Polarization State Tracing for Reflection Removal and Color-Consistent Reconstruction

Supplementary Material

When a camera captures an image through a glass medium, scenes on the same side as the camera may appear in the final photo. These reflection artifacts overlay the target scene and can significantly degrade the image's texture, color, polarization, and other optical information, further interfering with the execution of other visual tasks. In modern cities, for reasons such as aesthetics, insulation, or craftsmanship, colored glass is commonly used. Fig. 1 shows an example of a photo captured through glass containing reflections, as well as an example of colored glass.

Figure 1. Including reflective images and common examples of architectural exterior colored glass.

1. Polarization State Tracing Model (PSTM)

In this section, we provide a more detailed derivation and explanation of Sec. 3 of the main paper to more completely explain how wavelength–selective absorption modifies multi-bounce polarization propagation inside colored glass. Besides, we provide a detailed explanation of how to handle color bias and ghosting issues in this colored medium to address the problem of colored glass degradation. We describe the geometric laws that incident light propagation inside a glass slab, the spectral attenuation arising from colored media, and the polarization transformations at dielectric interfaces. These principles jointly show how reflection and transmission scenes propagate through multi-layer colored glass.

1.1. Basic Knowledge

Snell's Law and the Geometry Inside the Medium. Consider a homogeneous glass plate of refractive index n and physical thickness t, surrounded by air ($n_0 \approx 1$). Light from either the reflection side scene or transmission side

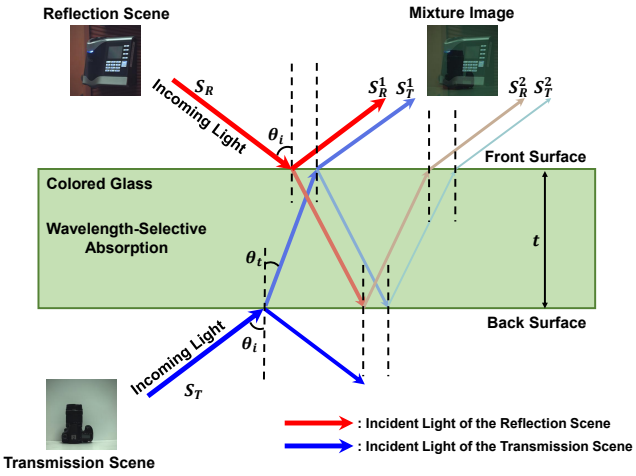

Figure 2. The proposed polarization new imaging model (PSTM) through colored glass.

scene impinges on the front interface with incident angle θ_i. The direction of refraction inside the glass is controlled by Snell's law

$$n_0 \sin\theta_i = n \sin\theta_t. \tag{1}$$

For oblique propagation, light travels along the effective geometric path,

$$L = \frac{t}{\cos\theta_t}, \tag{2}$$

which is strictly larger than t unless $\theta_t = 0$. After reaching the back interface, part of the light reflects internally and returns to the front surface. This bottom-up traversal results in a lateral displacement

$$\Delta x = 2t \tan\theta_t, \tag{3}$$

which is responsible for the spatial offset (ghost shadows) observed in the colored glass coordinate system.

Beer-Lambert Spectral Attenuation. Colored glass exhibits wavelength-dependent absorption, which increases exponentially with propagation distance. For the RGB channels, the diagonal attenuation matrix is obtained according to the Beer-Lambert law:

$$E(mL) = \mathrm{diag}\left(e^{-\beta_r mL}, e^{-\beta_g mL}, e^{-\beta_b mL}\right), \tag{4}$$

where $\beta_r, \beta_g, \beta_b$ represent the spectral absorption coefficients, and mL is the total distance traveled within the colored glass medium. Since absorption preserves polarization, the matrix in Eq.4 scales the three linear Stokes components $[S_0, S_1, S_2]$ uniformly for each color channel.

Fresnel Reflection and Transmission. At each dielectric interface, reflection and transmission depend on the direction of polarization relative to the plane of incidence. Let $R^1_{\parallel,\perp}$ and $T^1_{\parallel,\perp}$ denote the Fresnel power coefficients at the air to glass interface, and $R^2_{\parallel,\perp}$, $T^2_{\parallel,\perp}$ those at the glass to air interface. Removing the circular Stokes component yields the linear Mueller matrices. When light propagates from the outside (optically less dense medium, air) to the inside (optically denser medium, glass), the four orthogonal intensity coefficients are:

$$\begin{aligned} R^1_\perp(\theta_i, n) &= \frac{\sin^2(\theta_i - \theta_t)}{\sin^2(\theta_i + \theta_t)}, \\ R^1_\parallel(\theta_i, n) &= \frac{\tan^2(\theta_i - \theta_t)}{\tan^2(\theta_i + \theta_t)}, \\ T^1_\perp(\theta_i, n) &= \frac{\sin 2\theta_i \, \sin 2\theta_t}{\sin^2(\theta_i + \theta_t)}, \\ T^1_\parallel(\theta_i, n) &= \frac{\sin 2\theta_i \, \sin 2\theta_t}{\sin^2(\theta_i + \theta_t) \, \cos^2(\theta_i - \theta_t)}. \end{aligned} \tag{5}$$

When light propagates from an internal medium (optically denser, glass) to an external medium (optically rarer, air), the corresponding coefficients become:

$$\begin{aligned} R^2_\perp(\theta_i, n) &= \frac{\sin^2(\theta_t - \theta_i)}{\sin^2(\theta_t + \theta_i)}, \\ R^2_\parallel(\theta_i, n) &= \frac{\tan^2(\theta_t - \theta_i)}{\tan^2(\theta_t + \theta_i)}, \\ T^2_\perp(\theta_i, n) &= \frac{\sin 2\theta_t \, \sin 2\theta_i}{\sin^2(\theta_t + \theta_i)}, \\ T^2_\parallel(\theta_i, n) &= \frac{\sin 2\theta_t \, \sin 2\theta_i}{\sin^2(\theta_t + \theta_i) \, \cos^2(\theta_t - \theta_i)}. \end{aligned} \tag{6}$$

Here, the superscript 1 refers to the interface from air to glass (front surface), and 2 refers to the interface from glass to air (back surface). These scalar Fresnel coefficients will be used to construct the interface Mueller matrix in our PSTM derivation.

1.2. Derivation of the PSTM Degradation Model

We now derive the complete color glass degradation process based on Snell's geometry, Beer–Lambert absorption, and Fresnel–Muller polarization transmission. All derivations consider only the first two physically significant propagation orders. Higher-order contributions decay rapidly and can be safely ignored in practice.

Front-surface reflection is:

$$M^1_R = \begin{bmatrix} \frac{R^1_\parallel + R^1_\perp}{2} & \frac{R^1_\parallel - R^1_\perp}{2} & 0 \\ \frac{R^1_\parallel - R^1_\perp}{2} & \frac{R^1_\parallel + R^1_\perp}{2} & 0 \\ 0 & 0 & \sqrt{R^1_\parallel R^1_\perp} \end{bmatrix}. \tag{7}$$

Front-surface transmission is:

$$M^1_T = \begin{bmatrix} \frac{T^1_\parallel + T^1_\perp}{2} & \frac{T^1_\parallel - T^1_\perp}{2} & 0 \\ \frac{T^1_\parallel - T^1_\perp}{2} & \frac{T^1_\parallel + T^1_\perp}{2} & 0 \\ 0 & 0 & \sqrt{T^1_\parallel T^1_\perp} \end{bmatrix}. \tag{8}$$

Back-surface reflection is:

$$M^2_R = \begin{bmatrix} \frac{R^2_\parallel + R^2_\perp}{2} & \frac{R^2_\parallel - R^2_\perp}{2} & 0 \\ \frac{R^2_\parallel - R^2_\perp}{2} & \frac{R^2_\parallel + R^2_\perp}{2} & 0 \\ 0 & 0 & \sqrt{R^2_\parallel R^2_\perp} \end{bmatrix}. \tag{9}$$

Back-surface transmission is:

$$M^2_T = \begin{bmatrix} \frac{T^2_\parallel + T^2_\perp}{2} & \frac{T^2_\parallel - T^2_\perp}{2} & 0 \\ \frac{T^2_\parallel - T^2_\perp}{2} & \frac{T^2_\parallel + T^2_\perp}{2} & 0 \\ 0 & 0 & \sqrt{T^2_\parallel T^2_\perp} \end{bmatrix}. \tag{10}$$

The set of matrices $\{M^1_R, M^1_T, M^2_R, M^2_T\}$ completely shows the transformation of linear polarization at the two interfaces of the glass.

Two independent polarized scenes illuminate the glass: a *reflection-side scene* $S_R(x)$ and a *transmission-side scene* $S_T(x)$. Each scene generates its own set of reflection and transmission paths. The Stokes vector measured at position x is the sum of contributions from all valid paths of the two scenes.

A part of $S_R(x)$ will be reflected immediately at the front interface:

$$S^1_R(x) = M^1_R \, S_R(x). \tag{11}$$

The scene at the transmission end enters the glass, travels a distance L, undergoes back-surface transmission, and exits:

$$S^1_T(x) = E(L) \, M^2_T M^1_T S_T(x). \tag{12}$$

These first-order terms dominate the perceptual appearance of the foreground (through-glass) and background (front reflection) layers.

A fraction of S_R enters the glass, reaches the back interface, reflects, passes through the slab again, and exits at the front surface:

$$S^2_R(x) = E(2L) \, M^2_T M^2_R M^1_T \, S_R(x + \Delta x). \tag{13}$$

The lateral shift Δx arises from the two oblique traversals inside the slab.

Similarly, a component of S_T enters the glass, reflects at the back surface, then reflects again at the front surface before leaving:

$$S^2_T(x) = E(3L) \, M^2_T M^2_R M^1_R M^1_T \, S_T(x + \Delta x). \tag{14}$$

These second-order paths correspond exactly to the spatially shifted ghost shadows visible in real colored-glass images. Due to the longer propagation time in the glass ($2L$ or $3L$), these parts experience stronger spectral attenuation, resulting in reduced intensity and changes in color balance.

Collecting all physically relevant paths results in the complete PSTM degradation model:

$$S(x) = S_R^1(x) + S_T^1(x) + S_R^2(x) + S_T^2(x). \tag{15}$$

Higher-order paths (S^k, $k \geq 3$) are exponentially suppressed by Fresnel losses and Beer-Lambert absorption, and therefore omitted.

Eq. 15 forms the basis of the image formation model used throughout the main paper. It explicitly captures full-process interfacial polarization modulation, multi-path propagation, wavelength-selective attenuation, and geometric lateral displacement. This principled derivation distinguishes PSTM from all previous reflection-removal or polarization-based methods, which either treat layers as simple additive mixtures or ignore the interaction between polarization, geometry, and spectral absorption.

1.3. Sensitivity analysis of the PSTM model

The full PSTM can be written as **an infinite sum** over multi-bounce polarization paths as shown in Eq. 16, while each term decays exponentially with propagation order due to Fresnel losses and Beer-Lambert absorption. Our **2nd-order truncation (two rays leaving each side)** corresponds to retaining the dominant terms of this rapidly convergent series. **For extreme cases**, the model can be naturally extended by higher-order model.

$$\mathbf{S}(\mathbf{x}) = \sum_{k=0}^{\infty} \left(\mathbf{S}_R^{(k)}(\mathbf{x}) + \mathbf{S}_T^{(k)}(\mathbf{x}) \right). \tag{16}$$

We compare the 8th-order PSTM with the 2nd-order truncation over $t \in [2, 10]\, mm, \theta_i \in [0°, 60°]$, and absorption scales Ab_s $\{0.5, 1.0, 1.5\}$. Across all tested settings, the error is negligible, with a **minimum PSNR of 60.6 dB and a minimum SSIM of 0.99,** and the MAE remains below 1×10^{-3} even in ghost-shadow regions (Fig. 3(a)). Therefore, **the 2nd-order truncated modeling is effective.**

2. The GlassPol Dataset

Fig. 4 presents several representative examples from our GlassPol dataset. For each scene, the camera records four polarization channels (0°, 45°, 90°, 135°) in a single shot through the integrated micro–polarizer array, ensuring perfect spatial alignment without temporal inconsistencies. We include four typical indoor environments, *chair*, *dolls*, *kettle*, and *plants*, which collectively span a wide range of material properties, covering diffuse surfaces, glossy objects, and complex mixed reflection configurations. These

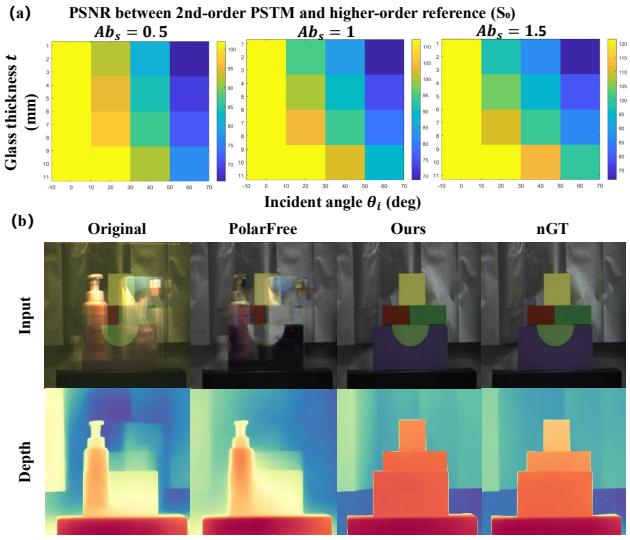

Figure 3. PSNR between the 2nd-order PSTM and a higher-order reference in the recovered S_0, evaluated under different glass thicknesses, incident angles θ_i, and absorption coefficients Ab_s. (b) Qualitative comparison on a real-world example about depth estimation.

scenes reflect common real–world settings and capture diverse polarization signatures, making GlassPol a suitable benchmark for studying colored glass degradation, polarization analysis, and reflection separation under physically realistic conditions.

As described in Sec. 5, all dataset images were captured using a custom polarization-based imaging device. The complete hardware configuration and data acquisition setup are shown in Fig. 5. We use a FLIR Blackfly BFS-U3-51S5P-C polarization camera based on a 5 MP Sony IMX250MZR sensor. The sensor incorporates an on-chip micro-polarizer array with four orientations (0°, 45°, 90°, 135°), so that all four polarization channels are recorded in a single exposure. An 8 mm fixed-focus industrial lens (selected according to the scene) is attached to provide a stable field of view. The camera is securely mounted on a vibration-isolated tripod and equipped with a fine-adjustment ball head to ensure precise alignment.

For our uncoated colored-glass dataset, we analyze per-angle channel intensity ratios using $r_\theta^{(A/B)} = \log\big((I_{\theta,A} + \epsilon)/(I_{\theta,B} + \epsilon)\big)$ for $\theta \in \{0°, 45°, 90°, 135°\}$ and $(A, B) \in \{(R, G), (R, B), (G, B)\}$. In low-DoLP regions, there is no systematic angle-dependent drift, with variations limited to about 1-2%, supporting polarization-independent attenuation in our setting.

Various colored glass samples, including orange-tinted acrylic, brown architectural coatings, and purple-tinted sheets, are placed between the camera and the object. Each plate introduces wavelength-selective transmission as well

Figure 4. Visualization of representative scenes from our GlassPol dataset. Each column corresponds to one of the four polarization orientations (0°, 45°, 90°, 135°) captured simultaneously by the micro–polarizer array in a single shot. From top to bottom: *chair*, *dolls*, *kettle*, and *plants*. These scenes cover diverse materials such as diffuse surfaces, glossy regions, and mixed reflection types, providing broad coverage of real–world polarization behaviors.

as multiple internal reflections, which produce the colored ghost reflections studied in this work. The glass is fixed using a kinematic mount to ensure repeatable and stable placement. During data acquisition, camera exposure, white balance, and illumination are kept fixed before and after removing the glass, and the paired captures are taken in quick succession to minimize environmental changes. We will add this strict capture protocol explicitly to the revised paper. Moreover, since the proposed pipeline operates on downsampled inputs, it is less sensitive to tiny residual misalignments between the paired captures.

3. Ablation Study

To understand the contribution of each component in our PANet, we conduct a comprehensive ablation study. Table 3 reports the quantitative results after removing individual modules, including the Window-Level Attention (WLA), Channel-Ring Attention (CRA), Local Attention Feed-forward Network (LAFN), the Polarization Feature-Cubic based Multi-head Self-Attention (PFC-MSA), and the full PFCubic Block.

Removing any component leads to a clear performance drop. In particular, eliminating PFC-MSA causes a severe decrease in PSNR/SSIM, indicating that modeling polarization-angle correlation is crucial. Removing the entire PFCubic Block results in the most severe degradation across all metrics, validating its central role in capturing both global polarization cues and local spatial structures. Fig. 6 shows the visual comparisons on our GlassPol dataset. Without WLA, the output becomes globally darker with lower contrast. Removing CRA introduces strong color shifts caused by improper compensation of wavelength-selective degradation. Without LAFN, local textures are noticeably blurred, and checkerboard edges lose sharpness. Removing PFC-MSA produces visible cross-angle inconsistency and ghost-like double edges. Eliminating the full PFCubic Block causes catastrophic failure, severe structural collapse and heavy color distortion. In contrast, our complete PANet generates the sharpest edges, natural brightness, and consistent color reproduction. These observations align well with theranking of PSNR/SSIM/LPIPS in Table 3.

4. Validation of recovered polarization quantities

We add **direct quantitative validation** of recovered polarization fields by reporting PSNR and SSIM on S_0, DoLP, and AoLP for both real-world and synthetic datasets (Table 1), which explicitly measures the error of physically meaningful metrics and supports **physical consistency.**

5. More Detailed Qualitative Comparison

To ensure fairness and reliability, we also added a white balance method [1] to attempt to address the attenuation problem caused by colored glass. As shown in Fig. 7,

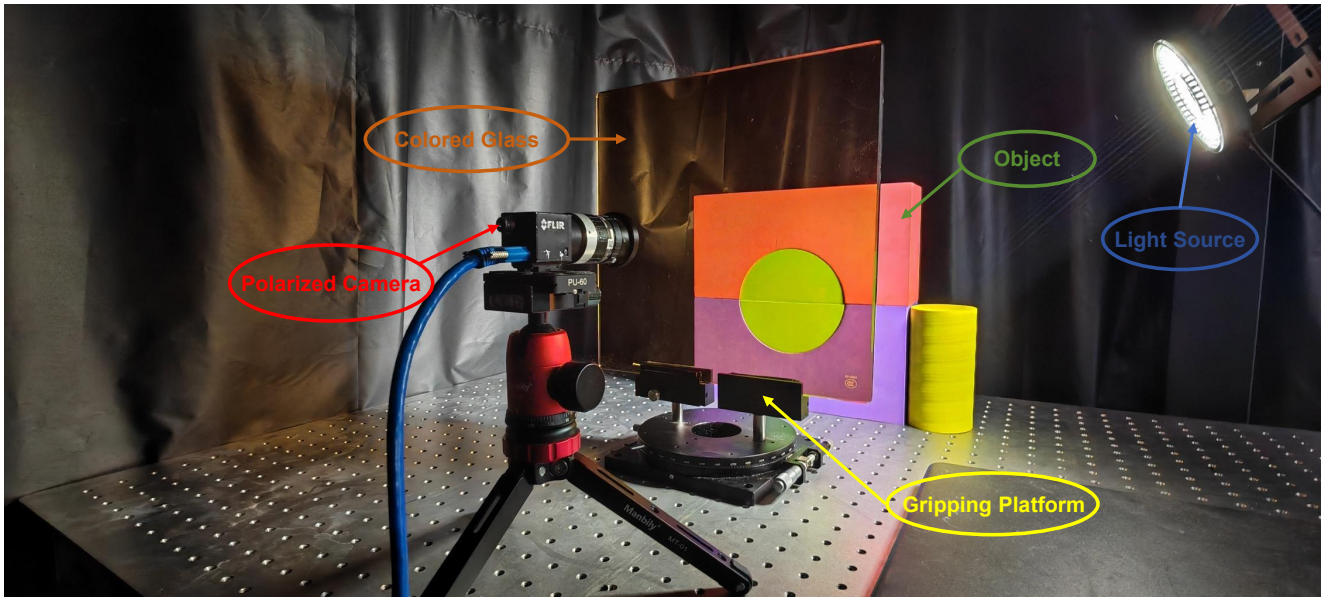

Figure 5. Set up a complete capture system for our colored glass reflection dataset.

Table 1. Results on real-world and synthetic datasets

Dataset	PSNR↑			SSIM↑		
	S_0	DoLP	AoLP	S_0	DoLP	AoLP
Real-world	31.11	25.25	14.16	0.895	0.715	0.532
Synthetic	32.80	27.87	14.88	0.947	0.756	0.652

this approach of applying white balance after reflection process [2, 3] does not effectively solve the problem we identified. As shown in Fig. 9 and Fig. 8, we use the polarization independent intensity S_0 as the input when comparing with RGB-based reflection removal methods. Since S_0 represents the polarization independent total light intensity, it is physically equivalent to the intensity captured by a standard RGB camera. RGB-based reflection removal and enhancement methods operate only on intensity and color, which correspond to S_0 but do not contain any polarization dependent information (S_1, S_2). Therefore, comparing our polarization-based method with RGB methods using S_0 ensures a fair and physically consistent evaluation on the shared intensity component.

To further show the advantages of recovering polarization resolved details, Fig. 10 and Fig. 11 provide full-resolution visualizations on two challenging scenes. Both figures contain strong reflections and heavy wavelength-selective attenuation caused by the colored glass. Despite these difficult conditions, our method preserves fine polarization-dependent structures, restores correct color tones, and effectively suppresses ghost layers. These results confirm that the proposed PSTM-based reconstruction remains effective even under severe colored-glass degradation.

Limitations and exploration. Our model fails under extreme specular highlights, where severe saturation and strong non-Lambertian reflections violate the assumptions of PSTM and make the polarization cues unreliable. As future work, it is promising to combine PSTM with generative priors (e.g., diffusion-based method PolarFree), PSTM can provide a physics-consistent conditioning, while the generative prior can regularize ill-posed regions.

References

[1] Mahmoud Afifi and Michael S. Brown. Deep white-balance editing. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1394–1403, 2020. 4

[2] Zheng Dong, Ke Xu, Yin Yang, Hujun Bao, Weiwei Xu, and Rynson W.H. Lau. Location-aware single image reflection removal. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4997–5006, 2021. 5

[3] Xin Feng, Wenjie Pei, Zihui Jia, Fanglin Chen, David Zhang, and Guangming Lu. Deep-masking generative network: A unified framework for background restoration from superimposed images. *IEEE Transactions on Image Processing*, 30: 4867–4882, 2021. 5

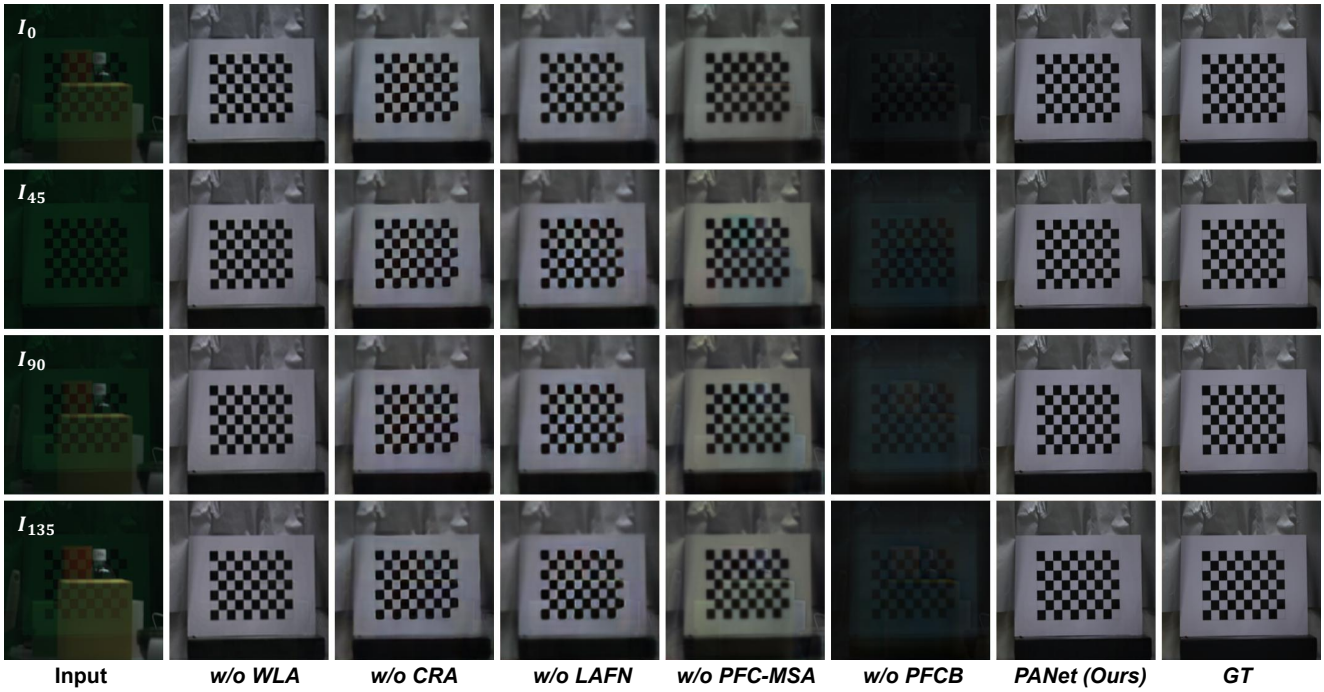

Figure 6. Visual comparisons for the ablation study on the GlassPol dataset. Each column corresponds to an ablated variant of PANet. PANet (Ours) produces the sharpest edges and the most stable color reproduction.

Figure 7. Comparison of applying a digital white-balance correction after reflection removal versus our PSTM-based method on colored-glass scenes. Although white balance slightly reduces color bias, it fails to compensate for the strong attenuation introduced by colored glass.

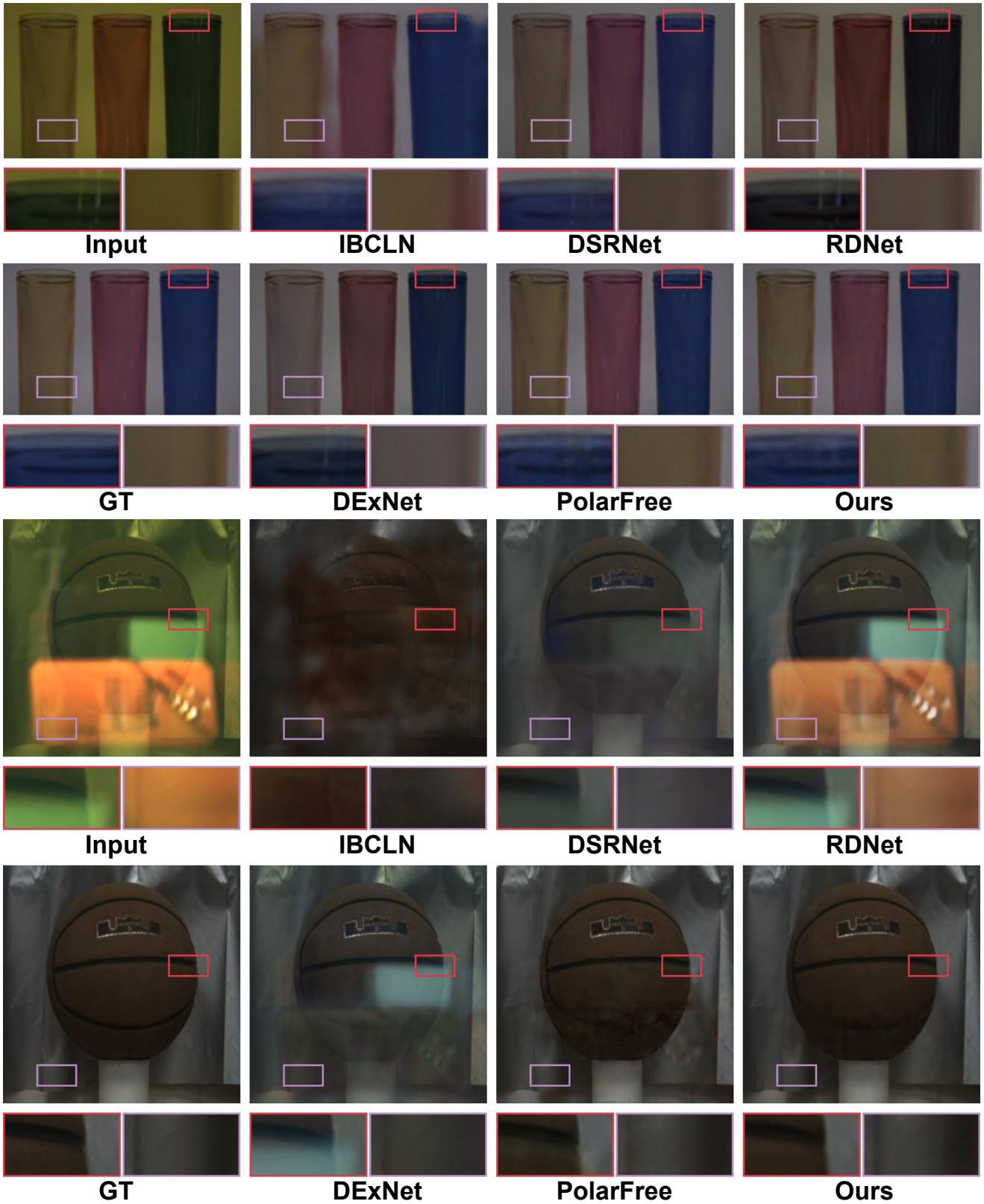

Figure 8. Additional visualization results.

Figure 9. Additional visualization results

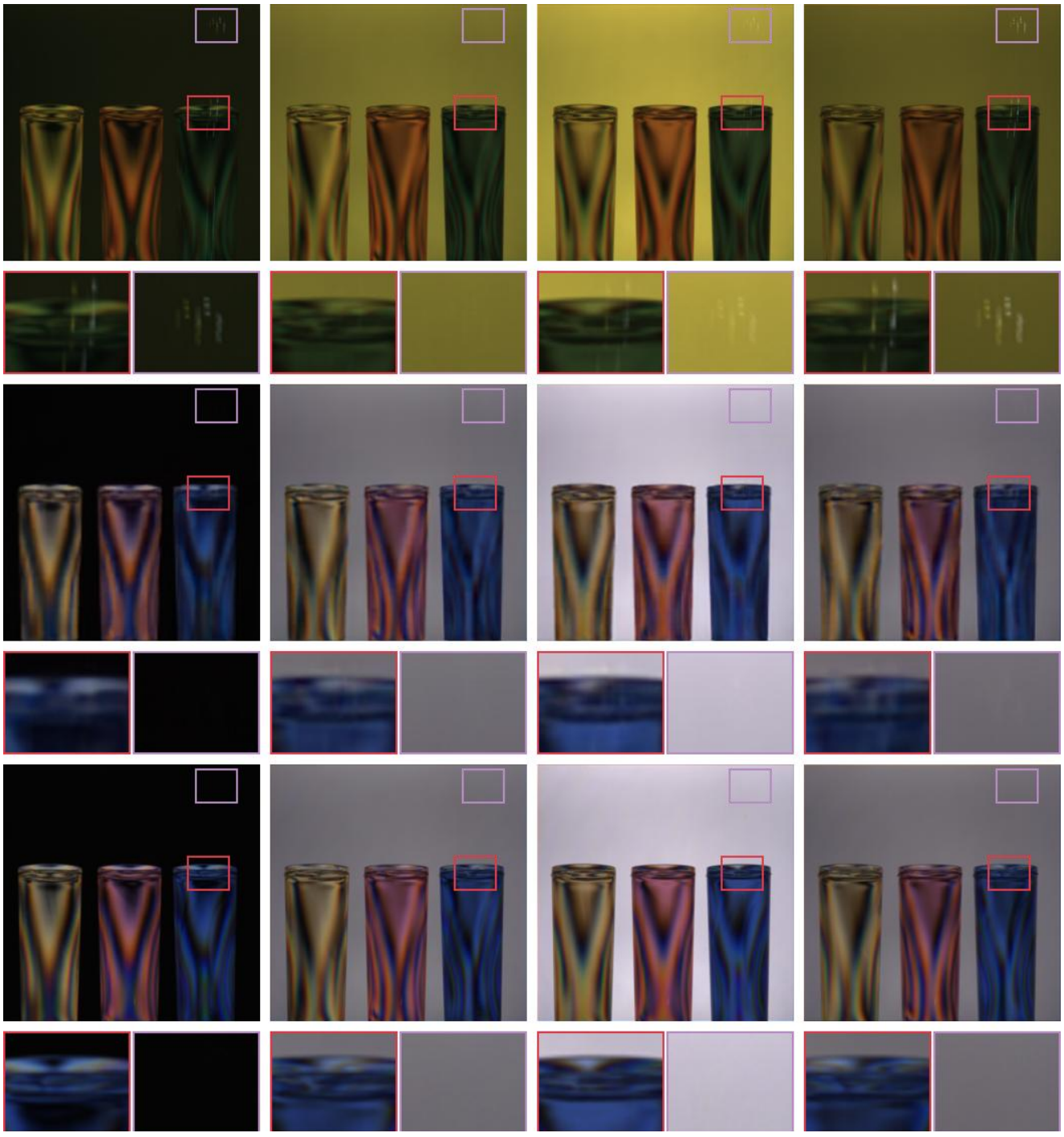

Figure 10. Full visualization of polarization resolved reconstruction on a challenging glass scene. The scene contains strong reflections and wavelength-selective absorption introduced by the colored glass. Our method recovers fine polarization-dependent details and correct color appearance while effectively suppressing double reflections and ghosting artifacts.

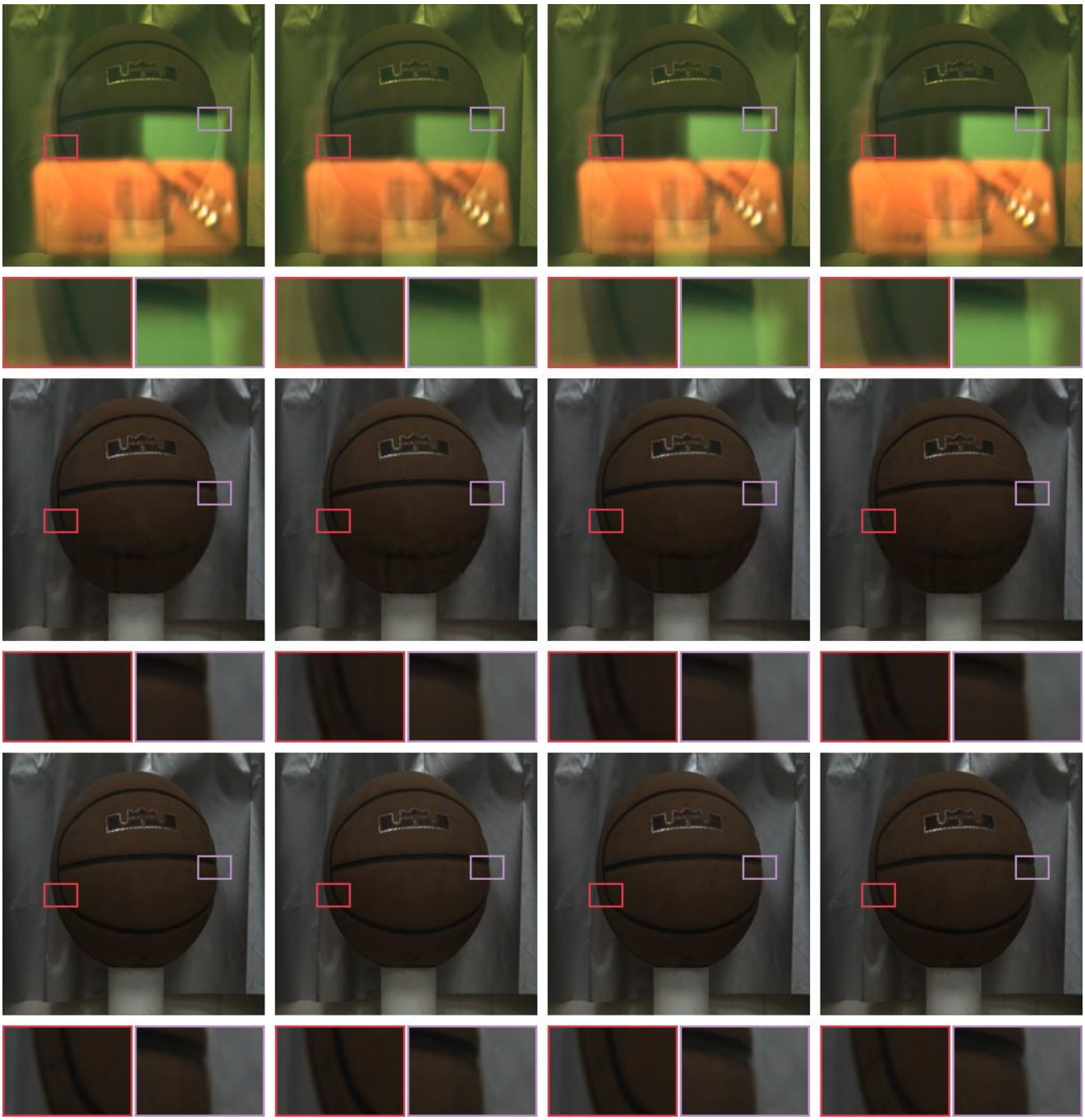

Figure 11. Additional polarization-detail visualization on another complex scene. Although the input observations suffer from strong reflection interference and heavy color distortion due to multi-path propagation inside the glass, our method successfully restores consistent intensity and polarization cues.