

RIS-GAN: Explore Residual and Illumination with Generative Adversarial Networks for Shadow Removal

Ling Zhang¹, Chengjiang Long^{2*}, Xiaolong Zhang¹ Chunxia Xiao^{3*}

¹Wuhan University of Science and Technology, Wuhan, Hubei, China

²Kitware Inc. Clifton Park, NY, USA

³School of Computer Science, Wuhan University, Wuhan, Hubei, China

{zhling, xiaolong.zhang}@wust.edu.cn, chengjiang.long@kitware.com, cxxiao@whu.edu.cn

Abstract

Residual images and illumination estimation have been proved very helpful in image enhancement. In this paper, we propose a general and novel framework RIS-GAN which explores residual and illumination with Generative Adversarial Networks for shadow removal. Combined with the coarse shadow-removal image, the estimated negative residual images and inverse illumination maps can be used to generate indirect shadow-removal images to refine the coarse shadow-removal result to the fine shadow-free image in a coarse-to-fine fashion. Three discriminators are designed to distinguish whether the predicted negative residual images, shadow-removal images, and the inverse illumination maps are real or fake jointly compared with the corresponding ground-truth information. To our best knowledge, we are the first one to explore residual and illumination for shadow removal. We evaluate our proposed method on two benchmark datasets, *i.e.*, SRD and ISTD, and the extensive experiments demonstrate that our proposed method achieves the superior performance to state-of-the-arts, although we have no particular shadow-aware components designed in our generators. Our source code is available at <https://github.com/zhling2020/RIS-GAN>.

Introduction

Shadow is a ubiquitous natural phenomenon, which is appeared when the light is partial or complete blocked, bringing down the accuracy and effectiveness of some computer vision tasks, such as target tracking, object detection and recognition(Mikic et al. 2000; Long et al. 2014; Cucchiara et al. 2002; Hua et al. 2013; Long and Hua 2015; Long and Hua 2017; Hua et al. 2018; Luo et al. 2019), image segmentation and intrinsic image decomposition (Li and Snavely 2018). Therefore, it is necessary to conduct shadow removal to improve the visual effect of image and video editing, such as film and television post-editing. However, it is still a very challenging problem to remove shadow in complex scenes due to illumination change, texture variation, and other environmental factors.

*This work was co-supervised by Chengjiang Long and Chunxia Xiao.

Copyright © 2020, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

A variety of existing works including traditional methods(Shor and Lischinski 2008; Xiao et al. 2013a; Zhang et al. 2019), and learning-based methods (Gryka, Terry, and Brostow 2015; Wang et al. 2018a; Le et al. 2018) have been developed to solve this challenging problem. Different from traditional methods that highly rely on some prior knowledge (*e.g.*, constant illumination and gradients) and often bring obvious artifacts on the shadow boundaries, learning-based methods especially recent deep learning methods like (Hu et al. 2018) and (Sidorov 2019) have achieved some advances. However, the effectiveness of these methods highly depends on the training dataset and the designed network architectures. When the training set is insufficient or the network model is deficient, they are insufficient to produce desired shadow detection masks and shadow-removal images. Also, most of existing deep learning methods just focus on shadow itself, without well exploring other extra information like residual and illumination for shadow removal.

In this paper, we propose a general framework RIS-GAN to explore both residual and illumination with Generative Adversarial Networks for shadow removal, unlike Sidorov's AngularGAN (Sidorov 2019) which just introduces an illumination-based angular loss without estimating illumination color or illumination color map. As illustrated in Figure 1, our RIS-GAN consists of four generators in the encoder-decoder structure and three discriminators. Such four generators are designed to generate negative residual images, intermediate shadow-removal images, inverse illumination maps, and refined shadow-removal images. In principle, unlike the existing deep learning methods (Qu et al. 2017; Hu et al. 2018; Zhu et al. 2018) which are designed particular for shadow-removal, any kinds of encoder-decoder structures can be used as our generators.

For the residual generator, we follow the idea of negative residual (Fu et al. 2017) and let the generator take a shadow image to generate a negative residual for detecting shadow area and recover a shadow-lighter or shadow-free image by applying a element-wise addition with the input shadow image indirectly. For the illumination generator, we design it based on the Retinex model (Fu et al. 2016; Guo, Li, and Ling 2016; Wang et al. 2019) where the ground-truth shadow-free image can be considered as a re-

flectance image, and the shadow image is the observed image. The output of the illumination generator is a inverse illumination map which can be used for shadow region detection and recovering a shadow-removal image by applying a element-wise multiplication with the input shadow image indirectly.

We shall emphasize that we use the refinement generator to refine the shadow-removal images obtained from the shadow removal generator in a coarse-to-fine fashion. Besides the coarse shadow-removal results, we also incorporate two indirect shadow-removal images via the explored negative residual and inverse illumination to recover the final fine shadow-removal image. With this treatment, the shadow-removal refinement generator has three complementary input sources, which ensures the high-quality shadow removal results.

Like all the Generative Adversarial Networks (GANs), our proposed RIS-GAN adopts the adversarial training process (Goodfellow et al. 2014) between the four generators and three discriminators alternatively to generate high-quality negative residual images, inverse illumination maps, and a shadow-removal images. It is worth mentioning that we design a cross loss function to make sure the recovered shadow-removal image is consistent with the explored residual and illumination. We also adopt a joint discriminator (He and Patel 2018) to ensure that three discriminators share the same architecture with the same parameter values to judge whether the generated results are real or fake compared with the corresponding ground-truth, which can make sure all the produced results are indistinguishable from the corresponding ground-truth. With the number of epochs increases, both generators and discriminators improve their functionalities so that it becomes harder and harder to distinguish a generated output from the corresponding ground-truths. Therefore, after a certain large number of training epochs, we can utilize the learned parameters in the generators to generate a negative residual image, a inverse illumination map, and a shadow-removal image.

Different from the existing shadow detection and removal methods, our main contributions can be summarized as three-fold: (1) we are the first one to propose a general and novel framework RIS-GAN with generators in an encoder-decoder structure to explore residual and illumination between shadow and shadow-free images for shadow removal; (2) the correlation among residual, illumination and shadow has been well explored within the cross loss function and the joint discriminator and we are able to get complementary input sources for better improving the quality of shadow-removal results; and (3) without any particular shadow-aware components in our encoder-decoder structured generators, our proposed RIS-GAN still achieves the outperformance to art-of-the-arts. Such experimental results clearly demonstrates the efficacy of the proposed approach.

Related Work

Shadow removal is to recover a shadow-free image. One typical group of traditional methods is to recover the illumination in shadow regions using illumination transfer (Xiao et al. 2013b; Guo, Dai, and Hoiem 2011; Zhang, Zhang, and

Xiao 2015; Khan et al. 2016), which borrow the illumination from non-shadow regions to shadow regions. Another typical group of traditional methods involves gradient domain manipulation (Finlayson et al. 2005; Liu and Gleicher 2008; Feng and Gleicher 2008). Due to the influence of the illumination change at the shadow boundary, both illumination and gradient based methods cannot well handle the boundary problems, especially in the presence of complex texture or color distortion.

Recently, deep neural networks are widely introduced for shadow removal through analyzing and learning the mapping relation between shadow image and the corresponding shadow-free image. Hu *et al.* (Hu et al. 2018) used multiple convolutional neural networks to learn image features for shadow detection and remove shadows in the image. Qu *et al.* (Qu et al. 2017) proposed an end-to-end DeshadowNet to recover illumination in shadow regions. Wang *et al.* (Wang et al. 2018a) proposed a stacked conditional generative adversarial network (ST-CGAN) for image shadow removing. Sidorov (Sidorov 2019) proposed an end-to-end architecture named AngularGAN oriented specifically to the color constancy task, without estimating illumination color or illumination color map. Wei *et al.* (Wei et al. 2019) proposed a two-stage generative adversarial network for shadow inpainting and removal with slice convolutions. Ding *et al.* (Ding et al. 2019) proposed an attentive recurrent generative adversarial network (ARGAN) to detect and remove shadow with multiple steps. Different from existing methods, our proposed RIS-GAN makes full use of the explored negative residual image and the inverse illumination map for generating more accurate shadow-removal results.

Approach

We explore the residual and illumination between shadow images and shadow-free images via Generative Adversarial Networks (GANs) (Goodfellow et al. 2014) due to the ability of GAN in style transfer and details recovery (Li and Wand 2016). The intuition behind is that the residual and illumination explored can provide informative additional details and insights for shadow removal.

The proposed framework RIS-GAN for shadow removal with multiple GANs is illustrated in Figure 1. Given an input shadow image I , three encoder-decoder structures are applied to generate residual image I_{res} , intermediate shadow-removal image I_{imd} , and inverse illumination map S_{inv} . With the element-wise addition with the input shadow image and the residual image, we are able to get an indirect shadow-removal image I_{rem}^1 . With the element-wise production with the input shadow image and the inverse illumination, we are able to get another indirect shadow-removal image I_{rem}^2 . We can apply another encoder-decoder structure to refine the coarse shadow-removal image I_{coarse} and the two indirect shadow-removal images I_{rem}^1 and I_{rem}^2 to produce a fine shadow-removal image I_{fine} .

Our RIS-GAN is composed of four generators in the same encoder-decoder structure and three discriminators. The four generators are residual generator, removal generator, illumination generator, detection generator, and refinement generator, denoted as G_{res} , G_{rem} , G_{illum} , and G_{ref} , respec-

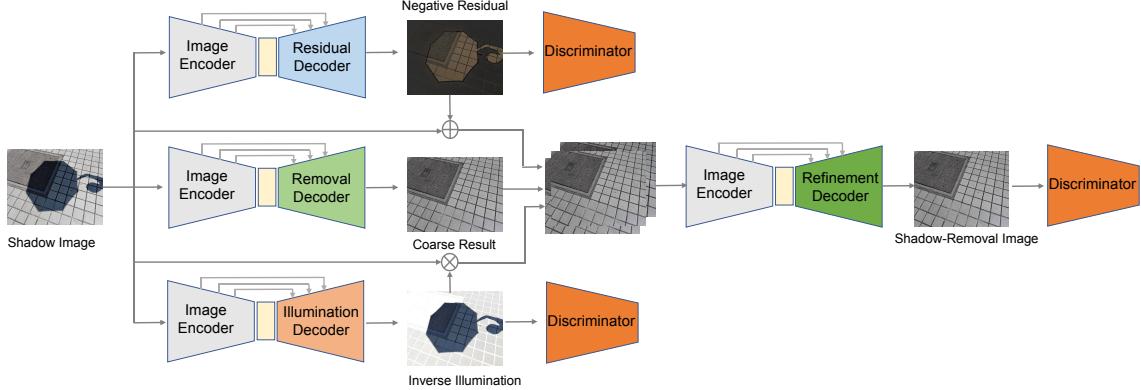


Figure 1: The framework of our proposed RIS-GAN, which is composed of four generators in the encoder-decoder structure and three discriminators. It takes a shadow image as input and outputs the negative residual image, inverse illumination map, and shadow-removal image in an end-to-end manner. Note that these four generators share the same architecture, and the three discriminators share the same parameters. No particular shadow-aware components are designed in the whole framework.

tively, for generating a negative residual image, a coarse shadow-removal image, an inverse illumination map, and a fine shadow-removal image. The three discriminators, D_{res} , D_{illum} and D_{ref} , share the same architecture and the same parameters to determine the generated residual images, inverse illumination maps, and final shadow-removal images to be real or fake, compared with the ground-truth residual images, inverse illumination maps, and the shadow-free images. For the readers convenience, we summarize the relations between the above mentioned notations as:

$$I_{res} = G_{res}(I), S_{inv} = G_{illum}(I) \quad (1)$$

$$I_{rem}^1 = I \oplus I_{res}, I_{rem}^2 = I \otimes S_{inv} \quad (2)$$

$$I_{coarse} = G_{rem}(I) \quad (3)$$

$$I_{fine} = G_{ref}(I_{coarse}, I_{rem}^1, I_{rem}^2) \quad (4)$$

Our network takes the shadow image as input and outputs the negative residual images, inverse illumination maps, and fine shadow-removal images in an end-to-end manner. The alternative training between generators and discriminators ensures the good quality of the prediction results.

In the following, we are going to describe the details of our generators, discriminators, loss functions, as well as the implementation details.

Encoder-Decoder Generators

In principle, any encoder-decoder structures can be used in our RIS-GAN framework. In this paper, we don't want to design any particular shadow-aware components in the framework, and just adopt the DenseUNet architecture (Raj and Venkateswaran 2018) as the implementation of each encoder-decoder generator. DenseUNet consists of a contracting path to capture context and a symmetric expanding path to upsample. Different with the conventional UNet architecture, DensUNet adds Dense Blocks in the network, which concatenate every layers output with its input, and feed it to the next layer. This enhances information and gradient flow in our four encoder-decoder generators:

- Residual Generator G_{res} is to get a residual image that is close to the ground-truth residual image I_{res}^{gt} obtained between shadow image and the corresponding shadow-free image I^{gt} , i.e., $I_{res} = I^{gt} - I$.
- Removal Generator G_{rem} is to produce a coarse shadow-removal image I_{coarse} .
- Illumination Generator G_{illum} is to estimate the inverse illumination map in the shadow image. Note that the ground-truth inverse illumination map is calculated based on the Retinex-based image enhancement methods (Fu et al. 2016; Guo, Li, and Ling 2016; Wang et al. 2019), i.e., $S_{inv}^{gt} = I^{gt} * I^{-1}$, where I^{gt} can be considered as a reflectance image, and I is the observed image.
- Refinement Generator G_{ref} is to refine the current intermediate shadow-removal image and two indirect shadow-removal images with the explored residual and illumination to formulate the final shadow-removal image.



Figure 2: The visualization of residual and illumination for shadow removal. From left to right are the shadow images I , the indirect shadow-removal image I_{rem}^1 by residual, the indirect shadow-removal image I_{rem}^2 by illumination, the fine shadow-free image I_{fine}^1 , and the ground-truth shadow-free image I^{gt} , respectively.

To better understand our detector generator and refinement generators, we visualize some examples in Figure 2. As we can observe, the indirect shadow-removal images obtained by residual and illumination have good quality and are complimentary to the intermediate shadow-removal image for further refinement to get the final shadow-removal

image.

Joint Discriminator

The discriminator is a convolutional network, which is used to distinguish the predicted residual image, the final shadow-removal image, and the estimated illumination produced by the generators to be real or fake, compared with the corresponding ground truth. To make sure all the produced results are indistinguishable from the corresponding ground truths, we make use of a GAN with joint discriminator (He and Patel 2018). The joint discriminator is trained to learn a joint distribution to judge whether the produced results are real or fake.

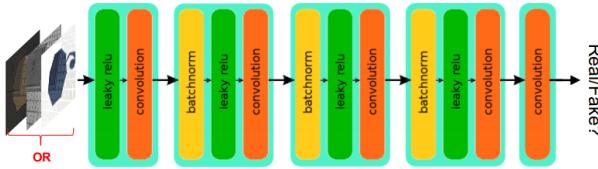


Figure 3: The architecture of the discriminator in our RIS-GAN. It consists of five convolution layers with batchnorm and leaky ReLU activations. For all these five convolution layers, all the kernel sizes are 4×4 ; the strides are 4×4 except the first convolution layer whose stride is 2×2 ; and the number of output channels is: $64 \rightarrow 128 \rightarrow 256 \rightarrow 512 \rightarrow 1$.

Our discriminator consists of five convolution layers, each followed by a batch normalization and a Leaky ReLU activation function, and one fully connected layer. The last fully connected layer outputs the probability value that the input image (result produced by generator) is a real image. Figure 3 gives details of the discriminator.

It is worth noting that we use the spectrum normalization method (Miyato et al. 2018) to stabilize the training process of discriminator network, because spectral normalization is a simple and effective standardized method for limiting the optimization process of the discriminator in GAN, and it can make the whole generators perform better.

Loss Functions

To get a robust parametric mode, the loss functions that we use to optimize the proposed RIS-GAN has five components: shadow removal loss \mathcal{L}_{rem} , residual loss \mathcal{L}_{res} , illumination loss \mathcal{L}_{illum} , cross loss \mathcal{L}_{cross} and adversarial loss \mathcal{L}_{adv} . The total loss \mathcal{L} is can be written as

$$\mathcal{L} = \lambda_1 \mathcal{L}_{res} + \lambda_2 \mathcal{L}_{rem} + \lambda_3 \mathcal{L}_{illum} + \lambda_4 \mathcal{L}_{cross} + \mathcal{L}_{adv}, \quad (5)$$

where $\lambda_1, \lambda_2, \lambda_3$, and λ_4 are hyperparameters.

Shadow removal loss is defined with visual-consistency loss and perceptual-consistency loss, i.e.,

$$\mathcal{L}_{rem} = \mathcal{L}_{vis} + \beta_1 \mathcal{L}_{percept}, \quad (6)$$

where β is the weight parameter. \mathcal{L}_{vis} is visual-consistency loss for removal generator which is calculated using L1-norm between the shadow removal result and the ground

truth, and $\mathcal{L}_{percept}$ is perceptual-consistency loss aiming to preserve image structure. To specify,

$$\mathcal{L}_{vis} = ||I^{gt} - I_{fine}||_1 + ||I^{gt} - I_{coarse}||_1. \quad (7)$$

$$\begin{aligned} \mathcal{L}_{percept} = & ||\text{VGG}(I^{gt}) - \text{VGG}(I_{fine})||_2^2 \\ & + ||\text{VGG}(I^{gt}) - \text{VGG}(I_{coarse})||_2^2. \end{aligned} \quad (8)$$

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

Residual loss can be perceived as the obscured brightness in shadow regions, i.e.,

$$\mathcal{L}_{res} = ||I_{res}^{gt} - G_{res}(I)||_1. \quad (9)$$

Illumination loss calculates L1-norm between the illumination result generated by G_{illum} and the ground truth of inverse illumination map S_{inv}^{gt} . Then illumination loss for illumination branch can be denoted as:

$$\mathcal{L}_{illum} = ||S_{inv}^{gt} - G_{illum}(I)||_1. \quad (10)$$

Cross loss is designed to ensure the consistency and correlation among residual, illumination and shadow information as

$$\mathcal{L}_{cross} = ||I^{gt} - (G_{res}(I) \oplus I)||_1 + \beta_2 ||I^{gt} - (G_{illum}(I) \otimes I)||_1. \quad (11)$$

Adversarial loss \mathcal{L}_{adv} is the joint adversarial loss for the network, and is described as:

$$\begin{aligned} \mathcal{L}_{adv} = & \mathbb{E}_{(I, I^{gt}, I_{res}^{gt}, S_{inv}^{gt})} [\log(D_{ref}(I^{gt})) \\ & + \log(1 - D_{ref}(G_{ref}(G_{res}(I), G_{rem}(I), G_{illum}(I)))) \\ & + \log(D_{res}(I_{res}^{gt})) + \log(1 - D_{res}(G_{res}(I))) \\ & + \log(D_{illum}(S_{inv}^{gt})) + \log(1 - D_{illum}(G_{illum}(I)))] \end{aligned} \quad (12)$$

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

Overall, our objective for the training task is solving a mini-max problem which aims to find a saddle point between generator and discriminator of our network.

Implementation Details

Our proposed method is implemented in Tensorflow in a computer with Intel(R) Xeon(R) Silver 4114 CPU@2.20GHz 192G RAM NVIDIA GeForce GTX 1080Ti. In our experiments, the input size of image is 256×256 . The learning rate value is set to 0.001. The parameters $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \beta_1$ and β_2 are set to 10, 100, 1, 1, 0.1 and 0.2 in our experiments, respectively. The minibatch size is 2. The initial learning rate is set as 0.001. We use Momentum Optimizer to optimize our generator and use Adam Optimizer for the discriminator. We alternatively train our generator and discriminator for 10,000 epochs.

Experiments

To verify the effectiveness of our proposed RIS-GAN, we conduct various experiments on the SRD dataset (Qu et al. 2017) and the ISTD dataset (Wang et al. 2018a). The SRD dataset has 408 pairs of shadow and shadow-free images publicly available. The ISTD dataset contains 1870 image

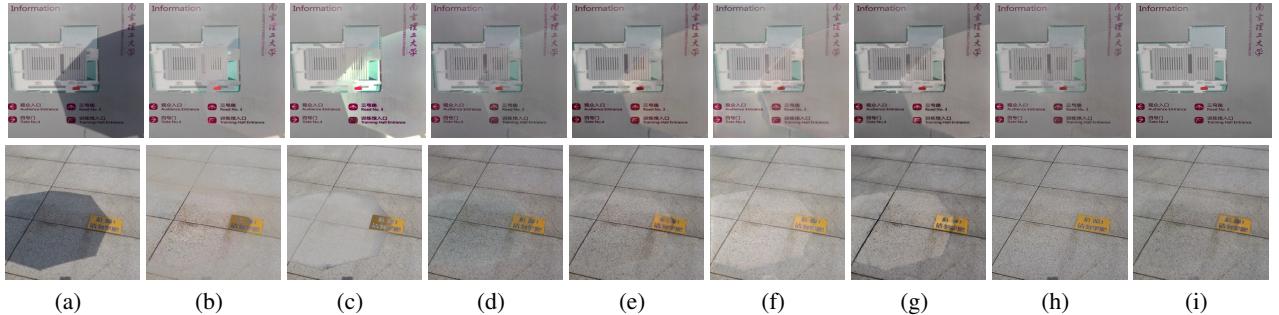


Figure 4: Shadow removal results. From left to right are: input images (a); shadow-removal results of Guo (b), Zhang (c), De-shadowNet (d), ST-CGAN (e), DSC (f), AngularGAN (g), and our RIS-GAN (h); and the corresponding ground truth shadow-free images (i).

triplets of shadow image, shadow mask and shadow-free image. Such a dataset has 135 different simulated shadow environments and the scenes are very diverse. Both these two datasets contains various kinds of shadow scenes. In this paper, we use the 1330 pairs of shadow and shadow-free images from the ISTD dataset for training, and use the rest 540 pairs for testing. We also use the trained RIS-GAN model to evaluate on the SRD dataset.

Regarding the metrics, 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. We also conduct a user study for comprehensive evaluation.

Comparison with State-of-the-arts

We compare our RIS-GAN with the state of-the-art methods including the two traditional methods, *i.e.*, Guo (Guo, Dai, and Hoiem 2011) and Zhang (Zhang, Zhang, and Xiao 2015) and the recent learning-based methods, *i.e.*, DeshadowNet (Qu et al. 2017), DSC (Hu et al. 2018), ST-CGAN (Wang et al. 2018a), and AngularGAN (Sidorov 2019). Note that shadow removal works on the pixel and recovers the value of the pixel, and therefore we add two frameworks, *i.e.*, Global/Local-GAN (Iizuka, Simo-Serra, and Ishikawa 2017) for image inpainting and Pix2Pix-HD (Wang et al. 2018b) for image translation, as another two shadow removal baselines for solid validation. To make the 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 1 and Table 2. From the table, we can observe that among all the competing methods, our proposed RIS-GAN achieves the best RMSE values in shadow regions, non-shadow regions, and the entire images on the two datasets, although we have no particular shadow-aware components designed in our generators. This suggests that the recovered shade-removal images obtained by our RIS-GAN is much closer to the corresponding ground-truth shadow-free images. As the main difference between our RIS-GAN and the state-of-the-art deep learning, exploring residual and illuminations demonstrates the great advantages in the task of shadow removal.

To further explain the outperformance of our proposed RIS-GAN, we provides some visualization results in Fig-

Table 1: Quantitative comparison results of shadow removal on the SRD dataset using the metric RMSE (the smaller, the better). S, N, and A represent shadow regions, non-shadow region, and the entire image, respectively.

Methods	Venue/Year	S	N	A
Guo	CVPR/2011	31.06	6.47	12.60
Zhang	TIP/2015	9.50	6.90	7.24
Global/Local-GAN	TOG/2017	19.56	8.17	16.33
Pix2Pix-HD	CVPR/2018	17.33	7.79	12.58
Deshadow	CVPR/2017	17.96	6.53	8.47
ST-CGAN	CVPR/2018	18.64	6.37	8.23
DSC	CVPR/2018	11.31	6.72	7.83
AngularGAN	CVPRW/2019	17.63	7.83	15.97
RIS-GAN	AAAI/2020	8.22	6.05	6.78

Table 2: Quantitative comparison results of shadow removal on the ISTD dataset in term of RMSE.

Methods	Venue/Year	S	N	A
Guo	CVPR/2011	18.95	7.46	9.30
Zhang	TIP/2015	9.77	7.12	8.16
Global/Local-GAN	TOG/2017	13.46	7.67	8.82
Pix2Pix-HD	CVPR/2018	10.63	6.73	7.37
Deshadow	CVPR/2017	12.76	7.19	7.83
ST-CGAN	CVPR/2018	10.31	6.92	7.46
DSC	CVPR/2018	9.22	6.50	7.10
AngularGAN	CVPRW/2019	9.78	7.67	8.16
RIS-GAN	AAAI/2020	8.99	6.33	6.95

ure 4 covering the traditional methods and the learning-based methods for shadow removal. As we can see in Figure 4(b), Guo can recover illumination in shadow regions and may produce unnatural shadow removal results especially for images with different textures in the shadow regions. Zhang cannot well handle the illumination change in shadow boundaries so that the recovered shadow-removal images have boundary problem, such as color distortion or texture loss, as shown in Figure 4(c). Compared with these two traditional methods, our proposed RIS-GAN not only effectively recovers illumination in shadow regions, but also reconstruct the illumination and texture in shadow boundaries, as shown in Figure 4(h).

As for the recent deep learning methods, DeshadowNet, ST-CGAN, and DSC deal with the images in aspect of color space, without considering the aspect of illumination. This

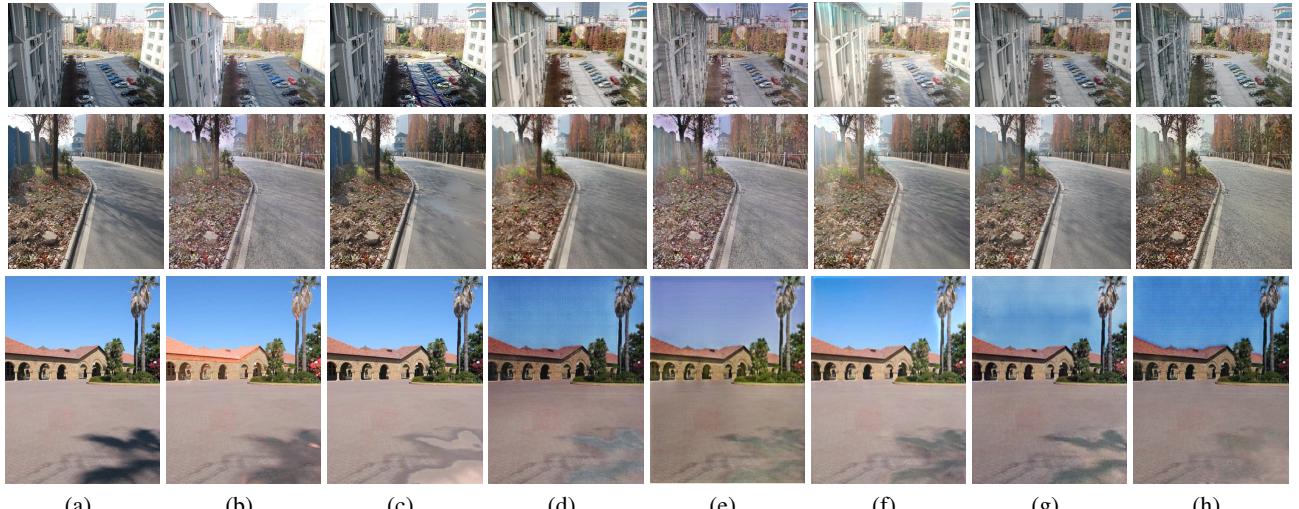


Figure 5: Shadow removal results. From left to right are: input images (a); shadow-removal results of Guo (b), Zhang (c), DeshadowNet (d), ST-CGAN (e), DSC (f), and AngularGAN (g); and shadow-removal results of our RIS-GAN.

may lead to unsatisfied shadow-removal results like color distortion or incomplete shadow removal, as shown in Figure 4 (d-f). AngularGAN introduces an angle loss which is defined based on the consideration of illumination, which makes the illumination of non-shadow regions very close to the corresponding ground-truth shadow-free images. It is worth mentioning here that the illumination is not well incorporated in the recovery of shadow-removal images, which causes some inconsistency between shadow-region and non-shadow region, as seen in Figure 4 (g). In contrast, taking residual negative residual images and the inverse illumination maps into consideration, our proposed RIS-GAN can effectively remove shadows and produce good result for both simple and complex scene image. The recovered illumination in shadow regions is consistent with surrounding environment and the texture details in shadow regions are well preserved, as shown in the Figure 4(h).

To further verify the robustness and the potential of our proposed RIS-GAN in the complicated scenes, we collect a few shadow images in the real-world life to run the experiments and report in Figure 5. Apparently, the shadow-removal images recovered by our proposed RIS-GAN look more realistic, with little artifacts. This observation demonstrates the robustness of our RIS-GAN for complex scenes.

User study We conduct a user-study with 100 random volunteers to evaluate the visual performance of our proposed RIS-GAN and some other shadow removal methods. We prepare 300 sets of images. Each set contains five shadow removal results by using methods of our RIS-GAN, DeshadowNet, ST-CGAN, DSC, and AngularGAN, respectively. For each volunteer, we randomly show them twenty sets images to choose which shadow removal image is the most natural in each set. Then there will be 2000 select results. Counting all the results, we find that 29.65% of shadow-removal images generated by our RIS-GAN are chosen as the most natural shadow removal result, while 14.85%, 19.55%, 20.35% and 15.60% of shadow removal results are chosen by Deshadow, ST-CGAN, DSC, and Ag-

ularGAN, respectively.

Ablation Study

To further evaluate some components of our proposed RIS-GAN, we design a series of variants as follows:

- **BASE:** take the input shadow images as the shadow-removal result.
- **R-GAN:** use G_{res} only and take I_{rem}^1 as the shadow-removal result.
- **I-GAN:** use G_{illum} only and take I_{rem}^2 as the shadow-removal result.
- **S-GAN:** use G_{rem} only and take I_{coarse} as the shadow-removal result.
- **RS-GAN:** remove G_{illum} , and G_{ref} takes I_{res} and I_{coarse} to get the fine shadow-removal image.
- **IS-GAN:** remove G_{res} , and G_{ref} takes S_{inv} and I_{coarse} to get the fine shadow-removal image.
- **RIS-GAN₁:** remove \mathcal{L}_{adv} from Equation 5.
- **RIS-GAN₂:** remove \mathcal{L}_{cross} from Equation 5.

We train the above seven GAN variants on the same training data and evaluate the shadow-removal results on both the SRD dataset and the ISTD dataset. The results are summarized in Table 3, from which we can observe: (1) all the GAN variants can recover shadow-light or shadow-free in the shadow regions when compared with BASE; (2) the negative residual image from the residual generator and the inverse illumination map can help improve the performance of the shadow-removal refinement, and the combination leads to the best performance; and (3) the loss functions \mathcal{L}_{adv} and \mathcal{L}_{cross} are necessary to ensure the high-quality shadow-removal results, this clearly demonstrates the advantage of the joint discriminator and cross learning among three outputs. We also provide the visualization in Figure 6, from which we can clearly see that our RIS-GAN recovers the best details of the shadow-removal regions and looks more realistic.

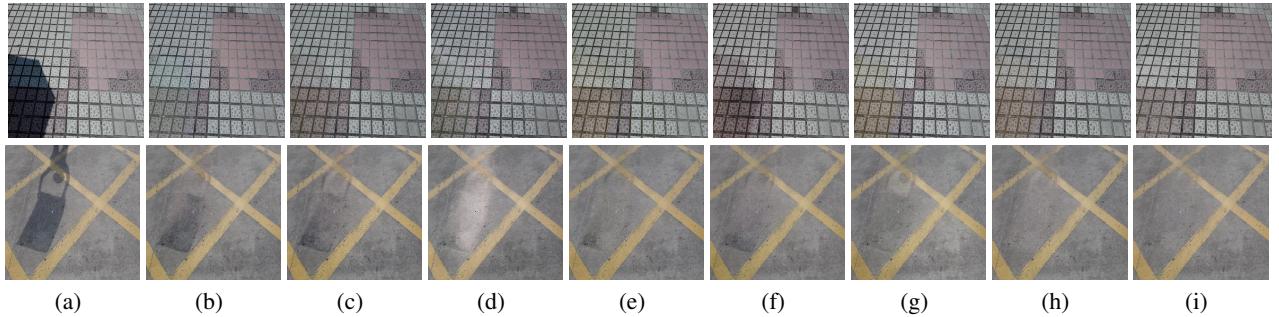


Figure 6: Shadow removal results. From left to right are: input images (a); shadow-removal results of R-GAN (b), S-GAN (c), I-GAN (d), RS-GAN (e), IS-GAN (f), RIS-GAN₁ (g), RIS-GAN₂ (h), and RIS-GAN (i), respectively.

Table 3: Quantitative shadow-removal results of ablation study on the SRD and ISTD datasets in term of RMSE.

Methods	SRD			ISTD		
	S	N	A	S	N	A
BASE	35.74	8.88	15.14	35.74	8.88	15.14
R-GAN	11.41	7.33	8.37	12.09	7.08	8.24
S-GAN	12.06	7.65	8.85	16.98	9.71	11.27
I-GAN	14.82	8.54	11.55	9.02	12.10	15.31
RS-GAN	10.33	6.44	7.35	9.43	6.26	7.03
IS-GAN	9.57	6.32	7.16	10.16	6.37	7.20
RIS-GAN ₁	9.37	6.64	7.32	9.17	7.16	7.59
RIS-GAN ₂	9.51	6.87	7.27	11.01	8.98	7.91
RIS-GAN	8.22	6.05	6.78	8.99	6.33	6.95

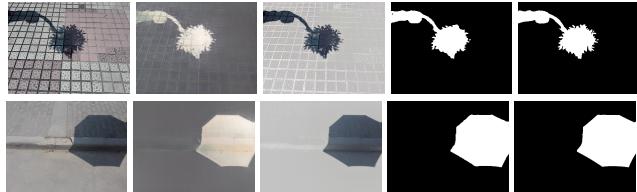


Figure 7: The visualization of detection results. From left to right are input images, negative residual images, inverse illumination maps, prediction shadow masks based on the explored negative residual images and inverse illumination maps, and ground-truth shadow masks, respectively.

Discussion

To better explore the potential of our proposed RIS-GAN, we also visualize the shadow detection masks, and extend the current approach for video shadow removal.

Shadow detection. Although we focus on shadow removal rather than detection, our RIS-GAN also can get the shadow detection masks based on the negative residual images and the inverse illumination maps. Figure 7 shows the promising detection results. We observe that both the generated negative residual images and inverse illumination maps effectively distinguish the shadow and non-shadow regions well.

Extension to video We apply our RIS-GAN to handle shadow videos by processing each frame in order. Figure 8 presents the shadow-removal results for the frames every 100 milliseconds. From this Figure we can observe that the video shadow-removal results by applying image-level shadow removal approach to video directly are not good

enough and there is still room for better improvement.



Figure 8: The visualization of shadow-removal results in a video. Note the frames are extracted every 100 milliseconds.

What's more, although our proposed RIS-GAN framework is designed for shadow removal, it is not limited to the shadow removal only. It is easy to be extended and applied to general image-level applications such as rain removal, image dehazing, intrinsic image decomposition, as well as other image style-transfer tasks.

Conclusions

In this paper, we have proposed a general and novel framework RIS-GAN to explore the residual and illumination for shadow removal. The correlation among residual, illumination and shadow has been well explored under a unified end-to-end framework. With the estimated negative residual image and inverse illumination map incorporating into the shadow refinement, we are able to get complementary input sources to generate a high-quality shadow-removal image. The extensive experiments have strongly confirmed the advantages of incorporating residual and illumination for shadow removal.

Our future work includes extending the current work to video-level shadow removal and applying the explored residual and illumination to solve the real challenging vision problems, such as image illumination enhancement.

Acknowledgments

This work was partly supported by Key Technological Innovation Projects of Hubei Province (2018AAA062), the National Natural Science Foundation of China (No. 61672390, No. 61972298, NO. 61902286, No.61972299, No.U1803262), the National Key Research and Development Program of China (2017yf-b1002600), China Postdoctoral Science Found (No. 070307), and the Joint laboratory Foundation of Xiaomi Company and Wuhan University. Chunxia Xiao is the corresponding author.

References

- [Cucchiara et al. 2002] Cucchiara, R.; Grana, C.; Piccardi, M.; Prati, A.; and Sirotti, S. 2002. Improving shadow suppression in moving object detection with hsv color information. In *T-ITS*.
- [Ding et al. 2019] Ding, B.; Long, C.; Zhang, L.; and Xiao, C. 2019. Argan: Attentive recurrent generative adversarial network for shadow detection and removal. In *ICCV*.
- [Feng and Gleicher 2008] Feng, L., and Gleicher, M. 2008. Texture-consistent shadow removal. In *ECCV*.
- [Finlayson et al. 2005] Finlayson, G. D.; Hordley, S. D.; Lu, C.; and Drew, M. S. 2005. On the removal of shadows from images. *T-PAMI*.
- [Fu et al. 2016] Fu, X.; Zeng, D.; Huang, Y.; Zhang, X.-P.; and Ding, X. 2016. A weighted variational model for simultaneous reflectance and illumination estimation. In *CVPR*.
- [Fu et al. 2017] Fu, X.; Huang, J.; Zeng, D.; Huang, Y.; Ding, X.; and Paisley, J. 2017. Removing rain from single images via a deep detail network. In *CVPR*.
- [Goodfellow et al. 2014] Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial nets. In *NeurIPS*.
- [Gryka, Terry, and Brostow 2015] Gryka, M.; Terry, M.; and Brostow, G. J. 2015. *Learning to Remove Soft Shadows*. ACM TOG.
- [Guo, Dai, and Hoiem 2011] Guo, R.; Dai, Q.; and Hoiem, D. 2011. Single-image shadow detection and removal using paired regions. In *CVPR*.
- [Guo, Li, and Ling 2016] Guo, X.; Li, Y.; and Ling, H. 2016. Lime: Low-light image enhancement via illumination map estimation. *TIP*.
- [He and Patel 2018] He, Z., and Patel, V. M. 2018. Densely connected pyramid dehazing network. *CVPR*.
- [Hu et al. 2018] Hu, X.; Fu, C. W.; Zhu, L.; Qin, J.; and Heng, P. A. 2018. Direction-aware spatial context features for shadow detection and removal. In *CVPR*.
- [Hua et al. 2013] Hua, G.; Long, C.; Yang, M.; and Gao, Y. 2013. Collaborative active learning of a kernel machine ensemble for recognition. In *ICCV*.
- [Hua et al. 2018] Hua, G.; Long, C.; Yang, M.; and Gao, Y. 2018. Collaborative active visual recognition from crowds: A distributed ensemble approach. *T-PAMI*.
- [Iizuka, Simo-Serra, and Ishikawa 2017] Iizuka, S.; Simo-Serra, E.; and Ishikawa, H. 2017. Globally and locally consistent image completion. *TOG*.
- [Khan et al. 2016] Khan, S. H.; Bennamoun, M.; Sohel, F.; and Togneri, R. 2016. Automatic shadow detection and removal from a single image. *T-PAMI*.
- [Le et al. 2018] Le, H.; Vicente, Y.; Tomas, F.; Nguyen, V.; Hoai, M.; and Samaras, D. 2018. A+ d net: Training a shadow detector with adversarial shadow attenuation. In *ECCV*.
- [Li and Snavely 2018] Li, Z., and Snavely, N. 2018. Learning intrinsic image decomposition from watching the world. *CVPR*.
- [Li and Wand 2016] Li, C., and Wand, M. 2016. Precomputed real-time texture synthesis with markovian generative adversarial networks. In *ECCV*.
- [Liu and Gleicher 2008] Liu, F., and Gleicher, M. 2008. Texture-consistent shadow removal. In *ECCV*.
- [Long and Hua 2015] Long, C., and Hua, G. 2015. Multi-class multi-annotator active learning with robust gaussian process for visual recognition. In *ICCV*.
- [Long and Hua 2017] Long, C., and Hua, G. 2017. Correlational gaussian processes for cross-domain visual recognition. In *CVPR*.
- [Long et al. 2014] Long, C.; Wang, X.; Hua, G.; Yang, M.; and Lin, Y. 2014. Accurate object detection with location relaxation and regionlets re-localization. In *ACCV*.
- [Luo et al. 2019] Luo, W.; Sun, P.; Zhong, F.; Liu, W.; Zhang, T.; and Wang, Y. 2019. End-to-end active object tracking and its real-world deployment via reinforcement learning. *T-PAMI*.
- [Mikic et al. 2000] Mikic, I.; Cosman, P. C.; Kogut, G.; and Trivedi, M. M. 2000. Moving shadow and object detection in traffic scenes.
- [Miyato et al. 2018] Miyato, T.; Kataoka, T.; Koyama, M.; and Yoshida, Y. 2018. Spectral normalization for generative adversarial networks. *arXiv*.
- [Qu et al. 2017] Qu, L.; Tian, J.; He, S.; Tang, Y.; and Lau, R. W. H. 2017. Deshadownet: A multi-context embedding deep network for shadow removal. In *CVPR*.
- [Raj and Venkateswaran 2018] Raj, N. B., and Venkateswaran, N. 2018. Single image haze removal using a generative adversarial network. *CVPR*.
- [Shor and Lischinski 2008] Shor, Y., and Lischinski, D. 2008. The shadow meets the mask: Pyramid-based shadow removal. In *CGF*.
- [Sidorov 2019] Sidorov, O. 2019. Conditional gans for multi-illuminant color constancy: Revolution or yet another approach? In *CVPRW*.
- [Wang et al. 2018a] Wang, J.; Li, X.; Hui, L.; and Yang, J. 2018a. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *CVPR*.
- [Wang et al. 2018b] Wang, T.-C.; Liu, M.-Y.; Zhu, J.-Y.; Tao, A.; Kautz, J.; and Catanzaro, B. 2018b. High-resolution image synthesis and semantic manipulation with conditional gans. In *CVPR*.
- [Wang et al. 2019] Wang, R.; Zhang, Q.; Fu, C.-W.; Shen, X.; Zheng, W.-S.; and Jia, J. 2019. Underexposed photo enhancement using deep illumination estimation. In *CVPR*.
- [Wei et al. 2019] Wei, J.; Long, C. L.; Zhou, H.; and Xiao, C. 2019. Shadow inpainting and removal using generative adversarial networks with slice convolutions. *CGF*.
- [Xiao et al. 2013a] Xiao, C.; She, R.; Xiao, D.; and Ma, K. L.

2013a. Fast shadow removal using adaptive multi-scale illumination transfer. *CGF*.

[Xiao et al. 2013b] Xiao, C.; Xiao, D.; Zhang, L.; and Chen, L. 2013b. Efficient shadow removal using subregion matching illumination transfer. *CGF*.

[Zhang et al. 2019] Zhang, L.; Yan, Q.; Zhu, Y.; Zhang, X.; and Xiao, C. 2019. Effective shadow removal via multi-scale image decomposition. *TVC*.

[Zhang, Zhang, and Xiao 2015] Zhang, L.; Zhang, Q.; and Xiao, C. 2015. Shadow remover: Image shadow removal based on illumination recovering optimization. *TIP*.

[Zhu et al. 2018] Zhu, L.; Deng, Z.; Hu, X.; Fu, C.-W.; Xu, X.; Qin, J.; and Heng, P.-A. 2018. Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. In *ECCV*.