Multi-view Dynamic Reflection Prior for Video Glass Surface Detection

Anonymous submission

In this supplement, we initially embarked on ablation experiments to conduct different hyperparameter configurations. In addition, we have unveiled a more expansive array of examples derived from the dataset, accompanied by more visualized results for comparison.

Additional Ablation Study

Within the context of the Location-aware Reflection Extraction (LRE) module, we conducted an ablation experiment to scrutinize the effect of varying the number of deformable attention blocks. As shown in Table 1, it was discerned that the performance was suboptimal with a singular block, demonstrating the least efficacy. Conversely, an incremental elevation in the number of blocks progressively enhanced performance. Remarkably, the optimization reached its peak with four blocks, beyond which the addition of further blocks (such as increasing the number to 6) ceased to contribute to any discernible augmentation in performance.

#Blocks	IoU↑	Accuracy↑	BER↓	MAE↓
1	0.795	0.896	10.16	0.103
2	0.798	0.897	9.91	0.101
4 (Ours)	**0.802**	**0.899**	**9.54**	**0.099**
6	0.799	0.899	9.78	0.100

Table 1: Ablation analysis of the number of MDA blocks.

In Table 2, the results reveal that a single Temporal Transformer Block (TTB) and a single Spatial Transformer Block (STB) exhibit the best performance. Intriguingly, an escalation in the number of blocks (such as an increase to two) does not culminate in a perceptible performance enhancement, indicating a saturation point in the efficiency gains achievable through this parameter.

In the original manuscript, the sequence is such that TTB precedes STB (as represented in Table 3 by "TTB+STB"). In this study, we opted to invert the order of these two blocks (denoted in Table 3 as "STB+TTB"). The results of this alteration demonstrate that the permutation of the sequence does not substantially influence the performance, signifying the robustness of our model.

#Blocks	IoU↑	Accuracy↑	BER↓	MAE↓
1 (Ours)	**0.802**	**0.899**	**9.54**	**0.099**
2	0.798	0.898	9.94	0.102

Table 2: Ablation analysis of the number of TTB and STB blocks.

Models	IoU↑	Accuracy↑	BER↓	MAE↓
STB+TTB	0.801	0.900	9.60	0.099
TTB+STB (Ours)	**0.802**	**0.899**	**9.54**	**0.099**

Table 3: Ablation analysis of the order of TTB and STB blocks.

Additional Visual Results

In this section, additional insights are provided into our VGSD-D dataset, along with the comparison with state-of-the-art techniques. Fig. 1 displays more examples of daily scenes. Meanwhile, Fig. 2 illustrates how our method stands against leading approaches in the following areas: salient object detection (MINet (Pang et al. 2020)), video salient object detection (UFO (Su et al. 2023)), semantic segmentation (SAM (Kirillov et al. 2023)), video shadow detection (SC-Cor (Ding et al. 2022)), video mirror detection (VMD (Lin, Tan, and Lau 2023)), glass detection (GlassNet (Lin, He, and Lau 2021), PGSNet (Yu et al. 2022)).

In the second scene of visual results of Fig. 2, a particularly challenging case is illustrated, where the boundaries of the glass are not distinctly delineated, rendering the identification of the corresponding glass area based solely on the edges unfeasible. Our methodology adeptly capitalizes on the reflection information from the glass surfaces across various images within the video sequence, yielding more accurate glass surface segmentation results.

References

Ding, X.; Yang, J.; Hu, X.; and Li, X. 2022. Learning shadow correspondence for video shadow detection. In *ECCV*, 705–722.

Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.;

Figure 1: Visual display of several proposed Video Glass Surface Detection dataset examples.

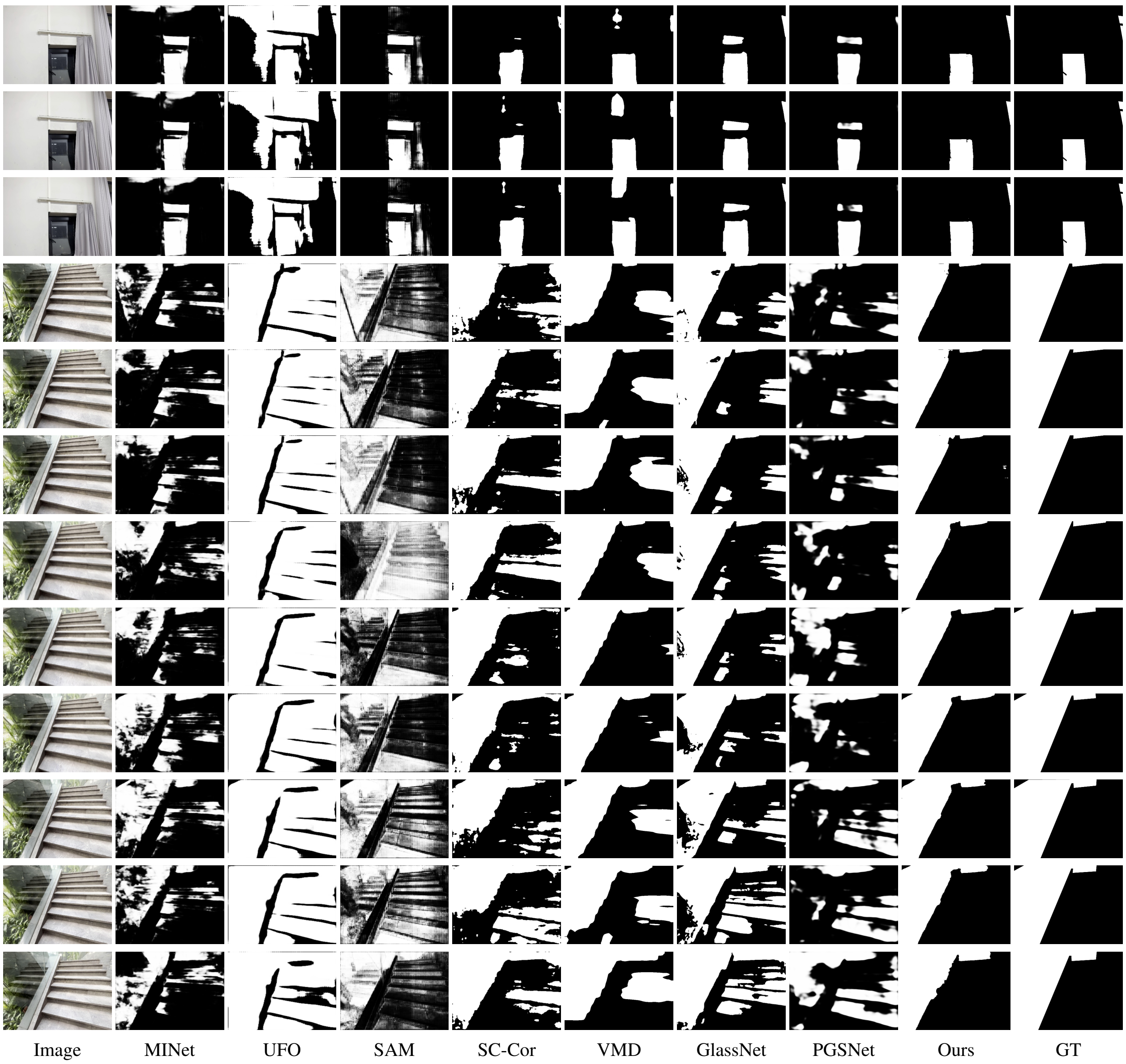

Figure 2: Qualitative comparison of selected state-of-the-art methods and our approach.

Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A. C.; Lo, W.-Y.; et al. 2023. Segment anything. In *ICCV*.

Lin, J.; He, Z.; and Lau, R. W. 2021. Rich context aggregation with reflection prior for glass surface detection. In *CVPR*, 13415–13424.

Lin, J.; Tan, X.; and Lau, R. W. 2023. Learning To Detect Mirrors From Videos via Dual Correspondences. In *CVPR*, 9109–9118.

Pang, Y.; Zhao, X.; Zhang, L.; and Lu, H. 2020. Multi-scale interactive network for salient object detection. In *CVPR*, 9413–9422.

Su, Y.; Deng, J.; Sun, R.; Lin, G.; Su, H.; and Wu, Q. 2023. A unified transformer framework for group-based segmentation: Co-segmentation, co-saliency detection and video salient object detection. *IEEE TMM*.

Yu, L.; Mei, H.; Dong, W.; Wei, Z.; Zhu, L.; Wang, Y.; and Yang, X. 2022. Progressive glass segmentation. *IEEE TIP*, 31: 2920–2933.