Supplementary Material

Revisiting Single Image Reflection Removal In the Wild

Yurui Zhu [1, 2], Xueyang Fu [1,] , Peng-Tao Jiang [2],
Hao Zhang [2], Qibin Sun [1], Jinwei Chen [2], Zheng-Jun Zha [1], Bo Li [2,]
[1] University of Science and Technology of China [2] vivo Mobile Communication Co., Ltd
zyr@mail.ustc.edu.cn, xyfu@ustc.edu.cn, libra@vivo.com

This supplementary material includes five parts:

- **Part 1** shows more visual comparison results in real-world reflection scenes.
- **Part 2** shows the quantitative comparison results on the subsets of SIR^2 [7].
- **Part 3** shows the model efficiency comparisons of different reflection removal methods.
- **Part 4** shows the data collection process with our proposed pipeline.
- **Part 5** shows the misalignment and artifacts brought by the previous dataset collection pipelines [4–6, 11].

Part 1: More visual vomparison results

In this part, we provide the visual comparisons with FRS [10], IBCLN [6], YTMT [2], LANet [1] and DSRNet [3]. All reflection images in this paper are from real-world reflection scenes. For example, Figure A1 A2 from the testing dataset [6]; Figure A3 A4 A5 A6 A7 A8 from [9],which provides various reflection images.

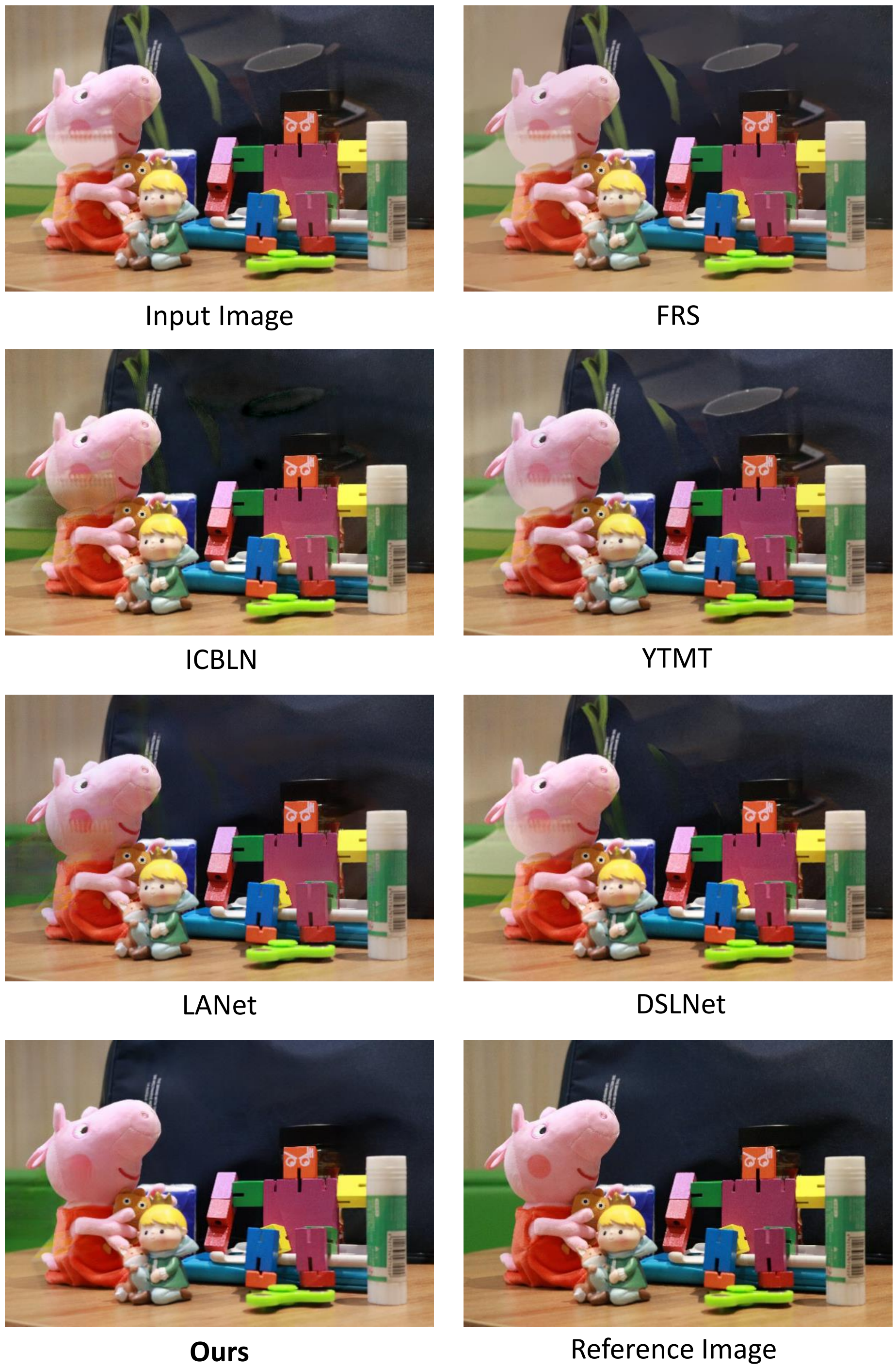

Figure A1. Visual comparisons with SOTA methods on real reflection scenes.

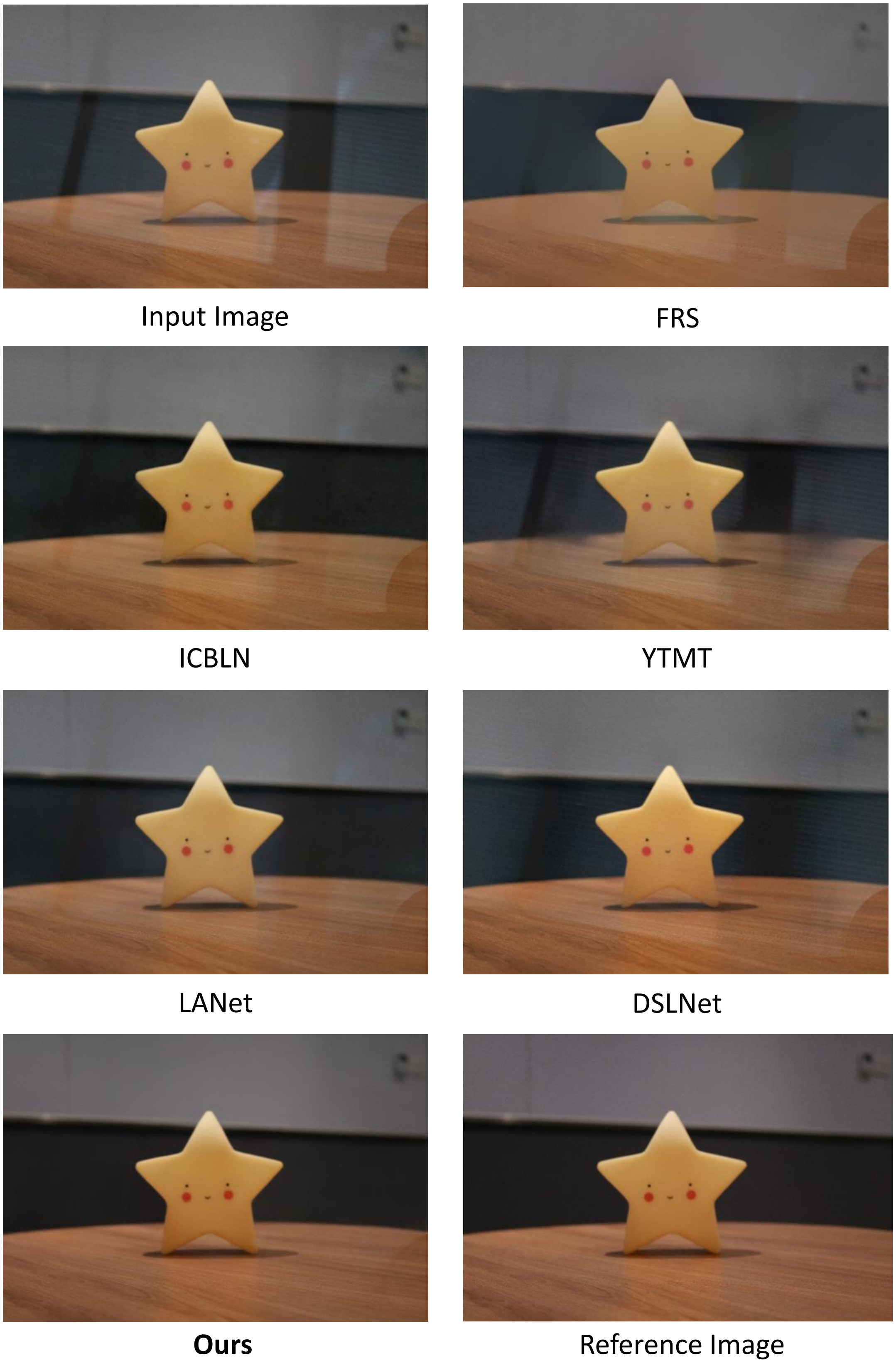

Figure A2. Visual comparisons with SOTA methods on real reflection scenes.

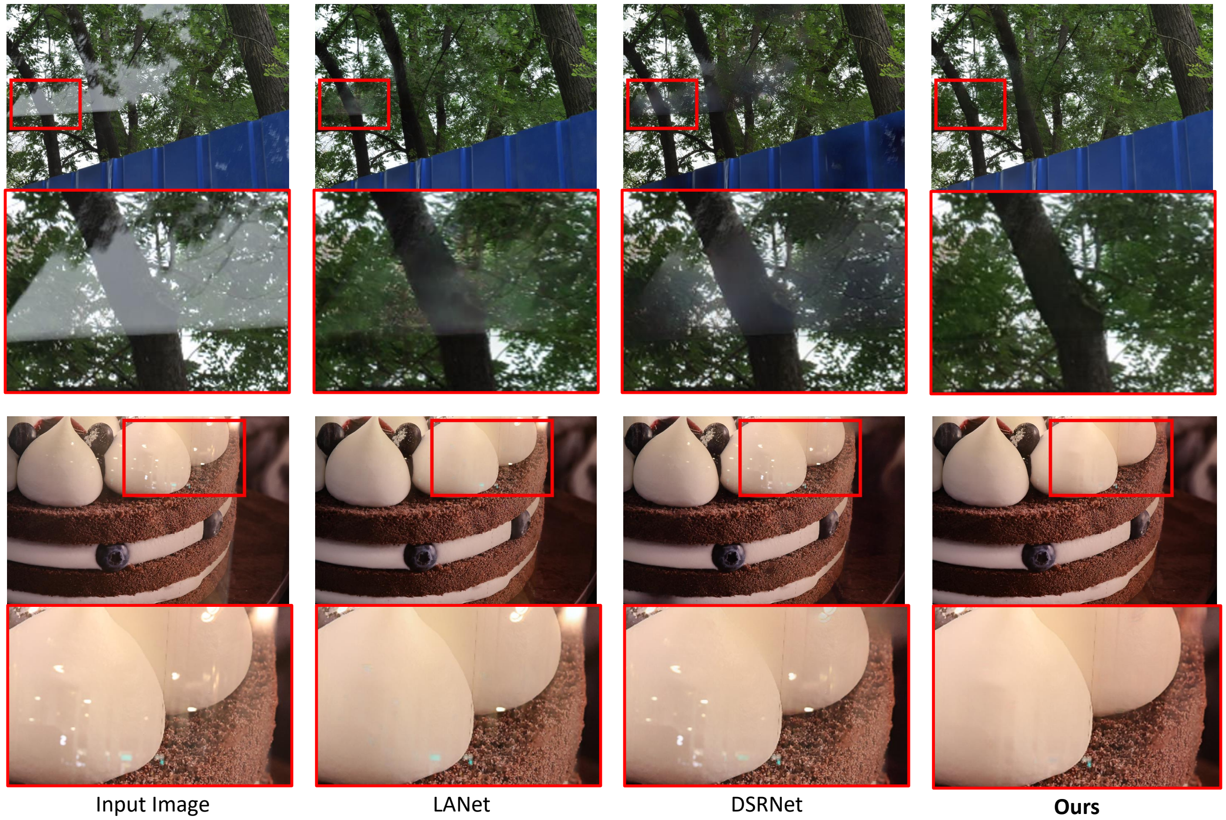

Figure A3. Visual comparisons with SOTA methods on real reflection scenes.

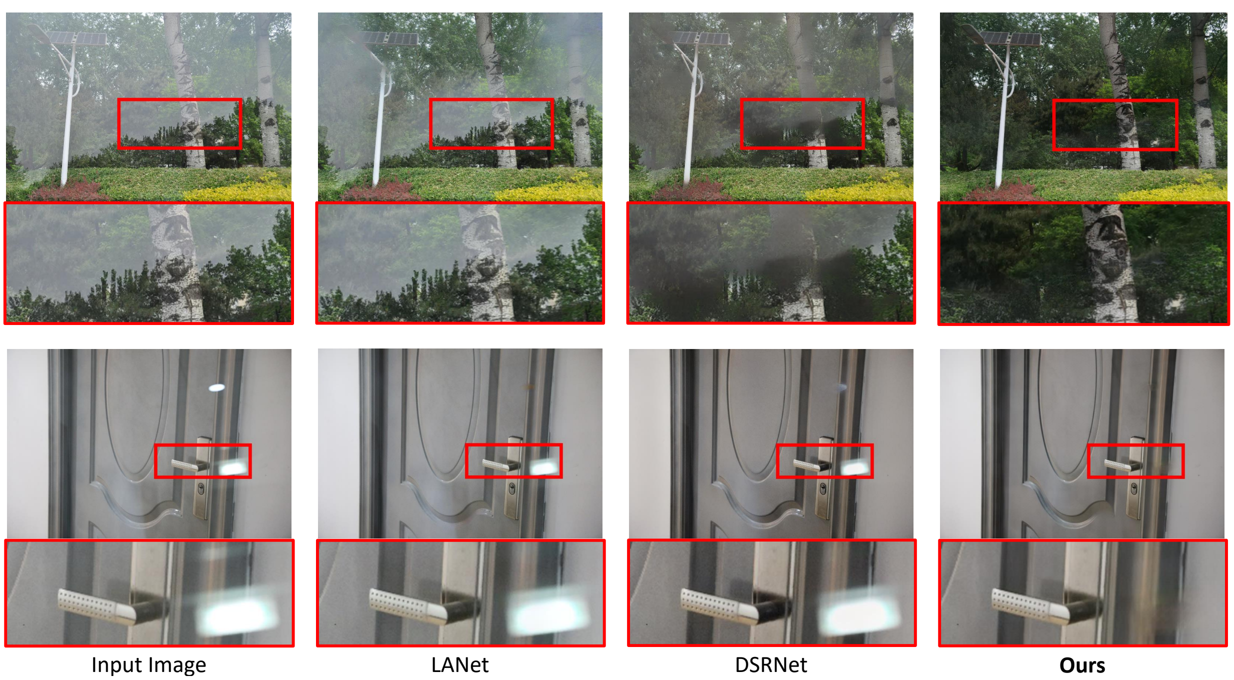

Figure A4. Visual comparisons with SOTA methods on real reflection scenes.

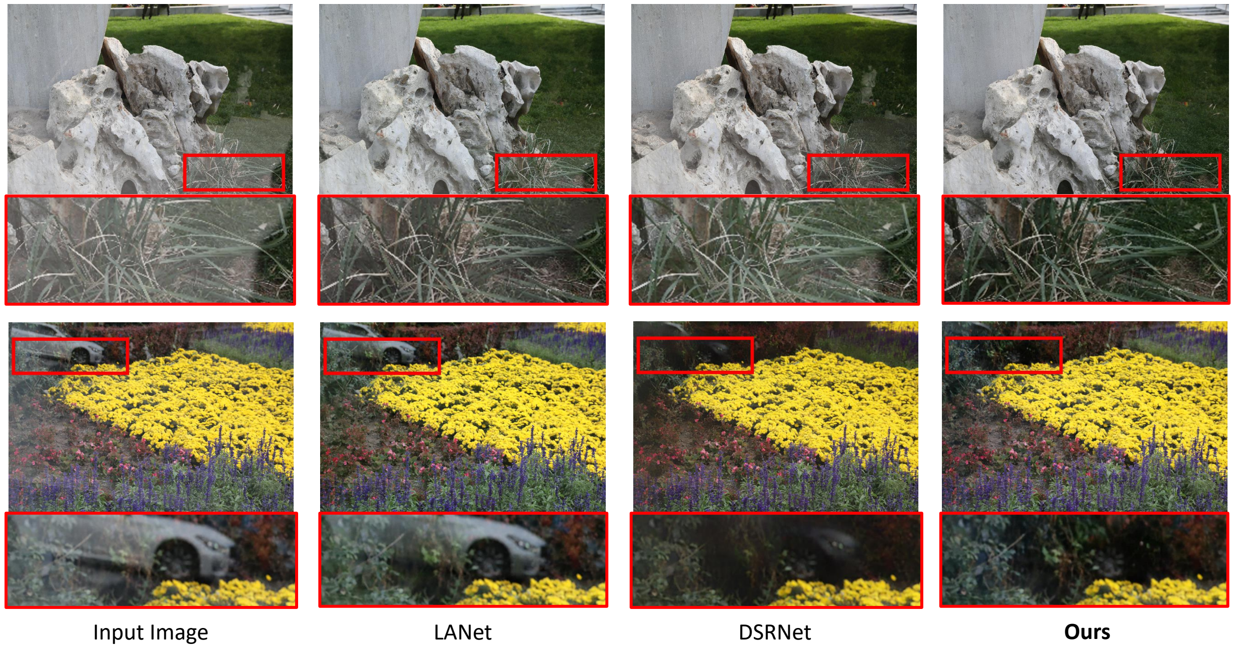

Figure A5. Visual comparisons with SOTA methods on real reflection scenes.

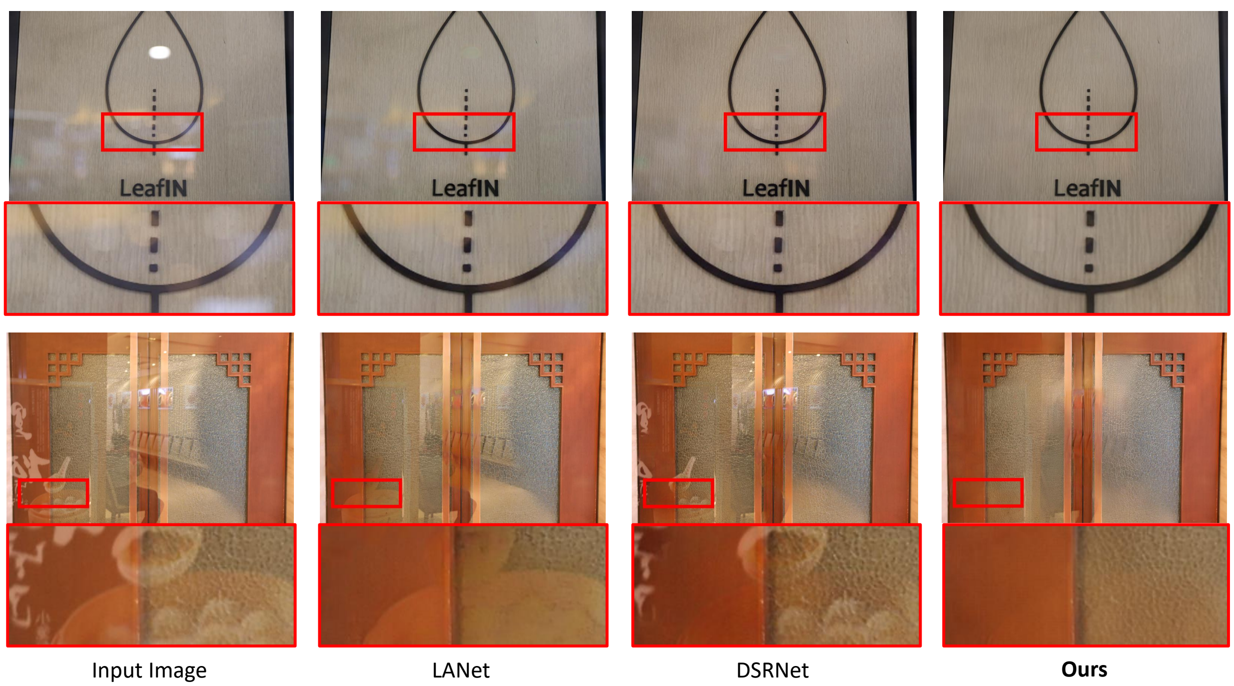

Figure A6. Visual comparisons with SOTA methods on real reflection scenes.

Input Image LANet DSRNet **Ours**

Figure A7. Visual comparisons with SOTA methods on real reflection scenes.

Row 1: The CDR dataset

Row 2: Casual photos

Figure A8. Visual results of ours on real reflection scenes. Left is the reflection image, and right is the estimated result.

Part 2: Quantitative comparison results on the subsets of SIR^2 [7]

SIR^2 [7] are specifically proposed for evaluating the performance of reflection removal in real-world scenes, including three subsets: $wild(55)$, $postcard(199)$ and $solid(200)$. These subsets include a wide range of reflection scenes, collected using different aperture sizes and varying glass thicknesses

Table A1. Quantitative comparisons on the subsets of real reflection benchmark [7]. The best results are in **bold**.

Methods	Venue	$wild(55)$		$postcard(199)$		$solid(200)$	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Input Image	-	25.97	0.90	20.94	0.87	23.68	0.89
FRS [10]	CVPR 20219	23.95	0.87	20.92	0.86	23.04	0.87
ICBLN [6]	CVPR 2020	24.41	0.89	23.25	0.88	24.75	0.89
YTMT [2]	NerIPS 2021	25.23	0.89	2.23	0.88	24.46	0.90
LANet [1]	ICCV 2021	25.89	0.90	21.27	0.89	24.00	0.90
PNACR [8]	ACM MM 2023	25.69	0.90	23.11	0.89	24.73	0.89
DSRNet [3]	ICCV 2023	24.36	0.89	23.96	**0.89**	26.01	0.92
Ours	-	**26.48**	**0.91**	**24.05**	**0.89**	**26.62**	**0.93**

Part 3: The model efficiency comparisons

Table A2. The model efficiency comparisons. FLOPs is calculated based on inputs with a resolution of $256 \times 256 \times 3$.

Methods	ERRNet [9]	YTMT [2]	LANet [1]	DSRNet [3]	Ours
Network Parameters (M: 10^6)	31.52	32.66	10.93	9.84	29.09
FLOPs (G: 10^9)	439.45	179.97	334.5	97.30	17.55

Part 4: Data collection process with our proposed pipeline

In Figure A10, we present a detailed visual illustration of our collection procedure using the proposed pipeline. Specifically, we would like to delineate each step of the collection process for enhanced clarity and understanding.

- *Step 1*: Set up the tripod and camera device in front of the reflective surfaces, like the glass scene shown in Figure A10.
- *Step 2*: Turn on the video recording mode on the camera device and set the camera to manual focus mode.
- *Step 3*: As illustrated in Figure A10(a), on the camera side, use a black cloth to block the reflective lights(including the *global ambient light* and *local object light* in Figure A9), then initiate video recording using a remote controller. Empirically, the frames from the first one or two seconds of the recording can be employed as the transmission image $\mathbf{T}$. At this step, only *background light* enters the camera device.
- *Step 4*: Remove the black cloth. As shown in Figure A10(b), we can capture the reflection images($\mathbf{I}$), and modulate the reflected interference content by manipulating the reflections through the manual obstruction.
- *Step 5*: In addition to the dynamic scenes in the reflective environment, we can also modify the objects within the environment or introduce new objects to adjust the reflection contents. This largely enriches the diversity of the reflection contents and obtain multiple reflection images ($\mathbf{I}_1$, ..., $\mathbf{I}_n$) corresponding to $\mathbf{T}$.
- *Step 6*: The captured videos typically range from 30 seconds to 150 seconds. For post-processing of recorded videos, we commonly select the average of the frames at the beginning as the transmission image. As for the reflection frames, we use a uniform sampling method to obtain reflection images with varying reflection contents. We Empirically sample every 20 frames or 30 frames to effectively avoid redundancy of similar reflection frames.

In contrast, to avoid misalignment and color distortion, prior pipelines [6, 7, 11] have to employ removable, relatively thin, and colorless glass for data collection. Furthermore, methods [4, 5] require data collection in RAW data format. By not imposing these constraints, our proposed pipeline offers a more cost-effective manner of acquiring large-scale aligned reflection pairs, which also is applicable to a variety of reflective surfaces such as building glasses, car glass windows, display glass, framing glass, and self-prepared glass.

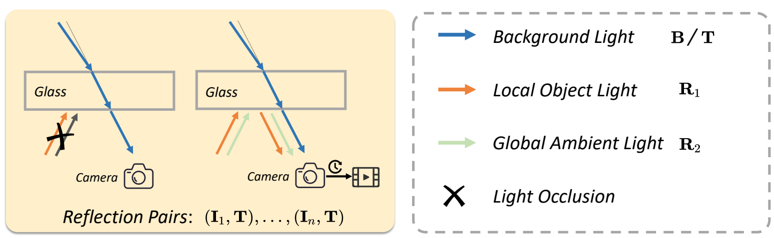

Figure A9. Simplified illustrations of our collection pipeline.($\mathbf{I}_i$: reflection image; $\mathbf{T}$: Transmission Image).

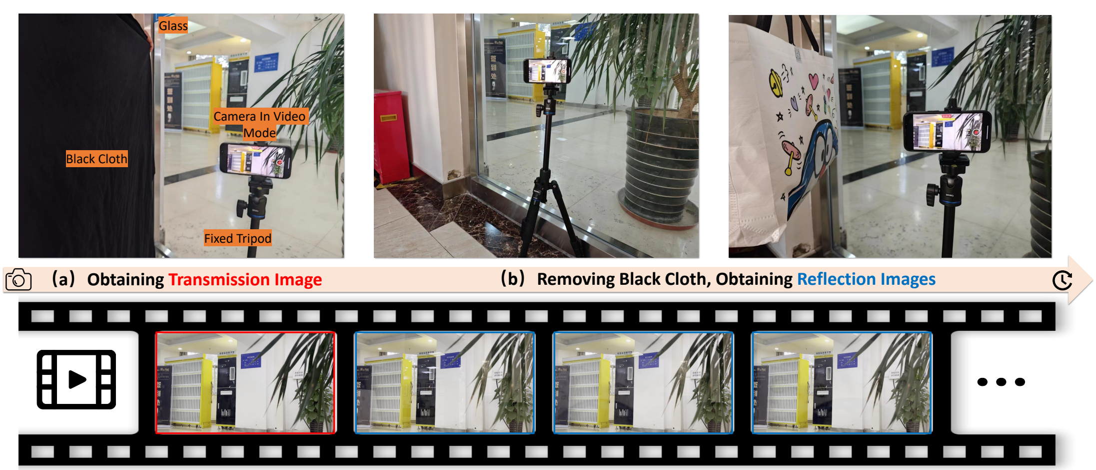

Figure A10. Illustration of our collection process with our proposed pipeline.

Part 5: Misalignment and artifacts brought by the previous pipelines

In our manuscript, we argue that the pipeline [6, 11] may introduce spatial misalignment in the reflection pairs due to glass refraction. To make the results clearer, we attempt to demonstrate this misalignment issue in the gradient domain. This is based on the fact that if there is a misalignment in the reflection image pair, then when subtracting the reflection image pair in the gradient domain, it is easy to produce double-edged results, as verified in Figure A11.

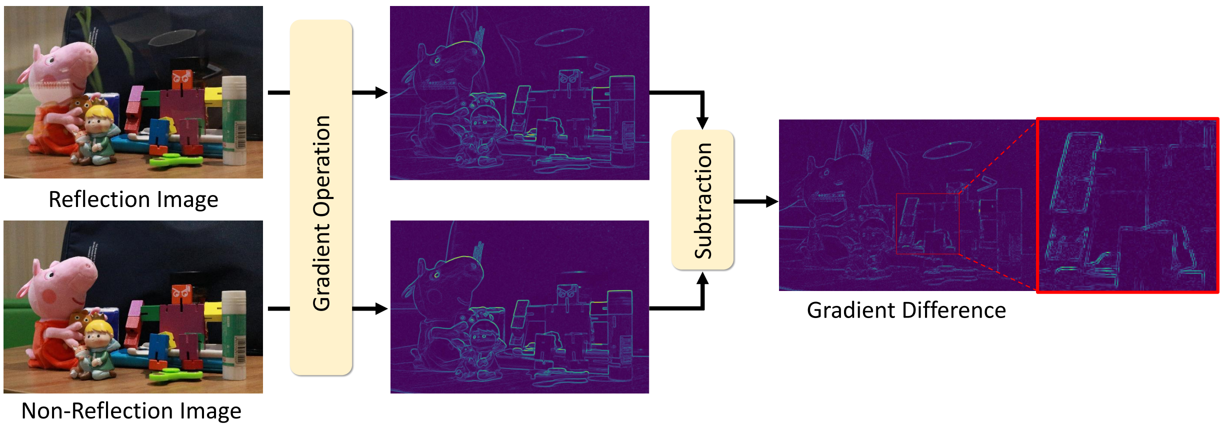

Figure A11. Illustration of misalignment issue in the pipeline [6, 11].

Moreover, we also noticed that their transmission images might leave minor reflection remnants with other data collection pipelines [4, 5]. Here, we provide some visual examples to verify the artifacts brought by their pipeline in Figure A12.

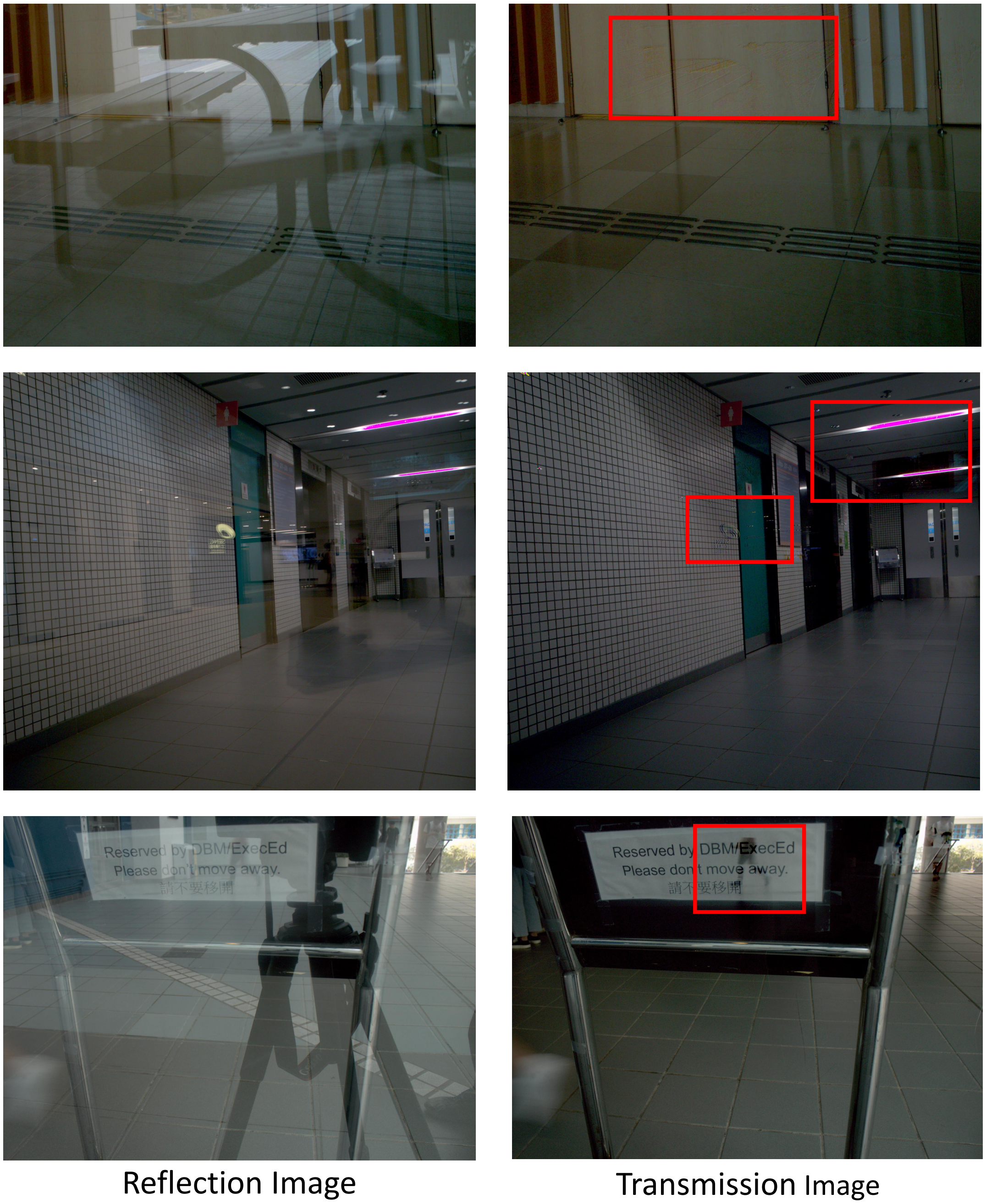

Figure A12. Illustration of the artifact issue in the pipeline [4, 5]. The reflection remnants and artifacts are obvious in the red boxes.

References

[1] Zheng Dong, Ke Xu, Yin Yang, Hujun Bao, Weiwei Xu, and Rynson WH Lau. Location-aware single image reflection removal. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5017–5026, 2021. 2, 7

[2] 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. 2, 7

[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. 2, 7

[4] 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, 8, 9, 10

[5] Chenyang Lei, Xuhua Huang, Chenyang Qi, Yankun Zhao, Wenxiu Sun, Qiong Yan, and Qifeng Chen. A categorized reflection removal dataset with diverse real-world scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3040–3048, 2022. 8, 9, 10

[6] 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, 2, 7, 8, 9

[7] Renjie Wan, Boxin Shi, Ling-Yu Duan, Ah-Hwee Tan, and Alex C Kot. Benchmarking single-image reflection removal algorithms. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3922–3930, 2017. 1, 7, 8

[8] 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. 7

[9] 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. 2, 7

[10] Yang Yang, Wenye Ma, Yin Zheng, Jian-Feng Cai, and Weiyu Xu. Fast single image reflection suppression via convex optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8141–8149, 2019. 2, 7

[11] 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. 1, 8, 9