Received: 5 July 2021 | Revised: 27 September 2021 | Accepted: 12 October 2021

IET Image Processing

DOI: 10.1049/ipr2.12377

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH PAPER

# Data augmentation and shadow image classification for shadow detection

**Guoquan Li**[1] | **Lingyun Wen**[1] | **Zhengwen Huang**[2] | **Ruiyang Xia**[1] | **Yu Pang**[1]

[1] School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, People's Republic of China

[2] Department of Electronic and Computer Engineering, Brunel University London, London, United Kingdom

**Correspondence**
Lingyun Wen, School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, People's Republic of China.
Email: s190101041@stu.cqupt.edu.cn

**Funding information**
National Key Research and Development Program of China, Grant/Award Number: 2019YFC1511300; the National Natural Science Foundation of China, Grant/Award Number: 61971079; the Basic Research and Frontier Exploration Project of Chongqing, Grant/Award Number: cstc2019jcyj-msxmX0666; the Innovative Group Project of the National Natural Science Foundation of Chongqing, Grant/Award Number: cstc2020jcyj-cxttX0002; Department of Science and Technology of Sichuan Province, Grant/Award Number: 2020YFQ0025

**Abstract**

Shadow detection is an important branch of computer vision. Recently, convolutional neural network (CNN)-based methods for shadow detection have achieved better performance than methods based on manually designed features. However, CNNs are extremely hungry for data and the training of CNN-based shadow detector requires time-consuming and expensive pixel-level annotations. To alleviate this problem in shadow detection, a method of data augmentation based on generative adversarial network (GAN), named ShadowGAN, has been proposed. Given a shadow mask and a shadow-free image, our ShadowGAN can generate shadow images with labels. To guide the training of ShadowGAN and get more realistic shadow images, $\mathcal{L}_1$ loss is further implemented to impose a restriction between real shadow images and generated shadow images. The effectiveness of ShadowGAN is demonstrated by training existing shadow detectors on enlarged dataset. In addition, to better make use of shadow-free images in shadow detection, shadow image classification task is added for the shadow detectors. Experiments show that this task can guide the feature extraction network to learn more robust shadow features. At last, these two methods are combined and a better performance of shadow detection is achieved.

## 1 | INTRODUCTION

Shadows are a common phenomenon in natural scenes, it may hamper or benefit some tasks of computer vision. On the one hand, shadows can provide auxiliary information such as light direction [1], camera location [2] and scene geometry [3]. On the other hand, objects covered by shadows can be regarded as adversarial samples, thus the performance of some tasks, such as object detection, tracking, semantic segmentation and recognition will suffer from shadows [4–8]. Therefore, shadow detection is widely used in computer vision tasks as one important pre-processing step. Early shadow detection methods rely on physical models of illuminations or manually designed shadow features [9–12], and require high imaging quality and several assumptions, such as Lambertian reflectance, Planckian lighting and narrowband camera sensors, which is not easy to be satisfied. As a result, these methods fail to detect shadows under different illuminations and environments. Convolutional neural network (CNN)-based shadow detection is demonstrated with higher accuracy and better generalization ability [13–16]. However, the training of CNN-based shadow detectors requires datasets with pixel-level annotations and labelling of shadow images is very time-consuming and expensive. Existing datasets for shadow detection include ISTD [17] and SBU [18], which only contain 1870 (1330 for training, 540 for testing) and 4727 (4089 for training, 638 for testing) image pairs respectively. The data is not enough for training CNN-based shadow detection to realize accurate feature extraction.

Generative adversarial networks (GAN) [19], which put two networks (a generator and a discriminator) competing with each other, is a powerful framework for learning data distribution and generating images from random noise distribution. Recently, researchers have carried out significant amount of works to improve its performance and expand its applications [20–23].

In this way, existing shadow detection dataset can be enlarged without labelling shadow images manually. These works make it possible to generate training samples of high quality via GAN. Inspired by [24], which implements Cycle-GAN [21] to generate shadow-free and shadow images with unpaired data, we propose ShadowGAN for generating shadow images only. Although the training of Cycle-GAN does not require paired data, we prefer a single-direction framework because a bidirectional network is distracted. To guide the training of ShadowGAN and get more realistic results, we further add $\mathcal{L}_1$ loss between real shadow images and generated shadow images. After ShadowGAN is trained, the discriminator is discarded. To generate new shadow samples, a randomly selected mask and a shadow-free image are fed into the generator. In this way, another 1330 shadow images with accurate annotations are collected and are added into the original training set. To improve the effectiveness of our ShadowGAN, we carried out extensive experiments and comparisons by training two recently proposed shadow detectors DSC [13] and BDRAR [14] on augmented datasets and original dataset, respectively. Results indicate that the performance of these two shadow detectors is improved by training on augmented datasets.

Moreover, existing shadow detection datasets, such as ISTD, contain shadow and shadow-free images. To make full use of these shadow-free images in shadow detection and improve shadow detection accuracy, in this work, we take a deeper step in exploring the importance of semantic information for shadow detection. Apart from augmenting the dataset with generated images, we further propose a shadow classification module and integrate this module with DSC and BDRAR. This module aims at classifying whether an image is a shadow image or not. Adding this module enables the feature extraction network of shadow detectors to extract more robust features. In other words, this module will suppress non-shadow features and enhance shadow features. On the other hand, the training of this module does not require pixel-level labels. Our proposed shadow classification module is light-weighted and easy to integrate with existing shadow detectors. The proposed classification module is simple yet effective and can augment the training set of shadow detection in an indirect way.

Finally, we combine our ShadowGAN and shadow classification module. In other words, we train shadow detectors integrated with shadow classification module on dataset augmented by images generated from ShadowGAN. Experiments show that our ShadowGAN and shadow classification module are compatible.

The main contributions are summarized as follows:

- A data augmentation method, ShadowGAN, is proposed and designed to enrich the training dataset for CNN-based shadow detection models, which does not require labelling of shadow images manually. Experiments demonstrate that the performance of shadow detectors trained on augmented dataset is improved.
- A shadow detection method assisted by shadow image classification is proposed, which is beneficial for the feature extraction network to extract more robust features. By joint training of shadow detection and shadow image classification, the results demonstrate that shadow image classification is effective for improving shadow detection.
- How semantic information affects shadow detection is deeply investigated and an insight for applying classification task to other computer vision tasks is provided.

## 2 | RELATED WORKS

### 2.1 | Shadow detection

#### 2.1.1 | Traditional methods

Early methods rely on physical characteristics of shadows to detect shadows, such as illuminant-invariant properties [10], texture and region properties [25–27]. In these works, the method based on entropy minimizing [10] is a typical one. This work proposed a method in which the illumination-invariant direction of RGB images is computed by entropy minimization, and then RGB images are projected into grey-scale images, named intrinsic images, in the illumination-invariant direction. This method assumed that shadow edges will be removed in illumination-invariant images and thus shadows can be located. Researchers further explored edge and pixel information of shadows. Zhu implemented shadow variant and shadow-invariant cues, which include pixel intensity, gradient and texture, to train a shadow classifier to detect shadows [25]. Besides, Huang trained a SVM as shadow boundary detector and implemented detected edges to recovery shadow regions [26]. After that, region-level information is considered. Guo addressed shadow detection by computing illumination information of segmented regions and applied graph-cut to label shadow and non-shadow regions [27].

#### 2.1.2 | CNN-based methods

The above methods have achieved considerable results, but CNN-based methods are proved to be more accurate and easier to generalize on other datasets. Khan is the first to explore shadow detection with CNNs, he proposed to use multiple convolutional neural networks to learn shadow features [28]. After shadows are detected, an algorithm is further applied to remove shadows. Shen proposed a structured CNN to exploit the local structure of shadow edges and formulated the recovery of shadow regions as least-square optimization problem [29]. Considering shadows contain abundant semantic information, Vicente proposed a semantic-aware patch-level stacked CNN model to detect shadows [18]. Nguyen proposed scGAN and imposed a sensitivity parameter to loss functions [30]. Recently, Hu proposed Direction-aware spatial context (DSC) module [13] and integrated it into feature pyramid network (FPN) [31] to detect shadows from different layers of FPN. Zhu claimed detecting shadows from bidirectional FPN network is beneficial to the task, he integrated recurrent attention residual (RAR) module into FPN from low layer to high layer and from high layer to low layer [14]. Zheng proposed Distraction-aware attention module, aiming at forcing FPN to distract or concentrate on features of false positive detections and false negative detections from previous methods [15]. Very recently, Chen proposed a multi-task semi-supervised model to make full use of unlabelled data [16] and this method can be regarded as a type of data augmentation. To get large high quality shadow detection/removal dataset, Guo [32] proposed a synthetic dataset for shadow detection/removal and proposed adversarial domain adaption to minimize domain bias. In [32] , GAN-based shadow detection is implemented, and the detected shadows are further used to guide the training of shadow removal. These CNN-based methods have achieved great improvement in shadow detection, but their ability may be limited by the scale of existing datasets. Besides, semantic information is significant for shadow detection. Existing datasets, such as ISTD dataset, contain shadow and shadow-free images. However, these works did not take a further step in making use of shadow-free images to strengthen their detectors to extract more useful semantic information. In this paper, we show that by adding shadow classification module, the FPN will do a better work in learning semantic information of shadows.

## 2.2 | Shadow generation

Shadows make synthesized images more realistic. There are several works exploring this. To generate shadows for 3D models inserted into an image, Zhang et al. [33] proposed their ShadowGAN. They use local and global adversarial discriminator to ensure realistic shadow shape and global consistency with existing shadows. Liu et al. [34] proposed ARShadowGAN to model shadow generation of virtual objects. They make use of attention mechanism to map the relation of shadows and environment. These two works aim at augmenting reality of virtual object. Liu et al. [35] proposed G2R-ShadowNet to generate shadow/shadow-free pairs from shadow images and shadow masks. They first cut out the original shadow region and then generate new shadows with a randomly selected shadow mask. They collected a new shadow removal dataset and trained a shadow removal network on this dataset. In this work, similar idea is adopted to generate shadow images.

## 2.3 | Data augmentation

There are common data augmentation methods for training CNNs, such as randomly cropping, randomly flipping and colour space transformation. However, some methods like randomly cropping and colour space transformation are not suitable to shadow detection datasets because such operations may break down semantic information. GAN is first proposed to learn the data distribution and experiments show that GAN can generate images from noises although these generated images are of low quality [19]. Inspired by this work, researchers carried out significant amount of works to improve the quality of generated images and expanded the idea of GAN to translate objects belong to a category to objects belong to another category, or domains to domains [20–23]. The powerful generative ability makes it possible to generate training data via GAN when training samples are difficult and expensive to collect. Zhong et al. [36] proposed a method based on Cycle-GAN to transform image style captured by a camera to the style captured by another camera. This method alleviates image variations caused by different cameras in person re-identification. As described in [37], existing datasets for threaten object detection in X-ray images are unbalanced because prohibited items seldom occur in security checkpoint while normal samples are abundant and easy to collect. CNN detectors trained on such datasets tend to predict input images as normal regardless of normal or abnormal samples. To solve this problem, a modified SAGAN [38] is proposed to generate new prohibited samples. Besides, to get more prohibited samples with different poses, this work implemented Cycle-GAN to translate real images containing prohibited items into X-ray style images. Considering the problem that a model trained on a specific domain hardly generalize to another domain, Huang et al. proposed a structure-aware network [39], named AugGAN, to achieve day-to-night domain adaptation and trained Faster-RCNN [40] and YOLO [41] on enlarged dataset to demonstrate the effectiveness of AugGAN.

In the field of shadow detection, Le et al. [42] proposed A+D Net, which contains a shadow detection network (D-Net) and a network (A-Net) generating hard-to-predict samples, to augment the training set and detect shadows. The A-Net focuses on generating adversarial samples to fool D-Net and the D-Net is trained on images from original dataset and A-Net. Very recently, by encouraging the output consistency of student and teacher network on unlabelled data, Chen et al. successfully improved shadow detection accuracy of their proposed multi-task shadow detector [16]. In shadow detection, researchers mainly focus on how to improve the detection accuracy, while this work concentrates on how to augment existing dataset without spending much time on labelling shadow images.

# 3 | PROPOSED METHODS

## 3.1 | ShadowGAN

Collecting images from real world is a common method to build a dataset but it requires more manual efforts, especially when pixel-level annotation is required. This work takes an initiative in collecting new samples via AGN. Our idea is inspired by [24], this work proposed unpaired shadow removal with Cycle-GAN by generating shadow-free and shadow images. However, this work aims at removing shadows and the generated shadow images are lack of reality. To solve this problem, we use a single-direction network instead of a network with bidirectional structure because we focus more on generating shadow images. Accordingly, we use paired data to train the network and add $\mathcal{L}_1$ loss to enable the generator to output more realistic shadows. We take the network in [24], the generator includes a down-sampling block and an up-sampling block with double convolutions. Between the two blocks, nine residual blocks with stride-two convolutions are inserted. Rather than output the

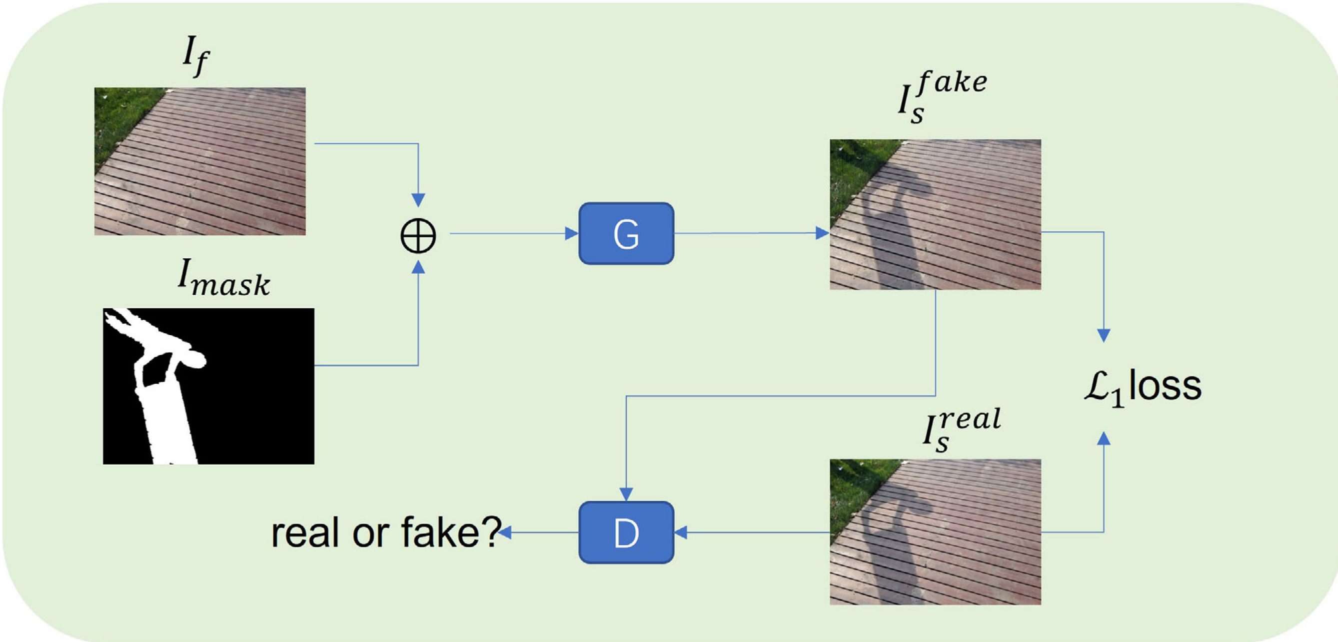


**FIGURE 1** The pipeline of ShadowGAN

feature map directly, the generator takes a residual learning strategy and adds up the learned feature map and input image. This strategy helps the generator to learn complementary shadow information and reserve original non-shadow information. The discriminator is a bunch of convolutions, its last layer takes the responsibility to classify the input as real or fake. In the generator and discriminator, after each convolution, instance normalization [43] is followed.

Unlike [24] which aims at learning to remove shadows from unpaired data, our goal is to generate new shadow images and enlarge the shadow detection dataset. Thus, we only use a pair of generator and discriminator. The pipeline of our ShadowGAN is shown in Figure 1.

As can be seen in Figure 1, during training phase, a shadow-free image and a shadow mask are fed into the generator. After that, the generator will generate shadows on the regions indicated by shadow mask. The generated image and corresponding real shadow image will be fed into discriminator. The discriminator is adopted to find out fake shadow images, while the generator is utilized to generate more realistic shadow images to cheat the discriminator. During such a process, the following objective function is optimized:

$$L(G,D) = E_{I_s^{real}\sim P_{data}(I_s^{real})}\left[\log\left(D\left(I_s^{real}\right)\right)\right] + E_{I_f\sim P_{data}(I_f)} \left[\log\left(1 - D\left(G\left(I_f \oplus I_{mask}\right)\right)\right)\right] + \|I_s^{real} - G\left(I_f \oplus I_{mask}\right)\|_1 \quad (1)$$

where $I_s^{real}$, $I_f$ and $I_{mask}$ denote real shadow images, shadow-free images and corresponding shadow masks respectively. $\oplus$ denotes the concatenation operation. $\|\cdot\|_1$ denotes $\mathcal{L}_1$ loss. $G$ is generator and $D$ is discriminator. Although adversarial loss is enough to fool the discriminator, the generated shadows tend to be artifacts. That is, the generator tends to only generate some black regions, but the discriminator will regard these regions as shadows. As pointed in [44], with only adversarial loss, the generator will fail to learn contextual information and thus the generated images will be less realistic. One way to solve this problem is to add additional restrictions, such as $\mathcal{L}_1$ loss and $\mathcal{L}_2$ loss. In view of $\mathcal{L}_1$ loss enable the network to generate less blurry images than $\mathcal{L}_2$ loss [21], in this work, $\mathcal{L}_1$ loss is imposed between generated shadow images and real shadow images. In testing phase, the discriminator will be discarded. To get new samples, a shadow mask and a shadow-free image are randomly sampled from ISTD training set and fed into the trained generator. Figure 2 shows several real shadow images and shadow images generated by our ShadowGAN. As can be seen in Figure 2, our ShadowGAN successfully generates realistic shadow images and these images are in various locations and shapes compared to original shadow images.

## 3.2 | Shadow image classification

To enable the feature extraction network to learn more robust shadow features, we follow a simple idea that a well-trained shadow detector can learn shadow features and it is supposed to be able to distinguish shadow and non-shadow images. Besides, the pixel-level supervision in shadow detection focuses more on details than semantic information while the category-level supervision focuses more on semantic information. Thus, the network can expect a stronger feature extraction ability by combining shadow detection and shadow image classification.

Based on the above assumption and understanding, a shadow image classification branch is added to shadow detectors, as shown in Figure 3.

**FIGURE 2** Generated shadow images and real shadow images. Column 1–5 are real shadow images in ISTD, column 6 is the image generated by our ShadowGAN, column 7 is the shadow mask to guide the generation of shadow images

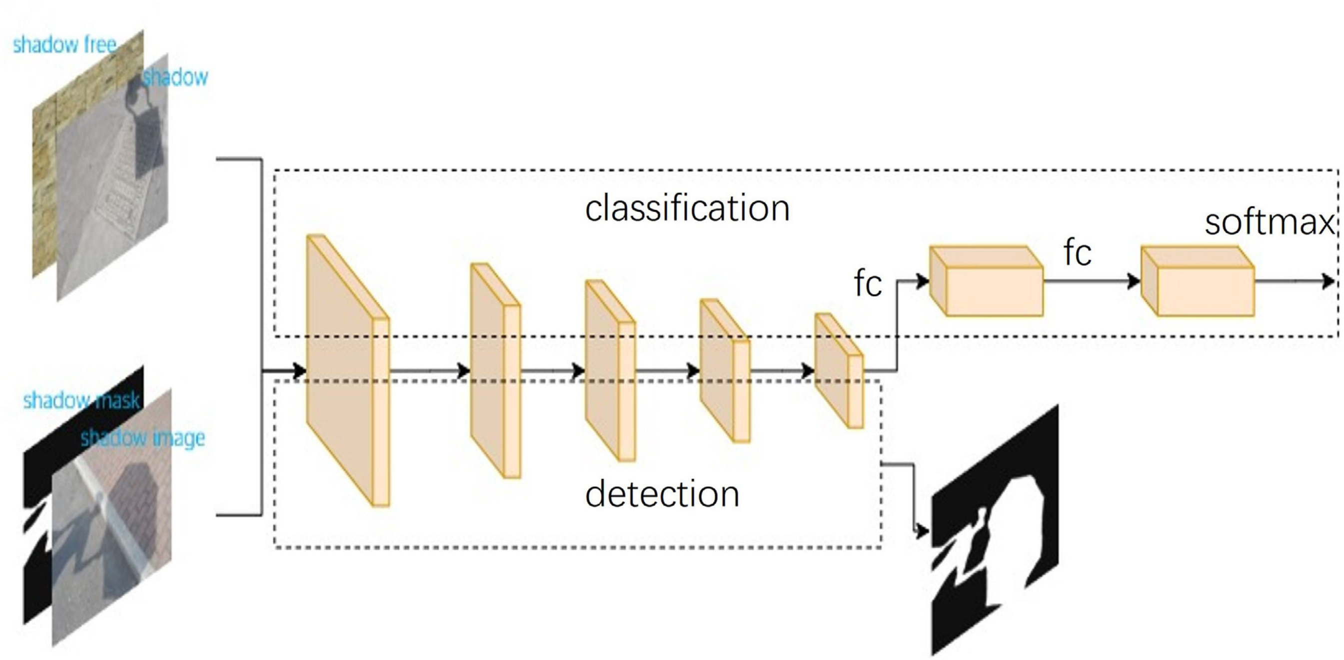


**FIGURE 3** The pipeline of ShadowGAN

The last layer of FPN contains useful semantic information of shadows, thus two fully connection layer are added behind this layer. The first fully connected layer is the original layer of ResNext101 [45] for 1000 category classification. It is followed by another fully connected layer for adapting this network to binary category (shadow and shadow-free) classification. We designed an end-to-end training strategy. The classification branch and detection branch share the same FPN. The detection branch is trained only once and the classification branch is trained twice in each round. Note that the FPN is active in the entire training process. During training, the classification branch is supervised by

$$\mathcal{L}_{cls} = -\left[y\log\tilde{y} + (1-y)\log(1-\tilde{y})\right] \tag{2}$$

where $y$ is the label of input image and $\tilde{y}$ is predicted label of categories. By denoting the loss function of shadow detecting branch as , the total loss can be computed as

$$\mathcal{L}_{total} = \lambda(i)\,\mathcal{L}_{det} + [1-\lambda(i)]\,\mathcal{L}_{cls}, \tag{3}$$

$$\lambda(x) = \begin{cases} 1, & x \bmod 3 = 0, \\ 0 & x \bmod 3 > 0. \end{cases} \tag{4}$$

where $i$ denotes current training iteration. It indicates that the parameters of feature extraction network will be updated by classification branch twice and updated by detection branch once in every three iterations.

An advantage of classification task is that it does not require pixel-level annotations, it only takes a glance to decide an image is shadow or non-shadow image. Our proposed shadow classification module is quite simple yet achieve a non-negligible improvement, as can be seen in Table 2. Such an improvement can be attributed to more robust shadow features extracted by feature extraction network. In other words, by combining classification module and shadow detection module, the feature extraction network goes even further in learning how to suppress non-shadow features, such as black objects, and enhance shadow features.

## 3.3 | Training of ShadowGAN and generation of shadow images

ISTD dataset [17] is adopted for training our network. ISTD is a dataset for shadow detection and shadow removal, it contains 1870 triplets including shadow images, shadow masks and shadow-free images. In ISTD dataset, 1330 triplets are assigned for training and the rest 540 triplets are assigned for testing.

We train our proposed ShadowGAN on the 1330 training triplets. After the model is trained, the discriminator is discarded and only the generator is tested to generate shadow images. The trained generator will infer on 1330 training triplets. That is, a shadow mask is randomly assigned to a shadow-free image, then the two images are fed into generator. Finally, the generator outputs a shadow image, the input shadow mask is the label of this shadow image. In this way, another 1330 annotated images with different shadows can be collected without manual labelling efforts.

**TABLE 1** BER of DSC and BDRAR trained on original dataset and enlarged dataset

| | Data augmentation | BER | PE | NE |
|---|---|---|---|---|
| DSC | × | 2.81 | 1.57 | 4.04 |
| | √ | 2.43 | 1.43 | 3.43 |
| BDRAR | × | 2.25 | 1.31 | 3.19 |
| | √ | 1.91 | 1.02 | 2.81 |

**TABLE 2** BER of DSC and BDRAR integrated with and without classification module

| | Cls | BER | PE | NE |
|---|---|---|---|---|
| DSC | × | 2.81 | 1.57 | 4.04 |
| | √ | 1.84 | 0.94 | 2.73 |
| BDRAR | × | 2.25 | 1.31 | 3.19 |
| | √ | 1.92 | 1.53 | 2.31 |

## 3.4 | End-to-end training of shadow classification and shadow detection

This section explains how to realize end-to-end training of shadow detection branch and shadow image classification branch. The whole network is divided into three parts, which are feature extraction network, shadow image classification branch and shadow detection branch. For simplicity, these three components are denoted as $F$, $C$ and $D$. In each training round, $C$ is trained twice, $D$ is trained only once. When $D$ is activated, shadow images and shadow masks will be fed into $F$, otherwise shadow-free images and shadow images will be fed into $F$ instead. It is worth noting that parameters of $F$ will be updated no matter whether $D$ or $C$ is activated. After the whole model is trained, $C$ is discarded during testing.

## 3.5 | Combination of ShadowGAN and shadow classification head

In this work, we propose ShadowGAN to augment the dataset and add shadow image classification task to the network. It is easy to implement this combination and does not require extra training tricks. First, we use ShadowGAN to generate 1330 shadow images and subsequently enlarge the shadow detection dataset. Then, we train shadow detection branch on the enlarged dataset and train shadow image classification branch on the original (real) shadow and shadow-free images. Experiments show that ShadowGAN and shadow image classification task are beneficial to shadow detection, as can be seen in Tables 1 and 2. Although these two methods can work independently, combining ShadowGAN and shadow image classification task can achieve a better performance, as can be seen in Table 4.

# 4 | EVALUATION

## 4.1 | Implementation details

All of our experiments are carried out on Pytorch1.2.0. DSC and BDRAR integrated with classification module are trained on four and two TiTAN V respectively.

During the training of Shadow detection module and ShadowGAN, default parameters are followed. To train shadow classification module, the batch-size is set to 32, the learning rate is 0.005.

Balanced error rate (BER) is a commonly-used metric for evaluating shadow detection results. It performs on detected shadow masks and corresponding ground-truth shadow masks. BER equally consider the accuracy on shadow regions and non-shadow regions. It can be computed by

$$BER = \left(1 - \frac{1}{2}\left(\frac{T_p}{N_p} + \frac{T_n}{N_n}\right)\right) \times 100\%, \tag{5}$$

where $T_p$, $N_p$, $T_n$ and $N_n$ are the number of correctly detected shadow pixels, the number of shadow pixels of ground truth, the number of correctly detected non-shadow pixels and the number of non-shadow pixels of ground truth, respectively. A lower BER indicates better shadow detection results. In the following experiments, results are rounded and accurate to two places of decimals. Accordingly, the positive error (PE) and negative error (NE) can be computed as

$$PE = \left(1 - \frac{T_p}{Np}\right) \times 100\%, \tag{6}$$

$$NE = \left(1 - \frac{T_n}{N_n}\right) \times 100\%. \tag{7}$$

## 4.2 | Evaluation of ShadowGAN

To evaluate the effectiveness of ShadowGAN, we train two recently proposed shadow detectors, DSC [13] and BDRAR [14], with default arguments on two datasets which are original 1,330 training pairs of ISTD and dataset enlarged by generated 1,330 images. These trained models are then tested on the 540 testing shadow images of ISTD. The results of BER are shown in Table 1.

As can be seen in the Table 1, both DSC and BDRAR gain a non-negligible improvement after training on enlarged datasets. Note that our aim is to demonstrate the effectiveness of data augmentation instead of comparing the performance of these two detectors.

## 4.3 | Failure case of ShadowGAN

Although images generated by ShadowGAN can improve the performance of shadow detection, from a subjective perspective, there are some failure cases. In Figure 4, the first column shows the generated shadows are lack of details. The second and third columns show that the generated shadows are unclear. The fourth column shows the loss of texture consistency between shadow and non-shadow regions. We attribute these failure cases to insufficient scenes in training set. Note that these images are also used to enlarge training set, no manual efforts are made to delete these images from generated samples.

**TABLE 3** Results of original BDRAR with batch size at 32

| Shadow detector | BER | PE | NE |
|---|---|---|---|
| BDRAR | 2.27 | 1.29 | 3.24 |

## 4.4 | Evaluation of shadow classification module

To evaluate the effectiveness of classification module, the classification module is integrated with DSC and BDRAR and then the modified shadow detectors are trained on ISTD dataset without data augmentation. We follow the above training strategy. Results are shown in Table 2.

In Table 2, the BER of BDRAR and DSC is improved by 0.33 and 0.97 after these two shadow detectors are integrated with classification module. It indicates that the detection module of DSC does a worse work in suppressing non-shadow features and enhancing shadow features. Although our proposed shadow classification module is very simple, it indeed forces the backbone to learn more robust shadow features.

As mentioned in the above training strategy, the batch-size for training the classification module is 32. To make it clear whether the improvement is contributed to classification module or a bigger batch size, we train original BDRAR (without classification module and data augmentation) by adjusting the batch size to 32 and learning rate to 0.02. The result is given in Table 3.

In fact, a bigger batch size does not mean a better shadow detection performance because the model will take the risk of overfitting [46].

For DSC, because it requires more GPUs to train the shadow detector with a batch size at 32, we omit this experiment. However, the classification module of DSC can be trained with a batch size at 32 because our classification module is light-weighted and 32 is just fine.

Another reason for the improvement is that our classification module can prevent the detection module from overfitting. To explicitly explain this, after each training iteration, the model will be validated on testing set with a batch size at 4. Note that parameters will not be updated during this phase. Training loss and validating loss of shadow detection module are shown in Figures 5, 6, 7 and 8.

In Figure 5, there is no distinct difference between the curves of training loss of DSC with and without classification module. This indicates that DSC learns well on the training set no matter it is integrated with classification module or not. However, the training of DSC without classification module tend to

**FIGURE 4** Failure cases of ShadowGAN

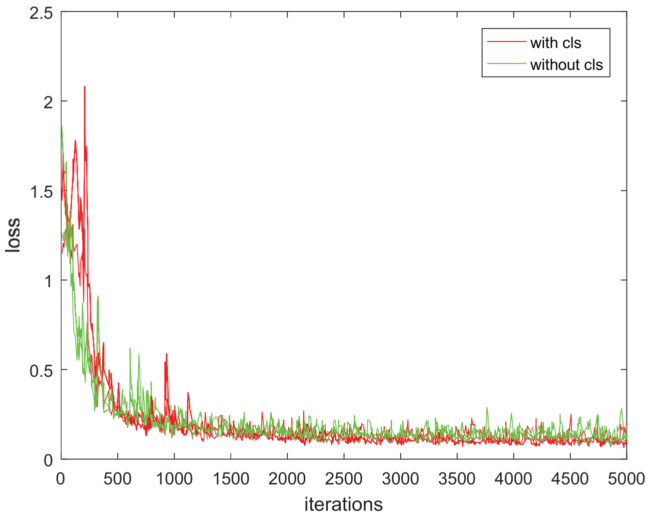


**FIGURE 5** Training loss of DSC

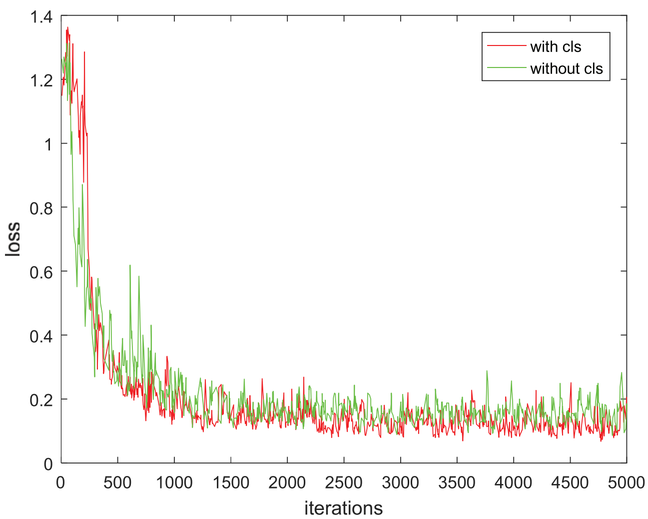


**FIGURE 6** Validation loss of DSC

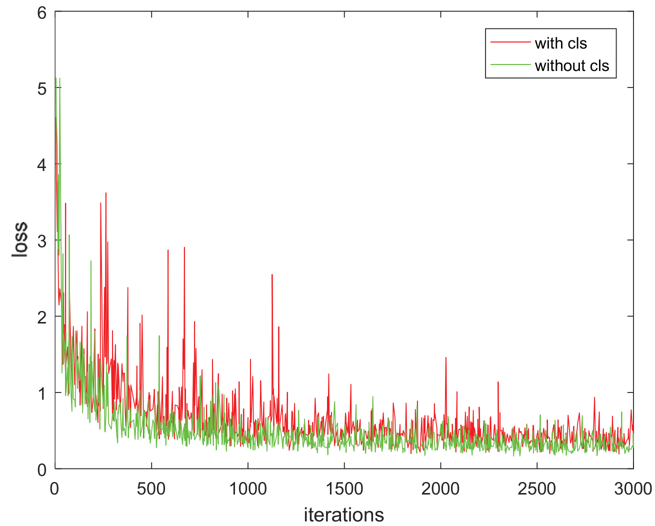


**FIGURE 7** Training loss of BDRAR

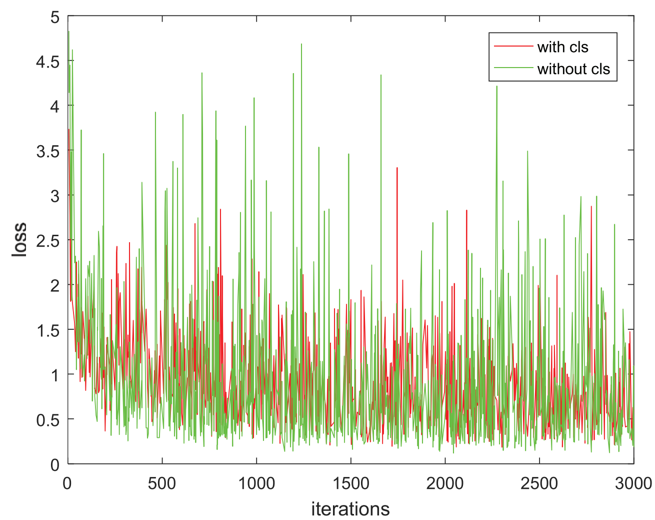


**FIGURE 8** Validation loss of BDRAR

overfitting. As can be seen in Figure 6, the curve of validation loss of DSC with classification module is under that of DSC without classification module because the classification module forces feature extraction network to learn more robust shadow features. As a result, classification module will prevent the detection module from overfitting after it is converged.

Similar to DSC, the training losses of BDRAR with and without classification module keep declining during training. However, the validation loss of BDRAR with classification module has a smaller lower bound than that of BDRAR without classification module. Thus, BDRAR with classification module can achieve a better performance.

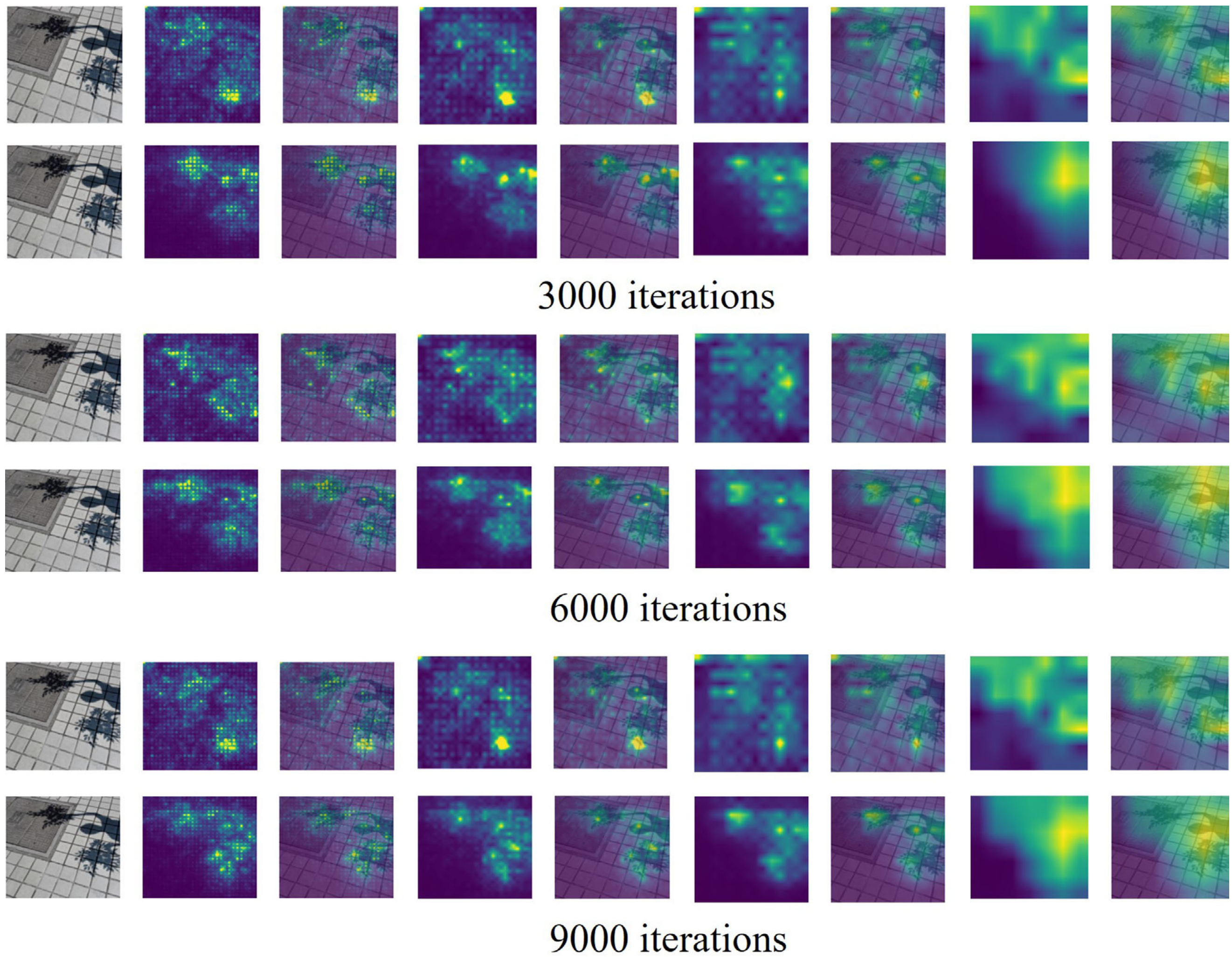


**FIGURE 9** Activation map of BDRAR trained by 9000 iterations

To visualize how classification branch effects results of shadow detection. We use LayerCAM [47] to visualize activation map at each layer. For comparison, BDRAR with/without classification are trained for 9000 iterations. Every 3000 iterations, LayerCAM is used to visualize the activation map.

In the activation maps of every 3000 iterations. The first row shows activation maps of BDRAR trained without classification branch, while the second row shows that with classification branch. In each row, the first column is input shadow image. In the remaining columns, every two columns denote the activation map and the original shadow image weighted by activation map of layer1, layer2, layer3, layer4 of ResNext.

Analogously, the activation maps of DSC with and without shadow image classification branch are shown as follows.

As mentioned above, layer1 concentrate on details, for example, boundaries of shadows and layer4 concentrate on semantic information, for example, the location of shadows. In these images, we can find that the feature extraction network fine-tuned by classification branch can learn better representations, especially semantic information, of shadows than that without classification branch.

## 4.5 | Evaluation of combination

We further carried out experiment on combining our shadow classification module and ShadowGAN. That is, the augmented dataset is used to train detection module and the classification module is trained on real shadow and non-shadow images. During testing phase, classification module is discarded, and the detector is tested on the testing set of ISTD. Results are given in Table 4.

Table 4 shows the shadow detection results of BDRAR and DSC trained on original training set and enlarged training set, with or without shadow classification module. As can be seen in Table 4, both ShadowGAN and shadow classification module can improve the performance of shadow detection. Besides, our ShadowGAN and shadow classification module are compatible and a better performance can be expected by combining these two methods.

In this section, we compare the quality of shadow images generated by our proposed ShadowGAN and Mask-ShadowGAN in [24]. Then we compare the performance of ShadowGAN with and without $\mathcal{L}_1$ loss. We slightly modified the code of

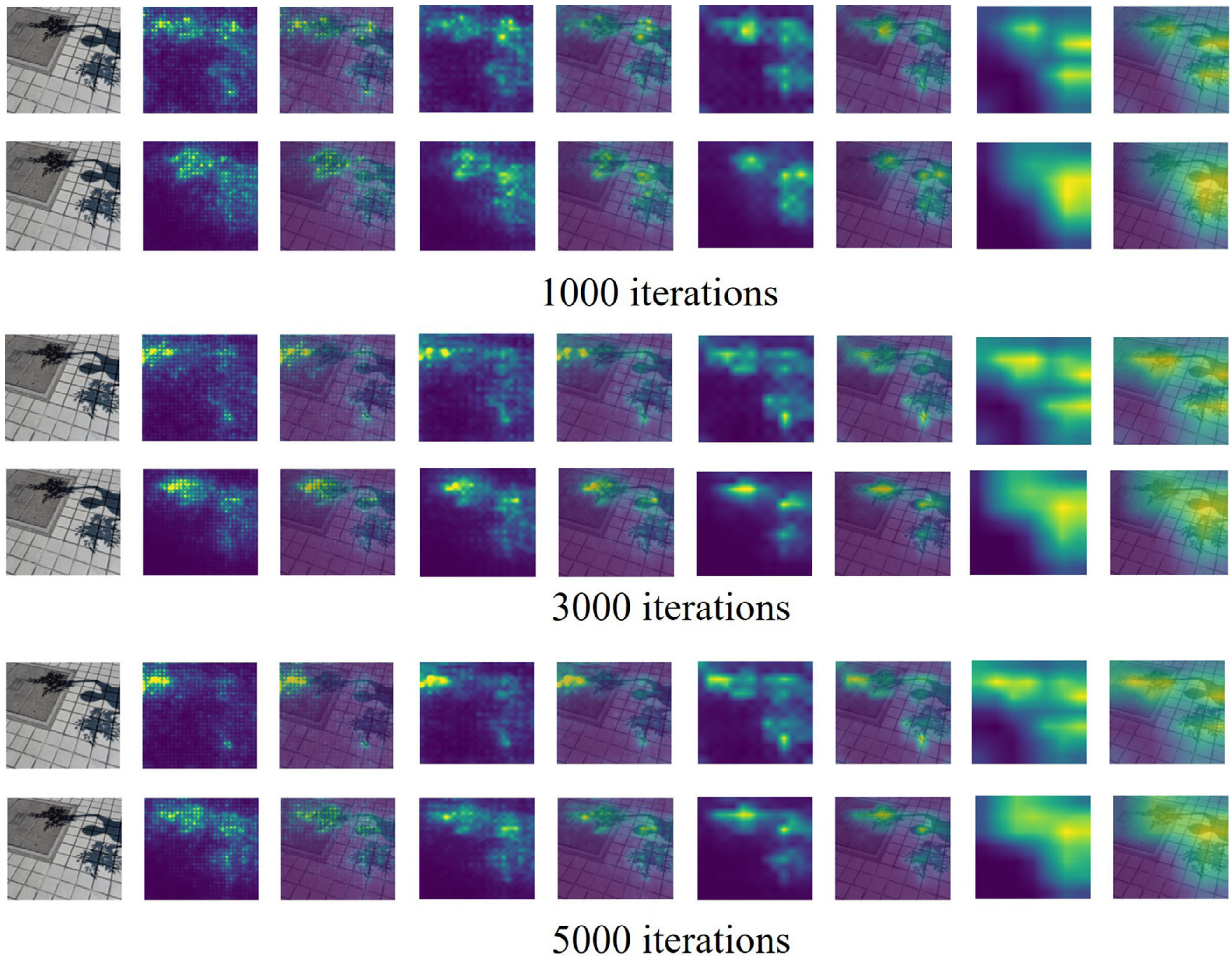


**FIGURE 10** Activation maps of feature extraction network with and without shadow image classification task under different iterations

**TABLE 4** Comparison of BER between shadow detectors with and without combination

| | Data augmentation | Cls | BER | PE | NE |
|---|---|---|---|---|---|
| BDRAR | × | × | 2.25 | 1.31 | 2.57 |
| | √ | × | 1.91 | 1.02 | 2.81 |
| | × | √ | 1.92 | 1.53 | 2.31 |
| | √ | √ | 1.87 | 1.65 | 2.09 |
| DSC | × | × | 2.81 | 1.57 | 4.04 |
| | √ | × | 2.43 | 1.43 | 3.43 |
| | × | √ | 1.84 | 0.94 | 2.73 |
| | √ | √ | 1.75 | 0.81 | 2.68 |

**TABLE 5** Ablation studies

| | Shadow GAN | L1 loss | Mask-Shadow GAN | BER | PE | NE |
|---|---|---|---|---|---|---|
| DSC | × | × | × | 2.81 | 1.57 | 4.04 |
| | × | × | √ | 2.57 | 1.24 | 3.89 |
| | √ | × | × | 2.38 | 1.43 | 3.33 |
| | √ | √ | × | 2.43 | 1.43 | 3.43 |
| BDRAR | × | × | × | 2.25 | 1.31 | 3.19 |
| | × | × | √ | 2.25 | 0.57 | 3.93 |
| | √ | × | × | 2.42 | 1.05 | 3.79 |
| | √ | √ | × | 1.91 | 1.01 | 2.80 |

Mask-ShadowGAN for comparison. That is, instead of generating shadow masks to guide the generation of shadow images, shadow masks are directly fed into the generator. Besides, paired images are used while training, and only the generator which transforms shadow-free images into shadow images is reserved while testing. The generated images are added into training set of ISTD. Then DSC and BDRAR without classification module are trained on these enlarged datasets. The results of BER are shown in Table 5.

In Table 5, for each method, the first row denotes the training set is not augmented, the second, third and fourth row

**TABLE 6** Comparison between different methods of data augmentation on BDRAR

| | Index | BER | PE | NE |
|---|---|---|---|---|
| ShadowGAN w/ L1 loss | 1 | 1.91 | 1.02 | 2.81 |
| | 2 | 1.81 | 1.42 | 2.20 |
| | 3 | 2.05 | 1.13 | 2.97 |
| | 4 | 2.62 | 0.57 | 4.69 |
| | 5 | 2.08 | 0.85 | 3.30 |
| | Average | **2.10** | 1.00 | 3.19 |
| ShadowGAN w/o L1 loss | 1 | 2.42 | 1.05 | 3.79 |
| | 2 | 2.59 | 1.13 | 4.05 |
| | 3 | 1.99 | 1.33 | 2.64 |
| | 4 | 2.21 | 1.00 | 3.42 |
| | 5 | 2.06 | 1.05 | 3.08 |
| | Average | 2.25 | 1.11 | 3.40 |
| Mask-ShadowGAN | 1 | 2.25 | 0.57 | 3.93 |
| | 2 | 2.14 | 0.59 | 3.69 |
| | 3 | 2.13 | 1.11 | 3.16 |
| | 4 | 2.45 | 0.96 | 3.94 |
| | 5 | 2.23 | 1.09 | 3.37 |
| | Average | 2.24 | 0.86 | 3.62 |
| naive L1 loss | 1 | 1.96 | 0.93 | 2.99 |
| | 2 | 2.04 | 0.89 | 3.18 |
| | 3 | 2.34 | 1.06 | 3.61 |
| | 4 | 2.35 | 0.81 | 3.89 |
| | 5 | 2.17 | 0.98 | 3.36 |
| | Average | 2.17 | 0.94 | 3.41 |

denote the training set is augmented by shadow images generated from Mask-ShadowGAN, ShadowGAN without $\mathcal{L}_1$ loss and our full ShadowGAN. The BER of DSC trained on dataset augmented by ShadowGAN without $\mathcal{L}_1$ loss is lower. As mentioned before, DSC can learn well on training set but have a bad performance on validation set. Thus, it is larger training set rather than higher quality of generated images that contributes to such an improvement. As for BDRAR, it learns more robust shadows on real shadow images and generated images with low quality will hamper its performance. Because shadow masks and shadow-free images are randomly selected during generation of shadow images, our 1330 new shadow images are randomly generated. For fair comparison, in this experiment, the generation of shadow images of each method is performed by five times and the shadow detector is trained on the enlarged dataset, note that in each round the same generator will generate different 1330 images. Because of training DSC is very time consuming, we only train BDRAR in this experiment. Results are given in Table 6.

Note that in the fourth experiment of ShadowGAN with $\mathcal{L}_1$ loss, the BER is 2.62. We think, as described in the former section, our ShadowGAN randomly assigned a mask to a shadow-free image and ShadowGAN accidentally generated more images of low quality.

## 5 | CONCLUSION

Shadow detection is an important pre-processing step of many computer vision tasks. Collecting training data for shadow detection is a challenging work because it requires pixel-level annotations. To augment existing shadow detection dataset without annotating shadow images manually, we propose ShadowGAN to generate new training samples from shadow-free images. Experiments of state-of-the-art shadow detectors trained on enlarged dataset show that ShadowGAN can augment existing dataset effectively and improve shadow detection results. To make full use of shadow-free images in shadow detection, we propose a light-weighted shadow classification module. This module helps feature extraction network to learn more robust shadow features. The classification module is light-weighted and easy to integrate with other shadow detectors. We believe future research will benefit from ShadowGAN and shadow classification module.

## ACKNOWLEDGEMENTS

This work was supported by the National Key Research and Development Program of China under Grant 2019YFC1511300; the National Natural Science Foundation of China under Grant 61971079; the Basic Research and Frontier Exploration Project of Chongqing under Grant cstc2019jcyj-msxmX0666 and the Innovative Group Project of the National Natural Science Foundation of Chongqing under Grant cstc2020jcyj-cxttX0002. Department of Science and Technology of Sichuan Province under Grant No. 2020YFQ0025.

## CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

*Lingyun Wen* https://orcid.org/0000-0002-9713-7366

## REFERENCES

1. Lalonde, J., Efros, A A., Narasimhan, S G.: Estimating natural illumination from a single outdoor image. 2009 IEEE 12th International Conference on Computer Vision. 183–190 (2009)
2. Wu, L., Gao, X., Foroosh, H.: Camera calibration and geo-location estimation from two shadow trajectories. Comput. Vision Image Understanding. 114(8), 915–927 (2010)
3. Paul, D.: Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques. 189–198 (1998)
4. Trivedi, M.M., Kogut, G.T., Cosman, P.C., et al.: Moving shadow and object detection in traffic scenes. Int J Automat Comput. 4, 38–46 (2007)
5. Cucchiara, R., Grana, C., Piccardi, M., et al.: Improving shadow suppression in moving object detection with HSV color information. 2001 IEEE Proceedings of Intelligent Transportation Systems. 334–339 (2001)
6. Guan, Ye-Peng.: Wavelet multi-scale transform based foreground segmentation and shadow elimination. Open Signal Process. J. 1 (2008)

7. Gao, F., Xu, Y., Ge, Y.: Property-based shadow detection and removal method for licence plate image. IET Image Process. 14, 1415–1425 (2020)
8. Wang, C., Xu, H., Zhou, L., et al.: Shadow Detection and Removal for Illumination Consistency on the Road. IEEE Trans. Intell. Veh. 5(4), 534–544 (2020)
9. Finlayson, G.D., Hordley, S.D., Chen, L., et al.: On the removal of shadows from images. IEEE Trans. Pattern Anal. Mach. Intell. 28(1), 59–68 (2006)
10. Finlayson, G.D., Drew, M.S., Chen, L.: Entropy minimization for shadow removal. Int. J. Comput. Vision. 85(1), 35–57 (2009)
11. Guo, R., Dai, Q., Hoiem, D.: Single-image shadow detection and removal using paired regions. Comput. Vision Pattern Recogn. 2033–2040 (2011)
12. Vicente, T., Hoai, M., Samaras, D.: Leave-one-out kernel optimization for shadow detection and removal. IEEE Trans. Pattern Anal. Mach. Intell. 40(3), 682–695 (2018)
13. Hu, X., Zhu, L., Fu, C., et al.: Direction-aware spatial context features for shadow detection. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 7454–7462 (2018)
14. Zhu, L., Deng, Z., Hu, X., et al.: Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. European Conference on Computer Vision. 122–137 (2018)
15. Zheng, Q., Qiao, X., Cao, Y., et al.: Distraction-aware shadow detection. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 162–5171 (2019)
16. Chen, Z., Zhu, L., Wan, L., et al.: A multi-task mean teacher for semi-supervised shadow detection. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 5610–5619 (2020)
17. Wang, J., Li, X., Yang, J.: Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 1788–1797 (2018)
18. Vicente, T., Hou, L., Yu, C.P., et al.: Large-scale training of shadow detectors with noisily-annotated shadow examples. European Conference on Computer Vision. 816–832 (2016)
19. Goodfellow, I.J., Pouget-Abadie, J., Jean, P.A., et al.: Generative adversarial nets. Proceedings of the 27th International Conference on Neural Information Processing Systems. 2, 2672–2680 (2014)
20. Isola, P., Zhu, J., Zhou, T.: Image-to-image translation with conditional adversarial networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 5967–5976 (2017)
21. Zhu, J., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. 2017 IEEE International Conference on Computer Vision (ICCV). 2242–2251 (2017)
22. Kim, T., Cha, M., Kim, H., et al.: Learning to discover cross-domain relations with generative adversarial networks. Proceedings of the 34th International Conference on Machine. 70, 1857–1865 (2017)
23. Yi, Z., Zhang, H., Tan, P.: DualGAN: Unsupervised dual learning for image-to-image translation. 2017 IEEE International Conference on Computer Vision (ICCV). 2868–2876 (2017)
24. Hu, X., Jiang, Y., Fu, C., et al.: Mask-ShadowGAN: Learning to remove shadows from unpaired data. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). 2472–2481 (2019)
25. Zhu, J., Samuel, G.G., Masood, S.Z., et al.: Learning to recognize shadows in monochromatic natural images. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 223–230 (2010)
26. Huang, X., Hua, G., Tumblin, J., et al.: What characterizes a shadow boundary under the sun and sky? 2011 International Conference on Computer Vision. 898–905 (2011)
27. Guo, R., Dai, Q., Hoiem, D.: Single-image shadow detection and removal using paired regions. Comput. Vision Pattern Recogn. 2033–2040 (2011)
28. Khan, S.H., Bennamoun, M., Sohel, F., et al.: Automatic shadow detection and removal from a single image. IEEE Trans. Pattern Anal. Mach. Intell. 38(3), 431-446 (2016)
29. Shen, L., Chua, T.W., Karianto, L.: Shadow optimization from structured deep edge detection. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2067–2074 (2015)
30. Nguyen, V., Vicente, T., Zhao, M., et al.: Shadow detection with conditional generative adversarial networks. 2017 IEEE International Conference on Computer Vision (ICCV). 4520–4528 (2017)
31. Lin, T.Y., Dollar, P., Girshick, R., et al.: Feature pyramid networks for object detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 936–944 (2017)
32. Guo, R., Ayinde, B., Sun, H.: Efficient shadow detection and removal using synthetic data with domain adaptation. 2020 25th International Conference on Pattern Recognition (ICPR). 5867–5874 (2021)
33. Zhang, S., Liang, R., Wang, M.: ShadowGAN: Shadow synthesis for virtual objects with conditional adversarial networks. Comput. Visual Media. 5(01), 106–116 (2019)
34. Liu, D., Long, C., Zhang, H., et al.: ARShadowGAN: Shadow generative adversarial network for augmented reality in single light scenes. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 8136–8145 (2020)
35. Liu, Z., Yin, H., Wu, X., et al.: From shadow generation to shadow removal. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 4927–4936 (2021)
36. Zhong, Z., Zheng, L., Zheng, Z., et al.: Camera style adaptation for person re-identification. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 5157–5166 (2018)
37. Zhu, Y., Zhang, Y., Zhang, H., et al.: Data augmentation of x-ray images in baggage inspection based on generative adversarial networks. IEEE Access. 8, 86536–86544 (2020)
38. Zhang, H., Goodfellow, I., Metaxas, D., et al.: Self-attention generative adversarial networks. arXiv. (2018). Available from: https://arxiv.org/abs/1805.08318v1
39. Huang, S.W., Lin, C.T., Chen, S.P., et al.: AugGAN: Cross domain adaptation with GAN-based data augmentation. European Conference on Computer Vision. 731–744 (2018)
40. Ren, S., He, K., Girshick, R., et al.: Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39(6), 1137–11149 (2017)
41. Redmon, J., Divvala, S., Girshick, R., et al.: You only look once: Unified, real-time object detection. Comput. Vision Pattern Recogn. 779–788 (2016)
42. Le, H., Vicente, T., Nguyen, V., et al.: A+D net: Training a shadow detector with adversarial shadow attenuation. arXiv 680–696 (2018). Available from: https://arxiv.org/abs/1712.01361
43. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Instance normalization: The missing ingredient for fast stylization [internet]. arXiv. (2016). Available from: https://arxiv.org/abs/1607.08022
44. Akcay, S., Atapour-Abarghouei, A., Breckon, T.P.: GANomaly: Semi-supervised anomaly detection via adversarial training. Asian Conference on Computer Vision. 622–637 (2019)
45. Xie, S., Girshick, R., Dollár, P., et al.: Aggregated residual transformations for deep neural networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 5987–5995 (2017)
46. Goyal, P., Dollár, P., Girshick, R., et al.: Accurate, large minibatch SGD: Training ImageNet in 1 hour. arXiv. (2017). Available from: https://arxiv.org/abs/1706.02677v1
47. Jiang, P.T., Zhang, C.B., Hou, Q., et al.: LayerCAM: Exploring hierarchical class activation maps. IEEE Trans. Image Process. 30, 5875–5888 (2021)


**How to cite this article:** Li, G., Wen, L., Huang, Z., Xia, R., Pang, Y.: Data augmentation and shadow image classification for shadow detection. IET Image Process. 1–12 (2021). https://doi.org/10.1049/ipr2.12377