

Generative Adversarial Networks 를 이용한 다중 도메인 간 이미지 변환 기술

염태선¹, 고강혁², 이민혁^{2*}

중앙대학교 기계공학부¹, 중앙대학교 전자전기공학부²
 taesun0415@cau.ac.kr, dogworld12@cau.ac.kr, *mlee@cau.ac.kr

A study on Image-to-Image Translation between Large Scale of Domains using Generative Adversarial Networks

Yeom Taesun¹, Ko Kanghyeok², Lee Minhyeok^{2*}
 School of Mechanical Engineering, Chung-Ang Univ¹
 School of Electrical and Electronics Engineering, Chung-Ang Univ²

요약

본 논문에서는 기존의 생성적 대립 신경망(Generative Adversarial Networks, GAN)을 기반으로 한 다중 도메인 간 이미지 변환 기술(multi-domain image-to-image translation)을 발전시키기 위해, 기존에 제시된 StarGAN 모델의 Discriminator 에서 Classifier 를 분리하여 기존의 문제점을 해결한 모델인 CstarGAN 을 제안한다. CstarGAN 은 기존의 모델과 다르게 데이터 증강(Data augmentation)을 사용하여 Classifier 를 효율적으로 학습시키는 과정이 존재하고, 이러한 과정을 통해 많은 수의 도메인에 대해서도 이미지 변환 시 좋은 결과물을 획득하게 된다. 또, 기존에 제대로 변환이 되지 않던 미세한 특성들에 대해서도 성능 향상을 보였다. 여러 평가 지표와 다른 모델들과의 비교를 통해 제시한 모델의 성능이 더 높음을 확인하였다.

1. 서론

딥러닝 기술이 발전함에 따라 이미지를 원하는 형태로 변환시켜주는 이미지 간 변환(image-to-image translation) 기술에 대한 연구가 활발히 이루어지고 있다. 그 중 생성적 대립 신경망(Generative Adversarial Networks, GAN)[1]을 기반으로 한 모델들이 중심을 이룬다.

GAN은 Generator와 Discriminator로 이루어져 있으며, 잠재 공간(latent space)의 noise vector를 입력으로 받아 Generator는 거짓 정보를 생성하고, Discriminator는 이를 참인지 거짓인지 구분하도록 학습이 진행된다.

StarGAN[2]은 GAN을 기반으로 한 image-to-image translation 모델 중 하나로, 기존의 연구들과 달리 하나의 Generator로 다중 도메인(multi-domain) 간 이미지 변환이 가능하다는 점이 특징이다. 하지만 이 모델은 도메인의 수가 늘어났을 때와 미세한 특징 변화 시 모델이 정상적으로 이미지를 생성하지 못한다는 문제점이 존재한다.

본 논문에서는 앞서 설명한 StarGAN의 한계점인 많은 도메인의 수에도 적용이 가능한 image-to-image translation 모델인 CstarGAN을 제안한다. 이는 StarGAN의 판별자에서 Classifier를 분리하는 방법으로 구현할 수 있으며, 이 때 Classifier에 입력되는 정보에 대하여 데이터 증강(Data augmentation)을 사용하였다.

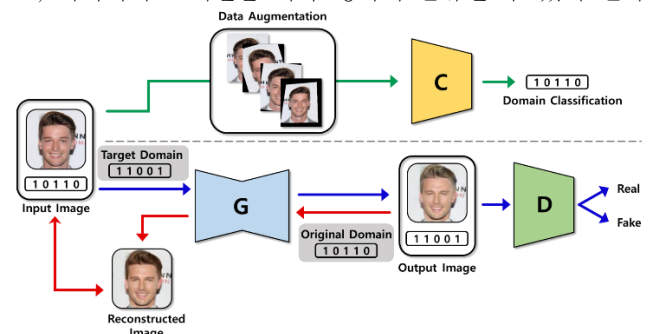
II. 본론

2.1. 모델 구조

StarGAN의 한계점을 해결하기 위해 본 논문은 Controllable Generative Adversarial Networks[3]에서 사용된 독립된 Classifier의 구조를 이용하였다. 본 논문의 모델 구조를 <그림 1>로 표현할 수 있다. 모델은 <그

림 1>의 아래 부분과 위 부분으로 나눌 수 있는데, 아래 부분에서는 입력 이미지를 목적 도메인(target domain)에 대한 이미지로 변환시키고 생성된 이미지를 참과 거짓으로 판별한다. 이에 대한 역의 과정으로 생성된 이미지를 기존 도메인(original domain)에 대하여 다시 변환시켜주고 이를 입력 이미지와의 손실 함수를 계산해 학습에 반영한다. 이 과정을 통해 Generator는 더욱 사실적인 이미지를 생성할 수 있게 된다.

위 부분에서는 입력 이미지를 Data augmentation을 사용하여 Classifier에 입력시켜준다. 이를 통해 Classifier는 많은 양의 이미지로 학습을 할 수 있게 되고, 이미지의 도메인을 더욱 정확히 분류할 수 있게 된다.



<그림 1. 본 논문의 모델 구조>

2.2. 학습 과정

모델의 학습에는 AFHQ[4] 데이터셋이 사용되었다. CstarGAN에서는 기존의 모델과 다르게 Discriminator의 각 층에 Spectral Normalization을 사용하여 학습의 안정성을 높일 수 있도록 했다. Data augmentation은 이미지를 무작위로 각도 조정하여 회전시키는 방법을 사용하였다.

다음 식들은 본 모델의 Generator, Discriminator, Classifier의 손실함수이다.

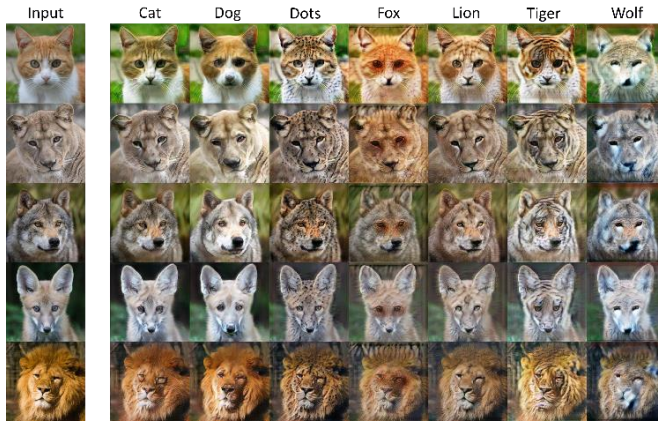
$$\hat{G} = \underset{\theta_G}{\operatorname{argmin}} \left\{ L_{adv}(r, D(G(X_{real}, L))) + \lambda_{cls} \cdot L_{cls}(L, C(G(X_{real}, L))) \right. \\ \left. + \lambda_{rec} \cdot L_{rec}(L, C(G(X_{real}, L), L')) \right\} \quad (1)$$

$$\hat{D} = \underset{\theta_D}{\operatorname{argmin}} \left\{ L_{adv}(r, D(X_{real})) + L_{adv}(f, D(G(X_{real}, L))) \right\} \quad (2)$$

$$\hat{C} = \underset{\theta_C}{\operatorname{argmin}} \{ L_{cls}(L, C(X_{real_aug})) \} \quad (3)$$

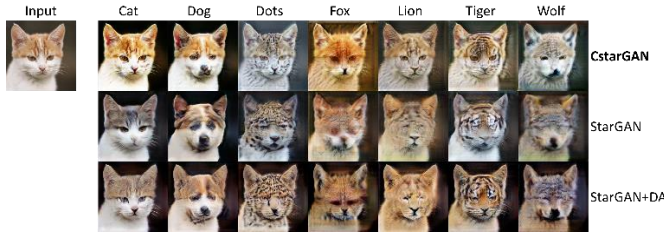
각 식에서 L 은 목적 도메인, L' 은 기존 도메인, X 는 입력 이미지를 의미한다. 식 (1)은 Generator의 손실함수로 adversarial loss, classification loss, reconstruction loss의 세 항으로 이루어져 있다. 식 (2)는 Discriminator의 손실함수로 이는 Discriminator가 참과 거짓을 정확하게 구분할 수 있도록 학습됨을 의미한다. 이 때 손실함수의 각 항은 기존의 모델과 다르게 hinge loss를 사용했다. 식 (3)은 Classifier의 손실함수로 입력 이미지에 대해 Data augmentation이 이루어진 후 이를 정확한 도메인에 분류할 수 있는 방향으로 학습됨을 보여준다. 이 세 손실함수는 학습 과정에서 역전파되어 각각 Generator, Discriminator, Classifier의 가중치를 업데이트 해주는 방식으로 학습이 진행된다.

2.3. 실험 결과



<그림 2. 각 도메인으로의 변환 결과>

모델의 실험에는 학습시와 동일하게 AFHQ 데이터셋이 사용되었다. <그림 2>는 입력 이미지를 일곱 개의 도메인으로 변환한 결과물이다. 이미지가 각 도메인 별 특징을 가진 채로 성공적으로 변환된 것을 살펴볼 수 있다.



<그림 3. 비교 모델들 간의 결과물 비교>

Method	FID↓	LPIPS↓
StarGAN	74.3	0.469
StarGAN+DA	70.1	0.420
CstarGAN	39.7	0.240

<표 1. 지표 평가 수치>

결과에 대한 평가 지표는 FID(Fr chet Inception Distance), LPIPS(Learned Perceptual Image Patch Similarity)가 사용되었다. 이 두 지표는 모두 낮을수록 모델이 좋은 성능을 보임을 알려주는 수치이다. <표 1>에서 CstarGAN이 비교 모델들에 비해 좋은 지표를 보이는 것을 확인할 수 있다. <그림 3>은 CstarGAN과 비

교 모델들의 입력 이미지에 대해서 도메인 간 변환을 한 결과물이다. 비교 모델은 두 종류로 하나는 StarGAN, 나머지 하나는 StarGAN에서 Data augmentation을 적용시킨 모델을 사용하였다. <표 1>에서의 결과와 마찬가지로 본 논문에서 제시한 모델이 가장 좋은 결과물을 얻은 것을 눈으로 확인할 수 있다.

III. 결론

본 논문에서는 StarGAN을 발전시킨 새로운 모델 CstarGAN을 제안하였다. Data augmentation을 이용해 기존 StarGAN의 Discriminator에서 분리된 Classifier를 학습시켰고, 그 결과 비교 모델들에 비해 좋은 결과를 얻을 수 있었다. 본 논문에서는 학습과 평가 과정에서 얼굴에 대한 데이터셋만 사용하였지만, 다른 종류의 데이터셋에도 적용하여 실험한다면 이미지 변환 기술의 발전에 도움이 될 것이다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2021R1F1A1050977).

참 고 문 헌

- [1] Goodfellow, Ian, et al. "Generative adversarial networks." *Communications of the ACM* 63.11 (2020): 139-144.
- [2] 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.
- [3] Lee, Minhyeok, and Junhee Seok. "Controllable generative adversarial network." *Ieee Access* 7 (2019): 28158-28169.
- [4] Choi, Yunjey, et al. "Stargan v2: Diverse image synthesis for multiple domains." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.