International Journal of Computer Vision (2023) 131:2471–2488
https://doi.org/10.1007/s11263-023-01823-9

MANUSCRIPT

Learning to Remove Shadows from a Single Image

Hao Jiang[1] · Qing Zhang[1] · Yongwei Nie[2] · Lei Zhu[3,4] · Wei-Shi Zheng[1]

Received: 27 November 2022 / Accepted: 15 May 2023 / Published online: 9 June 2023

Abstract

Recent learning-based shadow removal methods have achieved remarkable performance. However, they basically require massive paired shadow and shadow-free images for model training, which limits their generalization capability since these data are often cumbersome to obtain and lack of diversity. To address the problem, we present Self-ShadowGAN, a novel adversarial framework that is able to learn to remove shadows in an image by training solely on the image itself, using the shadow mask as the only supervision. Our approach is built upon the concept of histogram matching, by constraining the deshadowed regions produced by a shadow relighting network share similar histograms to the original shadow-free regions via a histogram-based discriminator. In order to speed up the single image training, we define the shadow relighting network to be lightweight multi-layer perceptions (MLPs) that estimate spatially-varying shadow relighting coefficients, where the parameters of the MLPs are predicted from a low-resolution input by a fast convolutional network and then upsampled back to the original full-resolution. Experimental results show that our method performs favorably against the state-of-the-art shadow removal methods, and is effective to process previously challenging shadow images.

Keywords Shadow removal · Image relighting · Generative adversarial network (GAN)

1 Introduction

Shadow removal has long been a fundamental problem in computer vision and image processing, because the presence of shadows in an image may not only degrade the overall visual quality (*e.g.*, undesirable shadows on portrait and document images), but also adversely affect various computer vision tasks including object detection and segmentation (He et al., 2016, 2017), intrinsic image decomposition (Li & Snavely, 2018; Wu et al., 2021), and image editing (Liu et al., 2020; Nestmeyer et al., 2020; He et al., 2021)).

Recent shadow removal methods are basically learning-based (Gryka et al., 2015; Guo et al., 2012; Khan et al., 2015), especially deep-learning-based (Xu et al., 2017; Qu et al., 2017; Wang et al., 2018; Hu et al., 2019; Zhang et al., 2020; Cun et al., 2020; Inoue & Yamasaki, 2020; Liu et al., 2021; Fu et al., 2021; Le & Samaras, 2021; Wan et al., 2022; Zhu et al., 2022). Most of them are trained in a fully supervised manner, relying on large-scale datasets consisting of paired shadow and shadow-free images. However, as described in Hu et al. (2019), such paired data is tedious to collect, usually covers very limited scene categories, and may be unreliable due to possible inconsistencies in color, luminance, and camera views, limiting the effectiveness and generalization capability of existing learning-based shadow removal methods (Fig. 1).

To address the issues arising from the dependency on paired training data, a recent research trend is to learn to remove shadows from unpaired data. A pioneering work

✉ Qing Zhang
zhangq93@mail.sysu.edu.cn

Hao Jiang
jiangh69@mail2.sysu.edu.cn

Yongwei Nie
nieyongwei@scut.edu.cn

Lei Zhu
leizhu@ust.hk

Wei-Shi Zheng
wszheng@ieee.org

[1] School of Computer Science and Engineering, Sun Yat-Sen University, Guangzhou 510006, Guangdong, China

[2] School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, Guangdong, China

[3] The Hong Kong University of Science and Technology (Guangzhou), Nansha, Guangzhou 51140, Guangdong, China

[4] The Hong Kong University of Science and Technology, Hong Kong SAR, China

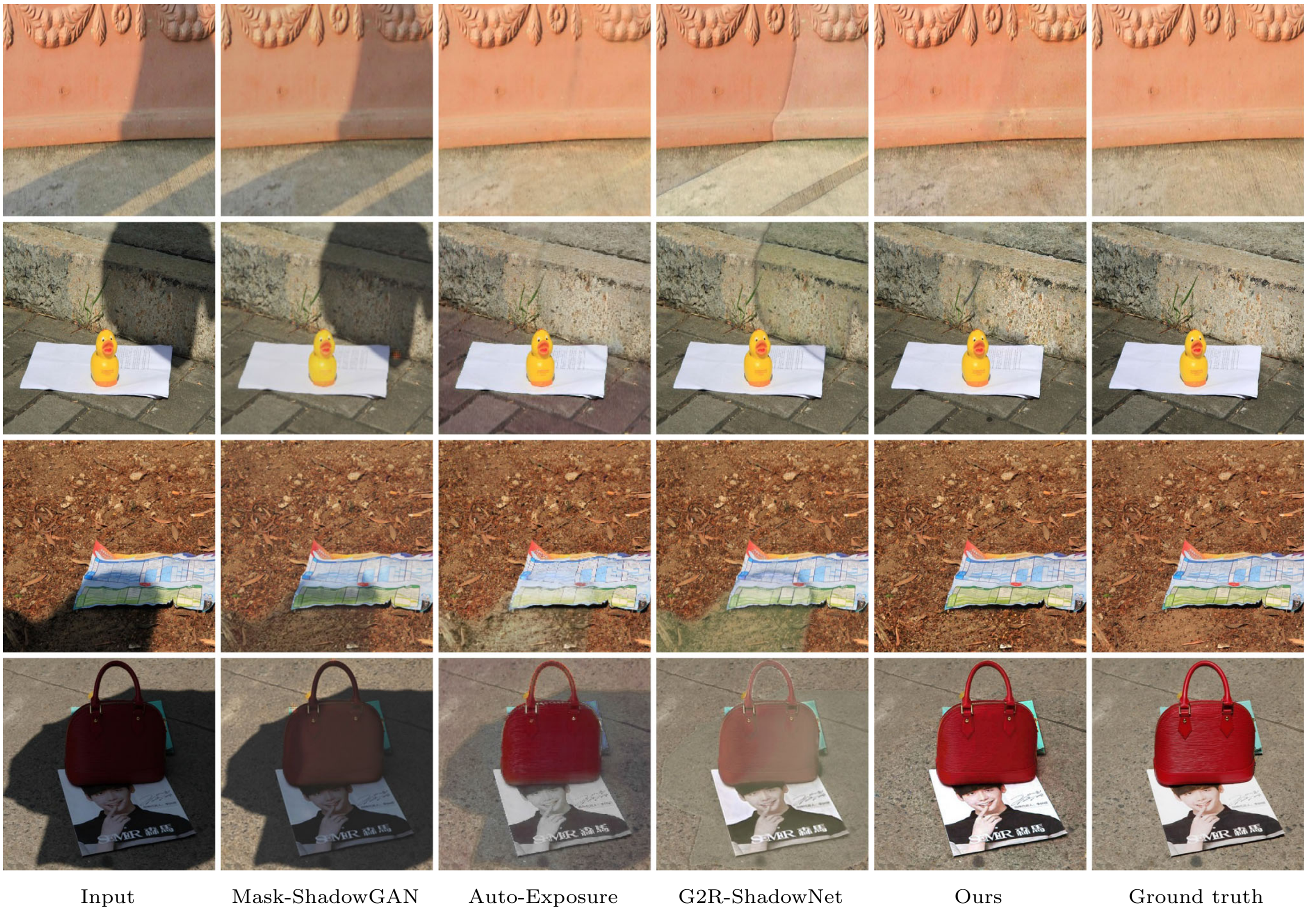

Fig. 1 Visual comparison of shadow removal results on images from the SRD dataset (Qu et al., 2017) with three state-of-the-art methods including Mask-ShadowGAN (Hu et al., 2019), Auto-Exposure (Fu et al., 2021), and G2R-ShadowNet (Liu et al., 2021)). As can be seen, there are different types of shadows in the presented input images, including: (i) shadows cast on highly textured background, (ii) shadows with inhomogeneous luminance, and (iii) shadow regions without texture-consistent shadow-free regions as reference

is Mask-ShadowGAN (Hu et al., 2019), which formulated a mask-guided CycleGAN (Zhu et al., 2017) based framework for learning shadow removal from unpaired shadow and shadow-free images. However, as analyzed by Le and Samaras (2020), it performs well only if there is sufficient statistical similarity between images from the shadow and shadow-free domains, *i.e.*, the domain gap between the unpaired data is assumed to be small.

To address the limitation of unpaired learning, Le and Samaras (2020) proposed to train an adversarial shadow removal framework using unpaired shadow and shadow-free patches cropped from a set of shadow images themselves. As the shadow and shadow-free patches are from the same images, the domain gap can be effectively reduced. However, this method assumes that shadows are homogeneous, which limits its effectiveness in handling complex shadows. Liu et al. (2021) presented G2R-ShadowNet, which learns to generate pseudo shadows for the shadow-free regions in an image and then trains a shadow removal generator based on the synthesized paired data. Since the realism of synthesized shadows is hard to guarantee, it may produce unnatural results.

In contrast to previous learning-based shadow removal methods which require either large-scale paired or unpaired training data for model training, we in this paper propose to avoid the training data dependency issue, by learning to remove shadows in an image through training on the image itself at test time. Our approach is formulated as an adversarial framework named as Self-ShadowGAN, which takes a single image and its shadow mask as the only training input. The key idea is to remove shadows based on histogram matching, *i.e.*, constraining the recovered deshadowed regions and the original shadow-free regions share the similar histograms. Note that although histogram matching has been adopted for shadow removal in an early work (Vicente & Samaras, 2014), we make the first attempt to enable superior shadow removal performance by implementing histogram matching in an adversarial framework trained on a single image.

Our proposed Self-ShadowGAN is comprised of a shadow relighting network as the generator for shadow removal, and two discriminators for ensuring that the deshadowed regions generated by the shadow relighting network are shadow-free and visually realistic. Inspired by Le and Samaras (2019), Le and Samaras (2020), to constrain the shadow removal procedure and stabilize the adversarial training, the relighting network is designed to predict pixel-adaptive shadow relighting coefficients characterized by a physical model of shadow formation. Based on the idea of histogram matching, a histogram-based discriminator is developed to perform illumination recovery for the shadow regions, using the histograms of original shadow-free regions as guidance. Besides, a patch-based discriminator is adopted for making textures in the deshadowed regions clear and consistent with those in original shadow-free regions. To reduce the training time cost while maintaining high visual quality for the shadow removal results, we define the shadow relighting network to be lightweight Multi-Layer Perceptions (MLPs) that estimate spatially-varying shadow relighting coefficients, where the parameters of the MLPs are predicted from a low-resolution input by a fast convolutional network and then upsampled back to the original full-resolution. In summary, our work makes the following contributions:

- To the best of our knowledge, this is the first work that allows to train an image-specific shadow removal network with superior performance from a single image.
- We present Self-ShadowGAN, a novel adversarial framework for shadow removal, where a histogram-based discriminator is designed to recover illumination for the shadow regions, and a shadow relighting network formed by lightweight MLPs is developed to allow effective yet efficient relighting coefficients learning.
- We compare our method with various state-of-the-art shadow removal methods on benchmark datasets. Experiments show that our approach performs favorably against previous shadow removal methods, and is highly robust to different types of shadow images.

2 Related Work

This section reviews existing shadow removal methods from the following two aspects, *i.e.*, traditional methods and deep-learning-based methods, with a focus on the latter.

Traditional methods Early shadow removal methods are mostly developed based on physical properties of shadow (Finlayson & Drew, 2001; Finlayson et al., 2002; Drew et al., 2003; Finlayson et al., 2005; Fredembach & Finlayson, 2005; Wu et al., 2007; Liu & Gleicher, 2008; Arbel & Hel-Or, 2010; Yang et al., 2012). As analyzed by Khan et al. (2015), these methods are usually less effective in handling shadows in complex real-world scenes. Another line of research directly takes low-level features (*e.g.*, edge/gradient (Wu & Tang, 2005; Shor & Lischinski, 2008; Finlayson et al., 2009; Wu et al., 2012), intensity (Guo et al., 2011; Xiao et al., 2013; Gryka et al., 2015; Gong & Cosker, 2014; Zhang et al., 2015), texture (Guo et al., 2012; Xiao et al., 2013; Ma et al., 2016), etc.) as cues for shadow removal. However, due to lack of high-level semantics, methods in this category may produce unnatural results.

Deep-learning-based methods Due to the advent of deep learning and the availability of large-scale paired training datasets (Qu et al., 2017; Wang et al., 2018), supervised shadow removal methods have received considerable research effort. Khan et al. (2015) firstly detected shadows using multiple convolutional neural networks (CNNs), and then removed shadows based on the shadow matte estimated from a Bayesian model. Qu et al. (2017) presented DeshadowNet, which extracts multi-context features to predict a matte layer for shadow removal. Wang et al. (2018) proposed to detect and remove shadows via a stacked conditional generative adversarial network. Hu et al. (2018), Hu et al. (2019) recovered direction-aware spatial contexts for shadow detection and removal. Le and Samaras (2019) removed shadows by learning to decompose shadow image into a linear combination of shadow-free image, shadow parameters and a matte layer. This work is later improved by incorporating an inpainting network for result refinement (Le & Samaras, 2021). Ding et al. (2019) designed an attentive recurrent generative adversarial framework for detecting and removing shadows. More recently, Fu et al. (2021) formulated shadow removal as an exposure fusion problem, while Chen et al. (2021) worked by transferring the contextual information of non-shadow regions to shadow regions. By coupling the learning procedures of shadow removal and shadow generation in a unified parameter-share framework, Zhu et al. (2022) developed a bijective mapping network for shadow removal.

In addition to the aforementioned supervised shadow removal methods, there also exists some works some works that remove the reliance on paired training data. Mask-ShadowGAN (Hu et al., 2019) learned to remove shadows from unpaired shadow and shadow-free images using a CycleGAN (Zhu et al., 2017) based framework. Liu et al. (2021) presented a lightness-guided shadow removal network based on unpaired data. Le and Samaras (2020) proposed to train shadow removal network from unpaired shadow and shadow-free patches cropped from a set of shadow images themselves. Liu et al. (2021) learned to generate pseudo shadows for original shadow-free regions to form paired training data, and then used the obtained paired data to train a shadow removal network. Jin et al. (2021) presented an unsupervised domain-classifier guided network to remove

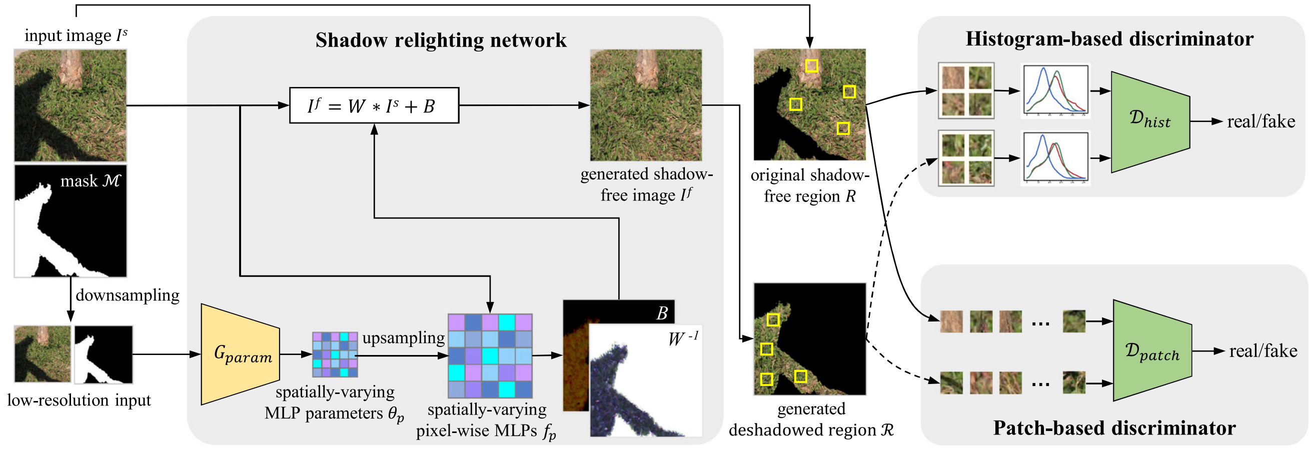
input image I^s
Shadow relighting network
$I^f = W * I^s + B$
mask $\mathcal{M}$
downsampling
low-resolution input
$\mathcal{G}_{param}$
upsampling
spatially-varying MLP parameters θ_p
spatially-varying pixel-wise MLPs f_p
B
W^{-1}
generated shadow-free image I^f
original shadow-free region R
generated deshadowed region $\mathcal{R}$
Histogram-based discriminator
$\mathcal{D}_{hist}$
real/fake
...
$\mathcal{D}_{patch}$
real/fake
Patch-based discriminator

Fig. 2 The schematic illustration of our shadow removal network. We first downsample the full-resolution input (shadow image and mask) to low-resolution for predicting spatially-varying MLP parameters θ_p via a convolutional network G_{param}. These parameters are then upsampled back to the original full-resolution for parameterizing pixel-wise, spatially-varying MLPs f_p, from which we are able to generate pixel-adaptive shadow relighting coefficients W and B at full-resolution (W^{-1} is shown here, since the value of W at each pixel is no less than 1), and recover a shadow-free image I^f based on Eq. (2). Next, the generated deshadowed regions $\mathcal{R}$ in I^f are fed to the discriminators to determine whether it is real shadow-free and visually consistent with the original shadow-free regions R, where the histogram-based discriminator D_{hist} ensures the luminance and color consistency between R and $\mathcal{R}$, while the patch-based discriminator D_{patch} enforces the texture consistency between R and $\mathcal{R}$

both hard and soft shadows. In contrast to these methods, we show in this work that a single shadow image itself provides sufficient cues for training an image-specific shadow removal network with superior performance.

3 Methods

This section describes the technical details of our approach. Figure 2 illustrates the network architecture of the proposed Self-ShadowGAN. As shown, our network takes as training input a single image and its shadow mask, and produce shadow removal result of the input image once the single-image-based adversarial training finished. Note that the requirement of shadow mask has limited impact on the practicability and robustness of our model, because the shadow mask utilized in our network can be easily obtained by either using recent automatic shadow detection methods (Nguyen et al., 2017; Hu et al., 2018; Zhu et al., 2018; Zheng et al., 2019; Wang et al., 2020) or existing interactive image matting techniques (Levin et al., 2007; Wang et al., 2007), and moreover, our approach has some tolerance for inaccurate shadow masks, as demonstrated in Sect. 4.2.

3.1 Pixel-Adaptive Shadow Relighting Model

As shown in Fig. 2, we remove shadows by learning to relight the shadow regions, we thus start by introducing the shadow relighting model employed in our network.

For a given shadow image I^s and its binary shadow mask $\mathcal{M}$ (shadow and non-shadow pixels are indicated by 1 and 0, respectively), it is assumed in Shor and Lischinski (2008), Le and Samaras (2019) that the desired shadow-free image I^f can be obtained by relighting the shadow regions via the following linear mapping model:

$$I_p^f = \omega * I_p^s + b, \tag{1}$$

where p denotes a shadow pixel. ω and b are the relighting coefficients, which are constants for each pixel in the shadow area. Under this assumption, the shadow removal problem can be reduced to the estimation of two constants ω and b. However, as mentioned in Vasluianu et al. (2021), the linear model in Eq. (1) with constant relighting coefficients ω and b can only represent shadows with homogeneous luminance, and is less effective in dealing with shadows with inhomogeneous luminance. To address the limitation, we propose to formulate the following pixel-adaptive shadow relighting model

$$I_p^f = W_p * I_p^s + B_p, \tag{2}$$

where W_p and B_p are relighting coefficients at each shadow pixel p, which are designed to be spatially varying to account for complex shadows with inhomogeneous luminance. Note, the values of W and B for the original shadow-free pixels are respectively fixed to 1 and 0, for ensuring that shadow-free regions are unchanged during the shadow relighting procedure.

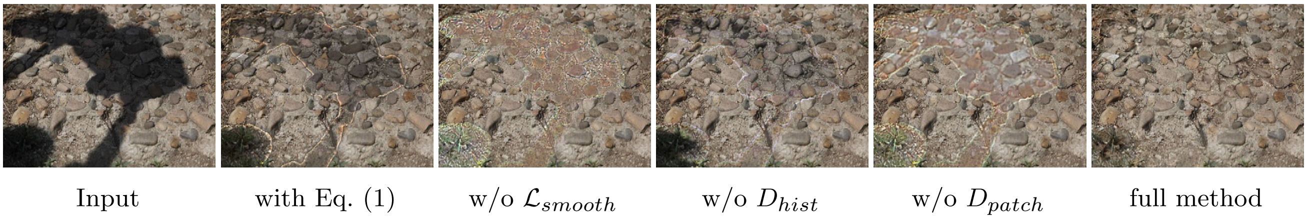

Fig. 3 Ablation study that validates the effectiveness of our proposed shadow relighting model in Eq. (2), the smoothness loss $\mathcal{L}_{smooth}$, the histogram-based discriminator D_{hist}, and the patch-based discriminator D_{patch} (w/o - without)

3.2 Shadow Relighting Network

Based on Eq. (2), we develop the shadow relighting network shown in Fig. 2, for illumination recovery of the shadow regions. Inspired by Gharbi et al. (2017), Shaham et al. (2021) and the local smoothness nature of the relighting coefficients (Le & Samaras, 2020; Shor & Lischinski, 2008), we argue that the estimation of relighting coefficients W and B need not be performed at full-resolution, and it is able to significantly reduce the computational cost yet produce high-quality results, by predicting relighting coefficients on downsampled low-resolution input with a fast convolutional network.

To build a shadow relighting network with reduced computational cost, we propose to model W and B at each pixel p as a spatially-varying pixel-wise nonlinear function f_p with respect to the full-resolution input image. Considering that the spatial context of an image determines the expressiveness of the pixel-wise function, each pixel-wise function f_p is thus designed to take as input the pixel's coordinates p, in addition to its color value I^s_p, and is parameterized with spatially-varying parameters θ_p and conditioned on the input image I^s. Specifically, we define each f_p as a Multi-Layer Perception (MLP) with ReLU activations, *i.e.*,

$$\left(W_p, B_p\right) = f_p(I^s_p, p) = f(I^s_p, p; \theta_p), \tag{3}$$

where f denotes the shared MLP architecture of all pixel-wise functions, and θ_p are the weights and biases for the MLPs. Note, we use MLPs with 5 layers and 64 channels per layer in all our experiments.

Efficient relighting coefficients prediction Naively predicting the parameters θ_p for each pixel would inevitably incur high computational cost. Based on the local smoothness nature of W and B, we instead predict a grid of parameter vectors θ_p by a convolutional network G_{param} that processes a much lower resolution image of the full-resolution input I^s. Next, the grid is upsampled to the full-resolution using nearest neighbor interpolation to obtain parameter vector θ_p for each high-resolution pixel p. The reason we employ the nearest neighbor upsampling scheme is that it can yield sufficiently good results without leading to additional arithmetic operations. G_{param} is a multi-layer convolutional network comprised of 3 convolutional layers with stride=2, which will further reduce the spatial dimensions of the downsampled image due to strided-convolution layers, and generate an output with a very low resolution of 16×16 corresponding to parameters of the full-resolution pixel-wise MLPs. Note that the computational efficiency of our model comes from both the downsampling of the full-resolution input and the spatial dimension reduction in G_{param}. We validate the effectiveness of our efficient prediction strategy in Table 1 and Fig. 12.

Our above efficient prediction strategy is essentially from HDRNet (Gharbi et al., 2017) and 3DLUT (Zeng et al., 2020) which also utilize the idea of performing main computation on downsampled input. Specifically, HDRNet and 3DLUT aim to predict color transformations that are more suitable for global image adjustment from a low-resolution input, while we predict a set of parameters that parameterize pixel-wise spatially-varying MLPs at a low-resolution to fit our pixel-adaptive shadow relighting model.

To regularize the relighting coefficients W and B, we enforce them to be locally smooth by imposing the following L_1 constraints:

$$\mathcal{L}_{smooth} = \|\nabla W\|_1 + \|\nabla B\|_1. \tag{4}$$

where ∇ is the gradient operator. By comparing the results in Fig. 3, we can notice that the proposed pixel adaptive shadow relighting model in Eq. (2) and the smoothness loss $\mathcal{L}_{smooth}$ can help produce results with better visual quality.

3.3 Histogram-Based Discriminator

It is common knowledge that shadow-free regions in an image provide important cues for illumination recovery of the shadow regions. A well-known shadow removal paradigm adopting this observation is to transfer the illumination of shadow-free regions to shadow regions. Following this paradigm, we develop a histogram-based discriminator based on the idea of histogram matching, for distinguishing whether the deshadowed regions $\mathcal{R}$ produced by the shadow relighting network have similar illumination distributions to the original shadow-free regions R. Different from previous works which basically perform illumination transfer between paired regions with similar textures (Guo et al., 2011; Chen et al., 2021), our histogram-based discriminator D_{hist} can not only avoid the time-consuming paired region search, but also enhance the method's robustness to images with complex textures, even the shadow and non-shadow regions have irrelevant texture or material (see our shadow result for the bottom input image in Fig. 1).

Our histogram-based discriminator D_{hist} is designed as a 6-layer convolutional network consisting of three 1D convolutional layers and three fully-connected layers, taking normalized RGB histograms with 64 bins per channel as training input. At each training iteration, 32 histograms are separately extracted from R and $\mathcal{R}$, and then fed to D_{hist} for discrimination. In order to account for the local diversity of illumination distributions, each histogram is calculated from a group of 16 randomly selected 20 × 20 patches, which is equivalent to computing histogram over a 80 × 80 image region. We define the following adversarial loss to optimize the generator G (*i.e.*, the shadow relighting network) and the histogram-based discriminator D_{hist}:

$$\begin{aligned}\mathcal{L}_{GAN}^{hist} &= \mathbb{E}_{H_R \sim p_{data}(H_R)}[\log D_{hist}(H_R)] \\ &\quad + \mathbb{E}_{H_{\mathcal{R}} \sim p_{data}(H_{\mathcal{R}})}[\log(1 - D_{hist}(H_{\mathcal{R}}))],\end{aligned} \tag{5}$$

where H_R and $H_{\mathcal{R}}$ are histograms extracted from R and $\mathcal{R}$, respectively. Note that as it is challenging to write $H_{\mathcal{R}}$ as a compact yet intuitive form with respect to the generator G, we omit G in the right side of Eq. (5).

Figure 3 validates the effectiveness of the histogram-based discriminator D_{hist}, where our method without D_{hist} fails to ensure the illumination consistency between the deshadowed regions and the original shadow-free regions. The reason we use three-channel RGB histogram rather than single-channel intensity histogram is because RGB histogram is beneficial to dealing with the color degradation led by shadows, and can help recover more natural color for shadow regions, as validated in Fig. 14. Naively matching RGB histogram may distort colors, but our method does not suffer from the problem because Eq. (4) would enforce the relighting coefficients W and B in the shadow relighting model to be locally smooth, which in turn help avoid possible color distortion artifacts arising from histogram matching.

3.4 Differentiable Histogram Construction

Conventional histogram constructed from hard-binning operation is not differentiable, and does not allow backward propagation of gradients during network training. To address the issue, we construct differentiable histogram based on kernel density estimation (KDE) to approximate the hard-binning process.

In contrast to conventional histogram that counts the number of pixels in each intensity interval, a KDE-based differentiable histogram is a function defined as the sum of a kernel function at each pixel. For simplicity, we describe the construction of differentiable histogram for a gray-scale image below, since the differentiable histogram of color image is similarly constructed.

For a gray-scale image I with a total number of N pixels, the gray level density $\mathcal{F}$ of the image can be formulated by kernel density estimation as the following form:

$$\hat{\mathcal{F}}_\tau(g) = \frac{1}{\tau N} \sum_p K\left(\frac{g - I_p}{\tau}\right), \tag{6}$$

where $g \in [0, 255]$ denotes a certain gray level, and I_p denotes the intensity at pixel p. $K(\cdot)$ is a kernel function. τ is the bandwidth used for controlling the smoothness of the kernel $K(\cdot)$. Similar to Avi-Aharon et al. (2020), we define $K(\cdot)$ as the derivative of the sigmoid function $\sigma(x)$ as follows

$$K(x) = \frac{d}{dx}\sigma(x) = \sigma(x) \cdot \sigma(-x), \tag{7}$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$. The reason for choosing the above kernel function is that it is a widely used non-negative real-valued integrable function, satisfying the basic normalization (*i.e.*, $\int_{-\infty}^{\infty} K(x)dx = 1$) and symmetry (*i.e.*, $K(x) = K(-x)$) requirements for a kernel.

Based on the density function in Eq. (6), we construct differentiable histogram H with discrete form, by calculating the probability of each pixel belonging to an interval centered at a certain gray level g. Specifically, the interval for each gray level g is $\Gamma_g = [g - \eta, g + \eta]$, where $\Delta L = 2\eta$ is the length of the interval, and $H(g)$ is calculated by

$$\begin{aligned}H(g) &= \int_{\Gamma_g} \hat{\mathcal{F}}_\tau(x)dx \\ &= \frac{1}{N}\sum_p \sigma\left(\frac{x - I_p}{\tau}\right)\Bigg|_{g-\eta}^{g+\eta} \\ &= \frac{1}{N}\sum_p \left[\sigma\left(\frac{g + \eta - I_p}{\tau}\right) - \sigma\left(\frac{g - \eta - I_p}{\tau}\right)\right].\end{aligned} \tag{8}$$

We empirically set $\Delta L = 1/64$ and $\tau = 0.4 \times \Delta L$ in all our experiments, since the parameter settings can not only represent image histogram in low mean square error, but also enable relatively high computational efficiency. Figure 4 compares conventional histograms and the differentiable histograms constructed by kernel density estimation.

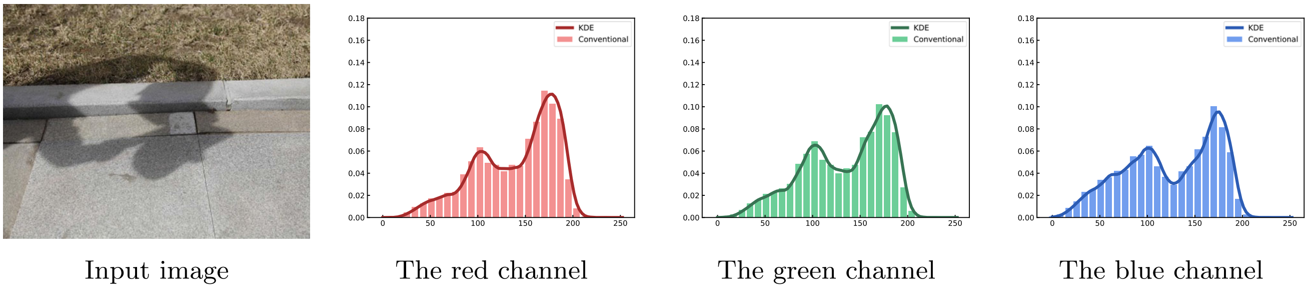

Fig. 4 Comparison of conventional histograms and the differentiable histograms calculated by kernel density estimation (KDE) on an input image. As shown, the differentiable histograms effectively approximate the conventional histograms

3.5 Patch-Based Discriminator

Although the histogram-based discriminator D_{hist} can help recover scene-consistent illumination for the shadow regions (see Fig. 3), it lacks the capability of texture recovery since it is designed to focus on illumination recovery. To address the problem, we introduce a patch-based discriminator D_{patch} into our adversarial framework.

The goal of D_{patch} is to make the deshadowed regions $\mathcal{R}$ visually share similar contexts to the original shadow-free regions R. To this end, we randomly select 200 patches with a small size of 32 × 32 from both $\mathcal{R}$ and R at each iteration, and feed them to D_{patch} to encourage context similarity via a adversarial loss defined as:

$$\begin{aligned}\mathcal{L}_{GAN}^{patch} &= \mathbb{E}_{P_R \sim p_{data}(P_R)}[\log(D_{patch}(P_R))] \\ &\quad + \mathbb{E}_{P_{\mathcal{R}} \sim p_{data}(P_{\mathcal{R}})}[\log(1 - D_{patch}(P_{\mathcal{R}})], \end{aligned} \tag{9}$$

where $P_{\mathcal{R}}$ and P_R denote patches from $\mathcal{R}$ and R, respectively. Note that a similar discriminator is also utilized in Hu et al. (2019), Le and Samaras (2020), but their goal is to recover illumination for shadow regions. In comparison, we use a histogram-based discriminator to recover scene-consistent illumination for shadow regions, while our patch-based discriminator designed on small patches aims to recover the fine-scale texture details. Figure 3 verifies the effectiveness of our patch-based discriminator D_{patch}.

3.6 Loss Function

The overall loss function for our Self-ShadowGAN is a weighted sum of the smoothness loss $\mathcal{L}_{smooth}$ in Eq. (4), and the adversarial losses for the two discriminators (*i.e.*, $\mathcal{L}_{GAN}^{hist}$ in Eq. (5) and $\mathcal{L}_{GAN}^{patch}$ in Eq. (9)), which is expressed as

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{smooth} + \lambda_2 \mathcal{L}_{GAN}^{hist} + \lambda_3 \mathcal{L}_{GAN}^{patch}, \tag{10}$$

where λ_1, λ_2, λ_3 are the weights, which are empirically set as $\lambda_1 = 1$, $\lambda_2 = 0.01$, $\lambda_3 = 0.02$.

3.7 Implementation Details

We build our Self-ShadowGAN on PyTorch and train it on an NVIDIA GeForce RTX 3090Ti GPU using the shadow mask produced by the method of Zhu et al. (2018). All parameters in our network (including the shadow relighting network and the two discriminators) are initialized by random noise following a zero-mean Gaussian distribution with standard deviation set as 0.02. Our network is optimized using the Adam optimizer with the first and second momentum values set as 0.5 and 0.999. The initial learning rate for the generator and the discriminators are 2×10^{-4}, which will be linearly decayed after 600 iterations until reduced to zeros. For the low-resolution input utilized in the shadow relighting network, it is created by bilinear downsampling while constraining it has at most 256 pixels on the short axis of the image. Considering that the shadow mask may fail to accurately locate the soft shadow boundaries in the penumbra area or detect shadows on complex background, the mask will be dilated before fed to the shadow relighting network to achieve better shadow removal effect. The reason for this operation is that our histogram-based discriminator can help lower our method's sensitivity to inaccurate masks, as we will demonstrate in Fig. 10.

In general, 1000 iterations are sufficient for our model to produce good results, which typically takes about 1 min for training on an image with 1 M pixels. Figure 5 shows how our results change with different number of training iterations. Although the inference performance of our method is inferior to existing shadow removal methods trained on large-scale datasets, it is about 10× faster than recent frameworks that are also trained on a single image (Ulyanov et al., 2018; Shaham et al., 2019; Gandelsman et al., 2019).

4 Experiments

Benchmark datasets We employ three benchmark datasets to evaluate the shadow removal performance of our approach. The first one is the ISTD+ dataset (Le & Samaras, 2019),

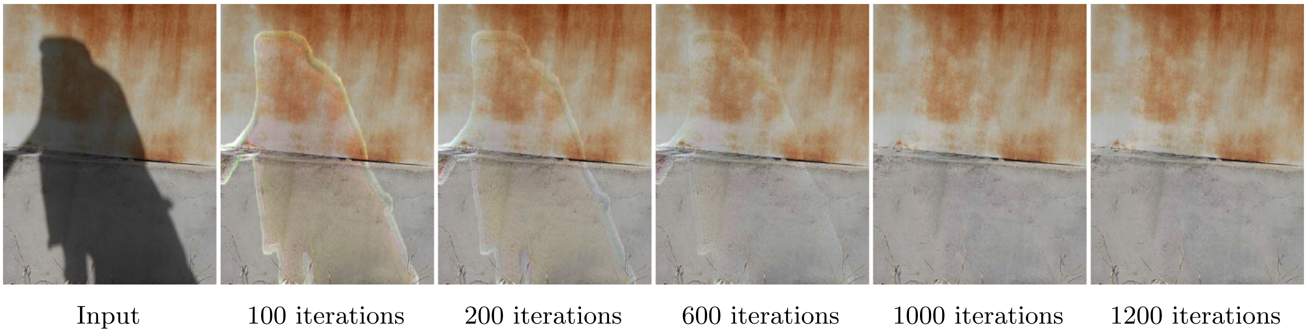

Fig. 5 Effect of different number of training iterations on the shadow removal result of an input image. As shown, our method produces good result at 1000 iterations, and more iterations (*e.g.*, result at 1200 iterations) lead to visually indistinguishable results

which is an adjusted version of the ISTD dataset (Wang et al., 2018) eliminating the underlying color inconsistency between the paired shadow and shadow-free images in ISTD. It contains 1870 triplets of shadow image, shadow mask, and the shadow-free image, where 540 triplets are split for testing. Note that we did not evaluate our method on the original ISTD dataset because the color inconsistency issue hinders reliable performance evaluation of our method, as well as other methods that are also trained in absence of the shadow-free ground truths in ISTD. The second dataset is the SRD dataset (Qu et al., 2017), which includes 2680 image pairs for training, and 408 image pairs for testing. For this dataset, we follow Fu et al. (2021) to use the shadow masks produced by Khan et al. (2015) for evaluation. The third one is the SBU dataset (Vicente et al., 2016). It includes 638 test images with shadow masks, but does not provide ground-truth shadow-free images.

Evaluation metrics We follow previous works (Le & Samaras, 2020; Fu et al., 2021) to resize all shadow removal results to 256 × 256 to facilitate comparison, and use the root mean square error (RMSE) in LAB color space on the shadow area, non-shadow area, and the entire image, as well as the PSNR and LPIPS (Zhang et al., 2018) in RGB color space, to evaluate the performance. In general, lower RMSE/LPIPS and higher PSNR values indicate better results.

4.1 Comparison with the State-of-the-Art Methods

We compare our method trained on a single image with twelve recent learning-based shadow removal methods. Based on the required training data, these compared methods can be classified into the following three groups:

- **Methods trained on paired shadow and shadow-free images** Guo et al. (2012), ST-CGAN (Wang et al., 2018), DSC (Hu et al., 2019), SP+M-Net (Le & Samaras, 2019), DHAN (Cun et al., 2020), Auto-Exposure (Fu et al., 2021);
- **Methods trained on unpaired shadow and shadow-free images** Mask-ShadowGAN (Hu et al., 2019), LG-ShadowNet (Liu et al., 2021), and DC-ShadowNet (Jin et al., 2021);
- **Methods trained on a set of shadow images with masks** Param+M+D-Net (Le & Samaras, 2020), G2R-ShadowNet (Liu et al., 2021), and wSP+M-Net (Le & Samaras, 2021).

Note, for fair comparison, we produce results of the compared methods using publicly-available implementation or trained models provided by the authors with recommended parameter setting. Following Fu et al. (2021), the quantitative results on the ISTD+ dataset of Le and Samaras (2020) are as reported in their paper.

Evaluation on ISTD+ Tables 1 and 2 give the quantitative comparison results on the ISTD+ dataset. As shown, in addition to Auto-exposure (Fu et al., 2021) which has slightly better performance on shadow area, our method outperforms other compared methods on all the three metrics, even some of them are directly trained on ISTD+. Figure 6 shows visual comparisons, where we can see that our method is effective to handle shadows cast over highly textured background and shadows with inhomogeneous luminance, and is able to recover clear texture details and natural colors for the shadow regions.

Evaluation on SRD As shown in Tables 1 and 2, our method achieves the overall best numerical results on the SRD dataset. Visual comparison results on the SRD dataset are shown in Fig. 7. As can be seen, our method robustly produces high-quality shadow removal results without leading to visual artifacts such as shadow residuals and color distortions, that are often appeared in the results of other methods.

Evaluation on SBU As the SBU dataset does not contain ground-truth shadow-free images, unlike the ISTD+ and SRD datasets, we perform visual comparisons for some complex cases from the SBU dataset. Comparing the visual results in Fig. 8, we notice two improvements of our method over the others. First, our method can effectively remove

Table 1 Comparison with the state-of-the-art shadow removal methods on the ISTD+ and SRD datasets in terms of RMSE (↑)

Method	ISTD+			SRD		
	Shadow	Non-shadow	All	Shadow	Non-shadow	All
Guo et al. (2012)	22.0	3.1	6.1	29.9	6.5	12.6
ST-CGAN (Wang et al., 2018)	13.4	7.7	8.7	16.5	14.6	15.5
DSC (Hu et al., 2019)	7.5	3.0	3.8	10.7	4.8	6.2
SP+M-Net (Le & Samaras, 2019)	8.1	2.8	3.6	12.8	5.4	7.9
DHAN (Cun et al., 2020)	11.2	7.1	7.7	8.9	4.8	5.6
Auto-Exposure (Fu et al., 2021)	**6.7**	3.8	4.2	8.6	5.8	6.6
Mask-ShadowGAN (Hu et al., 2019)	12.4	4.0	5.3	11.4	4.2	7.0
LG-ShadowNet (Liu et al., 2021)	9.9	3.4	4.4	15.8	5.6	9.5
DC-ShadowNet (Jin et al., 2021)	10.3	3.5	4.6	**8.0**	3.4	4.9
Param+M+D-Net (Le & Samaras, 2020)	9.7	3.0	4.0	15.1	5.5	9.2
G2R-ShadowNet (Liu et al., 2021)	8.9	2.9	3.9	13.2	5.5	8.1
wSP+M-Net (Le & Samaras, 2021)	9.1	2.6	3.6	12.0	5.0	6.9
Ours using Eq. (1)	10.9	2.7	4.5	13.3	3.3	6.5
Ours w/o $\mathcal{L}_{smooth}$	7.9	2.5	4.2	10.4	3.4	5.8
Ours w/o D_{hist}	10.6	2.6	4.4	12.4	3.5	6.2
Ours w/o D_{patch}	8.3	2.7	3.7	9.7	3.4	5.5
Ours w/o speedup	7.4	**2.5**	**3.4**	**8.0**	**3.2**	**4.7**
Ours (full method)	7.5	**2.5**	3.5	8.1	**3.2**	4.8

Best results are in bold

Table 2 Comparison with the state-of-the-art shadow removal methods on the ISTD+ and SRD datasets in terms of PSNR (↑) and LPIPS (↓)

Method	ISTD+ PSNR/LPIPS	SRD PSNR/LPIPS
Guo et al. (2012)	20.26/0.231	18.50/0.247
ST-CGAN (Wang et al., 2018)	24.52/0.157	19.34/0.210
DSC (Hu et al., 2019)	32.50/0.073	29.50/0.114
SP+M-Net (Le & Samaras, 2019)	33.93/0.061	24.96/0.229
DHAN (Cun et al., 2020)	25.78/0.057	31.24/**0.078**
Auto-Exposure (Fu et al., 2021)	29.28/0.187	27.87/0.191
Mask-ShadowGAN (Hu et al., 2019)	24.38/0119	24.66/0.159
LG-ShadowGAN (Liu et al., 2021)	29.45/0.093	21.95/0.180
DC-ShadowGAN (Jin et al., 2021)	27.36/0.172	28.00/0.160
Param+M+D-Net (Le & Samaras, 2020)	30.43/0.076	22.95/0.176
G2R-ShadowNet (Liu et al., 2021)	30.86/0.087	23.84/0.172
wSP+M-Net (Le & Samaras, 2021)	33.67/0.058	25.45/0.225
Ours	**34.14/0.054**	**31.93**/0.095

Best results are in bold

large-scale dark shadows casted on the cliff with varying depths, without noticeable shadow residuals (see the third row in Fig. 8). Second, we are able to produce shadow removal results with clear texture details and natural colors for shadows on complex backgrounds (see the first, second, and fourth rows in Fig. 8).

4.2 More Analysis

Ablation study Besides Fig. 3, we quantitatively evaluate the effectiveness of our pixel adaptive shadow relighting model, the smoothness loss $\mathcal{L}_{smooth}$, the histogram-based discriminator D_{hist}, and the patch-based discriminator D_{patch} in

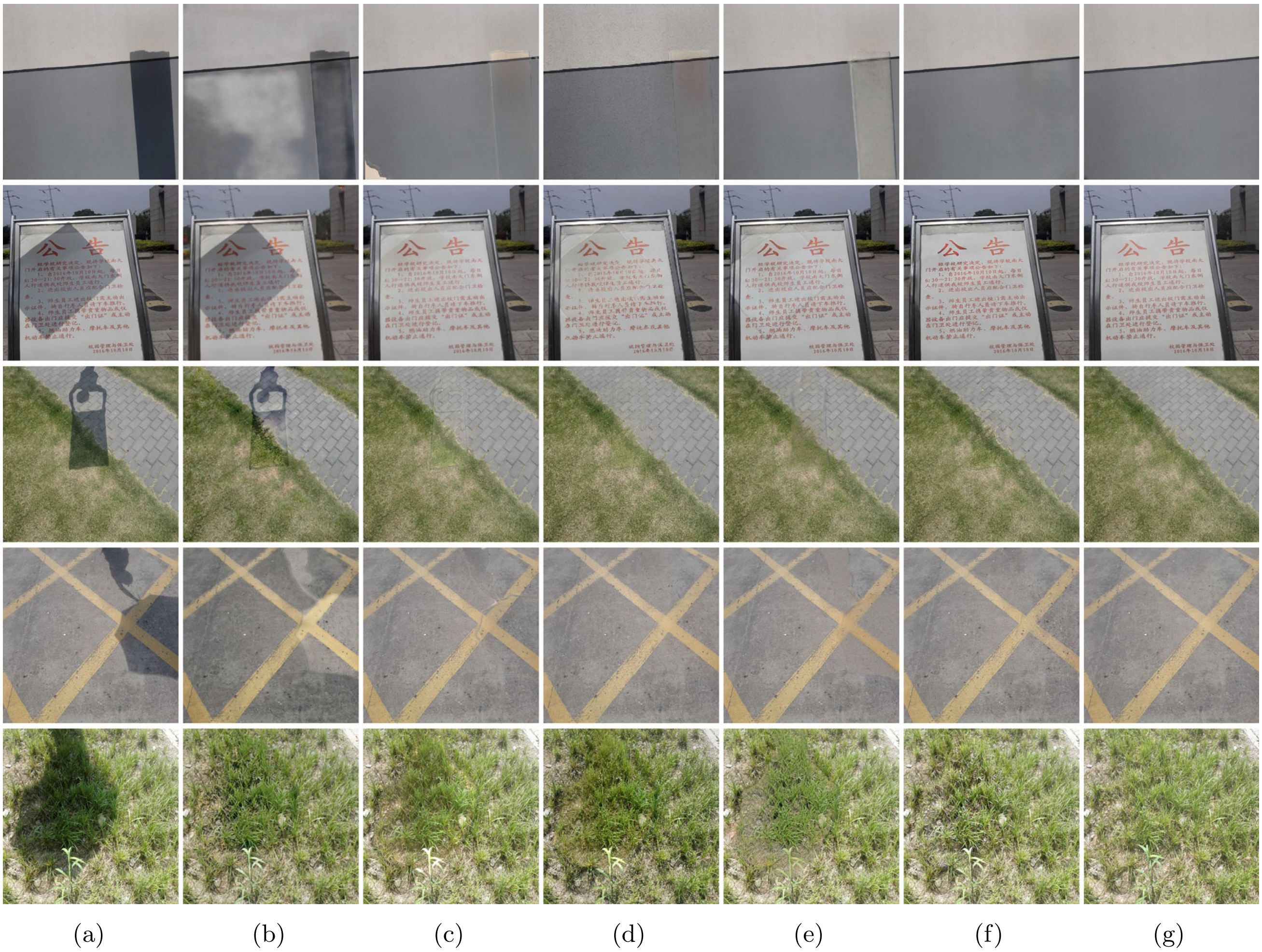

Fig. 6 Comparison with the state-of-the-art shadow removal methods on testing images from the ISTD+ dataset. **a** Input. **b**–**e** are results of Mask-ShadowGAN (Hu et al., 2019), SP+M-Net (Le & Samaras, 2019), Auto-Exposure (Fu et al., 2021), and G2R-ShadowNet (Liu et al., 2021). **f** Our result. **g** Ground truth

Table 1. By comparing the numerical results in the last row (ours) and the third to sixth row from the bottom, we observe clear performance improvement by adopting the above components. We also evaluate the necessity of the efficient relighting coefficients prediction described in Sect. 3.2 in Table 1 and Fig. 12. As shown, the shadow removal results produced by our method with and without the efficient prediction strategy are numerically very close and visually indistinguishable, while we achieve about 10× runtime speedup from the efficient prediction strategy.

Necessity of the shadow relighting model To verify the necessity of the shadow relighting model in Eq. (2), we replace the entire shadow relighting network with Pix2pix (Isola et al., 2017), which is a common image-to-image translation model. Figure 9 compares the results produced by our full method and the variant using Pix2pix as generator. As can be seen, due to lack of constraint from the shadow relighting model, the variant with Pix2pix recovers mismatched content for the shadow regions, while our method with the shadow relighting network produce high-quality results where the details attenuated by shadows are faithfully restored.

Effect of inaccurate shadow mask Figure 10 examines the effect of inaccurate mask on our shadow removal results. As shown, although the final dilated masks of the two images fail to accurately locate the shadows, our method still produces satisfying results, showing that our method is somewhat tolerance of inaccurate mask. This ability of our method comes from the histogram-based discriminator, because as long as the identified non-shadow regions are sufficiently reliable, the shadow regions can also be effectively processed by histogram matching even if some non-shadow regions are mistakenly treated as shadows. Figure 11 further compares the shadow removal results produced with GT shadow mask and detected mask. As shown, although the detected mask generated by Hu et al. (2018) is not as good as the GT mask,

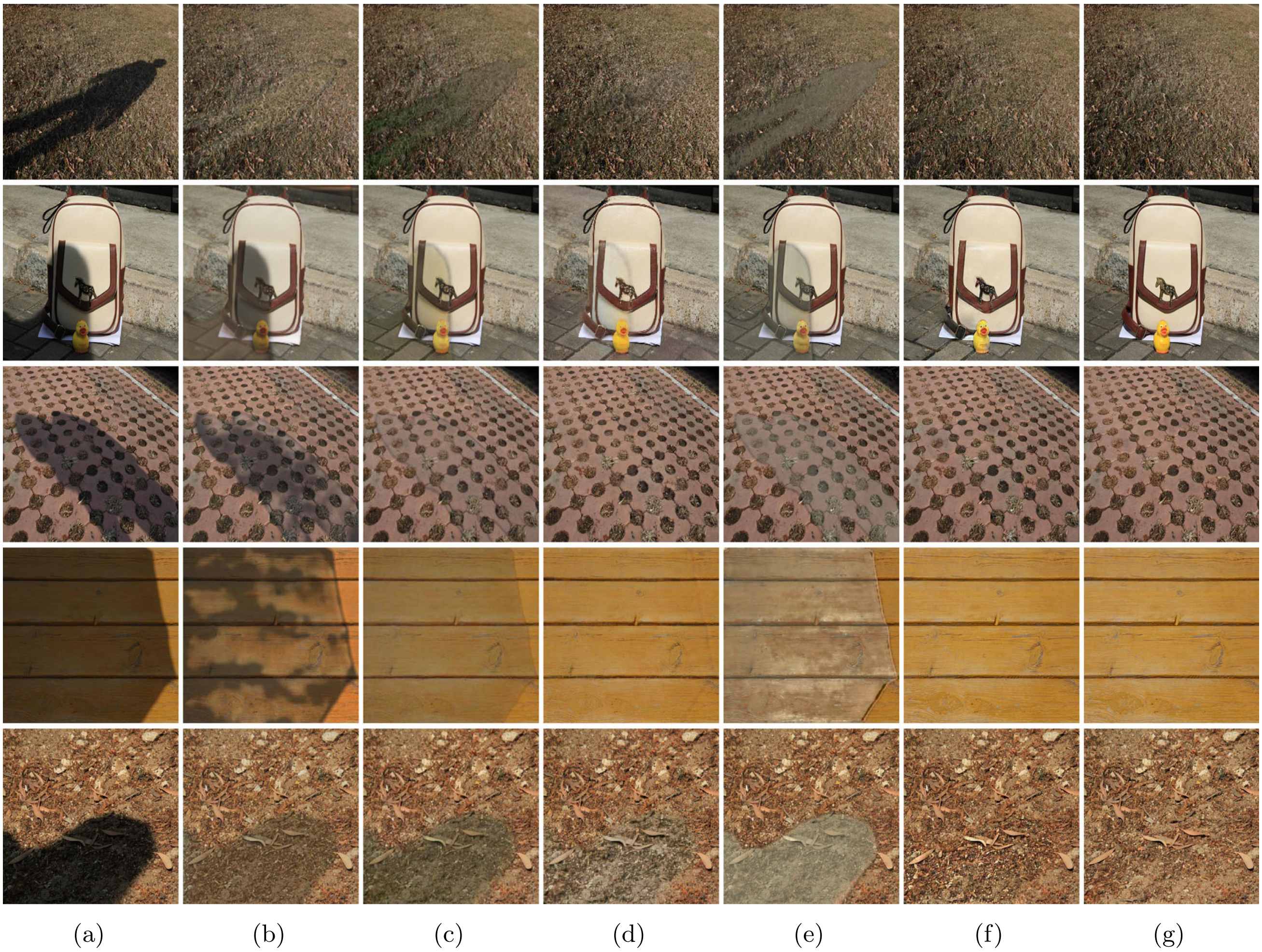

Fig. 7 Comparison with the state-of-the-art shadow removal methods on testing images from the SRD dataset. **a** Input. **b–e** are results of Mask-ShadowGAN (Hu et al., 2019), SP+M-Net (Le & Samaras, 2019), Auto-Exposure (Fu et al., 2021), and G2R-ShadowNet (Liu et al., 2021). **f** Our result. **g** Ground truth

the results produced from the detected mask and the GT mask are visually indistinguishable (Fig. 12).

Effect of varying patch sizes in D_{patch} Figure 13 explores how different patch sizes in our patch-based discriminator D_{patch} affect the performance of our method. As shown, we obtain shadow removal result with clear textures and natural illumination by using a relatively small patch size of 32 × 32, while larger patch sizes induce worse results because of lacking sufficient patches for model training.

Effect of single-channel histogram In Fig. 14, we conduct experiments to demonstrate the superiority of using three-channel RGB histogram over single-channel intensity histogram based on the Y channel in the YUV color space. By comparing the visual results, we can see that training with three-channel RGB histogram benefits recovering more natural illumination and color for the shadow regions, while single-channel intensity histogram may result in color distortion despite its effectiveness in illumination recovery.

Effect of different number of histogram bins We in Fig. 16 analyze the effect of different number of histogram bins on the shadow removal performance. As shown, 32 bins lead to result with weak shadow residuals while more bins ($\geq$ 64) produce good results that are visually indistinguishable from each other. To balance the shadow removal performance and the computational efficiency, we choose to construct histogram with 64 bins.

Necessity of single image training To verify the necessity of single image training for our network, we compare shadow removal results produced by our method trained on a single image and the SRD dataset in Fig. 15. As shown, our method based on single image training achieves better results. The reasons are explained as follows. First, by training on a single test image itself, we are able to avoid the domain gap issue encountered by training on the SRD dataset, and accordingly enhance the method's effectiveness in different types of shadow images. On the other hand, our network is designed

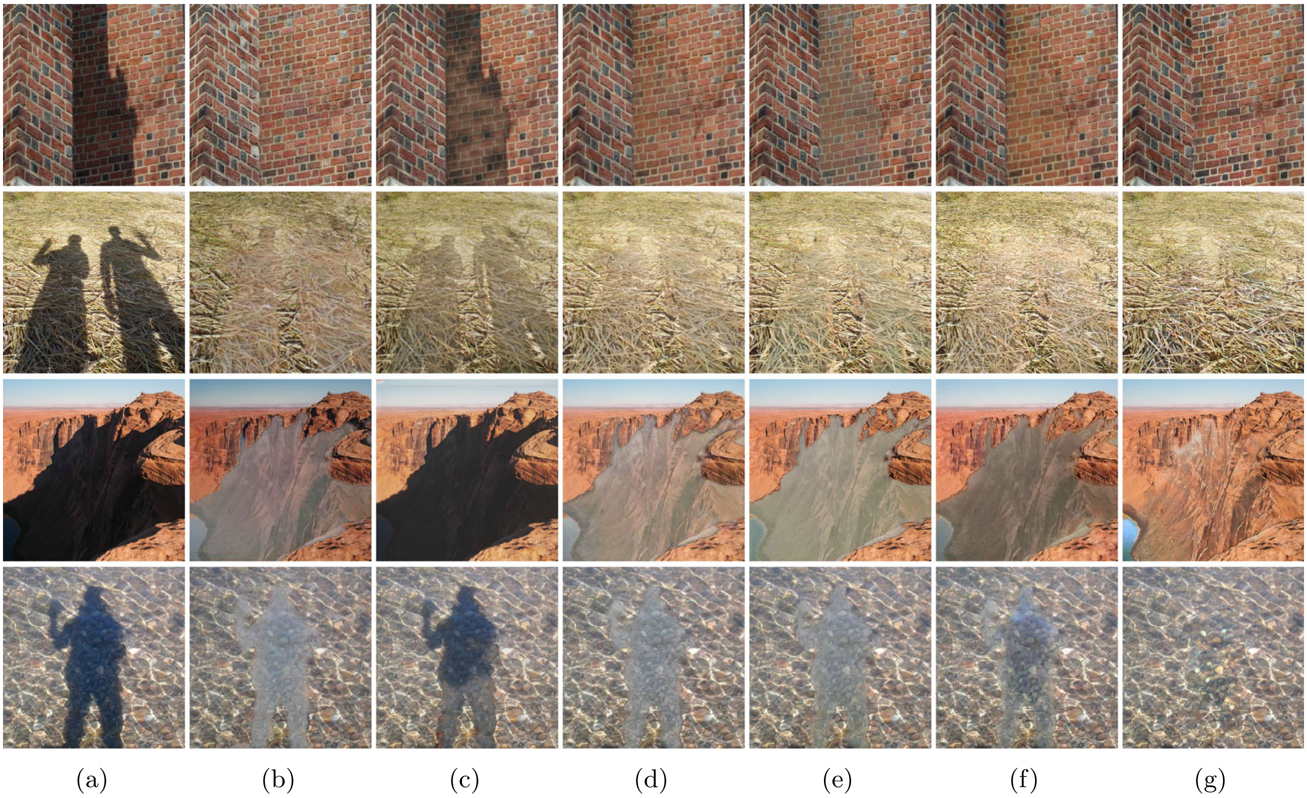

Fig. 8 Comparison with the state-of-the-art shadow removal methods on images from the SBU dataset. **a** Input. **b–f** are results of ST-CGAN (Wang et al., 2018), Mask-ShadowGAN (Hu et al., 2019), SP+M-Net (Le & Samaras, 2019), G2R-ShadowNet (Liu et al., 2021), and Auto-Exposure (Fu et al., 2021). **g** Our result

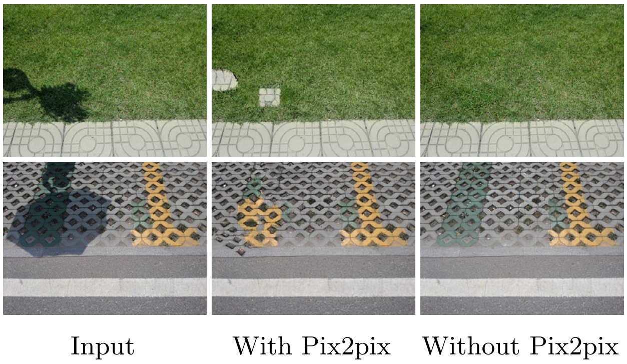

Fig. 9 Effect of the shadow relighting model in our method. Note, "Without Pix2pix" refers to our method with the shadow relighting network

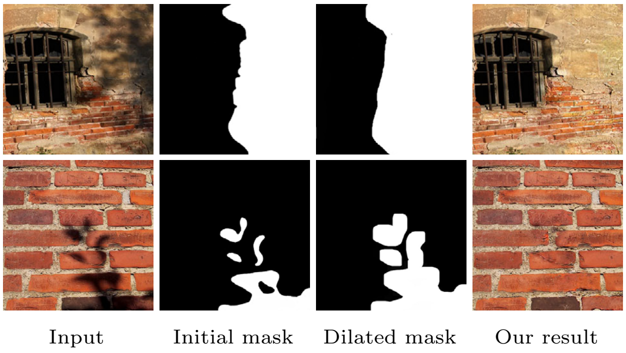

Fig. 10 Effect of inaccurate mask on our shadow removal results. As illustrated in Sect. 3.7, our initial shadow mask is automatically generated by Zhu et al. (2018), a dilation operation is then applied to the initial mask before fed it to our network

as a low-capacity lightweight model that is actually not suitable for training on large-scale datasets (Fig. 16).

Comparison to pretrained models with single image adaptation Adapting models trained on large-scale datasets to a single image is a common way to improve the performance on a specific image. Therefore, we compare our method to pretrained models with single image adaptation based on our histogram-based discriminator (*i.e.*, fine-tuing the pretrained models using the loss in Eq. (5) on a single image). As shown in Fig. 17, the single image adaptation operation applied to two recent pretrained models still fails to produce results as good as that of our method, manifesting that our method can not be replaced by single image adaptation of pretrained models.

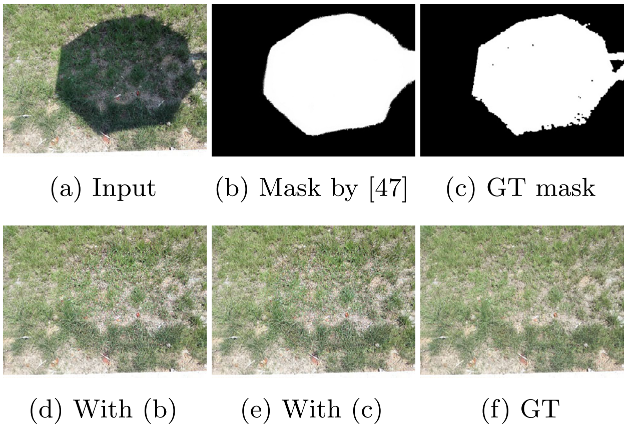

Fig. 11 Comparison of our shadow removal results produced with detected shadow mask and GT (ground truth) shadow mask

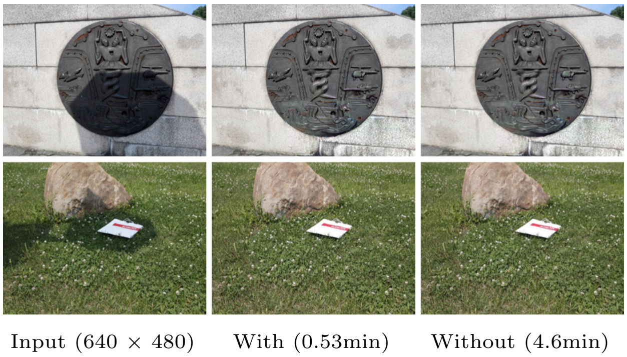

Fig. 12 Comparison of our shadow removal results and the required training time costs with and without using the efficient relighting coefficients prediction strategy on two images with the resolution of 640 × 480

Comparison to other single image based learning frameworks Training with single image has been explored by some previous works including DIP (Ulyanov et al., 2018), DoubleDIP (Gandelsman et al., 2019) and SinGAN (Shaham et al., 2019). However, their key idea is to learn internal statistics of an image by training a network to reconstruct the image from random noise input, which has essential difference from our proposed method, as we directly take an image as input and build our method upon histogram matching instead of image internal statistics. As can be seen in Fig. 18, our method produces clearly better shadow removal results than the above three single-image-based training alternatives, while the compared methods generated unpredictable results with distorted appearance.

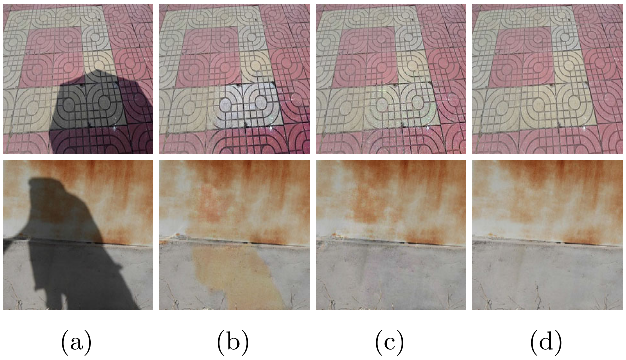

Fig. 14 Comparison of shadow removal results produced by our method using different histograms. **a** Input. **b** Result using single-channel intensity histogram. **c** Result using three-channel RGB histogram. **d** Ground truth

More results on complex scenes Figure 19 evaluates the shadow removal performance of our method on more complex scenes, including: (i) an image contains inhomogeneous shadows with irregular shape in the first row; (ii) an image that there exists weak texture similarity between non-shadow and shadow regions in the second row; (iii) an image that the non-shadow regions only occupy a very low percentage of the entire image in the third row. As can be seen, for all these challenging cases, our method clearly outperforms the compared methods, and is able to produce visually compelling results.

Results on different types of images We show in Figs. 20, 21 and 22 that our method can also effectively remove shadows in portrait, document, and remote sensing images, which are usually challenging for learning-based shadow removal methods targeting at natural scenes. As can be seen, our method produces good results, which are comparable or even better than the compared state-of-the-art shadow removal methods specifically designed for portrait (Zhang et al., 2020)

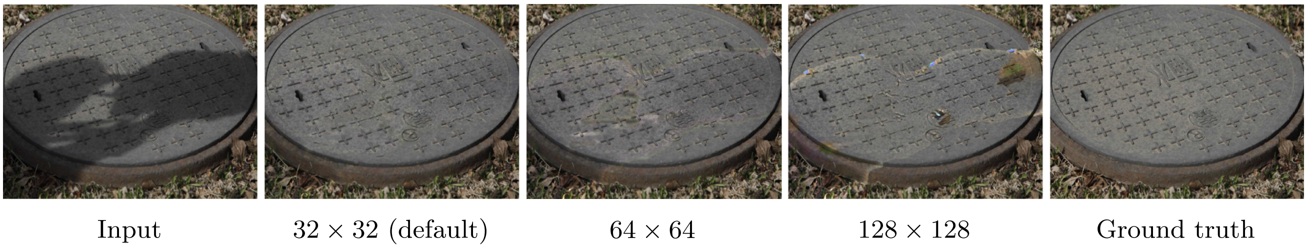

Fig. 13 Comparison of results produced with different patch sizes in the patch-based discriminator

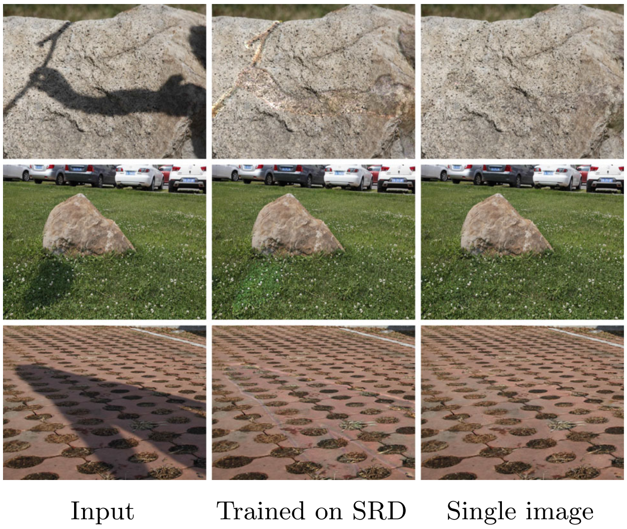

Fig. 15 Comparison of shadow removal results produced by our method trained on the SRD dataset and a single image. It is worth noting that, when an image set is employed for our network training, the trained generator (*i.e.*, the shadow relighting network) will be used to produce shadow removal results for the test images. In contrast, when a single image is adopted for network training, the shadow removal result of the image is obtained when the entire adversarial training finished

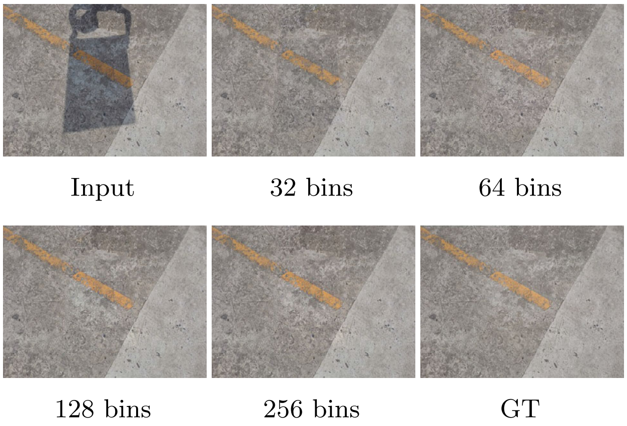

Fig. 16 Effect of different number of histogram bins in the histogram-based discriminator

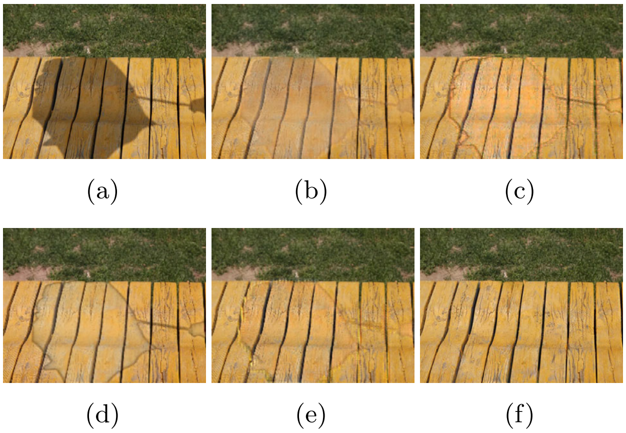

Fig. 17 Comparison between our method and models trained on the ISTD+ dataset with single image adaptation. **a** Input. **b** and **c** are results of SP+M-Net (Le & Samaras, 2019) without/with the adaptation. **d** and **e** are results of Auto-Exposure (Fu et al., 2021) without/with the adaptation. **f** Ours

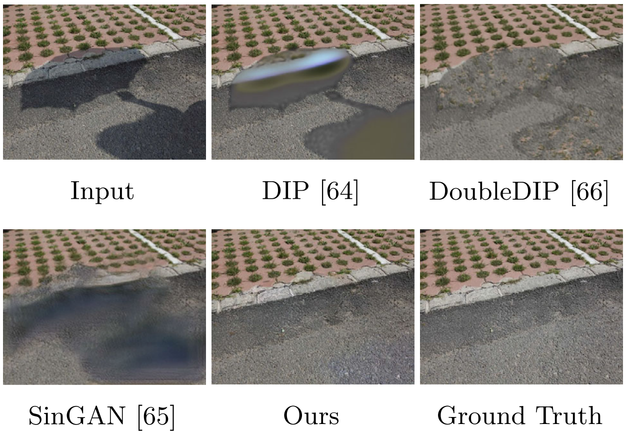

Fig. 18 Comparison of shadow removal results produced by other single image based learning frameworks including DIP (Ulyanov et al., 2018), DoubleDIP (Gandelsman et al., 2019), and SinGAN (Shaham et al., 2019)

and document (Lin et al., 2020) images, showing that our method works well for different types of shadow images.

Comparison on inference time Table 3 compares the inference time required by our method and other recent learning-based shadow removal methods on an image with the resolution of 1024 × 1024. As can be seen, as involving test-time training our inference time is much higher than that of other methods using pretrained models for testing, but the advantage is that our approach does not require the time-consuming pre-training and takes only a single shadow image as training input.

Limitations and future works Fig. 23 presents two examples where our method, as well as other state-of-the-arts (*e.g.*, Auto-Exposure (Fu et al., 2021)), all fail to produce visually compelling results. For the top image, we fail to recover visually natural and consistent texture for the dark hard shadows that barely contains visual information, while for the bottom input, our method does not completely remove the shadows around the dark branches and lead to noticeable color noises. Therefore, enhancing the effectiveness of our method in these hard cases will be our future goal. In addition, we are also interested in removing the requirement of shadow mask and accelerating our single image based model training to real-time performance. Another promising future work is to extend our method to enhancing underexposed photos (Wang et al., 2019; Zhang et al., 2019, 2020).

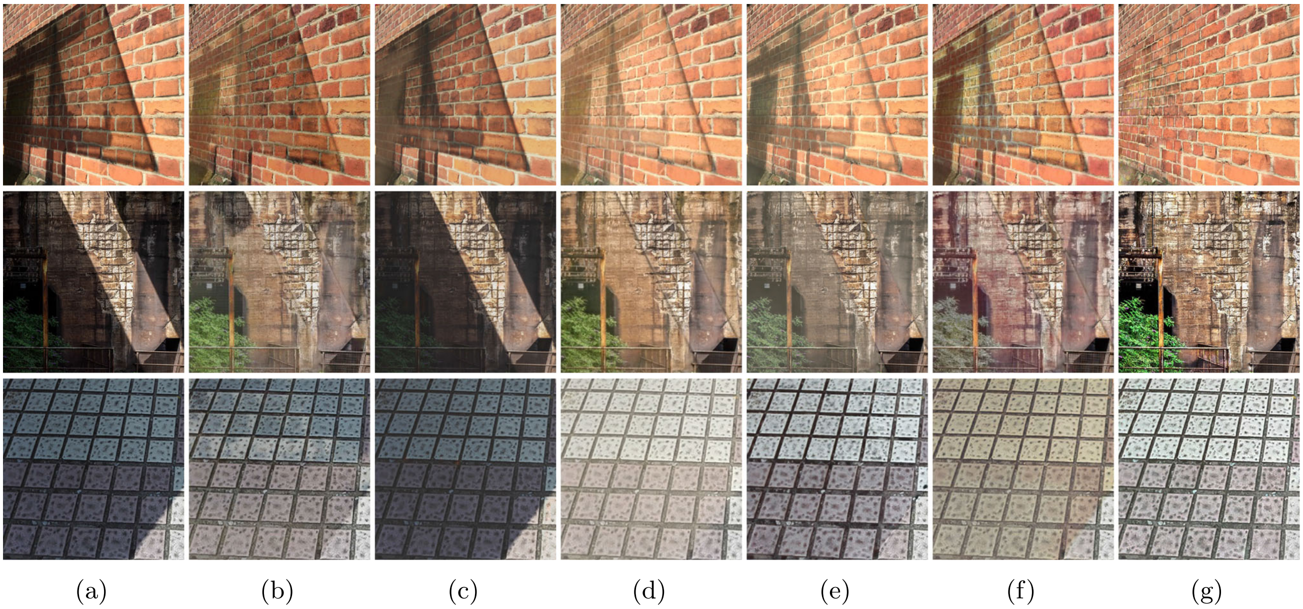

Fig. 19 More comparison with the state-of-the-art shadow removal methods on some complex scenes. **a** Input. **b–f** are results of ST-CGAN (Wang et al., 2018), Mask-ShadowGAN (Hu et al., 2019), SP+M-Net (Le & Samaras, 2019), G2R-ShadowNet (Liu et al., 2021), and Auto-Exposure (Fu et al., 2021). **g** Our result

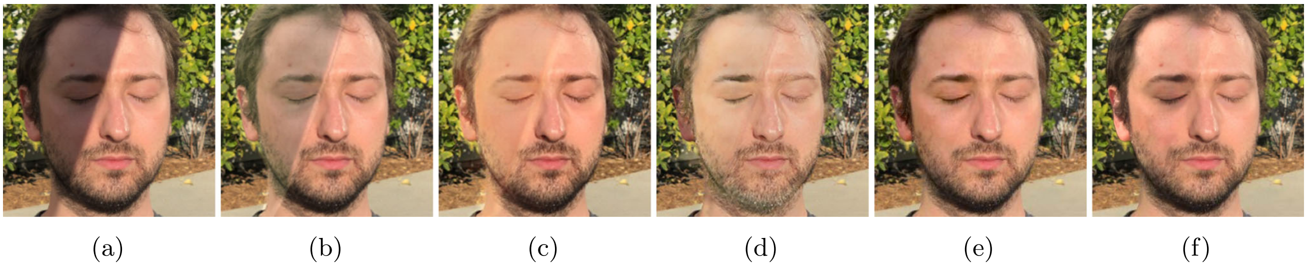

Fig. 20 Comparison with the state-of-the-art methods on a portrait shadow image (**a**). **b–d** are results of G2R-ShadowNet (Liu et al., 2021), Auto-Exposure (Fu et al., 2021), and DC-ShaodowNet (Jin et al., 2021). **e** Result of (Zhang et al., 2020), which is specially designed for portrait shadow removal. **f** Our result

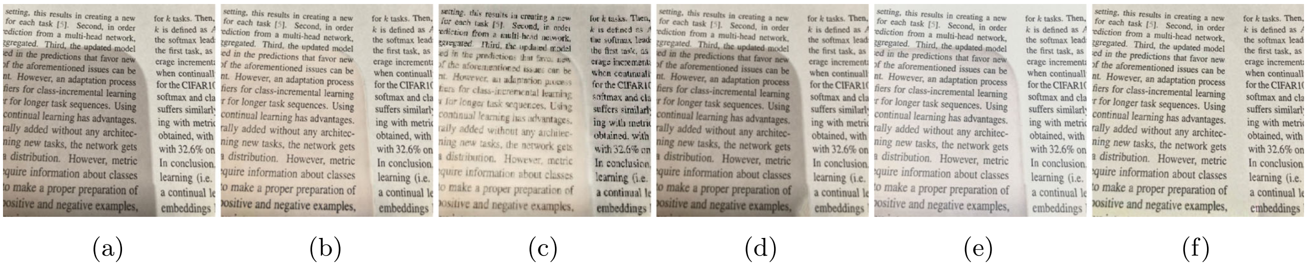

Fig. 21 Comparison with the state-of-the-art methods on a document shadow image (**a**). **b–d** are results of G2R-ShadowNet (Liu et al., 2021), Auto-Exposure (Fu et al., 2021), and DC-ShaodowNet (Jin et al., 2021). **e** Result of BEDSR-Net (Lin et al., 2020), which is specially designed for shadow removal of document image. **f** Our result

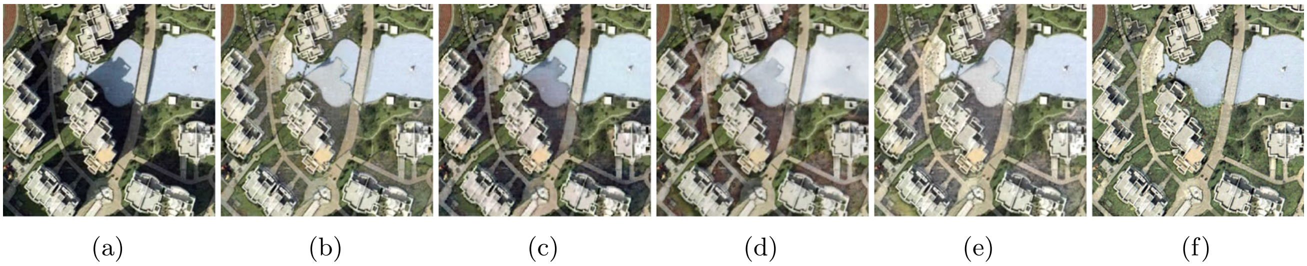

Fig. 22 Comparison with the state-of-the-art methods on a remote sensing shadow image (**a**). **b**–**e** are results of G2R-ShadowNet (Liu et al., 2021), Auto-Exposure (Fu et al., 2021), DC-ShaodowNet (Jin et al., 2021), and wSP+M-Net (Le & Samaras, 2021). **f** Our result

Table 3 Comparison with the state-of-the-art shadow removal methods on the inference time

Method	Pre-training	Inference
ST-CGAN (Wang et al., 2018)	> 12 h	0.97 s
DSC (Hu et al., 2019)		0.71 s
SP+M-Net (Le & Samaras, 2019)		0.68 s
DHAN (Cun et al., 2020)		0.88 s
Auto-Exposure (Fu et al., 2021)		0.56 s
Mask-ShadowGAN (Hu et al., 2019)		0.93 s
DC-ShadowNet (Jin et al., 2021)		0.86 s
G2R-ShadowNet (Liu et al., 2021)		0.78 s
wSP+M-Net (Le & Samaras, 2021)		0.70 s
Ours (single-image based)	–	1 min

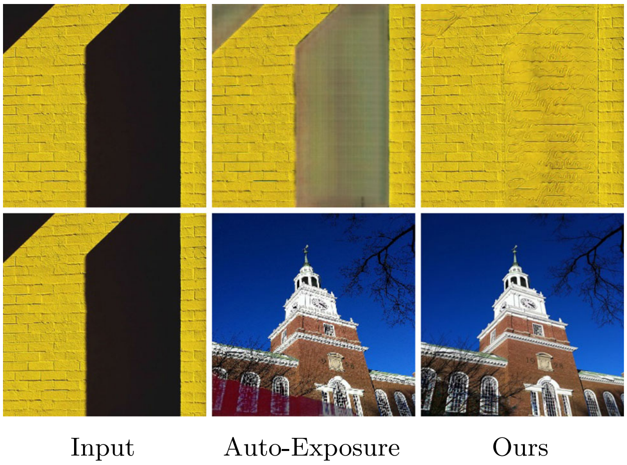

Fig. 23 Failure cases of our method

5 Conclusion

We have presented a novel adversarial shadow removal framework that can be trained on a single image. Our key idea is to transfer the illumination of shadow-free regions in an image to the shadow regions based on histogram matching. To do so, a pixel adaptive shadow relighting model is firstly introduced, with which we build a lightweight shadow relighting network for shadow removal. Next, a histogram-based discriminator is designed to ensure the illumination consistency by enforcing that there are similar histograms between the deshadowed regions and the original shadow-free regions, and a patch-based discriminator is introduced for texture recovery. Extensive experiments validate the effectiveness of our approach.

Acknowledgements This work was supported by the National Natural Science Foundation of China (U21A20471, 62072191), Guangdong Basic and Applied Basic Research Foundation (2023A1515030002, 2023B1515040025), and Guangzhou Municipal Science and Technology Project (Grant No. 2023A03J0671).

References

Arbel, E., & Hel-Or, H. (2010). Shadow removal using intensity surfaces and texture anchor points. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 33*(6), 1202–1216.

Avi-Aharon, M., Arbelle, A., & Raviv, T. R. (2020). *DeepHist: Differentiable joint and color histogram layers for image-to-image translation*. arXiv preprint arXiv:2005.03995

Chen, Z., Long, C., Zhang, L., & Xiao, C. (2021). CANet: A context-aware network for shadow removal. In *Proceedings of the international conference on computer vision* (pp. 4743–4752).

Cun, X., Pun, C.-M., & Shi, C. (2020). Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting gan. In *Proceedings of the association for the advancement of artificial intelligence* (vol. 34, pp. 10680–10687).

Ding, B., Long, C., Zhang, L., & Xiao, C. (2019). ARGAN: Attentive recurrent generative adversarial network for shadow detection and removal. In *Proceedings of the international conference on computer vision* (pp. 10213–10222).

Drew, M. S., Finlayson, G. D., & Hordley, S. D. (2003). Recovery of chromaticity image free from shadows via illumination invariance. In *Proceedings of The IEEE international conference on computer vision workshops* (pp. 32–39).

Finlayson, G. D., & Drew, M. S. (2001). 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).

Finlayson, G. D., Drew, M. S., & Lu, C. (2009). Entropy minimization for shadow removal. *International Journal of Computer Vision, 85*(1), 35–57.

Finlayson, G. D., Hordley, S. D., & Drew, M. S. (2002). Removing shadows from images. In *Proceedings of the European conference on computer vision* (pp. 823–836).

Finlayson, G. D., Hordley, S. D., Lu, C., & Drew, M. S. (2005). On the removal of shadows from images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28*(1), 59–68.

Fredembach, C., & Finlayson, G. (2005). Hamiltonian path-based shadow removal. In *The British machine vision conference* (vol. 2, pp. 502–511).

Fu, L., Zhou, C., Guo, Q., Juefei-Xu, F., Yu, H., Feng, W., Liu, Y., & Wang, S. (2021). Auto-exposure fusion for single-image shadow removal. In *Proceedings of the IEEE computer vision and pattern recognition* (pp. 10571–10580).

Gandelsman, Y., Shocher, A., & Irani, M. (2019). Double-DIP: Unsupervised image decomposition via coupled deep-image-priors. In *Proceedings of the IEEE computer vision and pattern recognition* (pp. 11026–11035).

Gharbi, M., Chen, J., Barron, J. T., Hasinoff, S. W., & Durand, F. (2017). Deep bilateral learning for real-time image enhancement. *ACM Transactions on Graphics, 36*(4), 1–12.

Gong, H., & Cosker, D. (2014). Interactive shadow removal and ground truth for variable scene categories. In *The British machine vision conference* (pp. 1–11).

Gryka, M., Terry, M., & Brostow, G. J. (2015). Learning to remove soft shadows. *ACM Transactions on Graphics, 34*(5), 1–15.

Guo, R., Dai, Q., & Hoiem, D. (2011). Single-image shadow detection and removal using paired regions. In *Proceedings of the IEEE computer vision and pattern recognition* (pp. 2033–2040).

Guo, R., Dai, Q., & Hoiem, D. (2012). Paired regions for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(12), 2956–2967.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. In *Proceedings of the international conference on computer vision* (pp. 2961–2969).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE computer vision and pattern recognition* (pp. 770–778).

He, Y., Xing, Y., Zhang, T., & Chen, Q. (2021). Unsupervised portrait shadow removal via generative priors. In *Proceedings of the ACM international conference on multimedia* (pp. 236–244).

Hu, X., Fu, C.-W., Zhu, L., Qin, J., & Heng, P.-A. (2019). Direction-aware spatial context features for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 42*(11), 2795–2808.

Hu, X., Jiang, Y., Fu, C.-W., & Heng, P.-A. (2019). Mask-ShadowGAN: Learning to remove shadows from unpaired data. In *Proceedings of the international conference on computer vision* (pp. 2472–2481).

Hu, X., Zhu, L., Fu, C.-W., Qin, J., & Heng, P.-A. (2018). Direction-aware spatial context features for shadow detection. In *Proceedings of the IEEE computer vision and pattern recognition* (pp. 7454–7462).

Inoue, N., & Yamasaki, T. (2020). Learning from synthetic shadows for shadow detection and removal. *IEEE Transactions on Circuits and Systems for Video Technology, 6*, 66.

Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1125–1134).

Jin, Y., Sharma, A., & Tan, R. T. (2021). DC-ShadowNet: Single-image hard and soft shadow removal using unsupervised domain-classifier guided network. In *Proceedings of the international conference on computer vision* (pp. 5027–5036).

Khan, S. H., Bennamoun, M., Sohel, F., & Togneri, R. (2015). Automatic shadow detection and removal from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 38*(3), 431–446.

Le, H., & Samaras, D. (2019). Shadow removal via shadow image decomposition. In *Proceedings of the international conference on computer vision* (pp. 8578–8587).

Le, H., & Samaras, D. (2020). From shadow segmentation to shadow removal. In *Proceedings of the European conference on computer vision* (pp. 264–281).

Le, H., & Samaras, D. (2021). Physics-based shadow image decomposition for shadow removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 01*, 1–1.

Levin, A., Lischinski, D., & Weiss, Y. (2007). A closed-form solution to natural image matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 30*(2), 228–242.

Li, Z., & Snavely, N. (2018). Learning intrinsic image decomposition from watching the world. In *Proceedings of the IEEE computer vision and pattern recognition* (pp. 9039–9048).

Lin, Y.-H., Chen, W.-C., & Chuang, Y.-Y. (2020). BEDSR-Net: A deep shadow removal network from a single document image. In *Proceedings of the IEEE computer vision and pattern recognition* (pp. 12905–12914).

Liu, A., Ginosar, S., Zhou, T., Efros, A. A., & Snavely, N. (2020). Learning to factorize and relight a city. In *Proceedings of the European conference on computer vision*.

Liu, F., & Gleicher, M. (2008). Texture-consistent shadow removal. In *Proceedings of the European conference on computer vision* (pp. 437–450).

Liu, Z., Yin, H., Mi, Y., Pu, M., & Wang, S. (2021). Shadow removal by a lightness-guided network with training on unpaired data. *IEEE Transactions on Image Processing, 30*, 1853–1865.

Liu, Z., Yin, H., Wu, X., Wu, Z., Mi, Y., & Wang, S. (2021). From shadow generation to shadow removal. In *Proceedings of the IEEE computer vision and pattern recognition* (pp. 4927–4936).

Ma, L.-Q., Wang, J., Shechtman, E., Sunkavalli, K., & Hu, S.-M. (2016). Appearance harmonization for single image. *Computer Graphics Forumshadow removal, 7*(35), 189–197.

Nestmeyer, T., Lalonde, J.-F., Matthews, I., & Lehrmann, A. (2020). Learning physics-guided face relighting under directional light. In *Proceedings of the IEEE computer vision and pattern recognition* (pp. 5124–5133) (2020)

Nguyen, V., Yago Vicente, T. F., Zhao, M., Hoai, M., & Samaras, D. (2017). Shadow detection with conditional generative adversarial networks. In *Proceedings of the international conference on computer vision* (pp. 4510–4518).

Qu, L., Tian, J., He, S., Tang, Y., & Lau, R. W. (2017). DeshadowNet: A multi-context embedding deep network for shadow removal. In *Proceedings of the IEEE computer vision and pattern recognition* (pp. 4067–4075).

Shaham, T. R., Dekel, T., & Michaeli, T. (2019). SinGAN: Learning a generative model from a single natural image. In *Proceedings of the international conference on computer vision* (pp. 4570–4580).

Shaham, T. R., Gharbi, M., Zhang, R., Shechtman, E., & Michaeli, T. (2021). Spatially-adaptive pixelwise networks for fast image translation. In *Proceedings of the IEEE computer vision and pattern recognition* (pp. 14882–14891).

Shor, Y., & Lischinski, D. (2008). The shadow meets the mask: Pyramid-based shadow removal. *Computer Graphics Forum, 27*(2), 577–586.

Ulyanov, D., Vedaldi, A., & Lempitsky, V. (2018). Deep image prior. In *Proceedings of the IEEE computer vision and pattern recognition* (pp. 9446–9454).

Vasluianu, F.-A., Romero, A., Van Gool, L., & Timofte, R. (2021). Shadow removal with paired and unpaired learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 826–835).

Vicente, T. F. Y., Hou, L., Yu, C.-P., Hoai, M., & Samaras, D. (2016). Large-scale training of shadow detectors with noisily-annotated shadow examples. In *Proceedings of the European conference on computer vision* (pp. 816–832).

Vicente, T. F. Y., & Samaras, D. (2014). Single image shadow removal via neighbor-based region relighting. In *proceedings of the European conference on computer vision* (pp. 309–320).

Wan, J., Yin, H., Wu, Z., Wu, X., Liu, Y., & Wang, S. (2022). Style-guided shadow removal. In *Proceedings of the European conference on computer vision* (pp. 361–378). Springer.

Wang, J., Agrawala, M., & Cohen, M. F. (2007). Soft scissors: An interactive tool for realtime high quality matting. *ACM Transactions on Graphics, 26*(3), 9.

Wang, J., Li, X., & Yang, J. (2018). Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *Proceedings of the IEEE computer vision and pattern recognition* (pp. 1788–1797).

Wang, R., Zhang, Q., Fu, C.-W., Shen, X., Zheng, W.-S., & Jia, J. (2019). Underexposed photo enhancement using deep illumination estimation. In *Proceedings of The IEEE conference on computer vision and pattern recognition* (pp. 6849–6857).

Wang, T., Hu, X., Wang, Q., Heng, P.-A., & Fu, C.-W. (2020). Instance shadow detection. In *Proceedings of the IEEE computer vision and pattern recognition* (pp. 1880–1889).

Wu, Q., Zhang, W., Kumar, B. V. (2012). Strong shadow removal via patch-based shadow edge detection. In *Proceedings of the IEEE international conference on robotics and automation* (pp. 2177–2182).

Wu, S., Makadia, A., Wu, J., Snavely, N., Tucker, R., & Kanazawa, A. (2021). De-rendering the world's revolutionary artefacts. In *Proceedings of the IEEE computer vision and pattern recognition* (pp. 6338–6347).

Wu, T.-P., & Tang, C.-K. (2005). A Bayesian approach for shadow extraction from a single image. In *Proceedings of the international conference on computer vision* (vol. 1, pp. 480–487).

Wu, T.-P., Tang, C.-K., Brown, M. S., & Shum, H.-Y. (2007). Natural shadow matting. *ACM Transactions on Graphics, 26*(2), 8.

Xiao, C., She, R., Xiao, D., & Ma, K.-L. (2013). Fast shadow removal using adaptive multi-scale illumination transfer. *Computer Graphics Forum, 32*(8), 207–218.

Xiao, C., Xiao, D., Zhang, L., & Chen, L. (2013). Efficient shadow removal using subregion matching illumination transfer. *Computer Graphics Forum, 32*(7), 421–430.

Xu, M., Zhu, J., Lv, P., Zhou, B., Tappen, M. F., & Ji, R. (2017). Learning-based shadow recognition and removal from monochromatic natural images. *IEEE Transactions on Image Processing, 26*(12), 5811–5824.

Yang, Q., Tan, K.-H., & Ahuja, N. (2012). Shadow removal using bilateral filtering. *IEEE Transactions on Image processing, 21*(10), 4361-4368.

Zeng, H., Cai, J., Li, L., Cao, Z., & Zhang, L. (2020). Learning image-adaptive 3d lookup tables for high performance photo enhancement in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 44*(4), 2058–2073.

Zhang, L., Long, C., Zhang, X., & Xiao, C. (2020). RIS-GAN: Explore residual and illumination with generative adversarial networks for shadow removal. In *Proceedings of the association for the advancement of artificial intelligence* (vol. 34, pp. 12829–12836).

Zhang, L., Zhang, Q., & Xiao, C. (2015). Shadow remover: Image shadow removal based on illumination recovering optimization. *IEEE Transactions on Image Processing, 24*(11), 4623–4636.

Zhang, Q., Nie, Y., & Zheng, W.-S. (2019). Dual illumination estimation for robust exposure correction. *Computer Graphics Forum, 38*(7), 243–252.

Zhang, Q., Nie, Y., Zhu, L., Xiao, C., & Zheng, W.-S. (2020). Enhancing underexposed photos using perceptually bidirectional similarity. *IEEE Transactions on Multimedia, 23*, 189–202.

Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 586–595).

Zhang, X., Barron, J. T., Tsai, Y.-T., Pandey, R., Zhang, X., Ng, R., & Jacobs, D. E. (2020). Portrait shadow manipulation. *ACM Transactions on Graphics, 39*(4), 78–1.

Zheng, Q., Qiao, X., Cao, Y., & Lau, R. W. (2019). Distraction-aware shadow detection. In *Proceedings of the IEEE computer vision and pattern recognition* (pp. 5167–5176) (2019)

Zhu, J.-Y., Park, T., Isola, P., Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the international conference on computer vision* (pp. 2223–2232).

Zhu, L., Deng, Z., Hu, X., Fu, C.-W., Xu, X., Qin, J., & Heng, P.-A. (2018). Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. In *Proceedings of the European conference on computer vision* (pp. 121–136).

Zhu, Y., Huang, J., Fu, X., Zhao, F., Sun, Q., & Zha, Z.-J. (2022). Bijective mapping network for shadow removal. In *Proceedings of the IEEE computer vision and pattern recognition* (pp. 5627–5636).

Zhu, Y., Xiao, Z., Fang, Y., Fu, X., Xiong, Z., & Zha, Z.-J. (2022). Efficient model-driven network for shadow removal. In *Proceedings of the AAAI conference on artificial intelligence*.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.