

ENHANCED SHADOW REMOVAL FOR SURVEILLANCE SYSTEMS.

19AD798 - DISSERTATION

Submitted by

CB.EN.P2AID19017 JISHNU P

in partial fulfillment of the requirements for the degree of

**MASTER OF TECHNOLOGY
IN
ARTIFICIAL INTELLIGENCE AND DATA SCIENCE**



**DEPARTMENT OF COMPUTER SCIENCE AND
ENGINEERING**

AMRITA SCHOOL OF ENGINEERING

AMRITA VISHWA VIDYAPEETHAM

COIMBATORE – 641 112

NOV 2020

AMRITA VISHWA VIDYAPEETHAM
AMRITA SCHOOL OF ENGINEERING, COIMBATORE
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING



BONAFIDE CERTIFICATE

This is to certify that 19AD798 – Dissertation entitled “Enhanced Shadow Removal for Surveillance Systems” submitted by JISHNU P (CB.EN.P2AID19017) in partial fulfillment of the requirements for the award of Degree of **Master of Technology in Artificial Intelligence and Data Science** is a bonafide record of the work carried out under my guidance and supervision at Amrita School of Engineering, Coimbatore.

JISHNU P

CB.EN.P2AID19017

Dr. Rajathilagam B

Thesis Advisor

Associate Professor

Department of CSE

Dr. (Col) P. N. Kumar

Chairperson

Department of CSE

Evaluated on:

INTERNAL EXAMINER

EXTERNAL EXAMINER

ABSTRACT

The presence of shadow is unavoidable while dealing with outdoor images in a variety of computer vision applications. Shadows always trouble computer vision tasks like visual navigation, object detection and tracking. In order to unveil the information occluded by shadow, it is essential to remove the shadow. This is a two-step process which involves shadow detection and shadow removal. However, the field of shadow detection greatly improved compared to the shadow removal task, but the field of shadow removal has been difficult because it needs to be restored after removing the shadow. In this paper, shadow-less image is generated using a modified conditional GAN (cGAN) model and using shadow image and the original image as the inputs. We propose a novel method using a discriminator that judges the local patches of the images. We not only use the residual generator to produce high-quality images, but we also use combined loss, which is the weighted sum of reconstruction loss and GAN loss for training stability. With the proposed model, we achieved significant improvements in the shadow removal task compared to the state of the art models. To verify the strength of our model, evaluation metrics such as Root Mean Squared Error(RMSE), Peak Signal to Noise ratio(PSNR) calculated. Higher PSNR and lower RMSE value of our model ensures the quality of the final shadow-less image generated.

Keywords : Shadow . GAN . Combined Loss

ACKNOWLEDGEMENT

I wish to record my deep sense of gratitude and profound thanks to my research supervisor Dr. Rajathilagam B, Associate Professor, Department of Computer Science and Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, for his keen interest, inspiring guidance, constant encouragement with my work during all stages, to bring this thesis into fruition.

I am extremely indebted to **Dr. (Col) P. N. Kumar**, Chairperson, Department of CSE, Amrita Vishwa Vidyapeetham, Coimbatore and panel members for their valuable suggestions and support during the course of my research work.

I also thank the faculty and non-teaching staff members of the Department of CSE, Amrita Vishwa Vidyapeetham, Coimbatore, for their valuable support throughout the course of my research work.

JISHNU P

TABLE OF CONTENTS

CHAPTER	TITLE	PAGE NO
	ABSTRACT	3
	LIST OF TABLES	7
	LIST OF FIGURES	8
	LIST OF ABBREVIATIONS	9
1	INTRODUCTION	10
1.1	Problem Definition	10
1.2	Motivation	11
1.3	Challenges	12
2	LITERATURE SURVEY	13
2.1	Survey	13
2.2	Findings	15
2.3	Objective of the work	16
3	PROPOSED WORK	17
3.1	Architecture of the System	17
3.1.1	Architecture of Generator	18
3.1.2	Architecture of Discriminator	19
3.2	Dataset for the Study and Platform	21
3.2.1	Dataset Information	21
3.2.2	Platform and Specification	22
4	RESULTS AND INFERENCES	23
4.1	Metrics for evaluation	23

4.2	Parameters settings	23
4.3	Results and discussion	24
5	CONCLUSION AND FUTURE ENHANCEMENTS	27
6	REFERENCES	28

LIST OF TABLES

TABLE NO	TITLE	PAGE NO
1	Comparing the models	25

LIST OF FIGURES

FIGURE NO	TITLE	PAGE NO.
1	Umbra and Penumbra Regions	10
2	Example: Shadow Removal	11
3	Architecture design of the System	17
4	U-Net Architecture of Generator	18
5	Architecture of local patch Discriminator	19
6	Generator and discriminator loss curve	25
7	Sample output	25
8	Output of proposed model on samples outside the ISTD dataset	26

LIST OF ABBREVIATIONS

Acronym	Full Name
GAN	Generative Adversarial Network
cGAN	Conditional Generative Adversarial Network
RMSE	Root Mean Squared Error
PSNR	Peak Signal to Noise Ratio
SRD	Shadow removal Dataset
DCNN	Deep Convolutional Neural networks
SVM	Support vector machine
UCF	University of Central Florida.

CHAPTER 1

INTRODUCTION

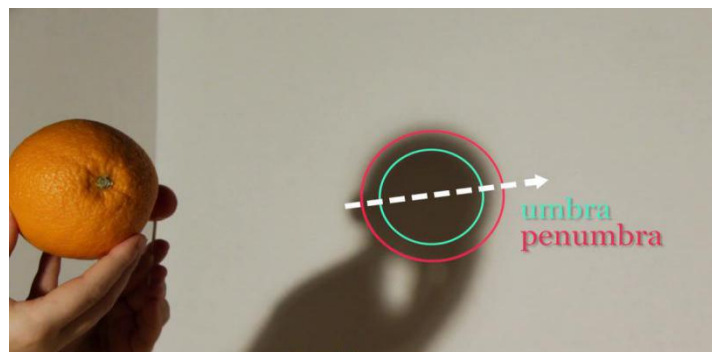
1.1 PROBLEM DEFINITION

Removing shadows from the images has been considered as a challenging task in the field of computer vision. The presence of opaque objects in the path of sunlight leads to the formation of shadows and depend on different factors such as the altitude of the sun and location of the object. For example, consider a bike and bus in traffic such that the bike is standing left side of the bus. There are chances for the shadow of the bus to cover the bike if the sunlight is from the right side of the bus. Shadow of different shapes distorts two different objects into a single object. This is called as the occlusion. This is a difficult situation in which we can't efficiently detect different objects. In this example, it will be difficult for us to distinguish between bike and bus. Probably the Bus and its shadow will merge together and form another shape which will be far different from the shape of a bus.

In traditional approaches, a common method to remove shadow consist of detecting shadows and using detected shadow masks as a clue for removing the shadows.

The field of shadow detection predicts the location of the shadowed region in the image and separates the shadowed and non-shadowed region of the original image in pixels.

This has been considered a challenging task to classify shadows in an image because shadows have various properties. Depending on the degrees of occlusion by the object, the brightness of shadow varies such as umbra and penumbra. The dark part of the shadow is called the umbra, and the part of a shadow that's a little lighter is called the penumbra.



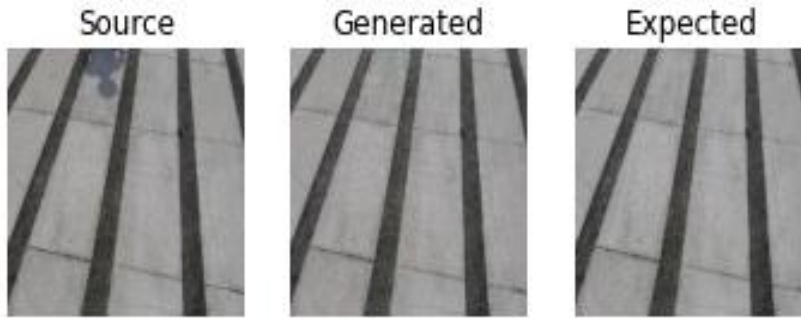
(Figure 1) Umbra and Penumbra regions

source : [<https://www.youtube.com/watch?v=huVCQXYy4jc>]

if there is an object with a black texture in the image, it can be misclassified as a shadow region.

Shadow removal is also a difficult task because we have to remove shadows and restore the information in that region according to the degree of occlusion.

In this paper we remove shadow based on GAN which generates realistic image.



(Figure 2) Example : Shadow removal (our model)

Our objective is to propose GAN based Shadow removal architecture which generates higher quality shadow-less images for the given shadowed image.

1.2 MOTIVATION

After the introduction of Generative Adversarial Networks (GAN) in 2014[7], the computer vision domain has taken leap at various tasks. Shadow removal is an important task which can be considered as an important preprocessing step for higher level Computer Vision tasks in surveillance systems.

The shadow detection phase has greatly improved, but the field of shadow removal has been difficult because it needs to be restored after removing the shadow. To the extent to my knowledge, the shadow removal part is open to further improvements even though there are papers which employ GANs for the same.

1.3 CHALLENGES

- Ensuring the higher quality in shadow-less image by using efficient evaluation metric and enhanced architecture.
- Addressing well known issues of GAN such as Vanishing Gradient, non-Convergence, Mode collapse by balancing the Generator and Discriminator training to avoid overfitting.

CHAPTER 2

LITERATURE SURVEY

2.1 SURVEY

Shadow detection and removal has been studied across a number of papers throughout the years. Shadow removal process extensively used for removing shadows from the images to make it suitable for higher level computer vision applications such as object detection, tracking. Shadow removal considered as the complex process compared to the shadow detection phase due to the difficulty while reconstructing the pixels in the detected shadow region. Jifeng Wang[1] introduced the ISTD dataset as part of the work titled “Stacked Conditional Generative Adversarial Networks for Jointly Learning Shadow Detection and Shadow Removal” and which considered as one of the benchmark dataset for the shadow removal process. From this view it is clear that the research in the shadow detection domain almost saturated and the focus is on the enhancements in the shadow removal phase. They proposed an architecture which contains two conditional GAN(cGAN) stacked together and performs shadow detection and removal tasks simultaneously. With this model RMSE value of 7.47 achieved on the ISTD dataset. Lack of considering the context information in the shadow removal phase was the drawback of this model.

In order to consider more details of the shadowed image like the illumination information, Ling Zhang[2] proposed a GAN based model which contains 4 generators and 3 discriminators in the work entitled “RIS-GAN: Explore Residual and Illumination with Generative Adversarial Networks for Shadow Removal”. This model achieved an RMSE value of 6.97 which is better than the model proposed by Jifeng Wang[1]. The model trained on two different benchmark datasets for shadow removal which are SRD and ISTD datasets. RMSE value of 6.78 achieved for the SRD dataset. Lower the RMSE value means better the quality of the generated shadow-less image.

The context information was still missing in these models. The model proposed by L. Qu[3] use the multi-context information for removing the shadow and reconstructing the shadowed region. Their work titled as “DeshadowNet: A Multi-context Embedding Deep Network for Shadow Removal ” uses only deep convolutional neural networks(DCNN) for shadow detection and shadow removal. Acquisition of multi-context information achieved by training three different networks named G-Net(Global Net), A-Net(Appearance net), S-Net(Semantic Net). Global context, appearance information and semantic context of the shadowed region captured by these networks. All these three networks together called as the DeshadowNet. The RMSE value scored on the SRD dataset is 6.64. Blurred effect was present in the shadowed region of the output images.

Generative Adversarial Network(GAN) base models produce high quality output images compared to the traditional model. Sidorov[4] proposed AngularGAN architecture to improve the color constancy on the images. Their work entitled “Conditional GANs for Multi-Illuminant Color Constancy: Revolution or Yet Another Approach?” used the GAN based model along with the angular information of the incident light on the object. The model trained on the synthetic dataset called GTAV dataset. Peak Signal to Noise Ratio(PSNR) used to evaluate the performance of the model. PSNR value for the GTAV dataset was 21dB. "Shadow Detection and Removal for Illumination Consistency on the Road" by Wang[5] showed promising results for shadow removal using non-linear SVM classifier. Proposed model was exclusively for traffic images and the model was not good for the large sized shadows in the scene. The model performs better for small sized shadows in the traffic images. Accuracy is used as the performance evaluation measure and achieved an accuracy of 83% on UCF dataset. They improved the accuracy of the previous work entitled "Shadow detection and removal for illumination consistency on the road"[9] by introducing adaptive variable scale regional compensation operator to remove the shadows.

Introduction of Generative Adversarial Networks[7] made drastic changes in the computer vision researches. Already implemented models migrated to GAN architectures as extension to their previous works to improve the results. Yun [6] Introduced GAN for shadow removal as an extension to the previous work entitled “Shadow Detection and

Removal From Photo-Realistic Synthetic Urban Image Using Deep Learning” and improved the performance. However, traditional modeled methods have limited ability to remove shadows when irregular illumination or objects with various colors are present.

In order to address the color inconsistencies in the shadowed region, Xiaodong Cun[11] proposed a novel network structure called dual hierarchically aggregation network (DHAN) which contains a series of convolutions as the backbone without any down-sampling and hierarchically aggregate multi-context features for attention and prediction for successful shadow removal. The model proposed in the work entitled “ Towards Ghost-free Shadow Removal via Dual Hierarchical Aggregation Network and Shadow Matting GAN” shows RMSE = 5.76 , PSNR = 34.48 dB for ISTD dataset and RMSE = 5.46, PSNR = 33.72 for SRD dataset respectively.

2.2 FINDINGS

The shadow removal phase is open to enhancements and not yet saturated. GAN based models show significant improvements in generating shadow-less images. Conditional Generative Adversarial Networks(cGAN) are very helpful for narrowing down the generated image space of the generator and thereby reducing the time for training the model. We can modify GAN architecture according to our needs and based on the inputs to the generator and the discriminator, the behaviour of the GAN changes significantly and produces good results.

Performance of the shadow removal models evaluated using the metrics RMSE, PSNR in most of the cases. Quantitative comparison between the pixels of the shadow-less image and the model generated image performed for evaluating the performance of the model.

2.3 OBJECTIVE OF THE WORK

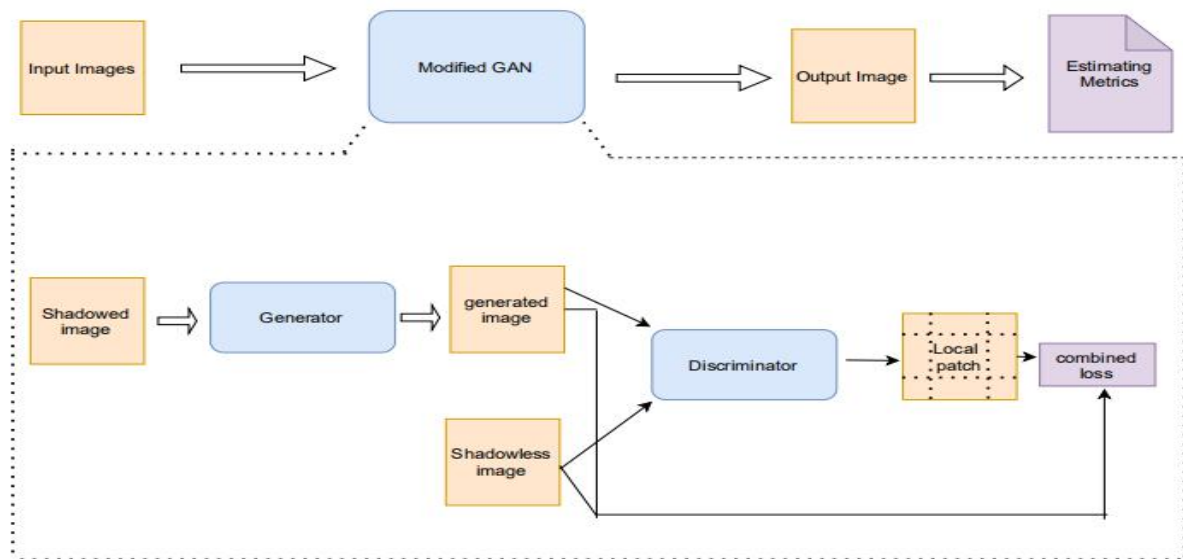
To propose GAN based Shadow removal architecture which generates higher quality shadow-less images for the given shadowed images.

CHAPTER 3

PROPOSED WORK

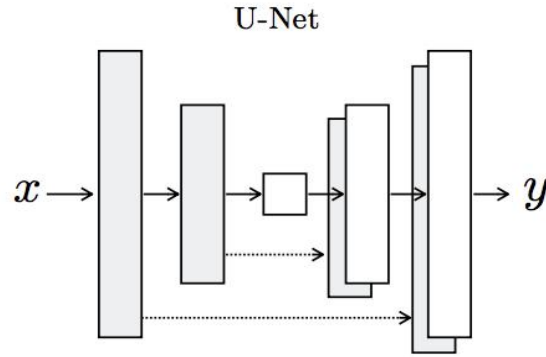
3.1 ARCHITECTURE OF THE SYSTEM

The proposed architecture for the shadow removal task is given in the figure below (figure 5). The generator model is provided with shadowed image as input and generates a translated version(shadow-less) of the image. The discriminator model is given an input image(shadow_less image) and a real or generated paired image and must determine whether the paired image is real or fake. Finally, the generator model is trained to both fool the discriminator model and to minimize the loss between the generated image and the expected target image.



(Figure 3) : Architecture Design of the system

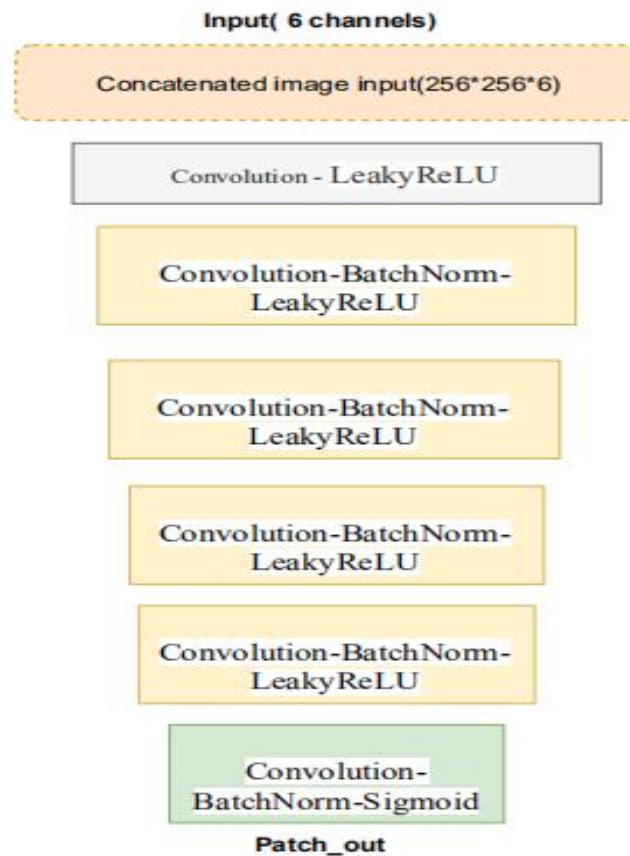
3.1.1 ARCHITECTURE OF GENERATOR



(figure 4) U-Net architecture of Generator.

Figure4 Shows the overall abstract architecture of the Generator. The generator is an encoder-decoder model using a U-Net architecture[12]. The encoder and decoder of the generator contain convolutional, batch normalization, dropout, and activation layers. The encoder encodes the information in the given input image and the context information generated in the bottleneck is used for reconstructing the image using the decoder block. Skip connections are used between corresponding encoder and decoder layers for improving the quality of the image generated by the generator.

3.1.2 ARCHITECTURE OF DISCRIMINATOR



(figure 5) Architecture of local patch Discriminator

Figure5 shows the architecture of local patch discriminator. The discriminator is a deep convolutional neural network that performs image classification. It takes both the source image (generated image) and the target image (shadow-less image) as input and predicts the likelihood of whether the shadow-less image is real or a fake translation of the generated image. The discriminator model is trained in the same way as a traditional GAN model. Adversarial training is used for training the discriminator model.

The training of discriminator is too fast compared to the generator. This will lead to the issues in the GAN such as Vanishing gradient, Mode collapse, Non - convergence.

Vanishing gradient :- If the discriminator is too good, then the the gradient becomes so small that the initial layers (generator) learn very slowly or stop learning completely.

Mode collapse :- The generator collapses which produces limited varieties of samples.

The obvious solution is to balance generator & discriminator training to avoid overfitting.

Combined Loss introduced to balance the training of Generator and Discriminator. Giving more importance to the reconstruction loss rather than the adversarial loss during the Generator training will reduce the effect of the fast training of Discriminator on the Generator.

Combined loss = λ_1 * reconstruction loss + λ_2 * adversarial loss

Where,

$$\lambda_1, \lambda_2 = 100, 1$$

The reconstruction loss is the loss function between the generated fake image and the shadow-less image.

$$L_{reconstruction} = \sum_{i=1}^n |Y_{true} - Y_{pred}|$$

Where,

Y_{pred} = The predicted value for the i th pixel.

Y_{true} = The observed(actual) value for the i th pixel

N = Total number of pixels.

Adversarial loss (Binary Cross Entropy loss)

$$L_{adversarial}(G,D) = E_{x,y}[\log D(x,y)] + E_{x,y}[\log(1 - D(x, G(x,y)))]$$

Where,

x - shadow-less image

y - shadowed image

3.2 DATASET FOR THE STUDY AND PLATFORM

3.2.1 DATASET INFORMATION

Since we are more focused on the shadow removal task, ISTD[1] shadow removal bench mark dataset is used. Dataset contains shadowed images, shadow mask, shadow-less images.

Train data - 1330 images (640 * 480 pixels)

Test data - 540 images(640 * 480 pixels)

During the data preprocessing phase, the images are loaded and re-scaled to (256*256) pixels for processing convenience and converted to numpy array and saved as compressed format (Example : ISTD(all)_256.npz)

3.2.2 PLATFORM AND SPECIFICATIONS

The model trained on the computing node with the following specifications

- RAM: 256Gb
- Processor: Intel(R) Xeon(R) Gold 5120 CPU @ 2.20GHz
- No of cores used: 4
- GPU: NVIDIA V100

CHAPTER 4

RESULTS AND INFERENCES

4.1 METRICS FOR EVALUATION

To measure the performance of the model, the selection of the proper evaluation metrics plays a major role. Most of the Baseline models use various metrics such as RMSE, PSNR. For efficient comparison with baseline models, we will be evaluating our model with RMSE, PSNR.

- **RMSE(Root Mean Square Error)**

The Mean Square Error (MSE) is the mean squared error between the generated shadow-less image and the original image.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_{true} - Y_{pred})^2}{n}}$$

- **PSNR(peak signal-to-noise ratio):**

This ratio is used as a quality measurement between the original and a compressed image. The higher the PSNR, better the quality of the reconstructed image. MSE - Mean square Error

$$PSNR = 10 * \log \frac{255 * 255}{MSE}$$

4.2 PARAMETERS SETTING

Model trained on the preprocessed ISTD dataset. Following parameters set in the proposed model.

- **Loss function:** $\lambda_2 * \text{Binary Cross Entropy} + \lambda_1 * \text{Mean Absolute Error(MAE)}$

- **Optimizer:** ADAM
- **Convolutional Layers:** Strides = (2,2), kernel size = (4,4)
- **Activation:**

DISCRIMINATOR

- **Internal Layers:** LeakyReLU, slope = 0.2
- **Output Layer:** Sigmoid

GENERATOR

- **Internal Layers :** relu
- **Output layers :** tanh
- **Decoder layers :** Dropout(0.5)

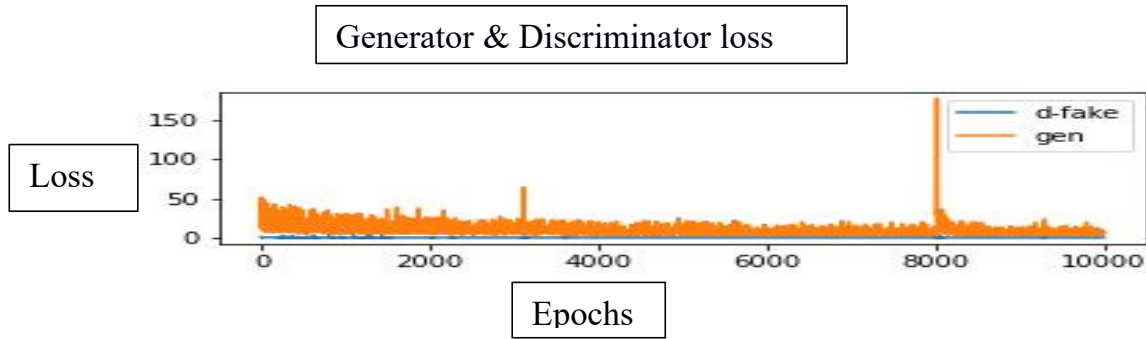
- **Learning rate:** 0.0002
- **Batch size:** 1
- **Epochs:** 100

4.3 RESULTS AND DISCUSSION

In this project, ISTD dataset is used to train our model and the metrics RMSE,PSNR are calculated. The state of the art models RIS-GAN[2], Stacked Conditional GAN(ST-GAN)[1] used as the base models for comparing the performance of our model(Table 1). The training graph based on the generator loss and discriminator loss is shown in figure 6.

Table 1 : Comparing the models

Model	RMSE	PSNR(dB)
ST-GAN[1]	7.47	-----
RIS-GAN[2]	6.67	-----
Our Model	1.997	53.38



(figure 6): Generator and Discriminator loss curve

From the loss curve(figure6), it is clear that the generator and discriminator training is like a min-max game. The generator tries to fool the discriminator and at the same time, the discriminator try not to fool by the generator's fake image. The loss never become same for the generator and discriminator for a good GAN[7] model.

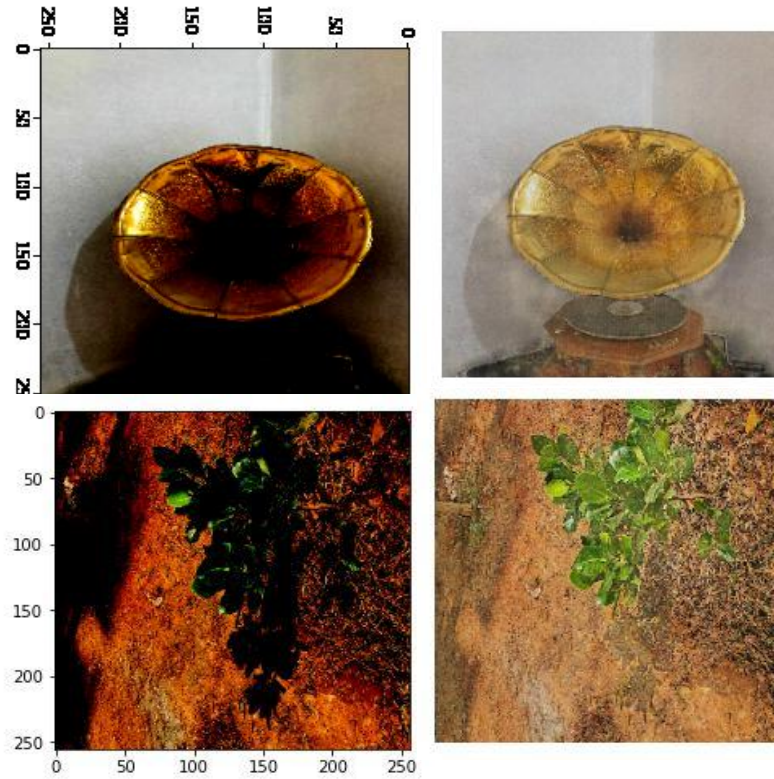
Sample results of our model shown in (figure 2). It is clear that the model performs well for the testing data in ISTD images.



(Figure 7) sample output. source, generated, expected images are arranged row wise.

In the above figure(figure 7), we can see that our model performs well for the indoor images and the outdoor images. First and third column corresponds to the outdoor images. RMSE = 0.045 and PSNR = 78.76dB for these two images. Second column

corresponds to the indoor image. It is an image of a black board in a classroom. The shadow successfully removed by our model. RMSE = 0.08 and PSNR = 74.58db for this particular indoor shadowed image sample.



(Figure 8) : output of the proposed model on samples outside the ISTD dataset.

First row corresponds to an indoor shadowed image(Figure 8) and corresponding output of our model. Second row corresponds to a outdoor shadowed image from a real life instance and corresponding shadow-less image produced by our model.

CHAPTER 5

CONCLUSION AND FUTURE ENHANCEMENTS

In this project, we proposed a GAN based shadow removal model for generating enhanced shadow-less images. Initially, we use basic conditional GAN model on ISTD dataset and analyzed the areas of improvements. Secondly, we modified the architecture and introduced a combined loss for training the model. Tuned the parameters by conducting repeated experiments and identified the appropriate set of parameters for the model. From the experiments, it was seen that the proposed model performs better than the existing model. The model showed promising results on the shadowed(outdoor and indoor) real-time images collected by us.

As a future enhancement, shadow-less image generated by the proposed model can be improved further by using image super-resolution techniques. Parameter tuning is a time-consuming task and also requires more domain knowledge. In addition to future enhancement, Neuroevolution techniques can be used to tune the hyperparameters to identify the best set of parameters for the model and thereby improving the performance of our model.

CHAPTER 6

REFERENCES

1. [1]Jifeng Wang, Xiang Li², Jian Yang, “Stacked Conditional Generative Adversarial Networks for Jointly Learning Shadow Detection and Shadow Removal”, CVPR 2017,<https://doi.ieeecomputersociety.org/10.1109/CVPR.2018.00192>
2. [2]Ling Zhang,¹ Chengjiang Long,^{2*} Xiaolong Zhang,¹ Chunxia Xiao³, “RIS-GAN: Explore Residual and Illumination with Generative Adversarial Networks for Shadow Removal”, AAAI 2020,<https://doi.org/10.1609/aaai.v34i07.6979>
3. [3]L. Qu, J. Tian, S. He, Y. Tang and R. W. H. Lau, "DeshadowNet: A Multi-context Embedding Deep Network for Shadow Removal," IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017,<https://ieeexplore.ieee.org/document/8099731>
4. [4]Sidorov, Oleksii, “Conditional GANs for Multi-Illuminant Color Constancy: Revolution or Yet Another Approach?”,The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2018,<https://arxiv.org/abs/1811.06604>
5. [5]Wang, H. Xu, Z. Zhou, L. Deng and M. Yang, "Shadow Detection and Removal for Illumination Consistency on the Road," in IEEE Transactions on Intelligent Vehicles, 2020, <https://ieeexplore.ieee.org/document/9068460>
6. [6]Yun, Heejin & Kang Jik, Kim & Chun, Jun-Chul. “Shadow Detection and Removal From Photo-Realistic Synthetic Urban Image Using Deep Learning”, Computers, Materials & Continua 2019, <https://www.techscience.com/cmc/v62n1/38123>
7. [7]Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, “Generative Adversarial Networks”, International Conference on Neural Information Processing Systems, 2014, <https://arxiv.org/abs/1406.2661>
8. [8]M. Mirza, S. Osindero, "Conditional Generative Adversarial Nets", In Arxiv 2014. <http://arxiv.org/abs/1411.1784>
9. [9]Wang, L. Deng, Z. Zhou, M. Yang and B. Wang, "Shadow detection and removal for illumination consistency on the road,"<https://ieeexplore.ieee.org/document/8304275>

10. [10]M. Wang, Y. C. Chung, C. L. Chang and S. W. Chen, "Shadow detection and removal for traffic images," IEEE International Conference on Networking, Sensing and Control, 2004, <https://ieeexplore.ieee.org/document/1297516>
11. [11]Xiaodong Cun, Chi-Man Pun, Cheng Shi, " Towards Ghost-free Shadow Removal via Dual Hierarchical Aggregation Network and Shadow Matting GAN" AAAI 2020, <https://arxiv.org/abs/1911.08718>
12. [12]Olaf Ronneberger, Philipp Fischer, Thomas Brox,"U-Net: Convolutional Networks for Biomedical Image Segmentation"MICCAI 2015, <https://arxiv.org/abs/1505.04597>