

# SHARDS: Efficient SHAdow Removal using Dual Stage Network for High-Resolution Images (Supplementary Material)

Mrinmoy Sen<sup>1</sup>, Sai Pradyumna Chermala<sup>1</sup>, Nazrinbanu Nurmohammad Nagori<sup>1</sup>,  
Venkat Peddigari<sup>1</sup>, Praful Mathur<sup>1</sup>, B H Pawan Prasad<sup>1</sup>, Moonhwan Jeong<sup>2</sup>

<sup>1</sup>Samsung R&D Institute India - Bangalore, <sup>2</sup>Samsung Electronics

{mrinmoy.sen, nazrin.n, p.mathur, pawan.prasad, mh0614.jeong}@samsung.com

{chs.pradyumna, venkatrp peddigari}@gmail.com

## 1. Overview

In this supplementary material we provide details on the architectures used for our shadow removal networks. We also provide additional visual results from the various networks proposed in our method.

## 2. Network Architecture

The architectures of our proposed Low-resolution Shadow Removal Network (LSRNet), Detail Refinement Network (DRNet) and the Discriminator used for the adversarial training are detailed in Table 1, 2 and 3 respectively. As described in the paper both LSRNet and DRNet follows the encoder-bottleneck-decoder architecture. DRNet shares the same architecture of LSRNet with the exception of attention mechanism in the network and uses bottleneck residual blocks instead of vanilla convolutional layers.

## 3. Additional Results

### 3.1. Results: Shadow Removal

In Table 1 from the paper, we quantitatively compared our proposed shadow removal network against the state-of-the-art baselines SID [6], DHAN [2] and AEF [3]. Our quantitative results are different from the claims of authors due to a change in the evaluation pipeline. As pointed out by the authors of [6] here, the calculated and reported metric in most of the exiting works is MAE but incorrectly reported as RMSE. We evaluated our results and the official results from published/pre-trained networks of above works using the below pipeline:

For each image in the test set:

1. Generate results using the network.
2. Convert the predicted image and ground-truth to LAB color space using the `rgb2lab` function from `skimage.color` module.

3. Calculate RMSE on the entire image and store the result.

4. Mask out the pixels in shadow-free region and shadow region for calculating RMSE on shadow and shadow-free regions respectively and store the results.

Return the average of stored values for all the images in the test set.

We present more visual comparisons of our model against the existing methods in Figure 1.

### 3.2. Results : Shadow Removal on High-Resolution images

Existing shadow removal methods are limited to low-resolution and do not scale well to high-resolution images. We provide additional examples in Figure 2, where we demonstrate that for existing techniques, shadow removal quality degrades as we progressively increase the image resolution. In contrast, our proposed method produces consistent artefact-free results even on high-resolution images.

### 3.3. Results : Shadow Detection

We provide additional qualitative comparison with our proposed shadow detector network against existing state-of-the-art techniques on the SBU dataset in Figure 3.

### 3.4. Results : End-to-end Shadow Removal results on High-Resolution images

In Figure 4 we showcase our end-to-end results on high-resolution images using the proposed shadow detection and shadow removal networks on ISTD-HQ (resolution: 2560x1920) and SFHQ (resolution: 4034x3024) datasets. SFHQ dataset consists of shadow annotations of only external-cast shadows. The network trained on SFHQ dataset is thus trained to remove external-cast shadows and leaves traces of self-cast shadows in some cases.

Layer	Configuration
conv	$k = 7 \times 7, c = 32$
conv	$k = 3 \times 3, c = 64, stride = 2$
conv	$(k = 3 \times 3, c = 128, stride = 2) \times 2$
conv	$k = 3 \times 3, c = 256, stride = 2$
bottleneck residual	$\begin{pmatrix} k = 1 \times 1, c = 64 \\ k = 3 \times 3, c = 64 \\ k = 1 \times 1, c = 256 \end{pmatrix} \times 4$
attention-bottleneck residual	$\begin{pmatrix} self - attention \\ k = 1 \times 1, c = 64 \\ k = 3 \times 3, c = 64 \\ k = 1 \times 1, c = 256 \end{pmatrix} \times 5$
cbam-up-conv	$\begin{pmatrix} CBAM \\ bilinear\ upsampling\ 2x \\ k = 3 \times 3, c = 128 \end{pmatrix} \times 2$
cbam-up-conv	$\begin{pmatrix} CBAM \\ bilinear\ upsampling\ 2x \\ k = 3 \times 3, c = 64 \end{pmatrix}$
cbam-up-conv	$\begin{pmatrix} CBAM \\ bilinear\ upsampling\ 2x \\ k = 3 \times 3, c = 32 \end{pmatrix}$
conv	$k = 7 \times 7, c = 3$

Table 1: LSRNet architecture

Layer	Configuration
conv	$k = 7 \times 7, c = 32$
bottleneck residual	$\begin{pmatrix} k = 1 \times 1, c = 16 \\ k = 3 \times 3, c = 16, stride = 2 \\ k = 1 \times 1, c = 64 \end{pmatrix}$
bottleneck residual	$\begin{pmatrix} k = 1 \times 1, c = 32 \\ k = 3 \times 3, c = 32, stride = 2 \\ k = 1 \times 1, c = 128 \end{pmatrix} \times 2$
bottleneck residual	$\begin{pmatrix} k = 1 \times 1, c = 32 \\ k = 3 \times 3, c = 32 \\ k = 1 \times 1, c = 128 \end{pmatrix} \times 9$
bottleneck residual	$\begin{pmatrix} upsampling\ 2 \times 2 \\ k = 1 \times 1, c = 32 \\ k = 3 \times 3, c = 32 \\ k = 1 \times 1, c = 128 \end{pmatrix}$
bottleneck residual	$\begin{pmatrix} upsampling\ 2 \times 2 \\ k = 1 \times 1, c = 16 \\ k = 3 \times 3, c = 16 \\ k = 1 \times 1, c = 64 \end{pmatrix}$
bottleneck residual	$\begin{pmatrix} upsampling\ 2 \times 2 \\ k = 1 \times 1, c = 8 \\ k = 3 \times 3, c = 8 \\ k = 1 \times 1, c = 32 \end{pmatrix}$
conv	$k = 7 \times 7, c = 3$

Table 2: DRNet architecture

Layer	Configuration
conv	$k = 4 \times 4, c = 32, stride = 2$
conv	$k = 4 \times 4, c = 64, stride = 2$
conv	$k = 4 \times 4, c = 128, stride = 2$
attention	$self - attention$
conv	$k = 4 \times 4, c = 256$
conv	$k = 4 \times 4, c = 1$

Table 3: Discriminator network architecture

## References

- [1] Zhihao Chen, Lei Zhu, Liang Wan, Song Wang, Wei Feng, and Pheng-Ann Heng. A multi-task mean teacher for semi-supervised shadow detection. In *CVPR*, 2020.
- [2] Xiaodong Cun, Chi-Man Pun, and Cheng Shi. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting gan. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(07):10680–10687, Apr. 2020.
- [3] Lan Fu, Changqing Zhou, Qing Guo, Felix Juefei-Xu, Hongkai Yu, Wei Feng, Yang Liu, and Song Wang. Auto-exposure fusion for single-image shadow removal. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10566–10575, 2021.
- [4] Xiaowei Hu, Tianyu Wang, Chi-Wing Fu, Yitong Jiang, Qiong Wang, and Pheng-Ann Heng. Revisiting shadow detection: A new benchmark dataset for complex world. *arXiv preprint arXiv:1911.06998*, 2019.
- [5] X. Hu, L. Zhu, C. Fu, J. Qin, and P. Heng. Direction-aware spatial context features for shadow detection. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7454–7462, 2018.
- [6] H. Le and D. Samaras. Shadow removal via shadow image decomposition. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8577–8586, 2019.
- [7] Quanlong Zheng, Xiaotian Qiao, Ying Cao, and Rynson W.H. Lau. Distraction-aware shadow detection. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [8] Lei Zhu, Zijun Deng, Xiaowei Hu, Chi-Wing Fu, Xuemiao Xu, Jing Qin, and Pheng-Ann Heng. Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision – ECCV 2018*, pages 122–137, Cham, 2018. Springer International Publishing.

Input

Mask

SID

DHAN

Exposure Fusion

Ours

Ground truth

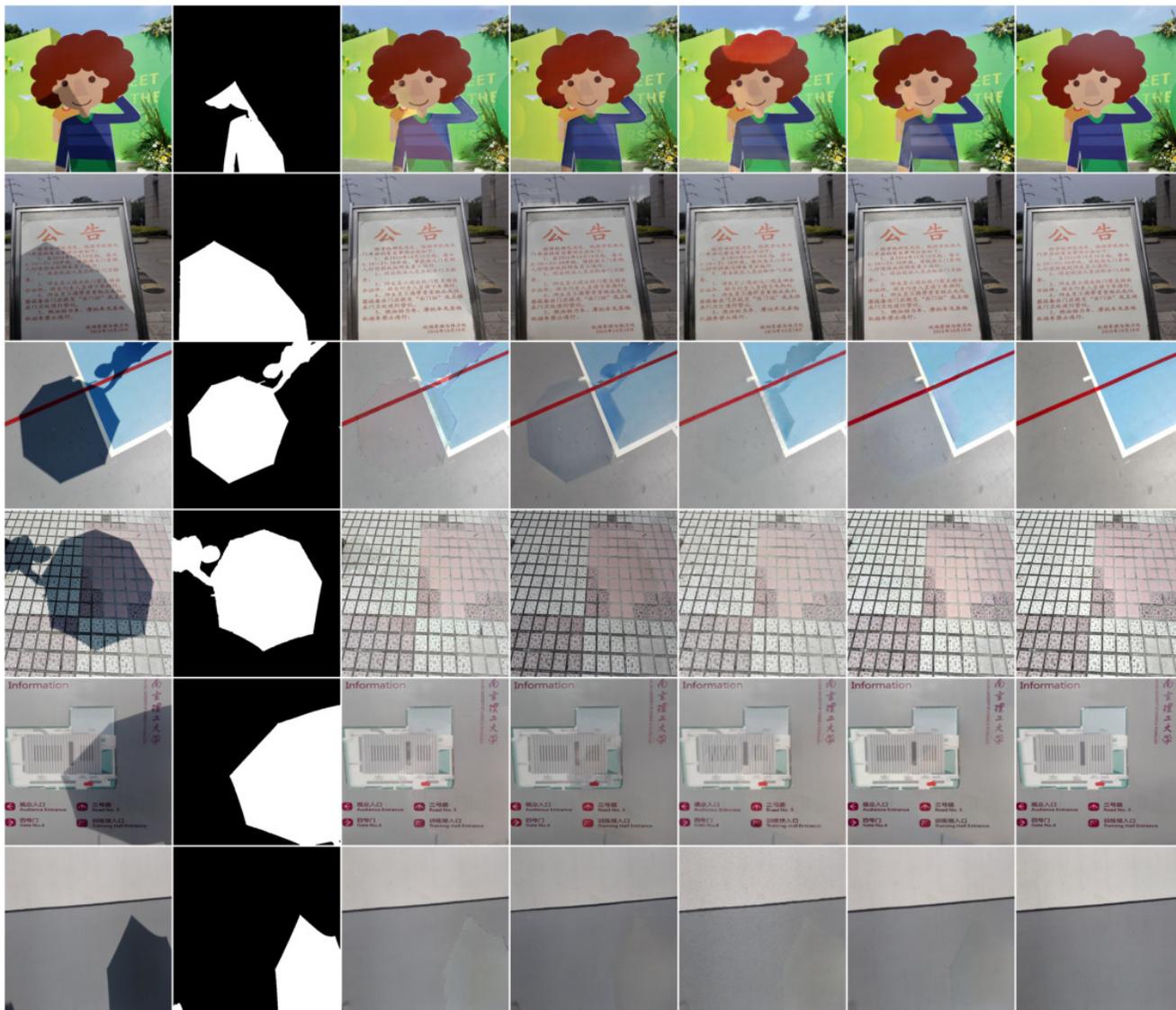




Figure 1: Qualitative results of our proposed Shadow Removal method, compared with existing state-of-the-art methods SID [6], DHAN [2], AEF [3] on ISTD and SFHQ datasets (**Zoom-in for better visualization**).

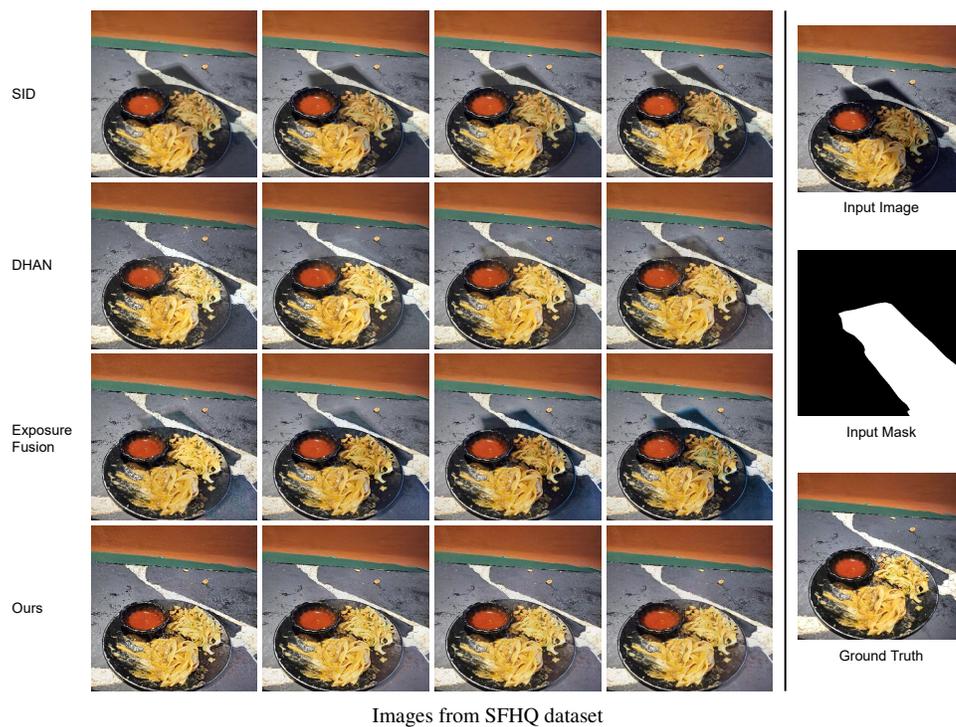
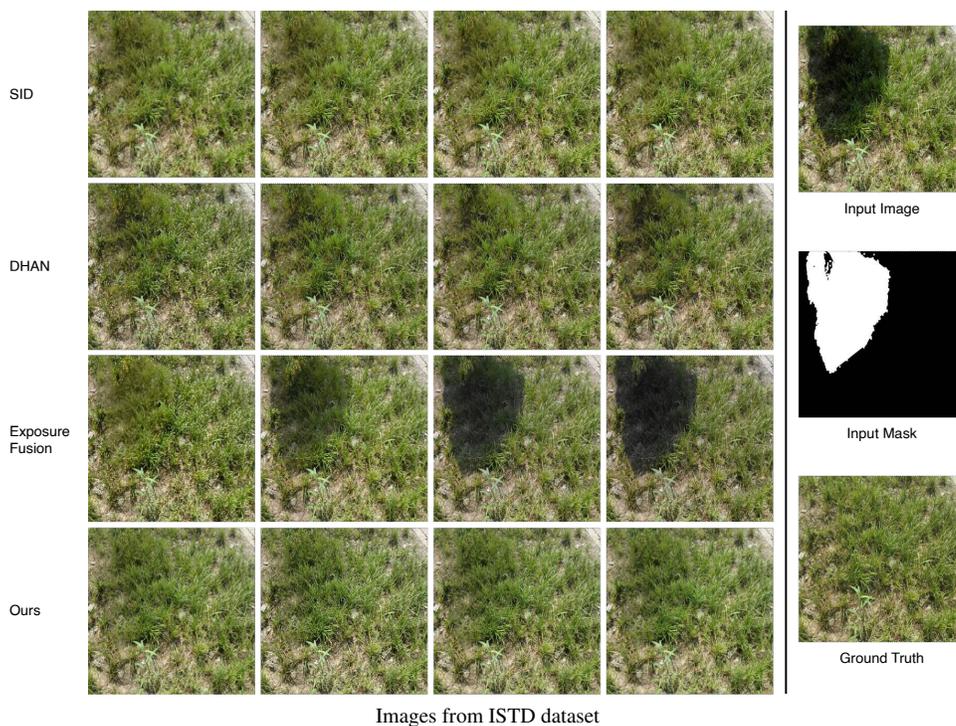
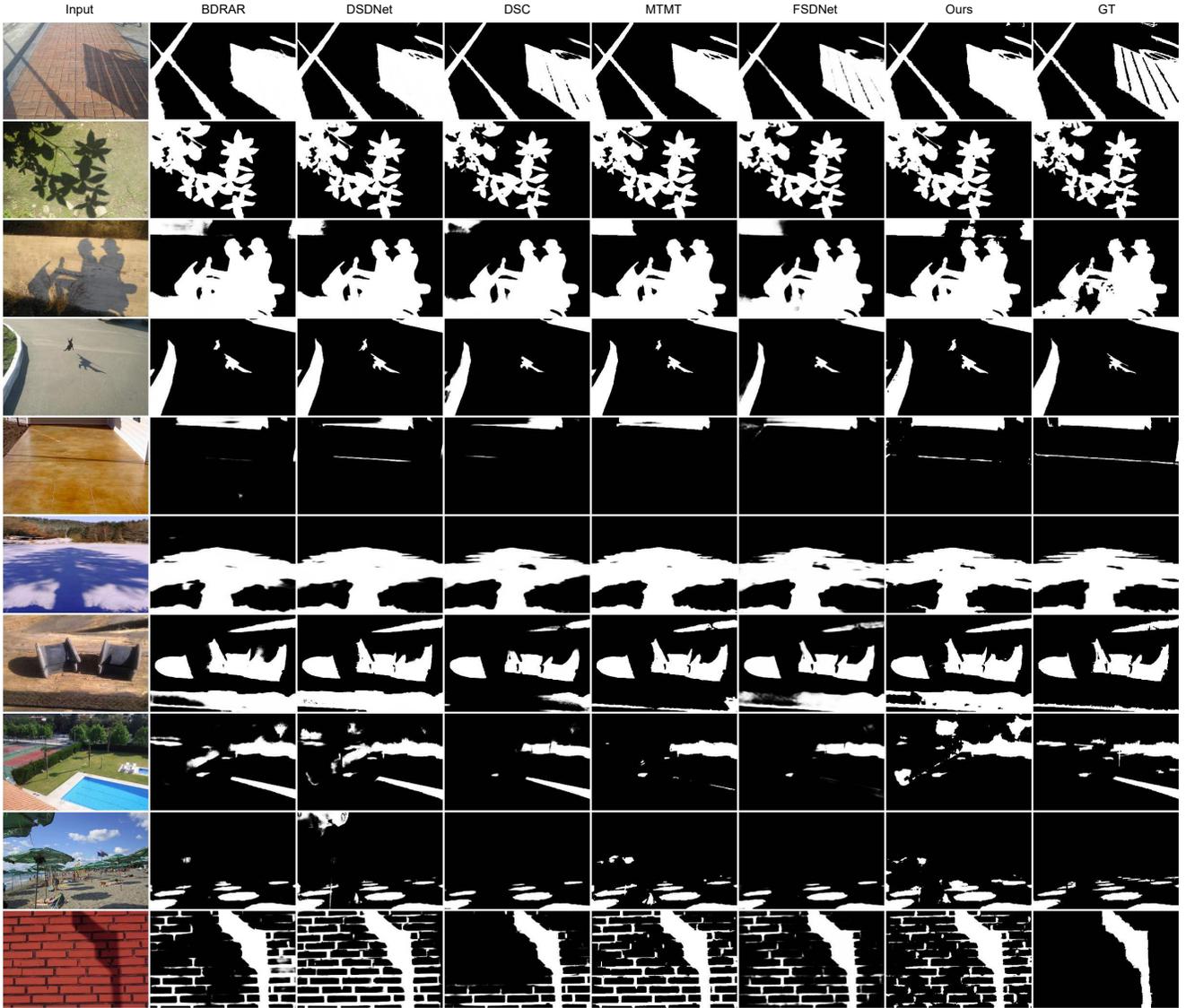


Figure 2: Qualitative results showing shadow removal results at progressively increasing image resolutions (left-to-right)(256, 512, 1024, 2048) between SID [6], DHAN [2], AEF [3] and the proposed method. As shown in the figure, shadow removal quality degrades as the image resolution is increased for existing methods. In contrast, our proposed method retains the shadow removal quality even at high-resolution. **(Zoom-in for better visualization)**



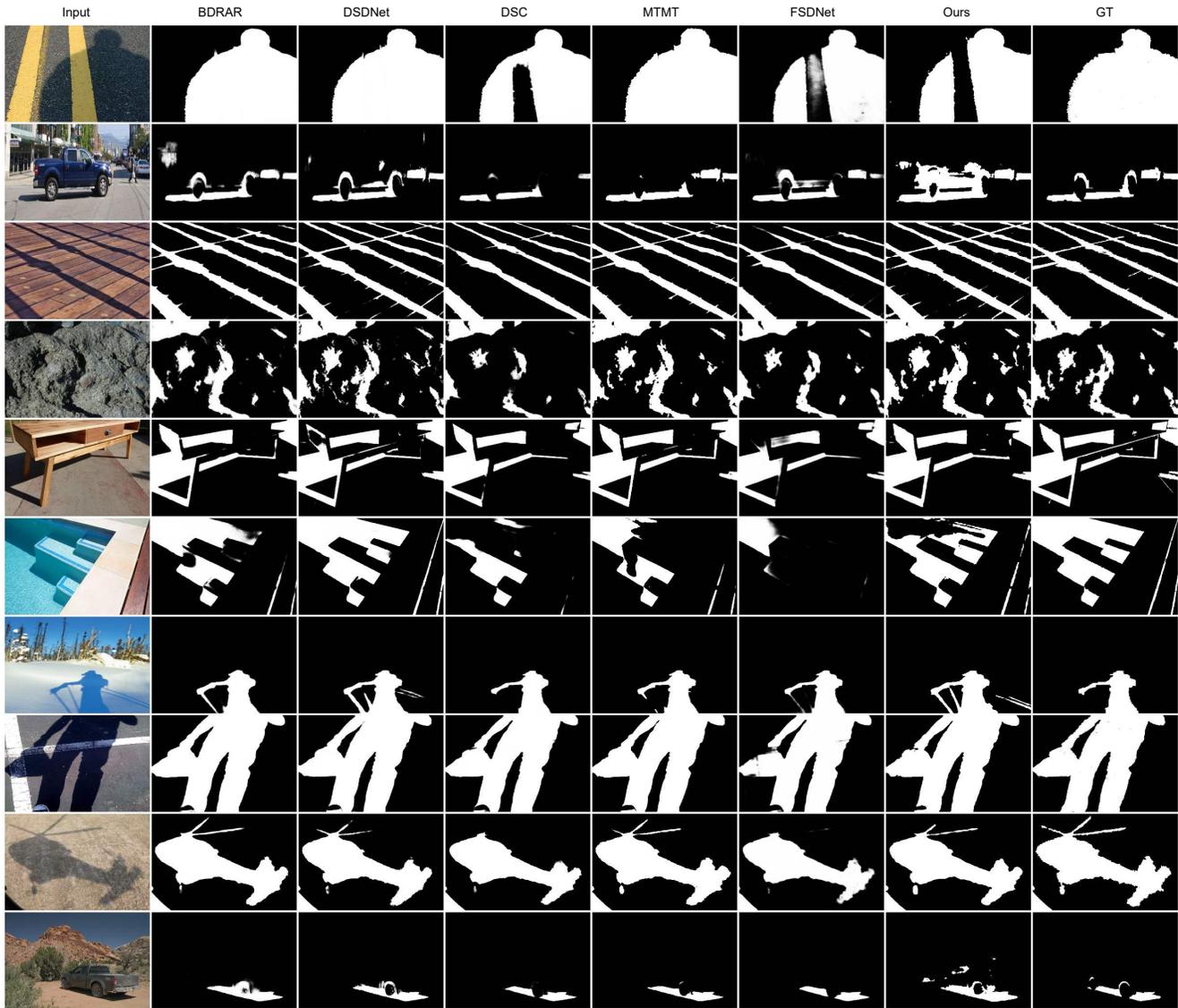


Figure 3: Qualitative results of our proposed Shadow Detection method, compared with existing state-of-the-art methods BDRAR [8], DSDNet [7], DSC [5], MTMT-Net [1], FSDNet [4].



Input



Output-Detector



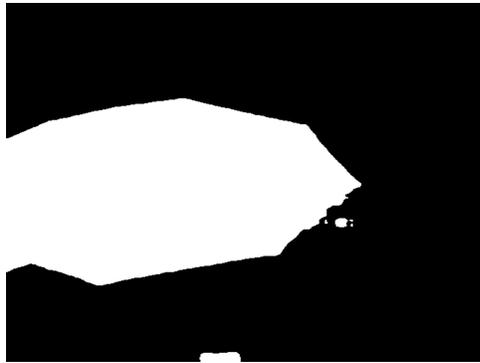
Output-Removal



Ground-Truth



Input



Output-Detector



Output-Removal



Ground-Truth



Input



Output-Detector



Output-Removal



Ground-Truth



Input



Output-Detector



Output-Removal



Ground-Truth

Figure 4: End-to-end Shadow Removal results using proposed Shadow Detector and Remover networks (SHARDS) on High-Resolution images from ISTD-HQ and SFHQ datasets (**Zoom-in for better visualization**).