# Image Shadow Removal via Multi-Scale Deep Retinex Decomposition

Xinchang Lu[a,b], Yuhui Quan[a,b], Yan Huang[a,*], Yong Xu[a,b], Hui Ji[c]

[a]*School of Computer Science and Engineering, South China University of Technology, 510000, Guangzhou, Guangdong, China*
[b]*Pazhou Laboratory, 510335, Guangzhou, Guangdong, China*
[c]*Department of Mathematics, National University of Singapore, 119076, Singapore*

**Abstract**

In recent years, deep learning has emerged as an important tool for image shadow removal. However, existing methods often prioritize shadow detection and, in doing so, they oversimplify the lighting conditions of shadow regions. Furthermore, these methods neglect cues from the overall image lighting when re-lighting shadow areas, thereby failing to ensure global lighting consistency. To address these challenges in images captured under complex lighting conditions, this paper introduces a multi-scale network built on a Retinex decomposition model. The proposed approach effectively senses shadows with uneven lighting and re-light them, achieving greater consistency along shadow boundaries. Furthermore, for the design of network, we introduce several techniques for boosting shadow removal performance, including a shadow-aware channel attention module, local discriminative and Retinex decomposition loss functions, and a multi-scale mechanism for guiding Retinex decomposition by concurrently capturing both fine-grained details and large-scale contextual information. Experimental results demonstrate the superiority of our proposed method over existing solutions, particularly for images taken under complex lighting conditions.

*Keywords:* Shadow removal, Retinex decomposition, Deep learning

## 1. Introduction

A shadow in an image is defined as a region where light intensity is diminished due to an object obstructing the light source, leading to a decrease in brightness and possible color alterations, as depicted in Fig. 1(a). Shadows in an image not only impair its aesthetic quality but also undermine the performance of related computer vision algorithms. Specifically, they can obscure object features, induce false positives or negatives in detection tasks, and create lighting inconsistencies. Image shadow removal seeks to rectify the lighting within these shadowed regions while maintaining the original colors and details of the entire image. This technique serves as a vital preprocessing step with extensive applications across various fields, including but

*Corresponding author

(a) shadowed image (b) Shadow mask (c) GT (d) SP-M-Net [5] (28.58/0.938/3.57) (e) UnfoldNet [1] (25.28/0.862/5.96) (f) MSRDNet [Ours] (36.75/0.978/2.31)

Figure 1: Illustration of image shadow removal. he PSNR (dB), SSIM, and RMSE values for each result are provided in brackets

not limited to object detection [1], medical image analysis [2], augmented reality [3], and satellite imaging [4].

Effective shadow removal in images presents numerous challenges, primarily due to the intricacies of lighting conditions and scene composition. Key issues include the ambiguity in distinguishing shadows from other dark regions, the need to handle both soft and hard shadows, and the preservation of original color and texture. Further complications arise from varying illumination conditions, multiple light sources, inter-reflections, and the non-uniformity of shadows. Lastly, another significant challenge is to maintain global illumination consistency when adjusting the lighting of shadow regions, as any oversight can lead to post-removal artifacts, such as anomalies along shadow boundaries. These multifaceted challenges make image shadow removal a complex problem to tackle.

In the era preceding deep learning, traditional methods for image shadow removal predominantly relied on handcrafted discriminative priors on both detect shadows and recover illumination in shadowed regions. These methods primarily focused on exploiting distinct attributes of shadow versus non-shadow regions, such as edges [6], intensities [7], geometries [8], and textures [9]. However, the oversimplified assumptions underlying these priors limit their applicability in complex real-world scenarios. In recent years, deep learning has made substantial contributions to the progress of shadow removal. Convolutional neural network (CNN)-based methods, often employing end-to-end training, focus on directly mapping a shadowed image to its shadow-free counterpart [10, 11, 12]. A variety of techniques utilize generative adversarial networks (GANs) to detect and subsequently remove shadows [13, 14]. Additionally, deep learning has made significant progress in shadow detection [15, 16], which can serve as a pre-processing step for subsequent shadow removal.

### *1.1. Motivation and Main Idea*

While deep learning has significantly enhanced the effectiveness and robustness of shadow removal techniques, current methods frequently fall short in fully leveraging the underlying illumination-related physics of shadows and the entire image. This limitation compromises their performance in complex lighting environments. Although end-to-end models offer a convenient solution, they often fail to generalize well across diverse real-world lighting conditions. Most existing physics-based models [5, 6], while exploiting certain physical properties of shadows, tend to rely on overly simplistic assumptions on shadows, such as constant illumination within shadow regions and apply basic linear transformations for illumination adjustments. This leads to undesirable artifacts along shadow boundaries, particularly under complex lighting conditions. For instance, Fig. 1 (d), illustrates that the SP-M-Net method [5] mistakenly darkens regions below the shadow boundary and excessively brightens the vertical gray stripe within shadowed areas. These issues motivate us to develop more advanced shadow removal techniques that can handle intricate lighting variations, particularly at shadow boundaries.

To illustrate, a shadowed image $\boldsymbol{Y}$ can be decomposed into shadow and non-shadow regions as follows:

$$\boldsymbol{Y} = (\mathbf{1} - \boldsymbol{M})\odot\boldsymbol{X} + \boldsymbol{M} \odot \boldsymbol{S}, \tag{1}$$

where $\odot$ denotes entry-wise multiplication, $\mathbf{1}$ is a matrix of all ones, $\boldsymbol{X}$ denotes the shadow-free image, $\boldsymbol{S}$ represents the shadow layer, and $\boldsymbol{M}$ is a binary mask, taking only $0, 1$ values, to indicate non-shadow and shadow region. Then, the shadow-free image $\boldsymbol{X}$ can be obtained by:

$$\boldsymbol{X} = (\mathbf{1} - \boldsymbol{M})\odot\boldsymbol{Y} + \boldsymbol{M} \odot \boldsymbol{Y}', \tag{2}$$

where $\boldsymbol{Y}'$ represents the re-light image obtained from $\boldsymbol{Y}$ which achieves natural illumination in shadow regions. While the model (2) is useful at identifying shadow regions, it has several limitations. First, it fails to consider the nuances of lighting and color, leading to possible erroneous illumination adjustment. Second, the binary nature of the mask often ignores soft transitions or gradations in shadows, resulting in visible artifacts along shadow boundary in the result. Lastly, the mask-based model overlooks the global illumination effects of shadows, potentially yielding results that appear inconsistent or unnatural. In summary, while useful for shadow identification, mask-based model (2) fall short in addressing the multifaceted challenge of restoring original color and illumination of images taken under complex illumination conditions.

To address these limitations, we utilize the well-known Retinex decomposition theory [17]. This theory has been widely used for explaining how the human visual system perceives colors under varying lighting condition, and it has also found extensive applications in tasks related to image enhancement [18, 19]. In Retinex decomposition, an image $\boldsymbol{I}$ is composed by an illumination map $\boldsymbol{L}$ and a reflectance map $\boldsymbol{R}$:

$$\boldsymbol{I} = \boldsymbol{L} \odot \boldsymbol{R}, \tag{3}$$

Here, the reflectance map $\boldsymbol{R}$ captures the inherent properties of the objects, for instance, their colors and patterns. It remains relatively invariant under different lighting conditions. The illumination $\boldsymbol{L}$ represents the light that illuminates the scene. Changes in the illumination can change the appearance of the scene without altering the inherent properties of the objects.

By leveraging the Retinex decomposition, the shadow-free image can be expressed as:

$$\boldsymbol{X} = (\boldsymbol{1} - \boldsymbol{M}) \odot \boldsymbol{L} \odot \boldsymbol{R} + \boldsymbol{M} \odot \boldsymbol{L}' \odot \boldsymbol{R}, \tag{4}$$

where $\boldsymbol{L}, \boldsymbol{L}'$ are the illumination maps of $\boldsymbol{Y}$ and $\boldsymbol{Y}'$ respectively, and the reflectance $\boldsymbol{R}$ is shared between $\boldsymbol{Y}$ and $\boldsymbol{Y}'$ due to its invariance to illumination changes caused by shadows. The adoption of a Retinex decomposition-based model for shadow removal is compelling for several reasons. First, it effectively models shadows as regions of altered illumination by employing separate illumination maps for shadowed and illuminated regions. Second, the model maintains reflectance invariance by using the same reflectance map for both shadowed and illuminated regions, aligning with the principles of Retinex theory. Third, the decoupling of illumination from reflectance enables more effective compensation and adjustment in shadowed areas, thereby enhancing re-lighting for shadow removal. Last, the decomposition allows flexible manipulation of $\boldsymbol{L}$ and $\boldsymbol{L}'$, facilitating more nuanced shadow removal.

Instead of directly incorporating the model (4) in the image domain, our approach, termed the Multi-Scale Deep Retinex Decomposition Network (MSRDNet), embeds the model (4) in feature space with a multi-scale fashion. There are two key components in MSRDNet for shadow removal: one performs Retinex decomposition in feature space for efficient differentiation between true shadows and other image artifacts and for better illumination correction with spatially-varying shadow parameters; and the other refines the model's focus on feature channels rich in shadow-related information, thereby enhancing shadow recovery. For further performance improvement, a local discriminative loss and a Retinex decomposition loss are introduced, concentrating the MSRDNet on shadow boundaries and improving illumination correction without introducing undesired artifacts.

The effectiveness of our approach is validated through extensive experiments on three benchmark datasets. For instance, as shown in Fig. 1, our approach successfully avoids over-brightening gray vertical streaks within shadowed regions and prevents erroneous removal of dark regions near the shadow boundary.

### *1.2. Main Contributions*

The primary contribution of this paper lies in addressing the limitations of existing shadow removal techniques, particularly those failing to account for the complexities of lighting conditions and shadow transitions around boundaries. While Retinex decomposition, as a widely used model explaining color perception under varying lighting conditions, has been utilized for shadow removal [15, 7], how to implement it as an effective physics-driven neural network remains a key to performance. Our proposed MSRDNet in this paper provided an efficient solution towards this goal. The main contributions of this paper are summarized as follows:

- An effective and interpretable deep neural network for image shadow removal is proposed based on a Retinex-based deshadowing model;
- Multi-scale shadow-aware processing modules are proposed, enabling better handling spatial variations of shadow illuminations across multiple scales;
- Specialized loss functions including a local discriminative loss and a Retinex decomposition loss are introduced to guide the model in focusing on shadow boundaries and ensuring the consistency of boundary between shadow and non-shadow regions.
- Our approach has achieved excellent performance and avoids common pitfalls like over-brightening and erroneous shadow removal.

## 2. Related Work

### 2.1. Non-learning Methods for Image Shadow Removal

Early studies on image shadow removal focus on non-learning methods that rely on handcrafted priors, *e.g.*, the distinct properties of shadow regions versus non-shadow regions in terms of edges [6, 20], colors and intensities [7, 21], or geometries and textures [8, 9]. Finlayson *et al.* [6] proposed to eliminate shadows by reintegrating image gradients from shadow-free regions. Yang *et al.* [21] proposed to derive intrinsic image structures from color information and then remove shadows by adjusting their brightness. Liu *et al.* [20] proposed to remove shadow by reconstructing the whole image from the estimation of shadow-free texture-consistent image gradients. Guo *et al.* [22] performed pairwise classification based on the information of segmented regions and then recovered the shadow-free image with the new illumination model. Grest *et al.* [23] introduced a method for segmenting scenes and adopted an adaptive thresholding strategy to remove shadows. While having their merits, these methods are constrained by their reliance on specific pre-defined priors that lack adaptability, limiting their performance, particularly on complex real-world data that deviate from those pre-defined rough assumptions.

### 2.2. Deep Learning Methods for Image Shadow Removal

In recent years, deep learning has emerged as a dominant approach for shadow removal, addressing the limitation of handcrafted prior-based methods by leveraging powerful modeling capability of deep neural networks. Le *et al.* [5] designed dual deep CNNs to predict shadow parameters and mask layers respectively, followed by a linear transformation to adjust the lighting conditions in shadow regions. Fu *et al.* [12] formulated shadow removal as an exposure fusion problem and introduced a shadow-aware fusion CNN to generate fusion weight maps for re-illuminated images. Gao *et al.* [11] proposed a two-stage method that sequentially uses a CNN for gray-scale enhancement and a CNN for colorization. Jie *et al.* [10] designed a random multi-level attentive CNN that fuses multi-level features via a self-attention mechanism, which enhances the adaptability to various shadow conditions.

Some studies focused on utilizing unpaired training samples. Hu *et al.* [14] employed a cycle GAN with a cycle consistency loss function for unpaired data learning. Le *et al.* [24] utilized an adversarial critic to train a shadow remover using unpaired shadow and non-shadow patches. A few studies works on the augmentation of training data. Cun *et al.* [25] developed a shadow matting GAN that synthesizes shadow images from corresponding shadow-free images and masks, thereby forming more realistic samples for training. Liu *et al.* [13] further relaxed requirement on unpaired data collection via introducing a weakly-supervised approach that trains models only with shadowed images and their corresponding shadow masks, without using ground-truth shadow-free images.

While having made progresses in improving shadow removal performance, these methods often employ simplistic strategies such as mere concatenation of shadow masks with shadow images and applying linear lighting adjustments solely in the image space. Moreover, many of these methods operate under the assumption that shadow regions exhibit constant variations, which limits their effectiveness in handling complex lighting conditions and varying shadow intensities.

### *2.3. Image Shadow Detection*

Image shadow detection itself is a problem with practical usages. An accurate shadow detector can significantly aid the shadow removal process by providing precise shadow masks. These masks are instrumental in identifying and localizing the shadows. Consequently, many shadow removal methods incorporate image shadow detection techniques as a preprocessing step to obtain inputs related to the location of shadows.

Early traditional methods rely on various clues for shadow detection [26], *e.g.*, the material and illumination relationships of region pairs [22], the comparison of texture descriptors and photometric properties [27], and the pairwise contextual cues [28]. Although these methods have shown promise in detecting shadows with reasonable accuracy in simplistic scenarios, they do not generalize well in complex environments.

Deep learning also has been extensively exploited in shadow detection. Hu *et al.* [16] employed a recurrent neural network to analyze spatial context in an orientation-aware manner and integrated this analysis into a CNN for enhancing shadow detection. Inoue *et al.* [15] extended a physically-based shadow lighting model

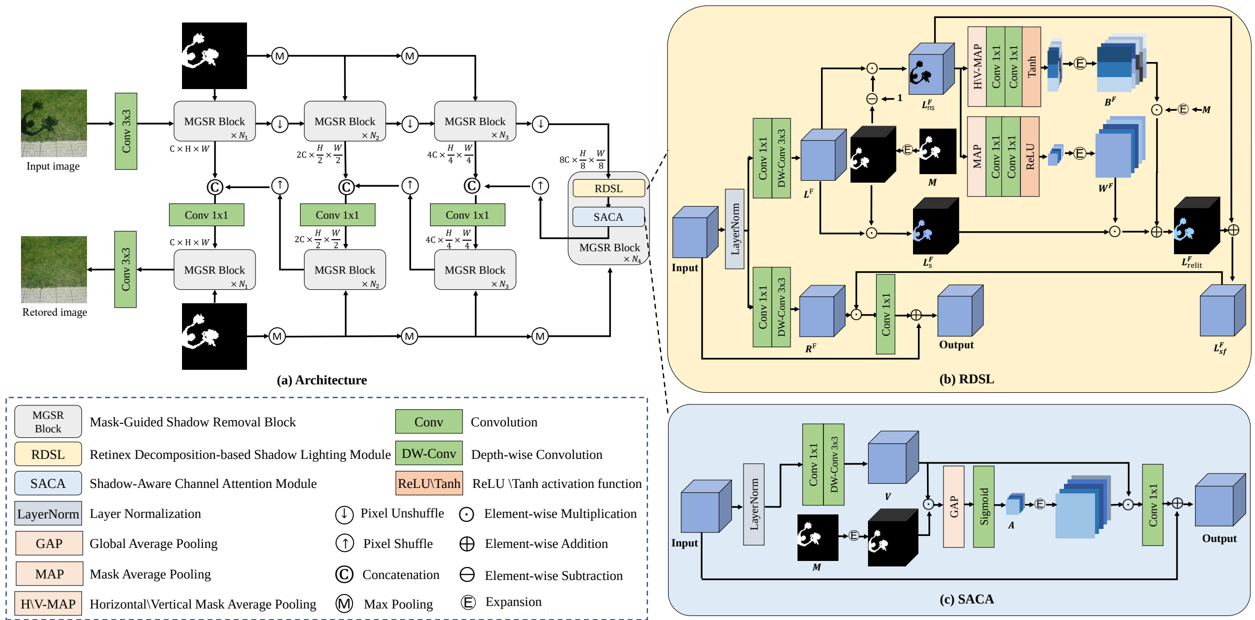



Figure 2: Architecture of proposed MSRDNet.

to synthesize shadow images, thereby improving the detection performance. Zhu *et al.* [29] implemented two parallel interactive branches to jointly generate shadow and non-shadow masks. Cun *et al.* [25] designed a dual hierarchical aggregation network that combines dilated multi-context features and attentions to further refine shadow detection.

## 3. Methodology

Our proposed MSRDNet for image shadow removal is inspired by the Retinex decomposition model described in (4). It operates within a multi-scale Retinex decomposition framework to effectively address shadow removal under complex illumination conditions. To ensure the seamless transition between shadow and non-shadow regions in the result, we augment conventional loss functions with a local discriminative loss and a Retinex decomposition loss that effectively guide the re-illumination of shadow regions.

### *3.1. Network Architecture*

As illustrated in Fig. 2(a), the MSRDNet is a 4-level symmetric encoder-decoder neural network enhanced with skip connections. Each level of this encoder-decoder structure comprises multiple Mask-Guided Shadow Removal (MGSR) blocks. The MGSR block performs efficient Retinex decomposition-based feature processing at multiple scales, consisting of a shadow-aware channel attention (SACA) module and a Retinex decomposition-based shadow lighting (RDSL) module, allowing them to extract features from both local and global perspectives. The early levels of the MGSR blocks have smaller receptive fields, allowing them to focus on local fine-grained details within the input data. As the data progresses through the MGSR blocks, the layers become progressively deeper, with larger receptive fields. This enables the later levels to capture more global, large-scale contextual information. Furthermore, the inclusion of skip connections facilitates the integration of features from earlier and later levels, effectively combining information from different scales. This results in a feature representation that adeptly captures both fine-grained local details and large-scale contextual information.

Specifically, the MSRDNet starts with a high-resolution image of size $C \times H \times W$, with $H \times W$ pixels and $C$ channels. A $3 \times 3$ convolution is initially applied to this input to extract low-level features. To perform a multi-scale analysis, through the encoder of MSRDNet, the spatial dimensions of feature maps are progressively halved, while the numbers of channels are doubled accordingly. This hierarchical downsampling serves to encapsulate and abstract important image features across varying scales. Conversely, the decoder iteratively restores the input lower-resolution features to higher-resolution representations. At each MGSR block of the decoder, the features from the previous MGSR block are concatenated with the corresponding features of the MGSR block in the encoder, and then a $1 \times 1$ convolution is employed. The resulting feature maps are used as the input features.

In addition, the shadow mask and its downsampled version are used as inputs for each MGSR block. The downsampled version is obtained through a $2 \times 2$ max pooling operation, which reduces the spatial dimensions while preserving the most prominent features of the mask. These shadow masks are employed in the RDSL module to identify shadow regions, enabling the model to leverage both global and local illumination information from non-shadowed areas to guide the illumination recovery process. They also serve as a reference in the SACA module, emphasizing shadow-rich channel features for more precise processing.

One side effect of the dimension reduction induced by the $2 \times 2$ pooling process is that it may introduce minor precision loss in shadow region identification. This raises concerns about the potential propagation of such errors. However, our approach effectively addresses this issue, and empirical experiments validate the robustness of our method against possible error propagation, as shown in Section 4.3.

### *3.2. Retinex Decomposition-based Shadow Lighting Module*

Our proposed MSRDNet leverages the concept of Retinex decomposition (3) within the feature space, particularly through the RDSL module. This module effectively distinguishes genuine shadow features from artifacts, thereby improving the visual quality of deshadowed regions. By learning spatially varying shadow parameters and integrating global and local information from non-shadow regions, MSRDNet handles spatially complex degradations within shadow areas more effectively, reducing boundary artifacts. As shown in Fig. 2(b), the input feature tensor $\boldsymbol{F} \in \mathbb{R}^{\hat{C} \times \hat{H} \times \hat{W}}$ undergoes decomposition into an illumination part $\boldsymbol{L}^{\boldsymbol{F}}$

and a reflectance part $\boldsymbol{R}^{\boldsymbol{F}}$, which is achieved via layer normalization and consecutive parallel convolutions, where $\hat{C} = C \times 2^{i-1}$, $\hat{H} = H/2^{i-1}$, $\hat{W} = W/2^{i-1}$, and $i$ represents the level in MSRDNet. With such a Retinex decomposition in feature space, the RDSL module then adjusts the lighting in shadow regions using the illumination map, guided by a shadow mask $\boldsymbol{M}$ expanded to match the dimensions of the target feature tenor.

The illumination map of features is partitioned into shadow regions $\boldsymbol{L}_{\text{s}}^{\boldsymbol{F}} = \boldsymbol{M} \odot \boldsymbol{L}^{\boldsymbol{F}}$ and non-shadow regions $\boldsymbol{L}_{\text{ns}}^{\boldsymbol{F}} = (\boldsymbol{1} - \boldsymbol{M}) \odot \boldsymbol{L}^{\boldsymbol{F}}$. Then, two types of coefficients, *i.e.*, global coefficients $\boldsymbol{W}^{\boldsymbol{F}}$ and local coefficients $\boldsymbol{B}^{\boldsymbol{F}}$, are derived from non-shadow regions to modify the lighting to process shadowed areas, as follows:

$$\boldsymbol{L}_{\text{relit}}^{\boldsymbol{F}} = \boldsymbol{L}_{\text{s}}^{\boldsymbol{F}} \odot \boldsymbol{W}^{\boldsymbol{F}} + \boldsymbol{M} \odot \boldsymbol{B}^{\boldsymbol{F}}. \tag{5}$$

The global coefficients $\boldsymbol{W}^{\boldsymbol{F}}$ integrate the global illumination information of non-shadow regions via mask average pooling (MAP), followed by two $1 \times 1$ convolutions, a ReLU activation function, and an expansion operator. The local coefficients $\boldsymbol{B}^{\boldsymbol{F}}$ address shadow degradation in a discriminative manner, obtained through horizontal/vertical MAP (H/V-MAP, alternating through the MGSR blocks), followed by two $1 \times 1$ convolutions, a Tanh activation function, and an expansion operator. The MAP and H/V-MAP are defined by:

$$\begin{aligned}
\text{MAP} : (\boldsymbol{L}_{\text{ns}}^{\boldsymbol{F}})(c,1,1) &= \Big(\sum_{i,j} \boldsymbol{1} - \boldsymbol{M}(1,i,j)\Big)^{-1} \sum_{i,j} \boldsymbol{L}_{\text{ns}}^{\boldsymbol{F}}(c,i,j), \\
\text{H-MAP} : (\boldsymbol{L}_{\text{ns}}^{\boldsymbol{F}})(c,i,1) &= \Big(\sum_{j} \boldsymbol{1} - \boldsymbol{M}(1,i,j)\Big)^{-1} \sum_{j} \boldsymbol{L}_{\text{ns}}^{\boldsymbol{F}}(c,i,j), \\
\text{V-MAP} : (\boldsymbol{L}_{\text{ns}}^{\boldsymbol{F}})(c,1,j) &= \Big(\sum_{i} \boldsymbol{1} - \boldsymbol{M}(1,i,j)\Big)^{-1} \sum_{i} \boldsymbol{L}_{\text{ns}}^{\boldsymbol{F}}(c,i,j).
\end{aligned}$$

After illumination adjustment in shadow regions, the modified illumination map $\boldsymbol{L}_{\text{sf}}^{\boldsymbol{F}}$ is synthesized by combined shadow regions $\boldsymbol{L}_{\text{relit}}^{\boldsymbol{F}}$ with the non-shadow regions $\boldsymbol{L}_{\text{ns}}^{\boldsymbol{F}}$. Finally, the modified illumination map $\boldsymbol{L}_{\text{sf}}^{\boldsymbol{F}}$ and the reflectance map $\boldsymbol{R}^{\boldsymbol{F}}$ are combined to reconstruct the enhanced features via a residual connection. Some visual examples of lighting adjustment are presented in Fig. 3, where the illumination features ($\boldsymbol{L}^{\boldsymbol{F}}, \boldsymbol{L}_{\text{s}}^{\boldsymbol{F}}, \boldsymbol{L}_{\text{sf}}^{\boldsymbol{F}}, \boldsymbol{L}_{\text{relit}}^{\boldsymbol{F}}$) are displayed in a single channel. The initial illumination map $\boldsymbol{L}^{\boldsymbol{F}}$ highlights the distinct contrast between shadowed and non-shadowed regions. After the RDSL module's adjustments, the illumination disparities between these regions in $\boldsymbol{L}_{\text{sf}}^{\boldsymbol{F}}$ are significantly minimized. This reduction is further enhanced by the attention mechanism in the SACA module and the local discriminative loss, which together result in a seamless transition between shadow and non-shadow regions. These results provide clear evidence of the Retinex model's effectiveness, particularly in reducing shadow boundary artifacts and ensuring smooth illumination recovery, thereby justifying its use in our approach.

### 3.3. Shadow-Aware Channel Attention Module

Shadows introduce considerable variations in color, texture, and illumination, thus complicating image shadow removal. To address these variations, the SACA module is specifically designed to operate synergistically within the concept of Retinex decomposition in feature channels, enabling the utilization of shadow lighting information for further enhancing the quality of the recovery of shadow regions.

The SACA module employs a shadow mask to identify channels rich in shadow features and directs attention towards them. By assigning different weights based on the similarity between the shadow mask and channel features, the module prioritizes shadow-affected areas. This mechanism helps to reduce the disparity at shadow boundaries, as shown in Fig.2 (c). Specifically, an input feature tensor is first mapped to an intermediate feature $\boldsymbol{V}$ via sequentially applying a layer normalization, a $1 \times 1$ convolution, and a $3 \times 3$ depth-wise convolution. Then, a shadow-aware channel attention vector is computed by measuring the similarity between $\boldsymbol{V}$ and $\boldsymbol{M}$, through element-wise multiplication, followed by global average pooling (GAP) and Sigmoid activation. This channel attention vector is expanded to re-weight each feature channel via element-wise product, guiding the process to focus on channels rich in shadow features. Finally, a $1 \times 1$ convolution further fuses the channel information, with a residual connection to obtain the output.

### 3.4. Loss Function

Let $\boldsymbol{X}$ and $\bar{\boldsymbol{X}}$ represent the ground-truth shadow-free image and the MSRDNet-restored image, respectively. The overall training loss function reads:

$$\mathcal{L} = \mathcal{L}_{\text{char}} + \lambda_1 \mathcal{L}_{\text{ld}} + \lambda_2 \mathcal{L}_{\text{retinex}}, \tag{6}$$

where $\mathcal{L}_{\text{char}}$ denotes the Charbonnier loss, $\mathcal{L}_{\text{ld}}$ denotes the local discriminative loss, and $\mathcal{L}_{\text{retinex}}$ denotes the Retinex decomposition loss. Both $\lambda_1$ and $\lambda_2$ are set as 1 in the experiments. The Charbonnier loss is commonly used in many computer vision tasks, particularly in image restoration problems [30]. It is a smooth approximation of the $\ell_1$ loss for robust training, which is defined as:

$$\mathcal{L}_{\text{char}} = \sqrt{\|\boldsymbol{X} - \bar{\boldsymbol{X}}\|_2^2 + \epsilon^2}, \tag{7}$$

with $\epsilon = 10^{-3}$ used in the experiments.

The local discriminative loss $\mathcal{L}_{\text{ld}}$ is inspired by the loss function proposed in [31] for image super-resolution, which is designed to more effectively capture and preserve image details and textures. Furthermore, our ap-

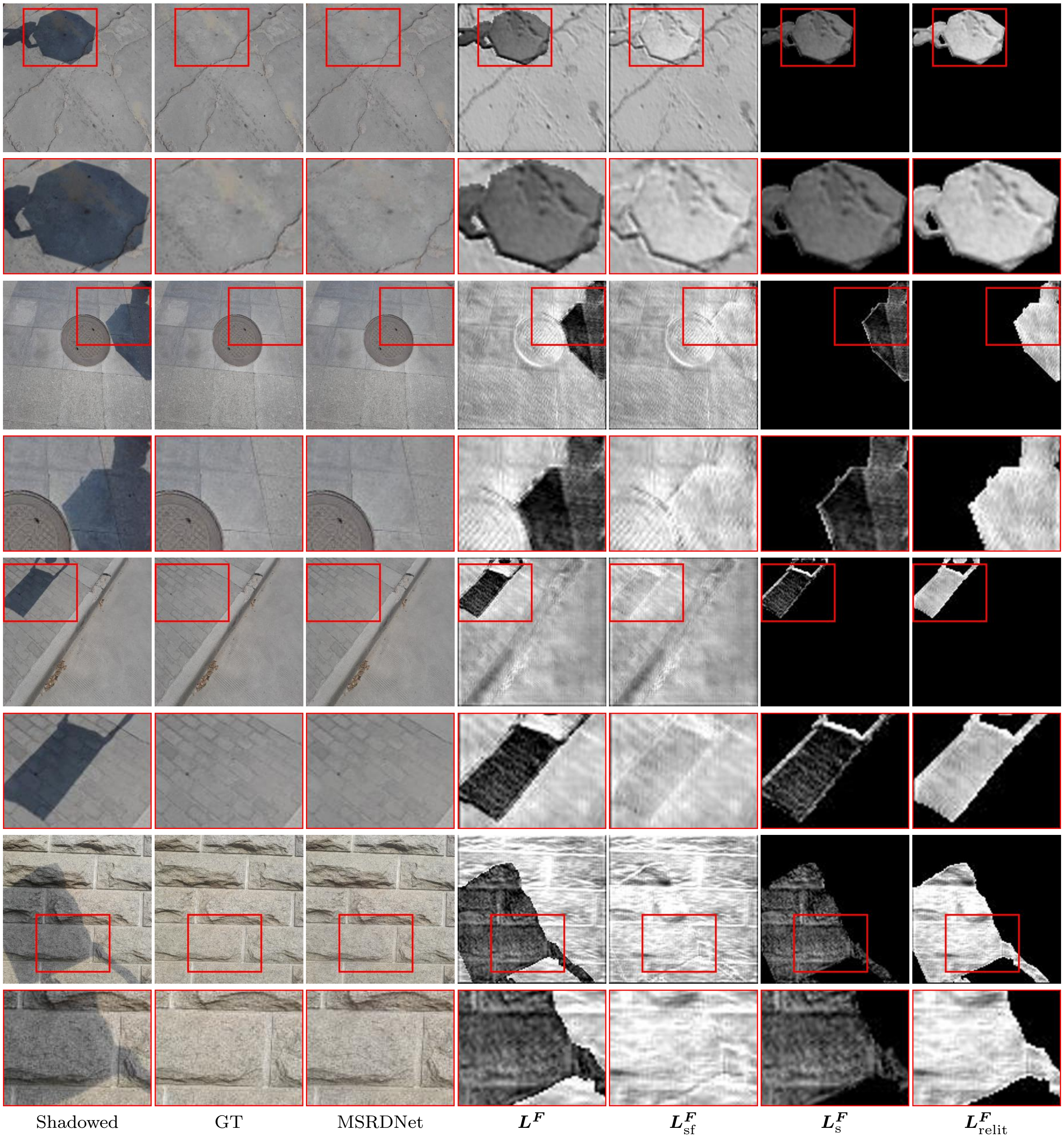



Figure 3: Visual examples of lighting adjustment in shadow regions. The illumination features from the first MGSR block at the second level of MSRDNet, within a single channel, are displayed

proach targets inaccuracies in the recovery process by utilizing the local statistical properties of the difference map, specifically focusing on regions where shadow boundary recovery is challenging. By emphasizing these regions, our method effectively reduces boundary discrepancies, improving overall deshadowing performance. The loss is defined as:

$$\mathcal{L}_{\mathrm{ld}} = \|\sigma \boldsymbol{W}(\boldsymbol{X} - \bar{\boldsymbol{X}})\|_1, \tag{8}$$

where $\boldsymbol{W}$ denotes the local variance of the residual map $\boldsymbol{E}$ between two images, and $\sigma$ scales its global variance. Concretely, the matrix $\boldsymbol{W}$ is defined as:

$$\boldsymbol{W}(i,j) = var(\boldsymbol{E}(i - \frac{s-1}{2} : i + \frac{s-1}{2}, j - \frac{s-1}{2} : j + \frac{s-1}{2})), \tag{9}$$

where $\boldsymbol{E} = \boldsymbol{X} - \bar{\boldsymbol{X}}$, and $\sigma = (var(\boldsymbol{E}))^{\frac{1}{\alpha}}$. The parameter $s$ and $\alpha$ are set as 7 and 5 respectively in the experiments. The local discriminative weight $\boldsymbol{W}$ penalizes inconsistencies between the restored and ground-truth images, particularly at shadow boundaries. The global coefficient $\sigma$ captures the overall error and aids in color and texture preservation.

The Retinex decomposition loss $\mathcal{L}_{\mathrm{retinex}}$ guides the model to embed Retinex decomposition in the feature domain. It is defined as:

$$\mathcal{L}_{\mathrm{retinex}} = \mathcal{L}_{\mathrm{R}} + \mathcal{L}_{\mathrm{LR}}, \tag{10}$$

where

$$\mathcal{L}_{\mathrm{R}} = \sum_{i=1}^{l} \sum_{j=1}^{N_i} \omega_i \|\phi(\boldsymbol{R}_{i,j}^{\boldsymbol{F}}) - \boldsymbol{R}_{\downarrow 2^{i-1}}^{\boldsymbol{X}}\|_1, \tag{11}$$

$$\mathcal{L}_{\mathrm{LR}} = \sum_{i=1}^{l} \sum_{j=1}^{N_i} \omega_i \|\psi((\boldsymbol{L}_{\mathrm{sf}}^{\boldsymbol{F}})_{i,j}) \phi(\boldsymbol{R}_{i,j}^{\boldsymbol{F}}) - \boldsymbol{X}_{\downarrow 2^{i-1}}\|_1. \tag{12}$$

Here, $l$ is the total number of levels in the MSRDNet, $N_i$ is count of MGSR blocks at the $i$-th level, $\omega_i$ is the balance coefficient, $\downarrow r$ denotes the down-sampling operator with a ratio of $r$, and $\psi$ and $\phi$ are functions mapping features of the illumination and reflectance maps to the image domain, respectively. The $(\boldsymbol{L}_{\mathrm{sf}}^{\boldsymbol{F}})_{i,j}$ and $\boldsymbol{R}_{i,j}^{\boldsymbol{F}}$ represent the features of illumination and reflectance map in $j$-th MGSR block of $i$-th level, and $\boldsymbol{R}^{\boldsymbol{X}}$ represents the reflectance map obtained by applying the classic Retinex decomposition method [32] on the ground-truth image $\boldsymbol{X}$.

The terms $\mathcal{L}_{\mathrm{R}}$ and $\mathcal{L}_{\mathrm{LR}}$ serve to align the feature maps with the reflectance and illumination layers extracted from the ground-truth image, respectively. By combining these losses, the model is steered towards accurate Retinex decomposition, thereby preserving essential details and achieving a reconstruction closely resembling the ground truth.

## 4. Experiments

### *4.1. Experimental Settings*

**Datasets**: Three commonly-used datasets are included for the performance evaluation of shadow removal models: ISTD [33], adjusted ISTD (ISTD+) [5] and SRD [34]. The ISTD includes 1330 training triplets and 540 test triplets. The ISTD+ has the same number of triplets as ISTD, except that the ISTD+ dataset reduces the lighting inconsistency between the shadowed images and the shadow-free images through post-processing. The SRD contains 3088 pairs of samples, of which 2680 pairs are used for training, and 408 pairs for testing. The ISTD and ISTD+ provide shadow masks. For SRD, following existing works, we use the masks from [25] for training and test. See Fig. 4 for example images on the datasets.


(a) ISTD (b) ISTD+ (c) SRD

Figure 4: Sample images from three benchmark datasets.

**Implementation details**: The number of MGSR blocks $N_1$ to $N_4$ are set as 2, 4, 6, 8 respectively, with the number of channel $C = 32$. The coefficients $\omega_1$ to $\omega_4$ are all set as $1/N$, where $N = N_1 + N_2 + N_3 + N_4$. The functions $\psi$ and $\phi$ are implemented using a $3 \times 3$ convolutional layer followed by a sigmoid activation. The training employs the AdamW optimizer on $256 \times 256$ random crops with a batch size of 4. The learning rate starts at $3e^{-4}$ and decays to $1e^{-6}$ after 150 epochs via cosine annealing. The MSRDNet is implemented in PyTorch on a single NVIDIA GTX 3090Ti GPU. The code will be public upon papaer's acceptance.

**Evaluation metrics**: Following existing works [5, 12, 33, 34, 25], we use there metrics to quantify the deshadowing performance: peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and rooted mean squared error (RMSE). The PSNR and SSIM are calculated in the original RGB space, while the RMSE is calculated in the LAB color space. Better performance generally leads to lower RMSE, higher PSNR, and higher SSIM. The three metrics are computed on the shadow regions, non-shadow regions, as well as the whole image, respectively.

### *4.2. Comparison with representative shadow removal methods*

We evaluate MSRDNet against several state-of-the-art shadow removal methods, including both non-learning ones (Guo *et al.* [7] and Gong *et al.* [35]) and deep learning-based ones (ST-CGAN [33], MS-GAN [14], ST-CGAN [33], MS-GAN [14], DeShadowNet [34], DSC [16], DHAN [25], G2R [13], FusionNet [12], CANet [36], UnfoldNet [1], BMNet [37], SP-M-Net [5], DC-ShadowNet [38], Param-Net [24], DMTN [39], RMLANet [10]). For fair comparison, all the results of these compared methods are quoted from existing literature whenever possible; or otherwise produced by their released codes. Following the experimental configuration of previous works [12, 1, 37], the results of shadow removal models are resized to $256 \times 256$ for evaluation.

**Quantitative Comparison**: The quantitative results on the three datasets are given in Tables 1, 2, 3, respectively. From the tables we can see that, MSRDNet excels on ISTD and ISTD+ across all metrics, notably in shadow regions. Compared to the second-best performer (DMTN) on ISTD, our MSRDNet improves PSNR by 1.53dB. On ISTD+, the PSNR of MSRDNet gains 1.06 dB against the second-best performer, BMNet. On SRD, our MSRDNet ranks the second. Furthermore, the tables also compare the model size and computational complexity [1]. We can see that the MSRDNet has a model size comparable to its top competitors such as DMTN [39] and UnfoldNet [1]. Moreover, the MSRDNet has a significant advantage over TBRNet and DMTN regarding the computational complexity, using only 9% and 14.2% #MACs of them respectively. These results have justified the superior performance of MSRDNet in image shadow removal.

**Qualitative Comparison**: The qualitative evaluation is done via visual comparisons of shadow removal results from different methods, as shown in Fig. 5, 6, 7, for the three datasets respectively. Clearly, MSRDNet consistently excels in visual quality. For instance, Fig. 5 shows MSRDNet's superior treatment of shadow boundaries, maintaining color continuity at the boundaries after post-removal. In Fig. 6, unlike other methods, MSRDNet not only effectively removes shadows of weak-intensity without misidentifying other regions,

[1] The number of Multipy-ACcumulate operations (#MACs) is used to evaluate the computational complexity, and the #MACs is analyzed by FVcore from `https://github.com/facebookresearch/fvcore`

Table 1: Quantitative comparisons on ISTD. **Boldface** indicates the best results and underline indicates the second best results.

| Method | Source | Params(M) | #MACs(G) | Shadow Region | | | Non-Shadow Region | | | Whole Image | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | PSNR | SSIM | RMSE | PSNR | SSIM | RMSE | PSNR | SSIM | RMSE |
| Input Image | n/a | n/a | n/a | 22.40 | 0.936 | 32.10 | 27.32 | 0.976 | 7.09 | 20.56 | 0.893 | 10.88 |
| Guo [7] | TPAMI'12 | n/a | n/a | 27.76 | 0.964 | 18.95 | 26.44 | 0.966 | 7.46 | 23.08 | 0.920 | 9.30 |
| Gong [35] | BMVC'14 | n/a | n/a | 30.14 | 0.973 | 14.98 | 26.98 | 0.972 | 7.29 | 24.71 | 0.926 | 8.53 |
| ST-CGAN [33] | CVPR'18 | 31.8 | 8.94 | 33.74 | 0.981 | 10.33 | 29.51 | 0.958 | 6.93 | 27.44 | 0.929 | 7.47 |
| DSC [16] | TPAMI'19 | 22.3 | 61.74 | 33.45 | 0.967 | 9.48 | n/a | n/a | 6.14 | n/a | n/a | 6.67 |
| MaskShadow-GAN [14] | ICCV'19 | 13.8 | 28.41 | n/a | n/a | 12.67 | n/a | n/a | 6.68 | n/a | n/a | 7.41 |
| SP-M-Net [5] | ICCV'19 | 141.2 | 27.82 | 32.16 | 0.981 | 10.30 | 26.40 | 0.970 | 7.47 | 25.08 | 0.943 | 7.79 |
| DHAN [25] | AAAI'20 | 21.8 | 131.44 | 34.98 | 0.984 | 7.52 | n/a | n/a | 5.43 | n/a | n/a | 5.76 |
| DC-ShadowNet [38] | ICCV'21 | n/a | n/a | 31.69 | 0.976 | 11.43 | 28.99 | 0.958 | 5.81 | 26.38 | 0.922 | 6.57 |
| G2R [13] | CVPR'21 | 22.76 | 56.94 | 31.63 | 0.975 | 10.72 | 26.19 | 0.967 | 7.55 | 24.72 | 0.932 | 7.85 |
| FusionNet [12] | CVPR'21 | 186.5 | 80.16 | 34.71 | 0.975 | 7.91 | 28.61 | 0.880 | 5.51 | 27.19 | 0.846 | 5.88 |
| CANet [36] | ICCV'21 | 358.2 | n/a | n/a | n/a | 8.86 | n/a | n/a | 6.07 | n/a | n/a | 6.15 |
| UnfoldNet [1] | AAAI'22 | 10.1 | 12.69 | 36.27 | 0.986 | 7.78 | 31.85 | 0.965 | 4.72 | 29.98 | 0.944 | 5.22 |
| BMNet [37] | CVPR'22 | **0.4** | **5.50** | 35.61 | 0.988 | 7.60 | 32.80 | 0.976 | 4.59 | 30.28 | 0.959 | 5.02 |
| TBRNet [40] | TNNLS'23 | 46.7 | 361.89 | 36.35 | 0.987 | 6.40 | 31.18 | 0.951 | 4.49 | 29.64 | 0.934 | 4.76 |
| DMTN [39] | TMM'23 | 22.8 | 230.22 | 35.83 | **0.989** | 7.19 | 33.01 | 0.979 | 4.18 | 30.42 | **0.965** | 4.62 |
| RMLANet [10] | TCSVT'23 | n/a | n/a | n/a | n/a | 7.68 | n/a | n/a | 4.24 | n/a | n/a | 4.80 |
| MSRDNet [Ours] | n/a | 19.7 | 32.75 | **37.36** | **0.989** | **6.38** | **33.38** | **0.980** | **4.09** | **31.16** | **0.965** | **4.48** |

Table 2: Quantitative comparisons on ISTD+. **Boldface** indicates the best results and underline indicates the second best results.

| Method | Source | Params(M) | #MACs(G) | Shadow Region | | | Non-Shadow Region | | | Whole Image | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | PSNR | SSIM | RMSE | PSNR | SSIM | RMSE | PSNR | SSIM | RMSE |
| Input Image | n/a | n/a | n/a | 20.83 | 0.927 | 36.95 | 37.46 | 0.985 | 2.4 | 20.46 | 0.894 | 8.4 |
| DeshadowNet [34] | CVPR'17 | n/a | n/a | n/a | n/a | 15.9 | n/a | n/a | 6.0 | n/a | n/a | 7.6 |
| SP-M-Net [5] | ICCV'19 | 141.2 | 27.82 | 35.72 | 0.986 | 7.5 | 36.55 | 0.978 | 2.8 | 32.37 | 0.956 | 3.5 |
| Param-Net [24] | ECCV'20 | n/a | n/a | n/a | n/a | 9.7 | n/a | n/a | 3.0 | n/a | n/a | 4.0 |
| DHAN [25] | AAAI'20 | 21.8 | 131.44 | 32.92 | 0.988 | 9.6 | 27.15 | 0.971 | 7.4 | 25.66 | 0.956 | 7.8 |
| FusionNet [12] | CVPR'21 | 186.5 | 80.16 | 36.04 | 0.978 | 6.6 | 31.16 | 0.892 | 3.8 | 29.45 | 0.861 | 4.2 |
| BMNet [37] | CVPR'22 | **0.4** | **5.50** | 37.87 | **0.991** | 5.8 | 37.51 | **0.985** | **2.4** | 33.98 | **0.972** | 3.0 |
| TBRNet [40] | TNNLS'23 | 46.7 | 361.89 | 36.34 | **0.991** | 6.5 | 35.57 | 0.977 | 3.3 | 31.91 | 0.963 | 3.8 |
| DMTN [39] | TMM'23 | 22.8 | 230.22 | 37.12 | **0.991** | 6.2 | 38.00 | 0.984 | 2.5 | 33.68 | **0.972** | 3.1 |
| RMLANet [10] | TCSVT'23 | n/a | n/a | n/a | n/a | 7.0 | n/a | n/a | **2.4** | n/a | n/a | 3.1 |
| MSRDNet[Ours] | n/a | 19.7 | 32.75 | **38.93** | **0.991** | **5.5** | **38.49** | **0.985** | **2.4** | **34.94** | **0.972** | **2.9** |

Table 3: Quantitative comparisons on SRD. **Boldface** indicates the best results and underline indicates the second best results.

| Method | Source | Params(M) | #MACs(G) | Shadow Region | | | Non-Shadow Region | | | Whole Image | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | PSNR | SSIM | RMSE | PSNR | SSIM | RMSE | PSNR | SSIM | RMSE |
| Input Image | n/a | n/a | n/a | 18.96 | 0.871 | 36.69 | 31.47 | 0.975 | 4.83 | 18.19 | 0.830 | 14.05 |
| Guo [7] | TPAMI'12 | n/a | n/a | n/a | n/a | 29.89 | n/a | n/a | 6.47 | n/a | n/a | 12.60 |
| DSC [16] | TPAMI'19 | 22.3 | 61.74 | 30.65 | 0.960 | 8.62 | 31.94 | 0.965 | 4.41 | 27.76 | 0.903 | 5.71 |
| SP-M-Net [5] | ICCV'19 | 141.2 | 27.82 | 32.67 | 0.970 | 8.61 | 32.04 | 0.964 | 5.32 | 28.71 | 0.919 | 6.23 |
| DHAN [25] | AAAI'20 | 21.8 | 131.44 | 33.67 | 0.978 | 8.94 | 34.79 | 0.979 | 4.80 | 30.51 | 0.949 | 5.67 |
| FusionNet [12] | CVPR'21 | 186.5 | 80.16 | 32.26 | 0.966 | 9.55 | 31.87 | 0.945 | 5.74 | 28.40 | 0.893 | 6.50 |
| UnfoldNet [1] | AAAI'22 | 10.1 | 12.69 | 34.94 | 0.980 | 7.44 | 35.85 | 0.982 | 3.74 | 31.72 | 0.952 | 4.79 |
| BMNet [37] | CVPR'22 | **0.4** | **5.50** | 35.05 | 0.981 | 6.61 | 36.02 | 0.982 | 3.61 | 31.69 | 0.956 | 4.46 |
| TBRNet [40] | TNNLS'23 | 46.7 | 361.89 | 35.53 | 0.981 | 6.38 | 34.97 | 0.979 | 4.21 | 31.64 | 0.952 | 4.86 |
| DMTN [39] | TMM'23 | 22.8 | 230.22 | **37.29** | **0.985** | **5.33** | **37.50** | **0.989** | 3.21 | **33.73** | **0.967** | **3.82** |
| RMLANet [10] | TCSVT'23 | n/a | n/a | n/a | n/a | 7.03 | n/a | n/a | **3.16** | n/a | n/a | 4.39 |
| MSRDNet[Ours] | n/a | 19.7 | 32.75 | 35.43 | 0.984 | 5.98 | 36.23 | **0.989** | 3.38 | 32.17 | 0.965 | 4.09 |

but also preserves color details well in non-shadow regions. In the more complex scenarios with irregular shadows, as in Fig. 7, most methods struggle leaving shadow residues or oversmoothed textures. In contrast, leveraging Retinex decomposition, MSRDNet not only successfully restores the illumination of all shadows, but also preserves the texture details of the image.

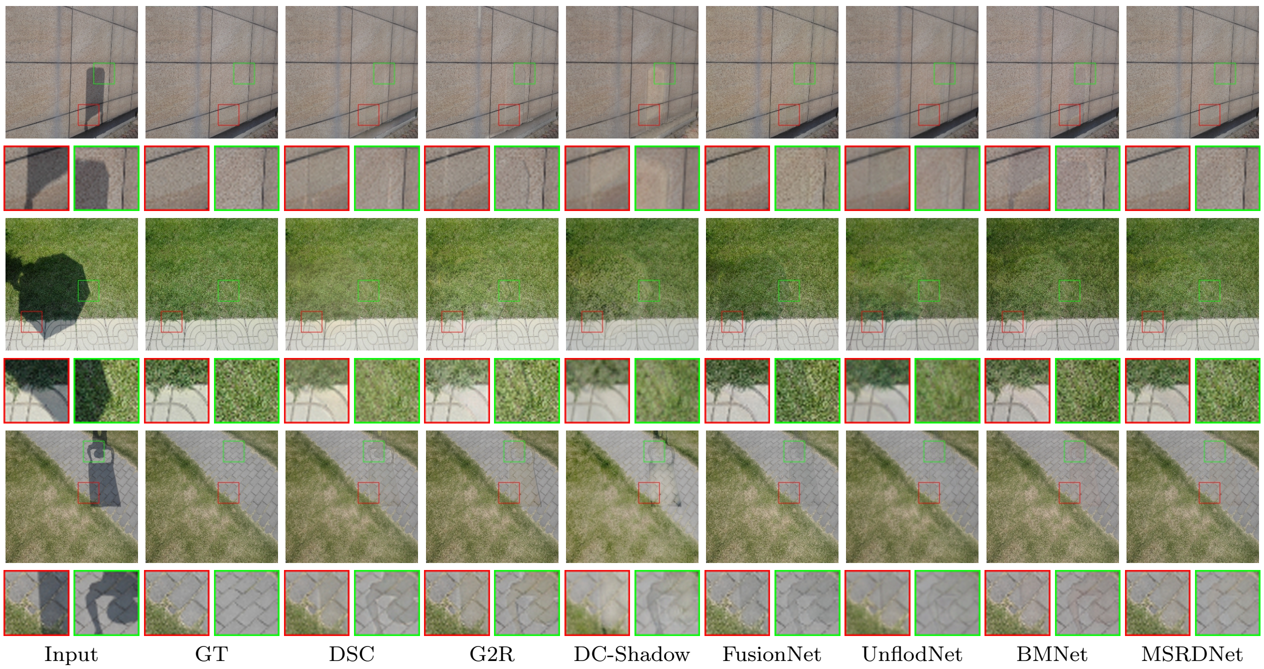



Figure 5: Shadow removal results of different methods on some images from ISTD.

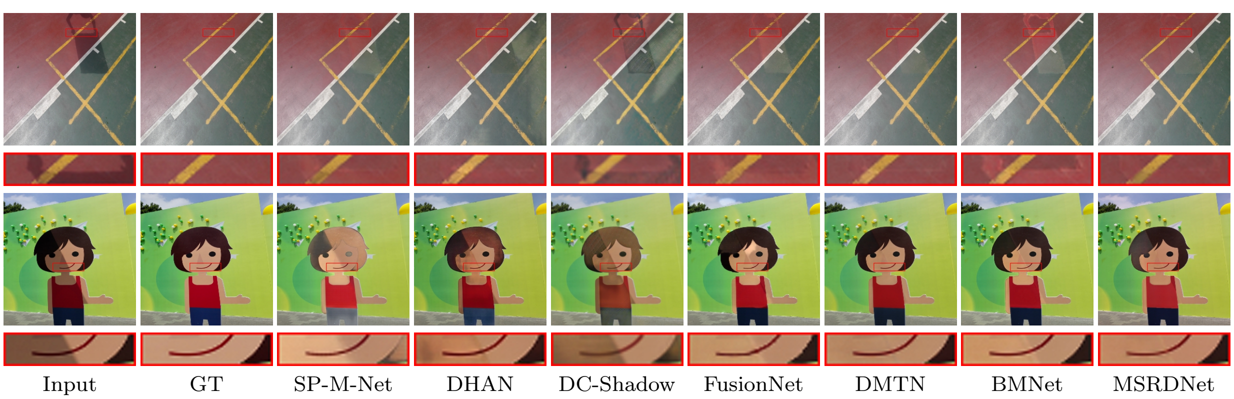



Figure 6: Shadow removal results of different methods on some images from ISTD+.

*4.3. Ablation Study*

To analyze the contribution of each key component in MSRDNet, we conduct ablation studies on ISTD by forming the following baseline models:


Figure 7: Shadow removal results of different methods on some images from SRD.

- “w/o SACA": It excludes the SACA module from the MGSR block. The number of MGSR blocks, $N_1$ to $N_4$, are set to $[4, 6, 6, 8]$, to maintain the original model size.
- “w/o $\boldsymbol{R^F}$": Removes the reflectance feature layer $R^f$ in the RDSL module.
- “w/o $\boldsymbol{B^F}$": It omits the branch calculating the local coefficient $\boldsymbol{B^F}$ in the RDSL module.
- “$\mathcal{L}_{\text{char}} \to \mathcal{L}_1$": It replaces Charbonnier loss $\mathcal{L}_{\text{char}}$ by the $\mathcal{L}_1$ loss for training.
- “w/o $\mathcal{L}_{\text{ld}}$": It excludes the local discriminative loss $\mathcal{L}_{\text{ld}}$ from the overall loss function in training.
- “w/o $\mathcal{L}_{\text{retinex}}$": It excludes the Retinex decomposition loss $\mathcal{L}_{\text{retinex}}$ from the overall loss function for training.

Table 4: Results of ablation study on ISTD.

| Model Setting | Shadow Region (S) | | All Image (ALL) | |
|---|---|---|---|---|
| | PSNR(dB) | RMSE | PSNR(dB) | RMSE |
| w/o SACA | 36.68 | 6.93 | 28.61 | 5.43 |
| w/o $\boldsymbol{R^F}$ | 35.99 | 7.81 | 30.22 | 4.96 |
| w/o $\boldsymbol{B^F}$ | 36.89 | 7.21 | 30.67 | 4.82 |
| $\mathcal{L}_{\text{char}} \to \mathcal{L}_1$ | 37.11 | 6.54 | 30.91 | 4.61 |
| w/o $\mathcal{L}_{\text{ld}}$ | 36.90 | 6.87 | 30.61 | 4.73 |
| w/o $\mathcal{L}_{\text{retinex}}$ | 37.02 | 6.59 | 30.90 | 4.59 |
| Original model | **37.36** | **6.38** | **31.16** | **4.48** |

The results of these baseline models are listed in Table 4, where MSRDNet notably outperforms all of them, demonstrating the effectiveness of the key components in our approach. The detailed analysis is as follows.

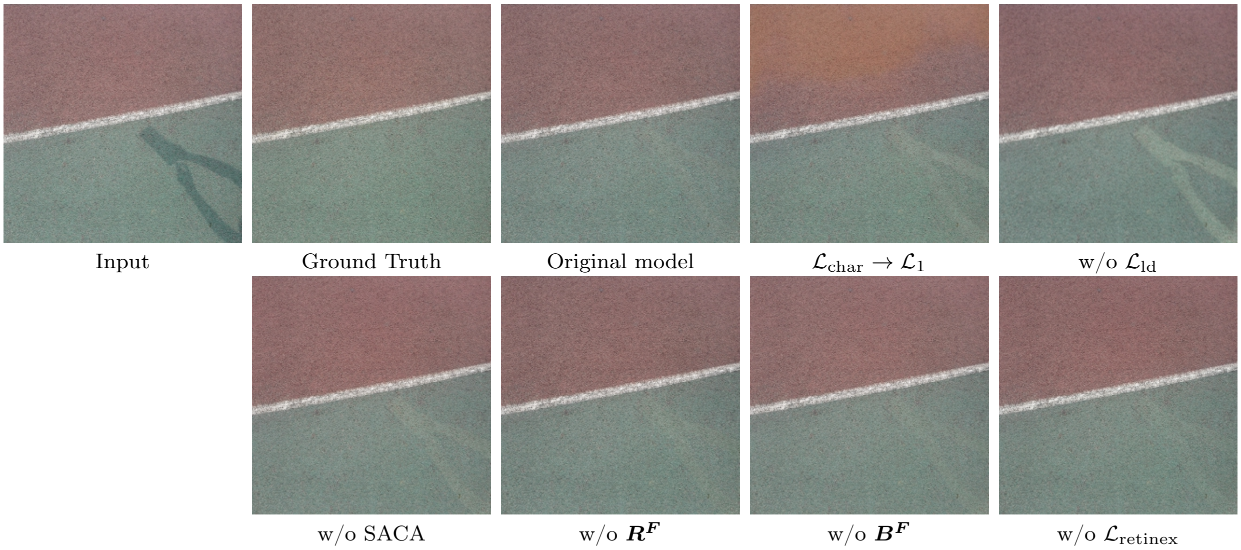



Figure 8: Shadow removal results of different methods on a shadowed image in the ablation study.

- Comparing "w/o SACA" to the original model reveals a PSNR drop of 2.55dB and an RMSE increase of 0.95, indicating SACA's importance to the performance.
- The model for "w/o $\boldsymbol{R^F}$" shows a PSNR drop of 1.37dB in shadow regions and 0.94dB overall, verifying the significance of Retinex decomposition in shadow removal.
- The model for "w/o $\boldsymbol{B^F}$" experiences a PSNR decrease from 37.36dB to 36.89dB in shadow regions and an RMSE increase from 6.38 to 7.21, highlighting the role of local coefficients in handling spatially-varying degradation in shadow removal.
- Replacing $\mathcal{L}_{\text{char}}$ with $\mathcal{L}_1$, lightly degrades performance. In addition, the local discriminative loss $\mathcal{L}_{\text{ld}}$ contributes a 0.55dB PSNR improvement, likely because enhanced boundary consistency brought by $\mathcal{L}_{\text{ld}}$.
- Omitting the Retinex decomposition loss $\mathcal{L}_{\text{retinex}}$ results in a PSNR drop of 0.34dB in shadow regions and 0.26dB overall, probably due to that the Retinex decomposition of features becomes less valid due to the lack of guidance from the loss.

The results of the baselines models on a shadowed image are shown in Fig. 8. The original MSRD model excels in shadow removal, while the "w/o SACA" and "w/o $R^{\boldsymbol{F}}$" models show deficiencies in shadow and non-shadow regions, respectively. Omitting $\boldsymbol{B^F}$ leaves some shadows in the result. Replacing or removing the original loss functions, like $\mathcal{L}_{\text{ld}}$, leads to performance degradation and noticeable artifacts along shadow boundaries. These inspections further validate the effectiveness of key components in our approach.

In practice, the shadow mask is rarely perfect, and there is possible precision loss of the mask through the MGSR blocks after pooling with down-sampling. However, several factors help mitigate this issue. First, the max pooling operation preserves the relative spatial relationships between shadowed and non-shadowed areas, reducing the impact of errors in the shadow mask. Additionally, our RDSL module utilizes both global and local illumination information from non-shadowed regions to guide the illumination recovery process. As a result, minor errors in the shadow mask have limited impact on the overall adjustments, as the global and local illumination information remains reliable.

To assess the robustness of our method to potential mask errors, we conducted experiments using shadow masks with varying degrees of accuracy. As illustrated in Fig. 9, the results show that our method consistently outperforms existing approaches, even with imperfect masks. This demonstrates that a perfectly accurate mask is not necessary, and that our method remains robust to varying levels of mask accuracy while still achieving satisfactory shadow removal.

## 5. Limitations and Potential Solutions

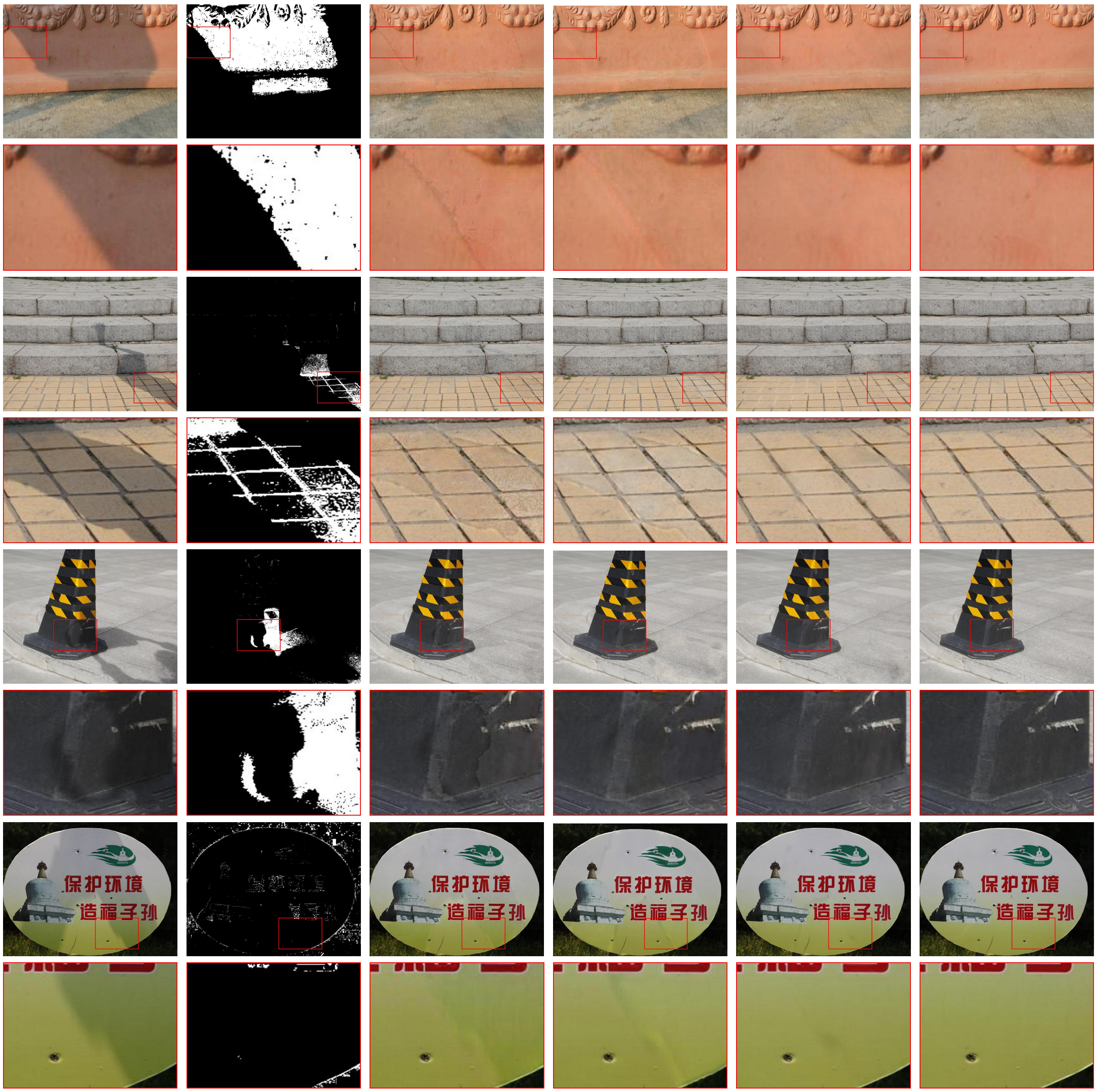
In this section, we discuss the limitations of our shadow removal method and propose potential solutions to address them.

### 5.1. *Accurate Treatment of Shadow Boundaries*

One of the key challenges in image shadow removal is accurately handle shadow boundaries, especially in cases with complex lighting conditions and soft shadows. While the proposed framework improves overall shadow removal quality, it does not work well for all situations. As shown in Fig. 3, some shadow boundary issues remain unresolved. To tackle this limitation, one potential solution is to introduce more flexible, spatially varying illumination parameters that can adaptively adjust to the unique characteristics of shadows, offering a more accurate depiction of how light interacts with surfaces near shadow boundaries. Additionally, incorporating soft shadow masks, which provide a more graduate transition between shadowed and non-shadowed regions, could improve boundary delineation and help the model achieve smoother transitions, preserving both texture and color consistency.

### 5.2. *Robustness to Low-Accuracy Shadow Detection*

The proposed method works well when shadow detection provides masks with reasonable accuracy, allowing for small to modest errors. However, the method struggles to handle large inaccuracies in the shadow mask, which can lead to suboptimal shadow removal results, as shown in Fig. 9. To address this limitation, one potential solution is to develop or integrate more robust shadow detection algorithms capable of handling




Input Shadow Mask UnfoldNet BMNet MSRDNet GT

Figure 9: The result of shadow removal with shadow masks of varying accuracy levels.

diverse lighting conditions, complex scenes, and various object types. Additionally, an iterative refinement process could be employed to improve the shadow mask based on feedback from the shadow removal outcomes. This would involve re-detecting shadows in regions where removal was less effective and adjusting the mask accordingly.

## 6. Conclusion

Image shadow removal remains a challenging task due to the diverse shapes, sizes, and complexities of shadows in real-world environments. In this paper, we proposed a deep Retinex decomposition-based approach to address these challenges, particularly in complex illumination scenarios.

Our method integrates a multi-scale Retinex decomposition model with components like the Retinex decomposition-based shadow lighting (RDSL) and shadow-aware channel attention (SACA). This design allows the model to capture fine-grained details and large-scale contextual information, enabling it to handle spatially varying shadow characteristics while ensuring global lighting consistency. Moreover, the RDSL module adapts shadow illumination using non-shadow information, while the SACA module enhances sensitivity to shadow-related features for accurate recovery. Additionally, the locally discriminative loss improves consistency between shadow and non-shadow regions, especially along boundaries. Extensive evaluations on benchmark datasets show that our approach outperforms existing methods, effectively handling various shadow shapes, sizes, and intensities, making it a robust solution for diverse shadow removal applications.

While our method demonstrated state-of-the-art performance, there are still areas for improvement. One bottleneck in performance lies in accurately addressing shadow boundary artifacts, particularly under complex lighting conditions and soft shadows, which may result in residual artifacts or incomplete shadow removal along edges. Additionally, the method's performance depends heavily on the precision of the input shadow mask, and inaccuracies in shadow mask detection can impact the final output quality.

Summarily, our work offers substantial benefits to the field of computer vision and image processing. By advancing the state-of-the-art in shadow removal, it provides a robust solution that can be integrated into various applications such as image editing, object recognition in shadowed environments, and autonomous navigation systems requiring accurate visual information. The approach also paves the way for further research into multi-scale decomposition techniques and attention mechanisms for better handling of complex lighting scenarios.

For future work, we aim to address the identified weaknesses by developing more adaptive illumination modeling techniques capable of handling complex lighting and soft shadows more effectively. We also plan to explore more robust shadow mask detection methods to reduce dependency on mask precision, potentially integrating shadow detection directly into our model. Additionally, extending our approach to handle dynamic scenes and real-time processing could significantly increase its practical utility in real-world applications.

## 7. Acknowledgment

This work was supported in part by National Natural Science Foundation of China (No.62372186), in part by Natural Science Foundation of Guangdong Province (Nos.2022A1515011755 and 2023A1515012841), in part by Science and Technology Plan Project of Guangzhou(No.2023A04J1681) , in part by Fundamental Research Funds for the Central Universities (No.x2jsD2230220), and in part by Singapore MOE AcRF Tier 1 (No.A-8000981-00-00).

## References

[1] Y. Zhu, Z. Xiao, Y. Fang, X. Fu, Z. Xiong, Z.-J. Zha, Efficient model-driven network for shadow removal, in: Proceedings of the AAAI conference on artificial intelligence, Vol. 36(3), 2022, pp. 3635–3643.

[2] H. Cheong, S. K. Devalla, T. H. Pham, L. Zhang, T. A. Tun, X. Wang, S. Perera, L. Schmetterer, T. Aung, C. Boote, et al., Deshadowgan: a deep learning approach to remove shadows from optical coherence tomography images, Translational Vision Science & Technology 9 (2) (2020) 23–23.

[3] E. Maltezos, A. Doulamis, C. Ioannidis, Improving the visualisation of 3d textured models via shadow detection and removal, in: 2017 9th International Conference on Virtual Worlds and Games for Serious Applications, IEEE, 2017, pp. 161–164.

[4] N. Su, Y. Zhang, S. Tian, Y. Yan, X. Miao, Shadow detection and removal for occluded object information recovery in urban high-resolution panchromatic satellite images, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9 (6) (2016) 2568–2582.

[5] H. Le, D. Samaras, Shadow removal via shadow image decomposition, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 8578–8587.

[6] G. D. Finlayson, S. D. Hordley, M. S. Drew, Removing shadows from images, in: Proceedings of European Conference on Computer Vision, Springer, 2002, pp. 823–836.

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

[8] R. K. Sasi, V. Govindan, Shadow removal using sparse representation over local dictionaries, Engineering Science and Technology, an International Journal 19 (2) (2016) 1067–1075.

[9] C. Xiao, D. Xiao, L. Zhang, L. Chen, Efficient shadow removal using subregion matching illumination transfer, in: Computer Graphics Forum, Vol. 32(7), Wiley Online Library, 2013, pp. 421–430.

[10] L. Jie, H. Zhang, Rmlanet: Random multi-level attention network for shadow detection and removal, IEEE Transactions on Circuits and Systems for Video Technology (2023).

[11] J. Gao, Q. Zheng, Y. Guo, Towards real-world shadow removal with a shadow simulation method and a two-stage framework, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 599–608.

[12] L. Fu, C. Zhou, Q. Guo, F. Juefei-Xu, H. Yu, W. Feng, Y. Liu, S. Wang, Auto-exposure fusion for single-image shadow removal, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 10571–10580.

[13] Z. Liu, H. Yin, X. Wu, Z. Wu, Y. Mi, S. Wang, From shadow generation to shadow removal, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 4927–4936.

[14] X. Hu, Y. Jiang, C.-W. Fu, P.-A. Heng, Mask-shadowgan: Learning to remove shadows from unpaired data, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 2472–2481.

[15] N. Inoue, T. Yamasaki, Learning from synthetic shadows for shadow detection and removal, IEEE Transactions on Circuits and Systems for Video Technology 31 (11) (2020) 4187–4197.

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

[17] E. H. Land, The retinex theory of color vision, Scientific american 237 (6) (1977) 108–129.

[18] Z. Zhao, H. Lin, D. Shi, G. Zhou, A non-regularization self-supervised retinex approach to low-light image enhancement with parameterized illumination estimation, Pattern Recognition 146 (2024) 110025.

[19] F. Jia, H. S. Wong, T. Wang, T. Zeng, A reflectance re-weighted retinex model for non-uniform and low-light image enhancement, Pattern Recognition 144 (2023) 109823.

[20] F. Liu, M. Gleicher, Texture-consistent shadow removal, in: Proceedings of European Conference on Computer Vision, Springer, 2008, pp. 437–450.

[21] Q. Yang, K.-H. Tan, N. Ahuja, Shadow removal using bilateral filtering, IEEE Transactions on Image Processing 21 (10) (2012) 4361–4368.

[22] R. Guo, Q. Dai, D. Hoiem, Single-image shadow detection and removal using paired regions, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2011, pp. 2033–2040.

[23] D. Grest, J.-M. Frahm, R. Koch, A color similarity measure for robust shadow removal in real time, in: VMV, 2003, pp. 253–260.

[24] H. Le, D. Samaras, From shadow segmentation to shadow removal, in: Proceedings of European Conference on Computer Vision, Springer, 2020, pp. 264–281.

[25] X. Cun, C.-M. Pun, C. Shi, Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting gan, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34(7), 2020, pp. 10680–10687.

[26] A. Sanin, C. Sanderson, B. C. Lovell, Shadow detection: A survey and comparative evaluation of recent methods, Pattern recognition 45 (4) (2012) 1684–1695.

[27] A. Leone, C. Distante, Shadow detection for moving objects based on texture analysis, Pattern Recognition 40 (4) (2007) 1222–1233.

[28] T. F. Y. Vicente, M. Hoai, D. Samaras, Leave-one-out kernel optimization for shadow detection and removal, IEEE Transactions on Pattern Analysis and Machine Intelligence 40 (3) (2017) 682–695.

[29] Y. Zhu, X. Fu, C. Cao, X. Wang, Q. Sun, Z.-J. Zha, Single image shadow detection via complementary mechanism, in: Proceedings of the ACM International Conference on Multimedia, 2022, pp. 6717–6726.

[30] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, L. Shao, Learning enriched features for real image restoration and enhancement, in: Proceedings of European Conference on Computer Vision, Springer, 2020, pp. 492–511.

[31] J. Liang, H. Zeng, L. Zhang, Details or artifacts: A locally discriminative learning approach to realistic image super-resolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 5657–5666.

[32] A. B. Petro, C. Sbert, J.-M. Morel, Multiscale retinex, Image Processing On Line (2014) 71–88.

[33] J. Wang, X. Li, J. 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, 2018, pp. 1788–1797.

[34] L. Qu, J. Tian, S. He, Y. Tang, R. W. Lau, Deshadownet: A multi-context embedding deep network for shadow removal, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4067–4075.

[35] H. Gong, D. Cosker, Interactive shadow removal and ground truth for variable scene categories., in: Proceedings of British Machine Vision Conference, 2014, pp. 1–11.

[36] Z. Chen, C. Long, L. Zhang, C. Xiao, Canet: A context-aware network for shadow removal, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 4743–4752.

[37] Y. Zhu, J. Huang, X. Fu, F. Zhao, Q. Sun, Z.-J. Zha, Bijective mapping network for shadow removal, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 5627–5636.

[38] Y. Jin, A. Sharma, R. T. Tan, Dc-shadownet: Single-image hard and soft shadow removal using unsupervised domain-classifier guided network, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 5027–5036.

[39] J. Liu, Q. Wang, H. Fan, W. Li, L. Qu, Y. Tang, A decoupled multi-task network for shadow removal, IEEE Transactions on Multimedia (2023).

[40] J. Liu, Q. Wang, H. Fan, J. Tian, Y. Tang, A shadow imaging bilinear model and three-branch residual network for shadow removal, IEEE Transactions on Neural Networks and Learning Systems (2023).