

적층 오토인코더 기반의 판별자 활용을 통한 적대적 생성 신경망 성능 향상 기법

조영채¹, 손봉기², 이재호*
덕성여자대학교^(1,*), 서원대학교²

¹20220691@duksung.ac.kr, ²bksohn@seowon.ac.kr, *izeho@duksung.ac.kr

Performance Improvement method on GAN exploiting Stacked Autoencoder-based Discriminator

Youngchae Cho¹, Bong-Ki Son², Jaeho Lee*
Duksung Women`s Univ.^(1,*), Seowon Univ.²

요약

오늘날 인공지능 이미지 처리 분야에서 생성 모델은 많은 관심을 받고 있고, 다양한 분야에서 발전을 이루고 있다. 그 중 가장 기본적인 생성모델인 GAN 은 많이 사용되고 있다. 하지만 제일 간단한 형태의 GAN 은 수준 높은 이미지의 생성에 한계가 있기에, GAN 에 사용하는 모델 중 하나인 판별자에 오토인코더를 사용하였다. 또한 오토인코더도 단층보다 여러 개를 쌓은 적층 오토인코더를 사용하여 더욱 품질이 좋은 이미지를 수행할 수 있게 된다.

I. 서론

최근 인공지능 생성 모델 중 이미지 생성 모델은 많은 분야에서 사용되고 있고 활발히 연구되고 있다. 그중 두가지 모델이 서로 경쟁하며 학습하는 모델인 GAN(Generative Adversarial Networks), 적대적 생성 신경망은 기존의 신경망과는 차별화된 구조로 평가받으며 생성 모델의 역사에 있어서 많은 의미를 지니고 있다.

물론 GAN 은 뛰어난 모델이지만 가장 기본적인 형태인 vanilla GAN 은 컬러 사진에서는 좋은 결과를 얻기 힘들어서 흑백 사진을 사용해야 학습이 수월하고 학습이 완료되는데 시간이 매우 오래 걸린다는 단점이 있다. 따라서 vanilla GAN 이 아닌 적층 오토인코더라는 모델을 GAN 에 적용하여 위의 한계를 보완할 것이다.

본논문에서는 GAN 의 판별자(Discriminator)부분에 기존의 CNN 만을 사용하지 않고 적층 오토인코더(Stacked Autoencoder)와 CNN(Convolution Neural Network)을 사용하여 이미지를 생성하는 모델을 제시한다. 이를 통해 컬러 이미지의 생성을 가능하게 하고 더욱 높은 품질의 이미지를 생성할 수 있다.

II. 연구배경

2.1 GAN

적대적 생성 신경망을 뜻하는 GAN 은 사용자가 원하는 이미지를 생성하는 데에 효과적인 모델이다. 이미지, 영상, 텍스트를 생성하는 등의 다양하게 응용되고 있다. 기존의 딥러닝 기술은 다층의 인공 신경망 하나를 학습시키는 방법을 활용하였지만, GAN 은

2 개의 인공 신경망의 상호작용을 활용하여 최종적으로는 인공 이미지를 생성하는 1 개의 생성 신경망을 실제로 사용한다.^[1]

위에 기술했던 2 개의 인공 신경망은 각각 판별자(Discriminator)와 생성자(Generator)를 의미한다. 이 두 네트워크가 적대적으로 학습하며 실제와 비슷한 데이터를 생성해내는 것이다. 여기서 판별자를 경찰, 생성자를 화폐 위조자에 비유할 수 있다. 생성자는 위조 화폐를 계속 생성하고, 판별자는 위조화폐와 진짜 화폐를 구분하며 두 모델 모두 학습되는 것이다. 여기서 생성자는 랜덤 노이즈를 신경망에 넣어 모방 데이터를 만들고, 판별자는 생성자가 만든 데이터와 진짜 데이터를 신경망에 넣어 구분한다.^[2] 이렇게 서로 적대적으로 학습하면 실제 이미지 데이터의 분포와 유사한 새로운 데이터를 생성할 수 있다.

2.2 Autoencoder & CNN

오토인코더(Autoencoder)란 비지도학습 모델로, 입력 데이터를 압축시킨 후 데이터의 특징을 추출하고 학습하여 본래의 입력 형태로 복원시키는 신경망을 의미한다. 여기서 데이터를 압축하는 모델은 인코더(Encoder), 원래 형태로 복원하는 모델은 디코더(Decoder)라 한다. 인코더는 매니폴드(manifold) 학습 방법을 이용하여 고차원 공간의 데이터를 저차원 manifold 공간으로 맵핑할 수 있도록 한다. 디코더는 인코딩 후 생성된 잠재 변수를 실제 데이터 분포와 연결할 수 있도록 한다.^[3] 최종적으로 입력 데이터와 복원 결과물의 차이가 최소화되도록 학습한다.

다음으로 CNN 은 합성곱 신경망을 뜻한다. CNN 은 필터를 통해 이미지에 합성곱연산을 통해 특징을 추출하여 Feature Map 을 학습한다.^[4]

이 논문에서는 앞서 언급했던 오토인코더와 CNN 을 결합해 인공 데이터와 진짜 데이터를 분류하는 데에 사용된다.

2.3 Stacked Autoencoder

적층 오토인코더, Stacked Autoencoder(SAE)란 위에서 언급했던 오토인코더 모델을 여러 개의 층으로 쌓은 구조로 이루어져 있다. 한 층이 하나의 오토인코더로 이루어져 있고, 각 오토인코더에서 복원한 이미지를 다음 오토인코더의 입력층에 주어 반복적으로 학습하며 입력 데이터와 출력 데이터의 차이를 최소화하는 방향으로 학습이 이루어진다.^[5]

이 논문에서는 5 개의 오토인코더를 쌓아 적층 오토인코더 모델을 구성하였다. 따라서 마지막 오토인코더의 인코더에서 추출한 특징을 CNN 에 입력 데이터로 넣어서 학습할 수 있도록 하였다.

III. 모델 설계

3.1 판별자 모델 구성

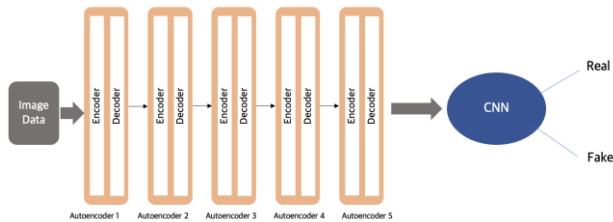


그림 1. 판별자 모델의 구성

기본적으로 GAN 모델은 판별자와 생성자, 이 두 모델로 구성 되어있다. 먼저 판별자는 그림 1 처럼 5 개로 연결된 적층 오토인코더와 CNN 을 결합하였다. 오토인코더를 여러 개 쌓음으로써 이미지의 특징을 더 정교하게 학습하여 더 정확히 복원된 이미지를 얻을 수 있고, 그냥 단순한 CNN 만을 사용했던 vanilla GAN 에 비해 3 개의 채널을 가진 컬러 사진의 사용이 용이해지고 더 많은 양의 데이터를 다룰 수 있게 된다. 이렇게 복원된 이미지를 CNN 에게 넘겨준다.

판별자의 마지막 단계에 있는 CNN 은 적층 오토인코더에 의해 반복적으로 압축되고 다시 복원된 이미지를 Fake 와 Real 을 구분한다. 그렇게 CNN 은 이미지 자체가 아닌 이미지 특징정보까지 반영하여 이미지를 이진분류하기 때문에 더욱 정확하게 분류할 수 있다.

각 오토인코더당 300 epochs 씩, CNN 은 400 epochs 를 pre-training 하여 모델로 저장했다.

3.2 GAN 모델 구성

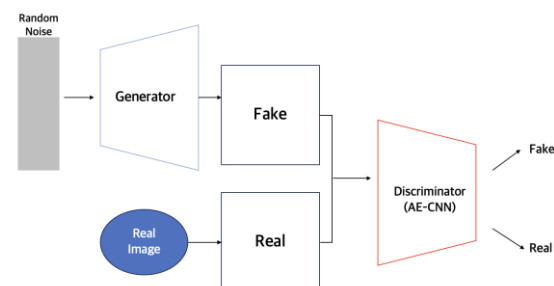


그림 2. GAN 모델의 구성

위의 생성자와 판별자로 GAN 모델을 구성하면 그림 2 와 같은 모델이 만들어진다. 최종적으로 랜덤 노이즈를 입력으로 받는 생성자는 인공 이미지를 생성해내고, 판별자는 생성자가 만든 인공 이미지와 진짜 이미지를 입력으로 받아 이 둘을 Fake/Real 로 구분한다.

생성자는 기본적인 CNN 모델로 구성되어 있다. 처음은 100 개의 랜덤 노이즈 벡터를 입력으로 한다. 그리고 LeakyReLu 활성화 함수와 배치 정규화 계층을 반복적으로 거치면서 실제 이미지의 크기에 맞도록 업샘플링하다가 마지막에 3 개의 필터를 가진 RGB 이미지를 생성한다.

그렇게 판별자는 loss 값을 계산하고 이를 바탕으로 생성자가 학습하여 실제 데이터와 유사한 이미지 데이터를 생성할 수 있게 된다. 이 GAN 모델은 총 25,000 epochs 를 학습하였다. 그리고 이렇게 학습된 GAN 모델에서 생성자 모델이 생성한 이미지를 확인하여 결과를 분석하였다.

3.3 Dataset

데이터는 이미지 생성 모델만큼 jpg 형식의 이미지를 사용하였다. 이미지에 있는 얼굴과 비슷한 이미지의 생성을 위해 본인의 얼굴이 중심으로 있는 사진을 사용하였다. 학습 시간과 차원을 조절할 때의 픽셀 수 계산을 고려해 이미지의 크기는 모두 200x200 으로 통일하였다. 적층 오토인코더와 CNN 으로 이루어져 있는 판별자를 pre-training 할 때에는 이미지 500 장을 사용하였다. 그리고 GAN 에 입력으로 넣는 Real 이미지는 이미지 200 장을 사용하여 학습을 진행했다.

3.4 결과

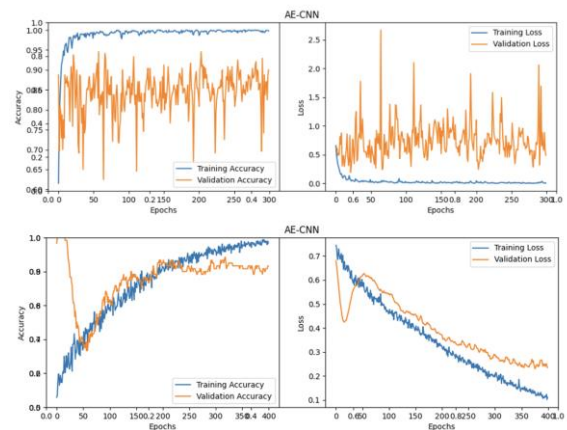


그림 3. 단층 오토인코더와 적층 오토인코더의 결과

먼저 판별자 모델을 pre-training 하였을 때, 그 결과는 그림 3 과 같다. 위의 표는 단층 오토인코더의 정확도와 loss, 아래는 적층 오토인코더의 정확도와 loss 를 나타낸다. 단층보다 오토인코더 여러 개를 쌓은 적층 오토인코더를 사용하였을 때 정확도가 일정하게 상승하고 loss 값이 일정하게 하강하는 것을 알 수 있다.



그림 4. vanilla GAN 과 적층 오토인코더를 사용한 GAN 의 결과

다음으로는 이렇게 학습한 판별자를 이용해 GAN 을 학습한 후 생성자가 생성한 이미지를 그림 4 를 통해 볼 수 있다. 위 이미지들은 모두 GAN 의 학습과정 중 초기 부분학습 이미지 생성 결과를 표시한 것이다. 왼쪽의 흑백 이미지들은 GAN 의 가장 기본 형태인 vanilla GAN 을 통해 생성한 이미지들이고, 오른쪽의 컬러 이미지들은 계속 언급해왔던 적층 오토인코더를 판별자에 적용한 GAN 의 생성자가 생성한 이미지들이다. vanilla GAN 은 학습 속도가 느려 컬러 이미지를 생성하는 데에는 한계가 있었다. 또한 결과물에서 얼굴의 형태를 찾기 어려웠다. 하지만 앞서 구성했던 GAN 모델은 완벽한 얼굴은 아니더라도 얼굴의 형태와 이목구비의 위치정도는 구분이 가능했다. 또한 머리카락의 색, 얼굴의 색, 입술의 색이 모두 확인이 가능해서 컬러 이미지도 잘 생성한다는 점을 알 수 있었다. 이는 충분한 학습시간을 투입할 경우 더욱 높은 품질의 이미지를 생성할 수 있을 것으로 예상된다.

IV. 결론

본 논문에서는 단층 오토인코더와 적층 오토인코더의 차이, 또 vanilla GAN 과 적층 오토인코더를 적용한 GAN 의 차이를 분석했다. GAN 의 판별자 모델을 구성할 때 단순한 CNN 만을 사용하기보다 오토인코더를 사용하는 것이 좋았고, 오토인코더도 단층보다 적층 오토인코더를 사용하는 것이 이미지의 품질을 더 향상시킬 수 있었다.

이러한 분석을 통해 수행한 실험으로 나의 얼굴을 생성할 때 흑백에 형태를 알아보기 어려웠던 사진에 비해 색과 형태가 구분 가능할 수 있을 정도로 생성이 가능했다.

만약 에포크 수를 더 늘린다면 더욱 정교한 얼굴 이미지를 생성할 수 있을 것이라는 기대를 할 수 있었고, 이 모델을 응용해 얼굴 이미지 뿐만 아니라 사용자가 원하는 다양한 이미지를 생성하는 것이 가능하다고 본다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(과제번호-2022R1A2C1009951).

참 고 문 헌

- [1] 조영주(Jo Y.J.);배강민(Bae K.M.);박종열(Park J.Y.), "GAN 적대적 생성 신경망과 이미지 생성 및 변환 기술 동향," 전자통신동향분석, 35 권, 4 호, 91-102, 2020, 10.22648/ETRI.2020.J.350409.
- [2] Goodfellow, Ian, et al. "Generative adversarial nets." Advances in neural information processing systems 27 (2014).
- [3] Bank, Dor, Noam Koenigstein, and Raja Giryes. "Autoencoders." Machine learning for data science handbook: data mining and knowledge discovery handbook (2023): 353-374.
- [4] LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324.
- [5] 노석범, 오성권. (2015). "Stacked Autoencoder 를 이용한 특징 추출 기반 Fuzzy k-Nearest Neighbors 패턴 분류기 설계." 전기학회논문지, 64(1), 113-120.