Technical Section

Estimating intrinsic characteristics of images for shadow removal☆

Yanting Liu [a,1], Zhihao Liu [b,1], Hui Yin [a,*], Jin Wan [c], Zhenyao Wu [d], Xinyi Wu [d], Song Wang [d]

[a] *Beijing Key Laboratory of Traffic Data Analysis and Mining, Beijing Jiaotong University, Beijing, 100044, China*
[b] *Artificial Intelligence and Intelligent Operation R&D Center, China Mobile Research Institute, Beijing, 100053, China*
[c] *Key Laboratory of Beijing for Railway Engineering, Beijing Jiaotong University, Beijing, 100044, China*
[d] *Department of Computer Science and Engineering, University of South Carolina, Columbia, 29201, United States of America*

ARTICLE INFO

ABSTRACT

Existing works show that shadow removal tasks can benefit from the physical illumination model on the formation of shadows. Inspired by prior works that recover the intrinsic characteristics by decomposing an image for its reflectance and illumination components, we study a variant of shadow illumination model that can better reflect the complexity in the real world — in this model, shadow-free pixels can be expressed by a translation formed of reflectance and illumination components. Based on the new illumination model, we develop a new LR-ShadowNet, which contains two sub-nets for estimating the illumination and reflectance components, respectively, and one sub-net for refining the shadow-removal result. Besides, several mask guidance module are incorporated into LR-ShadowNet for guiding components estimation and result refinement based on shadow region information. The whole network is trained in an end-to-end fashion guided by the shadow masks. Extensive experiments on the ISTD dataset and SBU-Timelapse dataset show that the proposed LR-ShadowNet achieves competitive performance with less computational cost and strong generalisation ability.

1. Introduction

Caused by the direct light on an object, shadows are a very common phenomenon in natural images and can introduce additional complexity and difficulty to many computer vision tasks [1–8]. To relieve this issue, many shadow-removal approaches have been developed in the past decades, aiming to transfer the shadow regions in an image to its non-shadow version [9–20].

Traditional methods [11,21–24] remove shadows based on the physical shadow models which represent a scene in terms of intrinsic characteristics. Recently, deep learning algorithms have shown remarkable performance on shadow removal by training on large-scale datasets [12–14,20]. However, most of them learn the transformation from a shadow image to its corresponding shadow-free image without considering the actual physical properties of shadows, leading to blurry textures and artefacts in the shadow-removal results.

To address this problem, the idea of image decomposition has recently been incorporated into shadow removal based on the illumination model [16,25] — a shadow image can be decomposed to a shadow-free image, shadow and a shadow density matte, while the shadow and the shadow density matte are estimated by deep neural networks. Based on a simplified illumination model, it is assumed that the colour transform from shadow to shadow free is affine and all shadow pixels in the umbra share the same set of components.

However, this simplified model does not consider the complicated illumination in real world, preventing the network from producing more accurate shadow removal. More specifically, illumination in a real scene can come from direct sunlight, diffuse skylight, and light reflected by other objects [26]. There is no guarantee that camera can record an ideal umbra shadow — shadow is formed by lacking direct illumination and a portion of the ambient illumination, and these two types of illumination are heterogeneous in natural scenes. To address this problem, in this paper we propose to estimate components for *each pixel* separately to produce more realistic shadow-free images.

In particular, we present a new version of shadow illumination model for achieving pixel-level transformation from shadow data to shadow-free data. This is achieved by estimating illumination and reflectance components, and assuming the intensity transform between shadow and shadow-free pixels is a translation formed of these two components in Lab colour space, as illustrated in Fig. 1. Based on the accurate and intrinsic-related (illumination and reflectance) shadow illumination model, a more realistic intensity transform is established,
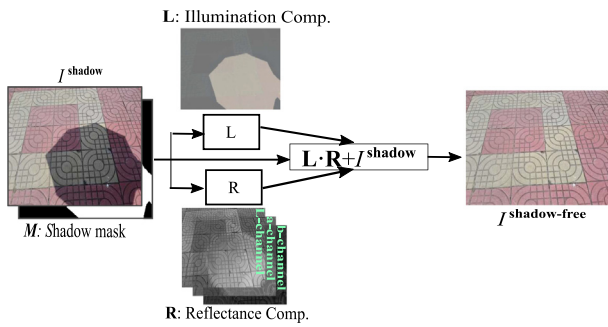

Fig. 1. An illustration of our basic idea of shadow removal via a novel version of shadow formation model. Given a shadow image I^{shadow} and its shadow mask M, we express the shadow-free image and reflectance components $\mathbf{R}$ in Lab colour space. $\mathbf{L}$ and $\mathbf{R}$ are predicted via two sub-nets $\mathcal{L}$ and $\mathcal{R}$ respectively, constrained by the physical properties of shadows.

and it is no longer necessary to assume that the umbra shadow is uniformly cast. Accordingly, we propose a novel LR-ShadowNet that consists of three sub-nets for the purpose of components estimation and result refinement. Specifically, given a shadow image and its shadow mask, two sub-nets are employed to estimate components of illumination and reflectance for shadow removal, respectively. And a third sub-net is included for refining shadow-removal results since the shadow illumination model may not always be rigorous in some special cases.

In addition, previous shadow removal methods [16,18,20,25] direct concatenate the image and mask as the input for shadow removal, and their performance depends on the quality of shadow masks because the shadow mask are not always exactly aligned with the shadow image. To alleviate this problem, we regard the shadow mask as a region prior and embed the mask at the feature level rather than the input level via several Mask Guidance Modules (MGM). Each MGM consists of three spatial feature transform blocks [27] conditioned on the shadow mask and are incorporated into the three sub-nets of our framework.

We evaluate the proposed shadow illumination model and shadow-removal method by training and testing on the ISTD dataset and show the generalisation ability of the trained model on the SBU-Timelapse dataset. Above all, the main contributions of this work can be summarised as:

- We present a new version of shadow illumination model for shadow removal, which accurately describes each shadow-free pixel by a translation model involving the illumination components, reflectance components, and the corresponding shadow pixel.
- We develop LR-ShadowNet, a novel supervised shadow removal network based on the proposed shadow illumination model, that consists of three sub-nets for illumination and reflectance components estimation and result refinement, respectively, together with new Mask Guidance Modules (MGM) for better utilising the shadow mask.
- Extensive experiments on two widely-used shadow removal datasets, ISTD and SBU-Timelapse, demonstrate that the proposed LR-ShadowNet achieves competitive performance with less computational cost.

2. Related work

2.1. Shadow illumination model

Our method is based on the image formation model proposed in [28] where the illumination of a shadow-free pixel is a combination of the diffuse and point-source illumination, and the intensity of the

shadow-free pixel is represented by a sum of the illumination and reflectance components. A line of works based on this model have been proposed for shadow removal. In [22], a sequence of images is decomposed into a time-varying reflectance image and its corresponding illumination images, with which the shadow effects can be eliminated. In [23], an affine illumination model for shadow removal is developed by expressing each shadow-free pixel as an affine transformation to the shadow pixel with the estimation of four unknown components. This affine illumination model has been further extended [24] by considering the reflectance variation, with which an adaptive illumination transfer approach is proposed for shadow removal. In [11], a simplified version of the affine illumination model for shadow removal is proposed by expressing each shadow-free pixel in terms of direct light, environment light, and surface reflectance of the pixel. Based on the affine illumination model [23], Le & Samaras [25] developed a deep-learning based approach to produce shadow-free images. Different from these works, in this paper we consider the intrinsic components of the image and use a translation illumination model on all the decomposed components for improving shadow-removal performance.

2.2. Deep-learning-based shadow removal

Deep-learning-based methods trained on large-scale datasets significantly boost the performance of shadow removal [13–15,25,29,30]. Many of them train their networks using paired shadow and shadow-free data. Wang et al. [13] proposed a stacked conditional generative adversarial network to jointly detect and remove shadows. Hu et al. [14] developed a direction-aware spatial context method by considering the direction information of shadows for improving shadow-removal performance. Le & Samaras [25] proposed a shadow removal network based on a shadow image decomposition model for helping estimate the shadow-free images, and later the network is improved by adding an inpainting network to refine the shadow-removal results [20]. Cun et al. [15] developed a dual hierarchical aggregation network and shadow matting GAN for shadow detection and removal, which can produce high-quality ghost-free images with the help of synthesised masks or scenes. Fu et al. [29] proposed an auto-exposure fusion network, which treats the shadow removal as an exposure fusion problem and automatically selects proper pixels from multiple over-exposure images to form the final output. [30] proposed a novel depth-aware shadow removal method, which estimates depth information from RGB images and leverages the depth feature as guidance to enhance shadow removal and refinement.

Due to the difficulty of constructing paired shadow and shadow-free data, recent works [17,31] also tried to train a deep network on unpaired data. Hu et al. [31] proposed Mask-ShadowGAN based on CycleGAN [32] to learn the mapping between the shadow-free domain and the shadow domain. Liu et al. [17] improved Mask-ShadowGAN by introducing lightness features to guide the learning for shadow removal. Another line of works trains the deep network without using shadow-free images [16,18,25,33]. Le & Samaras [16] constructed unpaired data in form of patches cropped from the shadow images based on shadow masks to train a weakly-supervised shadow removal model. Liu et al. [18] developed a shadow generation model to construct paired data using only shadow images and the corresponding shadow masks. [33] introduced a new unsupervised single-image shadow removal network DC-ShadowNet, which is guided by a domain classifier to focus on shadow regions. As mentioned earlier, in [16,25], a simplified physical shadow illumination model was incorporated into deep networks to help recover shadow-free images. In this paper, we continue along this line by proposing a novel shadow illumination model that can better reflect the complexity in real world and incorporating this model into deep neural networks for improving shadow-removal performance.
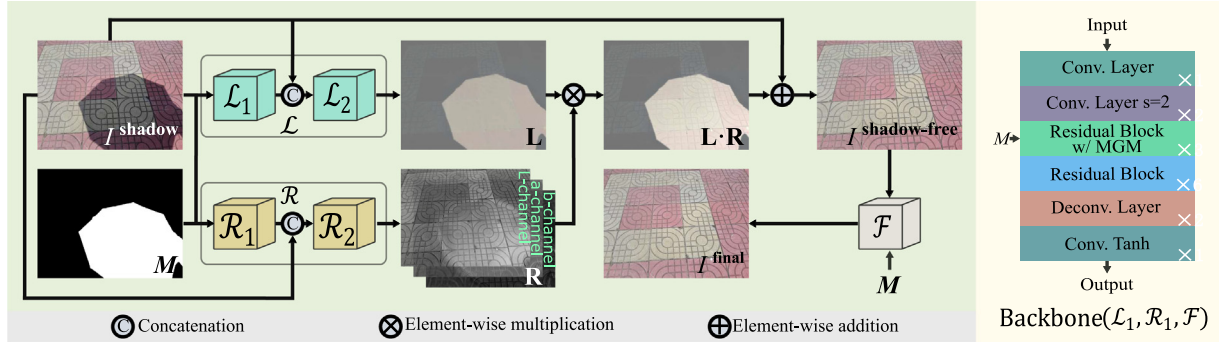
Fig. 2. Left: The network architecture of the proposed LR-ShadowNet. It takes as input a shadow image and the corresponding shadow mask, and finally produces the shadow-removal result. The training of the network involves one sub-net $\mathcal{L}$ for illumination components estimation, one sub-net $\mathcal{R}$ for reflectance components estimation, one sub-net $\mathcal{F}$ for result refinement, and four types of loss functions. Right: The backbone of $\mathcal{L}_1$, $\mathcal{R}_1$, and $\mathcal{F}$ with Mask Guidance Modules (MGM), where 's = 2' in 'Conv. Layer' represents the stride of convolution is 2.

2.3. Intrinsic image decomposition

Intrinsic image decomposition is a very classic task, and many related studies have emerged [28,34–37]. [28], as the theoretical basis of many shadow removal methods, mainly explained that the intensity of the shadow-free pixel is represented by a sum of the illumination and reflectance components, and the illumination of a shadow-free pixel is a combination of the diffuse and point-source illumination. [34] designed a new calculation method for sky, indirect, and sun illumination (ignoring projected shadows) by using incomplete and imprecise geometric shapes and a small set of input images for automatically reconstructing 3D points. In [35], the shading image is decomposed into a direct irradiance component, an indirect irradiance component, and a colour component, which account for different aspects of image formation. [37] proposed to decompose the shading component into direct (illumination) and indirect shading (ambient light and shadows) subcomponents. From the above discussions, we can see that intrinsic image decomposition shows powerful advantages in many tasks. Thus, in this paper, we proposed a new version of shadow illumination model for shadow removal. based on intrinsic image decomposition.

3. Methodology

3.1. Shadow illumination model

The seminal work about intrinsic image decomposition [28] laid the foundation for many existing shadow models [22,23,25]. Based on the one of the affine illumination models [23], Le & Samaras [25] developed a deep-learning based shadow-removal approach. Therefore, firstly revisiting the shadow illumination model proposed in [25], the intensity of a shadow-free pixel reflected from point x can be written as:

$$I_x^{\text{shadow-free}}(\lambda) = L_x^d(\lambda)R_x(\lambda) + L_x^a(\lambda)R_x(\lambda), \tag{1}$$

where L_x^d, L_x^a and R_x are the direct illumination, ambient illumination and reflectance, respectively, and λ is the wavelength of light. In the shadow illumination model [25], it is assumed that all direct illumination and some ambient illumination are blocked in the shadow area. Therefore, the intensity of a shadow pixel can be written as:

$$I_x^{\text{shadow}}(\lambda) = a_x(\lambda)L_x^a(\lambda)R_x(\lambda), \tag{2}$$

where $a_x(\cdot)$ represents the attenuation factor which indicates the remaining fraction of the ambient illumination that arrives at point x. According to Eqs. (1) and (2), the intensity of a shadow-free pixel can be expressed as a translation of the intensity of its shadow counterpart:

$$I_x^{\text{shadow-free}}(\lambda) = \left(L_x^d(\lambda) + (1 - a_x(\lambda))L_x^a(\lambda)\right)R_x(\lambda) + I_x^{\text{shadow}}(\lambda). \tag{3}$$

Following [25,38], we assume that Eq. (3) remains valid in the colour acquisition process with cameras. However, instead of representing the affine shadow components with fixed values as in [25], we learn the illumination and reflectance components for each pixel to construct the shadow illumination model as:

$$I^{\text{shadow-free}} = \mathbf{L} \cdot \mathbf{R} + I^{\text{shadow}}, \tag{4}$$

where $\mathbf{L} = L_x^d(\lambda) + (1 - a_x(\lambda))L_x^a(\lambda)$ and $\mathbf{L} \in \mathbb{R}^{C \times W \times H}$ stands for the illumination components, which can be treated as a component-wise amplification term. Similarly, $\mathbf{R} = R_x(\lambda)$, and $\mathbf{R} \in \mathbb{R}^{C \times W \times H}$ stands for the reflectance components. $I^{\text{shadow}} \in \mathbb{R}^{C \times W \times H}$ and $I^{\text{shadow-free}} \in \mathbb{R}^{C \times W \times H}$ represent shadow image and the corresponding shadow-free image, respectively. $\mathbf{L} \cdot \mathbf{R}$ is the residual between the shadow image and shadow-free image. C, W, and H represent the number of channels, the width, and the height, respectively. Here, the number of channels C is set as 3.

3.2. Shadow removal framework

The overall architecture of the proposed LR-ShadowNet is illustrated in Fig. 2, which contains three sub-nets for components estimation and prediction refinement. In the following paragraphs, we illustrate the detailed function of each sub-net, the backbone, and the mask guidance module. It is proved that the Lab colour space is more suitable for shadow removal tasks in [17]. Therefore, except for the illumination components, all the components involved in the proposed shadow illumination model (the shadow images, the shadow-free images and the reflectance components) are in the Lab colour space.

3.2.1. Reflectance estimation sub-net

$\mathbf{R}$ is the reflectance components, which can be regarded as imaging under normal illumination. To extract features related to the reflectance and predict the reflectance components $\mathbf{R}$, we train a sub-net $\mathcal{R}$ that takes as input the shadow image I^{shadow} and its shadow mask M and outputs $\mathbf{R}$. Here, the value range of $\mathbf{R}$ to be $(-1, 1)$ is approximately the same as the normalised intensity range of the input shadow image. The hyperbolic tangent function ($tanh(\cdot)$) is applied at the end of $\mathcal{R}$ to control the output range of the reflectance components $\mathbf{R}$.

3.2.2. Illumination estimation sub-net

The illumination components $\mathbf{L}$, are related to the missing illumination in the shadow region, which can be inferred from the illumination in the non-shadow region and the shadow region of the shadow image. $\mathbf{L}$ is a component-wise amplification term, which only scales the intensity of each Lab channel instead of being a colour itself with potentially negative a or b components.

The range of values of the shadow-removal result, the shadow image and reflectance components $\mathbf{R}$ are all in $(-1,1)$. Therefore, in order to

make Eq. (4) hold, the upper-bound of $\mathbf{L}$ cannot exceed 2, otherwise the three channel values of the output shadow-free image cannot be guaranteed to be in $(-1,1)$, that is, the expected shadow-free image in the Lab colour space cannot be obtained. At the same time, the value range of $\mathbf{L}$ is set to $[1,2]$ in our experiments based on three considerations: (1) Restricted by the physical property of illumination, the values of $\mathbf{L}$ are non-negative; (2) The value ranges of $I^{\text{shadow-free}}$, I^{shadow}, $\mathbf{R}$ are all $(-1,1)$. According to Eq. (4), the upper-bound of $\mathbf{L}$ should be at least 2; (3) The value range of $\mathbf{L}$ should not be overlapped with the value range of $\mathbf{R}$, i.e., $(-1,1)$ — otherwise it may cause ambiguity in the estimated illumination and reflectance components which are hard to distinguish when lacking supervision of ground-truth intrinsic components. When both $\mathbf{L}$ and $\mathbf{R}$ are negative, the residual between the shadow image and shadow-free image $\mathbf{L} \cdot \mathbf{R}$ may be forced to be non-negative and flipped to opposing colours, which is undesirable and even prohibited.

The illumination components $\mathbf{L}$ lack the corresponding supervised signal, so it is difficult to ensure that $\mathbf{L}$ does not deviate from the original input shadow image. Therefore, we employ the perceptual loss [39] to encourage the texture of the illumination components $\mathbf{L}$ to be the same as the ground-truth shadow-free image $\hat{I}^{\text{shadow-free}}$ during training. The perceptual loss is defined as:

$$l_{\text{perc}} = \ell_2(\phi_{\mathbf{L}}, \phi_{\hat{I}^{\text{shadow-free}}}), \tag{5}$$

where $\phi_{\mathbf{L}}$ and $\phi_{\hat{I}^{\text{shadow-free}}}$ are the feature maps extracted from a pre-trained VGG19 [40] model, which takes as input the reflectance components and the shadow-free image respectively. Since the numerical range of $\mathbf{L}$ and $\hat{I}^{\text{shadow-free}}$ are inconsistent, so we choose the feature maps from the final convolutional layer of VGG19 and empirically set the weight of this loss to 0.001 to alleviate this problem.

We use a sub-net $\mathcal{L}$ that takes as input the shadow image I^{shadow} and its shadow mask M to predict illumination components $\mathbf{L}$. The network architecture of $\mathcal{L}$ is the same as $\mathcal{R}$ except for the output range. Specifically, the sub-net $\mathcal{L}$ first converts I^{shadow} and M to features $f_{\mathcal{L}}$, and then $\mathbf{L}$ is obtained by

$$\mathbf{L} = 0.5 \cdot tanh(f_{\mathcal{L}}) + 1.5, \tag{6}$$

where $tanh()$ maps $f_{\mathcal{L}}$ to $(-1,1)$.

3.2.3. Intermediate shadow-removal result

According to Eq. (4), we can obtain the residual of the shadow image and the shadow-free image by multiplying the estimated illumination components $\mathbf{L}$ and reflectance components $\mathbf{R}$, and further calculate the intermediate shadow-removal result $I^{\text{shadow-free}}$. But in essence, decoupling the residual into $\mathbf{L}$ and $\mathbf{R}$ is ill-posed. Significantly, our proposed method provides a heuristic solution to the shadow removal task. Firstly, to ensure the accuracy of the estimated residual, a per-pixel reconstruction loss between $I^{\text{shadow-free}}$ and the ground-truth $\hat{I}^{\text{shadow-free}}$ are computed to jointly train the above two sub-nets:

$$l_{rec1} = \ell_1(I^{\text{shadow-free}}, \hat{I}^{\text{shadow-free}}). \tag{7}$$

To encourage the spatial consistency of the intermediate result $I^{\text{shadow-free}}$, we also apply the spatial consistency loss [41] for training the two sub-nets:

$$l_{\text{spa1}} = \frac{1}{K} \sum_{i=1}^{K} \sum_{j \in \omega(i)} \left(|(Y_i, Y_j)| - |(V_i, V_j)| \right)^2, \tag{8}$$

where K represents the number of local regions that are obtained by dividing the given image using a 4×4 sliding window without overlap. $\omega(i)$ denotes the four neighbouring regions (left, right, top, bottom) centred at the region i, Y and V are the average values of the local region, with a region size of 4×4, of $I^{\text{shadow-free}}$ and $\hat{I}^{\text{shadow-free}}$, respectively.

Since the shadow region and its adjacent region in the shadow removal result usually differ greatly, we further employ the area loss [18]

to train the two sub-nets, which ensures the content of $I^{\text{shadow-free}}$ and $\hat{I}^{\text{shadow-free}}$ to be the same in adjacent shadow areas:

$$l_{\text{area1}} = \ell_1\left(\psi(M) \cdot I^{\text{shadow-free}}, \psi(M) \cdot \hat{I}^{\text{shadow-free}}\right), \tag{9}$$

where ψ denotes the image dilation function with a kernel size of τ. Here, τ is set to 50 empirically. The dot operation between the dilated mask and the input image means we only keep the losses in the region of dilated mask.

3.2.4. Result refinement sub-net

According to Eq. (4) and the estimated components, we can obtain an intermediate shadow-removal result of the given shadow image. However, due to the lack of strict supervision for the illumination and reflectance components and the limitations of the physical model [25], there might be a deviation of components estimation, leading to inaccurate shadow removal results. Therefore, another sub-net $\mathcal{F}$ is applied to perform the result refinement by taking as input the intermediate shadow-removal result $I^{\text{shadow-free}}$ and shadow mask M.

The output of $\mathcal{F}$ is the final shadow-removal result I^{final}. The per-pixel loss L_{rec2}, the spatial consistency loss L_{spa2}, and the shadow area loss l_{area2} between I^{final} and $\hat{I}^{\text{shadow-free}}$ are computed for training all the three sub-nets:

$$l_{\text{rec2}} = \ell_1(I^{\text{final}}, \hat{I}^{\text{shadow-free}}), \tag{10}$$

$$l_{\text{spa2}} = \frac{1}{K} \sum_{i=1}^{K} \sum_{j \in \omega(i)} \left(|(X_i, X_j)| - |(V_i, V_j)| \right)^2, \tag{11}$$

$$l_{\text{area2}} = \ell_1(\psi(M) \cdot I^{\text{final}}, \psi(M) \cdot \hat{I}^{\text{shadow-free}}), \tag{12}$$

where X is the average values of the local region of I^{final}.

3.2.5. Backbone

The backbone of each proposed sub-net is based on the backbone in [17], which consists of three convolutional layers, nine residual blocks, two deconvolutional layers, and one convolutional layer with a hyperbolic tangent function $tanh()$. The illumination and reflectance sub-nets, $\mathcal{L}$ and $\mathcal{R}$, contain two backbones for each, as shown in the left of Fig. 2. Taking $\mathcal{L}$ as an example, $\mathcal{L}_1$ and $\mathcal{L}_2$ are used for region-aware representation extraction and components estimation, respectively. The architecture of $\mathcal{L}_1$ is shown in the right of Fig. 2. It takes as input the shadow image I^{shadow} and its shadow mask, and produces a region-aware representation $f'_{\mathcal{L}}$. $f'_{\mathcal{L}}$ is then concatenated with I^{shadow} and fed into $\mathcal{L}_2$ for components estimation. The refinement sub-net $\mathcal{F}$ contains one backbone.

3.2.6. Mask guidance module

The Mask Guidance Module (MGM) regards the shadow mask M as a prior and performs adaptive feature extraction in the feature space to better utilise the shadow mask. This allows it to be less sensitive to the accuracy of shadow masks on the shadow boundary. Specifically, We incorporate several MGM with M as guidance into each sub-net ($\mathcal{L}_1, \mathcal{R}_1, \mathcal{F}$). For each sub-net, we insert three MGM into the first three residual blocks (in the right of Fig. 2), which is similar to LG-ShadowNet [17] that inserts lightness guidance to the first three residual blocks of each sub-net. In each module, M is first transformed to three groups of modulation components, (α_1, β_1), (α_2, β_2), (α_3, β_3), and then applied to the corresponding feature maps (F_1, F_2, F_3) of each residual block via three spatial feature modulation (SFT) blocks [27], respectively. Each SFT block can be written as:

$$\hat{F}_i = (\alpha_i + 1) \cdot F_i + \beta_i, i = 1, 2, 3, \tag{13}$$

where $\hat{F}_i$ represents the output feature maps of the ith residual block and is fed to the next residual block. The SFT block consists of two groups of convolution, and each of them contains three convolutional layers with the channel number $\{C, 2C, C\}$, respectively, where C is same as the channel number of F_i. We choose the SFT block because it can help the network produce realistic results [27].

3.3. Loss function

Three sub-nets constitute the whole shadow removal framework LR-ShadowNet which is jointly trained with the following loss function:

$$l_{\text{total}} = l_{\text{rec1}} + l_{\text{rec2}} + l_{\text{area1}} + l_{\text{area2}} + \omega_1 l_{\text{perc}} + \omega_2(l_{\text{spa1}} + l_{\text{spa2}}), \quad (14)$$

where ω_1 and ω_2 are empirically set to 0.001 and 10 in experiments, respectively.

4. Experiments

4.1. Datasets and evaluation metrics

ISTD dataset [13,25] contains 1870 triplets of shadow, shadow-free, and shadow mask images that are collected under various illumination conditions and shadow shapes, where 1330 triplets are used for training and the rest 540 for testing. The resolution of each image is 480×640. Following [25], the shadow-free images in ISTD [13] are adjusted to reduce the colour difference between the shadow image and the corresponding shadow-free images. We use the BDRAR shadow detector [42] that is trained on the SBU [43] and ISTD datasets, to obtain the shadow masks for testing. The balance error rate between the obtained shadow masks and the ground-truth shadow masks in ISTD is 2.4.

SBU-Timelapse dataset [20] is used for evaluating the generalisation ability of shadow removal models. All 50 videos in this dataset are captured at static scenes with only moving shadows. Given a threshold of the intensity difference at each pixel location across the whole video, pixel locations whose intensity is larger than the threshold has both the shadow pixel and shadow-free pixel, and these pixel locations are used for evaluation. Following the setting of [20], we use a threshold of 80 to obtain the evaluation regions. We use BDRAR [42] trained on the SBU dataset to generate the shadow mask.

Evaluation metrics. For all experiments, we use the Mean Absolute Error (MAE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity (SSIM) as the evaluation metrics. The MAE is widely used in previous shadow removal methods as an evaluation metric [13,31], which computes the mean absolute error between the shadow-removal results and the shadow-free images in the Lab colour space for which the lower the better. PSNR and SSIM scores are the higher the better. All the images are resized to 256×256 for evaluation.

4.2. Implementation details

The proposed LR-ShadowNet is implemented using PyTorch and trained with a single NVIDIA GeForce GTX 2080ti GPU. The initial weights of the three sub-nets are based on Gaussian distribution with a mean of 0 and a standard deviation of 0.02. They are jointly optimised using Adam with the first and the second momentum setting to 0.5 and 0.999, respectively and a batch size of 1. The whole model is trained for 100 epochs in total with an initial learning rate of 2×10^{-4} for the first 50 epochs which is decayed linearly to 0 for the rest 50 epochs. For data augmentation, we perform random flipping and random cropping with a crop size of 400×400. In total, In total, it takes about 15 h to train the whole model and 9 min to test on the adjusted ISTD.

4.3. Ablation study

Shadow illumination model. To verify the effectiveness of the presented shadow illumination model in LR-ShadowNet, we try two variants based on two different physical illumination models: (1) the linear formation of shadow illumination model used in [23,25]:

$$I^{\text{shadow-free}} = \mathbf{w} I^{\text{shadow}} + \mathbf{b}, \quad (15)$$

Table 1
Ablation study to verify the effectiveness of the proposed shadow illumination model on the testing set of the adjusted ISTD using all three evaluation metrics. MAE is computed following the way of [20]. Hereafter, 'Shadow Region' and 'All' indicate the evaluations in the shadow region and over the whole image, respectively.

Model	Shadow Region			All		
	MAE	PSNR	SSIM	MAE	PSNR	SSIM
Input Image	39.0	20.83	0.927	8.4	20.46	0.894
Eq. (15)	8.2	33.96	0.987	4.0	30.41	0.956
Eq. (16)	6.3	37.01	0.989	3.4	32.92	0.962
LR-ShadowNet	**5.9**	**37.45**	**0.990**	**3.3**	**33.00**	**0.963**

Table 2
Quantitative results of the proposed model trained with different value ranges of L on the testing set of adjusted ISTD.

Value range	Shadow Region			All		
	MAE	PSNR	SSIM	MAE	PSNR	SSIM
$(-1, 1)$	6.7	36.13	0.989	3.5	32.18	0.961
$(0, 2)$	6.3	37.03	0.989	**3.3**	32.91	**0.963**
$(1, 2)$	**5.9**	**37.45**	**0.990**	**3.3**	**33.00**	**0.963**
$(1, 5)$	6.4	36.99	**0.990**	3.4	32.80	0.962

where $\mathbf{w}$ and $\mathbf{b}$ are pixel-wise parameters; and (2) a simpler version of our illumination model:

$$I^{\text{shadow-free}} = \mathbf{T} + I^{\text{shadow}}, \quad (16)$$

where $\mathbf{T}$ represents the compensation values for transforming shadow pixels to shadow-free pixels. We use one sub-net that has the same architecture as the $\mathcal{L}$ to estimate each of these components. The results are summarised in Table 1.

The result in the first row serves as a reference by directly taking the original shadow images for evaluation. Compared with this reference, all variants of our method can effectively remove shadows and significantly improve the performance in three metrics. From the second row, we observe that the first variant based on the affine illumination model (row 2, Eq. (15)) is inferior to the other two variants based on translation illumination models (rows 3 and 4), in both of which I^{shadow} has no pixel-wise multiplicative term. The results show the advantage of the translation illumination model over the affine one. Furthermore, we also find that the variant that learns two separate intrinsic components is superior to the variant that directly learns the whole translation term.

Value range of L. Next, we conduct ablation studies to justify the effectiveness of the predetermined value range of **L** on the testing set of adjusted ISTD. Quantitative results are shown in Table 2. It can be seen that setting the value range to be the same as the normalised image, $(-1, 1)$, achieves the worst results. This verifies the necessity of the non-negative range for **L**. Results of ranges $(0, 2)$ and $(1, 2)$ show that separating the value ranges of **L** and **R** is effective. Results in the last row demonstrate that the value range of **L** should have an appropriate upper bound for obtaining the best performance.

Mask guidance module. We study the impact of the shadow mask and the mask guidance modules used in LR-ShadowNet, and report the quantitative results of the following variants on the testing set of adjusted ISTD in Table 3: (1) removing all the mask guidance modules (trained without shadow masks), (2) removing all the mask guidance modules and concatenating the shadow image and its shadow mask as input. Results in the first row show that the proposed method trained without shadow masks is still effective. Results in rows 2–3 show that utilising region information in the feature level via mask guidance modules is superior to simply concatenating the shadow image and its shadow mask in the input level, which proves the effectiveness of the mask guidance module. In addition, we have added a set of visual comparison results on the ablation of the MGM in Fig. 3 and it can
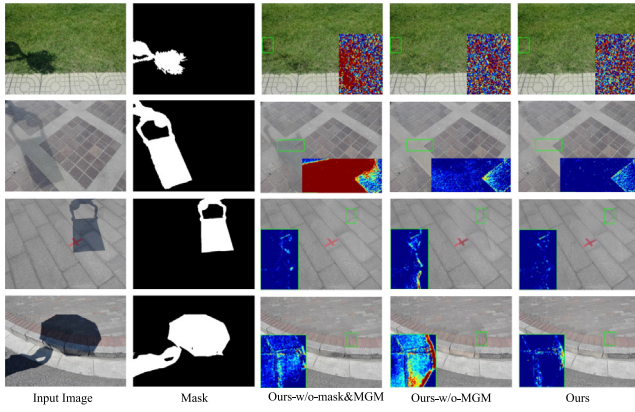
| Input Image | Mask | Ours-w/o-mask&MGM | Ours-w/o-MGM | Ours |

Fig. 3. Visualisation comparisons with and without MGM on ISTD. Results are shown with both the shadow-removal results and the heat maps of the difference between the shadow-removal result and shadow-free ground truth for better comparison.

Table 3
Ablation study on the effectiveness of Mask Guidance Modules (MGM) on the adjusted ISTD. M represents whether the model takes as input the shadow mask.

Model		Shadow Region			All		
M	**MGM**	MAE	PSNR	SSIM	MAE	PSNR	SSIM
×	×	6.6	36.72	0.989	3.8	32.10	0.959
√	×	6.4	36.83	0.989	3.4	32.77	0.962
√	√	**5.9**	**37.45**	**0.990**	**3.3**	**33.00**	**0.963**

Table 4
Quantitative results of the proposed model trained with different loss functions and network architectures on the testing set of adjusted ISTD.

Model		Shadow Region			All		
		MAE	PSNR	SSIM	MAE	PSNR	SSIM
Loss	w/o l_{perc}	6.4	37.06	0.989	3.4	32.89	0.962
	w/o l_{spa1}, l_{spa2}	6.3	37.01	0.989	3.4	32.86	0.962
	w/o l_{area1}, l_{area2}	6.2	37.16	**0.990**	3.4	32.94	**0.963**
Net. Arch.	w/o $\mathcal{L}, \mathcal{R}$	9.2	33.21	0.986	3.9	30.10	0.956
	w/o $\mathcal{L}_1, \mathcal{R}_1$	7.1	35.45	0.988	3.8	31.28	0.958
	w/o $\mathcal{L}_2, \mathcal{R}_2$	7.4	34.86	0.988	3.6	31.43	0.960
	w/o $\mathcal{L}_2$	7.3	35.21	0.988	3.6	31.55	0.960
	w/o $\mathcal{R}_2$	6.3	37.08	0.989	**3.3**	32.84	**0.963**
	w/o $\mathcal{F}$	6.7	36.73	0.989	3.4	32.73	0.962
LR-ShadowNet		**5.9**	**37.45**	**0.990**	**3.3**	**33.00**	**0.963**

be seen that the shadow removal results obtained from LR-ShadowNet with MGM are better, which also proves the effectiveness of MGM;

Loss functions and Network architectures. We also conduct ablation studies to verify the effectiveness of loss functions and the network architecture, as shown in Table 4. It can be seen from the first three rows that the perceptual loss l_{perc}, the spatial consistency loss l_{spa1}, l_{spa2}, and the shadow area loss l_{area1}, l_{area2} can improve the shadow-removal performance. Results in rows 4–10 prove the effectiveness of the proposed network architecture. We can find that the two backbones are important for estimating intrinsic components, without each of them the MAE drops significantly (rows 4–6). And the second backbones $\mathcal{L}_2, \mathcal{R}_2$, especially $\mathcal{L}_2$, are important for estimating intrinsic components, since illumination is more relevant to the shadow/non-shadow regions (rows 7–8). The results in row 9 show that $\mathcal{F}$ plays an important role in results refinement.

4.4. Comparison with state-of-the-arts

We first compare our method with several state-of-the-art methods on the adjusted ISTD: (1) traditional methods utilising image priors

for shadow removal: Yang et al. [44], Gong & Cosker [45]; (2) unpaired methods that learn shadow removal from unpaired shadow and shadow-free images: Mask-ShadowGAN [31], LG-ShadowNet [17]; (3) weakly-supervised methods trained with paired shadow images and shadow masks: Le & Samaras [16], G2R-ShadowNet [18]; (4) fully-supervised methods trained with paired shadow and shadow-free images, as well as shadow masks: ST-CGAN [13], DHAN [15], SP+M+I-Net [20], the fully supervised G2R *Sup.* [18], the auto-exposure fusion method [29] and G2C-DeshadowNet [47].

As shown in Table 5, our method achieves the best performance in terms of all the metrics over the whole image, and the MAE achieved by our method is 5.7% lower than the state-of-the-art method SP+M+I-Net. At the same time, our method achieves better scores on all evaluation metrics in non-shadow regions and the all image than SP+M+I-Net, which indicates that our method can improve the visualisation effect in the shadow regions without destroying the visualisation effect of the non-shadow regions, so the overall shadow removal effect of our method is better. In the shadow region, we obtain 19.2% gains of MAE than G2R *Sup.* [18] that used a similar backbone to our method. Besides, our method also has significant advantages in shadow regions than Einy et al. [46]. In the non-shadow region, the results produced by our method are closest to the shadow-free ground truth among all the deep-learning-based methods. The qualitative comparison results are shown in Fig. 4 with four samples from the testing set of adjusted ISTD. Compared with the results produced by other methods, the colour and edge of the shadow region in our results are more consistent with that of the ground truth for all four samples. Besides, we also present the visual comparisons of the results before and after $\mathcal{F}$ in Fig. 5. Especially from the second and fourth row of images, it can be seen that the shadow-removal results after $\mathcal{F}$ are smoother at the shadow edges and have better visual effects, which further demonstrated the effectiveness of our proposed result refinement sub-net $\mathcal{F}$. Finally, we provide some failure shadow removal samples on ISTD in Fig. 6, and it can be observed that our method does not perform well in some shadow areas that cover multiple colours (the 3–5 rows in Fig. 6), which may be due to the difficulty of removing complex shadows in our method.

We also design three variants with significant reductions in the number of parameters and FLOPs (Floating point operations) based on the proposed LR-ShadowNet to further justify the effectiveness of the proposed shadow illumination model and network architecture on the adjusted ISTD: (1) LRSN-16, this variant is built by reducing the channels of each layer to half of that in LR-ShadowNet; (2) LRSN-8, this variant is built by reducing the channels of this variant to a quarter of that in LR-ShadowNet; (3) LRSN-4, this variant is built by reducing the channels of this variant to one-eighth of that in LR-ShadowNet, and the maximum channels of each layer in these variant is 16. We report the MAE, the number of parameters, and FLOPs of three methods and the proposed LR-ShadowNet as well as the three variants in Table 6. We can see that the performance of each variant is gradually improved as the number of parameters of the variant increase. These results demonstrate that our method can achieve competitive performance with extremely less computational cost.

4.5. Generalisation ability

We compare the generalisation ability of our method and several state-of-the-art methods on the SBU-Timelapse dataset [20]. All methods are trained on the adjusted ISTD and tested on SBU-Timelapse. Quantitative and qualitative results are shown in Table 7 and Fig. 7, respectively. From the Table 7 we can see that the proposed LR-ShadowNet achieves the best MAE. Results in Fig. 7 show that our method can better enhances the brightness of the shadow regions.
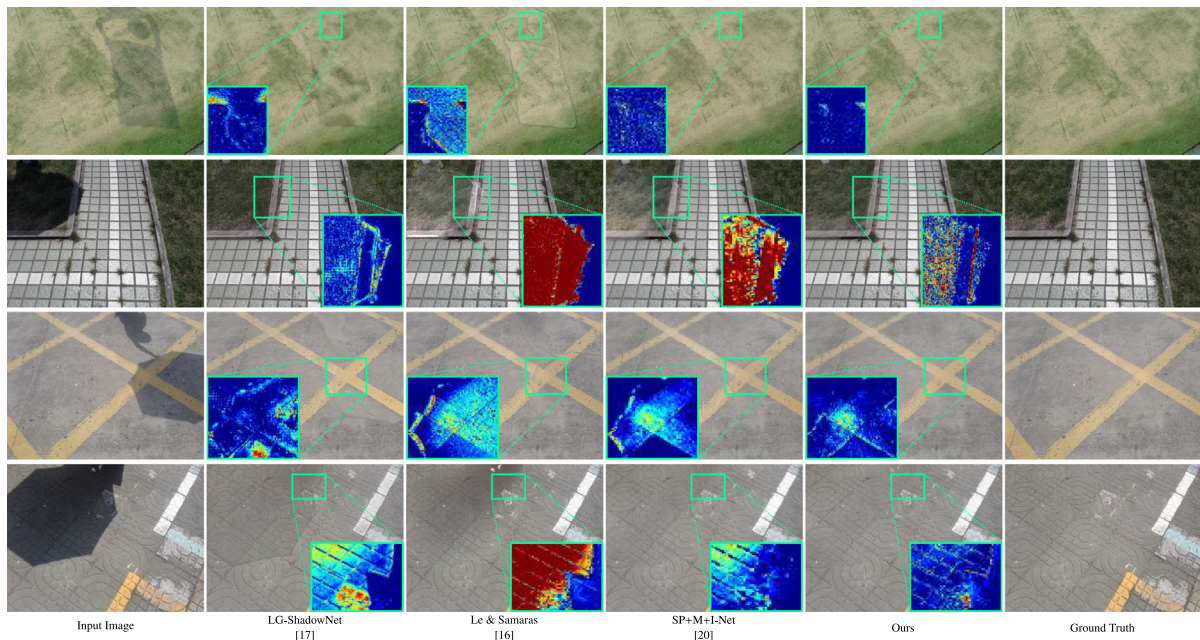
Fig. 4. Qualitative comparisons on four samples in the adjusted ISTD. Results are shown with both the shadow-removal results and the heat maps of the difference between the shadow-removal result and shadow-free ground truth for better comparison.

Table 5
Quantitative results of the proposed LR-ShadowNet and other state-of-the-art methods on the adjusted ISTD dataset. 'Non-Shadow Region' represents that the evaluation is performed on the non-shadow region of the image. Hereafter, the best two results are highlighted in red and blue, respectively.

Method	Shadow Region			Non-Shadow Region			All		
	MAE	PSNR	SSIM	MAE	PSNR	SSIM	MAE	PSNR	SSIM
Yang et al. [44]	24.7	21.57	0.937	14.2	22.25	0.782	15.9	20.26	0.706
Gong & Cosker [45]	13.3	30.53	0.972	2.6	36.63	0.982	4.3	28.96	0.943
Mask-ShadowGAN [31]	9.9	32.19	0.984	3.8	33.44	0.971	4.8	28.80	0.945
LG-ShadowNet [17]	9.7	32.44	0.982	3.4	33.68	0.974	4.4	29.20	0.946
Le & Samaras [16]	9.7	33.09	0.983	2.9	35.26	0.977	4.1	30.12	0.950
G2R-ShadowNet [18]	8.8	33.58	0.979	2.9	35.52	0.976	3.9	30.52	0.944
ST-CGAN [13]	13.4	31.70	0.979	7.9	26.39	0.956	8.6	24.75	0.927
DHAN [15]	11.2	32.92	0.988	7.1	27.15	0.971	7.8	25.66	0.956
G2R Sup. [18]	7.3	36.13	0.988	2.9	35.21	0.977	3.6	31.93	0.957
Fu et al. [29]	6.6	36.04	0.978	3.8	31.16	0.892	4.2	29.45	0.861
Einy et al. [46]	6.5	36.28	0.989	2.6	36.75	0.981	3.3	32.84	0.964
SP+M+I-Net [20]	6.0	37.59	0.990	3.1	35.96	0.975	3.5	32.91	0.961
G2C-DeshadowNet [47]	6.4	–	–	2.9	–	–	3.5	–	–
LR-ShadowNet	5.9	37.45	0.990	2.8	36.07	0.978	3.3	33.00	0.963

Table 6
Parameters, FLOPs, and MAE in the shadow region (MAE_S) and the whole image (MAE_A) on ISTD of several state-of-the-art methods and our models. The size of the input image for calculating FLOPs is set to 256×256 for all methods.

Method	# Params	FLOPs	MAE_S	MAE_A
LRSN-4	0.250M	1.6G	7.0	3.5
LRSN-8	0.981M	5.6G	6.9	3.5
LG-ShadowNet [17]	3.535M	17.7G	9.7	4.4
LRSN-16	3.884M	20.5G	6.4	3.4
G2R Sup. [18]	22.88M	113.9G	7.3	3.6
SP+M+I-Net [20]	195.6M	58.0G	6.0	3.5
LR-ShadowNet	15.45M	78.1G	5.9	3.3

Table 7
Quantitative comparison of the generalisation ability of the proposed LR-ShadowNet and the state-of-the-art methods on the SBU-Timelapse dataset. '–' represents the result is not publicly available.

Method	MAE_S	PSNR	SSIM
DHAN [15]	31.2	–	–
Mask-ShadowGAN [31]	27.3	–	–
wSP+M-Net (weakly trained) [20]	23.4	–	–
LG-ShadowNet [17]	23.3	21.29	0.9081
G2R-ShadowNet [18]	22.1	22.17	0.9138
SP+M+I-Net [20]	20.1	–	–
LR-ShadowNet	19.5	23.13	0.9224

5. Conclusion

In this paper, we studied a new translation illumination model involving multiple reflectance and illumination components for shadow removal. Based on this model, we proposed a novel LR-ShadowNet with three sub-nets for shadow removal. Two sub-nets are used for estimating illumination and reflectance components, respectively, and the last one is used for results refining. Via carefully designed mask guidance modules, LR-ShadowNet takes shadow masks as shadow region information to guide the feature learning and the entire network is supervisedly trained in an end-to-end fashion. Experimental results demonstrated the proposed LR-ShadowNet achieved competitive performance with less computational cost on the ISTD dataset and the SBU-Timelapse dataset.
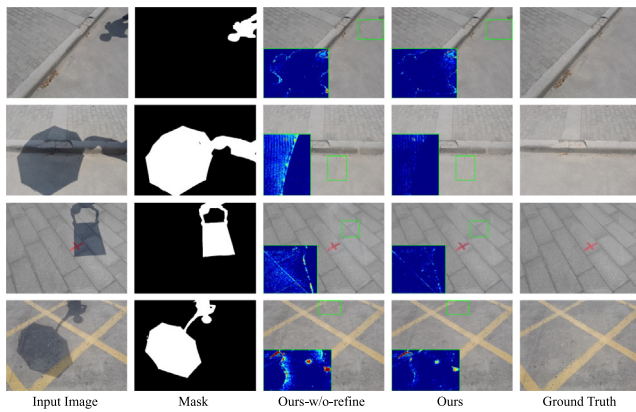
Fig. 5. Visualisation comparisons of the results before and after the result refinement sub-net $\mathcal{F}$ on ISTD. Results are shown with both the shadow-removal results and the heat maps of the difference between the shadow-removal result and shadow-free ground truth for better comparison.
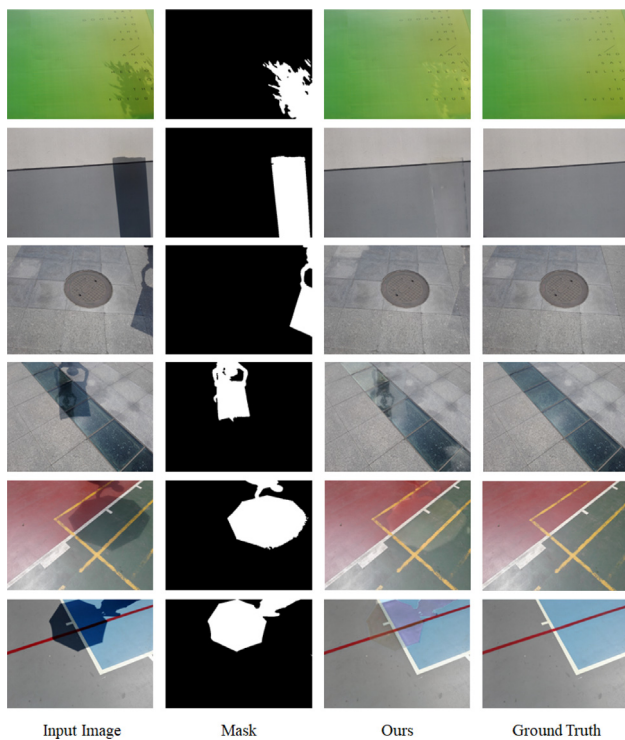

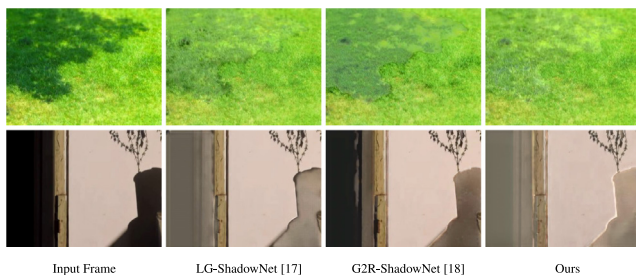**Fig. 6.** Visualisation results of some failure shadow removal examples on ISTD.

Fig. 7. Visualisation comparisons on two samples from SBU-Timelapse.

CRediT authorship contribution statement

Yanting Liu: Writing – review & editing, Writing – original draft, Visualization, Methodology. **Zhihao Liu:** Writing – review & editing, Writing – original draft, Visualization, Methodology. **Hui Yin:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology. **Jin Wan:** Writing – review & editing, Supervision. **Zhenyao Wu:** Writing – review & editing, Supervision. **Xinyi Wu:** Writing – review & editing, Supervision. **Song Wang:** Writing – review & editing, Writing – original draft, Supervision, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

References

[1] Nadimi S, Bhanu B. Physical models for moving shadow and object detection in video. IEEE Trans Pattern Anal Mach Intell 2004;26(8):1079–87.

[2] Jung CR. Efficient background subtraction and shadow removal for monochromatic video sequences. IEEE Trans Multimed 2009;11(3):571–7.

[3] Surkutlawar S, Kulkarni RK. Shadow suppression using rgb and hsv color space in moving object detection. Int J Adv Comput Sci Appl 2013;4(1).

[4] Su N, Zhang Y, Tian S, Yan Y, Miao X. Shadow detection and removal for occluded object information recovery in urban high-resolution panchromatic satellite images. IEEE J Sel Top Appl Earth Obs Remote Sens 2016;9(6):2568–82.

[5] Zhang W, Zhao X, Morvan J-M, Chen L. Improving shadow suppression for illumination robust face recognition. IEEE Trans Pattern Anal Mach Intell 2018;41(3):611–24.

[6] Le H, Vicente TFY, Nguyen V, Hoai M, Samaras D. A+D Net: Training a shadow detector with adversarial shadow attenuation. In: Eur. conf. comput. vis.. 2018.

[7] Müller T, Erdnüeß B. Brightness correction and shadow removal for video change detection with UAVs. In: Autonomous systems: sensors, processing, and security for vehicles and infrastructure 2019. Vol. 11009, International Society for Optics and Photonics; 2019, 1100906.

[8] Le H, Goncalves B, Samaras D, Lynch H. Weakly labeling the antarctic: The penguin colony case. In: IEEE conf. comput. vis. pattern recog. worksh.. 2019, p. 18–25.

[9] Finlayson GD, Hordley SD, Drew MS. Removing shadows from images. In: Eur. conf. comput. vis.. Springer; 2002, p. 823–36.

[10] Finlayson GD, Hordley SD, Lu C, Drew MS. On the removal of shadows from images. IEEE Trans Pattern Anal Mach Intell 2005;28(1):59–68.

[11] Guo R, Dai Q, Hoiem D. Paired regions for shadow detection and removal. IEEE Trans Pattern Anal Mach Intell 2012;35(12):2956–67.

[12] Qu L, Tian J, He S, Tang Y, Lau RW. Deshadownet: A multi-context embedding deep network for shadow removal. In: IEEE conf. comput. vis. pattern recog.. 2017.

[13] Wang J, Li X, Yang J. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In: IEEE conf. comput. vis. pattern recog.. 2018.

[14] Hu X, Fu C-W, Zhu L, Qin J, Heng P-A. Direction-aware spatial context features for shadow detection and removal. IEEE Trans Pattern Anal Mach Intell 2019.

[15] Cun X, Pun C-M, Shi C. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting GAN. In: AAAI. 2020, p. 10680–7.

[16] Le H, Samaras D. From shadow segmentation to shadow removal. In: Eur. conf. comput. vis.. 2020.

[17] Liu Z, Yin H, Mi Y, Pu M, Wang S. Shadow removal by a lightness-guided network with training on unpaired data. IEEE Trans Image Process 2021;30:1853–65.

[18] Liu Z, Yin H, Wu X, Wu Z, Mi Y, Wang S. From shadow generation to shadow removal. In: IEEE conf. comput. vis. pattern recog.. 2021.

[19] Wu W, Zhang S, Zhou K, Yang J, Wu X, Wan Y. Shadow removal via dual module network and low error shadow dataset. Comput Graph 2021;95:156–63.

[20] Le H, Samaras D. Physics-based shadow image decomposition for shadow removal. IEEE Trans Pattern Anal Mach Intell 2022;44(12):9088–101.

[21] Weiss Y. Deriving intrinsic images from image sequences. In: Int. conf. comput. vis.. Vol. 2, IEEE; 2001, p. 68–75.

[22] Matsushita Y, Nishino K, Ikeuchi K, Sakauchi M. Illumination normalization with time-dependent intrinsic images for video surveillance. IEEE Trans Pattern Anal Mach Intell 2004;26(10):1336–47.

[23] Shor Y, Lischinski D. The shadow meets the mask: Pyramid-based shadow removal. Comput. Graph. Forum 2008;27:577–86.

[24] Xiao C, She R, Xiao D, Ma K-L. Fast shadow removal using adaptive multi-scale illumination transfer. In: Comput. graph. forum. 2013.

[25] Le H, Samaras D. Shadow removal via shadow image decomposition. In: Int. conf. comput. vis.. 2019.

[26] Narasimhan SG, Nayar SK. Vision and the atmosphere. Int J Comput Vis 2002;48(3):233–54.

[27] Wang X, Yu K, Dong C, Change Loy C. Recovering realistic texture in image super-resolution by deep spatial feature transform. In: IEEE conf. comput. vis. pattern recog.. 2018, p. 606–15.

[28] Barrow H, Tenenbaum J, Hanson A, Riseman E. Recovering intrinsic scene characteristics. Comput Vis Syst 1978;2(3–26):2.

[29] Fu L, Zhou C, Guo Q, Juefei-Xu F, Yu H, Feng W, Liu Y, Wang S. Auto-exposure fusion for single-image shadow removal. In: IEEE conf. comput. vis. pattern recog.. 2021.

[30] Fu Y, Gai Z, Zhao H, Zhang S, Shan Y, Wu Y, Tang J. Depth-aware shadow removal. Comput Graph Forum 2022;41. URL: https://api.semanticscholar.org/CorpusID:254073300.

[31] Hu X, Jiang Y, Fu C-W, Heng P-A. Mask-ShadowGAN: Learning to remove shadows from unpaired data. In: Int. conf. comput. vis.. 2019.

[32] Zhu J-Y, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Int. conf. comput. vis.. 2017.

[33] Jin Y, Sharma A, Tan RT. DC-ShadowNet: Single-image hard and soft shadow removal using unsupervised domain-classifier guided network. In: Int. conf. comput. vis.. 2021, p. 5027–36.

[34] Laffont P-Y, Bousseau A, Drettakis G. Rich intrinsic image decomposition of outdoor scenes from multiple views. IEEE Trans Vis Comput Graphics 2012;19(2):210–24.

[35] Chen Q, Koltun V. A simple model for intrinsic image decomposition with depth cues. In: Proceedings of the IEEE international conference on computer vision. 2013, p. 241–8.

[36] Bonneel N, Kovacs B, Paris S, Bala K. Intrinsic decompositions for image editing. In: Computer graphics forum, vol. 36, Wiley Online Library; 2017, p. 593–609.

[37] Baslamisli AS, Das P, Le H-A, Karaoglu S, Gevers T. Shadingnet: image intrinsics by fine-grained shading decomposition. Int J Comput Vis 2021;129(8):2445–73.

[38] Finlayson G, Darrodi MM, Mackiewicz M. Rank-based camera spectral sensitivity estimation. J Opt Soc Amer A 2016;33(4):589–99.

[39] Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z, et al. Photo-realistic single image super-resolution using a generative adversarial network. In: IEEE conf. comput. vis. pattern recog.. 2017, p. 4681–90.

[40] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: Int. conf. learn. represent.. 2015.

[41] Guo C, Li C, Guo J, Loy CC, Hou J, Kwong S, Cong R. Zero-reference deep curve estimation for low-light image enhancement. In: IEEE conf. comput. vis. pattern recog.. 2020, p. 1780–9.

[42] Zhu L, Deng Z, Hu X, Fu C-W, Xu X, Qin J, Heng P-A. Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. In: Eur. conf. comput. vis.. 2018, p. 121–36.

[43] Vicente TFY, Hou L, Yu C-P, Hoai M, Samaras D. Large-scale training of shadow detectors with noisily-annotated shadow examples. In: Eur. conf. comput. vis.. Springer; 2016, p. 816–32.

[44] Yang Q, Tan K-H, Ahuja N. Shadow removal using bilateral filtering. IEEE Trans Image Process 2012;21(10):4361–8.

[45] Gong H, Cosker D. Interactive shadow removal and ground truth for variable scene categories. In: Brit. mach. vis. conf.. 2014.

[46] Einy T, Immer E, Vered G, Avidan S. Physics based image deshadowing using local linear model. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022, p. 3012–20.

[47] Gao J, Zheng Q, Guo Y. Towards real-world shadow removal with a shadow simulation method and a two-stage framework. In: IEEE conf. comput. vis. pattern recog. worksh.. 2022, p. 598–607.