데이터 특징을 이용하여 다양한 이미지를 생성하는 이미지 생성 네트워크 분석
Image Generation Network Analysis to Generate Various Images Using Data Features

이동규, 한동석

경북대학교 대학원 전자전기공학부

jasmindoe@knu.ac.kr, *dshan@knu.ac.kr

요 약

본 논문은 딥러닝 모델의 학습 데이터를 확보하기 위해 이미지 생성 네트워크에 관해 분석하였다. 이미지 생성 네트워크는 무작위 입력을 통해 특정한 결과물을 생성하기도 하지만 특정한 입력을 통해 그와 유사한 특징을 가지는 결과물을 만들어 낸다. 결과물 생성에 있어 필요로 하는 특징 정보만을 활용하게 되면 모델을 학습하는데 필요한 학습 데이터들을 확보할 수 있게 된다. 이미지 생성 네트워크 중 StyleGAN[2]은 입력 데이터의 특징을 변형하여 입력과 유사하지만 전혀 새로운 데이터를 생성할 수 있는 모델이다. 해당 모델은 데이터의 공간 상의 관계성을 유지하면서 각 특징들의 지각적 공간 관계를 유연하게 하기 위해 맵핑 네트워크와 추가적인 평균화 방식을 사용한다. 이런 특징 정보들을 통해 특정한 특징만을 조절하여 사용자가 목적으로 하는 결과물을 생성하게 된다. FFHQ(Flickr-faces-high-quality)와 COCO[4] 데이터셋을 사용하여 모델을 학습하여 사람의 데이터를 생성해서 성능을 분석하였다. 특정 입력으로부터 특징값을 조절하지 않는 무작위 결과물과 특징값을 조절한 결과물을 분석하고 이미지 퀄리티에 따른 결과물의 성능을 지각적 거리 수치와 생성 시간으로 분석하였다.

Ⅰ. 서 론

딥러닝 모델은 학습을 위해 데이터를 이용하여 원하는 결과를 얻어낼 수 있다. 학습 데이터는 모델에 사용되기 위해서는 목적에 따라 종류가 달라지고 가공이 되어야 한다. 데이터의 확보와 가공에는 많은 시간과 인력이 요구되다 보니 충분한 수의 데이터를 확보하기가 어렵다. 이러한 데이터를 확보하기 위해 딥러닝 방식으로 데이터를 생산하는 방안들이 연구되고 있다. 그 중 하나가 GAN(generative adversarial network)[1]로 생성 네트워크를 통해 새로운 종류의 이미지를 생성하는 딥러닝 모델이다.

GAN은 입력을 통해 이전에 없던 전혀 새로운 종류의 결과물을 만들어 낼 수 있다. 이는 기존에 수집한 학습 데이터와 다른 추가적인 데이터를 확보함에 있어 매우 유용하다. 그러나 GAN도 입력에 따라 원하는 결과물을 얻기가 힘들다. 생성 네트워크에서 우리가 필요로 하는 것을 골라 결과물을 얻어내야 하기 때문이다. CGAN(conditional GAN)[5]은 네트워크의 결과물을 통제하기 위한 모델이다. 해당 모델을 통해 어느 정도 사용자가 원하는 종류의 결과물을 얻을 수 있지만 분류할 수 있는 가짓수가 얼마되지 못하는 한계점이 있다. Pix2Pix[6]은 기존의 이미지 생성 네트워크들의 접근법과 다르게 사용자가 직접 원하는 것을 입력하여 결과물을 얻고자 한 모델이다. 기존의 GAN들이 무작위 입력값이나 기준이 되는 데이터를 입력으로 사용하였다면 해당 모델은 사용자로부터 단순한 입력 몇가지로 결과물을 생성한다. 그러나 이 모델 또한 학습을 위해서는 기존 모델들과는 다르게 쌍으로 존재하는 데이터가 요구되어 특정한 클래스의 데이터는 학습이 어려울 때가 있다.

이를 해결하고자 한 모델로 CycleGAN[9]이 있다. CycleGAN은 기존 GAN의 문제인 입력에 따른 결과물 붕괴 문제를 해결함과 동시에 학습의 용이성을 높이면서 원하는 결과물을 얻을 수 있게 입력을 유도한 모델이다. 이 모델이 학습한 카테고리와 객체에 따라 매우 뛰어난 성능을 보이면서 다른 스타일의 데이터를 얻을 수 있지만 다양한 변형을 하지 못하는 한계점이 존재한다. StyleGAN은 이미지의 특징들을 바로 생성 네트워크의 입력 공간에 매칭 하지 않고 추가적인 네트워크를 사용한다. 해당 네트워크에 입력의 특징들을 추출하여 다양한 변형을 가지는 결과물을 얻어낼 수 있는 모델이다. StyleGAN은 이전의 GAN들과는 달리 사용자가 원하는 종류의 특징만들 골라서 결과물을 생성할 수 있으며 동시에 뛰어난 결과물을 보여준다.

본 논문은 추가적인 학습 데이터를 확보하기 위해 다양한 특징을 가지는 데이터를 생성하는 이미지 생성 네트워크들인 styleGAN을 사용하여 결과물을 생성하고 다른 이미지 생성 네트워크와 비교하여 해당 방식이 학습 데이터를 확보하는데 적합한지 분석하였다.

Ⅱ. 테스트 구성과 결과

이미지 생성 네트워크는 GAN을 기반으로 한다. StyleGAN 또한 GAN을 기반으로 하는 모델로 일반적인 GAN이 아닌 DCGAN(deep convolution GAN)[7]에서 고화질의 이미지를 생성하는 모델인 PGGAN(progressive growing GAN)[]을 사용한다. PGGAN은 저해상도 이미지를 고해상도의 출력물로 생성할 수 있게 점진적으로 수축과 팽창을 하는 구조로 되어있다. 특정한 결과물을 가지는 이미지를 얻기 위해 추가적으로 8개의 층인 맵핑 네트워크를 사용한다. GAN에서 생성 네트워크가 결과물을 만들기 위해 입력 데이터를 공간 상에 투영을 해야 한다. 이 과정에서 styleGAN은 생성 네트워크로 바로 데이터를 투영하는 것이 아닌 맵핑 네트워크를 거치게 한다. 맵핑 네트워크를 거친 데이터들은 데이터들의 특징들이 가지는 공간상의 관계성이 유지되게 해주고 특정한 특징 정보를 사용할 수 있게 한다. 이 후 GAN으로 입력되는 이미지는 데이터 분포 공간에 맞게 데이터가 변형되고 해당 이미지의 특징 정보들은 데이터 공간에 분포하게 된다. 분류된 특징 벡터들은 AdaIN(adaptive instance normalization)[3] 방식과 추가적인 노이즈 입력이 합쳐져서 고유의 특징을 나타내는 이미지를 생성하게 된다. 해당 특징 벡터 값을 조절하여 기존의 이미지와 유사하면서 다른 정보를 가지는 이미지가 생성된다.

특징 벡터들은 완전히 개별로 분리된 것이 아닌 특징별로 연결성이 존재하기 때문에 모델의 학습 정도에 따라 생성되는 결과물이 달라진다. 특정한 색을 가지는 결과물을 만들기 위해 생성 네트워크에 해당 이미지와 특징 벡터 정보를 입력하여도 해당 특징 벡터 정보만이 아닌 다른 정보들 또한 영향을 받아 변화하게 된다.

생성하고자 하는 학습 데이터는 이미지 데이터로 사람을 생성하고자 하였다. 데이터셋은 FFHQ, COCO를 통해 학습하였다. 이미지 구별 네트워크는 ImageNet을 기반으로한 미리 학습된 모델을 사용하고 데이터를 공간상에 투영하기 위해 InceptionNet[8]을 사용하였다.

(a) (b) (c)

그림. 1. 원본 입력 (b)를 통해 생성된 무작위 결과물 (a), ©

아무런 특징 정보를 추출하지 않고 특정한 이미지 입력을 통해 나온 결과물은 그림 1과 같다. 그림 1의 (b)를 생성 네트워크에 입력으로 준 결과 (b)와는 전혀 다른 그림 1 (a)와 (c)가 생성되었다. 원본 이미지의 특징들 중 몇가지가 결과물에 나온 것을 볼 수 있는데 어린 나이대와 얼굴형태 등의 특징이 결과물에서 확인된다.

그림. 2. 특징 벡터를 조절하여 생성된 결과물들

그림 2는 중간의 원본 이미지에서 특정한 특징 정보들을 바꾼 결과물들이다. 위의 이미지는 중간 이미지에서 나이대 정보를 가지는 특징 벡터를 조절하였고 아래 이미지는 얼굴 형태 정보를 조절하였다. 결과물을 보면 조절한 특징 벡터만 결과물에 반영된 것이 아닌 다른 특징들도 변형된 것을 확인할 수 있다.

그림. 3. 이미지 화질에 따른 데이터 투영 결과물

표. 1. 이미지 화질에 따른 지각적 거리 결과 및 생성 시간

Resolution	Perceptual	Processing Time
128x128	0.2160	56s
512x512	0.2476	105s
1024x1024	0.2485	141s

이미지 정보를 공간상에 투영하기 위해 InceptionNet을 사용하여 투영한다. 이 때 이미지 화질에 따라 지각적 공간 거리값과 생성 시간이 달라지는데 보통 고화질일수록 생성 결과물 또한 좋아진다. 하지만 표 1에서 확인할 수 있듯 화질이 높아질수록 한 장의 이미지를 투영하는데 걸리는 시간이 기하급수적으로 증가하지만 생성 결과물이 뚜렷해짐을 볼 수 있다. 반대로 화질이 낮을수록 특징들 간에 지각적 공간 거리값이 줄어들어 원하는 특징 정보를 변형함에 있어 유용하게 된다.

Ⅲ. 결론

본 논문에서는 추가적인 학습 데이터를 확보하기 위해 이미지 생성 네트워크를 사용하기 적합한지 평가하고자 한다. StyleGAN은 사용자가 원하는 특징 정보들을 조절하여 생성 네트워크를 통해 원하는 결과물을 얻을 수 있는 모델이다. 이미지의 특징 정보에 따라 원본 입력과 유사하지만 다른 결과물을 얻거나 무작위 입력을 통해 전혀 새로운 결과물을 얻을 수 있다. 해당 이미지 생성 네트워크의 결과물은 학습 결과에 따라 퀄리티가 달라지고 완벽히 구분되는 특징만을 생성할 수 없기 때문에 추가적인 연구가 필요하다.

ACKNOWLEDGMENT

본 연구는 산업통상자원부와 한국산업기술진흥원이 지원하는 5G기반 자율주행 융합기술 실증 플랫폼 과제(과제고유번호 : 1415169669)의 지원을 받아 수행하였습니다.

참 고 문 헌

[1] Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. "Generative adversarial nets." Advances in neural information processing systems 27 (2014).

[2] Karras, Tero, Samuli Laine, and Timo Aila. "A style-based generator architecture for generative adversarial networks." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 4401-4410. 2019.

[3] Huang, Xun, and Serge Belongie. "Arbitrary style transfer in real-time with adaptive instance normalization." In Proceedings of the IEEE international conference on computer vision, pp. 1501-1510. 2017.

[4] Lin, Tsung-Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. "Microsoft coco: Common objects in context." In European conference on computer vision, pp. 740-755. Springer, Cham, 2014.

[5] Mirza, Mehdi, and Simon Osindero. "Conditional generative adversarial nets." arXiv preprint arXiv:1411.1784 (2014).

[6] Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[7] Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." arXiv preprint arXiv:1511.06434 (2015).

[8] Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[9] Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." Proceedings of the IEEE international conference on computer vision. 2017.