유미의 생성모델

CUAI 4기 모빌리티 A팀
조원(전자전기공학부), 이보림(소프트웨어학부), 최명균(전자전기공학부)

Abstract

현실적으로 많은양의 data를 크롤링하여 좋은 model을 학습시키는 것은 어렵다. 이의 해결책으로 많은 사람들은 많은 양의 dataset으로 학습된 pretrained model을 finetuning하는 방법을 사용하고있다. 가장 뜨거운 model인 GAN model중, latent vector를 control 하여 세부적인 feature들을 조종할 수 있게 만든 pretrained 된 다양한 stylegan model들을 활용하여 사람들이 관심을 가질만한 소량의 인기 웹툰의 얼굴 data를 가지고 다양한 model들을 train 해보았다.

1. Introduction

현재 Generative adversarial network(GANs)는 Computer vision의 가장 뜨거운 주제이며 다양한 task에서 뛰어난 성능을 보이고 있다. 다만, 몇몇 GANs들을 train하기 위해서는 많은 고화질의 data를 필요로 하며, 수렴하기 까지 매우 많은 시간과 자원이 소모된다.

보통 이러한 경우 transfer learning을 사용하여 data와 source의 부족함을 해결한다. 본 프로젝트에서는 새로운 mapping network를 사용하여 이전의 GANs에서 불가능 했던 scale-specific control을 가능하게 만들었던 StyleGAN과 이로부터 파생된 다양한 model들 사이의 성능 차이에 주목하였다. FFHQ, AFHQ, LSUN, CelebA-HQ와 같이 공개된 대규모 데이터셋으로 pre-trained된 모델을 가지고 제한된 데이터셋에서 얼마나 좋은 결과를 낼 수 있는가를 비교해 보았다. 추가적으로 다양한 augmentation 기법을 실험하여 overfitting을 막을 수있는 augmentation 조합을 탐색하였다. 새로운 mapping network를 사용하여 이전의 GANs에서 불가능 했던 scale-specific control을 가능하게 만들었던 StyleGAN과 이로부터 파생된 다양한 model들을 사용하여 이들의 차이에 주목하였다.

2. Related Work

1) StyleGAN & FreezeD

StyleGAN

고해상도 데이터 생성 학습이 어려웠던 기존의 GAN 의 문제점을 해결하기 위해 다른 새로운 학습방법을 적용시켰던 PGGAN[1]은 Generator 와 Discriminator 의 해상도를 점진적으로 늘려가는 방식으로 결과 sample image 의 해상도를 향상시켜 더욱 정교한 image 들을 생성해냈다.

그러나 GAN 의 image 생성과정은 아직 blackbox 였고, 어떻게 이미지가 합성되는지, 각각의 feature 들의 원인이 무엇인지에 대한 이해는 부족하였다. 또한 기존의 model 들은 normalized 된 latent vector 들이 바로 input 으로 들어가게 되어 학습이미지의 분포가 고정된(ex : Gaussian)분포에 non-linear 하게 mapping 되어 input vector 로 visual attribute 를 조절하기가 매우 어려웠다. 이러한 문제를 새로운 mapping network 로 해결한 model 이 바로 StyleGAN 이다. Input vector 들이 여러 개의 Fully-connected layer 들을 거쳐 다른 공간에 mapping 되어 더욱 유동적인 latent space 를 만들어 visual attribute 조절을 가능하게 만들었다.

FreezeD

StyleGAN 은 1024 x 1024 의 70,000 개 image 를 가지는 FFHQ 로 학습되었다. FreezeD 는 이러한 pretrained weight 를 바탕으로 효과적인 fine-tunning 방법론을 제시하였다. 단순히 간단하게 Discriminator 의 앞쪽 layer 들의 weight 만을 freeze 시킴으로써 이전의 fine-tuning technique 들의 성능을 압도하였다.

2) StyleGAN2

StyleGAN2 에서는 지난 버전에서 몇 가지 특징적인 결함들을 분석했고, 이에 대한 방법으로 모델 구조와, 훈련 방법에 있어 변화를 제안하였다. StyleGAN2 에서는 Generator 의 normalization 부분을 재디자인하였고, Progressive Growing 대신 skip connection 을 갖고 있는 hierarchical generator 를 사용하여 눈과 이의 stagnation 을 줄었다. 또한 model 의 용량 문제를 해결하고, 더욱 큰 model 을 사용하여 추가적인 품질을 향상시켰다.

3) StyleMapGAN

StyleMapGAN 은 실시간으로 이미지를 인코더로 latent space 에 projection 하고 로컬에서 이 이미지를 조작한다. StyleMapGAN 에서 Base Generator 는 spatial dimensions 로 실제 이미지를 projection 하고 local editing 을 가능하게 하여 stylemap 을 만들고 feature maps 과 stylemaps 사이 resolution 을 맞추기

위해 convolution 과 upsampling 이 들어있는 stylemap resizer 에 넣는다.

StyleMapGAN Discriminator 는 StyleGAN2 와 유사하며 Encoder 에서 stylemap 을 reconstruct 하며 MSE 로 에러를 매긴다. 이외에 다양한 evaluation metrics 를 사용하며 짧은 시간에 높은 성능을 가질 수 있게 하였다.

Reconstruction 이외에 W interpolation, Local editing, Unaligned transplantation, Random Generation, Style Mixing, Semantic Manipulation 과 같이 다양한 mixing type 을 이전 다른 모델들보다 더 높은 성능으로 보여준다.

3. Main

1) Dataset

성공적인 GAN의 fintuning을 위해서 dataset 선정 과정에서 몇가지 제한사항을 두었다. 가장 먼저 얼마나 많은 등장인물이 나오는 가이다. 두 번째로는 해당 웹툰만의 분별력 있는 스타일을 가지고 있는 가이다. 마지막으로 프로젝트에 사용할만한 인지도와 인기를 가지고 있는 가이다.

이러한 조건을 만족하는 유미의 세포들이라는 유명 웹툰을 선정하여 data들을 크롤링 하였다. 세포와 사람의 얼굴을 나누어 균등한 분포를 띄게끔 크롤링 하였다. Image의 scale은 통일하지 않았으며, Resize를 통해 256*256 scale로 바꾸어주었다. Resize 과정에서 비정상적인 aspect ratio를 가지는 image들은 원본의 모양이 심하게 훼손되어 학습을 저해하기 때문에 padding을 사용하여 비정상적인 aspect ratio를 가지는 image들을 정비율로 바꾸어 주는 방법을 시도하였지만, 배경이 padding의 색을 지정하는 부분에서 학습에 저해되는 요소가 발생하여 사용을 지양하였다. 최종적으로 600개의 face image들을 수집하여 학습에 사용하였다.

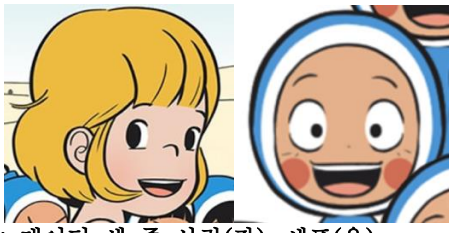
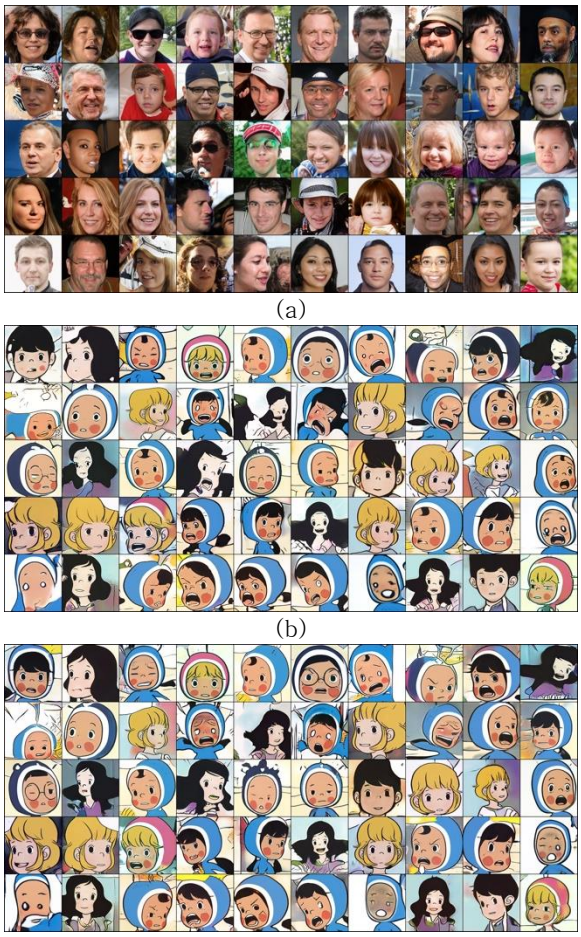


Fig 1: 데이터 셋 중 사람(좌), 세포(우)

2) Model

StyleGAN + FreezeD

가장먼저 StyleGAN과 FreezeD라는 finetining 기법을 활용하였다. FFHQ dataset으로 pretrained된 StyleGAN을 사용하였고, 유미의 세포들 dataset으로 fine-tuning하였다. 사람, 세포 2개의 class 모두를 사용하였고, 각각의 class들은 300개의 sample을 포함하고있다. 우리는 256x256의 resolution으로 pretrained된 model weight파일을 사용하였으며, 50000 iter까지 학습을 진행하였다.



(a) (b) (c)

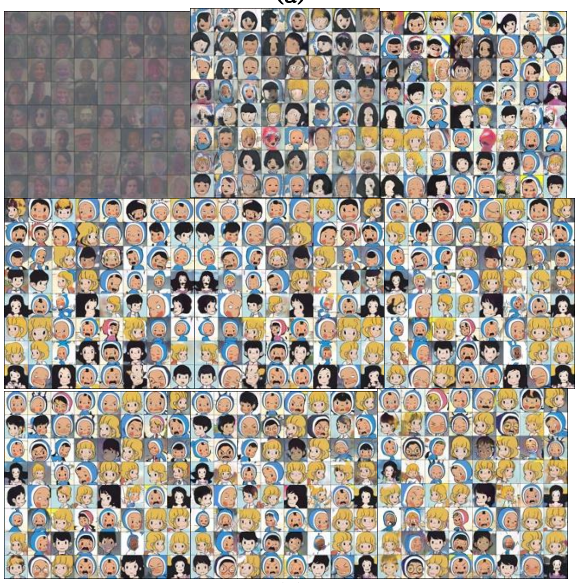
Fig 2: StyleGAN과 FreezeD으로 학습한 generator로 부터 생성된 samples.
(a) : 0 iter. (b) : 25000 iter. (c) : 50000 iter.

Figure 2에는 pretrained된 model의 generator에서 생성된 sample 이미지와, 유미의 세포들 data로 50000 iter 학습한 generator의 sample image이다.
꽤나 괜찮은 품질의 이미지들이 생성되었음을 확인할수 있다.

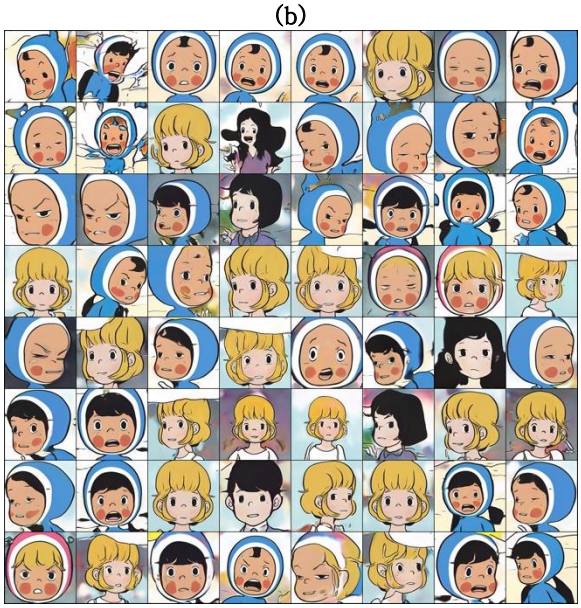
StyleGAN2

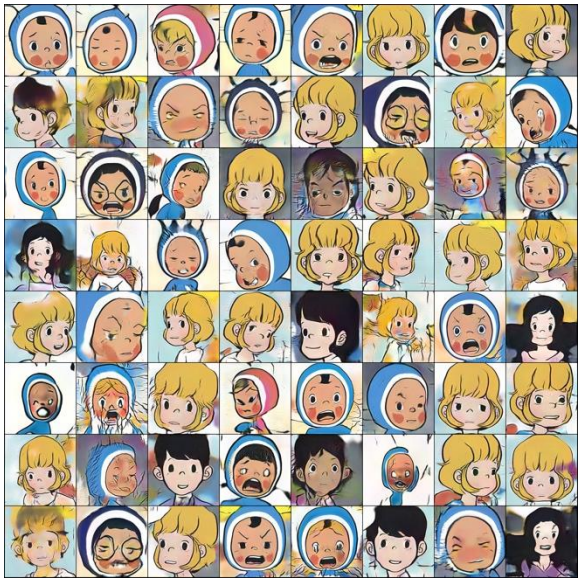
StyleGAN2를 활용한 training은 2가지 방향으로 나누어 진행하였다.
1. Pretrained weight을 사용하지 않고 noise 부터 image를 generate. 2. ffhq 이미지로 Pretrained된 model을 fintunning with FreezeG

(a)



(b)





(c)

Fig 3 (a) : train from base. (b) : finetuning with pretrained model(FFHQ) (c) : results

2가지 실험 모두 50000iter 까지 진행하였으며, 이외의 모든 실험조건은 동일하다.

일반적인 딥러닝 모델의 통념과는 달리 noise부터 train시킨 model이 좀더 깔끔한 image를 생성해 내는것을 확인 하였다. 이러한 결과의 원인으로는 많은 것들을 유추해 볼 수있는데, 가장 큰 원인으로 생각되는것은 데이터의 다양성 부족이였다. Pretrained 된StyleGAN의 mapping network는 굉장히 various한 data를 학습하여 복잡한 분포를 띄고 있을텐데, 이에반해 우리의 학습데이터는 비슷한 세포의 이미지들과 한쌍의 남여 캐릭터 뿐이라 이 과정에서 학습과정에서의 저해가 발생한다고 추측하였다.

StyleMapGAN

StyleMapGAN 에서 제공해주는 pretrained networks 는 CelebA-HQ 와 AFHQ, FFHQ 이렇게 2 가지가 있다. CelebA-HQ, AFHQ 2 가지 이미지 데이터셋 256x256 크기의 이미지 20M image 로 pretrained 모델들을 가지고 모두 train 을 해보았다.

FreezeD 나 FreezeG(StyleGAN1, StyleGAN2) 에 비해 train 이 비교적 오래 걸렸으나, train 이후 generate 할 때는 빠른 속도를 보여주었다.

Figure 5 같이 유미의 세포들을 보았을 때, 특징들과 이목구비가 CelebA HQ 사람 데이터와 너무 거리감 있다고 생각되어 AFHQ 동물 데이터로 pretrained model 로 다시 training 하였다.

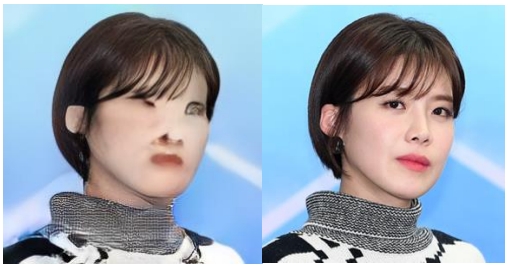


Figure 5 CelebA-HQ pre trained model 을 10000 iterations training 한 model 을 reconstruction 한 결과, 입력이미지(좌) 결과이미지(우)

(a)

(b)

Figure 6 (a) CelebA-HQ 데이터 (b)AFHQ 데이터

Fig 7 AFHQ pre-trained model 을 10000 iterations training 한 model random generation 한 결과

Figure 6,7 를 보아 AFHQ 데이터로 pretrained 된 모델을 학습시키면,앞서 예상했던 것과 같이 사람의 형태와 조금 떨어진 세포들은 잘 나오지만 사람의 모습과 가까운 유미와 같은 사람 캐릭터는 전보다 더 많은 에러를 가지고 생성된다는 것을 알 수 있다.

Fig 8,9 AFHQ pre-trained model 을 20000 iterations training 한 model random generation 한 결과

3) Train

Augementation

적은 data를 가지고 train해야 하기 때문에 다양한 augmentation 조합을 탐색하였다. 실험 결과 동일 조건에서 fid score과 좋은 sample generation을 나타내는 augmentation 조합을 찾아내었다. Cutout, blueness, RandomBrightness, CLAHE,

HorizontalFlip, GaussianBlur을 사용하였고. Geometric augmentation과 channel augmentation 등등 각 augmentation의 특징별로 나누어 Albumentation의 Oneof 함수를 적용하여 한번에 하나의 augmentation만을 적용해주도록 하였다.

Optimizer & lr & batch size

여러 optimizer를 실험해본 결과 AdamP와 AdamW를 사용하였때 좋은 수렴속도와 결과를 보여주었다.
Learning rate은 0.002를 사용하였고 별도의 scheduler은 적용해주지 않았다. Batch size는 32를 사용하였다.

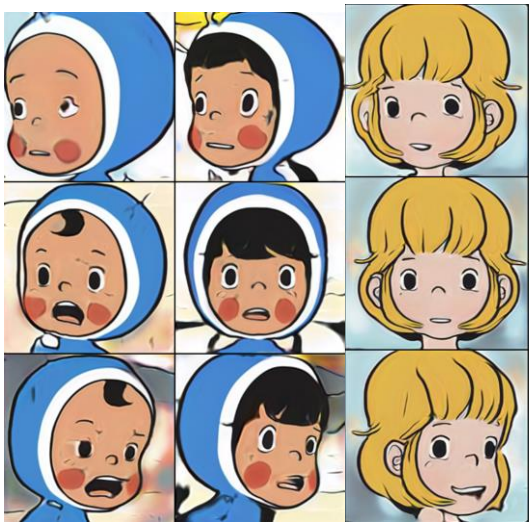
4) Apply

4.1) Control Visual Attribute

시각적으로 봤을때 가장 좋았던 StyleGAN2를 사용했던 model을 사용하여 몇가지 중요한 latent vector를 추출해보았다. 19개의 index로 이루어진 eigen vector를 사용하여 generator에 direction으로 넣어주었다.
Generation된 sample image들을 확인 해보면 각각의 eigen vector들이 어떠한 visual attibute를 조절해주는지 확인할 수 있었다.

Index 0을 조절해 주었을 때는 sample image들의 얼굴 각도가 변화 하였다. Index 5를 조절해 주었을 때는 세포의 이미지가 사람의 image로 변화하였다. Index 12를 조절해 주었을 때는 놀라는 표정을 가지는 이미지를 생성해 주었다.

이러한 실험을 통해 어떠한 latent vector 들이 어떠한 visual attribute를 조절해 주는지 알 수 있으며, 이로 인해 generator를 control 할 수 있게 되었다.



(a)

(b)

(c)

Fig10 Model에서 추출된 eigen vector에 따른 sample image (degree = ±5,10) (a) Index 0 (b) Index 5 (c) Index 12

4.2) StyleMixing

우리가 원하는 image를 학습된 data의 stytle에 맞게 transfer해주는 StyleMixing을 시도하였다. 사용자의 이미지를 projection 한 vector를 생성한뒤 이를 generator에 넣어 train 시켜준다.

이과정에서 사람의 사진을 조금 만화에 그림체에 맞게 변화 시켜주어야 generation이 잘될것이라고 판단하였다. FacialCartoonization을 사용하여

image를 만화에 가까운 style로 변형시켜준 뒤, 이를 StyleGAN2에 활용하였다.

Fig10 FacailCartoonization을 사용하여 이미지를 cartoonize 시킨 예시

하지만 기대와 달리 projection 결과가 좋지 못하였다.
유미의 세포들 data의 이미지들이 various 하지 못하기 때문에 사용자의 image를 잘 projecting 시키지 못하였다.
4.3) Local editing
StyleMapGAN 의 Local editing 을 가지고 세포들끼리 서로 합성하여 새로운 표정을 가지는 세포를 만드는 등 활용할 수 있다.

Fig 11 local editing 을 위한 reference image(왼), source image(오)

Fig 12 Fig 11 의 이미지들을 모델이 reconstruction 을 거친 결과

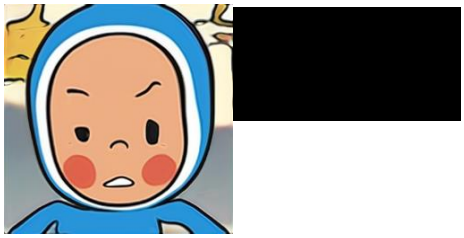


Figure 13 Figure 12 의 이미지들을 마스크 이미지(왼)에 맞게 합성한 결과(오)

Figure 11 은 local editing 에 입력할 데이터들이며 model 에서 합성하기 전 입력된 데이터들을 reconstruction 을 진행한다. Figure 12 들이 이에 대한 결과들이며 Figure 8 에 있는 mask image 에 따라 합성되어 새로운 표정을 가진 세포가 탄생한다.

3. Conclusion

이번 소논문에서는 세부적인 visual attribute를 조정할 수있는 StyleGAN base model들에 대하여 실험을 진행해 보았다. StyleGAN으로 부터 파생된 다양한 model들을 사용하여 상대적으로 적은양의 dataset을 가지고 각각의 model에 대한 효과적인 finetuning 방법과 효과적인 augmentation 방법들을 실험해 보았다.
다양한 feature 분포의 dataset이 구축되어있다면, 기존의 pretrained model을 finetuning 하는것이 효과적이지만, 이번 소논문의 data와 같이 various 하지 못한 data로 train 시킬시에는 noise 부터 학

습하는게 좋은 image들을 생성해 내었다.
학습된 best model들로 다양한 application을 진행해 보았다. Visual attribute를 control 하여 사용자가 원하는 feature를 가지는 이미지를 생성 하였으며, 2가지 data를 합성하여 새로운 image를 생성해 내기도 하였다.

참고 문헌

[1] Tero Karras, Samuli Laine, Timo Aila "A Style-Based Generator Architecture for Generative Adversarial Networks" CVPR Mar 2019
[2] Tero Karras "Analyzing and Improving the Image Quality of StyleGAN" Dec 2019
[3] Hyunsu Kim, Yunjey Choi, Junho Kim, Sungjoo Yoo, Youngjung Uh "Exploiting Spatial Dimensions of Latent in GAN for Real-time Image Editing" CVPR Apr 2021
[4]https://github.com/SystemErrorWang/FacialCartoonization
[5] https://github.com/rosinality/stylegan2-pytorch