

# IoT·인공지능·빅데이터의 실무응용 연구 1 코딩 기술서

## Various GANs and its Application : Face Aging

Song Ki Ung  
Department of Mathematical Sciences  
sk851@snu.ac.kr

### 콘텐츠 소개 (About Contents)

GAN(Generative Adversarial Network)은 딥러닝 기반의 대표적인 생성모델로서 주어진 Data의 분포를 학습하고 이로부터 실제와 같은 Data를 Sampling할 수 있는 모델의 학습을 목표로 한다. 그리고 오늘날에는 더 나아가 Style Transfer, Self-Supervised Learning, Anomaly Detection과 같은 여러 도메인의 Task에 적용되기 위한 많은 GAN Architecture들이 여러 연구를 통해 제안되고 있다. 본 프로젝트에서는 **다양한 GAN 모델들을 살펴보고 이를 Face Aging Task에 적용하여 그 결과를 확인하고자 하였다.**

### 1. Introduction

GAN(Generative Adversarial Network)은 주어진 Data의 분포를 학습하고 이로부터 실제와 같은 Data의 Sampling을 목표로 하는 대표적인 딥러닝 기반의 생성모델이다. Style Transfer, Self-Supervised Learning, Anomaly Detection과 같은 여러 도메인의 Task에 GAN을 적용하기 위한 여러 연구들이 활발히 이루어지고 있으며 본 프로젝트에서는 Face Aging Task에 GAN 모델을 적용하여 그 결과를 확인하고자 하였다.

#### 1.1. GAN

Goodfellow et al.(2014)에 의해 처음 제안되었던 생성 모델인 GAN(Generative Adversarial Network)은

10년이 채 되지 않은 짧은 기간 동안 비약적인 발전을 이루었다. GAN은 Figure 1에서 확인할 수 있듯이 Random noise  $z$ 가 주어졌을 때 이로부터 실제와 같은 가짜 Data  $G(z)$ 를 생성하는 Neural Network인 Generator  $G$ 와  $G$ 가 생성한 가짜 Data와 실제 Data를 구분하는 Neural Network인 Discriminator  $D$ 로 구성된다.  $G$ 와  $D$ 의 Loss function을 순차적으로 최소화하는 과정을 통해 학습이 진행되며 그 과정에  $G$ 는 실제와 같은 Data를 Sampling할 수 있는 성능을 가지도록 기대하게 된다. 학습의 안정성과 같은 문제점 해결과 Conditional input의 활용, 고해상도 Data의 생성과 같은 추가적인 목적을 위해 cGAN, WGAN, DCGAN, StyleGAN, CycleGAN과 같은 다양한 Architecture의 GAN 모델들이 제안되고 연구 중에 있다.

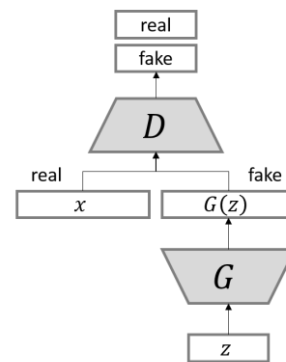


Figure 1: 기본적인 GAN Architecture

## 1.2. Face Aging

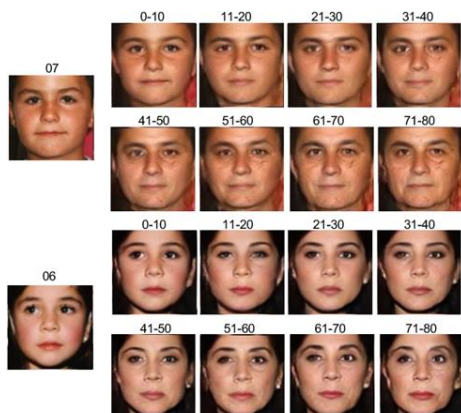


Figure 2: Face Aging Task

Face Aging Task는 Figure 2 와 같이 특정 연령대의 인물 얼굴이 Input Image로 주어졌을 때 이를 원하는 연령대로 바꾸는 Task를 의미한다. 이러한 Task는 단순히 즐거움만을 위한 Entertainment의 영역이 아니다. 예컨대 오래된 사진을 담고 있는 신분증을 소지한 사람의 신원 확인을 위한 Cross-Age Face Recognition에 Face Aging Task를 활용할 수 있을 것이다. 또한 KIST에서 개발하여 활용 중인 PoliSketch 모델처럼 실종 아동을 찾는 데 있어 Face Aging Task를 활용하여 효과적으로 실종자의 현재 연령대의 예상 얼굴을 생성하여 문제 해결에 도움을 줄 수 있을 것이라 생각된다. 이에 본 프로젝트에서는 GAN 모델을 기반으로 이러한 Face Aging Task를 과연 효과적으로 수행할 수 있을지를 확인하고자 하였다.

## 1.3. Dataset : UTKFace

Task를 위한 모델 학습에는 서버를 통해 주어진 Computing Resource의 성능과 GAN 모델 학습의 불안정성에 따른 충분한 학습 수 확보를 종합적으로 고려하여 상대적으로 크기가 작은, 약 100MB 정도의 "UTKFace - Large Scale Face Dataset Cropped" 데이터셋을 활용하여 학습을 진행하였다. 해당 데이터셋은 약 23000 여장의 이미지로 구성되며 1 세부터 100 세 이상의 다양한 연령대를 담고 있다. 또한 앞서 논의한 Computing Resource의 한계로

우선은 64 x 64 의 작은 이미지 사이즈를 위주로 다양한 모델 학습이 이루어졌다. 또한 16 세 이하의 청소년과 영유아 연령대의 경우 데이터셋의 크기가 충분하지 않아 학습이 유의미하게 이루어지지 않음을 확인했기에 Face Aging에 있어 대상이 되는 연령대를 17-29 세 / 30-39 세 / 40-59 세 / 60+세 의 4 개의 그룹으로 나누어 학습이 진행되었다.

## 1.4. FID(Fréchet Inception Distance)

서로 다른 GAN 모델들의 성능을 효과적으로 비교하여 평가하기 위해서는 정량적인 평가지표가 필요하다. 본 프로젝트에서는 이를 위한 정량적인 평가지표로 GAN이 생성한 Data와 실제 Data의 차이를 계산하는 FID(Fréchet Inception Distance)를 사용하였다. 이는 Pretrain된 Inception Network를 사용하여 실제 Target Data와 GAN이 생성한 Data의 Feature를 Network의 중간 Layer에서 계산을 하여 거리를 구하는 평가 지표이다. FID 값이 작을수록 실제 이미지와 생성된 이미지가 분포적으로 가까운 것을 의미하기에 GAN 모델의 좋은 성능을 의미한다.

## 2. Methodology

Face Aging Task를 수행하는 모델의 학습을 위해 본 프로젝트에서는 크게 4 가지의 서로 다른 GAN Architecture를 활용하고자 하였으며 3 가지 서로 다른 모델들을 Code로 구현하여 학습을 진행하였다.

### 2.1. Conditional GAN based approach

원하는 연령대로 Face Aging을 수행하기 위해서는 해당 연령대를 Conditional Input으로 하여 Data를 생성하는 GAN 모델을 1 차적으로 생각할 수 있다. Antipov et al.(2017)은 기존 GAN에 Conditional Input을 더하여 Data를 생성하기 위한 Conditional GAN(cGAN)을 활용한 Face Aging 모델을 구성하였다. Face Aging Task에 있어서는 Input 이미지의 얼굴을

적절한 연령대로 변화시키는 것도 중요하지만 동시에 Input Face의 Identity가 그대로 유지되는 것도 매우 중요한 부분이다. cGAN만을 활용한 Antipov et al.(2017)이 제안한 모델에서는 이러한 Identity를 유지하기 위해 Face Recognition을 수행하는 별도의 지도학습 모델을 학습하여 GAN의 Data 생성을 위해 사용되는 Feature를 최적화하는 과정에 사용하는 비효율성이 발생하였다. 따라서 본 프로젝트에서는 이러한 지도학습을 필요로 하는 비효율성을 줄이기 위해 GAN Architecture만을 활용할 수 있는 방법을 탐색하고자 하였다.

## 2.2. CycleGAN based approach

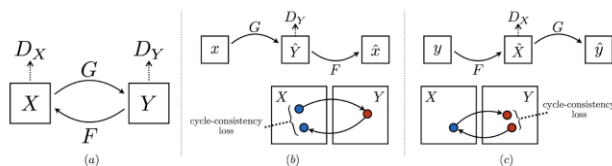


Figure 3: CycleGAN Architecture

Zhu et al.(2017)에 의해 제시된 CycleGAN의 경우 Figure 3 에서 확인할 수 있듯이 X 도메인에서 주어진 Data를 Y 도메인으로 옮긴 후 이를 다시 X 도메인으로 옮겨 기존 Data와의 거리를 Cyclic Loss term으로 활용하게 된다. 위와 같이 Cyclic Loss를 활용하여 2 개의 Generator로부터 학습이 이루어지는 형태를 통해 cGAN 모델을 활용할 때와 달리 지도학습 기반의 추가 모델을 필요로 하지 않으면서도 Face Aging에 있어 Identity를 유지할 수 있을 것이라 생각하여 CycleGAN을 Face Aging 모델로 이용하였다.

## 2.3. StarGAN based approach

CycleGAN만을 활용한다고 할 때 4 개의 연령대가 주어진 경우 임의의 A 연령대에서 B 연령대로 얼굴을 바꾸기 위해서는 총 12 개의 Generator, 즉 서로 다른 CycleGAN을 6 번 학습해야 하는 비효율이 발생한다. 이를 개선하기 위해 cGAN과 같이 Conditional Input을 활용하여 임의의 연령대로의 Translation을

수행하기 위한 Architecture 변형의 필요성이 있었다. Choi et al.(2018)에 의해 제안된 StarGAN 모델은 cGAN의 장점과 CycleGAN의 장점을 결합하여 Face Aging에서 Identity를 유지하기 위한 별도의 지도학습 모델이 필요없이 Single Generator로부터 Translation 하려는 연령대를 Conditional Input으로 하여 Face Aging Task의 학습이 가능할 것이라 판단하여 이를 CycleGAN의 개선 모델로서 학습시켜 비교해 보았다.

## 2.4. StarGAN based approach with Self-Attention

앞서 논의한 StarGAN만을 활용했을 경우 학습 데이터에 있어서는 유의미한 Face Aging 결과를 확인하였지만 Test Data에 대해서는 유의미한 Face Aging이 이루어지지 않음을 확인하였다. 일반적으로 GAN 모델의 학습에 있어 생성하고자 하는 Image의 Resolution이 크면 클수록 이는 Discriminator에서 유리하게 작용하며 반대의 경우는 Generator에게 유리하게 작용하는 경향이 있다. 본 프로젝트에서는 Resource의 한계로 인해 64 x 64 의 작은 사이즈 이미지로 학습이 진행되었기에 이는 Discriminator의 학습에 불리한 상황이었다. 이렇듯 제한된 사이즈의 이미지에서 Convolution만을 통해 Local한 범위로 제한된 연산은 Global한 Image 구조 학습에 더욱 장애가 될 수 있어 효율적인 학습 모델 구성을 위해 SAGAN(Self-Attention GAN)의 Self-Attention 모듈을 Network Architecture에 추가하였다.

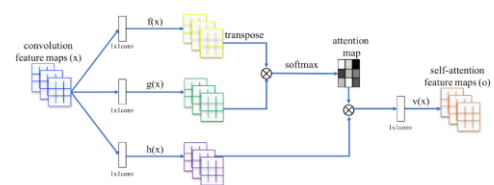


Figure 4: Self-Attention Module

Figure 4 는 사용된 Self-Attention 모듈의 Pipeline 을 나타낸다. Self-Attention은 자연어 처리 분야에서 제안된 개념으로 이를 주어진 Task에서의 과정으로 생각한다면 특정 얼굴의 영역, 예컨대 왼쪽 눈을 합성

혹은 구분한다고 할 때 이와 가장 관련되어 있는 영역, 예컨대 오른쪽 눈 영역에 Attention을 주는 형태로 이는 Image 구조의 전체적인 영역을 중요도와 함께 고려할 수 있도록 하는 기대효과가 있기에 이를 활용하였다.

### 3. Code Implementation

전체 Code는 총 4 개의 ipynb 파일들로 구성된다. CycleGAN, StarGAN, StarGAN with Self-Attention 모델을 각각 담고 있는 3 개의 ipynb 파일과 각 모델의 FID를 계산하기 위한 1 개의 ipynb 파일로 구성된다.

#### 3.1. CycleGAN based approach

Methodology의 2.2 에서 논의하였던 CycleGAN을 활용한 Face Aging 모델을 학습하고 확인하기 위한 Code는 CycleGAN.ipynb 파일에서 확인할 수 있다. CycleGAN은 두 개의 연령대가 주어졌을 때 이를 활용하여 학습을 진행하므로 서로 다른 두 연령대의 이미지가 담겨 있는 경로를 각각 설정해주면 된다.

#### 3.2. StarGAN based approach

비슷하게 2.3 에서 논의하였던 StarGAN을 활용한 Face Aging 모델의 Code는 StarGAN.ipynb 파일을 통해서 확인할 수 있다. StarGAN은 4 개의 연령대를 한 번에 학습에 활용할 수 있는 구조이기에 각각의 연령대의 이미지들을 담고 있는 폴더들의 상위 폴더의 경로를 학습 데이터의 경로로 설정하여 학습을 진행시킬 수 있다.

#### 3.3. StarGAN based approach with Self-Attention

마지막으로 2.4 에서 논의된 StarGAN 구조에 Self-Attention 모듈을 활용한 모델의 학습을 위한 Code는 StarGAN\_Self-Attention.ipynb 파일에서 확인할 수 있다. 학습 데이터의 경로의 설정은 StarGAN의 경우와 동일하다.

그러나 Self-Attention 모듈을 추가한 것만으로는 Test Data에 대해 CycleGAN을 넘어서는 성능이 얻어지지 못하고 학습 과정에서 Mode Collapse가 계속해서 발생하는 것을 확인할 수 있었다. 다양한 모델의 Parameter를 조절하여 실험한 결과, 4 개의 서로 다른 연령대를 Single Discriminator로 구분하고 Single Generator로 생성해야 하는 학습상황에서 64 x 64 의 작은 사이즈 이미지로 인해 Discriminator의 성능이 근본적으로 제한된 상황은 Self-Attention 모듈의 추가만으로는 해결할 수 없다고 판단하게 되었다. 또한 Sun et al.(2020)이 GAN 모델의 학습에서 Mode Collapse에 대한 이유를 Loss Surface의 모양을 통해 수학적으로 설명하고자 한 부분에서 아이디어를 얻어 기존의 GAN Loss에서 Discriminator와 관련된 부분을 Hinge Loss로 교체하여 학습에 적용하고자 했다.

따라서 128 x 128 크기로 학습 이미지의 사이즈를 개선하고 또한 Hinge Loss를 사용하여 모델 학습을 진행하였고 그 결과 64 x 64 크기로 학습 이미지를 구성할 때보다 Test Data에 있어 훨씬 더 우수한 결과를 정량적인 측면과 정성적인 측면에서 모두 확인할 수 있었다. 코드 상에서는 Hinge Loss를 활용하는 부분과 기존의 StarGAN Loss를 활용하는 부분을 둘 다 확인할 수 있다.

#### 3.4. Performance Evaluation

학습이 이루어진 각각의 3 가지 모델들 : CycleGAN, StarGAN, StarGAN with Self-Attention의 정량적인 평가지표로서 FID score를 계산하기 위한 Code는 Evaluation.ipynb 파일에서 확인할 수 있다. 학습된 모델의 Test를 위해서는 CycleGAN이 학습 데이터에 대해 가장 정성적으로 좋은 성능을 보였던 17-29 세 연령대에서 60+ 이상 연령대로의 Face Aging을 비교를 위한 기준점으로 설정하였다. 이를 위해 UTKFace 데이터셋이 아닌 한국 20 대 연예인들의 사진을 구글을 통해 Test Data로 확보하였다. Test

Data가 있는 폴더의 경로와 각 모델의 Weight 파일이 저장된 경로를 지정하면 각 모델별로 FID score가 계산되어 정량적인 지표를 확인함과 동시에 Face Aging의 결과 이미지 역시 확인하여 정성적인 비교도 가능하도록 구성되었다.

#### 4. Conclusion

본 프로젝트를 통해 Face Aging Task를 수행하는 GAN 모델을 학습해보고 그 결과를 확인하고자 했으며 단순한 CycleGAN에서부터 Self-Attention과 StarGAN을 함께 활용한 모델에 이르기까지 3 가지 종류의 모델을 다루었다. 모델 학습의 Trial and Error를 통해 Test Data에 대해 Generalized된 성능의 Face Aging Task를 수행하기 위해서는 학습 데이터의 Resolution이 커질 필요성을 확인하였고 이를 Mode Collapse와 관련한 논문을 통한 Hinge Loss의 도입과 결합하여 모델의 성능이 향상되는 결과를 확인할 수 있었다. 크기가 100MB에 불과한 작은 학습 데이터로 이러한 결과를 확인했다는 점 역시 인상 깊다고 할 수 있다. 그렇기에 현재보다 Computing Resource가 확보되는 상황에서 지금까지 논의된 부분을 64 x 64 크기의 작은 이미지가 아닌 더 큰 Resolution의 이미지, 예컨대 256 x 256 혹은 512 x 512 사이즈의 이미지로 더 큰 학습 데이터를 구성하여 모델들의 학습을 진행한 후 비교해 본다면 더욱 명확한 논의를 할 수 있을 것으로 기대된다.

#### References

- [1] Antipov, G., Baccouche, M., & Dugelay, J.-L. (2017). Face Aging With Conditional Generative Adversarial Networks.
- [2] Choi, Y., Choi, M., Kim, M., Kim, J. H. S., & Choo, J. (2018). StarGAN : Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation.
- [3] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014). Generative Adversarial Nets.
- [4] Sun, R., Fang, T., & Schwing, A. (2020). Towards a better global loss landscape of GANs. ArXiv, (NeurIPS).
- [5] Zhang, H., Goodfellow, I., Metaxas, D., & Odena, A. (2018). Self-Attention Generative Adversarial Networks.
- [6] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. Proceedings of the IEEE International Conference on Computer Vision, 2017-Octob, 2242–2251.