Auto-Exposure Fusion for Single-Image Shadow Removal

Lan Fu[1]*, Changqing Zhou[2]*, Qing Guo[2]†, Felix Juefei-Xu[3], Hongkai Yu[4],
Wei Feng[5], Yang Liu[2], Song Wang[1]

[1]University of South Carolina, USA, [2]Nanyang Technological University, Singapore
[3]Alibaba Group, USA, [4]Cleveland State University, USA, [5]Tianjin University, China

Abstract

Shadow removal is still a challenging task due to its inherent background-dependent[1] *and spatial-variant properties, leading to unknown and diverse shadow patterns. Even powerful deep neural networks could hardly recover traceless shadow-removed background. This paper proposes a new solution for this task by formulating it as an exposure fusion problem to address the challenges. Intuitively, we first estimate multiple over-exposure images w.r.t. the input image to let the shadow regions in these images have the same color with shadow-free areas in the input image. Then, we fuse the original input with the over-exposure images to generate the final shadow-free counterpart. Nevertheless, the spatial-variant property of the shadow requires the fusion to be sufficiently 'smart', that is, it should automatically select proper over-exposure pixels from different images to make the final output natural. To address this challenge, we propose the* **shadow-aware FusionNet** *that takes the shadow image as input to generate fusion weight maps across all the over-exposure images. Moreover, we propose the* **boundary-aware RefineNet** *to eliminate the remaining shadow trace further. We conduct extensive experiments on the ISTD, ISTD+, and SRD datasets to validate our method's effectiveness and show better performance in shadow regions and comparable performance in non-shadow regions over the state-of-the-art methods. We release the code in* https://github.com/tsingqguo/exposure-fusion-shadow-removal.

1. Introduction

Shadows are present in most natural images where the light source is blocked. Spatial-variant color and illumination distortion presented in the shadow region can hinder the performance of other computer vision tasks [3, 16, 24, 29, 36], such as object detection and tracking, object recognition, semantic segmentation, *etc*.

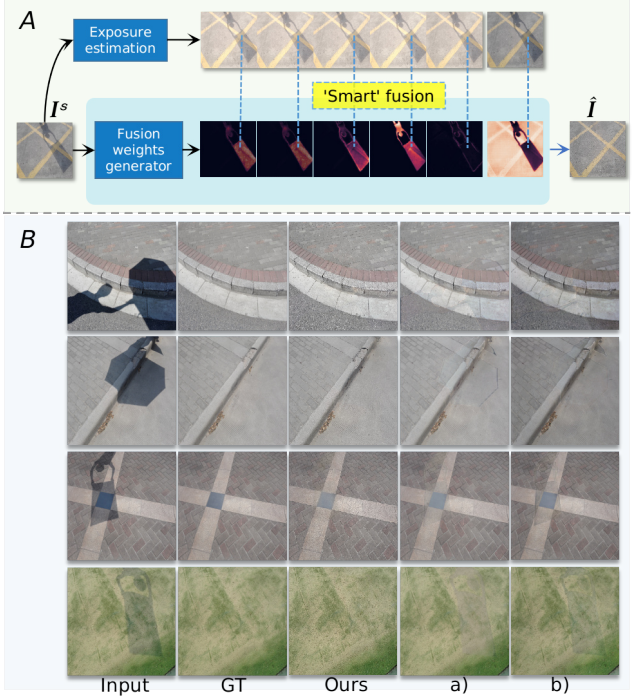

Figure 1: A: Illustration of the proposed auto-exposure fusion for shadow removal. B: Visualization results of our shadow removal results with the state-of-the-art methods. a) and b) are the shadow removal results of SP+M-Net [17] and DSC [14], respectively.

Previous shadow removal works either model this task based on physical shadow models for paired shadow and shadow-free images [17] or model it as an image-to-image translation problem based on the generative adversarial networks (GAN) for unpaired shadow and shadow-free images [15]. However, the learned shadow removal transformations by GAN-based methods, *e.g.*, MaskShadowGAN [15], tend to generate artifacts and image blur. They also suffer from data distribution requirements, where they expect the unpaired shadow and shadow-free image sets to share sta-

*Lan Fu and Changqing Zhou are co-first authors and contribute equally.
†Corresponding author: Qing Guo (tsingqguo@ieee.org)
[1]Background means the shadow-covered context in this paper.

tistical similarity [20], which is hard to be satisfied when data acquisition is unstable. On the other hand, the publicly available large-scale datasets of paired shadow and shadow-free images, such as SRD [27], ISTD [32], and ISTD+ [17], allow shadow removal tasks to learn a physically plausible transformation in a supervised way. In this paper, we focus on paired training data to perform the shadow removal task.

Shadow casting decreases the image quality with color and illumination degradation, over-exposure of the shadow image is an effective way to enhance the image quality. Intuitively, fusing the over-exposed one and the original shadow image could obtain the desired shadow-free image. Recent shadow decomposition works [17, 18], based on physical shadow models, mainly learn to relight the shadow image to a lit version and then fuse them together to acquire the desired shadow-free image via a shadow matte. However, since shadow casting degrades the color and illumination across the spatial region in a background-dependent and spatial-variant manner (*i.e*., the contiguous shadow cast on the background image may cause the shadow region to appear differently based on how the original shadow-free background region looks like, as well as where the shadow is cast spatially on the background image), we argue that multiple over-exposure fusion allows much higher level of flexibility and can provide a better solution to compensate the shadow region to have the same color and illumination with its non-shadow area, and better recovers the underlying content of the shadow region.

Shadow removal is still a challenging task for powerful state-of-the-art deep neural networks (DNN). Unknown and diverse shadow patterns pose two challenges to existing DNN based solutions: ❶ Shadow removal is a background-dependent task, which requires DNN to not only recover the illumination and color consistency with the shadow free area but also to preserve the content underlying the shadow. The spatial-variant property of shadow area requires that the fusion should be 'smart' enough to adaptively select the desired over-exposure pixels from various images to obtain the final shadow-free version. ❷ It is hard to obtain traceless background due to inconsistent shadow patterns along the boundary and inside the shadow region.

In this paper, we propose a novel method, named auto-exposure fusion network, for single image shadow removal, as shown in Fig. 1(A). We first utilize exposure estimation to learn multiple over-exposure images by compensating the shadow region with different exposure levels. Then we propose the *shadow-aware FusionNet* in Sec. 3.3 to produce fusion weight maps across all the over-exposure images for addressing the first challenge. It can 'smartly' select which over-exposed pixel is the best one to recover the position-specific background. The proposed method fuses the input image and its over-exposure versions in a pixel-wise way. Further, we propose a *boundary-aware RefineNet* in Sec. 3.4, to remove the remaining shadow trace for refining the removal result obtained in the previous step. Figure 1(B) shows that the proposed method can obtain traceless background image than the state-of-the-art methods SP+M-Net [17] and DSC [14]. The contributions of this paper are:

- To the best of our knowledge, this paper is the first work to study the shadow removal problem from the perspective of auto-exposure fusion.
- To accurately remove the shadow, we propose a new learning-based shadow-aware FusionNet followed by a boundary-aware RefineNet to accurately estimate, smartly fuse, and meticulously refine multiple over-exposure maps.
- The comprehensive experimental results on the public ISTD, ISTD+, and SRD datasets show that the proposed method achieved better performance in shadow regions and comparable performance in non-shadow regions over the state-of-the-art methods.

2. Related Work

Shadow removal. Traditional shadow removal methods employ prior information, *e.g*., gradient [9], illumination [35, 30, 33], and region [13, 31], for removing shadows. Recent deep learning based shadow removal methods boost the removal performance because of the available large-scale datasets of paired and unpaired shadow and shadow free images [17, 5, 15]. The Deshadow-Net [27] extracted multi-context features, involving global localization, appearance, and semantics, to predict a shadow matte layer for removing shadow in an end-to-end manner. Wang *et al*. proposed ST-CGAN [32] for joint shadow detection and removal by employing a stacked conditional GAN framework. The DSC [37] additionally utilized direction-aware context to improve shadow detection and removal. Le *et al*. [17] proposed to remove shadows from the perspective of shadow decomposition. On the other hand, the GAN based methods, *e.g*., MaskShadowGAN [15], made it possible to perform shadow removal on unpaired shadow and shadow free images by viewing it as an image-to-image translation problem. However, these methods suffered from artifacts and image blur. They also required the unpaired shadow and shadow free image sets to have similar statistical distribution.

We model the shadow removal problem from a novel direction, *i.e*., an auto-exposure fusion problem on paired shadow and shadow free images. Multiple over-exposure shadow images are generated to compensate the color and illumination degradation in the shadow region, then they are 'smartly' fused together to obtain the shadow free image.

Exposure fusion. Common imaging sensors' capture range is generally limited, a picture will often turn out to be under/over exposed in real world scene. Multi-exposure image fusion (MEF) can help to refine the image quality by fusing multi-exposure images into one. MEF algorithms aim to compute the fusion weight map for each image and fuse the

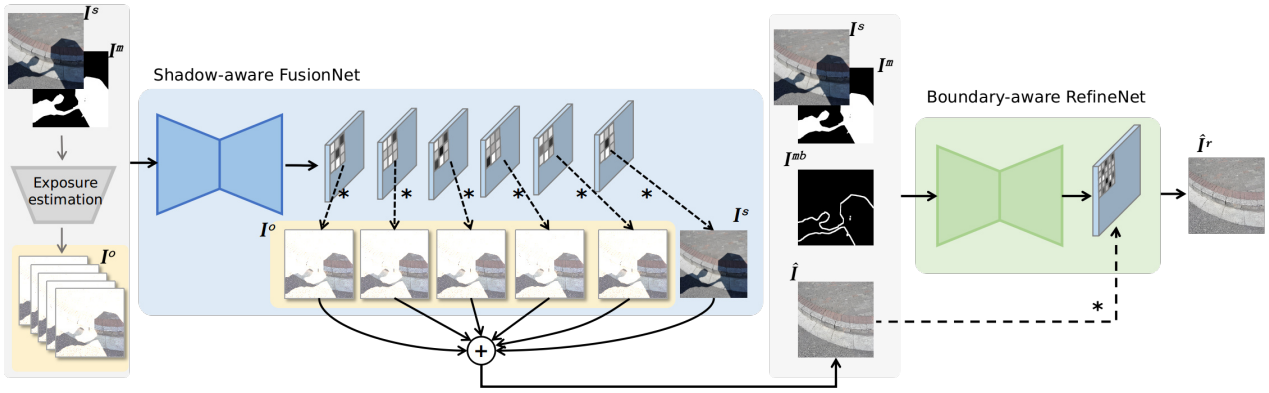

Figure 2: Illustration of the proposed framework for shadow removal with shadow-aware FusionNet and boundary-aware RefineNet.

input sequence with a weighted sum operation. Traditional MEF methods [8, 23] generally performed fusion locally or pixel wisely with hand-crafted features. Goshtasby *et al*. [8] proposed a patch-wise MEF by fusing uniform blocks with the best-exposed information chosen from each image. Mertens *et al*. [23] utilized perceptual factors, such as contrast and saturation, to construct an efficient pixel-wise MEF. Li *et al*. [19] proposed a guided filter based fusion approach, taking advantage of spatial consistency, with a two-scale decomposition. Ma *et al*. [21] performed image fusion by optimizing a structural similarity index (MEF-SSIM) with a novel gradient descent-based method. Recent deep learning based techniques have improved the fusion performance due to high representation abilities. DeepFuse network [26] performed multi-exposure fusion in an unsupervised manner by employing a loss function without reference image quality. MEF-Net [22], proposed by Ma *et al*., optimized the perceptually calibrated MEF-SSIM to predict and refine the fusion weight maps. In addition to these standard MEF methods for image enhancement, recent works also discussed the effects of MEF to the image classification [6, 2] from the angle of adversarial attack [12] by estimating the adversarial fusion weights with kernel prediction [10, 11].

In this paper, we utilize exposure fusion for the shadow removing task. Over-exposure is an effective way to enhance the image quality of shadow area. We employ pixel-wise fusion for a sequence of over-exposure images and shadow image to obtain the desired shadow-free image.

3. Methodology

In this section, we propose to formulate the shadow removal as an exposure fusion problem to recover traceless background in the shadow image. We introduce the whole framework in Sec. 3.1 and reveal the challenges. In Sec. 3.2, we explain how we generate the multi-exposure images for fusion. Then, our two main contributions, *i.e*., *shadow-aware FusionNet* in Sec. 3.3 and *boundary-aware RefineNet* in Sec. 3.4, help to address the challenges and achieve much better deshadowed images.

3.1. Exposure Fusion for Shadow Removal

We recast the shadow removal task as an exposure fusion problem and it can be formulated as

$$\hat{\mathbf{I}} = \phi(\mathbf{I}^{\mathrm{s}}), \tag{1}$$

where $\phi(\cdot)$ denotes a transfer function that can map the shadow image $\mathbf{I}^{\mathrm{s}}$ to the corresponding shadow free image $\hat{\mathbf{I}}$. A well-exposure image, *i.e*., shadow free image, could be obtained by exposure fusion of brackets of multi-exposure images to improve the image quality of shadow image. The purpose of employing image over exposure is to compensate the shadow region to have the same color and illumination as the non-shadow region. In this paper, we formulate the shadow region as an under exposed area of the shadow image. Then the problem left is to recover this area to its counterpart version which has the consistent color and illumination with the unshadowed area. Then, we can reformulate Eq. (1) to

$$\hat{\mathbf{I}} = \phi(\mathbf{I}^{\mathrm{s}}, \mathbf{I}^{\mathrm{o}}_i), i \in \{1, 2, \ldots, N\}, \tag{2}$$

where $\mathbf{I}^{\mathrm{o}}_i$ corresponds to the i-th over-exposure image of shadow image $\mathbf{I}^{\mathrm{s}}$. An intuitive way to solve it is to estimate an over-exposure version of the shadow image and then fuse them together to directly infer the desired shadow-free one. Nevertheless, shadow region is background-dependent and presents spatial-variant property, *i.e*., the color and illumination distortion across shadow region is variant, single over-exposure could not adaptively reflect the degradation in spatial space.

Therefore, we propose an auto-exposure fusion network for fusing shadow image with sequence of over exposed shadow images aiming to obtain the shadow free one. The whole framework of shadow removal is shown in Fig. 2. In Sec. 3.2, we employ a deep learning network to generate a sequence of over exposed shadow images. Then we propose the *shadow-aware FusionNet* in Sec. 3.3 to 'smartly' fuse brackets of exposed images by generating fusion weight maps across each pixel of the input image to adaptively recover the color and illumination. However, due to the existing partial shadowed region, it is hard to obtain traceless

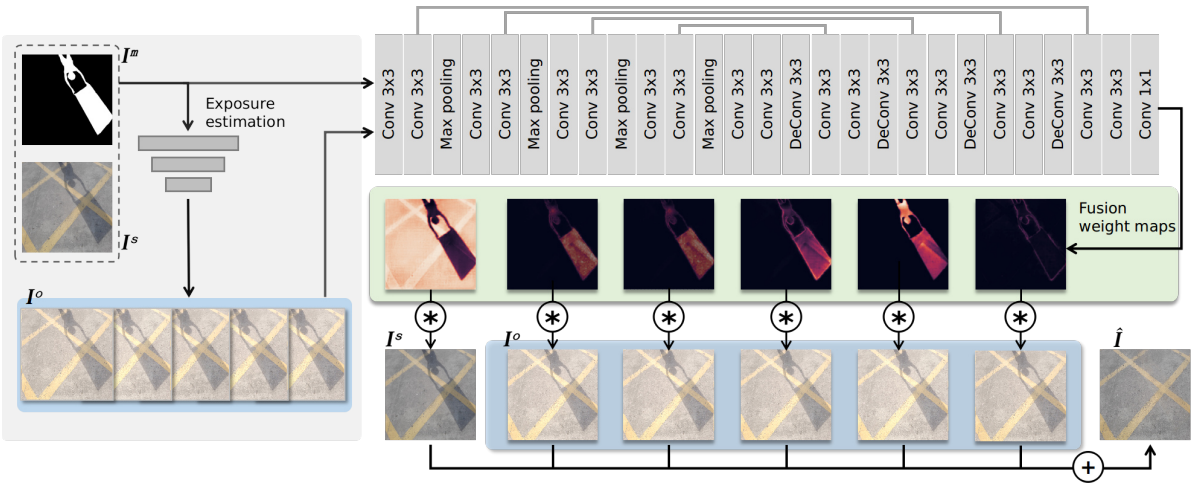

Figure 3: Illustration of the proposed shadow-aware FusionNet.

background due to the inconsistent shadow patterns along the boundary and inside the shadow area. Further, we propose a *boundary-aware RefineNet* in Sec. 3.4, to remove the residual shadow trace with the help of boundary mask.

3.2. Over-exposure Sequence Generation

We generate multiple exposure images through channel-wise weighting of the shadow image $\mathbf{I}^{\mathrm{s}}$ as following:

$$\mathbf{I}_i^{\mathrm{o}} = \alpha_i \mathbf{I}^{\mathrm{s}} + \beta_i, i \in \{1, 2, \ldots, N\}, \tag{3}$$

where $\alpha_i \in \mathbb{R}^{3\times1}$ controls the exposure degree of the i-th over-exposure image and $\beta_i \in \mathbb{R}^{3\times1}$ decides the potential intensity shifting. To realize the goal of shadow removal, we should estimate $\{\alpha_i\}$ and $\{\beta_i\}$ to make the shadow regions in the generated over-exposure images have the same color with the shadow-free regions in $\mathbf{I}^{\mathrm{s}}$. To this end, we aim to train a DNN to estimate the exposure parameters adaptively by taking the shadow image and shadow mask $\mathbf{I}^{\mathrm{m}}$ as input. Nevertheless, estimating all of the N exposure parameters directly via a DNN could let the training difficult. Instead, we adopt a two-stage way: first, we train a DNN to estimate the median exposure degree, *i.e.*, $\alpha_{\frac{N}{2}}$ and $\beta_{\frac{N}{2}}$

$$(\alpha_{\frac{N}{2}}, \beta_{\frac{N}{2}}) = \varphi(\mathbf{I}^{\mathrm{s}}, \mathbf{I}^{\mathrm{m}}), \tag{4}$$

where $\varphi(\cdot)$ denotes the DNN for exposure parameter estimation. Second, we generate all exposure images by performing a simple interpolation on $\alpha_{\frac{N}{2}}$ and $\beta_{\frac{N}{2}}$ with the assumption that the over-exposure sequence's images have similar color with minor difference

$$[\alpha_i, \beta_i] = \gamma_i [\alpha_{\frac{N}{2}}, \beta_{\frac{N}{2}}], i \in 1, 2, \ldots, N, \tag{5}$$

where $\{\gamma_i\}$ denotes the interpolation coefficients. Then, the key problem becomes how to train $\varphi(\cdot)$, which is a deep regression problem. The input data of exposure estimation is the shadow image and corresponding shadow mask. The ground truth of $\alpha_{\frac{N}{2}}$ $\beta_{\frac{N}{2}}$ is calculated by performing the least squares method [1] on the shadow mask covered regions of shadow image and its shadow-free counterpart. We optimize the exposure estimation by minimizing the mean squared error (MSE) between the estimated parameters and its ground truth. Note that, exposure parameters are estimated independently between color channels to adaptively adjust color distortion caused by shadow as well as camera sensor. We provide more details in the Sec. 3.5.

3.3. Shadow-aware FusionNet

In this section, we design the FusionNet to fuse the generated over-exposure images $\{\mathbf{I}_i^{\mathrm{o}}\}$ and produce the shadow-free image $\hat{\mathbf{I}}$. Intuitively, we can fuse $\{\mathbf{I}_i^{\mathrm{o}}\}$ by assigning each pixel a weight across different exposure degree

$$\hat{\mathbf{I}}[p] = \sum_{i=0}^{N} \mathbf{W}_i[p] \mathbf{I}_i^{\mathrm{o}}[p], \tag{6}$$

where $\mathbf{I}_0^{\mathrm{o}} = \mathbf{I}^s$, and $\mathbf{W}_i$ has the same size with $\mathbf{I}_i^{\mathrm{o}}$. Actually, such process means that each pixel of the final shadow-free image is the linear combination of N over-exposure images at the same pixel position and is fused independently. However, the fusion strategy ignores the local smoothness, leading to less natural or even noisy fusion results. Then, we further extend Eq. (6) by

$$\hat{\mathbf{I}}[p] = \sum_{i=0}^{N} (\mathbf{K}_i \circledast \mathbf{I}_i^{\mathrm{o}})[p] = \sum_{i=0}^{N} \sum_{q \in \mathcal{N}(p)} \mathbf{k}_i^p[p-q] \mathbf{I}_i^{\mathrm{o}}[q], \tag{7}$$

where $\circledast$ denotes the pixel-wise convolution, *i.e.*, each pixel is filtered by a kernel that is not shared by other pixels. Specifically, the p-th pixel of $\mathbf{I}_i^{\mathrm{o}}$ (*e.g.*, $\mathbf{I}_i^{\mathrm{o}}[p]$) and its neighboring pixels (*i.e.*, $\{\mathbf{I}_i^{\mathrm{o}}[q] | q \in \mathcal{N}(p)\}$) are linearly combined by an exclusive kernel (*i.e.*, $\mathbf{k}_i^p$ the p-th kernel in $\mathbf{K}_i$) as the combination weights and $\mathbf{k}_i^p[p-q]$ denotes $[p-q]$-th elements of $\mathbf{k}_i^p$. $\mathcal{N}(p)$ is the neighboring pixels of p. Compared

with Eq. (6), Eq. (7) considers the neighboring pixels' color and could avoid potential noisy results with better removal effect. We denote $\mathcal{K} = \{\mathbf{K}_i\}$ as pixel-wise fusion kernels.

Then, the key of generating the true shadow-free image is to estimate the fusion kernels accurately. Motivated by above process, we propose to estimate the fusion weight maps by training a CNN that takes the shadow image with shadow mask for guidance

$$\mathcal{K} = \text{FusionNet}(\mathbf{I}^{\text{m}}, \mathbf{I}^{\text{o}}), \tag{8}$$

where $\mathbf{I}^{\text{m}}$ is the shadow mask. The FusionNet is required to understand the shadow images and predict kernels that can spatially adapt to different shadow-covered contexts, thus can select suitable pixels from the multiple over-exposure images for shadow removal.

The pipeline of the shadow-aware FusionNet is shown in Fig. 3. FusionNet achieves shadow free recovery by 'smartly' selecting position-specific over-exposure pixels. The input data includes brackets of multiple exposure images, *i.e.*, the shadow image $\mathbf{I}^{\text{s}}$, corresponding shadow mask $\mathbf{I}^{\text{m}}$, and over-exposure images $\{\mathbf{I}_i^{\text{o}}\}$. FusionNet generates fusion weight maps, across all over-exposure images, to 'smartly' fuse the proper pixels from over-exposure versions with the shadow ones to the shadow free counterpart. Shadow mask $\mathbf{I}^{\text{m}}$ acts as a fusion guidance for FusionNet to let it assign low weights to non-shadow region and focus mostly on the shadow region, which is shown by the fusion weight maps in Fig. 3.

We employ L_1 distance to optimize our shadow-aware FusionNet. The loss function $\mathcal{L}_{\text{pix}}(\hat{\mathbf{I}}, \hat{\mathbf{I}}^*)$ is the pixel-wise L_1 distance between the ground truth shadow free image $\hat{\mathbf{I}}^*$ and the shadow removed image $\hat{\mathbf{I}}$

$$\mathcal{L}_{\text{pix}}(\hat{\mathbf{I}}, \hat{\mathbf{I}}^*) = \|\hat{\mathbf{I}}^* - \hat{\mathbf{I}}\|_1. \tag{9}$$

3.4. Boundary-aware RefineNet

Partially shadowed (penumbra) pixels exist along the shadow boundary. Inconsistent shadow patterns along the shadow boundary and inside the shadow region are still a challenge to state-of-the-art solutions to obtain traceless background. To solve this issue, we propose a boundary-aware RefineNet to eliminate the remaining shadow trace, which is shown in Fig. 4. It acts as a refinement of the shadow removal result obtained from FusionNet. Specifically, we model the boundary-aware RefineNet as

$$\mathcal{F} = \text{RefineNet}(\mathbf{I}^{\text{s}}, \mathbf{I}^{\text{m}}, \mathbf{I}^{\text{mb}}, \hat{\mathbf{I}}), \tag{10}$$

where $\mathbf{I}^{\text{mb}}$ is a penumbra mask, as shown in Fig. 4. Similar to Eq. (7), $\mathcal{F}$ is also pixel-wise refine kernels that integrate the context of pixel's $k \times k$ neighborhood region with that pixel to remove remaining trace. Then the refined shadow free image becomes

$$\hat{\mathbf{I}}^{\text{r}}[p] = (\mathbf{F} \circledast \hat{\mathbf{I}})[p] = \sum_{q \in \mathcal{N}(p)} \mathbf{f}^p[p-q]\hat{\mathbf{I}}[q] \tag{11}$$

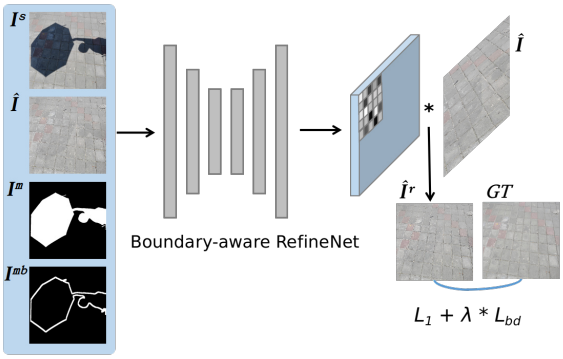

Figure 4: Illustration of the proposed boundary-aware RefineNet.

where $\mathbf{f}^p \in \mathbb{R}^{k \times k}$ is the exclusive kernel for performing convolution between the $k \times k$ neighboring pixels of the pixel p (*i.e.*, $\mathcal{N}(p)$) and the kernel weights in $\mathbf{f}^p$.

RefineNet's input data includes: the shadow image $\mathbf{I}^{\text{s}}$, shadow mask $\mathbf{I}^{\text{m}}$, penumbra mask $\mathbf{I}^{\text{mb}}$, and initial shadow removal result $\hat{\mathbf{I}}$. Penumbra mask acts as a guidance for RefineNet to keep color and illumination consistency of the shadow removed, shadow boundary, and the non-shadow regions. Penumbra mask $\mathbf{I}^{\text{mb}}$ is extracted by computing the difference between dilated shadow mask $\mathbf{I}^{\text{md}}$ and eroded shadow mask $\mathbf{I}^{\text{me}}$ for the penumbra region. We dilate/erode the shadow mask by 7 pixels to generate $\mathbf{I}^{\text{md}}$ and $\mathbf{I}^{\text{me}}$. The goal of RefineNet is to output a refined shadow removal image without trace.

The pixel-wise L_1 distance $\mathcal{L}_{\text{pix}}(\hat{\mathbf{I}}^{\text{r}}, \hat{\mathbf{I}}^*)$, between the ground-truth shadow-free image $\hat{\mathbf{I}}^*$ and the refined version of shadow removed image $\hat{\mathbf{I}}^{\text{r}}$, is utilized to optimize the boundary-aware RefineNet. In addition, inspired by Poisson image editing [25], we propose a boundary-aware loss $\mathcal{L}_{\text{bd}}(\hat{\mathbf{I}}^{\text{r}}, \hat{\mathbf{I}}^{\text{s}}, \hat{\mathbf{I}}^*, \mathbf{I}^{\text{m}})$ to seamlessly remove the shadow. It is defined as

$$\begin{aligned}\mathcal{L}_{\text{bd}}(\hat{\mathbf{I}}^{\text{r}}, \hat{\mathbf{I}}^{\text{s}}, \hat{\mathbf{I}}^*, \mathbf{I}^{\text{m}}) = \text{MSE}(\nabla\hat{\mathbf{I}}^{\text{r}}, \nabla\hat{\mathbf{I}}^{\text{s}}) * (1 - \mathbf{I}^{\text{m}}) \\ + \text{MSE}(\nabla\hat{\mathbf{I}}^{\text{r}}, \nabla\hat{\mathbf{I}}^*) * \mathbf{I}^{\text{m}}\end{aligned} \tag{12}$$

where ∇ denotes the Laplacian gradient operator. It aims to minimize the gradient domain along the shadow boundary. It keeps the same gradient domain of non-shadow region between predicted shadow-free image $\hat{\mathbf{I}}^{\text{r}}$ and shadow image $\hat{\mathbf{I}}^{\text{s}}$. At the same time, it reduces the difference of gradient domain between predicted shadow-free image $\hat{\mathbf{I}}^{\text{r}}$ and ground-truth one $\hat{\mathbf{I}}^*$ in the shadow region. The total loss of RefineNet is a weighted sum of $\mathcal{L}_{\text{pix}}(\hat{\mathbf{I}}^{\text{r}}, \hat{\mathbf{I}}^*)$ and $\mathcal{L}_{\text{bd}}(\hat{\mathbf{I}}^{\text{r}}, \hat{\mathbf{I}}^{\text{s}}, \hat{\mathbf{I}}^*, \mathbf{I}^{\text{m}})$, as shown in Fig. 4. We set λ to 0.1 in the experiment.

3.5. Implementation Details

The proposed pipeline is implemented in PyTorch. The details of network setting and training are:

1) Exposure estimation is trained together with FusionNet. Its goal is to estimate the median-exposure version of the input shadow image. We employ ResNeXt [34] as backbone

Table 1: Shadow removal results of our networks compared to state-of-the-art shadow removal methods on the ISTD [32] dataset.

Method \ RMSE	**Shadow**	**Non-Shadow**	**All**
Input Image	32.12	7.19	10.97
Guo *et al*. [13]	18.95	7.46	9.30
Gong *et al*. [35]	14.98	7.29	8.53
MaskShadow-GAN [15]	12.67	6.68	7.41
ST-CGAN [32]	10.33	6.93	7.47
DSC [14]	9.76	6.14	6.67
DHAN [4]	8.14	6.04	6.37
Ours	**7.77**	**5.56**	**5.92**

Table 2: Shadow removal results of our networks compared to state-of-the-art shadow removal methods on the ISTD+ [17] dataset.

Method \ RMSE	**Shadow**	**Non-Shadow**	**All**
Input Image	40.2	2.6	8.5
Guo *et al*. [13]	22.0	3.1	6.1
Gong *et al*. [7]	13.3	-	-
ST-CGAN [32]	13.4	7.7	8.7
DeshadowNet [27]	15.9	6.0	7.6
MaskShadow-GAN [15]	12.4	4.0	5.3
Param+M+D-Net [18]	9.7	**3.0**	4.0
SP+M-Net [17]	7.9	3.1	**3.9**
Ours	**6.5**	3.8	4.2

to do the estimation. We set the number of over-exposure images N to 5 by linearly interpolating the estimated exposure parameters with scaling factors between [0.95, 1.05]. For FusionNet, we employ a DNN with U-Net256 [28] as backbone.

2) Then boundary-aware RefineNet is to improve the shadow removal result with the same backbone as FusionNet. We train the RefineNet with exposure estimation and FusionNet together but freezing the latter two. Both FusionNet and RefineNet take the shadow mask as input, and we describe the setting of the datasets in Sec. 4.1.

In our experiments, same training parameters setting are employed for these three parts. The input image is resized to 256×256. The minibatch size is 8 and the initial learning rate is set to 0.0001. We use Adam optimizer for all the networks. We trained 400 epochs for each network.

4. Experiments

4.1. Datasets and evaluation measurement

Datasets. We train and evaluate the proposed method on three public datasets: ISTD [32], adjusted ISTD (ISTD+) [17], and SRD [27] datasets. They all have paired shadow and shadow-free images. Dataset ISTD and its adjusted version also have shadow masks. We introduce these three datasets as following:

1) The training set of ISTD dataset has 1,330 triplets of shadow, shadow free, and shadow mask images. The testing split consists of 540 triplets. The ISTD+ dataset has the same number of triplets with ISTD except that it adjusts the color inconsistency, between the shadow and shadow free image, with image processing algorithm [17]. The color mismatch results from the data acquisition setup. We use ground-truth shadow masks for training stage, while for inference, we compute the shadow masks by operating Otsu's algorithm to the difference between shadow and shadow free images, similar to MaskShadow-GAN [15]. We additionally refine these masks by a median filter to remove noises.

2) SRD dataset consists of 408 pairs of shadow and shadow free images without the ground-truth shadow mask. Here we simply use an adaptive threshold detection method, same as the one used in ISTD dataset, to extract the shadow mask from the difference between shadow free and shadow images. The extracted shadow masks are used both for training and testing. We utilize the public shadow masks provided by DHAN [4] for evaluation.

Evaluation measures. We utilize the root mean square error (RMSE) in LAB color space between the shadow removal result and the ground-truth shadow free image to evaluate the shadow removal performance, following previous works [32, 13, 27, 17, 4]. We directly compare our auto-exposure fusion framework with several state-of-the-art methods on the ISTD, ISTD+, and SRD datasets in quantitative and qualitative ways.

4.2. Shadow removal evaluation on ISTD dataset

We first report the shadow removal results of our method on ISTD dataset [32], as shown in Table 1. We compare the proposed method with the state-of-the-art algorithms: Guo *et al*. [13], Gong *et al*. [7], ST-CGAN [32], MaskShadow-GAN [15], DSC [14], and DHAN [4]. Different from other methods, MaskShadow-GAN utilizes unpaired shadow and shadow free images for training. The first row shows the RMSE values of the input shadow image and corresponding shadow free image without shadow removal operation. It shows that the proposed method obtains the best shadow removal performance in both shadow and non-shadow regions, leading to the lowest RMSE in the whole image. Specifically, the proposed method outperforms DSC [14] by 20.3% and 11.2% RMSE decreasing in shadow region and the whole image, respectively. The proposed method also outperforms the method DHAN [4] by reducing the RMSE from 8.14 to 7.77 in the shadow region. Training with unpaired data doesn't perform as well as training with paired version. Specifically, the proposed method outperforms MaskShadow-GAN by 38.6% and 20.1% RMSE decreasing in the shadow region and the whole area, respectively.

We also report the shadow removal performance of our proposed method on the adjusted ISTD (ISTD+) [17] dataset. As shown in Table 2, we compare the proposed method with state-of-the-art algorithms: Guo *et al*. [13], Gong *et al*. [7], ST-CGAN [32], DeshadowNet [27], MaskShadow-GAN [15], Param+M+D-Net [18], and SP+M-Net [17]. It

Table 3: Shadow removal results of our networks compared to state-of-the-art shadow removal methods on the SRD [27] dataset.

Method \ RMSE	**Shadow**	**Non-Shadow**	**All**
Input Image	40.28	4.76	14.11
Guo *et al*. [13]	29.89	6.47	12.60
DeshadowNet [27]	11.78	4.84	6.64
DSC [14]	10.89	4.99	6.23
DHAN [4]	8.94	**4.80**	**5.67**
Ours	**8.56**	5.75	6.51

turns out that the proposed method achieves the best shadow removal performance in the shadow region, outperforming SP+M-Net by 17.7% lower RMSE. It outperforms the DeshadowNet and ST-CGAN trained with paired shadow and shadow-free images, decreasing the RMSE by 59.1% and 51.4%, respectively. Compared to methods training with unpaired data, training with paired images still acquire better results. The proposed method outperforms Param+M+D-Net by about 32.9%, trained with unpaired shadow and shadow free patches. The proposed method achieves the comparable performance in the non-shadow and whole image region.

Figure 5 shows the visualization results of shadow removal from our methods and other state-of-the-art methods on the ISTD dataset. We can see that our result could recover traceless background in the shadow region. We can clearly see that traditional method, Guo *et al*. [13], suffers from severe artifacts and could not recover shadowed pixels successfully due to limited feature representation ability. ST-CGAN could improve the performance by training large-scale data, while it tends to generate blurry images, random artifacts, and incorrect colors, *e.g*., the fourth row shadow removed image. MaskShadowGAN and Param+M+D-Net also suffer from producing blurry images. Random artifacts along the shadow boundary can be easily spotted in the result of Param+M+D-Net, and it relights the boundary rather than removing it. Even though DSC and SP+M-Net could remove most of the shadow, their results still have trace along the shadow boundary, which does not exist in our result.

4.3. Shadow removal evaluation on SRD dataset

In this section, we show our shadow removal results on SRD dataset [27] in Table 3. We evaluate our result with the public masks provided by DHAN [4]. The proposed method obtains the best shadow removal results with the lowest RMSE in the shadow region. It reduces the RMSE from 8.94 to 8.56, compared to DHAN.

As shown in Table 3, the non-shadow region's RMSE values of different methods are very close (mean: 5.4, standard deviation: 0.6), which are similar to those of the Table 2 for ISTD+ dataset (mean: 4.4, standard deviation: 1.7). However, the standard deviations of the RMSE values in shadow region are significantly larger. This means that different methods including ours all perform well and very close on the non-shadow region, and the main difficulty of this problem comes from the shadow region. For the shadow region, our method obviously obtains the best performance.

Table 4: Ablation study of shadow removal on the ISTD+ [17] dataset.

Method \ RMSE	**Shadow**	**Non-Shadow**	**All**
Input Image	40.2	2.6	8.5
Fusion-N1	7.1	3.9	4.4
Fusion-N3	7.2	3.9	4.5
Fusion-N5	6.9	4.0	4.4
Fusion+RefineNet	6.6	3.8	4.3
Fusion+RefineNet+$\mathcal{L}_{\mathrm{bd}}$	**6.5**	**3.8**	**4.2**

4.4. Ablation study

We conduct ablation studies on ISTD+ dataset to evaluate the contribution of each step of our proposed method. For the effectiveness of per-pixel kernel fusion, *i.e*., Eq. (7) over Eq. (6), we perform *Fusion-N1* which fuses pairs of over-exposure and shadow images with the per-pixel kernel that considers 3×3 neighboring pixels and with pixel-wise fusion. It turns out that fusing image pair with neighboring information can boost the performance from 7.6 to 7.1 for RMSE in the shadow region, because neighboring region provides important spatial context information to represent the structure. We set 3×3 neighborhood for FusionNet.

Then, we conduct experiments to verify the effectiveness of multiple over-exposure by controlling the number of over-exposure images. In our implementation, we set the number N to 1, 3, and 5. The shadow removal models are denoted as *Fusion-N1*, *Fusion-N3*, and *Fusion-N5*, respectively. N is set to 5 for the remaining experiments. We test the effectiveness of boundary-aware RefineNet and loss $\mathcal{L}_{\mathrm{bd}}$ by models *Fusion+RefineNet* and *Fusion+RefineNet*+$\mathcal{L}_{\mathrm{bd}}$, respectively. The results are summarized in Table 4.

To estimate the effectiveness of multiple over-exposure to the shadow-aware FusionNet, we report the performance in shadow, non-shadow, and whole image regions with the metric RMSE. When N is 5, the shadow removal result in the shadow region reaches lower RMSE 6.9, compared to when $N = 1$. We set N to 5 for later ablation experiments. With the introducing of boundary-aware RefineNet, *Fusion+RefineNet* improves the shadow removal performance by about 0.3 RMSE decreasing. It verifies that penumbra region is a challenge for shadow removal task to get traceless background. The RMSE in non-shadow region also decreased, compared to *Fusion-N5*. Further, $\mathcal{L}_{\mathrm{bd}}$ loss optimized the shadow removal model *Fusion+RefineNet*+$\mathcal{L}_{\mathrm{bd}}$ better to reach the lowest RMSE 6.5, 3.8, and 4.2 in the shadow, non-shadow and the whole image regions.

To explain the small margin of shadow removal performance gain of *Fusion-N5* over *Fusion-N1*, we calculate the ground truth exposure for the p-th pixel in the shadow region by dividing the shadow-free pixel with its shadow counterpart for each testing example. Then, we count the average

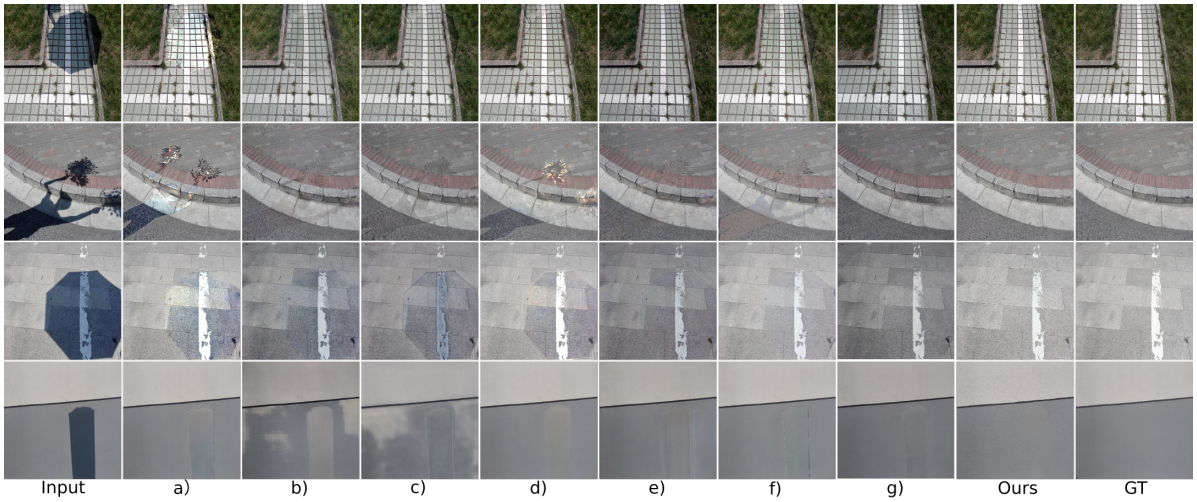

Figure 5: Illustration of the visualization results of shadow removal on dataset ISTD [32]. a) to g) are the results from comparison methods: Guo *et al*. [13], ST-CGAN [32], MaskShadow-GAN [15], Param+M+D-Net [18], DSC [14], SP+M-Net [17], and DHAN [4], respectively.

Table 5: Comparison of traceless background results in penumbra region on ISTD+ [17] dataset.

Method \ RMSE	**Penumbra**
SP+M-Net [17]	7.06
Ours	**5.96**

and the standard deviation (std. dev.) of GT exposures of all pixels in the shadow region for each example and show their relationship to the example's RMSE of the 3 variants in Fig. 6. We see that: 1) For the most examples, *Fusion-N5* and *Fusion-N3* have lower RMSE than *Fusion-N1&N3* and *Fusion-N1*, respectively. 2) Most examples' GT exposures have small variations (*i.e*., small std. dev.) across spatial coordinates, leading to similar RMSE on the three methods. 3) When the GT exposures' variation become larger, the advantages of *Fusion-N3&5* become more obvious.

We also compare our method with the state-of-the-art method SP+M-Net [17] about measuring the shadow removal result without residual trace. We evaluate the RMSE metric in the penumbra region by considering the penumbra mask $\mathbf{I}^{\text{mb}}$ as mentioned in Sec. 3.4. As shown in Table 5, our method performs better, decreasing RMSE by 15.5%. Visualizations are shown in Fig. 5(f) and ours. The SP+M-Net does not perform well to remove the residual trace.

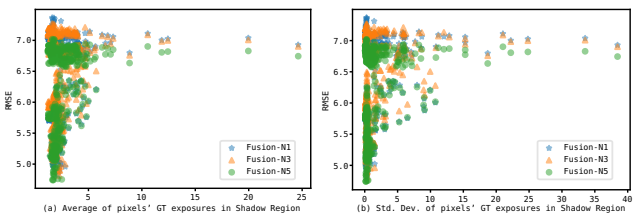

Figure 6: RMSE vs. average (*i.e*., (a)) and std. dev. (*i.e*., (b)) of pixels' GT exposures in shadow region for each testing example.

5. Conclusion

In this paper, we have proposed a novel and robust over-exposure fusion method for performing shadow removal task. Multiple over-exposure, relighting each pixel with different exposures, could compensate each pixel individually to tackle position specified color and illumination degradation. It benefits the shadow removal task by recovering the natural image from the spatial variant color and illumination degradation. Shadow-aware FusionNet smartly fuses brackets of over-exposure shadow images with shadow image by an adaptive per-pixel kernel weight map. It helps to fully recover the background content preserving the color and illumination details. The proposed boundary-aware RefineNet further eliminates the remaining trace caused by the penumbra area along the shadow boundary. With the boundary loss added, by optimizing to preserve the non-shadow region and recover the ground-truth shadow-free area of the shadow image, our work can obtain traceless background with the state-of-the-art shadow removal performance on the ISTD, ISTD+, and SRD datasets. In future, we plan to solve the challenging video shadow removal problem.

Acknowledgments: This work was supported by the NSFC under Grant U1803264, 61672376, 62072334, and 61671325. It was also supported in part by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG2-RP-2020-019), Singapore National Cybersecurity R&D Program No. NRF2018NCR-NCR005-0001, National Satellite of Excellence in Trustworthy Software System No.NRF2018NCR-NSOE003-0001, and NRF Investigatorship No. NRFI06-2020-0022. We gratefully acknowledge the support of NVIDIA AI Tech Center (NVAITC) and AWS Cloud Credits for Research Award.

References

[1] Samprit Chatterjee, Ali S Hadi, et al. Influential observations, high leverage points, and outliers in linear regression. *Statistical science*, 1(3):379–393, 1986.

[2] Y Cheng, F Juefei-Xu, Q Guo, H Fu, X Xie, SW Lin, W Lin, and Y Liu. Adversarial exposure attack on diabetic retinopathy imagery. *arXiv preprint arXiv:2009.09231*, 2020.

[3] Rita Cucchiara, Costantino Grana, Massimo Piccardi, and Andrea Prati. Detecting moving objects, ghosts, and shadows in video streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1337–1342, 2003.

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

[5] Bin Ding, Chengjiang Long, Ling Zhang, and Chunxia Xiao. Argan: Attentive recurrent generative adversarial network for shadow detection and removal. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 10213–10222, 2019.

[6] R Gao, Q Guo, F Juefei-Xu, H Yu, X Ren, W Feng, and S Wang. Making images undiscoverable from co-saliency detection. *arXiv preprint arXiv:2009.09258*, 2020.

[7] Han Gong and Darren Cosker. Interactive removal and ground truth for difficult shadow scenes. *JOSA A*, 33(9):1798–1811, 2016.

[8] A Ardeshir Goshtasby. Fusion of multi-exposure images. *Image and Vision Computing*, 23(6):611–618, 2005.

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

[10] Qing Guo, Felix Juefei-Xu, Xiaofei Xie, Lei Ma, Jian Wang, Bing Yu, Wei Feng, and Yang Liu. Watch out! motion is blurring the vision of your deep neural networks. *Advances in Neural Information Processing Systems*, 33, 2020.

[11] Qing Guo, Jingyang Sun, Felix Juefei-Xu, Lei Ma, Xiaofei Xie, Wei Feng, and Yang Liu. Efficientderain: Learning pixel-wise dilation filtering for high-efficiency single-image deraining. In *AAAI Conference on Artificial Intelligence*, 2021.

[12] Qing Guo, Xiaofei Xie, Felix Juefei-Xu, Lei Ma, Zhongguo Li, Wanli Xue, Wei Feng, and Yang Liu. SPARK: Spatial-aware Online Incremental Attack Against Visual Tracking. In *Proceedings of the European Conference on Computer Vision*, Aug 2020.

[13] Ruiqi Guo, Qieyun Dai, and Derek Hoiem. Paired regions for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(12):2956–2967, 2012.

[14] Xiaowei Hu, Chi-Wing Fu, Lei Zhu, Jing Qin, and Pheng-Ann Heng. Direction-aware spatial context features for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(11):2795–2808, 2020.

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

[16] Cláudio Rosito Jung. Efficient background subtraction and shadow removal for monochromatic video sequences. *IEEE Transactions on Multimedia*, 11(3):571–577, 2009.

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

[18] Hieu Le and Dimitris Samaras. From shadow segmentation to shadow removal. In *Proceedings of the IEEE European Conference on Computer Vision*, 2020.

[19] Shutao Li, Xudong Kang, and Jianwen Hu. Image fusion with guided filtering. *IEEE Transactions on Image processing*, 22(7):2864–2875, 2013.

[20] Yu Li, Sheng Tang, Rui Zhang, Yongdong Zhang, Jintao Li, and Shuicheng Yan. Asymmetric gan for unpaired image-to-image translation. *IEEE Transactions on Image Processing*, 28(12):5881–5896, 2019.

[21] Kede Ma, Zhengfang Duanmu, Hojatollah Yeganeh, and Zhou Wang. Multi-exposure image fusion by optimizing a structural similarity index. *IEEE Transactions on Computational Imaging*, 4(1):60–72, 2017.

[22] Kede Ma, Zhengfang Duanmu, Hanwei Zhu, Yuming Fang, and Zhou Wang. Deep guided learning for fast multi-exposure image fusion. *IEEE Transactions on Image Processing*, 29:2808–2819, 2019.

[23] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion: A simple and practical alternative to high dynamic range photography. In *Computer graphics forum*, volume 28, pages 161–171. Wiley Online Library, 2009.

[24] Sohail Nadimi and Bir Bhanu. Physical models for moving shadow and object detection in video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(8):1079–1087, 2004.

[25] Patrick Pérez, Michel Gangnet, and Andrew Blake. Poisson image editing. *ACM SIGGRAPH 2003 Papers*, pages 313–318, 2003.

[26] K Ram Prabhakar, V Sai Srikar, and R Venkatesh Babu. Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs. In *Proceedings of the International Conference on Computer Vision*, volume 1, page 3, 2017.

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

[28] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[29] Andres Sanin, Conrad Sanderson, and Brian C Lovell. Improved shadow removal for robust person tracking in surveillance scenarios. In *International Conference on Pattern Recognition*, pages 141–144. IEEE, 2010.

[30] Yael Shor and Dani Lischinski. The shadow meets the mask: Pyramid-based shadow removal. In *Computer Graphics Forum*, volume 27, pages 577–586. Wiley Online Library, 2008.

[31] Tomas F Yago Vicente, Minh Hoai, and Dimitris Samaras. Leave-one-out kernel optimization for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(3):682–695, 2017.

[32] Jifeng Wang, Xiang Li, and Jian Yang. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1788–1797, 2018.

[33] Chunxia Xiao, Ruiyun She, Donglin Xiao, and Kwan-Liu Ma. Fast shadow removal using adaptive multi-scale illumination transfer. In *Computer Graphics Forum*, volume 32, pages 207–218. Wiley Online Library, 2013.

[34] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1492–1500, 2017.

[35] Ling Zhang, Qing Zhang, and Chunxia Xiao. Shadow remover: Image shadow removal based on illumination recovering optimization. *IEEE Transactions on Image Processing*, 24(11):4623–4636, 2015.

[36] Wuming Zhang, Xi Zhao, Jean-Marie Morvan, and Liming Chen. Improving shadow suppression for illumination robust face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(3):611–624, 2018.

[37] Lei Zhu, Zijun Deng, Xiaowei Hu, Chi-Wing Fu, Xuemiao Xu, Jing Qin, and Pheng-Ann Heng. Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. In *Proceedings of the European Conference on Computer Vision*, pages 121–136, 2018.