ShadowHack: Hacking Shadows via Luminance-Color Divide and Conquer

Supplementary Material

7. Color Bias Analysis

We can observe that shadows in most images tend to appear bluish. This occurs because when an object blocks sunlight, the primary light source in the shadow area becomes the light scattered from the sky, which has shorter wavelengths and typically appears blue [7, 40]. Additionally, the color of the shadow is influenced by surrounding objects due to diffuse reflection. In Fig. 3 of the main manuscript, we present the color shift statistics of shadow regions in different datasets. Specifically, the color bias is obtained by subtracting shadow regions' reflectance map of the input from its GT counterpart. Through statistical analysis, we confirm the blue-red color shift in shadow regions (note that these statistics are primarily derived from outdoor scenes; in other complex scenarios such as indoor environments, the color bias may exhibit different distributions due to the influence of object surface colors); this observation guides our decomposition strategy as elaborated in the main manuscript.

8. Implementation Details

Our ShadowHack is implemented in Python with the PyTorch framework and is trained using NVIDIA RTX4090 GPUs. We utilize the AdamW optimizer with momentum set to (0.9, 0.999) and a weight decay of 10^{-2}. The initial learning rate is set to 2×10^{-4} and decays to 10^{-6} at the schedule's end in a cosine annealing manner. During training, we randomly crop the training data into patches of size 384×384, and apply data augmentation techniques, including rotation, flipping, mixup, and color jittering as in previous works [13, 32]. The input image is decoupled into the luminance and color components through a conversion from RGB color space to YCbCr color space via Kornia. For the LRNet, we set batch size to 4 and total iterations to 400k on the ISTD+ [27] dataset, while batch size to 8 and total iterations to 300k on the SRD [43] dataset. For the CRNet, the batch size is set to 4 for both datasets and trained for 100k iterations. Following most of the previous arts [13, 32, 33, 53, 65], we test our mode with full-resolution inputs and only resize the outputs for 256×256 to make a fair comparison with previous methods when evaluating. For the base network of each stage, we use a four-layer encoder-decoder structure (from L1 to L4), consisting of three downsampling and three upsampling operations. The dimension of the first layer is 32. In the LRNet, our RTA modules are integrated into L3 and L4, with an overlapping ratio set to 0.5 and a rectified mechanism λ_0 value of 0.7. For the CRNet, we bring a color encoder on the top of the U-shape network. The color encoder is an FC-MAE pretrained atto variant of ConvNext-v2, with merely 2M parameters.

λ_0	0.5	0.6	0.7	0.8
PSNR	33.77	33.78	33.89	33.71

Table 5. Ablation study of λ_0 on the ISTD+ dataset. Here we calculate PSNR on the luminance with full resolution.

Suggested by [58], in Eq. 10 of the main manuscript, λ can be re-parameterize by:

$$\lambda = \exp(\lambda_1^1 \cdot \lambda_1^2) - \exp(\lambda_2^1 \cdot \lambda_2^2) + \lambda_0, \tag{11}$$

where $\lambda_i^j, i, j \in \{1, 2\}$ and λ_0 are learnable and predefined parameters, respectively. λ_0 is set to 0.7 according to Tab. 5.

9. Failure Case Analysis

Multiple checkpoint training utilizes outputs from both premature and peak performance checkpoints of LRNet. It creates robustness for CRNet against failures in LRNet's Y prediction. The ConvNext-v2 color encoder further provides rich semantic clues, enabling the identification of same-object regions and referencing their respective colors. Fig. 12 shows CRNet adapting to up to 50% contamination in Y failures (contaminated by mixing input Y with LRNet's Y prediction). However, complete Y prediction failure (e.g., in extreme out-of-distribution scenarios) still impairs final shadow removal.

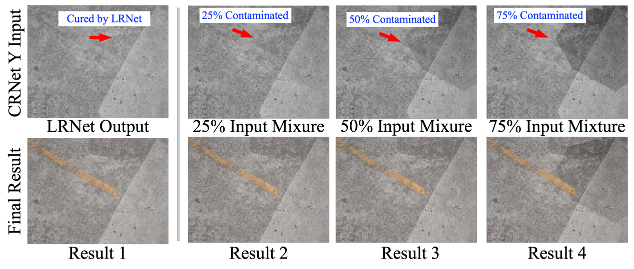


Figure 12. Generalization analysis for Failed Y Prediction

10. Additional experiment on ISTD Dataset

ISTD comparison as in Tab. 6. Our method achieves the best performance on all metrics. We will add this to the supplementary in the next revision.

Method	SF'23	RASM'24	HF'24	OmniSR+GM'25	Ours
PSNR↑	32.21	32.32	32.02	31.56	**33.15**
SSIM↑	**0.968**	**0.968**	**0.968**	0.965	**0.968**
RMSE↓	4.09	4.12	4.24	-	**3.80**

Table 6. Quantitative results on the ISTD dataset.

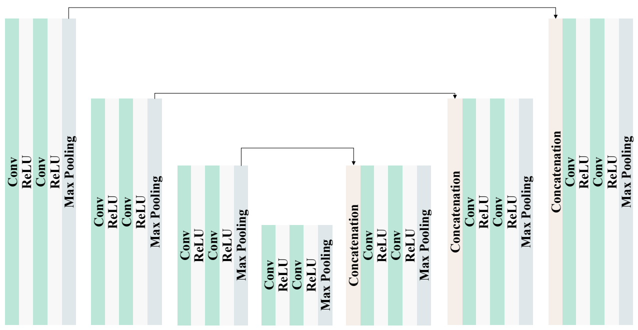


Figure 13. The decoupling model of Retinex-based framework.

11. Architectures of Extension Experiments

11.1. Ablation Study of Decoupling Framework

During the ablation study on our decoupling framework, we explore the end-to-end network and the Retinex-based framework. For the end-to-end network, we adopt the same baseline as our models. With an input concatenated by RGB images and its shadow masks, the network outputs shadow-free images directly. For the Retinex-based experiment, we first decompose the image to its illumination map and reflectance map. As shown in Fig. 13, the decomposing model consists of a U-Net [44] and is trained with the loss function of L1 loss and WTV loss [17]. We obtain illumination maps from the U-Net and their reflectance maps via division. Subsequently, we relight the illumination and remove the color bias in reflectance with the same architecture as LRNet.

11.2. Mask Refine Network

We make corrupted masks inspired by shadow counterfeiting in SILT [56], including expansion, corrosion, noise and distraction generated by random polygons. The mask refine network simply adopts a U-Net in Fig. 13. We take I_t and shadow mask as inputs and the refined mask will be generated. Considering the complexity of shadow characteristics, we inject ConvNext-v2 atto model [37] as our CRNet does. The extracted features of RGB inputs are fed into the U-Net decoder for a more precise mask.

Please note that we did NOT use mask refinement in any COMPARISON with competitors to keep it fair. It is only used in Sec. 5.4 of the main manuscript to show robustness and address handling inaccurate masks in real-world scenarios.

12. Comparison of Model Size

As shown in Tab. 7, we provide the total size of our models compared with state-of-the-art models. Our method achieves the best performance with a comparable parameter count.

Method	Params (M)	PSNR↑	RMSE↓
ShadowFormer [13]	11.4	32.90	4.04
ShadowDiffusion [14]	55.2	34.73	3.63
Li *et al.* [30]	23.9	33.17	3.83
DeS3 [24]	113.7	34.11	3.72
RASM [32]	5.2	34.46	3.37
Liu *et al.* [34]	117.9	33.48	3.66
Homoformer [53]	17.8	35.37	3.33
OmniSR+GM [55]	360.2	34.56	-
Ours	23.3	35.94	2.90

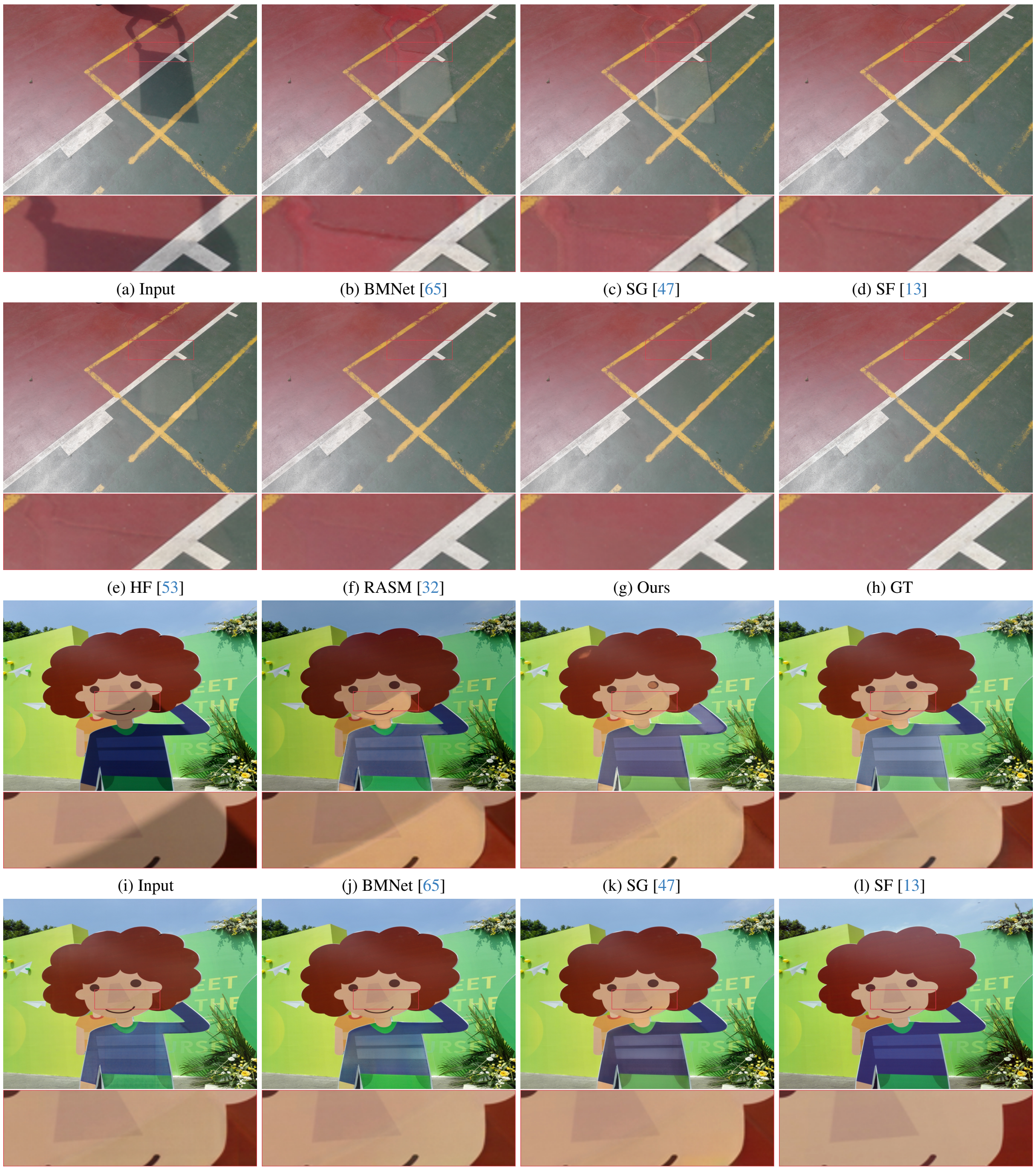
Table 7. Comparisons of model size and performance on the SRD dataset with state-of-the-art methods.

13. More Visual Results

We display more visual comparisons on the ISTD+ [27] and SRD [43] datasets in Figs. 14, 15, 16 and 17. Additionally, Fig 18 brings extra UIUC and UCF datasets results.

14. Flexible Shadow Removal

We evaluate our model with flexible shadow masks, as shown in Fig. 19, which demonstrates the high flexibility of our model in removing shadows based on user-specified masks.



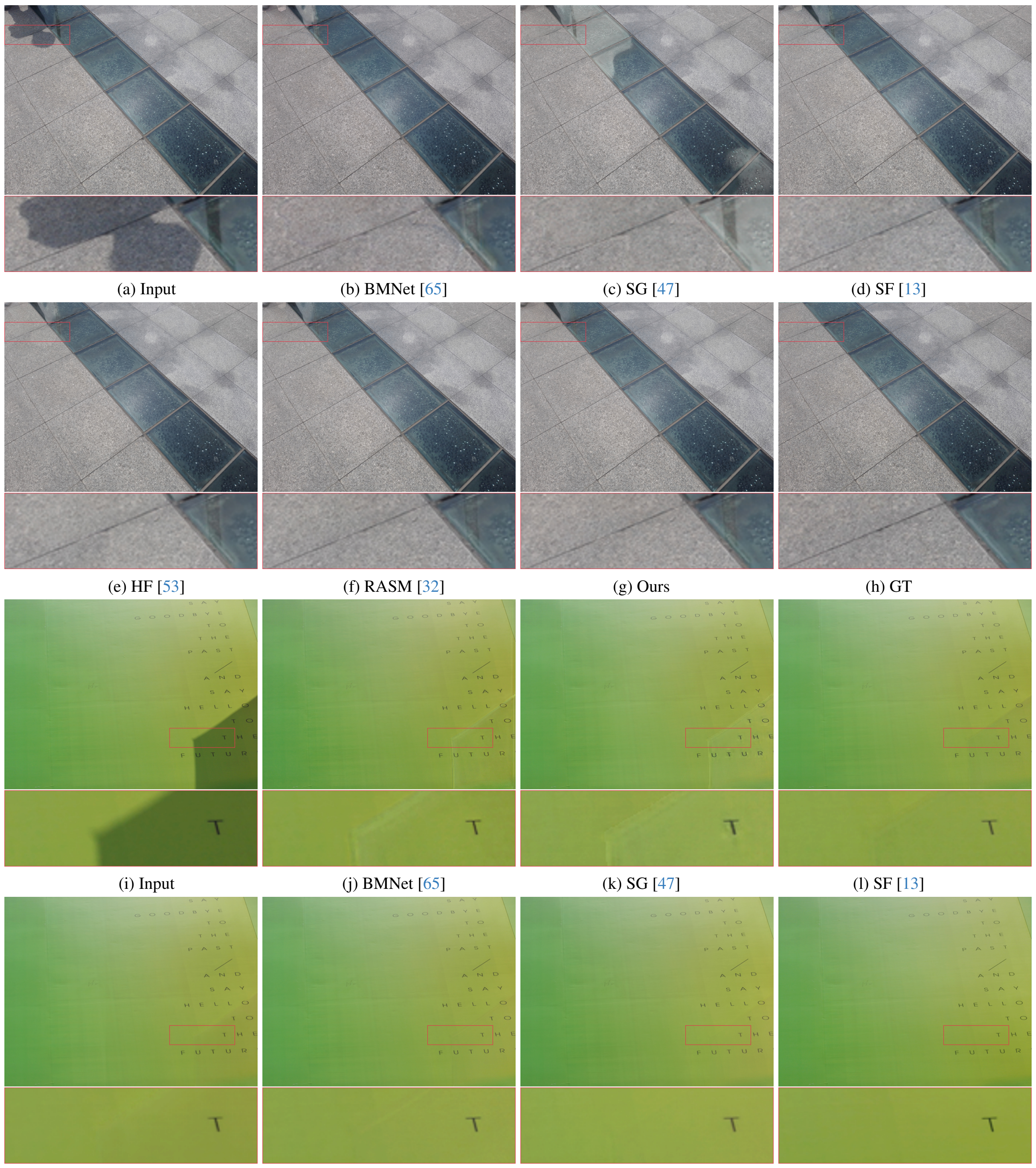
(a) Input (b) BMNet [65] (c) SG [47] (d) SF [13]

(e) HF [53] (f) RASM [32] (g) Ours (h) GT

(i) Input (j) BMNet [65] (k) SG [47] (l) SF [13]

(m) HF [53] (n) RASM [32] (o) Ours (p) GT

Figure 14. Visual comparisons on the ISTD+ dataset.



(a) Input (b) BMNet [65] (c) SG [47] (d) SF [13]

(e) HF [53] (f) RASM [32] (g) Ours (h) GT

(i) Input (j) BMNet [65] (k) SG [47] (l) SF [13]

(m) HF [53] (n) RASM [32] (o) Ours (p) GT

Figure 15. Visual comparisons on the ISTD+ dataset.

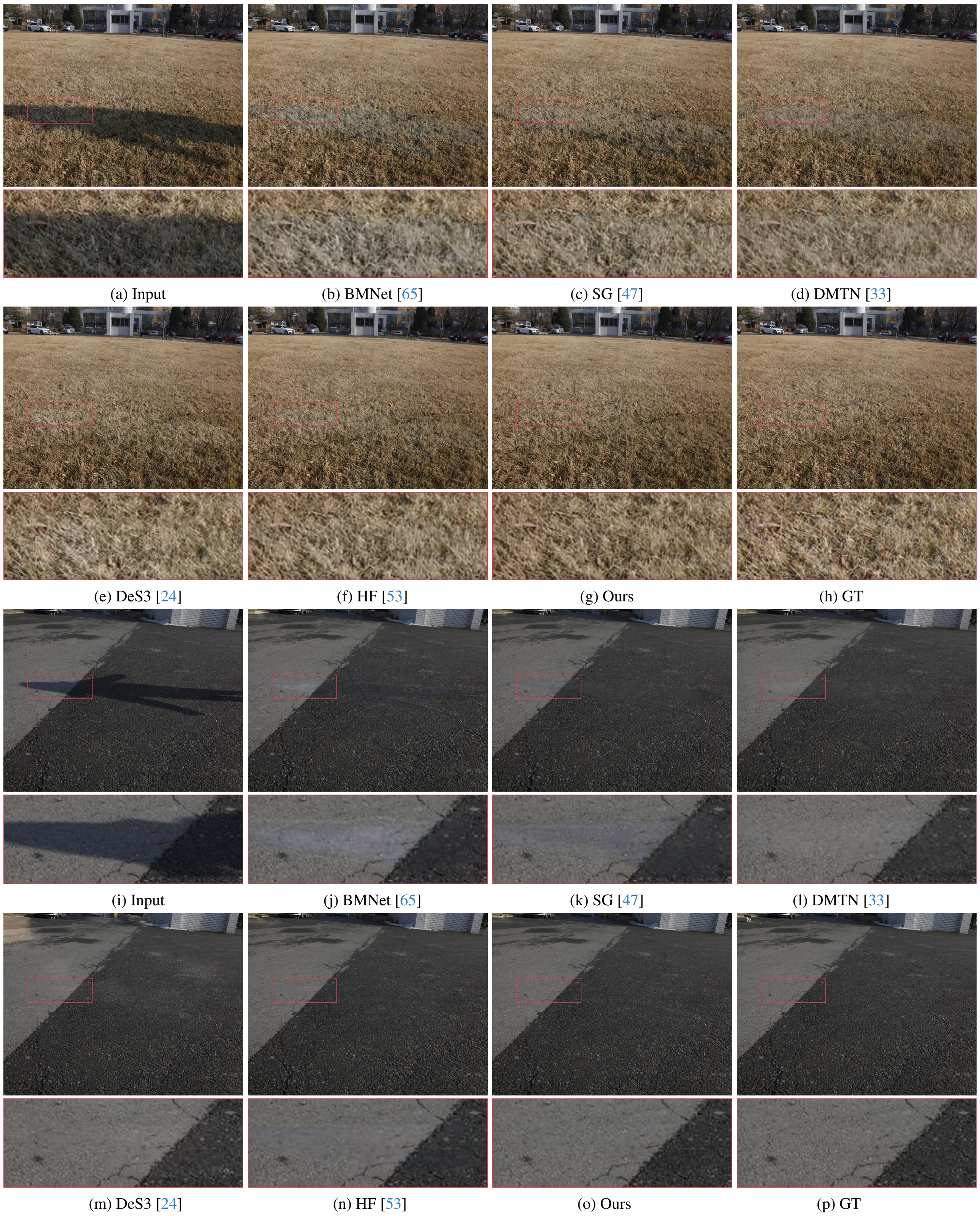
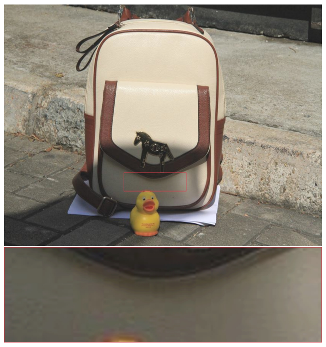
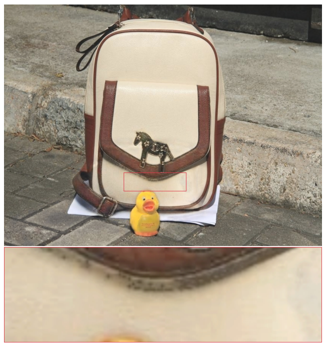
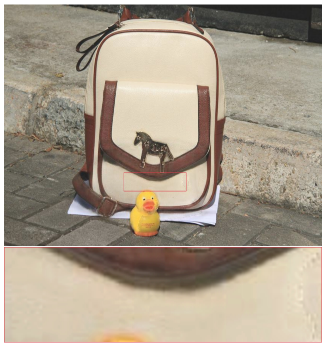
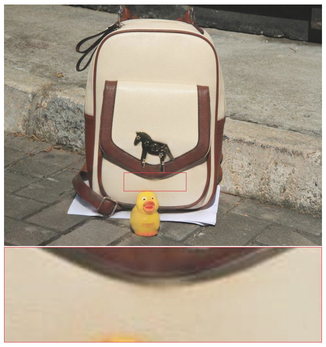


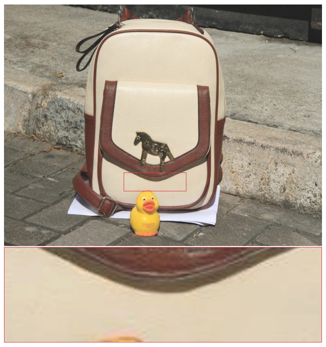
Figure 16. Visual comparisons on the SRD dataset.

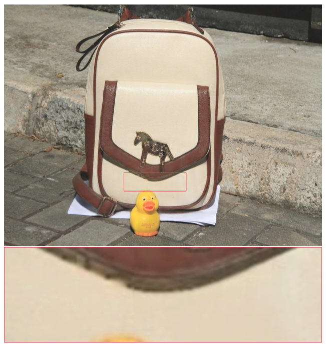

(a) Input

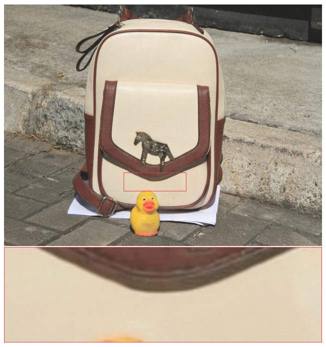

(b) BMNet [65]

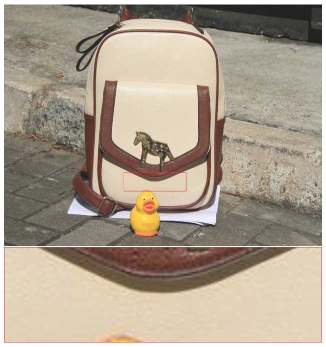

(c) SG [47]

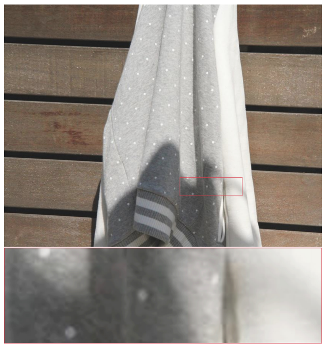

(d) DMTN [33]

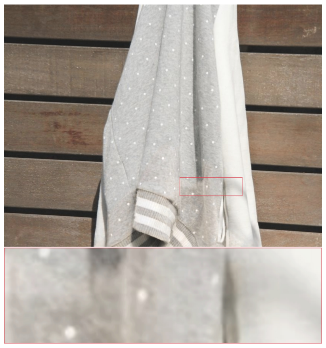

(e) DeS3 [24]

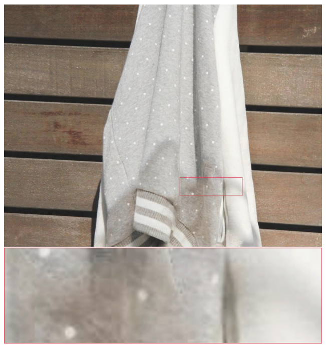

(f) HF [53]

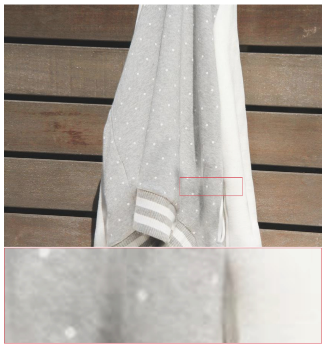

(g) Ours

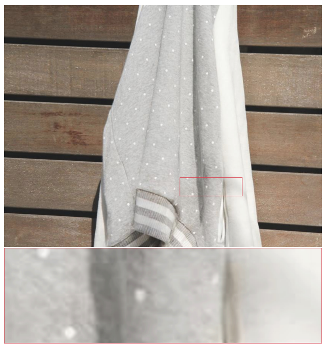

(h) GT

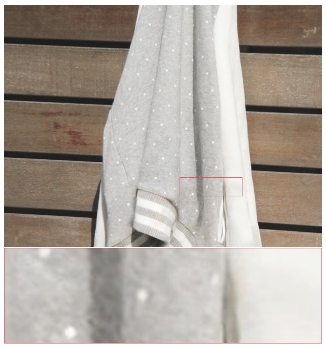

(i) Input

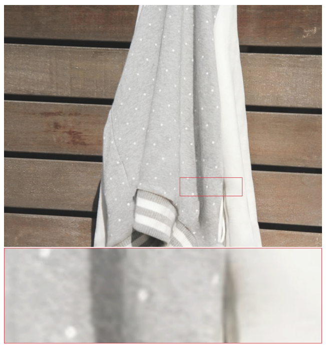

(j) BMNet [65]


(k) SG [47]


(l) DMTN [33]


(m) DeS3 [24]


(n) HF [53]


(o) Ours

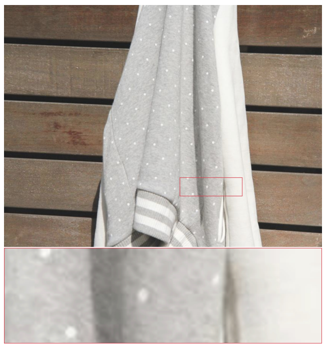

(p) GT

Figure 17. Visual comparisons on the SRD dataset.

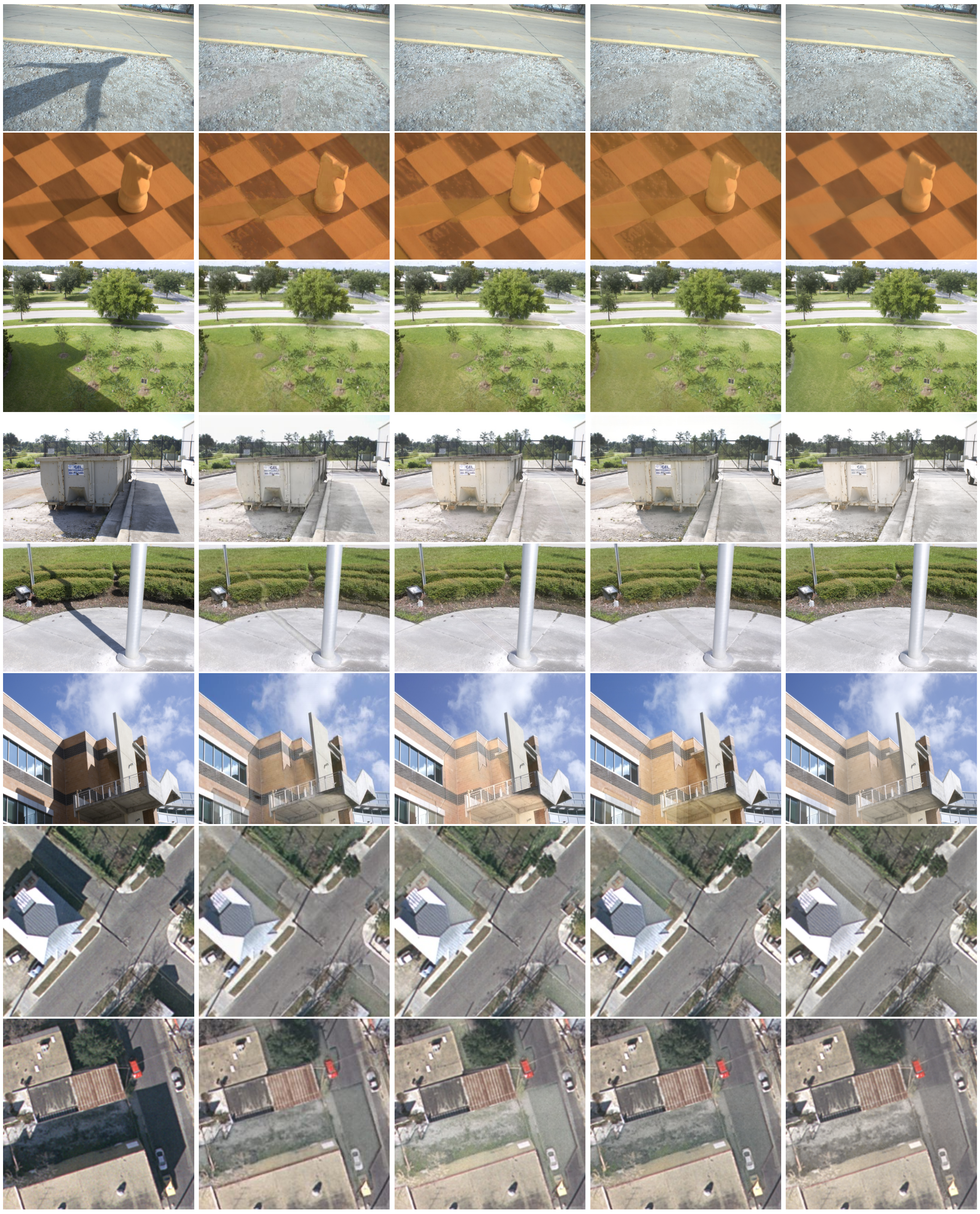

(a) Input (b) SG [47] (c) SF [13] (d) HF [53] (e) Ours

Figure 18. Visual comparisons on the UIUC and UCF datasets.

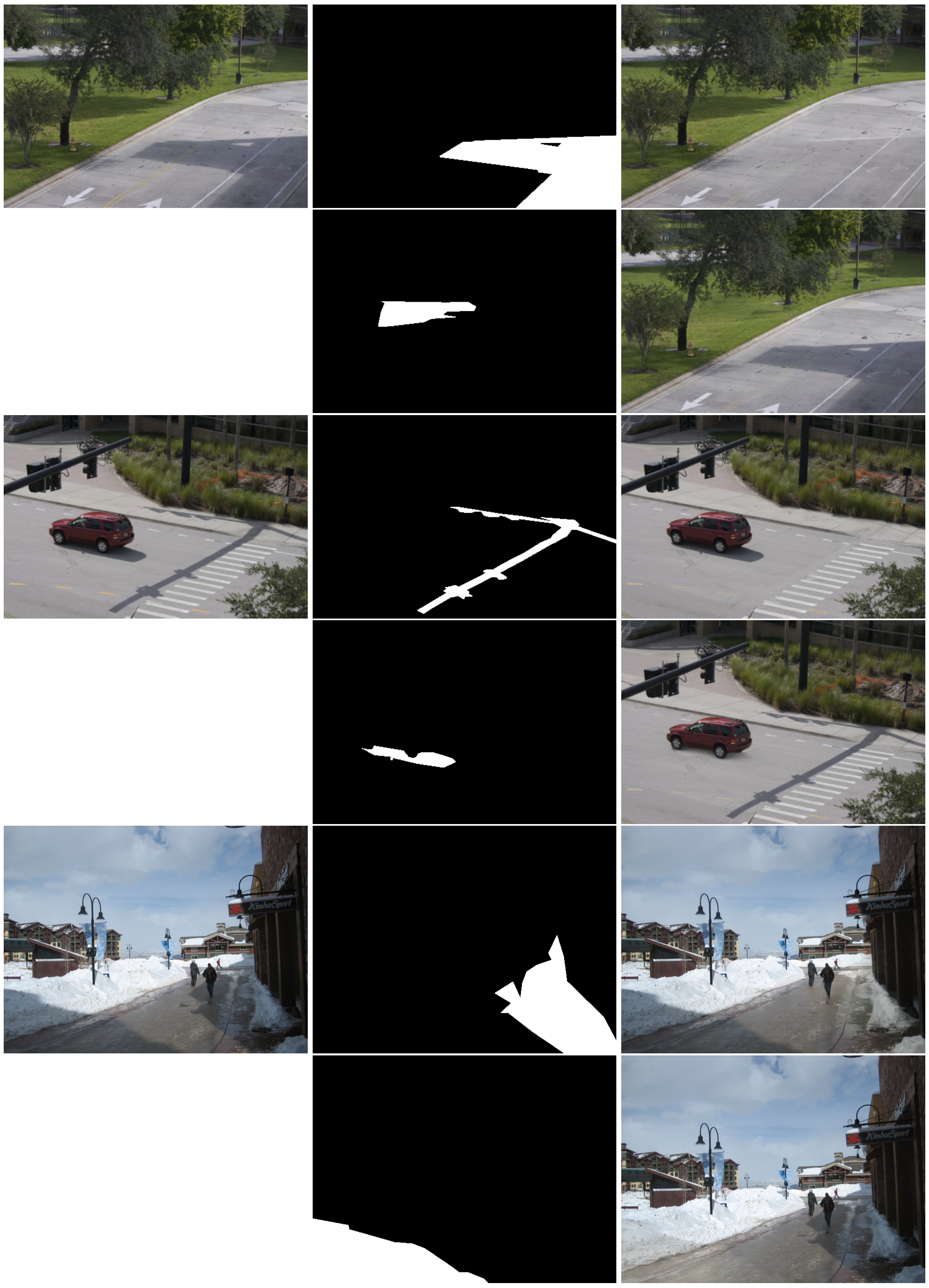


Figure 19. Visual results of selected shadow masks on the UCF dataset.