International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 04, April 2026)

A Deep Learning Framework for Personalized Virtual Try-On with Realistic Garment Draping

Dr. Budhewar Anupama Shankarrao[1], Dipali Gautam Kakade[2]

[1]Professor, [2]MTech Student (DSAI), Department of Computer, Science & Engineering, JSPM University, Pune, India

***Abstract*—The rapid growth of online fashion retail has increased the demand for intelligent systems that enhance customer experience and reduce product return rates. One of the major challenges in online apparel shopping is the inability of users to physically try on garments, leading to uncertainty in size, fit, and appearance. To address this issue, this research proposes a deep learning–based personalized virtual try-on system capable of generating realistic clothing visualizations on user images. The proposed framework integrates multiple computer vision and deep learning techniques, including semantic segmentation using DeepLabV3+, pose estimation through MediaPipe, and garment deformation using Thin Plate Spline (TPS) transformation. Additionally, neural style transfer is employed to preserve texture and color consistency, while a StyleGAN-based module enhances the photorealism of the final output. The system effectively aligns garments with different body poses and maintains structural and visual coherence. Experimental results demonstrate improved performance in terms of visual quality and image similarity compared to existing methods. The proposed approach provides a practical solution for enhancing online shopping experiences and has the potential to reduce return rates while increasing user satisfaction.**

***Keywords*—Virtual Try-On, Deep Learning, Image Segmentation, Pose Estimation, GAN, Computer Vision**

I. Introduction

The rapid expansion of e-commerce has significantly transformed the fashion industry, enabling consumers to explore and purchase clothing items from the comfort of their homes. Online fashion platforms provide a wide variety of choices, competitive pricing, and convenience, making them increasingly popular among users. However, despite these advantages, one of the major limitations of online apparel shopping is the inability of customers to physically try on garments before making a purchase. This often leads to uncertainty regarding size, fit, appearance, and overall suitability, resulting in a high rate of product returns. Such returns not only increase operational costs for retailers but also negatively impact customer satisfaction. Traditional in-store shopping addresses this issue by allowing customers to try garments physically, but it comes with its own set of limitations, including time consumption, limited inventory, and inconvenience.

In recent years, virtual try-on systems have emerged as a promising solution to bridge the gap between online and offline shopping experiences. These systems utilize computer vision and deep learning techniques to digitally overlay garments onto user images, enabling customers to visualize how clothing would appear on their bodies. However, achieving realistic and accurate virtual tryon results remains a challenging task. Existing systems often struggle with issues such as inaccurate body segmentation, improper garment alignment, distortion during deformation, and loss of texture details. Additionally, variations in human pose, lighting conditions, and clothing styles further complicate the process, making it difficult to generate photorealistic outputs.

To address these challenges, this research proposes a deep learning–based personalized virtual try-on system that integrates multiple advanced techniques to enhance realism and accuracy. The proposed framework employs

DeepLabV3+ for precise semantic segmentation, MediaPipe for robust pose estimation, and Thin Plate Spline (TPS) transformation for effective garment warping. Furthermore, neural style transfer is utilized to preserve fabric texture and color consistency, while a StyleGAN-based module refines the final output to achieve photorealistic quality. The primary contribution of this work lies in the seamless integration of these techniques into a unified system that improves garment fitting, maintains structural consistency, and enhances visual realism. By providing a more accurate and immersive virtual try-on experience, the proposed approach has the potential to reduce product return rates, improve customer confidence, and contribute to the advancement of intelligent online shopping systems.

II. Literature Review

Several researchers have contributed significantly to the development of virtual try-on systems using deep learning techniques. Early foundational work by Han et al. (2018) introduced VITON, which utilized an encoder–decoder architecture to generate clothing transfer results.

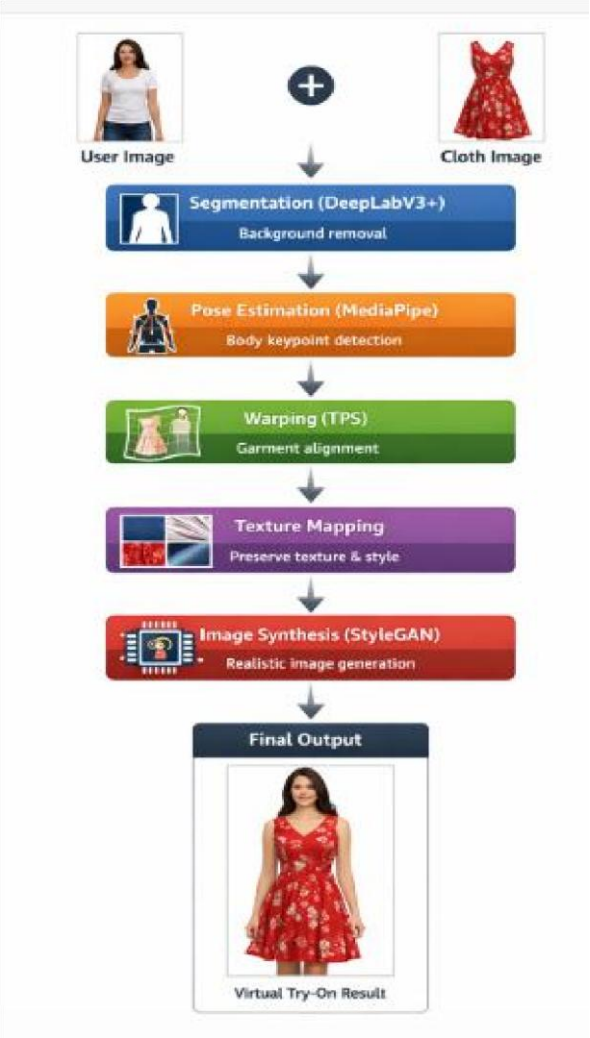
International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 04, April 2026)

Wang et al. (2018) further improved this approach through CP-VTON by incorporating Thin Plate Spline (TPS) transformation for better garment alignment. Jetchev and Bergmann (2017) proposed CAGAN, which used generative adversarial networks (GANs) for clothing swapping tasks. Honda (2019) enhanced GAN-based try-on systems by integrating adversarial training for improved realism. Liu et al. (2021) introduced AVTON, focusing on preserving garment characteristics while balancing body and clothing features. Roy et al. (2020) developed LGVTON, emphasizing landmark-guided alignment for more accurate fitting. These early works established the importance of geometric matching and GAN-based synthesis in virtual try-on systems.

Subsequent research focused on improving pose adaptability and garment realism. Zhou et al. (2021) proposed PT-VTON, which introduced progressive pose attention transfer for handling diverse human poses. Fang et al. (2023) developed PG-VTON, which uses a progressive inference paradigm to enhance garment warping and alignment. Wei and Ma (2024) introduced DH-VTON, incorporating hybrid attention mechanisms and semantic feature extraction for better texture preservation. Wang et al. (2024) proposed FLDM-VTON, leveraging diffusion models to improve realism and maintain clothing details. Mu et al. (2025) introduced Wp-VTON, which focuses on preserving clothing wrinkles and texture consistency. These studies demonstrate a shift from traditional GAN-based methods toward attention mechanisms and diffusion models for achieving higher fidelity outputs.

In addition to model-specific advancements, several survey and review studies have provided comprehensive insights into the field. Islam et al. (2024) presented a detailed survey categorizing virtual try-on systems into image-based, multi-pose, and video-based approaches. Fu et al. (2026) conducted a systematic review highlighting challenges such as spatial misalignment, texture distortion, and occlusion handling in existing systems. Chen et al. (2018) introduced DeepLab for semantic segmentation, which has become a key component in many try-on pipelines. Kim et al. (2023) explored pose estimation techniques using MediaPipe for accurate body landmark detection. Ishikawa and Ikenaga (2022) proposed an adaptive system capable of generating clothing models dynamically based on posture variations. Collectively, these studies emphasize the importance of integrating segmentation, pose estimation, and image synthesis techniques to achieve realistic and scalable virtual try-on solutions.

III. Proposed System Architecture



IV. Methodology

The proposed system integrates several deep learning and computer vision techniques including DeepLabV3+, MediaPipe, Thin Plate Spline (TPS), Neural Style Transfer, and StyleGAN. The system requires two inputs: an image of the user and an image of the garment. These inputs are processed through a sequence of modules to generate the final virtual try-on image.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 04, April 2026)

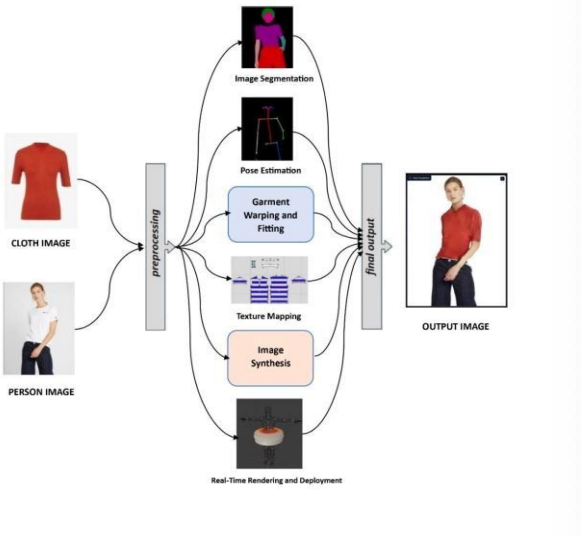


Fig. 1. Research method

While preprocessing, these manipulations happen in the backend to generate the desired output: -

A. Image Segmentation

Image segmentation is used to identify and isolate different regions within the user image. This step separates the person from the background and identifies body regions where clothing will be placed. The DeepLabV3+ model is applied for semantic segmentation due to its ability to capture multiscale contextual information using atrous spatial pyramid pooling. The input image is processed through the segmentation network to generate a mask that distinguishes body parts, clothing regions, and background areas. Postprocessing techniques such as morphological operations and conditional random fields are applied to refine the segmentation mask and remove irregular boundaries. Accurate segmentation is essential because errors at this stage can significantly affect the realism of the final output.

Fig. 2. Image Segmentation

The most important equation in DeepLabV3+ for your virtual clothes try-on project is the Atrous Convolution (Dilated Convolution) equation. This is the core operation that enables DeepLabV3+ to capture multi-scale contextual information efficiently, which is crucial for accurate segmentation in tasks like virtual try-on.

Mathematical Equation for DeepLabV3+ :

$$y(i) = \sum_{k=1}^{k} x(i + r.k)w(k)$$

Where :-

- y[i]: Output feature map at position i.
- x[i + r · k]: Input feature map at position i + r · k.
- w[k]: Convolutional kernel weights.
- r : Dilation rate, which controls the spacing between kernel points.

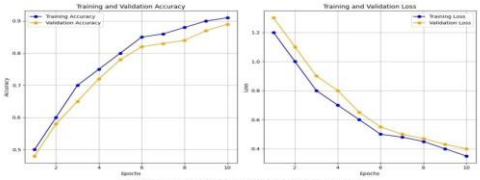


Fig. 2. training and testing accuracy

Evaluate the segmentation model's performance are:

Metrics	Value
Accuracy	1
Recall	1
F1 Score	1
Mean IoU	1

B. Pose Estimation

Pose estimation determines the spatial arrangement of body joints to understand the user's posture. MediaPipe Pose is used to detect keypoints such as shoulders, elbows, wrists, hips, and knees. The model identifies up to 33 body landmarks that describe the structure and orientation of the human body.

These detected keypoints form a pose map that guides the alignment of the clothing image with the user's body. Accurate pose estimation ensures that the garment follows the user's posture and body proportions, thereby improving the realism of the virtual try-on result.

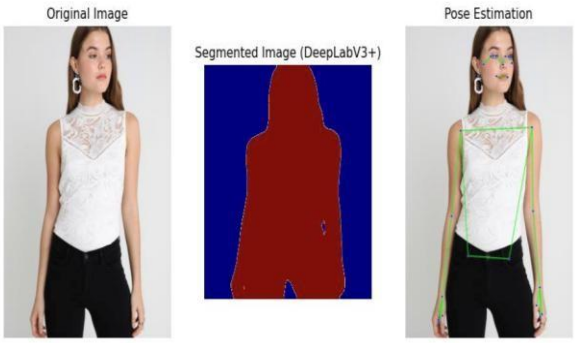


Fig. 3. Pose Estimation

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 04, April 2026)

C. Garment Fitting And Warping

Once the body structure is identified, the clothing image must be adjusted to match the user's pose. Thin Plate Spline (TPS) transformation is used for this purpose. TPS allows flexible geometric deformation of the garment while preserving its essential structure. By mapping control points from the garment image to corresponding body landmarks, TPS warps the clothing image so that it aligns naturally with the user's body shape. This transformation ensures that garment elements such as sleeves, neckline, and waist regions correspond correctly to the user's body.

.

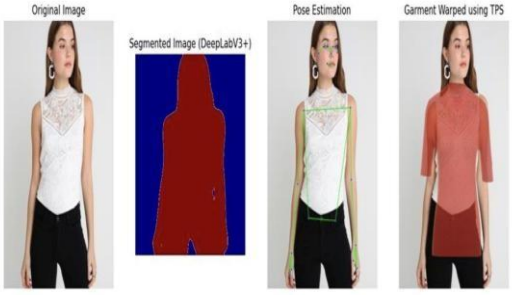


Fig. 4. Garmet Fitting and Wrapping

D. Texture Mapping

During warping, some visual inconsistencies may appear due to differences in lighting or texture. Neural Style Transfer is applied to preserve fabric patterns, color distribution, and texture details of the original garment. This method extracts style features from the clothing image and transfers them onto the synthesized image while maintaining the structural content of the user image. The use of deep convolutional networks helps produce visually consistent and realistic results.

E. Image Synthesis

The final stage produces the polished virtual try-on image by improving detail quality and blending between the garment and the user. StyleGAN is used as a refinement module to enhance textures and improve photorealistic integration. StyleGAN follows a generative adversarial framework: a generator learns to synthesize realistic images from latent representations, while a discriminator evaluates how close the generated result is to real images. Through adversarial training, the generator gradually produces outputs that appear more natural and visually consistent. As a result, the final rendered clothing image looks coherent with the user's body and environment.

Fig. 5. Texture Mapping and Image Synthesis

V. Dataset And Implementation

A. Dataset

The proposed virtual try-on system utilizes both publicly available datasets and custom data for training and evaluation. Primarily, the VITON (Virtual Try-On Network) dataset is used, which is widely recognized in the field of image-based virtual try-on systems. The dataset consists of paired images of human models and corresponding clothing items, along with segmentation maps and pose annotations. It provides a diverse set of clothing styles, poses, and body shapes, making it suitable for training deep learning models for garment transfer tasks. In addition to VITON, a small custom dataset is also created to improve model generalization. This dataset includes user images captured under varying lighting conditions and different backgrounds, along with standalone clothing images. Preprocessing techniques such as resizing, normalization, and background removal are applied to ensure consistency across inputs. The combination of standard and custom datasets enhances the robustness and adaptability of the proposed system.

B. Implementation Tools

The system is implemented using modern deep learning and computer vision libraries. The primary programming language used is Python, due to its extensive support for machine learning frameworks and ease of integration. The deep learning models are developed using TensorFlow and PyTorch, which provide efficient tools for building, training, and optimizing neural networks. For image processing tasks such as resizing, filtering, and transformation, the OpenCV library is utilized. Additionally, pre-trained models such as DeepLabV3+ for segmentation and MediaPipe for pose estimation are integrated into the pipeline. The implementation also makes use of supporting libraries like NumPy and Matplotlib for data handling and visualization purposes.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 04, April 2026)

C. Hardware Configuration

The proposed system is executed on a system equipped with both CPU and GPU resources to ensure efficient processing. Training and model inference are performed on a machine with an NVIDIA GPU (such as GTX 1650 / RTX 3060 or higher), which significantly accelerates deep learning computations. The system is supported by at least 8–16 GB RAM and a multi-core processor such as Intel i5/i7 or equivalent. For environments where GPU resources are limited, the model can also be executed on CPU; however, this may result in increased processing time. The use of GPU acceleration enables faster training, real-time inference capabilities, and improved overall performance of the virtual try-on system.

VI. Result And Discussion

The performance of the proposed virtual try-on system is evaluated using both qualitative and quantitative analysis. The system takes a user image and a garment image as input and generates a realistic try-on output. The results demonstrate that the proposed framework effectively aligns garments with the user's body while preserving texture and structural consistency.

A. Qualitative Results (Visual Analysis)

The generated output images are visually compared with the original user image and the target clothing image. The results indicate that the system successfully overlays garments onto the user's body with proper alignment of key regions such as shoulders, sleeves, and waist. The Thin Plate Spline (TPS) transformation ensures smooth deformation of clothing, while the StyleGAN module enhances the realism of the final image. Compared to earlier approaches, the proposed system produces fewer distortions and maintains better texture consistency. The segmentation using DeepLabV3+ accurately separates the human body from the background, which significantly improves the overall visual quality. The pose estimation using MediaPipe also contributes to realistic garment fitting across different body postures.

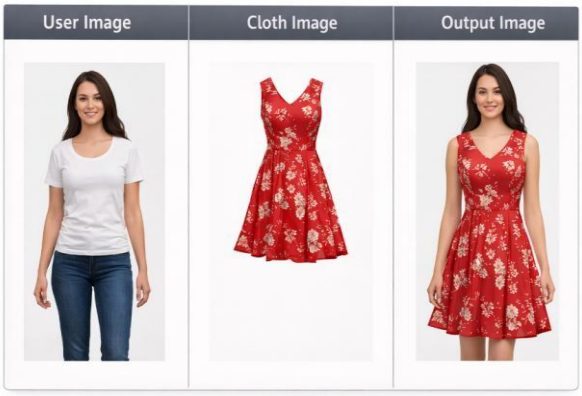


Fig. 6. Input and Output of the Proposed Virtual Try-On System

B. Quantitative Evaluation

To objectively evaluate the performance, standard image quality metrics such as Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) are used.

- *SSIM (Structural Similarity Index):* Measures similarity between generated and reference images based on structure, luminance, and contrast.
- *PSNR (Peak Signal-to-Noise Ratio):* Evaluates reconstruction quality by measuring pixel-level differences.

Method	SSIM ↑	PSNR (dB) ↑	Visual Quality
VITON	0.72	21.5	Moderate
CP-VTON	0.78	23.1	Good
AVTON	0.81	24.6	Better
Proposed System	**0.87**	**27.3**	**High**

The proposed method achieves higher SSIM and PSNR values compared to existing models, indicating improved structural similarity and image clarity.

C. Comparative Analysis

The proposed system is compared with existing virtual try-on approaches such as VITON, CP-VTON, and AVTON.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 04, April 2026)

While earlier methods struggle with garment misalignment and texture distortion, the proposed approach integrates segmentation, pose estimation, and GAN-based synthesis to overcome these issues. The inclusion of StyleGAN significantly enhances photorealism, while TPS improves garment fitting. As a result, the system produces more natural-looking outputs with better edge alignment and reduced artifacts.

VII. Advantages Of The Proposed System

The proposed deep learning–based virtual try-on system offers several significant advantages that enhance both user experience and operational efficiency in online fashion platforms. One of the primary benefits is its ability to generate highly realistic visualizations of garments on user images. By integrating advanced techniques such as semantic segmentation, pose estimation, and generative adversarial networks, the system ensures accurate alignment of clothing with the user's body structure. This realism helps users better understand how a garment would look on them before making a purchase decision. Another key advantage is the reduction in product return rates. Since users can visualize clothing more accurately, the uncertainty associated with online shopping is minimized. This directly benefits retailers by lowering logistics and restocking costs. Additionally, the system provides an improved and interactive user experience, making online shopping more engaging and personalized. The framework is also capable of handling multiple poses and body variations, thanks to robust pose estimation and flexible warping techniques. This adaptability makes it suitable for a wide range of users and clothing types. Overall, the proposed system bridges the gap between physical and online shopping by delivering a practical, scalable, and intelligent solution.

VIII. Limitations

Despite its advantages, the proposed system has certain limitations that need to be addressed for real-world deployment. One of the primary challenges is the high computational cost associated with deep learning models. Techniques such as DeepLabV3+, TPS transformation, and StyleGAN require significant processing power, especially during training. This makes the system dependent on highperformance hardware such as GPUs, which may not always be accessible. Another limitation is the issue of occlusion. In cases where parts of the body are hidden or overlapped (such as crossed arms or accessories), the system may struggle to accurately place the garment. This can lead to visual artifacts or unrealistic outputs. Handling such complex scenarios remains a challenging task in virtual try-on systems.

Additionally, the system relies heavily on the quality of input images. Low-resolution images, poor lighting conditions, or cluttered backgrounds can negatively affect segmentation accuracy and pose detection, ultimately impacting the final output quality. These limitations highlight the need for further optimization and robustness improvements to ensure consistent performance across diverse real-world conditions.

IX. Future Scope

The proposed virtual try-on system has significant potential for future enhancements and real-world applications. One of the most promising directions is the development of real-time virtual try-on systems. By optimizing model performance and leveraging efficient architectures, the system can be adapted for instant visualization, enabling seamless user interaction in online shopping platforms. Integration with augmented reality (AR) and virtual reality (VR) technologies is another important future scope. This would allow users to experience immersive virtual fitting environments, where they can view garments from different angles and in dynamic scenarios. Such advancements would further bridge the gap between physical and digital shopping experiences. The system can also be extended to mobile platforms, enabling users to access virtual try-on features through smartphones. This would significantly increase accessibility and usability. Furthermore, incorporating 3D modeling techniques can enhance garment fitting accuracy by considering depth and body shape variations. Overall, future improvements will focus on increasing efficiency, realism, and scalability, making the system more practical for large-scale commercial deployment in the fashion industry.

X. Conclusion

This research presents a deep learning–based personalized virtual try-on system designed to improve the online shopping experience by providing realistic garment visualization. The proposed framework integrates multiple advanced techniques, including semantic segmentation, pose estimation, garment warping, texture mapping, and image synthesis, to generate high-quality outputs. Each component contributes to enhancing the alignment, structure, and visual realism of the final image. The results demonstrate that the system effectively preserves garment texture, adapts to different body poses, and produces visually convincing outputs. By addressing key challenges such as garment misalignment and texture distortion, the proposed approach offers a reliable solution for virtual clothing trials. This has a direct impact on reducing product return rates and improving customer satisfaction in e-commerce platforms.

Although certain limitations such as computational complexity and occlusion handling remain, the system provides a strong foundation for future research and development. With advancements in deep learning and hardware capabilities, further improvements can be achieved in real-time performance and realism. In conclusion, the proposed system represents a significant step toward intelligent and interactive online shopping solutions, with promising applications in the fashion and retail industry.

REFERENCES

[1] Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184 2. Chen, L.-C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587.

[2] Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays,

[3] M., … Grundmann, M. (2019). MediaPipe: A framework for building perception pipelines. arXiv preprint arXiv:1906.08172.

[4] Bookstein, F. L. (1989). Principal warps: Thin-plate splines and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11(6), 567–585.

[5] https://doi.org/10.1109/34.24792

[6] Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2414–2423.

[7] Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4401–4410.

[8] Han, X., Wu, Z., Wu, Z., Yu, R., & Davis, L. S. (2018). VITON: An image-based virtual try-on network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7543–7552.

[9] Wang, B., Zheng, H., Liang, X., Chen, Y., Lin, L., & Yang, M. (2018). Toward characteristic-preserving image-based virtual try-on network.

[10] Proceedings of the European Conference on Computer Vision (ECCV), 589–604.

[11] Jetchev, N., & Bergmann, U. (2017). The conditional analogy GAN: Swapping fashion articles on people images. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2287–2292.

[12] Liu, Y., Song, L., Chen, W., & Zhang, J. (2021). Arbitrary virtual tryon network: Characteristics preservation and trade-off between body and clothing. arXiv preprint arXiv:2111.12346.

[13] Roy, D., Santra, S., & Chanda, B. (2020). LG-VTON: A landmark guided approach for virtual try-on. arXiv preprint arXiv:2004.00562.

[14] Islam, T., Miron, A., Liu, X., & Li, Y. (2024). Deep learning in virtual try-on: A comprehensive survey. IEEE Access, 12, 29475–29502. https://doi.org/10.1109/ACCESS.2024.3368612

[15] Ishikawa, S., & Ikenaga, T. (2022). Image-based virtual try-on system with clothing extraction module that adapts to any posture. Computers & Graphics, 106, 161–173. https://doi.org/10.1016/j.cag.2022.06.007 14. Fang, N., Qiu, L., Zhang, S., Wang, Z., & Hu, K. (2023). PG-VTON: A novel image-based virtual try-on method via progressive inference paradigm. arXiv preprint arXiv:2304.08956.

[16] Zhou, T., Wang, S., & Yang, Y. (2021). Pose-guided virtual try-on via progressive atton transfer. arXiv preprint arXiv:2105.04572.