

심층 적대적 생성 신경망의 오류 재학습을 이용한 얼굴 영상 생성 모델

하용욱 · 홍동진 · 차의영

부산대학교

Photo-realistic Face Image Generation by DCGAN with error relearning

Yong-Wook Ha · Dong-jin Hong · Eui-Young Cha

Pusan National University

E-mail : ywha12@gmail.com / eycha@pusan.ac.kr

요 약

본 논문에서는 계층형 적대적 생성 신경망(GAN: Generative Adversarial Network)에서 오류 판별자를 추가하여 영상 생성 성능을 개선하는 방안을 제안한다. 제안하는 영상 생성 방법에서는 영상 생성자가 빈번히 발생시키는 오류에 대해 별도로 학습을 수행하는 판별자를 모델에 추가하여 계층형 적대적 생성 신경망을 구성하였다. 본 논문에서 제안한 모델을 이용하여 생성한 영상의 효용성을 검증하는 방법으로는 Inception Score를 사용하였다. 학습 데이터로 celebA의 유명한 얼굴 이미지 중 정면 이미지 155,680장을 이용하였다. 본 논문의 모델로 생성한 10,000장의 얼굴 이미지를 Inception Score로 평가한 결과, 평균 1.742p의 성능을 나타내어 기존의 영상 생성 방법보다 높은 점수를 얻을 수 있었으며, 효용성을 확인할 수 있었다.

ABSTRACT

In this paper, We suggest a face image generating GAN model which is improved by an additive discriminator. This discriminator is trained to be specialized in preventing frequent mistake of generator. To verify the model suggested, we used *Inception score. We used 155,680 images of *celebA which is frontal face. We earned average 1.742p at Inception score and it is much better score compare to previous model.

키워드

AI, DCGAN, Image Generation, Face Image

I. 서 론

최근 인공지능 분야는 GPU의 비약적인 발전, 심층 학습(Deep Learning)과 빅데이터 같은 혁신 기술들로 인해 활발히 연구되고 있다. 특히 심층 학습은 영상, 음성, 자연어 등의 난해한 분야에서 뛰어난 성능을 보여주며 각광받고 있다. 최근에는 심층 학습이 사람의 정확성을 뛰어넘는 분류기에 사용될 뿐만 아니라 새로운 영상을 만들어내는 생성 모델의 영역에서도 활발히 연구되고 있다. 생성 모델 중 대표적으로 적대적 생성 신경망(Generative Adversarial Network)[1]으로 불리는 이 모델은 충분한 양의 데이터를 학습하면, 학습 데이터에 포함되지 않는 새로운 결과물을 생성한다. 하지만 창작 영역의 복잡성 때문에 이 모델이 내놓는 결과물은

신뢰성이 다소 떨어지는 경향이 있다. 본 논문에서는 생성 모델의 결과물 중, 유사한 실패물이 반복되는 경우, 이 경우를 학습하는 구별자를 따로 두어 더 좋은 영상을 생성할 수 있는 모델을 제안한다.

II. 심층 적대적 생성 신경망 구현

적대적 생성 신경망은 대표적인 생성 모델의 간접 학습방법이다. 주로 영상 생성과 해상도 개선 등의 분야에서 각광을 받고 있다.

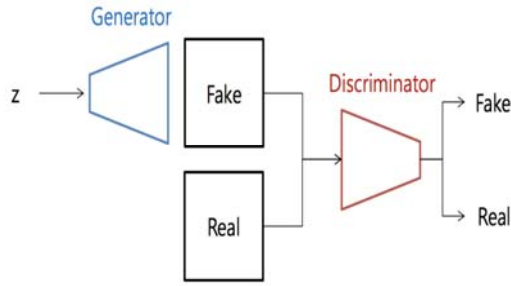


그림 1. 적대적 생성 신경망 구조

인공 신경망 중 하나인 컨볼루션 뉴럴 네트워크(CNN: Convolutional Neural Network)[2]는 인간의 뇌에서 일어나는 뉴런들의 상호작용을 모방하여 설계된 모형이다. 영상 분류와 생성 같이 통계적으로 많은 수의 가능성을 가진 모델의 학습에 있어 효과적이다.

심층 적대적 생성 신경망은 적대적 생성 신경망 구조(그림 2.)에서 생성자(Generator)와 구별자(Discriminator)에 여러 컨볼루션 층(Convolution Layer)을 두는 방법으로, 생성자의 성능에 큰 발전을 가져왔다.

본 논문이 제시하는 모델에서 생성자는 5X5의 필터 크기(kernel size)와 2의 보폭(stride)을 가지는 4개의 상승 컨볼루션 계층과 3X3의 필터 크기와 1의 보폭을 가지는 1개의 완전 연결계층으로 이루어져 있다. 마지막을 제외한 모든 상승 컨볼루션 층에 배치 정규화(Batch Normalization)[5]를 적용하였으며 모든 상승 컨볼루션 층에는 활성화 함수로 relu 함수가 사용되었다. 완전 연결계층은 활성화 함수로 tanh 함수를 사용하였다.

구별자는 LeakyReLU[6] 활성화 함수를 사용하는 4개의 컨볼루션 층을 사용하였으며 첫 번째 컨볼루션 층을 제외한 모든 층에 배치 정규화를 적용하였다. 마지막의 완전 연결계층에는 활성화 함수로 sigmoid 함수를 사용하였다.

III. 구별자 추가를 통한 모델 개선

본 논문의 모델에서는 생성자가 만든 가짜 영상과 진짜 영상을 판별하는 기존의 구별자 이외에 하나의 구별자를 더 추가했다. 이 추가적인 구별자는 기존의 구별자가 판별하지 못하는 생성 영상 사례(False Positive)에 대한 학습을 수행한다. 이 과정에서 컨볼루션 뉴럴 네트워크 분류기 중 하나인 vgg16[4] 네트워크를 사용한다.

생성자가 학습 과정에서 생성하는 영상 중 구별자를 통과한 영상을 vgg16 네트워크로 분류하여 그 분류 값이 레퍼런스 영상의 값과 비교했을 때 오차범위 밖에 있을 경우 별도로 저장한 후, 빈번히 발생하는 오류 영상을 추가된 구별자의 학습 데이터로 활용하였다. 기존의 구별자가 생성된 영상과 레퍼런스 영상을 판별하며 광범위한 학습을

하는 것과 달리, 추가된 구별자는 기존의 구별자의 빈번한 오류에 대해 협소한 학습 범위를 가진다.

IV. 실험결과

생성 모델의 성능을 객관적으로 평가해야 하는 경우 구분 모델의 평가 기준 등을 적용하기 어렵다.

본 논문에서 제안하는 영상 생성 방법에 대한 성능 평가 모델로는 Tim salimans가 제안한 Inception Score[8] 평가법을 이용하였다. 이 평가법은 분류 모델 중 하나인 Inception V3 모델을 사용하여 생성 모델이 생성한 영상에 대한 분류를 수행한 후, 분류 결과의 일관성을 기준으로 점수를 매긴다. 이때 Inception Score는 생성된 영상이 좋을수록 높은 값을 가지게 된다. 본 논문의 성능 평가에서는 생성된 영상 10,000개에 대한 Inception Score를 측정하였다.

표 1. 10 epoch에서 모델별 성능

모델	재학습 여부	Inception Score
VAE	X	1.456
DCGAN	X	1.694
DCGAN	O	1.742

학습 데이터로 클래스 정보등의 부차적인 정보를 사용하지 않고 영상만을 단독 사용하는 Variational Auto-Encoder[9]와 DCGAN[3]으로 실험하여 비교한 결과, 추가 구분자를 사용한 본 논문의 모델이 더 높은 Inception Score를 획득한 것을 알 수 있었다. DCGAN 모델끼리 비교하자면, 본 논문에서 제안한 모델은 기존의 모델보다 얼굴 생성 영상에서 눈, 코 입의 배치가 조금 더 자연스러워진 모습을 볼 수 있었다.

V. 결론

영상 특징을 학습하고 이를 재배치하는 신경망 구조를 이용한 영상 생성 모델에서 영상의 구성요소들이 불균형을 이루며 배치되는 현상은 고질적인 문제이다. 본 논문이 제안하는 영상 생성 모델은 GAN에서 구분자가 바르게 구별하지 못하는 영상의 종류를 기존의 분류 모델로 판별하고 이를 추가적인 구분자가 별도로 학습하는 방법으로 영상 생성의 일관성과 신뢰성을 높이는 결과를 도출하였다. 하지만 추가적인 구별자를 학습하기 위해 사용한 VGG16 네트워크 또한 신경망으로 학습한 분류 모델이며, 한정된 종류의 영상에만 분류 성능이 발휘되기에 광범위한 사례에 사용하기에는 한계가 남아있다.

References

- [1] Goodfellow, Ian, et al. "Generative adversarial nets." *Advances in neural information processing systems*. 2014.
- [2] Dumoulin, Vincent, and Francesco Visin. "A guide to convolution arithmetic for deep learning." *arXiv preprint arXiv:1603.07285* (2016).
- [3] Kim, Taehoon. "DCGAN-tensorflow." (2016).
- [4] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).
- [5] Salimans, Tim, and Diederik P. Kingma. "Weight normalization: A simple reparameterization to accelerate training of deep neural networks." *Advances in Neural Information Processing Systems*. 2016.
- [6] Xu, Bing, et al. "Empirical evaluation of rectified activations in convolutional network." *arXiv preprint arXiv:1505.00853* (2015).
- [7] Gauthier, Jon. "Conditional generative adversarial nets for convolutional face generation." *Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter semester 2014.5* (2014): 2.
- [8] Salimans, Tim, et al. "Improved techniques for training gans." *Advances in Neural Information Processing Systems*. 2016.
- [9] Kingma, Diederik P., and Max Welling. "Auto-encoding variational bayes." *_arXiv preprint arXiv:1312.6114_* (2013).