

정규화 기법 적용에 따른 GAN 모델의 성능 비교 연구

곽정기*, 고한석*

*고려대학교 전기전자공학과

e-mail : jgkwak@ispl.korea.ac.kr

A Study on the Performance Comparison of GAN Model According to the Normalization Techniques

Jeonggi Kwak*, Hanseok Ko*

*Dept. of Electrical Engineering, Korea University

요약

사람 얼굴 생성을 목적으로 하는 Generative Adversarial Network(GAN)에서 판별자(discriminator)의 각 레이어에 대한 스펙트럴 정규화(spectral normalization) 적용에 따른 출력 이미지의 결과를 비교하였다. 또한 생성자(generator)에 적용 인스턴스 정규화(Adaptive Instance Normalization) 모듈의 삽입에 따른 출력 이미지의 결과를 기존 모델과 비교하고 분석하였다.

1. 서론

최근 합성곱 신경망(Convolution Neural Network) [1]이 점점 발전되어 옴에 따라 다양한 방향의 연구가 제안되었다. 그 중 생성 모델에 따른 연구도 활발히 진행되었는데 Ian Goodfellow 가 생성자와 판별자가 서로 대립하며 학습을 진행하는 GAN [2] 모델을 처음으로 제안한 이후로 이를 활용한 생성 모델은 실제 사진과 비슷한 선명한 결과를 잘 출력하였고 학습 데이터 분포를 통해 학습 데이터와 유사한 새로운 샘플을 스스로 만들어 낼 수 있어 폭넓게 활용되며 우수한 성능을 보였다. GAN 모델은 여러 요인들에 의해 성능이 좌우되는데, 본 논문에서는 판별자의 생성자에 정규화 기법을 적용해보고 성능을 확인해보았다. 판별자의 각 레이어에는 스펙트럴 정규화 [3]를 적용하였고, 생성자의 각 레이어에는 배치 정규화(Batch Normalization) [4] 대신 적응형 인스턴스 정규화 [5]를 사용하여 얼굴 생성 GAN 모델에 적용하여 결과를 비교하였다.

2. 관련 연구 및 이론

GAN : 정규화 기법을 언급하기에 앞서 GAN 의 기본적 원리에 대해 설명하면, GAN 은 생성하는 가짜 샘플들의 확률 분포를 학습 데이터의 확률 분포와 가깝게 하는 것이 목적이다. 따라서 GAN 의 손실함수는 확률 분포간 거리를 측정하는 것에 대부분 기반한다. 초기의 GAN [2]은 KL-divergence 를 두 분포의 차이를 측정하는데 사용하였고, 그 후 많은 방법들이 제안되었는데, 그 중 Wasserstein distance 를 통해 두 확률 분포간 차이를 측정하고 학습하는 방법이 제안되었다 [6]. 다만 Wasserstein distance 를 GAN 에 활용하기 위

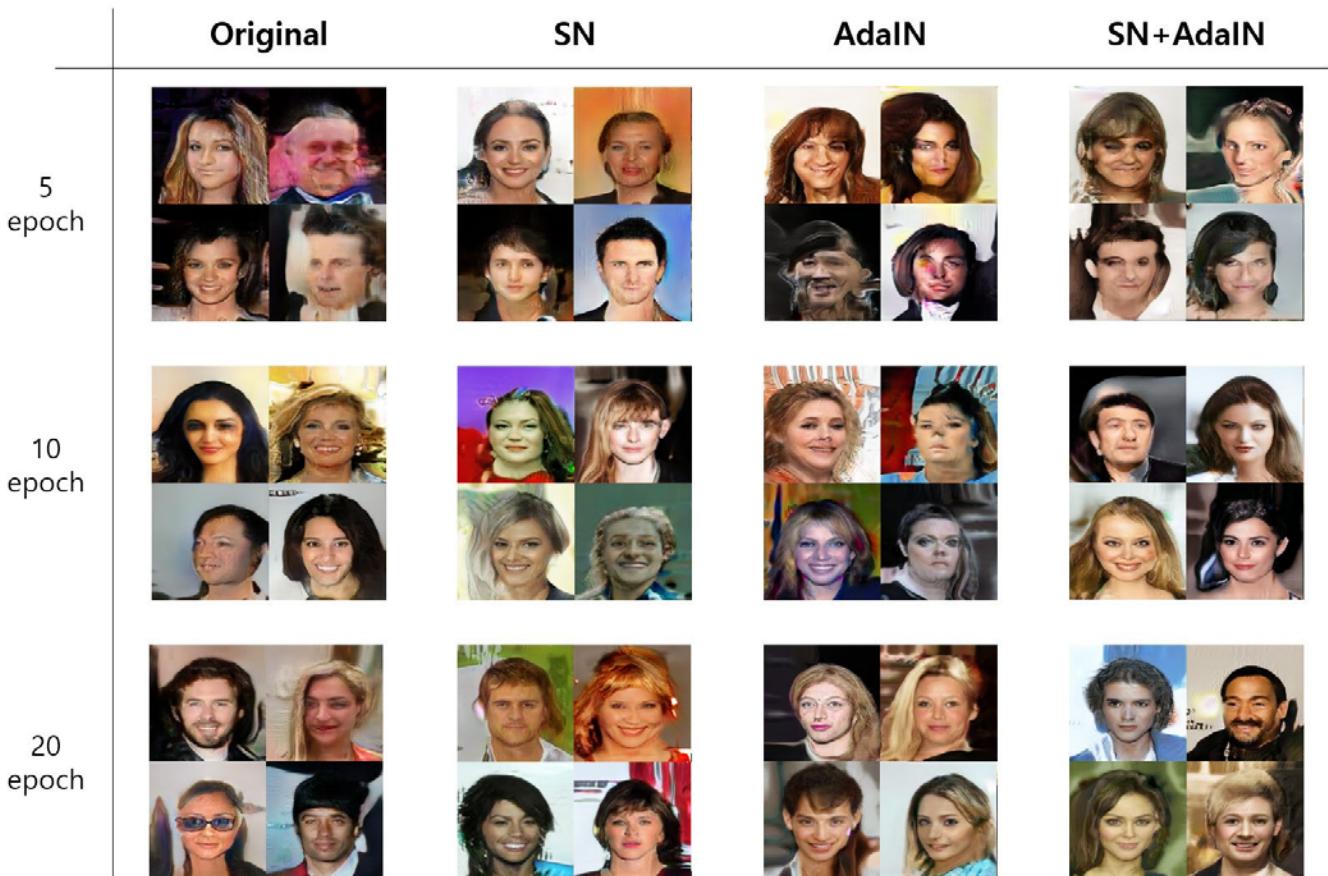
해서는 1-Lipschitz continuous 조건을 만족하여야 하였고, 그것을 만족시키기 위한 방법 중 하나로 제안된 것이 스펙트럴 정규화다.

스펙트럴 정규화 : 스펙트럴 정규화는 [3]에서 처음 제안되었다. 기존에는 판별자 내에서 1-Lipshitz continuous 를 만족시키기 위해서 그라디언트 패널티 [7]라는 방법을 주로 사용하였으나, 스펙트럴 정규화를 사용함으로써 해당 조건을 만족시키고 비교적 빠른 학습 시간을 보였다. 해당 기법을 적용시키는 방법으로는 판별자의 각 레이어에서 가중치 행렬의 스펙트럴 놈(spectral norm)의 최대값, 즉 특이값(singular value)으로 해당 행렬을 정규화 시켜준다.

$$\mathbf{W}_{SN} = \mathbf{W}/\rho(\mathbf{W}) \quad (1)$$

여기서 \mathbf{W} 는 각 레이어의 가중치 행렬을 의미하며, $\rho(\mathbf{W})$ 는 행렬 \mathbf{W} 의 스펙트럴 놈을 의미한다. 본 논문에서는 스펙트럴 정규화를 그라디언트 패널티와 따로 적용시키는 것이 아니라, 둘을 함께 사용하여 GAN 모델의 사람 얼굴 생성 성능을 비교해보았다.

적응형 인스턴스 정규화 : 해당 기법은 기존의 GAN 모델에서 처음으로 사용된 것은 아니고 스타일 변환(style transfer)을 수행하기 위한 방법으로 처음 제안되었다[5]. 기존의 방법과 같이 고정된 값들로 정규화를 하는 것이 아니라 훈련 데이터에 따라 다른 값들로 정규화가 이루어지는 것이 특징이다.



(그림 1.) 각 정규화 기법별 학습 단계에 따른 생성 결과 이미지. 각각 Original 은 기존의 모델, SN 은 스펙트럴 정규화를 적용하였을 때, AdaIN 은 적응형 인스턴스 정규화를 적용하였을 때, 가장 우측은 두 정규화 기법을 모두 적용하였을 때의 결과이다.

$$\text{AdaIN}(x, y) = \sigma(y) \left(\frac{x - \mu(x)}{\sigma(x)} \right) + \mu(y) \quad (2)$$

여기서 x 와 y 는 각각 기존의 특징 맵과 외부 데이터를 의미하고 기존 분포의 평균과 표준 편차로 정규화를 시키고 정규화된 값에 외부 데이터(입하고자 하는 스타일 이미지)의 표준 편차를 다시 곱하고 평균을 더해주어 계산하게 된다. 이 연산에 의해 생성되는 이미지는 주어진 학습 데이터의 스타일 정보를 반영 할 수 있게 된다. 기존 GAN 모델에서는 생성자에 주로 배치 정규화를 사용하여 실험하였는데, 본 논문에서는 배치 정규화 대신 적응형 인스턴스 정규화 모듈을 각 레이어에 삽입하였을 때 결과를 비교하였다.

3. 실험 및 결과

모델을 훈련에 있어서는, 우선 CelebA 데이터[8]를 사용하였다. 해당 데이터는 202,599 개의 얼굴 이미지를 제공한다. 출력 이미지의 크기는 128x128 이다. 따라서 모델을 훈련시키기에 앞서 학습 이미지들을 128x128 로 크기를 사전에 조정해 주었다. 순실 함수로는 WGAN-GP loss [7]를 사용하였다. 또한 학습에 있어서는 Adam optimizer 를 활용하여 최적화하였으며, 실험 pc 로는 GTX 1080Ti 를 사용하여 약 각 실험 별 약 24 시간 정도의 학습을 통해 생성된 이미지들을

비교하였다.

학습 과정에 따른 결과는 그림 1 에 나타나 있다. 기존의 WGAN [6] 모델에서 생성자에 적응형 인스턴스 정규화를 적용하였을 때, 판별자에 스펙트럴 정규화를 적용하였을 때, 그리고 두 가지를 모두 적용하였을 때 각각 5, 10, 20 에폭에서의 생성 결과를 나타낸다. 우선 생성자 부분에 있어서는 기존의 WGAN 생성자와 해당 모델에서 입력 레이어를 제외한 나머지 부분의 배치 정규화를 적응형 인스턴스 정규화로 대체한 새로운 모델의 성능을 확인하였다. 기존의 모델과 비교하였을 때, 모든 레이어에서 적응형 인스턴스 정규화 파라미터를 학습하기 때문에 학습 초기에는 디테일적으로 자연스럽지는 못하지만 학습이 점점 진행될수록 자연스럽거나 결함이 있는 부분들이 많이 개선되고 결과적으로 더 나은 이미지를 생성하는 것을 확인할 수 있었다. 이는 적응형 인스턴스 정규화를 생성자 각 레이어에 삽입함으로써 각 레이어 별로 입력 데이터의 정보를 더욱 효과적으로 활용하기 때문으로 볼 수 있다.

판별자 부분에서도 마찬가지로 기존의 판별자와 해당 모델에서 각 레이어에 스펙트럴 정규화 레이어를 추가한 새로운 모델의 출력 결과를 확인하였다. 스펙트럴 정규화를 추가할 경우, 출력 이미지 퀄리티의 개선뿐만 아니라 학습 모델의 안정화도 추가적으로

확인할 수 있었다. 스펙트럴 정규화를 사용하지 않았을 때는, 어느 시점 이후로는 GAN 모델들의 고질적인 문제인 mode collapse 가 발생하였지만 사용하였을 경우에는 학습을 중단시킬 때까지 안정적으로 이미지들이 생성되는 것을 확인할 수 있었다. 또한 둘을 같이 사용하였을 때도 각각을 사용하였을 때의 준하는 혹은 그 이상의 성능을 보였다.

4. 결론

본 논문에서는 기존 GAN 모델의 판별자와 생성자에 다른 정규화 기법을 각각 적용하여 성능을 확인해보는 실험을 진행하였다. 적응형 인스턴스 정규화를 통하여 생성자가 입력 데이터의 정보를 더 잘 학습할 수 있음을 확인하였고, 스펙트럴 정규화를 판별자에 적용함으로써 기존 모델 대비 성능 개선 및 학습 안정에 기여하는 것을 확인하였다. 다만 여전히 이미지 퀄리티에 있어서는 개선해야 할 부분이 남아있기에 제안한 방법을 넘어 더 많은 연구가 이뤄져야 할 것이다.

감사의 글

이 논문은 2019년도 BK21 플러스 사업에 의하여 지원되었음.

참고문헌

- [1] Krizhevsky, A.; Sutskever, I.; Hinton G. E. . Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems 2012.
- [2] Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Advances in neural information processing systems 2014.
- [3] Miyato, T., Kataoka, T., Koyama, M., and Yoshida, Y. Spectral normalization for generative adversarial networks. In ICLR, 2018.
- [4] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In ICML, 2015.
- [5] Huang, X.; Belongie, S. Arbitrary style transfer in real-time with adaptive instance normalization. In Proceedings of the IEEE International Conference on Computer Vision, 2017
- [6] Arjovsky, M.; Chintala, S. ; Bottou, L. Wasserstein Generative Adversarial Networks. ICML, 2017
- [7] Gulrajani, I.; Ahmed, F.; Arjovsky M., Dumoulin, V. ; Courville, A. C. Improved training of wasserstein gans. In Advances in neural information processing systems, 2017.
- [8] Liu, Z.; Luo, P.; Wang, X.; Tang, X. Deep learning face attributes in the wild. In Proceedings of the IEEE international conference on computer vision 2015.