Shadow Removal using Attention-based GANs

By

Mavrah Shahid

MSEE-2K20-DSSP 359601

Supervisor

Dr. Ahmed Salman

Department of Electrical Engineering

A thesis submitted in partial fulfillment of the requirements for the degree of Masters of Science in Electrical Engineering (MS EE)

In

School of Electrical Engineering & Computer Science (SEECS) ,

National University of Sciences and Technology (NUST),

Islamabad, Pakistan.

(May 2024)

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis entitled "Shadow Removal Using Attention-Based GANs" written by Mavrah Shahid, (Registration No 00000359601), of SEECS has been vetted by the undersigned, found complete in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS/M Phil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature: ______________________

Name of Advisor: Dr. Ahmad Salman

Date: 04-Jun-2024

HoD/Associate Dean: ______________________

Date: 04-Jun-2024

Signature (Dean/Principal): ______________________

Date: 04-Jun-2024

Approval

It is certified that the contents and form of the thesis entitled "Shadow Removal Using Attention-Based GANs" submitted by Mavrah Shahid have been found satisfactory for the requirement of the degree

Advisor : Dr. Ahmad Salman

Signature: ______________ ________

Date: 04-Jun-2024

Committee Member 1:Dr. Wajid Mumtaz

Signature: _ ________

04-Jun-2024

Committee Member 2:Dr. Salman Abdul Ghafoor

Signature: ______________ ________

Date: 04-Jun-2024

Signature: ______________________

Date: ________________________

FORM TH-4

National University of Sciences & Technology

MASTER THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by: (Student Name & Reg. #) Mavrah Shahid [00000359601]

Titled: Shadow Removal Using Attention-Based GANs

be accepted in partial fulfillment of the requirements for the award of Master of Science (Electrical Engineering) degree.

Examination Committee Members

1. Name: Wajid Mumtaz — Signature: 20-Jun-2024 12:04 PM

2. Name: Salman Abdul Ghafoor — Signature: 20-Jun-2024 12:04 PM

Supervisor's name: Ahmad Salman — Signature: 21-Jun-2024 12:42 AM

Salman Abdul Ghafoor
HoD / Associate Dean

20-June-2024
Date

COUNTERSINGED

26-June-2024
Date

M Ajmal
Muhammad Ajmal Khan
Principal

THIS FORM IS DIGITALLY SIGNED

Dedication

To my beloved **father and mother**,

This thesis is dedicated to you both, for your unwavering support, boundless love, and relentless encouragement throughout my educational journey. Your financial and emotional backing have been invaluable, providing me with the strength and determination to overcome challenges and pursue my dreams.

Thank you for being my pillars of strength, for believing in me even when I doubted myself, and for instilling in me the confidence to persevere through every obstacle. Your guidance and belief in my abilities have been instrumental in shaping my academic success.

I am forever grateful for your sacrifices, sacrifices you made to ensure my education, and for the countless sacrifices you continue to make for my well-being. This thesis is a testament to your enduring love and support, and I dedicate it to you with all my heart.

With deepest love and gratitude,

Mavrah Shahid

Certificate of Originality

I hereby declare that this submission titled "Shadow Removal Using Attention-Based GANs" is my own work. To the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any degree or diploma at NUST SEECS or at any other educational institute, except where due acknowledgement has been made in the thesis. Any contribution made to the research by others, with whom I have worked at NUST SEECS or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except for the assistance from others in the project's design and conception or in style, presentation and linguistics, which has been acknowledged. I also verified the originality of contents through plagiarism software.

Student Name:Mavrah Shahid

Student Signature:__________ _

Acknowledgments

All praises to Allah Almighty for giving me the strength and willingness to complete my thesis. Nothing is possible without His blessings. In particular,I would like to express gratitude to my Supervisor Dr. Ahmed Salman for their excellent guidance and continuous support. I am very fortunate that they have supervised my work. Our consistent meetings were filled with intellect, cooperation and a very friendly environment. I am very lucky to be a part of Erays Technologies, which is a wonderful place to work at. My Research work wouldn't have been possible without the cooperation of Mr. Zoolnurain and Mr. Hamza Tariq, who took the time out of their busy schedule to help me out.

Mavrah Shahid

Table of Contents

List of Figures

List of Tables

Abstract

Shadows are natural artifacts present in images that can hinder various Computer Vision tasks such as object detection, tracking, segmentation, and scene analysis. This research introduces an innovative approach to detect and remove shadows from single RGB images using Attention-based Generative Adversarial Networks (GANs). The proposed methodology employs a deep-learning model comprising Attention-based GANs, featuring two generators and two discriminators, to effectively identify and eliminate shadows. Subsequently, the shadow-free image generated by the GANs undergoes a post-processing step to refine shadow regions using a shadow mask. This post-processing stage combines traditional image processing techniques, including histogram matching, custom filters, and shadow boundary detection and estimation, to enhance the accuracy of shadow removal. Additionally, we used a large-scale benchmark dataset named "Extended ISTD," consisting of 5352 triplet images (shadow, shadow mask, shadow-free samples), facilitating both shadow detection and removal tasks. This dataset encompasses a diverse range of dark and hard shadow images, as well as multi-color contrast shadow images, serving as an extended version of the publicly available "ISTD Dataset." Upon training the Attention-based GANs on the provided dataset and applying the proposed post-processing step, an RMSE of 5.28 is achieved. The proposed methodology demonstrates efficient shadow removal capabilities, even in scenarios involving dark, hard shadows, and multi-color contrast shadow images.

CHAPTER 1

Introduction

Artefacts are the unwanted and undesired areas which are present in the image and tends to degrade the quality of entire image. Artefacts in digital images may results from the inner working of the digital camera such as blooming, aliasing, compression, noise etc. or artefacts may also arise in nature such as shadow, fog, haze, smoke, smog and many more. Presence of any kind of artefacts either from inner working of digital camera or from the nature, both of these tends to downgrade the quality of captured image. This downgraded image if further processed or feed into the deep-learning or computer-vision based application may tends to decrease the efficiency of the algorithm and hence increases the probability of miss-classification accuracy of deep-learning models. This is due to the fact that such deep learning based models are highly sensitive to the data on which they are trained.

If such unwanted artefacts are present in the training dataset than the deep-learning model might start learning these unwanted features or such unwanted features if present in real-time during the inference phase may tends to decrease the efficiency and accuracy of the models. So it is of core importance to perform some processing on images and such artefacts must be removed in order to achieve the state-of-the-art efficiency and accuracy of deep-learning models.

1.1 Shadow and Its Types

Shadows are the unwanted and undesired artefacts which are present in nature. Shadow is a darker region/area formed when the source of light is either partially or completely blocked. Shadows are formed when the light falls on an object, and is blocked by that object. The size and shape of the shadow formed depends upon the shape, position and the size of the object as well as the direction and intensity of light coming from the particular light source. Depending upon the dark area/region formed due to blockage of light, shadow can be further classified into three main categories i.e., Hard-Shadows, Soft-Shadows and Umbra/Penumbra. Different types of these shadows are discussed as below:

- **Hard-Shadow:** It is the dark area formed when the source of light is completely blocked. In case of hard shadow the surface texture is entirely vanished.

Figure 1.1: Dark/Hard Shadow; Surface texture is completely vanished

- **Soft-Shadow:** It is the area formed when the source of light is partially blocked. In case of soft shadows the background surface texture is partially vanished.

- **Umbra/Penumbra:** They are the two main parts of shadows. Umbra is darker region of shadow and occurs at the bottom area of shadow. On the other hand, penumbra is the lighter region of shadow and is mostly present at shadow boundaries.

Figure 1.2: Soft Shadow; Surface texture is partially vanished

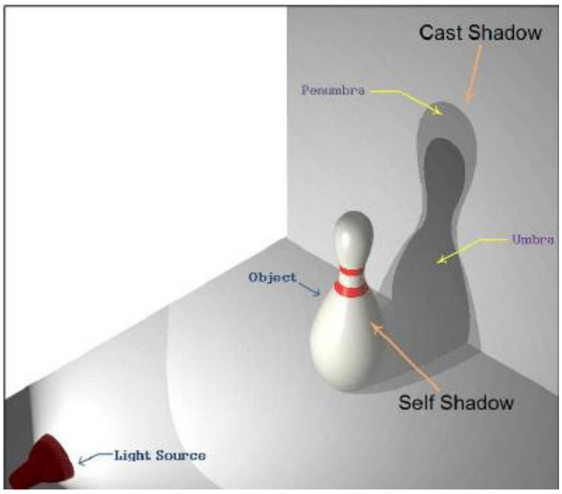
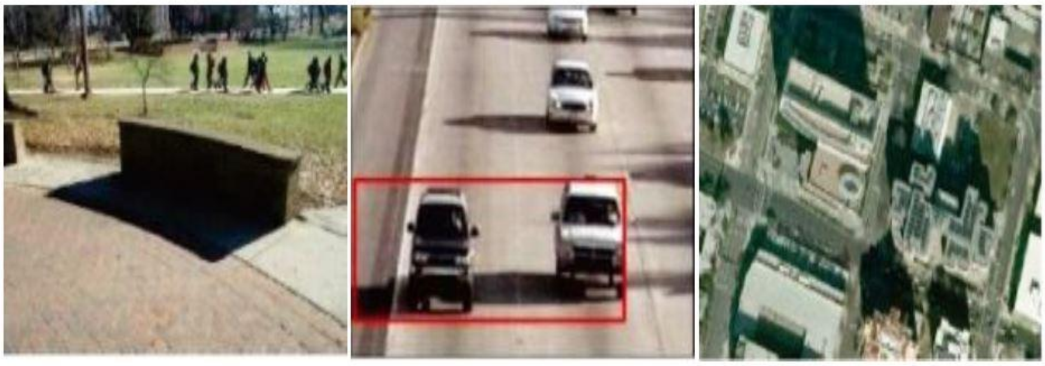


Figure 1.3: Umbra; darker region of shadow exists at bottom region of shadow, Penumbra; usually exist at shadow boundaries

1.2 Challenges Posed by Shadows

Shadows, if present in images may leads to decline the accuracy as well as efficiency of various deep-learning and computer-vision based algorithms. Presence of shadows in images can cause several issues such object merging, object losing, misinterpretations of remote sensing images and many more.

Figure 1.4 illustrates various issues that may be caused because of the presence of shadows in images. Figure 1.4a, consist of a very dark/hard shadow, in case if any



a. Object Detection b. Object Merging c. Mis-Interpretation of remote sensing images

Figure 1.4: Challenges in computer vision: navigating Shadow effects in images

object of interest is lying under the shadow area, it will be extremely difficult for an object detection algorithm to detect and recognise the object. Presence of shadow may also cause object merging as in Figure 1.4b, two vehicles in parallel are detected as a single vehicle because of the shadow in between. Shadows if present in remote sensing images may cause several misinterpretations. Similarly, shadows if present in images or video frames can be a problem for computer vision applications and algorithms. few of them are discussed in detail as below:

1.2.1 Object Detection:

Shadow if presents in images tends to produce false negative and false positive in the task of object detection. When a shadow is cast by an object, there is a high probability that the shadow might be detected as a separate object by the algorithm. Contrary to this, the object of interest may also not be detected if the presence of targeted object is significantly altered because of the presence of shadow.

1.2.2 Object Recognition:

Shadow if present in images or video frames can also highly affect the performance of object recognition task. The goal of object recognition task is to identify the type of object appearing in an image or video frame. Shadows, if present can alters the appearance of the object which may further leads the algorithm in failing to recognize it correctly.

1.2.3 Object Tracking:

The presence of shadows may also cause several problems in object tracking related tasks. If in case the desired object has moved in and out of the shadow, its appearance might change. This change makes it difficult for an object tracking algorithm to track it consistently.

1.2.4 Scene Understanding:

The capacity of computer vision algorithms to comprehend the scene in an image or video frame can also be impacted by shadows. It can be more challenging for the algorithm to correctly understand the image when there are shadows since they can hide crucial information or alter the illumination in the scene.So, keeping in account the various different issues caused by the presence of shadow, it is an important task to detect and remove shadows from images. Removing shadows from images also improves the visual effect of images so that various deep-learning and computer-vision based applications such as object detection, object classification, object tracking and many more can be carried out with high efficiency and accuracy.

1.3 Problem Definition

Shadows are the artefacts that frequently occurs in image and can highly affect the quality of images. Presence of shadows may result in distortion, loss of information, reduction in illumination/contrast etc., all of these issues make it a challenging task for a computer-vision and deep-learning based applications to correctly categorise and interpret the details in image. Shadow detection and removal is an important task for numerous computer vision based applications.

Although there has been significant research on shadow detection and removal, still it is an extremely difficult and challenging task to correctly detect, remove and efficiently predict the texture underneath the shadow region. This is because of the variation in illumination, different shapes and size of shadows. Existing shadow removal methods are capable of detecting and removing shadow only in case of soft shadows. In case of hard shadows, existing methodologies fails to correctly detect and remove shadows. Similarly the existing methodologies also fails to remove shadows in case of multi-color contrast images.

Therefore, there is a need for more effective and efficient methodology which is capable to detect and remove shadow in case of dark/hard shadows as well as multi-color contrast shadow images. The proposed methodology focuses on to fill this research gap of dark/hard shadows as well as multi-color contrast shadow images.

1.4 Our Contribution

The proposed research attempts to fill some gaps by addressing multiple issues and making important contributions in the field of shadow detection and removal from images. Firstly, we introduced Attention based GANs architecture for the shadow removal from a single RGB image.
Secondly, a new benchmark dataset is proposed which consist of 5352 triplet samples (shadow, shadow-mask, shadow-free) and is one of the largest publicly available dataset for shadow detection as well as for shadow removal purpose. Dark/hard shadow samples and multi color contrast shadow samples are specifically added to the dataset in order to increase the probability distribution of samples across the dataset.
Thirdly, a post processing step is also introduced which leverages the power of several different traditional image processing steps and is used to refine the shadow free images which are being predicted/generated by the deep learning based model.

1.5 Thesis Organization

The presented thesis report is organized such that Chapter 2, consist of detailed literature review about the already available methodologies and dataset for shadow detection and removal. Chapter 3, focuses on the proposed methodology in detail, this chapter also focuses on the preparation of the newly introduced dataset ”Extended ISTD” which consist of 5352 triplet samples. Chapter 4 discussed the details results and comparison of proposed methodology with other state-of-the-art methodologies already available. In the last, Chapter 5 concludes the presented research and also provides some suggestions regarding the future work, which can be carried out in the field of shadow detection and removal.

CHAPTER 2

Literature Review

This chapter of the report discusses about the various different methodologies which are already available with the purpose of shadow detection and removal. The later section of the report also discusses about the different datasets available for this purpose. Various different approaches which are already proposed for detecting and removing shadows from RGB images can be mainly divided into two main categories, i.e., traditional image processing techniques [1], [2], [3], [4], [5] and state-of-the-art deep-learning based models [6], [7], [8], [9], [10], [11], [12].

2.1 Related Work

Before the era of state-of-art deep-learning based models, traditional image processing techniques were used to detect and remove shadows from images. The drawbacks of traditional image processing algorithms was that these algorithms were not scalable and were able to produce shadow-free images only for specific scenarios. Some of the traditional techniques also depends on user input which is not feasible in different real-time scenarios and applications. In past few years deep-learning based models are emerging as they are capable of providing state-of-the-art results in terms of accuracy and efficiency, GANs [6] [7] [8] [9] are mostly used for sake of shadow detection and removal, [10] [11] tends to remove shadow by using different convolutions neural networks, [12] uses the latest concepts of transformers with the aim of shadow detection and removal via spatial attention map.

Deep-learning based models are highly sensitive to training data-set, the drawback for deep learning models is that there is no comprehensive data-set available for purpose of shadow detection and removal. Already available data-set contains few thousand images and hence covers very less probability distribution. If provided with millions of images in training data-set, deep-learning based model have ability to produce state-of-the-art result. Limitation of data-set leads to the unsatisfactory performance of the deep-learning models. This research purposes the state-of-the-art extended triplet data-set for shadow detection and removal purpose. Furthermore, the proposed method leverage's the power of deep-learning models followed by traditional image processing techniques in order to generate state-of-the-art shadow-free images while preserving the illumination, background texture, color combination and various important content present in an image.

Shadows if detected accurately and further processing applied on detected region may be able to produce the final shadow-free image. In earlier days, the main focus was to accurately detect shadows by leveraging different traditional image processing techniques. A sequence of effective approaches [13] [14] were proved efficient for detecting shadow boundaries on single RGB image. Using properties derived from shadow samples [15] dynamically generates feature space and calculates decision parameters followed by series of transformation to get the shadow-mask. Khan proposes [16], first state-of-the-art deep convolutional neural network, for the sake of shadow detection. This deep-learning model consist of 7-layer network architecture. Proposed methodologies were able to produce the desired shadow-mask accurately. Once the shadow was accurately detected, next step was to reconstruct the area lying beneath the shadow regions which can be done either by using the traditional image processing techniques or deep-learning based models.

2.1.1 Traditional Image Processing Techniques

Traditional image processing techniques can be used to get the shadow-free images. [1] and [5] both of these method uses three color lines to be manually drawn on the image to perform shadow detection and removal. Based on these three input color lines, the algorithm performs region matching. The algorithm tries to predict the texture beneath the shadow region based on region matching between the shadow and the shadow-free region of an image. Based on this region matching the algorithm tends to generate the shadow free image as an output.

Another state of the art traditional image processing based method, [2] uses the YCbCr color space and with the help of luminance Y-Channel, tries to output the shadow-free image. This methodology converts the input RGB image of shadow into YCbCr color space, computes the average of Y-Channel. All pixels intensities are compared with the average of Y-Channel, based on this comparison a binary image is generated which distinguishes between the shadow region and the shadow-free region, in this way the shadow detection is performed. Once the shadow region is detected, based on that ratio of average values of pixel intensities of shadow region and shadow-free region are computed and Cb, Cr pixels intensities are updated. The updated Cb, Cr channels is concatenated with the Y-channel and this YCbCr is converted to RGB image. Performing the color adjustment on the resultant RGB image tends to produce an image without shadows as the final result.
Similarly [3], [4] uses multi-channel thresholding for shadow detection purpose. Once the shadow is detected a shadow-matting technique is applied for sake of shadow-removal purpose. These traditional image processing methods performs well but still have few limitations. The limitations of traditional image processing techniques is that these methods are not scalable and are capable of producing shadow free images only for specific scenarios. Few of traditional techniques also depends on user input, this dependency of user input is not feasible in real time scenarios and applications.

2.1.2 Deep-Learning Based Models

In past few years deep-learning based convolutional neural networks models have emerged exponentially, as they are capable of providing state-of-the-art results in terms of accuracy and efficiency. Generative Adversarial Networks (GANs) [6], [7], [8] and [9] are most commonly used for the purpose of shadow detection and removal. GANs uses generators as well as Discriminators in order to detect and remove shadow. Generators tends to predict the shadow free image, whereas discriminators, on the other hand tends to classify that whether the image generated by the Generator is truly free of shadows or not.

Similarly, [10] and [11] uses Dual Hierarchical Networks and Direction Aware Spatial Context both are the state-of-the-art deep convolutional neural networks and tends to generate the shadow-free image.

[12] uses the latest state-of-the-art concepts of transformers with the purpose of shadow detection and removal via spatial attention map. It first generates the spatial attention map which is capable to distinguish between the shadow and the shadow-free regions of an image. With the help of the spatial attention map the shadow region is further refined and shadow-free image is generated at the output.

Limitations of deep-learning models is they demand millions of images in training data-set, whereas publicly existing data-set have very low probability distribution. Moreover, already existing deep-learning based models are performing well on monochromatic color images but fails to produce acceptable results on multi-color contrast images.

2.2 Available Dataset

Dataset is the collection of data in an ordered form and it contains a lot of separate samples of data and which are used to train the model with an objective of finding predictable patterns inside the whole dataset. Data is one of the most essential components for any Artificial Intelligence and deep learning based model as such models demands a lot of training data in order to perform well. Learning based models may require millions of training samples in dataset in order to make accurate predictions on real time unseen data. Besides quantity, quality of the data is also very important even though if state-of-the-art algorithms are implemented.

One of the quotes which best explains the working of machine learning, deep-learning based models is “Garbage in Garbage Out (GIGO)”, which simply means that if low quality training data is feed into the machine learning, deep-learning models during training, then the predicted output by the models will also be of the same low quality. Quality and quantity of data both plays an essential role in governing the performance of the machine learning based models.

Figure 2.1: Different types of dataset available

Existing datasets for the sake of shadow detection and removal are extremely limited with very smaller number of samples in training data which further leads to dropping the probability distribution of the samples in training data. Figure 2.1 shows the different types of dataset available.

- **Unpaired:** Figure 2.1a depicts the unpaired dataset. In this type of dataset the shadow image as well as the shadow-free image both are off different scenarios. [8] proposed this unique dataset. This dataset is can be used for shadow removal purpose only.

- **Paired-(Shadow Detection):** Figure 2.1b depicts the paired (shadow detection) dataset. This type of dataset consist of shadow image and shadow-mask image as a sample point and can be used for shadow detection purpose only.

- **Paired-(Shadow Removal):** Figure 2.1c visualizes the paired (Shadow Removal) dataset. This type of dataset consist of shadow image and shadow-free image of same scenarios as a sample point. This type of dataset can be used for shadow removal purpose only.

- **Triplet:** Figure 2.1d illustrates the Triplet dataset. This type of dataset consist of shadow, shadow-mask and shadow-free image as single sample in the dataset.

Such type of triplet have an advantage over other dataset as these dataset can be used for shadow detection as well as shadow removal purpose at the same time.

Publicly Existing Datasets				
Dataset	Quantity	Image context	Type	Purpose
USR[8]	4215	shadow, shadow-free	Unpaired	removal
ISTD[6]	1870	shadow, shadowmask, shadow-free	Triplet	removal
SRD[17]	3088	shadow, shadow-free	Pair	removal
LRSS[18]	37	shadow, shadow-free	Pair	removal
UIUC[19]	76	shadow, shadow-free	Pair	removal
UCF[20]	245	shadow, shadow-mask	Pair	detection
SBU[21]	4727	shadow, shadow-mask	Pair	detection

Table 2.1: Comparison of publicly accessible dataset for shadow detection or shadow removal purpose

Different dataset available for the purpose of either shadow detection or shadow removal are mentioned in Table 2.1. USR [8] is the only unpaired dataset available, which consist of 4215 shadow, shadow-free samples and is used only for the shadow removal purpose. ISTD [6] is the only triplet dataset publicly available and consist of 1870, shadow/shadow-mask/shadow free samples and hence be used for both shadow detection as well as for shadow-removal purpose. SRD [17], UIUC [19], LRSS [18] consist of 3088, 76 and 37 shadow/shadow-free training samples, and can be used for shadow removal purpose only. Similarly, SBU [21] and UCF [20] contains 4727 and 245, shadow/shadow-mask samples and hence can be used for shadow detection purpose only.

CHAPTER 3

Methodology

This chapter of the report discusses about the proposed methodology used for the sake of shadow detection and removal in detail. First section of this chapter discusses about the preparation of the new triplet (shadow/shadowmask/shadow-free) dataset named as ”Extended ISTD”, followed by the model selection, model modification, model training on the presented dataset, fine-tuning of hyper-parameters and environmental setup required to obtain the most optimal shadow-free image.

Furthermore, later section of this chapter discusses about the proposed post-processing step which leverages the power of traditional image processing techniques. The post processing step is introduced in order to refine the shadow-free image generated by deep-learning model and to get the most efficient and error free shadow removal image at the output.

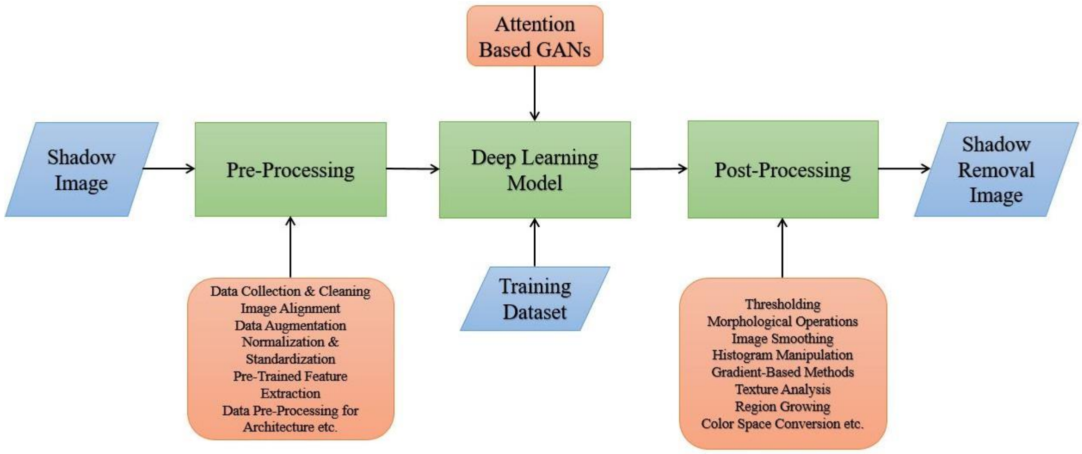


Figure 3.1: Block diagram of the proposed methodology

3.1 Introduction

The proposed methodology for the sake of Shadow detection and shadow removal can be split into two main steps; model utilizing deep neural networks "Attention based Generative Adversarial Networks" (GANs) followed by a post processing step which consist of traditional image processing techniques. The shadow image input is processed through a deep-learning model known as "Attention-based GANs". However, even after training, the GANs' output may not yield a completely shadow-free image. It often leaves behind a lighter shadow version or fails to adequately preserve shadow boundaries. Additionally, in some cases, dark shadow regions are not entirely removed.

In order to tackle these issues and to get a shadow-free image at the output, two steps are taken; proposing a new "Extended ISTD" dataset and introducing a post processing step which consist of different image processing techniques combined together to generate an image that is free from shadows. The presented dataset is a triplet (shadow/shadow-mask/shadow-free) dataset and consists of relatively high probability distribution by covering different multi-color shadow images as well as images consisting of darker/hard shadow regions.

The output shadow-free image predicted by deep-learning attention based GANs is further refined by leveraging the power of traditional image processing techniques such as; morphology operation, histogram matching, custom filters, shadow edge detection and many more. The proposed methodology uses Attention based GANs followed by traditional images processing techniques as a post-processing step as illustrated in Figure 3.1 and is capable of generating the shadow-free images with high efficiency.

Numerous deep-learning based models are available for the purpose of shadow detection and removal. All of the available deep-learning models are data driven and are capable to perform well in limited scenarios such as monochromatic shadow images, soft shadow images but they fails to produce satisfactory results on multi-color contrast images and images containing dark/hard shadow regions. This is due to the fact that the available dataset have a very low probability distribution.

The proposed methodology uses a modified version of ST-CGAN [6], as a deep-leaning based model. We added attention layer after the first convolutional layer of model. The advantage of using said model is that, it is capable of predicting shadow-mask as well as shadow-free image. The said model is trained on “ISTD” dataset, which consist of 1870 triplet (shadow/shadowmask/shadow-free) images and hence can be used for both shadow detection as well as for shadow removal purpose. Limitation of ISTD dataset is; firstly, it consists of extremely low number of training samples in the dataset. This leads to the fact of having a low probability distribution of samples across the dataset. Moreover, the ISTD dataset frequently consists of only monochromatic shadow images samples. All of these limitations leads to unsatisfactory performance of the trained model during inference phase on unseen data consisting of either multi-color shadow images or darker shadow regions in images.

In order to solve these challenges, “Extended ISTD” dataset is proposed which consist of 5352 triplet images samples i.e., shadow, shadow-mask and shadow-free. It is an extended version of already publicly available ISTD [6] dataset. “Extended ISTD” is presented with aim to increase number of samples in dataset and also to increase the targeted probability distribution by specifically adding multi-color shadow image and darker shadow images.

3.2 Presented Dataset - Extended ISTD

Dataset is one of the most essential components for deep-learning based models as such models demand a lot of training data in order to learn desired features and perform efficiently. Publicly existing dataset for purpose of shadow detection and removal are of four different types which are Unpaired, Paired (shadow-detection), Paired (shadow-removal) and Triplet. Unpaired dataset consists of shadow and shadow-free images from different scenarios for single training sample, USR [8] is the only unpaired dataset and contains 4215 samples, it is a complex dataset and requires more complex models and high computational models. Paired (shadow-detection), datasets consist of shadow and shadow-mask samples, it is used for shadow detection purpose.

SBU [21] and UCF [20] are the paired dataset used for purpose of shadow detection and contains 4727 and 245 training samples. Paired (shadow removal), dataset consist of shadow and shadow-free samples, they are used for shadow removal purpose only. SRD [17] contains 3088 samples, UIUC [19] and LRSS [18] both contains less than hundred samples and all the three datasets are used for shadow removal purpose. Triplet dataset consist of shadow, shadow-mask and shadow-free samples and can be simultaneously used for shadow detection as well as shadow removal purpose. ISTD [6] is the only triplet dataset publicly available and consist of 1870 samples in dataset.

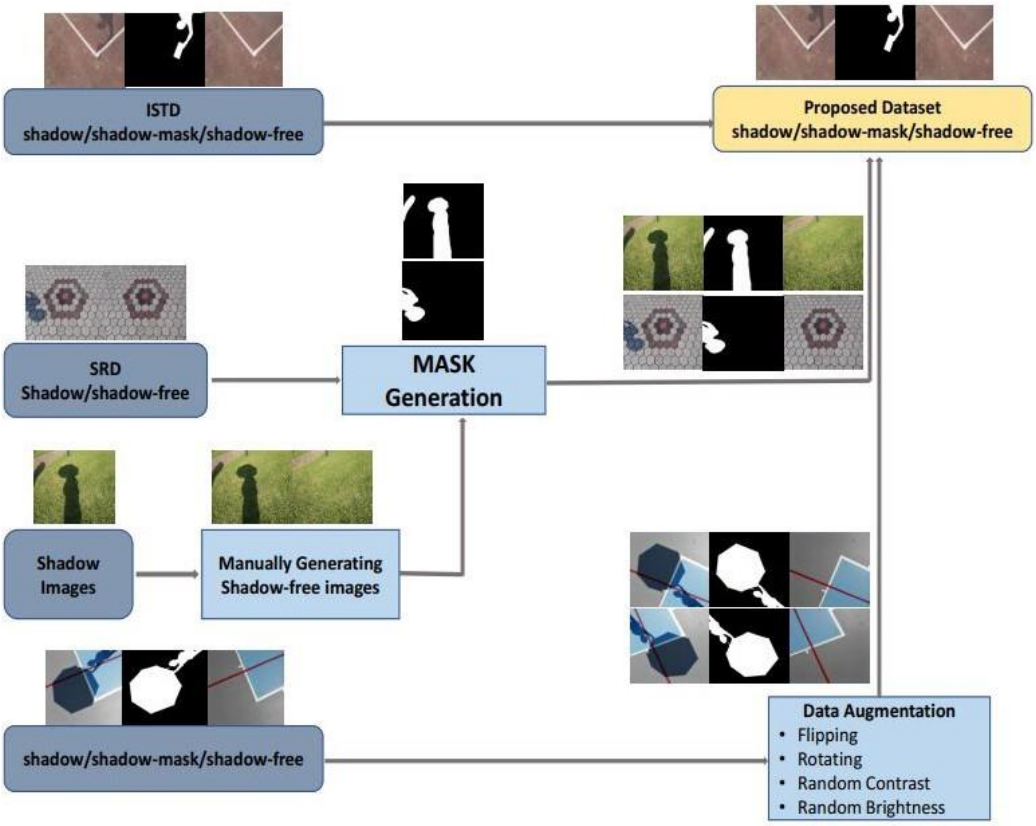


Figure 3.2: Flow diagram for preparation of dataset

To aid in the assessment of shadow understanding techniques, this research proposes a new extended version of ISTD dataset and added 3482 triplet (shadow/shadow-mask/shadow-free) samples in already publicly existing ISTD [6] dataset. The purposed extended version of ISTD dataset named as, "Extended ISTD" consist of 5352 triplet (shadow/shadow-mask/ shadow-free) samples. To the best of our awareness "Extended ISTD" is one of the frst extensive benchmark dataset which can be simultaneously used for shadow detection as well as for shadow removal purpose. In order to increase the probability distribution across the already existing ISTD dataset training dataset, the steps adopted are as follow:

1. Incremental Probability Distribution via Shadow / Shadow-free Images
2. Incremental Probability Distribution via Shadow Images
3. Data Augmentation

Figure 3.2 shows the flow process through which the presented dataset is compiled. Firstly, shadow/shadow-free images were take from SRD dataset, shadow-mask for each sample was generated and these triplet samples were merged into the existing ISTD. Secondly, random shadow images were gathered, their shadow-mask and shadow-free images were generated manually and these triplet samples were added into the dataset. Lastly, data augmentation was also carried out in order to increase the probability distribution across the presented dataset.

3.2.1 Incremental Probability Distribution via Shadow / Shadow-free Images

Shadow Removal Dataset (SRD) [17] is a pair dataset which consist of 3088 shadow and the shadow-free sample images and hence can be used only for shadow-removal purpose. The number of samples are relatively high when compared with other publicly available dataset. Limitation of dataset is that there are huge number of different samples from same scenarios. This leads to the fact of having low probability distribution across the dataset. SRD dataset is also incompatible with the purposed methodology as the purposed requires shadow, shadow-mask and the shadow-free images sample. The aim is to increase the probability distribution across the dataset. One or two random images from same scenarios were taken their shadow-mask was generated by using Algorithm 3.1.

Algorithm 3.1 This algorithm takes shadow image and shadow-free image as an input and aims to produce the shadow-mask at its output

---

Require: Shadow Image and Shadow-Free Image

I_1 : *ShadowImage*

I_2 : *ShadowFreeImage*

I_x : *NumberofPairImages*

$\mathrm{N}_x : 0$

O_m : *ShadowMaskImage*

$N_x \leq I_x$

$I \leftarrow I_2 - I_1$ {Subtracting shadow image from shadow free image}

$I \leftarrow I[:,:,1]$ {Converting 3-Channel Image to 1-Channel}

$I \leftarrow Thresholding[i > 200 \leftarrow 255; i < 200 \leftarrow 0]$ {Binary Image}

$I \leftarrow MedianFilter[I]$

$O_m \leftarrow MorphologicalClosing[I]$

Proposed methodology takes shadow, shadow-free samples of different scenarios. Shadow-mask is generated using different image processing techniques as shown in Algorithm 3.1. Firstly, pixel-wise subtraction is carried out between shadow and shadow-free image. 1-Channel of the resultant image is taken which undergoes the thresholding and binary image is generated.

Figure 3.3: Examples of samples added in training dataset from shadow and shadow-free Pair Images by generating the shadow-mask

3x3 median filter is applied on resultant binary image followed by morphological closing operation which gives the desired shadow-mask. These triplets set of samples i.e., shadow, shadow-free and newly generated shadow-mask are added in the dataset. Figure 3.3 illustrates few of the training samples added in the dataset from shadow and shadow-free images by generating their respective shadow mask via Algorithm 3.1.

Figure 3.4: Examples of samples added in training dataset from Shadow images only by generating Shadow-free and shadow-mask images

3.2.2 Incremental Probability Distribution via Shadow Images

Random images from internet consist of variety of shadow images with different scenarios, contrast, illumination, background texture and many more. Adding such images in the dataset can significantly increase the probability distribution. The core limitation in this case is that only shadow image is available and shadow-free, shadow-mask must be generated efficiently.

Shadow-free images were generated by selecting the shadow region manually and estimating the background, while removing the present shadows. Those shadow images whose shadow-free image was generated efficiently undergoes Algorithm 1 in order to get the shadow-mask. These triplets set of samples i.e., shadow, shadow-free image manually generated and shadow mask generated via proposed algorithm 1 are merged with the existing the dataset. Figure 3.4 illustrates few of the training samples added in the dataset from shadow image only by generating their shadow-free image and shadow-mask.

3.2.3 Data Augmentation

Data augmentation is a process which is commonly used to increase the probability distribution in dataset. Already available triplet samples (shadow/shadowmask/shadow-free) in dataset undergoes the augmentation, where different augmentations methods such as rotation, flipping, altering contrast and brightness were performed at random on triplet set (shadow/shadow-mask/shadowfree) of images. These augmented images might look similar to human eye but are totally different and unique sample for a deep convolutional neural network model. These triplet samples generated by means of augmentation method are than merged into the dataset. Hence, resulting in increase of probability distribution across the dataset.

3.3 Deep-Learning Based Model

The chosen model ST-CGAN [6], is a deep-learning based Generative Adversarial Network (GANs) model which consist of two generators (G1 and G2) and two discriminators (D1 and D2). The generator G1 receives the shadow image at its input and tends to predict the shadow mask. The discriminator D1 investigates that the generated shadow mask by G1 is predicted accurately or not. This generated shadow mask is then concatenated with the given input shadow image and is further feed into the generator G2. The generator G2 aims to predict the resultant shadow-free image. While the discriminator D2 investigates that either the predicted shadow free image is genuinely a shadow free version of input image or not.

The selected model demonstrates strong performance when handling monochromatic shadow images. However, it struggles to generate satisfactory results when faced with multi-color contrast shadow images and dark shadow images. To address this limitation, we introduced an attention mechanism following the first convolutional layer of G1. The purpose of incorporating the attention mechanism is to analyze the image pixel by pixel, allowing for more effective shadow detection.

Then this attention based model was trained on the proposed “Extended ISTD” dataset, in which the targeted probability distribution was increased by specifcally adding multi-color shadow samples and dark shadow samples. Presented dataset consist of 5352 triplet samples. During training, the dataset was splitted in such a way that it consists of 70 percent samples in training set, 20 percent samples in validation set and remaining 10 percent samples in test set.

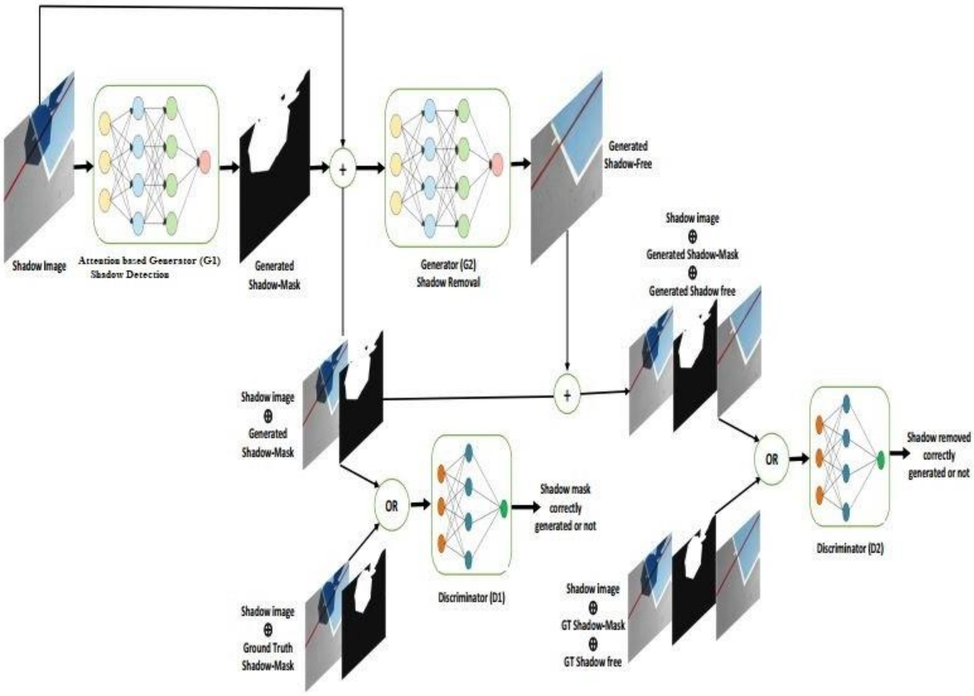


Figure 3.5: Model architecture of attention based GANs

As illustrated in Figure 3.5, the generator (G1) receives the shadow image as an input, and is used to produce the shadow mask. This generated shadow mask and the ground truth shadow mask is fed into the Discriminator (D1), which classify that either the shadow mask is generated accurately. The input shadow image and generated shadow mask are than concatenated and fed into the generator (G2), which tends to produce the shadow free image. This predicted shadow free image is then given into the discriminator (D2), which tends to categorize that the predicted shadow free image is actually a shadow free image or not.

3.3.1 Model Training

To train the model on the provided 'Extended ISTD' dataset, we had two options available: initiating model training from scratch or performing transfer learning. Training the model from scratch demands significant time and requires substantial computational resources. Conversely, transfer learning, a technique commonly employed in machine learning, utilizes a pre-trained model as a starting point. Transfer learning enables rapid progress, enhances efficiency, and conserves resources. However, due to the incorporation of the attention mechanism in G1, transfer learning was not feasible for our purposes. Hence, we proceeded to train our model from scratch. The training loss of generator and discriminator are shown in the Figure 3.6 for 1300 epochs.

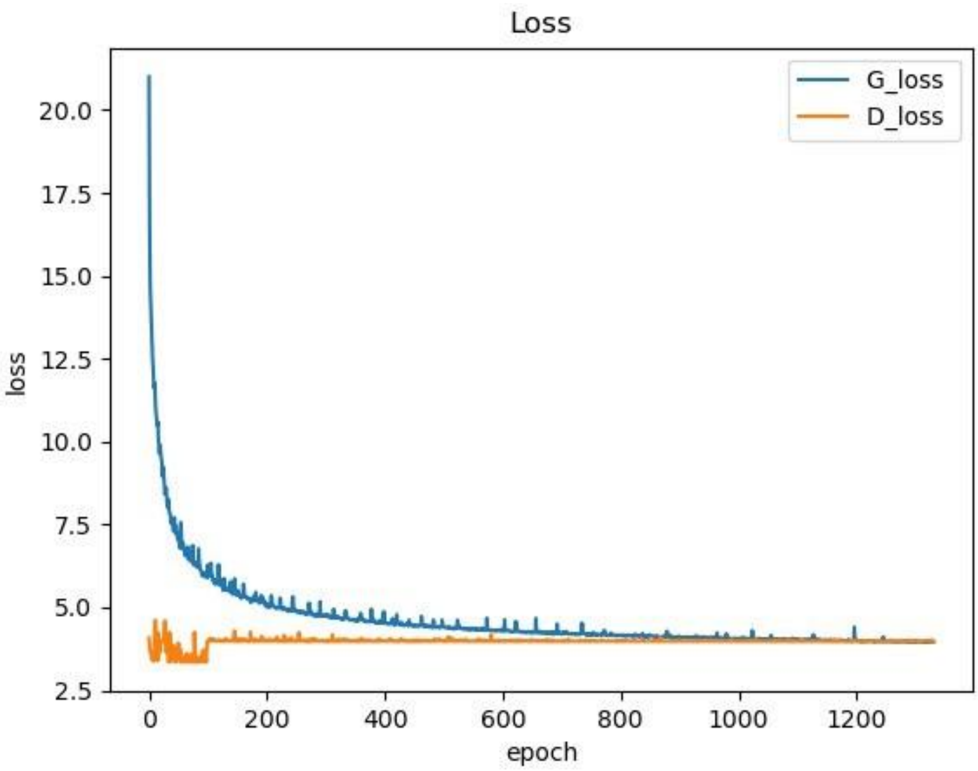


Figure 3.6: Attention based model training loss

At a point during model training, gradient descent was unable to converge at global minima and validation loss suddenly start increasing from “0.56” to “0.98”. This sudden increase in validation loss indicates the over-fitting of model on training set. This issue was tackled by fine tuning of hyper-parameters. Learning rate scheduling and increasing the number of samples in the batch was adopted simultaneously. After specific number of epochs learning rate was decreased via learning rate scheduling and batch size was gradually increased from 8, 16, 32 to 64 samples per batch during training. By fine tuning the hyper-parameter’s, an acceptable generator loss of “0.2507” and discriminator loss of “0.1918” was achieved as shown in the Figure 3.7.

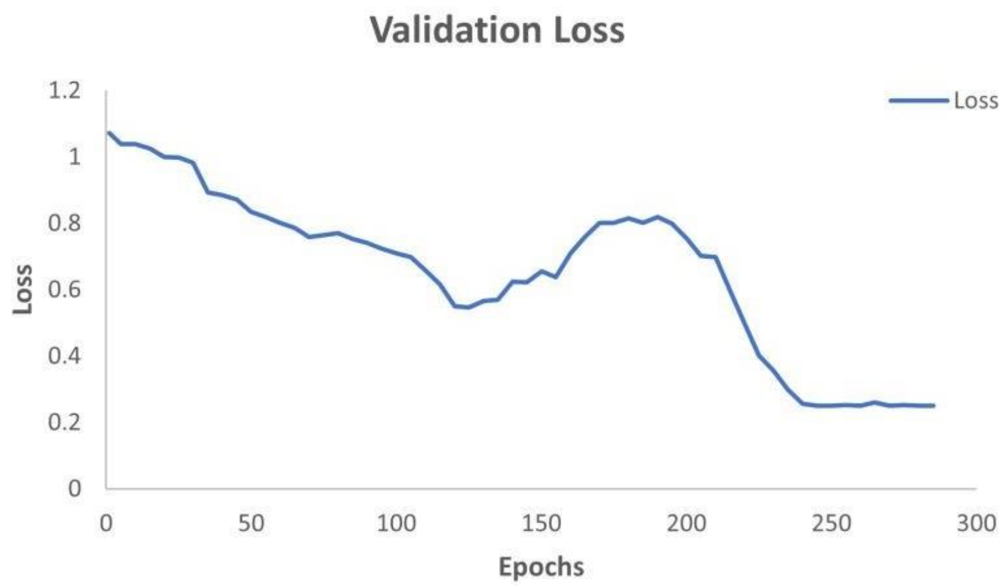


Figure 3.7: Validation loss while training

3.4 Post Processing - Traditional Image Processing Techniques

Deep-learning Attention based model, once trained on the presented "Extended ISTD" dataset was capable of removing shadow from images. But still in few scenarios either the lighter version of shadow was left behind or shadow boundaries were not preserved properly. So, in order to enhance the image quality and to remove the artefacts if present, the intermediate shadow free image generated by deep-learning model undergoes a post processing step. The proposed post processing step leverages the power of different image processing techniques combined together in order to produce a more refined shadow free image with efficiently removing the lighter version of shadows and preserving the shadow boundaries at the same time.

Shadow detection is one of the core importance in proposed post processing step. The post processing step is focused on predicting the texture beneath the shadow regions only. Shadows if detected correctly i.e., shadow mask if efficiently generated, then the post processing step is capable of producing the desired shadow free image efficiently, while preserving the texture beneath the shadow regions as well as shadow boundaries simultaneously. The post-processing step, which is outlined in Algorithm 3.2, is the purposed approach for further refining the results. This algorithm serves as a representation of the proposed method to be employed after the deep-learning based model. Its purpose is to enhance the quality of the image obtained from the preceding steps. By following the steps defined in Algorithm 3.2, the data can undergo additional transformations, adjustments, or filtering to ensure optimal results are achieved.

Algorithm 3.2 Proposed Post Processing step which is used to refine the intermediate shadow free image generated by deep learning model GANs

Require: Input Shadow Image and Intermediate Shadow-free Image by GANs

$I_{mask} \leftarrow Algorithm1\left[I_{sha}, I_{int-sha-free}\right]$

$I_{inv-mask} \leftarrow I_{mask}^{-1}$

$I_{sha-reg} \leftarrow I_{int-sha-free} - I_{inv-mask}$

$I_{sha-free-reg} \leftarrow I_{int-shad-free} - I_{mask}$

$I_{hist-match} \leftarrow Hist - Match\left[I_{sha-reg}, I_{sha-free-reg}\right]$

$I_{filt} \leftarrow \left(I_{hist-match} * n_1\right) + \left(I_{sha-reg} * n_2\right)$

$I_{filt-conc} \leftarrow I_{sha-free-reg} \oplus I_{filt}$

$I_{sha-bound} \leftarrow EdgeDetector\left[I_{filt-conc}, I_{mask}\right]$

$I_{sha-free} \leftarrow inpaint\left[I_{sha-bound}\right]$

The generated shadow mask by deep learning model is refned using the image processing techniques. Input shadow image I_{sha} and intermediate shadow free image $I_{int\text{-}sha\text{-}free}$ generated by deep learning model are used to generate the shadow mask I_{mask} via purposed Algorithm 3.2. The algorithm takes the shadow image I_{sha} and the intermediate shadow free image $I_{int\text{-}sha\text{-}free}$ generated by deep learning model as an input and generates the shadow mask I_{mask}. An inverted shadow mask $I_{inv\text{-}mask}$ is also generated by taking the inverse of the shadow mask I_{mask} generated by Algorithm 3.1. The shadow mask $I_{mask}$and inverted shadow mask $I_{inv\text{-}mask}$ are further used to extract the shadow region and shadow free region.

$$I_{mask} = \text{Algorithm1}[I_{sha}, I_{int\text{-}sha\text{-}free}]$$
$$I_{inv\text{-}mask} = I_{mask}^{-1}$$

The intermediate shadow free image $I_{int\text{-}sha\text{-}free}$, is further split into two separate images; shadow region $I_{sha\text{-}reg}$, which consist of the pixels belonging to shadow region only and shadow free region $I_{sha\text{-}free\text{-}reg}$; which consist of the pixels that belongs to the shadow free region only. Shadow region $I_{sha\text{-}reg}$ is extracted by pixel wise subtraction of inverted shadow mask $I_{inv\text{-}mask}$ from intermediate shadow free image $I_{int\text{-}sha\text{-}free}$. Similarly, the shadow free region $I_{sha\text{-}free\text{-}reg}$ is extracted through pixel wise subtraction of generated shadow mask I_{mask} from intermediate shadow free image $I_{\text{int-sha-free}}$.

$$I_{\text{sha-reg}} = I_{\text{int-sha-free}} - I_{\text{inv-mask}}$$
$$I_{\text{sha-free-reg}} = I_{\text{int-sha-free}} - I_{\text{mask}}$$

Once the shadow region $I_{\text{sha-reg}}$ and the shadow free regions $I_{\text{sha-free-reg}}$ are extracted, channel-wise histogram matching is applied on the extracted shadow region $I_{\text{sha-reg}}$ During the histogram matching process the extracted shadow free region $I_{\text{sha-free-reg}}$ is taken as a reference image.

$$I_{\text{hist-match}} = \text{HistMatching}[I_{\text{sha-reg}}, I_{\text{sha-free-reg}}]$$

The filtered image is obtained by performing pixel wise averaging histogram matched image $I_{\text{hist-match}}$ and the shadow region I_{sha}. n1 and n2 are the weights given to $I_{\text{hist-match}}$ and I_{sha}, where the total sum of n1 and n2 is equal to 1.

$$I_{\text{filt}} = (I_{\text{hist}} * n_1) + (I_{\text{sha-reg}} * n_2)$$

The resultant generated shadow free region by applying the filter I_{filt} is merged with the extracted shadow free region $I_{\text{sha-free-reg}}$ generated by a deep learning model. In this way, targeted histogram matching and average filters are applied on shadow region only.

$$I_{\text{filt-conc}} = I_{\text{sha-free-reg}} \oplus I_{\text{filt}}$$

The shadow boundaries $I_{\text{sha-bound}}$ are extracted by using the shadow mask. The shadow edges were detected with the help of shadow mask and the canny edge detector were used to detect the edges.

$$I_{\text{sha-bound}} = \text{EdgeDetector}(I_{\text{filt-conc}}, I_{\text{mask}})$$

The shadow boundaries are predicted via “cv2.inpaint”. The resultant image is a shadow-free image efciently generated with preserved shadow boundaries.

$$I_{\text{sha-free}} = \text{Inpaint}(I_{\text{sha-bound}})$$

The resultant $I_{\text{sha-free}}$ is the fnal generated shadow free image and is the refned version of intermediate shadow free image generated by deep learning model.

The Figure 3.7 input shadow image I_{sha} and intermediate shadow free image $I_{\text{int-sha-free}}$ are fed into the Algorithm 3.1, and shadow mask I_{mask} is generated. Using the generated shadow mask I_{mask}, shadow region $I_{\text{sha-reg}}$ and shadow free region $I_{\text{sha-free-reg}}$are then extracted from intermediate shadow free image $I_{\text{int-sha-free}}$. Histogram matching is applied on the extracted shadow region $I_{\text{sha-reg}}$, while taking the shadow region $I_{\text{sha-reg}}$ as source image and shadow free region $I_{\text{sha-free-reg}}$ as reference.

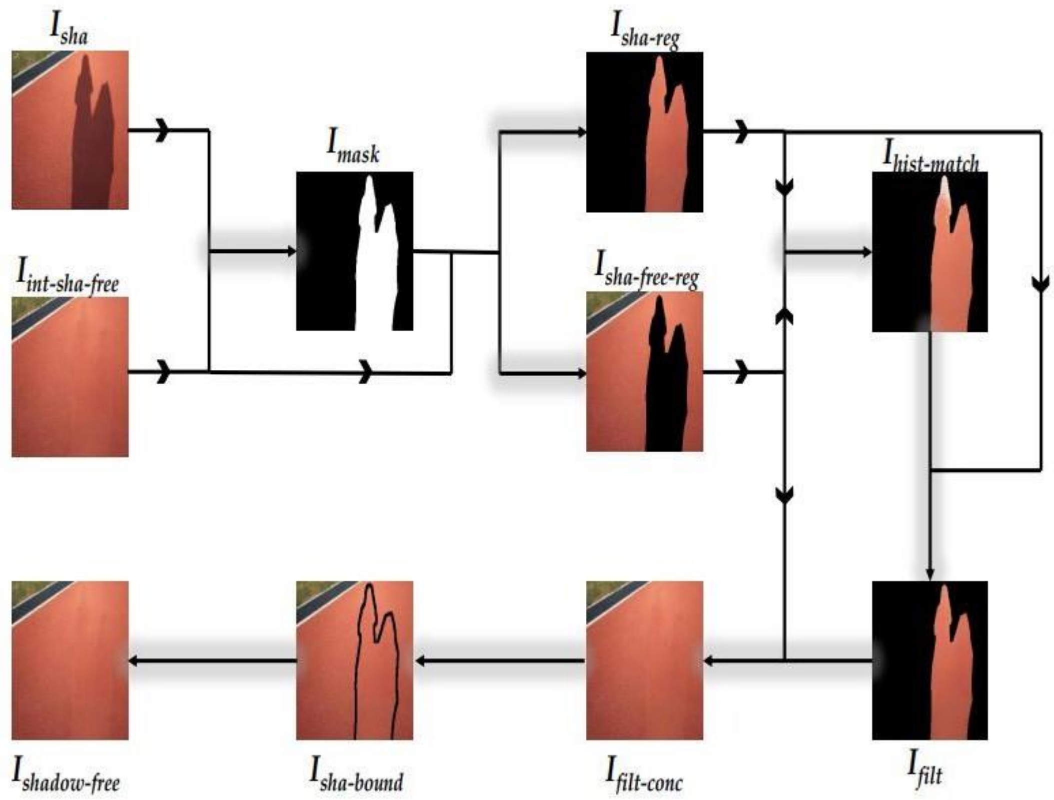


Figure 3.8: Purposed Post Processing Step

The proposed post processing step is used to refine the intermediate shadow free image generated by deep learning based model GANs. The achieved results shows that the presented post processing step is capable of generating the shadow free image efficiently by refining the intermediate shadow free image generated by GANs. Both qualitative and quantitative analysis of the achieved results shows that the presented methodology clearly outperforms the other SOTA methodologies in terms of shadow removal.

Figure 3.7 shows the flow diagram of the proposed post processing step. This step takes the shadow-free image generated by deep-learning based model as an input and tends to refine its by leveraging the power of various different traditional image processing techniques. The post processing steps aims to apply the processing on targeted shadow regions only. Various traditional image processing techniques including channel-wise histogram matching, customized weighted average filters are targeted applied to shadow region only. Furthermore, the shadow boundaries are detected by using the shadow mask via canny edge detector and are preserved by performing boundary estimation using ”cv2.inpaint”. The final output shadow-free image generated by the post-processing step is more refined with preserved shadow boundaries.

This chapter of the report delves into the detailed explanation of the proposed methodology adopted for the purpose of shadow detection and removal. The methodology encompasses several key components and techniques aimed at enhancing the accuracy and effectiveness of shadow removal.

The first major contribution of the proposed methodology is the introduction of a new dataset called ”Extended ISTD”. This dataset consists of 5352 triplet samples, making it a valuable resource for both shadow detection and shadow removal tasks. By utilizing this dataset, researchers and practitioners can train and evaluate their shadow removal models on a diverse range of realworld scenarios, thereby enhancing the generalizability and robustness of their approaches.

The proposed methodology further involves the use of Attention based Generative Adversarial Network (GAN) model. This model consists of two generators and two discriminators, which work in tandem to learn and generate high-quality shadow-free images. GANs have proven to be effective in various image generation tasks, and by employing them in the context of shadow removal, the proposed methodology aims to leverage their capabilities to produce superior results.

Additionally, the proposed methodology introduces a post-processing step to further refine and enhance the generated shadow-free images. This post-processing step integrates various traditional image processing techniques, combining their strengths to achieve more refined and efficient results. By leveraging these techniques, the proposed methodology aims to address potential artifacts or imperfections that may arise during the shadow removal process and improve the overall quality of the output image. By combining the extended dataset, the Attention based GAN-based model, and the post-processing step, the proposed methodology aims to tackle the challenges associated with shadow removal comprehensively. It addresses the limitations of existing methods by leveraging a rich dataset, advanced deep learning techniques, and additional refining steps. This comprehensive approach enhances the potential of the proposed methodology to produce high-quality shadow-free images that closely resemble the ground truth and effectively eliminate the presence of shadows.

CHAPTER 4

Results and Discussion

This chapter of the report discusses about the different evaluation metrics, which could be used to evaluate the performance of shadow removal result. Further sections of the chapter contains information about the system/environmental setup, training loop and hyper-parameter tuning of the deep-learning model in order to get the model converged at global minima. Whereas, later section of the chapter, compares the achieved results of proposed shadow removal method, visually as well as qualitatively with other state-of-the-art available methodologies.

4.1 Evaluation Metrics

Evaluation metrics are used in order to evaluate or to assess the performance of the deep-learning based models. There are several different evaluation metrics and the choice of metric highly depend on the problem being addressed. Choosing the appropriate evaluation metric depends on the particular task and goal of the deep-learning model. In order to evaluate the shadow removal problem, the evaluation metrics typically used are root mean square error (RMSE) and peak-signal-to-noise-ration (PSNR); which are discussed as below:

4.1.1 Root Mean Square Error (RMSE):

In order to evaluate the shadow removal performance, Root Mean Square Error (RMSE) is computed between the predicted images without shadows and the ground truth image without shadows. RMSE compares the two images by measuring the difference between them. Generally, smaller the value of RMSE, lower is the difference between two images and better is the performance of the algorithm.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}(I_{\mathrm{sf}} - I_{\mathrm{gt}})^2}$$

In above equation, I_{sf} is the shadow free image predicted by the proposed methodology, I_{gt} is ground truth shadow free image present in the test data and 'n' is the overall number of samples.

4.1.2 Peak-Signal-to-Noise-Ratio (PSNR)

It is also one of the most commonly used evaluation metric to check the similarity between two images. It compares the quality of a generated shadow-free image to the ground truth shadow-free image by calculating the ratio of the maximum possible power of the signal to the power of the noise that affects the fidelity of the image. PSNR is quantified in decibels (dB) and can be computed by using the following equation:

$$\mathrm{PSNR}(x, y) = 10\log_{10}\left(\frac{R^2}{\mathrm{MSE}}\right)$$

In above equation "R" represents the maximum fluctuation in the input image, while "MSE" stands for mean square error.Higher the value of PSNR, more will be the similarity between the generated shadow-free image to the ground truth version of the shadow-free image.

4.2 System Setup

The model was trained and post processing was performed using Core i7 12th generation CPU. Environmental setup for training, validating the proposed methodology includes, python 3.7.0, torch 1.5.0+CPU and torch-vision 0.6.0+CPU. The training was done for 1300 epochs and it toked "288" hours for the model to converge at global minima.The shadow free image predicted by the Attention based deep-learning model is then refined by a post processing step. This post processing step consist of different image processing techniques combined together in order to get the finial shadow free image.

Table 4.1: Environmental setup: Different Libraries/frameworks and their version used

Framework/Library	Version
PyTorch	1.5.0+CPU
Python	3.7.0
torchvision	0.6.0+CPU
Pillow	7.1.2
Matplotlib	3.2.2

4.3 Qualitative and Visual Comparison with State-of-art Methods

This section of the report thoroughly performs comparison of the proposed shadow removal methodology with the other available state-of-the-art methods. The proposed model is trained on presented "Extended ISTD" dataset which consist of 5352 triplet (shadow/shadow-mask/shadow-free) samples. Furthermore, a post processing step is performed to refine the shadow free image. In order to compare the qualitative results with other state-of-the-art methods, ISTD test set is chosen, which consist of 540 triplet samples.

4.3.1 Qualitative Comparison

For the purpose of qualitatively comparing the shadow removal results with other highly advanced shadow removal methodologies i.e., ST-CGAN [6], Mask-Shadow GAN [8], Towards Ghost free Shadow Removal [10], RISGAN [7] and SpA-Former [12] their results were attained from the authors published papers. Table 4.2 compares the root mean square error (RMSE) and peak-signal-to-noise-ratio (PSNR) of the proposed methodology with other state-of-the-art deep learning based models. The comparison shows that the proposed methodology clearly outperforms other available methodologies by achieving the RMSE (root mean square error) of "5.68" and a PSNR (Peak Signal to noise ratio) of "31.66". Low RMSE and High PSNR indicates that the predicted shadow-free image by the proposed methodology is much more similar to that present in the ground truth dataset.

Qualitative comparison on ISTD Test Dataset						
Model	RMSE ↓			PSNR ↑		
	S	NS	A	S	NS	A
ST-CGAN [6]	10.33	6.93	7.47	-	-	-
RIS-GAN [7]	8.99	6.33	6.95	-	-	-
Mask Shadow GAN [8]	-	-	7.61	-	-	-
Ghostfree Shadow Removal [10]	7.52	5.43	5.76	34.98	-	-
SpA-Former [12]	10.48	6.22	6.86	33.51	30.16	27.73
ARGAN [22]	6.65	5.41	5.89	-	-	-
Ours	2.30	4.12	5.28	46.19	33.91	31.66

Table 4.2: Qualitative comparison for the shadow removal of proposed methodology with state-of-the-art techniques

The Table 4.2 shows the Qualitative comparison between the purposed methodology and other available state of the art methodologies. Root Mean Square Error (RMSE), the lower the better and Peak-Signal-to-Noise Ratio (PSNR), the higher the better are used as the evaluation metric to compare the performance of the proposed methodology with other available state of the art methods. S represents the shadow region, NS represents the non shadow region or the shadow free region, while A represents the whole image. Qualitative comparison performed shows that the purposed methodology clearly outperforms the available SOTA methodology in terms of RMSE and PSNR.

4.3.2 Visual Comparison

The comparison and evaluation of the proposed shadow removal method against other state-of-the-art techniques can be visually observed in Figure 4.1 and Figure 4.2. These fgures provide a detailed illustration of the performance of the proposed method on the ISTD test dataset and a set of randomly chosen images with multi-color contrast shadows.
Figure 4.1 specifcally showcases the visual comparison of shadow removal outcomes achieved by the proposed methodology in contrast to other existing approaches on the test dataset of ISTD. On the other hand, Figure 4.2 demonstrates the results of the proposed method on a collection of randomly selected shadow images.

By observing the generated shadow-free images, it becomes evident that the proposed methodology surpasses other available methods in terms of effectiveness and efficiency. The proposed method exhibits a higher capability to produce shadow-free images that are visually appealing and of higher quality. Moreover, it demonstrates superior performance in scenarios involving complex backgrounds, dark/hard shadows, and multi-color contrast shadow images.
An important advantage of the proposed methodology is its ability to preserve the texture of the shadow background, even in challenging situations. This feature distinguishes it from the state-of-the-art methods, which often struggle to maintain the integrity of the shadow background when removing the shadows. The proposed methodology overcomes this limitation, thereby ensuring more realistic and accurate results.
In summary, the visual comparisons presented in Figure 4.1 and Figure 4.2 highlight the superiority of the proposed shadow removal method over existing state-of-the-art techniques. The method exhibits enhanced efficiency, produces visually pleasing shadow-free images, and effectively preserves the texture of the shadow background, even in complex scenarios.
In this chapter, the evaluation metrics relevant to shadow detection and removal were

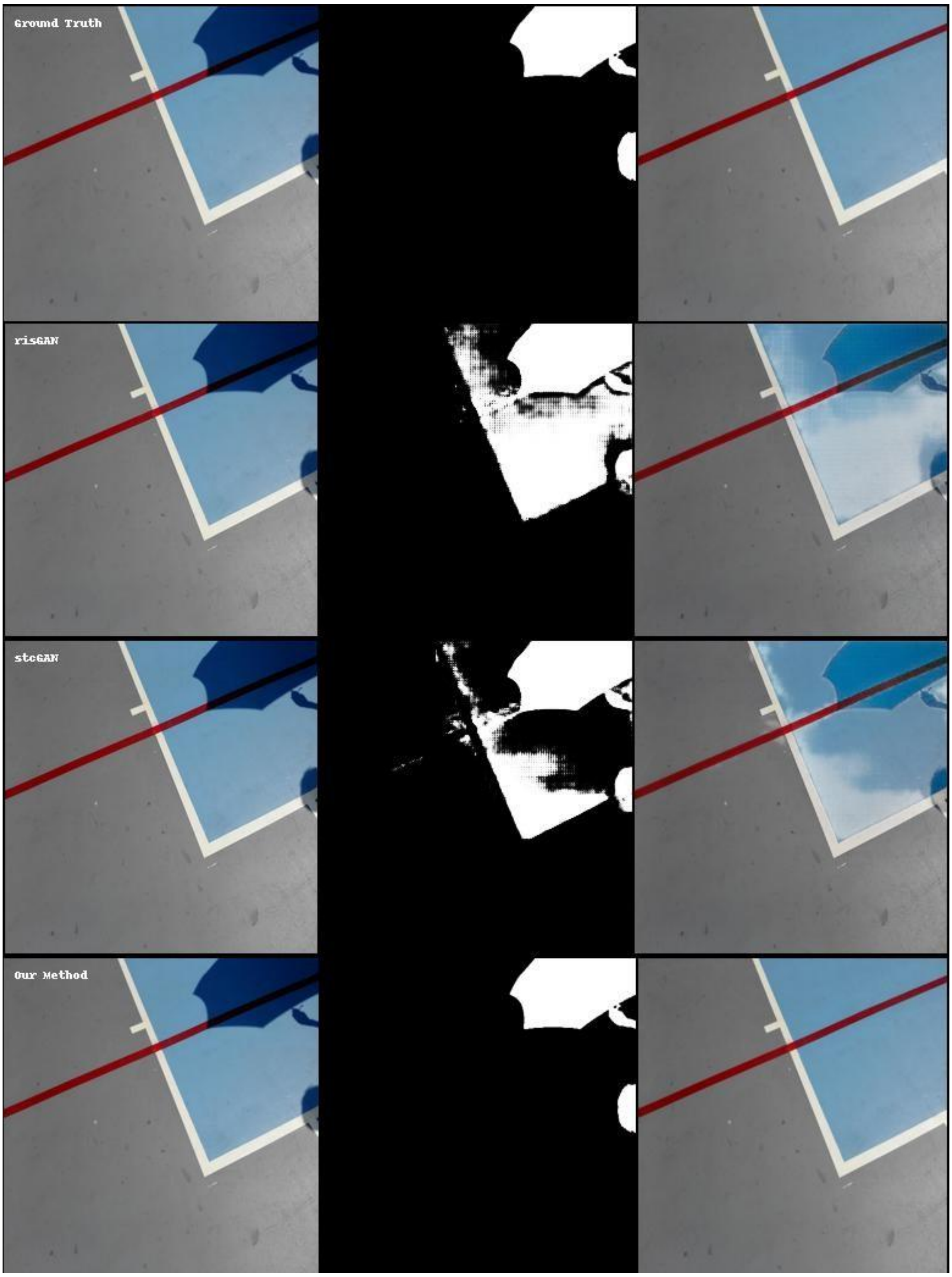


Figure 4.1: Inference result of the presented methodology with other highly advanced methodologies on ISTD Test Set

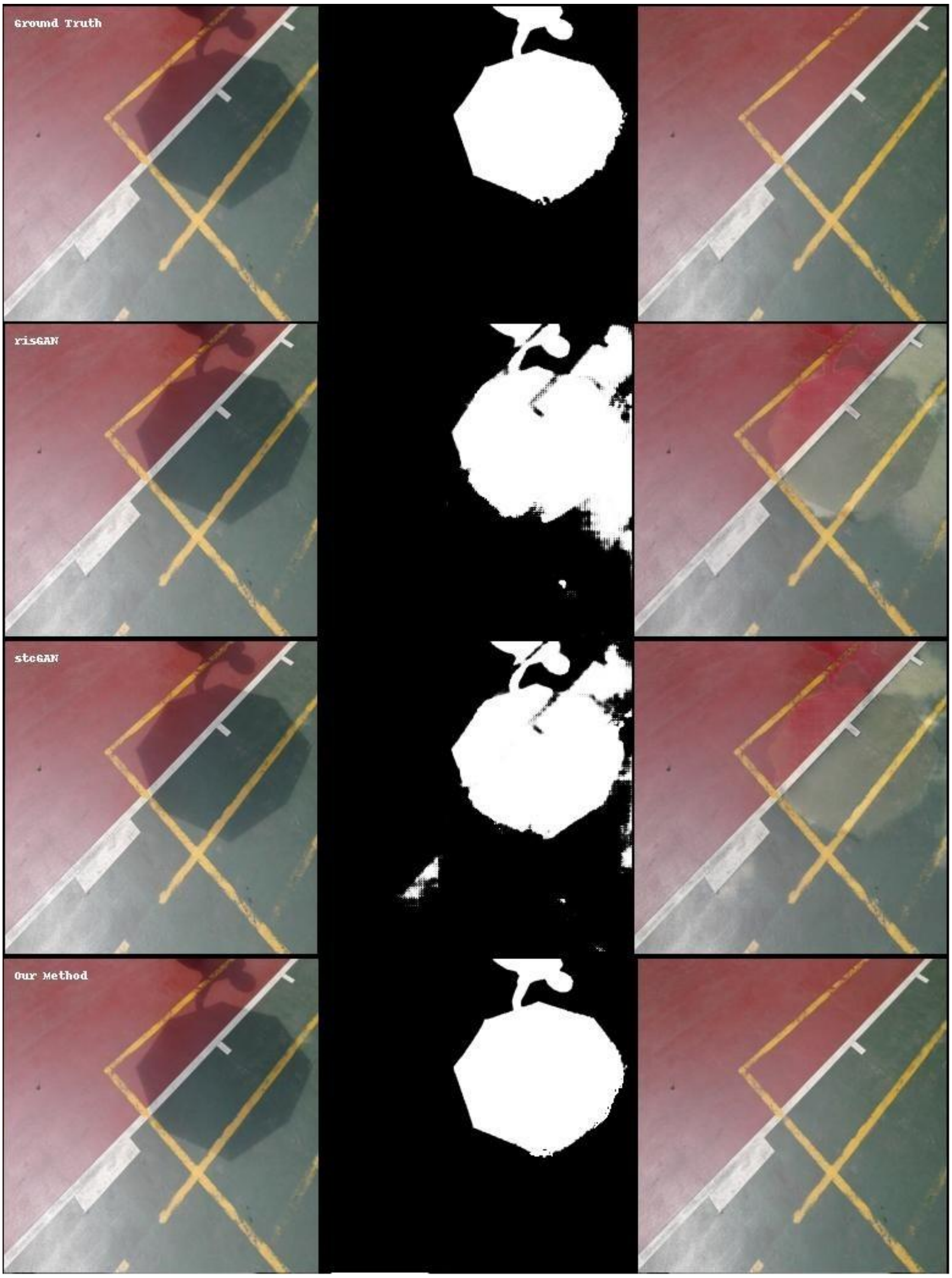


Figure 4.2: Inference result of the presented methodology with other highly advanced methodologies on Random multi-color contrast shadow image

thoroughly examined and discussed. These metrics serve as objective measures to assess the performance and effectiveness of different methods in these domains. By utilizing these evaluation metrics, researchers and practitioners can quantitatively evaluate the quality and accuracy of their shadow removal techniques. The latter part of the chapter focused on presenting a qualitative and visual comparison of the proposed shadow removal technique with other existing methodologies.
The purpose of this comparison was to demonstrate the superiority of the presented methodology in terms of its performance and outcomes. The qualitative comparison results clearly indicate that the proposed methodology achieves a lower Root Mean Square Error (RMSE) and a higher Peak Signal-to-Noise Ratio (PSNR) when compared to the other shadow removal techniques. A lower RMSE indicates that the proposed method produces results that are closer to the ground truth or desired output, while a higher PSNR signifies higher fidelity and quality of the generated shadow-free images. These quantitative measures provide strong evidence that the proposed methodology outperforms other existing methodologies in terms of accuracy and quality.

Moreover, visualizing the inference results further reinforces the superiority of the proposed methodology. By visually examining the shadow removal outcomes, it becomes evident that the proposed method consistently generates images that are visually superior to those produced by other shadow removal methodologies. The visual comparison provides tangible evidence of the improved performance and effectiveness of the proposed methodology. In summary, the chapter discussed the evaluation metrics applicable to shadow detection and removal and presented a comprehensive qualitative and visual comparison of proposed methodology with the other available techniques. The qualitative comparison demonstrated lower RMSE and higher PSNR values for the proposed method, indicating its superior accuracy and quality. Additionally, the visual comparison showcased the visually enhanced results obtained through the proposed methodology, further supporting its out performance over alternative shadow removal methodologies.

CHAPTER 5

Conclusion

This chapter in the report includes the conclusion/summary of the presented work. Few suggestions about the future work are also purposed in the later section of this chapter.

5.1 Summary

In the presented work, a novel methodology for the purpose of shadow detection and shadow removal is presented. The key idea is to detect and remove shadows from RGB images, while preserving the background texture and shadow boundaries. For this purpose, a large scale dataset "Extended ISTD" is presented, which consist of 5352 triplet samples and it can be used for both shadow detection as well as for shadow removal purpose.

The presented dataset is an extended version of already publicly available dataset ISTD. The presented dataset aims of increasing the probability distribution of samples across the dataset. This is done by specifically adding the dark/hard shadow samples and multi-color contrast shadow samples. Attention based Deep-learning based model GANs are trained on the presented dataset i.e., "Extended ISTD", fine-tuning of hyper-parameters was also performed to obtain the most efficient shadow-free image. The shadow-free image generated by the deep-learning model is then refined by a post processing step.

The purposed post-processing step leverages the power of different traditional image processing techniques combined together and aims to produce the final shadow free image while preserving the surface texture and shadow boundaries. The shadow free images generated by the proposed methodology has a low RMSE of 5.68 and a high PSNR of 31.66, and is capable of removing shadows efciently when compared with other highly advanced state-of-the-art methodologies.

5.2 Future Work

There have been significant progress in the field of shadow detection and removal in past few years but still there is much more to do in order to increase the efficiency and effectiveness of shadow removal techniques. Potentially possible future directions in the related field are as follow:

- **Dataset:** The publicly available datasets contains very few samples in training set. This leads to the fact of having an extremely low probability distribution across the dataset. Deep-learning based model requires millions of samples in training set to perform well. So there is a lot more work needed to be done in gathering the dataset related to shadow detection and removal.

- **Real-time Shadow detection and Removal:** Most of the work done in the field of shadow detection and removal is dealing with the images. Real-time applications requires low latency, high accuracy, efficient data processing, adaptability, robustness and many more. More over, real-time applications also requires real-time scenarios dataset.So, it is of core importance that the available algorithms may be optimized so that they should be efficient enough to detect and remove shadows with high efficiency and in less inference time.

Bibliography

[1] X. Yu, G. Li, Z. Ying, and X. Guo, "A new shadow removal method using color-lines," in International Conference on Computer Analysis of Images and Patterns, pp. 307–319, Springer, 2017.

[2] K. Deb and A. H. Suny, "Shadow detection and removal based on YCbCr color space," Smart Comput. Rev., vol. 4, no. 1, pp. 23–33, 2014.

[3] S. Anoopa, V. Dhanya, and J. J. Kizhakkethottam, "Shadow detection and removal using tri-class based thresholding and shadow matting technique," Procedia Technology, vol. 24, pp. 1358–1365, 2016.

[4] S. Khan, Z. Pirani, T. Fansupkar, and U. Maghrabi, "Shadow removal from digital images using multi-channel binarization and shadow matting," in 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), pp. 723–728, IEEE, 2019.

[5] S. Khan, M. Narvekar, T. Fansupkar, and U. Maghrabi, "Shadow removal using multi-channel binarization, color-line clustering and illumination estimation," in 2021 4th Biennial International Conference on Nascent Technologies in Engineering (ICNTE), pp. 1–6, IEEE, 2021.

[6] J. Wang, X. Li, and J. Yang, "Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1788–1797, 2018.

[7] L. Zhang, C. Long, X. Zhang, and C. Xiao, "RIS-GAN: Explore residual and illumination with generative adversarial networks for shadow removal," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 12829–12836, 2020.

[8] X. Hu, Y. Jiang, C.-W. Fu, and P.-A. Heng, "Mask-shadowgan: Learning to remove shadows from unpaired data," in Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2472–2481, 2019.

[9] T. Nagae, R. Abiko, T. Yamaguchi, and M. Ikehara, “Shadow detection and removal using GAN,” in 2020 28th European Signal Processing Conference (EUSIPCO), pp. 630–634, IEEE, 2021.

[10] X. Cun, C.-M. Pun, and C. Shi, “Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting GAN,” in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 10680–10687, 2020.

[11] X. Hu, C.-W. Fu, L. Zhu, J. Qin, and P.-A. Heng, “Direction-aware spatial context features for shadow detection and removal,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 11, pp. 2795–2808, 2019.

[12] X. F. Zhang, C. C. Gu, and S. Y. Zhu, “Spa-former: Transformer image shadow detection and removal via spatial attention,” arXiv preprint arXiv:2206.10910, 2022.

[13] X. Huang, G. Hua, J. Tumblin, and L. Williams, “What characterizes a shadow boundary under the sun and sky?,” in 2011 international conference on computer vision, pp. 898–905, IEEE, 2011.

[14] J.-F. Lalonde, A. A. Efros, and S. G. Narasimhan, “Detecting ground shadows in outdoor consumer photographs,” in European conference on computer vision, pp. 322–335, Springer, 2010.

[15] J. Liu, T. Fang, and D. Li, “Shadow detection in remotely sensed images based on self-adaptive feature selection,” IEEE Transactions on Geoscience and Remote Sensing, vol. 49, no. 12, pp. 5092–5103, 2011.

[16] S. H. Khan, M. Bennamoun, F. Sohel, and R. Togneri, “Automatic feature learning for robust shadow detection,” in 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1939–1946, IEEE, 2014.

[17] L. Qu, J. Tian, S. He, Y. Tang, and R. W. Lau, “Deshadownet: A multicontext embedding deep network for shadow removal,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4067–4075, 2017.

[18] M. Gryka, M. Terry, and G. J. Brostow, “Learning to remove soft shadows,” ACM Transactions on Graphics (TOG), vol. 34, no. 5, pp. 1–15, 2015.

[19] R. Guo, Q. Dai, and D. Hoiem, “Paired regions for shadow detection and removal,” IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 12, pp. 2956–2967, 2012.

[20] J. Zhu, K. G. Samuel, S. Z. Masood, and M. F. Tappen, “Learning to recognize shadows in monochromatic natural images,” in 2010 IEEE Computer Society conference

on computer vision and pattern recognition, pp. 223–230, IEEE, 2010.

[21] T. F. Y. Vicente, L. Hou, C.-P. Yu, M. Hoai, and D. Samaras, "Large-scale training of shadow detectors with noisily-annotated shadow examples," in European Conference on Computer Vision, pp. 816–832, Springer, 2016.

[22] B. Ding, C. Long, L. Zhang, and C. Xiao, "ARGAN: Attentive recurrent generative adversarial network for shadow detection and removal," in Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10213–10222, 2019.