An Improved Deep Learning Based Low Light Image Enhancement Model

By

Maj Arsal Bilal

(**Registration No:** 00000431946)

Department of Electrical Engineering

A thesis submitted in partial fulfilment of the requirements for the degree of Masters

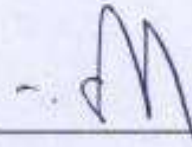in Electrical Engineering (MSEE)

In

Military College of Signals (MCS)

National University of Sciences and Technology (NUST)
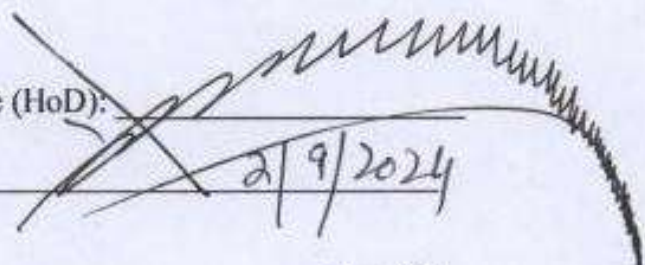
Islamabad, Pakistan

(September 2024)

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS thesis written by Mr **Major Arsal Bilal** Registration No. **00000431946** of **Military College of Signals** has been vetted by undersigned, found complete in all respect as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial, fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the student have been also incorporated in the said thesis.

Signature: _____

Name of Supervisor: **Professor Abdul Ghafoor, PhD**

Date: _____ 2| 9| 2024 _____

Signature (HoD): _____

Date: _____ 2| 9/2024 _____

Signature (Dean): _____

Date: _____ 3| 9/24 _____

Brig
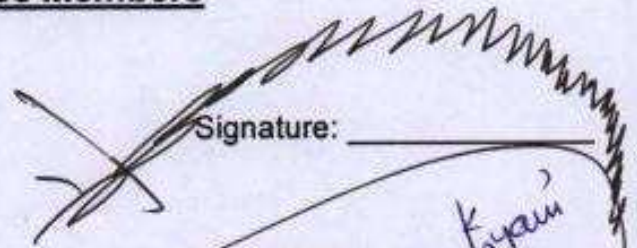Dean, MCS (NUST)
(Asif Masood, Phd)

ii

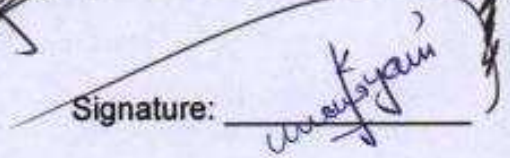# NATIONAL UNIVERSITY OF SCIENCES & TECHNOLOGY

MASTER THESIS WORK

We hereby recommend that the thesis prepared under our supervision by **Maj Muhammad Ishtiaq, MSEE-28 Course** Regn No **00000431956** Titled: **"Integration of Multi-Level Attention in U-Net with Generative Adversarial Network for Enhanced Underwater Visibility"** be accepted in partial fulfillment of the requirements for the award of **MS Electrical Engineering** degree.

Examination Committee Members

1. **Brig Adil Masood Siddiqui, PhD**   Signature: _____
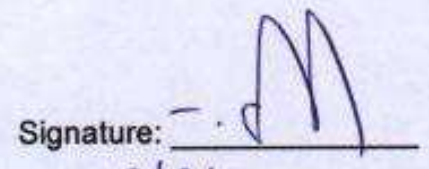
2. **Asst Prof Maemoona Farooq Kayani, PhD**   Signature: _____

Co-Supervisor: **Assoc Prof Muhammad Mohsin Riaz, PhD**   Signature: _____

Supervisor: **Prof Abdul Ghafoor, PhD**   Signature: _____

Date: 2/9/2024

Head of Department

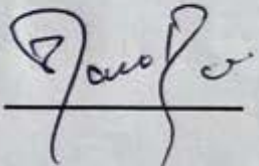2/9/2024
Date

COUNTERSIGNED

Brig
Dean, MCS (NUST)
(Asif Masood, Phd)

Dean

Date: 3/9/24

iii

CERTIFICATE OF APPROVAL

This is to certify that the research work presented in this thesis, entitled **"An Improved Deep Learning Based Low Light Image Enhancement Model"** was conducted by **Maj Arsal Bilal** under the supervision of **Prof Abdul Ghafoor, PhD**. No part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted to **Military College of Signal, National University of Sciences and Technology - Electrical Engineering Department** in partial fulfillment of the requirements for the degree of Master of Science in Field of **Electrical Engineering** Department of Electrical Engineering, National University of Sciences and Technology, Islamabad.

Student Name: **Maj Arsal Bilal**   Signature: _____
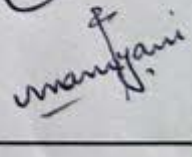
Examination Committee:
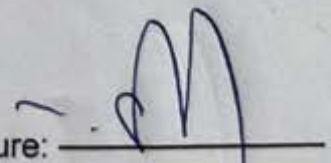
a) GEC Examiner 1: **Asst Prof Abdul Wakeel , PhD**   Signature: _____
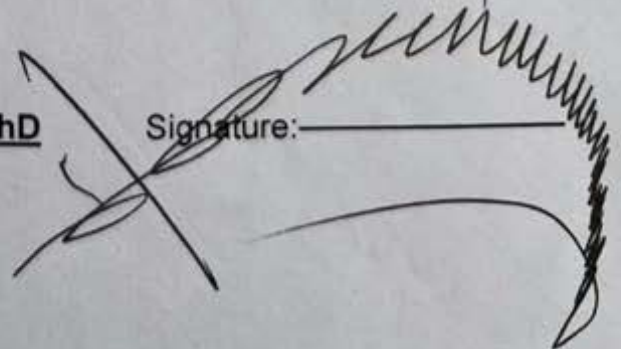
b) GEC Examiner 2: **Asst Prof Maemoona Farooq Kayani, PhD** Signature: _____

Name of Supervisor: **Prof Abdul Ghafoor, PhD**   Signature: _____
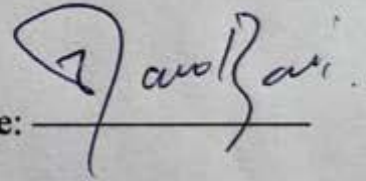
Name of Dean/HOD: **Brig Adil Masood Siddiqui, PhD**   Signature: _____

AUTHOR'S DECLARATION

I, **Major Arsal Bilal** hereby state that my MS thesis titled **"An Improved Deep Learning Based Low Light Image Enhancement Model"** is my own work and has not been submitted previously by me for taking any degree from National University of Sciences and Technology, Islamabad or anywhere else in the country/ world. At any time if my statement is found to be incorrect even after I graduate, the university has the right to withdraw my MS degree.

Student Signature: _____

Name: **Major Arsal Bilal**

Date: ___2/9/2024___
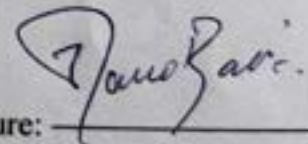
PLAGIARISM UNDERTAKING

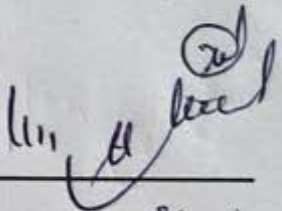I solemnly declare that research work presented in the thesis titled **"An Improved Deep Learning Based Low Light Image Enhancement Model"** is solely my research work with no significant contribution from any other person. Small contribution/ help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and National University of Sciences and Technology (NUST), Islamabad towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS degree, the University reserves the rights to withdraw/revoke my MS degree and that HEC and NUST, Islamabad has the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized thesis.

Student Signature: _____

Name: **Major Arsal Bilal**

Date: 2 | 9 | 2024

Dedication

"In the name of Allah, the most Beneficent, the most Merciful"

I dedicate this thesis to my family, friends, and teachers who supported me each step of the way, especially my parents.

This thesis is also dedicated to all the deserving children who do not have access to quality education especially young girls.

Acknowledgments

Glory be to Allah (S.W.A), the Creator, the Sustainer of the Universe. Who only has the power to honour whom He please, and to abase whom He please. Verily no one can do anything without His will. From the day, I came to NUST till the day of my departure, He was the only one Who blessed me and opened ways for me, and showed me the path of success. Their is nothing which can payback for His bounties throughout my research period to complete it successfully.

Abstract

Low-light image enhancement (LLIE) is important for various practical uses, aiming to improve their visual quality. Deep learning advances have been the driving force behind recent development in this field. Modern techniques utilize sophisticated neural networks to boost image brightness, contrast, and reduce noise, representing the cutting edge of image processing and computer vision. A new approach An Improved Deep Learning Model (AIDLM) is introduced, designed to tackle the challenges of LLIE. AIDLM consists of three main modules: a retinex decomposition network, an enhancement attention network, and a denoising network. The retinex decomposition network uses an explicit parametric regularized Retinex model tailored to individual pixels, while the enhancement attention network enhances the reflectance and illuminance of the V-channel in the HSV color space through spatial and channel attention mechanisms added to the UNet-like architecture , thus preventing color distortion. The denoising module further cleans the enhanced RGB image by eliminating noise from the H and S channels. Comprehensive experiments showed that AIDLM , greatly surpasses current baselines in LLIE, providing higher image quality and robustness.

Keywords: computer vision, Low-light image enhancement (LLIE), spatial and channel attention, perceptual loss.

Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

Abbreviations

NUST   National University of Sciences and Technology

MCS   Military College of Signals

AIDLM An Improved Deep Learning Model

DL   deep learning

CNN   convolutional neural network

LLIE   Low-light image enhancement

GHE   Global Histogram Equalisation

AHE   Adaptive Histogram Equalisation

CLAHE Contrast Limited AHE

ViTs   Vision Transformers

Introduction

1.1 Overview

In the realm of image processing, low light image enhancement (LLIE) has evolved as a critical area of research due to its profound impact on various applications, including surveillance, medical imaging, and autonomous driving. Low visibility, low contrast, noise, and colour deviation [79, 89, 56] are just a few of the degradations that frequently occur in images taken in low-light situations. These issues can significantly impact the effectiveness of advanced vision tasks such as lane detection [62], image recognition[69], remote sensing [47], object detection [42] and semantic segmentation [50].

LLIE is a crucial technique with significant effects for many high-stakes domains and applications. Enhancing these images improves visibility and detail, which is essential for security and surveillance, enabling better identification and monitoring. In medical imaging, enhanced low light images can lead to more accurate diagnoses and improved patient outcomes. For photographers and videographers, it allows for high-quality captures without intrusive lighting, preserving natural scenes. In the realm of autonomous vehicles [68], better low light imaging

enhances safety and navigation. Additionally, fields such as astronomy and augmented reality benefit from clearer, more detailed images. To counter these challenges developing advanced techniques for LLIE is imperative, as it enhances the reliability and effectiveness of visual data across numerous critical domains.

LLIE methods are generally fall ino two types: Traditional methods [4, 12, 7, 11, 8, 37] and deep learning (DL) methods [44, 40, 75, 74, 67, 57]. Traditional methods to improve low-light images primarily focused on improving image brightness and contrast through mathematical transformations and heuristic techniques but its often produce unnatural images, noise amplification and struggle with complex scenes. Color distortion and halo artifacts are common, making images look unrealistic. Overall, these methods struggle to balance visibility enhancement with maintaining a natural appearance. Nevertheless, new developments in deep learning (DL) have brought forth more advanced, data-driven methods that take advantage of neural networks' capabilities [52, 64] to learn intricate mappings from dimly lit to brightly illuminated pictures and produce better enhancement outcomes. They often mitigate common artifacts such as color distortion and noise, resulting in more natural-looking images. However deep learning (DL) based methods often struggle with extremely dark areas, can suffer from color distortion, over smoothing and loss of details. Overall, these methods can have issues with generalization, detail preservation, and efficiency. Both approaches face challenges in balancing enhancement with natural appearance and efficiency.

1.2   Problem Statement

- Many methods struggle with balancing the recovery of illumination, sharpness and texture details, especially under severely degraded conditions

- Effectively suppressing noise and artifacts while enhancing image quality remains a sig-

nificant hurdle

- Ensuring that enhancement techniques generalize well across diverse datasets is a persistent challenge

1.3   Research Motivation

LLIE is crucial for improving visibility and detail in images captured under poor lighting conditions. This improvement is essential for various high-stakes applications, including security and surveillance, medical imaging, photography, autonomous vehicles, astronomy, and augmented reality. Despite numerous advancements in the field, current methods face significant limitations that impede their effectiveness and reliability. The need for a robust solution to these issues justifies the selection of this topic. Some of the core challenges analyzed through the literature review are discussed below that laid the foundation of this thesis.

- Illumination and Texture Recovery: Many methods struggle with balancing the recovery of illumination and texture details, especially under severely degraded conditions.

- Noise and Artifact Suppression: Effectively suppressing noise and artifacts while enhancing image quality remains a significant hurdle.

- Generalization and Adaptability: Ensuring that enhancement techniques generalize well across diverse datasets and real-world conditions is a persistent challenge.

- Computational Efficiency: High computational demands and long inference times limit the scalability and real-time applicability of advanced methods.

1.4 Research Contribution

This research proposes several contributions:

- Innovative network (AIDLM) for LLIE, incorporating three distinct modules: Retinex decomposition network, enhancement attention network (*EAInet* and *EARnet*) and denoising network (*Dnet*).

- A multi-level attention mechanism (spatial and channel attention) within the enhancement attention network (*EAInet* and *EARnet*) which can more effectively capture and emphasize important features, improving overall image quality.

- Improved loss function by incorporation perceptual loss for denoising network (*Dnet*) for better denoising.

- Numerous benchmark datasets are used in lengthy tests to demonstrate the efficiency of the model. The findings demonstrate the superiority of the suggested model in LLIE by outperforming cutting-edge techniques in both qualitative and quantitative measures.

1.5 Thesis Organization

This thesis is divided into seven chapters:

Chapter 1: This chapter includes the basic introduction, background,research motivation and research contribution.

Chapter 2: This chapter provides an overview of relevant literature, encompassing articles pertinent to the scope of this study.

Chapter 3: This chapter presents the proposed methodology.

Chapter 4: This chapter delivers the experimental environment of proposed approach.

Chapter 5: This chapter presents ablation study and Results.

Chapter 6: This chapter presents limitations, conclusion and future work.

Literature Review

Image enhancement is key for improving visual quality and extracting information. Research in this field has developed along two main approaches: traditional image processing techniques and advanced deep learning methods. Traditional techniques apply established algorithms to refine images, while deep learning utilizes neural networks to achieve superior results. This section reviews both approaches to highlight their contributions and advancements in the field.

2.1   Traditional LLIE Methods

Traditional methods for LLIE have relied on various techniques to adjust image properties and improve visual quality. These methods can be broadly categorized into gray level transformation [4, 12] , histogram equalization [7, 11] and retinex-based methods [8, 37].

2.1.1   Gray Level Transformation

Gray level transformation includes linear / non-linear transformations and piece-wise transformation. Linear / Non-linear transformations adjust pixel values using linear or non-linear functions to enhance brightness and contrast. Techniques such as logarithmic and gamma trans-

formations [38, 66] are commonly used to expand dark pixel values and compress bright ones, thereby improving overall image visibility. However, these methods can often lead to unnatural-looking images and may fail to adequately enhance complex scenes. Where as piece-wise transformation involves adjusting gray values based on predefined intervals [13], which is effective for images with localized dark or bright areas. However, it can create abrupt transitions and halo artifacts in the enhanced image.

2.1.2 Histogram Equalization

Global Histogram Equalisation (GHE) [29], Adaptive Histogram Equalisation (AHE) [14], and Contrast Limited AHE (CLAHE) [7] are examples of Histogram Equalisation (HE) [5]. By balancing the histogram of pixel intensity values throughout the entire image, GHE improves contrast, especially in photos with concentrated grey values. However, GHE often amplifies noise and can cause over-enhancement, leading to loss of image details. Whereas AHE and CLAHE methods apply HE to localized regions of the image, improving local contrast and preventing noise amplification, which is a common issue with GHE. Despite their improvements, AHE and CLAHE can still introduce artifacts and may not work well on all types of low-light images.

2.1.3 Retinex-Based Methods

Single Scale Retinex (SSR)[10], Multi-Scale Retinex (MSR)[8], and Multi-Scale Retinex with Colour Restoration (MSRCR) [9] are examples of retinex-based techniques[3, 41, 43, 66]. SSR theory separates a picture into components related to illumination and reflectance. By enhancing the reflectance and reducing the effects of illumination, SSR improves image visibility under low-light conditions. By merging enhancing outcomes at several scales and taking into account both local and global information, MSR outperforms SSR. MSR preserves colour stability while

boosting contrast and detail in images. Despite these improvements, MSR can still encounter issues with color distortion and noise. MSRCR includes a color restoration function to address the color distortion issues present in SSR and MSR. MSRCR further enhances the visual quality of low-light images by preserving color fidelity. However, it can still suffer from halo effects and over-saturation in certain areas of the image. The two main stages of retinex-based LLIE techniques are enhancement and deconstruction. These techniques, which were first presented by [3], break down a picture into its illumination and reflectance components. The illumination component depicts the amount of light that strikes an object, whereas the reflectance component records the object's inherent characteristics. [17] introduced a Variational-based Framework (VF) for retinex, assuming a smooth illumination field but lacking detailed information about the reflectance. Subsequent approaches or variations [25, 22, 26, 41, 43, 66] expanded on this, introducing the Total Variation Model (TVM) [6, 23] for retinex, which assumes spatial smoothness in illumination and piecewise continuity in reflectance. This model ensures a smooth illumination field while maintaining the continuity of the reflectance component.In the enhancement phase, various techniques are applied to enhance the visual quality of decomposed components, leveraging the regularities identified in the decomposition phase. These methods strive to produce a more realistic and visually pleasing image by addressing the limitations of previous approaches.

2.2   Deep Learning LLIE Methods

In the realm of picture improvement, deep learning has gained popularity, especially for low-light photos. Improved visibility and quality of photos captured in low light has been demonstrated using deep learning approaches [48, 53, 58, 75, 70, 77, 82, 78]. Unlike traditional methods [4, 11, 8, 37] that often lead to issues such as noise amplification, detail loss, and

color distortion, deep learning-based techniques leverage data-driven models to automatically learn features from images under normal lighting. This enables the reduction of adverse effects caused by low light, resulting in significantly enhanced image quality. supervised learning methods [48, 46, 49, 52, 65, 84, 91] and unsupervised learning methods are examples of deep learning approaches for LLIE.

2.2.1 Supervised Learning Method

In supervised learning method we have end-to-end and decomposition-based learning methods.

By training a model directly on pairs of well-lit and low-light photos, end-to-end learning techniques enable the model to learn the mapping from enhanced to low-light images in a single step. This paper [56] proposed a hybrid network combining an encoder-decoder for global content estimation and a spatially variant RNN for edge details, improving visibility and edge preservation. In this paper [63] author introduced a multi-exposure fusion network that generates two enhanced images using sub-networks and reduces noise through average fusion and refinement. This paper [45] introduces a method utilizing deep Convolutional Neural Networks (CNNs) for reconstructing High Dynamic Range (HDR) images from a single exposure image. This method aims to address limitations in existing Inverse Tone-Mapping Operators (iTMOs) by enhancing the reconstruction quality of saturated regions in HDR images. However, it faces difficulty in reconstructing scenes with large saturated regions in all color channels, leading to the inability to infer structures and details. In this paper [72] author introduced Progressive Enhancement Network (PRIEN),that utilizes recursive units and a dual-attention model to efficiently extract features from low-light images, leading to significant enhancement.

Decomposition-based learning methods [53, 58, 75] decompose the image enhancement task into multiple sub-tasks, such as denoising and contrast adjustment, and address each sub-task

separately using specialized models.Inspired by the comprehensive explanation of retinex theory [1] numerous research efforts have been dedicated to image enhancement by integrating retinex decomposition concepts with deep learning algorithms. This paper [76] introduced the KinD++ network, which improves LLIE by decomposing images into reflectance and illumination layers and using a multi-scale illumination attention module to effectively restore the reflectance. The deep learning method known as URetinex-Net was presented by [82] as a way to improve low-light photos in practical settings. By framing the retinex decomposition problem as a model regularized with implicit priors and translating the optimization steps into a neural network, the authors utilize learning-based techniques to adaptively estimate both reflectance and illumination. The proposed methodology comprises three main modules: initialization, unfolding optimization, and illumination adjustment.According to experimental results, URetinex-Net effectively preserves features and reduces noise in low-light photos. However, learning-based methods still face challenges, such as limited interpretability and loss of details in reflectance restoration. This paper [86] proposed a illumination enhancement network, aiming to get better visual quality and computational efficiency of images captured in dark scenes. This network operates by effectively separating reflectance and illumination maps, resulting in significantly enhanced visual quality and noise reduction in low-light images . Additionally, the utilization of weight sharing and unsupervised training losses bolsters the model's generalization capability, ensuring robust performance across diverse datasets . However, the method's enhancement capability may become unbalanced when dealing with images that exhibit non-uniform illumination. This can lead to challenges such as color distortion and the loss of image integrity, particularly manifesting as a halo effect around brightly lit areas. This paper [91] proposed RFLLIE method that addresses key challenges in LLIE through its innovative light-head, heavy-tail architecture, which prioritizes feature reconstruction over extraction. This approach utilizes a decomposition-guided restoration loss to enhance contrast and suppress noise, enabling RFLLIE to outperform

existing methods both quantitatively and visually in image restoration tasks. The method's focus on feature reconstruction allows for superior performance in restoring fine details and improving overall image quality. However, RFLLIE faces limitations in effectively restoring background illumination and suppressing noise comprehensively. Additionally, its generalization capacity to tasks beyond low-light image restoration remains limited, indicating a need for further development to enhance its adaptability to a broader range of imaging challenges.A new technique for improving low-light images (LLIE) using brightness-aware attention modelling and residue quantisation was presented by [88]. This approach effectively enhances images while preserving details and maintaining parameter efficiency. This method employs a multi-stage framework, emphasizing hierarchical codebook construction, feature approximation, and brightness-aware attention modeling to improve image quality and minimize artifacts in low-light scenarios. Building on the strengths of VQ-based methods like VQ-VAE and VQ-GAN, which excel in learning textures and details, the paper addresses the common issue of detail loss and network instability caused by downsampling. The proposed LLIE method incorporates a three-stage framework centered on hierarchical codebook construction, feature approximation, and detail preservation, enhanced by brightness-aware attention modeling. However, the method primarily focuses on channel-wise self-attention and lacks spatial self-attention, which may limit its ability to effectively capture and preserve spatial details across the image.

2.2.2   Unsupervised Learning Method

Unsupervised learning methods do not require paired datasets of low-light and well-lit images. Instead, they use techniques such as adversarial training to learn the enhancement process from unpaired datasets, making them more flexible and widely applicable. EnlightenGAN [70] is a notable unsupervised learning method designed to tackle the LLIE problem, circumventing the need for paired low/normal-light image data, which is often impractical to obtain. This

innovative method leverages unpaired images for training, utilizing a global-local discriminator to guide the enhancement process and a self-regularized perceptual loss fusion to regularize the training based on the input data itself.

2.3   HSV Color Space

The HSV color space, representing hue, saturation, and value, is a pivotal color model in image processing due to its alignment with human visual perception. Unlike the RGB color space, HSV is designed to reflect how humans perceive colors, making it an intuitive and effective model for various applications. The independence of these three components allows for precise adjustments of each attribute without influencing the others, offering significant flexibility in image enhancement tasks. For instance, adjusting the value component can enhance brightness without altering the hue or saturation, making HSV particularly suitable for tasks such as retinex decomposition to improve low-light images. This flexibility and perceptual alignment make HSV a preferred choice for color adjustments in graphic design, video processing, and advanced image enhancement techniques. This paper [28] introduced an improved retinex algorithm for image enhancement, which converts images from RGB to HSV to leverage the independent characteristics of hue, saturation, and value. Through value component use of the retinex algorithm and using correlation coefficients to adjust saturation adaptively, they achieved enhanced brightness and color restoration without altering the original hue. Their experimental results showed substantial improvement in maintaining color fidelity and enhancing detail, especially useful for subsequent stripe information extraction and automatic coding. Similarly, [61] addressed issues of color distortion in LLIE by also utilizing the HSV color space. Their method enhances the brightness component, adjusts saturation based on brightness changes, and sharpens edges using the Laplace algorithm. This approach effectively enhances image details,

preserves the original color consistency, and improves visual quality.

Retinex decomposition in the RGB color space [58, 76, 75] often leads to significant color deviations because it processes all three color channels simultaneously, which can disrupt the natural balance of colors. This method can enhance brightness but may distort hues and saturations, resulting in unnatural-looking images. The interdependence of RGB channels makes it challenging to preserve true color fidelity during enhancement. Applying retinex decomposition on the V channel of the HSV color space [43, 66, 71] improves the illumination while preserving original hue and saturation, thereby reducing color distortions. This method benefits from the independent nature of the HSV components, ensuring more natural enhancements. Consequently, it produces images that maintain color fidelity and visual appeal. However, low-light images often have noisy hue and saturation, unlike normal-light images. This noise can carry over into the enhanced RGB images, resulting in visual artifacts. As a result, the enhancement process can introduce unwanted noise, affecting the overall image quality. To address this issue, proposed network performs retinex decomposition and enhancement on the V channel within the HSV space. After enhancing brightness, we apply noise reduction techniques to further denoise the image. This approach ensures clearer, more accurate image enhancements with minimized noise.

2.4 Attention Mechanism

Enhancing neural networks' performance, especially in LLIE tasks, requires attention methods. The network can focus on important elements by using spatial and channel attention processes, which improves the quality of the improved images. By directing focus to the most relevant aspects, these mechanisms enhance the overall effectiveness of the image improvement process. Spatial attention emphasizes significant spatial locations in the image, allowing the model to

concentrate on essential areas, while channel attention highlights important feature channels, enhancing the critical aspects of the image. Combining these mechanisms, either sequentially or in a combined block, enables the network to adaptively focus on both spatial and channel information simultaneously. This integration in LLIE models leads to improved feature representation, adaptive enhancement, and enhanced visual quality. By dynamically adjusting its focus based on the input image, the network can achieve superior performance, resulting in clearer and more visually appealing enhanced images with better sharpness, contrast, brightness and detail preservation.

The Convolutional Block Attention Module (CBAM) [54] introduces a novel approach to boosting the representational power of CNNs. CBAM enhances network performance by sequentially generating attention maps in both channel and spatial dimensions. These maps are then applied to the input feature map through multiplication, facilitating a refined and adaptive feature extraction process. Due to its lightweight design, CBAM integrates seamlessly into any CNN architecture, offering negligible computational overhead while being end-to-end trainable alongside the base networks. The spatial attention component of CBAM focuses on identifying informative regions within the feature map by leveraging inter-spatial relationships, while the channel attention component utilizes both max-pooled and average-pooled features to provide a finer level of attention than previous methods like the Squeeze-and-Excitation (SE) network [51]. This dual-attention mechanism allows CBAM to learn what features to emphasize or suppress, significantly refining intermediate feature representations. Visualizations from these experiments reveal that CBAM effectively directs the network's focus towards target objects, underscoring its potential as a critical component in future network architectures. However, a limitation of this model is that the sequential attention mechanisms might introduce additional latency during inference.

The Cascaded Network with Multi-Level Sub-Networks (CNMS) [93] presents an innovative method that integrates both localized modules and global networks. This approach enables more comprehensive semantic feature extraction and improves spatial accuracy by leveraging a hierarchical network structure. A key innovation in CNMS is the triple attention module (TAM), which refines features by focusing on both channel and spatial dimensions, thereby preserving detailed information. This module addresses the common issue of spatial information loss in traditional attention mechanisms. Experiments on multiple underwater datasets show that CNMS effectively reduces color deviation, increases contrast, and prevents over-enhancement, outperforming state-of-the-art methods.

Generally the problems are over-enhancement, noise amplification, color distortion, high computational complexity, and loss of natural image characteristics, leading to unnatural and artifact-prone results. By introducing an improved deep learning-based low light image enhancement model, this work attempts to overcome these drawbacks. This novel approach combines deep learning techniques with well-established image enhancement theories, integrating deep neural networks and retinex theory. By leveraging these cutting-edge techniques, it achieves high-quality images with enhanced visibility, reduced artifacts, and improved color accuracy.

Methodology

3.1 Proposed Network

Proposed network, named An Improved Deep Learning Model (A<small>IDLM</small>) integrates three core components: retinex decomposition network, enhancement attention network and denoising network as depicted in Figure. 1. Unlike traditional methods [73, 78, 82] that apply an implicit regularized retinex framework, the retinex decomposition network adopts a distinct approach. It uses an explicit parametric regularized retinex model, incorporating parameters specific to individual pixels. Proposed enhancement attention module aims to enhance the V-channel's reflectance and illuminance through two distinct networks (EAR net and EAI net) that incorporate both channel and spatial attention mechanisms. To mitigate color distortion, the initial two modules operate within the HSV color space thus effectively separating color information from illuminance. This approach yields a preliminary enhanced RGB image $\hat{E}$ by merging the adjusted reflectance and illuminance of the V channel with the original (unchanged) H and S channels. Given that the H and S channels may introduce noise, the denoising module (D net) is designed with addition of perceptual loss to eliminate this noise, ultimately producing a refined final image E .
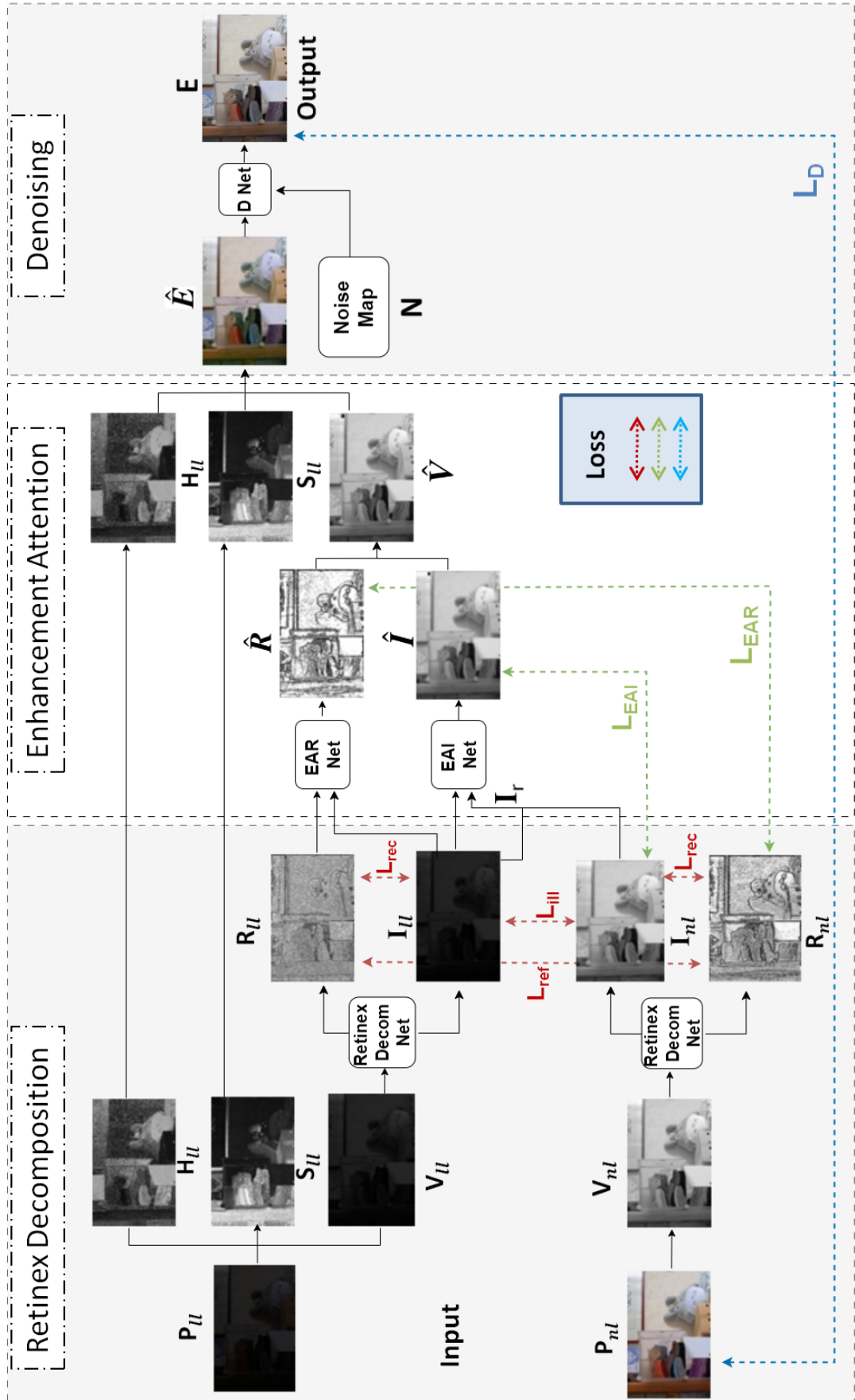
Figure 1: An Improved Deep Learning Model (AIDLM) for LLIE.

AIDLM model takes as input a low-light image $\boldsymbol{P}_{ll}$ along with its associated normal light image $\boldsymbol{P}_{nl}$. It first converts the RGB image to its HSV color space, which is denoted by $(\boldsymbol{H}_{ll}, \boldsymbol{S}_{ll}, \boldsymbol{V}_{ll})$ for the low-light image and $(\boldsymbol{H}_{nl}, \boldsymbol{S}_{nl}, \boldsymbol{V}_{nl})$ for the normal light image. To prevent color distortion, we restrict the retinex decomposition process solely to the V channel. The reflectance and illumination components of the low-light image are represented as $(\boldsymbol{R}_{ll}, \boldsymbol{I}_{ll})$, while those of the normal light image are denoted as $(\boldsymbol{R}_{nl}, \boldsymbol{I}_{nl})$. The Retinex Decomposition network is explained in Section 3.2 and shown in Figure. 2. $\boldsymbol{R}_{ll}$ and $\boldsymbol{I}_{ll}$ are the inputs to the enhance attention reflectance network (EAR Net), whereas $\boldsymbol{I}_{ll}$ and $\boldsymbol{I}_{r}$ are the inputs to the enhance attention illumination network (EAI Net) to produce the enhanced reflectance $\hat{\boldsymbol{R}}$ and enhanced illumination $\hat{\boldsymbol{I}}$, respectively. The proposed EAR Net and EAI Net have the same structure, explained in Section 3.3 and shown in Figure. 3. EAR and EAI Net both incorporate spatial and channel attention mechanisms, which are explained in Section 3.3.1 and shown in Figure. 4. EAI Net enhances brightness by adjusting the illuminance, guided by the following formula:

$$\boldsymbol{I}_r = \boldsymbol{I}_{ll} \oslash (\boldsymbol{I}_{nl} + \varepsilon) \tag{1}$$

where $\oslash$ represents element-wise division and ε is a small positive number to prevent division by zero. In contrast, EAR Net uses $\boldsymbol{I}_{ll}$ as a reference, since the luminance influences the noise in the reflectance. By using the enhanced reflectance $\hat{\boldsymbol{R}}$ and enhanced illumination $\hat{\boldsymbol{I}}$, we can obtain the enhanced V-channel $\hat{\boldsymbol{V}}$ by doing element-wise multiplication $\odot$ of $\hat{\boldsymbol{R}}$ and $\hat{\boldsymbol{I}}$,i.e.,

$$\hat{\boldsymbol{V}} = \hat{\boldsymbol{I}} \odot \hat{\boldsymbol{R}}. \tag{2}$$

As now we have the enhanced V-channel i-e $\hat{\boldsymbol{V}}$, we can get the HSV image by using $(\boldsymbol{H}_{ll}, \boldsymbol{S}_{ll}, \hat{\boldsymbol{V}})$. Then this HSV image is converted into the enhanced RGB image $(\hat{\boldsymbol{E}})$. As few images are taken under severe low light environment, the $(\boldsymbol{H}_{ll}, \boldsymbol{S}_{ll}, \hat{\boldsymbol{E}})$ may contain perceivable noise. Therefore,
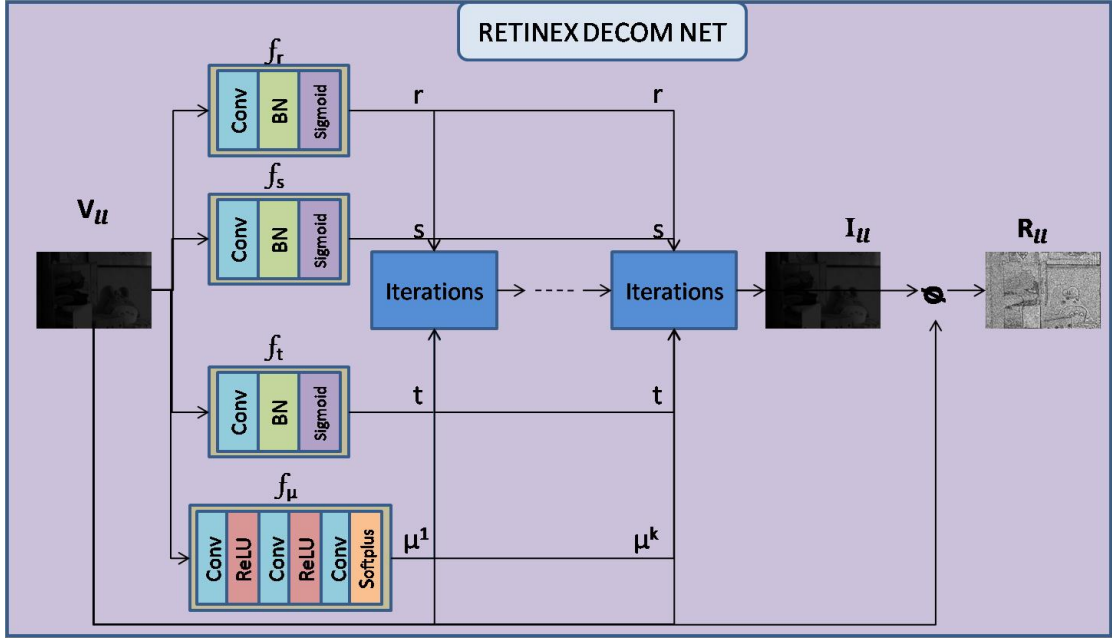
Figure 2: The Retinex Decom Network

we use a noise map ($\boldsymbol{N}$) and a denoising network (D net) to eliminate noise from the enhanced RGB image ($\hat{\boldsymbol{E}}$) and produce the refined enhanced RGB image ($\boldsymbol{E}$) using

$$\boldsymbol{E} = \text{D Net}(\hat{\boldsymbol{E}}, \boldsymbol{N}). \tag{3}$$

3.2 Retinex Decomposition Network

In models that employ analytical Retinex decomposition [43, 66], there is a necessity for manual parameter selection, where as regularizations are hindered by their limited data adaptability. While learning-based methods [75, 70] are more adaptive, they frequently struggle with a lack of interpretability. In this work, RDnet model [92] is used as retinex decom Net, its structure is illustrated in Figure. 2. The model employs explicit parametric regularization with parameters that vary for each pixel to execute Retinex decomposition on the V-channel, breaking it down into reflectance and illumination components. For images captured in low light and normal light conditions, the reflectance and illumination are denoted as ($\boldsymbol{R}_{ll}, \boldsymbol{I}_{ll}$) and ($\boldsymbol{R}_{nl}, \boldsymbol{I}_{nl}$), respectively.

The model parameters, denoted as r_i, s_i, t_i, will be predicted using corresponding modules where i denotes iterations:

$$r_i = f_r(\boldsymbol{V}) \tag{1}$$

$$s_i = f_s(\boldsymbol{V}) \tag{2}$$

$$t_i = f_t(\boldsymbol{V}) \tag{3}$$

which employ a simple architecture consisting of convolution, batch normalization and sigmoid (Conv+BN+Sigmoid) as shown in Figure. 2. In contrast, for the prediction of penalty parameters μ_i, it uses the module

$$\mu_i = f_\mu(\boldsymbol{V}) \tag{4}$$

that utilizes two convolution and ReLU layers followed by a Conv+Softplus layer as shown in Figure. 2. In explicit parametric Retinex decomposition model, $\boldsymbol{V}_{ll} \in \mathbb{R}^N$ represents the V channel of input low light image $\boldsymbol{P}_{ll}$, and $\boldsymbol{I}_{ll} \in \mathbb{R}^N$ denotes the illuminance of $\boldsymbol{V}_{ll}$. For simplicity, the reflectance $\boldsymbol{R}_{ll} \in \mathbb{R}^N$ of $\boldsymbol{V}_{ll}$ is computed as:

$$\boldsymbol{R}_{ll} = \boldsymbol{V}_{ll} \oslash \boldsymbol{I}_{ll} \tag{5}$$

The loss function for Retinex Decom Net $\mathscr{L}_{RET}$ is the combination of distinct loss components: the reconstruction loss $\mathscr{L}_{rec}$, the reflectance consistency loss $\mathscr{L}_{ref}$, and the illumination mutual consistency loss $\mathscr{L}_{ill}$:

$$\mathscr{L}_{RET} = \mathscr{L}_{rec} + \lambda_{ref}\mathscr{L}_{ref} + \lambda_{ill}\mathscr{L}_{ill} \tag{6}$$

where the reconstruction loss $\mathscr{L}_{rec}$ is given by:

$$\mathscr{L}_{rec} = \|\boldsymbol{V}_{ll} - \boldsymbol{I}_{ll} \odot \boldsymbol{R}_{ll}\|_1 + \|\boldsymbol{V}_{nl} - \boldsymbol{I}_{nl} \odot \boldsymbol{R}_{nl}\|_1 \tag{7}$$

The reflectance consistency loss $\mathscr{L}_{ref}$ is given by:

$$\mathscr{L}_{ref} = \|\boldsymbol{R}_{ll} - \boldsymbol{R}_{nl}\|_1 \tag{8}$$

The illumination mutual consistency loss $\mathscr{L}_{ill}$ is given by :

$$\mathscr{L}_{ill} = |\nabla \boldsymbol{I}_{ll}| + |\nabla \boldsymbol{I}_{nl}| \tag{9}$$

where $\|\cdot\|_1$ depict the l_1 norm of vectors, λ_{ref} and λ_{ill} are tuning parameters, and ∇ stands for the gradient operator. According to Retinex theory [3], both $\boldsymbol{V}_{ll}$ and $\boldsymbol{V}_{nl}$ should share the same reflectance. Therefore, $\mathscr{L}_{ref}$ aims to minimize the difference between $\boldsymbol{R}_{ll}$ and $\boldsymbol{R}_{nl}$. Additionally, $\mathscr{L}_{ill}$, as introduced by KinD [58], ensures the structural consistency between low and normal light illumination.

3.3 Enhancement Sub Network

The proposed enhancement networks, EAR net and EAI net, consists of UNet-like architecture [40], skip connections and attention mechanism [90] as illustrated in Figure. 3. UNet-like architecture [40] is well-known for its success in image-to-image translation tasks, and is optimized for efficient and rapid inference. Unlike the UNet [40] which uses concatenation, this approach employs the addition operator for the skip connection. Skip connections boost the efficacy of the UNet for LLIE by retaining crucial details, optimizing training processes, and facilitating the integration of features across different scales. Inspired by spatio-channel attention (SCA) mechanism used in [90], it is used as Attention mechanism in enhancement network. Attention mechanism consist of spatial and channel attentions added to UNet-like architecture to enhance the network's power to capture fine details and contextual information.

$\boldsymbol{I}_{ll}$ and $\boldsymbol{R}_{ll}$ are the random patches of input image with the dimensions of 48x48x1 each, which are first concatenated to produce a patch of dimension 48x48 x2 which is given as input to EAR Net. The EAR net consist of encoder, latent space and decoder block, in which there are 2 blocks for encoder, 1 block of latent space or bottle neck and 2 blocks for decoder as illustrated in Figure. 3. During encoding phase it gradually reduces spatial dimensions as 48, 24 and 12

21
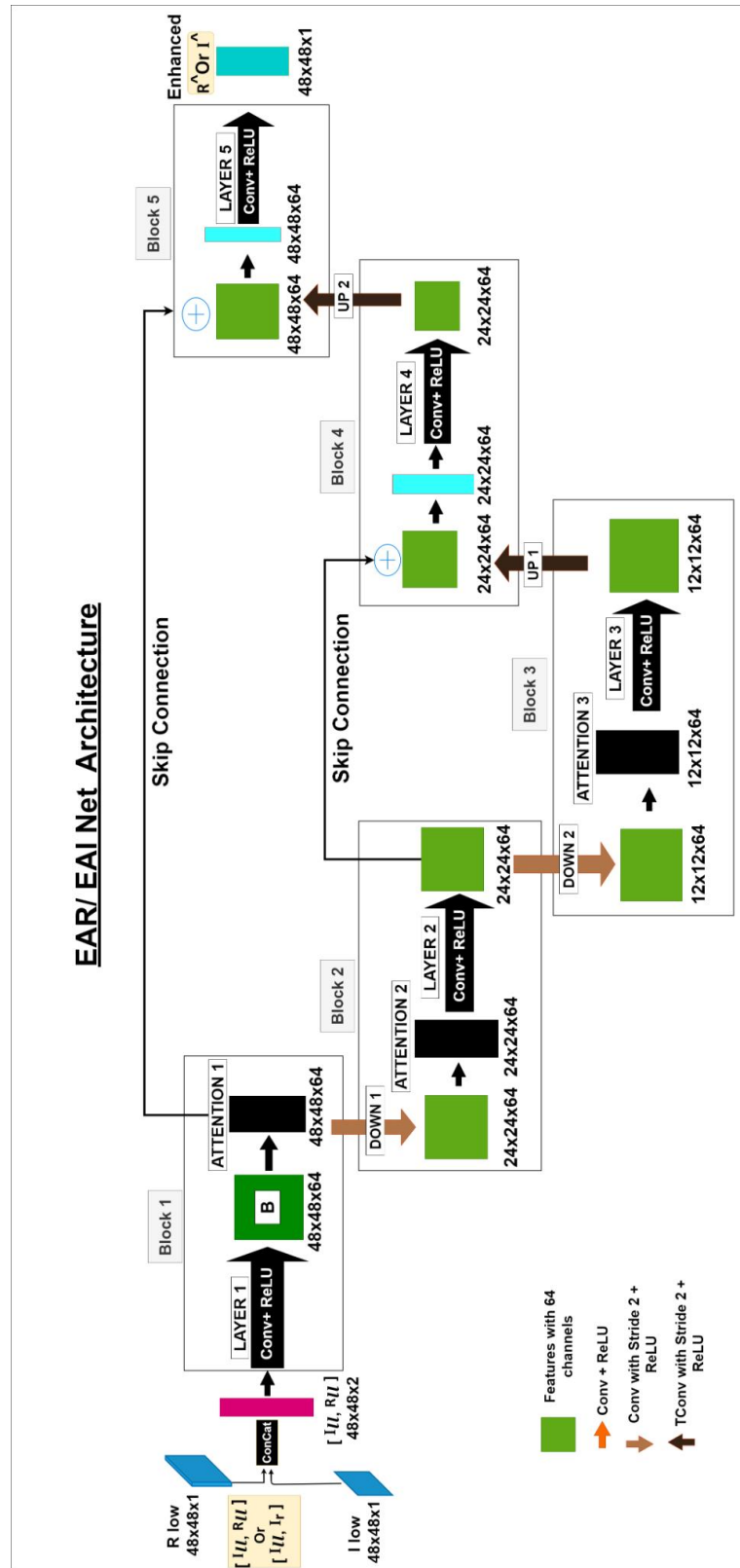
Figure 3: The architecture of proposed EAR/EAI Net.

while the channels remains fixed as 64 in each layer and a fixed kernel size of 3 and padding of 1 is used. In **Block1** the patch of 48x48x2 is passed through **Layer1** where two convolution layers (Conv) each followed by a ReLU activation is applied, which produces a patch of 48x48x64. The **Layer1's** output **B** is processed by the **Attention1** , which employs spatial and channel attention mechanisms(SCA) [90], to emphasize relevant features for image enhancement and produces a patch of dimension 48x48x64. These spatial and channel attention modules help the network concentrate on important spatial areas and specific channel information. The output of **Attention1** is used as a skip connection to **Block5** in decoder as well as it is processed by the **Down1** layer in Block2 which uses Conv + ReLU layers. In **Block2**, **Down1** layer reduces its dimensions to 24x24x64. The **Down1** layer's output is processed by the **Attention2**, which gives out a patch size of 24x24x64 which further will be passed through **Layer2** which uses Conv + ReLU. **Layer2** output of dimension 24x24x64 will be again used as a skip connection to **Block4** in decoder as well as it is given as input to **Down2** layer in Block3. In **Block3**, **Down2** layer uses Conv + ReLU to reduce its dimension to 12x12x64. At this bottle neck or latent space the output of **Down2** is given as input to **Attention3**. The output of **Attention3** with a dimension of 12x12x64 will be given as input to **Layer3** which apply Conv + ReLU to produce the output of 12x12x64. The decoder then increases the spatial resolution of the feature maps as 12, 24 and 48 while the channels remains fixed as 64 in each layer.In the decoder, information from the encoder is combined through skip connections at each scale. Each Up block in the decoder features a transposed convolution layer with a kernel size of 4, a stride of 2, and padding of 1 to enhance the resolution. This is followed by a ReLU activation function to generate the patch. In **Block5**, at the **Layer5**, Conv is applied on the patch 48x48x64 and produce a enhanced reflectance $\hat{R}$ with a dimension of 48x48x1. Similarly I_{ll} and I_r are the inputs for the EAI Net and it produces the enhanced illumination $\hat{I}$ with the same architecture as of EAR net.

3.3.1 Attention Mechanism

CNNs have dramatically improved the effectiveness of vision tasks [19, 33] by autonomously and hierarchically learning spatial feature hierarchies from data, thus minimizing the requirement for manual feature extraction and engineering. To further boost CNN performance, recent research has concentrated on key factors: depth, width, and attention. Previous studies [31, 32, 34, 35] have thoroughly examined the importance of attention, highlighting its vital role in human perception [15, 16]. A key characteristic of the human visual system is its capacity to analyze scenes through a series of selective glimpses, concentrating on essential parts to better comprehend visual structures [20]. Attention not only guides this focus but also enhances the representation of key features.

Expanding upon recent developments in attention-based low-level vision methods [51, 54, 81, 59], we applied a dual attention unit called spatial and channel attention mechanisms (SCA) [90]. This module extracts features from convolutional feature maps across various information streams. As illustrated in Figure. 4, the SCA Module efficiently filters out less relevant features, ensuring that only the most important features are transmitted to subsequent layers. Through channel attention [51] and spatial attention [54] mechanisms, these features are recalibrated.

As illustrated in Figure. 3, the output from the **Input Layer** in Block1 is regarded as a feature map B that contains 64 channels (48x48x64). This feature map B will then pass through **Attention1** layer. In Attention1 layer as shown in Figure. 4, this feature map B will first passes through a Covn+ ReLU+ Conv layer that extracts and enhances relevant features, generating an enhanced feature map, $\hat{\boldsymbol{B}}$. A spatial attention map, denoted as $\hat{\boldsymbol{f}}_{\text{SA}}$, is generated by analyzing the spatial relationships between features. Spatial attention focuses on "where" the informative regions are located in an image. To construct the spatial attention map $\hat{\boldsymbol{f}}_{\text{SA}}$, the spatial attention module first applies max pooling and global average pooling operations independently across
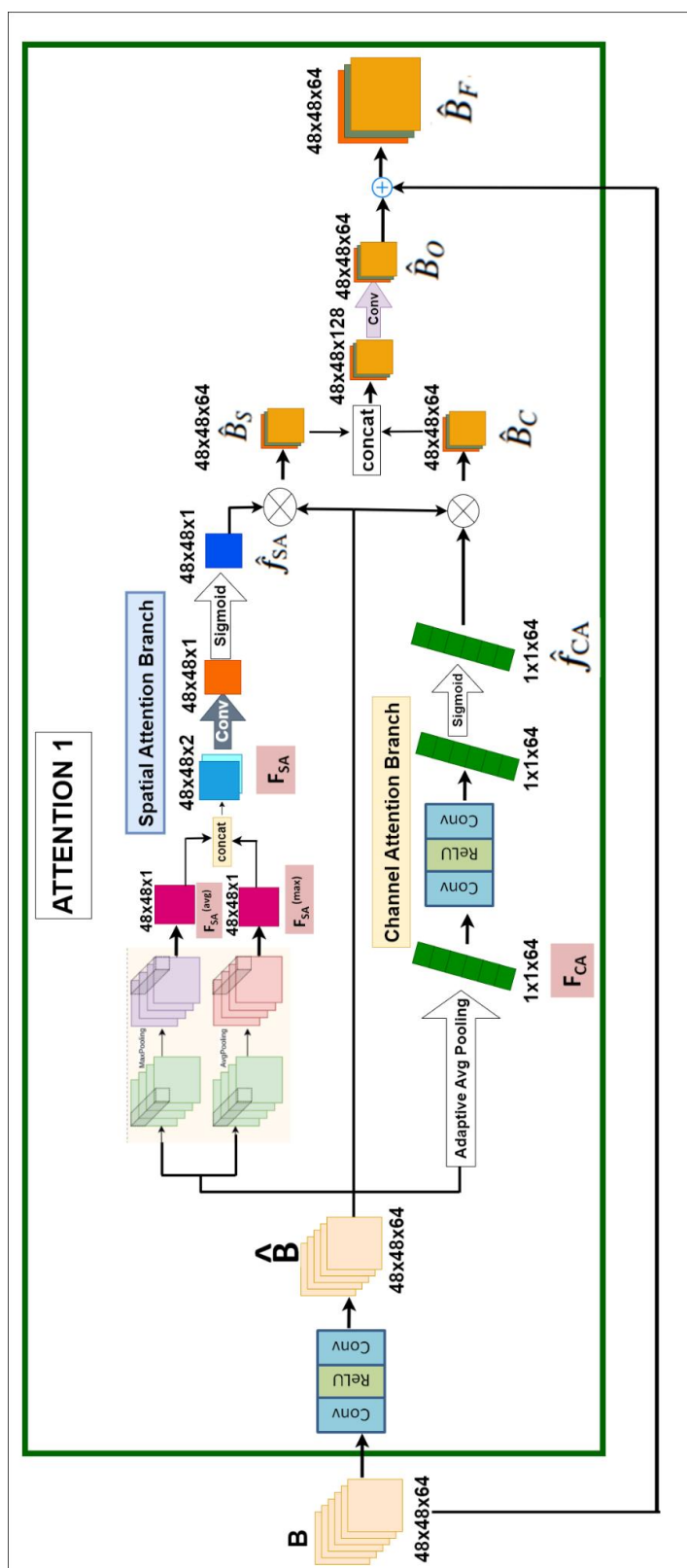
Figure 4: The architecture of the Attention1 Module.

the channel dimensions of the feature map $\hat{\boldsymbol{B}}$.

The max pooling is given as :

$$F_{SA}^{\max}(i,j) = \max_{c} \hat{B}(i,j,c) \tag{1}$$

This operation which highlights the most prominent features in the spatial domain of an image.

The global average pooling is given as :

$$F_{SA}^{\text{avg}}(i,j) = \frac{1}{C}\sum_{c=1}^{C} \hat{B}(i,j,c) \tag{2}$$

This operation providing a smoothed representation of spatial information of an image.

The feature descriptors $F_{SA}^{\max}$ and F_{SA}^{avg} correspond to the results of the max pooling and average pooling operations, respectively. These descriptors are concatenated element-wise to form a combined feature map, denoted as F_{SA}. This resulting feature map has dimensions $\mathbb{R}^{H\times W\times 2}$, which is $\mathbb{R}^{48\times 48\times 2}$ in Attention1. This feature map F_{SA} emphasizes important regions where significant features or anomalies in the reflectance and illuminance may be present.

$$F_{SA} = [f_{SA}^{\text{avg}}, f_{SA}^{\max}] \tag{3}$$

The feature descriptor $\boldsymbol{F}_{SA}$ is then processed through a convolution operation, followed by a sigmoid activation function. This procedure produces the spatial attention map $\hat{\boldsymbol{f}}_{\text{SA}}$, which has dimensions $\mathbb{R}^{H\times W\times 1}$ specifically $\mathbb{R}^{48\times 48\times 1}$ in Attention1.

$$\hat{f}_{SA} = \sigma(\text{Conv}(F_{SA})) \tag{4}$$

The final output, $\hat{B}_S$, from the spatial attention branch is achieved by performing element-wise multiplication of the original feature map $\hat{\boldsymbol{B}}$ with the spatial attention map $\hat{\boldsymbol{f}}_{\text{SA}}$. This step ensures that the refined spatial features are preserved and enhanced, making important regions of the

reflectance and illuminance more prominent.

$$\hat{B}_S = \hat{\boldsymbol{f}}_{\mathrm{SA}} \odot \hat{\boldsymbol{B}} \tag{5}$$

By applying this spatial attention mechanism, the neural network can focus on the most rele-
vant spatial regions, thereby enhancing the feature representation for subsequent layers. This
approach complements channel attention, which focuses on 'what' features are important, by
addressing 'where' these features are located within the spatial dimensions of the input.

Following this, a channel attention map $\hat{\boldsymbol{f}}_{\mathrm{CA}}$ is created by analyzing interactions between differ-
ent channels in the feature maps. Each channel of a feature map $\hat{\boldsymbol{B}}$ is treated as a distinct feature
detector. Primary objective of channel attention is to find and emphasize the most significant
information within the input image.

Channel attention [51] efficiently applied adaptive average pooling on the input feature map $\hat{\boldsymbol{B}}$,
resulting in a tensor $\boldsymbol{F}_{\mathrm{CA}}$ with dimensions $\mathbb{R}^{1 \times 1 \times C}$ that gathers global contextual details.

The global average pooling can be represented as:

$$F_{\mathrm{CA}} = \frac{1}{h \times w} \sum_{i=1}^{h} \sum_{j=1}^{w} \hat{F}(i, j, c) \tag{6}$$

Next, F_{CA} is passing through several convolutional layers followed by a sigmoid activation
function. This process generates channel attention map, denoted as $\hat{\boldsymbol{f}}_{\mathrm{CA}}$. This channel attention
map $\hat{\boldsymbol{f}}_{\mathrm{CA}}$ emphasizes channels that carry more critical features for reflectance and illuminance.
This operation can be represented as:

$$\hat{\boldsymbol{f}}_{\mathrm{CA}} = \sigma(\mathrm{Conv}(F_{\mathrm{CA}})) \tag{7}$$

Finally, the output $\hat{B}_C$ from the channel attention branch is derived by applying element-wise
multiplication between the feature map $\hat{\boldsymbol{B}}$ and channel attention map $\hat{\boldsymbol{f}}_{\mathrm{CA}}$, effectively scaling

the feature map.

$$\hat{B}_C = \hat{\boldsymbol{f}}_{\mathrm{CA}} \odot \hat{\boldsymbol{B}} \tag{8}$$

This step fine-tunes the feature map by focusing on the most informative channels, improving the representation of essential features in the reflectance and illuminance data.

The output of the spatial attention branch $\hat{B}_S$ and the channel attention branch $\hat{B}_C$ are combined through concatenation and then processed by a convolutional layer (Conv) to decrease the number of channels, resulting in $\hat{B}_O$.

$$\hat{B}_O = \mathrm{Conv}([\hat{B}_S, \hat{B}_C]) \tag{9}$$

This step integrates both spatial and channel attention enhancements, creating a more comprehensive refined feature map that incorporates improvements from both mechanisms.

This $\hat{B}_O$ is then added to a feature map B to get the final output of the attention module, which is denoted as $\hat{B}_F$.

$$\hat{B}_F = \hat{B}_O + B \tag{10}$$

This combination ensures that the final feature representation benefits from both spatial and channel attention, leading to a more detailed and accurate enhancement of the reflectance and illuminance.

The loss $\mathscr{L}_{EAR}$ for EAR net is

$$\mathscr{L}_{EAR} = \|\hat{\boldsymbol{R}} - \boldsymbol{R}_{nl}\|_1 + \|\boldsymbol{V}_{nl} - \boldsymbol{I}_{nl} \odot \hat{\boldsymbol{R}}\|_1 + \||\nabla\hat{\boldsymbol{R}}| - |\nabla\boldsymbol{R}_{nl}|\|_2^2. \tag{11}$$

The first two terms represent fidelity measures between the enhanced reflectance $\hat{\boldsymbol{R}}$ and the reflectance from normal-light images $\boldsymbol{R}_{nl}$. The third term quantifies the difference between the absolute gradients of $\hat{\boldsymbol{R}}$ and $\boldsymbol{R}_{nl}$.

The loss $\mathscr{L}_{EAI}$ for EAI net is

$$\mathscr{L}_{EAI} = \|\hat{\boldsymbol{I}} - \boldsymbol{I}_{nl}\|_1 + \|\boldsymbol{V}_{nl} - \hat{\boldsymbol{I}} \odot \boldsymbol{R}_{nl}\|_1 + \||\nabla\hat{\boldsymbol{I}}| - |\nabla\boldsymbol{I}_{nl}|\|_2^2. \tag{12}$$
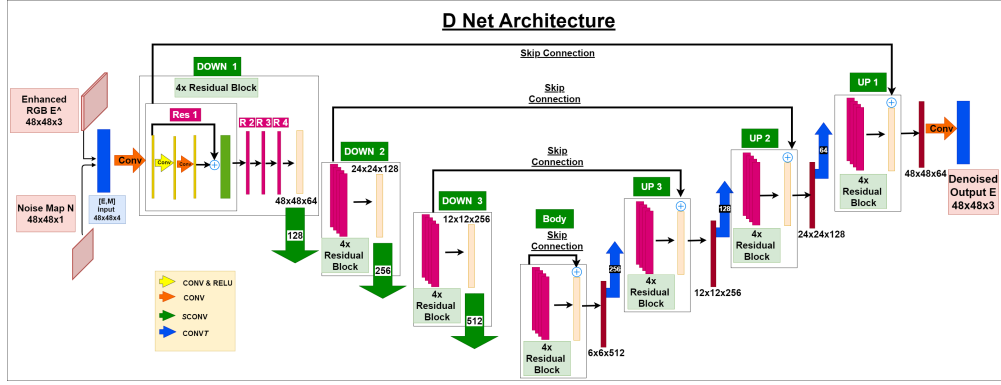
Figure 5: The architecture of the D Net.

3.4 Denoising Network

The DRU Net [83], known for its advanced capabilities in image denoising, is improved by adding perceptual loss and then utilized as D Net in AIDLM. Its structure is illustrated in Figure. 5. The D Net architecture is a highly advanced denoising network that combines the strengths of U-Net and ResNet architectures. This design is capable of managing a broad spectrum of noise levels, ensuring both robustness and adaptability.

$$E = \text{D Net}(\hat{E}, N) \tag{1}$$

In the above equation $\hat{\boldsymbol{E}}$ is the coarsely enhanced RGB image and N is a noise map and E is the denoised RGB image.

Although the DRUNet model achieves high PSNR and SSIM values, it often produces overly smooth or blurry images that lack fine details and may fail to preserve the semantic content and structure, particularly in complex regions. To address these issues, perceptual loss is incorporated. This results in more visually appealing images and enhances the overall performance of denoising networks.

The loss $\mathscr{L}_D$ for D Net is defined as

$$\mathscr{L}_D = \lambda_{MSE}\left(\|E - P_{nl}\|_2^2\right) + \lambda_{SSIM}\left(1 - \text{SSIM}(E, P_{nl})\right)$$

$$+ \lambda_{Per}\left(\text{PerceptualLoss}(E, P_{nl})\right). \tag{2}$$

Here, λ_{MSE}, λ_{SSIM}, and λ_{Per} are tuning parameters which are selected as 20, 1 and 0.5 respectively.

Experimental Environment of

Proposed Approach

This chapter offers a thorough rundown of the metrics used to assess the performance, the datasets that were chosen, and the experimental setup of the suggested method. Proposed method is constructed with various leading-edge learning-based approaches to showcase its efficacy. The eight supervised learning models are used for evaluation all are trained on the LOL dataset [53]. It includes RetinexNet [53], KinD [58], KinD++ [76], URetinex-Net [82], LLFlow [80], Retinexformer [85], RQ-LLIE [87] and DPRED [92]. For the five unsupervised learning methods, we evaluated EnlightenGAN [70], RUAS [73], SwinIR [71], Zero-DCE [60], and Zero-DCE++ [77] using their respective provided models.

4.1 Dataset Selection

The datasets utilized in study are pivotal for assessing the effectiveness of proposed LLIE method. We categorize them into two main types: reference datasets (with ground-truth) and non-reference datasets (without ground-truth). These categories offer a thorough framework for

evaluating the effectiveness of our method under various scenarios and lighting conditions.

Reference Datasets: Reference datasets are used as standards to assess how well algorithms and models function. They offer a consistent basis for comparing results and maintaining uniformity in research and development.The two ref datasets are as following :

- **LOL Dataset**: The LOL (Low-Light) dataset was created expressly for the purpose of improving low-light photos [53]. It is composed of five hundred image pairs, each containing an image taken in low light and a reference image taken in normal illumination. Fifteen pairs are set up for testing, and the remaining four hundred and eighty-five are designated for training. This dataset is particularly valuable for supervised learning models as it provides ground-truth images that allow for direct comparison and evaluation of enhancement algorithms. The LOL dataset is structured to offer a diverse range of scenarios and lighting situations, making it an excellent resource for developing and testing LLIE algorithms.

- **MIT-Adobe Five K Dataset**: The MIT-Adobe Five K dataset [21] is another significant dataset used for evaluating image enhancement techniques. It contains 5000 images, each paired with a reference image that has been retouched by professional photographers to enhance its visual appeal. We took 500 low-light/reference image pairs at random from this dataset for our analysis. This dataset is valuable for its high-quality reference images and the variety of lighting conditions, providing a robust benchmark for enhancement methods.

Non-Reference Datasets These datasets are important in evaluating the generalisation abilities of LLIE methods even though they don't contain ground-truth reference photos. They include a wide range of real world low-light images, each with unique qualities:

- **LIME Dataset**: The LIME (LLIE) dataset [41] has 10 images taken in poor light condi-
  tions. Despite its small size, this dataset is frequently used for benchmarking due to its
  challenging lighting conditions and diverse scenes.

- **NPE Dataset**: The NPE (Naturalness Preserved Enhancement) dataset [30] consists of
  84 low-light images. This dataset is a valuable tool for assessing how well enhancement
  algorithms maintain the aesthetics of photos because it concentrates on maintaining the
  natural appearance of images while improving visibility.

- **MEF Dataset**: The MEF (Multi-Exposure Fusion) dataset [39] includes 17 images. This
  dataset is used to evaluate methods that combine multiple exposure images to create a
  single, well-exposed image.

- **DICM Dataset**: The DICM (Dark Image Contrast Enhancement) dataset [27] contains
  64 images of low light. It is designed to evaluate contrast enhancement techniques and
  includes a different types of scenes taken in low-light situations.

4.2   Evaluation Matrices

For quantitative comparison, we employ several metrics: peak signal-to-noise ratio (PSNR)
[24], structural similarity index measure (SSIM) [18], Learned Perceptual Image Patch Simi-
larity (LPIPS) [55], DeltaE [2], Lightness-Order-Error (LOE) [30] and Lightness-Order-Error-
Reference (LOE_{ref}) [58].

- **PSNR and SSIM**: Metrics like PSNR [24] and SSIM [18] are frequently used to evaluate
  how closely enhanced images resemble ground-truth references. PSNR quantifies the
  peak error between images, where higher values signify superior quality. SSIM, on the
  other hand, compares structural information as seen by the human eye to evaluate image

quality. Better structural similarity and overall image quality are reflected by higher SSIM values.

- **LPIPS**: The LPIPS [55] metric measures the perceptual similarity of images. The difference between the images is assessed using deep neural network features; lower values indicate stronger perceptual similarity and higher quality.

- **DeltaE**: The colour difference between the improved image and the ground truth is measured using a metric called DeltaE [2]. It quantifies the perceptual difference in color, with lower values indicating a smaller color difference and better color accuracy.

- **LOE and** LOE_{ref}: The LOE [30] metric was created specifically to evaluate methods for improving low-light images. It uses the input low-light image as a reference. When ground-truth photographs are available, a modified version known as LOE_{ref} sets the reference to the normal-light image [58]. Lower values of LOE [30] and LOE_{ref} [58] signify a more accurate preservation of lightness order, indicating superior enhancement quality.

In other words, lower values for LPIPS, DeltaE, LOE, and LOE_{ref} indicate better performance in terms of perceptual similarity, colour accuracy, and lightness order preservation. Higher values for PSNR and SSIM, on the other hand, indicate better picture quality.

4.3 Experimantal Setup

The retinex decomposition network, the enhancement attention network, and the denoising network are the three separate networks that make up the suggested method. Training the entire system from scratch provides substantial hurdles because to the overall sophisticated architecture. As a result, we use a technique in which each sub-network is trained separately while the

other networks remain unchanged. We utilise a batch size of 10 for the retinex decomposition network, a patch size of 48×48, and a 10^{-4} learning rate that is modified every 10 epochs. In order to minimise the loss function $\mathscr{L}_{\text{RET}}$ as given in Equation 6, the network is trained for 100 epochs with the parameters λ_{ref} and λ_{ill} empirically fixed at 0.1. The enhancement attention network training consists of 100 epochs with fixed learning rate of 10^{-4}. The enhancement networks, EARnet and EAInet, are trained by optimising the losses $\mathscr{L}_{\text{EAR}}$ in Equation 11 and $\mathscr{L}_{\text{EAI}}$ in Equation 12, respectively. The batch size and patch size for the Enhancement Attention Network stay at 10 and 48×48, respectively. The patch size for the denoising network (Dnet) is 48×48, the batch size is 10, and the learning rate is 10^{-5}, which is fixed during training. Every five epochs, the 140 epochs of training for this network are evaluated. Using the default parameters, we employ the Adam optimiser [36]. We initialise the Denoising Network with pre-trained parameters from DRU Net [83], which can handle different kinds of noise, given the unknown noise distribution. Then, in order to better adapt to the dataset, we adjust the Dnet parameters using the loss $\mathscr{L}_D$ in Equation 2. The coarsely enhanced RGB image $\hat{\boldsymbol{E}}$ and a manually adjusted noise level map $\boldsymbol{N}$ make up Dnet's input.The noise level is set to 15 during training. For the non-reference datasets [41, 30, 39, 27], 15 for the LOL dataset [53], and 17 for the MIT-Adobe Five K dataset [21], the noise level is empirically adjusted during the testing stage. Our method is built in PyTorch and operates on an NVIDIA GeForce GTX 1660 Super GPU.

Ablation Study and Results

5.1 Ablation Study

In this chapter we give ablation studies to assess the contributions of various model components. First, we show why the enhancement module should include both spatial and channel attention. We then evaluate the contributions of the *EARnet* and *EAInet* modules empirically. Finally, we look into what the *Dnet* denoising module does. The LOL dataset [53] was used for these studies, and Table 1 summarises the results. Table 1 indicates that the full AIDLM model delivers the best performance, achieving a 26.551dB PSNR, 0.878 SSIM, 0.204 LPIPS, 7.546 DeltaE and 59.464 LOE$_{ref}$. When spatial and channel attention mechanisms are excluded, there is a decline in performance, with the 25.097dB PSNR, 0.861 SSIM, 0.248 LPIPS, 8.335 DeltaE and 80.66 LOE$_{ref}$, highlighting the importance of these components. Removing *EARnet* gives a 24.915dB PSNR, 0.873 SSIM, 0.213 LPIPS, 8.296 DeltaE and 220.81 LOE$_{ref}$, while excluding *EAInet* significantly reduces PSNR to 18.081dB, SSIM to 0.804, LPIPS to 0.288, DeltaE to 10.171 and LOE$_{ref}$ to 645.39. Finally, omitting *Dnet* decreases the PSNR to 23.681, causes a substantial reduction in SSIM to 0.540, LPIPS to 0.409, DeltaE to 9.390 and LOE$_{ref}$ to 203.03, highlighting the crucial role of each module in the enhancement network.

Table 1: Metrics averaged across 15 test images from the LOL dataset. Bold text indicates the top results.

Method	PSNR↑	SSIM↑	LPIPS↓	DeltaE↓	LOE$_{ref}$↓
AIDLM	**26.551**	**0.878**	**0.204**	**7.546**	**59.464**
AIDLM (without spatial and channel attention)	25.097	0.861	0.248	8.335	80.66
AIDLM (without *EARnet*)	24.915	0.873	0.213	8.296	220.81
AIDLM (without *EAInet*)	18.081	0.804	0.288	10.171	645.39
AIDLM (without *Dnet*)	23.681	0.540	0.409	9.390	203.03

Advantages of the spatial and channel attention in the enhancement module.

Table 1 shows the average metrics of AIDLM with and without spatial and channel attention in the improvement module in the first and second rows, respectively.The AIDLM achieves a PSNR of 26.551 dB and an SSIM of 0.878 when spatial and channel attention are integrated into the enhancement module, confirming its efficacy in enhancing image quality. However, if spatial and channel attention are omitted, the PSNR drops to 25.097 dB and the SSIM to 0.861, underscoring their crucial role in the enhancement network. The images presented in Figure 6 illustrate the performance of the AIDLM with and without the inclusion of spatial and channel attention modules in its enhancement network (EAI and EAR net). These modules play a important role in enhancing image quality, and their effectiveness is quantified using PSNR and SSIM.In Figure 6a, the AIDLM model operates without the spatial and channel attention modules, resulting in a PSNR of 24.2027 dB and an SSIM of 0.8907.Although the image's visual quality is enhanced by the improvement network, there are still certain locations where it is inadequate. Specifically, textures in the shadowed regions appear somewhat blurred, and the overall contrast is not as sharp. This is primarily due to the model's inability to effectively

(a) AIDLM without spatial and channel attention
PSNR = 24.2027 dB
SSIM = 0.8907

(b) AIDLM with spatial and channel attention
PSNR = 26.332 dB
SSIM = 0.9094
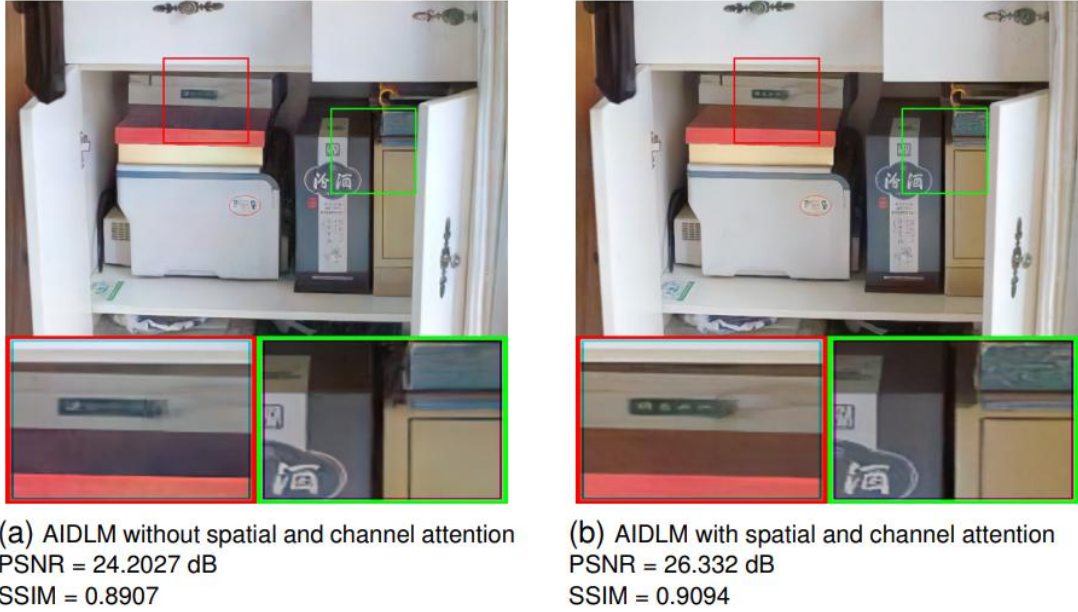
Figure 6: Visual comparison of AIDLM. For better understanding, objects are zoomed in red and green boxes.

highlight and enhance critical regions and features within the image, a function that the attention modules would typically perform.

Conversely, Figure 6b demonstrates the AIDLM model with integrated spatial and channel attention modules. This setup achieves a higher PSNR of 26.332 dB and an SSIM of 0.9094. The attention modules enable the network to focus on significant areas and features, thereby enhancing the overall clarity and details. The textures in the darker regions are more distinct, and there is a noticeable improvement in the contrast between different elements. This enhanced attention mechanism allows for more precise adjustments in brightness and color, resulting in a more visually appealing and accurately enhanced image. Overall, this ablation study underscores the importance of incorporating spatial and channel attention modules in the AIDLM's enhancement network. These modules significantly enhance the model's capability to improve image quality, as evidenced by higher PSNR and SSIM metrics. By emphasizing and amplifying critical image features, the attention mechanisms enable more detailed and accurate enhancement of low-light images, effectively preserving and improving visual details.

(a) Input
PSNR = 4.5212 dB
SSIM = 0.1365

(b) Ground-truth

(c) AIDLM
PSNR = 26.332 dB
SSIM = 0.9094

(d) AIDLM (without spacial and channel attention)
PSNR = 24.2027 dB
SSIM = 0.8907

(e) AIDLM (without EAR net)
PSNR = 23.5023 dB
SSIM = 0.8734

(f) AIDLM (without EAI net)
PSNR = 14.2070 dB
SSIM = 0.7861

(g) AIDLM (without D net)
PSNR = 25.0473 dB
SSIM = 0.3212

39

Figure 7: Visual comparison of different components of AIDLM model.

Effects of the enhancement modules *EARnet* **and** *EAInet***.**

The analysis of the AIDLM with and without specific enhancement modules—EARnet and EAInet—reveals their significant contributions to the network's performance.The average PSNRs and SSIMs over 15 test photos acquired using our network in two scenarios—one without *EARnet* and the other without *EAInet*—are shown in the third and fourth rows of Table 1, respectively. The AIDLM achieves a PSNR of 26.551dB and 0.878 SSIM when all modules are included, highlighting the model's capability to enhance image quality effectively. Removing EARnet reduces the PSNR to 24.915 dB and 0.873 SSIM, demonstrating its essential role in the enhancement network. Although the impact is not as severe as with other modules, it still signifies EARnet's contribution to refining and detailing the image enhancement process. Notably, the absence of EAInet results in a drastic reduction in metrics, with a PSNR of 18.081dB and 0.804 SSIM. This significant decline underscores the critical role of EAInet within the network. EAInet is pivotal in preserving structural integrity and enhancing overall image quality, and its removal severely degrades the network's performance. The stark difference in metrics when EAInet is excluded highlights its indispensable contribution to achieving high-quality enhancements. In Figure 7, the visual comparison provides further insights into the impact of excluding EARnet and EAInet. The input image (Figure 7a) and its ground truth (Figure 7b) set the baseline for comparison. The full AIDLM model (Figure 7c) achieves superior enhancement, reflected in a PSNR of 26.332dB and an SSIM of 0.9094. The exclusion of EARnet (Figure 7e) results in a PSNR of 23.5023 dB and 0.8734 SSIM, showing a noticeable decline in image quality, though not as severe as without EAInet (Figure 7f), which plummets to a PSNR of 14.2070dB and 0.7861 SSIM. This significant degradation underscores EAInet's crucial function in the enhancement process.

Effect of the Denoising Module *Dnet***.**

To understand the impact of the denoising module *Dnet*, we conducted tests on our AIDLM model using 15 images from the LOL dataset, excluding *Dnet*. The average metrics are presented in the fifth row of Table 1. Without *Dnet*, the performance drops significantly, with a 23.681dB PSNR and 0.540 SSIM, underscoring the importance of each component in the AIDLM network. Fig.7 shows the improved colour image of a test image sample. The resultant improved image without the denoising module has observable noise that reduces image quality, as shown in Fig.7g. Our network efficiently reduces noise and performs better after integrating the denoising module *Dnet*, proving the *Dnet* module's importance in our enhancement network.

5.2 Evaluation of The Proposed Method

We present a comprehensive evaluation of our proposed method, primarily trained on the LOL dataset [53], and extend its assessment to diverse datasets such as MIT-Adobe 5K, LIME, DICM, NPE, and MEF. Despite being trained exclusively on the LOL dataset, our method is evaluated for its generalization capability across these varied datasets. We rigorously compare its performance against established benchmarks, examining both quantitative metrics and qualitative results. Additionally, we explore its applicability in unsupervised settings, assessing its robustness and adaptability where labeled data may be limited or absent. Alongside performance evaluations, we analyze the computational complexities of our approach, providing insights into its efficiency and scalability relative to state-of-the-art methods.

5.2.1 Evaluation on LOL Dataset

The LOL dataset [53] includes 500 pairs of images taken in low and normal lighting conditions. Out of these, we utilize 485 pairs for training purposes and reserve 15 pairs for testing. The metrics for these tests are summarized in Table 1. Additionally, we offer a visual com-

parison of one of the images in Fig. 8. Based on the outcomes, we can derive the following insights. It also demonstrate the effectiveness of our AIDLM model. AIDLM achieves the highest SSIM (0.878), the lowest LPIPS (0.204), the lowest DeltaE (7.546), and the lowest LOE_{ref} (59.464), indicating superior image quality, color accuracy, and noise suppression. Our model, AIDLM, achieves a PSNR of 26.551, surpassing DPRED [92] with its PSNR of 26.092. However, Retinexformer [85] leads with the highest PSNR of 27.182, closely followed by AIDLM. LLFlow [80] achieves the second-best SSIM and LPIPS. DPRED ([92]) achieves the second-best LOE_{ref}. These results confirm that AIDLM provides a clear improvement over existing methods in LLIE.

As shown in Fig.8, DPRED [92] produces slightly blurry images with muted colors and a lack of sharpness. EnGAN [70] results in overly saturated colors and high contrast, leading to an artificial appearance. Jiang's [71] method is too dark overall, losing details in the shadows. KinD [58] and KinD++ [76] both suffer from noise and unnatural tints, with KinD++ slightly improved but still lacking in color fidelity and contrast. LLFlow [80] shows excessive sharpening and unrealistic colors. Retinexformer [85] and RetinexNet [53] struggle with exposure, leading to dark, dull images with significant noise and artifacts. RQ-LLIE [87] has uneven exposure, making parts of the image overly bright and others too dark. RUAS [73] is too dark, losing details in shadows, while URetinexNet [82] has uneven exposure and unnatural color balance. Zero-DCE [60] and Zero-DCE++ [77] both produce noisy images, with Zero-DCE++ having slightly better color correction but still struggling with exposure balance. The AIDLM method, in contrast, excels in image enhancement by maintaining a perfect balance between brightness and contrast, ensuring that details are clearly visible in both dark and light areas. The colors are natural and well-balanced, providing a realistic and pleasing look. The image is sharp and clear without any noticeable noise or artifacts. This method enhances the image effectively while preserving the natural appearance and intricate details, making it superior to the other methods
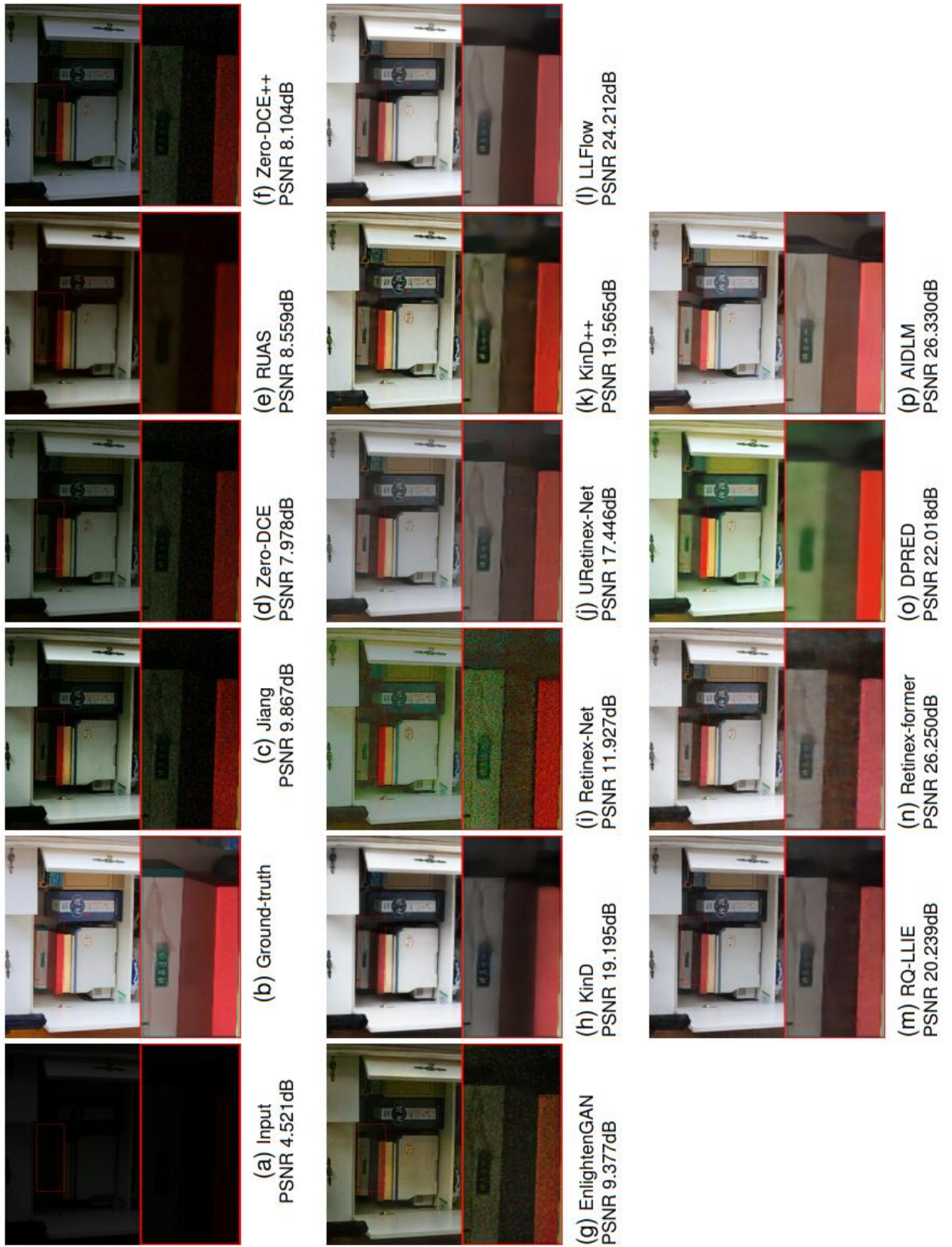
Figure 8: Visual comparison on LOL dataset. For better understanding, objects are zoomed in red box.

Table 2: Metrics averaged across 15 test images from the LOL dataset [53]. Bold text indicates the top results, while underlining indicates the second-best results.

Method	PSNR	SSIM	LPIPS	DeltaE	LOE	LOE$_{ref}$
EnlightenGAN [70]	17.483	0.655	0.391	10.305	355.321	395.591
RUAS [73]	16.445	0.505	0.382	10.414	**93.285**	218.909
Jiang et al. [71]	16.863	0.483	0.399	10.399	328.191	405.255
Zero-DCE [60]	14.861	0.562	0.385	10.446	190.711	272.609
Zero-DCE++ [77]	15.142	0.565	0.387	10.461	168.308	256.581
RetinexNet [53]	16.775	0.429	0.465	10.500	475.029	460.169
KinD [58]	20.381	0.831	0.270	9.711	208.718	276.427
KinD++ [76]	21.812	0.836	0.288	9.563	228.423	280.479
URetinex-Net [82]	21.329	0.835	0.238	10.040	<u>110.972</u>	197.015
LLFlow [80]	24.999	<u>0.873</u>	<u>0.224</u>	8.861	277.611	309.147
Retinexformer [85]	**27.182**	0.850	0.252	<u>7.855</u>	122.065	162.853
RQ-LLIE [87]	23.161	0.851	0.232	9.696	148.115	177.847
DPRED [92]	26.092	0.852	0.367	8.172	214.776	<u>88.150</u>
AIDLM (Proposed)	<u>26.551</u>	**0.878**	**0.204**	**7.546**	225.095	**59.464**

evaluated. The enhanced images demonstrate a high level of detail, accurate color reproduction, and a realistic appearance, showcasing the method's ability to handle a wide range of scenes and lighting conditions effectively.

5.2.2 Generalization

Without any further training or fine-tuning, we tested the suggested method's generalisation abilities on 500 photos taken from the MIT-Adobe Five K dataset [21]. The results, summarized

44

Table 3: Metrics averaged across 500 test images from MIT-Adobe Five K dataset [21]. Bold text indicates the top results, while underlining indicates the second-best results, and third best are with @.

Method	PSNR	SSIM	LPIPS	DeltaE	LOE	LOE$_{ref}$
EnlightenGAN [70]	15.6214	0.7403	0.2064	10.2969	552.1799	549.2273
RUAS [73]	7.7616	0.4675	0.4999	10.3984	825.8669	816.9359
Jiang et al. [71]	15.4296	0.7238	0.2246	10.0625	688.6929	689.0131
Zero-DCE [60]	15.4630	0.7011	0.2805	10.3984	365.0499	355.1534
Zero-DCE++ [77]	14.9995	0.6952	0.3107	10.5391	328.3433	333.8250
RetinexNet [53]	12.3053	0.6318	0.3027	10.1953	1377.3805	1352.7196
KinD [58]	21.2163	0.8042	0.2292	<u>9.8203</u>	212.8142	225.9808
KinD++ [76]	21.6492	0.8188	**0.1994**	9.8828	140.2021	154.8203
URetinex-Net [82]	13.6056	0.6876	0.2316	10.2344	355.6524	345.4250
LLFlow [80]	19.7882	0.7806	<u>0.2016</u>	10.0000	303.9430	318.0595
Retinexformer [85]	18.7697	0.7361	0.2404	10.1172	210.2383	223.9037
RQ-LLIE [87]	13.9310	0.7013	0.2290	10.2812	152.9560	161.1115
DPRED [92]	<u>24.4446</u>	<u>0.8336</u>	0.2317	@8.0859	**118.7620**	**61.3503**
AIDLM (Proposed)	**24.4581**	**0.8573**	@0.2047	**7.9414**	<u>131.01</u>	<u>83.99</u>

in Table 3, highlight the average metrics achieved by various methods, while Fig.9 offers a visual comparison using a sample image. Our method AIDLM consistently outperforms previous approaches as it is best in metrics like PSNR(24.4581), SSIM(0.8573), and DeltaE(7.9414) where as it is second best in LOE(131.01) and LOE$_{ref}$ (83.99) and third best in LPIPS (0.2047).

The improved pictures of sample photos from the MIT-Adobe Five K dataset are displayed in Fig.9 [21]. The analysis of supervised and unsupervised methods on MIT-Adobe Five K

dataset [21] reveals distinct issues across various methods for LLIE. DPRED ([92]) introduces minor color shifts and noise, whereas Jiang [71] enhances brightness but creates unnatural color changes and artifacts. KinD++ [76] has minor halo effects and could benefit from sharper output. LLFlow [80] maintains good color accuracy but does not sufficiently brighten the image. Retinexformer [85] exhibits significant color distortions and artifacts, particularly around balloons and the background. RetinexNet [53] results in color shifts and halo effects, producing overly processed and unnatural images. RQ-LLIE [87] improves brightness but introduces slight color shifts and artifacts, leading to some loss of detail. URetinexNet [82] performs fairly well but still has some artifacts and noise. Both Zero-DCE++ [77] and Zero-DCE [60] show color shifts, over-brightening, and unnatural appearances with visible artifacts and noise. EnGAN [70] balances brightness and color well but causes noticeable unnatural color shifts, particularly in skin tones. KinD [58] achieves good color accuracy and brightness but has minor artifacts. Finally, RUAS [73] causes significant over-brightening, leading to washed-out images with severe color distortion and loss of detail. RetinexNet [53] retains noise in the final enhanced images, while KinD [58], KinD++ [76], URetinex-Net [82], LLFlow [80], Retinexformer [85], RQ-LLIE [87], and our method AIDLM suppress noise well.

AIDLM excels in LLIE on MIT-Adobe Five K dataset [21] by maintaining excellent color accuracy and brightness levels. The method effectively illuminates low-light inputs while preserving natural colors and reducing noise and artifacts. The overall sharpness and detail in the image are well-preserved, and there are minimal visible artifacts. This balance of enhancement and accuracy makes AIDLM a strong performer among the evaluated models. The model demonstrates natural color representation, significant noise reduction, crisp sharpness, and minimal artifacts, ensuring high-quality image enhancement across diverse scenes and lighting conditions.

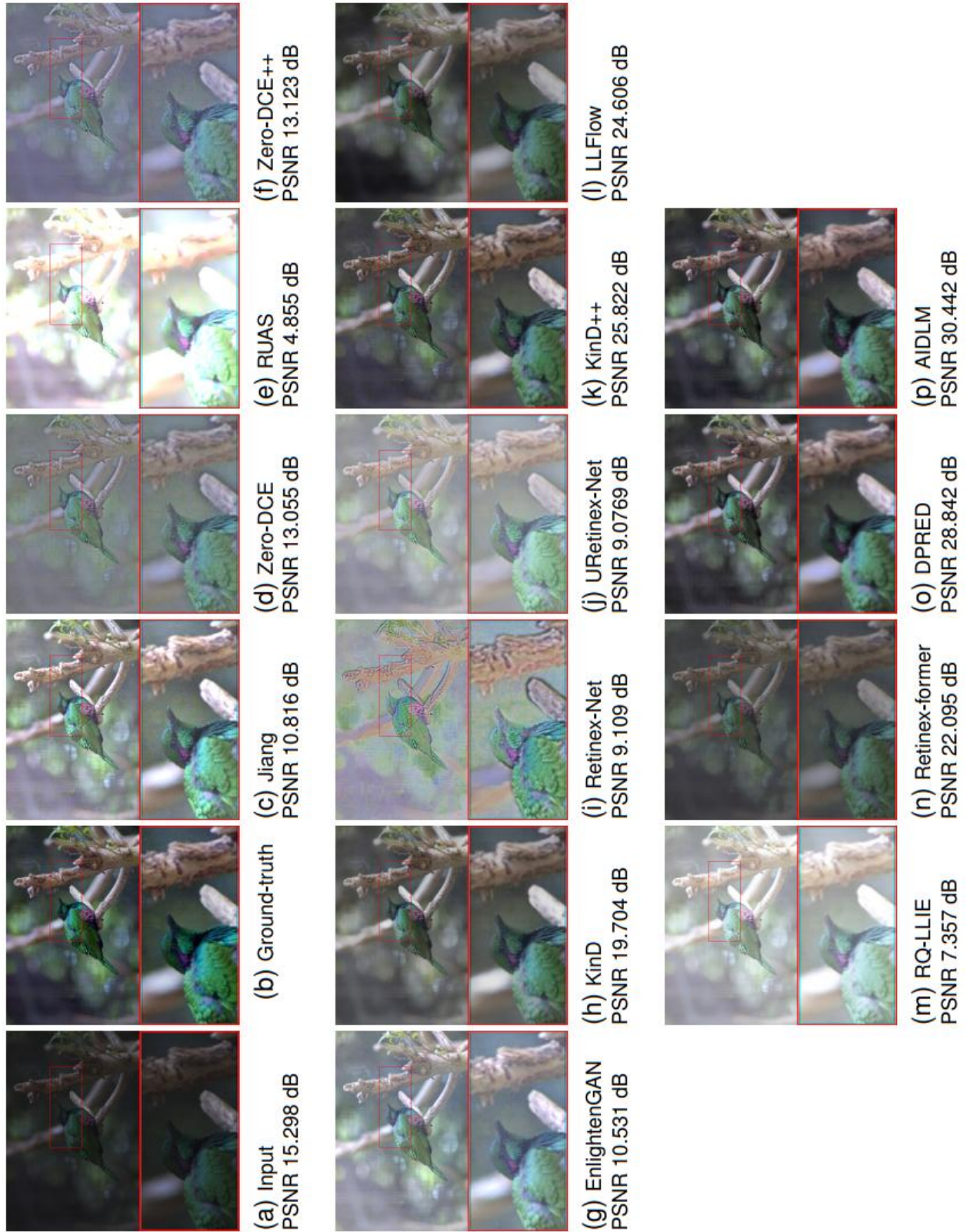Using a few non-reference datasets, we test AIDLM in order to confirm the suggested method's

Figure 9: Visual comparison on MIT-Adobe Five K dataset [21]. For better understanding, objects are zoomed in the red box.

Table 4: All non-reference datasets [41, 30, 39, 27] have average LOE [41] metrics. Bold text indicates the top results, while underlining indicates the second-best results.

Method	LIME [41]	NPE [30]	MEF [39]	DICM [27]
EnlightenGAN [70]	396.43	494.72	416.86	483.58
RUAS [73]	267.81	702.41	292.42	845.52
Jiang et al. [71]	512.49	707.47	479.45	644.28
Zero-DCE [60]	175.68	216.72	167.01	177.20
Zero-DCE++ [77]	126.01	212.63	125.40	195.12
RetinexNet [53]	542.64	700.57	615.53	447.46
KinD [58]	167.02	198.71	158.54	189.70
KinD++ [76]	353.73	457.67	358.43	337.19
URetinex-Net [82]	143.03	182.22	146.29	178.27
LLFlow [80]	193.17	239.41	253.55	300.53
Retinexformer [85]	112.51	242.16	145.11	524.20
RQ-LLIE [87]	125.45	195.81	136.03	238.77
DPRED [92]	<u>103.64</u>	<u>129.17</u>	<u>111.13</u>	**164.42**
AIDLM (Proposed)	**80.41**	**126.67**	**84.78**	<u>174.49</u>

generalisation performance [41, 30, 39, 27]. Table 4 presents the average LOE [30] metrics for non-reference datasets [41, 30, 39, 27]. Proposed method AIDLM consistently outperforms previous approaches as it is best in LOE metrics for three non reference datasets i-e LOE (80.41) for LIME, LOE (126.67) for NPE and LOE (84.78) for MEF. It is second best in LOE (174.49) for DICM dataset. The findings reveal that our proposed method outperforms others, highlighting its significant potential in the realm of LLIE. Fig.9 and 13 depicts enhanced image from the

datasets LIME [41], NPE [30], MEF [39] and DICM [27], respectively.

Fig.10 depicts enhanced image from the LIME[41] dataset. In the comparison of LLIE models, AIDLM consistently shows its strengths. Compared to EnlightenGAN, AIDLM preserves fine details and maintains natural color accuracy, avoiding reddish tint and uneven brightness. Over RUAS, AIDLM avoids over-saturation and maintains realistic color balance, whereas RUAS introduces excessive brightness and color distortion. AIDLM provides better noise reduction and maintains intricate details than Jiang et al., which exhibit noise and detail loss. Against Zero-DCE and Zero-DCE++, AIDLM offers superior brightness and contrast without artifacts, avoiding excessive brightening and color inaccuracies. Compared to RetinexNet and URetinex-Net, AIDLM preserves fine details, avoiding harsh contrasts, color shifts, and artifacts. AIDLM excels in noise reduction and maintains color fidelity compared to KinD and KinD++, avoiding color distortions and over-saturation. AIDLM provides a balanced enhancement over LLFlow, which fails in brightness enhancement. Against RetinexFormer and RQ-LLIE, AIDLM maintains detail and color balance, avoiding unrealistic artifacts and over-enhancement. Compared to DPRED, AIDLM offers better detail preservation and avoids over-saturation. Finally, AIDLM excels over UretinexNet, Zero-DCE Plus, and RUAS in maintaining finer details, realistic color balance, and avoiding over-enhancement.

Fig.11 depicts enhanced image from the NPE[30] dataset. In comparing AIDLM with other models on the NPE dataset, AIDLM excels in preserving fine details, maintaining natural color accuracy, and avoiding over-exposure. Unlike EnlightenGAN and RetinexNet, AIDLM avoids graininess, harsh contrasts, and color shifts. It provides better noise reduction and intricate detail maintenance compared to Jiang et al. and KinD. AIDLM offers superior brightness and contrast without artifacts, outperforming Zero-DCE and Zero-DCE++. It also maintains a balanced enhancement, avoiding over-saturation and unrealistic hues seen in RUAS and KinD++.

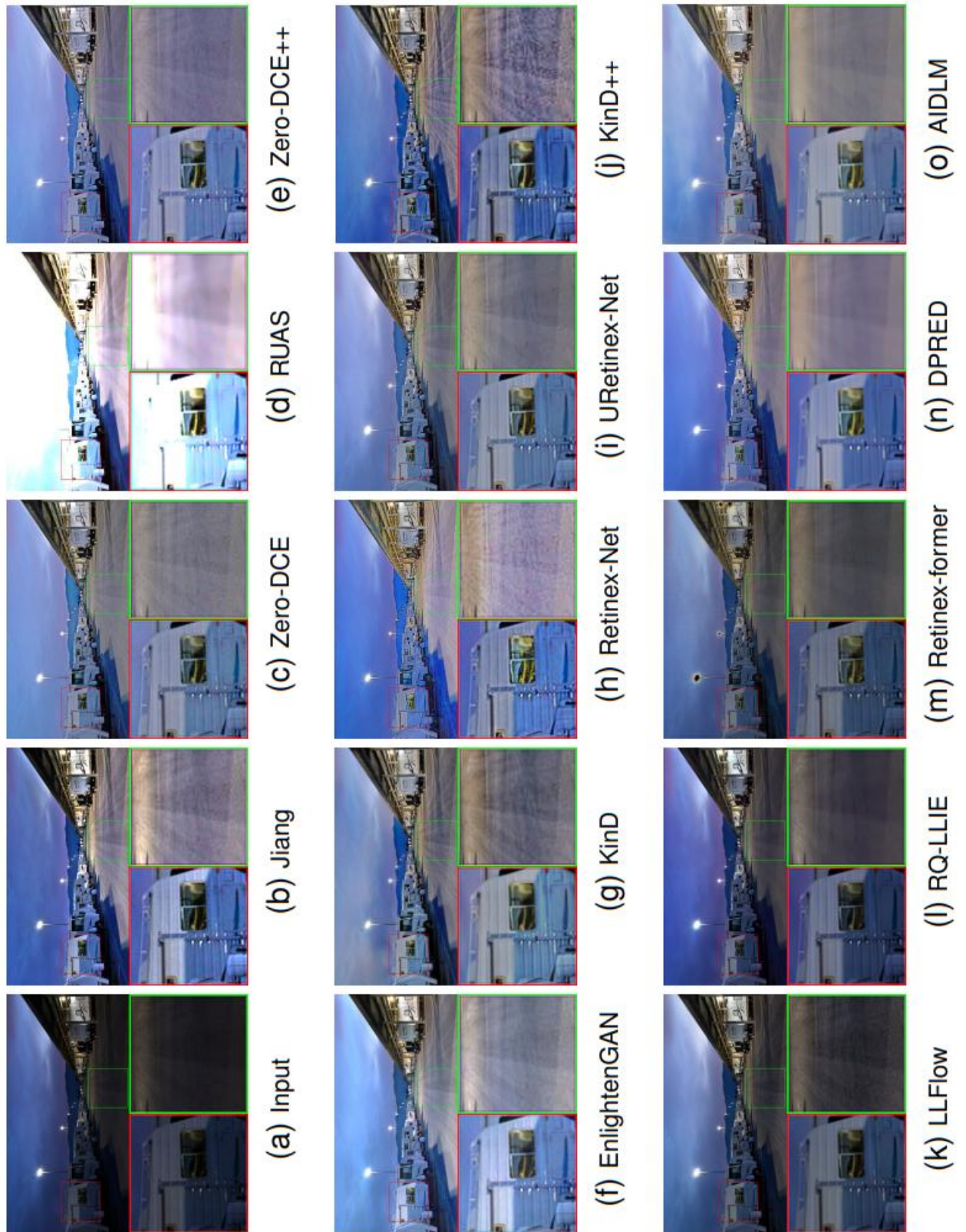**Figure 10:** Visual comparison on LIME dataset [41]. For better understanding, objects are zoomed in red and green boxes.
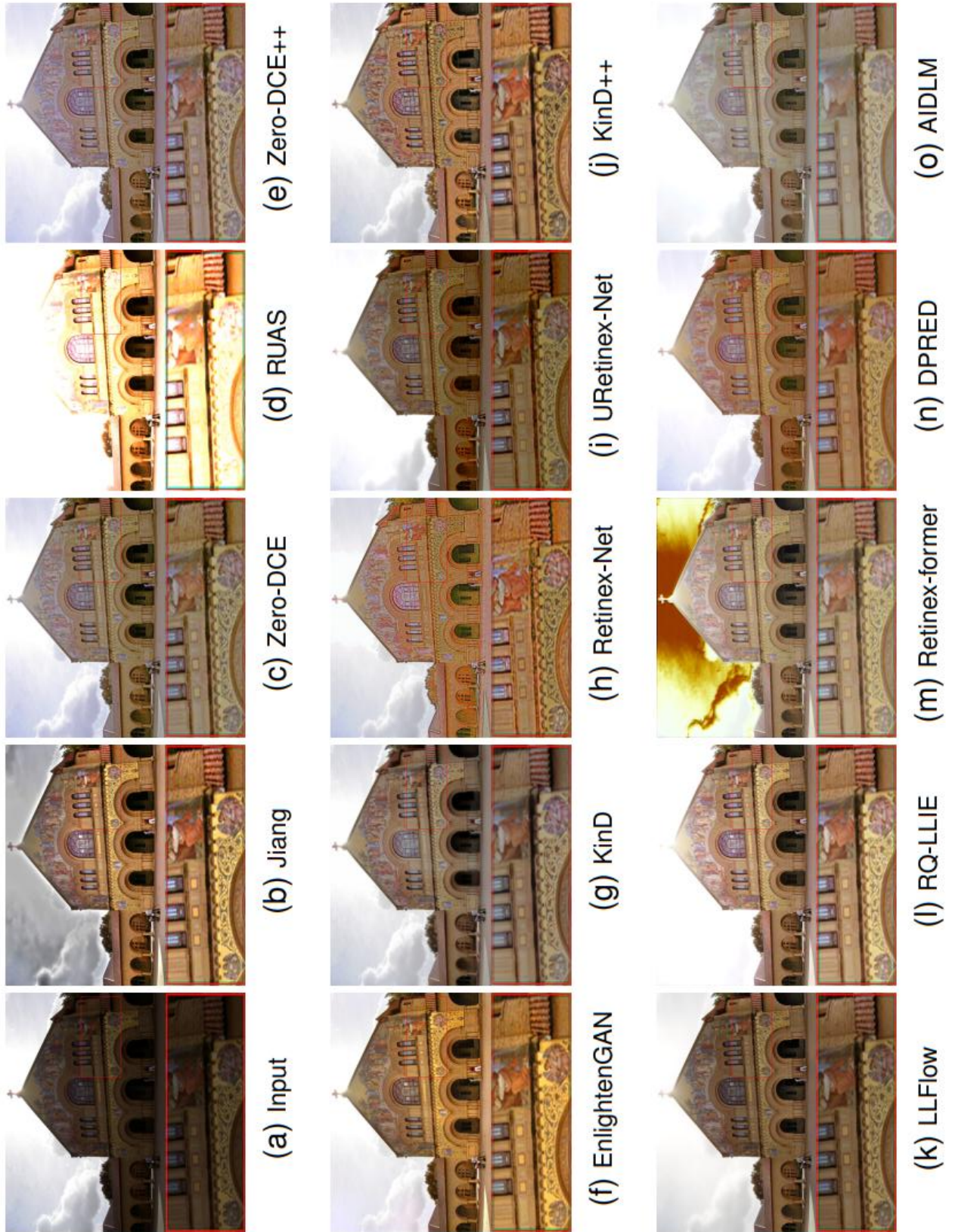
Figure 11: Visual comparison on NPE dataset [30]. For better understanding, objects are zoomed in red and green boxes.

Compared to LLFlow and RQ-LLIE, AIDLM provides superior brightness and color fidelity without losing details. Lastly, AIDLM ensures realistic enhancement, avoiding artifacts and inconsistencies present in Retinexformer and DPRED.

Figure 12: Visual comparison on MEF dataset [39]. For better understanding, objects are zoomed in red box.

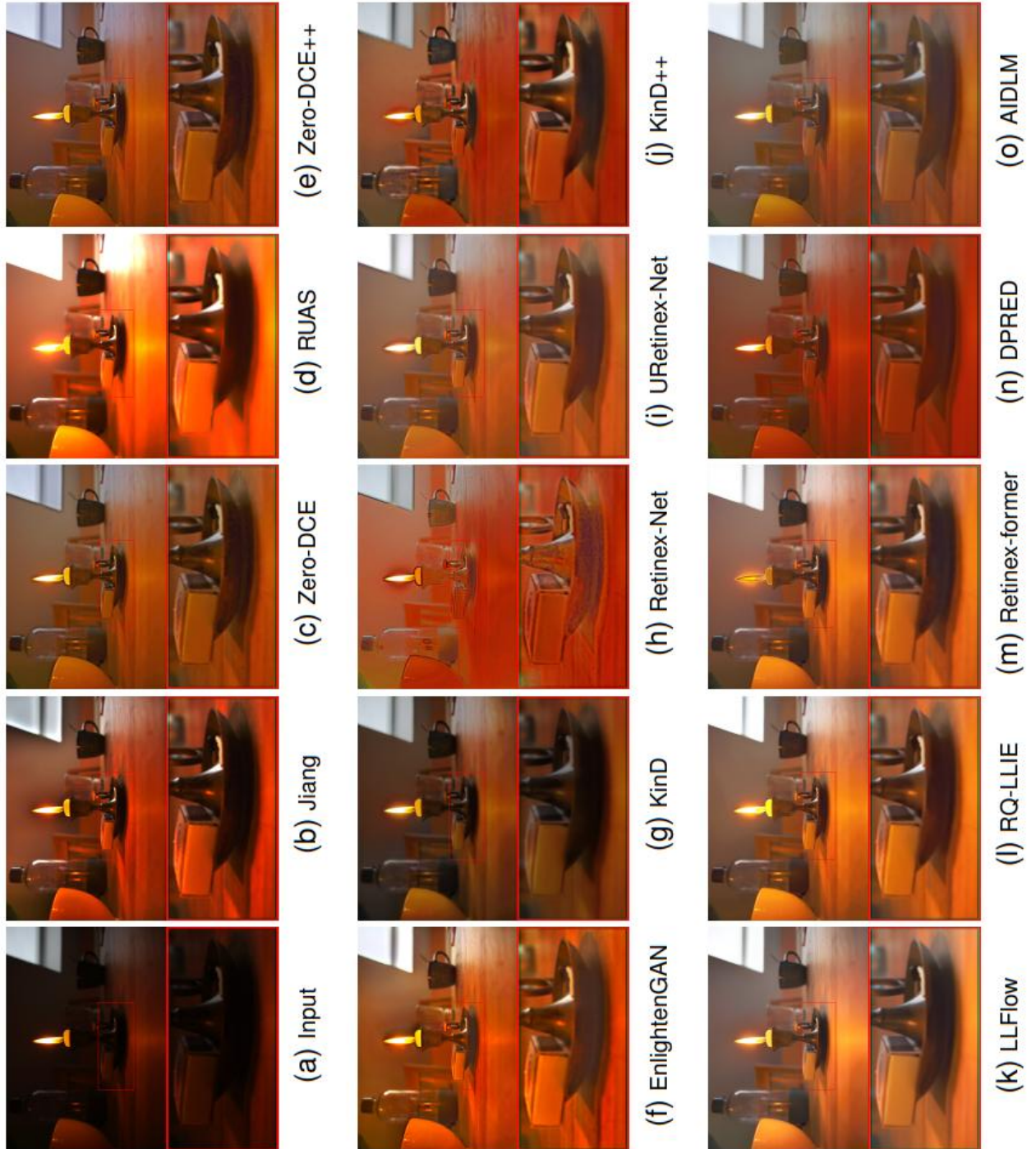Fig.12 depicts enhanced image from the MEF [39] dataset. The comparative analysis of unsupervised methods on MEF [39] dataset reveals distinct issues across various methods for LLIE. DPRED exhibited excessive noise and over-enhancement, leading to unnatural colors and loss of detail. EnGAN suffered from over-saturation and significant loss of detail in bright areas. Jiang's output had poor contrast and appeared washed out, losing details in shadows. KinD and KinD++ had insufficient enhancement, with the former being too dark and the latter showing over-saturation and uneven enhancement. LLFlow's image was flat with low contrast, failing to reveal details. Retinexformer introduced artifacts and uneven lighting, resulting in unnatural colors. RetinexNet produced a dark image lacking detail enhancement. RQ-LLIE suffered from over-enhancement, leading to unnatural colors and loss of detail. RUAS's image was excessively bright and over-saturated, causing detail loss. URetinexNet's image was overly smooth, losing fine details and appearing muted. Zero-DCE and Zero-DCE++ both failed to provide adequate enhancement, resulting in images that were too dark with insufficient brightness and detail. In contrast, AIDLM (proposed) demonstrated strong performance on MEF [39] dataset by providing balanced enhancement. The images produced maintained natural colors and well-preserved details in both dark and bright areas. The contrast and vibrancy were effectively managed, resulting in realistic and visually appealing images. This method stands out for its ability to enhance low-light images without introducing artifacts, excessive noise, or unnatural coloration's, highlighting its promising potential in LLIE applications.

Fig.13 depicts enhanced image from the DICM [27] dataset. In comparison to the other methods analyzed, AIDLM (Ours) addresses several key issues effectively. AIDLM excels in preserving fine details and natural color accuracy over EnlightenGAN. Additionally, AIDLM outperforms Jiang in noise reduction and maintaining intricate details without introducing artifacts. Moreover, compared to the Kind model, AIDLM shows better clarity and detail preservation, especially in the intricate designs. AIDLM surpasses the Kind Plus model in retaining finer
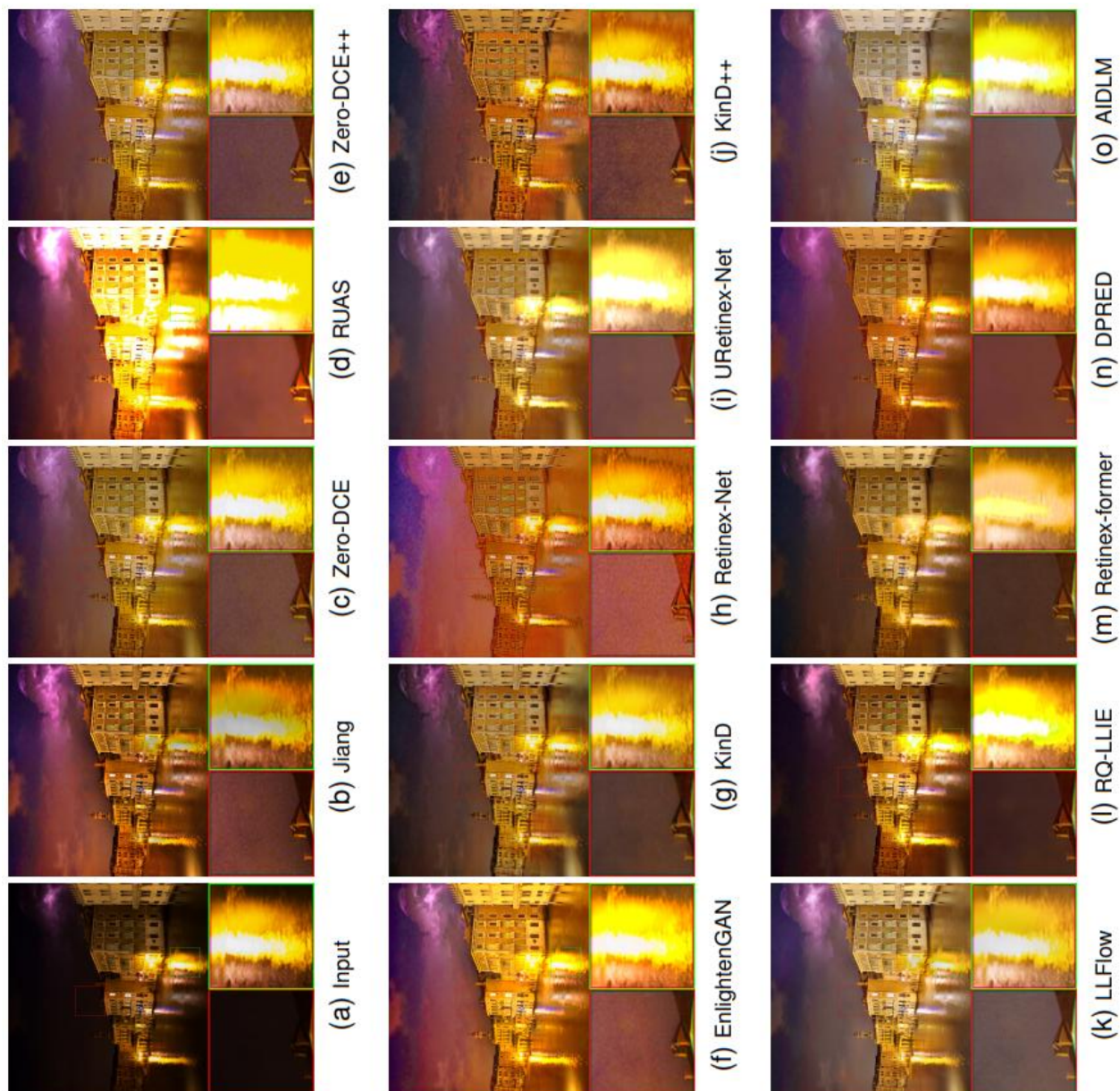
Figure 13: Visual comparison on DICM dataset [27]. For better understanding, objects are zoomed in red box.

details and avoiding over-saturation, providing a more realistic enhancement. Against LLFlow, AIDLM maintains superior detail and color fidelity, ensuring a balanced and natural enhancement. Furthermore, AIDLM outperforms RetinexFormer in providing a more consistent and natural enhancement, avoiding unrealistic artifacts. AIDLM offers better detail preservation and color accuracy compared to RetinexNet, ensuring a more realistic and visually pleasing result. In comparison to RQ-LLIE, AIDLM continues to excel in maintaining detail and achieving natural color balance without over-enhancement. AIDLM outperforms RUAS in avoiding over-exposure and maintaining a balanced enhancement that preserves both detail and natural colors. Finally, AIDLM excels over UretinexNet, Zero-DCE, and Zero-DCE Plus in maintaining fine details and realistic color balance without over-enhancement. However DPRED demonstrates superior brightness, contrast, and color vibrancy compared to AIDLM.

5.3 Computational Complexity

Table 5 provides a detailed comparison of the computational complexity of our proposed model, AIDLM, against various baseline methods. This comparison includes the number of trainable parameters and floating point operations (FLOPs) for images sized $600 \times 400 \times 3$. In terms of FLOPs, AIDLM sits at a median level, balancing computational demand with performance effectively. Our approach requires significantly fewer FLOPs than the supervised learning techniques LLFlow [80] and KinD++ [76]. However, it has more parameters and FLOPs than other models such as RetinexNet [53], URetinex-Net [82], KinD [58], DPRED [92], RQ-LLIE [87], and Retinexformer [85]. Despite this, AIDLM delivers competitive enhancement results with only a modest increase in computational cost. This highlights its ability to effectively balance computational complexity and performance, making it a practical choice for image enhancement tasks.

Table 5: Comparing computational complexity quantitatively, showing number of trainable parameters (#Parameters) in millions (M) as well as the floating-point operations per second (FLOPs) in gigaflops (G) or teraflops (T).

Method	#Parameters/M	FLOPs
EnlightenGAN [70]	8.637	61.010G
Jiang et al. [71]	4.317	36.717G
RUAS [73]	0.003	0.784G
Zero-DCE++ [77]	0.011	2.418G
RetinexNet [53]	0.555	135.997G
Zero-DCE [60]	0.079	19.008G
KinD [58]	8.160	127.768G
URetinex-Net [82]	0.340	208.497G
KinD++ [76]	8.275	2.532T
LLFlow [80]	17.420	1.050T
Retinexformer [85]	1.606	62.323G
RQ-LLIE [87]	11.383	593.499G
DPRED [92]	33.355	599.249G
AIDLM (Proposed)	33.853	651.414G

The detailed computational complexity of the AIDLM model and its individual components is outlined in Table 6. This table provides a breakdown of FLOPs and the number of trainable parameters for each part of AIDLM. The total FLOPs and parameters of AIDLM are presented alongside those of its constituent networks: Retinex Decom Net, EAI Net, EAR Net, and D Net.

Table 6: Detailed computational complexity of AIDLM and its components in terms of the number of

trainable parameters (#Parameters) in millions (M), and FLOPs in gigaflops (G).

Method	#Parameters/M	FLOPs
AIDLM	33.853	651.414G
Retinex Decom Net	0.005	0.009G
EAI Net	0.604	62.686G
EAR Net	0.604	62.807G
D Net	32.641	525.911G
Total	33.853	651.414G

Conclusion and Future Work

6.1 Conclusion

In this thesis, we propose a novel network, AIDLM, for LLIE. The proposed network consists of three modules: Retinex decomposition, enhancement attention, and denoising. The Retinex decomposition module combines the strengths of analytical models and learning methods, significantly improving illuminance and reflectance estimation. The enhancement networks, *EArnet* and *EAInet*, were developed using a UNet-like architecture [40]. Additionally, inspired by the SCA (spatio-channel attention) approach from [90], we integrated SCA into the UNet-like architecture. Our experimental results show that applying Retinex decomposition and enhancement to the V-channel in the HSV color space effectively prevents color distortion. The inclusion of SCA in the UNet-like architecture enhances the network's capability to capture fine details and contextual information. Furthermore, the denoising network efficiently reduces noise in the enhanced RGB image. Extensive experiments demonstrate that our proposed method surpasses state-of-the-art techniques and has strong generalization abilities.
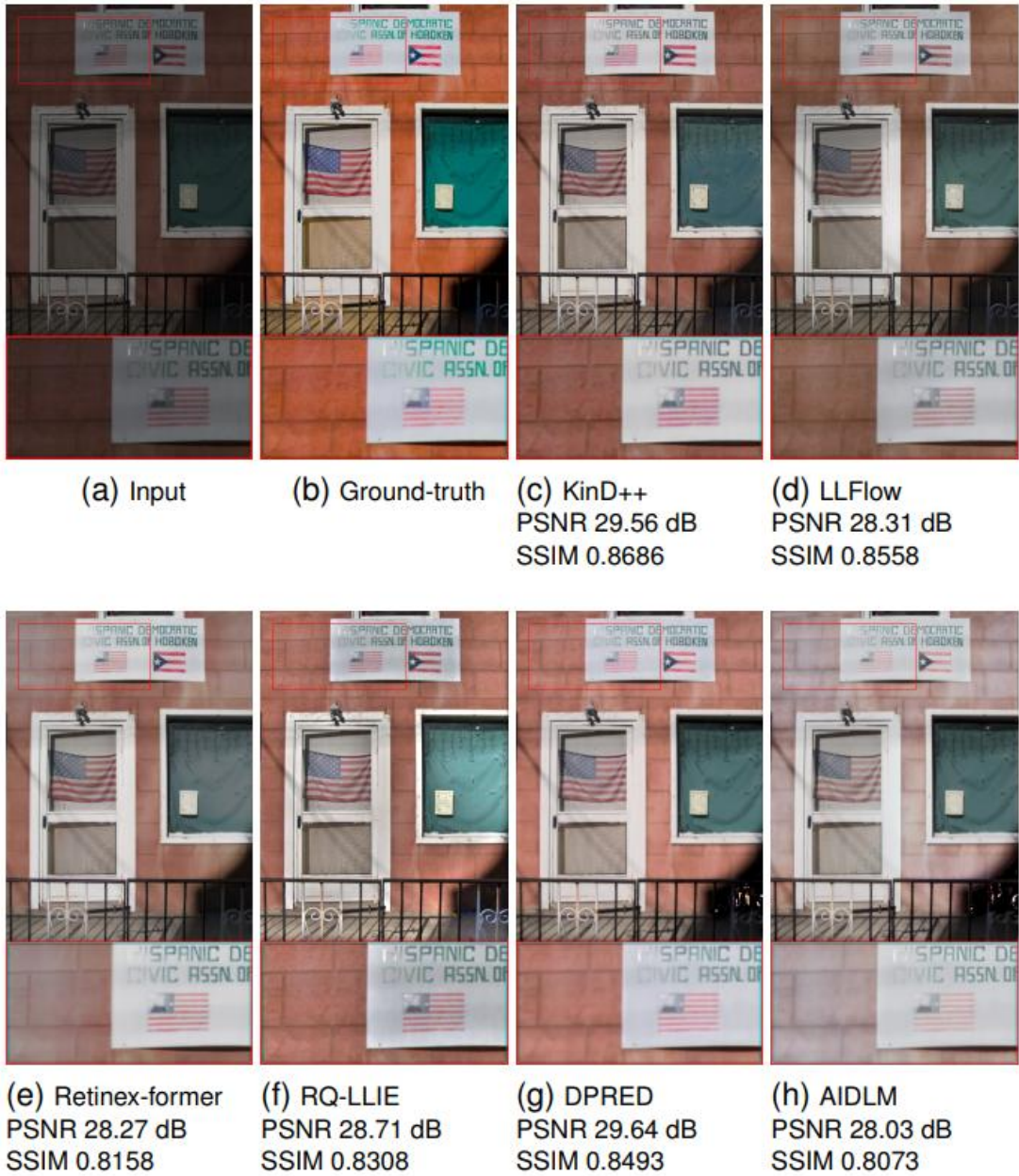
(a) Input

(b) Ground-truth

(c) KinD++
PSNR 29.56 dB
SSIM 0.8686

(d) LLFlow
PSNR 28.31 dB
SSIM 0.8558

(e) Retinex-former
PSNR 28.27 dB
SSIM 0.8158

(f) RQ-LLIE
PSNR 28.71 dB
SSIM 0.8308

(g) DPRED
PSNR 29.64 dB
SSIM 0.8493

(h) AIDLM
PSNR 28.03 dB
SSIM 0.8073

Figure 14: A visual comparison of an image for limitation case

6.2 Limitations

The primary drawback of proposed method is the extensive number of trainable parameters and FLOPs, as detailed in Table 6. By removing the Denoising module (*Dnet*), the trainable parameters and FLOPs are significantly reduced to 1.21207M and 125.50327G, respectively.

In Fig.14, we show an example where our method's results may not be optimal. The AIDLM

model, with 28.03dB PSNR and 0.8073 SSIM, performs reasonably well in reducing noise and maintaining structural details but requires improvement. It is outshined in color enhancement by models such as kind_plus (PSNR: 29.56, SSIM: 0.8686) [76], llflow (PSNR: 28.31, SSIM: 0.8558) [80], Retinexformer (PSNR: 28.27, SSIM: 0.8158) [85], rq_llie (PSNR: 28.71, SSIM: 0.8308) [87], and DPRED (PSNR: 29.64, SSIM: 0.8493) [92]. Although the AIDLM model is competent in managing noise and preserving structural integrity, it falls short in effectively retaining or improving colors. This results in less visually appealing images when compared to these higher-performing models. This observation highlights the need for further enhancement of the AIDLM model to improve color accuracy and overall visual quality.

6.3 Future Work

To advance low-light image enhancement, future work could focus on:

- Combining Vision Transformers (ViTs) and convolutional neural network (CNN) to leverage the strengths of both architectures for improved feature extraction and representation.

- Implementing advanced noise modeling tailored to extreme low-light conditions to reduce noise artifacts and enhance image quality.

Bibliography

[1]   Edwin H Land. "The retinex". In: *Ciba Foundation Symposium-Colour Vision: Physiology and Experimental Psychology*. Wiley Online Library. 1965, pp. 217–227.

[2]   K. McLaren. "XIII – The development of the CIE 1976 (L* a* b*) uniform colour space and colour-difference formula". In: *Journal of the Society of Dyers and Colourists* 92.9 (1976), pp. 338–341.

[3]   Edwin H Land. "The Retinex theory of color vision". In: *Scientific American* 237.6 (1977), pp. 108–129.

[4]   David CC Wang, Anthony H Vagnucci, and Ching-Chung Li. "Digital image enhancement: a survey". In: *Computer vision, graphics, and image processing* 24.3 (1983), pp. 363–381.

[5]   S.M. Pizer et al. "Contrast-limited adaptive histogram equalization: speed and effectiveness". In: *Proc. 1st Conf. Visualiz. Biomed. Comput.* 1990, pp. 337–345. DOI: 10.1109/VBC.1990.109340.

[6]   Leonid I Rudin, Stanley Osher, and Emad Fatemi. "Nonlinear total variation based noise removal algorithms". In: *Physica D: Nonlinear Phenomena* 60.1-4 (1992), pp. 259–268.

[7]   Karel Zuiderveld. "Contrast limited adaptive histogram equalization". In: *Graphics gems IV*. 1994, pp. 474–485.

[8] Zia-ur Rahman, Daniel J Jobson, and Glenn A Woodell. "Multi-scale retinex for color image enhancement". In: *Proceedings of 3rd IEEE international conference on image processing*. Vol. 3. IEEE. 1996, pp. 1003–1006.

[9] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. "A multiscale Retinex for bridging the gap between color images and the human observation of scenes". In: *IEEE Transactions on Image Processing* 6.7 (1997), pp. 965–976.

[10] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. "Properties and performance of a center/surround Retinex". In: *IEEE Transactions on Image Processing* 6.3 (1997), pp. 451–462.

[11] Yeong-Taeg Kim. "Contrast enhancement using brightness preserving bi-histogram equalization". In: *IEEE transactions on Consumer Electronics* 43.1 (1997), pp. 1–8.

[12] A Raji et al. "A gray-level transformation-based method for image enhancement". In: *Pattern Recognition Letters* 19.13 (1998), pp. 1207–1212.

[13] C Rezk-Salama et al. "Non-linear registration of pre-and intraoperative volume data based on piecewise linear transformations". In: *Vision, Modeling and Visualization (VMV)*. Vol. 99. 1999, p. 365.

[14] Hui Zhu, Francis HY Chan, and Francis K Lam. "Image contrast enhancement by constrained local histogram equalization". In: *Computer vision and image understanding* 73.2 (1999), pp. 281–290.

[15] Ronald A Rensink. "The dynamic representation of scenes". In: *Visual cognition* 7.1-3 (2000), pp. 17–42.

[16] Maurizio Corbetta and Gordon L Shulman. "Control of goal-directed and stimulus-driven attention in the brain". In: *Nature reviews neuroscience* 3.3 (2002), pp. 201–215.

[17] Ron Kimmel et al. "A variational framework for retinex". In: *International Journal of computer vision* 52 (2003), pp. 7–23.

[18] Zhou Wang et al. "Image quality assessment: from error visibility to structural similarity". In: *IEEE Transactions on Image Processing* 13.4 (2004), pp. 600–612.

[19] Jia Deng et al. "Imagenet: A large-scale hierarchical image database". In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee. 2009, pp. 248–255.

[20] Hugo Larochelle and Geoffrey E Hinton. "Learning to combine foveal glimpses with a third-order Boltzmann machine". In: *Advances in neural information processing systems* 23 (2010).

[21] Vladimir Bychkovsky et al. "Learning photographic global tonal adjustment with a database of input/output image pairs". In: *Proc. IEEE Conf. Comput. Vis. Pattern Recog.* 2011, pp. 97–104.

[22] Wenye Ma et al. "An L 1-based variational model for Retinex theory and its application to medical images". In: *CVPR 2011*. IEEE. 2011, pp. 153–160.

[23] Michael K Ng and Wei Wang. "A total variation model for retinex". In: *SIAM Journal on Imaging Sciences* 4.1 (2011), pp. 345–365.

[24] Jari Korhonen and Junyong You. "Peak signal-to-noise ratio revisited: Is simple beautiful?" In: *2012 Fourth international workshop on quality of multimedia experience*. IEEE. 2012, pp. 37–38.

[25] Wenye Ma, Stanley Osher, et al. "A TV Bregman iterative model of Retinex theory". In: *Inverse Probl. Imaging* 6.4 (2012), pp. 697–708.

[26] Xueyang Fu et al. "A variational framework for single low light image enhancement using bright channel prior". In: *2013 IEEE global conference on signal and information processing*. IEEE. 2013, pp. 1085–1088.

[27] Chulwoo Lee, Chul Lee, and Chang-Su Kim. "Contrast enhancement based on layered difference representation of 2D histograms". In: *IEEE Transactions on Image Processing* 22.12 (2013), pp. 5372–5384.

[28] Xujia Qin et al. "Structured light image enhancement algorithm based on retinex in HSV color space". In: *Journal of Computer-Aided Design & Computer Graphics* 25.4 (2013), pp. 488–493.

[29] V Rajamani, P Babu, and S Jaiganesh. "A review of various global contrast enhancement techniques for still images using histogram modification framework". In: *International Journal of Engineering Trends and Technology* 4.4 (2013), pp. 1045–1048.

[30] Shuhang Wang et al. "Naturalness preserved enhancement algorithm for non-uniform illumination images". In: *IEEE Transactions on Image Processing* 22.9 (2013), pp. 3538–3548.

[31] Jimmy Ba, Volodymyr Mnih, and Koray Kavukcuoglu. "Multiple object recognition with visual attention". In: *arXiv preprint arXiv:1412.7755* (2014).

[32] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate". In: *arXiv preprint arXiv:1409.0473* (2014).

[33] Tsung-Yi Lin et al. "Microsoft coco: Common objects in context". In: *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*. Springer. 2014, pp. 740–755.

[34] Karol Gregor et al. "Draw: A recurrent neural network for image generation". In: *International conference on machine learning*. PMLR. 2015, pp. 1462–1471.

[35] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. "Spatial transformer networks". In: *Advances in neural information processing systems* 28 (2015).

[36] Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *Int. Conf. Learn. Represent.* Ed. by Yoshua Bengio and Yann LeCun. 2015, pp. 1–15.

[37] Di Li et al. "A retinex algorithm for image enhancement based on recursive bilateral filtering". In: *2015 11th International Conference on Computational Intelligence and Security (CIS)*. IEEE. 2015, pp. 154–157.

[38] Zhetong Liang, Weijian Liu, and Ruohe Yao. "Contrast enhancement by nonlinear diffusion filtering". In: *IEEE Transactions on Image Processing* 25.2 (2015), pp. 673–686.

[39] Kede Ma, Kai Zeng, and Zhou Wang. "Perceptual quality assessment for multi-exposure image fusion". In: *IEEE Transactions on Image Processing* 24.11 (2015), pp. 3345–3356.

[40] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional networks for biomedical image segmentation". In: *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.* 2015, pp. 234–241.

[41] Xiaojie Guo, Yu Li, and Haibin Ling. "LIME: Low-light image enhancement via illumination map estimation". In: *IEEE Transactions on Image Processing* 26.2 (2016), pp. 982–993.

[42] Joseph Redmon et al. "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.

[43] Bolun Cai et al. "A joint intrinsic-extrinsic prior model for Retinex". In: *Proc. IEEE Int. Conf. Comput. Vis.* 2017, pp. 4000–4009.

[44] Qifeng Chen, Jia Xu, and Vladlen Koltun. "Fast image processing with fully-convolutional networks". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2017, pp. 2497–2506.

[45] Gabriel Eilertsen et al. "HDR image reconstruction from a single exposure using deep CNNs". In: *ACM transactions on graphics (TOG)* 36.6 (2017), pp. 1–15.

[46] Andrey Ignatov et al. "Dslr-quality photos on mobile devices with deep convolutional networks". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 3277–3285.

[47] Jiahang Liu et al. "An efficient contrast enhancement method for remote sensing images". In: *IEEE Geoscience and Remote Sensing Letters* 14.10 (2017), pp. 1715–1719.

[48] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. "LLNet: A deep autoencoder approach to natural low-light image enhancement". In: *Pattern Recognition* 61 (2017), pp. 650–662.

[49] Liang Shen et al. "Msr-net: Low-light image enhancement using deep convolutional network". In: *arXiv preprint arXiv:1711.02488* (2017).

[50] Yanrong Guo and Tao Chen. "Semantic segmentation of RGBD images based on deep depth regression". In: *Pattern Recognition Letters* 109 (2018), pp. 55–64.

[51] Jie Hu, Li Shen, and Gang Sun. "Squeeze-and-excitation networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 7132–7141.

[52] Feifan Lv et al. "MBLLEN: Low-Light Image/Video Enhancement Using CNNs." In: *Proc. Brit. Mach. Vis. Conf.* 2018, p. 13.

[53] Chen Wei et al. "Deep Retinex Decomposition for Low-Light Enhancement". In: *Proc. Brit. Mach. Vis. Conf.* 2018, p. 12.

[54] Sanghyun Woo et al. "Cbam: Convolutional block attention module". In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 3–19.

[55] Richard Zhang et al. "The unreasonable effectiveness of deep features as a perceptual metric". In: *Proc. IEEE Conf. Comput. Vis. Pattern Recog.* 2018, pp. 586–595.

[56]   Wenqi Ren et al. "Low-light image enhancement via a deep hybrid network". In: *IEEE Transactions on Image Processing* 28.9 (2019), pp. 4364–4375.

[57]   Yangming Shi, Xiaopo Wu, and Ming Zhu. "Low-light image enhancement algorithm based on retinex and generative adversarial network". In: *arXiv preprint arXiv:1906.06027* (2019).

[58]   Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. "Kindling the darkness: A practical low-light image enhancer". In: *Proc. 27th ACM Int. Conf. Multimedia.* 2019, pp. 1632–1640.

[59]   Yulun Zhang et al. "Residual non-local attention networks for image restoration". In: *arXiv preprint arXiv:1903.10082* (2019).

[60]   Chunle Guo et al. "Zero-reference deep curve estimation for low-light image enhancement". In: *Proc. IEEE Conf. Comput. Vis. Pattern Recog.* 2020, pp. 1780–1789.

[61]   Zhang Hongying and Zhao Jindong. "RetinexNet low illumination image enhancement algorithm in HSV space". In: *Laser Optoelectron. Prog* 57 (2020), p. 201504.

[62]   Tong Liu et al. "Lane detection in low-light conditions using an efficient data enhancement: Light conditions style transfer". In: *2020 IEEE intelligent vehicles symposium (IV).* IEEE. 2020, pp. 1394–1399.

[63]   Kun Lu and Lihong Zhang. "TBEFN: A two-branch exposure-fusion network for low-light image enhancement". In: *IEEE Transactions on Multimedia* 23 (2020), pp. 4093–4105.

[64]   Feifan Lv, Bo Liu, and Feng Lu. "Fast enhancement for non-uniform illumination images using light-weight CNNs". In: *Proceedings of the 28th ACM International Conference on Multimedia.* 2020, pp. 1450–1458.

[65] Li-Wen Wang et al. "Lightening network for low-light image enhancement". In: *IEEE Transactions on Image Processing* 29 (2020), pp. 7984–7996.

[66] Jun Xu et al. "STAR: A structure and texture aware Retinex model". In: *IEEE Transactions on Image Processing* 29 (2020), pp. 5022–5037.

[67] Anqi Zhu et al. "Zero-shot restoration of underexposed images via robust retinex decomposition". In: *2020 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE. 2020, pp. 1–6.

[68] Claudine Badue et al. "Self-driving cars: A survey". In: *Expert systems with applications* 165 (2021), p. 113816.

[69] Ao Chen et al. "Knowledge graph-based image recognition transfer learning method for on-orbit service manipulation". In: *Space: Science & Technology* (2021).

[70] Yifan Jiang et al. "EnlightenGAN: Deep light enhancement without paired supervision". In: *IEEE Transactions on Image Processing* 30 (2021), pp. 2340–2349.

[71] Zhuqing Jiang et al. "A switched view of Retinex: deep self-regularized low-light image enhancement". In: *Neurocomputing* 454 (2021), pp. 361–372.

[72] Jinjiang Li, Xiaomei Feng, and Zhen Hua. "Low-light image enhancement via progressive-recursive network". In: *IEEE Transactions on Circuits and Systems for Video Technology* 31.11 (2021), pp. 4227–4240.

[73] Risheng Liu et al. "Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement". In: *Proc. IEEE Conf. Comput. Vis. Pattern Recog.* 2021, pp. 10561–10570.

[74] Long Ma et al. "Learning deep context-sensitive decomposition for low-light image enhancement". In: *IEEE Transactions on Neural Networks and Learning Systems* 33.10 (2021), pp. 5666–5680.

[75] Wenhan Yang et al. "Sparse gradient regularized deep Retinex network for robust low-light image enhancement". In: *IEEE Transactions on Image Processing* 30 (2021), pp. 2072–2086.

[76] Yonghua Zhang et al. "Beyond brightening low-light images". In: *International Journal of Computer Vision* 129.4 (2021), pp. 1013–1037.

[77] Chongyi Li, Chunle Guo, and Chen Change Loy. "Learning to Enhance Low-Light Image via Zero-Reference Deep Curve Estimation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.8 (2022), pp. 4225–4238.

[78] Risheng Liu et al. "Learning with nested scene modeling and cooperative architecture search for low-light vision". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 01 (2022), pp. 1–17.

[79] Xiaoqian Lv et al. "BacklitNet: A Dataset and Network for Backlit Image Enhancement". In: *Computer Vision and Image Understanding* 218 (2022), p. 103403.

[80] Yufei Wang et al. "Low-Light Image Enhancement with Normalizing Flow". In: *AAAI Conf. Artif. Intell.* Vol. 36. 3. June 2022, pp. 2604–2612.

[81] Syed Waqas Zamir et al. "Learning Enriched Features for Fast Image Restoration and Enhancement". In: *arXiv e-prints* (2022), arXiv–2205.

[82] Wenhui Wu et al. "URetinex-Net: Retinex-Based Deep Unfolding Network for Low-Light Image Enhancement". In: *Proc. IEEE Conf. Comput. Vis. Pattern Recog.* 2022, pp. 5901–5910.

[83] Kai Zhang et al. "Plug-and-Play Image Restoration With Deep Denoiser Prior". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.10 (2022), pp. 6360–6376.

[84] Zhao Zhang et al. "Deep color consistent network for low-light image enhancement". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022, pp. 1899–1908.

[85] Yuanhao Cai et al. "Retinexformer: One-stage Retinex-based Transformer for Low-light Image Enhancement". In: *Proc. IEEE Int. Conf. Comput. Vis.* Oct. 2023, pp. 12504–12513.

[86] Yongqiang Chen et al. "A depth iterative illumination estimation network for low-light image enhancement based on retinex theory". In: *Scientific Reports* 13.1 (2023), p. 19709.

[87] Yunlong Liu et al. "Low-Light Image Enhancement with Multi-Stage Residue Quantization and Brightness-Aware Attention". In: *Proc. IEEE Int. Conf. Comput. Vis.* Oct. 2023, pp. 12140–12149.

[88] Yunlong Liu et al. "Low-light image enhancement with multi-stage residue quantization and brightness-aware attention". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 12140–12149.

[89] Manli Wang, Jiayue Li, and Changsen Zhang. "Low-Light Image Enhancement by Deep Learning Network for Improved Illumination Map". In: *Computer Vision and Image Understanding* 232 (2023), p. 103681.

[90] Ahsan B Bakht et al. "MuLA-GAN: Multi-Level Attention GAN for Enhanced Underwater Visibility". In: *Ecological Informatics* 81 (2024), p. 102631.

[91] Hai Jiang, Yang Ren, and Songchen Han. "Revisiting coarse-to-fine strategy for low-light image enhancement with deep decomposition guided training". In: *Computer Vision and Image Understanding* 241 (2024), p. 103952.

[92] Xiaofang Li et al. "Deep parametric Retinex decomposition model for low-light image enhancement". In: *Computer Vision and Image Understanding* 241 (2024), p. 103948.

[93]   Dehuan Zhang et al. "Robust underwater image enhancement with cascaded multi-level sub-networks and triple attention mechanism". In: *Neural Networks* 169 (2024), pp. 685–697.