Controllable Person Image Synthesis with Attribute-Decomposed GAN (ADGAN)

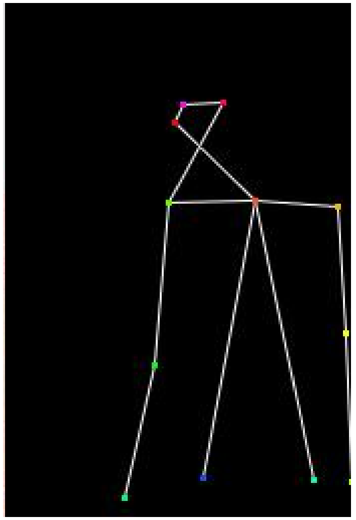
Yifang Men[1], Yiming Mao[2], Yuning Jiang[2], Wei-Ying Ma[2], ***Zhouhui Lian[1]****
[1]Wangxuan Institute of Computer Technology, Peking University
[2]Bytedance AI Lab

What We Can Do

- Synthesize person images in **arbitrary pose**

Reference

+



Target pose

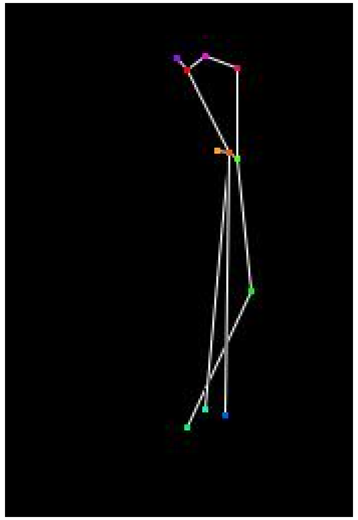
Target image

What We Can Do

- Synthesize person images in **arbitrary pose**

Reference

+



Target pose

→

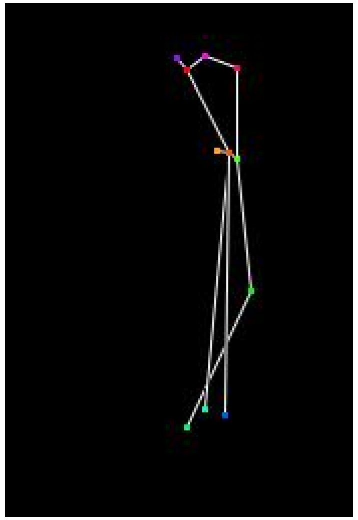
Target image

What We Can Do

- Achieve flexible and continuous control of **human attributes**, including **pose** and **component attributes** (i.e. **head**, **upper clothes** and **pants**)

Reference

+



Target pose

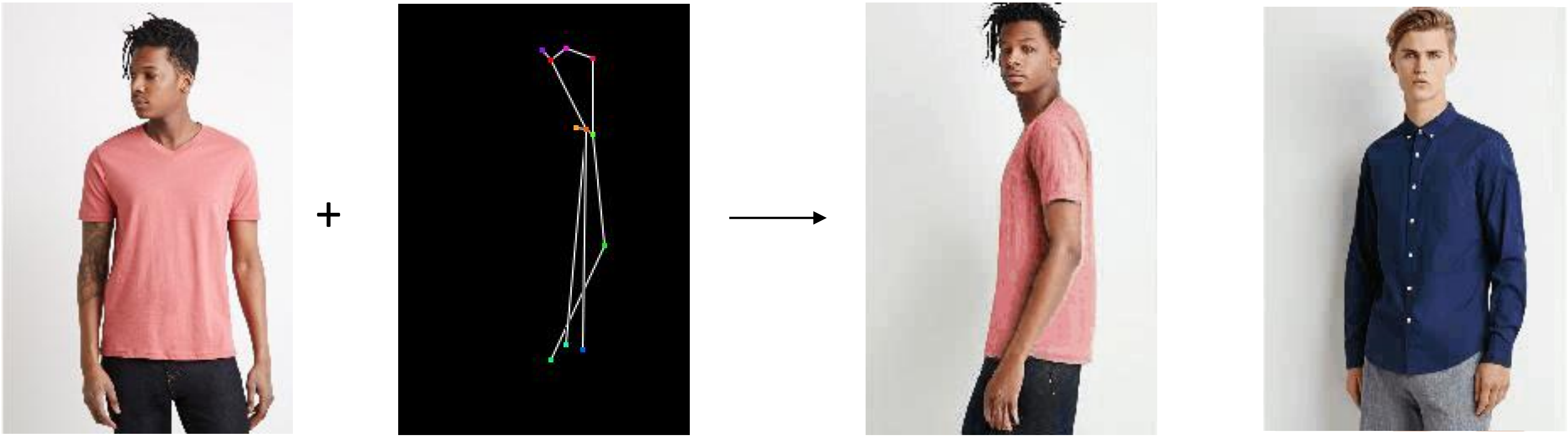
→

Target image

Head

What We Can Do

- Achieve flexible and continuous control of **human attributes**, including **pose** and **component attributes** (i.e. **head**, **upper clothes** and **pants**)



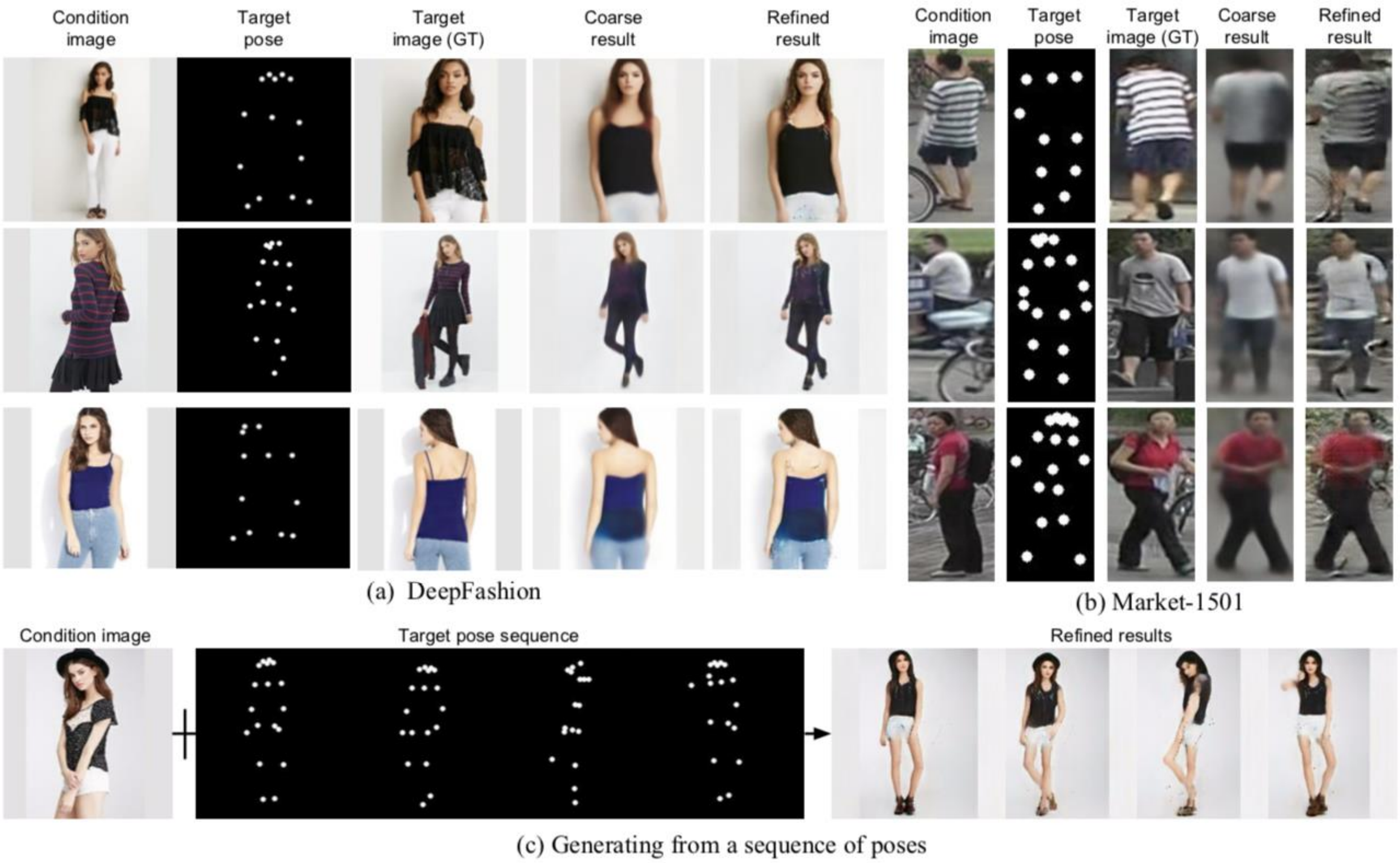
Reference | Target pose | Target image | **Upper clothes**

What We Can Do

- Also bring a significant **quality boost** compared with previous methods

Previous Work

Pose Transfer PG2 --> DPIG --> Def-GAN --> … --> PATN



"Pose Guided Person Image Generation." [Ma et al. NIPS 2017]

- Unsatisfied quality
- Limited controllability

Previous Work

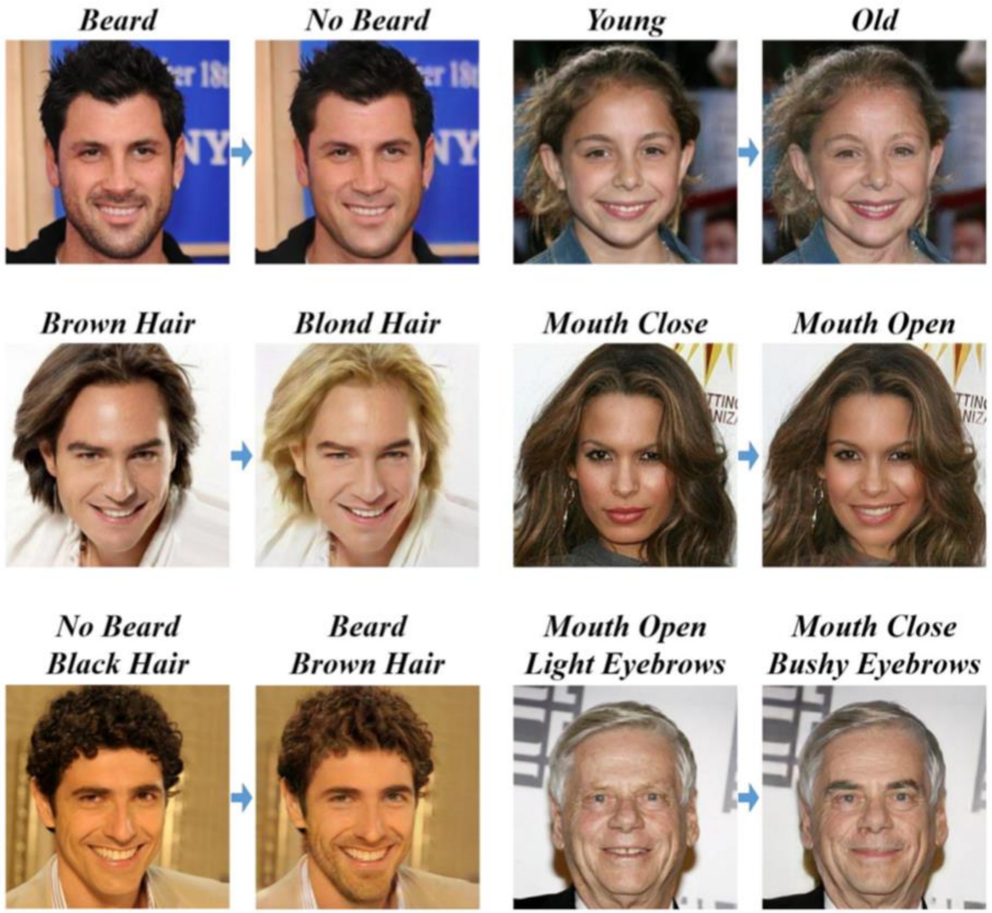
Appearance Transfer

“A Generative Model of People in Clothing.” [Lassner et al. ICCV 2017]

- Fail to model the intricate interplay of the inherent pose --> deformed textures

Previous Work

Facial Attributes Editing



- Require strict attribute annotation
 (e.g., smiling, beard and eyeglasses exist or not in the training dataset)

"AttGAN: Facial Attribute Editing by Only Changing What You Want." [He et al. TIP 2019]

Contributions

We introduce Attribute-Decomposed GAN, a neat and effective model

- It achieves flexible and continuous user control of human attributes, such as head, pant and upper clothes.
- It brings a significant quality boost for the original person image synthesis task.
-
- It makes an automatic and unsupervised attribute separation.

What We Have

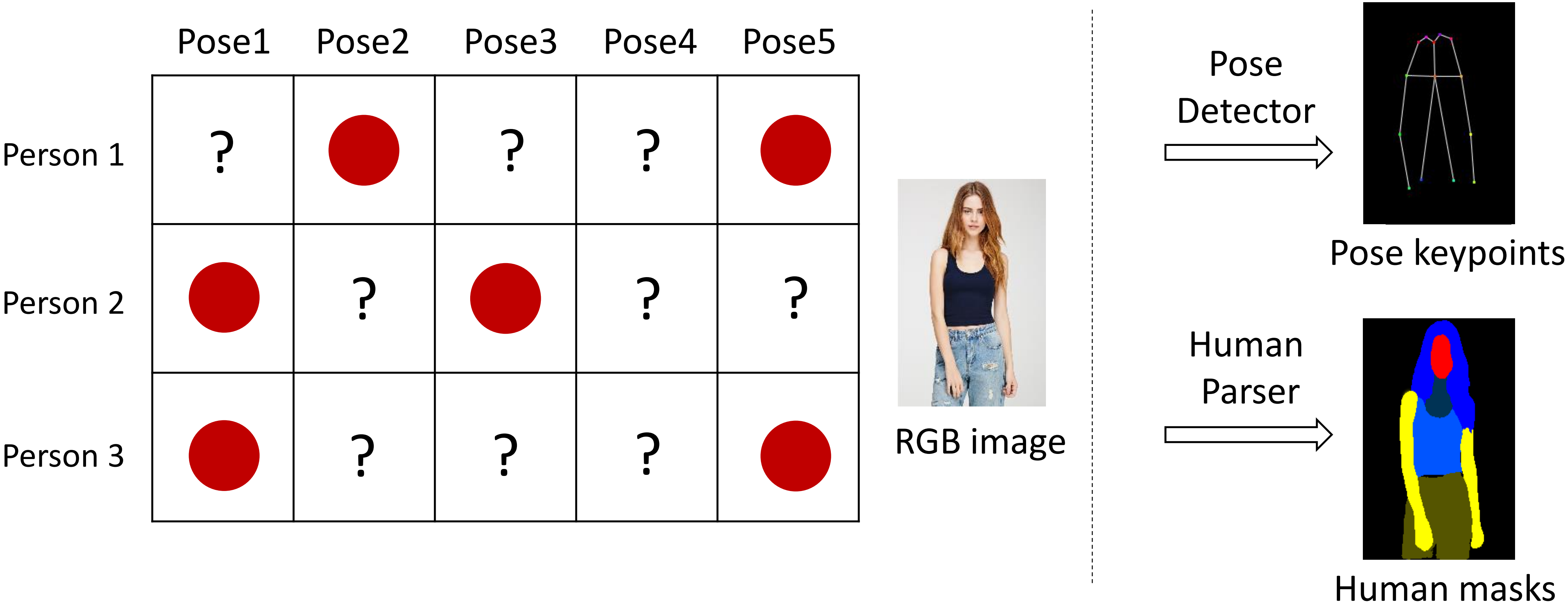
- A typical person image dataset (e.g. DeepFashion)

	Pose1	Pose2	Pose3	Pose4	Pose5
Person 1	?	●	?	?	●
Person 2	●	?	●	?	?
Person 3	●	?	?	?	●

RGB image

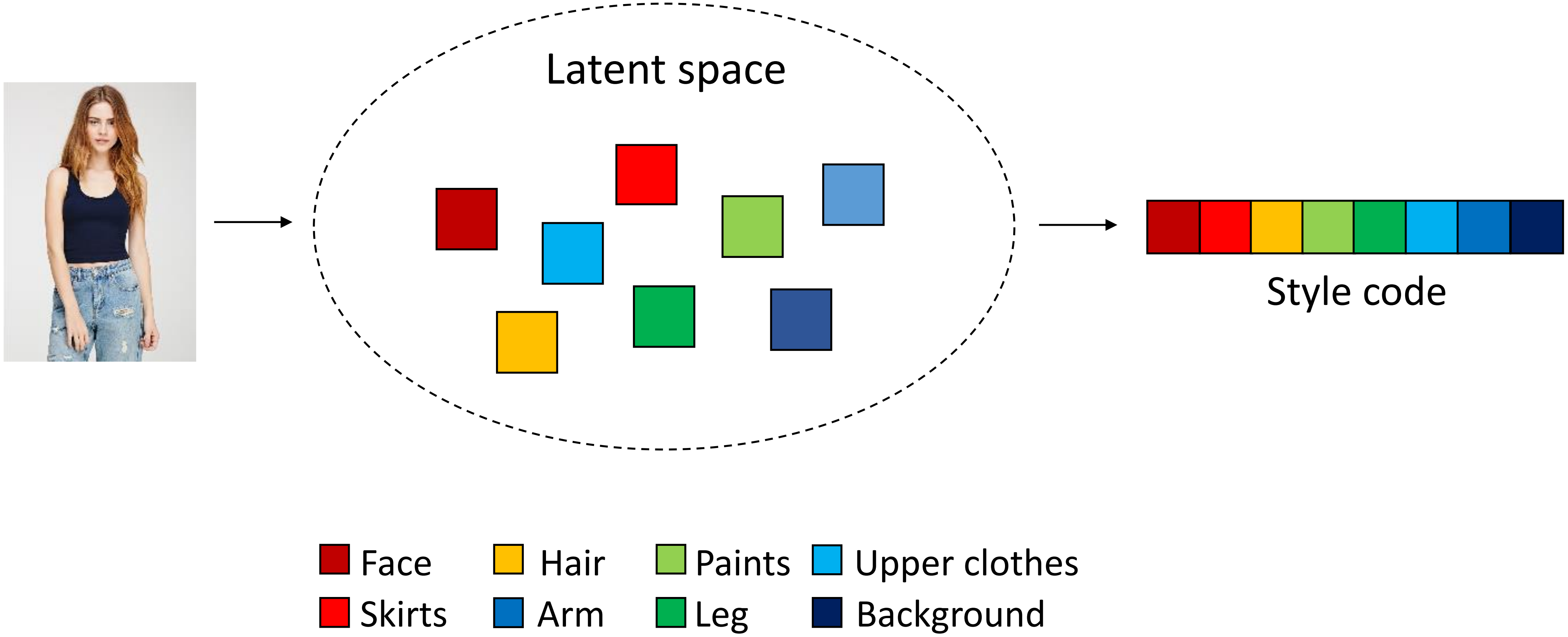
What We Have

- A typical person image dataset (e.g. DeepFashion)



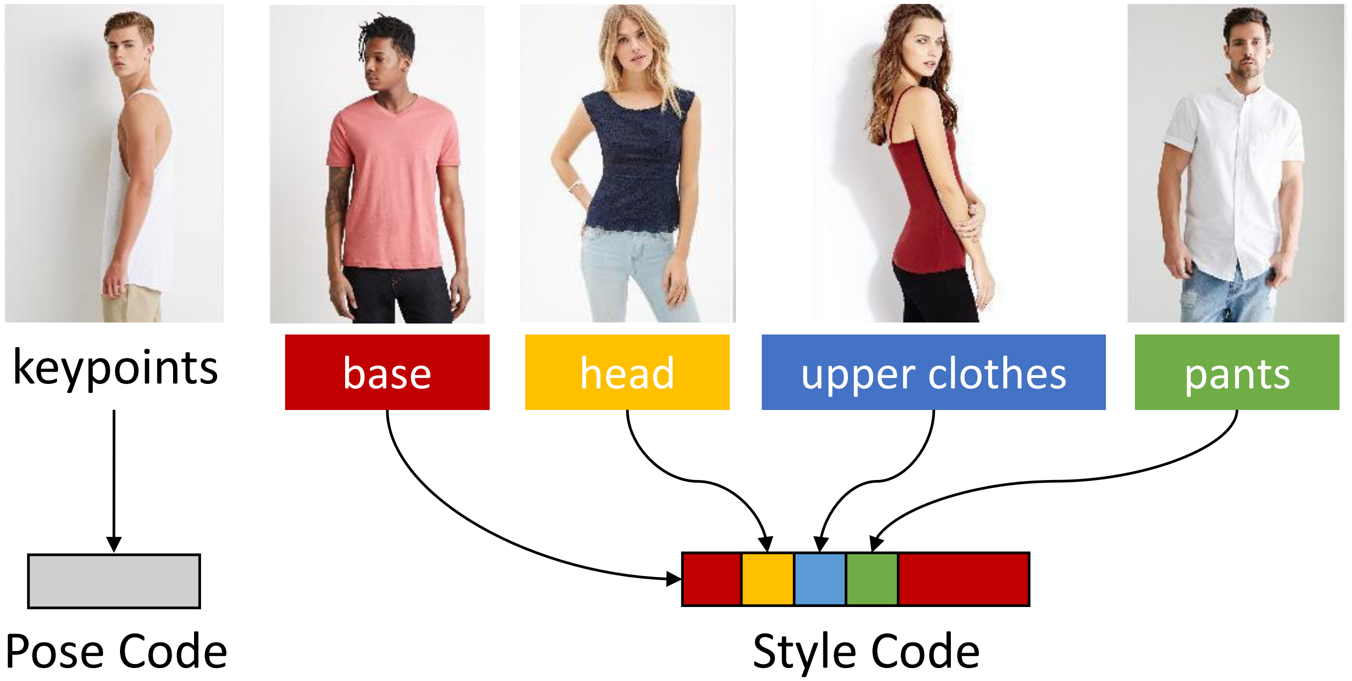
The Core Idea

- Learn an automatic, unsupervised attribute separation



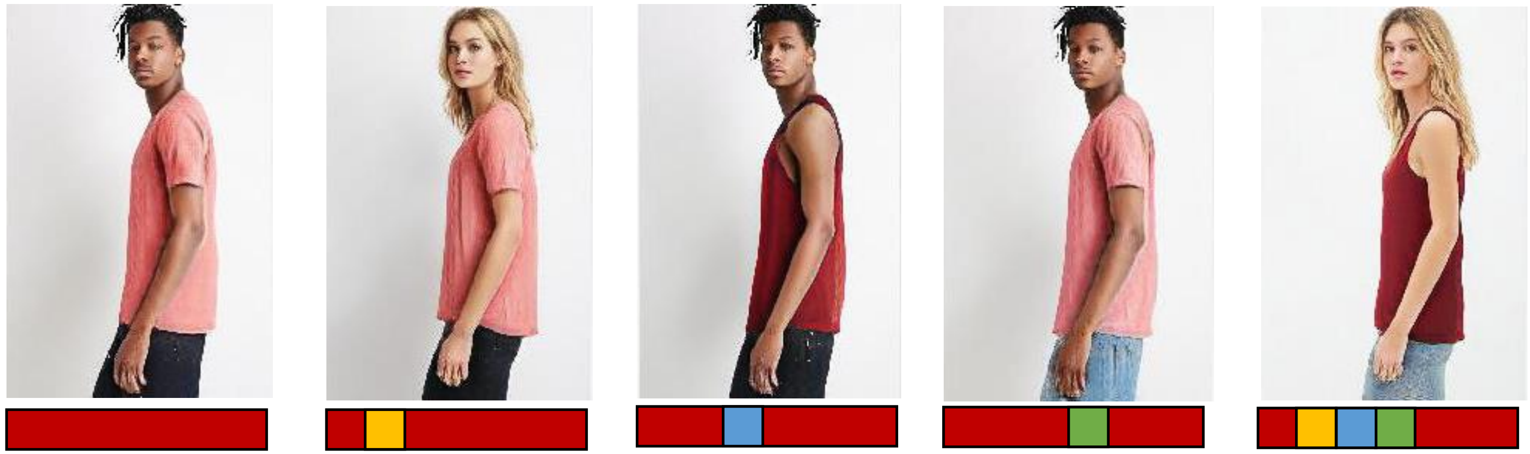
The Core Idea

- Multiple source persons with desired attributes -> style code



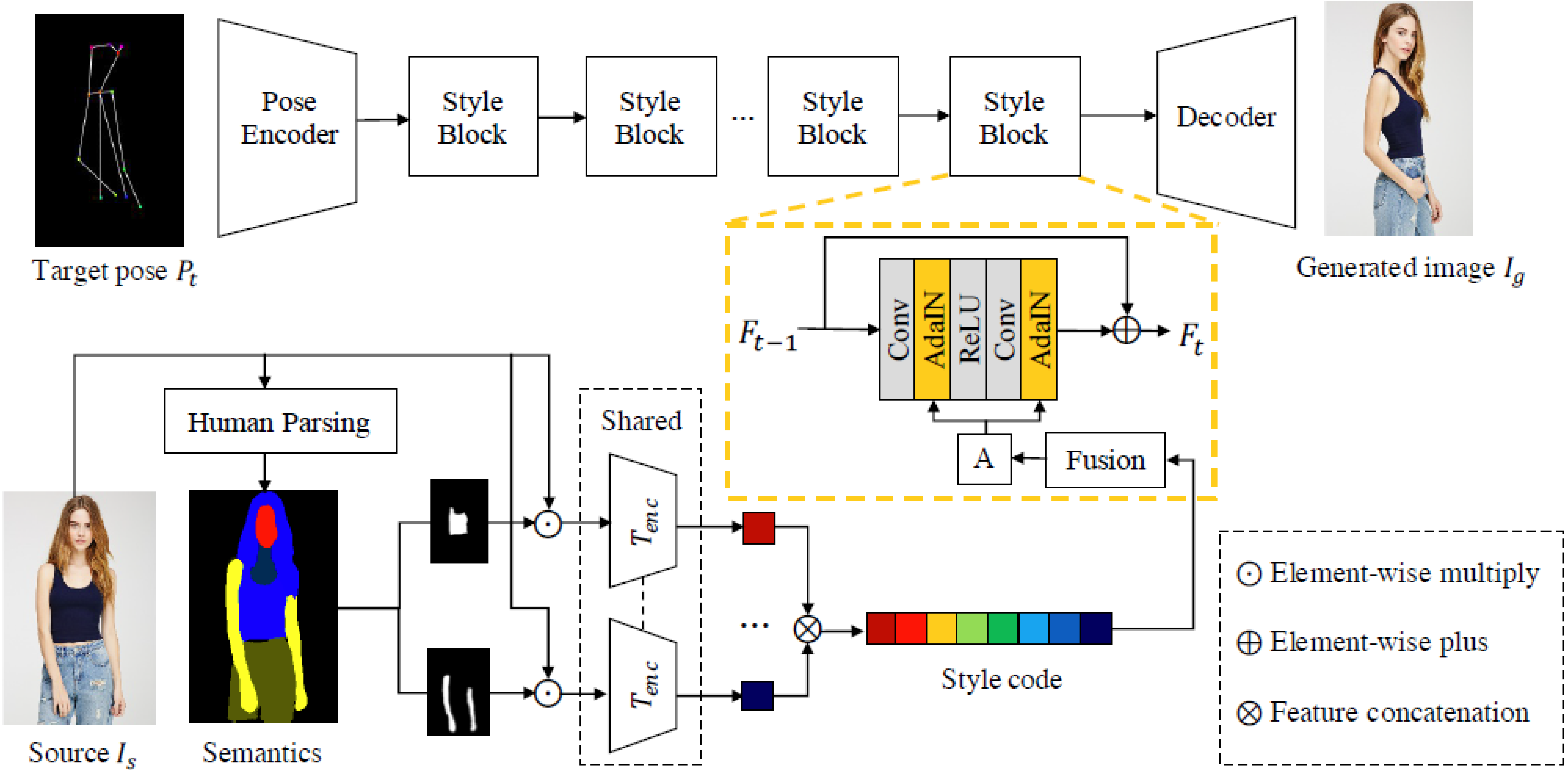
The Core Idea

- Control human attributes by editing the style code

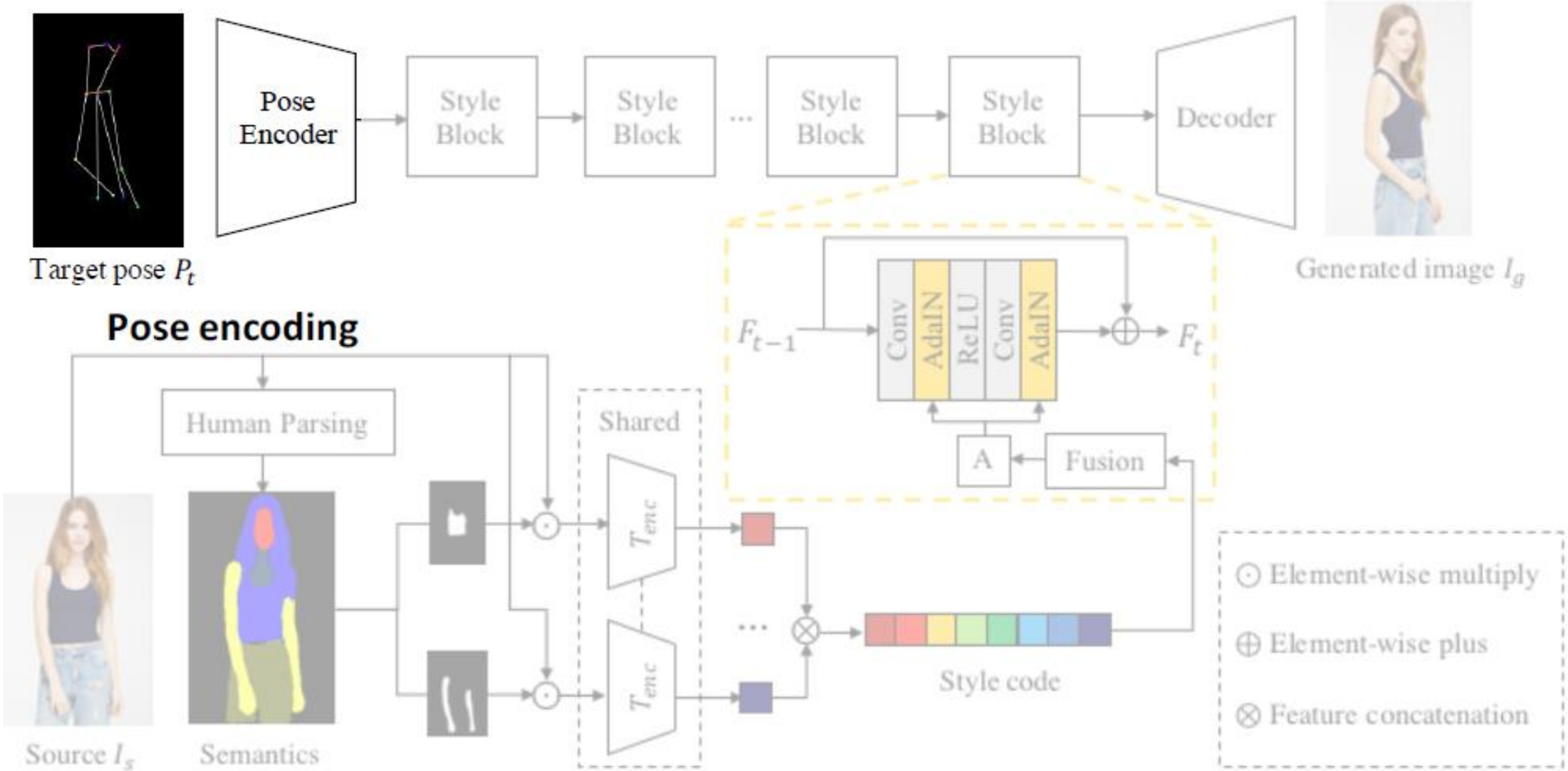


Generated images with varying style codes

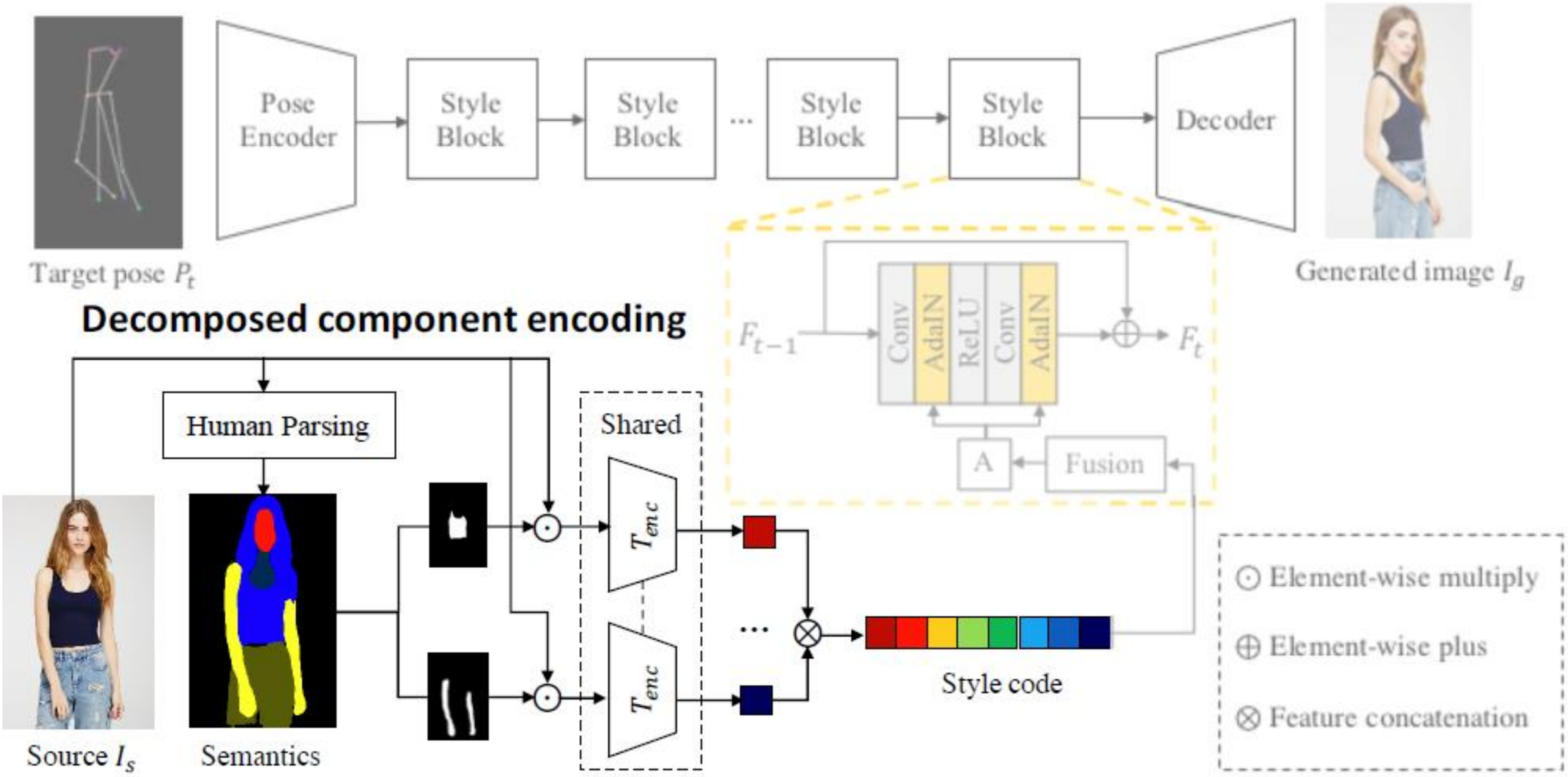
Generator Architecture



Generator Architecture

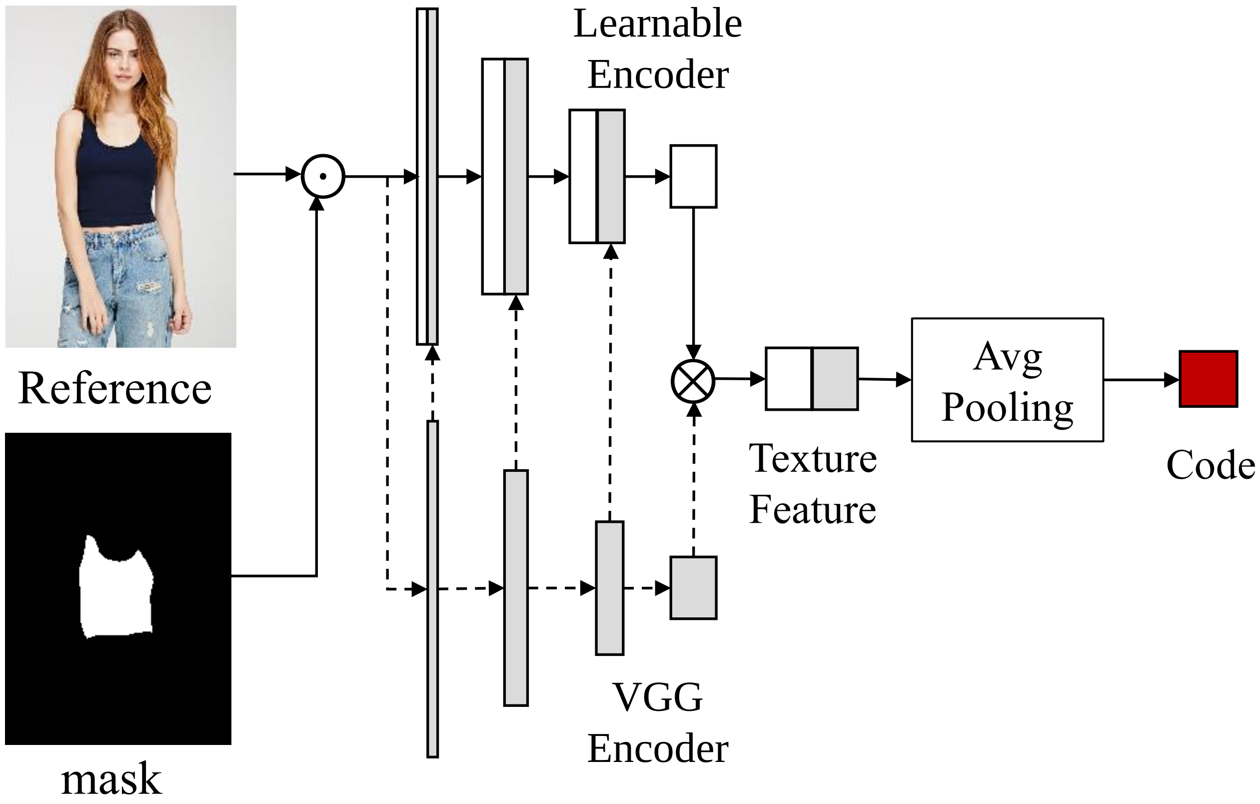


Generator Architecture



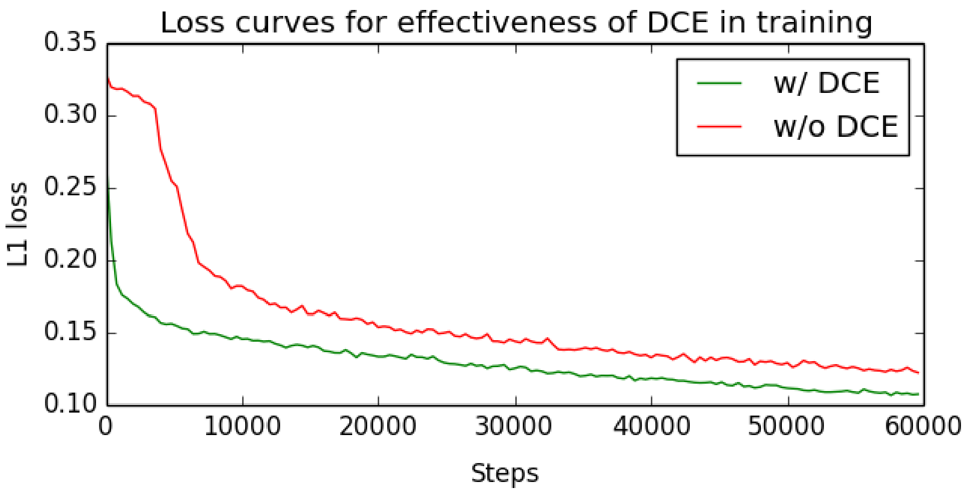
Generator Architecture

Details of global texture encoder T_{enc}



Generator Architecture

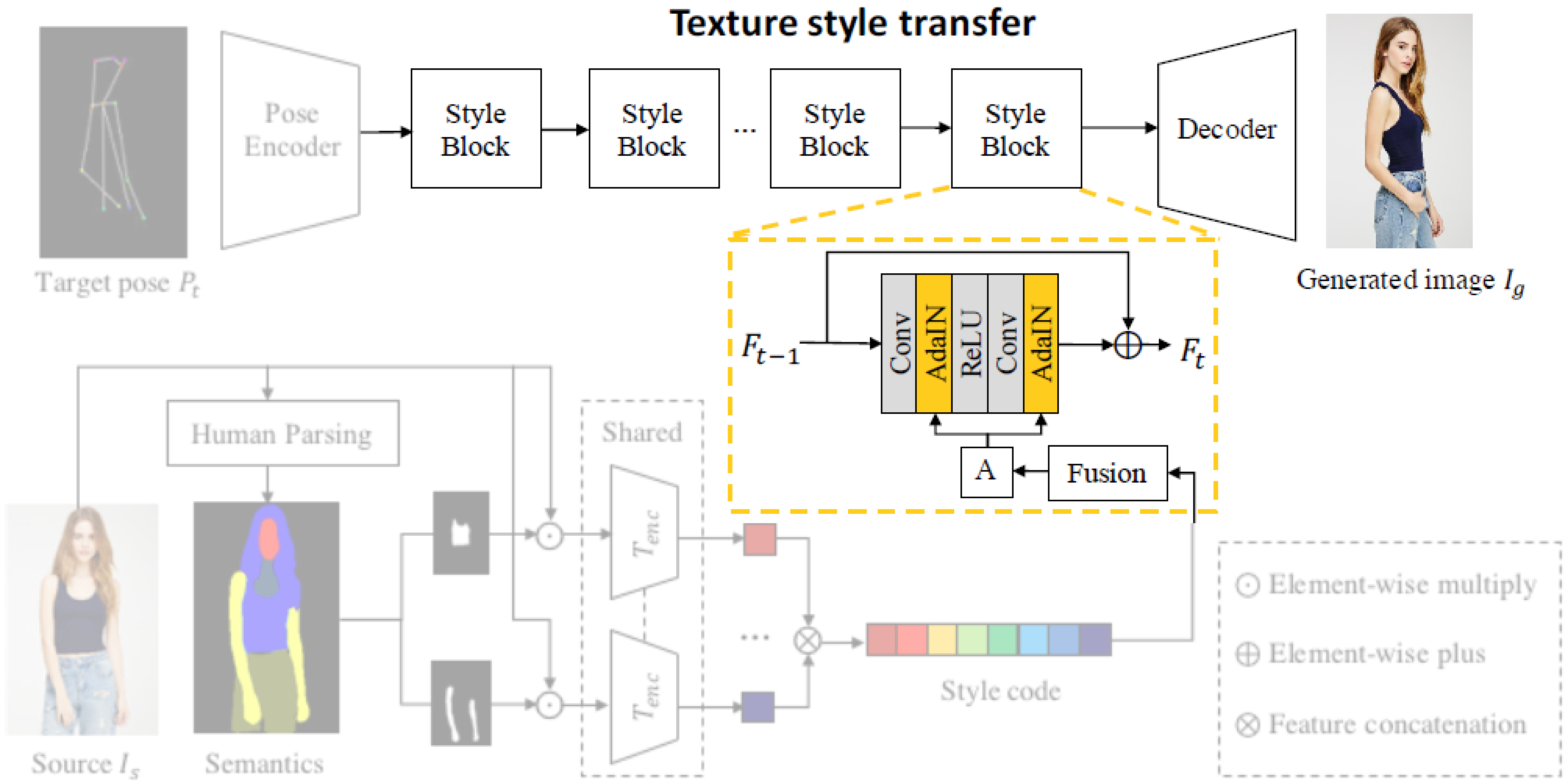
Effects of the decomposed component encoding (DCE) & global texture encoding (GTE)



(a) Source (b) Target pose (c) w/o DCE&GTE (d) w/o DCE (e) Full Model

speeds up the convergence of model and achieves more realistic results in less time.

Generator Architecture



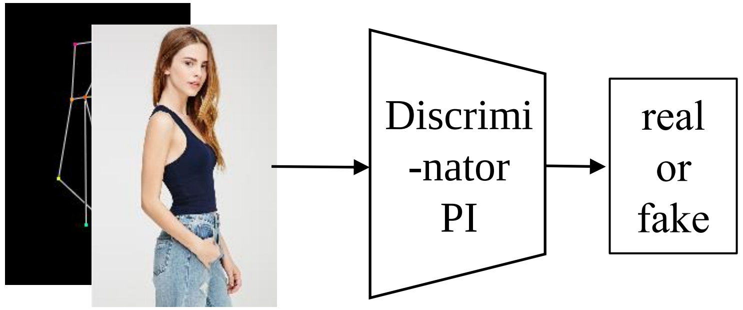
Generator Architecture

Auxiliary effects of the fusion module (FM) for DCE

(a) Source (b) Target pose (c) w/o DCE (d)w/ DCE, w/o FM (e)w/ DCE&FM

Discriminator

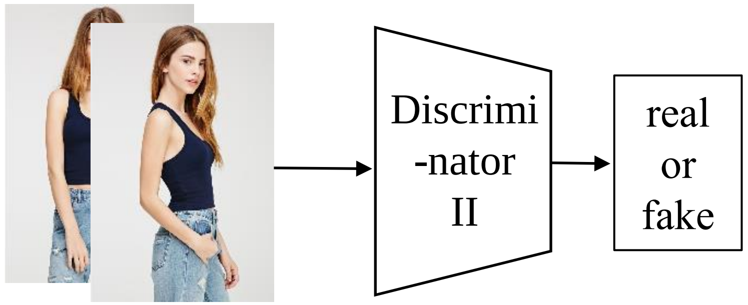
Discriminator D_p: pose consistence



$\{$target pose P_t, target image $I_t\}$—real pair

$\{$target pose P_t, generated image $I_g\}$—fake pair

Discriminator D_t: texture coherence



$\{$source image I_s, target image $I_t\}$—real pair

$\{$source image I_s, generated image $I_g\}$—fake pair

Loss Function

(1) Adversarial loss with two discriminator and one generator

$$\mathcal{L}_{adv} = \mathbb{E}_{P_t \in \mathcal{P},\, (I_s, I_g) \in \mathcal{I}} \{\log[D_{II}(I_s, I_g) \cdot D_{PI}(P_t, I_g)]\} + \mathbb{E}_{P_t \in \mathcal{P},\, I_s \in \mathcal{I}, I_t \in \hat{\mathcal{I}}} \{\log[(1 - D_{II}(I_s, I_t)) \cdot (1 - D_{PI}(P_t, I_t))]\}$$

(2) L1 loss between fake generated image and ground truth

$$\mathcal{L}_1 = \| I_g - I_t \|_1$$

(3) Perception loss with extracted feature via pretrained-VGG

$$\mathcal{L}_{per} = \| \mathcal{G}(\mathcal{F}^l(I_t)) - \mathcal{G}(\mathcal{F}^l(I_g)) \|_2$$

(4) **Contextual loss** for unaligned pairs (effective in person image synthesis task)

$$\mathcal{L}_{CX} = -\log(\mathrm{CX}(\mathcal{F}^l(I_t), \mathcal{F}^l(I_g)))$$

The Contextual Loss for Image Transformation with Non-Aligned Data. [Mechrez et al. ECCV 2018]

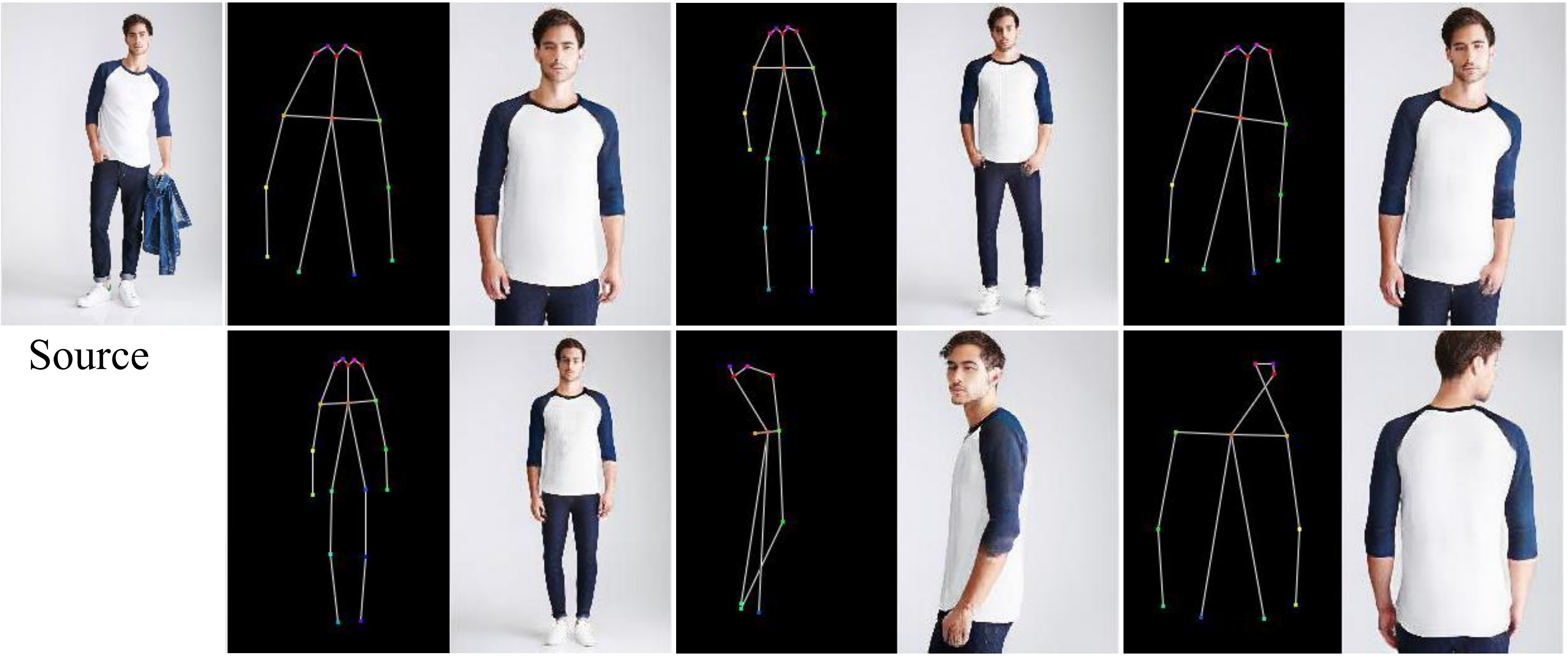
Loss Function

Effects of the contextual loss

(a) Source (b) Target pose (c) Ground-truth (d) w/o CX (e) w/ CX

Results

Pose Transfer (synthesize person images in arbitrary poses)



Source

Results

Pose Transfer (synthesize person images in arbitrary poses)

Source

Results

Motion Transfer (animate a single person image with driving action video)

Results

Comparison with state-of-the-arts

Results

Comparison with state-of-the-arts

Table 1: Quantitative comparison on DeepFashion.

Model	IS↑	SSIM↑	DS↑	CX-GS↓	CX-GT↓
PG^2	3.202	0.773	0.943	2.854	2.795
DPIG	3.323	0.745	0.969	2.761	2.753
Def-GAN	2.265	0.770	0.973	2.751	2.713
PATN	3.209	**0.774**	0.976	2.628	2.604
Ours	**3.364**	0.772	**0.984**	**2.474**	**2.474**

Table 2: Results of the user study (%).

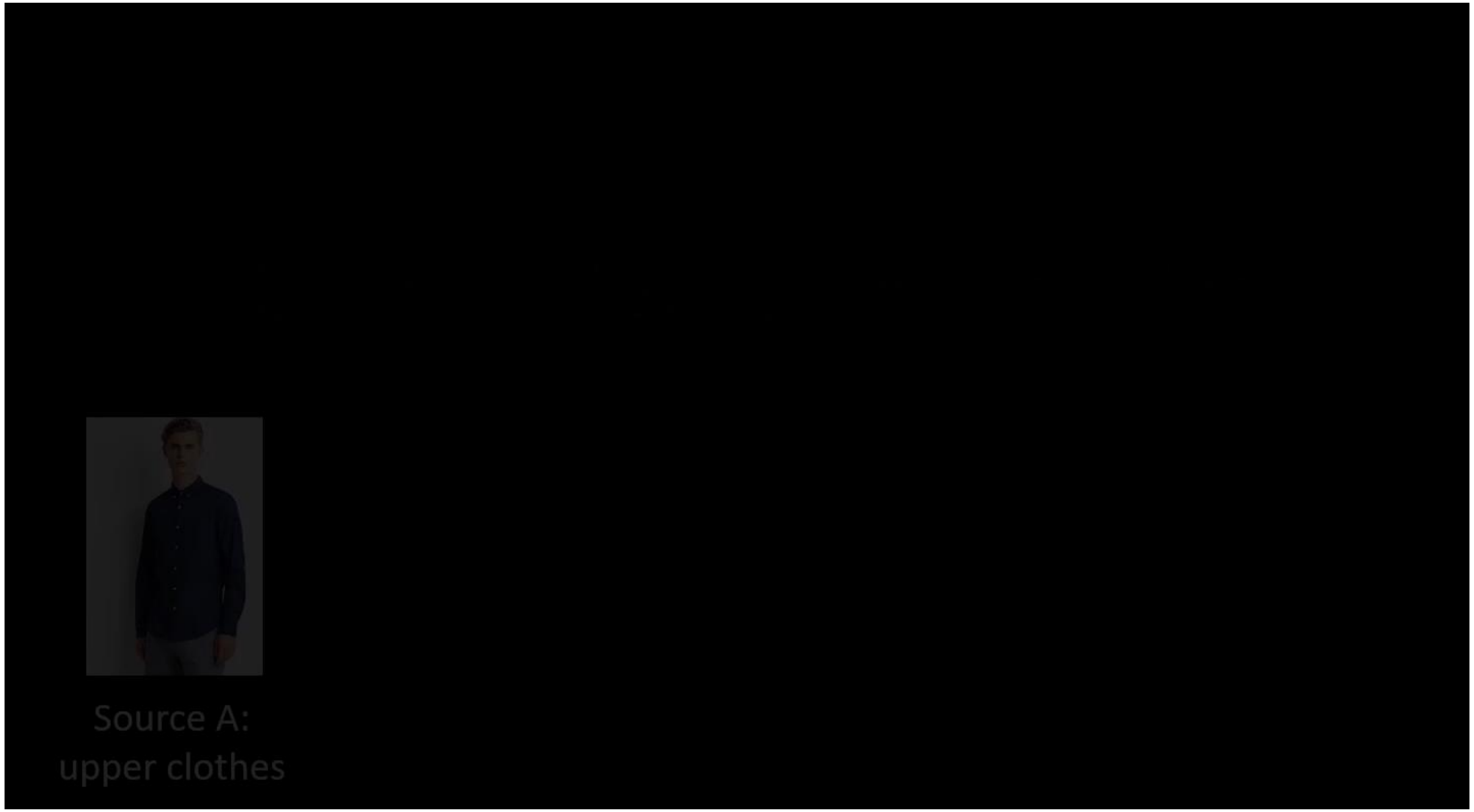
Indicator	PG^2	DPIG	Def-GAN	PATN	Ours
R2G	9.2	-	12.42	19.14	**23.49**
G2R	14.9	-	24.61	31.78	**38.67**
Prefer	1.61	1.35	16.23	7.26	**73.55**

Results

Component Attribute Transfer (transfer pants or upper clothes from left exemplar)

Results

Component Attribute Transfer (provide continuous user control)

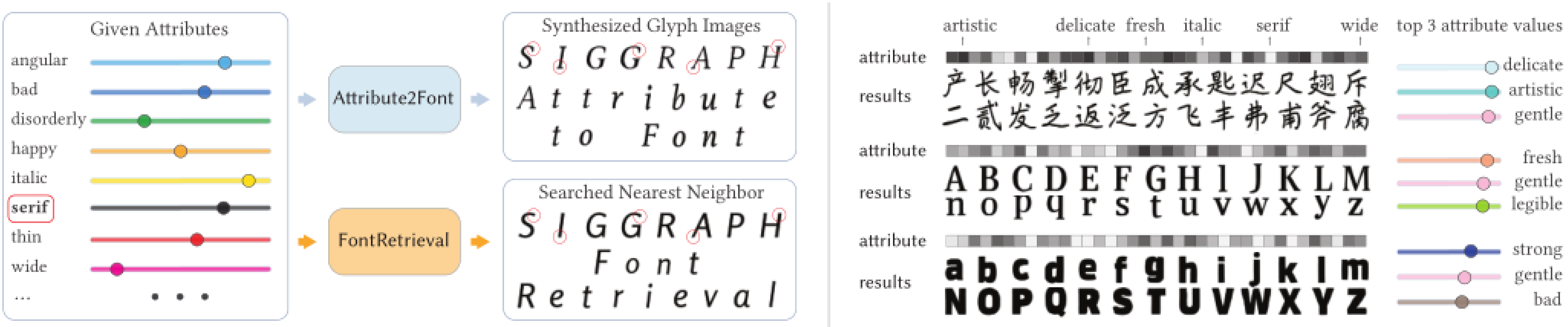


Another recently-published work

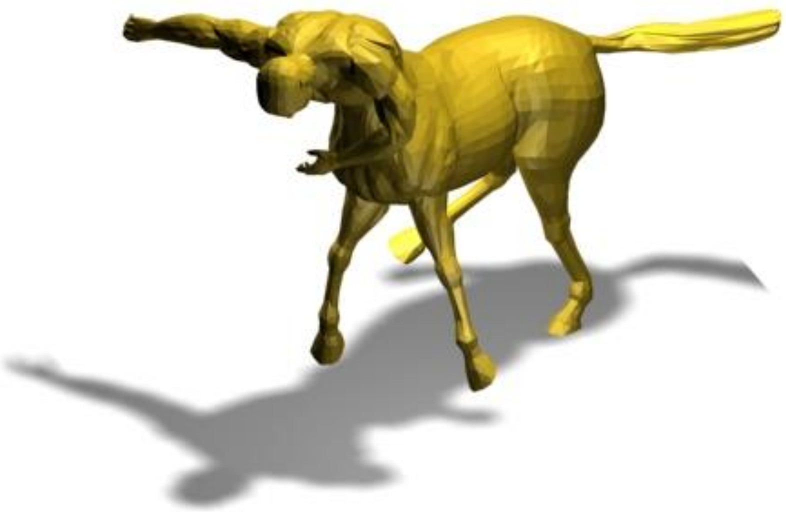
ACM TOG (SIGGRAPH 2020)

Attribute2Font: Creating Fonts You Want From Attributes

YIZHI WANG*, Wangxuan Institute of Computer Technology, Peking University, China
YUE GAO*, Wangxuan Institute of Computer Technology, Peking University, China
ZHOUHUI LIAN†, Wangxuan Institute of Computer Technology, Peking University, China



Thank you!



北京大学
PEKING UNIVERSITY
王选计算机研究所
字形计算技术实验室
http://www.wict.pku.edu.cn/cscl
lianzhouhui@pku.edu.cn