Holo-Relighting: Controllable Volumetric Portrait Relighting from a Single Image

Yiqun Mei[1] Yu Zeng[1†] He Zhang[2†] Zhixin Shu[2†] Xuaner Zhang[2] Sai Bi[2]
Jianming Zhang[2] HyunJoon Jung[2] Vishal M. Patel[1]
[1]Johns Hopkins University [2]Adobe Inc.

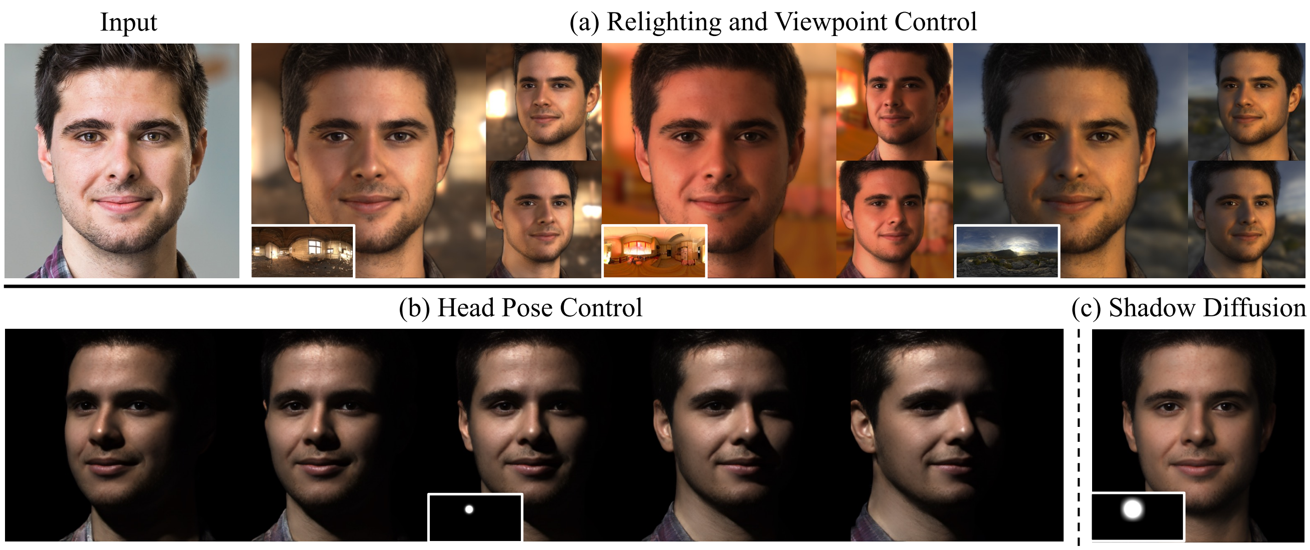

Figure 1. *Holo-Relighting* performs volumetric relighting on a single input portrait image, allowing users to individually control (1) lighting effect via an environment map, (2) camera viewpoint and (3) head pose. It is highly expressive and can render complex illumination effects on in-the-wild human faces with accurate view consistency (a). Controls are well disentangled to produce a realistic rendering of moving shadows cast by a point light while rotating the head (b). Practical photographic applications such as shadow diffusion (softening) are also made feasible with our system (c).

Abstract

At the core of portrait photography is the search for ideal lighting and viewpoint. The process often requires advanced knowledge in photography and an elaborate studio setup. In this work, we propose Holo-Relighting, a volumetric relighting method that is capable of synthesizing novel viewpoints, ***and*** *novel lighting from a single image. Holo-Relighting leverages the pretrained 3D GAN (EG3D) to reconstruct geometry and appearance from an input portrait as a set of 3D-aware features. We design a relighting module conditioned on a given lighting to process these features, and predict a relit 3D representation in the form of a tri-plane, which can render to an arbitrary viewpoint through volume rendering. Besides viewpoint and lighting control, Holo-Relighting also takes the head pose as a condition to enable head-pose-dependent lighting effects. With these novel designs, Holo-Relighting can generate complex non-Lambertian lighting effects (e.g., specular highlights and cast shadows) without using any explicit physical lighting priors. We train Holo-Relighting with data captured with a light stage, and propose two data-rendering techniques to improve the data quality for training the volumetric relighting system. Through quantitative and qualitative experiments, we demonstrate Holo-Relighting can achieve state-of-the-arts relighting quality with better photorealism, 3D consistency and controllability.*

1. Introduction

One essential challenge in portrait photography is to find the ideal lighting condition and viewpoint that best portray the subject, a process that often involves tedious adjustments of camera and lighting setup in a professional studio environment with expensive equipment [18, 54]. These require-

†The second authors

ments, however, are beyond the reach of most consumer photographers, especially with the increasing demands of quick selfies and candid photos. In this paper, we aim to meet the growing need of virtually synthesizing novel views and novel lighting for a given portrait to enable flexible portrait editing.

Several recent works [24, 25, 36, 45, 59, 61, 71, 79, 84, 87, 89] have made progress on single-image portrait relighting that enables post-capture lighting editing. However, simultaneous relighting and view synthesis for headshot portrait have received less attention. Such task inherently requires 3D understanding of an image, and is often addressed under a multi-view setting with specialized acquisition mechanisms [2, 20, 35, 37, 62, 88]. Despite great performance, these works are less approachable for average users and tend to fail on in-the-wild images, as they are designed for more controlled settings.

Precise view and lighting control requires a good estimation of physical properties such as material and geometry, which are fundamentally difficult given only a monocular 2D image. Previous works [12, 28, 44, 48, 63] usually rely on explicit physical modeling with simplified assumptions on the reflectance or/and lighting model to achieve the desired view and lighting control. For example, Lambertian and Phong models are often used as simplified reflectance models [44, 48, 63]; spherical harmonics [12, 28, 48] is commonly used as a lighting representation. Other works [28] make assumptions on the color of illumination to simplify their optimization space. While these assumptions provide a reasonable approximation to the actual light transport, they suffer from limited expressiveness and result in producing unrealistic shading effects and less accurate lighting effects.

In this paper, we propose ***Holo-Relighting***, a controllable volumetric relighting method that can render novel views and novel lighting from a single image. We show that challenging high-frequency lighting effects can be handled in a fully implicit manner without relying on any physical reflectance and lighting models. The key idea of our method is to represent 3D information from the input image as a set of 3D-aware features by exploiting the inversion of a pre-trained 3D GAN (EG3D [7]). We are then able to train a relighting model that is conditioned on the target lighting using these 3D-aware features. The output of the relighting model is a 3D representation embedded with the target illumination in the form of a tri-plane, from which an arbitrary view can be rendered using volume rendering. Besides viewpoint and lighting control, *Holo-Relighting* also takes the head pose as an input condition and enables head-pose-dependent lighting effects, which is under-explored in prior works.

We train the relighting module on a diverse set of portrait images under different illumination conditions, which are renderings using the "one-light-at-a-time" (OLAT) data captured with a light stage [11]. Our goal is to train the relighting module to learn arbitrary lighting effects, including ones that demand accurate geometry such as cast shadow. Therefore, it is crucial to ensure that the inverted input latent code retains more accurate geometry information. Simply relying on existing GAN inversion techniques [51, 52, 75, 80, 82] is sub-optimal due to the depth ambiguity when inferring geometry from a monocular input. To address this challenge, we extend the existing GAN inversion method [75] with multi-view regularization and camera pose optimization to encode a more accurate geometry. We additionally propose a shading transfer technique inspired by quotient image [57] to accommodate for the misalignment in high-frequency details induced by imperfect inversion. We show that training with data created using these strategies is important for synthesizing high-quality shading effects.

With extensive quantitative and qualitative experiments, we show that *Holo-Relighting* can produce relit portraits with more flexible control, superior photo-realism and multi-view consistency compared to the state-of-the-art alternatives. We summarize our contributions as follows:

- A novel controllable volumetric relighting method that renders free-view relit portraits with state-of-the-art photo-realism and view consistency.
- A novel relighting module capable of rendering complex shading effects including non-Lambertian reflections and cast shadows without imposing physical constraints.
- Two data-rendering techniques to make more effective use of light stage captures for training a volumetric relighting system.

2. Related Work

Portrait Relighting. There is a significant body of work studying 2D portrait relighting over past decades. In the pioneering work, Debevec *et al*. [11] design a special capture rig *i.e*. light stage to record reflectance field, which is then used to render images with novel lighting. Later methods remove hardware requirement by using techniques such as style transfer [58, 59], quotient image [47, 57], and intrinsic decomposition [1, 25, 27, 33, 34, 56]. Several recent approaches [36, 40, 45, 50, 61, 71, 79, 84, 88, 89] use neural networks to synthesize lighting effects and achieve higher photorealism. Our method is motivated by their success.

Beyond a 2D portrait, very few efforts investigate relighting and view synthesis in a unified framework. Many of them rely on multi-view setup to construct a physical relighting model [2, 20, 35, 69] or interpolate light transport [37, 49, 62, 88], thus are not suitable for in-the-wild images. For a single 2D input, the problem remains challenging due to its ill-posed nature. Some attempts use handcrafted face priors such as 3DMM [3] and apply physical or

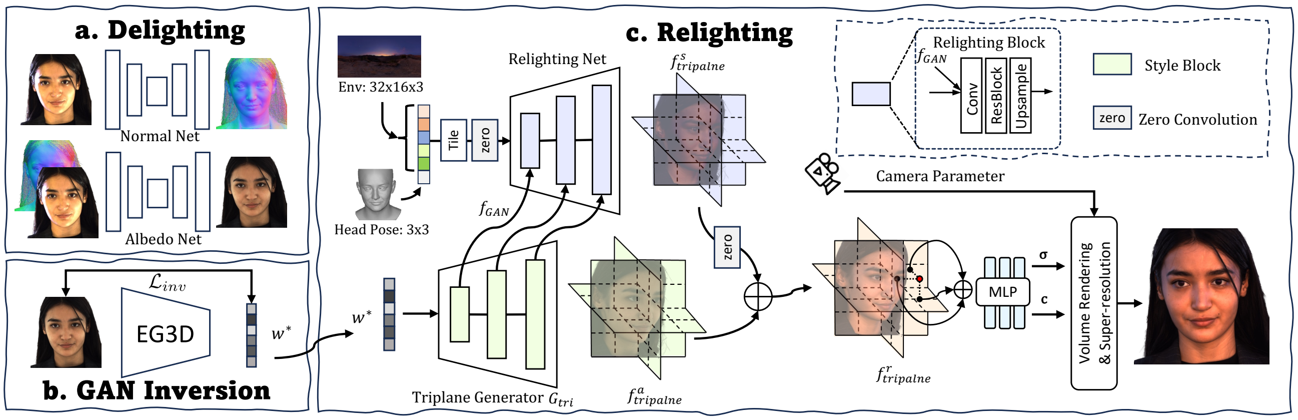

Figure 2. **An overview of *Holo-Relighting*.** Our method consists of three stages. (a) We first remove the shading from the input portrait and estimate an albedo image. (b) We then conduct GAN inversion upon EG3D to obtain a latent code w^* encoding 3D information of the subject. (c) The relighting network takes in the lighting condition, head pose as well as intermediate GAN features produced by EG3D's tri-plane generator G_{tri} using the inverted latent code w^*, and predicts a shading tri-plane $f^s_{triplane}$, which is summed to the albedo tri-plane $f^a_{triplane}$, resulting in the relit tri-plane $f^r_{triplane}$ with lighting embedded. High-resolution RGB images can be rendered from $f^r_{triplane}$ via volume rendering and a super-resolution network. During training, we freeze G_{tri} and only update the relighting net.

neural rendering [46, 66, 68] for relighting. These methods often lack geometric details and fail to synthesize hairs and mouth interior. As a result, their realism lags behind existing 2D approaches.

3D GANs. Recent 3D Generative Adversarial Networks (GANs) [16] can generate view-consistent images by training on 2D images. Early 3D GANs embed explicit 3D structure such as voxels [23, 41], and meshes [9, 22, 29] into their network but fail to achieve similar quality as those 2D GANs. More recent 3D GANs [6, 7, 14, 19, 42, 43, 55, 90] adopt implicit 3D representation such as radiance field [39]. Among them, EG3D [7] introduces an efficient tri-plane-based 3D representation which offers efficient and high-quality view synthesis on par with existing 2D GANs [5, 30]. Our method also adopts a tri-plane-based 3D representation for volume rendering.

Most 3D GANs are not relightable. To enable lighting control, recent efforts propose to embed physical reflectance models, such as Lambertian [44], Phong [48, 63], or Radiance Transfer [12], into their models to explicitly model lighting. While the simplified reflectance and illumination model (*i.e.* Spherical Harmonics lighting [17]) provide a good approximation to the light transport, they suffer from limited expressiveness and result in unrealistic rendering quality. Recently, Jiang *et al.* [28] propose to distill shading from 3D GANs. But their method assumes a fixed illuminant color to simplify their optimization space, thus is not applicable for general-purpose relighting.

GAN Inversion. GAN inversion is an essential step for GAN-based image editing which maps an image back to the latent space of a pretrained generator. It has numerous applications including inpainting [10, 51], style transfer [76, 77] and restoration [38, 67, 78]. 2D GAN inversion also enables viewpoint and lighting control [4, 8, 64, 65], but often lack view-consistency and/or limit to simple SH illumination. Recent efforts extend GAN inversion for 3D GANs to enable 3D face editing [75, 80, 82]. Our method exploits 3D GAN inversion [74] to retrieve 3D representations from an input image to help volumetric relighting.

Preliminary: EG3D Framework. Our method adopts pretrained EG3D [7] to extract 3D information from the input image through inversion. Here we briefly describe its framework. The core design of EG3D is to use an efficient tri-plane representation for volume rendering. A tri-plane consists of three individual 2D feature maps, where each of them represents three orthogonal plane f_{XY}, f_{YZ}, f_{XZ} in 3D space. These feature maps are split channel-wise from a single feature map produced by its StyleGAN2-like [30] tri-plane generator. For volume rendering, sampled 3D points along a ray are projected onto each plane to retrieve a summed 1-D feature f, from which a color feature c and density σ are decoded by an MLP. A multi-channel feature image I_c is then obtained via volume rendering [39]:

$$I_c(r) = \int_{t_n}^{t_f} T(t) \cdot \sigma(t) \cdot c(t)\, dt \tag{1}$$

where $T(t) = \exp\left(-\int_{t_n}^{t} \sigma(s)\, ds\right)$ is the transmittance. For efficiency, I_c is accumulated at low resolution and upsampled to the final high-resolution image through a super-resolution network.

3. Method

Holo-Relighting takes a single portrait image as input and predicts a novel image under desired lighting condition, viewpoint and head pose. Specifically, our system consists of three stages (Figure 2): (1) Delighting: given an input image, we remove the shading and shadow effects and predict an albedo image (Figure 2-(a)); (2) GAN Inversion: we reconstruct 3D information of the albedo image through GAN inversion, by projecting it to the latent space of a pretrained EG3D [7] (Figure 2-(b)). The obtained latent code w^* is used to retrieve a set of features f_{GAN} from EG3D's tri-plane Generator G and transmit them to a relighting network; (3) Relighting: we use a relighting module (Figure 2-(c)) conditioned on a given environment map, head pose and camera pose to process the GAN features f_{GAN} to produce a relit tri-plane $f^r_{triplane}$, from which a novel image can be synthesized through volume rendering.

3.1. Delighting stage

Following the popular delight-then-relight scheme in recent 2D relighting methods [36, 45, 71, 79], we start from a delighting stage, which predicts an albedo image using two separate networks, *i.e*. the *normal net* conditioned on the original input and the *albedo net* conditioned on the both inferred normal and input, as shown in Figure 2-(a). We use a U-Net architecture [53] for both networks. Training details can be found in Section 4.

3.2. GAN Inversion stage

We lift a 2D albedo image into 3D space by projecting it into the latent space of EG3D [7] through GAN inversion, which reconstructs 3D information of the subject. Specifically, given an albedo image $\mathcal{A}$ and a pretrained EG3D denoted as G and parameterized by θ, GAN inversion stage involves searching for a latent code w^* and fine-tuning G to best reconstruct $\mathcal{A}$. Formally,

$$w^*, \theta^* = \arg\min_{w,\theta} \mathcal{L}_{inv}(G(w;\theta), \mathcal{A}) \tag{2}$$

where $\mathcal{L}_{inv}$ is the reconstruction loss. And $G(w;\theta)$ is the generated albedo using weights θ. During optimization, we only update EG3D's tri-plane generator (denoted as G_{tri}) and freeze other parts. For simplicity, we refer θ as the parameters belonging to G_{tri} in the following paragraphs.

With optimized w^* and θ^*, we can retrieve a set of intermediate features f_{GAN} and an albedo tri-plane $f^a_{tripane}$ from $G_{tri}(w^*;\theta^*)$, which encodes the 3D information of the albedo image $\mathcal{A}$. We adapt an existing inversion method [75] during inference and refer readers to their paper for more detail. When creating training data, we propose two additional strategies to align encoded geometry and appearance with the target image. Details are described in Section 3.4.

3.3. Relighting stage

One core design of *Holo-Relighting* is the use of a learned neural network for relighting, that can generate complex non-Lambertian reflections and cast shadows without using any physical lighting priors. In the following, we first describe the designed relighting network in detail and then introduce two data-rendering techniques to facilitate training.

3.3.1 Relighting Net

As shown in Figure 2-(c), the relighting network takes a given lighting condition (*i.e*. environment map) and head pose (3×3 rotation matrix) as input. They are first reshaped into 1D tensors, and then tiled to 2D maps before feeding into the relighting net. The relighting net adopts a pyramidal structure which progressively increases its spatial resolution to match that of the tri-plane generator G_{tri}. At each resolution stage, the relighting net takes in an intermediate GAN feature f_{GAN} produced by $G_{tri}(w^*;\theta^*)$ of same resolution through concatenation, and outputs an upsampled feature for next resolution stage. The relighting net final produces a shading tri-plane $f^s_{triplane}$ which is added back to the albedo tri-plane $f^a_{triplane}$ to produce a relit tri-plane $f^r_{triplane}$ with the target illumination embedded. A relit image can then be rendered with volume rendering. In the following, we discuss several important aspects of the relighting net.

Feature-based Coarse-to-Fine Relighting. We build the relighting net upon GAN features, which encode rich 3D properties of the subject. Using f_{GAN}, the network managed to render complex lighting effects such as specular highlights and cast shadows (Figure 1) that require a good understanding of the geometry. The relighting network is naturally designed to consume GAN features in a coarse-to-fine manner, which enables to synthesize a global lighting distribution first and then refine local details based on a global context. As such, the network learns to synthesize both diffuse and high-frequency specular highlights well.

Stabilize Training with Zero-Convolution. As we add the shading tri-plane $f^s_{triplane}$ back to the albedo tri-plane $f^a_{triplane}$, the relighting network training is prone to diverge at the beginning. Inspired by [85], we apply two zero-initialized convolutions to the relighting network: one before the addition of shading tri-plane $f^s_{triplane}$ and albedo tri-plane $f^a_{triplane}$ and the other before taking in the lighting and pose condition. This strategy allows the model to gradually incorporate the given illumination signals and thus stabilize training.

Model Head-Pose Dependent Shading. Given a lighting condition, our relighting network can generate head-pose dependent effects, *e.g*. the moving specular highlights and cast shadows with changing head pose. This effect is often neglected in existing approaches [12, 44, 48, 63]. Although existing free-view relighting methods can render different views of a portrait, their view changes are always based on the movement of camera rather than head-pose.* In contrast, our method explicitly takes the head pose as an input

*Please refer to Fig. 9 (b) and (c) for the difference between the shading change caused by viewpoint change and head-pose change.

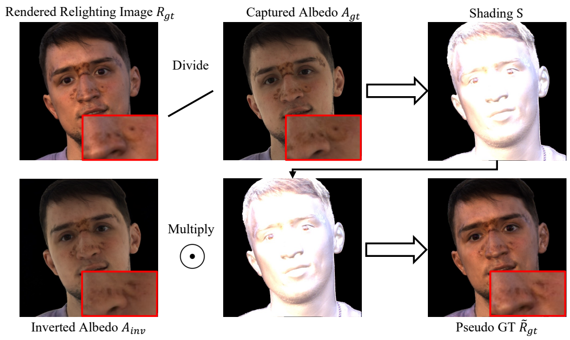

Figure 3. Illustration of proposed portrait shading transfer. We create a pseudo ground-truth image $\tilde{\mathcal{R}}_{gt}$ by transferring the shading from the OLAT renders. $\tilde{\mathcal{R}}_{gt}$ contains target illumination, and is consistent with the appearance encoded in f_{GAN} (*i.e.* $\mathcal{A}_{inv}$).

condition, and learns to generate head-pose dependent shading/shadows through training.

3.4. Training Data Creation

Similar to prior works [36, 45, 61], we train the relighting net using light stage [11] data rendered with diverse lighting environments. During training, the target image is a rendered relit image, and the input to the relighting net contains the target environment map, estimated head pose, camera pose, and GAN features f_{GAN} from $G(w^*; \theta^*)$. As we aim to use the relighting model to learn geometry-dependent lighting effects, we need to ensure that f_{GAN} encode accurate geometry of the input subject. We introduce a multi-view inversion strategy to leverage the multi-view light stage data for better geometry encoding. We then propose a shading transfer technique to deal with misalignment in high-frequency details induced by imperfect inversion.

Multi-View Regularization. Single-view inversion inherently suffers from depth ambiguity. The incorrectly encoded geometry prevents the relighting net from using 3D-aware features for relighting. To address this issue, we extend [75] into a multi-view scheme by leveraging our multi-view light stage captures. Formally, given a set of albedo images $\mathcal{A} = \{\mathcal{A}^1, \mathcal{A}^2, ..., \mathcal{A}^n\}$ captured from N viewpoints and associated camera poses $\mathbf{P} = \{p^1, p^2 ..., p^n\}$, we first freeze G and search for an optimal w^* shared among the frames and per-frame pose $\mathbf{P}^*$ by minimizing:

$$\mathcal{L}_{w,\mathbf{P}} = \frac{1}{n}\sum_{1}^{n} \|\phi_{VGG}(G(w, p^i)) - \phi_{VGG}(\mathcal{A}^i)\|_2 \quad (3)$$

where ϕ_{VGG} is a pretrained VGG [60] feature extractor. In the second step, we freeze the derived w^* and $\mathbf{P}^*$ and fine-tune the generator's parameter θ by optimizing

$$\mathcal{L}_{\theta} = \frac{1}{n}\sum_{1}^{n} \|G(\theta) - \mathcal{A}^i\|_2 + \mathcal{L}_{lpips}(G(\theta), \mathcal{A}^i) \quad (4)$$

where $\mathcal{L}_{lpips}$ is the perceptual loss defined in [86]. With multi-view regularization, the retrieved GAN features can reflect more accurate geometry.

Portrait Shading Transfer. As shown in Figure 3, even with advanced inversion techniques, the inverted albedo (which is encoded in f_{GAN}) cannot perfectly capture the details from the input. When training with per-pixel loss, such texture differences lead to blurry results. Inspired by quotient image [57], we resolve this with a simple yet effective strategy, *i.e.* by transferring the lighting from the target OLAT rendering to the inverted albedo image to create a pseudo ground-truth relighting image. Formally, given the real ground-truth relighting image $\mathcal{R}_{gt}$ and its associated albedo $\mathcal{A}_{gt}$, we compute a shading image $\mathcal{S} = \mathcal{R}_{gt}/\mathcal{A}_{gt}$ via element-wise division, a decomposition similar to [26]. We then transfer the shading $\mathcal{S}$ to the albedo image $\mathcal{A}_{inv}$ reconstructed from the latent code with a per-pixel multiplication $\mathcal{S}$ *i.e.* $\tilde{\mathcal{R}}_{gt} = \mathcal{S} \odot \mathcal{A}_{inv}$, where $\odot$ denotes the Hammond product, to produce a pseudo ground-truth image $\tilde{\mathcal{R}}_{gt}$, and use it to supervise the relighting module. As shown in Figure 3, $\tilde{\mathcal{R}}_{gt}$ preserves target shading while containing consistent facial details with the inverted albedo image, and thus is also consistent with encoded subject appearance in f_{GAN}.

3.5. Training Objective

To ensure both realness and fidelity, we use the following terms in our training objective function:

Reconstruction Loss $\mathcal{L}_{R_{norm}}$, $\mathcal{L}_{R_{alb}}$ and $\mathcal{L}_{R_{relit}}$: The standard $L1$ distance between the predicted normal, albedo, relit image and their ground-truth counterparts.

Perceptual Loss $\mathcal{L}_{P_{alb}}$ and $\mathcal{L}_{P_{relit}}$: The layer-wise feature difference between the predicted albedo/relit portrait and the ground truth albedo/relit portrait extracted by a pre-trained VGG [60] to increase perceptual quality.

Relighting Adversarial Loss $\mathcal{L}_{A_{relit}}$: The adversarial loss between the predicted relit portrait and ground truth to encourage the generation of high-frequency lighting details. We adopt a PatchGAN discriminator with spectral normalization [81] to compute this loss.

For delighting, We jointly train the albedo and normal net by optimizing:

$$\mathcal{L}_{delight} = \mathcal{L}_{R_{norm}} + \mathcal{L}_{R_{alb}} + \mathcal{L}_{P_{alb}} \quad (5)$$

For the relighting stage, the totally loss can be written as:

$$\mathcal{L}_{relight} = \mathcal{L}_{R_{relit}} + \mathcal{L}_{P_{relit}} + \mathcal{L}_{A_{relit}} \quad (6)$$

4. Data and Implementation Details

Data Preparation. Following [36, 45, 61], we use light stage captures [13] to render high quality datasets. Our light stage is structurally similar to [36, 61], which consists of 160 programmable LED lights and 4 frontal-view cameras. The light stage captures contain 69 subjects photographed with different poses, expressions and accessories, resulting in total 931 OLAT sequences. We reserve 15 subjects with different genders and races for testing. Following [45], we use the image captured with flat omnidi-

Table 1. Quantitative evaluations against free-view relighting methods.

Methods	LPIPS↓	NIQE↓	Deg↑	PSNR↑	SSIM↑
NeRFFaceLighting [28]	0.2179	8.132	0.6953	19.40	0.7214
FaceLit [48]	0.1542	8.017	0.7821	20.86	0.7281
Ours	**0.0917**	**5.273**	**0.8978**	**27.35**	**0.8511**

rectional lighting (*i.e.* all lights turned on) as ground-truth albedo and use photometric stereo [72] to calculate normals.

To create a high-quality training dataset, we collect 671 real environment maps from PloyHeaven [21]. We randomly select a subset of 500 environment maps for training and use the rest for testing. We augment lighting by randomly rotating the environment map and further include the original OLAT images into the training dataset, which results in a total of 520K images. We randomly pair testing OLAT sequence and environment maps to create test sets. More details can be found in the supplemental file.

Training Details. For training, we group 8 faces crops of size 512×512 as one batch. All networks are optimized with Adam [32] with a learning rate of $1e-5$. We joint train the albedo and normal net for 5 epochs, and train the relighting net for 10 epochs. The volume rendering resolution is set to 64×64. We adopt pre-trained EG3D and finetune their super-resolution network on our dataset. We freeze the triplane generator G_{tri} throughout the training. The proposed model is implemented using PyTorch and trained on 8 A100 GPUs. More details can be found in the supplement.

5. Experiments

In this section, we demonstrate the relighting quality and control capability of *Holo-Relighting* through extensive experiments. More results can be found in the supplemental file.

Evaluation Metrics. We follow [36] and report LPIPS [86] and NIQE [83] for perceptual quality and PSNR and SSIM [70] for fidelity. We also report identity metrics Deg (cosine similarity between LightCNN [73] features) to measure identity preservation. All metrics are calculated on the foreground subject with pre-computed masks using [31].

5.1. Comparisons with State-of-the-art Methods

We compare *Holo-Relighting* with state-of-the-art free-view relighting methods FaceLit [48], and NeRFFaceLighting [28]. In addition, as *Holo-Relighting* also supports conventional 2D portrait relighting by rendering to the input pose, we further show its effectiveness by comparing it with the state-of-the-art 2D relighting method Total Relighting [45](TR). For both experiments, we perform qualitative evaluations on in-the-wild images, and quantitative & qualitative evaluations on our test set.

Table 2. Quantitative evaluations against 2D relighting methods.

Methods	LPIPS↓	NIQE↓	Deg↑	PSNR↑	SSIM↑
TR [45]	0.1997	7.012	0.7638	19.58	0.6515
Ours	**0.0981**	**6.195**	**0.8793**	**26.17**	**0.8544**

5.1.1 Free-view Portrait Relighting.

To evaluate the performance on both view-synthesis and relighting, we compare our method with two state-of-the-art works: FaceLit [48] and NeRFFaceLighting [28]. For both methods, we follow their papers to estimate Spherical Harmonics coefficients from a reference image as target lighting (by adopting lighting estimator from [15] and [89] respectively). The reference image is rendered with environment maps and OLAT data [11] throughout our experiments. To perform relighting on an input image, we use [75] to perform inversion for FaceLit [48] as it does not have a native inversion module. For a fair comparison, we fix the head pose to frontal in our system throughout the evaluation to accommodate the setting in [48] and [28].

Qualitative results on in-the-wild images. We demonstrate the generalization capability of *Holo-Relighting* by performing relighting on in-the-wild portrait images, as shown in Figure 4. Since there is no ground truth, we provide a reference image rendered with the target environment map in the last column for perceptual comparison. *Holo-Relighting* generates the most convincing results, among all methods, with realistic shading and specularity. It is also robust to various lighting conditions, producing lighting effects that closely matches the reference images. In contrast, the results from NeRFFaceLighting and FaceLit contain visual artifacts and cannot fully reproduce the lighting effects.

Quantitative and qualitative results on OLAT test set. For quantitative evaluation, we create a test dataset of 235 images using OLAT data (frontal head pose) and sampled environment maps. Quantitative results are reported in Table 1. Our method consistently achieves the best perceptual quality, fidelity, and identity preservation quality. We report qualitative results in Figure 5. Given an input and a target lighting, our results match the target image most closely with faithful lighting effect and superior visual quality.

5.1.2 2D portrait relighting.

We compare *Holo-Relighting* with the state-of-the-art relighting method Total Relighting [45] (TR) on 2D portrait relighting. Total relighting is an image-to-image based system trained with OLAT data, and therefore does not allow view synthesis.

Qualitative results on in-the-wild images. We report visual comparison results in Figure 6. Our method produces photo-realistic diffuse and high-frequency specular reflections and robustly handles diverse subjects and illumination.

Figure 4. **Visual comparisons of free-view relighting on in-the-wild portraits.** Environment maps are shown as insets. We provide a reference image (last column) rendered using OLAT subject with the target lighting as guidance for comparison.

Figure 5. **Visual comparisons against free-view relighting methods on test set.** Our method produces more faithful target lighting effects.

Compared to TR [45], our method produces lighting effects that are more similar to the corresponding reference images.

Quantitative and qualitative results on OLAT test set. We create a test set of 940 images using OLAT data. From Figure 7, we can see that our method produces more faithful renderings than TR. As shown in Table 2, our method also yields better quantitative results.

5.2. Controllability

We demonstrate the flexible controllability of *Holo-Relighting* by illuminating an in-the-wild portrait photo † under a single directional light source. Using just one light

Figure 6. **Visual comparisons of 2D portrait relighting on in-the-wild images.** We provide a reference image rendered using OLAT subject and target lighting as guidance for comparsion.

source can effectively highlight intricate lighting effects such as specular highlights and cast shadows and showcase sophisticated light transport phenomena, which are crucial components for photorealism. Please refer to the supplemental video for more examples.

Lighting Control. *Holo-Relighting* can generate realistic cast shadows and specular highlights from a single light source, as shown in the first column of Figure 8. By softening the light source through blurring the environment map, the network can generate realistic shadow diffusion (softening) effects on the face. As shown in Figure 9-(a), *Holo-Relighting* produces consistent moving shadows and specu-

†The original image can be found in the supplemental file.

Figure 7. **Visual comparisons of 2D portrait relighting on test set**. Our method produces on-par or better results than Total Relighting [45] (TR).

Figure 8. **Illustration of lighting control using a single light source.** Our method produces realistic cast shadows and specular reflection. Important applications such as shadow diffusion (softening) can be easily achieved with *Holo-Relighting*.

lar reflections with a rotating light source.

Viewpoint Control. In addition to the view synthesis results shown in Figure 1, we further demonstrate our view control capability through the consistent lighting effect in Figure 9-(b): we can generate consistent cast shadows (*e.g.* under noise) and specular highlights (*e.g.* on the forehead and cheek) across different views.

Pose Control. We demonstrate our unique support for head pose control in Figure 9-(c). By comparing it with images of the same column in Figure 9-(b), one can see that *Holo-Relighting* synthesizes realistic moving highlights (*e.g.* on the forehead) and cast shadows (*e.g.* under the noise) with head rotation.

5.3. Ablation Study

We conduct ablation study on two proposed data-rendering methods: (1) Multi-view GAN Inversion and (2) Shading Transfer (ST). We report quantitative results in Table 3. We can conclude that (1) multi-view inversion significantly improves the visual quality of the result compared to vanilla single-view inversion during the training phase; (2) shading transfer helps to align the appearance of the target image with that encoded in GAN features and further improve the performance of *Holo-Relighting*. More details and qualitative results can be found in the supplemental material.

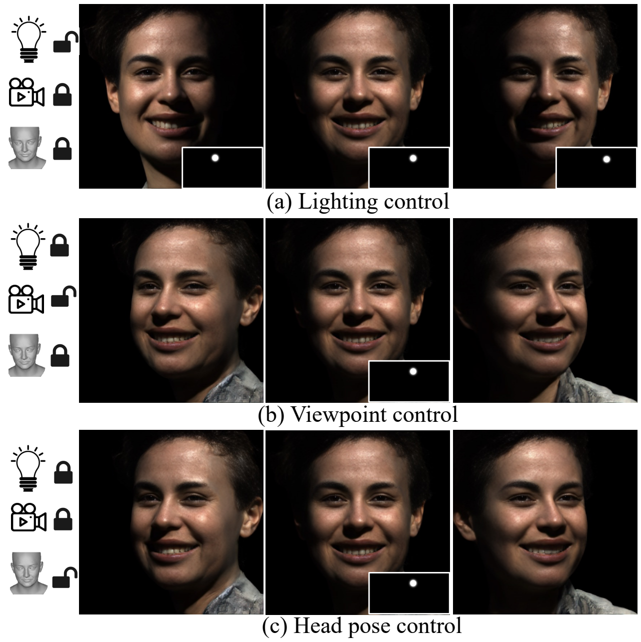

Figure 9. **Illustration of controllability using a single light source.** Our method produces consistent shadows and specular highlights with respect to a rotating (a) light source, (b) camera pose and (c) head pose.

Table 3. Ablation study on proposed data-rendering strategies.

Methods	LPIPS↓	NIQE↓	Deg.↑	PSNR↑	SSIM↑
Single-view	0.1443	7.381	0.8142	23.57	0.7845
w/o ST	0.1382	7.143	0.8387	24.03	0.7840
Holo-Relighting	**0.0997**	**6.330**	**0.8735**	**25.89**	**0.8479**

6. Conclusion

In this work, we propose *Holo-Relighting*, a novel volumetric relighting method that is capable of synthesizing renderings under novel viewpoints and novel lighting from a single input image. *Holo-Relighting* leverages the 3D representations from the pretrained 3D-aware GAN (EG3D) and predicts a lighting and pose-aware 3D representation (tri-plane features), from which a portrait image with controllable lighting, head pose, and viewpoint can be produced using volume rendering. We demonstrate the high controllability and state-of-the-art relighting quality of *Holo-Relighting* through extensive experiments. Discussion of limitations can be found in the supplemental file.

Acknowledgments This work was supported by NSF CAREER award 2045489. We thank Chaowei Company for the support of light stage data.

References

[1] Jonathan T Barron and Jitendra Malik. Shape, illumination, and reflectance from shading. *IEEE transactions on pattern analysis and machine intelligence*, 37(8):1670–1687, 2014. 2

[2] Sai Bi, Stephen Lombardi, Shunsuke Saito, Tomas Simon, Shih-En Wei, Kevyn Mcphail, Ravi Ramamoorthi, Yaser Sheikh, and Jason Saragih. Deep relightable appearance models for animatable faces. *ACM TOG*, 40(4):1–15, 2021. 2

[3] Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3d faces. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pages 157–164. 2023. 2

[4] Mallikarjun BR, Ayush Tewari, Abdallah Dib, Tim Weyrich, Bernd Bickel, Hans-Peter Seidel, Hanspeter Pfister, Wojciech Matusik, Louis Chevallier, Mohamed Elgharib, et al. Photoapp: Photorealistic appearance editing of head portraits. *ACM TOG*, 40(4):1–16, 2021. 3

[5] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. In *ICLR*, 2018. 3

[6] Eric R Chan, Marco Monteiro, Petr Kellnhofer, Jiajun Wu, and Gordon Wetzstein. pi-gan: Periodic implicit generative adversarial networks for 3d-aware image synthesis. In *CVPR*, pages 5799–5809, 2021. 3

[7] Eric R Chan, Connor Z Lin, Matthew A Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas J Guibas, Jonathan Tremblay, Sameh Khamis, et al. Efficient geometry-aware 3d generative adversarial networks. In *CVPR*, pages 16123–16133, 2022. 2, 3, 4

[8] Prashanth Chandran, Sebastian Winberg, Gaspard Zoss, Jérémy Riviere, Markus Gross, Paulo Gotardo, and Derek Bradley. Rendering with style: combining traditional and neural approaches for high-quality face rendering. *ACM TOG*, 40(6):1–14, 2021. 3

[9] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *CVPR*, pages 5939–5948, 2019. 3

[10] Giannis Daras, Joseph Dean, Ajil Jalal, and Alex Dimakis. Intermediate layer optimization for inverse problems using deep generative models. pages 2421–2432. PMLR, 2021. 3

[11] Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. Acquiring the reflectance field of a human face. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 145–156, 2000. 2, 5, 6

[12] Boyang Deng, Yifan Wang, and Gordon Wetzstein. Lumigan: Unconditional generation of relightable 3d human faces. *arXiv preprint arXiv:2304.13153*, 2023. 2, 3, 4

[13] Yu Deng, Jiaolong Yang, Sicheng Xu, Dong Chen, Yunde Jia, and Xin Tong. Accurate 3d face reconstruction with weakly-supervised learning: From single image to image set. In *CVPRW*, 2019. 5

[14] Yu Deng, Jiaolong Yang, Jianfeng Xiang, and Xin Tong. Gram: Generative radiance manifolds for 3d-aware image generation. In *CVPR*, pages 10673–10683, 2022. 3

[15] Yao Feng, Haiwen Feng, Michael J Black, and Timo Bolkart. Learning an animatable detailed 3d face model from in-the-wild images. *ACM Transactions on Graphics (ToG)*, 40(4): 1–13, 2021. 6

[16] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020. 3

[17] Robin Green. Spherical harmonic lighting: The gritty details. In *Archives of the game developers conference*, page 4, 2003. 3

[18] Christopher Grey. *Master lighting guide for portrait photographers*. Amherst Media, 2014. 1

[19] Jiatao Gu, Lingjie Liu, Peng Wang, and Christian Theobalt. Stylenerf: A style-based 3d aware generator for high-resolution image synthesis. In *ICLR*, 2021. 3

[20] Kaiwen Guo, Peter Lincoln, Philip Davidson, Jay Busch, Xueming Yu, Matt Whalen, Geoff Harvey, Sergio Orts-Escolano, Rohit Pandey, Jason Dourgarian, et al. The relightables: Volumetric performance capture of humans with realistic relighting. *ACM TOG*, 38(6):1–19, 2019. 2

[21] Poly Haven. Poly haven. 6

[22] Paul Henderson, Vagia Tsiminaki, and Christoph H Lampert. Leveraging 2d data to learn textured 3d mesh generation. In *CVPR*, pages 7498–7507, 2020. 3

[23] Philipp Henzler, Niloy J Mitra, and Tobias Ritschel. Escaping plato's cave: 3d shape from adversarial rendering. In *ICCV*, pages 9984–9993, 2019. 3

[24] Andrew Hou, Ze Zhang, Michel Sarkis, Ning Bi, Yiying Tong, and Xiaoming Liu. Towards high fidelity face relighting with realistic shadows. In *CVPR*, pages 14719–14728, 2021. 2

[25] Andrew Hou, Michel Sarkis, Ning Bi, Yiying Tong, and Xiaoming Liu. Face relighting with geometrically consistent shadows. In *CVPR*, pages 4217–4226, 2022. 2

[26] Alexandru Eugen Ichim, Sofien Bouaziz, and Mark Pauly. Dynamic 3d avatar creation from hand-held video input. *ACM TOG*, 34(4):1–14, 2015. 5

[27] Chaonan Ji, Tao Yu, Kaiwen Guo, Jingxin Liu, and Yebin Liu. Geometry-aware single-image full-body human relighting. 2022. 2

[28] Kaiwen Jiang, Shu-Yu Chen, Hongbo Fu, and Lin Gao. Nerffacelighting: Implicit and disentangled face lighting representation leveraging generative prior in neural radiance fields. *ACM TOG*, 42(3):1–18, 2023. 2, 3, 6

[29] Angjoo Kanazawa, Shubham Tulsiani, Alexei A Efros, and Jitendra Malik. Learning category-specific mesh reconstruction from image collections. In *ECCV*, pages 371–386, 2018. 3

[30] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *CVPR*, pages 8110–8119, 2020. 3

[31] Zhanghan Ke, Jiayu Sun, Kaican Li, Qiong Yan, and Rynson W.H. Lau. Modnet: Real-time trimap-free portrait matting via objective decomposition. In *AAAI*, 2022. 6

[32] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, San Diega, CA, USA, 2015. 6

[33] Ha A Le and Ioannis A Kakadiaris. Illumination-invariant face recognition with deep relit face images. In *IEEE Winter Conference on Applications of Computer Vision*, pages 2146–2155. IEEE, 2019. 2

[34] Chen Li, Kun Zhou, and Stephen Lin. Intrinsic face image decomposition with human face priors. In *ECCV*, pages 218–233. Springer, 2014. 2

[35] Wan-Chun Ma, Tim Hawkins, Pieter Peers, Charles-Felix Chabert, Malte Weiss, Paul E Debevec, et al. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. *Rendering Techniques*, 2007 (9):10, 2007. 2

[36] Yiqun Mei, He Zhang, Xuaner Zhang, Jianming Zhang, Zhixin Shu, Yilin Wang, Zijun Wei, Shi Yan, HyunJoon Jung, and Vishal M Patel. Lightpainter: Interactive portrait relighting with freehand scribble. In *CVPR*, pages 195–205, 2023. 2, 4, 5, 6

[37] Abhimitra Meka, Rohit Pandey, Christian Haene, Sergio Orts-Escolano, Peter Barnum, Philip David-Son, Daniel Erickson, Yinda Zhang, Jonathan Taylor, Sofien Bouaziz, et al. Deep relightable textures: volumetric performance capture with neural rendering. *ACM TOG*, 39(6):1–21, 2020. 2

[38] Sachit Menon, Alexandru Damian, Shijia Hu, Nikhil Ravi, and Cynthia Rudin. Pulse: Self-supervised photo upsampling via latent space exploration of generative models. In *CVPR*, pages 2437–2445, 2020. 3

[39] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 3

[40] Thomas Nestmeyer, Jean-François Lalonde, Iain Matthews, and Andreas Lehrmann. Learning physics-guided face relighting under directional light. In *CVPR*, pages 5124–5133, 2020. 2

[41] Thu Nguyen-Phuoc, Chuan Li, Lucas Theis, Christian Richardt, and Yong-Liang Yang. Hologan: Unsupervised learning of 3d representations from natural images. In *ICCV*, pages 7588–7597, 2019. 3

[42] Michael Niemeyer and Andreas Geiger. Giraffe: Representing scenes as compositional generative neural feature fields. In *CVPR*, pages 11453–11464, 2021. 3

[43] Roy Or-El, Xuan Luo, Mengyi Shan, Eli Shechtman, Jeong Joon Park, and Ira Kemelmacher-Shlizerman. Stylesdf: High-resolution 3d-consistent image and geometry generation. In *CVPR*, pages 13503–13513, 2022. 3

[44] Xingang Pan, Xudong Xu, Chen Change Loy, Christian Theobalt, and Bo Dai. A shading-guided generative implicit model for shape-accurate 3d-aware image synthesis. *NeurIPS*, 34:20002–20013, 2021. 2, 3, 4

[45] 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. *ACM TOG*, 40(4):1–21, 2021. 2, 4, 5, 6, 7, 8

[46] Foivos Paraperas Papantoniou, Alexandros Lattas, Stylianos Moschoglou, and Stefanos Zafeiriou. Relightify: Relightable 3d faces from a single image via diffusion models. In *ICCV*, 2023. 3

[47] Pieter Peers, Naoki Tamura, Wojciech Matusik, and Paul Debevec. Post-production facial performance relighting using reflectance transfer. *ACM TOG*, 26(3):52–es, 2007. 2

[48] Anurag Ranjan, Kwang Moo Yi, Jen-Hao Rick Chang, and Oncel Tuzel. Facelit: Neural 3d relightable faces. In *CVPR*, pages 8619–8628, 2023. 2, 3, 4, 6

[49] Pramod Rao, Mallikarjun B R, Gereon Fox, Tim Weyrich, Bernd Bickel, Hans-Peter Seidel, Hanspeter Pfister, Wojciech Matusik, Ayush Tewari, Christian Theobalt, and Mohamed Elgharib. Vorf: Volumetric relightable faces. 2022. 2

[50] Mengwei Ren, Wei Xiong, Jae Shin Yoon, Zhixin Shu, Jianming Zhang, HyunJoon Jung, Guido Gerig, and He Zhang. Relightful harmonization: Lighting-aware portrait background replacement. *CVPR*, 2024. 2

[51] Elad Richardson, Yuval Alaluf, Or Patashnik, Yotam Nitzan, Yaniv Azar, Stav Shapiro, and Daniel Cohen-Or. Encoding in style: a stylegan encoder for image-to-image translation. In *CVPR*, pages 2287–2296, 2021. 2, 3

[52] Daniel Roich, Ron Mokady, Amit H Bermano, and Daniel Cohen-Or. Pivotal tuning for latent-based editing of real images. *ACM TOG*, 42(1):1–13, 2022. 2

[53] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. pages 234–241. Springer, 2015. 4

[54] James Boniface Schriever and Thomas Harrison Cummings. *Complete Self-instructing Library of Practical Photography: Negative retouching; etching and modeling; encyclopedic index*. American school of art and photography, 1909. 1

[55] Katja Schwarz, Yiyi Liao, Michael Niemeyer, and Andreas Geiger. Graf: Generative radiance fields for 3d-aware image synthesis. *NeurIPS*, 33:20154–20166, 2020. 3

[56] Davoud Shahlaei and Volker Blanz. Realistic inverse lighting from a single 2d image of a face, taken under unknown and complex lighting. In *2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG)*, pages 1–8. IEEE, 2015. 2

[57] Amnon Shashua and Tammy Riklin-Raviv. The quotient image: Class-based re-rendering and recognition with varying illuminations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):129–139, 2001. 2, 5

[58] YiChang Shih, Sylvain Paris, Connelly Barnes, William T Freeman, and Frédo Durand. Style transfer for headshot portraits. *ACM TOG*, 33(4):1–14, 2014. 2

[59] Zhixin Shu, Sunil Hadap, Eli Shechtman, Kalyan Sunkavalli, Sylvain Paris, and Dimitris Samaras. Portrait lighting transfer using a mass transport approach. *ACM TOG*, 36(4):1, 2017. 2

[60] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 5

[61] Tiancheng Sun, Jonathan T Barron, Yun-Ta Tsai, Zexiang Xu, Xueming Yu, Graham Fyffe, Christoph Rhemann, Jay Busch, Paul E Debevec, and Ravi Ramamoorthi. Single image portrait relighting. *ACM TOG*, 38(4):79–1, 2019. 2, 5

[62] Tiancheng Sun, Kai-En Lin, Sai Bi, Zexiang Xu, and Ravi Ramamoorthi. Nelf: Neural light-transport field for portrait view synthesis and relighting. *arXiv preprint arXiv:2107.12351*, 2021. 2

[63] Feitong Tan, Sean Fanello, Abhimitra Meka, Sergio Orts-Escolano, Danhang Tang, Rohit Pandey, Jonathan Taylor, Ping Tan, and Yinda Zhang. Volux-gan: A generative model for 3d face synthesis with hdri relighting. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 1–9, 2022. 2, 3, 4

[64] Ayush Tewari, Mohamed Elgharib, Florian Bernard, Hans-Peter Seidel, Patrick Pérez, Michael Zollhöfer, and Christian Theobalt. Pie: Portrait image embedding for semantic control. *ACM TOG*, 39(6):1–14, 2020. 3

[65] Ayush Tewari, Mohamed Elgharib, Gaurav Bharaj, Florian Bernard, Hans-Peter Seidel, Patrick Pérez, Michael Zöllhofer, and Christian Theobalt. Stylerig: Rigging stylegan for 3d control over portrait images. In *CVPR*. IEEE, 2020. 3

[66] Ayush Tewari, Tae-Hyun Oh, Tim Weyrich, Bernd Bickel, Hans-Peter Seidel, Hanspeter Pfister, Wojciech Matusik, Mohamed Elgharib, Christian Theobalt, et al. Monocular reconstruction of neural face reflectance fields. In *CVPR*, pages 4791–4800, 2021. 3

[67] Xintao Wang, Yu Li, Honglun Zhang, and Ying Shan. Towards real-world blind face restoration with generative facial prior. In *CVPR*, pages 9168–9178, 2021. 3

[68] Youjia Wang, Kai He, Taotao Zhou, Kaixin Yao, Nianyi Li, Lan Xu, and Jingyi Yu. Free-view face relighting using a hybrid parametric neural model on a small-olat dataset. *IJCV*, 131(4):1002–1021, 2023. 3

[69] Yifan Wang, Aleksander Holynski, Xiuming Zhang, and Xuaner Zhang. Sunstage: Portrait reconstruction and relighting using the sun as a light stage. In *CVPR*, pages 20792–20802, 2023. 2

[70] Zhou Wang. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 6

[71] Zhibo Wang, Xin Yu, Ming Lu, Quan Wang, Chen Qian, and Feng Xu. Single image portrait relighting via explicit multiple reflectance channel modeling. *ACM TOG*, 39(6):1–13, 2020. 2, 4

[72] Andreas Wenger, Andrew Gardner, Chris Tchou, Jonas Unger, Tim Hawkins, and Paul Debevec. Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM TOG*, 24(3):756–764, 2005. 6

[73] Xiang Wu, Ran He, Zhenan Sun, and Tieniu Tan. A light cnn for deep face representation with noisy labels. *IEEE Transactions on Information Forensics and Security*, 13(11): 2884–2896, 2018. 6

[74] Weihao Xia, Yulun Zhang, Yujiu Yang, Jing-Hao Xue, Bolei Zhou, and Ming-Hsuan Yang. Gan inversion: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3121–3138, 2022. 3

[75] Jiaxin Xie, Hao Ouyang, Jingtan Piao, Chenyang Lei, and Qifeng Chen. High-fidelity 3d gan inversion by pseudo-multi-view optimization. In *CVPR*, pages 321–331, 2023. 2, 3, 4, 5, 6

[76] Shuai Yang, Liming Jiang, Ziwei Liu, and Chen Change Loy. Pastiche master: Exemplar-based high-resolution portrait style transfer. In *CVPR*, pages 7693–7702, 2022. 3

[77] Shuai Yang, Liming Jiang, Ziwei Liu, and Chen Change Loy. Vtoonify: Controllable high-resolution portrait video style transfer. *ACM TOG*, 41(6):1–15, 2022. 3

[78] Tao Yang, Peiran Ren, Xuansong Xie, and Lei Zhang. Gan prior embedded network for blind face restoration in the wild. In *CVPR*, pages 672–681, 2021. 3

[79] Yu-Ying Yeh, Koki Nagano, Sameh Khamis, Jan Kautz, Ming-Yu Liu, and Ting-Chun Wang. Learning to relight portrait images via a virtual light stage and synthetic-to-real adaptation. *ACM TOG*, 2022. 2, 4

[80] Fei Yin, Yong Zhang, Xuan Wang, Tengfei Wang, Xiaoyu Li, Yuan Gong, Yanbo Fan, Xiaodong Cun, Ying Shan, Cengiz Oztireli, et al. 3d gan inversion with facial symmetry prior. In *CVPR*, pages 342–351, 2023. 2, 3

[81] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Free-form image inpainting with gated convolution. In *ICCV*, pages 4471–4480, 2019. 5

[82] Ziyang Yuan, Yiming Zhu, Yu Li, Hongyu Liu, and Chun Yuan. Make encoder great again in 3d gan inversion through geometry and occlusion-aware encoding. *arXiv preprint arXiv:2303.12326*, 2023. 2, 3

[83] Lin Zhang, Lei Zhang, and Alan C Bovik. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*, 24(8):2579–2591, 2015. 6

[84] Longwen Zhang, Qixuan Zhang, Minye Wu, Jingyi Yu, and Lan Xu. Neural video portrait relighting in real-time via consistency modeling. In *ICCV*, pages 802–812, 2021. 2

[85] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *ICCV*, pages 3836–3847, 2023. 4

[86] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, pages 586–595, 2018. 5, 6

[87] Xuaner Zhang, Jonathan T Barron, Yun-Ta Tsai, Rohit Pandey, Xiuming Zhang, Ren Ng, and David E Jacobs. Portrait shadow manipulation. *ACM TOG*, 39(4):78–1, 2020. 2

[88] Xiuming Zhang, Sean Fanello, Yun-Ta Tsai, Tiancheng Sun, Tianfan Xue, Rohit Pandey, Sergio Orts-Escolano, Philip Davidson, Christoph Rhemann, Paul Debevec, et al. Neural light transport for relighting and view synthesis. *ACM TOG*, 40(1):1–17, 2021. 2

[89] Hao Zhou, Sunil Hadap, Kalyan Sunkavalli, and David W Jacobs. Deep single-image portrait relighting. In *ICCV*, pages 7194–7202, 2019. 2, 6

[90] Peng Zhou, Lingxi Xie, Bingbing Ni, and Qi Tian. Cips-3d: A 3d-aware generator of gans based on conditionally-independent pixel synthesis. *arXiv preprint arXiv:2110.09788*, 2021. 3