Pacific Graphics 2020
E. Eisemann, A. Jacobson, and F.-L Zhang
(Guest Editors)

Volume 39 (2020), Number 7

CLA-GAN: A Context and Lightness Aware Generative Adversarial Network for Shadow Removal

Ling Zhang[1], Chengjiang Long[2†], Qingan Yan[2], Xiaolong Zhang[1], Chunxia Xiao[3†‡]

[1]Hubei Key Laboratory of Intelligent Information Processing and Realtime Industrial System,
School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, China
[2] JD Digits, Mountain View, CA, USA
[3] School of Computer Science, Wuhan University, Wuhan, China
{zhling, xiaolong.zhang}@wust.edu.cn, {chengjiang.long, qingan.yan}@jd.com, cxxiao@whu.edu.cn

Abstract

In this paper, we propose a novel context and lightness aware Generative Adversarial Network (CLA-GAN) framework for shadow removal, which refines a coarse result to a final shadow removal result in a coarse-to-fine fashion. At the refinement stage, we first obtain a lightness map using an encoder-decoder structure. With the lightness map and the coarse result as the inputs, the following encoder-decoder tries to refine the final result. Specifically, different from current methods restricted pixel-based features from shadow images, we embed a context-aware module into the refinement stage, which exploits patch-based features. The embedded module transfers features from non-shadow regions to shadow regions to ensure the consistency in appearance in the recovered shadow-free images. Since we consider pathces, the module can additionally enhance the spatial association and continuity around neighboring pixels. To make the model pay more attention to shadow regions during training, we use dynamic weights in the loss function. Moreover, we augment the inputs of the discriminator by rotating images in different degrees and use rotation adversarial loss during training, which can make the discriminator more stable and robust. Extensive experiments demonstrate the validity of the components in our CLA-GAN framework. Quantitative evaluation on different shadow datasets clearly shows the advantages of our CLA-GAN over the state-of-the-art methods.

CCS Concepts

• ***Computing methodologies*** → *Context; Lightness; GAN; Coarse-to-fine; Shadow Removal;*

1. Introduction

Shadows are an important phenomenon in nature, appearing when light is partially or completely blocked [LLZ*20]. The brightness in shadow regions is lower than that in non-shadow regions. The low brightness in shadows will decrease the accuracy and effectiveness of some computer vision tasks, such as target tracking and recognition [MCKT00, CGP*02], image segmentation and intrinsic image decomposition [LS18]. Moreover, shadow removal and editing can improve the visual effect of images and videos [ZLW19], such as film and television post-editing. Removing shadows from images is now an important research topic in the fields of computer vision [LCSH19,HLYG18,LH17,LHK16,LH15, LHK13, HLYG13, ILBH20] and computer graphics [GTB15].

It is worth mentioning that a high-quality image with shadows removed should satisfy three aspects: (1) the illumination in shadow regions is effectively recovered while preserving the texture details; (2) the recovered illumination and color in shadow regions must be consistent with that of the surrounding environment; and (3) there are no artifacts in the image. However, due to the complexity in shadow regions and the indeterminacy of the light in the scene, producing high-quality shadow removal image is a challenging task.

A number of shadow removal approaches are already available, including the traditional methods based on prior information [SL08,XSXM13,XXZC13], and learning-based methods [GTB15, QTH*17, WLHY18, HFZ*18, WLZX19]. For traditional shadow removal methods, those utilizing illumination priors accurately rely on texture matching between shadow regions and non-shadow regions, while methods based on gradient priors cause obvious artifacts on shadow boundaries due to the irregular illumination changes on shadow boundaries.

More recently, learning-based methods have shown their potential for shadow removal. With proper network models and loss functions, these methods can directly remove shadow in images and produce good results. But they still have some limitations. For example, the designed models focus too much on pixel-based

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

[‡] Corresponding to Chunxia Xiao, Email: cxxiao@whu.edu.cn.

features. Such pixel-based analysis methods cannot effectively build spatial associations, which may reduce the robustness of these methods and produce undesired shadow removal results, especially for shadow images with complex scenes.

In this paper, we propose a new framework CLA-GAN to explore contextual and lightness information with Generative Adversarial Networks for shadow removal. Similar to [ZLZX20], we complete the shadow removal task in a coarse-to-fine fashion, which refines the coarse shadow removal result into a fine shadow-free image. At the coarse stage, we train an encoder-decoder structure to remove shadows in the image and produce a coarse shadow removal result. The role of the refinement stage is to correct the coarse prediction and predict a refined shadow removal result. However, different from the recent learning-based methods which solely focus on pixel-based features, we embed a context-aware module additionally exploring patch-based features. The contextual aware module transfers features from non-shadow regions into shadow regions, which uses features of patches in non-shadow regions as deconvolutional filters to reconstruct features in shadow regions. The use of the context-aware module can enhance the spatial association between pixels and facilitate the global appearance consistent for the whole image.

As lightness is an important characteristic for shadow removal, we add a lightness branch before refining the coarse result. Within the constraints of the lightness ground-truth (L channel of the Lab color space for the ground-truth shadow-free image), the lightness branch can produce a lightness map, which is fed into the following refinement branch. Specially, we concatenate the reconstructed features from the context-aware module as an additional input into the decoders of the lightness branch and the following refinement branch, as illustrated in Figure 1. From the experiments, we find that the context-aware module significantly improves the appearance consistency between shadow and non-shadow regions and makes the shadow removal result more natural.

Furthermore, to reduce the dependence between representations of the discriminator and the quality of the output result from generator, we augment the inputs of the discriminator by rotating the corresponding image pair in different degrees and add a rotation adversarial loss during training the discriminator. In addition, unlike the existing shadow removal methods using deep learning [ZLZX20, CPS20, DLZX19] which are designed with static parameters, we use dynamic weights. The dynamic weights ensure that the model pays more attention to shadow regions gradually, which is beneficial for generating high-quality shadow removal results especially in shadow regions.

To sum up, the main contributions of our work are three-fold as follows:

(1) We introduce a novel CLA-GAN to remove image shadows by jointly exploiting the context and lightness clues. The designed context-aware module, by exploring patch-based features, facilitates the global consistency between shadow and non-shadow regions in the shadow removal result.

(2) To make the discriminator more stable and effective, we augment the input data of the discriminator and add a rotation adversarial loss during training the discriminator.

(3) We use dynamic weights instead of static parameters to make the model pay more attention to shadow regions in training, which facilitates the production of high-quality shadow removal results.

2. Related Work

Several methods have been proposed for image shadow removal [AHO11,KBST16,LZS14,VHY*16,WT05,XTT14,YTA12,LS19, DLZX19, MGK19, YHWCYY20], which can be divided into two categories. One is the traditional shadow removal methods based on prior information. These methods do not need additional datasets, but they need to detect the shadow regions beforehand. The other is learning-based methods, which need to use additional datasets as training samples. These data-driven methods can directly recover the illumination of the image without detecting shadow regions.

One typical type of traditional shadow removal method recovers the illumination in shadow regions using illumination transfer [SL08, XSXM13, GDH11], which transfers the illumination from non-shadow regions to shadow regions. Shor et al. [SL08] build a linear mapping model between shadow regions and non-shadow regions to remove shadows in the images. This method is low time complexity but can only deal with images with consistent textures in shadow regions. Xiao et al. [XXZC13] proposed an illumination transfer method using the matched information of subregions to remove shadows in the image. This method can process the image with multiple textures. Since shadow regions are segmented beforehand and processed independently, the recovered illumination in different shadow areas may be inconsistent. To solve this limitation, Zhang et al. [ZZX15] decomposed the image into many overlapping image patches and proposed a local-to-global method to remove shadows using the proposed local illumination recovering operator. These methods based on illumination transfer [RAGS01] need to match an area in non-shadow regions to the area in shadow regions that they have similar texture, and the effectiveness of the method depends on the accuracy of texture matching and the illumination in the matched non-shadow region. However, texture matching is also a challenging task [YLY*16].

Another typical group of traditional shadow removal approaches removes shadows based on gradient domain manipulation [FHD02, FHLD05, LG08, MTC07]. Gradient information can describe the first-order variation of texture in images [LWZ*20, LXZ*20]. The common idea in these techniques is to redefine the gradient on shadow boundaries and reconstruct the shadow removal result utilizing the gradient information in shadow regions. Finlayson et al. [FHLD05] nullified the gradient on shadow boundaries and reconstructed the shadow-free images by solving a gradient-based Poisson equation. [FF04] removed shadows using gradient domain manipulation by employing Retinex algorithm. Liu et al. [LG08] constructed illumination variation lines at shadow boundaries to eliminate the gradient changes that were caused by illumination changes, and then reconstructed shadowless images using the illumination variation lines. However, due to the influence of the illumination change at the shadow boundary, methods using gradient information can cause boundary problems, such as texture detail losses or color distortions.

More recently, deep neural networks have been widely

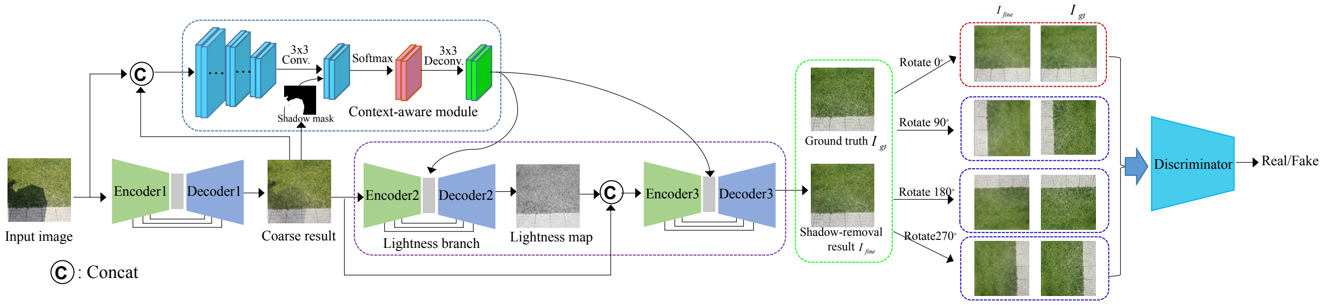

Figure 1: *The framework of the proposed CLA-GAN, which is composed of a generator with three encoder-decoder structures and a discriminator. The generated shadow removal result and the corresponding ground truth shadow-free image are rotated by four different degrees. The expanded data with corresponding labels are fed to the discriminator. The shadow mask in the context-aware module is used to distinguish shadow regions and non-shadow regions; this is obtained by binarizing the residual image between the coarse result and the input image.*

introduced for image processing tasks. Through optimizing the objective function, the models are taught to produce the desired result. Hu et al. [HFZ*18] proposed a Bayesian formulation to remove shadows in images. They first need to detect the shadows in the image using multiple convolutional neural networks. Qu et al. [QTH*17] proposed an end-to-end DeshadowNet to recover the illumination in shadow regions. This network integrates high-level semantic information, middle-level appearance information, and local image details. Wang et al. [WLHY18] proposed a stacked conditional generative adversarial network (ST-CGAN) for image shadow removal. Different from the commonly used multi-branch paradigms, they stacked all the tasks for multi-task learning. Lin et al. [YHWCYY20] proposed a BEDSR-Net to remove shadows from a document image. By combing the global background color of the image and the shadow attention map, this method can remove shadows using a U-net generator. Hu et al. [HJFH19] proposed a Mask-ShadowGAN for shadow removal using unpaired data. Methods using deep learning can produce better shadow removal results, but they need a large training dataset. Different from existing methods, we proposed a CLA-GAN which makes full use of the context feature. It can generate more accurate and natural shadow removal results.

3. Proposed method

3.1. Overview

In this paper, we present a new framework, CLA-GAN, to handle the shadow removal task, as illustrated in Figure 1. Like the generative adversarial network (GAN) architecture [GPAM*14], the proposed network also contains a generator and a discriminator. The generator consists of three encoder-decoders and an encoder structure used as the context-aware module. Our network uses a shadow image as the input and outputs the shadow removal result in an end-to-end manner.

With the input of a shadow image I, the generator network refines the coarse shadow removal result to the fine shadow-free image in a coarse-to-fine fashion. The first encoder-decoder, which is denoted as G_{coarse}, is used to produce a coarse shadow removal result I_{coarse}. The refinement stage is used to correct the coarse prediction and predict the fine shadow removal result. It contains two encoder-decoders, denoted as G_{light} and G_{fine}, respectively. The lightness branch G_{light} is used to produce a lightness map I_{light} which is fed to the following refinement branch G_{fine}. With the inputs of I_{coarse} and I_{light}, the refinement branch can produce the fine shadow removal image I_{fine}.

Specifically, we additionally embed a context-aware module $G_{context}$ at the refinement stage, which maps the features from non-shadow regions to shadow regions and generates reconstructed features $I_{feature}$ for the image. To produce more natural shadow removal results, we take the reconstructed features $I_{feature}$ as an auxiliary input for the decoders of G_{light} and G_{fine}, which ensures the global appearance consistency between shadow regions and non-shadow regions. The relations between the inputs and outputs can be summarized as:

$$I_{coarse} = G_{coarse}(I), \tag{1}$$

$$I_{feature} = G_{context}(I, I_{coarse}), \tag{2}$$

$$I_{light} = G_{light}(I_{coarse}, I_{feature}), \tag{3}$$

$$I_{fine} = G_{fine}(I_{light}, I_{coarse}, I_{feature}), \tag{4}$$

The discriminator D is designed to distinguish whether the generated image is a real image or not. Inspired by data augmentation by rotation, we rotate the shadow removal result I_{fine} and the corresponding ground-truth I_{gt} by four different degrees. By feeding the expanded images to the discriminator, the representations learned by the discriminator are more stable and useful, as shown in Figure 1.

3.2. Coarse-to-fine network

Our generator consists of four components: three encoder-decoder branches, which are used to produce the coarse result, the lightness map and the refined shadow removal result, respectively; and a context-aware module, which is used to transfer the appearance features from non-shadow regions to shadow regions.

Coarse network. The encoder-decoder used in coarse network is a U-Net architecture [RFB15], which is trained to obtain a coarse shadow removal result. Due to the dense block in DenseUNet can alleviate the gradient attenuation problem caused by the low illumination in shadow regions and enhance the input feature information, we apply the DenseUNet architecture [NN18] as the implementation of this encoder-decoder.

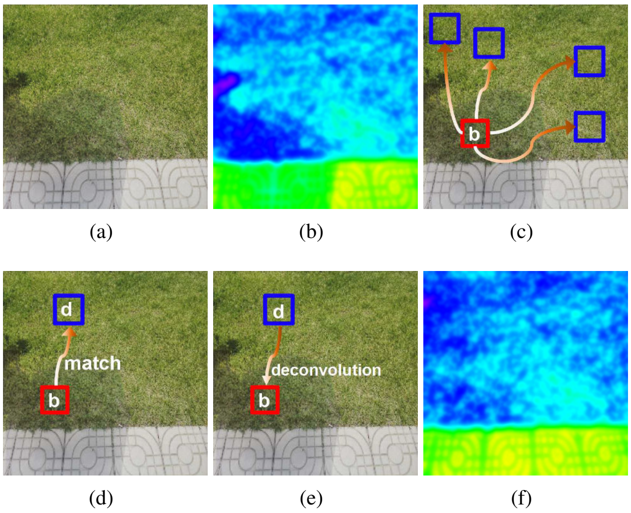

Figure 2: *Illustration of the context-aware module. (a) is the coarse shadow removal result for the input image in Figure 1. (b) is a visualization map for the extracted features from (a), which is plotted based on the similarity of patches cetered at each pixel in the image. In such a visualization map, pixels with similar colors have similar features and share the similar contextual information. By using features of patches in non-shadow regions as convolutional filters, we compute the cosine similarity between this patch and patch b in shadow regions, and find several similar patches (the blue boxes) in non-shadow regions for patch b (the red box in (c)). Then, we obtain the attention scores between each candidate patch and patch b using sotfmax function, and find the most similar neural patch d (the blue box in (d)) for patch in shadow regions. Finally, we use features of patch d as deconvolutional filters to reconstruct the feature of patch b, as shown in (e). (f) is the visualization map for the reconstructed features. This procedure operates on the feature layers.*

Refinement network. In general, we cannot obtain the desired result using the coarse network only. For example, there are still some shadows in the result, or there is color distortion in the image, as shown in Figure 4(b). To address this problem, we add a refinement stage after the coarse network, which is used to improve the quality of the coarse shadow removal results.

As lightness is a significant characteristic associated with shadow in an image, we introduce the lightness map to our refinement stage. As shown in Figure 1, our refinement stage contains a lightness branch and a refinement branch. The lightness branch is used to produce a lightness map without shadows. With the inputs of the coarse result and the lightness map, the following refinement branch is used to produce the desired shadow removal image. Similar to the coarse network, we also apply the DenseUNet architecture as the implementation of the two encoder-decoders at the refinement stage.

Inspired by [ZZX15], we can observe that areas with similar textures should have similar appearance information under the same light condition. The appearance information contain lightness and color. With this prior, we introduce a context-aware module which can deal with the image on feature layers and embed it into the refinement stage.

The goal of the context-aware module is to use the features of patches in non-shadow regions as convolutional filters to process the patches in shadow regions. Simply put, we transfer features from non-shadow regions to shadow regions. The output of the context-aware module is the reconstructed features of the image.

More specifically, we first apply downsampling operations to extract the features. The downsampling part has a similar structure as the encoder in DenseUNet. Then, we extract 3×3 patches centered at each pixel in non-shadow regions and use them as convolutional filters. For patch b centered at the pixel in shadow regions, we find several candidate patches for patch b using cosine similarity [NB10] between patch b and patch in non-shadow regions. Next, we apply a softmax operation to select a most similar non-shadow patch d for each shadow patch. Finally, we use features of the non-shadow patch as deconvolutional filters to reconstruct the shadow regions. The values of overlapped pixels are averaged.

As shown in Figure 2, the two visualization maps reveal that the context-aware module can borrow information from non-shadow regions to help shadow removal. Note that we use a shadow mask to distinguish shadow regions and non-shadow regions in the context-aware module during training. The shadow mask is obtained by binarizing the residual image between the coarse shadow removal result and the input image.

We concatenate the reconstructed features into the decoders of the lightness branch and the refinement branch, as shown in Figure 1. The use of the context-aware module can make our network utilize the surrounding features as references to correct the lightness and color in shadow regions, which make the lightness and color reconstruction in shadow regions more robust. From experiments, we find that the context-aware module significantly improves the illumination consistency between shadow and non-shadow regions, as shown in Figure 9(f).

3.3. Discriminator

The discriminator in the generative adversarial network (GAN) [GPAM*14] is used to distinguish the image produced by the generator is real or fake, compared with the ground truth. In the general case, the input of the discriminator is an image pair with ground truth. Inspired by the idea of data augmentation, to make sure that the discriminator becomes more stable and useful; to do this, we expand the input data of the discriminator.

Inspired by data augmentation by rotation [CZR*19], we rotate the real and fake images simultaneously with different degrees and produce additional label pairs. Similar to the original real and fake image pair, the additional image pairs are also fed to discriminators during training. In our work, we rotate the images by four different degrees. Let R be the set of possible rotations, and $R = \{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$. Let r be a rotation selected from R. When $r = 0^\circ$, the image pair is the original fake and real images; the remaining three pairs are the additional label pairs. Each image pair can produce an adversarial loss, and we consider the four adversarial losses as a rotational adversarial loss.

Our discriminator is a convolutional network. It consists of five convolution layers, each of which is followed by a batch normalization, a Leaky ReLU activation function, and one fully connected layer. The output of the last fully connected layer is the probability value that the image produced by the generator is a real image. We use the spectrum normalization method [MKKY18] to stabilize the training process of the discriminator network.

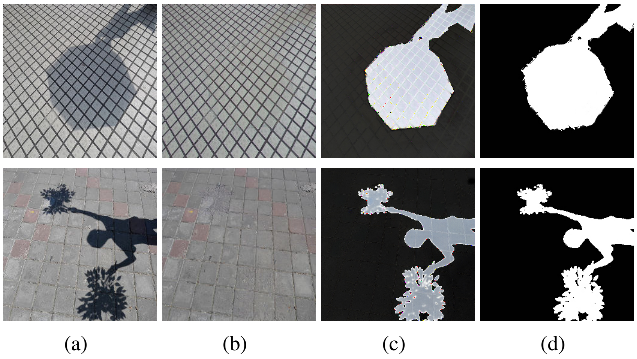

Figure 3: *Results of shadow mask from the residual image. From left to right are: input shadow image (a); shadow removal result (b); the residual image (c); the shadow mask (d).*

3.4. Loss function

To provide constraint information for the network and obtain a robust parametric model, we use a total loss L to optimize the proposed model. The total loss L includes four components: coarse loss L_{coarse}, light loss L_{light}, refined loss L_{fine} and rotational adversarial loss L_{adv}. The total loss L can be rewritten as:

$$L = L_{coarse} + L_{light} + L_{fine} + L_{adv}, \tag{5}$$

The corresponding loss components are described as follows.

Coarse loss. Due to the low illumination can weaken the texture details in shadow regions, we use visual-consistency loss and perceptual-consistency loss to train our coarse network. The perceptual-consistency loss can preserve the image structure. So the objective function for coarse loss can be denoted as:

$$L_{coarse} = \beta_1 L_{coarse\_image} + \beta_2 L_{coarse\_vgg}, \tag{6}$$

where β_1 and β_2 are two weighted parameters.

In our work, the visual-consistency loss is calculated using the L1-norm between the produced result and the corresponding ground truth, and the perceptual-consistency loss is the MSE error of the image features between the produced result and the corresponding ground truth. The image feature is extracted using the pre-trained VGG19 model on the ImageNet dataset. Specifically,

$$L_{coarse\_image} = ||I_{coarse} - I_{gt}||_1, \tag{7}$$

$$L_{coarse\_vgg} = ||VGG(I_{coarse}) - VGG(I_{gt})||_2^2, \tag{8}$$

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

Light loss. Light loss L_{light} is used to train the lightness branch. It is calculated to evaluate the light difference between the predicted light map I_{light} and the ground truth of the light map. We use the L channel in the Lab color space of the original ground truth as the ground truth of light map M_{gt}. The light loss for the lightness branch can be denoted as:

$$L_{light} = \beta_3 ||I_{light} - M_{gt}||_1, \tag{9}$$

where β_3 is the weight.

Refined loss. In our work, the refinement branch is trained using global loss L_{global} and local loss L_{local}. Similar to the coarse network, the global loss consists of the visual-consistency loss and the perceptual-consistency loss, which are denoted as L_{global_image} and L_{global_vgg} respectively. The local loss calculates the visual-consistency loss of shadow regions and non-shadow regions respectively, which are denoted as L_{shadow} and L_{lit}.

In later iterations during training, the loss value of L_{shadow} will get smaller, and the attention for shadow regions becomes smaller. However, we should pay more attention to shadow regions for the shadow removal task. To address this situation, we use dynamic weights to balance the losses of L_{shadow} and L_{lit}. That is, the weights for L_{shadow} and L_{lit} are dynamic during training, instead of using static loss weights.

To sum up, the objective function for the refined loss is written as:

$$\begin{aligned} L_{fine} &= L_{global} + L_{local} \\ &= \beta_4 L_{global\_image} + \beta_5 L_{global\_vgg} + \beta_6 L_{shadow} + \beta_6 L_{lit} \\ &= \beta_4 ||I_{fine} - I_{gt}||_1 + \beta_5 ||VGG(I_{fine}) - VGG(I_{gt})||_2^2 \\ &+ \beta_6 t||(I_{fine} - I_{gt})m||_1 + \beta_6(1-t)||(I_{fine} - I_{gt})(1-m)||_1, \end{aligned} \tag{10}$$

where β_4, β_5 and β_6 are the static parameters. t is the dynamic weight for local losses, and $t = 0.5 + \frac{(i-1)R+j}{E \times R}$, where $i \in \{1, 2, \cdots, E\}$, $i \in \{1, 2, \cdots, R\}$ and $R = \lceil \frac{N}{B} \rceil$. Epochs E and batch size B are the hyperparameters of network. N is the number of images in the training dataset. m is the shadow mask which is obtained by binarizing the residual image between the shadow removal result I_{fine} and the input image I, as shown in Figure 3.

Rotational adversarial loss. L_{GAN} is the rotational adversarial loss for the network, and it is described as:

$$L_{GAN} = \max_D E[\sum_r (log(D(I_{gt}^r)) + log(1 - D(I, I_{fine}^r)))]. \tag{11}$$

where $r \in R = \{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$ and D refers to the discriminator. I_{gt}^r and I_{fine}^r are the images that the ground truth I_{gt} and the shadow removal result I_{fine} rotate r degree, respectively.

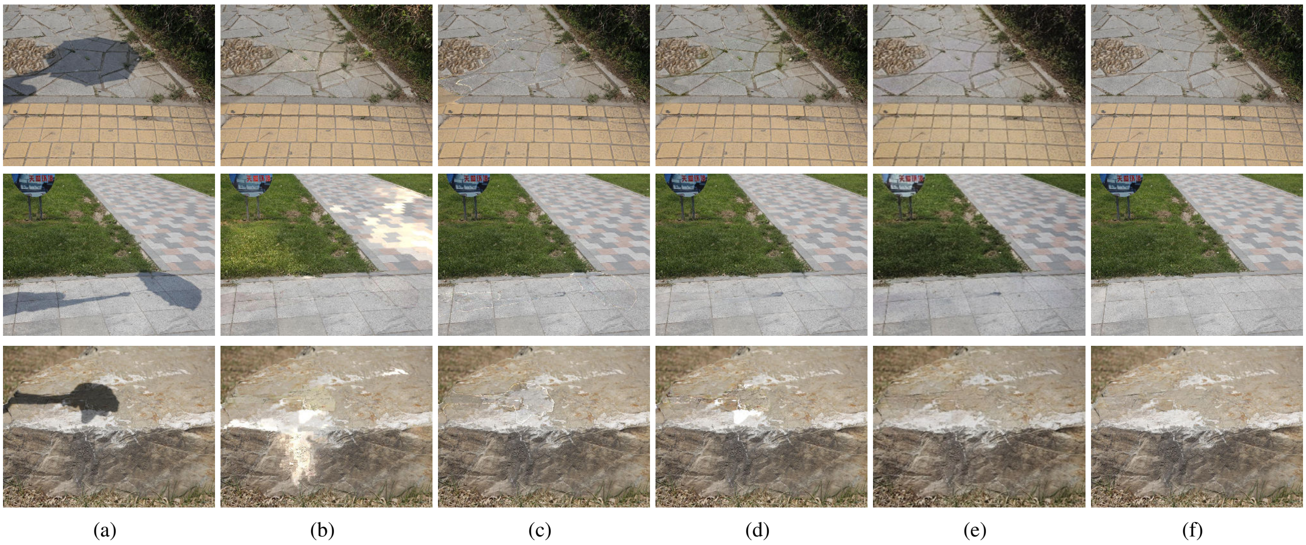

Figure 5: *Shadow removal results compared with traditional methods. From left to right are: input images (a); shadow removal results of Guo [GDH11] (b), Xiao [XXZC13](c), Zhang [ZZX15] (d), and our CLA-GAN (e); and the corresponding ground truth shadow-free images (f).*

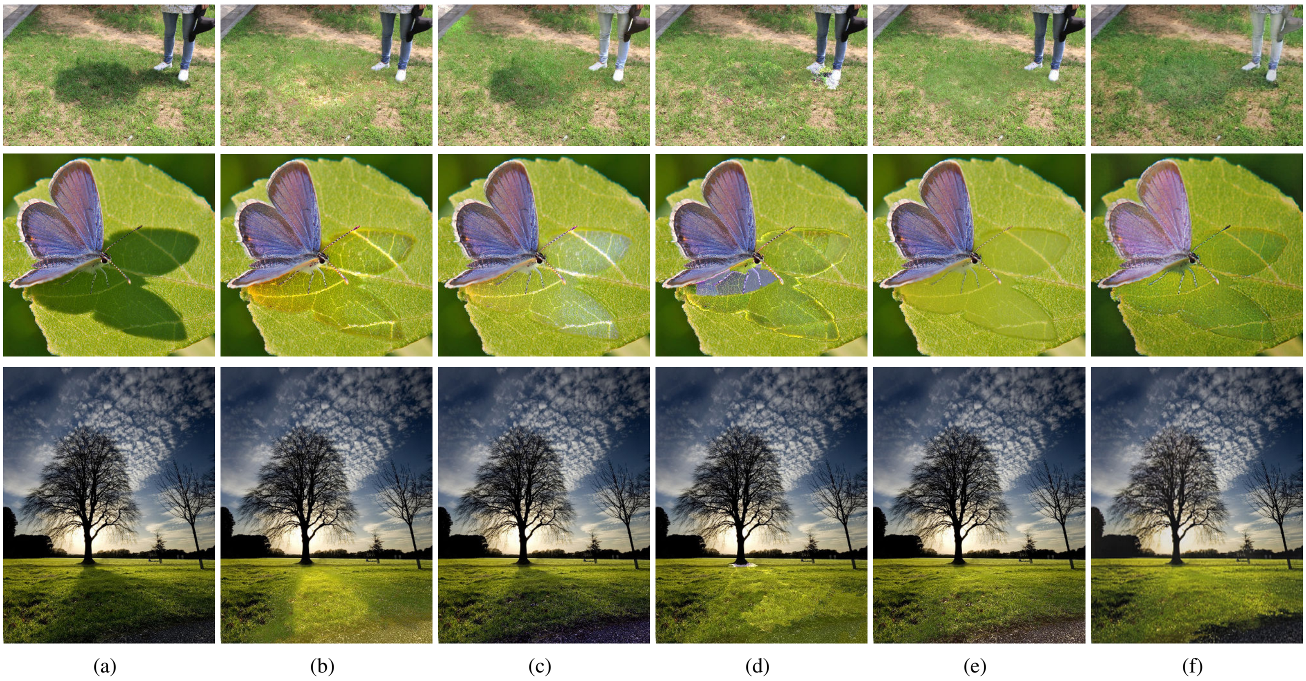

Figure 6: *Shadow removal results compared with traditional methods. From left to right are: input images (a); shadow removal results of Shor [SL08] (b), Guo [GDH11] (c), Xiao [XXZC13](d), Zhang [ZZX15] (e), and our CLA-GAN (f).*

4. Experiments

To verify the effectiveness of our CLA-GAN, we present various experimental results and compare them with the state-of-the-art shadow removal methods.

4.1. Implementation Details

Parameters. Our proposed method is implemented in TensorFlow on a computer with an Intel(R) Core(TM) i5 CPU @3.70GHz and a 16G RAM NVIDIA GeForce RTX 2080Ti. In our experiments, the input size of image is 256×256. The learning rate is set to 0.0001.

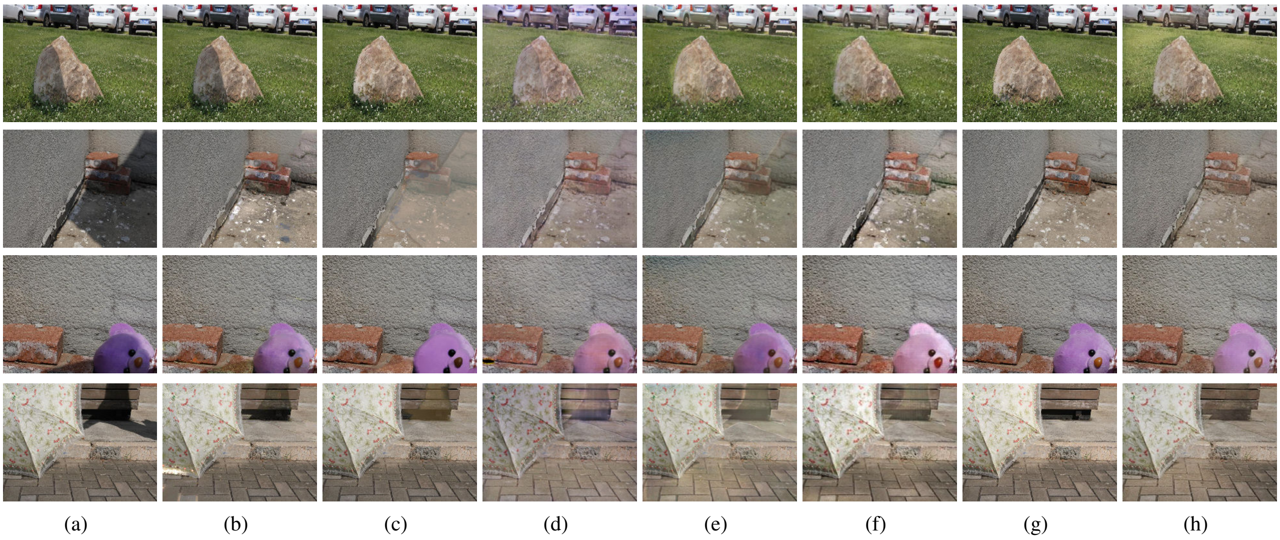

Figure 7: *Shadow removal results. From left to right are: input images (a); shadow removal results of [GDH11] (b), [ZZX15] (c), [WLHY18] (d), [HFZ* 18] (e), [QTH* 17] (f); ground truth (g), and our shadow removal results (h).*

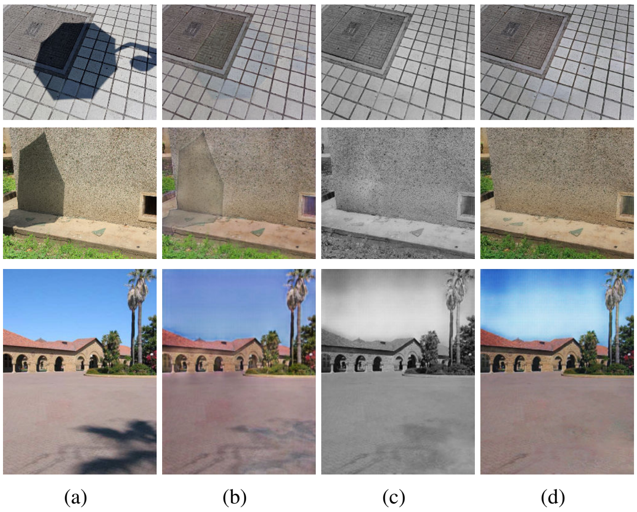

Figure 4: *Intermediate results of the proposed network. From left to right are: input images (a); coarse shadow removal results from coarse network (b); the lightness maps obtained from the lightness branch (c); and the refined shadow removal results (d).*

The parameters β_1, β_2, β_3, β_4, β_5, and β_6 are set to 50, 20, 50, 70, 20, and 70 in our experiments, respectively. We alternatively train the generative network and the discriminative network for 20,000 epochs.

Datasets. We train our model on the ISTD dataset [WLHY18]. The ISTD dataset contains 1870 image triplets of shadow image, shadow mask and shadow-free image. It is divided in two parts: 1330 image triplets for training and 540 image triplets for testing. In our work, we use the shadow images and the corresponding shadow-free images as our inputs during training. We evaluate the shadow removal effectiveness on the testing sets of SRD dataset and ISTD dataset. The testing set of SRD dataset [QTH*17] contains 408 pairs of shadow and shadow-free images.

Metrics. We use the root mean square error (RMSE) calculated in Lab space between the recovered shadow removal result and the ground truth shadow-free image as the metrics to evaluate the shadow removal performance. The smaller the value, the better the performance of this method.

4.2. Experiment and evaluation

Our method consist of four branches, and three of these branches can produce immediate visual results. They are the coarse result, the lightness map and the shadow removal result, as shown in Figure 4. The coarse result and the lightness map are used as inputs for the fine branch to produce the final shadow removal result. This strategy allows our network to obtain good results for both simple and complex scene image. The recovered illumination in shadow regions is consistent with the surrounding environment and the texture details in shadow regions are well preserved.

In the following, we compare our proposed method with the state-of-the-art methods including traditional methods, such as [GDH11], [XXZC13] and [ZZX15], and deep learning-based methods, such as DeshadowNet [QTH*17], DSC [HFZ*18], ST-CGAN [WLHY18], AngularGAN [Sid18] and RIS-GAN [ZLZX20]. To make the comparison fair, 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.

As shown in Table 1 and Table 2, we summarize the comparison results on test datasets of SRD [QTH*17] and ISTD [WLHY18], respectively. The two datasets contain various kinds of shadow scenes. The numerical results reveal the flexibility of the proposed

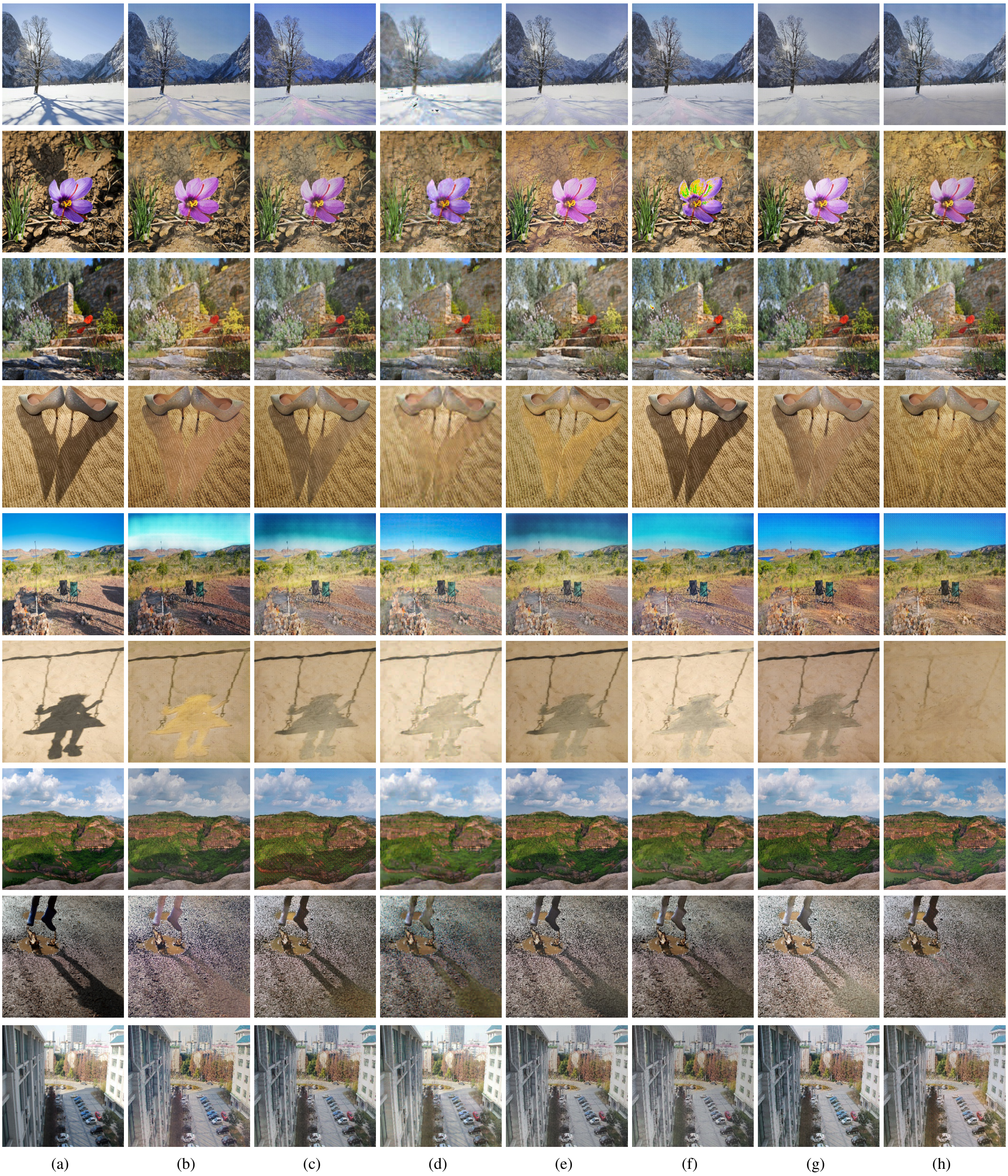

Figure 8: *Shadow removal results compared with deep learning methods. From left to right are: input images (a); shadow removal results of Deshadow [QTH*17] (b), ST-CGAN [WLHY18] (c), DSC [HFZ*18] (d), AngularGAN [Sid18] (e), ARGAN [DLZX19] (f), RIS-GAN [ZLZX20] (g), and our CLA-GAN (h).*

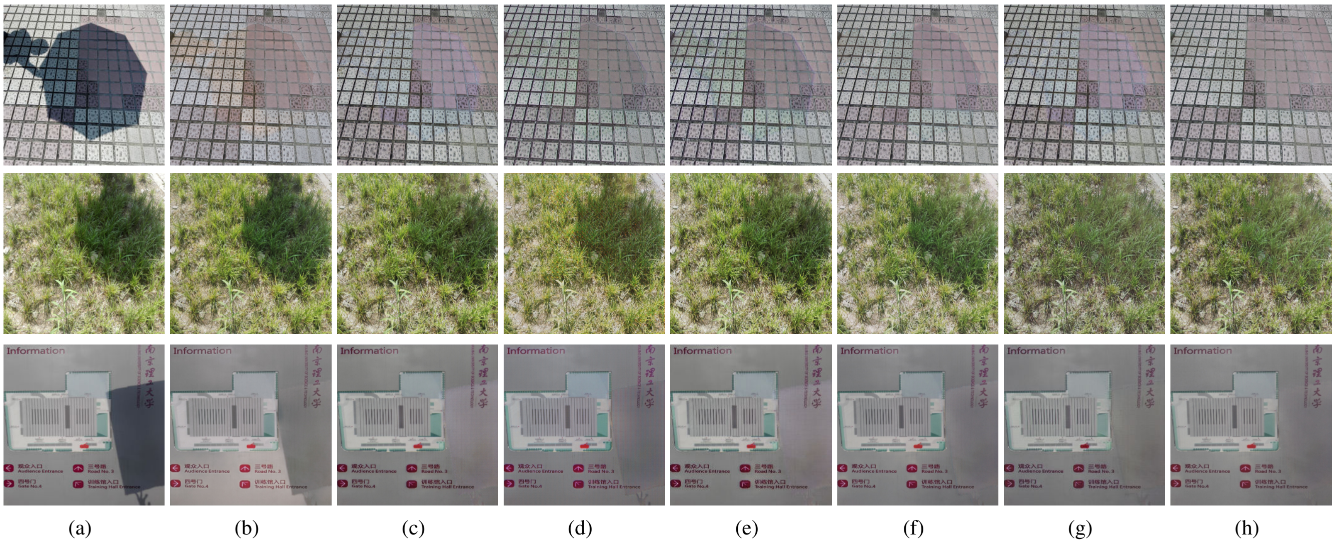

Figure 9: *Shadow removal results. From left to right are: input images (a); shadow removal results of C-GAN (b), LA-GAN (c), CLA-GAN$_1$ (d), CA-GAN (e), CLA-GAN$_2$ (f), CLA-GAN$_3$ (g), and our CLA-GAN (h).*

method. Compared with these methods, our method presents the best results in the whole image. This suggests that the proposed CLA-GAN is efficient and preponderant, which can produce results that are much closer to the ground-truth shadow-free images.

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
Xiao	PG/2013	13.71	6.88	8.94
Zhang	TIP/2015	9.50	6.90	7.24
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
AgularGAN	CVPRW/2019	17.63	7.83	15.97
RIS-GAN	AAAI/2020	8.22	6.05	6.78
CLA-GAN	PG/2020	8.10	6.01	6.59

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
Xiao	PG/2013	14.77	8.01	8.93
Zhang	TIP/2015	9.77	7.12	8.16
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
CLA-GAN	PG/2020	9.01	6.25	6.62

We also give some visualization results to further explain the outperformance of the proposed CLA-GAN. Visual comparisons with traditional methods are shown in Figure 5 and Figure 6. From the results, we find that [SL08, GDH11, XXZC13, ZZX15] can recover the illumination in shadow regions. However, these methods should detect shadow regions before shadow removal and require relative good shadow detection results. Moreover, due to the illumination changes in shadow boundaries, traditional methods may have boundary problems, such as color distortions or texture losses. Compared with these methods, our method not only effectively recovers the illumination in shadow regions, but also reconstructs the illumination and textures of shadow boundaries with less artifacts, as shown in Figure 5(e) and Figure 6(f).

Figure 7 and Figure 8 show some visualization results compared with learning-based methods. Because the method of Qu et al. [QTH*17] does not consider the aspect of illumination and lightness in the image, this method may lead to unsatisfactory shadow removal results such as color distortion or incomplete shadow removal. The same problem is present in the methods of [WLHY18, HFZ*18, DLZX19], as shown in Figure 8(b-d, f). AngularGAN and RIS-GAN employ illumination in their methods, which can makes illumination of non-shadow regions very close to the corresponding ground-truth shadow-free images for some images. But these methods do not focus more on shadow regions when training the models. This may cause incomplete shadow removal, as shown in Figure 8(e, g). In contrast, taking context information and dynamic weights for the loss function into consideration, our CLA-GAN can pay more attention to shadow regions and produce more natural and realistic shadow removal results, as shown in Figure 7(h) and Figure 8(h).

In additional, Figure 6 presents some shadow removal results with complex scenes. The recovered results by our proposed CLA-GAN look more natural and are more suitable for human

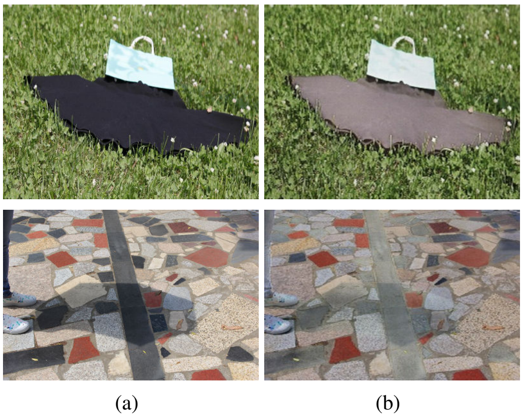

Figure 11: *Limitation. (a) Input images. (b) Results produced by our method.*

visual perception. The pleasant shadow removal results verify the robustness and the potential of the proposed CLA-GAN in these complicated scenes.

4.3. Ablation Study

Our network employs a two-stage strategy, and it can be divided into four components: the coarse branch, the context-aware module, the lightness branch and the refinement branch. Each component plays an important role in the whole network. Specifically, different from the existing learning-based methods for shadow removal, we use local loss with dynamic weights and rotational adversarial loss to training our model. To further evaluate and verify the effectiveness of these components and additional losses in the proposed CLA-GAN, we design a series of variants to get an ablation study. The variants are as follows:

BASE: take the input shadow images as the shadow removal result.

C-GAN: use G_{coarse} only and take I_{coarse} as the shadow removal result.

LA-GAN: remove $G_{context}$ and use $r = 0^\circ$ only; train the model without L_{local}.

CLA-GAN$_1$: add $G_{context}$ and use $r = 0^\circ$ only; train the model without L_{local}.

CA-GAN: CLA-GAN$_1$ without G_{light}.

CLA-GAN$_2$: add $G_{context}$ and use $r = 0^\circ$ only; train the model with L_{local}.

CLA-GAN$_3$: CLA-GAN$_2$ with t=0.5 in Equation 10.

We train the above variants on the same training data. Figure 9 presents some visual results for the mentioned different variants, from which we can clearly see that our CLA-GAN recovers the best details of the shadow removal regions and looks more realistic. We also evaluate the shadow removal results on SRD test dataset and ISTD test dataset. The results are summarized in Table 3, from which we can observe that (1) all the variants can recover the illumination in shadow regions, and get better results than BASE; (2) the context-aware module and the lightness branch are necessary to improve the performance of the shadow removal result and reduce the appearance difference between shadow regions and non-shadow regions; (3) the dynamic weights used in local loss L_{fine_local} and the rotational adversarial loss L_{GAN} are helpful for producing a natural shadow removal result.

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
C-GAN	12.06	7.65	8.85	16.98	9.71	11.27
LA-GAN	10.73	7.25	8.02	10.27	6.67	7.43
CLA-GAN$_1$	9.15	6.62	7.01	9.27	6.25	6.73
CA-GAN	10.62	7.31	7.93	9.64	6.53	6.91
CLA-GAN$_2$	8.96	6.27	6.78	9.11	6.27	6.65
CLA-GAN$_3$	9.23	6.57	7.02	9.20	6.25	6.70
CLA-GAN	8.10	6.01	6.59	9.01	6.25	6.62

Discussion. Shadow removal can be applied to some graphics applications, such as shadow editing. From [CGC*03], we can observe that an observed image C is a linear combination of the shadow-free image B and the shadow image F weighted by the visibility of the light source α, that is $C = \alpha F + (1 - \alpha)B$. With such a compositing equation, users can produce new images using the different shadow images based on their intent, as shown in Figure 10.

The proposed CLA-GAN can remove shadows and create more realistic shadow-free images. However, there is space for improvement. For example, our model does not distinguish black objects from shadows. When an object surface is very dark in color and appears black, our model may consider it as a shadow and perform shadow removal for this object, as shown in Figure 11, which is undesired in practice. Besides, since the environmental luminosity and camera exposure may vary, a training pair may have inconsistent tone and brightness in non-shadow regions [HFZ*18]. Given inconsistent training pairs, the network based on supervised data could produce biased results with slight color (tone or brightness) change in non-shadow regions.

5. Conclusions

In this paper, we propose a novel framework CLA-GAN exploring contextual and lightness information for shadow removal in a coarse-to-fine fashion. With the reconstructed feature information from the context-aware module embedded in the refinement stage, we can produce pleasant shadow removal results which have more consistent between shadow regions and non-shadow regions. Moreover, by applying the dynamic weights in the local loss, our CLA-GAN can pay more attention to shadow regions gradually during the training process. To make the discriminator more stable and useful, we augment the inputs of the discriminator and use rotational adversarial loss during training. The experimental results show that the proposed CLA-GAN can produce more natural shadow removal results.

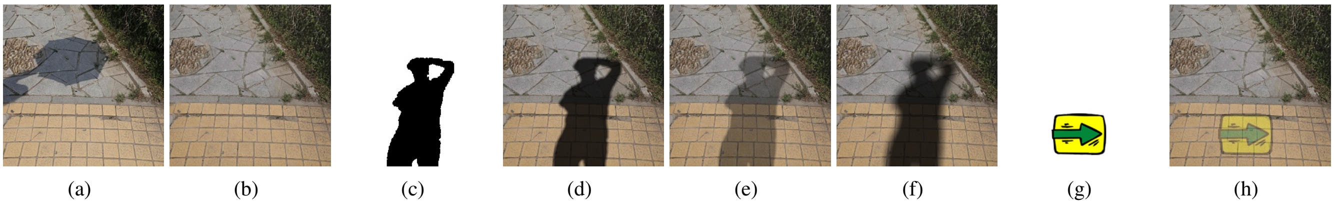

Figure 10: *Shadow editing. (a) is the input shadow image. (b) is our shadow removal result. (d), (e) and (f) are shadow editing results with different α using the new shadow image (c). (h) is the compositing image using image (g).*

6. Acknowledgments

This work was partly supported by the Key Technological Innovation Projects of Hubei Province (2018AAA062), NSFC (No.61902286, No.61672390, No.61972298, No.61972299, No.U1803262), China Postdoctoral Science Found (No.2018M642933), the National Key Research and Development Program of China (2017YFB1002600).

References

[AHO11] Arbel E., Hel-Or H.: Shadow removal using intensity surfaces and texture anchor points. *IEEE Transaction on Pattern Analysis and Machine Intelligence 33*, 6 (2011), 1202–1216. 2

[CGC*03] Chuang Y. Y., Goldman D. B., Curless B., Salesin D. H., Szeliski R.: Shadow matting and compositing. *ACM Transactions on Graphics 22*, 3 (2003), 494–500. 10

[CGP*02] Cucchiara R., Grana C., Piccardi M., Prati A., Sirotti S.: Improving shadow suppression in moving object detection with hsv color information. In *Intelligent Transportation Systems* (2002). 1

[CPS20] Cun X., Pun C., Shi C.: Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting gan. *AAAI Conference on Artificial Intelligence* (2020). 2

[CZR*19] Chen T., Zhai X., Ritter M., Lucic M., Houlsby N.: Self-supervised gans via auxiliary rotation loss. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019). 4

[DLZX19] Ding B., Long C., Zhang L., Xiao C.: Argan: Attentive recurrent generative adversarial network for shadow detection and removal. *IEEE Conference on Computer Vision and Pattern Recognition* (2019). 2, 8, 9

[FF04] Fredembach C., Finlayson G. D.: Fast re-integration of shadow free images. *Color imaging conference* (2004), 117–122. 2

[FHD02] Finlayson G. D., Hordley S. D., Drew M. S.: Removing shadows from images. In *European Conference on Computer Vision* (2002), vol. 2353, pp. 823–836. 2

[FHLD05] Finlayson G. D., Hordley S. D., Lu C., Drew M. S.: On the removal of shadows from images. *IEEE Transaction on Pattern Analysis and Machine Intelligence 28*, 1 (2005), 59–68. 2

[GDH11] Guo R., Dai Q., Hoiem D.: Single-image shadow detection and removal using paired regions. In *IEEE Conference on Computer Vision and Pattern Recognition* (2011), pp. 2033–2040. 2, 6, 7, 9

[GPAM*14] Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y.: Generative adversarial nets. In *Advances in neural information processing systems* (2014), pp. 2672–2680. 3, 4

[GTB15] Gryka M., Terry M., Brostow G. J.: *Learning to Remove Soft Shadows*. ACM Transactions on Graphics, 2015. 1

[HFZ*18] Hu X., Fu C. W., Zhu L., Qin J., Heng P. A.: Direction-aware spatial context features for shadow detection and removal. In *IEEE Conference on Computer Vision and Pattern Recognition* (2018). 1, 3, 7, 8, 9, 10

[HJFH19] Hu X., Jiang Y., Fu C., Heng P.: Mask-shadowgan: Learning to remove shadows from unpaired data. *IEEE International Conference on Computer Vision* (2019), 2472–2481. 3

[HLYG13] Hua G., Long C., Yang M., Gao Y.: Collaborative active learning of a kernel machine ensemble for recognition. In *IEEE International Conference on Computer Vision* (2013). 1

[HLYG18] Hua G., Long C., Yang M., Gao Y.: Collaborative active visual recognition from crowds: A distributed ensemble approach. *IEEE Transaction on Pattern Analysis and Machine Intelligence 40*, 3 (2018), 582–594. 1

[ILBH20] Islam A., Long C., Basharat A., Hoogs A.: Doa-gan: Dual-order attentive generative adversarial network for image copy-move forgery detection and localization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2020). 1

[KBST16] Khan S. H., Bennamoun M., Sohel F., Togneri R.: Automatic shadow detection and removal from a single image. *IEEE Transaction on Pattern Analysis and Machine Intelligence 38*, 3 (2016), 431–446. 2

[LCSH19] Long C., Collins R., Swears E., Hoogs A.: Deep neural networks in fully connected crf for image labeling with social network metadata. In *IEEE Winter Conf. on Applications of Computer Vision* (2019). 1

[LG08] Liu F., Gleicher M.: Texture-consistent shadow removal. In *European Conference on Computer Vision* (2008), pp. 437–450. 2

[LH15] Long C., Hua G.: Multi-class multi-annotator active learning with robust gaussian process for visual recognition. In *IEEE International Conf. on Computer Vision* (2015). 1

[LH17] Long C., Hua G.: Correlational gaussian processes for cross-domain visual recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 118–126. 1

[LHK13] Long C., Hua G., Kapoor A.: Active visual recognition with expertise estimation in crowdsourcing. In *IEEE International Conf. on Computer Vision* (2013). 1

[LHK16] Long C., Hua G., Kapoor A.: A joint gaussian process model for active visual recognition with expertise estimation in crowdsourcing. *International Journal of Computer Vision 116*, 2 (2016), 136–160. 1

[LLZ*20] Liu D., Long C., Zhang H., Yu H., Dong X., Xiao C.: Arshadowgan: Shadow generative adversarial network for augmented reality in single light scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2020). 1

[LS18] Li Z., Snavely N.: Learning intrinsic image decomposition from watching the world. 1

[LS19] Le H., Samaras D.: Shadow removal via shadow image decomposition. *IEEE Conference on Computer Vision and Pattern Recognition* (2019). 2

[LWZ*20] LIU Z., WANG W., ZHONG S., ZENG B., LIU J., WANG W.: Mesh denoising via a novel mumfordâĂŞshah framework. *Comput-Aided Des. (Proceedings of Solid and Physical Modeling) 126* (2020), 102858. 2

[LXZ*20] LIU Z., XIAO X., ZHONG S., WANG W., LI Y., ZHANG L., XIE Z.: A feature-preserving framework for point cloud denoising. *Comput-Aided Des. (Proceedings of Solid and Physical Modeling) 127* (2020), 102857. 2

[LZS14] LI H., ZHANG L., SHEN H.: An adaptive nonlocal regularized shadow removal method for aerial remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing 52*, 1 (2014), 106–120. 2

[MCKT00] MIKI I., COSMAN P. C., KOGUT G. T., TRIVEDI M. M.: Moving shadow and object detection in traffic scenes. In *International Conference on Pattern Recognition* (2000). 1

[MGK19] MURALI S., GOVINDAN V. K., KALADY S.: Single image shadow removal by optimization using non-shadow anchor values. *Computational Visual Media* (2019). 2

[MKKY18] MIYATO T., KATAOKA T., KOYAMA M., YOSHIDA Y.: Spectral normalization for generative adversarial networks. *arXiv preprint arXiv* (2018). 5

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

[NB10] NGUYEN H. V., BAI L.: Cosine similarity metric learning for face verification. *Asian Conference on Computer Vision* (2010), 709–720. 4

[NN18] N. B. R., N V.: Single image haze removal using a generative adversarial network. *IEEE Conference on Computer Vision and Pattern Recognition* (2018). 4

[QTH*17] QU L., TIAN J., HE S., TANG Y., LAU R. W. H.: Deshadownet: A multi-context embedding deep network for shadow removal. In *IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 2308–2316. 1, 3, 7, 8, 9

[RAGS01] REINHARD E., ADHIKHMIN M., GOOCH B., SHIRLEY P.: Color transfer between images. *IEEE Computer Graphics Applications 21*, 5 (2001), 34–41. 2

[RFB15] RONNEBERGER O., FISCHER P., BROX T.: U-net: Convolutional networks for biomedical image segmentation. *medical image computing and computer assisted intervention* (2015), 234–241. 4

[Sid18] SIDOROV O.: Conditional gans for multi-illuminant color constancy: Revolution or yet another approach? *arXiv: Computer Vision and Pattern Recognition* (2018). 7, 8

[SL08] SHOR Y., LISCHINSKI D.: The shadow meets the mask: Pyramid-based shadow removal. In *Computer Graphics Forum* (2008), pp. 577–586. 1, 2, 6, 9

[VHY*16] VICENTE T. F. Y., HOU L., YU C. P., HOAI M., SAMARAS D.: *Large-Scale Training of Shadow Detectors with Noisily-Annotated Shadow Examples*. Springer International Publishing, 2016. 2

[WLHY18] WANG J., LI X., HUI L., YANG J.: Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *IEEE Conference on Computer Vision and Pattern Recognition* (2018). 1, 3, 7, 8, 9

[WLZX19] WEI J., LONG C., ZOU H., XIAO C.: Shadow inpainting and removal using generative adversarial networks with slice convolutions. *Computer Graphics Forum 38*, 7 (2019), 381–392. 1

[WT05] WU T. P., TANG C. K.: A bayesian approach for shadow extraction from a single image. In *IEEE International Conference on Computer Vision* (2005), pp. 480–487. 2

[XSXM13] XIAO C., SHE R., XIAO D., MA K. L.: Fast shadow removal using adaptive multi-scale illumination transfer. *Computer Graphics Forum 32*, 8 (2013), 207–218. 1, 2

[XTT14] XIAO Y., TSOUGENIS E., TANG C.: Shadow removal from single rgb-d images. In *IEEE Conference on Computer Vision and Pattern Recognition* (2014), pp. 3011–3018. 2

[XXZC13] XIAO C., XIAO D., ZHANG L., CHEN L.: Efficient shadow removal using subregion matching illumination transfer. *Computer Graphics Forum 32*, 7 (2013), 421–430. 1, 2, 6, 7, 9

[YHWCYY20] YUN-HSUAN L., WEN-CHIN C., YUNG-YU C.: Bedsr-net: A deep shadow removal network from a single document image. *IEEE Conference on Computer Vision and Pattern Recognition* (2020). 2, 3

[YLY*16] YE M., LIANG C., YU Y., WANG Z., LENG Q., XIAO C., CHEN J., HU R.: Person reidentification via ranking aggregation of similarity pulling and dissimilarity pushing. *IEEE Transactions on Multimedia 18*, 12 (2016), 2553–2566. 2

[YTA12] YANG Q., TAN K. H., AHUJA N.: Shadow removal using bilateral filtering. *IEEE Transaction on Image Processing 21*, 10 (2012), 4361–4368. 2

[ZLW19] ZHANG S., LIANG R., WANG M.: Shadowgan: Shadow synthesis for virtual objects with conditional adversarial networks. *Computational Visual Media 5*, 01 (2019), 106–116. 1

[ZLZX20] ZHANG L., LONG C., ZHANG X., XIAO C.: Ris-gan: Explore residual and illumination with generative adversarial networks for shadow removal. In *AAAI Conference on Artificial Intelligence* (2020). 2, 7, 8

[ZZX15] ZHANG L., ZHANG Q., XIAO C.: Shadow remover: Image shadow removal based on illumination recovering optimization. *IEEE Transaction on Image Processing 24*, 11 (2015), 4623–36. 2, 4, 6, 7, 9