GraVITON: Graph Based Garment Warping with Attention Guided Inversion for Virtual Try-on

Sanhita Pathak, Vinay Kaushik, Brejesh Lall
IIT Delhi, IIIT Sonepat, India

IIT Delhi

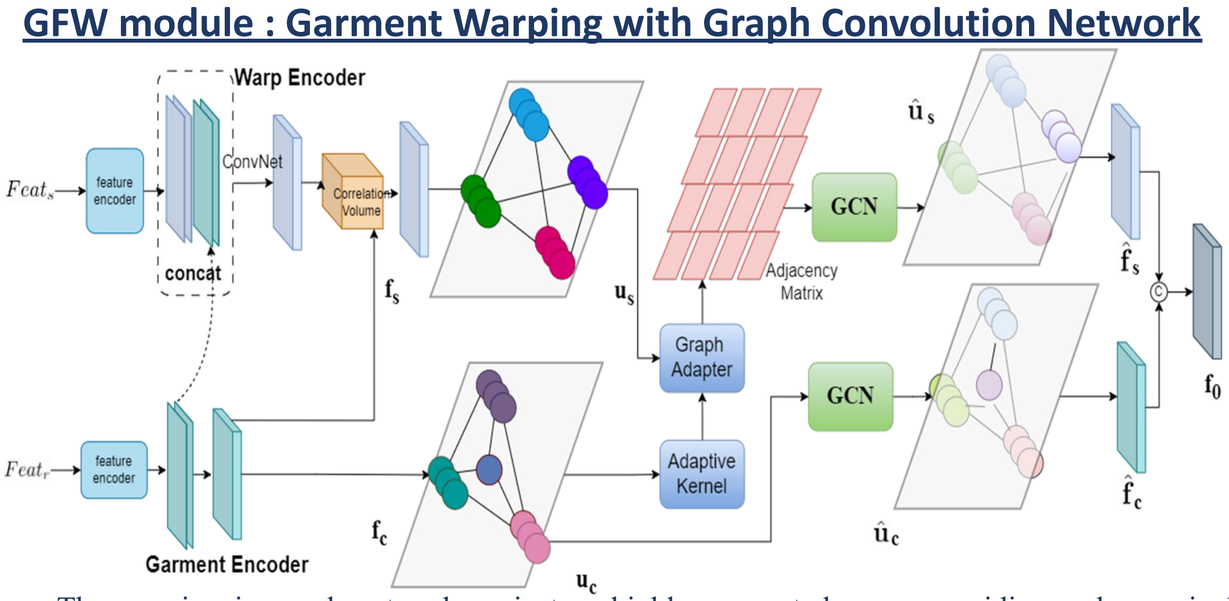
Abstract

Virtual try-on, a growing computer vision field, enhances e-commerce by allowing exact garment warping and seamless interaction with the human body. Existing approaches like TPS and flow handle garment warping but neglect contextual variables. The graph-based warping module warps the straight cloth which is utilised by the . The proposed technique uses a latent diffusion model to construct the final tryon, treating garment transfer as an inpainting task. The diffusion model is conditioned with decoupled cross attention based inversion of visual and textual information.

Contributions

- We introduce a Graph based flow warping module(GFW), that guides the f low warping by providing pixel neighbourhood context into source and reference. To the best of our understanding, we are the first to introduce graph based technique for garment warping.
- We propose Occlusion Aware warp Loss(OWL) to enable the complete warped garment learning in case of self-occlusion present in ground truth garments.
- We propose a Decoupled Cross Attention Adaptor (DCAA), enriching latent space inversion for a realistic tryon.
- Extensive experimentation and rigorous validation demonstrates that our method achieves state-of-the-art performance compared to existing prominent methods.

GFW module : Garment Warping with Graph Convolution Network



The warping in graph network projects a highly connected space providing a dense pixel context utilising graph's adjacency property. The features extracted are projected in the graph space the nodes being source and reference features. The process of node creation for both the source and context entails the computation of the adjacency matrix, which measures the similarity between all nodes. The offsets are computed and are utilised for warping.

Datasets

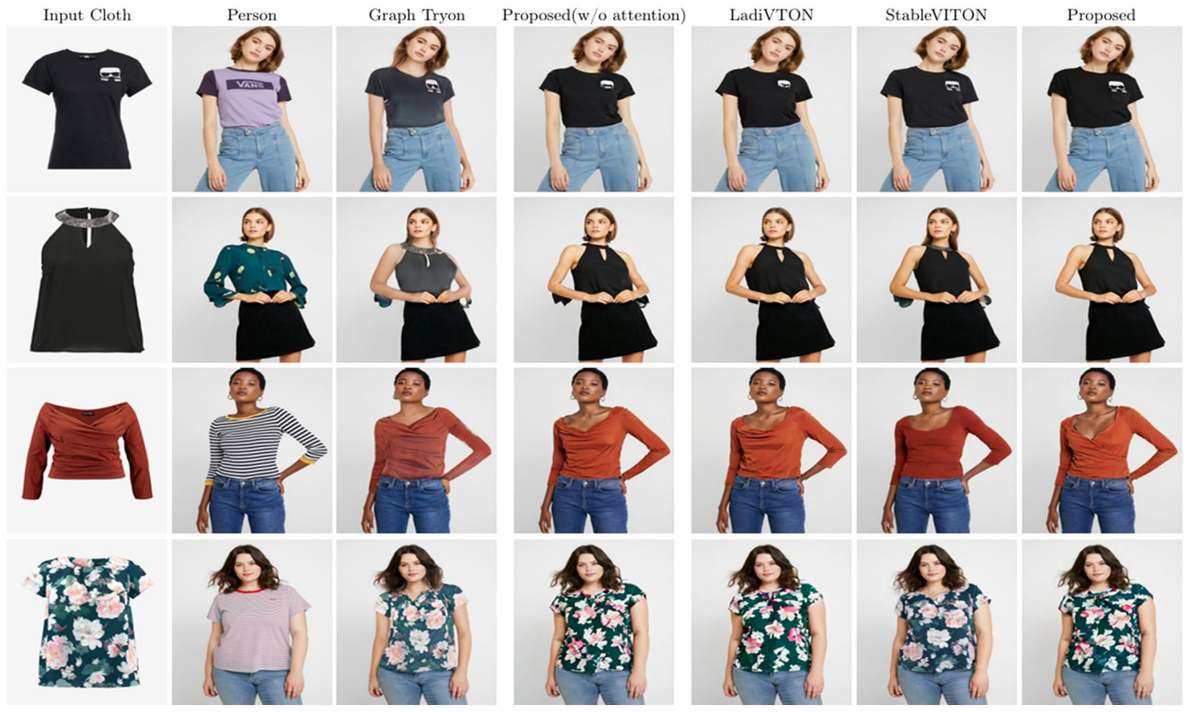
- VITON-HD : Test and Train sets containing upper garment and person images.
- Dresscode : Test and train set of upper-body, lower-body and dresses.

Methodology

The proposed methodology consists of two stages, namely the warping and try-on generation stage.

- Stage 1 : As shown in architecture diagram the stage one consists of the graph based warping module which utilises the offsets calculated from GFW module to warp the straight garment. This warped garment is utilised as the garment input for tryon stage.
- Stage 2 : This stage caters to the final tryon generation as an inpainting task utilising the warped garment input. The diffusion model utilises the textual inversion for garment type(upper/lower) and garment features which are utilised to compute decoupled coss attention guiding the generation in diffusion model. Along with the warped input the unet encoder utilises the human keypose points, binary human body mask and inpainting mask as the inputs to stage 2. The resulting inpainted image is the final try-on generation..

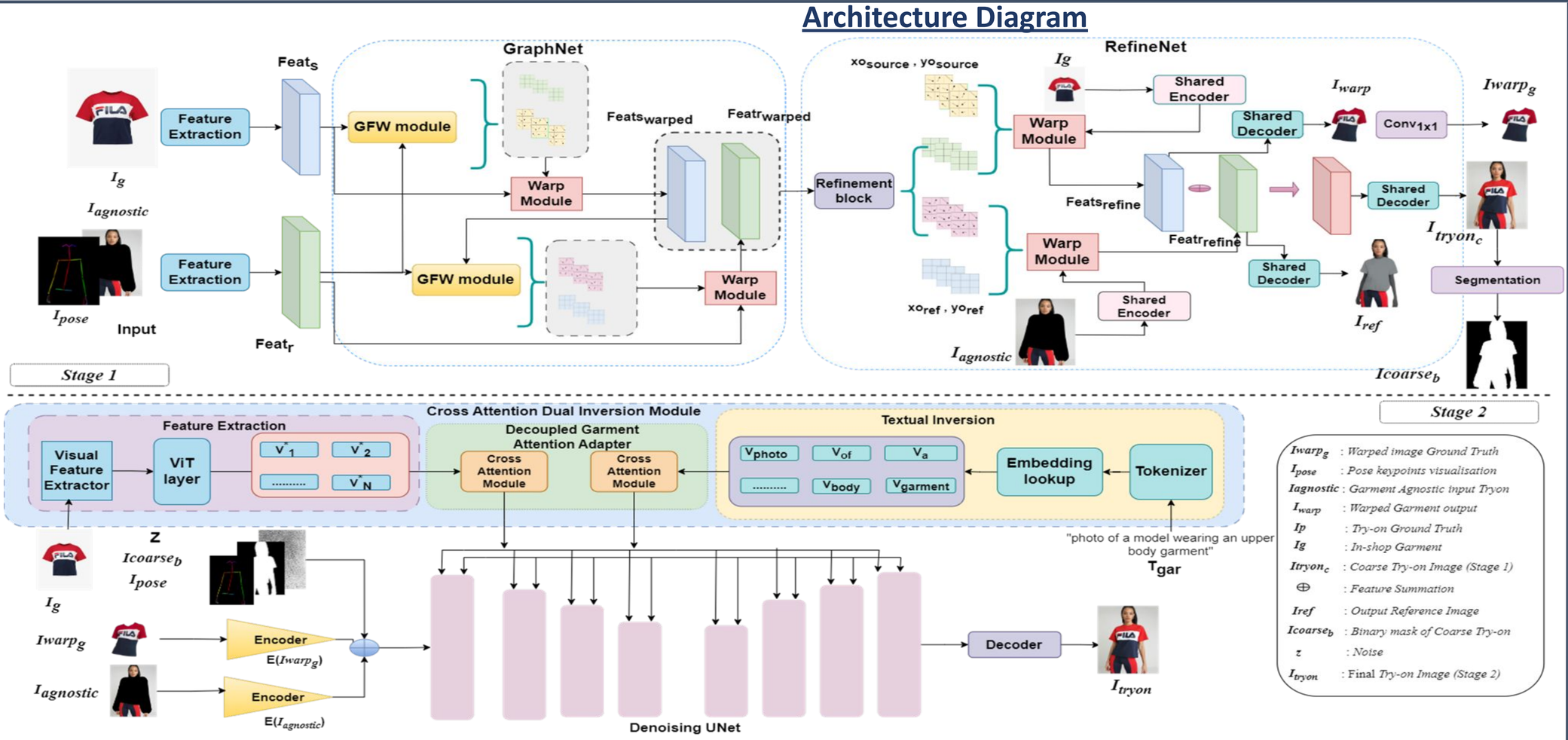
Ablation Qualitative Results on VITON HD Dataset



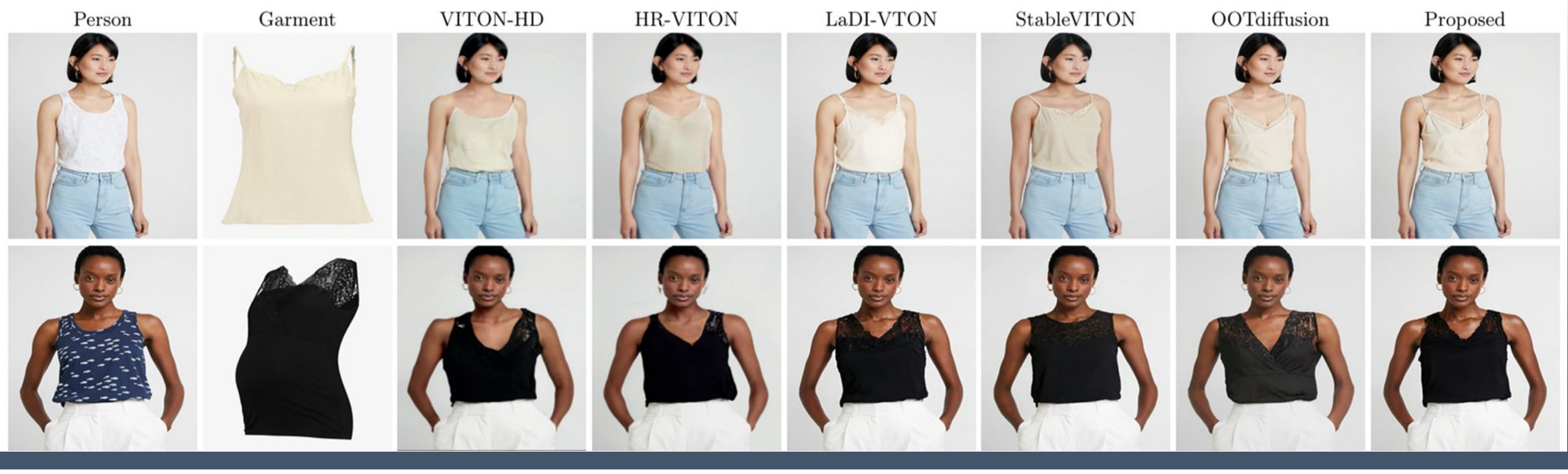
Quantitative Results

Model	LPIPS	SSIM	FID_p	KID_p	FID_u	KID_u
CP-VTON	-	0.791	-	-	30.25	40.12
ACGPN	-	0.858	-	-	14.43	5.87
VITON-HD	0.116	0.863	11.01	3.71	12.96	4.09
HR-VTON	0.097	0.878	10.88	4.48	13.06	4.72
LADIVTON	0.091	0.876	6.66	1.08	9.41	1.60
GraVITON	**0.088**	**0.891**	**6.57**	**1.06**	**9.20**	**1.46**

Architecture Diagram



Qualitative Results on VITON HD Dataset



Conclusion

Our paper introduces novel solutions to enhance virtual try-on technology, addressing critical challenges in garment warping and generation. By incorporating the Graph-based Flow Warping module (GFW), we achieve more accurate context reasoning, significantly reducing uncertainty in garment transfer. Graph-based Flow Warping module (GFW) aids in accurate context reasoning. Decoupled Cross-Attention Mechanism (DCAA) enriches latents pace information, leading to realistic try-on synthesis. Occlusion Aware Warp Loss (OWL) effectively handles self-occlusion, ensuring finer garment learning and seamless integration onto the human body.

Our contributions significantly improve the seminal work done by previous approaches by proposing novel graph-based framework for garment warping and introducing novel pose try-on synthesis using diffusion models.

References

- Morelli, D., Baldrati, A., Cartella, G., Cornia, M., Bertini, M., & Cucchiara, R. (2023, October). Ladi-vton: Latent diffusion textual-inversion enhanced virtual try-on. In *Proceedings of the 31st ACM International Conference on Multimedia* (pp. 8580-8589).
- Kim, J., Gu, G., Park, M., Park, S., & Choo, J. (2024). Stableviton: Learning semantic correspondence with latent diffusion model for virtual try-on. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 8176-8185).
- Xu, Y., Gu, T., Chen, W., & Chen, C. (2024). Ootdiffusion: Outfitting fusion based latent diffusion for controllable virtual try-on. *arXiv preprint arXiv:2403.01779*.