OBJECT RECOGNITION IN LOW LIGHT AND HAZY ENVIRONMENT

A Thesis Submitted
In Partial Fulfillment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

by

Kavinder Singh
(2K18/PHDCO/13)

Under the Supervision of
Prof. Anil Singh Parihar
Dept. of Computer Science & Engineering

Department of Computer Science & Engineering
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahabad Daulatpur, Main Bawana Road, Delhi-110042, INDIA

July 2024

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahabad Daulatpur, Main Bawana Road, Delhi-110042, INDIA

CANDIDATE'S DECLARATION

I **Kavinder Singh, (2K18/PHDCO/13)** hereby declare that the work which is being presented in the thesis entitled **"Object recognition in low light and hazy environment"** in the partial fulfillment of the requirements for the award of the Degree of Doctor of Philosophy, submitted in the Department of Computer Science & Engineering, Delhi Technological University is an authentic record of my own work carried out during the period from 2018 to 2024 under the supervision of Prof. Anil Singh Parihar.

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other Institute.

Kavinder Singh
(2K18/PHDCO/13)

This is to certify that the student has incorporated all the corrections suggested by the examiners in the thesis and the statement made by the candidate is correct to the best of my knowledge.

Signature of Supervisor **Signature of External Examiner**

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahabad Daulatpur, Main Bawana Road, Delhi-110042, INDIA

CERTIFICATE

Certified that **Kavinder Singh (2K18/PHDCO/13)** has carried out his research work presented in this thesis entitled **"Object recognition in low light and hazy environment"** for the award of **Doctor of Philosophy** from Department of Computer Science & Engineering, Delhi Technological University, Delhi, under my supervision. The thesis embodies results of original work, and studies are carried out by the student himself and the contents of the thesis do not form the basis of the award of any other degree to the candidate or to anybody else from this or any other institution.

Prof. Anil Singh Parihar
(Supervisor)
Department of Computer Science & Engineering
Delhi Technological University, Delhi

Date: 08/07/2024
Place: New Delhi

ACKNOWLEDGMENTS

I would like to express deep feelings of gratitude to all the people who have been pillars of support and encouragement. Completing a PhD is a significant milestone in my academic journey, and it would not have been possible without the support, guidance, and encouragement of many individuals.

First and foremost, I extend my deepest appreciation to my supervisor **Prof. Anil Singh Parihar**, Dept. of Computer Science & Engineering, DTU for his constant guidance, valuable suggestions and kind encouragement since my association with him. His patience, constant encouragement, immensely valuable ideas, suggestions, and unwavering support have been invaluable, not only in shaping this thesis but also in fostering my personal and academic growth. His substantial and thorough approach, together with his genuine interest in the research subject, turned my research work into a great experience. His suggestions will remain with me as an inexhaustible source of scientific learning throughout my life.

I am deeply indebted to all the faculty members of our department and institute for their support. Their dedication to sharing knowledge and providing constructive feedback has greatly enriched my experience. My sincere appreciation extends to Delhi Technological University, New Delhi for providing an environment conducive to research and learning.

A special note of gratitude is reserved for my labmates, Ashutosh Pandey, Aman Kumar Pandey, and Nihal Kumar for always being there and bearing with me the good and bad times during my wonderful days of Ph.D. Our journey together in the lab, filled with endless discussions, brainstorming sessions, and shared challenges, has been one of the most rewarding experiences of my academic life.

To my family, words fall short to express my gratitude. Their endless love, sacrifices, and belief in my dreams have been the foundation upon which i have built my aspirations. Their constant encouragement and unconditional support have been the bedrock of my resilience and perseverance. I owe them more than this thesis can convey; they are the unsung heroes of this academic achievement.

Date: 08/07/2024 **(Kavinder Singh)**
Place: Delhi

ABSTRACT

The thesis aims to develop new methods for object recognition in low-light and hazy environment. The performance of existing object recognition methods degrades while dealing with low-light/hazy environments due to poor visibility of details and other issues in the captured images. There are two approaches to deal with low-light/hazy environment: one approach to deal with low-light/hazy environment is to use enhancement/dehazing before object recognition, and the other approach focuses on direct object detection from the low-light/hazy images. The work attempted to improve object recognition using both approaches. Based on the first approach for low-light, this thesis presents two low-light image enhancement methods. The first method estimates structure-aware initial illumination from the input images based on the proposed multi-scale guided image filter. A multi-objective optimization function is formulated and solved to refine the illumination. The adaptive illumination adjustment is developed to improve the overall lightness of the low-light images leveraging the estimated illumination. The second method develops a deep-network for simultaneous estimation of illumination and reflectance from a single image. A branched encoder-decoder architecture is developed for the decomposition task. The estimated illumination is adjusted using the deep network to improve the overall lightness. The image refinement module is developed to improve the color, contrast and other details in the enhanced images. Similarly for hazy environment, We developed a variational optimization based method for image dehazing. It is difficult to estimate depth of a scene from RGB image. We introduced a notion that the objects with same structure at a depth contains similar transmission. Thus, we developed a new method for estimation of structure-aware initial transmission leveraging the scale-adaptive bilateral filter. We formulated a new variational optimization problem with regularization terms to preserve the structural details in the final transmission. The atmospheric light is not dependent on the color; thus, we developed a uniform atmospheric light estimation method.

The performance of the developed methods is compared with the various contemporary methods on a large set of images using visual and quantitative assessments. The analysis shows the superiority of the proposed methods over the existing methods in the enhancement/dehazing task. Furthermore, we analyzed the performance of object recognition using various enhancement methods. The enhancement method requires additional time for processing and the performance of object recognition depends greatly on the performance of enhancement methods. Thus, we aimed to develop

a direct object recognition methods for low-light and hazy environment that do not require explicit enhancement to be performed before object recognition.

This thesis present a new Multi-exposure refinement network for low-light object detection (MRN-LOD) to avoid the need for enhancement before recognition. The MRN-LOD contains: multi-exposure feature extractor, adaptive refinement network, and detection head. We introduced the notion of feature extraction from multi-exposure images for object recognition in low light. In addition, we proposed an adaptive refinement network to refine the features of low-light images for better recognition performance. The detection head uses the refined features to perform object recognition. Extensive experimentation on existing real-world and synthetic datasets shows the superiority of the proposed MRN-LOD. Furthermore, The performance of the object detection methods degrades in a hazy environment. To overcome this issue, we propose a Bi-stream feature fusion (BFF) network for object detection in a hazy environment. The BFF network consists of three modules: hybrid input, Bi-stream feature extractor (BFE), and multi-level feature fusion. We present the notion of hybrid input to extract features from the hazy images in an effective manner. The proposed BFE network extracts multi-level features from the hazy image and hybrid input. The multi-level feature fusion (MFF) network performs the convolution-based adaptive feature fusion and processes the extracted features. The proposed BFF model outperforms other state-of-the-art methods in hazy environments. Another challenge in hazy object detection is the unavailability of a dataset with sufficient samples and classes. In this work, we developed a synthetic object detection dataset for a hazy environment (DHOD).

PUBLICATIONS

Published/ Accepted

SCI-INDEXED JOURNALS

- **Kavinder Singh** and Anil Singh Parihar, "Variational optimization based single image dehazing," ***Journal of Visual Communication and Image Representation***, Volume 79, pp. 103241, August 2021. DOI: 10.1016/j.jvcir.2021.103241. **(Impact factor: 2.6)**

- **Kavinder Singh** and Anil Singh Parihar, "Illumination Estimation for Nature Preserving low-light image enhancement," ***The Visual Computer***, Volume 40, pp. 121-136, 2024. DOI: 10.1007/s00371-023-02770-9. (*Published*). **(Impact factor: 3.0)**

- **Kavinder Singh** and Anil Singh Parihar, "DSE-Net: Deep simultaneous estimation network for low-light image enhancement," ***Journal of Visual Communication and Image Representation***, Volume 91, pp. 103780, March 2023. DOI: 10.1016/j.jvcir.2023.103780. (*Published*). **(Impact factor: 2.6)**

- **Kavinder Singh** and Anil Singh Parihar,"BFF: Bi-Stream Feature Fusion for Object Detection in Hazy Environment," ***Signal, Image and Video Processing***, Volume 18, pp. 3097-3107, 2024. (*Published*).***(Impact factor: 2.3)***

- **Kavinder Singh** and Anil Singh Parihar, "MRN-LOD: Multi-exposure Refinement Network for Low-light Object Detection", ***Journal of Visual Communication and Image Representation***, Volume 99, pp. 104079, March 2024. DOI: 10.1016/j.jvcir.2024.104079. (*Published*) **(Impact factor: 2.6)**

SCOPUS-INDEXED CONFERENCES

- **Kavinder Singh** and Anil Singh Parihar, "A comparative analysis of illumination estimation based Image Enhancement techniques," ***2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)***, Vellore, India, 2020, pp. 1-5, doi: 10.1109/ic-ETITE47903.2020.195.

- **Kavinder Singh** and Anil Singh Parihar, "HDVF: Hazy object detection network based on variational features", ***International Conference on Artificial Intelligence, Advanced Materials, and Mechatronics Systems (AIAMMS-2023)***. (Accepted and Presented)

- **Kavinder Singh** and Anil Singh Parihar, "Image Decomposition for Object Detection in Low-light Environment", ***International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT-2023)***. (Accepted and Presented)

List of Figures

List of Tables

Table of Contents

Chapter 1

Introduction

The pervasive integration of computer vision across various applications and domains has intensified the requirement for robust and reliable algorithms. Object recognition is one of the essential and primary task in the field of computer vision. The performance of traditional object recognition methods is limited due to various reasons like, hand-crafted features, inadequate handling of occlusions, and limited contextual understanding. The integration of deep learning has improved the performance of object recognition systems. The ability of existing object recognition systems to identify and localize objects is inherently dependent on the quality of input images. The existing object detection systems effectively detects objects in the images captured in normal environment with clearly visible details. However, the efficacy of these systems is often compromised in challenging environment conditions, such as low-light, and hazy scenarios, which significantly hinders the image quality and hence, the performance of existing object recognition algorithms. The ability to identify and localize objects in images captured under these challenging circumstances is of paramount importance for various real-world applications, ranging from surveillance and autonomous vehicles to search and rescue operations.

The chapter includes the discussion related to object recognition, low-light image enhancement, and image dehazing. Further, This chapter discusses the challenges of object recognition in low-light and hazy environment, the significance of this research, and overview of the thesis.

1.1 Object Recognition

Object recognition is a fundamental task in computer vision that involves localization and classification of the objects within an image. Unlike image classification, where the goal is to assign a single label to an entire image, object recognition aims to identify the presence of multiple objects within the image and provide bounding boxes around them. Object recognition have widespread applications in fields such as image understanding, video analysis, and autonomous systems.

Traditional object recognition systems rely on feature extraction and pattern recognition. They often use techniques like edge detection, shape analysis, and texture analysis to distinguish objects. With the advancement of deep learning, convolution neural networks (CNNs) have become the cornerstone of modern object recognition. These networks automatically learns feature representation from the input images, making them highly effective in recognizing and classifying a wide variety of objects.

In a typical CNN-based object recognition system, an image is processed through multiple layers of convolution filters, pooling layers, and fully-connected layers. Each layer extracts and refine features, enabling the network to make informed decisions about the objects present in the image. These systems are often trained on a large datasets with annotated examples, allowing them to learn a wide range of object appearances and variations. Numerous object recognition methods have been proposed in the literature with remarkable performance as compared to the traditional object recognition methods. State-of-the-art object recognition methods such as Faster R-CNN (Region-based Convolutional Neural Network) and YOLO (You Only Look Once), have demonstrated impressive accuracy and efficiency. These methods typically leverage deep learning architectures, utilizing convolution neural networks (CNNs) to extract features from the images. Region proposal networks and anchor-based mechanisms are commonly employed for precise localization, while classification networks discern the object classes.

1.1.1 Object Recognition in challenging environment

The existing object recognition methods performs adequately for images with appropriate quality. However, these methods face considerable challenges when confronted with challenging environments characterized by low-light and haze. In this thesis, we focused on two challenging environments: low-light and hazy. One of the majorly used solution to deal with low-light and haze in the images is to use enhancement as a pre-processing step. We aim to develop enhancement algorithms to deal with the

problems related to the low-light environment and hazy environment. To develop a methods for better image quality, we explored various low-light image enhancement methods and image dehazing methods which are discusses in the following sections. In the solution based on enhancement before recognition, the performance of object recognition methods relies greatly on the performance of enhancement algorithms and requires additional processing time. Thus, we also developed direct object detection methods from hazy and low-light images to achieve better performance and detection speed.

1.1.2 Enhancement-based object recognition in low-light

Many times the environment is not conducive to capture an image with good illumination and results in a low-light image, which severely impact the performance of object recognition algorithms. The other cause of low-light images may be variable lightness in a scene, night-time imaging, improper camera setting, bad lighting sources, etc. The recapturing of the image is not feasible in many situations. Thus, low-light image enhancement plays a vital role in improving the details of an image suffering from poor lightness. Low-light image enhancement involves augmenting the brightness and contrast, reducing noise, and preserving the natural color and texture of the image.

In literature, many researchers have attempted to develop algorithms for low-light image enhancement. Low-light images contain most of the pixels with low intensity and can be loosely considered as low contrast images. To this end, one of the simplest and fundamental approaches is Histogram Equalization (HE) [9]. HE based approaches focus on performing intensity transformation using frequencies of pixel intensities. However, HE based methods may introduce saturation effect, over or under enhancement, halo effects, loss of local details, etc. The above algorithms give reasonably good performance in properly illuminated regions. However, these algorithms focus on contrast enhancement rather than improving the illumination of the image, thus sometimes fail to improve the low-light images and introduce artifacts like saturation effect, halo effect, etc.

It has been found that Retinex [10] based enhancement algorithms work reasonably well for low-light images. In the Retinex theory, an image is considered as the product of illumination and reflectance. In early Retinex based approaches [11, 12, 13], lightness is considered as the ratio of a pixel value to the average value of surrounding pixels and term it as center/surround operation. Jobson *et al.* [14] explored the properties of center/surround operation and used a Gaussian filter to develop Single Scale Retinex (SSR) for image enhancement. SSR has a major drawback that it provides

either dynamic range compression or color constancy but fails to provide both simultaneously. To overcome these limitations, Jobson *et al.* [15] used multiple scales of Gaussian and developed Multi-Scale Retinex (MSR). MSR works well for grayscale images, but for color images, it may lead to unnatural color appearance. The problem of color restoration is attempted in another work by Jobson *et al.* [16]. In these algorithms [14, 15, 16] reflectance is considered as the final enhanced image, which may look unnatural at times.

Another set of algorithms [1, 17] estimate the illumination component and use it to find the reflectance component. The final enhanced image is obtained as the product of refined illumination and reflectance. Wang *et al.* [1] proposed an algorithm Naturalness Preserved Enhancement Algorithm (NPEA) to deal with non-uniform illumination images. NPEA estimates illumination using a bright pass filter to find the reflectance component of the image. NPEA preserves naturalness at the cost of illumination improvement. Thus, darker regions may not enhance adequately. Fu *et al.* [17] proposed a fusion-based method to deal with weakly illuminated images. The algorithm uses a multi-fusion of Gaussian and Laplacian at multiple scales. The algorithm performs well in normal dark images but fails to provide significant lightness for a darker region. Guo *et al.* [3] proposed a low-light image enhancement algorithm using illumination map estimation (LIME). Similar to many other approaches, [1, 17] LIME estimates initial illumination by taking a pixel-wise maximum of all color channels of the low-light image. The algorithm computes refined illumination by optimizing a multi-objective problem based on initial illumination and its gradient. The performance of LIME is significantly good, however over enhancement occurs in properly illuminated regions of low-light images. Moreover, LIME sometimes fails to retain color constancy. An inferior enhancement algorithm may result in the loss of crucial details and introduce undesirable artifacts. Thus, designing an effective algorithm for the enhancement of low-light images is a challenging task.

A low-light image enhancement algorithm attempts to improve the illumination of an image to enhance the lightness in the final image. However, various challenges are faced while designing such algorithms. A major problem is inadequate enhancement of the non-uniformly illuminated images i.e. over-enhancement in the bright regions or under-enhancement in the dark regions. In practical scenarios, the low-light images may contain few regions with adequate lightness and other regions with inadequate lightness. Such images are termed as non-uniformly illuminated images. Few algorithms focuses on enhancement of low-light regions, which sometimes lead to the over-enhancement of the region with sufficient lightness. The algorithms which focus on maintaining the overall lightness or avoid over-enhancement sometimes fail to enhance

the details of the low-light regions. It is difficult to manage both over-enhancement and under-enhancement, simultaneously. Few algorithms fails to preserve the natural color in the enhanced image. The enhancement using few methods do not consider the color balance and hence lead to enhancement with unnatural colors. Similarly, the generation of undesired artifacts in the enhanced image is also a major challenge in low-light image enhancement. Sometimes algorithms attempt to reduce noise in the enhanced images. In attempt to avoid the noise, the algorithms fail to preserve the fine details in the enhanced images. Furthermore, the algorithms sometimes fail to restrict the dispersion effect in light dominant regions of the low-light images with varying illumination.

The simplest approach for object recognition in low-light is to apply low-light image enhancement (pre-processing) before object recognition. However, such an approach requires additional time and computation to enhance the image. Furthermore, the performance of recognition approach depends greatly on the performance of the enhancement method. The design of direct object recognition approach is challenging due to limited visiblity of details in the low-light images. The performance of existing methods degrade while dealing with low-light images. In many cases, the enhacement-before-recognition increases the time drastically. However, these methods provides only limited enhancement in the recognition performance. In addition to this, the availability of low-light object recognition dataset with sufficient number of samples is also another important challenge. The performance of deep learning methods is largly governed by the quality and quantity of samples. There is no dataset openly available for low-light object detection with a sufficient number of object classes and annotated RGB images.

1.1.3 Dehazing-based object recognition in hazy environment

The outdoor image acquisition may result in hazy images due to poor weather conditions. Due to the presence of suspended particles in the atmosphere, the reflected and incident light rays get scattered. The scattered light rays leads to the effects like blurring and contrast degradation. The acquisition of haze-free image in presence of haze/smoke/fog is quite difficult if not infeasible. The primary focus of image dehazing is to improve the visual quality of an image and recover haze-free image.

Image dehazing is particularly challenging due to numerous factors like the unclear distinction between the hazy pixels, natural scene, and even the camera error [18]. Furthermore, the lack of real ground truth and corresponding foggy images makes it difficult to generalize the techniques to real world data. To solve the problem researchers analyzed distinct haze relevant properties leveraging synthetic datasets and proposed

various image dehazing algorithms. A superior image dehazing algorithm requires appropriate haze relevant properties which implies the need for an analysis of the existing image dehazing algorithms. Wang and Yuan [18] classified a number of haze removal approaches and in provided a subjective review of output images from a number of traditional dehazing methods [19, 20, 21, 22] for analysis.

The primary challenge in dehazing is to estimate the depth of the scene and the amount of light scattering, which varies with the distance of objects from the camera. It can be inferred from physical image formation model that one of the methods for dehazing is by estimating global atmospheric light and depth information. Researchers [19, 20, 21, 22] proposed some assumptions or priors in order to facilitate the estimations and subsequently obtain haze free image. All these methods have inherent limitations to the assumptions with respect to specific scenes. He *et al.* [20] proposed Dark Channel Prior, which states that there is one color channel with pixel values nearing zero. Despite being able to produce fine results it should be noted that the assumption stands to be violated in multiple situations leading to an incorrect representation of the transmission map, hence the hazy image. Other methods tend to exploit multiple images from the same scene for haze free image restoration [23, 24, 25].

The significance of image dehazing is beyond mere aesthetic improvement of images. The feature extraction from haze-free image is comparably easier than hazy image. Thus, effective dehazing may help in performance improvement of various computer vision methods including object detection. However, the existing approaches suffer from various limitations, which may have an adverse effect on the performance of object recognition methods in hazy environment. Thus, the development of an effective algorithm for image dehazing is a challenging task.

The deep learning based image dehazing methods use specific datasets to learn haze relevant features. These methods perform well on the images which have distribution similar to the images of dataset. However, these methods fail to provide adequate dehazing for other images. The generalization capability of these methods is limited, which makes these methods less useful for real-world applications. Another major challenge is over-enhancement and under-enhancement. Few algorithms appropriately enhances the hazy images, and fail to deal with images with less haze density. On the other side, few algorithm are capable of performing dehazing with images that contains limited haze. Image dehazing algorithms attempts to reduce the haze density and generate the haze-free sample. However, algorithms sometimes fails to maintain the colors in the enhanced images. This distortion of color in the enhaced images is not desirable. In attempt to perform effective dehazing, sometimes algorithms loss the finer details in the dehazed images.

The pre-processing based object recognition approaches require additional time and computation to perform image dehazing. The dehazing algorithms sometimes fail to improve the recognition performance of the systems. Furthermore, the existing methods uses specific dataset for learning the haze-relevant characteristics. These methods performs well on the sample from the similar data. However, the performance of these methods degrade while dealing with samples from other datasets. The availability of supervised dataset is another important challenge in object recognition in hazy environment. There is no dataset available with sufficient number of object classes and annotated RGB images.

1.2 Image Formation Models

The concept of image formation models is fundamental in computer vision. The image formation models focus on describing how the observed image is formed through the interaction of light with objects and the environment. The models typically encompass the physical and geometric aspects of the scene illumination, object reflection properties, camera characteristics, and atmospheric conditions. These models are crucial for developing algorithms that can effectively process, enhance, and analyze images under various conditions.

Low-light image enhancement methods based on Retinex model performs impressively as it considers image as a composition of reflectance and illumination. Furthermore, the estimation and treatment of illumination helps in enhancement of low-light images. On the other side, the physical image formation model considers atmospheric light and scene depth. Researchers leveraged physical image formation model in many image dehazing algorithms. In this section, we discuss two image formation models, which are as follows:

- **Retinex model:** The model focuses on estimation of reflectance and illumiantion, which can help in enhancement of low-light images.
- **Physical image formation model:** It focuses on estimation of transmission map and atmospheric light. The model uses the estimated components to perform image dehazing.

These are the widely used models for in various papers. The detailed discussion is given below.

1.2.1 Retinex Model

Retinex [12] image formation model, which can be defined as:

$$\mathbf{I} = \mathbf{R} \circ \mathbf{L} \tag{1.1}$$

where **I** is an image, **R** is reflectance of the image, **L** is illumination, and symbol '$\circ$' represents element-wise multiplication. The reflectance component depends on the objects in images and their colors. The illumination component depends on the source of light present at the time of image acquisition. The source of light might be sunlight, moonlight, an artificial light source like LED, CFL, camera flashlight, etc. Change in the source of light may lead to a change in the properties of illumination. Illumination properties of low-light images are different from day-light images. It is mainly because sunlight provides uniform illumination except in shaded regions. Generally, illumination in low-light images is non-uniform due to the dominance of light sources in some regions of the scene. In many algorithms [1, 17, 3], illumination estimation is obtained from the image, and reflectance is computed using relation in (1.1). Thus, illumination estimation is a crucial step in the low-light image enhancement algorithm. Improper estimation of illumination leads to a degraded image and may produce artifacts.

1.2.2 Physical Image Formation Model

Haze formation occurs due to suspended particles like aerosol, water-vapor, dust or smoke in the air causing the light to scatter. This fundamental scattering based physical phenomenon is formulated mathematically by [26] and [27]. This has been used as the basis for dehazing by [19] and [28] wherein the origins, sizes and distributions of haze are also taken into account. The formulation is given by:

$$I(x) = A_\infty \rho(x) \exp^{-\beta d(x)} + A_\infty (1 - \exp^{-\beta d(x)}) \tag{1.2}$$

Where x is a pixel on a 2D coordinate plane, I is the input image intensity, A_∞ is the atmospheric light, d is the distance between object and lens, ρ is the reflectance of the object in frame, and β is the atmospheric attenuation constant.

I is the result of medium of transmission and light reflected from the object(s) in the frame, essentially representing the hazy image or the direct attenuation. And the transmission $exp^{-\beta d(x)}$ is represented by $t(x)$. Hence, reflectance times atmospheric light; $L_\infty \rho(x)$, is the actual haze free image given by $J(x)$. Simplifying the equation

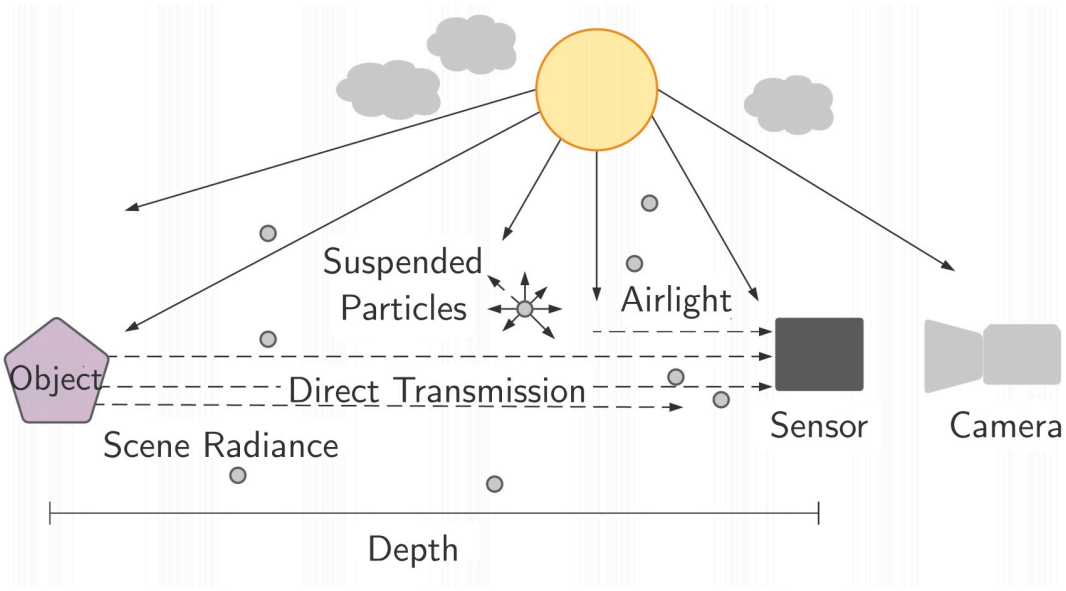


Fig. 1.1: The physical phenomenon of scattering and the process of imaging in hazy weather. The suspended particles lead to decrease in intensity of the reflected light causing the formation of hazy image on the camera sensor.

and replacing atmospheric light with A, the obtained expression is given as:

$$I(x) = J(x)t(x) + A(1 - t(x)) \tag{1.3}$$

From equations (4.1) and (4.2), we can infer that despite a physical model to represent image haze it is hard to estimate both unknowns, A and $t(x)$. However, if depth information or $d(x)$ is known, we can estimate J, A and t from I. This has primarily been done to solve the dehazing problem in recent years. It is ill-posed problem and effective estimation of the components is crucial.

1.3 Problem Statement

The performance of object recognition models degrades while dealing with images which are captured under the influence of low-light environment and hazy environment. The simplest approach is to use image enhancement/dehazing before performing object recognition. The enhancement algorithms for low-light image enhancement suffers from over enhancement and fails to preserve natural colors in images with varying illumination. Thus, nature preserving illumination estimation and deep-simultaneous estimation methods are proposed. Similarly, the image dehazing algorithms provide limited haze removal. We proposed a variational optimization method for transmis-

sion map estimation which focuses on structural details. The structural details are based on the notion that the objects at same depth level will have similar transmission map. Our developed algorithms improved the performance of existing object recognition methods. However, these methods require additional time for pre-processing step (enhancement or dehazing). To deal with this issue, we proposed direct detection algorithms for object recognition in low-light environment and hazy environment. The developed methods perform direct detection from the image without using the above-mentioned pre-processing step (enhancement or dehazing). Hence, the proposed recognition methods require comparably lesser time and are useful for real-time applications.

1.4 Main Contributions of the Thesis

The main contributions of this research are the development of new algorithms for low-light image enhancement, single image dehazing, object recognition in low-light, and object recognition in hazy environment.

1.4.1 Low-light Image Enhancement

In low-light image enhancement, the major challenge is to preserve the natural characteristics of the input image while avoiding the over-enhancement. In this research, we developed two new low-light image enhancement algorithms.

- First algorithm is based on the optimization approach for illumination estimation. In this work, a new structure-aware initial illumination estimation method is proposed leveraging the proposed multi-scale guided filter. We formulated and solved a new multi-objective optimization problem with a new regularization term to effectively preserve the structural details. Furthermore, a new adaptive illumination-adjustment is proposed to improve the overall lightness of the low-light image using the estimated illumination.

- The second algorithm is based on a deep learning architecture for image decomposition and enhancement. In this work, a novel branched encoder-decoder network is developed for simultaneous estimation of reflectance and illumination. The loss function is proposed, which considers the illumination and reflectance characteristics. Moreover, an image refinement module is developed to improve the color, contrast, and other details in the enhanced images.

1.4.2 Image Dehazing

In literature, various algorithms for single image dehazing are available. However, effective dehazing for distinct haze conditions remains a challenging task. A new algorithm is developed to perform dehazing using structural details based on the notion that the objects at same depth level will have similar transmission map.

- We developed a new method for the estimation of structure-aware initial transmission. We leverage the concept of adaptive bilateral filtering for the structure-aware smoothing. we present the notion that the structural details are same for the objects at same level of depth.
- We formulated a new variational optimization problem with regularization terms to preserve the structural details in the final transmission while smoothing the textural details. We used alternative direction minimization method to solve the variational optimization.
- We developed a technique for uniform atmospheric light estimation leveraging the dark channel.

1.4.3 Object Recognition in Low-light Environment

Low-light conditions present a myriad of intricacies for object recognition. Many approaches focus on enhancement before recognition, which requires additional time. In this research, we developed two direct object recognition methods.

- In the first approach, we developed a multi-exposure feature extractor. The multi-exposure images are generated using a single low-light image with distint exposure parameters. An adaptive refinement network is developed refine and combinedly process the features extracted from multi-exposure images. Finally, a new object recognition method is developed to directly recognize the objects from low-light images without using enhancement algorithms.
- In the second approach, we developed a dual-branch method for object detection in low-light conditions. We developed the notion of derived image to reinforce the low-light feature extraction. Furthermore, a feature adaptation network is developed to improve the relevant features from low-light images.

1.4.4 Object Recognition in Hazy Environment

The performance of object recognition methods degrades in hazy environment. Many approaches uses either dehazing before detection or domain adaptation, which have their own limitations. A new method is developed to perform object recognition in hazy environment.

- A new Bi-stream feature extractor network is developed to extract features at multiple-level from the hazy image and hybrid input. The notion of hybrid input is presented to facilitate the feature extraction of haze-relevant features.
- Th multi-level adaptive feature fusion network is developed to perform fusion of adequate features.
- A new dataset is developed to train object recognition models with annotated hazy images based on the Pascal-VOC dataset.

The developed methods are validated using the standard databased with other contemporary methods. The exhaustive experimentation and results demonstrates the superiority of the developed method with other contemporary algorithms.

1.5 Organization of the Thesis

The rest of the thesis is organized as follows: In **chapter 2**, the discussion related to the existing approaches for low-light image enhancement, image dehazing, object recognition, object recognition in low-light and object recognition in hazy environment. It also finds the research objectives. In **chapter 3**, nature preserving illumination estimation, and deep-simultaneous estimation network for low-light image enhancement are described. **Chapter 4** contains the discussion related to the variational optimization based single image dehazing. In **chapter 5**, the multi-exposure refinement network for object recognition in low-light environment is explained. In **chapter 6**, a bi-stream feature fusion method for object recognition in hazy environment is described. In **chapter 7**, the important conclusions are drawn from the proposed algorithms and it also gives the details of future works/applications.

Chapter 2

Literature Review

In this chapter earlier well-established and state-of-the-art algorithms related to object recognition in low-light and hazy environment are discussed. The majorly used strategy for object recognition in low-light and hazy environment uses enhancement/dehazing before object recognition. Thus, the literature covers Object recognition, Low-light image enhancement, Object recognition in low-light environment, Single image dehazing, and Object recognition in hazy environments. The chapter includes discussion of merits and demerits of existing algorithms, research objectives and brief of the proposed solutions.

2.1 Previous Works in Object Recognition

Object recognition involves indentifying and localizing objects within an image or a video. This field has evolved rapidly, particularly with the advent of deep learning. The methods can be broadly categorized as: traditional methods and deep learning based methods.

2.1.1 Traditional Object Recognition Methods

Sliding-window mechanisms, hand-crafted features, and classifiers are the base of most traditional object recognition methods. Viola and Jones [29] presented the cascaded classifier based on Haar features from an image. Agarwal and Roth [30] developed a sparse representation from side-views of cars for object recognition. Torralba *et al.* [31] proposed a mechanism for sharing common features using multitask learning among the classes to reduce the time complexity. Butko and Movellan [32] proposed a method to reduce the run-time of object detectors while dealing with images hav-

ing higher resolution. Felzenszwalb *et al.* [33] proposed a method to represent object classes that are highly variable using a combination of multi-scale deformable part models.

2.1.2 Deep learning based Object Recognition Methods

Deep learning based classifiers and detectors dominate recent object recognition methods. The DL-based object recognition methods are divided into two classes: two-stage methods (or region proposal based approaches) and one-stage methods (or single shot approaches).

2.1.2.1 Two-stage methods

Two stage methods are also known as region proposal based methods. The first stage in two-stage methods is responsible for generation of a sparse proposal set (e.g., Selective Search [34], EdgeBoxes [35], Deep-Mask [36, 37]). In contrast, the second stage is responsible for the classification and better object regions with improved bounding boxes. The two-stage detectors (like R-CNNs [38, 39, 40], and SPPnet [41]) achieve better accuracy than the one-stage detectors.

R-CNN (Region-based Convolutional Neural Network) [38] blended the potential of CNNs with the region proposal methods to enhance the performance of object recognition methods in an image. R-CNN identifies the potential object regions in an image ny region proposals and then use a CNN to extract features from each proposed region independently. These features are fed into classifier typically a Support Vector Machine (SVM) to determine the presence and class of the object. R-CNN requires high computational cost due to significant redundancy and not suitable for real-time applications. Fast R-CNN addressed several inefficiencies of the R-CNN, specifically in terms of speed and training process. Fast R-CNN leverages a shared convolutional feature map for the entire image. The region proposal are mapped to the feature maps rather than on input image. The feature are extracted using Region of Interest (RoI) pooling layer. These features are fed to the fully connected layers for object classification and localization. Faster R-CNN incorporated Region Proposal Network (RPN) into the architecture. The RPN is a fully convolutional network that predicts object bounds and obejtness scores at each position of the feature map. The RPN also introduces anchor boxes, predefined bounding boxes of various scales and aspect ratios, which aid in recognizing objects with different sized and shapes.

Further, many researchers attempted to enhance the performance of two-stage detectors by multi-layer exploitation [42, 43, 44, 45], contextual reasoning [46, 47, 48,

49], training strategies [50, 51, 52, 53], and architectural designs [54, 55, 56, 57, 58]. Two-stage methods are still less efficient than one-stage methods.

2.1.2.2 One-stage methods

One-stage object recognition methods provided a significant advancement in the field of computer vision, have revolutionized the approach to detecting objects within images. Unlike the two-stage methods that first generate region proposals and then classify them, one-stage methods streamline the process by simultaneously performing detection and classification in a single pass through the network. This approach dramatically enhances the computational efficiency, making it particularly suitable for real-time applications. One-stage recognition methods gained popularity due to their high efficiency.

You Only Look Once (YOLO) [59] is a pioneering architecture, which divides the input image into a grid, with each grid cell responsible for detecting objects within its bounds. YOLO model predicts both the bounding boxes and class probabilities in one evaluation, drastically reducing the processing time while maintaining considerable accuracy. YOLO improved detection speed by contemplating detection as a regression task and developing an end-to-end trainable method. Following YOLO, several iterations and improvements [60, 61, 62] have emerged, each enhancing aspects like detection accuracy, speed, and the ability to handle small objects. OverFeat [63] uses end-to-end trained CNN to detect and localize the objects. Single Shot MultiBox Detector (SSD) [64], which operates on multiple feature maps at different scales to improve the detection of objects of various sizes. SSD's ability to handle scale and aspect ratio variations makes it robust and versatile for diverse object sizes and shapes in an image.

RON [65] presented objectness-prior and reverse connection to reduce the search space, and detect objects using multiple levels of CNN. Based on SSD, various fundamentals [66, 67] are developed for training an object detection method from scratch. Object detection datasets [68, 69] contain samples of multiple classes; however, the unequal number of samples for each class creates a class imbalance. To deal with class imbalance, RetinaNet [70] proposed focal loss that reduces the weightage of easy negatives while increasing the weightage of hard examples. STOD [71] developed a scale-transfer layer that works on multiple scales for explicit exploration of consistency between the scales. SSDES [72] proposed a location-agnostic module and a semantic segmentation module for feature enrichment in the object detection method. DFPR [73] reconfigures the feature pyramid by developing presented global attention and lo-

cal reconfigurations. RFBN [74] proposed a receptive field block for enhancing the ability of feature discrimination and robustness. PFPN [75] developed a wide-width network for parallel feature pyramids to improve the detection process.

The above-discussed methods (one-stage, and two-stage) are developed and trained for object detection on images with normal light (where object-related details are clearly visible). The performance of the above-discussed methods (one-stage and two-stage) degrades while dealing with low-light images.

2.2 Low-Light Image Enhancement

In literature, many approaches for low-light image enhancement is developed to solve the issues related to low-light images. These methods can be broadly categorized as: traditional algorithms, retinex-based algorithms, deep learning based algorithms and other recent algorithms.

2.2.1 Traditional Algorithms

Many researchers have attempted to develop algorithms for image enhancement [76, 77, 78]. Low-light images contain most of the low-intensity pixels, and they are loosely considered low-contrast images. Histogram equalization-based approaches [79, 80] are the most straight-forward approaches. Several modifications of HE are developed to overcome these drawbacks. Abdullah *et al.* [81] proposed a dynamic HE (DHE) algorithm by dividing a histogram at local minima and keeping the dynamic range of each sub-histogram in proportion to the original image. Parihar and Verma [79] developed an entropy-based DHE and proposed an optimal division of histogram using entropy. Many algorithms [82, 83, 84] use a 2D histogram to incorporate local characteristics of the image to retain its natural characteristics. Parihar *et al.* [80] proposed an algorithm Fuzzy Contextual Contrast Enhancement (FCCE). FCCE uses fuzzy properties of the images to develop a fuzzy difference histogram (FDH). The authors further proposed a new intensity transfer function based on the intensity and local characteristics of the image to achieve natural-looking enhanced images. These algorithms focus on contrast enhancement rather than improving the illumination of the image. Thus, they sometimes fail to improve the contrast of low-light images. Contrast limited adaptive HE [85] deals with the saturation of bright regions and contrast enhancement. Another technique is gamma correction [86] which performs the mapping using a non-linear function. Chiu *et al.* [87] attempted to utilize the probability distribution of luminance pixels and adjust the contrast using modified gamma correction. These methods

consider the whole image to improve the lightness, which leads to artifacts and halo effects.

2.2.2 Retinex based Algorithms

Retinex [10, 13] based enhancement algorithms work reasonably well for low-light images. Many authors [2, 4, 88, 6, 89, 90] have attempted simultaneous estimation of illumination and reflectance. It is an ill-posed problem to estimate reflectance and illumination from a single image. Fu *et al.* [2] proposed a method for computing both components simultaneously with probabilistic-based prior (PIRE). The authors framed priors for both illuminance and reflectance to formulate maximum a posteriori (MAP). Fu *et al.* [91] proposed another approach for finding both components (i.e., illuminance and reflectance) from a single image based on the log of the Retinex model. These algorithm [2] works well for objects with proper illumination in a low-light image, but the enhancement is limited in darker regions. Li *et al.* [4] proposed SRLIME for simultaneous estimation of reflectance, illumination, and noise from an image. Hao *et al.* proposed semi-decoupled decomposition (LSDD) [6] for low-light image enhancement. These algorithms consider the inherent noise in the reflectance. SRLIME and LSDD algorithms provide good colour constancy. However, the algorithms [4, 6] result in degraded images due to over-smoothening in many images. Wang *et al.* [89] proposed a method for image decomposition using total generalizaed variation regularization and H^1 decomposition. In general, the complexity of algorithms estimating both illumination and reflectance is high. Thus, these algorithms may not be useful for real-time applications.

Another set of algorithms [1, 3, 17] estimate the illumination and use it to find the reflectance. The final enhanced image is obtained as the product of adjusted illumination and reflectance. Wang *et al.* [1] proposed Naturalness Preserved Enhancement Algorithm (NPEA) to deal with non-uniform illumination. NPEA preserves naturalness at the cost of illumination improvement. Thus, darker regions may not enhance adequately. Fu *et al.* [17] proposed a method that uses multi-fusion to deal with weak illumination .The algorithm uses a multi-fusion of Gaussian and Laplacian at multiple scales. The algorithm performs well in normal dark images but fails to provide significant lightness for a darker region. Guo *et al.* [3] proposed a low-light image enhancement algorithm using illumination map estimation (LIME). The performance of LIME is significantly good. However, over enhancement occurs in properly illuminated regions of low-light images. Moreover, LIME fails to retain color constancy.

2.2.3 Deep-learning based Algorithms

Researchers endeavored data-driven approaches in image dehazing, image deraining, underwater image enhancement and low-light image enhancement. To this end, Shen *et al.* [92] proposed MSR-net, a neural network-based approach to implement MSR and learn the mapping between low-light and ground-truth images in a supervised manner. LLNet [93] uses stacked sparse denoising auto-encoders to learn the signal features and then brightens the image in an adaptive technique. However, both MSR-net and LLNet generate the low-light training data synthetically using gamma correction and the addition of noise. Hence, they fail to characterize the natural low-light images and produce unnatural results ultimately. GLADNet [94] proposes a technique for global illumination estimation and then reconstructs the details to obtain the enhanced result. The detail reconstruction helps in the significant restoration of hidden details. However, GLADNet often leads to the introduction of artifacts and the loss of colors. Wei *et al.* [95] proposed Retinex-Net, for image decomposition and illumination mapping. Retinex-Net performs denoising of reflectance component. Though it works well, it employs an off-the-shelf denoising module for reflectance adjustment and ignores the color distortions. Zhang *et al.* [96] proposed kindling in darkness (KinD) for the decomposition of images. KinD [96] works on the limitations of Retinex-Net and flexibly adjusts the light levels while removing the visual defects. However, KinD is incapable of restoring the details hidden in extremely dark regions and smooths out the natural details as well while denoising. In contrast to other approaches, Zero-DCE [7] formulates the low-light image enhancement task as the problem of estimating image-specific curves and learns the curve parameters through a deep network. However, it ignores the inherent noise in the image. EnlightenGAN [97] is a generative adversarial network-based method that works without paired training data. It uses a global-local discriminator structure capable of handling varying lighting conditions. CSDGAN [8] is another technique that takes retinex theory as its basis. However, GAN-based methods generally amplify the noise and introduce various artifacts. Yu *et al.* [98] developed a modular low-light enhancement network using feature aggregation and attention. Li *et al.* [99] proposed a low-light image enhancement network based on attention method, and stacking method. Yu *et al.* [100] presented a low-light image enhancement method using two-stage decomposition and regulator method. Song *et al.* [101] proposed a spatial feature pyramid network for enhancement of low-light images. The method [101] uses three convolution kernels on spatial pyramid network to obtain features of distinct scales. The learning-based methods require a huge set of low-light and enhanced images. Moreover, the learning-based methods work well for

data similar to the test images. However, these methods may provide limited enhancement while dealing with significantly different images.

2.2.4 Other Recent Algorithms

Yu and Zhu [102] proposed a physical model-based algorithm, which considers the environmental light and scattering attenuation rate. Ren *et al.* [5] proposed LECARM, which uses a camera response model for the enhancement of low-light images. Ghosh *et al.* [103] proposed an adaptive bilateral filter-based textural smoothing. Wang *et al.* [52] proposed a fast method based on exposure fusion to enhance the details in multiple regions in the image using local laplacian filters. Rahman *et al.* [104] developed a wavelet transform based multi-scale decomposition and denoising method. Parihar *et al.* [88] proposed fusion-based [90] simultaneous estimation of reflectance and illumination. Tsai *et al.* [105] developed a multi-exposure fusion framework that fuses contrast limited histogram equalization, homomorphic filtering, and detail extraction for image enhancement. Xu *et al.* presented STAR [106] for simultaneous estimation of reflectance and illumination. It exploits the local derivatives to enhance the images. The methods show limited enhancement while dealing with variations in the lightness condition in an image.

2.3 Object Recognition in Low-light

Object recognition in low light environment is a challenging task, due to limited visibility of details in the images. Few techniques are developed to solve the issue of low-light environment for the object recognition methods.

Researchers [107, 108] attempted various approaches to detect objects from low-light images using enhancement or domain adaptation. Wu *et al.* [109] developed a method which uses cloud servers to perform enhancement of low-light images and detection using the edge computing mechanism. Huang *et al.* [110] presented a U-net-based image restoration for object recognition in low light. Sasagawa and Nagahara [108] developed the YOLO-in-the-dark method for the adaptation of Yolo-v3 [61] on RAW images to perform effectively under low light conditions. YOLO-in-the-dark method uses an existing enhancement method before detection using Yolo-v3. Hong *et al.* [111] synthesized a RAW image dataset using the MSCOCO and developed an enhancement method to improve visibility based on image quality. The detection is performed on the enhanced image for low-light images. The synthesis of RAW images for random RGB images is extremely difficult. LiteCortexNet [107] devel-

oped a decomposition method on a small synthetic dataset and then used the image reflectance for the detection process. Ren *et al.* [112] developed a U-net for enhancement and depthwise separable pyramid on an embedded platform for detection in low-light. Yuan *et al.* [113] developed a CLAHE-based enhancement method to improve low-light detection in transmission systems.

2.4 Single Image Dehazing

Researchers proposed various algorithms [114, 115, 116, 117] to perform image dehazing. Image dehazing algorithms can be broadly categorized as: traditional methods, prior-based methods, fusion-based methods, and learning-based methods.

2.4.1 Traditional Methods

Researchers [118, 119, 23, 120] developed multiple-input images based image dehazing algorithms. Nayar and Narasimhan [118, 119] proposed a binary scattering model for image dehazing. The authors analyse the scattering of atmospheric light to develop a method for recovering true scene radiance. Schechner *et al.* proposed a polarization-based multi-image algorithm [23] for image dehazing. This algorithm uses a polarization filter to estimate the range map, which helps in the estimation of true scene radiance. Shwartz *et al.* proposed a blind method [120] that separates airlight and recovers contrast. It estimates the degree of polarization using a probabilistic model to estimate the transmission map. These algorithms require multiple images to estimate true scene radiance. However, sometimes it is challenging to collect multiple images as per the requirement, which restricts the performance of these algorithms.

Another set of Researchers attempted single image dehazing. Tan [19] used Markov random field to develop a cost function. The cost function focuses on the contrast of the images and the smoothness of the airlight. Tan performed maximization of local contrast for image dehazing using the cost function. Fattal [121] proposed a mathematical model for image dehazing. It uses contrast and gradient of an image to estimate the enhanced image. Further, Fattal proposed color lines model [122] and automatic optical vector calculation [123]. Fattal used the property of pixels in small patches to develop color lines and estimate the atmospheric light. All of these algorithms result in the over-saturation of colors and the generation of halo effects.

2.4.2 Prior-based Methods

He *et al.* [20] proposed Dark channel prior (DCP) for single image dehazing. DCP estimates atmospheric light and transmission map from the dark channel. DCP uses the image matte of dark channel as refined transmission map. Image matte uses fine details to estimate the refined transmission map. The algorithm generates artifacts while dealing with the object having the same color as the airlight. Gibson and Nguyen [124] presented a new dark channel prior. The algorithm uses minimum volume ellipsoid approximation, which is inaccurate with pixels corresponding to bright objects. Zhu *et al.* [125] proposed color attenuation prior that focuses on depth map based estimation of transmission map. The estimation of the depth map from an image is ill-posed, which also affects the performance of CAP. Singh *et al.* proposed gradient channel prior [126, 127] that uses gradient for estimation of transmission map from a hazy image. Nair and Sankaran proposed an algorithm [128] that estimates transmission using surround filter and DCP. It is computationally simple; however, it fails to provide adequate lightness in dehazed images. The estimation of transmission map and atmospheric light is ill-posed problem. These algorithms sometimes lead to improper estimation of the transmission map, which causes the generation of artifacts and low-visible results.

2.4.3 Fusion-based Methods

Ancuti *et al.* [129, 130] proposed an algorithm that achieves image dehazing by fusing multiple versions of the input image. However, estimation of multiple versions from an input image for fusion is challenging task. Zhu *et al.* [131] proposed a multi-exposure image fusion method for image dehazing. It uses gamma correction to estimate multiple components of an image for fusion. Sometimes, multiple gamma-corrected images fail to capture the haze properties. Thus, these algorithm achieves limited enhancement while dealing with dense haze. Galdran *et al.* [132] presented a method that combines fusion and variational optimization for image dehazing. The algorithm estimates two components using variational optimization and then performs fusion to combine them for image dehazing.

2.4.4 Learning-based Methods

The application of deep learning provided significant improvement in image dehazing. Cai *et al.* [133] developed DehazeNet, a learning-based approach that estimates transmission map. DehazeNet works on the assumption that global airlight is constant, due

to which it fails to deal with both real outdoor and indoor images. Li *et al.* [134] proposed All-in-One Dehazing Network, a lightweight image dehazing network capable of producing haze-free image directly from the hazy image. Li *et al.* reformulated the atmospheric scattering model (4.2) to accommodate a K-estimation module. AOD-Net provides results with low brightness while dealing with images that are not from the dataset and fails to eradicate haze in some cases. Chen *et al.* [135] makes use of multi-scale dilated convolutions along with a encoder-decoder architecture and residual feature aggregation [136] from different levels. To address the problem of information loss during the convolution operation, [137] suggests the use of dilated convolutional blocks. GCANet fails to remove haze from object edges making the haze-free images look unnatural. Its performance on indoor images is much better than outdoor images which is indicated instances of abnormally contrasted patches. Liu *et al.* [138] proposed GridDehazeNet, an end-to-end multi scale framework for image dehazing which does not use the optical atmospheric scattering model (4.2) for predicting the haze-free image. The results produced contain significant number of patches with haze. GridDehazeNet performs significantly better on low haze indoor images compared to outdoor images, since the patchy output makes scene images look unnatural. Qin *et al.* [139] developed FFA-Net that treats different pixels and channels unequally by combining pixel attention and channel attention into a Feature Attention module. Dong *et al.* [140] attempts to restructure the well known denoising SOS (Strength Operate Subtract) [141] with a U-Net [142] for image-dehazing. Cycle-Dehaze [143] is an end to end pipeline which enhances the CycleGAN [144] architecture for for single image dehazing. This method does not employ any parameters of the classical optical atmospheric scattering model (4.1). Furthermore, it only requires the hazy input images and not the pristine ground truth images like most other pipelines do. The CycleGAN technique involves unsupervised training without paired examples, but requires two sets of images one hazed other clear although the images may not be paired. GCA [135], and DHS [145] methods uses genrative adversarial network (GAN) for image dehazing. GAN [135] based methods are difficult to optimize and have high chances of yielding an over-enhanced or under-enhanced image. Although learning-based strategies are effective, a bias can be observed with images similar to the training dataset. These techniques sometimes fail to deal with images that are distinct from the training images.

2.5 Object Recognition in Hazy Environment

Hazy environment generates various problems for a object recognition method. The performance of object recognition methods trained for normal images degrades while dealing with hazy images. Thus, few researchers attempted to overcome this issue by developing some specific solution for this problem. A very limited work has been done to solve this issue.

Object recognition in a hazy environment is less explored as compared to general object recognition. The existing methods for object recognition in hazy conditions focus on image dehazing [20, 76, 146] before detection. Another set of methods attempted to solve object recognition tasks in hazy environments by using domain adaptation. Sindagi *et al.* [147] developed domain adaptive object recognition using transmission map as the domain prior. Kumar *et al.* [148] presented the object detection after dehazing the image using dark channel prior [20] for autonomous vehicles. Kalwar *et al.* [149] presented a domain agnostic network by gated differentiable image processing to plug it into existing object recognition models. Liu *et al.* [150] developed image enhancement guided object recognition to refine the detection model with an enhancement branch. Li *et al.* [151] presented enhancement to dehaze image before detection. Tanwar *et al.* [152] presented an image dehazing model for object recognition. Qin *et al.* [153] proposed a detection-driven enhancement network which combines low-frequency and high-frequency components for multi-scale features. Li *et al.* [154] combined the dehazing and detection networks to propose detection-friendly dehazing.

2.6 Research Objectives and Proposed Solutions

Based on the analysis of the existing algorithms for low-light image enhancement, single image dehazing, object recognition for low-light, and object recognition for hazy environment, we have formulated the following objectives and proposed the solutions:

- Most of the low-light image enhancement techniques suffer from over-enhancement in the images with varying illumination. Thus, the algorithm for low-light image enhancement which does not over-enhance images with varying illumination is required. To overcome these issues, nature preserving illumination estimation algorithm and deep-simultaneous estimation network for low-light image enhancement are proposed. These methods are capable in dealing with the images with varying illumination.

- Existing image dehazing algorithms sometimes provide results with less details and halo artifacts. Thus, image dehazing method which can handle the effect of depth and atmospheric light simultaneously is required. To overcome the issues, we developed a variational optimization method for the estimation of transmission map. We presented the notion that objects at the same depth level have similar transmission characteristics. Furthermore, we develop a technique for uniform atmospheric light estimation leveraging the dark channel.

- Existing methods for object recognition in low-light uses image enhancement to deal with low-light conditions. Such type of methods involves two-steps: image enhancement and object recognition. The overall pipeline is bottlenecked by the performance and computation time of the enhancement algorithms. Thus, the algorithm for single-step object recognition in a low-light environment is required. To overcome these issues, a multi-exposure refinement network is developed for effective feature extraction and object recognition in low-light images.

- The challenge of object recognition for hazy images is relatively underexplored. The existing works for hazy object recognition focus on image dehazing task. Moreover, The computational cost of the algorithms to deal with hazy images is high. Thus, the algorithm for object recognition in a hazy environment is required. To address the problems of hazy object recognition, we developed a bi-stream feature fusion network for object recognition in hazy environment.

Chapter 3

Low-light Image Enhancement

This chapter presents a two methods for low-light image enhancement based on the Retinex theory. The chapter contains the discussion for Illumination estimation for Nature Preserving Low-light Image Enhancement (NPLIE) and Deep simultaneous estimation network (DSE-Net). The chapter includes the experimental results and analysis for both the methods and summary of the chapter.

3.1 Introduction

Images with good illumination are always desirable in the design of an effective object recognition system. However, many times the environment is not conducive to capture an image with good illumination and results in a low-light image. The other cause of low-light images may be variable lightness in a scene, night-time imaging, improper camera setting, bad lighting sources, etc. Sometimes an image may contain two types of regions: regions with properly illuminated objects and regions with poorly illuminated objects (i.e. the image with varying illumination). Low-light images pose challenges in object and feature detection and may degrade the performance of a object recognition system. low-light image enhancement may improve the visual quality of an image, and is the simplest solution to improve the effectiveness of an object recognition system. However, an inferior enhancement algorithm may result in the loss of crucial details and introduce undesirable artefacts. Thus, designing an effective algorithm for the enhancement of low-light images is a challenging task.

In literature, researchers have proposed low-light enhancement algorithms based on contrast enhancement [155, 156], retinex theory [10, 11, 12], deep learning [157, 158, 159, 160], and other approaches [5, 161, 162]. However, most of the low-light image enhancement algorithms suffer from one or more of the following limitations:

- Algorithms [2, 3, 4, 6] provide inadequate enhancement of non-uniformly illuminated images i.e. over-enhancement in the bright regions or under-enhancement in dark regions.
- Algorithms [2, 3, 4, 5] provide enhanced images with unnatural colors, i.e. fail to preserve natural colors.
- Algorithms [1, 3, 5] provide undesired artifacts in the enhanced images.
- Algorithms [3, 4, 6] are unable to restrict the dispersion effect in light dominant regions of the low-light image with varying illumination.
- Algorithms [3, 4, 6] result in loss of finer details.

In general, there is a trade-off between lightness enhancement and naturalness preservation, i.e. if an algorithm preserves naturalness, then the overall lightness is low and vice versa. To deal with the above-mentioned issues and improve the performance of object recognition, we developed two methods which are as follows:

- Illumination estimation for Nature Preserving Low-light Image Enhancement (NPLIE)
- Deep simultaneous estimation network for low-light image enhancement (DSE-Net)

The details related to the developed methods are discussed in the following sections.

3.2 Nature Preserving Low-light Image Enhancement (NPLIE)

In this section, we proposed a new low-light image enhancement based on the retinex model. In the retinex model, an image contains two components: illumination and reflectance. The estimation of illumination and reflectance from an image is ill-posed. The proposed algorithm estimates illumination of an image and leverages the retinex model to obtain reflectance component. We estimate the structure-aware initial illumination using the proposed multiple box (window) sizes in the guided filters [163]. We refine the initial illumination by formulating and solving a new multi-objective optimization function. The proposed objective function involves fidelity between the initial illumination and refined illumination. Thus, initial illumination plays a defining role in obtaining the refined illumination. At last, we apply the proposed adaptive

illumination adjustment to improve the lightness of the image based on the estimated illumination.

The importance of proposed NPLIE can be demonstrated using an image with varying illumination shown in Fig. 3.1, where bird is illuminated but other details are darker. It may be noticed that most algorithms are struggling to give natural enhancement. Some of the algorithms [5, 1, 3] try to enhance darker regions; however, already illuminated regions get over-enhanced in the process. The other algorithms [2, 6, 7, 8] try to maintain a natural look by avoiding over-enhancement but fail to enhance significant lightness in darker regions. On the other hand, the proposed NPLIE enhances darker regions of the image without over-enhancing already illuminated regions. The detailed analysis is shown in section 3.2.2.1.

The major contributions of this work are as follows:

- Proposed a new approach to estimate initial illumination leveraging the proposed multi-scale guided filter. The guided filter uses single box size due to which it is unable to achieve textural smoothening and structural preservation simultaneously. Thus, we propose multi-scale guided filter using multiple box sizes to achieve adequate textural smoothening while preserving the structural details.

- Formulated and proposed a new multi-objective optimization problem with a new regularization term to effectively preserve structural details in the refined illumination while smoothening the textural details. We solved the above optimization problem using Alternative Direction Minimization (ADM) algorithm [164].

- Proposed a new approach for adaptive illumination adjustment to improve the overall lightness of the low-light image using the estimated illumination. The proposed adaptive illumination adjustment increases the lightness in darker region, while maintaining the lightness of the bright regions by using the proposed illumination estimation to manage the lightness. Thus, it results in the natural enhancement of low light images.

- We performed an exhaustive analysis of the proposed approach on various datasets [95, 5, 1, 165, 166, 167] using qualitative and qualitative analysis and compared with state-of-the-art algorithms.

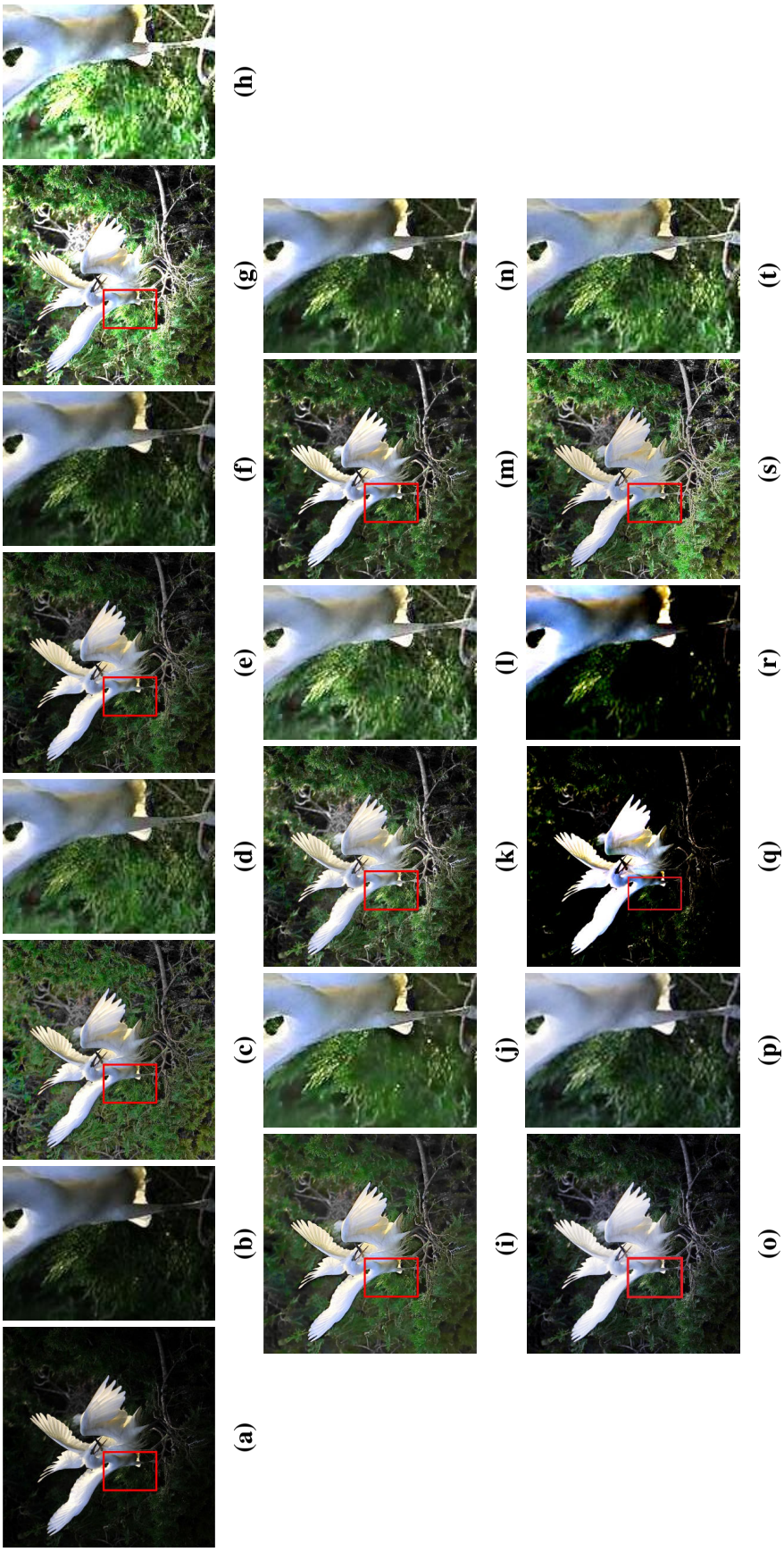
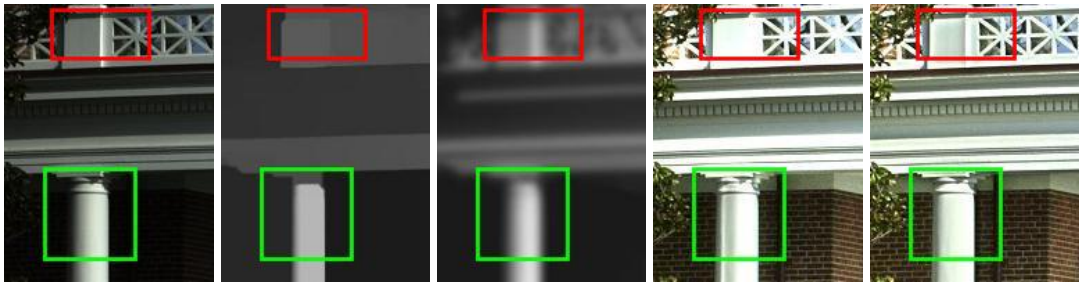


Fig. 3.1: Results of different algorithms on image birds. (a) Input image; (b) Patch of (a); (c) NPEA [1]; (d) Patch of (c); (e) PIRE [2]; (f) Patch of (e); (g) LIME [3]; (h) Patch of (g); (i) SRLIME [4]; (j) Patch of (i); (k) LECARM [5]; (l) Patch of (k); (m) LSDD [6]; (n) Patch of (m); (o) Zero-DCE [7]; (p) Patch of (o); (q) CSDGAN [8]; (r) Patch of (q); (s) Proposed method; (t) Patch of (s).



(a) (b) (c) (d) (e)

Fig. 3.2: Results of House. (a) Original Image; (b) Illumination of LIME; (c) Illumination of Proposed method; (d) LIME[3]; (e) Proposed Method.

3.2.1 The Proposed NPLIE

The algorithm uses three major modules for low-light image enhancement. First, NPLIE estimates initial illumination leveraging the proposed multi-scale guided filter. The algorithm uses multiple boxes to achieve textural smoothening while preserving the structural details. Second, we refine the initial illumination using the formulated multi-objective optimization function. NPLIE estimates reflectance using the refined illumination. Lastly, the algorithm adjusts the refined illumination using the proposed adaptive illumination adjustment to obtain the final illumination with corrected lightness. NPLIE combines the final illumination with the reflectance to provide an enhanced image.

In earlier work [14, 15], a smooth image is considered as an estimate of illumination. However, direct smoothening using a center/surround function does not capture the characteristics of non-uniform illumination. In dark regions, most of the information about illumination is available on the channel with the maximum intensity value [1, 17, 3]. A set of algorithms [1, 17, 3] use pixel-wise maximum across all color channels (bright-channel) as initial illumination. However, the bright-channel of low-light images may not represent true illumination. An image is indeed a 2D representation of 3D scenes (real world). Therefore, two adjacent objects in an image may be far distant in reality, and illumination in such cases may also vary. However, maximum intensity for two adjacent objects may be similar, and bright-channel will not distinguish their illumination. Moreover, in the surrounding region of a light source, the bright-channel estimation produces a scattering effect during low-light enhancement. It can be observed from Fig. 3.2 that the bright channel as initial illumination generates undesired artifacts. Thus, we proposed a structure-aware initial illumination instead of a bright-channel to address the above issue.

(a) b_{min} **(b)** b_m **(c)** b_{max} **(d)** $\hat{L}_{LIME}$ **(e)** $\hat{L}_{NPLIE}$

Fig. 3.3: Initial illumination with b_{min}, b_m, b_{max}, LIME (i.e. simple maximum intensity) and Proposed method in (3.3) respectively.

3.2.1.1 Structure-Aware Initial Illumination Estimation

In any optimization-based approach, initial illumination plays a defining role as it acts as a reference for refined illumination. One may notice in the red box of Fig. 3.2 that structural details are lost in Fig. 3.2(d) (in case of LIME), while details are preserved in Fig. 3.2(e) (in case of NPLIE). Similarly, the artifacts are prominently visible on the pillar in the green box of Fig. 3.2(d). The analysis of illumination estimation can help to understand these losses of details and artifacts, as shown in Fig. 3.2(b). One may notice that illumination is changing abruptly, which is quite unnatural. Thus, a structure-aware illumination estimation is highly desirable in images with non-uniform light conditions. The textural details of the image have almost no effect on the spatial variation of the illumination. Thus, the illumination of an image may be considered to be independent of textural details. The algorithm should smoothen textural details while preserving structural details to achieve an actual (or near actual) estimation of the illumination. In the proposed method, we develop an algorithm for structure-aware estimation of initial illumination by proposed multi-scale guided filtering.

In guided filtering [163], there is a fundamental assumption of linear relation between the guidance image (I) and the smoothen image (I_b). The linear model in guided filter ensures that the smoothen image contain an edge only if the input image has an edge. However, near strong edges gradient of each parameter of the filter is expected to be very small. Therefore, it is expected that the filtering should preserve structural details while smoothening the textural details. However, it also depends on the box size and image content inside the box.

In case of a box with high variance, the smoothen image retains maximum details from the guidance image which causes structural details preservation. In case of a box with low variance, the smoothen image captures the mean of the box filter that causes smoothening of textural details. The smaller box generally results in low variance as

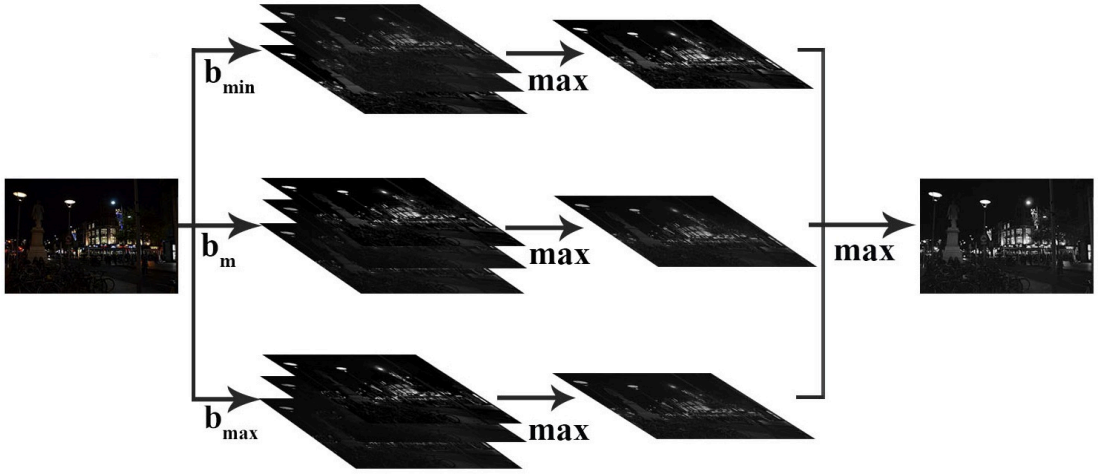


Fig. 3.4: Framework of Structure-Aware Initial Illumination Estimation.

the details do not change much within the box (except abrupt changes i.e. dominant structural details). Thus, it will perform textural smoothening with dominant structural details preservation. However, the overall smoothening will be low if we consider a very small box size (i.e. small neighborhood). As we increase the box size, the variance will increase primarily due to two reasons: (i) higher probability of including high detailed regions; and (ii) the variation inside the box is high due to larger spatial region coverage. The larger box size will result in better structural retention with slightly less contrast as it tries to pull all the values near the mean of the large region. From the above discussion, we noticed that smaller box size results in better textural smoothening with only dominant structural details preservation. On the other hand, larger box sizes result in better structural details preservation with compromised textural smoothening. Thus, it is intuitive to use some intermediate box size between b_{min} and b_{max}. However, it is empirically observed that the intermediate box also fails to give desired illumination estimation and contains some drawbacks of both b_{min} and b_{max}. It shows that the guided filter with a single box size is unable to perform the textural smoothening and structural preservation simultaneously, which is required for illumination. We propose to use three different box sizes b_{min}, b_m, and b_{max} to obtain structure-aware smoothening. However, the above relation is an approximate relation; thus, we performed exhaustive experimentation on a large set of images to analyse the effect of the box size in section 3.2.2.2.

We empirically observed that box size eight provides good textural smoothening as shown in Fig. 3.3(a). Thus, we select the minimum box size as eight, i.e. $b_{min} = 8$. However, b_{min} fails to preserve the structural details near the pillar region in Fig. 3.3(a). On the other hand maximum box size i.e. b_{max} preserves the structural details;

however, it fails to smoothen the textural details significantly as shown in Fig. 3.3(c). The maximum box size, b_{max} is computed as:

$$b_{max} = round(\min(H, W)/2) \tag{3.1}$$

where H and W represents number of rows and columns of the image respectively. Intuitively, we consider intermediate box size (b_m) for trade-off between b_{min} and b_{max}. However, intermediate box size also fails to give desired illumination and contains some drawbacks of both boxes. b_m is computed as:

$$b_m = round((b_{min} + b_{max})/2) \tag{3.2}$$

The framework of structure-aware initial illumination estimation is shown in Fig. 3.4. In first step, guided filters with box sizes b_{min}, b_m, and b_{max} are applied to the low-light image. Then, the pixel-wise max operation is performed across all three color channels on each output images of guided filters. This results in three max channels corresponding to each box size (max guided channels). Finally, the pixel-wise max operation is applied on above three max guided channels to achieve the structure-aware initial illumination. Fig. 3.3 shows the results of initial illumination for smoothened image (using b_{min}, b_m, and b_{max}), simple maximum operation (used in LIME), and the proposed method. We consider maximum operation [1, 5, 3] to obtain the effect of light in the initial illumination. It can be observed from Fig. 3.3(d) that proposed initial illumination contains the effect of light, structural preservation (at top near the grill), and low textural details. For notational convenience, let us consider $b_1 = b_{min}$, $b_2 = b_m$, and $b_3 = b_{max}$. The proposed algorithm computes initial illumination as given below:

$$\hat{\mathbf{L}} = \max_i \left\{ \bigcup_{i=1}^{3} \max_{c \in \{R,G,B\}} \left\{ \mathbf{I}_{b_i}^c \right\} \right\} \tag{3.3}$$

where $\hat{\mathbf{L}}$ denotes the initial illumination, $\mathbf{I}_{b_i}^c$ denotes c^{th} color channel of the output image of guided filter with box size b_i, max denotes pixel-wise maximum, and (R, G, B) denotes (red, green and blue) color channel of the image.

3.2.1.2 Formulated Optimization for Refined Illumination Estimation

The proposed initial illumination gives a reasonably good estimation of actual illumination and results in good enhancement. However, the estimated illumination still contains textural details and may result in over-enhancement. Thus, there is a need

for refining the initial illumination while preserving the structural details. To achieve a structure-aware smoothening, we formulate the following multi-objective optimization problem:

$$\min_{\mathbf{L}} ||\hat{\mathbf{L}} - \mathbf{L}||_2^2 + \alpha_n||\nabla\mathbf{I}_m - \nabla\mathbf{L}||_2^2 + \beta_n||\mathbf{G} \circ \nabla\mathbf{L}||_1 \tag{3.4}$$

where $\mathbf{L}$ is the refined illumination, $\hat{\mathbf{L}}$ is initial illumination as obtained by (3.3), ∇ denotes first-order derivative operator, $\nabla\mathbf{I}_m$ denotes maximum of ∇ from all channels of $\mathbf{I}$, $\mathbf{G}$ is a weight matrix based on initial illumination as discussed later in this section, $||\bullet||_p$ represents L^p norm operator, and α_n and β_n are the regularization parameters.

The first term (i.e. $||\hat{\mathbf{L}} - \mathbf{L}||_2^2$) of (3.4) provides fidelity between initial illumination and refined illumination. The second term (i.e. $||\nabla\mathbf{I}_m - \nabla\mathbf{L}||_2^2$) pulls the gradient of the refined illumination towards the maximum gradient of input image ($\nabla\mathbf{I}_m$). It tries to maximize the structural details in the illumination. The third term (i.e. $||\mathbf{G} \circ \nabla\mathbf{L}||_1$) provides smoothness to minimize textural details. The second term of (3.4) limits the smoothening effect introduced by the third term, while the third term prevents the second term from enhancing textural or noisy details. Thus, combining the second and third terms results in preserving structural details and smoothening textural details in the refined illumination.

In the third term, weight matrix $\mathbf{G}$ plays a vital role in smoothening of the textural details. The weights in matrix $\mathbf{G}$ are calculated as below:

$$\mathbf{G}_x = -log(\nabla_x\hat{\mathbf{L}}) \tag{3.5}$$

where ∇_x denotes first-order derivative along horizontal dimension.

$$\mathbf{G}_y = -log(\nabla_y\hat{\mathbf{L}}) \tag{3.6}$$

where ∇_y denotes first-order derivative along vertical dimension. The logarithm of derivatives assign a higher weight to textural details, while reducing the importance of structural details. Note that low gradient values characterize texture details, while high gradient values characterize structural details. Thus, the minimization of the multi-objective function significantly suppresses textural details and results in textural smoothening, while reducing the smoothening effect in structural details.

The efficient solution of (3.4) can be obtained using the Alternative Direction Minimization (ADM) algorithm [164]. The researchers [2, 4, 3] have proven the convergence of optimization problem like in (3.4) to the global optimum. Let us replace the

term $\nabla\mathbf{I}$ with an auxiliary variable $\mathbf{K}$ and rewrite objective function in (3.4) as:

$$\underset{\mathbf{L},\,\mathbf{K}}{\arg\min}\,||\hat{\mathbf{L}}-\mathbf{L}||_2^2+\alpha_n||\nabla\mathbf{I}_m-\mathbf{K}||_2^2+\beta_n||\mathbf{G}\circ\mathbf{K}||_1 \quad s.t.\nabla\mathbf{L}=\mathbf{K} \tag{3.7}$$

The multi-objective optimization which contains equality constraint can use Lagrangian multiplier to compute the optimum point. The equality constraint in (3.7) can be included in the objective function using Lagrangian multiplier ($\mathbf{M}$). The augmented Lagrangian function of (3.7) is given as:

$$\mathcal{L}=||\hat{\mathbf{L}}-\mathbf{L}||_2^2+\alpha_n||\nabla\mathbf{I}_m-\mathbf{K}||_2^2+\beta_n||\mathbf{G}\circ\mathbf{K}||_1+\varphi(\mathbf{M},\nabla\mathbf{L}-\mathbf{K}) \tag{3.8}$$

where $\varphi(\mathbf{M},\nabla\mathbf{L}-\mathbf{K})=\frac{\omega}{2}||\nabla\mathbf{L}-\mathbf{K}||_2^2+\langle\mathbf{M},\nabla\mathbf{L}-\mathbf{K}\rangle$ and $\langle\cdot,\cdot\rangle$ denotes element-wise multiplication, ω is a positive penalty term. The penalty term is required to control the rate of convergence to the solution [164]. The optimization problem can be solved easily by iteratively updating one variable while considering others as constant. The separation of the optimization problem (3.4) into sub-problems can simplify it. We divide the proposed multi-objective function into two sub-problems, sub-problem $\mathbf{L}$, and sub-problem $\mathbf{K}$. Sub-problem $\mathbf{L}$ is formed using all terms of (3.8) which contain illumination component. Sub-problem $\mathbf{K}$ is formed by considering the all terms of (3.8) which contain equality constraint. Let us solve both sub-problems using ADM[164] algorithm.

Sub-problem $\mathbf{L}$: Considering only terms related to the variable $\mathbf{I}$ of (3.8), we obtain the following sub-problem at i^{th} iteration:

$$\mathbf{L}^{i+1}=\underset{\mathbf{L}}{\arg\min}\,||\hat{\mathbf{L}}-\mathbf{L}||_2^2+\varphi(\mathbf{M}^i,\nabla\mathbf{L}-\mathbf{K}^i) \tag{3.9}$$

This is a well-known least square problem and can be solved by differentiating it with respect to $\mathbf{L}$ and putting it equal to 0:

$$\mathbf{L}=\frac{2\hat{\mathbf{L}}+\mathbf{D}^T(\omega\mathbf{K}-\mathbf{M})}{\mathbf{2}+\omega\mathbf{D}^T\mathbf{D}} \tag{3.10}$$

where D denotes the matrix containing Dx and Dy, Dx is the difference along rows and Dy is the difference along columns. Multiplication, transpose, and inverses of large matrices may be computationally inefficient. Thus, we use the 2D-FFT method by considering the condition of the circular boundary, and compute L as:

$$\mathbf{L}^{i+1}=\mathcal{F}^{-1}\left(\frac{\mathcal{F}\big(2\hat{\mathbf{L}}\big)+\sum_{d\in\{x,y\}}\mathcal{F}^c\big(\mathbf{D}_d(\omega^i\mathbf{K}^i-\mathbf{M}^i)\big)}{\mathbf{2}+\omega^i\sum_{d\in\{x,y\}}\mathcal{F}^c\big(\mathbf{D}_d\big)\cdot\mathcal{F}\big(\mathbf{D}_d\big)}\right) \tag{3.11}$$

where $\mathcal{F}$ denotes 2D-FFT operation, $\mathcal{F}^c$ denotes complex conjugate of 2D-FFT operation, $\mathcal{F}^{-1}$ denotes the inverse of 2D-FFT operation and **2** is a matrix equivalent to image size having all elements 2. Here, all operations are element-wise.
Sub-problem **K**: The sub-problem **K** is obtained from (3.8) by selecting terms related to **K** only. Thus, we have the following sub-problem at i^{th} iteration:

$$\begin{aligned}\mathbf{K}^{i+1} = \arg\min_{\mathbf{K}} \alpha_n||\nabla\mathbf{I}_m - \mathbf{K}||_2^2 + \beta_n||\mathbf{G}\circ\mathbf{K}||_1 \\ +\frac{\omega}{2}||\nabla\mathbf{L} - \mathbf{K}||_2^2 + \langle\mathbf{M}, \nabla\mathbf{L} - \mathbf{K}\rangle\end{aligned} \tag{3.12}$$

To solve the minimization problem (3.12), take derivative with respect to **K** and equate it to zero. The value of a pixel in an image cannot be negative. However, **K** can have negative pixel values because of higher weights, which is inappropriate while dealing with images. Thus, the computation of **K** requires normalization to deal with negative pixel values. Shrinkage operation [4, 3] helps in thresholding the values of the image. Hence, we can solve (3.12) using shrinkage operation:

$$\mathbf{K}^{i+1} = \mathcal{S}_{\frac{\beta\cdot\mathbf{G}}{2\alpha_n+\omega^i}}\left[\frac{2\alpha_n\nabla\mathbf{I}_m + \omega\nabla\mathbf{L}^{i+1} + \mathbf{M}^i}{2\alpha_n+\omega^i}\right] \tag{3.13}$$

where $\mathcal{S}_\phi[x] = sign(x)max(|x|-\phi, 0)$ and all operations are element-wise. The value of parameters α_n and β_n are chosen empirically. The parameters analysis shows that the best values of α_n and β_n are 0.5 and 0.1 respectively. The detailed analysis is discussed in section 3.2.2.6.
M and ω : The Lagrangian multiplier matrix **M** and penalty term ω can be updated as follows:

$$\begin{aligned}\mathbf{M}^{i+1} &= \mathbf{M}^i + \omega^i(\nabla\mathbf{L}^{i+1} - \mathbf{K}^{i+1});\\ \omega^{i+1} &= \omega^i\delta, \delta > 1.\end{aligned} \tag{3.14}$$

The above iterative algorithm needs suitable stopping criteria to achieve an optimal solution. We can either consider a threshold for the difference between $\mathbf{L}^{i+1}$ and $\mathbf{L}^i$ (i.e. L^i_{error}) or work out the maximum number of iterations. We considered the second option and found the maximum number of iterations using a large number of experiments with various image datasets. We analyze the results and found that after eight iterations, the proposed algorithm gives optimal results. Additionally, the algorithm often runs a larger number of iteration without significant improvement and results in the wastage of computation power if we consider the threshold difference. The whole procedure of estimating refined illumination is summarized in Algorithm 1,

Algorithm 1: Refined illumination estimation

Input: Proposed Structure-aware initial illumination $\hat{\mathbf{L}}$, α_n & β_n
Initialization: $\mathbf{L}^0 = \mathbf{K}^0 = \mathbf{M}^0 = 0, \omega^0$= 1, δ = 1.5.
while *not converged* **do**
    Update $\mathbf{L}^{i+1}$ via (3.11);
    Update $\mathbf{K}^{i+1}$ via (3.13);
    Update $\mathbf{M}^{i+1}$ via (3.14);
    i=i+1;
end
Output: Refined illumination $\mathbf{L}$=$\mathbf{L}^i$

which covers all details regarding variables initialization.

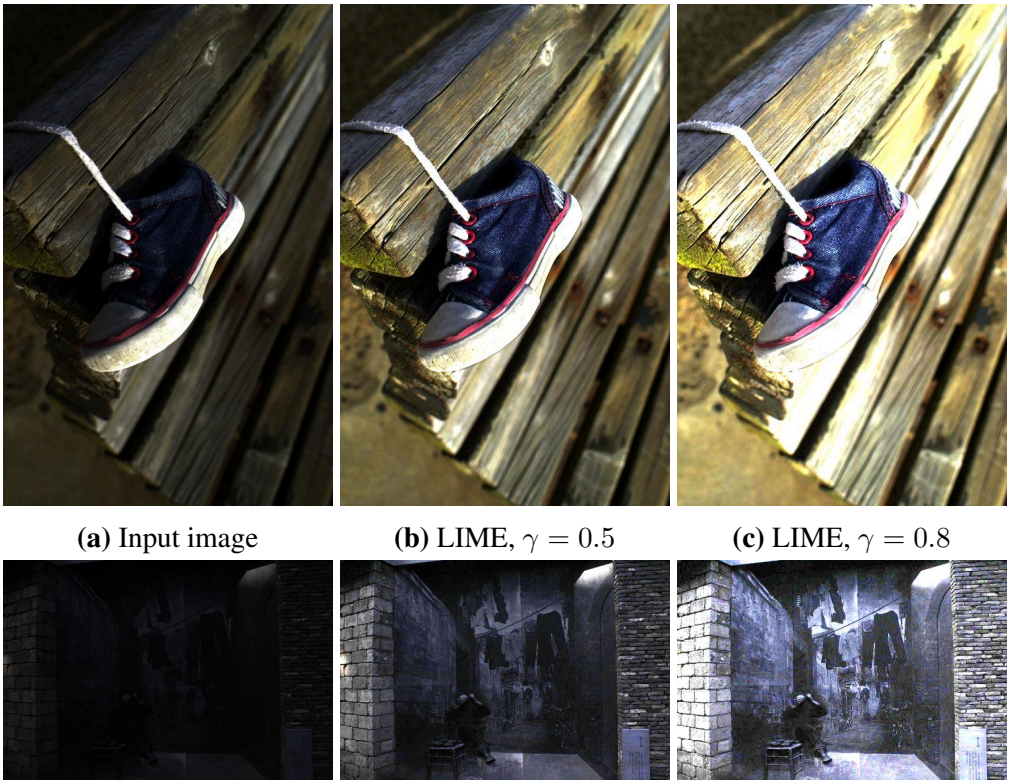
After estimation of illumination, we require to estimate reflectance. Remember illumination is independent of color channel; however, reflectance depends on color channel. The estimated reflectance is given as:

$$\mathbf{R}^c = \mathbf{I}^c / \mathbf{L} \tag{3.15}$$

where $\mathbf{R}^c$ is estimated reflectance of c^{th} color channel and division is element-wise.

3.2.1.3 Adaptive Illumination Adjustment

In traditional approaches for illumination adjustment, finding a suitable value for the adjusting factor (like gamma) for all kinds of images is not feasible because images with different illumination require different values of adjusting factor for the same algorithm. LIME and many other algorithms [3, 4, 2] use gamma correction for the adjustment of lightness in the image. For Fig. 3.5, LIME suggests suitable value of $\gamma = 0.8$, which works well for image 'Wall'. However, in the 'Shoe' image of Fig. 3.5, LIME results in over-enhancement around the white corners of the shoe. One may work out some lower value of gamma, say $\gamma = 0.5$. It addresses the problem of over-enhancement (ignoring some dark patches) in the image 'Shoe' but fails to improve illumination in case of the image 'Wall'. Thus, the same gamma value is unable to deal with different types of images. Even if an algorithm computes a different gamma value for different images, i.e. image-dependent gamma, it will not handle images with varying illumination. A global value of gamma is unable to produce faithful enhancement of all the regions in images with varying illumination. It can be observed from Fig. 3.6 that in one region (highlighted in the red box), LIME results in over-enhancement, and the other region (highlighted in the green box), it gives under-enhancement. Thus, to achieve natural enhancement, we require an adaptive illumination adjustment method



(a) Input image **(b)** LIME, $\gamma = 0.5$ **(c)** LIME, $\gamma = 0.8$

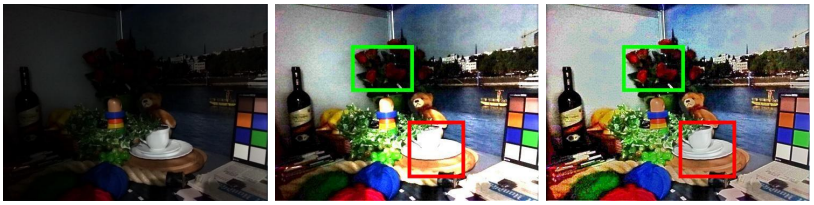
(d) Input image **(e)** LIME, $\gamma = 0.5$ **(f)** LIME, $\gamma = 0.8$

Fig. 3.5: Lightness comparison in Shoe(top row) image and Wall(bottom row) image after enhancement using LIME algorithm (with γ=[0.5,0.8]).

that adopts the value according to the local region of the low-light images. In this work, we developed an adaptive illumination adjustment method that uses the local illumination properties of the image. The adaptive illumination adjustment method is adapting automatically according to the required enhancement in the regions. The proposed approach eradicates over-enhancement and under-enhancement in different regions of the image with varying illumination. Thus, it helps in providing natural enhancement for all types of images irrespective of illumination property. The proposed illumination adjustment is given as:

$$\mathbf{L}_f = \mathbf{L}^{\cdot log(\mathbf{L})} \tag{3.16}$$

where $\mathbf{L}_f$ is the adjusted illumination estimation, and $\cdot$ represents the element-wise power operation. The logarithmic function improves consistency in illumination of the image. The proposed adaptive illumination adjustment uses the properties of estimated illumination properties to manage the lightness in the different regions of an image. The adaptive illumination adjustment improves the lightness of dark-regions



(a) Input image **(b)** LIME **(c)** Proposed method

Fig. 3.6: Input image, LIME with γ=0.8, and NPLIE with proposed adaptive illumination adjustment.

and restricts the lightness of regions with normal illumination. Note that the images are normalized between 0 to 1. The element-wise product of adjusted illumination and reflectance provides an enhanced image. The final enhanced image is given as:

$$\mathbf{I}_f^c = \mathbf{R}^c \circ \mathbf{L}_f \tag{3.17}$$

where $\mathbf{I}_f^c$ is final enhanced image.

3.2.2 EXPERIMENTAL RESULTS AND ANALYSIS

The proposed algorithm is analyzed and validated experimentally. The performance is evaluated with both quantitative and visual assessments. The time complexity is another important aspect to assess the performance of an algorithm. Thus, the computational cost is also analyzed for the proposed approach. We tested the proposed method using a large set of images (having varying illumination) taken from various image datasets: LECARM [5], NPEA [1], ExDark Dataset [165], LOL Dataset [95], EnlightenGAN [166], MIT-Adobe5k dataset [168] and NASA Database [167]. We compare the performance of the proposed algorithm with other state-of-the-art algorithms: NPEA [1], PIRE [2], LIME [3], SRLIME [4], LECARM [5], LSDD [6], Zero-DCE [7], and CSDGAN [8]. We have used codes and parameters (of other algorithms) as available on the author's websites/or as provided by authors for a fair comparison.

3.2.2.1 Quantitative Assessment

The quantitative assessment is always desirable to validate any experimental result. It may be easily verified that a quantitative measure occasionally may give better value for poorly enhanced image and inferior value for a high-quality image and vice versa. There is no single universally accepted quantitative assessment method available. There are two types of measures for quantitative assessments of the images:

referential and non-referential measures. In case of referential measures, good quality ground truth images are required. It is nearly infeasible to find an image database having pairs of good quality illuminated and dark (with varying illumination) version of the images. Even if such data is created, the enhancement is limited by the quality of the illuminated reference images. That is, the value of referential measure greatly depends upon the quality of the reference image. Thus, we have used four quantitative measures (frequently used in literature): Contrast Gain (CG) [2], Normalized Discrete Entropy (DE_N) [169], Information maximization based No-reference quality metric (NIQMC) [170], and Visual Information Fidelity (VIF) [171]. The measures for quantitative assessment was applied to all images in all datasets except MIT-Adobe5k dataset. In MIT-Adobe5k dataset, we have sampled 500 images that are either captured in night or contains low-light regions. Then, the mean metric value was computed and reported on all images of other datasets and sampled images of Adobe-MIT5k dataset. Higher structural details contribute to better contrast. Thus, higher CG is desirable for enhanced images. A higher value of DE_N represents more details and better enhancement. A higher value of NIQMC denotes better enhancement and more information in an image. Higher value of VIF indicates better enhancement and more nature preserving enhancement. The results of the quantitative assessment for various algorithms are shown in Table 3.1. The best values for quantitative metrics are bold. It may be noted from Table 3.1 that NPLIE achieves the highest mean value (for CG, DE_N, NIQMC, and VIF). LIME achieves comparable results; however, it may be observed from the visual analysis that LIME generates artifacts and provides poor color constancy. Therefore, CG, DE_N, NIQMC, and VIF metrics show that NPLIE is superior to the other state-of-the-art algorithms.

Statistical assessment: Analyzing all results manually may infer erroneous conclusions. Thus, we perform a statistical analysis of the results obtained by the proposed algorithm and other contemporary algorithms. We perform paired z-test [172] for the statistical analysis of the results. The pairs of proposed and other algorithms are used to perform the z-test. To assess the performance, the null hypothesis (H_0) and alternate hypothesis (H_1) are:

H_0: The performance of NPLIE and algorithm X is same.

H_1: The performance of NPLIE is better than algorithm X.

We conduct one-tail testing [173] to evaluate the enhancement. The probability of rejecting H_0 (i.e. p-value [174]) is obtained using z-distribution table [172]. We reject H_0 for a significantly small p-value and accept H_0 for a larger p-value. We test the hypothesis for all four measures (i.e. CG, DE_N, NIQMC, and VIF). Table 3.5 shows the z-statistics and p-value for the pair of NPLIE and other algorithms. It may be noted

Table 3.1: Mean values of Quantitative Measures for various algorithms

Metric	NPEA	PIRE	LIME	SRLIME	LECARM	LSDD	Zero-DCE	CSDGAN	Proposed
CG	0.57	0.54	0.61	0.42	0.55	0.55	0.88	0.91	**0.96**
DE_N	0.68	0.66	0.66	0.62	0.68	0.68	0.64	0.41	**0.77**
NIQMC	4.79	4.08	5.38	4.07	4.56	4.34	3.69	2.84	**5.45**
VIF	17.25	4.51	25.18	3.46	6.51	5.77	3.90	4.52	**32.31**

Table 3.2: Statistical analysis of one tail paired z-test for CG, DE_N, NIQMC, and VIF

Metric	Algorithm X	NPEA	PIRE	LIME	SRLIME	LECARM	LSDD	Zero-DCE	CSDGAN
CG	Z-stat	23.87	27.51	18.43	32.97	24.68	25.61	6.02	3.28
	p-value	0	0	0	0	0	0	8.89E-10	5.25E-04
DE_N	Z-stat	9.50	39.87	10.74	40.65	19.11	29.95	46.46	97.67
	p-value	0	0	0	0	0	0	0	0
NIQMC	Z-stat	68.25	62.23	4.44	61.07	40.43	44.62	66.17	67.02
	p-value	0	0	4.49E-06	0	0	0	0	0
VIF	Z-stat	5.67	12.69	2.13	13.21	11.64	11.95	12.59	12.21
	p-value	7.22E-09	0	0.017	0	0	0	0	0

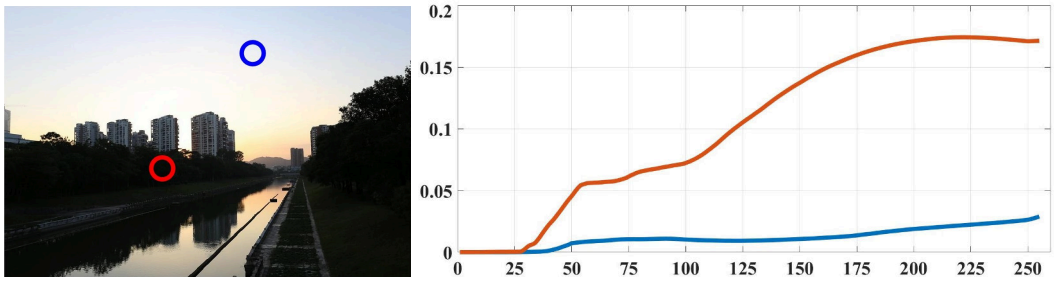
from Table 3.5 that all p-values are less than 0.05; thus, we can reject the null hypothesis (H_0) with at least a 0.05 level of significance. Hence, we can accept the alternate hypothesis that NPLIE achieves better enhancement than the other algorithms.

3.2.2.2 Effect of Box size on Initial Illumination

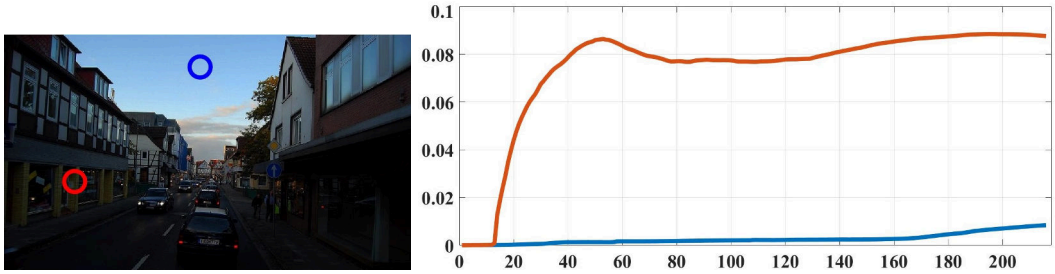
This section presents a detailed analysis of box size for initial illumination. Single box size is incapable of providing desirable initial illumination. We propose initial illumination estimation using three box size i.e. b_{min}, b_m and b_{max}. For proper estimation of initial illumination, we need to perform textural smoothening while preserving the structural details. We investigated the behavior of guided filters by varying the box size over 500 different types of images i.e. covering all kind of detail variations. We have shown three sample images for demonstration in Fig. 3.7. For the purpose of analysis, we have divided an image into three categories: low detailed regions, high detailed regions, and mixed regions (regions with low details for smaller boxes and high details for larger boxes, and vice versa).

Case I - Low detailed regions: In Fig. 3.7(a) and 3.7(b), the area highlighted with blue color shows the low detailed region in the image, and the blue line in the plot shows variance with respect to multiple box sizes. In such regions, variance is low for a smaller box, and there is a slight change in variance as we move from small box size to high box size because of fewer details. The low variance of the box results in high textural smoothening. In such regions, structural details are unavailable; thus, there is no effect of the box size. Smaller box size in such regions provides slightly higher smoothening than the larger box size, which can be observed from variance shown in the plot of Fig. 3.7(a) and 3.7(b). In low detailed regions, box size has less effect on variance as these regions do not contain structural details. Such regions perform textural smoothening for small and large box sizes due to low variance.

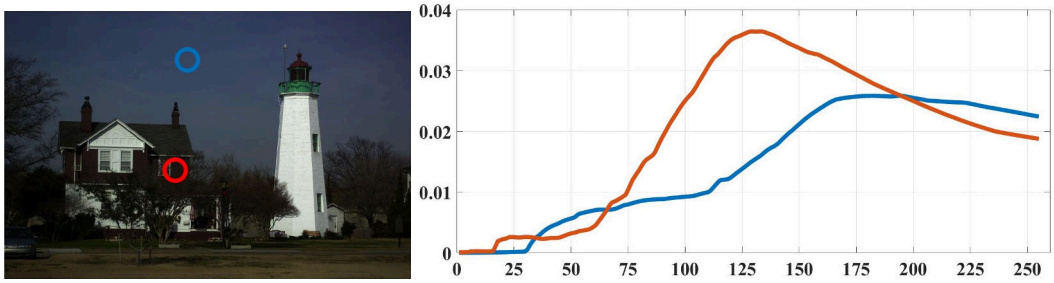
Case II - High detailed regions: In Fig. 3.7(a) and 3.7(b), the area highlighted with red color shows the high detailed region in the image, and the red line in the plot shows the variance with respect to multiple box sizes. Such regions contain textural and high structural details. Smaller boxes in such regions capture low variance, thus provides smoothening of the textural details. In high detailed regions of an image, the variance of a box increases sharply with an increase in box size up to maximum variance as shown by the red curve of the plot in Fig. 3.7(a) and 3.7(b). High variance results in low smoothening thus, results in the preservation of structural details. Moreover, the smoothened image with a large box size appears to be faded as it tries to pull all values towards the mean value of a large region. The smaller box size performs textural



(a) Sample 1



(b) Sample 2



(c) Sample 3

Fig. 3.7: Effect of box size on variance. The plot shows variance of the box with varying box size where x-axis denotes the box-size and y-axis denotes the variance of respective box. In the plot, red line and blue line shows the effect of box size in region highlighted with red color and blue color.

smoothening in the high detailed region, and the larger box size performs structural preservation.

Case III - Mixed Regions: In Fig. 3.7(c), the area highlighted with red and blue color shows the mixed region in the image. In Fig. 3.7(c), the area highlighted with blue color in the sky is a low detailed region. In this region, the variance is low for smaller box sizes, resulting in higher textural smoothening. However, when we increase the box size after some limit, it will capture structural details due to which variance will increase, resulting in structural preservation. On the other hand, the red area and red curve in Fig. 3.7(c) shows a region where initially the variance increases with an increase in box size, but once the box starts to capture the low detailed region, then the variance reduces. Thus, in such regions smaller box provides textural

(a) Refined 1 **(b)** Refined 2 **(c)** Refined 3 **(d)** Refined 4

Fig. 3.8: Assessment of refined illumination.

smoothening, and the larger box provides structural details preservation with slight textural smoothening as it captures comparably higher variance. Moreover, intermediate box size in such regions helps in structural preservation.

It is evident from the above analysis that the textural smoothening and structural preservation both cannot be achieved simultaneously using a single scale guided filter. As the textural smoothening and structural preservation depend on box size, characteristics of the image, and the neighborhood; thus, multiple box sizes are required to achieve textural smoothening and structural preservation simultaneously. The assessment of proposed initial and refined illumination is discussed in the next section.

3.2.2.3 Assessment of Initial and Refined Illumination

We analyzed the effect of the proposed initial and proposed refined illumination by comparing the performance of various combinations (listed below). The following combinations of refined illuminations are used for assessment:

- Refined 1 : the bright channel (used in [5, 1, 3]) with optimization used in [5, 3].
- Refined 2 : the proposed initial illumination with optimization used in [5, 3].
- Refined 3 : the bright channel (used in [5, 1, 3]) with the proposed optimization formulated in (3.4).
- Refined 4 : the proposed initial illumination with the proposed optimization formulated in (3.4).

These combinations (refined 1 to 4) are chosen to analyze the performance of the proposed initial illumination and the proposed optimization formulated in (3.4). The

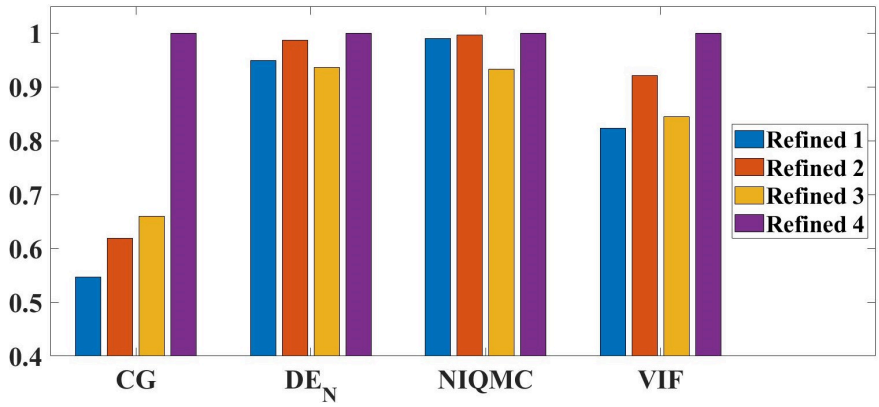
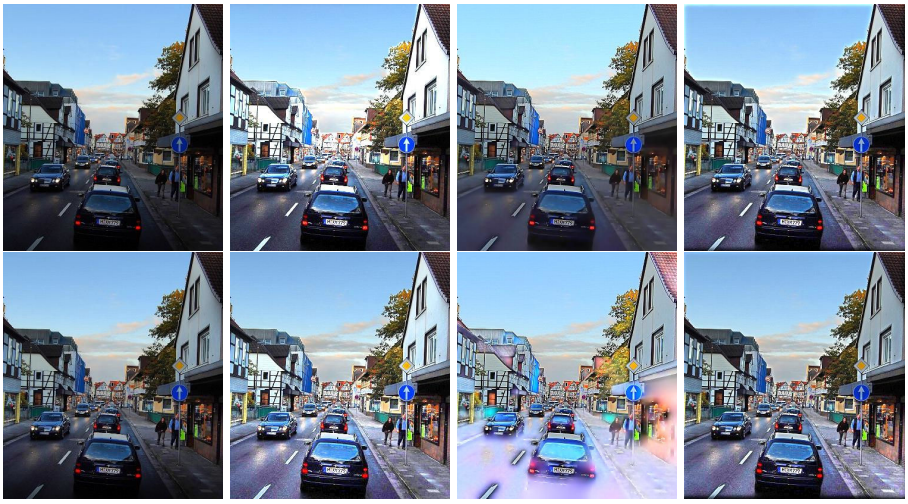


Fig. 3.9: Graph for quantitative comparison between several refined illumination

performance of the various combinations is analyzed by computing the performance metrics on enhanced images. We performed an exhaustive assessment on a large set of images, and some of the visual results are shown in Fig. 3.8. It is evident from Fig. 3.8(a) that refined 1 provides over-enhancement near the stands in the top image and near the cup in the bottom image. Refined 2 restricts the over-enhancement up to an extent, as shown in Fig. 3.8(b). It shows that the proposed initial illumination helps in restricting the over-enhancement. However, the overall enhancement is limited as refined 2 uses inadequate optimization. The result of refined 3 is shown in Fig. 3.8(c). Refined 3 provides comparably better performance than refined 1 and refined 2. It shows that the proposed optimized refinement restricts the over-enhancement after combining with the bright channel. It can be observed that the formulated optimization provides superior enhancement as compared to the optimization used in [5, 3] with the bright channel and the proposed initial illumination. Refined 4 achieves superior enhancement than other combinations. Refined 4 uses proposed initial illumination and proposed optimization, which helps in achieving better enhancement. Hence, the refined 4 (proposed method) is superior to any other combination of the initial illumination and optimization.

Further, we performed quantitative analysis of the resulting images using Contrast Gain (CG) [2], Normalized Discrete Entropy (DE_N) [169], Information maximization based No-reference quality metric (NIQMC) [170], and Visual Information Fidelity (VIF) [171]. For all performance measures, the higher value is the better. Details of the measures are discussed in section 3.2.2.1. Fig. 3.9 shows the quantitative assessment of the above-mentioned combinations (refined 1 to 4). Note: we normalized all measures using division with the maximum for representation purposes. Fig. 3.9 shows that refined 1 achieves good DE_N and NIQMC. However, CG and VIF are not good in case of refined 1. Refined 2 uses proposed initial illumination, which improves the



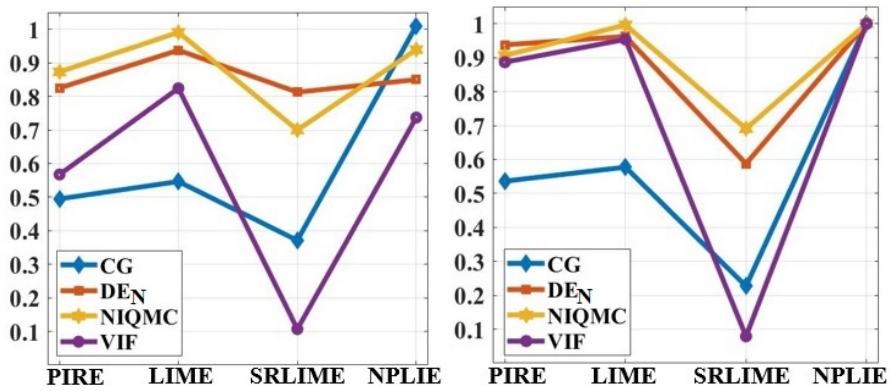
(a) PIRE **(b)** LIME **(c)** SRLIME **(d)** NPLIE

Fig. 3.10: Assessment of proposed adaptive illumination adjustment. Top row shows the results with traditional gamma (for illumination adjustment on various approaches), and bottom row shows the results with proposed adaptive illumination adjustment on various approaches.

value of CG, DE_N, NIQMC, and VIF as shown in Fig. 3.9. In case of refined 3, the initial illumination is not appropriate, which results in a decrease in the value of all performance measures except CG. It is evident from Fig. 3.9 that the proposed approach (i.e. refined 4) performs better than other combinations for all performance measures.

3.2.2.4 Assessment of proposed adaptive illumination adjustment

We replaced the traditional gamma correction for lightness improvement with the proposed adaptive illumination adjustment in PIRE, LIME, SRLIME, and NPLIE to analyze its effect. Fig. 3.10 shows the comparison of the proposed adaptive illumination adjustment and the traditional gamma with PIRE, LIME, SRLIME, and NPLIE. It is evident from Fig. 3.10 that the proposed adaptive illumination adjusment improves the performance of PIRE [2], LIME [3], and NLPIE. There is a slight improvement in lightness after using the proposed adaptive illumination adjustment with PIRE. Fig. 3.10(b) shows that there is a significant improvement in lightness when we use the proposed adaptive illumination adjustment with LIME. Fig. 3.10(c) shows results of SRLIME with traditional gamma and the proposed adaptive illumination adjusment. The results of SRLIME degrades with proposed adaptive illumination adjustment as its illumination estimation is inadequate. The proposed adaptive illumination adjustment improves the overall lightness in case of NPLIE as shown in Fig. 3.10(d).



(a) Traditional gamma (for illumination adjustment)

(b) Proposed adaptive illumination adjustment

Fig. 3.11: Quantitative assessment of proposed adaptive illumination adjustment.

The quantitative assessment of various algorithms with traditional gamma and proposed adaptive illumination adjustment is shown in Fig. 3.11. It is shown in Fig. 3.11(a) and 3.11(b) that the values of all performance measures improve for PIRE, LIME and NPLIE. However, the values of all measures decrease in case of SRLIME. The decomposed components from SRLIME algorithm are dependent on each other because of the optimization. In attempt to remove noise, SRLIME sometimes looses minor details which leads to inadequate estimation of the components. The proposed adaptive illumination adjustment uses the estimated illumination component for correction of lightness in the image, which is inadequate in SRLIME and may contain the details of reflectance. Thus, the illumination of SRLIME after the proposed illumination adjustment causes over-enhancement and distortion in the enhanced image. Hence, the proposed adaptive illumination adjustment fails to achieve better enhancement with SRLIME.

3.2.2.5 Visual Assessment

As discussed in the earlier section that there is no universally accepted quantitative measure for image assessment. Thus, visual assessment of the results becomes a necessity. An exhaustive visual assessment of the results from the proposed algorithm and other contemporary algorithms is performed. For easy reference, some of the regions of the images are highlighted in colored boxes, and the magnified view is shown adjacent to each image. The sample results are shown in Fig. 3.1 and Fig. 3.13-3.12.

The enhanced images of NPEA shows that it provides limited improvement of lightness, less contrast and over-enhancement in some regions. The highlighted region (red box) in Fig. 3.13(c) shows over-enhancement near cup. Fig. 3.14(c) shows less

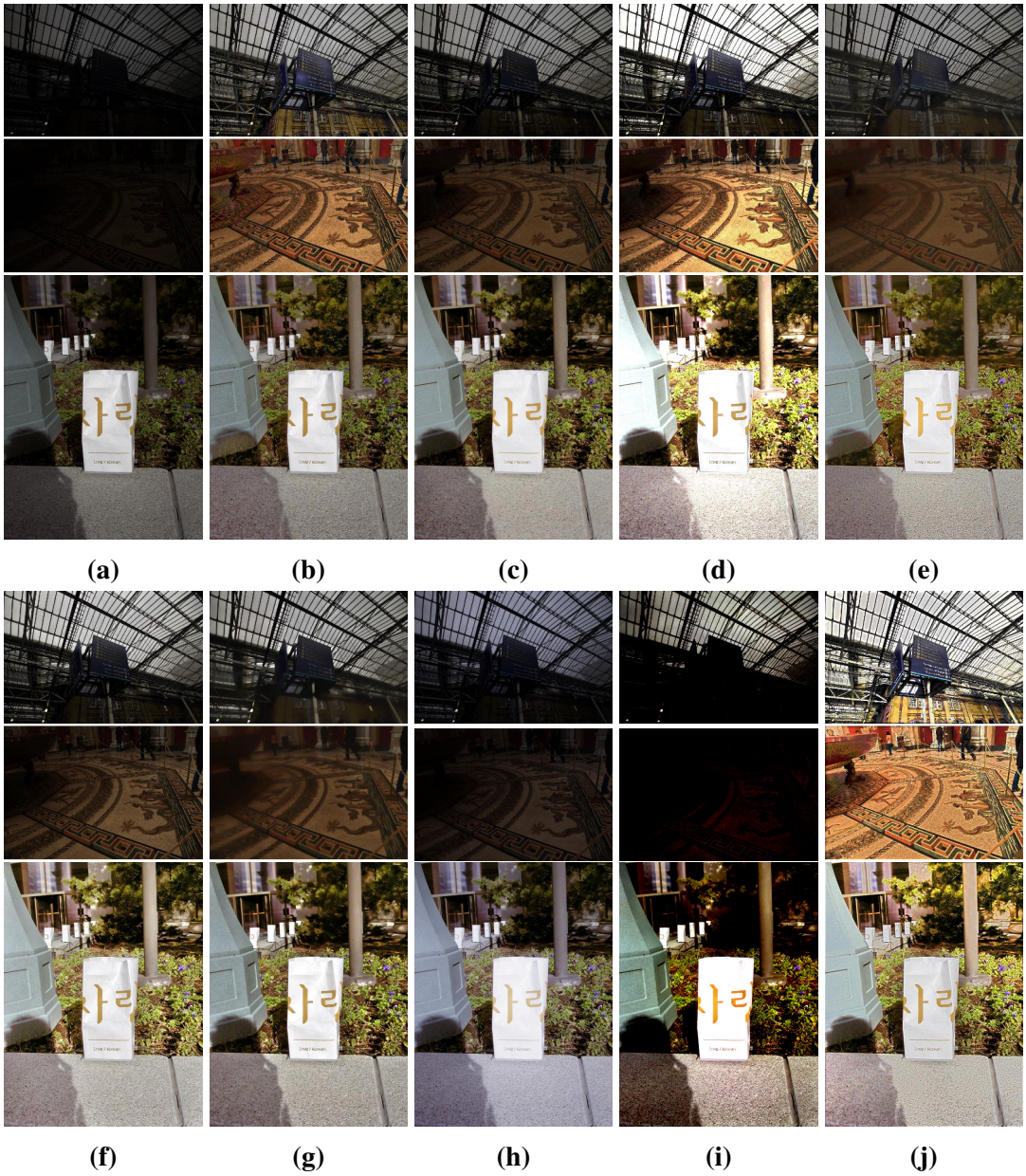


Fig. 3.12: Results of different algorithms on sample images. (a) Input image; (b) NPEA; (c) PIRE; (d) LIME; (e) SRLIME; (f) LECARM; (g) LSDD; (h) Zero-DCE; (i) CSDGAN; (j) Proposed method.

improvement in lightness. In Fig. 3.12(b), enhanced image of NPEA is acceptable but contrast is slightly low. PIRE provides limited enhancement as shown in Fig. 3.13(e), 3.14(e), and 3.12(c). It can be noted from the highlighted regions that PIRE fails to provide appropriate lightness in the dark regions. The enhanced images of LIME are shown in Fig. 3.13(g), 3.14(g), and 3.12(d). Fig. 3.13(g) shows LIME results in poor color constancy. Moreover, enhanced image of LIME shows over-enhancement in Fig. 3.13(h). SRLIME provides blurred details which causes poor

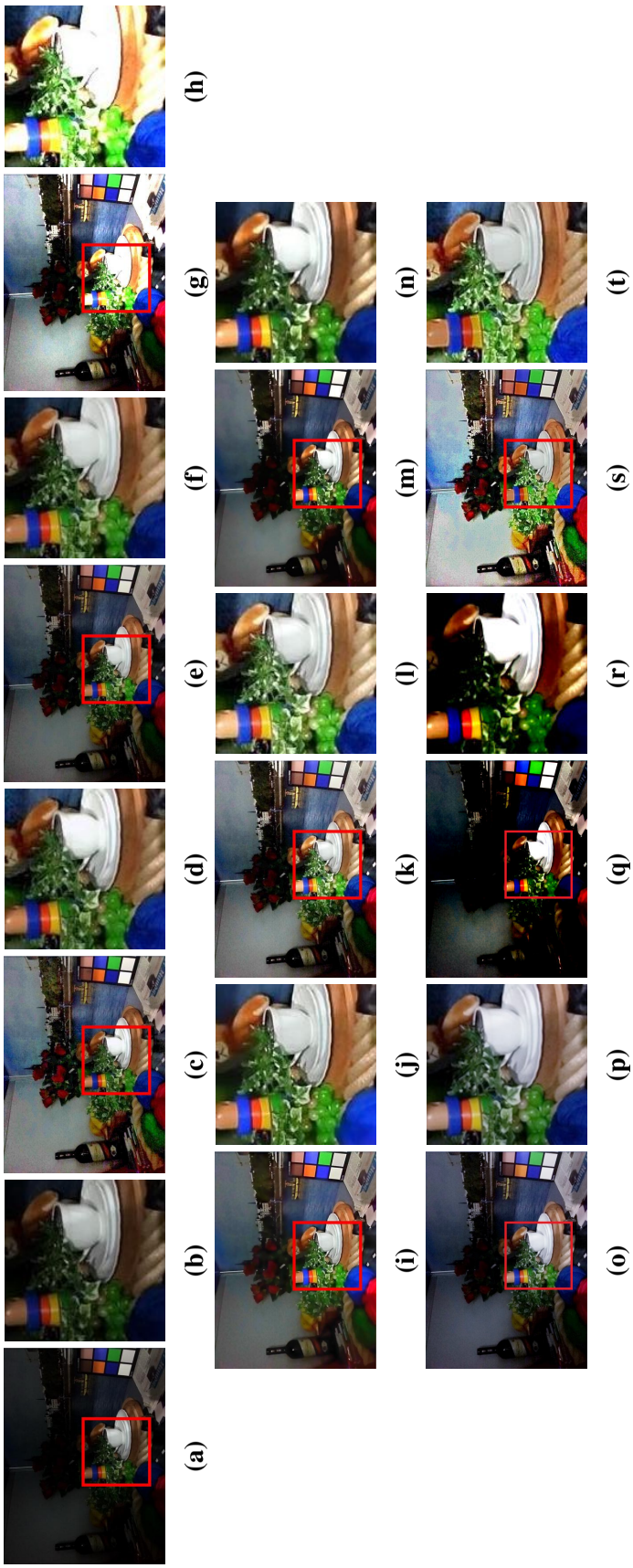


Fig. 3.13: Results of different algorithms on image corner. (a) Input image; (b) Patch of (a); (c) NPEA; (d) Patch of (c); (e) PIRE; (f) Patch of (e); (g) LIME; (h) Patch of (g); (i) SRLIME; (j) Patch of (i); (k) LECARM; (l) Patch of (k); (m) LSDD; (n) Patch of (m); (o) Zero-DCE; (p) Patch of (o); (q) CSDGAN; (r) Patch of (q); (s) Proposed method; (t) Patch of (s).

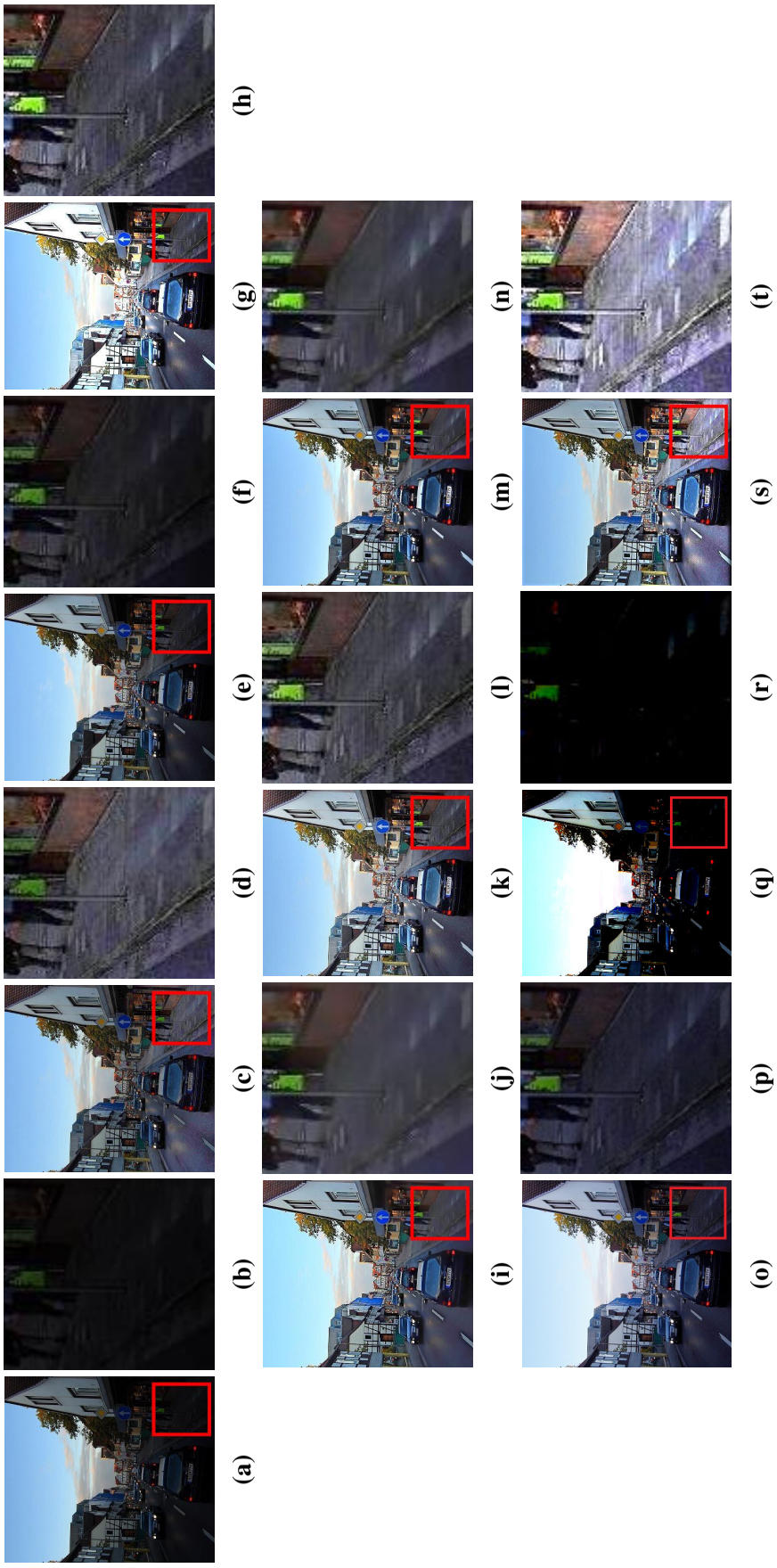


Fig. 3.14: Results of different algorithms on image road. (a) Input image; (b) Patch of (a); (c) NPEA; (d) Patch of (c); (e) PIRE; (f) Patch of (e); (g) LIME; (h) Patch of (g); (i) SRLIME; (j) Patch of (i); (k) LECARM; (l) Patch of (k); (m) LSDD; (n) Patch of (m); (o) Zero-DCE; (p) Patch of (o); (q) CSDGAN; (r) Patch of (q); (s) Proposed method; (t) Patch of (s).

visibility of details as shown in Fig. 3.13(i), 3.14(i), and 3.12(e). Fig. 3.14(i) shows that SRLIME provides limited lightness. Fig. 3.13(k), 3.14(k), and 3.12(f) shows the results of LECARM. It provides limited enhancement in low-light images and faded colors as shown in Fig. 3.14(l). In Fig. 3.13(m), 3.14(m), and 3.12(g), LSDD shows blurring effect; thus, the details in the image is unclear. Moreover, LSDD fails to achieve appropriate contrast. Zero-DCE provides limited enhancement as shown in Fig. 3.13(o) and 3.14(o). CSDGAN fails to improve the lightness in the dark regions as shown in Fig. 3.13(q) and Fig. 3.14(q). Both Zero-DCE and CSDGAN are deep-learning based methods and fails to achieve better results in generalized cases. Fig. 3.13(s), 3.14(s) and 3.12(h) shows the result of proposed algorithm. It can be noted that the proposed algorithm provides significant improvement of lightness in the low-light images as highlighted in Fig. 3.13(t) and Fig. 3.14(t). Moreover, the proposed algorithm restricts over-enhancement in light dominant region while enhancing the details of the dark regions in the images with varying illumination condition as shown in Fig.3.1(s), and 3.12(h). The visual comparison shows that the proposed algorithm achieves the superior enhancement of lightness with adequate contrast.

3.2.2.6 Parameter Analysis

The proposed algorithm involves two regularization parameters α_n and β_n. To analyze the effect of these parameters, we performed exhaustive experimentation, and it is found that the values of α_n and β_n should vary between 0 and 1. Since the optimization problem (3.4) is multi-objective optimization, considering negative values of α_n and

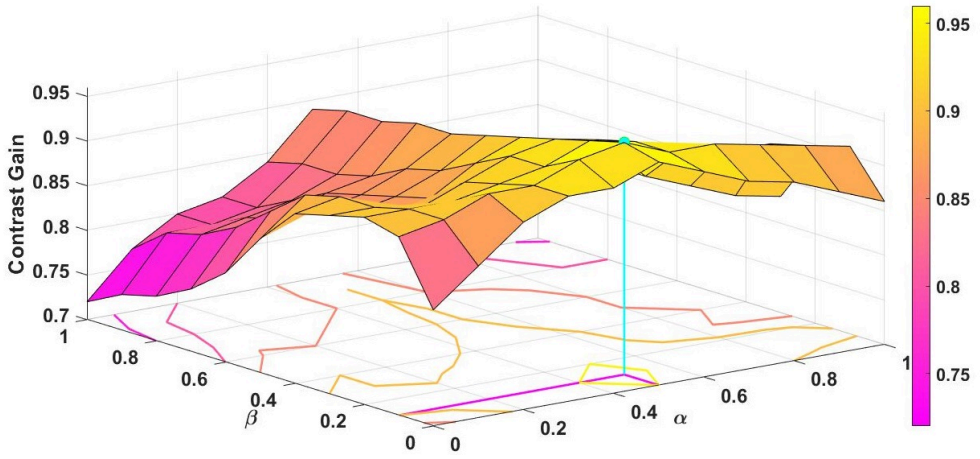


Fig. 3.15: Mean contrast gain for different pairs of (α_n,β_n). The contour denotes that pairs inside the contour have values based on the contour color. The colorbar represents the value of the contrast gain with respect to the color from the surface plot.

β_n does not make any sense. Parameters α_n and β_n regularize structural information and smoothening of textural details, respectively. Thus, if either of the parameter α_n and β_n have a value greater than 1, then one of the objectives (textural or structural details) will be dominant. For example, if parameter α_n is increased beyond value 1, the structural details will be dominant. Moreover, while capturing structural details, textural details may also get enhanced due to use of the gradients in the related term. It may result in unnatural enhanced images. Similarly, a large value of β_n may result in smoothening of structural details as well. This results in over-smoothen refined illumination. To avoid the dominance of the term related to the regularization parameter and limit the adverse effect, we consider values of α_n and β_n between 0 and 1. To arrive at optimal values of the parameters α_n and β_n, we perform experimentation over 500 images of diverse illumination and consider their mean contrast gain as a measure of enhancement. However, to have a better correlation between image qualities and values, we also analyze the best values using visual analysis. The results are plotted in Fig. 3.15. It may be noted that the best values of α_n and β_n are 0.5 and 0.1, respectively. Thus at α_n = 0.5, better structural details are captured without over-enhancing textural details. In case of α_n values less than 0.5, structural information of the image is not captured adequately. On the other hand, for α_n greater than 0.5, textural details are captured along with structural details. The parameter β_n at value 0.1 provides desirable smoothening of textural details without losing structural details. For case of β_n less than 0.1, smoothening of textural details is inadequate, while in case of β_n greater than 0.1 smoothening of structural details occurs as well. In our experimentation, the empirical values of parameters generate satisfactory results.

3.2.2.7 Computational Time

The computational time complexity plays an important role in the performance analysis of any algorithm. Thus, the computational time analysis of the various algorithm is performed on a set of 200 images with size 350x229. Table 4.4 shows the average computational time per image. In Table 4.4, (G) shows that the algorithm is running on ubuntu with titan-rtx GPU and 128GB RAM. Zero-DCE requires very lesser time then the proposed NPLIE. However, Zero-DCE is running on GPU whereas proposed NPLIE is using cpu for computations. Furthermore, the visual results of the proposed NPLIE is better than Zero-DCE. We use MATLAB 2018a on Windows10 running at 8GB RAM and core i5 processor @ 3.40GHz for CPU based algorithms. NPEA, SR-LIME, and LSDD are slower than other algorithms. These algorithms are not useful for real-time applications due to slow convergence. LECARM takes slightly lesser

Table 3.3: Average computational time per image (in seconds)

NPEA	PIRE	LIME	SRLIME	LECARM	LSDD	Zero-DCE	NPLIE
2.33	0.22	0.39	3.64	0.11	1.53	0.01(G)	0.21

time than the proposed algorithm. However, the visual quality of enhanced images produced by the proposed method is much better.

3.3 DSE-Net: Deep simultaneous estimation network

This section presents a novel branched encoder-decoder based deep simultaneous estimation network (DSE-Net) for low-light image enhancement. The proposed method first decomposes the input image into reflectance and illumination component using the proposed branched encoder-decoder model. Since the reflectance and illumination of an image contain different characteristics thus, both components require distinct focus for estimation. The proposed branched encoder-decoder model uses a separate decoder to estimate the reflectance and illumination. DSE-Net estimates new reflectance by element-wise division of the input image and the estimated illumination. After the decomposition of an image, the proposed DSE-Net aims at improving illumination using another deep network and then combines the improved illumination with the new reflectance. Finally, the proposed method performs image refinement to improve the visual context of the resultant image. The proposed model is capable of dealing with both low-light images and images with non-uniform illumination. The proposed low-light image enhancement is two-fold. First, the proposed approach deals with lightness by simultaneous decomposition and adjustment of the estimated illumination. The second module focuses on colour, contrast adjustment, and other visual aesthetics (like implicit noise) of the final enhanced image.

The **major contributions** of DSE-Net are summed up as follows:

1. We proposed a novel branched encoder-decoder architecture for simultaneous estimation of reflectance and illumination components. The proposed branched encoder and decoder network uses separate decoders to estimate illumination and reflectance from the encoded image.

2. We proposed a new loss function for image decomposition, which considers the characteristics of illumination and reflectance. The proposed loss function helps the decomposition network in textural smoothing of the illumination and estimating reflectance with finer details.

3. We proposed an image refinement module to improve the colour, contrast, and other details (by suppression of intrinsic noise) in the enhanced results to improve the quality further.

4. We conduct extensive experiments to show the effectiveness of our approach and its competitiveness with various state-of-the-art alternatives.

3.3.1 The Proposed Method

This section presents the proposed deep simultaneous estimation network (or DSE-Net) for low-light image enhancement. The framework of DSE-Net is illustrated in Fig. 3.16. The proposed DSE-Net contains three modules: image decomposition, illumination adjustment, and image refinement module. The image decomposition module works on the retinex model. The estimated illumination needs to be adjusted to adjust the lightness of a low-light image. DSE-Net achieves correction of lightness using the proposed illumination adjustment module. The proposed DSE-Net derives one more reflectance map, called $\mathbf{R}_{new}$, by performing element-wise division of the input image and the estimated illumination. It ensures that the details lost in the decomposition process are still preserved and used in further modules. Further, the proposed method combines the adjusted illumination and the estimated new reflectance to obtain the image with proper lightness. The lightness-adjusted image sometimes needs refinement to resolve artifacts that arise due to lightness correction. The image refinement module focuses on improving the color, contrast, and other visual aesthetics of an image. The following subsections contain a detailed description of the proposed method.

3.3.1.1 Image Decomposition

We propose a novel branched encoder-decoder architecture for image decomposition that inputs an image and decomposes it into reflectance and illumination simultaneously. The proposed image decomposition uses two decoder blocks that focus on the estimation of reflectance and illumination separately. Since reflectance and illumination have different characteristics; thus, both components are required to be treated separately using different decoders. The encoder of the proposed image decomposition module contains multiple down-blocks for feature extraction and encoding of an image. The down-block is a combination of convolution, max-pooling, and channel-attention [175], the detailed architecture is shown in Fig. 3.18. The proposed decoders are a set of up-blocks to estimate the reflectance and illumination of an image. The up-block contains upsampling, convolution, and channel attention; the detailed archi-

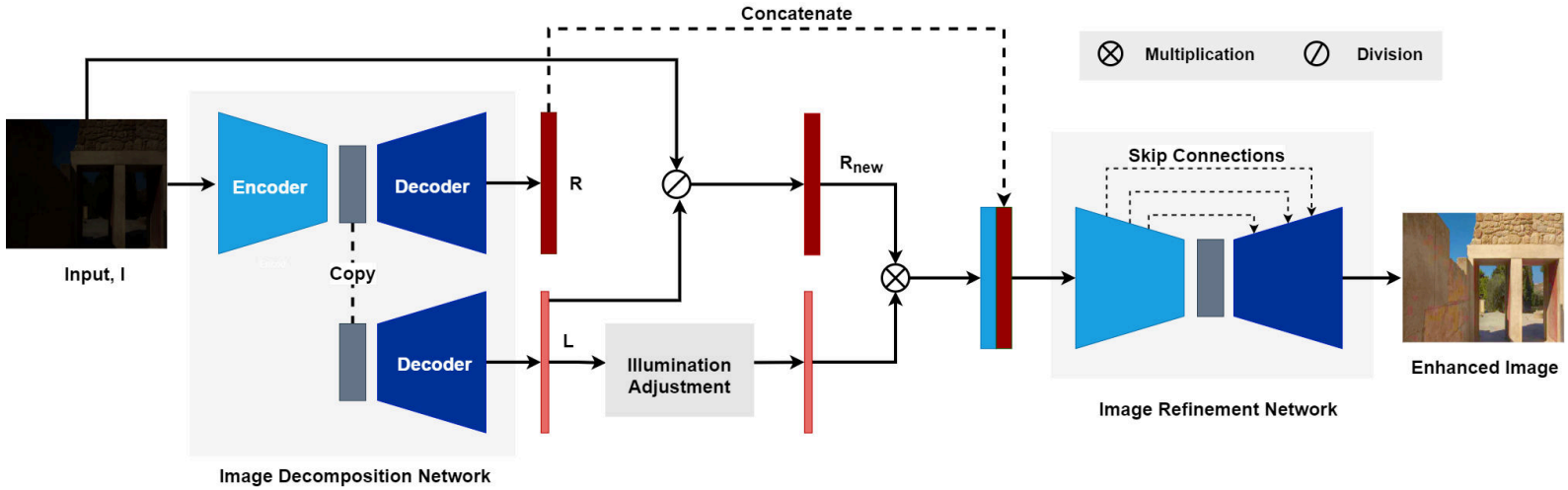


Fig. 3.16: The proposed framework of DSE-Net. From the functionality perspective, it is mainly divided into three modules, including image decomposition, illumination adjustment, and image refinement module.

tecture is shown in Fig. 3.18. The proposed DSE-Net uses encoder and branched decoder as shown in Fig. 3.17. The dotted line in Fig. 3.17 denotes the multiple numbers of intermediate blocks. DSE-Net contains five down-blocks and five up-blocks in the experiment.

Decomposing an image into two components is challenging as no ground-truth information is available for reflectance and illumination. To overcome the issue, we designed a loss function that incorporates various constraints related to reflectance and illumination. The proposed model uses the pairs of low-light images and ground truth images $[\mathbf{I}_l, \mathbf{I}_h]$ to train the decomposition network. Researchers [3, 1] discussed various illumination properties like local smoothness with structural-related details. Thus, we proposed the illumination constraint in the decomposition loss to maintain the relative structure of the illumination component with local smoothing, which can be denoted as:

$$\mathcal{L}_{ill} = \sum_{i \in \{l,h\}} \|max(\mathbf{I})_i - \mathbf{L}_i\|_1 + \left\| \frac{\nabla \mathbf{L}_i}{(\nabla \mathbf{P}_i + \epsilon)} \right\|_1 \tag{3.18}$$

where ∇ denotes the first order derivative ($\nabla = \nabla_h + \nabla_v$) across horizontal and vertical axis, max denotes the pixel-wise maximum operation, $\mathbf{P}$ denotes $max(\mathbf{R})$ and ϵ is the threshold value that restricts the value of the denominator term. Note that low gradient values characterize texture details, while high gradient values characterize structural details. The reflectance of an image contains properties related to the object details and textures. Thus, we introduced a reflectance constraint to maximize the contrast, structure, and texture-related details. The reflectance constraint focuses on the improvement of structural and textural details in the estimated reflectance, which can be expressed as:

$$\mathcal{L}_{ref} = \|\mathbf{R}_l - \mathbf{R}_h\|_2^2 + \sum_{i \in \{l,h\}} \|\nabla(max(\mathbf{I}_i)) * exp(c * \nabla max(\mathbf{R}_i))\|_1 \tag{3.19}$$

where $\|\cdot\|_2^2$ denotes the squared l^2 norm, $\mathbf{R}_l$ denotes the reflectance of low-light image, $\mathbf{R}_h$ denotes the reflectance of ground truth image. The term $\|\mathbf{R}_l - \mathbf{R}_h\|_1$ measures the dissimilarity between reflectance of low-light image ($\mathbf{R}_l$) and reflectance of ground-truth image ($\mathbf{R}_h$). Further, the reflectance contains the structure and texture-related details. Thus, the second term $\|\nabla(max(\mathbf{L}_i)) * exp(c * \nabla max(\mathbf{R}_i))\|_1$ maximizes the structural and textural details in estimation of the reflectance component. The proposed DSE-Net decomposes an image into two components and the combination of the component reconstructs the image. Thus, reconstruction loss must be used to maintain the

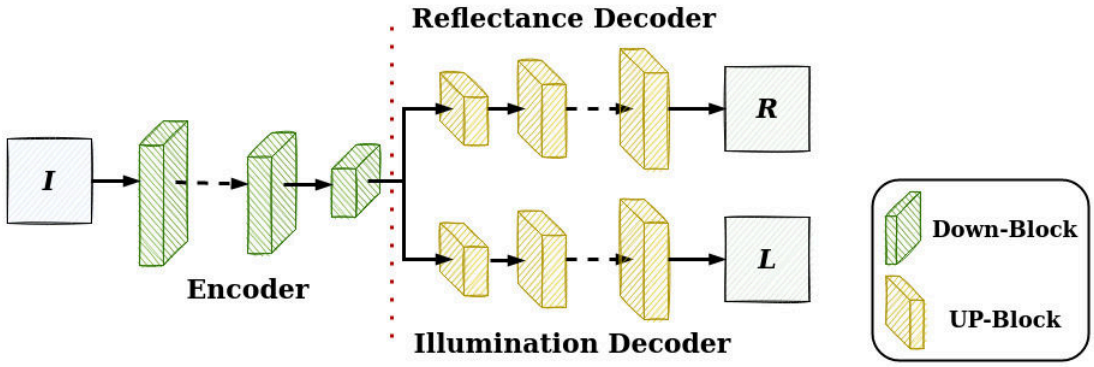


Fig. 3.17: Framework of proposed image decomposition module of DSE-Net. **I, R** and **L** denotes the input image, reflectance and illumination respectively.

properties of the retinex model. It can be expressed as:

$$\mathcal{L}_{recon} = \|\mathbf{I}_l - \mathbf{R}_l \cdot \mathbf{L}_l\|_2^2 + \|\mathbf{I}_h - \mathbf{R}_h \cdot \mathbf{L}_h\|_2^2 \tag{3.20}$$

In any data-driven approach, the loss function plays a vital role in obtaining effective results. In this work, we proposed illumination and reflectance constraints in the decomposition loss to achieve effective decomposition of a low-light image. The proposed loss function leverages the prior knowledge to effectively decompose an image into two components (i.e., reflectance and illumination) with the proposed novel branched encoder-decoder. The proposed combined loss for the image decomposition network can be expressed as:

$$\mathcal{L}_{total} = w_{ill}\mathcal{L}_{ill} + w_{ref}\mathcal{L}_{ref} + w_{recon}\mathcal{L}_{recon} \tag{3.21}$$

where w_{ill}, w_{ref}, and w_{recon} are the weights assigned to the losses which are decided empirically. The values of the reconstruction loss is 1, illumination constraint is 0.2, and reflectance constraint is 0.12.

The proposed decomposition network consists of two branches corresponding to the two desired outputs. We modified the bottleneck layers in the proposed branched encoder-decoder to propose a new architecture. The proposed encoder-decoder splits the bottleneck into two parts to preserve the properties of reflectance and illumination components. The architecture comprises convolutional (Conv) and rectified linear unit (ReLU) layers along with the pooling and upsampling operations. Finally, it uses the sigmoid activation function to estimate reflectance and illumination. The detailed architecture is shown in Fig. 3.16.

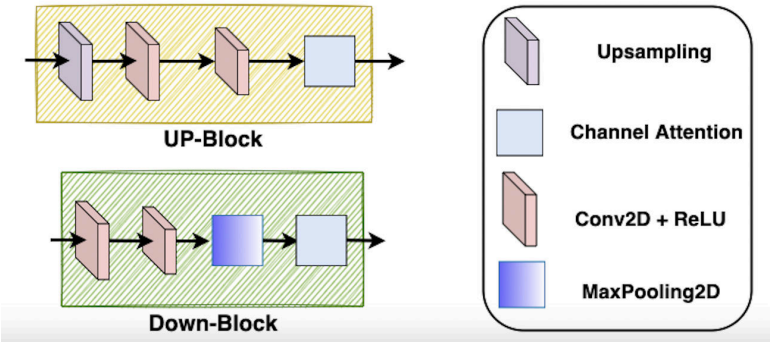


Fig. 3.18: Architecture design of UP-Block and Down-Block of DSE-Net.

In contrast to various illumination estimation techniques, the proposed method simultaneously estimates both the reflectance (**R**) and illumination (**L**) that, takes advantage of reconstruction, and produces better results. Further, we derive one more reflectance map, called $\mathbf{R}_{new}$, by performing an element-wise division operation on the input image and the estimated illumination. It makes sure that the details lost in the decomposition process are still preserved and can be used in further modules. After decomposition, DSE-Net requires lightness correction, for which we proposed the illumination adjustment module.

3.3.1.2 Illumination Adjustment

The adjustment of lightness plays a vital role in low-light image enhancement. The estimated illumination from the decomposition module contains low lightness, which needs adjustment to achieve enhancement. However, direct amplification may sometimes lead to over/under-enhanced results. Thus, an illumination adjustment module is required to treat lightness's dynamic nature in low-light images. We proposed a data-oriented illumination adjustment module, which can adaptively treat lightness. The proposed adjustment module analyses the illumination characteristics at multiple levels using the variation of receptive fields. Firstly, it processes the estimated illumination on a large (normal) scale; then, it uses downsampling to cover the larger region in the receptive field for illumination analysis. The proposed illumination adjustment module is U-Net-like architecture with channel attention. With the help of multi-level features of estimated illumination, the proposed model tries to capture the global and local distribution of the illumination. The adjustment in the estimated illumination of low-light image is needed to improve the lightness in the enhanced image. The illumination adjustment module requires a loss function to improve the lightness in the estimated illumination. Thus, the loss function must focus on improving the adjusted

illumination (i.e. from illumination adjustment module) with some ground-truth image. To maintain the properties of illumination, we are using a regularization term as well. The loss function to perform the illumination adjustment is:

$$\mathcal{L} = \|\mathbf{R}_l \cdot \mathbf{L}_{adj} - \mathbf{I}_h\|_2^2 + \left\| \frac{\nabla \mathbf{L}_{adj}}{(\nabla \mathbf{P}_l + \epsilon)} \right\|_1 \tag{3.22}$$

where, $\mathbf{L}_{adj}$ denotes the adjusted illumination, and $\mathbf{P}$ denotes $max(\mathbf{R})$. The first term $\|\mathbf{R}_l \cdot \mathbf{L}_{adj} - \mathbf{I}_h\|_2^2$ is used to reduce the difference between the illumination-adjusted image ($\mathbf{R}_l \cdot \mathbf{L}_{adj}$ i.e. product of reflectance and adjusted illumination) and the ground truth image ($\mathbf{I}_h$). Moreover, the smooth image is considered as illumination by many researchers [3, 1]. Thus, the second term $\|\nabla \mathbf{L}_{adj}/(\nabla \mathbf{P}_l + \epsilon)\|_1$ maximizes the smoothing in the illumination component by suppressing the textural details.

The algorithm then performs the element-wise multiplication of $\mathbf{R}_{new}$ and adjusted illumination to get the brightened image $\mathbf{I}_{new}$. In order to improve the color, contrast, and other visual details, the proposed method concatenates the image $\mathbf{I}_{new}$ with previously estimated $\mathbf{R}$ and feeds it to the image refinement module.

3.3.1.3 Image Refinement

The proposed method restores many details and obtains satisfactory results after the image decomposition. We propose the image refinement module to deal with various distortions introduced in the previous modules. It aims to produce sharp and natural-looking outputs while suppressing intrinsic noise. The image refinement network takes the concatenation of the lightness-adjusted image and the estimated reflectance component as input and produces the final enhanced image.

The proposed refinement module is an encoder-decoder architecture (similar to the decomposition module without the branched decoder) with skip connections and channel attention. The refinement module attempts to refine the unwanted artifacts and other issues from the lightness-adjusted image. The architecture uses multiple down-blocks and up-blocks. The down and up blocks are connected with a skip connection between them. With the help of down and up-block, the refinement module can focus on different parts of the image with varying receptive fields. The combination of down-blocks forms the encoder, and the combination of up-blocks forms the decoder. The layer details are somewhat similar to the image decomposition network except for the introduction of skip connections, which is illustrated in Fig. 3.16.

The proposed algorithm trains the image refinement network using a multi-objective loss function. The proposed loss function leverages three losses, namely, information

loss, color loss [176], and MSE loss. The three losses are computed between the final enhanced output and the ground truth image. In low-light images, it is challenging to enhance the details present in dark regions. Thus, to restore the details and preserve other information, the proposed method leverages the information loss [177]. We measure the information loss between the enhanced images ($\mathbf{I}_{enh}$) and the ground-truth images ($\mathbf{I}_h$). Instead of direct pixel-wise difference, the information loss uses the Euclidean distance between the activation maps obtained by the rectified linear unit (ReLU) activation of the VGG-16 network, which is pre-trained on the ImageNet dataset.

$$\mathcal{L}_{info} = \|\psi(\mathbf{I}_h) - \psi(\mathbf{R}_l)\|_2^2, \tag{3.23}$$

where $\psi()$ is the feature map obtained after $5th$ activation block of VGG-16 network.

The color loss [176] is obtained using Gaussian blur on images and then taking the pixel-wise difference. The final combined loss for the refinement unit can be expressed as:

$$\mathcal{L}_{total} = w_{info}\mathcal{L}_{info} + w_{color}\mathcal{L}_{color} + w_{mse}\mathcal{L}_{mse} \tag{3.24}$$

where w represents the weights assigned to the losses. This network can adjust color and contrast and deal with inherent noise to produce natural-looking output. The empirical values of w_{info}, w_{color} and w_{mse} are 2, 1, and 2 respectively.

3.3.2 Experiments

3.3.2.1 Implementation Details

The proposed method employs the LOL dataset [95] that provides 500 pairs of low-light images and corresponding ground-truth images. The dataset is split into two parts, 450 pairs for training and the remaining ones for evaluation. Moreover, we generate 1000 synthetic image pairs for training purposes. To generate synthetic images, we have corrected the lightness in the images using gamma values in 0.4 to 0.8. The patch size and the batch size are 128×128 and 10, respectively, for the image decomposition network. For the image refinement network, the patch size is taken as 256×256, and the batch size is taken as 10. The proposed approach employs an ADAM optimizer with a learning rate of 10^{-4} and other default parameters. We implement our work using the Tensorflow framework and train the entire network on NVIDIA-Titan-RTX GPU and Intel Xeon CPU with 128GB RAM.

3.3.2.2 Quantitative Comparisons

We performed comparison among the proposed DSE-Net and other state-of-the-art alternatives including LIME [3], STAR [106], GLADNet [94], KinD [96], ZeroDCE [7], and CSDGAN [8]. We perform the experiments by taking images from some widely used datasets, including LOL [95], LIME [3] MEF [179], NPEA [1], and DICM [180]. The LOL dataset comprises low-light and ground-truth image pairs. In contrast, the other datasets only include the images captured in diverse lighting environments and have no ground-truth. The results for various methods are obtained by using source codes that are available publicly following the suggested parameters. Further, We analyzed the proposed method in comparison to the state-of-the-art methods in an exhaustive manner.

We employ four metrics involving Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM) [178], Blind Image Spatial Quality Evaluator (BRISQUE) [181], and Universal Image Quality Index (UQI) [182] for quantitative assessments. A higher value for the metrics involving PSNR, SSIM, and UQI demonstrates better image quality, while a lower value is desirable for the BRISQUE score. PSNR, SSIM, and UQI are full-reference metrics and thus calculated using the testing image pairs available in the LOL dataset. On the other hand, BRISQUE is a no-reference metric and thus estimated using images from all the employed datasets.

Table 3.4 shows the quantitative analysis of PSNR, SSIM, BRISQUE, and UQI metrics for LIME [3], STAR [106], GLADNet [94], KinD [96], ZeroDCE [7], CSDGAN [8], and the proposed DSE-Net. It can be noted from Table 3.4 that the proposed DSE-Net outperforms the other state-of-the-art methods. SSIM value of the proposed method shows the higher quality of structural details in the enhanced image

Table 3.4: Quantitative comparisons by using various image quality assessment metrics. The best results are shown in red and the second best are shown in blue color.

Method	PSNR	SSIM	BRISQUE	UQI
LIME [3]	12.610	0.719	23.561	0.740
STAR [106]	10.070	0.413	25.506	0.345
GLADNet [94]	16.183	0.791	22.253	0.926
KinD [96]	15.225	0.815	22.334	0.914
ZeroDCE [7]	13.570	0.744	21.370	0.793
CSDGAN [8]	18.059	0.787	26.556	0.543
DSE-Net	21.109	0.823	17.183	0.928

Table 3.5: Statistical analysis of one tail paired t-test between algorithm X and DSE-Net

Metric	Algorithm X	**LIME** [3]	**STAR** [106]	**GLADNet** [94]	**KinD** [96]	**ZeroDCE** [7]	**CSDGAN** [8]
PSNR	t-stat	17.355	20.515	7.969	10.544	13.523	4.855
	p-value	7.24E-23	5.28E-26	1.07E-10	1.68E-14	1.84E-18	6.36E-06
SSIM	t-stat	9.888	14.358	2.796	0.938	3.954	3.39
	p-value	4.14E-13	1.74E-19	0.004	0.176	1.23E-04	6.77E-04
BRISQUE	t-stat	52.257	159.066	41.236	85.850	33.348	72.838
	p-value	6.15E-45	1.84E-68	5.23E-40	2.19E-55	1.21E-35	6.48E-52
UQI	t-stat	7.346	19.437	0.325	1.121	6.145	9.344
	p-value	9.69E-10	5.61E-25	0.373	0.134	6.99E-08	9.12E-13

compared to other methods. The proposed method achieves the best values for all the quantitative measures. The second best values for the different measures are achieved by different approaches, as shown in Table 3.4. Thus, the quantitative measures validate the superiority of the proposed DSE-Net over other techniques. The best results are shown in red, and the second-best is shown in blue color.

Statistical assessment: The manual analysis of all results may lead to erroneous conclusions. Thus, the statistical analysis of the other alternatives and DSE-Net is performed to analyse the results. We performed paired t-test [183] to analyse the results in statistical manner. The pairs of other methods (named, algorithm X) and proposed DSE-Net are used to perform t-test. The null and alternate hypothesis to perform the t-test and asses the performance are:
H_0: Performance of DSE-Net and algorithm X is same.
H_1: Performance of DSE-Net is better than algorithm X.
We evaluated the enhancement by conducting one-tail testing. we obtained p-value (i.e. probability for rejecting null hypothesis) using t-distribution table [183]. The H_0 is accepted for a large p-value and rejected for a smaller p-value. The hypothesis is tested using four measures (i.e. PSNR, SSIM, BRISQUE, and UQI). The t-statistics and p-value for the pair of other methods and proposed DSE-Net is shown in Table 3.5. It can be observed from Table 3.5 that most of the p-values are less than 0.05 (except SSIM in case of KinD, and UQI in case of GLADNet and Kind). Hence, the alternate hypothesis can be accepted which shows that DSE-Net achieves better enhancement.

3.3.2.3 Ablation Study

The proposed DSE-Net contains three sub-modules for low-light image enhancement. We perform various ablation studies to demonstrate the impact of each component of DSE-Net in achieving the desired output.

Ablation of image decomposition module: In the decomposition module of DSE-Net, the proposed branched encoder-decoder uses channel attention and multiple constraints (in loss). Table 3.6 shows the influence of each component in low-light image enhancement. The values in Table 3.6 show the mean PSNR and SSIM [178] score of the output image from DSE-Net without the mentioned component. The ablation of the respective model (as per Table) changes only the mentioned part of the model while keeping other modules the same as the final module. It can be observed from Table 3.6 that the proposed DSE-Net performs superiorly with respect to other variations. Further, it can be noted that each component contributes to the improvement of the proposed DSE-Net.

Table 3.6: Ablation study of Image Decomposition network. SSIM and PSNR values are shown below for DSE-Net without using mentioned module for training.

Method	PSNR	SSIM
without Channel Attention	17.43	0.78
Without Illumination Constraint	19.22	0.79
Without Reflectance Constraint	18.38	0.73
Without Reconstruction loss	17.09	0.72
DSE-Net	**21.11**	**0.82**

(a) **(b)** **(c)** **(d)** **(e)**

Fig. 3.19: Ablation study of Illumination adjustment module. (a) Input; (b) Estimated illumination; (c) Estimated reflectance; (d) Adjusted illumination; (e) Output.

Ablation of illumination adjustment module: As per the retinex model, multiplication of the estimated components generates the original image (low-light). To improve the lightness in an image, the estimated illumination needs adjustment. Thus, DSE-Net uses an illumination adjustment module to improve the lightness in an image. Fig. 3.19 shows the intermediate results i.e., estimated illumination, estimated reflectance, adjusted illumination, and the final enhanced image. It can be observed from Fig. 3.19 that the estimated illumination contains lightness properties that need correction to improve the lightness in the final image. The adjusted illumination shows that the proposed illumination adjustment module performs lightness correction to improve the overall lightness of the image. The final output shows that the enhanced image contains regions with proper lightness in the scene. Without the illumination adjustment module, the proposed model generates the original image without any correction in lightness.

Ablation of image refinement module: In the proposed DSE-Net, the image refinement module helps in the improvement of color, contrast, and other visual aesthetics of the final output. The refinement module uses a combination of loss and channel attention to improving the image. Table 3.7 shows the influence of each component of refinement module in low-light image enhancement. The values in Table 3.7 shows the mean PSNR and SSIM [178] score of the output image from DSE-Net without the mentioned component. The ablation of the respective model (as per Table) changes

Table 3.7: Ablation study of Image Refinement module. SSIM and PSNR values are shown below for DSE-Net without using mentioned module for training.

Method	PSNR	SSIM
without Channel Attention	19.11	0.80
Without information loss	18.84	0.78
Without color loss	19.22	0.80
Without MSE loss	17.28	0.77
DSE-Net	**21.11**	**0.82**

only the mentioned part of the model while keeping other modules the same as the final module. It can be noted that each component contributes to the enhancement of the final result and refinement module of the proposed DSE-Net. The proposed model achieves superior performance than other combinations mentioned in Table 3.7.

3.3.2.4 Visual Comparisons

We present the visual comparisons of various low-light images captured under diverse lighting conditions. The sample results are given along with the details.

Fig. 3.20 shows a 'castle' image with extremely dark surroundings and the corresponding enhanced results using various methods. It can be observed that all the methods, except STAR and KinD, give a competitive performance in the restoration of details. One may notice that the castle and its stand are entirely hidden in the input, as shown in Fig. 3.20(a). STAR fails in producing a bright output, as shown in Fig. 3.20(e), and similarly, the lightness improvement is limited in the case of LIME and CSDGAN. Though GLADNet restores the details, it produces unnatural output, as shown in the marked patch illustrated in Fig. 3.20(h). It can be observed that GLADNet introduces some artifacts near the light source. KinD smooths out natural details while handling noise and produces a blurred output, as visible in Fig. 3.20(i). The proposed approach outperforms other methods in preserving the naturalness and colors, as illustrated by Fig. 3.20(o). Further, the proposed method effectively deals with the problem of over-enhancement, as shown in Fig. 3.20(p), and produces a bright output.

Fig. 3.21 shows the results obtained for a 'landscape' image. It can be observed that GLADNet improves the visibility but alters the color of the sky, as shown in Fig. 3.21(g). STAR performs unsatisfactorily in detail restoration and brightness improvement, as shown in Fig. 3.21(e) and 3.21(f). KinD produces over-smooth results, as observed in Fig. 3.21(j), and it can be shown that the texture details are also wiped

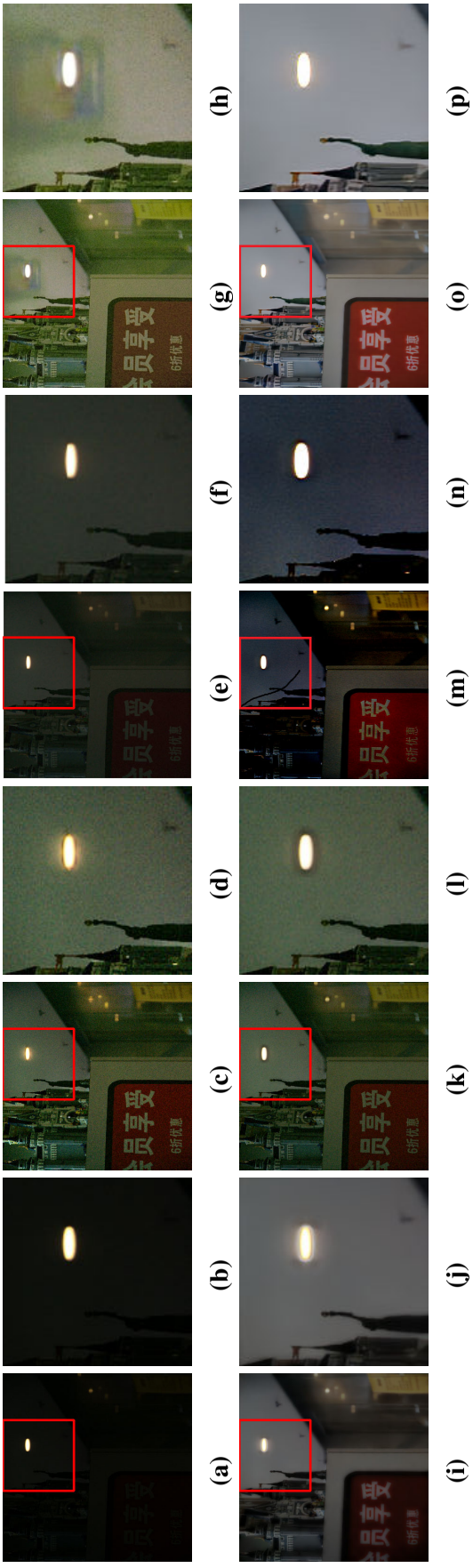


Fig. 3.20: Castle. (a) Input; (b) Marked patch of input; (c) LIME; (d) Marked patch of LIME; (e) STAR; (f) Marked patch of STAR; (g) GLADNet; (h) Marked patch of GLADNet; (i) KinD; (j) Marked patch of KinD; (k) ZeroDCE; (l) Marked patch of ZeroDCE; (m) CSDGAN; (n) Marked patch of CSDGAN; (o) Ours; (p) Marked patch of ours.

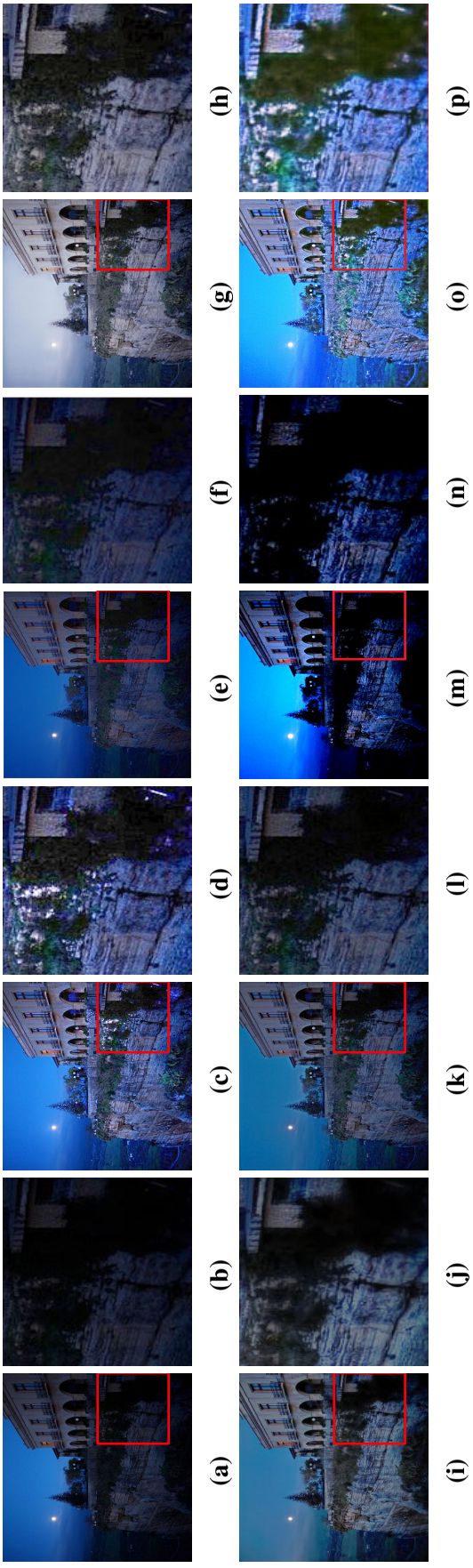


Fig. 3.21: Landscape. (a) Input; (b) Marked patch of input; (c) LIME; (d) Marked patch of LIME; (e) STAR; (f) Marked patch of STAR; (g) GLADNet; (h) Marked patch of GLADNet; (i) KinD; (j) Marked patch of KinD; (k) ZeroDCE; (l) Marked patch of ZeroDCE; (m) CSDGAN; (n) Marked patch of CSDGAN; (o) Ours; (p) Marked patch of ours.

Fig. 3.22: Toys. (a) Input; (b) LIME; (c) STAR; (d) GLADNet; (e) KinD; (f) ZeroDCE; (g) CSDGAN; (h) Ours.

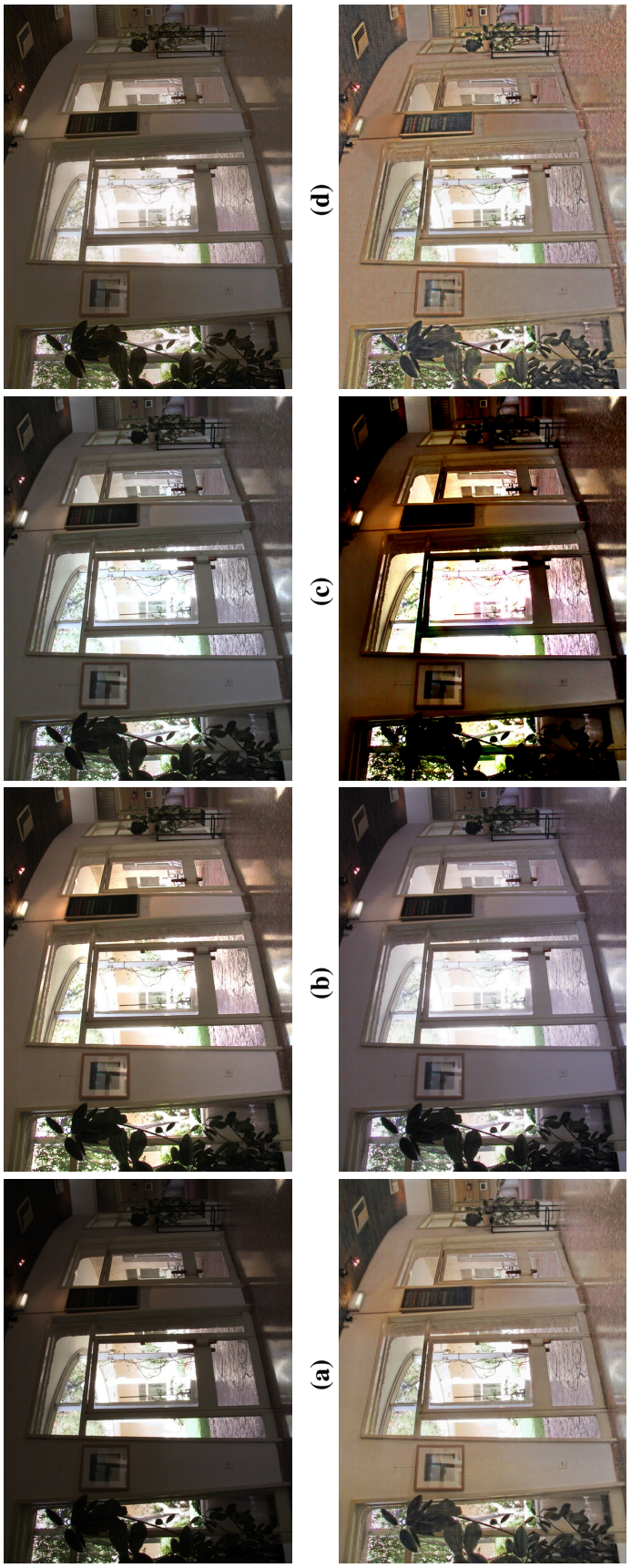


Fig. 3.23: Glass-Door. (a) Input; (b) LIME; (c) STAR; (d) GLADNet; (e) KinD; (f) ZeroDCE; (g) CSDGAN; (h) Ours.

out. CSDGAN tries to improve the contrast; however, the lightness improvement is not proper. The proposed method and LIME work best, as shown in Fig. 3.21(o) and 3.21(c), respectively. However, LIME often introduces artifacts and performs over-enhancement, as illustrated by the small white spots in Fig. 3.21(d). The proposed method achieves superior enhancement. The enhanced result contains more details and sharpness as compared to other methods.

Some more examples are provided in Fig. 3.22 and 3.23. From the observations, it can be observed that LIME can restore the details but often over-enhances the images. KinD focuses on noise removal, and as a result, it produces blurred outputs and restores a lesser amount of details in contrast to other methods. GLADNet and ZeroDCE perform well in terms of details restoration but often accompany noise and poor colors. Though STAR and CSDGAN give natural outputs, however, they somewhat fail in terms of lightness improvement. The proposed method shows excellent performance in terms of details restoration and outperforms other methods by producing less noisy, bright, and natural outputs.

3.3.2.5 Computational Time

The computational cost is an important performance aspect of any algorithm. Thus, the analysis of computational cost for various algorithm is performed on a set of 500 images with size 600×400. Table 4.4 shows the average computational time per image; (C) in table 4.4 shows that the algorithm runs on CPU. We run all algorithms (except LIME and STAR) on ubuntu running at 128GB RAM and Intel Xeon processor with nvidia titan-RTX gpu. For LIME and STAR, we used matlab running on windows 11 with i9 processor and 64GB RAM. STAR is slower than other algorithms. STAR algorithm is not useful for real-Time applications due to slow convergence. ZeroDCE takes slightly lesser time than the proposed DSE-Net because it uses light-weight network to estimate curve parameter. Further, the proposed DSE-Net produces results with better visual quality in the enhanced images.

Table 3.8: Average computational cost per image (in seconds)

LIME(C)	STAR(C)	GLADNet	KinD	ZeroDCE	CSDGAN	DSE-Net
0.47	2.02	0.39	0.37	0.06	0.11	0.09

3.4 Application in Computer Vision

One of the primary inspirations behind low-light image enhancement is to upgrade the performance of vision tasks. Thus, we demonstrate the potency of the proposed method in upgrading the performance of object detection and recognition task. We used the Tensorflow Object Detection API, which can detect the objects in images using machine learning and classify them into various categories. We test some low-light images and their corresponding enhanced results using the API. Further, the minimum support is taken as 0.25 for detecting an object.

Fig. 3.24 shows some of the examples. It can be seen that the API wrongly identifies the kitchen utensils and interprets them as mechanical fans in the first case. In contrast, things are correctly recognized after enhancement by the proposed method. In the second case, there is a significant improvement in the performance of object detection and recognition task. The object detection API could not recognize anything for the low-light image while it clearly labels the pillow and furniture for the enhanced counterpart. The developed enhancement methods help in improvement of object recognition; however, the time required is not suitable for real-time applications. Further, the quantitative analysis shows a limited improvement in performance under low-light conditions.

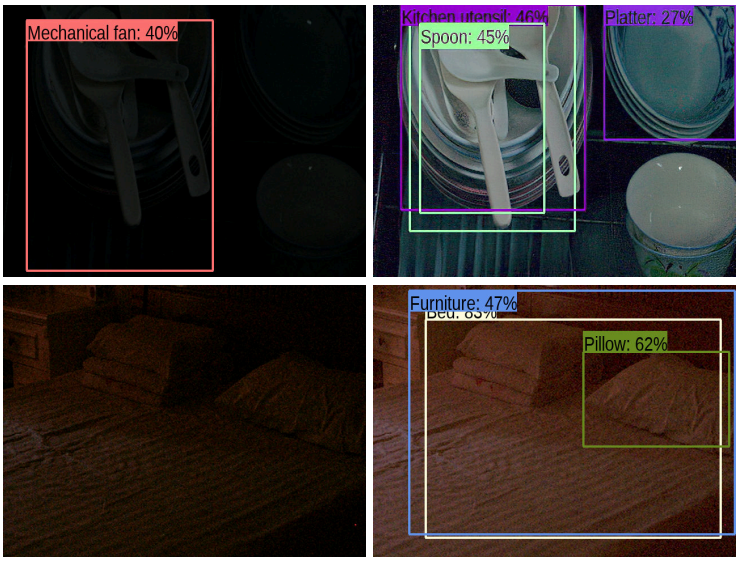


Fig. 3.24: The performance of Object Detection API on the images before (left) and after (right) enhancement by the proposed DSE-Net.

3.5 Summary

In this chapter, we presented two new approaches for low light image enhancement. The first approach is named as NPLIE which is based on estimation of illumination from a low light image. NPLIE gives natural contrast enhancement for low-light images. Further, we proposed a new adaptive illumination adjustment which helps in providing appropriate lightness to all regions in the images with varying illumination. The experimental analysis shows that the proposed algorithm improves dark regions of the image without over-enhancing properly illuminated areas of the same image. The exhaustive experimentation over a large number of images with different variations in the illumination conditions to check the performance of the proposed NPLIE is performed. The visual assessment shows that the proposed NPLIE maintains color constancy over the object under varying light conditions.

The second approach is named as DSE-Net which develops a deep network for simultaneous estimation of reflectance and illumination. DSE-Net uses a novel encoder-decoder architecture and decomposes images for low-light image enhancement and formulates a multi-objective loss function that considers the retinex theory. After decomposition of the image, the proposed method performs deep illumination adjustment for effective treatment of low-light regions. Further, we proposed an image refinement sub-module that performs the color and contrast adjustment with suppression of implicit noise in the image. The experimental results show that the proposed approach works adequately in restoring the details hidden in the dark areas. Furthermore, extensive experiments conducted on various challenging images show the superiority of the proposed method over the state-of-the-art approaches.

Chapter 4

Single Image Dehazing

This chapter presents a new single-image dehazing algorithm based on variational optimization. The chapter includes related work, estimation of transmission map, estimation of atmospheric light, experimental analysis, and conclusion.

4.1 Introduction

Computer vision tasks require images with appropriate visibility of the scene. However, sometimes the environment is not favorable to capture an image with good visibility, resulting in hazy images, rainy images, many other types of degraded images. In a real-world scenario, the atmosphere contains tiny particles that cause reflection, refraction, and light scattering due to bad weather conditions. Thus, outdoor images captured in lousy weather suffer from degraded visibility. While capturing an image in such conditions, the blending of atmospheric light and the transmission leads to a hazy image. Most vision-based tasks like object detection, fruit picking robot, and 3D reconstruction depend significantly upon the image quality. The hazy images contain faded colors, low contrast, and fewer details. Due to haze in an image, feature extraction becomes challenging that degrades the performance of the vision-based tasks. To improve the performance of vision-based tasks in a hazy environment, we must remove the effect of haze from the image. Image dehazing enhances the colors and contrast of an image by removing or reducing the effect of haze thus, improves the performance of the vision-based tasks. An inferior image dehazing method may cause undesirable artifacts and loss of crucial details. Therefore, developing a compelling image dehazing algorithm is a challenging task.

Most of the image dehazing algorithms suffer from one or more of the following limitations:

- limited generalization in case of learning-based dehazing methods.
- Over or under enhancement in dehazed image.
- Undesired artifacts and halo effects.
- Insignificant haze removal i.e. restricted image dehazing.
- Distortion of colors in dehazed images.
- Loss of finer details in the dehazed images.

In this chapter, we present a new single image dehazing algorithm using variational optimization-based transmission estimation to overcome these issues. The estimation of transmission map, atmospheric light, and scene radiance from an image without prior knowledge is ill-posed. The transmission map can be estimated from structural details (dominant edges of a scene, objects, and haze) of an image. To estimate the structure-aware transmission map, we leverage the notion of adaptive bilateral filtering [184]. Further, We formulate and solve a new variational optimization to estimate final transmission map. The objective function of variational optimization function helps in achieving the textual suppression and structural preservation. The significant contributions are summarized as follows:

- We propose a new method for the estimation of structure-aware initial transmission. We leverage the concept of adaptive bilteral filtering to achieve the strcuture-aware initial transmission.
- We formulated a new variational optimization problem with regularization terms to preserve the structural details in the final transmission while smoothing the textural details. We used the Alternative Direction Minimization (ADM) algorithm [164] to solve the formulated variational optimization.
- We performed an exhaustive analysis of the proposed approach with state-of-the-art algorithms on various datasets [20, 125, 185, 186, 187] using qualitative and visual analysis.

4.2 The Proposed Approach

This section presents the proposed variational optimization-based single image dehazing. Koschmieder *et al.* [26], and McCartney [27] presented the fundamental

scattering-based physical phenomenon which is used as the basis for dehazing. The formulation is given by:

$$I(x) = L_{\infty}\rho(x)\exp^{-\beta d(x)} + L_{\infty}(1 - \exp^{-\beta d(x)}) \tag{4.1}$$

where x is a pixel on a 2D coordinate plane, I is the observed image intensity, L_{∞} is the atmospheric light, d is the distance between object and lens, ρ is the reflectance of the object in the frame, and β is the atmospheric attenuation constant.

I is the result of the medium of transmission and light reflected from the object(s) in the frame, essentially representing the hazy image or the direct attenuation, and $t(x)$ represents the transmission $exp^{-\beta d(x)}$. Hence, reflectance times atmospheric light; $L_{\infty}\rho(x)$, is the actual haze-free image given by $J(x)$. Simplifying the equation and replacing atmospheric light with A, we obtain:

$$I(x) = J(x)t(x) + A(1 - t(x)) \tag{4.2}$$

From (4.1) and (4.2), we can infer that the estimation of two unknowns (i.e. A and $t(x)$) in the physical model is ill-posed. However, if depth information ($d(x)$) is known then we can estimate $J(x)$, A and $t(x)$ from $I(x)$. We have another image formation model which is known as Retinex model [10, 13, 188]. The retinex model considers an image as a product of reflectance and illumination. Retinex model is a special kind of physical model. If we consider Atmospheric light equal to zero then the second part of the physical model will become zero. We can write the physical model with A equal zero as:

$$I(x) = J(x)t(x) \tag{4.3}$$

It becomes similar to the retinex model. When the effect of atmospheric light is not present in an image then this model is used for image enhancement. In absence of atmospheric light, the scene radiance acts as reflectance and transmission map acts as illumination.

The atmospheric light can be considered to be constant through out a scene for all practical purposes. However, transmission varies with respect to space (medium). A hazy image captures the information about transmission map in terms of haze and object details present in it. Textural details (minor details of the scene, and texture of objects) of a scene does not affects the transmission map. However, structural details (dominant edges of a scene, objects, and haze) plays a major role in estimation of transmission map. The details of transmission map estimation is discussed in Section 4.2.1. Fig. 4.1 shows the framework of the proposed variational optimization-based

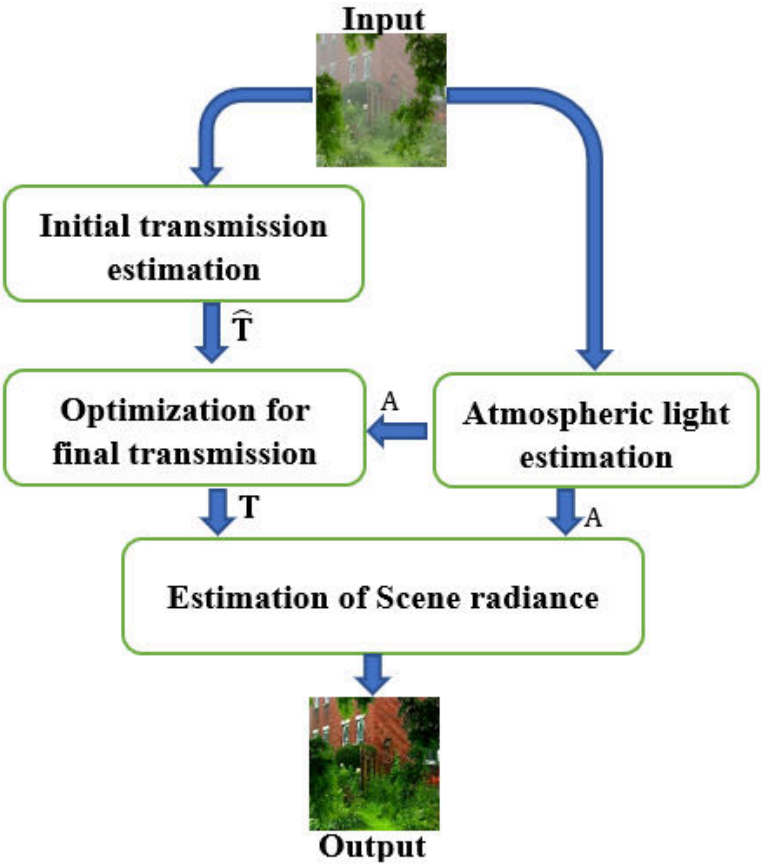


Fig. 4.1: Framework of Variational optimization based single image dehazing.

single image dehazing. Firstly, the proposed method estimates the structure-aware initial transmission. After estimating the initial transmission, we estimate the final transmission using the formulated variational optimization and initial transmission. We further estimate the uniform atmospheric light. Finally, the proposed method estimates the scene radiance for the dehazed output.

4.2.1 Estimation of Transmission map

In hazy images, the haze increases with increase in depth of the scene from the camera. The effect of haze is higher for the object at the higher depth and lesser to the closer object. The transmission map varies throughout the image (space). Thus, the transmission map capture the structural properties of the scene. The scene at larger depth will have different value of transmission than the scene which is at closer depth. Authors [20, 125] use our notion loosely by using foreground, background, and depth estimation to estimate the transmission map. However, these approaches have limited success. The information related to transmission map can be estimated in a better manner

by using the structural information of the haze. To achieve the structural information, we have to smooth the textural details without loosing the structural details. Thus, we propose a structure-aware transmission map estimation. We propose structure-aware transmission map, which uses structural-details, properties and position of a scene in the image. The proposed method uses variational optimization for the estimation of the final transmission map. The variational optimization-based method requires an initial transmission to estimate the final transmission map. It uses fidelity between the initial and final transmission map. Thus, the initial transmission has a defining effect on the estimation of the final transmission map. We estimate structure-aware initial transmission map that contains structural details with smooth texture. To achieve adequate textural smoothing with appropriate structural details, we use the Scale Adaptive Bilateral Filtering (SABF) [184]. SABF is a method of textural smoothing that uses bilateral filtering with adaptive scale. The proposed initial transmission can be expressed as:

$$\hat{\mathbf{T}} = \mathcal{S}(\max(\mathbf{I}^c)) \tag{4.4}$$

where $\hat{\mathbf{T}}$ is the initial transmission, $\max$ is the pixel-wise max operation, $\mathbf{I}$ is the input image, and $\mathcal{S}$ is the operator which performs Scale adaptive bilateral filtering.

The proposed approach leverages variational optimization to estimate the final transmission map from the initial transmission. Thus, we formulated the variational optimization function with initial transmission and regularization terms. The optimization function performs structure-preserving smoothing for transmission estimation. To achieve the structure-aware transmission map, we proposed to solve the following optimization problem:

$$\min_{\mathbf{T}} ||\hat{\mathbf{T}} - \mathbf{T}||_2^2 + \alpha||\nabla\mathbf{I}_m - \nabla\mathbf{T}||_2^2 + \beta||\mathbf{G}_T \circ \nabla\mathbf{T}||_1 \tag{4.5}$$

where $\mathbf{T}$ is the final transmission map, $\hat{\mathbf{T}}$ is initial transmission as obtained by (4.4), ∇ denotes first-order derivative operator, $\nabla\mathbf{I}_m$ denotes maximum of ∇ from all channels of $\mathbf{I}$ (i.e., input image), $\mathbf{G}_T$ is the weight based on the initial transmission, $||\bullet||_p$ is the L^p norm operator, and α & β is the parameter to stabilize the effect of both terms.

The first L^2 norm term (i.e. $||\hat{\mathbf{T}} - \mathbf{T}||_2^2$) of the optimization problem provides fidelity between initial and final transmission map. The second L^2 norm term (i.e., $||\nabla\mathbf{I}_m - \nabla\mathbf{T}||_2^2$) tries to maximize the structural details in the transmission. The L^1 norm term (i.e. $||\mathbf{G}_T \circ \nabla\mathbf{T}||_1$) endeavours to minimize the textural details and provide the structural-preserving smoothing. To achieve the structural-preserving smoothing

L^1 norm term uses weight map $\mathbf{G}_T$. The weight map $\mathbf{G}_T$ can be computed as-

$$\mathbf{G}_T = \sqrt{\mathbf{G}_T^2(x) + \mathbf{G}_T^2(y)} \tag{4.6}$$

where $\mathbf{G}_T(x)$ and $\mathbf{G}_T(y)$ are the weight map along the horizontal and vertical directions, which are defined as:

$$\mathbf{G}_T(x) = -log(\nabla_x \hat{\mathbf{T}}) \tag{4.7}$$

$$\mathbf{G}_T(y) = -log(\nabla_y \hat{\mathbf{T}}) \tag{4.8}$$

where ∇_x and ∇_y are the first-order derivatives along the horizontal and vertical directions. The weight map provides higher values for the textural details and lower values for structural details. Thus, the minimization of the optimization problem suppresses the textural details which helps in structure-preserving smoothing.

The Alternative Direction Minimization (ADM) method can provide an efficient solution to the optimization problem in (4.5). It can be seen in (4.5), both term (i.e. L^1 and L^2 norm) contains $\mathbf{T}$. Thus, an auxiliary variable $\mathbf{K}$ is introduced to separate both terms which makes it easy to solve. The introduction of $\mathbf{K}$ adds a constraint i.e. $\nabla\mathbf{T} = \mathbf{G}$. The equivalent optimization problem can be written as:

$$\min_{\mathbf{T}} ||\hat{\mathbf{T}} - \mathbf{T}||_2^2 + \alpha||\nabla\mathbf{I}_m - \mathbf{K}||_2^2 + \beta||\mathbf{G}_T \circ \mathbf{K}||_1 \qquad s.t. \nabla\mathbf{T} = \mathbf{K} \tag{4.9}$$

The lagrangian multiplier can provide a single optimization function by combining the equality constraint. The augmented lagrangian function can be represented as:

$$\mathcal{L} = ||\hat{\mathbf{T}} - \mathbf{T}||_2^2 + \alpha||\nabla\mathbf{I}_m - \mathbf{K}||_2^2 + \beta||\mathbf{G}_T \circ \mathbf{K}||_1 + \varphi(\mathbf{M}, \nabla\mathbf{T} - \mathbf{K}) \tag{4.10}$$

where $\varphi(\mathbf{M}, \nabla\mathbf{T} - \mathbf{K}) = \frac{\omega}{2}||\nabla\mathbf{T} - \mathbf{K}||_2^2 + \langle\mathbf{M}, \nabla\mathbf{T} - \mathbf{K}\rangle$ and $\langle\cdot,\cdot\rangle$ is pixel-wise multiplication, ω controls the convergence rate to the solution [164]. ADM can solve the optimization problem easily. The solution iteratively updates one term by keeping other as fixed at that time. The optimization problem is divided two into sub-problems, one for solution of $\mathbf{T}$, and other for solution of $\mathbf{K}$. The decomposition of (4.10) can be expressed as:

$$\mathbf{T}^{i+1} = \min_{T} ||\hat{\mathbf{T}} - \mathbf{T}||_2^2 + \varphi(\mathbf{M}^i, \nabla\mathbf{T} - \mathbf{K}^i) \tag{4.11}$$

$$\mathbf{K}^{i+1} = \min_{K} \alpha||\nabla\mathbf{I}_m - \mathbf{K}||_2^2 + \beta||\mathbf{G}_T \circ \mathbf{K}||_1 + \varphi(\mathbf{M}^i, \nabla\mathbf{T} - \mathbf{K}^i) \tag{4.12}$$

The sub-problem in (4.11) contains terms from (4.10) which contains transmission

map. Further, the sub-problem in (4.12) contains all term related to equality constraint in (4.10).

Gathering terms related to **T** for i^{th} iteration, It is a conventional least square problem. To solve the sub-problem we differentiate it w.r.t. to T and putting it equal to 0:

$$\mathbf{T} = \frac{2\hat{\mathbf{T}} + \mathbf{D}^T(\omega\mathbf{K} - \mathbf{M})}{\mathbf{2} + \omega\mathbf{D}^T\mathbf{D}} \tag{4.13}$$

where D is a matrix which contains Dx and Dy, Dx and Dy denotes the difference along horizontal and vertical direction. Inverses, transpose, and multiplication of large matrices may be computationally costly. Thus, we use the 2D-FFT method by considering the condition of the circular boundary, and compute T as:

$$\mathbf{T}^{i+1} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}\big(2\hat{\mathbf{T}}\big) + \sum_{d\in\{x,y\}} \mathcal{F}^c\big(\mathbf{D}_d(\omega^i\mathbf{K}^i - \mathbf{M}^i)\big)}{\mathbf{2} + \omega^i \sum_{d\in\{x,y\}} \mathcal{F}^c\big(\mathbf{D}_d\big) \cdot \mathcal{F}\big(\mathbf{D}_d\big)}\right) \tag{4.14}$$

where $\mathcal{F}$ denotes 2D-FFT operation, $\mathcal{F}^c$ denotes complex conjugate of 2D-FFT operation, $\mathcal{F}^{-1}$ denotes the inverse of 2D-FFT operation and **2** is a matrix equivalent to image size having all elements 2. Here, all operations are element-wise.

Considering only K- related terms from (4.10) to obtain (4.12) i.e. sub-problem related to **K** for i^{th} iteration. To solve the (4.12), we take derivative with respect to **K** and equate it to zero. Shrink operation maintains threshold of an image based on the base of shrink. Thus, we solved the K sub-problem using shrink operation:

$$\mathbf{K}^{i+1} = \mathcal{S}_{\frac{\beta\cdot\mathbf{G}_T}{2\alpha+\omega^i}}\left[\frac{2\alpha\nabla\mathbf{I}_m + \omega\nabla\mathbf{T}^{i+1} + \mathbf{M}^i}{2\alpha + \omega^i}\right] \tag{4.15}$$

where $\mathcal{S}_\phi[x] = sign(x)max(|x| - \phi, 0)$ and, all operations are element-wise. The values of parameters α and β are chosen empirically. The following computation provides the Lagrangian **M** and penalty term ω:

$$\begin{aligned} &\mathbf{L}^{i+1} = \mathbf{L}^i + \omega^i(\nabla\mathbf{I}^{i+1} - \mathbf{K}^{i+1}); \\ &\omega^{i+1} = \omega^i\delta, \delta > 1. \end{aligned} \tag{4.16}$$

To provide an optimal solution, the above iterative process requires an stopping criteria. Empirically it is noted that the proposed algorithm provides adequate estimation of transmission map after eight iterations. The other stopping criteria of threshold difference many times runs for a larger time without any improvement in the transmission. Thus stopping after eight iteration helps in achieving adequate transmission map and avoids the longer running time.

After estimation of the transmission map, the proposed algorithm estimates the final transmission as:

$$\mathbf{T}_1 = A - \mathbf{T} \tag{4.17}$$

The final transmission ($\mathbf{T}_1$) uses the variational optimization and the duality [189] to facilitate the estimation of transmission.

4.2.2 Estimation of Atmospheric light

He *et al.* [20] proposed DCP, a method for an atmospheric light estimation that helps in achieving dehazed results. The atmospheric light in DCP is distinct for all color channels, which sometimes causes artifacts in the enhanced image. Moreover, it uses a neighborhood approach in transmission estimation, which increases the time required to estimate atmospheric light. The atmospheric light is independent of the color and properties of the object. Thus, atmospheric light should be the same across all color channels. Therefore, we propose Dark-channel based uniform atmospheric light, which is independent of the color channel of the image. To estimate the proposed atmospheric light, we consider the maximum operation of the pixel-wise minimum of a color image. The proposed approach estimates the atmospheric light in fewer efforts. It is also constant across all color channels, which helps estimation of the transmission and the final dehazed image. The proposed atmospheric light can be represented as:

$$A = max\big(\min_{c \in R,G,B} \mathbf{I}^c, \epsilon\big) \tag{4.18}$$

where **I** is the input image, c denotes the color channel, and A denotes the proposed atmospheric light. The proposed estimation of atmospheric light requires less time as it does not use a kernel to estimate atmospheric light.

Finally, we estimate the dehazed image based on the estimated atmospheric light and final transmission map. The dehazed image can be expressed as:

$$\mathbf{J}^c = \frac{\mathbf{I}^c - (1 - \mathbf{T}_1)A}{max(\mathbf{T}_1, \epsilon)} \tag{4.19}$$

where ϵ is the threshold of the image, and c is the color channel of RGB color space.

4.3 Experimental results and Assessment

We analyzed the proposed algorithm on images from several databases: DCP [20], CAP [125], RESIDE database [185], O-Haze database [186], and HazeRD [187]. The quantitative and visual assessment is used to analyse the performance of proposed method with state-of-the-art algorithms: DCP [20], CAP [125], Dehazenet [133], AOD [134], FFA [139], GCA [135], and DHS [145]. The codes and parameters of the other algorithms are taken from the authors' provided source to ensure a fair comparison. Further, the discussion regarding parameter analysis ensures the superior performance of the proposed approach with the best values of the parameters.

4.3.1 Quantitative Analysis

Quantitative analysis is always desirable to validate any experimental result. However, it is a challenging task to develop a quantitative measure in the case of image dehazing. There have been several attempts to design such a measure. It may be easily verified that a quantitative measure may give better value for a poorly dehazed image and a low value for a high-quality enhanced image. Unfortunately, there is no single universally accepted quantitative assessment method available. We have used three quantitative methods (frequently used in literature): PSNR, SSIM [190], and VIF[171].

Peak Signal-to-Noise Ratio (PSNR) measures the ratio between the highest intensity of the image and corrupting noise. PSNR considers the strength of information with respect to the noise. Higher PSNR denotes adequate signal quality with respect to the noise. Thus, higher PSNR shows better enhancement. PSNR analyses the signal to noise ratio. It does not provide any additional assessment related to image content and dehazing. SSIM (Structural Similarity Index) [190] estimates similarity between two images based on the structural features. It measures similarity between two images based on mean, standard deviation, and variance. In addition, it analyses the structural degradation in the enhanced image with respect to the reference image. A higher SSIM value denotes better enhancement as it shows lesser degradation in the enhanced image. VIF (Visual Information Fidelity) [171] evaluates the naturalness of the enhanced image based on the image fidelity between reference and enhanced image. To evaluate the naturalness, it uses the naturalness scene statistics model. A higher VIF value denotes the lesser distortion and less amount of information loss. Thus, a higher VIF value shows better enhancement. We perform exhaustive experimentation on a large set of images. Table 4.1, Table 4.2, and Table 4.3 show the minimum, maximum, median, mean, and standard deviation values of SSIM, PSNR, and VIF, respectively.

Table 4.1: Quantitative assessment for various algorithms using statistical values of PSNR.

Metrices	DCP [20]	CAP [125]	DehazeNet [133]	AOD [134]	FFA [139]	GCA [135]	DHS [145]	**Ours**
Min	9.69	8.54	7.99	9.55	**12.71**	7.45	8.83	10.16
Max	20.53	20.17	20.64	20.04	22.14	19.24	22.88	**23.74**
Median	16.97	14.48	14.22	15.55	16.21	16.52	14.24	**17.10**
Mean	16.88	14.32	16.73	14.92	17.50	17.13	16.02	**17.60**
Std. Dev.	3.23	2.90	4.16	2.99	3.48	2.97	4.34	**2.84**

Table 4.2: Quantitative assessment for various algorithms using statistical values of SSIM.

Metrices	DCP [20]	CAP [125]	DehazeNet [133]	AOD [134]	FFA [139]	GCA [135]	DHS [145]	**Ours**
Min	0.53	0.45	0.35	0.40	0.47	0.37	**0.57**	0.55
Max	**0.94**	0.92	0.89	0.92	0.88	0.93	0.91	**0.94**
Median	0.82	0.80	0.75	0.76	0.76	0.85	0.78	**0.88**
Mean	0.81	0.83	0.77	0.77	0.69	0.82	0.83	**0.87**
Std. Dev.	0.12	0.11	0.13	0.12	0.15	0.11	**0.10**	**0.10**

Table 4.3: Quantitative assessment for various algorithms using statistical values of VIF.

Metrices	DCP [20]	CAP [125]	DehazeNet [133]	AOD [134]	FFA [139]	GCA [135]	DHS [145]	**Ours**
Min	0.18	0.16	0.21	0.22	0.27	0.23	0.28	**0.29**
Max	0.98	0.98	0.80	0.89	0.79	0.66	0.76	**1.03**
Median	0.49	0.46	0.43	0.51	0.56	0.44	0.42	**0.61**
Mean	0.51	0.44	0.47	0.48	0.54	0.49	0.44	**0.59**
Std. Dev.	0.19	0.18	0.17	0.19	0.17	**0.15**	0.17	0.16

Note best values in the table are bold, and the second-best values are underlined. The minimum statistical measure shows the worst-case performance of the method. The maximum statistical measure shows the best-case performance of the various method. The median, mean, and standard deviation show the statistics for average-case performance.

The statistical measure for PSNR values of various algorithms is shown in Table 4.1. It can be noted from Table 4.1 that the proposed method achieves the best mean PSNR value. The standard deviation shows that the deviation in PSNR values of the proposed method is less in comparison to other state-of-the-art methods. The minimum statistical measure shows that the proposed method achieves the second-best value in the worst case. Table 4.2 shows statistical measures for SSIM values of various algorithms. It can be observed from Table 4.2 that the proposed algorithm achieves the best mean value. The standard deviation of SSIM for the proposed method shows that most values remain near the mean value. The proposed method achieves the second-best SSIM value in the case of minimum statistical measures. Table 4.3 shows statistical measures for VIF values of various algorithms. It can be observed from Table 4.3 that the proposed algorithm achieves the best mean value. The proposed algorithm is second-best in the case of standard deviation. However, the mean value of the proposed method is better than GCA. The standard deviation of GCA with a lower mean VIF value means that most of the VIF values will remain near the mean statistics. The quantitative assessment shows that the proposed method is achieving superior values for most of the quantitative measures.

4.3.2 Visual Assessment

Due to the unavailability of universally accepted quantitative measures, visual assessment became mandatory for image assessment. The visual assessment for the proposed algorithm with other contemporary algorithms is performed extensively. Few sample results are shown in Fig. 4.2 to Fig. 4.6. For every test result image from Fig. 4.2 to Fig. 4.6, the input image is shown in (a), and the result of the proposed algorithm is shown in (q).

The input image 'Canon' as shown in Fig. 4.2(a), depicts a hazy image, and the zoomed-in section of the marked patch is shown in Fig. 4.2(b). The result of DCP is shown in Fig. 4.2(c) along with the zoomed-in image of the highlighted region. DCP provides adequate dehazing. However, it can be noticed from Fig. 4.2(d) that the image is slightly smoother and contains fewer details, which is undesirable. Fig. 4.2(e) shows the results of CAP. It can be observed from Fig. 4.2(f) that CAP is unable to

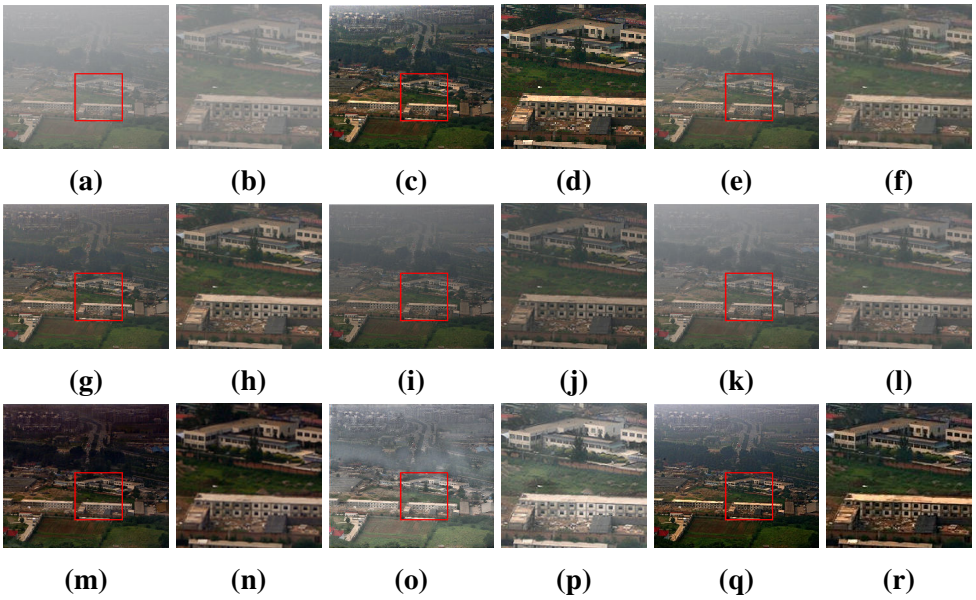


Fig. 4.2: Canon. (a) Input image; (b) Patch of input image; (c) DCP; (d) Patch of DCP; (e) CAP; (f) Patch of CAP; (g) DehazeNet; (h) Patch of DehazeNet; (i) AOD; (j) Patch of AOD; (k) FFA; (l) Patch of FFA; (m) GCA; (n) Patch of GCA; (o) DHS; (p) Patch of DHS; (q) Proposed method; (r) Patch of Proposed method.

remove haze properly and smoothed out the image, which results in the loss of details. The result of DehazeNet as shown in Fig. 4.2(g). Fig. 4.2(h) shows that the details are properly visible, and haze removal is also not adequate. AOD fails to eradicate the effect of haze from the image as shown in Fig. 4.2(i). The result of FFA is shown in Fig. 4.2(k). It can be observed from Fig. 4.2(l) that the haze removal using FFA is inadequate. The result of GCA is shown in Fig. 4.2(m). GCA achieves image dehazing but at the cost of image lightness which causes detail loss as shown in Fig. 4.2(n). The results of DHS is shown in Fig. 4.2(o). The artifacts are evident in the result of DHS as shown in Fig. 4.2(p). The result of the proposed algorithm is shown in Fig. 4.2(q). The proposed algorithm achieves superior image dehazing with proper contrast as shown in Fig. 4.2(r). The proposed algorithm achieves better color enhancement than the other algorithms.

The 'House' image and results of various algorithms are shown in Fig. 4.3. The result of DCP is shown in Fig. 4.3(c). It can be observed from Fig. 4.3(d) that DCP generates few undesired artifacts near the edges. Further, the details in the images are slightly smoother. Fig. 4.3(e) shows the results of CAP, which provides a limited enhancement in the image. CAP provides the blurred and dull details, which can be noticed from Fig. 4.3(f). It is evident from Fig. 4.3(d) that DehazeNet provides Over-enhanced results, which degraded the image quality. The result of AOD is shown in

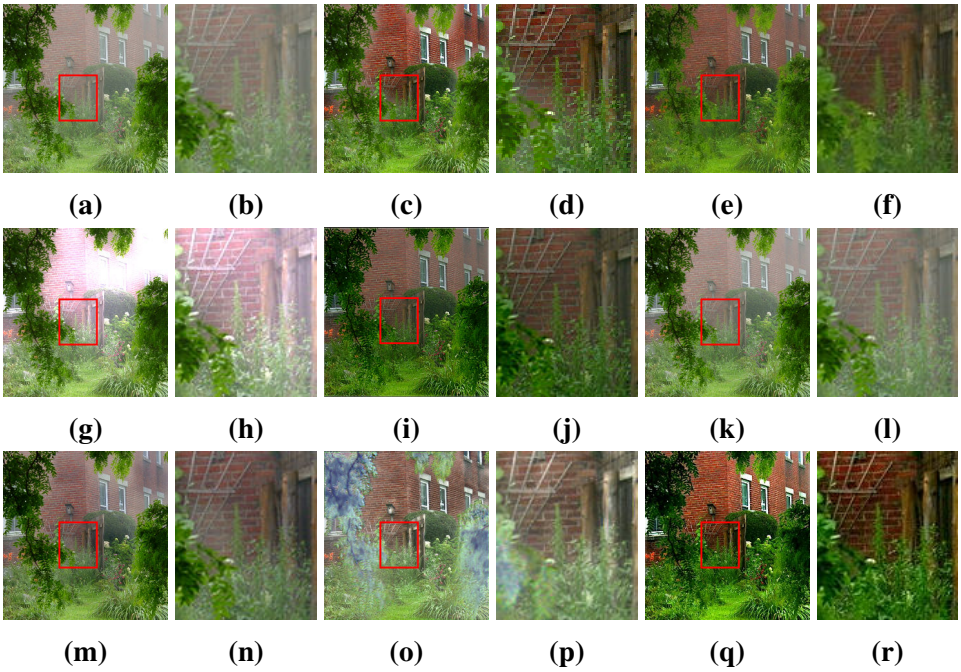


Fig. 4.3: House. (a) Input image; (b) Patch of input image; (c) DCP; (d) Patch of DCP; (e) CAP; (f) Patch of CAP; (g) DehazeNet; (h) Patch of DehazeNet; (i) AOD; (j) Patch of AOD; (k) FFA; (l) Patch of FFA; (m) GCA; (n) Patch of GCA; (o) DHS; (p) Patch of DHS; (q) Proposed method; (r) Patch of Proposed method.

Fig. 4.3(e). The haze removal in AOD is acceptable, as shown in Fig. 4.3(f). However, the colors in the enhanced image of AOD are faded. Haze can be observed from the results of FFA and GCA as shown in Fig. 4.3(k) and 4.3(m), respectively. The result of DHS is shown in Fig. 4.3(h); it generates various artifacts. Further, DHS distorts the color of the input image, which is undesirable. Fig.4.3(q) shows the result of the proposed algorithm. It can be noted from Fig. 4.3(r) that the proposed algorithm achieves better colors with adequate image dehazing. The superior details in the results of the proposed algorithm can be observed from Fig. 4.3(q).

The input image 'Flags' and its results from various contemporary algorithms is shown in Fig. 4.4. The result of DCP is shown in Fig. 4.4(c), which shows adequate image dehazing. However, the details in the image are not clearly visible due to the smoothing effect. Fig. 4.4(e) shows the results of CAP, which provides limited haze removal as thin haze is visible in the results. It can be observed from Fig. 4.4(f) that the details in the result of CAP are inadequate. In Fig. 4.4(g), the DehazeNet algorithm provides over-enhancement in the results due to over-exposure. AOD and FFA fail to remove the haze from the image as haze are evident in the results as shown in Fig. 4.4(e) and Fig. 4.4(f) respectively. The result of GCA is shown in Fig. 4.4(m), which shows that GCA fails in removing overall haze as haze is available in many regions.

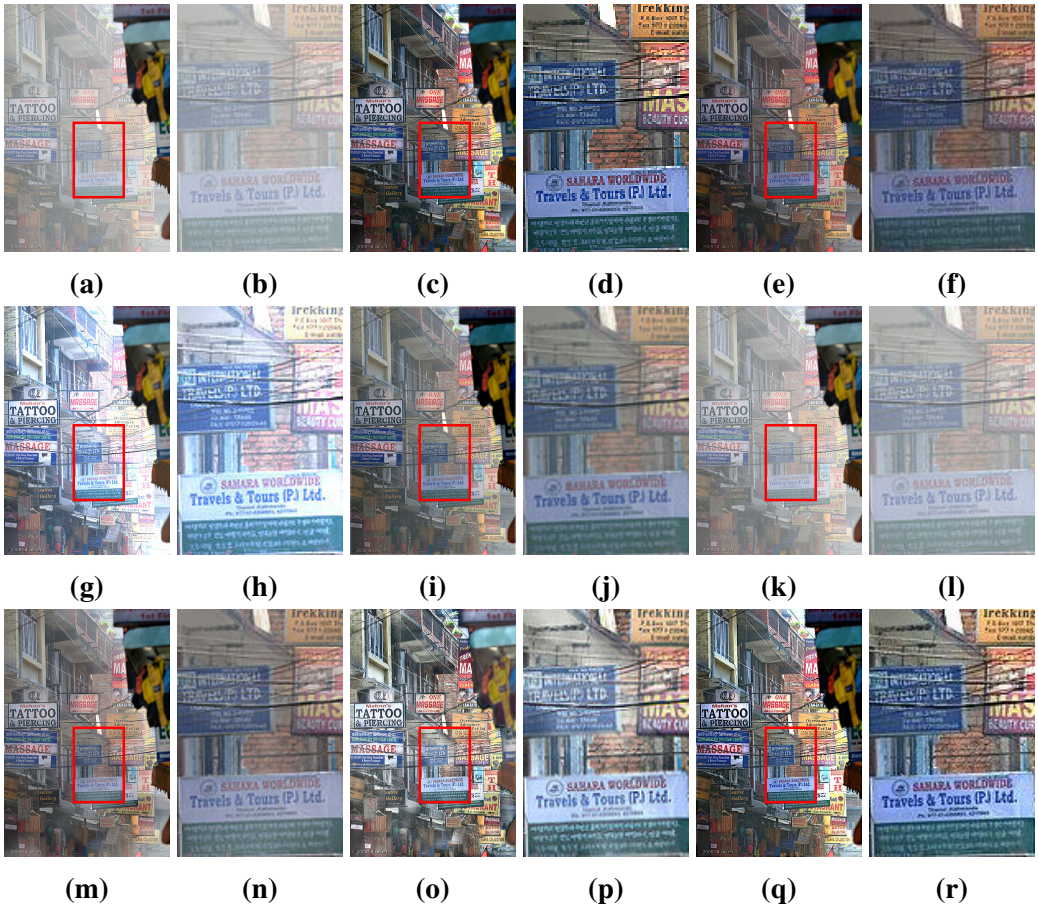


Fig. 4.4: Flags. (a) Input image; (b) Patch of input image; (c) DCP; (d) Patch of DCP; (e) CAP; (f) Patch of CAP; (g) DehazeNet; (h) Patch of DehazeNet; (i) AOD; (j) Patch of AOD; (k) FFA; (l) Patch of FFA; (m) GCA; (n) Patch of GCA; (o) DHS; (p) Patch of DHS; (q) Proposed method; (r) Patch of Proposed method.

In Fig. 4.4(o), DHS provides haze removal; however, artifacts and color distortion are evident. Fig. 4.4(i) shows the results of the proposed algorithm. The proposed algorithm provides better contrast and lightness. Further, the structural details of the output are better than the rest of the algorithms.

In Fig. 4.5, the input 'Forest' image and result from other algorithms is presented. The presence of artifacts is evident in the result of DCP as shown in Fig. 4.5(d). The results of CAP is shown in Fig. 4.5(e). It shows that the detail enhancement is limited. The result of DehazeNet shows over enhancement which causes loss of details as shown in Fig. 4.5(g). The result of AOD is shown in Fig. 4.5(i). In AOD, details and color enhancement is inadequate. FFA, GCA, and DHS is unable to perform image dehazing as shown in Fig. 4.5(k), 4.5(m) and 4.5(o) respectively. The haze is visible in the results of FFA, GCA, and DHS. The result of the proposed algorithm is shown in Fig. 4.5(q). The proposed algorithm performed superior image dehazing than other

Fig. 4.5: Forest. (a) Input image; (b) Patch of input image; (c) DCP; (d) Patch of DCP; (e) CAP; (f) Patch of CAP; (g) DehazeNet; (h) Patch of DehazeNet; (i) AOD; (j) Patch of AOD; (k) FFA; (l) Patch of FFA; (m) GCA; (n) Patch of GCA; (o) DHS; (p) Patch of DHS; (q) Proposed method; (r) Patch of Proposed method.

contemporary algorithms. Further, the proposed algorithm provides better colors in the enhanced image along with better details in the image.

The results of various contemporary algorithms for the 'Sea' input image are shown in Fig. 4.6. Fig. 4.6(c) shows the results of DCP. It can be observed from Fig. 4.6(c) and 4.6(d) that it provides a slight smooth output which results in fewer details. The result of CAP is shown in Fig. 4.6(e). The details in the result of CAP are not clearly visible, as shown in Fig. 4.6(f). DehazeNet algorithm provides high exposure in few regions of the image as shown in Fig. 4.6(g). Further, It fails to preserve the color properties of the image. The result of AOD is shown in Fig. 4.6(i). It can be noted from Fig. 4.6(i) that the lightness in the enhanced image is low that results in loss of crucial details. FFA fails to provide image dehazing as shown in Fig. 4.6(k). The result of GCA shows fewer details due to low lightness, as shown in Fig. 4.6(m). It can be noted from Fig. 4.6(o) that DHS fails to remove haze effectively. Fig. 4.6(q) shows the results of the proposed algorithm. The proposed algorithm achieves image dehazing with adequate contrast and color enhancement. The details in the result of the proposed algorithm are better than the result of other algorithms.

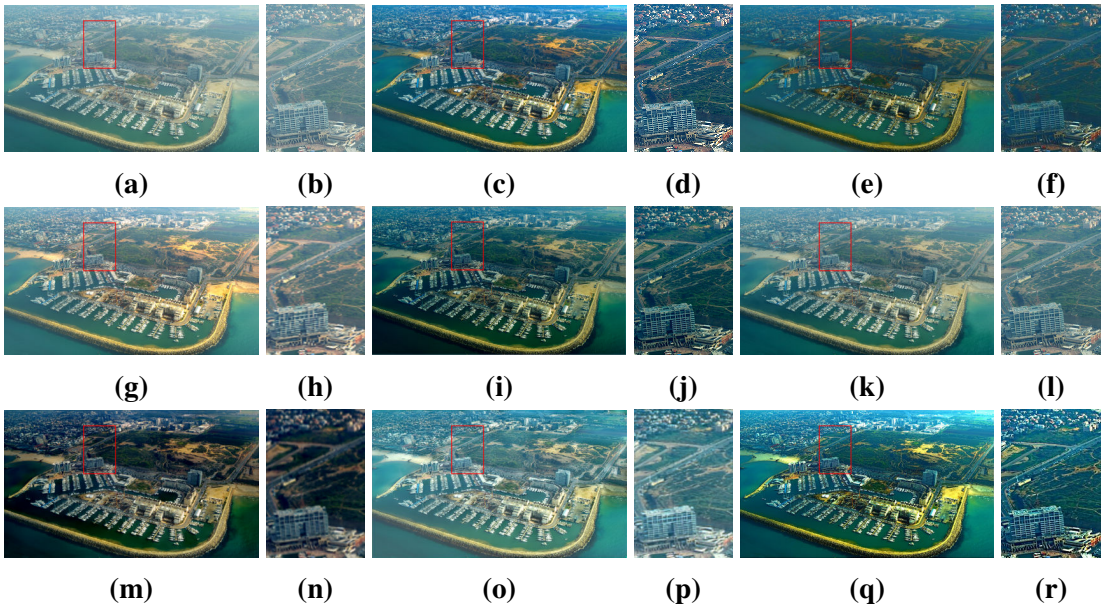


Fig. 4.6: Sea. (a) Input image; (b) Patch of input image; (c) DCP; (d) Patch of DCP; (e) CAP; (f) Patch of CAP; (g) DehazeNet; (h) Patch of DehazeNet; (i) AOD; (j) Patch of AOD; (k) FFA; (l) Patch of FFA; (m) GCA; (n) Patch of GCA; (o) DHS; (p) Patch of DHS; (q) Proposed method; (r) Patch of Proposed method.

4.3.3 Limitations

The discussion in the above section shows that the proposed method works better than other contemporary methods. The highlighted and zoomed-in areas of Fig. 4.5 shows that all methods fail to provide significant dehazing for smaller objects at larger depth (i.e. leaves in the background) in the image. The result of DCP and DHS is shown in Fig. 4.5(d) and 4.5(p), respectively. The highlighted area of DCP and DHS shows the generated artefacts near the edges which degrade the visibility of smaller objects at larger depth. The result of CAP, AOD, FFA, and GCA shown in Fig. 4.5(f), 4.5(j), 4.5(l), and 4.5(n), respectively. These algorithms fail to provide dehazing for the objects at larger depths. Dehazenet provides over-exposure in the final image which results in degraded visibility of the background as shown in Fig. 4.5(h). It can be observed from Fig. 4.5(r) that for smaller objects at larger depths, our method provides limited dehazing. The details are better than other contemporary algorithms, still it need some improvement for smaller objects at larger distance. The proposed method considers atmospheeric light to be constant across throughout the image. However, in night-time hazy images the atmospheric light is not same with-in the scene. Thus, the proposed algorithm may fail to deal with night-time hazy images with non-uniform atmospheric light.

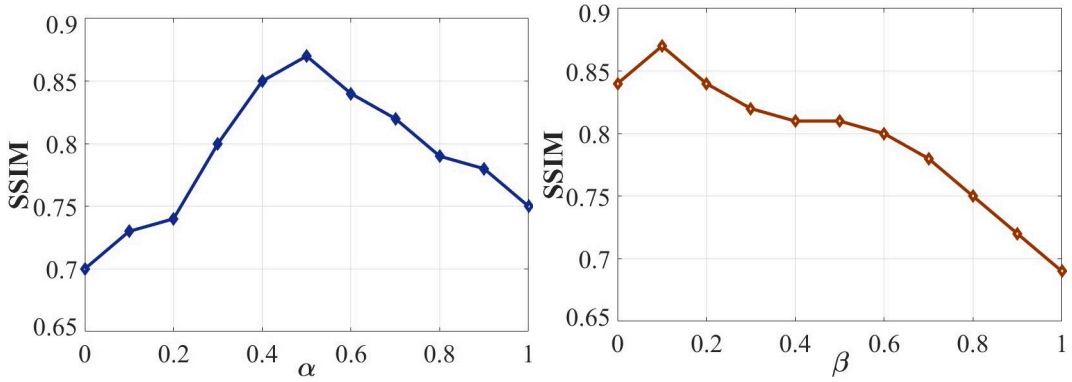


Fig. 4.7: Mean SSIM for different values of α and β.

4.3.4 Parameter Analysis

The proposed dehazing method uses α and β regularization parameters. We performed exhaustive experimentation to analyse the effect of regularization parameters. It is found that regularization parameters should vary between 0 and 1. Since positive values of regularization parameters contribute effectively in the variational optimization. The parameter α and β regulates the preservation of structural details and smoothing of textural details, respectively. Thus, if any of the regularization parameters have a value greater than one, then the respective term will dominate. If the value of α is increased beyond one, then structural details will dominate. In addition to the structural details, it may enhance the textural details due to the use of gradient in the respective term. On the other hand, if the value of β is greater than one, then it may result in the smoothing of structural details along with textural smoothing. We consider the value of regularization parameters between zero and one to avoid the dominance of the respective regularization term and limit the adverse effect. We selected SSIM to analyse the regularization parameter as it is one of the most widely used quantitative assessment measures. Fig. 4.7 shows the parameter analysis for various values by keeping the best value of other parameters. It may be observed from Fig. 4.7 that the proposed method works best for $\alpha = 0.5$ and $\beta = 0.1$. The performance of the proposed algorithm degrades for other values of α and β as they affect the structural details and textural smoothing. In our experimentation, the empirical values of parameters generate satisfactory results.

4.3.5 Computational Time

The run time of the various algorithm is analysed on a large set of images. The average run time per image is shown in Table 4.4. We analyze the run time of the algorithms

on a system with 8GB RAM, Tesla k80 GPU, and Core i5 processor @ 3.40GHz. We resized all images to 512x512 for a fair comparison. Further, we used codes and parameters from the authors' provided source. The algorithms which run on CPU are denoted with (C) with average running time in Table 4.4. On the other hand, (G) with the average run time in Table 4.4 shows that the algorithm uses GPU for computation. It can be observed from Table 4.4 that the proposed algorithm outperforms all algorithms except GCA [135]. GCA uses GPU for execution which provides faster computations to the algorithms. The above analysis shows that the proposed algorithm faster than most of the other contemporary algorithms.

Table 4.4: AVERAGE RUN TIME PER IMAGE (IN SECONDS)

DCP	CAP	DehazeNet	AOD	FFA	GCA	Ours
1.96 (C)	2.26 (C)	2.63 (C)	2.63 (C)	0.36 (G)	0.19 (G)	0.21 (C)

4.4 Summary

In this chapter, we proposed a single image dehazing algorithm based on the proposed variational optimization. The method uses the notion of same transmission for object at the same depth of any structure. The method develops a variational optimization for estimation of structure-aware transmission map using the initial transmission. The property of duality and atmospheric light is used to estimate the final transmission. The uniform atmospheric light is developed to provide better dehazing. The experimental assessment shows that the proposed algorithm improves the color and structure of the hazy images. Further, it restricts the over-enhancement of the regions with lesser haze. The exhaustive experimental assessment over a large number of images from several datasets shows that the algorithm yields dehazed images with natural visual quality. The visual analysis shows that the proposed method achieves better haze removal than the other algorithms.

Chapter 5

Object Recognition in Low-light Environment

This chapter presents a new multi-exposure refinement network for low-light object detection (MRN-LOD). The chapter includes the dataset for object detection in the low-light environment, multi-exposure feature extractor, Adaptive refinement network, experimental results and analysis, and the conclusion.

5.1 Introduction

Recent object recognition methods use a common architectural framework that includes the backbone, neck, and head as the components. The backbone focuses on feature extraction from the input image using pre-trained convolutional neural networks (CNN). The neck uses a feature pyramid network to process the extracted features to facilitate the recognition[1] of objects of different sizes. The processed features are used by the detection head to detect object classes and anchor boxes. Many variations of these components are developed to improve the performance and efficiency of object recognition methods. It is difficult to extract features from low-light images using existing backbones, which affects the detection performance. This chapter presents a new Multi-exposure refinement network for low-light object recognition to achieve better performance. MRN-LOD contains a multi-exposure feature extraction network (MFE) and an adaptive refinement network (ARN). We propose the notion of feature extraction from low-light images and multi-exposure images to facilitate low-light feature extraction. An adaptive refinement network is developed to refine the extracted features of low-light images using multi-exposed images. In MRN-LOD,

[1]Note: In this chapter, recognition and detection both means same i.e. object recognition.

feature extraction uses our MFE and refinement is performed using the ARN. The refined features are used by the detection head to recognize and localise the objects. The key contributions of the proposed MRN-LOD are as follows:

1. The notion of feature extraction from multi-exposure images to improve the performance of low-light object recognition. The multi-exposure images are generated using a single low-light image with distinct exposure parameters.

2. The development of an adaptive refinement network to refine the features of input low-light images. The ARN provides better features from low-light images to achieve improved detection performance in low-light conditions.

3. A new multi-exposure refinement network for low-light object detection is proposed, which extracts features directly from the low-light image without explicit use of enhancement and adaptively refines the extracted feature.

4. Extensive experimentation demonstrates the superior performance of the proposed method compared to other contemporary methods for low-light object recognition.

5.2 The Proposed Method

The proposed method focuses on object detection from low-light images, employing the combination of our multi-exposure feature extraction and adaptive refinement network. Unlike existing methods that predominantly prioritize image enhancement prior to detection, where detection performance heavily relies on the effectiveness of the enhancement algorithm, our method takes a different approach. It eliminates the need for image enhancement methods and significantly reduces both the time and computational resources required. To achieve this, we introduce a multi-exposure feature extraction technique that directly extracts features from low-light images, eliminating the need for separate enhancement methods. Additionally, we introduce an adaptive refinement network designed to process and refine the features extracted from low-light images using multi-exposed images. Finally, MRN-LOD employs a detection head to precisely localize objects within the images and assign them to their respective classes; the architecture of MRN-LOD is given in Fig. 5.1.

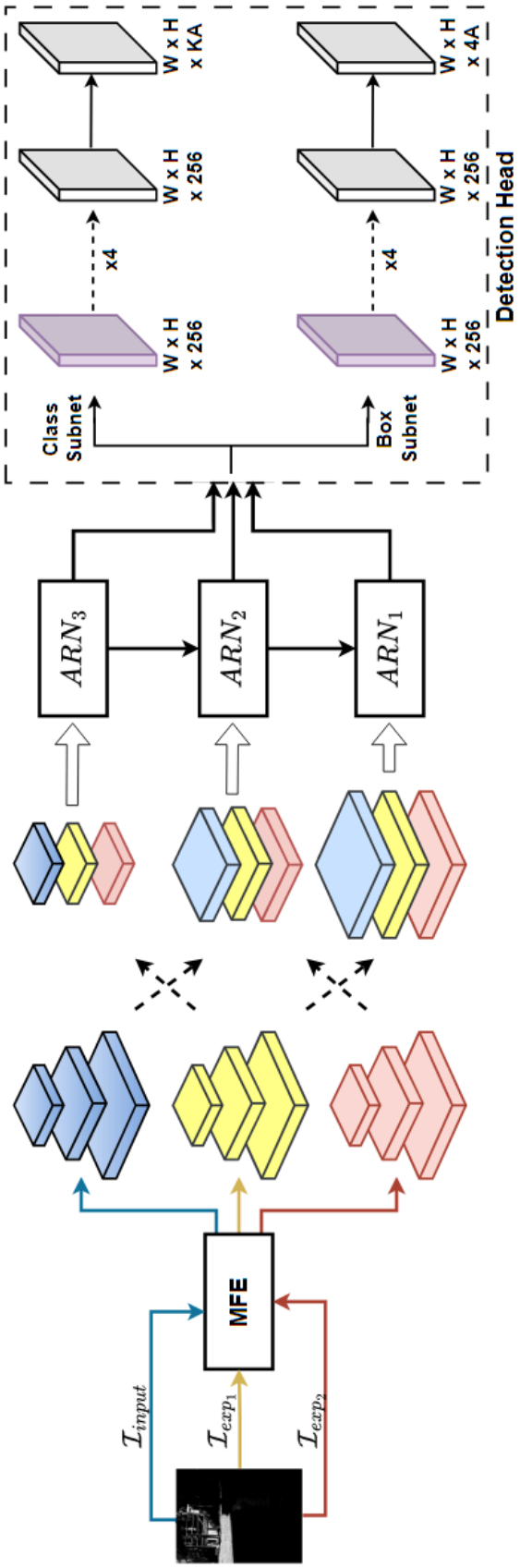


Fig. 5.1: MRN-LOD: Multi-exposure refinement network for low-light object detection. It uses the proposed MFE for extracting feature pyramids. Then, our method uses the proposed ARN for the refinement of the extracted features at all pyramidal levels. MRN-LOD uses the detection head for the prediction of bounding boxes with respective classes.

5.2.1 Multi-exposure feature extraction

Low-light images often exhibit a mixture of well-illuminated regions, thanks to local light sources, and areas plagued by poor visibility due to insufficient light. This variation in lighting conditions within a single image poses a formidable challenge for feature extraction. The available object detection methods focus on a single image for feature extraction and detection. These methods fail to detect objects in low-light images as feature extraction from a single low-light image is very difficult. The existing and straightforward approach is to enhance the low-light image and then perform feature extraction. However, the enhancement step requires additional computational and time. Moreover, the quality of the extracted features is contingent upon the effectiveness of the enhancement algorithm. The failure of the enhancement algorithm may lead to a degraded detection performance.

Another approach may be scaling exposure towards the brighter side. However, determining a universal exposure factor applicable to all types of image and region (i.e., local variation of light in an image), each with its unique lighting variations, is very difficult, if not impossible. Further, the exposure-scale version of the image suffers from the enhancement of the noise, saturation, and other artifacts. It makes feature extraction and object detection even worse. Thus, the direct use of a single-exposure image is inadequate for low-light object detection to achieve satisfactory performance. A new feature extraction paradigm for object detection in low-light images is required. We propose a new feature extractor, namely, the MFE network for low-light images. The MFE network generates multiple exposure instances of the input image to facilitate feature extraction. We introduce two exposures with the input image in MFE namely, an input image, baseline-exposure, and extended-exposure. The MFE network extracts features from the low-light image, it may effectively extract features from regions with normal lightness. The images with variations in exposure facilitate feature extraction from other regions of low-light images. The extended-exposure focuses on features of darker regions. The baseline-exposure helps in feature extraction from regions which are minorly affected with lightness issues. The MFE extracts feature pyramids from the above-mentioned multi-exposed inputs. We denote the feature pyramid of the input image as $\mathcal{P}_1$, baseline-exposure as $\mathcal{P}_2$, and extended-exposure as $\mathcal{P}_3$. The feature pyramid of distinct inputs ($\mathcal{P}_j$) can be represented as:

$$\mathcal{P}_j = [\mathcal{F}_j^3, \mathcal{F}_j^2, \mathcal{F}_j^1] \tag{5.1}$$

where j denotes the j^{th} input, and $\mathcal{F}_j^l$ denotes the feature map of l^{th} level of $\mathcal{P}_j$. The

extraction of the feature pyramid uses the ResNet architecture [136] in our method for all input with distinct exposures.

The MFE network is designed to capture features from various regions within low-light images by employing discrete exposure scales. Nevertheless, the challenge of inadvertent feature inclusion persists, attributable to factors such as noise amplification, saturation, and the emergence of unwanted artifacts. To mitigate these issues, we introduced an adaptive refinement process to obtain the appropriate features for object detection. This refinement procedure is strategically engineered to emphasize the extraction of pertinent features conducive to enhancing the effectiveness of low-light object detection. The adaptive refinement network is discussed in the next section.

5.2.2 Adaptive refinement network

We proposed an adaptive refinement network to refine the features of low-light images using multi-exposed images. The detailed architecture is given in Fig. 5.2. Separate processing of the features from multi-exposed images may process undesired features along with desired features. We propose to process the features in a combined manner to focus more on the adequate features from distinct exposures while minimizing the impact of inappropriate features. The refinement of features can be performed by simply adding the feature maps. However, the direct addition of feature maps assigns equivalent importance to all constituent feature maps. A weighted additive approach may provide appropriate importance to the respective feature maps. In this approach, the scalar weight parameter plays a pivotal role in regulating importance across the entirety of feature maps, thereby enabling adjustments that either augment or diminish the significance of these features. However, in the context of this refinement approach, it is noteworthy that undesirable features are given the same degree of attention as their more desirable counterparts, a configuration that is less than ideal.

To focus on the appropriate features, we propose using kernel-based adaptive refinement. Here, the weights of the kernel facilitate the attention of the appropriate features. We propose a convolution-based adaptive refinement network to focus on the appropriate features within and across distinct exposure inputs. The convolution-based refinement helps the different ARNs of MRN-LOD in focusing on the distinct region of the image. The learning phase of the convolution helps the ARN to focus on the appropriate region in the low-light image. The general expression for the proposed adaptive refinement is given as:

$$\mathcal{F}^l = \mathbf{W}^l * (\mathcal{F}^l_{\mathcal{S}_1} \odot \mathcal{F}^l_{\mathcal{S}_2} \odot \mathcal{F}^l_{\mathcal{S}_3} \odot \cdots) \tag{5.2}$$

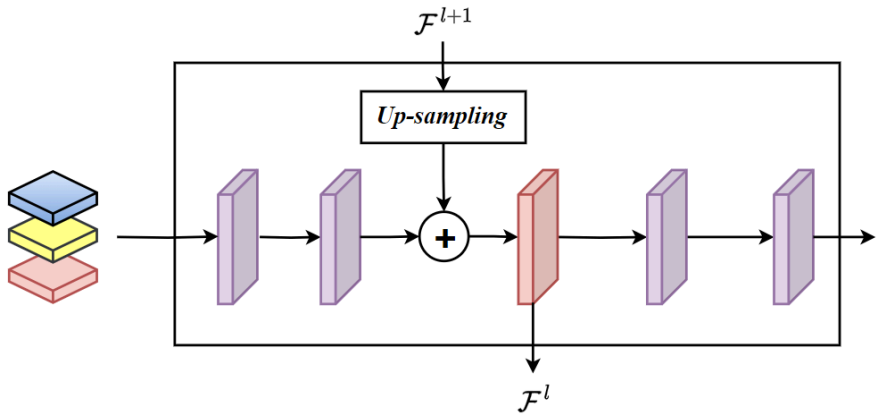


Fig. 5.2: Architecture of our Adaptive refinement network (ARN).

where $\odot$ denotes the concatenation of feature maps. The expression for convolution-based adaptive refinement (used in fig. 5.2) in the ARN is expressed as:

$$\mathcal{F}^{l} = conv(\mathcal{F}^{l}_{\mathcal{S}_1} \odot \mathcal{F}^{l}_{\mathcal{S}_2} \odot \mathcal{F}^{l}_{\mathcal{S}_3}) \tag{5.3}$$

To further improve the refined feature maps ($\mathcal{F}^{l}$), we perform multilevel refinement across ARN. The ARN performs multilevel feature refinement using the refined features of the level l^{th} and $(l-1)^{th}$. The multilevel refinement in the ARN can be expressed as:

$$\mathcal{F}^{l} = conv(\mathcal{F}^{l-1} + UP(\mathcal{F}^{l})) \tag{5.4}$$

where UP denotes the up-sampling operation with a scale factor of two. After multilevel refinement, the ARN network uses two convolution layers for processing the features, which can be expressed as:

$$\mathcal{F}^{l} = conv(conv(\mathcal{F}^{l})) \tag{5.5}$$

Finally, the proposed ARN generates refined feature maps. The feature map from the top level of ARN is used to generate two more feature maps after max pooling and convolution. The detection head contains two subnetworks for bounding box and classification which processes the refined feature maps. The generalized focal loss [191] is used to train the proposed MRN-LOD. The summary of the overall proposed MRN-LOD is discussed in the following section.

5.2.3 MRN-LOD: Multi-exposure refinement network for low-light object detection

The framework of proposed MRN-LOD is shown in Fig. 5.1. The MRN-LOD synthsizes two images with different exposures and then performs feature extraction from all images using the MFE network. The features extracted from multiple images are processed using the adaptive refinement network to refine the feature of low-light images with the help of other images with distinct exposure. The ARN helps the MRN-LOD to focus on the distinct regions in low-light images for better detection. The detector head processes the refined features to detect the class and location of the object. The architecture of the detection head is given in Fig. 5.1. We utilized the class subnetwork, and box subnetwork [191] to acquire the object class and localization, respectively. The refined feature map is processed by various convolution layers in class subnetwork as shown in Fig. 5.1 to detect the object classes. Another set of convolution layers in the box subnetwork are used to process the feature maps to improve the quality of bounding boxes. The subnetworks for class and box prediction use distinct sets of convolution layers to process the refined features ARN, which can be represented as:

$$x = conv(\mathcal{F}^l) \tag{5.6}$$

$$x = conv(x) \tag{5.7}$$

where, $\forall \mathcal{F}^l$ from the ARNs. The class and box subnet uses the feature of all levels of the ARN for the final prediction.

The proposed MRN-LOD uses the loss function of generalized focal loss [191] for training. The loss function ($\mathcal{L}$) is expressed as:

$$\mathcal{L} = -\alpha(1-q)^{\gamma} log(q) - \beta(1-d)^{\eta} log(d) \tag{5.8}$$

where α and β denote the focus on the classification and localisation, respectively. γ and η are the parameters for controlling the focus on hard examples. q denotes the probability of predicted objects and d denotes the predicted quality score.

5.3 Experimental results and assessment

5.3.1 Dataset

The availability of datasets is one of the important aspects of deep-learning based object detection methods for training and evaluation. To facilitate direct training of the object detection model for low-light conditions, we developed our synthetic dataset using the MS-COCO dataset [68]. The synthetic dataset contains all samples with synthetic low-light for eighty classes. We use the power law to synthesize the low-light image by using a random value of the exponent in the range 0.2 to 0.5. This approach produces low-light images with a large range of darkness. Moreover, the resulting images contain local variation in light according to the nature of the input image. We trained the proposed model on our synthetic dataset for training on low-light samples. To evaluate the performance of the proposed model in the real-world scenario, we used Ex-dark [165] dataset. The Ex-dark dataset [165] is publicly available for object detection in low-light conditions with 7,363 annotated low-light images. The Ex-dark dataset contains images with ten types of low-lightness for twelve kinds of objects. The annotated images help in the evaluation of object detection methods in low-light conditions. The model uses training and validation sets for training and a test set of images for evaluation on unseen images for robust comparison of the models. We used the model trained on our synthetic dataset for fine-tuning the Ex-dark dataset. The proposed model is evaluated on both datasets to show the superiority of the proposed model.

5.3.2 Implementation details

The proposed MRN-LOD has been implemented using the PyTorch framework in conjunction with MMDetection [192]. The MFE utilizes pre-trained ResNet architectures [136], specifically ResNet-50 and ResNet-101, considering hardware constraints. To optimize the MRN-LOD model, we conducted training for a total of 24 epochs employing the Stochastic Gradient Descent (SGD) optimizer. The learning rate was set to 0.01, with a decay factor of 0.1 applied at the 16^{th} and 22^{nd} epochs to enhance convergence. The training process follows a 2x training schedule, with each GPU accommodating a batch size of 3 samples. For our experiments, we employed two Titan-RTX GPUs to efficiently train the MRN-LOD model. The MRN-LOD model was trained using our synthetic dataset and Exdark dataset [165]. During the multi-scale training of MRN-LOD, we performed resizing of the shorter side of the input image to two specific scales, namely 480 and 800 pixels, to diversify the training process. To ensure optimal

model performance, we conducted an extensive analysis of exposure parameter using the grid search method in our experiments. In the context of the ablation study, we exclusively employed ResNet-50 for feature extraction using the MFE network. The ablation models underwent training for 12 epochs, and the batch size is six samples. These experimental setups and configurations were devised to thoroughly evaluate the performance and efficacy of the proposed MRN-LOD model under varying conditions.

5.3.3 Ablation study

In this section, we perform an ablation analysis to assess the contributions of the two primary components in our proposed MRN-LOD model, namely, the Multi-exposure Feature Extraction (MFE) and the Adaptive refinement Network (ARN).

We explored the effectiveness of MFE through several model configurations during the ablation study. We tweaked the configuration of MFE and developed two variants of the proposed model. The first variant uses one exposure-adjusted image along with another image (low-light or exposure-adjusted image with different parameters). This variant helped us assess the importance of distinct exposure-adjusted images in our MFE network. Furthermore, we altered the exposure parameter in the exposure-adjusted images to evaluate the exposure parameter for our model. This arrangement helps us investigate the influence of distinct exposure parameters on performance. Table 5.1 demonstrates the performance comparison among these different MFE configurations. Notably, the MFE networks (first variant) exhibit superior performance when compared to many existing state-of-the-art methods designed for low-light conditions, as demonstrated in Table 5.3. However, we observe that the proposed MFE network surpasses all other models, including our first variant. Consequently, we opt

Table 5.1: Ablation study of the proposed MFE for low-light object detection. Bold values demonstrate better performance.

Exposure	**AP**	**AP**$_{50}$	**AP**$_{75}$
[low-light, 4.0]	34.3	50.7	36.3
[4.0, 1.25]	34.1	50.1	35.9
[low-light, 1.25]	33.3	49.7	35.4
[2.85, low-light, 1.25]	34.7	51.6	37.1
[4.0, low-light, 1.43]	34.7	51.7	37.1
Proposed	34.9	51.9	37.4

Table 5.2: Ablation study of our ARN for low-light object detection.

MRN-LOD	**AP**	**AP**$_{50}$	**AP**$_{75}$
without ARN	31.4	48.7	32.9
with ARN	34.9	51.9	37.4

for the proposed MFE configuration (with three inputs) in our final model. The proposed MFE network excels by focusing on features extracted from both brighter and darker regions, leading to improved overall performance when compared to another alternative.

To assess the significance of the ARN, we investigated two model configurations: one with ARN and another without ARN. Table 5.2 presents a comparative analysis of these two configurations. Our findings reveal that the MRN-LOD model with ARN consistently outperforms its counterpart without ARN. The variant without ARN processes features independently, including undesired features, resulting in a notable performance degradation. This analysis underscores the critical role of adaptive refinement in the proposed model, which the proposed ARN addresses effectively. The ARN network facilitates the refinement of relevant features, leading to a notable enhancement in overall performance.

5.3.4 Quantitative analysis

This section presents an exhaustive quantitative analysis of the other contemporary methods and the proposed MRN-LOD on our synthetic dataset and the Ex-dark dataset. The analysis of the models with our synthetic dataset is given in Table 5.3 and Table 5.4. To maintain fairness in the comparison, we used the models and checkpoints from mmdetection [192]. The average precision (AP) is used to comparatively analyse the performance of state-of-the-art object detection models in a low-light environment in Table 5.3. The detection methods used for quantitative analysis in Table 5.3 use the ResNet-101 backbone. The proposed MRN-LOD 50 and MRN-LOD 101 use ResNet-50 and ResNet-101, respectively in the MFE for feature extraction. We analysed the detection performance with multiple pairs of the enhancement method and the detection methods. It can be observed from Table 5.3 that the object detection methods [70, 193, 194, 191] without image enhancement perform poorly under low light conditions. The use of a few image enhancement methods [2, 4, 1, 6, 7] helped the detection methods achieve comparably better performance. Sometimes, image enhancement [3, 5] leads to degrading performance. The improvement by the enhancement methods

Table 5.3: Quantitative Assessment of various methods for object detection in low-light conditions

Method	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
RetinaNet [70]	23.4	36.7	24.3	11.9	25.3	34.6
RepPoints [193]	27.7	44.1	29.1	13.3	29.7	41.1
FoveaBox [194]	24.8	39.4	25.7	13.3	27.7	34.7
GFL [191]	30.0	44.4	31.7	16.4	32.2	43.2
PIE [2] + RetinaNet [70]	25.4	39.6	26.7	11.8	27.6	37.3
PIE [2] + RepPoints [193]	29.9	47.3	31.4	14.1	32.3	44.3
PIE [2] + FoveaBox [194]	27.0	42.9	27.8	13.7	29.9	37.5
PIE [2] + GFL [191]	32.5	47.7	34.5	17.1	34.8	46.4
LIME [3] + RetinaNet [70]	21.7	34.3	22.6	9.3	23.9	33.0
LIME [3] + RepPoints [193]	25.2	40.1	26.3	10.7	27.1	38.6
LIME [3] + FoveaBox [194]	19.9	32.3	20.6	10.3	23.4	28.4
LIME [3] + GFL [191]	27.2	40.2	29.2	12.7	29.7	40.1
NPEA [1] + RetinaNet [70]	25.6	39.9	26.6	11.4	27.9	37.8
NPEA [1] + GFL [191]	32.8	48.1	34.8	16.6	35.3	47.0
SLIE [4] + RetinaNet [70]	25.2	39.5	26.3	10.7	27.4	37.1
SLIE [4] + GFL [191]	32.1	47.2	34.2	16.5	34.5	46.4
CRM [5] + RetinaNet [70]	22.8	35.3	23.7	10.7	25.4	33.3
CRM [5] + FoveaBox [194]	23.8	37.6	24.7	12.7	27.0	32.8
CRM [5] + GFL [191]	28.8	42.0	31.0	15.6	31.3	40.8
LSDD [6] + RetinaNet [70]	24.3	38.1	25.4	10.4	27.0	36.0
LSDD [6] + FoveaBox [194]	25.0	39.9	25.9	12.7	28.4	35.4
LSDD [6] + GFL [191]	31.2	45.9	33.5	15.5	33.8	45.2
ZDCE [7] + RetinaNet [70]	26.6	41.4	27.9	12.6	29.1	38.5
ZDCE [7] + FoveaBox [194]	28.0	44.2	29.3	14.5	31.2	38.3
ZDCE [7] + GFL [191]	33.7	49.4	35.8	17.5	36.5	47.6
Proposed MRN-LOD 50	<u>37.5</u>	<u>55.0</u>	<u>40.3</u>	<u>21.6</u>	<u>40.9</u>	<u>48.6</u>
Proposed MRN-LOD 101	**42.7**	**60.8**	**46.0**	**24.5**	**46.0**	**57.7**

is limited and depends greatly on the characteristics of samples in the dataset. The proposed methods outperform the other alternatives in Table 5.3 for object detection under low-light conditions. The proposed model (with ResNet-101-DCN) achieves 48.1 AP on the MS-COCO dataset as against 47.3 AP of GFL with ResNet-101-DCN. The proposed model also outperforms other models (with the same backbone) on the MS-COCO dataset.

To further analyse the proposed methods, we trained the contemporary methods with the help of data augmentation (with lightness variation) to improve the performance in low-light conditions. To ensure a fair comparison among the various models, we employed the mmdetection for training without any alterations to the hyperparameters. Given the constraints imposed by hardware limitations, it was not tenable to retrain all models with their specified configurations. Consequently, we retrained only the top-performing models under low-light conditions, which is sufficient to establish the effectiveness of our proposed model. We adopted ResNet-101 as the backbone architecture for all models considered for a fair comparison. The performance of various state-of-the-art models in low light with and without data augmentation is shown in Table 5.4. It can be seen from Table 5.4 that the performance of all models improved for low-light images after retraining with the help of data augmentation. Furthermore, the proposed model is achieving superior performance than all other models, as shown in Table 5.4.

The quantitative analysis of the contemporary methods and the proposed method on the Ex-dark [165] dataset is given in Table 5.5. The results of various methods (without enhancement) [40, 57, 193, 61] demonstrate that the existing methods are trained for normal-light images, and fail to deal with low-light conditions. The performance of the existing methods degrades under low-light conditions. The enhancement methods help the existing detection methods to achieve improved performance. However, the improvement in performance due to enhancement methods is limited. As the enhancement methods have their own constraints. The specialized methods for low-light object detection are also compared with the proposed MRN-LOD on the Ex-dark dataset. The low-light detection methods show better performance than the methods which focus on enhancement-before-detection. However, there is still scope for improvement. The proposed method achieves superior performance than the other contemporary methods.

Table 5.4: Performance of various object detection models (with ResNet-101 backbone) after training with and without Data Augmentation, separately. Models are evaluated in the ODLE dataset to verify performance in low light.

Method	Data Augmentation (DAg)	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
RetinaNet [70]	without DAg	29.0	45.1	30.5	14.4	31.3	41.8
	with DAg	30.7	48.7	33.7	15.6	34.9	43.1
GA [195]	without DAg	28.7	45.2	30.3	13.7	30.7	41.3
	with DAg	34.0	52.6	36.2	17.9	36.7	46.4
FSAF [196]	without DAg	29.8	46.1	31.2	14.2	32.1	42.5
	with DAg	31.5	49.3	33.1	15.9	33.5	42.5
ATSS [197]	without DAg	31.4	47.1	33.3	16.6	34.3	43.3
	with DAg	34.3	50.9	37.2	19.7	37.6	43.7
RepPoints [193]	without DAg	32.8	50.9	34.9	16.3	35.1	47.2
	with DAg	35.2	53.7	37.7	19.4	39.7	49.9
FoveaBox [194]	without DAg	31.5	49.1	32.9	16.4	34.4	43.9
	with DAg	32.6	51.4	34.4	16.2	35.7	44.5
FCOS [198]	without DAg	30.2	47.0	31.4	16.2	32.9	40.6
	with DAg	33.5	51.9	35.2	18.5	36.6	44.9
Generalized Focal Loss [191]	without DAg	35.9	52.4	38.5	19.2	38.7	50.2
	with DAg	39.5	57.3	42.8	22.1	43.1	53.3
Proposed MRN-LOD 101	-	**42.7**	**60.8**	**46.0**	**24.5**	**46.0**	**57.7**

Table 5.5: Quantitative assessment of distinct methods on Ex-dark dataset for low-light object detection.

Methods	Bicycle Chair	Boat Cup	Bottle Dog	Bus Motorbike	Car People	Cat Table	mAP
Faster RCNN [40]	70.3 52.9	56.2 58.8	63.5 69.1	73.4 54.3	66.6 66.9	55.6 32.3	60.0
Cascade RCNN [57]	70.4 48.3	58.0 59.3	60.7 75.4	80.3 55.0	61.0 66.2	63.4 36.6	61.2
YOLO [61]	71.8 49.7	64.5 56.8	63.9 63.8	81.6 61.8	76.8 65.7	55.4 40.5	62.7
KIND [96] + YOLO [61]	73.4 56.9	68.1 62.7	65.5 68.2	86.2 67.1	78.3 69.6	63.0 48.2	67.3
Zero-DCE [7] + YOLO [61]	79.5 65.7	71.3 68.6	70.4 75.4	89.0 67.2	80.7 76.2	68.4 51.1	72.0
EnlightenGAN [97] + YOLO [62]	86.6 71.5	65.0 68.7	69.8 77.6	92.7 74.3	69.7 69.1	72.9 56.8	72.9
LEDet [199]	77.0 59.6	73.3 64.5	60.5 76.4	86.5 61.6	73.8 63.5	68.7 46.2	67.6
MAET [200]	81.3 65.5	71.6 72.6	74.5 75.4	89.7 72.7	82.1 77.4	69.5 53.3	74.0
IAT-YOLO [201]	79.8 72.4	76.9 78.6	78.6 79.0	92.5 79.0	83.8 81.1	73.6 57.7	77.8
FRSE-Net [202]	85.9 74.4	70.7 78.7	77.6 80.9	92.8 74.3	78.7 80.6	72.0 61.0	77.3
Proposed MRN-LOD	87.2 78.6	77.9 80.5	79.5 83.4	92.6 72.6	85.9 82.1	73.8 64.3	**79.9**

5.3.5 Visual analysis

We provide visual samples in Fig. 5.3(a)-(i) to showcase the robustness and efficacy of the MRN-LOD model in low-light object recognition. These samples further illustrate that our MRN-LOD model consistently performs well in challenging low-light scenarios.

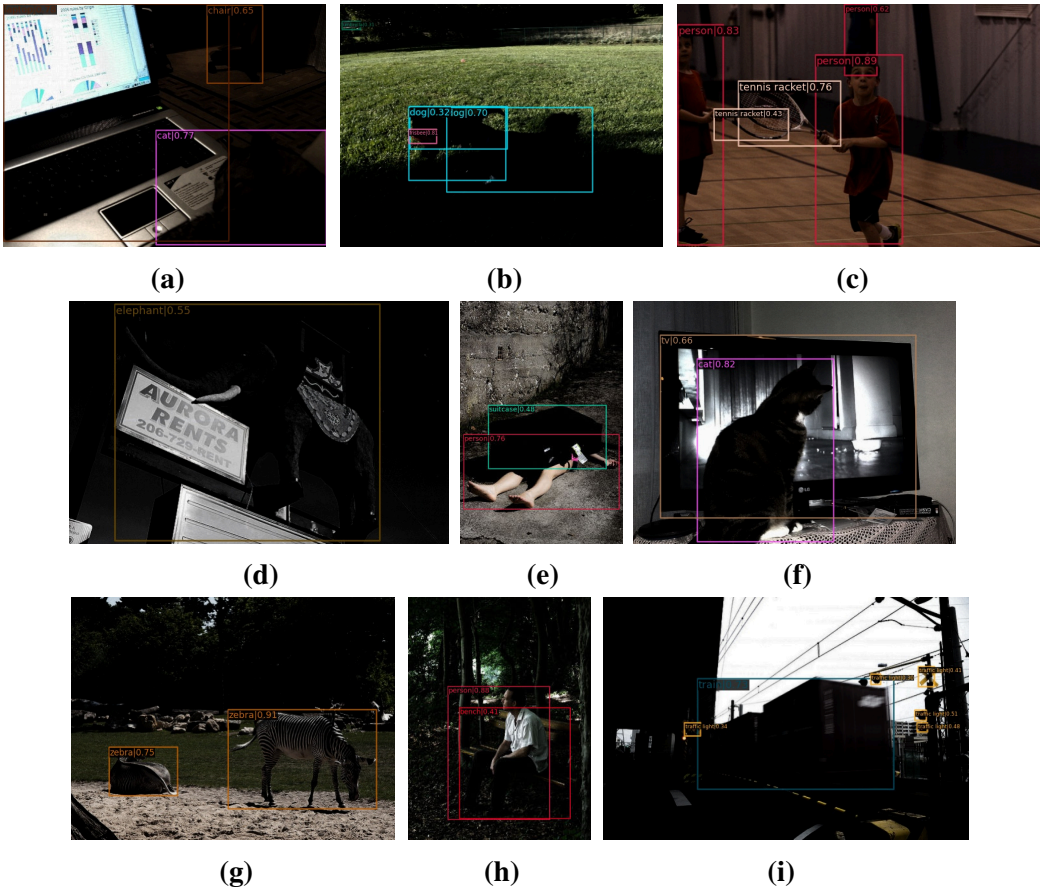


Fig. 5.3: Visual results of the proposed MRN-LOD network on sample images from validation set of the ODLE dataset.

5.3.6 Computational Time

In this section, we analyse the computational efficiency of the proposed MRN-LOD. Most of the existing approaches use enhancement before detection for low-light object recognition. Such methods require a higher computational time, as they require separate time for both enhancement and detection. The proposed MRN-LOD eliminates the requirement of an explicit enhancement step to be performed before detection. Thus, it reduces the time required by the enhancement step and requires the time only for detection step. Hence, it requires less computational time as compared to the approaches

based on enhancement before detection. The source codes and checkpoints are used from the authors' official source or MMdetection for a fair comparison. Existing object recognition methods [70, 193, 194, 191] without any image enhancement method require comparable or slightly less time than our MRN-LOD methods. However, it can be noted that the performance of these methods is very poor compared to our proposed MRN-LOD. In methods based on enhancement before detection, the enhancement step requires additional time to enhance image quality so that detection performance can be improved. We have considered both deep learning based and statistical image enhancement methods to assess the detection performance. The execution of statistical methods [2, 3, 4, 5, 6] is performed on i7 processor. The total time required by methods based on enhancement before detection is higher than the time required by the proposed method. To summarize, our MRN-LOD 50 method is the fastest among all tested methods with better detection performance. MRN-LOD 101 takes an additional 0.01 seconds but offers a 12% increase in AP compared to direct detection methods.

5.4 Summary

In this chapter, we present a new Multi-exposure refinement network for low-light object detection (MRN-LOD) to avoid the need for enhancement before detection. The MRN-LOD contains: multi-exposure feature extractor, adaptive refinement network, and detection head. The developed multi-exposure feature extractor extracts features from the multi-exposure images generated by the low-light image. We introduced the notion of feature extraction from muylti-exposure images for object detection in low light. In addition, we proposed an adaptive refinement network to refine the features of low-light images for better detection performance. The detection head uses the refined features to perform object detection. Extensive experimentation on real-world and synthetic datasets shows the superiority of the proposed MRN-LOD.

Chapter 6

Object Recognition in Hazy Environment

This chapter presents a new bi-stream feature fusion network for hazy object detection/recognition. The chapter includes the dataset for object detection in the hazy environment, hybrid input, bi-stream feature extractor, multi-level feature fusion network, experimental results and analysis, and the conclusion.

6.1 Introduction

For object detection in a hazy environment, the obvious solution [154, 152] is to apply image dehazing (as pre-processing) [76, 80] before the object detection[1] method. However, such an approach requires additional time and computation to enhance the image. Moreover, the performance of object detection relies highly on the performance of the dehazing algorithm in such an approach. Alternatively, some researchers [203, 204] used domain adaptation for object detection in a hazy environment. DA-based approaches attempt to learn haze-relevant features from an image-dehazing dataset to achieve better performance. In general, DA-based approaches have poor generalization across the datasets. A single-step solution i.e. direct object detection in a hazy environment is always desirable.

In this work, we developed an end-to-end trainable model for object detection in a hazy environment. The existing object detection model for hazy environments uses the image-dehazing dataset to learn haze-relevant features using domain adaptation. There are benchmark datasets available for evaluation of the performance. The presently available dataset [185, 205] contains samples of five to eight classes. To train a deep learning based end-to-end trainable method, we require a large amount of data. The performance of DL-based approaches is primarily governed by the quality and quantity

[1]Note: In this chapter, recognition and detection both means same i.e. object recognition.

of data samples along with the model. In this work, we develop a synthetic dataset for hazy object detection (DHOD) with twenty classes and more than twenty thousand annotated images. DHOD is synthesized using the Pascal-VOC dataset [69]. The DHOD contains annotated hazy images with distinct-level of haze. We used the developed DHOD dataset to train the proposed model and the benchmark RTTS dataset to test the performance and generalization.

We propose a novel architecture namely Bi-stream feature fusion (BFF) for object detection in the hazy environment. BFF consists of a Bi-stream feature extractor (BFE), and Multi-level feature fusion (MFF). The Bi-stream feature extractor uses two streams with inputs as a hazy image in one stream and a hybrid image in the other stream. The hybrid image is a combination of the min-channel, gradient, and max-channel of the hazy image. The hybrid image facilitates the BFE network in extracting appropriate features from hazy images. The proposed BFE network extracts multi-level features from the hazy image and hybrid image. Further, we propose the MFF network to perform the fusion of extracted features at multiple levels and process the extracted features. The BFF uses BFE for feature extraction, and MFF for feature fusion and processing. Finally, the proposed approach leverages the class subnetwork (subnet) and box subnet for object localization and classification. The major contributions of the proposed work are summarized as follows:

1. We proposed a novel Bi-Stream Feature Extraction (BFE) network. The BFE extracts features at multiple levels from the hazy image and hybrid image.

2. We proposed a multi-level adaptive feature fusion (MFF) network to perform the fusion of adequate features. MFF provides attention to appropriate features across the streams.

3. We developed a new dataset for hazy object detection (DHOD) to train the object detection models in the hazy environment based on the Pascal-VOC dataset.

4. We proposed an object detection model for the hazy environment namely Bi-stream feature fusion network, which outperforms state-of-the-art object detection models in a hazy environment.

6.2 The Proposed Approach

The existing object detection models are developed and trained to deal with images captured in normal environments. The existing models perform poorly in a hazy environment as such methods are not capable of dealing with hazy conditions. Thus, we

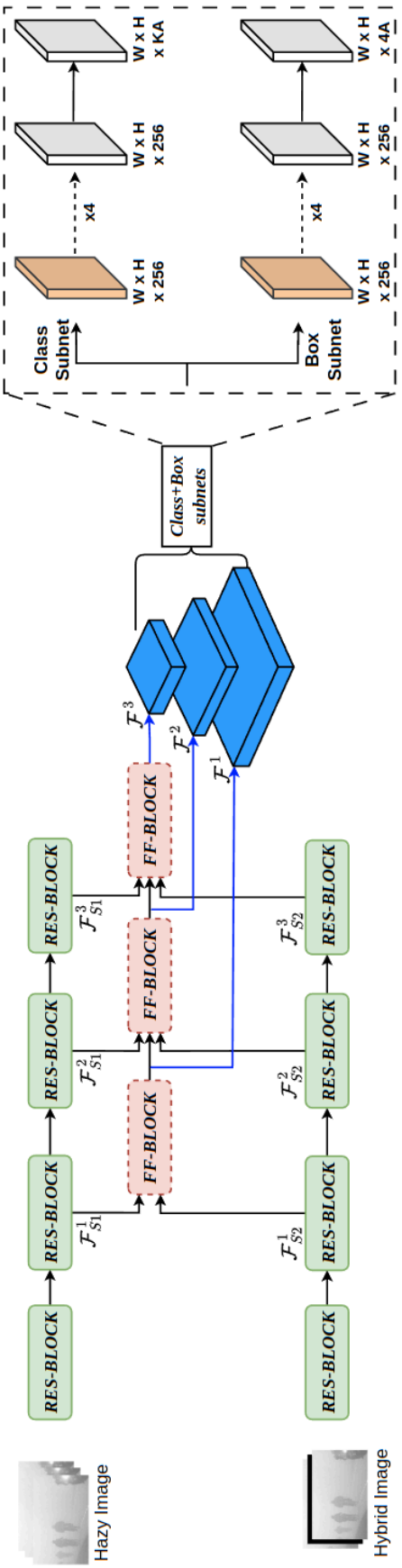


Fig. 6.1: Framework of Bi-Stream Feature Fusion for Object Detection in Hazy Environment.

propose a Bi-stream feature fusion (BFF) model for object detection in a hazy environment. The proposed model involves three sub-modules: Hybrid input, Bi-stream feature extractor, and feature-fusion block. The framework of proposed BFF network is shown in Fig. 6.1.

The proposed Bi-stream feature extractor is developed with a combination of two-streams (rather than a single stream) that leverages res-blocks in each stream. Each res-block contains a set of convolution, batch-norm and skip-connection to extract the features from the input image. In between each res-block, there is a pooling layer to reduce the size of the feature map. The set of res-blocks helps in the extraction of multi-level features from the hazy image and hybrid input. The feature-fusion block (FF-block) performs convolution-based and multi-level fusion to process the feature map in an appropriate manner. The feature fusion block is developed to perform an adaptive feature fusion, as shown in Fig. 6.2. The fused features are processed with the help of class and box subnets to detect the location of the object along with the class. The class and box subnets are two distinct convolution networks which generate the class and bounding box respectively. The detailed discussion is given in the following sections.

6.2.1 Hybrid image formation

Images captured in a hazy environment contain less visible and faded details, which degrades the performance of object detection. The existing object detection models use single-stream (i.e. only hazy images as input) for feature extraction and detection. Feature extraction from hazy images is a challenging task due to lesser details. Thus, we introduced the notion of hybrid input which to facilitate the feature extraction from the hazy image. Further, we propose the BFF model with two distinct streams to deal with the hazy image (haze-stream) and the hybrid image (hybrid-stream). The feature extraction from haze-stream and hybrid-stream uses pre-trained backbones which requires three-channels input; thus, we develop a hybrid input with three-channels.

The notion of hybrid input is to facilitate the feature extraction from hazy images. The dehazed image as the hybrid input may help the feature extraction; however, image dehazing techniques have their own limitations and require a large amount of computational time. Thus, we considered three such components as channels which helps in haze-related domains and require less computational time to restrict computational overhead. In DCP [20], the method uses a dark channel prior which uses window-based element-wise minimum to perform image dehazing. We considered the simple element-wise minimum pixel (min-channel) as the first channel of hybrid input. The

min-channel of hazy image is influenced by the airlight which assist the hybrid-stream in extraction of haze-relevant feature. The min-channel used in the hybrid image is represented as:

$$\mathcal{H}_1 = \min_{c \in \{R,G,B\}} (\mathbf{I}^c_{hazy}) \tag{6.1}$$

where $\mathbf{I}$ denotes the hazy image, c denotes the channels of the image (i.e. red, green, and blue channel), and $\mathcal{H}_1$ denotes the first channel of the hybrid image.

The second channel of hybrid input contains the gradient of the hazy image. Image edges are important to detect an object. However, it is difficult to extract features related to edges from the hazy image because hazy images contains faded details. Many dehazing methods [76, 206] focuses on improvement of gradient information to enhance the image. Thus, we leverages the gradient information as the second channel of our hybrid input to provide the textural-details for facilitating the feature extraction. The second channel of hybrid image $\mathcal{H}_2$ is represented as:

$$\mathcal{H}_2 = max(\max_{c \in \{R,G,B\}} (\nabla_h \mathbf{I}^c), \max_{c \in \{R,G,B\}} (\nabla_v \mathbf{I}^c)) \tag{6.2}$$

where ∇_h and ∇_v denote the first order derivative across horizontal and vertical directions of the image $\mathbf{I}$.

Many approaches consider the inverted hazy images as the low-light images to enhance the image quality. The element-wise maximum is used in many low-light image enhancement methods. Moreover, min-channel is an assumption-based consideration for the hybrid input and, it can have minor limitations. Thus, we considered the max-channel as the third channel to counter the limitations of the min-channel in the hybrid channel. The max-channel can be represented as:

$$\mathcal{H}_3 = \max_{c \in \{R,G,B\}} (\mathbf{I}^c) \tag{6.3}$$

where, $\mathcal{H}_3$ denote the third channel of the hybrid input. The final hybrid input ($\mathbf{I}_{hybrid}$) is a combination of the three drived channels i.e. $[\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3]$.

6.2.2 Bi-Stream feature extraction

The existing feature extractors perform poorly in a hazy environment. Thus, a new feature extraction paradigm for object detection in hazy images is required. We propose a novel feature extractor namely, a Bi-Stream feature extraction (BFE) network for hazy images with two distinct streams. We introduce two streams in BFE namely, haze-stream, and hybrid-stream to deal with distinct input separately. The haze-stream

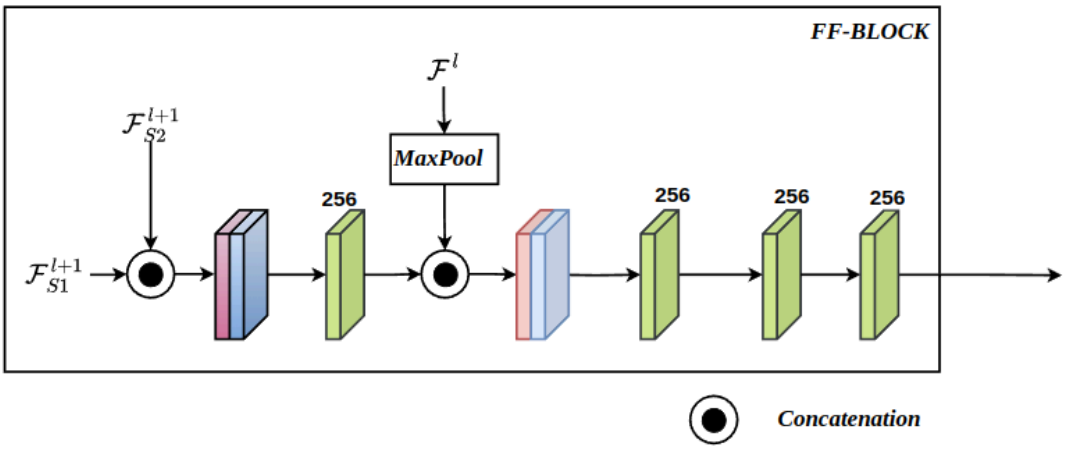


Fig. 6.2: Architecture of feature fusion block

facilitates the feature extraction from the input hazy images. The hybrid stream assists the haze-stream in extracting appropriate features from the hazy images. The framework of BFF is shown in Fig. 6.1. The BFE extracts multi-level features from the above-mentioned streams. We denote multi-level features of hazy-stream as $\mathcal{P}_{\mathcal{S}_1}$, and hybrid-stream as $\mathcal{P}_{\mathcal{S}_2}$. The multi-level features of j^{th} stream ($\mathcal{P}_{\mathcal{S}_j}$) is represented as:

$$\mathcal{P}_{\mathcal{S}_j} = [\mathcal{F}^3_{\mathcal{S}_j}, \mathcal{F}^2_{\mathcal{S}_j}, \mathcal{F}^1_{\mathcal{S}_j}] \tag{6.4}$$

where $\mathcal{S}_j$ denotes the j^{th} stream, and $\mathcal{F}^l_{\mathcal{S}_j}$ denotes the feature map of l^{th} level of $\mathcal{P}_{\mathcal{S}_j}$. The extraction of l^{th} level features in the j^{th} stream can be represented as:

$$\mathcal{F}^l_{\mathcal{S}_j} = \psi^k(|\mathbf{I}|_{\mathcal{S}_j}) \tag{6.5}$$

where $l = k - 1$ and k is the block number of ResNet-like architecture (ψ) [136], $\mathbf{I}$ denotes the input image, and $\mathcal{S}_j \in [hazy, hybrid]$ denotes the input to the stream.

The streams of BFE attempt to extract features from the hazy images and hybrid input. However, the problem of irrelevant features may still occur due to noise amplification, saturation, and other artefacts in the hybrid input, and suppressed details in the hazy image. Thus, we perform an adaptive feature fusion at multiple levels to extract the appropriate features for object detection. We propose an adaptive feature fusion for the features of the proposed bi-stream model. The proposed fusion focuses on the appropriate features which help in effective object detection in hazy images. The convolution-based adaptive feature fusion is used in the proposed feature fusion block, discussed in the next section.

6.2.3 Feature Fusion Block

The extracted multi-level features of the Bi-stream feature extraction network can be processed separately to detect the objects. However, such an approach processes the undesired features along with the appropriate features which degrade the performance of object detection as discussed in section 6.3.3. The multi-level features of both streams may contain appropriate features from distinct regions of images. Thus, we perform the adaptive feature fusion of the Bi-stream features. We propose the feature fusion network for the fusion and processing of the Bi-stream features. The detailed architecture of the feature fusion network is shown in Fig. 6.2. The simplest fusion approach can be given by considering the weighted sum, which can be expressed as:

$$\mathcal{F}^l = w_1 * \mathcal{F}^l_{\mathcal{S}_1} + w_2 * \mathcal{F}^l_{\mathcal{S}_2} \tag{6.6}$$

where $\mathcal{F}^l$ denotes the fused feature map, w_1 denotes weights for the first feature map, and w_2 denotes the weights of the second feature maps. The weighted sum strategy provides equal importance to the complete feature map. However, feature fusion is required to combine adequate features from distinct streams. The weighted sum of the feature maps is unable to fuse the adequate features based on the characteristics. However, the feature maps may contain adequate features in one region while having irrelevant features in others. Thus, we proposed an adaptive feature fusion to fuse the multi-level features. The proposed multi-level adaptive feature fusion approach is represented as:

$$\mathcal{F}^l = conv(\mathcal{F}^l_{\mathcal{S}_1} \odot \mathcal{F}^l_{\mathcal{S}_2}) \tag{6.7}$$

where $\mathcal{F}^l$ denotes the fused feature map, $conv$ denotes convolution, and $\odot$ denotes the concatenation of feature maps.

We perform multi-level fusion with the feature maps of the previous feature fusion block to further improve the fused feature maps ($\mathcal{F}^l$). The feature fusion network performs the multi-level fusion by using the fused features of $(l+1)^{th}$ and l^{th} block. The multi-level fusion in the MSPF network can be expressed as:

$$\mathcal{F}^l = conv(\mathcal{F}^{l+1} \odot MaxPool(\mathcal{F}^l)) \tag{6.8}$$

where $MaxPool$ denotes the max pooling operation with a kernel and stride of two. The multi-level fusion is used in all the levels of the proposed methods except the first level. After fusion, the feature fusion network uses three convolution layers for

processing the features, which can be expressed as:

$$\mathcal{F}^l = conv(conv(conv(\mathcal{F}^l))) \tag{6.9}$$

Finally, the proposed feature fusion block generates fused multi-level features $\mathcal{P}_{ff}$, which can be expressed as:

$$\mathcal{P}_{ff} = [\mathcal{F}^3, \mathcal{F}^2, \mathcal{F}^1] \tag{6.10}$$

To process the fused multi-level features, we use the detection approach which is similar to RetinaNet. The training of the proposed network uses generalized focal loss [191]. The summary of the overall proposed BFF network is discussed in the following section.

6.2.4 Bi-Stream feature fusion for object detection in hazy environment

The framework of the proposed BFF network for object detection in hazy environment is shown in fig. 6.1. The BFF network extracts multi-level features from a hazy image and hybrid input, then feeds into a feature fusion network. The feature fusion network performs a multi-level adaptive feature fusion to generate a fused feature pyramid. The fused feature pyramid requires processing for object detection and localization. We leveraged the class subnet and box subnet [191] to obtain the object class and localization, respectively. The class subnet processes each feature map of the fused pyramid using a convolution-network to detect the object classes from an image. Another convolution-network (i.e. box subnet) processes the feature maps to improve the quality of bounding boxes. The training of class and box subnets along with the feature extraction and fusion block is performed using the generalized focal loss. The class and box subnets use distinct sets of convolution layers to process the features from the fused feature pyramid, which can be represented as:

$$x = conv(conv(\mathcal{F}^l)) \tag{6.11}$$

where, $\forall \mathcal{F}^l \in \mathcal{P}_{ff}$. The class and box subnet uses the feature of all the levels from the fused feature pyramid for prediction. The loss function ($\mathcal{L}$) is expressed as:

$$\mathcal{L} = -\sum_{m=1}^{c} \alpha_m (1 - p_m^n)^\gamma log(p_m^n) \tag{6.12}$$

where α_m represents weights of m^{th} class, p_m^n represents the probability of class c for sample t, c represents the overall class number, and γ controls the degree of weights of samples.

6.3 Experimental results and analysis

6.3.1 Dataset

The existing object detection method for hazy images uses the pre-trained models and attempts to learn the haze-relevant features using domain adaptation. The models leverage either the RESIDE dataset or the foggy-cityscapes dataset to learn haze characteristics. Further, the validation of the object detection methods for a hazy environment uses the RTTS dataset [185] which is a subset of the RESIDE dataset. In such scenarios, the object detection methods tend to learn features which are similar to the features of RTTS dataset. Thus, such detectors perform well on the RTTS dataset and fail to generalize in the real-world scenario (or images of other datasets). In the Foggy-cityscapes dataset, there are eight object classes and three levels of haze for both training and validation. To perform end-to-end training and demonstrate the generalization capabilities of our proposed BFF model, we develop a new dataset which contains hazy images based on the images of Pascal-VOC dataset for training the object detection methods in a hazy environment. Our synthesized dataset is used to train the proposed model and then validation is performed using the datasets available for validation: RTTS dataset, Foggy-cityscapes, and Foggy-driving dataset.

The synthetic haze generation for our DHOD (dataset for hazy object detection) uses two approaches. In both approaches of haze generation, DHOD leverages the annotated images of pascal-VOC dataset. The first approach of haze generation is based on the physical image formation model [118, 76], which requires depth map and atmospheric light to generate haze in an image. The depth map is used in the estimation of transmission map with the aid of a beta parameter to control the haze in the depth. Further, the atmospheric light also controls the haze in an image. The variation of beta parameter and the atmospheric light helps in synthesizing the images with distinct variations of haze. The depth map is not available for the images. Thus, the estimation of depth map from a single image is performed using monodep [207]. The estimated depth is utilized for transmission map synthesis using the beta parameter. We used random values for atmospheric light and beta parameters to vary the level of haze in the images in a random manner. The discrete haze characteristics are not suitable for real-world scenarios. The value of atmospheric light and beta are in the range of 0.7

Table 6.1: Details of the datasets used for experimentation.

Dataset	Classes (instances)	Total samples
Synthetic	aeroplane, bicycle, bird, boat, bottle, bus, car, cat, chair, cow, dining table, dog, horse, motorbike, person, potted plant, sheep, sofa, train, tvmonitor	24,640
RTTS	Person, Car, Bus, Bike, Bicycle	4,322
Foggy-cityscapes	Person, Rider, Car, Truck, Bus, Train, Bike, Bicycle	5000
Foggy-Driving	Person, Rider, Car, Truck, Bus, Train, Bike, Bicycle	101

to 0.95. The second approach uses a generative method based on Cycle-GAN [208] to synthesize haze controlled by the generator. In the second approach of haze generation for DHOD, we developed a cycle-GAN to generate hazy to haze-free and haze-free to hazy samples. We trained the developed model on an unpaired set of hazy and normal images collected from distinct sources. The Cycle-GAN is used to generate haze in the samples. The final DHOD contains sixty percent of hazy images from the first approach and the forty percent of hazy images from the second approach. The training of the proposed model on our developed DHOD dataset and validation on the other datasets shows the superiority and generalization of the proposed model. The details of the datasets used for experimentation are given in Table 6.1.

6.3.2 Implementation details

The proposed BFF is implemented using the PyTorch framework. Both streams of the BFE network use pre-trained ResNet-101 in all experiments. The BFF network is trained for ten epochs using Stochastic Gradient Descent (SGD). The learning rate for the SGD optimizer is 0.01 for all experiments with decay at 8^{th} and 11^{th} epochs. During training, the model uses a batch size of 4 samples per GPU on the machine with two Titan-RTX GPUs and 128GB RAM. The proposed BFF model is trained on the synthesized DHOD dataset and validated on the RTTS dataset to analyze the performance in a hazy environment. In multi-scale training, we resized the shorter side of the input image into multiple scales between 360 to 640. In the ablation study, the BFF uses ResNet-101 in both streams of the BFE network. Further, the models are trained for twelve epochs during ablation with a batch size of eight.

6.3.3 Ablations

In this subsection, we analyzed the effectiveness of hybrid input and multi-level feature fusion. Table 6.2 shows the analysis of distinct components of the hybrid image. We have used three components in the hybrid image since the BFE network uses pre-trained backbone which can deal with three-channel input. Each component in the hybrid image is analysed to evaluate its impact on the overall performance. The mean average precision measure is used for analysis. Firstly, we evaluated the impact of individual components (i.e. x_{min}, x_{max}, and x_{delta}) by processing just one component in the hybrid frame and having other components as zero. The results shown in Table 6.2 show that only one component in the hybrid input reduces the overall mAP. The exclusion of two components from the hybrid input reduces the performance of the proposed methods by at least 1.6 mAP value. Further, the exclusion of a single component from the hybrid frame decimates the performance by at least 0.9 mAP value.

Table 6.3 shows the impact of multi-level fusion in the proposed method. we developed two alternative configurations, one with multi-level fusion and the other one without fusion. The BFF model with the proposed multi-level fusion outperforms the variant of our model without fusion. Direct detection using the distinct level of features from both streams is unable to process the features adequately. Thus, the BFF model without fusion results in a lesser mAP. On the other hand, the processing of bi-stream features using the proposed approach helps the proposed model in achieving better performance. This analysis shows the importance of proposed multi-level features in our BFF model. Our BFF model performs the fusion of appropriate features within and across the stream which results in a better performance.

Table 6.2: Ablation study of the hybrid input

X_{min}	X_{max}	X_{delta}	mAP
✓	✓	–	48.4
✓	–	✓	50.3
–	✓	✓	49.3
✓	–	–	47.9
–	✓	–	48.7
–	–	✓	49.6
✓	✓	✓	51.2

Table 6.3: Ablation study of multi-level feature fusion.

BFF model	mAP
without fusion	46.9
with fusion	51.2

6.3.4 Quantitative Analysis

This section presents an exhaustive quantitative analysis of the proposed BFF and state-of-the-art methods on the RTTS dataset, Foggy-cityscapes dataset, and Foggy-driving dataset. The values of quantitative measures for other methods are taken from the source provided by the authors for a fair comparison. The class-wise average precision and mean average precision of each method are used for quantitative comparison. State-of-the-art methods (used for comparison) use ResNet-50, ResNet-101, and VGG-16 as the backbones. The RTTS dataset, Foggy-cityscapes dataset, and Foggy driving dataset are widely used for the assessment of object detection methods in hazy environments. The proposed method outperforms the other methods in hazy conditions as well as normal conditions.

6.3.4.1 Validation on RTTS Dataset

The performance comparison of the proposed BFF with other contemporary methods on the RTTS dataset is given in Table 6.4. RTTS is a part of RESIDE (a well-known image dehazing dataset). It contains more than four thousand hazy images to assess the performance of object detection in a hazy environment. The images are annotated for five classes in RTTS namely, bicycle, bike, bus, car, and person. The RTTS dataset is not used for training the proposed method. Pascal-VOC dataset is used for warm-up training of the proposed model. Then, we used the synthesized DHOD dataset for training the model to deal with hazy object detection. The trained model is validated on the RTTS dataset. The contemporary methods are trained on the RESIDE dataset for adaptation of the haze in the detection model. The supervised training on RESIDE helps the existing models in learning the haze-relevant characteristics which are similar to the RTTS dataset. Table 6.4 shows that the proposed BFF-101 (with ResNet-101 backbone) outperforms the other state-of-the-art methods. Further, the BFF-50 model achieves second best for most of the classes. For the bicycle class, the BFF-50 achieves the third-best score. The mAP shows the superior performance of the proposed BFF network.

Table 6.4: Quantitative results of state-of-the-art methods in the hazy environment on RTTS dataset. The values in bold depict the best performance and the values in underlined depict the second best.

Method	Person	Car	Bus	Bike	Bicycle	mAP
Faster R-CNN [40]	46.6	39.8	11.7	19.0	37.0	30.9
DA-Faster [209]	42.5	43.7	16.0	18.3	32.8	30.7
SWDA [210]	40.1	44.2	16.6	23.2	41.3	33.1
SCL [211]	33.5	48.1	18.2	15.0	28.9	28.7
PBDA [147]	37.4	54.7	17.2	22.5	38.5	34.1
SADA [212]	37.9	52.7	14.5	16.1	26.2	29.5
DA-ODFS [213]	47.7	53.4	19.1	30.2	<u>49.3</u>	39.9
Proposed BFF-50	<u>71.3</u>	<u>52.5</u>	<u>26.4</u>	<u>43.8</u>	49.1	<u>48.6</u>
Proposed BFF-101	**74.0**	**55.2**	**28.4**	**46.9**	**51.6**	**51.2**

6.3.4.2 Validation on Foggy-Cityscapes Dataset

Another widely used dataset for performance analysis of object detection methods in hazy environments is Foggy-cityscapes. The cityscapes dataset is processed to add foggy conditions and generate foggy-cityscapes. Many methods attempted domain adaptation of the cityscapes dataset to the foggy-cityscapes dataset for object detection in a hazy environment. The existing models are finetuned on the foggy-cityscapes to learn the haze-relevant features for object detection in hazy environments. The foggy-cityscapes dataset contains seven classes: bicycle, bike, bus, car, person, rider, train, and truck. The proposed BFF network is finetuned on the foggy-cityscapes dataset to learn the features related to the new classes rider and truck. The proposed methods and other contemporary methods are validated on the images of foggy-cityscapes and results are shown in Table 6.6. The mAP shows that the proposed BFF-101 outperforms the existing methods for most of the classes. Further, it achieves the second best result for the truck and bus class. The overall mAP shows the superior performance of the proposed method on the foggy-cityscapes dataset.

6.3.4.3 Validation on Foggy-Driving Dataset

One of the widely used datasets for the validation of object detection methods in a hazy environment is Foggy-driving. The Foggy-driving depicts driving scenes from real-world foggy environments. The images are captured in a foggy environment with

Table 6.6: Quantitative results of state-of-the-art methods in the hazy environment on Foggy-cityscapes dataset. The values in bold depict the best performance and the values in underlined depict the second best.

Method	Person	Rider	Car	Truck	Bus	Train	Bike	Bicycle	mAP
Faster R-CNN [40]	32.0	39.5	36.2	19.4	32.1	9.4	23.3	33.2	28.1
DA-Faster [209]	37.2	46.8	49.9	28.2	42.3	30.9	32.8	40.0	38.5
SWDA [210]	29.9	42.3	43.5	24.5	36.2	32.6	30.0	35.3	34.3
SCL [211]	30.7	44.1	44.3	30.0	47.9	42.9	29.6	33.7	37.9
PBDA [147]	34.9	46.4	51.4	29.2	46.3	43.2	31.7	37.0	40.0
Mega-CDA [203]	37.7	49.0	52.4	25.4	49.2	46.9	34.5	39.0	41.8
UaDAN [214]	36.5	46.1	53.6	28.9	49.4	42.7	32.3	38.9	41.1
MEAA [215]	34.2	48.9	52.4	30.3	42.7	46.0	33.2	36.2	40.5
SADA [212]	48.5	<u>52.6</u>	62.1	29.5	50.3	31.5	32.4	45.4	44.0
CaCO [216]	38.3	46.7	48.1	33.2	45.9	37.6	31.0	33.0	39.2
MGA [217]	43.9	49.6	60.6	29.6	50.7	39.0	38.3	42.8	44.3
DA-ODFS [213]	39.9	51.6	59.0	**39.7**	**58.0**	<u>49.1</u>	<u>39.2</u>	45.1	47.6
Proposed BFF-50	<u>48.9</u>	52.1	<u>65.5</u>	35.1	56.8	47.8	38.6	<u>46.2</u>	<u>48.9</u>
Proposed BFF-101	**51.7**	**54.2**	**66.8**	<u>37.3</u>	<u>57.8</u>	**51.2**	**41.1**	**48.5**	**51.1**

the mobile phone and web. Many approaches use foggy-driving dataset to test the real-world generalization of their methods. The validation of different methods on the real-world foggy-driving dataset is shown in Table 6.5. Various methods have shown the performance analysis on synthetic datasets and their codes are unavailable to test on the real-world dataset. The performance analysis on the Foggy-driving dataset shows that the proposed BFF-101 can generalize on real-world scenes.

Table 6.5: Quantitative results of state-of-the-art methods in the hazy environment on Foggy-Driving dataset. The values in bold depict the best performance and the values in underlined depict the second best.

Method	mAP
Faster R-CNN	26.4
DA-Faster	31.6
SWDA	33.5
SCL	33.5
SADA	32.3
DA-ODFS	<u>34.6</u>
Proposed BFF-101	**36.4**

6.3.5 Computational time

We analysed the computational time of the proposed method and other contemporary methods [40, 209, 210, 211, 213, 215]. The codes and parameters are used as given by the authors for computational time analysis. We used images of size 1000×600 for the analysis. Most of the methods are focused on dehazing or its variants for adaptation. Further, the contemporary methods use the two-stage detection mechanism to achieve better accuracy in hazy conditions. The methods [40, 209, 210, 211, 213, 215] require different amounts of time for training; however, the time required for the inference is approximately the same. These methods process seven frames per second. The proposed method leverages a single-stage detection mechanism with BFE and feature fusion block. The proposed method with ResNet-101 processes fourteen frames per second. Further, our method with ResNet-50 processes eighteen frames per second. The proposed method achieves better detection performance than the other contemporary methods.

6.4 Summary

In this chapter, we propose a Bi-stream feature fusion (BFF) network for object detection in a hazy environment. The BFF network consists of three modules: hybrid input, Bi-stream feature extractor (BFE), and multi-level feature fusion. We present the notion of hybrid input to extract features from the hazy images in an effective manner. This paper leverages the hybrid input for feature extraction from the hazy images to avoid the requirement of enhancement in hazy object detection. The proposed BFE network extracts multi-level features from the hazy image and hybrid input. The multi-level feature fusion (MFF) network performs the convolution-based adaptive feature fusion and processes the extracted features. The proposed BFF model outperforms other state-of-the-art methods in hazy environments while achieving competitive performance in normal conditions. Another challenge in hazy object detection is the unavailability of a dataset with sufficient samples and classes. In this work, we developed a synthetic object detection dataset for a hazy environment (DHOD). The DHOD dataset contains twenty object classes with more than twenty thousand samples.

Chapter 7

Conclusion and Future Work

In the thesis, we developed new methods for object recognition in low-light and hazy environment. We attempted object recognition in low-light and hazy environment by two approaches. In the first approach, we developed the methods for low-light image enhancement and single image dehazing to deal with low-light and hazy image for better object recognition. In the second approach, we developed the methods for direct object recognition from the low-light/hazy images. We developed illumination estimation for nature preserving low-light image enhancement. The algorithm gives natural contrast enhancement for low-light images. Further, we proposed a new adaptive illumination adjustment which helps in providing appropriate lightness to all regions in the images with varying illumination. In addition to this, we developed a novel branched encoder-decoder based deep simultaneous estimation network called DSE-Net. The proposed method decomposes images into reflectance and illumination for low-light image enhancement. The proposed image decomposition architecture formulates a multi-objective loss function that considers the retinex theory. After decomposition of the image, the proposed method performs deep illumination adjustment for effective treatment of low-light regions. The developed enhancement methods are capable of dealing with varying illumination. Further, we developed a variational optimization for single image dehazing to improve recognition in hazy environment. We proposed an initial estimate of the transmission map that uses structure-aware smoothing. Further, we formulated a variational optimization which is used for the estimation of the final transmission map. The variational optimization refines the initial transmission with the help of various regularization terms. The proposed algorithm estimates dark-channel based uniform atmospheric light. These methods improve the performance of object recognition up to a certain limit. However, the computational requirement increases to deal with the low-light/hazy conditions using such approach which harms the real-time

performance of the object recognition systems. We developed direct object recognition methods for low-light environment (MRN-LOD) and hazy environment (BFF). These methods focuses on feature extraction from the low-light and hazy images using the developed techniques. The extracted features are refined and used for the classification and localization. We developed a DHOD dataset for training the object detection model directly on hazy annotated images. We showed the validation on the RTTS dataset after training on our DHOD dataset. Our extensive experimentation on both a synthetic dataset and the existing datasets demonstrates the superior performance of our proposed approach in challenging low-light and hazy conditions. Furthermore, it is noteworthy that our method maintains a competitive performance for object detection even under normal lighting conditions.

7.1 Future Work

The methods developed in this thesis represent a significant advancement in the field of computer vision, particularly for object recognition in low-light and hazy environments. However, as with any research endeavor, there are numerous opportunities for further exploration and enhancement. Below are several key directions for future work:

- The lack of a comprehensive dataset for night-time hazy conditions is not available. Future work will focus on creating such a dataset, which will include a diverse range of scenes and objects under varying levels of haze and low light. This dataset will be crucial for developing and benchmarking new algorithms tailored specifically for night-time hazy object detection.

- While the current methods demonstrate robust performance, further optimization is necessary to enhance their speed and computational efficiency without compromising accuracy. Future research will explore advanced techniques such as hardware acceleration, algorithmic refinement, and parallel processing. Additionally, leveraging techniques from hardware-specific optimizations, such as GPU and TPU acceleration, will be considered to achieve real-time processing capabilities.

- Building on the current work, future research will extend the scope to include object recognition under other adverse weather conditions, such as rain, snow, and extreme brightness. Each of these conditions presents unique challenges. Research will focus on adapting and expanding the existing algorithms to handle these new challenges effectively. This may involve developing novel feature

extraction techniques or enhancing existing models with new layers and architectures tailored to these conditions.

- Further studies will assess the impact of these advancements in practical applications, including urban planning, environmental monitoring, and public safety. For instance, object recognition in low-light and hazy conditions could significantly enhance surveillance systems, disaster management, and autonomous vehicle navigation. Research will aim to quantify the benefits and potential improvements in these areas, considering both technical performance and societal impact.

- Future research will also explore unconventional methods and emerging technologies that could further enhance object recognition capabilities. This may include investigating the potential of quantum computing, or novel machine learning paradigms. These explorations could lead to breakthroughs in handling the complex and dynamic conditions of low-light and hazy environments.

By addressing these areas, future research will not only enhance the existing methods but also pave the way for groundbreaking advancements in computer vision, ultimately contributing to safer, smarter, and more efficient technologies in various domains.

References

[1] S. Wang, J. Zheng, H. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3538–3548, Sep. 2013.

[2] X. Fu, Y. Liao, D. Zeng, Y. Huang, X. Zhang, and X. Ding, "A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4965–4977, Dec 2015.

[3] X. Guo, Y. Li, and H. Ling, "Lime: Low-light image enhancement via illumination map estimation," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 982–993, Feb 2017.

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

[5] Y. Ren, Z. Ying, T. H. Li, and G. Li, "Lecarm: Low-light image enhancement using the camera response model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 4, pp. 968–981, 2019.

[6] S. Hao, X. Han, Y. Guo, X. Xu, and M. Wang, "Low-light image enhancement with semi-decoupled decomposition," *IEEE Transactions on Multimedia*, pp. 1–1, 2020.

[7] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[8] L. Ma, R. Liu, J. Zhang, X. Fan, and Z. Luo, "Learning deep context-sensitive decomposition for low-light image enhancement," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2021.

[9] R. C. Gonzalez and R. E. Woods, *Digital image processing*. Pearson Education India, 2009.

[10] E. H. Land and J. J. McCann, "Lightness and retinex theory," *Journal of Optical Society of America*, vol. 61, no. 1, pp. 1–11, Jan 1971.

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

[12] ——, "Recent advances in retinex theory," in *Central and peripheral mechanisms of colour vision*. Springer, 1985, pp. 5–17.

[13] A. S. Parihar and K. Singh, "A study on retinex based method for image enhancement," in *2nd International Conference on Inventive Systems and Control*, Jan 2018, pp. 619–624.

[14] D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Transactions on Image Processing*, vol. 6, no. 3, pp. 451–462, March 1997.

[15] Z.-u. Rahman, D. J. Jobson, and G. A. Woodell, "Multi-scale retinex for color image enhancement," in *IEEE International Conference on Image Processing*, vol. 3. IEEE, 1996, pp. 1003–1006.

[16] D. J. Jobson, Z. 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*, vol. 6, no. 7, pp. 965–976, July 1997.

[17] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, and J. Paisley, "A fusion-based enhancing method for weakly illuminated images," *Signal Processing*, vol. 129, pp. 82–96, 2016.

[18] W. Wang and X. Yuan, "Recent advances in image dehazing," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, pp. 410–436, 01 2017.

[19] R. T. Tan, "Visibility in bad weather from a single image," in *Computer Vision and Pattern Recognition*, 06 2008, pp. 1–8.

[20] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341–2353, 2011.

[21] J.-P. Tarel and N. Hautière, "Fast visibility restoration from a single color or gray level image," 09 2009, pp. 2201–2208.

[22] Z. Rong and W. L. Jun, "Improved wavelet transform algorithm for single image dehazing," *Optik*, vol. 125, no. 13, pp. 3064 – 3066, 2014.

[23] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Instant dehazing of images using polarization," *Computer Vision and Pattern Recognition*, vol. 1, pp. I–I, 2001.

[24] S. G. Narasimhan and S. K. Nayar, "Chromatic framework for vision in bad weather," *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)*, vol. 1, pp. 598–605 vol.1, 2000.

[25] S. Shwartz, E. Namer, and Y. Schechner, "Blind haze separation," vol. 2, 02 2006, pp. 1984 – 1991.

[26] H. Koschmieder, *Luftlicht und sichtweite Naturwissenschaften*, 1938, vol. 26(32), pp. 521–528.

[27] E. McCartney, *Optics of the atmosphere: scattering by molecules and particles*. Wiley, New York, 1976.

[28] S. Nayar and S. Narasimhan, "Vision in bad weather," vol. 2, 02 1999, pp. 820 – 827 vol.2.

[29] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, 2001, pp. I–I.

[30] S. Agarwal and D. Roth, "Learning a sparse representation for object detection," in *Computer Vision—ECCV 2002: 7th European Conference on Computer Vision Copenhagen, Denmark, May 28–31, 2002 Proceedings, Part IV 7*. Springer, 2002, pp. 113–127.

[31] A. Torralba, K. P. Murphy, and W. T. Freeman, "Sharing visual features for multiclass and multiview object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 5, pp. 854–869, 2007.

[32] N. J. Butko and J. R. Movellan, "Optimal scanning for faster object detection," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 2751–2758.

[33] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, 2010.

[34] J. R. Uijlings, K. E. Van De Sande, T. Gevers, and A. W. Smeulders, "Selective search for object recognition," *International journal of computer vision*, vol. 104, pp. 154–171, 2013.

[35] C. L. Zitnick and P. Dollár, "Edge boxes: Locating object proposals from edges," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*. Springer, 2014, pp. 391–405.

[36] P. O. O Pinheiro, R. Collobert, and P. Dollár, "Learning to segment object candidates," *Advances in neural information processing systems*, vol. 28, 2015.

[37] P. O. Pinheiro, T.-Y. Lin, R. Collobert, and P. Dollár, "Learning to refine object segments," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 75–91.

[38] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.

[39] R. Girshick, "Fast r-cnn," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1440–1448.

[40] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[41] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[42] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[43] Z. Cai, Q. Fan, R. S. Feris, and N. Vasconcelos, "A unified multi-scale deep convolutional neural network for fast object detection," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*. Springer, 2016, pp. 354–370.

[44] T. Kong, A. Yao, Y. Chen, and F. Sun, "Hypernet: Towards accurate region proposal generation and joint object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 845–853.

[45] A. Shrivastava, R. Sukthankar, J. Malik, and A. Gupta, "Beyond skip connections: Top-down modulation for object detection," *arXiv preprint arXiv:1612.06851*, 2016.

[46] S. Bell, C. L. Zitnick, K. Bala, and R. Girshick, "Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2874–2883.

[47] S. Gidaris and N. Komodakis, "Object detection via a multi-region and semantic segmentation-aware cnn model," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1134–1142.

[48] A. Shrivastava and A. Gupta, "Contextual priming and feedback for faster r-cnn," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 330–348.

[49] X. Zeng, W. Ouyang, B. Yang, J. Yan, and X. Wang, "Gated bi-directional cnn for object detection," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII 14*. Springer, 2016, pp. 354–369.

[50] M. Najibi, M. Rastegari, and L. S. Davis, "G-cnn: an iterative grid based object detector," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2369–2377.

[51] A. Shrivastava, A. Gupta, and R. Girshick, "Training region-based object detectors with online hard example mining," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 761–769.

[52] C. Wang, C. He, and M. Xu, "Fast exposure fusion of detail enhancement for brightest and darkest regions," *The Visual Computer*, vol. 37, no. 5, pp. 1233–1243, 2021.

[53] D. Yoo, S. Park, J.-Y. Lee, A. S. Paek, and I. So Kweon, "Attentionnet: Aggregating weak directions for accurate object detection," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 2659–2667.

[54] J. Dai, Y. Li, K. He, and J. Sun, "R-fcn: Object detection via region-based fully convolutional networks," *Advances in neural information processing systems*, vol. 29, 2016.

[55] H. Lee, S. Eum, and H. Kwon, "Me r-cnn: Multi-expert r-cnn for object detection," *IEEE Transactions on Image Processing*, vol. 29, pp. 1030–1044, 2019.

[56] Y. Zhu, C. Zhao, J. Wang, X. Zhao, Y. Wu, and H. Lu, "Couplenet: Coupling global structure with local parts for object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4126–4134.

[57] Z. Cai and N. Vasconcelos, "Cascade r-cnn: high quality object detection and instance segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 5, pp. 1483–1498, 2019.

[58] K.-W. Cheng, Y.-T. Chen, and W.-H. Fang, "Improved object detection with iterative localization refinement in convolutional neural networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 9, pp. 2261–2275, 2017.

[59] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[60] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263–7271.

[61] ——, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[62] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[63] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," in *International Conference on Learning Representations*, 2014.

[64] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 21–37.

[65] T. Kong, F. Sun, A. Yao, H. Liu, M. Lu, and Y. Chen, "Ron: Reverse connection with objectness prior networks for object detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5244–5252.

[66] Z. Shen, Z. Liu, J. Li, Y.-G. Jiang, Y. Chen, and X. Xue, "Dsod: Learning deeply supervised object detectors from scratch," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 1937–1945.

[67] R. Zhu, S. Zhang, X. Wang, L. Wen, H. Shi, L. Bo, and T. Mei, "Scratchdet: Training single-shot object detectors from scratch," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 2263–2272.

[68] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[69] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision*, vol. 88, no. 2, pp. 303–338, 2010.

[70] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, 2020.

[71] P. Zhou, B. Ni, C. Geng, J. Hu, and Y. Xu, "Scale-transferrable object detection," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 528–537.

[72] Z. Zhang, S. Qiao, C. Xie, W. Shen, B. Wang, and A. L. Yuille, "Single-shot object detection with enriched semantics," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5813–5821.

[73] T. Kong, F. Sun, C. Tan, H. Liu, and W. Huang, "Deep feature pyramid reconfiguration for object detection," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 169–185.

[74] S. Liu, D. Huang *et al.*, "Receptive field block net for accurate and fast object detection," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 385–400.

[75] S.-W. Kim, H.-K. Kook, J.-Y. Sun, M.-C. Kang, and S.-J. Ko, "Parallel feature pyramid network for object detection," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 234–250.

[76] K. Singh and A. S. Parihar, "Variational optimization based single image dehazing," *Journal of Visual Communication and Image Representation*, vol. 79, p. 103241, 2021.

[77] P. Joshi and S. Prakash, "Image enhancement with naturalness preservation," *The Visual Computer*, vol. 36, no. 1, pp. 71–83, 2020.

[78] B. A. Karr, K. Debattista, and A. G. Chalmers, "Optical effects on hdr calibration via a multiple exposure noise-based workflow," *The Visual Computer*, vol. 37, no. 5, pp. 895–910, 2021.

[79] A. S. Parihar and O. P. Verma, "Contrast enhancement using entropy-based dynamic sub-histogram equalisation," *IET Image Processing*, vol. 10, no. 11, pp. 799–808, 2016.

[80] A. S. Parihar, O. P. Verma, and C. Khanna, "Fuzzy-contextual contrast enhancement," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1810–1819, April 2017.

[81] M. Abdullah-Al-Wadud, M. H. Kabir, M. A. Akber Dewan, and O. Chae, "A dynamic histogram equalization for image contrast enhancement," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 2, pp. 593–600, May 2007.

[82] T. Celik and T. Tjahjadi, "Contextual and variational contrast enhancement," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3431–3441, Dec 2011.

[83] T. Celik, "Spatial entropy-based global and local image contrast enhancement," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5298–5308, Dec 2014.

[84] C. Lee, C. Lee, and C. Kim, "Contrast enhancement based on layered difference representation of 2d histograms," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5372–5384, Dec 2013.

[85] G. Yadav, S. Maheshwari, and A. Agarwal, "Contrast limited adaptive histogram equalization based enhancement for real time video system," in *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2014, pp. 2392–2397.

[86] A. S. Parihar, "Fuzzy adaptive gamma correction for contrast enhancement," in *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, 2017, pp. 625–629.

[87] Y.-S. Chiu, F.-C. Cheng, and S.-C. Huang, "Efficient contrast enhancement using adaptive gamma correction and cumulative intensity distribution," in *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2946–2950.

[88] 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*, vol. 15, no. 7, pp. 1410–1423, 2021.

[89] C. Wang, H. Zhang, and L. Liu, "Total generalized variation-based retinex image decomposition," *The Visual Computer*, vol. 37, no. 1, pp. 77–93, 2021.

[90] A. S. Parihar, K. Singh, H. Rohilla, G. Asnani, and H. Kour, "A comprehensive analysis of fusion-based image enhancement techniques," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2020, pp. 823–828.

[91] X. Fu, D. Zeng, Y. Huang, X. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2016, pp. 2782–2790.

[92] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, and J. Ma, "Msr-net:low-light image enhancement using deep convolutional network," 2017.

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

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

[95] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *British Machine Vision Conference 2018, Newcastle, UK, September 3-6, 2018*. BMVA Press, 2018, p. 155.

[96] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 1632–1640.

[97] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "Enlightengan: Deep light enhancement without paired supervision," *arXiv preprint arXiv:1906.06972*, 2019.

[98] N. YU, J. LI, and Z. HUA, "Fla-net: multi-stage modular network for low-light image enhancement," *The Visual Computer*, pp. 1–20, 2022.

[99] M. Li, L. Zhao, D. Zhou, R. Nie, Y. Liu, and Y. Wei, "Aems: an attention enhancement network of modules stacking for lowlight image enhancement," *The Visual Computer*, pp. 1–17, 2021.

[100] X. Yu, H. Li, and H. Yang, "Two-stage image decomposition and color regulator for low-light image enhancement," *The Visual Computer*, pp. 1–11, 2022.

[101] X. Song, J. Huang, J. Cao, and D. Song, "Feature spatial pyramid network for low-light image enhancement," *The Visual Computer*, pp. 1–11, 2022.

[102] S. Yu and H. Zhu, "Low-illumination image enhancement algorithm based on a physical lighting model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 1, pp. 28–37, 2019.

[103] S. Ghosh, R. G. Gavaskar, D. Panda, and K. N. Chaudhury, "Fast scale-adaptive bilateral texture smoothing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 7, pp. 2015–2026, 2020.

[104] Z. Rahman, Y.-F. Pu, M. Aamir, and S. Wali, "Structure revealing of low-light images using wavelet transform based on fractional-order denoising and multiscale decomposition," *The Visual Computer*, vol. 37, no. 5, pp. 865–880, 2021.

[105] H.-C. Tsai, H.-J. Lin, and J.-J. Leou, "Multiexposure image fusion using intensity enhancement and detail extraction," *Journal of Visual Communication and Image Representation*, vol. 33, pp. 165–178, 2015.

[106] J. Xu, Y. Hou, D. Ren, L. Liu, F. Zhu, M. Yu, H. Wang, and L. Shao, "Star: A structure and texture aware retinex model," *IEEE Transactions on Image Processing*, vol. 29, pp. 5022–5037, 2020.

[107] S. Wang, J. Yang, D. Chen, J. Huang, Y. Zhang, W. Liu, Z. Zheng, and Y. Li, "Litecortexnet: toward efficient object detection at night," *The Visual Computer*, vol. 38, no. 9-10, pp. 3073–3085, 2022.

[108] Y. Sasagawa and H. Nagahara, "Yolo in the dark-domain adaptation method for merging multiple models," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16*. Springer, 2020, pp. 345–359.

[109] Y. Wu, H. Guo, C. Chakraborty, M. Khosravi, S. Berretti, and S. Wan, "Edge computing driven low-light image dynamic enhancement for object detection," *IEEE Transactions on Network Science and Engineering*, pp. 1–1, 2022.

[110] Y.-T. Huang, Y.-T. Peng, and W.-H. Liao, "Enhancing object detection in the dark using u-net based restoration module," in *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2019, pp. 1–9.

[111] Y. Hong, K. Wei, L. Chen, and Y. Fu, "Crafting object detection in very low light," in *BMVC*, vol. 1, no. 2, 2021, p. 3.

[112] K. Ren, Q. Tao, and H. Han, "A lightweight object detection network in low-light conditions based on depthwise separable pyramid network and attention mechanism on embedded platforms," *Journal of the Franklin Institute*, vol. 360, no. 6, pp. 4427–4455, 2023.

[113] Z. Yuan, J. Zeng, Z. Wei, L. Jin, S. Zhao, X. Liu, Y. Zhang, and G. Zhou, "Clahe-based low-light image enhancement for robust object detection in overhead power transmission system," *IEEE Transactions on Power Delivery*, vol. 38, no. 3, pp. 2240–2243, 2023.

[114] D. Singh and V. Chahar, "Comprehensive survey on haze removal techniques," *Multimedia Tools and Applications*, p. 9595–9620, 12 2017.

[115] G. Harish Babu and N. Venkatram, "A survey on analysis and implementation of state-of-the-art haze removal techniques," *Journal of Visual Communication and Image Representation*, vol. 72, p. 102912, 2020.

[116] A. Parihar, Y. Gupta, Y. Singodia, V. Singh, and K. Singh, "A comparative study of image dehazing algorithms," in *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, 06 2020, pp. 766–771.

[117] A. S. Parihar and G. Gupta, "A study on dark channel prior based image enhancement techniques," in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2020, pp. 1–7.

[118] S. Nayar and S. Narasimhan, "Vision in bad weather," in *Proceedings of IEEE Conference on Computer Vision*, vol. 2, 02 1999, pp. 820 – 827 vol.2.

[119] S. G. Narasimhan and S. K. Nayar, "Chromatic framework for vision in bad weather," *Computer Vision and Pattern Recognition*, vol. 1, pp. 598–605 vol.1, 2000.

[120] S. Shwartz, E. Namer, and Y. Schechner, "Blind haze separation," in *Computer Vision and Pattern Recognition*, vol. 2, 02 2006, pp. 1984 – 1991.

[121] R. Fattal, "Single image dehazing," *ACM Trans. Graph.*, vol. 27, no. 3, p. 1–9, Aug. 2008.

[122] ——, "Dehazing using color-lines," *ACM Trans. Graph.*, vol. 34, no. 1, Dec. 2015.

[123] M. Sulami, I. Geltzer, R. Fattal, and M. Werman, "Automatic recovery of the atmospheric light in hazy images," in *IEEE International Conference on Computational Photography (ICCP)*, 2014, pp. 1–11.

[124] K. B. Gibson and T. Q. Nguyen, "On the effectiveness of the dark channel prior for single image dehazing by approximating with minimum volume ellipsoids," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 1253–1256.

[125] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE transactions on image processing*, vol. 24, no. 11, pp. 3522–3533, 2015.

[126] D. Singh, V. Kumar, and M. Kaur, "Single image dehazing using gradient channel prior," *Applied Intelligence*, vol. 49, no. 12, pp. 4276–4293, 2019.

[127] M. Kaur, D. Singh, V. Kumar, and K. Sun, "Color image dehazing using gradient channel prior and guided l0 filter," *Information Sciences*, vol. 521, pp. 326–342, 2020.

[128] D. Nair and P. Sankaran, "Color image dehazing using surround filter and dark channel prior," *Journal of Visual Communication and Image Representation*, vol. 50, pp. 9–15, 2018.

[129] C. O. Ancuti, C. Ancuti, and P. Bekaert, "Effective single image dehazing by fusion," in *2010 IEEE International Conference on Image Processing*, 2010, pp. 3541–3544.

[130] C. O. Ancuti and C. Ancuti, "Single image dehazing by multi-scale fusion," *IEEE Transactions on Image Processing*, vol. 22, no. 8, pp. 3271–3282, 2013.

[131] Z. Zhu, H. Wei, G. Hu, Y. Li, G. Qi, and N. Mazur, "A novel fast single image dehazing algorithm based on artificial multiexposure image fusion," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–23, 2021.

[132] A. Galdran, J. Vazquez-Corral, D. Pardo, and M. Bertalmío, "Fusion-based variational image dehazing," *IEEE Signal Processing Letters*, vol. 24, no. 2, pp. 151–155, 2017.

[133] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.

[134] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 4780–4788.

[135] D. Chen, M. He, Q. Fan, J. Liao, L. Zhang, D. Hou, L. Yuan, and G. Hua, "Gated context aggregation network for image dehazing and deraining," *CoRR*, vol. abs/1811.08747, 2018.

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

[137] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 11 2016.

[138] X. Liu, Y. Ma, Z. Shi, and J. Chen, "Griddehazenet: Attention-based multi-scale network for image dehazing," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 7314–7323.

[139] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "Ffa-net: Feature fusion attention network for single image dehazing," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, pp. 11 908–11 915, Apr. 2020.

[140] H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, and M.-H. Yang, "Multi-scale boosted dehazing network with dense feature fusion," 2020.

[141] Y. Romano and M. Elad, "Boosting of image denoising algorithms," *SIAM Journal on Imaging Sciences*, vol. 8, pp. 1187–1219, 06 2015.

[142] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015.

[143] D. Engin, A. Genç, and H. K. Ekenel, "Cycle-dehaze: Enhanced cyclegan for single image dehazing," *CoRR*, vol. abs/1805.05308, 2018.

[144] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2242–2251, 2017.

[145] R. Malav, A. Kim, S. R. Sahoo, and G. Pandey, "Dhsgan: An end to end dehazing network for fog and smoke," in *Computer Vision – ACCV 2018*, C. Jawahar,

H. Li, G. Mori, and K. Schindler, Eds. Cham: Springer International Publishing, 2019, pp. 593–608.

[146] M. S. Akhtar, A. Ali, and S. S. Chaudhuri, "Mobile-unet gan: A single-image dehazing model," *Signal Image Video Process*, pp. 1–9, 2023.

[147] V. A. Sindagi, P. Oza, R. Yasarla, and V. M. Patel, "Prior-based domain adaptive object detection for hazy and rainy conditions," in *European Conference on Computer Vision*. Springer, 2020, pp. 763–780.

[148] B. Kumar, A. Mishra, A. Talesara, S. Kumar, S. Dey, Vyas, and R. Vyas, "Object detection for autonomous vehicle in hazy environment using optimized deep learning techniques," in *Proceedings of the 2022 Fourteenth International Conference on Contemporary Computing*, 2022, pp. 242–249.

[149] S. Kalwar, D. Patel, A. Aanegola, K. R. Konda, S. Garg, and K. M. Krishna, "Gdip: Gated differentiable image processing for object-detection in adverse conditions," *arXiv preprint arXiv:2209.14922*, 2022.

[150] W. Liu, G. Ren, R. Yu, S. Guo, J. Zhu, and L. Zhang, "Image-adaptive yolo for object detection in adverse weather conditions," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, 2022, pp. 1792–1800.

[151] B. Li, Y. Hua, and M. Lu, "Object detection in hazy environment enhanced by preprocessing image dataset with synthetic haze," in *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE, 2020, pp. 1618–1623.

[152] R. Tanwar, S. Verma, M. Kumar *et al.*, "Object detection using image dehazing: A journey of visual improvement," in *International Conference on Intelligent Technologies*. IEEE, 2022, pp. 1–8.

[153] Q. Qin, K. Chang, M. Huang, and G. Li, "Denet: Detection-driven enhancement network for object detection under adverse weather conditions," in *Asian Conf. on Computer Vision*, December 2022, pp. 2813–2829.

[154] C. Li, H. Zhou, Y. Liu, C. Yang, Y. Xie, Z. Li, and L. Zhu, "Detection-friendly dehazing: object detection in real-world hazy scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2023.

[155] K. G. Dhal, A. Das, N. Ghoshal, and S. Das, "Variance based brightness preserved dynamic histogram equalization for image contrast enhancement," *Pattern Recognition and Image Analysis*, vol. 28, no. 4, pp. 747–757, 2018.

[156] P. Shanmugavadivu, K. Balasubramanian, and A. Muruganandam, "Particle swarm optimized bi-histogram equalization for contrast enhancement and brightness preservation of images," *The Visual Computer*, vol. 30, no. 4, pp. 387–399, 2014.

[157] K. Xu, X. Yang, B. Yin, and R. W. Lau, "Learning to restore low-light images via decomposition-and-enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[158] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, and J. Jia, "Underexposed photo enhancement using deep illumination estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[159] W. Ren, S. Liu, L. Ma, Q. Xu, X. Xu, X. Cao, J. Du, and M.-H. Yang, "Low-light image enhancement via a deep hybrid network," *IEEE Transactions on Image Processing*, vol. 28, no. 9, pp. 4364–4375, 2019.

[160] Q. Zhang, G. Yuan, C. Xiao, L. Zhu, and W.-S. Zheng, "High-quality exposure correction of underexposed photos," in *Proceedings of the 26th ACM International Conference on Multimedia*, ser. MM '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 582–590.

[161] M. Gharbi, J. Chen, J. T. Barron, S. W. Hasinoff, and F. Durand, "Deep bilateral learning for real-time image enhancement," *ACM Trans. Graph.*, vol. 36, no. 4, jul 2017.

[162] Q. Zhang, Y. Nie, and W.-S. Zheng, "Dual illumination estimation for robust exposure correction," *Computer Graphics Forum*, vol. 38, no. 7, pp. 243–252, 2019.

[163] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397–1409, June 2013.

[164] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Advances in Neural Information*

Processing Systems, J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Q. Weinberger, Eds., vol. 24. Curran Associates, Inc., 2011, p. 612–620.

[165] Y. P. Loh and C. S. Chan, "Getting to know low-light images with the exclusively dark dataset," *Computer Vision and Image Understanding*, vol. 178, pp. 30 – 42, 2019.

[166] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "Enlightengan: Deep light enhancement without paired supervision," *arXiv preprint arXiv:1906.06972*, 2019.

[167] G. Woodell, "Retinex image processing," accessed:2019-08-13.

[168] V. Bychkovsky, S. Paris, E. Chan, and F. Durand, "Learning photographic global tonal adjustment with a database of input / output image pairs," in *The Twenty-Fourth IEEE Conference on Computer Vision and Pattern Recognition*, 2011.

[169] T. Celik, "Two-dimensional histogram equalization and contrast enhancement," *Pattern Recognition*, vol. 45, no. 10, pp. 3810 – 3824, 2012.

[170] K. Gu, W. Lin, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "No-reference quality metric of contrast-distorted images based on information maximization," *IEEE Transactions on Cybernetics*, vol. 47, no. 12, pp. 4559–4565, 2017.

[171] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[172] D. A. Lind, W. G. Marchal, and S. A. Wathen, *Statistical Techniques in Business & Economics*. 13th ed. New York, NY, USA: McGraw-Hill, 2014.

[173] S. Glen, "One tailed test: Definition & examples," 2024, [Online; accessed 22-June-2024]. [Online]. Available: https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/one-tailed-test/

[174] Wikipedia contributors, "P-value," 2023, [Online; accessed 22-June-2024]. [Online]. Available: https://en.wikipedia.org/wiki/P-value

[175] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 286–301.

[176] A. Ignatov, N. Kobyshev, R. Timofte, K. Vanhoey, and L. Van Gool, "Dslr-quality photos on mobile devices with deep convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 3277–3285.

[177] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision*. Springer, 2016, pp. 694–711.

[178] 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*, vol. 13, no. 4, pp. 600–612, 2004.

[179] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3345–3356, 2015.

[180] C. Lee, C. Lee, and C.-S. Kim, "Contrast enhancement based on layered difference representation," in *2012 19th IEEE International Conference on Image Processing*, 2012, pp. 965–968.

[181] A. Mittal, A. K. Moorthy, and A. C. Bovik, "Blind/referenceless image spatial quality evaluator," in *2011 conference record of the forty fifth asilomar conference on signals, systems and computers (ASILOMAR)*. IEEE, 2011, pp. 723–727.

[182] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE signal processing letters*, vol. 9, no. 3, pp. 81–84, 2002.

[183] D. A. Lind, W. G. Marchal, and S. A. Wathen, *Statistical Techniques in Business & Economics*. 13th ed. New York, NY, USA: McGraw-Hill, 2014.

[184] S. Ghosh, R. G. Gavaskar, D. Panda, and K. N. Chaudhury, "Fast scale-adaptive bilateral texture smoothing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 7, pp. 2015–2026, 2020.

[185] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, 2019.

[186] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer, "O-haze: A dehazing benchmark with real hazy and haze-free outdoor images," in *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018, pp. 867–8678.

[187] Y. Zhang, L. Ding, and G. Sharma, "Hazerd: An outdoor scene dataset and benchmark for single image dehazing," in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 3205–3209.

[188] K. Singh and A. S. Parihar, "A comparative analysis of illumination estimation based image enhancement techniques," in *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, 2020, pp. 1–5.

[189] A. Galdran, A. Alvarez-Gila, A. Bria, J. Vazquez-Corral, and M. Bertalmío, "On the duality between retinex and image dehazing," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018, pp. 8212–8221.

[190] Zhou 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*, vol. 13, no. 4, pp. 600–612, 2004.

[191] X. Li, C. Lv, W. Wang, G. Li, L. Yang, and J. Yang, "Generalized focal loss: Towards efficient representation learning for dense object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 3, pp. 3139–3153, 2023.

[192] K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Xu, Z. Zhang, D. Cheng, C. Zhu, T. Cheng, Q. Zhao, B. Li, X. Lu, R. Zhu, Y. Wu, J. Dai, J. Wang, J. Shi, W. Ouyang, C. C. Loy, and D. Lin, "MMDetection: Open mmlab detection toolbox and benchmark," *arXiv preprint arXiv:1906.07155*, 2019.

[193] Z. Yang, S. Liu, H. Hu, L. Wang, and S. Lin, "Reppoints: Point set representation for object detection," in *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2019.

[194] T. Kong, F. Sun, H. Liu, Y. Jiang, L. Li, and J. Shi, "Foveabox: Beyound anchor-based object detection," *IEEE Transactions on Image Processing*, vol. 29, pp. 7389–7398, 2020.

[195] J. Wang, K. Chen, S. Yang, C. C. Loy, and D. Lin, "Region proposal by guided anchoring," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

[196] C. Zhu, Y. He, and M. Savvides, "Feature selective anchor-free module for single-shot object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 840–849.

[197] S. Zhang, C. Chi, Y. Yao, Z. Lei, and S. Z. Li, "Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 9759–9768.

[198] Z. Tian, C. Shen, H. Chen, and T. He, "Fcos: A simple and strong anchor-free object detector," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 4, pp. 1922–1933, 2022.

[199] S. Hao, Z. Wang, and F. Sun, "LEDet: A Single-Shot Real-Time Object Detector Based on Low-Light Image Enhancement," *The Computer Journal*, vol. 64, no. 7, pp. 1028–1038, 05 2021.

[200] Z. Cui, G.-J. Qi, L. Gu, S. You, Z. Zhang, and T. Harada, "Multitask aet with orthogonal tangent regularity for dark object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021, pp. 2553–2562.

[201] W. Liu, G. Ren, R. Yu, S. Guo, J. Zhu, and L. Zhang, "Image-adaptive yolo for object detection in adverse weather conditions," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 2, 2022, pp. 1792–1800.

[202] Z. Jiang, D. Shi, and S. Zhang, "Frse-net: low-illumination object detection network based on feature representation refinement and semantic-aware enhancement," *The Visual Computer*, pp. 1–15, 2023.

[203] V. Vs, V. Gupta, P. Oza, V. A. Sindagi, and V. M. Patel, "Mega-cda: Memory guided attention for category-aware unsupervised domain adaptive object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 4516–4526.

[204] Z. Zhang, L. Zhao, Y. Liu, S. Zhang, and J. Yang, "Unified density-aware image dehazing and object detection in real-world hazy scenes," in *Asian Conf. on Computer Vision*, 2020.

[205] C. Sakaridis, D. Dai, and L. Van Gool, "Semantic foggy scene understanding with synthetic data," *Int. J. Comput. Vis.*, vol. 126, no. 9, pp. 973–992, Sep 2018.

[206] K. Singh and A. S. Parihar, "Illumination estimation for nature preserving low-light image enhancement," *The Visual Computer*, pp. 1–16, 2023.

[207] C. Godard, O. M. Aodha, M. Firman, and G. Brostow, "Digging into self-supervised monocular depth estimation," in *IEEE/CVF International Conference on Computer Vision*, 2019, pp. 3827–3837.

[208] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *IEEE International Conference on Computer Vision*, 2017, pp. 2242–2251.

[209] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. Van Gool, "Domain adaptive faster r-cnn for object detection in the wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3339–3348.

[210] K. Saito, Y. Ushiku, T. Harada, and K. Saenko, "Strong-weak distribution alignment for adaptive object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 6956–6965.

[211] Z. Shen, H. Maheshwari, W. Yao, and M. Savvides, "Scl: Towards accurate domain adaptive object detection via gradient detach based stacked complementary losses," *arXiv preprint arXiv:1911.02559*, 2019.

[212] Y. Chen, H. Wang, W. Li, C. Sakaridis, D. Dai, and L. Van Gool, "Scale-aware domain adaptive faster r-cnn," *Int. J. Comput. Vis.*, vol. 129, no. 7, pp. 2223–2243, 2021.

[213] X. Yang, M. B. Mi, Y. Yuan, X. Wang, and R. T. Tan, "Object detection in foggy scenes by embedding depth and reconstruction into domain adaptation," in *Asian Conf. on Computer Vision*, 2022, pp. 1093–1108.

[214] D. Guan, J. Huang, A. Xiao, S. Lu, and Y. Cao, "Uncertainty-aware unsupervised domain adaptation in object detection," *IEEE Trans. Multimedia.*, vol. 24, pp. 2502–2514, 2021.

[215] D.-K. Nguyen, W.-L. Tseng, and H.-H. Shuai, "Domain-adaptive object detection via uncertainty-aware distribution alignment," in *Proceedings of the 28th ACM international conference on multimedia*, 2020, pp. 2499–2507.

[216] J. Huang, D. Guan, A. Xiao, S. Lu, and L. Shao, "Category contrast for unsupervised domain adaptation in visual tasks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 1203–1214.

[217] W. Zhou, D. Du, L. Zhang, T. Luo, and Y. Wu, "Multi-granularity alignment domain adaptation for object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 9581–9590.

Author Biography

Kavinder Singh
Assistant Professor,
Department of Computer Science & Engineering
Delhi Technological University, Delhi, India
Email: Kavinder85@gmail.com; kavinder@dtu.ac.in

Kavinder Singh received his B.Tech. degree in Computer Science & Engineering in 2016 from A. K. Technical University, Lucknow, India, M.Tech. degree in Information systems from Delhi Technological University, New Delhi, India in 2018. He is pursuing his PhD from Delhi Technological University, New Delhi, India. He joined Department of Computer Science & Engineering at Delhi Technological University as Assistant Professor in 2020. Currently, he is Assistant Professor in the Department of Computer Science & Engineering at Delhi Technological University, New Delhi, India. His research interest includes Image Processing, Pattern Recognition, Object recognition and Computer Vision.