Exploiting Residual and Illumination with GANs for Shadow Detection and Shadow Removal

LING ZHANG, Wuhan University of Science and Technology, China
CHENGJIANG LONG, JD Finance America Corporation, USA
XIAOLONG ZHANG, Wuhan University of Science and Technology, China
CHUNXIA XIAO*, Wuhan University, China

Residual image and illumination estimation have been proven to be helpful for image enhancement. In this paper, we propose a general framework, called RI-GAN, that exploits residual and illumination using generative adversarial networks (GANs). The proposed framework detects and removes shadows in a coarse-to-fine fashion. At the coarse stage, we employ three generators to produce a coarse shadow-removal result, a residual image, and an inverse illumination map. We also incorporate two indirect shadow-removal images via the residual image and the inverse illumination map. With the residual image, the illumination map, and the two indirect shadow-removal images as auxiliary information, the refinement stage estimates a shadow mask to identify shadow regions in the image, and then refines the coarse shadow-removal result to the fine shadow-free image. We introduce a cross-encoding module to the refinement generator, in which the use of feature-crossing can provide additional details to promote the shadow mask and the high-quality shadow-removal result. In addition, we apply data augmentation to the discriminator to reduce the dependence between representations of the discriminator and the quality of the predicted image. Experiments for shadow detection and shadow removal demonstrate that our method outperforms state-of-the-art methods. Furthermore, RI-GAN exhibits good performance in terms of image dehazing, rain removal, and highlight removal, demonstrating the effectiveness and flexibility of the proposed framework.

CCS Concepts: • **Computing methodologies → Image processing**;

Additional Key Words and Phrases: Shadow detection, shadow removal, residual, illumination, RI-GAN

---

*Corresponding author: Chunxia Xiao (cxxiao@whu.edu.cn).
This work was co-supervised by Chengjiang Long and Chunxia Xiao.

Authors' addresses: Ling Zhang, Wuhan University of Science and Technology, School of Computer Science and Technology, Hubei Key Laboratory of Intelligent Information Processing and Realtime Industrial System, Wuhan, Hubei, China, zhling@wust.edu.cn; Chengjiang Long, JD Finance America Corporation, Mountain View, California, USA, cjfykx@gmail.com; Xiaolong Zhang, Wuhan University of Science and Technology, School of Computer Science and Technology, Wuhan, Hubei, China, xiaolong.zhang@wust.edu.cn; Chunxia Xiao, Wuhan University, School of Computer Science, Wuhan, Hubei, China, cxxiao@whu.edu.cn.

111

1 INTRODUCTION

Shadow removal is a low-level vision task that aims to recover the illumination in shadow regions while preserving the texture details. Low light in shadow regions can degrade the scene recognition capability of some vision tasks, such as satellite remote sensing image analysis [57], public security and military analysis [6, 32], industrial inspection [28–30], and computational photography [22, 31]. In addition, shadow detection and shadow removal are necessary to improve the visual effect of image and video editing, such as film and television post-editing. However, detecting and removing shadows in complex scenes remains challenging, due to illumination changes, texture variation, and other environmental factors.

Recently, a variety of works, including traditional methods [40, 50, 56] and learning-based methods [16, 24, 44], have focused on shadow detection and shadow removal. Unlike traditional methods that rely on prior knowledge (*e.g.*, constant illumination and gradients) and often bring artifacts on the shadow boundaries, learning-based methods have achieved significant advances. However, the effectiveness of these methods depends on the training dataset and the designed models. When the training data are insufficient or the designed model is deficient, they may produce undesirable results. Also, most existing learning-based methods focus more on shadow characteristics, without sufficiently exploring other image properties like residual images and illumination for model design.
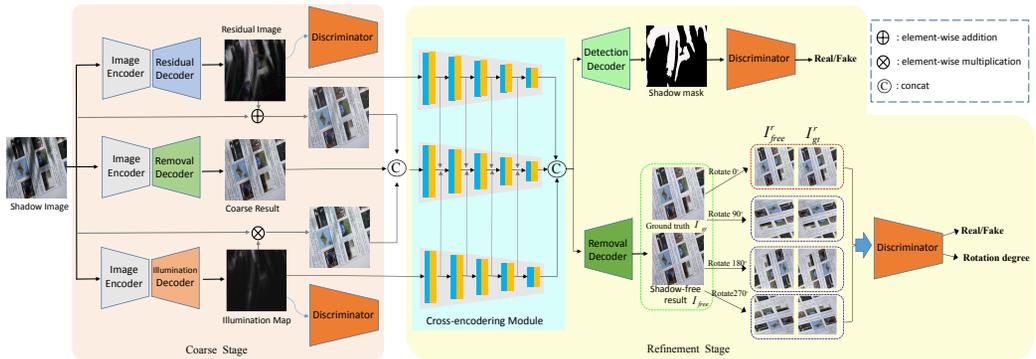

Fig. 1. Framework of our proposed RI-GAN, which is composed of a coarse stage and a refinement stage. RI-GAN takes a shadow image I as input and outputs a residual image I_{res}, an inverse illumination map S_{inv}, a shadow mask M, and a shadow-free image I_{free}. $< I_{free}^r, I_{gt}^r >$ denotes augmented data from rotating the image pair $< I_{free}, I_{gt} >$ by r degrees. The four discriminators are joint discriminators and share the same network and parameters. No particular task-aware components are designed in the framework.

In this paper, we propose a general framework, called RI-GAN, that exploits residual image and illumination map using generative adversarial networks (GANs) for shadow detection and removal by taking advantage of the intrinsic properties of images. Fig. 1 shows the framework of the proposed RI-GAN, which performs shadow detection and removal tasks in a coarse-to-fine fashion. The coarse stage employs three generators and two discriminators to obtain a residual image, a coarse shadow-removal image, and an inverse illumination map. The refinement stage uses a generator and two discriminators to produce a shadow mask and a fine shadow-free image. In principle, unlike existing deep learning methods [21, 36, 60], which are designed particularly for shadow removal, any encoder-decoder structures can be used as our generators.

The residual and illumination information between the shadow and shadow-free images can provide additional information for shadow detection and removal. We also incorporate two indirect shadow-removal images using the predicted residual image and illumination map. To take full advantage of the intrinsic properties of the image, we introduce a cross-encoding module in the refinement generator to learn more abundant features from the image. The cross-encoding module consists of a residual encoder, an illumination encoder, and an image encoder. Unlike the residual encoder and the illumination encoder, which have one input source, the image encoder uses the coarse shadow-removal result and the two indirect shadow-removal images as input sources. Feature-crossing in the three encoders is used to obtain richer feature information for high-quality results and to mitigate the reliance on the residual image and illumination map.

Our RI-GAN adopts an adversarial training process [15] between the generators and discriminators alternatively to generate a high-quality residual image, illumination map, shadow mask, and final shadow-removal image. We adopt joint discriminators [20] to ensure that the four discriminators share the same architecture as well as the same parameter values, such that all the produced results are indistinguishable from their corresponding ground-truth images. Moreover, we augment the inputs of the discriminator for shadow-removal result prediction at the refinement stage by rotating the input pair at different degrees. Besides judging the true or false classification for the predicted image, this discriminator also performs rotation-degree classification for the augmented data. Such treatment facilitates the stability of the discriminator and reduces the dependence between the representations of the discriminator and the quality of the result.

Given the commonality and generality of residual image and illumination map used in image processing, the proposed RI-GAN has no particular shadow-aware components for image shadow removal. Our RI-GAN is a general framework. Besides shadow removal, the RI-GAN model can be applied to other low-level image processing tasks, such as image dehazing, rain removal, and highlight removal. Fig. 2 shows different image processing results using our RI-GAN.

Our main contributions can be summarized as follows:

- We propose a general framework, called RI-GAN, with generators in encoder-decoder structures to exploit residual and illumination information for image shadow detection and removal.
- We introduce a cross-encoding module to make full use of the intrinsic properties of the image. As such, the network learns more useful information for producing high-quality results, and the joint discriminators and data augmentation make the network more stable and effective.
- Without any particular shadow-aware components in the proposed network, our RI-GAN nevertheless outperforms art-of-the-art methods. The results for image dehazing, rain removal, and highlight removal demonstrate the feasibility and flexibility of the proposed framework.

This paper is an extended version of our conference paper [55]. We extend [20] in several ways: (1) Here, we propose a general RI-GAN framework to perform shadow detection and removal tasks in a parallel joint model, whereas our previous framework, RIS-GAN [55], only focused on shadow removal. (2) We introduce a cross-encoding module in the refinement generator to exploit the intrinsic properties of the image. (3) We augment the input sources of the discriminator at the refinement stage for shadow-removal result prediction, and we

(a) Image dehazing

(b) Image highlight removal

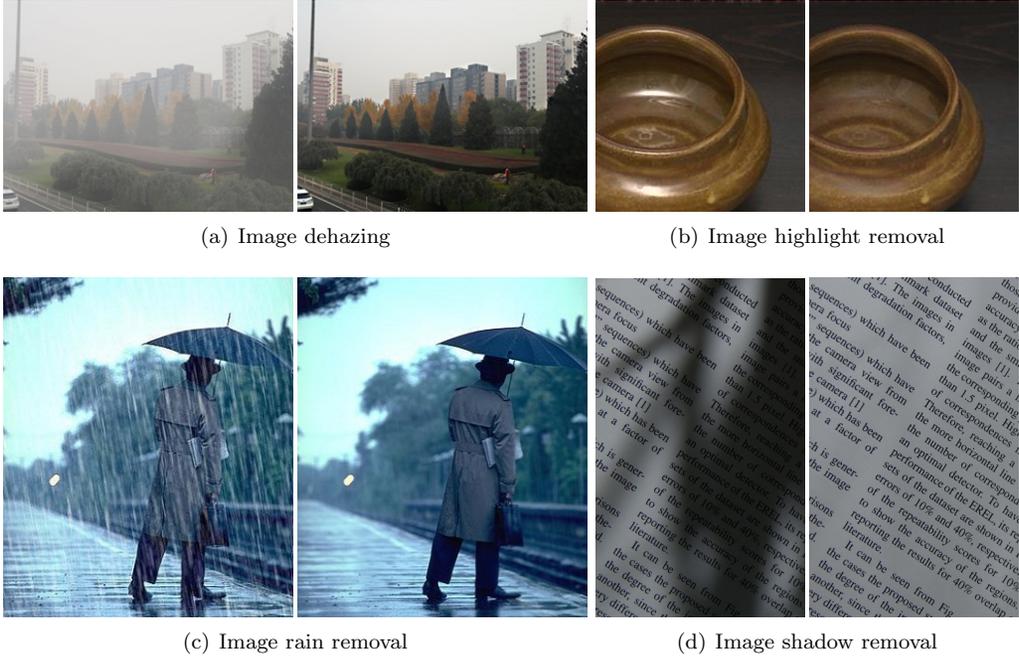(c) Image rain removal

(d) Image shadow removal

Fig. 2. Different image processing tasks using the proposed general RI-GAN. Given a corrupted image (left), we can restore a clean image (right).

train it with dual-task objectives. (4) We apply the proposed RI-GAN to other three image processing tasks to demonstrate the effectiveness and flexibility of our method.

The remainder of this paper is organized as follows. In Section 2, we introduce related work. In Section 3, we present the shadow image formation. In Section 4, we introduce the proposed general RI-GAN. Section 5 describes the experiments. Section 6 concludes the paper.

2 RELATED WORK

Shadow detection. Traditional methods often use the light and color difference between shadows and non-shadows for shadow detection. For example, Finlayson et al. [11] used light invariance for shadow detection, but this method only achieves satisfactory results on high-quality images and simple scenes. Zhang et al. [57] used manual interaction to mark shadow areas and non-shadow areas in images, and then separated the shadow area by combining a matting algorithm. This method is effective for shadow detection in simple scenes, but for complex scenes, it requires more interactive information to obtain relatively complete detection results, which is labor-intensive and cannot be processed in batches. Guo et al. [17] used a region-based method to detect shadows. Their algorithm is simple and effective, and suitable for hard shadow detection. But it risks dividing soft shadows in the image into non-shadow regions.

Deep learning is another common shadow detection strategy [5, 46]. Deep learning methods learn the appearance characteristics of the shadow areas from annotated datasets, and then use classifiers for shadow recognition. Nguyen et al. [34] proposed a shadow detector based

on an antagonism network that used a cascading convolutional neural network. Vicente et al. [42] used a block-based convolutional neural network to extract a shadow probability graph of the image. But their method needs interactive feature information. Zhu et al. [60] used global context and local context information to construct a deep convolutional neural network to detect shadows. Hu et al. [21] proposed a deep network with a direction-aware spatial context module for shadow detection through global semantic analysis. These learning-based shadow detection methods are very effective for simple scenes. For complex scene, however, the shadow detection results are often inaccurate. In addition, at present, these methods mainly detect the shadow region in the image, and cannot effectively distinguish hard shadows from soft shadows.

Shadow removal. Traditional methods for shadow removal usually utilize priors to build the models and solve them, such as illumination uniformity [17, 23, 40, 50, 51, 57] and gradient invariance [10, 11]. Methods based on illumination uniformity borrow the illumination from non-shadow regions to shadow regions. Shor et al. [40] proposed a linear mapping model to remove shadows in the image. Their model has low time complexity but can only deal with images with consistent texture in shadow regions. Xiao et al. [50] proposed a parameter-adaptive shadow-removal algorithm, which can process images with multiple textures in shadow regions. Since shadow regions are segmented beforehand and processed independently, the recovered illumination in different shadow areas may be inconsistent. Zhang et al. [57] decomposed images into overlapping image patches, and proposed a local-to-global method to remove shadows. However, such illumination transfer-based methods require taking texture matching for shadow regions and non-shadow regions before removing shadows, and the effectiveness of these methods depends on the illumination in the matched non-shadow region. Another typical methods are based on gradient domain manipulation to reconstruct the shadow-free result utilizing the gradient information on shadow regions. Finlayson et al. [11] reconstructed the shadow-free images by solving a gradient-based Poisson equation. Liu et al. [26] constructed illumination variation lines at the shadow boundary to eliminate the gradient change caused by the change of illumination, and reconstructed a shadowless image using the illumination variation lines. Due to the influence of illumination changes at the shadow boundary, both illumination- and gradient-based methods cannot effectively handle boundary problems, especially in the presence of complex textures or color distortions.

Recently, deep neural networks have been used for shadow removal by analyzing and learning the mapping relations between shadow images and their corresponding shadow-free images [54]. Learning-based methods for shadow removal have shown their potential. With suitable network models and datasets, these methods can produce good shadow-removal results [12, 21, 41, 54]. Hu et al. [21] used multiple convolutional neural networks to learn image features for shadow detection combined with multi-level color transfer. They proposed a Bayesian formulation to remove shadows in images. Qu et al. [36] proposed an end-to-end DeshadowNet to recover illumination in shadow regions. This network integrates high-level semantic information, middle-level appearance information, and local image details. Wang et al. [44] proposed a stacked conditional generative adversarial network (ST-CGAN) for image shadow removal. Sidorov [41] proposed an end-to-end architecture, named AngularGAN, oriented specifically to the color constancy task, without estimating illumination color or an illumination color map. Unlike the commonly used multi-branch paradigm, they stacked all the tasks for multi-task learning. Such deep learning methods can produce better shadow-removal results, but they require a large training dataset. Wei et al. [47] proposed a two-stage generative adversarial network for shadow inpainting and removal with slice convolutions.

Ding et al. [9] proposed an attentive recurrent generative adversarial network (ARGAN) to detect and remove shadows with multiple steps. These methods rely heavily on the designed model. Unlike existing methods, our RIS-GAN utilizes a explored residual image and an inverse illumination map to generate more accurate shadow-removal results.

3 SHADOW IMAGE FORMATION

Based on matting and compositing operations, [48] proposed the following image compositing equation:

$$I_{observe} = \alpha F + (1 - \alpha)B, \tag{1}$$

where $I_{observe}$, F, B, and α are the observed image, foreground image, background image, and fractional alpha, respectively. Inspired by this image compositing equation, we model the shadow image I as a fusion of the shadow-free image I_{free} and the shadow mask M:

$$I = (1 - \beta)I_{free} + \beta M, \tag{2}$$

where β is the fractional alpha. I_{free} and M can be considered the background image and the foreground image, respectively. The purpose of shadow detection and removal is to detect the shadow regions M and extract the shadow-free image I_{free} from the shadow image I.

4 THE GENERAL RI-GAN

4.1 RI-GAN Architecture

Residual image and illumination have been widely used in image processing. We exploit residual image and illumination map between the shadow image and the shadow-free image using generative adversarial networks (GANs) [15] for shadow detection and removal. The residual image denotes the difference between the shadow image and shadow-free image in color space, and the illumination map shows the brightness information for converting a shadow-free image to a shadow image. The intuition behind this is that the residual image and the inverse illumination map can provide additional information and insights for shadow detection and removal. This view also satisfies other low-level vision tasks, such as image rain removal and highlight removal. Fig. 3 presents some results of residual images and illumination maps for different image processing tasks.

The proposed RI-GAN for shadow detection and removal with multiple GANs is illustrated in Fig. 1. It is composed of a coarse stage and a refinement stage. The coarse stage S_{coarse} contains three generators and two discriminators: a residual generator G_{res}, a shadow removal generator G_{coarse}, an illumination generator G_{illum}, a residual discriminator D_{res}, and an illumination discriminator D_{illum}. The refinement stage S_{refine} contains a refinement generator G_{refine}, a detection discriminator D_{detect}, and a removal discriminator D_{ref}. The generators in our model are encoder-decoder structures.

Given an input shadow image I, the three generators at S_{coarse} generate a residual image I_{res}, a coarse shadow-removal image I_{coarse}, and an inverse illumination map S_{inv}. With the element-wise addition, we are able to get an indirect shadow-removal image I_{mid}^1 using the input shadow image I and the residual image I_{res}:

$$I_{mid}^1 = I_{res} + I. \tag{3}$$

With the element-wise multiplication, we can get another indirect shadow-removal image I_{mid}^2 using the input shadow image I and the inverse illumination S_{inv}:

$$I_{mid}^2 = S_{inv} * I. \tag{4}$$

Fig. 3. Example of residual images and illumination maps. The first, second, and third rows show input images, residual images, and illumination maps, respectively.

Then, by using the predicted residual image, the illumination map, and the two indirect shadow-removal images as supplementary information, the refinement generator refines the coarse shadow-removal image to a fine shadow-free image I_{free} and estimates a shadow mask M to indicate shadow regions in the image.

The four discriminators share the same architecture and the same parameters. The alternative training between generators and discriminators ensures the excellent quality of the prediction results.

4.2 Encoder-Decoder Generators

In principle, any encoder-decoder structures can be used in our RI-GAN framework. In this paper, we do not design task-specific components in the framework, and we adopt the DenseUNet architecture [37] for our encoder-decoder generators. DenseUNet consists of a contracting path to capture context and a symmetric expanding path to upsample. Unlike the conventional UNet architecture, DensUNet adds dense blocks in the network, which concatenate each layer's output with its input and feed it to the next layer. This enhances information and gradient flow in our encoder-decoder generators.

The residual generator G_{res} is designed to obtain a residual image that is close to the ground-truth residual image I_{gt}^{res} obtained between the shadow image and the corresponding ground-truth image I_{gt}:

$$I_{gt}^{res} = I_{gt} - I. \tag{5}$$

We design the illumination generator G_{illum} to estimate the inverse illumination map in the shadow image. Note that the ground-truth inverse illumination map is calculated based on Retinex-based image enhancement methods [13, 18, 45]:

$$S_{gt}^{inv} = I_{gt} * I^{-1}, \tag{6}$$

where I_{gt} can be considered a reflectance image, and I is the observed image.

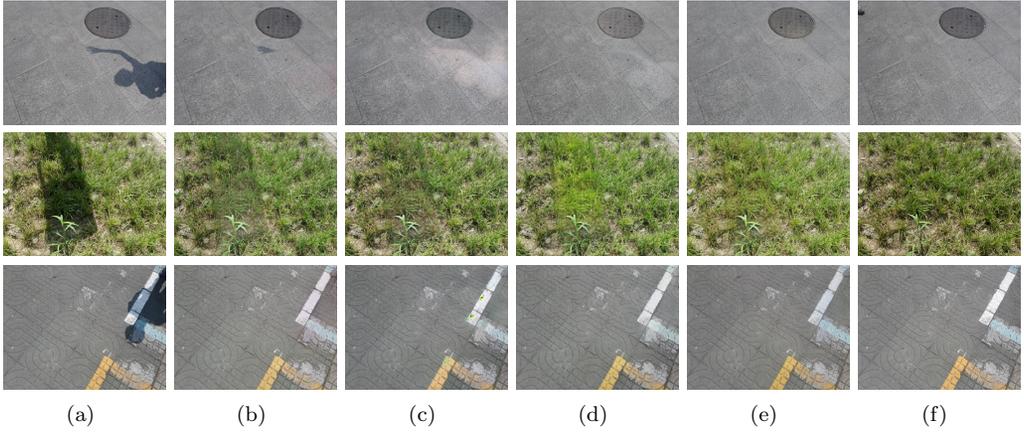| (a) | (b) | (c) | (d) | (e) | (f) |

Fig. 4. Visualization of shadow-removal results using our RI-GAN. From left to right: the shadow image I (a), the indirect shadow-removal image I_{mid}^1 by the residual image (b), the indirect shadow-removal image I_{mid}^2 by illumination (c), the coarse shadow-removal image I_{coarse} (d), the final shadow-free image I_{free} (e), and the ground-truth image I_{gt} (f).

Refinement generator. The predicted residual image and illumination map are used as supplementary information to help the refinement generator predict the shadow mask and shadow-free image. However, inaccurate residual images and illumination maps can misguide the network and produce undesirable results. To utilize residual image and illumination map effectively, we introduce a cross-encoding module to extract the features. The module consists of a residual encoder, an illumination encoder, and an image encoder. Unlike residual and illumination encoders with merely one input source, our image encoder uses two indirect shadow-removal images and the coarse shadow-removal image as input sources. Specifically, the downsampled extracted features from the residual encoder and the illumination encoder are fed into the image encoder, allowing the network to obtain richer feature information. Such treatment can relieve the reliance on the residual image and illumination map and produce more convincing results.

To better understand our generators, we visualize the two indirect shadow-removal images, the coarse shadow-removal image, and the final shadow-free result in Fig. 4. As we can observe, the indirect shadow-removal images obtained by the residual image and illumination map have good quality and are complimentary for further refinement to get the final shadow-free image.

4.3 Joint Discriminator

The discriminators are convolutional networks used to distinguish the predicted residual image, the estimated inverse illumination map, the shadow mask, and the final shadow-free image produced by the generators as either real or fake, compared with the corresponding ground-truth. We use GANs with joint discriminators [20] to ensure that all the predicted results are indistinguishable from the related ground-truth images. The four discriminators share the same network and parameters. The joint discriminators are trained to learn a joint distribution to judge whether the produced results are real or fake.

Inspired by data augmentation [3], we train the removal discriminator with dual-task objectives. We rotate the predicted shadow-free image I_{free} and the corresponding ground-truth image I_{gt} simultaneously at different degrees and produce augmentation pairs. Similar to the original real and fake image pair, the additional image pairs are also fed into the removal discriminator during training. Besides judging the predicted image as real or fake, this discriminator also performs rotation degree classification for all the inputs containing the augmentation images. Data augmentation for the discriminator renders the network more stable and useful. Note that only the original predicted images are considered for the true or false classification task. For data augmentation, we rotate the images to four different degrees. Let R be the set of possible rotations, and $R = \{0°, 90°, 180°, 270°\}$. $< I_{free}^r, I_{gt}^r >, r \in R$ denotes the augmented input pairs. When $r = 0°$, the image pair is the original true and false images.

Our discriminator consists of five convolution layers, each followed by a batch normalization, a leaky ReLU activation function, and one fully connected layer. After the last fully connected layer, there are two outputs. One is the probability value that the input image (the result produced by the generator) is a real image. The other is the probability value for the classification of the rotation degree. Fig. 5 gives details of the discriminator. Note that discriminators D_{res}, D_{detect} and D_{illum} only have the output path for judging the truth of the predicted image.
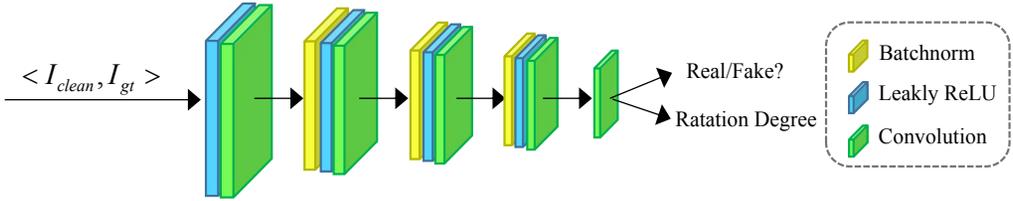

Fig. 5. Architecture of the discriminators in our RI-GAN, consisting of five convolution layers with batch normalization and leakly ReLU activations. For these five convolution layers, the kernel sizes are 4×4; the strides are 4×4, except the first convolution layer, whose stride is 2×2; and the number of output channels is $64 \rightarrow 128 \rightarrow 256 \rightarrow 512 \rightarrow 1$.

We use the spectrum normalization method [33] to stabilize the training process of the discriminator network, because spectral normalization is a simple and effective standardized method for limiting the optimization process of the discriminator in GANs. It can improve the performance of the generators.

4.4 Loss Functions

To obtain a robust parametric model, the loss function that we use to optimize the proposed RI-GAN for shadow detection and removal contains seven components: coarse removal loss $\mathcal{L}_{coarse}$, residual loss $\mathcal{L}_{resid}$, illumination loss $\mathcal{L}_{illum}$, cross-loss $\mathcal{L}_{cross}$, detection loss $\mathcal{L}_{detect}$, shadow removal loss $\mathcal{L}_{removal}$, and adversarial loss $\mathcal{L}_{adv}$. The total loss $\mathcal{L}$ can be written as

$$\mathcal{L} = \mathcal{L}_{coarse} + \mathcal{L}_{resid} + \mathcal{L}_{illum} + \beta_1 \mathcal{L}_{cross} + \beta_2 \mathcal{L}_{detect} + \beta_3 \mathcal{L}_{removal} + \beta_4 \mathcal{L}_{adv}, \quad (7)$$

where β_1, β_2, β_3, and β_4 are the weight parameters.

Coarse removal loss is defined with the visual-consistency loss and perceptual-consistency loss between the coarse shadow-removal result I_{coarse} and the corresponding ground-truth

I_{gt}:

$$\mathcal{L}_{coarse} = \mathcal{L}_{vis} + \mathcal{L}_{percept}$$
$$= ||I_{gt} - I_{coarse}||_1 + ||\text{VGG}(I_{gt}) - \text{VGG}(I_{coarse})||_2^2, \tag{8}$$

where $\mathcal{L}_{vis}$ is the visual-consistency loss between the predicted image and the corresponding ground-truth image. It is calculated using the L1-norm. $\mathcal{L}_{percept}$ is the perceptual-consistency loss aiming to preserve the image structure. VGG($\cdot$) is the feature extractor from the VGG19 model.

Residual loss calculates the L1-norm between the residual result generated by the residual generator at S_{coarse} and the ground-truth residual image I_{gt}^{res}:

$$\mathcal{L}_{resid} = ||I_{gt}^{res} - I_{res}||_1. \tag{9}$$

Illumination loss calculates the L1-norm between the inverse illumination result generated by the illumination generator at S_{coarse} and the ground-truth inverse illumination map S_{gt}^{inv}. Then, the illumination loss can be calculated as:

$$\mathcal{L}_{illum} = ||S_{gt}^{inv} - S_{inv}||_1. \tag{10}$$

Cross-loss is designed to ensure the consistency and correlation among the residual image, illumination, and image information. It calculates the L1-norm between the two indirect shadow-removal images I_{mid}^1 and I_{mid}^2 and the ground-truth image I_{gt}:

$$\mathcal{L}_{cross} = ||I_{gt} - I_{mid}^1||_1 + ||I_{gt} - I_{mid}^2||_1. \tag{11}$$

Detection loss is used to calculate the loss between the produced shadow mask M by the refinement generator and the corresponding ground-truth shadow mask M_{gt}. It is used as the supervision in the training processing for shadow detection. That is,

$$\mathcal{L}_{detect} = ||M - M_{gt}||_2^2. \tag{12}$$

Shadow removal loss is also defined with the visual-consistency loss and perceptual-consistency loss. That is,

$$\mathcal{L}_{removal} = ||I_{gt} - I_{free}||_1 + ||\text{VGG}(I_{gt}) - \text{VGG}(I_{free})||_2^2, \tag{13}$$

where VGG($\cdot$) is the feature extractor from the VGG19 model.

Adversarial loss is the joint adversarial loss for the network. It contains the direct adversarial loss $\mathcal{L}_{adv\_direct}$ and rotating adversarial loss $\mathcal{L}_{adv\_rotate}$, and is written as

$$\mathcal{L}_{adv} = \mathcal{L}_{adv\_direct} + \mathcal{L}_{adv\_rotate}. \tag{14}$$

Direct adversarial loss $\mathcal{L}_{adv\_direct}$ is the adversarial loss for the four discriminators to judge whether the produced results are real or fake:

$$\mathcal{L}_{adv\_direct} = \mathbb{E}_{(I,I_{free},I_{gt},M,M_{gt},I_{gt}^{res},S_{gt}^{inv})}[log(D_{detect}(M^{gt})) + log(1 - D_{detect}(M))$$
$$+ log(D_{res}(I_{gt}^{res})) + log(1 - D_{res}(I_{res}))$$
$$+ log(D_{illum}(S_{gt}^{inv})) + log(1 - D_{illum}(S_{inv})) \tag{15}$$
$$+ log(D_{ref}(I_{gt})) + log(1 - D_{ref}(I_{free}))],$$

where D_{res}, D_{illum}, D_{detect}, and D_{ref} are the four discriminators.

Rotating adversarial loss $\mathcal{L}_{adv\_rotate}$ is the adversarial loss for the discriminator D_{ref} to perform rotation-degree classification:

$$\mathcal{L}_{rotate} = \mathbb{E}_{(I,I_{gt},I_{free})}[- \sum_r log(D_{ref}(I_{gt}^r, I_{free}^r))], \tag{16}$$

where $r \in R = \{0°, 90°, 180°, 270°\}$. I_{gt}^r and I_{free}^r are the images that the ground-truth I_{gt} and the predicted shadow-removal result I_{free} rotate r degrees, respectively.

Overall, our objective for the training task is to solve a mini-max problem that aims to find a saddle point between the generators and discriminators of our network.

| (a) | (b) | (c) | (d) | (e) | (f) | (g) | (h) | (i) | (j) |
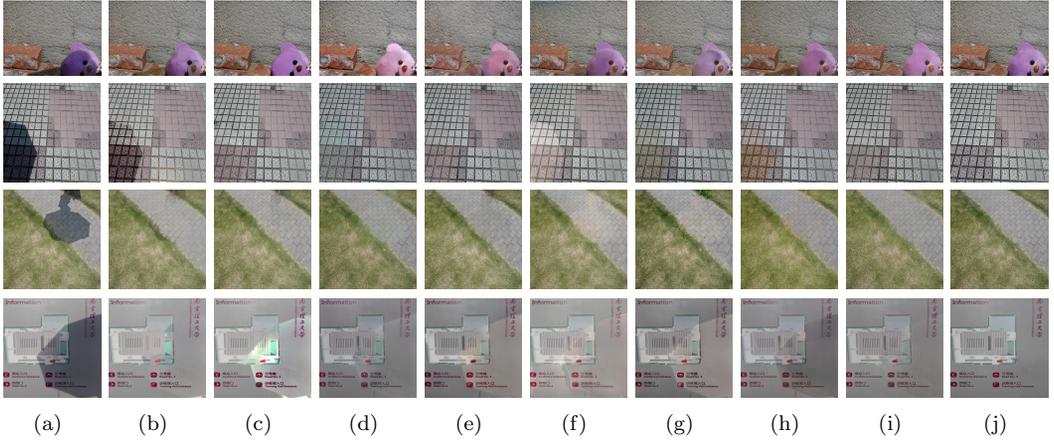
Fig. 6. Shadow removal results. From left to right: input images (a); shadow-removal results of Guo et al. [17] (b), Zhang et al. [57] (c), DeshadowNet [36] (d), ST-CGAN [44] (e), DSC [21] (f), Angu-larGAN [41] (g), ARGAN [9] (h), and our RI-GAN (i); and the corresponding ground-truth shadow-free images (j).

| (a) | (b) | (c) | (d) | (e) | (f) | (g) |

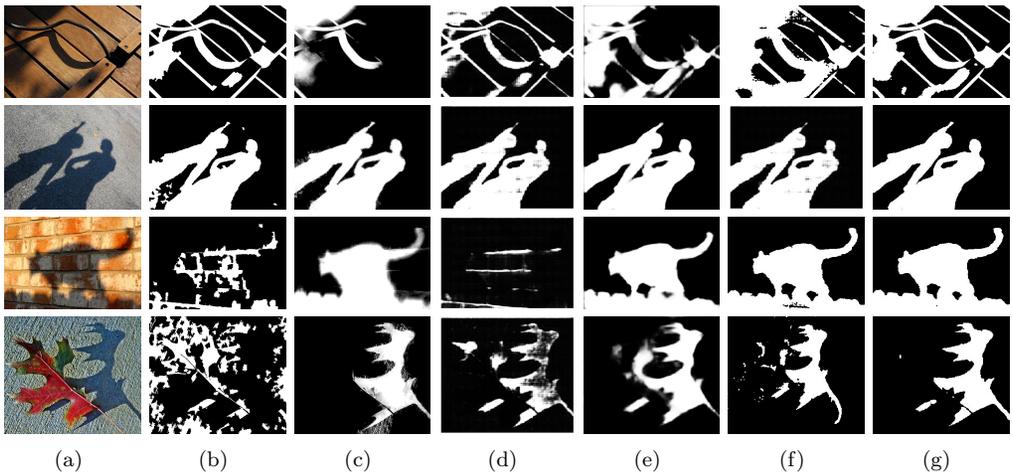Fig. 7. Shadow detection results. From left to right: input images (a); shadow detection results of Guo [17] (b), Zhang [57] (c), ST-CGAN [44] (d), DSC [21] (e), and BDRAR [59] (f); and shadow detection results of our method (g).

5 EXPERIMENTS

To evaluate and verify the effectiveness of the proposed RI-GAN, we present different experiments using our RI-GAN. The proposed model is implemented in Tensorflow in a

computer with NVIDIA GeForce RTX2080Ti. In our experiments, the input size of the image is 256×256. The minibatch size is 2. The initial learning rate is set at 0.001. We use the Momentum optimizer to optimize our generator, and we use the Adam optimizer for the discriminator. We alternatively train our generator and discriminator for 10,000 epochs. The parameters β_1, β_2, β_3, and β_4 are set as 1, 0.2, 10, and 1, respectively, in our image shadow detection and shadow removal experiments.

5.1 Comparisons with State-of-the-art Approaches

Datasets. For the shadow detection and removal task, we use the 1330 triplets of shadow images, shadow masks and shadow-free images from the ISTD dataset [44] for training, and we use the remaining 540 triplets for testing. We also evaluate the shadow detection results on the UCF dataset, which contains 110 pairs of shadow images and shadow masks. Besides the ISTD test data, we use data from the publicly available SRD dataset [36] for shadow-removal evaluation, which has 408 pairs of shadow and shadow-free images.

Table 1. Quantitative comparison results of shadow detection on the UCF and ISTD datasets in terms of BER (the smaller, the better).

Methods	Venue/Year	UCF	ISTD
Guo et al.	CVPR/2011	28.32	27.16
Zhang et al.	TIP/2015	9.21	8.56
ST-CGAN	CVPR/2018	17.69	3.84
DSC	CVPR/2018	8.73	2.40
BDRAR	ECCV/2018	9.45	2.20
DSDNet	CVPR/2019	7.59	2.17
MTMT-Net	CVPR/2020	7.34	1.72
FDRNet	ICCV/2021	7.28	**1.55**
RI-GAN	TOMM/2022	**7.25**	1.70

Metrics. We use the balance error rate (BER) [21] between the ground-truth mask and the predicted shadow mask to evaluate the shadow-detection performance. We use the root mean square error (RMSE) calculated in Lab space between the recovered shadow-removal image and the ground-truth shadow-free image to evaluate the shadow-removal performance.

Shadow detection. We compare our shadow-detection results with some state-of-the-art shadow-detection methods, namely, Guo et al. [17], Zhang et al. [57], ST-CGAN [44], DSC [21], BDRAR [59], DSDNet [58], MTMT-Net [5], and FDRNet [61]. Fig. 7 shows shadow-detection results for different methods. From these results, we can see that traditional methods of Guo et al. and Zhang et al. can effectively detect hard shadows with large areas, but for soft shadows or complex shadows, they do not obtain accurate results, as shown in Fig. 3(b, c). By considering both the residual and illumination information in our method, we can detect more accurate shadow regions (Fig. 3(g)) than other learning-based methods (Fig. 3(d-f)).

To further demonstrate the superiority of our approach, we evaluate the performance of shadow detection on the UCF dataset and ISTD test dataset. The results are summarized in Table 1. We can observe that our method achieves the optimal or suboptimal results on the two datasets, demonstrating the good performance of our method in terms of shadow detection.

Shadow removal. We compare our RI-GAN with some state-of-the-art shadow-removal methods, including the two traditional methods, viz., Guo et al. [17] and Zhang et al. [57], and recent learning-based methods, viz., DeshadowNet [36], DSC [21], ST-CGAN [44],
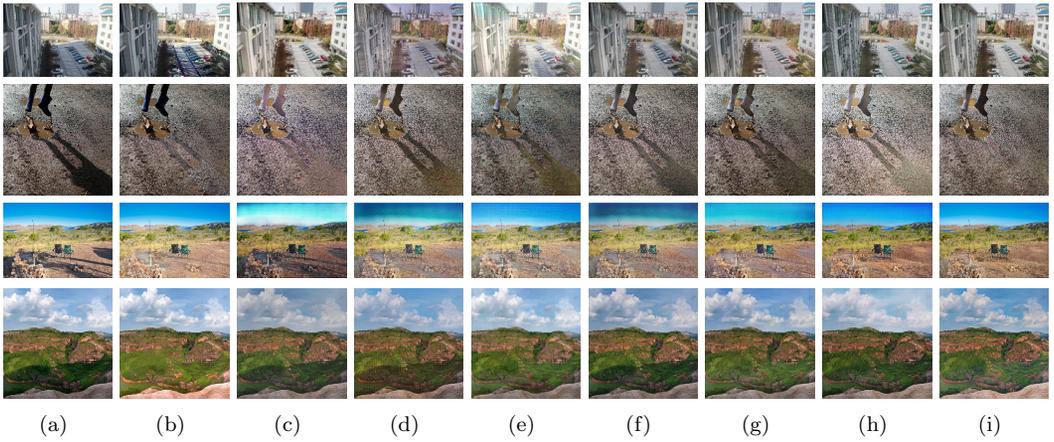
Fig. 8. Shadow-removal results for images with complex scenes. From left to right: input images (a); shadow-removal results of Zhang et al. [57] (b), DeshadowNet [36] (c), ST-CGAN [44] (d), DSC [21] (e), AngularGAN [41] (f), and ARGAN [9] (g), and CANet [4] (h); and shadow-removal results of our RI-GAN (i).

FusionNet [12], ARGAN [9], DHAN [7], and CANet [4]. For a fair comparison, we use the same training data with the same input size of images (256×256) to train all the learning-based methods on the same hardware. We summarize the comparison results in Table 2. From the table, we can observe that among all the competing methods, our proposed RI-GAN achieves smaller or similar RMSE values in shadow regions and the entire images on the two datasets, even though we have no particular shadow-aware components designed in our generators. This suggests that the recovered shadow-removal images obtained by our RI-GAN are much closer to the corresponding ground-truth shadow-free images.

Table 2. Quantitative comparison results of shadow removal on the SRD and ISTD datasets using the RMSE metric (the smaller, the better). S and A represent shadow regions and the entire image, respectively.

Methods	Venue/Year	SRD		ISTD	
		S	A	S	A
Guo et al.	CVPR/2011	31.06	12.60	18.95	9.30
Zhang et al.	TIP/2015	9.50	7.24	9.77	8.16
Deshadow	CVPR/2017	17.96	8.47	12.76	7.83
ST-CGAN	CVPR/2018	18.64	8.23	10.33	7.47
DSC	CVPR/2018	10.89	6.23	9.76	6.67
DHAN	AAAI/2020	8.94	**5.67**	8.14	6.37
CANet	ICCV/2021	8.62	6.42	8.86	6.15
FusionNet	CVPR/2021	8.56	6.51	8.14	5.92
RIS-GAN	AAAI/2020	8.22	6.78	8.99	6.95
RI-GAN	TOMM/2022	**8.17**	6.21	**8.12**	**5.91**

To further demonstrate the superiority of our proposed RI-GAN, we provide some visualization results in Fig. 6 and Fig. 8, covering traditional and learning-based methods for shadow removal. From these images, we can observe that the recovered shadow-removal images generated by traditional methods have boundary problems, such as color distortion or texture loss. For recent deep learning methods, DeshadowNet, ST-CGAN, DSC, ARGAN,

and CANet deal with the images in aspect of color space, without considering other image-intrinsic information. This may lead to unsatisfactory shadow-removal results like color distortion or incomplete shadow removal. In contrast, by fully utilizing image information such as residual images and inverse illumination maps, our proposed RI-GAN can effectively remove shadows and produce good results in both simple and complex scenes.

5.2   Application to Other Image Processing Tasks

(a)   (b)   (c)   (d)   (e)   (f)   (g)   (h)
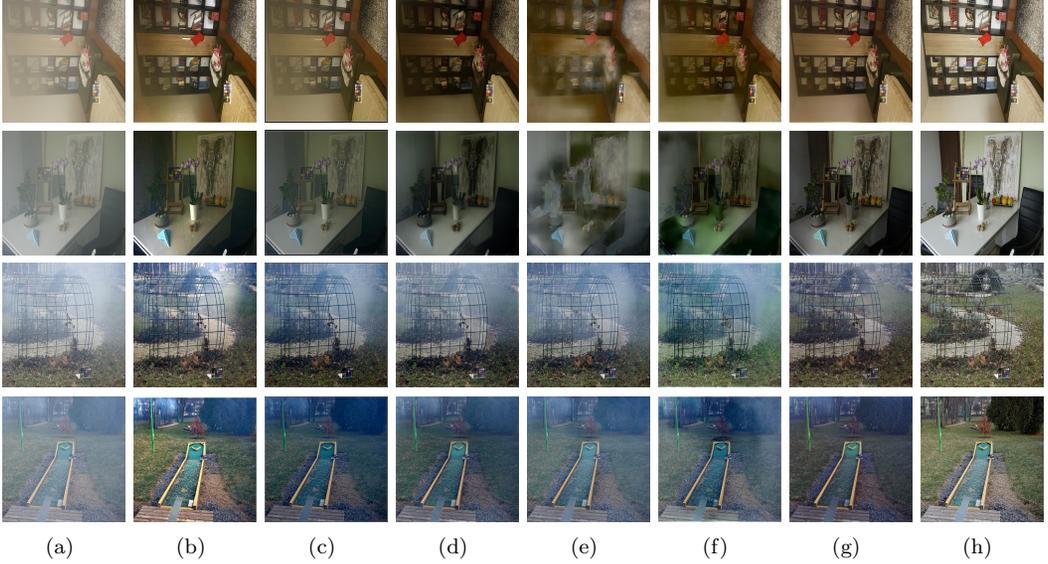
Fig. 9. Image dehazing results. From left to right: input images (a); dehazing results of He et al. [19] (b), AOD-Net [25] (c), DenseGAN [37] (d), GridDehazeNet [27] (e), LapDehazeNet [49] (f), our RI-GAN (g), and ground-truth haze-free images (h).

5.2.1   Image Dehazing. **Datasets.** We train the proposed RI-GAN for image dehazing task on the NYU dataset, which contains 1449 image pairs of haze images and haze-free images. Then, we evaluate the haze-removal results on test datasets of I-haze (30 image pairs with outdoor scenes) and O-haze (45 image pairs with indoor scenes).
Metrics. We use the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) between the ground-truth haze-free image and the predicted haze-removal result to evaluate the performance of our RI-GAN on haze removal.

Image dehazing is similar to image shadow removal. It aims to recover a haze-free image (I_{free}) from a haze image (I). Fig. 9 gives image dehazing results compared with the traditional method of He et al. [19], and five learning-based methods, viz., AOD-NET [25], DenseGAN [37], GridDehazeNet [27], DMPHN [8], and LapDehazeNet [49]. As shown in Fig. 9(b), although the haze-free results by He et al. [19] can extract the haze-free images effectively, some details may be missing compared with the ground-truth haze-free images. The results from AOD-NET, DenseGAN, and GridDehazeNet still exhibit some haze in the images, as shown in Fig. 9(c-e). However, our results remove most haze and preserve the details in the image, as shown in Fig. 9(f).

We also evaluate the haze-removal results on test datasets of I-haze and O-haze. For a fair comparison, we train all the learning-based methods on the NYU dataset. The results

are summarized in Table 3, from which we can observe that the proposed RI-GAN can be directly applied to image dehazing and achieves favorable results compared with the state-of-the-art methods.

Table 3. Quantitative comparison results of haze removal on the I-HAZE and O-HAZE datasets using the PSNR metric (the larger, the better) and the SSIM metric (the larger, the better).

Methods	Venue/Year	I-HAZE		O-HAZE	
		PSNR	SSIM	PSNR	SSIM
He et al.	PAMI/2011	12.86	0.54	15.06	**0.62**
AOD-NET	ICCV/2017	9.79	0.29	14.98	0.48
DenseGAN	CVPR/2018	10.16	0.37	15.01	0.58
GridDehazeNet	ICCV/2019	12.05	0.46	16.07	0.61
DMPHN	CVPR/2020	12.21	0.48	16.14	0.51
LapDehazeNet	CVPR/2021	11.60	0.45	15.67	0.52
RI-GAN	TOMM/2022	**12.88**	**0.56**	**16.32**	0.61

5.2.2 Image Rain Removal. **Datasets and Metrics.** For image rain removal, we train our RI-GAN on the Rain100H dataset, which contains 1800 image labels. Each label contains a rainy image and a corresponding rain-free image. We evaluate the results on the Rin100H (contains 100 image labels) and Rin100L (contains 200 image labels) test datasets. Similar to the haze-removal task, we also use the PSNR and SSIM between the ground-truth image and the predicted result to evaluate the rain-removal performance.

Image rain removal aims to remove the rain-streak layer from the input image and obtain a rain-free result. We use a rainy image as the corrupted input image and the corresponding ground-truth rain-free image as the supervision to train our model. Fig. 10 shows image rain-removal results compared with four state-of-the-art rain-removal methods, viz., CRDNet [35], RecDerain [38], RCDNet [43], and RLNet [2]. From the results, we can find that our methods can recover the rain-free images without unexpected patterns and better preserve image details.

We also evaluate our rain removal results compared with these two methods on the Rin100H and Rin100L test datasets. For a fair comparison, we train both these methods on the Rain100H dataset. The results are summarized in Table 4, from which we can observe that our proposed RI-GAN can also be directly used for rain removal and achieves preferable results compared with the two state-of-the-art methods.

Table 4. Quantitative comparison results of rain removal on the Rain100H and Rain100L testing dataset using the PSNR metric (the larger, the better) and the SSIM metric (the larger, the better).

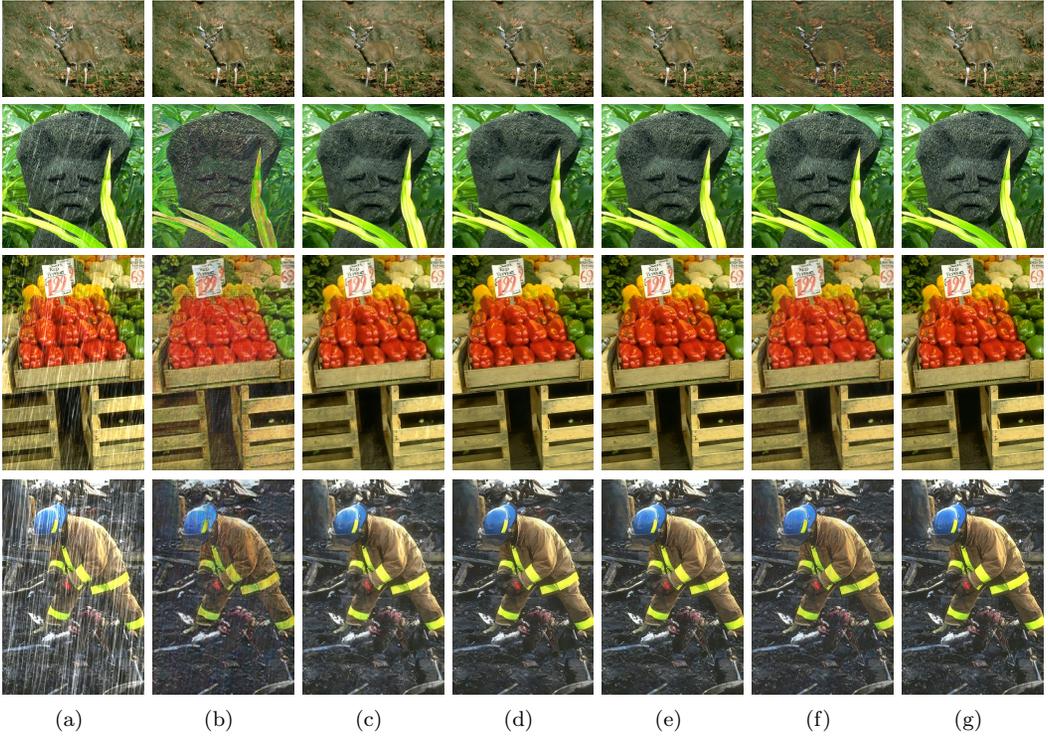| Methods	Venue/Year	Rain100H		Rain100L	
		PSNR	SSIM	PSNR	SSIM
CRDNet	SPL/2020	15.75	0.50	20.70	0.59
RecDerain	TIP/2020	14.40	0.51	23.36	0.71
RCDNet	CVPR/2020	28.70	0.88	33.41	0.96
RLNet	CVPR/2021	**29.50**	**0.89**	**33.98**	0.96
RI-GAN	TOMM/2022	28.94	**0.89**	**33.98**	**0.97**

Fig. 10. Image rain-removal results. From left to right: input images (a); rain-removal results of CRDNet [35] (b), RecDerain [38] (c), RCDNet [43] (d), RLNet [2] (e), our RI-GAN (f), and ground-truth rain-free images (g).

5.2.3 Highlight Removal. **Datasets and metrics.** We perform highlight-removal experiments on the SHIQ dataset and LIME dataset. The SHIQ dataset contains 16K image pairs of highlight images and highlight-free images. We select 6000 image pairs from the SHIQ dataset for training and 1000 image pairs for testing. The LIME dataset contains 80K+ image pairs, and we select 1000 image pairs for testing. Similar to the haze-removal task, we also use the PSNR and SSIM between the ground-truth image and the predicted result to evaluate the highlight-removal performance.

Highlights in the images can reduce the image quality and weaken the visual effect. Highlight removal aims to recover a highlight-free image. We train the proposed RI-GAN using the SHIQ dataset to perform the image highlight-removal task. We compare our method with five different highlight removal methods: Shen et al. [39], Yang et al. [53], Akashi et al. [1], Yamamoto et al. [52], and Fu et al. [14]. For a fair comparison, we use the same dataset to train all the learning-based methods. Fig. 11 shows some image highlight-removal results. From the images, we can observe that our results are closer to the ground-truth highlight-free images. Table 5 summarizes the comparison results on two testing datasets. From the table, we can observe that our proposed RI-GAN can obtain superior evaluation results compared with the state-of-the-art methods.
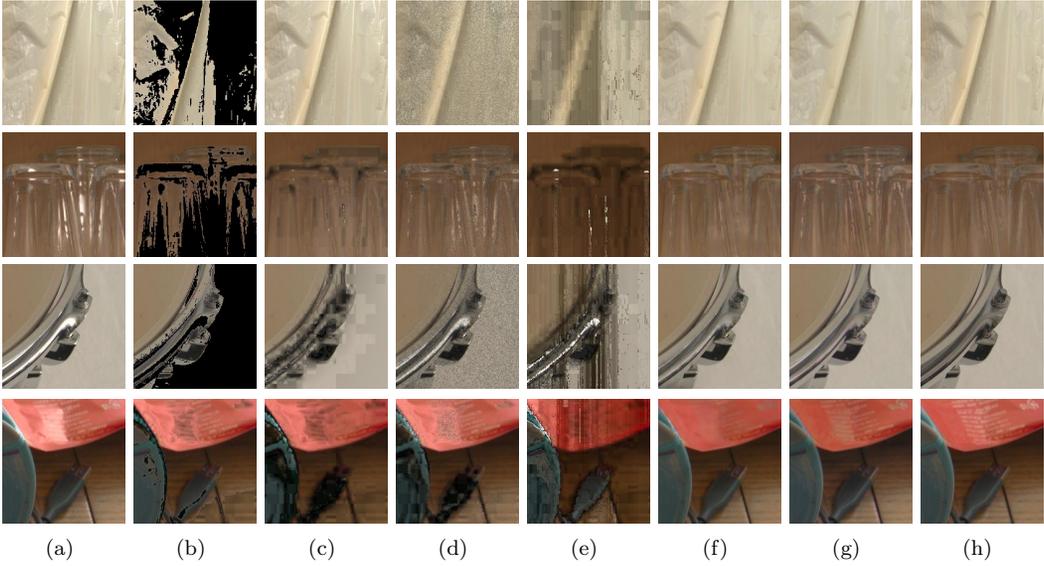
Fig. 11. Image highlight-removal results. From left to right: input images (a); highlight-removal results of Shen et al. [39] (b), Yang et al. [53] (c), Akashi et al. [1] (d), Yamamoto et al. [52] (e), Fu et al. [14] (f), our RI-GAN (g), and ground-truth highlight-free images (h).

Table 5. Quantitative comparison results of image highlight removal on the testing datasets of SHIQ and LIME using the metrics of PSNR (the larger, the better) and SSIM (the larger, the better).

Methods/Year	Venue	SHIQ		LIME	
		PSNR	SSIM	PSNR	SSIM
Shen et al.[39]	AO/2013	13.72	0.44	14.11	0.50
Yang et al.[53]	PAMI/2015	14.44	0.52	17.53	0.61
Akashi et al.[1]	CVIU/2016	14.27	0.57	16.33	0.57
Yamamoto et al.[52]	MTA/2019	20.21	0.64	21.17	0.69
Fu et al.[14]	CVPR/2021	32.37	0.87	34.28	0.90
RI-GAN	TOMM/2022	**33.64**	**0.94**	**35.83**	**0.92**

5.3 Ablation Study

To evaluate the importance of each component in our proposed RI-GAN, we perform ablation experiments on four different image processing tasks with and without each specific component. Specifically, we focus on four major components: the residual generator, the illumination generator, feature-crossing at G_{refine}, and data argumentation for D_{ref}. With these components, we reconstruct four different variants with or without the corresponding components, and test them with different image processing tasks. We mark the four variants as RI-GAN$_1$, RI-GAN$_2$, RI-GAN$_3$, and RI-GAN$_4$. Each variant removes one component from the full RI-GAN architecture. For example, RI-GAN$_1$ is the network without the residual generator at G_{res}, and RI-GAN$_2$ is the network without the illumination generator at G_{illum}.
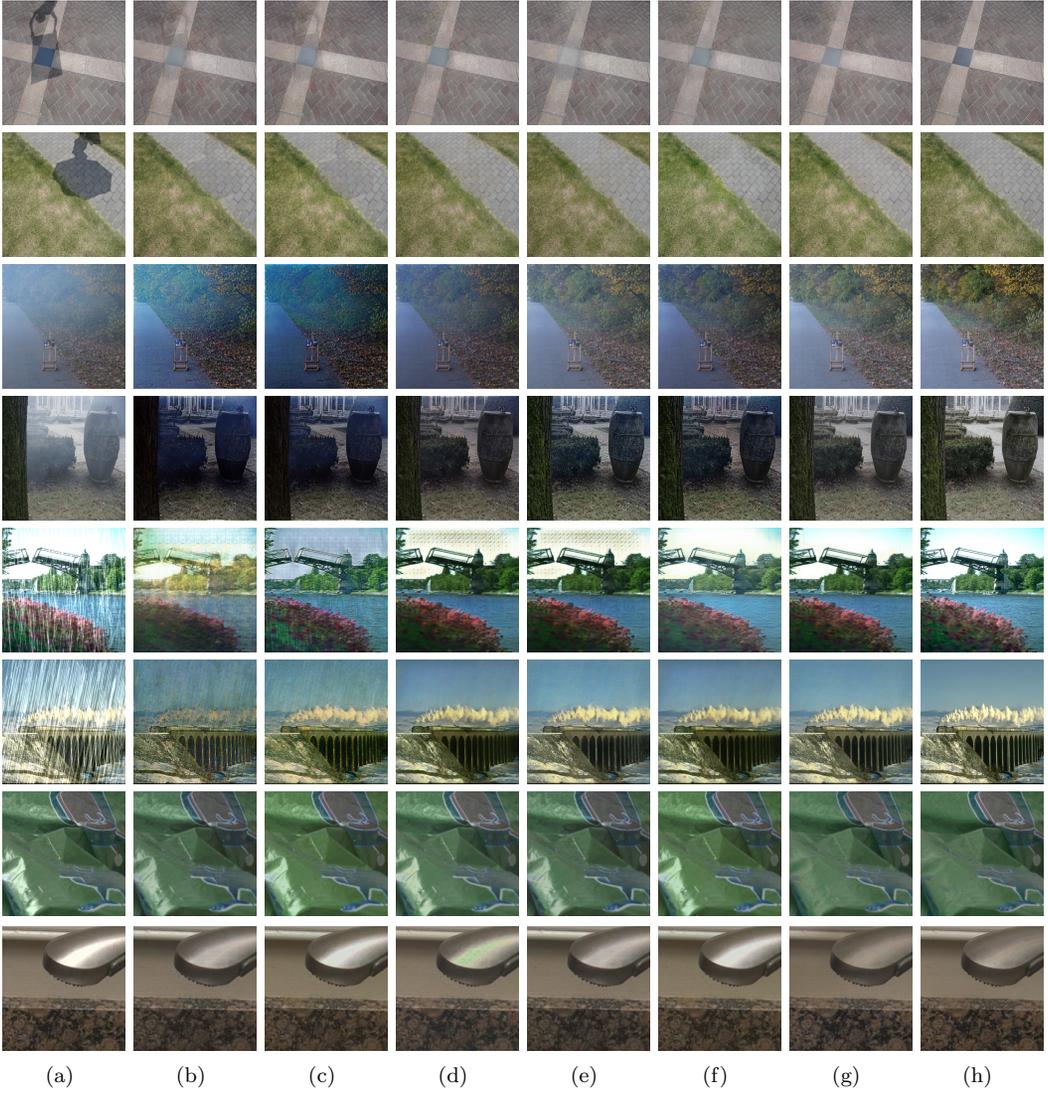
Fig. 12. Image processing results for different tasks. From left to right: input images (a); results of RI-GAN$_1$ (b), RI-GAN$_2$ (c), RI-GAN$_3$ (d), RI-GAN$_4$ (e), RIS-GAN (f), our RI-GAN (g), and the ground-truth (h).

The results for the ablation study are summarized in Table 6. From the table, we can observe that: (1) the four variants demonstrate lower performance than RI-GAN, and the combination leads to the best performance; and (2) the four components are necessary to ensure high-quality image processing results. We also provide visual results in Fig. 12 that show that our RI-GAN recovers the best clean images and makes the results more realistic.

We use a two-branch structure in the refinement generator for shadow detection and shadow removal. To illustrate the positive interaction between the two tasks, we design two variants for shadow detection and shadow removal separately, denoted as DetectionRI-GAN

Table 6. Quantitative results of ablation study on four different image processing tasks. ①, ②, ③, and ④ represent the residual generator, illumination generator, feature-crossing in G_{refine}, and data argumentation for D_{ref}, respectively. The test reconstructed network maintains the component marked with ✓ and ignores the component marked with ×. The test datasets for the four applications are ISTD, O-haze, Rain100H, and SHIQ in our quantitative experiments. Specifically, RMSE(S) and RMSE(A) in the table show the RMSE on shadow regions and on the whole image, respectively.

Methods	Components				Shadow Removal		Dehazing		Rain Removal		Highlight Removal	
	①	②	③	④	RMSE(S)	RMSE(A)	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
RI-GAN$_1$	×	✓	✓	✓	9.83	7.07	10.79	0.48	20.73	0.61	27.18	0.93
RI-GAN$_2$	✓	×	✓	✓	9.28	7.03	11.43	0.49	24.69	0.73	27.89	0.89
RI-GAN$_3$	✓	✓	×	✓	8.71	6.58	11.33	0.50	24.78	0.76	29.91	0.92
RI-GAN$_4$	✓	✓	✓	×	8.57	6.34	12.85	0.54	27.57	0.80	31.05	0.92
RI-GAN	✓	✓	✓	✓	**8.12**	**5.91**	**12.88**	**0.56**	**28.94**	**0.89**	**33.64**	**0.94**

Table 7. Quantitative results of a user study on four different image processing tasks. Values in the table show the percentage of the most natural images in the corresponding application results.

Shadow Removal		Dehazing		Rain Removal		Highlight Removal	
Methods	Percentage (%)	Methods	Percentage (%)	Methods	Percentage (%)	Methods	Percentage (%)
RI-GAN	26.51	RI-GAN	25.94	RI-GAN	26.48	RI-GAN	27.42
ST-CGAN	24.09	DenseGAN	23.66	CRDNet	23.58	Yang et al.	25.70
CANet	23.88	GridDehazeNet	24.87	RecDerain	23.62	Yamamoto et al.	19.71
FusionNet	25.52	LapDehazeNet	25.53	RLNet	26.32	Fu et al.	27.17

and ShadowRI-GAN. DetectionRI-GAN removes the shadow decoder in the refinement generator, and only performs the shadow-detection task. ShadowRI-GAN removes the detection decoder, and only conducts the shadow-removal task. We evaluate the performance of DetectionRI-GAN and ShadowRI-GAN on the ISTD test dataset. DetectionRI-GAN obtains a BER value of 1.97, and ShadowRI-GAN obtains an RMSE of 6.04. The two values are higher than that of RI-GAN. The superior values produced by our joint model demonstrate the positive interaction between the shadow-detection and shadow-removal tasks in our RI-GAN.

We adopt the joint discriminator to ensure that the four discriminators share the same architecture with the same parameter values. To evaluate the effectiveness of joint discriminator, we reconstruct our RI-GAN by applying four independent discriminators. The four discriminators have the same architecture, but they do not share the common parameters. The evaluation value of the RMSE on the ISTD dataset for the shadow-removal task is 6.01. This value is poorer than that of the RI-GAN using the joint discriminator. This confirms that the joint discriminator improves the performance of the proposed network.

5.4 User Study

To further evaluate the effectiveness of the proposed RI-GAN, we conducted a user study to evaluate the visual performance of the proposed RI-GAN and some other methods for the four image processing tasks. We selected 100 random volunteers to complete our user study. For each image processing task, we prepared 100 sets of images. Each set contained the image processing results using state-of-the-art methods and our RI-GAN. For each volunteer, we randomly showed them ten sets of images for each task. They selected the most natural image in each set. Thus, there were 1000 selected results for each task. The evaluated results are summarized in Table 7. From the table, we can observe that our RI-GAN obtains a higher

proportion of the most natural results compared with the corresponding state-of-the-art methods for each task. These results indicate that our proposed framework produces more pleasing results for users.

5.5 Discussion

Running time. In our testing experiments for image processing tasks, the time is mainly spent loading the pre-model. Generally, 50 to 60 seconds are required to load the trained model. After loading the pre-model, processing 100 images (256×256) consume 12 to 15 seconds.

Limitations. Our RI-GAN is designed for image-level image processing applications, where input and output images are the same size. So our RI-GAN is unfit for tasks that have different image resolutions between the input and the output, such as super-resolution. Furthermore, the proposed model is a supervised method, which requires different training datasets for different tasks.

6 CONCLUSION

In this paper, we have proposed a general framework, called RI-GAN, for shadow detection and removal that exploits residual image and illumination map. Our RI-GAN is composed of a coarse stage and a refinement stage. With the results predicted at the coarse stage, the refinement stage produces a shadow mask to identify shadow regions and remove shadows in the image. Specifically, the refinement generator employs a cross-encoding module to make full use of intrinsic properties of the image, such that the network learns more useful information for promoting high-quality results. Further, joint discriminators and data augmentation render the network more stable and effective. Our extensive experiments confirmed the advantages of incorporating residual image and illumination for image processing tasks. Also, the experiments in shadow removal and three other image processing tasks demonstrated the effectiveness and flexibility of the proposed framework.

The proposed RI-GAN is still a pixel-based analysis method. In the future, we would like to improve it and apply patch-based analysis to further improve the performance of the shadow-removal task. Moreover, we will extend the current work to video-level tasks and apply residual image and illumination to solve other vision-related problems.

ACKNOWLEDGMENTS

REFERENCES

[1] Y. Akashi, T. Okatani, and . 2016. Separation of reflection components by sparse non-negative matrix factorization. *Computer Vision and Image Understanding* 146 (2016), 77–85.

[2] C. Chen and H. Li. 2021. Robust Representation Learning with Feedback for Single Image Deraining. In *CVPR*.

[3] Ting Chen, Xiaohua Zhai, Marvin Ritter, Mario Lucic, and Neil Houlsby. 2019. Self-Supervised GANs via Auxiliary Rotation Loss. In *Conference on Computer Vision and Pattern Recognition (CVPR)*. 12146–12155.

[4] Zipei Chen, Chengjiang Long, Ling Zhang, and Chunxia Xiao. 2021. CANet: A Context-Aware Network for Shadow Removal. In *ICCV*. 4743–4752.

[5] Zhihao Chen, Lei Zhu, Liang Wan, Song Wang, and Pheng Ann Heng. 2020. A Multi-task Mean Teacher for Semi-supervised Shadow Detection. In *CVPR*.

[6] Rita Cucchiara, Costantino Grana, Massimo Piccardi, Andrea Prati, and Stefano Sirotti. 2002. Improving shadow suppression in moving object detection with HSV color information. In *IEEE Intelligent Transportation Systems. Proceedings.* 334–339.

[7] Xiaodong Cun, Chi-Man Pun, and Cheng Shi. 2020. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting GAN. In *AAAI*. 10680–10687.

[8] S. D. Das and S. Dutta. 2020. Fast Deep Multi-patch Hierarchical Network for Nonhomogeneous Image Dehazing. In *CVPR*.

[9] Bin Ding, Chengjiang Long, Ling Zhang, and Chunxia Xiao. 2019. ARGAN: Attentive Recurrent Generative Adversarial Network for Shadow Detection and Removal.. In *Conference on Computer Vision and Pattern Recognition (CVPR)*. 10212–10221.

[10] Liu Feng and Michael Gleicher. 2008. Texture-Consistent Shadow Removal. In *European Conference on Computer Vision (ECCV)*. 437–450.

[11] G. D. Finlayson, S. D. Hordley, C. Lu, and M. S. Drew. 2005. On the removal of shadows from images. *T-PAMI* (2005).

[12] Lan Fu, Changqing Zhou, Qing Guo, Felix Juefei-Xu, Hongkai Yu, Wei Feng, Yang Liu, and Song Wang. 2021. Auto-Exposure Fusion for Single-Image Shadow Removal. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.

[13] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. 2016. A weighted variational model for simultaneous reflectance and illumination estimation. In *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2782–2790.

[14] Fu Gang, Zhang Qing, Zhu Lei, Li Ping, and Chunxia Xiao. 2021. A Multi-Task Network for Joint Specular Highlight Detection and Removal. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.

[15] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, X. Bing, and Y. Bengio. 2014. Generative Adversarial Nets. In *Advances in neural information processing systems (NeurIPS)*.

[16] Maciej Gryka, Michael Terry, and Gabriel J. Brostow. 2015. *Learning to Remove Soft Shadows*. Vol. 34. 1–15 pages.

[17] Ruiqi Guo, Qieyun Dai, and D. Hoiem. 2011. Single-image shadow detection and removal using paired regions. In *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2033–2040.

[18] Xiaojie Guo, Yu Li, and Haibin Ling. 2017. LIME: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing* 26, 2 (2017), 982–993.

[19] K. He, J. Sun, and Xiaoou Tang. 2011. Single Image Haze Removal Using Dark Channel Prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, 12 (2011), 2341–2353.

[20] Zhang He and Vishal M. Patel. 2018. Densely Connected Pyramid Dehazing Network. In *Conference on Computer Vision and Pattern Recognition (CVPR)*. 3194–3203.

[21] X. Hu, C. W. Fu, Z. Lei, Q. Jing, and P. A. Heng. 2020. Direction-aware Spatial Context Features for Shadow Detection and Removal. 42, 11 (2020), 2795–2808.

[22] Gang Hua, Chengjiang Long, Ming Yang, and Yan Gao. 2018. Collaborative Active Visual Recognition from Crowds: A Distributed Ensemble Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40, 3 (2018), 582–594.

[23] S. H. Khan, M Bennamoun, F Sohel, and R Togneri. 2016. Automatic Shadow Detection and Removal from a Single Image. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38, 3 (2016), 431–446.

[24] Hieu Le, Yago Vicente, F Tomas, Vu Nguyen, Minh Hoai, and Dimitris Samaras. 2018. A+ D net: Training a shadow detector with adversarial shadow attenuation. In *European Conference on Computer Vision (ECCV)*. 680–696.

[25] B. Li, X. Peng, Z. Wang, J. Xu, and F. Dan. 2017. AOD-Net: All-in-One Dehazing Network. In *IEEE International Conference on Computer Vision (ICCV)*. 4780–4788.

[26] Feng Liu and Michael Gleicher. 2008. Texture-Consistent Shadow Removal. *ECCV* (2008).

[27] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. 2019. GridDehazeNet: Attention-Based Multi-Scale Network for Image Dehazing. In *International Conference on Computer Vision (ICCV)*. 7313–7322.

[28] Chengjiang Long and Gang Hua. 2015. Multi-class multi-annotator active learning with robust gaussian process for visual recognition. In *International Conference on Computer Vision (ICCV)*. 2839–2847.

[29] Chengjiang Long and Gang Hua. 2017. Correlational Gaussian Processes for Cross-Domain Visual Recognition. In *Conference on Computer Vision and Pattern Recognition (CVPR)*. 4932–4940.

[30] Chengjiang Long, Xiaoyu Wang, Gang Hua, Ming Yang, and Yuanqing Lin. 2014. Accurate object detection with location relaxation and regionlets re-localization. In *Asia Conference on Computer Vision*.

[31] Wenhan Luo, Peng Sun, Fangwei Zhong, Wei Liu, Tong Zhang, and Yizhou Wang. 2020. End-to-end Active Object Tracking and Its Real-world Deployment via Reinforcement Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42, 6 (2020), 1317–1332.

[32] Ivana Mikic, Pamela C Cosman, Greg Kogut, and Mohan M Trivedi. 2000. Moving shadow and object detection in traffic scenes. (2000).

[33] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. 2018. Spectral normalization for generative adversarial networks. *International Conference on Learning Representations* (2018).

[34] Vu Nguyen, Tomas F. Yago Vicente, Maozheng Zhao, Minh Hoai, and Dimitris Samaras. 2017. Shadow Detection with Conditional Generative Adversarial Networks. In *ICCV*.

[35] Long Peng, Aiwen Jiang, Qiaosi Yi, and Mingwen Wang. 2020. Cumulative Rain Density Sensing Network for Single Image Derain. *IEEE Signal Processing Letters* 27 (2020), 406–410.

[36] Liangqiong Qu, Jiandong Tian, Shengfeng He, Yandong Tang, and Rynson W. H. Lau. 2017. DeshadowNet: A Multi-context Embedding Deep Network for Shadow Removal. In *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2308–2316.

[37] N Bharath Raj and N Venkateswaran. 2018. Single Image Haze Removal using a Generative Adversarial Network. In *Conference on Computer Vision and Pattern Recognition (CVPR)*. 37–42.

[38] Dongwei Ren, Wei Shang, Pengfei Zhu, Qinghua Hu, and Wangmeng Zuo. 2020. Single Image Deraining Using Bilateral Recurrent Network. *IEEE Transactions on Image Processing* 29, 99 (2020), 6852–6863.

[39] H. Shen, Z. Zheng, and . 2013. Real-time highlight removal using intensity ratio. *Applied Optics* 52, 19 (2013), 4483–4493.

[40] Yael Shor and Dani Lischinski. 2008. The Shadow Meets the Mask: Pyramid-Based Shadow Removal. *Computer Graphics Forum* 27, 2 (2008), 577–586.

[41] Oleksii Sidorov. 2019. Conditional GANs for Multi-Illuminant Color Constancy: Revolution or yet Another Approach?. In *Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 1748–1758.

[42] Tomas F Yago Vicente, Le Hou, Chenping Yu, Minh Hoai, and Dimitris Samaras. 2016. Large-Scale Training of Shadow Detectors with Noisily-Annotated Shadow Examples. (2016).

[43] H. Wang, Q. Xie, Q. Zhao, and D. Meng. 2020. A Model-driven Deep Neural Network for Single Image Rain Removal. *CVPR*.

[44] Jifeng Wang, Xiang Li, Le Hui, and Jian Yang. 2018. Stacked Conditional Generative Adversarial Networks for Jointly Learning Shadow Detection and Shadow Removal. In *Conference on Computer Vision and Pattern Recognition (CVPR)*. 1788–1797.

[45] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. 2019. Underexposed Photo Enhancement Using Deep Illumination Estimation. In *Conference on Computer Vision and Pattern Recognition (CVPR)*. 6842–6850.

[46] Tianyu Wang, Xiaowei Hu, Qiong Wang, Pheng Ann Heng, and Chi Wing Fu. 2020. Instance Shadow Detection. In *Conference on Computer Vision and Pattern Recognition (CVPR)*. 1877–1886.

[47] Jingjiang Wei, Chengjiang Long, Hua Zhou, and Chunxia Xiao. 2019. Shadow Inpainting and Removal Using Generative Adversarial Networks with Slice Convolutions. *Computer Graphics Forum* 38, 7 (2019), 381–392.

[48] Tai Pang Wu, Chi Keung Tang, Michael S. Brown, and Heung Yeung Shum. 2007. Natural shadow matting. *Acm Transactions on Graphics* 26, 2 (2007), 8.

[49] B. Xiao, Z. Zheng, X. Chen, C. Lv, Y. Zhuang, and T. Wang. 2021. Single UHD Image Dehazing via Interpretable Pyramid Network. In *CVPR*.

[50] Chunxia Xiao, Ruiyun She, Donglin Xiao, and Kwan Liu Ma. 2013. Fast Shadow Removal Using Adaptive Multi-Scale Illumination Transfer. *Computer Graphics Forum* 32, 8 (2013), 207–218.

[51] Chunxia Xiao, Donglin Xiao, Ling Zhang, and Lin Chen. 2013. Efficient Shadow Removal Using Subregion Matching Illumination Transfer. *Computer Graphics Forum* 32, 7 (2013), 421–430.

[52] T. Yamamoto, A. Nakazawa, and . 2019. General Improvement Method of Specular Component Separation Using High-Emphasis Filter and Similarity Function. *Ite Transactions on Media Technology Applications* 7, 2 (2019), 92–102.

[53] Q. Yang, J. Tang, and N. Ahuja. 2015. Efficient and Robust Specular Highlight Removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37, 6 (2015), 1304–1311.

[54] Lin Yun-Hsuan, Chen Wen-Chin, and Chuang Yung-Yu. 2020. BEDSR-Net: A Deep Shadow Removal Network from a Single Document Image. *Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), 12902–12911.

[55] Ling Zhang, Chengjiang Long, Xiaolong Zhang, and Chunxia Xiao. 2020. RIS-GAN: Explore Residual and Illumination with Generative Adversarial Networks for Shadow Removal. *Proceedings of the AAAI Conference on Artificial Intelligence* 34, 7 (2020), 12829–12836.

[56] Ling Zhang, Qingan Yan, Yao Zhu, Xiaolong Zhang, and Chunxia Xiao. 2019. Effective shadow removal via multi-scale image decomposition. *The Visual Computer* 35, 6-8 (2019), 1091–1104.

[57] Ling Zhang, Qing Zhang, and Chunxia Xiao. 2015. Shadow Remover: Image Shadow Removal Based on Illumination Recovering Optimization. *IEEE Transactions on Image Processing* 24, 11 (2015), 4623–4636.

[58] Quanlong Zheng, Xiaotian Qiao, Ying Cao, and Rynson W.H. Lau. 2019. Distraction-Aware Shadow Detection. In *Conference on Computer Vision and Pattern Recognition (CVPR)*. 5162–5171.

[59] Lei Zhu, Zijun Deng, Xiaowei Hu, Chi Wing Fu, and Pheng Ann Heng. 2018. Bidirectional Feature Pyramid Network with Recurrent Attention Residual Modules for Shadow Detection. In *ECCV*.

[60] Lei Zhu, Zijun Deng, Xiaowei Hu, Chi-Wing Fu, Xuemiao Xu, Jing Qin, and Pheng-Ann Heng. 2018. Bidirectional Feature Pyramid Network with Recurrent Attention Residual Modules for Shadow Detection. In *European Conference on Computer Vision (ECCV)*. 122–137.

[61] Lei Zhu, Ke Xu, Zhanghan Ke, and Rynson W.H. Lau. 2021. Mitigating Intensity Bias in Shadow Detection via Feature Decomposition and Reweighting. In *ICCV*. 4682–4691.