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

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

Clova AI Research, NAVER Corp

박태우

Introduction

- 기존의 GAN 모델의 경우 1가지 도메인으로 이미지 생성을 진행했음.

(CycleGAN, DiscoGAN, pix2pix, cGAN)

- 만약 k개의 도메인을 원한다면 k(k-1)개의 generator를 학습시켜야함.

- 본 논문에서는 1개의 generator를 이용해 다수의 도메인 이미지 생성할 것을 제안.

- 데이터 셋 2개를 동시에 학습시키는 방법을 제안하고 학습시킴.

- Attribute : meaningful feature inherent. (hair color, gender, age …)

  Domain : a set of images sharing the same attribute value.

Introduction

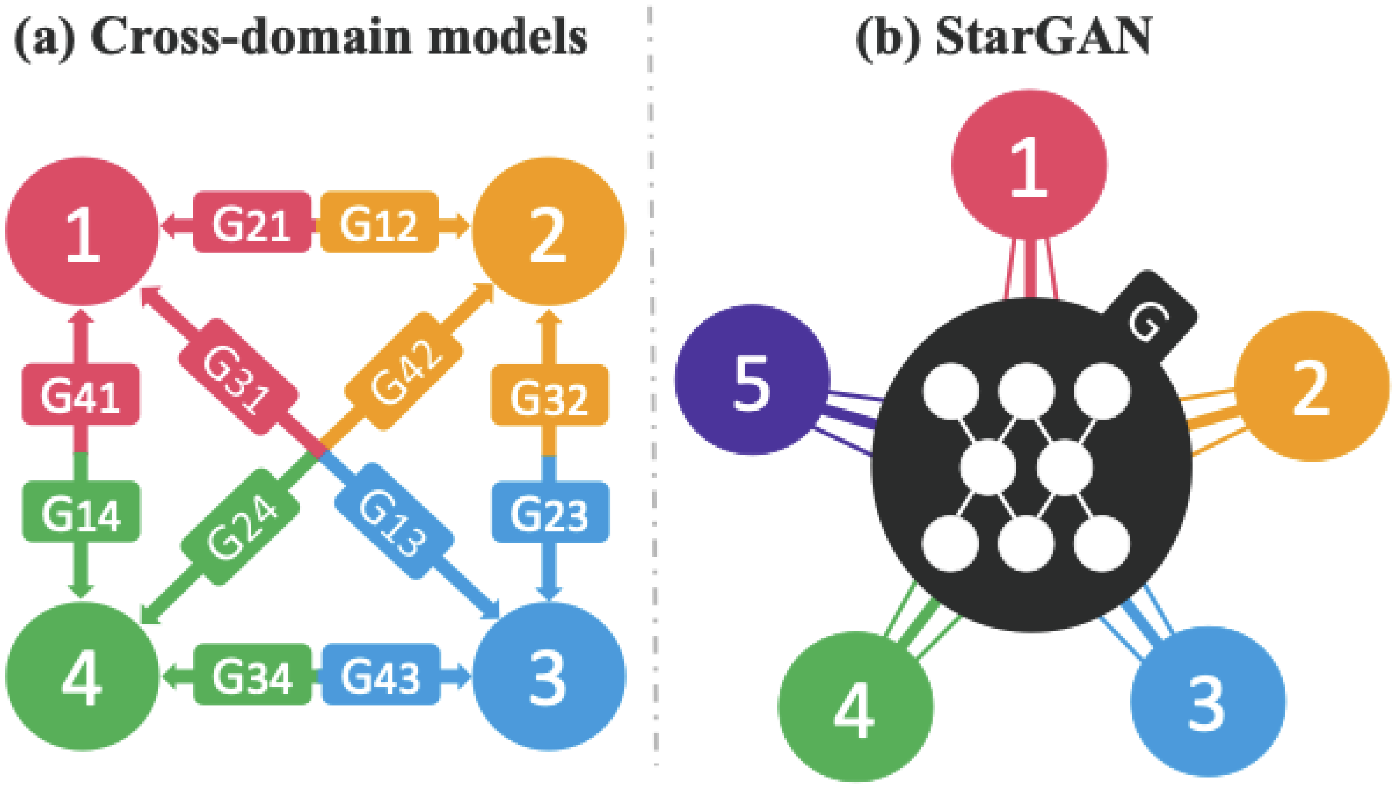
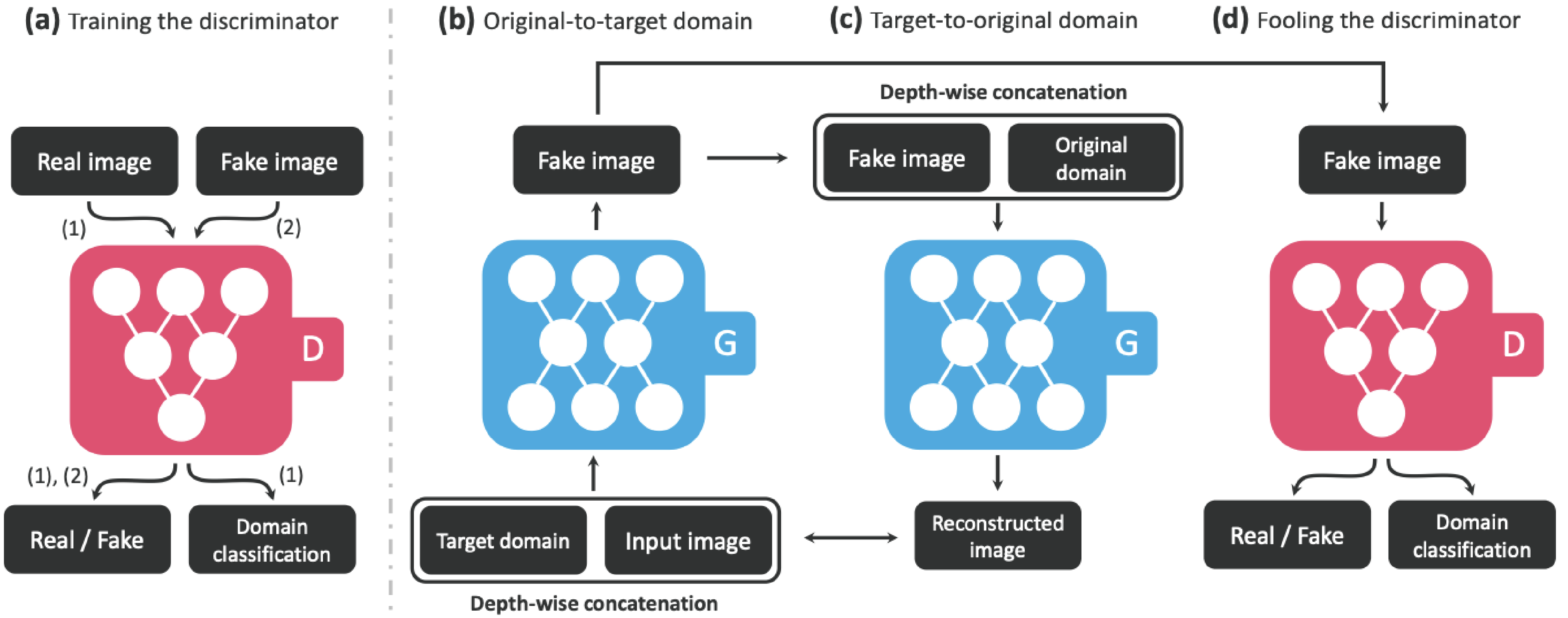


Figure 2. Comparison between cross-domain models and our proposed model, StarGAN. (a) To handle multiple domains, cross-domain models should be built for every pair of image domains. (b) StarGAN is capable of learning mappings among multiple domains using a single generator. The figure represents a star topology connecting multi-domains.

Architecture



- StarGAN은 기존 GAN 모델과 같이 Discriminator, Ganerator 2개의 모듈로 나눠짐.

- 기존 GAN 모델과 다른 점은 domain label이 추가, input이 원본 이미지인 점.

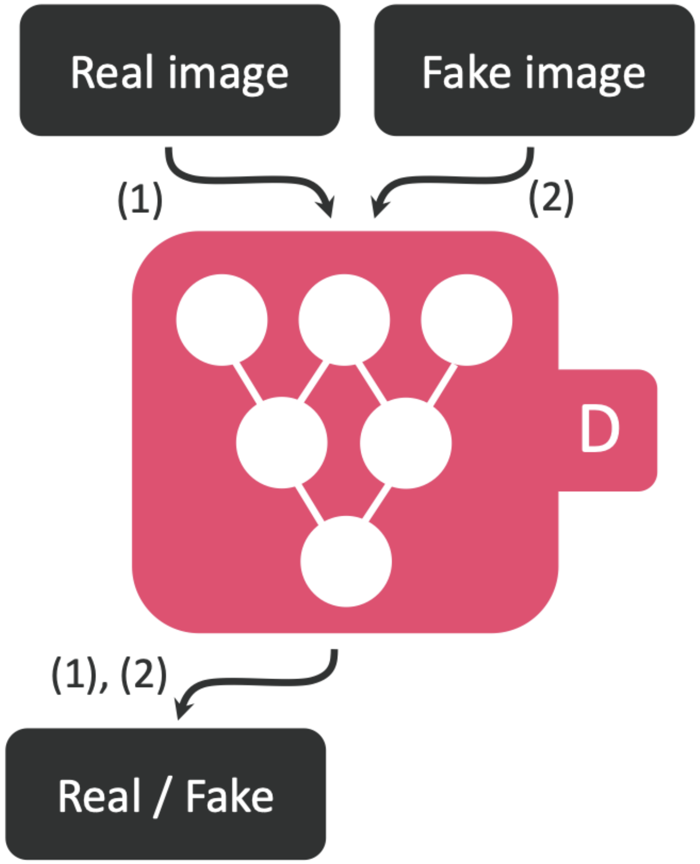
- Domain label c를 이용해 input image x를 output image y로 변환시킴. **G(x,c) -> y**

Training Loss

Adversarial Loss

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

Discriminator loss Generator loss



- Discriminator가 real image로 판단하면 D는 1, fake image로 판단하면 D는 0의 값을 가짐.

- 잘 판단할수록 두 항의 log은 모두 0에 가까워짐.

- 해당 loss는 label classification이 포함되지 않음.

Training Loss

Domain Classification Loss

1) Domain classification loss of real images

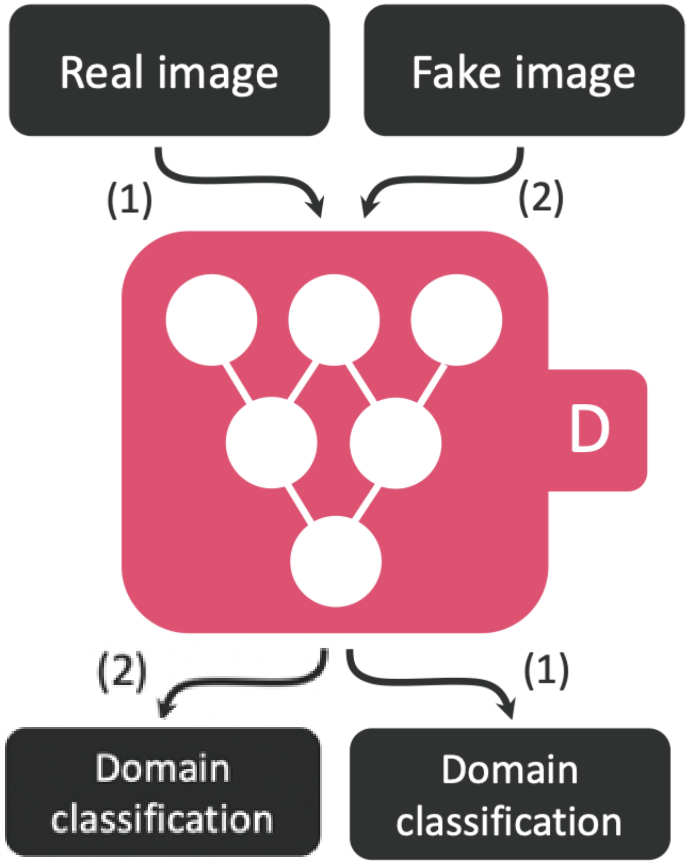
$$\mathcal{L}^{r}_{cls} = \mathbb{E}_{x,c'}[-\log D_{cls}(c'|x)]$$

Discriminator loss

2) Domain classification loss of fake images

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

Generator loss



- Real image의 경우 현재 domain의 확률분포 값.

- Fake image의 경우 target domain의 확률분포 값.

- 두 확률 모두 1에 가까울수록 loss 값은 0에 수렴.

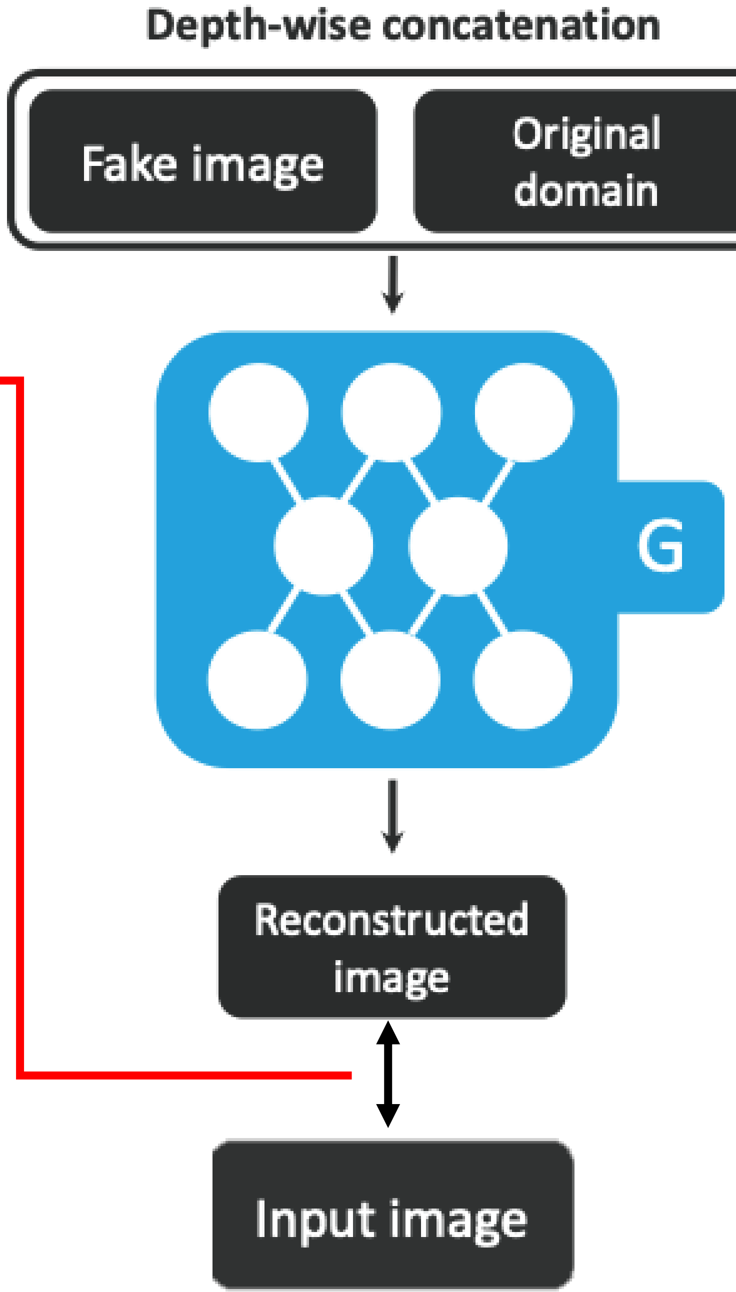
Training Loss

Reconstruction Loss

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

Generator loss

- 이전 loss들만 이용하면 target domain으로 변화할 때 input 이미지의 본래 형태를 잘 보존하지 못함.
- Fake image, original domain으로 이미지 생성
- 이렇게 만들어진 이미지를 input 이미지와 비교
- L1 norm을 마지막에 사용한다.



Training Loss

Full Objective

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

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

- λ_{cls} 와 λ_{rec} 는 hyper-parameter로 본 논문에서는 각각 1, 10으로 설정함.

- Domain 분류와 reconstruction loss들의 상대적인 중요도를 컨트롤함.

Training with Multiple Datasets

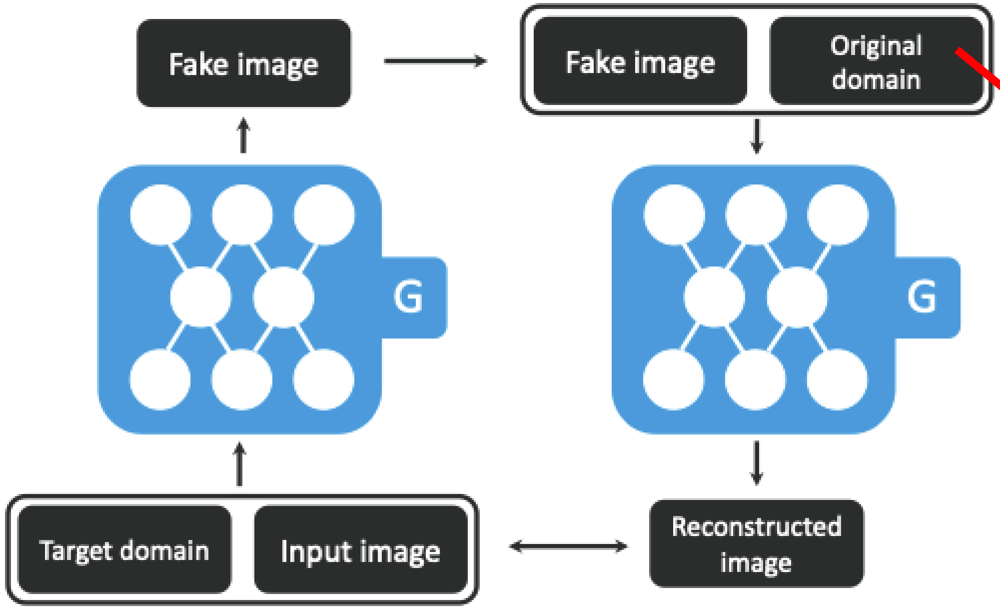
- Multiple datasets을 이용해 학습시키는 방법에 대해 언급함.

- 각각의 dataset이 서로 다른 부분적인 label을 가지는 것이 문제였음.

- CelebA의 경우 'hair color', 'gender'와 같은 domain label 존재.

- RaFD의 경우 'happy', 'angry'와 같은 domain label 존재

- Reconstructing 과정에서 input 이미지의 완벽한 label c' 정보가 필요하다는 점에서 문제가 됨.



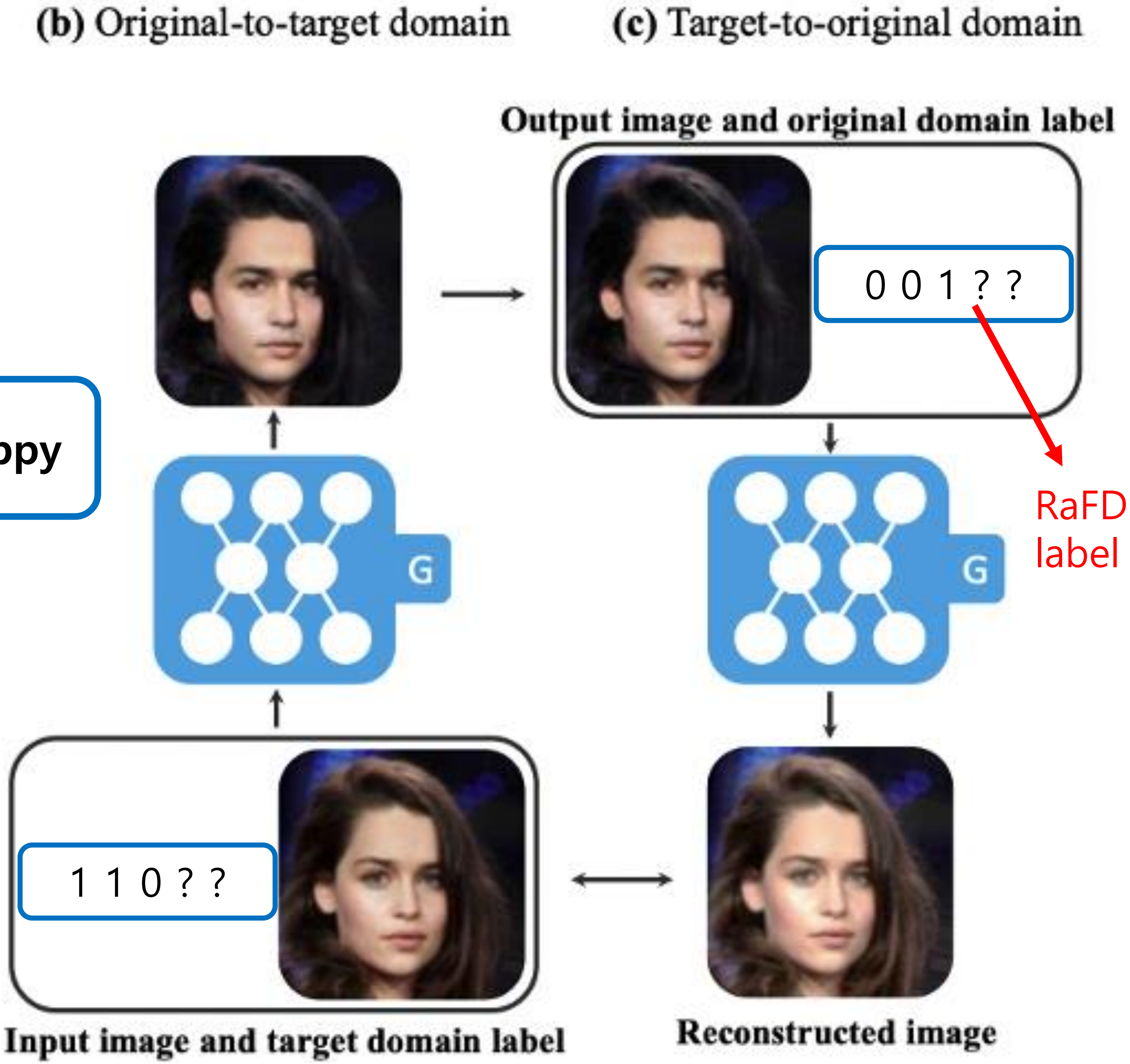
Input 이미지의 dataset은
domain label c'의 정보가 없음.

Training with Multiple Datasets

Training with CelebA

Hair color, Gender, Young, Sad, Happy

Domain Label



RaFD dataset의 label 정보가 없음

Training with Multiple Datasets

Mask Vector

- 앞의 문제를 해결하기 위해 mask vector m을 제안함.

- mask vector는 n-dimension one-hot vector로 표현되고 n은 dataset의 갯수.

- 통합된 label은 다음과 같이 표현된다.

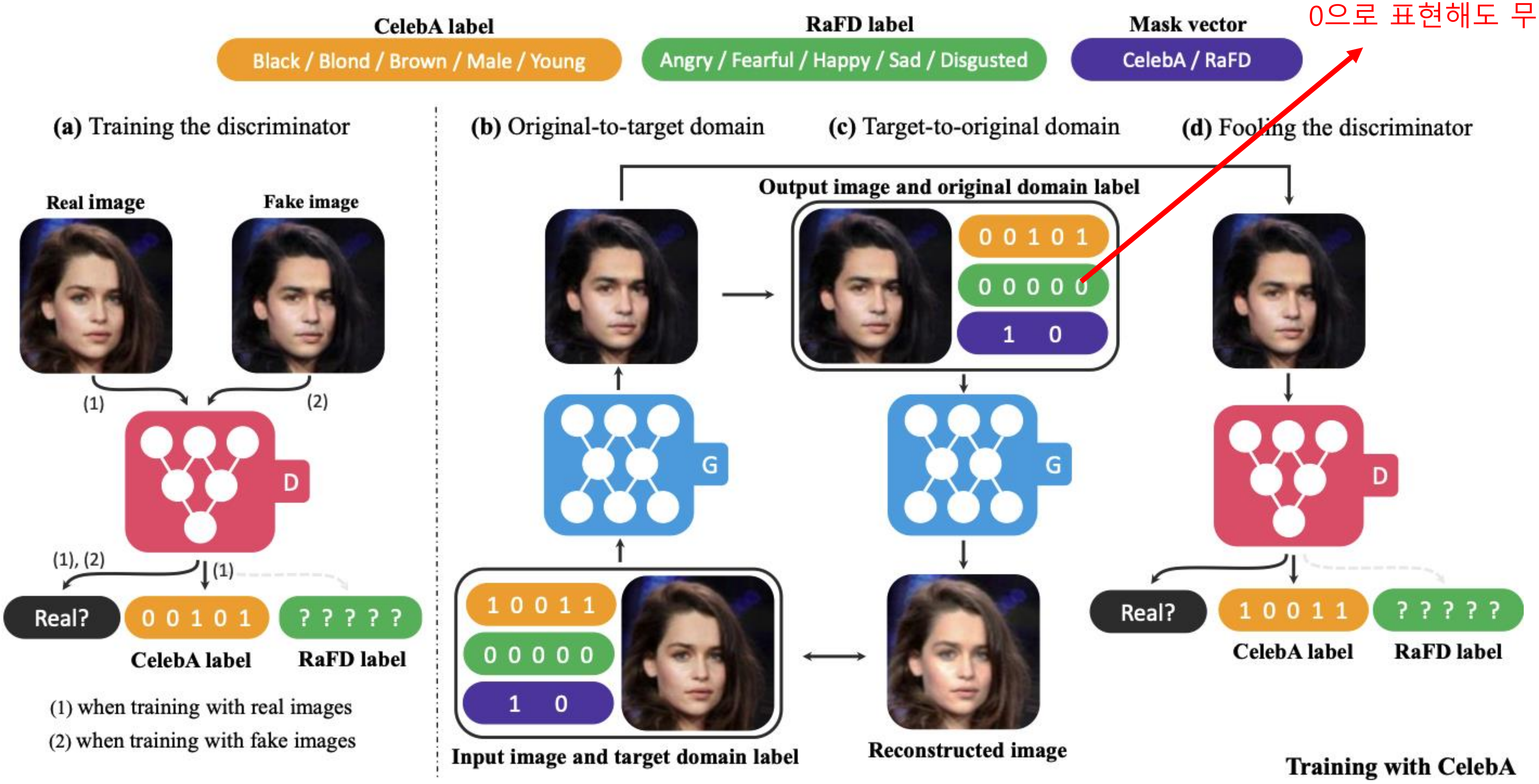
$$\tilde{c} = [c_1, ..., c_n, m]$$

- c_i 는 i번째 dataset이 가지는 attribute들을 표현한 binary vector.

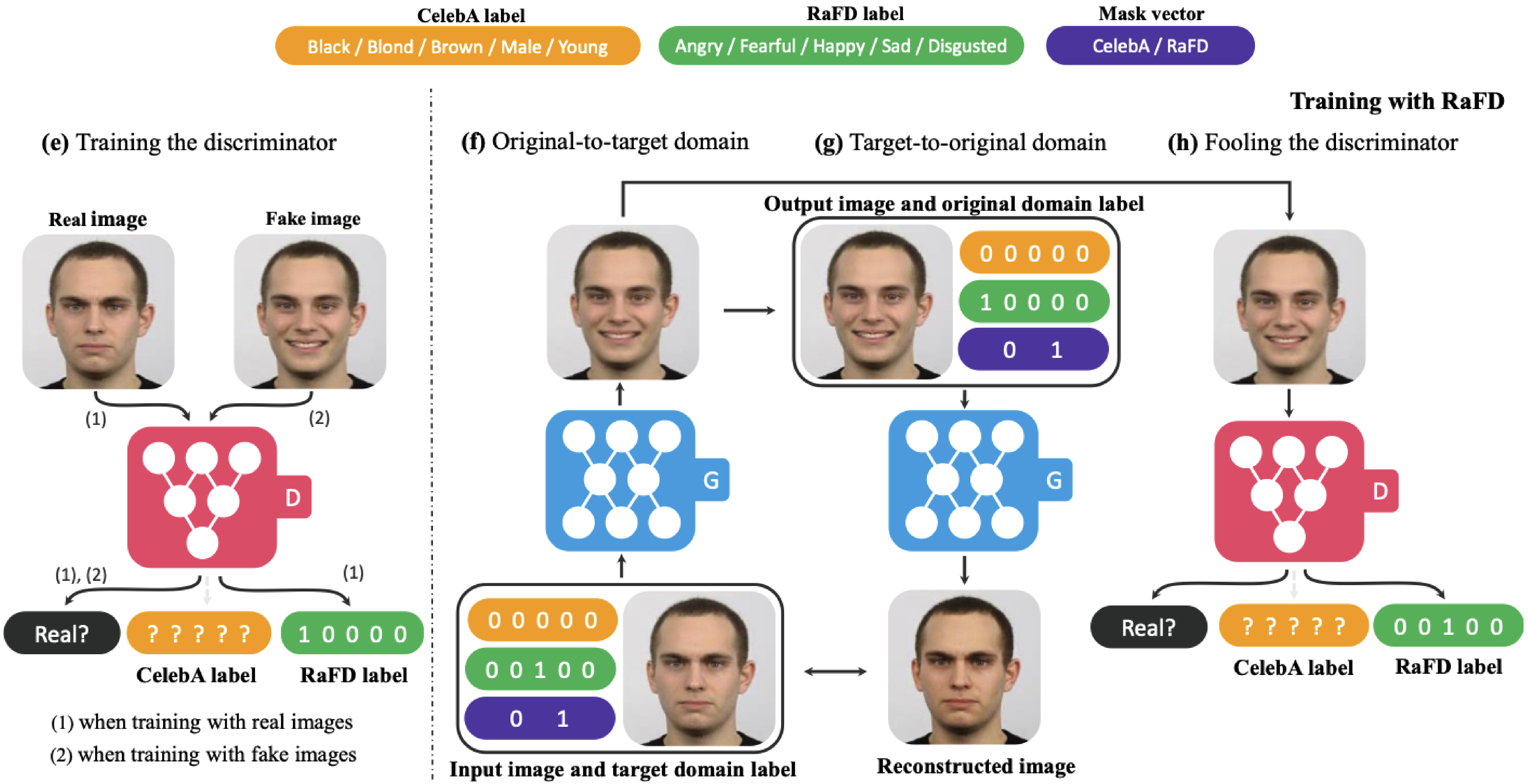
- 이를 통해 정보가 없는 label은 무시하게 되고 명시되어 있는 label에 집중한다.

Training with Multiple Datasets

Mask vector가 있기 때문에
0으로 표현해도 무시됨.



Training with Multiple Datasets



Result

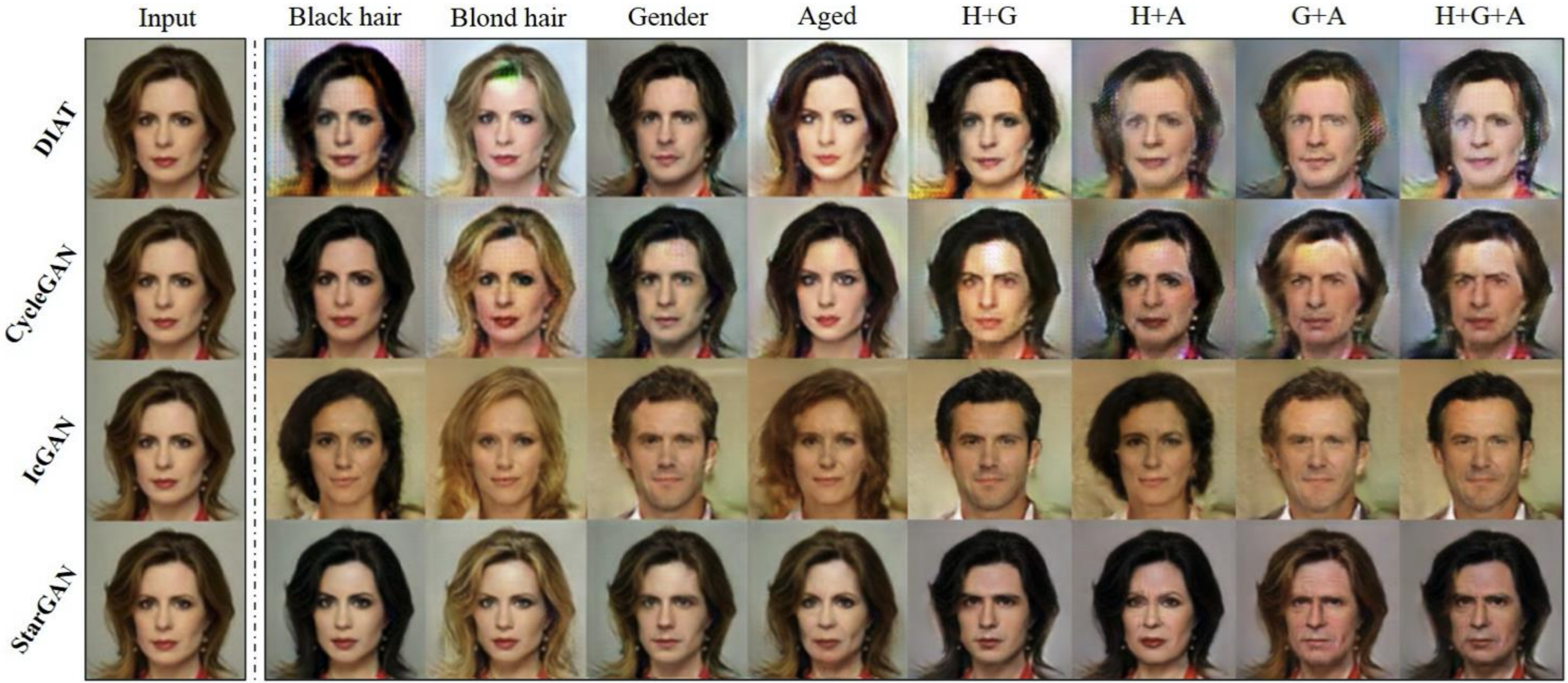


Figure 4. Facial attribute transfer results on the CelebA dataset. The first column shows the input image, next four columns show the single attribute transfer results, and rightmost columns show the multi-attribute transfer results. H: Hair color, G: Gender, A: Aged.

Result

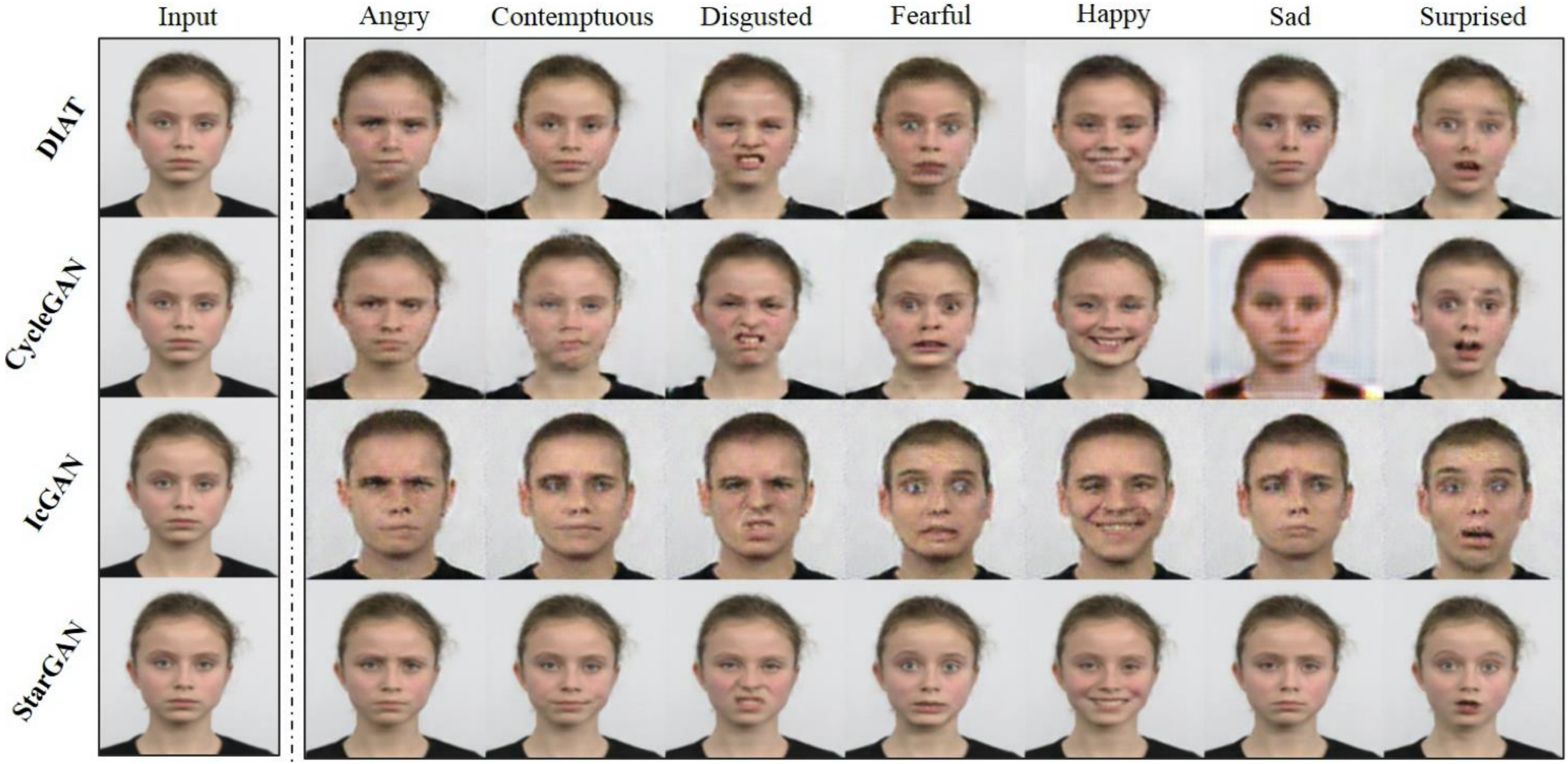


Figure 5. Facial expression synthesis results on the RaFD dataset.

Result

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.

Result

Figure 9. Single and multiple attribute transfer on CelebA (Input, Black hair, Blond hair, Brown hair, Gender, Aged, Hair color + Gender, Hair color + Aged, Gender + Aged, Hair color + Gender + Aged).

Reference

https://medium.com/curg/stargan-단일-모델로-다중-도메인-이미지-변환기-만들기-3b0fbdec121d

https://hichoe95.tistory.com/39