# From Shadow Segmentation to Shadow Removal

Hieu Le and Dimitris Samaras

Stony Brook University, Stony Brook, NY 11794, USA

**Abstract.** The requirement for paired shadow and shadow-free images limits the size and diversity of shadow removal datasets and hinders the possibility of training large-scale, robust shadow removal algorithms. We propose a shadow removal method that can be trained using only shadow and non-shadow patches cropped from the shadow images themselves. Our method is trained via an adversarial framework, following a physical model of shadow formation. Our central contribution is a set of physics-based constraints that enables this adversarial training. Our method achieves competitive shadow removal results compared to state-of-the-art methods that are trained with fully paired shadow and shadow-free images. The advantages of our training regime are even more pronounced in shadow removal for videos. Our method can be fine-tuned on a testing video with only the shadow masks generated by a pre-trained shadow detector and outperforms state-of-the-art methods on this challenging test. We illustrate the advantages of our method on our proposed video shadow removal dataset.

**Keywords:** Shadow Removal, GAN, Weakly-supervised, Illumination model, Unpaired, Image-to-Image.

## 1 Introduction

Shadows are present in most natural images. Shadow effects make objects harder to detect or segment [23], and scenes with shadows are harder to process and analyze [20]. Realistic shadow removal is an integral part of image editing [3] and can greatly improve performance on various computer vision tasks [32,41,56,24,21], getting increased attention in recent years [37,13,11]. Data-driven approaches using deep learning models have achieved remarkable performance on shadow removal [5,22,17,15,47,55] thanks to recent large-scale datasets [45,47].

Most of the current deep-learning shadow removal approaches are end-to-end mapping functions trained in a fully supervised manner. Such systems require pairs of shadow images and their shadow-free counter-parts as training signals. However, this type of data is cumbersome to obtain, lacks diversity, and is error-prone: all current shadow removal datasets exhibit color mismatches between the shadow images and their shadow-free ground truth (see Fig. 1 - left panel). Moreover, there are no images with self-cast shadows because the occluders are never visible in the image in the current data acquisition setups [47,37,15]. This dependency on paired data significantly hinders building large-scale, robust shadow-removal systems. A recent method trying to overcome this issue

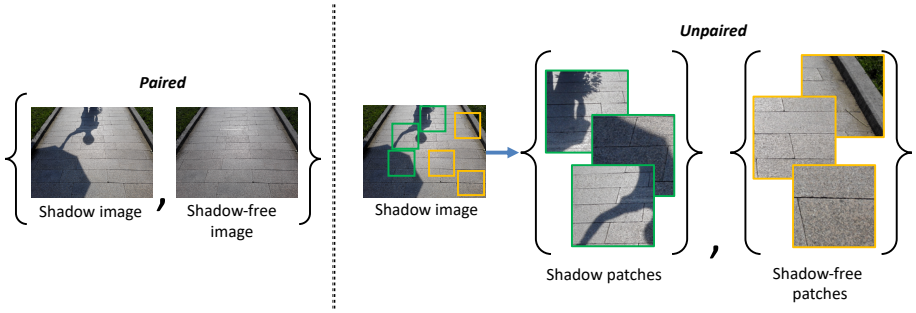


Fig. 1: Paired training data (left) consists of training examples {shadow, shadow-free} images which are expensive to collect, lack diversity, and are sensitive to errors due to possible color mismatches between the two images. Note the slightly different color tone between the two images. In this paper, we propose to learn shadow removal from unpaired shadow and non-shadow patches cropped from the same shadow image (right). This eliminates the need for shadow free images.

is MaskShadow-GAN [15], which learns shadow removal from unpaired shadow and shadow-free images. However, such cycle-GAN [58] based systems usually require enough statistical similarity between the two sets of images [25,2]. This requirement can be hard to satisfy when capturing shadow-free images is tricky, such as shadow-free images of urban areas [4] or moving objects [18,36].

In this paper, we propose an alternative solution to the data dependency issue. We first observe that image patches alongside the shadow boundary contain critical information for shadow removal, including non-shadow, umbra and penumbra areas. They sufficiently reflect the characteristics of the shadowing effects, including the color differences between shadow and non-shadow areas as well as the gradual changes of the shadow effects across the shadow boundary [34,33,14]. If we further assume that the shadow effects are fairly consistent in the umbra areas, a patch-based shadow removal can be used to remove shadows in the whole image. Based on this observation, we propose training a patch-based shadow removal system for which we use unpaired shadow and non-shadow patches directly cropped from the shadow images themselves as training data. This approach eliminates the dependency on paired training data and opens up the possibility of handling different types of shadows, since it can be trained with any kind of shadow image. Compared to MaskShadow-GAN, shadow and non-shadow patches cropped from the same image naturally ensure significant statistical similarity. The only supervision required in this data processing scheme are the shadow masks, which are relatively easy to obtain, either manually, semi-interactively [45,11], or automatically using shadow detection methods [5,59,57,23]. Automatic shadow detection is improving, with the main challenge being generalization across datasets. At some point, one can expect to get very good shadow masks automatically, which would allows training our shadow removal method with very little annotation cost.

In particular, to obtain shadow and shadow-free patches, we crop the shadow images into small overlapping patches of size $n \times n$ with a step size of $m$. Based

on the shadow masks, we group these patches into three sets: a non-shadow set ($\mathcal{N}$) containing patches having no shadow pixels, a shadow-boundary set ($\mathcal{B}$) containing patches lying on the shadow boundaries, and a full-shadow set ($\mathcal{F}$) containing patches where all pixels are in shadow. With small enough patch size $n$ and step size $m$, we can obtain enough training patches in each set. With this training set, we train a shadow removal system to learn a mapping from patches in the shadow-boundary set $\mathcal{B}$ to patches in the non-shadow set $\mathcal{N}$. Essentially, this mapping needs to infer the color difference alongside the shadow edges, including the chromatic attributes of the light source and the smooth change of the shadow effects across the shadow boundary, in order to transform a shadow patch to a non-shadow patch. This is, in spirit, similar to early shadow removal approaches that focus on shadow edges to remove shadows [38,9,8,44,46].

By simply cropping shadow images into patches, we are posing the shadow removal as an unpaired image-to-image cross-domain mapping [54,2,29] that can be estimated via an adversarial framework. In particular, we seek a mapping function $G$ that takes as input a shadow-boundary patch $x$ from the set $\mathcal{B}$, and outputs an image patch $\hat{x}$, such that a critic function $D$ cannot distinguish whether $\hat{x}$ was drawn from the non-shadow set $\mathcal{N}$ or generated by $G$. Note that one potential solution here is to use Cycle-GAN or MaskShadow-GAN to estimate this transformation. However, the mapping functions learned by these methods are not able to remove shadows from patches in the full-shadow set $\mathcal{F}$.

Training such an unpaired image-to-image mapping for shadow removal is challenging. The mapping is under-constrained and training can collapse easily. [12,28,27,30,42,31]. Here, we propose to systematically constrain the shadow removal process by a physical model of shadow formation [39] and incorporate a number of physical properties of shadows into the framework. We show that these physics-based priors define a transformation closely modelling shadow removal. Driven by an adversarial signal, our framework effectively learns physically-plausible shadow removal without any direct supervision from paired data. Specifically, we constrain the shadow removal process to a shadow image decomposition model [22] that extracts a set of parameters and a matting layer from the shadow image. This set of shadow parameters is responsible for removing shadows on the umbra areas of the shadows via a linear function. Thus, once we estimate these shadow parameters from shadow boundary patches, we can use them to remove shadows for patches fully covered by the same shadow under the assumption that they share the same set of shadow parameters. Based on the physical properties of shadows, we apply the following constraints to the model:

- We limit the search space of the shadow parameters and shadow matte to the appropriate value ranges that correspond to shadow removal.
- Our matting and smoothness losses ensure that shadow removal only happens in the shadow areas and transitions smoothly across shadow boundaries.
- Our boundary loss on the generated shadow-free image enforces color similarity between the inner and outer areas alongside shadow boundaries.

With these constraints and the adversarial signal, our method achieves shadow removal results that are competitive with state-of-the-art methods that were

trained in a fully supervised manner with paired shadow and non-shadow images [22,47,37]. We further compare our method to state-of-the-art methods on a novel and challenging video shadow removal dataset including static videos with various scenes and shadow conditions. This test exposes the weaknesses of data-driven methods trained on datasets lacking diversity. Our patch-based method seems to generalize better than other methods when evaluated on this video shadow removal test. Most importantly, we can easily fine-tune our pre-trained model on a single testing video to further improve shadow removal results, showcasing this advantage of our training scheme.

In short, our contributions are:

- We propose the use of an adversarial critic to train a shadow remover from unpaired shadow and non-shadow patches, providing an alternative solution to the paired data dependency issue.
- We propose a set of physics-based constraints that define a transformation closely modelling shadow removal, which enables shadow remover training with only an adversarial training signal.
- Our system trained without any shadow-free images has competitive results compared to fully-supervised state-of-the-art methods on the ISTD dataset.
- We collect a novel video shadow removal dataset. Our shadow removal system can be fine-tuned for free to better remove shadows on testing videos.

## 2 Related Work

Shadows are physical phenomena. Early shadow removal works, without much training data, usually focused on studying different physical shadow properties [8,7,9,6,1,10,26,53]. Many works look for cues to remove shadows starting from shadow edges. Finlayson *et al.*[9] used shadow edges to estimate a scaling factor that differentiates shadow areas from their non-shadow counterparts. Wu & Tang [51] imposed a smoothness constraint alongside the shadow boundaries to handle penumbra areas. Wu *et al.*[50] detected strong shadow-edges to remove shadows on the whole image. Shor & Lischinki [39] defined an affine relationship between shadow and non-shadow pixels where they used the areas surrounding the shadow edges to estimate the parameters of such affine transforms.

Shadow boundary effects can also be modeled via image matting [14]. Wu *et al.*[52] estimated a matte layer representing the pixel-wise shadow probability to estimate a color transfer function to remove shadows. Chuang *et al.* [3] computed a shadow matte from video for shadow editing. They computed the lit and shadow images by finding min-max values at each pixel location throughout all frames of a video captured by a static camera. We use this technique to create a video dataset for testing shadow removal methods in Sec. 4.4.

Current shadow removal methods [22,17,55,5,47] use deep-learning models trained with full supervision on large-scale datasets [47,37] of paired shadow and shadow-free images. Pairs are obtained by taking a photo with shadows, then removing the occluders from the scene to take the photo without shadows. Deshadow-Net [37] extracted multi-context features to predict a matte layer that

removes shadows. Some works use adversarial frameworks to train their shadow removal. In [47] a unified adversarial framework predicted shadow masks and removed shadows. Similarly, Ding *et al.*[5] used an adversarial signal to improve shadow removal in an iterative manner. Note that these methods use the shadow-free image as the main training signal while our method is trained only through an adversarial loss. In prior work [22] we constrained shadow removal by a physical model of shadow formation. We trained networks to extract shadow parameters and a matte layer to remove shadows. We adapt this model to patch-based shadow removal. Note that in [22], all shadow parameters and matting layers were pre-computed using paired training images and the network was trained to simply regress those values, whereas our model automatically estimates them through adversarial training. MaskShadow-GAN [17] is the only deep-learning method that learns shadow removal from just unpaired training data.

## 3 Method

We describe our patch-based shadow removal in Sec. 3.1. Our whole image pipeline for shadow removal is described in Sec. 3.2. For both image-level and patch-level shadow removal, we use shadow matting [3,35,40,49] to express a shadow-free image $I^{\text{shadow-free}}$ by:

$$I^{\text{shadow-free}} = I^{\text{relit}} \cdot \alpha + I^{\text{shadow}} \cdot (1 - \alpha) \tag{1}$$

with $I^{\text{shadow}}$ the shadow image, $\alpha$ the matting layer, and $I^{\text{relit}}$ the relit image. The relit image contains shadow pixels relit to their non-shadow values, computed via a linear function following a physical shadow formation model [22,39]:

$$I_i^{\text{relit}} = w \cdot I_i^{\text{shadow}} + b \tag{2}$$

The unknown factors in this shadow matting formula are the set of shadow parameters $(w, b)$ which define the linear function that removes the shadow effects in the umbra areas of the shadow, and the matte layer $\alpha$ that models the shadow effects on the shadow boundaries. We train a system of three networks to estimate these unknown factors via adversarial training. We use the annotated shadow segmentation masks for training. For testing, we obtain a segmentation mask for the image using the shadow detector proposed by Zhu *et al.* [59].

### 3.1 Patch-based Shadow Removal

Fig. 2 summarizes our framework to remove shadows from a single image patch, which consists of three networks: Param-Net, Matte-Net, and D-Net. Param-Net and Matte-Net predict the shadow parameters $(w, b)$ and the matte layer $\alpha$ respectively to jointly remove shadows. D-Net is the critic distinguishing between the generated image patches and the real shadow-free patches. With Param-Net and Matte-Net being the generators and D-Net being the discriminator, the

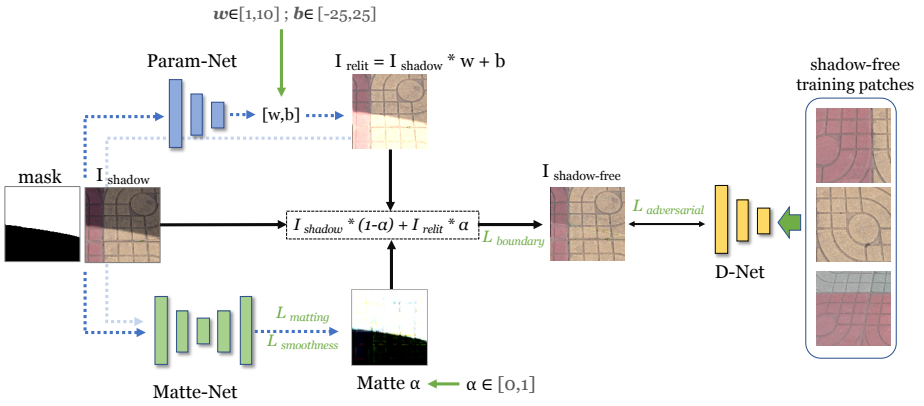


Fig. 2: **Weakly-supervised shadow decomposition.** Our framework consists of three networks: Param-Net, Matte-Net, and D-Net. Param-Net and Matte-Net predict the shadow parameters $(w, b)$ and the matte layer $\alpha$ respectively to jointly remove the shadow. Param-Net takes as input the input image patch and its shadow mask to predict three sets of shadow parameters $(w, b)$ for the three color channels, which is used to obtain a relit image. The input image patch, shadow mask, and relit image are input into Matte-Net to predict a matte layer. D-Net is the critic function distinguishing between the generated image patches and the real shadow-free patches. The only supervision signal is the set of shadow-free patches. The four losses guiding this training are the matting loss, smoothness loss, boundary loss, and adversarial loss.

three networks form an adversarial training framework where the main source of training signal is the set of shadow-free patches.

In theory, as D-Net is trained to distinguish patches containing shadow boundaries from patches without any shadows, a natural solution to fool D-Net is to remove the shadows in the input shadow patches to make them indistinguishable from shadow-free patches. However, such an adversarial signal from D-Net alone often cannot guide the generators, (Param-Net and Matte-Net) to actually remove shadows. The parameter search space is very large and the mapping is extremely under-constrained. In practice, we observe that without any constraints, Param-Net tends to output consistently high values of $(w, b)$ as they would directly increase the overall brightness of the image patches, and Matte-Net tends to introduce artifacts similar to visual patterns frequently appearing in the non-shadow areas. Thus, our main idea is to constrain this framework with physical shadow properties. Constraining the output shadow parameters, shadow mattes, and combined shadow-free images, forces the networks to only transform the input images in a manner consistent with shadow removal.

First, Param-Net estimates a scaling factor $w$ and an additive constant $b$, for each R,G,B color channel, to remove the shadow effects on the shadowed pixels in the umbra areas of the shadows via Eq. (2). Here we hypothesize that the main component that explains the shadow effects is the scaling factor $w$. Accordingly, we bound its search space to the range $[1; s_{max}]$. The minimum

value of $w = 1$ ensures that the transformation always scales up the values of the shadowed pixels. We set the search space for $b$ to the range $[-c, c]$ where we choose a relatively small value of $c = 25$ (the pixel intensity varies in the range [0,255]). Our intuition is to force the network to define the mapping mainly via the scaling factor $w$. We choose $s_{max} = 10$. This upper bound of $w$ prevents the network from collapsing as $w$ increases. As we show in the ablation study, the network fails to learn a shadow removal without proper search space limitation.

Matte-Net estimates a blending layer $\alpha$ that combines the shadow image patch and the relit image patch into a shadow-free image patch via Eq.1. The value of a pixel $i$ in the output image patch, $I_i^{\text{output}}$, is computed as:

$$I_i^{\text{output}} = I_i^{\text{relit}} \cdot \alpha_i + I_i^{\text{shadow}} \cdot (1 - \alpha_i) \tag{3}$$

We map the output of Matte-Net to [0,1] as $\alpha$ is being used as a matting layer and constrain the value of $\alpha_i$ as follows:

- If $i$ indicates a non-shadow pixel, we enforce $\alpha_i = 0$ so that the value of the output pixel $I_i^{\text{output}}$ equals its value in the input image $I_i^{\text{shadow}}$.
- If $i$ indicates a pixel in the umbra areas of the shadows, we enforce $\alpha_i = 1$ so that the value of the output pixel $I_i^{\text{output}}$ equals its relit value $I_i^{\text{relit}}$.
- We do not control the value of $\alpha$ in the penumbra areas of the shadows and rely on the training of the network to estimate these values.

where the umbra, non-shadow or penumbra areas can be roughly specified using the shadow mask. We define two areas alongside the shadow boundary, denoted as $\mathcal{M}_{in}$ and $\mathcal{M}_{out}$ - see Fig.3. $\mathcal{M}_{out}$ is the area right outside the boundary, computed by subtracting the shadow mask, $\mathcal{M}$, from its dilated version $\mathcal{M}_{dilated}$. The inside area $\mathcal{M}_{in}$ is computed similarly by subtracting an eroded shadow mask from the shadow mask. These two areas $\mathcal{M}_{in}$ and $\mathcal{M}_{out}$ roughly define a small area surrounding the shadow boundary, which can be considered as the penumbra area of the shadow. Then the above constraints are implemented as the matting loss $\mathcal{L}_{mat-\alpha}$ computed by the following formula for every pixel $i$:

$$\mathcal{L}_{mat-\alpha} = \sum_{i \in (\mathcal{M} - \mathcal{M}_{in})} |\alpha_i - 1| + \sum_{i \notin \mathcal{M}_{dilated}} |\alpha_i| \tag{4}$$

Moreover, since the shadow effects are assumed to vary smoothly across the shadow boundaries, we enforce an $L1$ smoothness loss on the spatial gradients of the matte layer, $\alpha$. This smoothness loss $\mathcal{L}_{sm}$ also prevents Matte-Net from producing undesired artifacts since it enforces local uniformity. This loss is:

$$\mathcal{L}_{sm-\alpha} = |\nabla \alpha| \tag{5}$$

Then, given a set of estimated parameters $(w, b)$ and a matte layer $\alpha$, we obtain an output image $I^{\text{output}}$ via the image decomposition formula (1). We penalize the $L1$ difference between the average intensity of pixels lying right outside and inside the shadow boundary, which are the two areas $\mathcal{M}_{in}$ and $\mathcal{M}_{out}$. This shadow boundary loss $\mathcal{L}_{bd}$ is computed by:

Input Image Shadow Mask $\mathcal{M}_{in}$ (green) & $\mathcal{M}_{out}$ (red)

Fig. 3: **The penumbra area of the shadow.** We define two areas alongside the shadow boundary, denoted as $\mathcal{M}_{in}$ (shown in green) and $\mathcal{M}_{out}$ (shown in red). These two areas roughly define a small region surrounding the shadow boundary, which can be considered as the penumbra area of the shadow.

$$\mathcal{L}_{bd} = \left| \frac{\sum_{i \in \mathcal{M}_{in}} I_i^{output}}{\sum_{i \in \mathcal{M}_{in}}} - \frac{\sum_{i \in \mathcal{M}_{out}} I_i^{output}}{\sum_{i \in \mathcal{M}_{out}}} \right| \tag{6}$$

Last, we compute the adversarial loss with the feedback from D-Net:

$$\mathcal{L}_{GAN} = \log(1 - D(I^{output})) \tag{7}$$

where $D(\cdot)$ denotes the output of D-Net.

The final objective function to train Param-Net and Matte-Net is to minimize a weighted sum of the above losses:

$$\mathcal{L}_{final} = \lambda_{sm}\mathcal{L}_{sm-\alpha} + \lambda_{mat}\mathcal{L}_{mat-\alpha} + \lambda_{bd}\mathcal{L}_{bd} + \lambda_{adv}\mathcal{L}_{GAN} \tag{8}$$

All these losses are essential for training our networks, as shown in our ablation study in Sec. 4.3. By using all the proposed losses together, our method is able to automatically extract a set of shadow parameters and an $\alpha$ layer from an input image patch. Fig. 4 visualizes the components extracted from our framework for two challenging input patches. In the first row, a dark shadow area is lit correctly to its non-shadow value. In the second row, the matte layer $\alpha$ is not affected by the dark material of the surface.

### 3.2 Image Shadow Removal using a patch-based model.

We estimate a set of shadow parameters and a matte layer for the input image to remove shadows via Eq. (1). First, we obtain a shadow mask using the shadow detector of Zhu *et al.* [59]. We crop the input shadow image into overlapping patches. All patches containing the shadow boundaries are then input into the three networks. We approximate the whole image shadow parameters from the patch shadow parameters, under the assumption that they share the same or very similar parameters. We simply compute the image shadow parameters as a linear combination of the patch shadow parameters. Similarly, we compute the values of each pixel in the matte layer by combining the overlapping matte patches. We set the matte layer pixels in the non-shadow area to 0 and those in

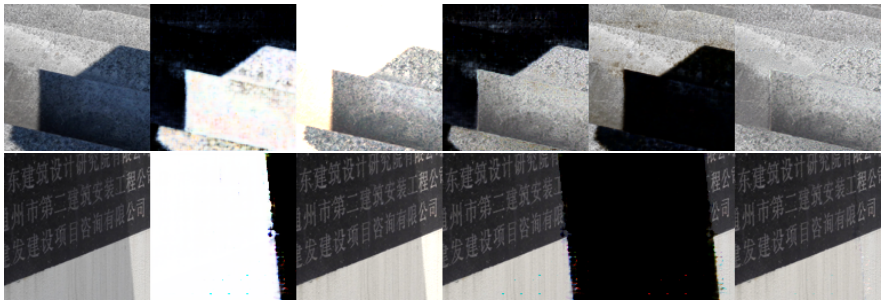


Fig. 4: **Weakly-supervised shadow image decomposition.** With only shadow mask supervision, our method automatically learns to decompose the shadow effect in the input image patch $I^{sd}$ into a matte layer $\alpha$ and a relit image $I^{relit}$. The matte layer $\alpha$ combines $I^{sd}$ and $I^{relit}$ to obtain a shadow-free image patch $I^{output}$ via Eq. (1).

the umbra area to 1. We observe that the classification scores obtained from the critic function D-Net correlate with the quality of the generated image patches. Thus, we normalize these scores to sum to 1 and use them as coefficients for the linear combinations that form the image shadow parameters and matte layer.

## 4 Experiments

### 4.1 Network Architectures and Implementation Details.

We use a VGG-19 architecture for Param-Net and a U-Net architecture for Matte-Net. D-Net is a simple 5-layer convolutional network. To map the outputs of the networks to a certain range, we use Tanh functions together with scaling and additive constants. We use stochastic gradient descent with the Adam solver [19] to train our model. The initial learning rate for Matte-Net and D-Net is 0.0002 and for Param-Net is 0.00002. All networks were trained from scratch. We experimentally set our training parameters ($\lambda_{bd}$, $\lambda_{mat-\alpha}$, $\lambda_{sm-\alpha}$, $\lambda_{adv}$) to (0.5, 100, 10, 0.5). We train our network with batch size 96 for 150 epochs. [1]

We use the ISTD dataset [47] for training. Each original training image of size $640 \times 480$ is cropped into patches of size $128 \times 128$ with a step size of 32. This creates 311,220 image patches from 1,330 training shadow images. This training set includes 151,327 non-shadow patches, 147,312 shadow-boundary patches, and 12,581 full-shadow patches.

### 4.2 Shadow Removal Evaluation

We first evaluate our method on the adjusted testing set of the ISTD dataset [47]. The testing set was adjusted [22] to mitigate the color mismatch between

[1] All code, trained models, and data are available at: `https://www3.cs.stonybrook.edu/~cvl/projects/FSS2SR/index.html`.

Table 1: **Shadow removal results of our networks compared to state-of-the-art shadow removal methods on the adjusted ISTD testing set [22,47]**. The metric is RMSE (the lower, the better). Best results are in bold.

| Methods | Training Data | Shadow | Non-Shadow | All |
|---|---|---|---|---|
| Input Image | - | 40.2 | 2.6 | 8.5 |
| Yang *et al.* [53] | - | 24.7 | 14.4 | 16.0 |
| Guo *et al.* [14] | Shd. Free + Shd. Mask | 22.0 | 3.1 | 6.1 |
| Gong *et al.* [11] | - | 13.3 | - | - |
| ST-CGAN [47] | Shd. Free + Shd. Mask | 13.4 | 7.7 | 8.7 |
| DeshadowNet [37] | Shd. Free | 15.9 | 6.0 | 7.6 |
| MaskShadow-GAN [15] | Shd. Free (Unpaired) | 12.4 | 4.0 | 5.3 |
| SP+M-Net [22] | Shd. Free + Shd.Mask | **7.9** | 3.1 | **3.9** |
| Ours | Shd. Mask | 9.7 | **3.0** | 4.0 |

the input shadow image and the ground-truth shadow-free image due to data acquisition setup. Following previous work [47,14,37,22], we compute the root-mean-square-error (RMSE) in the LAB color space on the shadow area, non-shadow area, and the whole image, where all shadow removal results are re-sized to $256 \times 256$. Note that our method can take any size image as input. We used the Zhu *et al.* [59] shadow detector, pre-trained on the SBU dataset and fine-tuned on the ISTD dataset, to obtain the shadow masks for our testing, as in [22].

In Table 1, we compare our weakly-supervised methods with the recent state-of-the-art methods of Guo *et al.* [14], Gong *et al.* [11], Yang *et al.* [53], ST-CGAN [47], DeshadowNet [37], MaskShadow-GAN [15], and SP+M-Net [22]. The second column shows the training data of each method. All other deep-learning methods require paired shadow-free images as training signal except MaskShadow-GAN, which is trained on unpaired shadow and shadow-free images from the ISTD dataset. ST-CGAN and SP+M-Net also require the training shadow masks. Our method, trained without any shadow-free image, got 9.7 RMSE on the shadow areas, which is competitive with SP+M-Net. However, SP+M-Net requires full supervision.

Our method outperforms MaskShadow-GAN by 22%, reducing the RMSE in the shadow area from 12.4 to 9.7 while also achieving lower RMSE on the non-shadow area. We outperform DeshadowNet and ST-CGAN, two methods that were trained with paired shadow and shadow-free images, reducing the RMSE by 38% and 26% respectively.

Fig. 5 compares qualitative shadow removal results from our method with other state-of-the-art methods on the ISTD dataset. Our method, trained with just an adversarial signal, produces clean shadow-free images with very few artifacts. On the other hand, ST-CGAN and MaskShadow-GAN tend to produce blurry images, introduce artifacts, and often relight the wrong image parts. Our method generates images which are visually similar to that of SP+M-Net. While SP+M-Net is less affected by the error in the shadow masks (shown in the 2nd


Fig. 5: **Comparison of shadow removal on ISTD dataset.** Qualitative comparison between our method and the state-of-the-art methods: ST-CGAN [47], MaskShadow-GAN [15], SP+M-Net[22]. Our method, trained without any shadow-free images, produces clean shadow-free images with very few artifacts.

row), our method generates images with more consistent colors between areas inside and outside the shadow boundaries (3rd and 4th rows). In all cases, our method preserves almost perfectly the textures beneath the shadows (last row).

### 4.3 Ablation Studies

We conduct ablation studies to better understand the effects of each proposed component in our framework. Starting from the original model with all the proposed features and losses, we train new models removing the proposed components one at a time. Table 2 summarizes these experiments. The first row shows the results of our model when we set the search space of the scaling factor $w$ to $[-10, 10]$ and the search space of the additive constant $b$ to $[-255, 255]$. In this case, the model collapses and consistently outputs uniformly dark images. Similarly, the model collapses when we omit the boundary loss $\mathcal{L}_{bd}$. We observe that this loss is essential in stabilizing the training as it prevents the Param-Net from outputting consistently high values.

Table 2: **Ablation Studies.** We train our network without a certain loss or feature and report the shadow removal performances on the ISTD dataset [47]. The metric is RMSE (the lower, the better). The table shows that all the proposed features in our model are essential in training for shadow removal.

| Methods | Shadow | Non-Shadow | All |
|---|---|---|---|
| Input Image | 40.2 | 2.6 | 8.5 |
| Ours w/o limiting search space | 47.5 | 2.9 | 9.9 |
| Ours w/o $\mathcal{L}_{bd}$ | 41.7 | 3.9 | 9.8 |
| Ours w/o $\mathcal{L}_{mat-\alpha}$ | 38.7 | 3.1 | 9.0 |
| Ours w/o $\mathcal{L}_{sm-\alpha}$ | 10.2 | 2.8 | 4.0 |
| Ours w/o $\mathcal{L}_{GAN}$ | 26.9 | 2.9 | 6.8 |
| Ours | 9.7 | 3.0 | 4.0 |

The matting loss $\mathcal{L}_{mat-\alpha}$ and $\mathcal{L}_{GAN}$ loss are critical for learning proper shadow removal. We observe that without the matting loss $\mathcal{L}_{mat-\alpha}$, the model behaves similarly to an image inpainting model where it tends to modify all parts of the images to fool the discriminator. Last, dropping the smoothness loss $\mathcal{L}_{sm}$ only results in a slight drop in shadow removal performance, from 9.7 to 10.2 RMSE on the shadow areas. However, we observe more visible boundary artifacts on the output images without this loss.

### 4.4 Video Shadow Removal

Video Shadow Removal is challenging for shadow removal methods. A video sequence has hundreds of frames with changing shadows. It is even harder for videos with a moving camera, moving objects, and illumination changes.

To better evaluate the performance of shadow removal methods in videos, we collected a set of 8 videos, each containing a static scene without visible moving objects. We cropped those videos to obtain clips with the only dominant motions caused by the shadows (either by direct light motion or motion of the unseen occluders). As can be seen from the top row of Fig. 6, the dataset includes videos containing shadows cast by close-up occluders, far distance occluders, videos with simple-to-complex shadows, and shadows on various types of backgrounds and materials. Inspired by [3], we propose a "max-min" technique to obtain a single pseudo shadow-free frame for each video: since the camera is static and there is no visible moving object in the frames, the changes in the video are caused by the moving shadows. We first obtain two images $V_{max}$ and $V_{min}$ by taking the maximum and minimum intensity values at each pixel location across the whole video. $V_{max}$ is then the image that contains the shadow-free values of pixels if they ever go out of the shadows. Similarly, their shadowed values, if they ever go into the shadows, are captured in $V_{min}$. Fig. 6 shows these two images for a video named "plant". From these two images, we can trivially obtain a mask, namely moving-shadow $\mathcal{M}$, marking the pixels appearing in both the shadow

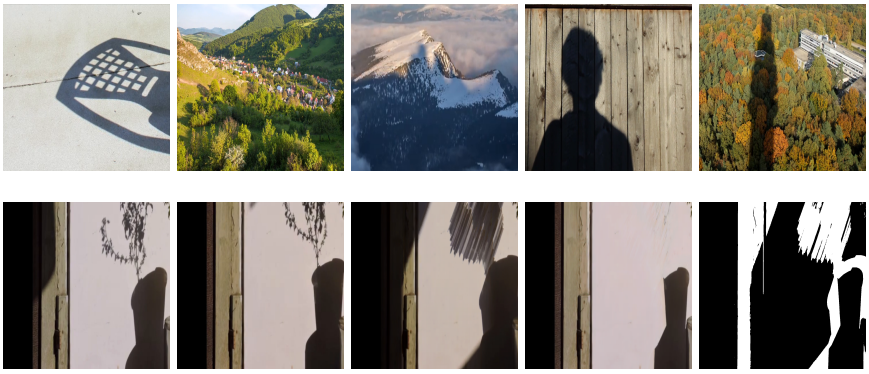


Fig. 6: **Examples of Video Shadow Removal dataset.** The dataset consists of videos where both the scene and the visible objects remaining static. The top row shows frames of different videos in our dataset. The second row visualizes our method to obtain the shadow-free frames for evaluating shadow removal.

and non-shadow areas in the video:

$$\mathcal{M}_i = \begin{cases} 1 & \text{if } V_{max,i} > V_{min,i} + \epsilon \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

where we set a small threshold of $\epsilon = 40$. This method allows us to obtain pairs of shadow and non-shadow pixel values in the moving-shadow mask, $\mathcal{M}$, for free.

To measure shadow removal performance, we input the frames of these videos into the shadow removal algorithm and measure the RMSE on the LAB color channel between the output frame and the image $V_{max}$ on the moving-shadow area $\mathcal{M}$. We compute RMSE on each video and take their average to measure the shadow removal performance on the whole dataset. Table 3 summarizes the performance of our methods compared to MaskShadow-GAN[15] and SP+M-Net[22] on these videos. Our method outperforms SP+M-Net and MaskShadow-GAN, reducing the RMSE by 5% and 11% respectively. As our method only needs shadow segmentation masks for training, we use a pre-trained shadow detection model [59] to obtain a set of shadow masks for each video. While these shadow mask sets are imperfect, fine-tuning our model using this free supervision results in a 10% error reduction, showing the advantage of our training scheme. Fig. 7 visualizes two example shadow removal results for different methods. We show a single input frame of each video. From left to right are the input frame, the shadow removal results of MaskShadow-GAN [15], the results of SP+M-Net [22], the results of our model trained on the ISTD dataset, and the result of our model fine-tuned with each testing video for 1 epoch. The top row shows an example where all methods perform relatively well. Our method seems to have better color balance between the relit pixels and the non-shadow pixels, although there is a visible boundary artifact due to imperfect shadow masks. After 1 epoch of fine-tuning, these artifacts are greatly suppressed. The bottom row shows a challenging case where all methods fail to remove shadows properly.

Table 3: **Shadow removal results on our proposed Video Shadow Removal dataset**. The metric is RMSE (the lower, the better), compared to the pseudo shadow-free frame on the moving shadow mask. All methods were pre-trained on the ISTD dataset. Ours+ denotes our model fine-tuned for one epoch on each video using the shadow masks generated by a shadow detector [59] pre-trained on the SBU dataset[43]

| Methods | Input Frame | [15] | [22] | Ours | Ours+ |
|---|---|---|---|---|---|
| RMSE | 32.9 | 23.5 | 22.2 | 20.9 | 18.0 |

Input Frame MaskShadow-GAN SP+M-Net Ours Ours+

Fig. 7: **Shadow Removal on Videos.**We visualize the shadow removal results of different methods on two frames extracted from our video dataset. "Ours+" denotes the results of our model fine-tuned with each testing video for 1 epoch. Top row shows an example where all methods perform relatively well. Bottom row shows a challenging case where all methods fail to remove shadow properly.

## 5 Conclusion

We presented a novel patch-based deep-learning model to remove shadows from images. This method can be trained on patches cropped directly from the shadow images, using the shadow segmentation mask as the only supervision signal. This obviates the dependency on paired training data and allows us to train this system on any kind of shadow image. The main contribution of this paper is a set of physics-based constrains that enable the training of this mapping. We have illustrated the effectiveness of our method on the standard ISTD dataset [47] and on our novel Video Shadow Removal dataset. As shadow detection methods mature with the aid of recently proposed shadow detection datasets [48,16], our method can be trained to remove shadows for a very low annotation cost.

**Acknowledgements.** This work was partially supported by the Partner University Fund, the SUNY2020 ITSC, and a gift from Adobe. Computational support provided by IACS and a GPU donation from NVIDIA. We thank Kumara Kahatapitiya and Cristina Mata for assistance with the manuscript.

## 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**, 1202–1216 (2011)
2. Choi, Y., Choi, M.J., Kim, M., Ha, J.W., Kim, S., Choo, J.: Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition pp. 8789–8797 (2017)
3. Chuang, Y.Y., Goldman, D.B., Curless, B., Salesin, D.H., Szeliski, R.: Shadow matting and compositing. ACM Transactions on Graphics **22**(3), 494–500 (July 2003), sepcial Issue of the SIGGRAPH 2003 Proceedings
4. Dare, P.: Shadow analysis in high-resolution satellite imagery of urban areas. Photogrammetric Engineering Remote Sensing **71**, 169–177 (02 2005). https://doi.org/10.14358/PERS.71.2.169
5. Ding, B., Long, C., Zhang, L., Xiao, C.: Argan: Attentive recurrent generative adversarial network for shadow detection and removal. 2019 IEEE/CVF International Conference on Computer Vision (ICCV) pp. 10212–10221 (2019)
6. Drew, M.S.: Recovery of chromaticity image free from shadows via illumination invariance. In: In IEEE Workshop on Color and Photometric Methods in Computer Vision, ICCV'03. pp. 32–39 (2003)
7. Finlayson, G., Drew, M.S.: 4-sensor camera calibration for image representation invariant to shading, shadows, lighting, and specularities. In: Proceedings of the International Conference on Computer Vision. vol. 2, pp. 473–480 vol.2 (July 2001). https://doi.org/10.1109/ICCV.2001.937663
8. Finlayson, G., Hordley, S., Lu, C., Drew, M.: On the removal of shadows from images. IEEE Transactions on Pattern Analysis and Machine Intelligence (2006)
9. Finlayson, G., Hordley, S.D., Drew, M.S.: Removing shadows from images. In: Proceedings of the European Conference on Computer Vision. pp. 823–836. ECCV '02, Springer-Verlag, London, UK, UK (2002), `http://dl.acm.org/citation.cfm?id=645318.649239`
10. Fredembach, C., Finlayson, G.D.: Hamiltonian path based shadow removal. In: BMVC (2005)
11. Gong, H., Cosker, D.: Interactive removal and ground truth for difficult shadow scenes. J. Opt. Soc. Am. A **33**(9), 1798–1811 (2016). https://doi.org/10.1364/JOSAA.33.001798, `http://josaa.osa.org/abstract.cfm?URI=josaa-33-9-1798`
12. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of wasserstein gans. In: Advances in Neural Information Processing Systems. pp. 5767–5777 (2017)
13. Guo, R., Dai, Q., Hoiem, D.: Single-image shadow detection and removal using paired regions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2011)
14. Guo, R., Dai, Q., Hoiem, D.: Paired regions for shadow detection and removal. IEEE Transactions on Pattern Analysis and Machine Intelligence (2012)
15. Hu, X., Jiang, Y., Fu, C.W., Heng, P.A.: Mask-ShadowGAN: Learning to remove shadows from unpaired data. In: ICCV (2019), to appear
16. Hu, X., Wang, T., Fu, C.W., Jiang, Y., Wang, Q., Heng, P.A.: Revisiting shadow detection: A new benchmark dataset for complex world. ArXiv **abs/1911.06998** (2019)

17. Hu, X., Zhu, L., Fu, C.W., Qin, J., Heng, P.A.: Direction-aware spatial context features for shadow detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2018)
18. KaewTrakulPong, P., Bowden, R.: An improved adaptive background mixture model for real- time tracking with shadow detection (2002)
19. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: Proceedings of the International Conference on Learning Representations (2015)
20. Le, H., Goncalves, B., Samaras, D., Lynch, H.: Weakly labeling the antarctic: The penguin colony case. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (June 2019)
21. Le, H., Nguyen, V., Yu, C.P., Samaras, D.: Geodesic distance histogram feature for video segmentation. In: ACCV (2016)
22. Le, H., Samaras, D.: Shadow removal via shadow image decomposition. In: Proceedings of the International Conference on Computer Vision (2019)
23. Le, H., Vicente, T.F.Y., Nguyen, V., Hoai, M., Samaras, D.: A+D Net: Training a shadow detector with adversarial shadow attenuation. In: Proceedings of the European Conference on Computer Vision (2018)
24. Le, H., Yu, C.P., Zelinsky, G., Samaras, D.: Co-localization with category-consistent features and geodesic distance propagation. In: ICCV 2017 Workshop on CEFRL: Compact and Efficient Feature Representation and Learning in Computer Vision (2017)
25. Li, Y., Tang, S., Zhang, R., Zhang, Y., Li, J., Yan, S.: Asymmetric gan for unpaired image-to-image translation. IEEE Transactions on Image Processing **28**, 5881–5896 (2019)
26. Liu, F., Gleicher, M.: Texture-consistent shadow removal. In: ECCV (2008)
27. Liu, H., Gu, X., Samaras, D.: Wasserstein gan with quadratic transport cost. In: The IEEE International Conference on Computer Vision (ICCV) (October 2019)
28. Liu, H., Xianfeng, G., Samaras, D.: A two-step computation of the exact gan wasserstein distance. In: International Conference on Machine Learning. pp. 3165–3174 (2018)
29. Liu, M.Y., Breuel, T., Kautz, J.: Unsupervised image-to-image translation networks. ArXiv **abs/1703.00848** (2017)
30. Mescheder, L., Nowozin, S., Geiger, A.: Which training methods for gans do actually converge? In: International Conference on Machine Learning (2018)
31. Miyato, T., Kataoka, T., Koyama, M., Yoshida, Y.: Spectral normalization for generative adversarial networks. In: International Conference on Machine Learning (2018)
32. Müller, T., Erdnüeß, B.: Brightness correction and shadow removal for video change detection with uavs. In: Defense + Commercial Sensing (2019)
33. Panagopoulos, A., Wang, C., Samaras, D., Paragios, N.: Estimating shadows with the bright channel cue (2010)
34. Panagopoulos, A., Wang, C., Samaras, D., Paragios, N.: Simultaneous cast shadows, illumination and geometry inference using hypergraphs. Pattern Analysis and Machine Intelligence, IEEE Transactions on **35**(2), 437–449 (2013). https://doi.org/10.1109/TPAMI.2012.110
35. Porter, T., Duff, T.: Compositing digital images. Proceedings of the ACM SIGGRAPH Conference on Computer Graphics **18**(3) (January 1984)
36. Prati, A., Mikic, I., Trivedi, M.M., Cucchiara, R.: Detecting moving shadows: Algorithms and evaluation. IEEE Trans. Pattern Anal. Mach. Intell. **25**, 918–923 (2003)

37. Qu, L., Tian, J., He, S., Tang, Y., Lau, R.W.H.: Deshadownet: A multi-context embedding deep network for shadow removal. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2017)
38. Shiting, W., Hong, Z.: Clustering-based shadow edge detection in a single color image. In: International Conference on Mechatronic Sciences, Electric Engineering and Computer. pp. 1038–1041 (Dec 2013). https://doi.org/10.1109/MEC.2013.6885215
39. Shor, Y., Lischinski, D.: The shadow meets the mask: Pyramid-based shadow removal. Computer Graphics Forum **27**(2), 577–586 (April 2008)
40. Smith, A.R., Blinn, J.F.: Blue screen matting. In: Proceedings of the ACM SIGGRAPH Conference on Computer Graphics (1996)
41. Su, N., Zhang, Y., Tian, S., Yan, Y., Miao, X.: Shadow detection and removal for occluded object information recovery in urban high-resolution panchromatic satellite images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **9**, 2568–2582 (2016)
42. Thanh-Tung, H., Tran, T., Venkatesh, S.: Improving generalization and stability of generative adversarial networks. In: International Conference on Learning Representations (2019)
43. Vicente, T.F.Y., Hoai, M., Samaras, D.: Noisy label recovery for shadow detection in unfamiliar domains. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016)
44. Vicente, T.F.Y., Hoai, M., Samaras, D.: Leave-one-out kernel optimization for shadow detection and removal. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(3), 682–695 (2018)
45. 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: Proceedings of the European Conference on Computer Vision (2016)
46. Vicente, T.F.Y., Samaras, D.: Single image shadow removal via neighbor-based region relighting. In: Proceedings of the European Conference on Computer Vision Workshops (2014)
47. 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 (2018)
48. Wang, T., Hu, X., Wang, Q., Heng, P.A., Fu, C.W.: Instance shadow detection. CVPR (2020)
49. Wright, S.: Digital compositing for film and video. In: Focal Press (2001)
50. Wu, Q., Zhang, W., Kumar, B.V.K.V.: Strong shadow removal via patch-based shadow edge detection. 2012 IEEE International Conference on Robotics and Automation pp. 2177–2182 (2012)
51. Wu, T.P., Tang, C.K.: A bayesian approach for shadow extraction from a single image. Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1 **1**, 480–487 Vol. 1 (2005)
52. Wu, T.P., Tang, C.K., Brown, M.S., Shum, H.Y.: Natural shadow matting. ACM Trans. Graph. **26**(2) (June 2007). https://doi.org/10.1145/1243980.1243982, `http://doi.acm.org/10.1145/1243980.1243982`
53. Yang, Q., Tan, K., Ahuja, N.: Shadow removal using bilateral filtering. IEEE Transactions on Image Processing **21**, 4361–4368 (2012)
54. Yi, Z., Zhang, H., Tan, P., Gong, M.: Dualgan: Unsupervised dual learning for image-to-image translation. 2017 IEEE International Conference on Computer Vision (ICCV) pp. 2868–2876 (2017)

55. Zhang, L., Long, C., Zhang, X., Xiao, C.: Ris-gan: Explore residual and illumination with generative adversarial networks for shadow removal. In: AAAI Conference on Artificial Intelligence (AAAI) (2020)
56. 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**, 611–624 (2019)
57. Zheng, Q., Qiao, X., Cao, Y., Lau, R.W.H.: Distraction-aware shadow detection. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 5162–5171 (2019)
58. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Computer Vision (ICCV), 2017 IEEE International Conference on (2017)
59. 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 (2018)