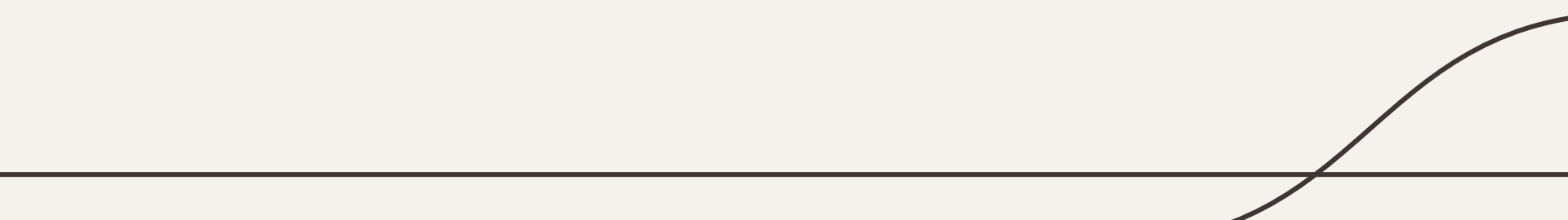




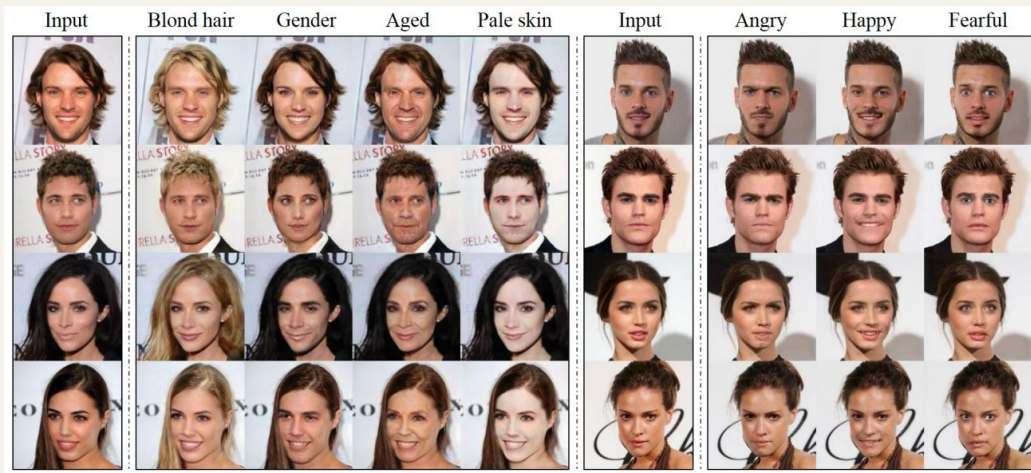
StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation

발표자 : 문경환



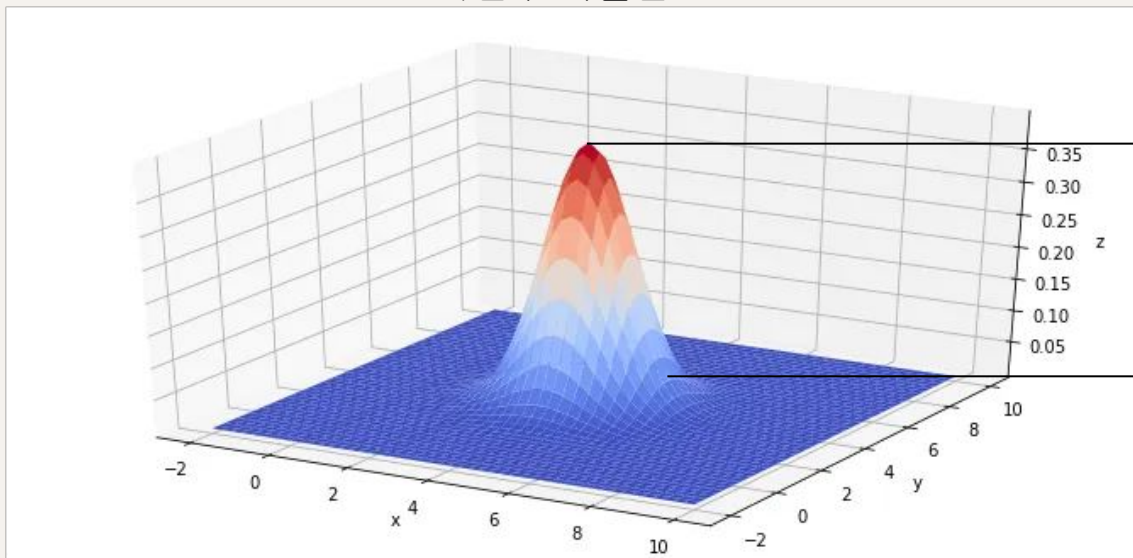
Abstract

- 기존의 Image-to-Image translation 모델에서 다양한 도메인간의 학습이 어려웠던 문제 해결.
- 데이터셋마다 가지고 있는 attribute의 차이로 인해 발생하는 문제를 mask vector를 사용해 해결.

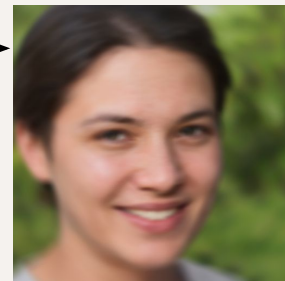
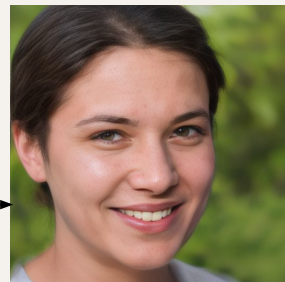


Generative model

다변수 확률분포

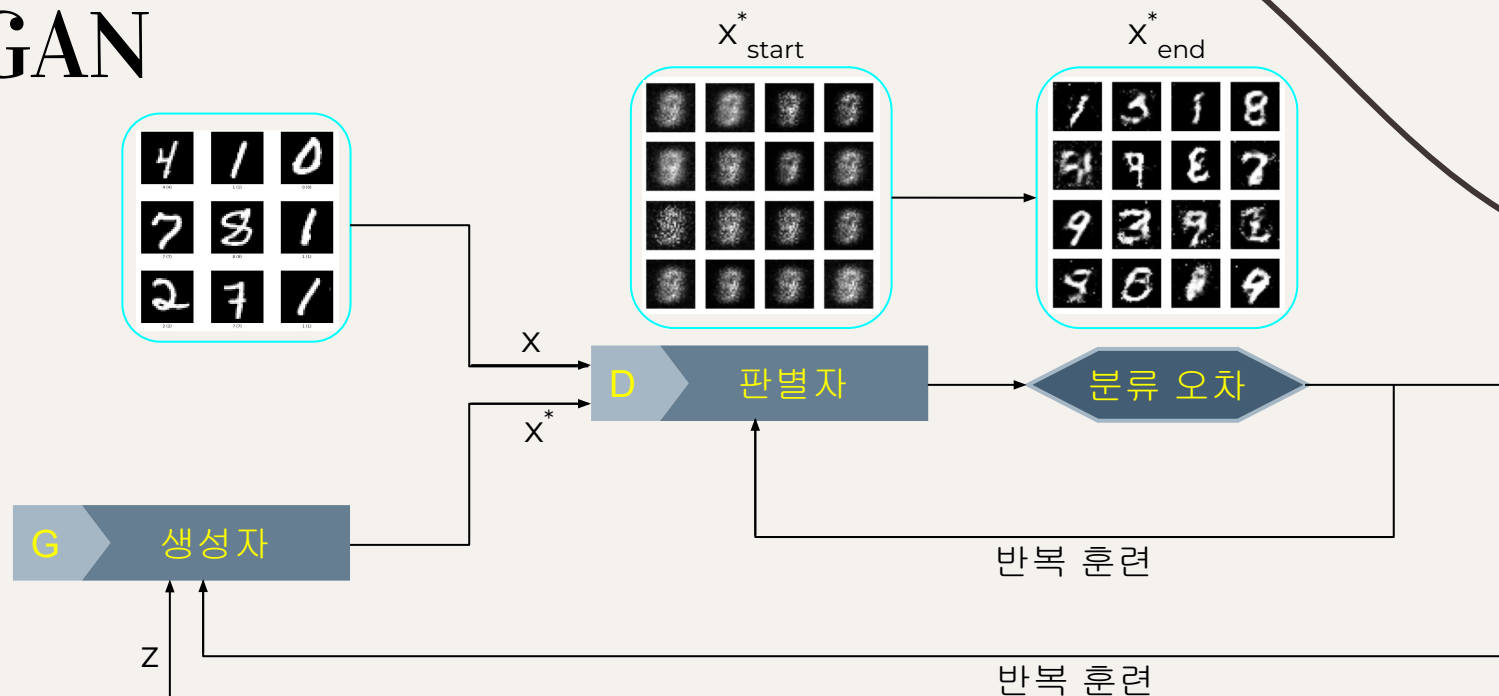


출처 : Aly Shmahell, medium



출처 : analyticsvidhya

GAN



Loss Function

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]$$



CGAN → Pix2Pix → CycleGAN

The image features a light gray background with two horizontal dark gray lines, one near the top and one near the bottom. A dark gray curve starts from the top right and arcs downwards towards the right edge.

CGAN \rightarrow Pix2Pix \rightarrow CycleGAN

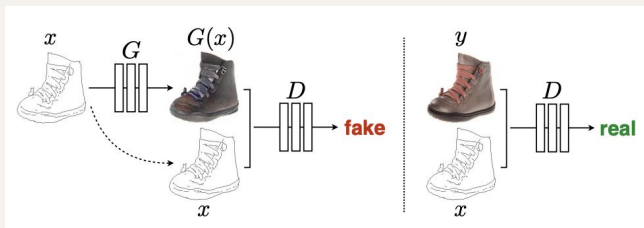


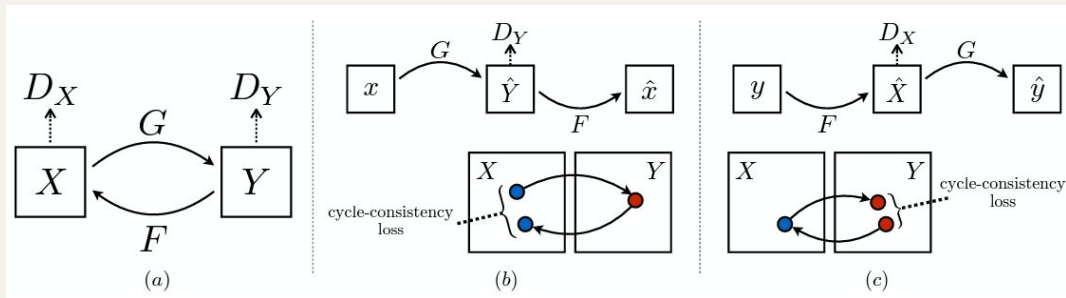
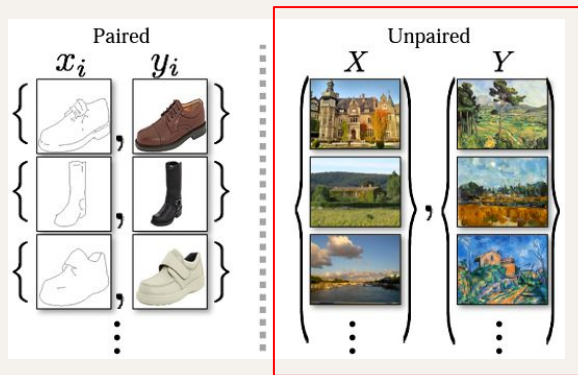
Figure 4: Different losses induce different quality of results. Each column shows results trained under a different loss. Please see <https://phillipi.github.io/pix2pix/> for additional examples.

The image features a light gray background with two horizontal dark gray lines, one near the top and one near the bottom. A dark gray curve starts from the top right and arcs downwards towards the right edge.

CGAN \rightarrow Pix2Pix \rightarrow CycleGAN

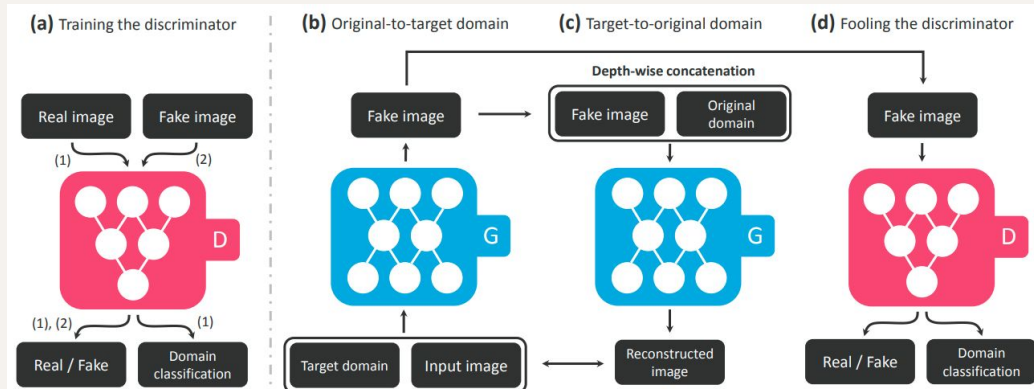
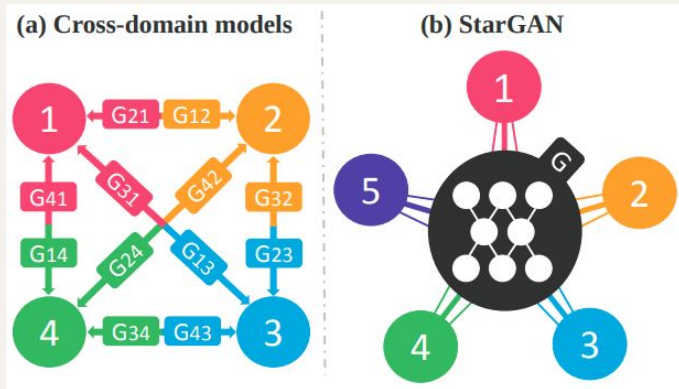
CycleGAN

- Unpaired 데이터셋으로 학습이 가능한 image-to-image translation 메서드.
- 기존의 Image-to-Image 모델에서 도메인 간의 학습이 이루어지지 않을 수 있었던 문제 해결.
- Cycle-consistency loss를 통해 다양한 Task에서 좋은 결과를 도출.



StarGAN

- 여러 도메인의 특징을 학습하는 generator G 를 학습시켜 입력 이미지인 x 를 타겟 도메인 c 에 대한 출력 이미지인 y 로 만들도록 학습함.
- Auxiliary classifier을 통해 discriminator이 여러 도메인을 컨트롤 할 수 있도록함 (이미지 source와 domain label에 대한 확률분포를 생성).



Mask Vector

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

- 데이터셋에 성별 등에 관한 label은 있었으나, 표정과 같은 label은 없었다. 이 문제를 해결하기 위해 mask vector을 제안하였음.
- n차원의 one-hot vector을 사용하였음.
- m : mask vector / n : 데이터셋의 수 / c_i : i번째 데이터셋의 label에 대한 vector
- Known label은 zero value를 배정함.
- 이를 통해 mask vector에 어떤 dataset인지를 명시하여 해당 dataset의 attribute에 관련된 label에 집중함.

Adversarial Loss

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

Adversarial Loss

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

Improved adversarial loss

- 생성된 이미지가 진짜 이미지와 구분되지 않도록 함.
- $G(x, c)$: G 는 이미지의 생성자로서, input image인 x 와 target domain label인 c 를 입력 받아 이미지를 생성함.
- $D_{src}(x)$: D 는 판별자로서, 진짜와 가짜 이미지를 판별함. 본 논문에서는 $D_{src}(x)$ 를 D 가 내보낸 source의 확률분포로 정의함.
- 결국 G 는 loss 값을 최소화 하려고 하며 D 는 loss 값을 최대화 하려고 함.
- Gradient penalty를 통해 더 퀄리티가 좋은 이미지를 생성해내도록 함.

Domain Classification/Reconstruction Loss

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

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

Domain Classification Loss

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

Reconstruction Loss

Domain Classification Loss

- 생성된 이미지가 적절하게 domain c 로 분류되도록 함.
- 이 식의 값이 작아지며 D 는 진짜 이미지를 원본 도메인으로 구분하는 것을 학습함.

Reconstruction Loss

- 생성된 이미지를 다시 원본 이미지로 돌렸을 때 원본이미지 x 와의 차이를 통해 loss 값을 책정함.

Full Objective & Network Architecture

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

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

- $\lambda_{cls}=1, \lambda_{rec}=10$

- Generator은 2개의 convolutional layer과, 6개의 residual block, 2개의 transposed convolutional layer로 구성됨.
- Generator은 Instance Normalization를 사용함 (Discriminator은 normalization을 사용하지 않음).
- Discriminator은 이미지를 patch 별로 나누어 real/fake를 판별함.

Performance

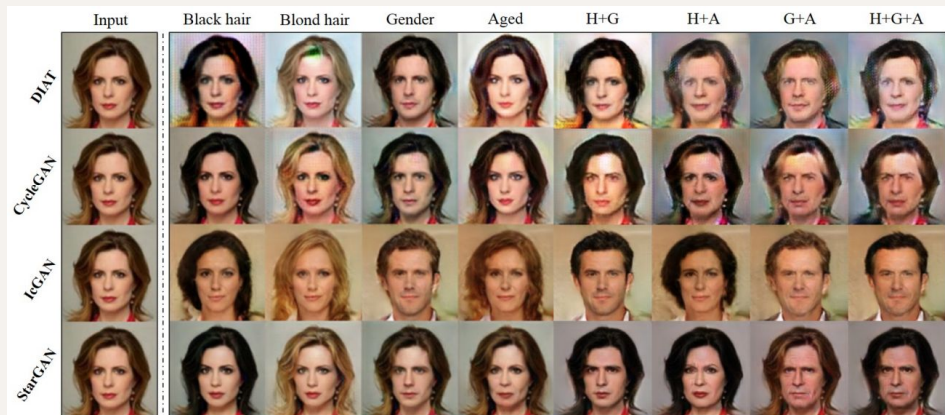


Figure 4. Facial attribute transfer results on the CelebA dataset. The first column shows the input image, next four columns show the single attribute transfer results, and rightmost columns show the multi-attribute transfer results. H: Hair color, G: Gender, A: Aged.



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

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.

How to improve StarGAN Performance

- L1 regularization을 추가함.
- StarGAN의 Generator을 추가해 서로 경쟁하는 형태로 이미지를 생성함.
- 더욱 많은 데이터셋으로 학습함.
- StarGAN에 다른 도메인을 학습할 수 있는 generator과 discriminator을 추가해 더욱 많은 도메인을 학습할 수 있도록함. 나아가 그 도메인을 학습한 generator을 통해 학습하는 최종 generator을 추가함.
- Reconstruction 부분을 여러 번 반복하여 도메인의 특징을 학습할 수 있도록함.

Q&A