



Integration of GAN and Adaptive Exposure Correction for Shadow Removal

Krunal Mehta, Manish Khare^(✉), and Avik Hati

Dhirubhai Ambani Institute of Information and Communication Technology,
Gandhinagar, Gujarat, India
{202011051,manish_khare}@daiict.ac.in

Abstract. Shadow removal from images and videos is an essential task in computer vision that concentrates on detecting the shadow generated by the obstructed light source, and obtains realistic shadow-free results. In this paper, we present a method based on generative adversarial networks (GANs) for shadow removal by supervised learning. Specifically, we train two generators and two discriminators to learn the mapping between shadow and shadow-free image domains. We employ generative adversarial constraints with cycle consistency and content constraints to learn the mapping efficiently. We also propose an adaptive exposure correction module to handle the over-exposure problem in the shadow area of the result. We additionally present a method for improving the quality of benchmark datasets and eventually achieving better shadow removal results. We also show ablation studies to analyze the importance of the ground-truth data with the adaptive exposure correction module in the proposed framework and explore the impact of using different learning strategies in the presented method. We validate the approach on the available large-scale benchmark Image Shadow Triplets dataset (ISTD), and show quantitative and visual improvements in the state-of-the-art results.

Keywords: Shadow removal · Shadow detection · GANs · Adaptive exposure correction · Benchmark dataset adjustment

1 Introduction

Shadow detection and removal is a fundamental and challenging task in computer vision and computer graphics. In an image, a shadow is a direct result of occluding a light source. The accuracy of several computer vision tasks, such as object segmentation [20], object recognition [2], and object tracking [13], can be influenced by the shadow since shadows have similar characteristics as objects, so they can get misclassified as part of an object.

In computer vision, the problem involving shadow detection and removal has received much attention. Early works related to this task [5, 6, 21, 26, 28, 29] used physical models of features like intensity, color, gradient, and texture. However, these hand-crafted feature-based methods suffer in understanding the high-level features and related semantic content. In recent years, deep learning-based

approaches for analyzing the mapping relation have made significant progress in this field. Khan *et al.* [11, 12] used convolutional neural networks (CNN) for shadow detection and Bayesian model for shadow removal. The model of Qu *et al.* [19] is based on an end-to-end multi-context embedding framework to extract essential characteristics from multiple aspects and accumulate them to determine the shadow matte. Fan *et al.* [4] employed a deep CNN structure containing an encoder-decoder and a refinement model for extracting features with local detail correction and learning the shadow matte. Bansal *et al.* [3] developed a deep learning model to extract features and directly detect the shadow mask. Hu *et al.* [7] presented a direction-aware spatial context (DSC) module, utilized with CNNs, to detect and remove the shadow.

The generative adversarial network (GAN) [1] and its extensions, presented in recent years, are dominant strategies for dealing with diverse image-to-image translation challenges. Conditional GANs (CGANs) [15] are significant GAN extensions that incorporate conditioning information into the generator and the discriminator. Nguyen *et al.* [17] demonstrated the first method of shadow detection with adversarial learning and constructed a CGAN-based architecture to output a shadow mask that can realistically correspond to the ground-truth mask. A shadow image with an adjustable sensitivity factor is used as the conditioning information to the generator and the discriminator. Wang *et al.* [27] presented a supervised model based on two Stacked-CGANs to tackle shadow detection and removal problems simultaneously in an end-to-end manner. Nagae *et al.* [16] developed a model based on the method in [27], with minor changes in the shadow removal CGAN, that estimates the illumination ratio and uses that estimation to produce the output. Although these approaches [7, 27] effectively remove the shadow, they tend to generate artifacts and inconsistent colors in the non-shadow area. Hu *et al.* [8] presented a Mask-ShadowGAN framework that enforces cycle consistency by the guidance of masks and learns a bidirectional mapping between the shadow and shadow-free domains. Tan *et al.* [24] developed a target-consistency GAN (TCGAN) for shadow removal that aims to learn a unidirectional mapping to translate shadow images into shadow-free images. These methods [8, 24] remove the shadow by maintaining a non-shadow region with cycle and target consistency but suffer from overexposure problems and random artifacts. Also, they require unpaired shadow and shadow-free datasets with the same statistical distribution for better learning.

In this paper, we propose a novel method based on GANs with cycle constraints, and introduce an adaptive exposure correction module for handling the overexposure problem. Figure 1 shows a shadow removal result of the proposed method compared with Mask-ShadowGAN [8], which suffers from over-exposure, particularly in the shadow area. However, our approach handles that problem and generates a result close to the ground-truth. The key contributions of this work are as follows.

- We present a framework that removes the shadow using *generative adversarial* constraints along with *cycle consistency* and *content* constraints.

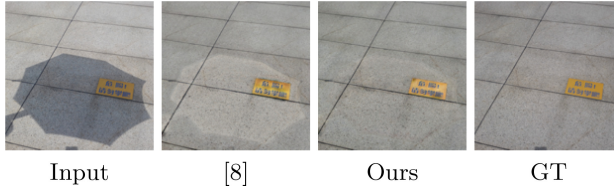


Fig. 1. Shadow removal results comparing the Mask-ShadowGAN [8] method with the proposed method.

- We introduce an adaptive exposure correction module for handling the over-exposure problem.
- We introduce a method for enhancing the quality of benchmark datasets and subsequently improving the shadow removal results.

The rest of the paper is organized as follows. Section 2 describes the proposed framework. Section 3 presents experimental results along with the ablation study, and we conclude the work in Sect. 4.

2 Proposed Method

The overall scheme of the proposed method is depicted in Fig. 2. The method is based on CycleGAN [30], in which each adversarial generator learns a mapping to another domain, and the corresponding discriminator guides the learning procedure. Apart from the adversarial and cycle constraints, we also employ content and identity constraints as guidance for better learning. Compared to the baseline Mask-ShadowGAN [8], which required unpaired data with an equal statistical distribution of shadow and shadow-free domains, our method utilizes available shadow, shadow-free, and shadow mask images to learn better mapping for shadow removal.

2.1 Generator and Discriminator Learning

The proposed method learns from both the shadow domain \mathbb{D}_x and the shadow-free domain \mathbb{D}_y . While learning from domain \mathbb{D}_x , the generator network G_f takes a real shadow image $I_s \in \mathbb{D}_x$ as input, and generates a shadow-free image \hat{I}_{f*} . The discriminator network D_f is used to differentiate whether the produced shadow-free image \hat{I}_{f*} is a real shadow-free image or not. To achieve the cycle-consistency, another generator G_s is used to reconstruct the shadow image \hat{I}_s from the generated shadow-free image \hat{I}_{f*} using a ground-truth shadow mask M_{gt*} for the image I_s as a guide.

In the process of learning from the shadow-free domain \mathbb{D}_y , the generator network G_s takes a real shadow-free image $I_f \in \mathbb{D}_y$ as input and a ground-truth shadow mask M_{gt} for the image I_f as a guide, and generates a shadow image \hat{I}_{s*} .

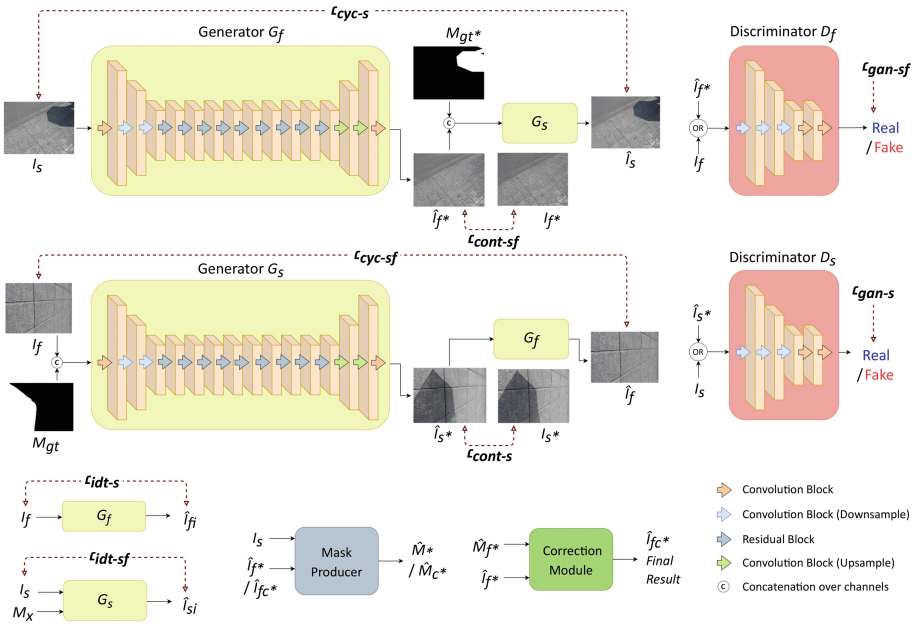


Fig. 2. Illustration of the architecture of the proposed method.

The discriminator network D_s determines if the created shadow image \hat{I}_{s*} is a real shadow image or not. To formulate the cycle-consistency loop, the generator G_f reconstructs the shadow-free image \hat{I}_f from the generated shadow image \hat{I}_{s*} .

To summarize, the discriminator network D_s takes either real sample I_s or fake sample \hat{I}_{s*} as input and discriminates whether the input is from \mathbb{D}_s or not. Similarly, discriminator D_f takes either real sample I_f or fake sample \hat{I}_{f*} as input and discriminates whether the input is from \mathbb{D}_f or not. We shall discuss the corresponding loss functions in Sect. 2.3.

2.2 Adaptive Exposure Correction Module

Given a shadow image, the generator network G_f is trained to produce a shadow-free image. But in the absence of any constraints, sometimes the generated shadow-free images are much brighter in the shadow area. To handle this over-exposure problem in the resulting shadow-free images, we propose to use an adaptive exposure correction module that takes the generated shadow-free image \hat{I}_{f*} and an intermediate shadow-free mask \hat{M}_* as inputs, and produces the final shadow-free result \hat{I}_{fc*} . The shadow mask \hat{M}_* for the input shadow image I_s is obtained as $\mathbb{B}(\hat{I}_{f*} - I_s, t)$, where the binarization operation \mathbb{B} is performed on the difference between \hat{I}_{f*} and the real input shadow image I_s , and t is a threshold obtained by Otsu’s algorithm [18]. \mathbb{B} sets the value as zero or one, where zero indicates non-shadow region (difference $\leq t$) and one indicates the

shadow region (difference $> t$). In the adaptive exposure correction module, we extract the shadow and non-shadow areas using \hat{M}_* and apply gamma correction (power-law transformations) in the shadow area. First, we transform the extracted shadow area to the HSV color space, then perform gamma correction on the value channel and convert it back to RGB color space. Finally, we combine the gamma-corrected shadow area with the non-shadow area to generate \hat{I}_{fc*} which is the final shadow-free image with exposure correction. To estimate the gamma value, we calculate the mean difference between the shadow and non-shadow areas and map that to the gamma value range 0 to 2. Ideally, for a non-overexposed image, the gamma value will be 1, and no correction will be done. Then the final shadow mask \hat{M}_{c*} is obtained as $\mathbb{B}(\hat{I}_{fc*} - I_s, t)$.

2.3 Objectives and Loss Functions

Adversarial Losses: The primary principle behind adversarial learning is that the discriminator will differentiate between real and generated results for both domains, encouraging the corresponding generator to deliver a better output concerning image qualities. The shadow-free adversarial loss and the shadow adversarial loss are given as:

$$\mathcal{L}_{gan-sf(G)} = MSE(P, D_f(\hat{I}_{f*})), \mathcal{L}_{gan-s(G)} = MSE(P, D_s(\hat{I}_{s*})) \quad (1)$$

$$\begin{aligned} \mathcal{L}_{gan-sf(D)} &= MSE(P, D_f(I_f)) + MSE(Q, D_f(\hat{I}_{f*})), \\ \mathcal{L}_{gan-s(D)} &= MSE(P, D_s(I_s)) + MSE(Q, D_s(\hat{I}_{s*})) \end{aligned} \quad (2)$$

where \hat{I}_{f*} (generated as $(G_f(I_s))$) and \hat{I}_{s*} (generated as $(G_s(I_f, M_{gt}))$) are the generated shadow-free and shadow images, respectively, with I_s and I_f being the input shadow and shadow-free images, respectively, and $P = 1, Q = 0$.

Cycle Consistency Losses: Cycle consistency L_1 losses defined in Eq. (3) and Eq. (4) are applied to encourage the reconstructed images to be comparable to the original input images and to effectively improve the bidirectional mapping in the G_f and G_s networks.

$$\mathcal{L}_{cyc-s} = \|\hat{I}_s - I_s\|_1 \quad (3)$$

$$\mathcal{L}_{cyc-sf} = \|\hat{I}_f - I_f\|_1 \quad (4)$$

Here, \hat{I}_s (generated as $G_s(G_f(I_s), M_{gt*})$) and \hat{I}_f (generated as $G_f(G_s(I_f, M_{gt}))$) are the reconstructed shadow and shadow-free images, respectively.

Identity Losses: The identity L_1 losses defined in Eq. (5) and Eq. (6) motivate generators G_s and G_f not to change the input image (a shadow image and a shadow-free image, respectively), and maintain color consistency.

$$\mathcal{L}_{idt-s} = \|\hat{I}_{si} - I_s\|_1 \quad (5)$$

$$\mathcal{L}_{idt-sf} = \|\hat{I}_{fi} - I_f\|_1 \quad (6)$$

where \hat{I}_{si} is the generated image using G_s from I_s and null mask M_x , and \hat{I}_{fi} is the generated image using G_f from I_f .

Content Losses: The L_1 constraint on content losses defined in Eq. (7) and Eq. (8) encourages generators to produce images that are closer to the ground-truth images.

$$\mathcal{L}_{cont-s} = \|\hat{I}_{s*} - I_{s*}\|_1 \quad (7)$$

$$\mathcal{L}_{cont-sf} = \|\hat{I}_{f*} - I_{f*}\|_1 \quad (8)$$

Here, I_{s*} and I_{f*} are the ground-truth shadow and shadow-free images, respectively, and \hat{I}_{f*} (generated as $(G_f(I_s))$) and \hat{I}_{s*} (generated as $(G_s(I_f, M_{gt}))$) are the generated shadow-free and shadow images, respectively.

Loss Function for Generators: The total generator loss for the proposed method is obtained as a weighted sum of the adversarial losses, cycle consistency losses, identity losses, and content losses, given as:

$$\mathcal{L}_G = \lambda_1(\mathcal{L}_{gan-s(G)} + \mathcal{L}_{gan-sf(G)}) + \lambda_2(\mathcal{L}_{cyc-s} + \mathcal{L}_{cyc-sf}) + \lambda_3(\mathcal{L}_{idt-s} + \mathcal{L}_{idt-sf}) + \lambda_4(\mathcal{L}_{cont-s} + \mathcal{L}_{cont-sf}) \quad (9)$$

where $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are appropriately chosen weights.

Loss Function for Discriminators: The discriminator loss for the shadow-free discriminator D_f and shadow discriminator D_s in the proposed method are given in Eq. (10) and Eq. (11), respectively.

$$\mathcal{L}_{D_f} = \lambda_5(\mathcal{L}_{gan-sf(D)}) \quad (10)$$

$$\mathcal{L}_{D_s} = \lambda_5(\mathcal{L}_{gan-s(D)}) \quad (11)$$

Here, λ_5 is the appropriately chosen weight.

2.4 Network Architecture and Training Strategy

We use the model of Johnson *et al.* [10] as the generator network, which consists of 3 convolutional layers, 9 residual blocks, and 2 deconvolution layers. After each convolution and deconvolution operation, the network employs instance normalization and the ReLU (rectified linear unit) activation function. For the discriminator network, we use PatchGAN [9], which focuses on classifying image patches as real or fake. Here, 4 convolutional layers are used with instance normalization and leaky ReLU activation function (slope = 0.2). Adam optimization [14] with a learning rate of 0.0002, with first and second order momentum as 0.5 and 0.999, is adopted during training. A zero-mean Gaussian distribution with a standard deviation of 0.02 initializes the network parameters. For data augmentation during training, images are resized to 286×286 and randomly cropped to 256×256 . The network is trained for 200 epochs keeping the mini-batch size as 1 with the PyTorch module and NVIDIA GeForce-RTX2080-Ti GPU.

2.5 Benchmark Dataset Adjustment

Ideally, in the benchmark dataset for the shadow removal task, the non-shadow area of the shadow and the corresponding shadow-free image should be the same. However, there is a significant difference in the color consistency, brightness, and contrast, since both shadow and shadow-free images were captured at different times of the day. On the whole testing dataset of ISTD [27], the root mean square error (RMSE) in the LAB color space between the shadow and shadow-free images in the non-shadow area is 6.83, which should ideally be close to 0. Figure 3 shows the sample triplets from the ISTD dataset, where the difference in the non-shadow area is clearly visible. Supervised models are trained to produce an output close to the ground-truth shadow-free image, and accordingly, the loss function is defined, and models are trained. However, methods yield color, brightness, and contrast inconsistent outputs compared to the non-shadow area of the shadow image. Hence, it is essential to adjust those ground-truth shadow-free images to achieve better results.

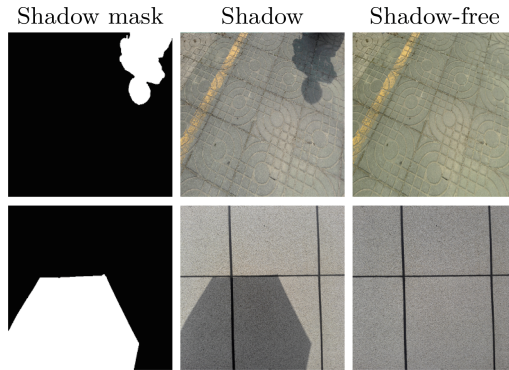


Fig. 3. ISTD triplets, showing issue in the non-shadow area.

To achieve this, we process each image individually to adjust the ground-truth shadow-free images using the regression technique. Following are the steps we used for this correction task.

- The non-shadow area of shadow and shadow-free images were extracted using the shadow mask.
- A regressor makes use of that extracted non-shadow area and learns to transform the non-shadow pixel values of shadow-free image into the corresponding pixel values of the shadow image.
- Finally, the trained regressor takes the shadow-free image as input and generates an adjusted shadow-free image.

We conducted various experiments by using three well-known regressors, Linear Regressor (LR), Decision-Tree Regressor (DTR), and K-Nearest-Neighbor

Regressor (KNNR). Further, we considered both RGB and LAB color spaces. Also, we executed experiments by using single-output regression, where regression is performed on three individual color channels, and by using multi-output regression, where regression is performed on three combined color channels. Finally, we used the optimal decision-tree multi-output regressor in RGB color space for the benchmark dataset adjustment. Following steps describe the algorithm of the decision-tree regressor.

- Given a training vector x and a label vector y , the decision tree divides the feature space in a recursive fashion, such that the samples with similar labels are grouped together.
- Let the data at node n be denoted by D_n having m_n samples. For each candidate split $\delta = (i, t_n)$, where i is feature and t_n is threshold, partition the data into $D_n^{left}(\delta)$ and $D_n^{right}(\delta)$ subsets according to following equations.

$$D_n^{left}(\delta) = \{(x, y) | x_i \leq t_n\} \quad (12)$$

$$D_n^{right}(\delta) = \{(x, y) | x_i > t_n\} \quad (13)$$

- The quality of a candidate split of node n is then measured using an impurity function G and loss function H according to Eq. (14) and Eq. (15), respectively. Here, \bar{y}_n is the mean value, and the mean squared error is used as the loss function.

$$H(D_n) = \frac{1}{m_n} \sum_{y \in D_n} (y - \bar{y}_n)^2, \bar{y}_n = \frac{1}{m_n} \sum_{y \in D_n} y \quad (14)$$

$$G(D_n, \delta) = \frac{m_n^{left}}{m_n} H(D_n^{left}(\delta)) + \frac{m_n^{right}}{m_n} H(D_n^{right}(\delta)) \quad (15)$$

- Parameters that minimize the impurity are selected for splitting, as follows:

$$\delta^* = \arg \min_{\delta} G(D_n, \delta) \quad (16)$$

- The algorithm is recursed for subsets $D_n^{left}(\delta^*)$ and $D_n^{right}(\delta^*)$ until $m_n = 1$.

3 Experimental Results

Database Description: To analyze the performance of the proposed framework, we experimented with the dataset containing image shadow triplets termed as ISTD [27] and trained models accordingly. ISTD contains 1870 triplets of shadow, shadow mask, and shadow-free image with 1330 image triplets in the training split and 540 in the testing split.

Evaluation Parameters: We followed [17, 25, 27] and used balance error rate (BER) for a quantitative comparison for shadow detection. Balance error rate is calculated as:

$$\text{BER} = 1 - \frac{1}{2} \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right) \quad (17)$$

where

- True Positive (TP) denotes the number of pixels that the predictive model has labeled as a shadow, and actually, it is a shadow.
- False Positive (FP) denotes the number of pixels that the predictive model has labeled as a shadow, and actually, it is a non-shadow.
- True Negative (TN) denotes the number of pixels that the predictive model has labeled as a non-shadow, and actually, it is a non-shadow.
- False Negative (FN) denotes the number of pixels that the predictive model has labeled as a non-shadow, and actually, it is a shadow.

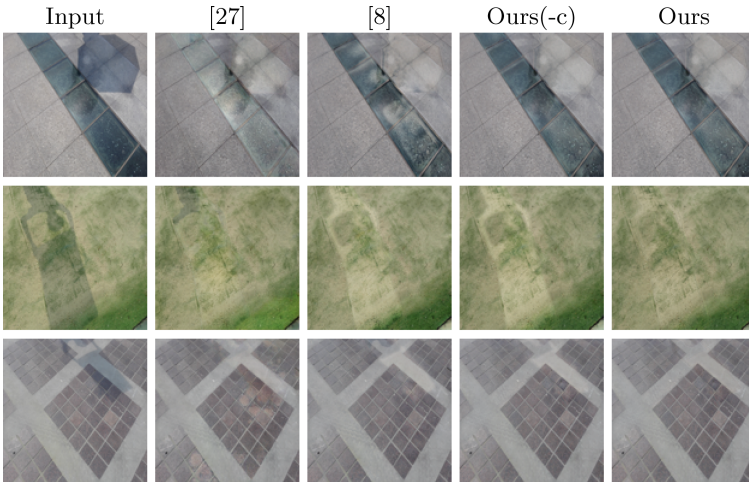
For the quantitative assessment of shadow removal, we followed recent procedures [7, 8, 24, 27] and used root mean square error (RMSE) in the LAB color space computed between the ground-truth and produced shadow-free images. We resized all images to 286×286 for a fair comparison. Additionally, we calculated the RMSE value in the four scenarios: RMSE value by comparing the resulting shadow-free image \hat{I}_{fc*} with the ground-truth shadow-free image I_{f*} (i) for all pixels (represented with ‘O’), (ii) for pixels in the shadow region (represented with ‘S’), (iii) for pixels in the non-shadow region (represented with SF), and (iv) by comparing \hat{I}_{fc*} with input shadow image I_s for pixels in the non-shadow region (represented with SF-I). In the experiments, the hyper-parameters $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5$ are set as 1, 10, 5, 5, 0.5, respectively. In the tables, best and second-best results are highlighted in bold and blue, respectively.

Evaluation on Removal: We compare the shadow removal performance of the proposed method with the methods in [5–8, 24, 27, 28] on the test dataset of ISTD. The results are shown in Table 1. Our method achieves the best performance in the O and SF scenarios, and the second-best performance in S and SF-I scenarios. TCGAN [24] achieves the best result in SF-I, but it has poor performance in S. Similarly, DSC [7] achieves the best result in S but performs poorly in SF and SF-I. Our approach achieves comparable results in all aspects and yields the best overall value O, compared to other methods. Figure 4 shows visual performance compared to methods ST-CGAN [27] and Mask-ShadowGAN [8]. While ST-CGAN [27] suffers from color-inconstancy and artifacts, and Mask-ShadowGAN [8] has over-exposure, our approach handles those issues and produces better output.

Evaluation on Detection: We evaluate the shadow detection performance with the recent methods [8, 14, 17, 27] on the ISTD test dataset. The quantitative results are shown in Table 2. The proposed method outperforms the baseline Mask-ShadowGAN [8] and methods CGAN [17], StackedCNN [25]. Methods SCGAN [17] and ST-CGAN [27] achieve better results since these methods

Table 1. Quantitative results of removal with RMSE on ISTD test dataset.

Method	Publication	O	S	SF	SF-I
Original	-	10.97	32.67	6.83	0
Yang [28]	IEEE TIP, 2012	15.63	19.82	14.83	-
Gong [5]	BMVC, 2014	9.3	18.95	7.46	-
Guo [6]	IEEE TPAMI, 2013	8.53	14.98	7.29	-
ST-CGAN [27]	IEEE CVPR, 2018	7.47	10.33	6.93	7.45
Mask-ShadowGAN [8]	IEEE, ICCV, 2019	6.99	11.41	6.17	6.75
TCGAN [24]	ARXIV, 2020	6.85	11.49	5.91	6.29
DSC [7]	IEEE TPAMI, 2020	6.67	9.22	6.39	6.61
Ours	-	6.54	10.03	5.88	6.49

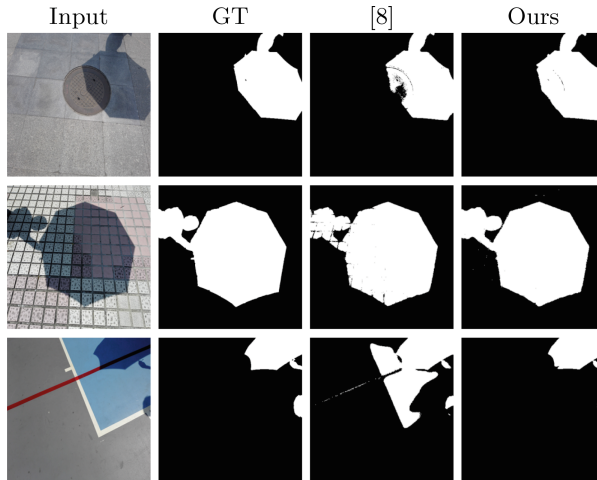
**Fig. 4.** Visual comparison of shadow removal results of ISTD test dataset.

specifically train their networks for the detection task. As our goal is shadow removal, we do not train any separate network for detection; instead, we extract the shadow mask from the final shadow-free image and input image as discussed in Sect. 2.2. Figure 5 shows the visual performance compared to state-of-the-art Mask-ShadowGAN [8]. Our approach produces a shadow mask result close to the ground-truth shadow mask.

Benchmark Dataset Adjustment: To adjust ground-truth shadow-free images, we experimented with Linear Regressor (LR), Decision-Tree Regressor (DTR), and K-Nearest-Neighbor Regressor (KNNR) in RGB and LAB color spaces. While performing regression in the LAB color space, both shadow and shadow-free images are transformed to the LAB space from the RGB space, and after performing regression and correction, they are again transformed back to the RGB space. Also, we have performed experiments by using a regression

Table 2. Quantitative results of detection with BER(%) on ISTD test dataset.

Method	BER	Method	BER
StackedCNN [25]	8.6	ST-CGAN [27]	3.85
CGAN [17]	9.64	Mask-ShadowGAN [8]	7.66
SCGAN [17]	4.7	Ours	6.48


Fig. 5. Visual comparison of shadow detection results on ISTD test dataset.

for each individual color channel (there will be three one-input to one-output regressor) and by using a regression for combined color-channel (multi-output regressor) (there will be one 3-input to 3-output regressor). For implementation, we used regression methods from the scikit-learn python library [22]. The results of the experiments are shown in Table 3.

Table 3. Quantitative results of ISTD test dataset adjustment task with RMSE.

	Original	Individual Channel					
		RGB LR	LAB LR	RGB DTR	LAB DTR	RGB KNNR	LAB KNNR
O	10.97	8.78	8.39	8.41	7.61	11.80	8.04
S	32.67	40.67	39.54	39.63	39.26	41.55	39.23
SF	6.83	2.81	2.55	2.56	1.68	6.23	2.20
	Original	Combined Channel					
		RGB LR	LAB LR	RGB DTR	LAB DTR	RGB KNNR	LAB KNNR
O	10.97	8.67	8.23	7.56	7.57	7.95	7.92
S	32.67	40.76	39.37	39.06	39.05	39.03	38.98
SF	6.83	2.66	2.39	1.66	1.67	2.12	2.11

Experimentally, we observed that the decision-tree combined channel regressor in RGB color space has a lower RMSE value in O and SF scenarios. So finally, we used that method and created a new adjusted ISTD training and testing dataset. Figure 6 shows the visual output of this database adjustment task by using the selected method.

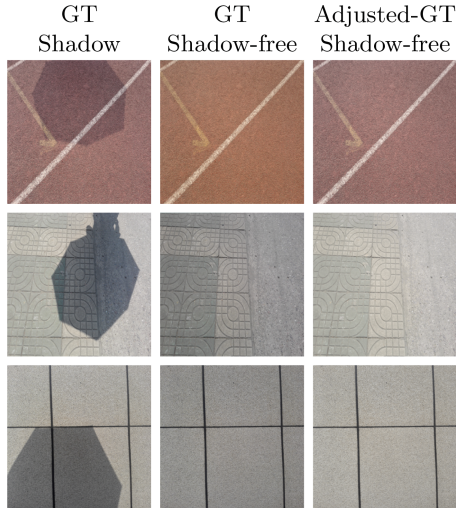


Fig. 6. Visual results of ISTD dataset adjustment task.

Evaluation on Removal with Adjusted Benchmark ISTD Dataset: We compare the shadow removal performance of the proposed method with the methods [8,27], trained and tested on the adjusted dataset of ISTD. Since the official code for the ST-CGAN method [27] is not available, we use the community code [23] for evaluation purpose. The results are shown in Table 4. The proposed method achieves the best performance in O, S, and SF scenarios compared to state-of-the-art methods.

Table 4. Quantitative shadow removal results with RMSE, trained and tested on adjusted ISTD dataset.

Method	O	S	SF	SF-I
Original	7.56	39.06	1.66	0
ST-CGAN [23,27]	8.79	11.35	8.31	4.28
Mask-ShadowGAN [8]	4.47	10.13	3.41	3.18
Ours	4.36	9.52	3.40	3.19

Ablation Study: We have done an ablation study on the presented framework by removing the exposure correction module (represented by -c) along with not using ground-truth shadow and shadow-free images (represented by -gt) and not using ground-truth masks (represented by -gtm). While performing an experiment with -gt, we ignored content losses, and for the -gtm experiment, initially, we generated masks by ground-truth shadow and shadow-free images according to Sect. 2.2. Removal and detection results for all the experiments are shown in Table 5. Visual performance for (-c) is shown in Fig. 4. Our approach achieves the best overall performance for removal and detection, and shows the importance of ground-truth data and correction module to achieve the best result.

Table 5. Ablation study.

Aspect	Removal				Detection
Method	O	S	SF	SF-I	BER
Ours	6.54	10.03	5.88	6.49	6.48
Ours(-gt)	6.98	11.07	6.22	6.54	8.37
Ours(-gtm)	6.85	10.11	6.23	6.93	6.64
Ours(-c)	6.57	10.62	5.82	6.43	6.76
Ours(-gt -c)	7.03	11.68	6.15	6.47	8.41
Ours(-gtm -c)	6.93	10.63	6.24	6.94	6.76

4 Conclusion

We proposed a method based on GAN to solve the shadow removal task in images. We used different constraints to effectively learn the bidirectional relationship between shadow and shadow-free domains under the paired setting. We also presented a novel process to handle the over-exposure problem after the training. As a result, the proposed method with an exposure correction module achieves the best or comparable performance compared to existing state-of-the-art methods, both quantitatively and visually. We explored the issue in benchmark datasets and introduced a technique for adjusting those benchmark datasets to additionally improve the shadow removal results. We also conducted various experiments to analyze the importance of ground-truth data and exposure correction module in generating better quality output.

Acknowledgements. This work was supported by the Science and Engineering Research Board (SERB), Department of Science and Technology (DST), New Delhi, India, under Grant No. CRG/2020/001982.

References

1. Goodfellow, I., et al.: Generative adversarial nets. In: *Advances in Neural Information Processing Systems*, vol. 27, pp. 2672–2680 (2014)
2. Russakovsky, O., et al.: Imagenet large scale visual recognition challenge. *Int. J. Comput. Vision (IJCV)* **115**(3), 211–252 (2015)
3. Bansal, N., Akashdeep, Aggarwal, N.: Deep learning based shadow detection in images. In: Krishna, C., Dutta, M., Kumar, R. (eds.) *Proceedings of 2nd International Conference on Communication, Computing and Networking. LNNS*, pp. 375–382. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-1217-5_37
4. Fan, H., Han, M., Li, J.: Image shadow removal using end-to-end deep convolutional neural networks. *Appl. Sci.* (2019)
5. Gong, H., Cosker, D.: Interactive shadow removal and ground truth for variable scene categories. In: *Proceedings of the British Machine Vision Conference* (2014)
6. Guo, R., Dai, Q., Hoiem, D.: Paired regions for shadow detection and removal. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(12), 2956–2967 (2013)
7. Hu, X., Fu, C., Zhu, L., Qin, J., Heng, P.: Direction-aware spatial context features for shadow detection and removal. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**(11), 2795–2808 (2020)
8. Hu, X., Jiang, Y., Fu, C., Heng, P.: Mask-ShadowGAN: learning to remove shadows from unpaired data. In: *IEEE International Conference on Computer Vision (ICCV)*, pp. 2472–2481 (2019)
9. Isola, P., Zhu, J., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1125–1134 (2017)
10. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) *ECCV 2016. LNCS*, vol. 9906, pp. 694–711. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46475-6_43
11. Khan, S.H., Bennamoun, M., Sohel, F., Togneri, R.: Automatic feature learning for robust shadow detection. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1939–1946 (2014)
12. Khan, S.H., Bennamoun, M., Sohel, F., Togneri, R.: Automatic shadow detection and removal from a single image. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(3), 431–446 (2016)
13. Khare, M., Srivastava, R.K., Khare, A.: Object tracking using combination of Daubechies complex wavelet transform and Zernike moment. *Multimedia Tools Appl.* **76**(1), 1247–1290 (2017)
14. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)
15. Mirza, M., Osindero, S.: Conditional generative adversarial nets. [arXiv:1411.1784](https://arxiv.org/abs/1411.1784) (2014)
16. Nagae, T., Abiko, R., Yamaguchi, T., Ikehara, M.: Shadow detection and removal using GAN. In: *Proceedings of 28th European Signal Processing Conference (EUSIPCO)*, pp. 630–634 (2021)
17. Nguyen, V., Vicente, T.F.Y., Zhao, M., Hoai, M., Samarasinghe, D.: Shadow detection with conditional generative adversarial networks. In: *IEEE International Conference on Computer Vision (ICCV)*, pp. 4510–4518 (2017)
18. Otsu, N.: A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **9**(1), 62–66 (1979)

19. Qu, L., Tian, J., He, S., Tang, Y., Lau, R.W.: DeshadowNet: a multi-context embedding deep network for shadow removal. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4067–4075 (2017)
20. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: IEEE Conference on Computer vision and Pattern Recognition (CVPR), pp. 779–788 (2016)
21. Sanin, A., Sanderson, C., Lovell, B.C.: Shadow detection: a survey and comparative evaluation of recent methods. *Pattern Recogn.* **45**(4), 1684–1695 (2012)
22. Scikit-learn: <https://scikit-learn.org/stable/>
23. ST-CGAN: https://github.com/IsHYuhi/ST-CGAN_Stacked_Conditional_Gen-erative_Adversarial_Networks
24. Tan, C., Feng, X.: Unsupervised shadow removal using target consistency generative adversarial network. *arXiv:2010.01291* (2020)
25. Vicente, T.F.Y., Hou, L., Yu, C.-P., Hoai, M., Samaras, D.: Large-scale training of shadow detectors with noisily-annotated shadow examples. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9910, pp. 816–832. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46466-4_49
26. Wang, B., Chen, C.L.P.: An effective background estimation method for shadows removal of document images. In: IEEE International Conference on Image Processing (ICIP), pp. 3611–3615 (2019)
27. Wang, J., Li, X., Yang, J.: Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1788–1797 (2018)
28. Yang, Q., Tan, K., Ahuja, N.: Shadow removal using bilateral filtering. *IEEE Trans. Image Process.* **21**(10), 4361–4368 (2012)
29. Yao, K., Dong, J.: Removing shadows from a single real-world color image. In: IEEE International Conference on Image Processing (ICIP), pp. 3129–3132 (2009)
30. Zhu, J., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: IEEE International Conference on Computer Vision (ICCV), pp. 2223–2232 (2017)