Pacific Graphics 2019
C. Theobalt, J. Lee, and G. Wetzstein
(Guest Editors)

Volume 38 (2019), Number 7

Shadow Inpainting and Removal Using Generative Adversarial Networks with Slice Convolutions

Jinjiang Wei[1], Chengjiang Long[2†], Hua Zou[1], Chunxia Xiao[1†]
[1]School of Computer Science, Wuhan University, Wuhan, Hubei, China
[2]Kitware Inc., Clifton Park, NY, USA
{weijinjiang, cxxiao, zouhua}@whu.edu.cn, chengjiang.long@kitware.com

Abstract

In this paper, we propose a two-stage top-down and bottom-up Generative Adversarial Networks (TBGANs) for shadow inpainting and removal which uses a novel top-down encoder and a bottom-up decoder with slice convolutions. These slice convolutions can effectively extract and restore the long-range spatial information for either down-sampling or up-sampling. Different from the previous shadow removal methods based on deep learning, we propose to inpaint shadow to handle the possible dark shadows to achieve a coarse shadow-removal image at the first stage, and then further recover the details and enhance the color and texture details with a non-local block to explore both local and global inter-dependencies of pixels at the second stage. With such a two-stage coarse-to-fine processing, the overall effect of shadow removal is greatly improved, and the effect of color retention in non-shaded areas is significant. By comparing with a variety of mainstream shadow removal methods, we demonstrate that our proposed method outperforms the state-of-the-art methods.

CCS Concepts
• ***Computing methodologies*** → *Shadow Inpainting; Shadow Removal; Top-down; Bottom-up; Slice Convolution; Non-local Block; Generative Adversarial Networks;*

1. Introduction

Shadows are common in natural scenes, and they are known to wreak havoc in many computer vision tasks such as image segmentation [QJL*19], object detection and recognition [LWH*14, TM19, HLYG13, LH15, LH17, HLYG18]. Therefore the ability to generate shadow-free images would benefit many computer vision algorithms. Furthermore, for aesthetic reasons, shadow removal can benefit image editing and computational photography algorithms [LKZF19]. Automatic shadow detection and removal from single images, however, are very challenging. A shadow is cast whenever an object occludes an illuminent of the scene, and it is the outcome of complex interactions between the geometry, illumination, and reflectance present in the scene. Identifying shadows is therefore difficult because of the limited information about the scene's properties.

There have been a number of approaches including physics-based method like illumination and texture [FHLD05, MTC07, LG08], traditional statistical learning-based methods with hand-crafted features [CGC*03, WTBS07, GTB15, AHO10], and recent deep learning-based methods [KBST15, VHS17, QTH*17, NVT*17] proposed for shadow removal. Although state-of-the-art shadow removal methods have been able to generate high quality shadow-removal images, the results are still far from perfect especially there are dark shadows in complex scenes.

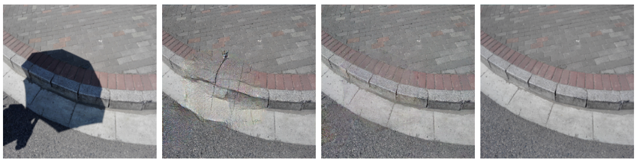

Figure 1: *Given a dark shadow image, we apply our 1st-stage TBGAN to inpaint the shadow area and get a coarse shadow-removal image. Then we apply our 2nd-stage TBGAN to further refine the coarse result from the 1st-stage TBGAN and achieve a high-quality and photo-realistic shadow-removal image. From left to right are input dark shadow image, result with the 1st-stage TBGAN, result with the 2nd-stage TBGAN, and the corresponding ground-truth shadow-free image.*

As a real-world example, Figure 1 provides a dark shadow image in which the dark shadow is cast on the ground. The umbrella shadow area is not clearly visible, and there is even no direct color and texture clue information about the background in the human body shadow area. It is very challenging to remove shadow and

[†] This work was co-supervised by Chengjiang Long and Chunxia Xiao.

recover a shadow-free image on such a kind of dark shadows which appear as “black holes" without providing any useful contextual information in the shadow area.

To handle dark shadows, we propose a novel tow-stage top-down and bottom-up Generative Adversarial Networks (TBGANs) in this paper for inpainting in shadow area and shadow removal in a coarse-to-fine fashion, as illustrated in Figure 2. We design the 1st-stage TBGAN for inpainting the shadow area which we can review as the partially or incompletely missing region, and the inpainting image generated from the network can be regarded as a coarse shadow-removal image. This treatment makes us possible to incorporate sufficient outsider training data to address the issue of lack sufficient training images for shadow-removal. With the shadow inpainting images obtained from the 1st-stage TBGAN, we combine them with the original input shadow images to feed into the 2st-stage TBGAN for further removing shadows as a refinement stage.

Another novelty of our approach is that we propose a top-down convolution module and a bottom-up convolution module which decomposes a 2D convolution/deconvolution kernel into two 1D convolution/deconvolution kernels separately so that it is able to extract long-rage contextual information for down-sampling and up-sampling. Such two modules have been successfully incorporated in the design of both generator and discriminator in the two-stage TBGANs.

It is worth noting that we also design a non-local block in the 2nd-stage TBGAN to explore both the local and global contextual spatial information for the recovery of a high-quality and photo-realistic shadow-removal image. We argue that our non-local block is able to capture long-range dependencies directly by computing interactions between any two positions and models the inter-dependencies of pixels. It is complementary to slice convolutions and helps with capturing long-range dependencies across shadow area and non-shadow area based on inter-dependencies of pixels.

Different from the recent work [DU17] which uses the two-stage networks within the only one GAN to generate missing part from coarse to fine, we use the 1st-stage GAN with slice convolutions/deconvolutions to inpaint shadow and then use the 2nd-stage GAN with the same structure plus a non-local block and 1×1 convolution to refine the recovered shadow-removal images. Besides the two-GAN structure with slice convolutions and non-local module, we shall highlight our biggest novelty is that we are able to explore the correlation between inpainting images and shadow images so that we can make full use of unlimited inpainting images to improve the shadow-removal performance. To our best knowledge, we are the first one to incorporate inpainting datasets for shadow removal.

To sum up, the contribution for this paper are three-fold:

- We propose a novel two-stage top-down and bottom-up Generative Adversarial Networks (TBGAN) for inpainting shadow area as a coarse shadow removal, and then continue to refine to obtain a high quality shadow-free image.
- In such two-stage TBGANs, our top-down encoder and bottom-up decoder with slice convolutions have been successfully adopted to extract long-range contextual spatial information for either down-sampling or up-sampling.
- We also design a non-local block in the 2nd-stage TBGAN to explore both the local and global contextual spatial information for better recovery of shadow-free images.

Our experiments conducted on multiple shadow-removal benchmark datasets have strongly demonstrated the efficacy of the proposed approach.

2. Related Work

A variety of shadow detection and removal methods have been proposed, in this section, we only review the most related works to our method.

Several shadow removal methods are proposed based on gradient domain manipulation [FHLD05, MTC07, LG08]. Finlayson *et al.* [FHLD05] removed the shadows by performing gradient operations for non-shadow regions. This method depends on accurate shadow edges detection and may not produce satisfactory results due to the inaccurate shadow edges detection. Mohan *et al.* [MTC07] removed shadows using gradient domain manipulation. This method works in the gradient domain and reintegrates the image after recognizing and removing gradients caused by shadows. This method requires much user interaction to specify the shadow boundary. Liu *et al.* [LG08] removed shadow by solving a Poisson equation, which constructed a shadow free and texture-consistent gradient field between the shadow and lit area. But this method also depends on accurate shadow boundaries. Shadow matting is also exploited in shadow detection and removal [CGC*03, WTBS07, GTB15]. Chuang *et al.* [CGC*03] proposed a method for shadow extracting and editing which considered the input image as a linear combination of a shadow-free image and a shadow matte image. Instead of considering shadow extraction as the conventional matting equation, Wu *et al.* [WTBS07] supposed shadow effect as a light attenuation problem, and applied user-supplied hints to identify shadow and non-shadow regions. Although these two methods tried to preserve the texture appearance under the extracted shadow, they still do not effectively recover the image detail in the shadow areas. Gryka *et al.* [GTB15] removed soft shadows applying a data-driven method, while this method requires accurate shadow annotation and specified initial shadow matte.

Several shadow removal methods are proposed based on illumination transferring. Inspired by the color transfer theory [RAGS01] Shor *et al.* [SL08] performed shadow removal by applying the illumination in the non-shadow sample region to shadow regions. This method requires that the shadow regions and the sample region share similar texture to produce satisfied results. By improving [SL08], Xiao *et al.* [XSXM13] recovered the illumination under the shadow regions using adaptive multi-scale illumination transfer. Later, Xiao *et al.* [XXZC13] removed shadow by performing illumination transfer between matched subregions. Guo *et al.* [GDH11] also detected as well as removing shadows based on paired regions. Zhang *et al.* [ZZX15] removed the shadows in image by using a coarse-to-fine illumination patch optimization strategy. Due to the size limitation of local patches, this method is difficult to provide satisfying results for shadows with large illumination variances. Given user annotation for shadow mask extraction, Arbel and Hel-Or [AHO10] fit a smooth thin-plate surface model in the shadow regions to produce an approximate shadow matte. With smooth thin-plate approximation, this method does not work well on shadows cast on different types of surfaces, also not handle the shadow boundaries well. As image depth is a useful cue for shadow detection and removal, Xiao *et al.* [XTT14] applied depth information provided by the depth sensor to remove shadows in RGB-D image. The performance of this method depends on the accuracy of the input depth map.

Recently, deep learning networks have been applied to the task of shadow detection [NVT*17, HFZ*18] and removal [KBST15, VHS17, QTH*17, NVT*17]. Nguyen *et al.* [NVT*17] detected shadow using conditional generative adversarial networks. Hu *et al.* [HFZ*18] detected shadow by analyzing image context in a direction-aware spatial context manner. However, these methods only work well on image with simple shadow. They cannot detect accurate shadow with complex scenes. Khan *et al.* [KBST15] used multiple convolutional neural networks to learn useful feature representations for shadow detection. With the shadow detection results, they further proposed Bayesian formulation for shadow removal. Vicente *et al.* [VHS17] considered shadow detection as a problem of labeling image regions and trained a kernel least-squares support vector machine for shadow detection and removal. Qu *et al.* [QTH*17] proposed an end-to-end DeshadowNet to recover illumination in shadow regions, which integrated high-level semantic information, mid-lever appearance information and local image details for shadow matte prediction. Wang *et al.* [WLY18] proposed a STacked Conditional Generative Adversarial Network (ST-CGAN) to perform the two tasks of shadow detection and shadow removal. Ding *et al.* [DLZX19] proposed an attentive recurrent Generative Adversarial Network for shadow detection and removal with multiple progressive steps in a coarse-to-fine fashion.

However, these deep learning based shadow detection and removal methods cannot handle complex shadows (for example, the images with both hard and soft shadow), and may create visible artifacts and in the shadow regions if the shadow type and surface material are not well represented in the training dataset, which greatly restricts possible application scenarios. In this paper, different from above methods, we first perform content completion in the shadow region, and then remove the shadow based on the shadow content inpainting results. We also borrow the STacked Conditional Generative Adversarial Network (ST-CGAN) for our task, while different from [WLY18], we perform the two tasks of shadow inpainting and shadow removal.

3. Method

As illustrated in Figure 2, our proposed two-stage framework for image shadow removal is composed of two TBGANs. The 1st-stage TBGAN is for inpainting in shadow area, which can be pre-trained using a large number of training data existing in image inpainting fields and then fine-tuned by a limited number of image pairs of shadow images and the corresponding shadow-free image. Such a GAN [KF19, LLK19, DLTLM19] can explore the low-frequency information to generate an inpainting image we consider as a coarse shadow-removal image. The 2nd-stage TBGAN is for shadow removal, which makes full use of shadow image and inpainting image obtained from the 1st-stage TBGAN to exploit the high frequency information for generating a shadow-free image. Both these two TBGANs use a symmetric encoder-decoder structure in which slice convolutions are used in both the top-down encoder and bottom-up decoder.

3.1. Slice convolutions in top-down encoder and bottom-up decoder

We propose to use a long slice convolution kernel for down-sampling that guarantees a long-range receptive field in a finite number of layers. As illustrated in Figure 3, each down-sampling consists of a slice convolution. This convolution method can replace the down-sampling process of any convolutional encoder. Each module convolution kernel consists of two components, vertical convolution and horizontal convolution. Suppose the shape of the input tensor is $C_{in} \times H \times W$, C_{in} is the number of input channels, H is the height of the image, W is the width, and the number of output channels is C_{out}. The shapes of the vertical convolution kernel and the horizontal convolution kernel are $W/2 \times 1$ and $1 \times H/2$. Respectively, The down-sampling step size is 2×1 and 1×2, and the margin is filled with 0 accordingly to ensure the size of the output tensor does not change. First, the horizontal convolution is performed, and the output tensor size is halved in the horizontal direction; then the vertical convolution is performed, and the output tensor size is halved in the vertical direction. Batch normalization and activation functions are added after each long and narrow convolution operation. After processing by a top-down module, the output tensor is $C_{out} \times \frac{H}{2} \times \frac{W}{2}$.

The slice convolution in this paper can have half of the receptive field in the initial stage of encoding of the generative network, and the extraction of the primary features can be considered more widely. In each down-sampling process, the convolution kernel can have more than half of the feature receptive range. In the process of halving the size of the convolution kernel, the high-dimensional features of the deep network can be obtained from the top down, and the features are gradually finely filtered and encoded. In order to prevent the convolution kernel parameter growth caused by the increase of the input image size, we use a convolution kernel decomposition operation similar to Google Inception V3 Net [SVI*16] for the large convolution kernel. Thus, even if the input image size is multiplied, the number of parameters of the convolution kernel only shows a logarithmic growth trend.

To keep the symmetry in the encoder-decoder structure, we design a slice deconvolution (see Figure 3) to up-sample the feature maps from $C_{in} \times H \times W$ to $C_{out} \times H \times W * 2$ by a deconvolution with the kernel size $W \times 1$ and then continue to $C_{out} \times H * 2 \times W * 2$ by another decovolution with the kernel size of $1 \times H$.

3.2. The 1st-stage TBGAN for inpainting in shadow area

We combine the images from the Places365 dataset [ZLK*18] and the Irregular Mask dataset [LRS*18]. For each raw image $\mathbf{x}$, we sample a binary mask $\mathbf{m}$ at a random location to obtain input image $\mathbf{z} = \mathbf{x} \odot \mathbf{m}$ with missing regions. We first use the image pairs $\{(\mathbf{z}, \mathbf{x})\}$ to train the 1st-stage TBGAN. Once it converges, we fine-tune the model with ISTD [WLY18], which is a shadow dataset with triplets of shadow image $\mathbf{x}_{shw}$, shadow mask $\mathbf{m}_{shw}$, and the corresponding shadow-free image $\mathbf{x}_{sfree}$. Note that we use the image pairs $\{(\mathbf{x}_{shw}, \mathbf{x}_{sfree})\}$ for training, and we consider $\mathbf{x}_{shw}$ as an

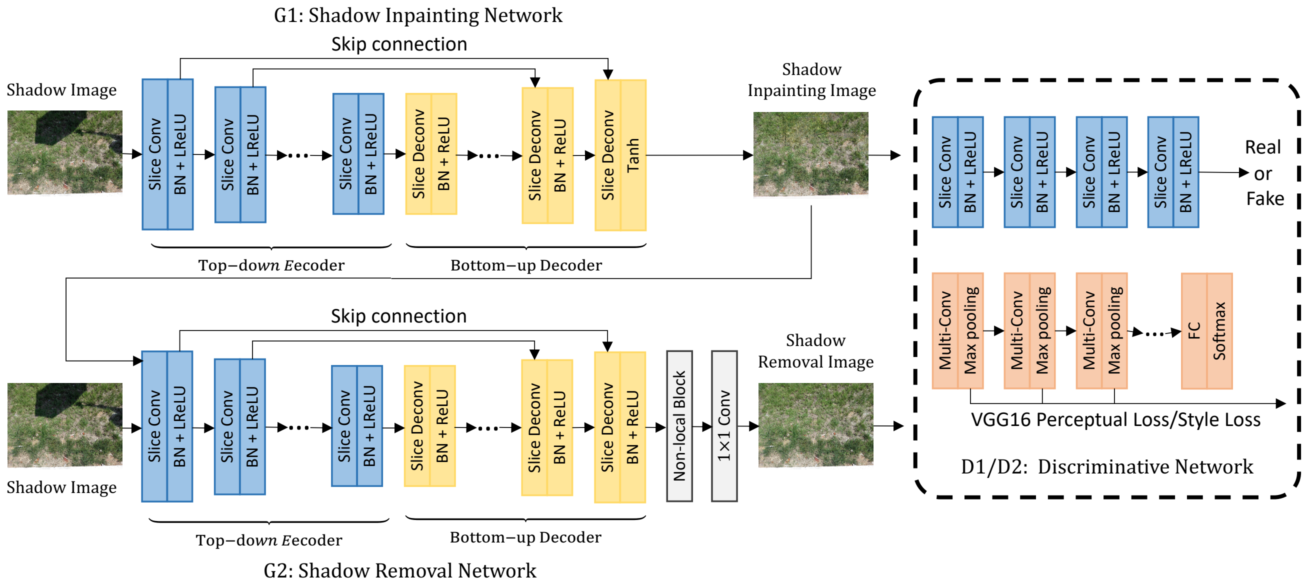

Figure 2: *The architecture of the proposed two-stage TBGANs for shadow inpainting and removal models. The shadow image is fed into the 1st-stage TBGAN in order to predict possible light intensity, color, and texture information for the shadow area and get a shadow inpainting image. The 2nd-stage TBGAN takes shadow inpainting image with high frequency information and the shadow image containing shadow area as input to recover a high-quality and photo-realistic shadow-removal image. Unlike the generator G1 of the 1st-stage TBGAN, the generator G2 of the 2nd-stage TGBAN incorporate an non-local block (see Figure 5) to explore both both local and global inter-dependencies of pixels. Both these two TGBANs share the same architecture of discriminator (D1/D2) with a two-branch network to incorporate adversarial loss, perceptual loss, and style loss.*

image with partially or incompletely missing regions and $\mathbf{x}_{sfree}$ as the ground-truth image to train the shadow inpainting network. The loss functions used to train the 1st-stage TBGAN include L1 loss with high-dimensional feature, perceptual loss, style loss and total variation loss, and a restored image of the output shadow area, as discussed in Section 3.4.

The intuition behind is that we want to actively identify the dark shadow areas as missing regions for recovery and take the inpaiting images as the preliminary shadow-removal results. That's why we propose to make full use of sufficient training data for image inpainting. We transfer the domain for image inpainting to the domain for shadow removal by fine-tuning the model that is trained on the Places365 dataset and the Irregular Mask dataset with the ISTD dataset. This treatment can effectively address the issue of lacking sufficient training data for training a shadow-removal model by incorporating sufficient outsider training data.

For images with hard-shadow areas in complex scenes, texture details may be lost due to weak illumination information. It is difficult to restore the corlor and texture details in shadow area by using the deep learning method only for shadow removal. Therefore, in this stage, we can apply the trained 1st-stage TBGAN to recover some texture details missing in the shadow regions. Meanwhile, the normal illumination in the non-shadow area can speculate and repair the low-frequency information such as light intensity and color type in the shadow area. As illustrated in Figure 4, our 1st-stage TBGAN is able to handle dark shadows and recover coarse shadow-removal images through shadow inpainting.

3.3. The 2nd-stage TBGAN for shadow removal

We take the 2nd-stage TBGAN for shadow removal which further recovers the shadow-free image from the coarse shadow-removal image at the 1st stage in a coarse-to-fine fashion. Note that the input for the network is the concatenation of the input image and inpainting image obtained from the 1st-stage TBGAN and we take the corresponding ground-truth shadow-free image for training. We argue that this treatment is reasonable because the inpainting image supplements the low frequency information, and the original shadow image contains high frequency information.

To make full use of both local and global spatial information for better recovery of shadow-free images, we introduce a non-local block to obtain the global correlation at the feature space to improve the overall shadow-removal effect. Inspired by Wang *et al.* [WGGH18], we design a non-local block to get a non-local feature map $\mathbf{y}$ to recovery a high-quality shadow-removal image from the feature map $\mathbf{x}$ as:

$$\mathbf{y}_i = \frac{1}{\mathcal{C}(\mathbf{x})} \sum_{\forall j} f\left(\mathbf{x}_i, \mathbf{x}_j\right) g\left(\mathbf{x}_j\right), \tag{1}$$

where $\mathcal{C}(\mathbf{x})$ is the normalization factor, and the response f for $\mathbf{x}_i$ and all $\mathbf{x}_j$ is computed to measure the correlation between each position pair $(\mathbf{x}_i, \mathbf{x}_j)$. And the function g computes a representation of the input signal at the position j. In this paper, we follow [BCM05, VSP*17] and define the function f as an embedded

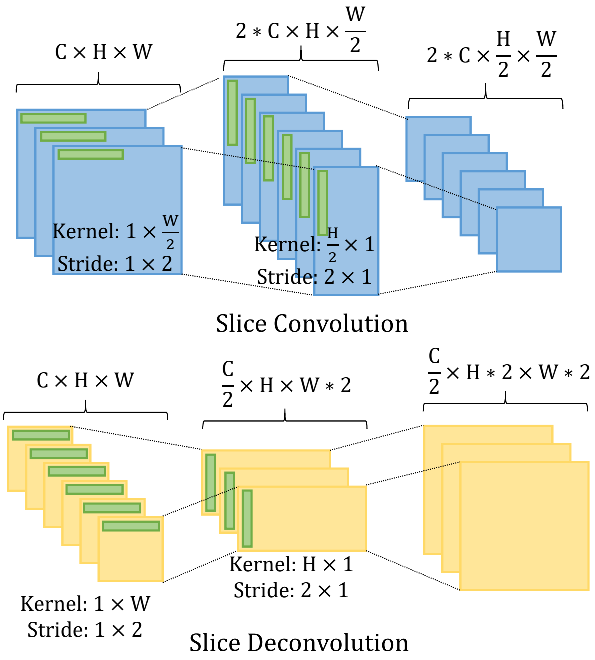

Figure 3: *The slice convolution and deconvolution are with long-range convolution size in horizontal and vertical directions separately, which is more effective to extract the long-range dependency information with less parameters.*

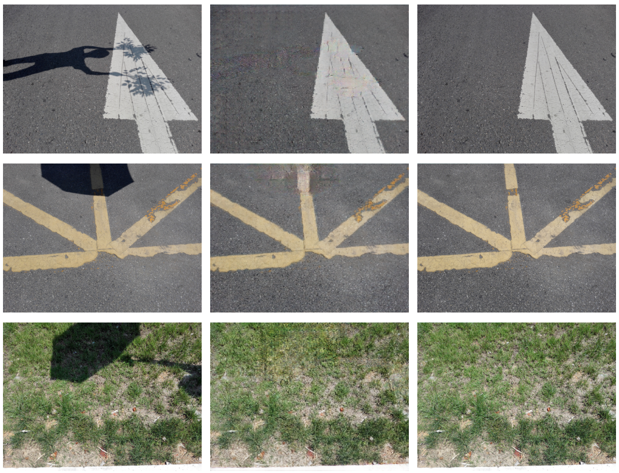

Figure 4: *Illustration of three dark shadow examples. From left to right are input dark shadow images, the results of shadow inpainting with our 1st-stage TBGAN, and the corresponding ground-truth shadow-free images, respectively.*

Gaussian function. Then we can rewrite Equation 1 as:

$$\mathbf{y}_i = softmax(\theta\left(\mathbf{x}_i\right)^T \phi\left(\mathbf{x}_j\right))g\left(\mathbf{x}_j\right), \qquad (2)$$

where θ, ϕ and g is implemented as 1×1 convolution in image space. For a given i, the $softmax$ computation along the demension j to replace $\frac{1}{C(\mathbf{x})}$.

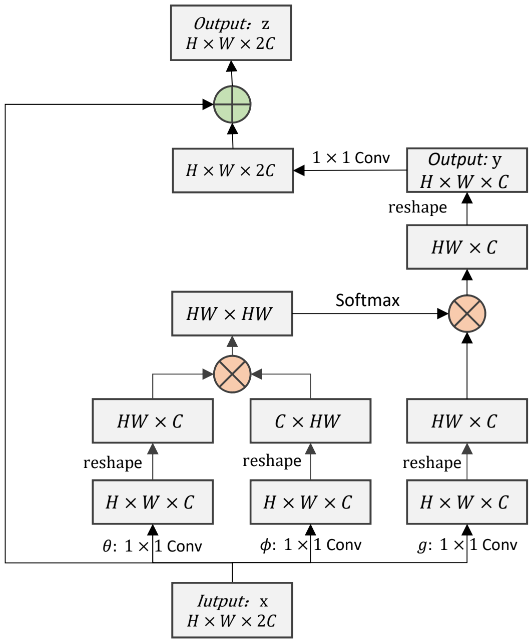

Figure 5: *The non-local block to explore both global and local spatial information for recovering shadow-free images. "⊗" and "⊕" represent the matrix multiply and the element-wise addition operations, respectively.*

As shown in Figure 5, we apply another 1×1 convolution on the final output $\mathbf{y}$ to get another feature map with the same size as $\mathbf{x}$ and then apply an element-wise addition operation to formulate the final feature map $\mathbf{z}$ to recovery the shadow-removal image with a 1×1 convolution.

We observe that our 2nd-stage TBGAN can further adjust the texture and illumination information in the shadow area to obtain more accurate shadow-removal results. The two-stage TBGAN can complement the complete frequency domain information of the shadow area for shadow removal. As we can see in Figure 6, for some bad shadow inpainting results, our 2nd-stage TBGAN can correct and recover better shadow-removal images.

3.4. Loss Functions

Our objective function combines the high-dimensional feature loss of the generative network with the loss function of the discriminative network to optimize the generative network, while the discriminative network is optimized using only the correct and wrong classification loss. Ultimately, the optimization function is a combination of high-dimensional feature loss of the generative network and the loss of GAN,

$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G,D) + \mathcal{L}_{total}, \qquad (3)$$

$$\begin{aligned}\mathcal{L}_{cGAN}(G,D) = {} & \mathbb{E}_{x,y\sim p_{data}(x,y)}[\log D(x,y)] \\ & + \mathbb{E}_{x\sim p_{data}(x)}[\log(1 - D(x,G(x)))],\end{aligned} \qquad (4)$$

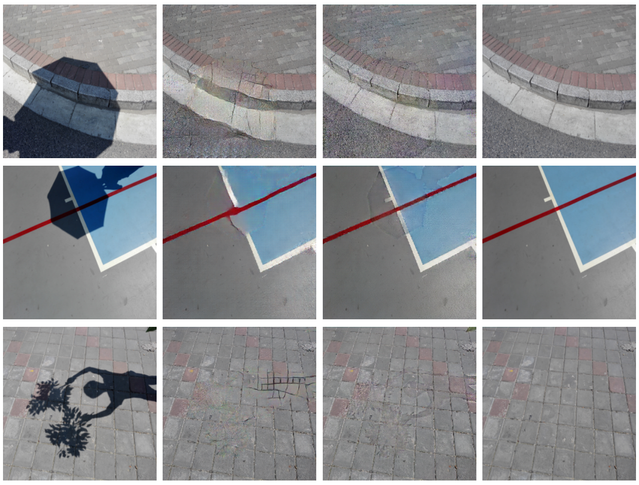

Figure 6: *Illustration of three bad shadow inpainting results. From left to right are input shadow images, the results of shadow inpainting with our 1st-stage TBGAN, the results of shadow removal with our 2nd-stage TBGAN, and the corresponding ground-truth shadow-free images, respectively.*

$$\mathcal{L}_{total} = \lambda\mathcal{L}_{L1} + \alpha\mathcal{L}_{perc} + \beta\mathcal{L}_{style} + \gamma\mathcal{L}_{tv}, \tag{5}$$

where we set $\lambda = 20$, $\alpha = 0.05$, $\beta = 100$, $\gamma = 0.1$ in the paper and each loss term is defined in the following description. In the 2nd-stage GAN, we adjust the parameter to $\lambda = 50$, $\alpha = 0.01$, $\beta = 50$, $\gamma = 0.01$.

We define L1 loss $\mathcal{L}_{L1}$ as

$$\mathcal{L}_{L1} = \frac{1}{C \times H \times W} \left\| I_{out} - I_{gt} \right\|_1, \tag{6}$$

where I_{in} is the input image, I_{out} is the output of the generative network, I_{gt} is the ground truth image, and the shape of the output tensor is $C \times W \times H$.

Perceptual loss $\mathcal{L}_{perc}$ can be defined as

$$\mathcal{L}_{perc} = \frac{1}{C_p \times H_p \times W_p} \sum_{h=1}^{H_p} \sum_{w=1}^{w_p} \left\| \Psi_p^{I_{out}} - \Psi_p^{I_{gt}} \right\|_1, \tag{7}$$

where $\Psi_p^{I_{out}}$ and $\Psi_p^{I_{gt}}$ are feature maps of the generated image I_{out} and the ground truth image I_{gt} respectively output by p-th layer in the pre-trained Convolutional Neural Network. $C_p \times H_p \times W_p$ is the number of output feature map shape, $\Psi_p^{I_{out}}$ and $\Psi_p^{I_{gt}}$ at the p-th layer in the CNN.

The style loss $\mathcal{L}_{style}$ is calculated based on two Gram matrices measuring the correlation of features as covariance matrix of the feature maps

$$\mathcal{L}_{style} = \frac{1}{C_p \times H_p \times W_p} \sum_{h=1}^{H_p} \sum_{w=1}^{W_p} \left\| \left(\Psi_p^{I_{out}}\right)^T \left(\Psi_p^{I_{out}}\right) - \left(\Psi_p^{I_{gt}}\right)^T \left(\Psi_p^{I_{gt}}\right) \right\|_1 \tag{8}$$

where we use the output feature maps of maxpooling1, maxpooling2, maxpooling3 of the pre-trained VGG-16 model to calculate the perceptual loss and style loss.

The total variation loss $\mathcal{L}_{tv}$ (Mahendran et al. [MV15]) is a smooth penalty term for the output image of the generative network and defined as

$$\begin{aligned} \mathcal{L}_{tv} = & \frac{1}{C \times H \times W} \sum_{(i,j) \in R} \left\| I_{out}^{i,j+1} - I_{out}^{i,j} \right\|_1 \\ & + \frac{1}{C \times H \times W} \sum_{(i,j) \in R} \left\| I_{out}^{i+1,j} - I_{out}^{i,j} \right\|_1 \end{aligned} \tag{9}$$

3.5. Implementation details

The input image size for our two-stage TBGANs is 256 × 256. The network is implemented in PyTorch and trained using the Adam optimizer [KB14] with parameters of beta1=0.9 and beta2=0.999. Note that the weights for our model are initialized from a Gaussian distribution with mean 0 and standard deviation 0.02. The learning rate for both the 1st-stage TBGAN and the 2nd-stage TBGAN is set as 10^{-4} initially and then gradually decayed linearly every 20 epochs to 10^{-6}. To make it simple, we use the same learning rate for both generators and the corresponding discriminators. We train 250 epochs for these two TBGANs.

Generators. The input of the inpainting generator G1 is a 3-channel image, and the input of the removal generator G2 is two 3-channel images. The size of the top-down slice convolution kernel is initially 128, halved each time until the slice convolution of minimum length 2. The size of the slice convolution kernel of the corresponding bottom-up module is doubled each time, with a maximum length of 128. The number of first convolution output channels is 64. The number of channels is doubled or halved after each top-down or bottom-up module. The maximum number of channels is 512. The slice convolution is no longer halved when it is at least 2. Each layer of convolution performs the following operations: convolution, batch normalization, and activation function. The activation function used in the top-down encoder for both G1 and G2 is Leaky ReLU [HZRS15] with slope 0.2. Except the final output of the generative network G1 followed by a tanh function, the activation function used in the bottom-up decoder is ReLU. Skip-connection is performed between the encoder and the decoder with the same size as the output tensor. The final output of the generative network G1 is followed by a tanh function. The final output is a 3-channel image, this changes the number of channels in the generators are as follows:

For G1, encoder: 3 or 6 -> 64 -> 128 -> 256 -> 512 -> 512 -> 512 -> 512 -> 512, decoder: 512+512 -> 512+512 -> 512+512 -> 512+512 -> 256+256 -> 128+128 -> 64+64 -> 3.

For G2, encoder: 3 or 6 -> 64 -> 128 -> 256 -> 512 -> 512 -> 512 -> 512 -> 512, decoder: 512+512 -> 512+512 -> 512+512 -> 512+512 -> 256+256 -> 128+128 -> 64+64 -> 6 -> 3.

Discriminators. Note that both the two-stage TBGANs share the same two-branch discriminative network architecture. The input for the two-branch discriminative network is a 3-channel image with the size of 256 × 256 for both the inpainting image from G1 and the shadow-removal image from G2. The first branch is a fully convolutional network with 4 slice convolutions and each slice convolution has a batch normalization and a Leaky ReLU with slope 0.2. The final layer of the first branch is a sigmoid function to produce a probability to judge as a real or fake inpainting or shadow-free image. The final output of this branch is 16×16. Note that the number of channels in the first branch are 3 -> 64 -> 128 -> 256 -> 512 -> 1. The 2nd branch is VGG16 which we use to derive perceptual loss and style loss to improve the performance of the generated network [JAFF16].

4. Experiments

In this section, we perform experiments on the ISTD dataset [WLY18] to verify the effectiveness of our proposed two-stage TBGANs. The ISTD dataset contains 1870 triples of shadow images, shadow masks, and shadow-free images. Such a dataset includes 135 different simulated shadow environments and the scenes are more diverse. We compare our proposed two-stage TBGANs against the current state-of-the-art methods including the traditional methods and the deep learning methods both qualitatively and quantitatively. To measure the shadow removal performance, we use the metric of root mean square error (RMSE) calculated in Lab space between the shadow-remova image and the corresponding ground-truth shadow-free images. To qualitatively and quantitatively measure the multifaceted factors of experimental results, we compare the RMSE on the shadow regions, non-shadow regions, and the full image, respectively.

4.1. Comparisons with the state-of-the-art methods

We compare our proposed two-stage TBGANs mehtod with five state-of-art methods including tradition methods, *i.e.*, Gong [GC14] and Guo [GDH12], the latest deep learning based methods, *i.e.*, DeshadowNet [QTH*17], ST-CGAN [WLY18], and DSC [HFZ*18]. To make the fair comparison, we use the same 1330 training triplets as DSC [HFZ*18] without any outsider dataset to train our propsoed two-stage TBGANs method and evaluate the shadow-removal performance on the same 540 testing triplets. And we have transferred the two frameworks of image inpainting Global/Local-GAN [ISSI17] and image translation Pix2Pix-HD [WLZ*18] for shadow removal. As mentioned in Seciton 3.2, we also incorporate the outsider inpainting dataset obtained from Places365 dataset with the Irregular Mask dataset to train an initial 1st-stage TBGAN model and then finetune it with the ISTD dataset. We denote the finetune verison of our proposed method as “two-stage TBGANs+finetune". The results are summarized in Table 1 and Figure 7. We also conduct a user study to further evaluate the comparison results.

Table 1: *Comparison of shadow removal results of different methods on the ISTD dataset in term of RMSE. The metric of RMSE directly measures the per-pixel error between the shadow removal images and the ground truth shadow-free images, and the smaller value of RMSE is better.*

	Shadow	Non-shadow	All
Guo [GDH12]	18.95	7.46	9.3
Gong [GC14]	14.98	7.29	8.53
Global/Local-GAN [ISSI17]	13.46	7.67	8.82
Pix2Pix-HD [WLZ*18]	10.63	6.73	7.37
Deshadow [QTH*17]	12.76	7.19	7.83
ST-CGAN [WLY18]	10.33	6.93	7.47
DSC [HFZ*18]	**9.22**	6.39	6.67
Two-stage TBGANs	10.14	5.91	6.70
Two-stage TBGANs+finetune	9.83	**5.58**	**6.39**

Quantitative comparisons. Table 1 shows the results of the quantitative comparisons on the ISTD dataset. As we can see, (1) all deep learning methods perform better than the traditional methods; (2) among the deep learning methods, our proposed two-stage TBGANs method achieves the smallest values of RMSE, 5.91 on non-shadow area and 6.70 on the entire images; (3) The RMSE of our proposed two-stage TBGANs on the shadow area is also comparable to that of DSC; (4) Through the transfer learning, both Global/Local-GAN and Pix2Pix-HD are able to be extended for shadow removal, but their performances are still worse than our proposed two-stage TBGANs; and (5) with the initial 1st-stage TBGAN model trained on the outsider inpainting dataset, our “two-stage TBGANs+finetue" achieves RMSE improvement from 10.14 to 9.83 on shadow areas, from 5.91 to 5.58 on non-shadow areas, and from 6.70 to 6.39 on the entire images, which experimentally proves our assumption that we can view shadow removal as a special case of image inpainting. All these observations strongly demonstrate the efficacy of the proposed two-stage TBGANs method for shadow inpainting and removal.

Qualitative comparisons. Figure 7 illustrates the results of the qualitative comparisons on the ISTD dataset. Apparently, compared with the state-of-the-art methods, our proposed two-stage TBGANs method is able to recover a higher quality and more photo-realistic shadow-free images in which the shadow area are more consistent with the non-shadow area in term of color and texture, and pretty closer to the ground-truth shadow-free images.

We observe that dark shadows exit in scenes with bright colors and excessive texture changes. For this kind of dark shadow images, our 1st-stage TBGAN is able to add the informative color and texture information as inpainting in shadow area, and then the 2nd-stage TBGAN continues to correct the color and texture information smoothly to reach high quality and photo-realistic images.

User study. We conduct a user-study by asking a total of 37 people to participate in a survey and check whether the shadow-removal images generated by our proposed method and DSC are shadow-free images without artifacts or not. We randomly sample one tenth of the images in the test set and ensure that at least one of the different scenes is included. Counting all the votes, the survey received a total of 1,155 valid votes. The survey shows 60.95% of the shadow-removal images generated by our proposed method are chosen as shadow-free images without artifacts, while only 39.05% of the shadow-removal images by DSC are chosen.

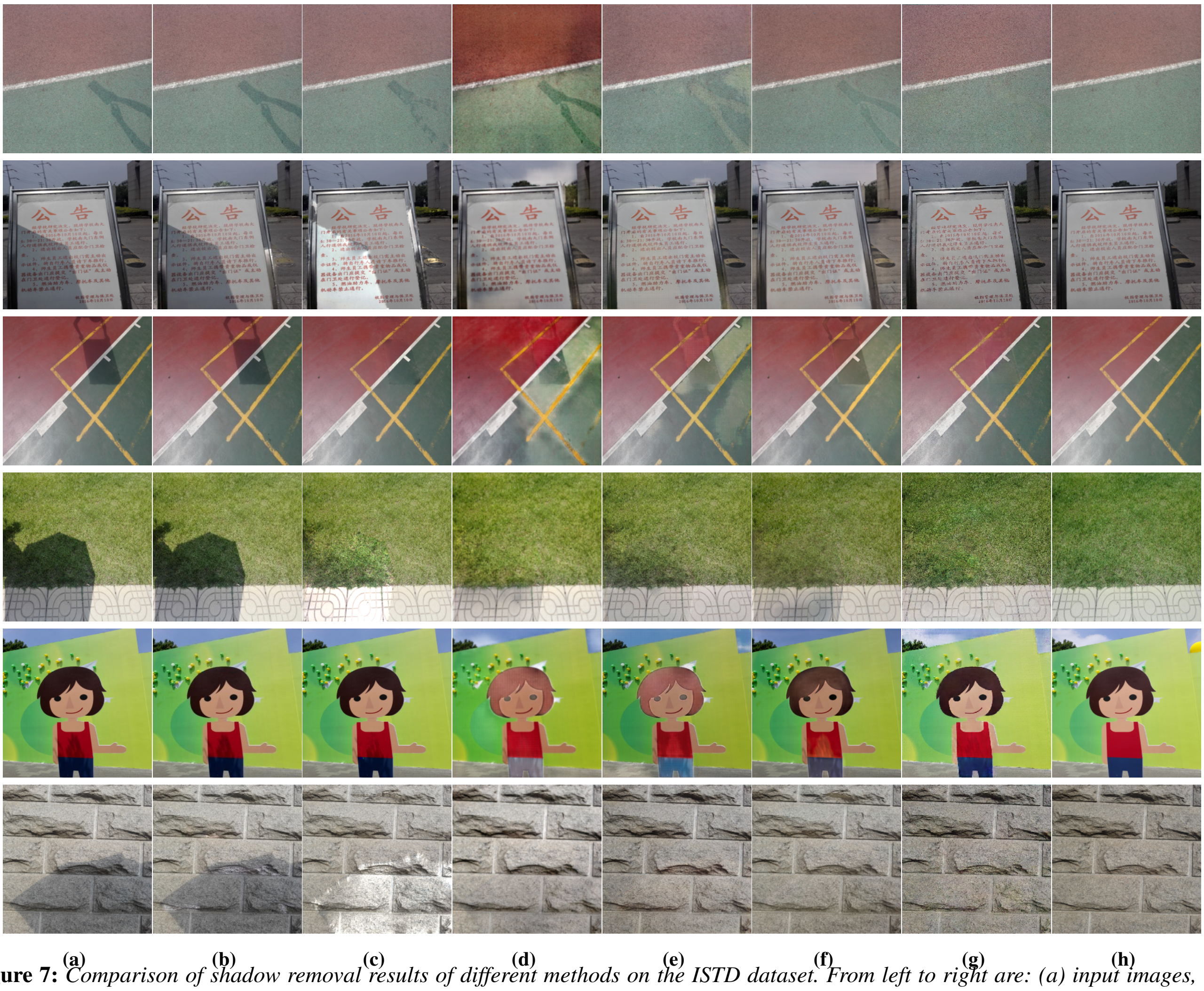

Figure 7: *Comparison of shadow removal results of different methods on the ISTD dataset. From left to right are: (a) input images, (b) Guo* et al. *[GDH12]'s results, (c) Gong* et al. *[GC14]'s results, (d) DeshadowNet [QTH*17]'s results, (e) ST-CGAN [WLY18]'s results, (f) DSC [HFZ*18]'s results, (g) our proposed tow-stage TBGANs' results, and (h) the corresponding ground-truth shadow-free images, respectively.*

4.2. Ablation Study

To further evaluate the network design, we conduct the ablation study to evaulate some components in our two-stage TBGANs. We denote the top-down encoder and bottom-up decoder as TDBU, and indicate the four variants of our proposed method as follows:

- *w/o TDBU & Inpainting*: no 1st-stage TBGAN and replace G2 in the 2nd-stage TBGAN in Figure 2 with a U-Net [RFB15].
- *w/o TDBU*: replace G1 and G2 in Figure 2 with a U-Net.
- *w/o Non-local block*: remove the non-local block from the 2nd-stage TBGAN.
- *1st-stage TBGAN*: only use the 1st-stage TBGAN, without apply the 2nd-stage TBGAN for further refinement.

For the fair comparison, we use the same 1330 training triplets without any outsider dataset to train the above four methods and evaluate the shadow-removal performance on the same 540 testing triplets. The results are summarized in Table 2 and Figure 8.

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

	Shadow	Non-shadow	All
w/o TDBU&Inpainting	18.72	16.33	16.77
w/o TDBU	17.72	16.16	16.51
w/o Non-local block	10.33	5.89	6.79
1st-stage TBGAN	12.37	**5.58**	7.44
Two-stage TBGANs	**10.14**	5.91	**6.70**

Quantitative comparisons. Table 2 shows the quantitative ablation analysis for the four variants. From this table we can observe:

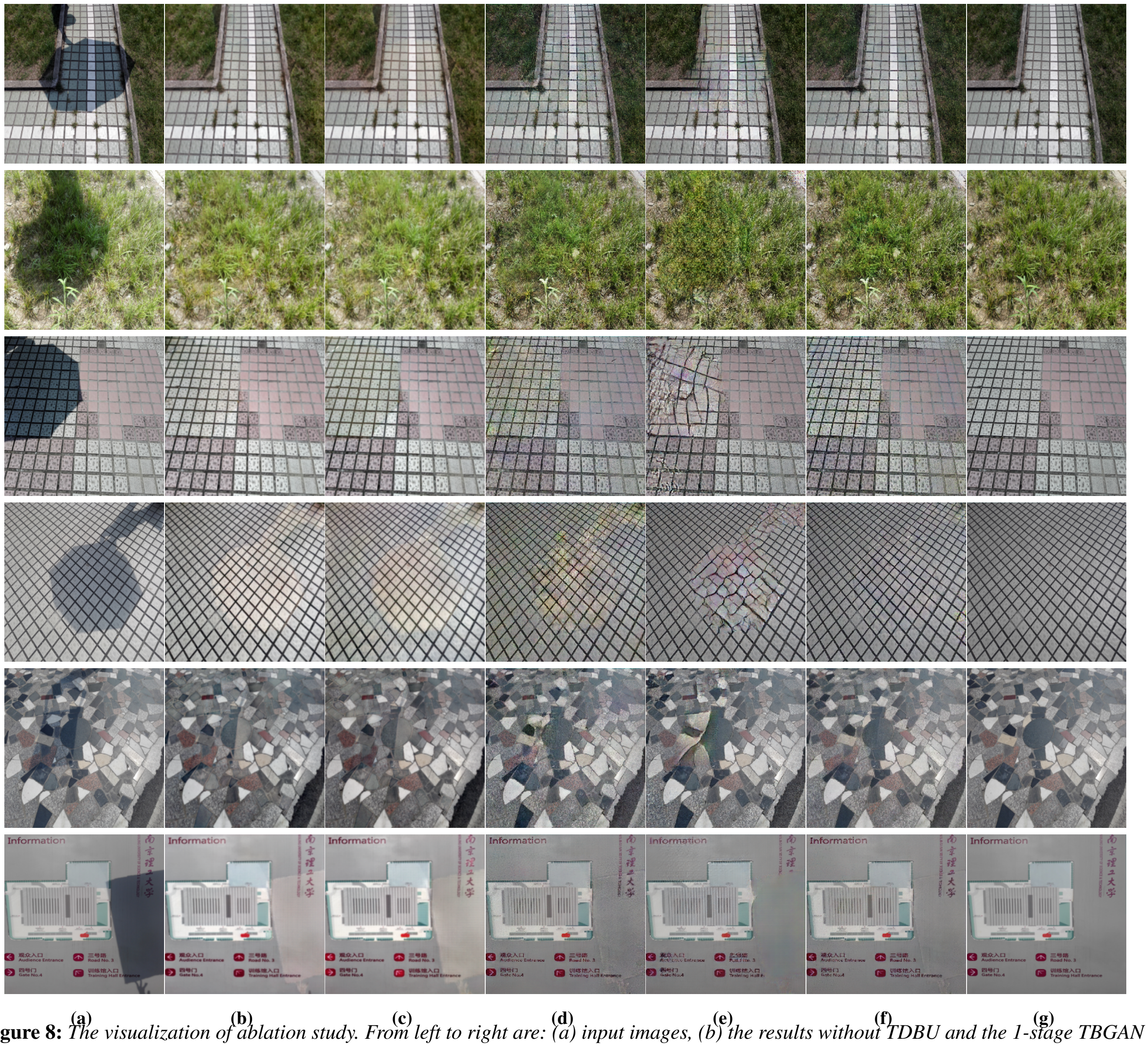

Figure 8: *The visualization of ablation study. From left to right are: (a) input images, (b) the results without TDBU and the 1-stage TBGAN for shadow inpainting, (c) the results without TDBU, (d) the results without Non-local block, (e) the results of the 1st-stage TBGAN, (f) the results of our proposed two-stage TBGANs, and (g) the corresponding ground-truth shadow-free images, respectively.*

(1) the design of TDBU structure with slice convolutions is able to enhance the receptive field of convolution in term of feature extraction for improving the performance of shadow removal; (2) with the non-local block, the quality of shadow-removal images can be further improved, which suggests that the non-local spatial information is beneficial to recover a higher quality shadow-free images; and (3) and although the 1st-stage TBGAN has achieved a good performance, our 2nd-stage TBGAN with non-local block is still able to further improve its performance. It is worth noting that the 1st-stage TBGAN only focues on inpainting the shadow area, which can explain why its recovery performance on the non-shadow area is better than both “w/o Non-local block" and our proposed two-stage TBGANs. All these observations demonstrate the reasonable design of our proposed two-stage TBGANs approach.

Qualitative comparisons. Figure 8 provides the visualization of the shadow-removal results on the ISTD dataset. As we can see, without TDBU modules and the non-local blocks, the recovered shadow-removal images look unrealistic and cause color and texture inconsistency between shadow and non-shadow areas. The 1st-stage TBGAN is able to recover coarse results without affecting the non-shadow area too much, which guarantees a low RMSE value in the non-shadow area, and then our 2nd-stage TBGAN with non-local block is able to correct the coarse shadow area and ensure color and texture consistency between shadow and non-shadow areas. Compared with these four variants, our proposed two-stage TBGANs methods is able to generate a high-quality and photo-realistic shadow-removal image which is much closer to the corresponding ground-truth shadow-free image.

To better explain why the non-local block works for shadow removal, we visualize the input feature map x ($H \times W \times 2C$), the residual feature map ($H \times W \times 2C$) obtained after applying a 1×1 Convolution on the output y ($H \times W \times C$), and the output feature map z ($H \times W \times 2C$) as shown in Figure 5. For the purpose of visualization, we apply a channel-wise average pooling on each feature map, as illustrated in Figure 9. We can observe that the residual feature map obtained from the non-local block is complementary to the input feature map, which leads to a much smoother feature map between the shadow region and the non-shadow region. In Figure 9, we also visualize the effect of the affinity matrix ($HW \times HW$) by reshaping to a tensor $H \times W \times HW$, extracting the channel-wise top $K = 5$ to get a tensor $H \times W \times K$, and then apply a channel-wise average pooling to get a matrix $H \times W$ in which the locations with high values indicate they are very similar to some other locations. Such observations can well explain why the non-local block is able to make the shadow regions and the non-shadow regions more consistent, which directly leads to better shadow-removal results.

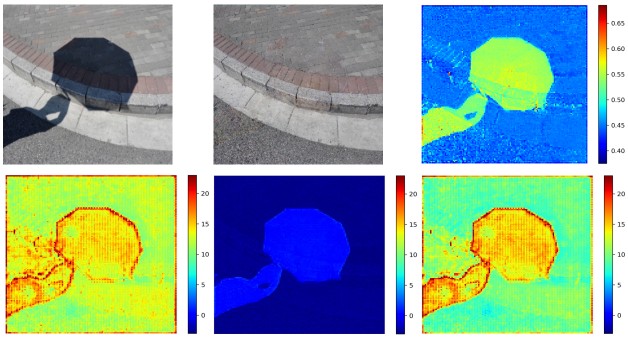

Figure 9: *The visualization of effect with the non-local block. On the top row, from left to right are the input shadow image, the shadow-removal image, and the effect of $HW \times HW$, respectively. On the bottom row, from left to right are the average feature map of input x, the average residual feature map, and the average feature map of output z in Figure 5, respectively.*

4.3. Discussion

To better explore the potential of our proposed two-stage TBGANs, we also visualize the shadow detection masks, extend the current approach for video shadow removal, as well as discuss the failure cases and its limitations.

Visualization of detection results. Although we focus on shadow removal rather than detection, we also can get the shadow detection mask by subtracting an input shadow image from the corresponding shadow-removal image. From the visualization of shadow detection results in Figure 10, we can see our proposed two-stage TBGANs is able to achieve reasonable detection results.

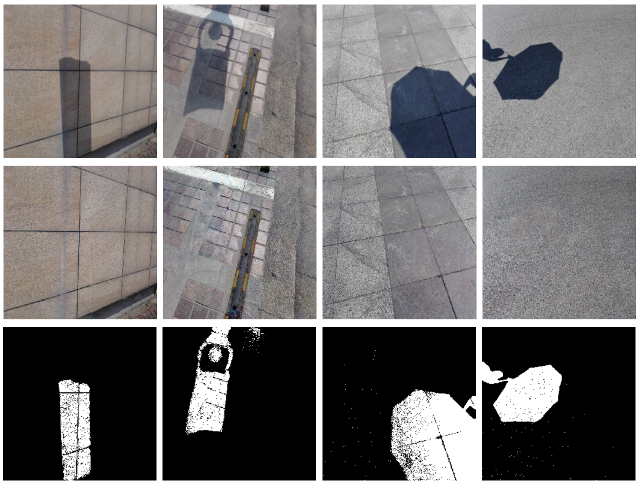

Figure 10: *The visualization of shadow detection masks (bottom) by subtracting the input shadow images (top) from the corresponding shadow-removal images (middle).*

Extension to video. We also apply the proposed two-stage TBGANs to handle shadow videos by processing each frame in order. As shown in Figure 12, it is suitable to insert videos in this paper. Instead, we show the shadow-removal results for 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.

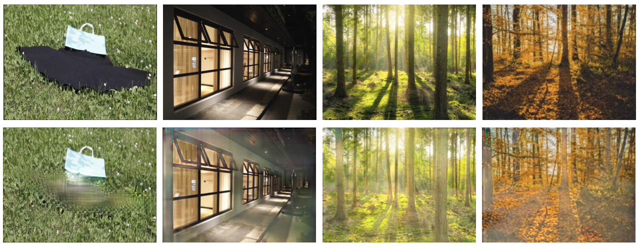

Figure 11: *The visualization of some failed examples. From top to down are input images and the shadow-removal results of our proposed two-stage TBGANs, respectively.*

Failure cases and limitation. To clarify, our proposed method is able to handle both soft and hard shadow in the current available shadow datasets. For shadow images with complex boundaries and complicated shapes in a forest environment, our proposed method may fail as most of the existing data-driving shadow removal methods. We observe that our proposed method will fail when the shadow area occupies a large proportion because the contextual information provided in the non-shadow area is limited. It also fails when there is inconsistency between shadow region and non-shadow region in complicated scene. The failure examples are visualized in Figure 11, from which we can observe the black skirt is recognized as shadow and the shadow in the other three complicated scenes cannot be removed completely.

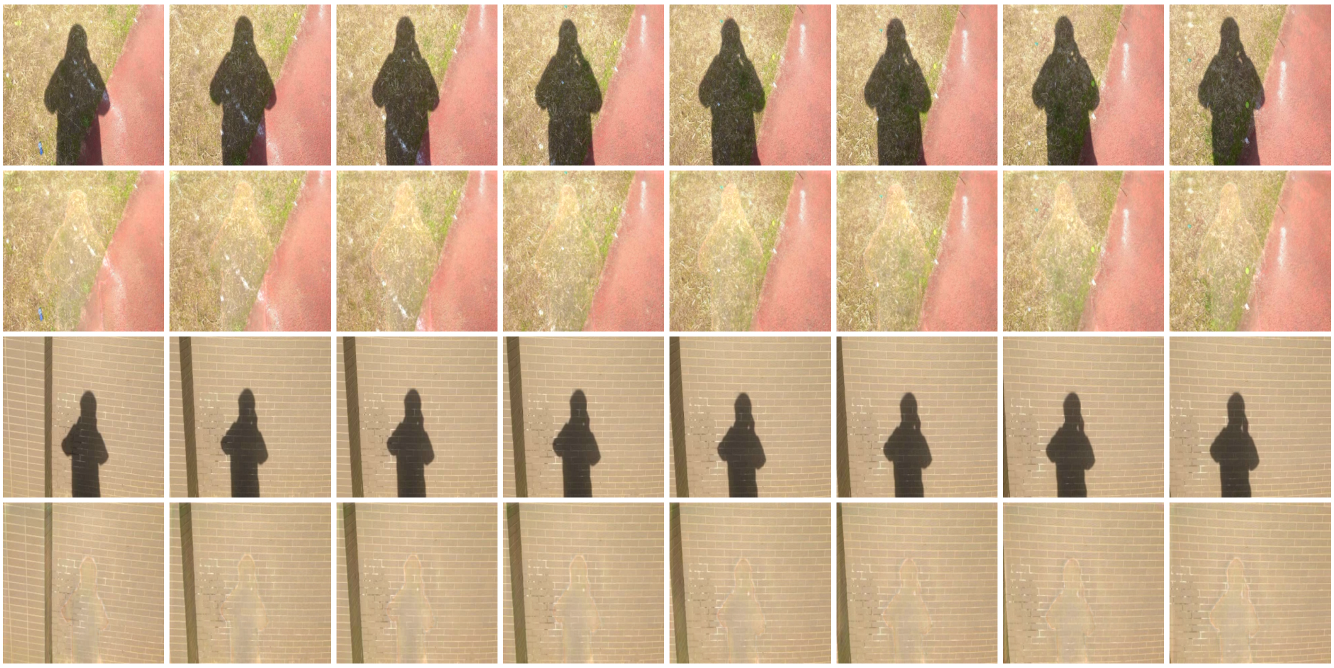

Figure 12: *The visualization of shadow-removal results in two videos. From top to bottom are frames of the first input video, the shadow-removal result for the first video, frames of the second input video, and the shadow-removal result for the second video, respectively. Note the frames are extracted every 100 milliseconds from these two videos.*

5. Conclusion

In this paper, we propose a two-sage top-down and bottom-up Generative Adversarial Networks (TBGANs) for shadow inpainting and removal using a novel top-down encoder and a bottom-up decoder with slice convolutions. The slice convolutions can effectively extract and restore the long-range spatial information. Shadow regions are first inpainted to get a coarse shadow-removal results by the 1st-stage TBGAN at the first stage and then the coarse shadow-removal results are further refined by the 2nd-stage TBGAN with a non-local block to achieve a better quality and photo-realistic shadow-removal results. With such a coarse-to-fine fashion, the overall effect of shadow removal is greatly improved, and the effect of color retention in non-shaded areas is significant. By comparing with a variety of mainstream shadow removal methods, it is found that our method is superior to the state-of-the-art methods.

In future, we plan to further explore top-down encoder and bottom-up decoder with slice convolutions and apply them to solve more real-world applications in the field of computer vision. We also plan to explore a more efficient shadow removal approach to handle videos [ZZLX17]. To obtain satisfying Illumination decomposition [ZYL*17] and image recoloring [ZXST17] results, effective shadow detection and removal is critical. Thus, in the future, we will exploit our method in these research directions.

6. Acknowledgment

This work was partly supported by the National Key Research and Development Program of China (2017YFB1002600), the NSFC (No. 61672390), Wuhan Science and Technology Plan Project (No.2017010201010109), and Key Technological Innovation Projects of Hubei Province (2018AAA062). Chunxia Xiao is the corresponding author of this paper.

References

[AHO10] ARBEL E., HEL-OR H.: Shadow removal using intensity surfaces and texture anchor points. *IEEE transactions on pattern analysis and machine intelligence 33*, 6 (2010), 1202–1216.

[BCM05] BUADES A., COLL B., MOREL J.-M.: A non-local algorithm for image denoising. In *CVPR* (2005), vol. 2, IEEE, pp. 60–65.

[CGC*03] CHUANG Y.-Y., GOLDMAN D. B., CURLESS B., SALESIN D. H., SZELISKI R.: Shadow matting and compositing. In *ACM Transactions on Graphics (TOG)* (2003), vol. 22, ACM, pp. 494–500.

[DLTLM19] DUKLER Y., LI W., TONG LIN A., MONTÚFAR G.: Wasserstein of wasserstein loss for learning generative models.

[DLZX19] DING B., LONG C., ZHANG L., XIAO C.: Argan: Attentive recurrent generative adversarial network for shadow detection and removal. In *ICCV* (2019).

[DU17] DEMIR U., UNAL G.: Deep stacked networks with residual polishing for image inpainting. *arXiv preprint arXiv:1801.00289* (2017).

[FHLD05] FINLAYSON G. D., HORDLEY S. D., LU C., DREW M. S.: On the removal of shadows from images. *IEEE transactions on pattern analysis and machine intelligence 28*, 1 (2005), 59–68.

[GC14] GONG H., COSKER D.: Interactive shadow removal and ground truth for variable scene categories. In *BMVC* (2014).

[GDH11] GUO R., DAI Q., HOIEM D.: Single-image shadow detection and removal using paired regions. In *CVPR 2011* (2011), IEEE, pp. 2033–2040.

[GDH12] GUO R., DAI Q., HOIEM D.: Paired regions for shadow detection and removal. *IEEE transactions on pattern analysis and machine intelligence 35*, 12 (2012), 2956–2967.

[GTB15] GRYKA M., TERRY M., BROSTOW G. J.: Learning to remove soft shadows. *ACM Transactions on Graphics (TOG) 34*, 5 (2015), 153.

[HFZ*18] HU X., FU C.-W., ZHU L., QIN J., HENG P.-A.: Direction-aware spatial context features for shadow detection and removal. *arXiv preprint arXiv:1805.04635* (2018).

[HLYG13] HUA G., LONG C., YANG M., GAO Y.: Collaborative active learning of a kernel machine ensemble for recognition. In *ICCV* (2013).

[HLYG18] HUA G., LONG C., YANG M., GAO Y.: Collaborative active visual recognition from crowds: A distributed ensemble approach. *IEEE transactions on pattern analysis and machine intelligence 40*, 3 (2018), 582–594.

[HZRS15] HE K., ZHANG X., REN S., SUN J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *ICCV* (2015), pp. 1026–1034.

[ISSI17] IIZUKA S., SIMO-SERRA E., ISHIKAWA H.: Globally and locally consistent image completion. *ACM Transactions on Graphics (ToG) 36*, 4 (2017), 107.

[JAFF16] JOHNSON J., ALAHI A., FEI-FEI L.: Perceptual losses for real-time style transfer and super-resolution. In *ECCV* (2016), Springer, pp. 694–711.

[KB14] KINGMA D. P., BA J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[KBST15] KHAN S. H., BENNAMOUN M., SOHEL F., TOGNERI R.: Automatic shadow detection and removal from a single image. *IEEE transactions on pattern analysis and machine intelligence 38*, 3 (2015), 431–446.

[KF19] KATHAROPOULOS A., FLEURET F.: Processing megapixel images with deep attention-sampling models. *arXiv preprint arXiv:1905.03711* (2019).

[LG08] LIU F., GLEICHER M.: Texture-consistent shadow removal. In *ECCV* (2008), Springer, pp. 437–450.

[LH15] LONG C., HUA G.: Multi-class multi-annotator active learning with robust gaussian process for visual recognition. In *ICCV* (2015).

[LH17] LONG C., HUA G.: Correlational gaussian processes for cross-domain visual recognition. In *CVPR* (2017).

[LKZF19] LIAO Z., KARSCH K., ZHANG H., FORSYTH D.: An approximate shading model with detail decomposition for object relighting. *International Journal of Computer Vision* (2019), 1–16.

[LLK19] LEE J., LEE I., KANG J.: Self-attention graph pooling. *arXiv preprint arXiv:1904.08082* (2019).

[LRS*18] LIU G., REDA F. A., SHIH K. J., WANG T.-C., TAO A., CATANZARO B.: Image inpainting for irregular holes using partial convolutions. In *Proceedings of the ECCV (ECCV)* (2018), pp. 85–100.

[LWH*14] LONG C., WANG X., HUA G., YANG M., LIN Y.: Accurate object detection with location relaxation and regionlets re-localization. In *ACCV* (2014).

[MTC07] MOHAN A., TUMBLIN J., CHOUDHURY P.: Editing soft shadows in a digital photograph. *IEEE Computer Graphics and Applications 27*, 2 (2007), 23–31.

[MV15] MAHENDRAN A., VEDALDI A.: Understanding deep image representations by inverting them. In *CVPR* (2015), pp. 5188–5196.

[NVT*17] NGUYEN V., VICENTE Y., TOMAS F., ZHAO M., HOAI M., SAMARAS D.: Shadow detection with conditional generative adversarial networks. In *ICCV* (2017), pp. 4510–4518.

[QJL*19] QI L., JIANG L., LIU S., SHEN X., JIA J.: Amodal instance segmentation with kins dataset. In *CVPR* (2019), pp. 3014–3023.

[QTH*17] QU L., TIAN J., HE S., TANG Y., LAU R. W.: Deshadownet: A multi-context embedding deep network for shadow removal. In *CVPR* (2017), pp. 4067–4075.

[RAGS01] REINHARD E., ADHIKHMIN M., GOOCH B., SHIRLEY P.: Color transfer between images. *IEEE Computer graphics and applications 21*, 5 (2001), 34–41.

[RFB15] RONNEBERGER O., FISCHER P., BROX T.: U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (2015), Springer, pp. 234–241.

[SL08] SHOR Y., LISCHINSKI D.: The shadow meets the mask: Pyramid-based shadow removal. In *Computer Graphics Forum* (2008), vol. 27, Wiley Online Library, pp. 577–586.

[SVI*16] SZEGEDY C., VANHOUCKE V., IOFFE S., SHLENS J., WOJNA Z.: Rethinking the inception architecture for computer vision. In *CVPR* (2016), pp. 2818–2826.

[TM19] TAFFAR M., MIGUET S.: Local appearance modeling for objects class recognition. *Pattern Analysis and Applications 22*, 2 (2019), 439–455.

[VHS17] VICENTE T. F. Y., HOAI M., SAMARAS D.: Leave-one-out kernel optimization for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence 40*, 3 (2017), 682–695.

[VSP*17] VASWANI A., SHAZEER N., PARMAR N., USZKOREIT J., JONES L., GOMEZ A. N., KAISER Ł., POLOSUKHIN I.: Attention is all you need. In *NeurIPS* (2017), pp. 5998–6008.

[WGGH18] WANG X., GIRSHICK R., GUPTA A., HE K.: Non-local neural networks. In *CVPR* (2018), pp. 7794–7803.

[WLY18] WANG J., LI X., YANG J.: Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *CVPR* (2018), pp. 1788–1797.

[WLZ*18] WANG T.-C., LIU M.-Y., ZHU J.-Y., TAO A., KAUTZ J., CATANZARO B.: High-resolution image synthesis and semantic manipulation with conditional gans. In *CVPR* (2018), pp. 8798–8807.

[WTBS07] WU T.-P., TANG C.-K., BROWN M. S., SHUM H.-Y.: Natural shadow matting. *ACM Transactions on Graphics (TOG) 26*, 2 (2007), 8.

[XSXM13] XIAO C., SHE R., XIAO D., MA K.-L.: Fast shadow removal using adaptive multi-scale illumination transfer. In *Computer Graphics Forum* (2013), vol. 32, Wiley Online Library, pp. 207–218.

[XTT14] XIAO Y., TSOUGENIS E., TANG C.-K.: Shadow removal from single rgb-d images. In *CVPR* (2014), pp. 3011–3018.

[XXZC13] XIAO C., XIAO D., ZHANG L., CHEN L.: Efficient shadow removal using subregion matching illumination transfer. In *Computer Graphics Forum* (2013), vol. 32, Wiley Online Library, pp. 421–430.

[ZLK*18] ZHOU B., LAPEDRIZA A., KHOSLA A., OLIVA A., TORRALBA A.: Places: A 10 million image database for scene recognition. *IEEE transactions on pattern analysis and machine intelligence 40*, 6 (2018), 1452–1464.

[ZXST17] ZHANG Q., XIAO C., SUN H., TANG F.: Palette-based image recoloring using color decomposition optimization. *IEEE Transactions on Image Processing 26*, 4 (2017), 1952–1964.

[ZYL*17] ZHANG L., YAN Q., LIU Z., ZOU H., XIAO C.: Illumination decomposition for photograph with multiple light sources. *IEEE Transactions on Image Processing 26*, 9 (2017), 4114–4127.

[ZZLX17] ZHANG L., ZHU Y., LIAO B., XIAO C.: Video shadow removal using spatio-temporal illumination transfer. In *Computer Graphics Forum* (2017), vol. 36, Wiley Online Library, pp. 125–134.

[ZZX15] ZHANG L., ZHANG Q., XIAO C.: Shadow remover: Image shadow removal based on illumination recovering optimization. *IEEE Transactions on Image Processing 24*, 11 (2015), 4623–4636.