

Image-to-Image translation GAN

김현성

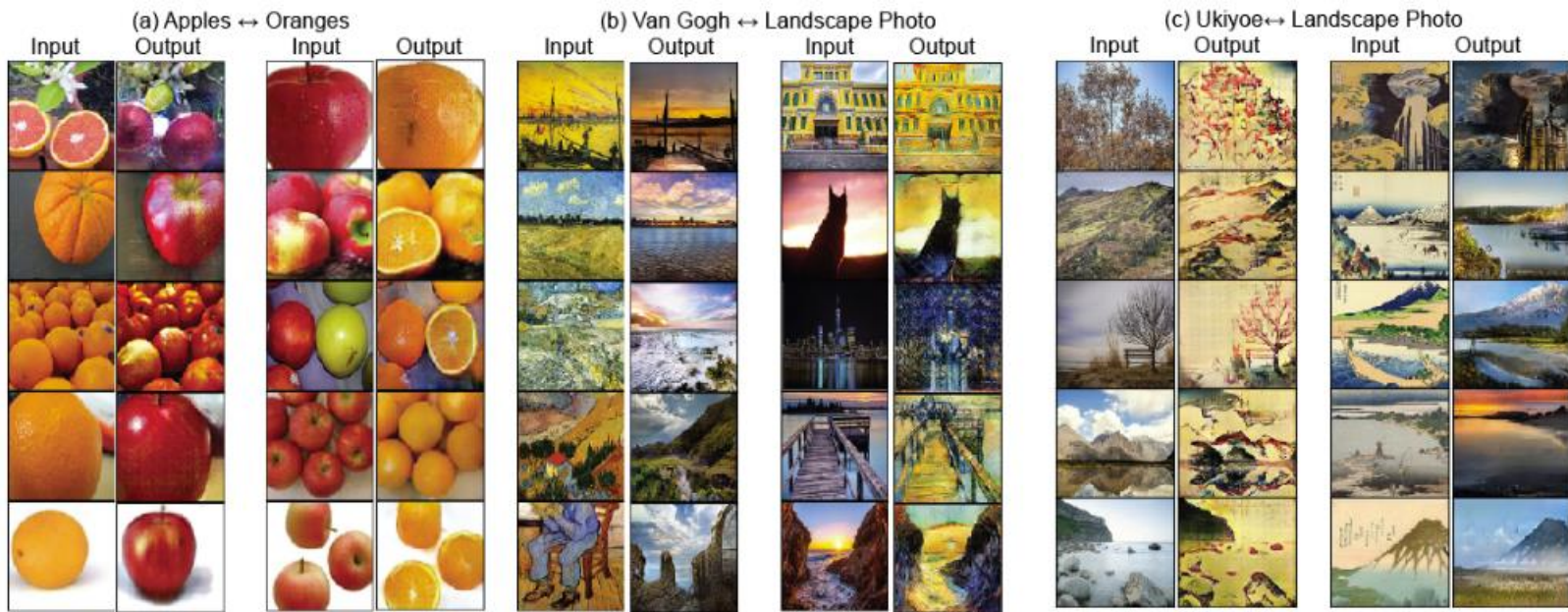
*Vision and Display System Lab.
Sogang University*

Outline

- Introduction
 - Image-to-Image translation GAN
- Associated Papers
 - U-GAT-IT
 - Attribute-Decomposed GAN
- Research

Image-to-Image Translation

- Goal
 - To learn a function that maps images within two different domain



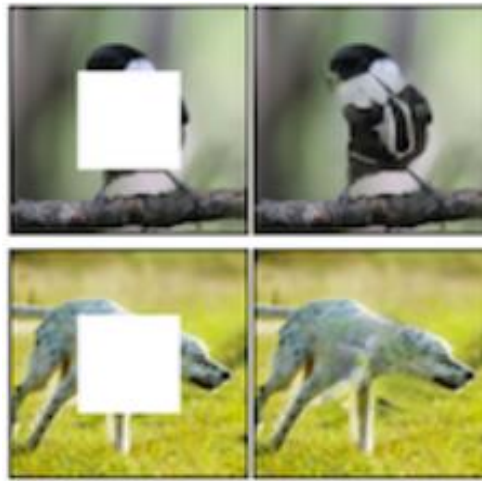
Domain ↔ Domain

Image-to-Image Translation

- Application



Super Resolution



Inpainting



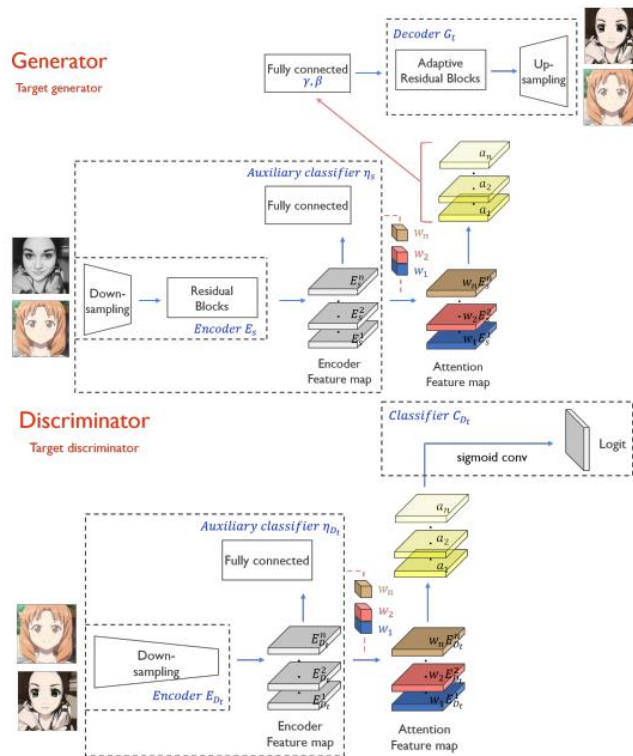
Translation

**Paired Data
(Supervised)**

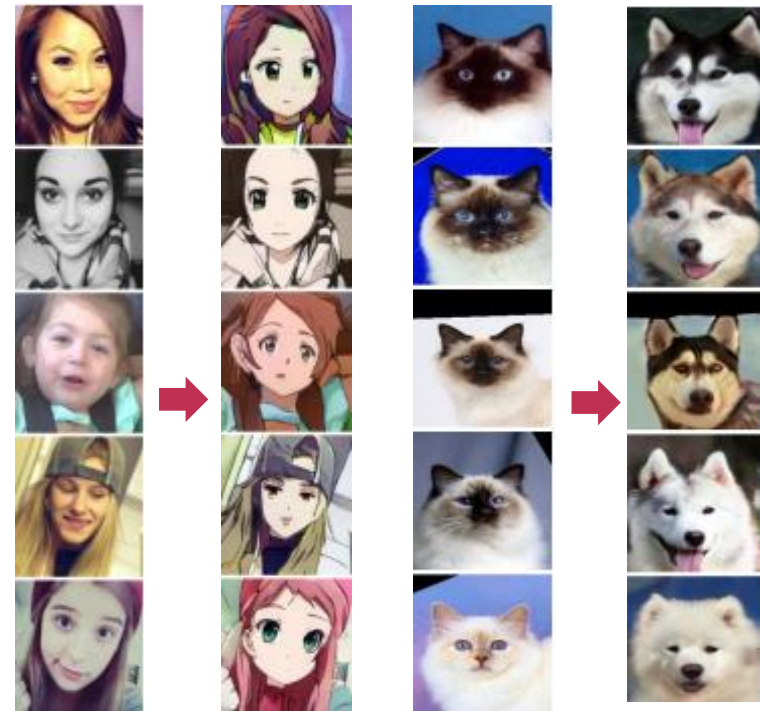
**Unpaired Data
(Unsupervised)**

Associated Papers(Unsupervised)

- U-GAT-IT
 - Overall Network

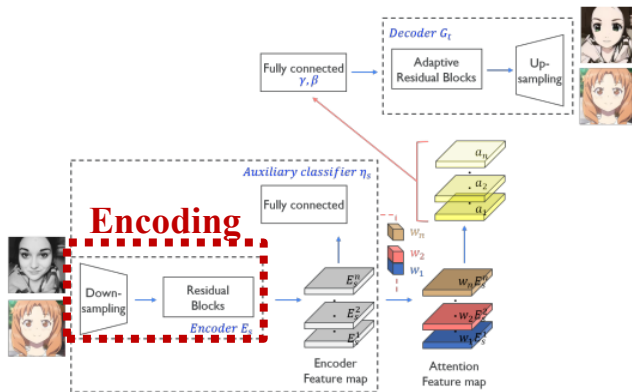


- Generated Images



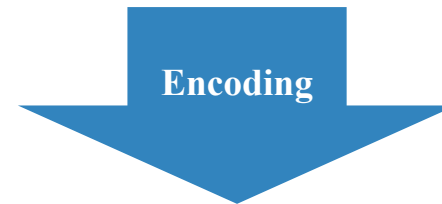
Associated Papers(Unsupervised)

- U-GAT-IT
 - Generator



$$x \in \{X_s, X_t\}$$

$$[B, C, H, W]$$



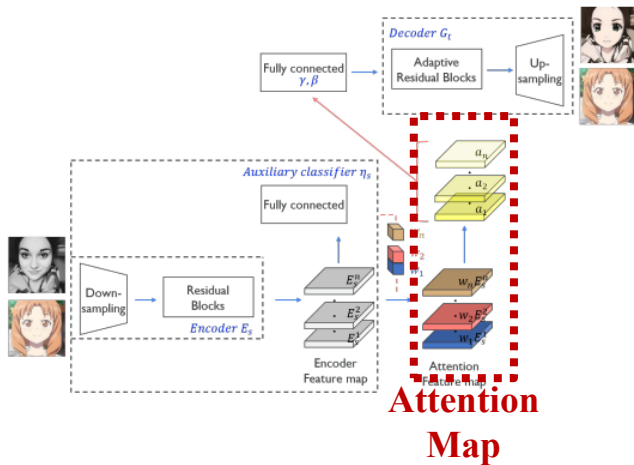
Feature Map

$$[B, 256, H/4, W/4]$$

$G_{S \rightarrow t}$: Generate from source to target domain

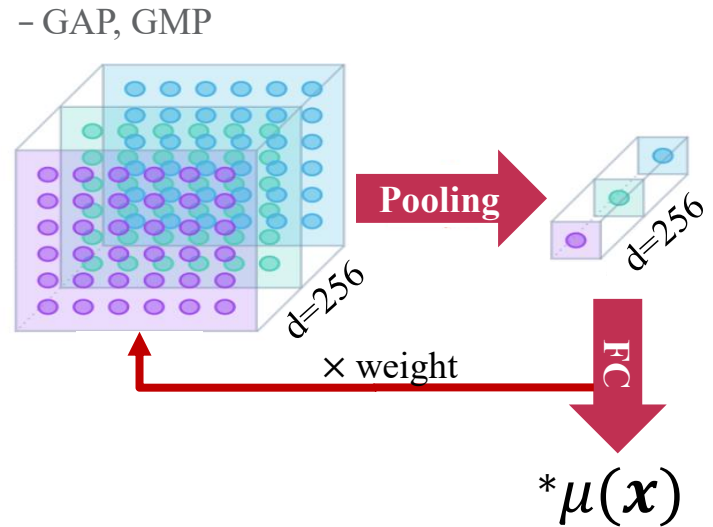
Associated Papers(Unsupervised)

- U-GAT-IT
 - Generator



$G_{S \rightarrow t}$: Generate from source to target domain

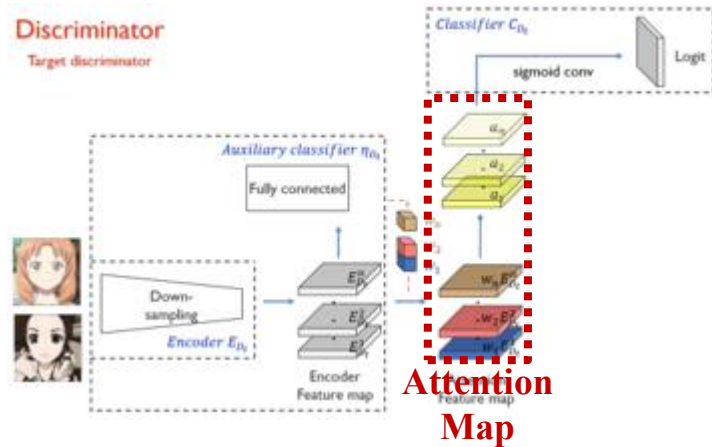
- Attention



*FC Layer is trained to make $\mu(x)$ represents the probability that x is from source Domain

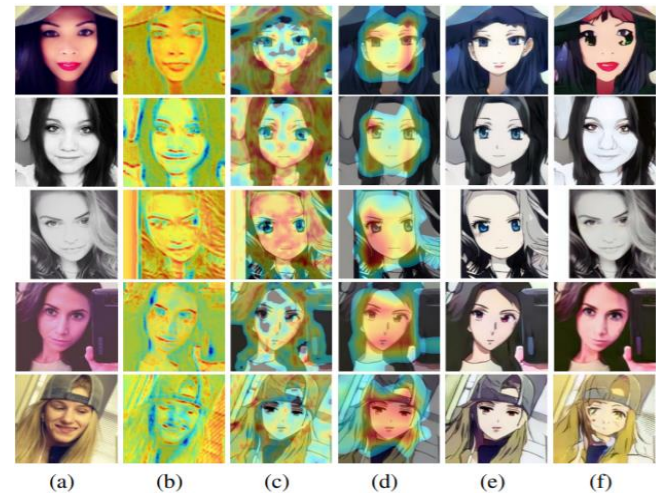
Associated Papers(Unsupervised)

- U-GAT-IT
 - Discriminator



D_t : Discriminator for target domain

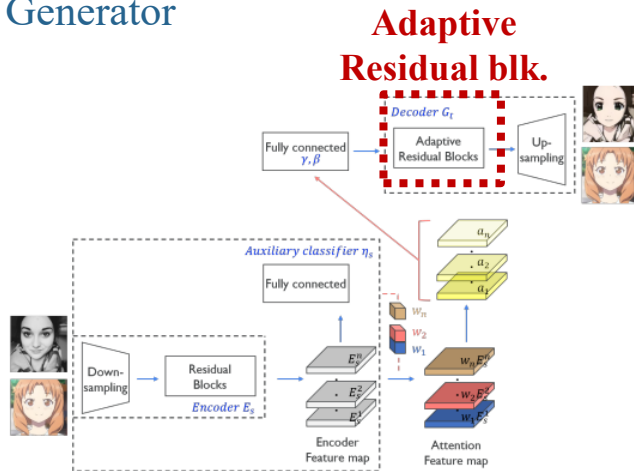
- Attention



(a) Input, (b)/(c)-(d) Visualization of Attention map of Generator/Discriminator, (e)/(f) generated images w/w/o attention

Associated Papers(Unsupervised)

- U-GAT-IT
 - Generator



- AdaLIN

- GAP, GMP

$$AdaLIN(\alpha, \beta, \gamma) = \gamma \cdot (\rho \cdot \hat{a}_I + (1 - \rho) \cdot \hat{a}_L) + \beta$$

$$\hat{a}_I = \frac{a - \mu_I}{\sqrt{\sigma_I^2 + \epsilon}}, \quad \hat{a}_L = \frac{a - \mu_L}{\sqrt{\sigma_L^2 + \epsilon}}$$

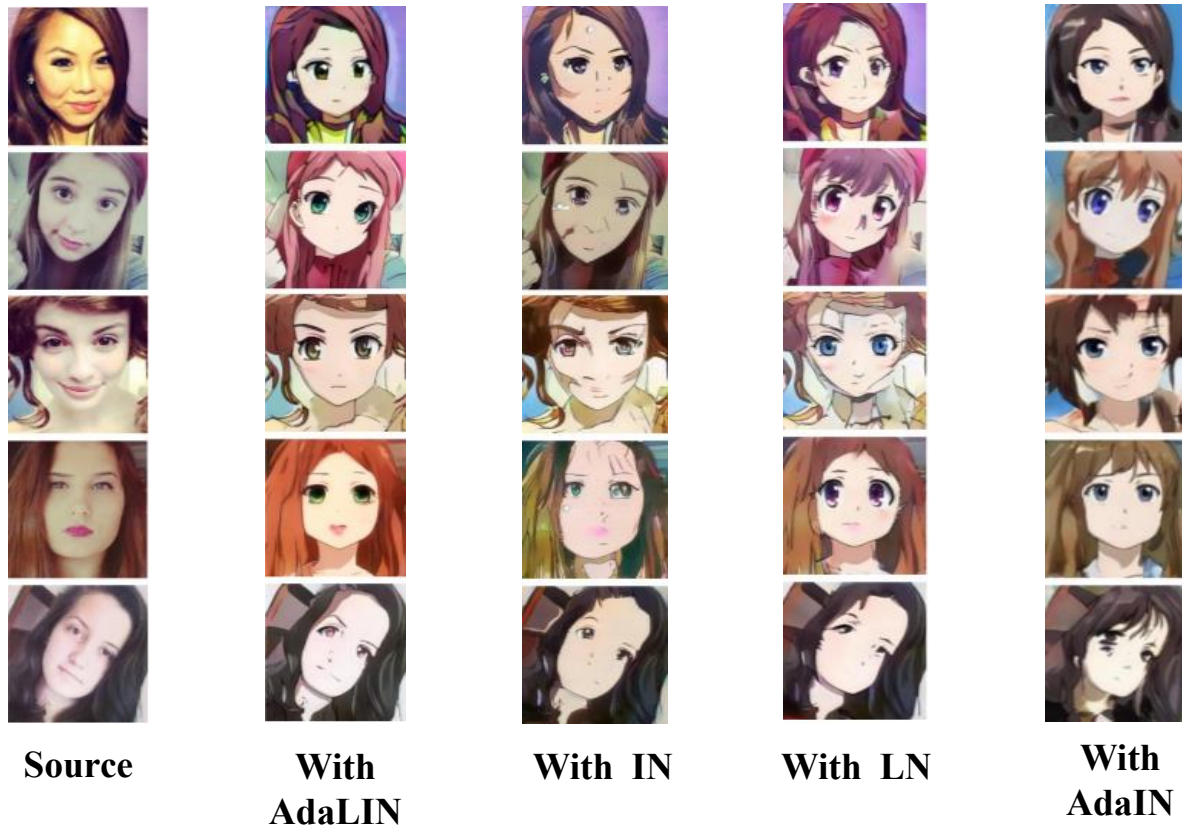
$$\rho \leftarrow clip_{[0,1]}(\rho - \tau \Delta \rho)$$

$G_{S \rightarrow t}$: Generate from source to target domain

ρ is a learnable parameter so that generator looks for the best normalization strategy

Associated Papers(Unsupervised)

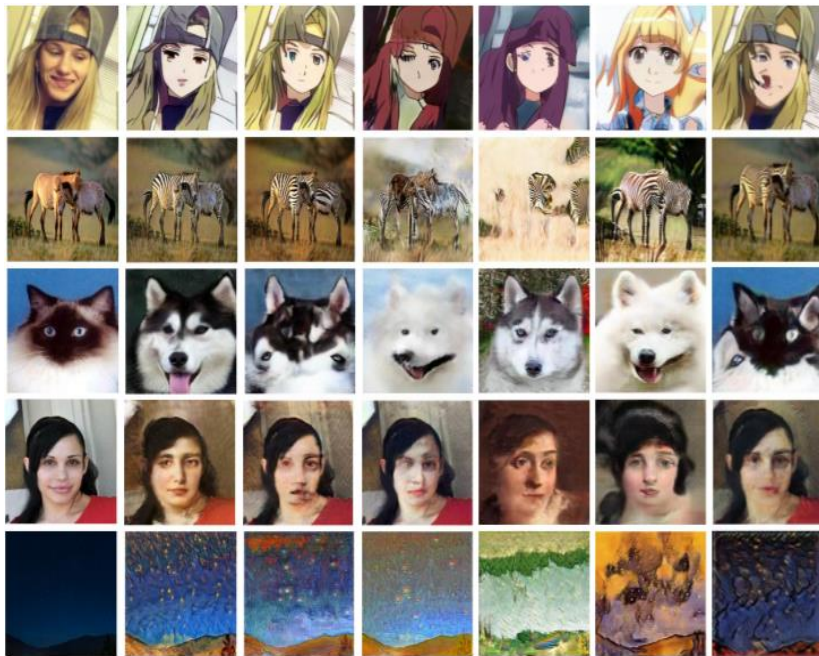
- U-GAT-IT
 - AdaLIN Experiment



Associated Papers(Unsupervised)

- U-GAT-IT

- Comparison



(a) Source (b) U-GAT-IT (c) CycleGAN (d) UNIT
 (e) MUNIT (f) DRIT (g) AGGAN

Model	selfie2anime	anime2selfie
U-GAT-IT	11.61 ± 0.57	11.52 ± 0.57
U-GAT-IT w/ IN	13.64 ± 0.76	13.58 ± 0.8
U-GAT-IT w/ LN	12.39 ± 0.61	13.17 ± 0.8
U-GAT-IT w/ AdaIN	12.29 ± 0.78	11.81 ± 0.77
U-GAT-IT w/ GN	12.76 ± 0.64	12.30 ± 0.77
U-GAT-IT w/o CAM	12.85 ± 0.82	14.06 ± 0.75
U-GAT-IT w/o G.CAM	12.33 ± 0.68	13.86 ± 0.75
U-GAT-IT w/o D.CAM	12.49 ± 0.74	13.33 ± 0.89

Model	selfie2anime	horse2zebra	cat2dog	photo2portrait	photo2vangogh
U-GAT-IT	73.15	73.56	58.22	30.59	48.96
CycleGAN	20.07	23.07	6.19	26.59	27.33
UNIT	1.48	0.85	18.63	32.11	11.93
MUNIT	3.41	1.04	14.48	8.22	2.07
DRIT	1.89	1.48	2.48	2.48	9.70

KID calculation(lower is better)

Associated Papers(Supervised)

- Controllable Person Image Synthesis with Attribute-Decomposed GAN
 - Generated Images



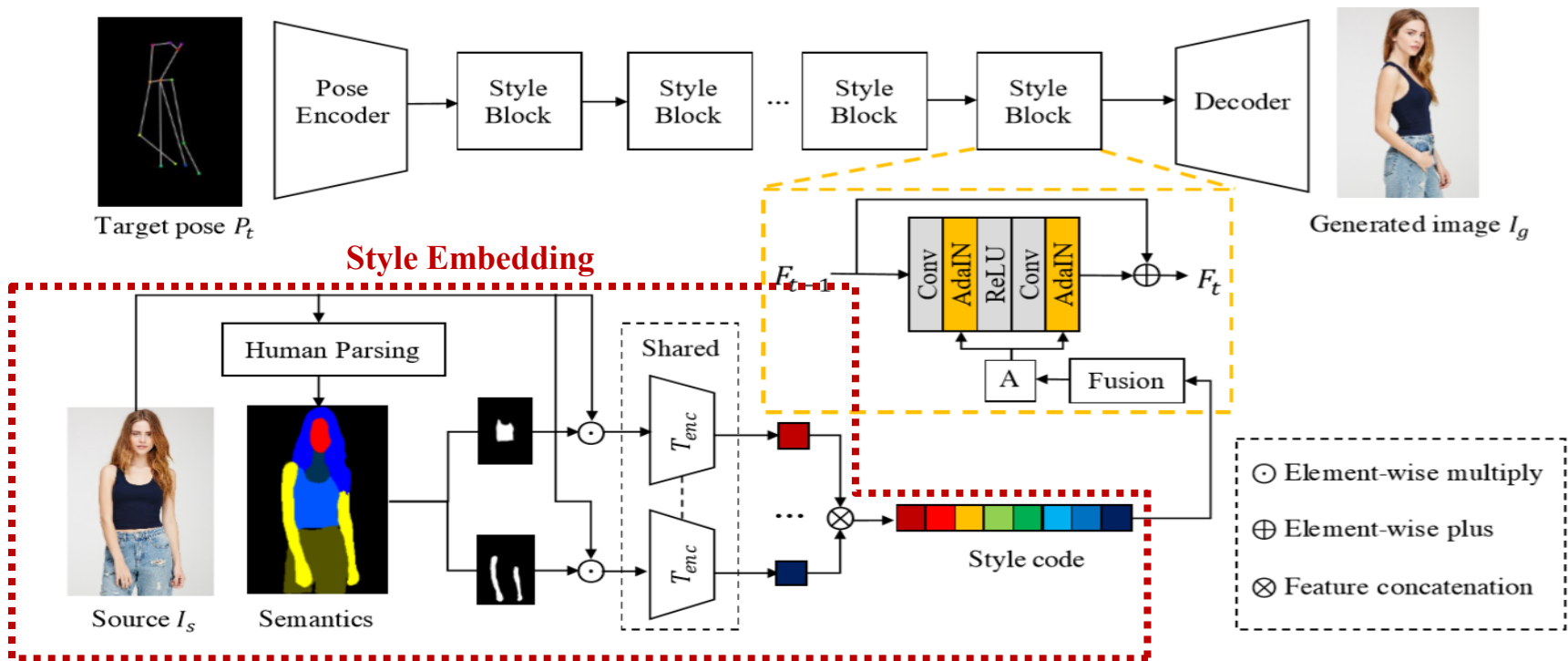
Generate person in the desired pose



Generate person in the desired fashion

Associated Papers(Supervised)

- Controllable Person Image Synthesis with Attribute-Decomposed GAN
 - Overall Network



Associated Papers(Supervised)

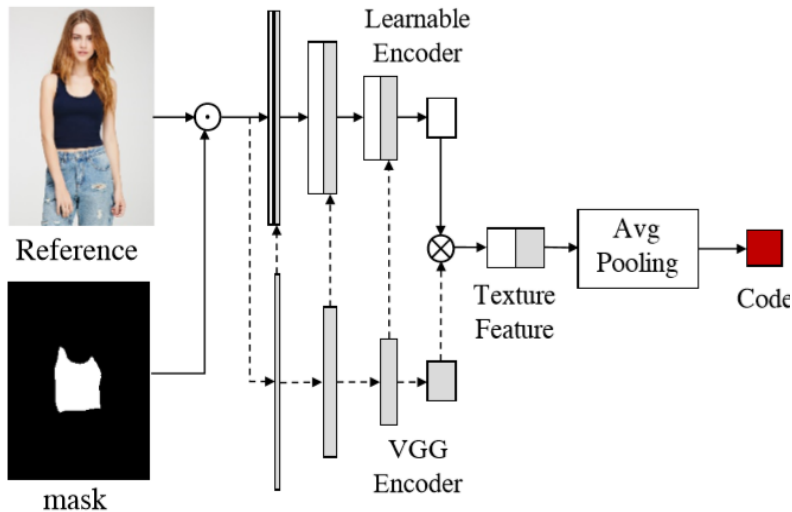
- Controllable Person Image Synthesis with Attribute-Decomposed GAN

- Decomposed component encoding

- Style Embedding

- Use pre-trained VGG19
- Encode styles using segmentation masks

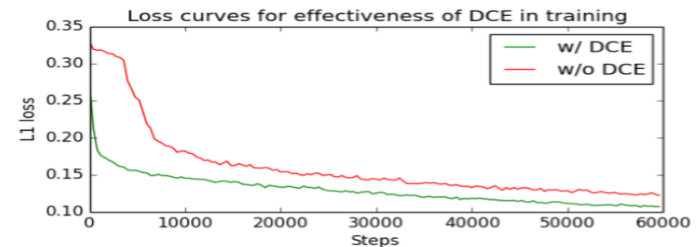
- Experimental Results



Details of the texture encoding

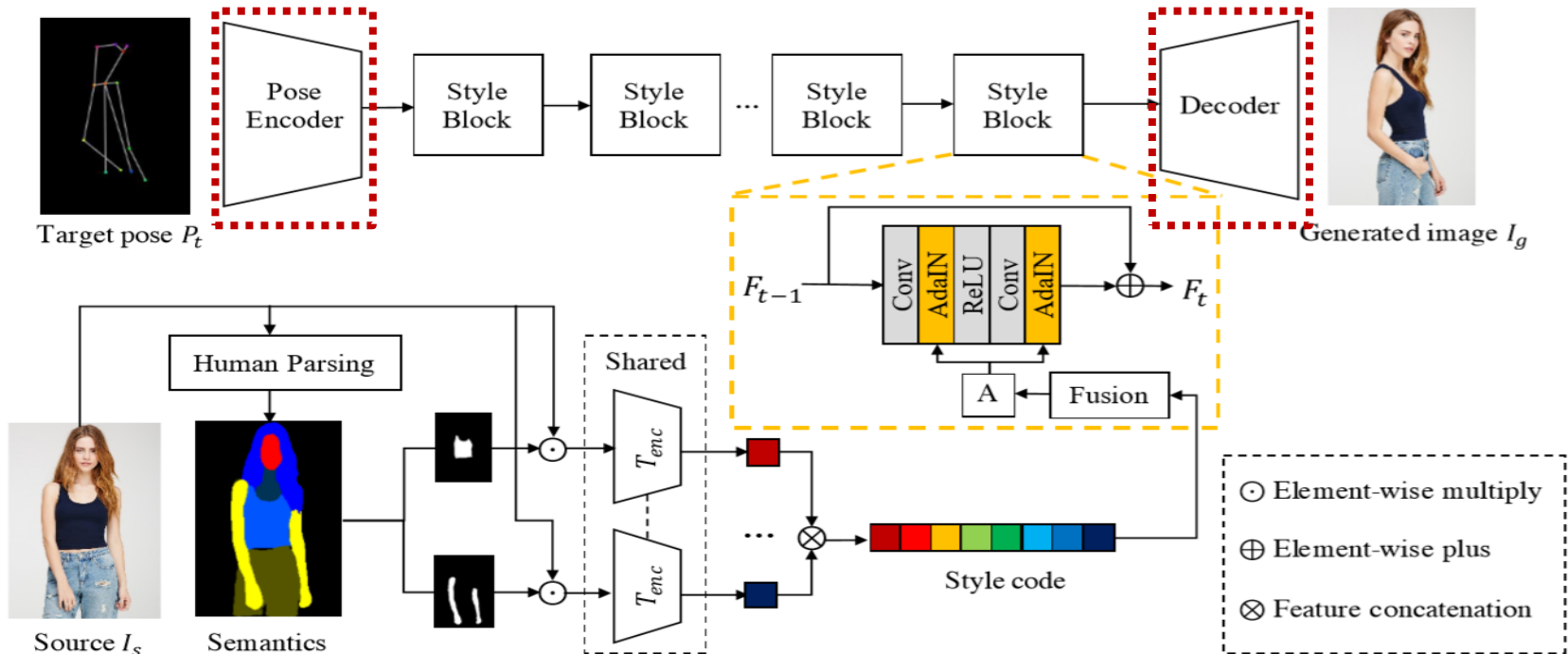


Source Target Pose W/O Embedding With Embedding



Associated Papers(Supervised)

- Controllable Person Image Synthesis with Attribute-Decomposed GAN
 - Encoder/Decoder – Same as CycleGAN



Associated Papers(Supervised)

- Controllable Person Image Synthesis with Attribute-Decomposed GAN

- Comparison



Model	IS \uparrow	SSIM \uparrow	DS \uparrow	CX-GS \downarrow	CX-GT \downarrow
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

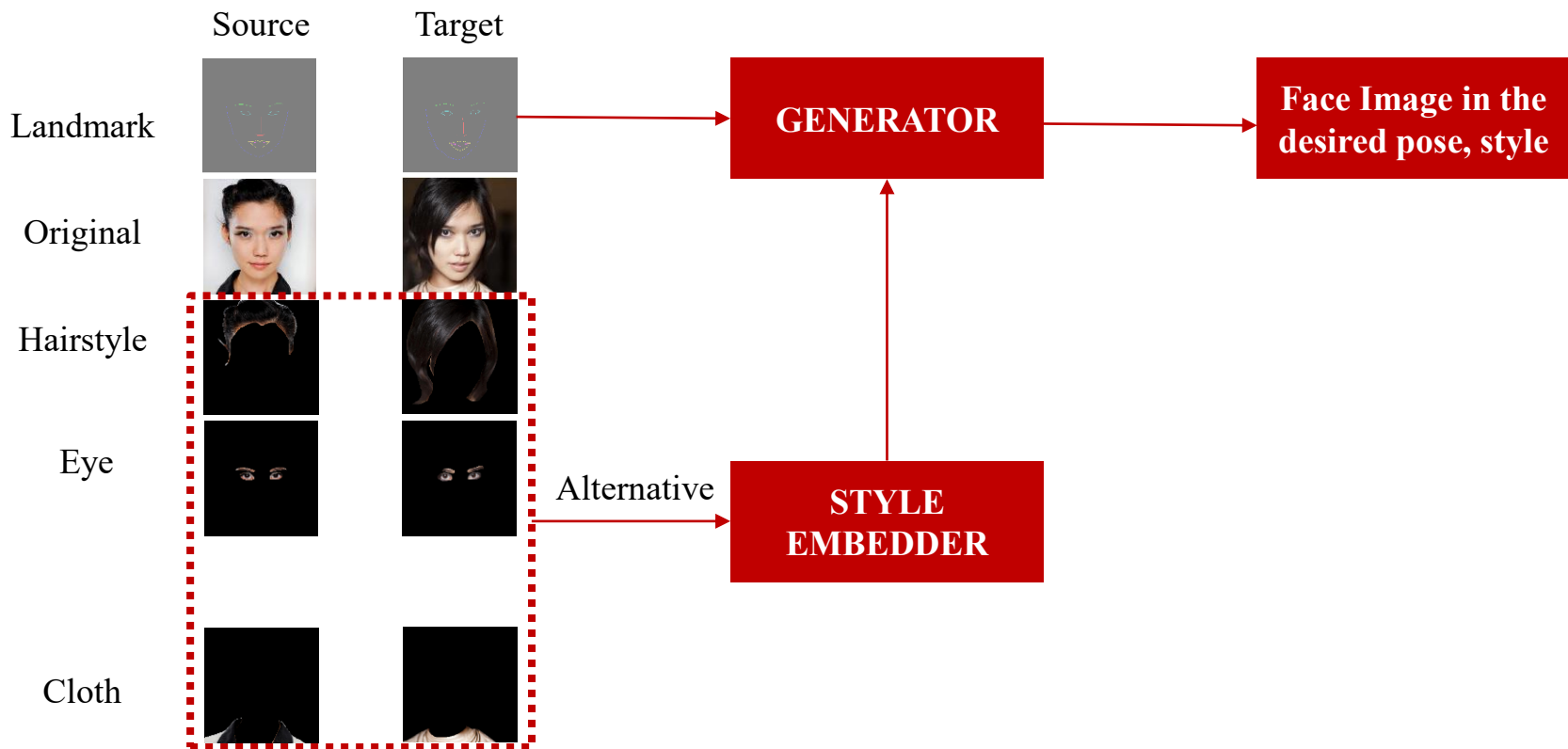
Quantitative comparison with SOTA methods on DeepFashion Dataset

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 of the user study

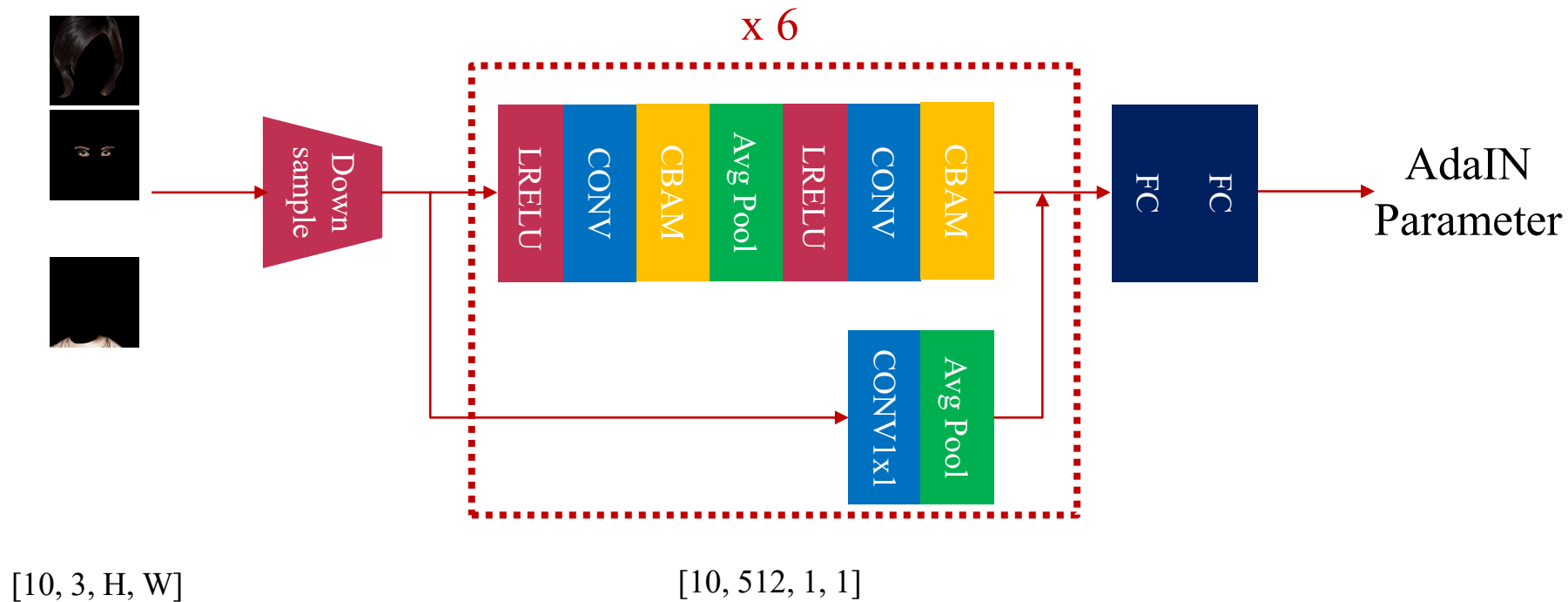
Experiment(Supervised)

- Controllable Face Image Synthesis with Attribute-Decomposed GAN



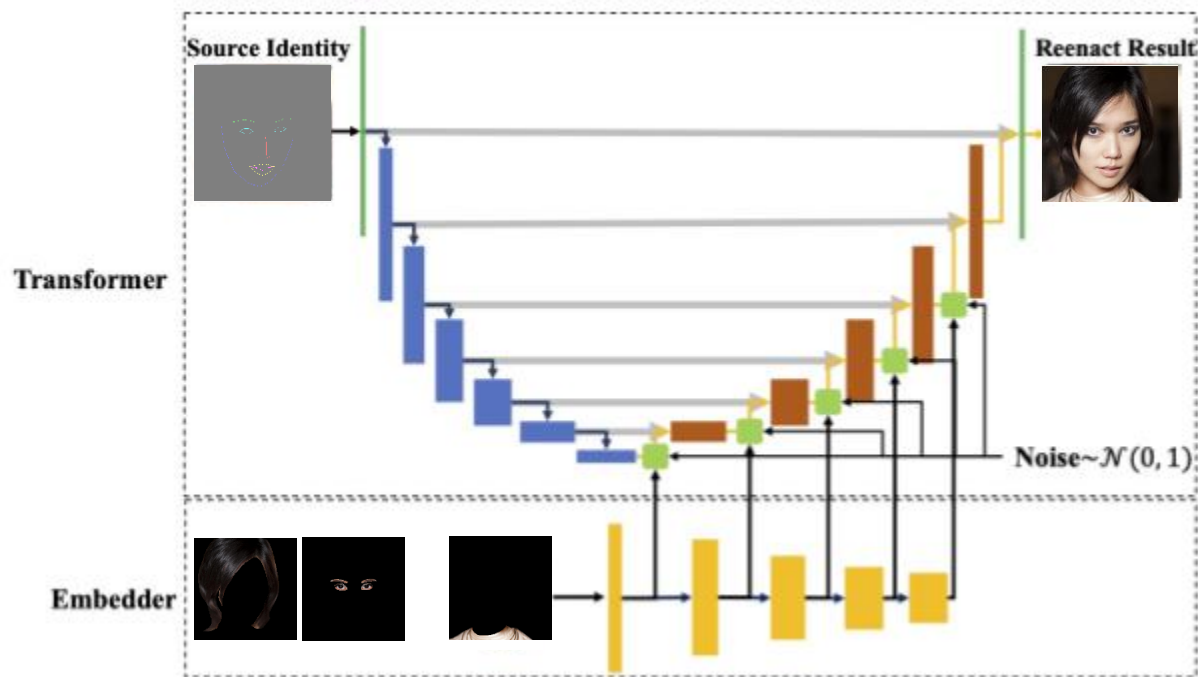
Research(Supervised)

- Controllable Face Image Synthesis with Attribute-Decomposed GAN
 - Embedder



Research(Supervised)

- Controllable Face Image Synthesis with Attribute-Decomposed GAN
 - Generator

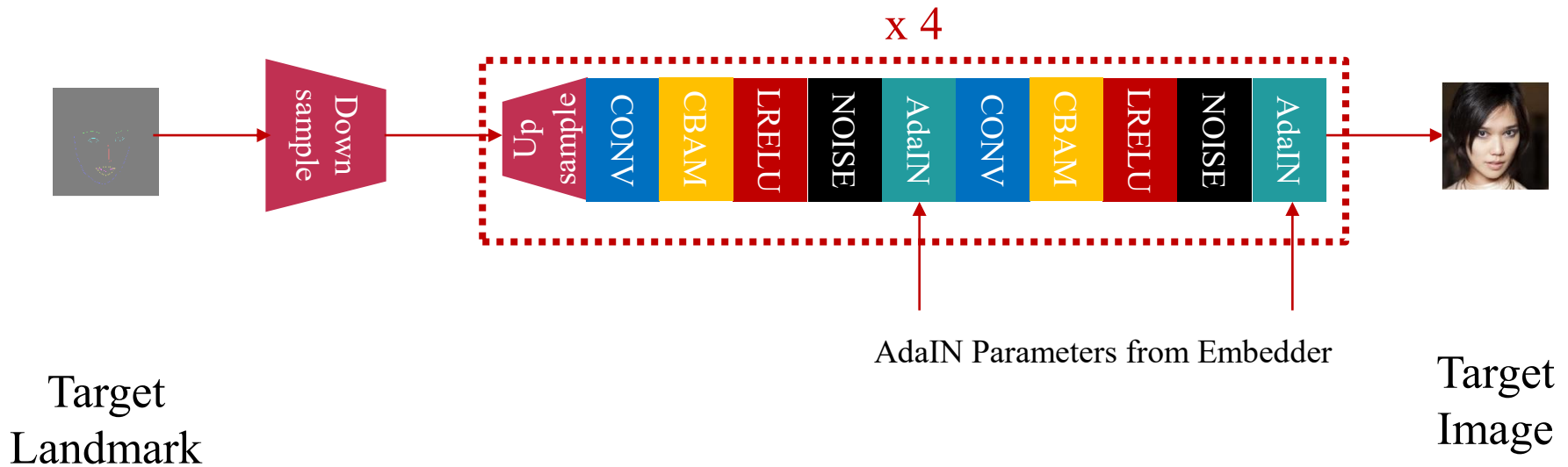


* AdaIN is used to inject embedded style

Research(Supervised)

- Controllable Face Image Synthesis with Attribute-Decomposed GAN

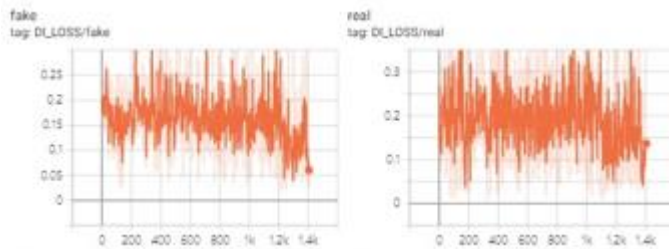
- Generator



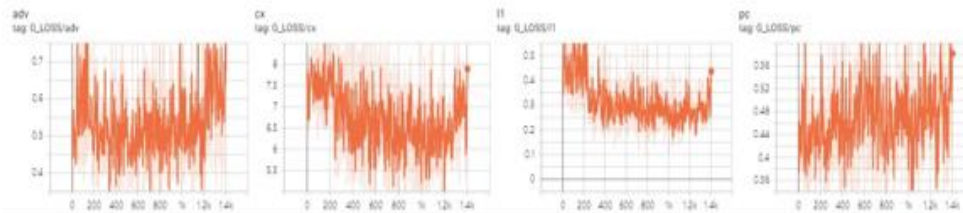
Research(Supervised)

- Controllable Face Image Synthesis with Attribute-Decomposed GAN

- Experiment



Discriminator Loss



Generator Loss



Source

Target

Generated

Q&A