생성적 적대 신경망 기반 Image-to-Image translation 기술 연구 동향

황도연, 김재석, 최윤호
부산대학교
h1d2y3@pusan.ac.kr, tjr3117@pusan.ac.kr, yhchoi@pusan.ac.kr

Research trend of Image-to-Image translation technology based on generative adversarial neural network

Hwang Do-Yeon, Kim Jae-Seok, Choi Yoon-Ho
Pusan National Univ.

요 약

인공지능 기술이 발달하고 다양한 분야에 적용됨에 따라 특정 도메인(domain)을 가진 이미지를 다른 도메인을 가진 이미지로 변환하는 Image-to-Image translation 기술이 발달하였다. 특히 생성적 적대 신경망(generative adversarial network, GAN)이 적용된 Image-to-Image translation 모델들은 성능 지표에서 뛰어난 향상을 보여주었으며 이를 통해 사용자들에게 편의를 제공하였다. 본 논문은 GAN을 적용한 Image-to-Image translation 모델들을 조사하였으며 해당 모델들의 목표와 작동 방식을 분석하였다.

Ⅰ. 서 론

Image-to-Image translation은 특정 도메인(domain)을 가진 입력 이미지를 목표 도메인을 가진 출력 이미지로 변환하는 것을 목표로 하는 분야이다. 최근 인공지능 기술이 발전하고 인공지능 모델들이 Image-to-Image translation 분야에 적용됨에 따라 해당 분야에서도 뚜렷한 성능 개선이 이루어지고 있다. 특히 인공지능 모델 중 하나인 생성적 적대 신경망(generative adversarial network, GAN)은 경쟁적 학습을 통해 양질의 이미지를 생성할 수 있어 이미지를 출력해야 하는 Image-to-Image translation 분야에서 적극적으로 사용되고 있다. 본 논문은 현재 뛰어난 성능을 보여주고 있는 GAN 기반 Image-to-Image translation 모델들의 목표 및 작동방식을 분석하였다.

본 논문의 구성은 다음과 같다. 2장에서는 GAN 기반 Image-to-Image translation 모델들을 간략하게 소개하고 해당 모델의 목표 및 작동 방식을 설명한다. 3장에서는 해당 내용을 바탕으로 본 논문의 결론을 맺는다.

Ⅱ. 본론

GAN이 적용된 Image-to-Image translation 모델들은 generator와 discriminator의 경쟁적 학습을 통한 입력 이미지와 출력 이미지 간 매핑 함수 학습을 목표로 한다. 본 장에서는 GAN의 학습과정을 바탕으로 각 모델들의 목표 및 작동 방식에 대해 분석하였다.

2.1 Pix2Pix

Convolutional neural network(CNN)를 사용한 기존 Image-to-Image translation 모델들은 출력 이미지의 평균을 구하는 방식으로 학습되어 흐릿한 출력 이미지를 생성하는 한계를 가지고 있다. 이러한 문제를 해결하기 위해 Isola 등은 GAN을 통해 Image-to-Image translation을 수행하는 pix2pix를 제안하였다[1]. Pix2pix는 generator와 discriminator로 구성되어 있으며 generator는 입력 이미지를 변환하고 discriminator는 이미지에 대해 실제/거짓 여부를 판단한다. Generator는 discriminator를 속이는 것을 목표로 하며 discriminator는 이미지를 정확히 감별하는 것을 목표로 한다. 이러한 두 모듈의 경쟁적 학습은 generator가 실제 이미지와 유사하게 이미지를 변환할 수 있도록 하며 이를 통해 출력 이미지가 흐릿하게 생성되는 문제를 해결한다. 또한 Pix2pix는 conditional GAN을 사용하여 여러 개로 세분화된 Image-to-Image translation의 다양한 문제들을 하나의 모델로 해결할 수 있다는 장점을 가지고 있다.

2.2 CycleGAN

Pix2pix는 비교적 선명한 출력 이미지를 생성할 수 있지만 학습을 위해서 paired data가 필요하다는 단점이 있다. 이러한 문제를 해결하기 위해 Zhu 등은 unpaired data로도 학습이 가능한 CycleGAN을 제안하였다[2]. Unpaired data는 ground truth가 없으므로 경쟁적 학습만을 사용하여 이미지 변환을 학습할 경우, discriminator는 속일 수 있지만 입력 이미지와 무관한 이미지로 변환되는 mode collapse가 발생할 수 있다. 이러한 문제를 해결하기 위해 CycleGAN은 두 개의 generator-discriminator 쌍으로 구성된 네트워크에서 cycle consistency loss를 적용하여 학습을 수행한다. 먼저, generator G는 도메인 X에 해당하는 실제 이미지 x를 도메인 Y의 분포를 따르는 $G(x)$로 변환한다. 반대로, generator F는 도메인 Y에 해당하는 이미지 y를 도메인 X의 분포를 따르는 $F(y)$로 변환한다. 이때, G와 F는 서로 역함수이므로 $G(x)$는 F를 통해 x로 복원될 수 있다. Cycle consistency loss는 복원된 $F(G(x))$와 x의 차이를 통해 계산되며 이를 통해 CycleGAN은 두 이미지가 최대한 유사하도록 학습된다. 두 이미지가 유사하기 위해서는 F로부터 x가 복원될 수 있도록 $G(x)$가 x의 특성을 보존하고 있어야 한다. 따라서 cycle consistency loss를 통해 학습이 수행될 경우, G는 Y의 분포를 따르는 동시에 x의 특성을 보존하는 변환을 수행할 수 있도록 학습된다.

2.3 UNIT

확률적 모델링 관점에서 unpaired data를 이용한 Image-to-Image translation 학습은 각 주변분포(marginal distribution)를 따르는 두 개의 이미지 집합을 사용하여 결합분포(joint distribution)를 추론하는 것으로 볼 수 있다. Liu 등은 해당 문제를 해결하기 위해서는 결합분포에 대한 추가적인 가정이 필요함을 지적하며 특수한 가정을 바탕으로 이미지를 생성하는 UNIT을 제안하였다[3]. 구체적으로, UNIT은 대응되는 한 쌍의 이미지가 shared-latent space에서 동일한 latent representation으로 사상될 수 있다는 가정을 사용한다. 해당 가정이 적용된 이미지 변환을 수행하기 위해 UNIT은 두 개의 generator-discriminator 쌍에 두 개의 encoder E_1, E_2를 추가하여 latent representation을 찾을 수 있도록 하였다. 이러한 구조에서 입력 이미지 x는 E_1을 통해 latent code인 z로 변환되며 generator G_1은 z를 입력으로 출력 이미지를 생성한다. 여기서 E_1-E_2, G_1-G_2는 각각 대칭되는 모듈과 일부 레이어의 가중치(weight)를 공유하며, 이를 통해 앞선 가정을 충족하는 이미지 변환이 수행될 수 있도록 한다.

2.4 AGGAN

CycleGAN은 unpaired data를 이용한 Image-to-Image translation을 가능하게 했으나 이미지 변환을 수행할 경우, 변환의 대상이 되는 전경(foreground) 뿐만 아니라 배경(background) 또한 변환되는 문제가 있다. 구체적으로 CycleGAN은 이미지 전체에 대한 기존(source) 도메인 이미지의 분포와 목표(target) 도메인 이미지의 분포 간의 차이를 줄이는 데만 집중하여 이미지 내부에서 변환의 목표가 되는 지역을 파악하기 힘들다는 한계를 가진다. 이러한 문제를 해결하기 위해 Mejatti 등은 attention 모듈을 적용한 Image-to-Image translation 모델인 AGGAN을 제안하였다[4]. AGGAN은 CycleGAN과 같은 두 개의 generator-discriminator 쌍에 두 개의 attention 모듈인 A_s, A_t를 추가하였으며 G를 통해 x를 $G(x)$로 변환하고 A_s를 통해 x에 대한 attention map인 $A_s(x)$를 생성한다. 최종 이미지 x'는 다음과 같은 방법을 통해 생성된다.

$$x' = A_s(x) \odot G(x) + (1 - A_s(x)) \odot x \quad (1)$$

(1)에서 $\odot$는 성분별 곱(element-wise product)을 의미하며 $A_s(x) \odot G(x)$는 이미지의 전경을 의미하고 $(1 - A_s(x)) \odot x$는 이미지의 배경을 의미한다. AGGAN은 attention 모듈을 통한 전경과 배경을 구분하는 이미지 변환을 수행함으로써 우수한 품질의 출력 이미지를 생성할 수 있도록 하였다.

2.6 StarGAN

언급된 Image-to-Image translation 모델들은 두 개의 도메인에 대해서만 이미지 변환이 가능하다는 점에서 확장성이 부족하였다. Choi 등은 이러한 문제를 해결하기 위해 하나의 모델로도 다양한 도메인에 대한 이미지 변환이 가능한 StarGAN을 제안하였다[5]. StarGAN은 기존 모델과는 다르게 generator에 입력 이미지뿐만 아니라 목표 도메인 레이블도 함께 입력으로 사용된다. Discriminator는 주어진 이미지에 대해 실제/거짓 여부뿐만 아니라 해당 이미지의 도메인 또한 판단한다. 새롭게 추가된 generator와 discriminator의 역할을 학습하기 위해 StarGAN은 다음과 같은 domain classification loss를 손실 함수로서 추가하였다.

$$L_{cls}^{r} = E_{x,c'}[-\log D_{cls}(c'|x)] \quad (2)$$

(2)는 domain classification loss에서 discriminator를 학습시키기 위한 부분이다. StarGAN의 discriminator는 실제 입력 이미지 x가 주어질 경우 해당 이미지에 대응하는 도메인인 c'을 정확히 예측하는 것을 목표로 학습된다.

$$L_{cls}^{f} = E_{x,c}[-\log D_{cls}(c|G(x,c))] \quad (3)$$

(3)은 domain classification loss에서 generator를 학습시키기 위한 부분이다. x를 입력으로 generator를 통해 c 도메인으로 변환된 이미지 $G(x,c)$가 주어질 경우, 해당 이미지는 discriminator에 의해 c로 분류되는 것을 목표로 학습된다. StarGAN은 이러한 학습과정을 통하여 단일 모델이 입력 이미지를 다양한 도메인의 이미지로 변환할 수 있도록 하였다.

Ⅲ. 결론

본 논문은 GAN을 사용한 Image-to-Image translation 모델들인 pix2pix, CycleGAN, UNIT, AGGAN, StarGAN을 분석하였다. 각 모델들은 발전과정에 따라 새로운 모듈과 학습방식을 제안하였으며 이를 통해 기존 모델들의 문제를 해결하고 성능을 개선하였다. 추후 이루어질 Image-to-Image translation 연구에서도 기존 모델의 한계를 진단하고 해당 문제를 해결하기 위한 합리적인 방법론이 제안되어야 할 것이다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원(2019-0-01343,융합보안핵심인재양성사업) 및 정보통신기획평가원의 대학ICT연구센터육성지원사업(IITP-2021-2020-0-01797)에 의하여 지원되었음.

참 고 문 헌

[1] ISOLA, P., Zhu, J. Y., Zhou, T., and Efros, A. A. “Image-to-image translation with conditional adversarial network,”. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125-1134, 2017.

[2] Zhu, J. Y., Park, T., Isola, P., and Efros, A. A. “Unpaired image-to-image translation using cycle-consistent adversarial networks,” In: Proceedings of the IEEE international conference on computer vision. pp. 2223-2232, 2017.

[3] Liu, M. Y., Breuel, T., and Kautz, J. “Unsupervised image-to-image translation networks. In: Advances in neural information processing systems,” pp. 700-708. 2017.

[4] Alami Mejjati, Y., Richardt, C., Tompkin, J., Cosker, D., and Kim, K. I. “Unsupervised Attention-guided Image-to-Image Translation,” Advances in Neural Information Processing Systems, pp. 3693-3703, 2018.

[5] Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., and Choo, J. “Stargan: Unified generative adversarial networks for multi-domain image-to-image translation,” In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 8789-8797, 2018.