Supplementary Materials: Reginal Attention For Shadow Removal

Anonymous Authors

1 MORE VISUAL RESULTS

In this section, we bring more visual results and comparisons with state-of-the-art methods on the ISTD+ dataset [7] (Please see Fig. 1, Fig. 2) and SRD dataset [8] (Please see Fig. 3).

2 DISCUSSION ON LOSSES

The loss function of our model turns out to be:

$$\mathcal{L} = \alpha_1 \mathcal{L}_{per} + \alpha_2 \mathcal{L}_{cont}, \tag{1}$$

where $\mathcal{L}_{cont}$ is Charbonnier Loss [6], and $\mathcal{L}_{per}$ is VGG perceptual loss. To verify the necessity of the loss terms, we conducted ablation studies on the ISTD+ dataset [7] by removing $\mathcal{L}_{vgg}$ term. As depicted in Tab. 1, the VGG perceptual loss boosts the performance by 0.38 dB PSNR in shadow region and 0.21 dB PSNR in all image.

Variant	Shadow Region (S)			All Image (ALL)		
	PSNR↑	SSIM↑	RMSE↓	PSNR↑	SSIM↑	RMSE↓
$\mathcal{L}_{cont}$	40.35	0.992	4.86	35.95	0.976	2.59
$\mathcal{L}_{cont} + \mathcal{L}_{vgg}$	**40.73**	**0.993**	**4.41**	**36.16**	**0.976**	**2.53**

Table 1: Ablation Experiments for losses with ISTD+ [7] dataset.

3 VISUALIZATION OF REGIONAL ATTENTION

To validate whether our proposed regional attention mechanism truly enables shadow areas to interact with their adjacent non-shadow areas, we selected several points on the image and visualized their attention weight allocation. As shown in Fig. 4, we can see that in completely illuminated areas or shadow areas, the attention weights of these points are relatively low and even, while points in shadows have a much larger attention weight when the attention area can encompass the surrounding non-shadow areas. Moreover, the second case and the third case of Fig. 4 show that not all non-shadowed regions are assigned high attention weights. Rather, non-shadowed region information which is similar to the selected points receives higher weights. This further illustrates that our regional attention mechanism enables each area in the shadow to select the surrounding non-shadowed region information that is helpful for its reconstruction.

4 SELECTIVE SHADOW REMOVAL

We also tested our model with human interaction. As shown in Fig.5, with the help of Segment Anything [5], the model permits the user to select the areas requiring shadow removal by simple straightforward clicks. It then generates high-quality shadow removal results for the specified shadow mask. This capability enables our model to respond effectively to selective shadow removal requirements.

REFERENCES

[1] Lanqing Guo, Siyu Huang, Ding Liu, Hao Cheng, and Bihan Wen. 2023. ShadowFormer: Global Context Helps Image Shadow Removal. In *AAAI*.

[2] Lanqing Guo, Chong Wang, Wenhan Yang, Siyu Huang, Yufei Wang, Hanspeter Pfister, and Bihan Wen. 2023. Shadowdiffusion: When degradation prior meets diffusion model for shadow removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14049–14058.

[3] Yeying Jin, Aashish Sharma, and Robby T Tan. 2021. DC-ShadowNet: Single-Image Hard and Soft Shadow Removal Using Unsupervised Domain-Classifier Guided Network. In *CVPR*. 5027–5036.

[4] Yeying Jin, Wei Ye, Wenhan Yang, Yuan Yuan, and Robby T. Tan. 2024. DeS3: Adaptive Attention-Driven Self and Soft Shadow Removal Using ViT Similarity. In *AAAI*. AAAI Press, 2634–2642.

[5] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloé Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross B. Girshick. 2023. Segment Anything. In *IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1-6, 2023*. IEEE, 3992–4003. https://doi.org/10.1109/ICCV51070.2023.00371

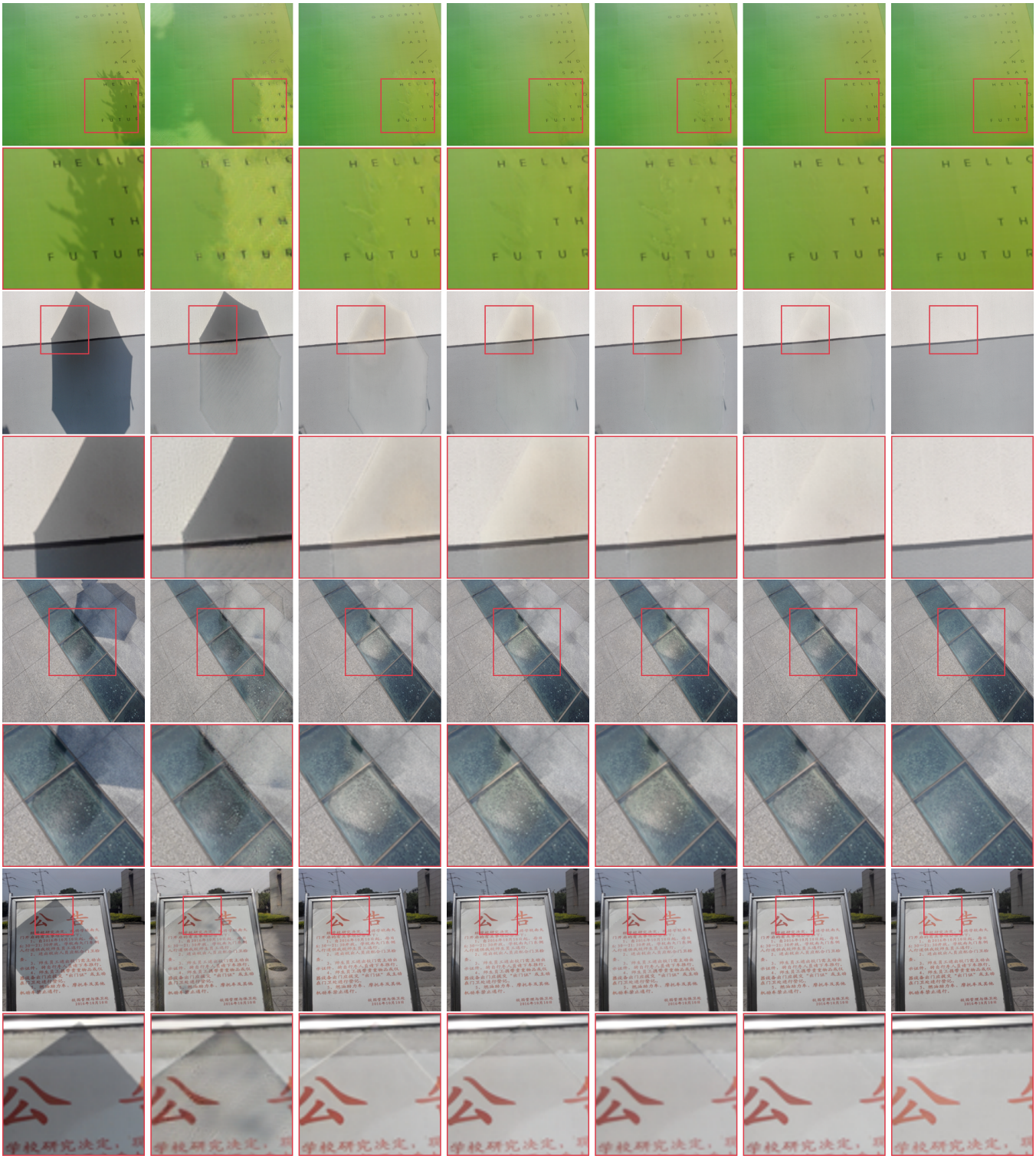
[6] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. 2019. Fast and Accurate Image Super-Resolution with Deep Laplacian Pyramid Networks. *IEEE TPAMI* 41, 11 (2019), 2599–2613.

[7] Hieu M. Le and Dimitris Samaras. 2019. Shadow Removal via Shadow Image Decomposition. In *ICCV*. IEEE, 8577–8586. https://doi.org/10.1109/ICCV.2019.00867

[8] Liangqiong Qu, Jiandong Tian, Shengfeng He, Yandong Tang, and Rynson W. H. Lau. 2017. DeshadowNet: A Multi-context Embedding Deep Network for Shadow Removal. In *CVPR*. IEEE Computer Society, 2308–2316. https://doi.org/10.1109/CVPR.2017.248

[9] Yurui Zhu, Jie Huang, Xueyang Fu, Feng Zhao, Qibin Sun, and Zheng-Jun Zha. 2022. Bijective Mapping Network for Shadow Removal. In *CVPR*. 5627–5636.

Unpublished working draft. Not for distribution.

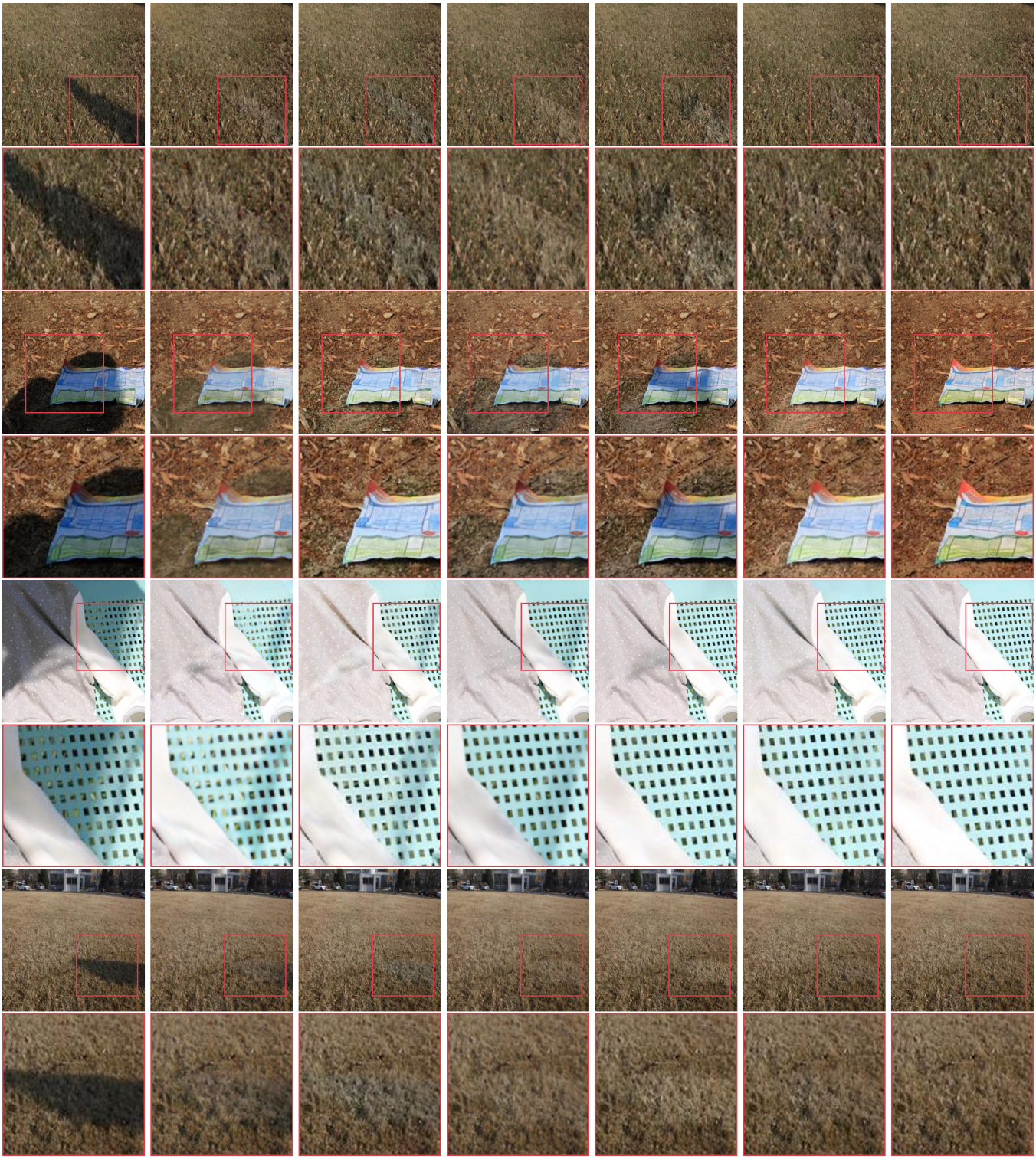


(a) Input (b) DC [3] (c) BMNet [9] (d) SF [1] (e) SD [2] (f) Ours (g) GT

Figure 1: A qualitive comparison on ISTD+ [7] dataset. Please zoom in to see the details.

(a) Input (b) DC [3] (c) BMNet [9] (d) SF [1] (e) SD [2] (f) Ours (g) GT

Figure 2: A qualitive comparison on ISTD+ [7] dataset. Please zoom in to see the details.



(a) Input (b) DC [3] (c) BMNet [9] (d) SD [2] (e) DeS3 [4] (f) Ours (g) GT

Figure 3: A qualitive comparison on SRD [8] dataset. Please zoom in to see the details.

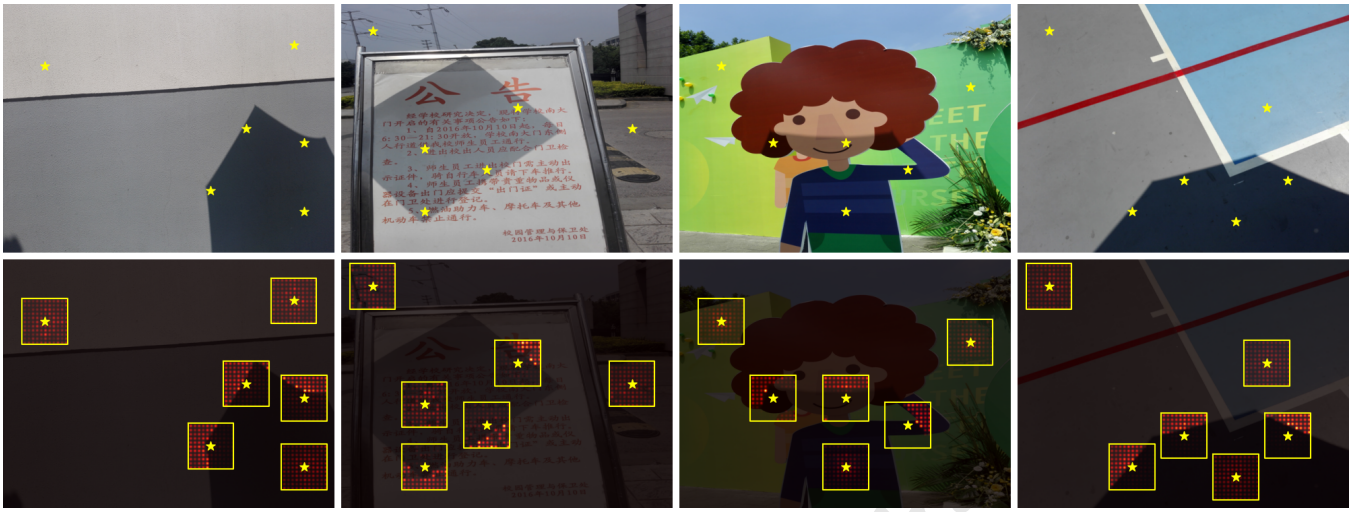
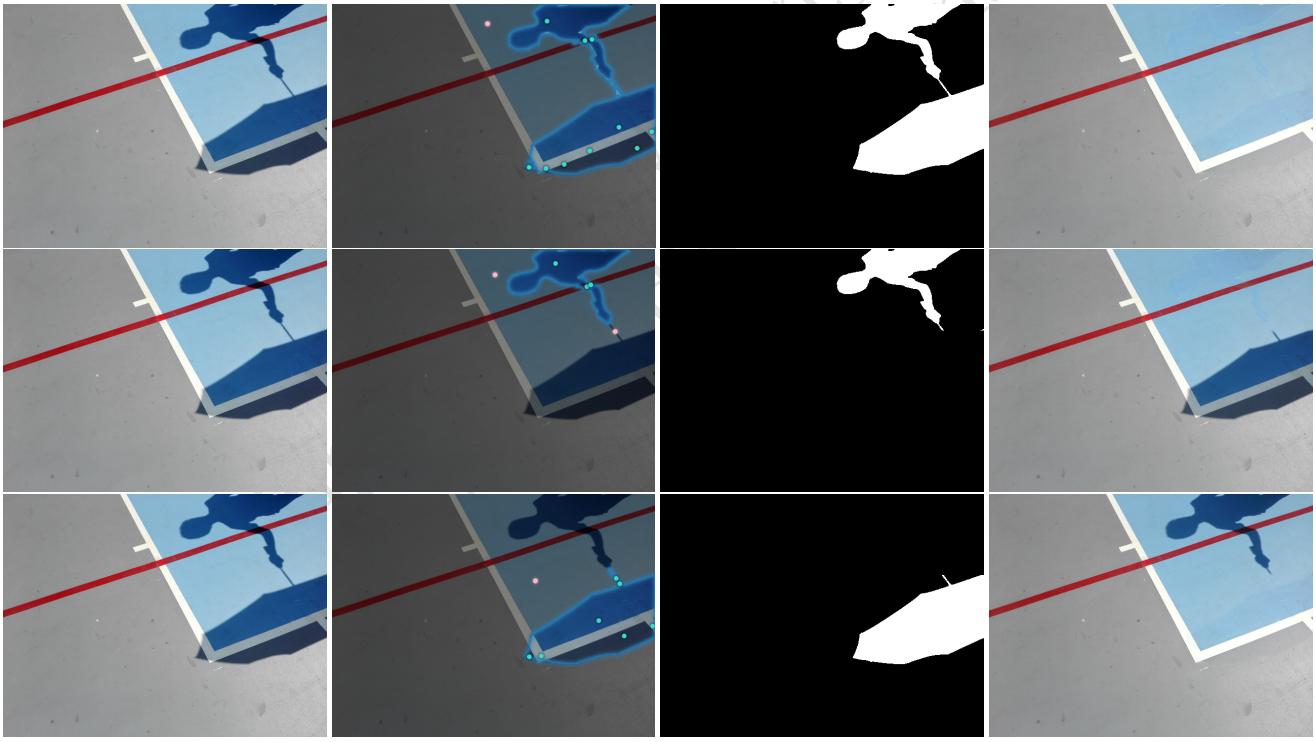


Figure 4: Visualization of our regional attention. The original image is on the first row, and the star marks the selected points. The heatmaps indicate the regional attention weight of the marked tokens. Brighter colors indicate a larger attention score.



(a) Input **(b) Shadow Area Selection** **(c) Mask** **(d) Output**

Figure 5: Selective shadow removal. In the second column of images, the green dots represent the areas we want to select, while the red dots represent the areas we wish to exclude from the selection. By selecting the shadow areas to be removed, our model can flexibly generate corresponding high-quality shadow removal results.

Unpublished working draft.
Not for distribution.

Unpublished working draft.
Not for distribution.