

# Shadow Detection And Removal From Images

by

**MEHTA KRUNAL KAMLESHKUMAR**  
**202011051**

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of

MASTER OF TECHNOLOGY  
in  
INFORMATION AND COMMUNICATION TECHNOLOGY  
to

**DHIRUBHAI AMBANI INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGY**

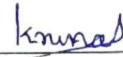


May, 2022

## Declaration

I hereby declare that

- i) the thesis comprises of my original work towards the degree of Master of Technology in Information and Communication Technology at Dhirubhai Ambani Institute of Information and Communication Technology and has not been submitted elsewhere for a degree,
- ii) due acknowledgment has been made in the text to all the reference material used.




Mehta Krunal Kamleshkumar

## Certificate

This is to certify that the thesis work entitled "Shadow Detection And Removal From Images" has been carried out by **Mehta Krunal Kamleshkumar (202011051)** for the degree of Master of Technology in Information and Communication Technology at *Dhirubhai Ambani Institute of Information and Communication Technology* under our supervision.

  
20/6/22

Prof. Manish Khare  
Thesis Supervisor

 20/06/2022

Prof. Avik Hati  
Thesis Co-Supervisor

# Acknowledgments

On the successful completion of my M.Tech thesis work, I am really grateful to all the individuals who have assisted me during this beautiful journey and without whom this work would not have been possible.

Firstly, I would like to express my sincere gratitude towards my thesis supervisor, Prof. Manish Khare, and my thesis co-supervisor, Prof. Avik Hati, for their constant support and guidance throughout the journey. I would like to convey my sincere thanks to Prof. Manish Khare for giving me a chance to explore this field and providing valuable assistance in carrying out this work. His friendly and helping nature and excellent domain knowledge enabled me to achieve the best outcome. I am also very thankful to Prof. Avik Hati for constantly providing his valuable input and giving proper direction to carry out this research work. His technical and domain knowledge genuinely helped me in solving any issue I had during this journey. The discussions made through continuous progress meetings truly motivated me to explore new things to achieve the desired outcome.

I would like to extend my gratitude to the Help Desk, DAIICT, for providing access to the required resources to carry out my work and giving all the technical assistance and help whenever needed. I would also like to thank Resource Center, DAIICT, for providing credentials to all the reading materials for the literature survey of this work.

Furthermore, I thank all my friends for their support and inspiration throughout this journey. A special thanks to my family for their unconditional love, care, encouragement, and confidence in me.

# Contents

<b>Abstract</b>	<b>v</b>
<b>List of Principal Symbols and Acronyms</b>	<b>vi</b>
<b>List of Tables</b>	<b>vii</b>
<b>List of Figures</b>	<b>viii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Objective . . . . .	2
1.3 Contributions . . . . .	2
1.4 Thesis Organization . . . . .	3
<b>2 Literature Survey</b>	<b>4</b>
2.1 GAN-Based Methods . . . . .	5
2.1.1 Introduction to GAN and CGAN . . . . .	5
2.1.2 Supervised Learning-Based Methods . . . . .	6
2.1.3 Unsupervised Learning-Based Methods . . . . .	8
2.2 Overview of Methods . . . . .	10
2.3 Overview of Datasets . . . . .	11
<b>3 Proposed Method</b>	<b>12</b>
3.1 Generators and Discriminators Learning . . . . .	12
3.2 Shadow Detection Using Mask Producer . . . . .	14
3.3 Adaptive Exposure Correction Module . . . . .	14
3.3.1 Importance of Correction Module . . . . .	14
3.3.2 Methodology for Exposure Correction . . . . .	15
3.3.3 Methodology for Gamma Estimation and Correction . . . . .	15
3.4 Objectives and Loss Functions . . . . .	16
3.4.1 Adversarial Losses . . . . .	16

3.4.2	Cycle Consistency Losses . . . . .	17
3.4.3	Identity Losses . . . . .	17
3.4.4	Content Losses . . . . .	18
3.4.5	Loss Function for Generators . . . . .	18
3.4.6	Loss Function for Discriminators . . . . .	18
3.5	Network Architecture and Training Strategy . . . . .	18
3.6	Benchmark Dataset Correction . . . . .	19
3.6.1	Importance of Correcting Benchmark Dataset . . . . .	19
3.6.2	Methodology for Benchmark Dataset Correction . . . . .	20
<b>4</b>	<b>Experimental Results</b>	<b>22</b>
4.1	Description of Datasets . . . . .	22
4.2	Evaluation Parameters . . . . .	22
4.3	Evaluation on Shadow Removal . . . . .	23
4.3.1	Evaluation on ISTD Dataset . . . . .	23
4.3.2	Evaluation on SRD Dataset . . . . .	24
4.4	Evaluation on Shadow Detection . . . . .	27
4.5	Benchmark Dataset Correction . . . . .	27
4.6	Evaluation on Shadow Removal with Corrected Benchmark ISTD Dataset . . . . .	29
4.7	Ablation Studies . . . . .	31
<b>5</b>	<b>Conclusions and Future Work</b>	<b>34</b>
5.1	Conclusions . . . . .	34
5.2	Future Work . . . . .	34
	<b>References</b>	<b>36</b>

# 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 obtaining realistic shadow-free results. The features of shadows are the same as those of objects. As a result, it has the potential to be misclassified as a part of the object, resulting in degrading the performance of many computer vision tasks. In recent years, several deep learning-based frameworks have been presented to solve this issue. This work presents 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 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 proposed approach on the two large-scale supervised benchmark datasets and show quantitative and visual improvements in the state-of-the-art results.

**Keywords:** Shadow detection, Shadow removal, Generative adversarial networks, Adaptive exposure correction, Benchmark dataset correction

# List of Principal Symbols and Acronyms

BCE	Binary Cross Entropy
BER	Balance Error Rate
CGAN	Conditional Generative Adversarial Network
CNN	Convolutional Neural Network
DL	Deep Learning
DTR	Decision-Tree Regressor
GAN	Generative Adversarial Network
GT	Ground-Truth
KNNR	K-Nearest-Neighbor Regressor
LR	Linear Regressor
MSE	Mean Square Error
MSGAN	Mask-ShadowGAN
ReLU	Rectified Linear Unit
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
YOLO	You Only Look Once

# List of Tables

2.1	Details of type and aspect of methods . . . . .	10
2.2	Details of data requirement of deep learning-based methods . . . .	11
2.3	Details of large-scale shadow datasets . . . . .	11
4.1	Quantitative shadow removal results with RMSE on ISTD test dataset . . . . .	24
4.2	Quantitative shadow removal results with RMSE on SRD test dataset . . . . .	24
4.3	Quantitative shadow detection results with BER on ISTD test dataset . . . . .	27
4.4	Quantitative results of ISTD test dataset correction with RMSE . . .	29
4.5	Quantitative shadow removal results with RMSE, trained and tested on corrected ISTD . . . . .	29
4.6	Ablation study on exposure correction module and ground- truth data . . . . .	31
4.7	Ablation study on adversarial loss . . . . .	32
4.8	Ablation study on using additional shadow mask as an input to shadow-free generator . . . . .	32
4.9	Ablation study on discriminator learning . . . . .	33



# List of Figures

1.1	Types of shadow . . . . .	1
1.2	Shadow effect in object detection task . . . . .	2
1.3	Comparison of shadow removal result . . . . .	3
2.1	Learning process of GAN and CGAN . . . . .	6
2.2	Nguyen <i>et al.</i> 's method . . . . .	7
2.3	Wang <i>et al.</i> 's method . . . . .	8
2.4	Hu <i>et al.</i> 's method . . . . .	9
2.5	Tan <i>et al.</i> 's method . . . . .	9
3.1	Illustration of the architecture of the proposed method . . . . .	13
3.2	Need for a correction module . . . . .	15
3.3	ISTD triplets, showing issue in non-shadow area . . . . .	20
3.4	Impact of using incorrect benchmark dataset during training . . . . .	20
3.5	Regressor learning for dataset correction . . . . .	21
4.1	Visual comparison of shadow removal results on ISTD test dataset . . . . .	25
4.2	Visual performance of shadow removal results on SRD test dataset . . . . .	26
4.3	Visual comparison of shadow detection results on ISTD test dataset . . . . .	28
4.4	Visual results of ISTD dataset correction task . . . . .	30
4.5	Visual comparison on GT and corrected GT training . . . . .	30
4.6	Visual performance of exposure correction module's ablation study . . . . .	32
5.1	Shadow edge issue . . . . .	35

## CHAPTER 1

# Introduction

Shadow detection and removal are fundamental and challenging tasks in the computer vision and computer graphics domain. It concentrates on detecting and removing the shadow, keeping the non-shadow area unchanged. Shadow is the darkest area in the image where the light from the light source is occluded by the object. Figure 1.1 shows different types of shadow, where the shadow is mainly classified into two categories: self shadow and cast shadow. Self shadow is formed by an occluded object portion not being exposed by the light, and cast shadow is produced by an object creating shadows on the area of another object. Moreover, the cast shadow is separated into umbra and penumbra regions. The Umbra region is a darker zone in which direct light is wholly obstructed, whereas the penumbra region is a lighter zone of the cast shadow.

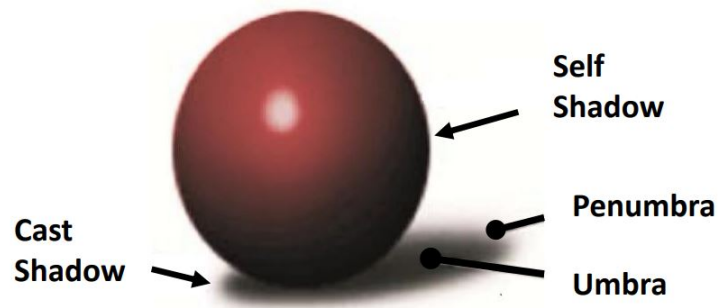


Figure 1.1: Types of shadow.

## 1.1 Motivation

Shadows have similar characteristics as the object, so sometimes, they can be misclassified as part of the object. Due to that, the accuracy of several computer vision tasks, such as object segmentation, object detection, and object tracking, can be badly affected.



Figure 1.2: Shadow effect in object detection task.

Figure 1.2 shows the example of object detection experiment in the shadow image and corresponding shadow-free image, where the pretrained model of YOLOv5 [1] was used as an object detector. In the experiment with a shadow-free image, the object is detected correctly, whereas, in the experiment with a shadow image, some portion of the shadow is detected as a part of an object. Also, it gives less confidence ( $0.48$ ) compared to shadow-free image object confidence ( $0.66$ ). Consequently, shadow detection and removal tasks are essential in many computer vision and computer graphics tasks to improve efficiency and achieve better results.

## 1.2 Objective

Given the domain of shadow images  $\mathbb{D}_x$  and the domain of shadow-free images  $\mathbb{D}_y$ , we are primarily focused on learning the mapping function  $G_f : \mathbb{D}_x \rightarrow \mathbb{D}_y$ , which transforms shadow domain image to the shadow-free domain image.

## 1.3 Contributions

In this work, we propose a novel method based on Generative Adversarial Networks (GANs) with cycle and content constraints, and introduce an adaptive exposure correction module for handling the common over-exposure problem. Figure 1.3 shows a shadow removal result of the proposed method compared with state-of-the-art Mask-ShadowGAN [2] method, 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.

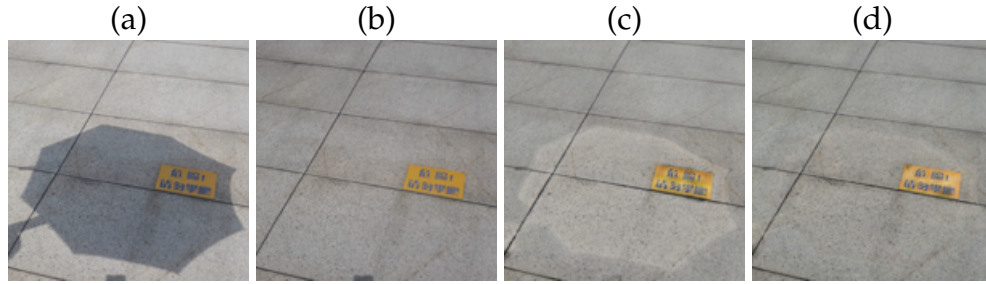


Figure 1.3: Comparison of shadow removal result; (a) input shadow image, (b) ground-truth shadow-free image, (c) output of Mask-ShadowGAN [2] method, (d) output of proposed method.

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.
- We introduce an *adaptive exposure correction module* for handling the common over-exposure problem.
- We introduce the method for enhancing the quality of benchmark datasets and subsequently improving the shadow removal results.
- We analyze the impact of using ground-truth data with the adaptive exposure correction module in the proposed framework. Also, we show the impact of using different learning strategies in the proposed methodology.

## 1.4 Thesis Organization

The rest of the thesis report is organized as follows. Chapter 2 shows the literature survey on shadow detection and removal. Chapter 3 explains the proposed framework. Chapter 4 presents experimental results along with the ablation studies. Finally, we conclude and provide future work in Chapter 5.

## CHAPTER 2

# Literature Survey

In computer vision, the problem involving shadow detection and removal has received much attention. Early works related to this task [3] used different hand-crafted features like texture, chromaticity, intensity, etc., for shadow detection and removal. Model-based strategies and property-based strategies are the two primary shadow detection and removal algorithms categories. Model-based techniques were further divided into physical-model-based and geometry-based methods, and property-based techniques were divided into texture-based and color-based methods. Several characteristics, such as camera localization, light source direction, and the shape of objects, are assumed to be known for approaches based on physical models [4]. With previous information on the ground surface, camera position, and object geometry, geometry-based techniques [5] may predict the orientation, size, and shape of shadows. The principle behind using texture-based techniques [6] is that the foreground item’s texture differs from the background’s texture, and the task is performed using texture descriptors. The color-based techniques [7, 8, 9, 10] attempt to analyze the color changes, intensity, and illumination for shadow detection and removal. However, these all 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. The model of Qu *et al.* [11] (DeshadowNet) is based on an end-to-end multi-context embedding framework to extract essential characteristics from multiple aspects and accumulate them to determine shadow matte. It basically learns the mapping function among the shadow image and its shadow matte and uses predicted shadow matte to reconstruct a shadow-free image with Equation (2.1).

$$I_s = I_{sf} \cdot S_m \quad (2.1)$$

where  $I_s$  is a shadow image, which is counted as a pixel-wise product of a shadow-free image  $I_{sf}$  and its shadow matte  $S_m$ .

Bansal *et al.* [12] developed an approach based on deep Convolutional Neural Network (CNN), which extracts features from the shadow image and uses them to detect shadows automatically in images. Fan *et al.* [13] presented a deep CNN network structure to learn the relation between the shadow image and its shadow matte. It is made up of two sub-network models: an encoder-decoder model for extracting features and reconstructing images, and a small refinement model that takes into account a multi-context strategy for local detail correction. Shadow-free image is recovered from the shadow matte according to Equation (2.1). Vicente *et al.* [14] developed StackedCNN, an approach that utilizes a Fully Connected Network (FCN) with a patchCNN for the shadow detection. Fu *et al.* [15] presented a network that uses exposure estimation to learn numerous over-exposure pictures by correcting the shadow region with varying exposure levels, then creating fusion weight maps and doing pixel-wise fusing the input image with its over-exposure counterparts to generate the final shadow-free result.

These CNN-based methods [11, 13, 15] achieve better performance compared to hand-crafted feature-based methods but suffer from random artifacts in non-shadow area and fail to obtain realistic shadow-free results.

## 2.1 GAN-Based Methods

### 2.1.1 Introduction to GAN and CGAN

The Generative Adversarial Networks (GANs) [16] and its extensions, presented in recent years, are dominant strategies for dealing with different image to image translation challenges. A GAN consists of two deep neural network architectures: a generator  $G$  and a discriminator  $D$ . The generator model  $G$  generates realistic fake images. The discriminator model  $D$  is used to determine if an input image was created by  $G$  or it is a real image from the training set. The two models trained together in an adversarial manner until the discriminator model fooled, meaning the generator model is generating samples such that it can fool the discriminator model.  $G$  focuses on generating images that make it difficult for  $D$  to discriminate  $G$ 's creation from the real set of data.  $D$ , on the other hand, wants to avoid getting a fool by  $G$ .

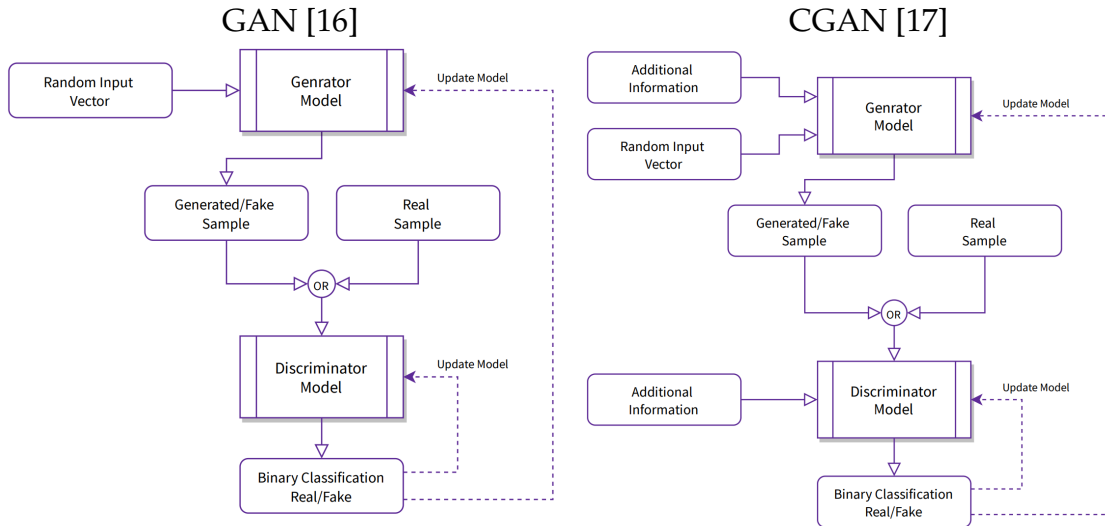


Figure 2.1: Learning process of GAN and CGAN.

The Conditional Generative Adversarial Networks (CGANs) [17] are significant GAN extensions that incorporate conditioning information to the generator  $G$  and the discriminator  $D$  models. It forces  $G$  to produce images that are indistinguishable from real images, and according to conditioning information. Figure 2.1 shows the learning process of GAN and CGAN in detail.

### 2.1.2 Supervised Learning-Based Methods

The Pix2PixGAN [18] is a well-known method for paired image-to-image translation. It is based on the CGAN, where the generator generates a target image, keeping the input image as conditional information, and the discriminator guides generator to produce a better result.

Nguyen *et al.* [19] demonstrated the first method of shadow detection with adversarial learning and constructed a CGAN-based architecture inspired by the Pix2PixGAN. Figure 2.2 shows the architecture, where the generator is trained to output a shadow mask that can realistically correspond to the ground-truth mask. The shadow mask is a binary image, where zero value indicates a shadow-free region and one value indicates a shadow region. The discriminator is trained to decide if the pair of shadow image and mask comes from the training data or not. Shadow image with an adjustable sensitivity factor is used as the conditioning information for the generator and discriminator.

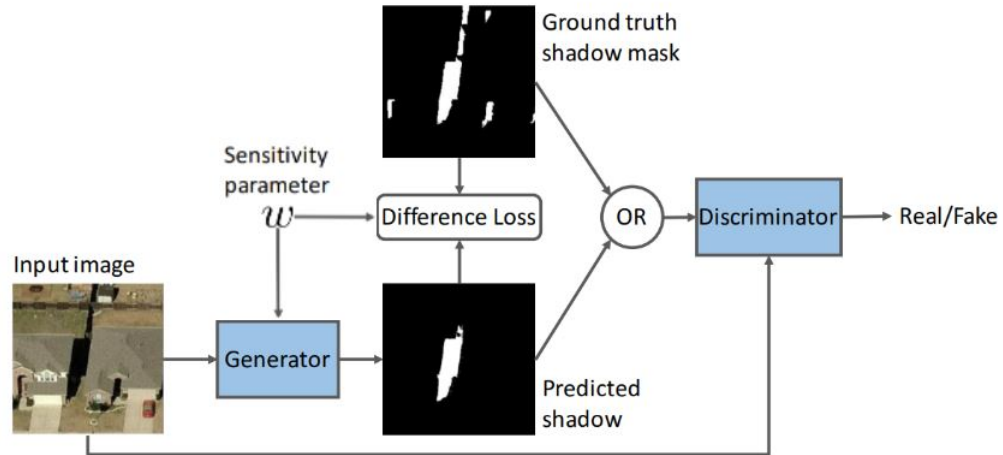


Figure 2.2: Nguyen *et al.*'s method. [19]

Wang *et al.* [20] presented a model based on Stacked Conditional Generative Adversarial Network (ST-CGAN) that uses paired data to tackle shadow detection and removal problems simultaneously in an end-to-end manner. The proposed method by the authors is shown in Figure 2.3. Two arranged CGANs, each one with a pair of generator and discriminator, build up the model. Pair  $G_1 - D_1$  is for shadow detection, and pair  $G_2 - D_2$  is for shadow removal. The shadow image is applied as input to produce a shadow mask image in the shadow detection generator  $G_1$ .  $G_2$  is a shadow removal generator that takes a four-channel input with a shadow image and mask and produces a shadow-free image. The shadow detection discriminator  $D_1$  takes a four-channel input, including the input shadow image and the mask generated by  $G_1$ . The shadow removal discriminator  $D_2$  takes a seven-channel input containing the input shadow image, mask from  $G_1$ , and the shadow-free image generated by  $G_2$ . Both discriminators will discriminate whether the generated sample is real or fake, and accordingly, weights and biases are updated. The loss function is made with the adversarial constraint along with the ground-truth content constraint. After the training, the model will be able to determine the optimum weights for detecting and removing shadows.

Nagae *et al.* [21] developed a model based on [20], with minor changes in the shadow removal CGAN, that estimates the illumination ratio and uses that estimation to produce the shadow-free output. Hu *et al.* [22] presented a Direction-aware Spatial Context (DSC) module based on Recurrent Neural Network (RNN) utilized with CNNs, to learn the spatial context in a direction-aware manner and subsequently detect and remove the shadow.



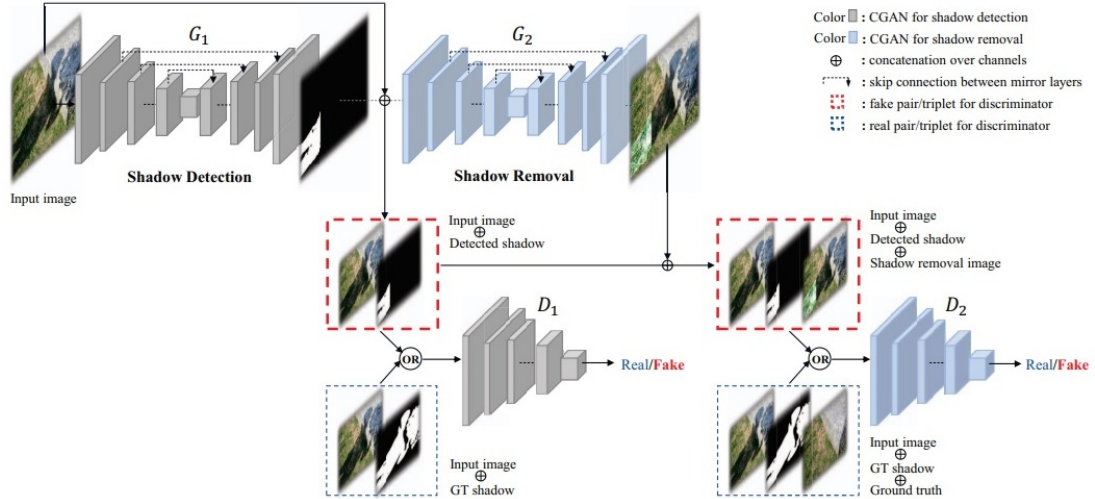


Figure 2.3: Wang *et al.*'s method. [20] (ST-CGAN)

Although these approaches [20, 21, 22] effectively remove the shadow, they tend to generate artifacts and inconsistent colors, particularly in the non-shadow area, resulting in performance degradation in the non-shadow area.

### 2.1.3 Unsupervised Learning-Based Methods

The CycleGAN [23] is a common method for unpaired image-to-image translation. The method is based on CGAN architecture. It uses two generative adversarial networks to establish cycle consistency restrictions and learn the mapping for translation using unpaired training data.

Hu *et al.* [2] presented a Mask-ShadowGAN (MSGAN) framework based on the CycleGAN [23] approach. Mask-ShadowGAN enforces cycle consistency by the guidance of the produced binary mask in the generative process. The key idea is to learn the bidirectional mapping within the shadow and shadow-free domains. Figure 2.4 depicts the Mask-ShadowGAN framework's overall network architecture. By using adversarial learning, the network  $G_f$  takes a shadow image as input and generates a shadow-free image that is indistinguishable from the shadow-free domain. In addition, network  $G_s$  is also trained, which takes a shadow-free image and shadow-mask image and generates a shadow image that is indistinguishable from the shadow domain. Thus cycle-consistency constraints are imposed to train  $G_f$  and  $G_s$ .  $D_f$  discriminates real and generated shadow-free images, and  $D_s$  discriminates real and generated shadow images. Along with the cycle consistency

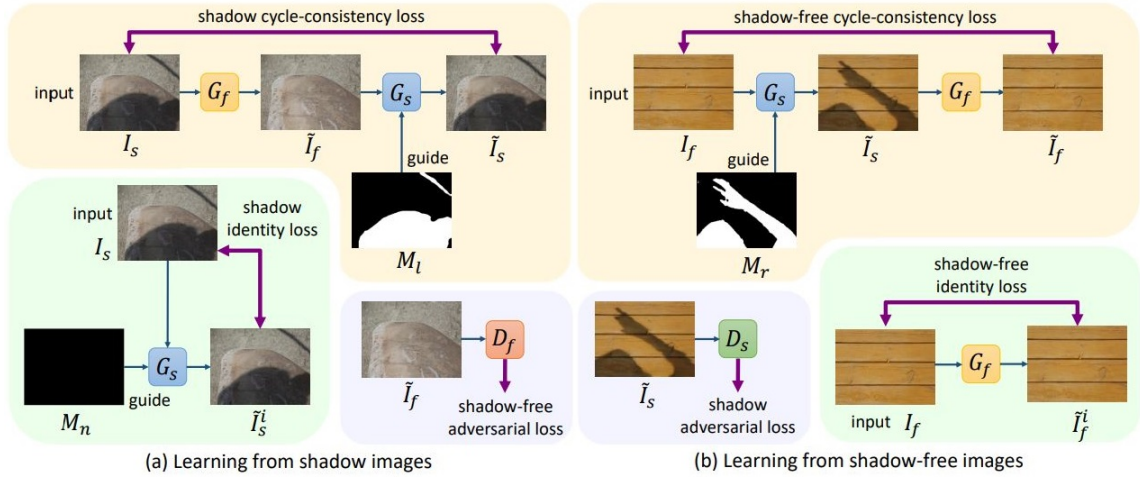


Figure 2.4: Hu *et al.*'s method. [2] (Mask-ShadowGAN)

and adversarial constraints, identity loss is also calculated to preserve the color combination. After the training, the model can remove shadows efficiently.

Tan *et al.* [24] developed a target consistency generative adversarial network (TCGAN) for the shadow removal using unpaired training data. TCGAN aims to learn a unidirectional mapping to translate shadow images into shadow-free images, as opposed to MaskShadowGAN, which learns bidirectional mapping. The architecture is shown in the Figure 2.5. TCGAN uses two generative adversarial

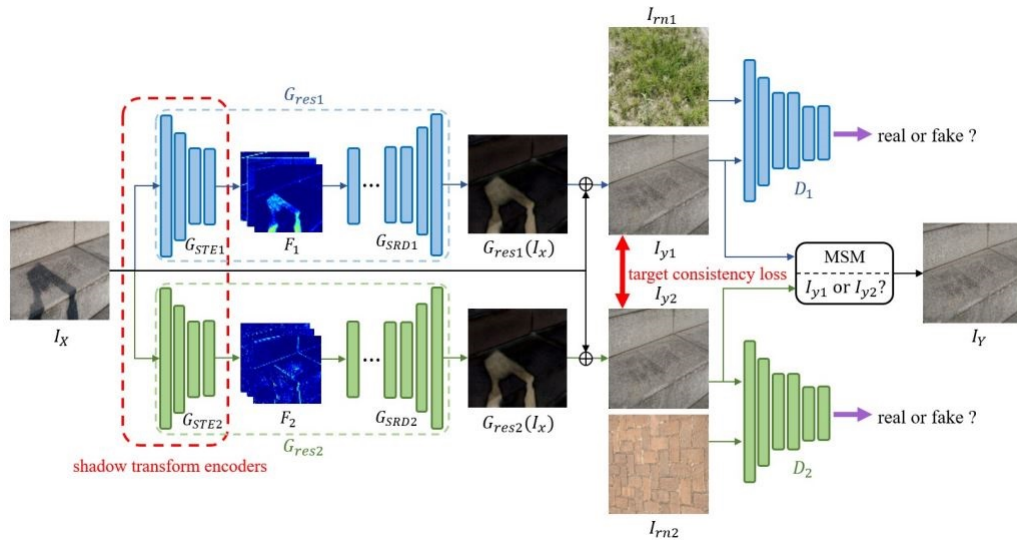


Figure 2.5: Tan *et al.*'s method. [24] (TCGAN)

networks, generators  $G_{res1}$  and  $G_{res2}$  to generate two feature images from a single input shadow image. Two different shadow-free images are produced from two feature images and a shadow image, and a target consistency constraint is

formed on those generated shadow-free images. Two discriminators,  $D_1$  and  $D_2$  discriminate input image as real or fake and guide corresponding generators to produce better realistic output. Additionally, the Model Selection Module (MSM) is presented at the end of GANs, which is a pretrained binary classifier model for shadow. It selects one image with the higher confidence from the two produced shadow-free images as the final shadow-free result.

These methods [2, 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.

## 2.2 Overview of Methods

Table 2.1 shows the overall details regarding the type of each method, describing whether the method is hand-crafted feature-based or deep learning-based. The table also shows the aspect of each method, telling whether the method does shadow detection and/or shadow removal.

Table 2.1: Details of type and aspect of methods.

Method	Type of method	Aspect
Prati [4]	Feature (Physical)	Shadow detection
Chang [5]	Feature (Geometry)	Shadow detection, Shadow removal
Leone [6]	Feature (Texture)	Shadow detection, Shadow removal
Guo [7]	Feature (Color)	Shadow detection, Shadow removal
Yang [8]	Feature (Color)	Shadow removal
Gong [9]	Feature (Color)	Shadow removal
Khare [10]	Feature (Color)	Shadow detection, Shadow removal
DeshadowNet [11]	DL (CNN)	Shadow removal
Bansal [12]	DL (CNN)	Shadow detection
Fan [13]	DL (CNN)	Shadow removal
StackedCNN [14]	DL (CNN)	Shadow detection
FusionNet [15]	DL (CNN)	Shadow removal
SCGAN [19]	DL (GAN)	Shadow detection
ST-CGAN [20]	DL (GAN)	Shadow detection, Shadow removal
Nagae [21]	DL (GAN)	Shadow detection, Shadow removal
DSC [22]	DL (GAN)	Shadow detection, Shadow removal
MSGAN [2]	DL (GAN)	Shadow detection, Shadow removal
TCGAN [24]	DL (GAN)	Shadow removal

Table 2.2 shows the type of images required and type of data needed for each deep learning-based method.

Table 2.2: Details of data requirement of deep learning-based methods.

Method	Data required	Type of data
DeshadowNet [11]	Shadow, Shadow-free	Paired
Bansal [12]	Shadow, Shadow-mask	Paired
Fan [13]	Shadow, Shadow-free	Paired
StackedCNN [14]	Shadow, Shadow-mask	Paired
SCGAN [19]	Shadow, Shadow-mask	Paired
ST-CGAN [20]	Shadow, Shadow-mask, Shadow-free	Paired
Nagae [21]	Shadow, Shadow-mask, Shadow-free	Paired
DSC [22]	Shadow, Shadow-mask, Shadow-free	Paired
FusionNet [15]	Shadow, Shadow-mask, Shadow-free	Paired
MSGAN [2]	Shadow, Shadow-free	Unpaired
TCGAN [24]	Shadow, Shadow-free	Unpaired

## 2.3 Overview of Datasets

Table 2.3 shows the details of large-scale available shadow datasets with the amount and type of data.

Table 2.3: Details of large-scale shadow datasets.

Dataset	Amount	Content	Type of data
SRD [11]	3088	Shadow, Shadow-free	Paired
ISTD [20]	1870	Shadow, Shadow-mask, Shadow-free	Paired
USR [2]	2445	Shadow, Shadow-free	Unpaired
SBU [14]	4723	Shadow, Shadow-mask	Paired
CUHK [25]	10500	Shadow, Shadow-mask	Paired

These large-scale datasets are generally used to train various deep learning-based models for shadow detection and/or shadow removal.

## CHAPTER 3

# Proposed Method

The overall scheme of the proposed method is depicted in Figure 3.1. The method is based on CycleGAN [23], 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 MaskShadowGAN [2], which required unpaired data with an equal statistical distribution of shadow and shadow-free domain, our method utilizes available shadow, shadow-free, and shadow-mask images to learn better mapping for shadow removal.

### 3.1 Generators and Discriminators 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  $\hat{I}_{s*}$ . 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  $\hat{I}_{s*}$ .

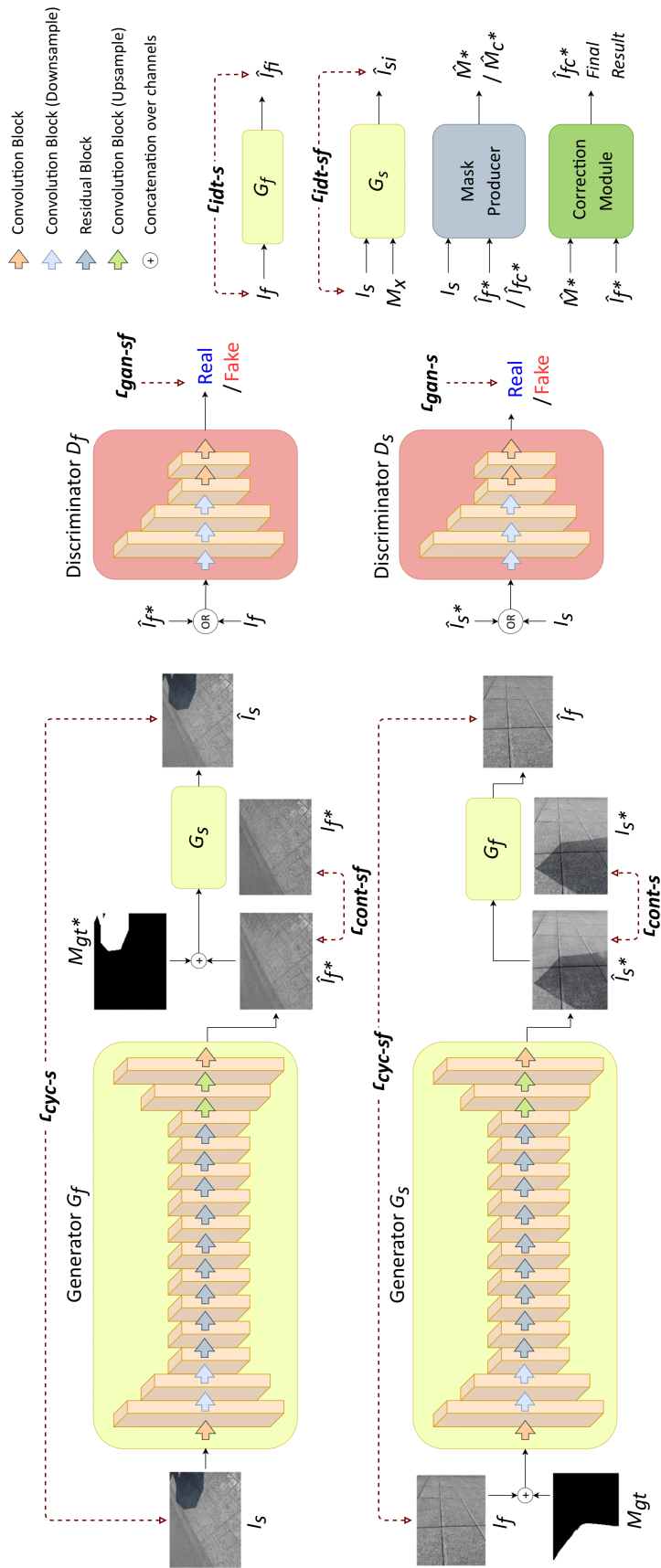


Figure 3.1: Illustration of the architecture of the proposed method.

To summarize, the generator network  $G_f$  generates the shadow-free result  $\hat{I}_{f*}$  for the  $I_s$ , and generator  $G_s$  generates the shadow result  $\hat{I}_{s*}$  for the  $I_f$ . 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.

## 3.2 Shadow Detection Using Mask Producer

The intermediate shadow mask  $\hat{M}_*$  for the input shadow image  $I_s$  is obtained by the Equation (3.1).

$$\hat{M}_* = \mathbb{B}(\hat{I}_{f*} - I_s, t) \quad (3.1)$$

where the binarization operation  $\mathbb{B}$  is performed on the difference between generated shadow-free image  $\hat{I}_{f*}$  and the real input shadow image  $I_s$ , and  $t$  is a threshold obtained by Otsu’s algorithm [26].  $\mathbb{B}$  sets the value as zero or one, where zero indicates non-shadow regions (difference  $\leq t$ ) and one indicates the shadow region (difference  $> t$ ).

The final shadow mask  $\hat{M}_{c*}$ , after applying the correction module, for the input shadow image  $I_s$  is obtained by the Equation (3.2).

$$\hat{M}_{c*} = \mathbb{B}(\hat{I}_{fc*} - I_s, t) \quad (3.2)$$

where the binarization operation  $\mathbb{B}$  is performed on the difference between the final shadow-free image  $\hat{I}_{fc*}$  (after applying the correction module, described in the next section) and the real input shadow image  $I_s$ .

## 3.3 Adaptive Exposure Correction Module

### 3.3.1 Importance of Correction Module

The generator network  $G_f$  is trained to produce a shadow-free result. But in the absence of any constraints, sometimes it produces results with a much brighter shadow area, resulting in a shadow-free result with an over-exposed shadow area. Figure 3.2 shows a shadow removal result of the proposed method without using correction module (denoted as *Ours (-c)*) compared with state-of-the-art Mask-ShadowGAN [2] method. While Mask-ShadowGAN suffers from over-exposure,



Figure 3.2: Need for a correction module.

particularly in the shadow area, *Ours (-c)* reduces the degree of over-exposure. However, still, it shows the need for a correction module to handle the over-exposure problem.

### 3.3.2 Methodology for Exposure Correction

In the adaptive exposure correction module, we apply gamma correction (power-law transformation) in the shadow area. Following are the steps we followed for the exposure correction.

- Step 1: Extract the shadow and shadow-free areas using the generated mask  $\hat{M}_*$  and transform the shadow area to HSV (Hue, Saturation, and Value) color space.
- Step 2: Apply gamma correction on the value channel of the shadow area. Here, the value channel is selected for gamma correction because it is responsible for the brightness in the image.
- Step 3: Transform the gamma-corrected HSV shadow area image back to the RGB (Red, Green, and Blue) color space.
- Step 4: Combine the gamma-corrected shadow area image with the original shadow-free area image to generate  $\hat{I}_{fc*}$  which is the final shadow-free image with exposure correction.

### 3.3.3 Methodology for Gamma Estimation and Correction

The following steps are performed to estimate the value of gamma and subsequently do gamma correction in the shadow area.

- Step 1: Transform the image to HSV color space.
- Step 2: Calculate shadow area mean of value channel (Let it denote as  $m_1$ ).



Step 3: Calculate shadow-free area mean of value channel (Let it denote as  $m_2$ ).

Step 4: Calculate mean difference of  $m_1$  and  $m_2$  as per Equation (3.3).

$$diff = m_1 - m_2 \quad (3.3)$$

Step 5: Calculate gamma value by the Equation (3.4). The expected range for the  $diff$  is between  $-128$  to  $128$ , and we are mapping the  $diff$  to the gamma value range  $0$  to  $2$ . Ideally, for a non-overexposed shadow area image,  $diff$  will be  $0$ , and the gamma value will be set as  $1$ . For an over-exposed shadow area image,  $m_1$  will be higher than  $m_2$ , making  $diff$  positive, resulting in the gamma value greater than  $1$ . Similarly, for an under-exposed shadow area image,  $m_1$  will be lower than  $m_2$ , making  $diff$  negative, resulting in the gamma value less than  $1$ .

$$gamma = (diff/128) + 1 \quad (3.4)$$

Step 6: Apply gamma correction on the value channel of the shadow area image with the calculated  $gamma$  by the Equation (3.5). The Equation (3.5) is formed such that the intensity value for a pixel remains in the range  $0$  to  $255$ .

$$val = [(val/255)^{gamma}] * 255 \quad (3.5)$$

## 3.4 Objectives and Loss Functions

### 3.4.1 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. Here, Mean Square Error (MSE) loss function is used to calculate adversarial losses. The shadow-free adversarial loss and the shadow adversarial loss for generators and discriminators are given as:

$$\mathcal{L}_{gan-sf(G)} = MSE(P, D_f(\hat{I}_{f*})) \quad (3.6)$$

$$\mathcal{L}_{gan-s(G)} = MSE(P, D_s(\hat{I}_{s*})) \quad (3.7)$$

$$\mathcal{L}_{gan-sf(D)} = MSE(P, D_f(I_f)) + MSE(Q, D_f(\hat{I}_{f*})) \quad (3.8)$$

$$\mathcal{L}_{gan-s(D)} = MSE(P, D_s(I_s)) + MSE(Q, D_s(\hat{I}_{s*})) \quad (3.9)$$

where  $P$  and  $Q$  are one and zero matrix, respectively, with a size as the discriminator's output size.  $\hat{I}_{f*}$  (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, and  $I_s$  and  $I_f$  are the input shadow and shadow-free images, respectively.

### 3.4.2 Cycle Consistency Losses

Cycle consistency  $L_1$  losses defined in Equation (3.10) and Equation (3.11) are applied to encourage the reconstructed images to be comparable to the original input images, to prevent random artifacts in the non-shadow area, and to effectively improve the bidirectional mapping in the  $G_f$  and  $G_s$  networks.

$$\mathcal{L}_{cyc-s} = \|\hat{I}_s - I_s\|_1 \quad (3.10)$$

$$\mathcal{L}_{cyc-sf} = \|\hat{I}_f - I_f\|_1 \quad (3.11)$$

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.

### 3.4.3 Identity Losses

The identity  $L_1$  losses shown in Equation (3.12) and Equation (3.13) will motivate generators not to change the input image if it is from the desired type and maintain color consistency.

$$\mathcal{L}_{idt-s} = \|\hat{I}_{si} - I_s\|_1 \quad (3.12)$$

$$\mathcal{L}_{idt-sf} = \|\hat{I}_{fi} - I_f\|_1 \quad (3.13)$$

Here,  $\hat{I}_{si}$  (generated as  $G_s(I_s, M_x)$ ) and  $\hat{I}_{fi}$  (generated as  $G_f(I_f)$ ) are the generated identity shadow and shadow-free images, respectively. Mask  $M_x$  with all zero values used as the input of  $G_s$ , along with the real shadow image  $I_s$ . Ideally, if the shadow-free image is given as input to generator  $G_f$ , it should generate the same image. Similarly, if the shadow image is given as input to  $G_s$ , it should generate the same image.

### 3.4.4 Content Losses

The  $L_1$  content constraints defined in Equation (3.14) and Equation (3.15) encourages generators to produce images that are closer to the ground-truth images.

$$\mathcal{L}_{cont-s} = \|\hat{I}_{s*} - I_{s*}\|_1 \quad (3.14)$$

$$\mathcal{L}_{cont-sf} = \|\hat{I}_{f*} - I_{f*}\|_1 \quad (3.15)$$

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.

### 3.4.5 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:

$$\begin{aligned} \mathcal{L}_G = & \lambda_1(\mathcal{L}_{gan-s(G)} + \mathcal{L}_{gan-sf(G)}) + \lambda_2(\mathcal{L}_{cyc-s} + \mathcal{L}_{cyc-sf}) \\ & + \lambda_3(\mathcal{L}_{idt-s} + \mathcal{L}_{idt-sf}) + \lambda_4(\mathcal{L}_{cont-s} + \mathcal{L}_{cont-sf}) \end{aligned} \quad (3.16)$$

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

### 3.4.6 Loss Function for Discriminators

The discriminator loss for the shadow-free discriminator  $D_f$  and shadow discriminator  $D_s$  in the proposed method are shown in Equation (3.17) and Equation (3.18), respectively.

$$\mathcal{L}_{D_f} = \lambda_5(\mathcal{L}_{gan-sf(D)}) \quad (3.17)$$

$$\mathcal{L}_{D_s} = \lambda_5(\mathcal{L}_{gan-s(D)}) \quad (3.18)$$

Here,  $\lambda_5$  is appropriately chosen weight.

## 3.5 Network Architecture and Training Strategy

We use the model of Johnson *et al.* [27] 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 Rectified Linear Unit (ReLU) activation function.

For the discriminator network, we use PatchGAN [18], 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 [28] with a learning rate of 0.0002, with first and second momentum as 0.5 and 0.999, is adopted during training. A zero-mean Gaussian distribution with a standard deviation of 0.02 initializes network parameters.

For data augmentation during training, images are resized to  $286 \times 286$  and randomly cropped to  $256 \times 256$ . We have chosen the median range image size  $256 \times 256$  for the processing since the size less than  $256 \times 256$  will result in information loss, and the size greater than  $256 \times 256$  will result in high computational cost. The network is trained for 200 epochs keeping the mini-batch size as 1 with the PyTorch module and NVIDIA GeForce-RTX2080-Ti GPU. In the experiments, the hyperparameters  $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5$  are set as 1, 10, 5, 5, 0.5. The models take 2~2.5 days of training to converge.

## 3.6 Benchmark Dataset Correction

### 3.6.1 Importance of Correcting Benchmark Dataset

Ideally, in the benchmark dataset for the shadow removal task, the shadow-free area of shadow and corresponding shadow-free image should be the same. Still, 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 [20], 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 3.3 shows the sample triplets from the ISTD dataset, where the difference in the non-shadow area is clearly visible.

Supervised models are trained to give output close to the ground-truth shadow-free image, and accordingly, the loss function is made, and models are trained. However, methods give the color, brightness, and contrast inconsistent output compared to the non-shadow area of the shadow image. Hence it is essential to correct those ground-truth shadow-free images to achieve better results.

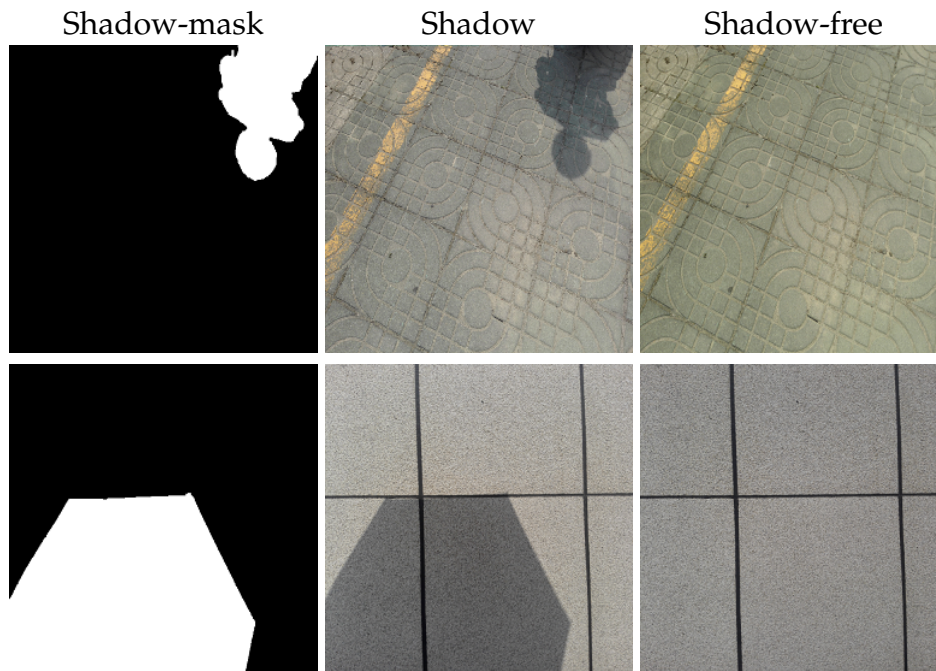


Figure 3.3: ISTD triplets, showing issue in non-shadow area.

Figure 3.4 shows the impact of using an incorrect ground-truth database during training. Our method tends to generate output that is close to the ground-truth shadow-free image. Since there is an inconsistent non-shadow area in the benchmark dataset, our approach also ends up providing an inconsistent non-shadow result compared with the input shadow image.



Figure 3.4: Impact of using incorrect benchmark dataset during training.

### 3.6.2 Methodology for Benchmark Dataset Correction

To correct the ground-truth shadow-free images, we processed each image individually and used the regression technique that transforms the pixel value of the non-shadow area of a shadow-free image into the corresponding pixel value of the

shadow image. Figure 3.5 shows the regressor learning, where the non-shadow area of shadow and shadow-free images were used as data to train the regressor, and finally, after training, a shadow-free image was given to the regressor to provide the corrected shadow-free image.

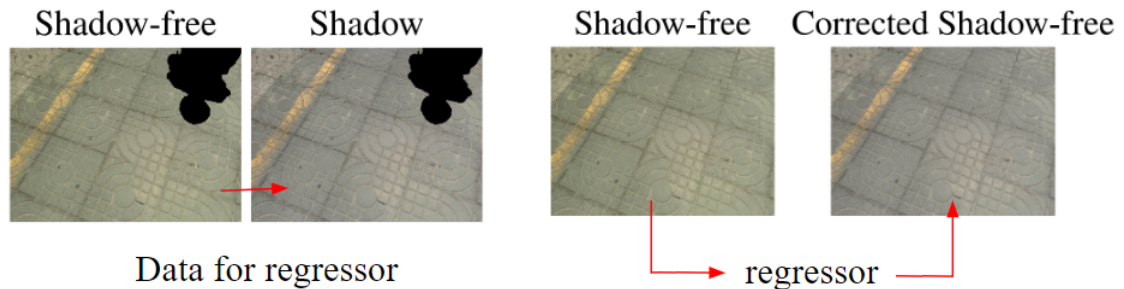


Figure 3.5: Regressor learning dataset correction.

We conducted various experiments by using three well-known regressors, Linear Regressor (LR), Decision-Tree Regressor (DTR), and K-Nearest-Neighbor Regressor (KNNR). We experimented with RGB and LAB color space. 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 multioutput regressor in RGB color space for the benchmark dataset correction.

## CHAPTER 4

# Experimental Results

In this chapter, we show experimental results of the proposed method on two benchmark datasets. We also show the experimental results of different benchmark dataset correction methods and subsequently show observed results of the proposed method on the corrected benchmark dataset. At the end of this chapter, we also show the results of various ablation studies.

### 4.1 Description of Datasets

To analyze the performance of the suggested framework, we experimented with the dataset containing image shadow triplets termed as ISTD [20] and trained models accordingly. ISTD has 1870 triplets of shadow, shadow-mask, and shadow-free image, with 1330 image triplets in the training split and 540 in the testing split. We also experimented with the shadow removal dataset termed as SRD [11] dataset, which contains 3088 pairs of shadow, and shadow-free image, with 2680 pairs for training and 408 pairs for testing purposes.

### 4.2 Evaluation Parameters

We followed [14, 19, 20] and used Balance Error Rate (BER) as an evaluation metric for a quantitative comparison of shadow detection. BER for a shadow-mask image is calculated as:

$$\text{BER} = 1 - \frac{1}{2} \left( \frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right) \quad (4.1)$$

where,

- True Positive ( $TP$ ) denotes, Predictive model labeled the pixel as a shadow, and actually, it is a shadow.

- True Negative ( $TN$ ) denotes, Predictive model labeled the pixel as a non-shadow, and actually it is a non-shadow.
- False Negative ( $FN$ ) denotes, Predictive model labeled the pixel as a non-shadow, and actually it is a shadow.
- False Positive ( $FP$ ) denotes, Predictive model labeled the pixel as a shadow, and actually it is a non-shadow.

For the quantitative assessment of shadow removal, we followed recent procedures [20, 22, 15, 2, 24] and used Root Mean Square Error (RMSE) in LAB color space computed among the ground-truth and produced shadow-free images. We resized all images to  $256 \times 256$  for a fair comparison.

Additionally, we calculated the RMSE value on the four scenarios: RMSE value by comparing the resulting shadow-free image  $\hat{I}_{fc*}$  with the ground-truth shadow-free image  $I_{f*}$  for all pixels (represented with **O**), for pixels in the shadow region (represented with **S**), for pixels in the non-shadow region (represented with **SF**), and by comparing  $\hat{I}_{fc*}$  with input shadow image  $I_s$  for pixels in the non-shadow region (represented with **SF-I**). The best and second-best results in all result tables are highlighted in bold and blue, respectively.

## 4.3 Evaluation on Shadow Removal

### 4.3.1 Evaluation on ISTD Dataset

We compare the shadow removal performance of the proposed method (trained with ISTD [20]) with the methods [8, 9, 7, 20, 2, 24, 22] on the test dataset of ISTD. The results are shown in Table 4.1, where the lowest RMSE value denotes a better one. Our method achieves the best performance in the **O** and **SF** scenarios and the second-best performance in **S** and **SF-I** scenarios. Although TCGAN [24] achieves the best result in **SF-I**, it has poor performance in **S**. Similarly, DSC [22] achieves the best result in **S** but performs poorly in **SF** and **SF-I**. Our approach achieves comparable results in all aspects and gets the best overall value **O**, compared to all other methods.

Figure 4.1 shows visual performance compared to methods [20, 2]. While STCGAN [20] suffers from color-inconstancy and random artifacts, and MaskShad-



Table 4.1: Quantitative shadow removal results with RMSE on ISTD test dataset.

Method	O	S	SF	SF-I
Original	10.97	32.67	6.83	0
Yang [8]	15.63	19.82	14.83	-
Gong [9]	9.3	18.95	7.46	-
Guo [7]	8.53	14.98	7.29	-
ST-CGAN [20]	7.47	10.33	6.93	7.45
Mask-ShadowGAN [2]	6.99	11.41	6.17	6.75
TCGAN [24]	6.85	11.49	5.91	6.29
DSC [22]	6.67	9.22	6.39	6.61
Ours	6.54	10.03	5.88	6.49

owGAN [2] has an over-exposure problem, particularly in the non-shadow area, our approach handles those issues and produces better visual output.

### 4.3.2 Evaluation on SRD Dataset

We compare the shadow removal performance of the proposed method (trained with SRD [11]) with the methods [8, 7, 9, 23, 2, 11, 22] on the test dataset of SRD. The results are shown in Table 4.2. Our method achieves the best performance in the overall **O** scenario. Note that we compared quantitative results only on **O** scenario, and scenarios **S**, **SF**, **SF-I** cannot be compared since ground-truth shadow-mask images are not available in this dataset.

Table 4.2: Quantitative shadow removal results with RMSE on SRD test dataset.

Method	O
Original	14.41
Yang [8]	22.57
Guo [7]	12.60
Gong [9]	8.73
CycleGAN [23]	9.14
Mask-ShadowGAN [2]	7.32
DeshadowNet [11]	6.64
DSC [22]	6.21
Ours	5.63

Figure 4.2 shows the visual performance of our method with the ground-truth data. Our approach produces a shadow-free result close to the ground-truth shadow-free image.

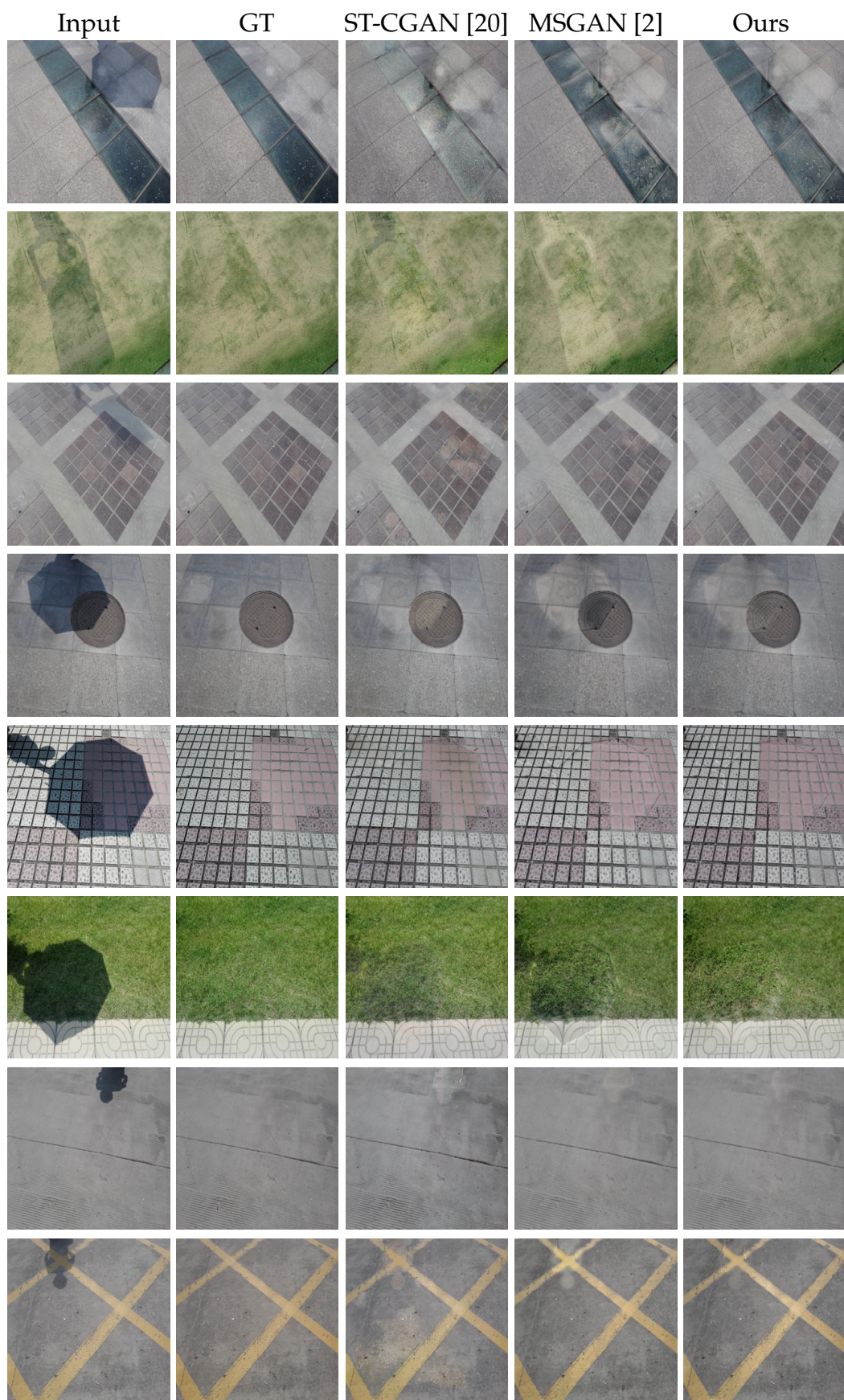


Figure 4.1: Visual comparison of shadow removal results on ISTD test dataset.

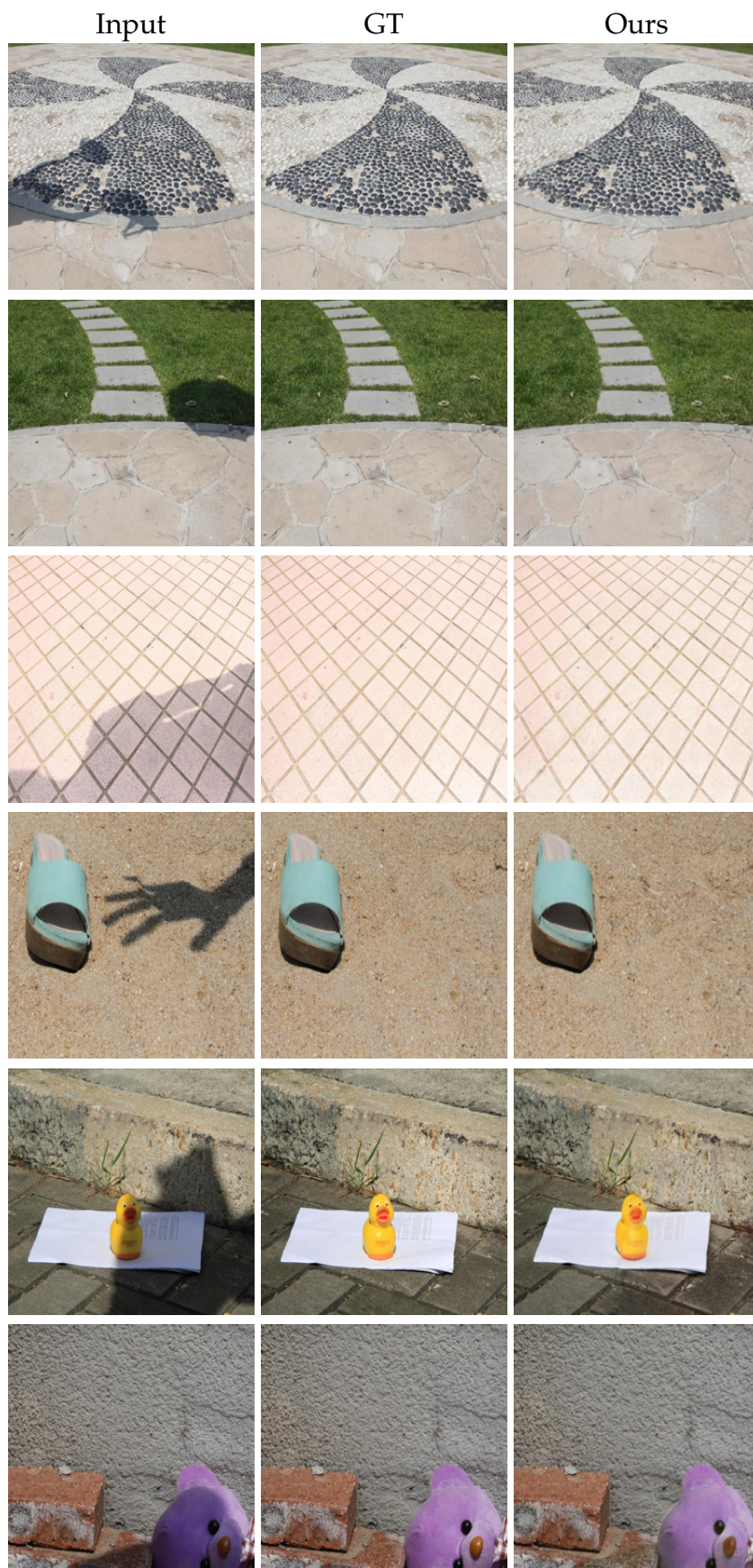


Figure 4.2: Visual performance of shadow removal results on SRD test dataset.

## 4.4 Evaluation on Shadow Detection

We evaluate the shadow detection performance of our method (trained with ISTD [20]) with the recent methods [19, 14, 20, 2] on the ISTD test dataset. The quantitative results are shown in Table 4.3. The proposed method outperforms the baseline Mask-ShadowGAN [2] and methods [14, 19]. Methods in [19, 20] achieve better results since these methods specifically train models 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 Section 3.2.

Table 4.3: Quantitative results of shadow detection with BER(%) on ISTD test dataset.

Method	BER
CGAN [19]	9.64
StackedCNN [14]	8.6
SCGAN [19]	4.7
ST-CGAN [20]	3.85
Mask-ShadowGAN [2]	7.66
Ours	6.48

Figure 4.3 shows the visual performance compared to state-of-the-art MaskShadowGAN [2]. Our approach produces a shadow-mask result close to the ground-truth shadow-mask image.

## 4.5 Benchmark Dataset Correction

To correct 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 space. While doing regression in LAB color space, both shadow and shadow-free images were transferred to the LAB image from the RGB image, and after doing regression and correction, they again moved back to the RGB image. Also, we have done experiments by using a regression for each individual color channel (there will be three 1 input to 1 output regressor) and by using a regression for combine color-channel (multioutput regressor) (there will be one 3 input to 3 output regressor). The results of the experiments are shown in Table 4.4.

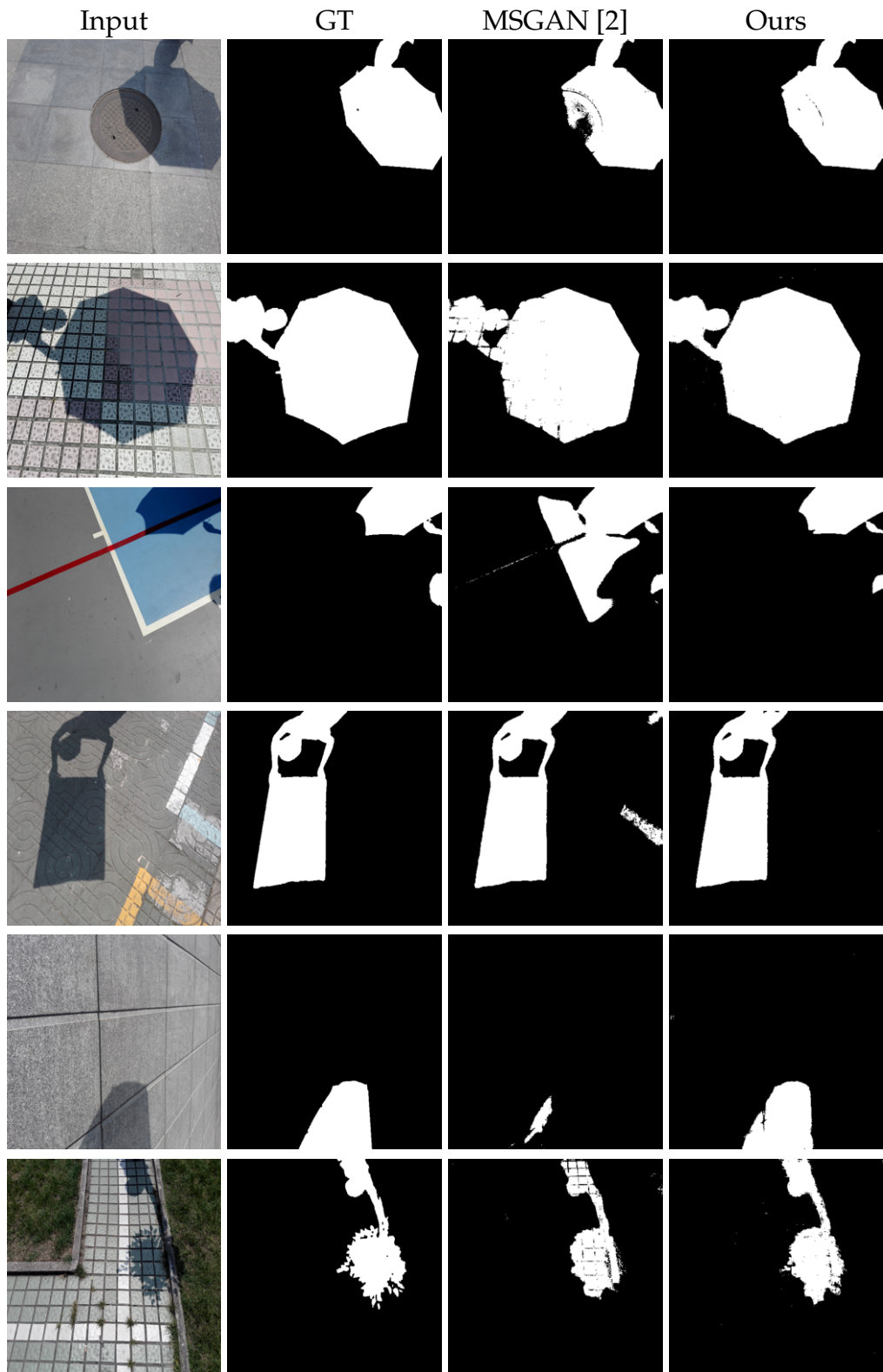


Figure 4.3: Visual comparison of shadow detection results on ISTD test dataset.

Table 4.4: Quantitative results of ISTD test dataset correction with RMSE.

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

Decision-tree combine channel repressor in RGB color space has lower RMSE value in **O** and **SF** scenarios. So finally, we used that method and created a new corrected ISTD training and testing dataset. Figure 4.4 shows the visual output of this database correction task by using the selected method.

## 4.6 Evaluation on Shadow Removal with Corrected Benchmark ISTD Dataset

We compare the shadow removal performance of the proposed method with the methods [20, 2], trained and tested on the corrected dataset of ISTD [20]. Since the official code for the ST-CGAN method [20] is not available, we use the community code [29] for evaluating purpose. The results are shown in Table 4.5. The proposed method achieves the best performance in **O**, **S**, and **SF** scenarios compared to state-of-the-art methods.

Table 4.5: Quantitative shadow removal results with RMSE, trained and tested on corrected ISTD.

Method	<b>O</b>	<b>S</b>	<b>SF</b>	<b>SF-I</b>
Original	7.56	39.06	1.66	0
ST-CGAN [29]	8.79	11.35	8.31	4.28
MSGAN [2]	<b>4.47</b>	<b>10.13</b>	<b>3.41</b>	<b>3.18</b>
Ours	<b>4.36</b>	<b>9.52</b>	<b>3.40</b>	<b>3.19</b>

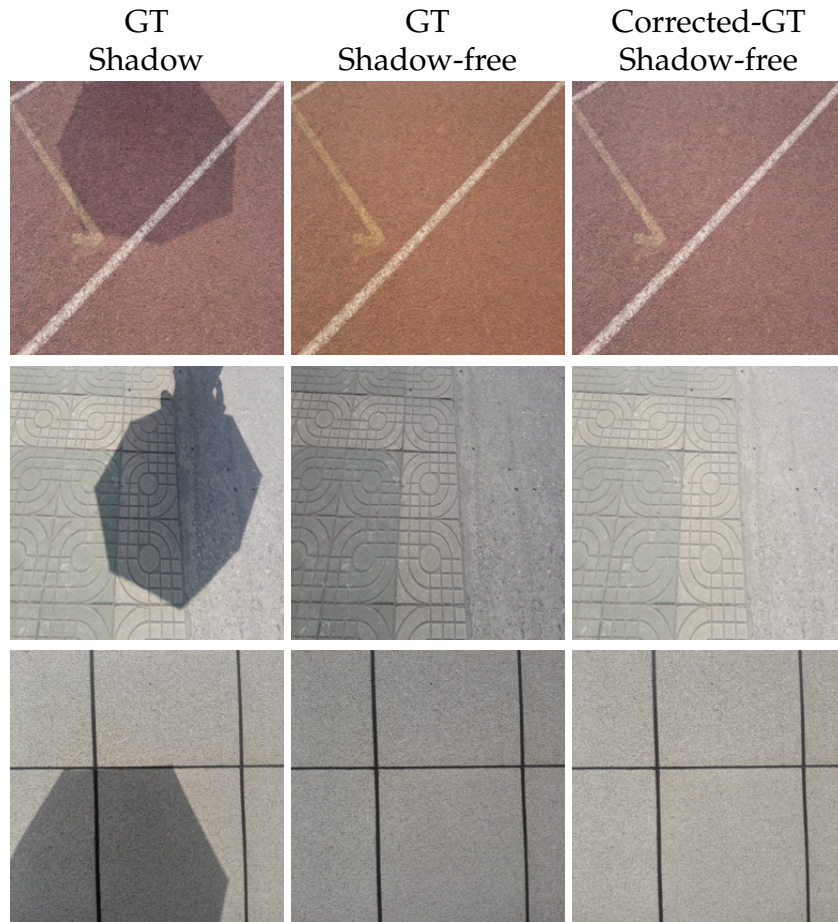


Figure 4.4: Visual results of ISTD dataset correction task.

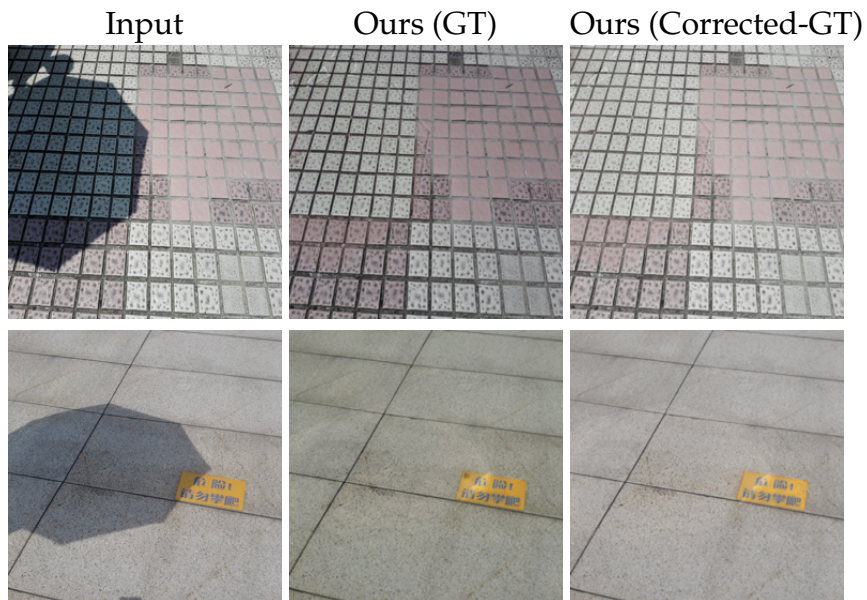


Figure 4.5: Visual comparison on GT and corrected GT training.

We also show the visual results of the proposed method in Figure 4.5, trained on both ground-truth and corrected ground-truth ISTD datasets. From the results, it is clear that the corrected ground-truth dataset helps to produce brightness consistent, contrast consistent, and color consistent output and tries to keep the non-shadow area of the original shadow image as it is in the final shadow-free output production.

## 4.7 Ablation Studies

We have done an ablation study on the proposed 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 Section 3.2. Removal and detection results for all the experiments are shown in Table 4.6. Visual performance for *-c* is shown in Figure 4.6. 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 4.6: Ablation study on exposure correction module and ground-truth data.

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

We also have done an ablation study on the adversarial loss of the presented method. The state-of-the-art MaskShadowGAN [2] method uses Binary Cross Entropy (BCE) loss as an adversarial loss, and our method uses Mean Squared Error (MSE) loss as an adversarial loss. We experimented with our method by using BCE adversarial loss (represented by *Ours (BCE)*). The results are in Table 4.7, showing our method with MSE adversarial loss (represented by *Ours (MSE)*) achieves the best performance.



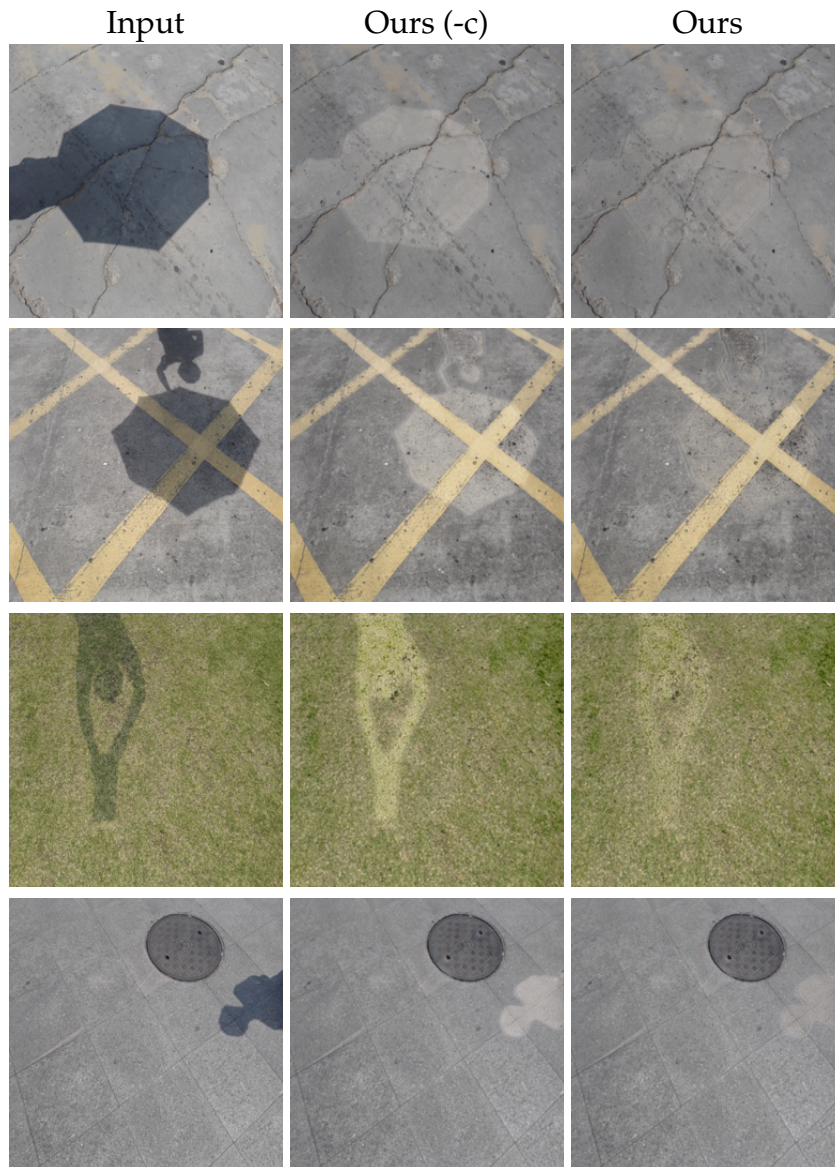


Figure 4.6: Visual performance of exposure correction module’s ablation study.

Table 4.7: Ablation study on adversarial loss.

Aspect	Removal				Detection
	<b>O</b>	<b>S</b>	<b>SF</b>	<b>SF-I</b>	<b>BER</b>
Ours (MSE)	<b>6.54</b>	<b>10.03</b>	<b>5.88</b>	<b>6.49</b>	<b>6.48</b>
Ours (BCE)	<b>6.68</b>	<b>10.29</b>	<b>6.01</b>	<b>6.69</b>	<b>6.97</b>

Table 4.8: Ablation study on using additional shadow mask as an input to shadow-free generator.

<b>Method</b>	<b>O</b>	<b>S</b>	<b>SF</b>	<b>SF-I</b>
Ours	6.54	10.03	5.88	6.49
FusionNet [15]	<b>5.95</b>	<b>7.82</b>	<b>5.60</b>	<b>6.77</b>
Ours (+sm)	<b>5.72</b>	<b>8.87</b>	<b>5.13</b>	<b>5.83</b>

For making a comparison with the method Auto-Exposure-FusionNet [15], which uses shadow and shadow-mask images as input during training and testing, we additionally used shadow mask as an input to Generator  $G_f$  in the proposed framework. The comparison result is depicted in Table 4.8, which shows our method with an additional shadow mask (represented by *Ours (+sm)*) performs better in **O**, **SF**, and **SF-I** scenarios.

We also experimented with the discriminator learning technique, in which we used only one discriminator instead of two discriminators, that will discriminate whether a pair of shadow image and shadow-free image is a real one (comes from the ground-truth training data) or a fake one (comes from the generator). The results of the experiment are shown in Table 4.9, where the approach with one discriminator performs well in the **S** region, while the original approach with two individual discriminators performs well in the **O**, **SF**, and **SF-I** aspects.

Table 4.9: Ablation study on discriminator learning.

Method	<b>O</b>	<b>S</b>	<b>SF</b>	<b>SF-I</b>
Ours	<b>6.54</b>	<b>10.03</b>	<b>5.88</b>	<b>6.49</b>
Ours (with 1 discriminator)	<b>6.86</b>	<b>9.83</b>	<b>6.31</b>	<b>6.66</b>

These ablation studies show the importance of ground-truth data with the over-exposure correction module in the proposed framework. Also, it shows the effectiveness of the selected learning strategy in generating better quality results.

## CHAPTER 5

# Conclusions and Future Work

## 5.1 Conclusions

We proposed a method based on a Generative Adversarial Network (GAN) to address the shadow removal task from 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 correcting those benchmark datasets to additionally improve the shadow removal results. We also conducted various experiments to analyze the importance of ground-truth data and the exposure correction module in the presented method. In addition, we experimented with different learning strategies like changing adversarial loss function, adding an extra shadow mask as an input to the generator, changing discriminator learning, etc., to study the proposed method's behavior for generating better quality output.

## 5.2 Future Work

The proposed approach generates the shadow-free result effectively, but still, the shadow is not appropriately removed on the shadow edges. Figure 5.1 shows the edge issue, where the shadow is clearly visible on the edges in the generated shadow-free results. In the future, a refinement module can be introduced to handle the shadow edge issue to obtain trace-less shadow-free results.



Figure 5.1: Shadow edge issue.

One discriminator learning strategy, as described in Section 4.7, achieves the best result in the shadow area compared with state-of-the-art methods along with the proposed method. At the same time, our approach earns the best result in the rest areas. The fusion method can be introduced in the future, which takes advantage of one discriminator learning strategy for the shadow area, and two discriminator learning strategy for the rest areas, to achieve the overall best performance.

Also, the proposed method can be tested on other shadow datasets with more evaluation parameters. Furthermore, this work can be extended for automated shadow detection and removal in video applications.

## References

- [1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer vision and Pattern Recognition (CVPR)*, pages 779–788, 2016.
- [2] X. Hu, Y. Jiang, C. Fu, and P. Heng. Mask-ShadowGAN: Learning to remove shadows from unpaired data. *IEEE International Conference on Computer Vision (ICCV)*, pages 2472–2481, 2019.
- [3] A. Sanin, C. Sanderson, and B. C. Lovell. Shadow detection: A survey and comparative evaluation of recent methods. *Pattern Recognit.*, 45(4):1684–1695, 2012.
- [4] A. Prati, I. Mikic, M.M. Trivedi, and R. Cucchiara. Detecting moving shadows: algorithms and evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7):918–923, 2003.
- [5] CJ. Chang, WF. Hu, JW. Hsieh, and YS. Chen. Shadow elimination for effective moving object detection with gaussian models. In *International Conference on Pattern Recognition*, volume 2, pages 540–543, 2002.
- [6] A. Leone and C. Distanto. Shadow detection for moving objects based on texture analysis. *Pattern Recognition*, 40(4):1222–1233, 2007.
- [7] R. Guo, Q. Dai, and D. Hoiem. Paired regions for shadow detection and removal. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(12):2956–2967, 2013.
- [8] Q. Yang, K. Tan, and N. Ahuja. Shadow removal using bilateral filtering. *IEEE Trans. Image Process.*, 21(10):4361–4368, 2012.
- [9] H. Gong and D. Cosker. Interactive shadow removal and ground truth for variable scene categories. In *Proceedings of the British Machine Vision Conference*, 2014.

- [10] M. Khare, RK. Srivastava, and A. Khare. Moving shadow detection and removal—a wavelet transform based approach. *IET Computer Vision*, 8(6):701–717, 2014.
- [11] L. Qu, J. Tian, S. He, Y. Tang, and R. W. Lau. DeshadowNet: A multi-context embedding deep network for shadow removal. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4067–4075, 2017.
- [12] N. Bansal, Akashdeep, and N. Aggarwal. Deep learning based shadow detection in images. In *Proceedings of 2nd International Conference on Communication, Computing and Networking*, pages 375–382. Springer Singapore, 2019.
- [13] H. Fan, M. Han, and J. Li. Image shadow removal using end-to-end deep convolutional neural networks. *Applied Sciences*, 2019.
- [14] T. F. Y. Vicente, L. Hou, C. Yu, M. Hoai, and D. Samaras. Large-scale training of shadow detectors with noisily-annotated shadow examples. *European Conference on Computer Vision (ECCV)*, pages 816–832, 2016.
- [15] L. Fu, C. Zhou, Q. Guo, F. Juefei-Xu, H. Yu, W. Feng, Y. Liu, and S. Wang. Auto-exposure fusion for single-image shadow removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10571–10580, 2021.
- [16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27:2672–2680, 2014.
- [17] M. Mirza and S. Osindero. Conditional generative adversarial nets. *arXiv:1411.1784*, 2014.
- [18] P. Isola, JY. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976, 2017.
- [19] V. Nguyen, T. F. Y. Vicente, M. Zhao, M. Hoai, and D. Samaras. Shadow detection with conditional generative adversarial networks. *IEEE International Conference on Computer Vision (ICCV)*, pages 4510–4518, 2017.
- [20] J. Wang, X. Li, and J. Yang. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1788–1797, 2018.

- [21] T. Nagae, R. Abiko, T. Yamaguchi, and M. Ikehara. Shadow detection and removal using GAN. *Proc. 28th European Signal Processing Conference (EU-SIPCO)*, pages 630–634, 2021.
- [22] X. Hu, C. Fu, L. Zhu, J. Qin, and P. Heng. Direction-aware spatial context features for shadow detection and removal. *IEEE Trans. Pattern Anal. Mach. Intell.*, 42(11):2795–2808, 2020.
- [23] JY. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *IEEE International Conference on Computer Vision (ICCV)*, pages 2223–2232, 2017.
- [24] C. Tan and X. Feng. Unsupervised shadow removal using target consistency generative adversarial network. *arXiv:2010.01291*, 2020.
- [25] X. Hu, T. Wang, CW. Fu, Y. Jiang, Q. Wang, and PA. Heng. Revisiting shadow detection: A new benchmark dataset for complex world. *IEEE Transactions on Image Processing*, 30:1925–1934, 2021.
- [26] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.*, 9(1):62–66, 1979.
- [27] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. *European Conference on Computer Vision (ECCV)*, pages 694–711, 2016.
- [28] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv:1412.6980*, 2014.
- [29] ST-CGAN. [https://github.com/IsHYuhi/ST-CGAN\\_Stacked\\_Conditional\\_Generative\\_Adversarial\\_Networks](https://github.com/IsHYuhi/ST-CGAN_Stacked_Conditional_Generative_Adversarial_Networks).