"Dark arts" : Algorithms for enhancement and interpretation of low light images

- Semester 8 Report -

Gihan Jayatilaka
Suren Sritharan
Harshana Weligampola

Department of Computer Engineering
University of Peradeniya

Final Year Project (courses CO421 & CO425) report submitted as a requirement of the degree of
B.Sc.Eng. in Computer Engineering

August 2020

Supervisors: Prof. Roshan Ragel, Dr. Roshan Godaliyadda, Dr. Vijitha Herath, and Dr. Parakrama Ekanayake

Declaration

I/We hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is our own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgments.

Gihan Jayatilaka
Suren Sritharan
Harshana Weligampola
August 2020

Abstract

Images taken in low light condition suffer from poor aesthetics and result in degraded performance for vision-based algorithms which are primarily designed for high-quality images. Therefore, Low light image enhancement is an important task in making computer vision algorithms work robustly under different environmental conditions. Through our studies, we have show that classical algorithms and supervised learning solutions have certain shortcomings in different environments. We analyze the reasons for these and shortcomings and propose a hybrid solution for this problem. Furthermore, we introduce an algorithmic testbed under which these algorithms can be studied in detail and we propose an evaluation metric which aims to enhance the visual aesthetics of the image. We also explore the possibility to detect a class of images in low light condition after learning them in well light conditions. In addition to these specific cases, the work reveals insights to the properties of digital images (and algorithmic interpretation of them) in different lighting conditions. On conclusion, this study suggests possible future research avenues on all these fronts.

Table of contents

List of figures vii

List of tables ix

Nomenclature x

1 Introduction 1
- 1.1 Background . . . 1
- 1.2 Research problem . . . 2
  - 1.2.1 Alternative solution . . . 3
- 1.3 Datasets . . . 3
  - 1.3.1 Paired . . . 4
  - 1.3.2 Unpaired . . . 4
- 1.4 Case studies . . . 4
  - 1.4.1 Low-light image enhancement . . . 4
  - 1.4.2 Controlled illumination enhancement . . . 4
  - 1.4.3 Object Detection in Low Light Conditions . . . 5

2 Low-light Image enhancement 6
- 2.1 Introduction . . . 6
- 2.2 Related works . . . 7
- 2.3 Proposed Method . . . 9
  - 2.3.1 Retinex decomposition network . . . 9
  - 2.3.2 Enhancement network . . . 10
  - 2.3.3 Patched Discriminator . . . 13
  - 2.3.4 Summary . . . 14
- 2.4 Results and discussion . . . 14
  - 2.4.1 Performance metrics . . . 14
  - 2.4.2 Retinex decomposition . . . 15

2.4.3 GAN based illumination enhancement . . . 15
2.4.4 Complete low-light image enhancement system . . . 16
2.4.5 Study of the inner-workings of the model. . . . 17
2.5 Conclusion . . . 17
2.6 Future work . . . 18

3 Controlled illumination enhancement 21
3.1 Related work . . . 22
3.2 Dataset . . . 23
3.3 Proposed Methodology . . . 24
3.3.1 Image enhancement under varying lightning . . . 25
3.3.2 Controlled illumination enhancement model . . . 28
3.4 Results and Analysis . . . 33
3.4.1 Comparative performance of two-class image enhancement model and the proposed image enhancement model . . . 33
3.4.2 Multi-class image generation performance . . . 34
3.5 Conclusion . . . 35
3.6 Future works . . . 36

4 Object Detection in Low Light Conditions 38
4.1 Related work . . . 38
4.1.1 Object detection . . . 38
4.1.2 Lighting condition enhancement . . . 39
4.2 One Class Problem . . . 39
4.2.1 OCGAN . . . 40
4.2.2 Retinex decomp . . . 41
4.3 One class novelty detection in low light images. . . . 41
4.3.1 Problem definition . . . 41
4.3.2 Proposed solution . . . 41
4.4 Experiments . . . 46
4.4.1 Implementation . . . 46
4.4.2 Tests . . . 46
4.5 Results . . . 46
4.6 Discussion of results . . . 47
4.7 Conclusion and Future Work . . . 47

5 Concluding remarks 49

References 51

List of figures

2.1 Forward cycle of the retinex-cycle-GAN model 10
2.2 Block diagram of the generator components. 12
2.3 Discriminator model 14
2.4 Results of decomposition network. Row 1 (L to R): $S_{low}, R_{low}, I_{low}$. Row 2 (L to R): $S_{high}, R_{high}, I_{high}$ 15
2.5 Comparison of the results from CycleGAN based illumination enhancement and GAN based illumination enhancement. 16
2.6 Results from the complete pipeline. Row 1: Low light input images. Row 2: Well-lit ground truth. Row 3: G^a Generic GAN enhanced image. Row 4: G^b Retinexnet enhanced image. Row 5: G^c retinex based GAN enhanced image. Row 5: G Retinex based CycleGAN (ours) enhanced image. . . . 19
2.7 Intermediate layer outputs of generator models. 20

3.1 Sample images from all classes 25
3.2 Image enhancement model. 26
3.3 Restoration-net architecture. 27
3.4 Illumination-net architecture. 29
3.5 Controlled illumination enhancement model. 29
3.6 Conditional illumination adjustment generator. 31
3.7 Conditional illumination adjustment discriminator. 31
3.8 Conditional illumination adjustment model training. 32
3.9 Comparison of the output with varying input image of varying illumination. 33
3.10 Variation of NIQE of the output with input images of varying illumination. 34
3.11 Error between ground truth and enhanced image under varying illumination. 35
3.12 Outputs of the controlled illumination enhancement model. 36
3.13 NIQE of the outputs of controlled illumination enhancement model. . . . 37

4.1 OCGAN Architecture proposed in [1] 41

4.2 One class detection neural network 42
4.3 Training one class detection neural network 44

List of tables

2.1 U-Net Specifications 11
2.2 Numerical comparision of algorithm performance (MSE - lower is better, NIQE ratio - higher is better) 17

3.1 Image synthesis parameters 24

4.1 Performance of the D^R (for dog class) with gradient descent based negative mining 47
4.2 Performance of D^R (for dog class) with random walk based negative mining 47

Nomenclature

Acronyms / Abbreviations

cGAN	Conditional Generative Adversarial Network
CIFAR	Canadian Institute For Advanced Research
CNN	Convolutional neural network
DCGAN	Deep Convolutional Generative Adversarial Network
DNN	Deep neural network
FFT	Fast Fourier transform
GAN	Generative Adversarial Network
HDR	High dynamic range
HSV	Human vision system
LLCNN	Low light convolutional neural network
LLIEnet	Low light image enhancement network
MNIST	Modified National Institute of Standards and Technology
MSR	Multi scale retinex
NN	Neural network
OCGAN	One Class Generative Adverserial Network
PDE	Partial differential equation
PSNR	Peak signal to noise ratio

QP	Quadratic programming
RGB	Red Green Blue
SSIM	Structural similarity index measure
TV	Total variation

Chapter 1

Introduction

1.1 Background

Imaging may be referring broadly to a wide range of activities. But for the scope of this study, we will be looking into imaging as the act of preserving visual information in the form of a two-dimensional picture. The science, art or the profession of acquiring images – photography was born with the film photography, in which the visual information (light pattern) was saved into a film of photosensitive material. These films could be developed back to printed images. With time, photography experienced its natural extensions – videos and photo editing. With this development, multiple types of researches were done in the field of image processing and many algorithms were developed. Furthermore, products were introduced into the market based on these researches. However, the application of a majority of these algorithms was limited to high-quality images

Low light images suffer from a range of issues such as low visibility, noise, colour distortion, etc. These issues prohibit them from being visually perceptible for the human eye and being useful for computer vision algorithms in terms of information richness. Enhancement of these low light images has been tried on hardware front (sensitive light-capturing mechanisms, longer exposure times, etc.) with varying degrees of success but at higher costs. The algorithmic approach to enhancing these images is considered to be an important research problem.

Digital imaging

Digital imaging refers to the acquisition of digital images. As opposed to analogue images, the key difference of digital images is the ability to form lossless copies due to their representation in the form of digital encoding.

Digital images

Digital images refer to any image that is saved as a matrix of numbers that are of a binary (or any other discrete base) number representation. These images can be stored in digital storage mediums and be processed by digital computers.

Digital cameras

Digital camera refers to a camera that captures photographs in digital memory. As opposed to analogue cameras, the images are stored in a digitally encoded format in a digital medium rather than through chemical techniques. Although analogue cameras exist, most of the cameras nowadays are digital including DSLR, phone, laptop, etc.

Digital Image processing

Image processing is a subcategory of signal processing. It refers to the manipulation of signals in order to improve image quality or convert to a different domain to solve a particular problem. Digital image processing refers to image processing done to digital images, usually using computer algorithms. It has many advantages over analogue image processing including efficiency, low cost, ease of implementation, etc.

Low light images

Low light images refer to images that that taken under low lighting conditions. A "good" image has the following characteristics.

- Visually pleasing
- Clear information

These conditions are preserved through contrast. The contrast is the colour difference between the objects in an image. The contrast is a function of brightness (intensity) of the colours in an image. When the brightness goes down, the contrast goes down and as a result, the images lose both the listed characteristics. In general, it is desirable to have "good images" and therefore images with normal light conditions (instead of low light condition).

1.2 Research problem

Many of the computer vision and image processing algorithms have been primarily designed for high-quality inputs. Thus when an image is taken under low lighting

conditions, apart from poor aesthetics, the poor quality may also significantly degenerate the performance of many computer vision algorithms.

1.2.1 Alternative solution

There are many alternative solutions which are being used to solve this issue. However, each of them has its own drawbacks.

External lighting

The use of a flasher is the most commonly used solution for the low light condition. This increases the illumination over the environment and therefore amplifies the brightness as a whole and the contrast.

Sophisticated hardware

Sophisticated hardware is included in devices such as night cameras. However, these are expensive and may not be suitable for different conditions.

HDR Imaging

High-dynamic-range imaging (HDRI) is an image processing technique which is used to produce images with high quality. This is usually done by combining images taken at different exposure setting within a short time period. Although it is not explicitly used to solve the problem of low light image enhancement, it is used in many cases as an alternative. However, one of the main issues is the "ghosting" effect which occurs due to any movements when taking multiple images. Therefore it is not suitable for dynamic environments.

1.3 Datasets

When classifying images into different classes, the same image taken under different lighting conditions can be labeled differently based on the illumination level. As a paired dataset refers to a dataset which contains contains the same image taken under different lighting conditions in multiple (or all) the classes. However, this is a practically challenging task and only two-class image datasets such as LOL are available in general with a few images. Since learning models require abundant data, synthetic images are required even though they present a few downsides.

On the other hand, unpaired dataset contain may not have the same image under different lighting conditions. This is comparatively easy to find, and such a dataset can be formed by combining different datasets containing low-light, well-lit and over-saturated images. However, training without paired images pose certain challenges. Therefore, unpaired images are only used in this work for testing.

1.3.1 Paired

There exist several options to collect a paired dataset of low/normal-light images, however none is efficient nor easily scalable. The images in LOL [2] dataset are captured in the daytime under normal light condition by changing the exposure time and ISO. The SID [3] dataset is generated by increasing the exposure time to generate well-lit images. However, it consists of raw sensor data under extremely low-light scenes, and this may limit its application for general low-light enhancement researches. Furthermore, due to the tedious experimental procedure, these datasets contain only a few images.

1.3.2 Unpaired

Unpaired datasets are easy to collect and they have abundant images. For example in the Exclusively Dark (ExDARK) dataset [4] contains a collection of 7,363 low-light images from very low-light environments to twilight (i.e 10 different conditions) with 12 object classes.

1.4 Case studies

We study the problem of image enhancement under 3 case studies. Although the problems are interconnected, they have been addressed them separately for clarity.

1.4.1 Low-light image enhancement

An algorithmic approach to low light enhancement is proposed in this work. We are more focused on utilizing unpaired datasets to train the model. Thus, leveraging the advantage of using large datasets for training.

1.4.2 Controlled illumination enhancement

While we study image enhancement as a two-class problem in the previous study, realistically the illumination may vary a lot. Our results show that the proposed models

suffer performance degradation in such varying lightning condition. Thus in addition to finding ways to improve the performance of the proposed model, we also analyze the ability to produce images of varying illumination levels in this case study.

1.4.3 Object Detection in Low Light Conditions

We develop algorithms to detect objects in low light conditions in this section. We try to learn from datasets in well light conditions and perform in low light conditions. The main focus of this study is to predict out of class images (during low light conditions) given a training set of a class of images (well light).

Chapter 2

Low-light Image enhancement

2.1 Introduction

Algorithmic approaches to low light image enhancement are built upon the understanding of photonics, digital image properties, and biology related to human vision. Retinex theory of human vision [5] has been the defacto inspiration for most of the algorithmic pipelines in this domain. The algorithms can be broadly broken down into non-trainable algorithms which have a fixed activity based on programmer's understanding of the process, trainable (learning) algorithms which learn the low light image enhancement from example images, and hybrids of both these approaches (which are most successful). When it comes to learning algorithms (or pipelines utilizing trainable sections) we see systems built with trainable filters on standard CNNs [6] and GANs [7].

These learning algorithms learn from datasets of two forms – paired (containing a low light and well-lit captures of the same scene/object) and unpaired (containing sets of low light and high light images without a counterpart). The paired datasets are more informative but are difficult to obtain. They are relatively rare in number and the variety of scenes in them. In contrast, unpaired datasets are not as informative as the low light datasets, are easier to obtain, abundant, and contain a wide variety of scenes/objects. Existing solutions depend on only one of these two types.

We introduced a DNN based algorithmic pipeline to enhance low light images with the following advantages.

- The system utilizing both paired and unpaired datasets. Standard CNNs are used to learn from the paired datasets while GANs are used to learn from unpaired datasets.

- The CNN and GAN architectures and their coordination are designed with the retinex model as an inspiration.

A brief summary of the proposed low-light image enhancement procedure is discussed in this chapter.

2.2 Related works

Classical Algorithms

Classical image processing algorithms are unsupervised algorithms that enhance low light images through well-formulated mathematical models. These algorithms are based on two schools of thought, namely intensity-based enhancement[8] and gradient-based enhancement[9]. They are computationally efficient and simple. But they are not robust enough to be used over different conditions without manual calibration.

Retinex theory

The Retinex Theory [5] is a biologically-motivated theory based on the colour constancy property of the human visual system (HVS). It states that any object has a lighting independent property known as the reflectance and the perceived colour of the object occurs due to the illumination on these objects. Thus, the retinex based algorithms focus on the decomposition of the image to obtain the reflectance which represents the "true colour" of the object. Many variations of the original retinex model have been proposed.[5, 10, 11] Recent works have made improvements to the retinex model for better performance and improved robustness [12] while other techniques such as LIME [13] employs a similar idea for the decomposition and enhancement of illumination maps. The model proposed in this work is also motivated by the retinex model.

Deep Convolution Networks

Deep learning techniques such as convolution neural networks (CNNs) [14] and autoencoder networks [15] have been proved to perform better than classical algorithms in many image processing tasks [16, 17]. In the field of low-light image enhancement, neural networks have been proposed in works such as LLNet [18] (based on stacked auto-encoder), LLCNN [19] (based on the retinex theory), MSRnet [20] (based on log transforms) and [21] (based on wavelet transform). Recently, the Retinex-net [2] and [3] have been proposed. These are full frameworks for low light image enhancement,

which brings impressive results. However, these techniques have certain shortcomings, the major one being that these architectures depend on training data in the form of paired datasets, which is difficult to obtain.

Adversarial Networks

Generative Adversarial Network [7] have proven to perform sufficiently well for many supervised and unsupervised learning problems. In [22] the authors propose a model through which the need for paired images has been elevated and image translation between two domains can be done through cycle-consistence loss. These techniques have been applied to many other applications including dehazing, super-resolution, etc. Lately, it has been applied to low light image enhancement in EnlightenGAN [23] with promising results and this has motivated our GAN model.

Dataset

The main requirement for many learning models is the existence of paired training data with low/normal-light images and several methods exist to collect such data. To train the retinex decomposition network we use the LOL dataset. For other purposes in this chapter we use unpaired datasets.

Supervised learning models depend on the availability of a large volume of training data but, the aforementioned techniques are neither efficient nor scalable. Thus synthetic image datasets are used as an alternative in many works [24, 25]. However, synthetic low light images are created by processing images taken under normal light conditions, and these algorithms may not consider other factors. This results in performance variation on low-light images taken in real-world conditions and therefore may not be suitable for non-synthetic dark images.

The introduction of learning models such as cyclic-GANs has removed the dependency on paired datasets. Therefore, low light images present in [4, 2] [1] can be mixed with high exposure images to create a dataset of unpaired low/normal light image.

Images generated may not reflect real-world dark images. However, these techniques have their own shortcomings, especially since the generated images do not represent the characteristics of low-light images, the performance of trained model under real-world conditions becomes unclear.

[1] LOL contains paired images, but images can be sampled without the need to keep any pair

2.3 Proposed Method

The proposed model splits the enhancement processes into multiple stages and each stage should fulfil one of the objectives specified below.

- The image's light level should be measured.
- The image's information should be extracted even in the poorly-light condition.
- The image's light should be increased while preserving and enhancing the information.
- The noise and deformations introduced to the image during the enhancement process should be taken care of.

Our work proposes a pipeline to achieve all those objectives through a series of neural networks. The first two objectives are achieved through the **decomp-net** (motivated by the retinex theory) which decomposes the images into the reflectance (colour information) and illumination (lighting information). The next two objectives are achieved through the **enhancement-net**. The first part of the process focuses on local neighbourhood of pixels while the second part focuses on both the local and global neighbourhoods.

The decomposition requires an understanding of how the images change under different lighting conditions. This relationship could be learned effectively using paired images (same scene captured under well-lit and low light condition). The enhancement requires an understanding of different types of scenes and objects in images. It is difficult to obtain a paired dataset for this task. Therefore we depend on an unpaired dataset. The overall architecture is given in Fig. 2.1 and the following subsections will describe the individual components.

2.3.1 Retinex decomposition network

The first part of the pipeline performs retinex decomposition on the images. Retinex idea could be extended to digital images (where S and R are 2 spatial dimensional matrices with 3 colour channels and I is a 2 spatial dimensional matrix with 1 channel) as $S = R \circ I$ where $\circ$ is the spatial element-wise multiplication. The images are decomposed into two parts as the reflectance R and the illumination I.

- Reflectance R : This part has the colour information of the image. This is a 3 channel image of dimensions similar to the original image. We assume this part to be consistent over different lighting conditions for a single scene/object.

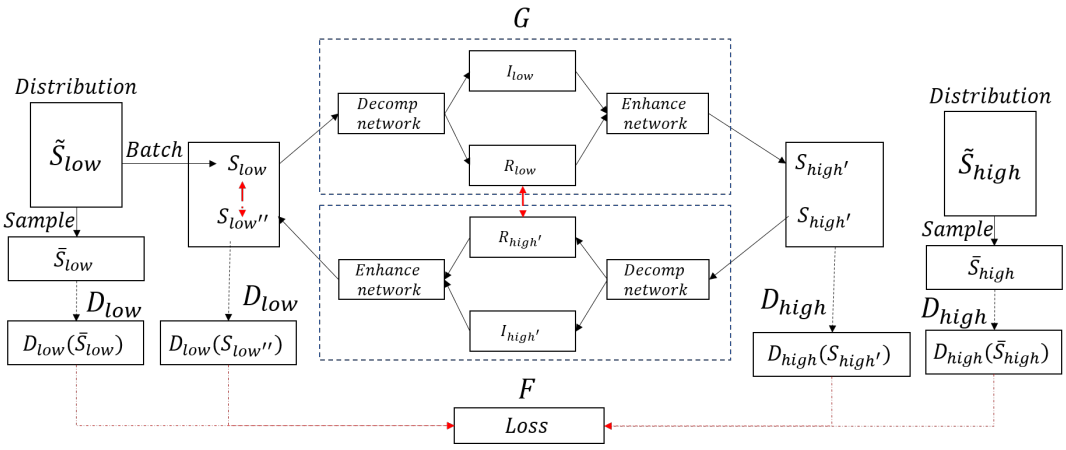


Fig. 2.1 Forward cycle of the retinex-cycle-GAN model

- Illumination I: This is a single channel image that represents the lighting effect of an image.

We define the generators used in the translation $S_{low} \rightarrow S_{high}$ as G_i and the generators used in the translation $S_{high} \rightarrow S_{low}$ as F_i. We define the NN decomposition of low light images as G_1 and the NN decomposition of high light images as F_1. But the symmetric training process ensures them both being the same $G_1 = F_1$. This is illustrated in Fig. 2.1 and Fig. 2.2.

- $G_1 : S_{low} \rightarrow [R_{low}, I_{low}]$
- $F_1 : S_{high} \rightarrow [R_{high}, I_{high}]$

For this, we use Deep Retinex Decomposition Network proposed in [2]. Both G_1 and F_1 were trained using LOL dataset [2] which consists of well-lit and low light, coloured image pairs. 16 random patches of size 64×64 are sampled from an image. Each image is randomly chosen to flip horizontally and/or vertically. The NN was trained to minimize the custom loss defined in [2] with the adam[26] optimizer at a learning rate of 0.001 and a decay factor of 0.9 for 100 iterations.

2.3.2 Enhancement network

The enhancement is done using a neural network. The architecture is inspired by U-Net [27] and its modifications over U-Net are given in Table 2.1. This CNN operates on the concatenation of R and I as the input (unlike U-Net). This will output an enhanced version of the illumination map. Since it is difficult to find a dataset with low light and

well-lit image pairs, using a paired dataset learning technique to train the enhancement network is not practical. Therefore, we use an unpaired dataset with a Generative Adversarial Network (GAN) to train the enhancement network.

Table 2.1 U-Net Specifications

Parameter	Value
Input size	256×256
Down/upsampling factor	$\frac{1}{2}, 2$
No. down/upsampling layers	$7, 7$
No channels (hidden layers)	4,128,256,512,512,512,512,512
Conv kernel size	3×3

We use the cycle consistency loss [22] to train the enhancement network with an unpaired dataset.

Cycle consistency

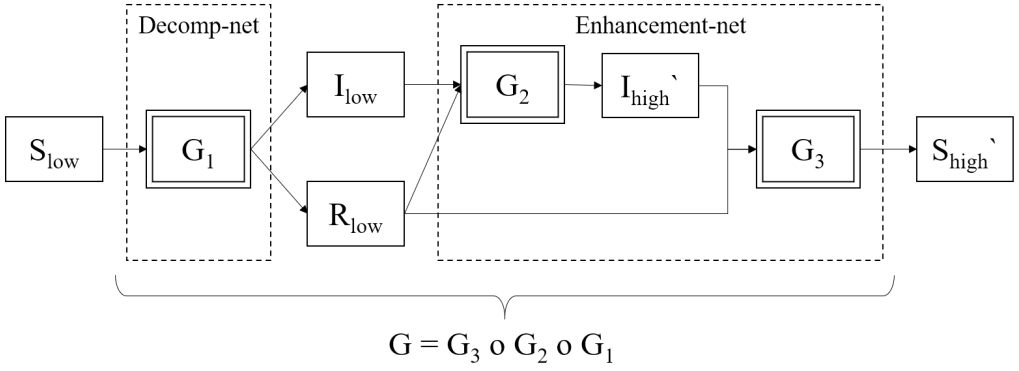
To enhance the training process we use two GANs[7] that will generate well-lit images from low light images and vice versa. We use these GANs as illustrated in Fig. 2.1 to preserve the cycle consistency using an adversarial loss [22]. This reduces the distance between the generated image and the expected image distributions.

The "cycle" consists of a forward pass $S_{low} \rightarrow S_{high'} \rightarrow S_{low''}$ and a backward pass $S_{low} \rightarrow S_{high'} \rightarrow S_{low''}$. This is further explained in Fig. 2.2, Fig. 2.1 and the list of equations given below.

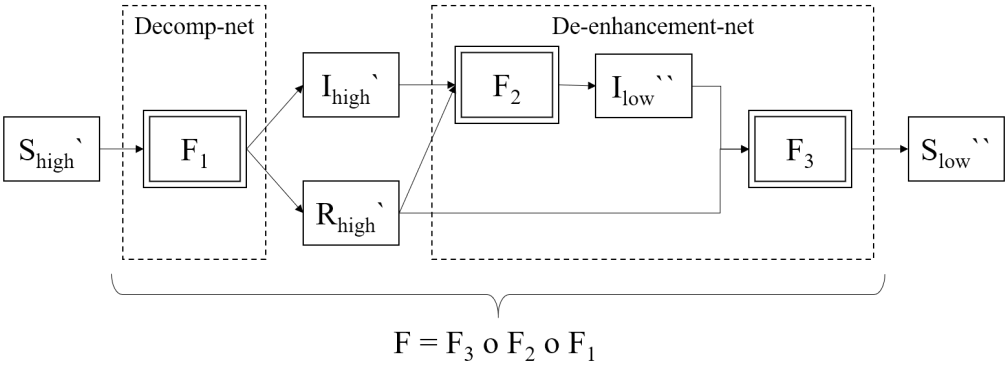
- $G_2 : [R_{low}, I_{low}] \rightarrow I_{high'}$
- $G_3 : [R_{low}, I_{high'}] \rightarrow S_{high'}$
- $S_{high'} = G(S_{low}) \approx S_{high}$
- $F_2 : [R_{high'}, I_{high'}] \rightarrow I_{low''}$
- $F_3 : [R_{high'}, I_{low''}] \rightarrow S_{low''}$
- $S_{low''} = F(S_{high'}) = F(G(S_{low})) \approx S_{low}$

G_1, G_2, F_1 and F_2 are trainable components (based on U-Net architecture) while G_3 and F_3 are non-trainable components.

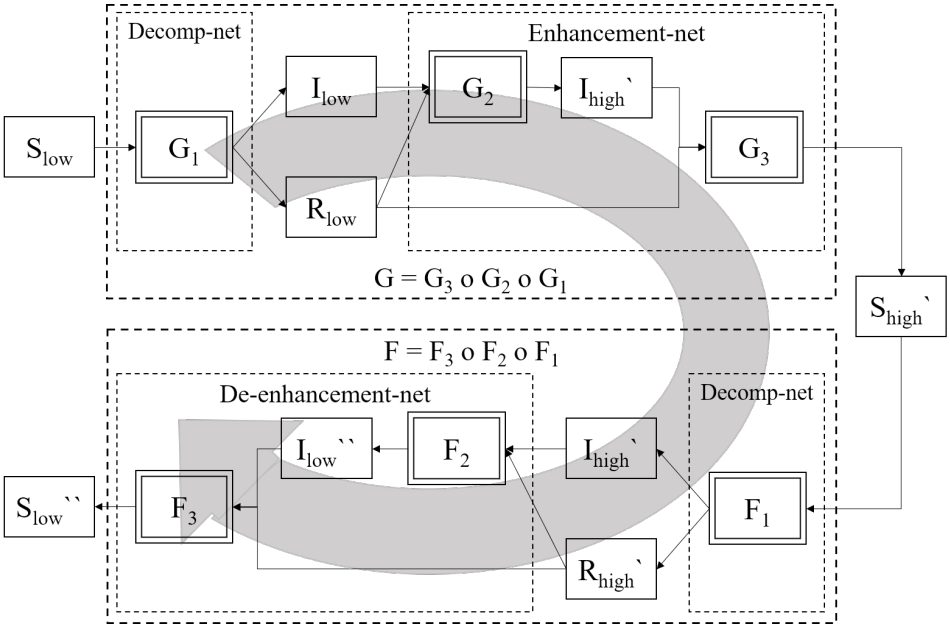
To preserve the cycle consistency of the forward cycle S_{low} and $S_{low''}$ should be the same. When considering the backward cycle, S_{high} and $S_{high''}$ should be the same. Thus,



(a) Forward generator.



(b) Reverse generator.



(c) Full cycle.

Fig. 2.2 Block diagram of the generator components.

we include a cycle consistency loss by taking the difference between S_{low}, $S_{low''}$ and S_{high}, $S_{high''}$. as shown in Eq. (2.1). In addition, using the retinex theory we know that reflectance map in each forward/backward cycle should be equal as well [2]. Therefore, we include the difference between R_{low}, $R_{high'}$ and R_{high}, $R_{low'}$ in the cycle consistency loss as well. This loss is given in Eq. (2.2). Then total cycle consistency loss is given by Eq. (2.3)

$$\begin{aligned}\mathcal{L}_{cyc_S} =& \mathbb{E}_{S_{low}\ p(S_{low})}\left[||F(G(S_{low})) - S_{low}||_1\right] + \\ & \mathbb{E}_{S_{high}\ p(S_{high})}[||G(F(S_{high})) - S_{high}||_1]\end{aligned} \tag{2.1}$$

$$\mathcal{L}_{cyc_R} = ||R_{low} - R_{high'}||_2 + ||R_{high} - R_{low'}||_2 \tag{2.2}$$

$$\mathcal{L}_{cyc} = \mathcal{L}_{cyc_S} + \mathcal{L}_{cyc_R} \tag{2.3}$$

Training of enhancement network

The total loss for the enhancement networks G and F including the generator loss is given in Eq. (2.4) and Eq. (2.5) where $H(p, q)$ is the binary cross entropy of distribution q relative to a distribution p.

$$\mathcal{L}_G = \mathcal{L}_{cyc} + H(D_{high}(G(S_{low})), \mathbf{1}) \tag{2.4}$$

$$\mathcal{L}_F = \mathcal{L}_{cyc} + H(D_{low}(F(S_{high})), \mathbf{1}) \tag{2.5}$$

These NNs (G_2, G_3, F_2, F_3) were trained against a dataset created by adding 485 well-lit images from LOL dataset. The generators were trained to minimize the loss Eq. (2.4) and Eq. (2.5). We use the Adam optimizer[26] at a learning rate of 0.0002 for the discriminator and generator. A decay factor of 0.5 is used. Each model is trained for 500 epochs with batch size 8.

2.3.3 Patched Discriminator

Usually, GANs use a discriminator to distinguish between real and generated data. When generating images, the whole image is processed and one scalar value between 0 and 1 is predicted. This method is not feasible for this problem because some regions in the image are more specific than the others. For example, an image with low-light in a small

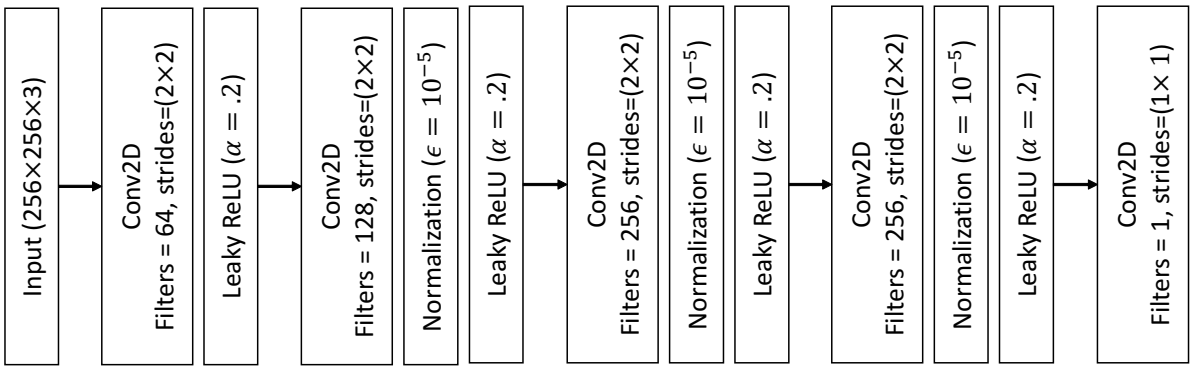


Fig. 2.3 Discriminator model

area can be identified as a well-light image by the discriminator if the whole image is considered.

Therefore, we use patches from the image and discriminate each patch. Then we take an average of those values. Using this method, we can improve the training of the discriminator. The discriminator model is explained in detail in Fig. 2.3. The loss for each discriminator is given in Eq. (2.6) and Eq. (2.7).

$$\mathcal{L}_{D_{high}} = H(D_{high}(G(S_{low})), \mathbf{0}) + H(D_{high}(S_{high}), \mathbf{1}) \tag{2.6}$$

$$\mathcal{L}_{D_{low}} = H(D_{low}(F(S_{high})), \mathbf{0}) + H(D_{low}(S_{low}), \mathbf{1}) \tag{2.7}$$

2.3.4 Summary

The first NN alone can produce R,I decomposition for both low light and well-lit images after training. The enhancement network performs the enhancement of the I and puts together R and I to generate enhanced image. The second network is optimized to minimize a wide range of errors.

2.4 Results and discussion

2.4.1 Performance metrics

In general, the preferred evaluation measure ranking based on visual feedback. For this purpose, this section contains examples of the results generated by the proposed pipeline. The results were also analysed for the pixel-wise mean squared error whenever possible to provide a numerical performance evaluation. Also, we are using the Structural Similarity

(SSIM) and Naturalness Image Quality Evaluator (NIQE) as the standard metrics to compare the results. Note that we are using the ratio between the NIQE values of the predicted image and the ground truth image. Furthermore, in order to validate the reasoning behind the specific components of the proposed pipeline, several results were generated to ensure how the inner workings of the system comply with our design goals (instead of treating NN as a black box).

2.4.2 Retinex decomposition

Figure 2.4 shows the performance of the decomposition network for a chosen image. The similarity of $R_{high} \approx R_{low}$ even when S_{high} and S_{low} are different is indicative of the performance of the decomposition network.

Fig. 2.4 Results of decomposition network. Row 1 (L to R): $S_{low}, R_{low}, I_{low}$. Row 2 (L to R): $S_{high}, R_{high}, I_{high}$

2.4.3 GAN based illumination enhancement

The GAN is used to enhance the images and illumination maps. The notation for these processes are: $\rightarrow$ denotes NNs trained under adversarial loss and $\rightleftharpoons$ denotes NNs trained under cycle consistency loss (with or without additional loss functions). Two separate experiments were done as per the following approaches.

By training the GAN on $[R, I]$, The performance of this network is shown in Figure 2.5. It is clearly visible how CycleGAN produced a better illumination enhancement compared to generic GAN.

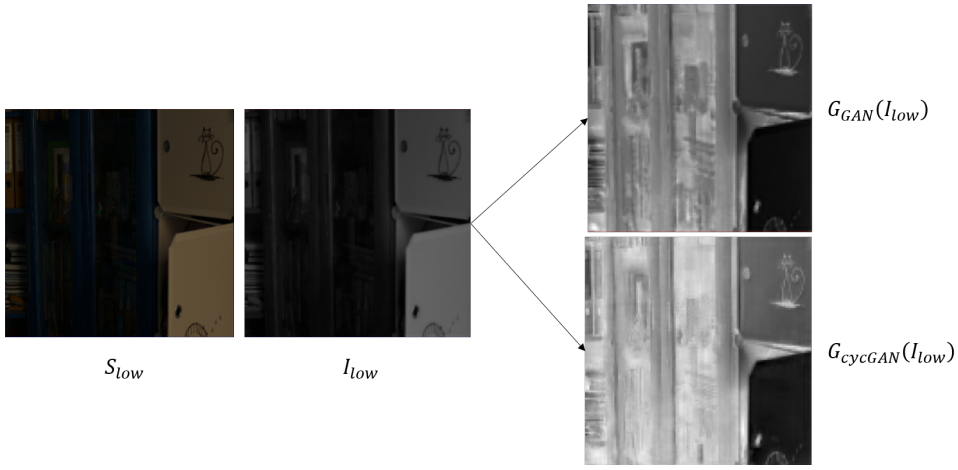


Fig. 2.5 Comparison of the results from CycleGAN based illumination enhancement and GAN based illumination enhancement.

2.4.4 Complete low-light image enhancement system

Our work studies the results from four approaches to low light enhancement (G and G_1 are explained in the proposed work section. G^a,G^b are different NNs whose architecture has been kept as close as to G for benchmarking).

1. $G^a : S_{low} \rightarrow S_{high}$
   Enhancing low light images using a generic GAN.

2. $G^b : S_{low} \rightarrow S_{high}$ with $G_1 : (R_{low}, I_{Low}) \rightarrow I_{high}$ using supervised training data.
   Enhancing low light images using Retinexnet.

3. $G^c : S_{low} \rightarrow S_{high}$ with $G_1 : (R_{low}, I_{Low}) \rightarrow I_{high}$ as an intermediate step. Enhancing low light images using a retinex aware GAN pipeline.

4. $G : S_{low} \rightleftharpoons S_{high}$ with $G_1 : (R_{low}, I_{Low}) \rightleftharpoons I_{high}$ as an intermediate step. Enhancing low light images using a retinex aware CycleGAN pipeline.

The results of all the four cases are given in Fig. 2.6. The comparison between them are given in Table. 2.2. As shown in the table both Retinexnet and proposed method have low MSE but when compared with NIQE our method shows better results overall.

Table 2.2 Numerical comparision of algorithm performance (MSE - lower is better, NIQE ratio - higher is better)

Algorithm	MSE	NIQE ratio
Ground truth	0.0000	1.0000
DCGAN	0.2171	1.4592
Retinex based DCGAN	0.0514	1.6967
Retinexnet	0.0090	1.7896
Proposed algorithm	0.0173	1.7921

2.4.5 Study of the inner-workings of the model.

The importance of each component of the model is analysed next. Figures 2.7a and 2.7b show how the NN works in the hidden layers. Layers (from left to right denoting layers from the input to the output) show how the light enhancement process happens sequentially. In figure 2.7b we observe that this results has been further enhanced through cycle consistency loss. In figure 2.7b, the similarity of the reflectance of the low light image (R_{low}) and well-lit image (R'_{high}) highlights the functionality of the cycle consistency loss we introduced. We observe further enhancement in the output S'_{high} which shows that the generative model plays a significant role in image enhancement.

When upsampling latent representation of the "U-net" model in G_2 generator, we resize the image using nearest neighbour method instead of using transpose convolution layers to avoid checkerboard like effect in the final image. Also, we use the tanh activation only at the end of the generator to normalize and rescale the final image. We use L2 regularizers for each kernel to avoid overwhelmingly increase weight values.

2.5 Conclusion

Ability to work under low light conditions is an important goal for computer vision. Among other things, enhancing the lighting in images is a crucial milestone in this. Classical image processing algorithms and traditional deep learning algorithms have their strengths and weaknesses.

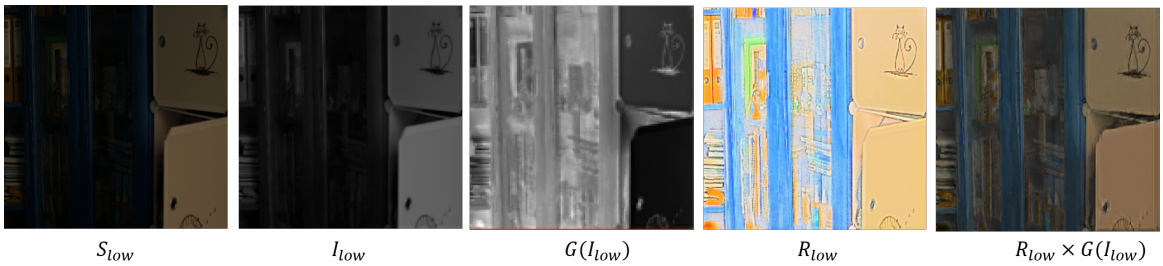
In this chapter we proposed a image processing pipeline to overcome several weaknesses of traditional deep learning algorithms by being able to use paired and unpaired datasets and use both traditional CNN and GAN architectures in a collaborative manner.

2.6 Future work

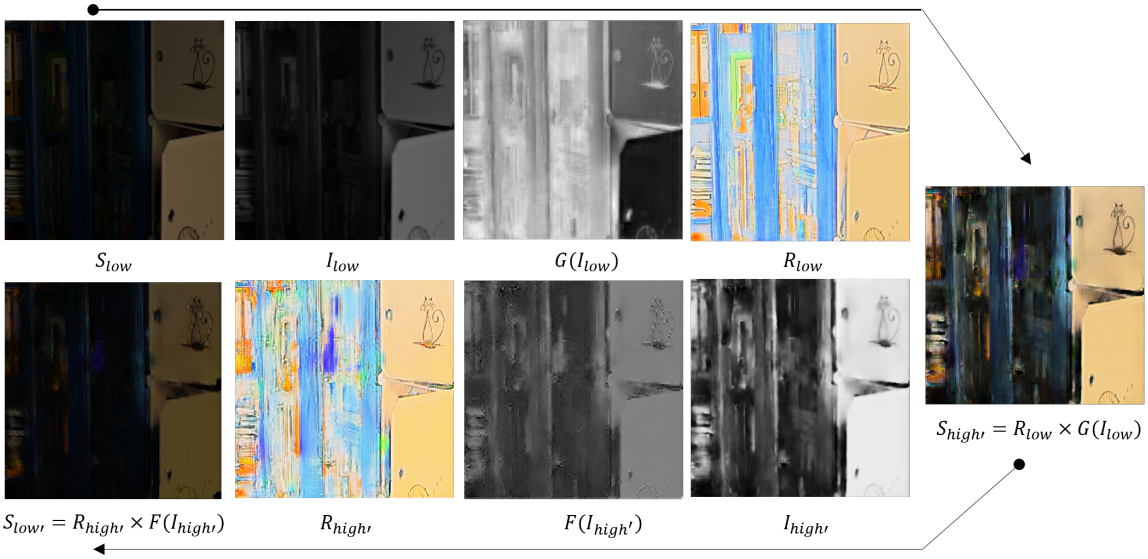
Even though this paper proposes a methodology to use both paired and unpaired datasets in the enhancement pipeline, the individual components use only one type of dataset. The future work should try to use both types of datasets in each step of the pipeline. The enhanced images of this work shows issues with respect to smoothness. Future work should explore the possibility of enhancing images while preserving natural-like smoothness.

The light enhancement problem is studied in this work as a two-class problem – low light and well-lit. The future work should try to realize different discrete levels of lighting enhancement followed by continuous levels of lighting enhancement. Furthermore, object detection, segmentation, captioning, etc. for low light images could be built upon this work.

Fig. 2.6 Results from the complete pipeline. Row 1: Low light input images. Row 2: Well-lit ground truth. Row 3: G^a Generic GAN enhanced image. Row 4: G^b Retinexnet enhanced image. Row 5: G^c retinex based GAN enhanced image. Row 5: G Retinex based CycleGAN (ours) enhanced image.



(a) Layer output from GAN with retinex model.



(b) Layer output from Cycle-GAN with retinex model.

Fig. 2.7 Intermediate layer outputs of generator models.

Chapter 3

Controlled illumination enhancement

Low-Light image enhancement is a well defined problem in the typical case study with two types of images: low-light and well lit images. Many models have been proposed for low light image enhancement and they give impressive results for low-light to well-lit image transformation. However, when translated to real world applications, these models often encounter performance degradation and the level of performance observed during experimentation is not maintained [28, 29].

This occurs mainly due to two reasons: (1) The two-class image enhancement problem only considers images under two type of lightning conditions rather than a range of lightning which we may encounter in real world applications. (2) The images in general have spatially uniform distribution of exposure, whereas in reality, the distribution of exposure may vary depending on many factors including the luminance, angle of incidence, artificial lightning sources, etc.

Therefore, the aforementioned models fail to maintain the performance level. Light enhancement under varying light conditions is not well defined as the two-class enhancement problem. Most of the prior works concentrate on the latter problem of adjusting exposure under exposure distortion while the former problem in which the dataset might consist of images from a range of lighting conditions is rarely addressed. In addition an interesting study would be the ability to create images of varying lighting conditions rather than a prelabeled “optimal” image using a model. These observations raise two interesting questions in regards to controlled illumination enhancement which are studied in the following sections.

1. Given a dataset containing images with different levels of illumination across images and varying levels of exposure within the same image, could we propose any improvements/modifications to perform "exposure-aware" image enhancement.

2. Would it be possible to control the light intensity/exposure of the output image using a dataset containing images taken under different lighting conditions.

3.1 Related work

Many prior works have studied and proposed models for exposure adjustment and illumination enhancement. However, traditionally, these two problems have been addressed separately. In terms of terminology brightness/illumination level refers to the light incident on the overall image while exposure focuses on the distribution of light in the image. Thus, it is evident that the lighting level adjustment cannot be handled without addressing the exposure distribution. In certain cases, an uneven exposure is preferred for visual/aesthetic value. However, in the presence of high exposure distortion, vision algorithms do not perform well.

In general images which suffer from exposure distortions can be enhanced using many of the prior works discussed. However, the quality of the output image degrades with the enhancement. Therefore, highly distorted images suffer from poor quality as we attempt to enhance the illumination. Therefore, illumination enhancement must be handled together with exposure adjustment rather than a separate problem.

In [30], the authors propose a model with controlled illumination adjustment based on the predicted exposure maps. The model selective enhances certain regions by partitioning the image based on the over/under-exposure of certain regions. In addition, the authors introduce a metric to measure the artificial illumination levels of an image

Even though the literature on hybrid techniques (exposure and illumination enhancement) is limited, the individual problems are well defined and studied in detail. Enhancement of images with exposure distortion have been studied in [30], while illumination enhancement has been studied in [13, 31]. We note that, though these techniques do not explicitly address both the issues, the models often implicitly handles both of them. However, by highlighting the inferior performance of these models, we emphasize on the need for improved models.

Recently learning algorithms have shown to improve the performance in many applications significantly. Motivated by this, we propose a deep learning model to handle controlled illumination and try to improve the performance in comparison to prior works.

3.2 Dataset

The dataset required for this problem must consist of a large set of images which vary in terms of both the brightness/illumination and the exposure distortion. Since there does not exist a single dataset with all these required variation, we combine multiple images from different dataset together with synthetic data to train and evaluate the model.

Multi-illumination levels

Controlled illumination enhancement is not a well defined problem. Thus the datasets available target other applications such as the low-light to well-lit image enhancement, analysis of images under different lightning conditions, study of image quality metrics, etc. Therefore, a labelled dataset with images of different illuminations is not available. However, diverse datasets can be combined with synthetic data as discussed in the following sections.

Discrete and Continuous illumination level

The illumination level can be expressed in terms of different metrics such as Value (V) in HSV spectrum, the luma component (Y) from the YCbCr spectrum or Lightness value (L^*) in CIELAB color space. The dataset can be labeled using one of these metrics. However, an important consideration is regarding the distribution of images. The labels suffer from low information loss when the images are labeled directly using the metric without separating them into different classes. However, generic supervised learning models cannot handle continuous data. Thus, discrete labels are preferred and the continuous label is divided into discrete ordered labels and the images are separated into classes depending on the label.

Dataset generation

As highlighted, the training set requires paired images. These are synthesised using the LOL dataset [32] and then classified into 10 classes based on the luma component (Y) from the YCbCr colour space. The synthesis algorithm is described below. The final training set contains 500 images distributed evenly over the 10 classes. The test is made by combining images from multiple dataset. These images are chosen at randoms from low-light image datasets [33], over-exposed image dataset (synthetic, HDR), exposure distorted image set [30], and other images from well-lit image dataset [34]. The final test set contains 100 images.

Image synthesis

The image generation methodology is based on the work in [25]. Realistic images of different light conditions are synthesized from normal-light images. A well-lit image differs from a low-light/high-exposure image in two ways: brightness/exposure and presence of noise. Thus, the visual quality of images vary with illumination level in the form of an "inverted U" function with a peak when the image has an optimal exposure. The synthesis function is based on a combination of linear and gamma transformation and by analysing different images, the authors verify that the synthesis algorithm performs well. The transformation function is given as:

$$I_{out} = \beta \times (\alpha \times I_{in})^{\gamma} \tag{3.1}$$

where I_{in} corresponds to the pixel values of the input image scale between 0 and 1 in the RGB colour space. α and β are linear scaling parameters and γ is the gamma transform parameter, and the three paramters are chosen from a uniform distribution. While the authors propose the transformation function for low-light image synthesis, we extend it to both low-light and high exposure synthetic images by changing the transformation parameters as shown in Table 3.1.

Table 3.1 Image synthesis parameters

Image synthesis type	α	β	γ
Low-light image synthesis	U(0.9, 1)	U(0.5, 1)	U(1.5, 5)
High-exposure image synthesis	U(0.9, 1)	U(0.9, 1)	U(0.3, 1)

Fig. 3.1 shows the few samples from the dataset belonging to each class. Through observation, we divide these classes of images into 3 sets as low-light (0-2), normal light (3-6) and over-exposed (7-9) images.

3.3 Proposed Methodology

In this study, the image enhancement model introduced in the previous case study was modified so that the model could learn from images of different levels of illuminations rather than from low-light/well-lit image pairs. The results show that this modification results in more equalized performance over a range of exposure/brightness at the expense of a slight drop in optimal performance. Next, to incorporate the ability to control the illumination of the enhancement process, a conditional generator was introduced. The

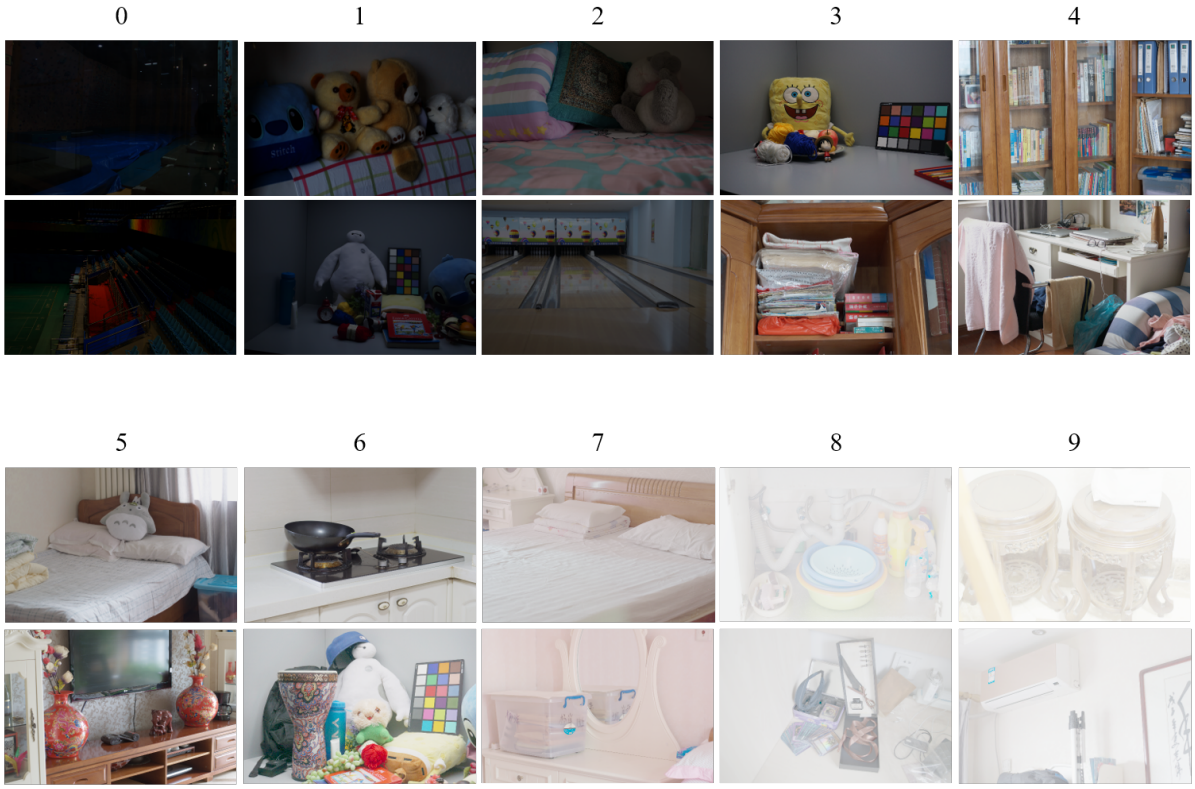


Fig. 3.1 Sample images from all classes

architecture and functionality of both these models are explained next. Note that the cyclic consistency check has been eliminated in these models since the models are trained on paired dataset. Since we use image synthesis to generate the dataset, the dataset size does not act as a bottleneck for training. However, using synthetic data has its downside when extending its application to real-world problems as discussed before and this issue must be addressed in future.

3.3.1 Image enhancement under varying lightning

The model studied in this section has a similar architecture as in the previous case study. However, in this case the dataset contains images taken under varying light conditions which are affected by extreme over/under-exposure. Furthermore, we observe that the original model suffers due to colour distortion when enhancing images with high exposure distortion. Thus, in order to handle this issue, the over/under-exposure maps are explicitly fed to the model.

Fig. 3.2 shows the architecture of the modified image enhancement model. The synthetic/well-lit image pair is first fed into the decomposition-net to obtain the corre-

sponding reflectance R and illumination I pairs. Next, the reflectance of the synthetic image is enhanced using the restoration-net . Next, the over-exposure and under-exposure maps are predicted using the exposure prediction model. Then, the over-exposure map and under-exposure map is concatenated with the illumination of the synthetic image and enhanced using the illumination-net. Finally, the enhanced reflectance $\hat{R}$ and illumination $\hat{I}$ is fused together to produce the output image S_{out}. The fusion operation in this case refers to the element wise multiplication, where $S_{out} = \hat{I} \circ \hat{R}$. Each component of the enhancement model is explained below.

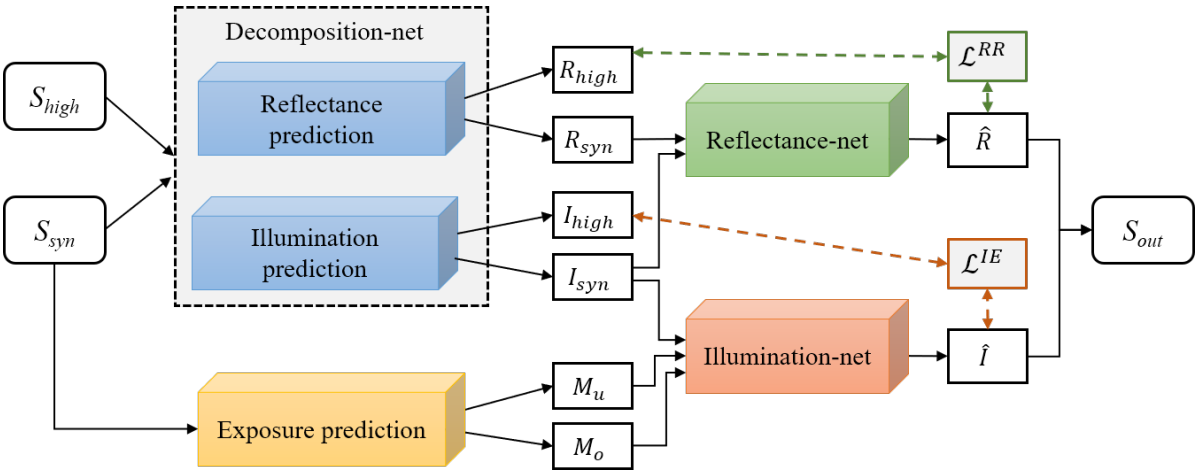


Fig. 3.2 Image enhancement model.

Decomposition-net

The decomposition-net is based on the work in [35], The model and its functionalities have been explained in Section 2.3.1. However, in this case, instead of the low-light image, the synthetic image is fed into the decomposition-net.

Restoration-net

Next, the reflectance is enhanced using the restoration-net. The restoration-net model is based on the illumination-aware restoration methodology proposed in [31]. When enhancing low-light images, the darker regions are polluted with noise. Thus, in the decomposed reflectance, the darker regions are heavily polluted. This degradation can be mathematically modeled as $S = R \circ I + E$, where E is the pollution component. Since pollution is illumination dependant, it may be decoupled as $S = R \circ I + \bar{E} \circ I$ where $\bar{E}$ is the illumination independent pollution component. However, studies have shown that modelling $\bar{E}$ is a complex problem, and removing the pollution E directly from the input

image S also poses certain challenges. Motivated by these finding the restoration-net tries to perform an illumination-aware reflectance restoration.

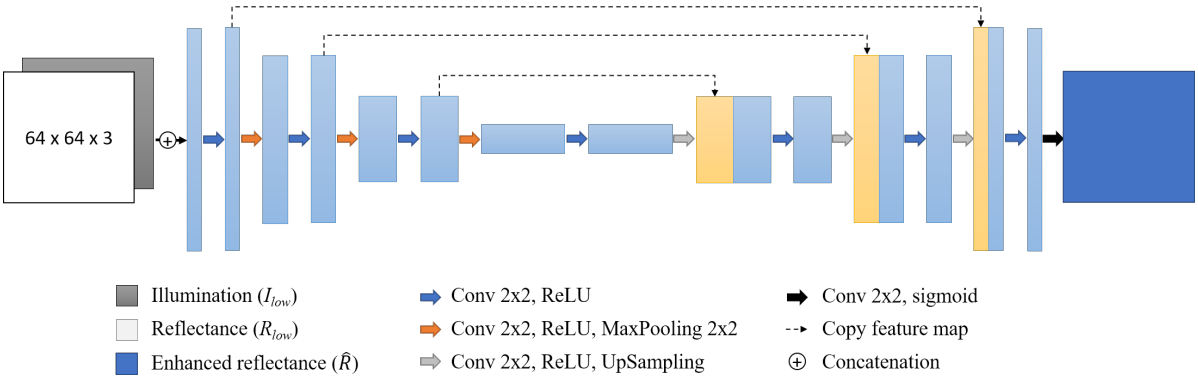


Fig. 3.3 Restoration-net architecture.

The architecture of the restoration-net is shown in Fig. 3.3. The loss function of the restoration-net is given as,

$$\mathcal{L}^{RR} = \left\|\hat{R} - R_{high}\right\|^2 + \left\|\nabla\hat{R} - \nabla R_{high}\right\|^2 - \text{SSIM}(\hat{R}, R_{high})$$

where $\text{SSIM}(\cdot,\cdot)$ is the structural similarity metric, $\hat{R}$ is the output of the restoration-net, R_{high} is the reflectance component of the well-lit image which is used as a ground truth. The first component $\left\|\hat{R} - R_{high}\right\|$ measures the error between the output and the ground-truth. The second component $\left\|\nabla\hat{R} - \nabla R_{high}\right\|$ measures the error of between the gradients which corresponds to the closeness of the texture/smoothness of the two reflectance. The SSIM measures the similarity between the reflectances.

Exposure prediction model

The under-exposure and over-exposure map is estimated based on the model proposed in [30]. The under-exposure and over-exposure maps are created based on pixel saturation and intensity. The pixel saturation is similar in both under-exposure and over-exposure conditions: low saturation pixels are perceived to be close to gray, while well-exposed pixels are perceived by their correct colour. On the other hand the intensity shows different variation with exposure : The over-exposed pixels have high intensity, while under-exposed pixels have low intensity. Based on these, the under-exposure map M_u

and over-exposure map M_o is modelled in the CIELAB color space as,

$$M_u = 0.5 + 0.5 \tanh\left(\delta_u \left(L_{max} - (\sqrt{a^2 + b^2} + G(L))\right)\right)$$
$$M_o = 0.5 + 0.5 \tanh\left(\delta_o \left(L_{min} - (\sqrt{a^2 + b^2} - G(L))\right)\right)$$

where L, a, b are the L^*, a^* and b^* values scaled between 0 and 1. δ is the scaling coefficient of the tanh function and through trial and error, it is set to $\delta_o = \delta_u = 5$. $G(\cdot)$ is a 10×10 Gaussian filter with $\sigma = 3$. Furthermore, $L_{max} = 1$ and $L_{min} = 0$

The objective of the model is to activate the tanh function depending on the saturation and the intensity. The saturation and the intensity are expressed by the terms $\sqrt{a^2 + b^2}$ and L respectively. Thus, the under-exposure map M_u is activated when both the saturation and intensity values are low, and the over-exposure map M_o is activated when the saturation is low while the intensity is high. The values of M_u and M_o varies between 0 and 1, and it represents their respective probabilities of over-exposure and under-exposure.

Illumination-net

Illumination adjustment in images can be done in multiple ways, even without decomposition. The gamma correction methodology used for image synthesis is one such example. However, we may be unable to learn the relationship between different exposure levels and their effect on other factors such as colour saturation, introduction of noise and other artifacts, etc. Therefore, we propose a supervised learning model to predict the enhanced illumination. There is no ground-truth for the illumination map. Therefore, we use the illumination of the well lit image I_{high} as the target illumination. The loss function of the model is given as

$$\mathcal{L}^{IE} = \left\|\hat{I} - I_{high}\right\|^2 + \left\|\nabla\hat{I} - \nabla I_{high}\right\|^2$$

where $\hat{I}$ is the output of the illumination-net. The first component $\left\|\hat{I} - I_{high}\right\|$ measures the error between the output and the ground-truth. The second component $\left\|\nabla\hat{I} - \nabla I_{high}\right\|$ measures the error of between the gradients which corresponds to the closeness of the smoothness of the two illumination. The architecture of the model is shown in Fig. 3.4.

3.3.2 Controlled illumination enhancement model

Next, the model in Fig. 3.2 is modified slightly to generate images conditionally based on the desired illumination level. Note that we initially labelled the dataset based on the

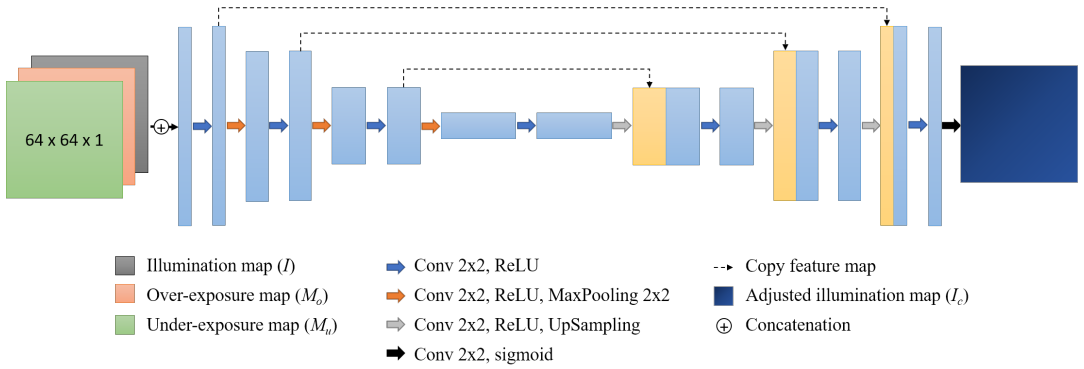


Fig. 3.4 Illumination-net architecture.

illumination level from 0 to 9. We use these class labels as an additional input to the controlled illumination enhancement model to control the illumination level of the output images. The architecture of the model is shown in Fig. 3.5 and the only difference is that the illumination enhancement-net has been replaced with the controlled illumination adjustment-net which adjusts the output illumination conditionally. The functionality of this component is explained next.

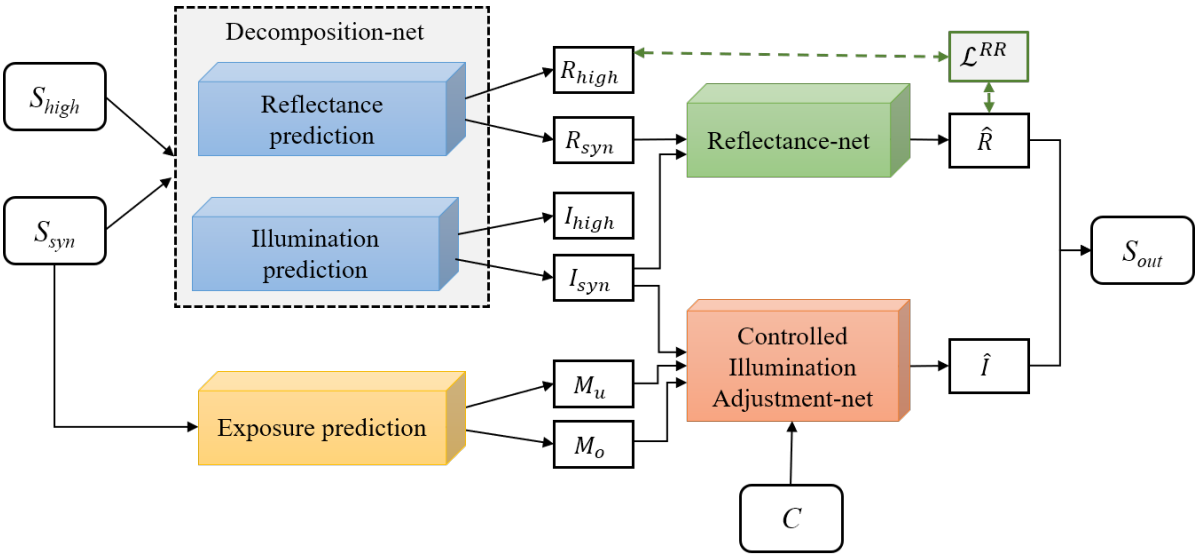


Fig. 3.5 Controlled illumination enhancement model.

Controlled Illumination Adjustment Network

In the previous study, the objective was to enhance the input image to a well-lit image with its well-lit counterpart as a reference/ground-truth. However, in this case, the objective is to produce an image of the desired illumination level. Thus, a given image

may not have an exact counterpart in the dataset with multiple illumination levels. Since we do not have a ground-truth/reference image with the desired illumination level, learning becomes comparatively challenging in this case study. On the other hand, we also note that through retinex theory, we could assume that the reflectance of these images stays the same and thus the only varying component is the illumination. Through this school of thought, we only modify the illumination enhancement component, and instead of using a fully supervised model to predict the image we introduce a conditional generative model which uses adversarial learning to predict the illumination map with the desired illumination level.

Every synthetic image only has a one counterpart, which is the well lit image. Even though this does not have a class label, through the same labeling mechanism used during data generation, we observe that the well-lit image is labelled as "5" frequently. Therefore the controlled illumination adjustment-net can be given as a function as $\text{CIA}(\overline{I}, C)$, where $\overline{I}$ refers to the input information : illumination map (I) and exposure map (M_u, M_o), and C refers to the target class of illumination level. This function must satisfy the following conditions.

$$\text{CIA}(\overline{I}_{syn}, C_{high}) = I_{high}$$
$$\text{CIA}(\overline{I}_{high}, C_{syn}) = I_{syn}$$

where I_{high} and I_{syn} refers to the illumination map of the well-lit image and synthetic image respectively. $\overline{I}_{syn}$ and $\overline{I}_{syn}$ are the inputs to the model, C_{syn} refers to the class label of the synthetic image and C_{high} refers to the class label of the well-lit image (usually 5). To model these conditions we employ the conditional generative adversarial network (cGAN) which is motivated by the works in [36]. As specified in prior sections GANs have two major components : Generator and Discriminator which are described below.

Conditional Generator

The generator is a deep convolutional generative adversarial network (DCGAN) with a U-net design. The architecture of the model is shown in Fig. 3.6. As shown in Fig. 3.6 the label map is fed into the network in addition to other information such as the illumination map, over-exposure map and under-exposure map. The label map is created by passing the label vector of size (10,1) into an embedding layer and then reshaping it to the required shape (64, 64, 1).

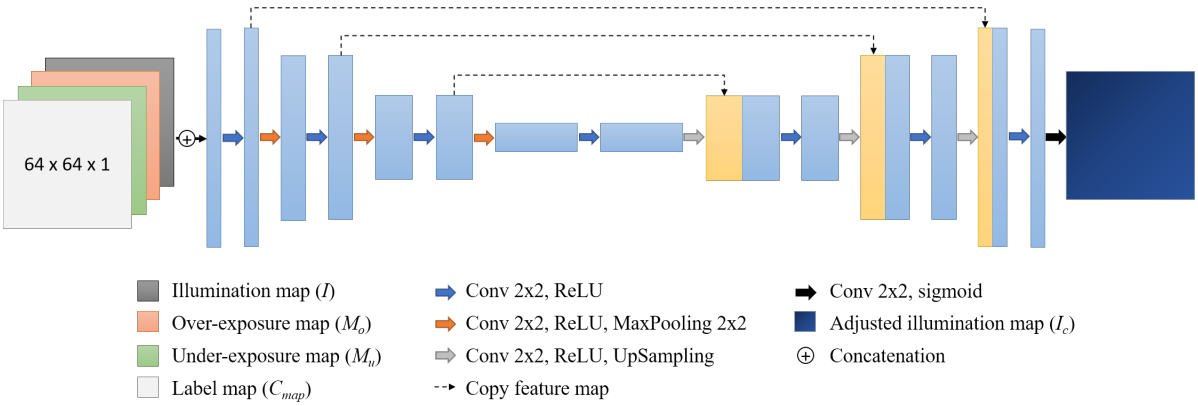


Fig. 3.6 Conditional illumination adjustment generator.

Conditional Discriminator

The discriminator is similar to the GAN discriminator in Fig. 2.3. However, in this case, the class label is passed together with the output image. The architecture of the model is shown in Fig. 3.7.

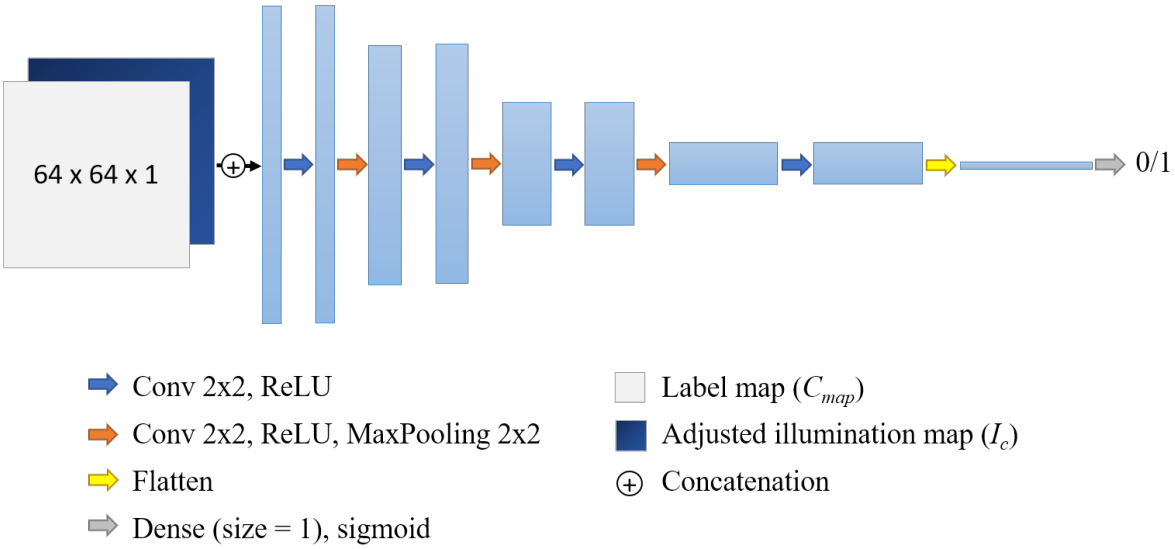


Fig. 3.7 Conditional illumination adjustment discriminator.

cGAN and Loss function

The combined generator and discriminator model is show in Fig. 3.8. For each image pair, the cGAN is trained in two stages. First, the illumination information of the synthetic image is fed to the generator together with the label of the well-lit image as shown in

Fig. 3.8a. The corresponding loss functions of the discriminator $\mathcal{L}_{D_{CIA}}$ and generator $\mathcal{L}_{G_{CIA}}$ are given by,

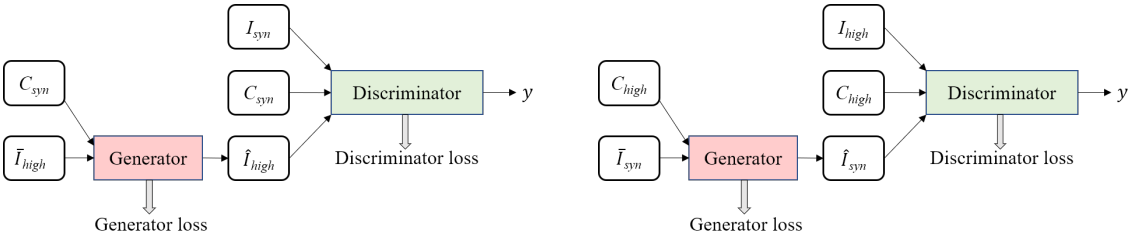
$$\begin{aligned}\mathcal{L}_{D_{CIA}} &= H\left(D_{CIA}\left(\text{CIA}(\overline{I}_{syn}, C_{high}\right), C_{high}), \mathbf{0}\right) + H\left(D_{CIA}\left(I_{high}, C_{high}\right), \mathbf{1}\right) \\ \mathcal{L}_{G_{CIA}} &= \mathbf{E}\left[\left\|\text{CIA}(\overline{I}_{syn}, C_{high}) - I_{high}\right\|^2\right] + \mathcal{L}_{D_{CIA}}\end{aligned} \tag{3.2}$$

Next, the illumination information of the well-lit image is fed to the generator together with the label of the synthetic image as shown in Fig. 3.8b. The corresponding loss functions of the discriminator and generator are given by,

$$\begin{aligned}\mathcal{L}_{D_{CIA}} &= H\left(D_{CIA}\left(\text{CIA}(\overline{I}_{high}, C_{syn}\right), C_{syn}), \mathbf{0}\right) + H\left(D_{CIA}\left(I_{syn}, C_{syn}\right), \mathbf{1}\right) \\ \mathcal{L}_{G_{CIA}} &= \mathbf{E}\left[\left\|\text{CIA}(\overline{I}_{high}, C_{syn}) - I_{syn}\right\|^2\right] + \mathcal{L}_{D_{CIA}}\end{aligned} \tag{3.3}$$

$D_{CIA}(I, C)$ refers to the discriminator model which takes the illumination I and the class label C as the input and outputs 1 if the illumination map belongs to the class and 0 otherwise. $\text{CIA}(\overline{I}, C)$ refers to the generator model which takes the illumination and the overexposure information $\overline{I}$ together with the class label C as the input and produces the desired illumination map as the output. $H(\cdot, \cdot)$ is the binary cross entropy measure between two distributions.

In the first stage the model convert the well-lit image illumination to the synthetic image illumination and in the second stage the reverse step is done. Note that the implementation of this training methodology doesn't require separate loss functions since the equations can be interchanged by swapping the input illumination information and the label.



(a) cGAN training - stage 1.

(b) cGAN training - stage 2.

Fig. 3.8 Conditional illumination adjustment model training.

3.4 Results and Analysis

3.4.1 Comparative performance of two-class image enhancement model and the proposed image enhancement model

First we analyze the performance of two-class image enhancement models on the synthetic dataset and compare them with the proposed model. We note that the results are divided into 3 sets based on their labels (i.e. light intensity) as follows : low-light (0-2), normal light (3-6), and over-exposed (7-9). Fig. 3.9 shows the output of different two-class illumination enhancement models with input images of varying brightness and exposure distributions. Fig. 3.10 shows the variation of quality of the output image from different models as a function of the illumination level of the images. The image quality is measured in terms of the NIQE metric [37]. We observed that the quality of the enhanced image degrades in extreme high-exposure/low-light conditions. The graph takes a "U" shape as noted before. In addition, the PSNR and SSIM values of the enhanced image with respect to the ground-truth is shown in Fig. 3.11.

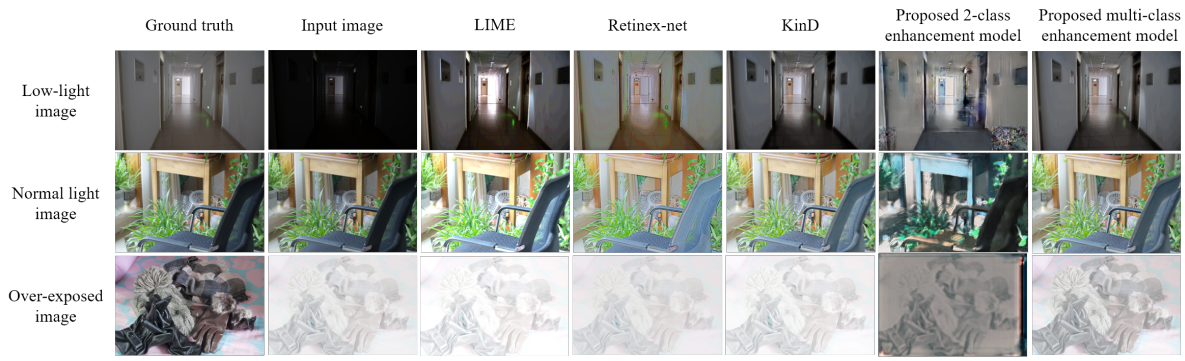


Fig. 3.9 Comparison of the output with varying input image of varying illumination.

Through these results we conclude that two-class enhancement models in general suffer from significant performance degradation in cases where the input image has high exposure distortion or when it is over/under exposed. Furthermore, it must be noted that these models have been proposed to improve the illumination of low light images. Therefore, in high exposure conditions, they further increase the exposure and as a result the the image quality degrades. In comparison to the 2-class enhancement models, the model proposed in this case study performs slightly better in high-exposure conditions, while maintaining a similar performance level at low-light case. Therefore, we can conclude that the additional exposure information provided to the model has helped improve the performance of the proposed model.

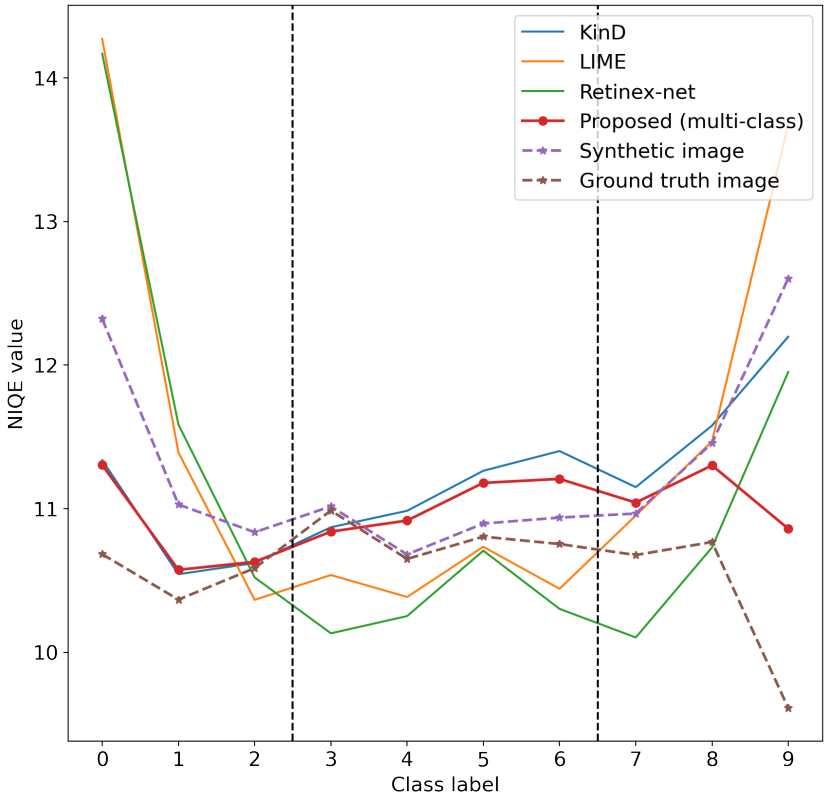
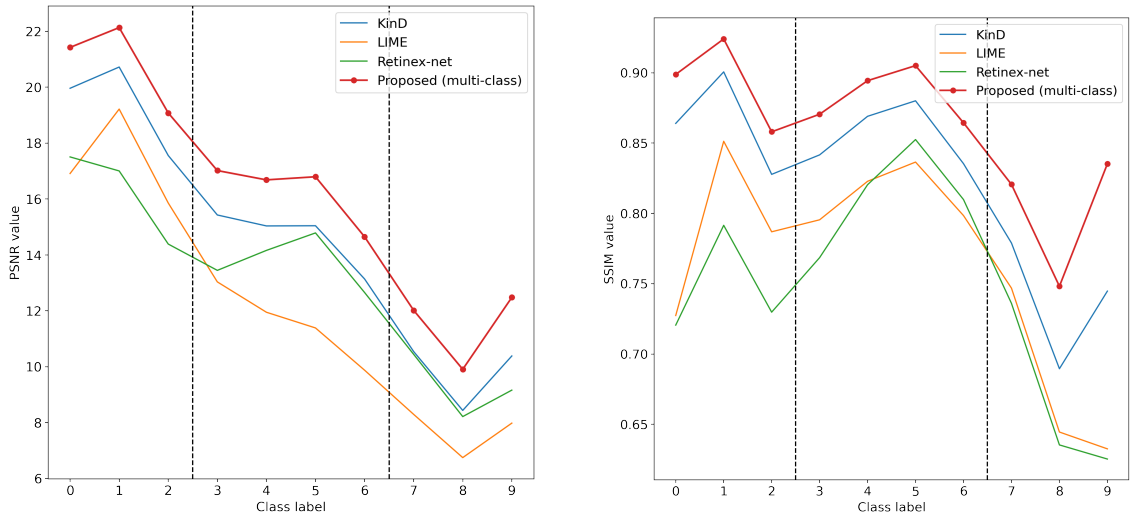


Fig. 3.10 Variation of NIQE of the output with input images of varying illumination.

3.4.2 Multi-class image generation performance

Next we study the performance of the proposed conditional illumination enhancement model which has the ability to produce output images of multiple classes. Similar to the previous case, the images are separated into 3 sets. We analyze the performance of the proposed model by studying the variation of its performance with different input images and expected output illumination classes. For each image the corresponding output from all 10 classes is produced. Fig. 3.12 shows few selected examples of the output produced by the controlled illumination enhancement model. The figure shows the output for each type of input image : low-light (row 1, row 4), normal light (row 2, row 5) and over-exposed (row 3, row 6). For these inputs the expected class label is varied from $C = 0, 1, \ldots, 9$ (column 2-6) and the output obtained is shown. Fig. 3.13 shows the variation of the quality of these output images. We analyze the variation of the quality of the output image as a function of both label of the input image and the label of the expected output image. The quality is measured in terms of the NIQE metric.

The results show that proposed model performs sufficiently well in converting the input image to an output with required illumination level. While we cannot make a



(a) PSNR (b) SSIM

Fig. 3.11 Error between ground truth and enhanced image under varying illumination.

definitive conclusion about the quality of the output image, we can observe that when The variation of NIQE shows that the quality of the output image is comparatively higher when the expected output class is closer to the label of the input image.

3.5 Conclusion

The study shows that the typical two-class enhancement model creates certain challenges when its application is extended to real-world problems. The result suggest that this occurs mainly in conditions with varying brightness and exposure distortion, and the model requires additional information for better performance. Thus, we propose a modification to the low-light illumination enhancement model proposed in the previous section. By explicitly feeding additional information to this model in regard to the exposure maps, we show that the performance on images of varying illumination can be improved. Especially the high exposure image enhancement shows significant improvement in quality compared to the two-class enhancement models

Then we address the controlled illumination enhancement problem and proposed to model to produce images of preferred illumination levels. By using a conditional generative network we eliminate the requirement for all-class pairs of the input image. Therefore, the model can be trained using images from only the well-lit image set and the corresponding multi-level illumination image set. The results show that the model performs sufficiently well. However, since this problem is not well defined there are

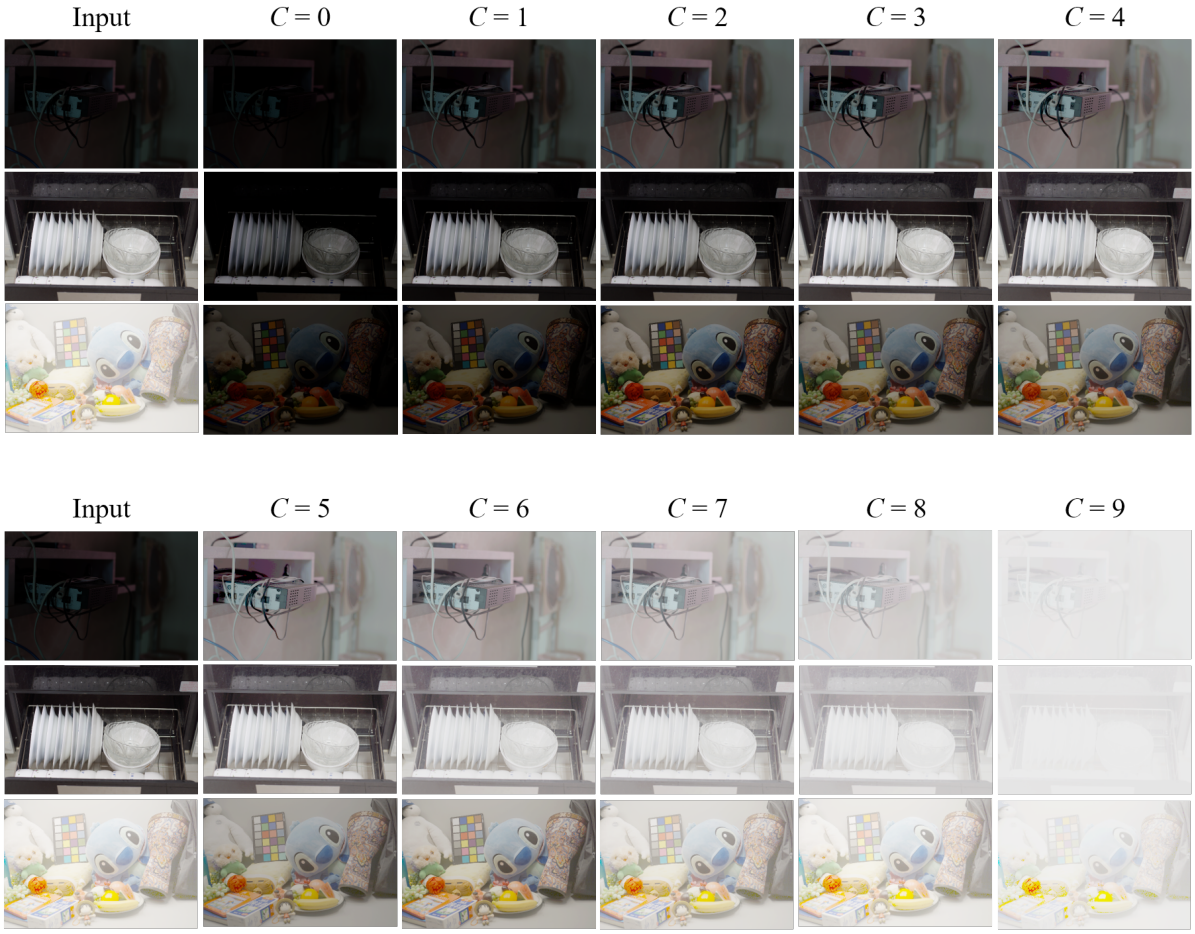


Fig. 3.12 Outputs of the controlled illumination enhancement model.

certain challenges associated with this problem including the availability of non-synthetic data, and the lack of evaluation metrics. These issues will be addressed in the future works.

3.6 Future works

We observed that the proposed modifications improved the results up to a certain extent. However, the model is not perfect and we observe that there is a need for improvement if it is to be extended to real-world applications. Recent works have shown that incorporating additional information into learning models result in significant performance improvement. Thus the proposed model could also be improved in such a way through reasoning, analysis, and ablation study.

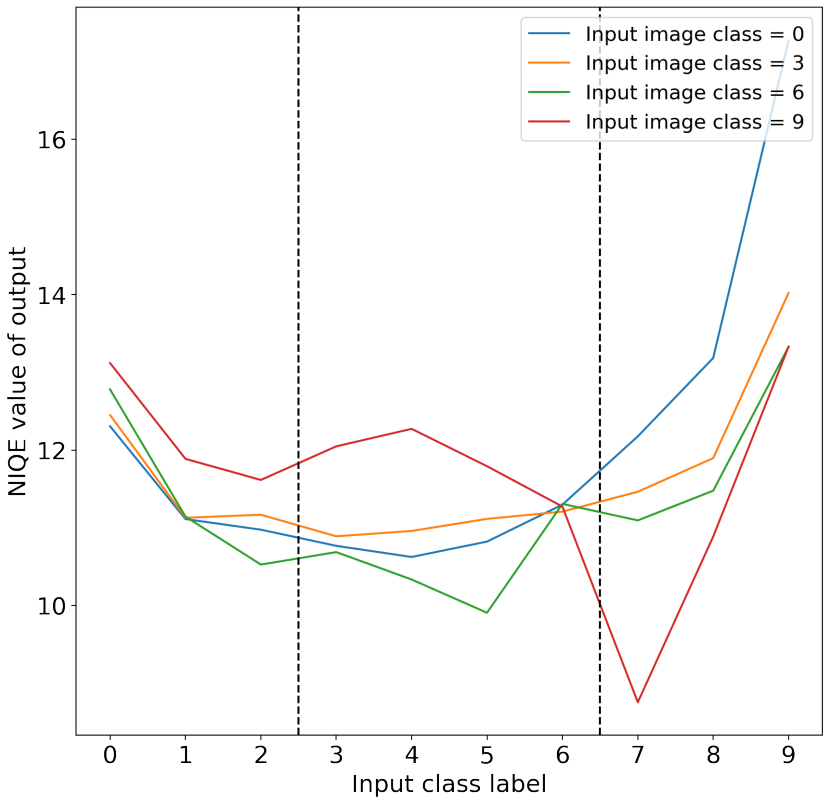


Fig. 3.13 NIQE of the outputs of controlled illumination enhancement model.

Furthermore, in the controlled illumination enhancement problem, since each image does not have a corresponding output from all other classes, error measurements such as SSIM or PSNR cannot be used as evaluation metrics in this case. The performance of the illumination enhancement model is highly dependant on the input dataset, and since synthetic data is used for training, the same performance level cannot be guaranteed in real world applications. Moreover, due to the lack of similar works, performing comparative studies and establishing metrics to analyze the model is a challenging task. These issues will be addressed in the future to make the research problem well defined.

Chapter 4

Object Detection in Low Light Conditions

In this section we explore how object detection performs in different lighting conditions. We start with standard object detection neural networks (trained on standard datasets), compare their performance on declining light levels and quantify the loss of accuracy with the drop of lighting conditions. Then we try to train these networks on low light images. Finally, we explore the possibility of training a neural network on well light images and adapting that knowledge to predict on low light conditions without a significant drop in accuracy.

4.1 Related work

4.1.1 Object detection

Object detection is the task of annotating multiple objects contained in the image. This includes identifying one or more objects in an image as well as determining their location in the image. The simplest form of object detection is image classification, where the task is to classify images where it is already known that an image only consist one object and you don't have to output the location of this object. This task is specifically known as image classification. In the scope of this chapter, both names object detection and image classification are used interchangeably.

Alexnet (2012) [16] is considered to be the first successful attempt to use deep learning for image classification. Alexnet outperformed the state of the art by a considerable margin proving the superiority of Neural Networks for image classification. Since then,

deeper neural networks have been proposed for image classification and they have surpassed Alexnet performance with the years. Inception network [38] is a successful model proposed for object detection in 2016.

Imagenet [39] is a dataset developed for training and testing image classifiers. The dataset contains around 3 million images (in the version that was first released. the latest version have much more). This dataset is considered to be a highly diverse collection since for each object type we get images with different backgrounds, angles, sizes and even occlusions. But most these images are captured during well light conditions.

4.1.2 Lighting condition enhancement

When detecting objects in low light images, the naive solution is to enhance the lighting condition and then use a standard object detection algorithm on the enhanced image. For this case, we can use multiple methods of low light image enhancement. These methods are discussed in section 2

4.2 One Class Problem

One class problem refers to the task of binary classification of images according to the presence of one class or not. Solutions to this problems are instrumental in detecting anomalies, etc.
Earliest research on one class problem have been with PCA, SVM [40] like linear models. Modern work tries to use neural networks [41].

An easier case of the one class problem is when we know the set of images that the class is a part of. Then we have to train the model to work only for the bigger set of images. This is just an application of binary classifiers. But the task becomes challenging when we don't have a set of images representative of what the input images for the neural network will be. We only have the set of images for the class we have to detect (or detect the absence of).

There have been multiple attempts to train algorithms to detect out of class images without a training set. Earlier attempts on this track have been using the idea about decision boundaries in algorithms.

Adverserial training of neural networks have been used in recent works for the one class problem. Here, the idea is to use two competing neural networks (generator – trying to generate out of class examples, discriminator – trying to decide is an image comes from the actual class or is it create by the generator).

4.2.1 OCGAN

OCGAN (One Class Generative adverserial Network) [1] is a recent attempt to tackle the one class detection problem (in the absence of non-class training examples) by a a GAN architecture equipped with negative mining.

Figure 4.1 denotes the architecture of the OCGAN network. Here, we have a few segments of the main NN pipeline.

- **Encoder:** This part of the NN takes the images to a low spatial dimensional representation with a higher number of spectral dimensions. The task of this part is to encode the spatial information in another representation.
- **Latent representation:** This is the representation that is generated in the middle by the encoder. This information should be enough to reconstruct the image. Therefore this can be considered as a spatially compressed representation. It is also important for this representation to have features that could've used by an discriminator.
- **Decoder:** Decoder is the part of the NN that can generate an actual image out of the latent representation. This section consists mainly of deconvolution operations.
- **Latent discriminator:** This is a classifier. Instead of working on actual images, it works on the latent representation of the images. This discriminator is considered to be better at handling noise and transformations in the input space because those information is removed in the encoding process.
- **Visual discriminator:** This is a classifier working on the actual images (which are output by the decoder). This NN is of same architecture as a standard image classifier.

The negative mining process is used to generate informative non-class images (images that are very different from the one class we are interested in). These images are generated as per the following algorithm.

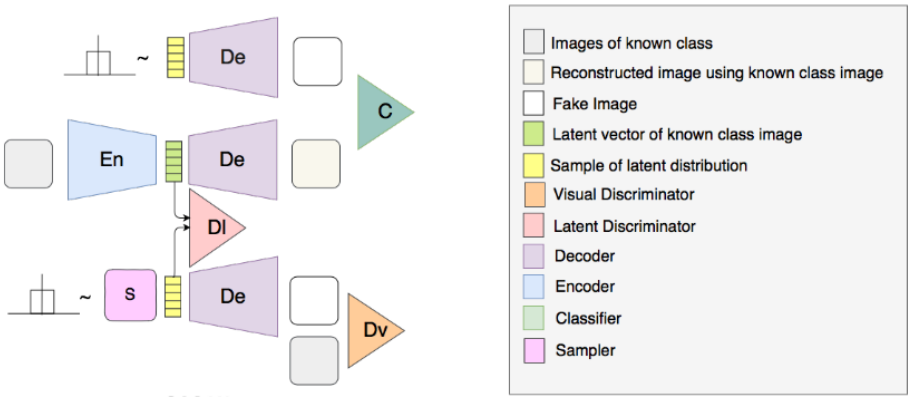


Fig. 4.1 OCGAN Architecture proposed in [1]

4.2.2 Retinex decomp

Retinex decomposition have been discusses in section 2.3.1. In this chapter, we are interested in using the decomposition network (G_1) that decomposes the image (S) into reflectance (R) and illumination (I).

4.3 One class novelty detection in low light images.

4.3.1 Problem definition

This section of the research deals with the problem

How to detect out of class images in low light conditions by training data in well-light conditions of the particular class?

In this section, we try to train an algorithm to detect the out of class images in a low light set of images. As training data we only use well light images of that class. In addition we use *two unpaired sets* of images in low and high light conditions.

4.3.2 Proposed solution

We propose a neural network pipeline motivated by Retinex Decomposition Network and OCGAN [1]. The higher level sections of the algorithm is as following. The pipeline is shown in Figure 4.2. Retinex decomp network is explained in 2.3.1. The discriminator is explained in the section below.

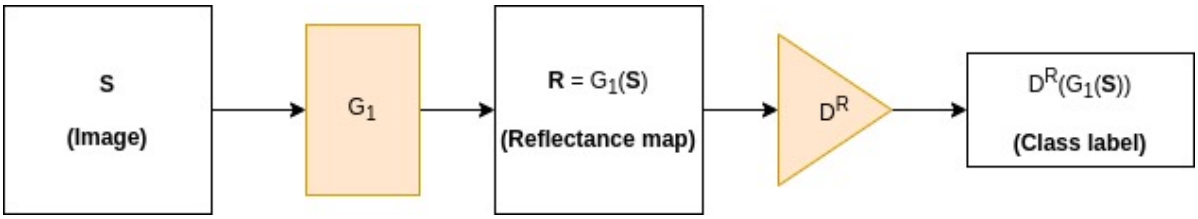


Fig. 4.2 One class detection neural network

- Retinex decomposition network G_1 : This network takes in an image and outputs the reflectance of map. This network is trained to generate the same reflectance map for images of multiple lighting conditions.
- Lighting condition agnostic discriminator D^R: This network is trained to detect the in class images from the reflectance map. This network's performance is invariant of the lighting condition of the original image.

,

Lighting condition agnostic discriminator D^R

This neural network is used to provide a binary classification for the reflectance maps. Sine reflectance maps are of 3 colour channels as well, an RGB image classifier can be used. We design this neural network inspired by Inception architecture in [38].

We identify four types of blocks and layers in the original inception-v3.

- Inception blocks which reduce the size of inputs.
- Inception blocks that does not change the size of inputs.
- flattening layers.
- 1000 class classification softmax layer.

The inception-v3 architecture is designed for classification of a large number of classes. Since one class problem does not require such output, we replace the classification layer. In addition, we opt for a shallower neural network than the original inception architecture by limiting the number of inception blocks in the network.

The final study is done by an inception like neural network with the following components.

- 1 inception block that reduces the size of input.

- 1 inception block that does not reduce the size of input.
- flattening layer.
- Sigmoid output (single scalar value)

The individual block architecture is as as the original work.

Training of neural networks

We have two neural networks in this pipeline. G_1 is trained as described in section 2.3.1. This training process is done by a paired dataset of low light and well light images. It should be noted that this training process does not require training data from the class we are interested in. We can simply use a pre trained G_1 here.

Training the lighting condition agnostic discriminator is the challenging task in this design. Since we already have images ($X_1 \rightarrow Y = 1$) for the "class", we can generate the reflectance maps of these images by sending them through the decomposition network.

Since we do not have images out of the class, we have to generate them. Since we are running the discriminator on the reflectance maps, what we need is negative (out of class) reflectance maps. We use a technique called negative mining for this. The idea is to randomly generate some reflectance map and make it as different as possible from the class. This algorithm will be discussed in detail in the next section.

Since negative mining is done using the discriminator itself, the negatives are more meaningful when the discriminator is trained properly. Therefore, we have to keep training the discriminator while mining negatives. This methodology is shown in Algorithm 1. An illustration is given in Figure 4.3.

Data: X_1, G_1, D^R
Result: updated D^R
$\mathrm{R}_1 \leftarrow G_1(X_1)$
$\mathrm{R}_0 \leftarrow randomly_generate()$
for $i = 0;\ i < 5;\ i = i + 1$ **do**
  $\mathrm{R}_0 \leftarrow negative_mine(R_0, G^R)$
  $\mathrm{Pred}_1 \leftarrow D^R(R_1)$
  $\mathrm{Pred}_0 \leftarrow D^R(R_0)$
  $\mathrm{Error}_1 \leftarrow 1 - Pred_1$
  $\mathrm{Error}_0 \leftarrow 0 - Pred_0$
  $\mathrm{D}^R \leftarrow back_propagate(Error_0, R_0, Error_1, R_1)$ **end**

Algorithm 1: Pseudocode for training D^R

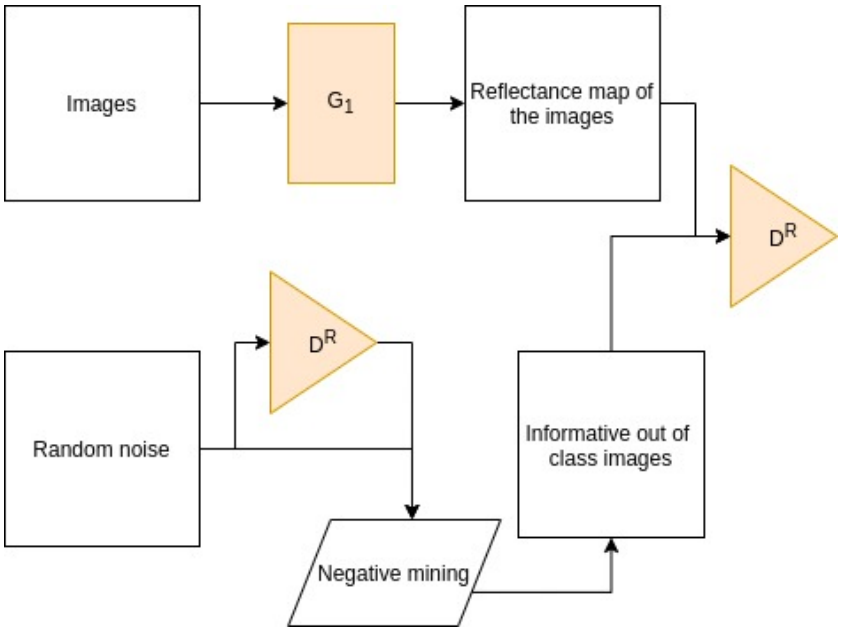


Fig. 4.3 Training one class detection neural network

Negative mining

Negative mining is the task of generating out of class examples to train a neural network. We have experimented with two algorithms for negative mining in our work as

1. Gradient descent based negative mining (as proposed in [1])
2. Random walk based negative mining (novel approach inspired by [42]).

Assuming that the reflectance maps are of the dimensionality $H \times W \times 3$, negative mining is a process of randomly sampling $\mathbb{R}^{H \times W \times 3}$ to get the desired out of class images. In both algorithms we start by randomly picking a point R_0 in this space. Then we use the existing discriminator D^R to judge if this point is considered to be in class or out of class. If this point is in close for with a very high probability, $(D^R(R_0) > 0.5)$, we ignore the point. If not, we try to make it even more out of class by **moving** it away from the class.

How points are moved is the difference of the two algorithms.

1. The gradient descent based algorithm tried to push the point through the gradient in such a way that it goes away from the class. This approach is computationally heavy (since this optimization step requires a lot of computation). This is given in Algorithm 2.

2. The random walk based technique tries to randomly move the point and asses its new value's likeliness of being inside the class. If the likeliness has increased, the move is cancelled. This is done several times to make sure that the point has moved away enough. This approach is not very computationally heavy since we are not running a gradient based optimization algorithm. The implementation is also easier. This is given in Algorithm 3.

Data: D^R
Result: R_0
$R_0 \leftarrow sample_from(\mathbb{R}^{H \times W \times 3})$
while $D^R(R_0) < 0.5$ **do**
  $R_0 \leftarrow sample_from(\mathbb{R}^{H \times W \times 3})$
end
for $i = 0;\ i < 5;\ i = i + 1$ **do**
  $Pred_0 \leftarrow D^R(R_0)$
  $R_0 \leftarrow gradient_descent(Pred_0, R_0, D^R)$
end

Algorithm 2: Negative mining using gradient descent

Data: D^R
Result: R_0
$R_0 \leftarrow sample_from(\mathbb{R}^{H \times W \times 3})$
while $D^R(R_0) < 0.5$ **do**
  $R_0 \leftarrow sample_from(\mathbb{R}^{H \times W \times 3})$ **end**
  $Pred_{00} = D^R(R_0)$ **for** $i = 0;\ i < 5;\ i = i + 1$ **do**
    $step \leftarrow sample_from(\mathbb{R}^{H \times W \times 3})$
    $Pred_{01} \leftarrow D^R(R_0 + step)$
    if $Pred_{01} < Pred_{00}$ **then**
      $R_0 \leftarrow R_0 + step$
      $Pred_{00} \leftarrow D^R(R_0)$
    end
  end

Algorithm 3: Negative mining using random walks

4.4 Experiments

4.4.1 Implementation

The algorithms were implemented using the python based tech stack (with open-cv for image handling, numpy for numerical computing, tensorflow for neural networks and optimization and matplotlib for visualization). The algorithms ran on Kepler server[1] with a NVIDIA Tesla K40 GPU.

Neural network weights were optimized using the adam optimizer in tensorflow library. The negative mining was done using the stochastic gradient descent optimizer in tensorflow. All computations were done with 32 bit floating point numbers.

4.4.2 Tests

We pick dogs class (200 from 800 images) and bicycle class (200 from 652 images) in the extremely dark dataset [33]. We also pick 20 images each from 10 other classes in the dataset. We use the annotations file to crop these images to hold only the object of interest and resize them to 300×300 to be fed into neural networks.

We use imagenet dataset [39] to obtain 1000 images each for dog class and bicycle class. These images are also cropped to 300×300 using their annotation files.

We use a trained G_1 and train two separate D^R's for dogs and bicycles as D^R_{dog} and D^R_{bic}. The discriminators are first trained from the images obtained from imagenet.

We test the two pipelines (with same G_1 and different D^R's) on the one class datasets (dogs and bicycles separately) and out of class datasets (generated by a collection of 10 other classes).

4.5 Results

Since the performance of the Retinex Decomp net is discussed earlier in 2.4.2, we look at the performance of the discriminator (D^R) in this section. Analyze the results of the experiments by a confusion matrix given in Tables 4.1 and 4.2.

The results obtained by the setup is unsatisfactory. Considering the cases of dogs as the class and the bicycle as the class, the average performance is not suitable for a binary classifier.

[1] kepler.ce.pdn.ac.lk

True Class	Predicted class	
	Dog	Other
Dog	98.35	1.65
Other	95.10	4.90

Table 4.1 Performance of the D^R (for dog class) with gradient descent based negative mining

True Class	Predicted class	
	Dog	Other
Dog	83.60	26.40
Other	87.00	13.00

Table 4.2 Performance of D^R (for dog class) with random walk based negative mining

4.6 Discussion of results

The performance of the retinex decomposition network has been discussed earlier. This section discusses the results of the discriminator D^R.

We see overfitting in both approaches (gradient descent based negative mining and random walk based negative mining). This means that the generated out of class examples are not representative of the out of class dataset. In generic terms, "dog images (in class) are closer to cat images (out of class in test dataset) compared to how different cat images are compared to generated images". This implies that the negative mining algorithm is unable to perform adequately in this experiment.

OCGAN was tested for MNIST dataset[43] and CIFAR-10 dataset[34]. Both these datasets are of spatial dimensions 32×32. Our application is on larger images. While it is possible to use negative mining to generate meaningful images in the $32 \times 32 \times 3$ spaces, generating a meaningful image in $300 \times 300 \times 3$ space is difficult.

OCGAN uses a latent representation of very small spatial dimensionality after the encoder. But out pipeline's intermediate representation is of the same size as the input image.

4.7 Conclusion and Future Work

This study on detecting novelty images (in low light conditions) given one class of images (in well light conditions) requires more exploration to come up with a well-performing model. Important future work includes

- Include an additional encoder that reduces the spatial dimensionality of the retinex decomposed object to be fed into the discriminator.
- Try a windowed approach to negative mining so that we can degenerate meaningful negative patches (which is an easier task than generating meaningful negative images).
- Introduce actual negative images (to bootstrap the training process). Actual negative images could be used even as starting points for negative mining. This way, better results could be produced for a reasonable number of optimization iterations in negative mining.

Object detection (even the one class problem) remains a challenging task in low light conditions. Research on this topic will build up an understanding of the light, imaging, perception and knowledge representation that would be important in building up robust image processing algorithms.

Chapter 5

Concluding remarks

Images taken in low light conditions suffer from both poor aesthetics and degraded performance with vision-based algorithms. Low-light image enhancement algorithms play a significant role in solving these issues overcoming the shortcoming of alternative solutions such as HDRI, Night Cameras, etc. The illumination enhancement problem is addressed under 3 cases in this study namely: low-light image enhancement, controlled illumination enhancement and low-light object detection.

In the first case study, we propose a generative learning model for low-light image enhancement. The proposed methodology has improved the performance when compared with the existing models. Apart from them, focus must also be made on finding new ways to incorporate model based solutions to learning models. Furthermore, the efficiency of these algorithms is not analyzed in detail. Also, we have evaluated the images with a resolution of 256×256. A study on the behavior with very large resolution images is necessary. By using efficient supervised learning model real-time performance can be promised and thus the idea of low light imaging can be extended to video processing as well. [1]

The model proposed in the first case study works well in two-class image enhancement problem. However, we observe that the performance degrades significantly in the presence of input images with varying illumination levels. We propose an improvement to the model to enhance its performance in varying lighting conditions, especially under high exposure. Next, we improve the proposed a model for controlled illumination enhancement. The cGAN based model performs sufficiently well in producing images with required illumination level. While the problem and its applications are not well defined, this case study mainly helps us in analyzing the variation of the image properties with illumination.

[1] The work related to this case study was published in [44]

Furthermore, its application could also be extended to automate image transformation to improve and modify image aesthetics.

We try to solve the one class novelty detection problem as an attempt to extend image recognition to low light conditions. We try to bring in the retinex based interpretation of images to extract information from images and detect them later. We also study the possibility of generating out of class images for a given class by an optimization process. The study has revealed a set of challenges in adopting these ideas for low light situations. Even the negative mining process shows issues when trying to scale for larger images. We attempt to modify the optimization technique used for out of class image generation. We also provide a set of future directions to improve upon the models and techniques to try to solve these issues.

The proposed models show improved performance in many cases. However, when considering their application for real-world applications, there are still certain barriers which must be overcome. First, we observe that conventional models may not be robust across different environments. Thus, training the model to be application-specific (e.g.: CCTV, industrial automation) and tailored to specific environment can result in significant performance gain. This could be done through biased dataset which is specific to the application. However, another avenue to explore is the ability to perform online training after deployment. Offline pre-training combined with online retraining has been shown to give superior performance in comparison to pre-trained models.

Finally, one of the major challenges associated with implementing machine learning models for real-world applications is the need for a high computational capability. Thus, the efficiency of the model should also be studied in addition to its performance. Improving efficiency and reducing the computational complexity can be done by improving the implementation and the architecture of the model. Furthermore, recent works have proposed different methodologies to embed learning models into low power portable devices, which would reduce the cost associated to implement these models significantly. While these avenues expand the bounds of this study significantly, they must be still addressed to extend the applications of these proposed models for real-world applications.

References

[1] P. Perera, R. Nallapati, and B. Xiang, "OCGAN: One-class novelty detection using gans with constrained latent representations," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2019-June, pp. 2893–2901, 2019.

[2] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *BMVC*, 2018.

[3] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3291–3300, 2018.

[4] Y. P. Loh and C. S. Chan, "Getting to know low-light images with the exclusively dark dataset," *Computer Vision and Image Understanding*, vol. 178, pp. 30–42, 2018.

[5] E. H. Land, "The Retinex Theory of Color Vision," *Sci. Am.*, vol. 237, no. 6, pp. 108–128, 1977.

[6] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, 2018.

[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, pp. 2672–2680, 2014.

[8] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Computer vision, graphics, and image processing*, vol. 39, no. 3, pp. 355–368, 1987.

[9] M. Tanaka, T. Shibata, and M. Okutomi, "Gradient-based low-light image enhancement," in *2019 IEEE International Conference on Consumer Electronics (ICCE)*, pp. 1–2, IEEE, 2019.

[10] E. H. Land, "Recent advances in retinex theory and some implications for cortical computations: color vision and the natural image.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 80, no. 16, p. 5163, 1983.

[11] W. Wang and C. He, "A variational model with barrier functionals for retinex," *Siam Journal on Imaging Sciences*, vol. 8, no. 3, pp. 1955–1980, 2015.

[12] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, "Structure-revealing low-light image enhancement via robust retinex model," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2828–2841, 2018.

[13] X. Guo, Y. Li, and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," *IEEE Transactions on image processing*, vol. 26, no. 2, pp. 982–993, 2016.

[14] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, "Object recognition with gradient-based learning," in *Shape, contour and grouping in computer vision*, pp. 319–345, Springer, 1999.

[15] D. H. Ballard, "Modular learning in neural networks," in *AAAI*, 1987.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[17] L. Xu, J. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," *Advances in Neural Information Processing Systems*, vol. 2, pp. 1790–1798, 01 2014.

[18] K. G. Lore, A. Akintayo, and S. Sarkar, "LLNet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognition*, vol. 61, pp. 650–662, 2017.

[19] L. Tao, C. Zhu, G. Xiang, Y. Li, H. Jia, and X. Xie, "LLCNN: A convolutional neural network for low-light image enhancement," in *2017 IEEE Visual Communications and Image Processing (VCIP)*, pp. 1–4, IEEE, 2017.

[20] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, and J. Ma, "Msr-net: Low-light image enhancement using deep convolutional network," *arXiv preprint arXiv:1711.02488*, 2017.

[21] Y. Guo, X. Ke, J. Ma, and J. Zhang, "A pipeline neural network for low-light image enhancement," *IEEE Access*, vol. 7, pp. 13737–13744, 2019.

[22] 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*, pp. 2223–2232, 2017.

[23] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "EnlightenGAN: Deep Light Enhancement without Paired Supervision," *Preprint*, 2019.

[24] W. Wang, C. Wei, W. Yang, and J. Liu, "GLADNet: Low-Light Enhancement Network with Global Awareness," *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pp. 751–755, 2018.

[25] F. Lv and F. Lu, "Attention-guided low-light image enhancement," *arXiv preprint arXiv:1908.00682*, 2019.

[26] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[27] 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* (N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds.), vol. 9351, pp. 234–241, Springer International Publishing, 2015.

[28] B. Arad, P. Kurtser, E. Barnea, B. Harel, Y. Edan, and O. Ben-Shahar, "Controlled lighting and illumination-independent target detection for real-time cost-efficient applications. the case study of sweet pepper robotic harvesting," *Sensors*, vol. 19, no. 6, p. 1390, 2019.

[29] W.-C. Kao, M.-C. Hsu, and Y.-Y. Yang, "Local contrast enhancement and adaptive feature extraction for illumination-invariant face recognition," *Pattern Recognition*, vol. 43, no. 5, pp. 1736–1747, 2010.

[30] C. Bai and A. R. Reibman, "Controllable image illumination enhancement with an over-enhancement measure," in *25th IEEE International Conference on Image Processing (ICIP)*, pp. 385–389, IEEE, 2018.

[31] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, (New York, NY, USA), p. 1632–1640, Association for Computing Machinery, 2019.

[32] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *BMVC*, 2018.

[33] Y. P. Loh and C. S. Chan, "Getting to know low-light images with the exclusively dark dataset," *Computer Vision and Image Understanding*, vol. 178, pp. 30–42, 2018.

[34] A. Krizhevsky, V. Nair, and G. Hinton, "The cifar-10 dataset," *online: http://www. cs. toronto. edu/kriz/cifar. html*, vol. 55, 2014.

[35] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *BMVC*, 2018.

[36] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125–1134, 2017.

[37] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "Completely Blind" Image Quality Analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.

[38] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.

[39] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.

[40] H. Hoffmann, "Kernel pca for novelty detection," *Pattern Recognition*, vol. 40, pp. 863–874, 03 2007.

[41] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," *arXiv preprint arXiv:1901.03407*, 2019.

[42] W. Wei, J. Erenrich, and B. Selman, "Towards efficient sampling: Exploiting random walk strategies," in *AAAI*, vol. 4, pp. 670–676, 2004.

[43] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[44] H. Weligampola, G. Jayatilaka, S. Sritharan, R. Godaliyadda, P. Ekanayaka, R. Ragel, and V. Herath, "A retinex based gan pipeline to utilize paired and unpaired datasets for enhancing low light images," 2020.