DC-ShadowNet: Single-Image Hard and Soft Shadow Removal Using Unsupervised Domain-Classifier Guided Network (Supplementary Paper)

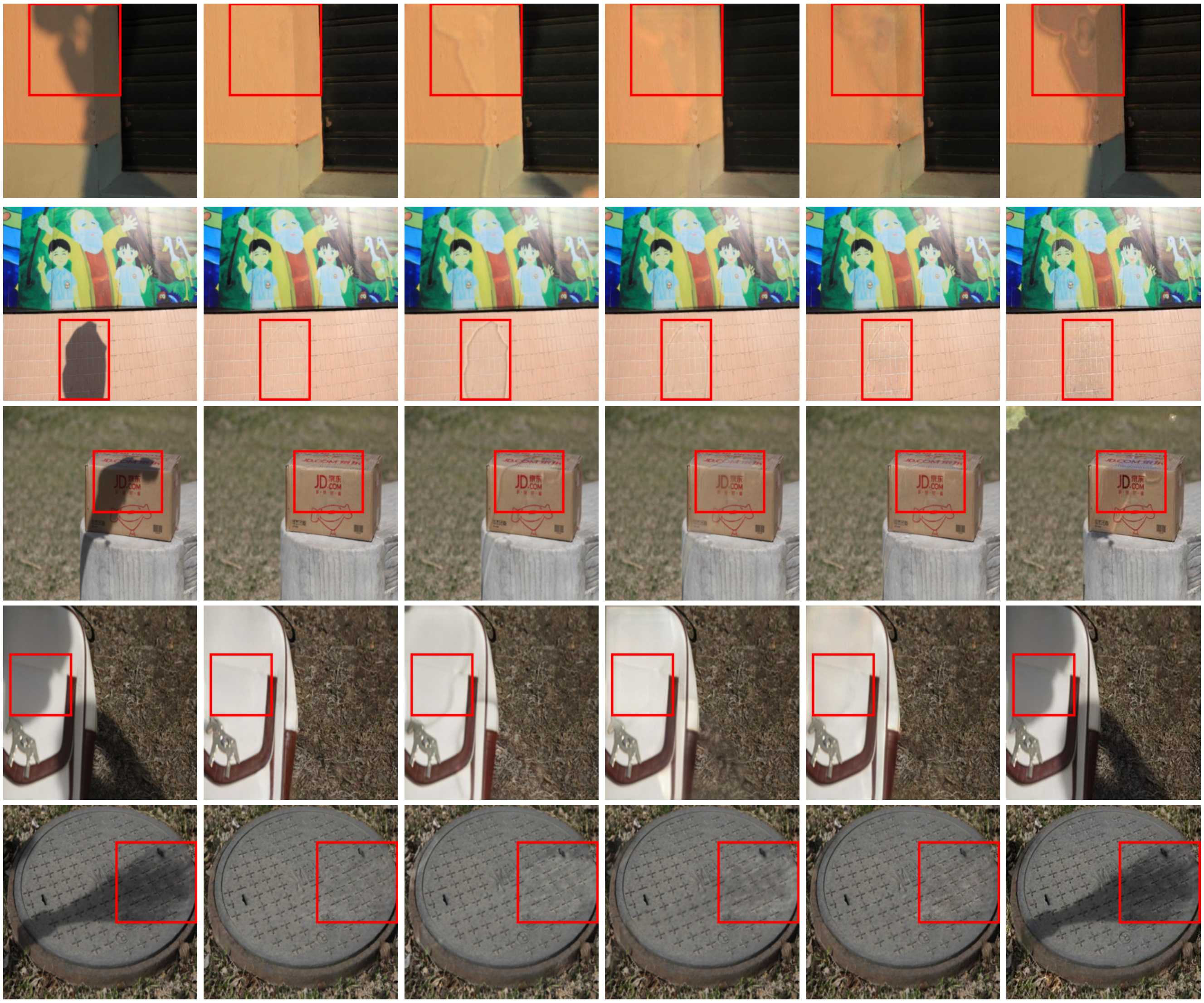
Yeying Jin[1], Aashish Sharma[1], and Robby T. Tan[1,2]
[1]National University of Singapore, [2]Yale-NUS College
jinyeying@u.nus.edu, aashish.sharma@u.nus.edu, robby.tan@{nus,yale-nus}.edu.sg

Figure F1. For shadow removal, existing state-of-the-art unsupervised method Mask-ShadowGAN [11], supervised methods DeShadowNet [17] and DSC [12], and traditional methods Gong [6] fail to remove shadows properly and create artifacts (see regions inside red boxes). Compared to them, our DC-ShadowNet generates a better shadow-free output.

1. Additional Results

We show additional results and comparisons of our method with the baseline methods. To evaluate shadow removal performance, we use the following datasets: SRD [17], AISTD [15], ISTD [18], USR [11] and LRSS [7] (a soft shadow dataset). We also provide more details of our network architecture and training procedure.

Quantitative Evaluation The details of our method and training parameters are provided in Sec. 3 and Sec. 4 of



(a) Input **(b) Ours** **(c)** MaskShadowGAN [11] **(d)** DSC [12] **(e)** DeshadowNet [17] **(f)** Gong *et al*. [6]

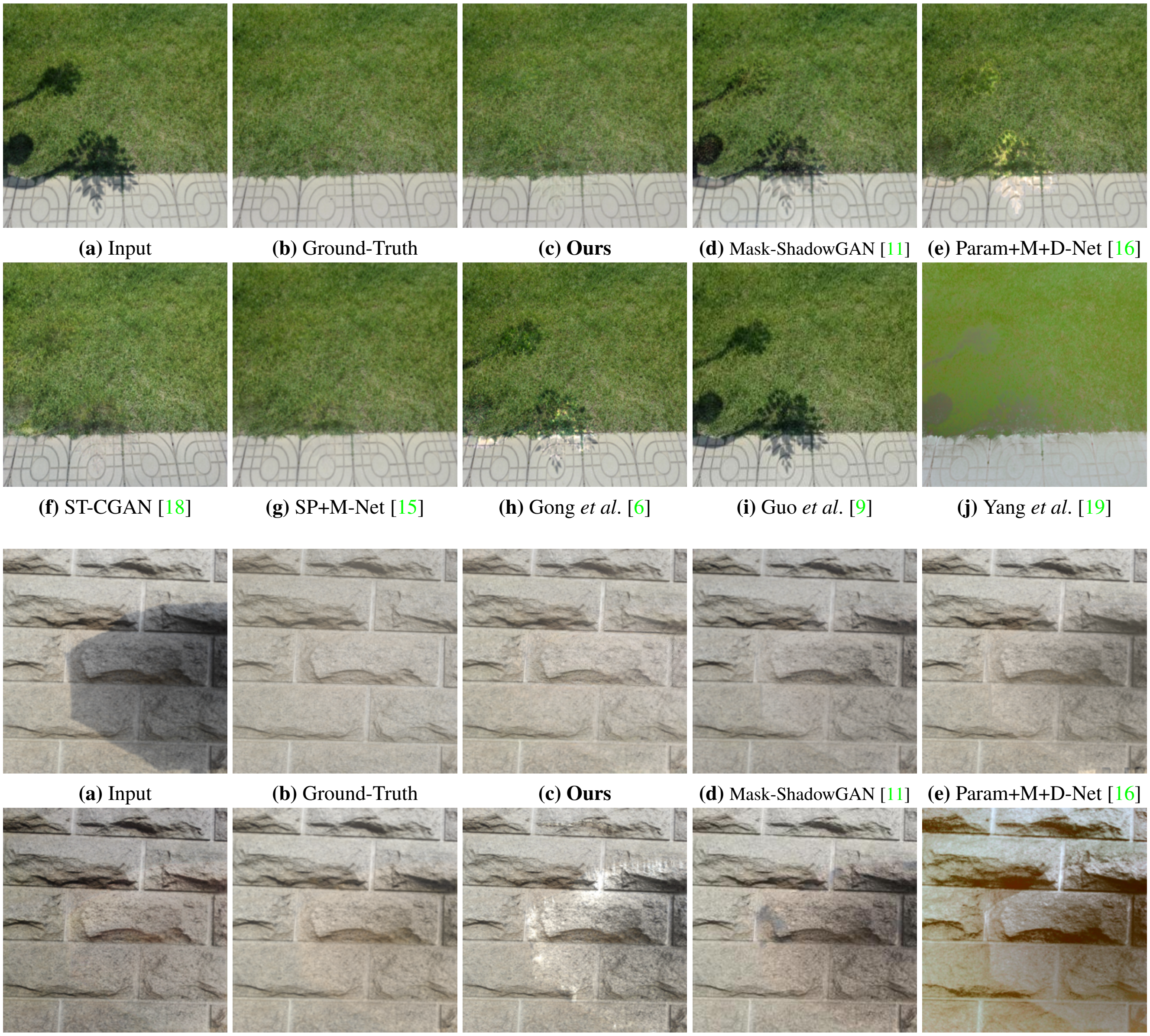
Figure F2. Comparison results on the SRD dataset. Qualitative comparisons of our DC-ShadowNet with the state-of-the-art unsupervised method (c) Mask-ShadowGAN [11], supervised methods (d) DSC [12], (e) DeShadowNet [17], and traditional method (f) Gong *et al*. [6]. Our DC-ShadowNet generates the best shadow-free output image.

the main paper. The quantitative evaluations on the SRD, AISTD, and LRSS datasets are provided in Sec. 4 of the main paper. For all the datasets, the image resolution used in our evaluations is set to 256×256. For the ISTD dataset, the quantitative evaluation is shown in Table T1. From the results, we can observe that our method outperforms all the baseline methods. Compared to the state-of-the-art unsupervised method Mask-ShadowGAN [11], our results for the shadow regions are better by $\sim$12%, showing the effectiveness of our method.

Qualitative Evaluation The qualitative results for the SRD dataset are shown in Figs. F2. For the AISTD dataset, the results are shown in Figs. F3 and F4. For the USR dataset, the results are shown in Fig. F5. For the LRSS soft-shadow dataset, the results are shown in Fig. F6.

We can observe that our DC-ShadowNet produces more robust and accurate shadow removal results than the competing baseline methods. Our results are robust on both hard and soft shadow images, and highlight the effectiveness of our method for shadow removal under diverse scenes and different shadow types.

Network Architecture and Training Details The encoders of our generators are based on ResNet [10] and have four residual blocks. The corresponding decoders use



(a) Input **(b)** Ground-Truth **(c) Ours** **(d)** Mask-ShadowGAN [11] **(e)** Param+M+D-Net [16]

(f) ST-CGAN [18] **(g)** SP+M-Net [15] **(h)** Gong *et al*. [6] **(i)** Guo *et al*. [9] **(j)** Yang *et al*. [19]

(a) Input **(b)** Ground-Truth **(c) Ours** **(d)** Mask-ShadowGAN [11] **(e)** Param+M+D-Net [16]

(f) ST-CGAN [18] **(g)** SP+M-Net [15] **(h)** Gong *et al*. [6] **(i)** Guo *et al*. [9] **(j)** Yang *et al*. [19]

Figure F3. Comparison results on the AISTD dataset. Qualitative comparisons of our DC-ShadowNet with the state-of-the-art unsupervised method (d) Mask-ShadowGAN [11], weakly-supervised method (e) Param+M+D-Net [16], supervised methods (f) ST-CGAN [18], (g) SP+M-Net [15], and traditional methods (h) Gong *et al*. [6], (i) Guo *et al*. [9] and (j) Yang *et al*. [19].

transpose convolutions to generate the output that is of the same resolution as the input. Our discriminators are based on global and local PatchGAN [13] architectures containing five convolution layers. The global discriminator processes an entire image of resolution 286×286 while the local discriminator processes small patches of resolution 70×70 cropped randomly from the image. For data augmentation, we resized images to 286×286, and randomly cropped them to 256×256.

In the training stage, our network is optimized using the Adam method [14] with a constant learning rate of 1×10^{-4} for the first-half of iterations and with a linearly decaying learning rate for the second-half of iterations. For test-time training, we initialize our network with the weights learned during the main training stage and use a smaller learning rate of 1×10^{-6}.

2. Method Details

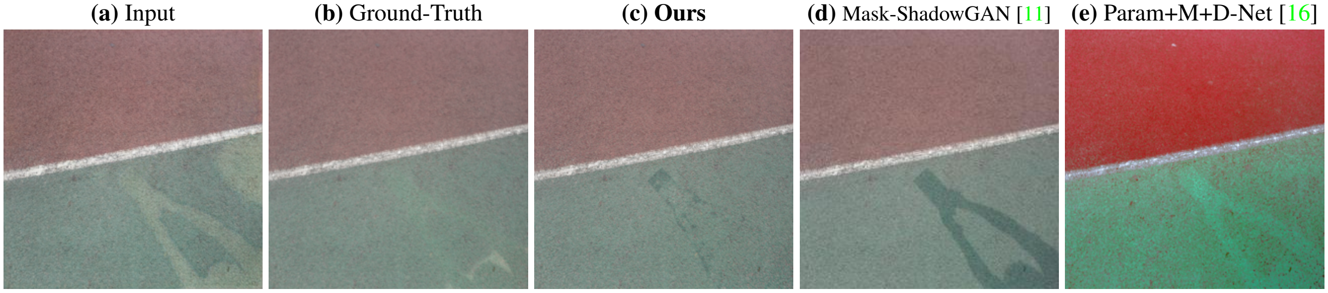
2.1. Shadow-Free Chromaticity

As described in Sec. 3.1 of the main paper, obtaining shadow-free chromaticity $\boldsymbol{\sigma}^{\text{phy}}_{sf}$ from the input shadow image

(a) Input **(b)** Ground-Truth **(c) Ours** **(d)** Mask-ShadowGAN [11] **(e)** Param+M+D-Net [16]

(f) ST-CGAN [18] **(g)** SP+M-Net [15] **(h)** Gong *et al*. [6] **(i)** Guo *et al*. [9] **(j)** Yang *et al*. [19]

(a) Input **(b)** Ground-Truth **(c) Ours** **(d)** Mask-ShadowGAN [11] **(e)** Param+M+D-Net [16]



(f) ST-CGAN [18] **(g)** SP+M-Net [15] **(h)** Gong *et al*. [6] **(i)** Guo *et al*. [9] **(j)** Yang *et al*. [19]

Figure F4. Comparison results on the AISTD dataset. Qualitative comparisons of our DC-ShadowNet with the state-of-the-art unsupervised method (d) Mask-ShadowGAN [11], weakly-supervised method (e) Param+M+D-Net [16], supervised methods (f) ST-CGAN [18], (g) SP+M-Net [15], and traditional methods (h) Gong *et al*. [6], (i) Guo *et al*. [9] and (j) Yang *et al*. [19].

$\mathbf{I}_s$ requires two steps: (1) Entropy Minimization, and (2) Illumination Compensation.

Entropy Minimization We first obtain the log-chromaticity representation of the input shadow image, and then use entropy minimization to find the projection direction θ, which is invariant to shadows [2, 4]. Projecting the log-chromaticity of the input shadow image into the direction orthogonal to θ, we can obtain a shadow-free chromaticity map $\boldsymbol{\sigma}_{sf}^{\text{ent}}$ that no longer contains any shadows (see Figs. F9b for some examples of hard and soft shadow images).

Illumination Compensation To obtain the shadow-free chromaticity $\boldsymbol{\sigma}_{sf}^{\text{ent}}$, we project the original input chromaticity into the direction that is orthogonal to the direction θ. Due to this projection, we lose the original light color; and, to compensate for this, we compute the light color (using about 30% of the brightest pixels from the input shadow image) to obtain the color-compensated shadow-free chromaticity $\boldsymbol{\sigma}_{sf}^{\text{phy}}$. The new shadow-free chromaticity map $\boldsymbol{\sigma}_{sf}^{\text{phy}}$ has the proper light colors (see Figs. F9c for hard and soft

(a) Input **(b) Ours** **(c)** Mask-ShadowGAN [12] **(d)** CycleGAN [20] **(e)** DHAN [1]

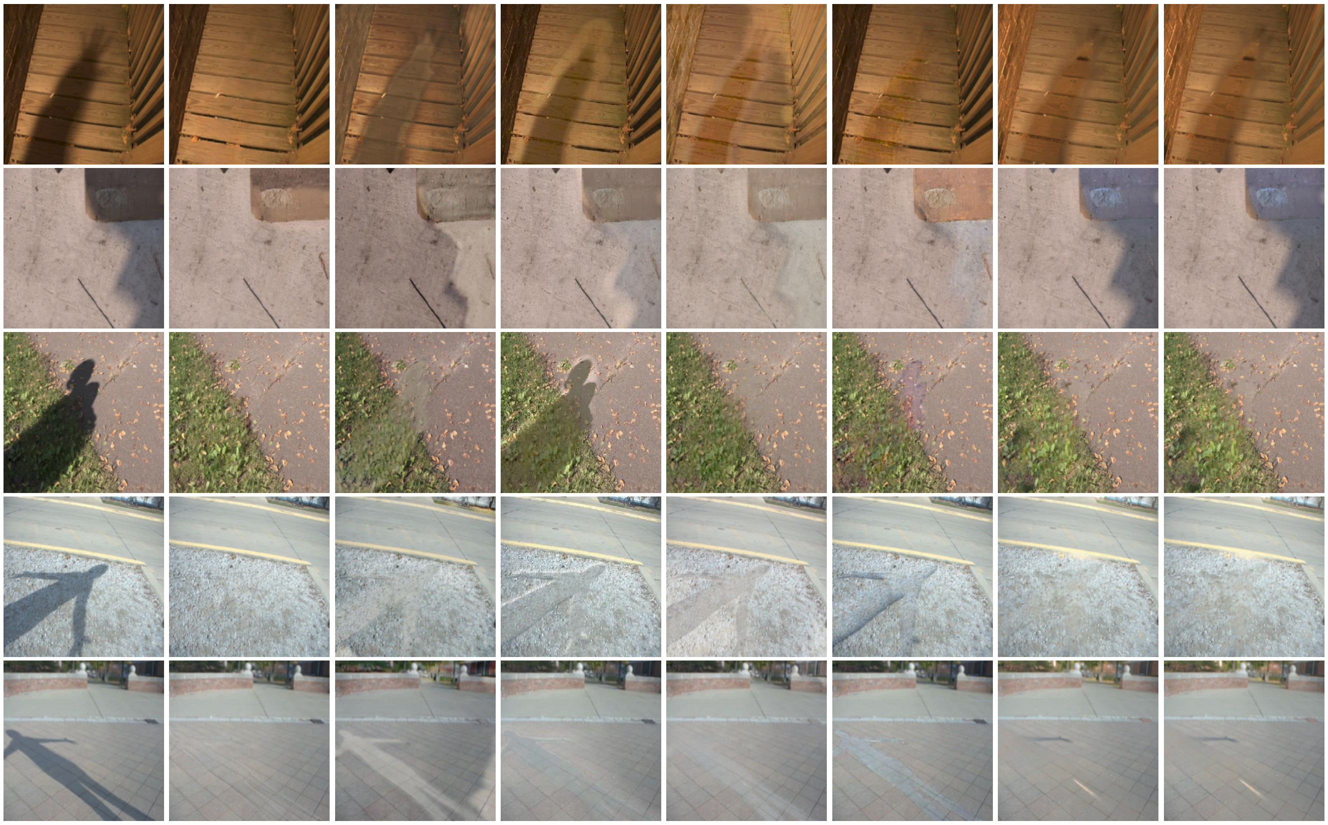
Figure F5. Comparison results on the USR dataset. (a) Input image, (b) Our result, (c) Unsupervised methods Mask-ShadowGAN [11] and (d) CycleGAN, Supervised method (e) DHAN [1]. The results show the better performance of our method.

shadow examples).

To our knowledge, our method is the first method to addresses both hard and soft shadows. While existing physics-based methods [5, 3] can handle hard shadows, they are not designed to handle soft shadows. While we employ the same physics-based techniques of these methods to obtain the shadow-free chromaticity, unlike these methods, we do not use the shadow-free chromaticity to detect shadow edges and use the detected edges for removing shadows [5, 3]. Instead, we use the shadow-free chromaticity to design our novel shadow-free chromaticity loss (see Sec. 3.1 in the main paper) that encourages the output chromaticity to be similar to the shadow-free chromaticity (see Figs. F9c & F9e for some hard and soft shadow examples). This makes our method perform better on hard shadows, and address soft shadows robustly compared to existing physics-based methods even though our method shares the same physics-based techniques with them.

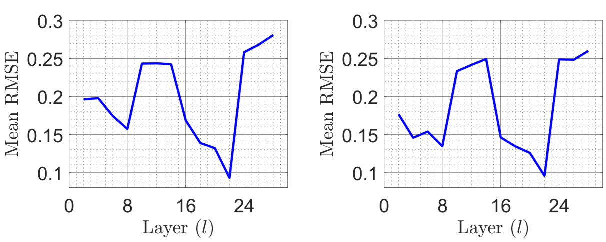
Handling Achromatic Surfaces To compute the correct invariant direction θ, similar to physics-based methods [5, 3, 8], our method assumes that the image surfaces are not achromatic (i.e. the image surfaces are not gray or white), since for achromatic surfaces the entropy minimization can be improper. This is shown in Fig. F10, where for the input shadow images that are nearly achromatic (see Fig. F10a), the corresponding entropy curves show multiple local minimas (see Fig. F10b) leading to improper entropy minimization and inaccurate recovery of shadow-free chromaticity maps (see Fig. F10c). Note that, (1) Given an entropy curve, we use MATLAB function ISLOCALMIN with its PROMINENCE parameter set to 0.05 to obtain the minimas. If multiple minimas are found with similar prominence values, it is assumed that the entropy minimization is improper and for such cases, we do not use our shadow-free chromaticity loss to avoid incorrect supervision to our method; and (2) Since our method also uses other losses such as adversarial and shadow-robust perceptual features (see Sec. 3 in the main paper) to guide our shadow-free output, our method can still do proper shadow removal for achromatic shadow images (see Fig. F10d).

In Fig. F10, we also show comparisons with chromatic shadow input images (see Fig. F10i), where we can see for such images, the entropy curves have prominent global minimas (see Fig. F10j). And, both the obtained shadow-free chromaticity maps and shadow-free outputs are proper (see Figs. F10k and F10l respectively). In sum, our method, DC-ShadowNet, can handle both chromatic and achromatic



(a) Input **(b) Ours** **(c)** [11] **(d)** SP+M-Net [15] **(e)** DHAN [1] **(f)** Gryka [7] **(g)** Guo [9] **(h)** Guo [9] (auto)

Figure F6. Comparison results on the soft shadow LRSS dataset (a) Input image, (b) Our result, (c) Unsupervised method Mask-ShadowGAN [11], Supervised methods (d) SP+M-Net [15] and (e) DHAN [1]. (f)∼(h) are the results of the traditional methods (auto means automatic detection). Our method, trained using unsupervised learning, generates better shadow-free results.



(a) Calibration on LRSS **(b)** Calibration on SRD

Figure F7. Calibration results on both the soft-shadow dataset LRSS and shadow dataset SRD show that the Conv22 layer (i.e. $l = 22$) generates the minimum RMSE error, indicating that it provides features that are most invariant to shadows.

shadow images properly due to the design and usage of our unsupervised loss functions and network architecture.

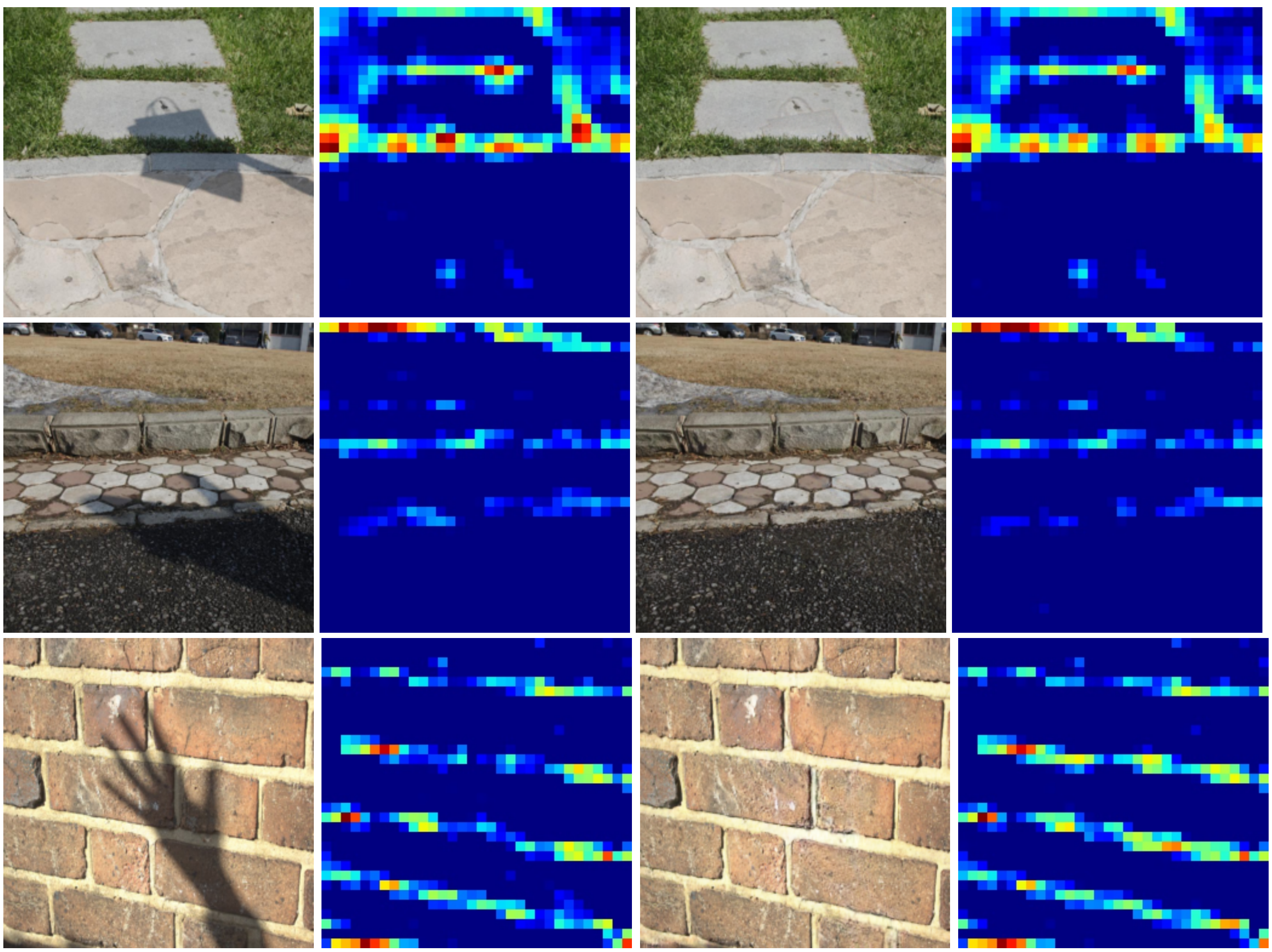
2.2. Shadow-Robust Features

To obtain the shadow-robust features for our shadow-robust feature loss (see Sec. 3.2 in the main paper), we perform a calibration experiment.

Calibration Experiment We take 62 shadow and shadow-free ground-truth image pairs from the LRSS soft-shadow dataset, and 408 shadow and shadow-free ground-truth image pairs from the SRD shadow dataset. For each dataset, given an input shadow image and its corresponding ground-truth shadow-free image, we select a layer l in the pre-trained VGG16 network, and calculate the RMSE score between the features obtained for the input shadow image and the features obtained for the ground-truth shadow-free image. This is repeated for all the image pairs in a dataset after which we compute the mean RMSE score of that dataset. The experimental results are shown Fig F7. The results shown in Fig F7 show that for both the SRD and LRSS datasets, the minimum mean RMSE score is obtained at the Conv22 layer (i.e. $l^* = 22$) of the VGG-16 network. This indicates that the Conv22 layer of the VGG-16 network provides features that are most invariant to shadows (also see Fig. F8 for some qualitative examples), which we use to design our shadow-robust feature loss.

Table T1. RMSE results of our DC-ShadowNet compared to the state-of-the-art shadow removal methods on the ISTD dataset. M shows that ground-truth shadow masks are also used in training.

Method	Training	All	Shadow	Non-Shadow
Our DC-ShadowNet	Unpaired	**5.88**	9.36	**5.19**
Mask-ShadowGAN [11]	Unpaired	6.66	10.68	5.86
Param+M+D-Net [16]	Unpaired + M	7.90	11.73	7.13
ST-CGAN [18]	Paired + M	6.59	9.38	6.03
SP+M-Net [15]	Paired + M	7.73	10.99	7.08
DAD [21]	Paired	6.72	**8.65**	6.39
Gong *et al*. [6]	-	8.10	14.18	6.89
Guo *et al*. [8]	Paired + M	9.59	18.56	7.81
Yang *et al*. [19]	-	15.53	19.54	14.73
Input Image	-	10.85	31.61	6.72



(a) Input Shadow Image ($\mathbf{I}_s$) **(b)** Feature Map ($V(\mathbf{I}_s)$) **(c)** Output Shadow-Free ($\mathbf{Z}_{sf}$) **(d)** Feature Map ($V(\mathbf{Z}_{sf})$)

Figure F8. (a) Input shadow image $\mathbf{I}_s$, (b) Sample feature map for $\mathbf{I}_s$, (c) Output shadow-free image $\mathbf{Z}_{sf}$, and (d) Sample feature map for $\mathbf{Z}_{sf}$. We can observe that features in (b) for the input shadow images are less affected by shadows and they are similar to the features in (d) owing to our shadow-robust feature loss.

3. Ablation Results

The quantitative results for our ablation experiments are shown in Sec. 5 of the main paper. Here we show the corresponding qualitative results. Fig. F11 demonstrates the effectiveness of the shadow-invariant chromaticity loss $\mathcal{L}_{\text{chroma}}$, shadow-robust feature loss $\mathcal{L}_{\text{feature}}$, boundary-smoothness loss $\mathcal{L}_{\text{smooth}}$, and the domain classifiers Φ_s^g and Φ_{sf}^d. As also supported by the qualitative results, each of these components is important in providing better performance to our method for both hard and soft shadow images.

Test-Time Training Test-time training is an effective technique to reduce any possible domain gap between the training and testing data. Since our method uses unsupervised learning, it can be used for test-time training which as we show, further improves the performance of our method. To perform our test-time training, given a test image as input, we finetune our generator G_s (with its weights initialized with the weights learned in the training stage) using our unsupervised shadow-invariant chromaticity, shadow-robust feature, and boundary-smoothness losses, since these losses are only dependent upon the test input image. The finetuning process is carried out until the total loss has reached convergence upon which the generated shadow-free image is taken as the output shadow-free image.

To evaluate the effectiveness of test-time training, we use 34 shadow and ground-truth shadow-free image pairs from the LRSS dataset that we do not use during training (see '**Results on Soft Shadows**' in Sec. 4 of the main paper). We perform the aforementioned finetuning process on each test image and observe that the overall shadow removal performance on the 34 images improves from 3.48 to 3.36 in RMSE and from 31.01 to 31.31 in PSNR. The qualitative results are shown in Fig. F12, and they confirm the improvement brought in by test-time training for our method.

References

[1] Xiaodong Cun, Chi-Man Pun, and Cheng Shi. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting gan. In *AAAI*, pages 10680–10687, 2020. 5, 6

[2] Graham D Finlayson, Mark S Drew, and Cheng Lu. Intrinsic images by entropy minimization. In *European conference on computer vision*, pages 582–595, 2004. 4

[3] Graham D Finlayson, Mark S Drew, and Cheng Lu. Entropy minimization for shadow removal. *International Journal of Computer Vision*, 85(1):35–57, 2009. 5

[4] Graham D Finlayson, Steven D Hordley, and Mark S Drew. Removing shadows from images. In *European conference on computer vision*, pages 823–836, 2002. 4

[5] Graham D Finlayson, Steven D Hordley, Cheng Lu, and Mark S Drew. On the removal of shadows from images. *IEEE transactions on pattern analysis and machine intelligence*, 28(1):59–68, 2005. 5

[6] Han Gong and Darren Cosker. Interactive shadow removal and ground truth for variable scene categories. In *BMVC*, pages 1–11, 2014. 1, 2, 3, 4, 7

[7] Maciej Gryka, Michael Terry, and Gabriel J Brostow. Learning to remove soft shadows. *ACM Transactions on Graphics (TOG)*, 34(5):1–15, 2015. 1, 6

[8] Ruiqi Guo, Qieyun Dai, and Derek Hoiem. Single-image shadow detection and removal using paired regions. In *CVPR 2011*, pages 2033–2040, 2011. 5, 7

[9] Ruiqi Guo, Qieyun Dai, and Derek Hoiem. Paired regions for shadow detection and removal. *IEEE transactions on pattern analysis and machine intelligence*, 35(12):2956–2967, 2012. 3, 4, 6

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2

[11] Xiaowei Hu, Yitong Jiang, Chi-Wing Fu, and Pheng-Ann Heng. Mask-shadowgan: Learning to remove shadows from unpaired data. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2472–2481, 2019. 1, 2, 3, 4, 5, 6, 7, 11

[12] Xiaowei Hu, Lei Zhu, Chi-Wing Fu, Jing Qin, and Pheng-Ann Heng. Direction-aware spatial context features for shadow detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7454–7462, 2018. 1, 2, 5

[13] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. 3

[14] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 3

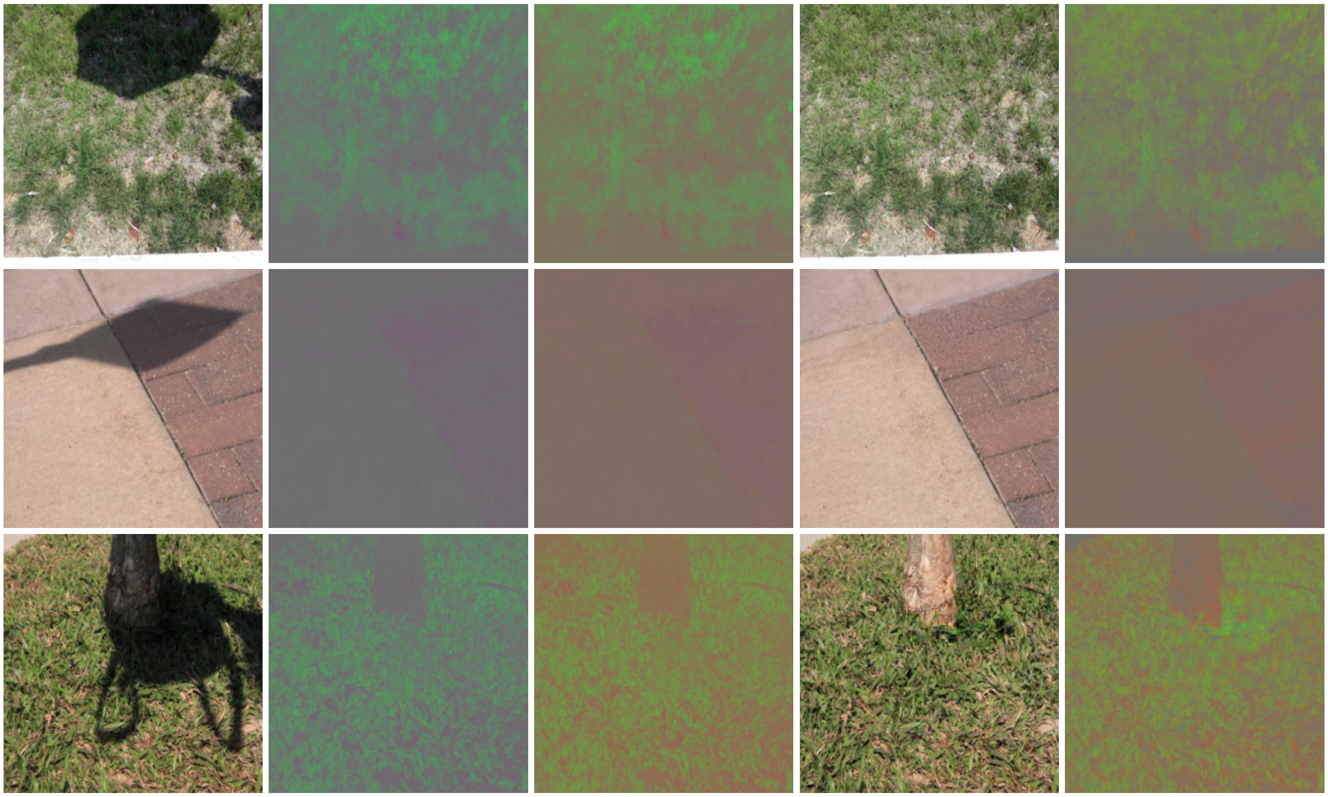
[15] Hieu Le and Dimitris Samaras. Shadow removal via shadow image decomposition. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 8578–8587, 2019. 1, 3, 4, 6, 7

[16] Hieu Le and Dimitris Samaras. From shadow segmentation to shadow removal. In *European Conference on Computer Vision*, pages 264–281. Springer, 2020. 3, 4, 7

[17] Liangqiong Qu, Jiandong Tian, Shengfeng He, Yandong Tang, and Rynson WH Lau. Deshadownet: A multi-context embedding deep network for shadow removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4067–4075, 2017. 1, 2

[18] Jifeng Wang, Xiang Li, and Jian Yang. 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*, pages 1788–1797, 2018. 1, 3, 4, 7

[19] Qingxiong Yang, Kar-Han Tan, and Narendra Ahuja. Shadow removal using bilateral filtering. *IEEE Transactions on Image processing*, 21(10):4361–4368, 2012. 3, 4, 7

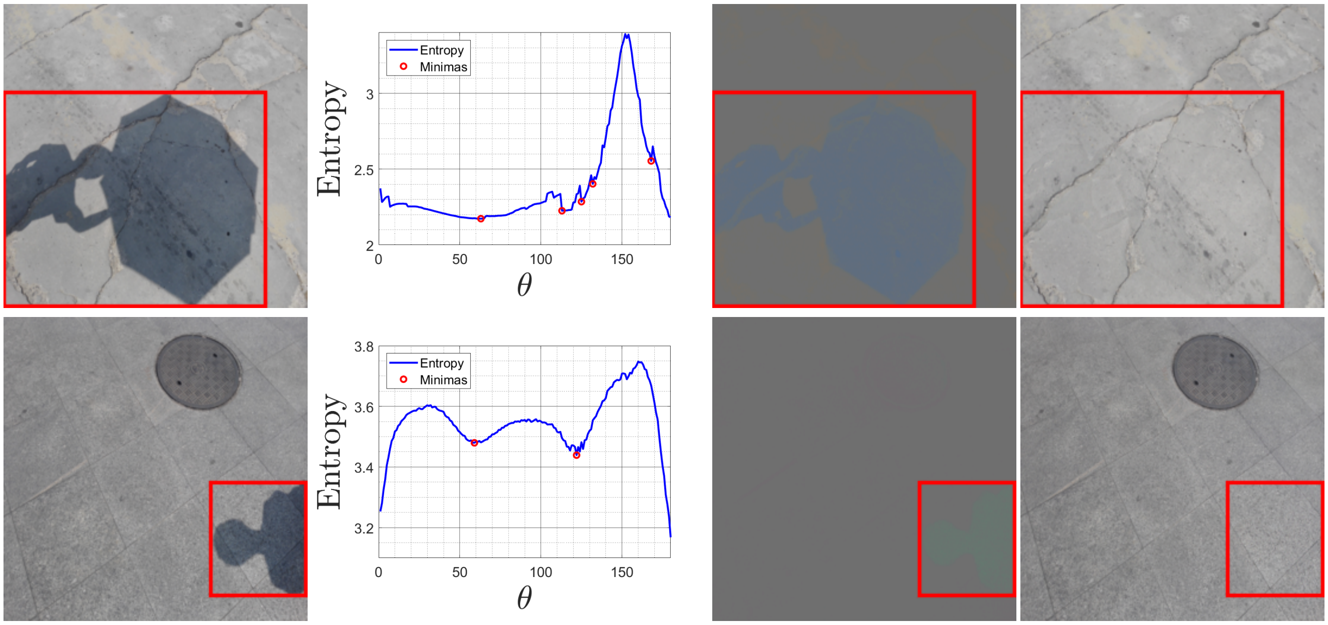


(a) Input $\mathbf{I}_s$ **(b)** $\boldsymbol{\sigma}_{sf}^{\text{ent}}$ **(c)** $\boldsymbol{\sigma}_{sf}^{\text{phy}}$ **(d)** Output $\mathbf{Z}_{sf}$ **(e)** $\boldsymbol{\sigma}_{SF}^{\text{Z}}$

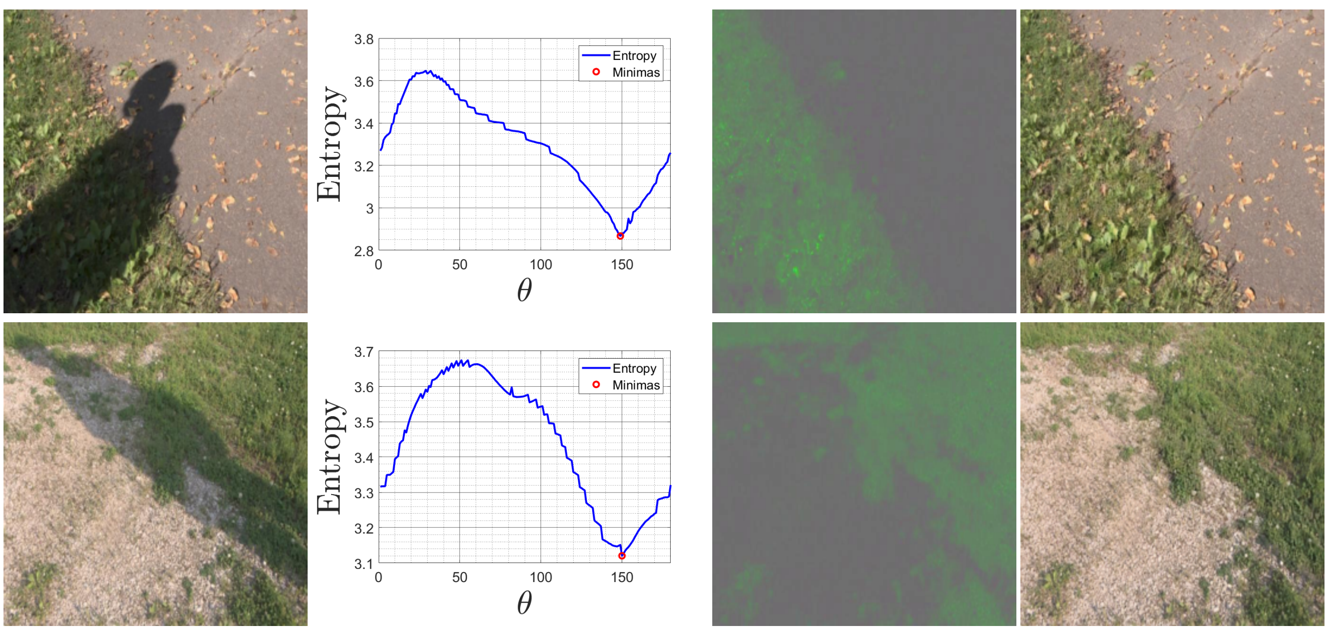
Figure F9. (a) Input shadow image, (b) Shadow-free chromaticity after entropy minimization $\boldsymbol{\sigma}_{sf}^{\text{ent}}$, (c) Shadow-free chromaticity after illumination compensation $\boldsymbol{\sigma}_{sf}^{\text{phy}}$, (d) Output shadow-free image, and (e) Chromaticity map the of output image $\boldsymbol{\sigma}_{sf}^{\mathbf{Z}}$. Our shadow-free chromaticity loss encourages (e) to be similar to (c) facilitating better shadow removal.

[20] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. 5

[21] Zhengxia Zou, Sen Lei, Tianyang Shi, Zhenwei Shi, and Jieping Ye. Deep adversarial decomposition: A unified framework for separating superimposed images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12806–12816, 2020. 7

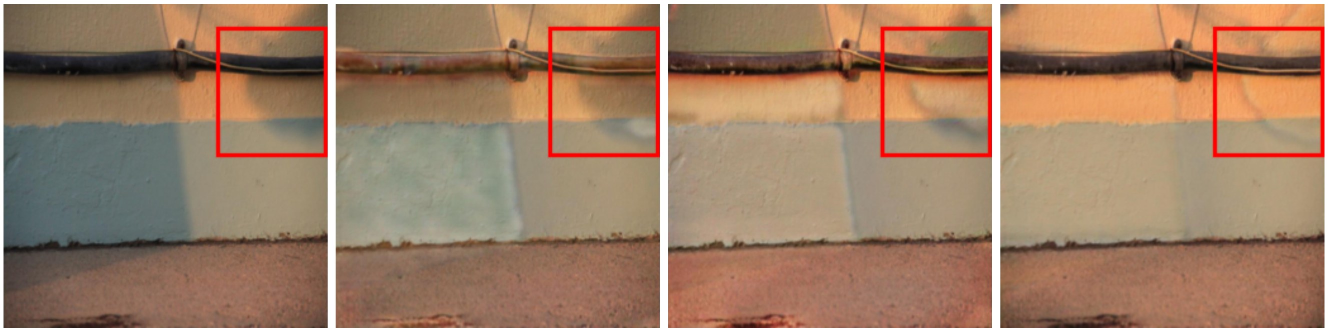


(a) Input Achromatic Image **(b)** Entropy Minimization **(c)** $\boldsymbol{\sigma}_{sf}^{\text{ent}}$ **(d)** Output Image

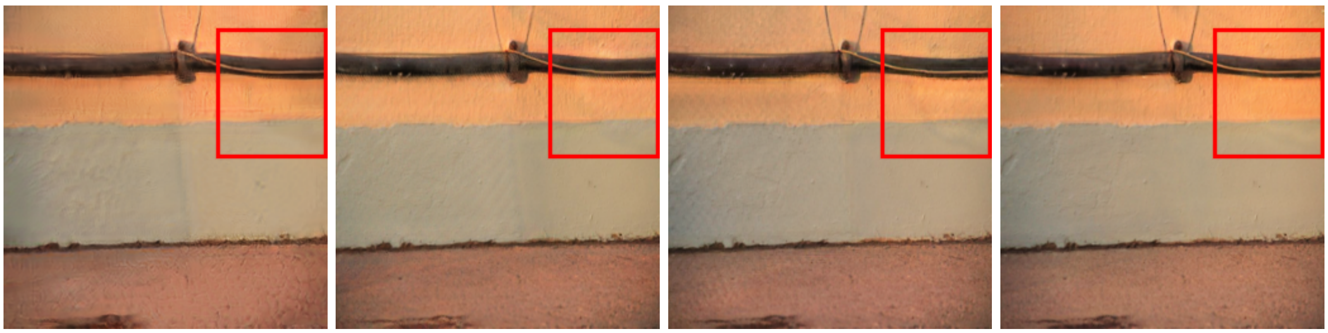


(i) Input Chromatic Image **(j)** Entropy Minimization **(k)** $\boldsymbol{\sigma}_{sf}^{\text{ent}}$ **(l)** Output Image

Figure F10. For achromatic input shadow images shown in (a), the entropy curves shown in (b) have multiple local minimas (due to the presence of inherent noise in the images) and the entropy minimization is improper. This leads to inaccurate recovery of the shadow-free chromaticity maps as shown in (c). However, as shown in (d), due to the presence of our other losses such as adversarial and shadow-robust perceptual features, our DC-ShadowNet is still able to remove shadows properly. For chromatic input images shadow images shown in (i), the entropy curves shown in (j) have distinct global minima leading to proper recovery of shadow-free chromaticity maps and shadow-free outputs shown in (k) and (l) respectively.



(a) Input **(b)** w/o Φ_s^g and Φ_{sf}^d **(c)** w/o Φ_{sf}^d **(d)** w/o Φ_s^g



(e) w/o $\mathcal{L}_{\text{chroma}}$ **(f)** w/o $\mathcal{L}_{\text{feature}}$ **(g)** w/o $\mathcal{L}_{\text{smooth}}$ **(h)** Output

Figure F11. Ablation Results (a) Input shadow image, (b), (c) and (d) are the shadow removal results without our domain classifiers Φ_s^g and Φ_{sf}^d, discriminator domain classifier Φ_{sf}^d, and generator domain classifier Φ_s^g, respectively. (e), (f) and (g) are the shadow removal results without our physics-based shadow-free chromaticity loss, shadow-robust feature loss and boundary smoothness loss respectively, (h) is the output from our method DC-ShadowNet that has all these components. As we can observe, the best results are obtained when all the components are used in our method.

(a) Input Soft Shadow **(b)** Ours (w/o test-time-train) **(c)** Ours (w/ test-time-train) **(d)** Mask-ShadowGAN [11]

Figure F12. Effectiveness of Test-Time Training (a) Input image, (b) and (c) show our results without and with test-time-training, (d) Result of Mask-ShadowGAN [11] for comparison.