Tag: TUE-PM-176

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Person Image Synthesis via Denoising Diffusion Model

Ankan Bhunia, Salman Khan, Hisham Cholakkal, Rao Anwer, Jorma Laaksonen, Mubarak Shah, Fahad Khan



Australian National University



UCF

A. Problem Formulation

Target pose

Source Image

Pose-guided Person Image Generation

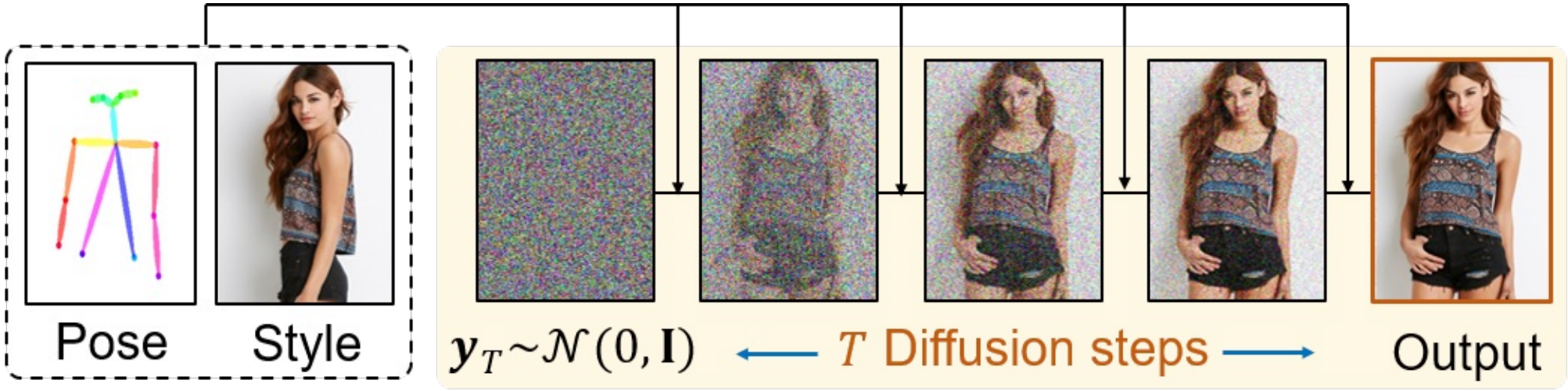
Output Image

B. Existing GAN-based Methods

1. Model attempts to generate the final image in a single forward pass.

2. Struggles to capture complex structure of the spatial transformation.

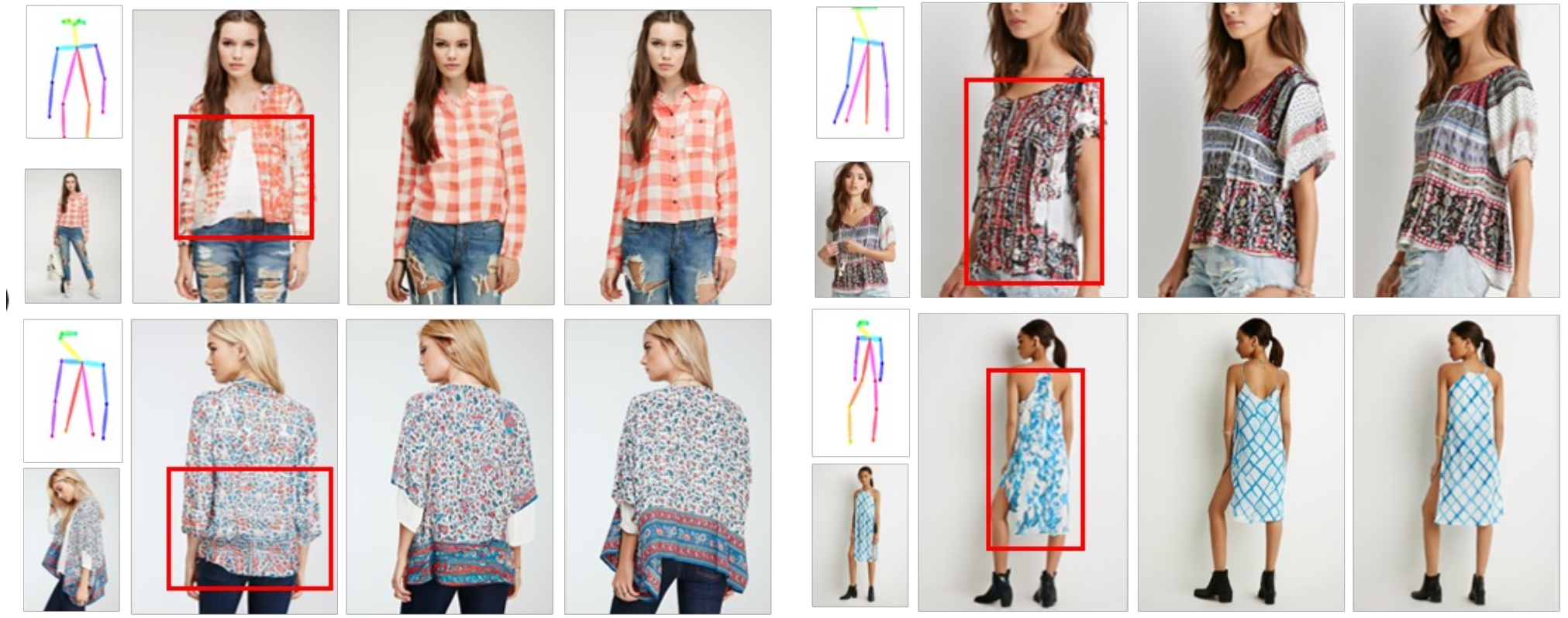
C. Our Diffusion-based Solution: PIDM



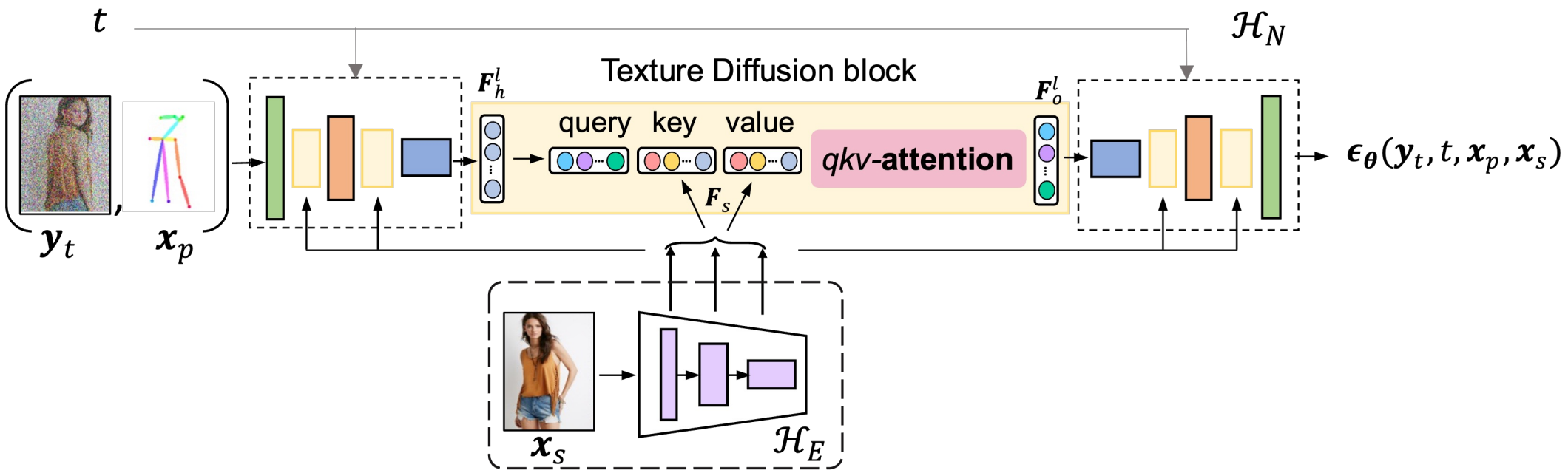
PIDM breaks down the generation process into several conditional denoising diffusion steps, each step being relatively simple to model.

C. Our Diffusion-based Solution: PIDM

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

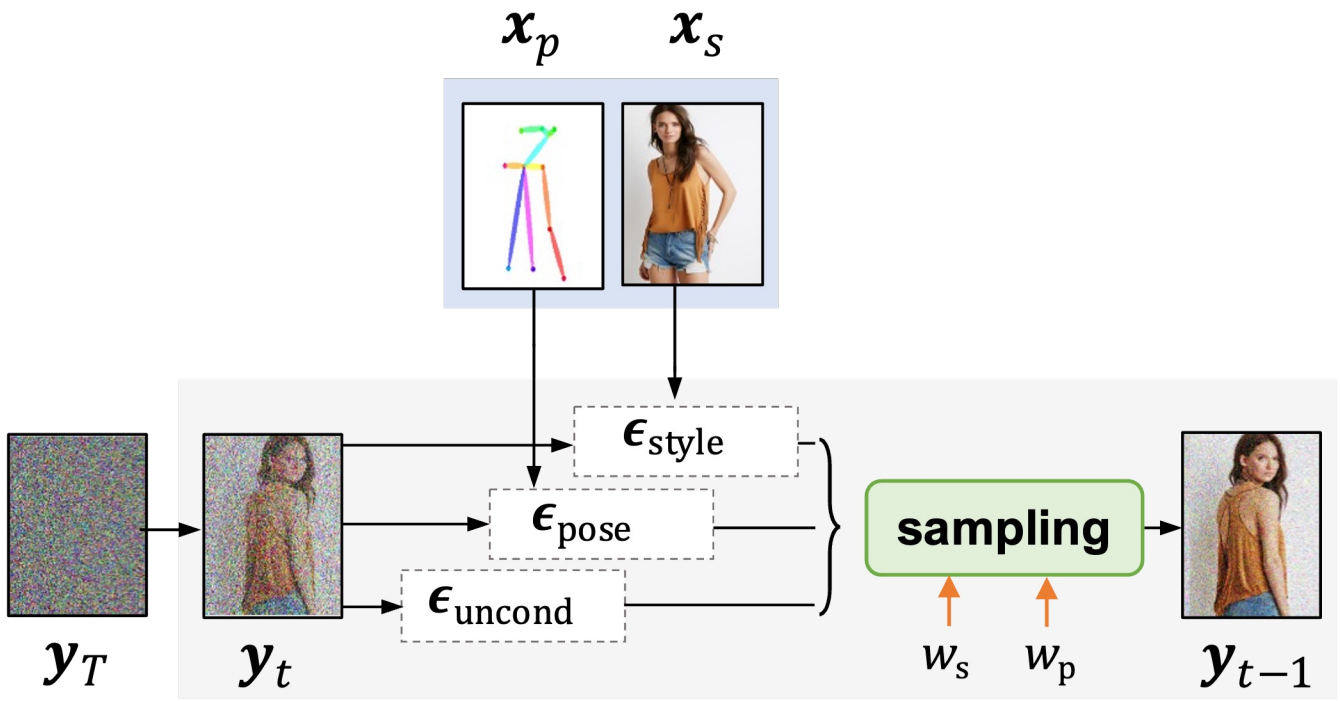


D. PIDM Framework



E. Disentangled Guidance sampling

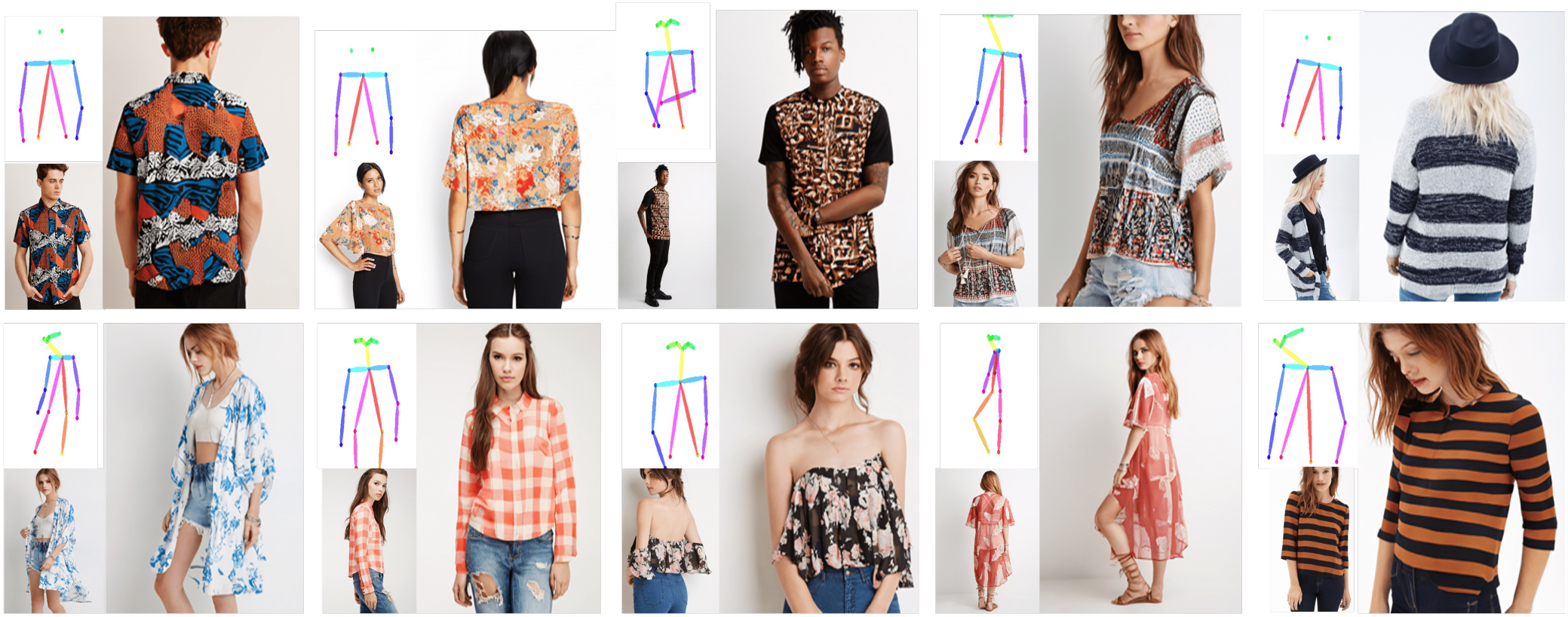
JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA



$$\epsilon_{\text{cond}} = \epsilon_{\text{uncond}} + w_p \epsilon_{\text{pose}} + w_s \epsilon_{\text{style}}$$

F. Results: DeepFashion Dataset

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

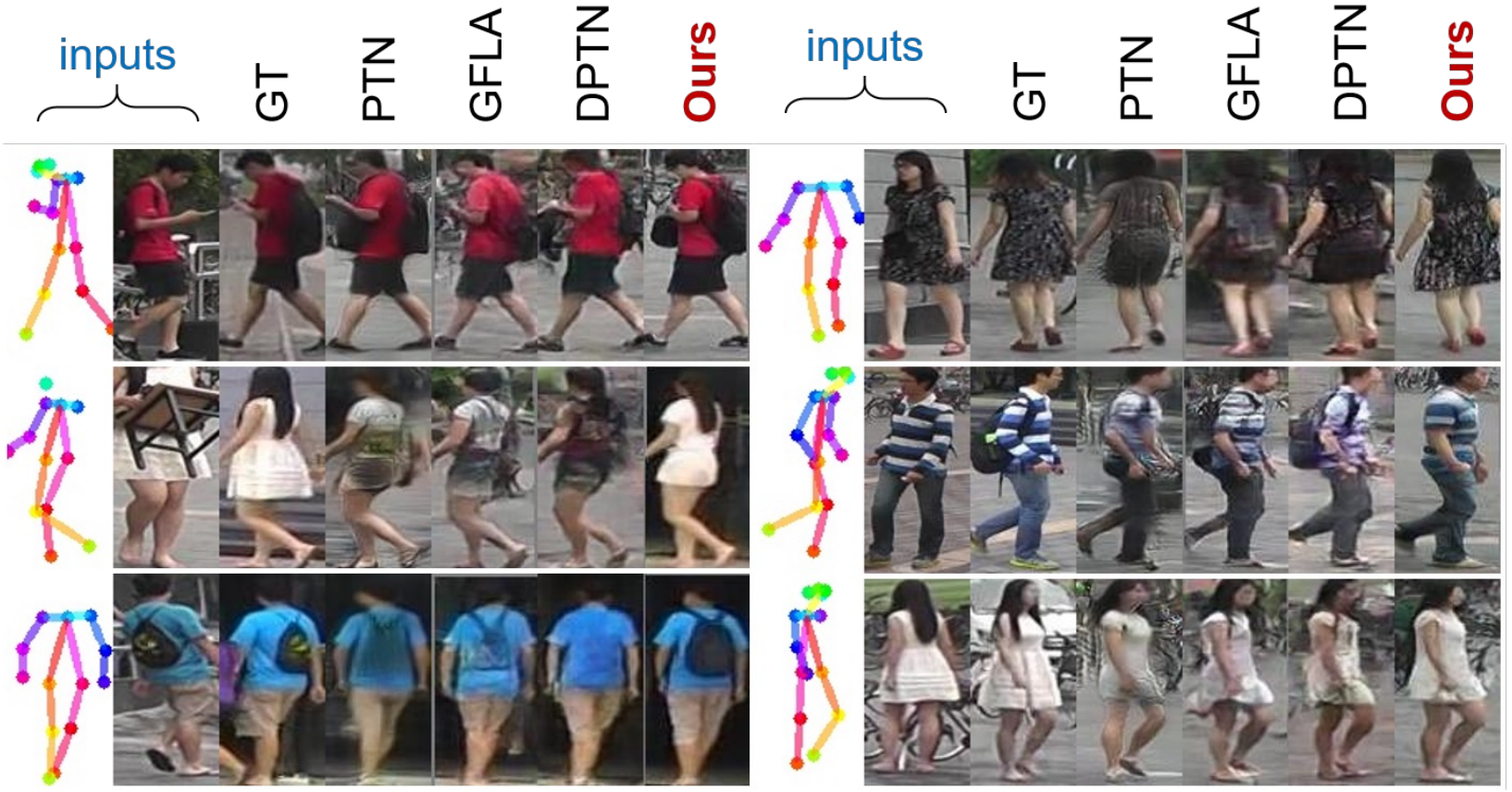


G. Comparisons: DeepFashion Dataset

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

H. Comparisons: Market-1501 Dataset

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

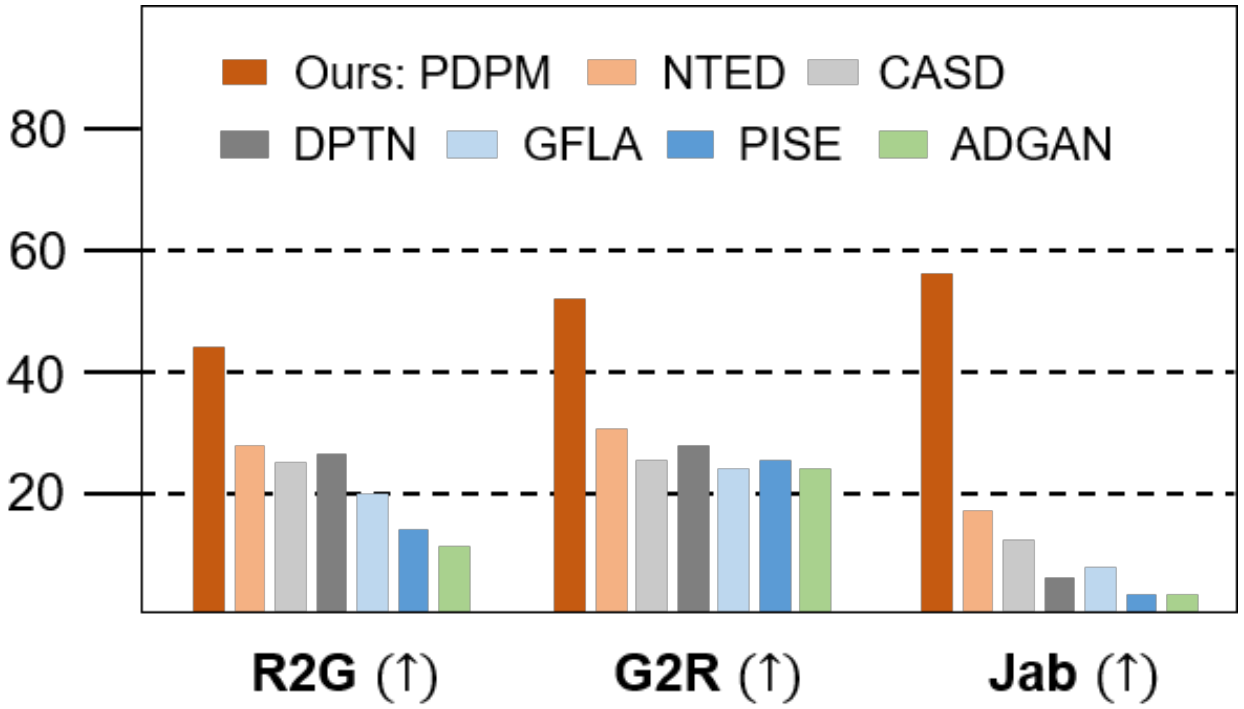


I. Quantitative Comparisons

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Dataset	Methods	FID(↓)	SSIM(↑)	LPIPS(↓)
DeepFashion [11] (256 × 176)	Def-GAN [20]	18.457	0.6786	0.2330
	PATN [30]	20.751	0.6709	0.2562
	ADGAN [14]	14.458	0.6721	0.2283
	PISE [23]	13.610	0.6629	0.2059
	GFLA [19]	10.573	0.7074	0.2341
	DPTN [24]	11.387	0.7112	0.1931
	CASD [28]	11.373	0.7248	0.1936
	NTED [18]	8.6838	0.7182	0.1752
	PIDM (Ours)	**6.3671**	**0.7312**	**0.1678**
DeepFashion [11] (512 × 352)	CocosNet2 [29]	13.325	0.7236	0.2265
	NTED [18]	7.7821	0.7376	0.1980
	PIDM (Ours)	**5.8365**	**0.7419**	**0.1768**
Market-1501 [27] (128 × 64)	Def-GAN [20]	25.364	0.2683	0.2994
	PTN [30]	22.657	0.2821	0.3196
	GFLA [19]	19.751	0.2883	0.2817
	DPTN [24]	18.995	0.2854	0.2711
	PIDM (Ours)	**14.451**	**0.3054**	**0.2415**

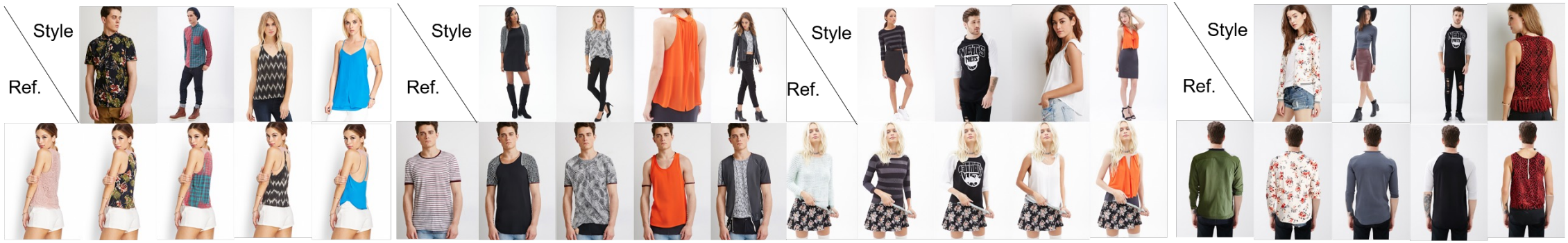
J. Human Evaluation



User study results on DeepFashion dataset in terms of R2G, G2R and Jab metric. Higher values indicate PIDM is preferred more often over the compared approaches.

K. Appearance Control

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA



$$\boldsymbol{y}_t^{ref} = \sqrt{\bar{\alpha}_t}\boldsymbol{y}^{ref} + \sqrt{1-\bar{\alpha}_t}\epsilon$$

$$\boldsymbol{y}_t = \boldsymbol{m} \odot \boldsymbol{y}_t + (1-\boldsymbol{m}) \odot \boldsymbol{y}_t^{ref}$$

L. Application to Person Re-identification

Methods	Percentage of real images				100%(+30K)
	20%	40%	60%	80%	
Standard	33.4	56.6	64.9	69.2	76.7
PTN [30]	55.6	57.3	67.1	72.5	76.8
GFLA [19]	57.3	59.7	67.6	73.2	76.8
DPTN [24]	58.1	62.6	69.0	74.2	77.1
PIDM (Ours)	**61.3**	**64.8**	**71.6**	**75.3**	**78.4**

M. In-the-wild Evaluation

Conclusion

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Scan the QR code for codes and demo

GitHub: https://github.com/ankanbhunia/PIDM/

Thank you.