

PolarFree: Polarization-based Reflection-Free Imaging

Mingde Yao¹, Menglu Wang³, King-Man Tam^{1,4}, Lingen Li¹, Tianfan Xue^{✉1,2}, Jinwei Gu¹

¹The Chinese University of Hong Kong, ²Shanghai AI Laboratory

³University of Science and Technology of China, ⁴Institute of Science Tokyo

mingdeyao@foxmail.com, tfxue@ie.cuhk.edu.hk

Abstract

Reflection removal is challenging due to complex light interactions, where reflections obscure important details and hinder scene understanding. Polarization naturally provides a powerful cue to distinguish between reflected and transmitted light, enabling more accurate reflection removal. However, existing methods often rely on small-scale or synthetic datasets, which fail to capture the diversity and complexity of real-world scenarios. To this end, we construct a large-scale dataset, **PolaRGB**, for Polarization-based reflection removal of RGB images, which enables us to train models that generalize effectively across a wide range of real-world scenarios. The PolaRGB dataset contains 6,500 well-aligned mixed-transmission image pairs, $8\times$ larger than existing polarization datasets, and is the first to include both RGB and polarization images captured across diverse indoor and outdoor environments with varying lighting conditions. Besides, to fully exploit the potential of polarization cues for reflection removal, we introduce **PolarFree**, which leverages diffusion process to generate reflection-free cues for accurate reflection removal. Extensive experiments show that PolarFree significantly enhances image clarity in challenging reflective scenarios, setting a new benchmark for polarized imaging and reflection removal. Code and dataset are available at <https://github.com/mdyao/PolarFree>.

1. Introduction

Reflection removal algorithms [1, 19, 30, 34, 38, 41] remove unwanted reflections in captured images, playing a critical role in applications such as autonomous driving [12] and photography [11, 43]. This problem commonly arises when imaging through semi-reflectors, like windows or glass, and overlapping reflections may obscure important details of scenes we want to capture. This problem is often formulated [10, 11] as a linear combination of the transmission layer T and the reflection layer R :

$$M = \alpha_t T + \alpha_r R, \quad (1)$$

✉Corresponding author.

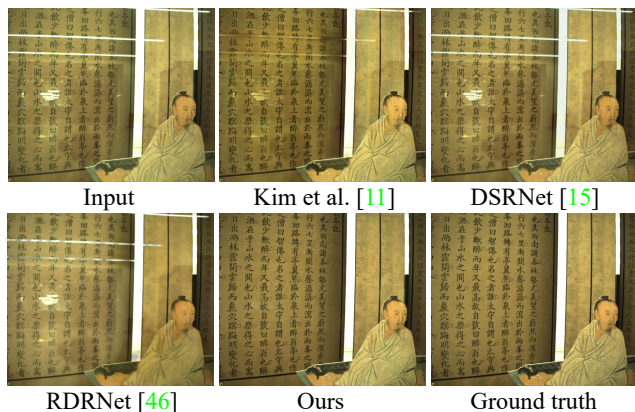


Figure 1. Our PolarFree effectively leverages polarization information to remove reflections, achieving superior performance in challenging scenes with complex backgrounds and highlights where previous methods [11, 15, 46] often fail.

where M is the mixed captured image, and α_t, α_r are blending coefficients resulting from light attenuation.

Polarization image sensors are becoming mainstream, allowing users to easily capture polarization images from a single shot in real-time [32]. However, existing methods [10, 11, 24, 46] typically rely on intensity-based cues, such as pixel brightness and color gradients, to distinguish transmitted and reflected layers. These methods face challenges because reflection removal is a highly ill-posed inverse problem [21] that recovers two unknown layers (reflection and transmission) from a single observation. Polarization provides valuable physics-based cues [27, 29] to alleviate ill-posed issues (Fig. 1), whereas transmitted light retains a distinct degree of polarization (Fig. 2b). This difference provides crucial signals for separating the two layers. Notably, at the Brewster angle (Fig. 2c), reflected light is fully polarized, enabling effective reflection removal [2].

Despite significant advances of polarization images, a key challenge in polarization-based reflection removal is the lack of large-scale, high-quality datasets. Existing polarized reflection removal datasets [18, 39] are limited in size and diversity, relying on small (<1000) or synthetic samples that fail to capture the complexity of real-world light-

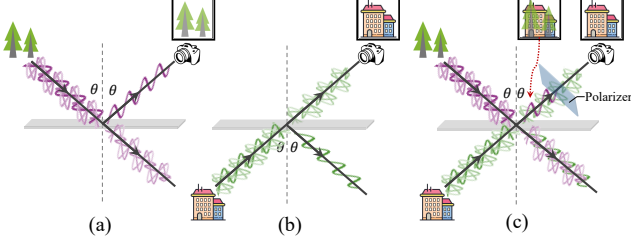


Figure 2. (a) & (b) A semi-reflector transforms unpolarized light into polarized light upon reflection and refraction, which is undetectable by standard RGB cameras but can be leveraged by polarization cameras for reflection-suppression tasks. (c) At the Brewster angle [2], a polarizer minimizes reflections.

ing conditions, materials, and scenes. Moreover, they typically exclude color information, reducing their applicability in real-world reflection removal tasks. Thus, there is a pressing need for a large-scale, comprehensive dataset that includes both RGB and polarization images, captured in diverse real-world environments, to advance polarization-based reflection removal.

To bridge this gap, we introduce PolaRGB, a novel dataset specifically collected for polarization-based reflection removal. As shown in Table 1, PolaRGB contains 6,500 high-quality, well-aligned RGB-polarization image pairs, $8\times$ larger than previous dataset [18]. Our dataset covers a diverse range of scenes, lighting conditions, and exposure settings, and is captured using off-the-shelf commercial cameras with polarized color patterns to ensure real-world applicability. PolaRGB provides both mixed images and ground-truth transmission layers, enabling accurate reflection separation and significantly enhancing the effectiveness of reflection removal across real-world scenarios.

Moreover, extracting reflection-free information from polarization data is challenging [18, 25, 39] due to the randomness of shooting angles, scene variations, and changing lighting conditions. To address this issue, we leverage the powerful generative capabilities of the diffusion model [3, 6, 8] to generate reflection-free cues. The diffusion model extracts and refines reflection-free priors from polarization images, effectively guiding the reflection removal and yielding precise and robust reflection-free results.

Specifically, PolarFree consists of two steps: a prior-generation step and a reflection removal step. First, in the prior-generation step, we use a diffusion model to generate reflection-free prior based on the polarization and RGB inputs. This strategy not only guides accurate reflection isolation in RGB images but also recovers background details, which previous methods [11, 15, 46] may miss. Next, the reflection removal step leverages the prior to effectively remove reflections, ensuring accurate transmission restoration. Additionally, we introduce a phase-based loss function in the frequency domain to mitigate color discrepancies caused by semi-reflections, guiding the network to focus

Table 1. Comparisons between PolaRGB and existing reflection removal datasets.

Dataset	Polarization	RGB	RAW	Data size	Resolution
SIR ² [36]	No	Yes	No	500	540×400
RRW [46]	No	Yes	No	14,952	2580×1460
ReflectNet [39]	Yes (Syn.)	Yes	No	-	1500×1000
Lei <i>et al.</i> [18]	Yes (Real)	No	Yes	807	1224×1024
PolaRGB (Ours)	Yes (Real)	Yes	Yes	6,500	1224×1024

on reflection removal rather than color adjustment. These components enable PolarFree to achieve robust reflection suppression while preserving the clarity and integrity of the transmission across diverse real-world scenes.

Extensive experiments on the PolaRGB dataset demonstrate the effectiveness of PolarFree, which outperforms existing methods [11, 15, 46] ~ 2 dB in terms of PSNR. Additionally, we performed real-world testing in environments commonly affected by reflections such as museums and galleries, demonstrating the method’s effectiveness in real-world scenarios. Our approach improves image clarity and preserves details better than previous methods, paving the way for more practical, reflection-robust applications.

2. Related Work

2.1. Reflection Removal

Reflection removal [1, 30, 34] has a longstanding history in computer vision, aiming to enhance the occluded objects caused by semi-reflectors like glass and windows. Early approaches [20, 24] typically relied on handcrafted priors, such as smoothness assumptions and gradient sparsity, to differentiate the reflection layer from the underlying scene. Some methods incorporated ghosting [31], flash cues [17, 19] and edge consistency [7], while others improved accuracy by leveraging multi-image setups [22, 28], utilizing temporal variation [26] and parallax [28] to separate reflections from background layers. However, these methods often require strict assumptions and fail in real-world scenes with complex textures or lighting [19, 38].

With the rise of deep learning [5, 35, 42, 44, 46], reflection removal methods advanced considerably. Techniques based on neural networks have shown promise by learning to separate reflection and transmission layers directly from data [10, 15, 35–37, 44]. However, without essential physical insights, these methods act as Bayesian regressions toward a learned average, often lead to suboptimal separation in complex environments with varying lighting, textures, and reflection intensities [35, 44]. Recent methods [9, 45] utilize language to interactively remove the reflection, which require additional efforts, limiting their practicality.

2.2. Polarization Reflection Removal

Polarization is an effective way to remove reflections due to the physics principle: the polarization of light behaves differently in the reflection and transmission layers [4, 13, 23, 27, 30], allowing separation, especially at the Brewster angle [4, 27]. However, capturing images exactly

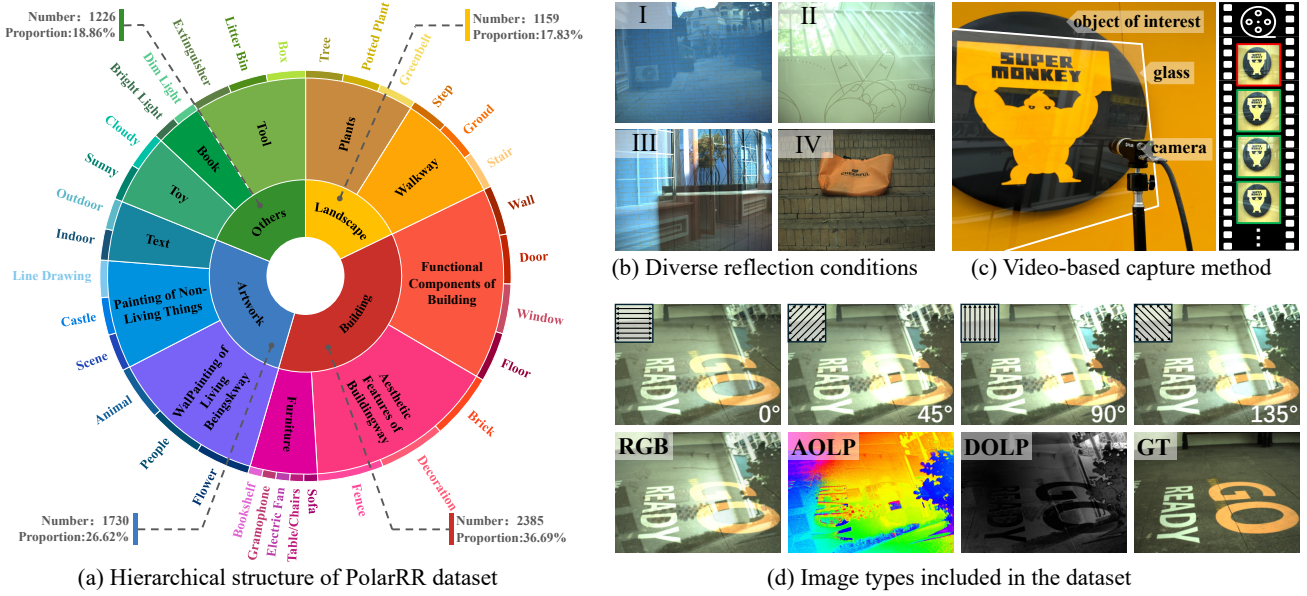


Figure 3. Overview of the PolarRR dataset. (a) Hierarchical structure of scenes is shown in the ring, with legends indicating sample counts and subset types. (b) Typical scenes illustrating varied reflection conditions: I. smooth blending of reflection and refraction, II. abrupt reflection with mixed components, III. reflection dominant over transmission, and IV. minimal or no reflection. (c) Video-based capture method (details in Sec. 3). (d) We provide polarized images at angles $\phi = 0^\circ, 45^\circ, 90^\circ$, and 135° , along with derived AoLP, DoLP, and a well-aligned unpolarized image. The dataset also includes ground truth transmission and estimated reflections, all available in both raw and RGB formats.

at the Brewster angle, as shown in Fig. 2c, is challenging. Therefore, in practice, methods typically rely on images taken at multiple polarization angles (e.g., $0^\circ, 45^\circ, 90^\circ$, and 135°) to capture polarization information [4, 27, 30].

Early polarized reflection removal methods [4, 27, 30] employ mathematical techniques like Principle Component Analysis (PCA) to separate reflection layers, which are effective when the transmission and reflection layers have significant content differences. Subsequently, Kong *et al.* [16] propose a multiscale scheme that automatically identifies the optimal separation of the reflection and background layers. Another advancements [39] introduce a neural network-based approach for polarization-guided reflection removal with synthetic dataset. Lyu *et al.* [25] utilize a pair of unpolarized and polarized images but lack of realistic polarization patterns. Lei *et al.* [18] further contribute by collecting a dataset of polarized images for reflection removal. However, this dataset only includes limited scene variations from pure polarization scenes, which limits its application scenario. In contrast, we present the first large-scale dataset captured for polarization-based RGB reflection removal, consisting of 6,500 well-aligned transmission-reflection pairs. We also propose a polarization-based reflection removal network based on diffusion models.

3. PolarRGB Dataset

3.1. Analysis

Although a few polarization-based reflection removal datasets exist [18, 39], they face major limitations: (1) they

are often synthetically generated [25, 39], which may not generalize well to real polarization, as perfectly simulating polarization phenomena is very hard; (2) they are typically limited in size and diversity ($<1,000$ samples), with limited scene variations, which constrains model robustness and generalizability [18, 39]; and (3) they lack RGB data [18, 39], reducing the practical value of these datasets.

Our polarized-based reflection removal dataset, PolarRGB, addresses all these limitations with the following key features, as shown in Fig. 3. It has following advantages: (1) It contains real-captured images with perfect spatial alignment. We achieve pixel-level alignment through a three-step process: careful capture setup, manual filtering, and homography transformations, all in raw space to avoid demosaicing artifacts (see Fig. 4b and more details in supplementary). (2) It contains large-scale diverse scenes. The collected dataset covers a wide range of indoor and outdoor settings with four distinct reflection types, including smooth, sharp, high-brightness, and subtle reflections (see Fig. 3b). (3) It has extensive data modalities. Our dataset offers paired mixed and transmission images with both polarization and RGB captures, precisely aligned and available in both raw and RGB formats (see Fig. 3d).

3.2. Processing Pipeline

We leverage an efficient video-based capture flow for dataset collection, as shown in Fig. 3c. First, we capture a transmission-only image T_{raw} as the ground truth. Then, we place a semi-reflective glass plate in front of the scene and rotate it continuously to capture mixed images

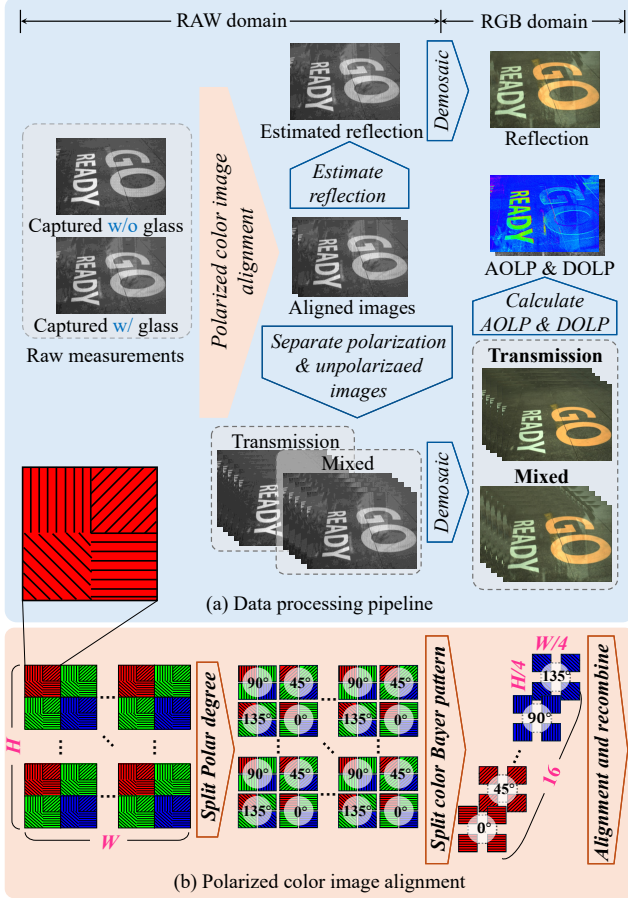


Figure 4. Data processing pipeline for obtaining aligned mixed and transmission images, and polarized images.

M_{raw} . A division-of-focal-plane polarization camera with color Bayer pattern is used to simultaneously capture both polarized and RGB information, as shown in Fig. 4.

After obtaining the raw mixed image M_{raw} and the transmission image T_{raw} , we process the images sequentially in the raw and RGB domains as shown in Fig. 4a. In the raw domain, we first align M_{raw} and T_{raw} to correct spatial misalignment caused by light refraction. This is done by separating the images into different polarization angles and color channels, then applying affine transformation matrices to each of the channels to avoid aliasing from direct alignment, as shown in Fig. 4b. Next, we perform polarization separation on the aligned images to obtain four polarization images (0° , 45° , 90° , and 135°), and use Malus’s law [27, 29] to sum them, producing the unpolarized image of the scene. Finally, we estimate the reflection image by searching for the optimal blending coefficient α_r and α_t in edge space, as illustrated in Eq. 1. For detailed steps and proof, please refer to the supplementary material.

Next, we apply demosaicking to the processed raw images to obtain the mixed, transmission, and reflection images, each consisting of four polarization images and the unpolarized RGB image. Additionally, we compute the

Stokes parameters (see Sec. 4.1) to estimate the Angle of Linear Polarization (AOLP) and Degree of Linear Polarization (DOLP), which are then used in reflection removal.

4. Method

4.1. Problem Formulation

Object. Given an RGB image and its corresponding polarization images, we aim to recover the transmission layer of the RGB image by utilizing the distinct polarization characteristics. Our measurement consists of spatially-aligned RGB and polarization images, where the polarization images captured at four distinct angles (0° , 45° , 90° , and 135°). These four angles provide a comprehensive polarization measurement [30], which allows us to compute polarization features crucial for separating layers.

Polarization Preliminary. For semi-reflective surfaces that produce mixed images, the observed intensity $I_M^\phi(x)$ can be decomposed into reflected intensity $I_R^\phi(x)$ and transmitted intensity $I_T^\phi(x)$ as

$$I_M^\phi(x) = \alpha_R I_R^\phi(x) + \alpha_T I_T^\phi(x). \quad (2)$$

If a linear polarizer is placed in front of the camera at an angle ϕ , the captured intensity $I^\phi(x)$ can be represented as

$$I_M^\phi(x) = \alpha(\theta; \phi; \phi_\perp) I_R^\phi(x) + (1 - \alpha(\theta; \phi; \phi_\parallel)) I_T^\phi(x), \quad (3)$$

where $\alpha(\theta; \phi; \phi_\perp)$ and $(1 - \alpha(\theta; \phi; \phi_\parallel))$ are polarization-dependent mixing coefficients based on the angle of incidence θ , ϕ_\perp and ϕ_\parallel are the canonical polarization directions for the reflection and transmission, respectively. Directly solving for these parameters is challenging because $\alpha(\cdot)$, $I_R^\phi(x)$, and $I_T^\phi(x)$ are all unknown [27, 39]. Additionally, relationship between the observed intensity and the underlying reflected and transmitted components is highly non-linear and influenced by various factors [39]. To address this, we utilize Stokes parameters [18] that provide a efficient way to represent and analyze polarized light, enabling more robust separation of reflection and transmission.

Stokes Parameters. To capture the polarization effects in the scene, we use Stokes parameters $[S_0, S_1, S_2]$, which can be derived from intensity measurements at specific polarization angles (0° , 45° , 90° , and 135°):

$$\begin{aligned} S_0 &= (I_{0^\circ} + I_{45^\circ} + I_{90^\circ} + I_{135^\circ})/2, \\ S_1 &= I_{0^\circ} - I_{90^\circ}, S_2 = I_{45^\circ} - I_{135^\circ}. \end{aligned} \quad (4)$$

Here, S_0 represents the total intensity of light, and S_1 and S_2 provide information about the linear polarization state of the light based on intensity differences at these key angles.

Using the Stokes parameters, we compute the Degree of Linear Polarization (DOLP) and Angle of Linear Polarization (AOLP) as follows [2]

$$\begin{aligned} \text{DOLP}(x) &= \sqrt{S_1(x)^2 + S_2(x)^2} / S_0(x), \\ \text{AOLP}(x) &= \frac{1}{2} \text{atan2}(S_2(x)/S_1(x)). \end{aligned} \quad (5)$$

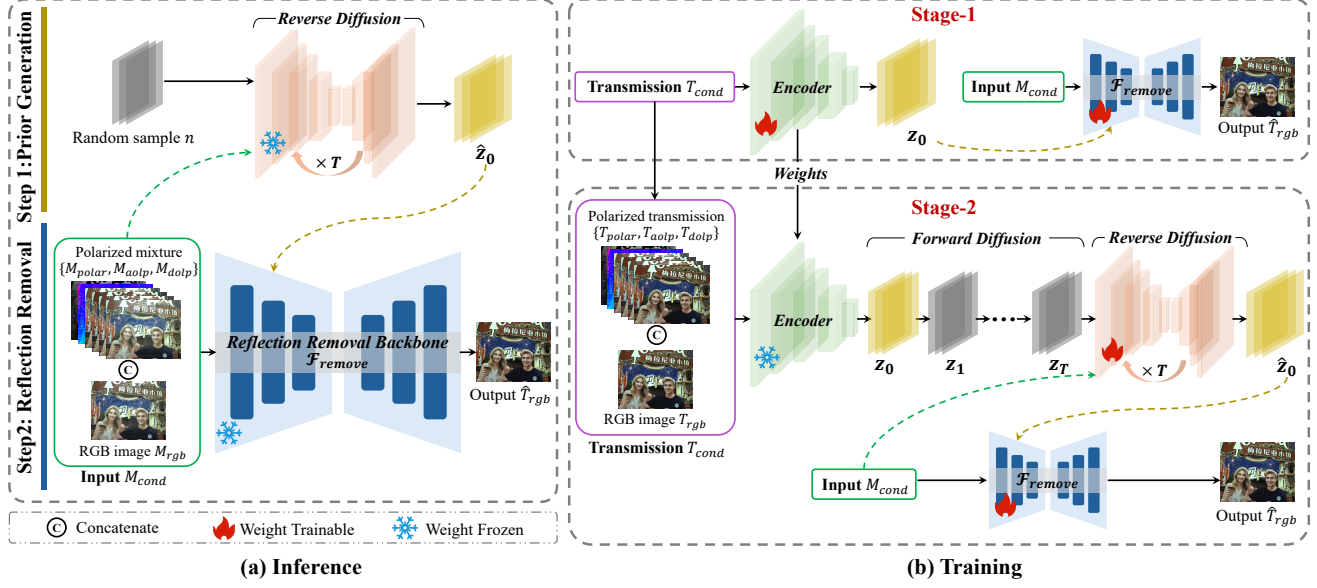


Figure 5. Pipeline of PolarFree. (a) During inference, PolarFree leverages polarized and RGB images as inputs, which are feeds into a conditional diffusion model to generate the prior \hat{z}_0 . The generated prior, along with the inputs, is then passed to the reflection removal backbone $\mathcal{F}_{\text{remove}}$ to remove reflections. (b) PolarFree is trained in two stages. (1) A prior encoder extracts a reflection-free prior z_0 from clean transmission images and polarization cues, which serves as the supervision for the conditional diffusion model in stage two. (2) The conditional diffusion model is trained to progressively denoise noisy images, supervised by the prior from stage one, ensuring robust reflection separation.

In this formulation, DOLP describes the proportion of polarized light relative to total intensity, which helps indicate the degree of reflection in the scene. AOLP, on the other hand, reveals the orientation of the polarized light, allowing us to distinguish between reflection and transmission components more effectively, as shown in Fig. 3d. By leveraging intensity measurements at 0° , 45° , 90° , and 135° , these parameters provide valuable cues for separating mixed reflection and transmission layers in semi-reflective scenes.

4.2. PolarFree Network

To achieve high-quality reflection removal with effective utilization of polarization information, we introduce PolarFree, a dedicatedly designed two-step network, where each step addresses a distinct aspect of the reflection removal challenge. Inspired by [3], we leverage diffusion model to generate the prior for reflection removal. As shown in Fig. 5, during inference, the first step utilizes a conditional diffusion model to extract reflection-free priors, effectively isolating essential details from polarization data. The second step integrates these priors with RGB inputs, guiding the network to accurately separate reflections and enhance clarity, even in complex, real-world environments.

Prior Generation. As shown in Fig. 5a, in the first step, we start with a randomly initialized noise n , which is gradually denoised through a conditional diffusion model $\mathcal{F}_{\text{diff}}$

$$\hat{z}_0 = \mathcal{F}_{\text{diff}}(n|M_{\text{cond}}), \quad (6)$$

where \hat{z}_0 is the generated prior, $M_{\text{cond}} = \{M_{\text{polar}}, M_{\text{aolp}}, M_{\text{dolp}}, M_{\text{rgb}}\}$ represents the mixed im-

ages, and n is the initial noise. Through an iterative denoising process, the diffusion model progressively refines a noise image, conditioned on the polarization and RGB data, to generate a reflection prior.

The denoising process follows the denoising diffusion probabilistic model (DDPM) framework [8] and employs a U-Net architecture to predict noise. At each timestep t , the U-Net receives a noisy intermediate z_t and outputs a noise estimate $\epsilon_\theta(z_t, t)$, predicting the noise added at that timestep. This process can be formulated as

$$z_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(z_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(z_t, M_{\text{cond}}, t) \right) + \sigma_t z, \quad (7)$$

where α_t and β_t control the noise schedule across the timesteps, σ_t is the standard deviation, and z is noise sampled from a standard Gaussian distribution. $\bar{\alpha}_t$ is the cumulative product of α_t , indicating noise level to timestep t .

In this way, the U-Net gradually removes the noise in each step, conditioning on the polarization-based measurement input M_{cond} . This process continues until noise is fully removed, yielding a sample \hat{z}_0 that represents the reflection-free prior distribution extracted from the conditioned inputs.

Reflection Removal. As shown in Fig. 5a, once the prior \hat{z}_0 has been obtained, the second step of PolarFree is separating the transmission and reflection layers. By leveraging the polarization cues provided by M_{polar} , M_{aolp} , and M_{dolp} , the model $\mathcal{F}_{\text{remove}}$ separates the transmission features from the reflection ones, which can be represented as

$$\hat{T}_{\text{rgb}} = \mathcal{F}_{\text{remove}}(\hat{z}_0, M_{\text{cond}}), \quad (8)$$

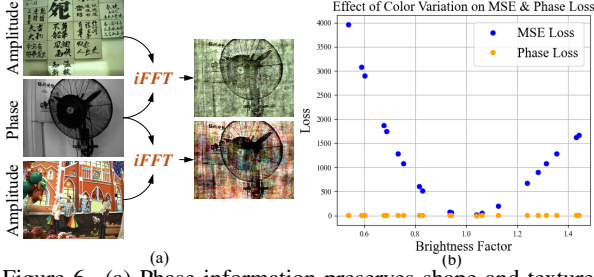


Figure 6. (a) Phase information preserves shape and texture details, while color primarily affects the amplitude. (b) We apply two types of random color perturbations to the image and compute the perturbation errors. Phase-based loss is less sensitive to color changes.

where $\mathcal{F}_{\text{remove}}$ is a reflection removal neural network. This step ensures that the final output contains distinct transmission and reflection components, facilitating high-quality image reconstruction under challenging reflective scenarios.

4.3. Training

To train our PolarFree, particularly to address the challenge of the diffusion model lacking a suitable ground truth, we adopt a two-stage training strategy, as illustrated in Fig. 5b. This strategy consists of two sequential objectives: *learning to extract a reflection-free prior* and *learning to generate a reflection-free prior*.

First Stage. In the first stage, we train an encoder to extract a prior z_0 for reflection-free information, which will serve as the supervision signal for the diffusion model in the second stage. Specifically, we feed the ground truth transmission images $T_{\text{cond}} = \{T_{\text{polar}}, T_{\text{aolp}}, T_{\text{dolp}}, T_{\text{rgb}}\}$ into the encoder \mathcal{E} , obtaining $z = \mathcal{E}(M_{\text{cond}}, T_{\text{rgb}})$. This z contains reflection-free information enriched with polarization-related cues. Subsequently, z_0 conditions the reflection removal backbone $\mathcal{F}_{\text{remove}}$ to predict the clean transmission image $\hat{T}_{\text{rgb}} = \mathcal{F}_{\text{remove}}(z, M_{\text{cond}})$.

Second Stage. In the second stage, we train the diffusion model $\mathcal{F}_{\text{diff}}$ to generate a reflection-free prior z_0 from noisy input images, and finetune the reflection removal backbone $\mathcal{F}_{\text{remove}}$. The key challenge here is that the diffusion model lacks of direct supervision. Therefore, we leverage the prior z_0 extracted in the first stage as the supervision signal to guide the model.

We start with the extracted reflection-free prior z_0 and add noise over multiple timesteps. This process transforms the “ground-truth” prior z_0 into noisy versions z_t at each timestep t , which can be expressed as

$$q(z_t|z_0) = \mathcal{N}(z_t; \sqrt{1 - \beta_t}z_0, \beta_t\mathbf{I}), \quad (9)$$

where \mathcal{N} represents a Gaussian distribution, β_t is the noise schedule controlling the noise level added at each timestep, \mathbf{I} represents the identity matrix.

In the reverse diffusion process, the model is trained to progressively remove the noise step by step to recover the

clean prior z_0 , which can be represented as

$$p_\theta(z_{t-1}|z_t) = \mathcal{N}(z_{t-1}; \mu_\theta(z_t, t, M_{\text{cond}}), \sigma_t^2\mathbf{I}), \quad (10)$$

where $\mu_\theta(z_t, t, M_{\text{cond}})$ is the mean function of the current state z_t , timestep t , and a conditional input M_{cond} .

Throughout the reverse process, the model conditions on the polarization measurements M_{cond} , which provide essential information about the scene’s physical properties. These measurements help the model accurately generate reflection-free components during the denoising process.

After the reverse process completes, the model outputs a final reflection-free prior \hat{z}_0 , which contains clean transmission information. This prior is used as the guidance signal for the reflection removal backbone $\mathcal{F}_{\text{remove}}$. The backbone takes this reflection-free prior and the polarization data to predict the clean transmission image \hat{T}_{rgb} .

4.4. Losses

Basic Loss. To optimize PolarFree, we follow [46] to utilize three basic losses: L1 loss, VGG perceptual loss [14], and total variation (TV) loss [33]. The L1 loss minimizes the pixel-wise difference between the predicted transmission image \hat{T}_{rgb} and the ground truth transmission image T_{rgb} as $\mathcal{L}_1 = \|\hat{T}_{\text{rgb}} - T_{\text{rgb}}\|_1$. The VGG perceptual loss compares feature activations from a pre-trained VGG network [14], weighted by λ_l : $\mathcal{L}_{\text{VGG}} = \sum_l \lambda_l \|\phi_l(\hat{T}_{\text{rgb}}) - \phi_l(T_{\text{rgb}})\|_1$, where ϕ_l denotes the activations of the l -th VGG layer. We also use TV loss to constrain consistency via the gradient operator $\mathcal{L}_{\text{TV}} = \|\nabla \hat{T}_{\text{rgb}} - \nabla T_{\text{rgb}}\|_1$.

Phase loss. While basic losses help the network match color and intensity values to the ground truth, they struggle with color discrepancies caused by semi-reflective surfaces during dataset collection. These variations, due to reflection and transmission properties of materials, affect the color and intensity, leading to mismatches between the predicted and target images and hindering the model’s ability to learn the correct transmission map.

To address this issue, we introduce a phase loss to focus on the structural information of the transmission, which is less sensitive to color variations. As shown in Fig. 6, phase information primarily captures the geometry and texture of the image, independent of color changes. The phase loss is formulated as

$$\mathcal{L}_{\text{phase}} = \|\angle(FFT(\hat{T})) - \angle(FFT(T_{\text{rgb}}))\|_1, \quad (11)$$

where FFT is the Fourier transform, and $\angle(\cdot)$ represents the phase angle of the Fourier coefficients.

Diffusion Loss. The diffusion loss follows the standard DDPM formulation [8], where the model predicts the noise at each step t and minimizes the difference between the predicted noise and the true noise as

$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{q(z_t|z_{t-1})} [\|\epsilon_\theta(z_t, t) - \epsilon_{\text{true}}(z_t)\|_2^2]. \quad (12)$$

Here, $\epsilon_{\text{true}}(z_t)$ is the added noise obtained by Eq. 9, and $\epsilon_\theta(z_t, t)$ is the predicted noise, conditioned on the polariza-

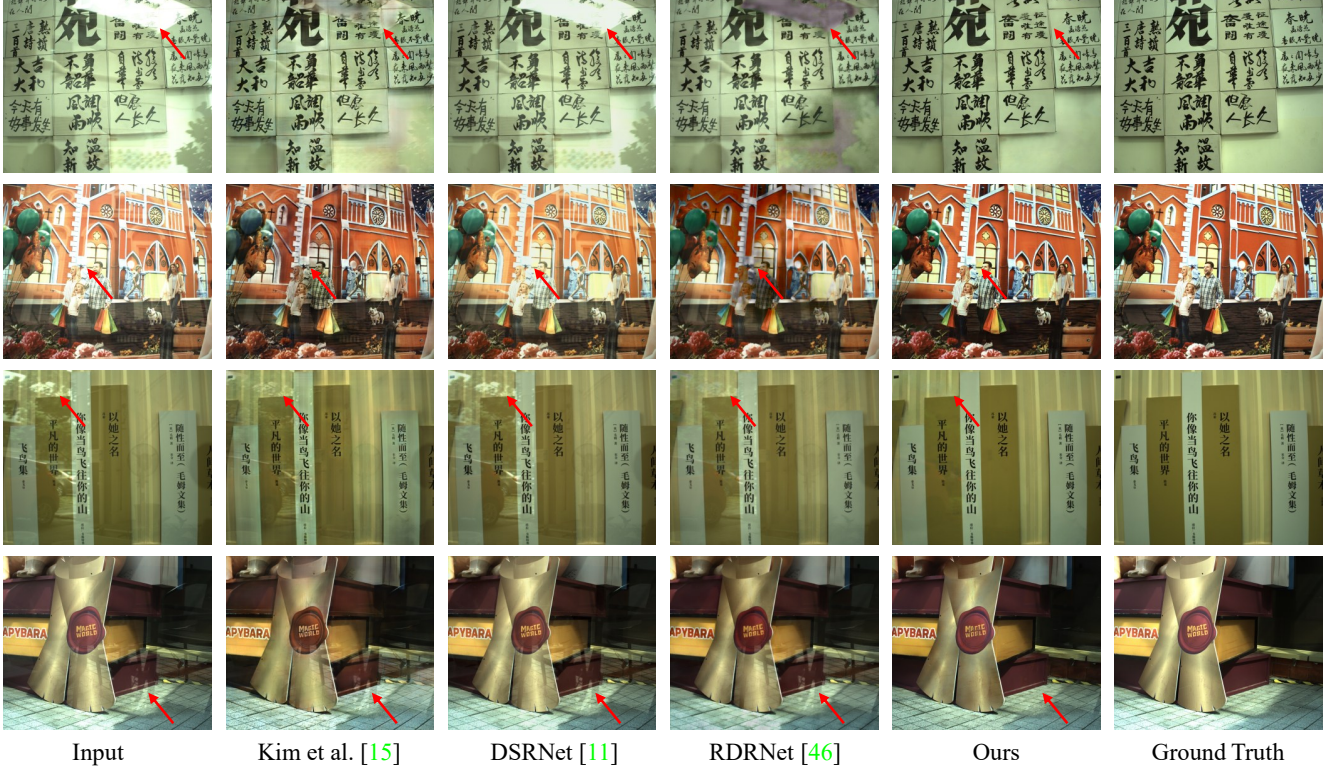


Figure 7. Qualitative comparisons on the PolaRGB dataset.

tion measurements and RGB input M_{cond} .

Total Loss. The above losses are weighted summed to serve as the supervision for the first stage as

$$\mathcal{L}_{\text{stage1}} = \gamma_1 \mathcal{L}_1 + \gamma_2 \mathcal{L}_{\text{VGG}} + \gamma_3 \mathcal{L}_{\text{TV}} + \gamma_4 \mathcal{L}_{\text{phase}}. \quad (13)$$

In the second stage, we use a combined loss function consisting of the diffusion loss and the reconstruction loss as

$$\mathcal{L}_{\text{stage2}} = \gamma_5 \mathcal{L}_{\text{diff}} + \gamma_6 \mathcal{L}_{\text{recon}}. \quad (14)$$

where $\mathcal{L}_{\text{recon}}$ has the same format as $\mathcal{L}_{\text{stage1}}$, and $\gamma(\cdot)$ are the coefficients. This combined loss helps refine the reflection-free image while ensuring it aligns with the ground truth.

5. Experiments

5.1. Implementation Details

We train and evaluate PolarFree on the PolaRGB dataset. The whole dataset consists of 67 scenes, with a total of 6,500 paired images. For each scene, we keep the background (transmission) and camera fixed, and varying the glass position to capture images with reflections. The 67 scenes are then randomly split into 56 training scenes and 11 testing scenes, containing 6,312 and 188 paired images, respectively. This division ensures no data leakage between the training and testing sets, with each set containing only a subset of all categories.

We also test PolarFree in real-world scenes like museums and galleries, where ground truth is unavailable. We implement PolarFree using PyTorch on a single NVIDIA RTX 4090 GPU. Training is conducted with a batch size of

2 and an AdamW optimizer with a learning rate of 2×10^{-4} , 30k iterations on the PolaRGB dataset for both stages.

We compare our method with recent advanced reflection removal methods, including Lei *et al.* [18], IBCLN [21], DSRNet [11], YTMT [10], and RDRNet [46]. For a fair comparison, we modify the input settings to align with ours and use only the transmission layer for supervision. We have re-trained baseline methods on the PolaRGB dataset. The evaluation is conducted using both objective metrics (PSNR, SSIM), a perceptual metric (LPIPS), and a language-based non-reference metric (Q-Align [40]).

5.2. Results

Results on PolaRGB. We present the quantitative results in Tab. 2. It can be seen that, our method outperforms other methods across multiple metrics, demonstrating the effectiveness of PolarFree for high-quality reflection removal. Notably, for fairness, we modified the input layer of all baselines to accept polarization information. Visual results in Fig. 7 show that PolarFree provides cleaner reflection removal with sharper edges and better color preservation. In contrast, previous methods often suffer from color distortions or imperfect reflection removal, especially in areas with complex reflections. Our method maintains high fidelity to the ground truth, especially in challenging regions with low light and subtle reflections.

Real-captured without Ground Truth. To demonstrate the generalization ability to unseen and more complex re-

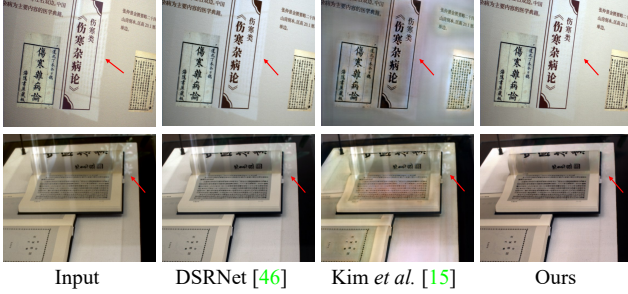


Figure 8. Visual comparison of reflection removal results in real-world museum scenes with glass display cases.

Table 2. Quantitative comparison on the PolaRGB dataset, evaluated using objective metrics (PSNR, SSIM), perceptual metric (LPIPS), and the Language-based non-reference metric (Q-align [40]). Best results are in **bold**, second-best results are underlined.

Method	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	Q-align \uparrow
Lei et al. [18]	18.73	0.7962	0.3804	3.2109
Kim et al. [15]	20.67	<u>0.8399</u>	0.2714	3.7148
IBCLN [21]	19.73	<u>0.8173</u>	0.2488	3.0938
YTMT [10]	16.86	0.7544	0.4489	3.0938
DSRNet [11]	16.84	0.7913	0.2828	3.6992
RDRNet [46]	15.88	0.6964	0.5250	2.9531
Ours	22.44	0.8681	0.1325	3.8867

reflection, We further evaluate our model on real-captured images at museum with complex reflections from glass enclosures. Here, we only provide qualitative results to visually assess the model’s capability. As shown in Fig. 8, our approach effectively reduces reflections while preserving fine details, despite the complex lighting and material variations often present in museum environments. This demonstrates the practical robustness of our method and its potential for real-world applications.

5.3. Ablation Study

We perform ablation studies to evaluate the effectiveness of key components in our method. More experimental results and analyses can be found in the supplementary.

Polarization Information. To assess the importance of polarization cues, we conduct an experiment where the polarization information (AOLP, DOLP, and polarization images) is removed, and the model is trained solely using RGB data. As shown in Table 3 and Fig. 9, the absence of polarization significantly impairs the model’s ability to differentiate between reflection and transmission layers, leading to noticeably poorer reflection removal performance. These results demonstrate the critical role of polarization in accurate reflection removal.

Diffusion Prior. We also evaluate the effectiveness of diffusion prior by removing the conditional diffusion model. While without diffusion prior achieves reasonable results, as indicated by the comparison metrics in Table 3, it lacks fine-grained control and high-quality output offered by the

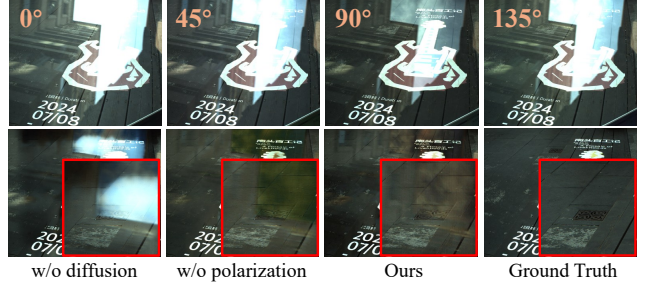


Figure 9. Highly challenging reflection removal scenes with complex reflections and highlights. Top: Polarization images from different angles provide complementary information for effective reflection removal, note the third (90°) image. Bottom: The diffusion module effectively handles highlights, while polarization better restores the color and details in such challenging scenarios.

Table 3. Ablation study on PolarFree framework.

Method	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	Q-align \uparrow
w/o \mathcal{L}_{phase}	22.41	0.8622	0.1402	3.8524
w/o polarization	21.56	0.8620	0.1483	3.7949
w/o diffusion prior	20.21	0.8627	0.1552	3.7949
Ours	22.44	0.8681	0.1325	3.8867

diffusion model, especially in complex scenes (Fig. 9). The conditional diffusion process significantly improves separation accuracy and preserves finer transmission details.

Phase-based Loss. We explore the impact of our phase-based loss function. We remove this loss from the training procedure and observe the model’s performance without it. The results show a noticeable increase in color discrepancies and less accurate reflection removal, particularly in scenarios involving semi-reflective surfaces (Table 3). The phase-based loss function ensures the model focuses on reflection removal rather than compensating for color inconsistencies, leading to improved overall results.

6. Conclusion

In this work, we propose PolaRGB, a large-scale, comprehensive dataset for polarization-based reflection removal. With 6,500 well-aligned RGB-polarization image pairs, PolaRGB is 8× larger than existing polarization datasets and is the first to include both RGB and polarization images, uniquely captured across diverse real-world scenes and lighting conditions. Additionally, we present PolarFree, a novel reflection removal model that leverages the generative capabilities of diffusion models. This approach enables PolarFree to generate reflection-free cues, enhancing separation accuracy while preserving fine details in the transmission layer. A phase-based loss function is also introduced to further improve the model’s performance. Comprehensive experimental results on realistic scenarios demonstrate the effectiveness of our method. We believe these contributions set a foundation for further advancements in polarization-based reflection removal, paving the way for more sophisticated applications in complex real-world environments.

Acknowledgments

This work was supported in part by the National Key R&D Program of China (No.2022ZD0160201) and Shanghai Artificial Intelligence Laboratory, and the RGC Early Career Scheme (ECS) No. 24209224.

References

- [1] Amit Agrawal, Ramesh Raskar, Shree K Nayar, and Yuanzhen Li. Removing photography artifacts using gradient projection and flash-exposure sampling. *ACM Transactions on Graphics*, 24(3):828–835, 2005. [1](#), [2](#)
- [2] Max Born and Emil Wolf. *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Elsevier, 2013. [1](#), [2](#), [4](#)
- [3] Zheng Chen, Yulun Zhang, Ding Liu, Jinjin Gu, Linghe Kong, Xin Yuan, et al. Hierarchical integration diffusion model for realistic image deblurring. *Advances in neural information processing systems*, 36, 2024. [2](#), [5](#)
- [4] Hany Farid and Edward H Adelson. Separating reflections and lighting using independent components analysis. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, volume 1, pages 262–267. IEEE, 1999. [2](#), [3](#)
- [5] Yuanbiao Gou, Boyun Li, Zitao Liu, Songfan Yang, and Xi Peng. Clearer: Multi-scale neural architecture search for image restoration. *Advances in neural information processing systems*, 33:17129–17140, 2020. [2](#)
- [6] Yuanshen Guan, Ruikang Xu, Mingde Yao, Ruisheng Gao, Lizhi Wang, and Zhiwei Xiong. Diffusion-promoted hdr video reconstruction. *arXiv preprint arXiv:2406.08204*, 2024. [2](#)
- [7] Byeong-Ju Han and Jae-Young Sim. Reflection removal using low-rank matrix completion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5438–5446, 2017. [2](#)
- [8] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. [2](#), [5](#), [6](#)
- [9] Yuchen Hong, Haofeng Zhong, Shuchen Weng, Jinxiu Liang, and Boxin Shi. L-differ: Single image reflection removal with language-based diffusion model. In *Proceedings of the european conference on computer vision (ECCV)*, 2024. [2](#)
- [10] Qiming Hu and Xiaojie Guo. Trash or treasure? an interactive dual-stream strategy for single image reflection separation. *Advances in Neural Information Processing Systems*, 34, 2021. [1](#), [2](#), [7](#), [8](#)
- [11] Qiming Hu and Xiaojie Guo. Single image reflection separation via component synergy. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13138–13147, 2023. [1](#), [2](#), [7](#), [8](#)
- [12] Zhiyong Huang, Biao Xiong, Cao Tian, Jing Zhan, Xiang Fei, and Nazaraf Shah. Dust and reflection removal from videos captured in moving car. In *2016 IEEE 13th International Conference on e-Business Engineering (ICEBE)*, pages 182–187. IEEE, 2016. [1](#)
- [13] Yujin Jeon, Eunsue Choi, Youngchan Kim, Yunseong Moon, Khalid Omer, Felix Heide, and Seung-Hwan Baek. Spectral and polarization vision: Spectro-polarimetric real-world dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22098–22108, 2024. [2](#)
- [14] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016. [6](#)
- [15] Soomin Kim, Yuchi Huo, and Sung-Eui Yoon. Single image reflection removal with physically-based training images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5164–5173, 2020. [1](#), [2](#), [8](#)
- [16] Naejin Kong, Yu-Wing Tai, and Joseph S Shin. A physically-based approach to reflection separation: from physical modeling to constrained optimization. *IEEE transactions on pattern analysis and machine intelligence*, 36(2):209–221, 2013. [3](#)
- [17] Chenyang Lei and Qifeng Chen. Robust reflection removal with reflection-free flash-only cues. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14811–14820, 2021. [2](#)
- [18] Chenyang Lei, Xuhua Huang, Mengdi Zhang, Qiong Yan, Wenxiu Sun, and Qifeng Chen. Polarized reflection removal with perfect alignment in the wild. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1750–1758, 2020. [1](#), [2](#), [3](#), [4](#), [7](#), [8](#)
- [19] Chenyang Lei, Xudong Jiang, and Qifeng Chen. Robust reflection removal with flash-only cues in the wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. [1](#), [2](#)
- [20] Anat Levin and Yair Weiss. User assisted separation of reflections from a single image using a sparsity prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(9):1647–1654, 2007. [2](#)
- [21] Chao Li, Yixiao Yang, Kun He, Stephen Lin, and John E Hopcroft. Single image reflection removal through cascaded refinement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3565–3574, 2020. [1](#), [7](#), [8](#)
- [22] Tingtian Li, Yuk-Hee Chan, and Daniel PK Lun. Improved multiple-image-based reflection removal algorithm using deep neural networks. *IEEE Transactions on Image Processing*, 30:68–79, 2020. [2](#)
- [23] Xiaobo Li, Lei Yan, Pengfei Qi, Liping Zhang, François Goudail, Tiegeng Liu, Jingsheng Zhai, and Haofeng Hu. Polarimetric imaging via deep learning: A review. *Remote Sensing*, 15(6):1540, 2023. [2](#)
- [24] Yu Li and Michael S Brown. Single image layer separation using relative smoothness. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2752–2759, 2014. [1](#), [2](#)
- [25] Youwei Lyu, Zhaopeng Cui, Si Li, Marc Pollefeys, and Boxin Shi. Reflection separation using a pair of unpolarized and polarized images. *Advances in neural information processing systems*, 32, 2019. [2](#), [3](#)

- [26] Ajay Nandoriya, Mohamed Elgharib, Changil Kim, Mohamed Hefeeda, and Wojciech Matusik. Video reflection removal through spatio-temporal optimization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2411–2419, 2017. 2
- [27] Shree K Nayar, Xi-Sheng Fang, and Terrance Boult. Separation of reflection components using color and polarization. *International Journal of Computer Vision*, 21(3):163–186, 1997. 1, 2, 3, 4
- [28] Simon Niklaus, Xuaner Cecilia Zhang, Jonathan T Barron, Neal Wadhwa, Rahul Garg, Feng Liu, and Tianfan Xue. Learned dual-view reflection removal. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3713–3722, 2021. 2
- [29] Yoav Y Schechner, Joseph Shamir, and Nahum Kiryati. Polarization-based decorrelation of transparent layers: The inclination angle of an invisible surface. In *Proceedings of the seventh IEEE international conference on computer vision*, volume 2, pages 814–819. IEEE, 1999. 1, 4
- [30] Yoav Y Schechner, Joseph Shamir, and Nahum Kiryati. Polarization and statistical analysis of scenes containing a semireflector. *JOSA A*, 17(2):276–284, 2000. 1, 2, 3, 4
- [31] YiChang Shih, Dilip Krishnan, Fredo Durand, and William T Freeman. Reflection removal using ghosting cues. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3193–3201, 2015. 2
- [32] Sony Semiconductor Solutions. Polarsens: Polarization sensing technology, 2024. Accessed: 2024-11-15. 1
- [33] David Strong and Tony Chan. Edge-preserving and scale-dependent properties of total variation regularization. *Inverse problems*, 19(6):S165, 2003. 6
- [34] Robby T Tan and Katsushi Ikeuchi. Separating reflection components of textured surfaces using a single image. *IEEE transactions on pattern analysis and machine intelligence*, 27(2):178–193, 2005. 1, 2
- [35] Renjie Wan, Boxin Shi, Ling-Yu Duan, Ah-Hwee Tan, and Alex C Kot. Crnn: Multi-scale guided concurrent reflection removal network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4777–4785, 2018. 2
- [36] Renjie Wan, Boxin Shi, Haoliang Li, Ling-Yu Duan, Ah-Hwee Tan, and Alex C Kot. Corrn: Cooperative reflection removal network. *IEEE transactions on pattern analysis and machine intelligence*, 42(12):2969–2982, 2019. 2
- [37] Mengyi Wang, Xinxin Zhang, Yongshun Gong, and Yilong Yin. Personalized single image reflection removal network through adaptive cascade refinement. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 8204–8213, 2023. 2
- [38] Kaixuan Wei, Jiaolong Yang, Ying Fu, David Wipf, and Hua Huang. Single image reflection removal exploiting misaligned training data and network enhancements. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8178–8187, 2019. 1, 2
- [39] Patrick Wieschollek, Orazio Gallo, Jinwei Gu, and Jan Kautz. Separating reflection and transmission images in the wild. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 89–104, 2018. 1, 2, 3, 4
- [40] Haoning Wu, Zicheng Zhang, Weixia Zhang, Chaofeng Chen, Liang Liao, Chunyi Li, Yixuan Gao, Annan Wang, Erli Zhang, Wenxiu Sun, et al. Q-align: Teaching lms for visual scoring via discrete text-defined levels. *arXiv preprint arXiv:2312.17090*, 2023. 7, 8
- [41] Tianfan Xue, Michael Rubinstein, Ce Liu, and William T Freeman. A computational approach for obstruction-free photography. *ACM Transactions on Graphics (TOG)*, 34(4):1–11, 2015. 1
- [42] Mingde Yao, Ruikang Xu, Yuanshen Guan, Jie Huang, and Zhiwei Xiong. Neural degradation representation learning for all-in-one image restoration. *IEEE Transactions on Image Processing*, 2024. 2
- [43] Jae-Seong Yun and Jae-Young Sim. Reflection removal for large-scale 3d point clouds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4597–4605, 2018. 1
- [44] Xuaner Zhang, Ren Ng, and Qifeng Chen. Single image reflection separation with perceptual losses. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4786–4794, 2018. 2
- [45] Haofeng Zhong, Yuchen Hong, Shuchen Weng, Jinxiu Liang, and Boxin Shi. Language-guided image reflection separation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24913–24922, 2024. 2
- [46] Yurui Zhu, Xueyang Fu, Peng-Tao Jiang, Hao Zhang, Qibin Sun, Jinwei Chen, Zheng-Jun Zha, and Bo Li. Revisiting single image reflection removal in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25468–25478, 2024. 1, 2, 6, 7, 8