

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

Yeying Jin¹, Aashish Sharma¹, and Robby T. Tan^{1,2}

¹National University of Singapore, ²Yale-NUS College

jinyeying@u.nus.edu, aashish.sharma@u.nus.edu, roddy.tan@{nus,yale-nus}.edu.sg

Abstract

Shadow removal from a single image is generally still an open problem. Most existing learning-based methods use supervised learning and require a large number of paired images (shadow and corresponding non-shadow images) for training. A recent unsupervised method, Mask-ShadowGAN [13], addresses this limitation. However, it requires a binary mask to represent shadow regions, making it inapplicable to soft shadows. To address the problem, in this paper, we propose an unsupervised domain-classifier guided shadow removal network, DC-ShadowNet. Specifically, we propose to integrate a shadow/shadow-free domain classifier into a generator and its discriminator, enabling them to focus on shadow regions. To train our network, we introduce novel losses based on physics-based shadow-free chromaticity, shadow-robust perceptual features, and boundary smoothness. Moreover, we show that our unsupervised network can be used for test-time training that further improves the results. Our experiments show that all these novel components allow our method to handle soft shadows, and also to perform better on hard shadows both quantitatively and qualitatively than the existing state-of-the-art shadow removal methods.

1. Introduction

Shadow removal from a single image can benefit many applications, such as image editing, scene relighting, etc., [19, 17, 16]. Unfortunately, in general, removing shadows from a single image is still an open problem. Existing physics-based methods for shadow removal [7, 6, 10] are based on entropy minimization that can capture the invariant features of shadow and non-shadow regions belong to the same surfaces in the log-chromaticity space. These methods, however, tend to fail, particularly when the image surfaces are close to achromatic (e.g. gray or white surfaces), and are not designed to handle soft shadow images.

[†]This work is supported by MOE2019-T2-1-130.

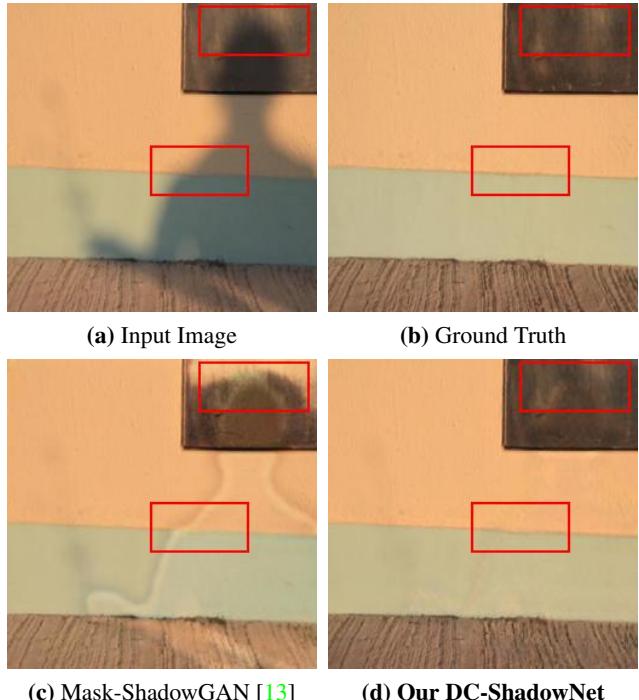


Figure 1. Existing state-of-the-art shadow removal methods like Mask-ShadowGAN [13] fail to remove soft shadows properly and create artifacts (see regions inside red boxes). Compared to it, our method generates a better shadow-free output.

Unlike physics-based methods, deep-learning methods, e.g. [24, 27, 14, 20, 1, 21], are more robust to different conditions of image surfaces and lighting. However, most of these methods are based on fully-supervised learning, which means that for training, they require pairs of shadow and their corresponding non-shadow images. To collect these image pairs in a large amount, particularly for images containing diverse scenes and shadows can be considerably expensive.

Recently, Hu *et al.* propose an unsupervised method, Mask-ShadowGAN [13], the network architecture of which is based on CycleGAN [34]. To remove shadows, the method mainly relies on adversarial training that employs

a discriminator to check the quality of the generated output. Unfortunately, due to the absence of ground truth, the discriminator relies solely on unpaired non-shadow images, which can cause the generator to produce incorrect outputs. Moreover, the method uses a binary mask to represent shadow regions present in the input image, making it inapplicable to soft shadow images. Fig. 1 shows an example where for the given soft-shadow input image, the output generated by the method [13] is improper.

In this paper, our goal is to remove both hard and soft shadows from a single image. To achieve this, we propose *DC-ShadowNet*, an unsupervised network guided by the shadow/shadow-free domain classifier. Specifically, we integrate a domain classifier (that classifies the input image to either shadow or shadow-free domain) into our generator and its corresponding discriminator. This allows our generator and discriminator to focus on shadow regions and thus perform better shadow removal. Unlike the existing unsupervised method [13], which only relies on adversarial training based on an unpaired discriminator (i.e. using unpaired non-shadow images as reference images), our method uses additional novel unsupervised losses that enable our method to achieve better shadow removal results. Our new losses are based on physics-based shadow-free chromaticity, shadow-robust perceptual features, and boundary smoothness.

Our physics-based shadow-free chromaticity loss employs a shadow-free chromaticity image, which is obtained from the input shadow image by performing entropy minimization in the log-chromaticity space [7]. Our shadow-robust perceptual features loss uses shadow-robust features obtained from the input shadow image using the pre-trained VGG-16 network [15]. We also add a boundary smoothness loss to ensure that our output shadow-free image has smoother transitions in the regions that contained shadow boundaries. All these ideas enable our method to better deal with hard and soft shadow images compared to existing methods like [13] (see Fig. 1 for an example showing the better performance of our method). Furthermore, we show that our method being unsupervised can be used for test-time training to further improve the performance of our method. As a summary, here are our contributions:

1. We introduce DC-ShadowNet, a new unsupervised single-image shadow removal network guided by a domain classifier to focus on shadow regions.
2. We propose novel unsupervised losses based on physics-based shadow-free chromaticity, shadow-robust perceptual features, and boundary smoothness losses for robust shadow removal.
3. To our knowledge, our method is the first unsupervised method to perform shadow removal robustly for both hard and soft shadow in a single image.

2. Related work

Physics-based shadow removal methods (e.g. [4, 3, 5, 7, 6]) are based on the physics models of illumination and surface colors. These methods assume that the surface colors in the input image are chromatic, and hence they are erroneous when this assumption does not hold. These methods are designed to remove hard shadows only. In contrast, our method is based on unsupervised learning and is designed to handle both hard and soft shadows. Also, our method is more robust in dealing with achromatic surfaces.

Some other non-learning-based methods rely on user interaction. Gryka *et al.* [9] propose a regression model to learn a mapping function of shadow image regions and their corresponding shadow mattes. However, they need the user to provide brush strokes to relight shadow regions. Guo *et al.* [10, 11] use annotated ground truth to learn the appearances of shadow regions. Unlike these methods, our method is learning-based and does not rely on hand-crafted feature descriptors, making it more robust. Moreover, our method does not need any annotated ground truth and user interaction; hence, it is more practical and efficient.

To address the aforementioned limitations of non-deep learning methods, many deep learning methods are proposed. Wang *et al.* [27] use a stacked conditional GAN (ST-CGAN) to detect and remove shadows jointly. Le *et al.* [20, 21] propose SP+M-Net do shadow removal using image decomposition. Hu *et al.* [14, 12] propose to add global and direction-aware context into the direction-aware spatial context (DSC) module. Ding *et al.* [2] introduce an LSTM-based attentive recurrent GAN (ARGAN) to detect and remove shadows. All these methods are trained on paired data using supervised learning. Hence, training them using various soft shadows and complex scenes is difficult, since obtaining the ground truths is intractable. In contrast, our method is based on unsupervised learning and does not need any paired data.

Recently, Hu *et al.* [13] propose an unsupervised deep-learning method Mask-ShadowGAN. Unfortunately, since it mainly relies on adversarial training for shadow removal, it cannot guarantee that the generated output images are shadow-free since there is no strong guidance for the network to do so. Moreover, it cannot handle soft shadows due to the use of binary masks. In contrast, our method DC-ShadowNet uses new additional unsupervised losses and domain-classifier guided network that helps our method to more effectively deal with hard and soft shadows.

3. Proposed Method

Fig. 2 shows the architecture of our network, DC-ShadowNet. Given a shadow input image, \mathbf{I}_s , we use a generator, G_s , to transform it into a shadow-free output image \mathbf{Z}_{sf} . Also, given an unpaired shadow-free input image, \mathbf{I}_{sf} ,

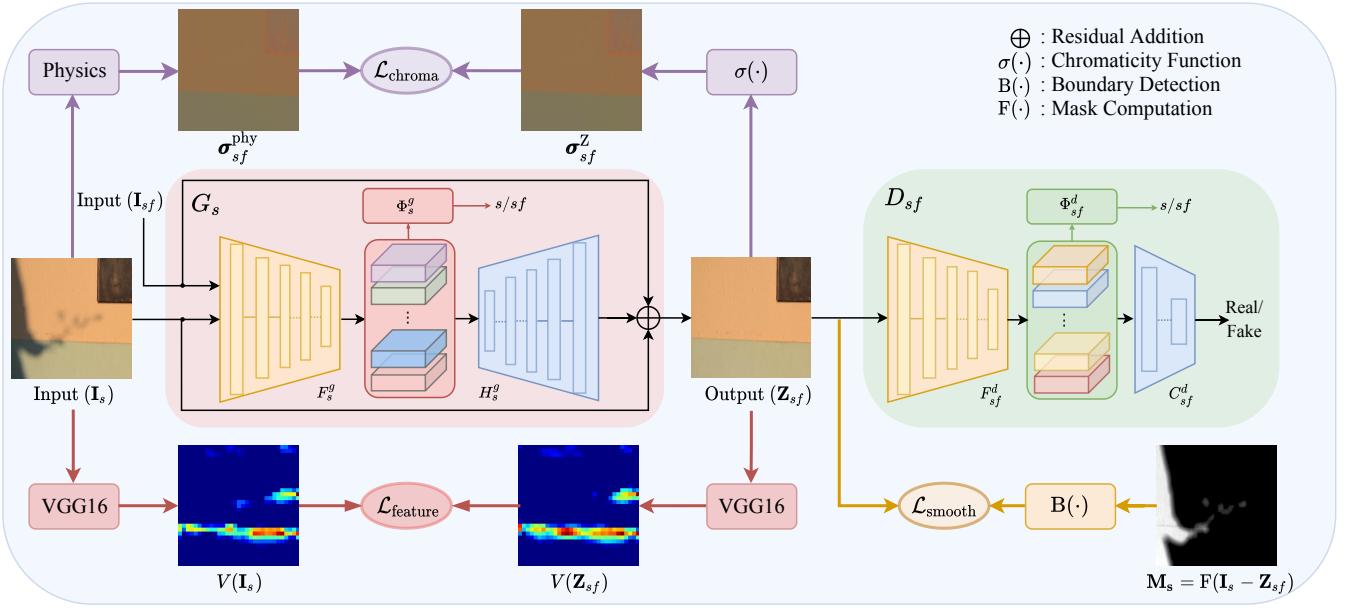


Figure 2. Network Architecture of Our DC-ShadowNet. We have two domains: shadow, s , and shadow-free, sf . Our shadow removal generator is represented by G_s . It consists of an encoder F_s^g , a decoder H_s^g , and a domain classifier Φ_s^g . We also use a discriminator D_{sf} that consists of its own encoder F_{sf}^d , a classifier C_{sf}^d and a domain classifier Φ_{sf}^d . For the input shadow image \mathbf{I}_s , its corresponding output shadow-free image is represented by \mathbf{Z}_{sf} . Also, for the unpaired input shadow-free image \mathbf{I}_{sf} , G_s reconstructs the image back. The domain classifiers, Φ_s^g and Φ_{sf}^d , are used to classify whether the inputs to their respective networks, G_s and D_{sf} , belong to shadow (s) or shadow-free (sf) domain. To guide our generator G_s to do shadow removal, other than adversarial loss from the discriminator D_{sf} , we include novel losses: shadow-free chromaticity loss \mathcal{L}_{chroma} (purple) guided by the physics-based shadow-free chromaticity σ_{sf}^{phy} obtained from \mathbf{I}_s ; shadow-robust feature loss $\mathcal{L}_{feature}$ (red) guided by the shadow-robust perceptual features $V(\mathbf{I}_s)$ obtained from \mathbf{I}_s , and boundary smoothness loss \mathcal{L}_{smooth} (orange) guided by the boundary detection of our generated soft shadow mask \mathbf{M}_s .

we expect the generator, G_s , to simply reconstruct the image back. Therefore, the generator G_s , whether its input is a shadow or shadow-free image, always generates a shadow-free output image. Note that, in our method, we have two domains: shadow, s , and shadow-free, sf .

Our generator G_s consists of an encoder (F_s^g), decoder (H_s^g) and a domain classifier (Φ_s^g). We use a discriminator D_{sf} to assess the quality of the shadow removal output. It consists of an encoder (F_{sf}^d), a classifier (C_{sf}^d) and a domain classifier (Φ_{sf}^d). Both the domain classifiers, Φ_s^g and Φ_{sf}^d , are used to classify the inputs of their respective modules, G_s and D_{sf} , belonging to either shadow or shadow-free domain. However, unlike Φ_s^g , which is trained together with G_s , Φ_{sf}^d is pre-trained, and its weights are kept frozen while training D_{sf} . The underlying idea of integrating the domain classifier into our generator and its discriminator is to guide our network to focus on shadow regions. The reference images of our discriminator are the unpaired shadow-free real images. Our discriminator's classifier, C_{sf}^d , outputs the real/fake binary label, where real refers to the label given to an image that belongs to the reference images.

While not shown in Fig. 2, for the sake of clarity, we employ another generator G_{sf} and the shadow mask to transform the shadow-free output image back to a shadow image, in order to enforce reconstruction consistency [34] and lo-

cate the shadow regions. Also, another discriminator D_s is used to distinguish whether the generated shadow image is real or not. Our method, DC-ShadowNet, is trained in an unsupervised manner using our losses, which are described in the following sections.

3.1. Shadow-Free Chromaticity Loss

Given a shadow input image \mathbf{I}_s , we obtain a physics-based shadow-free chromaticity image σ_{sf}^{phy} , which is used to guide our shadow removal generator G_s , through our shadow-free chromaticity loss function. Obtaining σ_{sf}^{phy} from \mathbf{I}_s requires two steps: (1) Entropy Minimization, and (2) Illumination Compensation.

Entropy Minimization Following [6], as shown in Fig. 3, we plot the input shadow image \mathbf{I}_s onto the log-chromaticity space, calculate the entropy, and use the entropy minimization to find the projection direction θ , which is specific to \mathbf{I}_s . From θ , we can obtain a shadow-free chromaticity map σ_{sf}^{ent} that no longer contains any shadows (see Figs. 3 and 4b). However, owing to the projection, there is a color shift present in σ_{sf}^{ent} , which can be corrected by using the illumination compensation procedure.

Illumination Compensation To correct the color of the shadow-free chromaticity map σ_{sf}^{ent} , following [3], we add

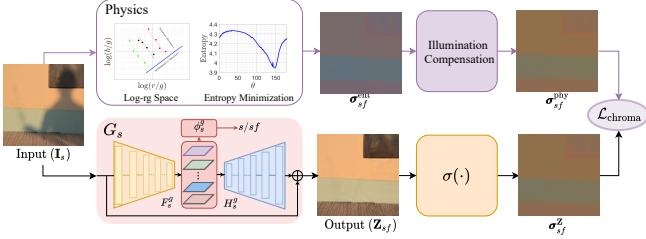


Figure 3. Shadow-Free Chromaticity Loss. The upper part is the physics-based pipeline where we use entropy minimization followed by illumination compensation to generate the shadow-free chromaticity image σ_{sf}^{phy} from the input image I_s . The lower part shows our shadow removal generator G_s guided by σ_{sf}^{phy} through our shadow-free chromaticity loss \mathcal{L}_{chroma} .

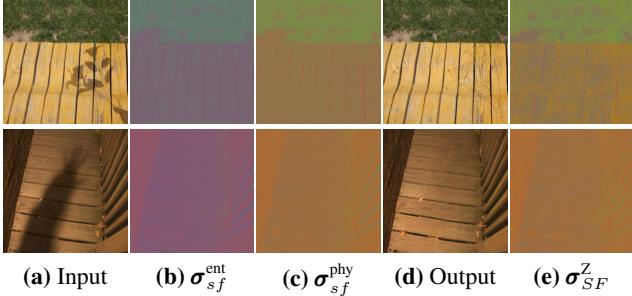


Figure 4. (a) Input shadow image I_s , (b) Shadow-free chromaticity after entropy minimization σ_{sf}^{ent} , (c) Shadow-free chromaticity after illumination compensation σ_{sf}^{phy} , (d) Output shadow-free image Z_{sf} , and (e) Chromaticity map of output image σ_{sf}^Z . Our shadow-free chromaticity loss constrains (e) to be similar to (c) facilitating better shadow removal.

back the original illumination color of the non-shadow regions to the map. For this, we use uniformly sampled 30% of the brightest pixels from the input image I_s based on the assumption that these pixels are located in the non-shadow regions of I_s . Once we reinstate the illumination color, we can obtain a new shadow-free chromaticity map σ_{sf}^{phy} , (see Figs. 3 and 4c).

Having obtained our shadow-free chromaticity, σ_{sf}^{phy} , for the output shadow-free image Z_{sf} , we compute its chromaticity map σ_{sf}^Z by:

$$\sigma_{sf_c}^Z = \frac{Z_{sf_c}}{(Z_{sf_r} + Z_{sf_g} + Z_{sf_b})}, \quad (1)$$

where $c \in \{r, g, b\}$ represents a color channel, $Z_{sf} = [Z_{sf_r}, Z_{sf_g}, Z_{sf_b}]$, and $\sigma_{sf}^Z = [\sigma_{sf_r}^Z, \sigma_{sf_g}^Z, \sigma_{sf_b}^Z]$. We can now define our shadow-free chromaticity loss as:

$$\mathcal{L}_{chroma}(G_s) = \mathbb{E}_{I_s} [\|\sigma_{sf}^Z - \sigma_{sf}^{phy}\|_1]. \quad (2)$$

Using the loss function expressed in Eq. (2), we enforce the chromaticity of the output shadow-free image, σ_{sf}^Z , to be the same as our physics-based shadow-free chromaticity

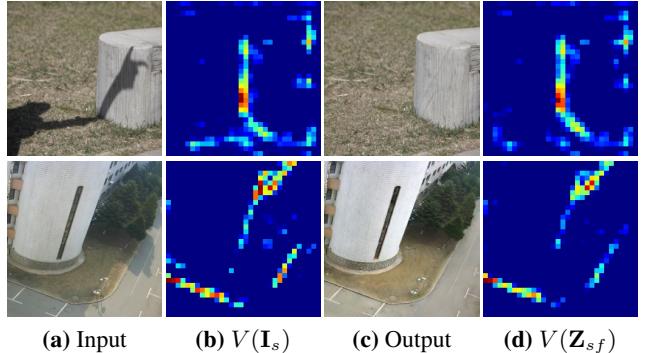


Figure 5. (a) Input shadow image I_s , (b) Sample feature map for I_s , (c) Output shadow-free image Z_{sf} , and (d) Sample feature map for Z_{sf} . We can observe that features in (b) for the input shadow images are less affected by shadows, and they are similar to the features in (d) owing to our shadow-robust feature loss.

σ_{sf}^{phy} , which can be observed in the results shown in Fig. 4 for both hard shadow and soft shadow images¹.

3.2. Shadow-Robust Feature Loss

Our shadow-robust feature loss is based on the perceptual features obtained from the pre-trained VGG-16 network [15, 26]. Since we do not have ground truth to obtain the correct shadow-free features, to guide the shadow-free output, we use features from the input shadow image itself. Our underlying idea is that, since with some degree of shadows and lighting conditions, object classification using the pre-trained VGG-16 is known to be robust [28], there should be some features in the pre-trained VGG-16 that are less affected by shadows. Based on this, we perform a calibration experiment and find that the Conv22 layer in the VGG-16 network provides features that are least affected by shadows.

Hence, from the input shadow image, we obtain the shadow-robust features and use them to guide our shadow-free output image. Specifically, given an input shadow image I_s and the corresponding shadow-free output image Z_{sf} , we define our shadow-robust feature loss as:

$$\mathcal{L}_{feature}(G_s) = \mathbb{E}_{I_s} [\|V(Z_{sf}) - V(I_s)\|_1], \quad (3)$$

where $V(I_s)$ and $V(Z_{sf})$ denote the feature maps extracted from the Conv22 layer of the pre-trained VGG-16 network for I_s and Z_{sf} respectively. Fig. 5 shows some examples where we can observe that the features $V(I_s)$ are less affected by shadows and represent more structural information (like edges).

¹For surfaces that are close to being achromatic, the entropy minimization can fail, which can lead to the improper recovery of the shadow-free chromaticity map. However, due to the presence of our other unsupervised losses, our method can still generate proper shadow removal results.

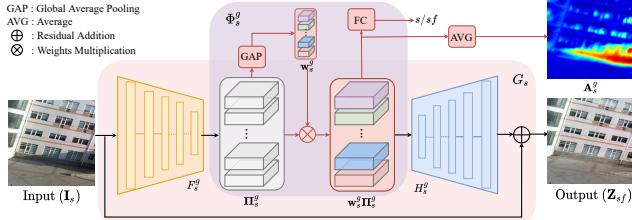


Figure 6. Domain Classification and Shadow Attention. In the generator G_s , its encoder F_s^g extracts feature maps Π_s^g from the input shadow image I_s . As in [33], using global average pooling (GAP), the domain classifier Φ_s^g is trained to learn the weights w_s^g of the feature maps. Averaging the weighted feature maps generates an attention map A_s^g , i.e. $A_s^g = \frac{1}{n} \sum_{i=1}^n w_s^g_i \Pi_s^g$ (n being the total number of feature maps), which clearly shows that the network is focusing on shadow regions.

3.3. Domain Classification Loss

We incorporate an attention mechanism that allows our DC-ShadowNet to know the shadow removal/restoration regions [33, 23, 18]. To achieve this, we create a domain classifier Φ_s^g and integrate it with the generator G_s . We train Φ_s^g to classify whether the input to G_s is from the shadow or shadow-free domain. Fig. 6 shows the integration of Φ_s^g into G_s to obtain an attention map A_s^g that highlights shadow regions. We also add a similar domain classifier Φ_{sf}^d to the discriminator D_{sf} . This allows our network to selectively focus on important shadow regions and generate better shadow removal results (see Fig. 7).

Since the generator can accept either a shadow or shadow-free image as input, it allows us to train it together with its domain classifier. However, for the discriminator, the domain of its input image, which is the output of the generator, can be ambiguous². For this reason, we pre-train the domain classifier of the discriminator using the following classification loss:

$$\mathcal{L}_{\text{domcls}}(D_{sf}) = \mathbb{E}_{I_s} \left[-\log (\Phi_{sf}^d(F_{sf}^d(I_s))) \right] + \mathbb{E}_{I_{sf}} \left[-\log (1 - \Phi_{sf}^d(F_{sf}^d(I_{sf}))) \right], \quad (4)$$

and after pre-training, we freeze its weights during the main training cycle that trains our entire network (see Fig. 2). To train the domain classifier of the generator, we use a similar classification loss:

$$\mathcal{L}_{\text{domcls}}(G_s) = \mathbb{E}_{I_s} \left[-\log (\Phi_s^g(F_s^g(I_s))) \right] + \mathbb{E}_{I_{sf}} \left[-\log (1 - \Phi_s^g(F_s^g(I_{sf}))) \right]. \quad (5)$$

²In the early stage of training, shadow removal can be improper, and the output of the generator can still have shadows. Hence, it is difficult to ensure that the domain of the output is always shadow-free.

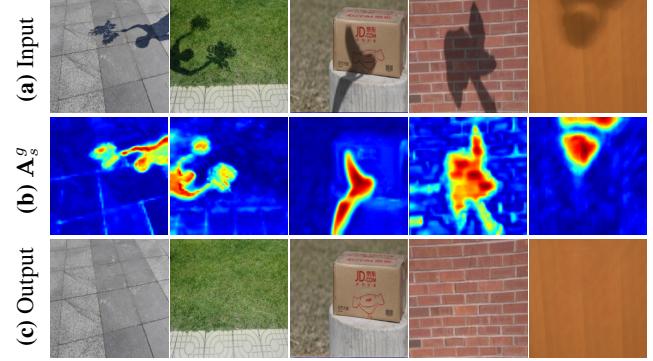


Figure 7. (a) Input shadow image I_s , (b) Attention map A_s^g , and (c) Output shadow-free image Z_{sf} . The attention maps clearly indicate the shadow regions of the input shadow images.

3.4. Boundary Smoothness Loss

To ensure that the output shadow-free image Z_{sf} have smoother transitions in the boundaries defined by the shadow regions of the input shadow image I_s , we also use a boundary smoothness loss:

$$\mathcal{L}_{\text{smooth}}(G_s) = \mathbb{E}_{I_s} \left[\|B(M_s) * |\nabla(Z_{sf})||_1 \right], \quad (6)$$

where ∇ is the gradient operation, B is a noise-robust function [29, 25, 31] to compute the boundaries of the shadow regions from our shadow mask M_s . To obtain M_s , we compute the difference between the input shadow image I_s and output shadow-free image Z_{sf} , and apply our mask detection function F on the difference:

$$M_s = F(I_{sc} - Z_{sf_c}) = \sum_{c \in \{r, g, b\}} \frac{1}{3} \left| (N(I_{sc} - Z_{sf_c})) \right|, \quad (7)$$

where the function N is a normalization function defined as $N(\mathbf{I}) = (\mathbf{I} - \mathbf{I}_{\min}) / (\mathbf{I}_{\max} - \mathbf{I}_{\min})$, where \mathbf{I}_{\max} and \mathbf{I}_{\min} are the maximum and minimum values of \mathbf{I} , respectively. Note that, our shadow mask M_s is a soft map and have the values in the range of $[0, 1]$. See Fig. 8b for some examples.

The noise-robust function B is defined as: $B(M_s) = B_{sx} + B_{sy}$ where $B_{sx}(p) = \left| \sum_{q \in R_p} g_{p,q} \partial_x(M_s(q)) \right|$ and $B_{sy}(p) = \left| \sum_{q \in R_p} g_{p,q} \partial_y(M_s(q)) \right|$, ∂_x and ∂_y are partial derivatives in horizontal and vertical directions respectively, p defines a pixel, R_p is a 3×3 window around p , and $g_{p,q}$ is a weighing function measuring spatial affinity defined as $g_{p,q} = \exp \left(\frac{-(p-q)^2}{2\tau^2} \right)$, where τ is set to 0.01 by default. See Fig. 8(c) for some examples of our soft boundary detection.

3.5. Adversarial, Consistency and Identity Losses

For shadow removal, we use the generator G_s , which is coupled with a discriminator D_{sf} . To ensure reconstruction consistency, we use another generator G_{sf} coupled with its

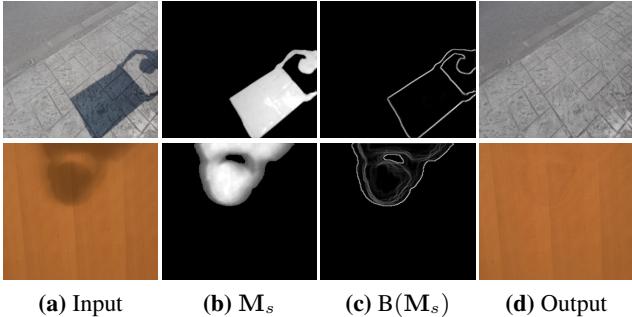


Figure 8. (a) Input shadow image, (b) Soft shadow mask M_s , (c) Detected shadow boundaries, and (d) Our output shadow-free results. We can observe that our boundary smoothness loss helps in having smoother outputs in the shadow boundary regions.

own discriminator D_s . We use adversarial losses to train our DC-ShadowNet:

$$\mathcal{L}_{\text{adv}}(G_s, D_{sf}) = \mathbb{E}_{\mathbf{I}_{sf}} [\log(D_{sf}(\mathbf{I}_{sf}))] + \mathbb{E}_{\mathbf{I}_s} [\log(1 - D_{sf}(G_s(\mathbf{I}_s)))], \quad (8)$$

$$\mathcal{L}_{\text{adv}}(G_{sf}, D_s) = \mathbb{E}_{\mathbf{I}_s} [\log(D_s(\mathbf{I}_s))] + \mathbb{E}_{\mathbf{I}_{sf}} [\log(1 - D_s(G_{sf}(\mathbf{I}_{sf}, M_s)))]. \quad (9)$$

During training, the losses expressed in Eqs. (8) and (9) are actually minimized as $\min_{G_s} \max_{D_{sf}} (\mathcal{L}_{\text{adv}}(G_s, D_{sf}))$ and $\min_{G_{sf}} \max_{D_s} (\mathcal{L}_{\text{adv}}(G_{sf}, D_s))$ respectively. Note that, unlike generator G_s , the generator G_{sf} takes the mask M_s (from Eq. 7) as input to help render more proper shadow images [13]. Following [34, 30], we define our reconstruction consistency losses by:

$$\mathcal{L}_{\text{cons}}(G_s) = \mathbb{E}_{\mathbf{I}_s} [| | G_{sf}(G_s(\mathbf{I}_s), M_s) - \mathbf{I}_s | |_1], \quad (10)$$

$$\mathcal{L}_{\text{cons}}(G_{sf}) = \mathbb{E}_{\mathbf{I}_{sf}} [| | G_s(G_{sf}(\mathbf{I}_{sf}, M_s)) - \mathbf{I}_{sf} | |_1]. \quad (11)$$

While our G_s is designed to remove shadows from shadow input image \mathbf{I}_s , we also encourage it to output the same image as input, if the input is a shadow-free image \mathbf{I}_{sf} . We achieve this by using the following identity losses [34]:

$$\mathcal{L}_{\text{idem}}(G_s) = \mathbb{E}_{\mathbf{I}_{sf}} [| | (G_s(\mathbf{I}_{sf})) - \mathbf{I}_{sf} | |_1], \quad (12)$$

$$\mathcal{L}_{\text{idem}}(G_{sf}) = \mathbb{E}_{\mathbf{I}_s} [| | (G_{sf}(\mathbf{I}_s, M_0)) - \mathbf{I}_s | |_1]. \quad (13)$$

where M_0 represents a mask with all zero values.

Overall Loss We multiply each loss function with its respective weight, and sum them together to obtain our overall loss function. The weights of the losses, $\{\mathcal{L}_{\text{chroma}}, \mathcal{L}_{\text{feature}}, \mathcal{L}_{\text{smooth}}, \mathcal{L}_{\text{domcls}}, \mathcal{L}_{\text{adv}}, \mathcal{L}_{\text{cons}}, \mathcal{L}_{\text{idem}}\}$, are represented by $\{\lambda_{\text{chroma}}, \lambda_{\text{feat}}, \lambda_{\text{sm}}, \lambda_{\text{dom}}, \lambda_{\text{adv}}, \lambda_{\text{cons}}, \lambda_{\text{idem}}\}$.

4. Experiments

To evaluate our method, we use five datasets: SRD [24], adjusted ISTD (AISTD) [20], ISTD [27], USR [13] and

Table 1. RMSE results on the SRD dataset. All, S and NS represent entire, shadow and non-shadow regions respectively.

Method	Training	All	S	NS
Our DC-ShadowNet	Unpaired	4.66	7.70	3.39
Mask-ShadowGAN [13]	Unpaired	6.40	11.46	4.29
DSC [14]	Paired	4.86	8.81	3.23
DeShadowNet [24]	Paired	5.11	3.57	8.82
Gong <i>et al.</i> [8]	-	12.35	25.43	6.91
Input Image	-	13.77	37.40	3.96

Table 2. RMSE results on the AISTD dataset. All, S and NS represent entire, shadow and non-shadow regions respectively. M shows that ground truth shadow masks are also used in training.

Method	Training	All	S	NS
Our DC-ShadowNet	Unpaired	4.6	10.3	3.5
Mask-ShadowGAN [13]	Unpaired	5.3	12.5	4.0
DeshadowNet [24]	Paired	7.6	15.9	6.0
ST-CGAN [27]	Paired+M	8.7	13.4	7.7
Gong <i>et al.</i> [8]	-	-	13.3	-
Guo <i>et al.</i> [10]	Paired+M	6.1	22.0	3.1
Yang <i>et al.</i> [32]	-	16.0	24.7	14.4
Input Image	-	8.5	40.2	2.6

LRSS [9], where LRSS is a soft shadow dataset. To ensure fair comparisons, all the unsupervised baselines, including ours are trained and tested on the same datasets. For the SRD dataset, for Table 1 and Fig. 9 rows 2-4, we use 2680 shadow images and 2680 shadow-free images for training. We use 408 shadow images that have shadow-free ground truth for testing. Similarly, for Table 2, we use 1330 training and 540 testing AISTD images; Fig. 9 row 1, we use 1330 training and 540 testing ISTD images. For the USR dataset, we use 1956 shadow, 1770 shadow-free images for training, 489 shadow images for testing. However, for testing, the USR dataset does not provide paired shadow and shadow-free images.

Our DC-ShadowNet is trained in an unsupervised manner (Sec. 3). The weights of our losses $\{\lambda_{\text{chroma}}, \lambda_{\text{feat}}, \lambda_{\text{sm}}, \lambda_{\text{idem}}, \lambda_{\text{adv}}, \lambda_{\text{cons}}, \lambda_{\text{dom}}\}$ are set empirically to $\{1, 1, 1, 10, 1, 10, 1\}$. Following the baselines [11, 13], to evaluate shadow removal performance³, we use root mean squared error (RMSE) between the ground truth and the predicted shadow-free image⁴. Hence, lower numbers show better performance.

³Results of [13, 14, 27, 8, 20, 1] are taken from their official implementations. Results of [9, 11] are obtained from their project website: <http://visual.cs.ucl.ac.uk/pubs/softshadows/>. The quantitative results are taken from the paper [21].

⁴As mentioned in [22], the default RMSE evaluation code used by all methods (including ours) actually computes mean absolute error (MAE).

Table 3. RMSE (lower is better) and PSNR (higher is better) results on the LRSS dataset (soft shadow dataset). M and S respectively show that ground truth shadow masks and synthetic paired data are used in training. P and UP denote paired and unpaired training, respectively.

Method	Input	[11]	[11] (auto)	[9]	[1]	[20]	[13]	Ours
RMSE	12.26	6.02	5.87	4.38	7.92	7.48	7.13	3.48
PSNR	18.05	27.88	28.02	29.25	25.57	23.93	25.12	31.01
Training	-	P+M	P	P+M+S	P+M+S	P+M	UP	UP

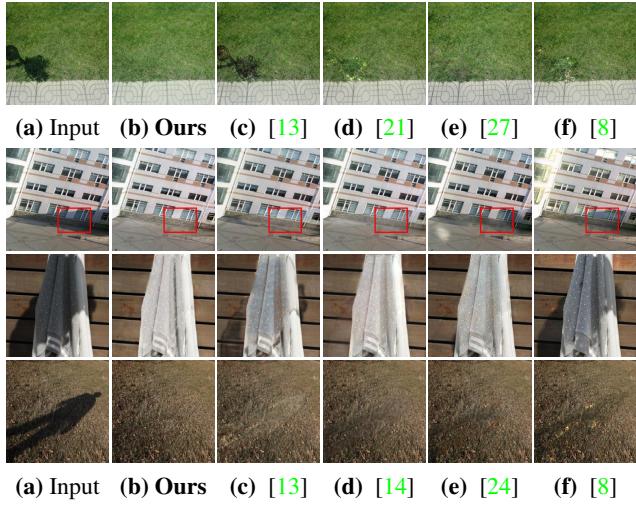


Figure 9. Comparison results on the ISTD (top row) and SRD (bottom three rows) datasets. (a) Input image, (b) Our method, unsupervised method (c) Mask-ShadowGAN [13], weakly-supervised method (d) Param+M+D-Net [21] (top row), supervised methods DSC [14], (e) ST-CGAN [27] (top row), DeshadowNet [24], and traditional method (f) Gong *et al.* [8]. Our method trained using unsupervised learning provides the best performance.

Results on Hard Shadows We conduct quantitative evaluations on the SRD and AISTD datasets, and the corresponding results are shown in Table 1 and Table 2, respectively. For comparisons, we use the state-of-the-art unsupervised shadow removal method Mask-ShadowGAN [13], weakly-supervised method Param+M+D-Net [21], supervised methods DSC [14], DeshadowNet [24], ST-CGAN [27], and traditional methods Gong *et al.* [8], Guo *et al.* [11], and Yang *et al.* [32]. From Tables 1 and 2, our DC-ShadowNet trained in an unsupervised manner achieves the best performance compared to the baseline methods. Compared to the state-of-the-art unsupervised method Mask-ShadowGAN [13], our results for the shadow regions are better by $\sim 33\%$ and $\sim 18\%$ on the SRD and AISTD datasets, respectively.

The qualitative results for the SRD (rows 2-4) and ISTD (top row) datasets are shown in Fig. 9, which include challenging conditions and diverse objects. For example, the shadow image contains shadows casted on semantic objects (i.e., building, wall). In Fig. 9, the method [13] alters the

Table 4. Ablation experiments of our method using the SRD dataset. All, S and NS represent entire, shadow and non-shadow regions, respectively. The numbers represent RMSE.

Method	All	S	NS
Our DC-ShadowNet	4.66	7.70	3.39
w/o \mathcal{L}_{smooth}	4.72	7.80	3.43
w/o $\mathcal{L}_{feature}$	4.83	8.04	3.50
w/o \mathcal{L}_{chroma}	5.05	8.50	3.61
w/o Φ_s^g	5.20	8.94	3.65
w/o Φ_{sf}^d	5.49	9.42	3.87
w/o Φ_s^g and Φ_{sf}^d	8.12	16.10	4.80
Input Image	13.77	37.40	3.96

colors of the non-shadow regions and cannot properly handle shadow boundaries. For the method [8], the recovery of the shadow-free images is unsatisfactory. In comparison, our DC-ShadowNet performs better, showing the effectiveness of our domain classification network and our novel unsupervised losses.

Results on Soft Shadows The LRSS dataset has 134 shadow images, mainly contains soft-shadow images. We pre-trained our DC-ShadowNet on the SRD training set, then we use 100 LRSS images for training it in an unsupervised manner. The remaining 34 LRSS images with their corresponding shadow-free images are used for testing. The quantitative results are shown in Table 3. We compare our DC-ShadowNet with the following methods: unsupervised method Mask-ShadowGAN [13], supervised methods SP+M-Net [20] and DHAN [1], automatic method Guo [11], and interactive method [9] which requires user-annotations of shadow regions. As shown in Table 3, our method achieves the lowest RMSE and highest PSNR.

The qualitative results covering a diverse set of images such as indoor/outdoor scenes, shadow regions, etc., are shown in Fig. 10. While the state-of-the-art methods can remove shadows to some extent, the results are still improper. Mask-ShadowGAN [13] fails to handle soft-shadows since it uses binary masks to represent shadow regions. Moreover, it mainly relies on adversarial training that cannot guarantee proper shadow removal. Supervised methods like DHAN [1] and SP+M-Net [20] have artifacts in the shadow regions as they suffer from the domain gap problem. Guo [11] fails due to the difficulty in automatically identifying soft shadow regions. Compared to all the baseline methods, our results are more proper, and the image surfaces are better-restored.

Test-Time Training We show that our method being unsupervised can be used for test-time training to further improve the results on the test images. For this, we use the 34 shadow images from the test set used in the soft shadow

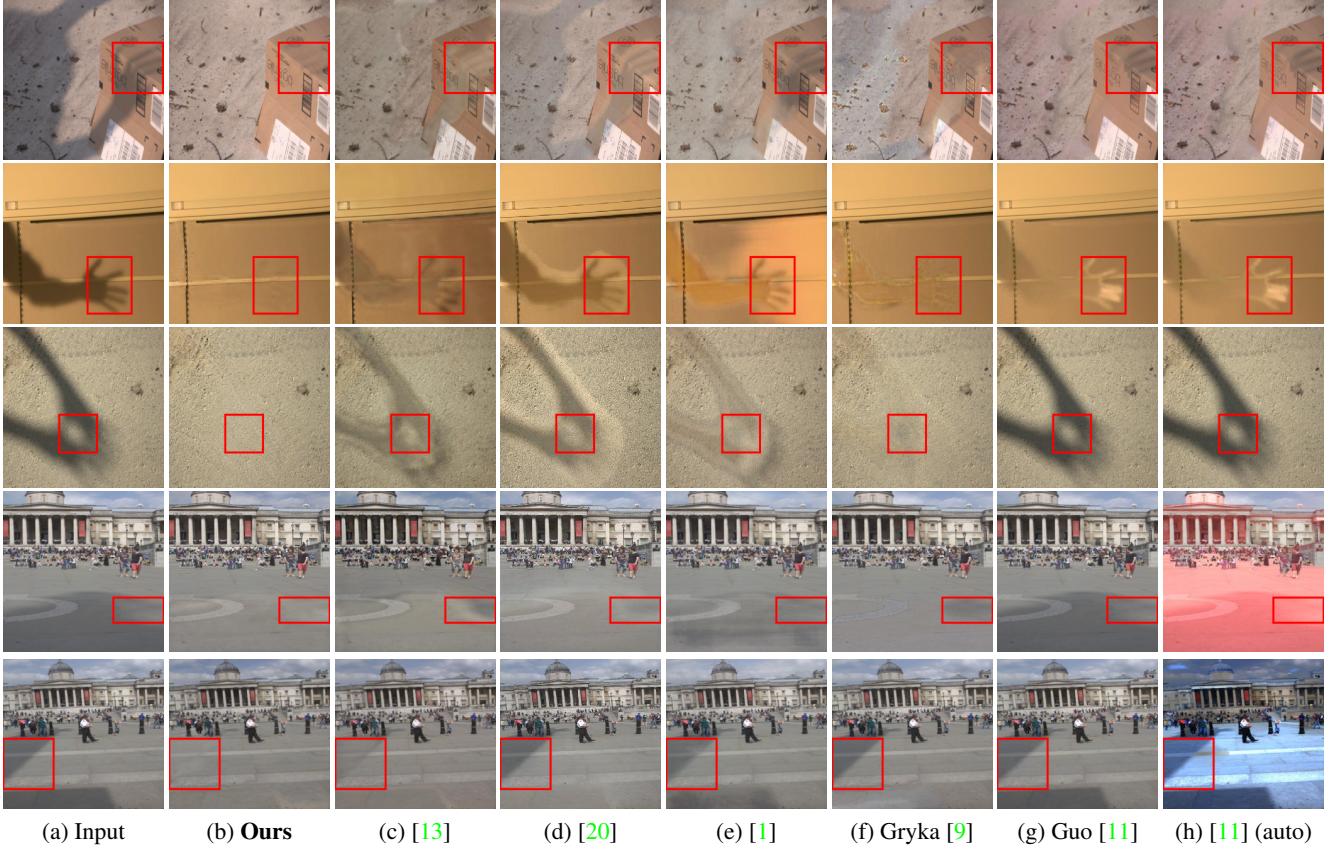


Figure 10. Comparison results on the soft shadow LRSS dataset (a) Input image, (b) Our result, (c) Unsupervised method Mask-ShadowGAN [13], Supervised methods (d) SP+M-Net [20] and (e) DHAN [1]. (f)~(h) are the results of the traditional methods (auto means automatic detection). Our method, trained using unsupervised learning, generates better shadow-free results.



Figure 11. (a) Input image, (b) and (c) show our results without and with test-time-training, (d) Result of Mask-ShadowGAN [13].

evaluation above, and employ our unsupervised losses to train our method. To evaluate shadow removal performance, we use the corresponding shadow-free images; and the performance in terms of RMSE and PSNR improves from 3.48 and 31.01 to 3.36 and 31.31, respectively. See Fig. 11 for a qualitative example showing the effectiveness of test-time training.

5. Ablation Study

We conduct ablation studies to analyze the effectiveness of different components of our method such as 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 classifier Φ_s^g and Φ_{sf}^d . We use the SRD dataset for our experiments and the corresponding quantitative results are shown in Table 4. Each component of our method is important and contributes to the better performance of our method.

6. Conclusion

We have proposed DC-ShadowNet, an unsupervised learning-based shadow removal method guided by domain classification network, shadow-free chromaticity, shadow-robust feature and boundary smoothness losses. Our method can robustly handle both hard and soft shadow images. We integrate a domain classifier with our generator and its corresponding discriminator, enabling our method to focus on shadow regions. To train DC-ShadowNet, we use novel unsupervised losses that enable it to directly learn from unlabeled (no ground truth) real shadow images. We also showed that we could employ test-time refinement that can further improve our performance. Experimental results have confirmed that our method is effective and outperforms the state-of-the-art shadow removal methods.

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. 1, 6, 7, 8
- [2] Bin Ding, Chengjiang Long, Ling Zhang, and Chunxia Xiao. Argan: Attentive recurrent generative adversarial network for shadow detection and removal. In *Proceedings of the IEEE international conference on computer vision*, pages 10213–10222, 2019. 2
- [3] Mark S Drew, Graham D Finlayson, and Steven D Hordley. Recovery of chromaticity image free from shadows via illumination invariance. In *IEEE Workshop on Color and Photometric Methods in Computer Vision, ICCV'03*, pages 32–39, 2003. 2, 3
- [4] Graham D Finlayson and Mark S Drew. 4-sensor camera calibration for image representation invariant to shading, shadows, lighting, and specularities. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 473–480, 2001. 2
- [5] Graham D Finlayson, Mark S Drew, and Cheng Lu. Intrinsic images by entropy minimization. In *European conference on computer vision*, pages 582–595, 2004. 2
- [6] Graham D Finlayson, Mark S Drew, and Cheng Lu. Entropy minimization for shadow removal. *International Journal of Computer Vision*, 85(1):35–57, 2009. 1, 2, 3
- [7] 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. 1, 2
- [8] Han Gong and Darren Cosker. Interactive shadow removal and ground truth for variable scene categories. In *BMVC*, pages 1–11, 2014. 6, 7
- [9] Maciej Gryka, Michael Terry, and Gabriel J Brostow. Learning to remove soft shadows. *ACM Transactions on Graphics (TOG)*, 34(5):1–15, 2015. 2, 6, 7, 8
- [10] Ruiqi Guo, Qieyun Dai, and Derek Hoiem. Single-image shadow detection and removal using paired regions. In *CVPR 2011*, pages 2033–2040, 2011. 1, 2, 6
- [11] 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. 2, 6, 7, 8
- [12] Xiaowei Hu, Chi-Wing Fu, Lei Zhu, Jing Qin, and Pheng-Ann Heng. Direction-aware spatial context features for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(11):2795–2808, 2019. 2
- [13] 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, 6, 7, 8
- [14] 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, 6, 7
- [15] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016. 2, 4
- [16] Rei Kawakami, Katsushi Ikeuchi, and Robby T Tan. Consistent surface color for texturing large objects in outdoor scenes. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 2, pages 1200–1207. IEEE, 2005. 1
- [17] Salman H Khan, Mohammed Bennamoun, Ferdous Sohel, and Roberto Togneri. Automatic shadow detection and removal from a single image. *IEEE transactions on pattern analysis and machine intelligence*, 38(3):431–446, 2015. 1
- [18] Junho Kim, Minjae Kim, Hyewon Kang, and Kwang Hee Lee. U-gat-it: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. In *International Conference on Learning Representations*, 2020. 5
- [19] Jean-François Lalonde, Alexei A Efros, and Srinivasa G Narasimhan. Estimating natural illumination from a single outdoor image. In *2009 IEEE 12th International Conference on Computer Vision*, pages 183–190, 2009. 1
- [20] 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, 2, 6, 7, 8
- [21] Hieu Le and Dimitris Samaras. From shadow segmentation to shadow removal. In *European Conference on Computer Vision*, pages 264–281. Springer, 2020. 1, 2, 6, 7
- [22] Official Github page of 'Shadow Removal via Shadow Image Decomposition' ICCV 2019. <https://github.com/cvlab-stonybrook/SID>. 6
- [23] Rui Qian, Robby T Tan, Wenhan Yang, Jiajun Su, and Jiaying Liu. Attentive generative adversarial network for raindrop removal from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2482–2491, 2018. 5
- [24] 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, 6, 7
- [25] Aashish Sharma and Loong-Fah Cheong. Into the twilight zone: Depth estimation using joint structure-stereo optimization. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 103–118, 2018. 5
- [26] Aashish Sharma, Loong-Fah Cheong, Lionel Heng, and Robby T Tan. Nighttime stereo depth estimation using joint translation-stereo learning: Light effects and uninformative regions. In *2020 International Conference on 3D Vision (3DV)*, pages 23–31. IEEE, 2020. 4
- [27] 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, 2, 6, 7

- [28] Brandon Richard Webster, Samuel E Anthony, and Walter J Scheirer. Psyphy: A psychophysics driven evaluation framework for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 41(9):2280–2286, 2018.
- [29] Li Xu, Qiong Yan, Yang Xia, and Jiaya Jia. Structure extraction from texture via relative total variation. *ACM transactions on graphics (TOG)*, 31(6):1–10, 2012.
- [30] Wending Yan, Aashish Sharma, and Robby T Tan. Optical flow in dense foggy scenes using semi-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13259–13268, 2020.
- [31] Wending Yan, Robby T Tan, and Dengxin Dai. Nighttime defogging using high-low frequency decomposition and grayscale-color networks. In *European Conference on Computer Vision*, pages 473–488. Springer, 2020.
- [32] Qingxiong Yang, Kar-Han Tan, and Narendra Ahuja. Shadow removal using bilateral filtering. *IEEE Transactions on Image processing*, 21(10):4361–4368, 2012.
- [33] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2921–2929, 2016.
- [34] 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.

4

5

6

5

6, 7

6

6, 7

1, 3, 6