HDR-Like Image Generation to Mitigate Adverse Wound Illumination Using Deep Bi-directional Retinex and Exposure Fusion

A Master's Thesis

by

Songlin Hou

Department of Computer Science

Worcester Polytechnic Institute

Advisor: Professor Emmanuel Agu ______________________________

Co-Advisor: Dr. Clifford Lindsay______________________________

Reader: Professor Michael Gennert ______________________________

Dept. Head: Professor Craig Wills ______________________________

Abstract

Periodic assessment is necessary to monitor the healing progress of chronic wounds. Image analysis using computer vision algorithms has recently emerged as a viable alternative as demonstrated by prior work. However, the performances of such image analysis methods degrade on images captured in adverse illumination that occurs in many indoor environments. To mitigate these lighting problems, High Dynamic Range (HDR) image enhancement techniques can be used to mitigate over- and under-exposure issues and preserve the details of scenes captured in non-ideal illumination. In this paper, we address over- and under-exposure simultaneously using a deep learning-based bi-directional illumination enhancement network based on the Retinex theory. Over- and under-exposed images are generated, which are then fused into a final image with enhanced illumination in an exposure fusion step. Our proposed method outperformed the state-of-the-art on various metrics including structure similarity, peak signal-noise ratio and changes in segmentation accuracy (SSIM scores $0.76 \pm 0.04/0.69 \pm 0.08$ on bright/dark images, PSNR scores 28.60 ± 0.70 on dark images and DSC scores $0.76 \pm 0.09/0.74 \pm 0.09$ on bright/dark images).

Acknowledgements

I started my research project at WPI in the summer of 2020, when most of the on-campus work turned online. Without in-person activities, getting into a research project became a rather big challenge to me. Even now when I look back, my research journey cannot be considered as smooth by any standards: spending months searching for ideas, constantly getting stuck on experiments, staying up numerous nights to record and compare experimental results. It has been a circle that I have repeated for almost one year in order to find a workable solution. Being a teaching assistant and taking classes at the same time made things even harder–I had to learn to find the right balance between research, teaching tasks, and coursework. All these experiences pushed me to learn how to better manage my time.

While the research journey has not been easy, I never expected to receive so much help from my advisor Professor Emmanuel Agu and co-advisor Professor Clifford Lindsay. Professor Emmanuel gave me tremendous support during my graduate research and I always got motivated by him when I ever felt down. Professor Clifford offered me lots of invaluable advice and is always patient to answer any questions. So, above all, I would like to thank Professor Emmanuel Agu and Professor Clifford Lindsay for everything they have done. It would not have been possible to generate ideas, finish all experiments or write this thesis without their generous help and support.

I am also grateful to Professor Michael Gennert for being my thesis reader. Further, I would like to thank professors and fellow researchers at the WPI Smartphone Wound Analysis and Decision Support (SmartWAnDS) research group for their feedback and help on my thesis. I would also like to acknowledge the Turinig High-Performance Cluster provided by the Academic Research Computing Group at Worcester Polytechnic Institute. Last, I would like to thank my friends at WPI

for their support and love. Without doubts, I am always indebted to my family for their unconditional love and support to make me who I am and give me the dreams that I can pursue.

Contents

1 Introduction 1

1.1 Motivation . . . 1

1.2 Challenges . . . 3

1.3 Related Work . . . 5

1.4 Our Approach . . . 6

1.5 Novelty and Contributions . . . 8

1.6 Thesis Outline . . . 9

2 Background 11

2.1 Smartphone-Based Wound Assessment . . . 11

2.2 Image Illumination Enhancement . . . 12

2.2.1 Histogram-Based Methods . . . 12

2.2.2 Frequency-Based Methods . . . 13

2.2.3 Retinex-Based Methods . . . 13

2.3 LDR and HDR . . . 14

2.4 HDR Image Generation . . . 16

2.5 Tone Mapping . . . 18

2.6 Image Fusion . . . 19

2.7 Retinex Theory . . . 19

2.8 Image Segmentation . . . 20

3 Related Work **22**

3.1 Direct Enhancement . . . 22

3.2 Multi-Exposure Fusion . . . 23

3.3 Inverse Tone Mapping . . . 24

4 Bracket Image Generation **26**

4.1 Assumptions in Decomposition . . . 27

4.2 Optimization Formation . . . 28

4.3 Construction of Illumination Weight Matrix . . . 30

4.4 Data-Driven Approach to Solve the Optimization Problem . . . 31

4.4.1 Training of DecomNet . . . 33

4.4.2 Evaluation of DecomNet . . . 35

4.5 Illumination Adjustment . . . 35

4.5.1 Training of SynthNet . . . 37

4.5.2 Evaluation of SynthNet . . . 37

4.6 Bi-Directional DRM Neural Network . . . 38

4.6.1 Network Design . . . 39

4.6.2 Discussion . . . 41

5 Exposure Fusion **42**

5.1 Illumination-Based Fusion . . . 43

5.2 Direct Fusion . . . 46

5.2.1 Fusion Concept . . . 46

5.2.2 Limitations of Direct Fusion . . . 47

5.3 Detail Enhancement . . . 48

5.3.1 Transformer Functions . . . 48

5.3.2 Output of Transformer Functions 50
5.4 Detail-Recovered Fusion 51
5.4.1 Global Illumination Adjustment 52
5.4.2 Visual Comparison of Fused Results 53

6 Experiments and Evaluation 55

6.1 Dataset Preparation 55
6.2 Data Preprocessing 56
6.2.1 Splitting Images Based on Illuminations 57
6.2.2 Setting Up Train/Test Images 60
6.3 HDR Target Generation 61
6.4 Training of Bi-Directional Network 63
6.5 Image Measuring Metrics 64
6.5.1 Structural Similarity Index Measure 65
6.5.2 Mean Squared Error 66
6.5.3 Peak Signal-to-Noise Ratio 66
6.6 Dice Coefficient Segmentation Metric 67
6.7 Performance Evaluation Using SSIM 67
6.8 Performance Evaluation Using PSNR 69
6.9 Performance Evaluation Using Dice Similarity Coefficient (U-Net) . . 71
6.10 Comparison with State-Of-Art Methods 72
6.10.1 Visual Comparison 72
6.10.2 Quantitative Comparison 73
6.10.3 Enhancement Effect with Various Illuminations 74
6.11 Discussions on Experiments 75

7 Conclusion and Future Work 79

List of Figures

1.1 Segmentation Results (Using U-Net) on Wound Images with Different lighting conditions . . . 4

1.2 Basic Components in SmartWAnDS System Architecture . . . 7

2.1 Images with Different Exposures[1] . . . 15

2.2 Traditional HDR Generation . . . 16

2.3 Example of Semantic and Instance Segmentation [2] . . . 21

4.1 Structure of Bi-Directional Network . . . 27

4.2 Network Structure of DecomNet [51] . . . 33

4.3 Network Structure of SynthNet [51] . . . 36

5.1 Exposure Fusion and HDR-like Image Generation . . . 43

5.2 Membership Functions for Different Exposures. 1)left: under-exposed images. 2) middle: well-exposed image 3) right: over-exposed image . 44

5.3 The comparison between the original input image (a) and the directly fused result (b). We can clearly notice that some texture details in the wound area are lost after direct fusion. . . . 48

5.4 Intermediate Results and Enhanced Result (over-exposed image) . . . 50

5.5 Intermediate Results and Enhanced Result (under-exposed image) . . 51

5.6 Visual Comparison(Over-Exposed Images) . . . 53

5.7 Visual Comparison(Under-Exposed Images) 53

6.1 Moulage Used in IVDS Creation[20] 56

6.2 Hardware Settings Used in IVDS[20] 56

6.3 Two Images with Different Light Source Intensities 58

6.4 15 Images in IVDS only different in light source intensities(light source position 7, light temperature index 136). Light intensities from 10 to 255 in 15 steps. 58

6.5 Four Thresholds on Illumination Layer 59

6.6 LDR Images Used for HDR Target Creation 62

6.7 Inverse Camera Response Function for RGB Channels. The exposure times we used are 1/30, 8/30, 14/30 and 25/30 seconds. 62

6.8 HDR Targets Generated with Three Tone Mappers. 1) Drago (gamma=1.6, saturation=1.45); 2) Mantiuk(gamma=1.6, scale=0.8, saturation=1.45); 3) Reinhard(gamma=1.6, intensity=-8, lightAdapt=0.5, colorAdapt=0); The HDR-like image generated by **Drago** was used as the target to calculate image similarities. 63

6.9 Distribution of SSIM Scores 69

6.10 Distribution of PSNR Scores 70

6.11 Training Loss of U-Net 71

6.12 Testing Loss of U-Net 72

6.13 Visual Comparison on Enhanced Images. The first column contains the input images (moulage) from the IVDS dataset. The first 3 rows include HDR results generated from 3 over-exposed images and the last 3 rows include HDR results with under-exposed images. Except for our method, all other enhanced images(column 2-4) are tone-mapped with *Drago* using the same settings (gamma=1.6, saturation=1.45). 73

6.14 SSIM with Different L-values. Results are processed with moving average(window size equals to 10) for better visualization. 75

6.15 PSNR with Different L-values. Results are processed with moving average(window size equals to 10) for better visualization. 76

6.16 DSC with Different L-values. Results are processed with moving average(window size equals to 10) for better visualization. 76

List of Tables

6.1 The Illumination Varying Dataset (IVDS) 57

6.2 Conditions of Image Types 59

6.3 Hardware Specification and Software Platforms 64

6.4 Comparison with Other Methods 74

Chapter 1

Introduction

In this section, we mainly focus on the background of our research. We first discussed the thesis motivation and challenges of our work. The related work and our proposed method will also be briefly discussed in this section. At last, we listed the contributions in our work and described the structure of this thesis.

1.1 Motivation

Chronic wounds are commonly defined as wounds that take longer to heal than other wounds. Compared to most non-chronic wounds, chronic wounds usually take more than 3 months to heal. Some chronic wounds may take up more than 1 year to heal, with 60-70% probability of recurrence[42]. Due to the long-lasting pain both physically and mentally chronic wounds can cause to the patients, treatment for the chronic wounds has always been an urgent and emerging issue[34]. In the process of chronic wound treatments, patient care usually involves several visual inspections by wound experts and requires frequent visits to hospitals. However, the visual inspection processes are still quite rudimentary without an objective measuring metric. The quality of these visual inspection processes can be also challenging to maintain especially given the current situation that the number of chronic wound patients is huge compared to the limited number of wound experts.

To alleviate those issues in traditional visual inspection, our group has been working on new approaches in order to provide more reliable and objective solutions to chronic wound treatment since 2011. We developed the Smartphone Wound Analysis and Decision-Support (SmartWAnDS) system to provide automatic decision-making based on wound images taken using a smartphone and Electronic Health Records (EHR). This system is incorporated in the traditional chronic wound treatment and used to provide aided or supplementary suggestions for caregivers such as nurses and wound experts. This system is designed in such a way that caregivers who may not have professional wound treatment knowledge, such as nurses or social service works, can offer objective decisions with the guidance of the SmartWAnDS system. It reduces the workload of caregivers by simplifying the processes of routine checking-ups and also provides timely objective wound assessments to patients, which also reduce the risks of amputations caused by untimely treatments or un-objective assessments.

To obtain a persuasive assessment which can best reflect the healing process of wounds, the quality of wound images plays a significant role in our SmartWAnDS system. The information provided in the wound image can be crucial to making unbiased wound assessment and proper decision-making. However, wound images taken using smartphone cameras usually suffer from issues such as adverse global illumination, ill-exposures, image blur and etc, which can lead to degradation of image qualities. In this thesis, we mainly focus on the issues related to improper exposures in wound images taken using smartphone cameras.

1.2 Challenges

Wound images are an important type of data to use for wound assessment, which are also used in our SmartWAnDS system to provide objective assessments. The wound images our system uses are mostly taken by smartphones since they are most accessible to most patients. While smartphone images of chronic wounds are easy to acquire, it poses several challenges. First, unlike medical images generated by dedicated medical devices such as MRI scanners, the qualities of images taken by smartphones can be influenced by many environmental factors, such as the lighting condition, reflection, blur caused by motion, and etc. Second, the quality of the images can also be affected by the hardware/software configuration of smartphone cameras, such as the sensitivity of imaging sensors, the exposure settings of camera applications and blur caused by loss of focus. The quality of wound images can be of key importance for most machine vision tasks such as semantic segmentation, image understanding, and etc. According to [47] and [48], the variation in lighting and camera angles can cause up to 38% errors in wound area measurements. Wound images with poor qualities can also negatively affect our SmartWAnDS system since such wound images can lead to inaccuracy in most vision-related tasks, which in turn fails to provide persuading and objective solutions for chronic wound patients.

The issue of degradation in performance caused by poor quality wound images can be demonstrated in Figure 1.1. Without proper lighting and exposure settings, the segmentation results can degrade severely.

A key problem we aim to address is: how can we design a method to enhance the wound images taken by smartphones by converting the image into an **HDR**(high dynamic range) image to minimize the negative effect (such as loss of texture information) caused by adverse lighting condition or ill-exposure to improve the visual

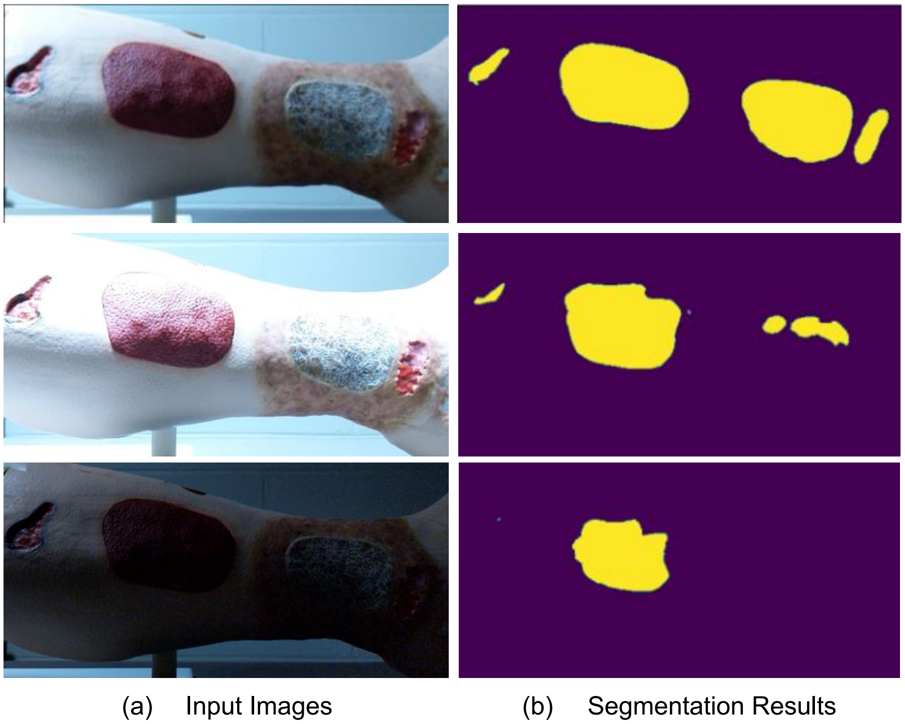


Figure 1.1: Segmentation Results (Using U-Net) on Wound Images with Different lighting conditions

quality of images in terms of texture details, image metrics and finally lead to higher performance in machine vision tasks (such as image segmentation) which are of great importance in image-based wound assessment?

In this thesis, we mainly focus on the exposure issues of wound images. Images with exposure issues can appear very dark (under-exposure) or very bright (over-exposure) in a significant amount of regions. Usually, images with ill-exposure are caused by the difference between the rather limited dynamic range of light cameras are able to capture and the high dynamic range of lights in the real world. Due to the limited illumination range of most smartphone cameras, only a very subset of illumination range can be captured and recorded (usually 0 - 255) by imaging sensors. The image is generated by filling each pixel value by estimating the amount of Photon captured by each sensor in the sensor array in a very short time(exposure time) with additional steps such as quantization. As a result of the difference in two light ranges, some sensors which record an excessive amount of photons can be saturated and thus leading to saturated pixels. Negative-saturated pixels can occur if some sensors cannot record enough photons to display the scene properly. Most images with ill-exposures are due to the direct sunlight or very dark environment when the images are taken, but improper camera settings can also lead to ill exposure.

1.3 Related Work

Researches on HDR (high dynamic range) image conversion as a way to mitigate exposure issues in images has been discussed in many literatures. One of the most popular methods is synthesizing HDR images using multiple photographic exposures[10]. Several images taken with different exposures are taken and the HDR image is generated by estimating the irradiance map. Apart from that, latest methods tend to synthesize HDR images using one exposure, especially with the booming of convolutional neural networks. For example, Endo et al. [14] used Camera Response Functions (CRFs) sampled from Grossberg and Nayar's Database of Response Function [15] to synthesize images with different exposures and used a deep network to learn the mapping between the input image and bracket images before synthesizing the HDR images. Liu et al. [31] model the HDR-to-LDR image generation process and adopt dedicated deep models to inverse each step.

An important and related field is image fusion. Image fusion is similar to HDR image generation as a way of image enhancement. However, image fusion methods generally do not preserve the relationship between HDR pixel intensities and actual irradiance intensity in the resulted image. It works by integrating visual information in several images into a single image. Song et al.[43] proposed a fusion method based on illumination using a probabilistic model. Vonikakis et al. [46] introduced membership functions to images based on exposure rates and pixel illumination values.

1.4 Our Approach

Our approach takes two phases. First, we adopt a deep learning approach to enhance the image exposure of the original image. Two counterparts will be generated from the original image. One image is the version with higher exposure (up-exposed) image and the other is the version with lower exposure (down-exposed) image. Then, we use these two images together with the original image to go through the image fusion pipeline. We proposed a detail-recovered fusion method that can address the detail loss in traditional image fusion processes. In this thesis, the final enhanced image we plan to generate is the tone-mapped version of the HDR image instead of the real HDR image, so the generated image has the same dynamic range as the input image (0-255). But for the sake of simplicity, we denote the tone-mapped HDR image as the HDR-like image or HDR image.

To validate our method, we include three types of evaluation experiments. First, we compare our generated HDR image with the ground truth HDR image which we synthesize with the traditional HDR generation pipeline. Several metrics are used, such as structure similarity score (SSIM) and peak signal-noise ratio (PSNR). Then, we train a UNet network using our prepared dataset and test the segmentation performance on both the original image and the enhanced image. Last, we compare our method with the state-of-art methods in HDR generation.

Figure 1.2 illustrates the system architecture of our SmartWAnDS system. It also visualizes how our method (in red dotted box) can be integrated into the whole pipeline of the SmartWAnDS system. Patients can receive recovery treatments from caregivers at patients' homes while caregivers can collect patients' health records(EHR and wound images) using the smartphone app. During the treatment process, EHR and wound images enhanced using our proposed method will be uploaded to a central server for analysis. With our proposed enhancement method, the enhanced wound image will provide higher visual quality and recover any texture information loss in the original image to facilitate the analysis process. When the wound analysis is done, the clinicians at hospitals will receive the analysis results from the central server and the assessment information will be published on the smartphone app as the feedback. At last, caregivers can be notified and give instructed treatments to patients according to the feedback.

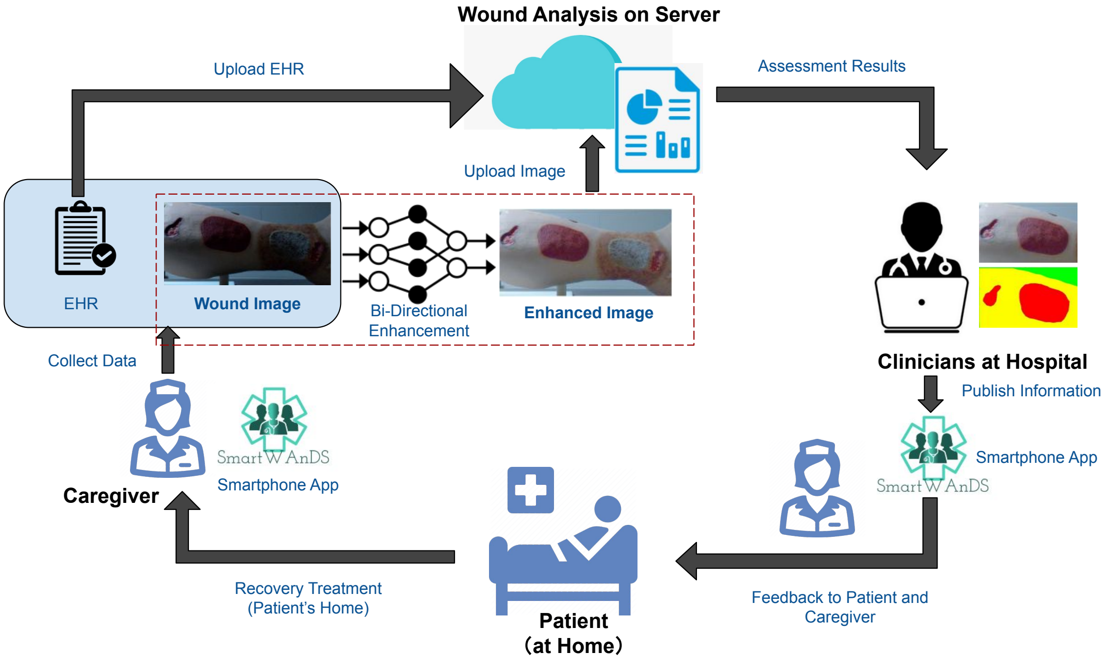


Figure 1.2: Basic Components in SmartWAnDS System Architecture

1.5 Novelty and Contributions

To the best of our knowledge, there is very little research in techniques to enhance both over-exposed and under-exposed images at the same time. Moreover, our work is the first to mitigate the over-exposure and under-exposure issues of chronic wound images. While there are many works on image enhancement, most of them are focused on low-light image enhancement and high-light removal. As for the HDR conversion, most of the methods are focused on exposure expansion of the images without exposure issues. Also, the image enhancement methods and the HDR conversion are designed for images which are not directly related to medical images. Our previous work includes 1) creation of the Illumination Varying Dataset of Wounds (IVDS) dataset, which is a large dataset consisting of 55440 wound images captured under controlled and different light conditions, smartphone cameras, and camera settings; 2) correlation analysis between the lighting conditions and segmentation accuracy and 3) enhancement of under-exposed images using Deep Retinex Model. In this thesis, 1) we designed the Bi-Directional Retinex model based on the Deep Retinex Model to allow both up and down-exposure operation, 2) explored different fusion methods for HDR generation and 3) proposed the detail-recovered fusion for HDR generation from ill-exposed images to enhance the image illumination condition. We also placed more emphasis on the part of over-exposed image enhancement since it is far less studied in image analysis compared with under-exposed image enhancement.

The contributions of our work are three-fold:

1. We designed a Bi-Directional Retinex Model based on the Deep Retinex Model to allow both up and down-exposure operation at the same time. The Deep Retinex Model is only designed for low-light image enhancement (up expo-

sure). In our work, we evaluate the performance of DRM in high-light image enhancement (down exposure). We also optimize the decomposition part in the DRM so the two operations (up/down exposure) can be performed simultaneously while reducing training and inference time. The bi-directional model can support end-to-end training and generate up and down-exposed images simultaneously with our proposed network structure.

2. We proposed a two-level image fusion pipeline to fuse images with different exposures into a detail-recovered HDR image. The issue of information loss in illumination-based fusion is addressed in our approach. Compared with direct fusion (illumination-based fusion), our fusion method is proved to preserve more texture details in ill-exposed regions.

3. We compared our proposed HDR-like generation method with several state-of-art methods. The proposed method in this thesis obtained most highest scores in several metrics(SSIM scores $0.76 \pm 0.04/0.69 \pm 0.08$ on bright/dark images, PSNR scores 28.60 ± 0.70 on dark images and DSC scores $0.76 \pm 0.09/0.74 \pm 0.09$ on bright/dark images). We also showed our method has consistent performance on images with various illumination conditions.

1.6 Thesis Outline

This thesis is structured in the following ways. In the first chapter, we discussed the motivations of this thesis, the question and proposed method in general as well as our contributions. In chapter 2, we listed the concepts and components which we adopted in our research. In chapter 3, we mainly surveyed the recent image enhancement work related to ours. In chapter 4 and 5, we discussed our proposed approaches in detail.(Chapter 4 focuses on how the bi-directional network is designed to create bracket images and chapter 5 focuses on how we use detail-recovered fusion to generate the HDR images.) In chapter 6, we mainly followed the steps of our experiments and described how we evaluate our proposed method and compare it with state-of-art methods. The conclusion and future work are included in chapter 7.

Chapter 2

Background

2.1 Smartphone-Based Wound Assessment

Smartphone-based wound assessment refers to a group of methods which utilizes the built-in sensors to help monitor, collect information of wounds to provide more symptom-related information to doctors, nurses and caregivers in the treatment procedure. As a supplement to the traditional therapy process, these types of assessment methods are getting increasingly popular largely due to the extensive use of smartphones as well as constantly evolving smartphone-related techniques. Besides, smartphone-based assessments also provide a low-cost and flexible way[3] for fast diagnosis compared with traditional methods which usually take a much longer time and are required by trained clinicians.

Different from dedicated medical equipment, the applications of smartphone-based wound assessment mainly focus on the early stage of diagnosis[4] which does not require operations that are too complex or demand high precision. Most wound assessments are focused on vision-related tasks. For example, Wang et al.[49] proposed a system which can provide image-based wound analysis on diabetic foot ulcers using built-in cameras. It can effectively analyze the wound healing status with a guided image capture procedure in the app. Apart from that, some works [45, 26] combined image processing and supervised learning models to detect and classify the types of wounds on wound images. Models such as neural networks, random forest, support vector machines as well as K-Nearest Neighbors(KNN) were used in wound tissue classification. Bhelonde et al.[8] implemented a vision-based system solely on smartphones to assess diabetic ulcers. It utilized the accelerated mean-shift algorithm to segment the wound areas and the Red-Yellow-Black color evaluation model to assess the wound healing process.

Our work does not conduct wound assessment on patients directly, however, our proposed method is closely related to it. In this thesis, we aim to enhance the wound images taken using smartphone cameras to mitigate the adverse effect caused by ill-exposures, which is typically found in wound images taken with smartphone cameras. The enhancement in wound images can be beneficial to common wound analysis tasks (such as wound image segmentation).

2.2 Image Illumination Enhancement

Image illumination enhancement refers to methods which enhance the original image in order to have a more preferable illumination condition. We categorize the methods in 3 types, which are histogram-based, frequency-based and Retinex-based methods.

2.2.1 Histogram-Based Methods

Histogram-based image enhancement methods are widely studied in many literatures. Many methods are based on histogram equalization[19], which works by remapping the colors of the input image based on the histogram of the pixel values (gray value). The goal of using histogram-based methods is to enhance the contrast of the original image, which does not put much constraint on preserving the original color information of the input image. These methods can be further divided into global and local methods based on the image regions they take effects.

2.2.2 Frequency-Based Methods

Frequency-based methods[39, 5] are sometimes referred to as enhancement methods in the frequency domain, which are contrary to methods in the spatial domain. Unlike spatial domain methods which directly consider images as matrices of pixel values, frequency domain methods take images as signals and generally require Fourier transform. The frequency domain is converted back into the spatial domain after operations are applied. Apart from illumination enhancement, frequency domain methods are also widely used in image smoothing, image sharpening, de-noising, and etc.

2.2.3 Retinex-Based Methods

Retinex-based methods[37, 51, 16] focus on separating an image into two components which are illumination and reflectance map. They are a group of methods based on the Retinex theory[27]. The effect of illumination enhancement is directly related to the accuracy in estimating the illumination maps, which ideally only reveal information of the illumination condition in the real world.

Our proposed method belongs to the Retinex-based methods. The Retinex Theory is used in bracket image generation (in Section 5.4).

2.3 LDR and HDR

The dynamic range of images refers to the ratio between the brightest and darkest part of an image [36]. Most of the images we encounter and use are LDR (low dynamic range) images. LDR images, which are contrary to the HDR (high dynamic range) images, are taken with most commercial cameras and stored as 8-bit formats. The most common LDR images use 8 bits (0-255) integers to present the pixel values in each color channel. Due to the fixed values every pixel can vary, the range of pixel values in each channel is limited. The HDR images, on the other hand, usually store the pixel values in floating decimal numbers, which results in a much larger range compared with the LDR images[40]. HDR images capture a much wider range of light intensities than LDR images do. In this way, HDR images can more accurately record the illumination changes in natural scenarios, especially in very dark and very bright environments. Capturing HDR images directly can be rather expensive since it requires dedicated imaging devices to support enough dynamic range in sensors. Generally, HDR images are synthesized with several LDR images taken with different exposures, and it has been a standard way for most photographers and artists to generate HDR products. Displaying HDR images directly can also be costly since most commercial displays and printers are limited to low dynamic ranges. So instead of directly displaying HDR content, tone mapping is used to convert an HDR image by mapping the high dynamic range back to the range of LDR images for display.

Due to the very limited range of illumination change the LDR images are capable to display, LDR images tend to suffer from exposure issues, typically over-exposure and under-exposure. An LDR image with over-exposure issues usually contains a large portion of very bright areas (pixel values approaching 255). These pixels be-

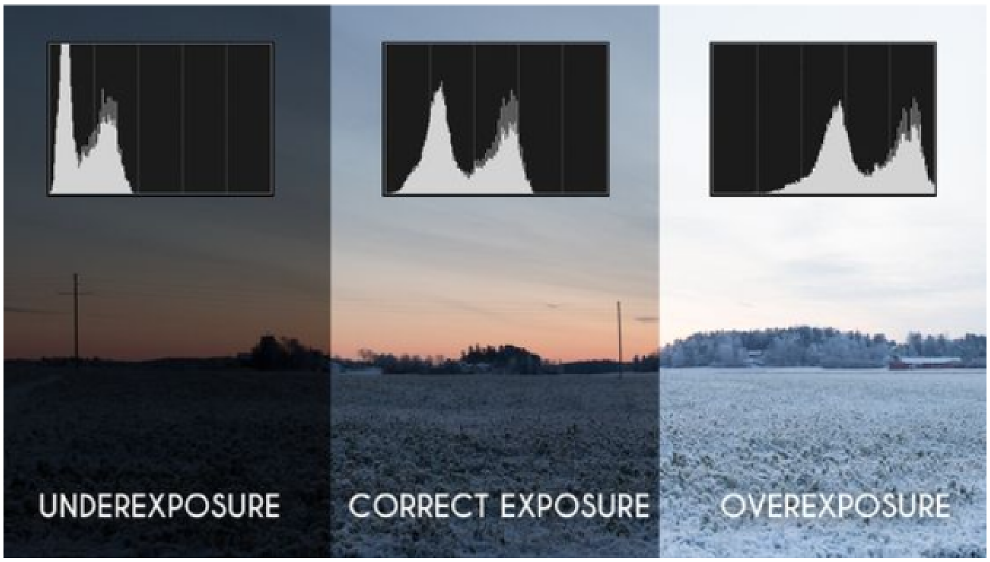


Figure 2.1: Images with Different Exposures[1]

come saturated because the number of photons received at these sensor locations exceeds the maximum that pixel values are able to display. On the other hand, a large portion of dark areas (pixel values approaching) usually can be found in under-exposed images. These pixels are negatively saturated because the number of photons received is too small that they fail to provide enough contrast. Figure 2.1 shows an example of LDR images of different exposures. The histogram of the illumination layer of each image is also included. We can clearly notice that the image on the left (under-exposed) contains lots of low-intensity pixels and pixel values in the image on the right (over-exposed) generally have high intensities. Most pixel values in the image with appropriate exposure (correct-exposed) are concentrated in the center part of the histogram and such images can reveal more texture information than the under- and over-exposed ones.

Converting LDR image into HDR image can enhance the contrast of the original image since the range of pixel is expanded to allow more dynamic range. Mostly, HDR images are tone mapped instead of being displayed directly. The tone-mapped

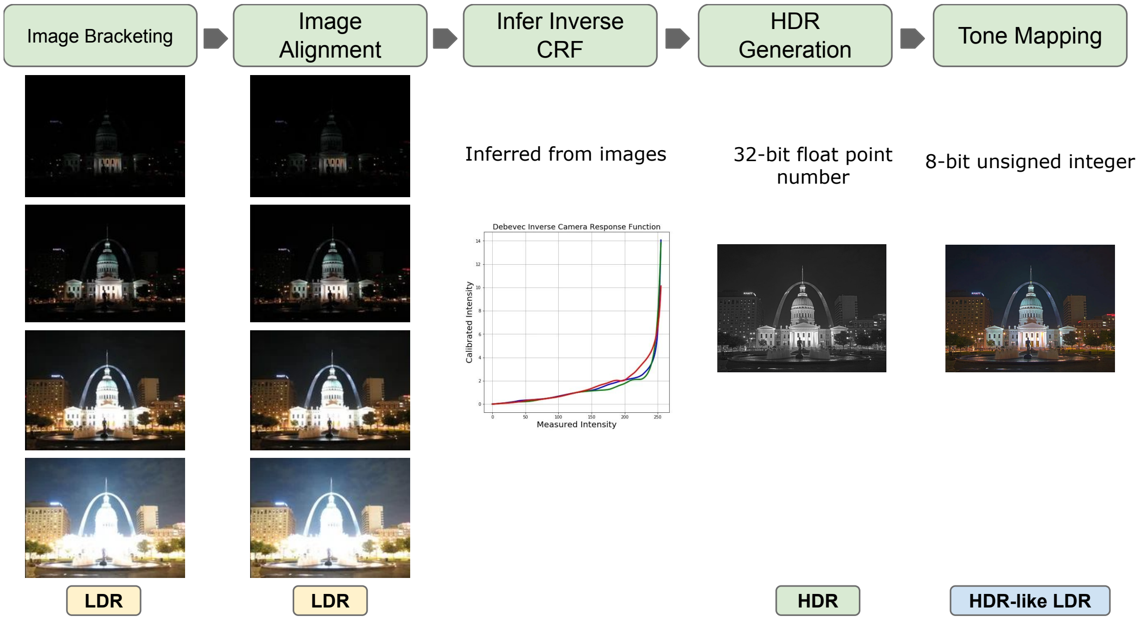


Figure 2.2: Traditional HDR Generation

HDR image has the same dynamic range as the LDR image, but the color and illumination contrast is enhanced compared with the original LDR image. In our work, we aim to generate the tone-mapped HDR image directly to improve the illumination condition of the original image.

2.4 HDR Image Generation

HDR images can be obtained directly, but most of the HDR images we encounter are synthesized from LDR images. Traditional HDR image generation[10] pipeline (see Fig.2.2) requires several LDR images taken with different exposures of the same subject. For most smartphone cameras, 3 images are usually taken. These images are then aligned to make the subject appear in the same position in each photo.

The method which combines several lower range photos taken with different exposure values is usually referred to as exposure bracketing[Kalantari 2017].

Ideally, the pixel values will have a linear relationship with the radiance received, which means the value of each pixel should be linearly proportional to the radiance received at the corresponding sensor. However, in reality, the relationships are usually non-linear in most cameras. And different cameras usually have different mapping. This mapping (radiance to pixel value) is usually referred to as the camera response function (CRF). To recover the radiance distribution of the scene where photos are taken, an inverse operation should be used to map the pixel values back to the radiance values. This function, which is the inverse operation of CRF, is named inverse CRF, or invCRF in most literatures. With the different exposures used as well as the photos, we can estimate the invCRF curves of the camera. The invCRF will map R, G and B channels separately, so three curves will be produced.

The invCRF, exposures, as well as photos, will be used to construct the HDR image. The HDR image can be seen as the radiance map of the scene since it reflects the degrees of radiance in each pixel. Also, due to the high dynamic range in nature, the difference of the pixel values in the HDR can be several magnitudes larger than the original LDR's. Typically, the pixel values of HDR images are stored using 32-bit floating-point numbers. In order to accommodate for most non-HDR displays, tone mapping is usually performed on the HDR image to narrow the range down and convert it to one LDR image. The LDR image processes lots of advantages compared with the original LDR, most notably the increase of the contrast of colors. It can also mitigate the exposure issues of the original LDR image.

Apart from the traditional way of generating HDR images, there are also many researches which attempt to generate HDR content with fewer restrictions. For example, some researchers have developed methods of generating HDR images using single image. These methods, which are sometimes referred to as single-exposure HDR generation, are usually implemented in two categories. The first category of single-exposure methods[31, 14] will first synthesize additional images with different exposures, and go through the traditional HDR generation process with the original and synthesized images. The second category of methods[13, 33] work more like a direct mapping function. It learns the correlations between the LDR image and the HDR counterparts and directly maps the LDR image to the HDR image.

The LDR-to-HDR conversion produces images with greater luminosity than standard imaging techniques to emulate the experience of human vision systems. Unlike histogram-based images or Retinex-based images, HDR is not typically used in enhancement over existing photos.

In our work, we generate the HDR target image using the traditional HDR image generation pipeline (in Section 6.3). The image is used as the ground truth when we test the generated images on several metrics.

2.5 Tone Mapping

Tone mapping is the process of converting an HDR image back into an LDR image for the ease of visualization in commercial displays and printers. Unlike the camera response function, tone mappers are designed to enhance the visual quality instead of solely expressing the positive relationship between radiance and pixel values. The most common tone mappers are Drago[12], Mantiuk,[32] and Reinhard[41].

Tone mapping was used to generate the HDR image target in Section 6.3. We also used tone mapping to convert HDR images generated from methods we compete against (in Section 6.10) into 8-bit LDR images in performance comparison.

2.6 Image Fusion

Image fusion[53] is very similar to the HDR generation as it takes several images to generate one image with better visual quality. One key difference between image fusion and HDR generation is, image fusion works without knowledge of the exposure, so it is not related to the recovery of the radiance map. In other words, the relationship of pixel intensities and the radiance values is usually not consistent since no information of environmental illumination is required when doing fusion. As a result, image fusion will not necessarily preserve the positive relationship between the pixel values and radiance levels. Most image fusion algorithms work in spatial space or frequency space.

Other than enhancing the visual quality of images, the purpose of image fusion varies on different scenarios, but in general, image fusion is a way to combine or gather information from multiple images to create an output image with ideally all required information for analysis, which is considered as a branch in data fusion.

While there are many standard techniques in medical image fusion[21], lots of works are focused on images generated with specialized devices, such as MRI scanners, and most of these methods are not designed for user-taken photos. In this thesis, we only focus on photos taken by users.

In our work, we apply image fusion techniques on bracket images (in Section 5.4) to generate the enhanced image.

2.7 Retinex Theory

Proposed by Edwin Land, Retinex Theory[27] provides a conceptual model on human color vision to account for color sensations in real scenes. The main idea behind the Retinex Theory is color constancy, which means the perceived color of objects will remain relatively constant even under varying illumination conditions, or different ambient light levels. According to this theory, an image can be decomposed and represented using two subcomponents, which are usually termed “reflectance” (or “reflectance map”) and “illumination” (or “illumination map”). Reflectance map can be considered the intrinsic color property of objects. The value of the reflectance map is determined by properties (such as material) of the object itself, and will always remain invariable when observed. Illumination map describes the ambient illumination condition which is directly related to the light level. The two subcomponents are independent of each other, and an image taken at a certain level of illumination condition can be preserved as the element-wise multiplication of these two components. This relationship can be concluded using Equation 2.1. denotes the input image, while and denote the reflectance map and illumination map respectively.

$$S = R \otimes I \tag{2.1}$$

A large portion of our method is based on the Retinex theory. To be specific, the assumptions of reflectance and illumination map of input images we made, optimization we used as well as the bi-directional network structure we proposed are all based on it.

2.8 Image Segmentation

Image segmentation is a general process to group each pixel of input images. Pixels belonging to the same object will be assigned with the same label(usually referred to as image segments). There are generally two categories of image segmentation, which are semantic segmentation and instance segmentation. These two segmen-

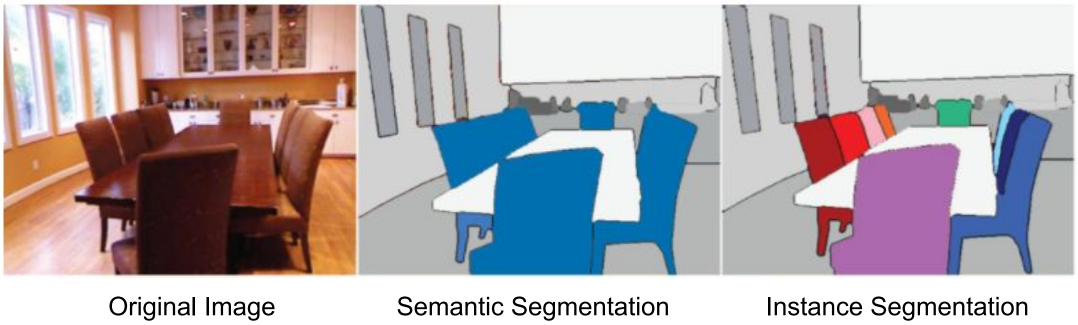


Figure 2.3: Example of Semantic and Instance Segmentation [2]

tation methods are similar in the way that both provide pixel-level prediction on the input images (which object does the current pixel belong to). The difference is instance segmentation will differentiate objects within a same category, which, in other words, can predict each pixel to which object it belongs. On the other, semantic segmentation will not differentiate objects and can only output which category the current pixel belongs to.

Fig.2.3 shows the segmentation results generated by semantic segmentation and instance segmentation. Pixels with the same color are considered to belong to the same type. In semantic segmentation, pixels related to all the chairs will have the same label (chair). However, the segmentation result doesn't directly provide information on different chairs–they are instead treated as the same type. In instance segmentation, all chairs are identified while every chair is treated as a single instance labeled as chair.

Image segmentation is an important technique in image-based wound assessment. In our work, we use image segmentation as a way to help evaluate the effectiveness of our proposed method in image enhancement. The change in segmentation accuracy is used as an indicator to show how much performance gain we can get in machine vision tasks using the proposed enhancement method.

Chapter 3

Related Work

In this chapter, we discussed methods which are related to our work, which are Direct Enhancement, Multi-Exposure Fusion and Inverse Tone Mapping.

3.1 Direct Enhancement

In this section, we mainly discuss two classic types of methods which enhance image contrast or color without additional information (such as multiple exposures) or effect on dynamic range, which are histogram-based methods and methods using illumination map estimation.

In histogram-based methods, Histogram Equalization[19] is a classic method which remaps the colors of the input image based on the histogram of the pixel values (gray value) so the cumulative distribution function of pixel values is approximately linear. By making the pixel values more evenly distributed in an image, it boosts the contrast of colors. It serves as the basis of many other histogram-based methods. According to the area in which the method takes effects, there are global histogram equalization methods (such as the Brightness Preserving Bi Histogram Equalization, Dualistic Sub image Histogram Equalization, Recursive Mean

Separate Histogram Equalization and etc) as well as local histogram equalization methods (such as Adaptive Histogram Equalization and Contrast Limited Adaptive Histogram Equalization).

Methods using illumination map estimation[37, 51, 16] are generally based on the Retinex Theory[27, 28], which assumes that one image can be represented as the pixel-wise product of two components, the reflectance map and the illumination map. The reflectance map conserves the color information of the scene which is not affected by the environmental light conditions, while the illumination map represents the distribution of environmental lights in the image. A classic approach is Single Scale Retinex[9] which uses one scale for surround function when doing convolution. Multi Scale Retinex[22, 38] is similar but multiple surround functions are used. There are many variations which are based on Muti Scale Retinex but extend to other situations, such as improvements on color restoration[22], or on low-light images[30], and etc.

3.2 Multi-Exposure Fusion

Exposure fusion techniques utilize information from images captured or synthesized at multiple exposures to infer the HDR-like image with optimal exposure. Song et al.[43] use a probabilistic model that preserves the calculated image luminance levels and suppresses reversals in the image luminance gradients. Mertens et al.[35] compute a perceptual quality measure for each pixel in the multi-exposure sequence encoding desirable qualities such as saturation and contrast for fusion. Liu et al. [29] perform signal decomposition using independent component analysis to restore texture and color information during fusion. Vonikakis et al. [46] introduce membership functions, which assign weights to every pixel in images based on exposure rates and pixel illumination values. In our work, we merge images using a membership function similar to Vonikakis et al. [46] to merge images. A key difference is that using exposure fusion to generate the final output, an intermediate fusion is performed to enhance the texture details in highly saturated and negatively saturated regions. Our method ensures that texture details are preserved maximally during exposure fusion.

3.3 Inverse Tone Mapping

Inverse tone mapping or reverse tone mapping are referred as inverse Tone Mapping Operator (iTMO) [7] or reverse Tone Mapping Operator (rTMO) in the literature. It is the dual of tone mapping and infers an HDR from LDR image, which is the opposite of tone mapping [7]. iTMO is similar to the conventional HDR imaging pipeline, except that only one LDR image taken with single exposure level is utilized. Endo et al. [14] propose a technique for synthesizing bracket images using Camera Response Functions (CRFs) selected from 201 curves in Grossberg and Nayar's Database of Response Function [15]. They use k-means clustering and an encoder-decoder model to learn the mapping between the input LDR image and synthesized bracket images. Liu at al. [31] model the HDR-to-LDR image generation process with three steps (dynamic range clipping, non-linear mapping and quantization) and inverse each step by designing a dedicated neural network. [33] design a multi-scale CNN architecture named ExpandNet that learns the direct mapping from LDR image to its HDR counterpart. HDRCNN [13] predicts missing details in over-exposed regions during HDR generation. Apart from directly generating an HDR image, some works [52, 23] generate tone-mapped images or HDR-like images instead of the original HDR images. Besides, some iTMOs [6, 25] focus on reconstructing saturated or over-exposed regions to recover lost information in highlight areas. Similar to other iTMO methods, our proposed method only requires one LDR as input, however, we focus on generating the tone-mapped image instead of the HDR image itself.

Chapter 4

Bracket Image Generation

In this chapter, we will describe our algorithmic pipeline using the Retinex model to generate images with higher/lower exposures (bracket images). There are two steps in generating images with different exposure. First, the image will be decomposed into reflectance map and illumination map according to the Retinex Theory. Then the illumination map will be adjusted (higher/lower exposure) and merged with reflectance map to produce a bracket image which is only different in illumination conditions from the input image.

The structure of our proposed pipeline in generating bracket images is shown in Figure 4.1. Two types of deep networks are deployed in our pipeline, which are $DecomNet$ and $SynthNet$. $DecomNet$ is used to decompose image into reflectance map and illumination map based on the Retinex Theory, and $SynthNet$ is an encoder-decoder model used to generate images with higher and lower exposure. The pipeline is optimized in terms of maximized data sharing and consistency in decomposition for images with a higher range of illumination densities.

The details in the design of the pipeline are discussed in the rest of the chapter. The following sections in this chapter are structured in the following way. First, we

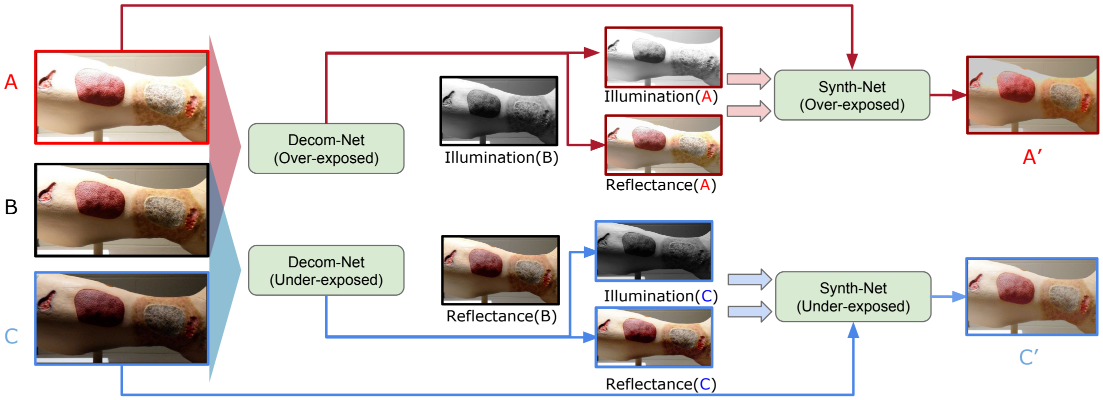


Figure 4.1: Structure of Bi-Directional Network

discuss the assumptions we made in image decomposition and how an optimization problem can be formalized based on them. We later describe how this optimization can be achieved using a data-driven approach. Next, we discuss the methodology used to synthesize bracket images given the reflectance map and illumination map. At last, we wrap up and describe in detail the whole process in bracket image generation in our proposed pipeline.

4.1 Assumptions in Decomposition

Directly decomposing an image into reflectance map and illumination map is a classic ill-posed question, so we formalize the decomposition as an optimization problem based on several assumptions. The assumptions used in image decomposition are discussed in detail in this section.

An image S can be decomposed into reflectance map R and illumination map I. The relationship (see Equation. 2.1) holds true for any given image S. For a grayscale image with only one channel, there is only one reflectance map responding to that channel. For color images with multiple color channels (such as RGB images), the number of reflectance maps is equal to the number of color channels. In this section, we discuss the assumptions proposed in the Retinex Theory as well as the ones on R and I.

Color Constancy and Consistency of Reflectance. When observing an object, human color perception system will ensure the color perceived stays relatively constant (on the surface) under varying illumination conditions. Color constancy is usually considered ideal since it is not affected by the illumination condition. We use reflectance map Rto represent the objective color independent from illumination conditions. In this way, we can assume the values of the reflectance map remain consistent in images only different in illumination conditions.

Independence in Color Channels. For color images (RGB image), the same computation is performed on each color channel. In other words, each color channel can be considered as the element-wise multiplication between the unique reflectance map and the illumination. The reflectance map on each channel is usually different from each other, but the same reflectance map is used across all channels.

Smoothness of Illumination. The intensities of illumination tend not to change drastically in consecutive regions in images, which is intuitive in most cases. In other words, The overall variation of values in illumination map I should be low if the pixels are on the same surfaces. However, this assumption will not hold if the observed object is highly complex in structure.

4.2 Optimization Formation

Given the relation between S, R and I as well as the assumptions mentioned above, we can translate the decomposition problem into an optimization problem. We define the re-construction error in the Equation 4.1, given the target image S, estimation of reflectance map $\hat{R}$ and estimation of illumination map $\hat{I}$.

$$L_{recon}(\hat{R}, \hat{I}, S) = \|\hat{R} \otimes \hat{I} - S\|_1 \tag{4.1}$$

Based on the assumption on the consistency of reflectance, difference between predicted reflectance map $\hat{R}$ and the ground-truth reflectance map R should be minimized, so the consistency loss of the reflectance map can be defined in Equation 4.2, given R and $\hat{R}$. In general, the value of ground-truth reflectance map R is unknown. To solve this problem, we will discuss how the value of R can be approximated with a data-driven approach.

$$L_{refle}(R, \hat{R}) = \|R - \hat{R}\|_1 \tag{4.2}$$

By using the assumption on the smoothness of illumination, high variation in the predicted illumination map I should be penalized. However, since illumination should only be considered smooth for pixels on the same plane, minimizing gradient variation of the illumination map without considering structure information does not result in reasonable solutions. In the most extreme case, the values in the illumination map will all be equal and result in 0 in variance, which is far from reality.

To resolve this issue, the weight matrix can be used to imply the structural property of objects. A typical way to include structural information in the estimated illumination map is to construct a weight matrix W(usually we use 2 matrice to represent weights in horizontal and vertical axis separately) using the information provided in the reflectance map R or the input image S. In this way, we can use Equation 4.3 to represent the smoothness loss of illumination. The calculation of smoothness loss is performed in both horizontal and vertical directions. ΔI_h and

ΔI_v denote the average gradient in horizontal and vertical directions respectively. ΔW_h and ΔW_v denote the weight matrix along the horizontal/vertical axis. The construction of the weight matrix is discussed in the next section.

$$L_{smooth}(I) = \|\Delta I_h \otimes W_h + \Delta I_v \otimes W_v\| \tag{4.3}$$

We combine the 3 loss function as the target loss function. The target loss function is illustrated in Equation. 4.4. λ_1, λ_2 and λ_3 are positive coefficients used to balance the weights in optimization. By minimizing the target loss function, we can obtain the $\hat{R}$ and $\hat{I}$ from the given image.

$$L(\hat{R}, \hat{I}) = \lambda_1 L_{recon} + \lambda_2 L_{refle} + \lambda_3 L_{smooth} \tag{4.4}$$

4.3 Construction of Illumination Weight Matrix

We usually consider the horizontal weight W_h and the vertical weight W_v separately when constructing the weight matrix. In this section, we will discuss several strategies to construct the weight matrix.

Uniform Weight. By setting the values of W_h and W_v all equal to 1, L_{smooth} can be treated as the minimization problem without any structural information involved. This is a trivial case which is not used in our experiments.

Gradient of Illumination Map. Guo et al.[16] proposed an iterative way to calculate the value of illumination map I. According to their method, the initial illumination map $\hat{I}$ can be used as the weight in approximation. The initial illumination map $\hat{I}$ can be calculated as the maximum intensity in R, G and B channels for all the pixels (see Equation 4.5). The weight in both axes can be computed using Equation 4.6 and 4.7 .

$$\hat{I}(x) = \max(S_i(x)), i \in r, g, b \tag{4.5}$$

$$W_h = \frac{1}{\Delta_h \hat{I} + \epsilon} \tag{4.6}$$

$$W_v = \frac{1}{\Delta_v \hat{I} + \epsilon} \tag{4.7}$$

Gradient of Reflectance Map. The reflectance map can imply the structure information of the input image. More precisely, the gradient of the reflectance map can be used as an indicator of the existence of strong structures (such as edges) in the image. We can assume that the level of illumination can change drastically in pixels where such strong structures exist. In this way, the weight can be considered as the gradient of reflectance map in both axes, which is shown in Equation 4.8 and 4.9. λ is a scalar controlling the effect of weight.

$$W_h = exp(-\lambda \Delta R_h) \tag{4.8}$$

$$W_v = exp(-\lambda \Delta R_v) \tag{4.9}$$

4.4 Data-Driven Approach to Solve the Optimization Problem

As we have discussed in the previous section, to calculate the loss in the generated reflectance map L_{refle}, we need to know the ground-truth value of the reflectance map R, which is not possible in general. To address this issue, we can adopt a data-driven approach to approximate the value of R. With the data-driven approach, we can provide a workable solution to minimize the defined loss function (see Equation 4.4) and finally solve the optimization problem to get the values for reflectance $\hat{R}$ and illumination map $\hat{I}$.

Instead of only relying on one input image S, we introduce a way to take image pairs as input. One image pair consists of two or more images. In the case of only two images, we use a reference image S' apart from the input S we hope to decompose. The reference image S' should only be different from the input image S on the illumination conditions. In other words, the reflectance maps of both images should be the same, as indicated in Equation 4.10.

$$R(S) = R(S') \tag{4.10}$$

Without other prior knowledge of reflectance maps of both images, we can simulate the process of generating R and I using a deep learning network and use the assumptions discussed above as constraints to solve the optimization problem. The network takes into a pair of images as input and applies the same transformation on both images. The high-level representation is shown in Equation 4.11 and 4.12.

$$\hat{R}, \hat{I} \leftarrow DecomNet(S) \tag{4.11}$$

$$\hat{R'}, \hat{I'} \leftarrow DecomNet(S') \tag{4.12}$$

We adopt the *DecomNet* structure from Wei et al[51].The structure of the *DecomNet* is illustrated in Figure 4.2. The input RGB image (pixel range between 0 and 1) is used as the initial reflectance map (shape = $W \times H \times 3$) while the initial illumination map (shape = $W \times H \times 1$) was first inferred from the input

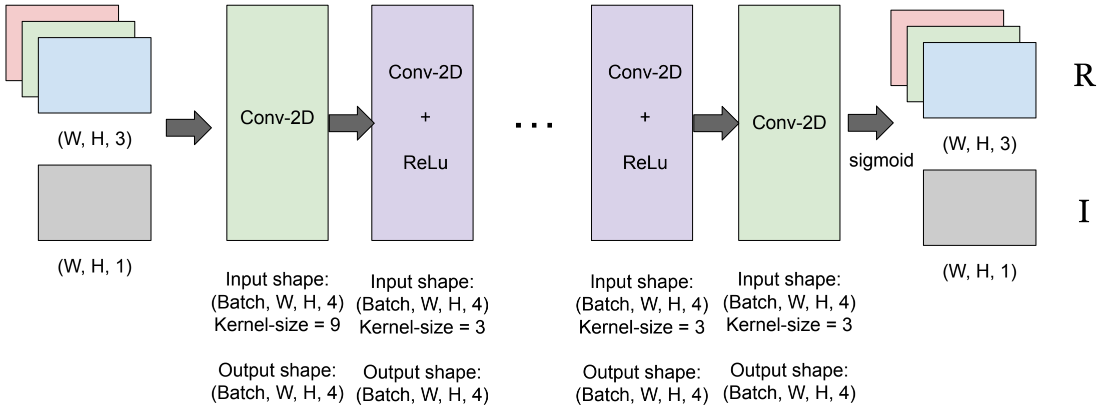

(W, H, 3)
(W, H, 1)
Conv-2D
Conv-2D + ReLu
. . .
Conv-2D + ReLu
Conv-2D
sigmoid
(W, H, 3)
(W, H, 1)
R
I
Input shape: (Batch, W, H, 4) Kernel-size = 9
Output shape: (Batch, W, H, 4)
Input shape: (Batch, W, H, 4) Kernel-size = 3
Output shape: (Batch, W, H, 4)
Input shape: (Batch, W, H, 4) Kernel-size = 3
Output shape: (Batch, W, H, 4)
Input shape: (Batch, W, H, 4) Kernel-size = 3
Output shape: (Batch, W, H, 4)

Figure 4.2: Network Structure of DecomNet [51]

RGB image. Similar to LIME, the initial illumination is the maximum intensity in R, G and B channels for all the pixels (see Equation 4.5). The initial reflectance and illumination map are stacked (shape = $W \times H \times 4$) on the channel dimension and used as input to the network. The input layer of the *DecomNet* is a convolutional layer which takes $W \times H \times 4$ tensors as input. Followed by the input layer is a series of same network cells. Each network cell consists of a convolutional layer and ReLu function. The output layer is a convolutional layer without activation functions. Since all the convolution operations are padded to maintain the input size, the shape of the output tensor will not be changed. To obtain the output reflectance and illumination map, we split the output tensor into a $W \times H \times 3$ and a $W \times H \times 1$ tensor. Sigmoid function is applied on both tensors to make the pixel range between 0 and 1.

4.4.1 Training of DecomNet

Training the *DecomNet* can be a little different from most deep learning networks. To be able to calculate the loss function (see Equation 4.4), we need to provide the ground-truth reflectance map R. Without other information, we deploy two exact *DecomNet*s to train simultaneously, the parameters are shared across the two networks. The reflectance map obtained from one network is used as the ground-truth R, while another reflectance map is considered as the estimation $\hat{R}$.

While the parameters of these two networks are shared, we pass slightly different data for training (Or there will be no difference than only training one network). As we discussed before, a pair of image (S, S') is used as inputs for the two networks (S is used in the first *DecomNet* while S' is used in another). And we update the loss function as Equation 4.13.

$$\begin{aligned} L(\hat{R}, \hat{I}, S, \hat{R}', \hat{I}', S') = \lambda_1 L_{AvgRecon}(\hat{R}, \hat{I}, S, \hat{R}', \hat{I}', S') + \\ \lambda_2 L_{AvgMutRecon}(\hat{R}, \hat{I}, S, \hat{R}', \hat{I}', S') + \\ \lambda_3 L_{AvgSmooth}(\hat{R}, \hat{I}, S, \hat{R}', \hat{I}', S') + \\ \lambda_4 L_{refle}(\hat{R}, \hat{R}') \end{aligned} \tag{4.13}$$

where

$$L_{AvgRecon}(\hat{R}, \hat{I}, S, \hat{R}', \hat{I}', S') = \frac{1}{2}(L_{recon}(\hat{R}, \hat{I}, S) + L_{recon}(\hat{R}', \hat{I}', S')) \tag{4.14}$$

$$L_{AvgMutRecon}(\hat{R}, \hat{I}, S, \hat{R}', \hat{I}', S') = \frac{1}{2}(L_{recon}(\hat{R}, \hat{I}', S') + L_{recon}(\hat{R}', \hat{I}, S)) \tag{4.15}$$

$$L_{AvgSmooth}(\hat{R}, \hat{I}, S, \hat{R}', \hat{I}', S') = \frac{1}{2}(L_{smooth}(\hat{I}, \hat{R}) + L_{smooth}(\hat{I}', \hat{R}')) \tag{4.16}$$

To calculate the $L_{AvgSmooth}$ (see Equation 4.16), we use Equation 4.17.

$$L_{smooth}(\hat{I}, \hat{R}) = \|\Delta \hat{I}_h \otimes exp(-\lambda \Delta \hat{R}_h) + \Delta \hat{I}_v \otimes exp(-\lambda \Delta \hat{R}_v)\| \tag{4.17}$$

4.4.2 Evaluation of DecomNet

Evaluation of $DecomNet$ is straight-forward: take either one of the network (since there share the same parameters) and feed one image as input. The output of the network contains the reflectance map and the illumination map.

4.5 Illumination Adjustment

Given an input image, we aim to generate images with different exposures. We have discussed the pipeline of decomposing the image into reflectance map and illumination map. In this section, we will discuss how these two components can be used to synthesize images with different exposures.

As we discussed before, the illumination map controls the light information of the image. To change the exposure of a given image, we need to update the illumination map of the original image, and use the new illumination map to merge with the original reflectance map. We can represent the operation as Equation 4.18. We use $f(\hat{I})$ to denote the transformation of the estimation of illumination map $\hat{I}$.

$$S^* = \hat{R} \otimes f(\hat{I}) \tag{4.18}$$

Although we can use specific rules to design the transformation function f, a better way is to simulate real scenarios. In this way, we can learn the transformation function by using a pair of images (similar to decomposition).

We can again use a neural network ($SynthNet$) to learn the transformation. Given the input image S, we choose an image S^+ which is taken with higher expo-

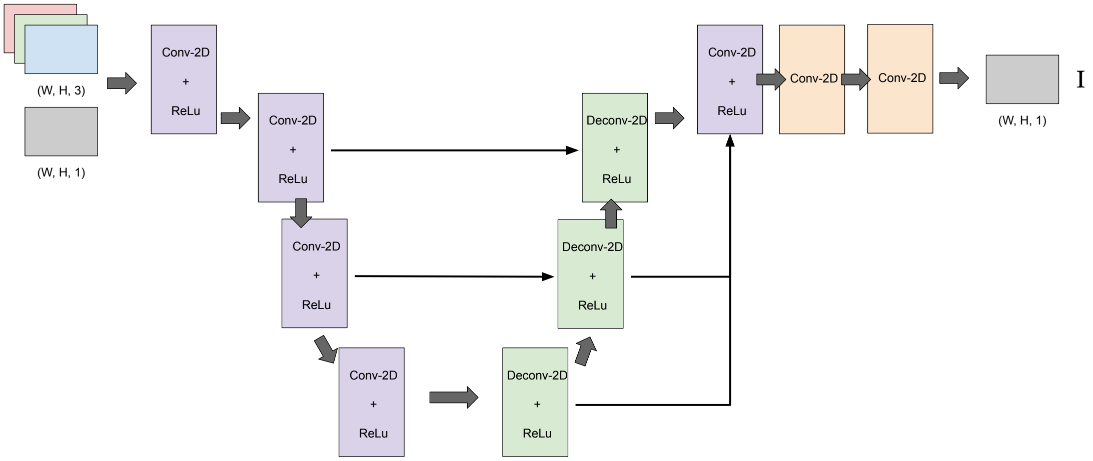


Figure 4.3: Network Structure of SynthNet [51]

sure. In theory, after decomposing these two images, the reflectance maps should be identical while the illumination map of S^+ reflects a brighter light condition than the illumination map of S. The network structure is illustrated in Figure 4.3.

SynthNet acts as the transformer function which can transform the original illumination map of the input image. Different from how we represent the transformer function in Equation 4.18, we also use the reflectance map as input to help us design the loss function (discussed in the next section). The main structure resembles the structure of U-Net, which contains a series of convolution operations followed by a series of de-convolution operations. Skip connections are also added to avoid information loss during the de-convolution steps. To solve the shape mismatching issues when adding the skip connections, up-sampling is performed before de-convolution operations. The results of de-convolution operations are concatenated and processed with several convolution layers to output a tensor with one channel (illumination map).

4.5.1 Training of SynthNet

SyntheNet takes the output of *DecomNet* as input. For the original image S, $\hat{R}$ and $\hat{I}$ are obtained using *DecomNet*. The output $\hat{I*}$ denotes the illumination map which is transformed in *SynthNet*. The relationship is shown in Equation 4.19.

$$\hat{I*} \leftarrow SynthNet(\hat{R}, \hat{I}) \tag{4.19}$$

We consider two metrics for the loss function for *SynthNet*: 1) the similarity of the synthesized image S^* and a similar image S^+ with higher exposure; 2) the smooth loss of the generated illumination map. The loss function we use to train *SyntheNet* is shown in Equation 4.20.

$$L(\hat{I*}, \hat{R}, S^+) = \lambda_1 L_{recon}(\hat{R}, \hat{I*}, S^+) + \lambda_2 L_{smooth}(\hat{I*}, \hat{R}) \tag{4.20}$$

4.5.2 Evaluation of SynthNet

Evaluation of *SynthNet* is straightforward. We can use Equation 4.19 to get the new illumination map. By merging the illumination map with the reflectance map (see Equation 4.18) we can generate an image with a different exposure. While we illustrate the training process of *SynthNet* using an image S^+ with higher exposure than the original image S as the reference image, an image with lower exposure S^- can also be used. The difference is: when we use S^+ as the reference image, the light level of the generated illumination map $\hat{I*}$ will be higher than before $\hat{I}$ (closer to the illumination map of S^+); however, the light level of the generated illumination map $\hat{I*}$ will be lower than before $\hat{I}$ if we use S^- as the reference image.

4.6 Bi-Directional DRM Neural Network

Given an input image S and a reference S', the $DecomNet$ and $SynthNet$ can work together to transform the input image S to a new image $S*$ which is similar to the reference image S' on the illumination levels. It provides a feasible way to either up-expose (increase the global illumination) or down-expose (decrease the global illumination) the original image by varying the reference images. However, to acquire both up-exposed and down-exposed images, we need to repeat the training process with different input-reference image pairs and use multiple sets (2 sets at least) of $DecomNet$s and $SynthNet$s to generate images with different exposures. Besides, these training processes are isolated, which means one input-reference image pair will only be used in one training and evaluation process.

To be able to synthesize images with different exposures during a single training and evaluation phase. We propose a bi-directional network structure which can take 3 images as input based on the $DecomNet$ in the training phase, and generate two images simultaneously with higher and lower exposures using one image in the evaluation phase. In this section, we will discuss the details of the bi-directional network structure and loss function. We also adopt a network structure similar to $SynthNet$ to generate enhanced illumination maps. These two network structures can be trained and evaluated end-to-end. The bracket image generation using the bi-directional network and methods for illumination map enhancement will be discussed in detail in this section. We will also discuss how this method can be generalized to enable multiple-image (more than 2 images) generation.

4.6.1 Network Design

The structure of the bi-directional network is illustrated in Figure 4.1. To train the network, three images with different exposures are required. We use three moulage images of different exposures from the IVDS dataset. Among the three images, image A has the highest illumination level (over-exposed) and image C has the lowest illumination level (under-exposed). The illumination level of image B sits between image A and C. Image B is used as the reference image for both image A and C.

In the training phase, we deploy 2 *DecomNet*s with all parameters shared. Image A and C are assigned to separate networks in training, while image B (reference image) can be assigned to either of the 2 networks. Since we do not modify the structure of *DecomNet*, the structure of each network stays unchanged (see Figure 4.2). In every training iteration, we calculate the combined loss of the 3 networks. The loss function is defined in Equation 4.21. For the simplicity of the equation, we use $\hat{I}_x$ and $\hat{R}_x$ to denote the illumination and reflectance map generated by the *DecomNet* using x as input. While $\lambda_n, n = 1, 2, ...$ are used as coefficients to balance each importance of each loss components.

$$\begin{aligned} L_{bi\_direc} = \lambda_1(L_{reconn}(\hat{R}_A, \hat{I}_A, A) + L_{reconn}(\hat{R}_B, \hat{I}_B, B) + L_{reconn}(\hat{R}_C, \hat{I}_C, C)) + \\ \lambda_2(L_{reconn}(\hat{R}_B, \hat{I}_A, A) + L_{reconn}(\hat{R}_C, \hat{I}_A, A) + \\ L_{reconn}(\hat{R}_A, \hat{I}_B, B) + L_{reconn}(\hat{R}_C, \hat{I}_B, B) + \\ L_{reconn}(\hat{R}_A, \hat{I}_C, C) + L_{reconn}(\hat{R}_B, \hat{I}_C, C)) + \\ \lambda_3(L_{smooth}(\hat{I}_A, \hat{R}_A) + L_{smooth}(\hat{I}_B, \hat{R}_B) + L_{smooth}(\hat{I}_C, \hat{R}_C)) \end{aligned} \tag{4.21}$$

Inspired by the training process of *DecomNet*, the loss we aim to minimize also contains three parts. The first part describes the quality of the reconstruction using the decomposed maps. This loss is designed based on the basic assumption of the Retinex Theory (see Equation 2.1). The second part describes how well the reflectance maps generated from images with different illumination conditions can be used to construct images. This loss is based on the assumption that the reflectance maps should be identical for images different only from illuminations (see Equation 4.10). The last part is used to ensure the illumination maps should be smooth in general. This loss can also be seen as the smoothing operation on the illuminance map.

The decomposed maps of the three images are used as inputs for the *SynthNet*. To be able to generate two illumination maps (both up-exposed and down-exposed versions), two instances of *SynthNet* are required. Contrary to the training process of *DecomNet*, the two *SynthNet*s are trained separately. The first instance uses decomposed maps from image A as input while the second one uses maps from image C. Both instances use the illumination map of image B as ground-truth. In this way, the first instance learns the transformation function to reduce the illumination density (down-exposure) while the second one learns how illumination density can be increased (up-exposure). The same loss function (see Equation 4.20) can be used in this process. In the final step, the illumination map is used together with the corresponding reflectance map to generate bracket images. We call the network the bi-directional network since it integrates the up- and down-exposure operation at once and can be used to generate bracket images with higher and lower exposures than the original images.

4.6.2 Discussion

Compared with training two completely different models for bracket image generation, our proposed bi-directional network has several advantages: 1) Data sharing. The original design of *DecomNet* can only be trained using two images (input and reference images). Our bi-directional design makes training multiple in a single phase possible. 2) Stable training. By using the information from three images, we can apply more constraints in the training process, which can lead to a more stable training process. 3) Decomposition consistency. Decomposition can yield more consistent results for images taken with a higher range of illumination densities.

Chapter 5

Exposure Fusion

In this chapter, we will discuss how the bracket images(images with different exposures) can be merged into an HDR-like image using exposure fusion. The generated image will take into consideration all texture information of bracket images and give preferable illumination conditions.

In order to prevent texture information loss and also enhance details in ill-exposed regions, we propose the detail-recovered fusion pipeline in exposure fusion is shown in Figure 5.1. Three images (down-exposed, up-exposed and the original image) are used as inputs. At the first stage, we use two transformer functions together with illumination-based fusion to generate a detail-enhanced version of each bracket image(labeled as *Enhanced Bracket Images*). The detail-enhanced bracket images are then fused again using illumination-based fusion to synthesize the HDR-like image.

We discuss the design of our proposed pipeline in the rest of the chapter. The following sections in this chapter are structured in the following way. First, we discuss illumination-based fusion, which is central in our implementation. We later show the limitation of applying illumination-based fusion directly in image generation. To

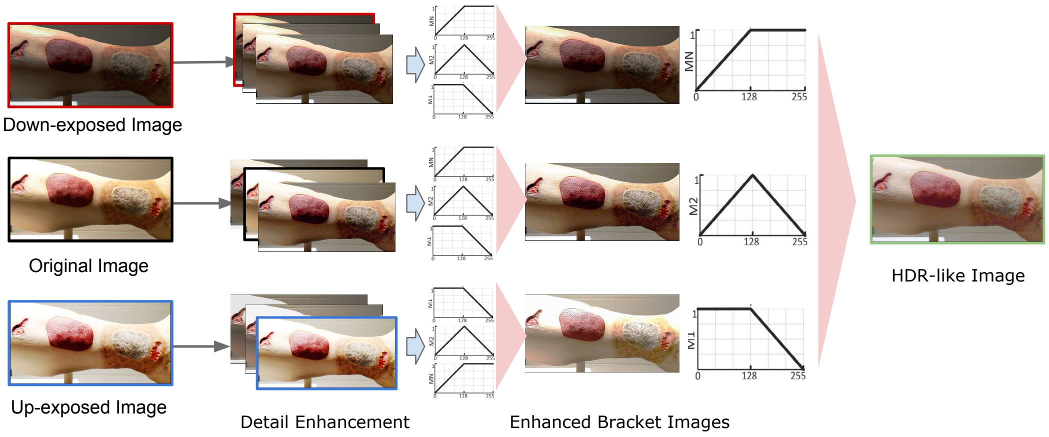


Figure 5.1: Exposure Fusion and HDR-like Image Generation

overcome the limitation of illumination-based fusion, we describe how we enhance image details using our designed transformer functions according to different image exposures. At last, we describe in detail how HDR-like images can be generated using our proposed pipeline.

5.1 Illumination-Based Fusion

Illumination-based fusion[46] is a kind of image fusion method which takes several images with different illumination conditions (or different exposures) and merges them into one image. Given three images S^-, S and S^+ with increasing exposures, we can design three membership functions (weights) for each image and use the weighted sum of the three images as the fused result.

The curves of the three membership functions are shown in Figure 5.2. The horizontal axis denotes pixel intensity in the illumination layer in any 8-bit images. For under-exposed images (images with low exposure), the bright regions are more emphasized than the dark regions since the bright regions contain texture information of the scene while the dark regions are mostly negatively saturated. On the other

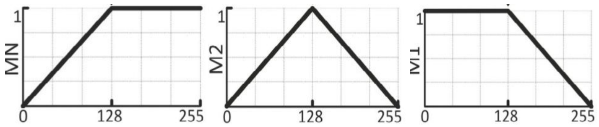


Figure 5.2: Membership Functions for Different Exposures. 1)left: under-exposed images. 2) middle: well-exposed image 3) right: over-exposed image

hand, for over-exposed images (images with high exposure), the saturated pixels are mostly included in the bright regions while texture details are more significant in the dark regions. Thus the dark regions are emphasized relative to the bright regions in the over-exposed images. For well-exposed images (exposure between over- and under-exposure), regions either too bright or too dark are penalized while we only concentrate on the regions with moderate illumination intensities.

To apply the illumination-based fusion, we need to obtain the illumination layer from each of the three images. Similar to the illumination map which is generated from image decomposition based on the Retinex Theory, illumination layer can also be used as an estimator of the illumination. However, illumination layer doesn't follow the assumptions (relationship between reflectance, smoothing assumptions, and etc.) made in the Retinex Theory, also it doesn't convey information on the intrinsic properties of the objects in the scene. However, it serves the fusion purpose well because of its simplicity and compatibility in our pipeline.

Three images with different exposures are first converted into the $YCbCr$ space. The first channel (the Y channel) encodes the information of the illumination, while the other two channels control the color information in the image. We take the Y channel only and use it to calculate the weight matrix using the membership functions mentioned above. For an image S with size $h \times w \times 3$, the resulted weight matrix can be calculated using Equation 5.1.

$$\boldsymbol{W}_{(i,j)} = F_{mem}(Y(\boldsymbol{S})_{(i,j)}), 1 \leq i \leq h, 1 \leq j \leq w \tag{5.1}$$

$Y(\boldsymbol{S})$ denotes the Y channel (shape of Y channel is $h \times w$) of the image S in the $YCbCr$ space. We use $Y(\boldsymbol{S})_{(i,j)}$ to represent the scalar value on the row i and column j from the Y channel. This value is then passed to one of the three membership functions based on the illumination intensity level. The calculated value from the membership function is used to construct the weight matrix, with the scalar value on the row i and column j from the weight matrix W equals to the value calculated from the membership function.

The weight matrices for the three images (S^{-}, S and S^{+}) are denoted as $\boldsymbol{W}^{-}$, $\boldsymbol{W}$ and $\boldsymbol{W}^{+}$. They all have the same shape ($h \times w$) since the shapes of the three input images are the same ($h \times w \times 3$). We calculate the weighted sum using the three weight matrices. This operation is shown in Equation 5.2.

$$\boldsymbol{S}_{weighted} = (\boldsymbol{W}^{-} \otimes \boldsymbol{S}^{-} + \boldsymbol{W} \otimes \boldsymbol{S} + \boldsymbol{W}^{+} \otimes \boldsymbol{S}^{+}) \otimes (\frac{1}{\boldsymbol{W}^{-} + \boldsymbol{W} + \boldsymbol{W}^{+}}) \tag{5.2}$$

We use the symbol $\otimes$ to represent the element-wise multiplication. The image matrices are multiplied with their corresponding weight matrices in an element-wise fashion. The latter part $\frac{1}{\boldsymbol{W}^{-}+\boldsymbol{W}+\boldsymbol{W}^{+}}$ means we first add the three weight matrices, and calculate the element-wise reciprocal (calculate the reciprocal for each element in the result matrix), which is different from the inverse matrix of the sum of three weight matrices.

The final step is to normalize the image to make the all the pixel values are within a certain range (e.g. 0-255). The pseudo-code of the fusion process is shown in Algorithm 5.1.

Algorithm 1 Illumination-based Fusion

```
procedure FUSE($img_+, img, img_-$)        ▷ RGB images with high/median/low exposures
    $rgb\_list \leftarrow \{img_+, img, img_-\}$
    $weight\_list \leftarrow \{\}$
    for $rgb$ in $rgb\_list$ do
        $ycbcr \leftarrow$ convert $rgb$ to YCbCr space.
        $\boldsymbol{Y}ch \leftarrow$ get the Y channel from $ycbcr$.
        $weight\_list$.add(get_weight_matrix($\boldsymbol{Y}ch$))
    end for
                                    ▷ weight_list contains weights for rgb_list
    $img^* \leftarrow$ weighted sum of $rgb\_list$
    $img^* \leftarrow$ normalized($img^*$)
    return $img^*$▷ Fused image. Clip the pixel values within the range of $(0-1)$
end procedure
```

5.2 Direct Fusion

5.2.1 Fusion Concept

To generate the HDR-like image, we synthesize two bracket images with the bidirectional network given an input image. The input image S is up-exposed to an image with a higher illumination level S^+ and down-exposed to an image with a lower illumination level S^-. Using the illumination-based fusion, we can directly generate the fused result.

The process of direct fusion is shown in Equation 5.3. We use S_F to denote the resulted image from direct fusion. The implementation of function $FUSE()$ is described in Algorithm 5.1.

$$S_F = FUSE(S^+, S, S^-) \tag{5.3}$$

5.2.2 Limitations of Direct Fusion

While directly applying the illumination-based fusion can integrate the most textural informative regions into the resulting image, there are some limitations for over- and under-exposed images. In over-exposed images, a large portion of the image regions can be saturated (the values of pixels are approaching 255). The saturated regions cannot provide any viable information for analysis. However, these highly saturated pixels can still have a positive weight when being used for fusion as long as the pixel value is not as high as 255. Similar problems occur on the under-exposed images. When fusing the under-exposed images, some of the dark regions are negatively saturated (the values of pixels are approaching 0), but as long as the pixel values are not 0, a small positive weight will be applied in the fusion process.

Every pixel value of the fused result $\boldsymbol{S}_{weighted}$ can be seen as the weighted sum of three pixel values taken from the three images at the same pixel location. The relationship can be represented in Equation 5.4. For pixel located at (i, j), if it is saturated in S^+ with no information gain for the full image, the pixel value $\boldsymbol{S}_{weighted(i,j)}$ computed using $\boldsymbol{S}^-_{(i,j)}$ and $\boldsymbol{S}_{(i,j)}$ can be seen as diluted because of the effect of weighted average. Similarly, if the pixel is negatively-saturated on S^-, the pixel value will also be diluted because only $\boldsymbol{S}_{(i,j)}$ and $\boldsymbol{S}^-_{(i,j)}$ are providing texture information and the averaged value is affected by $\boldsymbol{S}^-_{(i,j)}$.

$$\boldsymbol{S}_{weighted(i,j)} = \frac{\boldsymbol{W}^-_{(i,j)}\boldsymbol{S}^-_{(i,j)} + \boldsymbol{W}_{(i,j)}\boldsymbol{S}_{(i,j)} + \boldsymbol{W}^+_{(i,j)}\boldsymbol{S}^+_{(i,j)}}{\boldsymbol{W}^-_{(i,j)} + \boldsymbol{W}_{(i,j)} + \boldsymbol{W}^+_{(i,j)}} \tag{5.4}$$

The dilution of the pixels causes the degradation of the contrast for the whole image. It is most obvious if the over- or under-exposed regions are large. Because of this limitation, the fused image may degrade heavily on contrast in highly saturated or negatively saturated regions. Figure 5.3 shows the comparison between the

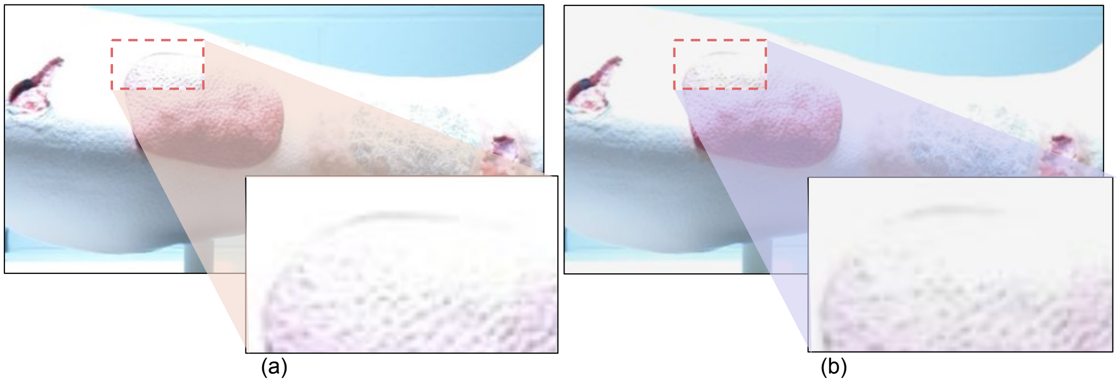


Figure 5.3: The comparison between the original input image (a) and the directly fused result (b). We can clearly notice that some texture details in the wound area are lost after direct fusion.

original image S and the direct fused image using illumination-based fusion. The original image is highly over-exposed in some regions. In these regions, the fused image even loses more texture details because of the effect of dilution.

5.3 Detail Enhancement

To address the texture information loss using illumination-based fusion, we propose a method to enhance the texture details before performing fusion. The detail enhancement is used as one additional step to reduce the contrast loss in illumination-based fusion. Two transformer functions are used to enhance the details, for over- and under-exposed images respectively.

5.3.1 Transformer Functions

The transformer function for the over-exposed images is defined in Equation 5.5 while the one for the under-exposed images is defined in Equation 5.6. As gamma correction is commonly used to improve contrast in dark images[24], the design of TF_- and TF_+ are based on gamma correction with TF_- enhancing the details in over-exposed and TF_+ enhancing the under-exposed regions.

Parameter img is a $h \times w \times 3$ matrix representing the input RGB image. And λ is a small custom parameter. As for $TF_+(img, \lambda)$, λ are the inverse of γ in gamma correction. $TF_-(img, \lambda)$ works by applying gamma correction on the color-inverted version of the input rather than the input itself. Then, we invert the color after gamma correction to generate the output.

$$TF_-(img, \lambda) = 1 - (1 - img)^{\frac{1}{\lambda}} \tag{5.5}$$

$$TF_+(img, \lambda) = img^{\frac{1}{\lambda}} \tag{5.6}$$

These two transformer functions are used to generate detail-enhanced copies of the original image. For an over-exposed image, $TF_-(img, 1)$, $TF_-(img, 2)$ and $TF_-(img, 3)$ are generated and similarly, we will generate $TF_+(img, 1)$, $TF_+(img, 2)$ and $TF_+(img, 3)$ for an under-exposed image. One thing to notice is: when the value of λ equals to 1, either $TF_-(img, \lambda)$ or $TF_+(img, \lambda)$ will return the original image without any modification. So, in our case, there are two newly-generated images for any input image, and we can use the three images (2 newly-generated ones and the original one) to synthesize the detail-enhanced image.

Since the two transformer functions will also change the illumination of the input image when enhancing the details, we use the illumination-based fusion again in the last step of detail enhancement. The three images ($TF_-(img, 1)$, $TF_-(img, 2)$ and $TF_-(img, 3)$ for over-exposed image, $TF_+(img, 1)$, $TF_+(img, 2)$ and $TF_+(img, 3)$ for under-exposed image) are synthesized using the illumination-based fusion. The fusion process for over- and under-exposed images can be represented in Equation 5.7 and 5.8 respectively. $FUSE$ is the illumination-based function which we defined

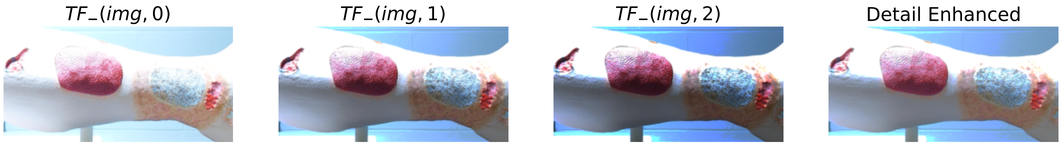


Figure 5.4: Intermediate Results and Enhanced Result (over-exposed image)

in Algorithm 5.1. img_E^+ and img_E^- mean the enhanced versions of the over- and under-exposed images. This process is illustrated in Figure 5.1, where images with various exposures are processed with the two transformer functions before being fused using the illumination-based fusion. The three generated images (*Enhanced Bracket Images*) are detail-enhanced versions of the former ones.

$$img_E^+ = FUSE(TF_-(img, 1), TF_-(img, 2), TF_-(img, 3)) \tag{5.7}$$

$$img_E^- = FUSE(TF_+(img, 3), TF_+(img, 2), TF_+(img, 1)) \tag{5.8}$$

5.3.2 Output of Transformer Functions

Figure 5.4 and 5.5 show the original images, images transformed using the transformer functions together with the detailed enhanced image. In Figure 5.4, the input image $TF_-(img, 0)$ is over-exposed, so we use function TF_- to recover the details in the over-exposed regions. Similarly, in Figure 5.5, we use function TF_+ to recover the details in the under-exposed region of the input image $TF_+(img, 0)$. We can clearly see the detail-enhanced images contains more texture information in the ill-exposed regions for over- and under-exposed image input.

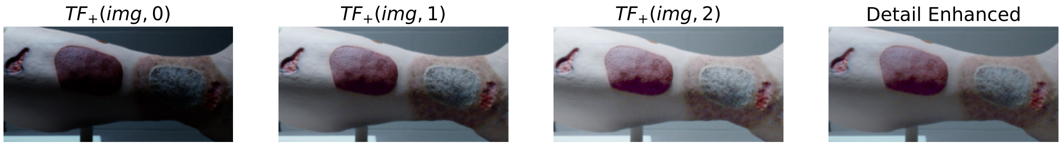


Figure 5.5: Intermediate Results and Enhanced Result (under-exposed image)

5.4 Detail-Recovered Fusion

To generate the HDR-like image with preferable illumination and texture details recovered in the ill-exposed region, we propose the detail-recovered fusion which can take 3 bracket images as input. With the knowledge of detail enhancement for both over- and under-exposed images as well as illumination-based fusion, the fusion process is relatively straightforward. For the 3 bracket images (S^+, S, S^-), we first apply different detail enhancement methods based on their illumination conditions. TF_+ is used to process the image S^- and the detail-enhanced version of S^- can be obtained. Similarly, we obtain the detail-enhanced version of S^+ using TF_-. For image img, we can use both TF_+ and TF_- to enhance details in both over- and under-exposed regions (see Equation 5.9).

$$img_E = FUSE(TF_+(img, 1), img, TF_-(img, 1)) \tag{5.9}$$

When we acquire the detail-enhanced versions of three images (img_E^-, img_E and img_E^+), we apply the illumination-based fusion on these enhanced images to get the detail-recovered result (see Equation 5.10).

$$img_{DR} = FUSE(img_E^+, img_E, img_E^-) \tag{5.10}$$

The pseudo-code of this algorithm is shown in Algorithm 2. And the pipeline of detail-recovered fusion is illustrated in Figure 5.1.

Algorithm 2 Detail-recovered Exposure Fusion

```
procedure DR-FUSE($img_+, img, img_-$)    ▷ Up/Original/Down exposed images
    $img\_list_+ \leftarrow \{\}$
    $img\_list_- \leftarrow \{\}$
    $img\_list \leftarrow \{\}$
    for $\lambda$ in 1..3 do
        $img\_list_-$.add($TF_-(img_-, \lambda)$)
        $img\_list_+$.add($TF_+(img_+, \lambda)$)
    end for
    $img\_list$.add($TF_+(img, 2)$)
    $img\_list$.add($img$)
    $img\_list$.add($TF_-(img, 2)$)
                                        ▷ Perform first level fusion on 3 images
    $img_+^{(1)} \leftarrow$ FUSE($img\_list_+$)
    $img_-^{(1)} \leftarrow$ FUSE(reversed($img\_list_-$))
    $img^{(1)} \leftarrow$ FUSE($img\_list$)
    $img^* \leftarrow$ FUSE($img_+^{(1)}, img^{(1)}, img_-^{(1)}$)    ▷ Perform second level fusion
    return $img^*$                                                 ▷ Fused image.
end procedure
```

5.4.1 Global Illumination Adjustment

To mitigate the change in global illumination in detail-recovered fusion, we introduce a global illumination multiplier ψ to adjust the lighting condition. We define ψ as follows.

$$\psi = \mathbf{min}\left(\frac{1}{\mathbf{max}(img_{DR})}, \frac{\mathbf{max}(\mathbf{avg}(img), C)}{\mathbf{avg}(img_{DR})}\right) \tag{5.11}$$

img is the original image and img_{DR} is the fused image. C is a constant that controls the average pixel values of the fused image in case the original image is too dark or too bright. We also limit pixels of the adjusted image to the $(0-1)$ range by not allowing the value of ψ exceeding $1/\mathbf{max}(img_{DR})$. We set $C = 0.6$ for all

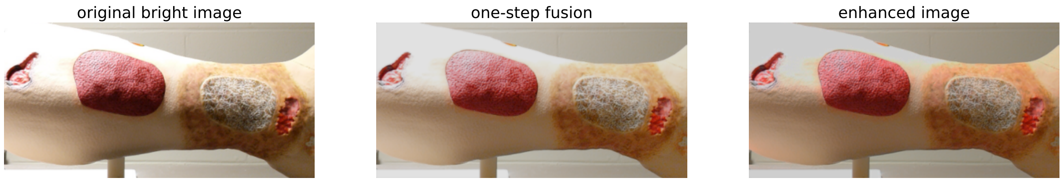


Figure 5.6: Visual Comparison(Over-Exposed Images)

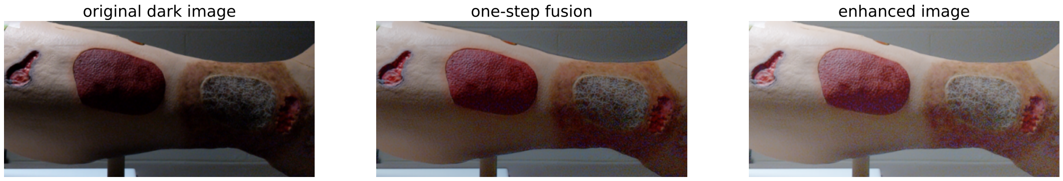


Figure 5.7: Visual Comparison(Under-Exposed Images)

our experiments.

The multiplier ψ changes the illumination by scaling the pixel values linearly. It also ensures every scaled pixel in the resulted image will always stay in the range $(0 - 1)$, so no information will be lost due to value clipping.

The fused result after we adjust the global illumination can be denoted as $img_{DR(\psi)}$ (see Equation 5.12).

$$img_{DR(\psi)} = \psi \times img_{DR} \tag{5.12}$$

5.4.2 Visual Comparison of Fused Results

We visually compare the fused images generated using the direct approach (one-step fusion) and the detail-recovered approach. We use over- and under-exposed images as inputs and compare the fused results. The visual comparisons are shown in Figure 5.6 and 5.7. More detailed comparisons will be discussed in the next chapter.

We can notice the images generated using detail-recovered are visually better than the directly fused images in over- and under-exposed illumination conditions. Specifically, for the over-exposed image, the recovered fused image provides more smooth-lit texture details in the wound areas, a larger portion of over-exposed regions is also recovered. For under-exposed images, the recovered fused image is not only more smooth in light, but it also contains far less noise than the directly fused image.

Chapter 6

Experiments and Evaluation

In this chapter, we discuss the experiments we have conducted in detail. We also included performance evaluations of our proposed method using several metrics such as SSIM, PSNR, dice coefficient, and etc.

6.1 Dataset Preparation

In our previous work[20], the Illumination Varying Dataset (IVDS) is created to analyze the correlation between image illumination and segmentation accuracy. It consists of 7 separated sub-datasets and each sub-dataset contains 7920 images of a wound moulage captured under various illumination conditions/smartphone cameras/parameters. The moulage is illustrated in Figure 6.1.

The illumination included 33 different light sources, 16 light color temperatures ranging from 3200K to 5600K and 15 light intensities ranging from 10 to 255. The location of the light source is attached on a controllable moving platform and the color temperature is controlled using Python. The hardware settings are shown in Figure 6.2. And the details of IVDS dataset are listed in Table 6.1.

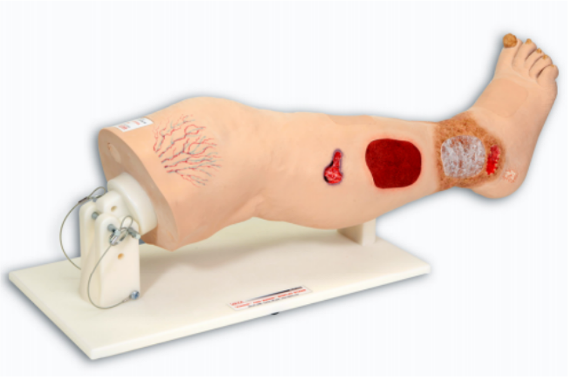


Figure 6.1: Moulage Used in IVDS Creation[20]

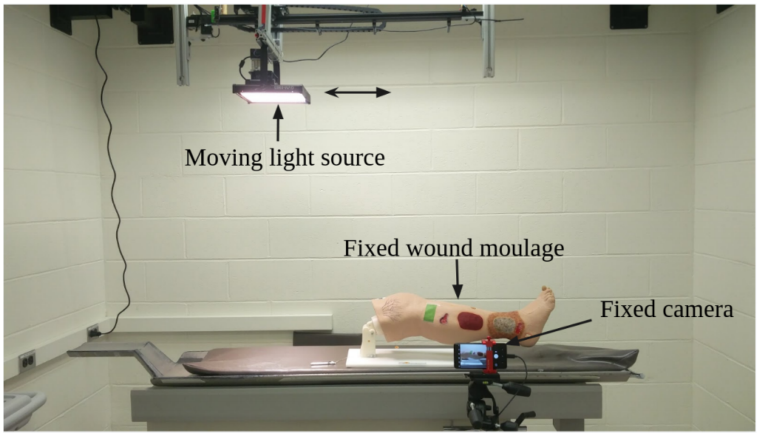


Figure 6.2: Hardware Settings Used in IVDS[20]

6.2 Data Preprocessing

The datasets we used in our experiments are from the **Wound Illumination Varying Dataset** (IVDS) which was created in our previous work[20].

We aim to test the performance of our proposed method on both over- and under-exposed images. For sub-datasets created using automatic configurations, the exposure was automatically adjusted based on the environmental lighting conditions so the images taken in different light source intensities do not show significant variations on the image illumination. On the other hand, the images in sub-datasets created with manual configuration preserve a relatively positive correlated relation-

Table 6.1: The Illumination Varying Dataset (IVDS)

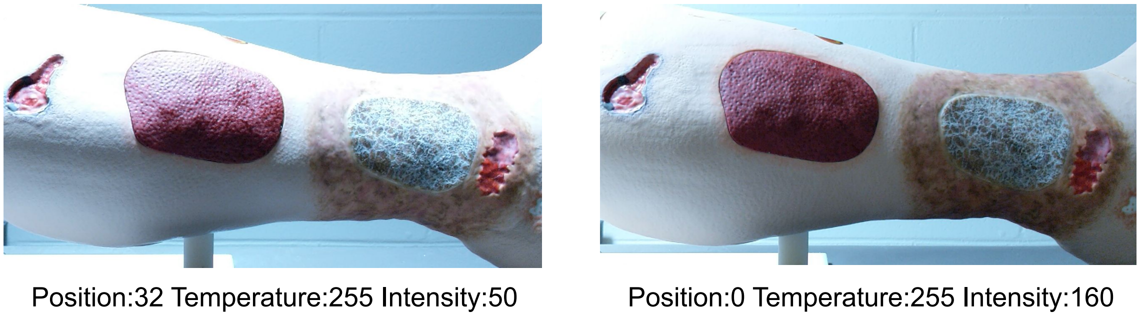
Dataset Name	**Camera Phone**	**Mode**	**No.of Images**
Pixel-M1	Pixel2 XL	Manual	7920
Pixel-M2	Pixel2 XL	Manual	7920
Pixel-Auto	Pixel2 XL	Automatic	7920
Pixel-WBDL	Pixel2 XL	Automatic	7920
HTC-Auto	HTC Desire 816	Automatic	7920
Moto-Auto	Moto G5 Plus	Automatic	7920
Samsung-Auto	Samsung J7 V	Automatic	7920

ship between light source intensity and image illumination. So in order to reduce the bias caused by the automatic adjustment from the smartphone camera and also be able to test images with high illumination variations, we chose *Pixel-M1* and *Pixel-M2* which were the only two sub-datasets captured in manual mode.

6.2.1 Splitting Images Based on Illuminations

Dataset *Pixel-M1* and *Pixel-M2* contain images different in light source intensities and parameters. Since we want to treat images with over- and under-exposed issues separately, the first step was to split the images based on the image illumination.

Please note we cannot classify images solely based on the light source intensities since the light source position can play an essential role in the image illumination. In other words, when the light source is placed away from the target wound moulage, the captured image would still be dark in general even though the light source intensity is high. And when the light source is placed near the center of the wound moulage, the captured images tend to be over-exposed even the light source intensity is relatively moderate. We use Figure 6.3 to illustrate this issue. The image on the left side looks much brighter (contains more over-exposed regions) than the image on the right side, however, the light source intensity of the left-sided image is much smaller than the intensity of the image on the right side.

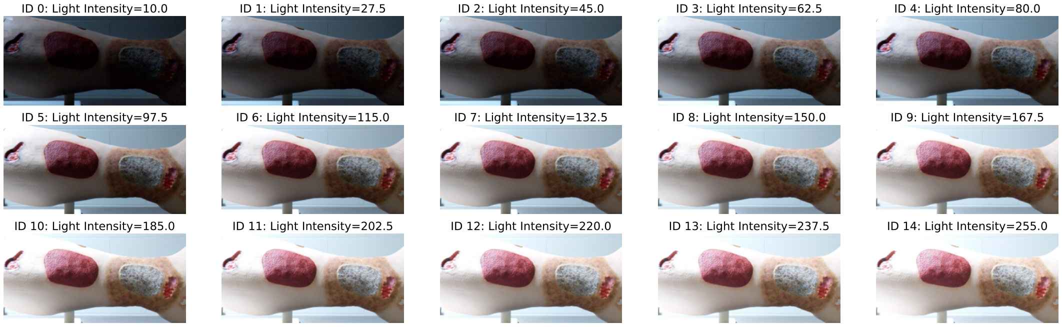


Position:32 Temperature:255 Intensity:50

Position:0 Temperature:255 Intensity:160

Figure 6.3: Two Images with Different Light Source Intensities

We group images based on the (camera, parameters) pair. In each group, there are 15 images which are only different in their light source intensities, and we can label images as "bright", "normal" and "dark". Figure 6.4 shows all 15 images in one (camera, parameters) group captured in light source position 7 and light temperature index 136 using the Pixel 2 smartphone camera. In this typical case, we labelled image $0-2$ as "dark", image 5 as "normal" and image $9-14$ as "bright".



ID 0: Light Intensity=10.0
ID 1: Light Intensity=27.5
ID 2: Light Intensity=45.0
ID 3: Light Intensity=62.5
ID 4: Light Intensity=80.0
ID 5: Light Intensity=97.5
ID 6: Light Intensity=115.0
ID 7: Light Intensity=132.5
ID 8: Light Intensity=150.0
ID 9: Light Intensity=167.5
ID 10: Light Intensity=185.0
ID 11: Light Intensity=202.5
ID 12: Light Intensity=220.0
ID 13: Light Intensity=237.5
ID 14: Light Intensity=255.0

Figure 6.4: 15 Images in IVDS only different in light source intensities(light source position 7, light temperature index 136). Light intensities from 10 to 255 in 15 steps.

An alternative way[18] to automate this process is to classify images based on the histogram in the illumination layer. We need to set up four thresholds ($Th1-Th4$) and the four thresholds can help us determine which type of image it should be. The four thresholds are fixed values which divide the illumination histogram into 5

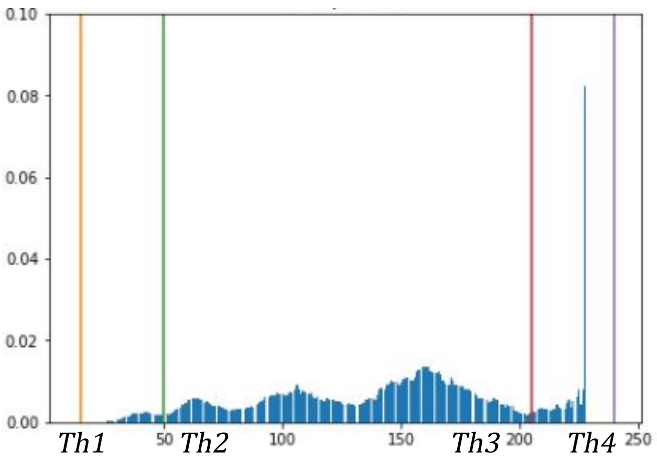


Figure 6.5: Four Thresholds on Illumination Layer

Table 6.2: Conditions of Image Types

Type	**Description**
dark	$Lo > 2Hi$
bright	$Hi > 2Lo$
extreme	$\frac{1}{2} \leq \frac{Hi}{Lo} \leq 2$ and Hi and $Lo > \frac{P_{total}}{4}$
normal	others

intervals (see Figure 6.5). The 4 types can be inferred from this approach are *dark*, *bright*, *extreme* and *normal*, which are determined by two parameters (Hi and Lo). The conditions of these four types are summarized in Table 6.2.

The calculation of the two parameters (Hi and Lo) are shown in Equation 6.1 and 6.2. The function hi means the number of pixels whose values are equal to i. In our experiment, we set $Th1 = 15$, $Th2 = 50$, $Th3 = 205$ and $Th4 = 240$.

$$Hi = \sum_{i=Th3}^{Th4} h(i) + \sum_{i=Th4}^{255} h(i) \times 2 \tag{6.1}$$

$$Lo = \sum_{i=0}^{Th1} h(i) \times 2 + \sum_{i=Th1}^{Th2} h(i) \tag{6.2}$$

Similar to how we manually label each image, we first label all the images using the histogram-based approach. We then draw two samples with the same size from the ”dark” and ”bright” images. One drawback is for some images with the same (camera, parameters) pair, there is no image in the 15 images labeled as ”normal”. In such cases, we need to either manually pick one image as the ”normal” image or ignore the 15 images with the same (camera, parameters) pair completely. In our following experiments, we labeled images manually.

6.2.2 Setting Up Train/Test Images

To set up the training and test images, we use 80% of the images for training and the remaining 20% for evaluation. In order to keep the relationship of images sampled from the same (camera, parameter) group, we don't sample the images directly, instead, we only sample the (camera, parameter) groups, and within each sampled group, we take samples from the ”bright” and ”dark” images. This operation can be described in the following steps.

1. From each dataset, we randomly choose 80% groups from all groups (422 out of 528 groups are chosen).

2. In each group we have chosen, we take N random samples from ”dark” images and N random samples from ”bright” images. Both are without replacement. Of the 15 images, we pick the image with the best exposure as the ”normal”. The ”normal” image should contain the least over- and under-exposed regions compared with other images.

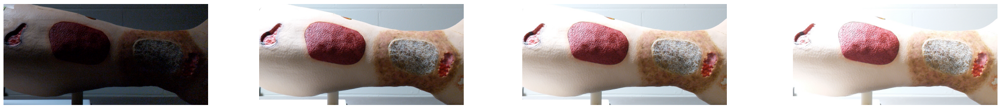
3. Use all sampled images from the chosen groups as the training set.

4. For the rest 20% groups, we perform the similar step (step 2) and use the images as the test set.

6.3 HDR Target Generation

To evaluate the performance of our proposed method, we need to generate the HDR wound image as the ground truth. Four LDR images were used to synthesize the HDR target. In our experiment, the light source intensities of selected four LDR images are 10, 80, 140 and 250. All of them are from the *Pixel-M1* dataset with the same light source position(index number 9) and light temperatures(index number 102). These four images are shown in Figure6.6. The inverse camera response function for the RGB channels is shown in Figure6.7.

We compared the *Drago*, *Mantiuk* and *Reinhard* tone mappers for converting the HDR wound image into a HDR-like image (shown in Fig.6.8 with parameters). *Drago* performed the best, with minimal over/under exposed areas and it preserved most texture details. Although it's feasible to generate the HDR target using images from each (camera, parameter) group and use different HDR targets just for that group, we noticed that in our experiments many generated HDR targets did not provide enough details. This issue happened frequently because images in a same group can suffer from similar exposure issues (both over- and under-exposed issues) simultaneously, which leads to texture information loss in a common image region for all images in the group. With little information in some common regions in all images, it becomes very challenging for the generated HDR target to recover the lost information. Besides, to obtain the tone-mapped version of the HDR target image, the parameters used in the tone-mapper are usually different for HDR images generated from LDR images with various illumination conditions, which can make it hard to choose the best image in each group. Apart from that, comparing with different HDR targets in different groups is actually measuring how close our synthesized result is relative to the HDR target in each group, no matter how much



(a) intensity=10 (b) intensity=80 (c) intensity=140 (d) intensity=250

Figure 6.6: LDR Images Used for HDR Target Creation

texture information is preserved in the generated HDR targets, which is different from our goal that we want to restore the lost texture information in over- and under-exposed images.

On the other hand, the single HDR target makes the bench-marking process much more consistent. All HDR images converted from ill-exposed images using our method as well as other methods can be compared using the same target HDR image. Besides, the HDR target is generated with the idea that as much as texture details should be preserved in the image. So the degree of resemblance between the synthesized image and the HDR target can be used as an indicator to tell how much texture details are recovered from the original ill-exposed image.

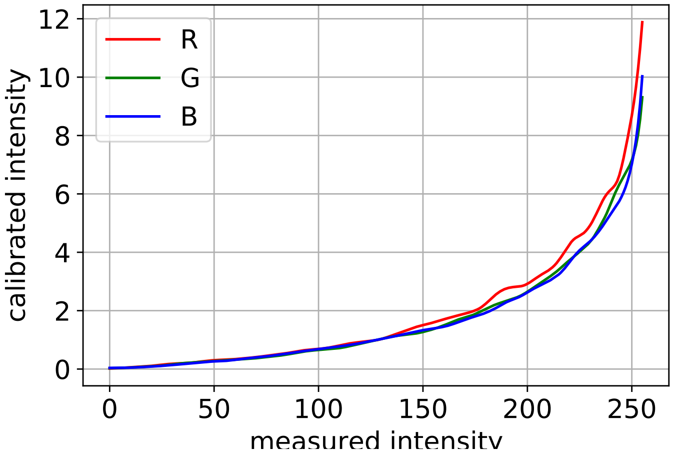
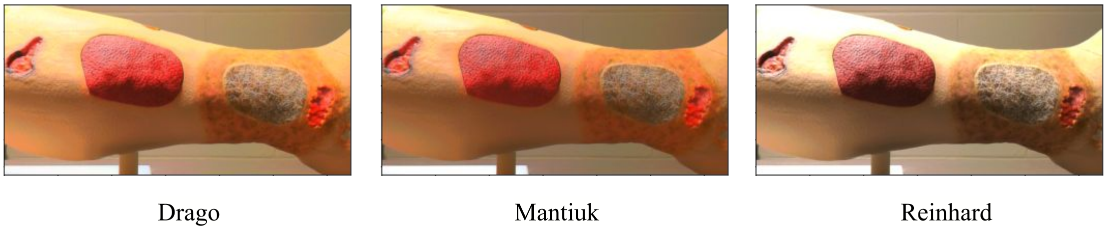


Figure 6.7: Inverse Camera Response Function for RGB Channels. The exposure times we used are 1/30, 8/30, 14/30 and 25/30 seconds.


Drago

Mantiuk

Reinhard

Figure 6.8: HDR Targets Generated with Three Tone Mappers. 1) Drago (gamma=1.6, saturation=1.45); 2) Mantiuk(gamma=1.6, scale=0.8, saturation=1.45); 3) Reinhard(gamma=1.6, intensity=-8, lightAdapt=0.5, colorAdapt=0); The HDR-like image generated by **Drago** was used as the target to calculate image similarities.

6.4 Training of Bi-Directional Network

The Bi-Directional network (see Figure 4.1) can convert an input image into two images with higher and lower exposures. The network can be logically split into two models when training, which are up-exposure model and down-exposure model. These two models have the exact network components (decom-net + synth-net), and the parameters of decom-nets of both models are shared (same network parameters), but they have its own synth-net with different parameters. And up- and down-exposure model contain the workflows of increasing or decreasing the exposures of the original image respectively.

We generate training pairs using the labeled data with permutation. Each pair consists of one "bright" image, one "normal" image, and one "dark" image. We always make sure the three images belong to the same (camera, parameter) group, which means except for the light intensities, all the configurations remain the same for the three images in each pair. The "bright"/"normal" images are used to train the down-exposure model while the "dark"/"normal" images are used to train the up-exposure model. In our experiment, we use 1267 pairs for training and 317 pairs for testing. All the images are re-scaled to $512 * 256$ for fast training and testing.

Table 6.3: Hardware Specification and Software Platforms

Type	**Description**
Processor Model	Intel(R) Xeon(R) CPU
Processor Count	4
Processor Frequency	2.30GHz
Socket Count	1
Core(s) per socket	1
Thread(s) per core	2
L1d cache	32K
L1i cache	32K
L2 cache	256K
L3 cache	46080K
GPU Model	Tesla P100-PCIE-16GB
GPU Count	1
GPU Video Memory	16 GB
RAM	25.5 GB
Disk	150 GB SSD
Operating System	Ubuntu 18.04 LTS 64-bit
Deep Learning Platform	Tensorflow 2.4.1

The specs of the machine and software platforms used for training and test are listed in Table 6.3. We trained the bi-directional network for 100 epochs in total that took approximately 20 minutes. After the hyper-parameter search, we utilized a batch size of 16, patch size of 48, the Adam optimizer and a learning rate of 0.001.

6.5 Image Measuring Metrics

The purpose of image measuring metrics is to compare how similar or how close two images are to each other. The common use case is to measure the quality of reconstructed images. It is also considered a method to assess image quality.

6.5.1 Structural Similarity Index Measure

Structural Similarity Index Measure (SSIM)[50] is a very popular metric to measure the similarity between two images designed by modeling image distortion as a combination of three factors: loss of correlation, luminance distortion and contrast distortion[17]. Given a test image S and the ground-truth image (reference image) G, these three factors are multiplied together to get the SSIM value (see Equation 6.3). In the equation, $l(S, G)$ measures the closeness of two images' average illumination. $c(S, G)$ measures similarity in contrast between these two images, while $s(S, G)$ compares the structural information. For these three parameters, a higher value (maximal value is 1) indicates higher resemblance(or more similarity) while a lower value (minimal value is 0) means higher distortion (larger difference).

$$SSIM(S, G) = l(S, G)c(S, G)s(S, G) \tag{6.3}$$

The definitions of these three parameters are listed in Equation 6.4. The average luminance of both images (μ_S and μ_G) are used to measure the closeness of lumination in $l(S, G)$. When the values of $\mu_S and \mu_G$ are equal, $l(S, G)$ reaches its maximal value (which is 1). $c(S, G)$ is quite similar in structure to $l(S, G)$, however, the standard deviation of both images (σ_S and σ_G) are used to measure the differences in contrast. In $s(S, G)$, the covariances $\sigma_{S,G}$ between the two images and their standard deviations are used to calculate the correlation coefficient. C_1, C_2 and C_3 are positive constants.

$$\begin{aligned} l(S, G) &= \frac{2\mu_S\mu_G + C_1}{\mu_S^2 + \mu_G^2 + C_1} \\ c(S, G) &= \frac{2\sigma_S\sigma_G + C_2}{\sigma_S^2 + \sigma_G^2 + C_2} \\ s(S, G) &= \frac{\sigma_{S,G} + C_3}{\sigma_S\sigma_G + C_3} \end{aligned} \tag{6.4}$$

6.5.2 Mean Squared Error

Mean squared error (MSE) is not a metric typically utilized in assessing image quality. However, it is also a widely used method to calculate the differences between two images[44, 17]. The definition of MSE is straightforward – differences of intensities in every pixel are compared between images S and G, both of the same size $H \times W$. Equation 6.5 shows how MSE of these two images is calculated.

$$MSE(S,G) = \frac{\sum_{i=1}^{H}\sum_{j=1}^{W}(S_{(i,j)} - G_{(i,j)})^2}{HW} \tag{6.5}$$

One main difference between SSIM and MSE is: MSE calculates the absolute errors in images while SSIM is not directly related to differences in pixel intensities. In our experiments, MSE is not used to compare the generated images and the ground-truth image. A similar metric Mean Absolute Error (MAE) is used in the design of our reconstruction loss in *DecomNet* (see Equation 4.2).

6.5.3 Peak Signal-to-Noise Ratio

Peak Signal-to-Noise Ratio (PSNR) is another metric to measure the quality of reconstructed images in comparison to the ground truth. Using the reference image (noise-free image) G and the image S (encoded using 8 bits), PSNR is defined in Equation 6.6.

$$PSNR(S,G) = 10\log_{10}(\frac{255^2}{MSE(S,G)}) \tag{6.6}$$

Image S is usually a noisy approximation of image G. PSNR is used as a metric to reflect how much noise is contained in the image S. This metric also measures the absolute errors(based on MSE),

6.6 Dice Coefficient Segmentation Metric

Dice Similarity Coefficient(DSC)[11] is a metric for measuring segmentation accuracy compared to ground truth. Unlike the other metrics, DSC is not a direct method to compare two images. Instead, DSC takes the segmentation result S' generated from the test image S, and compares it with the a prepared segmentation result G'.

The segmentation results (which are sometimes called masks) can be viewed as a binary matrix which has the same shape as the input image. Pixels are labeled as either 1 or 0 manually or from model prediction. All the pixels with value 1 constitute the region of interest.

Calculation of DSC is illustrated in Equation 6.7. TP, FP and FN are abbreviations of true positive, false positive and false negative, which can be obtained by comparing the segmentation result S' with the ground-truth result G'.

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{6.7}$$

We use DSC obtained from U-Net trained on the IVDS dataset to locate wound areas in images. Taking wound segmentation as a typical example in wound image assessment, we aim to demonstrate improvement in performance in wound image assessment methods by comparing the DSC scored using original images and images enhanced using our proposed method.

6.7 Performance Evaluation Using SSIM

In this section, we evaluate the performance of our proposed bi-directional model using Structural Similarity Index (SSIM)[50] as the metric. SSIM is a popular method to compare the similarity between two images. The value of SSIM ranges from 0 to 1, while a higher value means higher resemblance. We calculate the SSIM scores before and after HDR conversion and compare the changes in distribution.

The experiments are conducted in the following steps.

1. For an ill-exposed (either over- or under-exposed) image img, we calculate the SSIM score between the HDR target img_T and img.

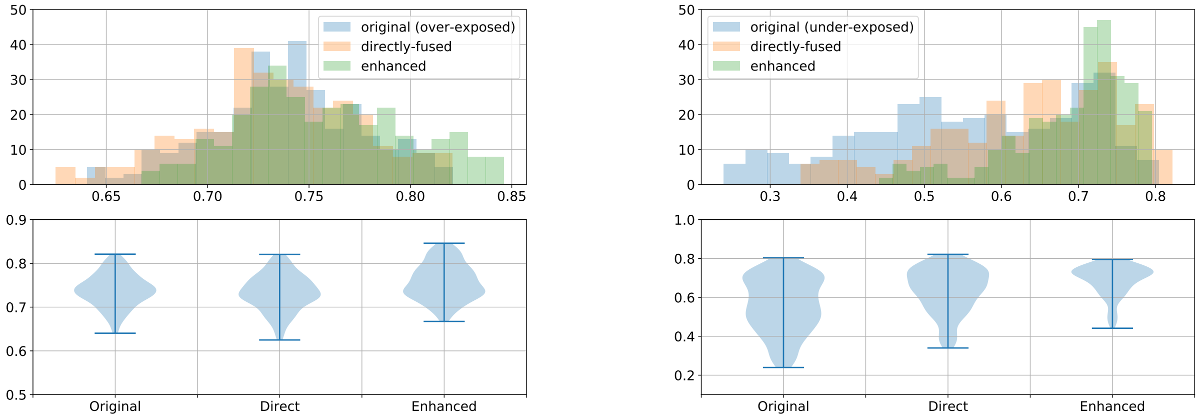
2. We generate the bracket images using the trained bi-directional network, which are img^+ (up-exposed) and img^- (down-exposed).

3. The generated bracket images together with the original image img are fused using detail-recovered fusion. The fused result (HDR-like image) is img_{DR}.

4. The SSIM score between the HDR target img_T and the fused result img_{DR} is calculated.

5. We compare the two SSIM scores and record the results.

Figure 6.9 shows the comparison of SSIM scores for over- and under-exposed images. The values labeled with *original* are SSIM scores calculated using the original ill-exposed image img and the ones labeled with *enhanced* are SSIM scores using the fused result img_{DR} (HDR-like image). We also included the SSIM scores between the HDR target img_T and the image synthesized directly using illumination-based fusion, which are labeled as *directly-fused.* The HDR-like images generated achieved higher similarity scores, indicating a higher resemblance to the HDR target. Also, compared with direct fusion, the detail-recovered exposure fusion method achieves significantly higher similarity scores for both over- and under-exposed images.



(a) SSIM(over-exposed)

(b) SSIM(under-exposed)

Figure 6.9: Distribution of SSIM Scores

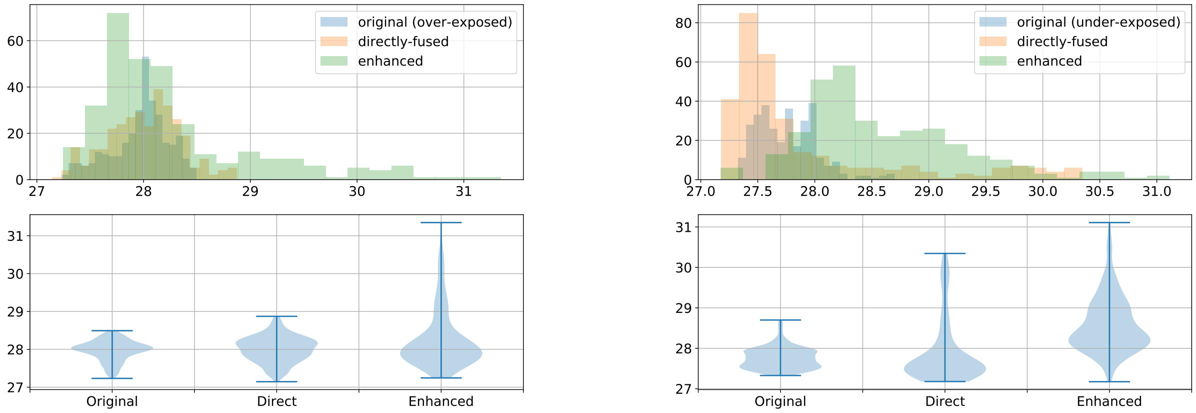
6.8 Performance Evaluation Using PSNR

Similar to performance evaluation using SSIM, we use a different metric named Peak Signal-Noise Ratio (PSNR)[17] to measure the difference between the HDR target and synthesized images. Unlike SSIM, PSNR is calculated based on the mean square error (MSE) between two compared images. Most of the steps are similar to what we used to compare SSIM scores. The steps of evaluation using PSNR are shown as follows.

1. For an ill-exposed (either over- or under-exposed) image img, we calculate the PSNR score between the HDR target img_T and img.

2. We generate the bracket images using the trained bi-directional network, which are img^+ (up-exposed) and img^- (down-exposed).

3. The generated bracket images together with the original image img are fused using detail-recovered fusion. The fused result (HDR-like image) is img_{DR}.



(a) PSNR(over-exposed)

(b) PSNR(under-exposed)

Figure 6.10: Distribution of PSNR Scores

4. The PSNR score between the HDR target img_T and the fused result img_{DR} is calculated.

5. We compare the two PSNR scores and record the results.

The comparison of PSNR scores is illustrated in Figure 6.10. The PSNR scores of original images img (labeled as *original*), images synthesized directly using illumination-based fusion (labeled as *directly-fused*) and the fused results img_{DR} (HDR-like image) are compared. We can notice a similar trend as we have seen in the distribution of SSIM scores (see Figure 6.9). When the original images (either over- or under-exposed) are converted into HDR-like images, there is a significant increase in PSNR scores. Besides, compared with illumination-based fusion, images using detail-recovered fusion generally have higher PSNR scores.

6.9 Performance Evaluation Using Dice Similarity Coefficient (U-Net)

We also used the change in semantic segmentation accuracy as an evaluation metric since semantic segmentation is an important task in medical image analyses including wounds. *Dice Similarity Coefficient*(DSC)[11] is used in the experiment as the metric for measuring segmentation accuracy compared to ground truth.

In our experiments, we trained a U-Net model to differentiate the chronic wound regions(positive) and the rest(negative) on *Pixel-M1* and *Pixel-M2* datasets. 80% of the data is used in training and the remaining 20% is used for evaluation. Figure 6.11 and 6.12illustrate the training and testing loss of the U-Net model.

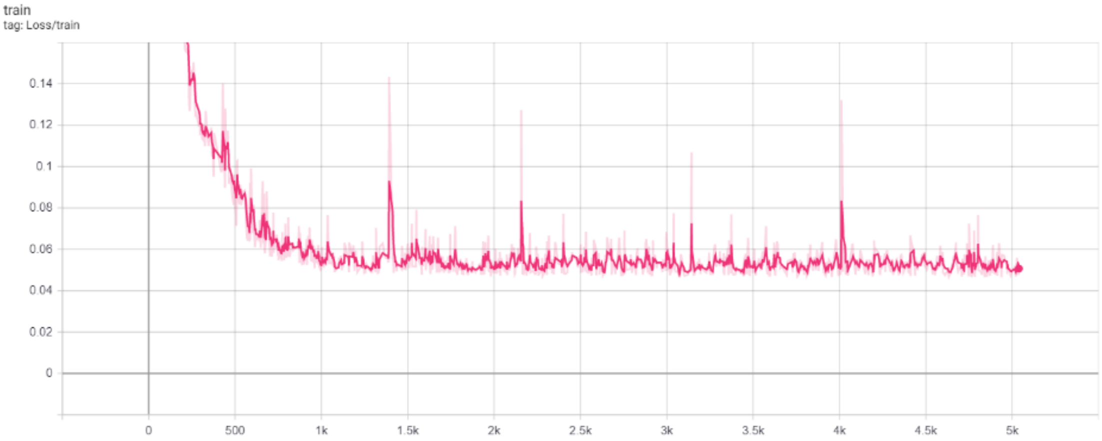


Figure 6.11: Training Loss of U-Net

The comparison of DSC scores is listed in Table 6.4. **DSC-** and **DSC+** are scores calculated on under and over-exposed images respectively. We can find our method achieves the highest DSC scores for both under and over-exposed among all methods compared against.

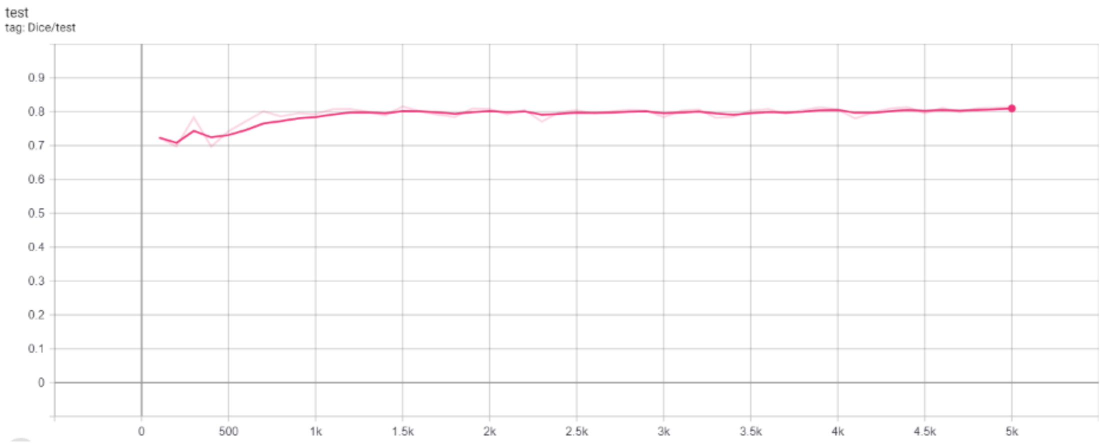


Figure 6.12: Testing Loss of U-Net

6.10 Comparison with State-Of-Art Methods

6.10.1 Visual Comparison

We compared results generated with our proposed method with those generated from HDRCNN[13], ExpandNet[33] and SingleHDR[31]. Sample results are shown in Figure6.13. The 1st column in Figure6.13 shows the original images that are used as inputs (the first 3 images are over-exposed and the last 3 images are under-exposed). Each row shows the HDR-like images generated using different methods. HDRCNN mitigates the over-exposed issues at the cost of reducing the color contrast. SingleHDR is prone to saturated issues for both over and under-exposed images. The performance of ExpandNet is roughly equivalent to our method in over-exposed images. However, ExpandNet may introduce significant artifacts in saturated regions. Overall, the enhanced results generated by our method have the best visual quality.

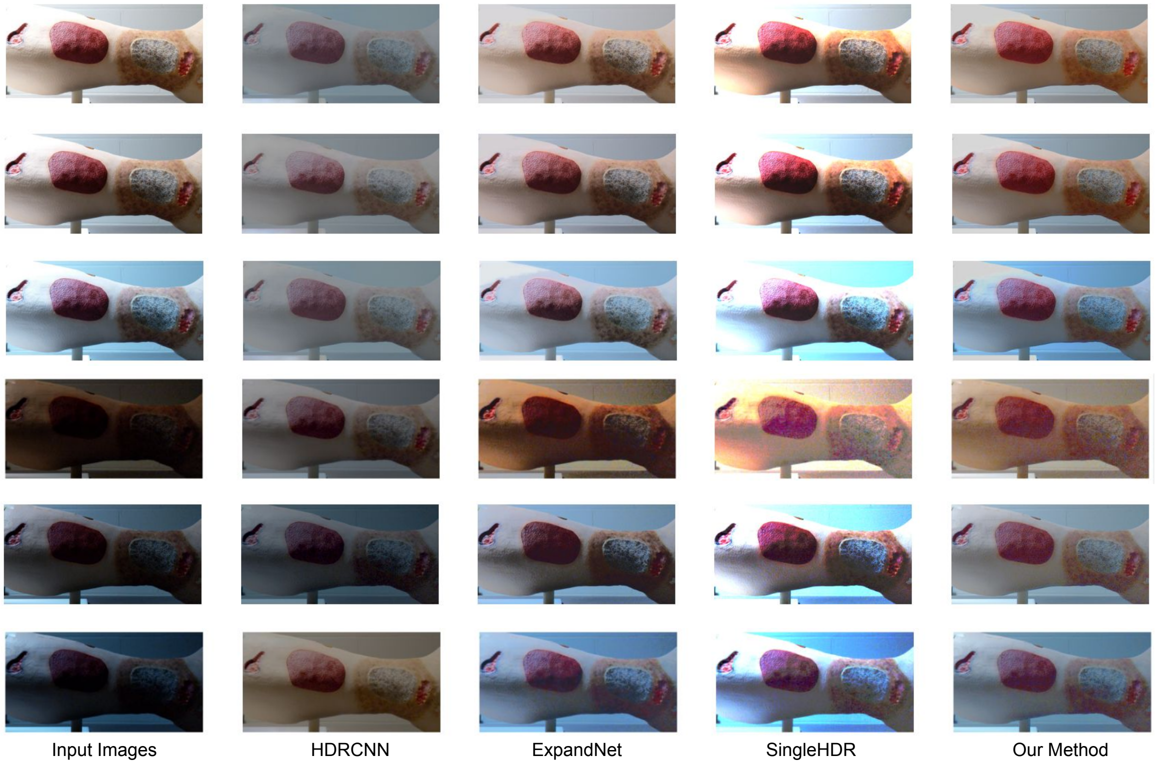


Figure 6.13: Visual Comparison on Enhanced Images. The first column contains the input images (moulage) from the IVDS dataset. The first 3 rows include HDR results generated from 3 over-exposed images and the last 3 rows include HDR results with under-exposed images. Except for our method, all other enhanced images(column 2-4) are tone-mapped with *Drago* using the same settings (gamma=1.6, saturation=1.45).

6.10.2 Quantitative Comparison

We also used SSIM. PSNR and DSC as our metrics to quantify the qualities of images after enhancement (see Table 6.4). All statistics are computed on the same test set, in which the over- and under-exposed images are tested and recorded separately. We use **SSIM+/SSIM-** to represent SSIM scores calculated using over- and under-exposed test images respectively, which is similar to how we calculate **PSNR+/PSNR-** and **DSC+/DSC-**. It's easy to find our method achieves the highest DSC scores for both under and over-exposed among all methods compared against.

Table 6.4: Comparison with Other Methods

Methods	**SSIM+**	**SSIM-**	**PSNR+**	**PSNR-**	**DSC+**	**DSC-**
Original	0.74 ± 0.04	0.56 ± 0.14	27.97 ± 0.27	27.76 ± 0.26	0.75 ± 0.07	0.58 ± 0.14
HDRCNN	0.73 ± 0.02	0.66 ± 0.06	$\mathbf{28.46 \pm 0.29}$	27.53 ± 0.28	0.41 ± 0.18	0.55 ± 0.11
ExpandNet	0.74 ± 0.04	0.68 ± 0.11	27.97 ± 0.35	27.97 ± 0.84	0.75 ± 0.11	0.73 ± 0.11
SingleHDR	0.66 ± 0.06	0.60 ± 0.11	27.90 ± 0.21	28.02 ± 0.22	0.60 ± 0.12	0.55 ± 0.14
Our Method	$\mathbf{0.76 \pm 0.04}$	$\mathbf{0.69 \pm 0.08}$	28.25 ± 0.76	$\mathbf{28.60 \pm 0.70}$	$\mathbf{0.76 \pm 0.09}$	$\mathbf{0.74 \pm 0.09}$

Compared with the original images (not enhanced using any method), we can also notice that not all image enhancement methods can achieve higher scores in quantitative assessments. For example, all enhancement methods produced images with higher SSIM scores for under-exposed images (**SSIM-**), but for other metrics on images with different exposures, the effects of enhancement are not significant. Our proposed method, however, got higher scores after enhancement on both over- and under-exposed images on all metrics compared with the original image, which indicates our method is a viable and relative stable method for image enhancement.

6.10.3 Enhancement Effect with Various Illuminations

We further analyzed the effect of enhancement on images with different lighting conditions using SSIM, PSNR and DSC. We use image L-values to estimate their average illumination. The L-value is the average value of the luminance channel $L(S)$ of an image S in CIELAB or LAB space can be expressed as $Lvalue = avg(L(S))$.

Figure 6.14, 6.15 and 6.16 show how SSIM, PSNR and DSC changes with different L-values respectively. For clearer visualization, all data points plotted in these three figures are smoothed using moving average with window size 10. As the L-value increases, SSIM scores of all methods increase in general. Beyond an L-value of 70, the SSIM scores of the original images, images generated by ExpandNet and our method starts to decrease. SSIM scores of HDRCNN are the most steady among all methods and SingleHDR has the best SSIM scores for very bright images (L-value larger than 80). Our method achieves the highest PSNR scores for dark (L-value around 35) and bright (L-value around 65) images. And ExpandNet achieves better scores in the middle range of L-values. Our method generally achieves better DSC scores on dark images (L-value below 55), and ExpandNet has roughly equivalent performance with our method on bright images.

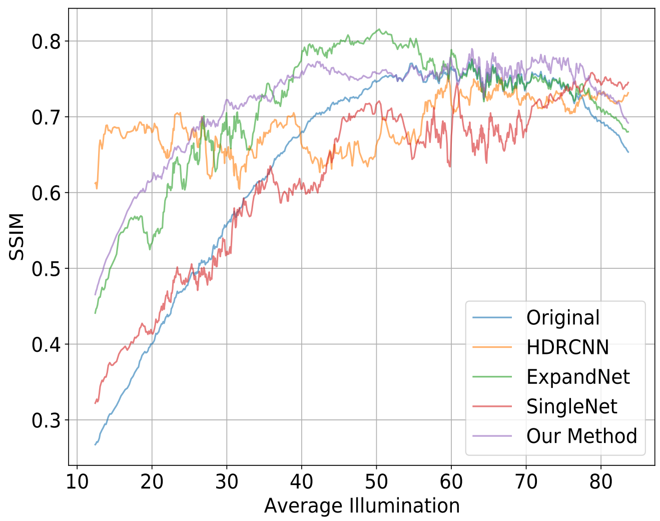


Figure 6.14: SSIM with Different L-values. Results are processed with moving average(window size equals to 10) for better visualization.

6.11 Discussions on Experiments

With all the experiments and evaluations we have conducted, we can see our method provides a viable solution to enhance both over- and under-exposed wound images. And it remains most stable compared with all the methods we compare against in terms of SSIM, PSNR and DSC. An interesting fact we also notice is: segmentation accuracy (DSC) of U-Net can be strongly affected by under-exposure and less affected by images with over-exposed issues(After enhancement using our pro-

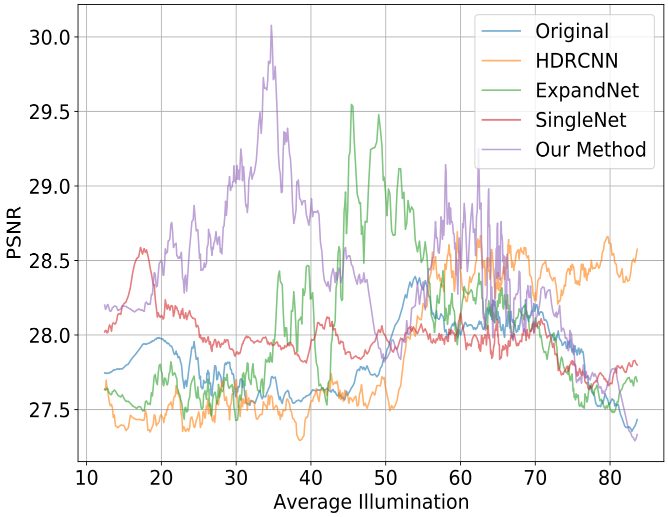


Figure 6.15: PSNR with Different L-values. Results are processed with moving average(window size equals to 10) for better visualization.

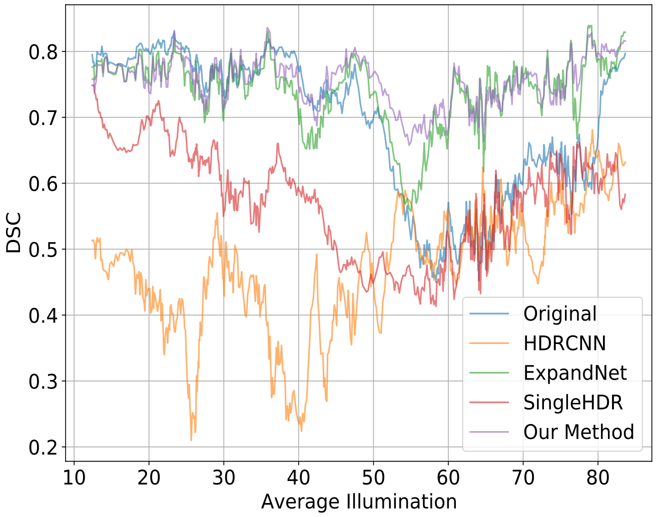


Figure 6.16: DSC with Different L-values. Results are processed with moving average(window size equals to 10) for better visualization.

posed method, the average DSC increased 16% on under-exposed images, while the increase on over-exposed images is only 1%). A similar conclusion can also be derived from our previous work[20]. The concrete conclusions of our experiments are summarized as follows.

- Compared with direct fusion(applying illumination-based fusion directly[46]), our proposed method(detail-recovered fusion) can effectively preserve more texture information, which can be testified by visual comparison and higher resemblance to the HDR-target in terms of SSIM and PSNR.
- Compared with the original wound image, our enhancement method can improve segmentation performance of U-Net in terms of DSC, while the enhancement effects vary for under- and over-exposed images.
- Our method achieves the highest scores in most metrics compared with state-of-art methods. And it has relatively consistent enhancement effects on image with various illuminations (characterized with L-Values).

Our work demonstrated the feasibility of improve performance of machine vision tasks (segmentation accuracy using U-Net) by mitigating adverse illuminations on wound images. However, there are still many limitations we did not cover in our work. 1) Due to the limitation of available datasets we can choose, the whole experiments and evaluations were conducted on the moulage image (instead of real wound images). 2) In the training process, we noticed the inconsistency of light temperatures between training and reference images can lead to large artifacts in bracket image generation, which means training the bi-directional network would require images with the same or close color temperature. 3) We notice the enhancement of over-exposed images is generally harder than under-exposed images due to texture information likely completely lost in large regions, and our method does not provide significant improvement in terms of DSC.

Apart from these limitations, we believe the proposed method achieves our goal– it successfully mitigates texture information loss of ill-exposed wound images by converting the original LDR image into the HDR-like image, which leads to improvement in performance in the machine vision tasks(wound segmentation using U-Net).

Chapter 7

Conclusion and Future Work

Wound assessment using images taken with smartphone cameras largely reduces the workload of caregivers in routine checking-ups in traditional treatment and provides a responsive and objective way to monitor the healing process of wounds, which can reduce the risks of amputations. However, wound images taken using smartphones are prone to exposure issues, which may lead to severe loss in texture information. The texture information loss can result in degradation in machine vision tasks which are common in image-based wound assessments and finally lead to inaccuracy in wound assessment. In order to address this issue, we proposed a novel deep network structure (Bi-Directional Illumination Enhancement Network) to mitigate texture information loss of ill-exposed wound images by converting the original LDR image into the HDR-like image.

In this thesis, a deep network structure was proposed to generate bracket images. To mitigate information loss in over and under-exposed image regions when fusing, we applied two transformer functions to recover the textural details. Illumination-based fusion was used in the 2-step fusion to generate the final output. Experiments show our method achieves higher scores in most metrics (SSIM, PSNR, DSC) for both dark and bright images on chronic wound images in the IVDS dataset.

The method proposed in our thesis provides a viable solution to mitigate the issues related to exposures taken using smartphone cameras, however, there are still lots of aspects in image-based wound assessments we want to explore. In our future work, we aim to improve accuracy in Photographic Wound Assessment Tool (PWAT) prediction by integrating the image enhancement pipeline into the modeling of a deep convolutional network. Since we are only working on images with synthetic wounds captured in the IVDS, we plan to conduct more experiments on real wounds in the next step in order to explore the feasibility of applying image enhancement techniques to facilitate a broader range of wound assessment tasks.

Bibliography

[1] Using your cameras histogram to take better photos. https://blog.banggood.com/using-your-cameras-histogram-to-take-better-photos-29932.html. Accessed: 2021-4-27.

[2] What is "semantic segmentation" compared to "segmentation" and "scene labeling"? https://stackoverflow.com/questions/33947823/what-is-semantic-segmentation-compared-to-segmentation-and-scene-labeling. Accessed: 2021-4-27.

[3] E. Agu, P. Pedersen, D. Strong, B. Tulu, Q. He, L. Wang, and Y. Li. The smartphone as a medical device: Assessing enablers, benefits and challenges. In *2013 IEEE International Workshop of Internet-of-Things Networking and Control (IoT-NC)*, pages 48–52. IEEE, 2013.

[4] T. Aledavood, J. Torous, A. M. T. Hoyos, J. A. Naslund, J.-P. Onnela, and M. Keshavan. Smartphone-based tracking of sleep in depression, anxiety, and psychotic disorders. *Current psychiatry reports*, 21(7):1–9, 2019.

[5] G. Anbarjafari, S. Izadpanahi, and H. Demirel. Video resolution enhancement by using discrete and stationary wavelet transforms with illumination compensation. *Signal, Image and Video Processing*, 9(1):87–92, 2015.

[6] F. Banterle, P. Ledda, K. Debattista, M. Bloj, A. Artusi, and A. Chalmers. A psychophysical evaluation of inverse tone mapping techniques. In *Computer Graphics Forum*, volume 28, pages 13–25. Wiley Online Library, 2009.

[7] F. Banterle, P. Ledda, K. Debattista, and A. Chalmers. Inverse tone mapping. In *Proceedings of the 4th international conference on Computer graphics and interactive techniques in Australasia and Southeast Asia*, pages 349–356, 2006.

[8] A. Bhelonde, N. Didolkar, S. Jangale, and N. L. Kulkarni. Flexible wound assessment system for diabetic patient using android smartphone. In *2015 International Conference on Green Computing and Internet of Things (ICGCIoT)*, pages 466–469. IEEE, 2015.

[9] D. H. Choi, I. H. Jang, M. H. Kim, and N. C. Kim. Color image enhancement using single-scale retinex based on an improved image formation model. In *2008 16th European Signal Processing Conference*, pages 1–5. IEEE, 2008.

[10] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH 2008 classes*, pages 1–10. 2008.

[11] L. R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

[12] F. Drago, K. Myszkowski, T. Annen, and N. Chiba. Adaptive logarithmic mapping for displaying high contrast scenes. In *Computer graphics forum*, volume 22, pages 419–426. Wiley Online Library, 2003.

[13] G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk, and J. Unger. Hdr image reconstruction from a single exposure using deep cnns. *ACM transactions on graphics (TOG)*, 36(6):1–15, 2017.

[14] Y. Endo, Y. Kanamori, and J. Mitani. Deep reverse tone mapping. *ACM Trans. Graph.*, 36(6):177–1, 2017.

[15] M. D. Grossberg and S. K. Nayar. What is the space of camera response functions? In *Proc. IEEE CVPR*, volume 2, pages II–602. IEEE, 2003.

[16] X. Guo, Y. Li, and H. Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2):982–993, 2016.

[17] A. Hore and D. Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th international conference on pattern recognition*, pages 2366–2369. IEEE, 2010.

[18] S.-C. Hsia and T.-T. Kuo. High-performance hdr image generation by inverted local patterns. *IET Image Proc.*, 9(12):1083–1091, 2015.

[19] Y. C. Hum, K. W. Lai, and M. I. Mohamad Salim. Multiobjectives bihistogram equalization for image contrast enhancement. *Complexity*, 20(2):22–36, 2014.

[20] A. B. Iyer. Let there be light... characterizing the effects of adverse lighting on semantic segmentation of wound images and mitigation using a deep retinex model. *Masters thesis, Worcester Polytechnic Institute*, 2020.

[21] A. P. James and B. V. Dasarathy. Medical image fusion: A survey of the state of the art. *Information fusion*, 19:4–19, 2014.

[22] D. J. Jobson, Z.-u. Rahman, and G. A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image processing*, 6(7):965–976, 1997.

[23] N. K. Kalantari and R. Ramamoorthi. Deep high dynamic range imaging of dynamic scenes. *ACM Trans. Graph.*, 36(4):144–1, 2017.

[24] A. Khunteta, D. Ghosh, et al. Fuzzy rule-based image exposure level estimation and adaptive gamma correction for contrast enhancement in dark images. In *IEEE Int'l Conf. Signal Processing*, volume 1, pages 667–672. IEEE, 2012.

[25] R. P. Kovaleski and M. M. Oliveira. High-quality reverse tone mapping for a wide range of exposures. In *Proc. SIBGRAPI*, pages 49–56. IEEE, 2014.

[26] K. S. Kumar and B. E. Reddy. Wound image analysis classifier for efficient tracking of wound healing status. *Signal & Image Processing*, 5(2):15, 2014.

[27] E. H. Land. Retinex theory. *Journal of the Optical Society of America*, 61(1):1–11, 1971.

[28] E. H. Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977.

[29] H. Li, S. M. Smith, S. Gruber, S. E. Lukas, M. M. Silveri, K. P. Hill, W. D. Killgore, and L. D. Nickerson. Denoising scanner effects from multimodal mri data using linked independent component analysis. *Neuroimage*, 208:116388, 2020.

[30] H. Lin and Z. Shi. Multi-scale retinex improvement for nighttime image enhancement. *Optik*, 125(24):7143–7148, 2014.

[31] Y.-L. Liu, W.-S. Lai, Y.-S. Chen, Y.-L. Kao, M.-H. Yang, Y.-Y. Chuang, and J.-B. Huang. Single-image hdr reconstruction by learning to reverse the camera pipeline. In *Proceedings IEEE CVPR*, pages 1651–1660, 2020.

[32] R. Mantiuk, S. Daly, and L. Kerofsky. Display adaptive tone mapping. In *ACM SIGGRAPH 2008 papers*, pages 1–10. 2008.

[33] D. Marnerides, T. Bashford-Rogers, J. Hatchett, and K. Debattista. Expandnet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content. In *Computer Graphics Forum*, volume 37, pages 37–49. Wiley Online Library, 2018.

[34] L. Martinengo, M. Olsson, R. Bajpai, M. Soljak, Z. Upton, A. Schmidtchen, J. Car, and K. Järbrink. Prevalence of chronic wounds in the general population: systematic review and meta-analysis of observational studies. *Annals of epidemiology*, 29:8–15, 2019.

[35] T. Mertens, J. Kautz, and F. Van Reeth. Exposure fusion: A simple and practical alternative to high dynamic range photography. In *Computer graphics forum*, volume 28, pages 161–171. Wiley Online Library, 2009.

[36] K. Myszkowski, R. Mantiuk, and G. Krawczyk. High dynamic range video. *Syn. Lectures Computer Graphics & Animation*, 1(1):1–158, 2008.

[37] A. S. Parihar, K. Singh, H. Rohilla, and G. Asnani. Fusion-based simultaneous estimation of reflectance and illumination for low-light image enhancement. *IET Image Processing*.

[38] Z.-u. Rahman, D. J. Jobson, and G. A. Woodell. Multi-scale retinex for color image enhancement. In *Proceedings of 3rd IEEE International Conference on Image Processing*, volume 3, pages 1003–1006. IEEE, 1996.

[39] P. Rasti, M. Daneshmand, F. Alisinanoglu, C. Ozcinar, and G. Anbarjafari. Medical image illumination enhancement and sharpening by using stationary wavelet transform. In *2016 24th Signal Processing and Communication Application Conference (SIU)*, pages 153–156. IEEE, 2016.

[40] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010.

[41] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda. Photographic tone reproduction for digital images. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 267–276, 2002.

[42] N. A. Richmond, A. D. Maderal, and A. C. Vivas. Evidence-based management of common chronic lower extremity ulcers. *Dermatologic therapy*, 26(3):187–196, 2013.

[43] M. Song, D. Tao, C. Chen, J. Bu, J. Luo, and C. Zhang. Probabilistic exposure fusion. *IEEE Transactions on Image Processing*, 21(1):341–357, 2011.

[44] H. L. Tan, Z. Li, Y. H. Tan, S. Rahardja, and C. Yeo. A perceptually relevant mse-based image quality metric. *IEEE Transactions on Image Processing*, 22(11):4447–4459, 2013.

[45] F. J. Veredas, R. M. Luque-Baena, F. J. Martín-Santos, J. C. Morilla-Herrera, and L. Morente. Wound image evaluation with machine learning. *Neurocomputing*, 164:112–122, 2015.

[46] V. Vonikakis, O. Bouzos, and I. Andreadis. Multi-exposure image fusion based on illumination estimation. In *Proc. IASTED SIPA*, pages 135–142, 2011.

[47] L. Wang. System designs for diabetic foot ulcer image assessment. *System*, 2016:03–07, 2016.

[48] L. Wang, P. C. Pedersen, E. Agu, D. M. Strong, and B. Tulu. Area determination of diabetic foot ulcer images using a cascaded two-stage svm-based classification. *IEEE Transactions on Biomedical Engineering*, 64(9):2098–2109, 2016.

[49] L. Wang, P. C. Pedersen, D. M. Strong, B. Tulu, E. Agu, and R. Ignotz. Smartphone-based wound assessment system for patients with diabetes. *IEEE Transactions on Biomedical Engineering*, 62(2):477–488, 2014.

[50] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[51] C. Wei, W. Wang, W. Yang, and J. Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018.

[52] J. Zhang and J.-F. Lalonde. Learning high dynamic range from outdoor panoramas. In *Proc. IEEE Int'l Conf. Computer Vision*, pages 4519–4528, 2017.

[53] Y. Zheng, E. Blasch, and Z. Liu. *Multispectral image fusion and colorization*, volume 481. SPIE Press, 2018.