

꼼꼼한 딥러닝 논문 리뷰와 코드 실습

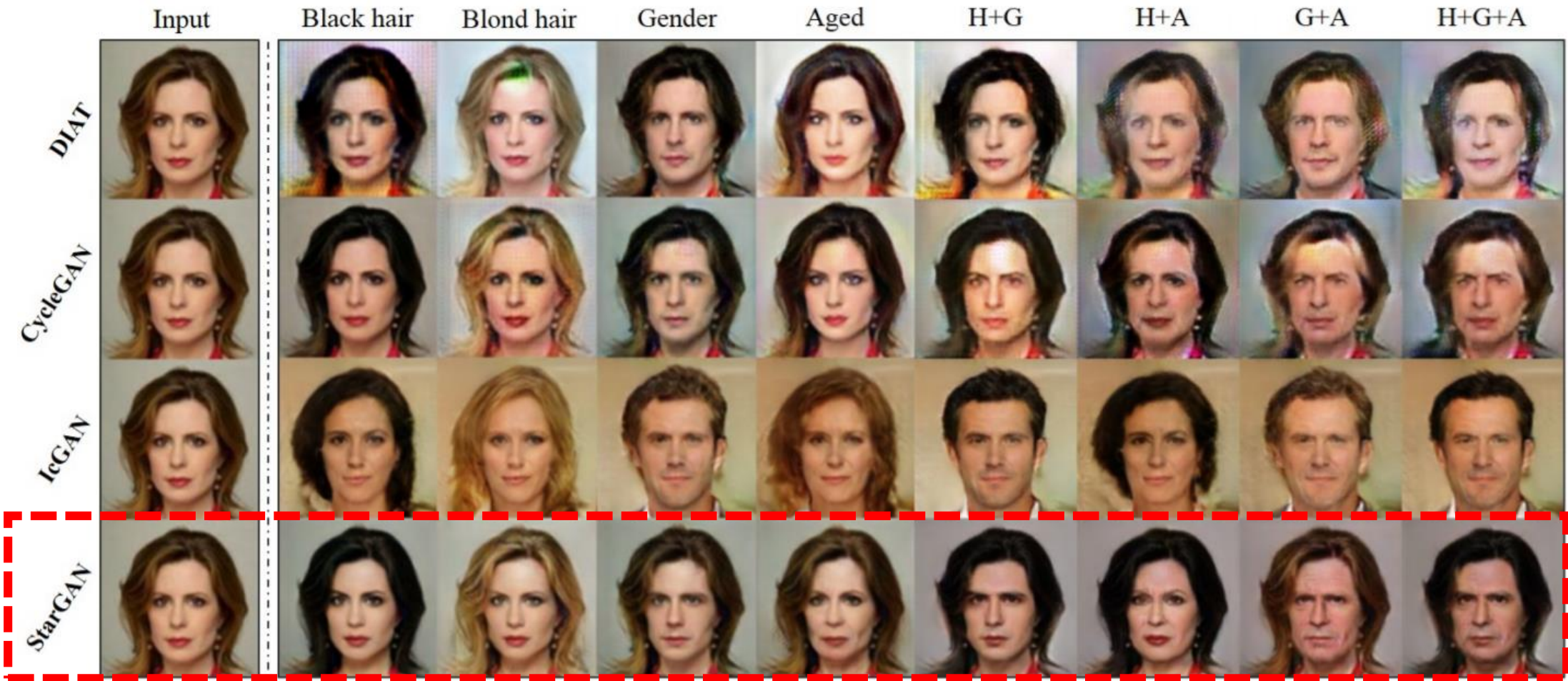
Deep Learning Paper Review and Code Practice

나동빈(dongbinna@postech.ac.kr)

Pohang University of Science and Technology

StarGAN (CVPR 2018 Oral)

- 다중 도메인에서의 효율적인 image-to-image translation 네트워크인 **StarGAN**을 제안합니다.

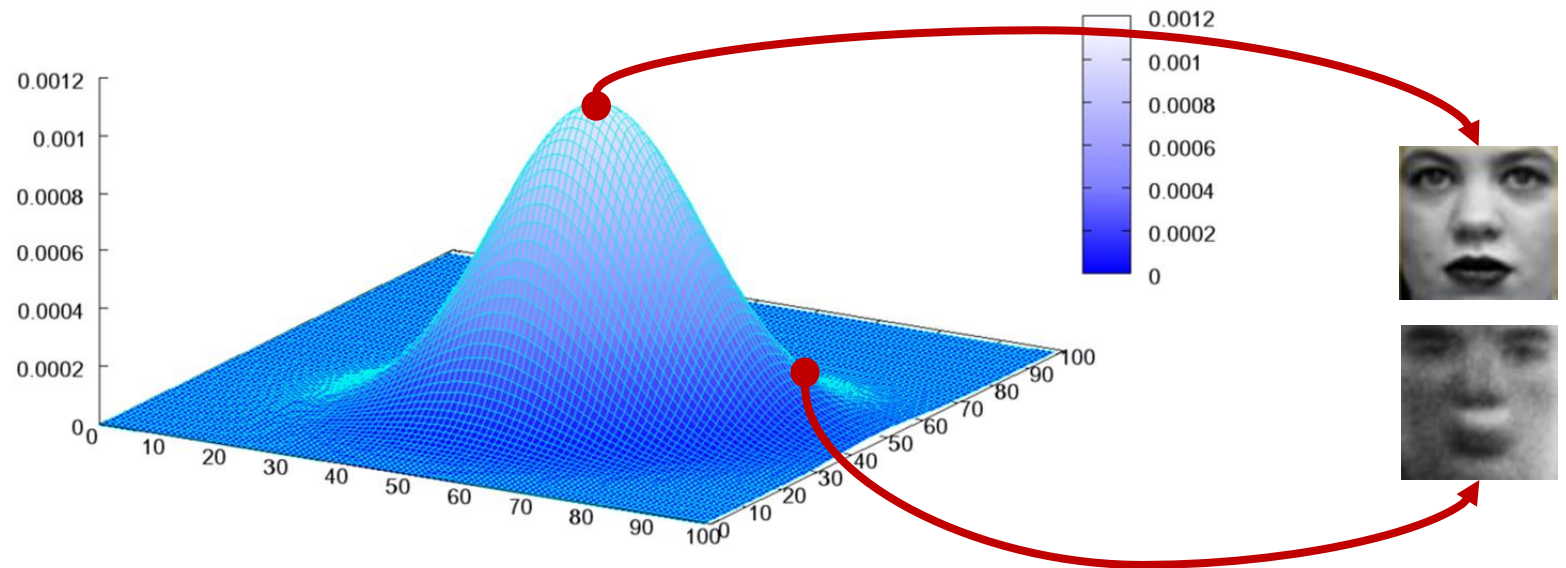
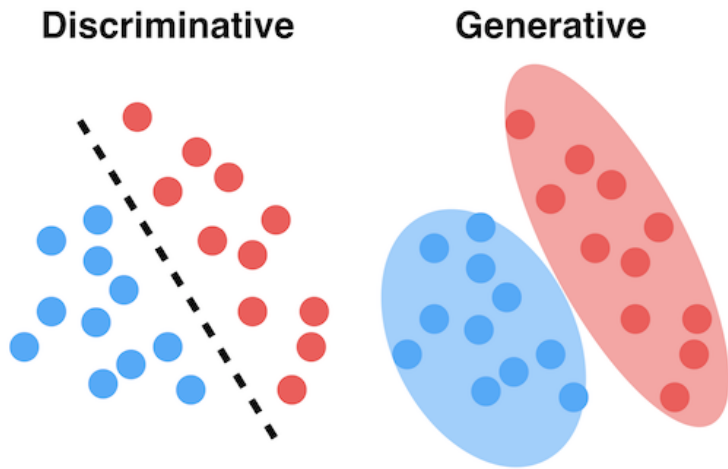


연구 배경: 생성 모델 (Generative Models)

- 생성 모델은 실존하지 않지만 있을 법한 이미지를 생성할 수 있는 모델을 의미합니다.

Generative Model (produce) → An image that does not exist but is likely to exist

- A statistical model of the joint probability distribution
- An architecture to generate new data instances



연구 배경: Generative Adversarial Networks (GAN)

- 생성자(generator)와 판별자(discriminator) 두 개의 네트워크를 활용한 생성 모델입니다.
- 다음의 목적 함수(objective function)를 통해 생성자는 이미지 분포를 학습할 수 있습니다.

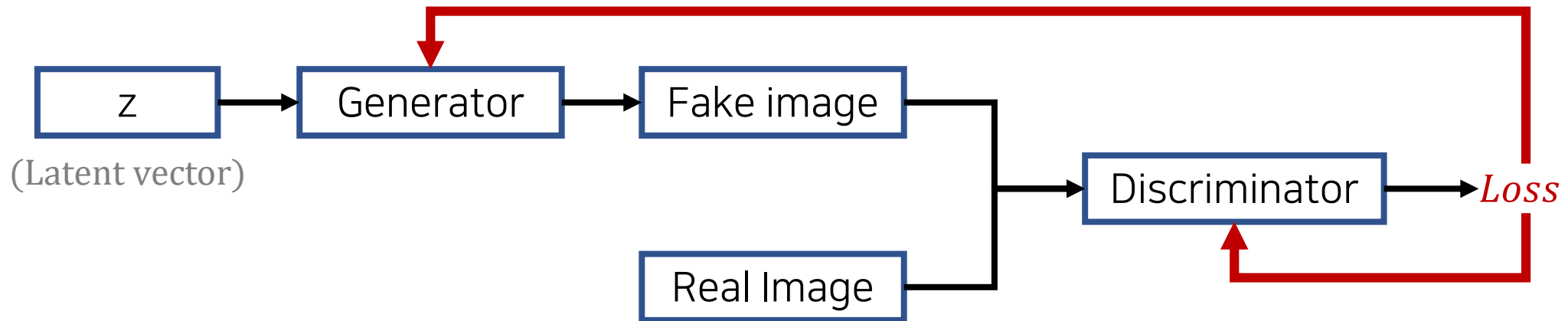
$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log \{1 - D(G(z))\}]$$

Generator

$G(z)$: new data instance

Discriminator

$D(x)$ = Probability: a sample came from the real distribution (Real: 1, Fake: 0)

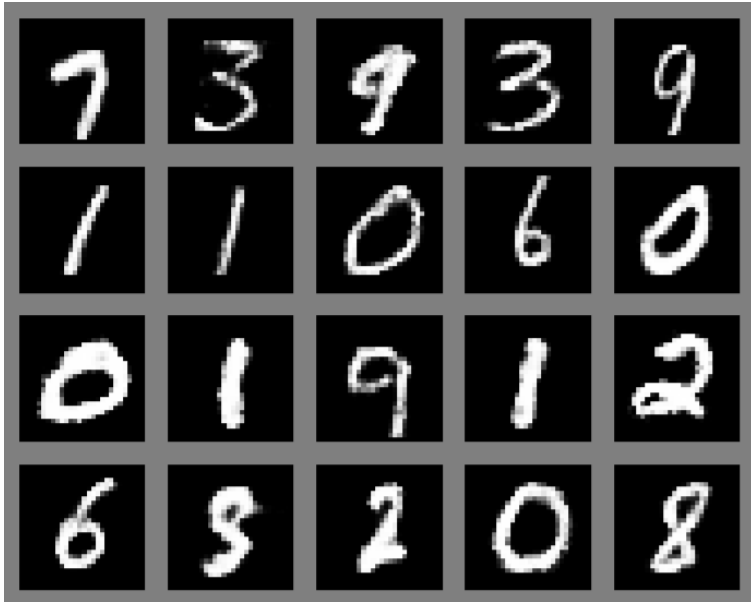


연구 배경: Generative Adversarial Networks (GAN)

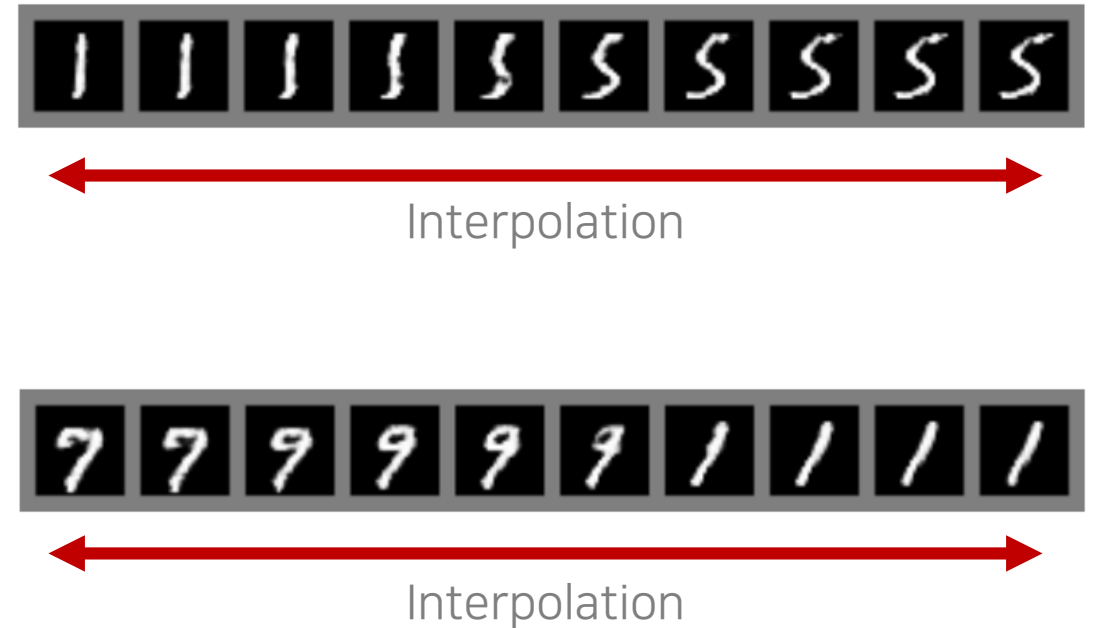
- 생성자(generator)와 판별자(discriminator) 두 개의 네트워크를 활용한 생성 모델입니다.

① 다양한 형태의 있을 법한 가짜 이미지 생성

$G(z)$ →



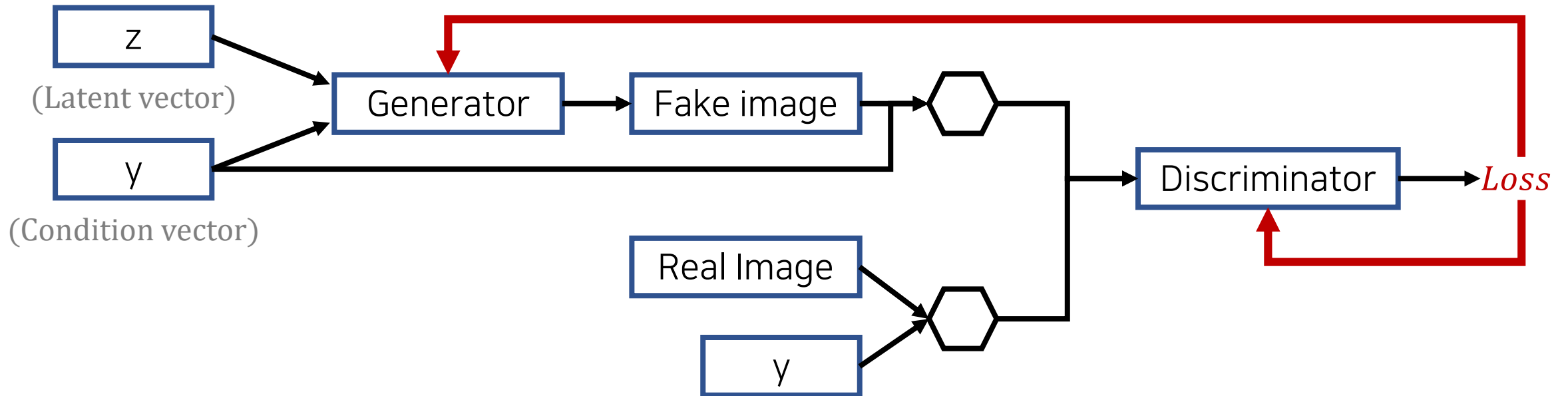
② Latent vector 사이의 interpolation 진행



연구 배경: Conditional GAN (cGAN)

- 데이터의 모드(mode)를 제어할 수 있도록 조건(condition) 정보를 함께 입력하는 모델입니다.

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x|y)] + E_{z \sim p_z(z)} [\log \{1 - D(G(z|y))\}]$$



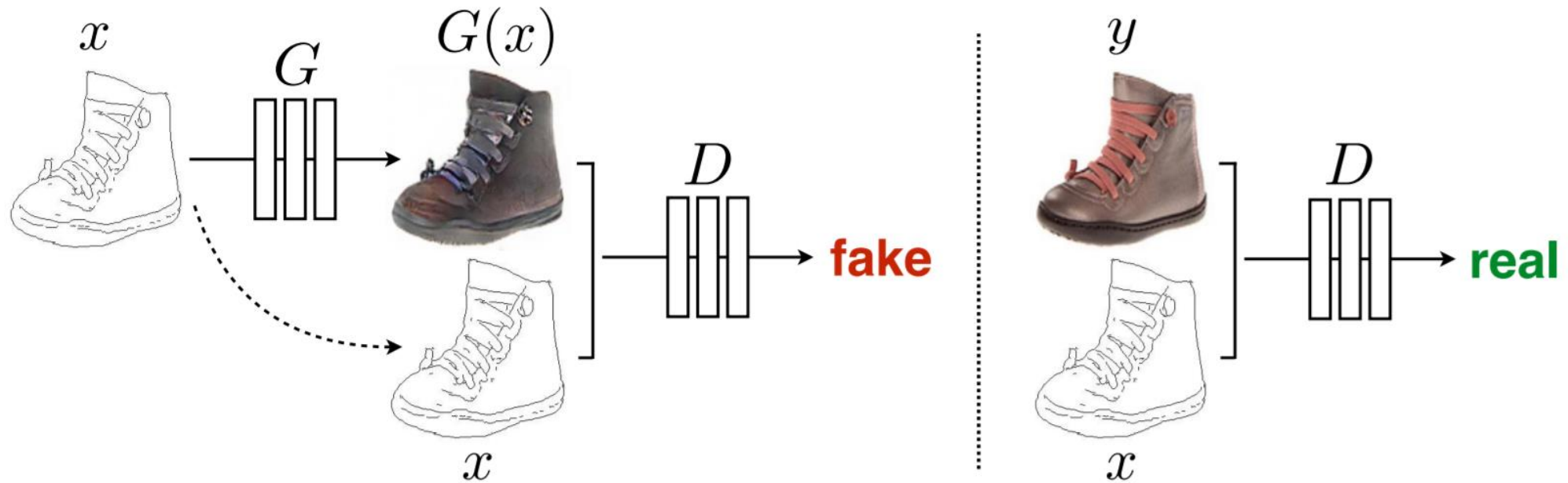
연구 배경: Conditional GAN (cGAN)

- 데이터의 모드(mode)를 제어할 수 있도록 조건(condition) 정보를 함께 입력하는 모델입니다.



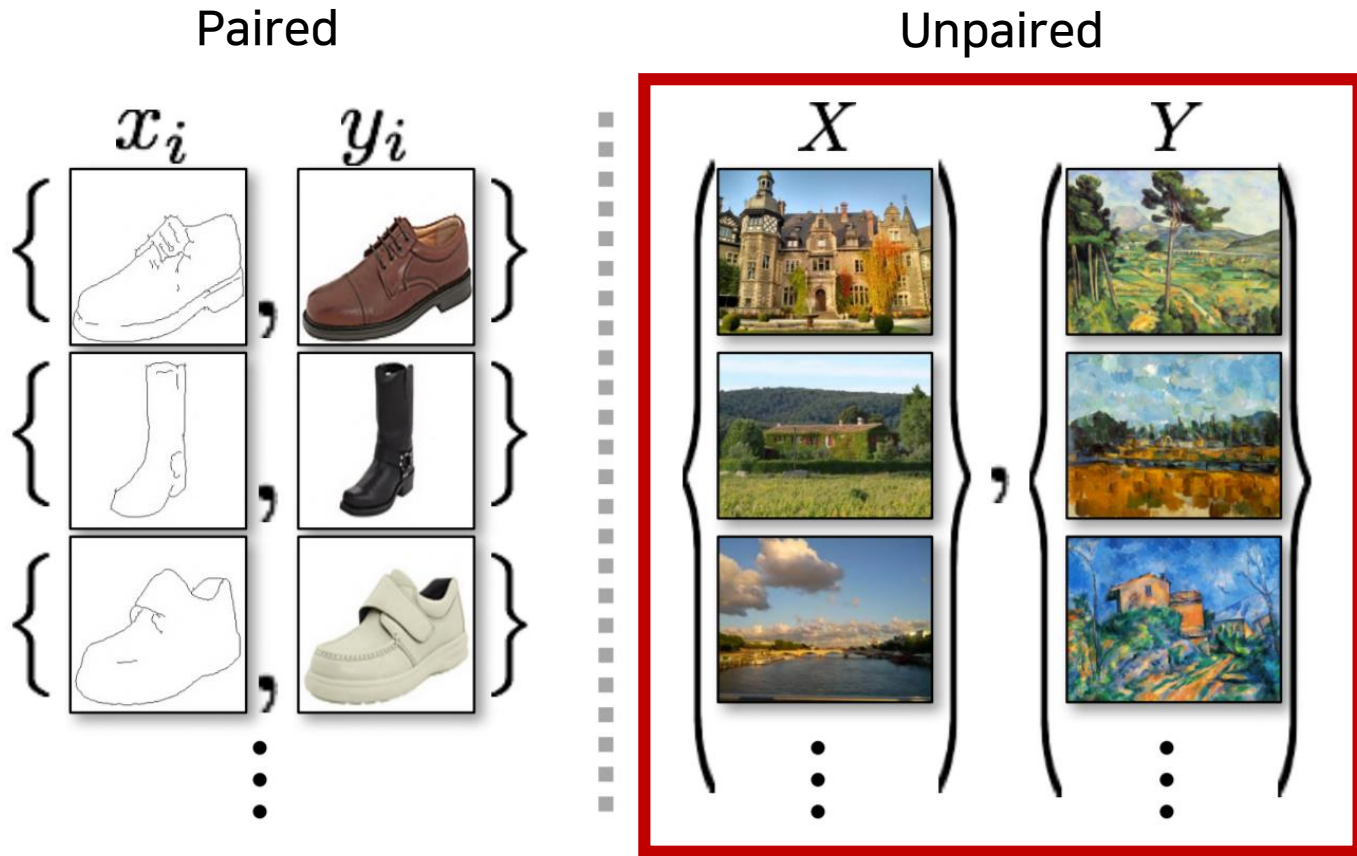
연구 배경: Image-to-Image Translation ① Pix2pix 개요

- The task of **I2I translation** is to change a particular aspect of a given image to another.
- 대표적인 Image-to-image translation 아키텍처로 pix2pix가 있습니다.
 - Pix2pix는 학습 과정에서 이미지를 조건(condition)으로 입력하는 cGAN의 유형으로 볼 수 있습니다.



연구 배경: Image-to-Image Translation ① Pix2pix 한계점

- Pix2pix는 서로 다른 두 도메인 X, Y 의 데이터를 한 쌍으로 묶어 학습을 진행합니다.
 - 다만 Colorization과 같은 태스크에서는 데이터셋을 구성하기 쉬우나 그렇지 않은 경우도 존재합니다.



한 쌍으로 묶이지 않은(Unpaired)
데이터 셋에 대해서도 적용이 가능할까요?



CycleGAN을 이용해 해결 가능

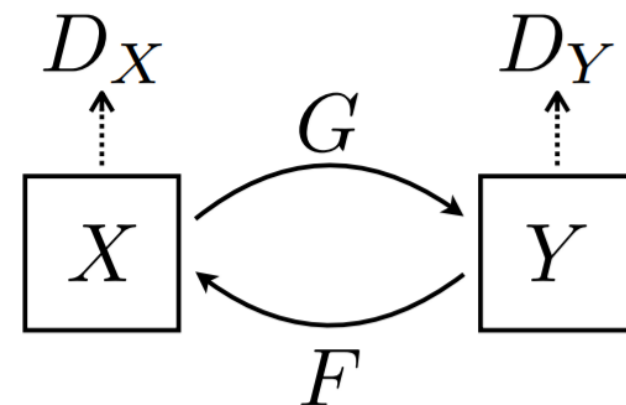
연구 배경: Image-to-Image Translation ② CycleGAN의 문제 상황

- 별도의 제약 조건 없이 단순히 입력 이미지 x 의 일부 특성을 타겟 도메인 Y 의 특성으로 바꾸고자 한다면 어떤 입력이든 상관없이 특정한 도메인에 해당하는 하나의 이미지만 제시하게 될 수도 있습니다.

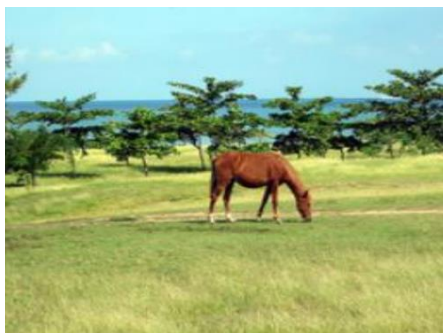


연구 배경: Image-to-Image Translation ② CycleGAN

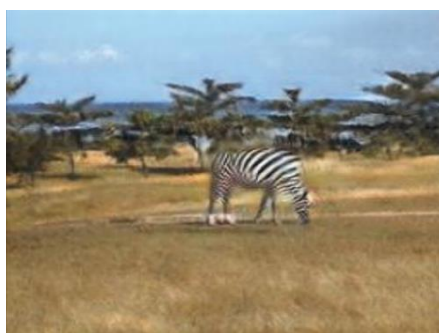
- CycleGAN은 $G(x)$ 가 다시 원본 이미지 x 로 재구성(reconstruct)될 수 있는 형태로 만들어지도록 합니다.
 - 이를 통해 **원본 이미지의 content는 보존(preserve)**하고 도메인과 관련한 특징을 바꿀 수 있습니다.
- 이를 위해 2개의 변환기(translator)를 사용합니다.
 - $G: X \rightarrow Y$
 - $F: Y \rightarrow X$
- Cycle-consistency loss를 사용합니다.



Input x



Output $G(x)$



Reconstruction $F(G(x))$



목표

$$\left[\begin{array}{l} F(G(x)) \approx x \\ G(F(y)) \approx y \end{array} \right]$$

연구 배경: WGAN-GP

- WGAN은 함수가 1-Lipshichtz 조건을 만족하도록 하여 안정적인 학습을 유도합니다.
 - 본래 WGAN 논문은 weight clipping을 이용하여 제약 조건을 만족하도록 합니다.
- **WGAN-GP**에서는 gradient penalty를 이용하여 WGAN의 성능을 개선합니다.

$$L = \underbrace{\mathbb{E}_{\tilde{\mathbf{x}} \sim \mathbb{P}_g} [D(\tilde{\mathbf{x}})] - \mathbb{E}_{\mathbf{x} \sim \mathbb{P}_r} [D(\mathbf{x})]}_{\text{Original critic loss}} + \underbrace{\lambda \mathbb{E}_{\hat{\mathbf{x}} \sim \mathbb{P}_{\hat{\mathbf{x}}}} [(\|\nabla_{\hat{\mathbf{x}}} D(\hat{\mathbf{x}})\|_2 - 1)^2]}_{\text{Gradient penalty}}$$

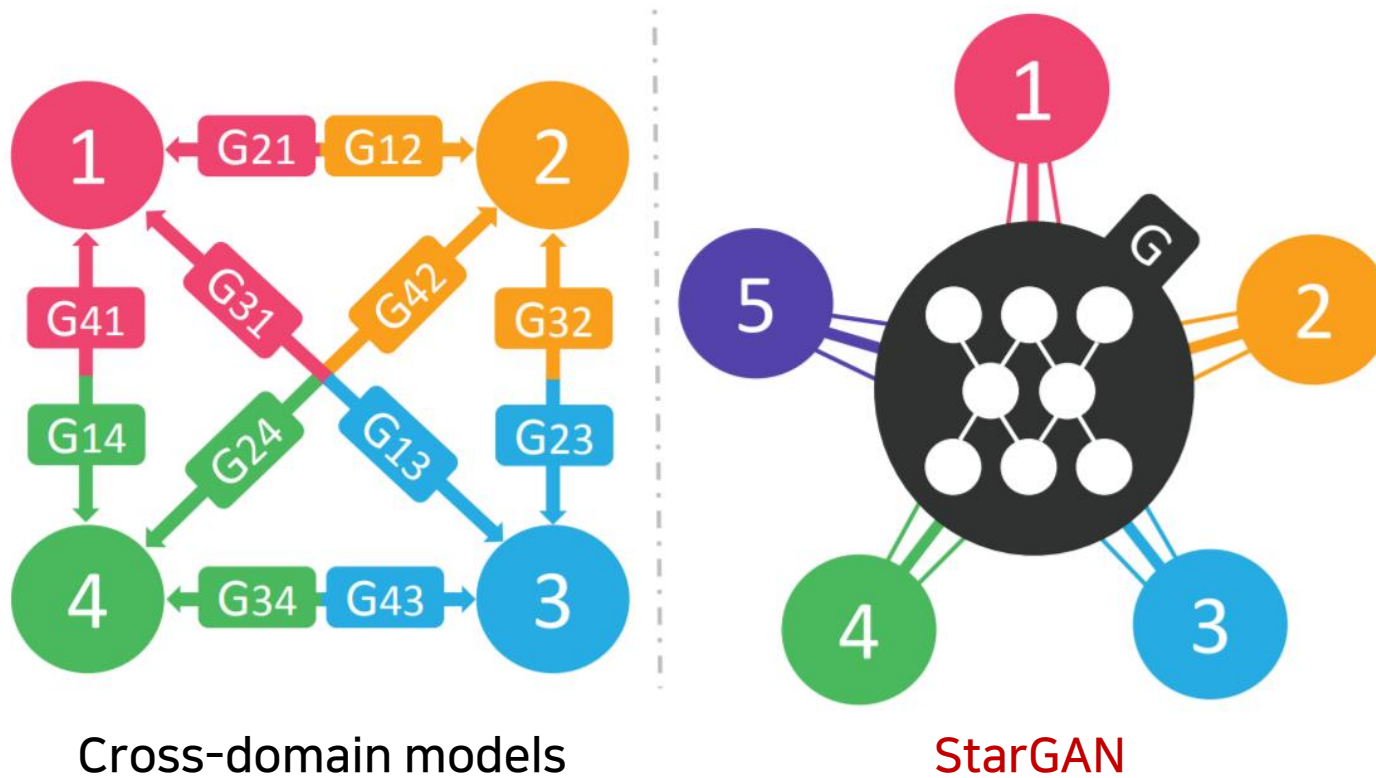
(Unsupervised case)
Inception Score 측정 결과



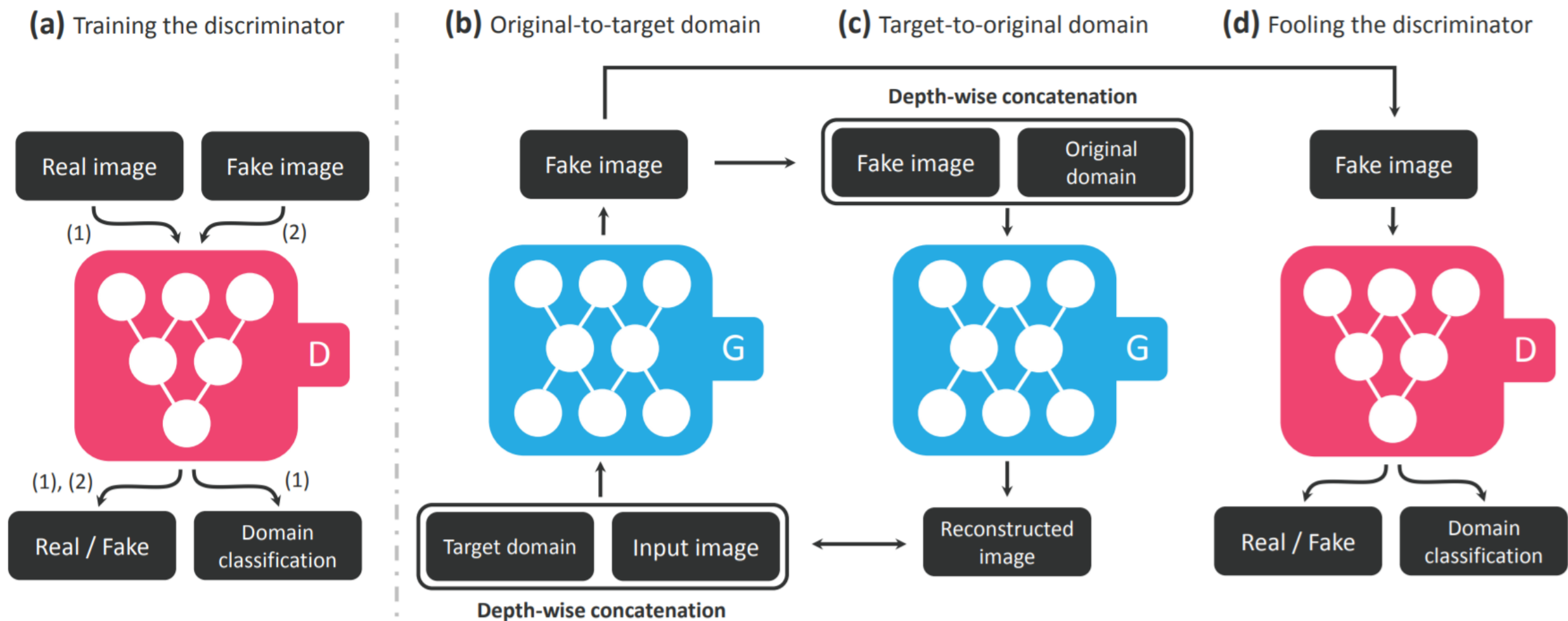
Method	Score
ALI [8] (in [27])	5.34 ± .05
BEGAN [4]	5.62
DCGAN [22] (in [11])	6.16 ± .07
Improved GAN (-L+HA) [23]	6.86 ± .06
EGAN-Ent-VI [7]	7.07 ± .10
DFM [27]	7.72 ± .13
WGAN-GP ResNet (ours)	7.86 ± .07

StarGAN: 다중 도메인을 위한 하나의 모델

- 하나의 뉴럴 네트워크를 이용해 다중 도메인(multi domain) 사이에서의 이미지 변환이 가능합니다.
 - 기존 image-to-image translation 아키텍처를 그대로 이용한다면 여러 개의 네트워크가 필요합니다.



StarGAN Overview



StarGAN Loss Function

- StarGAN: ① Adversarial loss + ② Domain classification loss + ③ Reconstruction loss

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

Domain classification $\left[\begin{array}{l} \mathcal{L}_{cls}^f = \mathbb{E}_{x,c} [-\log D_{cls}(c|G(x, c))] \\ \mathcal{L}_{cls}^r = \mathbb{E}_{x,c'} [-\log D_{cls}(c'|x)] \end{array} \right]$

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

최종 목적 함수 $\left[\begin{array}{l} \mathcal{L}_D = -\mathcal{L}_{adv} + \lambda_{cls} \mathcal{L}_{cls}^r \\ \mathcal{L}_G = \mathcal{L}_{adv} + \lambda_{cls} \mathcal{L}_{cls}^f + \lambda_{rec} \mathcal{L}_{rec} \end{array} \right]$

Mask Vector

- Multiple dataset에서의 학습을 위하여 Mask vector m 를 사용할 수 있습니다.

$$\tilde{c} = [c_1, \dots, c_n, m] \quad \leftarrow [\cdot]: \text{Concatenation}$$

CelebA label

Black / Blond / Brown / Male / Young

RaFD label

Angry / Fearful / Happy / Sad / Disgusted

Mask vector

CelebA / RaFD

Input

Angry

Disgusted

Fearful

Happy

Neutral

Sad

StarGAN
(SNG)



StarGAN
(JNT)



더 많은 데이터를
활용한 쪽이 높은
성능

Training with Multiple Datasets

