	University of Engineering and Technology

The Digital Wardrobe: An SLR on Virtual Try-On Algorithm

Submitted to

Dr. Tauqir Ahmad

Submitted by

Abdullah Mustafa   **(2023-MSCS-242)**
Muhammad Ali   **(2023-MSCS-211)**

Problem Statement:

The research papers discussed various challenges in the domain of virtual try-on systems. The first paper focused on enhancing virtual try-on results by addressing issues related to challenging poses and self-occlusions during the transfer of clothing items. The second paper aimed to develop a Squeezing-Averse Virtual Try-On via Sequential Deformation, targeting challenges like complex warping, clothing size adjustment, and the need for an appropriate clothing elimination policy. Another paper delved into the creation of a virtual try-on system allowing versatile garment styling while maintaining identity and ensuring high-quality image production. Additionally, the tasks of multimodal-conditioned fashion image editing and high-fidelity virtual try-on image synthesis were explored, addressing challenges related to multimodal prompts and preserving garment shapes and patterns. The alignment improvement in virtual try-on applications was the focus of another study, while a separate paper tackled the synthesis of detailed, realistic visualizations of garments on individuals, accommodating significant body pose and shape changes.

Research GAP:

Several challenges and limitations were identified across different virtual try-on research papers. Geometric matching issues were noted, affecting garment details and resulting in poor quality parsing, leading to unconvincing images. The limitations of one study included crack artifacts in samples, amplified by the up-sizing of appearance flows, and difficulty in fully removing clothes textures behind the neck. Another paper highlighted the need for improvements in accessories segmentation and expanding the system to include more clothing categories. Some studies did not explicitly list limitations, while others suggested differences between generated and real looks. Specific challenges for models included inaccurate generation of hands in limited areas, dependence on given sketches, and potential artifacts due to geometric warping module failures. Several methods were reported to exhibit garment leaking artifacts, identity preservation concerns, and unknown performance in complex backgrounds. The limited exploration of full body try-on in upper body-focused studies was also acknowledged.

Literature Review:

Table 1 Comparison with existing Studies

Sr.	Name	Year	Dataset	Methodology	Results	Future Work
[1]	C-VTON: Context-Driven Image-Based Virtual Try-On Network.	2022	VITON and MPV datasets	Two-stage pipeline with a Body-Part Geometric Matcher (BPGM) aligning clothing with subject pose and a Context-Aware Generator (CAG) synthesizing the final try-on image.	C-VTON outperformed state-of-the-art models in terms of FID and LPIPS scores, producing photo-realistic and visually convincing results.	Addressing garment detail loss in geometric matching and improving realism in try-on images.
[2]	SD-VITON-Virtual-Try-On	2024	The paper does not explicitly have its dataset.	This research proposes a novel method, Cascaded Appearance Flow for High-Resolution Virtual Try-On (CVTON), to address challenges in generating realistic virtual try-on images.	The proposed method, CVTON, avoids sleeve-squeezing (preserving sleeve patterns) and waist-squeezing (maintaining consistent gaps between clothing elements) artifacts in virtual try-on images. This is in contrast to other methods that suffer from these issues.	A slightly reducible way to address the crack issue will be to infer the last AFs with more-closer to an image resolution. A solution for the issue of clothes behind the neck would be to mask out such regions when pre-processing the inputs. However, this additional technique was not applied since it was not included in the dataset.
[3]	The paper is titled "Size Does Matter: Size-aware	2023	The authors created the "Pure Cotton" dataset, a high-	The authors propose the "Clothing-Oriented Transformation Try-On Network (COTTON)"	COTTON outperforms state-of-the-art methods in high-resolution	The authors plan to enhance the segmentation of accessories and

			resolution outfit dataset, and also used the public "Dress Code" dataset for evaluation.	which uses landmarks and segmentation for precise clothing deformation and size adjustment3.	virtual try-on tasks both qualitatively and quantitatively45.	expand the system to include shoes, hats, and other accessories.
	Virtual Try-on via Clothing-oriented Transformation Try-on Network."1					
[4]	GP-VTON: Towards General Purpose Virtual Try-on via Collaborative Local-Flow Global-Parsing Learning1	2023	VITON-HD and DressCode	The GP-VTON framework employs a Local-Flow Global-Parsing (LFGP) warping module and a Dynamic Gradient Truncation (DGT) training strategy23.	GP-VTON outperforms existing state-of-the-art methods in generating photo-realistic try-on results, even with challenging inputs.	Not specified in the provided context.
[5]	The paper is titled "Wearing the Same Outfit in Different Ways – A Controllable Virtual Try-on Method."1	2022	The research utilizes the "OVNet dataset," which contains a variety of garments worn in different styles2.	The method involves predicting control points on garments, enabling instance-independent editing of garment drape, and using an adversarially trained renderer for realistic image generation.	The proposed method allows for interactive control over how garments are worn, such as tucking or untucking shirts, and generates state-of-the-art quality images4.	The authors aim to extend controllability to dynamic body poses and support different body sizes6.
[6]	Multimodal Garment Designer: Human-Centric Latent Diffusion Models for Fashion Image Editing	2023	Dress Code and VITON-HD, extended with multimodal annotations12	The approach involves a novel architecture based on latent diffusion models, capable of following multimodal prompts while preserving the model's characteristics.	The experiments demonstrate that the proposed approach generates high-quality images based on given multimodal inputs and outperforms competitors in terms of realism.	The document does not provide explicit details about future work. However, addressing the mentioned limitations could be a potential area for future research and improvements.
[7]	Pose-Garment Keypoints Guided Inpainting (KGI)1	2023	VITON-HD dataset	KGI uses a graph-based model to predict pose-oriented garment keypoints for accurate garment warping and segmentation map estimation45. It also includes a semantic-conditioned inpainting scheme to fill in missing regions6.	KGI has been extensively tested on the VITON-HD dataset, showing significant improvements over prior methods in both qualitative and quantitative results.	The paper suggests potential extensions of KGI for other types of garments in future work7. However, specific details are not provided.

[8]	The algorithm presented in the document is called SAL-VTON.	2023	The document mentions a new dataset with a unified labeling rule for diverse garment styles, but it doesn't provide a specific name for this dataset12.	SAL-VTON uses semantically associated landmarks in a three-stage framework: landmark localization, a warping model for alignment refinement, and a generator that uses landmarks to capture local semantics and control try-on results3.	Extensive experiments demonstrate that SAL-VTON handles misalignment effectively and outperforms state-of-the-art methods both qualitatively and quantitatively45.	The document doesn't mention any specific plans for future work.
[9]	TryOnDiffusion: A Tale of Two UNets	2023	The paper mentions a dataset of 4 million image pairs of the same person in different poses2. However, the exact name of the dataset is not provided.	The authors propose a diffusion-based architecture called Parallel-UNet, which unifies two UNets to preserve garment details and warp the garment for significant pose and body changes in a single network4. The method involves implicit warping of the garment through cross attention mechanisms and combines warping and blending in a single network pass5.	The method was trained on 4 million image pairs and evaluated both quantitatively and qualitatively against state-of-the-art methods, showing superior performance2. In user studies, TryOnDiffusion was chosen as the best in over 92% of cases.	The authors believe that the Parallel-UNet architecture will be impactful for the general case of image editing and plan to extend the architecture to videos in the future6. They also plan to experiment with full body try-on3.

References:

[1] https://openaccess.thecvf.com/content/ICCV2023/papers/Chen_Size_Does_Matter_Size-aware_Virtual_Try-on_via_Clothing-oriented_Transformation_Try-on_ICCV_2023_paper.pdf

[2] GitHub - benquick123/C-VTON: C-VTON: Context-Driven Image-Based Virtual Try-On Network

[3] https://github.com/SHShim0513/SD-VITON

[4] https://arxiv.org/pdf/2303.13756.pdf

[5] https://arxiv.org/pdf/2211.16989.pdf

[6] https://arxiv.org/pdf/2304.02051.pdf

[7] https://openaccess.thecvf.com/content/ICCV2023/papers/Li_Virtual_Try-On_with_Pose-Garment_Keypoints_Guided_Inpainting_ICCV_2023_paper.pdf

[8]https://openaccess.thecvf.com/content/CVPR2023/papers/Yan_Linking_Garment_With_Person_via_Semantically_Associated_Landmarks_for_Virtual_CVPR_2023_paper.pdf

[9] https://arxiv.org/pdf/2306.08276.pdf