

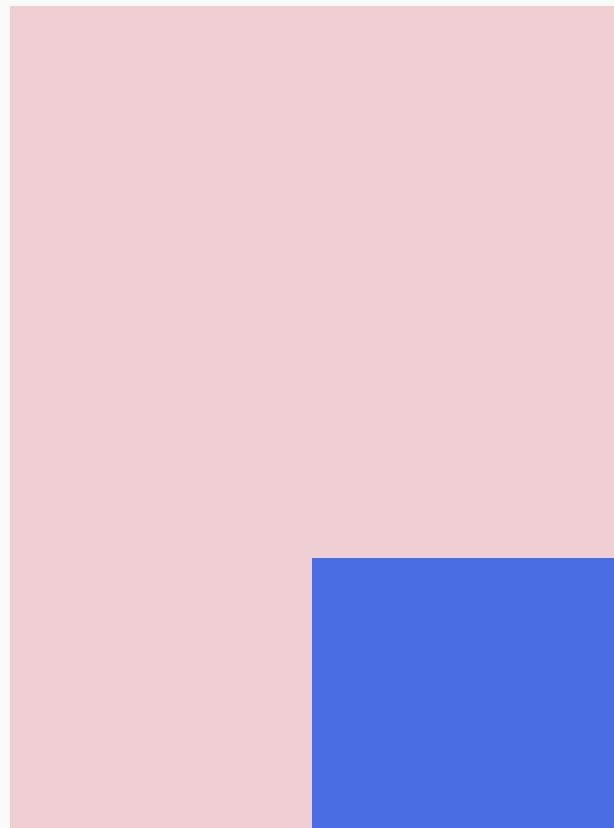


ESTsoft

3D를 활용한 감성분석

2조 - 김토일 심재은 이상윤 허영민

https://github.com/LeesY99/text_to_3D_person



Contents

1 프로젝트 소개

2 기획 의도

3 타겟 고객 (Persona)

4 DATASET

5 프로젝트 설계

6 모델 설명

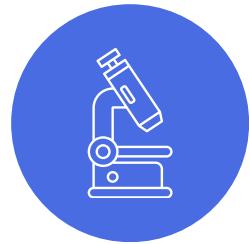
7 프로젝트 결과 (demo 영상)

8 도전과제 및 해결방안

9 향후 계획 및 발전 방향

10 참고 문헌

1 프로젝트 소개



3D를 활용한 감성분석

GAN 모델을 사용해 다양한 표정의 이미지를 생성

이를 3D 모델링으로 변환

동화책의 텍스트 감성 분석에 따라 맞춤형 감정 표현을 제공하는 시스템을 개발

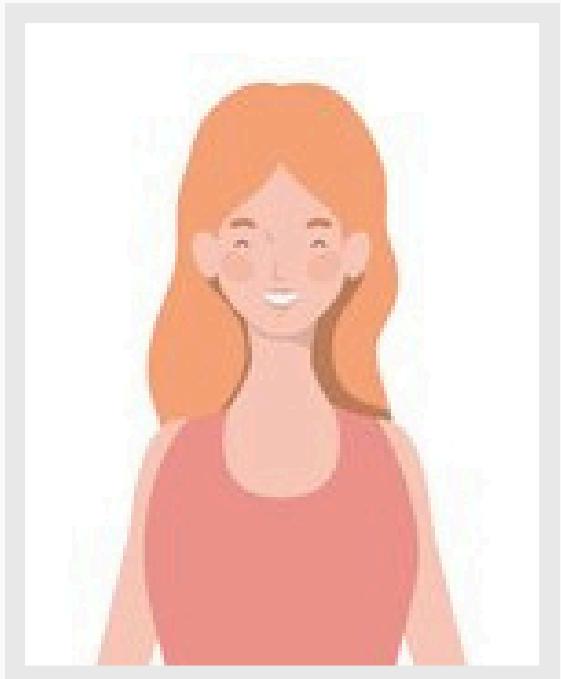
2 기획 의도



**아이들이 동화책의 감정을 더 깊이 이해하고
몰입할 수 있도록**

- 텍스트의 감정을 시각적으로 표현하는 것이 필요
- 감정 시각화를 통해 이야기의 **감정선**을 보다 생생하게 전달
- 이는 아이들의 공감 능력과 상상력을 크게 향상

3 타겟 고객 (Persona)



 이름 자스민
나이 35
직업 전업주부,
워킹맘
Family 기혼

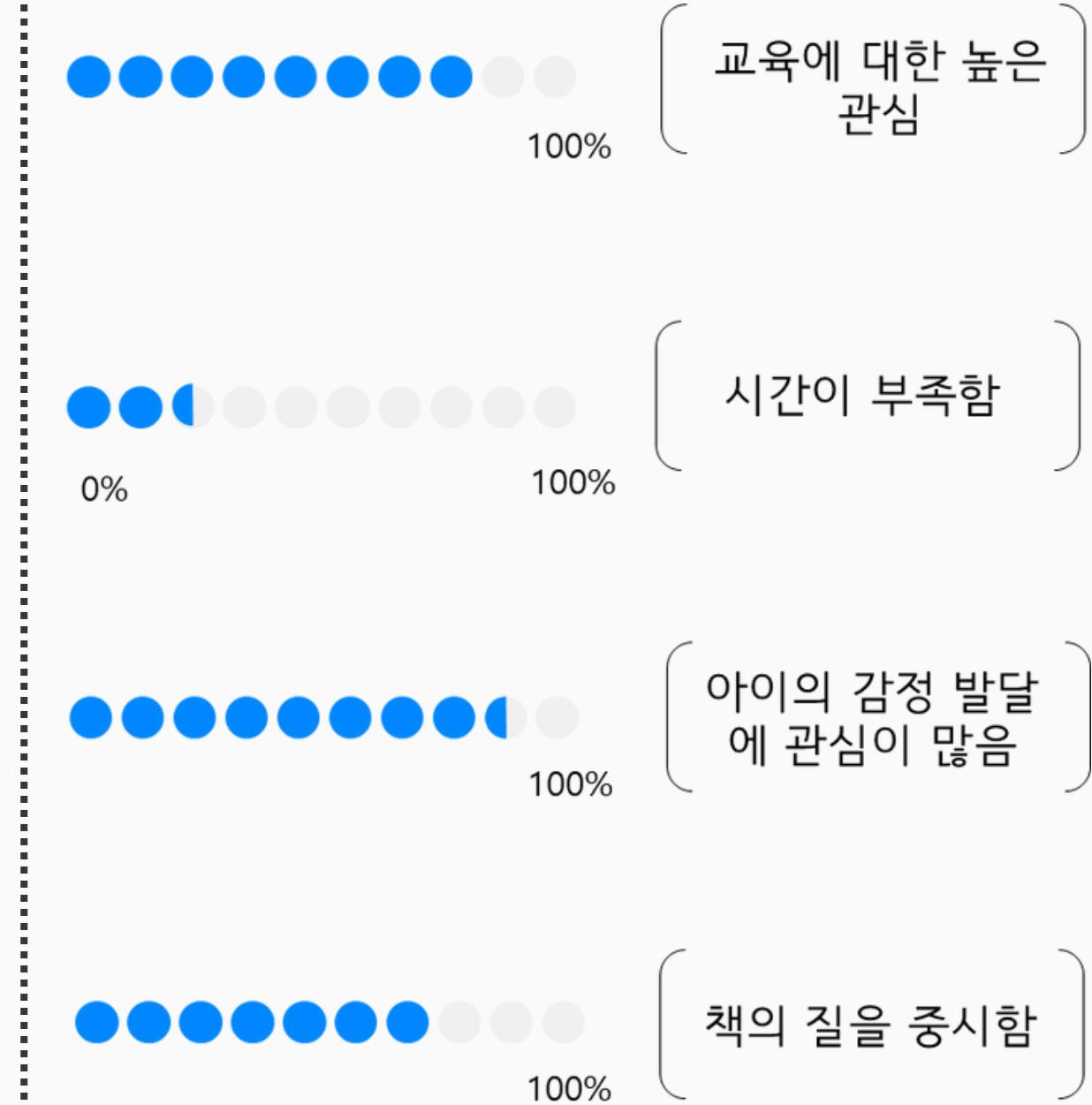
Key Characteristics

- **가족 구성**
3세 아들을 둔 학부모
- **직업 및 생활**
워킹맘으로 직장 생활과 가사를 병행
바쁜 일상 속에서도 아이의 교육과 정서 발달에 많은 관심을 가짐

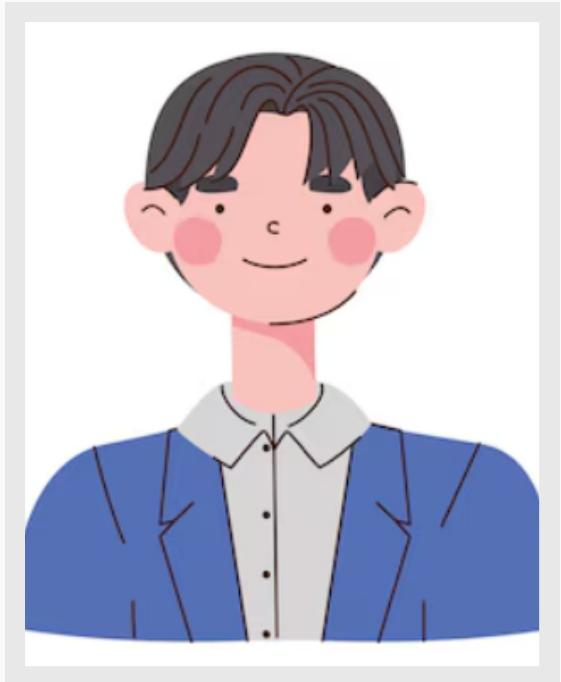
Goals

- **아이의 감정 이해 증진**
감정을 시각적으로 표현한 동화책을 통해 아이가 다양한 감정을 더 잘 이해하고 표현할 수 있도록 도움
- **교육적 가치 제공**
감정 교육과 정서 발달을 중시하는 콘텐츠를 통해 학부모들이 신뢰할 수 있는 교육 자료를 제공
- **독서 습관 형성**
아이가 책 읽는 즐거움을 느끼고 지속적으로 독서 습관을 기를 수 있도록 매력적인 동화책을 제공

Behaviors



3 타겟 고객 (Persona)



이름 류선재

나이 37

직업 동화책 출판
편집자

Background 10년
경력의 출판 전문가

Key Characteristics

- 가족 구성

10년 경력의 출판 전문가

- 직업 및 생활

동화책 출판 편집자로서 기획, 편집, 출판 전 과정 관리

- 창의적 협업

작가 및 일러스트레이터와의 창의적인 협업을 통해 독창적인
동화책 제작

Behaviors



시장 트렌드 분석



콘텐츠 품질 중시



기술 활용 및
혁신 추구



마케팅 및 프로
모션 전략

Goals

- 우수한 동화책 출판

감정을 시각적으로 표현한 동화책을 통해 독자들에게 깊은
인상을 남기고 교육적 가치를 제공

- 독자 만족도 향상

부모와 아이들이 만족할 수 있는 고품질의 동화책을 출판하여
독자 충성도를 높임

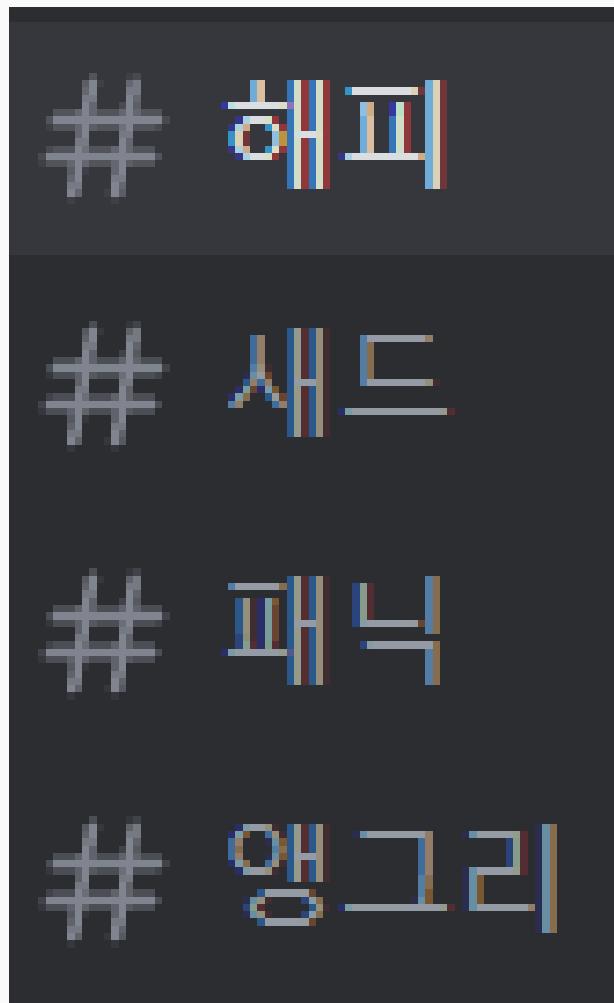
- 디지털 출판 강화

전자책, 오디오북 등 디지털 출판 채널을 강화하여 새로운 시장 개척

4 DATASET

EDA / 전처리

1) 토의를 통해 맞지 않다고
판단된 이미지 제거
(classification)



2) 중복 이미지 제거

4) 얼굴부분 crop

3) annot_A 기준으로 적절하지
않은 좌표를 가진 이미지 제거
(classification)

```
"boxes": {  
    "maxX": 1500.189177,  
    "maxY": 669.5033842,  
    "minX": 872.5148770000001,  
    "minY": -204.08019579999998  
},
```



4 DATASET

데이터 증강 (Data Augmentation)

anger 1495 -> 4712
happy 1493 -> 4979
panic 1479 -> 4928
sadness 1481 -> 5123



1
face detection
face alignment

2
pose
augmentation

Add 20°
in Pitch



Minus 20°
in Yaw



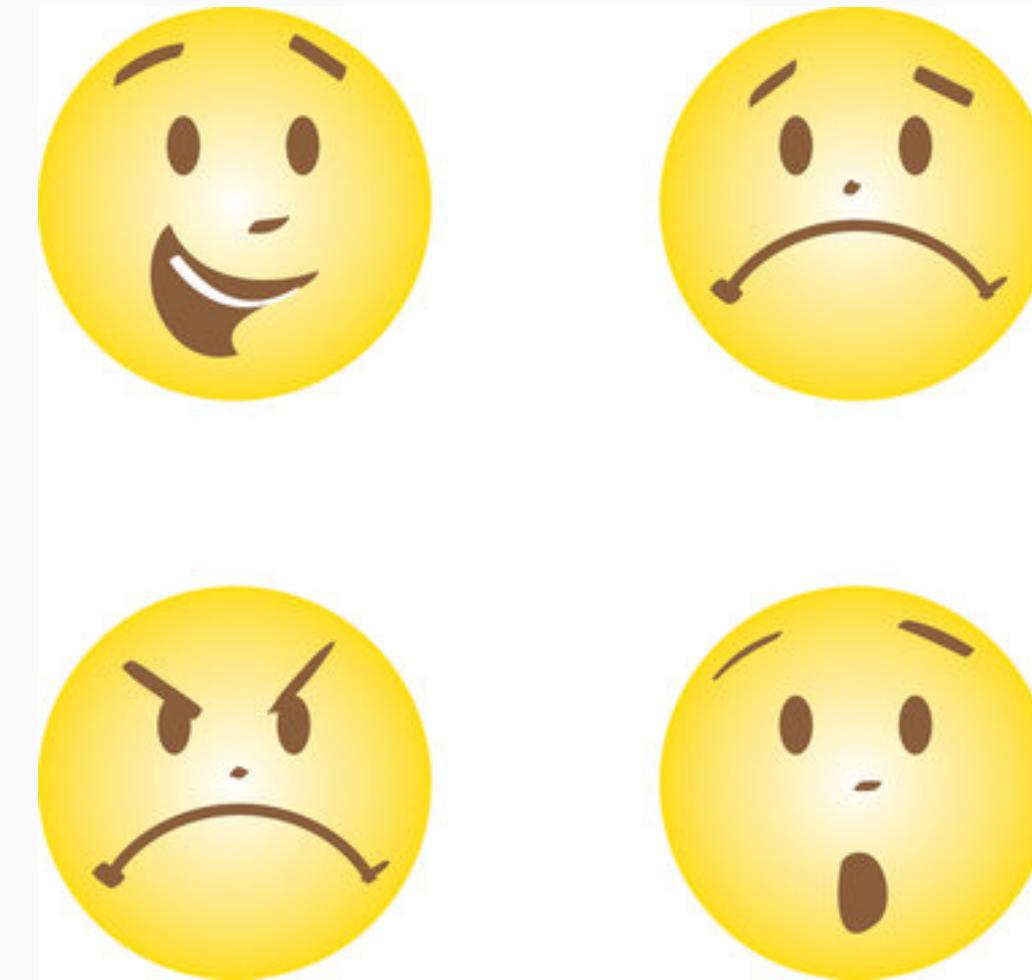
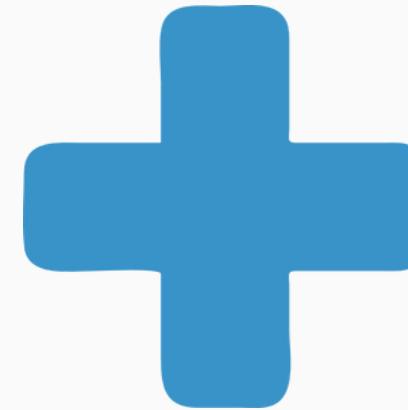
Minus 20°
in Pitch



Add 20°
in Roll

4 DATASET

starGAN 데이터 셋



CelebA(Celebrity Attributes) 데이터 셋

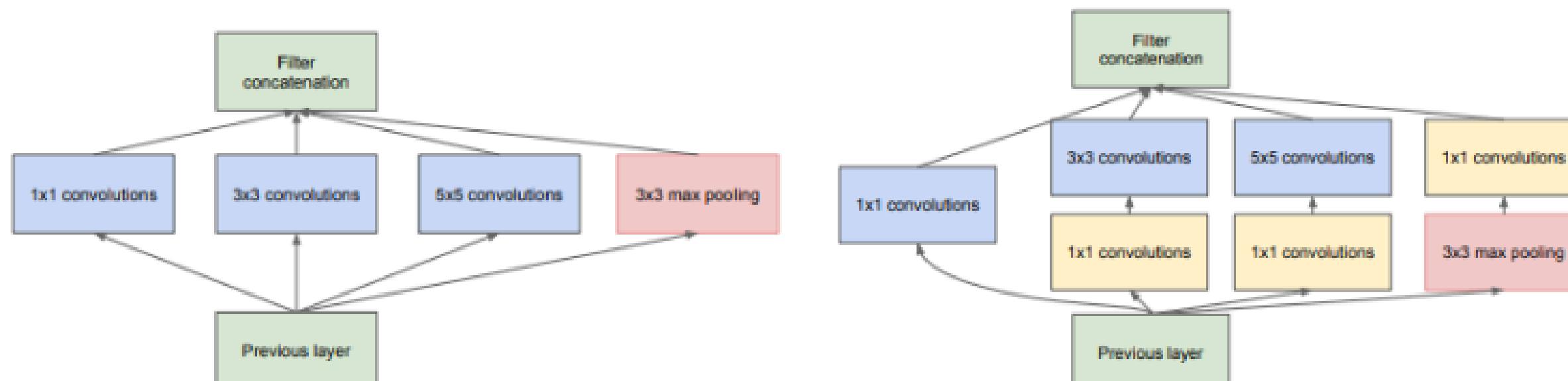
4개의 감정 이미지 데이터 셋

5 프로젝트 설계



6 모델 설명 - 분류 모델

- **googleNet**



(a) Inception module, naïve version

(b) Inception module with dimension reductions

Figure 2: Inception module

- 1x1, 3x3, 5x5 세 개의 Conv layer와 1개의 Max-pooling

- 1x1 Conv layer 사용

* 1x1 Conv은 입력의 채널 수를 줄이는 역할
=> 차원을 축소하여 연산량을 줄여줌

6 모델 설명 - 분류 모델

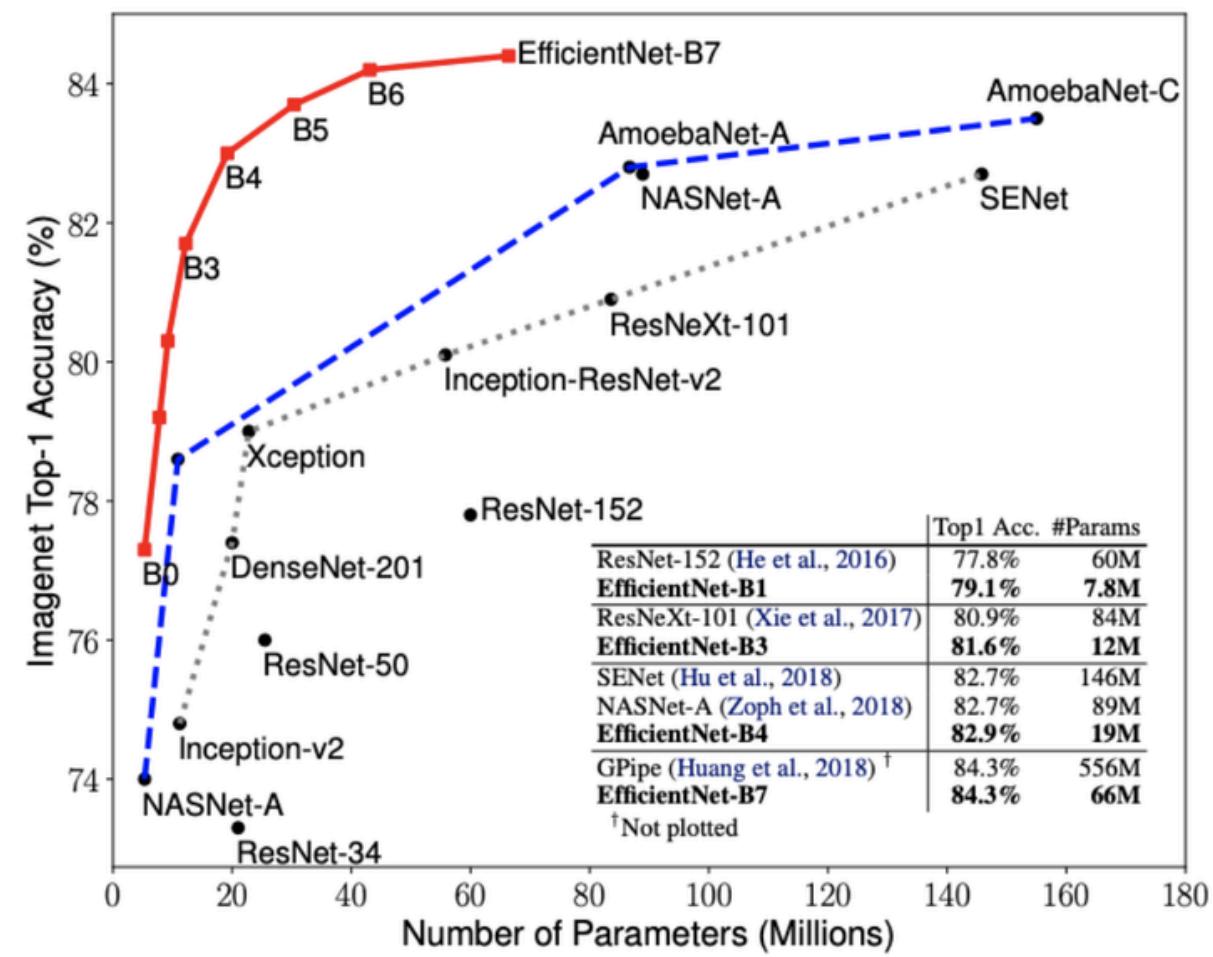
• 구조

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

Table 1: GoogLeNet incarnation of the Inception architecture

6 모델 설명 - 분류 모델

● EfficientNet



효율적인 모델 크기&성능

연산량과 파라미터 수를 줄임에도 높은 성능 유지
=> 감정 이미지 분류 작업에서 빠르고 정확한 예측 가능

전이 학습 지원

대규모 데이터 셋으로 사전 학습 된 모델(pre-trained) 제공

복합적 스케일링 방법

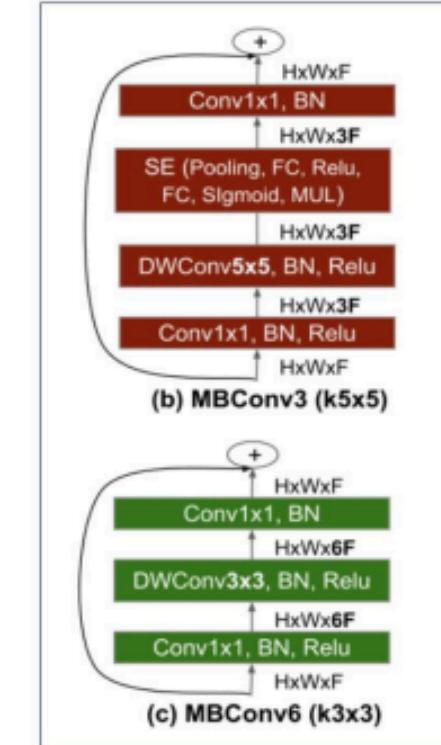
너비, 깊이, 해상도를 균형 있게 확장하는 복합적 스케일링 방법 사용
=> 다양한 해상도에서 이미지 처리 가능

6 모델 설명 - 분류 모델

- 구조



MBConv Block 구조



※ MBConv 블록:
EfficientNet의 주요 구성 요소
로, 경량화된 딥러닝 모델을 구
축하기 위해 제안된 블록

Stage i	Operator $\hat{\mathcal{F}}_i$	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels \hat{C}_i	#Layers \hat{L}_i
1	Conv3x3	224×224	32	1
2	MBConv1, k3x3	112×112	16	1
3	MBConv6, k3x3	112×112	24	2
4	MBConv6, k5x5	56×56	40	2
5	MBConv6, k3x3	28×28	80	3
6	MBConv6, k5x5	14×14	112	3
7	MBConv6, k5x5	14×14	192	4
8	MBConv6, k3x3	7×7	320	1
9	Conv1x1 & Pooling & FC	7×7	1280	1

B1~B7은 B0에서 Depth, Width를
증가시켜서 모델 생성.
Depth는 기존 Block의 개수를 증가
시킴
Width는 기존 filter수를 증가 시킴

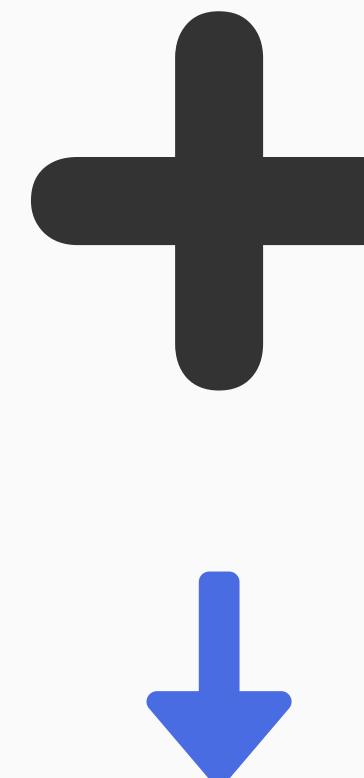
6 모델 설명 – 이미지 분류 모델

GoogleNet

- 효율적인 멀티 스케일 특징 추출
- 깊이와 계산 효율성의 좋은 균형

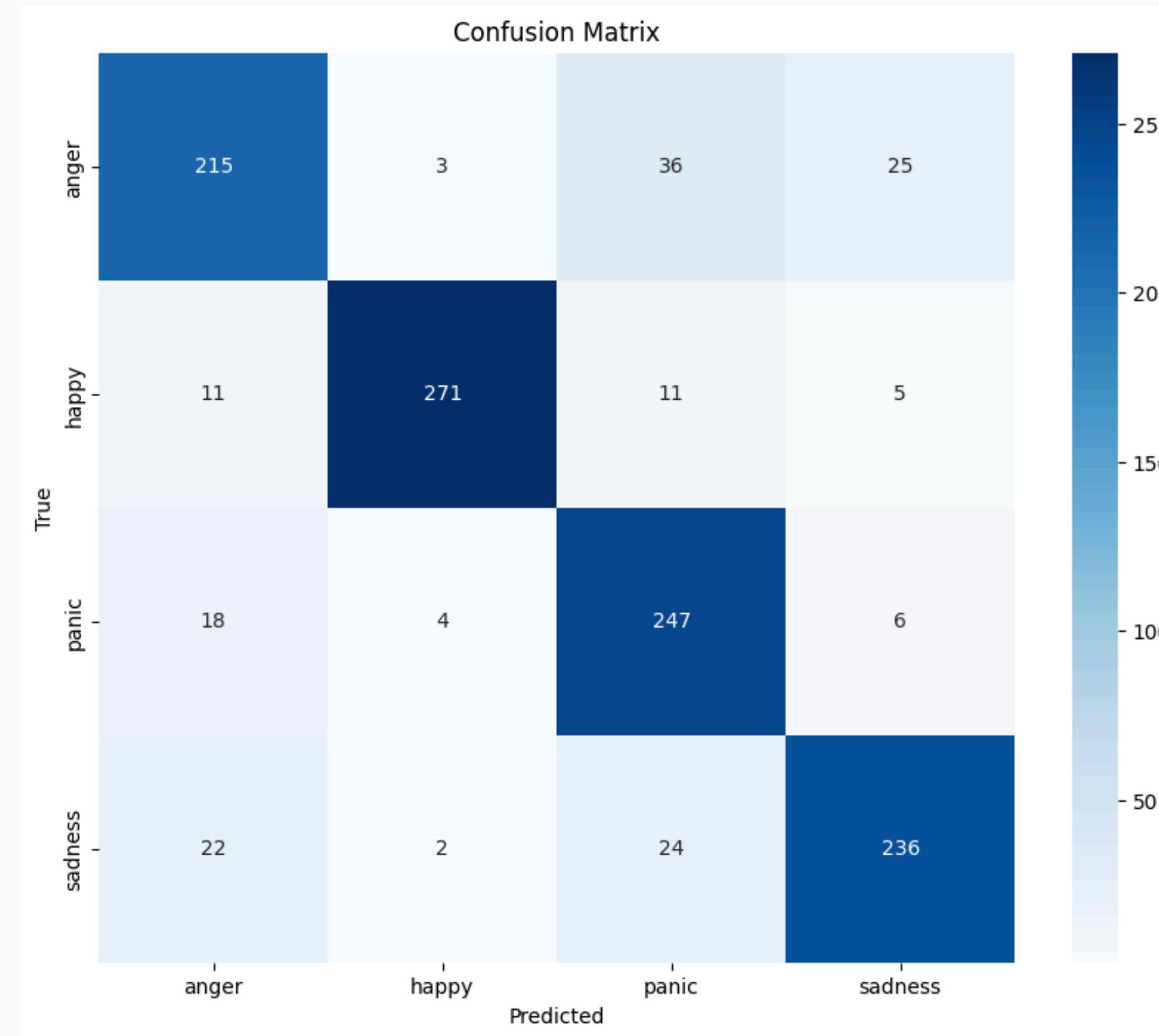
EfficientNet

- 최적화된 파라미터 사용과 확장성
- 다양한 이미지 분류 벤치마크에서 최첨단 성능



약 86% 성능

6 모델 설명 - 이미지 분류 모델



6 모델 설명 – 생성 모델 (StarGAN)

Image-to-image translation : 특정한 특징을 갖는 이미지를 또 다른 특징을 갖는 이미지로 변환
ex) 흑발을 금발로, 웃는 얼굴에서 화난 얼굴로 등등

attribute : 이미지에 있는 의미있는 특징. (머리색, 성별, 나이 등)

attribute value : attribute의 값. (머리색-흑발/금발, 성별-남/여)

domain : 같은 attribute value를 공유하는 이미지들의 집합. (여성의 이미지들은 하나의 domain을 구성)

기존 모델의 한계

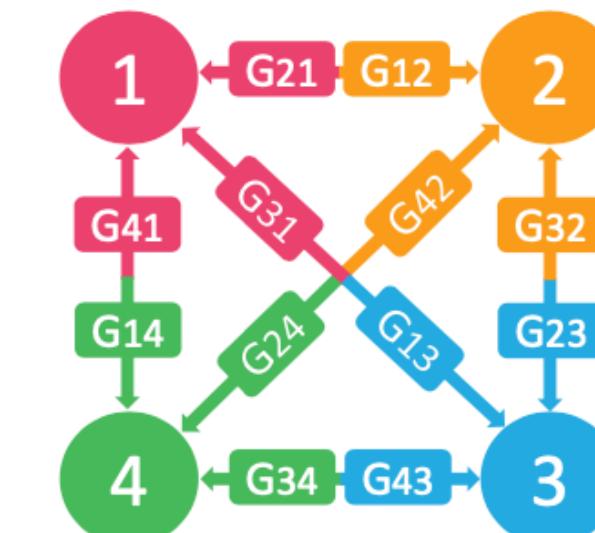
- 1) 여러 가지의 특징 변환을 위해서 여러 네트워크를 학습 해야 했음
- 2) 각 데이터 셋들은 라벨이 같지 않기 때문에 모두 결합해 학습할 수 없음



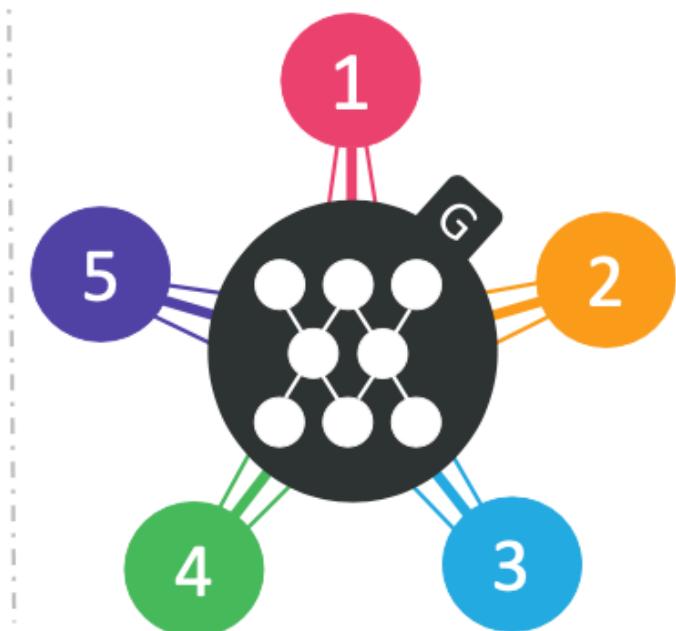
StarGAN

- 1) 하나의 모델로 여러 가지 도메인에 image-to-image translation 적용 가능
- 2) 여러 데이터 셋에 대해 동시에 학습할 수 있음 (label 달라도 가능)
mask vector를 활용해 각기 다른 데이터셋 간 겹치지 않는 라벨은 무시할 수 있게끔 학습

(a) Cross-domain models



(b) StarGAN



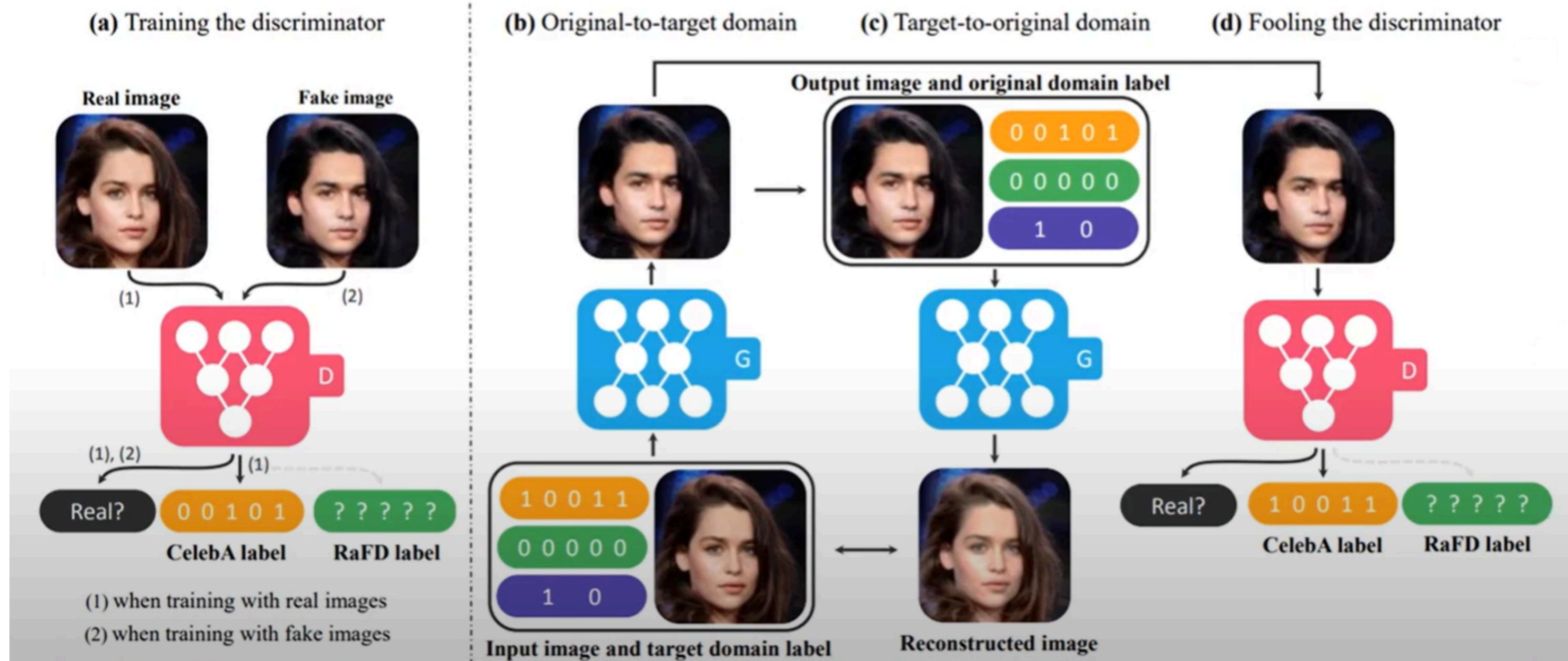
6 모델 설명 – 생성 모델 (StarGAN)

- StarGAN - test셋 결과



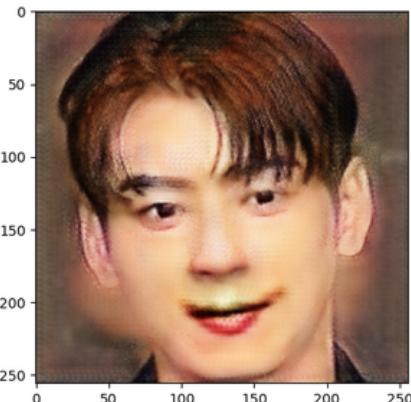
6 모델 설명 – 생성 모델 (StarGAN)

```
[Black_Hair, Blond_Hair, Brown_Hair, Male, Young, Anger, Happy, Panic, Sadness, CelebA, Custom]  
[[ 1 , 0 , 0 , 1 , 1 , 0 , 0 , 0 , 0 , 1 , 0 ]]
```



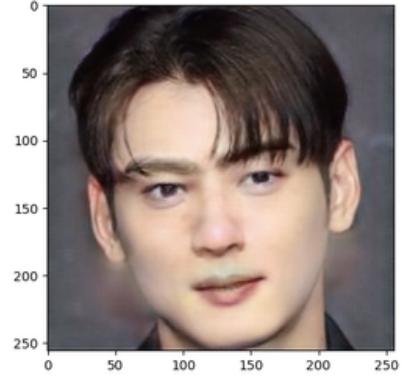
6 모델 설명 – 생성 모델 (StarGAN)

제공받은
감정 데이
터만 사용
했을 때



angry

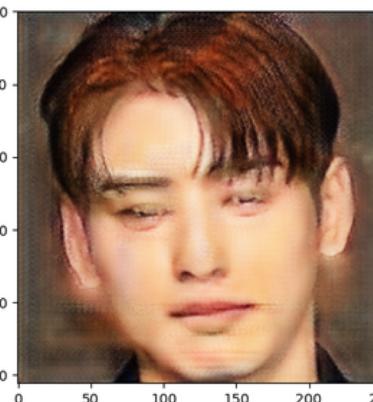
CelabA
데이터와
결합했을
때



happy

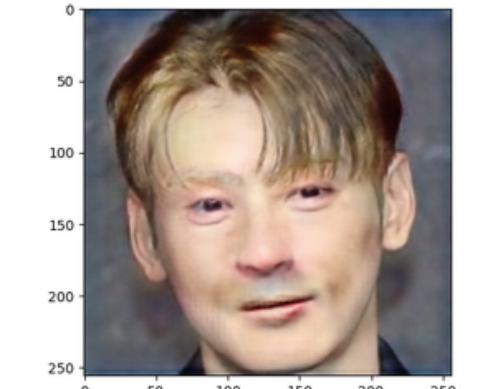


panic



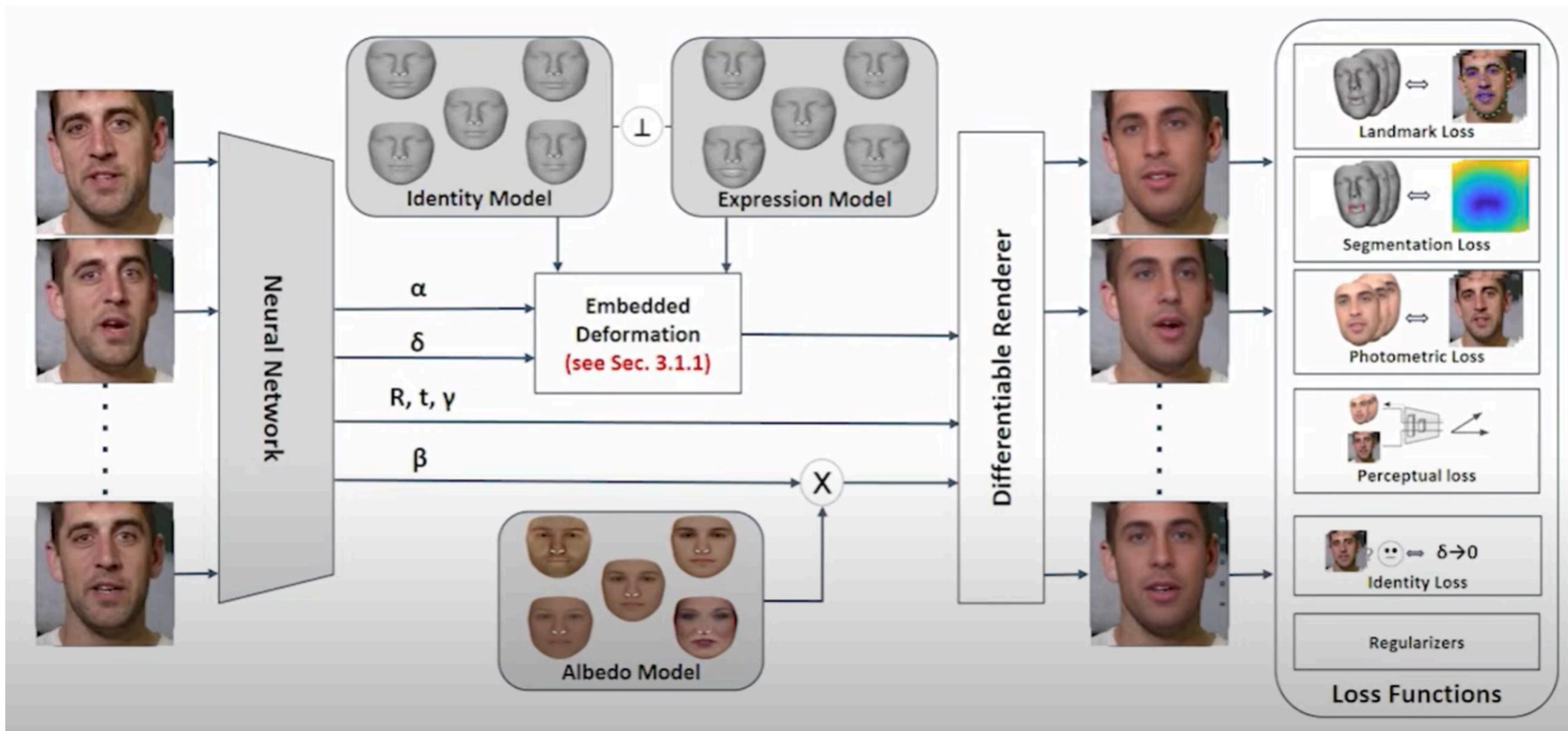
sadness

틀린 mask vector를
주었을 경우



```
# [ Black_Hair, Blond_Hair, Brown_Hair, Male, Young, Anger, Happy, Panic, Sadness, CelabA, Custom]
c_trg = [[ 0 , 1 , 0 , 1 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 1 ]]
```

6 모델 설명 – 3D 모델 (3DMM)



3DMM

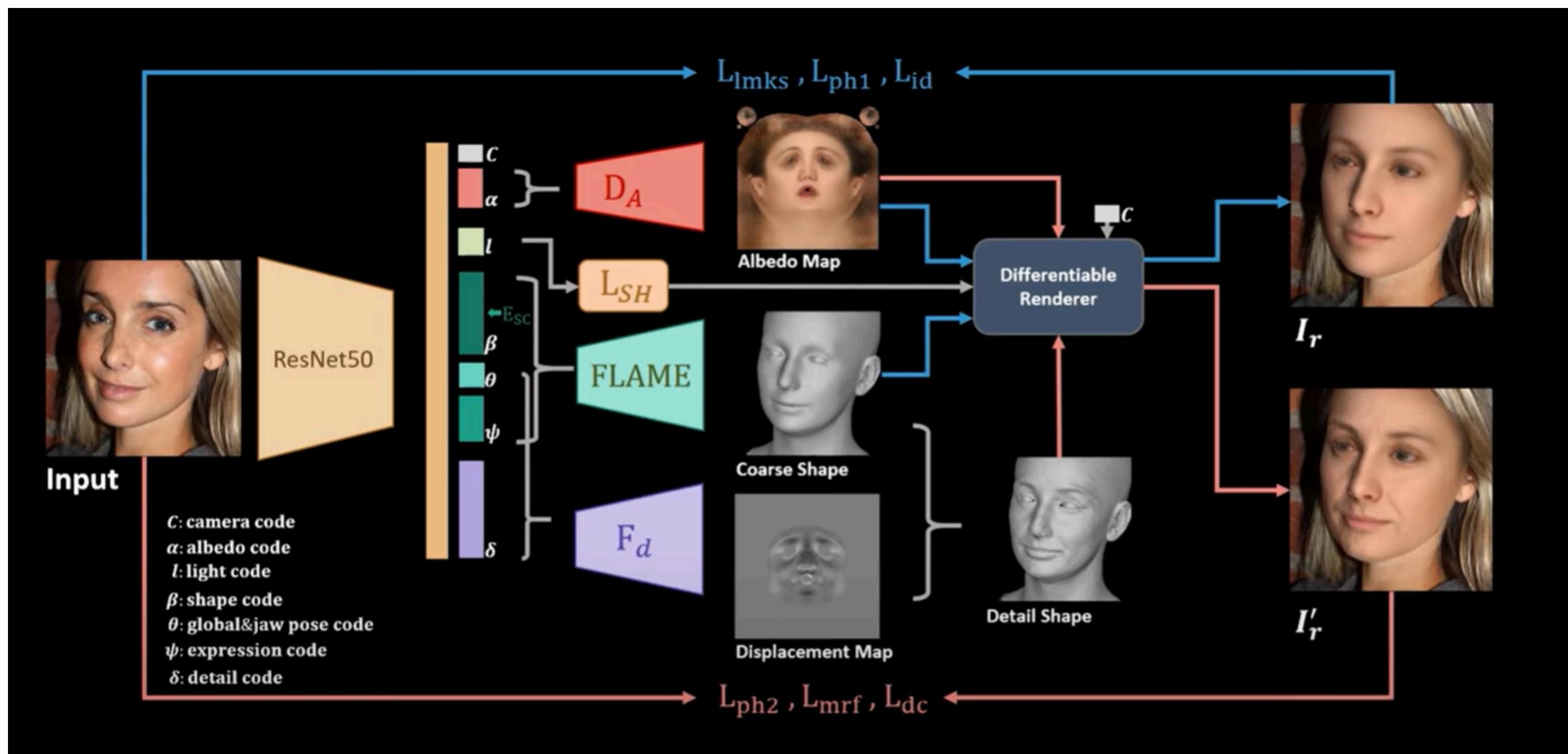


MESH



인코더 디코더 구조 (랜드마크, 세그멘테이션, 픽셀강도, 알베도) 손실을 줄이며 원본 이미지와 비슷하게 3D face를 만드는 모델

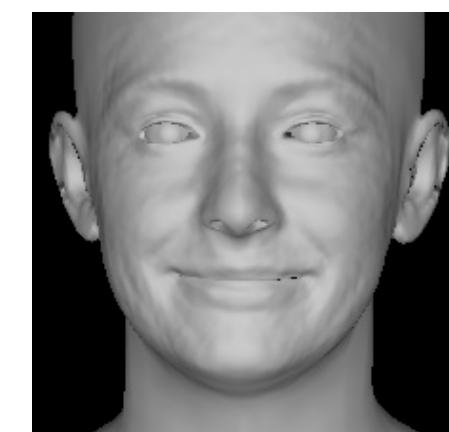
6 모델 설명 – 3D 모델 (DECA)



DECA

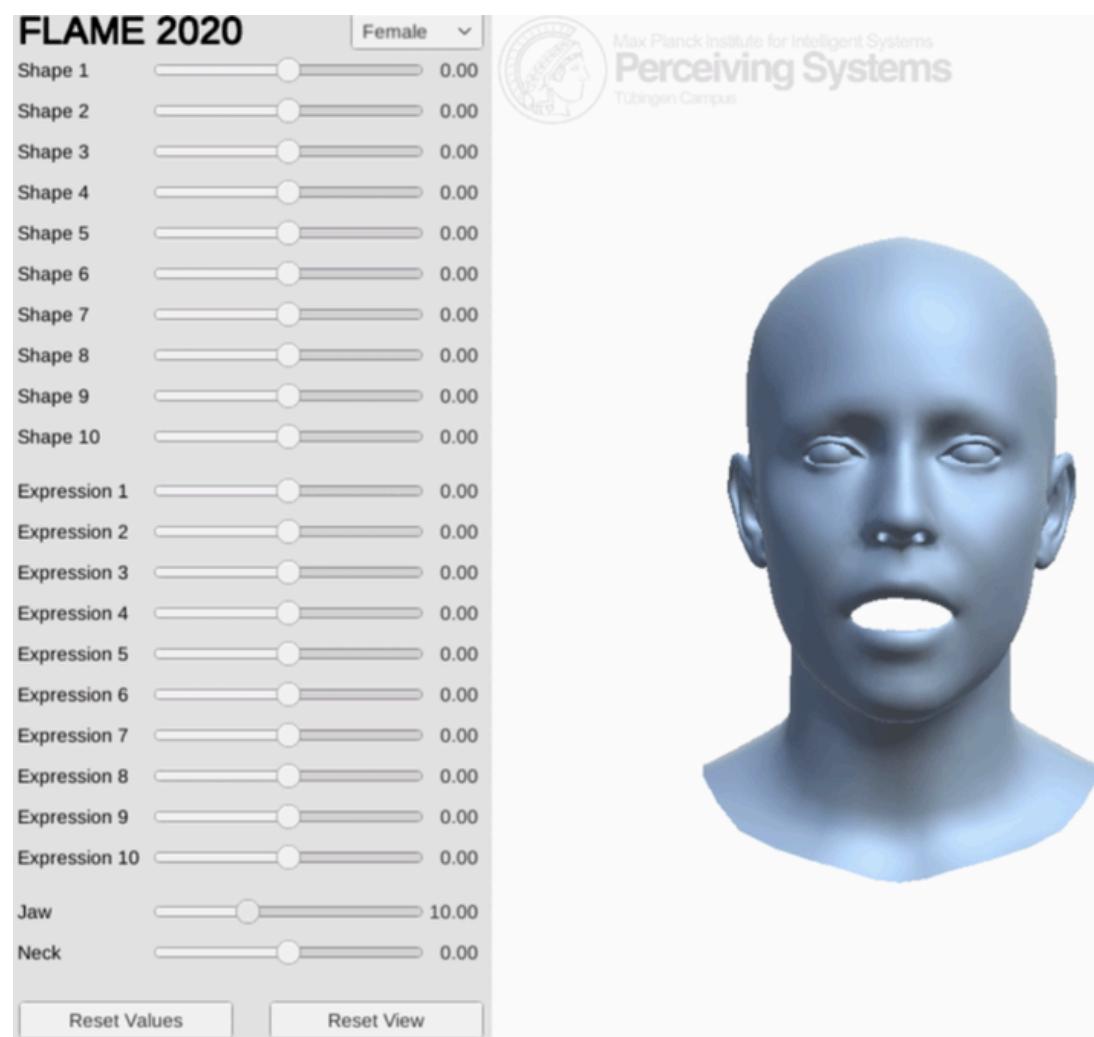


MESH

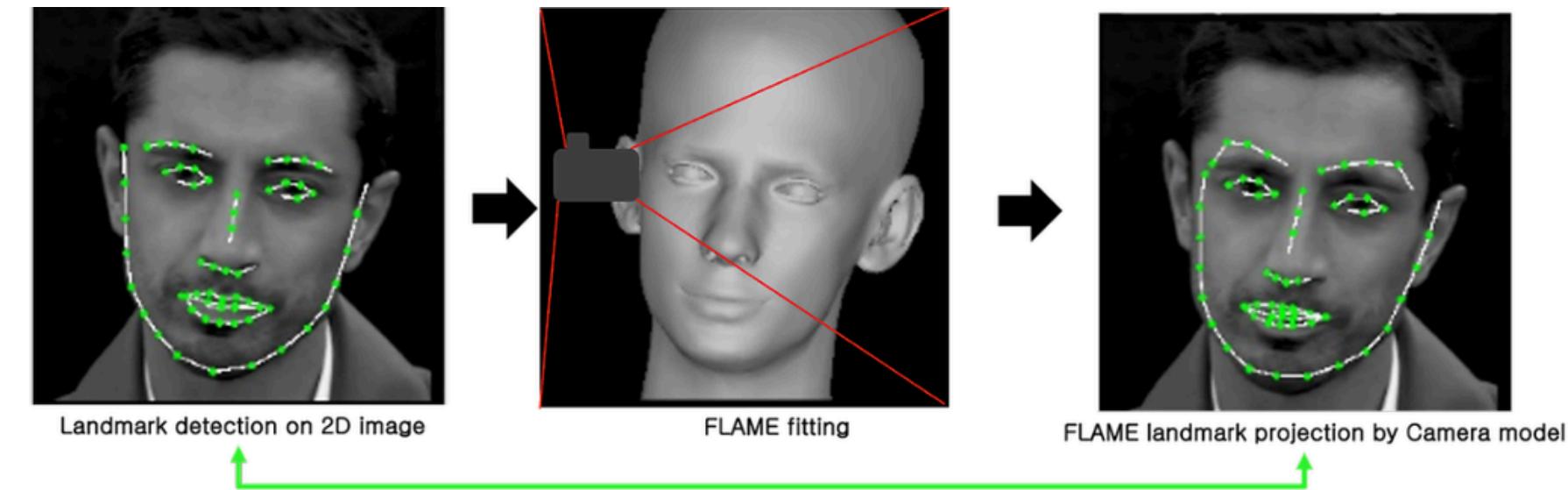


6 모델 설명 – 3D 모델 (DECA – course stage)

FLAME



1) Landmark re-projection loss



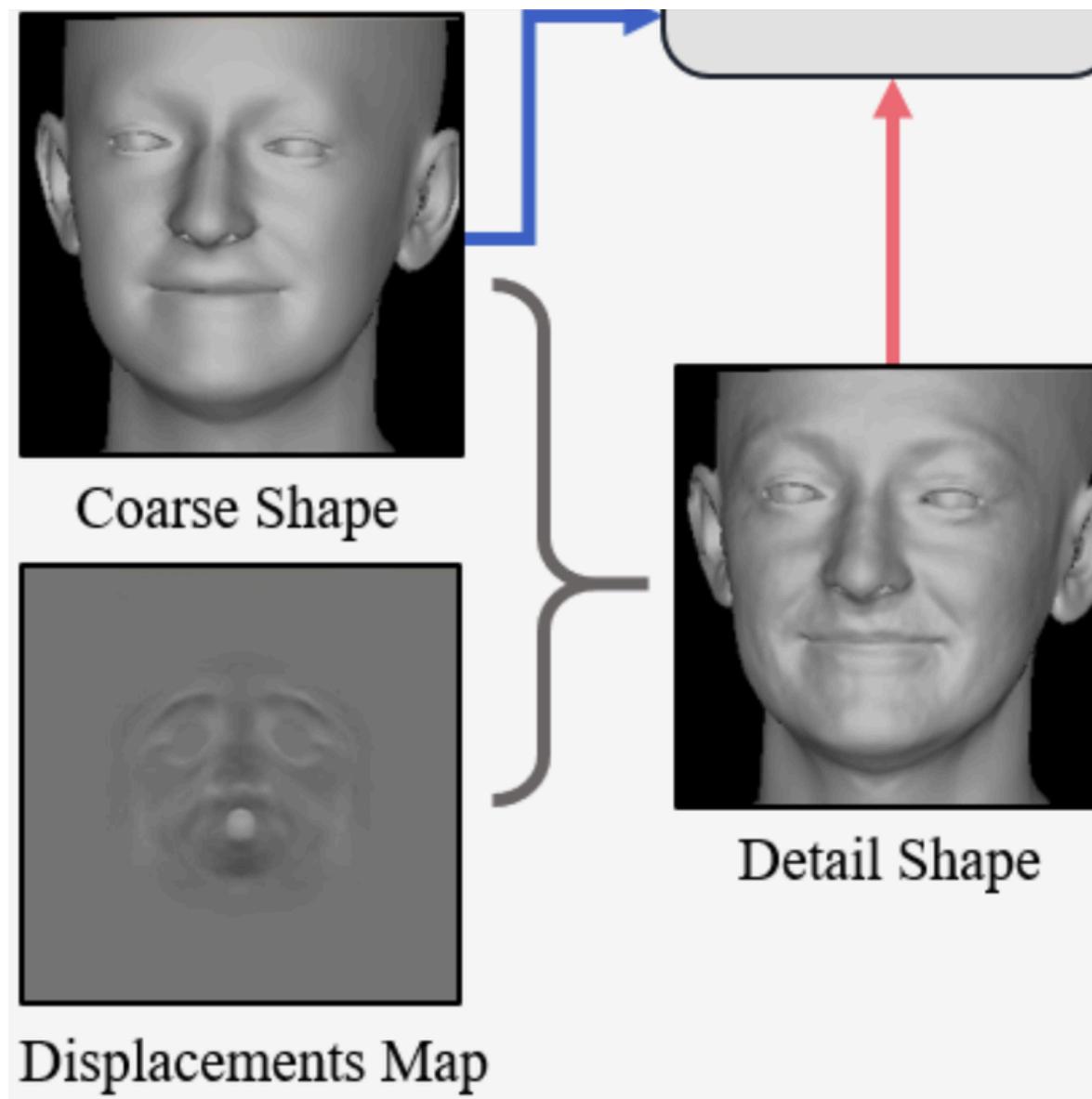
2) Photometric loss

3) Identity loss

shape(100), expression(50), pose(6) 파라미터와, albedo(50), camera(3), lighting(27)

6 모델 설명 – 3D 모델 (DECA – detail stage)

UV displacement map



- 1) **Detail photometric loss**
course stage의 photometric loss와 동일
단, detailed shape 적용
- 2) **Soft symmetry loss**
큰 포즈나 머리카락 장신구 등에 의한
가려짐에도 방해받지 않고 Robust한
displacement map을 출력하기 위함

6 모델 설명 - 텍스트 분류 모델

Estsoft 시계열 3조
(김토일,이재익,정치훈)

KoLECTRA-base-v3 모델

시계열 발표 모델 그대로 사용

판별기 인코더 층 : 12개

self-attention : 12개

생성기 인코더 층 : 12개

self-attention : 4개

vocab size : 35000

max_len = 75

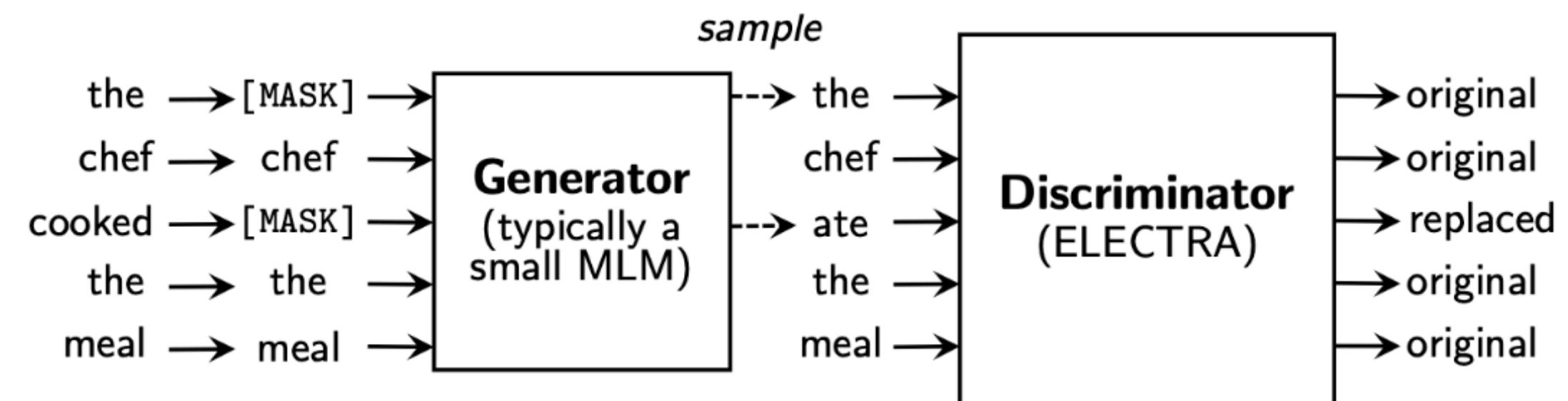
batch_size = 16

learning_rate = 5e-6

activation function : gelu, softmax

optimizer : adamw

KoELECTRA



INPUT

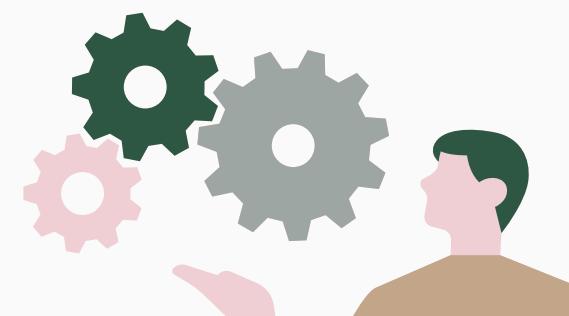
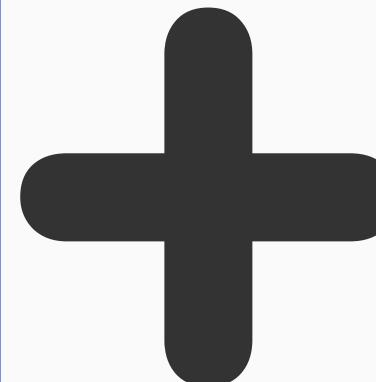
정면사진 한 장



동화책 텍스트

ex)미운 아기 오리

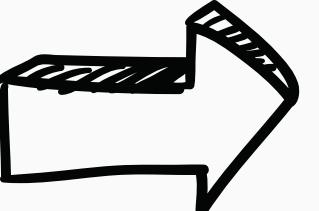
따스한 햇살이 내려앉은 개울가 덤불 숲
에서, 어미오리가 알을 품고 있었어.
툭! 톡! 알에 금이 가더니, 빠직 빠지직! 아
기오리들이 하나둘 고개를 내밀었어.



IMAGE

STARGAN(celabA domain) Hair color, gender, age 도메인 선택

정면사진



female young



male young



female old



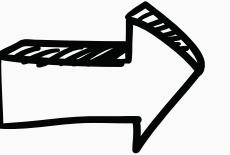
male old



IMAGE

STARGAN
(custom domain)

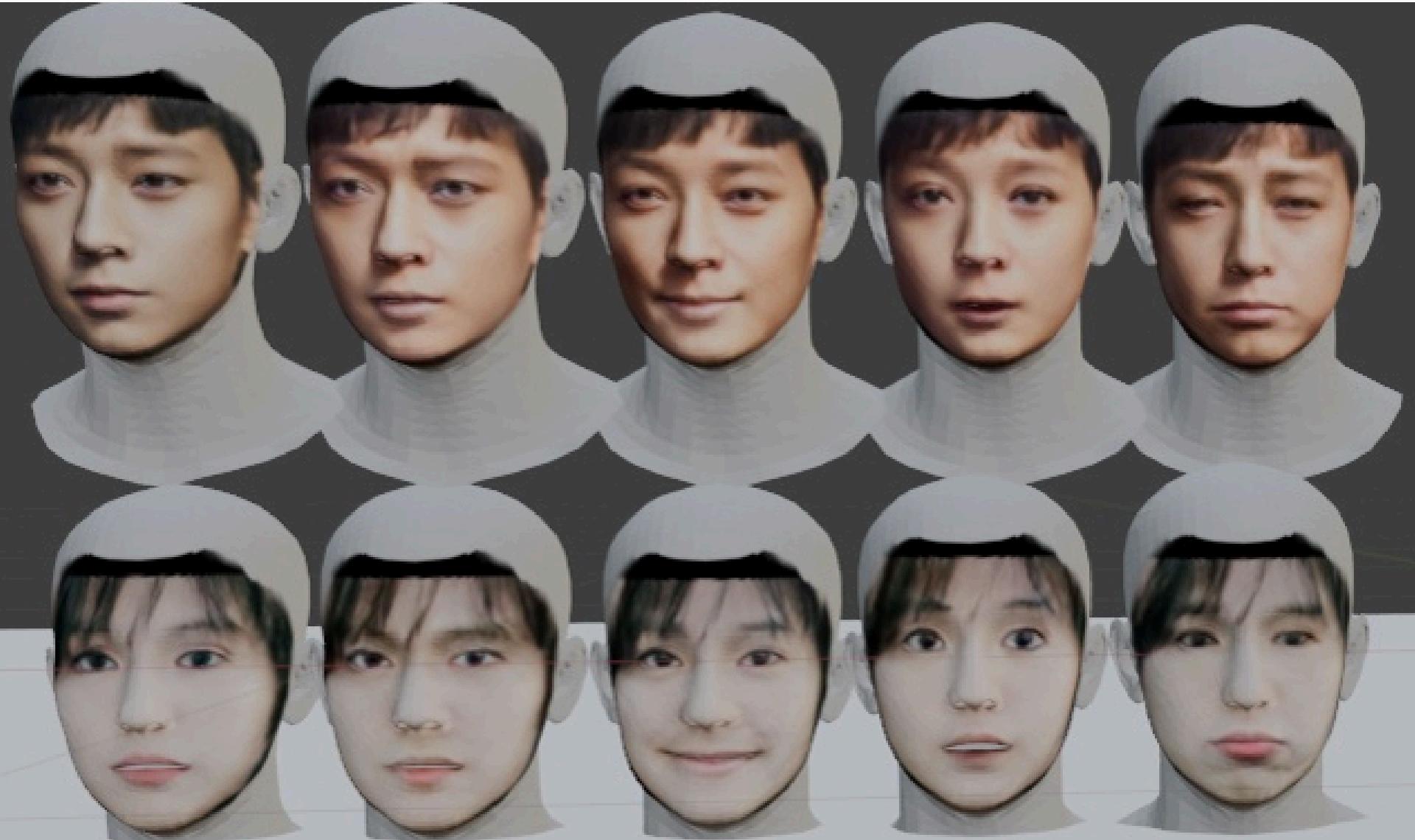
선택한 사진



Deca



3D Model Avatar

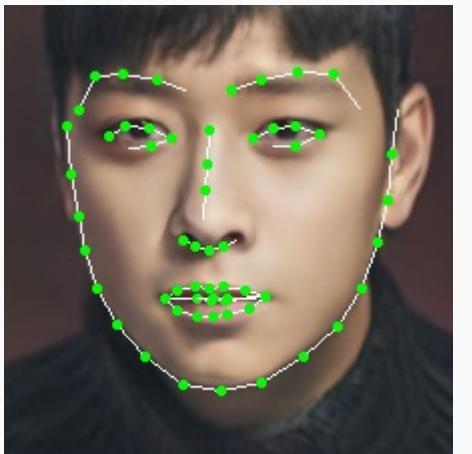


3D Model Avatar

image



68 landmarks

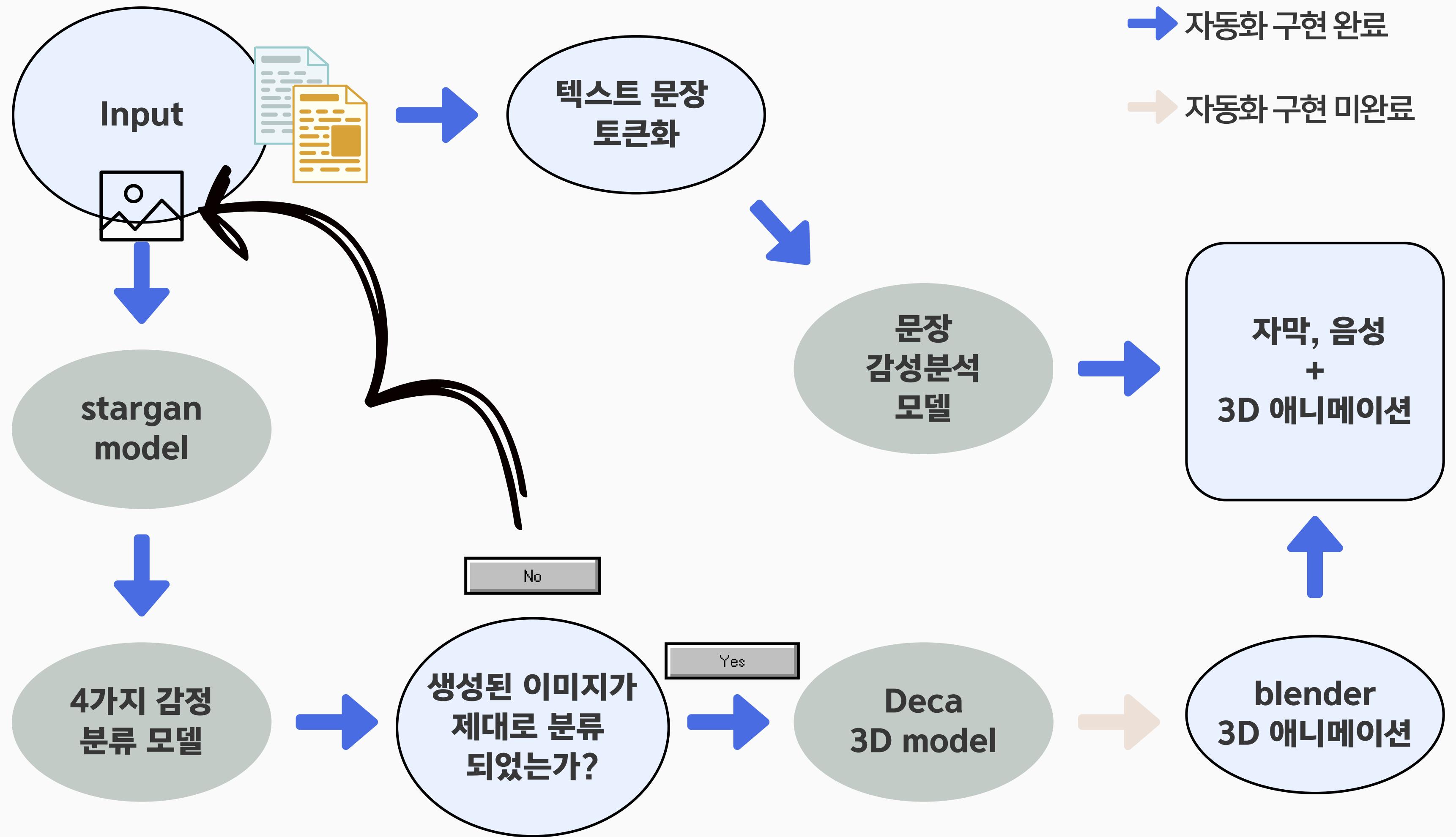


3D mesh



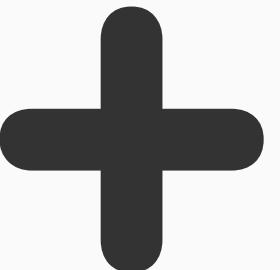
displacement map Albedo map





```
# 동화책 텍스트 입력  
input_text = """  
이스트소프트 최종 발표날이 왔어요.  
훌륭한 심사위원님들 앞이라 너무 긴장되네요.  
발표를 하다가 치명적인 실수를 해서 슬펐어요.  
실수한 나 자신에게 너무 화가 났어요.  
비록 실수가 있었지만 성장할 수 있는 좋은 기회였습니다.  
부트캠프를 하며 많이 배우고 성장한 것 같아서 행복합니다.  
"""  
# 문장 토큰화  
sentences = tokenize_sentences(input_text)
```

동화책 텍스트 입력



원본 사진



문장: 이스트 소프트 최종 발표 날이 왔어요.
예측된 감정: 중립
문장: 훌륭한 심사위원님들 앞이라 너무 긴장되네요.
예측된 감정: 불안
문장: 발표를 하다가 치명적인 실수를 해서 슬펐어요.
예측된 감정: 슬픔
문장: 실수한 나 자신에게 너무 화가 났어요.
예측된 감정: 분노
문장: 비록 실수가 있었지만 성장할 수 있는 좋은 기회였습니다.
예측된 감정: 중립
문장: 부트캠프를 하며 많이 배우고 성장한 것 같아서 행복합니다.
예측된 감정: 행복
Moviepy - Building video final_video_with_audio.mp4.
MoviePy - Writing audio in final_video_with_audioTEMP_MPY_wvf_snd.mp3
MoviePy - Done.
Moviepy - Writing video final_video_with_audio.mp4

감성분석후 자막, 음성을 넣은 동영상 출력

자막, 동영상 (moviepy 모듈)
음성 (gTTS 패키지)

7 프로젝트 결과 (demo 영상)



훌륭한 심사위원님 틀 앞이라 너무 긴장되네요.(불안)

8 도전 과제 및 해결 방안

문제점 1

부족한 4가지 감정의 데이터



1. 다른 도메인을 가진 CelebA 데이터랑 같이 학습시켜 성능을 올림

문제점 2

결과물의 감정 표현력



1. 다양한 데이터 셋 확보
2. 고해상도 3D 스캔 및 모델링 기술 적용
3. 감정 변화의 애니메이션 처리

문제점 3

환경 설정 및 버전 호환 문제로
고성능의 3D 모델 적용 못함



1. 가상 환경을 이용한 버전관리
2. 공식문서 및 커뮤니티 리소스 활용

9 향후 계획 및 발전 방향

개선점 및 추가 기능

더 다양한 감정 표현을 위해 GAN 모델 업그레이드
실시간으로 감정을 분석하고 반영하는 기능을 추가할 계획

확장 가능성

교육용 애플리케이션, 심리 치료 보조 도구 등으로 확장

시장 점유율 확대

차별화된 감정 시각화 동화책을 통해 동화책 시장에서 경쟁력을 확보

10 참고 문헌

<https://chickenrush.tistory.com/16>

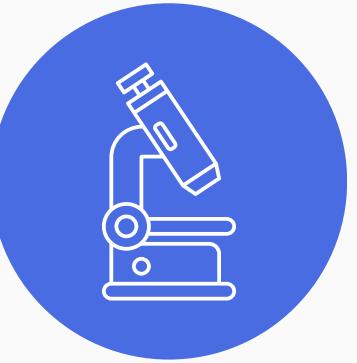
https://youtu.be/-r9M4Cj9o_8?feature=shared

<https://github.com/yfeng95/DECA>

<https://github.com/ascust/3DMM-Fitting-Pytorch>

[Yunjey Choi 외 5인. \(2018, september 21\). StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation.](#) Korea University, Clova AI Research, NAVER Corp, The College of New Jersey, Hong Kong University of Science & Technology

[Radek Danevcek 외 2인. \(2022\). EMOCA: Emotion Driven Monocular Face Capture and Animation.](#) Max Planck Institute for Intelligent Systems, Tübingen, Germany



감사합니다.
Thank You
