[9] Hao Tang, Bin Ren, Pingping Wu, Nicu Sebe (2024) . Hierarchical Cross-Attention Network for Virtual Try-On

[10] Yingmao Miao, Zhanpeng Huang , Rui Han , Zibin Wang , Chenhao Lin, Chao Shen (2023). Shining Yourself: High-Fidelity Ornaments Virtual Try-on with Diffusion Model

