Adaptive Illumination Mapping for Shadow Detection in Raw Images

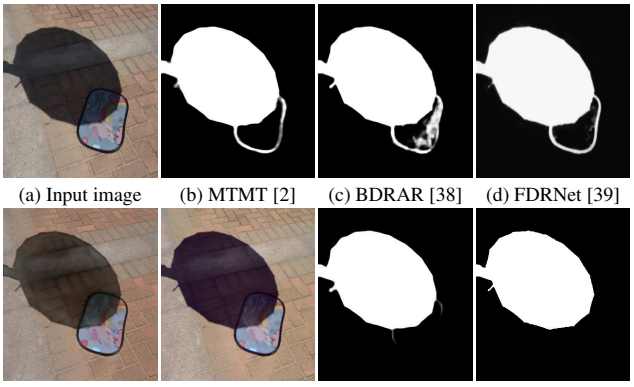
Jiayu Sun[1,2†] Ke Xu[2†] Youwei Pang[1] Lihe Zhang[1‡]
Huchuan Lu[1] Gerhard Hancke[2] Rynson Lau[2‡]
[1]School of Information and Communication Engineering, Dalian University of Technology
[2]Department of Computer Science, City University of Hong Kong
{jiayusun666, kkangwing}@gmail.com, lartpang@mail.dlut.edu.cn,
{zhanglihe, lhchuan}@dlut.edu.cn, {gp.hancke,rynson.lau}@cityu.edu.hk

Abstract

Shadow detection methods rely on multi-scale contrast, especially global contrast, information to locate shadows correctly. However, we observe that the camera image signal processor (ISP) tends to preserve more local contrast information by sacrificing global contrast information during the raw-to-sRGB conversion process. This often causes existing methods to fail in scenes with high global contrast but low local contrast in shadow regions. In this paper, we propose a novel method to detect shadows from raw images. Our key idea is that instead of performing a many-to-one mapping like the ISP process, we can learn a many-to-many mapping from the high dynamic range raw images to the sRGB images of different illumination, which is able to preserve multi-scale contrast for accurate shadow detection. To this end, we first construct a new shadow dataset with $\sim$ 7000 *raw images and shadow masks. We then propose a novel network, which includes a novel adaptive illumination mapping (AIM) module to project the input raw images into sRGB images of different intensity ranges and a shadow detection module to leverage the preserved multi-scale contrast information to detect shadows. To learn the shadow-aware adaptive illumination mapping process, we propose a novel feedback mechanism to guide the AIM during training. Experiments show that our method outperforms state-of-the-art shadow detectors. Code and dataset are available at* `https://github.com/jiayusun/SARA`.

1. Introduction

Whenever there are light and objects in a scene, there are shadows. Although usually unnoticed, shadows can tell a lot of information about the scene, *e.g.*, the shapes, volumes and locations of objects, and the sources, directions of lights. The analysis of shadows can facilitate a lot of applications varying from scene understanding to the creation and assessment of novel scenes. Hence, it is essential to design effective shadow detection models.



(a) Input image (b) MTMT [2] (c) BDRAR [38] (d) FDRNet [39]
(e) sRGB image A (f) sRBG image B (g) Ours (h) Ground truth

Figure 1. When a non-shadow region has low contrast to a shadow region but high contrast to another non-shadow region (*e.g.*, the black boundary of the object), existing methods may fail to detect the shadow region correctly (b-d). We propose to detect shadows in the raw images. By learning to project raw images into sRGB images of different intensity ranges adaptively, our method can detect the shadow region correctly (g).

There are many methods proposed for shadow detection. Conventional methods [5, 4, 14, 37, 7] typically handcraft statistical features (*e.g.*, pixel color, intensity, gradient, and/or a combination of them) to represent the characteristics of shadows, and design heuristic models to detect shadows. The key limitation of these methods is that their low-level features often fail to represent shadows in real-world scenes. Deep learning-based shadow detectors avoid the hand-craft feature engineering. In order to maximize the discrepancies of deep features learned from shadow and non-shadow regions, they typically focus on designing neural networks to mine both global and local contrast information via, *e.g.*, directional context aggregation [10], cross-layer feature fusion [38, 33], negative example supervi-

[†]Equal Contribution
[‡]Corresponding Authors

sion [35], semi-supervised learning [2] and shadow feature decomposition and reweighing [39].

Despite their success, these methods may still fail to detect shadow correctly. As shown in Figure 1(b-d), they wrongly detect black boundary regions as shadows. Such error-prone regions typically have high contrasts to surrounding non-shadow regions and low contrasts to shadow regions. While capturing these images, we realize that humans are able to perceive the small-scale contrast information, *i.e*., the black boundary outside the shadow region should still have notable contrast to the shadow regions. However, such contrast information is diminished in the camera-finishing sRGB image, as the camera ISP is essentially a many-to-one operation of information reduction. In order to preserve the global contrast in the output sRGB image, the raw-to-sRGB conversion of ISP has to sacrifice the local contrasts. This often fails existing shadow detection methods and cannot be easily addressed by using the camera-finishing sRGB images.

In this paper, we propose a novel method that detects shadows from raw images. The key insight is that we can learn a many-to-many mapping between the high dynamic range raw data and the standard dynamic range sRGB images of different illumination conditions, which can preserve different scales of contrast information for shadow detection. To this end, we first construct the first ShAdow RAw (SARA) dataset consisting of $\sim$ 7000 raw image and mask pairs for training and evaluation. We then propose a novel network to detect shadows from raw images. Our network has a novel adaptive illumination mapping (AIM) module, which learns to project the linear raw images into sRGB images of different intensity ranges (Figure 1(e,f)). A shadow detection module then detects shadows by modeling multi-scale contrast information derived from the AIM. We propose a feedback mechanism to guide the AIM during training to render sRGB images of different illumination in a shadow-aware manner, which maximizes the discrepancies between shadow and non-shadow regions and facilitates shadow detection performance. As shown in Figure 1(g), our method detects the shadow regions more accurately compared to existing methods.

This work has the following three main contributions:

- We propose a novel method to detect shadows from raw images, which, unlike previous methods, is able to model multi-scale shadow/non-shadow contrast information derived from the linear raw data for robust shadow detection.
- We propose a novel network with two novelties: (1) a novel AIM to produce images of different scales of intensity ranges; and (2) a feedback mechanism to guide the AIM to generate multi-scale contrast information in a shadow-aware manner.
- We construct the first SARA dataset with 7019 raw images and their corresponding shadow masks, to facilitate the learning process.

Extensive experiments show that the proposed method performs favorably against state-of-the-art methods.

2. Shadow Detection Methods

Conventional shadow detection methods typically design physical-based shadow models [5, 4] and leverage classical machine learning techniques to classify shadow pixels [14, 37, 7]. These methods hand-craft low-level features (*e.g*., intensity [7, 37, 11, 30], edge [5, 14, 37, 11], chromacity [5, 4, 14, 7, 30], and texture [37, 7, 30]) to represent the shadows, which may fail in complex real-world scenes.

Deep learning has advanced shadow detection performance significantly due to its capability of learning shadow representations from a large number of images. Existing methods typically design different network architectures to model global and local contrast information for shadow detection. Nguyen *et al*. [21] mine shadow features by formulating the shadow detection in the generative adversarial manner [19]. Wang *et al*. [32] propose a stacked conditional GAN [19] network to learn shadow detection and removal jointly. Hu *et al*. [10, 9] propose a direction-aware attention mechanism to aggregate spatial information in a local-to-global way. Zhu *et al*. [38] propose the bidirectional feature pyramid network that fuses features of every two adjacent layers to obtain local and global contrast information. Zheng *et al*. [35] propose a distraction-aware method to improve the shadow detection performance by learning to predict false positive and true negative areas. Chen *et al*. [2] propose a multi-task learning method to exploit shadow counts and edge, and formulate a semi-supervised method to exploit additional unlabeled images. Zhu *et al*. [39] show that previous deep methods are biased to the intensity contrasts and propose a feature decomposition and reweighting mechanism to learn robust shadow representations by re-adjusting the importance of shadow cues.

All these methods detect shadows from sRGB images. In this paper, we propose the first method to detect shadows from raw images with a new dataset, by learning a many-to-many mapping to model multi-scale contrast information for detecting shadows.

3. SARA Dataset

We propose the first ShAdow RAw (SARA) dataset with 7019 raw images and corresponding shadow masks for training and evaluation. The raw images are split into 6143 images for training and 876 images for test. Figure 2 show some examples of our dataset.

Diversity. We consider diverse scenes with different foreground objects casting shadows and background objects

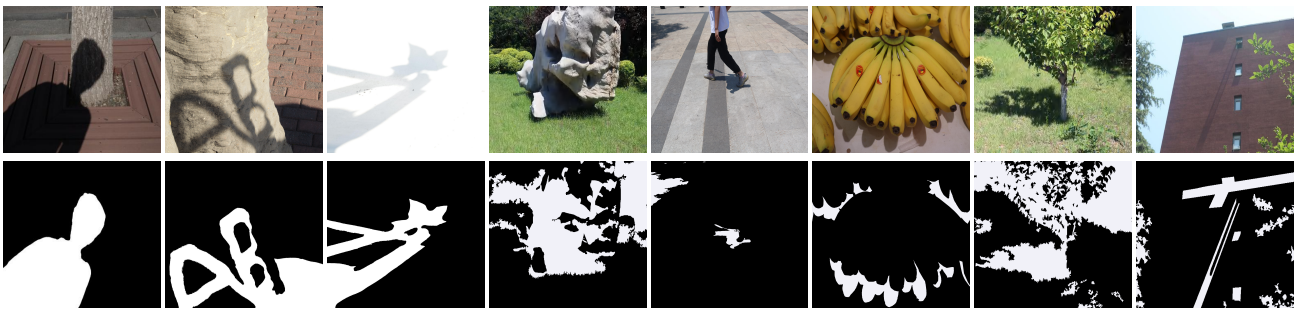


Figure 2. Examples of our dataset.

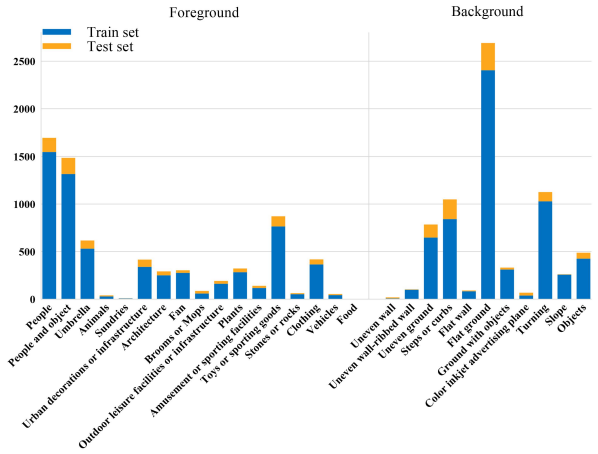


Figure 3. Object categories distribution of our dataset.

that shadows are projected to. To cast shadows of diverse shapes, we consider 17 categories including humans, animals and a variety of objects such as umbrellas, toys, and trees. We consider 11 different background scenes, including flat (*e.g.*, roads, sports fields and flat walls) and non-planar background surfaces (*e.g.*, grass fields, stairs and trees). Figure 3 summarizes the object categories of our dataset.

Imaging Setting. To avoid introducing additional noise into the capturing, we take the following measures: (1) We use a heavy-duty tripod (8 kg) to ensure the camera stability and avoid possible camera shakes. (2) To reduce the slight displacement of the camera caused by pressing the shutter, we trigger the camera shutter button using a mobile phone in remote. (3) To maintain the exposure consistency when taking images, we use the full manual shooting mode of the camera (canceling the automatic metering, focusing, exposure compensation and white balance) and manually adjust the aperture and shutter speed. (4) In order to reduce noise, we use the low ISO (100).

If the foreground object is portable, we take one shadow image and one shadow-free image (by removing the object(s) that casts shadow). For scenes with unportable shadowing objects, we only take the shadow image.

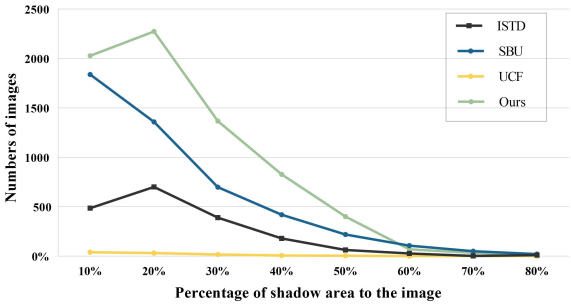


Figure 4. Shadow area proportions of different datasets.

Label Generation. For scenes with both shadow and shadow-free images, we first compute and binarize the difference maps between the shadow and shadow-free image pairs, to obtain the initial shadow masks. We then ask two groups of volunteers, one group for refining the labels and the other group for checking the accuracies. For scenes with shadow images only, their shadow masks are directly labeled via one group of volunteers and checked by the other group.

Dataset Statistics. We analyse three kinds of statistical information for a better understanding of our dataset.

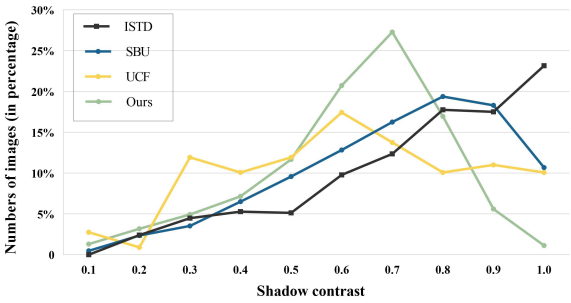
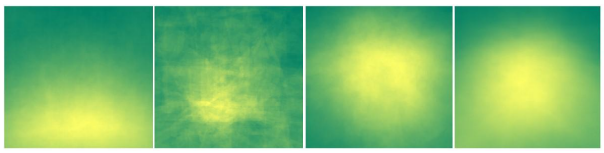


Figure 5. Shadow contrast distributions of different datasets.

1) Shadow Area Proportion. We compute the proportions of shadow pixels to the whole image. Figure 4 shows the line plot of shadow area ratios in our dataset, ISTD [32], SBU [29], and UCF [37], from which we can see that our



(a) SBU [29] (b) UCF [37] (c) ISTD [32] (d) Ours

Figure 6. Shadow location distribution of different datasets.

dataset contains more shadows of small and medium areas.

2) Shadow Contrast Distribution. We compute the $\mathcal{X}^2$ distance of the color histograms between shadow areas and non-shadow areas. Figure 5 plots the shadow contrast distributions of our dataset and existing datasets, from which we can see that our dataset tends to have less shadows with high contrasts, indicating that our dataset is more challenging.

3) Shadow Location Distribution. We also study the spatial locations distribution of shadows in Figure 6. We can see that while shadows in SBU [29] and UCF [37] tend to concentrate on the bottom of images, those in ISTD [32] are more cluttered in the center of images. In contrast, shadows in our dataset occupy the images in a mroe smooth and even manner.

4. Proposed Method

We propose to detect shadows in the raw domain due to its two main advantages against the camera-finishing sRGB images. First, raw images are high dynamic ranges that preserve more information than low dynamic range sRGB images. Second, the intensities of raw images are linearly proportional to the scene radiance, which can be manipulated to produce sRGB images of different intensity ranges.

To this end, we propose a novel neural approach as shown in Figure 7. Our network consists of a novel shadow-aware adaptive illumination mapping (AIM) module and a simple and effective shadow detector. The AIM module takes a demosaiced 4-channel raw image I_r as the input and predicts two intermediate sRGB images (I_{s1} and I_{s2}) with different intensity ranges. The detector then predicts the single-channel shadow map S_m by exploiting multi-scale contrast information from the intermediate sRGB images. We also exploit a feedback mechanism to guide the AIM to generate multi-scale contrast information in a shadow-aware manner.

Adaptive Illumination Mapping Module. The intensity of the raw image is linearly proportional to the scene radiance. While the dynamic range of the raw image is high, which preserves more scene information than the sRGB image, the contrast is lower. We propose the adaptive illumination mapping (AIM) module to project the raw images into sRGB images to enhance the contrast. Unlike previous tone mapping algorithms [20, 25, 31, 26, 28, 3, 6, 23] that sacrifice the global contrast for more local contrast, our AIM module is able to predict a wider range of contrast by learning to predict sRGB images of different intensity ranges.

Specifically, we build the AIM module with an encoder-decoder architecture and adopt VGG-16 [27] as the encoder to generate a compact latent representation from the demosaiced 4-channel raw image I_r. We then assign two decoders of identical structures with skip links to decode the latent representation into a pair of sRGB images. In the last decoder layer, we first expand the summation of I_r and the frontal decoder feature to 64 channels and then shrink it to 3 channels for sRGB prediction by applying pixel-wise convolution layers. To diversify the illumination, we propose the following loss term to enhance and constrain the output images as:

$$\mathcal{L}^t_{aim} = \sum_{i=1}^{2} L_1(I^t_{si} - I_{srgb}) + \lambda \sum_{i=1}^{2} L_1(M^\tau_i I^t_{si} - M^\tau_i I_{srgb}), \tag{1}$$

where t represents the situation before and after the feedback mechanism. I_{srgb} is the sRGB image that corresponds to the raw image I_r in our dataset, λ is the balancing hyper-parameter. The first term is the L_1 loss between the individual contrast-enhanced image I^t_{si} and the sRGB image, aiming to guide the module to perform the lumination mapping process. The second term is utilized to enhance their diversities by emphasizing the loss on different intensity levels of images, respectively. We use M^τ_i to indicate the different intensity levels on images, as it is a soft binary mask generated by a learnable threshold τ on the luminance space Y of the image, as:

$$M^\tau_1 = Sig(\beta(Y - \tau)), \tag{2}$$

$$M^\tau_2 = 1 - M^\tau_1, \tag{3}$$

where Sig is the sigmoid function. And β is an amplification factor. M^τ_2 is the reverse mask of M^τ_1, which separate the intensity level. Guided by $\mathcal{L}_{aim}$, AIM brings sRGB images of different intensity ranges. We visualize M^τ_1, I^1_{s1}, M^τ_2, I^1_{s2} and the intensity histograms of I^1_{s1} and I^1_{s2}, respectively in Figure 8.

Shadow Detection Module. Unlike previous shadow detection methods, our shadow detection module aims to exploits the high dynamic range and wide color gamut, and the multi-scale contrasts of different intensity ranges, images derived from the raw images. To this end, we concatenate the demosaiced raw image I_r and the contrast-enhanced sRGB images I^1_{s1} and I^1_{s2} to form a 10-channel input to our shadow detection module. Our shadow detection module also adopts the encoder-decoder architecture that first transforms the input into a compact latent representation, then decodes it into a two-channel feature, which is then splited into initial shadow mask S^1_m and a shadow error map S_e inspired by [12].

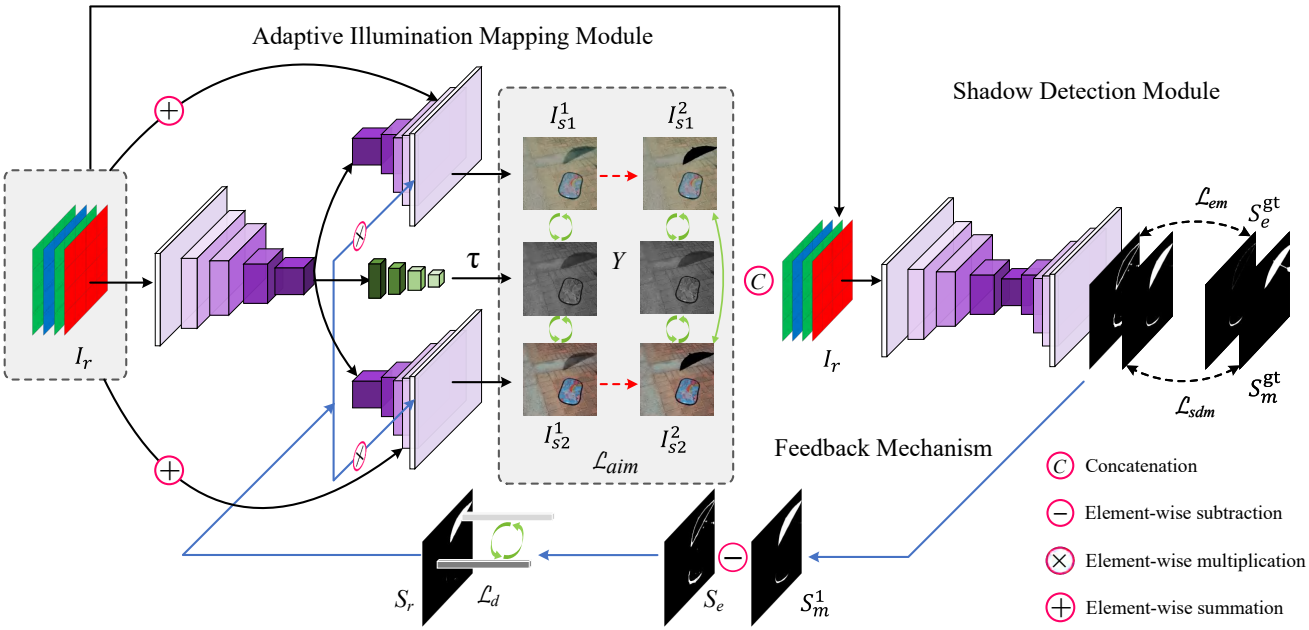
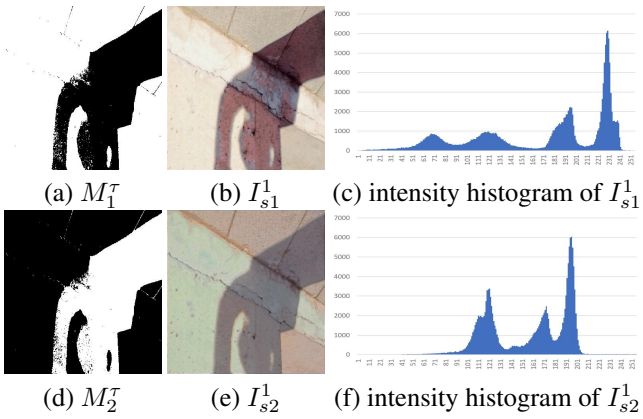


Figure 7. Overview of the proposed method. Given a demosaiced 4-channel raw image, we first feed it to our AIM module to produce two sRGB images (I_{s1}, I_{s2}) of different intensity ranges. They are then concatenated with raw image and are fed into the shadow detection module to detect shadows. A feedback mechanism is proposed to train the AIM module to be shadow-aware, with a diversity loss to avoid I_{s1}, I_{s2} being identical.



(a) M_1^τ (b) I_{s1}^1 (c) intensity histogram of I_{s1}^1

(d) M_2^τ (e) I_{s2}^1 (f) intensity histogram of I_{s2}^1

Figure 8. Visualization of M_1^τ (a), I_{s1}^1 (b), M_2^τ (d), and I_{s2}^1 (e) of Eq.1. (c) and (f) are the intensity histograms of (b) and (e), respectively. It shows that the AIM module learns to project the raw image into sRGB images of different intensity distributions.

The combination of the initially predicted mask and error map is used to guide the AIM module to perform shadow-aware illumination mapping process, which results in another pair of sRGB images I_{s1}^2 and I_{s2}^2 that maximize the discrepancies between the shadow and non-shadow regions. These images are then fed into the shadow detection module to produce the final shadow mask S_m^2.

Feedback Mechanism. The feedback mechanism aims to leverage the initial shadow detection map S_m^1 and the shadow error map S_e to guide the AIM module to maximize the contrast information between shadow and non-shadow regions.To this end, we first subtract the initial shadow detection map S_m^1 with the shadow error map S_e, to obtain a more robust shadow map S_r. And the feedback mechanism can be formulated as follows: (1) We first count the numbers of shadow and non-shadow pixels in S_r via:

$$N_s = Avgpool(S_r) \times H \times W, \tag{4}$$

$$N_{ns} = Avgpool(1 - S_r) \times H \times W, \tag{5}$$

where N_s, N_{ns}, H and W indicate the numbers of shadow and non-shadow pixels, height and width of the shadow map, respectively. $Avgpool$ is global average pooling.

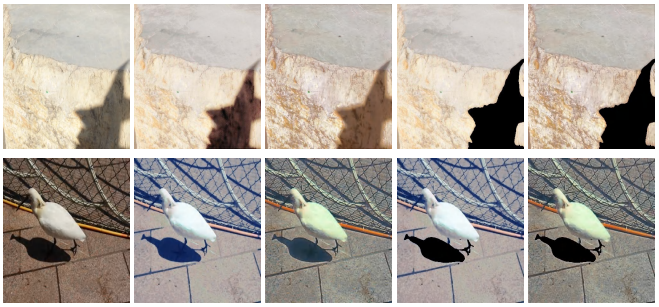
(2) We then extract the features inside and outside the shadow regions for two illumination mapping decoders, respectively, by multiplying the S_r with the features of the last layer of the decoder. (3) For each pair of illumination mapping features inside/outside the shadow regions, we compute two compact representations V_s and V_{ns} as:

$$V_s = \frac{Avgpool(f \times S_r) \times H \times W}{N_s}, \tag{6}$$

$$V_{ns} = \frac{Avgpool(f \times (1 - S_r)) \times H \times W}{N_{ns}}, \tag{7}$$

where f represents the 4th layer's features from one illumination mapping decoder.

(4) Given the two compact representations of shadow and non-shadow regions, we propose to use the following



(a) I_{rgb} (b) I^1_{s1} (c) I^1_{s2} (d) I^2_{s1} (e) I^2_{s2}

Figure 9. Results of our adaptive illumination mapping module before (the second and third s) and after (the fourth and fifth columns) the feedback mechanism.

loss term to maximize their discrepancies in the feature domain as:

$$\mathcal{L}_d = \sum_{j=1}^{3} Mean(Abs(V_s^j + V_{ns}^j)). \quad (8)$$

To supervise the learning of shadow predictions and error maps, we first compute the ground truth error map S_e^{gt} by subtracting the initial shadow detection map S_m^1 with the ground truth shadow map S_m^{gt}. We then compute the L_2 loss between S_e and S_e^{gt}. To synthesize the local structure information, We supervise the learning of our shadow detection module via L_{sdm} which is the sum of weighted balanced binary cross entropy loss and weighted IoU loss [34].

Figure 9 show the results of our AIM module before (the second and third columns) and after (the fourth and fifth columns) the feedback mechanism. We can see that the each group (each columns) has the diversity, and the contrasts between shadow and non-shadow regions are effectively enhanced by the feedback mechanism.

Inference. The inference process of our model can be described as: Given a demosaiced raw image I_r as input, the AIM predicts a pair of sRGB images I^1_{s1} and I^1_{s2} with different intensity ranges. These sRGB images and I_r are concatenated together and fed into the shadow detection module to obtain an initial shadow map S_m^1 and a shadow error map S_e. S_m^1 and S_e are used to predict a more robust shadow map S_r to guide the AIM to produce another pair of sRGB images I^2_{s1} and I^2_{s2}. They are concatenated with I_r as the input of the shadow detection module to produce the final shadow detection map S_m^2. The whole inference takes around $0.03s$ on one GTX 3090 GPU when processing a raw image of 400×400 resolution.

5. Experiments

Implementation Details. The proposed model is implemented on the Pytorch framework [24], and tested on a PC with an i7 4GHz CPU and a GTX 3090 GPU. The shadow detection module uses the ConvNeXt [17] as the backbone. As we train our model, the input images are resized to a resolution of 400 × 400. The network parameters are initialized randomly. In addition to the standard augmentation strategies, *i.e.*, cropping and flipping, we use multi-scale training to make the image fluctuates between image size 0.75 and 1.5. Batch size is set to 4. For loss minimization, we first train our network without feedback mechanism for 25 epochs. It adopts the ADAM optimizer [13]. Both AIM and SDM use an initial learning rate of $1e^{-4}$ and AIM divided by 10 for every 10 epochs. SDM divided by 10 at the 20^{th} epoch. We then train our network with feedback mechanism for another 25 epochs. The λ in Eq. 1 is set to 10 in our experiment.

Table 1. Quantitative comparison of our method with the state-of-the-art methods on the proposed dataset. We report the error rates for shadow region and non-shadow region as well as the balanced error rate (BER). The best and second best results are marked in **bold** and underlined. Method* indicates that the method is trained using raw images.

Methods	Year	Evaluation Metrics		
		BER↓	Shadow↓	Non-shadow↓
ST-CGAN [32]	2018	10.93	13.75	8.11
DSC [10]	2018	10.35	14.23	6.46
ADNet [15]	2018	7.58	10.33	4.83
BDRAR [38]	2018	3.91	4.07	3.76
BDRAR* [38]	2018	4.34	**2.82**	5.87
SID [1]	2018	7.71	8.83	6.59
SID* [1]	2018	8.9	12.15	5.65
DSD [35]	2019	5.63	4.20	7.07
ITSD [36]	2020	3.85	6.12	1.59
MTMT [2]	2020	3.65	4.36	2.95
MTMT* [2]	2020	4.79	6.61	2.94
FDRNet [39]	2021	7.55	9.14	5.97
FDRNet* [39]	2021	5.46	8.31	2.61
SAMNet [16]	2021	9.13	16.93	1.34
LGSL [18]	2021	14.59	27.97	1.22
Ours	-	2.97	4.77	**1.16**
Ours*	-	**2.61**	3.74	1.47

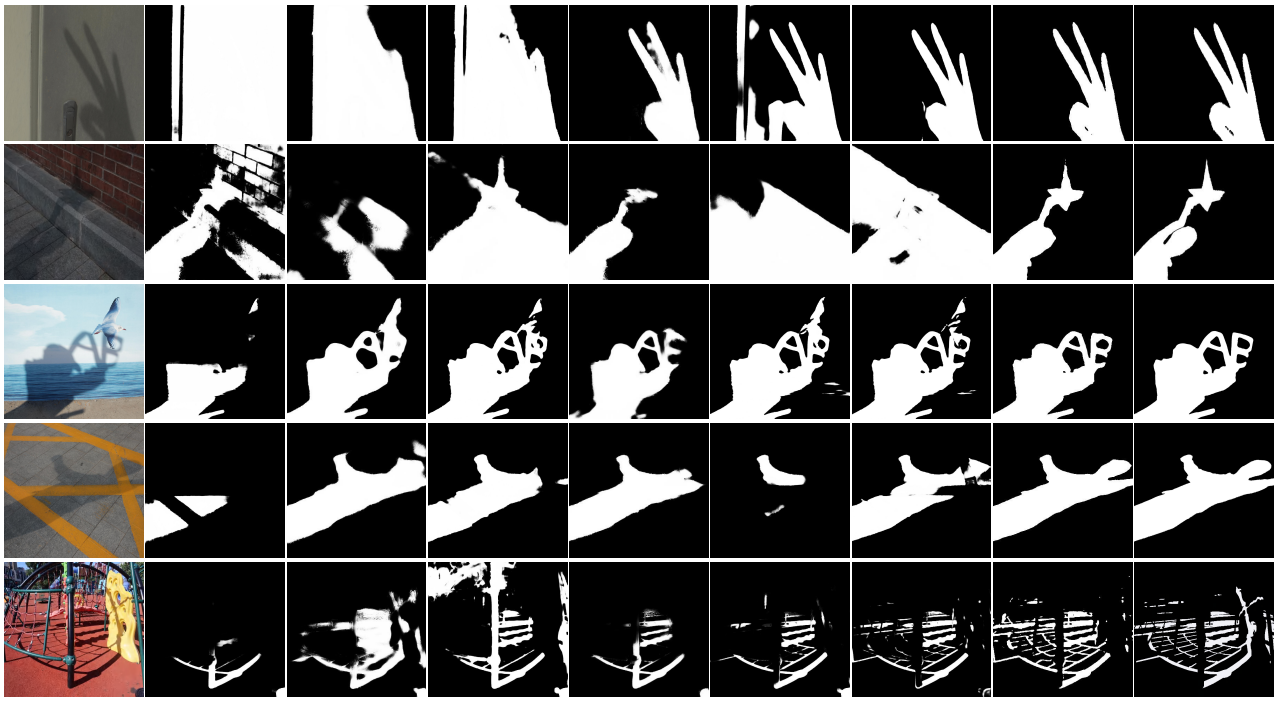
Evaluation Methods, Dataset and Metric. We compare our method to 11 state-of-the-art deep methods, including 7 shadow detection methods: ST-CGAN [32], DSC [10], BDRAR [38], ADNet [15], DSD [35], MTMT [2], FDRNet [39]; 3 salient object detection methods: SAMNet [16], LGSL [18], ITSD [36]; and 1 raw image tone mapping and denoising method: SID [1]. We evaluate our method on the proposed dataset, which contains both raw image-mask and sRGB image-mask pairs (5718 pairs respectively). Following previous shadow detection methods, we adopt the Balanced Error Rate (BER) as the evaluation metric:

$$BER = (1 - \frac{1}{2}(\frac{TP}{TP+FN} + \frac{TN}{TN+FP})) \times 100, \quad (9)$$

where FN, FP, TN and TP indicates the numbers of false negative, false positive, true negative and true positive shadow pixels, respectively.

Table 2. Quantitative comparison between ours and the state-of-the-art methods on three sRGB image-based shadow detection datasets. For each dataset, we list the error rates for shadow region and non-shadow region, and the balanced error rate (BER). The best results are marked in **bold**. (*) MTMT is trained with extra unlabelled data; (**) DSD is trained with extra supervision from other models.

Methods	Year	SBU [29]			UCF [37]			ISTD [32]		
		BER↓	Shadow↓	Non Shad.↓	BER↓	Shadow↓	Non Shad.↓	BER↓	Shadow↓	Non Shad.↓
Unary-Pariwise [7]	2011	25.03	36.26	13.80	-	-	-	-	-	-
stacked-CNN [29]	2016	11.00	8.84	12.76	13.00	8.84	12.76	8.60	7.69	9.23
scGAN [22]	2017	9.10	8.39	9.69	11.50	7.74	15.30	4.70	3.22	6.18
patched-CNN [8]	2018	11.56	15.60	7.52	-	-	-	-	-	-
ST-CGAN [32]	2018	8.14	3.75	12.53	11.23	**4.94**	17.52	3.85	2.14	5.55
DSC [10]	2018	5.59	9.76	**1.42**	10.54	18.08	**3.00**	3.42	3.85	3.00
BDRAR [38]	2018	3.64	3.40	3.89	7.81	9.69	5.44	2.69	**0.50**	4.87
ADNet [15]	2018	5.37	4.45	6.30	9.25	8.37	10.14	-	-	-
DC-DSPF [33]	2019	4.90	4.70	5.10	7.90	6.50	9.30	-	-	-
DSD** [35]	2019	3.45	3.33	3.58	7.59	9.74	5.44	2.17	1.36	2.98
MTMT* [2]	2020	3.15	3.73	2.57	7.47	10.31	4.63	1.72	1.36	2.08
FDRNet [39]	2021	3.04	**2.91**	3.18	7.28	8.31	6.26	1.55	1.22	1.88
Ours	-	**2.87**	3.64	2.10	**7.01**	9.43	4.61	**1.18**	1.05	**1.31**



(a) Input Image (b) ADNet [15] (c) BDRAR [38] (d) DSD [35] (e) FDR [39] (f) MTMT [2] (h) ITSD [36] (i) Ours (j) GT

Figure 10. Qualitative comparison of the proposed method with the most recent state-of-the-art methods on our dataset.

5.1. Comparing to the State-of-the-arts

Quantitative Comparisons. We conducted our experiments on our dataset, as well as three existing sRGB image-based datasets, *i.e.*, SBU [29], UCF [37] and ISTD [32]. Table 1 reports the quantitative comparisons between our method and 11 existing methods. We first demonstrate the effectiveness of detecting shadows in raw images against that in the sRGB images. We train two versions of proposed model, *i.e.*, one using the sRGB images and the other one using raw images, denoted as Ours and Ours* (last two rows in Table 1). We can see that our raw model outperforms our sRGB model with a notable advantage in the shadow regions (3.74 *v.s.* 4.77). Our sRGB model performs slightly better than our raw model in non-shadow regions (1.16 *v.s.* 1.47) as the sRGB images are better in preserving contrast information in the regions of medium intensities (*i.e.*, not too dark or too bright) than raw images. Nonetheless, our raw model achieves a higher shadow detection performance in terms of the overall BER score (2.61 *v.s.* 2.97), which verifies the effectiveness of using raw images in our method.

Third, we note that not all models are suitable for using raw images to detect shadows. While BDRAR [38] would

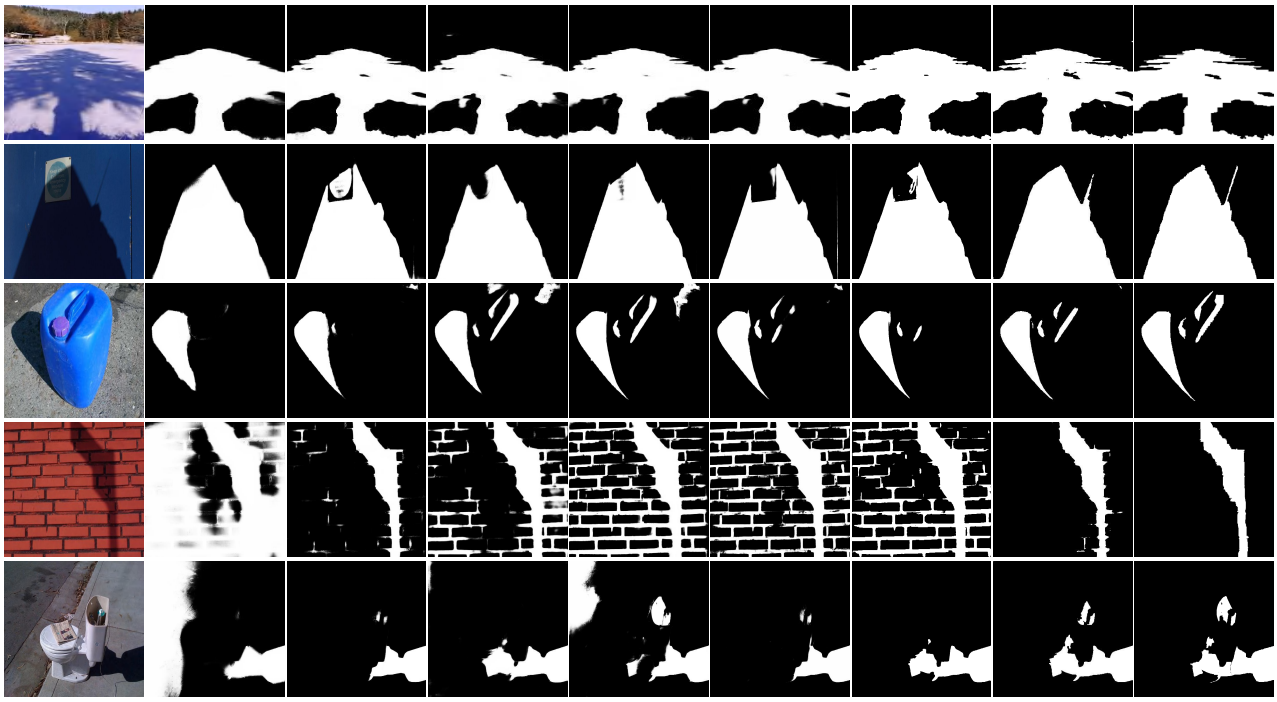


Figure 11. Qualitative comparison of the proposed method with the most recent state-of-the-art methods on sRBG image-based shadow detection dataset.

also enjoy the benefits of using raw images (comparing the 4th and 5th rows), the detection performance of MTMT [2] using raw images is worse than that of using sRGB images. We note that the key reason to the success of MTMT [2] is to incorporate additional unlabeled shadow images in a semi-supervised manner. This explains the performance drop as our dataset does not contain additional unlabelled data for performing the semi-supervised learning.

Last, we can see that one state-of-the-art salient object detection method LGSL [18] (the last third row) can performs quite good in the non-shadow regions (1.22) but surprisingly worse in the shadow regions (14.59). We hypothesize that since salient object detection methods are designed to detect visually distinctive objects, they tend to favor objects/regions of medium intensities where most colors/textures exist. A similar phenomenon can be seen in the last fourth row, where the SAMNet [16] behaves very similar to the LGSL [18]. This shows that shadow detection task is fundamentally different from the salient object detection, and directly training salient object detection models for shadow detection cannot address the shadow detection problem. Similarly, directly applying the classic raw image processing method SID [1] is also not an ideal solution (see the sixth row). These comparisons generally verify the effectiveness of our model design. Table 2 shows the quantitative results on the three sRGB-based benchmarks. We can see that our method achieves the best BER scores over all state-of-the-art methods, on the three datasets. Compared to the second best-performing method, FDRNet [39], our method reduces the BER scores by 5.59%, 3.71%, and 23.8%, respectively. This demonstrates the effectiveness of our proposed method on sRGB images.

Table 3. Ablation studies on the proposed adaptive illumination mapping module (AIM), feedback mechanism (FM), the uses of $\mathcal{L}_d$ and error map (EM). We report the overall BER scores on our test set of raw images.

	AIM	**FM**	L_d	**EM**	**BER↓**
SDM	×	×	×	×	3.19
SDM+AIM	✓	×	×	×	2.92
SDM+AIM+FM	✓	✓	×	×	2.86
SDM+AIM+FM	✓	✓	✓	×	2.70
Ours	✓	✓	✓	✓	**2.61**

Visual Comparisons. Figure 10 shows some challenging scenes in that shadows have low contrasts to the non-shadow regions. We can see that our method has clear visual advantages over existing shadow detection methods on challenging scenes, *e.g.*, the scene where shadows with complex shapes projected onto a non-flat wall corner in a dark scene (second row). While these scenes are typically challenging for existing methods, our method can detect the shadow regions correctly. These visual comparisons generally verify our idea of detecting shadows from raw images. In Figure 11, we provide visual comparison

Table 4. Ablation studies on the proposed adaptive illumination mapping module (AIM), feedback mechanism (FM), the uses of $\mathcal{L}_d$ and error map (EM). We report the overall BER scores on the test set of SBU [29] dataset.

	AIM	**FM**	L_d	**EM**	**BER↓**
SDM	×	×	×	×	3.39
SDM+AIM	✓	×	×	×	3.01
SDM+AIM+FM	✓	✓	×	×	3.00
SDM+AIM+FM	✓	✓	✓	×	2.89
Ours	✓	✓	✓	✓	**2.87**

on sRBG image-based datasets, which further demonstrated our method can be extended to sRBG images.

5.2. Ablation Study

In order to verify our raw model designs, we performs ablation studies on the proposed adaptive illumination mapping module (AIM), feedback mechanism (FM), the use of $\mathcal{L}_d$ and the use of error map (EM). Table 3 reports the performance. We can see that by adding the adaptive illumination mapping module (AIM) with the shadow detection module (SDM), BER is reduced by around 8.46% (comparing the first two rows). This shows that the sRGB images of different intensity ranges produced by our adaptive illumination mapping module can significantly facilitate shadow detection performance. By comparing the 2nd and 3rd rows, we can see that the use of initial shadow detection results to guide the adaptive illumination mapping module further brings 2.05% performance gains. By further introducing the $\mathcal{L}_d$, the performance continuously grows by 5.59%. These two comparisons verify the design of the feedback mechanism in guiding the adaptive illumination mapping module to be shadow-aware. Last, by introducing the error map, a 3.33% performance gain is obtained. as the error map helps build robust shadow and non-shadow regions representations, which further facilitate the shadow detection process. We also report the ablation results on the sRGB image-based SBU [29] dataset, in Table 4.

6. Conclusion

In this paper, we have proposed a novel neural approach to detect shadows from raw images. Our network has a novel adaptive illumination mapping module to predict sRGB images of different intensity ranges, and a shadow detection module to exploit such illumination information to detect shadows. We have also proposed a novel feedback mechanism to guide the illumination mapping process in a shadow-aware manner. To facilitate the learning process, we have constructed a new dataset with raw images and corresponding shadow masks. Extensive experiments demonstrate that our method outperforms state-of-the-art shadow detection approaches.

Acknowledgements

This work was supported by the National Natural Science Foundation of China #62276046.

References

[1] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *CVPR*, 2018.

[2] 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.

[3] Zeev Farbman, Raanan Fattal, Dani Lischinski, and Richard Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM TOG*, 2008.

[4] Graham D Finlayson, Mark S Drew, and Cheng Lu. Entropy minimization for shadow removal. *IJCV*, 2009.

[5] Graham D Finlayson, Steven D Hordley, Cheng Lu, and Mark S Drew. On the removal of shadows from images. *IEEE TPAMI*, 2005.

[6] Bo Gu, Wujing Li, Minyun Zhu, and Minghui Wang. Local edge-preserving multiscale decomposition for high dynamic range image tone mapping. *IEEE TIP*, 2012.

[7] Ruiqi Guo, Qieyun Dai, and Derek Hoiem. Single-image shadow detection and removal using paired regions. In *CVPR*, 2011.

[8] Sepideh Hosseinzadeh, Moein Shakeri, and Hong Zhang. Fast shadow detection from a single image using a patched convolutional neural network. In *IROS*, 2018.

[9] 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. *IEEE TIP*, 2021.

[10] Xiaowei Hu, Lei Zhu, Chi-Wing Fu, Jing Qin, and Pheng-Ann Heng. Direction-aware spatial context features for shadow detection. In *CVPR*, 2018.

[11] Xiang Huang, Gang Hua, Jack Tumblin, and Lance Williams. What characterizes a shadow boundary under the sun and sky? In *ICCV*, 2011.

[12] Zhanghan Ke, Di Qiu, Kaican Li, Qiong Yan, and Rynson WH Lau. Guided collaborative training for pixel-wise semi-supervised learning. In *ECCV*, pages 429–445. Springer, 2020.

[13] P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv:1412.6980*, 2014.

[14] Jean-François Lalonde, Alexei A Efros, and Srinivasa G Narasimhan. Detecting ground shadows in outdoor consumer photographs. In *ECCV*, 2010.

[15] Hieu Le, Tomas F. Yago Vicente, Vu Nguyen, Minh Hoai, and Dimitris Samaras. A+d net: Training a shadow detector with adversarial shadow attenuation. In *ECCV*, 2018.

[16] Yun Liu, Xin-Yu Zhang, Jia-Wang Bian, Le Zhang, and Ming-Ming Cheng. SAMNet: Stereoscopically attentive multi-scale network for lightweight salient object detection. *IEEE TIP*, 2021.

[17] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *CVPR*, pages 11976–11986, 2022.

[18] Mingcan Ma, Changqun Xia, and Jia Li. Pyramidal feature shrinking for salient object detection. In *AAAI*, 2021.

[19] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. In *NeurIPS Workshop*, 2014.

[20] Rico Montulet, Alexia Briassouli, and N Maastricht. Deep learning for robust end-to-end tone mapping. In *BMVC*, 2019.

[21] Vu Nguyen, Tomas F. Yago Vicente, Maozheng Zhao, Minh Hoai, and Dimitris Samaras. Shadow detection with conditional generative adversarial networks. In *ICCV*, 2017.

[22] Vu Nguyen, Tomas F Yago Vicente, Maozheng Zhao, Minh Hoai, and Dimitris Samaras. Shadow detection with conditional generative adversarial networks. In *ICCV*, 2017.

[23] Sylvain Paris, Samuel W Hasinoff, and Jan Kautz. Local laplacian filters: edge-aware image processing with a laplacian pyramid. *ACM TOG*, 2011.

[24] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In *NeurIPS Workshop*, 2017.

[25] Aakanksha Rana, Praveer Singh, Giuseppe Valenzise, Frederic Dufaux, Nikos Komodakis, and Aljosa Smolic. Deep tone mapping operator for high dynamic range images. *IEEE TIP*, 2019.

[26] Erik Reinhard and Kate Devlin. Dynamic range reduction inspired by photoreceptor physiology. *IEEE TVCG*, 2005.

[27] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *ICLR*, 2015.

[28] Jack Tumblin and Holly Rushmeier. Tone reproduction for realistic images. *IEEE CGA*, 1993.

[29] Tomás Vicente, Le Hou, Chen-Ping Yu, Minh Hoai, and Dimitris Samaras. Large-scale training of shadow detectors with noisily-annotated shadow examples. In *ECCV*, 2016.

[30] Tomas F Yago Vicente, Minh Hoai, and Dimitris Samaras. Leave-one-out kernel optimization for shadow detection and removal. *IEEE TPAMI*, 2017.

[31] Yael Vinker, Inbar Huberman-Spiegelglas, and Raanan Fattal. Unpaired learning for high dynamic range image tone mapping. In *ICCV*, 2021.

[32] Jifeng Wang, Xiang Li, and Jian Yang. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *CVPR*, 2018.

[33] Yupei Wang, Xin Zhao, Yin Li, Xuecai Hu, and Kaiqi Huang. Densely cascaded shadow detection network via deeply supervised parallel fusion. In *IJCAI*, 2018.

[34] Jun Wei, Shuhui Wang, and Qingming Huang. F^3net: fusion, feedback and focus for salient object detection. In *AAAI*, 2020.

[35] Quanlong Zheng, Xiaotian Qiao, Ying Cao, and Rynson W.H. Lau. Distraction-aware shadow detection. In *CVPR*, 2019.

[36] Huajun Zhou, Xiaohua Xie, Jian-Huang Lai, Zixuan Chen, and Lingxiao Yang. Interactive two-stream decoder for accurate and fast saliency detection. In *CVPR*, 2020.

[37] Jiejie Zhu, Kegan GG Samuel, Syed Z Masood, and Marshall F Tappen. Learning to recognize shadows in monochromatic natural images. In *CVPR*, 2010.

[38] 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 *ECCV*, 2018.

[39] Lei Zhu, Ke Xu, Zhanghan Ke, and Rynson WH Lau. Mitigating intensity bias in shadow detection via feature decomposition and reweighting. In *ICCV*, 2021.