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

> Choi, Yunjey, et al. "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

Abstract

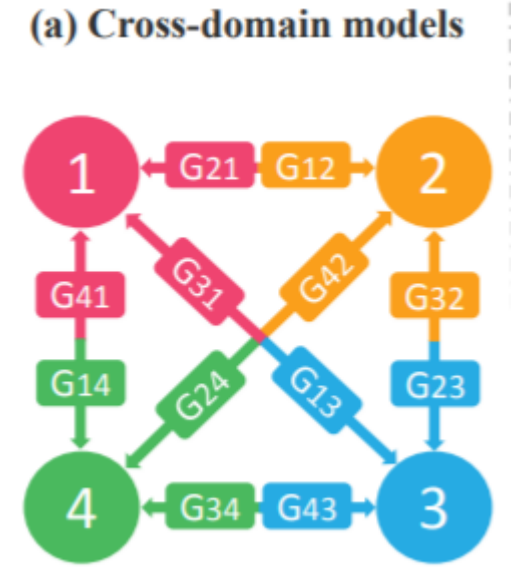
기존 연구에서는 2개 도메인 안에서 각 쌍으로 모델을 만들어 이미지 변환을 수행하였지만 그 이상을 수행하기위해 여러 도메인에 대해 1개의 모델로 이미지 변환을 수행할 수 있는 StarGAN을 제안한다.

학습시 image x 와 target domain c를 D가 동시에 참고하며 분류함으로 1개의 모델로 이미지를 변환 가능함.

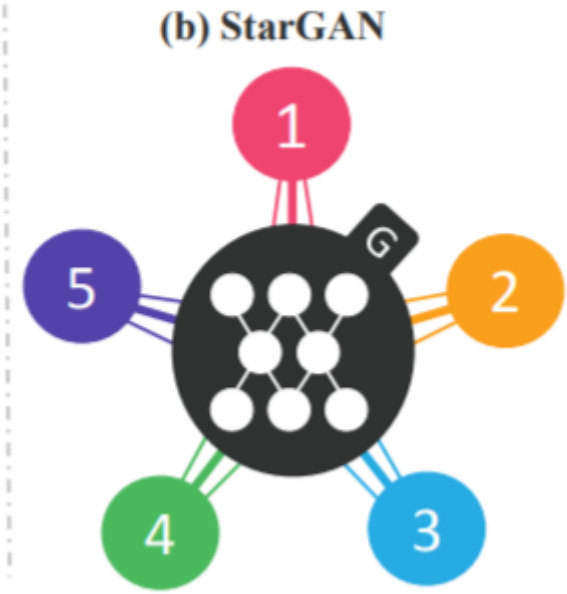
1. Introduction

StarGAN은 여러 도메인에서의 속성에 따른 이미지 변환 작업을 수행한다.

기존의 모델들은 k 개의 도메인이 존재한다면, k(k-1)개의 generator가 학습되어야 한다.
그 이유는 각 도메인 간에 서로 변환하는 모델(ex. 금발 -> 흑발, 흑발 ->금발)이 있어야 되기 때문.



위와 같은 경우 k개의 도메인중에서 2개의 도메인만 학습 가능하다. 데이터를 충분히 활용 못하면, 성능이 제한됨.



이러한 문제를 해결하기위해 위 사진과 같은 구조인 StarGan을 제안하였음. 이는 1개의 generator로 여러 도메인을 학습한다. 이는 간단히 generator에 이미지와, 도메인을 input으로 사용하고 학습 시, 무작위로 target도메인을 생성하고, 이미지를 target도메인으로 변환하도록 학습함으로써 이미지 변환이 이루어질 수 있다.

또한 도메인label에 mask vector를 더함으로써 다른 데이터 세트간 학습을 통합시는 방법을 제안한다. 이 방법은 모델이 특정 데이터 셋에서 제공된 label에 집중하고, unknown label(mask vector가 더해진 것?)을 무시함으로써 보증될 수 있다.

2. Related Work

- GAN
- Conditional GANs
- Image-to-Image Translation

3. Star Generative Adversarial Networks

Input 이미지 x 에 target 도메인 c로 변환시켜 output 이미지 y를 생성하는 모델 G는 다음과 같다.

$$G(x, c) \rightarrow y$$

이 때 G는 x의 매끄러운 변환을 위해서 c를 무작위로 발생시킨다.
여러 도메인의 제어를 위한 *1*개의 *D*를 만들기 위해서 *auxiliary classifier*를 선언함.
이를 통해 *D*는 *source*와 *domain label*의 확률 분포를 만들어낸다. -? 어떤 의미인지- c를 학습하기 위해?

Adversarial Loss

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

D는 진짜, 가짜 이미지를 구분하는 역할을 하며, G는 새로운 이미지 G(x,c)를 생성한다. D는 최대화, G는 최소하 함으로 학습이 진행된다.

Domain Classification Loss

이미지 x 를 domain label c에 맞게 변환시켜 y를 얻기 위해 D의 위에 auxiliary classifier를 둠으로써, 도메인 분류를 한다. 즉, domain classification loss는 D와 G로 나뉜다.

$$\mathcal{L}_{cls}^{r} = \mathbb{E}_{x,c'}[-logD_{cls}(c'|x)] \qquad (2)$$

위(real_image의 loss)를 최소화 하기 위해 D는 real image x와 그에 상응하는 original domain c'을 학습한다.

$$\mathcal{L}_{cls}^{f} = \mathbb{E}_{x,c'}[-logD_{cls}(c|G(x,c))] \qquad (3)$$

위 2,3 번 식의 최소화를 통해 G 는 타겟 도메인 c가 분류 될 수 있는 이미지를 생성하게 된다.

얼마나 많이 target domain으로 image translation을 할 것인가.

Reconstruction Loss

위의 1,2,3번 식(loss)를 최소화 함으로써 G는 이미지가 진짜인지, 올바른 target domain인지 분류 할수 있도록 학습된다. 하지만, 위 loss들의 최소화는 input image를 domain에 맞게 변경하는 동안 이미지 원본을 보존하는 것을 보장할 수 없다. 이러한 문제를 해결하기 위해 **cycle consistency loss**를 적용하였다.

$$\mathcal{L}_{rec} = \mathbb{E}_{x,c,c'}\left[||x - G(G(x,c),c')||_1\right] \qquad (4)$$

- Cycle consistency loss란?
  Mode Collapse(D,G의 학습 불균형 문제로 인해 G가 한 종류의 이미지만 생성하게 되는 문제) 해결을 위해서 Cycle GAN 논문에서 제안한 loss로 y -f(x) 의 형식이 아닌 x - g(f(x))를 사용하는 형태

위 처럼 원본 이미지인 (x,c)를 G를 통과 시켜 image translation 후 다시 원본이미지로 변화 시킨다.

즉 원본 이미지를 얼마나 보존할 것인가를 의미.

Full Objective

전체 목적함수는 다음과 같다.

$$\mathcal{L}_D = -\mathcal{L}_{adv} + \lambda_{cls}\mathcal{L}_{cls}^{r}, \qquad (5)$$

$$\mathcal{L}_G = \mathcal{L}_{adv} + \lambda_{cls}\mathcal{L}_{cls}^{f} + \lambda_{rec}\mathcal{L}_{rec} \qquad (6)$$

$$\lambda_{cls} = 1, \quad \lambda_{rec} = 1$$

위와 같은 목적함수에서의 hyper parameter를 통하여 이미지를 얼마나 보존할 지, 변화 시킬지 결정 가능하다.

3.2 Training with Multiple Datasets

이처럼 StarGAN은 여러 다른데이터셋에서 학습을 하는 동안의 문제점이 있는데, 이는 facial expression인 행복, 화남과 같은 domain을 포함하지 않지만, 머리색, 성별같은 속성들은 포함하는 경우를 의미한다.

이는 translated image로 input image를 학습 하는 과정에서 label이 부족하기 때문에 발생할 수 있다.

Mask Vector

위의 문제를 해결하기 위해서, 저자는 특정 데이터 셋에서의 label에 focus를 두고 그 외의 label은 무시할수 있도록 하기 위해 mask vector m을 추가하였다.

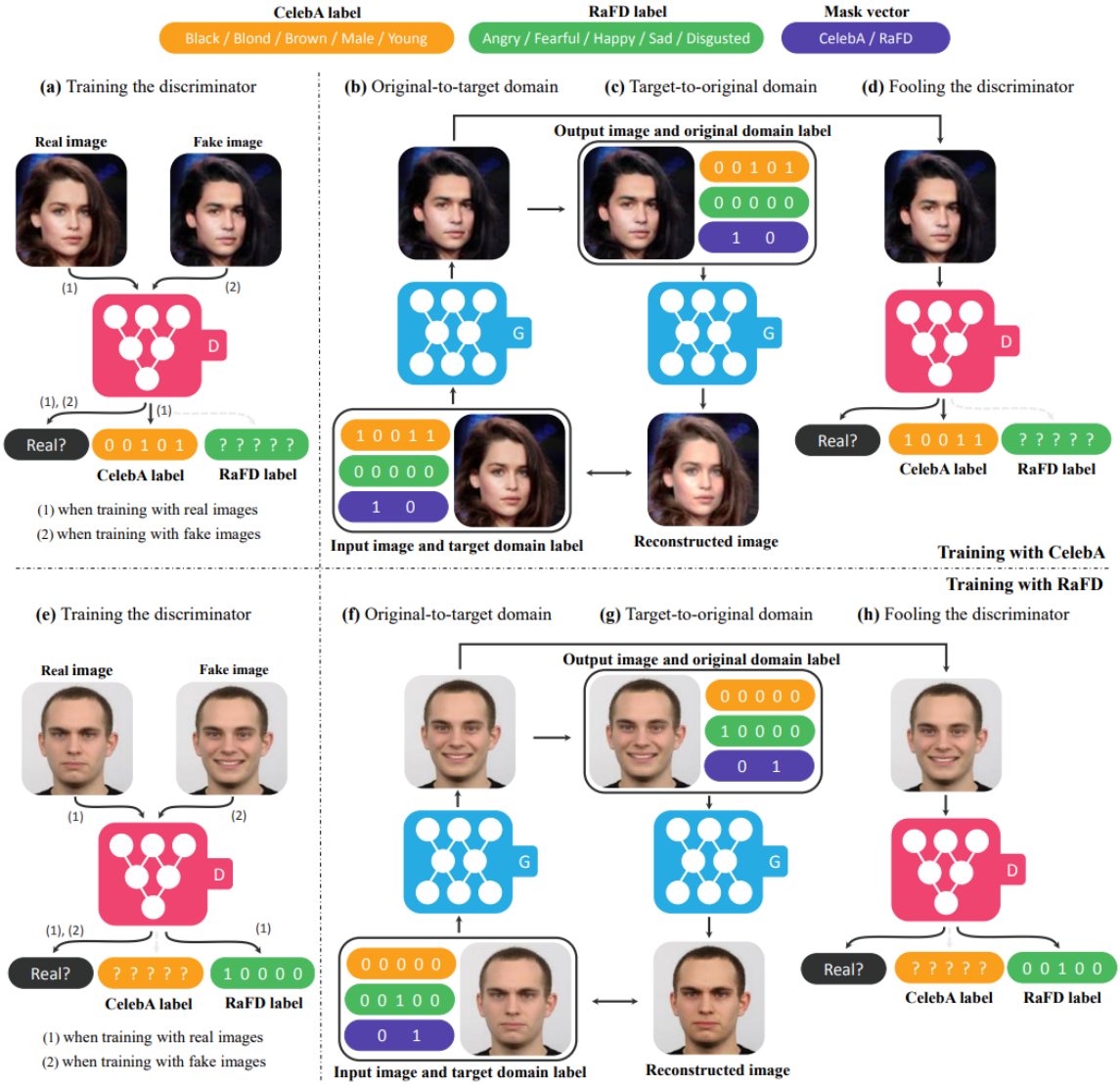
$$\tilde{c} = [c_1, \cdots\cdots, c_n, m] \qquad (7)$$

위에서의 c_n는 각 데이터 셋을 의미하고 어떤 데이터셋인지 명시하기 위해 m이라는 mask vector를 추가한다.

그래서 본 논문에서는 CelebA와 RaFD 데이터 셋을 사용하기 때문에 n=2가 된다.

Training Strategy

모든 데이터셋에서 전반적인 label에 맞는 확률 분포를 만들기 위해 D의 auxiliary classifier를 확장시킨다.
이를 통해 D는 명확한 label에 관해서만 classification error를 최적화 시키도록 하게 된다.



4. Implementation

GAN의 학습 안정성을 위해 Eq(1)대신에 gradient penalty와 Wasserstein GAN objective로 대체 하였다.

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

여기서 gradient penalty에 대한 파라미터로는 10을 사용했다고 한다.

5. Experiments

Learned role of mask vector

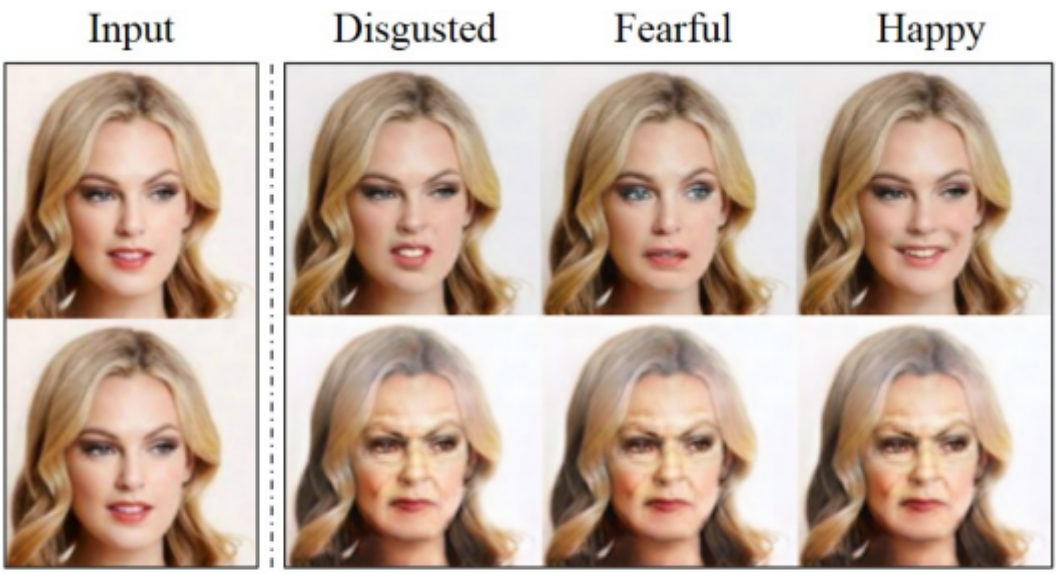


Figure 7. Learned role of the mask vector. All images are generated by StarGAN-JNT. The first row shows the result of applying the proper mask vector, and the last row shows the result of applying the wrong mask vector.

위의 그림은 mask vector의 효과를 알아보기 위해 잘못된 mask vector를 사용했던 실험의 결과로, 그림을 보면 각 감정에 맞는 표정합성에 실패하고 나이를 조작하고 있는 것을 볼 수 있다. 그런 이유로는 표정에 관한 벡터들을 무시하고, 나이에 관한 벡터들을 입력으로 받았기 때문이라고 한다. 이를 통해, 원하는 label에 focus를 둘 수 있도록 하려는 mask vector의 역할을 제대로 수행하고 학습했음을 확인할 수 있다.