PolarFree: Polarization-based Reflection-free Imaging ——Supplementary Material——

Mingde Yao[1], Menglu Wang[3], King-Man Tam[1,4], Lingen Li[1], Tianfan Xue[1,2], Jinwei Gu[1]
[1]The Chinese University of Hong Kong, [2]Shanghai AI Laboratory
[3]University of Science and Technology of China, [4]Institute of Science Tokyo

The supplementary material is organized as follows:

• Section 1 provides details of dataset, including hardware configuration, unpolarized RGB image acquisition, reflection estimation, and alignment.

• Section 2 provides details of experiments, including implementation details, number of parameters, and running time.

• Section 3 provides additional results, including training on synthetic datasets, comparisons with more baselines, and more visual results.

• Section 4 provides further discussions, particularly addressing the limitations and failure cases.

1. Dataset Details

In this section, we provide a detailed overview of the PolaRGB dataset, including the hardware configuration, the process of capturing RGB images, and the method for estimating reflection images.

1.1. Hardware configuration

To capture high-quality polarization data, we employed the BFS-U3-51S5PC-C USB 3.1 Blackfly®S Polarized Color Camera from Edmund Optics. This camera offers adjustable aperture and focal length, enabling precise control over image quality. For stability, the camera was mounted on a tripod during all captures. Instead of fixing the camera parameters, we use the automatical setting to ensure clear and sharp images under varying conditions, aligning with real-world practices and challenges.

For reflective surfaces, we chose acrylic panels with a high transmittance rate of 95%. This design minimizes secondary reflections caused by material thickness while maintaining a realistic representation of reflection scenarios. The effectiveness of this setup is evident in our real-world reflection removal results, particularly in museum and gallery environments, where complex reflections often occur.

We randomly position the semi-reflective glass plate in front of the camera, with the distance ranging from 5 cm to 100 cm and the angle randomly adjusted.

1.2. Obtain RGB image

We use a division-of-focal-plane (DoFP) polarization camera to capture polarized images. By combining four angles (0°, 45°, 90°, 135°), we generate an approximate RGB image I_0, to mimic the output of a standard RGB camera, avoiding the need for alignment with RGB images.

1.3. Reflection estimation

We provide our process for estimating reflection images and summarize it in Algorithm 1. Examples of estimated reflection are shown in Fig. 1.

Given a transmission image T and a mixed image M, our goal is to estimate the reflection image R_{best}, which represents the portion of the image containing the reflected layer, excluding the transmission components. The key to our approach is iterating over a range of weight factors and optimizing the estimated reflection based on feature matching and edge detection. The detailed steps are as follows:

Iterative Weight Adjustment. We begin by defining a weight factor j that ranges from 0.6 to 1.2 emperically, with a step size $\Delta j = 0.02$. In our work, the parameter j is empirically set between 0.6 and 1.2 to better recover the structure of residual reflection, especially in cases where the standard weight of 1 does not adequately capture the reflection's structure.

For each value of j, we compute a reflection candidate R^j using the formula $R^j = |M - j \cdot T|$. This step allows us to explore different weightings of the transmission and mixed images to uncover the best possible reflection estimate.

Edge Detection. We apply Canny edge detection to both the reflection candidate R^j and the transmission image T to reduce noise and minimize color variations. This step highlights key structural features, making the feature matching process more robust and accurate.

Feature Matching. We employ the ORB (Oriented FAST and Rotated BRIEF) feature detector to extract keypoints and descriptors from the edge-detected images. The BFMatcher (Brute Force Matcher) is then used to find the best matches between the reflection candidate R^j_{edges} and the transmission image T_{edges}. The number of matches serves as an indicator of the quality of the match, with fewer matches signaling a better estimate of the reflection.

Final Output. After performing feature matching for each weight factor j, we select the reflection image R^j with the lowest number of accurate matches. This reflection image is considered the best estimate, as it is the one that deviates the most from the transmission image, capturing the true reflection. The selected reflection is then set as R_{best}.

This method provides a robust solution to the challenging problem of reflection estimation by leveraging iterative weight adjustment, edge detection, and feature matching. By working directly in the raw image domain, we are able to avoid the potential pitfalls of image processing artifacts like demosaicing and ISP effects.

Algorithm 1: Reflection Estimation Algorithm

Input: Transmission image T, mixed image M, weight range $j \in [0.6, 1.2]$, step size Δj
Output: Optimized reflection images R_{best}
// Initialize best match score
best_matches $\leftarrow \infty$;
// Loop over weight factor j
For each $j \in [0.6, 1.2]$ *with step* Δj
  // Compute reflection image R^j
  $R^j \leftarrow |M - j \cdot T|$;
  // Apply Canny edge detection
  $R^j_{\text{edges}}, T_{\text{edges}} \leftarrow \text{Canny}(R^j), \text{Canny}(T)$;
  // Feature matching using ORB and BFMatcher
  $desR \leftarrow$ ORB_detect_and_compute(R^j_{edges}) ;
  $desT \leftarrow$ ORB_detect_and_compute(T_{edges}) ;
  matches $\leftarrow$ BFMatcher($desR$, $desT$) ;
  // Get the number of matches
  num_matches $\leftarrow$ len(matches) ;
  // Update best match if current is better
  if num_matches $<$ best_matches **then**
    best_matches $\leftarrow$ num_matches ;
    $R_{\text{best}} \leftarrow R^j$;
Save R_{best} to output.

1.4. Alignement

We employ three key techniques to ensure high-quality alignment of transmission and mixed images. First, during the capture phase, we carefully minimize motion from both the camera and the scene, effectively reducing background blur and preserving image clarity. Second, after the capture process, we conduct a manual selection step to filter out images with noticeable motion blur, ensuring that only the sharpest and most stable images are retained. Finally, we apply a robust algorithmic alignment to address image shifts caused by glass refraction. This alignment operates directly on raw Bayer sub-images by estimating homography and applying affine transformations, resulting in precise and consistent alignment. These techniques together enable us to produce well-aligned image pairs critical for downstream tasks.

Mixed image Transmission Estimated reflection

Figure 1. Visualization of estimated reflection. We also show the captured mixed and transmission images.

2. Experimental Details

2.1. Detail implementations

In our framework, we utilize a prior generation model alongside a transformer-based reflection removal model. For the prior generation, we adopt the standard DDPM [2], conditioning it on polarization information. For the reflection removal component, we employ the widely used Restormer [7] as the backbone and modify its input channels to accommodate polarization-based inputs. This design ensures effective integration of polarization information across both stages.

To ensure a fair comparison, we use only the transmission layer as the supervision signal for all algorithms, without relying on the reflection layer. This is because the reflection layer is estimated, and there may be some errors in this estimation process. Using the estimated reflection layer as a supervision signal could introduce inaccurate labels, which would affect the model training process. Additionally, the estimation of the reflection layer depends on multiple factors, such as polarization information and other prior assumptions, which may vary across different methods, further complicating the fairness of the comparison. Therefore, to eliminate these potential confounding factors, we chose to use only the transmission layer as the sole supervision signal.

2.2. Parameters and running time

The overall runtime of our model is measured at 0.9148 seconds on a single GTX4090Ti with a spatial resolution 1224 $\times$ 1024, demonstrating its efficiency in handling complex tasks. Specifically, the diffusion sampling process for prior generation is completed in just 0.0698 seconds. Meanwhile, the reflection removal stage, which involves more intensive processing, takes 0.8450 seconds. As observed, even at a resolution of 1k, our model maintains a fast processing speed. Our design is flexible, allowing for the replacement of the reflection removal backbone. In the future, we will explore more lightweight approaches to further accelerate the model's performance.

In terms of computational complexity, the diffusion component, using a U-Net, has 2.62M parameters and a computational cost of 0.0423 GFLOPs (at a 128 $\times$ 128 patch size). The reflection removal stage, based on Restormer [7], has 30.03M parameters and a computational cost of 72.47 GFLOPs (also at 128 $\times$ 128 patch size).

3. More Results

In this section, we present additional results, including training on synthetic datasets, comparisons with a broader range of methods, and more qualitative comparisons.

3.1. Training with synthetic dataset.

To demonstrate the effectiveness of our proposed real-world dataset, we train the reflection removal methods with synthetic polarization dataset used in ReflectNet [6]. The dataset simulates polarized images with reflections by modeling HDR processes, introducing displacement perturbations, and simulating glass curvature. Using the official implementation of ReflectNet, we synthesize data on the Places dataset.

We train both the Kim *et al.* [4] and the ours model using this synthetic dataset and evaluate performance on our captured real-world dataset. Table 1 presents a quantitative comparison of the results, highlighting the significant performance gap between models trained on synthetic versus real-world data. Our findings show that synthetic data alone cannot fully model intricate reflections and interactions, reinforcing the need for real-world data to develop robust reflection removal methods.

Table 1. Quantitative comparison of reflection removal performance using models trained on synthetic and real-world data.

Method	Training Data	PSNR↑	SSIM↑	LPIPS↓	Q-align↑
Kim *et al.* [4]	Synthetic [6]	16.32	0.6210	0.4974	3.5271
	PolaRGB (Ours)	**20.67**	**0.8399**	**0.2714**	**3.7148**
PolarFree (Ours)	Synthetic [6]	17.80	0.7039	0.3395	3.6387
	PolaRGB (Ours)	**22.44**	**0.8681**	**0.1325**	**3.8867**

3.2. Comparison with more methods

To further evaluate the effectiveness of our method, we compare it against several traditional reflection removal approaches on our real-world dataset. The compared methods include Fraid *et al.* [1], Schechner *et al.* [5], and ReflectNet [6].

For a fair comparison, we use the standard input format required by each method. Specifically, the methods proposed by Schechner *et al.* [5] and ReflectNet [6] leverage polarized images as input, utilizing polarization information to enhance their performance in separating reflection and transmission layers. In contrast, the method by Fraid *et al.* [1] operates solely on RGB images without the benefit of polarization cues.

Table 2 presents the quantitative results of this comparison. Our method demonstrates superior performance across all metrics, significantly outperforming the other approaches. This highlights the effectiveness of our model in handling real-world complexities and leveraging both the polarization information and our robust backbone design for reflection removal.

The results also emphasize the limitations of traditional methods, particularly those relying solely on RGB inputs, in capturing and separating intricate reflection patterns present in real-world scenes. This comparison further underscores the necessity of advanced techniques and real-world datasets to achieve state-of-the-art performance in reflection removal.

Table 2. Quantitative comparison of our method with traditional reflection removal approaches on the real-world dataset.

Method	PSNR↑	SSIM↑	LPIPS↓	Q-align↑
Fraid *et al.* [1]	15.84	0.7514	0.2597	3.5508
Schechner *et al.* [5]	8.06	0.1176	0.8742	1.8262
ReflectNet *et al.* [6]	<u>19.11</u>	<u>0.8334</u>	<u>0.2113</u>	<u>3.8027</u>
PolarFree (Ours)	**22.44**	**0.8681**	**0.1325**	**3.8867**

3.3. Additional visual comparisons of final results

We present additional visualization results on polarization-based reflection removal in Fig. 2–4, with Kim *et al.* [4], DSRNet [3], and RDRNet [8]. These examples demonstrate the effectiveness of our method across a variety of challenging scenarios, including scenes with strong specular highlights, complex backgrounds, and intricate reflection patterns caused by non-uniform glass surfaces.

In Fig. 5, we present additional reflection removal results in museum and gallery settings, which represent an important application scenario for this task. From the results, it is evident that our method consistently outperforms existing approaches, achieving a cleaner separation of reflection and transmission layers. In contrast, other methods face challenges in preserving fine details when dealing with strong reflections and often struggle to handle scenes with highly textured or cluttered backgrounds. These comparisons highlight the robustness of our model in addressing the complexities of real-world applications, particularly in scenarios where accurate reflection removal is critical.

4. Further Discussions

Here, we discuss some of the limitations of our work.

Currently, our dataset generation relies on fixed glass sheets to produce reflections. However, real-world reflections are often more complex, involving multi-layer reflections (which can cause ghosting), defocus, and color variations. While our dataset covers a range of scenarios, the types of glass sheets used remain limited. In future work, we plan to include a wider variety of glass types to better handle more complex reflection types.

Additionally, our method currently integrates polarization information in a relatively straightforward manner. We aim to explore more efficient ways to leverage polarization data, such as embedding the imaging principles of polarization directly into the network (e.g., similar to the unfold network approach). Moreover, the inclusion of polarization information is orthogonal to many existing reflection removal priors based on intensity (such as defocus and ghosting). Future research will explore how to jointly utilize these priors for improved reflection removal.

In Fig. 6, we showcase several particularly challenging scenarios where reflections cause significant information loss and color distortion. Existing models fail in these situations, as reflected by their poor performance. Although our model still has some limitations (e.g., poor recovery of highlights and slight color shifts), it outperforms previous methods by recovering more information. Moving forward, we are considering the use of diffusion models or large language models (LLMs) to better remove reflections and reconstruct the transmission layer.

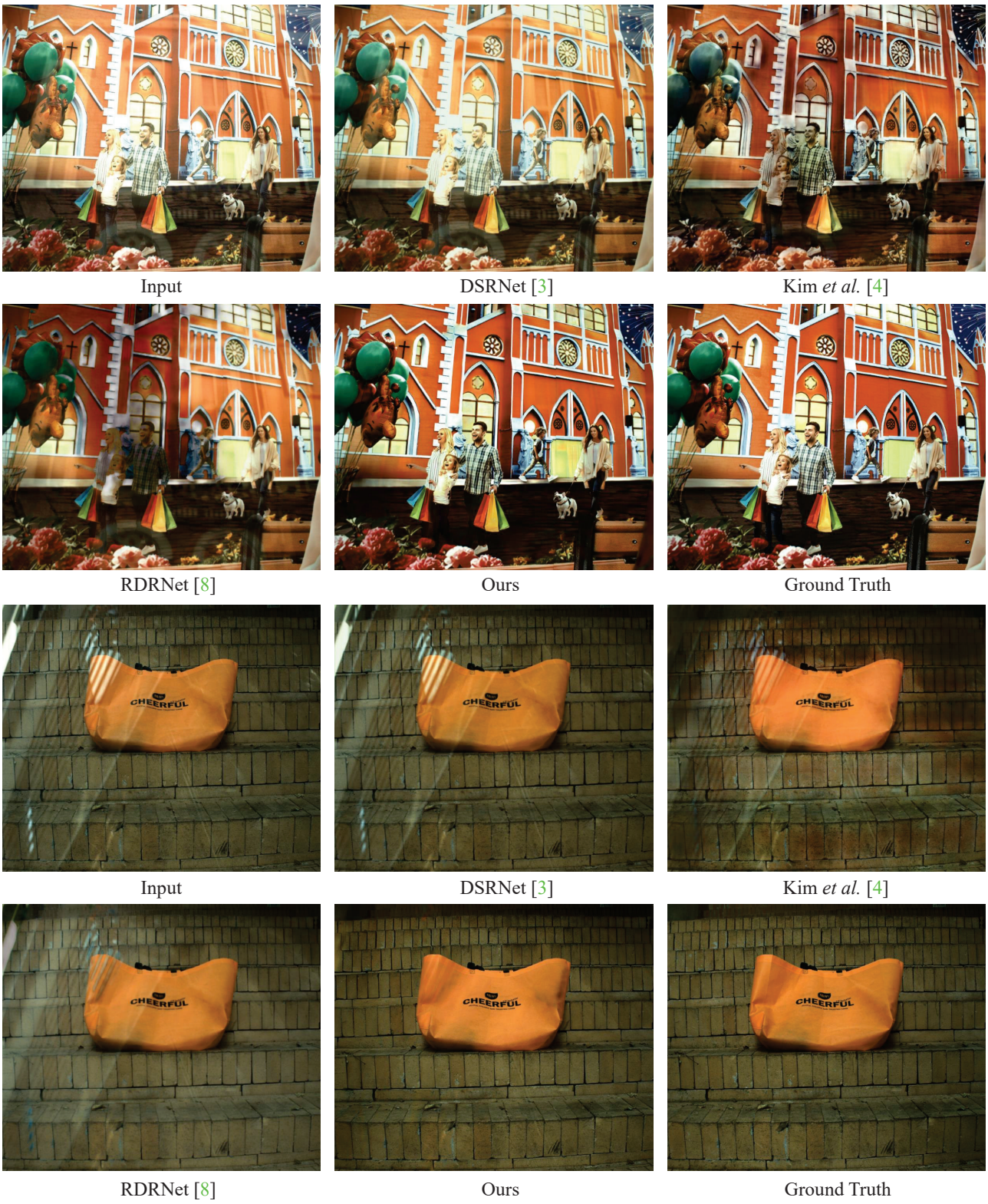


Figure 2. Qualitative comparisons on the PolaRGB dataset.

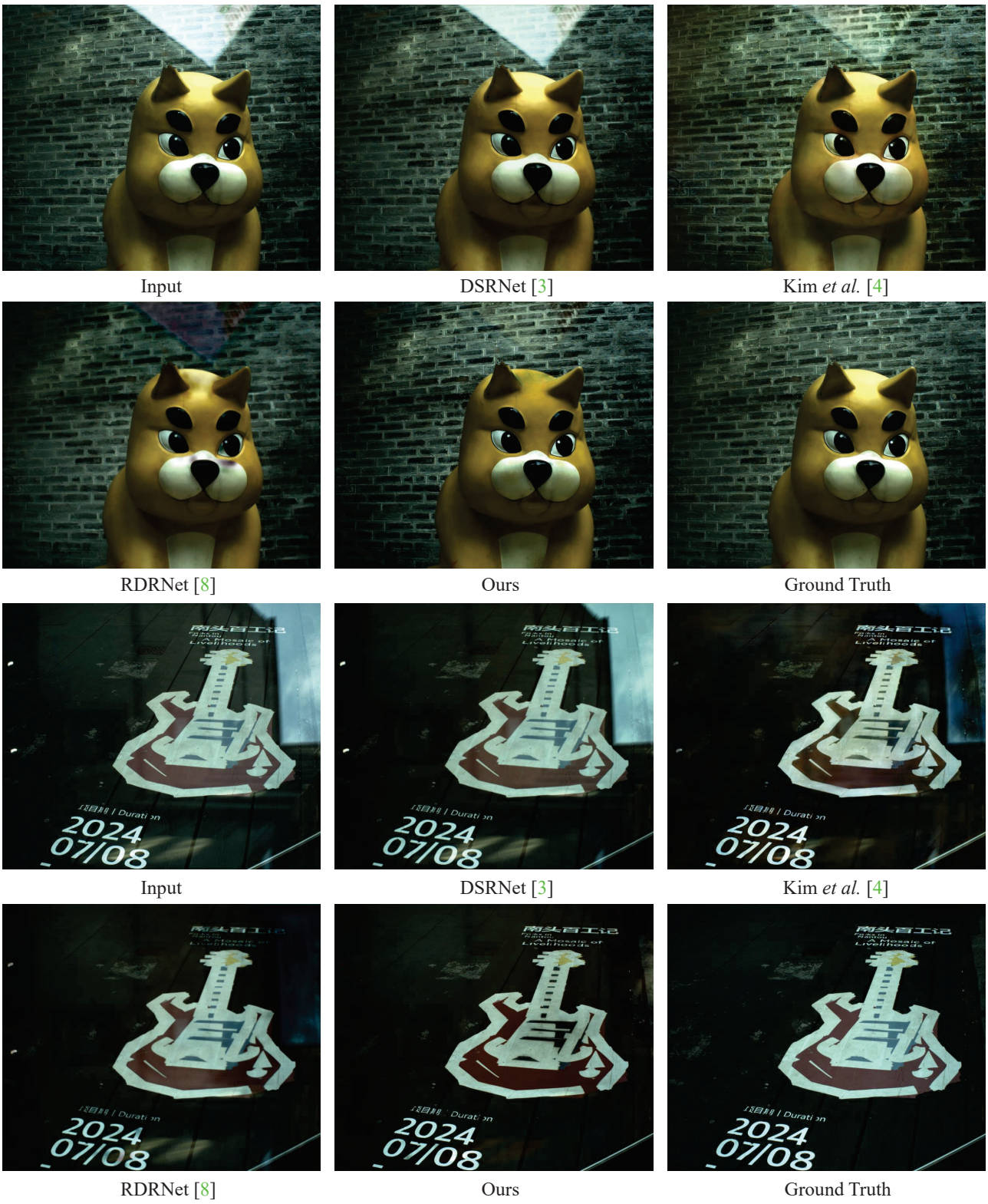


Figure 3. Qualitative comparisons on the PolaRGB dataset.

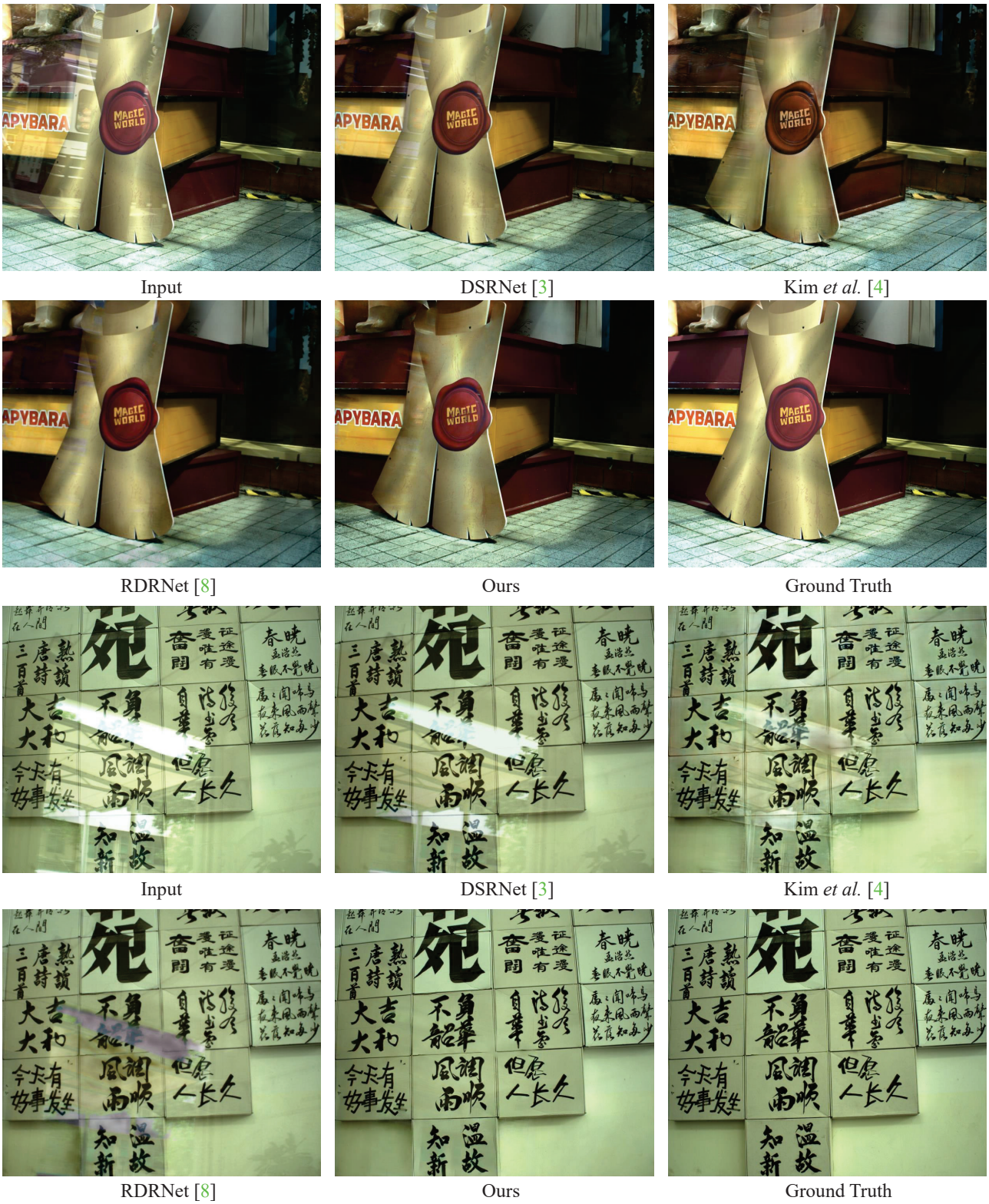
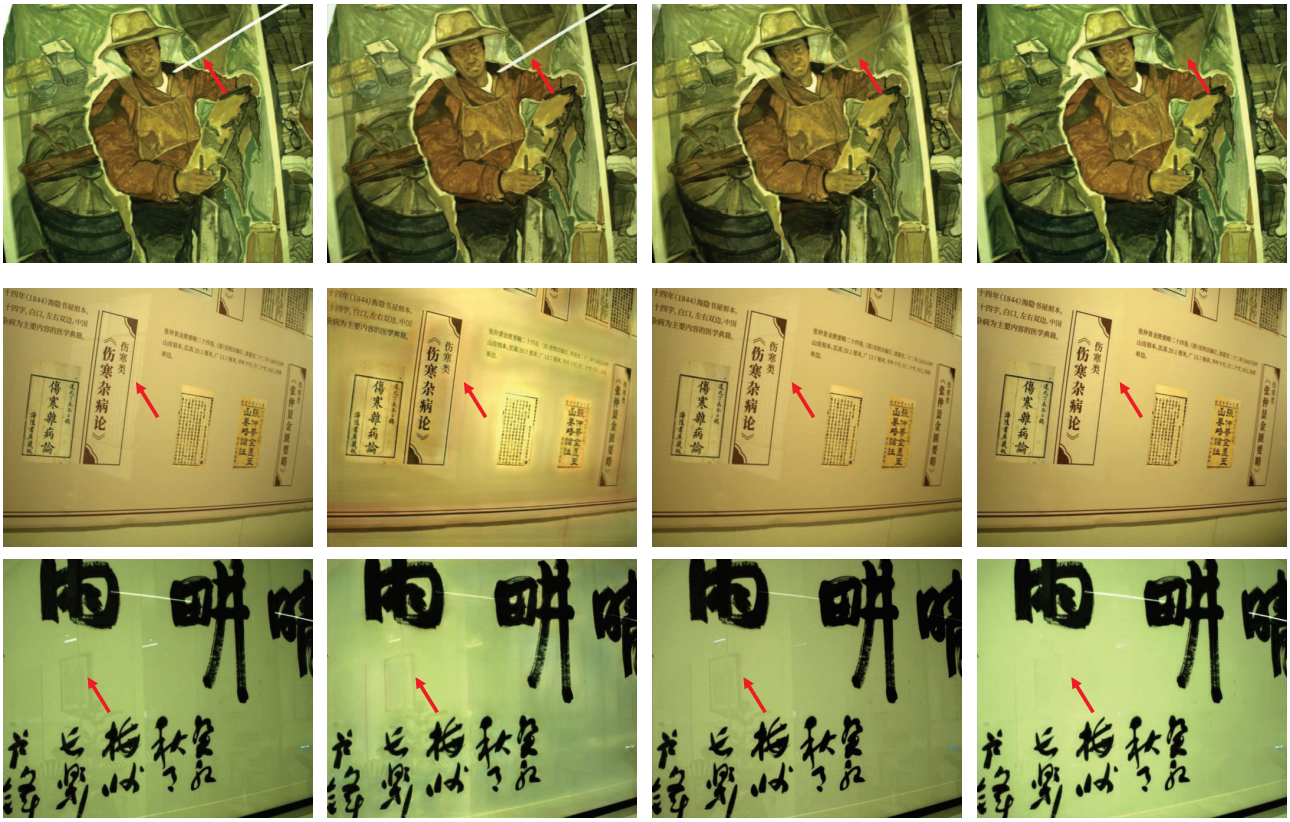
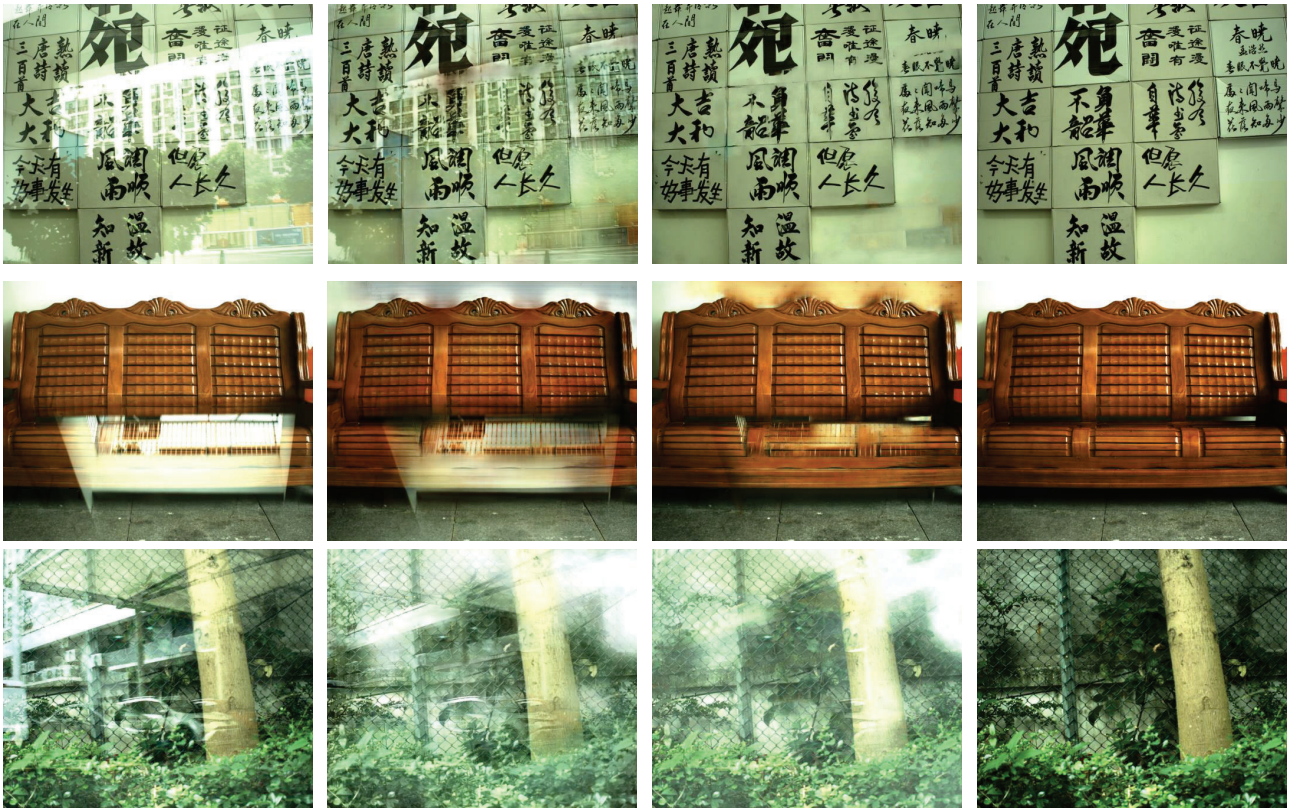


Figure 4. Qualitative comparisons on the PolaRGB dataset.



Input | Kim *et al.* [4] | DSRNet [3] | Ours

Figure 5. Reflection removal results in museum and gallery scenarios without ground truth. Our method can handle highlights and intricate reflections.



Input | Kim *et al.* [4] | Ours | Ground Truth

Figure 6. Extremely challenging reflection removal scenarios, where reflections cause significant information loss and color distortion. Existing models fail in these situations, while our method recovers more details, despite some remaining imperfections in highlight restoration and slight color shifts.

References

[1] 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. 4, 5

[2] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 4

[3] 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. 5

[4] 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. 4, 5

[5] Yoav Y Schechner, Joseph Shamir, and Nahum Kiryati. Polarization and statistical analysis of scenes containing a semireflector. *JOSA A*, 17(2):276–284, 2000. 4, 5

[6] 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. 4, 5

[7] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022. 4

[8] 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. 5