경희대학교 데이터 분석 동아리

KHUDA 2023

StyleGAN2를 이용한 이미지 생성

한상진, 황동민

배경

배경

- '이 사람이 축구 게임 속 선수였다면 어떤 모습일까'라는 궁금증에서 시작.

-> StyleGAN2를 이용해 축구게임 속 이미지를 만들어 보자!

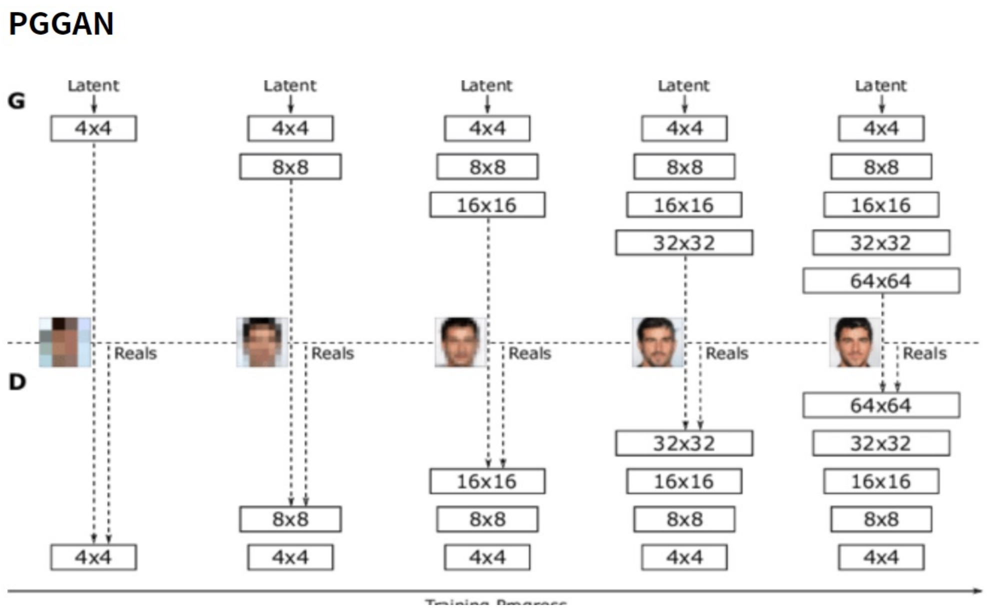
손흥민 선수와 fifa23 손흥민 선수 페이스온 이미지

https://youtu.be/S5ghaeeocCs

StyleGAN2

PGGAN

- 낮은 해상도부터 높은 해상도까지 차근차근 점진적으로 생성하는 대표적인 생성 모델.
- 이미지의 특징 제어가 어렵다는 한계점이 존재



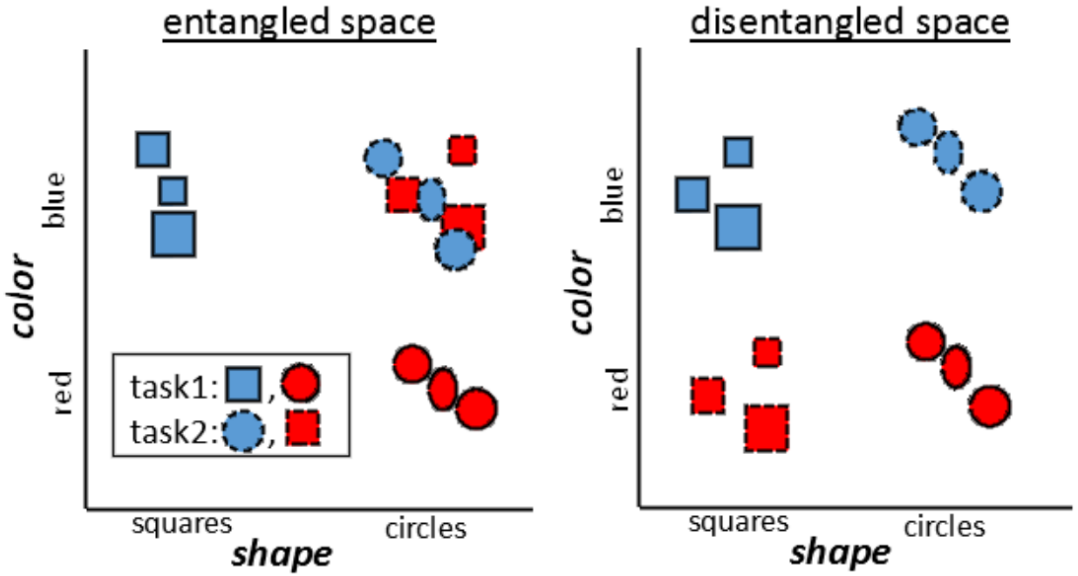
KHUDA

StyleGAN

- PGGAN을 바탕으로 고화질 이미지를 생성.
- Disentanglement 특성을 향상시켰다.

Entangle : 각 특징들이 서로 얽혀 있어서 구분이 안됨.

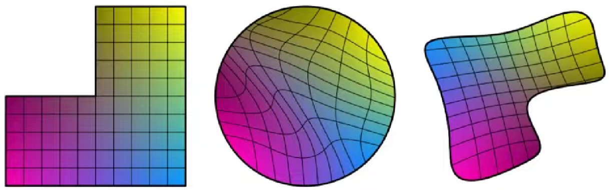
Disentangle : 다양한 특징들이 잘 분리되어 있다.



StyleGAN

StyleGAN의 핵심 아이디어: 매핑 네트워크 (Mapping Network)

- 512차원의 z 도메인에서 w 도메인으로 매핑을 수행합니다.
- 가우시안 분포에서 샘플링한 z 벡터를 직접 사용하지 않습니다.
  - 계산된 w 벡터를 사용할 때가 효과가 좋습니다.



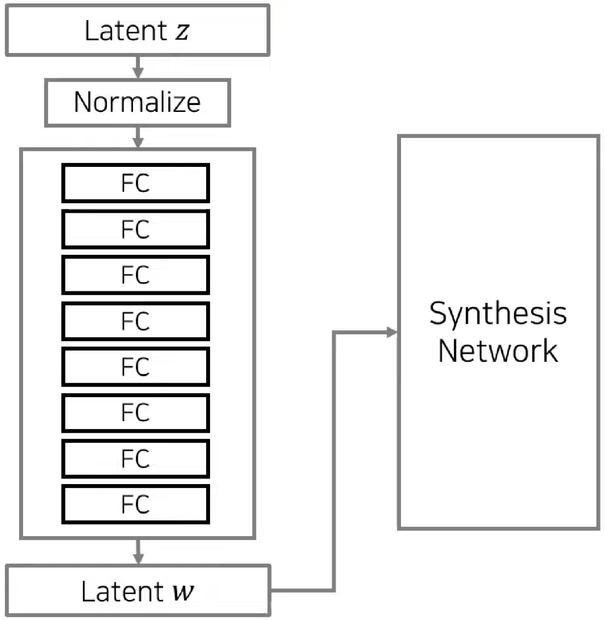
(a) Distribution of features in training set

(b) Mapping from $\mathcal{Z}$ to features

(c) Mapping from $\mathcal{W}$ to features

In W space, the factors of variation become more linear.

Z: Fixed distribution
Learned mapping $f: z \rightarrow w$



A Style-Based Generator Architecture for Generative Adversarial Networks (CVPR 2019)

나동빈

출처 : 나동빈 유튜브

KHODA

StyleGAN2 논문 리뷰

- 'Analyzing and Improving the Image Quatliy of StyleGAN'이라는 논문을 리뷰함.
- 목적 : StyleGAN에서의 blob-like artifacts, Phase artifact 를 고치고 Result Quality를 더 증가시킴.

Analyzing and Improving the Image Quality of StyleGAN

Tero Karras NVIDIA · Samuli Laine NVIDIA · Miika Aittala NVIDIA · Janne Hellsten NVIDIA
Jaakko Lehtinen NVIDIA and Aalto University · Timo Aila NVIDIA

arXiv:1912.04958v2 [cs.CV] 23 Mar 2020

Abstract

The style-based GAN architecture (StyleGAN) yields state-of-the-art results in data-driven unconditional generative image modeling. We expose and analyze several of its characteristic artifacts, and propose changes in both model architecture and training methods to address them. In particular, we redesign the generator normalization, revisit progressive growing, and regularize the generator to encourage good conditioning in the mapping from latent codes to images. In addition to improving image quality, this path length regularizer yields the additional benefit that the generator becomes significantly easier to invert. This makes it possible to reliably attribute a generated image to a particular network. We furthermore visualize how well the generator utilizes its output resolution, and identify a capacity problem, motivating us to train larger models for additional quality improvements. Overall, our improved model redefines the state of the art in unconditional image modeling, both in terms of existing distribution quality metrics as well as perceived image quality.

1. Introduction

The resolution and quality of images produced by generative methods, especially generative adversarial networks (GAN) [16], are improving rapidly [23, 31, 5]. The current state-of-the-art method for high-resolution image synthesis is StyleGAN [24], which has been shown to work reliably on a variety of datasets. Our work focuses on fixing its characteristic artifacts and improving the result quality further.

The distinguishing feature of StyleGAN [24] is its unconventional generator architecture. Instead of feeding the input latent code $\mathbf{z} \in \mathcal{Z}$ only to the beginning of a the network, the *mapping network* f first transforms it to an intermediate latent code $\mathbf{w} \in \mathcal{W}$. Affine transforms then produce *styles* that control the layers of the *synthesis network* g additional random noise maps to the synthesis network. It has been demonstrated [24, 38] that this design allows the intermediate latent space $\mathcal{W}$ to be much less entangled than the input latent space $\mathcal{Z}$. In this paper, we focus all analysis solely on $\mathcal{W}$, as it is the relevant latent space from the synthesis network's point of view.

Many observers have noticed characteristic artifacts in images generated by StyleGAN [3]. We identify two causes for these artifacts, and describe changes in architecture and training methods that eliminate them. First, we investigate the origin of common blob-like artifacts, and find that the generator creates them to circumvent a design flaw in its architecture. In Section 2, we redesign the normalization used in the generator, which removes the artifacts. Second, we analyze artifacts related to progressive growing [23] that has been highly successful in stabilizing high-resolution GAN training. We propose an alternative design that achieves the same goal — training starts by focusing on low-resolution images and then progressively shifts focus to higher and higher resolutions — without changing the network topology during training. This new design also allows us to reason about the effective resolution of the generated images, which turns out to be lower than expected, motivating a capacity increase (Section 4).

Quantitative analysis of the quality of images produced using generative methods continues to be a challenging topic. Fréchet inception distance (FID) [20] measures differences in the density of two distributions in the high-dimensional feature space of an InceptionV3 classifier [39]. Precision and Recall (P&R) [36, 27] provide additional visibility by explicitly quantifying the percentage of generated images that are similar to training data and the percentage of training data that can be generated, respectively. We use these metrics to quantify the improvements.

Both FID and P&R are based on classifier networks that have recently been shown to focus on textures rather than shapes [12], and consequently, the metrics do not accurately capture all aspects of image quality. We observe that the

1. Introduction

본 논문의 목적: StyleGAN에서의 Artifact를 고치고 Result Quality를 더 증가시킴.

기존 Stylegan에서 Artifacts에 2가지 문제가 있음

1. blob-like artifacts

- Generator가 설계적 결함을 피하기 위해 Blob-like artifact를 생성하는 것을 발견했기 때문에 section2에서 generator에 사용되는 normalization을 재디자인해서 artifact를 제거한다.

2. Phase artifact

- Progressive growing과 관련된 Artifact를 분석하고 전체적인 구조는 변화하지 않고 점진적으로 고해상도로 전환시키는 새로운 대안의 design을 제안.

Quantative analysis of quality of images와 관련해서

FID와 Precision and Recall은 Shape 보다는 Texture에 중점을 두기 때문에 이미지 품질의 모든 측면을 정확하게 평가하기 어렵다. PPL metric이 shape의 일관성 및 안정성과 관련이 있음을 관찰하였다. 따라서 PPL Metric을 규제로 사용할 예정.

2. Removing normalization artifacts

Stylegan에서 생성된 대부분의 이미지에서 blob-shaped artifact를 발견할 수 있다.

이는 AdaIN의 문제로 발생한다고 본다. -> AdaIN은 각 feature map의 평균/표준편차를 개별적으로 정규화해서 서로 연관된 feature들의 정보가 손실될 수 있음

StyleGAN2 논문 리뷰

- 'Analyzing and Improving the Image Quatliy of StyleGAN'이라는 논문을 리뷰함.
- 목적 : StyleGAN에서의 blob-like artifacts, Phase artifact 를 고치고 Result Quality를 더 증가시킴.

Figure 1. Instance normalization causes water droplet -like artifacts in StyleGAN images. These are not always obvious in the generated images, but if we look at the activations inside the generator network, the problem is always there, in all feature maps starting from the 64x64 resolution. It is a systemic problem that plagues all StyleGAN images.

Blob-like Artifacts

Phase Artifacts

KHODA

StyleGAN2 논문 리뷰

- 'Analyzing and Improving the Image Quatliy of StyleGAN'이라는 논문을 리뷰함.
- 목적 : StyleGAN에서의 blob-like artifacts, Phase artifact 를 고치고 Result Quality를 더 증가시킴.

Analyzing and Improving the Image Quality of StyleGAN

Tero Karras NVIDIA
Samuli Laine NVIDIA
Miika Aittala NVIDIA
Janne Hellsten NVIDIA
Jaakko Lehtinen NVIDIA and Aalto University
Timo Aila NVIDIA

arXiv:1912.04958v2 [cs.CV] 23 Mar 2020

Abstract

The style-based GAN architecture (StyleGAN) yields state-of-the-art results in data-driven unconditional generative image modeling. We expose and analyze several of its characteristic artifacts, and propose changes in both model architecture and training methods to address them. In particular, we redesign the generator normalization, revisit progressive growing, and regularize the generator to encourage good conditioning in the mapping from latent codes to images. In addition to improving image quality, this path length regularizer yields the additional benefit that the generator becomes significantly easier to invert. This makes it possible to reliably attribute a generated image to a particular network. We furthermore visualize how well the generator utilizes its output resolution, and identify a capacity problem, motivating us to train larger models for additional quality improvements. Overall, our improved model redefines the state of the art in unconditional image modeling, both in terms of existing distribution quality metrics as well as perceived image quality.

1. Introduction

The resolution and quality of images produced by generative methods, especially generative adversarial networks (GAN) [16], are improving rapidly [23, 31, 5]. The current state-of-the-art method for high-resolution image synthesis is StyleGAN [24], which has been shown to work reliably on a variety of datasets. Our work focuses on fixing its characteristic artifacts and improving the result quality further.

The distinguishing feature of StyleGAN [24] is its unconventional generator architecture. Instead of feeding the input latent code $\mathbf{z} \in \mathcal{Z}$ only to the beginning of a the network, the *mapping network* f first transforms it to an intermediate latent code $\mathbf{w} \in \mathcal{W}$. Affine transforms then produce *styles* that control the layers of the *synthesis network* g additional random noise maps to the synthesis network. It has been demonstrated [24, 38] that this design allows the intermediate latent space $\mathcal{W}$ to be much less entangled than the input latent space $\mathcal{Z}$. In this paper, we focus all analysis solely on $\mathcal{W}$, as it is the relevant latent space from the synthesis network's point of view.

Many observers have noticed characteristic artifacts in images generated by StyleGAN [3]. We identify two causes for these artifacts, and describe changes in architecture and training methods that eliminate them. First, we investigate the origin of common blob-like artifacts, and find that the generator creates them to circumvent a design flaw in its architecture. In Section 2, we redesign the normalization used in the generator, which removes the artifacts. Second, we analyze artifacts related to progressive growing [23] that has been highly successful in stabilizing high-resolution GAN training. We propose an alternative design that achieves the same goal — training starts by focusing on low-resolution images and then progressively shifts focus to higher and higher resolutions — without changing the network topology during training. This new design also allows us to reason about the effective resolution of the generated images, which turns out to be lower than expected, motivating a capacity increase (Section 4).

Quantitative analysis of the quality of images produced using generative methods continues to be a challenging topic. Fréchet inception distance (FID) [20] measures differences in the density of two distributions in the high-dimensional feature space of an InceptionV3 classifier [39]. Precision and Recall (P&R) [36, 27] provide additional visibility by explicitly quantifying the percentage of generated images that are similar to training data and the percentage of training data that can be generated, respectively. We use these metrics to quantify the improvements.

Both FID and P&R are based on classifier networks that have recently been shown to focus on textures rather than shapes [12], and consequently, the metrics do not accurately capture all aspects of image quality. We observe that the

1. Introduction

본 논문의 목적: StyleGAN에서의 Artifact를 고치고 Result Quality를 더 증가시킴.

기존 Stylegan에서 Artifacts에 2가지 문제가 있음

1. blob-like artifacts

- Generator가 설계적 결함을 피하기 위해 Blob-like artifact를 생성하는 것을 발견했기 때문에 section2에서 generator에 사용되는 normalization을 재디자인해서 artifact를 제거한다.

2. Phase artifact

- Progressive growing과 관련된 Artifact를 분석하고 전체적인 구조는 변화하지 않고 점진적으로 고해상도로 전환시키는 새로운 대안의 design을 제안.

Quantative analysis of quality of images와 관련해서

FID와 Precision and Recall은 Shape 보다는 Texture에 중점을 두기 때문에 이미지 품질의 모든 측면을 정확하게 평가하기 어렵다. PPL metric이 shape의 일관성 및 안정성과 관련이 있음을 관찰하였다. 따라서 PPL Metric을 규제로 사용할 예정.

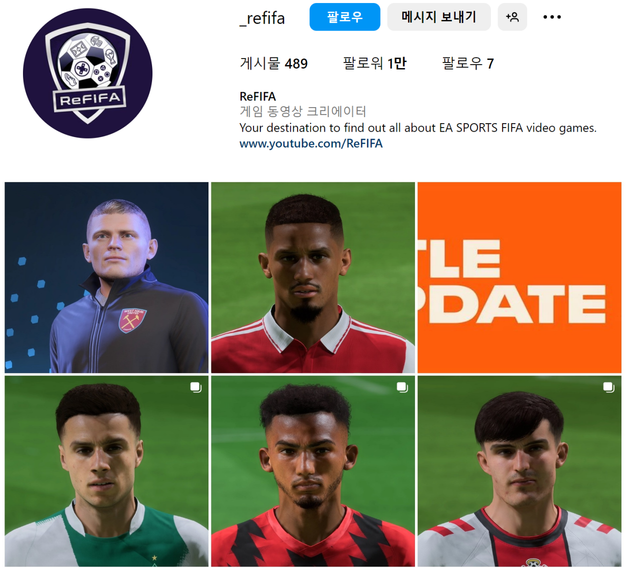
2. Removing normalization artifacts

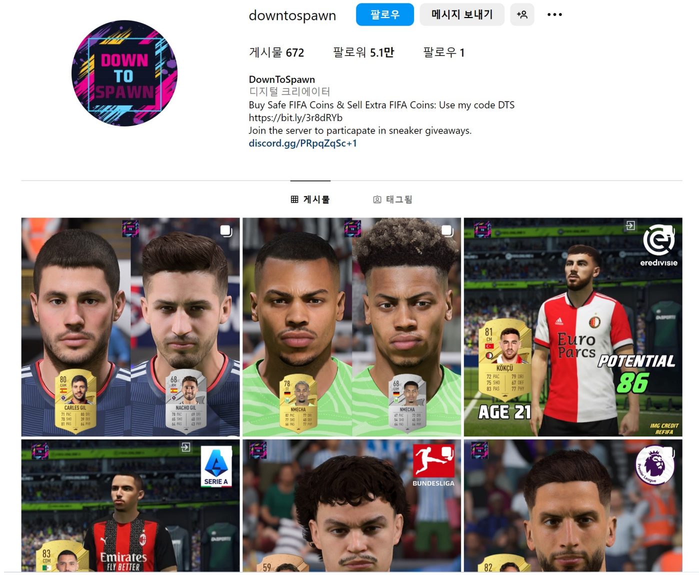
Stylegan에서 생성된 대부분의 이미지에서 blob-shaped artifact를 발견할 수 있다.

이는 AdaIN의 문제로 발생한다고 본다. -> AdaIN은 각 feature map의 평균/표준편차를 개별적으로 정규화해서 서로 연관된 feature들의 정보가 손실될 수 있음

데이터셋

-Selenium과 BeautifulSoup을 이용한 인스타그램 페이지 크롤링

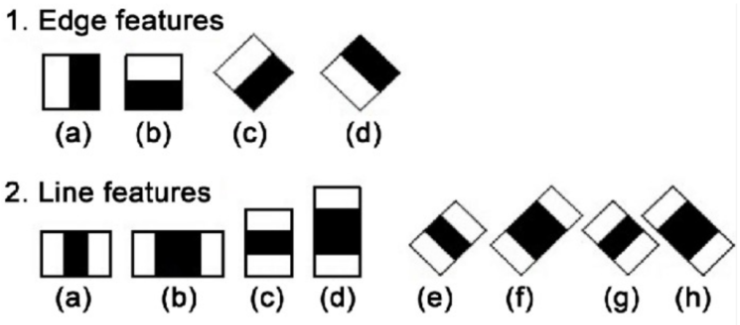




420장 정도의 축구선수 이미지 다운로드

데이터셋

Face Detection: Haar-based Cascade
사각형의 필터 마스크를 이용해 얼굴 인식



KHODA

데이터셋

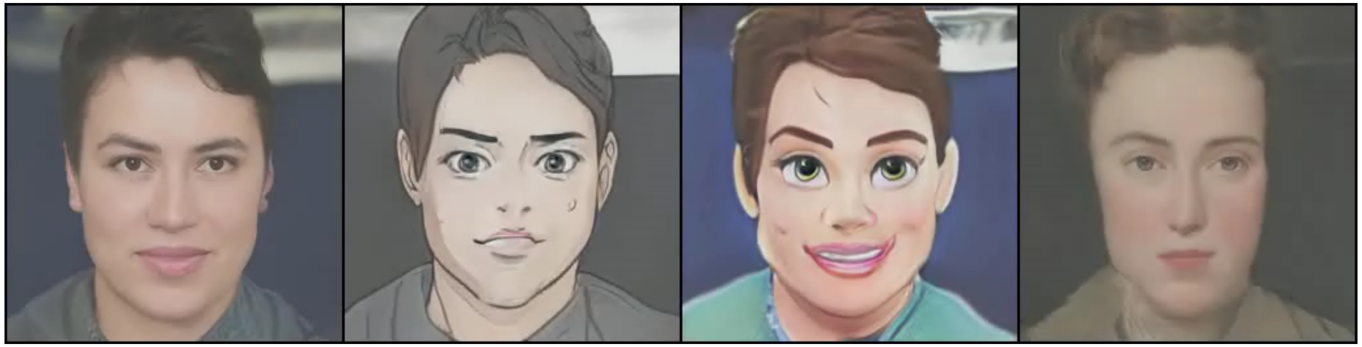
Python CV2 모듈에 내장되어 있음.

얼굴 검출(약 430장) -> 정사각형 형태로 이미지 분리 -> 1024*1024 크기로 resize

Transfer Learning

다양한 전이학습 기법: StyleGan2 – Cartoon

서로 다른 데이터 셋과 기법으로 학습된 모델



KHUDA

전이학습(Transfer Learning) – Freeze-D

- FFHQ를 통해 Pretrained된 StyleGAN2 모델을 이용해 학습을 진행
- Freeze-D : Custom한 데이터셋 혹은 다른 도메인의 데이터셋으로 학습할 때 Discriminator의 특정 layer를 freezing 시켜 학습 시키는 방법
- 기존의 전체 layer를 미세 조정하는 finetuning 방법보다 훨씬 더 우수한 성능을 나타냄

Freeze the Discriminator: a Simple Baseline for Fine-Tuning GANs

Sangwoo Mo
KAIST
swmo@kaist.ac.kr

Minsu Cho
POSTECH
mscho@postech.ac.kr

Jinwoo Shin
KAIST
jinwoos@kaist.ac.kr

Abstract

Generative adversarial networks (GANs) have shown outstanding performance on a wide range of problems in computer vision, graphics, and machine learning, but often require numerous training data and heavy computational resources. To tackle this issue, several methods introduce a transfer learning technique in GAN training. They, however, are either prone to overfitting or limited to learning small distribution shifts. In this paper, we show that simple fine-tuning of GANs with frozen lower layers of the discriminator performs surprisingly well. This simple baseline, FreezeD, *significantly outperforms previous techniques used in both unconditional and conditional GANs. We demonstrate the consistent effect using StyleGAN and SNGAN-projection architectures on several datasets of Animal Face, Anime Face, Oxford Flower, CUB-200-2011, and Caltech-256 datasets. The code and results are available at https://github.com/sangwoomo/FreezeD.*

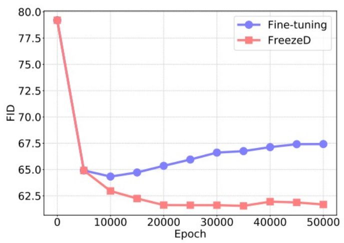


Figure 1: Trends of FID [15] scores of fine-tuning and our proposed baseline, FreezeD, on 'Dog' class in the Animal Face [36] dataset. While fine-tuning suffers from overfitting, FreezeD shows consistent stability in training GANs.

data and resources. Indeed, most of the recent success in

Transfer Learning 결과

Source vs Transfer

Source vs Transfer

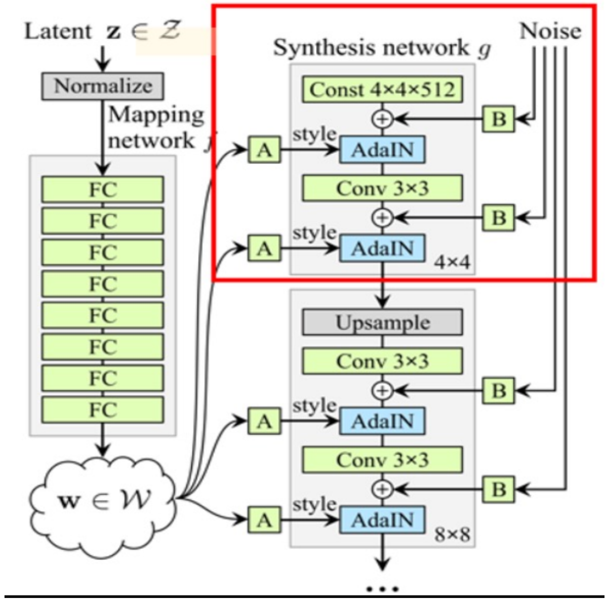
Source vs Transfer

KHODA

Source vs Transfer

Model Blending - Layer Swap

서로 다른 모델의 구조를 섞는 기법
초반 블록은 FFHQ로 학습된 기존 모델의 블록
그 이후의 블록은 축구 선수 이미지로 전이학습한 모델의 블록

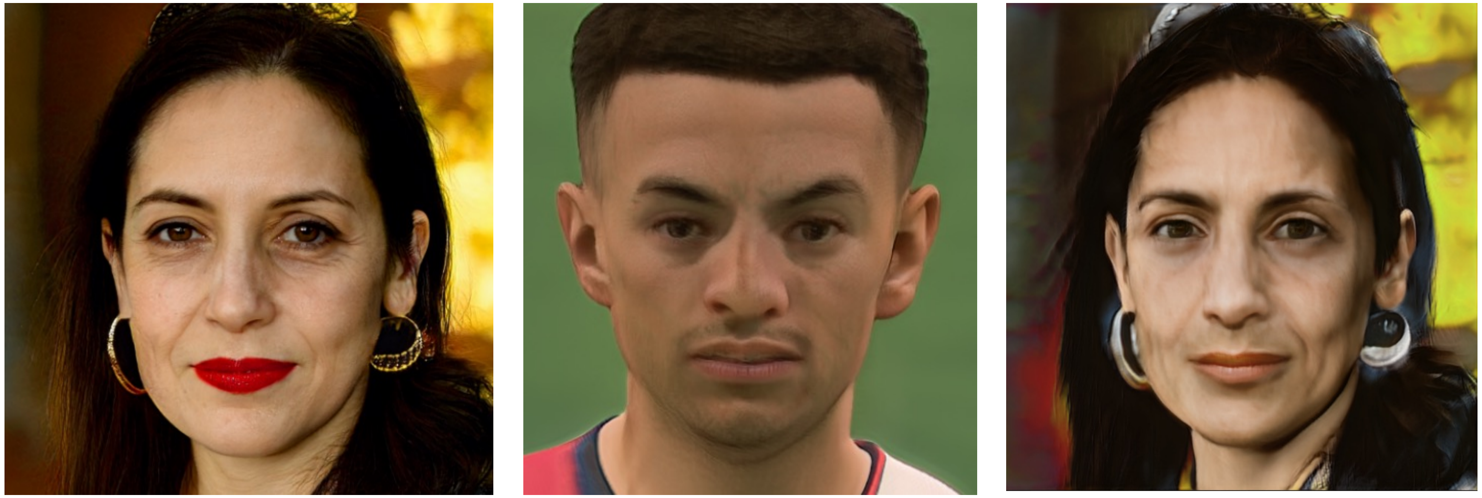


소스 이미지에서 얼굴의 전체적인 형태를 유지
(얼굴의 구조, 방향 등)

세부적인 표현은 게임 내 이미지와 비슷하게 함
(피부 묘사, 광원 등)

FFHQ : 앞의 4개의 블록(64*64)
Transfer model: 뒤의 5개의 블록

Source vs Transfer vs Blended



Source vs Transfer vs Blended

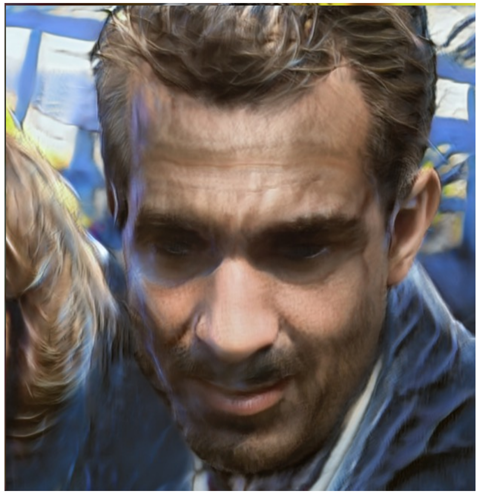


Source vs Transfer vs Blended

KHODA

Source vs Transfer vs Blended



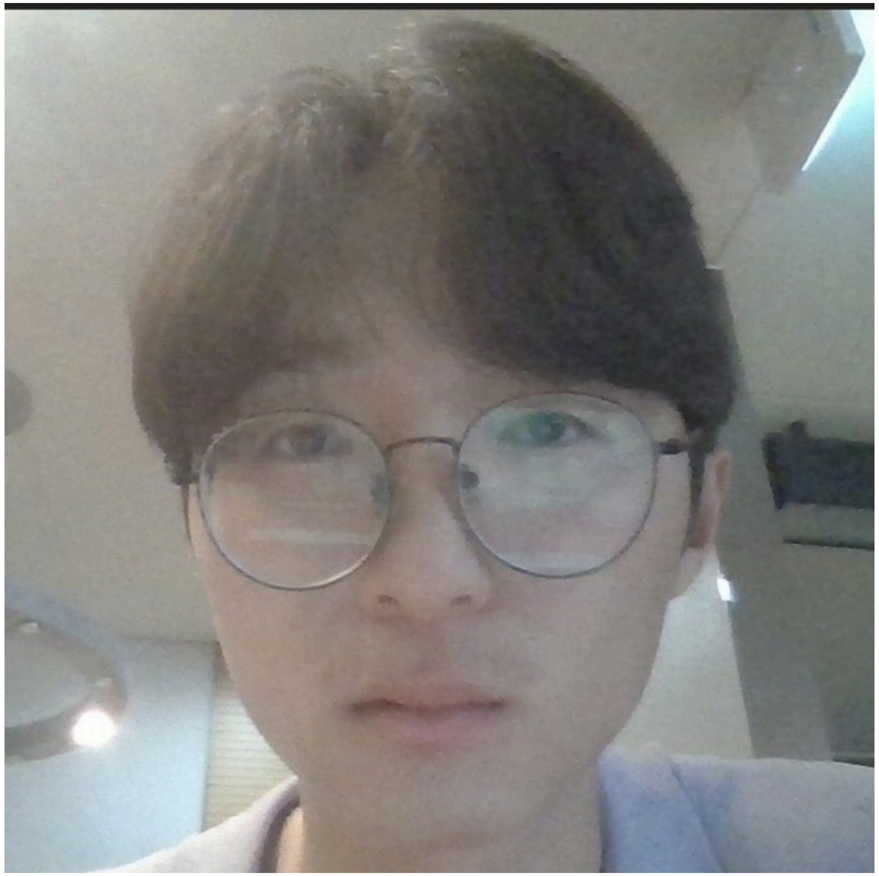
Model Blending - Layer Swap

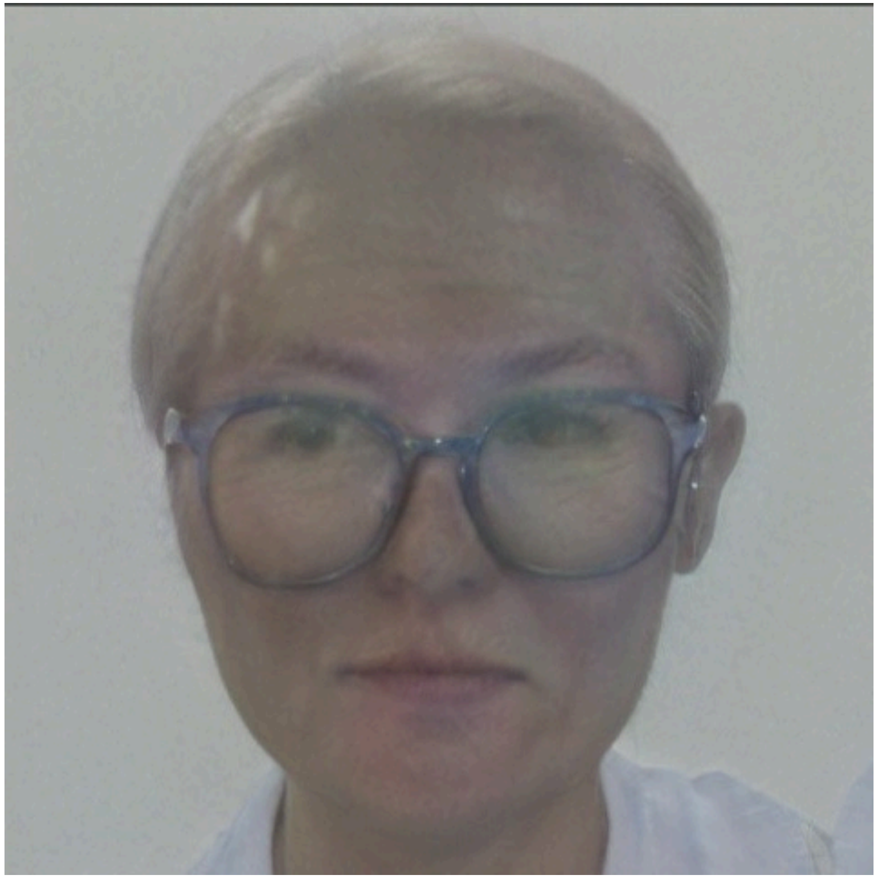
1. 남자 축구 선수 이미지 밖에 없음

   -> 여성 이미지가 입력되어도 남성으로 바꿈

2. 대부분이 짧은 머리 또는 빡빡이

   -> 머리카락에 대한 표현이 아쉬움





KHODA

ThanK You