

3DPR: Single Image 3D Portrait Relighting with Generative Priors

PRAMOD RAO, Max Planck Institute for Informatics, SIC & VIA Research Center, Germany

ABHIMITRA MEKA*, Google Inc., USA

XILONG ZHOU*, Max Planck Institute for Informatics & SIC, Germany

GEREON FOX, Max Planck Institute for Informatics & SIC, Germany

MALLIKARJUN B R, Max Planck Institute for Informatics & SIC, Germany

FANGNENG ZHAN, Max Planck Institute for Informatics & SIC, Germany

TIM WEYRICH, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

BERND BICKEL, ETH Zürich, Switzerland

HANSPETER PFISTER, Harvard University, USA

WOJCIECH MATUSIK, Massachusetts Institute of Technology, USA

THABO BEELER, Google Inc., Switzerland

MOHAMED ELGHARIB, Max Planck Institute for Informatics & SIC, Germany

MARC HABERMANN, Max Planck Institute for Informatics, SIC & VIA Research Center, Germany

CHRISTIAN THEOBALT, Max Planck Institute for Informatics, SIC & VIA Research Center, Germany

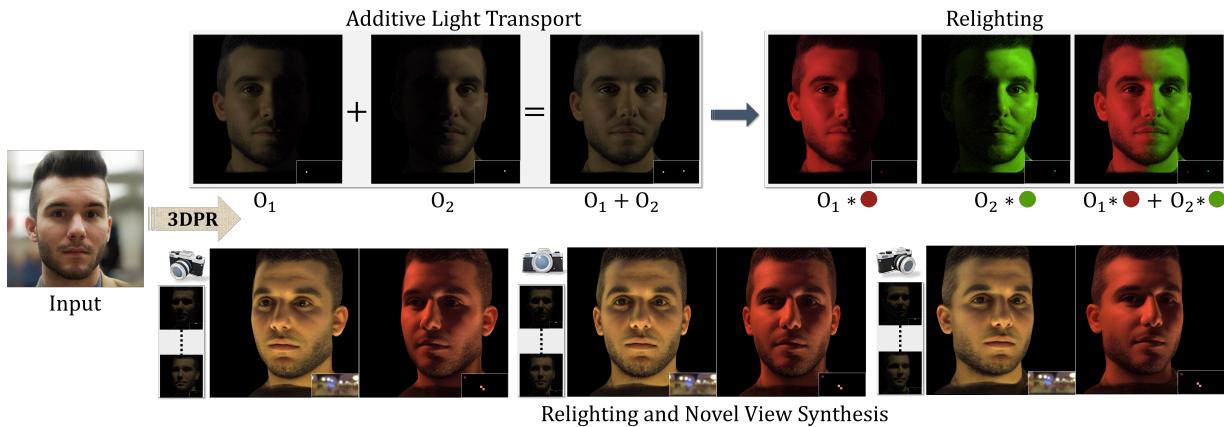


Fig. 1. We present 3DPR, a monocular 3D portrait relighting method that can synthesize novel views under desired illumination. Given a monocular input image, 3DPR predicts a reflectance basis in the form of One-Light-At-a-Time (OLAT) images of the subject (top row). By linearly combining the OLAT basis based on given HDRI map, the subject can be placed in novel relit environments (bottom row). Moreover, the OLATs can be rendered for a desired novel camera viewpoint, facilitating 3D-consistent portrait relighting.

Rendering novel, relit views of a human head, given a monocular portrait image as input, is an inherently underconstrained problem. The traditional graphics solution is to explicitly decompose the input image into geometry, material and lighting via differentiable rendering; but this is constrained by the multiple assumptions and approximations of the underlying models

*Equal contribution

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SA Conference Papers '25, December 15–18, 2025, Hong Kong, Hong Kong

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

and parameterizations of these scene components. We propose 3DPR, an image-based relighting model that leverages generative priors learnt from multi-view One-Light-at-A-Time (OLAT) images captured in a light stage. We introduce a new diverse and large-scale multi-view 4K OLAT dataset of 139 subjects to learn a high-quality prior over the distribution of high-frequency face reflectance. We leverage the latent space of a pre-trained generative head model that provides a rich prior over face geometry learnt from in-the-wild image datasets. The input portrait is first embedded in the latent manifold of such a model through an encoder-based inversion process. Then a novel triplane-based reflectance network trained on our lightstage data is used to synthesize high-fidelity OLAT images to enable image-based relighting. Our reflectance network operates in the latent space of the generative head model, crucially enabling a relatively small number of lightstage images to train the reflectance model. Combining the generated OLATs according to a given HDRI environment maps yields physically accurate environmental relighting results. Through quantitative and qualitative evaluations, we demonstrate that 3DPR outperforms previous methods, particularly in preserving identity

and in capturing lighting effects such as specularities, self-shadows, and subsurface scattering.

CCS Concepts: • Computing methodologies → *Image manipulation; Image-based rendering*.

ACM Reference Format:

Pramod Rao, Abhimitra Meka, Xilong Zhou, Gereon Fox, Mallikarjun B R, Fangneng Zhan, Tim Weyrich, Bernd Bickel, Hanspeter Pfister, Wojciech Matusik, Thabo Beeler, Mohamed Elgharib, Marc Habermann, and Christian Theobalt. 2025. *3DPR: Single Image 3D Portrait Relighting with Generative Priors*. In *SIGGRAPH Asia 2025 Conference Papers (SA Conference Papers '25), December 15–18, 2025, Hong Kong, Hong Kong*. ACM, New York, NY, USA, 25 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnn>

1 INTRODUCTION

Modern computer graphics applications, such as Augmented and Virtual Reality, demand blending of real and synthetic assets together seamlessly into a single image. Human faces are of particular importance in such applications and require converting few-shot or even a single monocular face image into 3D assets that can be rendered under novel environments from desired viewpoints to achieve visual immersion. However, achieving accurate relighting and novel view synthesis in a single unified rendering framework is a highly ill-posed challenge due to the underconstrained problem of 3D modeling from a monocular input and the complexity of the underlying light transport.

Recently, many data-driven methods have been proposed to learn a 3D prior over the underconstrained solution space of this problem. Some methods [Chan et al. 2022; Deng et al. 2022; Gu et al. 2022] propose generative volumetric representations to synthesize portraits in a 3D-consistent manner. Such methods are aimed at solving novel view synthesis and do not necessarily tackle the problem of face relighting. More recent methods [Deng et al. 2023; Jiang et al. 2023; Ranjan et al. 2023] build on these volumetric representations and learn a generative model that can also extract reflectance information from a given portrait image. These methods are trained using physically-based rendering models and hence suffer from several issues such as the unknown camera and illumination conditions in their in-the-wild training sets. The gap between the underlying physical rendering model and real-world images also significantly affects the quality of photorealism and lighting reproduction. To address these issues, image-based relighting methods [Haotian et al. 2024; Li et al. 2023; Sarkar et al. 2023; Yang et al. 2023] have been developed using One-Light-at-A-Time (OLAT) [Debevec et al. 2000] datasets, which offer ground truth supervision without the need to explicitly model physical light transport. However, these works are either subject-specific [He et al. 2024; Saito et al. 2024; Sarkar et al. 2023; Yang et al. 2023] or they require multi-view input during inference [Haotian et al. 2024; Li et al. 2023], and hence do not support generalization to monocular in-the-wild portraits as input.

To relight a single portrait image in a physically-accurate manner, several other works [B R et al. 2021a; Deng et al. 2023; Jiang et al. 2023; Rao et al. 2022, 2024a] train a generative model using such an OLAT dataset or GAN-based training[Goodfellow et al. 2014]. However, several challenges limit the fidelity of the resulting relighting: Some such methods [Deng et al. 2023; Jiang et al. 2023; Mei et al. 2024; Rao et al. 2023] require test-time optimization for each subject,

which is very time-consuming and impractical for most AR/VR applications. Other methods [B R et al. 2021a; Rao et al. 2024a] propose lightweight solutions, that utilize 2D generative models like StyleGAN[Karras et al. 2020] or EG3D [Chan et al. 2022] that restricts detailed face reflectance modelling and struggle with complex lighting effects, such as accurate shadows and specularities (Fig. 6). VoRF [Rao et al. 2022], also a volumetric relighting technique, excels at capturing shadows as it models face reflectance through a set of OLAT basis functions. However, its results are over-smooth and it struggles to generalize to in-the-wild portraits as it is trained on a limited set of face OLAT images that do not capture a rich geometry and appearance prior. No large-scale light stage dataset of sufficient diversity is currently publicly available.

To address these challenges, we present *3DPR*, a 3D portrait relighting method that leverages volumetric generative models combined with a novel light stage dataset. We introduce *FaceOLAT*, a large-scale, high-quality, human face OLAT dataset that will be publicly released for the benefit of the research community. *3DPR* takes a monocular portrait as input and renders 3D-consistent novel views under any given lighting, accurately simulating complex light transport effects (Fig. 1). To achieve this, our method synthesizes OLAT images for the desired viewpoint, that are then linearly combined according to an HDRI map. We build on EG3D [Chan et al. 2022], which provides a robust generative prior for facial geometry and appearance, enabling our approach to effectively generalize to unseen faces. The input portrait is embedded into EG3D’s latent space via encoder-based GAN inversion [Yuan et al. 2023]. Crucially, we combine EG3D with our novel reflectance model, efficiently encoding face reflectance into a triplane representation, which allows rendering high-resolution OLAT images.

Our dataset, *FaceOLAT*, offers 40 camera viewpoints at 4K resolution and 331 point light sources, surpassing all publicly available datasets (see Tab. 1) as well as the *non-public* OLAT dataset of Weyrich et al. [2006], which is widely used for evaluating face reflectance modelling. *FaceOLAT* includes subjects of different skin tones, hair colors, eye colors, ethnicities and ages, providing demographic diversity. Training *3DPR* on *FaceOLAT* leads to state-of-the-art results, both quantitatively and qualitatively.

In summary, we contribute:

- An image-based 3D portrait relighting method leveraging a combination of pretrained generative prior and an OLAT dataset to enable physically accurate editing of both illumination and viewpoint of a monocular input image.
- *FaceOLAT*, a large face OLAT dataset, comprising 139 subjects captured by 40 cameras under 331 point light sources. Our dataset will be publicly available.

Comprehensive quantitative and qualitative evaluations shows that our method achieves state-of-the-art performance. Code and pre-trained checkpoints is available under <https://vcai.mpi-inf.mpg.de/projects/3dpr/>.

2 RELATED WORK

Many portrait relighting methods employ illumination models that are trained on synthetic data [Chandran et al. 2022; Lattas et al. 2021; Sengupta et al. 2018; Shu et al. 2017; Zhou et al. 2019]. While

Table 1. *FaceOLAT* is the first large-scale, publicly available multi-view HDR OLAT face dataset. It includes 139 subjects captured under 3 expressions, illuminated with 331 dense OLAT lighting conditions from 40 viewpoints at 4K resolution. This setup enables high-fidelity full-head reflectance modeling, including hair. None of the existing datasets that are *publicly* available offers this combination of subject diversity, dense illumination, and multi-view coverage at this scale. The ✓ symbol for RGCA [Saito et al. 2024] indicates the use of grouped OLATs [Wenger et al. 2005], intended for dynamic capture. ICT-3DRFE [Stratou et al. 2011] and Ultrastage [Zhou et al. 2023] provide only gradient illumination, which is not optimal for high-quality relighting.

| Dataset | # Illuminations | # Subjects | # Views | Resolution | Image-based Relighting |
|------------------------------------------------|-----------------|------------|---------|------------|------------------------|
| ICT-3DRFE [Stratou et al. 2011] | 3 | 23 | 2 | 1K | ✗ |
| Ultrastage [Zhou et al. 2023] | 3 | 100 | 32 | 8K | ✗ |
| RGCA [Martinez et al. 2024; Saito et al. 2024] | 460 | 4 | 110 | 4K | ✓ |
| Dynamic OLAT [Zhang et al. 2021b] | 114 | 4 | 1 | 1K | ✓ |
| <i>FaceOLAT</i> | 331 | 139 | 40 | 4K | ✓ |

these methods do generalize to novel identities, their photorealism and overall quality leave room for improvement [Sengupta et al. 2018; Shu et al. 2017; Zhou et al. 2019]. Modeling the complex light transport effects exhibited by human faces, as well as the sub-surface material properties of skin [Klehm et al. 2015] is a challenging task. This is why a different line of research, image-based relighting [Debevec et al. 2000] uses OLAT images captured with a light stage as a basis for relighting according to any given HDRI environment map. This approach has been extended to estimating reflectance from monocular images [B R et al. 2021b; Yamaguchi et al. 2018], based on parametric face models: FRF [B R et al. 2021b] aims to regress an OLAT basis for a given camera view but the parametric face model limits it to the face interior. Similarly, many methods are limited to portrait relighting without view synthesis [Meka et al. 2019; Nestmeyer et al. 2020; Pandey et al. 2021; Sun et al. 2020; Zeng et al. 2024; Zhang et al. 2025, 2021b, 2020] or subject-specific relighting [Bi et al. 2021].

Some methods learn face priors using 2D generative models, adapting them for photorealistic editing of pose, expression, and lighting [B R et al. 2021a; Buehler et al. 2021; Tewari et al. 2020a,b]. Specifically, PhotoApp [B R et al. 2021a] combines the advantages of lightstage OLAT data and a generative StyleGAN model, resulting in impressive identity generalization, simultaneous relighting and novel view synthesis of the full head. Nevertheless it suffers from view inconsistency and fails to preserve the original identity due to the absence of a consistent 3D facial geometry representation. In contrast, our method leverages a prior in volumetric space. This results in improved view consistency and ensures the preservation of the original identity throughout the editing process.

Neural field techniques have achieved high photorealism in view synthesis, but relighting remains an open problem. Srinivasan et al. [Srinivasan et al. 2021] show relighting in general scenes using co-located camera and light source, for a dense set of input images. More recent work [Boss et al. 2021; Rudnev et al. 2022; Zhang et al. 2021a] has extended this to images captured under unknown lighting, but these methods are scene-specific and cannot generalize to monocular inputs for an object category like faces or heads. Hong et al. [Hong et al. 2022] build a parametric head model conditioned on a lighting latent code. They disentangle lighting and reflectance by supervision on a multi-light dataset, but are limited by the sparse lighting variation in the training data. Kwak et al. [Kwak et al. 2022] attempt to decouple semantic attributes (including lighting) but suffer from significant view inconsistency due to the underlying

unsupervised training scheme. Other methods [Rao et al. 2022; Sun et al. 2021] achieve view synthesis and relighting of real people from sparse images: NeLF [Sun et al. 2021] relies on a pixelNeRF-inspired [Yu et al. 2021] architecture and thus struggles to capture global features. Holo-relighting [Mei et al. 2024] also leverages an EG3D prior to disentangle, delight and then relight a volumetric face from a single image, using lightstage data. However, it does not estimate intermediate OLAT images, but relies on neural networks to fully interpret an environment map, giving more opportunity for physically implausible results. It also requires test-time optimization which is computationally expensive and time consuming.

Some approaches [Ranjan et al. 2023; Tan et al. 2022] focus on relighting synthetic identities sampled from a learned latent space, but cannot relight a given real image. In contrast, LumiGan [Deng et al. 2023] can very well relight a given image, but while its adversarial self-supervised training leads to plausible-looking outputs, it does not supervise actual physical accuracy.

Both Lite2Relight [Rao et al. 2024a] and NeRFFaceLighting [Jiang et al. 2023] use a triplane representation [Chan et al. 2022]. While the former trains on a light stage dataset and produces physically accurate relighting, the latter trains on an in-the-wild dataset. However, NeRFFaceLighting uses spherical harmonics (SH), restricting its results to low-resolution lighting conditions and Lite2Relight samples the target lighting from the latent space of the 3D generator. In contrast, our method explicitly synthesizes OLAT images, which are then linearly combined according to an HDRI map.

The method that overcomes most of the aforementioned challenges is VoRF [Rao et al. 2023, 2022]. It is the closest related work in terms of problem setting. VoRF builds on a light stage dataset to learn physically accurate lighting. However, VoRF's face prior, learned from a relatively small number of light stage subjects, struggles to generalize to monocular inputs of unseen faces. Our method addresses this problem by combining a generative 3D face prior [Chan et al. 2022] with light transport learned from a lightstage dataset. This leads to 3D-consistent novel-view synthesis and physically accurate relighting.

3 FACEOLAT: A NEW LARGE-SCALE OLAT DATASET

Image-based relighting methods benefit from directly leveraging captured reflectance data without relying on any predefined material models. They achieve this by linearly combining One-Light-At-a-Time (OLAT) images. Given an HDR environment map specifying

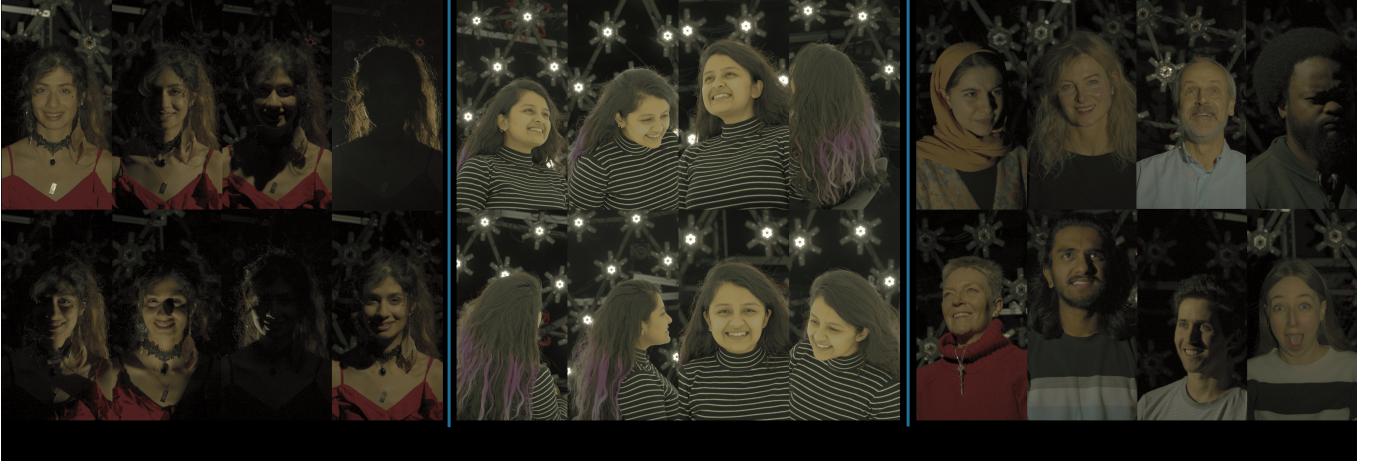


Fig. 2. Overview of the Dataset. The *FaceOLAT* lightstage dataset comprises 139 subjects captured from 40 camera viewpoints, resulting in 331 OLAT images per subject, illuminated by point light sources. Each OLAT image is captured at 4K resolution. A snapshot of the dataset is shown in the figure. A detailed description along with dataset demographics is provided in the supplemental material.

illumination, the relit image C is computed as: $C \approx \sum_{l \in I} f_l \cdot O(l)$, where I denotes the set of OLAT lighting directions, $O(l)$ is the OLAT image for lighting from direction l , and f_l represents the environment map weights.

To effectively train our model to synthesize OLAT images (Sec. 4), a high-quality multi-view OLAT dataset is essential. However, existing publicly available datasets are limited in scale and diversity [Saito et al. 2024; Zhang et al. 2021b]. Addressing this significant gap, we introduce *FaceOLAT*, a novel, large-scale OLAT dataset comprising 139 diverse subjects captured using a well-calibrated lightstage system. Each subject was recorded from 40 uniformly distributed viewpoints at 4K resolution under 331 OLAT lighting conditions, capturing four distinct facial expressions. The Fig. 2 provides an overview of the dataset. Our detailed dataset capture pipeline resolves practical challenges, such as minor involuntary subject movements during the 7s capture duration, by interleaving fully lit reference frames every 21 OLAT captures and employing optical flow-based alignment [Teed and Deng 2020]. Additional pre-processing includes precise calibration, detailed 3D reconstruction, and efficient background segmentation using BGMv2 and RMBGv2 [Lin et al. 2020; Zheng et al. 2024]. We partition the dataset into training and evaluation subsets, with 129 subjects designated for training and the remaining 10 for evaluation. Our dataset, which includes the preprocessing results, such as 3D reconstructions, will be publicly accessible. Additionally, the supplemental document contains additional information on demographics, preprocessing techniques, and dataset acquisition. We now detail our proposed methodology that leverages this dataset

4 METHOD

Given a single portrait image, our goal is to edit both viewpoint and illumination in a photorealistic and 3D-consistent manner. To achieve this, *3DPR* is trained on an OLAT dataset and operates in two stages: In the first stage, the input portrait is embedded into

the latent space of EG3D via an encoder-based GAN inversion process (see Sec. 4.1), enabling our framework to benefit from EG3D’s strong 3D generative prior. In the second stage, we introduce an OLAT-based reflectance module that synthesizes OLAT images using an efficient volumetric triplane representation [Chan et al. 2022]; this stage is described in detail in Sec. 4.2. During training, the reflectance module is supervised using ground-truth OLAT images from the lightstage dataset (see Sec. 4.3 and supplemental for additional training details). At inference time, *3DPR* takes a single RGB input image and synthesizes OLAT images for novel viewpoints, which are then linearly combined to approximate the target lighting condition (see Sec. 4.4).

4.1 3D Inversion

EG3D transforms a noise vector $\mathbf{z} \in \mathbb{R}^{1 \times 512}$ into an intermediate latent code $\mathbf{w} \in \mathbb{R}^{14 \times 512}$, which is passed to a StyleGAN2 generator G_{gen} [Karras et al. 2020] to produce tri-planar features $F_g \in \mathbb{R}^{96 \times 256 \times 256}$. These features encode both geometry and appearance, and serve as a compact 3D representation of the scene. They can be rendered to images from arbitrary viewpoints, by volume rendering. To obtain this feature representation from a single portrait image C , we employ a pre-trained encoder-based inversion network \mathcal{E} [Yuan et al. 2023], such that $F_g = \mathcal{E}(C)$ (see Fig. 3). The tri-plane features F_g are decoded by EG3D’s MLP-based decoder G_{dec} and rendered volumetrically from a given camera viewpoint v to produce a low-resolution RGB image $c_{\text{rgb}} \in \mathbb{R}^{3 \times 128 \times 128}$ and a high-frequency feature image $c_{\text{hf}} \in \mathbb{R}^{29 \times 128 \times 128}$. We specifically leverage c_{hf} , which encodes high-frequency appearance details, as input to the next stage of our pipeline for synthesizing high-resolution OLATs.

4.2 Learning Face Reflectance

To model facial reflectance, we aim to generate OLAT images for any light direction $\omega_i \in \mathbb{R}^3$. Given the encoded tri-plane feature map F_g from the inversion stage, we concatenate it with the light

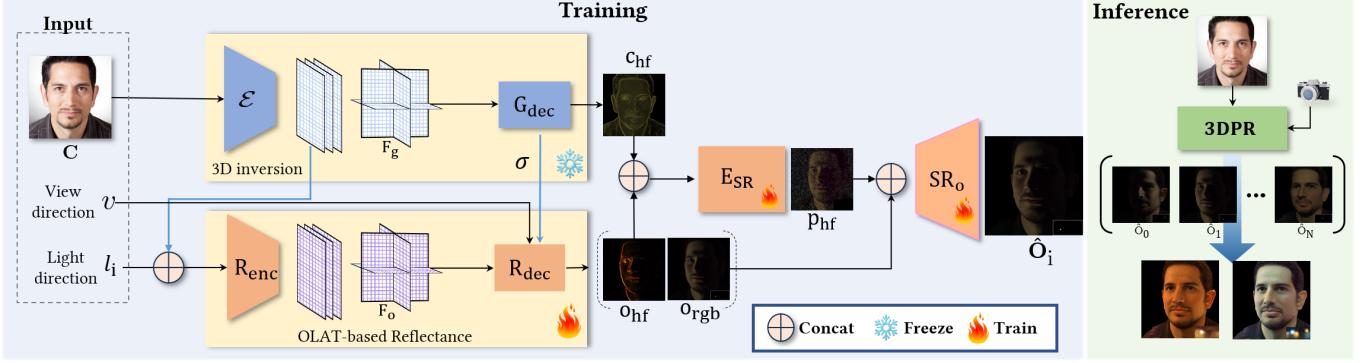


Fig. 3. Given one portrait image C , 3DPR renders the subject from a novel viewpoint under new lighting. **Left:** During the training stage, C is first fed into the 3D-aware encoder \mathcal{E} [Yuan et al. 2023] to produce tri-planar features F_g , concatenated with a given light direction l_i . Then, the concatenated features are fed into the *Reflectance Encoder* R_{enc} and *Reflectance Decoder* R_{dec} to render a low-resolution OLAT image o_{rgb} and high-frequency reflectance features o_{hf} . We further combine o_{hf} with the corresponding appearance features c_{hf} , to be fed into *SR Encoder* E_{SR} to obtain fused high frequency features p_{hf} . At the end, the *OLAT Super-Resolution* network SR_o produces a high-resolution OLAT image \hat{O}_i . In our architecture, we use pre-trained networks [Chan et al. 2022] for StyleGAN and G_{dec} and keep these modules frozen, and only train R_{enc} , R_{dec} , E_{SR} and SR_o on light stage dataset. **Right:** During inference, 3DPR takes a single portrait image, view and light direction as input, and synthesizes OLATs, which are then linear combined for novel illumination.

direction ω_i and pass the result to our OLAT encoder R_{enc} , which predicts reflectance-aware tri-plane features $F_o \in \mathbb{R}^{96 \times 256 \times 256}$ as: $F_o = R_{enc}(F_g, \omega_i)$. R_{enc} is based on a ResNet architecture [He et al. 2016], and the depth of 96 channels is critical for modeling complex skin-light interactions such as specularities, hard shadows, and subsurface scattering.

To synthesize a specific OLAT image from these features, we also incorporate the view direction v and use our OLAT decoder R_{dec} , a lightweight single-layer MLP, which outputs both a low-resolution RGB image $o_{rgb} \in \mathbb{R}^{3 \times 128 \times 128}$ and a high-frequency reflectance feature map $o_{hf} \in \mathbb{R}^{29 \times 128 \times 128}$ via NeRF-style volume rendering: $o_{rgb}, o_{hf} = R_{dec}(F_o, v)$. If we directly fed o_{rgb} and o_{hf} into the super-resolution module SR_o , that module could easily overfit to the relatively small number of subjects in the OLAT dataset. To prevent this, we introduce a feature fusion module E_{SR} , which combines o_{hf} with the high-frequency identity features c_{hf} obtained from the inversion stage (see Sec. 4.1). The fused feature map $p_{hf} \in \mathbb{R}^{29 \times 128 \times 128}$ is computed as: $p_{hf} = E_{SR}(c_{hf} \oplus o_{hf})$, where \oplus denotes channel-wise concatenation. Since SR_o was pretrained to use c_{hf} , the module E_{SR} quickly learns to forward much of the information from c_{hf} to SR_o . Only for lighting information is SR_o forced to rely on o_{rgb} and o_{hf} . Information about the identity of the subject however, is contained in c_{hf} from the start of training, which acts as a kind of regularization that prevents SR_o from trying to derive the identity from o_{rgb} and o_{hf} . Our experiments show (see Sec. 5.3 and Tab. 4) that introducing E_{SR} does improve quality. Finally, we combine o_{rgb} and p_{hf} and pass them through the OLAT super-resolution network SR_o to synthesize the final high-resolution OLAT image $\hat{O} = SR_o(o_{rgb} \oplus p_{hf}) \in \mathbb{R}^{3 \times 512 \times 512}$. This pipeline enables accurate and generalizable OLAT image synthesis, which is used for relighting under arbitrary environment maps.

4.3 Loss Functions

Reconstruction Loss. Given ground-truth OLAT images O_i from the light stage, we supervise our predicted OLATs \hat{O}_i using the $L1$ loss $\mathcal{L}_O = \|\hat{O}_i - O_i\|_1$. This encourages direct correspondence between the predicted and reference images at the pixel level.

ID-MRF Loss. Relying solely on \mathcal{L}_O is insufficient, as misalignments introduced by the inversion stage lead to subtle errors that cannot be corrected by pixel-wise losses. Adversarial losses are also unsuitable here, given the limited number of subjects in the dataset (130), which risks discriminator overfitting and unstable training. To address this, we adopt the Implicit Diversified Markov Random Field (ID-MRF) loss introduced by Wang *et al.* [Wang et al. 2018]. This loss encourages local feature-level similarity by minimizing the patch-wise nearest-neighbor distances between \hat{O}_i and O_i in a feature space extracted from a pre-trained VGG19 network [Simonyan and Zisserman 2015]. The ID-MRF loss is computed as $\mathcal{L}_{MRF} = \mathcal{L}_M(\Phi_1(\hat{O}_i, O_i)) + \mathcal{L}_M(\Phi_2(\hat{O}_i, O_i))$, where Φ_1 and Φ_2 correspond to activations from the conv3_2 and conv4_2 layers of VGG19, respectively, and \mathcal{L}_M denotes the matching function.

Final Objective. The total loss is given by $\mathcal{L} = \mathcal{L}_O + 0.3\mathcal{L}_{MRF}$, where the ID-MRF term is weighted to balance reconstruction accuracy with local structural detail. As shown in Sec. 5.3, this loss formulation recovers high-frequency details more effectively than commonly used perceptual losses such as LPIPS [Zhang et al. 2018].

4.4 Testing

At test time, given a monocular portrait image, we employ the encoder-based inversion network to derive F_g . This feature map is processed through the reflectance network, which synthesizes OLAT images corresponding to a specified lighting direction and camera viewpoint. The design of the 3D generative model enables

the generation of OLAT images in a single forward pass, eliminating the need for computationally intensive test-time optimization processes like in VoRF [Rao et al. 2023, 2022] or NFL [Jiang et al. 2023]. Finally, exploiting the additivity of light transport, the predicted OLAT images can be linearly combined with the desired HDR environment maps (see Sec. 3). This allows relighting of the portrait under the desired illumination conditions while supporting novel viewpoints. Our inference pipeline is visualized in Fig. 3.

5 RESULTS AND DISCUSSION

We evaluate *3DPR* on two categories of datasets: For qualitative evaluation, we assess simultaneous view synthesis and relighting on in-the-wild subjects from RAVDESS [Livingstone and Russo 2018], Flickr [Shih et al. 2014], and FFHQ [Karras et al. 2020]. For quantitative evaluation, we use *WeyrichOLAT* [Weyrich et al. 2006] and *FaceOLAT*, where we construct input-reference image pairs by selecting 10 unseen subjects and relighting them using 10 novel HDR environment maps (Sec. 3). Sec. 5.1 presents qualitative results of simultaneous relighting and view synthesis on in-the-wild data using *3DPR*. We further compare our approach both qualitatively and quantitatively to state-of-the-art methods in Sec. 5.2. Finally, we analyze key design choices through ablation studies in Sec. 5.3.

5.1 Qualitative Evaluation

Fig. 4 presents qualitative results for simultaneous view synthesis and relighting. *3DPR* preserves the linear nature of light transport and accurately reproduces illumination effects such as hard shadows, specular highlights, and self-shadowing, all consistent with the reference images. For instance, observe the shading details on the nose and cheek regions in the OLAT renderings (2nd to 4th columns). Our method leverages a rich 3D generative prior and models facial reflectance through OLAT-based reflectance module, enabling it to capture the complex interplay between light, face geometry and skin. This design allows *3DPR* to faithfully relight subjects while preserving facial structure and expressions from the input. Overall, our qualitative results show that *3DPR* produces accurate relighting that is 3D-consistent across diverse subjects.

5.2 Baseline Comparisons

We compare *3DPR* against several state-of-the-art approaches for simultaneous view synthesis and relighting:

- **PhotoApp** [B R et al. 2021a] leverages the generative prior of StyleGAN2 [Karras et al. 2020] to learn a latent space transformation for portrait relighting.
- **VoRF** [Rao et al. 2023, 2022] trains an autodecoder-based NeRF [Mildenhall et al. 2020] to learn a volumetric reflectance field of human heads.
- **NeRFFaceLighting (NFL)** [Jiang et al. 2023] disentangles lighting and appearance using EG3D-based design principles and performs relighting via an SH-based representation.
- **Lite2Relight (L2R)** [Rao et al. 2024b] employs an MLP-based reflectance network that probes the latent space of EG3D to enable controllable relighting.

For a fair comparison, we evaluate all methods on *WeyrichOLAT*, a well-established (non-open-source) benchmark, using the same

train-test split as L2R to ensure standardized evaluation. In Tab. 2, all baselines and our method ($3DPR_w$) are trained and evaluated on *WeyrichOLAT*. Thus, the observed improvements of *3DPR* arise from the effective way of combining 3D generative priors with OLAT representation. Further, in Tab. 3, we benchmark our method against the strongest baselines on *FaceOLAT* and retrain L2R on our dataset to ensure a fair comparison.

We measure relighting accuracy using multiple metrics: LPIPS [Zhang et al. 2018], RMSE, DISTS [Ding et al. 2020], PSNR, SSIM, and identity consistency (ID), computed as the cosine similarity between MagFace [Meng et al. 2021] embeddings of the relit and ground-truth images.

Quantitative comparisons with all baselines on *WeyrichOLAT* are shown in Tab. 2 and Fig. 5. To further validate generalization and performance, we also evaluate *3DPR* against the two strongest baselines, NFL and L2R, on our new *FaceOLAT* dataset; results are presented in Tab. 3 and Fig. 6. For this evaluation, we retrain L2R following its original training protocol.

Table 2. Quantitative Comparisons: NeLF [Sun et al. 2021], PhotoApp [B R et al. 2021a], VoRF [Rao et al. 2023, 2022], NeRFFaceLighting [Jiang et al. 2023] and Lite2Relight [Rao et al. 2024b]. Metrics evaluated on the *WeyrichOLAT* test set, for simultaneous view synthesis and relighting.

| | SSIM↑ | LPIPS↓ | RMSE↓ | DISTS↓ | PSNR↑ | ID↑ |
|-----------------------------------|-------------|---------------|---------------|---------------|--------------|--------------|
| NeLF | 0.75 | 0.4874 | 0.2466 | 0.2212 | 19.72 | 0.798 |
| PhotoApp | 0.72 | 0.4163 | 0.1988 | 0.2031 | 29.13 | 0.853 |
| VoRF | 0.69 | 0.3253 | 0.1967 | 0.1934 | 20.21 | 0.860 |
| NeRFFaceLighting | 0.79 | 0.2171 | 0.2393 | 0.2107 | 27.24 | 0.892 |
| Lite2Relight | 0.83 | 0.2492 | 0.1841 | 0.1719 | 28.27 | 0.936 |
| Ours ($3DPR_w$) | 0.87 | 0.1828 | 0.1332 | 0.1689 | 28.69 | 0.942 |

Table 3. Quantitative Comparisons: NeRFFaceLighting [Jiang et al. 2023], Lite2Relight [Rao et al. 2024b]. Performance metrics are evaluated on the *FaceOLAT* test dataset, for simultaneous view synthesis and relighting.

| | SSIM↑ | LPIPS↓ | RMSE↓ | DISTS↓ | PSNR↑ | ID↑ |
|-----------------------------------|-------------|---------------|---------------|---------------|--------------|--------------|
| NeRFFaceLighting | 0.77 | 0.2385 | 0.2926 | 0.2193 | 16.97 | 0.906 |
| Lite2Relight | 0.79 | 0.2506 | 0.2619 | 0.20861 | 16.72 | 0.910 |
| Ours ($3DPR_o$) | 0.83 | 0.1996 | 0.1801 | 0.1751 | 21.02 | 0.943 |

Relighting and Novel View Synthesis. Tabs. 2 and 3 report quantitative comparisons, while Fig. 5 shows qualitative examples. These results confirm that *3DPR* outperforms competing methods both numerically and visually. PhotoApp lacks an explicit 3D representation, often resulting in identity inconsistencies (e.g., altered jawlines) under novel viewpoint and fails to capture accurate illumination effects. Despite this, it achieves surprisingly high PSNR scores, largely due to the high visual quality of StyleGAN2-generated images. However, PSNR is not well-suited for measuring the variations in complex illuminations and thus cannot properly account for the nuances of human visual perception [Zhang et al. 2018]. VoRF struggles with accurate OLAT synthesis due to inference-time optimization that modifies the learned volumetric reflectance representation. Furthermore, its limited face prior restricts its ability to generalize to unseen identities. NFL also suffers from challenges in relighting



Fig. 4. Simultaneous view synthesis and relighting. The top row shows a reference portrait from *FaceOLAT*, rendered under selected OLAT directions and corresponding environment map-based relighting, all computed from the 331 OLATs. In the following rows, the first column presents the “in-the-wild” input. Columns 2–4 show OLAT renderings from novel viewpoints, with the light source direction illustrated in the inset. Columns 5–7 show relit outputs from *3DPR*, under novel viewpoints and HDRI environment maps, as shown in the insets. This visualization demonstrates *3DPR*’s ability to simultaneously perform relighting and viewpoint editing on “in-the-wild” images, producing sharp specular highlights, self-shadows and subsurface scattering.



Fig. 5. Baseline Comparisons with PhotoApp [B R et al. 2021a], VoRF [Rao et al. 2023, 2022], NeRFFaceLighting[Jiang et al. 2023] and Lite2Relight [Rao et al. 2024a]. We compare these approaches against *3DPR_w* trained on *WeyrichOLAT* [Weyrich et al. 2006]. We demonstrate that *3DPR* is more effective in preserving the identity of the subjects and produces relighting that more closely resembles the ground truth than other approaches.

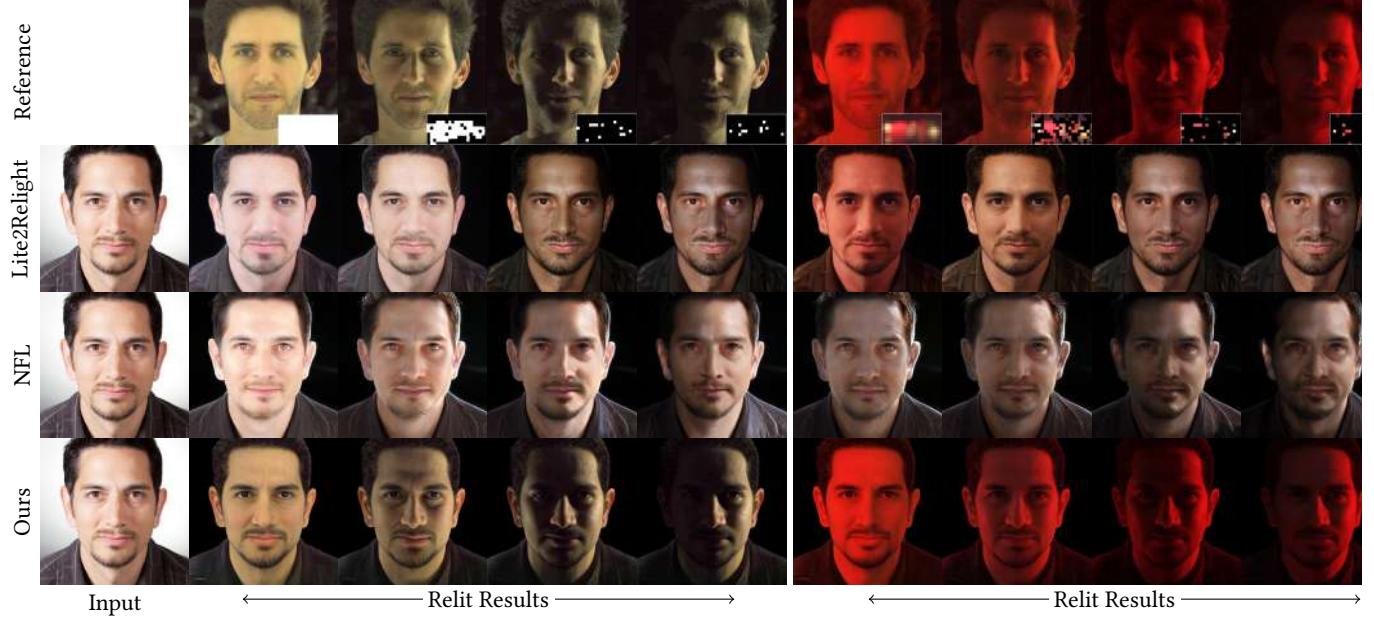


Fig. 6. Benefits of OLAT-based Relighting We compare *3DPR* with Lite2Relight (L2R) and NeRFFaceLighting (NFL) under increasingly sparse and colored lighting. Each row shows relit results from one method. The first column is the input *in-the-wild* image; the top row shows the target lighting from a reference subject. The next four columns correspond to sparse white lighting, created by progressively blacking out pixels in an all-white HDR environment map. The final four columns correspond to an indoor environment map, which are made increasingly sparse in a similar manner. As lighting becomes sparser, both baselines degrade noticeably. Generally, NFL suffers from shading artifacts and color inconsistencies due to its low-frequency SH representation and inaccurate lighting-albedo disentanglement. This is more pronounced under colored lights. L2R also struggles to handle sparse conditions, failing to reproduce sharp shadows or specular highlights and thus yielding inconsistent relighting. In contrast, *3DPR* maintains identity and reproduces shadows and specular highlights consistently across all scenarios, thanks to its explicit OLAT-based reflectance modeling. These results highlight the robustness and generalization ability of our approach.

accuracy and identity preservation, mainly due to its two-stage optimization pipeline. As visualized in Fig. 5, input lighting is often baked into the albedo, and NFL fails to capture high-frequency details. Finally, L2R relights portraits by predicting latent vectors through probing EG3D’s manifold, but lacks precise lighting control. As shown in Fig. 5, it fails to reproduce soft shadows and yields inconsistent facial illumination relative to the ground truth. In contrast, *3DPR* leverages the additive nature of light transport, enabling fine-grained lighting control while faithfully preserving identity – even under novel viewpoints. This demonstrates the method’s robustness in modeling complex light interactions and its ability to maintain photorealism and consistency across a wide range of subjects and conditions. Please refer to supplementary materials for additional comparisons.

Significance of OLAT-based Relighting. Both NFL and L2R leverage the EG3D generative prior to directly predict relit portraits. In contrast, our approach not only incorporates the EG3D prior but also explicitly models facial reflectance via OLAT prediction and linear combination with environment maps. This design offers fine-grained control over lighting and enables relighting under *any* lighting condition, including artistic, sparse, or non-natural illumination setups commonly used in cinematic and indoor environments.

We hypothesize that such conditions fall outside the training distribution of EG3D, which is primarily trained on *in-the-wild* images with natural illuminations. To evaluate this, we create increasingly sparse lighting conditions by randomly replacing environment map pixels with zero (see the first row in Fig. 6). As the lighting becomes sparser (left to right), both NFL and L2R exhibit noticeable performance degradation, confirming our hypothesis. While L2R performs reasonably under dense lighting, it fails under sparse set-ups due to out-of-distribution target illuminations. NFL, limited by its SH-based lighting representation and inaccurate albedo-lighting disentanglement, struggles to reproduce high-frequency effects and breaks down under colored light. In contrast, *3DPR* remains robust across lighting conditions, accurately reproducing shadows, specularities, and other complex effects even under sparse or unconventional illumination. See the supplementary material and caption of Fig. 6 for detailed analysis and visual examples.

Quality of OLATs. We quantitatively evaluate the accuracy of OLAT renderings produced by *3DPR*, obtaining significantly improved results (SSIM: 0.88, LPIPS: 0.1753, PSNR: 28.70) compared to the state-of-the-art method VoRF (SSIM: 0.71, LPIPS: 0.3148, PSNR: 20.43). Qualitative results in Fig. 7 demonstrate that our synthesized OLATs generalize robustly to both our evaluation dataset and “*in-the-wild*” subjects. Our method effectively preserves the additive

Table 4. Quantitative Results: Design Ablations: We report the influence of various losses and E_{SR} . The performance metrics are evaluated for relighting performance across 10 unseen subjects [Weyrich et al. 2006].

| | SSIM \uparrow | LPIPS \downarrow | RMSE \downarrow | DISTS \downarrow | PSNR \uparrow |
|---------------------------------------|-----------------|--------------------|-------------------|--------------------|-----------------|
| \mathcal{L}_O | 0.70 | 0.2563 | 0.1631 | 0.2745 | 21.43 |
| $\mathcal{L}_O + \mathcal{L}_{LPIPS}$ | 0.75 | 0.1978 | 0.1441 | 0.2005 | 23.26 |
| w/o E_{SR} | 0.85 | 0.2046 | 0.1465 | 0.1809 | 28.68 |
| $\mathcal{L}_O + \mathcal{L}_{MRF}$ | 0.87 | 0.1828 | 0.1332 | 0.1689 | 28.69 |

properties of light transport, and accurately reproduces complex illumination effects, including specular highlights, hard shadows, and subsurface scattering effects.

5.3 Ablation Study

Timing Evaluations: On an NVIDIA 3090 GPU, 3DPR synthesizes the complete set of 331 OLAT images in approximately 30.49 s. We observe that this number can be reduced to 150 OLATs with minimal degradation in quality, reducing the runtime to around 13.8 s. Since OLAT synthesis is fully parallelizable, using an H100 GPU reduces the time for generating all 331 OLATs to just 7.74 s. Importantly, in 3DPR, OLATs for a given subject and viewpoint are rendered only once. Once these are generated, relighting under a novel environment map takes just 0.24 s. Although our method is not as fast as Lite2Relight, we believe it offers a practical balance between efficiency and quality. Compared to optimization-based baselines, it is competitively fast and significantly outperforms all baselines, including Lite2Relight, in relighting fidelity. Furthermore, because our approach models a continuous reflectance field, it naturally supports flexible upsampling of lighting resolution. For instance, we can synthesize 1324 OLAT images in just 34.64 s on a H100 GPU, demonstrating the scalability of our method.

Significance of SR Encoder: Given the relatively small size of the lightstage training dataset with its limited subject diversity, the reflectance module, particularly the SR_o network, tends to overfit during relighting. This overfitting is evident in artifacts observed on subjects not included in the lightstage dataset (refer to supplementary materials). To address this issue and improve generalization, we combine robust high-frequency face prior features c_{hf} with high-frequency reflectance features o_{hf} using E_{SR} . This integration mitigates the memorization of training subjects and enhances overall performance, as summarized in Tab. 4.

Significance of \mathcal{L}_{MRF} : We quantitatively analyze the impact of different loss functions employed in training 3DPR and summarize the results in Tab. 4. It is evident that supervision with only the L_1 loss is insufficient to produce high-quality relighting results. While combining \mathcal{L}_{LPIPS} [Zhang et al. 2018] with per-pixel L_1 loss is a commonly adopted approach, this combination still leads to suboptimal performance, primarily due to this metric missing high-frequency details. While \mathcal{L}_{MRF} alone produces satisfactory results, we find that combining it with \mathcal{L}_O further enhances performance and accelerates convergence.

6 LIMITATIONS

Despite strong results, 3DPR has several limitations that suggest directions for future work. (i) Although *FaceOLAT* provides full-head coverage, our relighting quality degrades on the back of the head; this stems from the EG3D prior, whose representation does not reliably cover regions outside the front-face region. Integrating a more comprehensive 3D generative prior with *FaceOLAT* could address this limitation and enable full-head relighting. (ii) The scope of this work is limited to facial reflectance (face, eyes, scalp hair). Consequently, headgear and accessories (e.g., helmets, sunglasses) are out of domain, as 3DPR does not synthesize OLATs for these materials. Extending the approach with reflectance priors for a broader set of objects and materials is a promising direction. (iii) Our method inherits EG3D’s difficulty in consistently modeling fine hair fibers: novel-view synthesis can exhibit local inconsistencies in the hair region; small misalignments between OLAT renderings may accumulate into noise or flicker when linearly combined; and the super-resolution stage can introduce strand “popping” under head rotation. Addressing these effects will require stronger high-frequency priors and alignment strategies tailored to hair. (iv) Finally, despite conditioning the OLAT decoder R_{dec} on the viewing direction, view-dependent effects (e.g., on the nose bridge and cheeks) are relatively subdued (see supplementary video). While our OLAT quality (Fig. 7) and overall relighting fidelity surpass the baselines, these subtle view-dependent cues contribute weakly to the training objective and are therefore not strongly expressed; improving supervision and objectives for view dependence remains important future work.

7 CONCLUSION

In this paper, we presented 3DPR, a unified framework that addresses the challenge of editing both illumination and viewpoint in portrait images using a single monocular input. Our method draws on the strength of a pre-trained 3D-aware generator, enabling it to learn a rich facial prior. We used *FaceOLAT*, a new lightstage dataset, in the training of a novel reflectance network, which allows 3DPR to accurately capture facial reflectance through HDR OLAT images. To enhance the quality of full-head portrait relighting, we use a combination of a reconstruction loss and ID-MRF loss. Our quantitative and qualitative evaluations show that our method exhibits promising advantages over the existing state-of-the-art approaches, particularly in terms of achieving 3D-consistent editing, simulating accurate light transport effects and controlling novel illumination. We believe that our work contributes to the ongoing research in this field, and we hope it will inspire further exploration and advancements in monocular portrait image editing.

ACKNOWLEDGMENTS

This work was supported by the ERC Consolidator Grant 4DReply (770784) and Saarbrücken Research Center for Visual Computing, Interaction, and AI. We thank Oleksandr Sotnychenko for helping us with setting up data capture. Finally, we thank Shrisha Bharadwaj for discussions, proofreading and innumerable support.



Fig. 7. OLAT evaluation. The top row shows ground-truth OLAT images from the testset of *FaceOLAT*. The second row shows 3DPR’s prediction of this ground truth, showing high accuracy. Furthermore, OLAT renderings for in-the-wild portraits are displayed in Rows 3 and 4. Notably, our method’s predictions exhibit a close resemblance to the actual light direction of the reference. Additionally, our approach consistently captures the intricate details of hard shadows, subsurface scattering effects (see orange boxes) and specular highlights (see blue dashed boxes) across different subjects.

REFERENCES

- Mallikarjun B R, Ayush Tewari, Abdallah Dib, Tim Weyrich, Bernd Bickel, Hans Peter Seidel, Hanspeter Pfister, Wojciech Matusik, Louis Chevallier, Mohamed A Elgarib, and Christian Theobalt. 2021a. PhotoApp: Photorealistic appearance editing of head portraits. *ACM Transactions on Graphics* 40, 4 (2021).
- Mallikarjun B R, Ayush Tewari, Tae-Hyun Oh, Tim Weyrich, Bernd Bickel, Hans-Peter Seidel, Hanspeter Pfister, Wojciech Matusik, Mohamed Elgarib, and Christian Theobalt. 2021b. Monocular Reconstruction of Neural Face Reflectance Fields. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Sai Bi, Stephen Lombardi, Shunsuke Saito, Tomas Simon, Shih-En Wei, Kevyn Mcphail, Ravi Ramamoorthi, Yaser Sheikh, and Jason Saragih. 2021. Deep Relightable Appearance Models for Animatable Faces. *ACM Transactions on Graphics*, Article 89 (2021), 15 pages.
- Mark Boss, Raphael Braun, Varun Jampani, Jonathan T. Barron, Ce Liu, and Hendrik P.A. Lensch. 2021. NeRD: Neural Reflectance Decomposition from Image Collections. In *The IEEE International Conference on Computer Vision (ICCV)*.
- Marcel C. Buehler, Abhimitra Meka, Gengyan Li, Thabo Beeler, and Otmar Hilliges. 2021. VariTex: Variational Neural Face Textures. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.
- Adrian Bulat and Georgios Tzimiropoulos. 2017. How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks). In *International Conference on Computer Vision*.
- Eric R. Chan, Connor Z. Lin, Matthew A. Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas Guibas, Jonathan Tremblay, Sameh Khamis, Tero Karras, and Gordon Wetzstein. 2022. Efficient Geometry-aware 3D Generative Adversarial Networks. In *CVPR*.
- Sreenithy Chandran, Yannick Hold-Geoffroy, Kalyan Sunkavalli, Zhixin Shu, and Suren Jayasuriya. 2022. Temporally Consistent Relighting for Portrait Videos. In *The IEEE Winter Conference on Applications of Computer Vision (WACV) Workshops*. 719–728.
- Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. 2000. Acquiring the reflectance field of a human face. In *Annual conference on Computer graphics and interactive techniques*.
- Boyang Deng, Yifan Wang, and Gordon Wetzstein. 2023. LumiGAN: Unconditional Generation of Relightable 3D Human Faces. In *arXiv*.
- Yu Deng, Jiaolong Yang, Jianfeng Xiang, and Xin Tong. 2022. GRAM: Generative Radiance Manifolds for 3D-Aware Image Generation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Keyan Ding, Kede Ma, Shiqi Wang, and Eero P. Simoncelli. 2020. Image Quality Assessment: Unifying Structure and Texture Similarity. *CoRR* abs/2004.07728 (2020). <https://arxiv.org/abs/2004.07728>
- Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680.
- Jitao Gu, Lingjie Liu, Peng Wang, and Christian Theobalt. 2022. StyleNeRF: A Style-based 3D Aware Generator for High-resolution Image Synthesis. In *International Conference on Learning Representations*.
- Yang Haotian, Zheng Mingwu, Ma ChongYang, Lai Yu-Kun, Wan Pengfei, and Huang Haibin. 2024. VRMM: A Volumetric Relightable Morphable Head Model. In *SIGGRAPH 2024 Conference Proceedings*.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (Las Vegas, NV, USA) (CVPR '16)*. IEEE, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- Mingming He, Pascal Clausen, Ahmet Levent Tasel, Li Ma, Oliver Pilarski, Wenqi Xian, Laszlo Rikker, Xueming Yu, Ryan Burgert, Ning Yu, and Paul Debevec. 2024. DiffRelight: Diffusion-Based Facial Performance Relighting. In *SIGGRAPH Asia 2024 Conference Papers (SA '24)*. Association for Computing Machinery, New York, NY, USA, Article 11, 12 pages. <https://doi.org/10.1145/3680528.3687644>

- Yang Hong, Bo Peng, Haiyao Xiao, Ligang Liu, and Juyong Zhang. 2022. HeadNeRF: A Real-time NeRF-based Parametric Head Model. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Kaiwen Jiang, Shu-Yu Chen, Hongbo Fu, and Lin Gao. 2023. NeRFFaceLighting: Implicit and Disentangled Face Lighting Representation Leveraging Generative Prior in Neural Radiance Fields. *ACM Transactions on Graphics (TOG)* (2023).
- Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. 2018. Progressive Growing of GANs for Improved Quality, Stability, and Variation. In *International Conference on Learning Representations (ICLR)*.
- Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. Analyzing and Improving the Image Quality of StyleGAN. In *Proc. CVPR*.
- Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.). <http://arxiv.org/abs/1412.6980>
- Oliver Klehm, Fabrice Rousselle, Marios Papas, Derek Bradley, Christophe Hery, Bernd Bickel, Wojciech Jarosz, and Thabo Beeler. 2015. Recent Advances in Facial Appearance Capture. *Computer Graphics Forum (Proceedings of Eurographics - State of the Art Reports)* 34, 2 (May 2015), 709–733. <https://doi.org/10.1111/cgf.12779>
- Jeong-gi Kwak, Yuanming Li, Dongsik Yoon, Donghyeon Kim, David Han, and Hanseok Ko. 2022. Injecting 3D Perception of Controllable NeRF-GAN into StyleGAN for Editable Portrait Image Synthesis. In *European Conference on Computer Vision*. Springer, 236–253.
- Alexandros Lattas, Stylianos Moschoglou, Stylianos Ploumpis, Baris Gecer, Abhijeet Ghosh, and Stefanos P Zafeiriou. 2021. AvatarMe++: Facial Shape and BRDF Inference with Photorealistic Rendering-Aware GANs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021).
- Junxuan Li, Shunsuke Saito, Tomas Simon, Stephen Lombardi, Hongdong Li, and Jason Saragih. 2023. MEGANE: Morphable Eyeglass and Avatar Network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 12769–12779.
- Shanchuan Lin, Andrey Ryabtsev, Soumyadip Sengupta, Brian Curless, Steve Seitz, and Ira Kemelmacher-Shlizerman. 2020. Real-Time High-Resolution Background Matting. *arXiv* (2020), arXiv–2012.
- Steven R. Livingstone and Frank A. Russo. 2018. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVEDS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PLOS ONE* 13, 5 (05 2018), 1–35. <https://doi.org/10.1371/journal.pone.0196391>
- Julieta Martinez, Emily Kim, Javier Romero, Timur Bagautdinov, Shunsuke Saito, Shouyu Yu, Stuart Anderson, Michael Zollhöfer, Te-Li Wang, Shaojie Bai, Chenghui Li, Shih-En Wei, Rohan Joshi, Wyatt Borsos, Tomas Simon, Jason Saragih, Paul Theodosis, Alexander Greene, Anjani Josyula, Silvio Mana Maeta, Andrew I. Jewett, Simon Venshtain, Christopher Heilmann, Yueh-Tung Chen, Sidi Fu, Mohamed Ezzeldin A. Elshaer, Tingfang Du, Longhua Wu, Shen-Chi Chen, Kai Kang, Michael Wu, Youssef Emad, Steven Longay, Ashley Brewer, Hitesh Shah, James Booth, Taylor Koska, Kayla Haidle, Matt Andromalos, Joanna Hsu, Thomas Dauer, Peter Selednik, Tim Godisart, Scott Ardisson, Matthew Cipperly, Ben Humberston, Lon Farr, Bob Hansen, Peihong Guo, Dave Braun, Steven Krenn, He Wen, Lucas Evans, Natalia Fadeeva, Matthew Stewart, Gabriel Schwartz, Divam Gupta, Gyeongsik Moon, Kaiwen Guo, Yuan Dong, Yichen Xu, Takaaki Shiratori, Fabian Prada, Bernardo R. Pires, Bo Peng, Julia Buffalini, Autumn Trimble, Kevyn McPhail, Melissa Schoeller, and Yaser Sheikh. 2024. Codec Avatar Studio: Paired Human Captures for Complete, Driveable, and Generalizable Avatars. *NeurIPS Track on Datasets and Benchmarks* (2024).
- Yiqun Mei, Yu Zeng, He Zhang, Zhixin Shu, Xuaner Zhang, Sai Bi, Jianming Zhang, HyunJoon Jung, and Vishal M Patel. 2024. Holo-Relighting: Controllable Volumetric Portrait Relighting from a Single Image. *arXiv preprint arXiv:2403.09632* (2024).
- Abhimitra Meka, Christian Häne, Rohit Pandey, Michael Zollhöfer, Sean Fanello, Graham Fyffe, Adarsh Kowdle, Xueming Yu, Jay Busch, Jason Dourgarian, Peter Denny, Sofien Bouaziz, Peter Lincoln, Matt Whalen, Geoff Harvey, Jonathan Taylor, Shahram Izadi, Andrea Tagliasacchi, Paul Debevec, Christian Theobalt, Julien Valentini, and Christoph Rhemann. 2019. Deep Reflectance Fields: High-Quality Facial Reflectance Field Inference from Color Coded Illumination. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)* (2019).
- Qiang Meng, Shichao Zhao, Zhida Huang, and Feng Zhou. 2021. MagFace: A universal representation for face recognition and quality assessment. In *CVPR*.
- Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *European Conference on Computer Vision (ECCV)*.
- Thomas Nestmeyer, Jean-François Lalonde, Iain Matthews, and Andreas M Lehrmann. 2020. Learning Physics-guided Face Relighting under Directional Light. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Rohit Pandey, Sergio Orts-Escalano, Chloe LeGendre, Christian Haene, Sofien Bouaziz, Christoph Rhemann, Paul Debevec, and Sean Fanello. 2021. Total Relighting: Learning to Relight Portraits for Background Replacement. *ACM Transactions on Graphics (Proceedings SIGGRAPH)* (2021).
- Anurag Ranjan, Kwang Moo Yi, Jen-Hao Rick Chang, and Oncel Tuzel. 2023. FaceLit: Neural 3D Relightable Faces. In *CVPR*. <https://arxiv.org/abs/2303.15437>
- Pramod Rao, Mallikarjun B. R, Gereon Fox, Tim Weyrich, Bernd Bickel, Hanspeter Pfister, Wojciech Matusik, Fangneng Zhan, Ayush Tewari, Christian Theobalt, and Elgharib Mohamed. 2023. A Deeper Analysis of Volumetric Relightable Faces. *International Journal of Computer Vision* (10 2023), 1–19. <https://doi.org/10.1007/s11263-023-01899-3>
- 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. 2022. VoRF: Volumetric Relightable Faces. (2022).
- Pramod Rao, Gereon Fox, Abhimitra Meka, Mallikarjun B R, Fangneng Zhan, Tim Weyrich, Bernd Bickel, Hans-Peter Seidel, Hanspeter Pfister, Wojciech Matusik, Mohamed Elgharib, and Christian Theobalt. 2024a. Lite2Relight: 3D-aware Single Image Portrait Relighting. (2024).
- Pramod Rao, Gereon Fox, Abhimitra Meka, Mallikarjun B R, Fangneng Zhan, Tim Weyrich, Bernd Bickel, Hans-Peter Seidel, Hanspeter Pfister, Wojciech Matusik, Mohamed Elgharib, and Christian Theobalt. 2024b. Lite2Relight: 3D-aware Single Image Portrait Relighting. (2024).
- Viktor Rudnev, Mohamed Elgharib, William Smith, Lingjie Liu, Vladislav Golyanik, and Christian Theobalt. 2022. NeRF for Outdoor Scene Relighting. In *European Conference on Computer Vision (ECCV)*.
- Shunsuke Saito, Gabriel Schwartz, Tomas Simon, Junxuan Li, and Giljoo Nam. 2024. Relightable Gaussian Codec Avatars. In *CVPR*.
- Kripasindhu Sarkar, Marcel C. Bühlert, Gengyan Li, Daoye Wang, Delio Vicini, Jérémie Rivière, Yinda Zhang, Sergio Orts-Escalano, Paulo Gotardo, Thabo Beeler, and Abhimitra Meka. 2023. LitNeRF: Intrinsic Radiance Decomposition for High-Quality View Synthesis and Relighting of Faces. In *SIGGRAPH Asia 2023 Conference Papers (<conf-loc>, <city>Sydney</city>, <state>NSW</state>, <country>Australia</country>, </conf-loc>) (SA '23)*. Association for Computing Machinery, New York, NY, USA, Article 42, 11 pages. <https://doi.org/10.1145/3610548.3618210>
- Soumyadip Sengupta, Angjoo Kanazawa, Carlos D. Castillo, and David W. Jacobs. 2018. SISNet: Learning Shape, Refractance and Illuminance of Faces in the Wild. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- YiChang Shih, Sylvain Paris, Connally Barnes, William T. Freeman, and Frédéric Durand. 2014. Style transfer for headshot portraits. *ACM Trans. Graph.* 33, 4, Article 148 (July 2014), 14 pages. <https://doi.org/10.1145/2601097.2601137>
- Z. Shu, E. Yumer, S. Hadap, K. Sunkavalli, E. Shechtman, and D. Samaras. 2017. Neural Face Editing with Intrinsic Image Disentangling. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *International Conference on Learning Representations*.
- Vanessa Sklyarova, Egor Zakharov, Otmar Hilliges, Michael J Black, and Justus Thies. 2023. HAAR: Text-Conditioned Generative Model of 3D Strand-based Human Hairstyles. *ArXiv* (Dec 2023).
- Pratul P. Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T. Barron. 2021. NeRV: Neural Reflectance and Visibility Fields for Relighting and View Synthesis. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Giota Stratou, Abhijeet Ghosh, Paul Debevec, and Louis-Philippe Morency. 2011. Effect of illumination on automatic expression recognition: A novel 3D relightable facial database. In *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 611–618. <https://doi.org/10.1109/FG.2011.57711467>
- Tiancheng Sun, Kai-En Lin, Sai Bi, Zexiang Xu, and Ravi Ramamoorthi. 2021. NeLF: Neural Light-transport Field for Portrait View Synthesis and Relighting. In *Eurographics Symposium on Rendering*.
- Tiancheng Sun, Zexiang Xu, Xiuming Zhang, Sean Fanello, Christoph Rhemann, Paul Debevec, Yun-Ta Tsai, Jonathan T. Barron, and Ravi Ramamoorthi. 2020. Light Stage Super-Resolution: Continuous High-Frequency Relighting. In *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)*.
- Feitong Tan, Sean Fanello, Abhimitra Meka, Sergio Orts-Escalano, Danhang Tang, Rohit Pandey, Jonathan Taylor, Ping Tan, and Yinda Zhang. 2022. VoLux-GAN: A Generative Model for 3D Face Synthesis with HDRi Relighting. *arXiv:2201.04873 [cs.CV]*
- Zachary Teed and Jia Deng. 2020. RAFT: Recurrent All-Pairs Field Transforms for Optical Flow. In *European Conference on Computer Vision*.
- Ayush Tewari, Mohamed Elgharib, Gaurav Bharaj, Florian Bernard, Hans-Peter Seidel, Patrick Pérez, Michael Zöllhofer, and Christian Theobalt. 2020a. StyleRig: Rigging StyleGAN for 3D Control over Portrait Images. *CVPR 2020*. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.
- Ayush Tewari, Mohamed Elgharib, Mallikarjun BR, Florian Bernard, Hans-Peter Seidel, Patrick Pérez, Michael Zöllhofer, and Christian Theobalt. 2020b. PIE: Portrait Image Embedding for Semantic Control. *ACM Transactions on Graphics (Proceedings SIGGRAPH Asia)* 39, 6 (December 2020). <https://doi.org/10.1145/3414685.3417803>
- Yi Wang, Xin Tao, Xiaojuan Qi, Xiaoyong Shen, and Jiaya Jia. 2018. Image Inpainting via Generative Multi-column Convolutional Neural Networks. In *Advances in Neural*

- Information Processing Systems.* 331–340.
- Zhibo Wang, Xin Yu, Ming Lu, Quan Wang, Chen Qian, and Feng Xu. 2020. Single Image Portrait Relighting via Explicit Multiple Reflectance Channel Modeling. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)* (2020).
- Andreas Wenger, Andrew Gardner, Chris Tchou, Jonas Unger, Tim Hawkins, and Paul Debevec. 2005. Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM Trans. Graph.* 24, 3 (July 2005), 756–764. <https://doi.org/10.1145/1073204.1073258>
- Tim Weyrich, Wojciech Matusik, Hanspeter Pfister, Bernd Bickel, Craig Donner, Chien Tu, Janet McAndless, Jinho Lee, Addy Ngan, Henrik Wann Jensen, and Markus Gross. 2006. Analysis of Human Faces using a Measurement-Based Skin Reflectance Model. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)* (2006).
- Shuco Yamaguchi, Shunsuke Saito, Koki Nagano, Yajie Zhao, Weikai Chen, Kyle Olszewski, Shigeo Morishima, and Hao Li. 2018. High-fidelity facial reflectance and geometry inference from an unconstrained image. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)* (2018).
- Haotian Yang, Mingwu Zheng, Wanquan Feng, Haibin Huang, Yu-Kun Lai, Pengfei Wan, Zhongyuan Wang, and Chongyang Ma. 2023. Towards practical capture of high-fidelity relightable avatars. In *SIGGRAPH Asia 2023 Conference Papers*. 1–11.
- Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. 2021. pixelNeRF: Neural Radiance Fields from One or Few Images. In *CVPR*.
- Ziyang Yuan, Yiming Zhu, Yu Li, Hongyu Liu, and Chun Yuan. 2023. Make Encoder Great Again in 3D GAN Inversion through Geometry and Occlusion-Aware Encoding. *arXiv preprint arXiv:2303.12326* (2023).
- Egor Zakharov, Vanessa Sklyarova, Michael J Black, Giljoo Nam, Justus Thies, and Otmar Hilliges. 2024. Human Hair Reconstruction with Strand-Aligned 3D Gaussians. *ArXiv* (Sep 2024).
- Chong Zeng, Yue Dong, Pieter Peers, Youkang Kong, Hongzhi Wu, and Xin Tong. 2024. DiLightNet: Fine-grained Lighting Control for Diffusion-based Image Generation. In *ACM SIGGRAPH 2024 Conference Papers*.
- Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2025. Scaling In-the-Wild Training for Diffusion-based Illumination Harmonization and Editing by Imposing Consistent Light Transport. In *The Thirteenth International Conference on Learning Representations*. <https://openreview.net/forum?id=u1cQYxR1H>
- Longwen Zhang, Qixuan Zhang, Minye Wu, Jingyi Yu, and Lan Xu. 2021b. Neural Video Portrait Relighting in Real-Time via Consistency Modeling. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 802–812.
- Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *CVPR*.
- Xuaner Zhang, Jonathan T. Barron, Yun-Ta Tsai, Rohit Pandey, Xiuming Zhang, Ren Ng, and David E. Jacobs. 2020. Portrait Shadow Manipulation. In *ACM Transactions on Graphics (TOG)*.
- Xiuming Zhang, Pratul P. Srinivasan, Boyang Deng, Paul Debevec, William T. Freeman, and Jonathan T. Barron. 2021a. NeRFactor: Neural Factorization of Shape and Reflectance under an Unknown Illumination. *ACM Transactions on Graphics* (2021).
- Peng Zheng, Dehong Gao, Deng-Ping Fan, Li Liu, Jorma Laaksonen, Wanli Ouyang, and Nicu Sebe. 2024. Bilateral Reference for High-Resolution Dichotomous Image Segmentation. *CAAI Artificial Intelligence Research* (2024).
- Yang Zheng, Menglei Chai, Delio Vicini, Yuxiao Zhou, Yinghao Xu, Leonidas Guibas, Gordon Wetzstein, and Thabo Beeler. 2025. GroomLight: Hybrid Inverse Rendering for Relightable Human Hair Appearance Modeling. *arxiv*.
- Hao Zhou, Sunil Hadap, Kalyan Sunkavalli, and David W. Jacobs. 2019. Deep Single-Image Portrait Relighting. In *The IEEE International Conference on Computer Vision (ICCV)*.
- Taotao Zhou, Kai He, Di Wu, Teng Xu, Qixuan Zhang, Kuixiang Shao, Wenzheng Chen, Lan Xu, and Jingyi Yu. 2023. Relightable Neural Human Assets from Multi-view Gradient Illuminations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4315–4327.

SUPPLEMENTARY MATERIAL

We encourage readers to view the supplementary video for a comprehensive display of additional relighting results and OLAT renderings. Upon acceptance of this paper, we will make the *FaceOLAT* dataset, source code and pre-trained weights available. This supplementary document is organized as follows:

- In the Sec. 8 we provide details of our lightstage dataset - *FaceOLAT*.
- In Sec. 9, we provide extensive qualitative and quantitative comparisons of *3DPR* with state-of-the-art methods in 2D and 3D portrait relighting, demonstrating our robust performance.
- In Sec. 10 we provide training details.
- Sec. 11 provide details regarding design ablations, demonstrating how various configurations influence the effectiveness of our method.

8 FACEOLAT DATASET

Capturing accurate, high-fidelity human facial reflectance is essential for advancing research in facial relighting, skin & hair reflectance modeling, and realistic rendering, particularly for augmented and virtual reality applications. The pioneering work by Debevec et al. [2000] first demonstrated the feasibility and effectiveness of One-Light-At-a-Time (OLAT) setups using specialized hardware, known as a lightstage. Subsequently, the community widely adopted lightstages due to their effectiveness in detailed reflectance capture [B R et al. 2021a,b; Bi et al. 2021; Mei et al. 2024; Meka et al. 2019; Pandey et al. 2021; Rao et al. 2023, 2024a; Saito et al. 2024; Sarkar et al. 2023; Sun et al. 2020; Wang et al. 2020; Zhang et al. 2021b]. Nevertheless, the complexity, high costs, and demanding calibration procedures of such setups have restricted widespread accessibility. As a consequence, publicly available comprehensive OLAT datasets remain limited, constraining significant progress largely to closed-source initiatives and hindering broader advancements in the field.

Currently, the most extensive publicly available multi-view OLAT dataset is provided by the Codec Avatar Studio [Martinez et al. 2024; Saito et al. 2024], consisting of only four relightable subjects captured under dynamic illumination conditions (see Tab. 1 in main paper). This lack of diversity poses significant challenges for data-driven approaches, limiting their ability to generalize effectively to varied facial features and conditions.

To overcome these limitations, we present *FaceOLAT*, an extensively collected and meticulously calibrated OLAT dataset, distinguished by its scale, diversity, and quality (Fig. 12). Unlike prior datasets (refer to Tab. 1 in main paper), *FaceOLAT* uniquely features a large number of multi-view HDR captures at 4K resolution, a substantial variety of subjects, and diverse lighting conditions. These attributes make it particularly suitable for detailed reflectance analysis and the establishment of comprehensive reflectance priors.

In parallel, recent advancements in hair modeling [Sklyarova et al. 2023; Zakharov et al. 2024; Zheng et al. 2025] have highlighted the critical need for high-quality, multi-view hair datasets, which remain sparse. *FaceOLAT* addresses this specific shortcoming by capturing detailed scalp hair reflectance from multiple viewpoints, encompassing diverse hairstyles, colors, and lengths (Fig. 8a, Fig. 8b,

Fig. 12). Such comprehensive representation provides valuable priors for improving hair modeling and relighting methods.

In the subsequent sections, we detail various aspects of *FaceOLAT*, including the configuration of the lightstage (Sec. 8.1), data preprocessing pipeline (Sec. 8.2), demographic diversity (Sec. 8.3), and comparative analysis with existing datasets (Sec. 8.4). These descriptions aim to clearly illustrate the extensive efforts and specific challenges involved in dataset creation, emphasizing the significant contribution *FaceOLAT* makes to the research community.

8.1 Lightstage Configuration

Our capture setup utilizes a 2-meter diameter spherical lightstage equipped with 331 programmable light sources. Each source consists of six LEDs capable of emitting five specific wavelengths: red ($\lambda = 630\text{nm}$), green ($\lambda = 530\text{nm}$), royal blue ($\lambda = 450\text{nm}$), amber ($\lambda = 600\text{nm}$), and daylight white ($5650K$). During the capture sessions, subjects are centrally positioned and uniformly illuminated by white light. The captures are performed synchronously from 40 viewpoints at 4K resolution using RED Komodo cameras operating at 60 FPS.

8.2 Data Preprocessing

The captured data undergo rigorous preprocessing to ensure its suitability for model training:

8.2.1 Optical Flow Alignment. Each subject undergoes a sequence of 331 illumination conditions per expression, defined as a *take*, lasting approximately 6 seconds. Minor involuntary subject movements during this interval lead to motion blur when OLAT frames are linearly combined for relighting (Fig. 10, left). To address this issue, we intersperse fully lit reference frames every 21 OLAT captures. Optical flow computed using RAFT [Teed and Deng 2020] between these reference frames is interpolated linearly to align all OLAT frames, significantly mitigating motion blur, as illustrated in Fig. 9.

8.2.2 Background Segmentation. Accurate foreground-background segmentation is crucial for effective facial reflectance modeling. We generate segmentation masks primarily using BGMv2 [Lin et al. 2020], employing reference background captures without subjects. When BGMv2 produces suboptimal results, we manually select robust alternative masks from RMBGv2 [Zheng et al. 2024].

8.2.3 Calibration and 3D Reconstruction. Camera parameters and 3D meshes are reconstructed using Agisoft Metashape¹, with calibration guided by static reference objects and alignment markers placed in the capture volume. We use fully lit frames for reconstruction, as these are devoid of harsh shadows and specular highlights, which are prominent in OLAT images and tend to degrade feature matching quality. Using uniformly illuminated frames improves the robustness of Metashape's feature matching algorithm for both multi-view stereo reconstruction and camera calibration, resulting in an average re-projection error of 0.91 pixels. Further, we will provide a white-balance correction code to correct colors as shown in Fig. 11 along with *FaceOLAT*.

Despite the superior accuracy achievable through multi-view calibration, we utilize monocular face tracking approaches consistent

¹<https://www.agisoft.com/>



(a) Multi-view captures of hair under fully-lit illumination.

(b) Hair captures of 10 randomly selected subjects showcasing diversity.

(c) OLAT captures of hair from a fixed viewpoint.

Fig. 8. Overview of Hair Captures in FaceOLAT. (a) Multi-view images under fully-lit conditions facilitate high-quality hair reconstruction. (b) Our dataset includes a diverse range of hairstyles, colors, and lengths, providing rich priors for modeling. (c) OLAT captures from a fixed viewpoint support detailed hair reflectance modeling.

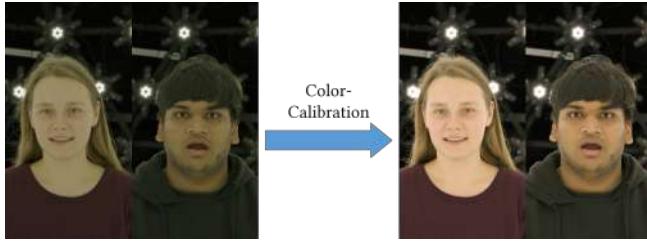


Fig. 11. Color Correction. We perform color correction on FaceOLAT. Shown are representative examples before (left) and after (right) calibration.

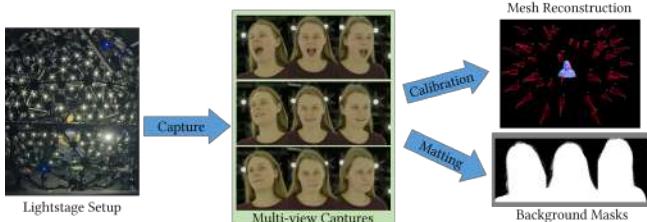


Fig. 9. Data Processing Pipeline. Each subject is captured under 350 lighting conditions (331 OLAT and 19 fully lit) from 40 viewpoints, across 3 or 4 distinct facial expressions. We then estimate the camera parameters and undistort the images. Finally, background matting techniques are applied to extract background masks.

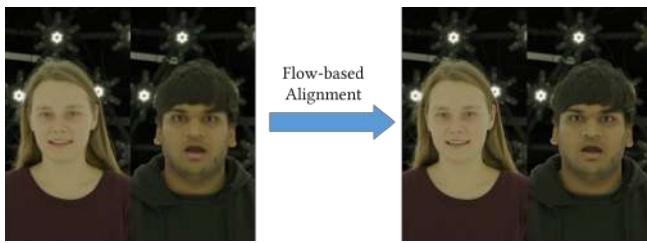


Fig. 10. Impact of Optical Flow-Based Alignment. We demonstrate that subjects involuntarily tend to introduce minor motion between frames during the capture process, resulting in noticeable blurring when OLAT images are combined (see left column). The results after applying optical flow (see right column) illustrate that correction using RAFT [Teed and Deng 2020] significantly reduces such misalignments.

with monocular portrait relighting methods for our training and evaluations, following EG3D [Chan et al. 2022].

8.3 Demographics

We collected a novel multiview lightstage dataset, *FaceOLAT*, comprising 139 subjects. We made an effort to include participants representing diverse demographics across age, gender, skin color, and hair color. Out of the 139 subjects, 122 participants completed the demographic survey, and their responses are summarized in Fig. 12.

51.25% of the participants are aged between 24 and 30 years, 17.9% fall within the 31–40 years, 17.9% are between 18 and 23 years, and 9.8% are between 51 and 60 years old. Regarding gender, our dataset includes approximately 59% male, 37% female, and 1.6% non-binary participants, with the remaining subjects opting not to disclose their gender.

For skin color, we utilize the Fitzpatrick scale². The dataset comprises 34.1% participants with Type 1 (Light, pale white) skin, 21.1% with Type 2 (White, fair) skin, 20.3% with Type 3 (Medium, white to olive) skin, 14.6% with Type 4 (Olive, moderate brown) skin, 4.9% with Type 5 (Brown, dark brown), and 4.9% with Type 6 (Black, very dark brown to black) skin.

Hair color is reported using a hair color scale³, where higher values correspond to lighter hair colors, and lower values represent darker hair tones.

8.4 Comparison with Existing Datasets

In this section, we highlight the key features that make *FaceOLAT* a unique contribution compared to existing publicly available light stage datasets. *FaceOLAT* offers the following advantages:

²<https://emergetulsa.com/fitzpatrick/>

³<https://satinhaircolor.com/education/>

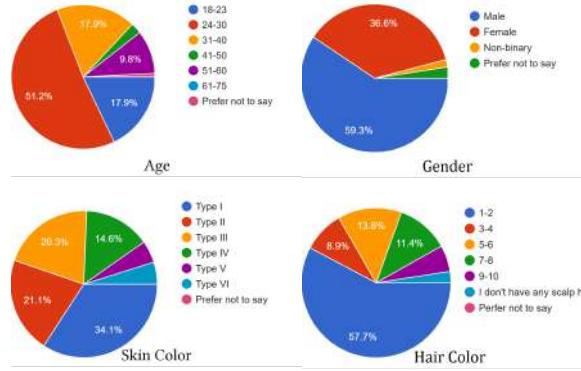


Fig. 12. **Dataset Summary.** We summarize age, gender, skin color and hair color demographics.

- Illumination Coverage:** 331 lighting conditions, each corresponding to a unique incident direction, enabling detailed relighting effects such as specular highlights, subsurface scattering, self-shadowing, and hard shadows.
- Scale and Diversity:** Our corpus comprises 139 subjects, providing a diverse set of facial geometries and appearances. This results in enhanced generalizability and constitutes the **largest database** of relightable heads available to date.
- Multi-view Resolution:** Consistent 4K resolution across 40 viewpoints, capturing high-frequency details in facial features, hair, and eye regions.

The most comparable dataset to ours is the dynamic face OLAT dataset from Codec Avatar Studio [Martinez et al. 2024; Saito et al. 2024], which includes only 4 subjects captured under grouped OLAT lighting conditions. This dataset suffers from two primary limitations: (1) the small number of subjects limits its ability to support robust face reflectance prior modeling, and (2) grouped OLAT captures are less suitable for static image-based relighting due to reduced directional specificity.

Another relevant dataset is Ultrastage [Zhou et al. 2023], which captures full-body subjects under gradient illumination. While gradient lighting is effective for estimating photometric normals, it is suboptimal for high-quality relighting. Gradient-based relighting approaches, such as Ultrastage, rely on simplified BRDF assumptions and fail to capture complex light-skin interactions. In contrast, our OLAT-based acquisition captures the complete reflectance field, making it highly suitable for physically accurate, image-based relighting.

Additional publicly available datasets, including ICT-3DFE [Stratou et al. 2011] and the Dynamic OLAT dataset [Zhang et al. 2021b], capture subjects from only one or two viewpoints, thereby limiting their utility for 3D face modeling. Furthermore, these datasets contain few subjects and are limited in resolution.

Finally, although not publicly available, the OLAT dataset by Weyrich et al. [Weyrich et al. 2006] has long served as a standard benchmark for evaluating face reflectance models [B R et al. 2021a,b; Rao et al. 2023, 2022, 2024a; Sun et al. 2021; Weyrich et al. 2006]. Hence, we train and compare 3DPR and prior methods on this dataset to ensure standardized and fair evaluation. However, it

presents several key limitations: (1) it is a closed-source dataset, limiting reproducibility and broader research adoption; (2) all subjects exhibit a single, neutral expression with closed eyes, making it unsuitable for modeling eye reflectance, expressions, and mouth interior; and (3) it captures only the frontal face hemisphere from 14 views, with no coverage of hair or the back of the head. *FaceOLAT* addresses each of these shortcomings, offering public access, diverse expressions with open-eye captures, and full-head coverage including hair, thus providing a more comprehensive resource for developing and evaluating face reflectance models.

In summary, *FaceOLAT* addresses critical limitations of existing datasets by providing a large-scale, high-resolution, and multi-view dataset for full-head reflectance modeling, making it a valuable resource for the community.

9 ADDITIONAL BASELINES

In this section, we first provide qualitative comparisons with Holo-Relighting [Mei et al. 2024] (HR). Next, we evaluate our approach against a 2D portrait relighting method, Total Relighting [Pandey et al. 2021]. Additionally, we present further qualitative evaluations for Light2Relight, NeRFFaceLighting, VoRF, and NeLF.

9.1 Holo-Relighting

We provide a comparison with Holo-Relighting [Mei et al. 2024]. Their method also relies on an EG3D prior to reconstruct the facial volume for view synthesis and is trained on lightstage data for relighting. However, unlike our approach, their pipeline directly inputs an environment map and produces the relit result without generating OLATs, leading to physically implausible outputs. As shown in Fig. 13, their technique often produces oversaturated results and lacks finer details, such as facial hair. Furthermore, as evidenced in Fig. 14, Holo-Relighting struggles to accurately relight under point light conditions, despite being trained on a lightstage dataset. For example, it fails to reproduce complex skin-light interactions, such as subsurface scattering (see the ear regions in columns 1 and 2). Additionally, under a point light source positioned below the face, Holo-Relighting fails to generate coherent relit results. In contrast, 3DPR effectively handles subsurface scattering effects and achieves accurate relighting, as evidenced by its close match to the reference image.

Since neither the code nor implementation details of HR are publicly available, retraining their method with our data for quantitative evaluation was not feasible. Instead, we shared our evaluation data with the authors. Upon discussion, we found that quantitative comparisons would be challenging due to differences in the radiance scaling factors for the HDR environment maps and variations in lightstage capture setups. Therefore, we provide only qualitative comparisons by evaluating Holo-Relighting against 3DPR trained on *FaceOLAT*.

9.2 Total Relighting

Total Relighting [Pandey et al. 2021] (TR) explicitly decomposes face reflectance by employing dedicated networks to separately extract surface normals, albedo, and specular maps. This method utilizes the Phong reflectance function to facilitate the decomposition process. We conduct both qualitative and quantitative comparisons of

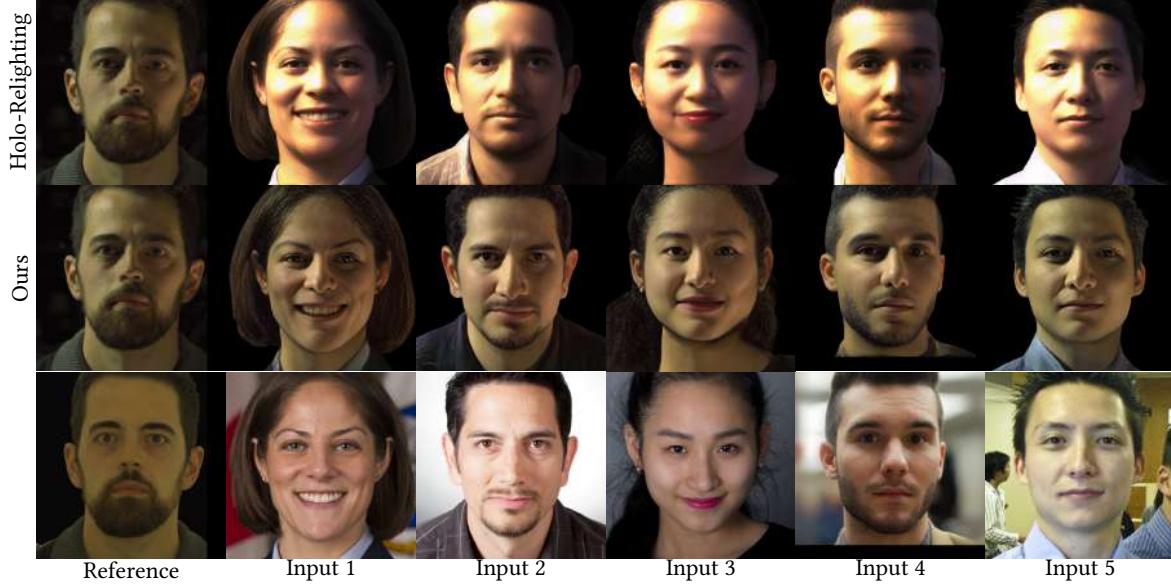


Fig. 13. Qualitative Results: 3DPR vs. Holo-Relighting [Mei et al. 2024] (in-the-wild Subjects). The first row presents relit results for HR. We can see that relit results suffer from oversaturation artifacts during relighting. Second row shows 3DPR relighting results, which have the closest resemblance to the ground truth (first column) for all the inputs.

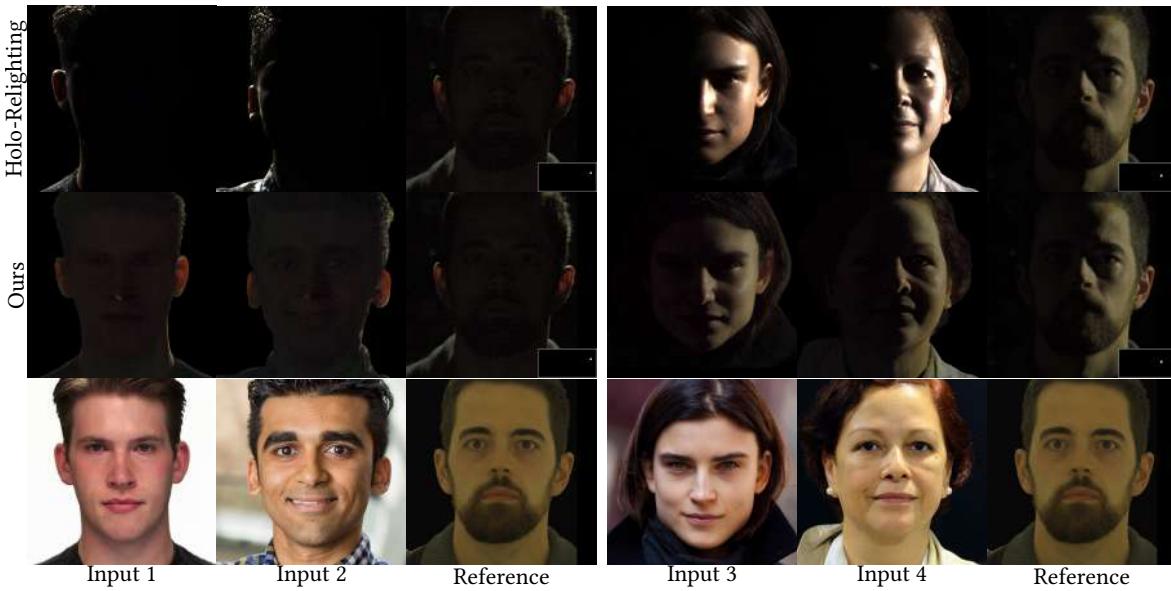


Fig. 14. Qualitative Results: 3DPR vs. Holo-Relighting [Mei et al. 2024] (In-the-Wild Subjects). We present qualitative results using point light sources. The first row shows results from Holo-Relighting, where the method struggles to capture effects such as subsurface scattering (see the ear lobes regions) and accurate light directions (Inputs 1 and 2) and show oversaturated specular artifacts (Inputs 3 and 4). In contrast, 3DPR demonstrates precise light control, accurately relighting for point light sources (see Input 3 and 4), and effectively capturing complex light-skin interactions, including subsurface scattering (see Inputs 1 and 2).

3DPR with Total Relighting (TR) to assess the effectiveness of our proposed relighting approach. In our quantitative analysis, we evaluate the performance using metrics such as LPIPS [Zhang et al. 2018], RMSE, DISTS [Ding et al. 2020], SSIM, and PSNR across 10 unseen

lightstage subjects under 27 different illumination conditions. For TR, due to the lack of publicly available code or models, evaluations were performed by original authors using their pre-trained models using our inputs.

Table 7. Quantitative Results: OLAT Evaluation We report the quality of synthesized OLATs for simultaneous view-synthesis and relighting.

| | SSIM \uparrow | LPIPS \downarrow | PSNR \uparrow |
|------|-----------------|--------------------|-----------------|
| VoRF | 0.71 | 0.3148 | 20.43 |
| Ours | 0.88 | 0.1753 | 28.70 |

Table 5. Quantitative comparison with Total Relighting[Pandey et al. 2021]: Metrics are evaluated for relighting performance across 10 unseen lightstage subjects [Weyrich et al. 2006] under 27 different illumination conditions.

| | LPIPS \downarrow | RMSE \downarrow | DISTS \downarrow | SSIM \uparrow | PSNR \uparrow |
|-------------------|--------------------|-------------------|--------------------|-----------------|-----------------|
| Total Relighting | 0.2991 | 0.2203 | 0.2228 | 0.8612 | 29.25 |
| Ours ($3DPR_w$) | 0.2599 | 0.1660 | 0.1933 | 0.8772 | 28.73 |

The results, detailed in Tab. 5, demonstrate that $3DPR$ surpasses the TR baseline in all metrics except PSNR, which achieves a score closely competitive with TR. This underscores our method’s high-quality performance, aligning closely with the state-of-the-art while potentially offering additional advantages *i.e.* relighting under novel views. Building on our quantitative assessments, qualitative evaluations from Fig. 15, and Fig. 16 further illustrate the advantages of $3DPR$ over TR, in handling both lightstage and in-the-wild subjects. Our method is successful in capturing accurate shadows and specular highlights, in contrast to TR, which often exhibits issues with inaccurate lighting and produces cloudy artifacts on facial regions. These observations highlight the critical role of employing a rich face prior and utilizing implicit face reflectance modeling through OLATs. These strategies enable $3DPR$ to achieve precise relighting across diverse illumination conditions.

9.3 IC-Light

We compare against the recent diffusion-based relighting method IC-Light [Zhang et al. 2025]. Diffusion models provide strong priors for generalization; however, we observe that IC-Light, while producing plausible results, struggles to reproduce physically accurate lighting effects such as specular highlights, subsurface scattering, and cast shadows. These limitations are more pronounced under sparse, highly directional illumination (see input 2 in Fig. 17). We also evaluate cross-view relighting consistency by first generating novel viewpoints with EG3D and then relighting them using IC-Light; the results reveal inconsistent lighting across viewpoints (Fig. 17). Additionally, quantitative results on FaceOLAT (Tab. 6) show that $3DPR$ outperforms IC-Light, including stronger identity preservation during relighting. We attribute the superior performance of $3DPR$ to its 3D-consistent generative prior coupled with explicit OLAT-based reflectance modeling notably absent in IC-Light.

Table 6. Quantitative comparison with IC-Light[Zhang et al. 2025]: Metrics are evaluated for relighting performance on *FaceOLAT* under 10 different illumination conditions.

| | SSIM \uparrow | LPIPS \downarrow | RMSE \downarrow | DISTS \downarrow | PSNR \uparrow | ID \uparrow |
|-----------------------|-----------------|--------------------|-------------------|--------------------|-----------------|---------------|
| IC-Light (input-view) | 0.70 | 0.309 | 0.375 | 0.284 | 20.36 | 0.913 |
| IC-Light (novel-view) | 0.67 | 0.313 | 0.470 | 0.309 | 20.24 | 0.874 |
| Ours | 0.83 | 0.199 | 0.180 | 0.175 | 21.02 | 0.943 |

9.4 NeRFFaceLighting and Lite2Relight

In this section, we provide additional qualitative comparisons for: NeRFFaceLighting (NFL) [Jiang et al. 2023] and Lite2Relight (L2R) [Rao et al. 2024b] in Fig. 19 and Fig. 18. We observe that both L2R and NFL begin to degrade under sparse illumination conditions. NFL struggles significantly with identity preservation and lighting realism. Its spherical harmonics (SH)-based lighting representation captures only low-frequency cues, and it fails to model high-frequency effects such as sharp shadows or specular highlights. In addition, we observe that lighting tends to be baked into the albedo during its inaccurate inversion process, further limiting generalization under colored lighting conditions.

L2R performs reasonably well under dense natural lighting but fails under sparse or highly directional environments, likely due to its latent-space probing strategy, which is biased by the EG3D training distribution. This leads to artifacts such as inconsistent illumination, lack of shadows, and distorted identity features in extreme cases. To verify this observation, we contacted the authors of Lite2Relight, who confirmed that this is a known limitation of their method, as documented in the original paper.

In contrast, our method explicitly models face reflectance using OLATs and linearly combines these basis images with arbitrary environment maps. This allows for faithful and consistent relighting, even under sparse or colored light setups. The figure clearly illustrates our method’s ability to maintain photorealism, preserve facial structure, and accurately reproduce lighting effects across a wide range of conditions. These results highlight the importance and effectiveness of our OLAT-based formulation for generalizable relighting.

9.4.1 NeLF. NeLF [Sun et al. 2021] proposes a pixelNeRF-based [Yu et al. 2021] method that leverages image features to infer the underlying 3D geometry and light transport. We present qualitative results for NeLF in Fig. 20. We can see that despite utilizing three input views, NeLF encounters difficulties in achieving coherent face reconstruction. This issue arises from its reliance on the aggregation of local image features to infer geometry, which results in poor facial geometry reconstruction. Consequently, NeLF scores poorly as well as reported in the Tab.2 of the main paper.

9.5 VoRF

For qualitative analysis, we compare $3DPR_w$ with VoRF, both trained on *WeyrichOLAT*, and present the corresponding quantitative results in Tab. 7, where $3DPR$ significantly outperforms VoRF across all metrics. Since both methods rely on explicit OLAT representations to model facial reflectance, it is crucial to assess the accuracy of the synthesized OLATs. As shown in Fig. 21, $3DPR$ generalizes effectively to both our evaluation dataset (2nd row) and to “in-the-wild” subjects (3rd and 4th rows), producing higher-fidelity relighting results that capture complex lighting effects more reliably than VoRF. Note that for evaluating VoRF against $3DPR$, we use $3DPR_w$. Since the lightstage dataset of Weyrich et al. [2006] contains subjects with closed eyes, $3DPR_w$ fails to accurately model the reflectance of the eye region. Consequently, we mask the eye region in Fig. 21 and Fig. 22.



Fig. 15. Qualitative Results - 3DPR vs. Total Relighting (Lightstage Subjects). The first column presents the input image, followed by columns showing relighting results for the same pose. The 2nd and 5th columns feature results from Total Relighting (TR), which often displays inaccurate relighting and cloudy artifacts in the face regions. The 3rd and 6th columns show results from 3DPR, and the 4th and 7th columns display the ground truth. These subjects are sourced from the test set of the Lightstage dataset [Weyrich et al. 2006], demonstrating accurate relighting of 3DPR in portrait relighting compared to TR.

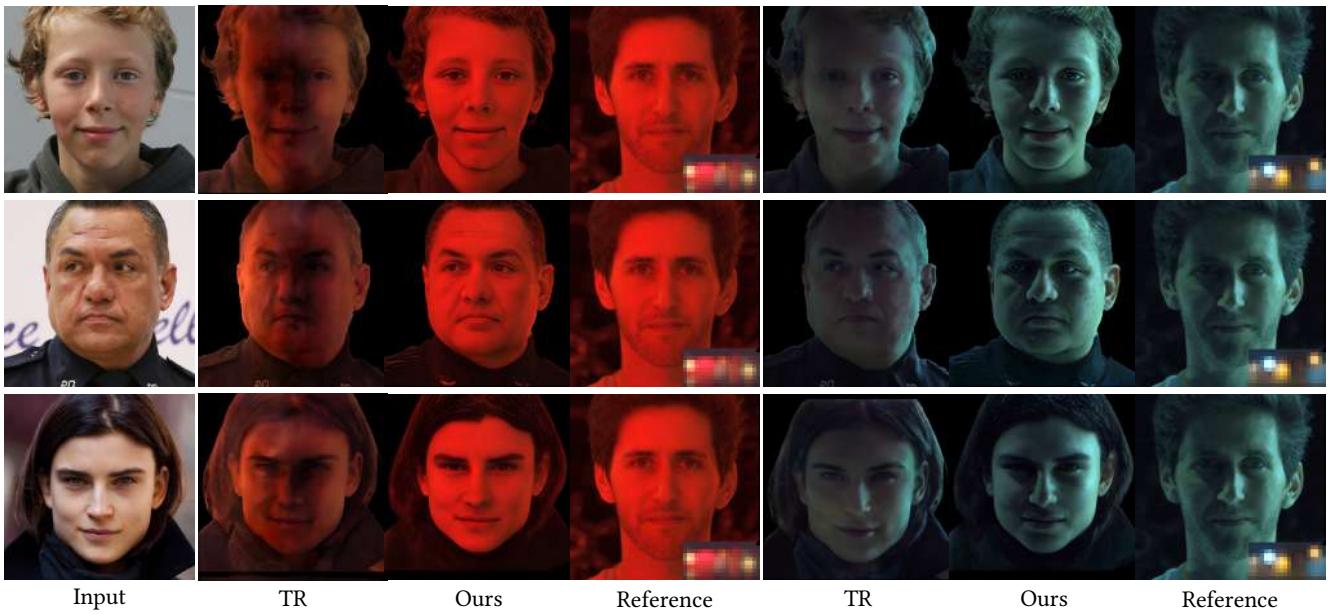


Fig. 16. Qualitative Results - 3DPR vs. Total Relighting (in-the-wild Subjects). The first column presents the input image, followed by columns displaying relighting results for the same pose. The 2nd and 5th columns showcase results from Total Relighting (TR), which often exhibits inaccurate relighting and cloudy artifacts in the facial regions. The 3rd and 6th columns illustrate results from 3DPR, while the 4th and 7th columns display reference images relit under target illumination, as ground truth is unavailable for in-the-wild samples. These subjects are picked from CelebA-HQ [Karras et al. 2018] and Ravdess [Livingstone and Russo 2018], highlighting the robust relighting performance of 3DPR compared to TR. We use $3DPR_o$ for this evaluation.

9.6 LumiGAN

LumiGAN [Deng et al. 2023] decomposes a given face into geometry, albedo, specular, and visibility maps. Next, it combines the decomposed rendering with spherical harmonic projections of desired environment maps to perform relighting. The method relies entirely on adversarial training which produces plausible results, but, due to the lack of relit ground truth supervision during training, LumiGAN lacks physically accurate lighting. For instance, by observing **Figure S3** [Deng et al. 2023] we can see that the direction of illumination

and specularities (observe the highlights on forehead and nose region) are inconsistent across subjects for the same environment map. Moreover, the specular highlights appear in a few subjects, but are missing in others. We hypothesize that these occurrences are due to inaccurate decomposition of face reflectance components.

In contrast, 3DPR is based on *volumetric reflectance fields*. Our face reflectance modeling is built using OLATs and these OLATs are captured in a lightstage [Weyrich et al. 2006] setup that can represent

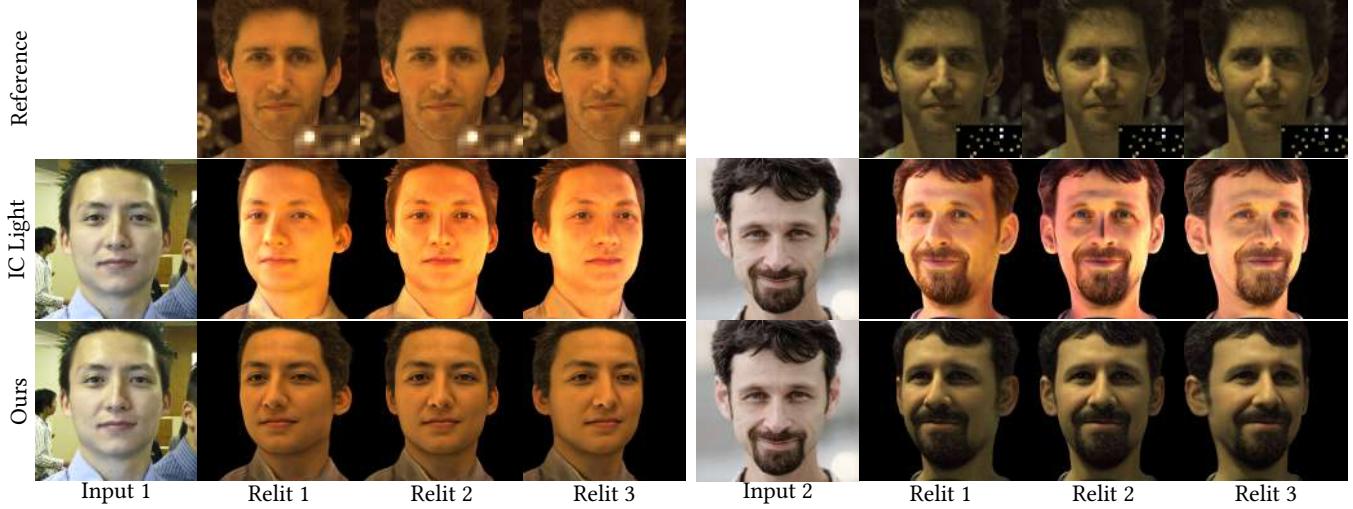


Fig. 17. Qualitative results: 3DPR vs. IC-Light [Zhang et al. 2025] on “in-the-wild” subjects. Each panel uses an “in-the-wild” input (first column). The reference image (top row) indicates the target lighting from a different subject. Each row shows a different camera viewpoint of the same subject; columns correspond to the same viewpoint across methods. IC-Light often introduces inaccurate lighting effects and struggles to synthesize soft self-shadowing (see input 2). In addition, lacking an explicit 3D representation, IC-Light yields inconsistent relighting across viewpoints (e.g., observe specular highlights on the nose and cheeks across views). In contrast, 3DPR preserves 3D consistency across viewpoints and produces coherent relighting even under sparse illumination.

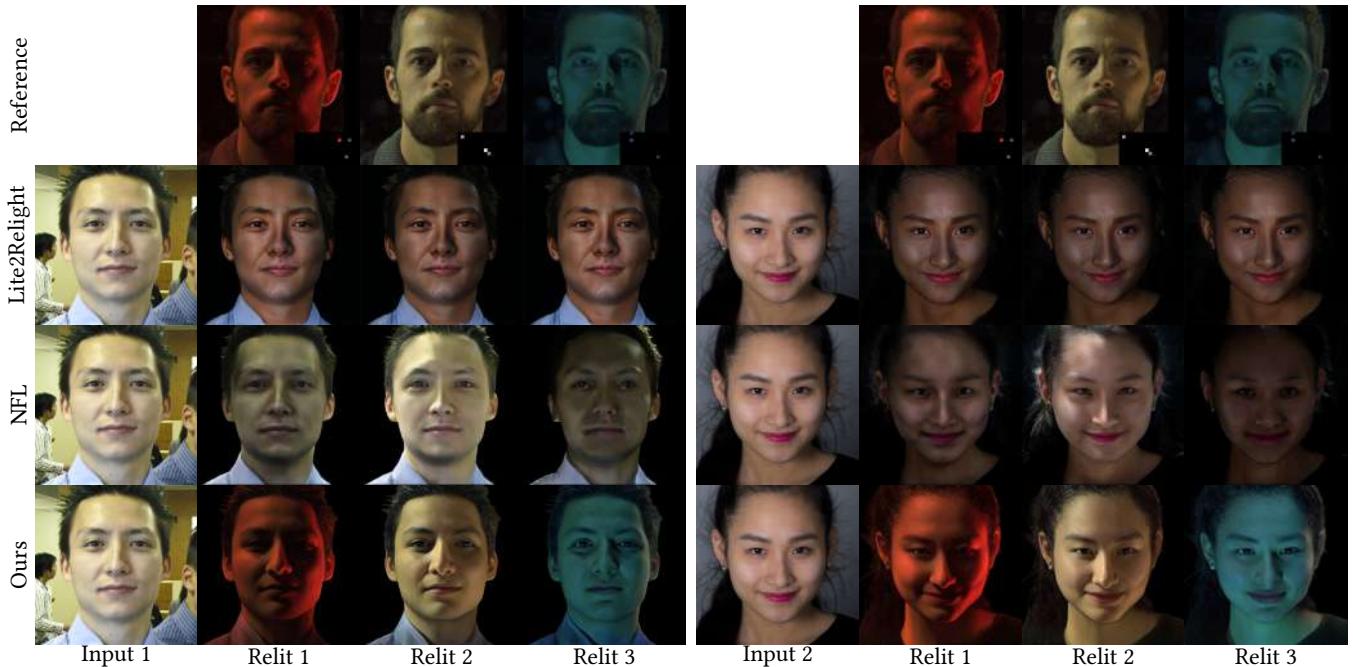


Fig. 18. Baseline comparison with NFL vs L2R. The input images (first column) are in-the-wild images. The reference image (top row) depicts the target lighting conditions from a different subject for comparison. Both L2R and 3DPR are trained on FaceOLAT. The remaining rows show relit versions of the input subject. L2R and NFL exhibit shading artifacts in the face region, highlighting difficulties in achieving accurate relighting. In contrast, 3DPR demonstrates consistent relighting under challenging illumination conditions while preserving identity. Furthermore, our method maintains relighting consistency across different subjects.

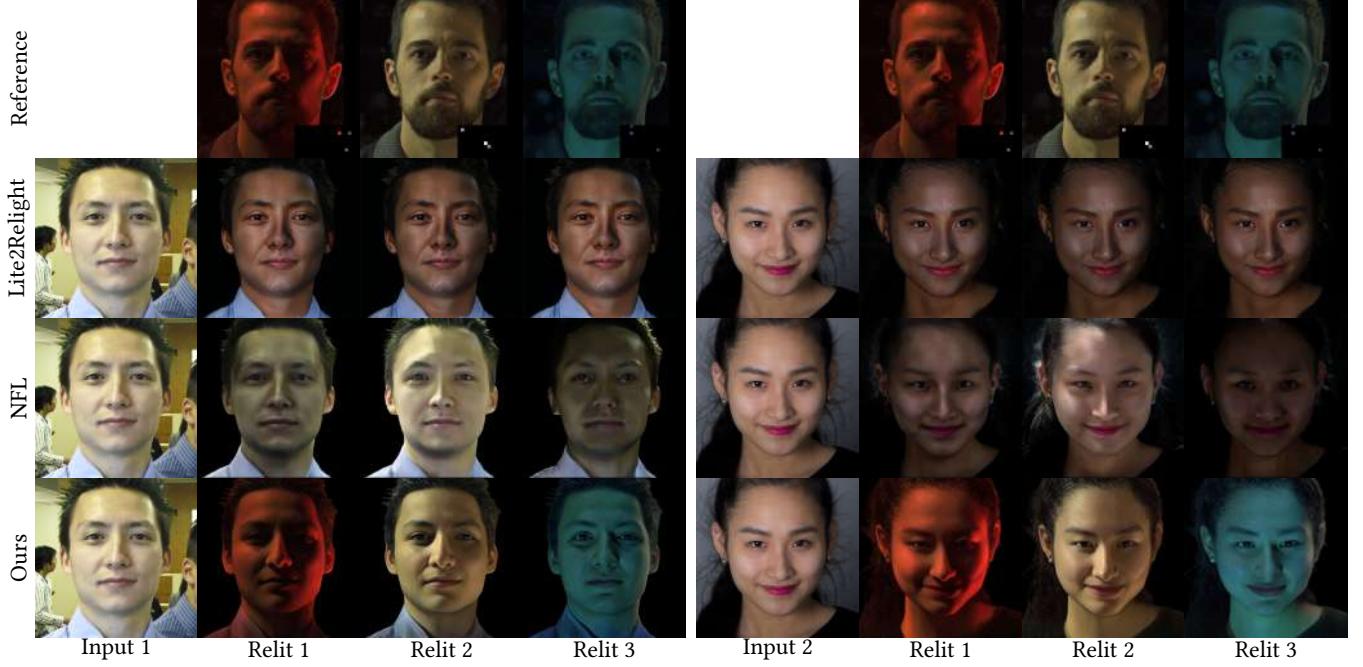


Fig. 19. Qualitative Results: 3DPR vs. NFL vs. L2R (In-the-Wild Subjects). This figure uses in-the-wild input images (first column) for evaluation. The reference image represents the target lighting conditions from a different subject for comparison. Both L2R and 3DPR are trained on FaceOLAT. The subsequent rows show the same subject under different views (*i.e.*, each column represents the same view). L2R and NFL exhibit shading artifacts in the face region, indicating challenges with accurate relighting. In contrast, 3DPR successfully relights under challenging illumination conditions while preserving 3D consistency and synthesizing coherent novel viewpoints.

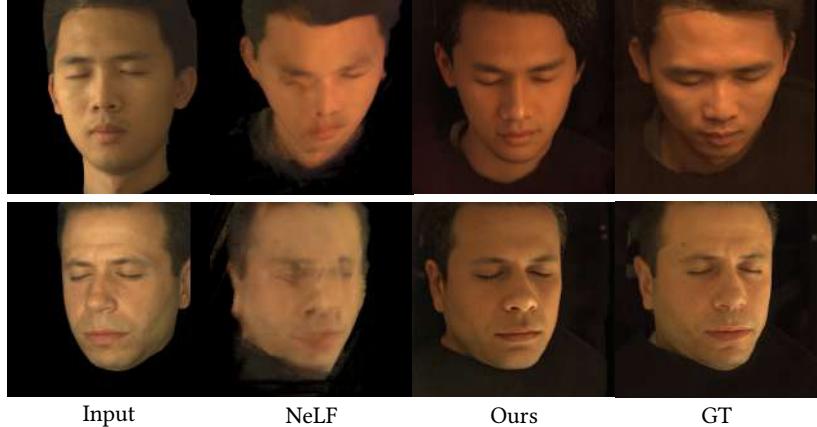


Fig. 20. Qualitative Results: Baseline Comparisons with NeLF [Sun et al. 2021]. We use three input views to generate the results of NeLF and a single input view (first column) for our method. NeLF struggles to preserve the identity of the subject due to its 3D representation and we can see that 3DPR can synthesize the subject well under novel viewpoints. These subjects are sourced from the test set of the Lightstage dataset [Weyrich et al. 2006].

physically accurate lighting. Thus, by supervising our method with OLAT dataset as ground truth, we learn physically accurate lighting. Since neither the code, nor the pretrained weights of LumiGAN are publicly available, we have compared our method with a state-of-the-art GAN-based approach that is similar to LumiGAN in formulation: NeRFFaceLighting [Jiang et al. 2023] (see main paper for the results).

9.7 VoLux-GAN

VoLux-GAN [Tan et al. 2022] is a 3D-aware generator capable of relighting arbitrary samples, but cannot reconstruct and relight a specific given portrait image. Conversations with the VoLux-GAN authors confirmed that the rich StyleGAN latent space, which facilitates the embedding of test subjects, is not preserved in their



Fig. 21. Qualitative Results: 3DPR vs. VORF (OLAT Renderings) For a lightstage subject from the test set, we compare our synthesized results (Row 2) and VoRF directly with ground truth (Row 1). Our renderings are much sharper and closely resemble the ground truth. Additionally, OLAT renderings for ‘in-the-wild’ portraits by 3DPR are shown in Rows 5 and 7, while VoRF’s renderings are in Rows 6 and 8. Notably, 3DPR’s predictions closely match the actual light directions of the references (Row 4). Furthermore, our method consistently captures intricate details such as hard shadows, specular highlights, and subsurface scattering effects across various subjects. Finally, the last two columns demonstrate relit results under novel viewpoints. We use $3DPR_w$ for evaluation.

approach due to the disentangled representation. This is further reinforced by the fact that they do not show any inversion or reconstruction results, but only generated identities sampled from the latent space. 3DPR, however, is not just another 3D-GAN approach, but specifically designed for portrait relighting, capable of reconstructing and relighting given portrait images, which VoluxGAN cannot do.

10 TRAINING

We train 3DPR with 100 HDRIs (60 indoor, 40 outdoor) per subject, covering a wide range of color temperatures and lighting conditions, which makes it robust to in-the-world lighting during inference as demonstrated by our superior performance in both qualitative and quantitative evaluations.

3DPR is implemented using PyTorch and integrates a pre-trained 3D generator [Chan et al. 2022] alongside a geometry-aware inversion framework [Yuan et al. 2023]. We train our method on two

Table 9. Quantitative Results: Ablation Study on F_o dimension. The performance metrics are evaluated for relighting performance across 10 unseen Lightstage subjects [Weyrich et al. 2006].

| | SSIM ↑ | PSNR ↑ | LD ↓ |
|----------|-------------|--------------|--------------|
| $D = 9$ | 0.84 | 22.62 | 10.89 |
| $D = 48$ | 0.86 | 28.74 | 11.57 |
| $D = 96$ | 0.87 | 28.69 | 10.30 |

Table 8. Quantitative Results: Number of Training Subjects. The performance metrics are evaluated for relighting performance across 10 unseen Lightstage subjects [Weyrich et al. 2006].

| | SSIM ↑ | PSNR ↑ | LD ↓ |
|-----------|-------------|--------------|-------------|
| $n = 5$ | 0.85 | 28.66 | 11.18 |
| $n = 25$ | 0.86 | 28.62 | 10.71 |
| $n = 100$ | 0.86 | 28.61 | 10.56 |
| $n = 250$ | 0.87 | 28.69 | 10.3 |

different lightstage datasets: *FaceOLAT* and *WeyrichOLAT* [Weyrich et al. 2006]. The trained models are denoted as $3DPR_o$ and $3DPR_w$, respectively, in our results. During the first 60k training iterations, we only train R_{enc} and R_{dec} , supervised by low-resolution OLAT images o_{rgb} . This “warm-up” phase is crucial for stabilizing the training process. Following this, E_{SR} is initialized with weights from the original EG3D super-resolution network. We then proceed to training all modules jointly, including E_{SR} , for an additional 150k iterations. Training uses a batch size of 8 on 4 × NVIDIA H100 GPUs over a span of 2 days. We employ the Adam optimizer [Kingma and Ba 2015] with a learning rate of 0.00015. The training dataset consists of 130 subjects, each re-illuminated under 20 randomly sampled natural illumination conditions from three different viewpoints.

11 DESIGN ABLATION STUDY

In this section, we ablate various design choices of $3DPR$ qualitatively and quantitatively. We conduct all the ablations using $3DPR_w$ that was trained on *WeyrichOLAT* by Weyrich et al. [2006].

11.1 Significance of SR Encoder:

We summarize qualitative analysis in the Fig. 22 to highlight the importance of E_{SR} .

11.2 Significance of \mathcal{L}_{MRF} :

We present the qualitative results Fig. 23 clearly shows that including \mathcal{L}_{MRF} in training leads to best results.

11.3 Effect of Training Subjects Number

We evaluate the task of simultaneous view synthesis and relighting using multiple models trained with 250, 100, 25 and as few as 5 subjects. We observe that with as few as 5 subjects we achieve reasonable relighting shown in Fig. 24. Although we achieve the best performance with 250 subjects. From the quantitative results in Tab. 8 we observe a similar performance with 100 and 25 training subjects. Further, we achieve relighting without involving any tedious optimization process compared to VoRF [Rao et al. 2023]

or NFL[Jiang et al. 2023]. This demonstrates the significance of the EG3D prior that helps in generalization towards unseen subjects including in-the-wild captures (see Fig. 3 in the main paper). In, Tab. 8 we report an average facial landmarks deviation (LD) metric to evaluate facial geometry consistency by measuring the average deviation of 68 facial key points [Bulat and Tzimiropoulos 2017].

11.4 Dimensionality of *Reflectance Encoder*

The pre-trained EG3D triplane representation, F_g , consists of 96 dimensions. In this ablation study, we investigate the impact of varying the dimension size of F_o on relighting quality. We start with a depth of F_o at 9 dimensions. At this level, without sufficient o_{hf} to encode high-frequency details, the relighting results appear blurry, as illustrated in Fig. 25. As the depth of F_o increases, we observe enhancements in the quality of relighting and facial features. The best results are achieved when F_o is extended to 96 dimensions, as detailed in Tab. 9. This configuration allows for effective encoding of complex lighting effects, including subsurface scattering (visible in the first row’s third and fourth columns) and specular highlights on the nose (noted in the second row’s fourth column).

REFERENCES

- Mallikarjun B R, Ayush Tewari, Abdallah Dib, Tim Weyrich, Bernd Bickel, Hans Peter Seidel, Hanspeter Pfister, Wojciech Matusik, Louis Chevallier, Mohamed A Elgarib, and Christian Theobalt. 2021a. PhotoApp: Photorealistic appearance editing of head portraits. *ACM Transactions on Graphics* 40, 4 (2021).
- Mallikarjun B R, Ayush Tewari, Tae-Hyun Oh, Tim Weyrich, Bernd Bickel, Hans-Peter Seidel, Hanspeter Pfister, Wojciech Matusik, Mohamed Elgarib, and Christian Theobalt. 2021b. Monocular Reconstruction of Neural Face Reflectance Fields. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Sai Bi, Stephen Lombardi, Shunsuke Saito, Tomas Simon, Shih-En Wei, Kevyn McPhail, Ravi Ramamoorthi, Yaser Sheikh, and Jason Saragih. 2021. Deep Relightable Appearance Models for Animatable Faces. *ACM Transactions on Graphics*, Article 89 (2021), 15 pages.
- Mark Boss, Raphael Braun, Varun Jampani, Jonathan T. Barron, Ce Liu, and Hendrik P.A. Lenzsch. 2021. NeRD: Neural Reflectance Decomposition from Image Collections. In *The IEEE International Conference on Computer Vision (ICCV)*.
- Marcel C. Buehler, Abhimitra Meka, Gengyan Li, Thabo Beeler, and Otmar Hilliges. 2021. VariTex: Variational Neural Face Textures. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.
- Adrian Bulat and Georgios Tzimiropoulos. 2017. How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks). In *International Conference on Computer Vision*.
- Eric R. Chan, Connor Z. Lin, Matthew A. Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas Guibas, Jonathan Tremblay, Sameh Khamis, Tero Karras, and Gordon Wetzstein. 2022. Efficient Geometry-aware 3D Generative Adversarial Networks. In *CVPR*.
- Sreemithy Chandran, Yannick Hold-Geoffroy, Kalyan Sunkavalli, Zhixin Shu, and Suren Jayasuriya. 2022. Temporally Consistent Relighting for Portrait Videos. In *The IEEE Winter Conference on Applications of Computer Vision (WACV) Workshops*. 719–728.
- Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. 2000. Acquiring the reflectance field of a human face. In *Annual conference on Computer graphics and interactive techniques*.
- Boyang Deng, Yifan Wang, and Gordon Wetzstein. 2023. LumiGAN: Unconditional Generation of Relightable 3D Human Faces. In *arXiv*.
- Yu Deng, Jiaolong Yang, Jianfeng Xiang, and Xin Tong. 2022. GRAM: Generative Radiance Manifolds for 3D-Aware Image Generation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Keyan Ding, Kede Ma, Shiqi Wang, and Eero P. Simoncelli. 2020. Image Quality Assessment: Unifying Structure and Texture Similarity. *CoRR* abs/2004.07728 (2020). <https://arxiv.org/abs/2004.07728>
- Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680.
- Jiatao Gu, Lingjie Liu, Peng Wang, and Christian Theobalt. 2022. StyleNeRF: A Style-based 3D Aware Generator for High-resolution Image Synthesis. In *International Conference on Learning Representations*.

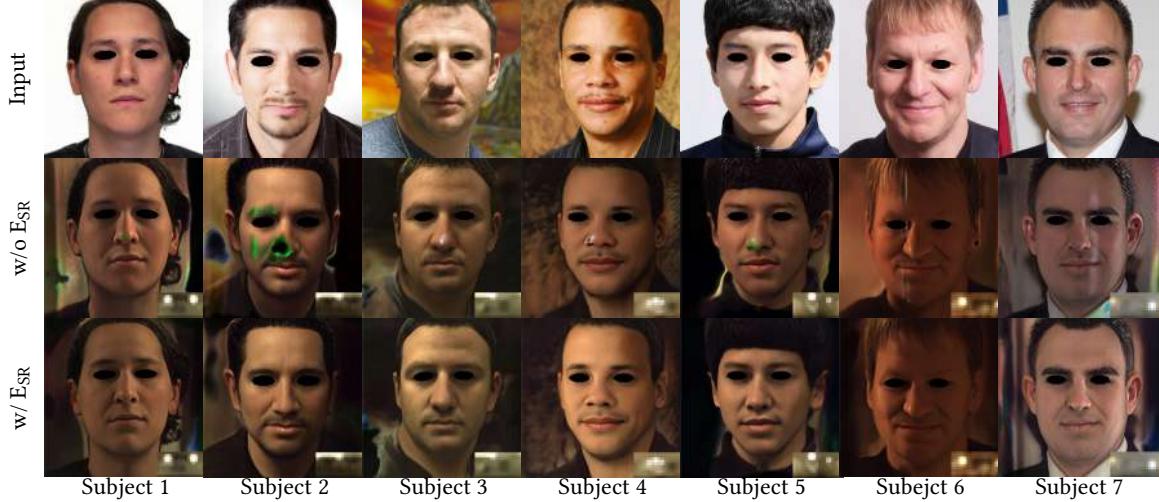


Fig. 22. Ablative Study on Influence of ESR: Without incorporating F_g , 3DPR struggles to generalize to unseen subjects, leading to noticeable artifacts in the face regions, particularly around the nose and mouth as shown in the middle row. Conversely, with F_g , there is a faithful preservation of the input identity, demonstrating the importance of this component in maintaining accuracy. We use $3DPR_w$ for evaluation.

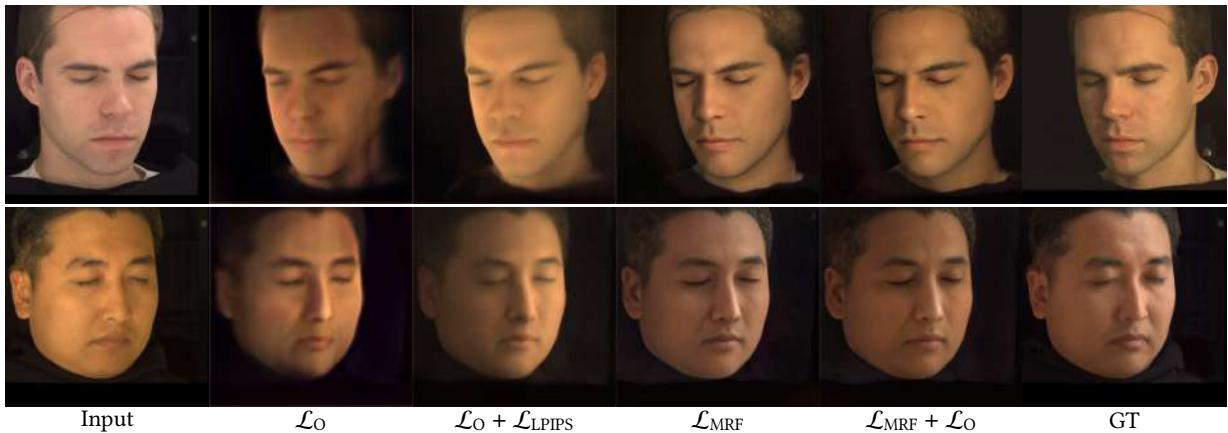


Fig. 23. Ablation Study on Losses. We show the impact of various loss functions on relighting quality. Specifically, \mathcal{L}_{MRF} leads to significant improvements. These subjects are sourced from the test set of the Lightstage dataset [Weyrich et al. 2006].

- Yang Haotian, Zheng Mingwu, Ma ChongYang, Lai Yu-Kun, Wan Pengfei, and Huang Haibin. 2024. VRMM: A Volumetric Relightable Morphable Head Model. In *SIGGRAPH 2024 Conference Proceedings*.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV, USA) (CVPR '16). IEEE, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- Mingming He, Pascal Clausen, Ahmet Levent Taşel, Li Ma, Oliver Pilarski, Wengi Xian, Laszlo Rikker, Xueming Yu, Ryan Burgert, Ning Yu, and Paul Debevec. 2024. DiffRelight: Diffusion-Based Facial Performance Relighting. In *SIGGRAPH Asia 2024 Conference Papers* (SA '24). Association for Computing Machinery, New York, NY, USA, Article 11, 12 pages. <https://doi.org/10.1145/3680528.3687644>
- Yang Hong, Bo Peng, Haiyao Xiao, Ligang Liu, and Juyong Zhang. 2022. HeadNeRF: A Real-time NeRF-based Parametric Head Model. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Kaiwen Jiang, Shu-Yu Chen, Hongbo Fu, and Lin Gao. 2023. NeRFFaceLighting: Implicit and Disentangled Face Lighting Representation Leveraging Generative Prior in Neural Radiance Fields. *ACM Transactions on Graphics (TOG)* (2023).
- Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. 2018. Progressive Growing of GANs for Improved Quality, Stability, and Variation. In *International Conference on Learning Representations (ICLR)*.

- Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. Analyzing and Improving the Image Quality of StyleGAN. In *Proc. CVPR*. Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.). <http://arxiv.org/abs/1412.6980>
- Oliver Klehm, Fabrice Roussel, Marios Papas, Derek Bradley, Christophe Hery, Bernd Bickel, Wojciech Jarosz, and Thabo Beeler. 2015. Recent Advances in Facial Appearance Capture. *Computer Graphics Forum (Proceedings of Eurographics - State of the Art Reports)* 34, 2 (May 2015), 709–733. <https://doi.org/10.1111/cgf.12540>
- Jeong-gi Kwak, Yuanming Li, Dongsik Yoon, Donghyeon Kim, David Han, and Hanseok Ko. 2022. Injecting 3D Perception of Controllable NeRF-GAN into StyleGAN for Editable Portrait Image Synthesis. In *European Conference on Computer Vision*. Springer, 236–253.
- Alexandros Lattas, Stylianos Moschoglou, Stylianos Ploumpis, Baris Gecer, Abhijeet Ghosh, and Stefanos P Zafeiriou. 2021. AvatarMe++: Facial Shape and BRDF Inference with Photorealistic Rendering-Aware GANs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021).
- Junxuan Li, Shunsuke Saito, Tomas Simon, Stephen Lombardi, Hongdong Li, and Jason Saragih. 2023. MEGANE: Morphable Eyeglass and Avatar Network. In *Proceedings*

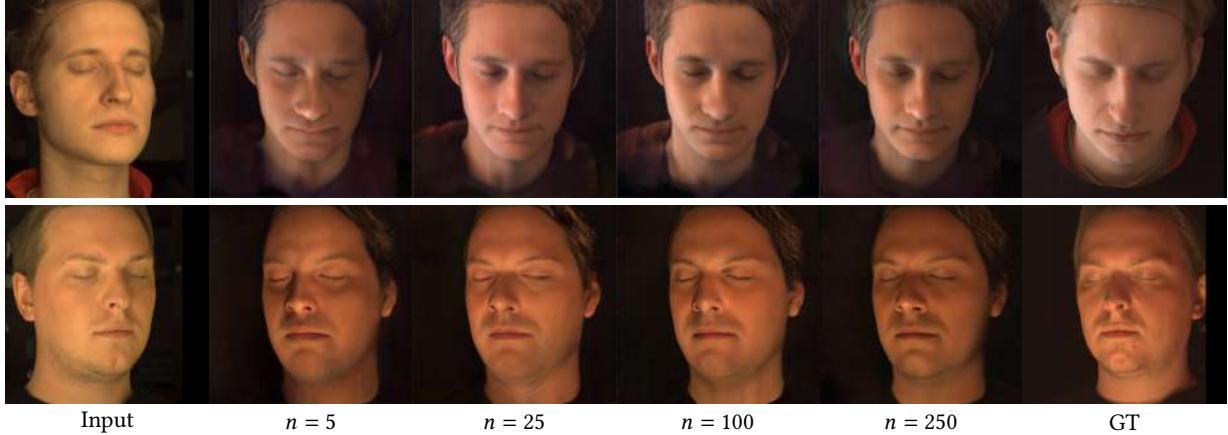
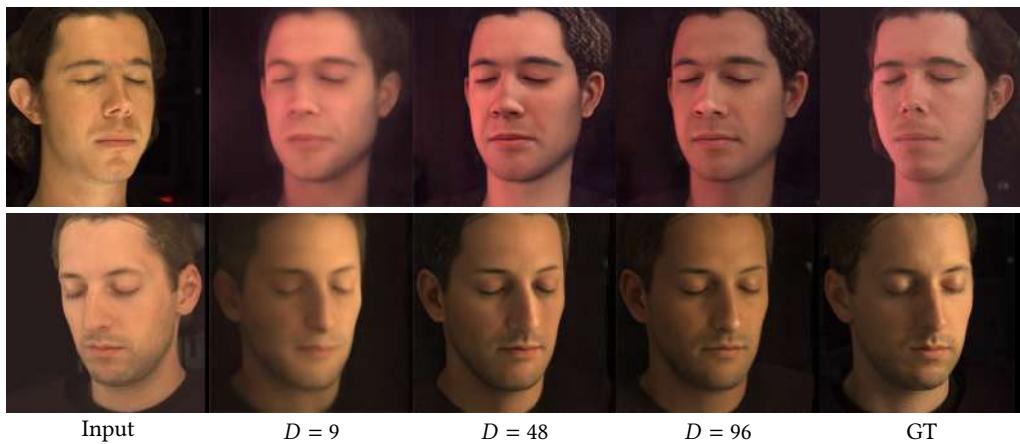


Fig. 24. Qualitative Results: Ablative Study on Number of Training Subjects.

Fig. 25. Qualitative Results: Ablative Study on dimensionality of F_o .

- of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 12769–12779.
- Shanchuan Lin, Andrey Ryabtsev, Soumyadip Sengupta, Brian Curless, Steve Seitz, and Ira Kemelmacher-Shlizerman. 2020. Real-Time High-Resolution Background Matting. *arXiv* (2020), arXiv–2012.
- Steven R. Livingstone and Frank A. Russo. 2018. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PLOS ONE* 13, 5 (05 2018), 1–35. <https://doi.org/10.1371/journal.pone.0196391>
- Julieta Martinez, Emily Kim, Javier Romero, Timur Bagautdinov, Shunsuke Saito, Shouyu Yu, Stuart Anderson, Michael Zollhöfer, Te-Li Wang, Shaojie Bai, Chenghui Li, Shih-En Wei, Rohan Joshi, Wyatt Borsos, Tomas Simon, Jason Saragih, Paul Theodosis, Alexander Greene, Anjani Josyula, Silvio Mana Maeta, Andrew I. Jewett, Simon Venshtain, Christopher Heilman, Yueh-Tung Chen, Sidi Fu, Mohamed Ezzeldin A. Elshaer, Tingfang Du, Longhua Wu, Shen-Chi Chen, Kai Kang, Michael Wu, Youssef Emad, Steven Longay, Ashley Brewer, Hitesh Shah, James Booth, Taylor Koska, Kayla Haidle, Matt Andromalos, Joanna Hsu, Thomas Dauer, Peter Selednik, Tim Godisart, Scott Ardisson, Matthew Cipperly, Ben Humberston, Lon Farr, Bob Hansen, Peihong Guo, Dave Braun, Steven Krenn, He Wen, Lucas Evans, Natalia Fadeeva, Matthew Stewart, Gabriel Schwartz, Divam Gupta, Gyeongsik Moon, Kaiwen Guo, Yuan Dong, Yichen Xu, Takaaki Shiratori, Fabian Prada, Bernardo R. Pires, Bo Peng, Julia Buffalini, Autumn Trimble, Kevyn McPhail, Melissa Schoeller, and Yaser Sheikh. 2024. Codec Avatar Studio: Paired Human Captures for Complete, Driveable, and Generalizable Avatars. *NeurIPS Track on Datasets and Benchmarks* (2024).
- Yiqun Mei, Yu Zeng, He Zhang, Zhixin Shu, Xuaner Zhang, Sai Bi, Jianming Zhang, HyunJoon Jung, and Vishal M Patel. 2024. Holo-Relighting: Controllable Volumetric Portrait Relighting from a Single Image. *arXiv preprint arXiv:2403.09632* (2024).
- Abhimitra Meka, Christian Häne, Rohit Pandey, Michael Zollhöfer, Sean Fanello, Graham Fyffe, Adarsh Kowdle, Xueming Yu, Jay Busch, Jason Dougarian, Peter Denny, Sofien Bouaziz, Peter Lincoln, Matt Whalen, Geoff Harvey, Jonathan Taylor, Shahram Izadi, Andrea Tagliasacchi, Paul Debevec, Christian Theobalt, Julien Valentin, and Christoph Rhemann. 2019. Deep Reflectance Fields: High-Quality Facial Reflectance Field Inference from Color Gradient Illumination. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)* (2019).
- Qiang Meng, Shichao Zhao, Zhida Huang, and Feng Zhou. 2021. MagFace: A universal representation for face recognition and quality assessment. In *CVPR*.
- Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *European Conference on Computer Vision (ECCV)*.
- Thomas Nestmeyer, Jean-François Lalonde, Iain Matthews, and Andreas M Lehrmann. 2020. Learning Physics-guided Face Relighting under Directional Light. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Rohit Pandey, Sergio Orts-Escalano, Chloe LeGendre, Christian Häne, Sofien Bouaziz, Christoph Rhemann, Paul Debevec, and Sean Fanello. 2021. Total Relighting: Learning to Relight Portraits for Background Replacement. *ACM Transactions on Graphics (Proceedings SIGGRAPH)* (2021).
- Anurag Ranjan, Kwang Moo Yi, Jen-Hao Rick Chang, and Oncel Tuzel. 2023. FaceLit: Neural 3D Relightable Faces. In *CVPR*. <https://arxiv.org/abs/2303.15437>
- Pramod Rao, Mallikarjun B. R, Gereon Fox, Tim Weyrich, Bernd Bickel, Hanspeter Pfister, Wojciech Matusik, Fangneng Zhan, Ayush Tewari, Christian Theobalt, and Elgarib Mohamed. 2023. A Deeper Analysis of Volumetric Relightable Faces. *International Journal of Computer Vision* (10 2023), 1–19. <https://doi.org/10.1007/s11263-023-01899-3>

- 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. 2022. VoRF: Volumetric Relightable Faces. (2022).
- Pramod Rao, Gereon Fox, Abhimitra Meka, Mallikarjun B R, Fangneng Zhan, Tim Weyrich, Bernd Bickel, Hans-Peter Seidel, Hanspeter Pfister, Wojciech Matusik, Mohamed Elgharib, and Christian Theobalt. 2024a. Lite2Relight: 3D-aware Single Image Portrait Relighting. (2024).
- Pramod Rao, Gereon Fox, Abhimitra Meka, Mallikarjun B R, Fangneng Zhan, Tim Weyrich, Bernd Bickel, Hans-Peter Seidel, Hanspeter Pfister, Wojciech Matusik, Mohamed Elgharib, and Christian Theobalt. 2024b. Lite2Relight: 3D-aware Single Image Portrait Relighting. (2024).
- Viktor Rudnev, Mohamed Elgharib, William Smith, Lingjie Liu, Vladislav Golyanik, and Christian Theobalt. 2022. NeRF for Outdoor Scene Relighting. In *European Conference on Computer Vision (ECCV)*.
- Shunsuke Saito, Gabriel Schwartz, Tomas Simon, Junxuan Li, and Giljoo Nam. 2024. Relightable Gaussian Codec Avatars. In *CVPR*.
- Kripasindhu Sarkar, Marcel C. Bühl, Gengyan Li, Daoye Wang, Delio Vicini, Jérémie Rivière, Yinda Zhang, Sergio Orts-Escalano, Paulo Gotardo, Thabo Beeler, and Abhimitra Meka. 2023. LitNeRF: Intrinsic Radiance Decomposition for High-Quality View Synthesis and Relighting of Faces. In *SIGGRAPH Asia 2023 Conference Papers (<conf-loc>, <city>Sydney/<city>, <state>NSW/<state>, <country>Australia/<country>, </conf-loc>) (SA '23)*. Association for Computing Machinery, New York, NY, USA, Article 42, 11 pages. <https://doi.org/10.1145/3610548.3618210>
- Soumyadip Sengupta, Angjoo Kanazawa, Carlos D. Castillo, and David W. Jacobs. 2018. SfsNet: Learning Shape, Reflectance and Illuminance of Faces in the Wild. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- YiChang Shih, Sylvain Paris, Connelly Barnes, William T. Freeman, and Frédéric Durand. 2014. Style transfer for headshot portraits. *ACM Trans. Graph.* 33, 4, Article 148 (July 2014), 14 pages. <https://doi.org/10.1145/2601097.2601137>
- Z. Shu, E. Yumer, S. Hadap, K. Sunkavalli, E. Shechtman, and D. Samaras. 2017. Neural Face Editing with Intrinsic Image Disentangling. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *International Conference on Learning Representations*.
- Vanessa Sklyarova, Egor Zakharov, Otmar Hilliges, Michael J Black, and Justus Thies. 2023. HAAR: Text-Conditioned Generative Model of 3D Strand-based Human Hairstyles. *ArXiv* (Dec 2023).
- Pratul P. Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T. Barron. 2021. NeRV: Neural Reflectance and Visibility Fields for Relighting and View Synthesis. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Giota Stratou, Abhijeet Ghosh, Paul Debevec, and Louis-Philippe Morency. 2011. Effect of illumination on automatic expression recognition: A novel 3D relightable facial database. In *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 611–618. <https://doi.org/10.1109/FG.2011.5771467>
- Tiancheng Sun, Kai-En Lin, Sai Bi, Zexiang Xu, and Ravi Ramamoorthi. 2021. NeLF: Neural Light-transport Field for Portrait View Synthesis and Relighting. In *Eurographics Symposium on Rendering*.
- Tiancheng Sun, Zexiang Xu, Xiuming Zhang, Sean Fanello, Christoph Rhemann, Paul Debevec, Yun-Ta Tsai, Jonathan T. Barron, and Ravi Ramamoorthi. 2020. Light Stage Super-Resolution: Continuous High-Frequency Relighting. In *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)*.
- Feitong Tan, Sean Fanello, Abhimitra Meka, Sergio Orts-Escalano, Danhang Tang, Rohit Pandey, Jonathan Taylor, Ping Tan, and Yinda Zhang. 2022. VoLux-GAN: A Generative Model for 3D Face Synthesis with HDRI Relighting. *arXiv:2201.04873 [cs.CV]*
- Zachary Teed and Jia Deng. 2020. RAFT: Recurrent All-Pairs Field Transforms for Optical Flow. In *European Conference on Computer Vision*.
- Ayush Tewari, Mohamed Elgharib, Gaurav Bharaj, Florian Bernard, Hans-Peter Seidel, Patrick Pérez, Michael Zöllhofer, and Christian Theobalt. 2020a. StyleRig: Rigging StyleGAN for 3D Control over Portrait Images, CVPR 2020. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.
- Ayush Tewari, Mohamed Elgharib, Mallikarjun BR, Florian Bernard, Hans-Peter Seidel, Patrick Pérez, Michael Zöllhofer, and Christian Theobalt. 2020b. PIE: Portrait Image Embedding for Semantic Control. *ACM Transactions on Graphics (Proceedings SIGGRAPH Asia)* 39, 6 (December 2020). <https://doi.org/10.1145/3414685.3417803>
- Yi Wang, Xin Tao, Xiaojuan Qi, Xiaoyong Shen, and Jiaya Jia. 2018. Image Inpainting via Generative Multi-column Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*. 331–340.
- Zhibo Wang, Xin Yu, Ming Lu, Quan Wang, Chen Qian, and Feng Xu. 2020. Single Image Portrait Relighting via Explicit Multiple Reflectance Channel Modeling. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)* (2020).
- Andreas Wenger, Andrew Gardner, Chris Tchou, Jonas Unger, Tim Hawkins, and Paul Debevec. 2005. Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM Trans. Graph.* 24, 3 (July 2005), 756–764. <https://doi.org/10.1145/1073204.1073258>
- Tim Weyrich, Wojciech Matusik, Hanspeter Pfister, Bernd Bickel, Craig Donner, Chien Tu, Janet McAndless, Jinho Lee, Addy Ngan, Henrik Wann Jensen, and Markus Gross. 2006. Analysis of Human Faces using a Measurement-Based Skin Reflectance Model. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)* (2006).
- Shuco Yamaguchi, Shunsuke Saito, Koki Nagano, Yajie Zhao, Weikai Chen, Kyle Olzewski, Shigeo Morishima, and Hao Li. 2018. High-fidelity facial reflectance and geometry inference from an unconstrained image. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)* (2018).
- Haotian Yang, Mingwu Zheng, Wanquan Feng, Haibin Huang, Yu-Kun Lai, Pengfei Wan, Zhongyuan Wang, and Chongyang Ma. 2023. Towards practical capture of high-fidelity relightable avatars. In *SIGGRAPH Asia 2023 Conference Papers*. 1–11.
- Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. 2021. pixelNeRF: Neural Radiance Fields from One or Few Images. In *CVPR*.
- Ziyang Yuan, Yiming Zhu, Yu Li, Hongyu Liu, and Chun Yuan. 2023. Make Encoder Great Again in 3D GAN Inversion through Geometry and Occlusion-Aware Encoding. *arXiv preprint arXiv:2303.12326* (2023).
- Egor Zakharov, Vanessa Sklyarova, Michael J Black, Giljoo Nam, Justus Thies, and Otmar Hilliges. 2024. Human Hair Reconstruction with Strand-Aligned 3D Gaussians. *ArXiv* (Sep 2024).
- Chong Zeng, Yue Dong, Pieter Peers, Youkang Kong, Hongzhi Wu, and Xin Tong. 2024. DiLightNet: Fine-grained Lighting Control for Diffusion-based Image Generation. In *ACM SIGGRAPH 2024 Conference Papers*.
- Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2025. Scaling In-the-Wild Training for Diffusion-based Illumination Harmonization and Editing by Imposing Consistent Light Transport. In *The Thirteenth International Conference on Learning Representations*. <https://openreview.net/forum?id=u1cQYxRIIH>
- Longwen Zhang, Qixuan Zhang, Minye Wu, Jingyi Yu, and Lan Xu. 2021b. Neural Video Portrait Relighting in Real-Time via Consistency Modeling. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 802–812.
- Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *CVPR*.
- Xuaner Zhang, Jonathan T. Barron, Yun-Ta Tsai, Rohit Pandey, Xiuming Zhang, Ren Ng, and David E. Jacobs. 2020. Portrait Shadow Manipulation. In *ACM Transactions on Graphics (TOG)*.
- Xiuming Zhang, Pratul P. Srinivasan, Boyang Deng, Paul Debevec, William T. Freeman, and Jonathan T. Barron. 2021a. NeRFactor: Neural Factorization of Shape and Reflectance under an Unknown Illumination. *ACM Transactions on Graphics* (2021).
- Peng Zheng, Dehong Gao, Deng-Ping Fan, Li Liu, Jorma Laaksonen, Wanli Ouyang, and Nicu Sebe. 2024. Bilateral Reference for High-Resolution Dichotomous Image Segmentation. *CAAI Artificial Intelligence Research* (2024).
- Yang Zheng, Menglei Chai, Delio Vicini, Yuxiao Zhou, Yinghao Xu, Leonidas Guibas, Gordon Wetzstein, and Thabo Beeler. 2025. GroomLight: Hybrid Inverse Rendering for Relightable Human Appearance Modeling. *arxiv*.
- Hao Zhou, Sunil Hadap, Kalyan Sunkavalli, and David W. Jacobs. 2019. Deep Single-Image Portrait Relighting. In *The IEEE International Conference on Computer Vision (ICCV)*.
- Taotao Zhou, Kai He, Di Wu, Teng Xu, Qixuan Zhang, Kuixiang Shao, Wenzheng Chen, Lan Xu, and Jingyi Yu. 2023. Relightable Neural Human Assets from Multi-view Gradient Illuminations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4315–4327.