



StarGan

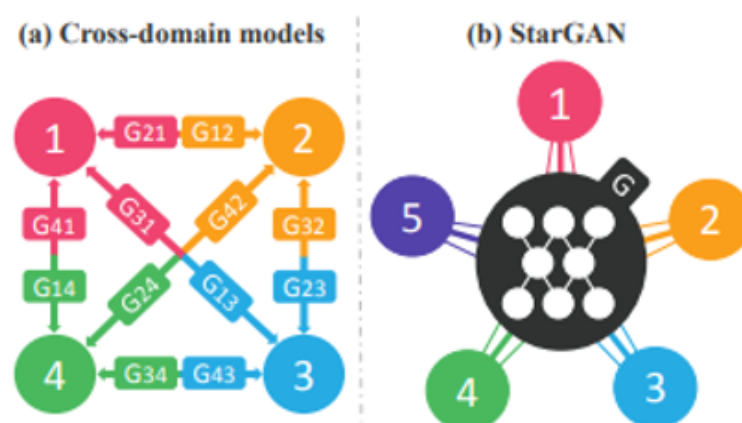
StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation

Abstract

이전 모델인 CycleGAN은 이미지에서 도메인 쌍을 바꿔주는 역할을 해내기 위해서 두 개의 생성기를 통해 이를 해결할 수 있었습니다. 하지만 두 개의 도메인에 국한하지 않고 여러 도메인에 접근 가능한 StarGAN을 제안합니다. StarGAN을 통해 여러 도메인을 가진 여러 데이터 세트를 동시에 훈련이 가능해진다.

Introduction

예를 들어, 남성에서 여성이라는 특징으로 변경할 시 남성 이미지와 여성 이미지를 쌍으로 하여 데이터를 학습할 수가 있다. 하지만, CelebA 데이터 세트에는 머리색, 성별, 나이와 같은 얼굴 속성과 관련된 40개의 레이블 얼굴 표정과 관련된 8개의 레이블 등이 존재하여 각각 학습해주는 것은 비효율적이다.



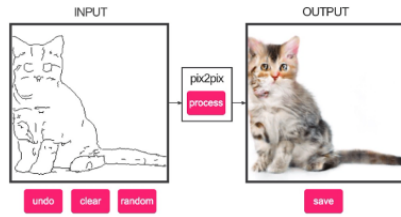
그렇기에 한 번에 학습할 수 있는 StarGAN을 제안한다. StarGAN은 여러 도메인의 학습 데이터를 가져와 단일 생성기를 사용하여 모든 도메인 간의 매핑을 학습한다.

또한, 마스크 벡터를 활용하여 여러 데이터 세트 간의 다중 도메인 이미지 변환이 가능하다.

기존 Cycle Gan 혹은 Style Gan의 경우에는 angry라는 특징을 줄 때, 모델 안에서 angry의 정도만을 설정할 수 있지 이를 happy라는 특징을 줄 수는 없는 것이다! 그렇기 때문에 특정 특성에서 다른 특성으로 넘어갈 수 있도록 Star Gan이 제시된다.

Related Work

2.1 pix2pix



2.2 CycleGAN

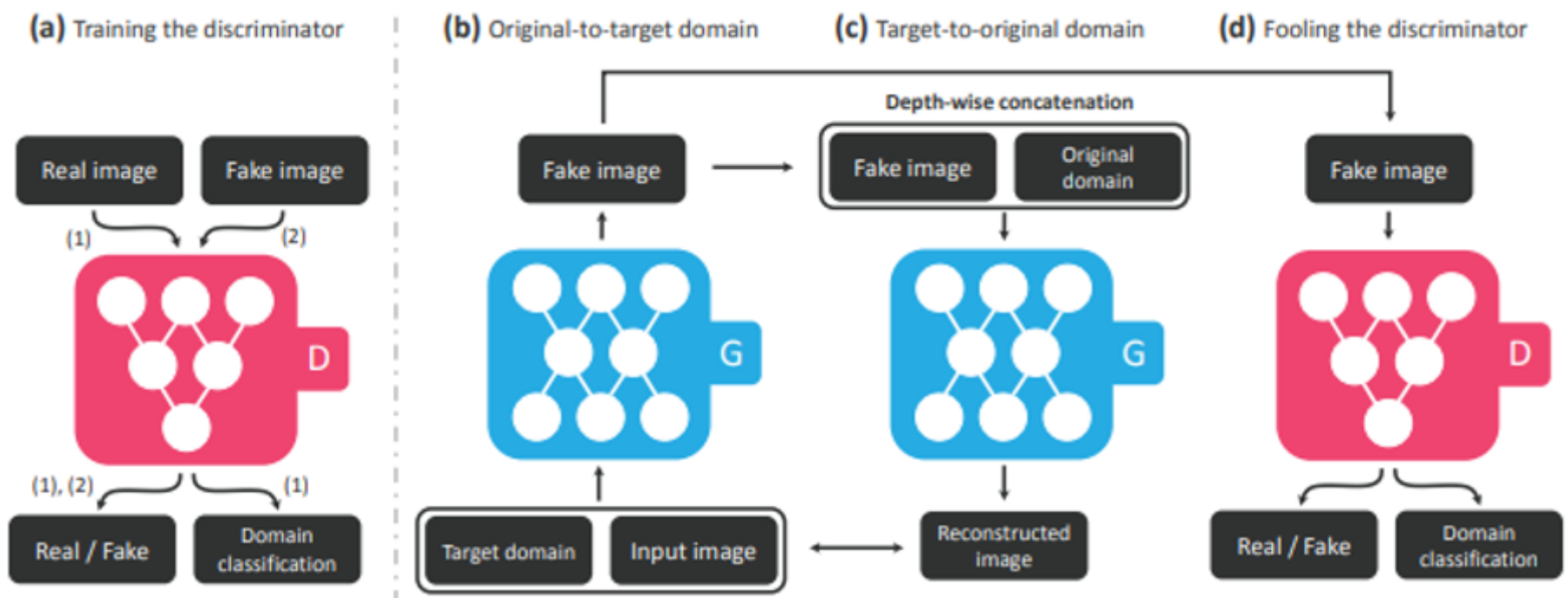


pix2pix cGAN을 사용한 방식이고 쌍을 이루는 데이터 샘플이 필요합니다.

CycleGAN 프레임워크는 한 번에 두 개의 서로 다른 도메인 간의 관계만 학습할 수 있습니다.

Star Generative Adversarial Networks

3.1. Multi-Domain Image-to-Image Translation



1. Discriminator는 진짜 이미지와 가짜 이미지를 판단할 뿐 아니라 Domain classification을 통해 이미지가 어떠한 도메인을 가졌는지 분류를 한다.
2. Original to target domain G는 이미지와 타겟 도메인을 인풋으로 받는다. 이 때 타겟 도메인은 원 핫 인코딩으로 만들며 랜덤으로 생성한다.
3. Target to original domain G는 B에서 생성한 가짜 이미지를 인풋으로 받아 다시 진짜 도메인과 G에 들어가 학습을 한다. 이 때의 아웃풋을 Reconstructed image라 하며 원래의 이미지와 비교해 차이를 줄이도록 학습한다.
4. Fooling the discriminator는 b에서 만든 가짜 이미지를 인풋으로 받아 진짜 이미지인지 가짜 이미지인지를 구분하고 a에서와 마찬가지로 도메인을 구분한다. 이 결과는 G에 피드백되어 G는 더욱 진짜 같은 이미지를 만들게 된다.

모델의 목표는 생성자 G를 학습하는 것이다. G는 입력 이미지 x 와 도메인 c 를 이용하여 출력 이미지 y 를 목표로 학습하게 된다.

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

생성자 G는 이미지를 통해 새롭게 생성된 $G(x, c)$ 를 만들고 판별자 D는 만들어진 이미지와 구분하고자 합니다.

여기서 생성자는 로스를 최대화하고자 하고, 판별자는 로스를 최소화하고자 합니다.

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

실제 이미지의 도메인 분류 손실을 나타냅니다.

$$\mathcal{L}_{cls}^f = \mathbb{E}_{x,c}[-\log D_{cls}(c|G(x,c))]. \quad (3)$$

가짜 이미지의 도메인 분류 손실을 나타냅니다.

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

만든 이미지를 다시 인풋 이미지와 비교했을 때의 손실을 나타냅니다.

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

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

최종적으로 생성자 G와 판별자 D의 손실 값 입니다.

여기서 λ 는 하이퍼 매개변수로 상대적 중요성을 나타내고 논문에서는 λ 을 사용하였습니다.

3.2. Training with Multiple Datasets

StarGan의 큰 장점 중 하나는 여러 데이터셋을 한번에 넣어줄 수 있다는 것입니다. 각각의 데이터 셋은 특정 성질을 각각 학습 할 수 있도록 구성되어있습니다.

예를 들어, 1번째 데이터 셋은 표정을, 2번째 데이터 셋은 얼굴의 모형을 다룬다면 이 두가지 데이터셋을 한번에 모델에 집어넣고 마스크 벡터를 사용하여 데이터마다 학습할 부분을 집어넣어 줄 수 있습니다.

A 데이터 셋의 특징이 3 가지고 이고 B 데이터 셋의 특징이 4가지라면 이 벡터를 결합한 뒤 학습할 부분만을 열어 두는 방식입니다.

A [1, 0, 1] + B [0, 0, 0, 0] 단, 각각의 특징은 원 핫 인코딩을 통하여 표현합니다.

데이터 셋을 구분하는 vector를 추가적으로 집어넣고 데이터 셋 마다 학습할 부분을 미리 지정해주는 방식을 통해 데이터 셋 마다 학습할 부분을 지정해 줄 수 있다.

A [1, 0, 1, 0] + B [0, 0, 0, 0] + Mask vector [1, 0]

Experimental

데이터의 훈련 방식은 Adam을 사용하였고, 데이터의 증대를 위해 뒤집기를 사용하였습니다.

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%

모델의 평가를 위해 아마존 설문을 사용하였고, 이를 통해 StarGAN은 다른 모델에 비해 좋은 평가를 받았습니다. 다른 모델에 비해 성능이 좋은 점은 하나의 생성자에 보다 많은 데이터를 학습할 수 있기 때문입니다.

각 열의 합은 100%.

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

또한, 다양한 속성을 사용하는 태스크에 있어 StarGan은 하나의 생성자와 판별자만을 통해 구현이 가능하므로 파라미터의 수 또한 다른 모델에 비해 적은 것을 볼 수 있습니다. 그리고 분류에 대한 오류 또한 작습니다.

Conclusion

Stargan은 단일 생성기와 단일 판별기를 사용하여 여러 도메인 간의 확장 가능합니다.

그리하여 기존 방법에 비해 작업이 쉬워졌으며, 보다 높은 품질의 이미지를 생성하게 되었습니다.

또한, 마스크 벡터를 사용하여 서로 다른 도메인 레이블 세트를 가진 여러 데이터 세트를 활용할 수 있습니다.

네이버 클로바 여러분들 감사합니다.