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

- Semester 7 Report -

Gihan Jayatilaka
Suren Sritharan
Harshana Weligampola

Department of Computer Engineering
University of Peradeniya

CO421 : Final Year Project I report submitted as a requirement of the degree of
B.Sc.Eng. in Computer Engineering

November 2019

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
November 2019

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 around the clock in all environments. Through our studies, we have 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. On conclusion, the paper suggests possible future research avenues.

Table of contents

List of figures vi

List of tables vii

Nomenclature viii

1 Introduction 1

1.1 Background . . . 1

1.1.1 Terminology . . . 1

1.2 The problem . . . 2

1.2.1 Alternative solution . . . 3

1.3 Proposed Solution . . . 3

2 Related work 4

2.1 Schools of thought . . . 4

2.1.1 Intensity-based . . . 4

2.1.2 Gradient-based . . . 4

2.2 Classical image processing for low light image enhancement . . . 5

2.2.1 Retinex theory . . . 5

2.3 Deep learning for low light image enhancement . . . 8

2.4 Retinex aware machine learning solutions . . . 9

2.4.1 Retinex model-based neural networks . . . 9

2.4.2 Log scale NN solutions . . . 10

2.4.3 Wavelet transform . . . 10

3 Methodology 12

3.1 Algorithm testbed . . . 12

3.2 Classical algorithms . . . 13

3.2.1 Variational retinex . . . 14

3.3 DNN models 14
3.3.1 LLCNN 15
3.3.2 LLIEnet 16
3.3.3 RetinexNet 17

4 Experimental Setup and Implementation 18
4.1 Python ecosystem 18
4.1.1 Tensorflow implementation 19
4.1.2 Keras implementation 19
4.2 Dataset 19
4.2.1 Synthetic data 20
4.2.2 Original data 20
4.3 Evaluation Metric 21
4.3.1 PSNR 21
4.3.2 Modified PSNR 22

5 Results and Analysis 23
5.1 Visualization tool 23
5.2 Classical algorithms 23
5.2.1 Intensity and Gradient-based technique 23
5.2.2 Retinex-based techniques 26
5.2.3 Analysis 26
5.3 DNN-based algorithms 28
5.3.1 LLCNN 28
5.3.2 MSRnet 29
5.3.3 LLIEnet pipeline - DWT 33
5.3.4 RetinexNet 33

6 Conclusions and Future Works 36

References 37

List of figures

2.1 Flow chart of the Retinex algorithm 6
2.2 LLCNN convolution block with residual connections 9
2.3 MSRnet architecture 10
2.4 Wavelet transform aware neural network 11

3.1 Decomposition network in RetinexNet 17

5.1 Colour transformations in RGB space 24
5.2 Colour transformations in HSV space 24
5.3 Colour distribution transformations in RGB space 25
5.4 Colour distribution transformations in HSV space 25
5.5 Image enhancement using CLAHE and Gradient-based algorithms 26
5.6 Image enhancement using Variational retinex model and LIME 27
5.7 Variation of the output of the Retinex algorithm with tunable parameters 28
5.8 Results of LLCNN using SSIM loss function 29
5.9 Results of LLCNN using MSE loss function 30
5.10 Results of LLCNN without residual connections using MSE loss function 30
5.11 Results of the MSRnet 31
5.12 Results of the MSRnet. Enhanced A: original MSRnet architecture. Enhanced B: MSRnet with sigmoid activation 32
5.13 Results of the DWT operation.
a. dark image, b. DWT output, c. normal light image. 33
5.14 Results of the Decompose network on LOL dataset 34
5.15 Results of the Decompose network on our dataset 35

List of tables

4.1 Actual datasets 21

Nomenclature

Acronyms / Abbreviations

API	Application programming interface
CLAHE	Contrast Limited Adaptive Histogram Equalization
CNN	Convolutional neural network
DNN	Deep neural network
FFT	Fast Fourier transform
GAN	Generative Adversarial Network
HDR	High dynamic range
HSV	Human vision system
LIME	Low-light Image Enhancement via Illumination Map Estimation
LLCNN	Low light convolutional neural network
LLIEnet	Low light image enhancement network
MSR	Multi scale retinex
NN	Neural network
PDE	Partial Differential Equations
PDE	Partial differential equation
PSNR	Peak signal to noise ratio
QP	Quadratic programming

RNN	Recurrent Neural Network
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.

1.1.1 Terminology

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 The 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 Proposed Solution

An algorithmic approach to low light enhancement is proposed in this work. It is elaborated in detail in the following sections.

Chapter 2

Related work

2.1 Schools of thought

Looking at the evolution of low light enhancement algorithm through history, we observe that there are 2 basic schools of thought.

2.1.1 Intensity-based

The intensity-based techniques consider the intensity of the image pixels and enhance the pixel intensity values to obtain a lighter image. The most common and well known pixel-based technique is the histogram equalization method. This is a contrast stretching technique that increases the spread of intensity values of the image. There are multiple variants of the Histogram equalization technique [1].

2.1.2 Gradient-based

Gradient-based techniques consider the difference of the pixel intensity values as the gradient of the image to enhance the intensity. They enhance low light images by increasing the gradient of the low light regions of the image. These are comparatively newer low light image enhancement techniques which are gaining popularity since the human visual system (HVS) is more sensitive to the gradient that the absolute intensity values. Furthermore, these algorithms can be simultaneously applied with other gradient-based techniques such as edge detection and smoothing, thus resulting in improved efficiency. However, there are drawbacks such as intensity saturation which makes their application challenging.

2.2 Classical image processing for low light image enhancement

Classical image processing algorithms are unsupervised algorithms which enhance low light images through well-formulated mathematical models. As highlighted previously, these algorithms are based on two schools of thought, namely intensity-based enhancement and gradient-based enhancement. These algorithms are computationally efficient, simple and each variant of the technique can be used for a specific purpose. However, these methods only consider a single aspect (intensity or gradient) for light enhancement. However, the retinex based algorithms (discussed in the following section) on the other hand consider multiple aspects and thus performs comparatively well. Even though it is not explicitly stated, retinex based algorithms compromise of both intensities as well as gradient information in their model. Therefore, the literature study on classical algorithms was mainly based on the retinex based solutions.

2.2.1 Retinex theory

The Retinex[1] Theory [2, 3] was a significant step in the field of vision based algorithms. It explains the way, the human vision perceives different colours under different lighting conditions. Through the results of the "Mondarian" experiment [3] it was suggested that the sensory signals from the same colours remain consistent regardless of the lighting condition and thus the colour perceived by the cortex remains constant. This feature of the human visual system (HVS) is known as colour constancy and it aids humans to identify the same colours under different lighting conditions. Furthermore, the retinex theory states that colour constancy occurs due to the ability of the HVS to approximate the composition of different illuminations on an object and remove them to obtain the "true colour" of the object known as the reflectance.

The retinex algorithms thereby developed are based on this theory. The retinex algorithms are based on the assumption 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.

The primary objective of these algorithms is to decompose the image S into the reflectance R and illumination L as $S = R * L$. However, to decompose and obtain the

[1] formed by the amalgamation of words retina and cortex

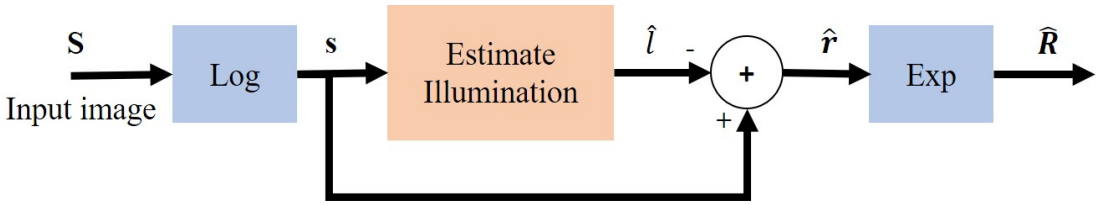

Fig. 2.1 Flow chart of the Retinex algorithm

reflectance, the illumination must be extracted and this is an ill-posed problem. The retinex based algorithms use different techniques to estimate this illumination. Note that the estimation of illumination is a necessary but not compulsory step. The illumination effect can be removed without explicit decomposition of the image.

The initial step of any retinex based algorithm is the same. The problem is first converted to the log domain as $s = \log S$, $l = \log L$, $r = \log R$ and thus $s = l+r$. This step is motivated both mathematically (preferring additions over multiplications), and physiologically, (referring to the sensitivity of our visual system) [3]. The different Retinex algorithms usually have the same flow chart as shown in Figure 2.1, but the method through which the illumination is estimated is different. The following subsections give a brief introduction of the variations of the retinex algorithm.

Since the introduction of retinex theory, many variations of this model have been proposed for the decomposition. Some of these proposed solutions are closer to the initial theory [3–6], whereas some others have made drastic changes but are still based on the same idea of retinex [7–15]. The different Retinex algorithms usually have the same flow chart as shown in Figure 2.1, however the method through which the illumination is estimated is different. The following subsections give a brief introduction of the variations of the retinex algorithm.

Path-based algorithm

The original retinex model uses a path based algorithm. There are many variations of the path based implementation of retinex decomposition including random walks [2], piece-wise linear paths [4], dual spiral pixel paths [5], etc. All these variations are based on multiple agents walking along the pixels to find the illumination eventhough their implementations vary.

Recursive algorithms

Recursive algorithms were developed as an alternative for random walk in [6]. Although it is comparatively efficient, the number of required recursion steps is not fixed and thus convergence is not always guaranteed.

PDE based model

These techniques use differential analysis to analyze the variation of smoothness in an image to estimate the illumination. Based on this analysis many partial differential equation (PDE) based models have been proposed. These models vary on the method of calculation of PDEs such as Laplacian transform [7], discontinuity extraction [8], divergence calculation [9, 10], etc. However, all PDE formulations have the following similarity : the formulation uses the logarithmic form of the original model and it analyzes the gradient with the assumption of smoothness of illumination to extract the reflectance image.

Variational models

The variation model proposed by Kimmel et al. in [11] became the basis for many retinex based solutions. The model is based on the same assumptions as the previous models. However, the authors proposed converted to retinex decomposition problem into a mathematically well-posed constrained optimization problem. Through an accelerated project normalized steepest descent (PNSD) technique and iterative estimation based on gaussian pyramids, the authors arrived at a definite solution for the proposed model.

Future works based[12–15] on this variational model introduced improvements to this model. The authors added additional optimization objectives and constraints to make the model more rigid and well-posed and thus improve the performance of the model.

Illumination map estimation

Illumination map estimation introduced in [16] takes a different path compared to the variational model. Low-light Image Enhancement via Illumination Map Estimation (LIME) is a technique that estimates the illumination maps in each of the RGB colour channels separately based on structural information. Furthermore as opposed to the variational model which mainly focuses on decomposition, LIME proposes additional illumination enhancement steps which further improve the lighting conditions of the image. In terms of classical algorithms, LIME is considered as one of the state-of-the-art work and is used as a benchmark for this study.

2.3 Deep learning for low light image enhancement

Machine learning techniques have experienced an increase in interest during recent times. The popularity of using neural networks for image processing came after breakthrough techniques such as convolution neural networks (CNNs) [17], auto encoder networks [18]. In recent years, neural networks have been proved to perform better than classical algorithms in many image processing tasks [19, 20]. As such the neural networks have been introduced for low light image enhancement problem as well.

LLNet [21] is one of the initial method using deep neural networks to enhance low-light images. The network is a variant of the stacked-sparse denoising autoencoder using a multi-layer perceptron. Authors use a non-linear transformation to generate training data for the proposed neural network. This method is further improved by Chen et.al [22] by building datasets of raw short-exposure low-light images, with corresponding long-exposure reference images and develop a network to learn the enhancement function. But the results perform well only on the constructed datasets.
Recently, convolutional neural network (CNN) achieves impressive progress in several computer vision applications. As to low-level image processing applications, CNN makes several breakthroughs in super-resolution [23], image denoising [24], etc. In these problems, the pixel values in degraded images are around the true values, and the average pixel values almost don't change. But in low light image enhancement problem, the range of the pixel values varies. Also, for low-light image enhancement, texture preserving is more important and the brightness is allowed to be fluctuant around the ground truth. Therefore, the loss function should capture these features in order to converge to a good solution.

These characteristics can be modelled using CNNs and other deep learning techniques to obtain the best performance for image enhancement. Using different filters and residual connections it is possible to combine multi-scale feature maps together to generate enhanced images. These residual connections preserve original features and textures. Also using a loss function that depends on the structural similarity, these textures in the image can be retained. Furthermore, deep learning can be used to merge low level preprocessing steps like denoising into the model using a deeper pipeline. This can replicate some of the classical methods like retinex decomposition in a more generic way. These methods are discussed in the next section.

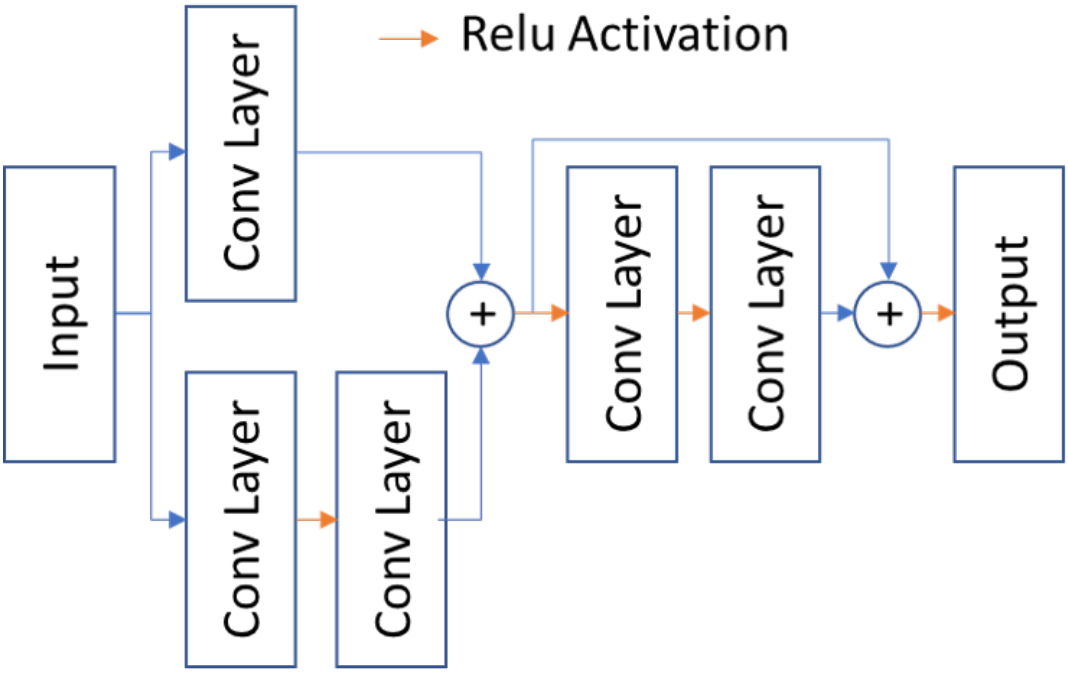

Fig. 2.2 LLCNN convolution block with residual connections

2.4 Retinex aware machine learning solutions

It was evident that both classical techniques and machine learning techniques have different strengths and weaknesses. Therefore an approach was developed to merge both these schools of thought. The work in this domain attempts to merge convolution neural networks [17] with (1) Retinex model (2) Log domain (3)Wavelet transform.

2.4.1 Retinex model-based neural networks

Retinex model's flowchart suggests that the image should undergo decomposition to the reflectance and the illumination (of the same size as the image), each of them undergoes different image processing steps (denoise the reflectance and amplify the illumination) and then be merged into one image.

LLCNN [25] draws motivation from Inception architecture [26] and Resnet [27] architecture. LLCNN uses *convolution blocks* in its pipeline similar to inception. The idea of residual connections from resnet is used to give two separate paths for reflectance and illumination. The 1x1 convolution (which was first put forward in inception architecture) is present in LLCNN. As shown in Figure 2.2 the LLCNN residual connections are shorter than the two pathways proposed in the retinex flowchart. This can be both

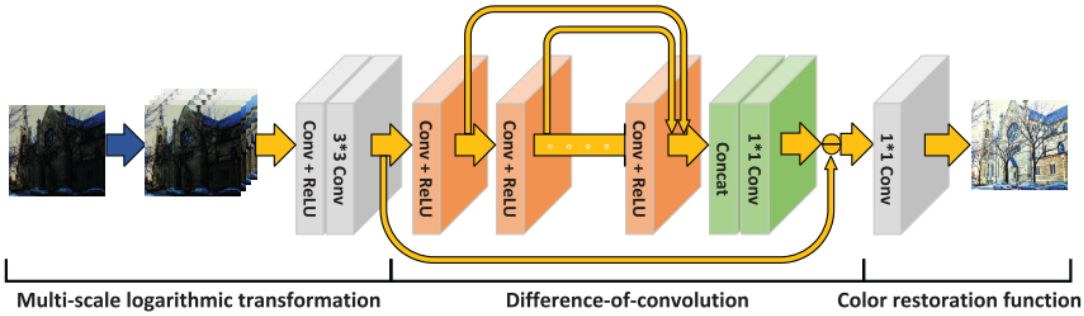

Fig. 2.3 MSRnet architecture

advantageous (able to learn several iterations of Retinex like light enhancement) and disadvantageous (cannot perform a complete Retinex like light enhancement).

2.4.2 Log scale NN solutions

A solely neural network solution cannot mimic the transformation between the RGB scale and the log(RGB) scale. Therefore an explicit log transform was introduced to the NN pipeline starting from MSRnet architecture [28].

MSRnet proposes transforming the RGB scale to multiple ranges in the log scale to create a higher dimensional representation of the image consisting of the RGB scale and all the log scales. This higher dimensional *image* is fed into two pipelines (one a *residual like* connection and another with multiple convolution steps). The results (reduced to 3 colour channels) are merged at the end and undergo a colour restoration operation as shown in Figure 2.3.

2.4.3 Wavelet transform

Wavelet transform (discrete version applied across the spatial domain) is able to extract information out of the image [29]. This idea is used in [30] as another pathway between the input image and the output in the neural network. This allows the neural network to make use of another classical image processing technique to improve its performance as shown in Figure 2.4.

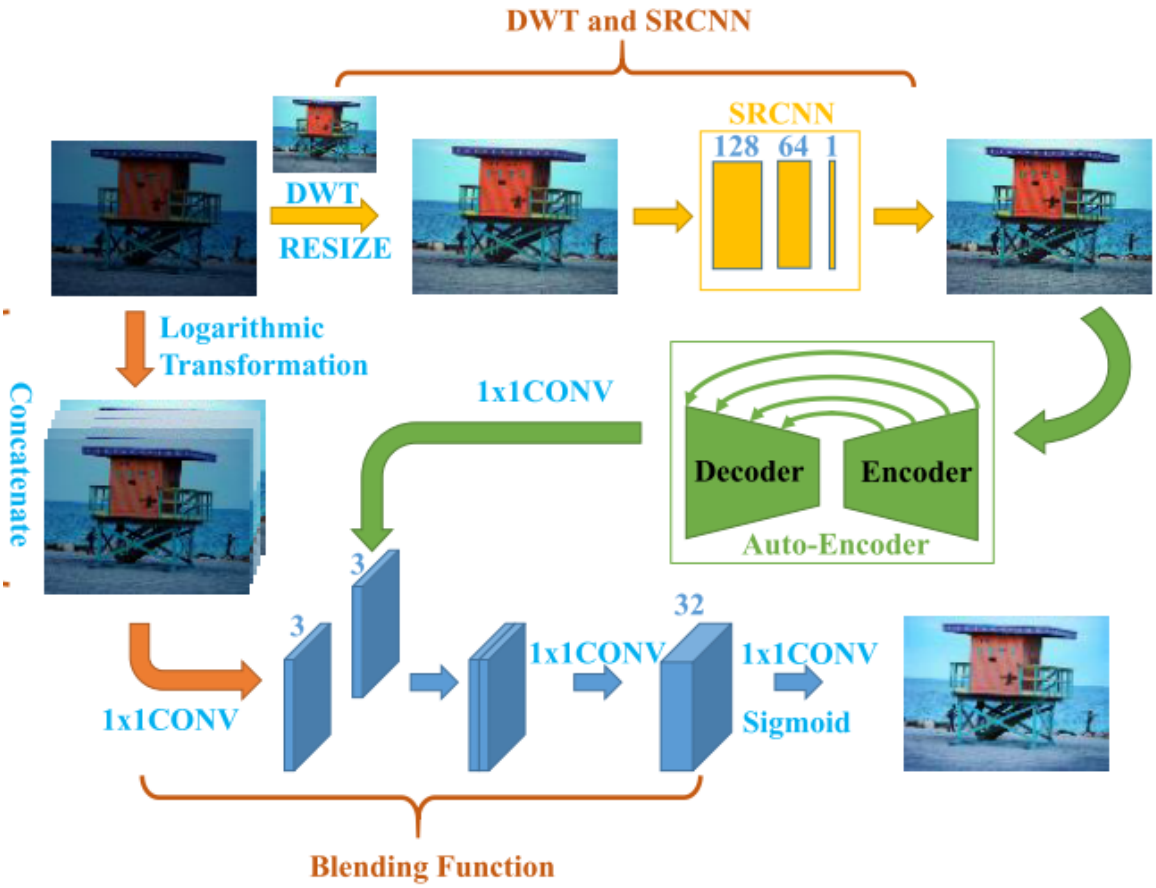

Fig. 2.4 Wavelet transform aware neural network

Chapter 3

Methodology

3.1 Algorithm testbed

The first part of the project is to implement a unified testbed for testing all low light imaging algorithms. This requires the following components.

1. **Image handling**
   This section should be able to load images, crop them to algorithm feedable resolution pieces and stitch the results at the output end.

2. **Dataset synthesis**
   Synthetic datasets are often used in this project. Synthetic data generation requires a generator that can be parameterized. Even though the data generation has a stochastic element (the random noise added, random gamma changes) everything else should be deterministic (range for random noise, range for gamma changes). It is desirable to have a fixed seed for pseudo-randomness so that all tests are repeatable.

3. **Algorithm interface**
   An interface was designed to fit any algorithm (classical and NN) in the following two forms.

   - PythonScript.py : This script should have the following functions.
     - fit()
     - predict()
   - KerasModel.h5

4. **Evaluation**
   This utility is supposed to compare 3 image sets (low light, true normal light and enhanced) and output the evaluation metrics such as
   - SSIM
   - MSE
   - PSNR
   - Modified PSNR.
5. **Visualization**
   This utility is designed to be able to visualize the transformations of the enhancement process. This will take in two or three images from the set (low light, normal light or enhanced) and plot the colour transformation analysis.
   This tool can work in both RGB, HSV colour spaces.

3.2 Classical algorithms

The contrast limited adaptive histogram equalization (CLAHE) [1] and Gradient enhancement algorithm proposed in [31] were analyzed using the testbed. These techniques are computationally efficient and provide real-time performance. However, the efficacy of these algorithms was lower compared to the retinex based models. Therefore, more concern was given to retinex based algorithms and its variants. Although there are multiple variations of retinex-based low-light image enhancement algorithms, we focus our attention on [11] and [16] which are two of the landmark researches done in this field.

The experiments were designed to answer the following questions in regard to classical algorithms:

1. What is the basis of the algorithm (theory, model) ?
2. What are the shortcomings of the solution?
3. Can the problems be solved?
4. Can the model be improved?

3.2.1 Variational retinex

Introduction

The variational retinex model proposed in [11] deals with the compensation for illumination effect in images. The authors propose a well-formulated mathematical model for image decomposition. The specific solution does not deal with low light image enhancement. However, the decomposed reflectance can be considered to be the "true colour" of the image. Furthermore, gamma correction is applied to improve the aesthetics of output reflectance.

The retinex model is based on the following assumptions:

- An image S can be decomposed into the reflectance R and illumination L.
- The illumination image L is spatially smooth.
- The reflectance image R is piece-wise continuous (Except at the object edges).

These proposed model combines these assumptions together with other optimization objectives such as the visually pleasing nature, to generate a constrained optimization function, and then is solved to obtain the reflectance image. Since its introduction, many other improvements have been proposed including the works in [12–15]. These models incorporate additional objective functions to improve the efficacy of the solution. However, the basic assumptions of the model remain the same in all these studies.

Experiments

Experiments were designed to answer the following additions questions related to this specific case study.

1. Do the assumptions act as limitations?
2. What are the effects of varying the parameters of each optimization objective?
3. Could any additional objective functions be included to improve the efficacy?

3.3 DNN models

Machine learning techniques have experienced an increase in interest during recent times. The popularity of using neural networks for image processing came after breakthrough techniques such as convolution neural networks (CNNs) [17], autoencoder networks

[18]. In recent years, neural networks have been proved to perform better than classical algorithms in many image processing tasks [19, 20]. As such the neural networks have been introduced for low light image enhancement problem as well.

3.3.1 LLCNN

Introduction

LLCNN uses *convolution blocks* in its pipeline similar to inception. The idea of residual connections from resnet is used to give two separate paths for reflectance and illumination.

Loss function

In [25] authors propose to use the Structural Similarity (SSIM) given in (3.1) as the loss function for the DNN. They claim that this gives the best results compared with other approaches such as [16], [32]. We studied the effect of the loss function by changing the loss function and observing the results. We used Mean Square Error (MSE) as the loss function for our comparison.

$$SSMI(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{3.1}$$

where μ_x average of x, μ_y average of y, σ_x^2 variance of x, σ_y^2 variance of y, σ_{xy} co-variance of x and y, c_1 and c_2 are two variables to stabilize the division with weak denominator.

Effect of residual connections

Authors of LLCNN [25] claim that the residual connections will learn to mimic the retinex model due to two separate paths in the model architecture. Thus, decomposing the image into reflectance and illumination in the neural network itself. Also, they claim that serial convolutions blocks reflect the Multi-Scale Retinex (MSR) model. These claims were tested by varying the number of convolutional blocks and residual connections in each block. Our objective was to find the effect of the residual connections on the training and the final result. First, we removed all the residual connections from the convolution block and trained the LLCNN. Stochastic Gradient Descent (SGD) was used as the optimizer and SSMI (3.1) is used as the loss function.

3.3.2 LLIEnet

Introduction

LLIEnet (low light image enhancement network) is a pipeline proposed by drawing motivation from MSRnet [33]. This pipeline contains the following components.

- Discrete wavelet transform from [34]
- autoencoders as proposed in [35]
- Super-resolution from [36]
- Multiscale logarithmic transforms from [28]
- Blending functions from [28]

The methodology was designed to answer the following questions.

1. Is discrete wavelet transform doing any image enhancement?
2. Is the discrete wavelet transform contributing to the light enhancement done by the next layers in any way?
3. What is the rise of the number of parameters in LLIEnet compared to MSRnet?
4. Is the performance increment of LLIEnet totally due to the parameter number increase (raw computation) or because of a clever architecture?

The experiments are as follows.

1. Compared the low light image, DWT (resized) output and the normal light image for MSE, SSIM and the proposed metric.
2. Replace the DWT branch by a downsized low light image and compare the NN performance.
3. Calculation
4. Compare the results of MSRnet and LLIEnet. Try to justify the resulting improvement against the increase of trainable parameters.

3.3.3 RetinexNet

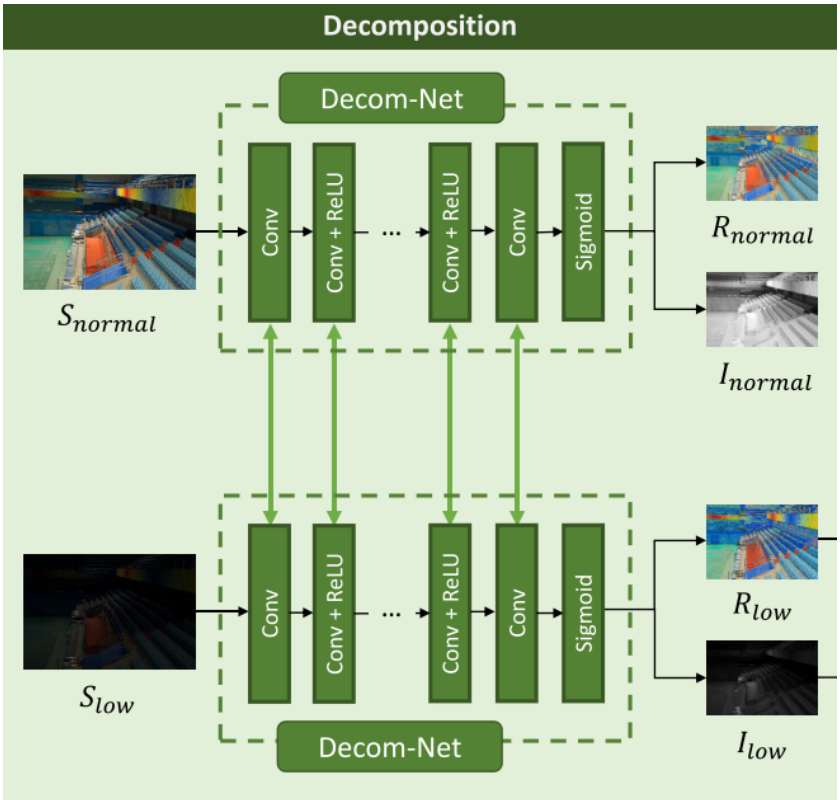

Fig. 3.1 Decomposition network in RetinexNet

As mentioned earlier the Retinex model is a strong idea that revolutionized the way we look at images. Authors of the RetinexNet [37] propose a method to decompose an image into reflectance and illumination without labelled data from classical/other algorithms. Most existing Retinex-based methods have carefully designed hand-crafted constraints and parameters for this highly ill-posed decomposition, which may be limited by model capacity when applied in various scenes. Also, the authors use another neural network RelightNet after the decomposition to enhance the image using the decomposed reflectance and illumination. Our main focus is to study the decomposition of the image and use it to devise our own approach for enhancement. Thus, we analyze the decomposition network which is illustrated in Figure 3.1.

Chapter 4

Experimental Setup and Implementation

4.1 Python ecosystem

Python was chosen as the platform of implementation for the project due to the following reasons,

- Ease of scripting in python.
- Availability of image handling libraries for python.
- Availability of symbolic mathematics libraries
- Availability of GPU support for accelerated mathematical operations.
- Availability of high-level libraries for neural network implementation.
- Most of these tools being free.

Classical algorithms can be implemented using mathematical packages such as scipy, numpy, etc. However, it is easy to comprehend and use if more high level libraries are used for machine learning algorithms. Python has multiple machine learning libraries which can be used for the implementation of machine learning algorithms. Out of these, Tensor flow and Keras were chosen for implementation since these packages provide a good balance between ease of usage and customizability.

4.1.1 Tensorflow implementation

Tensorflow is a symbolic mathematics toolkit that can create a "flow" (a procedure for manipulation) for "tensors" (high dimensional matrices). This "flow" is represented as a computational graph and then it is run by feeding in data for tensors. The package has tensor operations of a large spectrum (from basic arithmetic to complicated optimizers).

Tensorflow was chosen due to the following reasons.

1. Ease of implementing and tweaking algorithms on a higher level with Python interpreted programming language.
2. Ability to harness the efficiency of compiled and statically typed C++ code for computation.
3. The GPU acceleration is supported on the Computer Engineering Department servers.

4.1.2 Keras implementation

Keras is a higher level deep learning API, which can run on top of Tensorflow, Theano, or CNTK. It was developed with the intention of easier and faster prototyping and thus can implement neural networks such as DNN, CNN and RNN out of the box. Keras was implemented with Tensorflow backend in this study. Keras was chosen as a supplementary tool since it enables faster experimentation and runs efficiently on both the CPU and GPU due to its optimized implementation. However, it must be noted that Tensorflow allows more customization and therefore must not be disregarded.

4.2 Dataset

The datasets consist of light and dark images for both training and evaluation. There are 2 ways of creating these images. The light and dark images can be either generated using image processing algorithms such as gamma correction [38] or, they could be obtained under different lighting conditions and different exposures. The following section explains specific techniques used for data generation and collection.

4.2.1 Synthetic data

A large amount of data is required to train the neural networks because of the large number of trainable parameters in the NN. It is practically impossible to get enough actual training data to train a NN.

We have identified that a NN should learn a few different operations in the full light enhancement process.

1. Amplify the brightness.
2. Amplify the contrast.
3. Denoise.
4. Fill colours to places where colour is not seen.
5. Some additional operations that cannot be explicitly named.

Synthetic data can be created to emulate 1, 2 by gamma adjustments and 3 by adding random noise. Other processes require actual data for the NN to learn from. Gamma adjustment is defined by (4.1)

$$\bar{v} = \left(\frac{v}{255}\right)^{\frac{1}{\gamma}} \times 255 \tag{4.1}$$

where v is the original pixel value, $\bar{v}$ is the gamma encoded value and γ is the gamma value. The drawback of using gamma adjustment is that this value can be decoded using the inverse function of (4.1) given by

$$v = \left(\frac{\bar{v}}{255}\right)^{\gamma} \times 255 \tag{4.2}$$

Thus if synthetic data was used to train, the DNN may simply learn the inverse function (4.2) instead of the actual relationship between well-lit images and low light images. Thus the synthetic data was only used for initial experimental purposes and was not used for further analysis.

4.2.2 Original data

Original data refers to data acquired by changing the lighting condition or the exposure duration. We work on the assumption of both these operations having the same effect on an image.

The algorithms are tested with several low light image enhancement datasets as given in Table 4.1.

Dataset	No image pairs	Resolution
LOL[1]	485	600×400
Brightening train[2]	1000	384×384

Table 4.1 Actual datasets

4.3 Evaluation Metric

The evaluation metrics used in the study are as follows.

- Mean squared error
- Structural similarity
- Peak signal to noise ratio
- Mean window peak signal to noise ratio.

MSE and SSIM are explained in the previous sections. PSNR and Modified PSNR will be explained in this section.

4.3.1 PSNR

Peak signal-to-noise ratio, (PSNR), is the ratio between the maximum power of a signal and the power of noise. This metric is used in our study by taking the difference between the predicted image and the true normal light image as the noise.

$$PSNR_{Image} = 10 \log_{10} \left(\frac{max(Intensity)^2_{image}}{MSE_{image}} \right)$$

The metric is not suitable for the performance assessment of low light image enhancement algorithms because the "maximum intensity" of the image is not a relevant metric on a grid-like signal.

[1]Obtained from [37]
[2]Obtained from [39]

4.3.2 Modified PSNR

A modified version of PSNR is proposed. This metric is assessing the PSNR on different windows of the image. And the maximum intensity is replaced by the range of intensities inside the window. The final answer is obtained by averaging the values across all windows.

$$PSNR_{Image} = 10\log_{10}\left(\sum_{win \in Image} \frac{(max(Intensity)_{win} - min(Intensity)_{window})^2}{MSE_{window}}\right)$$

We speculate that this metric matches the human perception of the quality of a low light image enhancement than the PSNR. Confirming this speculation requires a user study. This user study will be conducted at a later point in time.

Chapter 5

Results and Analysis

5.1 Visualization tool

The colour transformation between the low light images (dark images) and the well-light images (high light images, true images) was observed by the tool. The analysis was done on both the RGB colour space and the HSV colour space as shown in figure Figure 5.1, Figure 5.2, Figure 5.3, and Figure 5.4. The following trends were observed in the visualization.

- Figure 5.1 ⇒ The light enhancement for all 3 colour channels R,G,B should be linear processes for the very low brightness pixels. The relationship is a function of individual pixels.
- Figure 5.1 ⇒ The light enhancement is a non-linear transformation. The relationship is a function of individual pixels and the neighbouring pixels.
- Figure 5.2 and Figure 5.4 ⇒ The major contribution for the light enhancement happens in the S channel.
- Figure 5.2 and Figure 5.4 ⇒ The least contribution for the light enhancement happens in the H channel.

5.2 Classical algorithms

5.2.1 Intensity and Gradient-based technique

The experimental study was initially done for the CLAHE and the gradient enhancement algorithm. Figure 5.5 shows the output of these algorithms for two samples. While

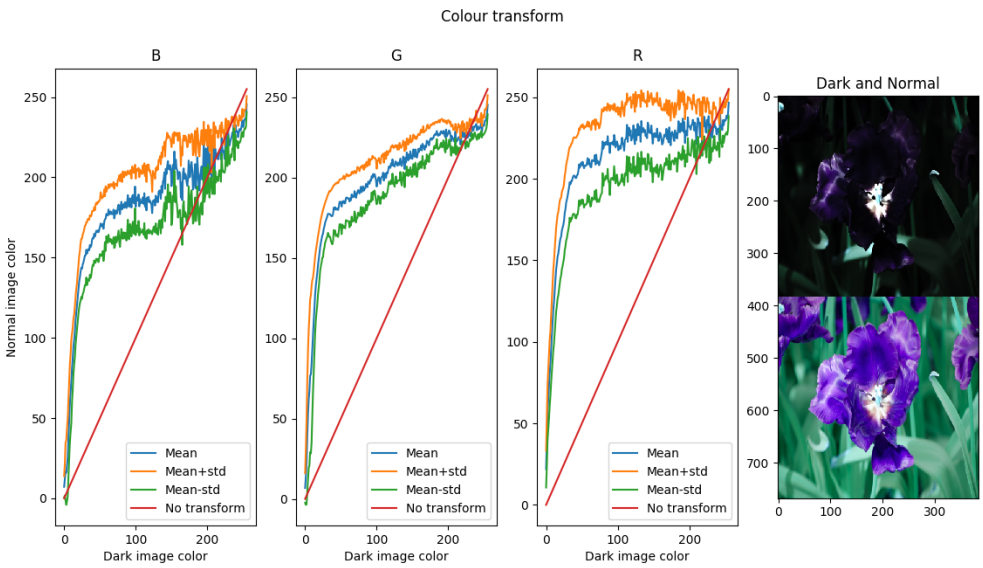

Fig. 5.1 Colour transformations in RGB space

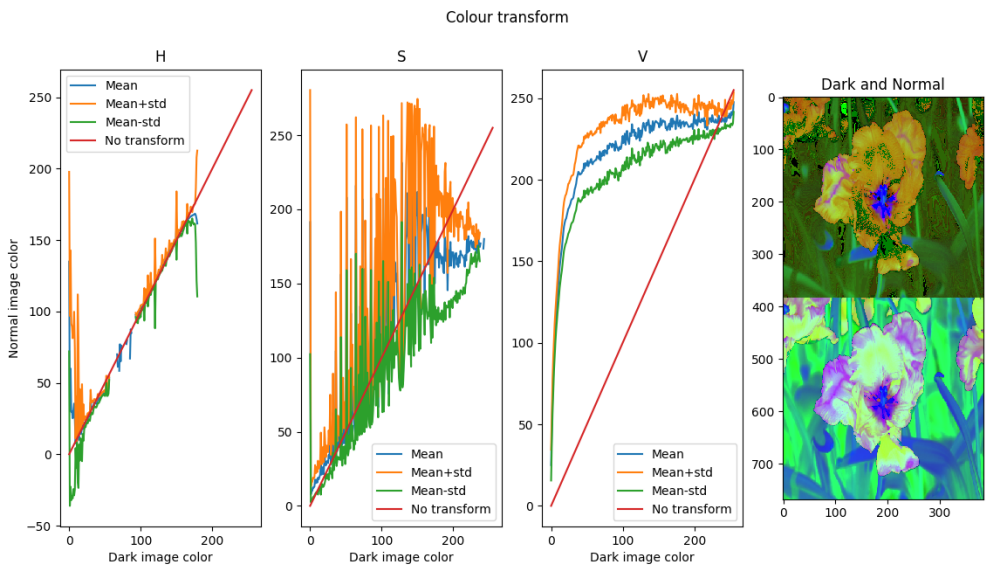

Fig. 5.2 Colour transformations in HSV space

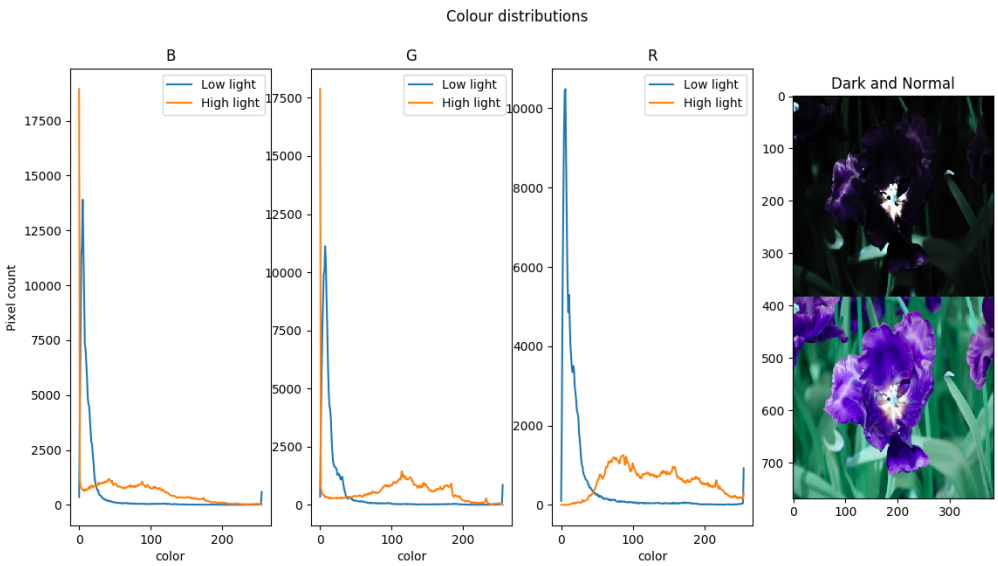

Fig. 5.3 Colour distribution transformations in RGB space

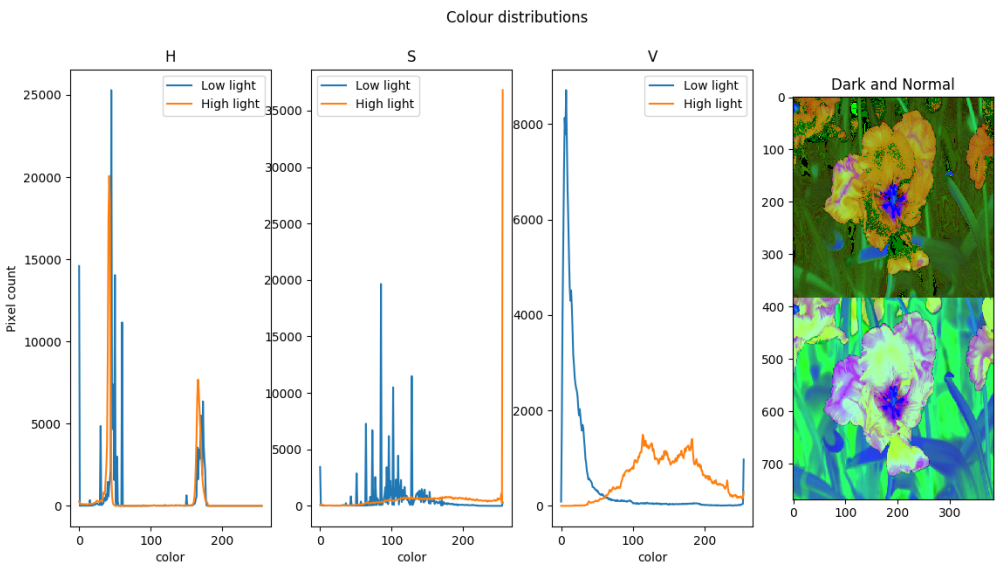

Fig. 5.4 Colour distribution transformations in HSV space

(a) Low light image (b) CLAHE (c) Gradient (d) Well-lit image

Fig. 5.5 Image enhancement using CLAHE and Gradient-based algorithms

gradient enhancement performs better for the first sample, CLAHE gave better results for the second. Thus, the major downside of these algorithms is that they are not robust enough to perform well in different environments. The parameters of these algorithms must be adjusted manually to suit specific environments. These limitations make the algorithms unsuitable across different dynamic environments.

5.2.2 Retinex-based techniques

The results from the variational retinex model [11] and LIME [16] are shown in Figure 5.6. We observe, that these algorithms do not provide perfect results, they show significant improvement over the previous techniques. Furthermore, it can be observed that the variational retinex outputs generally form blurry images, whereas LIME results in over-saturation. The reasons for these results are analyzed in the following section

5.2.3 Analysis

It can be observed that the retinex-based model shows superior performance compared to gradient enhancement and histogram equalization in general. This can be attributed to the following reasons.

(a) Low light image (b) Variational Retinex (c) LIME (d) Well-lit image

Fig. 5.6 Image enhancement using Variational retinex model and LIME

- Retinex-based models consider both the intensity and gradient information to enhance the images.
- The retinex model is mathematically well formulated and specific to the problem of image enhancement.
- The problem-specific nature of retinex algorithms enables parameters to be adjusted such that an optimal result can be obtained based on the environment.

However, one of the major downsides of retinex-based models is parameter tuning. Although the model can be custom-made for a specific problem, the parameter of the model must be adjusted specifically to each image. Figure 5.7 shows the variation of the performance of the variational retinex model with 2 parameters which represent the pixel enhancement and luminance enhancement. These parameters are dependant on the pixel density and the average illumination of the original image and thus must be adjusted specifically to the given input image. Failure to do so will result in an image with poor aesthetics / lighting or over-saturated images.

Furthermore, the low efficacy of the retniex-based algorithms can be attributed to the assumptions of the retinex model. These act as limitations in cases where these assumptions do not hold (e.g. : The spatial smoothness of illumination may not be conserved in the presence of multiple lighting sources with different colours).

Fig. 5.7 Variation of the output of the Retinex algorithm with tunable parameters

Apart from these, retinex models are also computationally inefficient. This occurs usually due to the complexity of the model, and the inability to accelerate / speed up the algorithm and it's implementation. To solve this issue a better model needs to be produced. However, another significant reason is the high number of features that have to be considered in classical algorithms. Feature extraction/ reduction are usually not addressed in classical algorithms and this could lead to high inefficiencies in the presence of larger images or when considering multiple frames from a video.

To address these issues a learning-based model could be a suitable alternative since supervised learning model have the ability to recognize patterns in the dataset through training. Therefore in the next few section we analyze the exiting DNN-based solutions and how classical models and theories could be incorporated into them to design context aware learning models.

5.3 DNN-based algorithms

5.3.1 LLCNN

The results of the LLCNN model with SSIM loss function is shown in Figure 5.8. Even though SSIM preserve the structural information of the image, we can see some parts of the image are over exposed. Also, some colour ranges are over-saturated. As shown in Figure 5.8 even if the background is closer to the normal image the skier is not well constructed. This is also seen in the picture with horses in Figure 5.8.

Fig. 5.8 Results of LLCNN using SSIM loss function

Then we use the MSE loss function to estimate the well-lit image using low lit image. The results obtained are shown in Figure 5.9. It shows much better clear picture than Figure 5.8. This is because the MSE calculates the loss using pixel-wise comparison. But it is difficult to jump to conclusions because it might as well overfit to the dataset. But it is clear that the structure of the image is not well reconstructed when using MSE. Because there exist a blurriness in the predicted images in Figure 5.9.

Here onward we will use MSE as the loss function unless otherwise stated. Even though authors claim that the residual connections (two parallel paths of layered connections) act as a decomposition technique that will decompose image into reflectance and illumination, we found out that the predicted image is much similar to the LLCNN with residual connections. The results from the LLCNN without residual connections is shown in Figure 5.10. But the convergence time significantly increased from 10 epochs to 100 epochs when we remove the residual connections.

5.3.2 MSRnet

The results obtained by MSRnet showed good structural information enhancement. But they suffered from an error of colouring as seen in Figure 5.11. These results were obtained by exact replication of the original paper [28].

Fig. 5.9 Results of LLCNN using MSE loss function

Fig. 5.10 Results of LLCNN without residual connections using MSE loss function

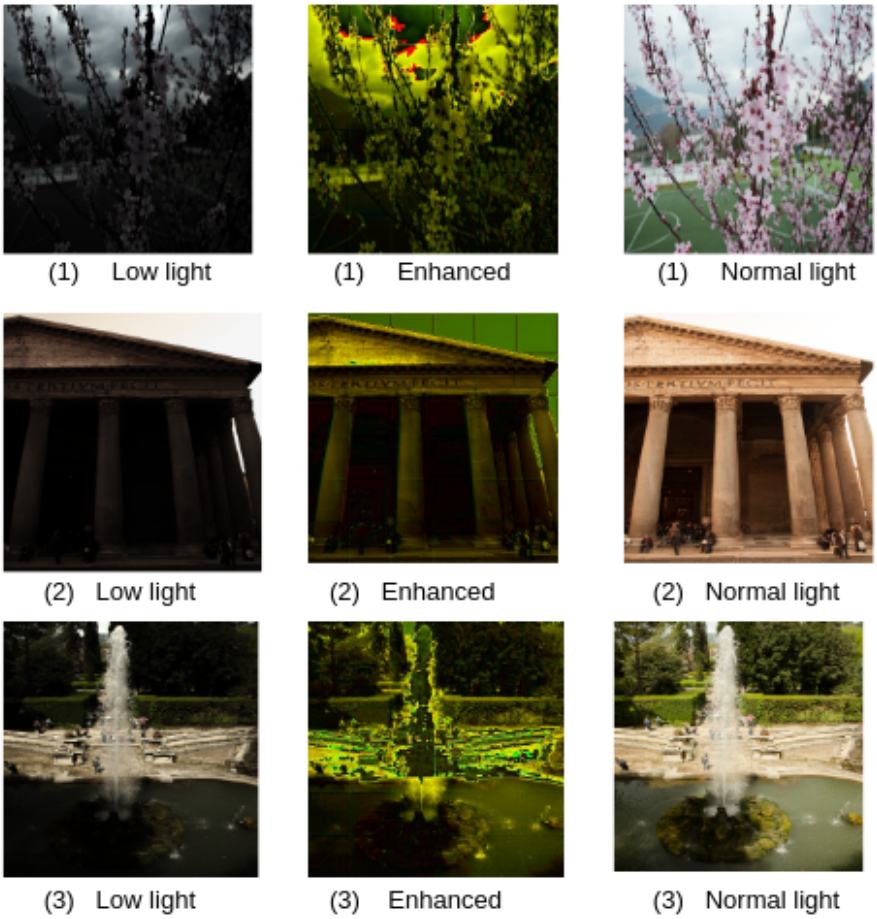

Fig. 5.11 Results of the MSRnet

It is understood that colour correction happens in the last layers of a neural network [25]. It was observed that the MSRnet architecture has no non-linear function in the last few layers. This reduces the activity of a large part of the neural network to a single linear operation.

Better results were obtained by introducing a non-linear activation to the output layer of the NN. The best results (shown in Figure 5.12) were obtained with the sigmoid activation.

Fig. 5.12 Results of the MSRnet. Enhanced A: original MSRnet architecture. Enhanced B: MSRnet with sigmoid activation

5.3.3 LLIEnet pipeline - DWT

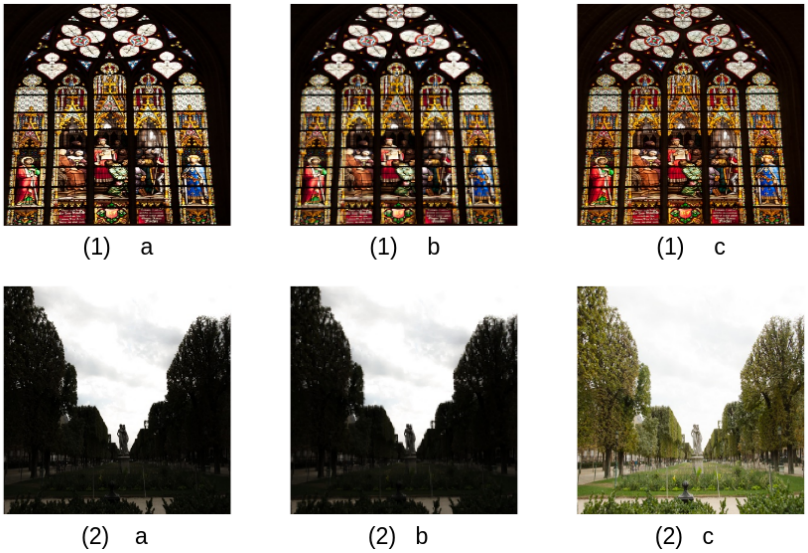

Fig. 5.13 Results of the DWT operation.
a. dark image, b. DWT output, c. normal light image.

It is evident that DWT has no significant brightness improvement. The DWT output looks equal to a down sampled version of the original image as per the results given in Figure 5.13.

It is also observed that LLIEnet architecture has several times the trainable parameters as MSRnet. The performance increment of the LLIEnet over MSRnet could be easily attributed to a large number of trainable parameter rather than any improvement of architecture.

5.3.4 RetinexNet

We experimented with different types of DNN to estimate the enhanced image. But they did not directly tackle the retinex problem. We use the decompose network of this RetinexNet and observe the behaviour with different sets of images. This decomposition network gave the best results so far. Once the image is decomposed into reflectance and illumination, the problem of image enhancement is significantly simplified. Thus, we focused more on decompose network.

We observed the decomposition in LOL dataset Figure 5.14 as well as our own dataset Figure 5.15. They show good results when compared with other classical techniques.

One significant result obtained by experimenting with the decompose network is that the even if we do not share the weights in two parallel networks, the model manages

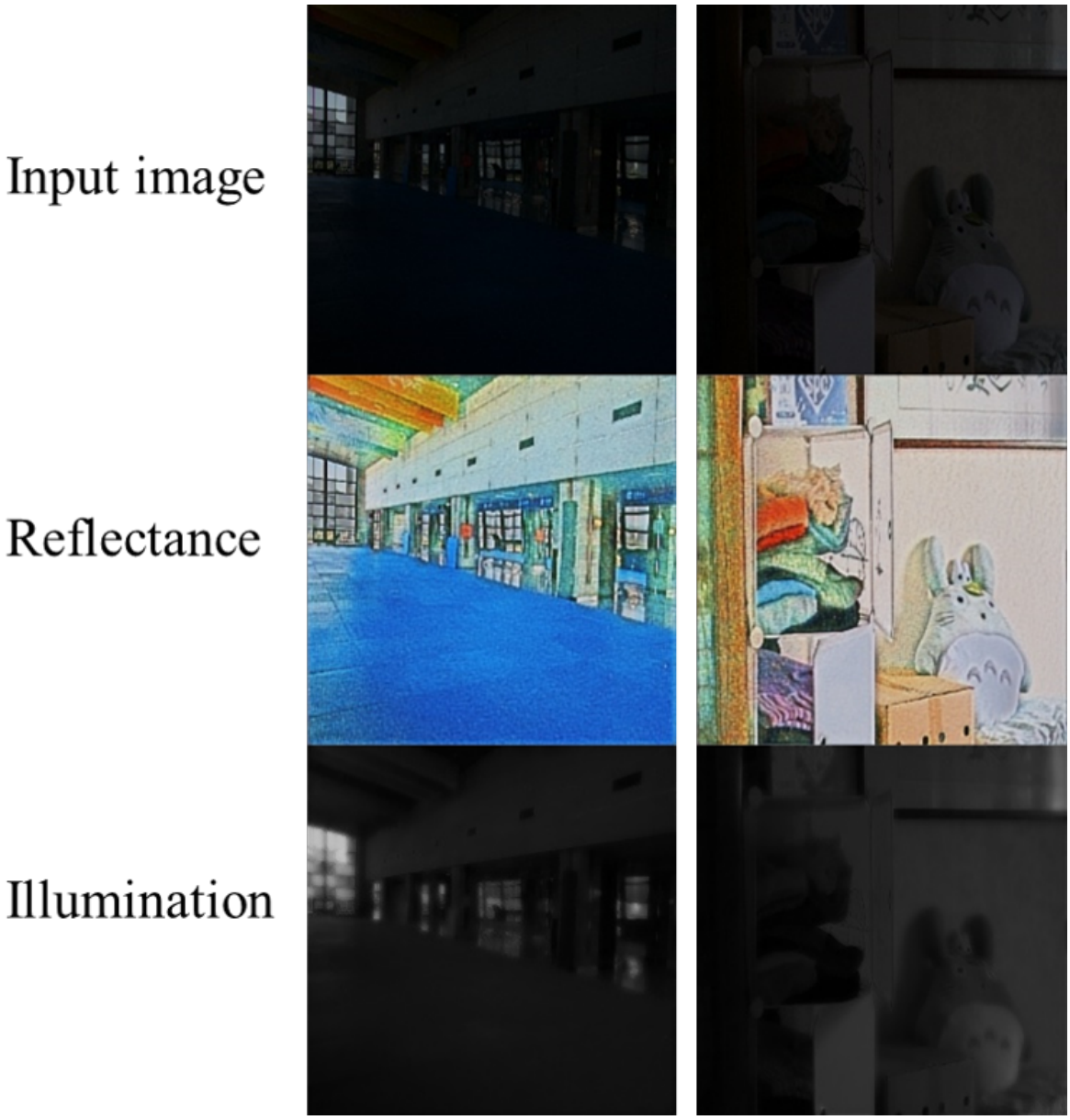

Fig. 5.14 Results of the Decompose network on LOL dataset

Fig. 5.15 Results of the Decompose network on our dataset

to converge to the solution obtained when the two parallel networks are trained while sharing the weights. We expected that the two networks to be identical, but surprisingly they were not. We are planning to further investigate this phenomenon.

Chapter 6

Conclusions and Future Works

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.

Although classical retinex-based algorithm perform well, the assumptions made act as a limitation in specific cases. Furthermore, the inability to automate feature selection and parameter tuning are two of the major downsides of these algorithms. DNN based solutions are lucrative due to the rise in their interest. However, a context unaware model may not provide good performance. From our observations we can conclude that context-aware, retinex model based solutions show significant improvement in performance compared to the alternatives.

However, the existing solutions have specific limitation. For example, the SSIM loss function is not robust enough to handle dynamic resolution image datasets. Thus, it is limited to a given resolution. This can be improved by using convolution layers to estimate SSIM. Then, we can calculate the loss without constraining on resolution.

The proposed improvement have improved the performance of 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. 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.

References

[1] 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.

[2] E. H. Land and J. J. McCann, "Lightness and retinex theory," *Josa*, vol. 61, no. 1, pp. 1–11, 1971.

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

[4] 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.

[5] T. J. Cooper and F. A. Baqai, "Analysis and extensions of the frankle-mccann retinex algorithm," *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 85–93, 2004.

[6] E. H. Land, "An alternative technique for the computation of the designator in the retinex theory of color vision," *Proceedings of the national academy of sciences*, vol. 83, no. 10, pp. 3078–3080, 1986.

[7] B. K. Horn, "Determining lightness from an image," *Computer graphics and image processing*, vol. 3, no. 4, pp. 277–299, 1974.

[8] A. Blake, "Boundary conditions for lightness computation in mondrian world," *Computer vision, graphics, and image processing*, vol. 32, no. 3, pp. 314–327, 1985.

[9] J.-M. Morel, A. B. Petro, and C. Sbert, "Fast implementation of color constancy algorithms," in *Color Imaging XIV: Displaying, Processing, Hardcopy, and Applications*, vol. 7241, p. 724106, International Society for Optics and Photonics, 2009.

[10] J. M. Morel, A. B. Petro, and C. Sbert, "A pde formalization of retinex theory," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2825–2837, 2010.

[11] R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel, "A variational framework for retinex," *International Journal of computer vision*, vol. 52, no. 1, pp. 7–23, 2003.

[12] M. Elad, "Retinex by two bilateral filters," in *International Conference on Scale-Space Theories in Computer Vision*, pp. 217–229, Springer, 2005.

[13] M. K. Ng and W. Wang, "A total variation model for retinex," *SIAM Journal on Imaging Sciences*, vol. 4, no. 1, pp. 345–365, 2011.

[14] W. Ma and S. Osher, "A tv bregman iterative model of retinex theory," *Ucla Cam Report*, vol. 10, p. 13, 2010.

[15] 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.

[16] X. Guo, Y. Li, and H. Ling, "Lime: Low-light image enhancement via illumination map estimation," *IEEE Transactions on Image Processing*, vol. 26, pp. 982–993, Feb 2017.

[17] 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.

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

[19] 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.

[20] 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.

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

[22] 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.

[23] J. K. L. Jiwon Kim and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1646–1654, 2016.

[24] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, pp. 3142–3155, July 2017.

[25] 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, 12 2017.

[26] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[28] 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.

[29] X. Qing, X. Shuai, T. Bing, L. Jiansheng, and G. Zexun, "Complex wavelets and its application to image fusion," in *20th ISPRS Congress Technical Commission III*, vol. 35, pp. 1–6, 2004.

[30] 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.

[31] 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.

[32] M. Abdullah-Al-Wadud, M. H. Kabir, M. A. Akber Dewan, and O. Chae, "A dynamic histogram equalization for image contrast enhancement," *IEEE Transactions on Consumer Electronics*, vol. 53, pp. 593–600, May 2007.

[33] 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.

[34] H. Demirel and G. Anbarjafari, "IMAGE resolution enhancement by using discrete and stationary wavelet decomposition," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1458–1460, 2011.

[35] B. Du, S. S. Member, W. Xiong, S. S. Member, J. Wu, L. Zhang, L. Zhang, and D. Tao, "Stacked Convolutional Denoising Auto-Encoders for Feature Representation," *IEEE Transactions on Cybernetics*, vol. 47, no. 4, pp. 1017–1027, 2017.

[36] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8692 LNCS, no. PART 4, pp. 184–199, 2014.

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

[38] C. Poynton, *Digital Video and HDTV Algorithms and Interfaces*. Morgan Kaufmann Publishers Inc., 1 ed., 2003.

[39] W. Wang, C. Wei, W. Yang, and J. Liu, "Gladnet: Low-light enhancement network with global awareness," in *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, pp. 751–755, May 2018.