This ACCV 2024 paper, provided here by the Computer Vision Foundation, is the author-created version.
The content of this paper is identical to the content of the officially published ACCV 2024
LNCS version of the paper as available on SpringerLink: https://link.springer.com/conference/accv

ADSP: Advanced Dataset for Shadow Processing, enabling visible occluders via synthesizing strategy.

Chang-Yu Hsieh[1 [0009−0005−5616−456X]] and Jian-Jiun Ding[1 [0000−0003−4510−2273]]

Graduate Institute of Electronics Engineering, National Taiwan University, Taiwan
darkrepulser.ray@gmail.com jjding@ntu.edu.tw

Abstract. Shadows can lead to malfunctions in computer vision, making shadow removal an essential task for restoring underlying information. For a long time, researchers have proposed hand-crafted methods based on observing shadow formation models. Then, deep-learning-based solutions have further advanced performance in restoration quality. However, existing datasets have several limitations, such as lacking occluders, restricted camera angles, and inconsistency. In this paper, a novel benchmark called the Advanced Dataset for Shadow Processing (ADSP) is introduced. Through the synthesizing strategy, the ADSP becomes the first dataset containing outdoor images with occluders. Statistical analysis and experiments demonstrate that the ADSP has the advantages of lower domain shifting, matching real-world scenarios, and sufficient generalizing capability. Moreover, as a reference for the removal task, we also propose the Segmented Refinement Removal Network (SRRN), which includes three subnets for shadow removal, color adjustment, and boundary smoothing, respectively. It achieves state-of-the-art performance and can be set as a reference for shadow removal.

Keywords: Shadow Removal · Dataset Creation · Restoration

1 Introduction

Shadows are cast when objects occlude light and then generate the corresponding regions with relatively low intensity. In the shadow regions, most image features (edge, textures, color, *etc.*) are weakened, resulting in challenges on many computer vision tasks, such as recognition [51], detection [5,39], segmentation [34,38], tracking [4,24,7], intrinsic image decomposition [28,11], dehazing [37], and *etc.* [53]. Intrinsic diversity of shadows [1] aggravates the difficulty of shadow removal. The shadow comes from multiple interactions between the environment and occluders and has complicated properties. For example, its shape depends on the profiles of occluders, and its intensity depends on the light source and the aperture for image acquisition. The sunlight in the evening may appear yellowish due to atmospheric refraction. A coarse surface will lead to non-smooth shadow edges.

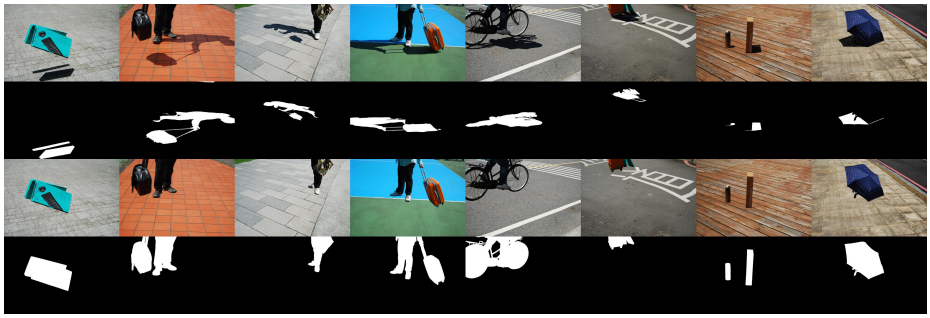
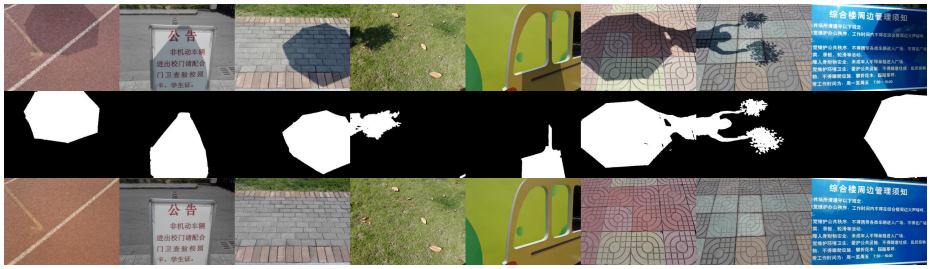


Fig. 1: Samples from the proposed ADSP benchmark. From top to bottom are shadow image, shadow mask, the corresponding shadow-free image and the occluder mask.

Shadow removal aims to recover the lost information from a shadow image. Moreover, shadow detection, which predicts the shadow mask, is often an auxiliary task of shadow removal. Early shadow removal/detection works adopted physical models [50,43,17,16,1,25,9,30,8], which rely on the prior information of shadow. Recently, with the large-scale datasets [35,45,42,41,36,22], the shadow detection/removal problem can be formulated as a regression model and adopted learning-based networks as solutions [55,26,20,27,33,12,49,44,14,3]. Most of them applied supervised learning. Although some works adopted unsupervised learning [21,32,40] and used unpaired datasets for training, supervised learning with paired data is still the mainstream for shadow removal. In addition, due to the importance of the shadow position information, some datasets contain shadow masks as well, which makes them able to be used for shadow detection and results in the development of multi-task training. Collecting a shadow removal dataset takes significant effort and time. Therefore, most existing datasets have similar limitations on contents and diversity, like (1) little visible occluder in shadow images, (2) unignorable inconsistency problems among non-shadow regions, and (3) the restricted range of the camera depression angle. In Fig. 2, examples from two existing benchmarks, the SRD [35] and the ISTD [45], are presented.

In this paper, a novel benchmark, named the Advanced Dataset for Shadow Processing (ADSP[1]), is proposed. Through the new synthesizing strategy, the visible occluder can be extracted singularly by carefully labeled masks and generates the shadow-free ground truth to form paired data. In addition, well-labeled shadow masks allow the proposed ADSP to be utilized in most shadow tasks, *e.g.* shadow detection and generation. Figure 1 provides examples of image pairs in the ADSP. The proposed ADSP provides shadow image pairs with highest quality and resolution currently, which can be adopted to tackle the situation in which both the occluder and the corresponding shadow are visible in the outdoor scene. To our knowledge, it is the first benchmark designed explicitly for such situation. We also propose a novel shadow removal network, the Segmented

[1] https://github.com/ElucidatorRay/ADSP



(a) ISTD[45].From top to bottom are shadow images, shadow masks, and the corresponding shadow-free images, respectively.

(b) SRD[35]. First row are shadow images and the second are shadow-free images.

Fig. 2: Samples from two existing shadow removal benchmarks.

Refinement Removal Network (SRRN). It achieves state-of-the-art performance on the ADSP and can be set as a reference for shadow removal.

In summary, the contributions of this work are as follows:

- We proposed the ADSP, a new benchmark that contains 1220 image pairs of shadow-affected images, shadow-free images, shadow masks, and occluder masks. It provides excellent supervision for shadow removal.
- Statistical analysis shows that the proposed ADSP outperforms existing datasets by more challenging data and fewer inconsistency problems. On the other hand, domain shift experiments prove that the ADSP provides excellent generalization capability with fewer data pairs.
- We proposed the SRRN for single-image shadow removal. It has three subnets representing distinct stages, which perform preliminary removal, color adjustment, and boundary smoothing. Comprehensive experiments show that the SRRN achieved state-of-the-art performance on shadow removal.

2 Related Works

Shadow removal is a challenging task in computer vision. In this section, we review the development history of shadow removal methods and describe existing (1) large-scale benchmarks, (2) prior-based methods, and (3) deep-learning-based methods.

2.1 Large-scale Benchmarks

The success of deep-learning-based methods relies on large-scale datasets to provide supervision. Therefore, researchers had proposed various benchmarks [35,45,41] for training supervised shadow removal.

Qu et al. [35] proposed the SRD, a large-scale shadow removal dataset. The authors captured corresponding shadow-free images by passively waiting for sunlight variation or actively removing objects. They placed the camera with a tripod, shot with a wireless remote controller, and fixed exposure parameters to mitigate the inconsistency. The SRD exhibited great diversity and quantity. However, the visible occluder was infeasible during active acquiring, or there would be remaining shadows, as in the left three columns of Fig. 2b. Moreover, there is some inconsistency, as shown in the fifth column in Fig. 2b.

Wang et al. [45] proposed the ISTD, which is the first triplet dataset consisting of shadow/shadow-free images and shadow masks. It enabled the multi-task training strategy, which had been proven helpful in many works [56,12]. However, it also lacked the visible occluder. In addition, most of the scenes have restricted view, either on the ground with a high camera depression angle or on a vertical plane, as shown in Fig. 2a.

Recently, Vasluianu et al. [41] proposed the WSRD, which was built fully around a set of controllable conditions, including a directional light for casting shadow and a diffusive light to generate uniform light distribution. Image pairs were collected by turning the directional light on and off during shadow/shadow-free acquisition. The WSRD is a diverse benchmark, however, such a setup is not available for outdoor collection. Furthermore, it still suffers from inconsistency problems and relatively unrealistic semantic information.

Besides the above benchmarks, some datasets focused on other aspects of shadow but did not conform to the requirement of supervised shadow removal. Datasets such as SBU [42], UIUC [16], UCF [54], and CUHK-Shadow [22] were designed for shadow detection, lacking the shadow-free ground truth. Wang et al. [46] introduced the novel Instance Shadow Detection and the SOBA benchmark, including masks and bounding boxes of shadow-object pairs but no shadow-free images. Shadow-AR [29] and DESOBA series [19,31] focused on generating reasonable shadows for objects in the scene and contained partial-shadow-free information. Sen et al. [36] constructed a private SFHQ with high-resolution data, and Hu et al. [21] collected the USR with unpaired images.

2.2 Prior Based Methods

Early shadow-removal works were mainly based on the priors, which utilized the underlying information behind shadow and physical models. They used the properties of shadow, like gradients [8,30,9], illumination [10,48,50], color [43], regions [16,17,43], texture [25,43], and intensity [1] to build shadow detection/removal formulation. Prior-based methods usually have weaker robustness and lower generalizability due to the strong hypothesis on shadow.

2.3 Deep Learning Based Methods

Since deep learning achieved outstanding performance on many CV tasks, most of the recent works on shadow removal applied multiple networks to recover shadow-free images from corrupted ones.

Qu et al. [35] proposed an end-to-end automatic DeshadowNet that applies high-level semantic information. Wang et al. [45] introduced a multi-task perspective, which jointly learned both detection and removal models simultaneously. Le et al. [26] viewed a shadow-free image as the linear combination of shadow and relit images and adopted image decomposition for shadow removal. Hu et al. [20] analyzed the image context in a direction-aware manner and developed a direction-aware module for shadow detection and removal. Fu et al. [12] formulated shadow removal as an exposure fusion problem and restored the shadow-free image by fusing the image with several over-exposure ones. Yu et al. [49] proposed the CNSNet to restore the shadow-free image. Wan et al. [44] considered the style consistency of de-shadowed and non-shadow regions and applied a two-stage removal process. Guo et al. [14] proposed the ShadowFormer, a transformer-based network, to exploit the global correlation between shadow and non-shadow regions. They also proposed ShadowDiffusion [15], which was a diffusion-based network, and integrated image and degradation priors.

In addition, some works aimed to deal with the problems of limited datasets. Hu et al. [21] proposed the Mask-ShadowGAN, which was based on a mask-guided cycle-consistency constraint. However, it needs that the shadow and shadow-free images should share similar statistical properties, which is hard to satisfy. In order to deal with it, Le et al. [27] proposed to crop two kinds of patches from a single shadow image and introduced a patch-based shadow removal system. Liu et al. [33] further proposed the G2R-ShadowNet containing three sub-networks focusing on generation, removal, and refinement, respectively. It can avoid the requirements of strict physics-based constraints, high computation, and carefully cropping, *etc*.

3 Methodology

3.1 Advanced Dataset for Shadow Processing (ADSP)

We summarize the following requirements that an advanced benchmark for shadow removal must fulfill. First, it should contain the visible occluder since shadow images with whole or parts of the occluder are universal scenarios in practical applications. Thus, the visible occluder helps to improve the robustness of models to in-the-wild scenarios. Second, it should be composed of at least triplet data because the multi-task training strategy and mask-enhanced removal have become increasingly popular. Furthermore, additional attributes about shadow, like position, expand the potential of the proposed dataset. Third, the image inconsistency problem should be considered at the acquisition stage because data with lower deviation helps to train unbiased models and accurately evaluate performance.

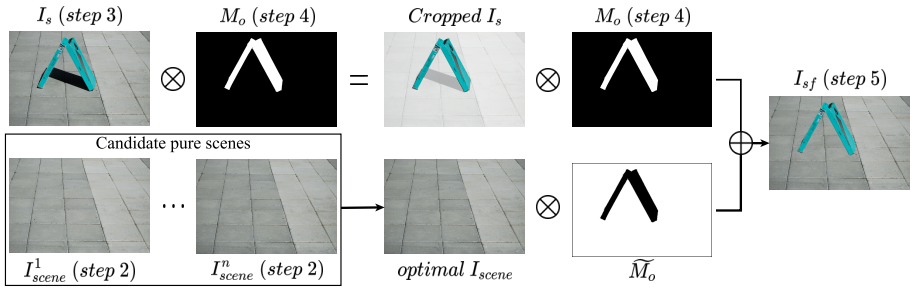


Fig. 3: The adopted construction process of the proposed ADSP, where $\otimes$ is the Hadamard product and $\oplus$ is the element-wise addition.

To achieve these requirements, unlike existing datasets that directly shoot two kinds of pictures to form a pair, we propose the synthesizing strategy and formulate a series of rigorous steps for data acquisition. The process for constructing a single data pair is as follows:

1. We first determine a shooting scene without any shadow in advance. To maximize diversities, we choose it based on the surface texture and the direction of the light source.
2. Then, we collect the pure scene I_{scene} (w/o shadow or occluder). I_{scene} is highly maneuverable for post-processing.
3. Next, we add occluders into the environment to cast shadows and collect the shadow-affected image I_s (w/ shadow and occluder).
4. After image shooting, we label the I_s with LabelMe and get the shadow mask M_s and the occluder mask M_o.
5. With the position indicated by M_o, we crop the occluder from I_s, paste it onto I_{scene}, and synthesize the shadow-free image I_{sf}.
6. Last, we integrate four images (I_s, I_{sf}, M_s, M_o) to form the quadruplet data.

The entire construction process has three stages: image acquisition (steps 1 to 3), labeling (step 4), and synthesizing (steps 5 to 6). In stage 1, we used the camera Nikon D60 with lenses of Nikon DX SWM ED Aspherical 0.28/0.92ft 52. We also adopted a tripod and a wireless controller to avoid inconsistency caused by non-expected movement. Furthermore, since a clear presentation of background information is more critical than the shadow itself, between steps 1 and 2, we conduced multiple trial shots to determine the optimal shutter speed to produce a visually pleasant result. Such operation also led to a broader intensity range of the shadow region. We will analyze this later in Section 4.1. Subsequently, we switched the camera temporarily to the automatic mode for auto-focusing and then back to the manual mode to guarantee an identical setting. To ensure consistency, we consciously gathered more I_{scene} randomly for the later synthesizing stage.

In stage 2, we labeled I_s and achieved pixel-level accuracy on both masks. In stage 3, we performed synthesizing to generate I_{sf}. Figure 3 presents the process

of the proposed synthesizing strategy. Notably, for each scene, there was a set of candidate I^i_{scene}, where i is the index of each photo. We selected the optimal I^*_{sf} based on the non-shadow region error between I^i_{sf} and I^i_{scene}.

Currently, using the synthesizing strategy is inevitable to collect the outdoor shadow scenes with occluders. Without a controllable environment, actively gathering I_{sf} by disabling directional light like WSRD [41] is not feasible for outdoor scenes. In addition, passively waiting for sunlight change and acquiring I_{sf} will result in severe inconsistency problems. More specifically, compared with I_s, the non-shadow region of I_{sf} would be darker, as shown in Fig. 2. In general, the advantages of using the synthesizing method are apparent. First, the complex interaction between the occluder and the environment is simplified, and the shadow cast by any movable object is collectable. Second, I_s and I_{sf} can be acquired under the same illumination condition, mitigating the inconsistency problem. Furthermore, without the limitation of no occluder, the camera depression angle has a higher degree of freedom. Compared with existing outdoor datasets, the scenes in the ADSP dataset are not restricted to the top view or with shadows on the vertical plane.

Overall, the ADSP dataset was constructed using the synthesizing strategy. Shadow-free occluder photos are available, providing semantic information that is more practical and realistic. Two kinds of masks with pixel-level accuracy make it a quadruplet dataset. At the same time, the rigorous collecting process ensures excellent consistency. It contains 1220 high-resolution (2592 x 3872 px) image pairs. Figure 1 shows some samples of the proposed ADSP dataset.

3.2 Segmented Refinement Removal Networks (SRRN)

Inspired by Chang et al. [2], we built the removal model with multi-stage processing and expanded the regular two-stage network (*i.e.* removal and refinement) into a three-stage one. We first observed that two factors decrease the reliability of removal results. The first one is the boundary effect around the edge of shadow and the second one is the bias of color in the shadow region. To address these problems, we divided the refinement stage into two more specialized steps and proposed the Segmented Refinement Removal Network (SRRN). The SRRN contains three subnets aiming at (i) preliminary shadow removal, (ii) shadow area color adjustment, and (iii) shadow region boundary smoothing, respectively. Distinct loss functions were applied to supervise three subnets.

Figure 4 presents a schematic overview of the proposed SRRN. In the initial stage, we used the ShadowFormer [14] as the backbone of the removal subnet θ_{sr}. It comprises a linear projection layer and several channel attention modules in the encoder and decoder sections to capture multi-scale hierarchical global features. Within the bottleneck layers, the Shadow-Interaction Module [14] was employed to recover shadowed areas with the assistance of contextual features from non-shadow regions. θ_{sr} is constrained by the Charbonnier Loss.

On the other hand, for the second and third stages, the color adjustment θ_{ca} and the boundary smoothing θ_{bs} subnets, we adopted the two-stage SG-ShadowNet[44] as the backbone. The θ_{ca} is structured as a U-net [18]. It is

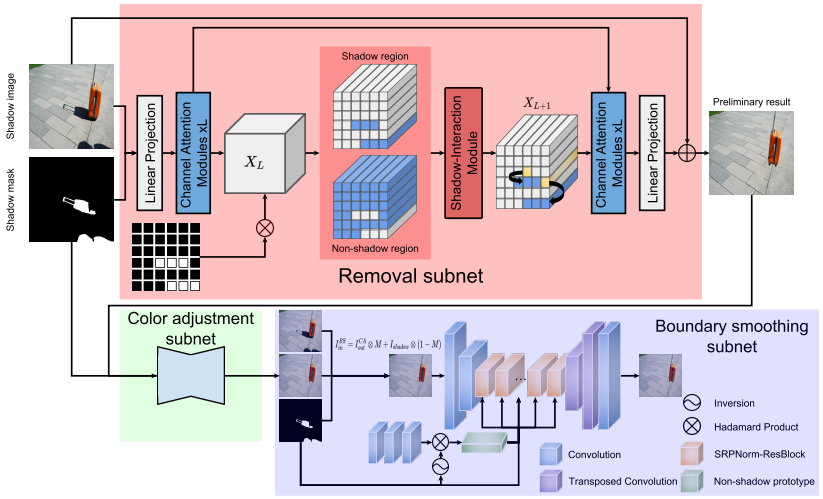
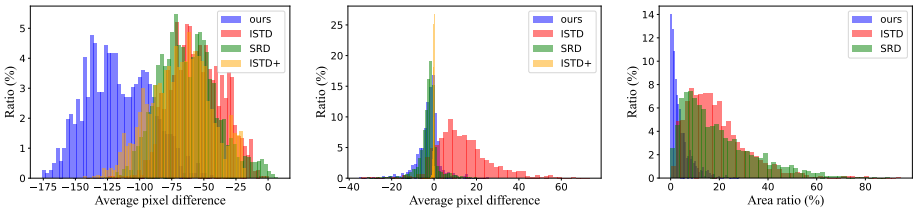


Fig. 4: Overview of the Segmented Refinement Removal Network (SRRN). The removal subnet (the top red region) applies the ShadowFormer[14] as the backbone for preliminary removing. The color adjustment subnet (the bottom-left green region) and the boundary smoothing subnet (the bottom-right blue region) uses the SG-ShadowNet[44] as the backbone for refinements.

equipped with the regular reconstruction loss and the shadow area loss to correct the color bias in the shadow region. θ_{bs} comprises the region style estimator and the boundary refinement network. The former employs several 1×1 convolution layers to extract non-shadow prototypes. The second one, which follows the U-Net [32] architecture, adopts 9 SRPNorm-ResBlocks[44] in bottleneck to perform smoothing. In this stage, besides the reconstruction loss and shadow area loss, we also applied the spatial consistency loss [13] and the penumbra loss to enhance the performance of the penumbra region. In Section 4.3, we will perform the ablation study to analyze the effect of the added penumbra loss and discuss its optimal parameter. More details on training will be presented in the Reproduction section of the supplementary material.

4 Experiments

We conduct the following experiments to evaluate the proposed benchmark and the baseline model. First, we performed statistical analysis on the proposed ADSP and three existing datasets, revealing the properties of the proposed ADSP and verifying that it prevails in both difficulty and consistency. Second, we applied domain shift experiments on two state-of-the-art (SOTA) algorithms, the ShadowFormer [14] and the SG-ShadowNet [44], in which the models were trained and evaluated on different datasets. The results show that the ADSP is more challenging and capable of generalization. Third, we perform the ablation



(a) Shadow density **(b)** Non-shadow consistency **(c)** Shadow area ratio

Fig. 5: Histograms of three statistical analyses. Figure 5a represents the average intensity attenuation on the shadow region, Figure 5b shows the average bias on the non-shadow region, and Figure 5c shows the distribution of the area ratio of the shadow region. For the SRD, we adopted shadow masks predicted by Cun et al. [6]. For fairness, the occluder region was excluded when computing Figs. 5a and 5b.

Table 1: Distribution of three kinds of occluder size of images. Where R_o means the area ratio of the occluder mask.

Size	Small	Medium	Big
R_o	∼10%	10%∼30%	30%∼
ratio	65.98%	31.39%	2.62%

Table 2: Distribution of three kinds of distance between shadow and camera.

Type of view	close	middle	long
range (m)	0∼2	2∼6	6∼
freq.	17.21%	77.05%	5.66%

Table 3: The appearance count of eight types of occluder. Note that there may be more than two types in single image.

Type of occluder	suitcase	keyboard case	human	camera bag	umbrella	carton	bike	Others
frequency	214	288	608	103	536	76	35	13

study on the proposed SRRN, verifying the reasonability of our design. Last, we compare the SRRN with SOTA models and show that our SRRN achieves superior performance.

4.1 Statistical Analysis

We give statistical analysis based on (1)the pixel difference, (2)the shadow mask ratio, (3) the occluder type and size, and (4) the type of view.

First, Fig. 5a represents the average intensity attenuation on the shadow region, *i.e.*, the density of collected shadow. It reveals that the ADSP is a challenging dataset with more significant decay, resulting in weaker underlying information. Note that such stronger degradations stem from both the collecting environment and the occluder. As mentioned in Section 3.1, we acquired a clear background via customized shutter speed. Hence, darker shadows are not made on purpose but come from stronger sunlight. Furthermore, the visible occluder contributes to the challenge, which will be shown in Section 4.2. On the other hand, Fig. 5b shows the bias on the non-shadow region, reflecting the inconsistency. According to the result, the proposed process has excellent potential

to reduce biases and suppress the inconsistency. Compared with the ISTD, the ADSP ensures the correctness of most pairs, *i.e.*, the distribution has a peak value around zero. Compared to the SRD, it achieves a higher accuracy with an acceptable precision trade-off. Specifically, our dataset-level difference mean is about -2.49, better than -2.52 of the SRD. However, the SRD has a lower standard deviation (about 4.66) than ours (about 5.36). We attribute these to the fact that we did not customize I_{scene} for each I_s but used a randomly augmented candidates set. Even so, it still verifies that the synthesizing strategy can generate I_{sf} with stable quality and no exceptional adjustments is needed.

Second, Fig. 5c shows the ratio of the shadow area. We ascribed the lower shadow area ratio to the trade-off of getting a wider camera depression angle. As mentioned in Section 2.1, the restricted camera view derives from the requirement of no occluder. Therefore, top-view photos with low depression angles undoubtedly consist of very large shadow areas. Third, as mentioned in Section 1, the ADSP was designed explicitly for visible occluders, thus we conducted analyses on occluders especially. For the occluder size, we used the ratio of the mask to the whole image as indicators and dividing them into three major types (Small, Medium and Big), as Table 1 showed. For the occluder type, we made the collected objects as diverse as possible and split them into eight types, as shown in Table 3. Fourth, we also analyzed the distance between the shadow and the camera, hoping that it would be helpful to train the shadow removal algorithms with specific interest ranges. We subjectively estimate the distances of photos and classify them into three categories, as shown in Table 2.

4.2 Domain Shift Validation

In Section 4.1, we verified the advantages of the ADSP from the statistical perspective. Here, we prove it again via domain shift experiments. Cross-evaluations on distinct sets with different distributions reveal each benchmark's difficulty and generalization capability. We applied two SOTA models on three existing datasets and our ADSP. We used official splitting on the formers and randomly divided the latter into training (1100 pairs) and testing (120) sets. The official implementation of each model was re-trained after minimal modifications. Down-sampled ADSP images were adopted to provide a closer image size range with existing data. We reported PSNRs of the whole image, the shadow region, and the non-shadow region, respectively. We also improve the calculation method of region metrics to prevent unreasonable high values, as shown in the supplementary material.

Table 4 shows the results of domain shift experiments. According to it, we summarized four comparisons to prove the proposed ADSP. We use $A \rightarrow B$ to represent the experiment, having A as the source domain and B as the target domain. We use $\{\}$ to represent the candidates for each domain, *e.g.* $\{A, B\} \rightarrow \{C, D\}$ means a collection of four experiments: $A \rightarrow C$, $A \rightarrow D$, $B \rightarrow C$, and $B \rightarrow D$. We first separate these four datasets into two groups by their content. The proposed ADSP and the DESOBAv2 contain visible occluder, while the SRD and the ISTD seldom do. We found that experiments of

Table 4: The quantitative result of domain shift experiments. We evaluated each model on the validation set of each benchmark and reported PSNR on the entire image/shadow region/non-shadow region and SSIM of the entire image. Different colors of underline and cell indicate included values in distinct comparisons, which we did in Section 4.2. Red and Blue represent the best or second best value among each row, except for experiments in the diagonal, which were not shifted.

Metrics	PSNR:(whole/shadow region/non-shadow region), SSIM:(whole)			
Eval\Train	SRD [35]	ISTD [45]	DESOBAv2 [31]	ADSP (ours)
ShadowFormer [14]				
SRD	30.34/27.07/32.45 0.8885	18.37/15.70/19.92 0.8272	22.51/17.12/27.10 0.7839	22.39/16.96/27.54 0.8228
ISTD	21.79/17.11/23.83 0.8976	30.48/27.84/31.55 0.9273	24.71/21.92/25.81 0.8662	21.18/16.47/23.29 0.9020
DESOBAv2	27.72/13.58/30.29 0.9371	21.00/13.09/21.55 0.9143	35.27/24.76/36.67 0.9606	29.45/18.83/30.72 0.9414
ADSP	25.51/12.57/30.77 0.8985	20.77/12.78/21.60 0.8679	28.70/18.28/30.99 0.8967	32.41/24.90/33.25 0.9190
SG-ShadowNet [44]				
SRD	25.69/20.51/29.88 0.8536	20.94/17.14/23.36 0.8224	20.92/15.17/27.04 0.7685	22.73/17.54/26.88 0.8222
ISTD	24.70/22.85/25.20 0.9089	28.95/26.23/30.10 0.9213	24.73/21.84/25.89 0.8647	23.34/19.91/27.74 0.8961
DESOBAv2	26.39/11.90/29.42 0.9328	24.49/13.66/25.92 0.9257	32.20/19.29/34.23 0.9483	27.52/14.57/29.41 0.9346
ADSP	24.98/12.67/29.23 0.8837	22.41/13.27/24.42 0.8699	28.21/17.58/30.88 0.8955	31.61/21.09/33.73 0.9195

$\{SRD, ISTD\} \rightarrow \{ADSP, DESOBAv2\}$ have much lower shadow region PSNRs. The interval of all eight results is $[11.90, 13.66]$. By contrast, those of $\{DESOBAv2, ADSP\} \rightarrow \{SRD, ISTD\}$ is $[15.17, 21.92]$. It reveals that images with visible occluder can provide better robustness to deal with diverse shadow images. Second, among eight $ADSP \rightarrow DESOBAv2$ metrics, there are seven first-place results. The leftover one also earned second place with an insignificant degradation of 0.01. It proves that the visible occluder is also one of the critical reasons causing high difficulty because the metric decay of domain shift experiments between datasets with similar contents is slighter.

Based on the same idea, we further focus on the results between two popular benchmarks (the SRD and the ISTD) and our ADSP. The shadow region PSNRs interval of $\{SRD, ISTD\} \rightarrow ADSP$ is $[12.57, 13.27]$, and that of $ADSP \rightarrow \{SRD, ISTD\}$ is $[16.47, 19.91]$. Milder domain shift effect and completely non-overlapped intervals prove the ADSP has greater difficulty.

Last, from the performance ranking of all metrics, ten of the sixteen $ADSP \rightarrow \{SRD, ISTD\}$ results achieve the first or second place, confirming that the ADSP provides excellent generalization capabilities. Although some results do not behave impressively, we should realize that the ADSP achieves these under

Table 5: The quantitative result of the ablation study using our models with different hyper-parameters. Reported metrics were computed in the validation set of the proposed ADSP. Red and Blue indicate the best and second-best results in such metrics.

Model	θ_{sr}	θ_{ca}/θ_{bs}	SRRN							
$\mathcal{L}_p^{bs}$	✗		✓							
λ	0		1	1.5	2	2.5	3	5	10	20
Train on.	Combine		Combine (θ_{sr})/ADSP(θ_{ca}/θ_{bs})							
$psnr_{all}$	32.44	29.42	33.20	33.24	33.18	33.19	33.23	33.16	33.18	33.10	33.04
$psnr_{s}$	25.08	18.56	25.39	25.71	25.70	25.70	25.61	25.73	25.79	25.80	25.82
$psnr_{ns}$	33.23	31.83	34.17	34.14	34.06	34.09	34.15	34.04	34.05	33.92	33.88
$ssim_{all}$	0.9202	0.9011	0.9280	0.9275	0.9270	0.9271	0.9269	0.9264	0.9263	0.9252	0.9236

the disadvantage of data quantity. Our ADSP has 1100 training and 120 testing pairs, fewer than the SRD (2680/408) and the ISTD (1330/540). The smaller scale stems from a more rigorous and complex construction process and the data diversity, *i.e.*, the novel quadruplet data the ADSP provides. Moreover, for the DESOBAv2 containing a considerable quantity (20000/296), due to the partial-shadow-free information, it gains an advantage only on the relatively simple ISTD, which verifies the potential of the proposed synthesizing strategy.

4.3 Ablation Study

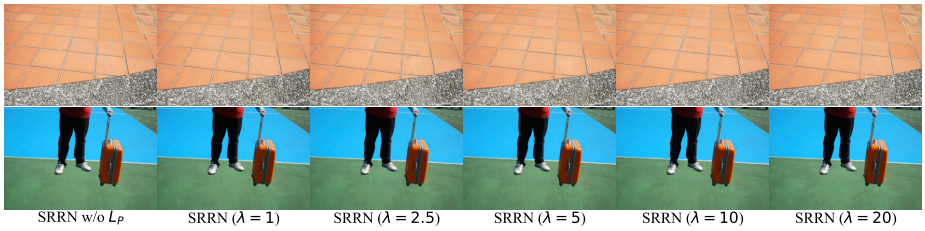
As mentioned in Section 3.2, we applied the penumbra loss in the boundary smoothing subnet θ_{bs} to focus on the penumbra region. The total adopted loss functions used in stage two training (θ_{ca}, θ_{bs}) are as follows:

$$\mathcal{L}_1 = ||I_{gt} \otimes M - I_{out} \otimes M||_1 \tag{1}$$

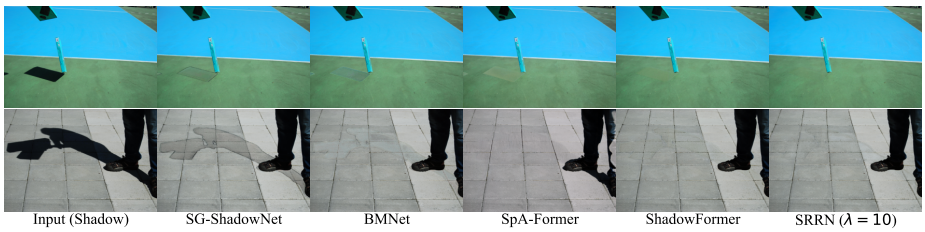
$$\mathcal{L}_{overall} = \mathcal{L}_R^{ca} + \mathcal{L}_A^{ca} + \mathcal{L}_R^{bs} + \mathcal{L}_A^{bs} + 10 \times \mathcal{L}_{spa}^{bs} + \lambda \times \mathcal{L}_p^{bs} \tag{2}$$

where the subscripts mean the type of loss functions, the superscripts represent the target model, $\otimes$ is the Hadamard product, and $\mathcal{L}_R$, $\mathcal{L}_A$, and $\mathcal{L}_p^{bs}$ are all variants of the $\mathcal{L}_1$ loss with different masks M, where R means the reconstruction loss with M_1 of one, A means the shadow area loss with the shadow mask M_s, and *spa* means the spatial consistency loss [13]. In this section, we tried to search for an optimal λ, which controls the ratio of the penumbra loss $\mathcal{L}_p$ with a penumbra mask M_p.

Table 5 compares two individual stages and nine ablation studies with different λ. We can see that all results of the complete SRRN outperformed the two individual stages. Moreover, $\mathcal{L}_p$ shows great help on the performance of the shadow region R_s. Figure 6a show the qualitative results of six different λ. Among all results, the shadow contours gradually become lighter as the λ increases. Especially for the $\lambda \geq 10$, they have been imperceptible. We chose the SRRN with $\lambda = 10$ as the final version because it has the best overall performance.



(a) Qualitative results of ablation studies of hyper-parameter λ.



(b) Examples of shadow removal results on ADSP from five best models.

Fig. 6: Qualitative results of ablation study and comparions with state-of-the-art models.

4.4 Comparison with State-of-the-Art Methods

In this subsection, we compare the proposed SRRN with eight SOTA models. We followed most settings mentioned in Section 4.2, except for the training set. We re-trained each SOTA model on the combined dataset (SRD[35], ISTD[45], and the proposed ADSP) and evaluated it on the testing set of the down-sampled ADSP. We reported two results with λ equal to 1 and 10 to emphasize that added penumbra loss is a trade-off between two regions.

Table 6 presents quantitative results of comparison. Two SRRNs surpass all competing methods among every included metric. Compared with the input image, our methods recover underlying information on the shadow region and keep the non-shadow region from generating artifacts. All of the seven metric values of our methods have positive changes to the original inputs, which some SOTAs did not. On the other hand, Fig. 6b demonstrates the visual examples of the best five models. We can also find that despite existing competing methods being able to reconstruct the original shadow-free image, they usually leave some visually unnatural areas, *e.g.*, residual shadows or visible shadow boundaries. Our method, in contrast, fixes those areas as much as possible and makes them imperceptible.

5 Limitation

There are still some limitations of the ADSP since it mainly handles the problems of 1) occluders, 2) limited camera angles, and 3) inconsistency in outdoor scenarios. There are some trade-offs for these goals. First, the outdoor light

Table 6: The quantitative results of shadow removal using our models and recent methods on the proposed ADSP. Red and Blue indicate the best and second-best results in such metrics.

Method	RMSE ↓			PSNR ↑			SSIM ↑
	Whole	shadow	non-shadow	Whole	shadow	non-shadow	Whole
Input Image	23.552	115.778	8.754	21.68	7.09	31.08	0.8842
Mask-ShadowGan [21]	15.564	61.955	9.268	25.11	13.21	30.10	0.8825
DC-ShadowNet [23]	18.284	61.219	11.444	23.75	13.46	27.95	0.8594
Fu et al. [12]	12.828	42.907	9.635	27.05	16.37	30.03	0.8912
SG-ShadowNet [44]	9.498	31.851	7.688	29.42	18.56	31.83	0.9011
BMNet [56]	8.588	23.116	7.537	30.62	21.34	32.34	0.9086
SpA-Former [52]	13.154	32.640	9.670	26.92	18.72	30.16	0.8988
ShadowFormer [14]	6.988	15.335	6.649	32.44	25.08	33.23	0.9202
SADC [47]	10.263	35.483	8.158	28.80	17.85	31.66	0.9053
Ours ($\lambda = 1$)	6.220	14.157	5.821	33.24	25.71	34.14	0.9275
Ours ($\lambda = 10$)	6.316	14.066	5.951	33.10	25.80	33.92	0.9252

source is usually the sunlight. Thus, soft or overleaped shadows are unavailable. Second, the proposed method has higher requirement on environment. Collected scene images should not contain any intrinsic shadows, which limits the acquiring on some situation, *e.g.* non-planar plane or a vertical wall. Third, as mentioned in Section 4.1, adopting a broader camera angle unavoidably causes shadows to have a small image ratio. Last, the proposed post-processing construction can not deal with self-shadows, *i.e.*, the shadow cast on the occluder itself.

6 Conclusion

In this work, a novel benchmark for shadow removal was proposed. With the synthesizing strategy, our proposed ADSP mitigated the limitation of existing datasets. The images in the ADSP dataset contain occluders in outdoor images, have wider camera depression angles, and well avoid inconsistency. We conducted statistical analysis and domain shift experiments to evaluate the proposed benchmark. Moreover, as a reference for the removal task on the ADSP, we proposed the SRRN, a novel three-stage network. An ablation study presented that the newly added penumbra loss could effectively improve the performance of shadow region recovery. Compared to existing methods, the proposed SRRN achieved state-of-the-art performance on shadow removal.

Acknowledgments. We thank to the anonymous reviewers for their constructive feedback. Additionally, we would like to express our gratitude to Shih-Fen Shia for assisting data labeling, Yi-Hsien Chen, Chun-Han Lin for supporting data collection and Yu-Hsiang Kung for providing partial computational resources.

Disclosure of Interests. The authors declare no competing interests.

References

1. Arbel, E., Hel-Or, H.: Shadow removal using intensity surfaces and texture anchor points. IEEE Transactions on Pattern Analysis and Machine Intelligence **33**(6), 1202–1216 (2011). https://doi.org/10.1109/TPAMI.2010.157
2. Chang, H.E., Hsieh, C.H., Yang, H.H., Chen, I.H., Chen, Y.C., Chiang, Y.C., Huang, Z.K., Chen, W.T., Kuo, S.Y.: Tsrformer: Transformer based two-stage refinement for single image shadow removal. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 1436–1446 (June 2023)
3. Chen, Z., Long, C., Zhang, L., Xiao, C.: Canet: A context-aware network for shadow removal. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 4723–4732 (2021). https://doi.org/10.1109/ICCV48922.2021.00470
4. Cucchiara, R., Grana, C., Piccardi, M., Prati, A.: Detecting moving objects, ghosts, and shadows in video streams. IEEE Transactions on Pattern Analysis and Machine Intelligence **25**(10), 1337–1342 (2003). https://doi.org/10.1109/TPAMI.2003.1233909
5. Cucchiara, R., Grana, C., Piccardi, M., Prati, A., Sirotti, S.: Improving shadow suppression in moving object detection with hsv color information. In: ITSC 2001. 2001 IEEE Intelligent Transportation Systems. Proceedings (Cat. No.01TH8585). pp. 334–339 (2001). https://doi.org/10.1109/ITSC.2001.948679
6. Cun, X., Pun, C.M., Shi, C.: Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting gan. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 10680–10687 (2020)
7. Danelljan, M., Shahbaz Khan, F., Felsberg, M., van de Weijer, J.: Adaptive color attributes for real-time visual tracking. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2014)
8. Finlayson, G., Hordley, S., Lu, C., Drew, M.: On the removal of shadows from images. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(1), 59–68 (2006). https://doi.org/10.1109/TPAMI.2006.18
9. Finlayson, G.D., Drew, M.S., Lu, C.: Entropy minimization for shadow removal. International Journal of Computer Vision **85**(1), 35–57 (2009)
10. Finlayson, G.D., Hordley, S.D., Drew, M.S.: Removing shadows from images. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) Computer Vision — ECCV 2002. pp. 823–836. Springer Berlin Heidelberg, Berlin, Heidelberg (2002)
11. Fu, G., Zhang, Q., Xiao, C.: Towards high-quality intrinsic images in the wild. In: 2019 IEEE International Conference on Multimedia and Expo (ICME). pp. 175–180 (2019). https://doi.org/10.1109/ICME.2019.00038
12. 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: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10571–10580 (June 2021)
13. Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S., Cong, R.: Zero-reference deep curve estimation for low-light image enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2020)
14. Guo, L., Huang, S., Liu, D., Cheng, H., Wen, B.: Shadowformer: Global context helps shadow removal. In: Williams, B., Chen, Y., Neville, J. (eds.) Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference

on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023. pp. 710–718. AAAI Press (2023). https://doi.org/10.1609/AAAI.V37I1.25148, https://doi.org/10.1609/aaai.v37i1.25148
15. Guo, L., Wang, C., Yang, W., Huang, S., Wang, Y., Pfister, H., Wen, B.: Shadowdiffusion: When degradation prior meets diffusion model for shadow removal. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 14049–14058 (June 2023)
16. Guo, R., Dai, Q., Hoiem, D.: Single-image shadow detection and removal using paired regions. In: CVPR 2011. pp. 2033–2040 (2011). https://doi.org/10.1109/CVPR.2011.5995725
17. Guo, R., Dai, Q., Hoiem, D.: Paired regions for shadow detection and removal. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**(12), 2956–2967 (2013). https://doi.org/10.1109/TPAMI.2012.214
18. Guo, S., Yan, Z., Zhang, K., Zuo, W., Zhang, L.: Toward convolutional blind denoising of real photographs. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019)
19. Hong, Y., Niu, L., Zhang, J.: Shadow generation for composite image in real-world scenes. AAAI (2022)
20. Hu, X., Fu, C.W., Zhu, L., Qin, J., Heng, P.A.: Direction-aware spatial context features for shadow detection and removal. IEEE Transactions on Pattern Analysis and Machine Intelligence **42**(11), 2795–2808 (2020). https://doi.org/10.1109/TPAMI.2019.2919616
21. Hu, X., Jiang, Y., Fu, C.W., Heng, P.A.: Mask-shadowgan: Learning to remove shadows from unpaired data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (October 2019)
22. Hu, X., Wang, T., Fu, C.W., Jiang, Y., Wang, Q., Heng, P.A.: Revisiting shadow detection: A new benchmark dataset for complex world. IEEE Transactions on Image Processing **30**, 1925–1934 (2021). https://doi.org/10.1109/TIP.2021.3049331
23. Jin, Y., Sharma, A., Tan, R.T.: Dc-shadownet: Single-image hard and soft shadow removal using unsupervised domain-classifier guided network. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 5027–5036 (October 2021)
24. KaewTraKulPong, P., Bowden, R.: An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection, pp. 135–144. Springer US, Boston, MA (2002). https://doi.org/10.1007/978-1-4615-0913-4_11, https://doi.org/10.1007/978-1-4615-0913-4_11
25. Lalonde, J.F., Efros, A.A., Narasimhan, S.G.: Detecting ground shadows in outdoor consumer photographs. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) Computer Vision – ECCV 2010. pp. 322–335. Springer Berlin Heidelberg, Berlin, Heidelberg (2010)
26. Le, H., Samaras, D.: Shadow removal via shadow image decomposition. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (October 2019)
27. Le, H., Samaras, D.: From shadow segmentation to shadow removal. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M. (eds.) Computer Vision – ECCV 2020. pp. 264–281. Springer International Publishing, Cham (2020)
28. Li, Z., Snavely, N.: Learning intrinsic image decomposition from watching the world. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2018)

29. Liu, D., Long, C., Zhang, H., Yu, H., Dong, X., Xiao, C.: Arshadowgan: Shadow generative adversarial network for augmented reality in single light scenes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2020)
30. Liu, F., Gleicher, M.: Texture-consistent shadow removal. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) Computer Vision – ECCV 2008. pp. 437–450. Springer Berlin Heidelberg, Berlin, Heidelberg (2008)
31. Liu, Q., You, J., Wang, J., Tao, X., Zhang, B., Niu, L.: Shadow generation for composite image using diffusion model. CoRR (2024)
32. Liu, Z., Yin, H., Mi, Y., Pu, M., Wang, S.: Shadow removal by a lightness-guided network with training on unpaired data. IEEE Transactions on Image Processing **30**, 1853–1865 (2021). https://doi.org/10.1109/TIP.2020.3048677
33. Liu, Z., Yin, H., Wu, X., Wu, Z., Mi, Y., Wang, S.: From shadow generation to shadow removal. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4927–4936 (June 2021)
34. M Le, H., Goncalves, B., Samaras, D., Lynch, H.: Weakly labeling the antarctic: The penguin colony case. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (June 2019)
35. Qu, L., Tian, J., He, S., Tang, Y., Lau, R.W.H.: Deshadownet: A multi-context embedding deep network for shadow removal. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017. pp. 2308–2316. IEEE Computer Society (2017). https://doi.org/10.1109/CVPR.2017.248, https://doi.org/10.1109/CVPR.2017.248
36. Sen, M., Chermala, S.P., Nagori, N.N., Peddigari, V., Mathur, P., Prasad, B.H.P., Jeong, M.: Shards: Efficient shadow removal using dual stage network for high-resolution images. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 1809–1817 (January 2023)
37. Shin, J., Park, H., Paik, J.: Region-based dehazing via dual-supervised triple-convolutional network. IEEE Transactions on Multimedia **24**, 245–260 (2022). https://doi.org/10.1109/TMM.2021.3050053
38. Stander, J., Mech, R., Ostermann, J.: Detection of moving cast shadows for object segmentation. IEEE Transactions on Multimedia **1**(1), 65–76 (1999). https://doi.org/10.1109/6046.748172
39. Sultana, M., Mahmood, A., Jung, S.K.: Unsupervised moving object detection in complex scenes using adversarial regularizations. IEEE Transactions on Multimedia **23**, 2005–2018 (2021). https://doi.org/10.1109/TMM.2020.3006419
40. Vasluianu, F.A., Romero, A., Van Gool, L., Timofte, R.: Shadow removal with paired and unpaired learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 826–835 (June 2021)
41. Vasluianu, F.A., Seizinger, T., Timofte, R.: Wsrd: A novel benchmark for high resolution image shadow removal. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 1826–1835 (June 2023)
42. 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: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) Computer Vision – ECCV 2016. pp. 816–832. Springer International Publishing, Cham (2016)
43. Vicente, T.F.Y., Samaras, D.: Single image shadow removal via neighbor-based region relighting. In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) Computer

Vision - ECCV 2014 Workshops. pp. 309–320. Springer International Publishing, Cham (2015)
44. Wan, J., Yin, H., Wu, Z., Wu, X., Liu, Y., Wang, S.: Style-guided shadow removal. In: European Conference on Computer Vision. pp. 361–378. Springer (2022)
45. 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 (CVPR) (June 2018)
46. Wang, T., Hu, X., Wang, Q., Heng, P.A., Fu, C.W.: Instance shadow detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2020)
47. Xu, Y., Lin, M., Yang, H., Chao, F., Ji, R.: Shadow-aware dynamic convolution for shadow removal. Pattern Recognition **146**, 109969 (2024). https://doi.org/https://doi.org/10.1016/j.patcog.2023.109969, https://www.sciencedirect.com/science/article/pii/S0031320323006672
48. Yang, Q., Tan, K.H., Ahuja, N.: Shadow removal using bilateral filtering. IEEE Transactions on Image Processing **21**(10), 4361–4368 (2012). https://doi.org/10.1109/TIP.2012.2208976
49. Yu, Q., Zheng, N., Huang, J., Zhao, F.: Cnsnet: A cleanness-navigated-shadow network for shadow removal. In: European Conference on Computer Vision. pp. 221–238. Springer (2022)
50. Zhang, L., Zhang, Q., Xiao, C.: Shadow remover: Image shadow removal based on illumination recovering optimization. IEEE Transactions on Image Processing **24**(11), 4623–4636 (2015). https://doi.org/10.1109/TIP.2015.2465159
51. Zhang, W., Zhao, X., Morvan, J.M., Chen, L.: Improving shadow suppression for illumination robust face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **41**(3), 611–624 (2019). https://doi.org/10.1109/TPAMI.2018.2803179
52. Zhang, X., Zhao, Y., Gu, C., Lu, C., Zhu, S.: Spa-former:an effective and lightweight transformer for image shadow removal. In: 2023 International Joint Conference on Neural Networks (IJCNN). pp. 1–8 (2023). https://doi.org/10.1109/IJCNN54540.2023.10191081
53. Zhong, Y., Liu, X., Zhai, D., Jiang, J., Ji, X.: Shadows can be dangerous: Stealthy and effective physical-world adversarial attack by natural phenomenon. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 15345–15354 (June 2022)
54. Zhu, J., Samuel, K.G.G., Masood, S.Z., Tappen, M.F.: Learning to recognize shadows in monochromatic natural images. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 223–230 (2010). https://doi.org/10.1109/CVPR.2010.5540209
55. 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 (ECCV) (September 2018)
56. Zhu, Y., Huang, J., Fu, X., Zhao, F., Sun, Q., Zha, Z.J.: Bijective mapping network for shadow removal. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5627–5636 (June 2022)