Towards Racially Unbiased Skin Tone Estimation via Scene Disambiguation

Haiwen Feng, Timo Bolkart, Joachim Tesch, Michael J. Black, and
Victoria Abrevaya

Max Planck Institute for Intelligent Systems, Tübingen, Germany
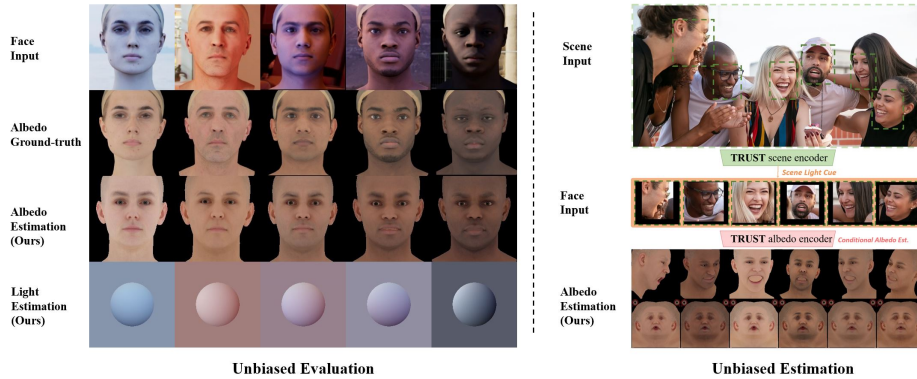{hfeng,tbolkart,jtesch,black,vabrevaya}@tuebingen.mpg.de

Fig. 1: **FAIR** (left) is a new dataset for unbiased albedo estimation that uses high-quality facial scans, varied lighting, and a new evaluation metric to benchmark current approaches in terms of accuracy and fairness. We also propose a new network, **TRUST** (right), for facial albedo estimation that reduces bias by addressing the light/albedo ambiguity using scene disambiguation cues.

Abstract. Virtual facial avatars will play an increasingly important role in immersive communication, games and the metaverse, and it is therefore critical that they be inclusive. This requires accurate recovery of the albedo, regardless of age, sex, or ethnicity. While significant progress has been made on estimating 3D facial geometry, appearance estimation has received less attention. The task is fundamentally ambiguous because the observed color is a function of albedo and lighting, both of which are unknown. We find that current methods are biased towards light skin tones due to (1) strongly biased priors that prefer lighter pigmentation and (2) algorithmic solutions that disregard the light/albedo ambiguity. To address this, we propose a new evaluation dataset (FAIR) and an algorithm (TRUST) to improve albedo estimation and, hence, fairness. Specifically, we create the first facial albedo evaluation benchmark where subjects are balanced in terms of skin color, and measure accuracy using the Individual Typology Angle (ITA) metric. We then address the light/albedo ambiguity by building on a key observation: the image of the full scene –as opposed to a cropped image of the face– contains important information about lighting that can be used for disambiguation. TRUST regresses facial albedo by conditioning on both the face region

and a global illumination signal obtained from the scene image. Our experimental results show significant improvement compared to state-of-the-art methods on albedo estimation, both in terms of accuracy and fairness. The evaluation benchmark and code are available for research purposes at https://trust.is.tue.mpg.de.

1   Introduction

For critical systems such as face recognition, automated decision making or medical diagnosis, the development of machine-learning-based methods has been followed by questions about how to make their decisions fair to all sectors of the population [1, 8, 32, 46]. For example, Buolamwini and Gebru [8] identify race and gender biases in face analysis methods, which disproportionately misclassify dark-skinned females; Rajkomar et al. [50] note an influence of historical biases in machine learning methods for health care; and Kim et al. [30] point out how rendering algorithms in computer graphics are designed mostly for light-skinned people.

However, no such analysis exists for the task of 3D facial avatar creation. Here the problem involves estimating the 3D geometry and albedo from one or more images, where albedo is defined as a texture map corresponding to the diffuse reflection of the face (including skin tone, lips, eye, etc). With growing interest in on-line virtual communication, gaming, and the metaverse, the role of facial avatars in our lives will likely increase. It is then critical that such technology is equally accessible to all and that every person can be represented faithfully, independent of gender, age, or skin color.

In this paper we address, for the first time, the problem of fairness in 3D avatar creation from images in-the-wild, and show that current methods are biased towards estimating albedo with a light skin color. While skin-tone bias has been extensively studied in the field of face recognition [18, 62], here we examine a different problem. Specifically, we consider bias within methods that *regress* facial albedo (of which skin tone is only one aspect) and discuss the particular challenges that this involves. We analyze three main sources for this bias. First, existing statistical albedo models are trained from unbalanced datasets, producing strongly biased priors. Second, existing albedo regression methods are unable to factor lighting from albedo. When combined with a biased prior, they simply infer dim illumination to compensate for dark skin. Finally, there is, to date, no standard evaluation protocol that quantitatively reveals bias in albedo estimation. This has led to the field largely ignoring issues of fairness in albedo estimation.

This work makes several key contributions to advance fairness in facial albedo estimation. First, we create a new dataset of realistic synthetic images of faces in varying lighting conditions. We use this dataset, together with specifically designed metrics, as a benchmark for evaluating accuracy and fairness of facial albedo reconstruction (Fig. 1, left); we call the benchmark FAIR for Facial Albedo Independent of Race. With this we are able, for the first time, to quan-

tify the bias of existing methods. We find that all existing methods are biased to light skin tones, and analyze the technical reasons behind this.

Finally, we introduce a new neural network model called TRUST (Towards Racially Unbiased Skin Tone estimation) that produces state-of-the-art results in terms of accuracy, as well as less biased albedo estimates (Fig. 1, right). The key to TRUST is the design of a novel architecture and losses that explicitly address the light/albedo ambiguity. Specifically, we propose a two-branch network to leverage scene cues for disambiguation. The first branch recovers the illumination signal in the form of Spherical Harmonics (SH) [51], both from the entire scene image and the individual facial crops. The second branch estimates diffuse albedo from the individual crops, conditioned on the intensity of the SH vector. This provides cues about the global illumination and helps the albedo network to better decide on the overall skin tone. The network is further coupled with a new statistical albedo model trained using a balanced sample of subjects. The proposed approach achieves 56% improvement over the best alternative method when evaluated on the new FAIR benchmark, highlighting the importance of tackling fairness both from the dataset and the algorithmic point of view.

In summary, our contributions are: (1) We identify, analyze and quantify the problem of biased facial albedo estimation. (2) We propose a new synthetic benchmark, as well as new metrics that measure performance in terms of skin tone and diversity. (3) We propose a solution for albedo estimation that significantly improves fairness by explicitly addressing the light/albedo ambiguity problem through scene disambiguation cues. (4) The benchmark, code, and model are publicly available for research purposes.

2   Related work

3D face and appearance estimation. Face albedo estimation is typically approached as an inverse rendering problem, in which face geometry, reflectance and illumination parameters are estimated from a single image. Methods can be roughly categorized into optimization-based [2, 6, 7, 56, 66] and learning-based [11, 16, 22, 26, 29, 58, 65]; see Egger et al. [20] for a review. The majority of these estimate parameters of a 3D morphable model (3DMM) [7, 27, 48] trained from two-hundred white European subjects. This lack of diversity results in a strong bias towards light-skinned appearance when jointly estimating reflectance and lighting.

To obtain more varied appearance, Smith et al. [61] build an albedo model from light-stage data, and Gecer et al. [25] build a neural generative model from $10K$ texture maps. Several other methods learn flexible appearance models directly from images, while jointly learning to reconstruct 3D shape and appearance [10, 40, 54, 57, 60, 63, 64, 67, 68, 70]. Another line of work synthesizes high-frequency appearance details from images [37, 55, 72]. All of these methods treat the face in isolation by only analyzing a tightly cropped face region. In contrast, our approach exploits cues from the entire scene that help decouple albedo from lighting in unconstrained in-the-wild images.

A related line of work directly estimates *skin tone* from images – i.e. a single vector value (e.g. RGB) that summarizes the color of the face. This can be used as a proxy for color correction [5, 12, 43] and for cosmetology purposes [33, 34]. Here we focus on estimating the full albedo map, but evaluate the accuracy of both albedo and skin tone, with the goal of identifying potential fairness issues.
Disambiguating appearance and lighting. Given the RGB values of a pixel, the proportion of color explained by light versus intrinsic appearance cannot be recovered without further knowledge [51]. While the use of a statistical appearance prior limits the variability of the intrinsic face color, it does not fully resolve the ambiguity. To address this, Aldrian et al. [2] regularize the light by imposing a "gray world" constraint that encourages the environment to be, on average, gray. Hu et al. [28] regularize the albedo by imposing symmetry, thus preventing illumination variation from strong point lights being explained by the appearance. Egger et al. [19] learn a statistical prior of in-the-wild SH coefficients and use this to constrain the light in an inverse rendering application. These methods impose priors or regularizers over the estimated light or appearance, using heuristics to complement the statistical face model. Instead of using the tightly cropped face in isolation, we estimate the environment light from the scene and use this in learning to disambiguate light and albedo.
Racial face bias. Several works identify demographic bias (sex, age, race) for algorithms in the context of biometrics (e.g., face verification, age estimation, race classification, emotion classification, etc.); see Drozdowski et al. [18] for a review. Dedicated benchmark datasets were proposed in [53, 69] to support the study of racial biases in face recognition. Wang et al. [69] show bias in four commercial face recognition systems, and propose an algorithmic solution to mitigate this. Buolamwini and Gebru [8] report bias in gender classification systems correlated to skin tone and sex, where darker-skinned females are misclassified most frequently. Kim et al. [30] describe racial bias in existing quantitative metrics used to group skin colors and hair types, and propose an evaluation scheme based on complex numbers to alleviate this bias. No other previous work has quantitatively assessed bias in facial albedo estimation methods, nor is there any existing dataset for this task. Hence, FAIR provides an important new resource to the community.

3   Dataset and Metrics for Quantifying Skin Tone Bias

To develop unbiased algorithms it is important to first identify potential problems. Since it is difficult to acquire ground-truth appearance, there is to date no albedo evaluation dataset available, much less one that covers a wide range of ethnicities and scenes. To address this, we describe a new dataset for evaluating single-image albedo reconstruction methods, constructed using high-quality facial scans. We additionally propose a set of metrics that measure fidelity of the estimated albedo, as well as accuracy and diversity of the overall skin tone.

Balancing skin type. There are many types of bias but here we focus on skin color, which means we need an evaluation dataset that is balanced across this
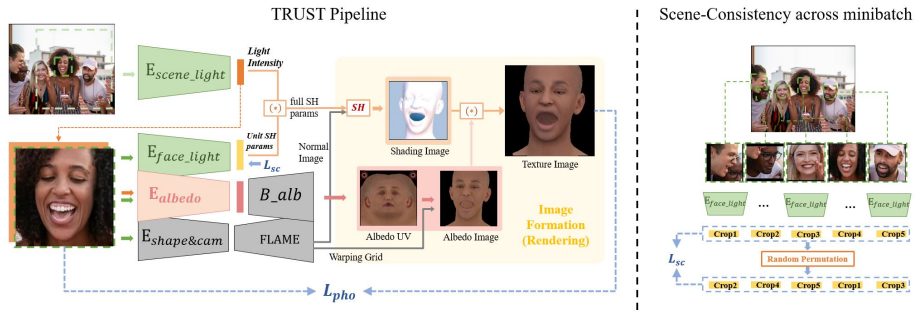
Fig. 2: We address biased albedo estimates by tackling the light/albedo ambiguity using scene disambiguation cues. Given an image of a scene with faces, we first obtain the 3-channel global light intensity factor using the E_{scene_light} encoder. We next extract facial crops, and condition the albedo encoder E_{albedo} by concatenating the intensity as an extra channel to the input. Finally, a crop-based light encoder E_{face_light} estimates normalized SH parameters, which are combined with the color intensity to obtain the final light estimate for the crop. The network is trained using a scene consistency loss, which requires the light from all faces in an image to be the same (i.e. permutation invariant).

factor. There are several ways to quantify the skin color of a person. While self-reported ethnicity labels are commonly used in the literature (e.g. [38, 69]), ethnicity is not well defined and there can be a large variety of skin types within ethnic groups. In dermatology, an extensively used system is the Fitzpatrick scale [23], employed also in computer vision research for balancing datasets [8, 17, 36, 71]. However, the scale is based on subjective assessments of skin sensitivity and tanning ability, and it has been shown to work sub-optimally for certain populations [49, 73]. Instead, we employ the Individual Typology Angle (ITA) [9], also recently considered in [45]. The metric is based on the L* (lightness) and B* (yellow/blue) components of the CIE L*a*b* color space, and is computed as

$$\text{ITA}(L^*, b^*) = \frac{\arctan(\frac{L^*-50}{b^*}) \times 180}{\pi}. \tag{1}$$

We consider the ITA score of an albedo map to be the average of all pixel-wise ITA values within a pre-computed skin region area (defined on the UV map). ITA can be used to classify the skin according to six categories, ranging from very light (category I) to dark (category VI) [9, 15]. It has the advantage of being an objective metric, easily computable from images, and significantly correlated with skin pigmentation [14]. More details can be found in Sup. Mat.

Dataset construction. The dataset is constructed using 206 high-quality 3D head scans purchased from Triplegangers[1]. The scans were selected such that they cover a relatively balanced range of sexes and skin colors (ITA skin group).

---

[1] https://triplegangers.com/

Ages range from 18 to 78 years old. All of the scans were captured under neutral expression and uniform lighting. We obtain UV texture maps compatible with FLAME [39] by registering the FLAME model to the scans using the approach in [39] and we treat the texture maps as approximate ground-truth albedo.

To create a photo-realistic dataset we rendered the scans in complex scenes. Specifically, we used 50 HDR environment maps from Poly Haven[2] covering both natural and artificial lighting. In each scene we rendered three head scans under the same illumination. To ensure a balanced distribution of skin types, each image was constructed by first randomly selecting a skin type, and then randomly selecting a sample scan within the type. Our final testing dataset contains 721 images and 2163 facial crops under different illumination, with approximate ground-truth UV maps. Details and examples are provided in Sup. Mat.

Evaluation metrics. To focus on both accuracy and fairness, we use the following metrics:

- ITA error. We compute fidelity of skin tone by taking the average error in ITA (degrees, see Eq. 1) between predicted and ground-truth UV maps, over a skin mask region (see Sup. Mat.). We report average error per skin type (I to VI), as well as average ITA error across all groups.

- Bias score. We quantify the *bias* of a method in terms of skin color by measuring the standard deviation across the *per-group* ITA errors (note that this is not the same as the standard deviation over the full dataset). A low value indicates roughly equal performance (i.e. unbiased) across all skin tones.

- Total score. We summarize average ITA and bias score into a single score, which is the sum of the two.

- MAE. We also report results with a commonly employed image metric, namely mean average error. Here we calculate errors over the entire UV map, as opposed to only the skin region.

Fairness of albedo estimates of current methods. Using the above evaluation criteria, we benchmark several recent methods for facial albedo estimation from images in the wild. These are summarized in the top rows of Tab. 1. When observing the "ITA per skin type" column, we note that the accuracy of all methods varies with skin type, performing best on those values that are better represented by the statistical model[3]. In particular, extreme skin types such as I and VI tend to have noticeably larger errors. Low (bold) numbers for some types and high ones for others on each line indicates bias. An unbiased algorithm should have roughly equal errors across skin types.

---

[2] https://polyhaven.com/

[3] For GANFIT [25], the albedos contain a significant amount of baked-in lighting, and were captured with lower light conditions, hence the tendency to do well on dark skin tones.

4   Reasons Behind Racially Biased Albedo Estimations

Recovering the 3D shape and appearance of an object from a single RGB image is a highly ill-posed problem due to fundamental ambiguities arising from the interplay between geometry [4, 21] and lighting [51] in the image formation process. For known objects such as the human face, a standard strategy is to use a strong prior in the form of a low-dimensional statistical model, e.g. the 3D Morphable Model (3DMM) [7, 48] and its variants [13, 39, 61], to constrain the solution to the space of valid shapes and appearances. This idea has been widely adopted and has led to impressive advances in the field [20]. Yet, no careful consideration has been given to ensure that the models cover a balanced demographic distribution; indeed, the most widely used appearance model [48] was built using two-hundred white subjects. Several 3DMM variants have been made available (e.g. [38, 61]) but none of them ensure a balanced distribution of skin tones. It is worth noting that a biased albedo model, employed within a neural network, can still be trained to produce outputs that are far away from the statistical mean. This is the case for example of MGCNet [58] (see Tab. 1), which uses low regularization weights to extrapolate results for type V skin tones. However, these are noisy estimates that do not faithfully represent the albedo, and the model still cannot extrapolate to type VI skin tones.

Even if one had an unbiased statistical model of face albedo, the problem remains ill-posed, leading to an *algorithmic* source of bias [44]. There is a fundamental ambiguity between scene lighting and albedo that cannot be resolved without strong assumptions [51]. Even with a good statistical model of face albedo, there are an infinite number of valid combinations of light and albedo that can explain the image. For example, an image of a darked-skin person can be explained by both dark skin and a bright light, or light skin and a dim light. This can be easily observed by looking at the shading equation commonly used for diffuse objects:

$$I_R = I_A \odot I_S, \tag{2}$$

where I_R is the final rendered image, I_S is the shading image, I_A is the albedo image, and $\odot$ denotes the Hadamard product. When both the albedo and light are unknown there is a scale ambiguity: for a fixed target image I_R, an increase in I_S by a factor of s results in a decrease in I_A by a factor of $1/s$. Given that skin tone dominates the albedo color in I_A, an overly bright estimate of the light will result in an overly dark estimate of the skin tone, and vice-versa. To address this we next propose a method that exploits scene lighting to reduce ambiguity in conjunction with an improved prior.

5   Unbiased Estimation via Scene Disambiguation Cues

Resolving the ambiguity above requires additional information. While most methods in the literature work with a cropped face image, our key insight is that the larger image contains important disambiguation cues.

Table 1: Comparison to state of the art on the FAIR benchmark. We show: average ITA error over all skin types; bias score (standard deviation); total score (avg. ITA+Bias); mean average error; and avg. ITA score per skin type in degrees (I: very light, VI: very dark). Our method achieves more balanced estimates, as can be seen in the bias score, as well as accurate skin color predictions.

Method	Avg. ITA ↓	Bias ↓	Score ↓	MAE ↓	ITA per skin type ↓					
					I	II	III	IV	V	VI
Deng et al. [16]	22.57	24.44	47.02	27.98	**8.92**	9.08	8.15	10.90	28.48	69.90
GANFIT [25]	62.29	31.81	94.11	63.31	94.80	87.83	76.25	65.05	38.24	**11.59**
MGCNet [58]	21.41	17.58	38.99	25.17	19.98	12.76	8.53	9.21	22.66	55.34
DECA [22]	28.74	29.24	57.98	38.17	9.34	11.66	11.58	16.69	39.10	84.06
INORig [3]	27.68	28.18	55.86	33.20	23.25	11.88	**4.86**	9.75	35.78	80.54
CEST [70]	35.18	12.14	47.32	29.92	50.98	38.77	29.22	23.62	21.92	46.57
Ours (BFM)	16.19	15.33	31.52	21.82	12.44	**6.48**	5.69	9.47	16.67	46.37
Ours (AlbedoMM)	17.72	15.28	33.00	19.48	15.50	10.48	8.42	**7.86**	**15.96**	48.11
Ours (BalancedAlb)	**13.87**	**2.79**	**16.67**	18.41	11.90	11.87	11.20	13.92	16.15	18.21

5.1 Model

We begin by describing the image formation model employed throughout this work.

Geometry. We reconstruct geometry using the FLAME [39] statistical model, which parameterizes a face/head mesh using identity $\boldsymbol{\beta} \in \mathbb{R}^{|\boldsymbol{\beta}|}$, pose $\boldsymbol{\theta} \in \mathbb{R}^{3k+3}$ (with $k = 4$ the number of joints), and expression $\boldsymbol{\psi} \in \mathbb{R}^{|\boldsymbol{\psi}|}$ latent vectors.

Albedo. We use a low-dimensional linear model to represent diffuse albedo. To avoid the biases present in current publicly available models, we purchased 54 uniformly lit 3D head scans from 3DScanStore[4], covering the full range of skin types as measured by the ITA score. We converted these into the $d \times d$ FLAME UV texture space, and used Principal Component Analysis (PCA) to learn a model that, given albedo parameters $\boldsymbol{\alpha} \in \mathbb{R}^{|\boldsymbol{\alpha}|}$, outputs a UV albedo map $A(\boldsymbol{\alpha}) \in \mathbb{R}^{d \times d \times 3}$. The albedo image is reconstructed as $I_A = W(A(\boldsymbol{\alpha}))$, where W is a warping function that converts the UV map into camera space.

Illumination. We take the standard approach of approximating environment lighting using spherical harmonics (SH). Using this, the color of pixel (i, j) is computed as

$$
\begin{aligned}
I_R(i,j) &= I_A(i,j) \cdot I_S(i,j) \\
&= I_A(i,j) \cdot \sum_{k=1}^{B^2} \gamma_k \mathbf{H}_k(N(i,j)) \\
&= I_A(i,j) \cdot \gamma \cdot \mathbf{h}(i,j)
\end{aligned}
\tag{3}
$$

---

[4] https://www.3dscanstore.com/

where $I_R(i,j)$ is the intensity at pixel (i,j) (computed once for each channel); $B = 3$ is the number of SH bands; $\gamma = (\gamma_1, \ldots, \gamma_{B^2})$ is the vector containing the SH parameters; $\mathbf{H}_k$, $k = 1 \ldots B^2$, are the orthogonal SH basis functions, $\mathbf{h}(i,j) \in \mathbb{R}^{B^2}$ is the vectorised version of $\mathbf{H}_k(N(i,j))$, and N is the normal image.

Since the overall intensity of the image can be described by either the albedo I_A or the shading I_S, we mitigate this fact by decomposing the SH coefficient vector of each color channel into a unit-scale vector and its norm:

$$I_R(i,j) = ||\gamma|| \cdot I_A(i,j) \cdot \frac{\gamma}{||\gamma||} \cdot \mathbf{h}(i,j). \tag{4}$$

The scale factor $||\gamma||$ can now be regarded as the overall light intensity (one value per RGB channel), while the unit-scale SH parameters $\frac{\gamma}{||\gamma||}$ contain the light directional information. Given a fixed directional SH parameter, the scale is the key variable that modulates the skin tone. We estimate this value independent of the albedo and the normalized SH vector.

5.2 TRUST Network

An image of a face is typically only a small part of a larger image. Our key novelty here is to address the light/albedo ambiguity problem by leveraging *scene cues* to regularize the ill-posed problem. This is implemented by three design choices: (1) a novel *scene consistency* loss, (2) a two-branch architecture that exploits the SH decomposition from Eq. 4, and (3) a light-conditioned albedo estimation network.

Scene consistency loss. First, we observe that a scene containing multiple faces can provide hints about the illumination. Taking inspiration from human color constancy, we can assume a single model of illumination for the entire scene, such that when we observe a variety of skin colors we know that the difference is due to albedo and not lighting[5]. While other works have considered albedo consistency among different views as a cue for disambiguation [59, 63], *light consistency* within an image has been mostly unexplored. We formalize the idea by requiring the SH parameters of different faces in a same image to be close to each other. Specifically, given the set of all the estimated SH vectors $\{\gamma^i\}_{i=\{1..N\}}$, where N is the number of faces in the image, we penalize the difference between any two facial crops as $L_{sc} = ||P(\{\gamma^i\}) - \{\gamma^i\}||_1$, where $P(\cdot)$ is a random permutation function (illustrated in Fig. 2 right).

However, we note that this loss alone cannot ensure a correct albedo reconstruction: it can only work when there is a variety of skin tones present in the scene, otherwise the estimated albedos can still be consistently brighter or darker than the true albedo.

---

[5] There are exceptions to this, such as a scenes where some faces are in shadow or where the lighting is high-frequency.

Two-branch architecture. As a complementary cue, we leverage the decomposition of SH into norm and direction described in Eq. 4, and exploit the fact that the global illuminant can provide additional information about the overall skin tone. To see why, consider the example of a dark-skinned person during daylight. The albedo can be correctly predicted as a dark skin under bright light, or incorrectly predicted as a light skin under dark light; the choice will depend largely on the bias of the prior. Yet, a coarse estimate of the global illuminant will reveal that the first case is the correct one. This provides a strong cue about where the output should lie in the skin color palette.

We implement this observation by proposing a two-branch architecture, shown in Fig. 2, left. First, the input image is passed through a *scene light encoder* E_{scene_light} that predicts the norm of the scene spherical harmonics $||\gamma||$ for each RGB channel, thus capturing the overall light intensity and color. We next obtain facial crops using an off-the-shelf face detector (or ground-truth values at training time). From these crops, an *albedo encoder* E_{albedo} predicts the albedo parameters of the model $\boldsymbol{\alpha}$, while a *crop-based light encoder* E_{face_light} predicts the normalized SH vector $\gamma' = \gamma/||\gamma||$. The outputs of E_{scene_light} and E_{face_light} are then combined to obtain the final SH prediction γ.

Light-conditioned albedo estimation. We note that during test time, the albedo estimator E_{albedo} does not contain information about the scene, which can still introduce ambiguities. To address this, we propose to condition E_{albedo} on the estimated light intensity. For this, we broadcast the intensity factor from the scene light encoder into an image of the same size as the facial crop, and concatenate it with the input as an additional channel.

Semi-supervised training. Given that unsupervised disentanglement cannot be solved without proper inductive biases [41] or without a limited amount of potentially imprecise supervision [42], here we use a semi-supervised learning strategy. Specifically, we generate a synthetic training set of 50k images using 1170 scans acquired from Renderpeople[6], combined with 273 panoramic HDR images from Poly Haven[7]. We train the networks using a combined synthetic/real dataset to ensure generalization.

Training. TRUST is trained using the following loss function:

$$\mathcal{L} = \lambda_{pho}L_{pho} + \lambda_{sc}L_{sc} + \mathbb{I}\lambda_{SH}L_{SH} + \mathbb{I}\lambda_{alb}L_{alb}. \tag{5}$$

Here, $L_{pho} = ||I - I_R||_1$ is the L1 photometric loss between the input image and rendered image, and L_{sc} is the scene consistency loss. $\mathbb{I}$ is an indicator function with value 1 for supervised training data and 0 for real. When training with synthetic data we also employ an L1 loss $L_{SH} = ||\gamma - \tilde{\gamma}||_1^1$ between ground-truth SH coefficients γ and the estimates $\tilde{\gamma}$, as well as an L1 loss between the rendered

---

[6] https://renderpeople.com/

[7] Note that these scenes are completely different from those used in the evaluation benchmark.

Input   GANFIT   INORig   MGCNet   Deng et al   CEST   DECA   TRUST   GT-albedo
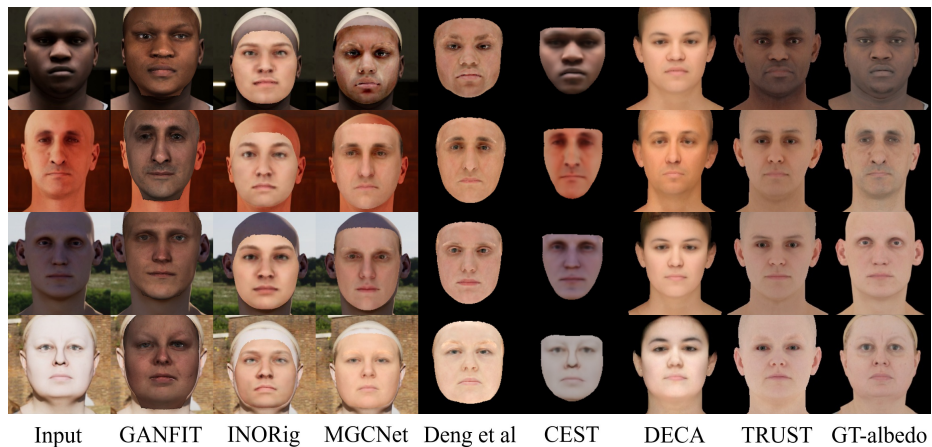
Fig. 3: Comparison with recent face reconstruction methods on the proposed albedo benchmark. From left to right: input image, GANFIT [24], INORig [3], MGCNet [58], Deng et al. [16], CEST [70], DECA [22], TRUST (ours) and ground-truth albedo rendering.

albedo and ground-truth. Note that we include the self-supervised loss on synthetic data since the ground-truth light and albedo are only approximations to the physical ground-truth. We set $\lambda_{pho} = 10, \lambda_{sc} = 10, \lambda_{SH} = 20, \lambda_{alb} = 20$ to weight the loss terms based on validation-set performance.

To compute the photometric loss we need an estimate of the geometry and camera. For this we use a pre-trained state-of-the-art geometry estimation network, DECA [22], which provides FLAME shape and expression coefficients, as well as weak perspective camera parameters. This module is fixed during training.

5.3   Implementation details

We implemented our approach in PyTorch [47], using the differentiable rasterizer from Pytorch3D [52] for rendering. We use Adam [31] as the optimizer with a learning rate of $1e-4$. All of the encoders use Resnet-50 architectures with input images of size 224×224 (in the case of full scene images, we first randomly crop a square and then resize). The UV space size is $d = 256$. To get shape, expression and camera parameters we use DECA [22] to regress the FLAME parameters $|\boldsymbol{\beta}| = 100$ and $|\boldsymbol{\psi}| = 50$. To train with real data, we use the subset of the OpenImages dataset [35] that contains faces.

We apply a two-stage training strategy. The first stage employs fully supervised training with a batch size of 32 using the synthetic dataset for one epoch. The second stage uses semi-supervised training with a batch size of 48 and all proposed losses for 4 epochs. More details can be found in Sup. Mat.

Table 2: **Ablation study.** We show comparisons to the following alternatives:
(1) light estimation from crops ("faceSH") with self-supervised (self) and semi-
supervised (semi) training sets; (2) SH intensity estimation from the scene, with
SH directional estimation from crops ("fuseSH"); (3) fuseSH with scene consis-
tency loss ("sc"); (4) fuseSH with conditioning ("cond"); (5) Ours: fuseSH with
scene consistency and conditioning.

Method	Avg. ITA ↓	Bias ↓	Score ↓	MAE ↓
faceSH + self	24.17	11.46	35.62	25.81
faceSH + semi	14.70	6.60	21.31	17.64
fuseSH	15.22	11.08	26.31	**17.02**
fuseSH + sc	15.82	3.50	19.32	20.37
fuseSH + cond	**14.16**	6.56	20.72	17.05
Ours	14.18	**2.63**	**16.81**	19.08

6   Evaluation

We compare TRUST qualitatively and quantitatively with several SOTA meth-
ods. MGCNet [58], Deng et al. [16], INORig [3] and DECA [22] use the Basel Face
Model (BFM) [48] for albedo estimation; GANFIT [25] uses its own GAN-based
appearance model; and CEST [70] is a model-free approach.

6.1   Qualitative results

We provide qualitative results on both real images (Fig. 4a) and synthetic images
from the FAIR benchmark (Fig. 3), showing faces with a variety of skin colors
and scene illuminations. We observe that (1) GANFIT [25], INORig [3], and
DECA [22] produce albedo maps with low variety, hence achieving low ITA
values on specific skin types, but high ITA values for other skin types; (2) the
model-free approach CEST [70] bakes in a significant amount of light, since
they cannot properly disentangle it from the albedo; (3) MGCNet [58] and, to a
certain degree, Deng et al. [16] produce more diverse albedos, but since the BFM
model does not include dark skin tones in the training dataset, these can only be
encoded by extrapolating considerably, introducing noise. For more qualitative
examples, see Sup. Mat.

   To evaluate robustness we apply TRUST on images of a same subject un-
der different lighting, and with varying backgrounds. Fig. 4b shows that the
estimated albedo, per subject, is consistent across lighting and background vari-
ations, and that skin tones are well captured.

6.2   Quantitative comparisons

We quantitatively compare several albedo estimation methods on the FAIR
benchmark as described in Sec. 3. Table 1 shows that the proposed approach

| (a) | (b) |

Fig. 4: **(a)** Qualitative comparisons on *real world images*. From left to right: 1) Input, 2) INORig [3], 3) MGCNet [58], 4) DECA [22], 5) TRUST (Ours). **(b)** Given images of the same subject (from the FAIR benchmark test set) under varying lighting (Rows 1 & 3), TRUST outputs similar albedos (Rows 2 & 4).

(using the balanced albedo model) outperforms existing methods on all aggregated measures (i.e. Avg. ITA, Bias, Score, and MAE), and produces a more uniform (low) error across skin types. Note that while Deng et al. [16], INORig [3], and GANFIT [25] obtain the lowest scores for individual skin types, they have large errors for others, indicating their bias towards particular skin tones, which results in higher aggregated errors.

It is worth noting that, while shading information acts locally and provides gradients for geometry reconstruction, the skin tone is a global property of the albedo (low-frequency component). Hence, correct shape estimation does not necessarily imply good skin tone estimation (and vice-versa), which explains why methods such as DECA can achieve state-of-the-art results on shape estimation, even with strong regularization on the albedo coefficients that lead to incorrect skin tone estimates.

6.3   Ablation studies

Effect of albedo model space. We investigate how much a biased albedo space affects the final performance. To this end, we train two additional versions of our final network, but using instead the BFM albedo space [48] and AlbedoMM [61]. Both model variants show large improvements on most skin types compared to prior work (see Tab. 1). This demonstrates that addressing the light/albedo ambiguity is important for unbiased albedo estimation, and allows to push the albedo space to its representational limit. Both model variants perform similarly (see Tab. 1) in Avg. ITA error and Bias scores (std), but around 20% and 580% higher than our final model with a balanced albedo space. Most errors of these two variants come from type V and VI, which shows the importance of using a balanced albedo model to cover the full range of skin tones.

Light and albedo estimation from facial crops (faceSH): Here, both light and albedo encoders use facial crops as input, which is comparable to prior methods. We evaluate this setting using self-supervised training and semi-supervised training, to test the importance of the synthetic vs in-the-wild training set for reducing racial bias. Results are shown in the first two rows of Table 2, where we see that all metrics are significantly improved with access to synthetic data.
Scene light estimation (fuseSH): We next consider the case where the light intensity is estimated from the scene image, as in our approach, but with the following differences: (1) without conditioning and without scene consistency (fuseSH), (2) with scene consistency alone (fuseSH+sc), (3) with conditioning alone (fuseSH + cond). We observe in Table 2 that each of these components improves a different aspect: while conditioning results in better skin color predictions, the scene consistency encourages fairer estimates. Our final model, which uses both scene-consistency and conditioning, achieves the lowest total score.

7   Conclusions

This work presented initial steps towards unbiased estimation of facial albedo from images in the wild, using two main contributions. First, we built a new evaluation benchmark that is balanced in terms of skin type. We used this benchmark to highlight biases in current methods. We further proposed a new albedo estimation approach that addresses the light/albedo ambiguity problem, hence encouraging fairer estimates. Our solution is built on the idea that the scene image can be exploited as a cue to disambiguate light and albedo, resulting in more accurate predictions. The experimental results confirm that scene information helps to obtain fairer albedo reconstructions. We hope that this work will encourage the development of methods that are designed not just towards realism, but also towards fairness.

Bibliography

[1] Adamson, A.S., Smith, A.: Machine learning and health care disparities in dermatology. JAMA dermatology **154**(11), 1247–1248 (2018) 2

[2] Aldrian, O., Smith, W.A.: Inverse rendering of faces with a 3D morphable model. Transactions on Pattern Analysis and Machine Intelligence (PAMI) **35**(5), 1080–1093 (2012) 3, 4

[3] Bai, Z., Cui, Z., Liu, X., Tan, P.: Riggable 3D face reconstruction via in-network optimization. In: Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6216–6225 (June 2021) 8, 11, 12, 13

[4] Bas, A., Smith, W.A.P.: What does 2D geometric information really tell us about 3D face shape? International Journal of Computer Vision (IJCV) **127**(10), 1455–1473 (2019) 7

[5] Bianco, S., Schettini, R.: Adaptive color constancy using faces. Transactions on Pattern Analysis and Machine Intelligence (PAMI) **36**(8), 1505–1518 (2014) 4

[6] Blanz, V., Romdhani, S., Vetter, T.: Face identification across different poses and illuminations with a 3D morphable model. In: International Conference on Automatic Face & Gesture Recognition (FG). pp. 202–207 (2002) 3

[7] Blanz, V., Vetter, T.: A morphable model for the synthesis of 3d faces. In: SIGGRAPH. pp. 187–194 (1999) 3, 7

[8] Buolamwini, J., Gebru, T.: Gender shades: Intersectional accuracy disparities in commercial gender classification. In: Conference on Fairness, Accountability and Transparency. pp. 77–91. PMLR (2018) 2, 4, 5

[9] Chardon, A., Cretois, I., Hourseau, C.: Skin colour typology and suntanning pathways. International journal of cosmetic science **13**(4), 191–208 (1991) 5

[10] Chaudhuri, B., Vesdapunt, N., Shapiro, L., Wang, B.: Personalized face modeling for improved face reconstruction and motion retargeting. In: IEEE European Conference on Computer Vision (ECCV) (2020) 3

[11] Chen, A., Chen, Z., Zhang, G., Mitchell, K., Yu, J.: Photo-realistic facial details synthesis from single image. In: International Conference on Computer Vision (ICCV). pp. 9429–9439 (2019) 3

[12] Choi, H., Choi, K., Suk, H.: Performance of the 14 skin-colored patches in accurately estimating human skin color. In: Computational Imaging XV. pp. 62–65 (2017) 4

[13] Dai, H., Pears, N., Smith, W., Duncan, C.: Statistical modeling of craniofacial shape and texture. International Journal of Computer Vision (IJCV) **128**(2), 547–571 (2019) 7

[14] Del Bino, S., Sok, J., Bessac, E., Bernerd, F.: Relationship between skin response to ultraviolet exposure and skin color type. Pigment cell research **19**(6), 606–614 (2006) 5

[15] Del Bino, S., Bernerd, F.: Variations in skin colour and the biological consequences of ultraviolet radiation exposure. British Journal of Dermatology **169**, 33–40 (2013) 5

[16] Deng, Y., Yang, J., Xu, S., Chen, D., Jia, Y., Tong, X.: Accurate 3D face reconstruction with weakly-supervised learning: From single image to image set. In: Conference on Computer Vision and Pattern Recognition Workshops (CVPR-W). pp. 0–0 (2019) 3, 8, 11, 12, 13

[17] Dooley, S., Downing, R., Wei, G., Shankar, N., Thymes, B., Thorkelsdottir, G., Kurtz-Miott, T., Mattson, R., Obiwumi, O., Cherepanova, V., et al.: Comparing human and machine bias in face recognition. arXiv preprint arXiv:2110.08396 (2021) 5

[18] Drozdowski, P., Rathgeb, C., Dantcheva, A., Damer, N., Busch, C.: Demographic bias in biometrics: A survey on an emerging challenge. Transactions on Technology and Society **1**(2), 89–103 (2020) 2, 4

[19] Egger, B., Schönborn, S., Schneider, A., Kortylewski, A., Morel-Forster, A., Blumer, C., Vetter, T.: Occlusion-aware 3D morphable models and an illumination prior for face image analysis. International Journal of Computer Vision (IJCV) **126**(12), 1269–1287 (2018) 4

[20] Egger, B., Smith, W.A.P., Tewari, A., Wuhrer, S., Zollhöfer, M., Beeler, T., Bernard, F., Bolkart, T., Kortylewski, A., Romdhani, S., Theobalt, C., Blanz, V., Vetter, T.: 3D morphable face models - past, present, and future. Transactions on Graphics (TOG) **39**(5), 157:1–157:38 (2020) 3, 7

[21] Egger, B., Sutherland, S., Medin, S.C., Tenenbaum, J.: Identity-expression ambiguity in 3D morphable face models. arXiv preprint arXiv:2109.14203 (2021) 7

[22] Feng, Y., Feng, H., Black, M.J., Bolkart, T.: Learning an animatable detailed 3D face model from in-the-wild images. Transactions on Graphics, (Proc. SIGGRAPH) **40**(4), 88:1–88:13 (2021) 3, 8, 11, 12, 13

[23] Fitzpatrick, T.B.: The validity and practicality of sun-reactive skin types i through vi. Archives of dermatology **124**(6), 869–871 (1988) 5

[24] Gecer, B., Deng, J., Zafeiriou, S.: Ostec: One-shot texture completion. In: Conference on Computer Vision and Pattern Recognition (CVPR). pp. 7628–7638 (2021) 11

[25] Gecer, B., Ploumpis, S., Kotsia, I., Zafeiriou, S.: Ganfit: Generative adversarial network fitting for high fidelity 3d face reconstruction. In: Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1155–1164 (2019) 3, 6, 8, 12, 13

[26] Genova, K., Cole, F., Maschinot, A., Sarna, A., Vlasic, D., Freeman, W.T.: Unsupervised training for 3D morphable model regression. In: Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8377–8386 (2018) 3

[27] Gerig, T., Morel-Forster, A., Blumer, C., Egger, B., Luthi, M., Schönborn, S., Vetter, T.: Morphable face models - an open framework. In: International Conference on Automatic Face & Gesture Recognition (FG). pp. 75–82 (2018) 3

[28] Hu, G., Mortazavian, P., Kittler, J., Christmas, W.: A facial symmetry prior for improved illumination fitting of 3d morphable model. In: 2013 International Conference on Biometrics (ICB). pp. 1–6. IEEE (2013) 4

[29] Kim, H., Zollhöfer, M., Tewari, A., Thies, J., Richardt, C., Theobalt, C.: InverseFaceNet: Deep monocular inverse face rendering. In: Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4625–4634 (2018) 3

[30] Kim, T., Rushmeier, H., Dorsey, J., Nowrouzezahrai, D., Syed, R., Jarosz, W., Darke, A.: Countering racial bias in computer graphics research. arXiv preprint arXiv:2103.15163 (2021) 2, 4

[31] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014) 11

[32] Kinyanjui, N.M., Odonga, T., Cintas, C., Codella, N.C., Panda, R., Sattigeri, P., Varshney, K.R.: Fairness of classifiers across skin tones in dermatology. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 320–329. Springer (2020) 2

[33] Kips, R., Gori, P., Perrot, M., Bloch, I.: Ca-gan: Weakly supervised color aware gan for controllable makeup transfer. In: European Conference on Computer Vision. pp. 280–296. Springer (2020) 4

[34] KIPS, R., TRAN, L., MALHERBE, E., PERROT, M.: Beyond color correction: Skin color estimation in the wild through deep learning. Electronic Imaging **2020**(5), 82–1 (2020) 4

[35] Krasin, I., Duerig, T., Alldrin, N., Ferrari, V., Abu-El-Haija, S., Kuznetsova, A., Rom, H., Uijlings, J., Popov, S., Veit, A., et al.: Openimages: A public dataset for large-scale multi-label and multi-class image classification. Dataset available from https://github. com/openimages **2**(3), 18 (2017) 11

[36] Krishnapriya, K.S., Albiero, V., Vangara, K., King, M.C., Bowyer, K.W.: Issues related to face recognition accuracy varying based on race and skin tone. IEEE Transactions on Technology and Society **1**(1), 8–20 (2020). https://doi.org/10.1109/TTS.2020.2974996 5

[37] Lattas, A., Moschoglou, S., Gecer, B., Ploumpis, S., Triantafyllou, V., Ghosh, A., Zafeiriou, S.: AvatarMe: Realistically renderable 3D facial reconstruction. In: Conference on Computer Vision and Pattern Recognition (CVPR). pp. 760–769 (2020) 3

[38] Li, R., Bladin, K., Zhao, Y., Chinara, C., Ingraham, O., Xiang, P., Ren, X., Prasad, P., Kishore, B., Xing, J., et al.: Learning formation of physically-based face attributes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3410–3419 (2020) 5, 7

[39] Li, T., Bolkart, T., Black, M.J., Li, H., Romero, J.: Learning a model of facial shape and expression from 4D scans. ACM Transactions on Graphics, (Proc. SIGGRAPH Asia) **36**(6), 194:1–194:17 (2017), https://doi.org/10.1145/3130800.3130813 6, 7, 8

[40] Lin, J., Yuan, Y., Shao, T., Zhou, K.: Towards high-fidelity 3D face reconstruction from in-the-wild images using graph convolutional networks. In: Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5891–5900 (2020) 3

[41] Locatello, F., Bauer, S., Lucic, M., Rätsch, G., Gelly, S., Schölkopf, B., Bachem, O.: Challenging common assumptions in the unsupervised learning of disentangled representations. arXiv preprint arXiv:1811.12359 (2018) 10

[42] Locatello, F., Tschannen, M., Bauer, S., Rätsch, G., Schölkopf, B., Bachem, O.: Disentangling factors of variation using few labels. arXiv preprint arXiv:1905.01258 (2019) 10

[43] Marguier, J., Bhatti, N., Baker, H., Harville, M., Süsstrunk, S.: Assessing human skin color from uncalibrated images. International Journal of Imaging Systems and Technology **17**(3), 143–151 (2007) 4

[44] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., Galstyan, A.: A survey on bias and fairness in machine learning. ACM Comput. Surv. **54**(6) (Jul 2021). https://doi.org/10.1145/3457607, https://doi.org/10.1145/3457607 7

[45] Merler, M., Ratha, N., Feris, R.S., Smith, J.R.: Diversity in faces. arXiv preprint arXiv:1901.10436 (2019) 5

[46] Osoba, O.A., Welser IV, W.: An intelligence in our image: The risks of bias and errors in artificial intelligence. Rand Corporation (2017) 2

[47] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems (NeurIPS) (2019) 11

[48] Paysan, P., Knothe, R., Amberg, B., Romdhani, S., Vetter, T.: A 3d face model for pose and illumination invariant face recognition. In: 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance. pp. 296–301. Ieee (2009) 3, 7, 12, 13

[49] Pichon, L.C., Landrine, H., Corral, I., Hao, Y., Mayer, J.A., Hoerster, K.D.: Measuring skin cancer risk in african americans: is the fitzpatrick skin type classification scale culturally sensitive. Ethn Dis **20**(2), 174–179 (2010) 5

[50] Rajkomar, A., Hardt, M., Howell, M.D., Corrado, G., Chin, M.H.: Ensuring fairness in machine learning to advance health equity. Annals of internal medicine **169**(12), 866–872 (2018) 2

[51] Ramamoorthi, R., Hanrahan, P.: A signal-processing framework for inverse rendering. In: Pocock, L. (ed.) SIGGRAPH. pp. 117–128 (2001) 3, 4, 7

[52] Ravi, N., Reizenstein, J., Novotny, D., Gordon, T., Lo, W.Y., Johnson, J., Gkioxari, G.: Pytorch3d. https://github.com/facebookresearch/pytorch3d (2020) 11

[53] Robinson, J.P., Livitz, G., Henon, Y., Qin, C., Fu, Y., Timoner, S.: Face recognition: too bias, or not too bias? In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 0–1 (2020) 4

[54] Sahasrabudhe, M., Shu, Z., Bartrum, E., Güler, R.A., Samaras, D., Kokkinos, I.: Lifting autoencoders: Unsupervised learning of a fully-disentangled 3D morphable model using deep non-rigid structure from motion. In: International Conference on Computer Vision Workshops (ICCV-W). pp. 4054–4064 (2019) 3

[55] Saito, S., Wei, L., Hu, L., Nagano, K., Li, H.: Photorealistic facial texture inference using deep neural networks. In: Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5144–5153 (2017) 3

[56] Schönborn, S., Egger, B., Morel-Forster, A., Vetter, T.: Markov chain monte carlo for automated face image analysis. International Journal of Computer Vision (IJCV) **123**(2), 160–183 (2017) 3

[57] Sengupta, S., Kanazawa, A., Castillo, C.D., Jacobs, D.W.: SfSNet: Learning shape, reflectance and illuminance of facesin the wild'. In: Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6296–6305 (2018) 3

[58] Shang, J., Shen, T., Li, S., Zhou, L., Zhen, M., Fang, T., Quan, L.: Self-supervised monocular 3D face reconstruction by occlusion-aware multi-view geometry consistency. In: European Conference on Computer Vision (ECCV). vol. 12360, pp. 53–70 (2020) 3, 7, 8, 11, 12, 13

[59] Shi, F., Wu, H.T., Tong, X., Chai, J.: Automatic acquisition of high-fidelity facial performances using monocular videos. ACM Transactions on Graphics (TOG) **33**(6), 1–13 (2014) 9

[60] Shu, Z., Yumer, E., Hadap, S., Sunkavalli, K., Shechtman, E., Samaras, D.: Neural face editing with intrinsic image disentangling. In: Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5541–5550 (2017) 3

[61] Smith, W.A.P., Seck, A., Dee, H., Tiddeman, B., Tenenbaum, J., Egger, B.: A morphable face albedo model. In: Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5010–5019 (2020) 3, 7, 13

[62] Terhörst, P., Kolf, J.N., Huber, M., Kirchbuchner, F., Damer, N., Moreno, A.M., Fierrez, J., Kuijper, A.: A comprehensive study on face recognition biases beyond demographics. IEEE Transactions on Technology and Society **3**(1), 16–30 (2021) 2

[63] Tewari, A., Bernard, F., Garrido, P., Bharaj, G., Elgharib, M., Seidel, H.P., Pérez, P., Zollhofer, M., Theobalt, C.: Fml: Face model learning from videos. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 10812–10822 (2019) 3, 9

[64] Tewari, A., Zollhöfer, M., Garrido, P., Bernard, F., Kim, H., Pérez, P., Theobalt, C.: Self-supervised multi-level face model learning for monocular reconstruction at over 250 hz. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2549–2559 (2018) 3

[65] Tewari, A., Zollöfer, M., Kim, H., Garrido, P., Bernard, F., Perez, P., Christian, T.: MoFA: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction. In: International Conference on Computer Vision (ICCV) (2017) 3

[66] Thies, J., Zollhöfer, M., Stamminger, M., Theobalt, C., Nießner, M.: Face2Face: Real-time face capture and reenactment of RGB videos. In: Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2387–2395 (2016) 3

[67] Tran, L., Liu, F., Liu, X.: Towards high-fidelity nonlinear 3d face morphable model. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1126–1135 (2019) 3

[68] Tran, L., Liu, X.: Nonlinear 3d face morphable model. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7346–7355 (2018) 3

[69] Wang, M., Deng, W., Hu, J., Tao, X., Huang, Y.: Racial faces in the wild: Reducing racial bias by information maximization adaptation network. In: International Conference on Computer Vision (ICCV). pp. 692–702 (2019) 4, 5

[70] Wen, Y., Liu, W., Raj, B., Singh, R.: Self-supervised 3D face reconstruction via conditional estimation. In: International Conference on Computer Vision (ICCV). pp. 13289–13298 (2021) 3, 8, 11, 12

[71] Wilson, B., Hoffman, J., Morgenstern, J.: Predictive inequity in object detection. arXiv preprint arXiv:1902.11097 (2019) 5

[72] Yamaguchi, S., Saito, S., Nagano, K., Zhao, Y., Chen, W., Olszewski, K., Morishima, S., Li, H.: High-fidelity facial reflectance and geometry inference from an unconstrained image. Transactions on Graphics (TOG) **37**(4), 1–14 (2018) 3

[73] Youn, J., Oh, J., Kim, B., Suh, D., Chung, J., Oh, S., Kim, J., Kang, S.: Relationship between skin phototype and med in korean, brown skin. Photodermatology, photoimmunology & photomedicine **13**(5-6), 208–211 (1997) 5