Controllable Person Image Synthesis with Attribute-Decomposed GAN

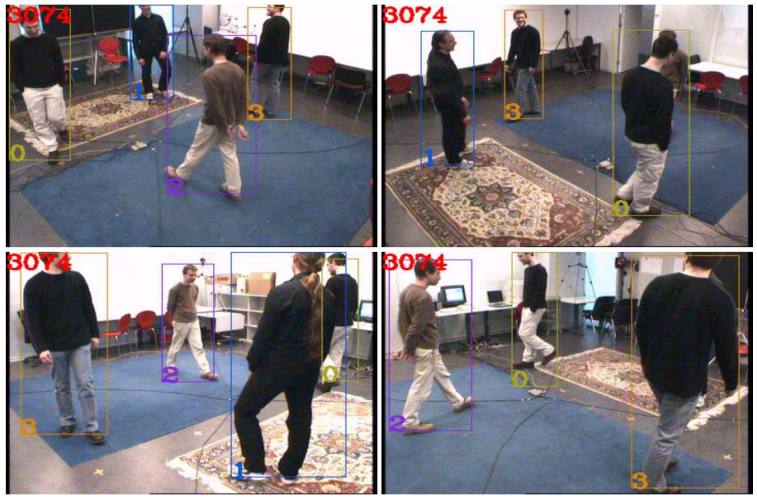
Muhammad Kashif Ali

Visual Intelligence Lab.

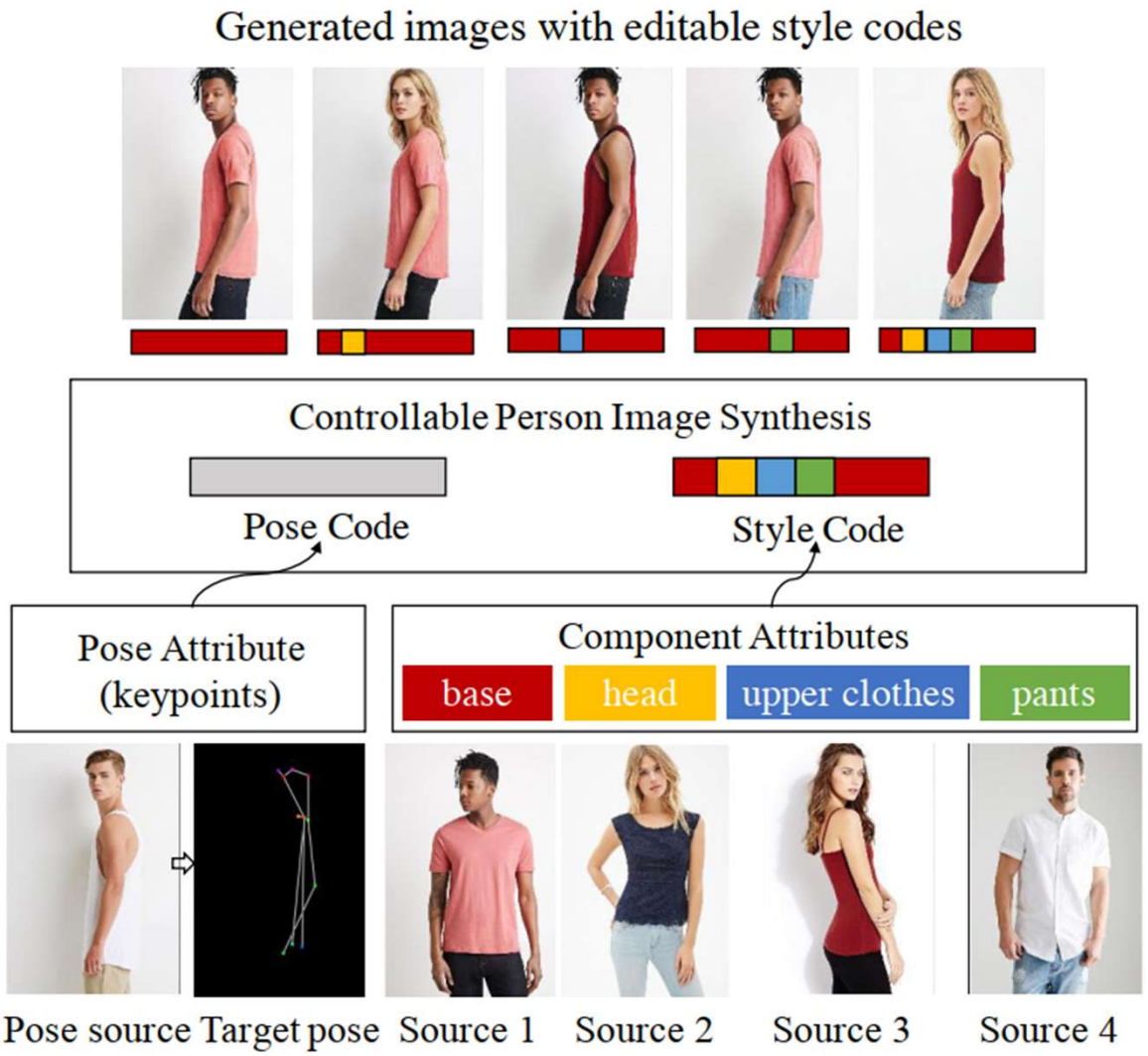
Contents

- Introduction + Problems + Scope of this research
- Method
  - Observations
  - Proposed solutions to fill the research gap (Main contributions: Architecture + Attribute targetting)
  - Implementation details
- Results
- My thoughts + What we can learn from this paper

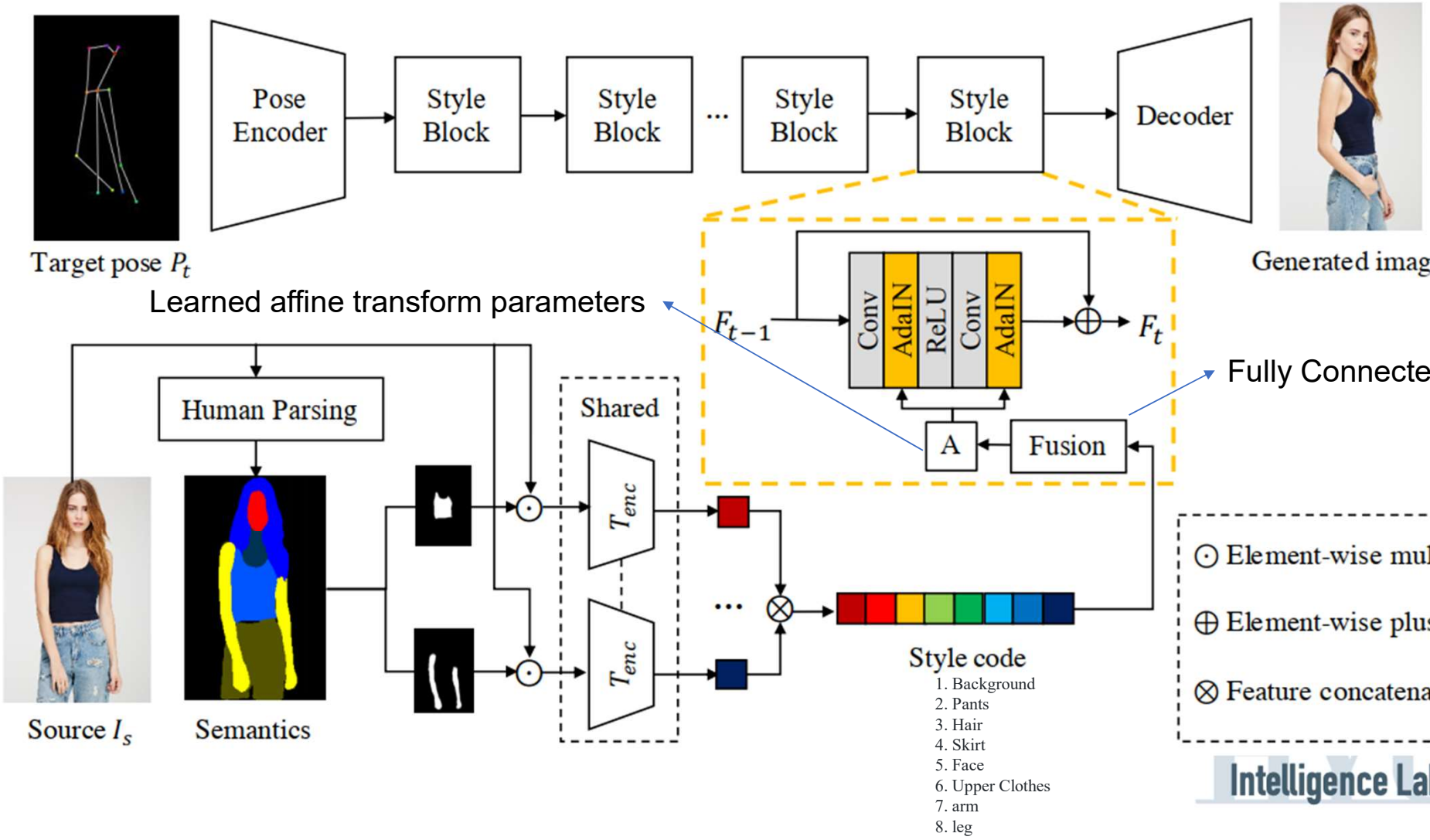
Visual Intelligence Lab.

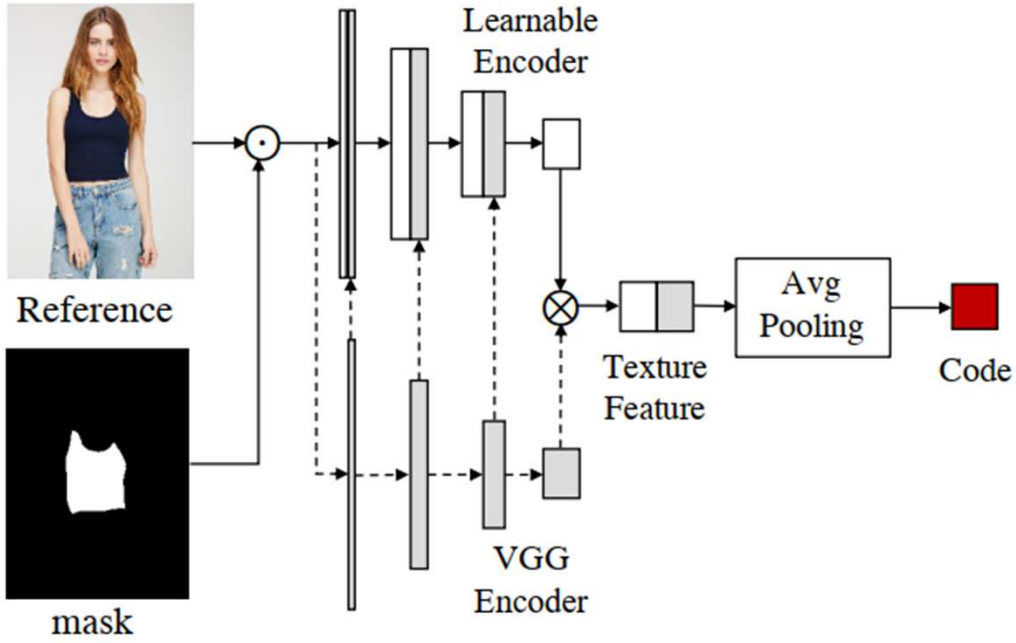


Visual Intelligence Lab.



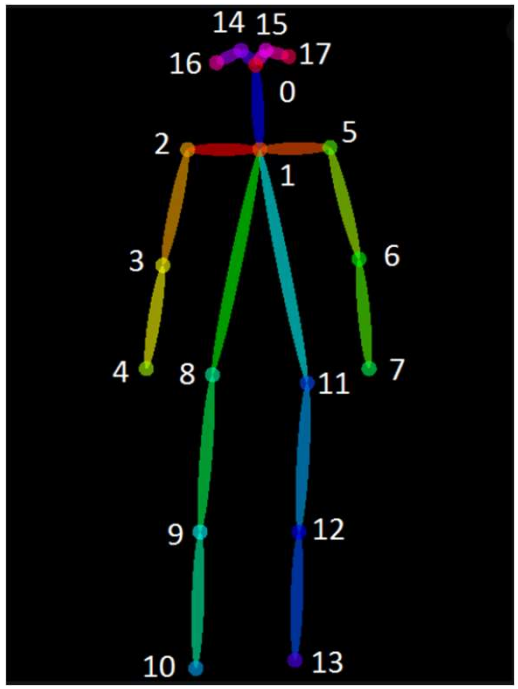
Visual Intelligence Lab.


Target pose P_t
Pose Encoder
Style Block
Style Block
...
Style Block
Style Block
Decoder
Generated image I_g
Learned affine transform parameters
F_{t-1}
Conv
AdaIN
ReLU
Conv
AdaIN
F_t
A
Fusion
Fully Connected Layer
Human Parsing
Shared
T_{enc}
T_{enc}
Source I_s
Semantics
Style code
1. Background
2. Pants
3. Hair
4. Skirt
5. Face
6. Upper Clothes
7. arm
8. leg
⊙ Element-wise multiply
⊕ Element-wise plus
⊗ Feature concatenation
Intelligence Lab.



Texture Encoder Network

Visual Intelligence Lab.



Open-pose for target pose estimation

Cut down version of a human parser limited to 8 categories

Visual Intelligence Lab.

$$\mathcal{L}_{total} = \mathcal{L}_{adv} + \lambda_{rec}\mathcal{L}_{rec} + \lambda_{per}\mathcal{L}_{per} + \lambda_{CX}\mathcal{L}_{CX}$$

$$\begin{aligned} L_{adv} =& \mathbb{E}_{I_s,P_t,I_t}[log(D_t(I_s, I_t) \cdot D_p(P_t, I_t))] + \\ & \mathbb{E}_{I_s,P_t}[log((1 - D_t(I_s, G(I_s, P_t))) \\ & \cdot (1 - D_p(P_t, G(I_s, P_t))))]. \end{aligned}$$

$$L_{rec} = ||G(I_s, P_t) - I_t||_1$$

$$\mathcal{L}_{per} = \frac{1}{W_l H_l C_l} \sum_{x=1}^{W_l} \sum_{y=1}^{H_l} \sum_{z=1}^{C_l} \parallel \phi_l(I_g)_{x,y,z} - \phi_l(I_t)_{x,y,z} \parallel_1$$

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

HYU Visual Intelligence Lab.

Results

Visual Intelligence Lab.

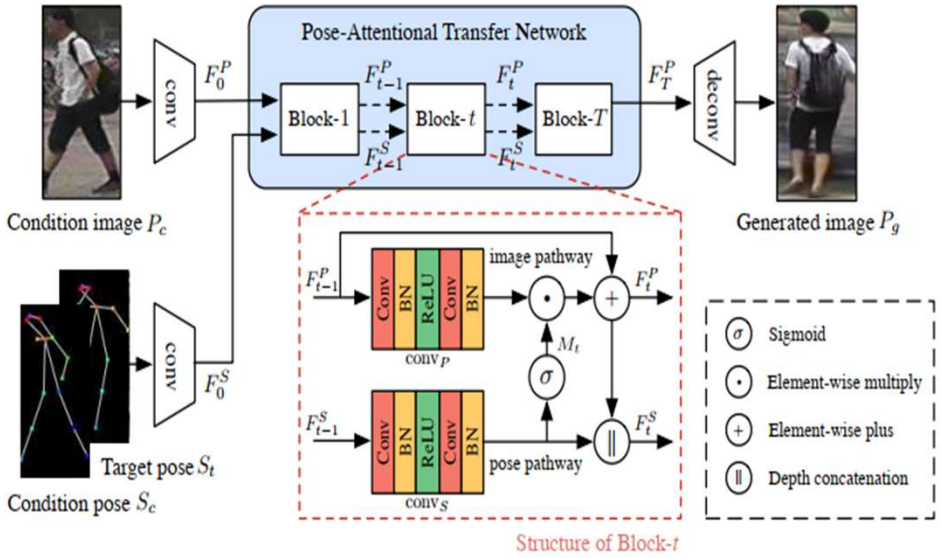
(a) Source (b) Target pose (c) Ground-truth (d) w/o CX (e) w/ CX

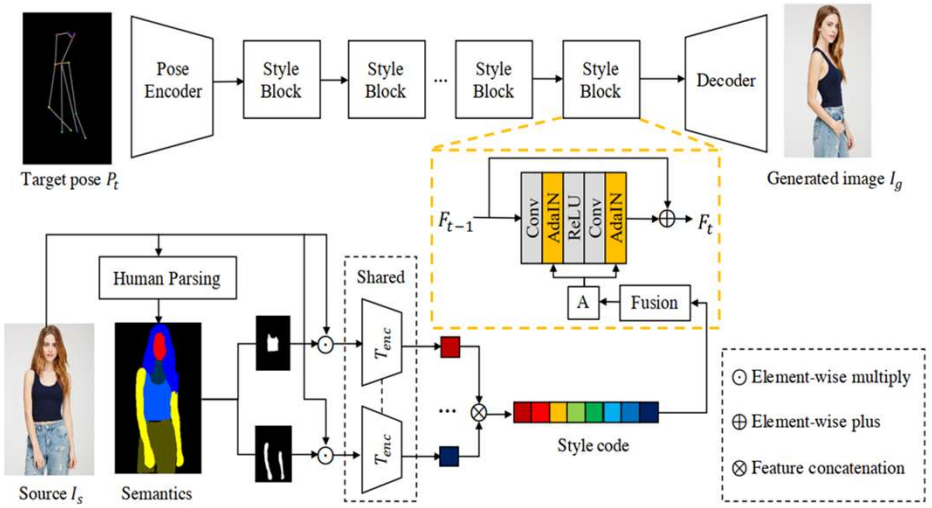
Visual Intelligence Lab.

Visual Intelligence Lab.

Visual
Intelligence Lab.

My Thoughts

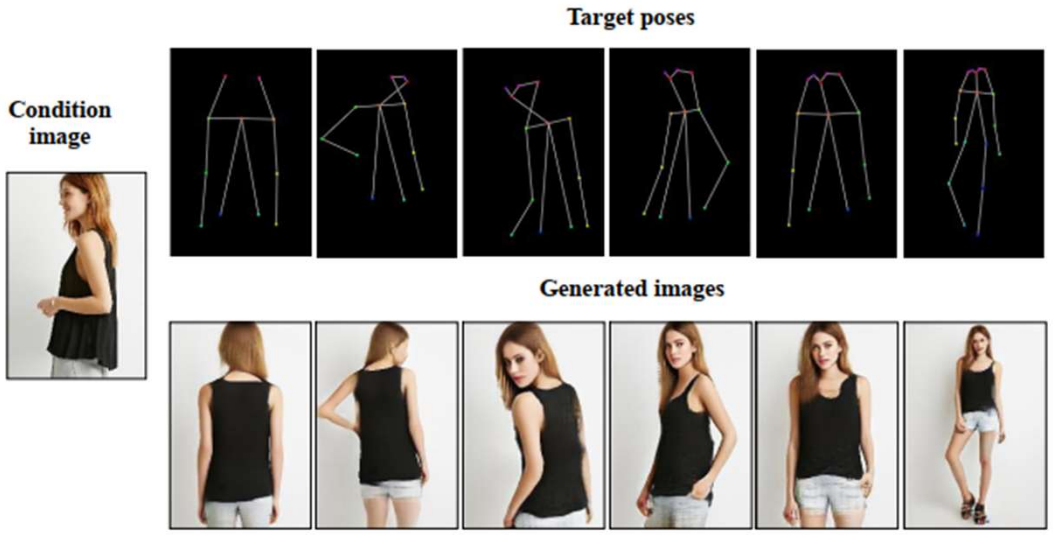




Zhen Zhu, Tengteng Huang, Baoguang Shi, Miao Yu, Bofei Wang, and Xiang Bai. Progressive pose attention transfer for person image generation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2347–2356, **2019**.

This paper

Visual Intelligence Lab.



Zhen Zhu, Tengteng Huang, Baoguang Shi, Miao Yu, Bofei Wang, and Xiang Bai. Progressive pose attention transfer for person image generation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2347–2356, **2019**.

Q & A