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

PRAMOD RAO, Max Planck Institute for Informatics, Germany and Saarbrücken Research Center for Visual Computing, Interaction and Artificial Intelligence, Germany
ABHIMITRA MEKA*, Google Inc., USA
XILONG ZHOU*, Max Planck Institute for Informatics, Germany
GEREON FOX, Max Planck Institute for Informatics, Germany
MALLIKARJUN B R, Max Planck Institute for Informatics, Germany
FANGNENG ZHAN, Harvard University, USA
TIM WEYRICH, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Germany
BERND BICKEL, ETH Zürich, Switzerland and IST Austria, Switzerland
HANSPETER PFISTER, Harvard University, USA
WOJCIECH MATUSIK, Computer Science and Artificial Intelligence Laboratory (CSAIL), USA and Massachusetts Institute of Technology (MIT), USA
THABO BEELER, Google Inc., USA
MOHAMED ELGHARIB, Max Planck Institute for Informatics, Germany
MARC HABERMANN, Max Planck Institute for Informatics, Germany and Saarbrücken Research Center for Visual Computing, Interaction and Artificial Intelligence, Germany
CHRISTIAN THEOBALT, Max Planck Institute for Informatics, Germany and Saarbrücken Research Center for Visual Computing, Interaction and Artificial Intelligence, Germany

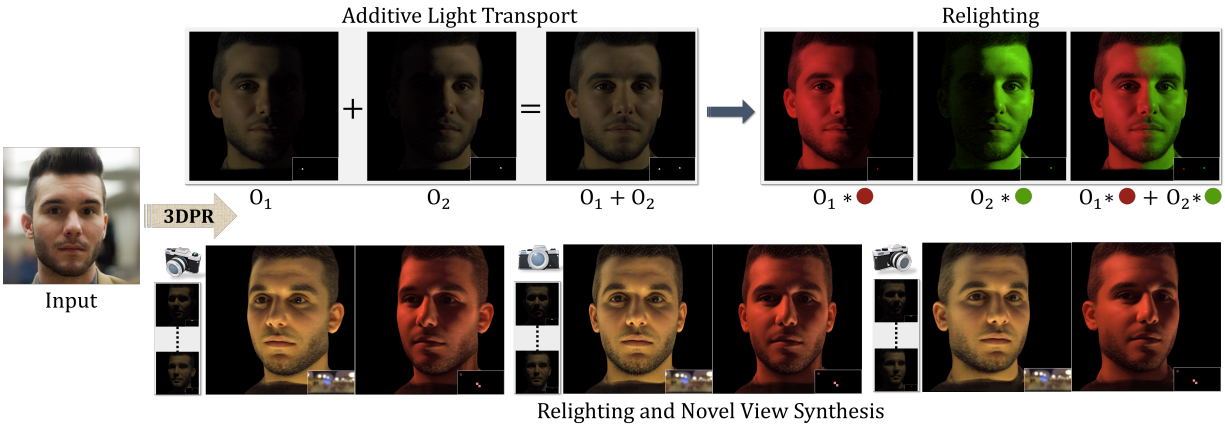

Fig. 1. We present *3DPR*, a monocular 3D portrait relighting method that can synthesize novel views under desired illumination. Given a monocular input image, *3DPR* predicts a reflectance basis in the form of One-Light-At-a-Time (OLAT) images of the subject (top row). By linearly combining the OLAT 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.

*Equal contribution

Authors' Contact Information: Pramod Rao, Max Planck Institute for Informatics, Saarbrücken, Germany and Saarbrücken Research Center for Visual Computing, Interaction and Artificial Intelligence, Saarbrücken, Germany, prao@mpi-inf.mpg.de; Abhimitra Meka, Google Inc., San Fransisco, USA, abhim@google.com; Xilong Zhou, Max Planck Institute for Informatics, Saarbrücken, Germany, xzhou@mpi-inf.mpg.de; Gereon Fox, Max Planck Institute for Informatics, Saarbrücken, Germany, gfox@mpi-inf.mpg.de; Mallikarjun B R, Max Planck Institute for Informatics, Saarbrücken, Germany, mbr@mpi-inf.mpg.de; Fangneng Zhan, Harvard University, Cambridge, USA, fnzhan@mit.edu; Tim Weyrich, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Nürnberg, Germany, tim.weyrich@fau.de; Bernd Bickel, ETH Zürich, Zürich, Switzerland and IST Austria, Zürich, Switzerland, bernd.bickel@ist.ac.at; Hanspeter Pfister, Harvard University, Cambridge, USA, pfister@g.harvard.edu; Wojciech Matusik, Computer Science and Artificial Intelligence Laboratory (CSAIL), Cambridge, USA and Massachusetts Institute of Technology (MIT), Cambridge, USA, wojciech@csail.mit.edu; Thabo Beeler, Google Inc., Zürich, USA, tbeeler@google.com; Mohamed Elgharib, Max Planck Institute for Informatics, Saarbrücken, Germany, mohamedaelgharib@gmail.com; Marc Habermann, Max Planck Institute for Informatics, Saarbrücken, Germany and Saarbrücken Research Center for Visual Computing, Interaction and Artificial Intelligence, Saarbrücken, Germany, mhaberma@mpi-inf.mpg.de; Christian Theobalt, Max Planck Institute for Informatics, Saarbrücken, Germany and Saarbrücken Research Center for Visual Computing, Interaction and Artificial Intelligence, Saarbrücken, Germany, theobalt@mpi-inf.mpg.de.

SA Conference Papers '25, Hong Kong, Hong Kong

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 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, 12 pages. https://doi.org/10.1145/3757377.3763962

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

ACM ISBN 979-8-4007-2137-3/25/12
https://doi.org/10.1145/3757377.3763962

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	✓
FaceOLAT	331	139	40	4K	✓

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 pretrained 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 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 illumination, the relit image $\mathbf{C}$ is computed as: $\mathbf{C} \approx \sum_{l \in I} f_l \cdot \mathrm{O}(l)$, where I denotes the set of OLAT lighting directions, $\mathrm{O}(l)$ 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 preprocessing 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 $\mathbf{z} \in \mathbb{R}^{1\times512}$ into an intermediate latent code $\mathbf{w} \in \mathbb{R}^{14\times512}$, which is passed to a StyleGAN2 generator $\mathrm{G_{gen}}$ [Karras et al. 2020] to produce tri-planar features $\mathrm{F_g} \in \mathbb{R}^{96\times256\times256}$. 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 $\mathbf{C}$, we employ a pre-trained encoder-based inversion network $\mathcal{E}$ [Yuan et al. 2023], such that $\mathrm{F_g} = \mathcal{E}(\mathbf{C})$ (see Fig. 3). The tri-plane features $\mathrm{F_g}$ are decoded by EG3D's MLP-based decoder $\mathrm{G_{dec}}$ and rendered volumetrically from a given camera viewpoint $\mathbf{v}$ to produce a low-resolution RGB image $\mathrm{c_{rgb}} \in \mathbb{R}^{3\times128\times128}$ and a high-frequency feature image $\mathrm{c_{hf}} \in \mathbb{R}^{29\times128\times128}$. We specifically leverage $\mathrm{c_{hf}}$, 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 $\mathrm{F_g}$ from the inversion stage, we concatenate it with the light direction ω_i and pass the result to our OLAT encoder $\mathrm{R_{enc}}$, which predicts reflectance-aware tri-plane features $\mathrm{F_o} \in \mathbb{R}^{96\times256\times256}$ as: $\mathrm{F_o} = \mathrm{R_{enc}}(\mathrm{F_g}, \omega_i)$. $\mathrm{R_{enc}}$ 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 $\mathbf{v}$ and use our OLAT decoder $\mathrm{R_{dec}}$, a lightweight single-layer MLP, which outputs both a low-resolution RGB image $\mathrm{o_{rgb}} \in \mathbb{R}^{3\times128\times128}$ and a high-frequency reflectance feature map $\mathrm{o_{hf}} \in \mathbb{R}^{29\times128\times128}$ via NeRF-style volume rendering: $\mathrm{o_{rgb}}, \mathrm{o_{hf}} = \mathrm{R_{dec}}(\mathrm{F_o}, \mathbf{v})$. If we directly fed $\mathrm{o_{rgb}}$ and $\mathrm{o_{hf}}$ into the super-resolution module $\mathrm{SR_o}$, that module could easily overfit to the realitvely small number of subjects in the OLAT dataset. To prevent this, we introduce a feature fusion module $\mathrm{E_{SR}}$, which combines $\mathrm{o_{hf}}$ with the high-frequency identity features $\mathrm{c_{hf}}$ obtained from the inversion stage (see Sec. 4.1). The fused feature map $\mathrm{p_{hf}} \in \mathbb{R}^{29\times128\times128}$ is computed as: $\mathrm{p_{hf}} = \mathrm{E_{SR}}(\mathrm{c_{hf}} \oplus \mathrm{o_{hf}})$, where $\oplus$ denotes channel-wise concatenation. Since $\mathrm{SR_o}$ was pretrained to use $\mathrm{c_{hf}}$, the module

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.

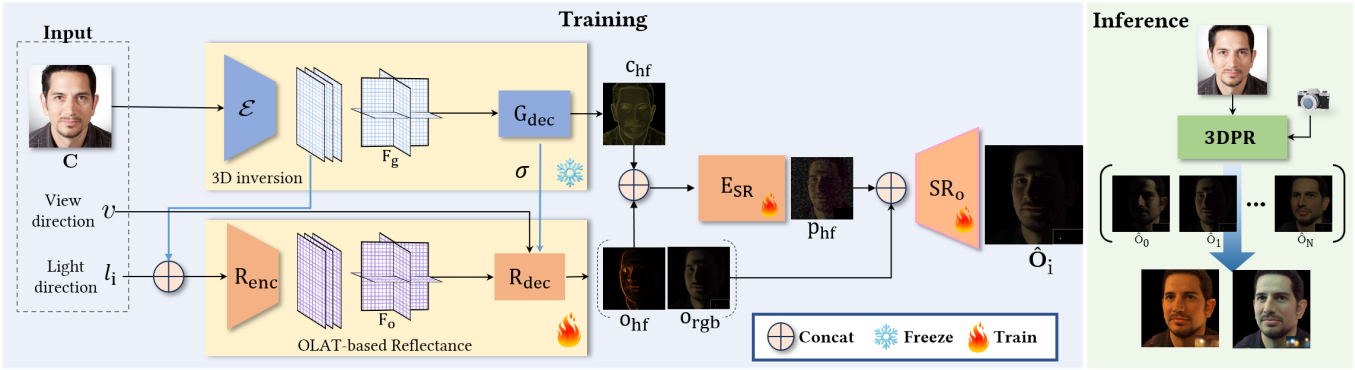

Fig. 3. Given one portrait image C, *3DPR* renders the subject from a novel viewpoint under new lighting. **Left**: During the training stage, C is first fed into the 3D-aware encoder $\mathcal{E}$ [Yuan et al. 2023] to produce tri-planar features F_g, concatenated with a given light direction l_i. Then, 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 $\hat{\mathbf{O}}$. 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 synthesize OLATs, which are then linear combined for novel illumination.

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 $\hat{\mathbf{O}} = SR_o(o_{rgb} \oplus p_{hf}) \in \mathbb{R}^{3\times512\times512}$. 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 $\mathbf{O}_i$ from the light stage, we supervise our predicted OLATs $\hat{\mathbf{O}}_i$ using the $L1$ loss $\mathcal{L}_O = \|\hat{\mathbf{O}}_i - \mathbf{O}_i\|_1$. 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 $\hat{\mathbf{O}}_i$ and $\mathbf{O}_i$ in a feature space extracted from a pre-trained VGG19 network [Simonyan and Zisserman 2015]. The ID-MRF loss is computed as $\mathcal{L}_{\text{MRF}} = \mathcal{L}_{\text{M}}(\Phi_1(\hat{\mathbf{O}}_i, \mathbf{O}_i)) + \mathcal{L}_{\text{M}}(\Phi_2(\hat{\mathbf{O}}_i, \mathbf{O}_i))$, where Φ_1 and Φ_2 correspond to activations from the `conv3_2` and `conv4_2` layers of VGG19, respectively, and $\mathcal{L}_{\text{M}}$ denotes the matching function.

Final Objective. The total loss is given by $\mathcal{L} = \mathcal{L}_{\text{O}} + 0.3\mathcal{L}_{\text{MRF}}$, 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**

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,

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.

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**

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 ↑	LPIPS ↓	RMSE ↓	DISTS ↓	PSNR ↑
$\mathcal{L}_O$	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**

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 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 setups 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 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

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.

the results in Tab. 4. It is evident that supervision with only the $L1$ loss is insufficient to produce high-quality relighting results. While combining $\mathcal{L}_{\text{LPIPS}}$ [Zhang et al. 2018] with per-pixel $L1$ loss is a commonly adopted approach, this combination still leads to suboptimal performance, primarily due to this metric missing high-frequency details. While $\mathcal{L}_{\text{MRF}}$ alone produces satisfactory results, we find that combining it with $\mathcal{L}_{\text{O}}$ 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

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.

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 Comput- ing, 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.

References

Mallikarjun B R, Ayush Tewari, Abdallah Dib, Tim Weyrich, Bernd Bickel, Hans Peter Seidel, Hanspeter Pfister, Wojciech Matusik, Louis Chevallier, Mohamed A Elgharib, 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 Elgharib, 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*.

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.

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*.

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, 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, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. Analyzing and Improving the Image Quality of StyleGAN. In *Proc. CVPR*.

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/f7mb4b

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 (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, Shoou-I 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 Mano 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 Dourgarian, 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-Escolano, 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. Bühler, Gengyan Li, Daoye Wang, Delio Vicini, Jérémy Riviere, Yinda Zhang, Sergio Orts-Escolano, 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, Refectance 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 Frédo 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*.

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-Escolano, 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.

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).

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=u1cQYxRI1H

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).

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.