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

Supplementary Material

Figure A. Light stage. *POLAR* is captured using a calibrated light-stage setup that provides controlled illumination and high-fidelity appearance capture.

Figure B. Viewpoints. *POLAR* provides 32 synchronized camera views offering diverse visual perspectives.

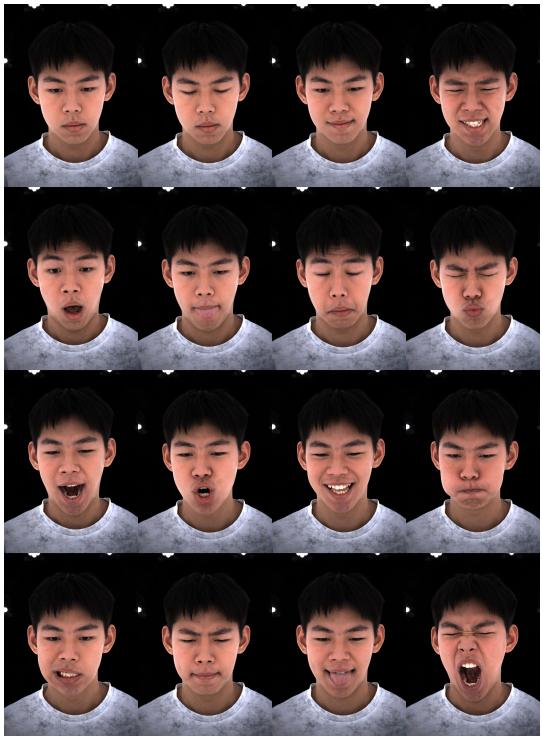

Figure C. Expressions. *POLAR* includes 16 distinct facial expressions capturing a wide range of appearance variations.

A. The *POLAR* Dataset

A.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. A. Each LED is activated sequentially to record One-Light-at-a-Time (OLAT) captures. The lights are photometrically calibrated to ensure consistent intensity and color temperature. Each unit has a radiation angle of 30°, with illumination designed such that the effective range coincides with the spherical light field radius of the stage. Beyond this range, light intensity drops sharply, ensuring minimal crosstalk between neighboring directions. The geometric positions of all light sources are registered in a global spherical coordinate system (θ, ϕ), enabling precise annotation of incident illumination. The 156 directions provide dense and nearly uniform sampling of the frontal hemisphere (Fig. A). Each camera is equipped with a **35 mm fixed-focal-length lens**, which minimizes geometric distortion while preserving facial proportions. We minimized subjects' motion via a fast 2.8s acquisition, a rigid headrest, and breath-holding. We also insert fully-lit anchor frames every 20 frames for drift monitoring. Any sequence exceeding the threshold is immediately re-captured. These ensure inherent alignment, as we found that further optical flow alignment yielded only marginal gains while risking unnecessary interpolation artifacts. We demonstrate portrait images from 32 different

Figure D. Synthetic relit portraits in our POLAR dataset.

viewpoints, as shown in Fig. B. Images are recorded at **4K resolution** in linear color space with 16-bit precision, retaining both fine-scale details and high dynamic range. Multi-view coverage ensures that relighting can be studied not only for frontal images but also across a range of viewpoints, supporting applications in 3D reconstruction and view-consistent relighting.

A.2. Facial Expression

The real OLAT subset includes 16 controlled facial expressions spanning neutral, mild, moderate, and extreme deformations, as demonstrated in Fig. C. These expressions cover a wide range of anatomically meaningful configurations, including: neutral and relaxed states; eye-closed and gaze-down variants; lip-compression and cheek-inflation motions; symmetric and asymmetric mouth deformations (closed-mouth smile, open-mouth smile, wide mouth-open, extreme yawning); as well as brow-raising and frowning behaviors. This curated expression set ensures comprehensive coverage of both subtle muscle activations and large-amplitude shape changes.

A.3. HDR-relit Examples

To further illustrate the diversity and coverage of our dataset, Fig. D presents a representative subset of HDR-relit portraits generated using a wide variety of high-dynamic-range environment maps. The examples span subjects of different skin tones and ethnicities, including Black, White, and Asian individuals, and cover multiple viewing angles as well as a range of facial expressions. Across these variations, the relit results exhibit coherent shading behavior, realistic highlight placement, and consistent shadow geometry under complex outdoor and indoor illumination. These HDR-relit samples complement the real OLAT captures by exposing each subject to rich, spatially varying light fields that cannot be reproduced with point-light acquisition alone. For external HDRs, we provide download scripts and directory listings to ensure reproducibility. Through this process, every subject expands from OLAT captures to thousands of HDR-lit portraits under diverse and physically consistent lighting, providing rich supervision for downstream tasks.

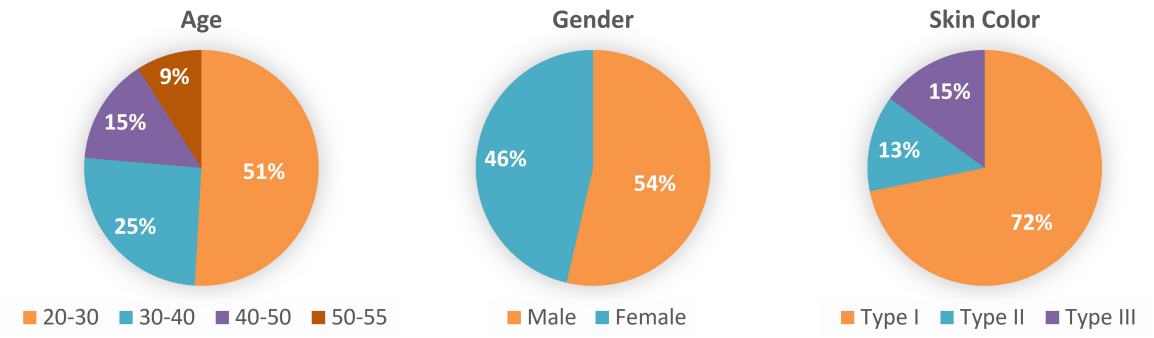

Figure E. Dataset Summary. We summarize age, gender, skin color in our dataset.

A.4. Dataset Statistics

To better characterize the demographic distribution of our OLAT dataset, we provide an analysis across age, gender, and skin-type attributes, as shown in Fig. E. The dataset primarily consists of young to middle-aged adults, with the majority falling within the 20–30 age range (51%), followed by 30–40 (25%), 40–50 (15%), and 50–55 (9%).

Gender distribution is relatively balanced, with 54% female and 46% male participants. Such balance helps mitigate gender-related bias and ensures more stable generalization when learning appearance or reflectance priors.

For skin type, the dataset predominantly contains subjects of Type I–III participants. Although the distribution is naturally skewed toward the primary demographic region where data collection was conducted, the inclusion of multiple skin tones improves the dataset's applicability to cross-ethnicity appearance modeling.

A.5. Supplementary Data Processing Pipeline

Detailed foreground matting.

Accurate foreground extraction is a crucial step in our processing pipeline, as it directly determines the realism of synthesized relit images. In particular, hair strands and semi-transparent boundaries are extremely sensitive to segmentation quality: hard binary masks often produce halo artifacts or clipped silhouettes when composited under new illumination, while high-quality alpha mattes preserve fine details and lead to significantly more natural relighting results.

We adopt the *Matte-Anything* framework for alpha matting. Since interactive scribbles or clicks are impractical for our large-scale batch processing, we design an automatic initialization strategy: facial keypoints are detected to provide foreground seeds, and a text prompt (`"person"`) is supplied to guide the model toward the subject region.

As all 156 OLAT images plus the uniform-light portrait are captured with the subject in a fixed pose, we only need to generate a single high-quality matte per subject, which can be reused across lighting conditions. However, directly using the captured uniform-light images is problematic: in some viewpoints, bright light sources appear near the face, leading to segmentation failures. To address this, we instead generate uniform-light images synthetically by averaging OLAT responses, ensuring a consistent background where stage lights do not appear as artifacts.

Another challenge arises from the optical setup: because illumination intensity decays sharply beyond the light field radius, the background remains very dark. In this setting, dark regions of hair or clothing often merge with the background, making separation difficult. To mitigate this, we apply strong gamma correction to the uniform-light composites before matting, which enhances contrast and improves foreground-background discrimination in low-light regions.

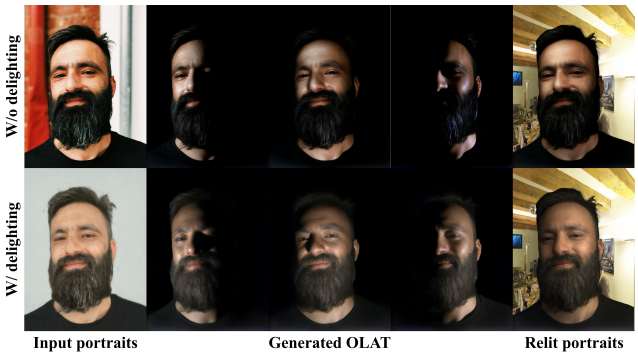

Figure F. Our failure case and delighting solution. Without delighting, the predicted OLAT images entangle the illumination present in the input, leading to biased shading and inconsistent light responses. After applying the delighting module, the input is normalized to an illumination-neutral appearance, enabling the model to generate more accurate OLAT predictions.

B. Additional Ablation Study

B.1. Delighting Module

A typical failure case occurs when the input portrait contains strong non-uniform illumination, for example when

Table 1. Quantitative ablation study.

Method	PSNR ↑	SSIM ↑	LPIPS ↓	Infer. time ↓
W/o L_{energy}	21.05	0.80	0.283	-
W/o delighting	20.83	0.80	0.263	-
Ours (diffusion)	18.92	0.76	0.278	9s
Ours (LBM)	**22.12**	**0.82**	**0.115**	**0.3s**

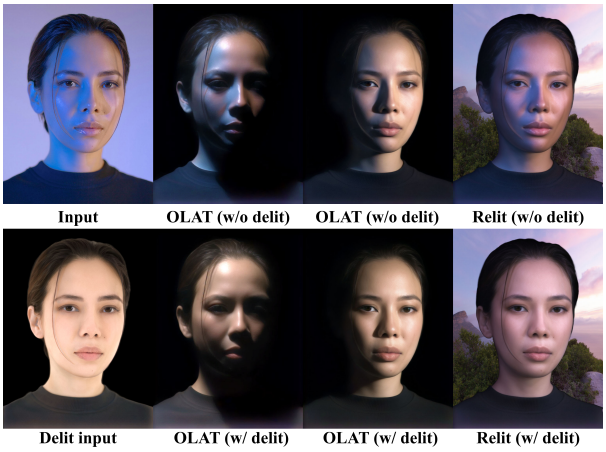

Figure G. Delighting module to solve arbitrary inputs.

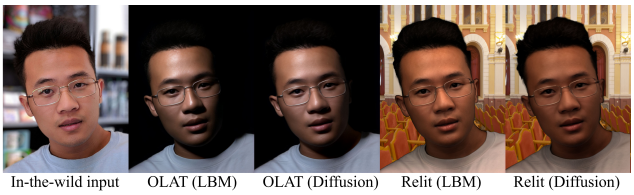

Figure H. **Ablation study of LBM vs. Diffusion.** While the diffusion-based approach often results in over-smoothed textures and muted lighting, our LBM-based framework effectively preserves fine-grained facial details and exhibits better physical consistency in both OLAT and complex relighting scenarios, despite being trained on the same dataset.

one side of the face is significantly brighter than the other. Since POLARNet assumes a uniformly lit input image, such uneven lighting can be partially preserved in the latent representation and may propagate to the generated OLAT outputs,as shown in Fig. F. As a result, the predicted OLAT set may exhibit an unintended global shading bias that affects the quality of the synthesized relit portraits.

To address this issue, we introduce a delighting module that attempts to restore an approximately uniform-light appearance before OLAT prediction. By providing a more illumination-neutral input, the resulting OLAT estimates become more consistent across directions and the relighting quality improves accordingly. Tab. 1 and Fig. G show that this module restores OLAT and relit image quality from strong directional lights and shadows.

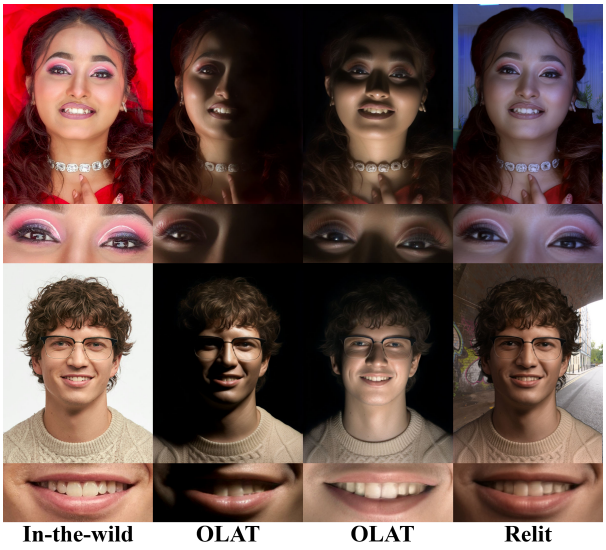

Figure I. More qualitative cases on accessories.

B.2. Diffusion *vs.* LBM

As quantified in Tab. 1 and qualitatively compared in Fig. H, our LBM-based approach demonstrates superior performance over the diffusion-based alternative. Diffusion models often suffer from over-smoothing artifacts and inconsistent lighting, as evidenced by the blurred shadow boundaries. In contrast, our LBM-based method produces sharper, more physically-grounded relighting results with enhanced facial details and realistic light-material interactions. Crucially, our LBM achieves these high-fidelity results while being approximately 30× faster than the diffusion counterpart (0.3s vs. 9s), providing a more practical solution for real-time applications without compromising visual quality.

B.3. Energy Loss

To evaluate the contribution of the L_{energy}, we conduct an ablation study by removing it from our full objective. As shown in Tab. 1, the exclusion of L_{energy} leads to a noticeable performance degradation, where the PSNR drops from 22.12 dB to 21.05 dB and the LPIPS increases significantly to 0.283. These results demonstrate that L_{energy} plays a critical role in regularizing the learning process and ensuring high-fidelity reconstruction with better perceptual quality.

B.4. More Qualitative Results

In our data capturing setting, subjects are requested to remove glasses, heavy make-up, and other accessories to minimize the interference of uncontrolled specular reflections. However, as illustrated in Fig. I, our model exhibits robust generalization capabilities even when confronted with these challenging cases. Despite the lack of such diverse attributes in the training constraints, our model maintains

Figure J. Real captured OLAT sequence in our POLAR dataset (selected 48 frontal LEDs).

high fidelity and physical consistency.

C. OLAT Visualization

C.1. Real Data and POLARNet Predictions

To further demonstrate the quality of the POLAR dataset and the effectiveness of our proposed POLARNet for single-image OLAT generation, we provide a visual comparison between ground-truth OLAT captures and the corresponding OLAT predictions produced by our model. Specifically, we select 48 representative light directions that uniformly sample the frontal hemisphere and display. Fig. J presents the ground-truth OLAT samples and Fig. K shows POLARNet predictions.

Figure K. Generated OLAT sequence of test set by our POLARNet (selected 48 frontal LEDs).

Across all directions, POLARNet produces directionally accurate and physically meaningful illumination responses, including the movement of highlights, shading variations, and overall light–geometry interaction. The predicted OLAT images exhibit plausible specular and diffuse behavior as well as consistent energy falloff, while maintaining stable identity and global facial structure. Although some fine-scale details appear slightly smoothed, the model reliably captures the dominant lighting characteristics and avoids common artifacts such as deformation, overexposure, or inconsistent photometric shifts.

In addition to the studio examples, we also provide OLAT sequences generated from *in-the-wild* portrait images in Fig. L. It demonstrate that POLARNet generalizes

Figure L. Generated OLAT sequence of in-the-wild portraits by our POLARNet (selected 48 frontal LEDs).

beyond Light Stage conditions and is able to produce directionally consistent illumination responses on real, unconstrained images. The model preserves identity and global facial structure while delivering plausible per-light shading and highlight behaviors.

Using only a single uniform-light input, POLARNet can generate a complete set of directionally aligned, physically meaningful OLAT responses without multi-light input or multi-step diffusion sampling. These results highlight the practicality and effectiveness of our approach for controlled relighting tasks and for enabling physically interpretable illumination modeling.

C.2. Relighting Consistency

To further validate the physical consistency of our OLAT-based relighting pipeline, we use the generated OLAT from

Figure M. Relighting results under a rotating HDR map.

Fig. L to synthesize relit results under a set of rotating outdoor environment maps, as shown in Fig. M. As the environment map rotates, the synthesized portraits exhibit coherent and physically meaningful illumination changes. Highlights shift smoothly across facial regions, cast shadows reposition according to the dominant light direction, and the global shading pattern evolves consistently with the movement of the sun in the environment map. Importantly, POLARNet maintains stable facial identity.

D. Implementation Details

Training setup. We train our POLARNet on a single NVIDIA A100 (40GB) GPU using mixed-precision (FP16). Training uses a batch size of 4, an input resolution of 1024×768, and the AdamW optimizer with a learning rate of 5×10^{-5}. The model is trained for 40k steps, which takes approximately 8 hours in total. Our training set contains 154 subjects, covering diverse identities, facial geometry, and illumination conditions.

Inference performance. POLARNet performs single-step OLAT prediction. Given one uniform-light portrait, each OLAT image is generated in 0.35s, and producing the full set of 157 directions requires about 54s on a single A100 GPU. This efficiency comes from our flow-based latent transport formulation, which avoids iterative denoising or multi-step sampling used in diffusion models. The fast inference allows POLARNet to support interactive relighting and large-scale OLAT synthesis.