Flash-Splat: 3D Reflection Removal with Flash Cues and Gaussian Splats – Supplement

In this supplementary material, we show results on three additional scenes (Sec. 1), describe each scene's setup (Sec. 2), conduct more comparison experiments (Sec. 3), report quantitative performance (Sec. 4), and provide more details of an ablation study (Sec. 5). Additionally, our project webpage shows the **rendered videos** of the transmitted and reflected 3D scenes separated by our proposed method, which outperforms the separation by NeRFReN [3].

1 Additional Scenes

In addition to the 4 scenes we presented in Section 5, we show results on three additional scenes. We evaluate our proposed method and the baselines on these scenes in the same way as described in Section 4. As shown in Figure 15, 16, 17, our proposed method significantly outperforms the baselines in terms of reflection transmission separation.

2 Scene Descriptions

In cases of strong specular reflections, such as the scenes in our captured dataset, it is challenging even for humans to identify which objects belong to the transmitted scene, and which belong to the reflected scene. Therefore, to help readers understand the setup of our scenes, we briefly describe the transmitted and reflected scenes for each of our 7 scenes (4 in the main paper, 3 in the supplement).

- **Figure 5, Office** (main paper). The transmitted scene is a bookcase in an office with a glass wall. We set up our camera in the corridor facing inside the office. The reflected scene is a study area at the end of the corridor.
- **Figure 6, Game Controller** (main paper). The transmitted scene is a game controller in a black case with a glass cover. Note that the glass surface is horizontal to the ground. The reflected scene is a door with a glass window (you can also see the corridor through the door's window). The door is upside down due to reflection.
- **Figure 7, Lens Stage** (main paper). The transmitted scene is a lens stage with several lenses on it. The lens stage is covered with a glass case. The reflected scene consists of tables and chairs.
- **Figure 8, Outdoor Scene** (main paper). We took photos of a toy and a power bank inside a glass window from outdoors. The reflected scene includes some bags on an outdoor table, with plants and another building's windows (mildly defocused) 20 meters away in the background.
- **Figure 15, Shelf** (supplement). The transmitted scene is a shelf with boxes, bags, and batteries on it. The reflected scene consists of tables and chairs.

– **Figure 16, Poster** (supplement). The transmitted scene is a poster on the wall of a corridor. The reflected scene is the corridor itself.
– **Figure 17, Lab** (supplement). The transmitted scene is a cabinet with some boxes on it and a lamp's pole (black) in front of it. The reflected scene includes a door, a lamp, and a table with various items on it.

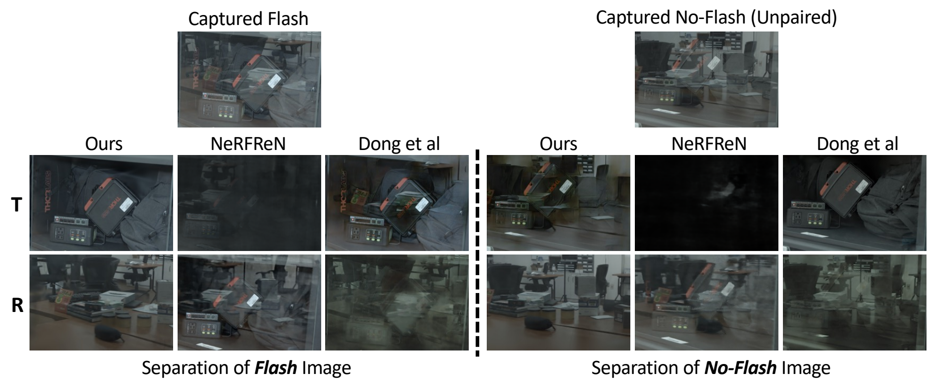

Fig. 15: Comparison with NeRFReN [3] and Dong et al [2] on Shelf scene. Top, middle, and bottom rows are the captured images, separated transmissions, and separated reflections, respectively. Our reflection separation approach is far more effective.

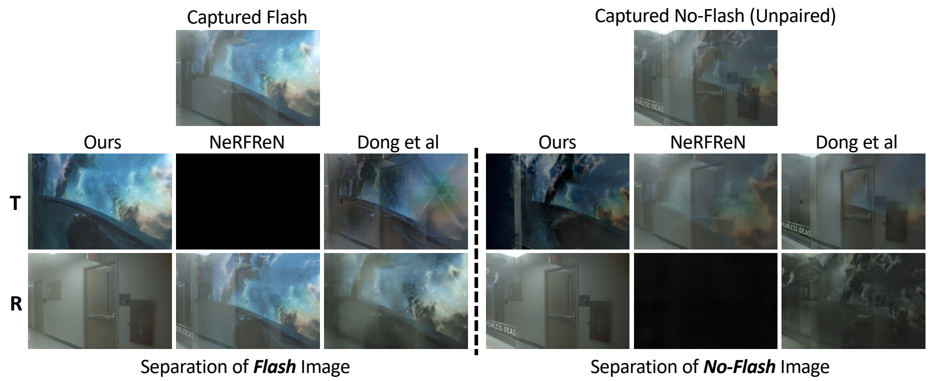

Fig. 16: Comparison with NeRFReN [3] and Dong et al [2] on Poster scene. Top, middle, and bottom rows are the captured images, separated transmissions, and separated reflections, respectively. Our reflection separation approach is far more effective.

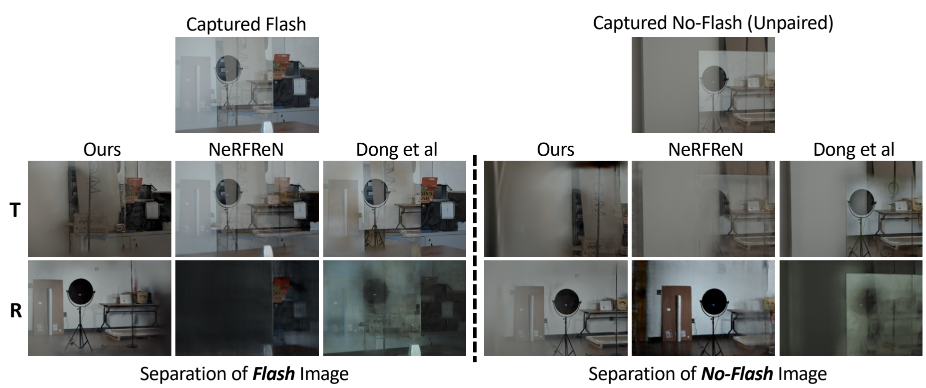

Fig. 17: Comparison with NeRFReN [3] and Dong et al [2] on Lab scene. Top, middle, and bottom rows are the captured images, separated transmissions, and separated reflections, respectively. Our reflection separation approach is far more effective.

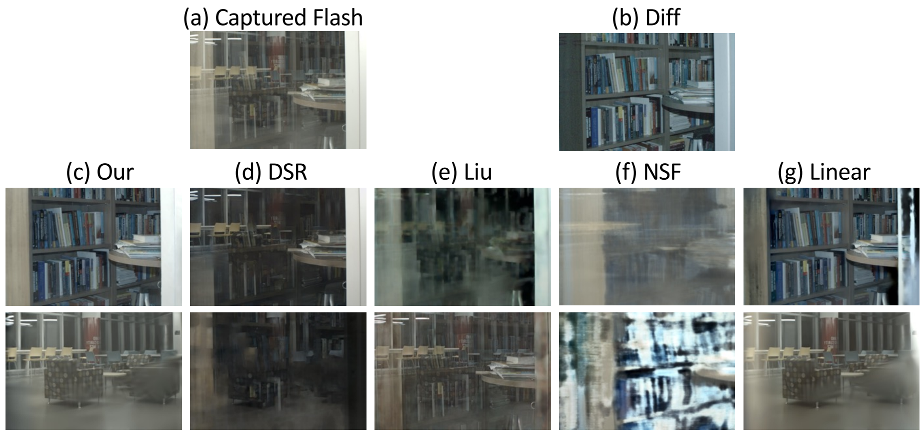

Fig. 18: Additional Comparisons. Our method (c) outperforms (d) DSRNet [4] : ICCV 2023, supervised, single image-based; (e) Liu [5] : CVPR 2020, supervised, burst-based; and (f) Neural Spline Fields (NSF) [1]: CVPR 2024, unsupervised, burst-based. Our reconstructed transmission is close to (b), the paired flash/no-flash difference (Diff), which requires paired images captured with a tripod. Additionally, in (g), we trained a pure linear representation by enforcing $T_F = cT_N$. This model results in imperfect reflection separation (notice the right side) compared to our soft-linear system with the Pearson loss.

3 Additional Comparisons

We conduct 4 more comparison experiments. As shown in Figure 18, our method (c) also outperforms (d) DSRNet [4] : ICCV 2023, supervised, single image-based; (e) Liu [5] : CVPR 2020, supervised, burst-based; and (f) Neural Spline Fields (NSF) [1]: CVPR 2024, unsupervised, burst-based. Note that none of these methods can take advantage of our unpaired flash/no-flash data. Our reconstructed transmission is close to (b), the paired flash/no-flash difference (Diff), which requires paired images captured with a tripod.

Additionally, as shown in Figure 18 (g), we trained a pure linear representation by enforcing $T_F = cT_N$. This model results in imperfect reflection separation (notice the right side) compared to our soft-linear system with the Pearson loss, since the relationship between T_F and T_N is not perfectly linear.

4 Quantitative Evaluation

To compute quantitative metrics, we need to have a ground truth transmission scene as a reference. While it is difficult (oftentimes impossible) to remove the glass from a scene, we can instead compute the paired flash/no-flash difference as the reference transmission scene. In Table 1, we report the averaged PSNR and LPIPS between the difference image and each method's separated transmission scene. We find that our method performs the best.

Metric	Methods				
	DSR [4]	**Liu** [5]	**NSF** [1]	**NeRFReN** [3]	**Ours**
PSNR ↑	13.02	11.16	9.40	10.09	**20.42**
LPIPS ↓	0.5754	0.6765	0.7452	0.7153	**0.2868**

Table 1: Averaged Quantitative Evaluations. We calculate the PSNR and LPIPS between each method's separated transmissions and the paired flash/no-flash differences, which serve as references for the ground truth transmissions. Our method has a huge quantitative advantage over the other methods, which corresponds with our huge qualitative advantage shown in the visual comparisons in Figure 5-8, 15-17. Granted, our method's transmission is not perfect as it exhibits a slightly different color tone compared to the difference image, e.g., Figure 18 (b, c). Nevertheless, our result successfully obtains structural information that is very close to the reference image, outperforming other methods by a large margin.

5 Details of the Ablation Study in Section 6.1

In Section 6.1 of the main paper, we design and test a flashless framework, where we remove the flash cues from our proposed framework and keep everything else the same. Figure 19 shows the detailed architecture of this flashless framework.

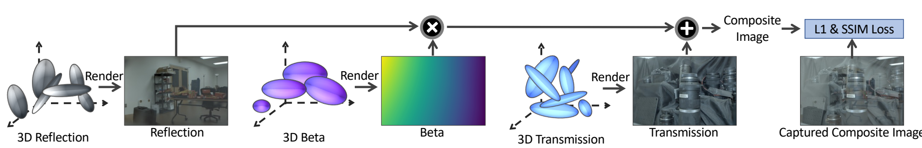

Fig. 19: Ablation: the flashless version of our proposed framework. To demonstrate the importance of flash cues, we design an ablation study where we remove the flash cues from our proposed framework. This "flashless" framework is still a 3DGS-based approach, but does not utilize flash/no-flash photography at all. It uses 3 3DGSs to represent the reflected scene, the transmitted scene, and the reflection factor β. The loss is calculated between the captured images and images rendered from these 3 3DGSs. More descriptions can be found in Section 6.1 in the main paper.

References

1. Chugunov, I., Shustin, D., Yan, R., Lei, C., Heide, F.: Neural spline fields for burst image fusion and layer separation. CVPR (2024) 4, 5
2. Dong, Z., Xu, K., Yang, Y., Bao, H., Xu, W., Lau, R.W.: Location-aware single image reflection removal. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 5017–5026 (2021) 2, 3

3. Guo, Y.C., Kang, D., Bao, L., He, Y., Zhang, S.H.: Nerfren: Neural radiance fields with reflections. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18409–18418 (2022) 1, 2, 3, 5
4. Hu, Q., Guo, X.: Single image reflection separation via component synergy. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 13138–13147 (2023) 4, 5
5. Liu, Y.L., Lai, W.S., Yang, M.H., Chuang, Y.Y., Huang, J.B.: Learning to see through obstructions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14215–14224 (2020) 4, 5