

# Supplementary Materials: Regional 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:

$$\mathcal{L} = \alpha_1 \mathcal{L}_{per} + \alpha_2 \mathcal{L}_{cont}, \quad (1)$$

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.

## 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, Hui Cheng, and Bihai Wen. 2023. ShadowFormer: Global Context Helps Image Shadow Removal. In *AAAI*.
- [2] Lanqing Guo, Chong Wang, Wenhao Yang, Siyu Huang, Yufei Wang, Hanspeter Pfister, and Bihai 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, Wenhao 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 Dollar, 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.

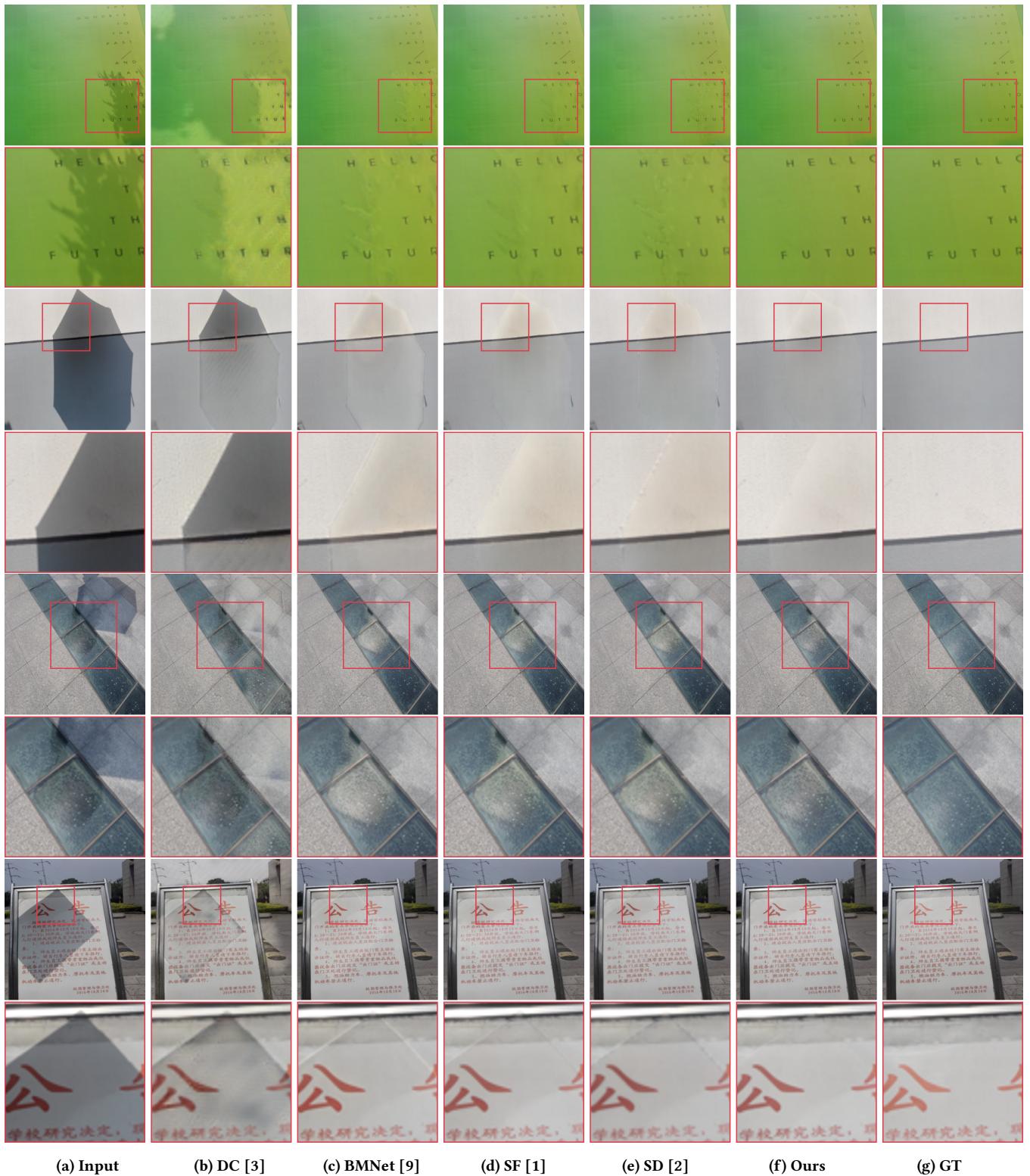


Figure 1: A qualitative comparison on ISTD+ [7] dataset. Please zoom in to see the details.

117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232



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

233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348

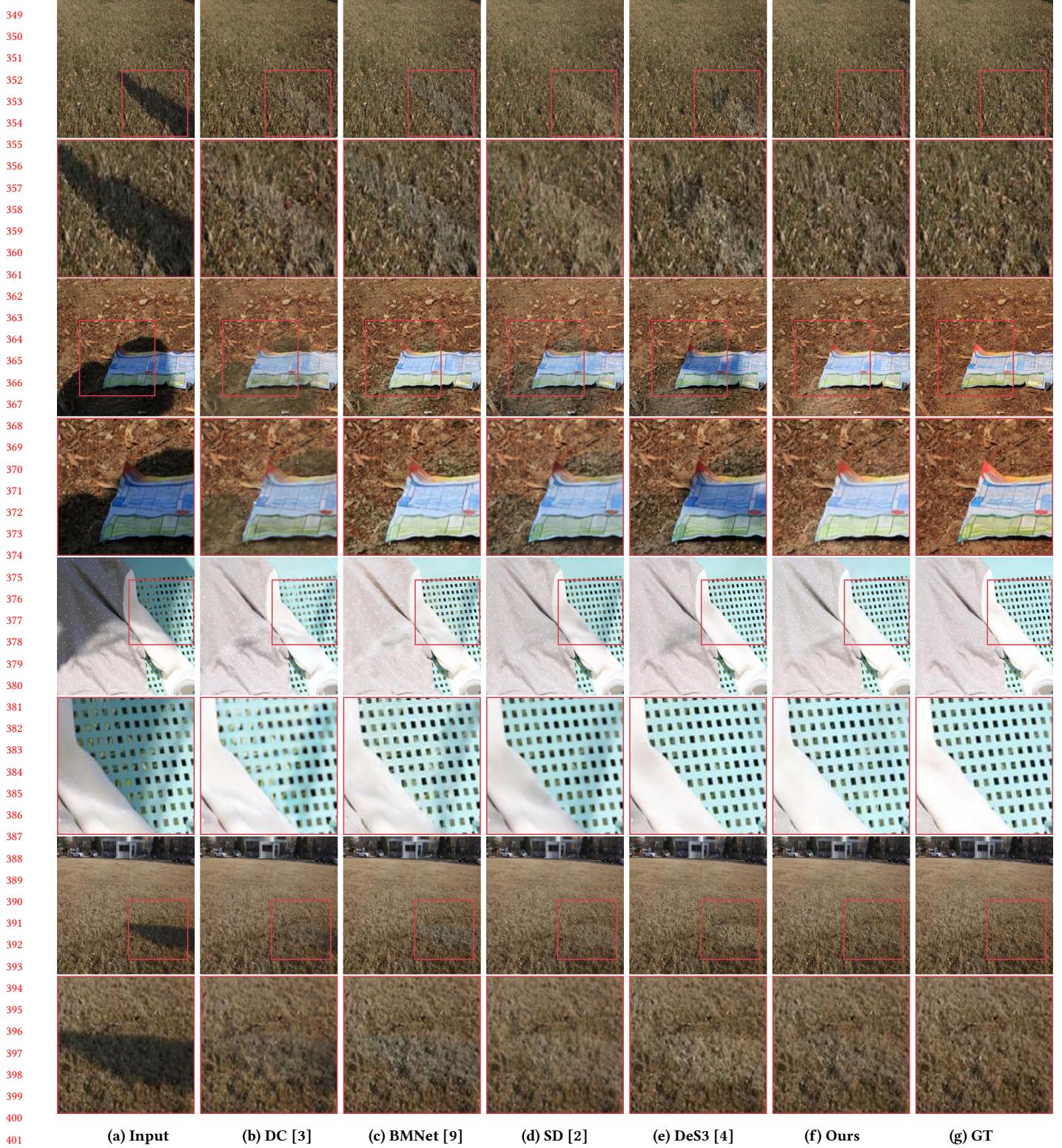


Figure 3: A qualitative comparison on SRD [8] dataset. Please zoom in to see the details.

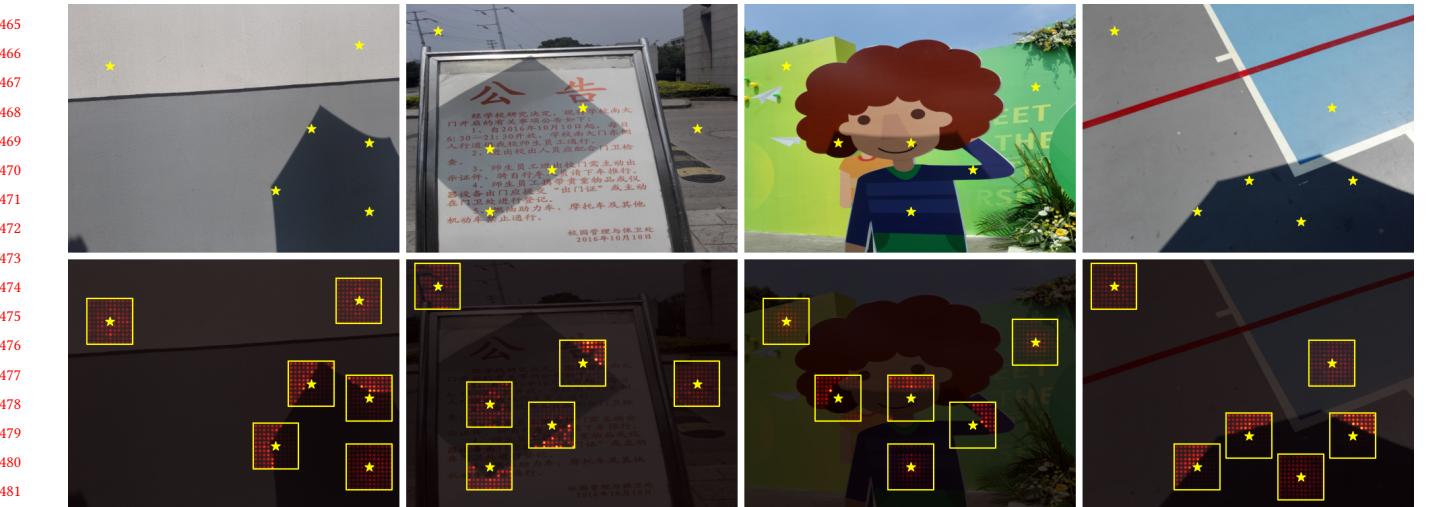


Figure 4: Visualization of our regional attention. The original image is on the first row, and the star marks the selected points. The heatmaps indicate the regional attention weight of the marked tokens. Brighter colors indicate a larger attention score.

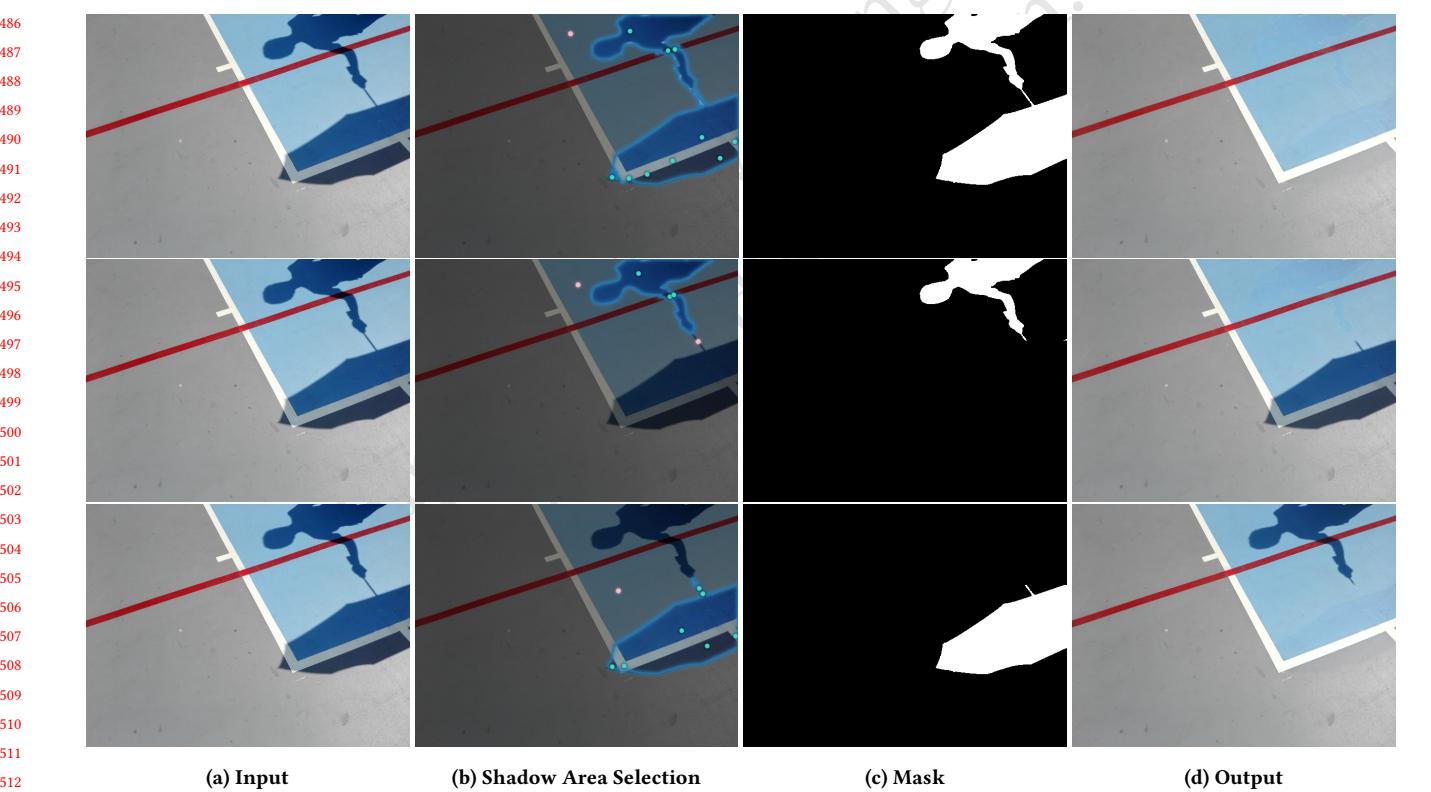


Figure 5: Selective shadow removal. In the second column of images, the green dots represent the areas we want to select, while the red dots represent the areas we wish to exclude from the selection. By selecting the shadow areas to be removed, our model can flexibly generate corresponding high-quality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

523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580

581	639
582	640
583	641
584	642
585	643
586	644
587	645
588	646
589	647
590	648
591	649
592	650
593	651
594	652
595	653
596	654
597	655
598	656
599	657
600	658
601	659
602	660
603	661
604	662
605	663
606	664
607	665
608	666
609	667
610	668
611	669
612	670
613	671
614	672
615	673
616	674
617	675
618	676
619	677
620	678
621	679
622	680
623	681
624	682
625	683
626	684
627	685
628	686
629	687
630	688
631	689
632	690
633	691
634	692
635	693
636	694
637	695
638	696

697	755
698	756
699	757
700	758
701	759
702	760
703	761
704	762
705	763
706	764
707	765
708	766
709	767
710	768
711	769
712	770
713	771
714	772
715	773
716	774
717	775
718	776
719	777
720	778
721	779
722	780
723	781
724	782
725	783
726	784
727	785
728	786
729	787
730	788
731	789
732	790
733	791
734	792
735	793
736	794
737	795
738	796
739	797
740	798
741	799
742	800
743	801
744	802
745	803
746	804
747	805
748	806
749	807
750	808
751	809
752	810
753	811
754	812