Supplementary Material for "Style-Guided Shadow Removal"

Jin Wan[1], Hui Yin[1,✉], Zhenyao Wu[2], Xinyi Wu[2], Yanting Liu[1], and Song Wang[2,✉]

[1] Beijing Jiaotong University, Beijing, China
{jinwan, hyin, 19112024}@bjtu.edu.cn
[2] University of South Carolina, Columbia, SC, USA
{zhenyao, xinyiw}@email.sc.edu, songwang@cec.sc.edu

In this supplementary material, we first introduce the details of the coarse deshadow network (CDNet) in Sec. A. Then we report additional quantitative results for performance comparison and complementary study in Sec. B. More visualization results in image- and video-level are presented in Sec. C. Finally, we discuss the failure cases in Sec. D.

A Coarse Deshadow Network

The detailed architecture of the proposed coarse deshadow network (CDNet) is shown in Fig. 1, which is built upon the Unet structure in [2]. We further remove skip connections and half the number of the convolution filters in [2] to reduce the computational cost of the network. The shadow image I^{S} and the corresponding shadow mask M are fed into the CDNet as inputs to predict a coarse de-shadowed image I^{C}, which serves as the input of the subsequent style-guided re-deshadow network.

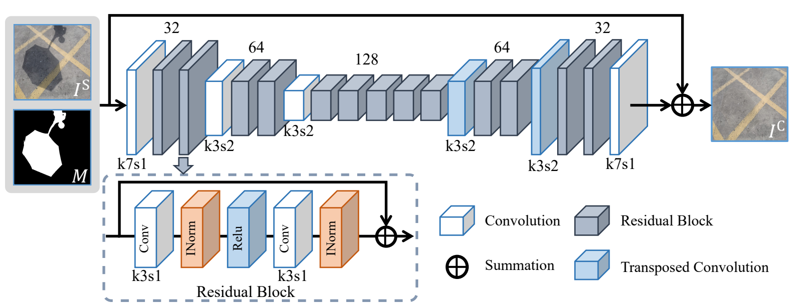

Fig. 1. The architecture of the coarse deshadow network with k and s denoting the kernel size and stride for the convolutional layers.

Table 1. Quantitative results of the proposed method compared to state-of-the-art methods on the SBU-Timelapse dataset [7].

Method	RMSE↓
SP+M-Net [5]	24.3
wSP+M-Net [7]	23.4
DC-ShadowNet [4]	25.1
SP+M+I-Net [7]	20.1
SG-ShadowNet (Ours)	**19.6**

Table 2. Quantitative results of G2R-shadowNet *Sup.* [9] and G2R-ShadowNet-*Sup.*+SRNet on the ISTD+ dataset [5].

Method	**Shadow** RMSE↓	**Non-Shadow** RMSE↓	**All** RMSE↓
G2R-ShadowNet *Sup.* [9]	7.32	2.87	3.62
G2R-ShadowNet *Sup.*+SRNet	6.57	2.86	3.47

B Additional Analysis

B.1 Quantitative results

To further justify the practicality and generalization capacity of our method, we further evaluate on SBU-Timelapse dataset [7] with the network trained on ISTD+ [5] and shadow masks generated by the pre-trained BDRAR [10]. It can be observed from Table 1 that our method still achieves the best shadow-removal performance in the moving-shadow regions – RMSE of 19.6. compared to SP+M+I-Net [7] (20.1) and DC-ShadowNet [4] (25.1), which further proves the strong practicality and generalization capacity of our method.

B.2 Complementary study

We also investigate the proposed style-guided re-deshadow network (SRNet) as a post-processing module to complement the existing shadow removal methods. Specially, we first generate coarsely de-shadowed images by using the official pre-trained model provided by G2R-shadowNet *Sup.* [9], which then is fed into the SRNet for another 200 training epochs with the corresponding ground truths. Note that we utilize the same shadow masks as in [9] for testing to make a fair comparison. It can be seen from the quantitative results in Table 2 and the qualitative comparisons in Fig. 2 that the proposed SRNet can bring additional improvements to the existing method.

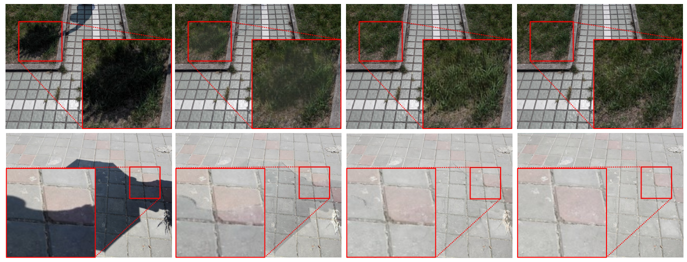

Fig. 2. Qualitative comparisons on two sample images from the ISTD+ dataset [5], by using G2R-shadowNet *Sup.* [9] and G2R-ShadowNet *Sup.*+SRNet.

C Qualitative Comparisons

In this section, we provide more qualitative comparison results on the ISTD+ dataset [5] in Fig. 3 and utilize the heatmaps to show the difference between the result and ground truth for the best view. It can be found that our method achieves better visual consistency between the shadow and non-shadow regions. From the heatmaps, we can also observe that the de-shadowed results generated by the proposed method are more faithful to the ground truths. Moreover, we also present visual comparison on the Video Shadow Removal dataset [6] in Fig. 4. From these video samples, we find that our method shows stronger generalization capacity by leveraging style guidance from the non-shadow region of the current frame for shadow removal.

D Failure Cases

In this section, we discuss a failure case of our method. As shown in Fig. 5, the shadow region (the region in the red rectangle of the images, zoom in for the best view.) is incorrectly divided into the non-shadow region, leading to unsatisfactory de-shadowed results. Mask-free shadow removal approach might be a good future direction to address this shortcoming.

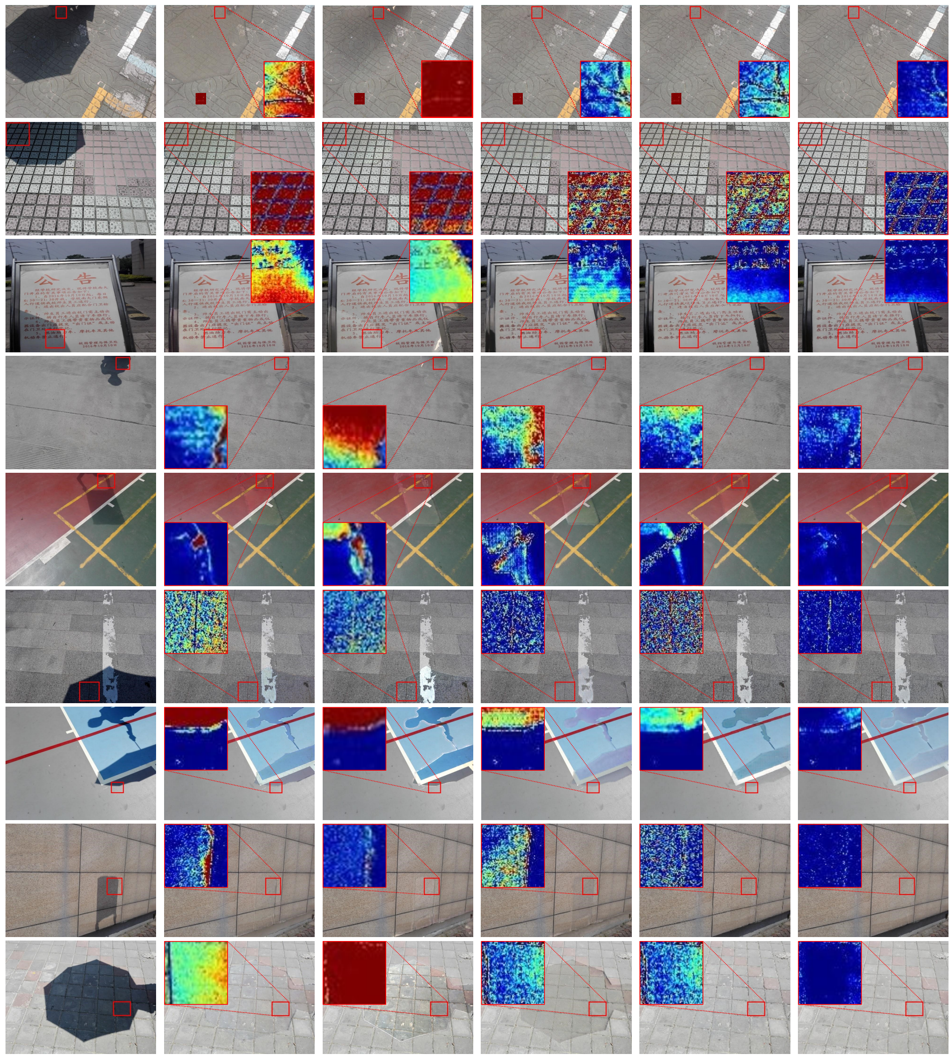

Fig. 3. Qualitative comparisons with the state-of-the-art methods on the ISTD+ dataset [5].

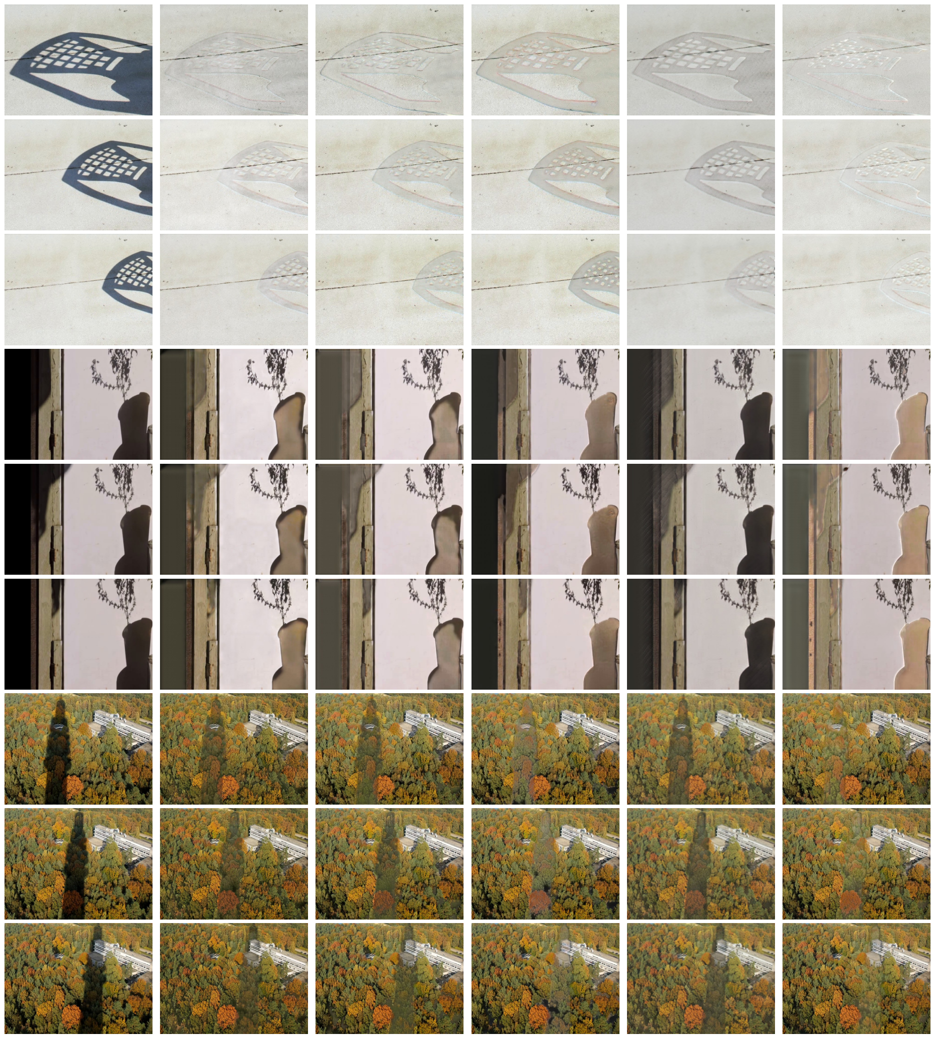

Fig. 4. Qualitative comparisons with the state-of-the-art methods on the Video Shadow Removal dataset [6].

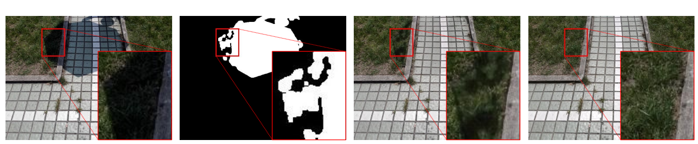

Fig. 5. Failure case arises in the wrong shadow mask.

References

1. Fu, L., Zhou, C., Guo, Q., Juefei-Xu, F., Yu, H., Feng, W., Liu, Y., Wang, S.: Auto-exposure fusion for single-image shadow removal. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 10571–10580 (June 2021) 4
2. Guo, S., Yan, Z., Zhang, K., Zuo, W., Zhang, L.: Toward convolutional blind denoising of real photographs. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 1712–1722 (2019) 1
3. Hu, X., Jiang, Y., Fu, C.W., Heng, P.A.: Mask-shadowgan: Learning to remove shadows from unpaired data. In: Int. Conf. Comput. Vis. (2019) 5
4. Jin, Y., Sharma, A., Tan, R.T.: Dc-shadownet: Single-image hard and soft shadow removal using unsupervised domain-classifier guided network. In: Int. Conf. Comput. Vis. pp. 5027–5036 (2021) 2, 5
5. Le, H., Samaras, D.: Shadow removal via shadow image decomposition. In: Int. Conf. Comput. Vis. (2019) 2, 3, 4
6. Le, H., Samaras, D.: From shadow segmentation to shadow removal. In: Eur. Conf. Comput. Vis. (2020) 3, 4, 5
7. Le, H., Samaras, D.: Physics-based shadow image decomposition for shadow removal. IEEE Trans. Pattern Anal. Mach. Intell. (2021) 2
8. Liu, Z., Yin, H., Mi, Y., Pu, M., Wang, S.: Shadow removal by a lightness-guided network with training on unpaired data. IEEE Trans. Image Process. **30**, 1853–1865 (2021) 5
9. Liu, Z., Yin, H., Wu, X., Wu, Z., Mi, Y., Wang, S.: From shadow generation to shadow removal. In: IEEE Conf. Comput. Vis. Pattern Recog. (2021) 2, 3, 4, 5
10. Zhu, L., Deng, Z., Hu, X., Fu, C.W., Xu, X., Qin, J., Heng, P.A.: Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. In: Eur. Conf. Comput. Vis. pp. 121–136 (2018) 2