도메인 변환을 위한 DoubleRoundTrip 모델

김인수
전자전기공학부
고려대학교
서울특별시 성북구 안암로 145
dlstn0910@korea.ac.kr

태동현
전자전기공학부
고려대학교
서울특별시 성북구 안암로 145
hyunnibal@korea.ac.kr

석준희*
전자전기공학부
고려대학교
서울특별시 성북구 안암로 145
jseok14@korea.ac.kr

요 약

딥러닝 기술의 발전에 따라 개발된 적대적 인공 생성 신경망 (GAN)은 여러 분야에서 활용되고 있다. 특히 다양한 분야에서 활용될 수 있는 도메인 변환에 특화된 순환 적대적 인공 생성 신경망 (CycleGAN)의 개발 이후 이미지 변환 문제에서 GAN 은 훌륭한 성능을 선보였다. 다만, 기존의 CycleGAN 모델은 학습이 불안정하다는 점과 더불어 제대로 변환되지 않은 이미지가 다수 존재한다는 한계를 가지고 있다. 본 논문에서는 이러한 CycleGAN 의 한계를 개선하기 위해 가변 오토인코더(VAE)를 GAN 구조에 이미지 변환 모델인 이중 양방향 생성 모델 (DoubleRoundTrip)을 제시하고 모델의 성능을 견본 이미지 데이터셋에서 확인하였다.

키워드: 적대적 인공 생성 신경망, 순환 적대적 인공 생성 신경망, 가변 오토인코더, 이중 양방향 생성 모델

1. 서론

도메인 변환을 이용하는 이미지 변환 연구는 인공지능 분야, 특히 컴퓨터 비전 분야에서 주요하게 연구되던 항목 중 하나이다.[1] 최근 적대적 인공 생성 신경망 (GAN)의 개발 이후 이러한 GAN 을 기반으로 하는 다양한 도메인 변환 모델들이 제시되었다.[2]

이러한 GAN 기반 도메인 변환 모델들 중 두 쌍의 GAN 모델을 연결하여 학습하는 순환 적대적 인공 생성 신경망 (CycleGAN)은 도메인 변환 문제를 해결하는데 있어서 훌륭한 성능을 보이는 것이 확인되었다.[3] 다만, 이러한 CycleGAN 모델은 GAN 모델이 지니는 고질적인 문제인 불안정한 학습의 문제와 도메인 변환에 있어서 제대로 된 도메인에 매핑이 되지 않는 경우가 다수 존재한다는 한계로부터 자유롭지 못하다.

본 논문에서는 이러한 CycleGAN 의 한계를 해결하기 위해 변환하고자 하는 다른 도메인들의 이미지들이 갖는 공통 공간을 정의하고 이러한 공통 공간을 가변 오토인코더(VAE)를 통해 거치는 도메인 변환 모델인 이중 양방향 생성 모델 (DoubleRoundTrip)을 제시한다. [4]

2. 관련연구

2.1 적대적 인공 생성 신경망 (GAN)

적대적 인공 생성 신경망 (GAN)은 두 개의 인공 신경망이 서로 결합하여 학습하는 구조의 딥러닝 모델이다. 생성기와 판별기로 명명된 두 적대적 인공 신경망 간의 미니맥스 게임을 통하여 GAN 모델은 학습된다. 생성기는 무작위의 노이즈로부터 실제의 학습 데이터와 분간할 수 없는 그럴듯한 샘플을 생성하는 방향으로 학습하고 판별기는 생성기가 생성한 샘플과 실제 데이터 샘플을 판별하는 방향으로 학습한다. 이러한 두 인공 신경망들의 경쟁적 학습은 아래의 손실 함수 1 과 2 를 통해 구현된다.

$$L_D = -\left(\log\left(D(x)\right) + \log\left(1 - D\left(G(z)\right)\right)\right) \quad (1)$$

$$L_G = -\log\left(D\left(G(z)\right)\right) \quad (2)$$

그러나, 원본 데이터를 다른 도메인의 데이터로 변환하는 도메인 변환 문제에 노이즈로부터 그럴듯한 가짜를 생성하도록 설계된 GAN 모델의 손실함수를 그대로 적용할 수는 없다.

2.2 순환 적대적 인공 생성 신경망 (CycleGAN)

순환 적대적 인공 생성 신경망 (CycleGAN)은 이미지의 도메인 변환에 특화된 GAN 모델로서 순환 안정성 손실함수 (Cycle Consistency Loss)와 일치 손실함수 (Identity Loss) 를 통해 이미지의 도메인 변환을 훌륭히 수행하는 GAN 모델을 구현하였다. CycleGAN 모델의 도메인 변환을 위한 학습은 아래 손실함수 3 과 4 를 통해 구현된다.

$$L_{Cycle}(G,F) = E_{x\ p_{data}(x)}[\left|F\left(G(x)\right) - x\right|¿¿1] + E_{y\ p_{data}(y)}[\left|G(F\right.$$

(3)

$$L_{Identity}(G,F) = E_{x\ p_{data}(x)}[\left|F(x) - x\right|¿¿1] + E_{y\ p_{data}(y)}[\left|G(y)\right.$$

* 교신저자

(4)

손실함수 3과 4에 손실함수 1과 2를 조합한 손실함수를 통해 CycleGAN은 적대적 학습을 수행하여 이미지의 도메인 변환을 구현한다.

3. 실험방법

본 논문에서는 CycleGAN의 학습 불안정성과 원하지 않는 다른 이미지로 변환되는 경우가 잦다는 한계를 극복하기 위해 VAE를 생성기 구조에 추가한 DoubleRoundTrip 모델을 제시한다.

도메인 변환이란 특정 도메인의 데이터를 다른 도메인의 데이터로 변환하는 연구를 일컫는다. 이러한 도메인 변환 문제에서 다른 도메인으로 변환된 데이터와 원본 데이터 사이에는 공통된 특성이 존재하는 것을 확인할 수 있다.

DoubleRoundTrip 모델에서는 변환된 이미지 사이의 공통된 특성이 존재하는 점에 주목하여 공통 잠재 공간을 생성기에 추가하는 것으로 기존의 CycleGAN의 문제점을 보완한다. 이러한 생성기의 공통 잠재 공간은 VAE 구조를 통해 실현된다. 이러한 DoubleRoundTrip 모델의 구조는 아래의 그림 1과 같다.

그림 1의 X, Y는 각각 데이터들이 존재하는 도메인을 의미하며 G_X, G_Y 는 도메인 변환을 위한 생성기를 의미한다. D_X, D_Y, D_Z 는 생성기와 적대적 학습을 하는 판별기를 의미한다. Z는 도메인 X와 Y 사이의 잠재공간을 의미하며 이렇게 도메인 간 공통의 잠재공간은 H_X, H_Y 라는 VAE 구조를 통해 구현된다.

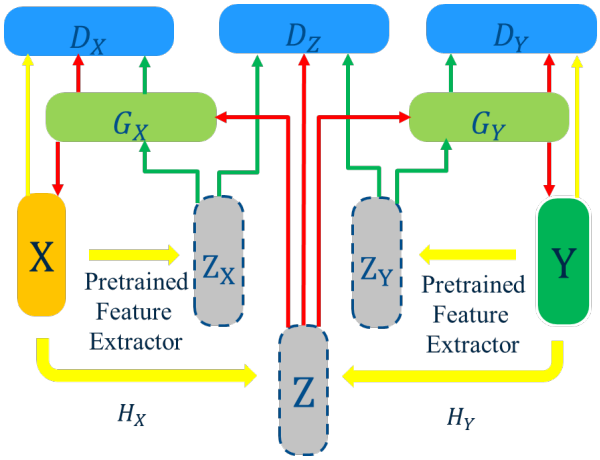

그림 1. DoubleRoundTrip 모델의 구조도

이러한 DoubleRoundTrip 모델의 학습은 손실함수 5를 통해 구현된다.

$$L_{DRT}=\alpha_1 E_{x\ p_x}\left|x-G_X\left(H_X(x)\right)\right|+\alpha_2 E_{(x,y)\ (p_x,p_y)}\left|x-G_Y\left(H_Y(y\right)\right.$$

(5)

4. 실험결과

본 실험은 앞서 제시한 DoubleRoundTrip 구조를 CycleGAN 학습에 사용된 사진-모네 그림 데이터와 말-얼룩말 데이터를 통해 학습하는 방향으로 진행되었다. ADAM 옵티마이저와 WGAN Loss를 사용하여 학습하였다.

학습한 결과를 따라 도메인 변환을 한 이미지는 그림 2를 통해 확인할 수 있다. VAE를 도입한 DoubleRoundTrip 모델의 경우 도메인 변환에 있어서 어느정도 성능을 보임을 확인할 수 있었다.

(a) (b) (c)

그림 2. DoubleRoundTrip 모델을 통해 도메인 변환을 한 결과.

그림 2의 (a), (b), (c)는 각각 원본 사진, 원본 사진 도메인으로 VAE를 통해 복구한 결과 타겟 고흐 도메인으로 이미지를 변환한 결과이다.

5. 결론

본 논문에서는 CycleGAN의 한계를 극복하기 위해 VAE 구조를 생성기에 도입한 DoubleRoundTrip 모델을 제시한다. 본 연구진이 제시하는 DoubleRoundTrip 모델은 DRT 손실함수를 도입하는 것으로도메인 간 공통 잠재 공간을 생성기가 포착할 수 있도록 구현하였다. 다만, 실제 학습에서 학습 불안정성이 크게 개선되지 않았기에 향후 연구를 통해 개선을 꾀하는 것이 바람직할 것으로 사료된다.

Acknowledgement

본 논문은 2023년도 한국연구재단(NRF-2022R1A2C2004003)과 산업통상자원부의 재원으로 한국산업기술진흥원(P0012725)의 지원을 받아 수행된 연구임.

참 고 문 헌

[1] Karras, Tero, Samuli Laine, and Timo Aila. "A style-based generator architecture for generative adversarial networks." *Proceedings of the IEEE/CVF conference on*

computer vision and pattern recognition. 2019.

[2] Goodfellow, Ian, et al. "Generative adversarial networks." *Communications of the ACM* 63.11 (2020): 139-144.

[3] Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." *Proceedings of the IEEE international conference on computer vision*. 2017.

[4] Kingma, Diederik P., and Max Welling. "Auto-encoding variational bayes." *arXiv preprint arXiv:1312.6114* (2013).