대조 학습 및 단면 레이블 스무딩을 이용한 AttnGAN 개선 방법

이정은, 이정우
서울대학교

j_lee@snu.ac.kr, junglee@snu.ac.kr

Improving AttnGAN with Contrastive Learning and One-sided Label Smoothing

Lee Jung Eun, Lee Jung Woo
Seoul National Univ.

요 약

본 논문은 텍스트로부터 고품질 이미지를 생성하는 최신 알고리즘인 AttnGAN(Attentional Generative Adversarial Network)에 동일 data 간의 비교를 통해 학습하는 self-supervised learning 인 대조 학습과 출력 분포의 신뢰도를 낮춤으로써 모델의 일반화 성능을 개선하는 데 도움이 되는 레이블 스무딩 기법을 사용하여 AttnGAN 을 향상시켰다.

Ⅰ. 서 론

텍스트 기반 이미지 합성은 주어진 텍스트로 시각적 이미지를 생성하는 AI 작업이고 다양한 응용 분야가 있어 인기를 끌고 있다. 가장 많이 쓰이는 모델은 GAN(Generative Adversarial Network)[3] 기반의 생성 모델이며 서로 다른 두 개의 네트워크를 적대적으로 학습시키며 실제 데이터와 비슷한 데이터를 생성하는 모델이다. AttnGAN (Attentional GAN)[4]는 GAN 의 변형으로, 어텐션 메커니즘과 컨볼루션 신경망 작업을 사용하여 보다 상세하고 사실적인 이미지를 생성해내어 최근 각광받고 있는 모델이다.

한편, GAN 을 훈련하는 데 있어 주요 과제 중 하나는 품질이 낮은 이미지밖에 생성할 수 없도록 생성자와 판별자가 더 이상 최적화되지 않을 수 있다는 것이다. 즉, 판별자는 더 이상 생성자에 유용한 피드백을 제공할 수 없고 이로 인해 생성자가 개선되지 않은 채 학습 프로세스가 중단될 수 있다. 이러한 문제를 개선하기 위해 본 논문에서는 대조 학습(Contrastive Learning)과 단면 레이블 스무딩(One-sided Label Smoothing)을 AttnGAN 에 적용하여 성능을 비교해보았다.

Ⅱ. 본론

대조 학습은 둘 이상의 서로 다른 유형의 입력 데이터를 구별하도록 모델을 교육하는 기법이다.[6] GAN 에서는 동물과 풍경과 같은 다양한 유형의 실제 이미지를 구별하는 훈련으로 판별자를 향상시키는 데 사용할 수 있다. 대조 학습 관련 논문 들 중 구현이 쉬우면서 비선형 변환이 없고 배치 크기가 작아 계산 비용이 적은 GAN 모델 논문[2]에서 알고리즘을 참고하여 AttnGAN 에 대조 학습을 적용시켰다.

레이블 스무딩은 신경망을 훈련하는 데 사용되는 레이블에 소량의 노이즈를 추가하는 방법이다. GAN 에서는 실제 이미지의 레이블은 1 로, 생성된 이미지의 레이블은 0 으로 설정하여 학습을 진행하는데 레이블 스무딩을 적용하면 0 과 1 이 아닌 0.1 과 0.9 와 같이 약간 조정된 값을 사용하게 된다. 단면 레이블 스무딩은 여기서 실제 이미지의 레이블인 1 에만 스무딩을 적용한 것이다.[5] 심층 신경망은 정답 클래스에 대해서는 1 에 근접한 확률을 생성하여 확실하게 분류하도록 훈련되는 경향이 있기 때문에 단면 레이블 스무딩을 적용하여 레이블 0 은 그대로 유지하게 하고 레이블 1 에만 소량의 노이즈를 추가하여 판별자가 1 에 매우 근접한 값을 예측했을 때 패널티를 받도록 한다. 이렇게 하면 과적합의 위험을 줄이고 새로운 데이터에 대한 모델의 일반화 능력을 향상시킬 수 있다. 본 논문에서는 AttnGAN 에 단면 레이블 스무딩을 적용하여 실제 이미지의 레이블 1 대신 0.9 를 사용하였다.

Method	*IS*	*R-Precision(%)*
AttnGAN	2.95±0.03	11.64
AttnGAN + CL	3.00±0.03	12.03
AttnGAN + LS	3.30±0.03	11.99
AttnGAN + CL and LS	**3.44±0.05**	**12.12**

표 1. Data Set CUB 에 대한 평가 지표 비교 결과.
여기서 CL 은 대조 학습 LS 는 단면 레이블 스무딩이다.

실험 시 사용한 데이터 세트는 CUB[1]이며 새(Birds) 이미지와 그 이미지를 설명하는 텍스트 들로 구성되어있다. 평가 지표는 Inception Score(IS)[5]와 R-Precision[4]을 사용했다. IS 는 사전 학습되고 CUB 데이터 세트로 fine-tune 된 인셉션 네트워크를 사용하여 클래스 분포들의 KL-다이버전스를 계산한다. R-Precision 은 생성된 이미지가 주어진 텍스트 설명과 얼마나 일치하는 지를 표현한다. IS 와 R-Precision 모두 값이 높을수록 좋은 모델임을 의미한다.

실험 결과, 표 1 에서 보이는 것과 같이 대조 학습 자체로는 IS 와 R-Precision 을 각각 1.7%, 3.4% 향상시켰고, 단면 레이블 스무딩은 IS 를 11.9%, R-Precision 을 3.0% 향상시켰다. 한편, 두 기법을 동시에 적용했을 경우 IS 는 16.6%, R-Precision 은 4.0% 상승했다. 기본 모델인 AttnGAN 과 비교하여 단일 기법으로도 성능이 충분히 향상되지만 두 기법을 따로 적용했을 경우보다 동시에 적용한 경우가 성능이 더 좋음을 확인할 수 있다.

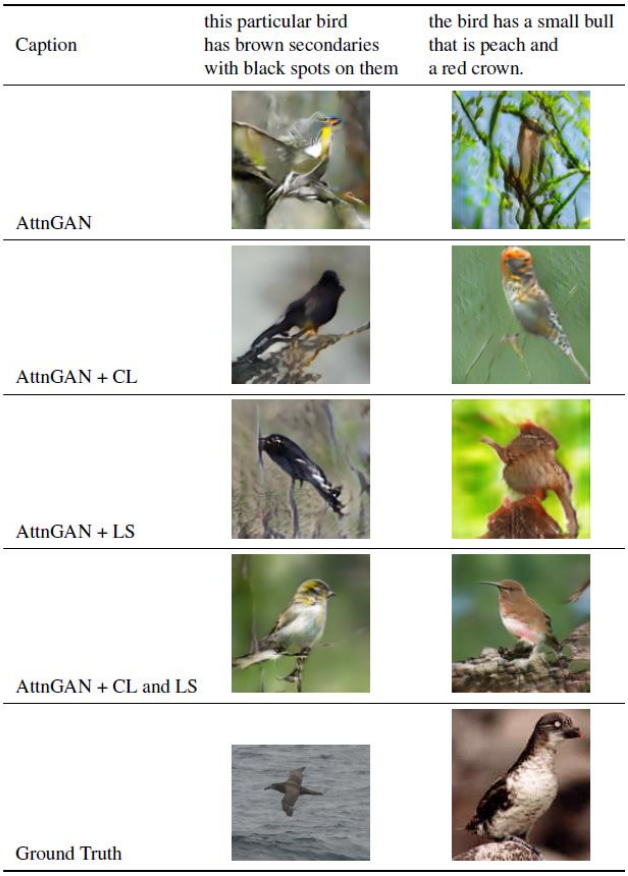

그림 1. 생성된 이미지 비교
여기서 CL 은 대조 학습 LS 는 단면 레이블 스무딩이고 Ground Truth 는 학습에 사용된 실제 이미지이다.

평가 지표로 확인하는 것 외에 생성된 이미지들끼리 비교하는 것으로도 대조 학습과 단면 레이블 스무딩을 모두 적용한 모델이 성능이 더 좋다는 것을 확인할 수 있다. 그림 1 은 각각의 모델들을 학습시킨 후 주어진 Caption 들로 이미지를 생성한 결과이다. AttnGAN 자체, 혹은 대조 학습만 적용하거나 단면 레이블 스무딩만 적용한 모델이 생성한 이미지와 비교하여 두 기법을 동시에 적용한 모델의 이미지가 더 품질이 좋은 것을 확인할 수 있다. 또한 Ground Truth 이미지보다 Caption 에 더욱 일치한다.

실험을 통해 대조 학습은 Caption 과 더욱 일치되는 이미지를 생성할 수 있고 일반화에도 효과가 있으며 단면 레이블 스무딩은 더 다양한 이미지를 생성하게 하고 일반화 측면에서는 더욱 효과적임을 알 수 있다. 이 두 가지를 동시에 사용하면 시너지 효과로 이미지의 품질과 일반화 측면에서 성능이 크게 향상되는 것을 확인하였다.

Ⅲ. 결론

대조 학습 및 단면 레이블 스무딩은 보다 좋은 이미지를 생성하고 새로운 데이터에도 좋은 성능을 낼 수 있도록 일반화 능력을 향상시키는 데에 효과적인 기법이다. AttnGAN 에 두 기법을 적용한 결과 성능을 크게 향상시킬 수 있었다. 한편 AttnGAN 은 GAN 의 응용 알고리즘이므로 다른 GAN 응용 알고리즘에 적용시킨다면 마찬가지로 이미지의 품질 및 일반화 측면에서 성능 향상을 기대할 수 있을 것이다.

참 고 문 헌

[1] P. Welinder P. Perona C. Wah, S. Branson and S. Belongie. Caltech-ucsd birds-200-2011 dataset., 2011. The Caltech-UCSD Birds-200-2011 Dataset. California Institute of Technology.

[2] Martin Takac Rajshekhar Sunderraman Hui Ye, Xiulong Yang and Shihao Ji. Improving text-to-image synthesis using contrastive learning. arXiv:2107.02423, 2021.

[3] Mehdi Mirza Bing Xu David Warde-Farley Sherjil Ozair Aaron C Courville Ian J Goodfellow, Jean Pouget-Abadie and Yoshua Bengio. Generative adversarial nets. NeurIPS, 2014.

[4] Qiuyuan Huang Han Zhang Zhe Gan Xiaolei Huang Tao Xu, Pengchuan Zhang and Xiaodong He. Attngan: Fine-grained text to image generation with attentional generative adversarial networks. CVPR, 2018.

[5] Wojciech Zaremba Vicki Cheung Alec Radford Xi Chen Tim Salimans, Ian Goodfellow. Improved techniques for training gans. NeurIPS, 2016.

[6] Mohammad Norouzi Ting Chen, Simon Kornblith and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. ICML, 2020.