StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation

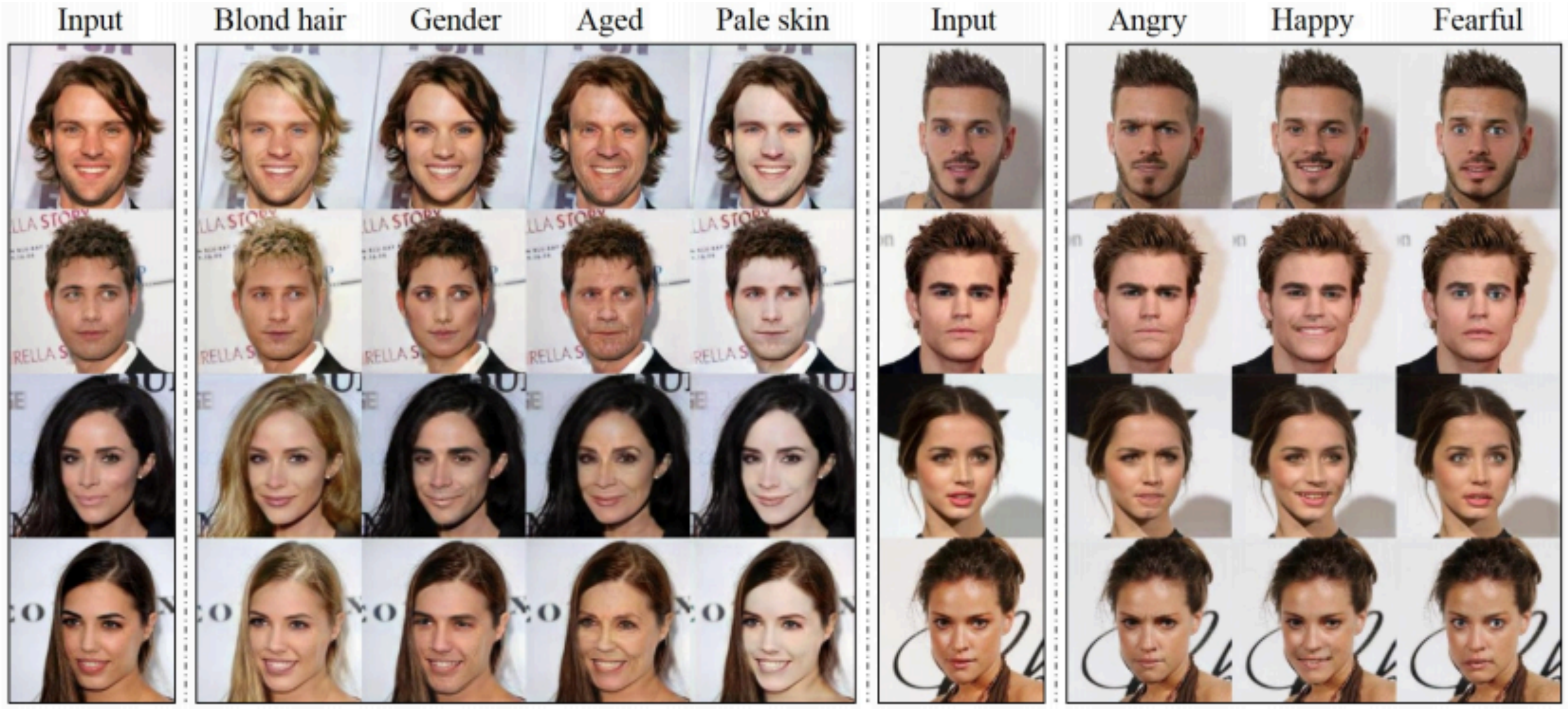
Yunjey Choi Minje Cho. Munyoung Kim Jung-Woo Ha Sunghun Kim Jaegul Choo

Jan, 28, 2021
구재원

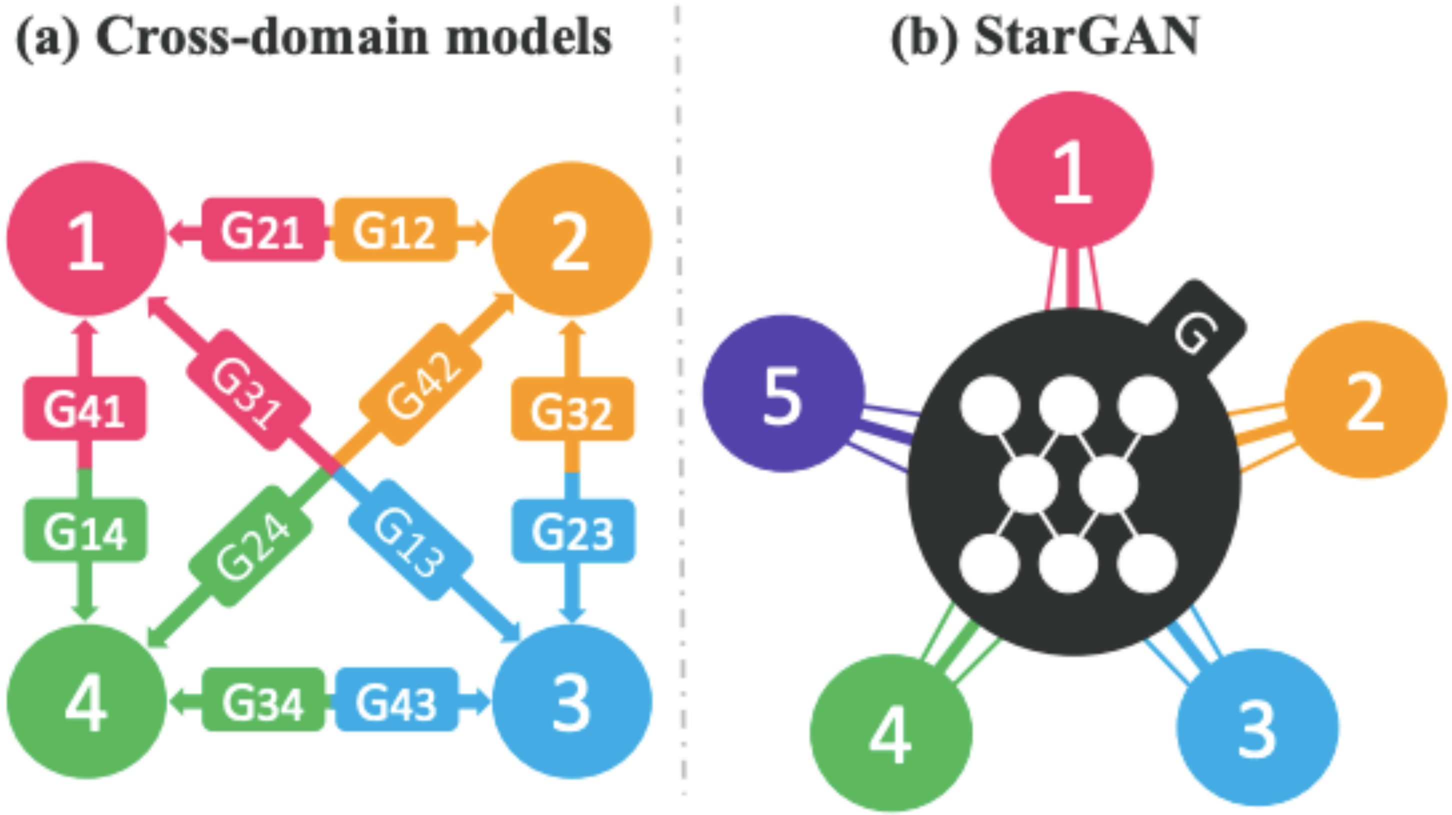
Introduction

>> Image to Image translation method: learns to translate images from one domain to the other

ex. Pix2pix, cycleGAN…



Introduction

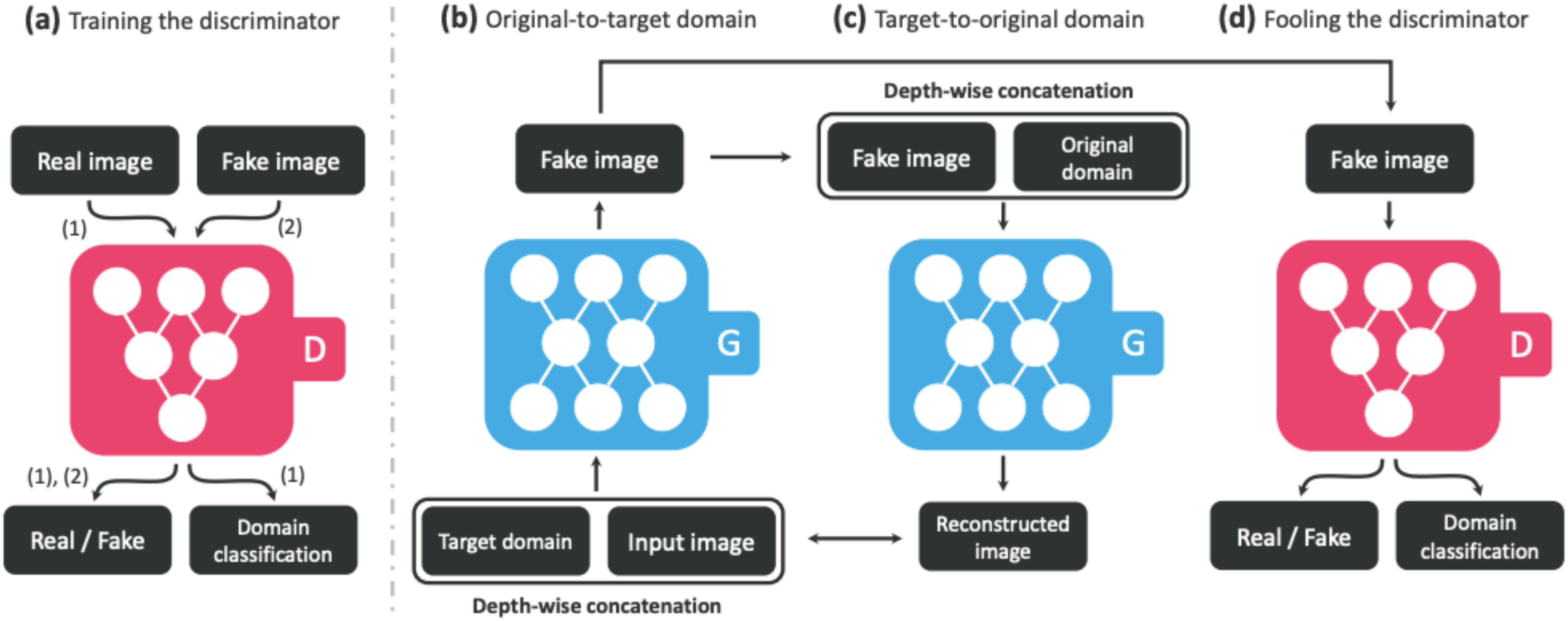


K(k-1) generator

single generator

>> multi-domain에서의 효율적인 image to image translation model

Overall Network



Network: Loss Functions

>> Adversarial Loss

$$\mathcal{L}_{adv} = \mathbb{E}_x\left[\log D_{src}(x)\right] + \mathbb{E}_{x,c}[\log(1 - D_{src}(G(x,c)))],$$

Target domain Label

$$\mathcal{L}_{adv} = \mathbb{E}_x[D_{src}(x)] - \mathbb{E}_{x,c}[D_{src}(G(x,c))] - \lambda_{gp}\,\mathbb{E}_{\hat{x}}[(||\nabla_{\hat{x}} D_{src}(\hat{x})||_2 - 1)^2],$$

>> Domain Classification Loss

$$\mathcal{L}^r_{cls} = \mathbb{E}_{x,c'}[-\log D_{cls}(c'|x)],$$

$$\mathcal{L}^f_{cls} = \mathbb{E}_{x,c}[-\log D_{cls}(c|G(x,c))].$$

>> Reconstruction Loss

$$\mathcal{L}_{rec} = \mathbb{E}_{x,c,c'}[||x - G(G(x,c),c')||_1],$$

>> Full Objective

$$\mathcal{L}_D = -\mathcal{L}_{adv} + \lambda_{cls}\,\mathcal{L}^r_{cls},$$

$$\mathcal{L}_G = \mathcal{L}_{adv} + \lambda_{cls}\,\mathcal{L}^f_{cls} + \lambda_{rec}\,\mathcal{L}_{rec},$$

Mask Vector

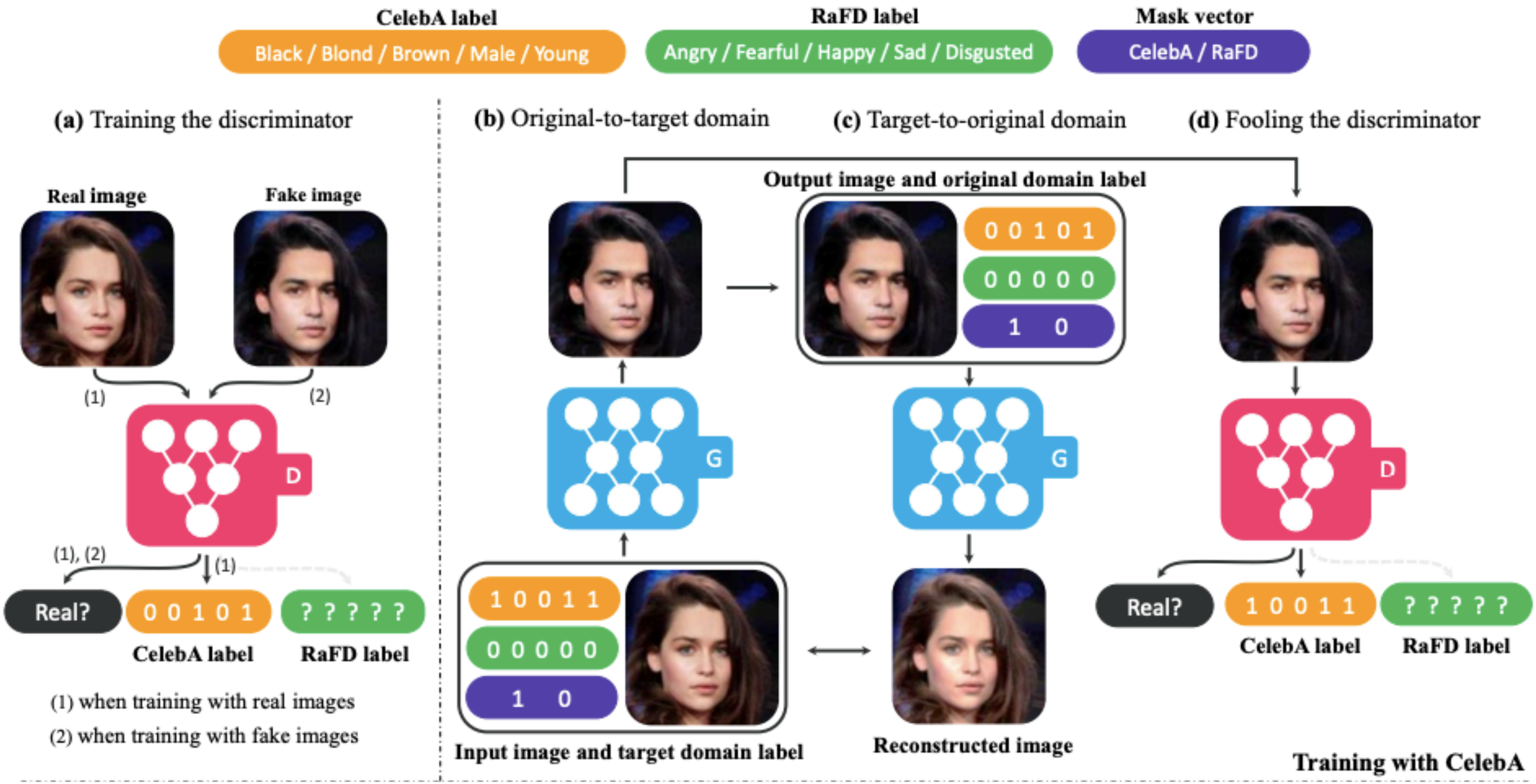
>> label information is only partially known to each dataset

>> introduce mask vector m that allows to ignore unspecified labels

$$\tilde{c} = [c_1, ..., c_n, m],$$

[.] : concatenation

Mask Vector



CelebA label
Black / Blond / Brown / Male / Young
RaFD label
Angry / Fearful / Happy / Sad / Disgusted
Mask vector
CelebA / RaFD
(a) Training the discriminator
Real image
Fake image
(1)
(2)
D
(1), (2)
(1)
Real?
0 0 1 0 1
? ? ? ? ?
CelebA label
RaFD label
(1) when training with real images
(2) when training with fake images
(b) Original-to-target domain
(c) Target-to-original domain
(d) Fooling the discriminator
Output image and original domain label
0 0 1 0 1
0 0 0 0 0
1 0
G
G
1 0 0 1 1
0 0 0 0 0
1 0
Input image and target domain label
Reconstructed image
D
Real?
1 0 0 1 1
? ? ? ? ?
CelebA label
RaFD label
Training with CelebA

Experiments

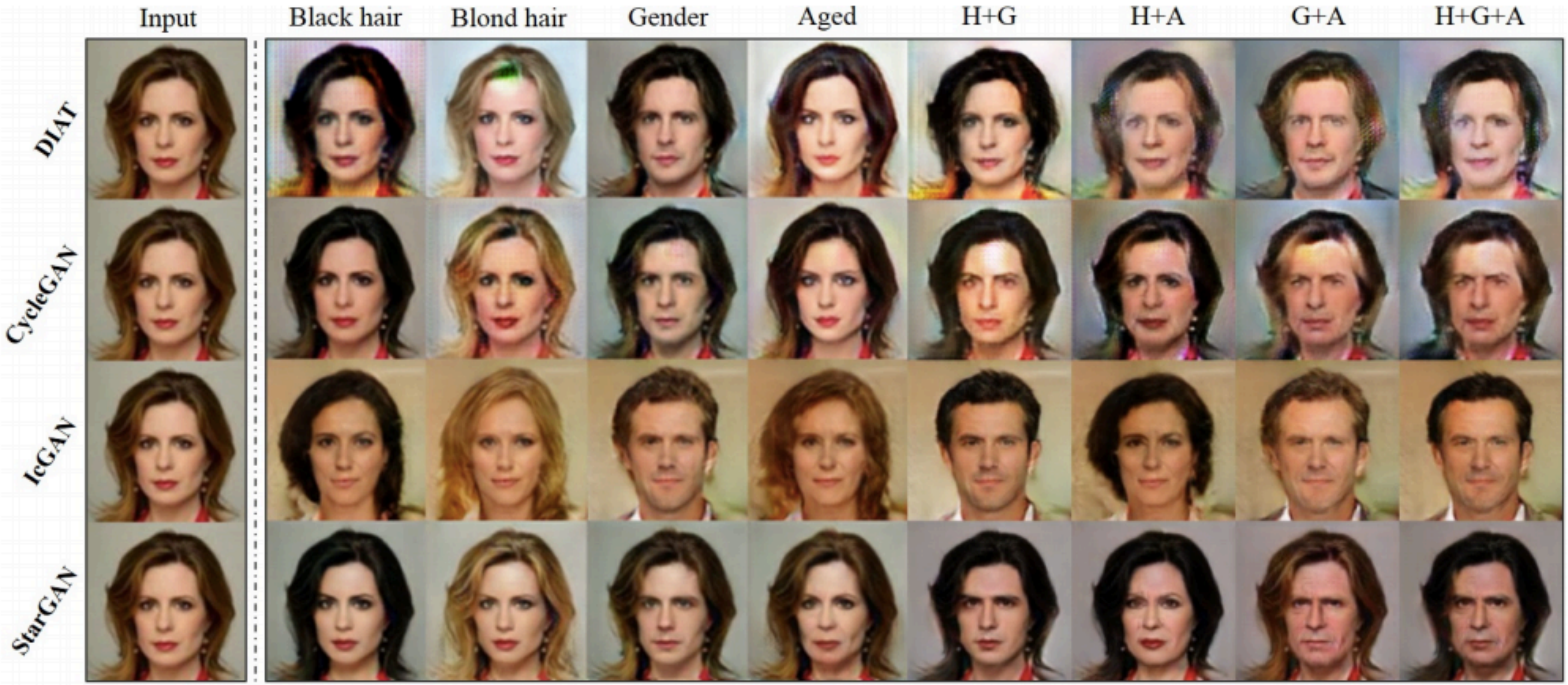


Figure 4. Facial attribute transfer results on the CelebA dataset. The first column shows the input image, next four columns show the single attribute transfer results, and rightmost columns show the multi-attribute transfer results. H: Hair color, G: Gender, A: Aged.

Experiments

>> Quantitative results on CelebA

Method	Hair color	Gender	Aged
DIAT	9.3%	31.4%	6.9%
CycleGAN	20.0%	16.6%	13.3%
IcGAN	4.5%	12.9%	9.2%
StarGAN	**66.2%**	**39.1%**	**70.6%**

Table 1. AMT perceptual evaluation for ranking different models on a single attribute transfer task. Each column sums to 100%.

Method	H+G	H+A	G+A	H+G+A
DIAT	20.4%	15.6%	18.7%	15.6%
CycleGAN	14.0%	12.0%	11.2%	11.9%
IcGAN	18.2%	10.9%	20.3%	20.3%
StarGAN	**47.4%**	**61.5%**	**49.8%**	**52.2%**

a

Table 2. AMT perceptual evaluation for ranking different models on a multi-attribute transfer task. H: Hair color; G: Gender; A: Aged.

Experiments

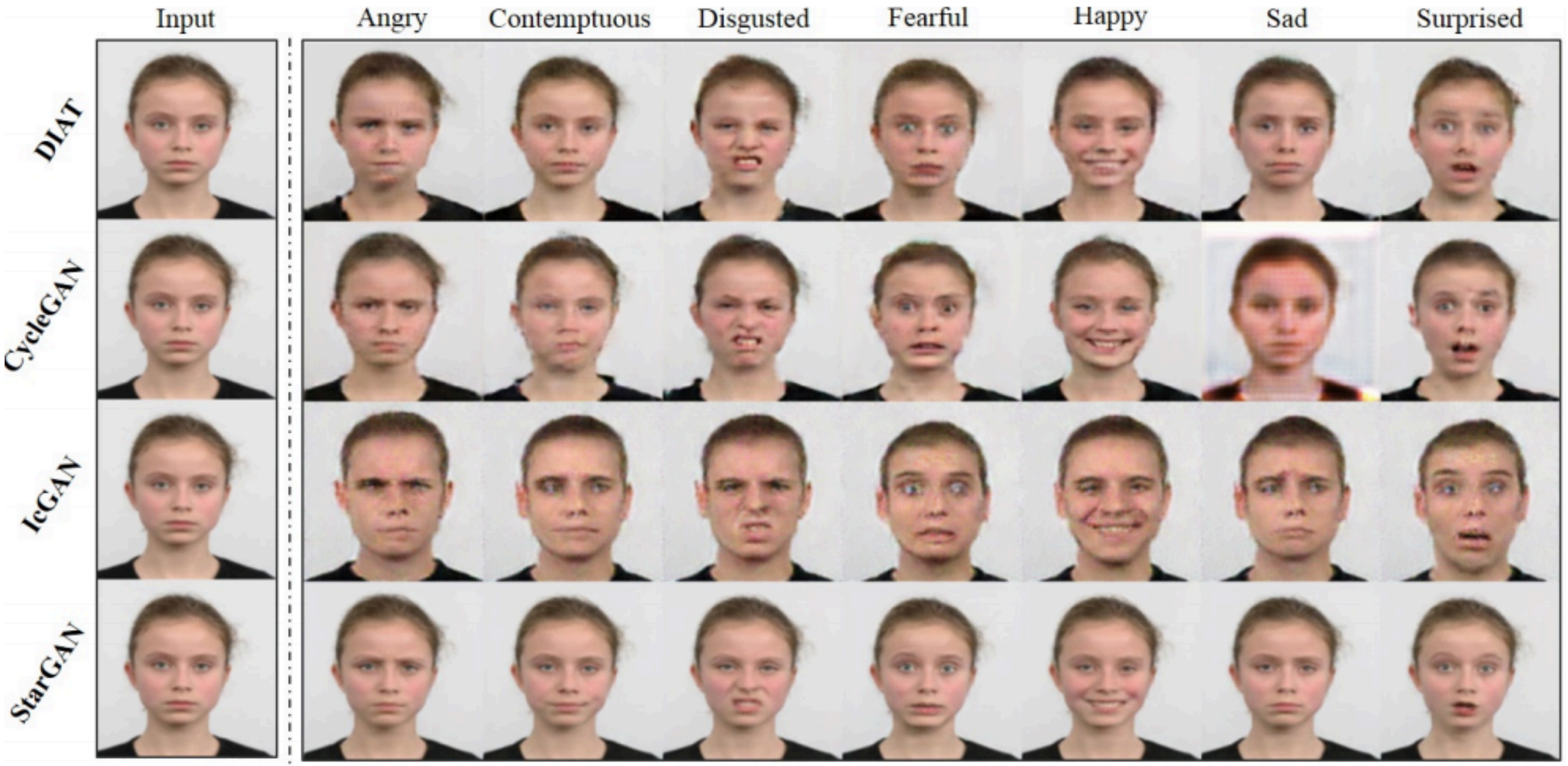


Figure 5. Facial expression synthesis results on the RaFD dataset.

Experiments

>> Quantitative results on RaFD

Method	Classification error	# of parameters
DIAT	4.10	52.6M $\times$ 7
CycleGAN	5.99	52.6M $\times$ 14
IcGAN	8.07	67.8M $\times$ 1
StarGAN	**2.12**	**53.2M** $\times$ **1**
Real images	0.45	-

Table 3. Classification errors [%] and the number of parameters on the RaFD dataset.

Experiments

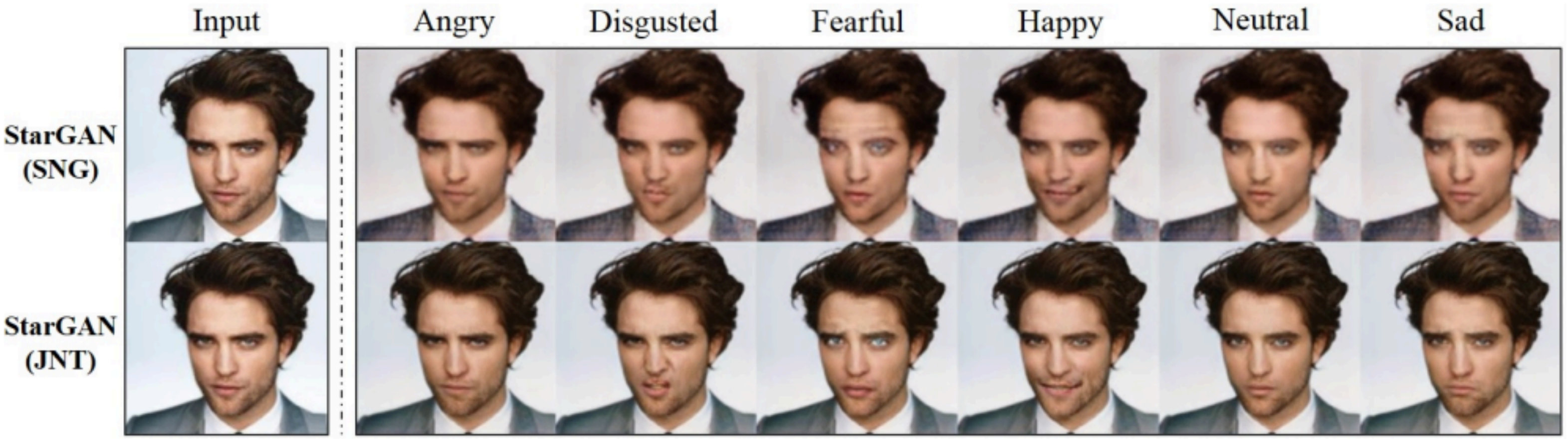


Figure 6. Facial expression synthesis results of StarGAN-SNG and StarGAN-JNT on CelebA dataset.

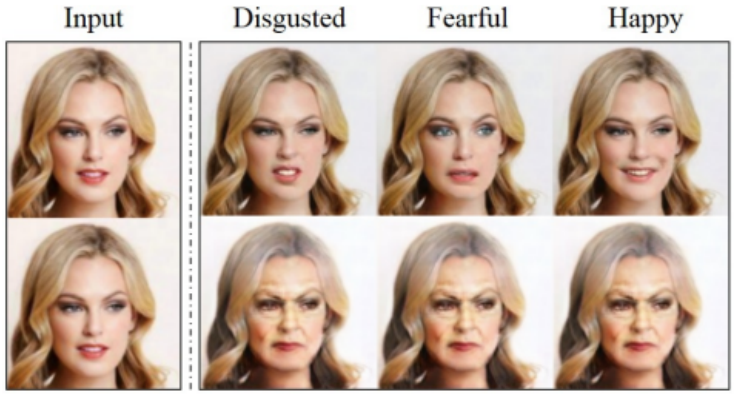


Figure 7. Learned role of the mask vector. All images are generated by StarGAN-JNT. The first row shows the result of applying the proper mask vector, and the last row shows the result of applying the wrong mask vector.