This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

POLAR: A Portrait OLAT Dataset and Generative Framework for Illumination-Aware Face Modeling

Zhuo Chen[1,2†] Chengqun Yang[1†] Zhuo Su[2*] Zheng Lv[2] Jingnan Gao[1]
Xiaoyuan Zhang[2] Xiaokang Yang[1] Yichao Yan[1*]

[1]Shanghai Jiao Tong University [2]PICO

{ningci5252, ycq0191, gjn0310, xkyang, yanyichao}@sjtu.edu.cn,
suzhuo13@gamil.com, {lvzheng.101, zhangxiaoyuan.1001}@bytedance.com

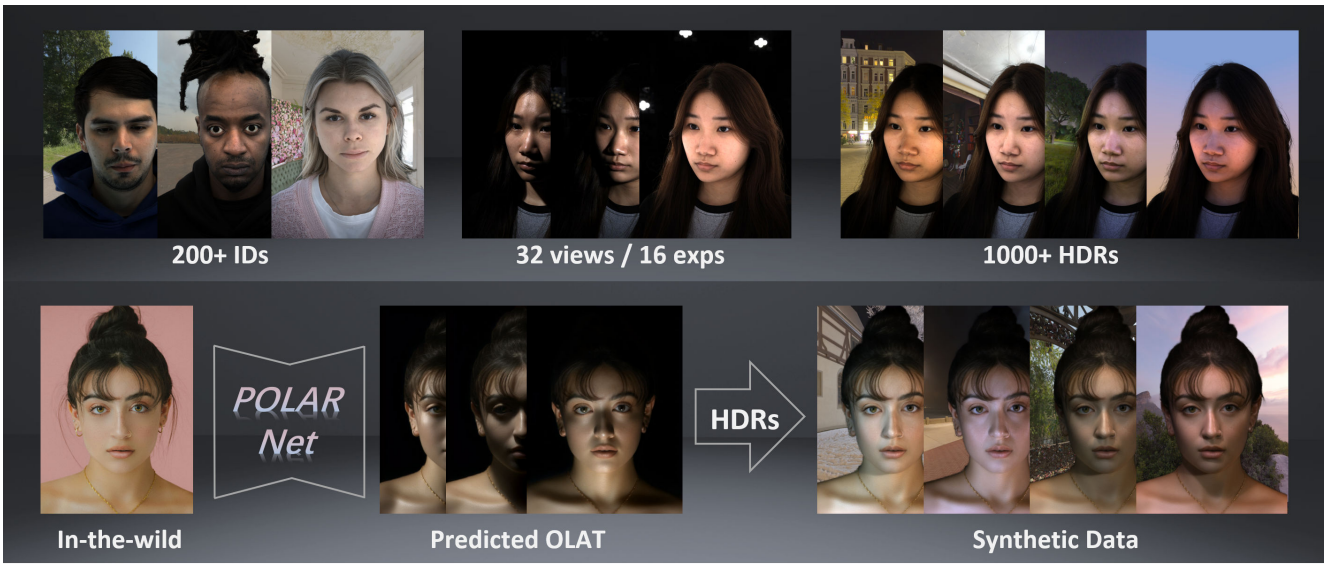


Figure 1. **POLAR** captures high-resolution OLAT facial data with diverse subjects and expressions, from which we synthesize large-scale HDR-relit portraits. **POLARNet** further learns to generate per-light OLAT responses from a single portrait, enabling scalable and physically consistent relighting under arbitrary HDR environments. Our project page: https://rex0191.github.io/POLAR/.

Abstract

*Face relighting aims to synthesize realistic portraits under novel illumination while preserving identity and geometry. However, progress remains constrained by the limited availability of large-scale, physically consistent illumination data. To address this, we introduce **POLAR**, a large-scale and physically calibrated One-Light-at-a-Time (OLAT) dataset containing over 200 subjects captured under 156 lighting directions, multiple views, and diverse expressions. Building upon POLAR, we develop a flow-based generative model **POLARNet** that predicts per-light OLAT responses from a single portrait, capturing fine-grained and direction-aware illumination effects while preserving facial identity. Unlike diffusion or background-conditioned methods that rely on statistical or contextual cues, our formulation models illumination as a continuous, physically interpretable transformation between lighting states, enabling scalable and controllable relighting. Together, POLAR and POLARNet form a unified illumination learning framework that links real data, generative synthesis, and physically grounded relighting, establishing a self-sustaining "chicken-and-egg" cycle for scalable and reproducible portrait illumination.*

1. Introduction

Face relighting is a long-standing problem in computer vision and computer graphics, with broad applications in photography, cinematography, virtual production, and digital humans. Illumination plays a central role in visual perception and storytelling, shaping how light and shadow convey emotion and intent in portrait imagery. Although neural ren-

[1]* Corresponding author, † Equal contribution

dering [43] and diffusion-based generation [20, 31, 64] have greatly advanced the realism of portrait relighting, progress is ultimately limited by the scarcity of large-scale, physically consistent illumination data. Robust relighting models depend not only on network design but also on the diversity and physical fidelity of the data used for training.

Among various illumination data, **One-Light-at-a-Time (OLAT)** [11] captures provide the most faithful measurement of facial light transport. Each capture records the subject's response to a single directional light, and since light transport is linear with illumination, arbitrary lighting can be reconstructed by linearly combining these single-light bases. This property makes OLAT a bridge between physics-based rendering and data-driven learning. However, existing OLAT datasets remain limited in both scale and accessibility. Most light-stage datasets [6, 20, 33–35, 37] used in production are closed-source, while academic releases [43, 48, 63] typically cover only a few subjects or expressions and are constrained to low-resolution settings. As a result, there is still a substantial gap between the growing need for high-fidelity, relightable facial data and the scarcity of publicly available resources.

To fill this gap, we introduce **POLAR** (Portrait OLAT for Relighting), a large-scale and physically calibrated open dataset for illumination-controlled facial capture. It contains over 200 subjects recorded under 156 calibrated lighting directions, 32 camera views, and 16 expressions, providing the most diverse portrait OLAT resource available to the open-source community. In addition to raw captures, POLAR offers relit portraits synthesized under diverse HDR environments, forming a comprehensive platform for data-driven illumination research. POLAR thus serves as a reliable and physically calibrated resource for developing and validating light transport models.

Although POLAR captures accurately model light transport through physically grounded single-light responses, they are inherently limited to the captured subjects, since each sequence must be recorded under controlled lightstage conditions and cannot generalize to arbitrary individuals. To overcome this limitation, we develop a flow-based generative model **POLARNet** that predicts per-light OLAT responses directly from a single uniformly lit portrait. The model learns that illumination variations follow consistent physical patterns rather than random image changes, allowing it to capture fine-grained, direction-aware relighting effects. Unlike conventional diffusion models [24, 31, 64] that denoise from Gaussian noise, our formulation learns a continuous trajectory that supports **one-step** inference between the uniform-light images and its corresponding OLAT responses, ensuring that only illumination varies while facial identity remain consistent. Conditioned on light direction, the model enables fast and physically controllable synthesis of single-light responses, which can be linearly recombined into full environment relighting results. Compared to background-conditioned approaches, our method explicitly models illumination structure and preserves physical interpretability. By generating per-light responses, the model provides a scalable and low-cost way to synthesize physically consistent illumination data.

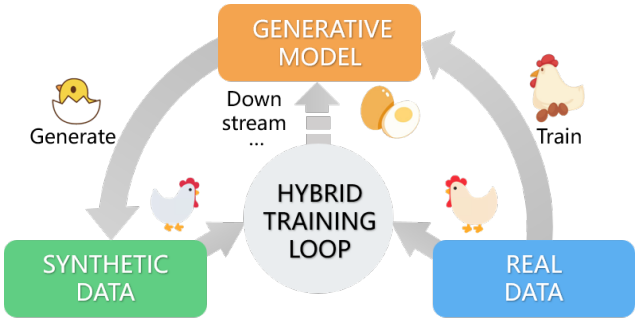


Figure 2. Overview of our "chicken-and-egg" co-evolution loop, where OLAT data guide model learning, and the generative model expands data diversity to benefit downstream tasks.

POLAR and its corresponding POLARNet form a unified framework for illumination learning that connects data creation, model training, and physically grounded synthesis. As shown in Fig. 2, this interaction forms a self-sustaining "chicken-and-egg" cycle, where real data guide the model toward physical accuracy, and the model in turn expands the diversity of illumination data beyond what can be captured in the lab. By releasing both the dataset and the model, we aim to promote an open ecosystem for portrait illumination research, supporting not only relighting but also broader downstream tasks such as lighting-aware video generation and digital human rendering. Our main contributions are summarized as follows:

- **An open, large-scale OLAT dataset.** We present POLAR, a physically calibrated and publicly available OLAT dataset containing 220 subjects, 32 camera views, and 16 expressions, with HDR-relit portraits.
- **A direction-conditioned model for OLAT generation.** We introduce a pair-wise supervised latent bridge formulation that learns physically consistent illumination transitions across light directions, enabling stable and identity-preserving one-step OLAT synthesis.
- **A unified illumination learning framework.** By combining real OLAT captures with generative synthesis, we bridge the gap between costly lightstage acquisition and scalable, physically grounded illumination learning.

2. Related Work

2.1. Light Stage

Light Stage [11] is an active illumination system enabling fine-grained lighting control. This capture system enables collection of detailed multi-view imagery under known illu-

mination, laying the foundation for developing and benchmarking realistic relighting methods [8, 16, 25, 28, 37, 43, 66]. Compared to other setups for capturing real-world multi-illumination data [4, 14, 32], One-Light-at-a-Time (OLAT) images are facilitated by the Light Stage's precise, programmable lighting control, enabling acquisition of detailed reflectance data. Leveraging the linearity of light transport, these OLAT captures can be linearly combined to achieve physically accurate image-based relighting under arbitrary illumination [11, 16, 38, 42]. A number of Light Stage OLAT datasets for static objects [30, 54, 67], human faces [43, 48, 63], hands [8], and full-body humans [52] have been released. However, existing resources are often constrained by limited accessibility or limited diversity, such as small identity counts and lack of expressive variations. Thus, we collect a fully open, high-quality and diverse OLAT dataset named **POLAR**, offering a large-scale, high-fidelity foundation for face relighting research. Recently, the concurrent FaceOLAT Dataset [40] has also been released, highlighting the continued scarcity of open OLAT data and complementing POLAR in scale and accessibility.

2.2. Portrait Relighting Methods

Portrait relighting has been extensively studied in both 2D images domain [7, 9, 10, 13, 15, 17, 18, 21, 22, 24–26, 31, 33, 34, 37, 41, 46, 49, 55–57, 60, 61, 68] and 3D avatars [2, 39, 43–45, 50, 51, 62]. In 2D domain, several methods leverage physics-guided decomposition frameworks to separate input portraits into intrinsic components and re-render them under HDR environment maps [25, 37, 49, 57]. While physically interpretable, these approaches often base on simplified BRDF reflectance models, failing to capture complex light transport phenomena such as subsurface scattering [13, 26]. To mitigate reliance on physical priors, recent works adopt diffusion-based models, enabling flexible, photorealistic relighting [24, 41, 61]. Meanwhile, general relighting frameworks tailored for objects and scenes utilize 3D reconstruction or background-based environmental cues [1, 3, 19, 23, 27, 36, 47, 58, 60, 64]. However, they are limited in identity-preserving portrait relighting. Alternatively, 3D avatar researches focus on constructing relightable human representations that support both environmental lighting and novel viewpoints. Despite offering strong geometric priors, 3D approaches often rely on parametric templates or specific identities, limiting their generalization and expressiveness [53, 65]. In contrast, **POLARNet** can generate OLAT data to achieve physically reliable relighting with cross-identity generalization capability.

3. The *POLAR* Dataset

We now introduce **POLAR**, a large-scale dataset of human faces captured under OLAT illumination. We first describes our acquisition setup (Sec. 3.1) and details the data modalities (Sec. 3.2), and then explains our relighting synthesis pipeline, (Sec. 3.3), and finally compare our POLAR with existing dataset (Sec. 3.4).



Figure 3. Our Light Stage consists of 156 LEDs and 32 synchronized cameras uniformly covering the full sphere. Each light is sequentially activated to produce OLAT captures.

3.1. Acquisition Setup

We constructed a dedicated Light Stage to acquire high-quality facial OLAT data. We employed a multi-view imaging system of **32** synchronized cameras. The Light Stage is equipped with **156** individually controllable LED light sources, distributed in a near-spherical configuration around the subject to cover the full sphere, as shown in Fig. 3. Each LED is activated sequentially to record One-Light-at-a-Time (OLAT) captures. For each subject, we captured a comprehensive set of **16 distinct expressions**, including neutral, smiling, surprise, and other variations. This adds intra-subject diversity and allows analysis of illumination effects under both static and expressive conditions. Participants were instructed to avoid glasses, makeup, or reflective accessories to minimize confounding effects. In total, our setup captures $\mathbf{157 \times 32 \times 16}$ images per subject (one all-light frame), yielding a dense, high-resolution, and multi-view OLAT dataset tailored for face relighting research. **Details are included in the supplementary material.**

3.2. Data Modalities and Annotations

The POLAR dataset provides multiple modalities supporting both relighting and general face modeling tasks:

- **Raw OLAT captures.** Each subject is recorded under 156 illuminations across 32 views and 16 expressions at 4K resolution.
- **Lighting annotations.** Each OLAT image has calibrated light directions (θ, ϕ) for illumination reconstruction.
- **Alpha mattes.** Per-pixel alpha maps are provided for precise compositing and boundary preservation.
- **Relit portraits.** From captured OLATs, we generate both uniformly lit and HDR-relit portraits.

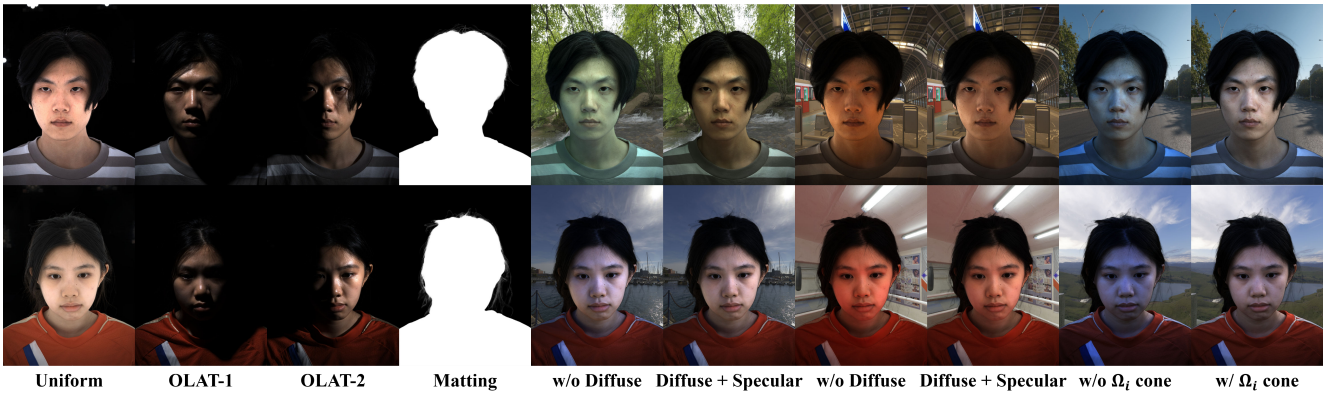


Figure 4. Comparison of OLAT relighting strategies in the POLAR dataset. We evaluate the effect of diffuse–specular separation and light-cone sampling range, showing that both refinements improve shading realism and highlight consistency.

3.3. Data Processing Pipeline

Foreground matting. High-quality foreground extraction is essential for realistic relighting. We employ an automatic alpha matting strategy based on the Matte-Anything [59] framework. To ensure robustness under varying illumination, uniform-light composites are used for matte generation, and mild gamma correction is applied to enhance foreground-background separation.

Relighting Synthesis. A central component of POLAR is the generation of relit portraits under arbitrary HDR environment maps, which transforms the raw OLAT captures into a scalable supervision resource. OLAT-based relighting relies on the linear relationship between illumination and appearance. Given single-light captures $\{I_i\}$ and their corresponding light directions $\{l_i\}$, the appearance under an environment map $E(\mathbf{l})$ can be approximated by

$$I_E \approx \sum_i w_i I_i, w_i = \int_{\Omega_i} E(\mathbf{l})\, d\mathbf{l}, \tag{1}$$

where $E(\mathbf{l})$ denotes the HDR radiance along light direction $\mathbf{l}$ on the unit sphere, and Ω_i denotes the solid-angle region centered at direction l_i, corresponding to a 30° illumination cone for each light. This linear compositing is widely used and physically sound for diffuse transport, but specular reflections can introduce exposure imbalance and mild color shifts when simulating colored illuminations.

To enhance the realism of relit results, we first separate diffuse and specular contributions to prevent global color bias [25] (Fig. 4). Diffuse weights are computed from grayscale intensity, ensuring illumination color does not overly tint skin tones, while specular weights preserve RGB information to reproduce colored highlights. The final relit image is a balanced combination of the two components:

$$I_E \approx \alpha \sum_i w_i^{\text{diff}} I_i + (1-\alpha) \sum_i I_i \odot w_i^{\text{spec}}, \tag{2}$$

where $\odot$ denotes element-wise multiplication between RGB specular weights and OLAT images, and α is a scene-dependent blending factor. We normalize the accumulated weights to match the total illumination energy of the HDR environment, maintaining consistent global exposure across relit results. A perceptual tone-mapping is applied to avoid over-saturation under high-intensity illumination, and HDR intensities are calibrated against measured light energy to ensure consistent exposure across relit results. We further compare light-cone sampling and point-light approximation in Fig. 4, where cone-based integration better captures high-intensity local lights and covers full HDR map.

In practice, we compute lighting weights by projecting each HDR environment map onto the 156 calibrated OLAT directions. For each subject, we produce relit portraits under a diverse set of HDR environments, sampled from both procedurally designed "studio-style" setups and publicly available sources such as Poly Haven. This synthesis expands each subject's illumination coverage, producing a diverse set of physically consistent relit portraits that serve as scalable supervision for downstream tasks.

Background Compositing. To enhance realism, each relit portrait is composited with a physically consistent background rendered from the corresponding HDR environment map. The background is precomputed using the subject camera parameters (viewpoint and focal length) to ensure geometric alignment.

3.4. Dataset Comparison

To highlight the contribution of POLAR, we compare it with existing datasets that provide illumination-controlled captures of human subjects in Table 1. Most large-scale illumination datasets developed within industry (e.g., by Adobe [6, 33, 34], NetFlix [20, 35], or Google [37]) remain proprietary and unavailable to the public due to privacy and licensing constraints. Academic efforts have begun to release open OLAT datasets, but most remain limited in scale, diversity, or accessibility. ICT-3DRFE [48] contains 23 identities under gradient and polarized light-

Table 1. Comparison of POLAR with existing OLAT-style face datasets. For Accessibility, ✓and ✗ indicate whether the dataset is or is not open source, while ○ means partially open source. POLAR uniquely combines large-scale OLAT captures, calibrated HDR-relit portraits, and full open-source availability, offering a more complete resource for illumination learning.

Dataset	Subjects	Views	Expressions	Frames	Resolution	Light Types	Accessibility
Total Relighting [37]	70	6	9	10.6M	4K	OLAT + HDR-relit	✗
Adobe Data [6, 33, 34]	59	4	5-15	1.2M	1K	OLAT + HDR-relit	✗
NetFlix Data [20, 35]	67	36	-	12M	4K	OLAT + HDR-relit	✗
ICT-3DRFE [48]	23	2	15	14K	1K	Gradient / Polarized	✓
Dynamic OLAT [63]	< 10	4	1	603K	4K	OLAT	○
Goliath-4 [43]	4 (public)	100+	-	> 1M	Mixed	Relightable captures	○
BecomingLit [45]	15	16	**23**	221K	3K	OLAT	✓
FaceOLAT [40]	139	**40**	4	5.5M	4K	OLAT	✓
POLAR (ours)	**220**	32	**16**	**28.8M**	4K	**OLAT + HDR-relit**	✓

ing, while the Dynamic OLAT dataset [63] records fewer than ten subjects, limiting its coverage for diverse relighting learning. Goliath-4 [43] focuses on multi-human or multi-modal assets, yet its full dataset remains under restricted access. The concurrent FaceOLAT [40] provides a multi-view face OLAT dataset emphasizing 3D reconstruction and evaluation. In contrast, our **POLAR** dataset contains **220** subjects captured under **156** calibrated light directions, across **32** views and **16** expressions at **4K** resolution. Combined with large-scale HDR relighting synthesis, POLAR includes more than **28.8 million** images in total, offering a substantial increase in both data volume and diversity over existing open OLAT resources. Along with calibrated meshes, HDR-relit portraits, and a synthesis pipeline, it provides a comprehensive foundation for portrait relighting and illumination-aware modeling. These features make POLAR one of the first open and large-scale OLAT datasets that bridge the gap between proprietary lightstage data and limited academic releases.

4. Flow-based OLAT Generative Model

Direct lightstage acquisition is costly and inaccessible to most users, and many effective relighting systems remain closed-source. Our **POLAR** dataset bridges this gap by providing a large-scale, publicly available OLAT resource. Beyond data release, we further leverage POLAR to train a generative model that learns per-light responses directly from a single uniformly-lit portrait, thus enabling low-cost and scalable OLAT-style data creation for arbitrary subjects. The overview is shown in Fig. 5.

4.1. Bridge Matching for Light Transport

Motivation. Relighting differs from conventional image translation tasks: illumination variations follow physically structured changes rather than random appearance shifts. Uniform-light and OLAT images of the same subject share identical appearance, while differing in light direction and intensity. However, standard diffusion models learn from noise-to-image mappings and often entangle illumination with texture or identity, leading to inconsistent shading and identity drift. To address this, we adopt a **flow-based latent transport**, which models a continuous trajectory between two semantically aligned endpoints, the uniform light source and the target OLAT image, avoiding random Gaussian initialization. This constrains the trajectory to evolve only along the illumination dimension while preserving the intrinsic identity shared by both endpoints.

Pair-wise latent transport. Unlike the original Flow Matching [29] or LBM [5], which learn a distributional mapping between two unaligned domains, we leverage *pair-wise aligned* supervision from OLAT captures, providing deterministic and physically grounded guidance for illumination transport. Each training pair $(x_u, x_l^{(\theta,\phi)})$ consists of a uniformly lit portrait and its OLAT portrait illuminated from direction (θ, ϕ). We use a lightweight variational autoencoder (E, D), following [5], to map both images into latent space $z_u = E(x_u)$ and $z_l = E(x_l^{(\theta,\phi)})$. The latent interpolation along the bridge is defined as

$$z_t = (1-t)z_u + tz_l + \sigma\sqrt{t(1-t)}\,\epsilon, \tag{3}$$

where $\epsilon \sim \mathcal{N}(0, I)$ introduces limited stochasticity, σ controls the noise level and $t \in [0, 1]$ parameterizes the illumination path. During training, a velocity field $v_\theta(z_t, t, c_{\text{dir}})$ predicts the instantaneous drift that transports z_t toward z_l, effectively learning a conditional light transport process.

Direction-conditioned velocity field. To achieve controllable illumination, the velocity network $v_\theta(\cdot)$ is conditioned on the light direction $c_{\text{dir}} = (\sin\theta, \cos\theta, \sin\phi, \cos\phi)$, represented via sinusoidal positional embeddings. The training objective minimizes the discrepancy between the predicted and target drift:

$$\mathcal{L}_{\text{LBM}} = \mathbb{E}_{t,\epsilon}\Big[\|v_\theta(z_t, t, c_{\text{dir}}) - (z_l - z_t)/(1-t)\|_2^2\Big]. \tag{4}$$

This formulation encourages the model to capture structured illumination priors and generalize beyond paired data,

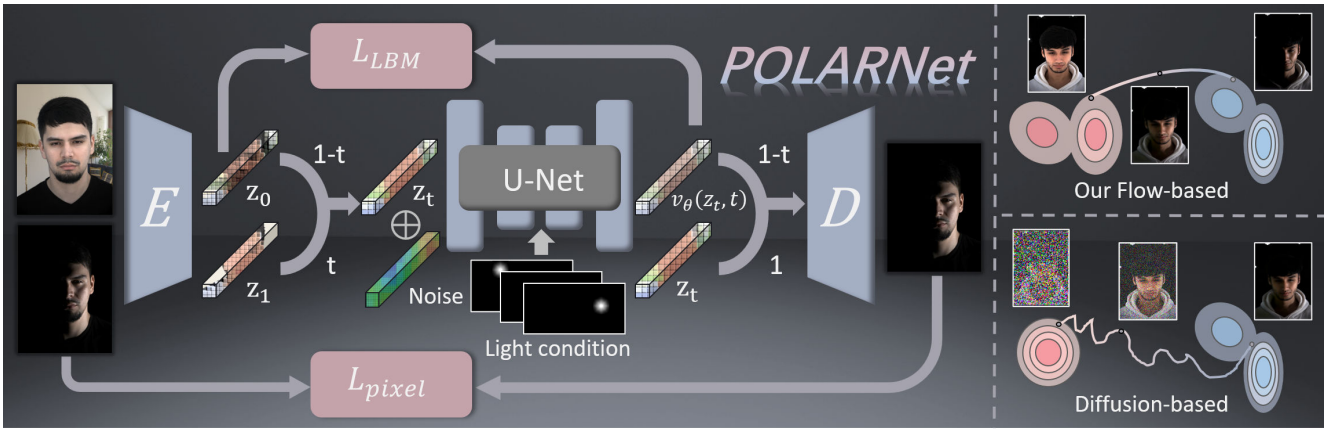


Figure 5. **Overview of our flow-based OLAT data generation framework.** Given a uniformly lit portrait, the encoder–decoder pair $(\mathbf{E}, \mathbf{D})$ maps both the input and its target OLAT image into latent space. Latent Bridge Matching learns a continuous, direction-conditioned transport between these endpoints, supervised by the velocity field loss $\mathcal{L}_{\text{LBM}}$. A conditional U-Net predicts the latent drift $v\theta(z_t, t, c_{\text{dir}})$ using the encoded light direction as input. During inference, a single forward step transports the latent z_u toward the illumination-specific latent z_l, enabling efficient generation of per-light OLAT responses for all calibrated directions. These synthesized OLATs can be linearly composed to render realistic relighting under arbitrary HDR environments.

enabling one-step generation of OLAT responses for unseen subjects under arbitrary lighting directions.

Inference. Given a uniformly lit portrait I_{uni} of an arbitrary person, we encode it into the latent space as $z_u = E(I_{\text{uni}})$. For a target light direction $\mathbf{L}{=}(\theta, \phi)$, the model directly predicts the transported latent $\hat{z}_l = z_u + (1{-}t)\, v_\theta(z_u, t{=}0, c_{\text{dir}})$ in a **single forward pass**, bypassing the iterative integration typically required in diffusion-based sampling. This one-step inference achieves fast and stable generation of OLAT responses while preserving the physical consistency learned during training. The decoded output $\hat{I}_{\text{olat}}(\mathbf{L}) = D(\hat{z}_l)$ represents the subject illuminated from direction $\mathbf{L}$, and by repeating this process across all calibrated lights, we can synthesize a complete OLAT sequence for the subject.

Once per-light responses are obtained, we can further compose full-environment relighting using the same physically grounded synthesis procedure described in Sec. 3.3, by linearly integrating the generated OLAT images according to the target HDR environment map. The inference and synthesis stages together form a unified, low-cost pipeline for generating realistic portrait illumination under arbitrary lighting conditions.

4.2. Training Details

Following Sec. 4.1, each training pair (x_0, x_1) consists of a uniformly lit portrait I_{uni} and its OLAT counterpart $I_{\text{olat}}(\mathbf{L})$ illuminated from direction $\mathbf{L}{=}(\theta, \phi)$. All images are encoded by the pre-trained VAE (E, D) into latent features (z_0, z_1), and intermediate samples z_t are constructed via stochastic interpolation. The model is optimized using the light-conditioned **LBM objective** in Eq. 4.

Identity consistency. Since OLAT captures often contain strong shadows, we introduce an identity loss computed in image space to preserve facial characteristics:

$$\mathcal{L}_{\text{id}} = \| f_{\text{id}}(D(z_t)) - f_{\text{id}}(D(z_0)) \|_1, \tag{5}$$

where f_{id} denotes a pre-trained face encoder (ArcFace [12]) and $D(\cdot)$ is the VAE decoder.

Energy and uncertainty-aware pixel loss. To balance photometric contribution across illumination intensities, we re-weight the pixel loss using an energy-aware mask, where each pixel is scaled by its relative brightness $w(x) = \min(1, \kappa I_{\text{olat}}(x)/\bar{I}_{\text{olat}})$, with $\bar{I}_{\text{olat}}$ denoting the mean intensity and κ controlling clipping. This weighting suppresses dark, low-signal regions and emphasizes well-lit areas, resulting in a more perceptually balanced pixel loss:

$$\mathcal{L}_{\text{pix}} = \| w \odot (\hat{I}_{\text{olat}} - I_{\text{olat}}) \|_1. \tag{6}$$

To further ensure global exposure consistency and prevent brightness drift between prediction and ground truth, we introduce an additional energy regularization term:

$$\mathcal{L}_{\text{energy}} = \Big\| \frac{\|\hat{I}_{\text{olat}}\|_1}{\|I_{\text{olat}}\|_1} - 1 \Big\|_1. \tag{7}$$

Together, these losses encourage physically plausible exposure balance and perceptually stable optimization under varying illumination.

Total objective. The overall training objective combines the LBM loss with identity, pixel, and energy regularization:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{LBM}} + \lambda_{\text{id}}\mathcal{L}_{\text{id}} + \lambda_{\text{pix}}\mathcal{L}_{\text{pix}} + \lambda_{\text{energy}}\mathcal{L}_{\text{energy}}. \tag{8}$$

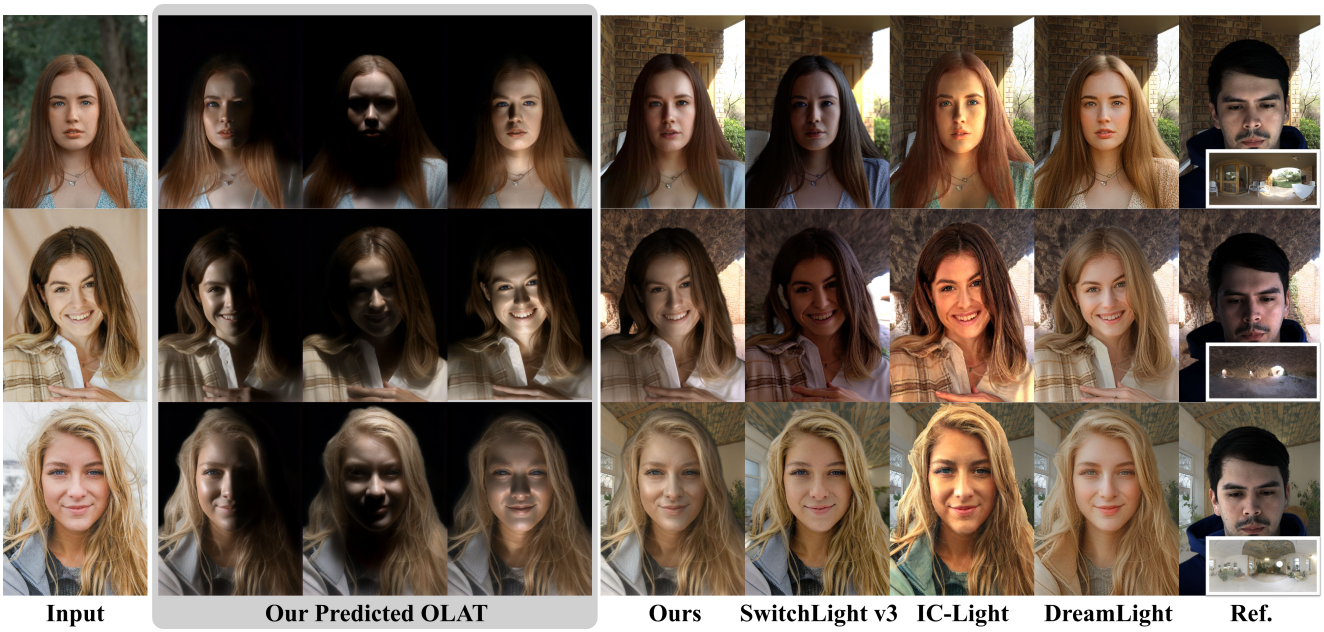


Figure 6. Qualitative comparison of portrait relighting in in-the-wild scenarios. From left to right: input portrait, our generated OLAT responses, relit results from our method, and comparisons with SwitchLight, IC-Light, and DreamLight. The rightmost column shows the reference relit image synthesized from real captured OLAT data.

5. Experiments

5.1. Comparison of Relit Image Synthesis Pipelines

To evaluate the fidelity of our OLAT generation model, we compare the synthesized OLAT images with the corresponding captured OLAT data from the POLAR dataset.

1. **Ground-truth OLAT composition.** We directly synthesize relit portraits by linearly combining captured OLAT responses with HDR environment maps.
2. **Generated OLAT → composition.** We first generate a full OLAT sequence for a new subject using our conditional diffusion model, and then compose relit portraits with environment maps.

As shown in Fig. 7, the generated OLAT results exhibit consistent illumination directionality and realistic shading transitions that closely resemble the captured ground truth. When these generated OLAT images are further combined to render environment-lit portraits, the resulting relit images align well with those produced from real OLAT captures, demonstrating that our model effectively learns the underlying light transport behavior.

5.2. Baseline Comparisons

We conduct qualitative comparisons on in-the-wild portraits to assess the visual quality and relighting accuracy between ours and three representative methods. **IC-Light** [64], and **DreamLight** [31] are background-condition relighting methods. **SwitchLight** [25] adopts a physics-inspired intrinsic decomposition to model relightable portraits.

Table 2. Quantitative comparison with state-of-the-art image-based relighting methods.

Methods	LPIPS↓	PSNR↑	SSIM↑
SwitchLight [25]	0.168	20.69	**0.84**
IC-Light [64]	0.314	18.47	0.702
DreamLight [31]	0.175	19.87	0.79
Ours	**0.115**	**22.12**	0.82

As illustrated in Fig. 6, the first three columns show the predicted OLAT images generated by our baseline model from a single input portrait. These results demonstrate that our model can faithfully reconstruct directional lighting responses, including smooth shading and realistic specular highlights. The following columns present relit results synthesized from arbitrary environment illuminations using the predicted OLAT basis. Compared with previous relighting approaches, our method better maintains consistent facial appearance, avoids color drift, and exhibits more natural light transitions. These qualitative results verify that learning an OLAT decomposition not only enables faithful recovery of light transport but also supports high-quality relighting of in-the-wild portraits.

5.3. Quantitative Comparisons

We conduct quantitative evaluations to assess the visual fidelity and perceptual realism of our relighting results. Tab. 2 reports the comparison against **SwitchLight** [25], **IC-Light** [64], and **DreamLight** [31]. We use LPIPS↓ for perceptual similarity and naturalness, and PSNR↑ and

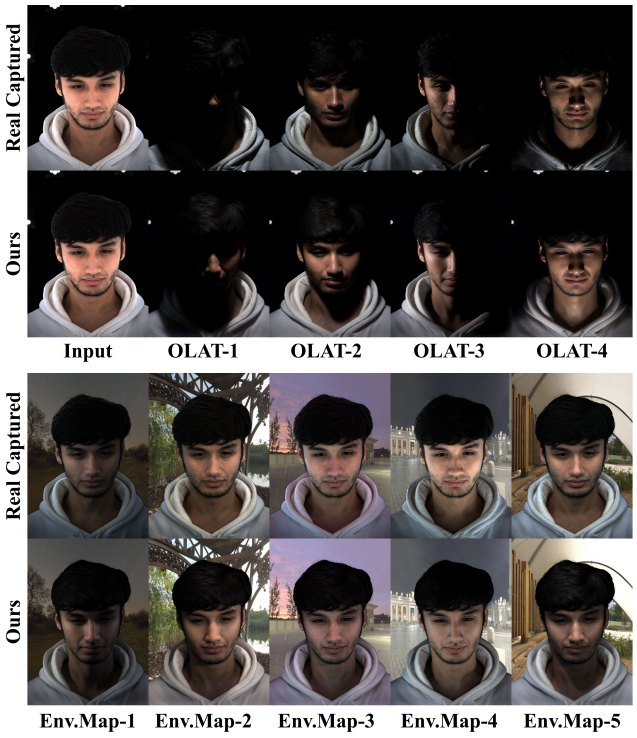


Figure 7. Comparison between generated and captured OLAT data, and their corresponding environment-lit synthesis results. The top part shows OLAT images produced by our model compared with the captured OLAT data. The bottom part presents relit portraits synthesized using the generated OLATs and real OLAT captures under different environment illuminations.

SSIM↑ for pixel-wise reconstruction fidelity. All models are evaluated on the test set of relit portraits. Our method achieves the lowest perceptual error and the highest structural similarity , indicating better preservation of facial details and lighting consistency.

5.4. Ablation Study

Physical consistency under environment rotation. We compare our physical OLAT-based relighting with a background-conditioned lighting model [5] under rotating HDR environments (Fig. 8). As the HDR map rotates, our method maintains physically consistent shading and highlight movement, consistent with the actual change in lighting direction. In contrast, the background-conditioned approach relies primarily on contextual cues from the background, leading to inconsistent lighting and shadow discontinuities when the background rotates.

Energy and uncertainty-aware loss. We compare models trained with and without the proposed energy-aware terms (Fig. 9). Without ω and $\mathcal{L}_{\text{energy}}$, the model tends to overfit low-intensity regions, producing globally darker results and losing contrast in illuminated areas. By introducing per-pixel weighting $\omega(x)$ that emphasizes bright regions and an

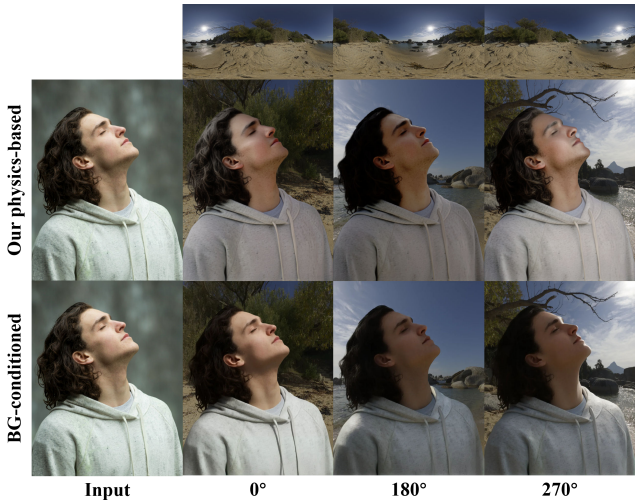


Figure 8. Comparison of relighting consistency under environment rotation. Our Physical OLAT-based relighting preserves consistent shading and highlight movement, while background-conditioned methods show appearance-driven inconsistencies.

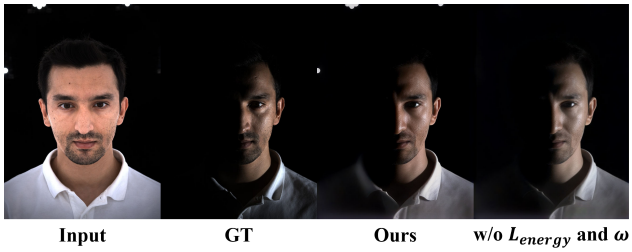


Figure 9. Ablation study on energy and uncertainty-aware loss.

energy consistency constraint $\mathcal{L}_{\text{energy}}$, the network learns to preserve overall exposure and recover high-frequency shading details under strong illumination.

6. Conclusion and Discussion

We have presented **POLAR**, a large-scale and physically calibrated portrait dataset offering diverse OLAT captures across identities, expressions, and viewpoints. Built upon this foundation, our **POLARNet** learns to generate per-light responses from a single portrait, enabling controllable and physically consistent relighting. The synthesized OLATs closely approximate real captures, supporting realistic portrait relighting. This unified dataset–model framework provides a scalable path toward illumination learning and future advances in downstream tasks.

Limitation. POLARNet still has several limitations. The generated OLAT data may lose high-frequency details around specular highlights and shadow boundaries, and performance degrades under extreme facial poses or lighting conditions. We currently address arbitrary lighting using a delighting module, shown in Supplementary PDF. In future work, we plan to extend the framework to video-based OLAT synthesis for temporally consistent illumination.

Acknowledgment

The authors of Shanghai Jiao Tong University were supported in part by the Science and Technology Commission of Shanghai Municipality under Grant 25511103000, and NSFC (62571314).

References

[1] Sai Bi, Zexiang Xu, Pratul Srinivasan, Ben Mildenhall, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. Neural Reflectance Fields for Appearance Acquisition. *ArXiv*, 2020. 3

[2] Sai Bi, Stephen Lombardi, Shunsuke Saito, Tomas Simon, Shih-En Wei, Kevyn Mcphail, Ravi Ramamoorthi, Yaser Sheikh, and Jason Saragih. Deep relightable appearance models for animatable faces. *ACM Transactions on Graphics (ToG)*, 40(4):1–15, 2021. 3

[3] Mark Boss, Raphael Braun, V. Jampani, Jonathan T. Barron, Ce Liu, and Hendrik P. A. Lensch. NeRD: Neural Reflectance Decomposition from Image Collections. In *ICCV*, 2021. 3

[4] Diogo Carbonera Luvizon, Vladislav Golyanik, Adam Kortylewski, Marc Habermann, and Christian Theobalt. Relightable neural actor with intrinsic decomposition and pose control. In *European Conference on Computer Vision*, pages 465–483. Springer, 2024. 3

[5] Clément Chadebec, Onur Tasar, Sanjeev Sreetharan, and Benjamin Aubin. Lbm: Latent bridge matching for fast image-to-image translation. *arXiv preprint arXiv:2503.07535*, 2025. 5, 8

[6] Sumit Chaturvedi, Mengwei Ren, Yannick Hold-Geoffroy, Jingyuan Liu, Julie Dorsey, and Zhixin Shu. Synthlight: Portrait relighting with diffusion model by learning to re-render synthetic faces. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 369–379, 2025. 2, 4, 5

[7] Jianqi Chen, Yilan Zhang, Zhengxia Zou, Keyan Chen, and Zhenwei Shi. Dense pixel-to-pixel harmonization via continuous image representation. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1, 2023. 3

[8] Zhaoxi Chen, Gyeongsik Moon, Kaiwen Guo, Chen Cao, Stanislav Pidhorskyi, Tomas Simon, Rohan Joshi, Yuan Dong, Yichen Xu, Bernardo Pires, et al. Urhand: Universal relightable hands. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 119–129, 2024. 3

[9] Wenyan Cong, Jianfu Zhang, Li Niu, Liu Liu, Zhixin Ling, Weiyuan Li, and Liqing Zhang. Dovenet: Deep image harmonization via domain verification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8394–8403, 2020. 3

[10] Wenyan Cong, Xinhao Tao, Li Niu, Jing Liang, Xuesong Gao, Qihao Sun, and Liqing Zhang. High-resolution image harmonization via collaborative dual transformations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18470–18479, 2022. 3

[11] Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. Acquiring the reflectance field of a human face. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 145–156, 2000. 2, 3

[12] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019. 6

[13] Craig Donner and Henrik Wann Jensen. A spectral bssrdf for shading human skin. *Rendering techniques*, 2006:409–418, 2006. 3

[14] Abhijeet Ghosh, Graham Fyffe, Borom Tunwattanapong, Jay Busch, Xueming Yu, and Paul Debevec. Multiview face capture using polarized spherical gradient illumination. In *SIGGRAPH Asia 2011 Conference Papers*, pages 1–10, 2011. 3

[15] Julian Jorge Andrade Guerreiro, Mitsuru Nakazawa, and Björn Stenger. Pct-net: Full resolution image harmonization using pixel-wise color transformations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5917–5926, 2023. 3

[16] Kaiwen Guo, Peter Lincoln, Philip Davidson, Jay Busch, Xueming Yu, Matt Whalen, Geoff Harvey, Sergio Orts-Escolano, Rohit Pandey, Jason Dourgarian, et al. The relightables: Volumetric performance capture of humans with realistic relighting. *ACM Transactions on Graphics (ToG)*, 38(6):1–19, 2019. 3

[17] Zonghui Guo, Dongsheng Guo, Haiyong Zheng, Zhaorui Gu, Bing Zheng, and Junyu Dong. Image harmonization with transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 14870–14879, 2021. 3

[18] Zonghui Guo, Haiyong Zheng, Yufeng Jiang, Zhaorui Gu, and Bing Zheng. Intrinsic image harmonization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16367–16376, 2021. 3

[19] Jon Hasselgren, Nikolai Hofmann, and Jacob Munkberg. Shape, Light, and Material Decomposition from Images using Monte Carlo Rendering and Denoising. In *NeurIPS*, 2022. 3

[20] Mingming He, Pascal Clausen, Ahmet Levent Taşel, Li Ma, Oliver Pilarski, Wenqi Xian, Laszlo Rikker, Xueming Yu, Ryan Burgert, Ning Yu, et al. Diffrelight: Diffusion-based facial performance relighting. In *SIGGRAPH Asia 2024 Conference Papers*, pages 1–12, 2024. 2, 4, 5

[21] Andrew Hou, Michel Sarkis, Ning Bi, Yiying Tong, and Xiaoming Liu. Face relighting with geometrically consistent shadows. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4217–4226, 2022. 3

[22] Yifan Jiang, He Zhang, Jianming Zhang, Yilin Wang, Zhe Lin, Kalyan Sunkavalli, Simon Chen, Sohrab Amirghodsi, Sarah Kong, and Zhangyang Wang. Ssh: A self-supervised framework for image harmonization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4832–4841, 2021. 3

[23] Haian Jin, Isabella Liu, Peijia Xu, Xiaoshuai Zhang, Songfang Han, Sai Bi, Xiaowei Zhou, Zexiang Xu, and Hao Su. Tensoir: Tensorial inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 165–174, 2023. 3

[24] Haian Jin, Yuan Li, Fujun Luan, Yuanbo Xiangli, Sai Bi, Kai Zhang, Zexiang Xu, Jin Sun, and Noah Snavely. Neural gaffer: Relighting any object via diffusion, 2024. 2, 3

[25] Hoon Kim, Minje Jang, Wonjun Yoon, Jisoo Lee, Donghyun Na, and Sanghyun Woo. Switchlight: Co-design of physics-driven architecture and pre-training framework for human portrait relighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25096–25106, 2024. 3, 4, 7

[26] Theodore Kim, Holly Rushmeier, Julie Dorsey, Derek Nowrouzezahrai, Raqi Syed, Wojciech Jarosz, and AM Darke. Countering racial bias in computer graphics research. In *ACM SIGGRAPH 2022 Talks*, pages 1–2. 2022. 3

[27] Zhengfei Kuang, Kyle Olszewski, Menglei Chai, Zeng Huang, Panos Achlioptas, and Sergey Tulyakov. Neroic: Neural rendering of objects from online image collections. *ACM Transactions on Graphics (TOG)*, 41(4):1–12, 2022. 3

[28] Junxuan Li, Chen Cao, Gabriel Schwartz, Rawal Khirodkar, Christian Richardt, Tomas Simon, Yaser Sheikh, and Shunsuke Saito. Uravatar: Universal relightable gaussian codec avatars. In *SIGGRAPH Asia 2024 Conference Papers*, pages 1–11, 2024. 3

[29] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022. 5

[30] Isabella Liu, Linghao Chen, Ziyang Fu, Liwen Wu, Haian Jin, Zhong Li, Chin Ming Ryan Wong, Yi Xu, Ravi Ramamoorthi, Zexiang Xu, et al. Openillumination: A multi-illumination dataset for inverse rendering evaluation on real objects. *Advances in Neural Information Processing Systems*, 36:36951–36962, 2023. 3

[31] Yong Liu, Wenpeng Xiao, Qianqian Wang, Junlin Chen, Shiyin Wang, Yitong Wang, Xinglong Wu, and Yansong Tang. Dreamlight: Towards harmonious and consistent image relighting. *arXiv preprint arXiv:2506.14549*, 2025. 2, 3, 7

[32] Wan-Chun Ma, Tim Hawkins, Pieter Peers, Charles-Felix Chabert, Malte Weiss, Paul E Debevec, et al. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. *Rendering Techniques*, 9(10):2, 2007. 3

[33] Yiqun Mei, He Zhang, Xuaner Zhang, Jianming Zhang, Zhixin Shu, Yilin Wang, Zijun Wei, Shi Yan, HyunJoon Jung, and Vishal M Patel. Lightpainter: Interactive portrait relighting with freehand scribble. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 195–205, 2023. 2, 3, 4, 5

[34] Yiqun Mei, Yu Zeng, He Zhang, Zhixin Shu, Xuaner Zhang, Sai Bi, Jianming Zhang, HyunJoon Jung, and Vishal M Patel. Holo-relighting: Controllable volumetric portrait relighting from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4263–4273, 2024. 3, 4, 5

[35] Yiqun Mei, Mingming He, Li Ma, Julien Philip, Wenqi Xian, David M George, Xueming Yu, Gabriel Dedic, Ahmet Levent Taşel, Ning Yu, et al. Lux post facto: Learning portrait performance relighting with conditional video diffusion and a hybrid dataset. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 5510–5522, 2025. 2, 4, 5

[36] Jacob Munkberg, Jon Hasselgren, Tianchang Shen, Jun Gao, Wenzheng Chen, Alex Evans, Thomas Müller, and Sanja Fidler. Extracting triangular 3d models, materials, and lighting from images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8280–8290, 2022. 3

[37] Rohit Pandey, Sergio Orts-Escolano, Chloe Legendre, Christian Haene, Sofien Bouaziz, Christoph Rhemann, Paul E Debevec, and Sean Ryan Fanello. Total relighting: learning to relight portraits for background replacement. *ACM Trans. Graph.*, 40(4):43–1, 2021. 2, 3, 4, 5

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

[39] Pramod Rao, Gereon Fox, Abhimitra Meka, Mallikarjun BR, Fangneng Zhan, Tim Weyrich, Bernd Bickel, Hanspeter Pfister, Wojciech Matusik, Mohamed Elgharib, et al. Lite2relight: 3d-aware single image portrait relighting. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–12, 2024. 3

[40] Pramod Rao, Abhimitra Meka, Xilong Zhou, Gereon Fox, Fangneng Zhan, Tim Weyrich, Bernd Bickel, Hanspeter Pfister, Wojciech Matusik, Thabo Beeler, et al. 3dpr: Single image 3d portrait relight using generative priors. *arXiv preprint arXiv:2510.15846*, 2025. 3, 5

[41] Mengwei Ren, Wei Xiong, Jae Shin Yoon, Zhixin Shu, Jianming Zhang, HyunJoon Jung, Guido Gerig, and He Zhang. Relightful harmonization: Lighting-aware portrait background replacement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6452–6462, 2024. 3

[42] Mark Sagar. Reflectance field rendering of human faces for" spider-man 2". In *ACM SIGGRAPH 2005 Courses*, pages 14–es. 2005. 3

[43] Shunsuke Saito, Gabriel Schwartz, Tomas Simon, Junxuan Li, and Giljoo Nam. Relightable gaussian codec avatars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 2, 3, 5

[44] Kripasindhu Sarkar, Marcel C Bühler, Gengyan Li, Daoye Wang, Delio Vicini, Jérémy Riviere, Yinda Zhang, Sergio Orts-Escolano, Paulo Gotardo, Thabo Beeler, et al. Litnerf: Intrinsic radiance decomposition for high-quality view synthesis and relighting of faces. In *SIGGRAPH Asia 2023 Conference Papers*, pages 1–11, 2023.

[45] Jonathan Schmidt, Simon Giebenhain, and Matthias Niessner. Becominglit: Relightable gaussian avatars with hybrid neural shading. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. 3, 5

[46] Soumyadip Sengupta, Brian Curless, Ira Kemelmacher-Shlizerman, and Steven M Seitz. A light stage on every desk.

In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2420–2429, 2021. 3

[47] Pratul P Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T Barron. Nerv: Neural reflectance and visibility fields for relighting and view synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7495–7504, 2021. 3

[48] Giota Stratou, Abhijeet Ghosh, Paul Debevec, and Louis-Philippe Morency. Effect of illumination on automatic expression recognition: a novel 3d relightable facial database. In *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, pages 611–618. IEEE, 2011. 2, 3, 4, 5

[49] Tiancheng Sun, Jonathan T Barron, Yun-Ta Tsai, Zexiang Xu, Xueming Yu, Graham Fyffe, Christoph Rhemann, Jay Busch, Paul E Debevec, and Ravi Ramamoorthi. Single image portrait relighting. *ACM Trans. Graph.*, 38(4):79–1, 2019. 3

[50] Tiancheng Sun, Kai-En Lin, Sai Bi, Zexiang Xu, and Ravi Ramamoorthi. Nelf: Neural light-transport field for portrait view synthesis and relighting. *arXiv preprint arXiv:2107.12351*, 2021. 3

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

[52] Timo Teufel, Pulkit Gera, Xilong Zhou, Umar Iqbal, Pramod Rao, Jan Kautz, Vladislav Golyanik, and Christian Theobalt. Humanolat: A large-scale dataset for full-body human relighting and novel-view synthesis. *arXiv preprint arXiv:2508.09137*, 2025. 3

[53] Ayush Tewari, Tae-Hyun Oh, Tim Weyrich, Bernd Bickel, Hans-Peter Seidel, Hanspeter Pfister, Wojciech Matusik, Mohamed Elgharib, Christian Theobalt, et al. Monocular reconstruction of neural face reflectance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4791–4800, 2021. 3

[54] Marco Toschi, Riccardo De Matteo, Riccardo Spezialetti, Daniele De Gregorio, Luigi Di Stefano, and Samuele Salti. Relight my nerf: A dataset for novel view synthesis and relighting of real world objects. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 20762–20772, 2023. 3

[55] Yi-Hsuan Tsai, Xiaohui Shen, Zhe Lin, Kalyan Sunkavalli, Xin Lu, and Ming-Hsuan Yang. Deep image harmonization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3789–3797, 2017. 3

[56] Ke Wang, Michaël Gharbi, He Zhang, Zhihao Xia, and Eli Shechtman. Semi-supervised parametric real-world image harmonization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5927–5936, 2023.

[57] Zhibo Wang, Xin Yu, Ming Lu, Quan Wang, Chen Qian, and Feng Xu. Single image portrait relighting via explicit multiple reflectance channel modeling. *ACM Transactions on Graphics (ToG)*, 39(6):1–13, 2020. 3

[58] Xiaoyan Xing, Konrad Groh, Sezer Karaoglu, Theo Gevers, and Anand Bhattad. Luminet: Latent intrinsics meets diffusion models for indoor scene relighting. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 442–452, 2025. 3

[59] Jingfeng Yao, Xinggang Wang, Lang Ye, and Wenyu Liu. Matte anything: Interactive natural image matting with segment anything model. *Image and Vision Computing*, 147: 105067, 2024. 4

[60] Yu-Ying Yeh, Koki Nagano, Sameh Khamis, Jan Kautz, Ming-Yu Liu, and Ting-Chun Wang. Learning to relight portrait images via a virtual light stage and synthetic-to-real adaptation. *ACM Transactions on Graphics (TOG)*, 41(6): 1–21, 2022. 3

[61] Chong Zeng, Yue Dong, Pieter Peers, Youkang Kong, Hongzhi Wu, and Xin Tong. Dilightnet: Fine-grained lighting control for diffusion-based image generation. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–12, 2024. 3

[62] Dongbin Zhang, Yunfei Liu, Lijian Lin, Ye Zhu, Kangjie Chen, Minghan Qin, Yu Li, and Haoqian Wang. Hravatar: High-quality and relightable gaussian head avatar. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 26285–26296, 2025. 3

[63] Longwen Zhang, Qixuan Zhang, Minye Wu, Jingyi Yu, and Lan Xu. Neural video portrait relighting in real-time via consistency modeling. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 802–812, 2021. 2, 3, 5

[64] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 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*, 2025. 2, 3, 7

[65] Xuaner Zhang, Jonathan T Barron, Yun-Ta Tsai, Rohit Pandey, Xiuming Zhang, Ren Ng, and David E Jacobs. Portrait shadow manipulation. *ACM Transactions on Graphics (TOG)*, 39(4):78–1, 2020. 3

[66] Taotao Zhou, Kai He, Di Wu, Teng Xu, Qixuan Zhang, Kuixiang Shao, Wenzheng Chen, Lan Xu, and Jingyi Yu. Relightable neural human assets from multi-view gradient illuminations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4315–4327, 2023. 3

[67] Xilong Zhou, Jianchun Chen, Pramod Rao, Timo Teufel, Linjie Lyu, Tigran Minasian, Oleksandr Sotnychenko, Xiaoxiao Long, Marc Habermann, and Christian Theobalt. Olatverse: A large-scale real-world object dataset with precise lighting control. *arXiv preprint arXiv:2511.02483*, 2025. 3

[68] Jun-Yan Zhu, Philipp Krahenbuhl, Eli Shechtman, and Alexei A Efros. Learning a discriminative model for the perception of realism in composite images. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3943–3951, 2015. 3