Bijective Mapping Network for Shadow Removal

Yurui Zhu[†], Jie Huang[†], Xueyang Fu[*], Feng Zhao, Qibin Sun, Zheng-Jun Zha
University of Science and Technology of China, China
{zyr, hj0117}@mail.ustc.edu.cn, {xyfu, fzhao956, qibinsun, zhazj}@ustc.edu.cn

Abstract

Shadow removal, which aims to restore the background in the shadow regions, is challenging due to its highly ill-posed nature. Most existing deep learning-based methods individually remove the shadow by only considering the content of the matched paired images, barely taking into account the auxiliary supervision of shadow generation in the shadow removal procedure. In this work, we argue that shadow removal and generation are interrelated and could provide useful informative supervision for each other. Specifically, we propose a new Bijective Mapping Network (BMNet), which couples the learning procedures of shadow removal and shadow generation in a unified parameter-shared framework. With consistent two-way constraints and synchronous optimization of the two procedures, BMNet could effectively recover the underlying background contents during the forward shadow removal procedure. In addition, through statistical analysis of real-world datasets, we observe and verify that shadow appearances under different color spectrums are inconsistent. This motivates us to design a Shadow-Invariant Color Guidance Module (SICGM), which can explicitly utilize the learned shadow-invariant color information to guide network color restoration, thereby further reducing color-bias effects. Experiments on the representative ISTD, ISTD+ and SRD benchmarks show that our proposed network outperforms the state-of-the-art method [11] in de-shadowing performance, while only using its 0.25% network parameters and 6.25% floating point operations (FLOPs).

1. Introduction

Shadows are cast when light sources are fully or partially blocked by objects and are common in various natural scenes. However, shadows often present challenges for a variety of existing computer vision tasks, *e.g*., object tracking [38] and detection [35], face recognition [49], *etc*. Consequently, shadow removal has been studied for a long time as one of the fundamental computer vision tasks.

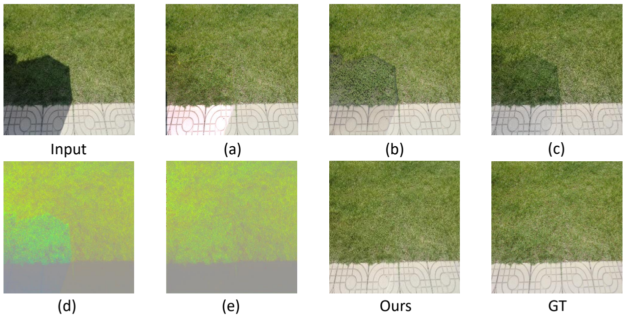

Figure 1. Visual comparisons with state-of-the-art methods on a real-world shadow scene. (a) to (c): Param+M+D-Net [28], Fu *et al*. [11], G2R [34]. (d) and (e) are the color map of the input image and our learned shadow-invariant color map, respectively.

Currently, much attention has been drawn to recovering the shadow region contents from a shadow image. Existing methods broadly come in two flavors: traditional model-based techniques and deep learning-based methods. Traditional shadow removal methods rely on the priors of shadow images , *e.g*., image gradients [15], illumination [44] and regions [16, 39]. However, due to these priors limitations, the traditional methods often are not effective to handle the shadows in complicated real shadow scenes [23].

Recently, deep learning has achieved remarkable success in various computer vision tasks [6, 12, 13, 21, 22, 36], which also includes shadow removal and gradually dominated this field. Le *et al*. [27, 28] attempt to build a linear shadow illumination model to characterize the mapping relationship between the shadow image $\mathbf{I}_s$ and the shadow-free image $\mathbf{I}_{sf}$. DSC [19] and DHAN [5] exploit the global and multi-context features to better remove shadows by designing the direction-aware spatial attention module or the growth dilated convolutions. Moreover, recently some methods [5, 34] attempt to exploit the shadow generation to obtain a large number of pseudo shadow pairs as network training data to boost the de-shadowing performance. However, most of

[*] Corresponding author, † co-first authors contributed equally. This work was supported by the National Key R&D Program of China under Grant 2020AAA0105702, the National Natural Science Foundation of China (NSFC) under Grants U19B2038 and 61901433, the University Synergy Innovation Program of Anhui Province under Grants GXXT-2019-025, the USTC Research Funds of the Double First-Class Initiative under Grant YD2100002003.

these methods may easily ignore the shadow generation procedure, as the inverse operation of shadow removal, could also provide auxiliary informative supervision for the shadow removal procedure.

In this paper, we propose a Bijective Mapping Network (BMNet) to accurately recover the underlying content in the shadow regions. Specifically, our BMNet is composed of the forward and backward mapping procedures with the help of conditional inputs $\mathbf{C}$, *i.e.*, shadow masks and color maps. Similar to previous methods [11, 34], the forward mapping process aims at learning a nonlinear function $F(\cdot)$ to transfer the shadow images $\mathbf{I}_s$ to their shadow-free version $\mathbf{I}_{sf}$. Ideally, if the forward mapping function $\mathbf{I}_{sf} = F(\mathbf{I}_s, \mathbf{C}; \theta)$ is optimal, the input shadow image $\mathbf{I}_s$ should be reconstructed closely through the reverse mapping $F^{-1}(\mathbf{I}_{sf}, \mathbf{C}; \theta)$. Such reverse mapping procedure could provide the regular constraint and informational supervision to improve the forward mapping performance [17]. Our method utilizes such a supplementary mechanism to push the estimated shadow-free images $\hat{\mathbf{I}}_{sf}$ close to the ground truth. With respect to the above key idea, we naturally implement it based on the advanced invertible frameworks [2, 4, 50]. In addition, we notice the shadow removed results of the many shadow removal methods exist obvious color-bias effect, as shown in Figure 1. By conducting statistical analysis on real-world shadow datasets, we observe that shadow appearances are different under different color spectrums. This observation motivates us to devise a Shadow-Invariant Color Guidance Module (SICGM) to explicitly employ the learned shadow-invariant color information to guide the color restoration. Different from previous methods [11, 27, 34] using only a single scale of shadow mask information, our SICGM integrates both the color and mask information into the network in a multi-scale fashion. With the guidance of the additional shadow-invariant color cues, our method could further reduce the color-bias effect. In summary, our contributions are as follows:

- We propose a new shadow removal framework, which couples the procedures of shadow removal (forward mapping) and generation (reverse mapping) in the same parameters-shared bijective mapping network (BMNet). Two procedures are synchronously optimized with consistent two-way constraints, which could benefit from each other and improve the overall de-shadowing performance.

- We propose a Shadow-Invariant Color Guidance Module in the BMNet, which explicitly incorporates the shadow-invariant color information to guide the desired color restoration of shadow regions, therefore further solving the color-bias issue.

- Comprehensive experiments on the public ISTD, ISTD+, and SRD datasets demonstrate that our proposed method achieves superior performance with very few network parameters and computational costs, *e.g.*, only 0.25% parameters and 6.85% the floating point operations of the SOTA method [11].

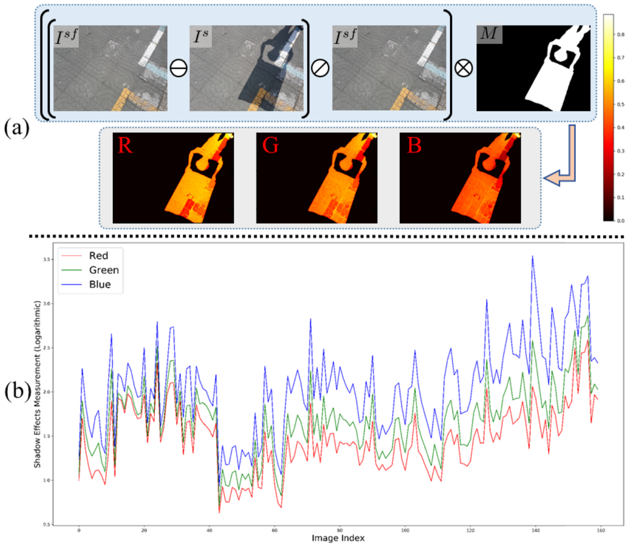

Figure 2. **(a)** A simplified visual representation of one pair of real-world shadow images statistics process based on the Eqn. (1) and (2). We could find that the attenuation appearances of the shadows under different RGB color spectrums are obviously different. **(b)** Statistics of shadow effects of different RGB basic color spectrums on the real-world ISTD test dataset. Due to the space limitation, here we present the statistical results of the top 30% image pairs in order based on the Eqn. (3). To highlight the difference, the logarithmic transformation is performed. It is clear that the shadow effects measurement values on the three RGB spectrums corresponding to most images are obviously different. Red, Green, and Blue are the three primary colors of color space. Hence, we argue the degrees of shadow effects under different color spectrums are different based on the statistical results. **(Best viewed on screen.)**

2. Related work

Shadow Generation. Previous shadow generation methods [29, 48] mainly aimed at generating shadows for the virtual objects. Besides, some shadow removal algorithms contain the shadow generation network, adopting generative adversarial techniques. G2R [34] and DHAN [5] synthesize a large number of pseudo shadow pairs through the designed shadow generators for network training. In the Mask-ShadowGAN [20, 33], the corresponding shadow generators are trained to ensure that the generated shadows and real shadows have a similar distribution by adversarial learning. Mask-ShadowGAN [20] is inspired by CycleGAN [1]. Cycle consistency losses are proposed to avoid the mode collapse issue of GANs and help minimize the distribution divergence. However, we leverage the complementary of shadow generation and removal processes to better obtain the de-shadowing performance and rely on the supervised training manner to achieve shadow generation.

Shadow Removal. Pioneer approaches [9, 10, 14, 16, 46, 47] usually implement shadow removal using various handcrafted priors, *e.g.*, image gradients, regions or user interaction. Finlayson *et al.* [9, 10] restore the shadow-free images based on the gradient consistency property. Guo *et al.* [16] attempt to build relative illumination conditions among individual regions to recover shadow contents. Gong *et al.* [14] employ the two rough user interactive inputs to design the robust algorithm for shadow removal.

Recently, Deep Neural Networks (DNNs) methods have achieved remarkable progress in the shadow removal field based on the publicly available large-scale datasets [6, 36, 40]. More precisely, DeshadowNet [36] integrates context embedding information to predict shadow matte for shadow removal. DSC [19] novelly designs a direction-aware spatial attention module to capture global information for shadow detection and removal. CANet [3] attempts to transfer the contextual information of non-shadow regions to shadow regions to achieve shadow removal. Fu *et al.* [11] formulate the shadow removal task as a multiple exposure images fusion problem. Meanwhile, generative adversarial network techniques [6, 22, 30, 40] are applied for enhancing the reality of shadow removed results or unpaired dataset training.

Previous methods [20, 34] also make use of the typical inverse procedures of shadow removal and generation to construct a close-loop constraint for their framework. However, instead of training two individual generators with various losses, we involve these two procedures into the same parameter-shared network and exploit the shadow generation procedure to serve as a regular constraint and informational supervision to maximize the overall performance.

Invertible Neural Network. Due to the reversible property, Invertible Neural Networks (INN) have drawn much attention in many computer vision tasks. INN could implement the forward and backward propagation operations within the same framework. In other words, the forward process learns a mapping function $y = f_\theta(x)$ from the source domain x to the target domain y, while the reverse mapping process can be written as $x = f_\theta^{-1}(y)$.

Pioneering researches about normalizing flow-based INN [7, 8, 25] mainly focus on the image generation tasks, which can transform a posterior distribution to another distribution with information lossless [31]. Because of the powerful network representation capacity or reversible property, INNs are also applied to many inference vision tasks, *e.g.*, Image Colorization [2], Image Rescaling [45], Image Denoising [32], Video Super-Resolution [53] and Low-Light Image Enhancement [50]. In this paper, we take advantage of the invertible mechanism of INN to tightly couple the inverse procedures of shadow removal and generation with consistent two-way constraints.

3. Method

3.1. Motivation

Here we illustrate the motivation behind the two core designs of our shadow removal algorithm. The first core design is our Bijective Mapping Network for shadow removal. Currently many previous shadow removal methods [11, 20, 27, 28, 34] exploit powerful DNNs to lean a non-linear transformation function $\mathbf{I}_{sf}$ = $F(\mathbf{I}_s, \mathbf{C}; \theta)$ to estimate $\hat{\mathbf{I}}_{sf}$ to get close to the reference $\mathbf{I}_{sf}$ based on the conditional inputs $\mathbf{C}$ (*i.e.*, shadow masks $\mathbf{M}$).

Ideally, if the forward mapping $\mathbf{I}_{sf}$ = $F(\mathbf{I}_s, \mathbf{C}; \theta)$ is optimal, we could obtain the input $\mathbf{I}_s$ through the corresponding reverse mapping $F^{-1}(\mathbf{I}_{sf}, \mathbf{C}; \theta)$. Shadow removal (forward) and generation (reverse) procedures are the two sides of the same coin. Introducing the reverse mapping process could provide a regular constraint on the $\mathbf{I}_s$ to improve the forward mapping performance [17]. In this paper, we employ such a bijective mapping manner to push the estimated shadow-free images $\hat{\mathbf{I}}_{sf}$ close to the reference. Finally, we naturally implement such a bijective mapping learning process based on the latest INNs [2, 4, 50].

Another core design of our method is based on our observation and statistical analysis. Intuitively, shadows often bring obvious color degradation to images, which shows that the original ratios of RGB spectrums have changed. It is likely that shadows have different effects on different RGB spectrums, leading to different shadow appearances under different colors. Moreover, we conduct the statistical analysis to verify this viewpoint. We first leverage the ratio between the pixel difference (values attenuation) between $\mathbf{I}_{sf}$ and $\mathbf{I}_s$ in the shadow area and the original pixel value to measure shadow appearances, which can be written as

$$A_x = \frac{I_x^{sf} - I_x^s}{I_x^{sf}}, \tag{1}$$

where A_x indicates the attenuation degree of shadow at point x, and we also call it the shadow effect at point x in this paper. For the commonly used RGB color space, there are three basic color spectrums. We respectively calculate the shadow effects of each color spectrum (red, blue, and green). For example, the shadow effects of Red spectrum at point x can be expressed as

$$A_x^R = \frac{R_x^{sf} - R_x^s}{R_x^{sf}}. \tag{2}$$

Moreover, we obtain the average shadow effects of one pair images on Red spectrum through

$$A_{mean}^R = Mean((\frac{\mathbf{R}^{sf} - \mathbf{R}^s}{\mathbf{R}^{sf}}) * \mathbf{M}), \tag{3}$$

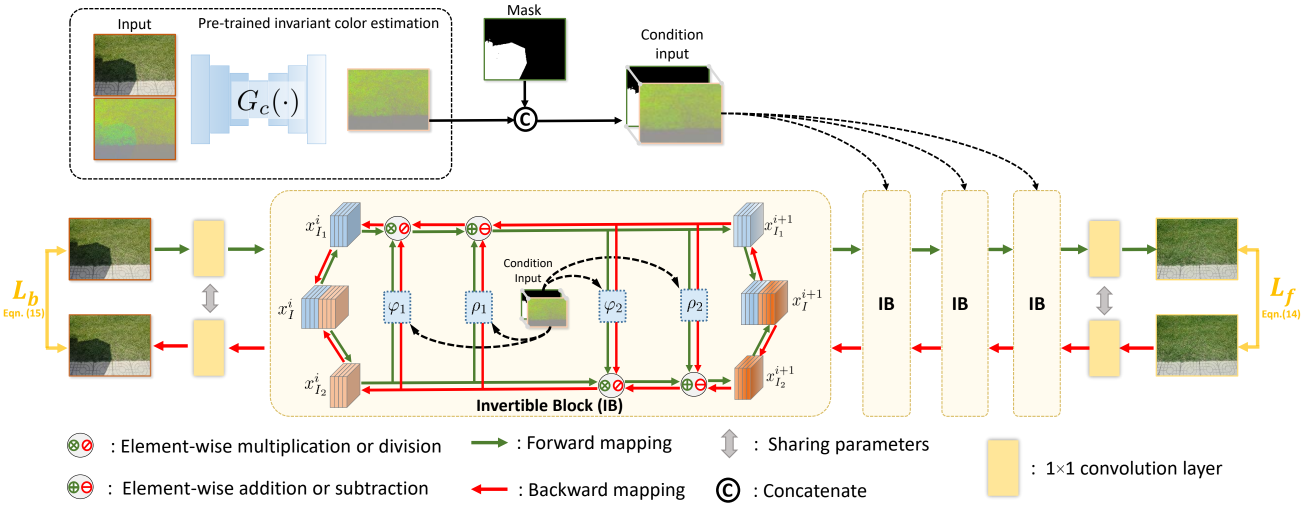

Figure 3. Illustration of our proposed Bijective Mapping Network (BMNet) for shadow removal. BMNet implements the two-way mapping procedures with the help of the condition inputs in the same parameters-shared framework. The green arrow indicates the shadow removal procedure (forward mapping), and the red arrow indicates the shadow generation procedure (reverse mapping). The reverse mapping procedure could provide the regular constraint and informational supervision to improve the de-shadowing performance.

where $Mean$ indicates the average operation and $\mathbf{M}$ indicates the shadow mask, which is provided by the original dataset. The shadow mask $\mathbf{M}$ is used in Eqn. 3, because we only need to calculate the statistical values in the shadow region. Likewise, the shadow effects on the blue and green spectrums are calculated using Eqn. (2) and (3) as well.

We further present the average shadow effects of each pair of images in the three basic color spectrums on the ISTD dataset, as shown in Figure 2. In conclusion, different color spectrums will be affected by shadows, but the degrees of shadow effects are different under different color spectrums. This conclusion motivates us to additionally employ the color cue to guide the network to reconstruct shadow-free images and reduce the color-bias effect.

3.2. Bijective Mapping Network

As shown in Figure 3, our proposed Bijective Mapping Network (BMNet) is a symmetrically designed framework, consisting of forward and backward mapping procedures during the training process. In the forward mapping process, BMNet receives the shadow image $\mathbf{I}_s$ as input and learns forward mapping from $\mathbf{I}_s$ to $\hat{\mathbf{I}}_{sf}$ based on the condition input. Specifically, we first obtain the shallow feature representation x^0_I through the plain convolutional layer with the kernel size of 1×1. Then x^0_I and the auxiliary condition information $\mathbf{C}$ (include the shadow mask and color map) are sent into n cascaded invertible blocks (IBs) for further features affine transformation. Afterward, we apply another 1×1 convolutional layer to generate $\hat{\mathbf{I}}_{sf}$. Consequently, the detailed forward mapping operation can be expressed as:

$$\begin{aligned} x^0_I &= Convs_{1\times1}(\mathbf{I}_s), \\ x^n_I &= IBs(x^0_I, \mathbf{C}), \\ \hat{\mathbf{I}}_{sf} &= Convs_{1\times1}(x^n_I), \end{aligned} \tag{4}$$

where $\mathbf{C}$ indicates the shadow mask and the learned shadow-invariant color map in our paper.

Deepening into the i-th invertible block, the input features x^i_I are equally divided into two parts $[x^i_{I_1}, x^i_{I_2}]$ along the channel dimension and pass through the affine transformation to obtain the output x^{i+1}_I as

$$x^{i+1}_{I_1} = x^i_{I_1} \otimes \varphi_1(x^i_{I_2}, \mathbf{C}) \oplus \rho_1(x^i_{I_2}, \mathbf{C}), \tag{5}$$

$$x^{i+1}_{I_2} = x^i_{I_2} \otimes \varphi_2(x^{i+1}_{I_1}, \mathbf{C}) \oplus \rho_2(x^{i+1}_{I_1}, \mathbf{C})), \tag{6}$$

$$x^{i+1}_I = Concat[x^{i+1}_{I_1}, x^{i+1}_{I_2}], \tag{7}$$

where $\otimes$ and $\oplus$ indicate element-wise multiplication and addition; φ_i and ρ_i $(i = 1, 2)$ are the affine transformation functions and we adopt the proposed SICGM to present them (details refer to Section 3.3).

When the reverse mapping propagation happens, the ground truth $\mathbf{I}_{sf}$ is reversibly transferred into its shadow-degraded version $\hat{\mathbf{I}}_s$. The affine transformations are naturally inverted in the $i-$th invertible block, which is easy to present as

$$x^i_{I_2} = (x^{i+1}_{I_2} \ominus \rho_2(x^{i+1}_{I_1}, \mathbf{C})) \oslash \varphi_2(x^{i+1}_{I_1}, \mathbf{C}), \tag{8}$$

$$x^i_{I_1} = (x^{i+1}_{I_1} \ominus \rho_1(x^i_{I_2}, \mathbf{C})) \oslash \varphi_1(x^i_{I_2}, \mathbf{C}), \tag{9}$$

$$x^i_I = Concat[x^i_{I_1}, x^i_{I_2}], \tag{10}$$

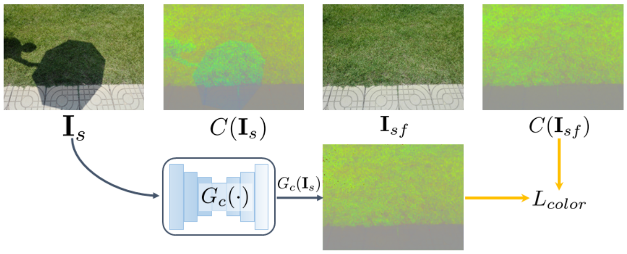

Figure 4. The first row is the shadow image pair ($\mathbf{I}_s$, $\mathbf{I}_{sf}$) and their corresponding color maps ($C(\mathbf{I}_s)$, $C(\mathbf{I}_{sf})$); The second raw is the learned color map $G_c(\mathbf{I}_s)$ with the L_{color} constraint.

where $\ominus$ and $\oslash$ indicate element-wise subtraction and division, respectively.

3.3. Shadow-Invariant Color Guidance Module

Referring to the Retinex theory [26] and [42], we could obtain the corresponding color map of an image $\mathbf{I}$ as:

$$C(\mathbf{I}) = \frac{\mathbf{I}}{Mean_c(\mathbf{I})}, \tag{11}$$

where the $Mean_c$ operation indicates the average operation of each pixel among RGB channels. We provide visual comparisons between the color maps from $\mathbf{I}_s$, $\mathbf{I}_{sf}$ and the estimated color map from the encoder $G_c(\cdot)$ in the Figure 4. The network architecture of $G_c(\cdot)$ employs classical UNet [37] structure. We can find that utilizing encoder $G_c(\cdot)$ could obtain a high-quality color map close to reference. In other words, $G_c(\cdot)$ aims at learning the shadow-invariant color information from $\mathbf{I}_s$. In our proposed Shadow-Invariant Color Guidance Module (SICGM), we explicitly inject the learned invariant color information $G_c(\mathbf{I}_s)$ and shadow mask to help the network to reconstruct shadow-free images and reduce color-bias.

As shown in Figure 5, SICGM is built on the pyramid multi-scale architecture to improve the network capability. At each scale, color maps and masks are used as conditional inputs to integrate with features F_s pass through the designed Conditional Coupling Layer (CC-Layer). More precisely, CC-Layer employs the Spatial Feature Transform (SFT) [41] to inject the conditional inputs with scaling and shifting feature transformation operation. Consequently, the detailed operation at each scale can be expressed as:

$$\begin{aligned} (\boldsymbol{\gamma}, \boldsymbol{\beta}) &= Convs_{1\times 1}(G_c(\mathbf{I}_s), \mathbf{M}), \\ \tilde{F}_s &= CCLayer(F_s \mid \boldsymbol{\gamma}, \boldsymbol{\beta}) \\ &= \boldsymbol{\gamma} \otimes F_s \oplus \boldsymbol{\beta}, \\ \tilde{F}_s &= Resblocks(\tilde{F}_s), \end{aligned} \tag{12}$$

where $Convs_{1\times 1}$ represents the vanilla convolutional layers with the kernel size of 1×1; $\otimes$ and $\oplus$ refer to element-wise multiplication and addition; $Resblocks$ is derived from the ResNet [18]. Finally, the features of different scales are aggregated together as the module final outputs.

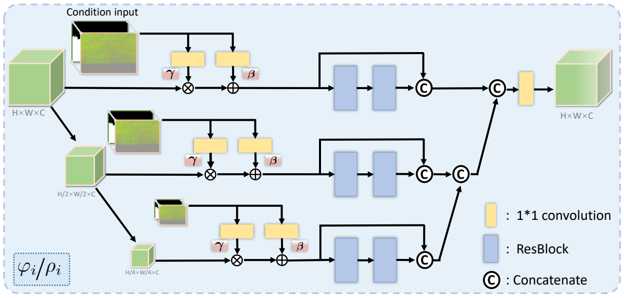

Figure 5. Illustration of our proposed Shadow-Invariant Color Guidance Module (SICGM), which is used as affine transformation functions φ_i and ρ_i ($i = 1, 2$) in the invertible blocks.

3.4. Loss function

During the training phase, we first train the color encoder $G_c(\cdot)$ to obtain the shadow-invariant color map through L1 distance loss by

$$L_{color} = \|G_c(\mathbf{I}_s) - G_c(\mathbf{I}_{sf})\|_1. \tag{13}$$

Then we train the BMNet based on the pre-trained color encoder $G_c(\cdot)$ outputs. We also adopt the L1 distance as our BMNet loss function in the bijective mapping process. The forward mapping process loss is defined as:

$$L_f = \|\hat{\mathbf{I}}_{sf} - \mathbf{I}_{sf}\|_1. \tag{14}$$

Similar, the loss function of backward mapping process is :

$$L_b = \|\hat{\mathbf{I}}_s - \mathbf{I}_s\|_1. \tag{15}$$

Therefore, the total loss of the BMNet is a weighted sum of the aforementioned losses:

$$L_{total} = L_f + \lambda L_b, \tag{16}$$

where λ indicates the weight factor.

4. Experiments

Implementation Details. Our proposed method is implemented in the PyTorch platform on the PC with a single GPU (NVIDIA GeForce GTX 3090). During our training phase, we adopt the Adam [24] optimizer with the batch size of 4 and the patch size of 256×256. For the three benchmark datasets, the total number of iterations is set as $1.5e5$. The initial learning rate of our BMNet is 0.0004 and the learning rate is reduced by half by every $5e4$ iteration. The weight factor λ of L_{total} is empirically set to 0.4.

Table 1. Quantitative comparisons with the SOTA methods on the ISTD datasets. The best and the second results are boldfaced and underlined, respectively. S, NS, and ALL indicate the shadow region, non-shadow region, and all the image, respectively.

Region	Metrics	Methods								
		Guo *et al.* [16]	ST-CGAN [40]	Mask-ShadowGAN [20]	DSC [19]	DHAN [5]	G2R [34]	Fu *et al.* [11]	Jin *et al.* [22]	Ours
S	PSNR ↑	27.76	33.74	-	34.64	34.65	31.63	34.71	31.69	**35.61**
	SSIM ↑	0.964	0.981	-	0.984	0.983	0.975	0.975	0.976	**0.988**
	RMSE ↓	18.65	9.99	12.67	8.72	8.26	10.72	7.91	11.43	**7.60**
NS	PSNR ↑	26.44	29.51	-	31.26	29.81	26.19	28.61	28.99	**32.80**
	SSIM ↑	0.975	0.958	-	0.969	0.937	0.967	0.880	0.958	**0.976**
	RMSE ↓	7.76	6.05	6.68	5.04	5.56	7.55	5.51	5.81	**4.59**
ALL	PSNR ↑	23.08	27.44	-	29.00	28.15	24.72	27.19	26.38	**30.28**
	SSIM ↑	0.919	0.929	-	0.944	0.913	0.932	0.945	0.922	**0.959**
	RMSE ↓	9.26	6.65	7.41	5.59	6.37	7.85	5.88	6.57	**5.02**
#Parameters (M: 10^6)		-	29.24	11.38	22.30	21.75	22.76	143.01	21.16	**0.37**
# FLOPs (G: 10^9)		-	17.88	56.83	123.47	262.87	113.87	160.32	105.00	**10.99**

Table 2. Quantitative comparisons with the SOTA methods on the SRD datasets.

Region	Metrics	Methods							
		Input Images	Guo *et al.* [16]	DeshadowNet [36]	DSC [19]	DHAN [5]	Fu *et al.* [11]	Jin *et al.* [22]	Ours
S	PSNR ↑	18.96	-	-	30.65	33.67	32.26	34.00	**35.05**
	SSIM ↑	0.871	-	-	0.960	0.978	0.966	0.975	**0.981**
	RMSE ↓	36.69	29.89	11.78	8.62	8.94	8.55	7.70	**6.61**
NS	PSNR ↑	31.47	-	-	31.94	34.79	31.87	35.53	**36.02**
	SSIM ↑	0.975	-	-	0.965	0.979	0.945	0.981	**0.982**
	RMSE ↓	4.83	6.47	4.84	4.41	4.80	5.74	3.65	**3.61**
ALL	PSNR ↑	18.19	-	-	27.76	30.51	28.40	31.53	**31.69**
	SSIM ↑	0.8295	-	-	0.903	0.949	0.893	0.955	**0.956**
	RMSE ↓	14.05	12.60	6.64	5.71	5.67	6.50	4.65	**4.46**

Benchmark Datasets. We conduct shadow removal experiments on the three representative ISTD [40], adjusted ISTD (ISTD+) [27] and SRD [36] benchmarks. ISTD dataset is composed of 1870 images triples (shadow images, shadow-free images, and shadow mask). This dataset has been divided into 1330 training triplets and 540 testing triplets. ISTD+ is proposed in [27], which has decreased the color inconsistency of the ISTD dataset through their designed color adjustment algorithm. Hence, the number of training and testing triplets is the same as ISTD. SRD dataset contains 2680 training pairs and 408 testing pairs, respectively. Because SRD does not provide the ground truth shadow masks, we directly utilize the public SRD shadow masks provided by DHAN [5] during the training and testing phase.

Evaluation Metrics. Following the previous methods [11, 22, 34] to evaluate the shadow removal performance, we employ the shadow removed results with a resolution of 256 × 256. We have utilized the root mean square error (RMSE) between the estimated shadow-free images $\hat{\mathbf{I}}_{sf}$ and ground truth $\mathbf{I}_{sf}$ in the LAB color space. For the RMSE metric, the lower values indicate better results. Moreover, we also adopt the Peak Signal-to-Noise Ratio (PSNR) and the structural similarity (SSIM) [43] to measure the deshadowing performance of various methods in the RGB color space. For the PSNR and SSIM metrics, higher values represent better results.

Table 3. Quantitative comparisons with the SOTA methods on the ISTD+ datasets. The best and the second results are boldfaced and underlined, respectively.

Method \RMSE ↓	Shadow	Non-Shadow	ALL Image
Input Images	40.2	2.6	8.5
Guo *et al.* [16]	22.0	3.1	6.1
ST-CGAN [40]	13.4	7.7	8.7
DeshadowNet [36]	15.9	6.0	7.6
ShadowGAN [20]	12.4	4.0	5.3
Param+M+D-Net [28]	9.7	3.0	4.0
G2R [34]	7.3	2.9	3.6
SP+M-Net [27]	7.9	3.1	3.9
Fu *et al.* [11]	6.5	3.8	4.2
Jin *et al.* [22]	10.3	3.5	4.6
Ours (w detected mask)	6.1	2.9	3.5
Ours (w GT mask)	**5.6**	**2.5**	**3.0**

4.1. Shadow Removal Evaluation on ISTD Dataset

In Table 1, we report the comparison results with recent state-of-the-art (SOTA) methods on the ISTD dataset, including ST-CGAN [40], Mask-ShadowGAN [20], DSC [19], DHAN [5], G2R [34], Fu *et al.* [11], and Jin *et al.* [22]. Apart from the deep learning (DL) based methods, we also provide one of the traditional shadow removal methods: Guo *et al.* [16]. In order to ensure the fairness of comparison, the results of these SOTA methods are provided by the authors or obtained from the original paper. Obvi-

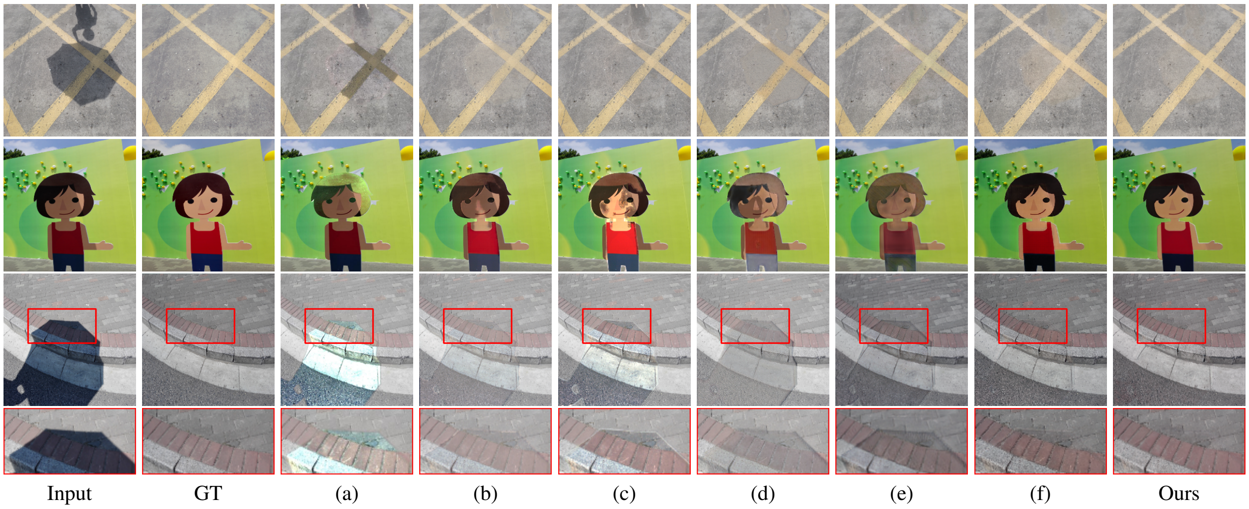

Figure 6. Visual comparison results of shadow removal on the ISTD and ISTD+ dataset. (a) to (f) are the estimated results from SOTA methods : Guo *et al.* [16], SP+M-Net [27], Param+M+D-Net [28], G2R [34], Jin *et al.* [22], and Fu *et al.* [11], respectively.

ously, our method achieves the best shadow removal performance among the shadow regions (S), non-shadow regions (NS), and all the images (ALL). Our method is also the method with the smallest amount of network parameters and the smallest amount of computational cost among all DL-based methods. Specifically, from the PNSR metric values, our method is 1.28dB higher than DSC [19]. Our proposed method also outperforms method [11] by decreasing the RMSE value from 7.91 to 7.60, only using its 0.25% network parameters and 6.25% floating point operations (FLOPs).

In addition, we have provided the visual comparison results in the Figure 6 . Obviously, the traditional method [16] cannot successfully eliminate shadows in these relatively complex scenes, and it will also greatly affect the non-shadow regions of the input image, *e.g.*, the result of the second row and third column. We can see that the results of [27, 28] suffer from obvious artifacts and color-bias effect. They fail to recover the shadowed contents probably because of their simplified linear shadow model. G2R [34] is prone to generate the blurry results and inaccurate colors, *e.g.*, the sixth column estimated results. Although method [11, 22] could remove most of the shadows, their results still exist obvious color-bias, especially in shadow regions. Compared to previous methods, our method could better restore the contents and color of shadow regions and less boundary trace.

We report the de-shadowing performance of our method on the ISTD+ dataset in Table 3. We compare the SOTA methods: Guo *et al.* [16], ST-CGAN [40], DeshadowNet [36], ShadowGAN [20], Param+M+D-Net [28], SP+M-Net [27], Fu *et al.* [11] and Jin *et al.* [22]. For the RSME metric, our method also delivers the superior performance in the S, NS and ALL regions among the SOTA methods, outperforming method [11] by 13.54 % lower RMSE value.

4.2. Shadow Removal Evaluation on SRD Dataset

In the Table 1, we report the comparison results with recent SOTA methods on the ISTD dataset, including Guo *et al.* [16], DeshadowNet [36], DSC [19], DHAN [5], G2R [34], Fu *et al.* [11], and Jin *et al.* [22]. Our method also delivers the best de-shadowing performance with the lowest RMSE and the highest PSNR values. Compared with DHAN [5], the RMSE value of our method is reduced from 8.55 to 6.61 in the shadow region. Moreover, we also provide the visual comparisons in Figure 7.

4.3. Ablation Study

Analysis of the effects of the reverse mapping procedure. In our paper, we argue the shadow generation procedure (reverse mapping) could act as a regular constraint and provide useful supervision for the shadow removal procedure (forward mapping). We conduct experiments to verify the effects of the reverse mapping process on the ISTD dataset. In Table 4, we provide the de-shadowing performance without the reverse mapping process (equivalent to no backward loss L_b). We found that the de-shadowing performance has dropped significantly based on the PSNR and RMSE metric values. In addition, we present the visual comparisons when canceling the reverse mapping procedure in Figure 8. We can see that there still exist obvious shadow residues and boundary traces.

Analysis of the Effects of Modules. We evaluate the effects of each module of our proposed network on the ISTD dataset. For the SICGM module, we change the multi-scale design into a single-scale design. We replace the SFT operation with simple concatenation. For the condition inputs, we cancel the shadow-invariant color cue to verify

Table 4. Ablation study of the each module effects. ⇒ indicates the replacement operation.

Models	S		NS		ALL	
	PSNR ↑	RMSE ↓	PSNR ↑	RMSE ↓	PSNR ↑	RMSE ↓
w/o backward loss L_b	35.13	7.97	31.92	4.79	29.42	5.33
w/o multi-scale	35.33	7.78	32.10	4.73	29.77	5.17
w/o color guidance	35.54	7.68	31.67	4.76	29.64	5.16
SSLayer (SFT ⇒ concat)	35.22	7.89	32.11	4.96	29.68	5.16
Ours (default)	35.61	7.60	32.80	4.59	30.28	5.02

Table 5. Ablation study of the number of invertible blocks (IB). We empirically set 4 IBs as default in our paper.

Metrics		Numbers of IB						
		1	2	3	4	5	6	7
ALL	PNSR ↑	28.79 (-1.49)	29.48 (-0.80)	29.95 (-0.33)	30.28 (+0.00)	30.36 (+0.08)	30.53 (+0.25)	30.61 (+0.33)
	RMSE ↓	5.63 (+0.61)	5.17 (+0.15)	5.04 (+0.02)	5.02 (+0.00)	4.93 (-0.09)	4.78 (-0.24)	4.71 (-0.31)
#Parameter (M: 10^6)		0.16	0.23	0.30	0.37	0.44	051	0.58

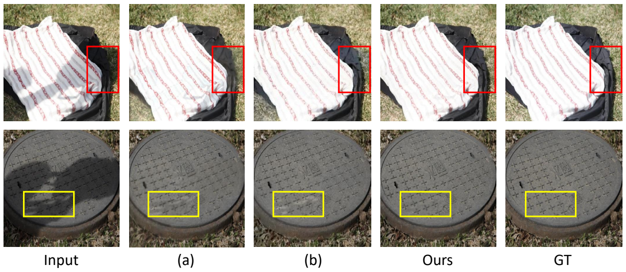

Figure 7. Visual comparison results of shadow removal on the SRD dataset. (a) and (b) are the estimated results from SOTA methods : DSC [19] and Fu *et al*. [11].

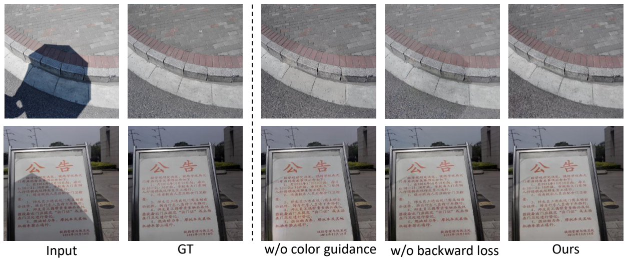

Figure 8. Visual illustration of effects of the color guidance and backward loss L_b.

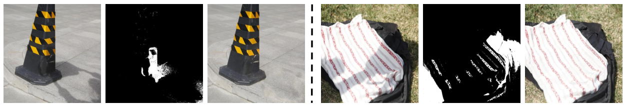

Figure 9. Our method tolerates the inaccurate masks as input. From left to right in the above two examples are: input, inaccurate mask, and output.

the effectiveness of the color information. The PSNR and RMSE values of each variation in Table 4 illustrate the contributions of each module to the de-shadowing performance improvement. We also provide visualization comparisons without shadow-invariant color cues guidance. As shown in Figure 8, there exists an obvious color-inconsistency between the shadow and surrounding non-shadow regions.

Analysis of the Numbers of Invertible Blocks. We conduct experiments to verify the shadow removal performance of BMNet with different numbers **N** of invertible blocks (IB). We report the PSNR and RMSE results and corresponding network parameters on the ISTD dataset in Table 5. Increasing the number of IB will increase the number of model parameters and computational costs. When **N** > 4, increasing the same amount of parameters will bring a limited performance improvement. Hence, we empirically set $\mathbf{N} = 4$ as the default setting, considering the trade-off between network performance and parameters.

Analysis of the Effects of the Shadow Masks. Following the previous methods [11, 27, 28, 34], shadow masks have been used as additional auxiliary information of the network to provide shadow locations. With the improvement of the shadow detection method [19, 51, 52], the detected mask can also be used as auxiliary information to reduce the cost of shadow removal [28]. Hence, our method also employs the shadow mask as default input to assist the shadow generation procedure. Here, we further verify our method performance when the network takes the imperfect detected shadow masks from [52] as condition inputs on the ISTD+ dataset in Table 3. Using detected masks, our method shows a slight decrease in de-shadowing performance. Moreover, masks of SRD dataset from DHAN [5] contain inaccurate shadow masks, our method still could robustly remove these shadows, as shown in Figure 9.

5. Limitation

Our proposed BMNet can effectively remove shadows in the images. However, it still has limitations. In the second row of Figure 8, our result suffers from slight shadow boundary traces. The shadow boundary pixels often are partially shadowed (penumbra). The shadow degradation degree of penumbra pixels exists difference compared to shadow region pixels (umbra), which may bring inconsistent traces along the boundary during the processing.

6. Conclusion

In this paper, we propose a symmetrically designed bijective mapping network, which exploits the auxiliary supervision of shadow generation (reverse mapping) for the shadow removal procedure (forward mapping). Moreover, through conducting statistical analysis on real-world datasets, we observe and verify that shadow appearances under different color spectrums are different. We specifically develop a Shadow-Invariant Color Guidance Module, explicitly embedding the shadow-invariant color information to guide the network to better remove shadows and reduce color-bias effect. Finally, comprehensive experiments demonstrate the superiority of our method, using the least parameters and FLOPs to achieve the best performance.

References

[1] Amjad Almahairi, Sai Rajeshwar, Alessandro Sordoni, Philip Bachman, and Aaron Courville. Augmented cyclegan: Learning many-to-many mappings from unpaired data. In *ICML*, pages 195–204. PMLR, 2018. 2

[2] Lynton Ardizzone, Carsten Lüth, Jakob Kruse, Carsten Rother, and Ullrich Köthe. Guided image generation with conditional invertible neural networks. *arXiv preprint arXiv:1907.02392*, 2019. 2, 3

[3] Zipei Chen, Chengjiang Long, Ling Zhang, and Chunxia Xiao. Canet: A context-aware network for shadow removal. In *ICCV*, 2021. 3

[4] Ka Leong Cheng, Yueqi Xie, and Qifeng Chen. Iicnet: A generic framework for reversible image conversion. In *CVPR*, 2021. 2, 3

[5] Xiaodong Cun, Chi-Man Pun, and Cheng Shi. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting gan. In *AAAI*, 2020. 1, 2, 6, 7, 8

[6] Bin Ding, Chengjiang Long, Ling Zhang, and Chunxia Xiao. Argan: Attentive recurrent generative adversarial network for shadow detection and removal. In *ICCV*, 2019. 1, 3

[7] Laurent Dinh, David Krueger, and Yoshua Bengio. Nice: Non-linear independent components estimation. *arXiv preprint arXiv:1410.8516*, 2014. 3

[8] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016. 3

[9] Graham D Finlayson, Mark S Drew, and Cheng Lu. Entropy minimization for shadow removal. *IJCV*, 2009. 3

[10] Graham D Finlayson, Steven D Hordley, Cheng Lu, and Mark S Drew. On the removal of shadows from images. *PAMI*, 2005. 3

[11] Lan Fu, Changqing Zhou, Qing Guo, Felix Juefei-Xu, Hongkai Yu, Wei Feng, Yang Liu, and Song Wang. Auto-exposure fusion for single-image shadow removal. In *CVPR*, 2021. 1, 2, 3, 6, 7, 8

[12] Xueyang Fu, Menglu Wang, Xiangyong Cao, Xinghao Ding, and Zheng-Jun Zha. A model-driven deep unfolding method for jpeg artifacts removal. *TNNLS*, 2021. 1

[13] Xueyang Fu, Xi Wang, Aiping Liu, Junwei Han, and Zheng-Jun Zha. Learning dual priors for jpeg compression artifacts removal. In *ICCV*, pages 4086–4095, 2021. 1

[14] Han Gong and Darren Cosker. Interactive removal and ground truth for difficult shadow scenes. *JOSA A*, 2016. 3

[15] Maciej Gryka, Michael Terry, and Gabriel J Brostow. Learning to remove soft shadows. *TOG*, 2015. 1

[16] Ruiqi Guo, Qieyun Dai, and Derek Hoiem. Paired regions for shadow detection and removal. *PAMI*, 2012. 1, 3, 6, 7

[17] Yong Guo, Jian Chen, Jingdong Wang, Qi Chen, Jiezhang Cao, Zeshuai Deng, Yanwu Xu, and Mingkui Tan. Closed-loop matters: Dual regression networks for single image super-resolution. In *CVPR*, 2020. 2, 3

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 5

[19] Xiaowei Hu, Chi-Wing Fu, Lei Zhu, Jing Qin, and Pheng-Ann Heng. Direction-aware spatial context features for shadow detection and removal. *PAMI*, 2019. 1, 3, 6, 7, 8

[20] Xiaowei Hu, Yitong Jiang, Chi-Wing Fu, and Pheng-Ann Heng. Mask-shadowgan: Learning to remove shadows from unpaired data. In *ICCV*, 2019. 2, 3, 6, 7

[21] Yukun Huang, Zheng-Jun Zha, Xueyang Fu, Richang Hong, and Liang Li. Real-world person re-identification via degradation invariance learning. In *CVPR*, pages 14084–14094, 2020. 1

[22] Yeying Jin, Aashish Sharma, and Robby T Tan. Dc-shadownet: Single-image hard and soft shadow removal using unsupervised domain-classifier guided network. In *ICCV*, 2021. 1, 3, 6, 7

[23] Salman H Khan, Mohammed Bennamoun, Ferdous Sohel, and Roberto Togneri. Automatic shadow detection and removal from a single image. *PAMI*, 2015. 1

[24] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[25] Diederik P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. *arXiv preprint arXiv:1807.03039*, 2018. 3

[26] Edwin H Land and John J McCann. Lightness and retinex theory. *Josa*, 1971. 5

[27] Hieu Le and Dimitris Samaras. Shadow removal via shadow image decomposition. In *ICCV*, 2019. 1, 2, 3, 6, 7, 8

[28] Hieu Le and Dimitris Samaras. From shadow segmentation to shadow removal. In *ECCV*, 2020. 1, 3, 6, 7, 8

[29] Daquan Liu, Chengjiang Long, Hongpan Zhang, Hanning Yu, Xinzhi Dong, and Chunxia Xiao. Arshadowgan: Shadow generative adversarial network for augmented reality in single light scenes. In *CVPR*, 2020. 2

[30] Daquan Liu, Chengjiang Long, Hongpan Zhang, Hanning Yu, Xinzhi Dong, and Chunxia Xiao. Arshadowgan: Shadow generative adversarial network for augmented reality in single light scenes. In *CVPR*, 2020. 3

[31] Yang Liu, Zhenyue Qin, Saeed Anwar, Sabrina Caldwell, and Tom Gedeon. Are deep neural architectures losing information? invertibility is indispensable. In *ICONIP*. Springer, 2020. 3

[32] Yang Liu, Zhenyue Qin, Saeed Anwar, Pan Ji, Dongwoo Kim, Sabrina Caldwell, and Tom Gedeon. Invertible denoising network: A light solution for real noise removal. In *CVPR*, 2021. 3

[33] Zhihao Liu, Hui Yin, Yang Mi, Mengyang Pu, and Song Wang. Shadow removal by a lightness-guided network with training on unpaired data. *TIP*, 2021. 2

[34] Zhihao Liu, Hui Yin, Xinyi Wu, Zhenyao Wu, Yang Mi, and Song Wang. From shadow generation to shadow removal. In *CVPR*, 2021. 1, 2, 3, 6, 7, 8

[35] Sohail Nadimi and Bir Bhanu. Physical models for moving shadow and object detection in video. *PAMI*, 26(8):1079–1087, 2004. 1

[36] Liangqiong Qu, Jiandong Tian, Shengfeng He, Yandong Tang, and Rynson WH Lau. Deshadownet: A multi-context embedding deep network for shadow removal. In *CVPR*, 2017. 1, 3, 6, 7

[37] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*. Springer, 2015. 5

[38] Andres Sanin, Conrad Sanderson, and Brian C Lovell. Improved shadow removal for robust person tracking in surveillance scenarios. In *ICPR*, 2010. 1

[39] Tomas F Yago Vicente, Minh Hoai, and Dimitris Samaras. Leave-one-out kernel optimization for shadow detection and removal. *PAMI*, 2017. 1

[40] Jifeng Wang, Xiang Li, and Jian Yang. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *CVPR*, 2018. 3, 6, 7

[41] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *CVPR*, 2018. 5

[42] Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex C Kot. Low-light image enhancement with normalizing flow. *arXiv preprint arXiv:2109.05923*, 2021. 5

[43] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *TIP*, 2004. 6

[44] Chunxia Xiao, Ruiyun She, Donglin Xiao, and Kwan-Liu Ma. Fast shadow removal using adaptive multi-scale illumination transfer. In *CGF*, 2013. 1

[45] Mingqing Xiao, Shuxin Zheng, Chang Liu, Yaolong Wang, Di He, Guolin Ke, Jiang Bian, Zhouchen Lin, and Tie-Yan Liu. Invertible image rescaling. In *ECCV*. Springer, 2020. 3

[46] Qingxiong Yang, Kar-Han Tan, and Narendra Ahuja. Shadow removal using bilateral filtering. *TIP*, 2012. 3

[47] Ling Zhang, Qing Zhang, and Chunxia Xiao. Shadow remover: Image shadow removal based on illumination recovering optimization. *TIP*, 2015. 3

[48] Shuyang Zhang, Runze Liang, and Miao Wang. Shadowgan: Shadow synthesis for virtual objects with conditional adversarial networks. *CVM*, 2019. 2

[49] Wuming Zhang, Xi Zhao, Jean-Marie Morvan, and Liming Chen. Improving shadow suppression for illumination robust face recognition. *PAMI*, 41(3):611–624, 2018. 1

[50] Lin Zhao, Shao-Ping Lu, Tao Chen, Zhenglu Yang, and Ariel Shamir. Deep symmetric network for underexposed image enhancement with recurrent attentional learning. In *ICCV*, 2021. 2, 3

[51] 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 *ECCV*, 2018. 8

[52] Lei Zhu, Ke Xu, Zhanghan Ke, and Rynson WH Lau. Mitigating intensity bias in shadow detection via feature decomposition and reweighting. In *ICCV*, 2021. 8

[53] Xiaobin Zhu, Zhuangzi Li, Xiao-Yu Zhang, Changsheng Li, Yaqi Liu, and Ziyu Xue. Residual invertible spatio-temporal network for video super-resolution. In *AAAI*, 2019. 3