Supplementary Materials
From Shadow Segmentation to Shadow Removal

Hieu Le and Dimitris Samaras

Stony Brook University, Stony Brook, NY 11794, USA

1 Overview

We provide additional analysis and qualitative shadow removal results of our proposed framework. In particular:

- We provide more examples of the components extracted from our framework [5].
- Ablation Studies: We visualize the learned components of our framework for the ablation studies described in the main paper.
- We evaluate the shadow removal performance of our method using different sets of shadow masks.
- We provide the shadow removal results of our model in comparison with state-of-the-art methods using ground truth images at the original scale.
- We provide a direct comparison of our method to the state-of-the-art method SP-M+Net [4].
- In addition to this document, we provide two videos showing the shadow removal results of our method in comparison with MaskShadow-GAN[3], and SP+M-Net [4] on two testing videos.

2 Components Visualization

Fig. 1 visualizes the components extracted from our framework. From the left to right are the input shadow patches, the relit images computed from the shadow parameters, the matte layers, and the output image patches. The figure shows that our method can automatically learn to accurately decompose the shadow effect in the input image patch into a matte layer and a relit image. This is achieved with only the adversarial loss obtained from D-Net. It can be seen from the figure that our framework performs shadow removal properly: the relit images I^{relit} have pixels in the shadow areas correctly lit to their non-shadow values and the matte layers α accurately model the shadow effects on the penumbra areas.

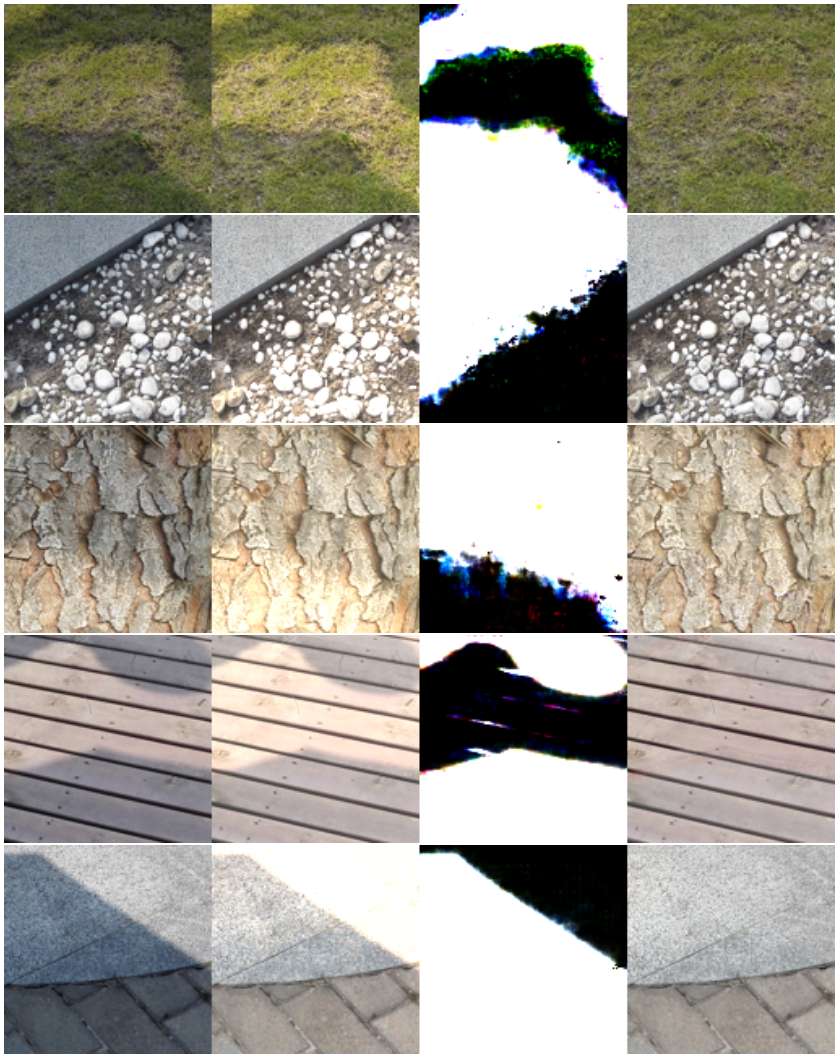


Fig. 1: **Shadow Image Decomposition.** Our method automatically learns to decompose the shadow effect in the input image patch I^{sd} into a matte layer α and a relit image I^{relit}. The matte layer α combines I^{sd} and I^{relit} to obtain a shadow-free image patch I^{output}.

3 Ablation Studies.

We visualize the components extracted from our framework trained with each proposed components removed, one at a time. The quantitative results of these ablation studies are presented in the main paper - Table 2.

Matting Loss - $\mathcal{L}_{mat-\alpha}$. Fig. 2 visualizes the extracted relit images and the matte layers of our framework when we remove the matting loss $\mathcal{L}_{mat-\alpha}$. Without this matting loss, Param-Net outputs consistently high values while Matte-Net tends to perform image inpainting to disguise the shadow effects, modifying both the shadow and non-shadow areas of the images.

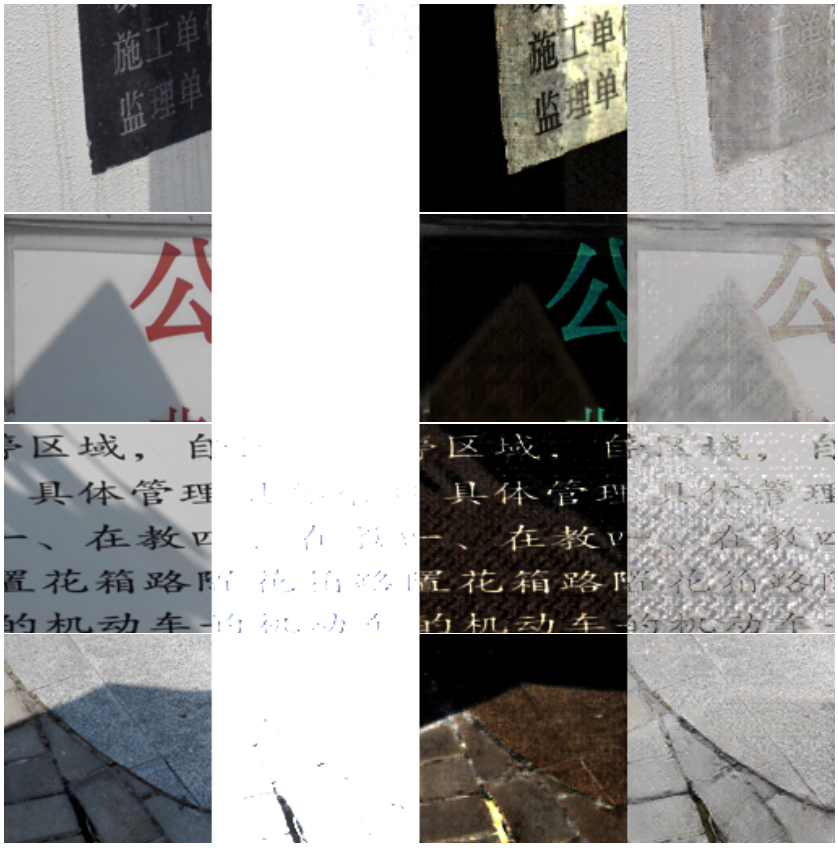


Fig. 2: **Ablation Study - Matting Loss $\mathcal{L}_{mat-\alpha}$.** The components extracted from our model trained when omitting the matting loss from the original setting. From left to right: Input shadow image patch, the relit image, the matte layer, and the output image obtained when the matting loss is omitted.

GAN Loss - $\mathcal{L}_{GAN}$. Fig. 3 shows the components learned by our model when omitting the GAN loss. This model is trained using the proposed physical constraints, which limit the framework to only modify the images in a manner that is similar to shadow removal, i.e., all shadowed pixels are lit uniformly without introducing new artifacts and the textures beneath the shadow areas

are preserved. There are visible boundary artifacts and the shadowed pixels are not relit to the appropriate values.

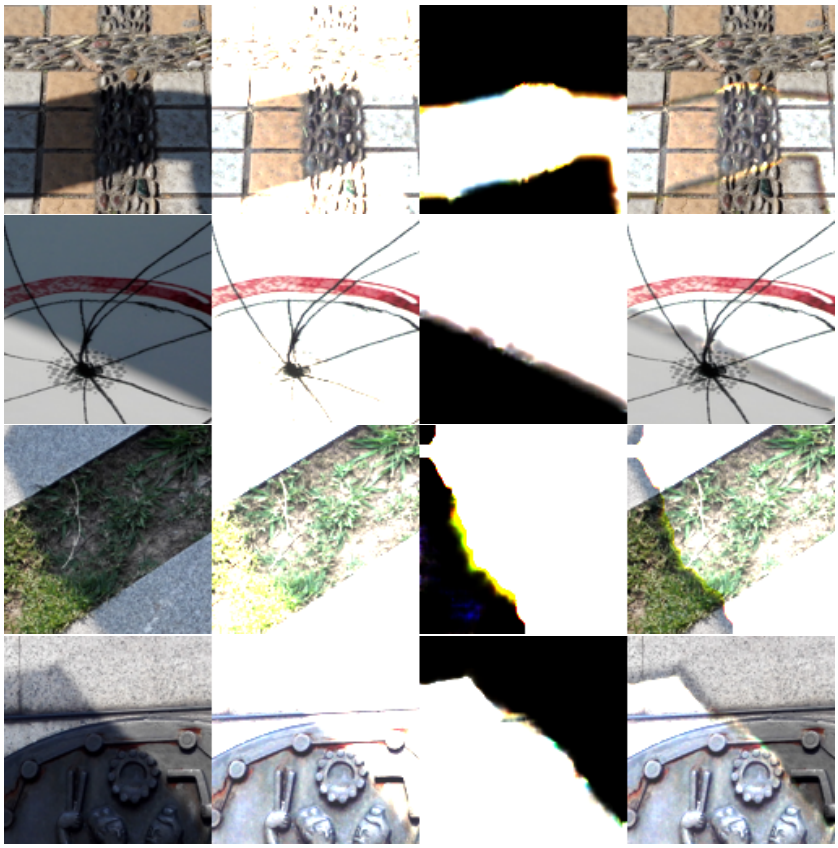


Fig. 3: **Ablation Study - GAN Loss $\mathcal{L}_{GAN}$**. The components extracted from our model trained when omitting the GAN loss from the original setting. From left to right: Input shadow image patch, the relit image, the matte layer, and the output image obtained when the GAN loss is omitted. There are visible boundary artifacts and the shadowed pixels are not relit to the appropriate values.

Smoothness Loss - $\mathcal{L}_{sm-\alpha}$. Fig. 4 shows the components learned by our model when omitting the smoothness loss. Note that omitting the smoothness loss $\mathcal{L}_{sm}$ only results in a slight drop in shadow removal performance, from 9.7 to 10.2 RMSE on the shadow areas. However, we observe more visible boundary artifacts on the output images without this loss, as clearly shown in Fig. 5.

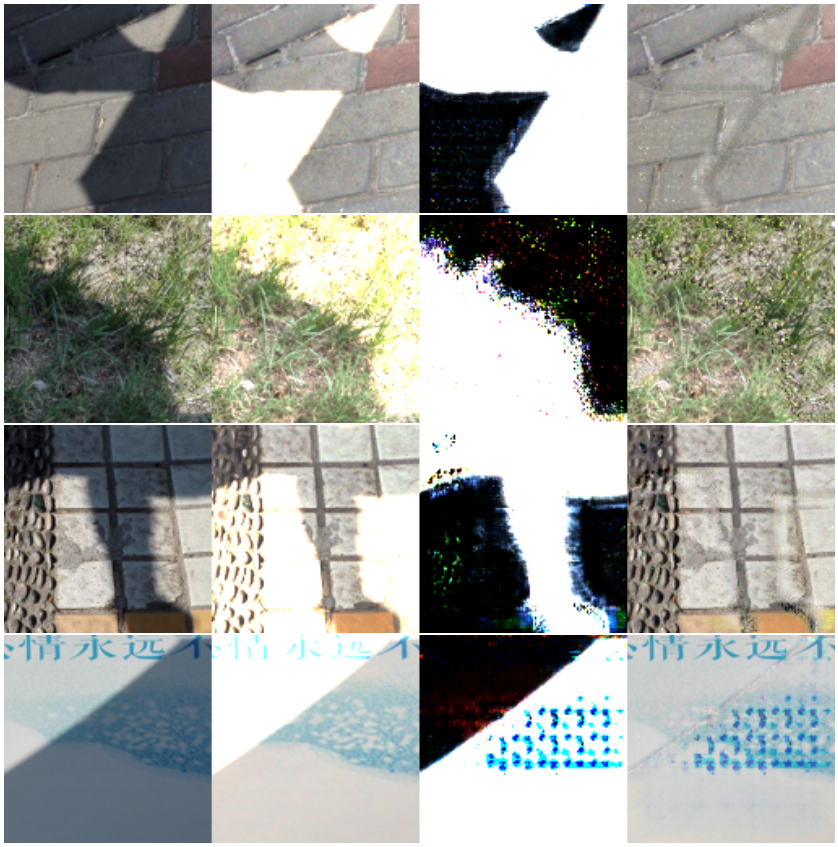


Fig. 4: **Ablation Study- smoothness Loss $\mathcal{L}_{sm-\alpha}$**. The components extracted from our model trained when omitting the smoothness loss from the original setting. From left to right: Input shadow image patch, the relit image, the matte layer, and the output image obtained when the smoothness loss is omitted. There are more visible boundary and image artifacts without this loss. More shadow removal examples for this ablation study are provided in Fig. 5

Boundary Loss - $\mathcal{L}_{bd}$ and limiting the parameters search space. Figs. 6 and 7 show that the model collapses when we omit the boundary loss $\mathcal{L}_{bd}$ or when we do not bound the search space of the shadow parameters to a proper range. In Fig. 6, Param-Net outputs consistently high values. In Fig.7, there is no meaningful mapping being learned.

4 Analysis of Shadow Mask Quality.

We evaluated the shadow removal performance of our method using different sets of shadow masks. We used as shadow detector the model proposed by Zhu *et al.*[10] pre-trained on the SBU dataset [7] (publicly available by the authors). We fine-tuned this model on the ISTD training set for 3000 epochs, and acquired a 2.2 balance-error-rate (BER) on the ISTD testing set. We used two snapshots of the model during the fine-tuning at epochs 1000 and 2000 to output two sets of shadow masks, which achieved a 4.1 BER and 3.1 BER respectively on the ISTD testing set. Last, we evaluated the performance of our model using the ground truth shadow masks.

Table 1 reports the shadow removal performance of our models using different sets of shadow masks. The results of our model improve with better sets of shadow masks.

Table 1: **Our Shadow Removal using different sets of shadow masks.** The shadow masks are generated by three snapshots of the shadow detector proposed by Zhu *et al.*[10] pre-trained on the SBU dataset [7] and fine-tuned on the the ISTD dataset for 1000, 2000, and 3000 epochs. The RMSE are shown for the shadow area, non-shadow area, and the whole image. All images are resized to 256×256. The last row shows the results of our model using the actual ground truth shadow masks.

Epoch	BER	Shad.	Non-Shad.	All
1000	4.1	10.0	3.7	4.7
2000	3.1	9.6	3.0	4.0
3000	2.2	9.7	3.0	4.0
GT Shad. Masks	0	**9.1**	**2.6**	**3.6**

I^{sd} W/o $\mathcal{L}_{sm-\alpha}$ With $\mathcal{L}_{sm-\alpha}$ Ground Truth

Fig. 5: **The effect of the smoothness loss $\mathcal{L}_{sm-\alpha}$ on the output images of our framework.** Without the smoothness loss, there are more visible boundary artifacts on the output images.

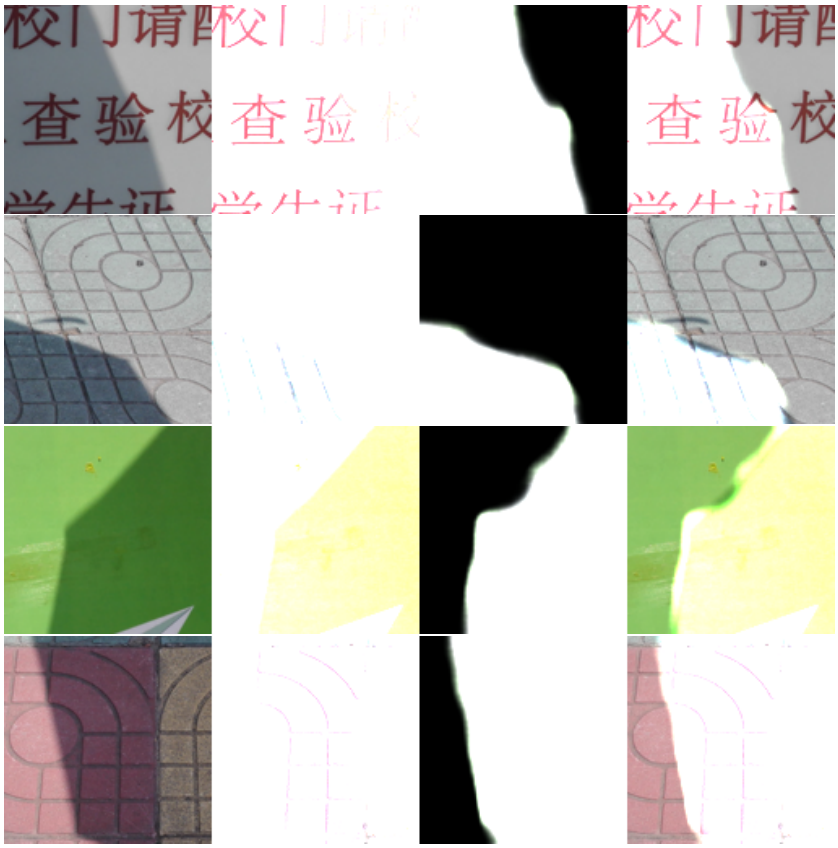


Fig. 6: **Ablation Study - boundary Loss $\mathcal{L}_{bd}$.** The components extracted from our model trained when omitting the boundary loss from the original setting. From left to right: Input shadow image patch, the relit image, the matte layer, and the output image obtained when the boundary loss is omitted.

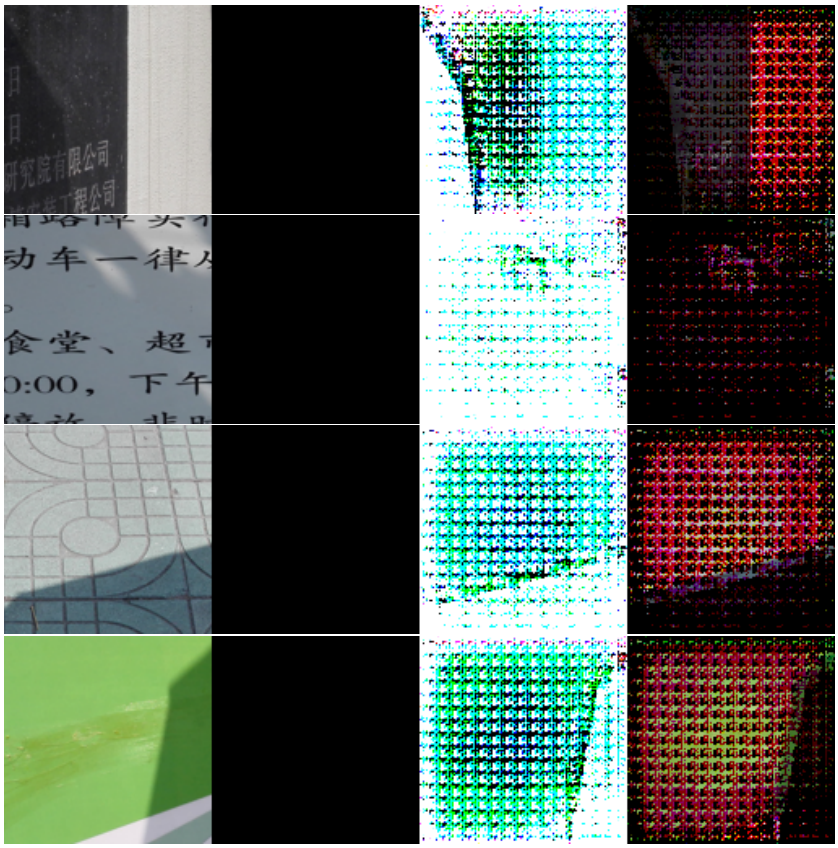


Fig. 7: **Ablation Study: no limits on search space**. The components extracted from our model trained when we set the search space of the scaling factor w to $[-10, 10]$ and the search space of the additive constant b to $[-255, 255]$. From left to right: Input shadow image patch, the relit image, the matte layer, and the output image obtained when the search space is not limited.

5 Shadow Removal Evaluation on the original size ground truth images.

Previous works [8, 6, 4] evaluate shadow removal performances on ground truth images resized to 256× 256. However, different resizing techniques might cause inconsistent shadow removal evaluations. To avoid this, we compare the shadow removal performance of our methods to other state-of-the-art methods on the ground truth images at the original scale (480×640) in Table 2. Each method originally outputs the results at a different size, as shown in the second column. We resize all images to 480×640 using bi-linear interpolation.

Table 2: **Shadow removal results of our networks in comparison with state-of-the-art methods for shadow removal on the ground truth images at the original scale (480×640).** The metric is RMS error (the lower, the better). The RMS errors are shown for the shadow area, non-shadow area, and the whole image. All images are resized to 480×640 via bi-linear interpolation.

Methods	Output Size	Shadow	Non-Shadow	All
Input Image	-	40.4	3.3	9.1
Yang *et al.* [9]	600×600	25.3	15.2	16.8
Guo *et al.* [2]	640×640	22.3	4.3	7.1
Gong *et al.* [1]	480×640	14.4	-	-
ST-CGAN *et al.*[8]	256×256	14.0	8.6	9.5
DeshadowNet [6]	256×256	16.3	7.5	8.9
MaskShadow-GAN [3]	480×640	13.1	5.0	6.3
SP+M-Net [4]	512×512	**8.5**	**3.6**	**4.4**
Ours	480×640	10.5	3.7	4.8

6 Direct comparison to SP+M-Net.

We provide a direct comparison to SP+M-Net. Fig. 8 shows a histogram of per-image RMSE differences on the shadow areas between our method and SP+M-Net. The values are computed by taking $RMSE_{ours} - RMSE_{SP+M-Net}$ for all images in the ISTD testing set [8] (540 images in total). A negative value means our method is better and a positive value means the opposite. It shows that we are comparable to SP+M-Net even though our method is trained with only shadow masks and SP+M-Net is trained fully-supervised: Most of the values are in the range [-2,2]. We perform better than SP+M-Net in 156/540 images. Fig. 9 shows some of these examples. It is clear that our method relit the shadowed pixels more accurately.

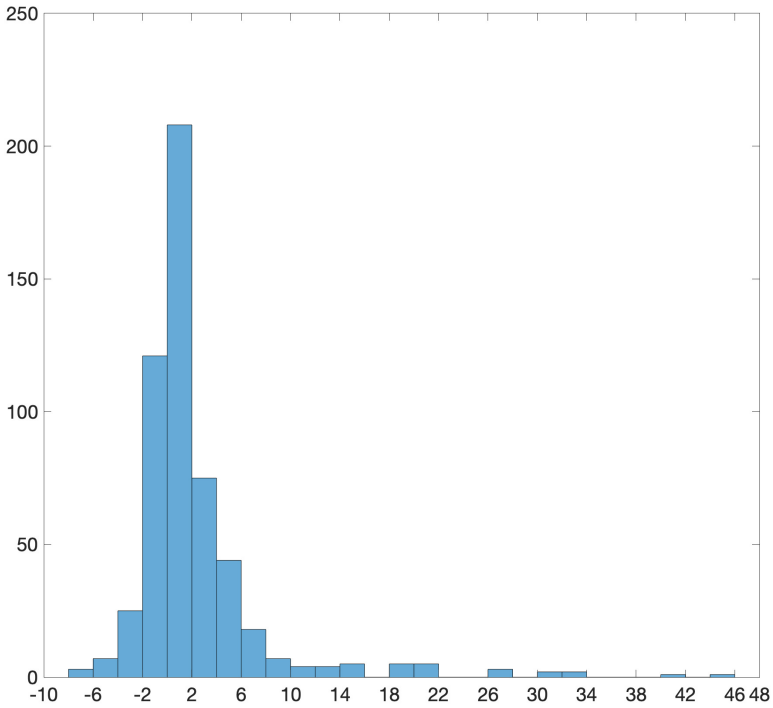


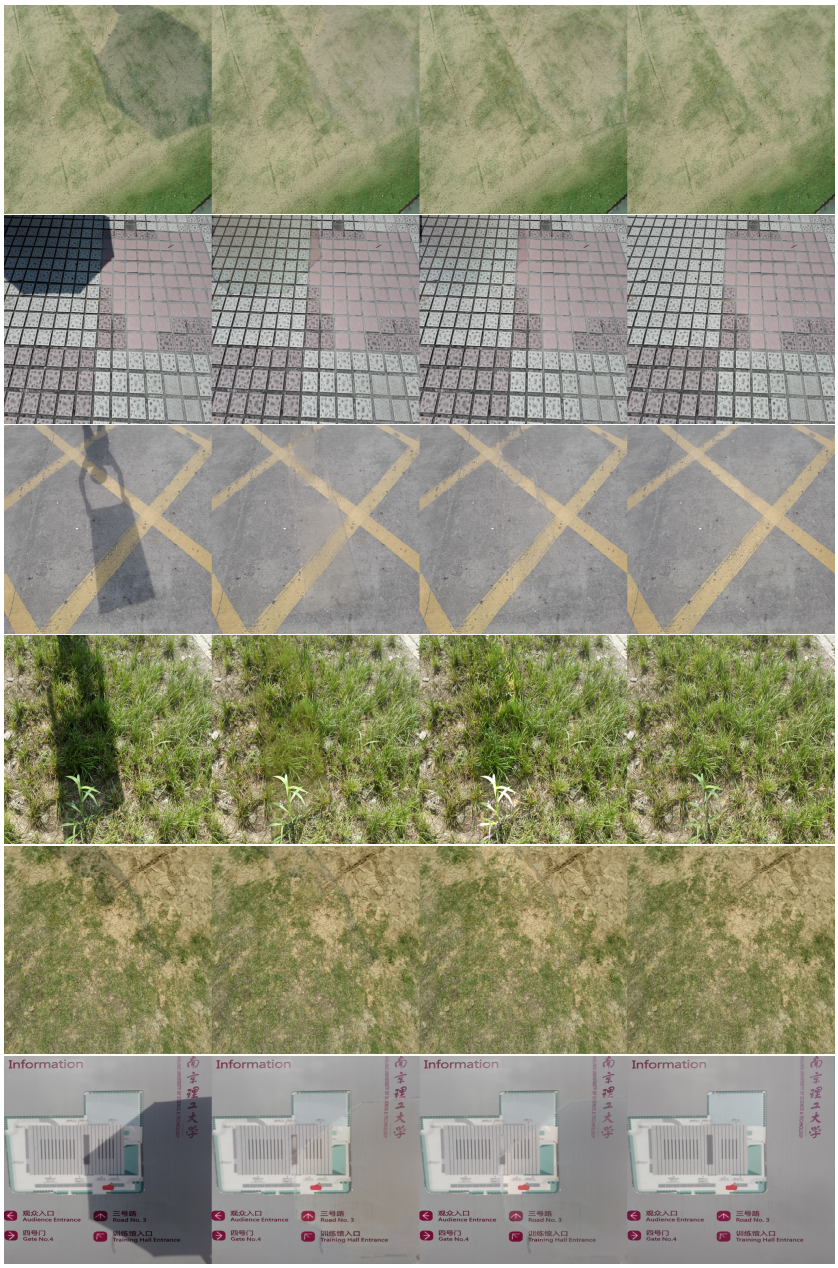
Fig. 8: **Comparison between our method and SP+M-Net [4]**. The histogram of per-image RMSE differences between our method and SP+M-Net. The values are computed by taking $RMSE_{ours} - RMSE_{SP+M-Net}$ for all images in the ISTD testing set [8]. A negative value means our method is better and a positive value means the opposite. It shows that we are comparable to SP+M-Net even though our method is trained with only shadow masks and SP+M-Net is trained fully-supervised.

7 Video Shadow Removal Results.

In additional to this document, we include two videos showing the shadow removal results of our method in comparison with MaskShadow-GAN [3] and SP+M-Net [4] on two videos, namely "tower" and "slovenkia". Both videos show that our proposed video shadow removal dataset is an extremely challenging test for shadow removal methods. All methods fail to remove the shadows or handle the shadow boundaries completely.

The results shown in the videos were obtained by frame-by-frame shadow removal. No motion cue was used. No temporal coherence was enforced. Our method, albeit not perfect, is able to remove shadows from these videos better

than the other methods. MaskShadow-GAN only mildly removes shadows in both videos while SP+M-Net is not able to perform in the several last frames of the video "slovenkia" where the shadow areas occupy the majority of the images. On the other hand, our method is affected by the imperfect shadow masks as can be seen in "tower" where some shadow areas are not detected consistently throughout the videos, creating the "flickering" effects on the bottom left corner. We believe this is an interesting problem for future research.



I^{sd} SP+M-Net [4] Ours Ground Truth

Fig. 9: **Comparison between our method and SP+M-Net.** SP+M-Net is trained fully supervised while we only need shadow masks. Our method outperforms SP+M-Net in these casse.

References

1. Gong, H., Cosker, D.: Interactive removal and ground truth for difficult shadow scenes. J. Opt. Soc. Am. A **33**(9), 1798–1811 (2016). https://doi.org/10.1364/JOSAA.33.001798, http://josaa.osa.org/abstract.cfm?URI=josaa-33-9-1798
2. Guo, R., Dai, Q., Hoiem, D.: Paired regions for shadow detection and removal. IEEE Transactions on Pattern Analysis and Machine Intelligence (2012)
3. Hu, X., Jiang, Y., Fu, C.W., Heng, P.A.: Mask-ShadowGAN: Learning to remove shadows from unpaired data. In: ICCV (2019), to appear
4. Le, H., Samaras, D.: Shadow removal via shadow image decomposition. In: Proceedings of the International Conference on Computer Vision (2019)
5. Le, H., Samaras, D.: From shadow segmentation to shadow removal. In: ECCV (2020)
6. Qu, L., Tian, J., He, S., Tang, Y., Lau, R.W.H.: Deshadownet: A multi-context embedding deep network for shadow removal. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2017)
7. Vicente, T.F.Y., Hou, L., Yu, C.P., Hoai, M., Samaras, D.: Large-scale training of shadow detectors with noisily-annotated shadow examples. In: Proceedings of the European Conference on Computer Vision (2016)
8. Wang, J., Li, X., Yang, J.: Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2018)
9. Yang, Q., Tan, K., Ahuja, N.: Shadow removal using bilateral filtering. IEEE Transactions on Image Processing **21**, 4361–4368 (2012)
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: Proceedings of the European Conference on Computer Vision (2018)