KiDRS Korea Institute of Design Research Society 2016, Vol. 01, No. 1, 001-010

Generative adversarial nets를 이용한 빈센트 반 고흐 이미지 생성 시스템

GAN(Generative Adversarial Nets) 기법을 중심으로

Vincent Van Gogh image generation system using generative adversarial nets

Focusing on the Gan Method

주 저 자 : 조수영 (Cho, Sooyoung) 광운대학교 전자통신공학과

공동저자 : 김문석 (Kim, Moon Seok) 광운대학교 동북아문화산업학부

교신저자 : 손채봉 (Sohn, Chae-Bong) 광운대학교 전자통신공학과
cbsohn@kw.ac.kr

접수일자 2017. 12. 04 / 심사완료일자 2017. 12. 20 / 게재확정일자 2017. 12. 22

Abstract

CNN and RNN are classifiers for image and speech recognition, and are used in many computer vision. However, this model alone does not produce images or voices. So we used GAN (Generative Adversarial Nets), which is a non-biped learning, to create and restore objects. To determine the distribution of the assumed model data, we use the generative model G and the Discriminator model D to judge the probability of each case in order to distinguish whether the sample came from the training data from the actual or model G Respectively. This is called minimax two - player game. The model G makes a mistake maximize in order to make a mistake, and the model D makes a mistake minimize in order not to make a mistake. Model D uses technology such as recognition technology, which is a conventional prayer learning method. Through this technique, the damaged or modified objects are compared with the training sample to classify and confirm what kind of object is the first type. There is no need for another network to generate the sample through training and it will be evaluated and confirmed through experiments.

Keyword

Generative Adversarial Nets, CNN, DCGANs

요약

CNN과 RNN은 이미지 및 음성 인식 등을 위한 분류자로서 많은 컴퓨터 비전에 사용된 지도 학습 모델이다. 지도학습 기반의 딥러닝 기술이 발전할 수 있었던 이유는 여러 가지 라벨링 된 공공의 데이터 셋이 존재했기 때문이다. 이러한 모델만으로는 이미지나 음성을 생성하는데 어렵다. 비지도 학습인 GAN(Generative Adversarial Nets)을 사용하여 이미지의 색감을 변형하는 형식의 작업을 통해 이미지와 음성을 쉽게 생성한다. Target 이미지는 Vincent van gogh의 대표적인 작품에서 인물 자화상에 대해 object model로 상정하였다. 상정한 model 데이터의 분포를 알아내기 위해서 DCGANs(Deep convolutional generative adversarial networks)을 사용하여 학습을 시키고 후보자들에 대해서 판별을 했다. Generative를 통해서 생성된 이미지를 판단하기 위해 Discriminator를 이용하여 sample이 training data에서 온 것이 실질적인지 혹은 만들어진 것인지를 구별하기 위해 각각의 경우에 대한 확률을 추정하였다. 이미지 데이터 세트를 통해서 convolutional adversarial pair의 생성 자와 판별 자의 객체 부분에서 학습한다는 증거를 보여주게 된다. 또한 training을 통해 나온 sample 생성에 있어서 다른 비교 군과 비교 및 증명하기 위해 mnist와 다른 다양한 이미지를 같은 모델에 비교하며 적용 가능성을 보여준다. Vincent van gogh의 자화상을 이용하여 생성된 이미지를 확인이 가능하였다.

목차

1. 서론

1-1. 연구의 배경 및 목적

1-2. 연구 방법

2. 이론적 배경

2-1. GAN의 개념과 유형

2-2. DCGAN의 개념

3. GAN의 활용 사례

3-1. MNIST Dataset

3-2. 얼굴 인식 및 표현 분야

4. 실험

4-1. 데이터 모델 샘플

4-2. 실험환경

4-3. n-step 학습

5. 결론

참고문헌

1. 서론

1-1. 연구의 배경 및 목적

최근 컴퓨터 비전에 관해서 많은 Deep leaning에 관련된 자료 및 알고리즘들이 개발이 되었다. 그 중 인식 분야가 크게 개발이 되어 가고 있는 시점이다. 인식 분야는 학습된 데이터를 이용하여 물체의 영상을 통해서 물체의 종류, 크기, 방향 위치 등 3차원적인 공간정보를 실시간으로 알아내는 기술이다. 이러한 기술들은 고성능 시스템을 이용해서 크게 향상되고 있다. 이러한 영상처리 기법을 이용해서 이미지의 영상을 처리하고 표현을 할 수 있는데 시각적인 요소를 객체 유형에 따라 상세한 시각정보를 각각의 속성으로 표현을 하였다.(객체 범주가 벡터로 인코딩 된 특징을 구별한다.) 그래서 이것을 실질적으로 이미지 영상에서 표현을 하는데 이러한 모든 정보를 Deep learning 기술을 이용하여 학습을 시킬 수 있다. 그러나 이미지 영상은 매우 복잡하고 고차원적으로 구성되어 있기 때문에 여러 방법을 사용하더라도 모델의 정보를 이용하여 object의 모델링을 한다는 것은 매우 어려운 점을 포함하고 있다, 그래서 Convolutional Networks(ConvNets)를 통해 많은 문제를 해결 할 수가 있는데, 이는 고 해상도나 여러 영상에서 처리해야 할 이미지 및 object 모델링하는 것을 대부분의 이미지를 필터를 생성하거나 여러 네트워크를 적용시키는 작업을 한다. 그리고 그 정보를 가지고 최적화를 하거나 정보에 적합한 알고리즘을 사용하여 문제점을 해결한다. 이는 이미지 분류 및 인식 말고도 다양한 부분에서도 적용이 되고 있다. 물체를 인식하여 검출 및 추적하는 기술은 높은 인식률과 판단 율이 중요하다. 그래서 많은 연구 부분에서 여러 Deep learning 기술들을 사용하는 방식으로 진행 되고 있다. 객체를 처리하기 위한 시스템을 영상처리 기술만 가지고 한다면 시스템 부분에서 많은 단계 및 연산을 작업하는 부분이 생겨 이 점에 관한 단점이 생긴다. 그리고 영상을 바꾸는 과정에서 손실이 생기기 때문에 실시간 뿐 아니라 다른 정지된 영상 내에서도 인식 및 추적 기술에 관한 판단 율이 저하되기 때문에 많은 부분을 수정을 한다는 과정이 생긴다. 그리고 자연스러운 이미지의 구조를 활용하는 부분에서 특정한 부분에서 이미지 구조를 캡처하거나 laplacian 혹은 gaussian distribution을 사용하는 등 여러 가지 방법으로 일련의 생성 모델을 구축하는 경우도 종종 있다. 위와 같은 방법 중 하나로 GAN (Generative Adversarial Networks)이라고 불리는 네트워크를 이용하여 이미지의 새로운 영상을 생성할 것이다. GAN은 샘플을 만들 시에 저주파 이미지를 시작으로 coarse to fine 방식으로 그려진다. 이러한 과정을 통해 다음 이미지를 샘플링을 하고 최종적인 레벨에 도달할 때까지 이전 스케일의 출력을 조절하는 방식으로 나아간다. 위와 같은 접근 방식은 객체 분류에 있어서 시스템적 구분과 비슷하나 시각적인 면에 있어서 매우 높은 효과를 볼 수 있다. 그리고 높은 결과 값을 표현하여 원본 이미지와 동일한 수준의 값을 얻는다. 본 논문에서는 GAN 네트워크를 이용하여 데이터들을 학습을 시키는데 Vincent van gogh의 인물 자화상을 이용하여 최종 data의 생성과 분석을 하는 방식을 제안한다.

1-2. 연구 방법

GAN에서 생성된 이미지 모델은 parametric과 non-parametric으로 구성이 된다. Non-parametric란 흔히 이미지 패치와 일치하는 기존 이미지의 데이터베이스에서 매칭하고 각각의 해상도와 페인팅을 수행하는 방식이다. 반대로 이미지 생성을 위한 parametric model은 mnist와 비슷한 형태로 많은 형태로 연구가 되어가고 있다. 하지만 현실에서 자연스러운 이미지를 생성하는데 크게 어려움을 느낄 수 있다. 이미지 생성에 대한 다양한 sampling의 접근 방식이 (Kingma & Welling, 2013)[1] 성공을 케이스가 있지만 아직까지도 흐릿하게 보이거나 뭉개지는 형상이 보인다. 또 다른 접근 방식은 이미지의 방향에 따른 반복적인 프로세스를 학습시켜서 이미지를 생성하는 방식이 존대한다.(Sohl-Dickstein et al., 2015)[2]

반면에 GAN은 난해하고 이해하기 어려운 이미지들을 생성을 한다. 이 접근법은 laplacian parameter를 확장하는 형식으로 이미지에 대한 높은 품질의 이미지를 보여 주지만 여러 다양한 모델을 학습하여 연결을 할 경우에는 노이즈가 발생하는 형식이 나타

1) Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.

2) Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., & Ganguli, S. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. arXiv preprint arXiv:1503.03585.

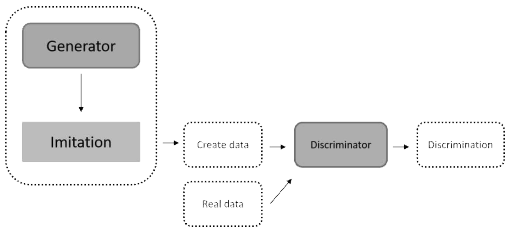
나게 된다. (Goodfellow et al., 2014)[3)]

Recurrent network approach와 de-convolution network를 통해서 노이즈가 발생하는 형상을 없애며, 자연 영상을 생성하였다. 그리고 입력 이미지에 대한 검사부분에 있어서 gradient descent를 이용하여 검사를 실시 할 수 있다.

2. 이론적 배경

2-1. GAN의 개념과 유형

지도학습(supervised learning)은 데이터와 label을 이용하여 집중적으로 학습을 시키는 방법이다. 예시로 어떤 한 사물을 인식을 시키기 위해서 여러 가지의 사물의 정보를 입력을 시켜 답을 알려주고 지속적으로 학습을 시키는 경우이다. 얼굴이나 사물, 음성인식 등 여러 방향으로 알려져 있는 인식 분야에서는 지도학습 방법을 사용한다. 하지만 이러한 방법에는 한계가 존재를 하는데 지속적으로 data를 입력을 시켜야 하며 label을 통해서 학습을 시켜야 한다. 얼굴 인식 분야에서 보면 특정 인물에 관해서 인식을 시키기 위해서는 labeling을 통해서 학습을 시키는데 data 수가 부족하거나 잘못된 information 입력하게 되면 치명적인 문제점이 발생한다는 점이 지도학습의 한계를 바라 볼 수가 있다. 하지만 비지도 학습(unsupervised learning)에서는 label이 아닌 스스로 학습을 하는 모델이다. 비지도 학습이 주목받는 이유는 지도 학습에 비해 존재하지 않는 label을 배우기 때문이다. 지금까지 deep learning 기술은 지도학습 기반의 CIFAR 같은 labeling이 되어 있는 공공의 데이터의 존재로 개발이 되어 왔다. 그런데 이 deep learning의 기술을 적용시키는 방법이 어렵기 때문에 이 부분을 보완을 해주는 것이 비지도 학습 부분이다. 비지도 모델 중 GAN을 이용하여 문제점을 개선할 수 있다. GAN은 인공지능의 문제점인 수동적인 인식 부분에서 능동적으로 행동하는 방향으로 개선이 되었다. 그리고 GAN은 모조품을 통해 배우는 학습 알고리즘이라고 불려진다. 주요 목표는 최대한 진짜 같은 샘플을 만드는 것을 목표로 하는 것이다. 비지도 학습 방법은 실제 데이터를 토대로 함수로 표현을 하고 이를 다시 일반화 하여 새로운 데이터로 만드는 것이다. 데이터가 사람의 얼굴을 재창조 한다면, 여러 다양한 얼굴의 학습 데이터를 입력해주고, 입력한 데이터들에 관해서 변화되는 과정을 통해 적절한 모델을 만들어 다른 형상의 얼굴을 표현을 할 수가 있다. 대부분의 확률 모델 네트워크는 다양한 데이터들의 평균점을 합산하여 표현이 되기 때문에 blur한 이미지를 많이 얻게 되는데, GAN을 통해서 학습을 시키게 된다면 generator와 discriminator의 두 model이 각각 학습을 하면서 표현되기 때문에 이미지의 생성 되는 과정에서 위와 같은 문제점을 보완을 한다.



[그림 1] Basic structure of GAN

위 그림 (그림 1)과 같이 generator가 하나의 data를 만들고 discriminator가 구분을 하면서 발전해 나가기 때문에 learning time이 길어질수록 좋은 결과물을 얻게 되는 것이다. 여기서 discriminator는 기존 지도학습 방법의 '인식 기술'과 같은 방법으로 사용한다. 따라서 generator가 discriminator를 속이기 위해 더 좋은 결과물을 표현하기 때문에 스스로 학습이 되는 방향으로 네트워크가 만들어 진다. 두 가지의 모델을 사용을 하면서 학습을 하는 GAN의 아이디어는 기존 비지도 학습의 많은 문제점을 피하면서도 지도학습의 기술을 가지고 활용하기 때문에 기존 생성 모델들의 문제점을 보완하고 새로운 이미지를 낼 수가 있으며, 사용자가 입력한 조건에 가까운 이미지를 생성을 할 수가 있다.

2-2. DCGAN의 개념

GAN에 관한 많은 연구가 진행이 되면서 network의 불안정한 구조로 인해 최근 많은 연구 주제가 '안정화' 작업이었다. Minimax 혹은 saddle problem을 풀어야하는 GAN은 두 가지 모델을 가지고 있기 때문에 이에 관한 불안정한 구조로 만들어질 수밖에 없다. 이론적으로는 fixed solution으로

3) Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).

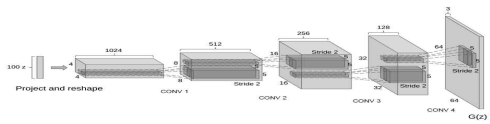
수렴하는 것이 보장이 되지만, 여러 논문에서 실제 적용하는 단계에서 불안정한 구조 때문에 단점이 나타났다. 이 문제를 극복하고자 만든 것이 DCGAN이다. (NIPS 2016 Tutorial by Ian goodfellow) 그리고 생성 영상의 해상도를 높이기 위해 많은 연구가 진행이 되어 있으며 (Denton et al.,2015)[4], CNNs의 black box methods 부분과 sample의 평가를 측정하기 위한 부분을 해결하고자 DCGAN을 사용을 하게 되었다. DCGAN은 Convolutional 구조로 만들어 졌으며, 그 구조는 CNN에 많은 영향을 끼치게 되었는데, 지도학습과 비지도 학습에도 많은 영향을 끼쳤다. 이론적으로 DCGAN은 대부분 상황에서 안정적으로 학습이 되는 구조로 제안이 되었으며, generator가 벡터 산술 연산이 가능한 설정을 가지고 있으며 semantic 수준에서의 sample generation을 할 수 있다는 점과 학습이 완료된 네트워크를 통해서 시각화 하고 특정 filter들이 이미지의 특정 물체를 학습하는 점과 학습이 완료된 discriminator가 다른 비지도 학습 알고리즘들과 비교하여 비등한 성능을 보인다는 점이 DCGAN의 매우 큰 특징이다. DCGAN과 GAN의 차이는 Convolutional의 구조의 유무이다.

[표 1] GAN module

Algorithm	Convolutional structure
GAN	not apply
DCGAN	apply

3. GAN의 활용 사례

3-1. MNIST Dataset[5]



[그림 2] Deep Convolutional Generative Adversarial Networks

4) Denton, E. L., Chintala, S., & Fergus, R. (2015). Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks. In Advances in neural information processing systems (pp. 1486-1494).

5) http://yann.lecun.com/exdb/mnist/ THE MNIST DATABASE of handwritten digits

GAN에는 loss function이 손실을 표현을 하지만 각 모델의 성취도 혹은 성능을 나타낸다. 그래서 각 모델의 loss function을 최대화 하는 것이 학습의 목표로 한다. discriminator의 성능은 실제 이미지를 잘 구분할수록, 1에 가까운 수가 부여 된다. 그리고 generator의 성능은 자기 자신이 만든 생성된 이미지를 분류기가 구분하지 못하게 해야 한다. 그 때가 낮은 수에 가까울수록 좋은 성능이라고 볼 수 있다. 이때를 d_loss (real result)와 g_loss (fake result)로 표현 할 수 있다. 두 model (표 1)을 학습 시키는 과정을 통해서 최대화 시킨다. 이러한 과정을 mnist에 적용을 시키는데 입력 data를 넣어서 생성과 분류를 통해 output을 뽑는 형태이다. (그림 3)

[그림 3] MNIST training results

[표 1] GAN module

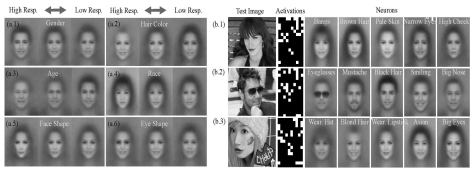
module	function
generator	2 layer로 구성된 이미지 생성기
discriminator	2 layer로 구성된 binary 분류기
random noise	random normal한 batch size 개수의 noise를 생성

3-2. 얼굴 인식 및 표현 분야

얼굴 인식 및 표현 분야에서는 많은 학습 네트워크의 발전에도 불구하고 제한되지 않는 환경에서는 큰 효과를 볼 수가 없었다. 이는 많은 요소가 자리 잡고 있는데, 그중 환경의 큰 변화에 있어서 정확도가 매우 불안정하기 때문이다. 그 중 labelling pose invariant feature가 하나의 solution이지만 대규모적인 환경과 데이터가 포함이 된다면 학습 할 알고리즘에 있어서 연산 량과 속도가 매우 떨어지게 되고 비용이 매우 커지게 된다. 이러한 문제점을 DCGAN을 이용하여 손실되는 고주파 정보를 복구하고 이외의 시각적 품질을 유지 및 향상시킴으로 문제점을 해결을 한다. 이는 이미지 뿐 아니라 활자 인식에서도 사용되는 방법으로써, 네트워크에 다양한 data를

학습을 시키게 되면 사용자보다 풍부하고 정교한 이미지나 언어 등을 사용이 가능하다. 대부분의 데이터 셋은 CelebA[6] 데이터 셋을 사용을 하는데 200,000개로 구성이 되어 있으며 유명인사와 여러 다양한 인물이 포함된 대규모 얼굴 속성 데이터이다. dataset 이미지는 얼굴 (그림 4)뿐 아니라 다양한 자세도 포함이 되어 있어서 학습에 매우 유용하며 (그림 5), 풍부한 특수 효과를 바라볼 수 있다. 그리고 얼굴 인식 및 표현의 학습 데이터 뿐 아니라 다른 컴퓨터 비전 작업에서도 사용이 가능 하다[7,8,9].

[그림 4] Asian face dataset



[그림 5] Various angles

4. 실험

4-1. 데이터 모델 샘플

CelebA dataset에서 교육 된 모델을 샘플로 보여준다. DCGAN을 이용해서 (그림 4)를 통해서 훈련된 모델의 샘플을 보여준다. 학습된 데이터와 비교를 위해 Vincent van gogh의 인물 자화상으로 130장의 data로 상정하였다. Vincent van gogh의 paintings방법이나 900여 점의 그림들과 1100여 점의 습작 등 중에서 자화상 및 인물의 표현들이 비슷하게 분포 되어 있기에 상정하였다.

4-2. 실험 환경

GPU 1080Ti를 사용하여 epoch 80,000로 돌렸다. CelebA의 데이터와 비교 분석 위해 같은 Hyper-parameter (표 2)로 학습을 시켰다. 이미지 사이즈로 178*218의 구성하여 학습을 시켰다.

[표 2] Hyper-parameter

parameter	value
learning rate	0.0002
batch size	64
beta1	0.5 (momentum term of adam)

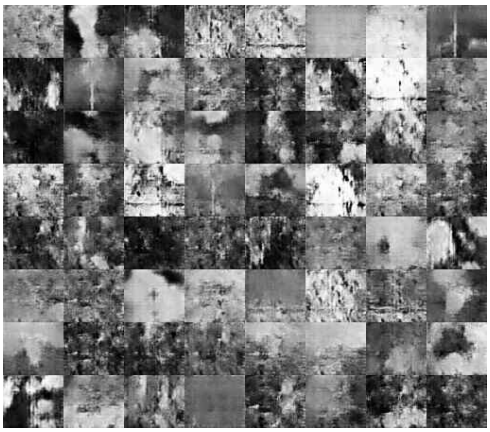
4-3. n-step 학습

학습은 1,000, 10,000, 50,000, 80,000번의 학습 시켰다 (표 3). Epoch 1,000 step (그림 6)이 지난 후의 loss function의 average가 가지고 있는 값은 d_loss는 1 g_loss가 5일 때의 결과 이미지가 표현이 되지 않았음을 확인이 가능했다. 10,000 step (그림 7)에서부터 시작하여 어느 정도의 이미지가 형성이 되기 시작하였으며, 50,000 step (그림 8)으로 넘어가면서 이미지가 형성이 시작이 되었다. 단순한 생성 모델로 높은 샘플 이미지를 생성을 할 수가 있었으며, 적은 수의 데이터 중에서 다양성을 가진 이미지 데이터 분포를 사용했기 때문에 모델의 생성 확인이 가능하였으나 잠재적으로 낮은 확률로 이미지의 블러(Blur) 현상을 가지는 이미지도 포함이 되었다.

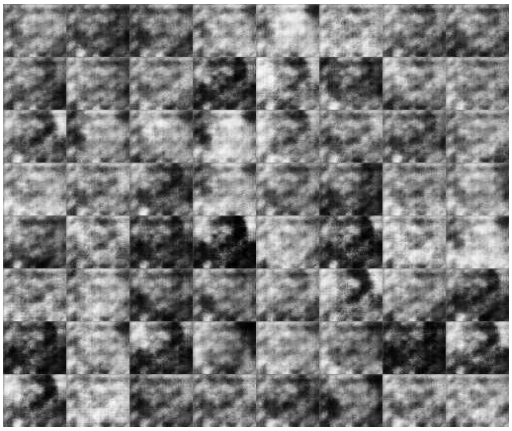
[표 3] learning to n-step

step	d_loss (average)	g_loss (average)
1,000	1	5
10,000	0.09	4.4
50,000	0.0013	4.1
80,000	0.0006	3.8

6) http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html Large-scale CelebFaces Attributes (CelebA) Dataset



[그림 6] learning to 1,000 step



[그림 7] learning to 10,000 step

[그림 8] learning to 50,000 step

[그림 9] learning to 80,000 step

5. 결론

본 연구는 DCGAN을 이용하여 얼굴에 관한 데이터를 생성하는데 후기 인상주의의 대표적인 화기인 Vincent van gogh의 인물 자화상을 학습을 시켰다. 전체적으로 원본 이미지와 비슷한 현상을 띠었으나, 생성된 이미지의 패턴이 비슷한 양상을 보이는 것이 데이터 수가 적었기 때문이다. 또한 잠재적으로 낮은 확률로 이미지의 블러 현상을 가지는 이미지가 포함되었다. 적은 수의 데이터가 블러 현상을 가지고 왔으나, 이미지가 원본 이미지의 형상과 비슷한 것은 원본 이미지의 형태를 본떠 만들어지는 형상이기 때문이고 그 외의 배경이나 옷 등의 오브젝트들은 model에 의해서 틀려지고 색상이 변화 한다는 점에 있어서는 생성된 이미지로 판단을 할 수가 있다. 위와 같은 실험 방법을 토대로 더 많은 데이터 수와 분포를 가지고 있다면 더욱 다양한 모습과 여러 색상을 가진 이미지 데이터를 생성을 할 수 있다. Vincent van gogh 같은 이미지 뿐 아니라 위와 같은 실험 방법을 사용하여 더욱 풍부하고 다양한 디자인 이미지 데이터를 가지고 있고 이것을 활용할 환경을 구축한다면 실험 결과에서 보이는 단방향성인 얼굴 이미지에 국한 되지 않을 것이며 다양한 데이터를 수집을 할 수가 있을 것이다. 하나의 예로서 건축 디자인에 관련된 책자를 만든다고 가정을 하고 각 시대마다의 건축 디자인을 학습을 시킨다면 여러 건축 디자인을 새로 생성하여 결과를 도출할 것이다. 이 부분에 있어서 많은 데이터가 필요하기 때문에 위와 같은 결과보다 더욱 확실하게 변화된 결과를 얻을 수 있을 것이며, 다른 디자인도 사용이 가능하다.

참고문헌

1. Kingma, Diederik P., and Max Welling. "Auto-encoding variational bayes." arXiv preprint arXiv:1312.6114, 2013.
2. Goodfellow, Ian, et al. "Generative adversarial nets." Advances in neural information processing systems, 2014.
3. Sohl-Dickstein, Jascha, et al. "Deep unsupervised learning using nonequilibrium thermodynamics." arXiv preprint arXiv:1503.03585, 2015.
4. Denton, Emily L., Soumith Chintala, and Rob Fergus. "Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks." Advances in neural information processing systems. 2015.
5. http://yann.lecun.com/exdb/mnist/ THE MNIST DATABASE of handwritten digits
6. http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html Large-scale CelebFaces Attributes (CelebA) Dataset
7. Goodfellow, Ian. "NIPS 2016 tutorial: Generative adversarial networks." arXiv preprint arXiv:1701.00160, 2016.
8. Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." arXiv preprint arXiv:1511.06434, 2015.
9. Liu, Ziwei, et al. "Deep learning face attributes in the wild." Proceedings of the IEEE International Conference on Computer Vision. 2015.