


Supplementary Materials of "ADSP: Advanced Dataset for Shadow Processing, enabling visible occluders via synthesizing strategy."

Chang-Yu Hsieh¹ and Jian-Jiun Ding¹ 

Grad. Inst. Commun. Eng., National Taiwan University, Taiwan
darkrepulser.ray@gmail.com jjding@ntu.edu.tw

1 Experimental setup details

In this section, we provide implementation details about Domain Shift Validation (*i.e.* Sec 4.2. in main text) and Comparison with State-of-the-Art Methods (*i.e.* Sec 4.4. in main text).

Dataset division There are four benchmarks (SRD[[Qu+17](#)], ISTD[[WLY18](#)], DESOBAv2[[Liu+24](#)], and our ADSP) included in our work. First, we used the official splitting for two popular datasets for the removal tasks (SRD and ISTD), as shown in Tab. 1. Second, because the DESOBAv2 was built for shadow generation, its original form is not feasible for training supervised shadow removal algorithms. Specifically, for shadow images in the DESOBAv2, there might be multiple separate shadows in a single image. On the other hand, the provided shadow-free data might be a set of images. In such sets, each image consists of incomplete shadow-free information, *i.e.*, only shadow in a specific target region was removed. Moreover, every shadow-free set's union might contain some shadows that have not been removed. Figure 1 demonstrate a example of shadow-free image set in DESOBAv2. We found that all six images contained limited shadow-free information, only the golf balls, while the most evident shadows cast by two people were not included. Thus, we stated this condition in the main text as "partial-shadow-free information". In our work, we combined shadow-free images/shadow masks with the same prefix and synthesized ground truth as cleanly as possible. There are 20296 pairs after combination (originally 26407 shadow-free images). We randomly pick 296 images from them as the testing set and the remaining 20000 as the training set. Third, for the proposed ADSP, we adopted random splitting to form training (1100 pairs) and testing (120) sets.

Improved metrics calculation As stated in Sec 4.2. in the main text, we improved the metrics calculation method and reported the results of the new evaluation algorithm. Many previous papers adopt RSME, PSNR, and SSIM of



Fig. 1: A shadow-free image set example from the DESOBv2. The first row is all shadow-free images with the prefix 21, and the second is the corresponding shadow masks. Six removed shadows belong to golf balls, and the large shadows cast by two men were not eliminated.

Table 1: Number of pairs in training/testing set of each used benchmark. Note that there are initially 408 pairs in the testing set of the SRD. We excluded 15 images that do not have the corresponding masks.

Set\Benchmark	SRD[Qu+17]	ISTD[WLY18]	DESOBv2[Liu+24]	ADSP (ours)
Training	2680	1330	20000	1100
Testing	393	540	296	120

the whole image, shadow region, and non-shadow region to indicate the performance of different parts of the results. Where the commonly steps for calculating region metrics is as follows:

1. Determining the target (interest) region by the mask, usually shadow mask.
2. Assigning a substitute value (usually 0) to replace the original RGB value of the non-interest region.
3. Calculating each metric value on such post-processed image.

The first, second, and fifth columns of Fig. 2 show the evaluation results following the above steps. Evaluating the shadow region will assign the non-shadow region with zero value (*i.e.* the black area) and vice-versa. The metrics value of the shadow region seems to be better than those of the whole image. It is unreasonable because the bias of the two images primarily comes from the shadow region. We ascribe this to the inappropriate prior of viewing the non-interested region of the image as totally correct, which affects the precision of evaluation, especially for images with a low ratio of shadow region. Therefore, we adopted an improved procedure to do evaluation, as follows:

1. Determining the target (interest) region by the mask, usually shadow mask.
2. Filtering the non-interest region out and calculating each metric value only on region which was not filtered.

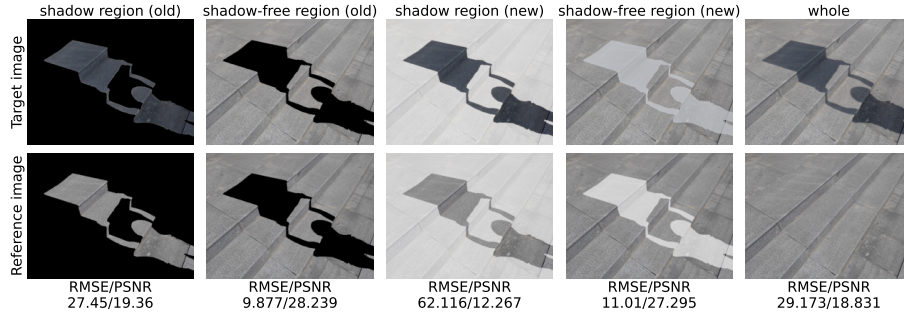


Fig. 2: An evaluation sample using two algorithms.

The third and fourth columns of Fig. 2 show the results of the improved procedure. The excluded region (*i.e.*, the transparent area) will not participate in the metrics calculation. In this way, both RSMs and PSNRs follow the correct order of non-shadow, whole, and shadow. However, this improved process has the side effect of relatively lower shadow region metrics value than previous works.

2 Reproduction

In this section, we give detailed experimental setups about the proposed SRRN, hoping that is helpful to the reader who wants to reproduce our result. Note that we adopted a different reference style to present the citations in the supplementary document to avoid confusion with indexes of citations in the main text.

Overview The SRRN contains three subnetworks. Stage one (θ_{sr}) concentrates on conducting preliminary removal. Stage two (θ_{ca} and θ_{bs}) aims to refine the results from Stage one in two aspects (shadow area color adjustment and shadow boundary smoothing). As mentioned in the main text, we trained the SRRN with two stages. The detailed steps are as follows:

1. We first trained θ_{sr} by the mixed data from the SRD[Qu+17], the ISTD[WLY18], and the proposed ADSP, where the mask information of the SRD applied the results of shadow detection from Cun et al. [CPS20].
2. We conduct preliminary removal with pre-trained θ_{sr} on both training and validation sets of the ADSP and generate elementary removal results.
3. We trained θ_{ca} and θ_{bs} on the preliminary recovered ADSP in step 2. At the same time, the penumbra loss was included to constrain the output of θ_{bs} .

Stage 1. The hyper-parameters to train θ_{sr} are summarized in Table 2:

θ_{sr} contains the input of three shadow patches and the corresponding shadow mask patches in a single batch. They were randomly cropped from the original

Table 2: Detailed setups of θ_{sr} in the proposed SRRN.

Backbone	ShadowFormer [Guo+23]	
Basic	Epoch	500
	Batch size	3
Input	Resizing size	-
	Cropping size	320 x 320 x 3
	Float point accuracy	float32
	Value range	[0, 1]
	Augmentation 1	Randomly rotation/vertical flip
	Augmentation 2	Randomly mix images within batch
Optimizer	AdamW[LH19]	$lr = 2e - 4, \beta = [0.9, 0.999]$ $eps = 1e - 8, weight_decay = 0.02$
Scheduler	Gradual Warmup Cosine Annealing	$multiplier = 1, total_epoch = 3$ $T_max = 497, eta_min = 1e - 6$
Loss function	Total model	Charbonnier Loss

images in the combined dataset. The cropping patch is square with a side length of 320 pixels. Thus, shadow patches and masks have the shape of (3, 3, 320, 320) and (3, 1, 320, 320), respectively. The adopted data augmentation method consists of two parts. The first part has regular random rotation (0, 90, 180, or 270 degrees) and random flip (no flip or vertical flip). Therefore, there are eight possible combinations based on the above two kinds of augmentation methods. Every batch applied one of the above eight augmentations in the entire training process. The second part is to perform patch mixing within a single batch after epoch 5. The detailed steps are as follows:

1. Generate a copy of a single batch and shuffle its order.
2. Sample a set of numbers from the Beta distribution, where two concentration parameters are all 1.2, *i.e.* $\alpha = \beta = 1.2$.
3. Mix the original and permuted batch using the weights from the sampled set above.

Two schedulers make the learning rate rise from 0 first and then reduce. The peak is at epoch 3, *i.e.*, $total_epoch$ of the Gradual Warmup Scheduler. Then, the learning rate decays following the Cosine Annealing Scheduler. θ_{sr} was constrained by the Charbonnier Loss, as shown in Eq. (1).

$$\mathcal{L}_{Charbonnier}(I_{gt}, I_{deshadow}) = \sqrt{(I_{gt} - I_{deshadow})^2 + \epsilon^2}, \epsilon = 10^{-3}. \quad (1)$$

Stage 2. The hyper-parameters to train θ_{ca} and θ_{bs} are as in Table 3:

In this stage, θ_{ca} and θ_{bs} process training pairs 1×1 . Every pair was first resized into 448 by 448 and then randomly cropped as the patches with size 400 by 400. Then, shadow and shadow-free images are converted from RGB to the LAB color space and normalized into $[-1, 1]$. In the data augmentation

Table 3: Detailed setups of θ_{ca} and θ_{bs} in the proposed SRRN.

Backbone	SG-ShadowNet[Wan+22]	
Basic	Epoch	200
	Batch size	1
Input	Resizing size	448 x 448
	Cropping size	400 x 400
	Float point accuracy	float32
	Preprocessing Augmentation	Normalized on LAB space Randomly horizontal flip
Optimizer	Adam[KB15]	$lr = 2e - 4, \beta = [0.5, 0.999]$
Scheduler	Handcraft	Linearly decay to 0 from the epoch 50
Loss function	θ_{ca}	L1 loss, shadow area loss
	θ_{bs}	L1 loss, shadow area loss, spatial consistency loss, penumbra loss

process, we only adopted random horizontal flips. The scheduler is handcrafted linearly attenuated, making the learning rate the same as the initial value, *i.e.*, $2e-4$, during the first 50 epochs and reduced linearly to 0 during the remaining epochs.

Two subnets were supervised by a composite loss function $\mathcal{L}_{overall}$ containing six terms from three kinds of loss functions shown as follows. The pixel-level reconstruction loss \mathcal{L}_R is the L_1 loss of the deshadowed image $I_{deshadow}$ and the ground truth I_{gt} . The area loss \mathcal{L}_A is another L_1 loss with a mask M indicating a specific target region. The last one is the spatial consistency loss \mathcal{L}_{spa} [Guo+20].

$$\mathcal{L}_R(I_{deshadow}, I_{gt}) = \|I_{gt} - I_{deshadow}\|_1, \quad (2)$$

$$\mathcal{L}_A(I_{deshadow}, I_{gt}, M) = \|I_{gt} \otimes M - I_{deshadow} \otimes M\|_1, \quad (3)$$

$$\mathcal{L}_{spa} = \frac{1}{K} \sum_{i=1}^K \sum_{j \in \Omega(i)} (|Y_i, Y_j| - |V_i, V_j|)^2, \quad (4)$$

where \otimes denotes the Hadamard product and M is the mask of the target region. For \mathcal{L}_{spa} , K is the number of local areas, $\Omega(i)$ represents the 4-adjacent areas of area i , and Y and V are the average intensity values of these local areas on $I_{deshadow}$ and I_{gt} , respectively.

Thus, the complete composite loss function $\mathcal{L}_{overall}$ is as follows:

$$\begin{aligned} \mathcal{L}_{overall} &= \mathcal{L}_R^{ca} + \mathcal{L}_R^{bs} + \mathcal{L}_A^{ca} + \mathcal{L}_A^{bs} + 10 \mathcal{L}_{spa}^{bs} + \lambda \mathcal{L}_{penumbra}^{bs} \\ &= \mathcal{L}_R(I_{out}^{ca}, I_{gt}) + \mathcal{L}_R(I_{out}^{bs}, I_{gt}) \\ &+ \mathcal{L}_A(I_{out}^{ca}, I_{gt}, M_s) + \mathcal{L}_A(I_{out}^{bs}, I_{gt}, M_s) \\ &+ 10 \times \mathcal{L}_{spa}^{bs} \\ &+ \lambda \times \mathcal{L}_A(I_{out}^{bs}, I_{gt}, M_p) \end{aligned} \quad (5)$$



Fig. 3: Examples of the generated penumbra mask. The kernel size and the number of iterations of dilation and erosion are 5 and 2, respectively.

where each term has a subscript to indicate the type and a superscript to indicate the model. Therefore, I_{out}^{ca} and I_{out}^{bs} mean the outputs of θ_{ca} and θ_{bs} , respectively, M_s is the shadow mask, and M_p is the penumbra mask. λ is the weight to control the penumbra loss and the hyper-parameter. Their optimization has been discussed in ablation studies.

Penumbra mask. As stated in the main text, we computed the penumbra masks M_p by subtracting the eroded shadow mask from the dilated shadow mask shown as follows.

$$M_p = Dilation(M_s; K = 5; iter = 2) - Erosion(M_s; K = 5; iter = 2) \quad (6)$$

where $Dilation(\cdot; K; iter)$ and $Erosion(\cdot; K; iter)$ represent morphological dilation and erosion operations with kernel K of an adjustable size $k \times k$ and the iteration number of $iter$, respectively. In our implementation, we set $k = 5$, $iter = 2$.

Figure 3 shows some examples of the generated penumbra mask. The resultant masks cover most transition bands between the shadow and non-shadow regions where the boundary effect often appears.

Comparison with SOTA Methods on popular ISTD/SRD In this section, we provide comparisons with SOTAs on two popular benchmarks. Table 4 and Tab. 5 show comparison results on the ISTD and the SRD, respectively. Apparently, the SRRNs do not achieve the same leading place as Sec 4.4. We ascribe this to the second stage of the SRRN. As mentioned above, in the second stage of training, we applied only the proposed ADSP as the training set to acquire better refinement results. However, even so, two SRRNs also surpass most of the SOTAs, proving the effectiveness of our three-stage design.

Table 4: The quantitative results of shadow removal using our models and recent methods on the ISTD[WLY18].

Method	RMSE ↓			PSNR ↑			SSIM ↑
	Whole shadow	non-shadow		Whole shadow	non-shadow		Whole
Input Image	26.826	58.165	15.904	20.33	13.35	25.67	0.8843
Mask-ShadowGan	15.988	22.863	14.292	24.83	21.76	26.06	0.8978
DC-ShadowNet	19.582	26.169	17.760	23.07	20.45	24.09	0.8764
Fu et al.	15.712	17.252	15.561	25.33	24.77	25.77	0.8946
SG-ShadowNet	9.886	13.196	9.017	29.29	26.71	30.32	0.9225
BMNet	10.777	15.487	9.702	28.50	25.16	29.70	0.9211
SpA-Former	14.778	19.925	13.465	25.78	23.32	26.75	0.8917
ShadowFormer	8.733	11.468	7.985	30.64	28.00	31.66	0.9289
SADC	10.816	16.226	9.409	28.39	24.74	29.99	0.9232
Ours ($\lambda = 1$)	9.265	13.053	8.224	29.91	26.80	31.19	0.9227
Ours ($\lambda = 10$)	9.399	13.316	8.321	29.69	26.47	30.98	0.9229

Table 5: The quantitative results of shadow removal using our models and recent methods on the SRD[Qu+17].

Method	RMSE ↓			PSNR ↑			SSIM ↑
	Whole shadow	non-shadow		Whole shadow	non-shadow		Whole
Input Image	37.294	71.964	13.254	17.85	11.51	27.16	0.7896
Mask-ShadowGan	20.416	30.555	13.692	23.59	19.65	26.81	0.8222
DC-ShadowNet	25.645	39.087	16.296	21.35	17.31	25.09	0.7834
Fu et al.	16.330	25.435	10.974	25.35	21.05	28.80	0.8405
SG-ShadowNet	16.644	28.176	10.409	25.23	20.13	29.34	0.8412
BMNet	11.303	16.278	9.028	28.35	24.71	30.61	0.8567
SpA-Former	16.052	22.836	13.042	25.22	21.77	27.31	0.8174
ShadowFormer	9.524	13.099	7.764	29.90	26.66	31.98	0.8752
SADC	14.758	24.005	11.104	25.82	21.23	28.59	0.8267
Ours ($\lambda = 1$)	11.048	16.066	8.464	28.33	24.56	30.98	0.8616
Ours ($\lambda = 10$)	10.831	15.419	8.549	28.48	24.95	30.82	0.8640

References

- [KB15] Diederik P. Kingma and Jimmy Ba. “Adam: A Method for Stochastic Optimization”. In: *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*. Ed. by Yoshua Bengio and Yann LeCun. 2015. URL: <http://arxiv.org/abs/1412.6980> (cit. on p. 5).
- [Qu+17] Liangqiong Qu et al. “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*. IEEE Computer Society, 2017, pp. 2308–2316. DOI: [10.1109/CVPR.2017.248](https://doi.org/10.1109/CVPR.2017.248). URL: <https://doi.org/10.1109/CVPR.2017.248> (cit. on pp. 1–3, 7).

- [WLY18] Jifeng Wang, Xiang Li, and Jian Yang. “Stacked Conditional Generative Adversarial Networks for Jointly Learning Shadow Detection and Shadow Removal”. In: *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*. 2018, pp. 1788–1797. DOI: [10.1109/CVPR.2018.00192](https://doi.org/10.1109/CVPR.2018.00192). URL: http://openaccess.thecvf.com/content%5C_cvpr%5C_2018/html/Wang%5C_Stacked%5C_Conditional%5C_Generative%5C_CVPR%5C_2018%5C_paper.html (cit. on pp. 1–3, 7).
- [LH19] Ilya Loshchilov and Frank Hutter. “Decoupled Weight Decay Regularization”. In: *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. URL: <https://openreview.net/forum?id=Bkg6RiCqY7> (cit. on p. 4).
- [CPS20] Xiaodong Cun, Chi-Man Pun, and Cheng Shi. “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. 07. 2020, pp. 10680–10687 (cit. on p. 3).
- [Guo+20] Chunle Guo et al. “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 (cit. on p. 5).
- [Wan+22] Jin Wan et al. “Style-Guided Shadow Removal”. In: *Computer Vision - ECCV 2022 - 17th European Conference, Tel Aviv, Israel, October 23-27, 2022, Proceedings, Part XIX*. 2022, pp. 361–378. DOI: [10.1007/978-3-031-19800-7_21](https://doi.org/10.1007/978-3-031-19800-7_21). URL: https://doi.org/10.1007/978-3-031-19800-7_21 (cit. on p. 5).
- [Guo+23] Lanqing Guo et al. “ShadowFormer: Global Context Helps Shadow Removal”. In: *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*. Ed. by Brian Williams, Yiling Chen, and Jennifer Neville. AAAI Press, 2023, pp. 710–718. DOI: [10.1609/AAAI.V37I1.25148](https://doi.org/10.1609/AAAI.V37I1.25148). URL: <https://doi.org/10.1609/aaai.v37i1.25148> (cit. on p. 4).
- [Liu+24] Qingyang Liu et al. “Shadow Generation for Composite Image Using Diffusion model”. In: *CoRR* (2024) (cit. on pp. 1, 2).