

HiFace: High-Fidelity 3D Face Reconstruction by Learning Static and Dynamic Details

저자: Zenghao Chai, Tianke Zhang, Tianyu He, Xu Tan 등

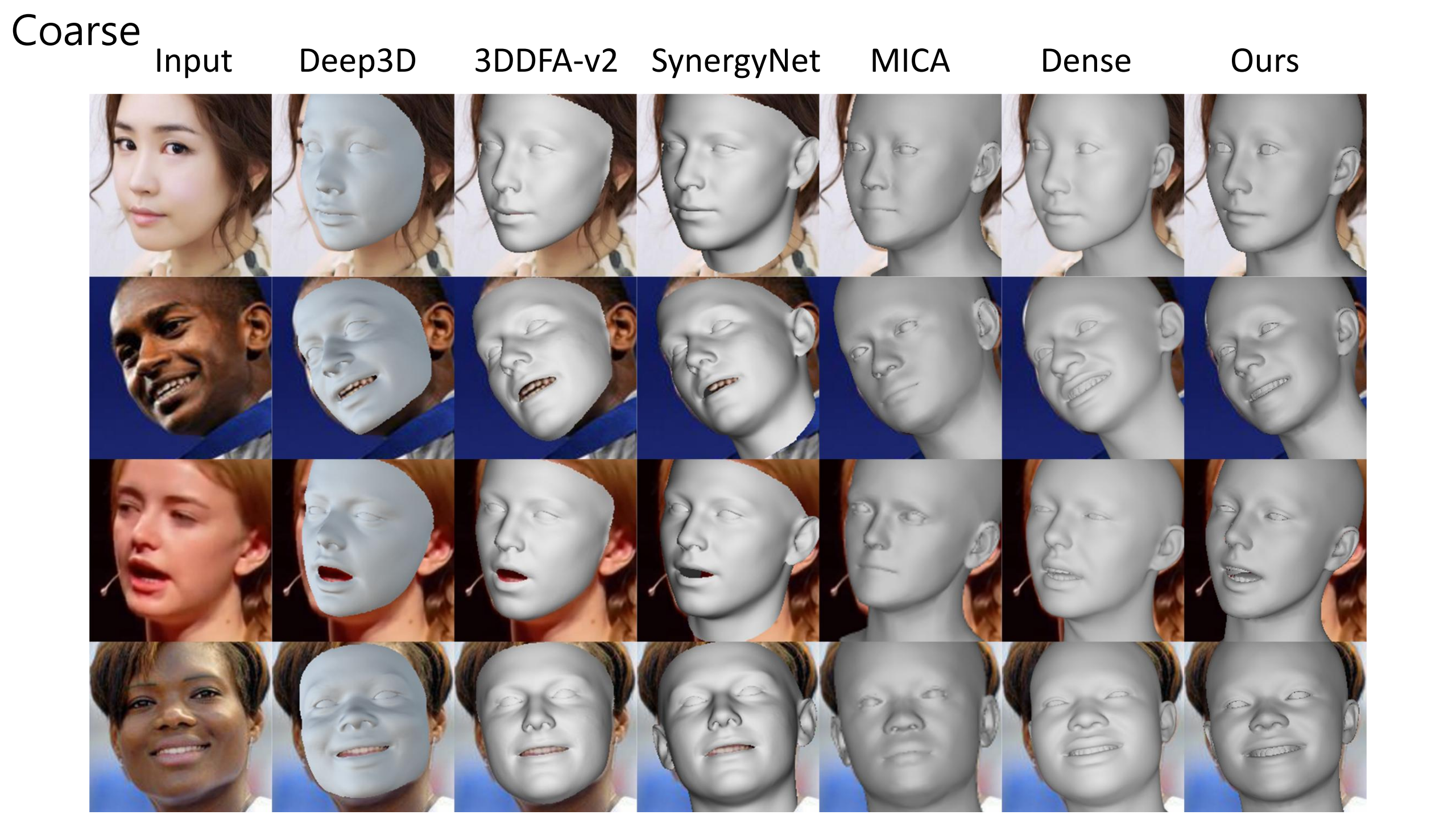
학회: ICCV 2023

3D face Reconstruction?

- 2D 얼굴 이미지로부터 3D 얼굴 mesh를 복원(재구성) 하는 것.
- VR, 게임, 페이스 리깅, AI 아바타 등에서 사용
- HiFace 기술은 그 일부

3DMM 기반 접근법

- 얼굴을 선형 모델로 단순화해서 복원.
- Coarse shape 만 복원 가능
 - 주름 등의 미세한 detail 표현 힘들
- 최근 Displacement Map 활용해 Detail 복원 시도
 - Displacement Map: 기본 얼굴 mesh(3D) 를 2D로 전개한 UV 좌표계에
서, 얼굴의 어느 표면이 오목해지고 볼록해졌는지 기록한 것.



Detail

Input

FaceScape

Unsup

DECA

EMOCA

FaceVerse

Ours



Detail 복원의 어려움의 원인

- 얼굴의 3D 복원에는 coarse shape 뿐만 아니라, static detail과 dynamic detail이 관여함.
 - coarse shape: 기본적인 얼굴 형태를 identity, expression, albedo 요소들을 선형 조합해 나타낸 것. 3DMM으로 생성.
 - 정적 디테일: 주름, 점, 흉터 등 개인 특징
 - 동적 디테일: 표정에 따라 변하는 특징
- 기존 방법들은 **학습 과정에서 정적 디테일과 동적 디테일이 섞여 부자연스러운 결과 발생.**
- HiFace: 두 디테일을 분리하여 학습하는 방법을 제안

HiFace?

- 3DMM 기반
- 정적(static) 및 동적(dynamic) 디테일을 명확하게(explicitly) 분리하는 고품질 3D face reconstruction 기술
- 단일 이미지에서 사실적이고 애니메이션이 가능한 3D 얼굴을 복원한다.
- Coarse shape과 Detail을 동시에 복원.



HiFace의 결과물

Reconstruction

첫째 줄: input 2D image

둘째 줄: Coarse shape

셋째 줄: Coarse shape + Details



HiFace의 결과물

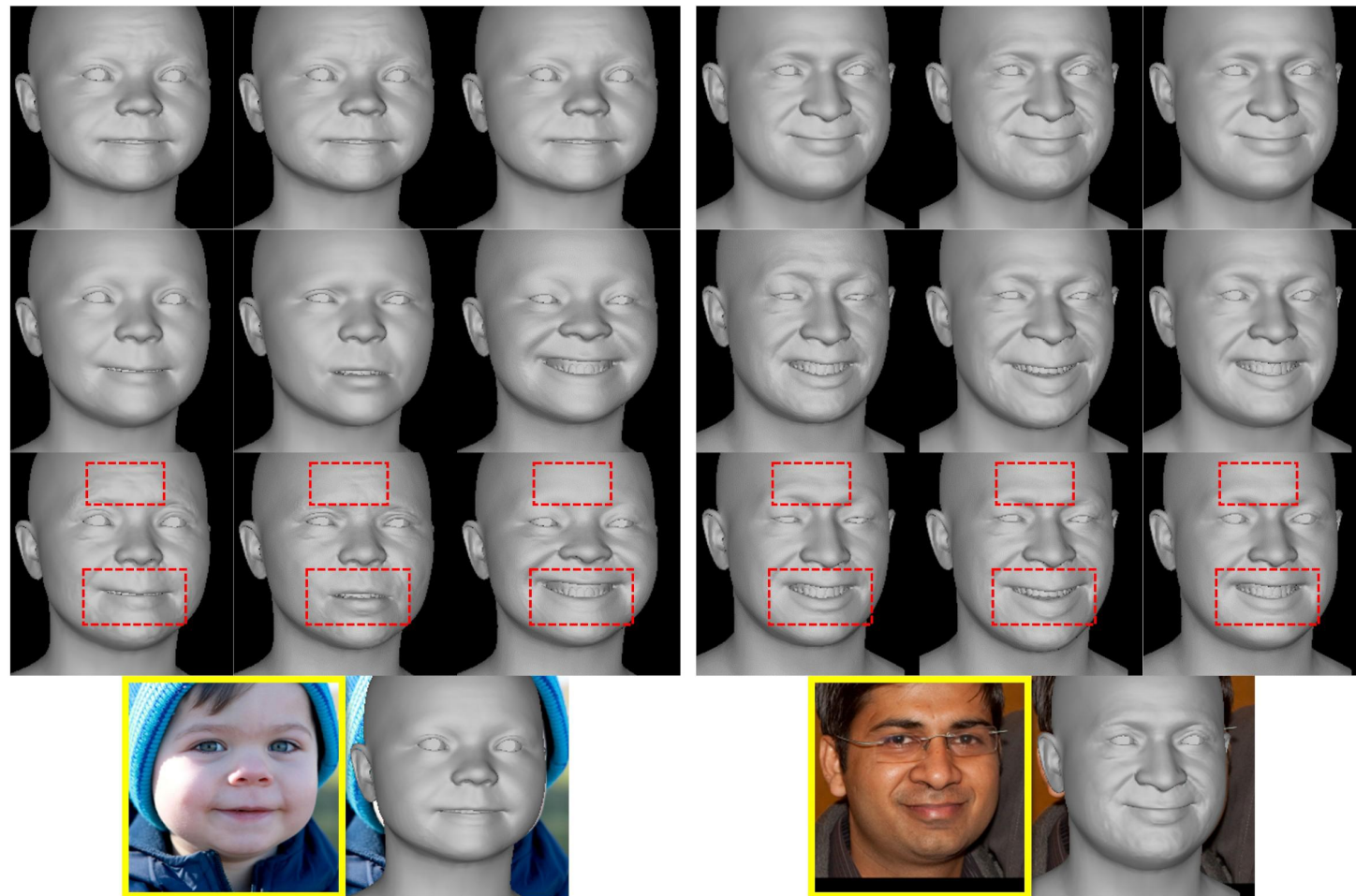
Animated (다양한 표정 적용)

맨 아래 줄: input & coarse shape

1번째 줄: static detail (개인적 특징)

2번째 줄: dynamic detail (표정적 특징)

3번째 줄: static + dynamic



HiFace의 기여

- ✓ 정적 & 동적 디테일을 분리하여 학습하는 HiFace 모델 제안
- ✓ 합성 데이터 & 실제 데이터를 병행 학습 → 일반화 성능 향상
- ✓ 표정 변형 시 '압축 & 이완' 주름을 선형 보간하여 자연스러운 결과 도출
- ✓ 최신 SOTA 대비 15% 이상 성능 개선 (REALY 벤치마크 결과)

3DMMs(3D Morphable Models)란?

- 얼굴 형상을 선형 결합 (Linear Combination)을 이용하여 나타내며, 여러 요소(예: 정체성, 표정, 조명 등)를 개별적인 성분으로 분리하여 조작 가능.
- 다양한 인종, 성별, 나이의 이미지에서, PCA를 통해 각 요소를 조작하는 주성분벡터들을 얻는다.

3DMM의 구성 요소

- 얼굴 형상을 다음과 같은 요소로 분해

1. **Identity (정체성) 요소:** 개인의 고유한 얼굴 구조를 나타내며, 선형 결합 형태로 표현됨.

$$S = \bar{S} + \sum_i \beta_i B_{id,i}$$

- \bar{S} : 평균 얼굴 형상
- $B_{id,i}$: Identity 기반이 되는 주성분 벡터
- β_i : Identity 계수

2. **Expression (표정) 요소:** 얼굴의 표정 변화를 나타내는 성분.

$$S_{exp} = \sum_j \xi_j B_{exp,j}$$

- $B_{exp,j}$: 표정을 결정하는 주성분 벡터
- ξ_j : Expression 계수

3. **Texture (텍스처) 요소:** 얼굴의 피부 색상 및 질감을 나타내는 요소.

$$A = \bar{A} + \sum_k \alpha_k B_{alb,k}$$

3DMM으로 새로운 얼굴 형상 구성하기

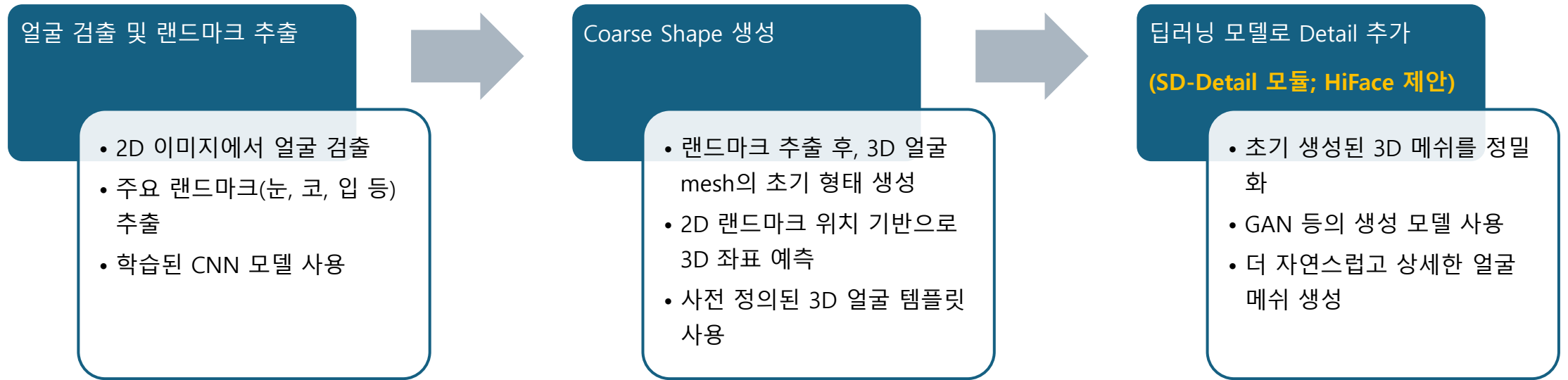
$$\begin{aligned} \mathbf{S} &= \bar{\mathbf{S}} + \beta \mathbf{B}_{\text{id}} + \xi \mathbf{B}_{\text{exp}} \\ \mathbf{A} &= \bar{\mathbf{A}} + \alpha \mathbf{B}_{\text{alb}} \end{aligned},$$

- 새로운 얼굴 S = 평균 얼굴 + (주성분 벡터 * 계수)들로 나타낸다.

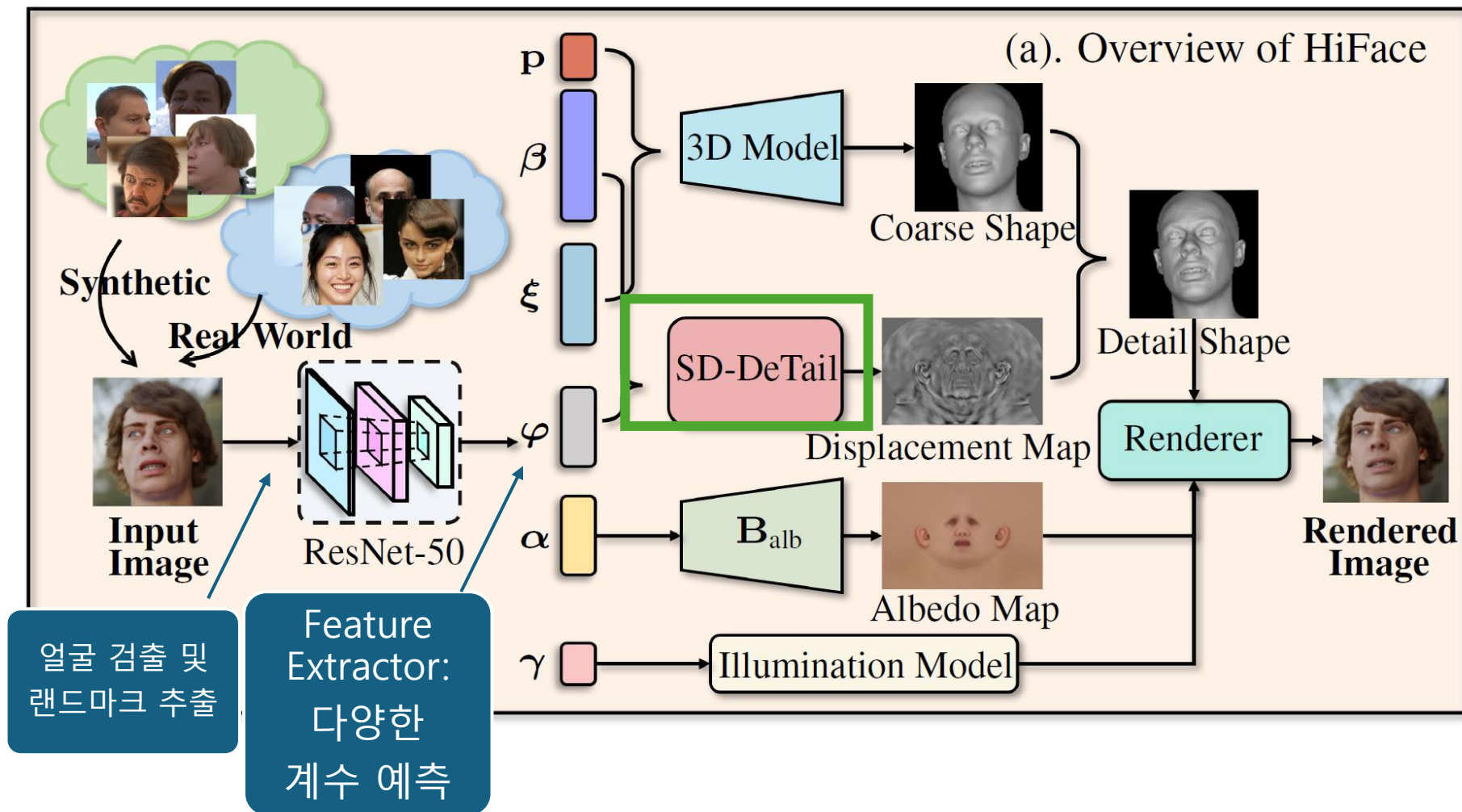
합성 데이터의 필요성

- 실제 이미지에서 주름 라벨링 어려움 → 합성 데이터 활용
- 최근 연구에서는 고품질 합성 데이터 생성 기술 발전
- HiFace는 합성 & 실제 데이터를 결합하여 학습, 도메인 적응력 향상

참고: 3D face reconstruction 과정

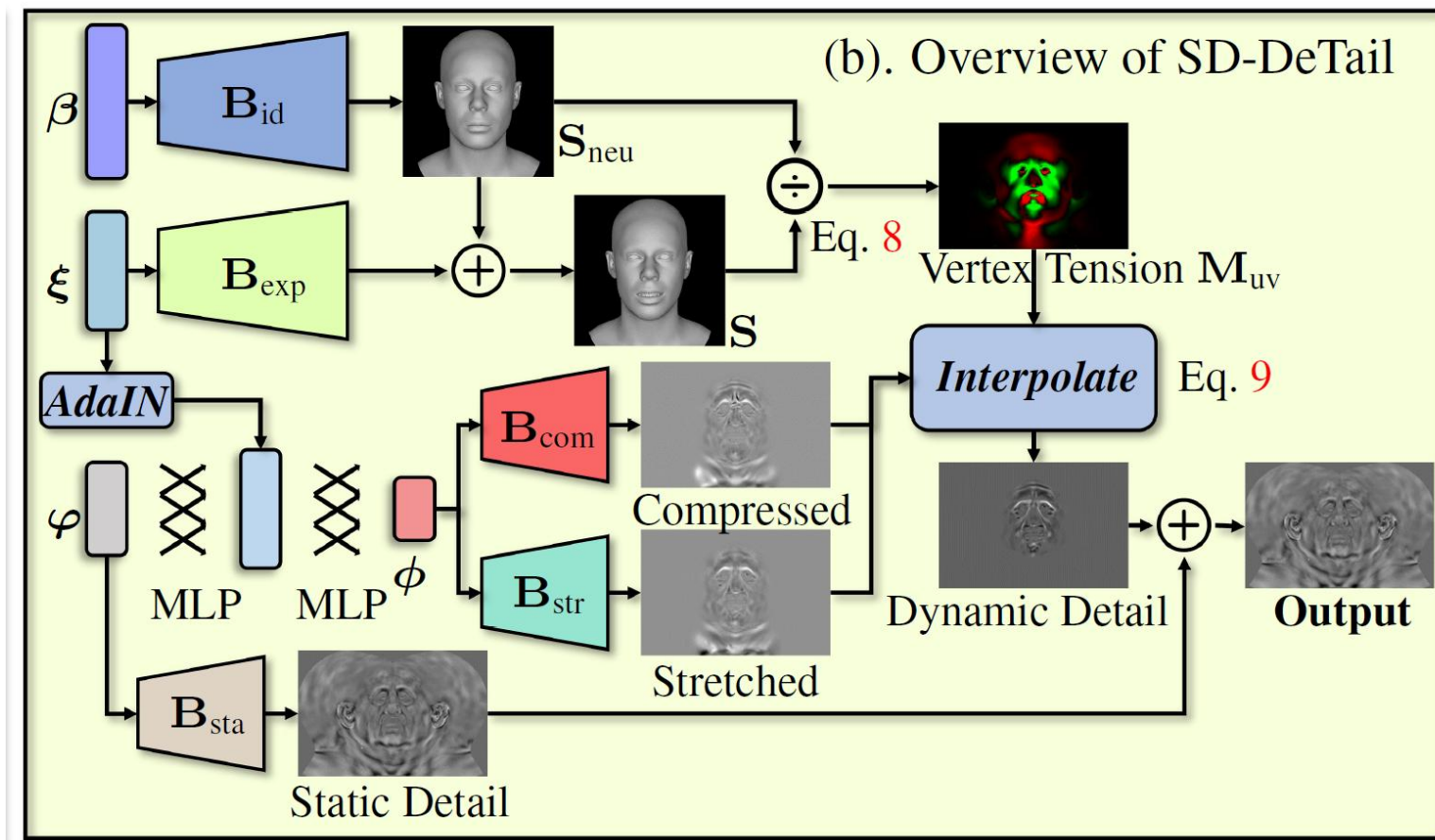


Learning Framework of HiFace



p : pose coef β : identity coef ξ : expression coef φ : static coef α : albedo coef γ : SH coef ϕ : dynamic coef

Pipeline of SD-DeTail

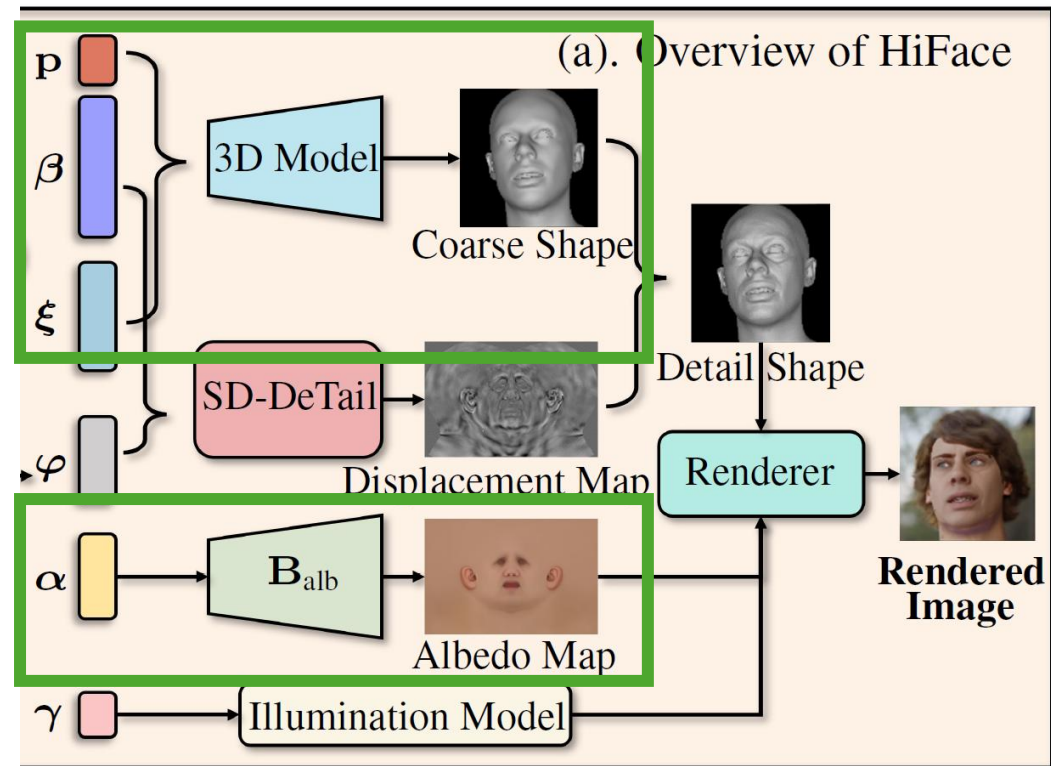


p: pose coef β : identity coef ξ : expression coef φ : static coef α : albedo coef γ : SH coef ϕ : dynamic coef

Coarse Shape, Albedo

- 3DMM 방식을 따라 3D face model 표현
- 얼굴 형상(Shape S)와 원본색상(Albedo A)를 선형 모델로 표현
- 얼굴 형상은 개체성(Identity, β), 표정(Expression, ξ) 계수를 통해 성분 조절

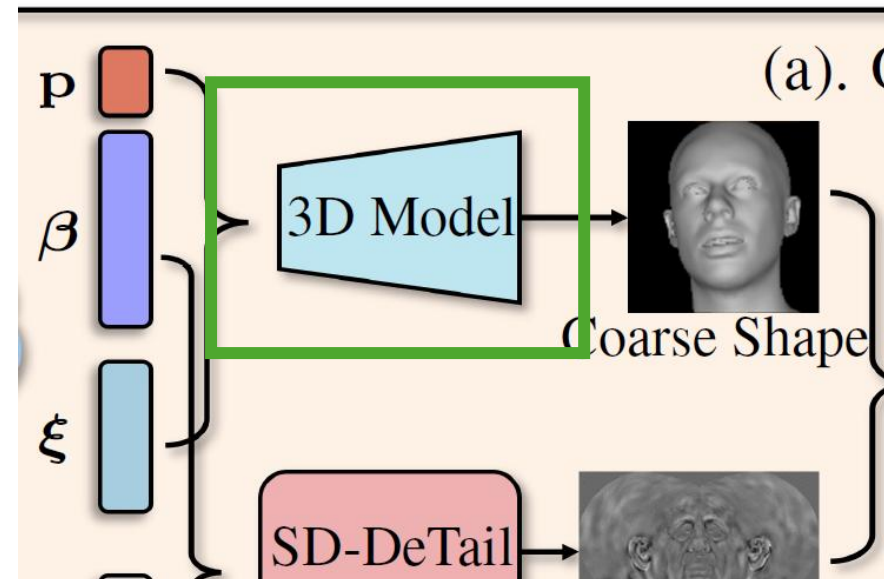
$$\mathbf{S} = \bar{\mathbf{S}} + \beta \mathbf{B}_{\text{id}} + \xi \mathbf{B}_{\text{exp}},$$
$$\mathbf{A} = \bar{\mathbf{A}} + \alpha \mathbf{B}_{\text{alb}}$$



Pose & Camera

- Linear Blend Skinning(LBS)을 사용
- 얼굴 회전(Rotation, θ), 평행이동(Translation, t) 구현
- $p = [\theta \mid t]$: 예측된 포즈 파라미터.
- 얼굴이 다른 방향을 바라보게 할 수 있음.

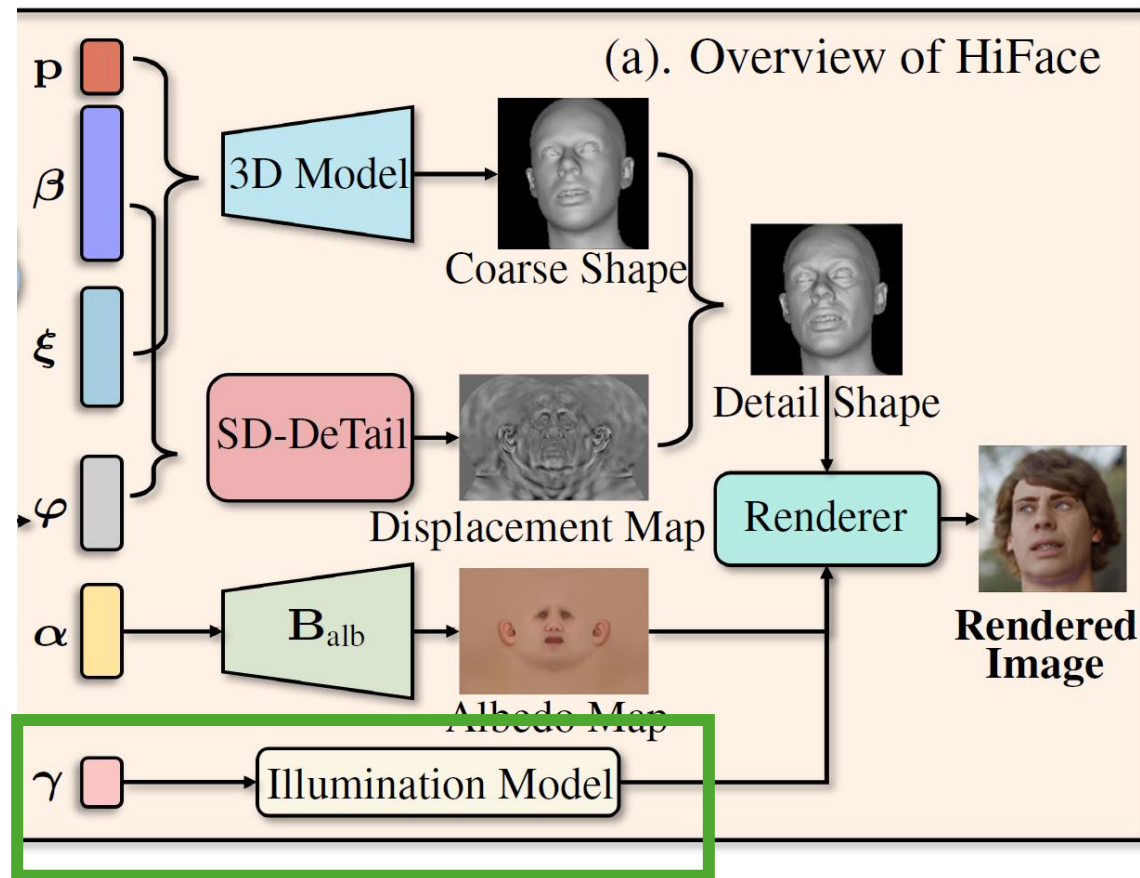
$$S_p = \text{LBS}(S, p, J; W),$$



Illumination Model

- Spherical Harmonics(SH) 모델을 사용해서 사진에 나온 조명을 근사
- 조명에 따른 피부 밝기, 그림자, 하이라이트 등을 계산

$$\mathbf{T} = \mathbf{A} \odot \sum_{k=1}^9 \gamma_k \Psi_k(\mathbf{N}),$$

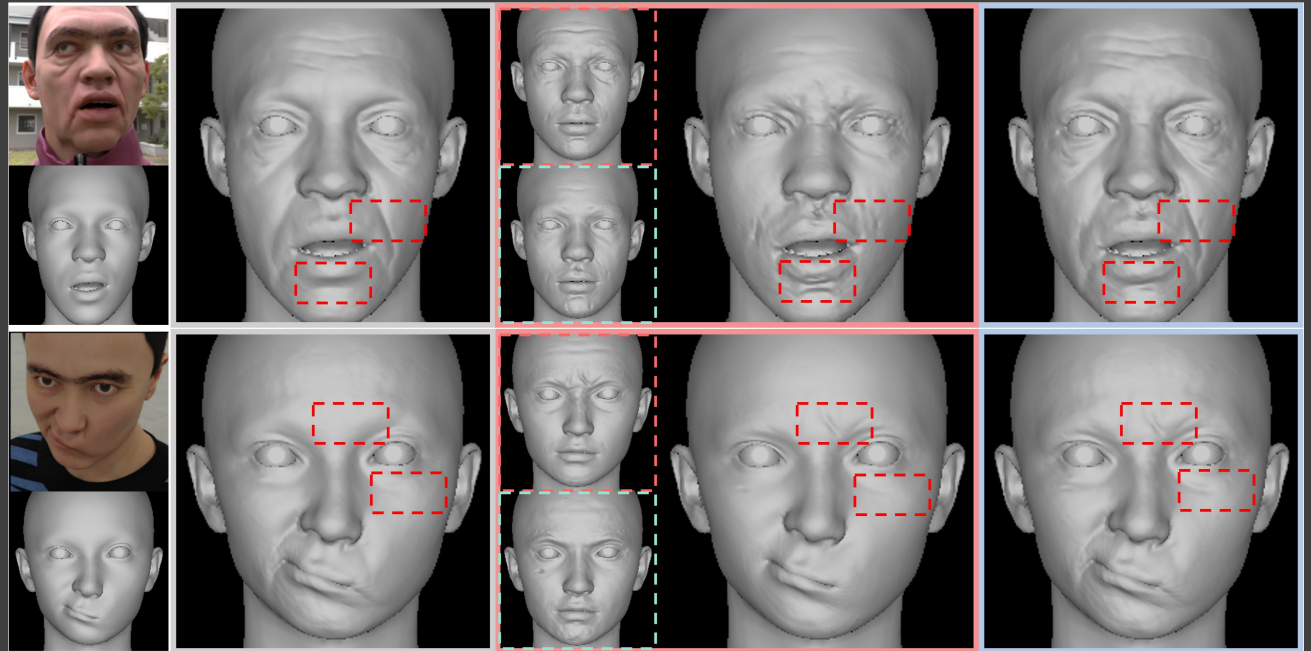


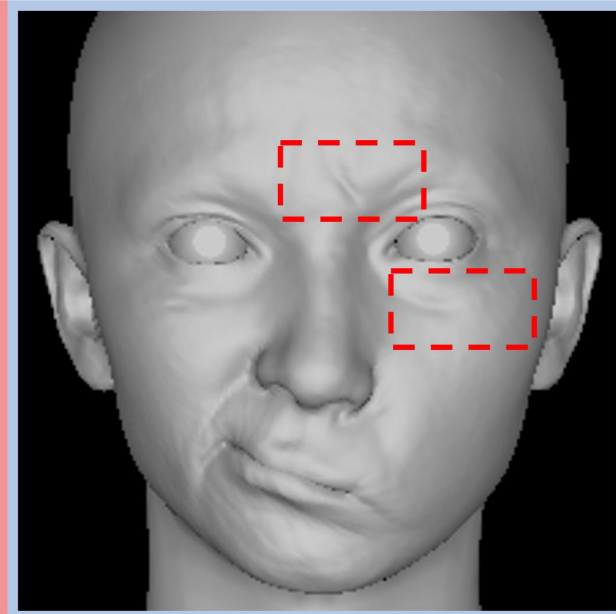
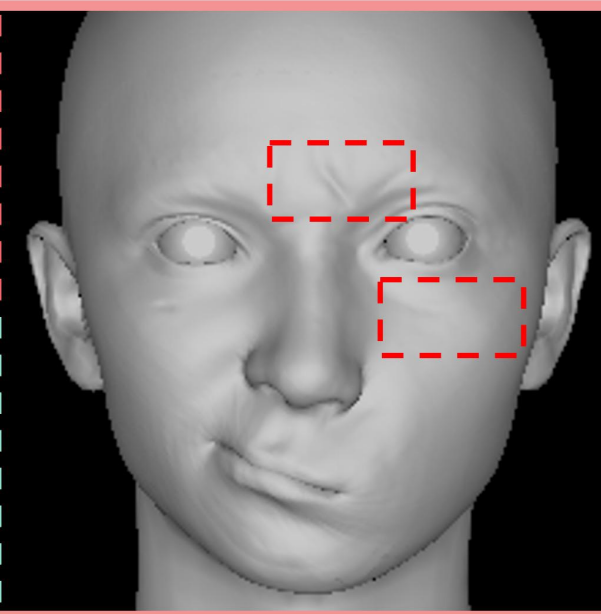
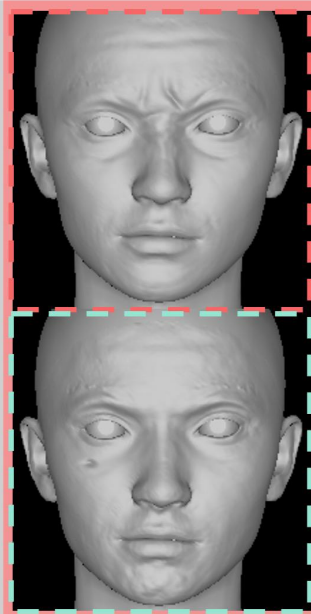
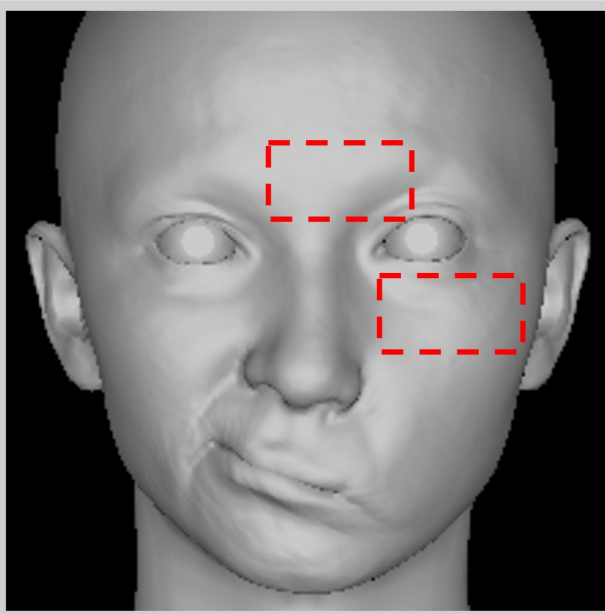
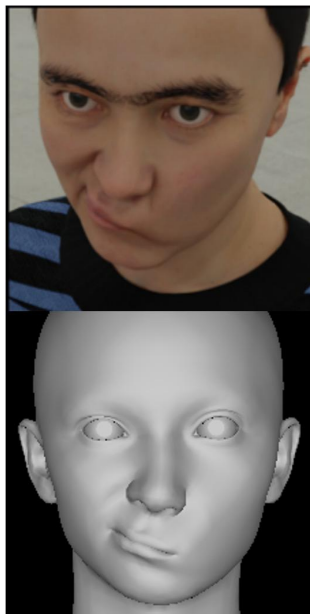
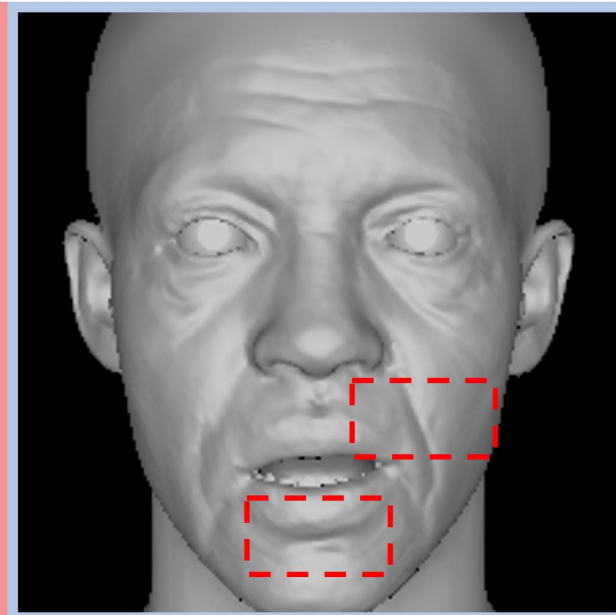
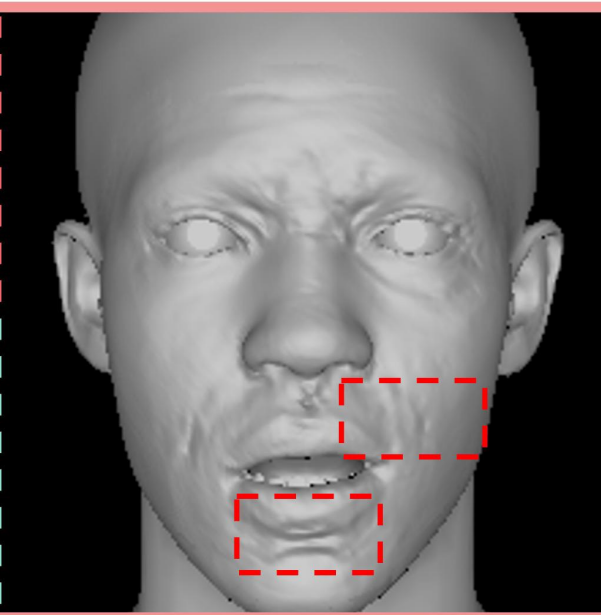
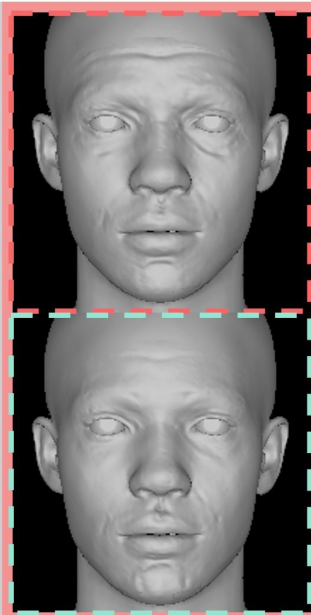
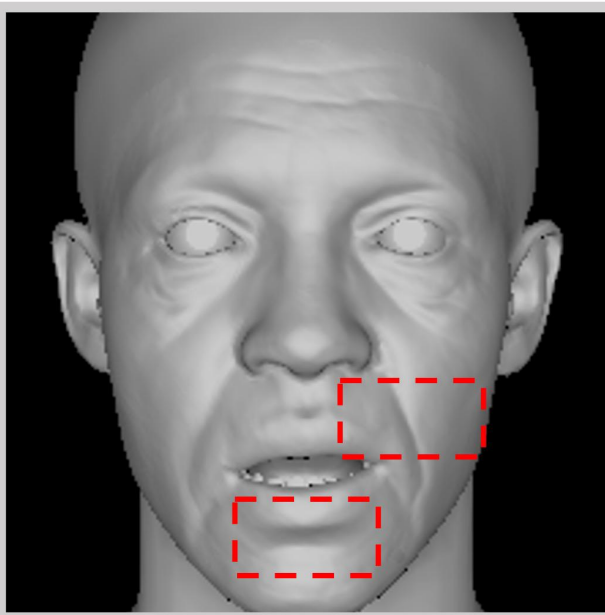
'정적' 과 '동적' 디테일

Static Detail (정적 디테일): 개인
고유의 주름 (나이, 피부결)

Dynamic Detail (동적 디테일):
표정 변화에 따른 주름 (미간, 눈
가 등)

$$\mathbf{D} = \mathbf{D}_{\text{sta}} + \mathbf{D}_{\text{dyn}},$$





Coarse

Static

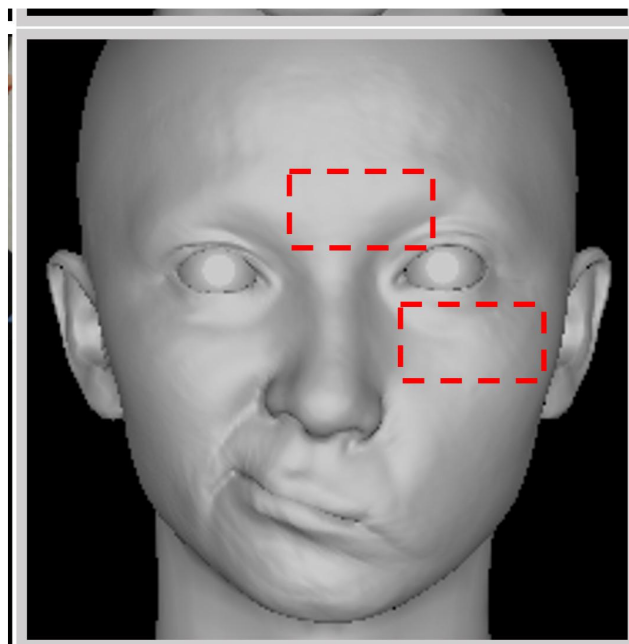
Compressed,
Stretched

Dynamic
(Interpolated)

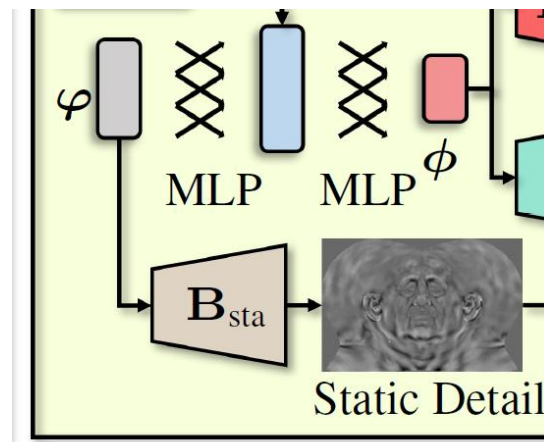
Static + Dynamic

Static Detail

- 정적 디테일 (Static Detail) → 사람마다 고유한 주름 & 피부 결
- PCA 기반 Displacement Basis B_{sta} 를 사용해 정적 주름 생성

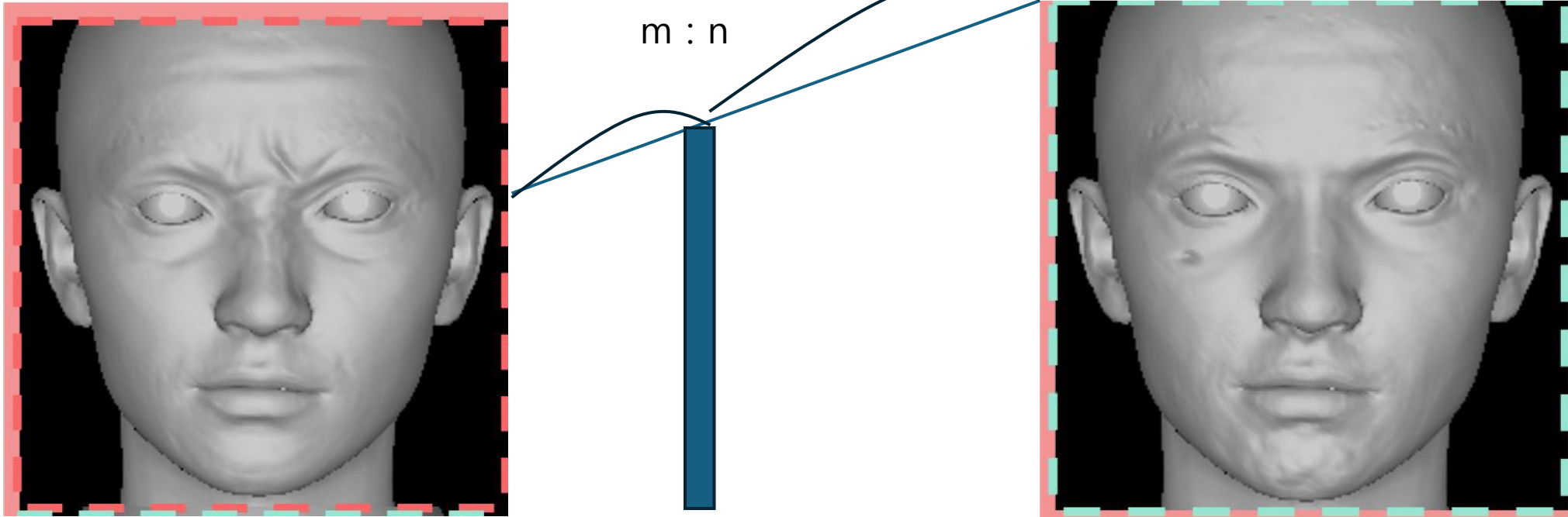


$$D_{sta} = \bar{D}_{sta} + \varphi B_{sta},$$



압축과 이완

- 높은 복잡도로 인해, dynamic detail을 이미지로부터 바로 얻어내는 것은 어려움
- 압축(Compressed) & 이완(Stretched) 상태를 선형 보간 (비율적 배합)



Dynamic Detail



- Dynamic Detail: 표정 변화에 따라 생성되는 주름
- 압축(Compressed) & 이완(Stretched) 상태를 기반으로 선형 보간
- AdaIN을 사용하여 표정 계수(ξ)를 기반으로 동적 디테일 계수(ϕ) 생성
 - AdaIN을 사용하면 한 확률변수의 분포를 다른 확률변수의 분포에 맞게 재조정할 수 있음.

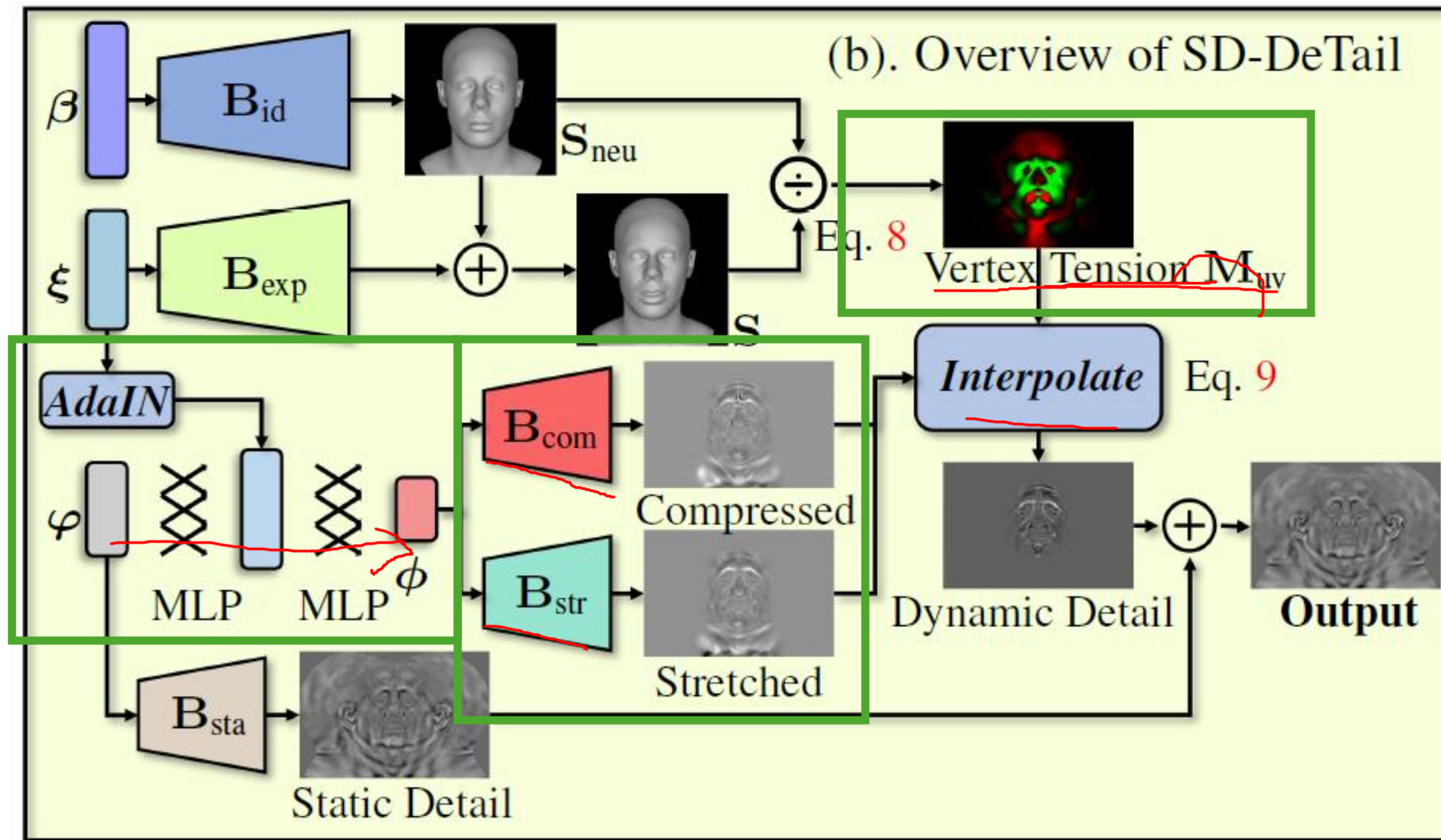
$$\phi = \Phi \left(\sigma(\tilde{\xi}) \left(\frac{\varphi - \mu(\varphi)}{\sigma(\varphi)} \right) + \mu(\tilde{\xi}) \right)$$

$$D_{com} = \bar{D}_{com} + \phi_{com} B_{com}$$

$$D_{str} = \bar{D}_{str} + \phi_{str} B_{str}$$

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

(b). Overview of SD-DeTail



Vertex Tension을 계산해 Map 생성

- 얼굴의 각 부위는 표정 변화에 따라 다르게 움직임
- Vertex Tension을 계산하여 주름 변화량을 조절

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

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

Loss functions (L2)

- **Detail Loss (L_{detail})**

Static & Dynamic Detail의 Ground-Truth와 비교

- **Shape Loss (L_{shp})**

Coarse Shape의 Ground-Truth vertex 정점과 비교

- **Self-Supervised Loss (L_{self})**

- **Knowledge Distillation Loss (L_{kd})**

- **Regularization Loss (L_{reg})**

학습 파라미터들의 크기가 너무 커지거나 작아지는 것을 방지