Fashion Try-On using Poly-GAN

Shruti Nair[1], Rajvee Pisey[2], Meera Narale[3], Dr. Mrudul Dixit[4]
[1, 2, 3, 4]Department of Electronics and Telecommunications Engineering
MKSSS's Cummins College of Engineering for Women, Pune, India
[1]shruti.nair@cumminscollege.in, [2]rajvee.pisey@cumminscollege.in, [3]meera.narale@cumminscollege.in,
[4]mrudul.dixit@cumminscollege.in
1, 2, 3 - Students 4 - Assistant Professor and Dean Alumni

Abstract*: Virtual clothing try-on is in major demand in recent years, with the objective of transferring an input garment image onto a reference person. The goal of the typical try-on assignment is to match the target clothing item with the provided individual's physique and therefore present the person's try-on look. On the basis of different generative adversarial networks, a proposal is made for posture-guided virtual try-on method to attain this objective. Experiments would be conducted on the VITON dataset and a few of our own images. It features pairs of images; each pair comprises of a women's front-view and an image of top piece of clothes image. The Structural Similarity Index Measure (SSIM) will be used to assess the accuracy of the model. New outfits are generated onto existing human model images while retaining the body shape and pose of the wearer.

Keywords: Poly-GAN, Generative adversarial networks, Try-On, Fashion, Clothes, Models

I. INTRODUCTION

Virtual clothing try-on is in major demand in recent years, with the intent of moving an ideal apparel picture onto a reference individual. The objective of the typical try-on assignment is to match the target clothing item with the gave person's build and thus present a try-on look of the individual. Practically speaking, however, individuals might be more intrigued by their take a try-on looks with different postures.

Based on generative adversarial networks (GANs), the objective to suggest a posture directed virtual try-on strategy in light of to accomplish. By first foreseeing the semantic layout of the reference image, determining whether it will be changed after try-on.

Two models comprise GAN, G and D which is the generative model and discriminative model respectively. The G model is responsible in capturing the information distribution, whereas the D model predicts whether a sample was selected from the training data rather than the G data. G's training approach is designed to reduce the likelihood of D making a mistake. GANs are used to visualise designs, convert words to pictures, transfer artwork style, generate face images and video game scenarios, and create super-resolution images, among other things.

The model predicts semantic format of the input picture and the changes will be seen after try-on, and afterward deciding if its picture content must be created or preserved by the expected format of the semantic, which will result in photorealistic try-on and rich attire details. Fashion synthesis may be a difficult process that needs the placement of a reference article of clothing on an input model in an arbitrary posture wearing a special garment.

Experiments would be conducted on the VITON dataset and a few of our own images. It features pairs of images; each pair comprises of a women's front-view and an image of top piece of clothes image. The project will demonstrate the effectiveness of our approach through Structural Similarity Index Measure (SSIM) and with generative adversarial learning, create a customer interface for developing novel attire based on the pictures of the wearer. The main motive of this project arose from our interest in the domain of Artificial Intelligence and Machine Learning to enhance our knowledge around it. Due to the COVID-19 pandemic, going to a store and trying on clothes had become a greater problem. With consumer demands changing, the stores of the future will look vastly different. Generative Adversial Network is an unsupervised modelling approach in Machine Learning. After researching more about GAN and found it to be interesting and compelling to learn new technologies. Thus, inspired to do a real-life implementation of a virtual try-on in deep learning.

Consumers may virtually try on garments using digital solutions, which not only improves their shopping experience by changing the way people browse for clothes, but is also cost effective for retailers. In recent years, virtual clothing try-on has grown in popularity, with the purpose of placing a desired clothing picture onto a reference person. However, individuals might be keener on their try-on looks with different postures.

Among various GANs available, the implementation of the TRY-ON is done using Poly-GAN. First, train PolyGAN with data, where inputs are the images of a model wearing a garment and another input garment (top only). In the following stage, to infer the appearance flows between person representation (RGB model (Parsed)) and the garment image, conjugating a mask containing the lower body clothes region (garment mask), hair and face, and the individual's body segmentation outcome, and the human posture assessment outcome as the individual representation. Then, at that point, the skeleton is utilized to generate the shape of the wrapping garments. Trained a generative module with these wrapped garments, the conserved areas on the model picture, and human posture estimates along channels as inputs to get the person image with the new outfit as our result.

II. LITERATURE SURVEY

Nilesh Pandey[1] stated that Poly-GAN is a multi-input conditioning algorithm that may be used for a variety of applications, including picture alignment, stitching, and inpainting. On the basis of the model's skeleton in any posture, Poly-GAN may conduct a spatial modification of the clothing. Using the DeepFashion dataset, the Inception Score and Structural Similarity Index yielded measurable findings.

Yuying Ge[2] stated that the "Teacher-tutor-student" knowledge distillation is a novel approach for producing profoundly photograph sensible pictures without the requirement for human parsing. The technique uses the parser-based method's artificial photos as" tutor knowledge," with the artifacts being rectified in a self-supervised way by actual" Teacher knowledge" acquired from real human photographs. Extensive evaluations show that the method outperforms previous arts approaches by a wide margin.

Shizhan Zhu[6] proposed a model for solving an image problem: generating fresh clothing for people's photos. This job appears to be critical for the offline or online retail and fashion industries. Changing the garments on people's photos isn't easy. The image created should be of good quality and free of blur. The drawback is inability to produce long sleeves on photos, such as with T-shirts.

Srivatsan Varadharajan[7] demonstrates that internal interpolation coefficients per layer can be learned using StyleGAN2 to create an experience of try-on. While encouraging, the technique falls short in postures and clothing that are underrepresented. Because interpolation assumes perfect projection, poor projection of real images has a direct impact on interpolation results.

Built on generative adversarial learning, Yevhen Pozdniakov[9] has offered a unique technique for producing new clothes on a person. The model" redresses" a person based on a picture of the person and a statement suggesting an alternative attire. At the same time, the wearer's position and attitude remain unaltered. Virtual Try-on Network and Liquid Warping GAN models were used to change the apparel on the people's photos.

Tim Salimans[11] facilitates the understanding of the instability of GAN's most successful theory. Many strategies are provided to stabilize DCGAN training. That includes matching feature, discrimination of minibatch, historical averaging and virtual batch normalization. Adding these recommendations to a basic implementation of DCGANs might be a wonderful way to learn more about GANs.

Jun-Yan Zhu[12] states that rather than picture synthesis from a random vector, the CycleGAN tackles the challenge of image-to-image translation. It addresses the situation where paired training examples for image-to-image translation are unavailable. Cycle-Consistency loss formulation insights on how this stabilises GAN training. It can be utilised for super resolution, style transfer, and among other things.

Aladdin Masri[13] showed one of the most important steps in solving the problem is detecting the user and body parts. In the user interface, for detecting body parts, monitoring skeletal movements, and estimating posture. The Project is written in C# and is intended to be a real-time, Kinect hacking application. In conjunction with Microsoft Kinect, the middleware of the Kinect driver is utilized for a range of core operations as well as the tracking process. Presents a virtual dressing room application that makes use of the Microsoft Kinect sensor.

III. METHODOLOGY

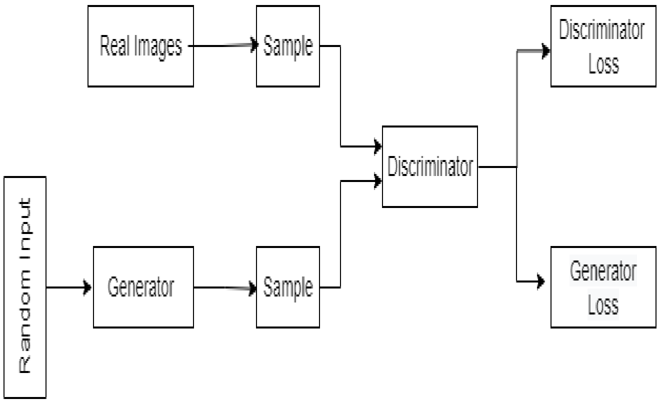

Figure 1: GAN Architecture

Figure 1 shows Generative Adversarial Network (GAN) which comprises of G and D models – generative and discriminative model. The discriminative model, rather than the generative model, assesses the probability that a sample came from the training data, whereas the generative model captures the data distribution. The architecture is similar to a two-player minimax game, using G's training approach maximising the probability of D committing an error.

During the literature survey, various types of GAN were introduced, few being DCGAN, CycleGAN, StyleGAN, ACGPN and Poly-GAN. DCGAN is a Deep Convolutional GAN and it is one of the most widely used, powerful, and successful GAN architecture types. ConvNets are used instead of a Multi-layered perceptron to implement it. ConvNets are built with a convolutional stride and no max pooling, and the layers in this network are not completely connected.

The CycleGAN model is intended to solve the problem of image-to-image conversion. The image-to-image translation challenge's purpose is to use a training set of matched picture pairings so that mapping between an input image and an output image can be learned. Getting matched samples, on the other hand, isn't always possible. CycleGAN uses cycle-consistent adversarial networks to learn this mapping without requiring paired input-output images.

A GAN for virtual try-on apparel applications is the Adaptive Content Generating and Preserving Network (ACGPN). It is divided into three sections. First, following try-on, a semantic layout generating module follows the anticipated semantic structure using semantic segmentation of the input image. Second, the clothes warping function warps images of clothes based on the semantic structure established, employing a second-order difference criterion to ensure consistency during training. Lastly, a contents union inpainting function combines every necessary information (for example, input image, semantic structure, deformed garments) to generate every semantic structure of the human body in an adaptable manner. When compared to current approaches, ACGPN can produce photorealistic images with significantly greater precision and richer fine-details.

The StyleGAN is an expansion of the continuously evolving GAN that is a methodology for preparing generator models fit for orchestrating extremely enormous high-quality pictures during the training cycle through the steady expansion of both discriminator and generator models from small to large images.

Current solutions employ a three-stage channel to align the garment with the human posture, sewing the aligned garment, and then refining the resulting image. Poly-GAN simplifies the procedure by doing all three jobs in a single architecture. This design upholds the criteria on the encoders' layers and employs skip connections from the coarse layers of the encoder to the decoder's appropriate layers.

The peculiarity of Poly-GAN permits us to work on various inputs and ideal for a wide variety of applications including inpainting, stitching and image alignment. Poly-GAN can conduct a dimensional change of the clothing based on the model's skeleton in any posture. When the garment mask contains irregular holes, Poly-GAN may also execute image stitching and inpainting on it, inconsiderate of the clothes orientation. The Try-On model would anticipate the input image's sematic layout and decide if its picture contents need to be developed or maintained, resulting to virtual photorealistic try-on.

A. Detailed Design

Among the various GANs available, Poly-GAN has been used to implement the TRY-ON.

1) Input Model: The image of the reference model is taken as one input.
2) Input Garment: The image of the reference garment is taken as another input.

3) Pose Estimator (Skeleton): Leaving the head while keeping the hip and hands, an array of bone list created which is drawn to create the skeleton.
4) Wrapped Clothes: The input garment takes the shape of the input model.

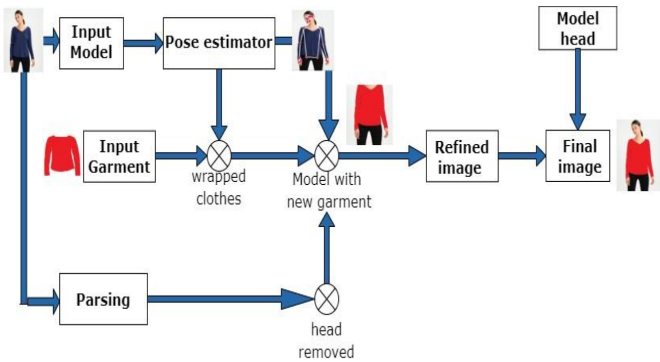

Fig. 2: Block Diagram

5) Parsing: The input model is parsed by converting the image (which is in string format) into an integer array. The resultant image is converted to Cuda Tensor and its contrast is improved.
6) Head Removal: The head is removed after parsing the input model image.
7) Model with new Garment: The output of the shape stage is concatenated with the removed head image to get the stitched stage output.
8) Refined image using Poly-GAN.
9) Model head: The model head which was removed earlier is added to the refined model image.
10) Final Image: After adding the head to the refined image, the final output as the model with new garment is obtained.

First, the data to train Poly-GAN, with inputs being pictures of a model and another input garment (top clothing only) were used. In the following stage, to infer the appearance flows between person representation (RGB model (Parsed)) and the garment image, followed with conjugating a mask containing the lower body clothes region (garment mask), hair and face, and the individual's body segmentation outcome, and the human posture assessment outcome as the individual representation. The warped clothes with garment images are generated using the skeleton. The generative module is trained to get the model with new garment by concatenating these warped garments, the preserved areas on the model picture, and human posture estimation as inputs.

IV. RESULTS AND DISCUSSION

(a) (b) (c) (d) (e)

Figure 3: Poly-GAN results from left to right: (a) Input Image; (b) Estimated Pose (skeleton); (c) Input Garment; (d) Shape and Stitched Output; (e) Refined Final Output

Fig. 3(b) shows the pose estimated (skeleton) of the input image of the model (as in Fig. 3(a)). Fig. 3(c) is the input image of the garment which takes the shape of the input model using the key points of the skeleton and is stitched on the input model as shown in Fig. 3(d). Fig. 3(e) depicts the final output image after refinement which is produced with the input model wearing the input garment.

TABLE I: Accuracy Scores

Sr. No.	Stage	Accuracy
(i)	Shape Stitch	0.86
	Final Refined	0.82
(ii)	Shape Stitch	0.80
	Final Refined	0.84
(iii)	Shape Stitch	0.79
	Final Refined	0.76
(iv)	Shape Stitch	0.81
	Final Refined	0.87

Table 1 shows the Structural Similarity Index Measure (SSIM) of the two stages – Shape Stitch and Final Refined Output, for the four test cases.

V. CONCLUSION

The Poly-GAN is trained with inputs being model images wearing clothing piece and another input garment (top clothing only). In the following stage, to infer the appearance flows between person representation (RGB model (Parsed)) and the garment image, conjugating a mask containing the lower body clothes region (garment mask), hair and face, and the individual's body segmentation outcome, and the human posture assessment outcome as the individual representation. Then, at that point, the skeleton is utilized to generate the shape of the wrapping garments. Trained a generative module with these wrapped garments, the conserved areas on the model picture, and human posture estimates along channels as inputs to get the person image with the new outfit as our result.

The accuracy using the Structural Similarity Index Measure (SSIM) of the model was found to be between 75% to 90% which indicates the model can effectively work as a virtual try-on.

Concluding by highlighting the robustness of the implemented model and the high quality of the produced images. Planning to train the model on other datasets in the future, such as for men's clothing. Would also want to improve the image generation for the reference person in various poses and views (side view, back view). Another factor that could be considered in future work and will be beneficial to consumers is the size of the clothing and the size of the reference person.

This model could be used to create photo-realistic try-on pictures for online shopping and other websites. The model has a long way to go before it can produce acceptable results for all types of clothing and reference poses.

However, the field of machine learning is doing very well, thanks to the ever-decreasing cost of computational power and the ever-increasing brain power invested in its research.

REFERENCES

[1] Nilesh Pandey, Andreas Savakis, "Poly-GAN: Multi - conditioned GAN for fashion synthesis", published in Neurocomputing Volume 414, 13, Pages 356-364, Nov-2020 https://arxiv.org/pdf/1909.02165.pdf

[2] Yuying Ge, Yibing Song, Ruimao Zhang, Chongjian Ge, Wei Liu and Ping Luo, "Parser-Free Virtual Try-on via Distilling Appearance Flows", published 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) https://cutt.ly/YGkVIDl

[3] Rahul Singh, Aman Bindal, Md Azad Khan, Rakshith MR and Prof. Divakara N, "Virtual Clothing Try-on Using Generative Adversarial Networks", published in IJARIIE-ISSN(O)-2395-4396, 2021 https://cutt.ly/aGkVFx0

[4] Han Yang, Ruimao Zhang, Xiaobao Guo, Wei Liu, Wangmeng Zuo and Ping Luo, "Towards Photo-Realistic Virtual Try-On by Adaptively Generating Preserving Image Content" published in Cornell University, March 2020 https://arxiv.org/pdf/2003.05863.pdf

[5] Shion Honda, "VITON-GAN: Virtual Try-on Image Generator Trained with Adversarial Loss", published in Eurographics Digital Library, 2019 https://cutt.ly/XGkVX9Y

[6] Shizhan Zhu, Sanja Fidler, Raquel Urtasun, Dahua Lin Chen Change Loy, "2017 IEEE International Conference on ComputerVision (ICCV)", published in IEEE International Conference on Computer Vision (ICCV), 2017 https://cutt.ly/MGkV15W

[7] Kathleen M Lewis, Srivatsan Varadharajan, Ira Kemelmacher- Shlizerman," TryOnGAN: Body-Aware Try-On via Layered Interpolation", published in Cornell University, Jan 2021 https://arxiv.org/abs/2101.02285

[8] Gaurav Kuppa, Andrew Jong, Xin Liu, Ziwei Liu, Teng-Sheng Moh, " ShineOn: Illuminating Design Choices for Practical Video-based Virtual Clothing Try-on", published in 2021 IEEE Winter Conference on Applications of Computer Vision Workshops (WACVW) https://www.computer.org/csdl/proceedingsarticle/wacvw/2021/196700a191/1sZ3rKHyrHa

[9] Yevhen Pozdniakov, Orest Kupyn," Changing clothing on people images using generative adversarial networks", Master Thesis under Ukrainian Catholic University https://cutt.ly/OGx8p2Y

[10] Xintong Han, Zuxuan Wu, Zhe Wu, Ruichi Yu, Larry S. Davis," VITON: An Image-based Virtual Try-on Network", published in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) https://cutt.ly/5GvxiS3

[11] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, Xi Chen," Improved Techniques for Training GANs", published in NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems, December 2016 https://arxiv.org/abs/1606.03498

[12] Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros," Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", published in 2017 IEEE International Conference on Computer Vision (ICCV), Oct 2017 https://arxiv.org/abs/1703.10593

[13] Aladdin Masri; Muhannad Al-Jabi," Virtual Dressing Room Application", published in IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), April 2019 https://ieeexplore.ieee.org/document/8717410

[14] VITON Dataset, https://www.kaggle.com/datasets/rkuo2000/vitondataset