

Shadow Detection and Removal from video using Deep Learning

by

Krutika Dodiya
202111079

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

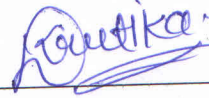


May, 2023

Declaration

I hereby declare that

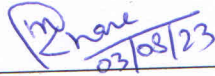
- i) the thesis comprises of my original work towards the degree of Master of Technology in Information and Communication Technology at Dhirubhai Ambani Institute of Information and Communication Technology and has not been submitted elsewhere for a degree,
- ii) due acknowledgment has been made in the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



Krutika Dodiya

Certificate

This is to certify that the thesis work entitled **Shadow Detection and Removal From Video using Deep Learning** has been carried out by **Krutika Dodiya** for the degree of Master of Technology in Information and Communication Technology at *Dhirubhai Ambani Institute of Information and Communication Technology* under my/our supervision.



Dr. Manish Khare
Thesis Supervisor

Dr. Bakul Gohel
Thesis Co-Supervisor

Acknowledgments

First and foremost, I would like to thank my parents, Mr Dilipbhai Dodiya and Mrs Kusumben Dodiya, for always supporting and believing that I can achieve what I deserve. It would not have been possible to complete my MTech without their support.

I would like to express heartfelt gratitude to my thesis supervisor Dr. Manish Khare for their untiring guidance and constant support throughout my research. I am grateful to my thesis co-supervisor Dr. Bakul Gohel for their valuable input, encouragement, and support throughout this journey. Their expertise and guidance have been pivotal in enriching my understanding and enhancing the overall quality of my work. I am extending my gratitude to Intelligence Surveillance Research Lab for providing me necessary resources and the GPU system, and I am also grateful for having supporting lab members.

I am appreciative of my friends Divya, Nisarg, Dhairya, Ruchita, Hemani, Harsh P, Parth, Hardi and all of my classmates for their unwavering support and insightful comments. I was able to get through because of their never-ending support and encouragement. In closing, I want to express my gratitude to everyone at the Institute who made this time of learning possible for me.

Contents

Abstract	v
List of Tables	vii
List of Figures	viii
1 Introduction	1
1.1 Types of Shadow	1
1.2 Cast Shadow and Self Shadow	2
1.3 Hard Shadow and Soft Shadow	3
1.4 Motivation	4
1.5 Objective	4
1.6 Contribution	5
1.7 Organization of Thesis	6
2 Literature Survey	7
2.1 Past Research	7
2.2 Properties-Based Methods	8
2.3 Deep Learning Based Methods	8
2.4 Generative Adversarial Network based methods	10
2.5 U-Net	12
2.6 Details of Data Requirement	15
2.7 Overview of Datasets	15
3 Proposed Method	17
3.1 Shadow Detection Using U-Net	17
3.2 Shadow Removal Using Cascade U-Net	19
3.3 Loss Function	22
4 Experimental Results	23
4.1 Experimental Dataset	23

4.2	Evaluation Parameters	25
4.3	Shadow Detection Evaluation	26
4.4	Shadow Detection Results on ISTD	26
4.5	Shadow Removal Results on ISTD	27
4.6	Shadow Removal Results on SRD	29
4.7	Ablation studies	31
5	Conclusion and Future Scope	33
	References	34

Abstract

The removal of shadow from images is crucial in computer vision as it can enhance the interpretability and visual quality of images. This research work proposes a cascade U-Net architecture for the shadow removal, consisting of two stages of U-Net Architecture. In the first stage, a U-Net is trained using the shadow images and their corresponding ground truth to predict the shadow free images. The second stage uses the predicted shadow free images and ground truth as input to another U-Net, which further refines the shadow removal results. This cascade U-Net architecture enables the model to learn and refine the shadow removal progressively, leveraging both the initial predictions and ground truth.

Experimental evaluations on benchmark datasets demonstrate that our approach achieves notably good performance in both qualitative and quantitative evaluations. By using both objective metrics such as Structural Similarity Index(SSIM), and Root mean Square Error (RMSE), and subjective evaluations where human observers rate the quality of the shadow removal results, our approach was found to outperform other state-of-the-art methods. Overall, our proposed cascade U-Net architecture offers a promising solution for the shadow removal that can improve image quality and interpretability.

Key Terms— shadow removal, cascade U-Net, deep learning, computer vision, image processing.

Nomenclature

i_{shadow}	Shadow Image
$L(i_{gt}, i_{pred})$	Loss of i_{gt} and i_{pred}
i_{gt}	Ground Truth of Shadow Image
i_{pred}	At Intermediate Predicted Shadow Free Image
BCE	Binary Cross Entropy
CGAN	Conditional Generative Adversarial Network
CNN	Convolution Neural Network
CycGAN	Cyclic Generative Adversarial Networks
GAN	Generative Adversarial Network
GT	Ground Truth
HSV	Hue Saturation Value
ISTD	Image Shadow Triplets Dataset
MSGAN	Mask-Shadow Generative Adversarial Network
ReLU	Rectified Linear Unit
RMSE	Root Means Square Error
SRD	Shadow Removal dataset
SSIM	Structural Similarity Index
ST-CGAN	Stacked Conditional Generative Adversarial Networks
TCGAN	Target-Consistency Generative Adversarial Network

List of Tables

2.1	Methods for shadow detection and removal	8
2.2	Data Requirements and Type of Data for below approaches	15
2.3	Datasets for Shadow Detection and Removal	16
3.1	Summary of the shadow detection training parameter for U-Net architecture.	19
3.2	Summary of training parameter for cascade U-Net architecture. . .	20
4.1	Quantitative Shadow Detection results with RMSE and Dice coefficient on ISTD test dataset.	26
4.2	Evaluation of removal effectiveness using RMSE on the ISTD test dataset, showcasing quantitative outcomes.	28
4.3	Quantitative shadow removal results with RMSE on SRD [16] test dataset.	29

List of Figures

1.1	Object with Shadow	1
1.2	Types of Shadow.	2
1.3	Hard and Soft Shadow	3
1.4	Transforms shadow domain image to shadow free domain image.	4
1.5	Results on ISTD	5
2.1	Hieu <i>et al</i> [12] method framework	10
2.2	Generative Adversarial Networks Architecture.	11
2.3	Conditional Generative Adversarial Networks Architecture.	11
2.4	Image Segmentation using U-Net	13
2.5	Standard U-Net Architecture	14
3.1	U-Net for the shadow Detection	18
3.2	Proposed architecture:Cascade U-Net	21
4.1	ISTD Benchmark Dataset	24
4.2	SRD Dataset	24
4.3	Shadow detection on benchmark ISTD	27
4.4	Comparison of Shadow removal results of different methods on ISTD dataset	28
4.5	Visual Results of SRD [14] dataset	30
4.6	Ablation study 1	31
4.7	Ablation study 2	32
4.8	Ablation study 3	32

CHAPTER 1

Introduction

A shadow is an area of darkness created when an opaque object blocks the passage of light from a source, resulting in a contrasting absence of illumination. Figure 1.1 shows occurrence of the shadow due to blockage of light by object.



Figure 1.1: Object with Shadow

1.1 Types of Shadow

Several factors influence the occurrence and characteristics of different types of shadow, There are primarily two distinct types of shadow that occur depending on the position of the lighting source and the intensity of the light, namely hard and soft shadow. Additionally, we can categorize shadow as cast shadow and self shadow.

1.2 Cast Shadow and Self Shadow

Cast shadow[21] occur when an object blocks a light source, resulting in the formation of a shadow on the surface or background behind it. These shadow typically appear as darker regions with well-defined edges and can provide important cues about the shape, position, and orientation of objects in an image. As

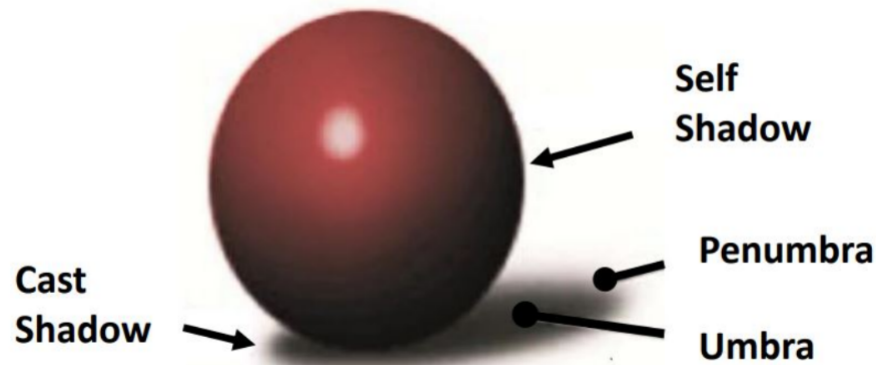


Figure 1.2: Types of Shadow.

shown in Figure 1.2, Cast shadow [21] also have two variant are named as umbra and penumbra. The umbra [21] is characterized by its high level of darkness and the absence of any discernible details or illumination. It has well-defined edges and represents the area of total shadow. The penumbra [21] is the transitional region surrounding the umbra. It is an area where partial illumination occurs, as the light source is only partially blocked by the object casting the shadow. In the penumbra, there is a gradual transition from the darkest region (umbra) to the fully illuminated area outside the shadow. The penumbra appears less dark than the umbra and typically exhibits softer edges due to the partial obstruction of the light source.

In Figure 1.2, Self shadow [10] refer to shadow that objects cast on themselves due to the presence of multiple light sources or complex lighting conditions. These shadow often occur on curved or irregular surfaces and can introduce variations in illumination and color across an object, making it challenging to analyze and interpret.

1.3 Hard Shadow and Soft Shadow

Well-defined, crisp edges characterize hard shadow [10], which arise when the light source is relatively small, intense, and the objects that cast the shadow possess distinct boundaries. Hard shadow are typically seen when the light source is direct, such as sunlight on a clear day or a focused artificial light. The edges of the shadow are sharp, creating a high contrast between the shaded area and the illuminated surroundings. In terms of appearance, hard shadow produce a clear separation between the shadow and the object, resulting in a distinct dark region.

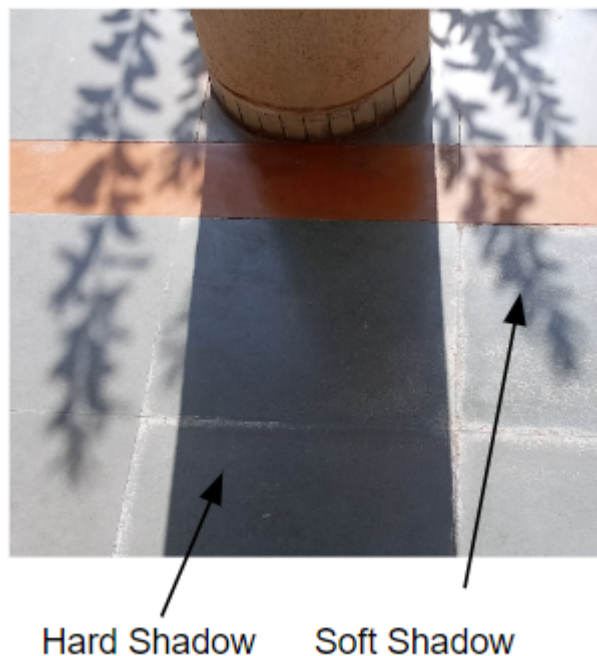


Figure 1.3: Hard and Soft Shadow

In Figure 1.3, Soft shadow[21] on the other hand, have blurred or diffuse edges. They occur when the light source is larger in size or when the object casting the shadow is in close proximity to the surface it falls upon. Soft shadow can be observed in various lighting conditions, such as an overcast sky or when light passes through a translucent material. Unlike hard shadow, soft shadow exhibit a gradual transition from the shaded area to the illuminated region, resulting in smoother and more subtle effect. The intensity of the shadow gradually decreases as it moves away from the object, leading to a more gentle and diffused appearance.

1.4 Motivation

Accurate shadow detection and removal techniques have crucial applications in domains like autonomous driving, surveillance, augmented reality, image/video editing, medical imaging etc. Shadow pose challenges for object detection, tracking, behaviour analysis, visual coherence, and diagnostic accuracy. By providing robust and reliable solutions, our research aims to enhance object recognition and scene understanding in autonomous driving systems, improve tracking accuracy and reliability in surveillance, enhance visual realism in augmented reality, refine visual quality in image/video editing, and aid accurate analysis in medical imaging. These advancements will contribute to safer autonomous vehicles, more effective surveillance, immersive augmented reality experiences, enhanced visual content, and improved medical diagnostics.

1.5 Objective

Given the domain of shadow images I_x and the domain of shadow free images I_y , we are primarily focused on learning the mapping function $I_{SF}: I_x \rightarrow I_y$, which transforms shadow domain image to shadow free domain image. Here Figure 1.4 shows visual representation of the research work objective.

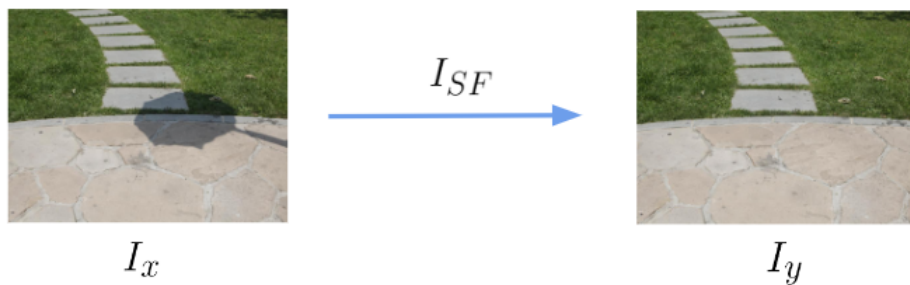


Figure 1.4: Transforms shadow domain image to shadow free domain image.

1.6 Contribution

In this study, we introduce a novel approach for shadow removal using a cascade U-Net architecture. Our proposed method consists of two U-Net architectures connected in sequence. The initial U-Net takes the shadow image and the corresponding ground truth as input, producing an initial output referred to as i_{pred} . Subsequently, the second U-Net leverages this predicted output i_{pred} and the ground truth i_{gt} as input to generate a refined shadow removal image, termed $i_{finalPred}$. By employing this cascade architecture, our method aims to enhance the accuracy and quality of shadow removal results.

The key contributions of this work are as follows:

- We present a pipeline that removes the shadow using deep learning approach.
- We introduce Cascade U-Net that effectively remove the complex shadow from images.
- Proposed cascade U-Net architecture gives comparatively good results benchmark datasets like ISTD[20] and SRD [16].



Figure 1.5: Shadow_img is Original shadow image, Ground_truth is represent Ground truth of shadow image and Predicted_img is represent final output of cascade U-Net.

1.7 Organization of Thesis

An Overview of the organization of my thesis is given below:

Chapter 2 Literature Survey which discusses work done in the field of shadow detection and removal using various Properties-based, Deep learning based approaches. **Chapter 3** explains in detail the proposed Method, In which the overall scheme is discussed by the proposed model architecture for shadow detection using U-Net and shadow removal using cascade U-Net. Also discussed about loss function, which is used in the model. **Chapter 4** Contain results of the proposed method on the ISTD and SRD dataset with comparisons with other approaches. Evaluation parameters for shadow removal are SSIM and RMSE, also discussed. Visual results of shadow detection and, quantitative results and ablation studies of this thesis work. **Chapter 5** shows conclusion and future scope.

CHAPTER 2

Literature Survey

Chapter 2 presents an overview of methods for shadow detection and removal, including Feature-based and Deep learning-based approaches. Feature-based methods leverage properties such as geometry, color, and texture to identify and eliminate Shadow. Deep learning-based methods utilize neural networks and techniques like GANs for accurate shadow detection and removal. The chapter also discusses the data requirements for these methods and highlights available datasets for training and evaluating shadow removal algorithms.

2.1 Past Research

In the domain of shadow detection and removal from image/video, many notable work has been done in last. In the early era, properties-based approaches were used, which involved utilizing properties such as chromaticity, color, features, and texture. Deep learning and GAN-based methods have shown promising results for shadow removal. By leveraging the power of neural networks and adversarial training, these approaches can automatically extract meaningful features and generate visually appealing outputs. Compared to traditional properties-based approaches, deep learning and GAN-based methods offer more accurate and effective solutions for shadow removal tasks.

Table 2.1 presents an overview of methods for shadow detection and removal. Table 2.1 outlines various techniques categorized by method type, such as feature-based and deep learning approaches and GAN based approaches. These methods include Prati *et al.*[15], Guo *et al.*[6], Yang *et al.*[22] , Gong *et al.*[5], Khare *et al.*[11], DshadowNet [16], Fan *et al.*[3], StackedCNN [19], FusionNet[4], SCGAN [14], ST-CGAN [20], and TCGAN [18] methods overview.

Table 2.1: Methods for shadow detection and removal

Method	Type of Method	Aspect
Prati <i>et al.</i> [15]	Feature (Physical)	shadow detection
Guo <i>et al.</i> [6]	Feature (Color)	shadow detection, shadow removal
Yang <i>et al.</i> [22]	Feature (Color)	shadow removal
Gong <i>et al.</i> [5]	Feature (Color)	shadow removal
Khare <i>et al.</i> [11]	Feature (Wavelet Feature)	shadowdetection, shadow removal
DeshadowNet [16]	DL (CNN)	shadow removal
Fan <i>et al.</i> [3]	DL (CNN)	shadow removal
StackedCNN [19]	DL (CNN)	shadow detection
FusionNet [4]	DL (CNN)	shadow removal
SCGAN [14]	DL (GAN)	shadow detection
ST-CGAN [20]	DL (GAN)	shadow detection, shadow removal
TCGAN [18]	DL (GAN)	shadow removal

2.2 Properties-Based Methods

Shadows are a real-world occurrence. Early efforts in shadow detection and removal mostly concentrated on researching several physical shadow features, such as texture, chromaticity, intensity, etc. The term "Properties-based technique" refers to these. Techniques like colour-based and texture-based approaches are examples of Properties-based strategy. Guo *et al.*[6] simplification of this model uses a linear system to depict the relationship between shadow pixels and shadow free pixels. They achieve this by matching regions that have Shadow and those that don't have shadow. Khare *et al.* [11] proposed method on discrete wavelet transform (DWT). DWT is used to suggest a novel method for shadow identification and removal. The multi-resolution capability of DWT, which divides an image into four distinct bands without sacrificing the spatial information.

2.3 Deep Learning Based Methods

In the field of deep learning, there have been notable advancements in recent years in methods that assess the mapping relationship between shadow and shadow free domains.

Wang *et al.*[20] introduced a solution that combined the colour-based and model-based approaches to shadow identification. The shadow is first identified in its moving region using a properties-based method, after which a coarse region

is created using a model-based method for moving shadow. This coarse region is then employed in a shadow detection method based on HSV color space. The suggested method worked effectively at detecting moving shadow, especially those that fell within the vehicle’s boundary. However, it has been determined that this strategy is not appropriate for shadow that are cast by or beneath moving vehicles.

To map shadow pixels with shadow free pixels, Hieu *et al.*[12] uses the shadow illumination model. Framework of Hieu *et al.*[12] method is in Figure 2.1 estimate the physical illumination model’s parameters using deep neural networks. To anticipate the shadow parameter and determine the shadow matte, Hieu *et al.*[12] uses deep networks like SP-NET, M-NET and I-NET. I-NET is another deep network used to refine the outcome.

The system models the shadow free image using the shadow image, the shadow parameter, and the shadow matte. The shadow free image can be expressed as shown in equation 2.1,

$$I_{\text{shadow free}} = I_{\text{relit}} \cdot \alpha + I_{\text{shadow}} \cdot (1 - \alpha) \quad (2.1)$$

where shadow image and shadow free image I_{shadow} and $I_{\text{shadow free}}$ respectively. α is the matting layer, and relit image represent as I_{relit} . Each pixel i of the relit image I_{relit} is computed by equation 2.2,

$$I_{\text{relit}_i} = w \cdot I_{\text{shadow}_i} + b \quad (2.2)$$

The shadow free image is created by linearly combining the relit image with the input shadow image. The matting layer α represents the per-pixel coefficients of this process. The value of α should ideally be 1 in the non-shadow area and 0 at the shadow area’s umbra. Near the shadow boundary, the value of progressively changes for the pixels in the penumbra of the shadow. The matting layer α computed by equation 2.3,

$$\alpha_i = \frac{I_{\text{shadow free}_i} - I_{\text{relit}_i}}{I_{\text{shadow}_i} - I_{\text{relit}_i}} \quad (2.3)$$

The matting layer α represents the per-pixel coefficients of the linear combination of the relit image and the input shadow image that results in the shadow free image. Ideally, the value of α should be 1 at the non-shadow area and 0 at

the umbra of the shadow area. For the pixels in the penumbra of the shadow, the value of α gradually changes near the shadow boundary.

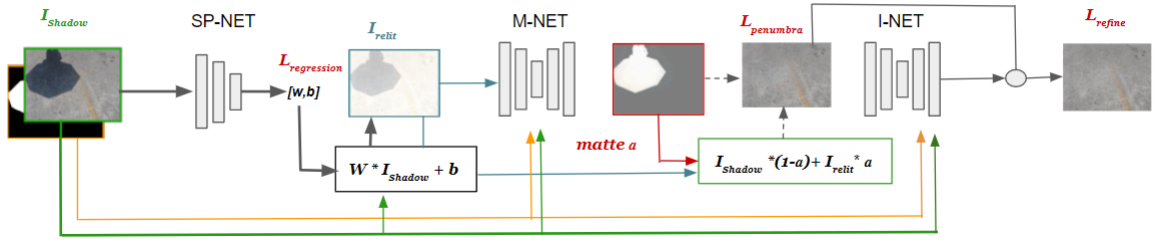


Figure 2.1: Framework of Hieu *et al*[12] method.

Deep learning-based methods for shadow detection and removal are currently encouraged by the availability of large-scale datasets like ISTD[20], SBU[19], USR[7], and others. The most significant result in the detection and removal of Shadow is obtained using GAN. Deshadow-Net, which is trained end-to-end, is introduced by Quet *al.*[16]. In order to forecast shadow matte, the model produced results by extracting multi-context features at each layer of the network. The main function of this is to learn the mapping between the shadow image and its shadow matte, and then utilize the predicted shadow matte to reconstruct an image without shadow.

2.4 Generative Adversarial Network based methods

In recent years, the utilization of Generative Adversarial Networks (GANs) and their subsequent advancements has emerged as a prominent approach for addressing various challenges related to translating images. Illustrated in Figure 2.2, GAN consists of two Deep Neural Network architectures: a generator denoted as G , and a discriminator denoted as D . The generator model G is capable of producing authentic-looking synthetic images. The purpose of the discriminator model D is to determine whether an input image originates from G or belongs to the genuine set of training images. Both models are trained together in an adversarial manner, wherein the objective is for G to generate samples that deceive D , while D strives to avoid being tricked by G . The ultimate goal is for G to generate images that closely resemble the authentic dataset, making it challenging for D to differentiate between the creations of G and real images.

Unlike traditional generative models, CGANs introduce a conditioning variable to control the generation process, allowing the generation of samples tailored to specific attributes or classes. Figure 2.3 shows, The conditioning variable serves

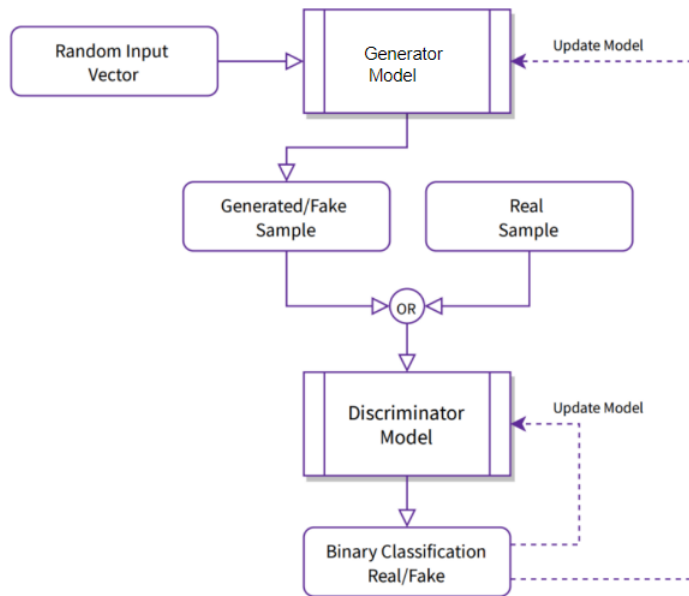


Figure 2.2: Generative Adversarial Networks Architecture.

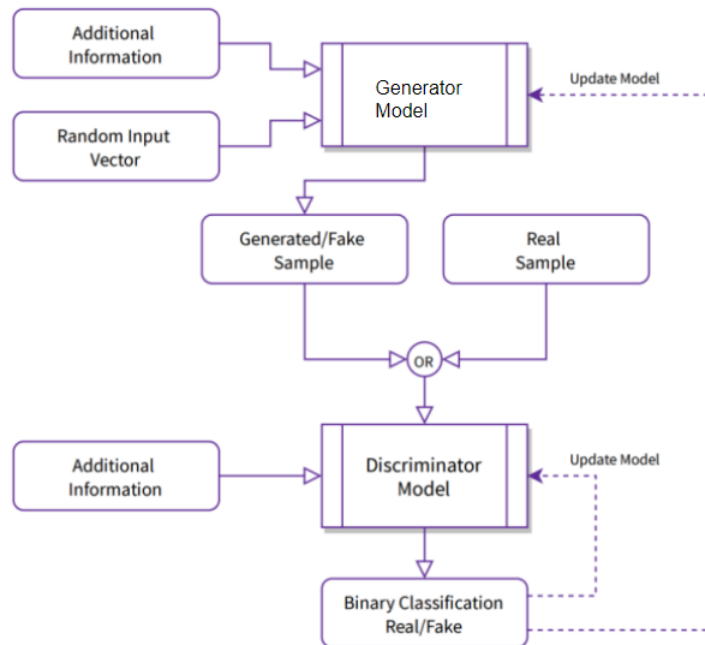


Figure 2.3: Conditional Generative Adversarial Networks Architecture.

as additional information for both the generator and discriminator networks, enabling them to learn the conditional distribution of the data.

Wang *et al.*[20] proposed model is based on a novel STacked Conditional Generative Adversarial Network (ST-CGAN), which consists of two stacked Conditional Generative Adversarial Networks (CGANs) with a generator and a discriminator in each. Using Conditional GAN, this model is capable of both shadow detection and shadow removal. The architecture uses GAN loss to increase model accuracy and uses shadow mask and shadow free images for taring. In particular, the first generator receives a shadow image and outputs a shadow detection mask. This shadow picture is combined with its expected mask and then passed through the second generator to get its shadow free image. Additionally, the two associated discriminators for the detected shadow region and reconstruction by removing shadow, respectively, are very likely to reflect higher level relationships and overall image properties.

Recently, by using properties of shadow with GAN for shadow removal task. Channel Attention GAN CANet[1], it employs two networks for shadow detection and removal. The approach takes into account the physical characteristics of shadow and the camera's image acquisition system. Network architecture considers the relationship between color channels to enhance performance. During training, the authors modify the color and introduce artifacts to the training images, increasing the complexity of the dataset.

2.5 U-Net

The U-Net architecture is a popular Conventional Neural Network (CNN) model designed explicitly for image segmentation tasks. In 2015, it was introduced by Ronneberger *et al.*[17] and has since been widely adopted in various medical imaging applications. For image segmentation, U-Net architecture is frequently employed. It contains both an encoder and a decoder. Because of how all the layers are stacked, the architecture of the U-Net, where the encoder comes first and is followed by the decoder, is known as a U shape.

The standard U-Net architecture is represent by Figure 2.5. The encoder is identical to the standard convolutional network architecture, also known as the U-Net contraction path. A Rectified Linear Unit (ReLU), stride two downsam-

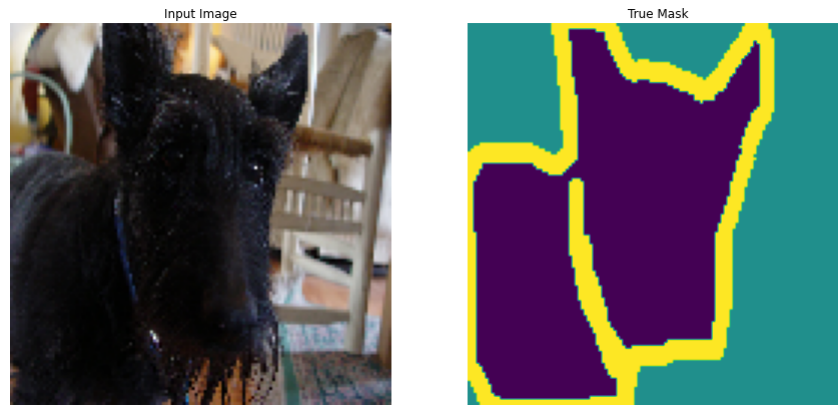


Figure 2.4: Image Segmentation using U-Net

plings, and a max pooling operation divides the four convolution layers. Two 2D convolutions with a 3×3 size kernel make up each layer. The decoder, also known as the expansion path, comes after the contraction path, which consists of four deconvolution layers. Each deconvolution layer (ReLU) consists of a 2D transposed convolution operation for upsampling, concatenation with a corresponding cropped feature map from the contraction path, two convolutions, and then a rectified linear unit.

Upsampling and regaining the spatial data that was lost during encoding are tasks of the decoder path. The feature maps are gradually enlarged, and they are combined with the skip connections from the respective encoder layers. By combining data from many scales, the model is able to accurately segment data while also capturing local and global contexts. Upsampling operations are typically followed by convolutional layers in the decoder route. The feature maps spatial resolution is increased by the upsampling techniques, enabling the model to retrieve finer features. Transposed convolution, commonly referred to as deconvolution, or bilinear interpolation are frequent methods for upsampling. The specific needs of the task and the available processing resources may influence the upsampling method selection.

The output of the extended path is then applied to a further 1×1 convolution to obtain the final output. For the training process, *Adam's* optimizer and *Binary Cross Entropy* as the loss function were optimised, with an 80:20 split between the training and validation stages. The training images were initially reduced in size to 512×512 patches to fit within the network design.

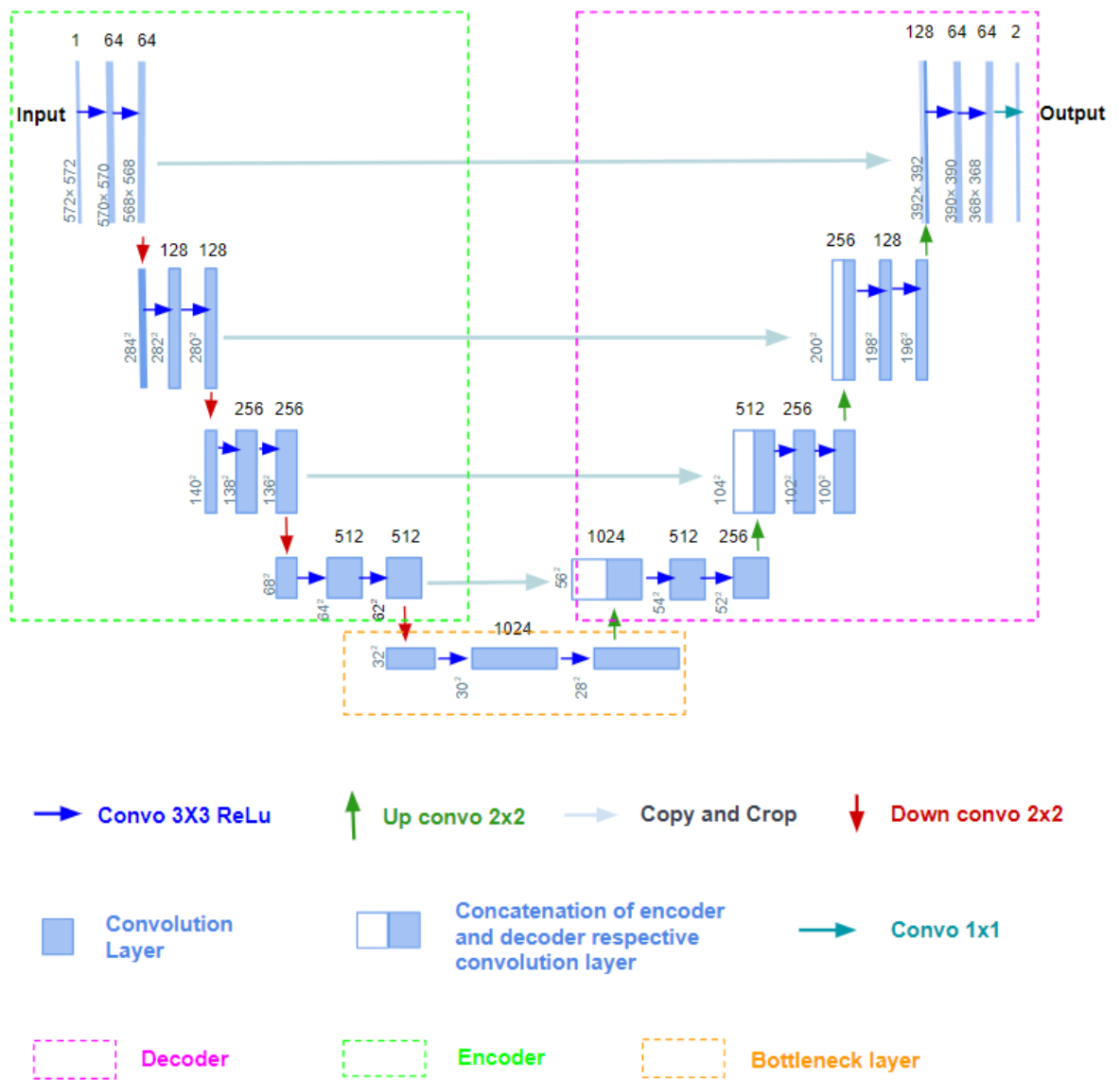


Figure 2.5: Standard U-Net Architecture[17].

2.6 Details of Data Requirement

Table 2.2 highlights the data requirements for these methods, specifying the types of data needed, including shadow, shadow-mask, and shadow free images. Some methods require paired data, while others can work with unpaired data. This summary provides a glimpse into the methods and data requirements for shadow detection and removal.

Table 2.2 provides a comprehensive overview of the data requirements and types needed for various approaches in shadow detection and removal. For instance, DeshadowNet [16], Bansal *et al.*[2], Fan *et al.*[3], StackedCNN [19], SCGAN [14], ST-CGAN [20], Nagae *et al.*[13], DSC [9], FusionNet [4], and MSGAN [7] rely on shadow and shadow-mask images, along with shadow free images, which need to be paired. On the other hand, MSGAN [7] and TCGAN [18] can work with unpaired data consisting of shadow and shadow free images. This Table serves as a valuable resource, shedding light on the different techniques employed and the corresponding data requirements for effective shadow detection and removal.

Table 2.2: Data Requirements and Type of Data for below approaches

Method	Data Required	Type of Data
DeshadowNet [16]	shadow, shadow free	Paired
Bansal <i>et al.</i> [2]	shadow, shadow-mask	Paired
Fan <i>et al.</i> [3]	shadow, shadow free	Paired
StackedCNN [19]	shadow, shadow-mask	Paired
SCGAN [14]	shadow, shadow-mask	Paired
ST-CGAN [20]	shadow, shadow-mask, shadow free	Paired
Nagae <i>et al.</i> [13]	shadow, shadow-mask, shadow free	Paired
DSC [9]	shadow, shadow-mask, shadow free	Paired
FusionNet [4]	shadow, shadow-mask, shadow free	Paired
MSGAN [7]	shadow, shadow free	Unpaired
TCGAN [18]	shadow, shadow free	Unpaired

2.7 Overview of Datasets

Table 2.3 shows the details of large-scale available shadow datasets with the amount and type of data. There are several large-scale shadow datasets that have been published for training and evaluating shadow removal methods. Table 2.3 contain benchmark dataset which include the Image Shadow Triplets Dataset

(ISTD) [20], Shadow Remove Dataset with shadow images and shadow free image(SRD)[16], the Shadow Remove Dataset with shadow images and masks (SBU) [19], and the Unpaired Shadow Removal Dataset (USR) [7]. These datasets provide diverse and extensive shadow images, ground truth shadow free images, and shadow masks for training and testing shadow removal algorithms. They enable researchers and practitioners to develop and evaluate shadow removal methods on a large scale, facilitating the advancement of shadow removal techniques in the field of computer vision.

Table 2.3: Datasets for Shadow Detection and Removal

Dataset	Amount	Content	Type of Data
SRD[16]	3088	shadow, shadow free	Paired
ISTD [20]	1870	shadow, shadow-mask, shadow free	Paired
USR [7]	2445	shadow, shadow free	Unpaired
SBU [19]	4723	shadow, shadow-mask	Paired
CUHK [8]	10500	shadow, shadow-mask	Paired

CHAPTER 3

Proposed Method

Chapter 3 presents a novel cascade U-Net architecture for the shadow removal in computer vision. The proposed method utilizes two sequentially connected U-Net Architectures to iteratively refine the shadow removal process. The architecture captures both local and global contextual information, improving the accuracy and quality of results. The experimental evaluations on benchmark datasets demonstrate promising outcomes, also a section discussed shadow detection using U-Net. The chapter also discusses the training parameters and loss function used, emphasizing the significance of *Binary Cross Entropy* for optimizing the model.

3.1 Shadow Detection Using U-Net

For detecting the shadow from images U-Net [17] is used. U-Net is a model which has take input from the shadow image paired with the shadow mask. At first, it is tried with the ISTD dataset containing the shadow image and the shadow mask to predict on testing dataset image mask. ISTD dataset images contain hard shadow. For the training process, *Adam* as optimizer and *Binary Cross Entropy* as the loss function were used for optimization, while the train and validation split ratio was set to be 80:20. Before the training, the training images resize as 128×128 to fit into the network architecture. Figure 3.1 shows the shadow detection using single U-Net. And Table 3.1 gives all information about training parameter for the shadow detection

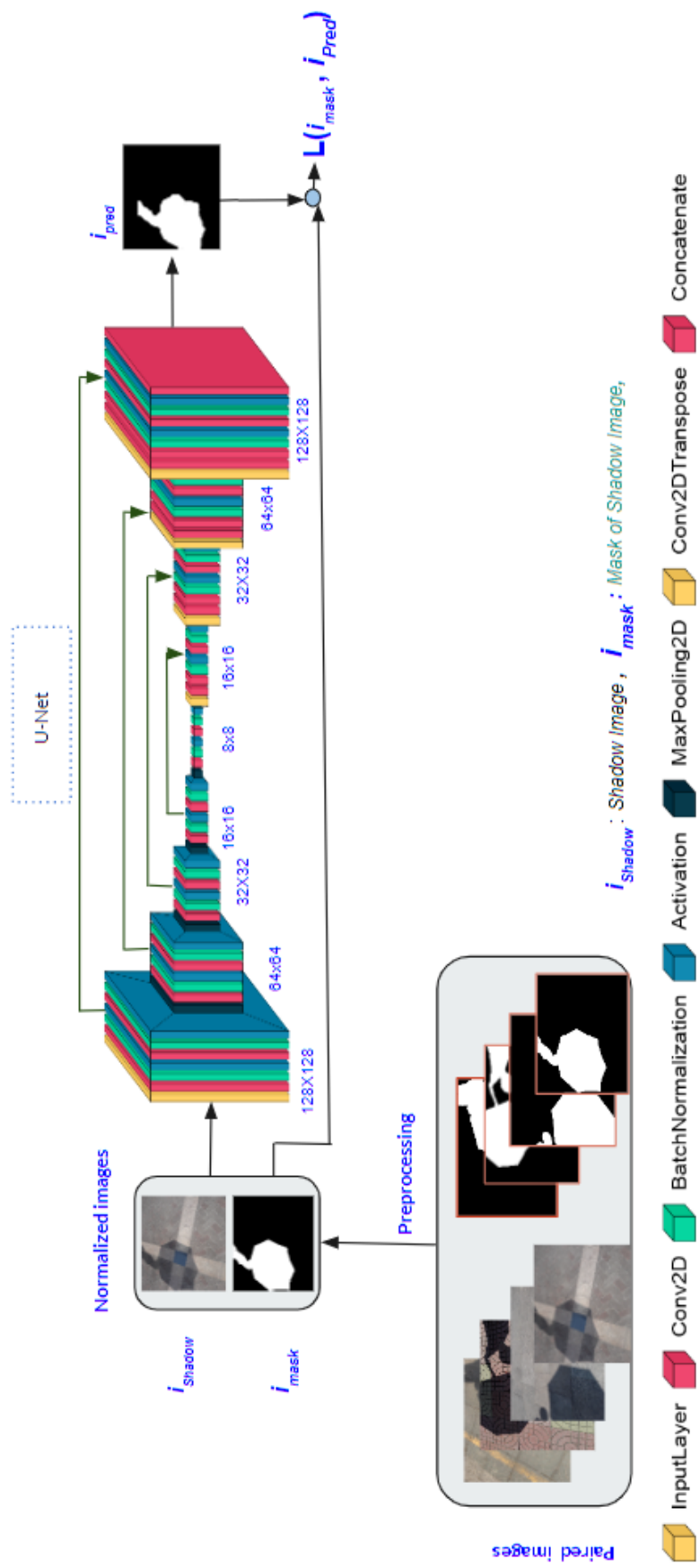


Figure 3.1: U-Net for Shadow Detection

Table 3.1: Summary of the shadow detection training parameter for U-Net architecture.

Training Parameter	Type/value
input	$128 \times 128 \times 3$
Initial Learning rate	1e-7
Filter size	3×3
Pooling size	2×2
Batch size	8
Number of epochs	200
Loss Function	BCE
output	$128 \times 128 \times 3$

3.2 Shadow Removal Using Cascade U-Net

The shadow removal is a critical task in computer vision that aims to enhance image quality and improve the accuracy of downstream image processing tasks by eliminating the shadow. Convolutional neural networks (CNNs), specifically U-Net, have shown great potential in addressing this challenge due to their ability to capture local and global contextual information effectively. We propose a novel cascade U-Net architecture for the shadow removal. This architecture comprises two U-Net networks connected in sequence, offering the potential to effectively eliminate hard shadows.

In this approach, proposed a two-step process for achieving partial shadow removal from input shadow images. Initially, employ the first U-Net model, which takes as input the shadow image along with its corresponding ground truth. Once the first U-Net has been trained, and predict partially shadow removal image. Next, the second UNet model takes as input the primary shadow removal image obtained from the output of the first U-Net and also the original shadow image. The main purpose of the second U-Net is to further refine the partial shadow removal image generated in the previous step.

The first U-Net takes the shadow image and the corresponding ground truth as input and predicts an initial output, denoted as i_{pred} . This predicted output, along with the ground truth, is then fed as input to the second U-Net, which generates a refined shadow removal image, denoted as $i_{finalPred}$.

During the training process, the RMSE (Root Mean Squared Error) is com-

puted to measure the pixel-wise difference between the predicted output and the ground truth, denoted as $L(i_{gt}, i_{pred})$. The ISTD[20] dataset is used for training, which contains the shadow images paired with the shadow masks for the initial training. The *Adam* optimizer and *Binary Cross Entropy* loss function are used for training, with a train-validation split ratio of 80:20. The input image resize to 128x128 to fit the network architecture. Table 3.2 contain information regarding training parameters. After the training of the first U-Net, the loss of after first U-Net achieved 0.5124. The second U-Net is then trained using the shadow free images predicted by the first U-Net and the corresponding shadow free images from the ISTD dataset and achieved loss of 0.3784. Additionally, experiments with the unpaired SRD dataset [16] are also conducted.

In Figure 3.2 , the input images are denoted as i_{shadow} and i_{gt} , which represent the shadow image and the shadow free image, respectively. i_{pred} and $i_{finalPred}$ are the primary result generated by the first U-Net and the final shadow removal image, respectively. $L(i_{gt}, i_{pred})$ and $L(i_{gt}, i_{finalPred})$ represent the error functions, i.e., the RMSE between the ground truth and the predicted output for the first U-Net and the final output, respectively.

Overall, the proposed cascade U-Net architecture for the shadow removal demonstrates promising results, as shown by the experimental evaluations on benchmark datasets. The use of ground truth information and the iterative refinement approach in the cascade U-Net can potentially improve the accuracy and quality of the shadow removal results, contributing to the advancement of computer vision techniques in addressing the challenges of shadow removal in various applications.

Table 3.2: Summary of training parameter for cascade U-Net architecture.

Training Parameter	First U-Net	second U-Net
Initial Learning rate	1e-7	1e-7
Filter size	3×3	3×3
Pooling size	2×2	2×2
Batch size	8	8
input	$128 \times 128 \times 3$	$128 \times 128 \times 6$
output	$128 \times 128 \times 3$	$128 \times 128 \times 3$
Number of epochs	50	50
Loss Function	BCE	BCE

3.3 Loss Function

For the task of shadow removal, various loss functions such as *SSIM* (Structural Similarity Index), *Gradient-based loss*, and *Binary Cross Entropy* (BCE) can be employed. Among these options, *Binary Cross Entropy* has been found to yield favorable results. In the specific architecture of cascade U-Net, *Binary Cross Entropy* is utilized for both U-net components. This choice of loss function helps to effectively train the model and achieve improved performance in the shadow removal task.

Binary Cross Entropy

The *Binary Cross Entropy* (BCE) is a loss function commonly used in machine learning for binary classification tasks. It measures the dissimilarity between predicted probabilities and true labels, enabling the training of models to accurately classify binary outcomes. The BCE loss is computed over all instances in the dataset and then averaged to obtain a single loss value. BCE is essential for optimizing models in various applications, including my thesis on the shadow detection and removal.

$$BCE = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (3.1)$$

- N represents the total number of samples.
- y_i is the true label (either 0 or 1) for the i^{th} sample.
- p_i is the predicted probability for the i^{th} sample.

CHAPTER 4

Experimental Results

In this Chapter, two benchmark datasets ISTD and SRD are used, we demonstrate experimental findings from the proposed approach presented in this chapter. We also present the experimental outcomes of these two benchmark dataset obtained using the proposed approach. Additionally, we present the outcomes of numerous ablation trials at the end of this chapter.

4.1 Experimental Dataset

In order to assess the effectiveness of the proposed framework, we conducted experiments using a dataset called ISTD, which consists of Image Shadow Triplets (ISTD) [20]. This dataset is widely recognized as a significant benchmark for evaluating the shadow detection and removal techniques. It comprises 2410 triplets of images, including the shadow, shadow-mask, and shadow free images, all captured from various scenes. The resolution of these images is 640x480 pixels. The reason for using ISTD for the shadow detection is that it provides shadow masks for the corresponding shadow images. On the other hand, the SRD dataset[16] lacks the shadow masks for the corresponding shadow images. Therefore, both the ISTD and SRD datasets were utilized for shadow removal evaluation purposes. The research work is mainly based on the shadow removal from video, since both dataset contain images which are frames extracted from small videos clips.

Below Figure 4.1 shows sample images of ISTD[20] and Figure 4.2 shows sample images of SRD dataset[16]. SRD[16] contains 2680 pairs of the shadow images and the shadow free images. With an image resolution of 840x640 pixels, which is split in an 80:20 ratio for training and testing.

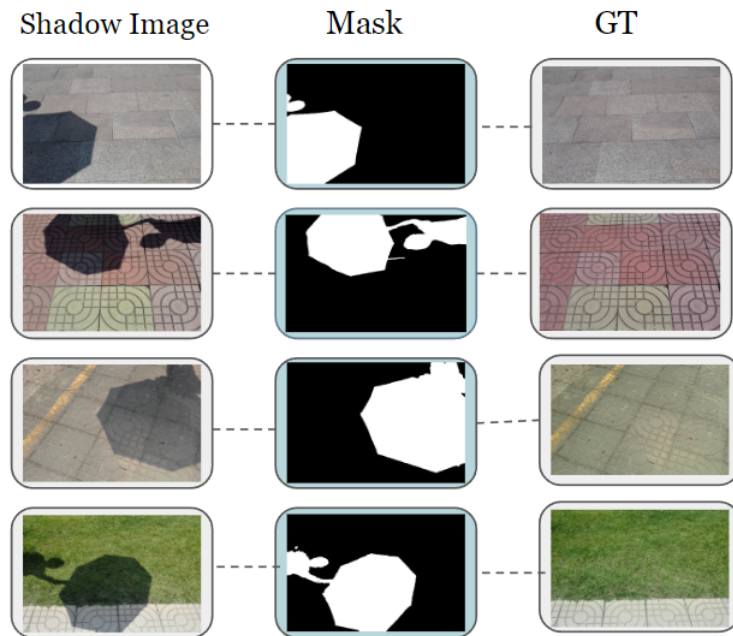


Figure 4.1: Sample of ISTD Benchmark Dataset [20]

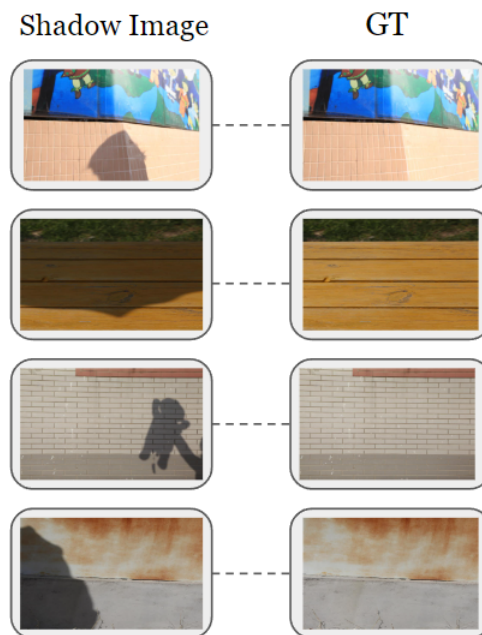


Figure 4.2: Sample of SRD Dataset[16]

4.2 Evaluation Parameters

To assess the generated results, the evaluation metric utilized is the Root Mean Square Error (RMSE) measured in the RGB color space. It calculates the disparity between the ground-truth images and the produced shadow free images. For the quantitative analysis of shadow removal, the pixel-wise concurrence between a predicted image and its corresponding ground truth is measured using the Structural Similarity Index (SSIM). The value of SSIM ranges between 0 and 1, where a higher value indicates better similarity between the images. Equations 4.1 and 4.2 are for SSIM and RMSE respectively, which are shown below. SSIM for the benchmark ISTD dataset on the proposed method is 0.7092

$$\text{SSIM}(\hat{x}, x) = \frac{(2\mu_{\hat{x}}\mu_x + c_1)(2\sigma_{\hat{x}x} + c_2)}{(\mu_{\hat{x}}^2 + \mu_x^2 + c_1)(\sigma_{\hat{x}}^2 + \sigma_x^2 + c_2)} \quad (4.1)$$

$$\text{RMSE}(\hat{x}, x) = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{x}_i - x_i)^2} \quad (4.2)$$

- N is the total number of pixels in the images.
- \hat{x} is the predicted image being compared.
- x is the original image
- $\mu_{\hat{x}}$ and μ_x are the mean values of \hat{x} and x respectively.
- $\sigma_{\hat{x}^2}$ and σ_x are variances of \hat{x} and x respectively
- $\sigma_{\hat{x}x}$ is the covariance between \hat{x} and x
- c_1 and c_2 are small constants added for numerical stability, usually set to $c_1 = (k_1L)^2$ and $c_2 = (k_2L)^2$ Where L is the dynamic range of the pixel values (e.g., 255 for 8-bit images) and k_1 and k_2 are constants typically set to small values.

For both U-Net, RMSE is used as an evaluation metric. SSIM is used as the final result evaluation metric along with RMSE. Both evaluation metrics RMSE and SSIM are based on pixel to pixel wise operation.

4.3 Shadow Detection Evaluation

The evaluation metrics employed to assess the quality of shadow free image involves the utilization of Root Mean Square Error (RMSE). For the quantitative analysis of shadow detection, the dice coefficient was employed to measure the level of agreement at the pixel level between the predicted image and its corresponding ground truth. The equation 4.3 shows the dice coefficient. The Dice coefficient yields a value between 0 and 1.

$$Dice - coefficient = \frac{2 * |A \cap B|}{|A| + |B|} \quad (4.3)$$

Here A and B can be considered as the shadow free predicted image and the ground truth image. Here the below Table 4.1 gives quantitative results on ISTD. The best and second-best results in the Tables are highlighted in blue and bold, respectively.

Table 4.1: Quantitative Shadow Detection results with RMSE and Dice coefficient on ISTD test dataset.

Methods	Dice coefficient	RMSE overall
ST-CGAN [20]	0.5632	7.47
Mask-Shadow GAN [7]	0.6768	6.99
Ours	0.8553	6.45

4.4 Shadow Detection Results on ISTD

The evaluation of shadow detection is conducted on the ISTD test dataset. Our proposed method is compared with the two state-of-the-art approaches, MaskShadowGAN [7] and ST-CGAN [20]. Figure 4.3 displays the shadow detection results on ISTD. Table 4.1 presents the quantitative performance of shadow detection, measured in terms of RMSE. A lower RMSE value indicates superior performance.

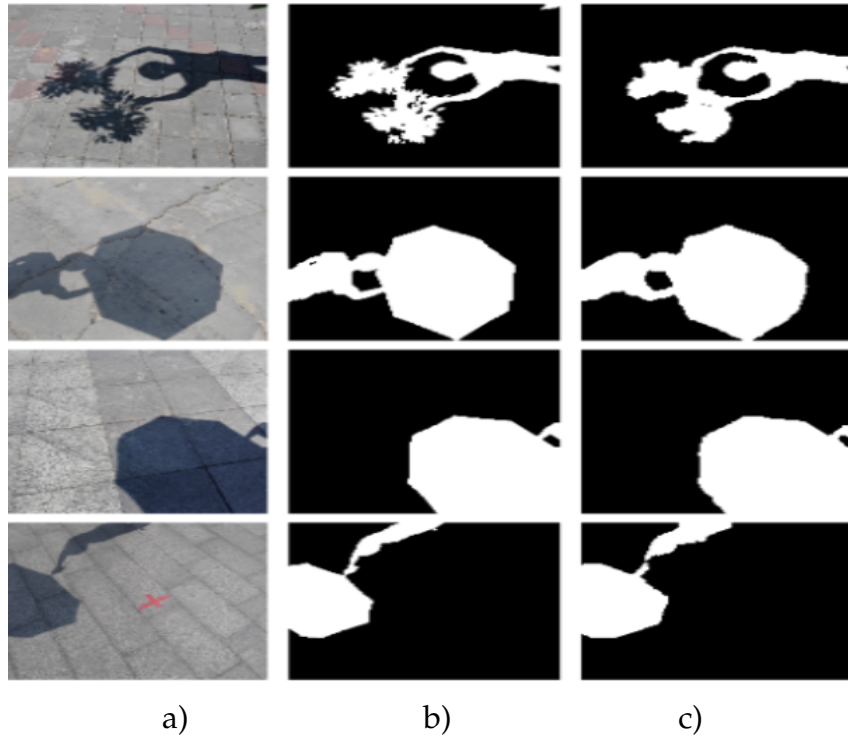


Figure 4.3: shadow detection using U-Net. a) represents the shadow image, b) represents ground truth and c) represents the shadow detection using U-Net.

4.5 Shadow Removal Results on ISTD

The results displayed in Figure 4.4 exhibit the performance of various techniques on shadow removal in ISTD. It is important to highlight that our innovative cascade U-Net approach accomplishes the shadow removal without relying on a shadow image mask, whereas other methods are specifically designed for either shadow detection or shadow removal, but not both of them simultaneously.

In the ISTD test dataset, various approaches for shadow removal were evaluated, and their quantitative results are presented in Table 4.2. Among these methods, the proposed cascade U-Net stands out by achieving an RMSE (Root Mean Squared Error) of 7.05 without relying on a shadow image mask. This outperforms other techniques such as those introduced by Yang *et al.*[22], Gong *et al.*[6], Guo *et al.*[5], ST-CGAN [20], and DSC *et al.*[3]. The original RMSE on the other hand, exhibits 10.97. ST-CGAN[20] has complex architecture that first generate shadow mask and then use it for shadow removal, whereas cascade U-Net has simple architecture.

The visual results on the ISTD test dataset are shown in Figure 4.4 compression

Table 4.2: Evaluation of removal effectiveness using RMSE on the ISTD test dataset, showcasing quantitative outcomes.

Methods	RMSE overall
Original	10.97
Yang <i>et al.</i> [22]	15.63
Gong <i>et al.</i> [5]	9.3
Guo <i>et al.</i> [6]	8.53
Ours	7.05

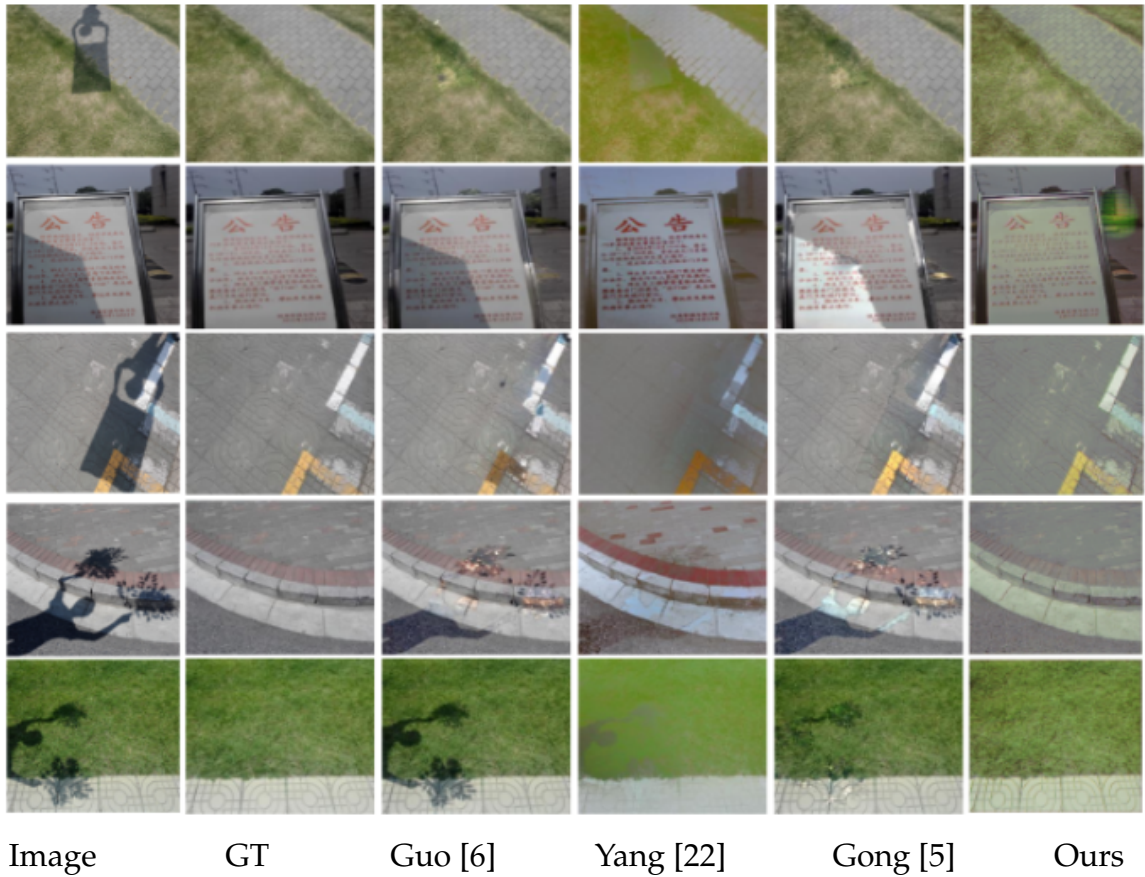


Figure 4.4: Comparison of Shadow removal results of different methods on ISTD dataset

with other methods like Guo *et al.*[5], Yang *et al.*[22], Gong *et al.*[5] with GT of shadow image and Ours proposed model results are shown below. By visualizing Figure 4.4, our method gives comparatively good results than others. Gong *et al.*[5] method gives good results among Guo *et al.*[5] and Yang *et al.*[22].

4.6 Shadow Removal Results on SRD

The performance of our proposed method, trained with SRD [16], is compared with other methods [5, 6, 7, 22, 23] on the SRD test dataset. The visual results of SRD dataset on the proposed cascade U-Net architecture is shown in Figure 4.5. The results, shown in Table 4.3, demonstrate that our method outperforms others in the overall scenario. It is important to note that the comparison of quantitative results is limited to the scenario only, as shadow-mask images are unavailable for the dataset.

Table 4.3: Quantitative shadow removal results with RMSE on SRD [16] test dataset.

Method	RMSE
Original	14.41
Yang <i>et al.</i> [22]	22.57
Guo <i>et al.</i> [6]	12.60
Gong <i>et al.</i> [5]	8.73
CycleGAN [23]	9.14
Mask-ShadowGAN [7]	7.32
Ours	7.85

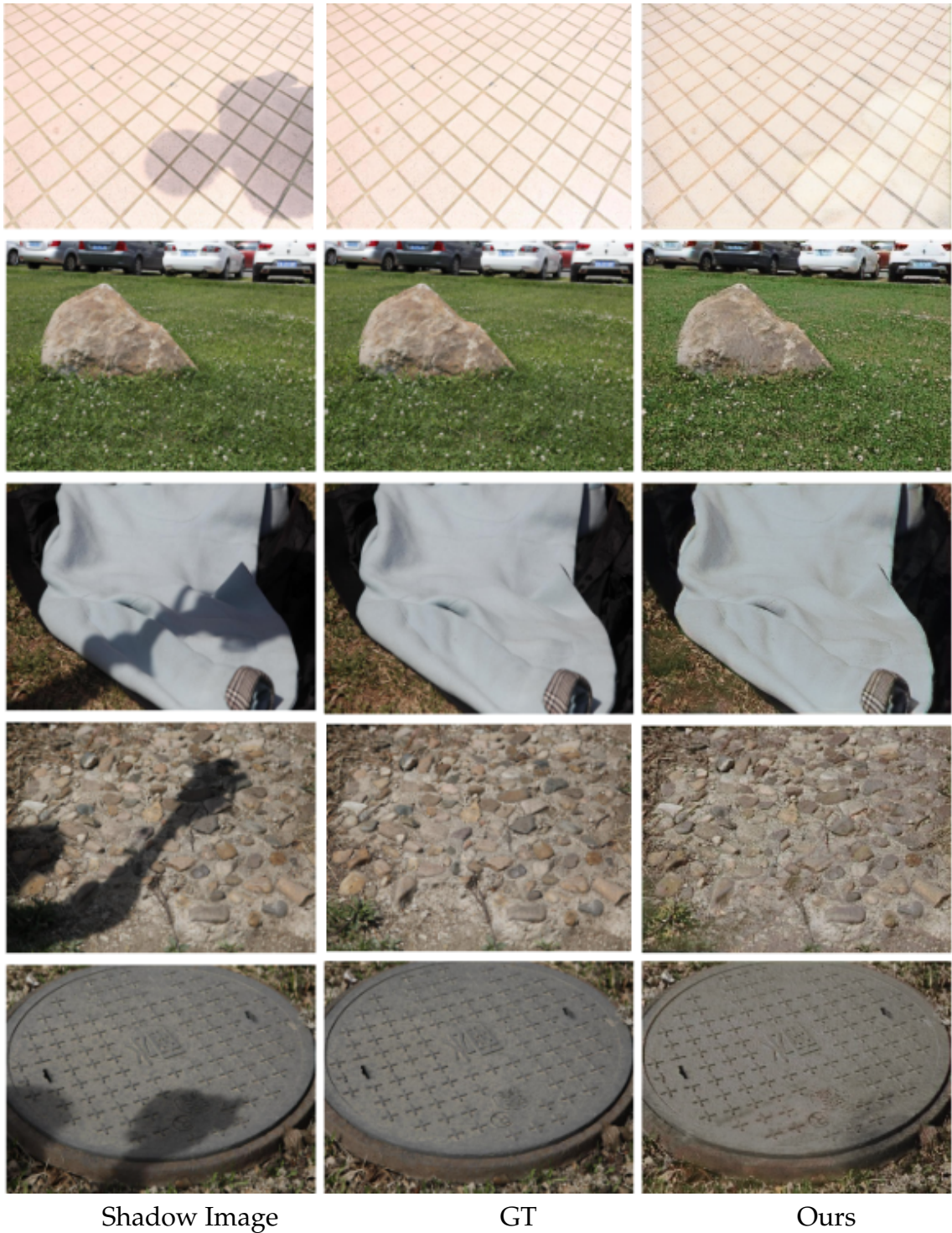


Figure 4.5: Visual performance of shadow removal results on SRD test dataset.

4.7 Ablation studies

In order to remove the shadow from Images, various experiments have been done on the U-Net model and by changing Loss functions.

Result Using Single U-Net with Binary Cross Entropy Loss:

For shadow removal, single U-Net is used with Binary Cross Entropy as loss function. The visual result on ISTD shows in below Figure 4.6. Using single U-Net, the shadow is not properly removed, edges are visible. The single U-Net with BCE loss has not given desirable output.

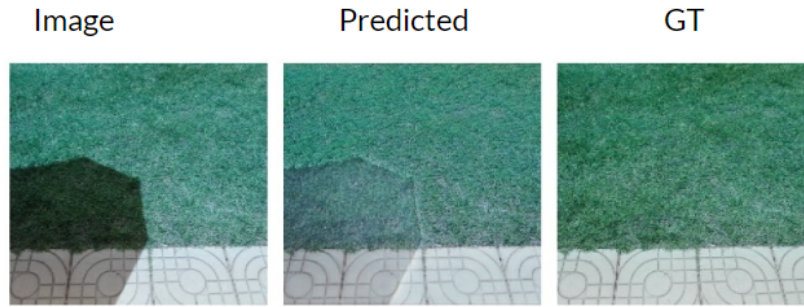


Figure 4.6: BCE Loss on ISRD

Cascade U-Net with Gradient Loss:

In the context of the shadow removal, the gradient loss can be used as a regularization term to encourage smoothness in the estimated shadow free image. The gradient loss penalizes abrupt changes or sharp edges in the output image, promoting a more visually pleasing and realistic result.

$$\text{Gradient Loss} = \sum_{i=1}^n \|\nabla_{\mathbf{x}}f(\mathbf{x}_i) - \nabla_{\mathbf{x}}g(\mathbf{x}_i)\|^2 \quad (4.4)$$

In the above formula 4.4, f and g are functions, n is the number of data points, \mathbf{x}_i represents the i -th data point, and $\nabla_{\mathbf{x}}$ denotes the gradient with respect to \mathbf{x} . The $\|\cdot\|^2$ represents the squared L2 norm of the difference between the gradients. Figure 4.7 shows the result of ISTD with Gradient Loss.

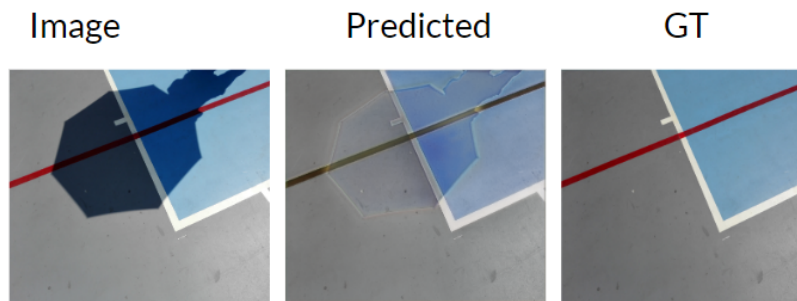


Figure 4.7: Gradient Loss on ISRD

Cascade U-Net with SSIM loss

SSIM loss, a frequently used technique for removing the shadow, assesses the structural similarity of two images. It directs model training to improve the shadow while maintaining structure, producing outputs that are correct and pleasing to the eye. Figure 4.8 shows Predicted image has color imbalance red color is transformed. SSIM loss with cascade U-Net has not given competitive results.



Figure 4.8: SSIM Loss on ISRD

CHAPTER 5

Conclusion and Future Scope

This thesis proposed a cascade U-Net architecture for the shadow removal in computer vision. The architecture consists of two stages of U-Net architectures that progressively learn and refine the shadow removal results. The first stage trains a U-Net using the shadow images and ground truth image to predict shadow free images. The second stage utilizes the predicted shadow free images and ground truth image as input to refine the results further.

The experimental evaluations on benchmark datasets demonstrate the effectiveness of our approach, outperforming state-of-the-art methods in both qualitative and quantitative evaluations. The evaluation metrics such as structural similarity index (SSIM) and Root Mean Square Error, along with subjective evaluations by human observers, confirm the superiority of our method. The proposed cascade U-Net architecture offers a promising solution for enhancing image quality and interpretability by removing the shadow. It leverages both initial predictions and ground truth images, enabling progressive improvement in the shadow removal.

Future Scope: This research opens avenues for further refinements and optimizations in the shadow removal techniques to advance image analysis in various computer vision applications. Future research can explore further improvements and extensions of the cascade U-Net architecture for other related tasks in computer vision. The scope of this study focuses on the development and evaluation of a cascade U-Net architecture for the shadow removal in computer vision. The study aims to enhance the interpretability and visual quality of images by effectively removing the shadow. The proposed approach shows promising results in generating the shadow free images, but it is acknowledged that there is an issue with the shadow edges, where Shadow are still visible and need to reduce RMSE for predicted Shadow Free Image.

References

- [1] R. Abiko and M. Ikehara. Channel attention gan trained with enhanced dataset for single-image shadow removal. *IEEE Access*, 10:12322–12333, 2022.
- [2] N. Bansal, Akashdeep, and N. Aggarwal. Deep learning based shadow detection in images. In *Proceedings of 2nd International Conference on Communication, Computing and Networking: ICCCN 2018, NITTTR Chandigarh, India*, pages 375–382. Springer, 2019.
- [3] H. Fan, M. Han, and J. Li. Image shadow removal using end-to-end deep convolutional neural networks. *Applied Sciences*, 9(5):1009, 2019.
- [4] 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*, pages 10571–10580, 2021.
- [5] H. Gong and D. Cosker. Interactive shadow removal and ground truth for variable scene categories. In *BMVC*, pages 1–11, 2014.
- [6] R. Guo, Q. Dai, and D. Hoiem. Paired regions for shadow detection and removal. *IEEE transactions on pattern analysis and machine intelligence*, 35(12):2956–2967, 2012.
- [7] 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*, pages 2472–2481, 2019.
- [8] X. Hu, T. Wang, C.-W. Fu, Y. Jiang, Q. Wang, and P.-A. Heng. Revisiting shadow detection: A new benchmark dataset for complex world. *IEEE Transactions on Image Processing*, 30:1925–1934, 2021.
- [9] X. Hu, L. Zhu, C.-W. Fu, J. Qin, and P.-A. Heng. Direction-aware spatial context features for shadow detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7454–7462, 2018.

- [10] C. Jiang and M. Ward. Shadow identification. In *Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 606–612, 1992.
- [11] M. Khare, R. K. Srivastava, and A. Khare. Moving shadow detection and removal—a wavelet transform based approach. *IET Computer Vision*, 8(6):701–717, 2014.
- [12] H. Le and D. Samaras. Physics-based shadow image decomposition for shadow removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):9088–9101, 2021.
- [13] T. Nagae, R. Abiko, T. Yamaguchi, and M. Ikehara. Shadow detection and removal using gan. In *2020 28th European Signal Processing Conference (EUSIPCO)*, pages 630–634. IEEE, 2021.
- [14] V. Nguyen, T. F. Yago Vicente, M. Zhao, M. Hoai, and D. Samaras. Shadow detection with conditional generative adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4510–4518, 2017.
- [15] 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.
- [16] L. Qu, J. Tian, S. He, Y. Tang, and R. W. Lau. Deshadownet: A multi-context embedding deep network for shadow removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4067–4075, 2017.
- [17] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.
- [18] C. Tan and X. Feng. Unsupervised shadow removal using target consistency generative adversarial network. *arXiv preprint arXiv:2010.01291*, 2020.
- [19] 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 *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI 14*, pages 816–832. Springer, 2016.

- [20] 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*, pages 1788–1797, 2018.
- [21] C. Wyman and C. D. Hansen. Penumbra maps: Approximate soft shadows in real-time. In *Rendering techniques*, pages 202–207, 2003.
- [22] Q. Yang, K.-H. Tan, and N. Ahuja. Shadow removal using bilateral filtering. *IEEE Transactions on Image processing*, 21(10):4361–4368, 2012.
- [23] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.