


2025년 컴퓨터정보공학부 졸업작품/논문 중간보고서

프로젝트 팀	<input type="checkbox"/> 졸업작품팀 <input checked="" type="checkbox"/> 졸업논문팀			
프로젝트명	생성형 AI를 이용한 3D 얼굴 표정 디테일 모델링			
팀명	DeepThinkers			
지도교수명	이혁준			
팀원 현황	학번	성명	캡스톤설계 교과목 이수학기	비고
	2021202087	장현웅	2025-1	2학기 휴학
	2020202067	나웅재	2025-1	

신청일자 : 2025. 06. 30.

신청자(팀장) : 장현웅 

지도교수 확인 : 이혁준 (인)



광운대학교
KwangWoon University

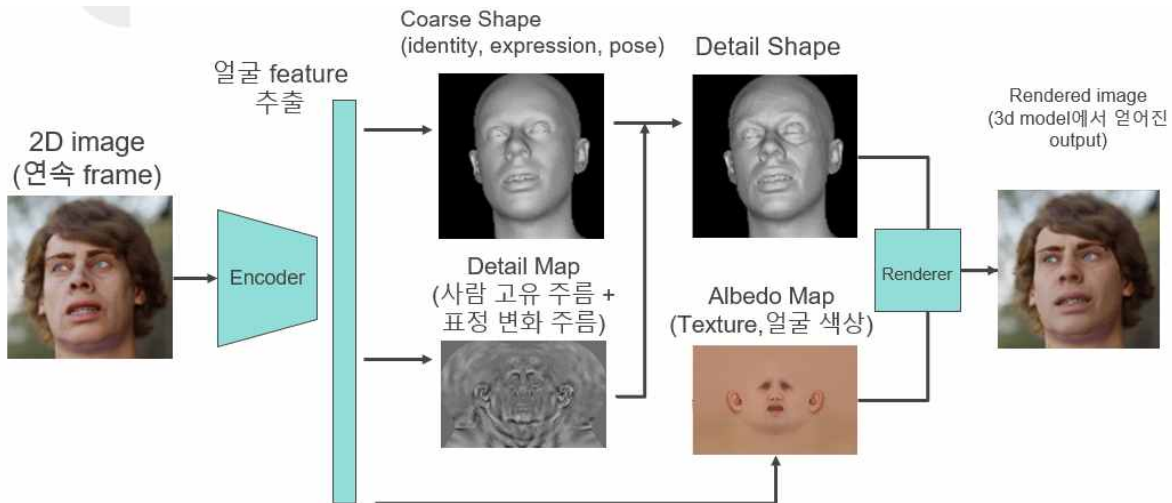
목 차

1. 과제의 개요	3
가. 배경 및 필요성	3
나. 과제의 목표 및 내용	4
다. 시스템 개요	4
2. 과제 수행 결과	5
가. 관련 기술	5
나. 중간 결과	10
다. 향후 계획	13
3. 과제의 평가	14
가. 계획 대비 진행도 평가.....	14
나. 개선 방안	15
다. 기타 보고사항	15
라. 과제 기대 효과	16
4. 별첨	17

1. 과제의 개요

가. 배경 및 필요성

2D 얼굴 이미지로부터 3D 얼굴 모델을 복원하는 3D 얼굴 재구성(3D Face Reconstruction)기술은 VR/AR, 애니메이션, 게임, 얼굴 인식, 의료 분야 등 다양한 분야에서 활용되고 있다. 이 기술은 일반적으로 대략적 얼굴 형상(Coarse Shape)을 구성한 후, 표정 정보를 추가하여 얼굴 움직임을 자연스럽게 구현하며, 디테일 맵을 통해 피부의 미세한 주름 및 표면 굴곡을 추가하고, 마지막으로 텍스처를 통해 피부 색상 및 전반적인 표면 특성을 반영하는 과정을 거친다.



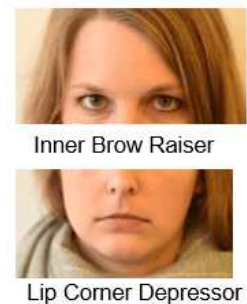
3DMM 방법은 PCA 기반 선형 계수를 사용해서 얼굴 mesh를 구성한다.

$$S = S_{\text{평균}} + B_{\text{exp}} * \beta + B_{\text{exp}} * \xi + \dots$$

이때 β 는 신원(identity), ξ 는 표정(expression)...등에 대응하는 성분이며 B 는 고정된 basis vector를 PCA로 미리 구한 것이다.

기존 3DMM(3D Morphable Model) 기반 기법은 PCA(Principal Component Analysis) 등으로 생성된 선형 모델을 사용한다. 이는 전체 얼굴의 전역 특징(평균적인 특징)만 재구성하며, 얼굴의 각 영역에서 나타나는 국소적이고 미세한 표정 변화를 충분히 모델에서 복원해내지 못했다. 이를 위해서 AU(Action Unit)을 통해 얼굴 각 부분의 움직임을 해석하는 방법이 제시되었다.

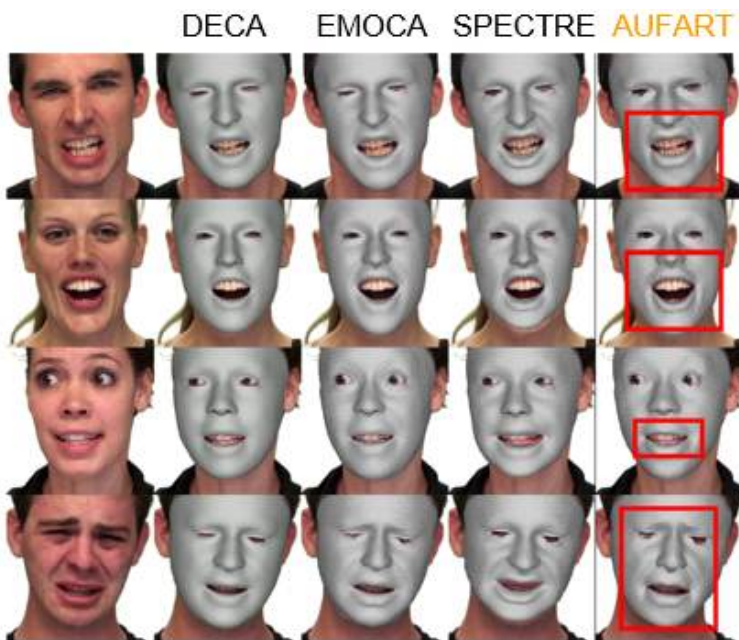
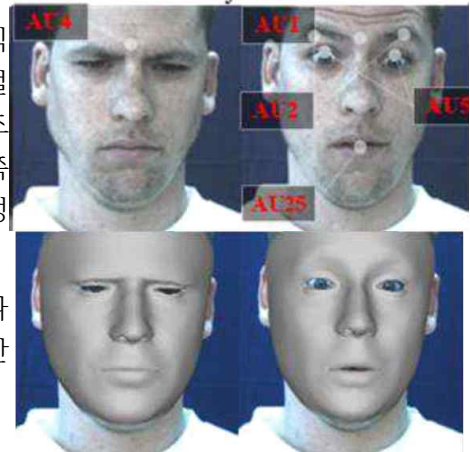
AU는 국소적 근육 집단의 움직임을 뜻한다. 얼굴 이미지의 관심 부위(AU-specific region)에서 각 AU 노드의 활성화 특징을 추출하여, AU1(눈썹 위로 올림), AU6(눈가 잔주름) 등 노드로 대응되어 그래프로 연결하며, 각 움직임이 나타나면 1, 아니면 0으로 표현되는 데이터이다. 예를 들어서, 오른쪽 그림과 같이 inner brow raiser와 lip corner depressor가 활성화 되어있으면, 화난 표정으로 볼 수 있다.



기존에 단순히 AU 정보를 도입하는 기존 방법은 각 AU 간의 상호 관계와 복합적 조합을 고려하지 않아, 실제 얼굴에서 관찰되는 미묘한 근육 움직임을 포착하는 데 한계가 있었다. 또한, 입력 이미지와 재구성된 얼굴 사이의 AU 활성화 차이를 최소화하는 AU 기반 loss 함수가 충분히 적용되지 않아 표정의 자연스러운 재현이 어려웠다.

AUFART는 이러한 한계를 해결하기 위해, 사전 학습된 AU 인식 모델에서 추출한 AU-specific feature와 얼굴 전체의 글로벌 feature를 함께 입력받아 Transformer의 cross-attention 구조를 통해 AU 간 관계와 표정의 의미적 구성을 학습한다. 오른쪽과 같이 활성화된 AU 간의 관계를 학습하여, 자연스러운 표정 정보를 반영하는 coarse shape를 복원하게 한다.

더불어, AU 활성도의 일치성을 유지하도록 설계된 AU 기반 다중 손실 함수를 도입하여, 대조군 기술들보다 자연스럽게 세밀한 표정 복원이 가능하도록 한다.



하지만 AUFART기술은 coarse shape만을 복원하므로 아래 사진에서 표정 변화로 인해 턱에 생기는 미세 주름 등을 구체적으로 표현하기 어렵다. 또한, 텍스처를 생성하지 않기 때문에 기존에 쓰이는 pretrained 모델을 사용해야 하였고, 기존 pretrained 모델 기반 Albedo map은 얼굴의 본연의 색상을 충분히 재현하지 못하며, 주름에 따른 색상 변화를 고려하지 않아 결과적으로 텍스처의 자연스러움이 떨어지는 문제가 있다.



이러한 한계점들을 극복하고, 더욱 사실적이고 정교한 3D 얼굴 모델을 복원하기 위한 통합 프레임워크 제안의 필요성이 대두되었다.

나. 과제의 목표 및 내용

본 과제의 최종 목표는 AUFART, HiFace, cGAN 딥러닝 기술들을 분석하고, 각 기술의 장점을 융합하여 기존 3D 얼굴 복원 기술의 한계를 극복하며, 사실적인 세부 표현이 적용된 3D 얼굴 모델을 생성하는 새로운 프레임워크를 제안하는 것이다.

1. AUFART의 AU 파라미터를 활용한 디테일 맵 생성 모듈 도입

HiFace의 SD-DeTail 모듈을 AUFART에 접목하여 디테일 맵(정적 및 동적 주름)을 추가한다.

이때, AUFART의 Transformer에서 얻은 표정 계수나 AU 파라미터를 SD-DeTail 모듈의 입력에 활용하여, 기존 DECA 기반의 표정 계수보다 더 정확한 동적 디테일을 생성한다.

2. cGAN을 활용한 텍스처 생성 모듈 도입

기존 DECA의 알베도(albedo) 계수 및 조명 계수 대신, cGAN 기반 텍스처 생성 모듈을 도입하여 더 사실적인 피부 질감과 색상을 가진 텍스처(Albedo Map)를 생성한다.

cGAN의 Generator에 원하는 근육 움직임을 AU 조건으로 도입함으로써, 각 근육 별 움직임 정보가 들어간 텍스처를 생성할 수 있도록 한다.

3. AUFART 학습에 사용되는 프레임 수 증가

연속적인 이미지 시퀀스로 학습되는 AUFART의 입력 프레임 수를 증가시켜, Transformer가 AU와 표정 변화의 시간적 연속성을 더 세밀하게 학습하도록 유도한다. 이를 통해 표정 계정의 정확도를 높이고 전반적인 재구성 성능을 개선한다.

다. 시스템 개요

본 시스템은 입력되는 2D 이미지 sequence로 부터 AUFART를 통해 표정 정보가 확실하게 드러나는 coarse shape을 생성한다. 이후 얼굴에 구체적으로 드러나는 사람 고유의 주름(정적 디테일) 및 얼굴 표정 변화 시 얼굴 표면에 생기는 표정에 따른 주름(동적 주름)의 미세한 주름을 SD-DeTail 모듈을 통해 coarse shape에 반영하며, cGAN 을 통해 메쉬에 사실적인 텍스처를 생성하여 렌더링한다.

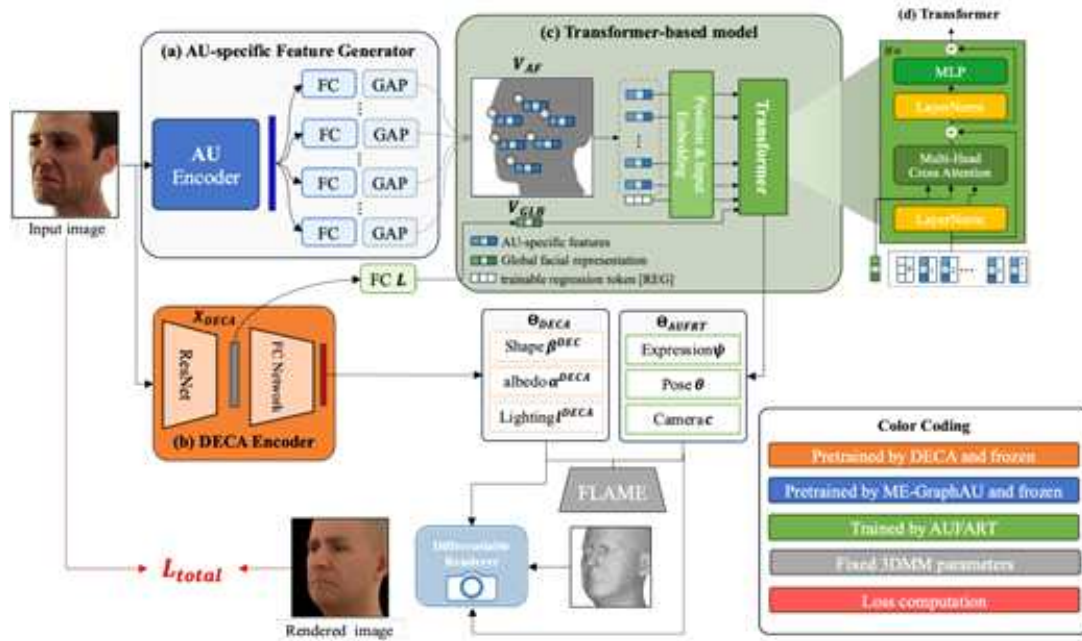
얼굴의 국소 부위의 움직임인 AU 정보를 사용해 각 모듈을 통합하여, 2D 이미지로부터 3D 메쉬를 복원하는 톨을 구성하도록 한다.

2. 과제 수행 결과

가. 관련기술

본 프로젝트에서 활용하고자 하는 주요 관련 기술은 다음과 같다.

AUFART (Action Unit based 3D Face Reconstruction Technology)



AUFART는 단일 2D 얼굴 이미지로부터 AU 정보에 기반하여 3D 얼굴을 복원하는 딥러닝 기반 프레임워크로, AU-specific Feature와 Global Facial Feature 간의 상호작용을 Transformer를 통해 학습함으로써, 주름과 같은 미세 표정 재현 능력을 향상시킨다. 주요 구성 요소는 다음과 같다.

(a) AU-specific Feature Generator

입력 이미지로부터 AU-specific Feature를 추출하는 모듈로, CNN + General Average Pooling + Fully Connected 레이어 구조로 이루어진다. 총 27개의 특징 벡터 $\{v_1, v_2, \dots, v_{27}\}$ 을 생성하며 각 AU 벡터는 512차원이다. AU-specific Feature는 transformer의 Query로 전달되게 된다.

(b) DECA Encoder

입력 이미지로부터 DECA 기반 ResNet + FC Network 구조를 통해 전역 얼굴 특징(global feature)을 추출한다. 여기서 추출되는 파라미터는 다음과 같다:

β : 신원(ID; 고유 얼굴 형상) 계수 / α : 피부 텍스처(albedo) 계수 / δ : 조명 (lighting) 계수

해당 파라미터들은 모두 DECA에서 사전학습된 고정 파라미터로 사용되며 학습되지 않는다. AUFART는 AU에 따른 미세 표정(expression) 변화가 어떻게 coarse shape에 재현되는지를 목적으로 하기 때문이다.

(c),(d) Transformer-based model

Transformer는 AU-specific feature와 전역 얼굴 특징(global feature)를 입력으로 하여, 표정(Expression, ψ), 자세(Pose, θ), 카메라(Camera, c) 파라미터를 예측하게 된다. 이 파라미터들이 최종적으로 추후 FLAME 모델에 들어가 3D 메쉬 구조를 생성하게 된다.

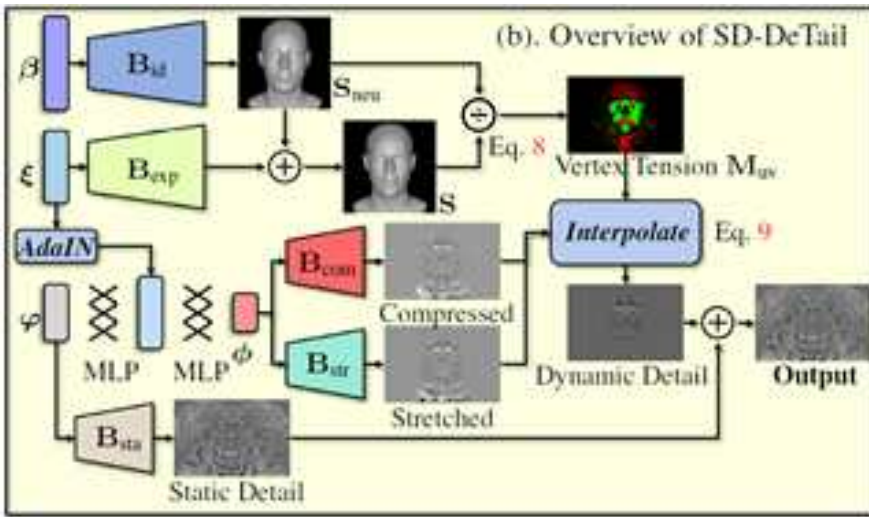
과정은 다음과 같다. 우선, 앞 단에서 받은 총 27개의 AU가 Position Embedding을 통해 위치 정보를 가지도록 한다. 맨 앞쪽에 [REG] 토큰을 추가하며, 이것은 transformer 전체가 예측한 결과가 된다. 이후, Multi-head Cross Attention을 통해 AU-specific Feature(Query)와 전역 얼굴 특징(Key, Value) 간 관계를 학습하게 된다.

Transformer 내부는 MLP 블록을 여러 회 반복하는 구조로 되어 있으며, AU-specific feature와 Global Feature를 동시에 처리하도록 한다.

(e) flame decoder & renderer

위에서 얻은 표정, 자세, 카메라 계수와, DECA에서 얻은 신원, 피부 텍스처, 조명 계수값을 FLAME 모델에 대입해 3D mesh를 생성한다. 이후, differentiable renderer로 3D 메쉬를 2D 이미지로 렌더링하고, 입력 이미지와의 자가지도학습을 수행한다.

SD-DeTail (Static and Dynamic Decoupling for DeTail reconstruction)



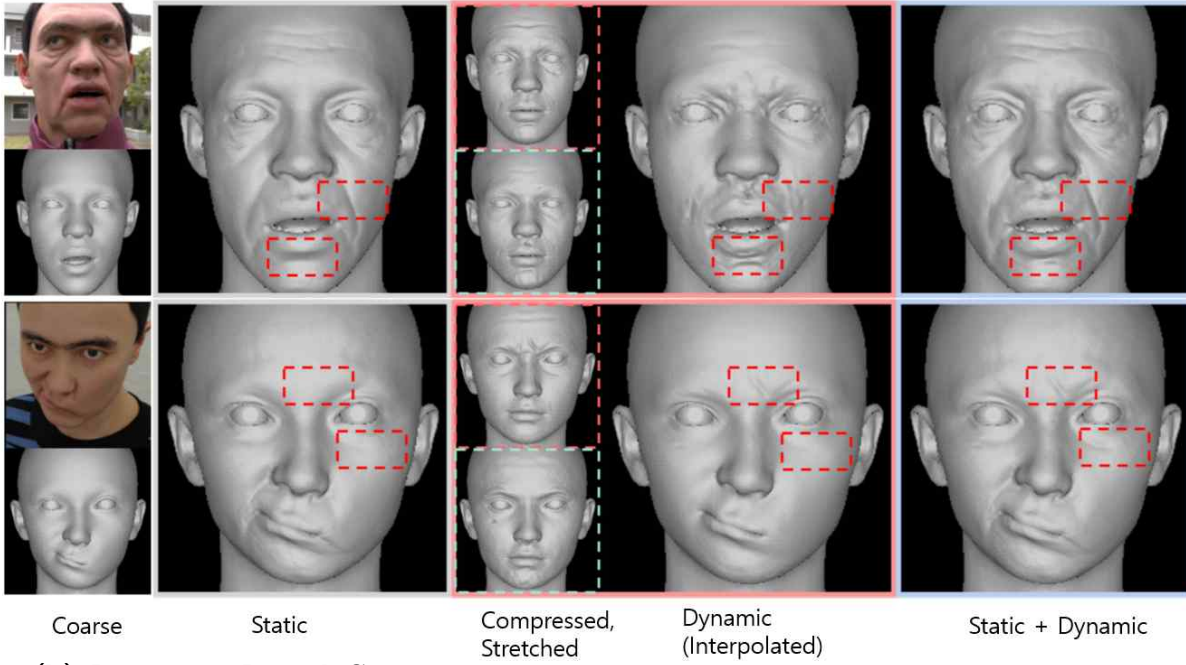
SD-DeTail은 HiFace에서 제안된 정적/동적 디테일 분리 모듈로, 3D 얼굴 재구성에서 정적 디테일(사람 얼굴 고유의 주름)과 동적 디테일(표정 변화에 따른 주름)을 분리해 고해상도의 디테일을 사실적으로 재구성하기 위해 고안되었다. 기존에는 두 가지 미세 주름이 혼동되어 나타나, 얼굴 표정의 변화 시 미세 주름의 재현이 사실성이 떨어진다. 본 모듈은 두 종류의 디테일을 다음과 같이 분리하여 처리한다.

(a) Static Detail generator

정적 디테일(사람 고유의 주름이나 피부 특성)은 PCA 기반 displacement basis B_{sta} 를 구성하여, 입력 이미지로부터 추출된 정적 디테일 계수 ϕ 를 통해

$$D_{sta} = \bar{D}_{sta} + \phi B_{sta}$$

형태로 복원된다. 이때 정적 디테일 계수는 coarse shape에 따라 3D displacement map 상에서 위치가 결정되는 디테일 맵 형태로 표현된다.



(b) Dynamic Detail Generator

표정에 따라 변하는 동적 디테일은 두 개의 극단적 표정 상태, 즉 압축(Compressed)와 신장(Stretched) 표정을 선형 보간해서 구성한다. 높은 복잡도로 인해, dynamic detail을 2D 이미지로부터 바로 얻어내기 어렵기 때문이다. 두 극단적 표정 상태에 대응하는 displacement basis B_{com} , B_{str} 를 각각 구성한 후, 표정 계수(위 그림의 ξ)의 분포에 맞게 앞에서 얻은 정적 계수 ϕ 의 분포를 AdaIN으로 변형하고 MLP로 통과시켜 계수 ϕ 를 생성한다.

이로부터 압축 및 신장 displacement map D_{com} , D_{str} 를 생성한 뒤, vertex tension map을 사용해 두 압축된 표정과 신장된 표정을 선형 보간한다.

vertex tension map은 얼굴 위 각 부위를 정점으로 만들어, 정점 간 연결을 그래프화 한 뒤, 정점 간 거리가 얼마나 멀어지고 가까워지는 기반으로 장력의 개념을 형상화한 것이다. 즉, 국소적으로 얼굴에서 압축되는 부분과 이완되는 부분을 계산해, 이완된 표정과 신장된 표정의 displacement map을 선형 보간한다.

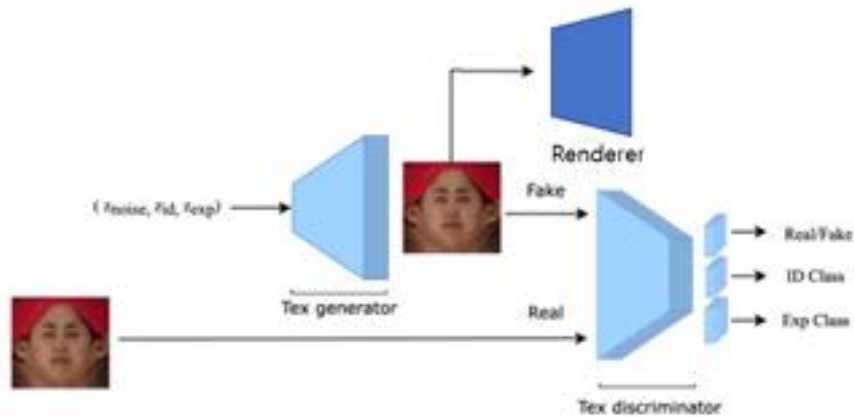
$$t_{vi} = 1 - \frac{1}{K} \sum_{k=1}^K \frac{\|e_k\|}{\|e'_k\|}$$

$$D_{dyn} = M_{uv}^+ \odot D_{com} + M_{uv}^- \odot D_{str}$$

SD-DeTail 모듈로 AUFART에서 처리한 Coarse Shape 뒤에서 Detail map을 생성하면, 얼굴에 표현되지 않는 미세 주름을 정밀하게 표현할 수 있을 것이다.

cGAN (Conditional Generative Adversarial Networks)

기존의 GAN(Generative Adversarial Network)는 무작위 벡터 $z \sim N(0,1)$ 를 입력으로 받아, Generator가 이를 해석해 데이터(얼굴 이미지)를 생성하고, Discriminator는 해당 데이터가



실제(real)인지 생성(fake)인지 판단하면서 경쟁적으로 학습하는 구조로 설계된다. 이때 입력되는 z 는 데이터 생성의 다양성을 부여하는 잠재 벡터(latent vector)로서, 동일한 구조 안에서 여러 결과물을 무작위적으로 생성하도록 하는 확률 변수이다. 기본 GAN은 어떠한 제어 변수도 없기 때문에, 생성자가 어떤 종류의 데이터를 생성할지 사용자가 직접 개입할 수 없다.

cGAN(conditional GAN)은 기존 GAN 구조에 조건 벡터 c 를 추가해, Generator가 단순히 무작위 z 만이 아니라 특정 조건을 만족하는 데이터를 생성하도록 학습한다. c 는 위 그림에서 인코더로부터 추출된 z_exp , z_id 등이며, 클래스 라벨, 텍스트 설명, 얼굴 표정 등을 포함하고 있다. cGAN에서 Generator는 다음과 같은 수식으로 표현된다.

$$G: (z, c) \rightarrow x'$$

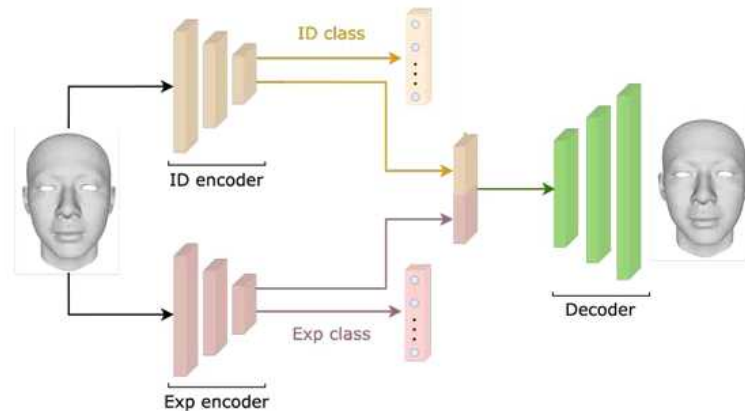
Discriminator는 마찬가지로 이 조건을 함께 입력된 데이터 x 와 받아, 입력된 데이터 x 가 실제 데이터셋에 속하는지, 생성된 데이터셋에 속하는지 판단하게 된다.

즉, Generator는 조건을 만족하는 다양한 형태의 데이터를 만들고, Discriminator는 조건을 위반하는 위조 데이터를 판별하며, 양쪽이 경쟁적으로 학습함으로써 조건에 부합하면서도 현실감 있는 데이터를 생성할 수 있게 된다.

논문에서는 z_id , z_exp , 그리고 z_noise 를 generator에 동시에 입력한다. z_id , z_exp 는 SAE 의해 고차원 얼굴 mesh 데이터를 잠재 공간으로 매핑한 잠재 벡터로, 각각 얼굴의 신원, 표정 feature를 가진다. z_noise 는 원래의 GAN에서 사용되는 순수 무작위 벡터이다

cGAN을 통해 본 연구에서는 실제 이미지와 구분되지 않는, 재현율 높은 텍스처를 생성하는 것을 목표로 한다.

SAE (Supervised AutoEncoder):



SAE는 conditional GAN에서 제안한 3D 얼굴 메시 데이터를 보다 효율적으로 인코딩하기 위한 방법으로, 신원(identity)와 표정(expression) 특징을 서로 독립된 잠재공간에 분리하여 학습하는 특징을 가지고 있다.

일반 오토인코더는 단일 인코더로 입력을 잠재 벡터 z 하나로 압축한다. 이 방식은 내부의 차원 중 어떤 차원 ID이고, 어떤 차원이 표정인지 명확하게 구분하지 않는다. 따라서 z 중 일부 값만 바꾸고 싶어도 어떠한 값이 ID이고 어떤 값이 표정인지 정확하게 알지 못해 바꾸지 못하는 문제가 있었다. 모든 정보가 얹혀 있기 때문에, 제어가 불가능하다. 예를 들어, z 의 3번째 원소를 0.5에서 0.7로 바꾸면, 얼굴, 피부색, 표정이 모두 바뀌어 버린다.

기존 3DMM 방법들에서 identity 계수, expression 계수 등이 PCA의 방법으로 서로 basis가 나뉘어졌긴 하지만, 각 basis가 수직이라고 해서 의미까지 분리되는 건 아니다. PCA의 결과로 이미지 픽셀상에서 구한 성분에 실제로 표정, 조명, 나이 등 정보가 모두 섞여 있을 수 있다.

따라서 학습 과정에서 SAE는 처음부터 비선형 신경망으로 각 파라미터를 분리된 잠재 공간에서 학습시키며, ID encoder는 라벨로 ID 분류 클래스를, Expression encoder는 expression 분류 클래스로 정답 라벨을 붙여 학습시킨다.

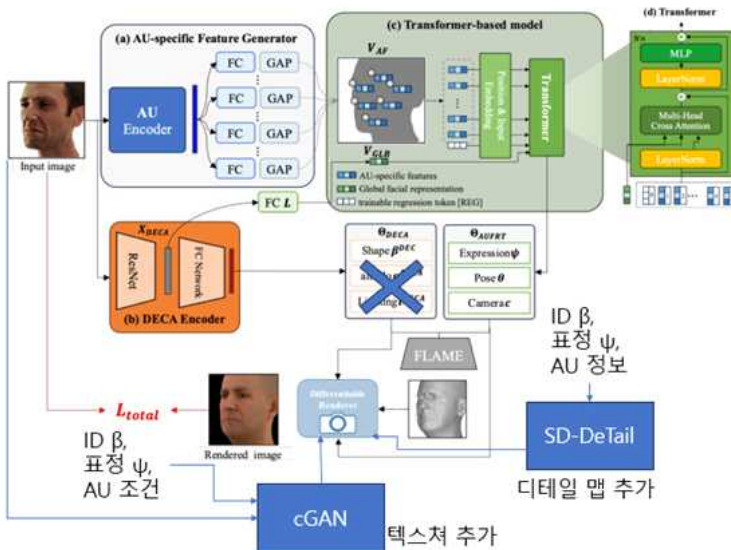
이를 통해, 인코더에서 최종적으로 z_{id} 와 z_{exp} 는 각각 추출되어 id와 표정을 잘 분류하는 벡터로 학습되며, PCA에서 단순히 데이터의 통계적 특징만 보고 어떤 성분이 의미있는 정보를 담는지 이해하지 않는 방법과 차이를 보인다.

Conditional GAN 논문에서는 SAE를 사용하면 정보가 명확하게 분리되도록 할 수 있어, 표정 계수 조절 시 신원이 바뀌는 등의 문제를 해결할 수 있음을 밝힌다.

나. 중간 결과

상위 설계

현재 프로젝트는 각 모듈을 통합하는 아이디어를 구체적으로 설계하고 구현하기에 앞서 각 모듈의 원리를 이해하고 각 모듈 간 수학적인 접합점을 모색하고 있는 데에 있다. 따라서 각 모듈의 구동 및 작동에 대한 아이디어는 향후 변경될 수 있다.



본 프로젝트에서 제안하는 시스템은 기존 3D 얼굴 복원 파이프라인에 새로운 모듈들을 통합하여 사실적인 디테일과 텍스처를 추가하는 것을 목표로 한다.

1. 시퀀스 입력

연속된 2D 얼굴 이미지 프레임이 입력된다. 입력은 정적 이미지가 아닌 sequence 단위로 구성되어 있어 시간적 연속성을 고려한 학습이 가능하다.

2. 특징 추출 및 Coarse Shape 복원 (AUFART 기반)

입력된 이미지로부터 얼굴의 Identity(신원), Pose(자세), Expression(표정)에 해당하는 Coarse Shape 정보와 AUFART가 학습한 Action Unit 기반의 표정 계수(ψ)를 추출한다. AUFART의 Transformer는 AU 간의 상호 관계를 모델링하여 미세 표정 변화 정보를 포착하게 된다.

3. 디테일 맵 생성 (HiFace SD-DeTail + AU)

복원된 Coarse Shape와 표정 계수(ψ)는 HiFace 기반 SD-DeTail 모듈로 전달된다.

정적 디테일(고유 주름 등)은 PCA 기반 displacement basis로 생성되며, 동적 디테일(표정 변화에 따른 주름)은 두 개의 표현 상태 간 보간을 통해 얻어진다. 이 과정에 AUFART의 AU 파라미터를 함께 활용하여, 국소적인 얼굴 변화에 따른 미세한 표정 변화를 정확하게 복원하는데 기여한다.

4. 텍스처 생성 (cGAN)

추출된 ID 및 Expression (또는 AU) 계수를 조건으로 하는 cGAN이 Albedo Map을 생성한다. 설정에 따라 다양한 조명 조건을 조절할 수 있으며, AU 계수를 조건으로 받아 주름 디테일에 일관성 있는 사실적인 피부 텍스처를 만들어낸다.

5. 3D 모델 통합 및 렌더링

생성된 Coarse Shape, Detail Map, Albedo Map이 Renderer에 전달된다. Renderer는 이 세 요소를 통합하여 완전한 3D 얼굴 모델을 구축하고, 이를 실제와 같은 2D 이미지로 렌더링하여 최종 결과를 출력한다.

상세 설계

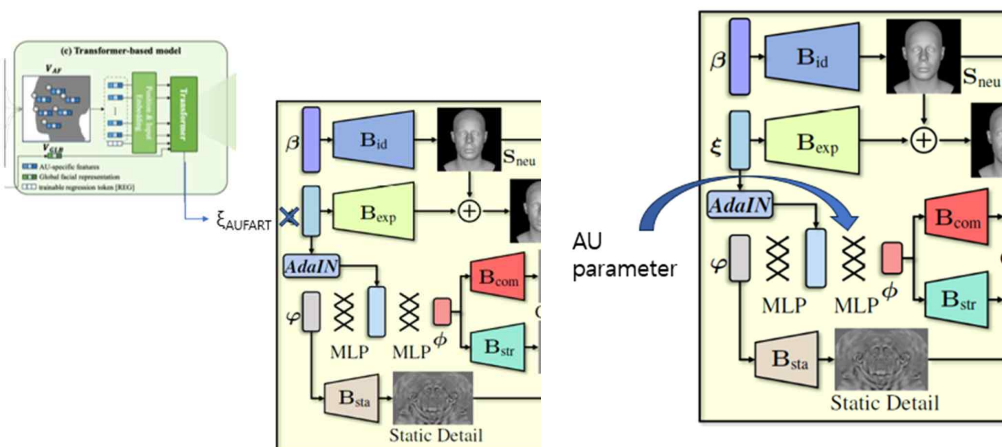
현재까지 논의된, 모듈과 모듈 간 접합점에 대한 구조를 설명하도록 한다.

1. AUFART의 AU 파라미터를 활용한 디테일 맵 생성 모듈 도입

AUFART의 렌더링 전에 SD-DeTail 모듈을 도입해 디테일 맵을 추가할 수 있다. 이때, AU 파라미터를 활용해 동적 디테일 맵의 품질을 향상시킨다. HiFace에서는 동적 디테일 계수를 구할 때 정적 디테일 계수의 분포에 변형된 표정 계수의 분포를 반영해(AdaIN), 이를 MLP에 통과시킨다.

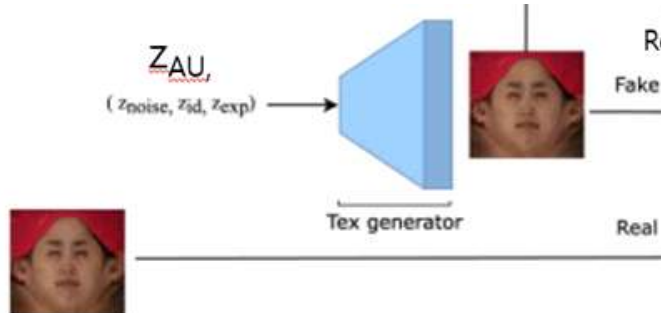
$$\phi = \Phi\left(\sigma(\xi)\left(\frac{\varphi - \mu(\varphi)}{\sigma(\varphi)} + \mu(\xi)\right)\right)$$

이때 사용되는 표정 계수는 기존 기술 DECA에서 PCA를 통해 생성된 것으로, 얼굴 전체적인 근육의 움직임을 반영할 뿐 각 근육의 정확한 움직임을 반영하지 않았다. 따라서, AdaIN에서 AUFART의 Transformer에서 얻은 표정 계수를 사용하거나, AU 파라미터를 MLP의 입력에 같이 사용한다면, 더 정확한 동적 디테일을 생성할 수 있을 것이다.



2. cGAN을 활용한 텍스처 생성

앞선 AUFART의 그림에서 텍스처를 생성하기 위해 기존 기술 DECA의 알베도 계수 α 와 조명 계수 1을 사용하는 대신, cGAN을 사용한 텍스처 생성 모듈을 도입하여 더 사실적인 텍스처를 생성할 수 있다. 특히, cGAN의 Generator에 원하는 근육 움직임을 AU 조건으로 도입함으로써, 각 근육 별 움직임 정보가 들어간 텍스처를 생성할 수 있을 것이다.



3. AUFART의 학습에 사용되는 프레임 수 증가

AUFART는 연속적인 이미지 시퀀스로 학습시킬 수 있다. 학습을 위해 한번에 입력되는 프레임 수를 증가시키면, Transformer가 어떤 AU가 어떤 표정과 연관되어 있는지 더 세밀하게 학습시킬 수 있으므로, 더욱 정확한 표정 계수를 생성할 수 있을 것이다.

기능 테스트 (중간 결과)

현재까지의 프로젝트 진행 상황은 주로 cGAN 오픈소스 코드 구동 환경 설정 및 결과물 분석에 집중되어 있다. AUFART와 HiFace 모듈의 오픈소스 코드를 구하지 못하여 직접 구현이 지연되었다.

1. cGAN을 활용한 expression intensity 변화 실험

표정의 강도를 단계적으로 변화시키면서, 얼굴을 재구성하였다. 별첨 4.1의 그림에서는 다양한 표정 강도에 대해 입이 벌려진 정도를 연속적으로 조절하면서, generator가 자연스러운 텍스처 생성이 가능한지 확인하기 위한 실험을 진행하였다.

2. cGAN을 활용한 expression 종류 변화 실험

표정의 종류를 변화하면서, 관련연구 논문에서 제시한 여러 종류의 표정에 대해 텍스처가 자연스럽게 생성되는지를 확인하고자 하였다. 별첨 4.2의 그림에서는 다양한 표정에 대해 3D 재구성을 진행해보면서, generator가 자연스러운 텍스처 생성이 가능한지 확인하기 위한 실험을 진행하였다.

3. 실험 결과

눈가 부분에 검은 색 얼룩이 공통적으로 생기며, 매쉬의 표정 생성과 텍스처의 표정 생성 부분 사이 정합성이 떨어지는 실험 결과도 4.3에 첨부하였다. 이러한 문제의 원인에 대하여 추가적으로 조사하면서, cGAN 모듈을 활용한 텍스처 생성을 도입하는 시도 중에 있다.

다. 향후 계획

1. 논문 재현 및 평가 (계속 진행):

현재 40% 완료된 cGAN의 Experiment 부분에서 대조군과의 비교 실험을 진행할 것이다. HiFace와 AUFART에 대한 코드 작성을 계속하며, 각 논문에서 제시한 실험을 재현하고 방법론의 타당성 검증하고자 한다.

2. 학습 프레임 수 증가 실험:

HiFace 모듈 통합 후, AUFART가 처리하는 입력 프레임 수를 확장하여 표정 변화의 시간적 연속성을 더욱 잘 처리하는지 확인하는 실험을 수행할 것이다.

3. 모듈 통합 및 성능 개선:

AUFART의 AU 파라미터를 활용하여 HiFace의 SD-DeTail 모듈을 강화하는 인터페이스를 설계하고 구현할 것이다. AU 파라미터를 조건으로 활용하는 cGAN 구조를 설계하고 Loss function을 최적화하여야 한다.

4. Prototype 구축 및 학습:

설계된 통합 프레임워크를 기반으로 프로토타입을 구축하고, 실제 얼굴 이미지/영상 및 합성 데이터를 혼합하여 학습을 진행한다.

5. 성능 측정 및 개선:

REALY benchmark를 통해 coarse shape 및 디테일 맵 복원 성능을 정량적으로 평가하고, SSIM, FID score를 통해 텍스처(albedo map) 복원 성능을 평가한다. Optical flow consistency를 통해 애니메이션에서의 시간적 일관성을 평가한다. 시각적으로 주름/모공 디테일, 표정 자연스러움 등 정성 평가를 수행하여 기존 방법 대비 개선 사항을 확인할 것이다.

5	중간 보고서 작성 및 발표	6월 말
6	기존 방식과 성능 비교, 모델 개선	7월~8월
7	실험 고도화 및 최적화	8월~9월
8	다양한 데이터셋 적용, 성능 평가	9월
9	최종 모델 평가 및 추가 연구 수행	10월
10	결과 정리 및 논문 작성	10월~11월
11	발표 자료 제작	11월

3. 과제의 평가

가. 계획 대비 진행도 평가

기존 일정을 진행하면서 완료한 세부 달성 목표와, 진행중인 세부 달성 목표를 각각 요약하면 다음과 같다.

완료한 세부 달성 목표

활용할 수 있는 오픈소스 코드를 찾고 코드 구조를 파악

현재 cGAN 오픈소스를 확보하고 내용을 분석하였으며, 어떤 방식으로 구조가 구성되어있는지 파악하였다. 이를 통해 cGAN을 AUFART, HiFace와 합할 때 코드를 용이하게 다룰 수 있게 되었다.

cGAN 논문에서 제시된 바가 어떻게 구현되었는지 이해

실제로 환경을 구성하여 코드를 구동시켜 보면서, 핵심 모듈이 논문에서 제시된 바와 같이 어떻게 구성되었는지 이해할 수 있게 되었다.

Colab에서 구동하는 환경에 익숙해지기 / inference 코드 구동

cGAN 모델 환경을 구축하고, Colab에서 구동하는 환경에 익숙해지면서 기초적인 python inference 실험을 진행할 수 있게 되었다.

진행 중인 세부 달성 목표

HiFace코드 구축 & AUFART 코드 확보

HiFace의 경우 오픈소스를 확보할 수 없어, 학기 중 머신러닝 프로젝트를 진행한 후 논문 내용을 토대로 코드를 기초부터 작성하여야 한다. 이로 인해 계획에 차질이 생겼으며 추후 진행 예정이다.

cGAN 논문에서 제시한 각 방법론의 타당성 검증 (40% 완료)

cGAN의 코드를 통해 inference를 진행하면서, 논문의 Experiment 부분에서 실제로 다른 대조군 기술들과 비교하여 얼마만큼의 성능차이가 생기는데 대해 실제 실험으로 검토할 필요성이 있다. 더불어, HiFace, AUFART에 대한 code는 아직 작성 중이기 때문에, 해당 기술들에 대한 논문 내용의 실험적인 타당성 검토는 지연되고 있다.

AUFART 코드 확보 및 학습 Frame 수 증가 실험

AUFART의 입력으로 들어가는 학습 Frame 수를 늘려보면서, AUFART 모델의 성능이 더 좋아지는지에 대한 실험을 진행하여야 한다. 하지만 AUFART 역시 코드를 확보하는 과정 상에 있으며, 이에 따라 추후 HiFace 코드 합할 경우 차후 진행하기로 하였다.

나. 개선 방안

원래 계획을 진행하면서 하기로 한 내용 외에도, 다음의 주요 논의 대상에 대한 개선방안을 고려해보도록 하였다.

표정 제어 및 성분 분리도 향상을 위한 Auto Encoder 비교 실험

AUFART, HiFace 등 기존 3D 얼굴 복원기술은 DECA나 EMOCA에서 제공하는 선형 PCA 기반 표정 계수를 사용하는데, cGAN 논문에서는 표정 성분과 ID(신원) 성분이 완전히 분리되지 않는 것을 지적하며 이것을 해결할 새로운 방안으로 SAE(Supervised Auto Encoder)를 제안한다.

따라서, 표정 조절 시 신원 정보가 함께 변형되는 문제가 확인되고 SAE가 유의미하게 성능 향상을 보인다면, 해당 기술을 채택해 AUFART 또는 cGAN 조건 계수로 도입할 예정이다.

cGAN 기반 텍스처의 부자연스러움 해결 방안

현재 cGAN 생성 결과에서 나타나는 눈가 및 턱 밑의 검은 얼룩/음영 문제 해결이 시급하다. 논문에서 제안된 코드를 제안된 외부 3D 렌더러로 렌더링해보았으나, cGAN 모델 자체의 학습 한계로 인해 텍스처의 품질이 저하되는 것으로 보인다. 학습해보았던 generator - discriminator 아키텍처를 분석하고, 고해상도에서 문제없이 텍스처를 생성하도록 구조를 수정해야 한다. SD-DeTail 모듈에서 처리되는 Static & Dynamic Detail (정적& 동적 주름)에 대한 정보를 받아서, 이것을 cGAN에서 조건으로 하여 텍스처를 렌더링할 때 사용할 수 있는지 생각해보도록 한다.

AU 정보의 HiFace 모듈 사용 구체화

AUFART의 AU정보를 HiFace의 SD-DeTail 모듈의 Dynamic Detail 생성에 어떻게 효과적으로 통합시킬 것인지에 대한 구체적인 인터페이스 설계가 필요하다. AUFART의 transformer에서 생성된 표정 계수(ψ)를 HiFace의 AdaIN 또는 MLP 입력에 통합하는 방법, 또는 Vertex tension ap에 직접적으로 사용하는 방법을 검토한다.

다. 기타 보고사항

오픈소스 코드 실제 구현에 대한 예상 기간

HiFace 논문의 구조를 정리하여 논문에서 제시된 구조를 Prototype으로 구축하는 데에는 1달 정도의 시간이 걸릴 것으로 예상되나, 하이퍼파라미터 조정, 코드 구조 변경, 최적화 등에 예상보다 어느 정도의 시간이 더 필요한 것으로 논의 중이다.

라. 과제 기대 효과

더욱 사실적인 얼굴 표현 재현

기존 DECA나 EMOCA 기반 모델들이 정적 표현에 치우쳐 자연스러운 표정 재현이 어렵다는 한계를 가졌던 반면, 본 프로젝트는 AU 기반 표정 계수와 SD-Detail 모듈의 정적·동적 분리를 통해 개개인의 주름과 감정 표현을 보다 섬세하게 구현할 수 있을 것이다. 아바타 생성, 표정 인식 기반 인터랙션 시스템 등에서 유용하게 활용될 수 있다.

3D 모델 제작 비용 절감

인코더에서 도출된 잠재 벡터(z_{id} , z_{exp})와 AU 파라미터를 조건 입력으로 활용함으로써, 사용자는 명시적으로 표정을 조절하거나, ID를 고정한 채 다양한 표정을 생성하는 제어가 가능해진다. 이로써 하나의 인물 이미지로부터 다양한 표현 상태의 고품질 3D 모델을 생성할 수 있으며, 이는 게임, 애니메이션, 메타버스 아바타 자동 생성 등에서 제작 비용 및 인력 소모를 절감할 수 있는 기반이 된다.

프레임 기반 학습을 통한 시계열 표현의 자연스러움 향상

AUFART의 입력 프레임 수를 확장함으로써 Transformer 기반 구조가 시간적 연속성을 더 정밀하게 학습하도록 유도하는 것이 성공할 경우, 급격하게 달라지는 얼굴 표현에서도 프레임 간 표정의 부드러운 전이가 가능해진다. 표정이 변화하는 비디오클립을 제작하거나, 감정 추론 등에서 조금 더 안정적으로 사용될 수 있을 것이다.

멀티모달 통합

마지막으로, 본 프레임워크는 오디오 기반 표정 생성, 립 리딩을 통한 자연스러운 입 움직임 등의 멀티모달 통합에서도 기반 기술로 활용 가능하다. AU 기반 조건 제어, 정적/동적 디테일 분리, Texture와 Detail 간 연계 방식 등은 모두 모듈화 가능한 구조로 되어 있어, 다양한 modality와의 연동 연구에도 기여할 수 있다.

4. 별첨

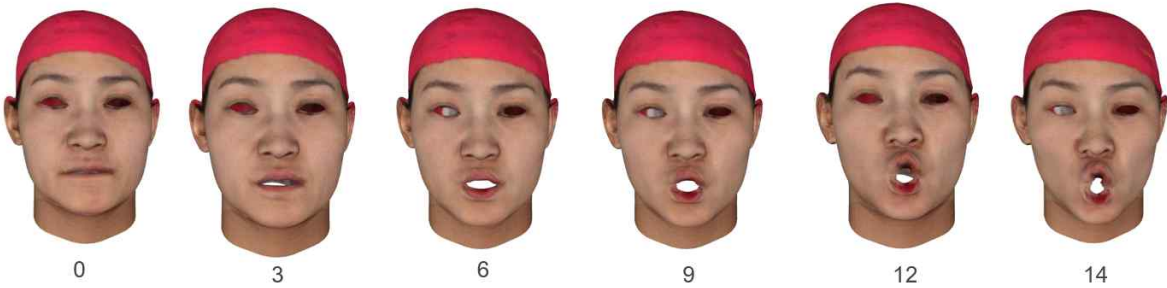
cGAN inference 코드

[3DFaceCAM \(fork\).ipynb](#)

Expression intensity 변화

표정의 intensity 를 단계적으로 변화시키며 얼굴을 재구성.

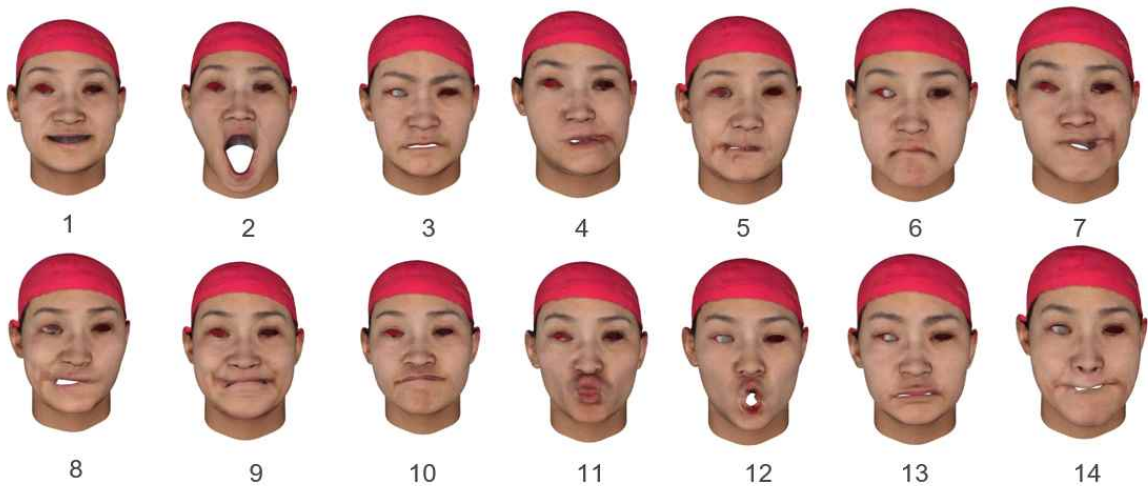
표정의 강도를 0.0~1.5로 조절하며 동일한 표정의 생성 결과 변화를 확인



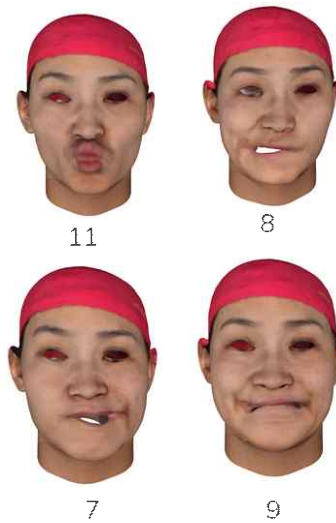
Expression 종류 변화

동일한 Identity에서 One Hot Encoding으로 표정을 구분하여 결과 도출.

Id-Expression Feature를 Auto Encoder로 분리하여 Id를 유지한 채 표정만 변화시킬 수 있음을 확인.



정합성이 다소 떨어지는 실험 결과



위의 11, 9번 모델의 경우 복잡한 입 위치의 표정에 대해서도 텍스처와 메쉬 간 위치가 정합하지만, 7,9번 (쏠린 입, 보조개 표정) 의 경우 얼굴 메쉬의 위치에 대해 실제 이미지와 비교해보았을 때 텍스처가 적용되어야 하는 부분이 다소 차이가 있음을 반복해서 확인하였다.