



face chAlnge

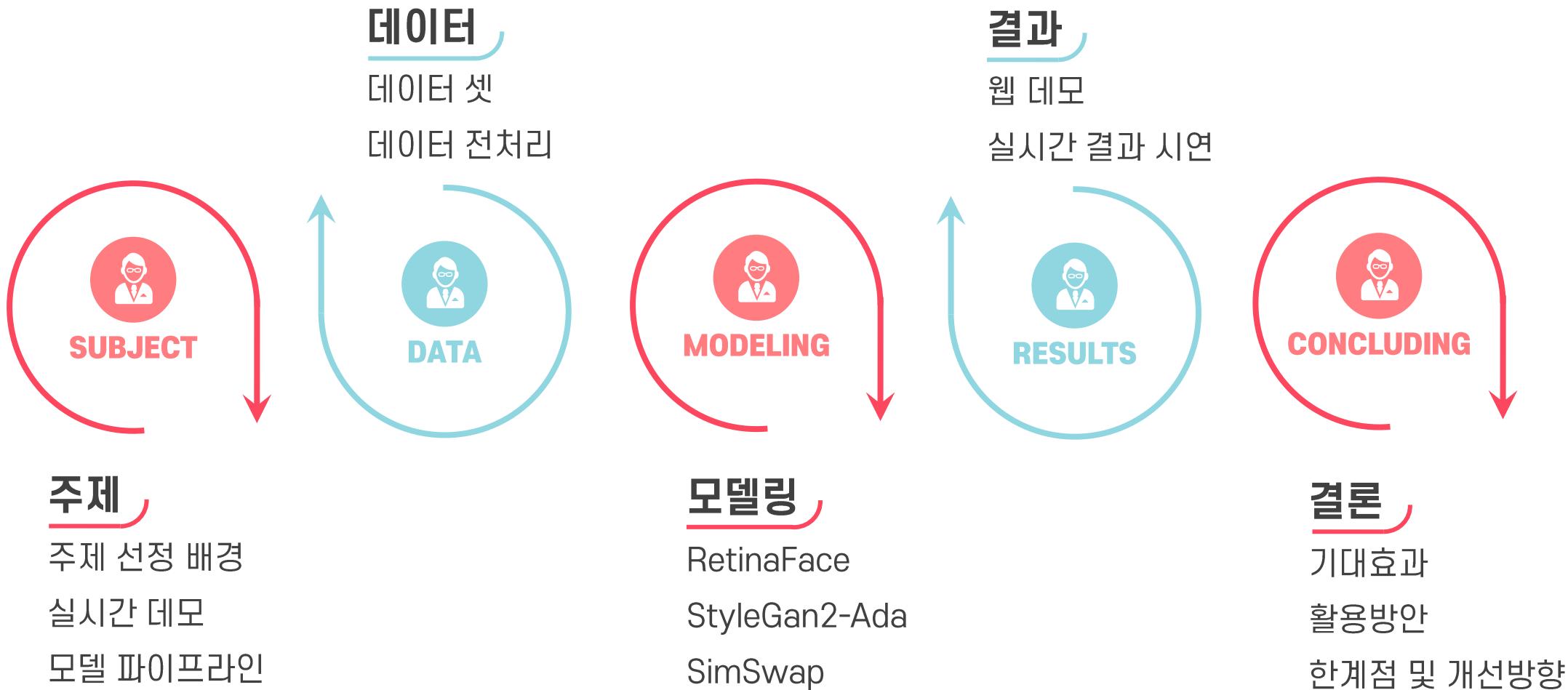
사진 속 배경인물 얼굴 변경

김상홍, 마민정, 송창용, 오주영, 조혁준





목차





SUBJECT



주제 선정 배경



포토샵을 통한
자르기, 모자이크, 블러



사람이 많은 장소 (ex. 길거리, 여행지, 공공장소 등)에서 찍은 사진
→ 초상권 문제로 다양한 처리를 해주어야 됨



세상에 존재하지 않는 다른 **사람의 얼굴로 대체해** 주는 것은 어떨까?



주제 선정 배경



딥러닝을 통한
다른 사람으로 대체



IDEA!



같은 성별, 비슷한 나이대의 얼굴을 만들어서 바꿔주자!



실시간 데모

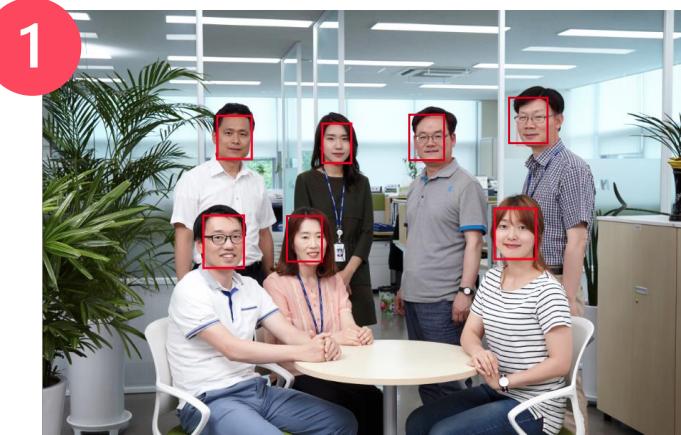


[실시간 데모]

1. QR 코드를 통해 오픈채팅방 접속
2. 배경에 다른 사람이 나온 사진 전송
3. 대체하지 않을 얼굴 선택



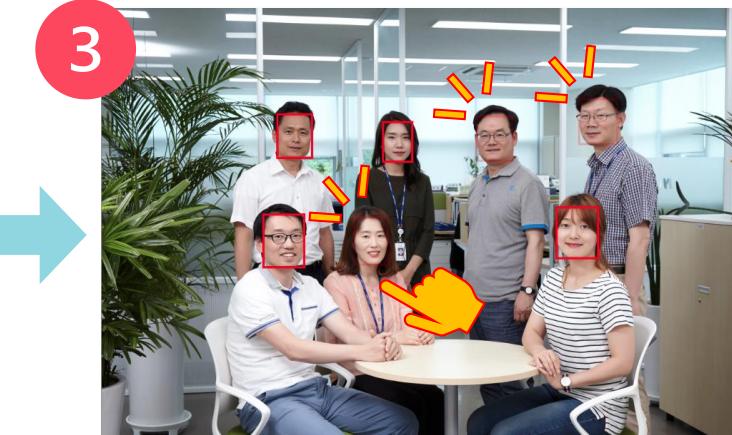
모델 파이프라인



[RetinaFace를 사용한 얼굴 탐지]



[Age, Gender Prediction]



[대체하지 않을 얼굴 선택]



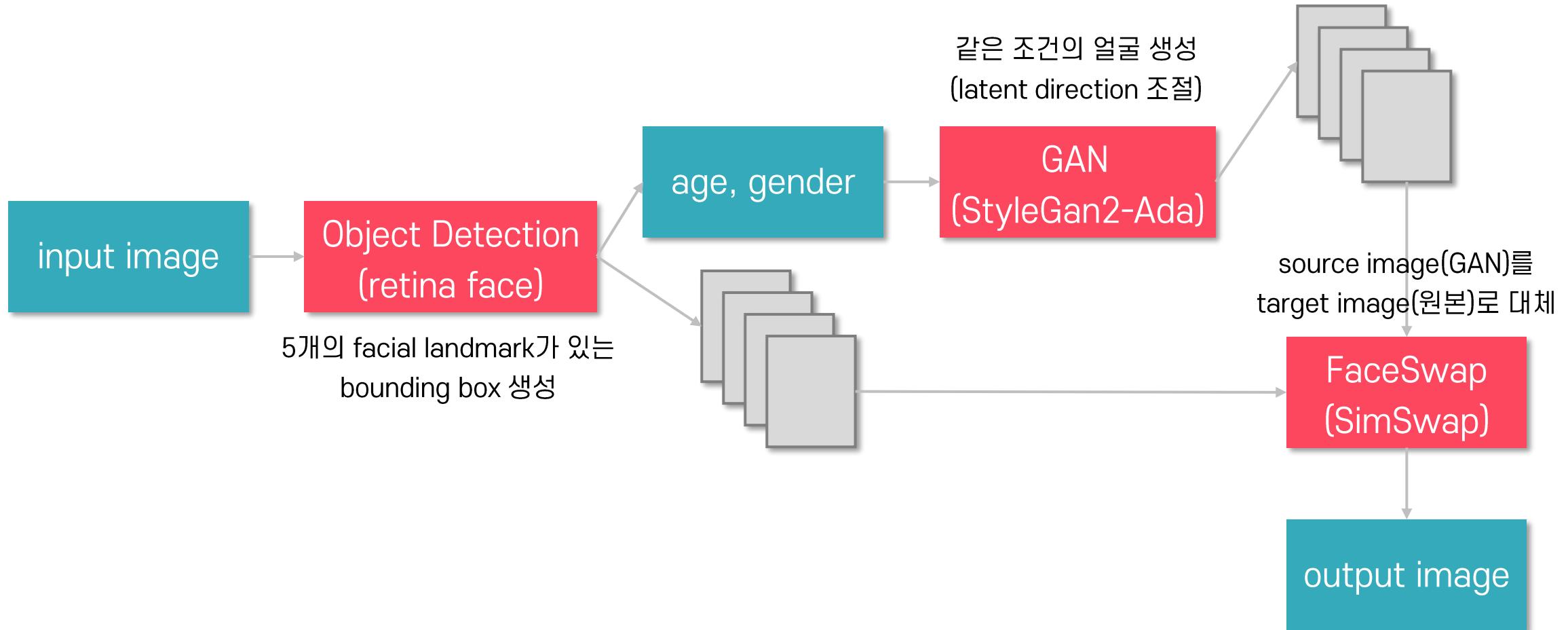
[SimSwap을 사용해 얼굴 대체]



[StyleGan2-Ada로 조건에 맞게 얼굴 생성]



모델 파이프라인





DATA



데이터셋 및 전처리



<http://shuoyang1213.me/WIDERFACE/>

WIDER FACE

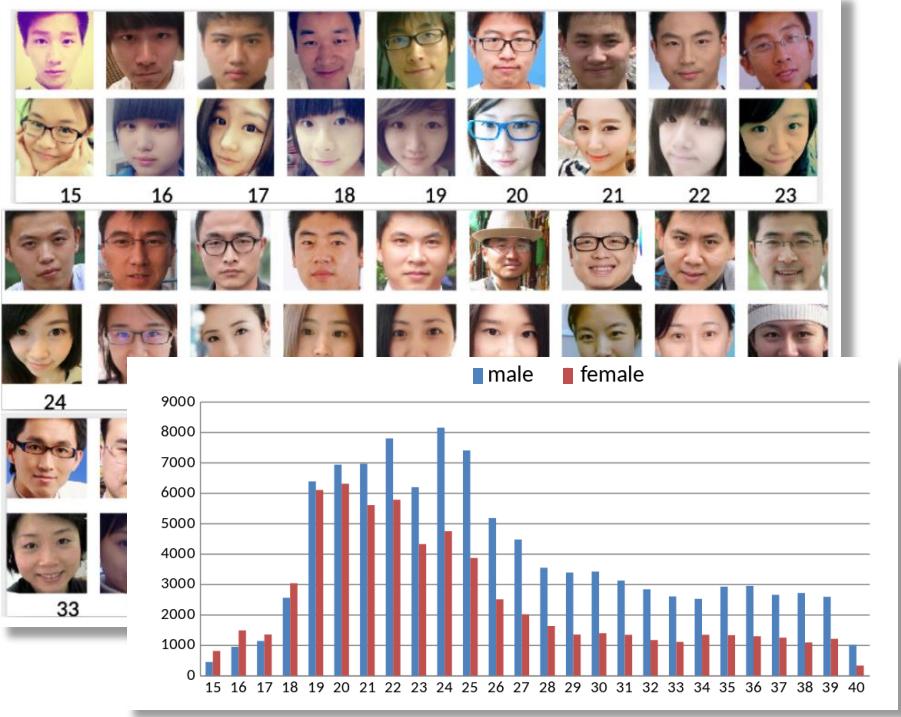
- face detection 분야의 benchmark 데이터
- 32,203장의 이미지와 393,703개의 얼굴
- 61가지 event classes

전처리 : RetinaFace

gender, age pretrained model 사용해서 widerface annotation에 gender, age pseudo label 추가



데이터 셋 및 전처리



AFAD(Asian Face Age Dataset)

나이와 성별 label이 있는 160K 얼굴 이미지 데이터
총 164,432장(female:63,680, male:100,752,
15-40세)



<https://pythonawesome.com/all-age-faces-dataset-mostly-asian>

AAFD(All Age Faces Dataset)

대부분 아시아인으로 구성된 얼굴 이미지 데이터
총 13,322장(female:7,381, male:5,941, 2-80세)

전처리 : StyleGan2-ada

image align

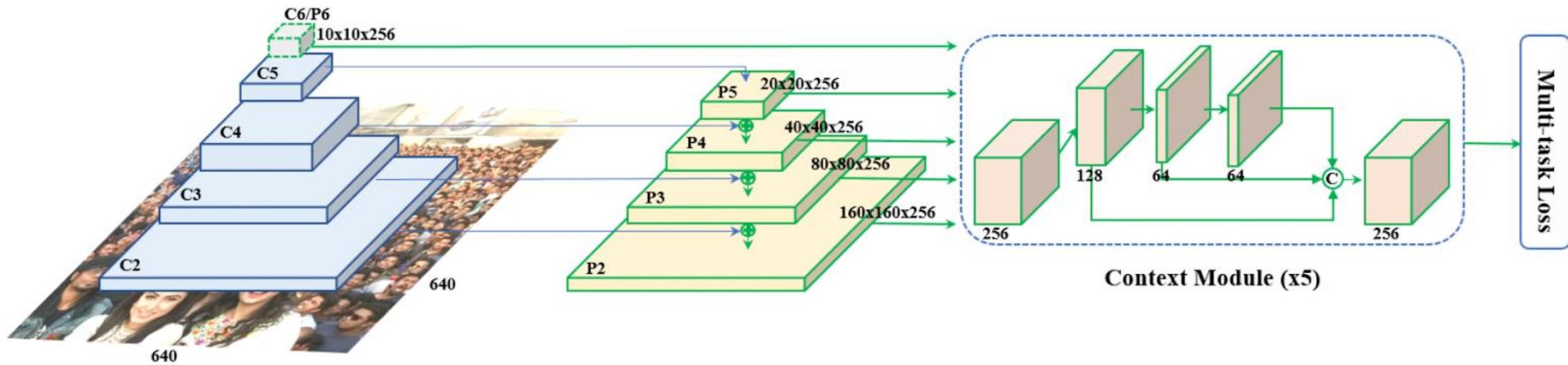


MODELING



모델링 : RetinaFace

RetinaFace란?



- ResNet + FPN을 거친 single stage object detection 모델
- 5개의 pyramid level에 개별의 context module을 거쳐 receptive field가 커지고, contextual reasoning power가 강화됨
- two-stage detector와는 달리 ROI를 sampling하지 않기 때문에 훨씬 빠른 inference 속도

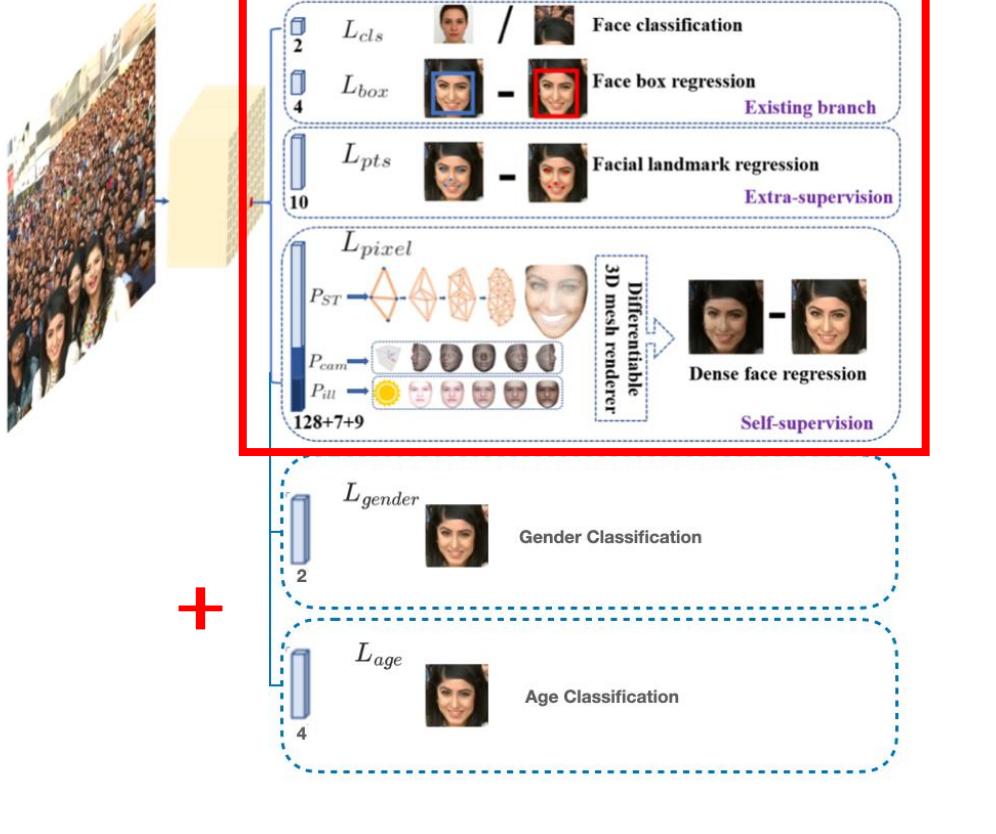


모델링 : RetinaFace

새롭게 구현한 RetinaFace



기존 RetinaFace



기존 RetinaFace 구조에 Gender Classification Head, Age Classification Head 추가

각 anchor에서 Face classification, bbox regression, facial landmark regression, gender classification age classification의 multi-task learning을 수행

- backbone은 더 robust한 feature를 추출
- 별도의 network 없이 face detection과 gender/age classification을 end-to-end로 수행



모델링 : RetinaFace

모델의 최적 Loss Function



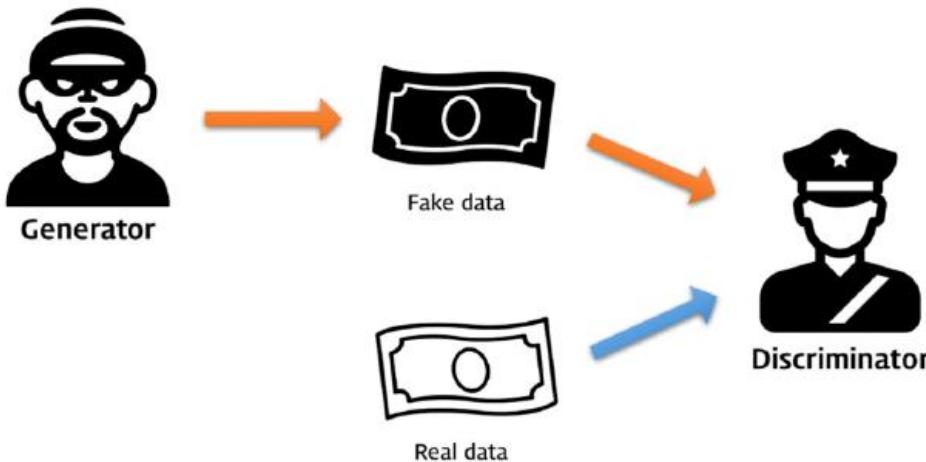
$$L = L_{cls}(p_i, p_i^*) + \lambda_1 p_i^* L_{box}(t_i, t_i^*) + \lambda_2 p_i^* L_{pts}(l_i, l_i^*) \\ + \lambda_4 p_i^* L_{gender}(g_i, g_i^*) + \lambda_5 p_i^* L_{age}(a_i, a_i^*).$$

1. face classification) softmax loss for binary classes (face / not face)
2. face box regression) 각 box의 크기가 상이, anchor별 target을 normaliz한 후 smooth-L1 loss 사용
3. facial landmark regression) target을 normaliz한 후 smooth-L1 loss 사용
4. gender classification) positive anchor에 대해 softmax loss for binary classification (male/female)
5. age classification) positive anchor에 대해 softmax loss for four classes
ex) <10, 10<&<30, 30<&<60, 60>



모델링 : StyleGan

GAN이란?

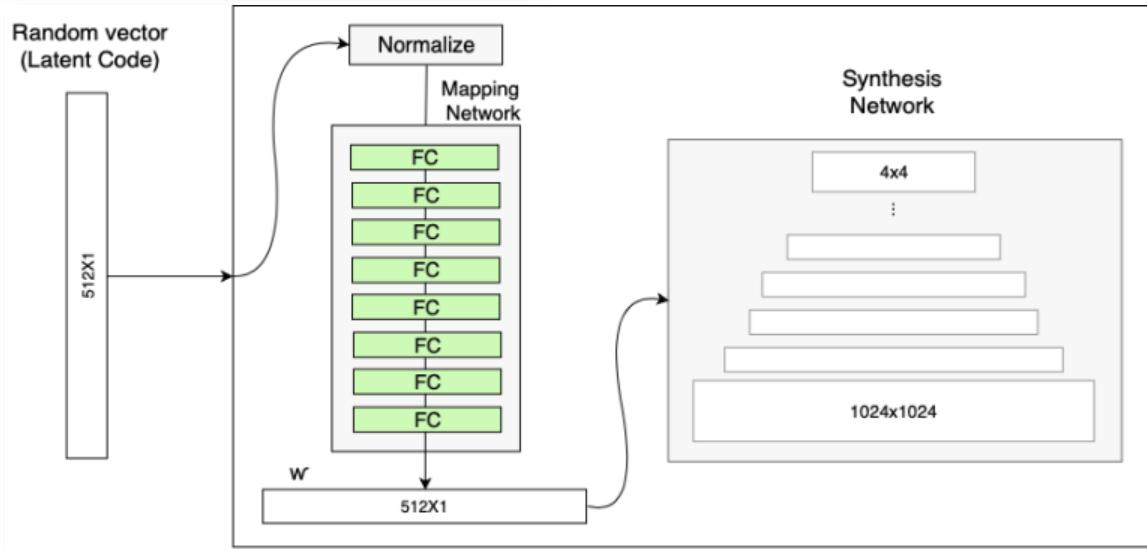


- Generator(생성자)와 Discriminator(판별자) 두 개의 모델이 동시에 학습
- 생성자는 실제 데이터가 가지고 있는 분포를 학습하고 판별자는 생성자가 생성한 데이터가 Real data 인지 Fake data인지를 구별하면서 적대적으로 학습
- 학습 과정에서 생성자의 목표는 판별자를 속일 수 있는 ‘실제와 같은 이미지’를 생성하는 것
- 학습 과정에서 판별자의 목표는 2개의 이미지가 주어졌을 때 실제 이미지가 맞는지 구분하는 것



모델링 : StyleGan

StyleGan이란?

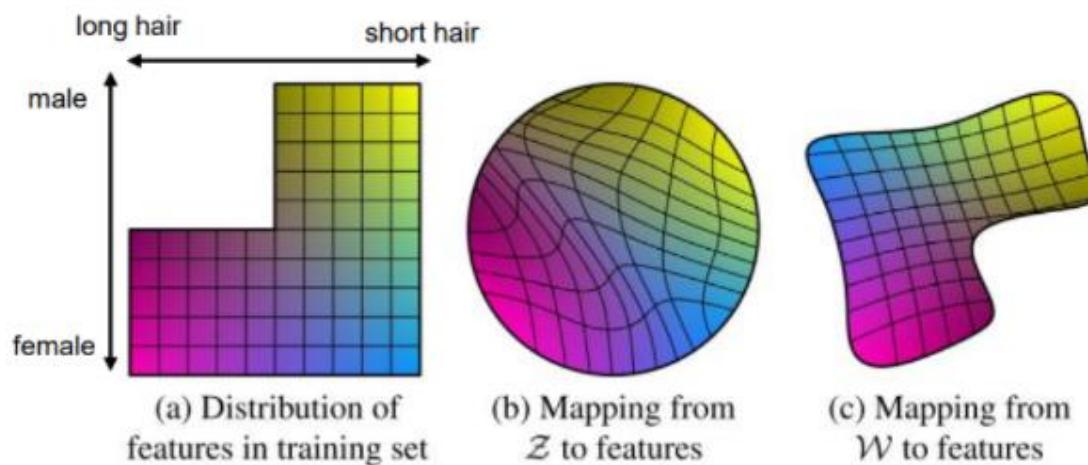


- 기존의 generator는 input latent vector Z가 직접 convolution, up-sampling 등을 거쳐 이미지로 변환
- StyleGan은 Z가 fully-connected layer로 구성된 nonlinear mapping network을 거쳐 intermediate latent vector W로 먼저 변환
- W는 constant tensor가 이미지로 변환되는 과정에서 스타일을 입하는 역할 수행



모델링 : StyleGan

StyleGan2-ada 특징



- Vanilla GAN은 고정된 분포에서 latent space를 구성하기 때문에 entangle
- 이는 latent space에서 선형적인 변환이 일어났을 때 결과물은 그렇지 않을 수 있음을 의미
- stylegan에서는 latent vector Z 를 generator의 input으로 바로 사용하지 않고 non-linear mapping network를 사용하여 W 를 input으로 사용하여, disentanglement하게 함
- 이는 주어진 얼굴과 비슷한 속성을 가진 얼굴을 생성해야 하는 우리 프로젝트에서 매우 필요한 특징

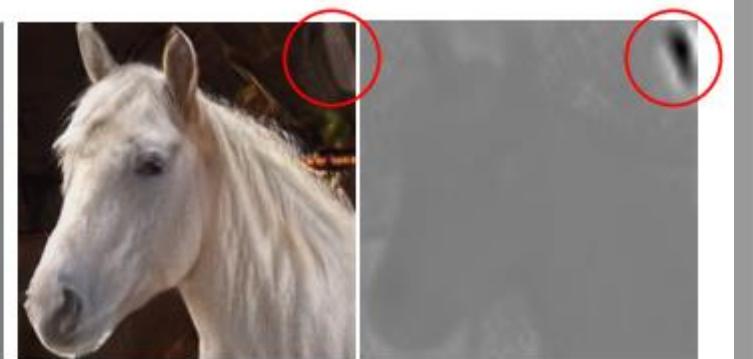


모델링 : StyleGan

StyleGan vs. StyleGan2 vs. StyleGan2-ada



	StyleGAN	StyleGAN2	StyleGAN2-ada
단점	정체불명의 물방울 모양도 같이 생성됨	많은 학습시간, 많은 데이터가 필요	-
개선점	-	물방울 모양을 생성하는 것으로 의심되는 AdaIN을 instace Normalization으로 대체	data augmentation을 적용하여 더 적은 학습시간, 적은 데이터



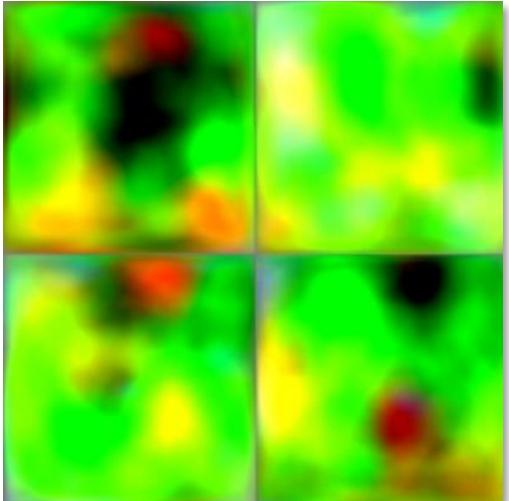


모델링 : StyleGan

학습 시간 별 시각화



T4 또는 P100 GPU로 [3, 256, 256] size의 이미지 65047장을 24시간 학습시킨 결과



2h 27m 32s

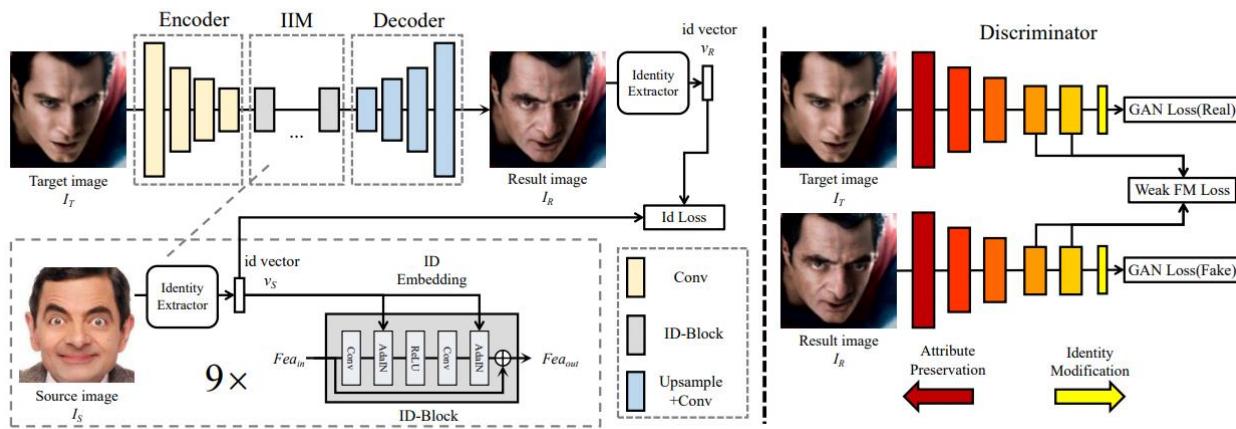
7h 19m 34s

23h 08m 34s



모델링 : SimSwap

SimSwap이란?



- 기존 딥페이크는 별도의 Identity extractor가 없음
→ identity를 target image에 주입하는 과정에서 하나의 특정 identity밖에 주입하지 못함
- SimSwap는 Identity extractor를 별도로 학습시켜 임의의 source image가 들어와도 id vector를 잘 추출

- 추출된 id vector는 encoder에서 추출된 정보와 함께 decoder입력으로 주입되며, decoder는 최종적으로 source image를 target image에 합성
- 합성된 image는 다시 한번 Identity extractor를 거쳐 id vector를 출력하게 되며, 이때의 id vector와 source image의 id vector의 cosine similarity가 낮도록 강제
→ 합성된 image와 source image가 닮도록 강제할 수 있음



모델링 : SimSwap

모델의 최적 Loss Function



$$\lambda_{Id}L_{Id} + \lambda_{Recon}L_{Recon} + L_{Adv} + \lambda_{GP}L_{GP} + \lambda_{wFM}L_{wFM_sum}$$

	수식	설명
Identity Loss	$L_{Id} = 1 - \frac{v_R \cdot v_S}{\ v_R\ _2 \ v_S\ _2}$	source image의 id vector와 합성된 image의 id loss의 cosine similarity로 정의되며, 두 이미지의 identity가 유지되도록 강제함
Reconstruction Loss	$L_{Recon} = \ I_R - I_T\ _1$	합성된 image와 target image 와의 L1 distance로 정의되며, 합성된 image가 target image를 닮도록 강제함
Adversarial Loss	-	합성된 image가 real image의 분포를 따를 수 있도록 강제함
Gradient Penalty	-	GAN의 불안정한 학습과정을 해소하도록 함
Weak Feature Matching Loss	$L_{wFM_sum} = \sum_{i=1}^2 L_{wFM}(D_i)$ $L_{wFM}(D) = \sum_{i=m}^M \frac{1}{N_i} \ D^{(i)}(I_R) - D^{(i)}(I_T)\ _1$	Discriminator의 전체 layer에서 target image와 합성된 image에서 추출된 feature가 유사하도록 강제함으로써 두 image가 닮도록 강제함



RESULTS



결과 예시





결과 예시



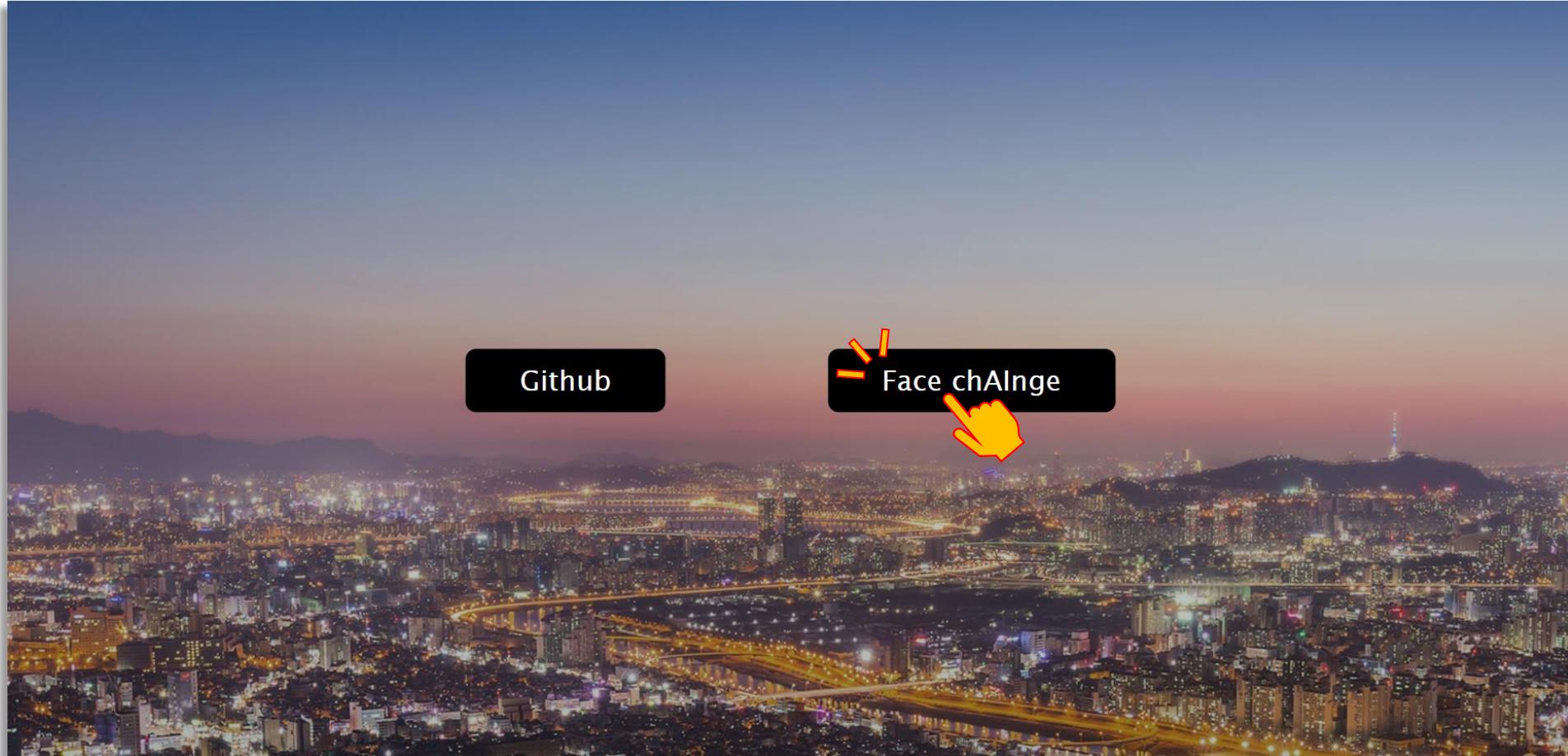


결과 예시





웹 데모



<https://github.com/Jy0923/Face-chAlnge>



CONCLUDING



기대 효과



01

제약 없는 사진 촬영

사람이 많은 곳에서 자유롭게 사진을 찍을 수 있음

ex. 길거리, 여행지, 공공장소, 파티, 행사 등



02

빠르고 자연스러운 보정

포토샵 등 따로 시간을 내어 보정할 필요 없음

동양인 데이터셋을 활용함으로써 보다 자연스러운 얼굴 생성 가능



03

초상권 침해 방지

다른 사람의 얼굴을 맘대로 찍는 경우 초상권 침해 문제

인스타그램, 페이스북 등 SNS에 올릴 경우 더욱 조심해야 함



활용 방안

정준영, 승리, 범죄자 등 사회적으로 물의를 일으킨 사람/ 전 여친·남친 얼굴 변형





활용 방안

정준영, 승리, 범죄자 등 사회적으로 물의를 일으킨 사람/ 전 여친·남친 얼굴 변형





활용 방안

정준영, 승리, 범죄자 등 사회적으로 물의를 일으킨 사람/ 전 여친·남친 얼굴 변형

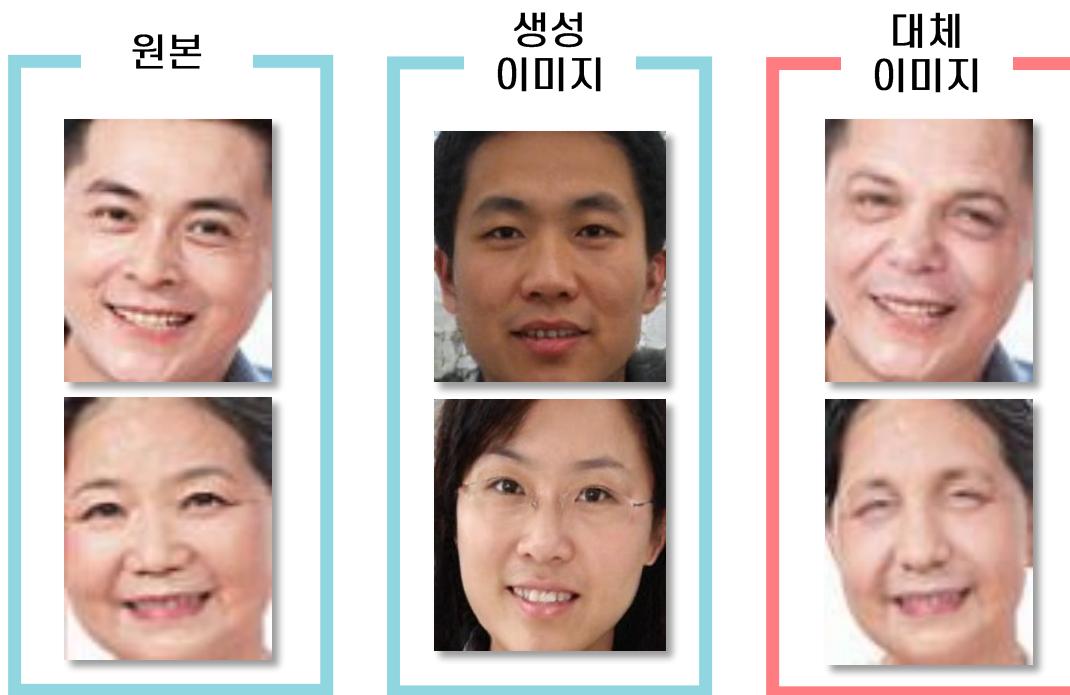




한계점 및 개선방안

- 총 2-80세 얼굴 데이터 중 10세 미만, 60세 이상이 적은 불균형 데이터
→ 더 많은 아시아인 데이터셋 수집하기

- 생성한 이미지의 해상도가 낮을 경우 바뀐 인물의 얼굴이 부자연스러움
→ 고해상도의 아시아인 데이터셋 수집하기





팀원 소개



김상호

SimSwap 구현
백엔드 구현



송창용

StyleGan2 구현
프론트엔드 구현



조혁준

StyleGan2 구현
PPT 제작



마민정

SimSwap 구현
PPT 제작 및 발표



오주영

RetinaFace 구현
PPT 제작



face chAlnge
THANK YOU FOR LISTENING