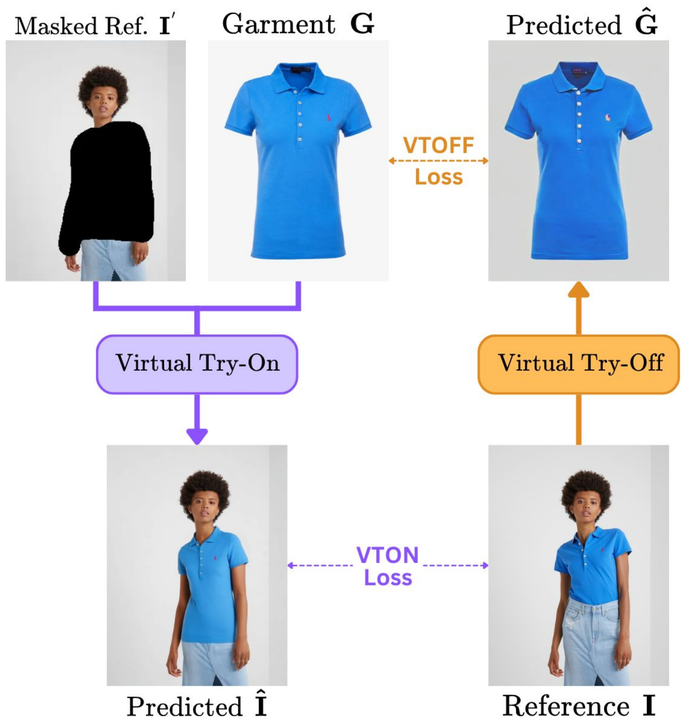
TryOffDiff: Virtual-Try-Off via High-Fidelity Garment Reconstruction using Diffusion Models

Riza Velioglu, Petra Bevandić, Robin Chan, Barbara Hammer

UNIVERSITÄT BIELEFELD

BMVC 2025

Problem

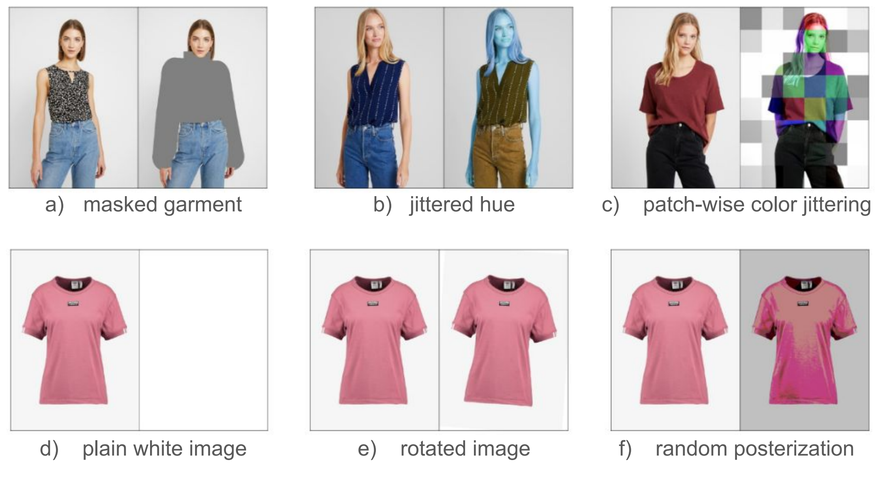


The two tasks form a cycle, where one's output can serve as other's input.

- We introduce the **Virtual Try-Off (VTOFF)** task.
- Motivations:
  - replace costly traditional product photography,
  - potentially reduce product return rates,
  - support image retrieval applications.

Key Finding: We analyze popular image generation metrics and find that DISTS [1] is the most suitable for assessing VTON & VTOFF performance.

Code, models, and demo are open-source!



Case	SSIM ↑	MS^{SSIM} ↑	CW^{SSIM} ↑	LPIPS ↓	DISTS ↓
a	82.6	71.3	40.3	35.1	21.4
b	98.8	98.9	96.4	12.1	14.4
c	90.2	78.7	58.9	33.8	19.8
d	73.1	66.1	0	45.7	69.5
e	81.5	77.2	72.4	17.4	10.0
f	77.5	92.6	73.4	20.9	29.9

Method

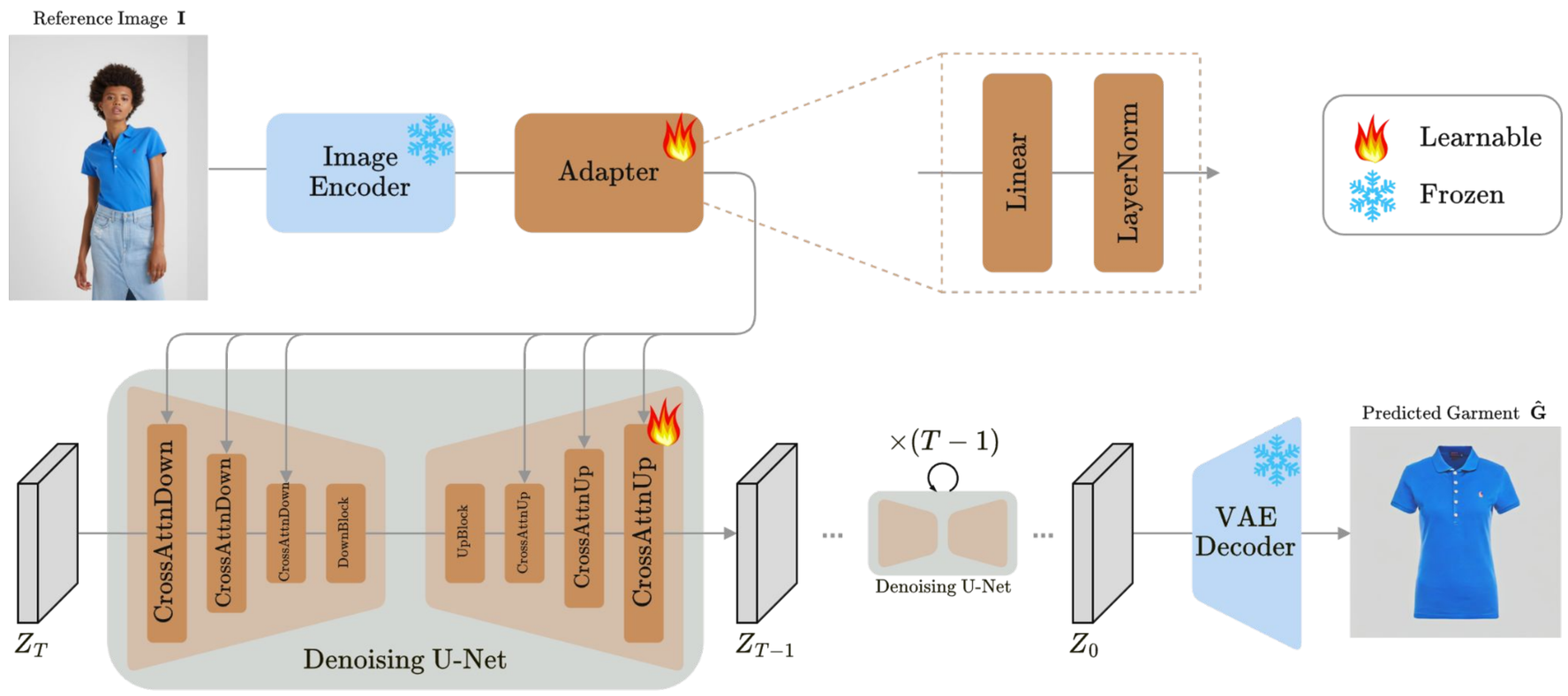


Image Encoder: SigLIP [9] | **Denoising U-Net & VAE**: Stable Diffusion-v1.4 [10]

SigLIP extracts features from a reference image, which are processed by adapter modules and embedded into Stable Diffusion v1.4 by replacing text features within the cross-attention layers. This enables image-guided generation. The adapter and diffusion model are trained jointly.

Results

Method	CW^{SSIM} ↑	LPIPS↓	FID↓	DISTS↓
GAN-Pose [36]	<u>32.5</u>	<u>44.2</u>	73.2	30.4
ViscoNet [6]	28.9	54.0	42.3	31.2
OOTDiff. [58]	26.1	49.5	54.0	32.4
CatVTON [9]	32.0	45.9	<u>31.4</u>	<u>28.2</u>
IP-Adapter [59]	30.1	51.5	36.1	30.9
TryOffDiff	**49.2**	**30.1**	**14.5**	**20.3**

Evaluation on VITON-HD shows consistent improvement across all metrics.

Train/Test	CW^{SSIM} ↑	LPIPS↓	FID↓	DISTS↓
HD / HD	49.2	30.1	14.5	20.3
DC / HD	37.0	38.3	22.9	26.6
DC / DC	47.8	31.6	17.1	21.6
HD / DC	34.7	41.5	29.6	28.1

Within- and cross-dataset evaluation of TryOffDiff. Trained and tested on VITON-HD[7] (HD) or DressCode[8] (DC, upper-body) in within-dataset (HD/HD, DC/DC) and cross-dataset (DC/HD, HD/DC) settings.

Acknowledgement

This work has been funded by Horizon Europe program under grant agreement 101134447-ENFORCE, and by the German federal state of North Rhine-Westphalia as part of the research funding program KI-Starter.

[1] Ding et al., Image Quality Assesment: Unifying Structure and Texture Similarity, TPAMI, 2022
[2] Roy et al. Multi-scale attention guided pose transfer, PR, 2023
[3] Cheong et al. Visconet: Bridging, harmonizing visual and textual conditioning for controlnet, ECCVW 2024
[4] Xu et al., Ootdiffusion: Outfitting fusion based latent diffusion for controllable virtual try-on, AAAI, 2025.
[5] Chong et al., Catvton: Concatenation is all you need for virtual try-on with diffusion models, ICLR, 2025
[6] Ye at al. Ip-adapter: Text compatible image prompt adapter for text-to-image diffusion models, arXiv 2023
[7] Choi et al., Viton-hd: High-resolution virtual try-on via misalignment-aware normalization, CVPR, 2021
[8] Morelli et al., Dress code: High-resolution multi-category virtual try-on, CVPR, 2022
[9] Rombach et al., High-Resolution Image Synthesis with Latent Diffusion Models, CVPR, 2022
[10] Zhai et al., Sigmoid Loss for Language Image Pre-Training, ICCV, 2023