Face Relighting with Geometrically Consistent Shadows

Andrew Hou[*,1], Michel Sarkis[2], Ning Bi[2], Yiying Tong[1], Xiaoming Liu[1]

[1]Michigan State University, [2]Qualcomm Technologies Inc.

{houandr1, ytong, liuxm}@msu.edu, {msarkis, nbi}@qti.qualcomm.com

https://github.com/andrewhou1/GeomConsistentFR

Abstract

Most face relighting methods are able to handle diffuse shadows, but struggle to handle hard shadows, such as those cast by the nose. Methods that propose techniques for handling hard shadows often do not produce geometrically consistent shadows since they do not directly leverage the estimated face geometry while synthesizing them. We propose a novel differentiable algorithm for synthesizing hard shadows based on ray tracing, which we incorporate into training our face relighting model. Our proposed algorithm directly utilizes the estimated face geometry to synthesize geometrically consistent hard shadows. We demonstrate through quantitative and qualitative experiments on Multi-PIE and FFHQ that our method produces more geometrically consistent shadows than previous face relighting methods while also achieving state-of-the-art face relighting performance under directional lighting. In addition, we demonstrate that our differentiable hard shadow modeling improves the quality of the estimated face geometry over diffuse shading models.

1. Introduction

Single image face relighting is a problem of great interest among the computer vision and computer graphics communities. Relighting consumer photos has been a major driving factor in motivating the problem given widespread interest in photo editing. Face relighting also has applications in other areas such as Augmented Reality (AR) [29], where it can be used to modify facial illuminations to match the environment lighting, and face recognition [16, 33], where it can relight images to frontal illuminations. It is thus relevant both for consumer interests and entertainment and for security applications such as authentication.

Earlier relighting methods [35, 43, 54] tend to make the simplified assumption that light is naturally scattered by the environment and thus diffuse in nature, and that human skin is a lambertian material. While this is sufficient to model

Figure 1. **Overview**. We introduce a novel face relighting method that produces geometrically consistent shadows. By proposing a differentiable algorithm based on the principles of ray tracing that directly uses the face geometry for modeling hard shadows, our method produces physically correct hard shadows which the state-of-the-art face relighting method, Hou *et al.* [10], cannot produce.

general lighting directions and soft shadows, it does not account for non-lambertian effects such as hard shadows from strong directional lights. This is highly problematic since many light sources in the real world (*e.g.* the sun) are best modeled as directional lights. In AR/VR, the environment lighting is often also set to be directional lights. In order to enhance photorealism both for in-the-wild consumer photos and in AR/VR, proper hard shadow modeling is a necessity.

One important problem in face relighting is thus handling hard shadows. Most existing methods do not handle non-diffuse lighting and are unable to synthesize realistic hard shadows [35, 43, 54]. They generally use smooth lighting conditions such as low-order Spherical Harmonics (SH) and train on images with diffuse lighting. While many illumination conditions in the wild are ambient or area-based, these assumptions do not account for the interactions of strong directional lights and point lights, which produce hard shadows. Among current methods that do model hard shadows [10, 29, 30], none are able to guarantee geometrically consistent cast shadows since they do not directly utilize the estimated face geometry to generate them. Without using the geometry directly, the shape of the cast shadows, such as those cast by the nose, may be incorrect.

We introduce a novel differentiable algorithm to estimate

the locations of cast shadows using the principles of ray tracing. We rely on the principle that cast shadows will be located on parts of the face where the projected ray to the light source intersects some occluding surface, such as the nose. Our method can thus leverage the estimated face geometry to produce geometrically consistent hard shadows (see Fig. 1). We further demonstrate that differentiably modeling hard shadows can improve the quality of the face geometry, especially in regions that produce hard shadows (*e.g.* the nose and near the boundary of the face), compared to models that assume diffuse shading. We therefore show that differentiable hard shadow modeling not only benefits the realism of the relit image, but also the intrinsic component estimation which can benefit other downstream tasks.

Our proposed method thus has four main contributions:

⋄ We propose a single image face relighting method that can produce geometrically consistent hard shadows.

⋄ We introduce a novel differentiable algorithm to estimate facial cast shadows based on the estimated geometry.

⋄ We achieve SoTA relighting performance on 2 benchmarks quantitatively/qualitatively under directional lights.

⋄ Our differentiable hard shadow modeling improves the estimated geometry over models that use diffuse shading.

2. Related Work

Face Relighting Prior works can be categorized into 4 groups: intrinsic decomposition [2,3,5–7,16–19,22,23,29, 30, 35, 36, 40, 44–47, 51], image-to-image translation [10, 43, 49, 53, 54], style transfer [21, 27, 34, 38, 39], and ratio images [32, 37, 42, 50]. Intrinsic decomposition estimates the geometry, albedo, and lighting and renders the image with a new lighting. Image-to-image translation instead directly estimates the relit image. Style transfer transfers the lighting of the reference image as a style to the input image. Ratio image methods relight by estimating the ratio between the source and target images or illuminations. A summary of our method compared to recent works is shown in Tab. 1.

Nestmeyer *et al.* [29] model hard shadows using a binary visibility map estimated from a U-Net. The visibility map is not directly constrained by the face geometry and therefore has a large amount of freedom, which can lead to geometrically inconsistent shadows. Also, since it is binary, their cast shadows are black whereas the true intensity of a shadow should match the environment's ambient light.

Hou *et al.* [10] utilize shadow masks to assign higher weights along hard shadow borders. While this improves the shadows, they do not use the geometry and thus the shadows can have any shape. Our model produces hard shadows where rays cast from the face to the light source intersect other parts of the face geometry, which ensures that the shadows are geometrically consistent.

Pandey *et al.* [30] model both diffuse and specular lighting, and can generate non-lambertian effects such as hard

Method	Lighting Model	Model Category	Handles Hard Shadows	Geom. Consistent Hard Shadows
SfSNet [35]	SH	Intrinsic	X	X
DPR [54]	SH	Im2Im	X	X
SIPR [43]	Environment Map	Im2Im	X	X
Nestmeyer [29]	Directional Light	Intrinsic	✓	X
Hou [10]	SH	Im2Im	✓	X
Total Relighting [30]	Environment Map	Intrinsic	✓	X
Proposed	Directional Light	Intrinsic	✓	✓

Table 1. **Method Comparison**. A summary of our proposed method compared to recent face relighting methods.

shadows. However, they rely on a shading network to produce the relit image given the albedo and light maps, which will have network estimation error. Thus, there is no guarantee that they produce geometrically consistent shadows.

Differentiable Rendering and Ray Tracing In recent years, multiple differentiable renderers have been proposed [14, 20, 24, 26, 28] that are suitable for inverse rendering tasks. However, the majority of differentiable renderers do not explicitly model shadows, particularly hard shadows.

The most similar work to ours is Li *et al.* [20], which proposes a differentiable ray tracer to model hard shadows. Their method operates on meshes and introduces a novel Monte Carlo edge sampling algorithm to handle the non-differentiability along triangle edges. Our shadow modeling is also a form of differentiable ray tracing, but operates on the 2.5D points generated from a face depth map rather than a mesh. Instead of integrating over edges of mesh triangles to determine shading, we find it sufficient to sample points between each point on the face and the light source and assign a cast shadow to the point if one of the sampled points intersects with some facial parts, such as the nose.

Recently, Srinivasan *et al.* propose NeRV [41], which can model scene-specific hard shadows from any directions more efficiently than prior work. However, they rely on a visibility MLP to predict the locations of shadows, which leaves the possibility of generating geometrically inconsistent shadows due to network estimation error. Furthermore, NeRV requires hundreds of images with known lighting and pose and can only be trained on one static scene at a time, whereas our model is generalizable due to leveraging public face datasets in training, only requires a single image per subject, and can be applied in inference to any face image.

3. Proposed Method

3.1. Problem Formulation

Our relighting method relies on intrinsic decomposition, and is thus motivated by the rendering equation [11]:

$$L_o(\mathbf{x}, \omega_o) = \int_{\omega_i \in \Omega} f(\mathbf{x}, \omega_i, \omega_o) L_i(\mathbf{x}, \omega_i) \langle \mathbf{n}, \omega_i \rangle \, d\omega_i,$$

(1)

Figure 2. **Model Overview**. Given a single image $\mathbf{I}_t$ and target lighting direction ω_t, our model generates a relit image $\mathbf{I}_p$ with geometrically consistent cast shadows. The geometric consistency is achieved thanks to our shadow mask estimation module, which estimates shadow mask $\mathbf{M}_{shadow}$ using depth map $\mathbf{D}_p$ (the face geometry). $\mathbf{M}_{shadow}$ incorporates non-diffuse cast shadows into our shading $\mathbf{S}_p$.

where $\mathbf{x}$ is a point on the 3D surface, $\mathbf{n}$ is the surface normal at $\mathbf{x}$, ω_i and ω_o are the incoming and outgoing lighting directions respectively, Ω is the unit hemisphere centered around $\mathbf{n}$ with all possible values of ω_i, $L_i(\mathbf{x}, \omega_i)$ and $L_o(\mathbf{x}, \omega_o)$ are the incoming and outgoing radiances respectively, and $f(\mathbf{x}, \omega_i, \omega_o)$ is the bidirectional reflectance distribution function (BRDF) determining the material's reflectance. If only diffuse reflection is considered, *i.e.* $f(\mathbf{x}, \omega_i, \omega_o) = a(\mathbf{x})/\pi$, the rendering equation becomes:

$$L_o(\mathbf{x}, \omega_o) = \frac{a(\mathbf{x})}{\pi} \int_{\omega_i \in \Omega} L_i(\mathbf{x}, \omega_i) \langle \mathbf{n}, \omega_i \rangle \, d\omega_i = a(\mathbf{x})s(\mathbf{x}).$$
(2)

Here $a(\mathbf{x})$ is the diffuse albedo and $s(\mathbf{x})$ the diffuse shading.

Assuming there is a dominant directional light in the scene, without considering the visibility of the light source from $\mathbf{x}$, the diffuse shading $s(\mathbf{x})$ can be approximated by:

$$s(\mathbf{x}) = i_a + i_d \langle \mathbf{n}, \omega_d \rangle,$$
(3)

where i_a is the intensity of the ambient light in the scene and i_d and ω_d is the intensity and direction of the directional light, respectively. In other words, $L_i(\mathbf{x}, \omega_i) = i_a + i_d \pi \, \delta(\omega_i, \omega_d)$, where $\delta(\cdot, \cdot)$ is the Dirac delta function.

To model *differentiable* cast shadows, we introduce the shadow mask $\mathbf{M}_{shadow}$ to model the visibility of the directional light. For each point, $\mathbf{M}_{shadow}$ stores a value close to 0 if the point is under a cast shadow and close to 1 otherwise. The intensity under a cast shadow should be i_a, since the directional component is blocked by some part of the face and thus only the ambient component should contribute to the shadow's intensity. To model cast shadows in the shading, we represent the modified shading s' as:

$$\begin{aligned} s'(\mathbf{x}) &= i_a + \mathbf{M}_{shadow}(\mathbf{x})i_d \langle \mathbf{n}, \omega_d \rangle \\ &= \mathbf{M}_{shadow}(\mathbf{x})s(\mathbf{x}) + (1 - \mathbf{M}_{shadow}(\mathbf{x}))i_a. \end{aligned}$$
(4)

Our reformulation of the rendering equation is thus,

$$L_o(\mathbf{x}, \omega_o) = a(\mathbf{x})s'(\mathbf{x}),$$
(5)

which overcomes the limitations of the diffuse shading s and models cast shadows.

Unlike prior face relighting methods that model cast shadows [10, 29, 30], our formulation ensures geometrically consistent cast shadows by computing $\mathbf{M}_{shadow}$ directly using the estimated face geometry. We discuss this in Sec. 3.3.

3.2. Architecture

Given a single image $\mathbf{I}_t$ as input, our model estimates the intrinsic components: depth map $\mathbf{D}_p$ (geometry), albedo $\mathbf{A}_p$, the lighting direction ω_p, and the ambient lighting intensity $i_{a,p}$. Our architecture is largely adopted from the hourglass network used by DPR [54], but we replicate the decoder and form two branches to estimate $\mathbf{D}_p$ and $\mathbf{A}_p$ respectively. ω_p and $i_{a,p}$ are estimated using a multilayer perceptron (MLP) following the encoder. The surface normals $\mathbf{N}_p$ are then computed from $\mathbf{D}_p$, and the shading $\mathbf{S}_p$ is computed via Eqn. 4. We will discuss how $\mathbf{M}_{shadow}$ is computed in the next section. The final rendered image $\mathbf{I}_p$ is then generated following Eqn. 5 as:

$$\mathbf{I}_p = \mathbf{A}_p \mathbf{S}_p.$$
(6)

During training, we render $\mathbf{I}_p$ using ω_p, the estimated lighting direction of the input image $\mathbf{I}_t$. We supervise the intrinsic component estimation by enforcing that $\mathbf{I}_p$ reproduces the input image $\mathbf{I}_t$. During inference, our model instead accepts a target lighting direction ω_t as input, which allows us to perform relighting. We compute $\mathbf{M}_{shadow}$ using $\mathbf{D}_p$ and ω_t, which allows us to generate relit images with geometrically consistent cast shadows from any lighting direction. We illustrate our overall model architecture in Fig. 2.

Figure 3. **Shadow Mask Estimation**. We generate $\mathbf{M}_{shadow}$ using $\mathbf{D}_p$, ω_t, and the principles of ray tracing. For every point $\mathbf{x}_i \in \mathbf{D}_p$, we sample points from $\mathbf{D}_p$ along the direction $\overrightarrow{\mathbf{x}_i \omega_t}$. If there exists a sampled point whose distance to $\overrightarrow{\mathbf{x}_i \omega_t}$ is close to 0, then $\overrightarrow{\mathbf{x}_i \omega_t}$ intersects a surface (*e.g.* the nose) along its path and $\mathbf{x}_i$ is under cast shadow. If there is no such point among the sampled points, then $\mathbf{x}_i$ is not under cast shadow. We show 2 points $\mathbf{x}_1$ and $\mathbf{x}_2$, marked as green and red asterisks respectively. Among the sampled points for $\mathbf{x}_1$ (green points), there exists a point (marked by a yellow arrow) that intersects a surface (the nose) and thus $\mathbf{x}_1$ is under a cast shadow. For $\mathbf{x}_2$, none of the sampled points (red points) intersect a surface, so $\mathbf{x}_2$ is not under a cast shadow.

3.3. Shading Estimation

One of our key contributions is producing geometrically consistent cast shadows using shadow mask $\mathbf{M}_{shadow}$. It is generated directly from the estimated geometry and used to produce shading $\mathbf{S}_p$ that models cast shadows. We now discuss the motivation and formulation.

Ray Tracing To motivate our algorithm for generating $\mathbf{M}_{shadow}$, we first discuss the traditional ray tracing algorithm. For every point $\mathbf{x}_i$ on the 3D object, the ray tracing algorithm casts a *shadow ray* towards the light source [1]. If the shadow ray intersects with a surface along its path, then $\mathbf{x}_i$ is under cast shadow. In our setting, the shadow rays determine which points are under cast shadow based on whether the ray intersects with some parts of the estimated 3D face, such as the nose.

Shadow Mask Estimation Our method incorporates the principles of ray tracing to generate $\mathbf{M}_{shadow}$ using the estimated depth map $\mathbf{D}_p$ and the target lighting direction ω_t (see Fig. 3). Each pixel in $\mathbf{D}_p$ corresponds to a 3D point $\mathbf{x}_i$, and we represent ω_t as a unit vector in 3D space. We can thus represent the shadow ray for point $\mathbf{x}_i$ as an order pair $(\mathbf{x}_i, \omega_t)$. To determine whether it intersects the surface, we sample $m = 160$ points $\mathbf{x}_{s1}, \mathbf{x}_{s2}, ..., \mathbf{x}_{sm}$ along its direction from $\mathbf{D}_p$ at regular intervals. We then provide a differentiable visibility function based on the following observation: if there exists a sampled point whose distance to the ray is close to 0, then the current ray or a nearby ray would intersect with the surface and the point $\mathbf{x}_i$ is under cast shadow. Conversely, if none of the sampled points have a distance close to 0 to the ray, the shadow ray does not intersect the surface and $\mathbf{x}_i$ is not under cast shadow. We therefore compute the minimum distance d_{min} between

the sampled points and the ray by

$$d_{min} = \min_{j \in [1,m]} |\overrightarrow{\mathbf{x}_i \mathbf{x}_{sj}} \times \omega_t|, \qquad (7)$$

where $\times$ is the cross product. If d_{min} is close to 0, we set the corresponding shadow mask value $\mathbf{M}_{shadow}(\mathbf{x}_i)$ to be close to 0, indicating $\mathbf{x}_i$ is under a cast shadow. Otherwise, it should be close to 1. To achieve this while ensuring that computing $\mathbf{M}_{shadow}(\mathbf{x}_i)$ is a differentiable operation, we define $\mathbf{M}_{shadow}$ as a Sigmoid function of d_{min}:

$$\mathbf{M}_{shadow}(\mathbf{x}_i) = \frac{-4e^{-d_{min}}}{(1 + e^{-d_{min}})^2} + 1. \qquad (8)$$

We apply our algorithm to all points $\mathbf{x}_i$ in depth map $\mathbf{D}_p$ to generate the shadow mask $\mathbf{M}_{shadow}$, which indicates where cast shadows lie on the face. Since we use $\mathbf{D}_p$ to compute $\mathbf{M}_{shadow}$, we directly leverage the 3D geometry of the face to synthesize our cast shadows, ensuring that they are geometrically consistent with respect to the face.

3.4. Training Losses

We utilize multiple loss functions to supervise the intrinsic decomposition. To supervise the depth estimation, we define $\mathcal{L}_{depth} = \frac{\Sigma \, \mathbf{M}_{depth} \|\mathbf{D}_p - \mathbf{D}_t\|_1}{\Sigma \, \mathbf{M}_{depth}}$, where $\mathbf{D}_t$ is the groundtruth depth map, and mask $\mathbf{M}_{depth}$ defines where we have depth supervision in the image. The groundtruth depth is obtained using the method of Bai *et al.* [4] to first estimate the face mesh, and subsequently apply z-buffering.

To supervise the albedo estimation, we define $\mathcal{L}_{albedo} = \frac{\Sigma \, \mathbf{M}_{face} \|\mathbf{A}_p - \mathbf{A}_t\|_1}{\Sigma \, \mathbf{M}_{face}}$, where $\mathbf{A}_t$ is the groundtruth albedo, and $\mathbf{M}_{face}$ is the full face mask. We generate our groundtruth albedo using SfSNet [35]. Since SfSNet's estimated albedo does not generalize perfectly to our training data, we only apply $\mathcal{L}_{albedo}$ in grayscale to give our model more freedom in estimating the RGB albedo.

To supervise our lighting estimation, we define two additional losses: $\mathcal{L}_{ambient}$ and $\mathcal{L}_{light}$. We define $\mathcal{L}_{ambient} = \|i_{a,p} - i_{a,t}\|_1$, where $i_{a,t}$ is the groundtruth ambient intensity. Since determining the groundtruth ambient intensity of an image is challenging, we set $i_{a,t}$ to be the same value for all training images. We also define $\mathcal{L}_{light} = 1 - \langle \omega_p, \omega_t \rangle$, where $\langle \omega_p, \omega_t \rangle$ is the inner product between the predicted and the groundtruth lighting direction ω_t. We obtain ω_t using SfSNet, and convert the estimated SH representation into a dominant lighting direction.

To ensure that the estimated intrinsic components as a whole represent a plausible decomposition, we define a reconstruction loss $\mathcal{L}_{recon} = \frac{\Sigma \, \mathbf{M}_{face} \|\mathbf{I}_p - \mathbf{I}_t\|_2^2}{\Sigma \, \mathbf{M}_{face}}$ between the rendered image $\mathbf{I}_p$ and the input $\mathbf{I}_t$. Finally, to improve the perceptual quality, we employ a PatchGAN [8] discriminator that operates on 70×70 patches. We define our adversarial loss as $\mathcal{L}_{GAN}$ and treat our rendered images as the

Method	SfSNet [35]	DPR [54]	SIPR [43]	Nestmeyer [29]	Hou [10]	Proposed
LPIPS	0.5222±0.0743	0.2644±0.0808	0.2764±0.0736	0.3795±0.2294	0.2013±0.0676	**0.1622±0.0490**
MSE	0.0961±0.0495	0.0852±0.0515	0.0166±0.0107	0.0588±0.0538	0.0303±0.0162	**0.0150±0.0112**
DSSIM	0.2918±0.0375	0.1599±0.0558	0.1539±0.0452	0.2226±0.1356	0.1186±0.0388	**0.0990±0.0381**

Table 2. **Relighting Evaluation on Multi-PIE Images with Target Lighting (mean± standard deviation)**. We compare our model against methods that accept a single image and a target lighting. Our method achieves the best performance across all metrics (bold).

(a) Input Image (b) Target Image (c) SfSNet [35] (d) DPR [54] (e) SIPR [43] (f) Nestmeyer [29] (g) Hou [10] (h) Proposed

Figure 4. **Qualitative Relighting Performance on Multi-PIE (Target Lightings)**. Each method performs relighting given a single input image and a target lighting. Our method's cast shadows much more closely match the target image compared to Hou *et al.* [10] and Nestmeyer *et al.* [29], two baselines modeling cast shadows. SIPR [43], DPR [54], and SfSNet [35] are unable to produce cast shadows.

Method	Shih [38]	Shu [39]	Hou [10]	Proposed
LPIPS	0.2446±0.0750	0.1548±0.0482	**0.1499±0.0444**	0.1580±0.0485
MSE	0.0529±0.0361	0.0188±0.0177	0.0192±0.0119	**0.0176±0.0127**
DSSIM	0.1998±0.0827	0.0994±0.0415	**0.0942±0.0360**	0.0962±0.0381

Table 3. **Lighting Transfer Evaluation on Multi-PIE (mean± standard deviation)**. Each input image is assigned a random reference image. The reference image is a different subject and a different lighting from the input image.

fake distribution and the input images as the real distribution. We also utilize a DSSIM loss to further improve the perceptual quality similar to Hou *et al.* [10] and Nestmeyer *et al.* [29] defined as $\mathcal{L}_{DSSIM} = \frac{(1-SSIM(\mathbf{I}_p, \mathbf{I}_t))}{2}$.

Our final loss function is thus defined as:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{depth} + \lambda_2 \mathcal{L}_{albedo} + \lambda_3 \mathcal{L}_{ambient} + \lambda_4 \mathcal{L}_{light} + \lambda_5 \mathcal{L}_{recon} + \lambda_6 \mathcal{L}_{GAN} + \lambda_7 \mathcal{L}_{DSSIM}, \quad (9)$$

where λ_i are the weights for each loss function.

Implementation Details We train using PyTorch [31] for 100 epochs with the Adam Optimizer [15] and a learning rate of 0.0001 on one GeForce RTX 2080 Ti GPU. In training, we set $\lambda_1 = 1$, $\lambda_2 = 5$, $\lambda_3 = 2.5$, $\lambda_4 = 1$, $\lambda_5 = 20$, $\lambda_6 = 0.01$, $\lambda_7 = 8$, the groundtruth ambient intensity to $i_{a,t} = 0.5$, and the directional intensity to $i_d = 0.5$.

4. Experiments

Since our training objective is to minimize the difference between the rendered image $\mathbf{I}_p$ and the input image $\mathbf{I}_t$, we are ultimately free to use any face dataset with lighting variation. We train our model on the CelebA-HQ dataset [12], containing 30, 000 in-the-wild face images from the CelebA dataset [25]. Following the testing protocol of Hou *et al.* [10], we evaluate our relighting performance quantitatively

on the Multi-PIE [9] dataset, which contains 18 images per subject each with a unique directional light.

4.1. Quantitative Evaluations

Multi-PIE Evaluation with Target Lightings As our model relights using a target lighting ω_t, we randomly select 1 of the 18 directional lights in Multi-PIE as ω_t for each subject and also randomly select 1 of the 18 images as the input image. Since Multi-PIE captures each subject under every lighting condition, we have the relighting groundtruth and can quantitatively compare our relit image with the groundtruth image under the target lighting ω_t. We thus compare with prior methods that accept a target lighting as input [10, 29, 35, 43, 54]. We evaluate using 3 metrics: MSE, DSSIM [29], and LPIPS [52]. Both DSSIM and LPIPS are metrics that are highly correlated with perceptual quality [29, 52]. DSSIM = $\frac{1}{2}(1 - \text{SSIM})$ is an error metric defined based on SSIM [48]. During evaluation, we compute the metrics for all methods only in the face region indicated by $\mathbf{M}_{depth}$. This ensures a fair comparison with our method, since $\mathbf{M}_{depth}$ represents where our images receive depth supervision from Bai *et al.* [4], which does not estimate the depth outside of the face. Our method is thus intended to relight the face region, not the hair or background.

We report the results of our evaluation in Tab. 2 and note that our model achieves the best performance across all 3 metrics, with the largest gains in DSSIM and LPIPS, indicating that the perceptual quality of our relit images significantly improves over prior work. This is largely due to our differentiable hard shadow modeling that generates more appropriately shaped hard shadows. We also explicitly model both ambient and directional light which helps to produce more well-balanced colors in our relit images

(a) Input Image   (b) Reference   (c) Target Image   (d) Shih [38]   (e) Shu [39]   (f) Hou [10]   (g) Proposed

Figure 5. **Qualitative Relighting Performance on Multi-PIE (Lighting Transfer)**. The target lighting is estimated from the reference image and used to relight the input image. Notice that our model estimates the correct target lightings from the reference images whereas Shih *et al.* [38] transfers the wrong lightings. Furthermore, neither Shih *et al.* [38] nor Shu *et al.* [39] can produce cast shadows through lighting transfer, whereas our model can. Hou *et al.* [10] fails to transfer an appropriate cast shadow for the subject in the top row and the cast shadow for the subject in the second row is noticeably worse than ours in terms of shape and boundary.

than prior work, where the shadows may be too dark or the illuminated face may be too bright.

Multi-PIE Evaluation Using Lighting Transfer Some methods require both an input and a reference image and relight by transferring the style of the reference to the input, known as lighting transfer. To evaluate our lighting transfer performance, we sample a random lighting for each Multi-PIE subject to serve as the input and a random reference image from the entire dataset. The reference image is a different subject from the input and under a different lighting. For lighting transfer, we first feed the reference image to our model to estimate the target lighting direction ω_t and the ambient intensity $i_{a,p}$. We then pass the input image along with ω_t and $i_{a,p}$ to our model to generate the relit image. The groundtruth target image is readily available in Multi-PIE. We compare with Shih *et al.* [38], Shu *et al.* [39], and Hou *et al.* [10] and report the results in Tab. 3. We achieve the best performance in MSE and comparable performance to Hou *et al.* in terms of DSSIM and LPIPS. We believe that a large reason why the performance is lower is the imperfect lighting supervision from SfSNet [35], which limits our model's ability to estimate the correct lighting from the reference image and lowers the relighting performance.

4.2. Qualitative Evaluations

Multi-PIE Results We show qualitative relighting results of Multi-PIE [9] subjects for target lightings in Fig. 4 and for lighting transfer in Fig. 5. When using target lightings, our model produces cast shadows that much more closely match the shape of the shadows in target images compared to prior work. This is due to our shadow mask estimation that incorporates the face geometry to synthesize cast shadows, improving their geometric consistency. Hou *et al.* [10] and Nestmeyer *et al.* [29] also model cast shadows, but neither synthesize cast shadows directly from the face geometry and instead regress them from CNNs. Thus, they have no guarantee that their cast shadows correctly match the face geometry, as seen in Fig. 4. SIPR [43], DPR [54], and

a) Hou [10]   b) Proposed   c) Target Lighting

Figure 6. **Relighting Error Maps**. We show the average L_1 error map between our relit images and the groundtruth test images of Multi-PIE for each lighting and compare with Hou *et al.* [10]. As shown in b), we have lower error in the shadowed regions, including shadows cast around the nose. Hou *et al.* has higher errors around the cast shadows, demonstrating that our method produces more geometrically consistent shadows across all subjects.

SfSNet [35] primarily model diffuse lightings and generally cannot produce cast shadows. When performing lighting transfer, we notice that our model can accurately estimate the target lighting from the reference image and is superior to the baselines in transferring over cast shadows (see Fig. 5). Shih *et al.* [38] and Shu *et al.* [39] largely fail to transfer over cast shadows and Hou *et al.* [10] produces cast shadows that do not match the shape of the groundtruth.

FFHQ Results We evaluate our performance on in-the-wild faces from the FFHQ [13] dataset. We show in Fig. 7 that we produce more geometrically consistent shadows than prior work across several subjects and lightings. We also show in Fig. 8 that our cast shadows are geometrically consistent as we rotate the target lightings around the face. Compared to Hou *et al.* [10], our cast shadows have much more plausible shapes and shadow boundaries.

4.3. Ablations and Additional Experiments

Reconstruction Error Analysis To better understand the distribution of reconstruction errors during relighting, Fig. 6

Figure 7. **Qualitative Relighting Performance on FFHQ**. Across multiple in-the-wild subjects and target lightings, our model produces more geometrically consistent cast shadows than prior methods while achieving noticeably better visual quality. Best viewed if enlarged.

Figure 8. **Comparison of Geometric Consistency of Cast Shadows**. We compare the geometric consistency of our cast shadows across 7 target lightings with Hou *et al.* [10], another face relighting method that models cast shadows. Notice that our model's cast shadows shown in row b) are more plausible in terms of shape and shadow boundaries than the cast shadows of Hou *et al.*, shown in row a).

visualizes the average L_1 error map between our relit images and the groundtruth target images in our Multi-PIE test set. We generate an error map for each target lighting separately and compute the average across all test subjects with that target lighting. We compare our error maps with Hou *et al.* [10], the SoTA face relighting method, and notice that our error maps have much lower error in the shadowed face regions, including the shadows cast around the nose. This further demonstrates that our method produces more geometrically consistent shadows across all test subjects.

Geometry Error Analysis One benefit of modeling hard shadows differentiably is that the end-to-end training may improve the intrinsic components, such as geometry, in face regions that cast hard shadows. To demonstrate this, we compare our surface normal errors on the Multi-PIE test images with two baselines: SfSNet [35], an intrinsic decomposition method with a diffuse SH lighting model, and DFNRMVS [4], which provides our geometry supervision. We choose surface normal error as the metric since

the rendering equation uses surface normals, rather than the depth, to compute the shading. Although Multi-PIE lacks groundtruth 3D shapes, we use DFNRMVS [4] to estimate face meshes given 3 multi-view faces per subject as input, from which we compute the groundtruth surface normals. A dataset with large lighting variation and 3D groundtruth shapes is still lacking, partially due to the sensitivity of 3D scanners to illumination. We train on Multi-PIE subjects 1-250 and test on subjects 251-346. As for the geometry supervision in training, we use the face meshes from DFN-RMVS provided only a single frontal image as input, which produces lower quality shapes than 3 views.

As shown in Tab. 4, across all test images and lightings, our model achieves the lowest average angular error in surface normal estimation. Improving over DFNRMVS shows that our model is not upper bounded by the quality of our shape supervision. Our end-to-end training incorporating differentiable shadow modeling can yield further improvements to the geometry. Improving over SfSNet also

Figure 9. **Per Lighting Surface Normal Error on Multi-PIE**. We compute the surface normal errors (degrees) across all test images of the same lighting. Our proposed model (P) achieves lower errors across all lightings than DFNRMVS [4] (D) and SfSNet [35] (S). We record the improvement percentage of P over the best baseline, where higher percentages are highlighted in brighter green. A reference image and its lighting direction are provided for each lighting. P improves the normals the most for lightings with hard shadows (*e.g.* lightings 4, 6, 8, 9, 10, 15, 17, and 18), which highlights the benefit of our hard shadow modeling in improving the face geometry.

Method	Surface Normal Angular Error (Degrees)
SfSNet [35]	14.2796 ± 2.1442
DFNRMVS [4]	12.4505 ± 2.3939
Proposed	$\mathbf{11.0672 \pm 1.9489}$

Table 4. **Surface Normal Errors on Multi-PIE (mean± standard deviation)**. We compare with SfSNet and DFNRMVS. Our model produces more accurate surface normals than SfSNet, which assumes a diffuse SH lighting model, and DFNRMVS, our shape supervision, which shows the ability of our differentiable hard shadow modeling in improving the geometry.

Figure 10. **Surface Normal Improvement**. We visualize our surface normal improvement over the best baseline (DFNRMVS [4]) for test images in 4 lightings with hard shadows. The first row shows a reference image for each lighting. Notice the large improvements along and near the nose bridge and the face boundary, which cast hard shadows. This shows the contribution of our differentiable hard shadow modeling in improving the geometry.

highlights the contribution of our shadow modeling, as SfSNet uses a diffuse SH lighting model and thus has no incentive to improve the geometry in regions producing hard shadows. Fig. 9 further demonstrates that the largest improvements are achieved for lightings with significant hard shadows. Fig. 10 visualizes that our model improves the geometry of the nose and especially the nose bridge significantly, which is where hard shadows are cast from. It also improves near the boundary of the face, which also tends to produce hard shadows. This demonstrates that our differentiable hard shadow modeling improves the geometry estimation, especially in regions that cast hard shadows.

5. Conclusion

We have proposed a novel face relighting method that produces geometrically consistent hard shadows. Unlike prior work, our approach is the first to directly synthesize cast shadows from the geometry, which improves the shadow's shape and boundary. We have shown on the Multi-PIE and FFHQ datasets that our method achieves state-of-the-art face relighting performance quantitatively and qualitatively under directional lighting. We have also shown that our differentiable hard shadow modeling improves the geometry, especially around the nose, compared to prior work that assumes diffuse shading. We hope that our work will motivate future physics-driven relighting methods, and provide insights for handling hard shadows.

Limitations Since we use SfSNet's [35] imperfect estimated lighting as supervision, our model accumulates more error during lighting transfer. The RGB albedo from SfSNet also does not generalize well to our training data, limiting the quality of our estimated albedo. Training on a dataset where we know the groundtruth lightings and could compute the albedo from photometric stereo similar to [29] would improve our model's performance. In addition, although CelebA-HQ [12] contains some images under directional lighting, it primarily contains images under diffuse lighting. Our model would benefit from a publicly available in-the-wild dataset with primarily directional lights.

Broader Impact Creating deepfakes or affecting surveillance by adding shadows are major concerns. We acknowledge these risks but argue that our model only adds self shadows, which are generally limited in size and primarily around the nose. The user cannot freely manipulate the image or add shadows to any location they desire, which limits malicious use cases. Moreover, our method can synthesize images with self shadows for training, which can improve the robustness of face methods to self-shadowed images.

References

[1] Arthur Appel. Some techniques for shading machine renderings of solids. In *AFIPS*, 1968. 4

[2] Yousef Atoum, Mao Ye, Liu Ren, Ying Tai, and Xiaoming Liu. Color-wise attention network for low-light image enhancement. In *CVPRW*, 2020. 2

[3] Mallikarjun B R, Ayush Tewari, Tae-Hyun Oh, Tim Weyrich, Bernd Bickel, Hans-Peter Seidel, Hanspeter Pfister, Wojciech Matusik, Mohamed Elgharib, and Christian Theobalt. Monocular reconstruction of neural face reflectance fields. In *CVPR*, 2021. 2

[4] Ziqian Bai, Zhaopeng Cui, Jamal Ahmed Rahim, Xiaoming Liu, and Ping Tan. Deep facial non-rigid multi-view stereo. In *CVPR*, 2020. 4, 5, 7, 8

[5] Jonathan T Barron and Jitendra Malik. Shape, illumination, and reflectance from shading. *PAMI*, 2015. 2

[6] Bernhard Egger, Sandro Schonborn, Andreas Schneider, Adam Kortylewski, Andreas Morel-Forseter, Clemens Blumer, and Thomas Vetter. Occlusion-aware 3D morphable models and an illumination prior for face image analysis. *IJCV*, 2018. 2

[7] Kyle Genova, Forrester Cole, Aaron Maschinot, Aaron Sarna, Daniel Vlasic, and William T. Freeman. Unsupervised training for 3D morphable model regression. In *CVPR*, 2018. 2

[8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *NeurIPS*, 2014. 4

[9] Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, and Simon Baker. Multi-PIE. *Image and Vision Computing*, 2010. 5, 6, 11, 12

[10] Andrew Hou, Ze Zhang, Michel Sarkis, Ning Bi, Yiying Tong, and Xiaoming Liu. Towards high fidelity face relighting with realistic shadows. In *CVPR*, 2021. 1, 2, 3, 5, 6, 7, 11, 12

[11] James T. Kajiya. The rendering equation. In *SIGGRAPH*, 1986. 2

[12] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of GANs for improved quality, stability, and variation. In *ICLR*, 2018. 5, 8, 12

[13] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *CVPR*, 2019. 6, 11, 12

[14] Hiroharu Kato, Yoshitaka Ushiku, and Tatsuya Harada. Neural 3D mesh renderer. In *CVPR*, 2018. 2

[15] Diederik Kingma and Jimmy Ba. Adam. A method for stochastic optimization. In *ICLR*, 2014. 5

[16] Ha Le and Ioannis Kakadiaris. Illumination-invariant face recognition with deep relit face images. In *WACV*, 2019. 1, 2

[17] Gun-Hee Lee and Seong-Whan Lee. Uncertainty-aware mesh decoder for high fidelity 3D face reconstruction. In *CVPR*, 2020. 2

[18] Jinho Lee, Raghu Machiraju, Baback Moghaddam, and Hanspeter Pfister. Estimation of 3D faces and illumination from single photographs using a bilinear illumination model. In *EGSR*, 2005. 2

[19] Chen Li, Kun Zhou, and Stephen Lin. Intrinsic face image decomposition with human face priors. In *ECCV*, 2014. 2

[20] Tzu-Mao Li, Miika Aittala, Frédo Durand, and Jaakko Lehtinen. Differentiable Monte Carlo ray tracing through edge sampling. In *SIGGRAPH Asia*, 2018. 2

[21] Yijun Li, Ming-Yu Liu, Xueting Li, Ming-Hsuan Yang, and Jan Kautz. A closed-form solution to photorealistic image stylization. In *ECCV*, 2018. 2

[22] Jiangke Lin, Yi Yuan, Tianjia Shao, and Kun Zhou. Towards high-fidelity 3D face reconstruction from in-the-wild images using graph convolutional networks. In *CVPR*, 2020. 2

[23] Feng Liu, Luan Tran, and Xiaoming Liu. Fully understanding generic objects: Modeling, segmentation, and reconstruction. In *CVPR*, 2021. 2

[24] Shichen Liu, Tianye Li, Weikai Chen, and Hao Li. Soft rasterizer: A differentiable renderer for image-based 3D reasoning. In *ICCV*, 2019. 2

[25] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *ICCV*, 2015. 5

[26] Matthew M. Loper and Michael J. Black. OpenDR: An approximate differentiable renderer. In *ECCV*, 2014. 2

[27] Fujun Luan, Sylvain Paris, Eli Shechtman, and Kavita Bala. Deep photo style transfer. In *CVPR*, 2017. 2

[28] Linjie Lyu, Marc Habermann, Lingjie Liu, Mallikarjun B R, Ayush Tewari, and Christian Theobalt. Efficient and differentiable shadow computation for inverse problems. In *ICCV*, 2021. 2

[29] Thomas Nestmeyer, Jean-Francois Lalonde, Iain Matthews, and Andreas Lehrmann. Learning physics-guided face relighting under directional light. In *CVPR*, 2020. 1, 2, 3, 5, 6, 7, 8, 11

[30] Rohit Pandey, Sergio Orts-Escolano, Chloe LeGendre, Christian Haene, Sofien Bouaziz, Christoph Rhemann, Paul Debevec, and Sean Fanello. Total relighting: Learning to relight portraits for background replacement. In *SIGGRAPH*, 2021. 1, 2, 3

[31] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in PyTorch. In *NeurIPSW*, 2017. 5

[32] Pieter Peers, Naoki Tamura, Wojciech Matusik, and Paul Debevec. Post-production facial performance relighting using reflectance transfer. In *SIGGRAPH*, 2007. 2

[33] Laiyun Qing, Shiguang Shan, and Xilin Chen. Face relighting for face recognition under generic illumination. In *ICASSP*, 2004. 1

[34] Mallikarjun B R, Ayush Tewari, Abdallah Dib, Tim Weyrich, Bernd Bickel, Hans-Peter Seidel, Hanspeter Pfister, Wojciech Matusik, Louis Chevallier, Mohamed Elgharib, and Christian Theobalt. Photoapp: Photorealistic appearance editing of head portraits. *TOG*, 2021. 2

[35] Soumyadip Sengupta, Angjoo Kanazawa, Carlos D. Castillo, and David W. Jacobs. SfSNet: Learning shape, refectance and illuminance of faces in the wild. In *CVPR*, 2018. 1, 2, 4, 5, 6, 7, 8, 11, 12

[36] Davoud Shahlaei and Volker Blanz. Realistic inverse lighting from a single 2D image of a face, taken under unknown and complex lighting. In *FG*, 2015. 2

[37] Amnon Shashua and Tammy Riklin-Raviv. The quotient image: Class-based re-rendering and recognition with varying illuminations. *PAMI*, 2001. 2

[38] YiChang Shih, Sylvain Paris, Connelly Barnes, William T. Freeman, and Fredo Durand. Style transfer for headshot portraits. In *SIGGRAPH*, 2014. 2, 5, 6

[39] Zhixin Shu, Sunil Hadap, Eli Shechtman, Kalyan Sunkavalli, Sylvain Paris, and Dimitris Samaras. Portrait lighting transfer using a mass transport approach. *TOG*, 2017. 2, 5, 6

[40] Zhixin Shu, Ersin Yumer, Sunil Hadap, Kalyan Sunkavalli, Eli Shechtman, and Dimitris Samaras. Neural face editing with intrinsic image disentangling. In *CVPR*, 2017. 2

[41] Pratul P. Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T. Barron. NeRV: Neural reflectance and visibility fields for relighting and view synthesis. In *CVPR*, 2021. 2

[42] Arne Stoschek. Image-based re-rendering of faces for continuous pose and illumination directions. In *CVPR*, 2000. 2

[43] Tiancheng Sun, Jonathan T Barron, Yun-Ta Tsai, Zexiang Xu, Xueming Yu, Graham Fyffe, Christoph Rhemann, Jay Busch, Paul Debevec, and Ravi Ramamoorthi. Single image portrait relighting. In *SIGGRAPH*, 2019. 1, 2, 5, 6, 7

[44] Luan Tran, Feng Liu, and Xiaoming Liu. Towards high-fidelity nonlinear 3D face morphable model. In *CVPR*, 2019. 2

[45] Luan Tran and Xiaoming Liu. Nonlinear 3D face morphable model. In *CVPR*, 2018. 2

[46] Luan Tran and Xiaoming Liu. On learning 3D face morphable model from in-the-wild images. *PAMI*, 2019. 2

[47] Yang Wang, Lei Zhang, Zicheng Liu, Gang Hua, Zhen Wen, Zhengyou Zhang, and Dimitris Samaras. Face relighting from a single image under arbitrary unknown lighting conditions. *PAMI*, 2009. 2

[48] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *TIP*, 2004. 5

[49] Zhibo Wang, Xin Yu, Ming Lu, Quan Wang, Chen Qian, and Feng Xu. Single image portrait relighting via explicit multiple reflectance channel modeling. In *SIGGRAPH Asia*, 2020. 2

[50] Zhen Wen, Zicheng Liu, and Tomas Huang. Face relighting with radiance environment maps. In *CVPR*, 2003. 2

[51] Shuco Yamaguchi, Shunsuke Saito, Koki Nagano, Yajie Zhao, Weikai Chen, Kyle Olszewski, Shigeo Morishima, and Hao Li. High-fidelity facial reflectance and geometry inference from an unconstrained image. In *SIGGRAPH*, 2018. 2

[52] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 5

[53] Xuaner Zhang, Jonathan T. Barron, Yun-Ta Tsai, Rohit Pandey, Xiuming Zhang, Ren Ng, and David E. Jacobs. Portrait shadow manipulation. *TOG*, 2020. 2

[54] Hao Zhou, Sunil Hadap, Kalyan Sunkavalli, and David W Jacobs. Deep single-image portrait relighting. In *ICCV*, 2019. 1, 2, 3, 5, 6, 7, 11

Face Relighting with Geometrically Consistent Shadows
(Supplementary Materials)

Method	MSE	DSSIM	LPIPS
DPR [54]	0.0171	0.0796	0.1286
Proposed	**0.0080**	**0.0562**	**0.1268**

Table 1. **Quantitative Evaluation for General (Diffuse) Relighting** We outperform DPR quantitatively across all metrics in diffuse relighting on the Multi-PIE dataset.

Input   DPR [54]   Proposed   Target

Figure 1. **General (Diffuse) Relighting**. We outperform DPR qualitatively in diffuse relighting on the Multi-PIE dataset, where each input image is relit by averaging the predictions of 3 randomly selected target lightings. The groundtruth is the average of the 3 groundtruth Multi-PIE images.

1. Diffuse Relighting Evaluation

To compare with DPR [54] on more general, diffuse lightings, we follow their protocol and generate diffuse lighting groundtruth by averaging 3 random directional lighting images per Multi-PIE [9] subject. For both DPR and our method, we feed each of the 3 target lightings separately and average the predictions to generate the final relit image. We outperform DPR in general relighting both quantitatively and qualitatively (Tab. 1 and Fig. 1).

2. Geometric Consistency Comparison with Nestmeyer *et al.* [29]

To compare with the SoTA relighting method Hou *et al.* [10], we used the average L_1 error for each Multi-PIE lighting's test subjects to verify that our model was improving primarily around the hard shadow region in Fig. 6 of the main paper. We show the same error map to compare with Nestmeyer *et al.* [29] in Fig. 2. Our method has particularly low error in the hard shadow region (nose and cheek), whereas Nestmeyer *et al.* has high error in and around the shadow, especially for the first row's lighting. Our method

Nestmeyer [29]   Proposed   Target Lighting

Figure 2. **Error Maps**. We visualize the average L_1 error for each Multi-PIE lighting's test subjects. Our method has significantly lower error around the hard shadow regions (nose and cheek) compared to Nestmeyer *et al.* [29], which demonstrates that our method produces more geometrically consistent hard shadows.

thus produces more geometrically consistent hard shadows.

3. Albedo Comparison

Our albedo supervision from SfSNet [35] is far from perfect, as shown in Fig. 3, which is why we define the albedo loss in grayscale and not RGB. We adopt this supervision primarily because albedo supervision has limited options for single image in-the-wild datasets besides PCA, which often does not preserve facial details well. However, our model's estimated albedo clearly improves over SfSNet.

4. Comprehensive FFHQ Relighting Results

We strongly believe in diversity and the representation of all groups in the computer vision community. We therefore show a wide variety of relighting results with diversity and inclusion in mind. Our results cover as many racial groups as possible, as well as other factors such as different ages, genders, poses, expressions, subjects with facial hair, and the presence of glasses (See Fig. 4). We also increased the lighting diversity to demonstrate that our model can handle many different desired illuminations.

5. FFHQ Relighting Video

We include a video with 4 FFHQ [13] subjects where we rotate the light around the face, move the light horizontally, and move the light vertically. From left to right, we visualize the target lighting, the relighting results of Hou

| Input Image | SfSNet [35] | Proposed |

Figure 3. **Albedo Comparison**. Our method is able to produce high quality albedo despite the imperfect supervision from SfSNet [35] by keeping the albedo loss $\mathcal{L}_{albedo}$ in grayscale, which gives our model more freedom in the RGB space.

et al. [10], and our proposed method's relighting results. Our video demonstrates our high relighting quality as well as the geometric consistency of our shadows across many lightings. Compared to [10], it is clear that the shape of our shadows is superior, especially when comparing the first subject. We also modify the tone of the image significantly less, while [10] seems to frequently produce overly dark shadows. The video can be viewed here.

6. Licenses for Face Related Datasets

Although we don't collect any face data ourselves in this work, we do make use of existing face datasets, including Multi-PIE [9], FFHQ [13], and CelebA-HQ [12]. The Multi-PIE database was collected at Carnegie Mellon University, where all subjects agreed that their data would be used for research purposes. We only use the database internally for our work and primarily for evaluation. FFHQ consists of images published on Flickr, which are all under multiple licenses that allow free use, adaptation, and redistribution for noncommercial purposes. The creators also provide a way to remove an individual's photo from the dataset if they so desire. CelebA-HQ consists entirely of images collected from the internet. Although there is no associated IRB approval, the authors assert in the dataset agreement that the dataset is only to be used for noncommercial research purposes, which we strictly adhere to. Users must also agree not to sell, reproduce, or exploit any of the data and can only make copies of the data within their own organization, which we also adhere to.

Figure 4. **Comprehensive and Diverse Relighting Performance on FFHQ**. Every two rows (*e.g.* c, d) shows the input image in the first row and our relighting results in the second row. We demonstrate our relighting performance on a wide variety of racial groups, genders, ages, expressions, and poses and also include subjects with facial hair and glasses. We find that our model is able to generalize to a wide range of subjects across many different lightings. Best viewed if enlarged.