JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY | 人工智能研究院 Artificial Intelligence Institute

Improving Fairness in Facial Albedo Estimation via Visual-Textual Cues

CVPR 2023

TUE-PM-036 highlight

Xingyu Ren, Jiankang Deng, Chao Ma, Yichao Yan, Xiaokang Yang

SJTU

Preview

- Inferring Albedo from Single Image is ill-posed!

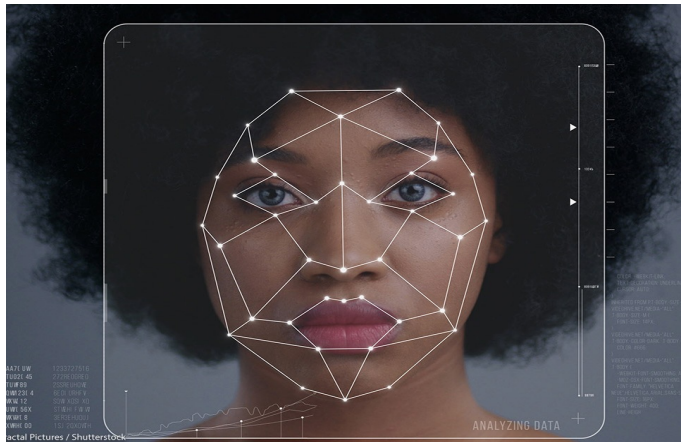
$$\mathbb{R} = \mathbb{L} \cdot \mathbb{A}$$

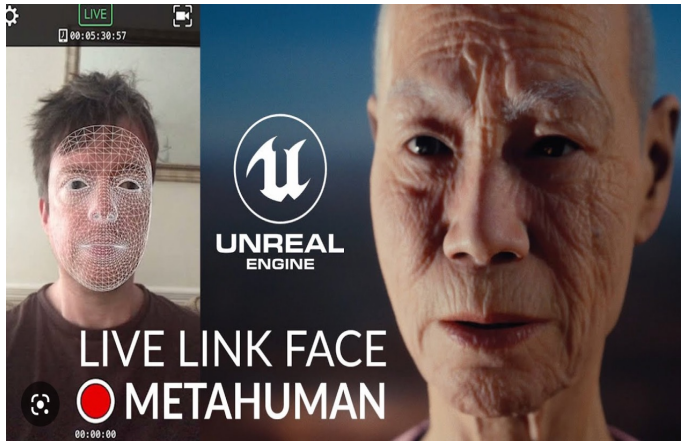
Illumination Estimation → Biased Albedo

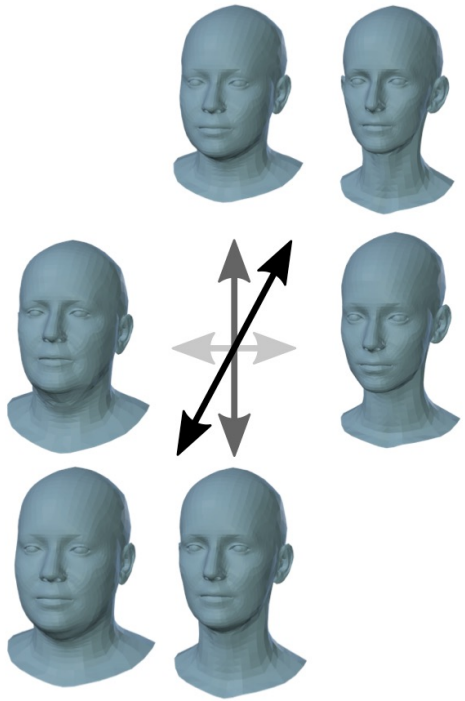
- Insight: Semantic Attributes will control Albedo!

$\mathbb{R} = \mathbb{L} \cdot \mathbb{A}$ → Facial Attribute Prior → Race: Asian, Age: Young, Skin: Brown, ... → Albedo

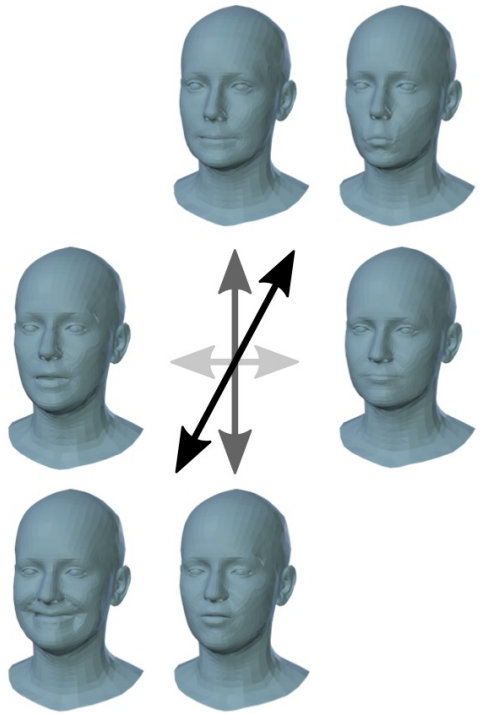
Background: 3D Face Reconstruction







shape



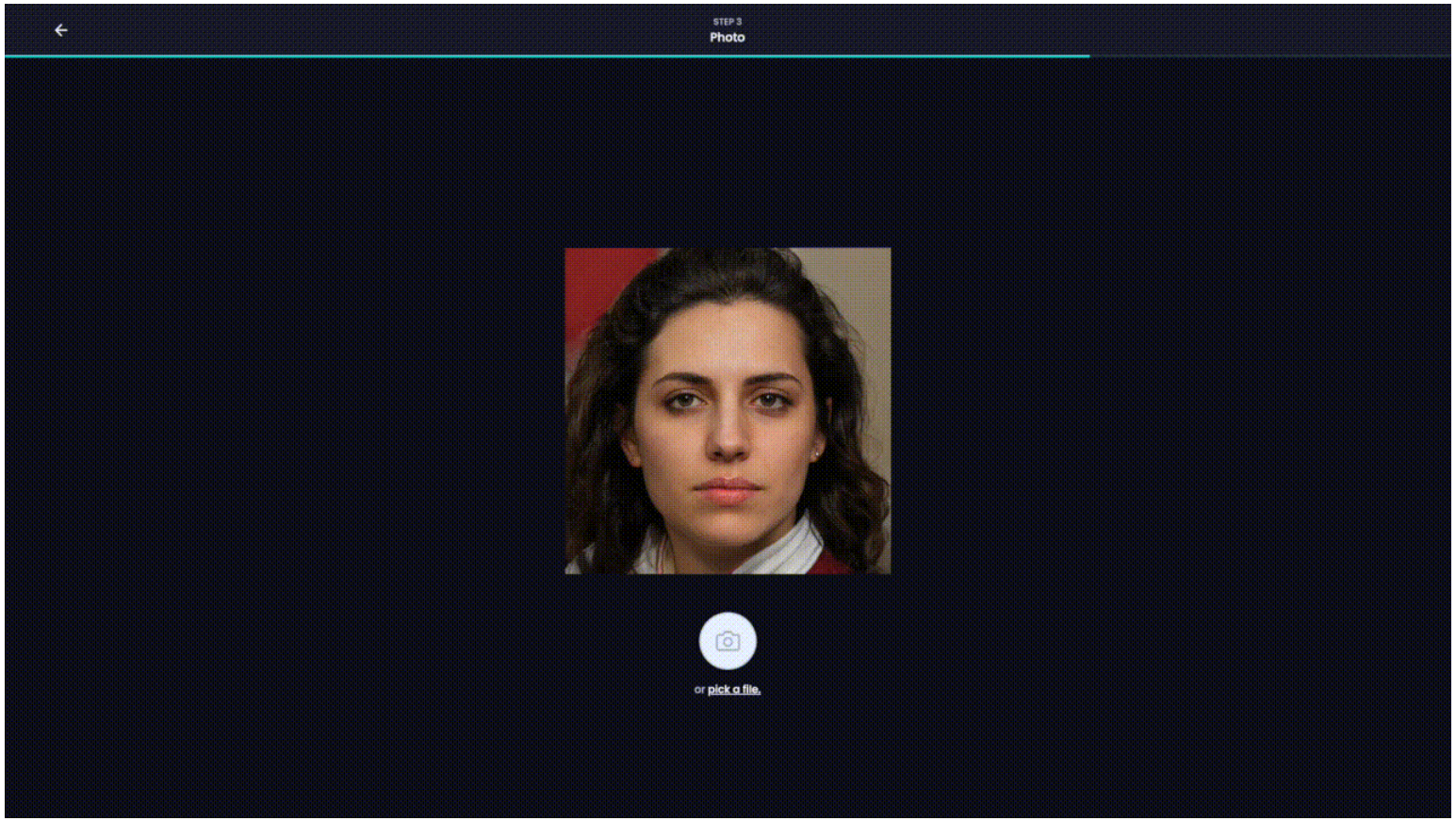
expression

Applications:

- Face Recognition
- Manipulation
- Facial animation
- …

Tianye Li, Timo Bolkart, Michael J. Black, Hao Li, and Javier Romero ,Learning a model of facial shape and expression from 4D scans. In ToG, 2017.

Background: 3D Face Reconstruction



Applications:

- Avatar Creation
- AR / VR
- …

Challenge

- Inferring albedo from pixels is an ill-posed problem

$$\mathcal{R} = \mathcal{A} \odot \mathcal{S}$$

? ?

Challenge

- Inferring albedo from pixels is an ill-posed problem

$$\mathcal{R} = \mathcal{A} \odot \mathcal{S}$$

Albedo Model

Lighting Constrain

Challenge

- Inferring albedo from pixels is an ill-posed problem

$$\mathcal{R} = \mathcal{A} \odot \mathcal{S}$$

Albedo Model

Lighting Constrain

Biased! *Biased!*

Challenge

- Inferring albedo from pixels is an ill-posed problem

$$\mathcal{R} = \mathcal{A} \odot \mathcal{S}$$

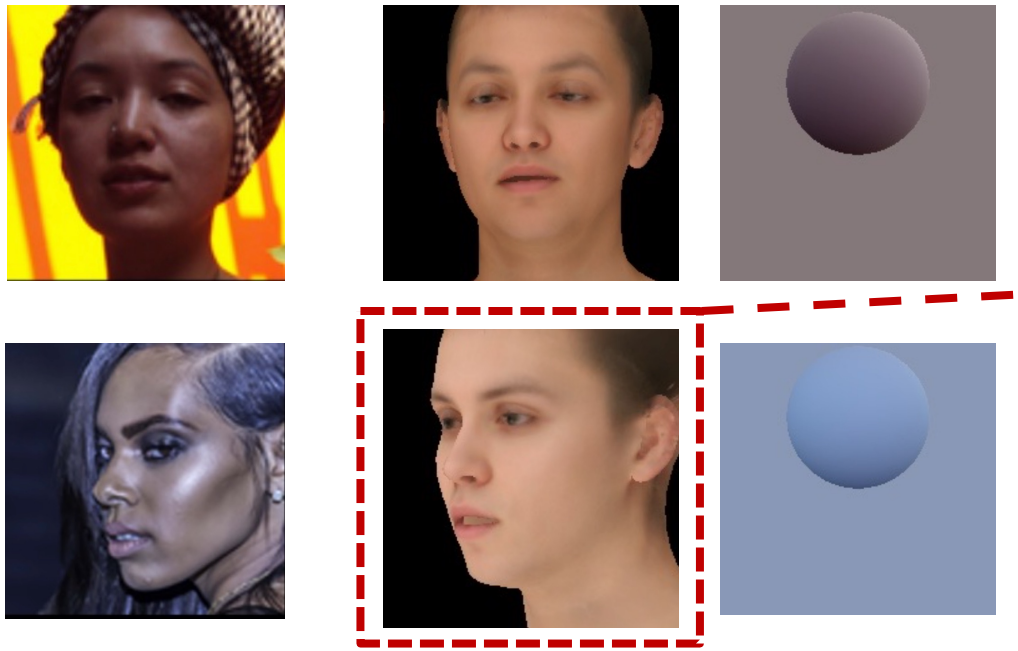
Unbiased Albedo Model

Scene Lighting Constrain

Challenge

- Inferring albedo from pixels is an ill-posed problem

$$\mathcal{R} = \mathcal{A} \odot \mathcal{S}$$



Motivation

- How do people disentangle light from a single image?
  - Prior Knowledge
    - Caucasian → White
    - …
    - African → Dark
  - Facial Attribute → Albedo

Motivation

- Inferring albedo from pixels is an ill-posed problem

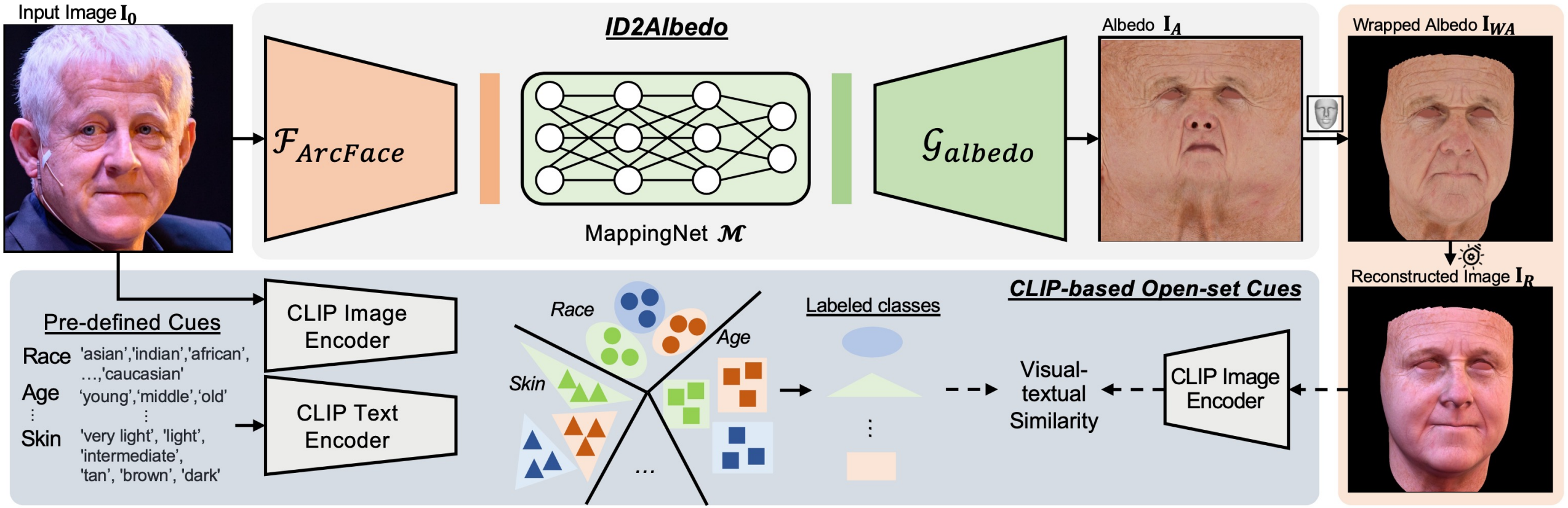
$$\mathcal{R} = \mathcal{A} \odot \mathcal{S}$$

Facial Attribute

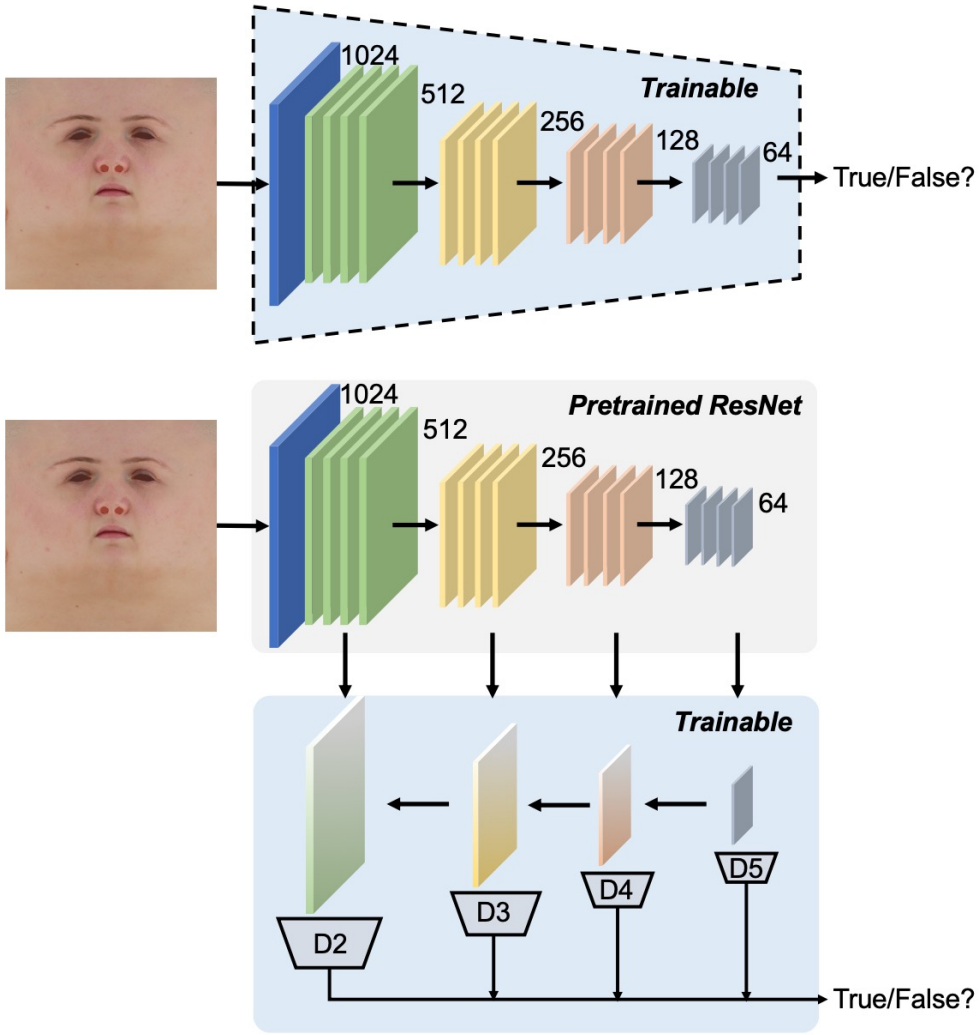
Unbiased Albedo Generator

No Constrain

Method: ID2Albedo

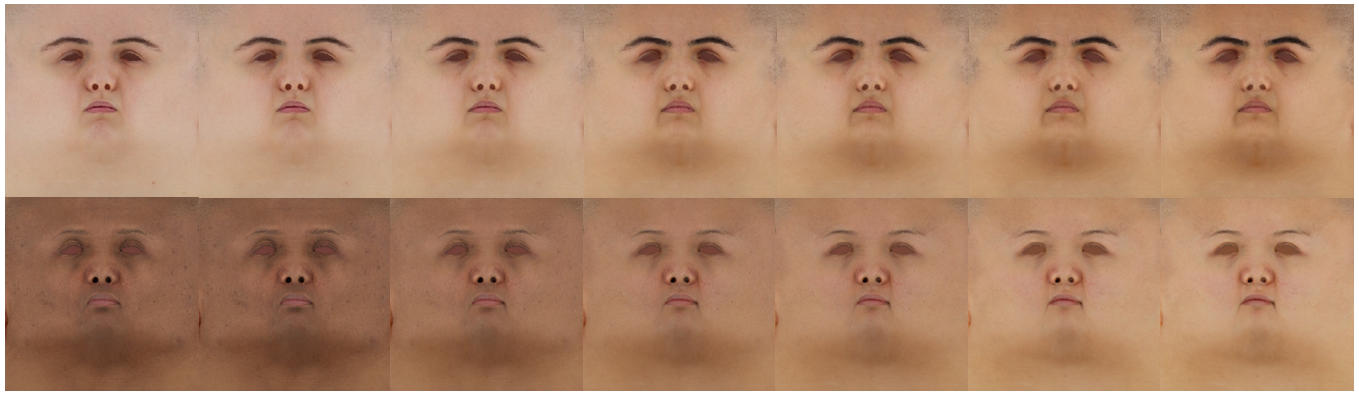


Method: Albedo Generator



Random Seeds

Interpolation

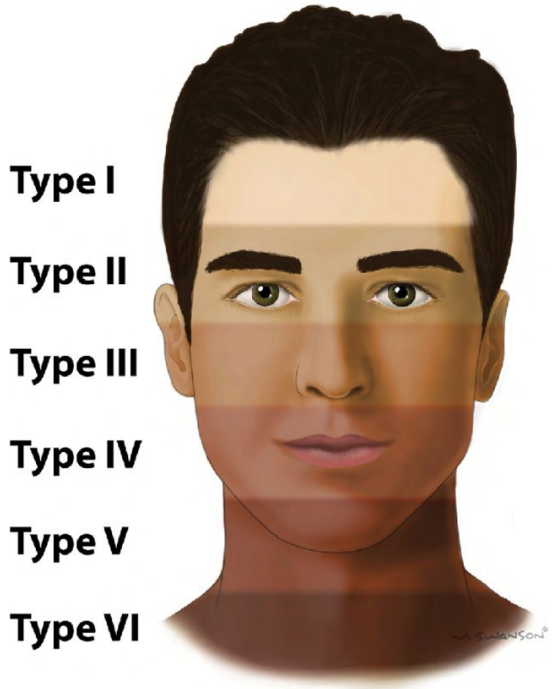


Evaluation

- ITA Score

$$\mathrm{ITA}(L^*, b^*) = \frac{180}{\pi} \times \arctan\left(\frac{L^* - 50}{b^*}\right),$$

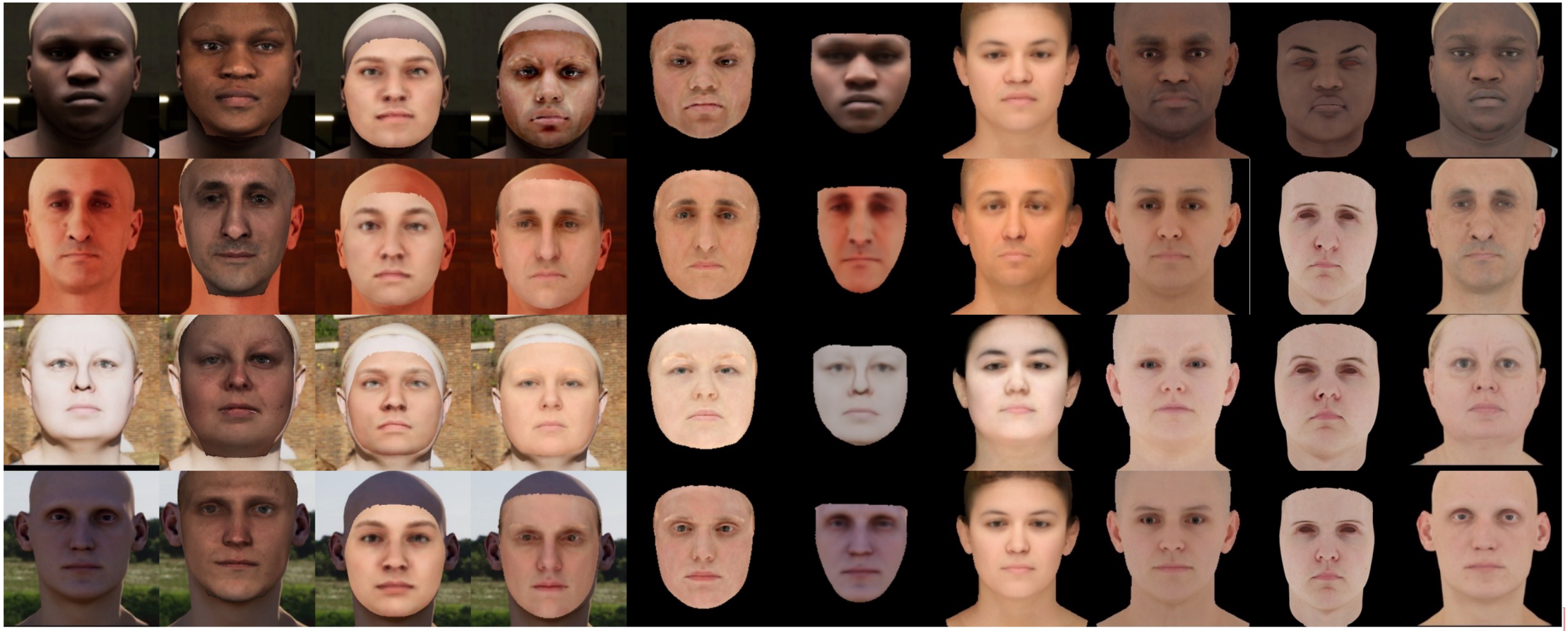
SKIN TYPE	SKIN COLOR	REACTION TO SUN: SUNBURN	REACTION TO SUN: TANNING
I	Light, pale white	Always burns	Never Tans
II	White, fair	Usually burns	Tans with difficulty
III	Medium, white to olive	Sometimes mild burns	Gradually tans to olive
IV	Beige olive, moderate brown	Rarely burns	Easy tan to moderate brown
V	Brown, dark brown	Very rarely burns	Tans very easily
VI	Very dark brown to black	Never burns	Always tans



Results: FAIR Benchmark

Method	Avg. ITA ↓	Bias ↓	Score ↓	MAE ↓	ITA per skin type ↓					
					I	II	III	IV	V	VI
Deep3D [10]	22.57	24.44	47.02	27.98	**8.92**	9.08	8.15	10.90	28.48	69.90
GANFIT [19]	62.29	31.81	94.11	63.31	94.80	87.83	76.25	65.05	38.24	11.59
MGCNet [46]	21.41	17.58	38.99	25.17	19.98	12.76	8.53	9.21	22.66	55.34
DECA [15]	28.74	29.24	57.98	38.17	9.34	11.66	11.58	16.69	39.10	84.06
INORig [2]	27.68	28.18	55.86	33.20	23.25	11.88	**4.86**	9.75	35.78	80.54
CEST [55]	35.18	12.14	47.32	29.92	50.98	38.77	29.22	23.62	21.92	46.57
TRUST [14] (BFM)	16.19	15.33	31.52	21.82	12.44	**6.48**	5.69	9.47	16.67	46.37
TRUST [14] (AlbedoMM)	17.72	15.28	33.00	19.48	15.50	10.48	8.42	**7.86**	**15.96**	48.11
TRUST [14] (BalancedAlb)	13.87	**2.79**	**16.67**	**18.41**	11.90	11.87	11.20	13.92	16.15	18.21
Ours (ID2Albedo)	**12.07**	4.91	16.98	23.33	18.30	9.13	5.83	9.46	19.09	**10.59**

Results: FAIR Benchmark



Results: In-the-wild Images

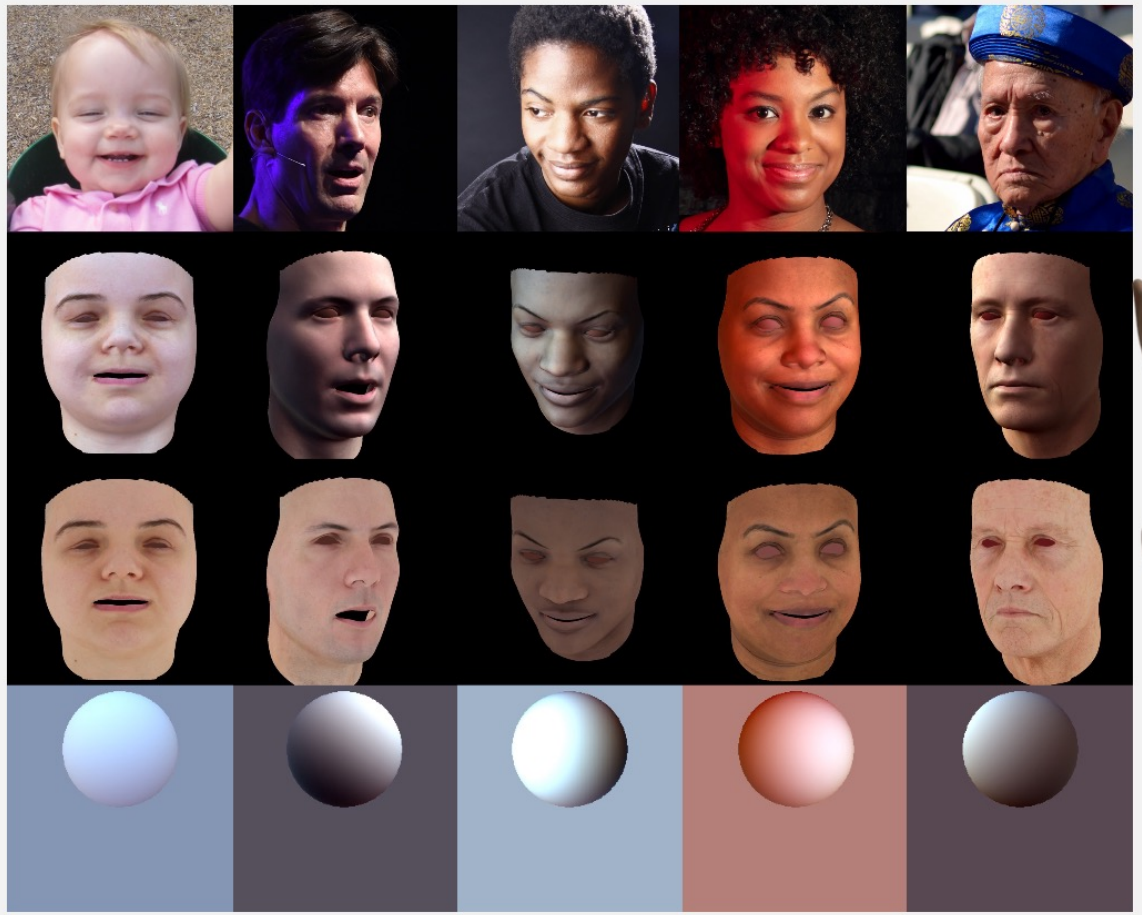
Methods	M-SSIM↑	LPIPS↓	FID↓	ID↑
Deep3D [10]	0.73	0.1933	74.41	0.712
DECA [15]	0.61	0.2089	98.13	0.585
TRUST [14]	0.64	0.2112	97.37	0.603
Ours	**0.87**	**0.1549**	**45.56**	**0.867**

Ablations

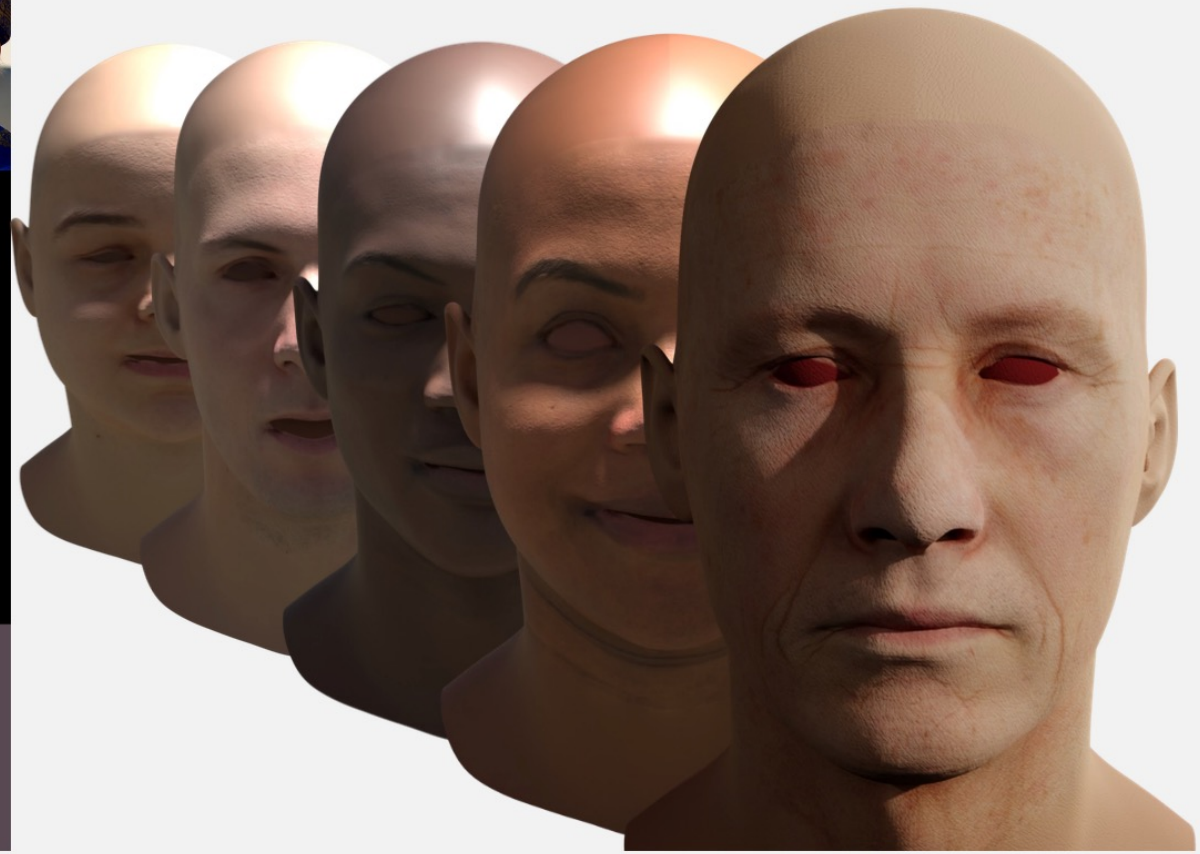
Albedo Encoder	Avg. ITA ↓	Bias ↓	Score ↓
ResNet-100 [23] (Scratch)	58.46	32.59	91.05
ResNet-100 [23] (ImageNet)	31.63	15.48	47.11
ArcFace [9] (fully trainable)	41.63	19.81	61.44
ArcFace [9] (L2 + L3 + L4)	28.75	11.87	40.62
ArcFace [9] (L3 + L4)	19.52	9.46	28.98
ArcFace [9] (L4)	14.58	6.79	21.37
ArcFace [9] (Frozen)(Ours)	13.46	5.86	19.32

Configs	Avg. ITA ↓	Bias ↓	Score ↓
w/o any cues	25.66	23.51	49.17
Manual labeled races	18.13	10.46	28.59
CLIP [40] cues (only races)	16.21	7.44	23.65
CLIP cues all (ours)	13.46	5.86	19.32

Improving **Fairness** in Facial Albedo Estimation via **Visual-Textual Cues**



Unbiased Albedo Reconstruction



Realistic Head Rendering