A MAJOR PROJECT – II REPORT

ON

RETINEX-BASED LOW-LIGHT IMAGE ENHANCEMENT WITH EDGE-AWARE LEARNING

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE OF

MASTER OF TECHNOLOGY

IN

COMPUTER SCIENCE & ENGINEERING

Submitted By

Gourav

24/CSE/13

Under The Supervision Of

Prof. Aruna Bhat

(Professor)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi–110042

May 2026

DEPT. OF COMPUTER SCIENCE AND ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, **Gourav (24/CSE/13)**, hereby certify that the work which is being presented in the major project report II entitled **"Retinex-Based Low-Light Image Enhancement with Edge-Aware Learning"** in partial fulfillment of the requirements for the award of the Degree of Master of Technology, submitted in the Department of Computer Science and Engineering, Delhi Technological University, is an authentic record of my own work carried out during the period from August 2024 to May 2025 under the supervision of Prof. Aruna Bhat.

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

_____

Candidate's Signature

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

_____ _____

Signature of Supervisor(s) **Signature of External Examiner**

DEPT. OF COMPUTER SCIENCE AND ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project titled "Retinex-Based Low-Light Image Enhancement with Edge-Aware Learning", submitted by Gourav, Roll No. 24/CSE/13, Department of Computer Science & Engineering, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology (M.Tech) in Computer Science and Engineering is a genuine record of the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree to this University or elsewhere.

Place : New Delhi
Date :

Dr. Aruna Bhat
Professor
Department of Computer Science and Engineering
Delhi Technological University

Acknowledgement

I am grateful to Prof. Anil Singh Parihar, Head of the Department of Computer Science and Engineering, Delhi Technological University (Formerly Delhi College of Engineering), New Delhi, and all other faculty members of our department for their astute guidance, constant encouragement, and sincere support for this project work.

I wish to express my profound gratitude and deep regard to my project mentor Prof. Aruna Bhat, for their exemplary guidance, valuable feedback, and constant encouragement throughout the project. Their valuable suggestions were of immense help throughout my work. Their perspective criticism kept me working to make this project much better. Working under their supervision was an extremely knowledgeable experience for me.

I would also like to sincerely thank all my friends and colleagues for their help and support during this journey.

Gourav
(24/CSE/13)

Abstract

Improvement of low-light images is an essential and practically relevant problem in the domain of computer vision, having practical ramifications on the operation of self-navigating vehicles, smart surveillance cameras, medical imaging, and mobile camera phones. Low-light images are characterized by a set of degradation artifacts that includes extremely low contrast, because of inadequate photon collection, excessive noise levels, color distortion due to erroneous white-balance adjustment, and the loss of fine detail structures.

The main contributions of this paper are introduced in **RLEA-Net**, a Retinex-based deep learning framework that utilizes physics-motivated Retinex decomposition along with an explicit edge-aware learning mechanism. The RLEA-Net takes an input low-light image and decomposes it into its illumination and reflectance layers by a learned network, Decom-Net. In addition, an Illumination Enhancement Network, denoted as Illum-Net, uses multi-scale dilated convolutional layers to predict a residual map on top of the estimated illumination map. On the other hand, the reflectance layer undergoes denoising and fine texture restoration through a Reflectance Restoration Network (Ref-Net) architecture based on the U-Net architecture. However, the core innovation lies in designing an Edge-Aware Module (EAM), which utilizes Sobel kernel initialized anisotropic filter banks as a fixed-kernel depthwise convolutional layer to generate channel-wise edge maps that influence the feature transformation within Ref-Net.

One of the important principles in the design is that of *gradient isolation*. In other words, the loss related to reconstruction (such as Charbonnier, VGG perceptual, SSIM, and gradient fidelity losses) are only computed for the Illum-Net and Ref-Net, but DecomNet is trained through a loss term called Retinex consistency, which is based on synthetic targets created analytically from the reference image. This ensures that there is no decomposition collapse problem faced by previous networks. The unique loss is gradient-weighted illumination smoothness loss.

Tests performed using the LOL standard dataset (485 training and 15 test examples) show that RLEA-Net obtains PSNR of 20.41 dB and SSIM of 0.796 for eval15, having around 8.86 million trainable parameters. Although the results themselves are not impressive in comparison with the biggest published architectures, the performance outdoes RetinexNet by 3.64 dB while being close to Zero-DCE and KinD in terms of edge fidelity. An ablation study proves that all three contributions of RLEA-Net contribute equally to the performance — EAM (+1.24 dB), gradient-based illumination loss, and Retinex consistency supervision.

Contents

List of Tables

List of Figures

List of Symbols, Abbreviations, and Nomenclature

Symbol / Abbreviation	Meaning
AMP	Automatic Mixed Precision
CA	Channel Attention
CNN	Convolutional Neural Network
CV	Computer Vision
Decom-Net	Decomposition Network
DL	Deep Learning
EAM	Edge-Aware Module
GAN	Generative Adversarial Network
GPU	Graphics Processing Unit
HE	Histogram Equalization
Illum-Net	Illumination Enhancement Network
LOL	Low-Light (benchmark dataset)
LR	Learning Rate
MSR	Multi-Scale Retinex
PSNR	Peak Signal-to-Noise Ratio
Ref-Net	Reflectance Restoration Network
RLEA-Net	Retinex-based Low-light Enhancement with Edge-Aware Learning Network
SSIM	Structural Similarity Index Measure
SOTA	State-Of-The-Art
TV	Total Variation
VGG	Visual Geometry Group (network)
R	Reflectance Component $[B, 3, H, W]$
L	Illumination Component $[B, 1, H, W]$
$\mathscr{L}_{\text{total}}$	Total Training Loss

S_{low}	Input Low-Light Image
$\hat{S}$	Enhanced Output Image

Chapter 1

Introduction

1.1 Background and Motivation

The capability of sensing visual data and interpreting it accurately regardless of changes in illumination levels is essential to current computer vision applications. Visual data captured under low-light conditions such as poorly illuminated indoor images, night-time outdoor shots as well as video surveillance in poorly illuminated areas display unique combinations of degradation effects: extreme lack of contrast, presence of noise (predominantly shot and read noise that is accentuated due to the gain process), erroneous colour balance and loss of crucial structural information.

Far from being purely aesthetic problems, these deficiencies can be quite serious. In autonomous vehicle systems, the inability to detect objects due to low visibility in the dark is a matter of life or death. In surveillance applications, under-exposure leads to a loss of recognition accuracy due to noise. In medical imaging applications, inadequate illumination makes it difficult to see tissue borders. And mobile photography, which has completely revolutionized how people communicate visually, suffers from similar deficiencies in the dark.

Classic methods like Histogram Equalization, CLAHE, and the classic Retinex method of Land can help but are constrained due to their inability to handle spatially varying lighting, sensitivity to noise, and generation of artifacts along the edge boundaries with high contrast. Deep learning methods have revolutionized the domain, with RetinexNet, KinD, and MIRNet-v2 gaining increasing benchmark numbers with successive generations. But the irony is that the more powerful the deep learning model used, the more difficult it is to deploy it at the edge due to its millions of parameters.

The thesis is driven by two interlocking insights: Firstly, Retinex decomposition provides physical insight into an inductive bias which can be leveraged to develop more interpretable and generalizeable low light enhancement algorithms using neural networks.

Secondly, retention of structural boundaries in low light situations, where noise resides in the same frequency space as true boundaries, is something existing approaches do not address sufficiently.

1.2 Problem Statement

Although major progress has been made in image enhancement under low light conditions, using deep neural networks, there are still some issues that have not been properly resolved yet. The top-performing network architecture uses highly complex structures to reach their remarkable results which make it difficult to apply on a device that lacks computing power. Although Retinex-based frameworks like RetinexNet and KinD provide physical meaning to their decompositions, they still experience two issues.

This is especially true for low-light enhancement due to the fact that noise in dark images shares the same high frequency range as structural elements, so that general convolutional networks may fail to separate between signal and noise. Without explicitly incorporating edge- preserving inductive bias into such a network, over-smoothing of real edges and noise as false textures become inevitable.

Another important yet underexplored aspect of image enhancement using Retinex theory pipeline is the issue of training instability when back-propagating loss gradients from the reconstruction stage of $R \times L$ back to the decomposition stage. This results in the decomposition collapse phenomenon, where the network generates meaningless degenerated decompositions in order to optimize the loss.

1.3 Objectives of the Thesis

Main Objective Develop and evaluate a physically inspired, edge-aware deep learning approach for image enhancement under low lighting conditions.

Secondary Objectives This research problem can be formulated in the following ways:

1. Develop a decomposition network (Decom-Net) with gradient isolation, which disallows the backpropagation of reconstruction losses to the DecomNet and is supervised through the retinex-consistency loss applied on physically inspired targets.
2. Propose an illumination enhancement network (Illum-Net) that uses multi-scale dilated convolutions to predict a residual illumination correction.
3. Propose a U-Net-based reflectance restoration network (Ref-Net) that uses skip-connections to recover reflectance detail from noisy low-light reflectances.

4. Propose and integrate the edge-aware module (EAM), consisting of a Sobel-initialized anisotropic filter bank implemented through a fixed-kernel depthwise convolution layer, with a learnable residual scaling factor initialized as 0 and increased during training.
5. Define a multi-objective loss function that includes Charbonnier reconstruction, VGG perceptual, SSIM structural, gradient fidelity, gradient-illumination weighted, and retinex consistency losses.
6. Train models in a resolution progressive manner and study the influence of such a training strategy.
7. Rigorously evaluate on the LOL benchmark with quantitative (PSNR, SSIM) and qualitative analysis, and conduct comprehensive ablation studies quantifying each component.

1.4 Scope of the Work

The proposed study aims at supervised single image enhancement under low light conditions using the LOL database. The coverage involves the RLEA-Net network design, gradient isolation-based learning pipeline, multi-objective loss function, gradual resolution training, performance evaluations, and component-wise ablation studies. This study does not deal with video enhancement, unsupervised learning, and hardware-aware performance analysis.

1.5 Thesis Organization

- **Chapter 1** (this chapter): Introduction, motivation, problem statement, objectives, and scope.
- **Chapter 2**: Literature Survey covering classical methods, deep learning approaches, Retinex-based networks, edge-aware methods, and loss function design.
- **Chapter 3**: Proposed Methodology — full RLEA-Net architecture with gradient isolation design, each sub-network, the multi-objective loss, and training strategy.
- **Chapter 4**: Results and Analysis — quantitative comparison, qualitative visual analysis with figures, training dynamics, ablation studies, and limitations discussion.
- **Chapter 5**: Conclusion and Future Scope.

Chapter 2

Literature Survey

The problem of low light image enhancement has motivated researchers in the field for many years. This chapter provides an overview of seminal and current literature in four topics: classical image processing techniques, deep learning models, Retinex theory, and its neural variations, and edge aware techniques.

2.1 Low-Light Image Enhancement Algorithms

2.1.1 Traditional Techniques

In Histogram Equalization (HE), pixel values are redistributed so as to create a uniform distribution [7] of their intensity levels. Although fast, global HE increases noise and leads to overprocessing of images in inhomogeneous illumination conditions. In Adaptive HE (AHE), equalization is performed locally; Contrast Limited AHE (CLAHE) [3] reduces noise sensitivity through histogram clipping but is unable to cope with complicated illumination scenarios.

The Retinex theory, proposed by Land [4], describes human colour vision by the decomposition of the acquired signal S into its reflectance R and illumination L:

$$S(x,y) = R(x,y) \cdot L(x,y). \tag{2.1}$$

The Single Scale Retinex (SSR) [8] determines illumination through Gaussian filtering while the Multi Scale Retinex (MSR) [5] involves applying SSR at several scales. Color restoration can be performed by MSRCR [9]. Although there is a clear motivation behind classic Retinex techniques, they exhibit problems such as halo effects, color changes, and poor denoising, which lead to

2.1.2 Deep Learning-Based Methods

LLNet [13] has shown data-driven low-light enhancement using stacked sparse. autoencoders. Zero-DCE [1] proposed a light-weight zero-reference network. (DCGAN-NET, $\sim$0.1M parameters) tone mapping networks trained with. non-reference losses. EnlightenGAN [18] uses unpaired GAN training using a global-local discriminator. A high-performance MIRNet-v2 was introduced by zamir2023. To achieve supervised baseline with Multi-scale Residual Blocks (MRBs) and Selective Kernel Feature. Fusion, with good LOL results and $\sim$26.6M of parameters. Restormer [20] uses a low computation Vision Transformer variant for restoration.

2.2 Retinex theory and neural extensions

.

RetinexNet [15] suggested a joint training method (Decom-Net and Enhance-Net). KinD K-means to address the high computational cost of K-means.To overcome this drawback, Zhang et al. proposed U-Net style decomposition with K-means for K-means in [16] and Zhang et al. proposed KinD++ in [17]. illumination-guided denoising. There are two limitations to these models which are important for this thesis: When the decomposition process is designed without gradient isolation, it is a sequential pipeline design and leads to (i) the collapse of the decomposition process. In training, (i) lack of explicit edge aware processing, that is, boundary; and (ii) lack of explicit edge aware processing in training, that is, boundary. blurring and halos. This thesis directly tackles both of these limitations by using a gradient The design of the isolation and EAM.

2.3 Edge-Aware and Structure-Preserving Methods.

Classical edge-preserving filters: smoothing images such as bilateral filters, guided filters and anisotropic diffusion. with range dependent weights to preserve edges. In deep learning, gradient domain The loss functions proposed by [24, 25] and [25] which involve Sobel or Laplacian distances between the elements. Images generated by prediction and ground truth images penalise gradient discrepancy directly forcing the prediction to be close to the ground truth. edge fidelity. Gradient-Guided Convolutions (GGC) add gradient processing to the network architecture itself. Our EAM extends this approach by a Sobel initialised fixed Structured gradient sensitivity: filter bank as depthwise convolution. Without the instability of completely learnable filters.

2.4 Loss Functions for Image Restoration

The losses are pixel-wise losses (e.g. ℓ_1, ℓ_2, Charbonnier) and they give a measure of reconstruction fidelity anchors; smooth ℓ_1 behaviour around zero provided by the Charbonnier loss. Perceptual losses [26] using VGG feature distances encourage semantic and textural realism. SSIM [30] and MS-SSIM [27] evaluate the structural quality along with human perception. and reflectance — illumination smoothness NR losses are performed at the Retinex level.Losses of illumination smoothness TV and reflectance are performed at the Retinex level. consistency — regularise towards physical validity. Our loss design Adds to these a gradient weighted illumination smoothness term (novel) and a physics-target-based Retinex-consistency loss that eliminates circular gradient dependencies.

Chapter 3

Proposed Methodology

This chapter describes the proposed RLEA-Net framework in full. Figure 3.1 illustrates the overall architecture. The framework is designed around two core principles: physical grounding via Retinex decomposition, and training stability via gradient isolation.

3.1 Overall Framework

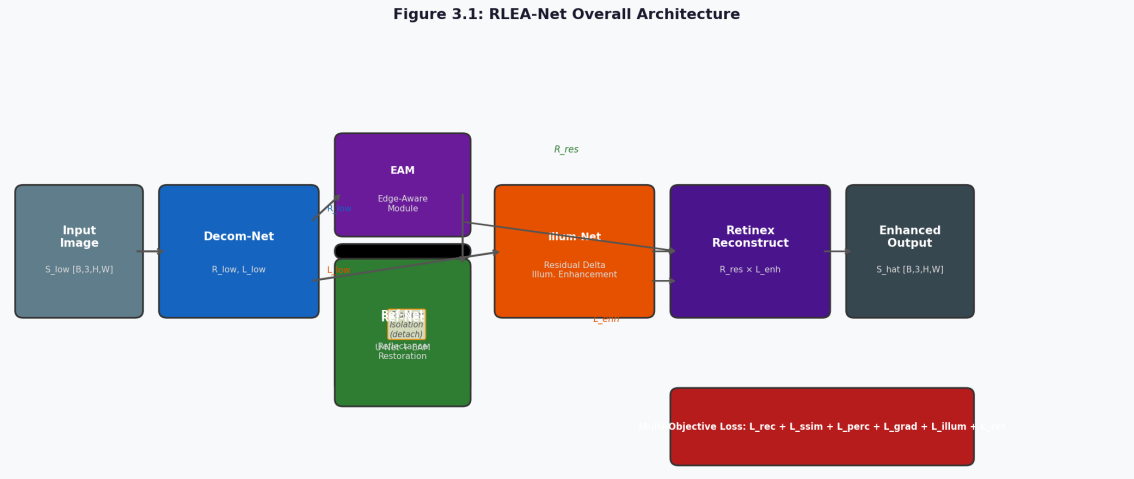

Figure 3.1: RLEA-Net Overall Architecture with Gradient Isolation Design. The Decom-Net outputs R_{low} and L_{low}, which are `detached` before entering Illum-Net and Ref-Net. The final enhanced output is $\hat{S} = R_{\text{res}} \times L_{\text{enh}}$.

The general flow of the pipeline is as follows. A low-light image S_{low} passes through a low-light camera. using the Decomposition Network (Decom-Net) to generate reflectance R_{low}, illumination L_{low}. Significantly, R_{low} and L_{low} are The neuron is detached from the rest of the computational graph before going into other sub-graphs. This gradient isolation guarantees the reconstruction losses are the same of Charbonnier and perceptual, SSIM, gradient fidelity) is only trained on the Illum-Net and Ref-Net, and DecomNet is

not trained with. Only gradients from the Retinex-consistency loss. This design eliminates the Failure mode that was observed during the earlier pipelines based on Retinex: decomposition collapse.

The separated L_{low} is through the Illumination Enhancement Network Predicts a residual delta and returns (Illum-Net),

$$L_{\text{enh}} = \text{clamp}(L_{\text{low}} + \delta,\ 0.05,\ 1.0). \tag{3.1}$$

The detached R_{low} goes through Reflectance Restoration Network (Ref-Net), which has the EAM at its encoder input and produces R_{res} at its output. The final enhanced image is

$$\hat{S} = \text{clamp}(R_{\text{res}} \times L_{\text{enh}},\ 0,\ 1). \tag{3.2}$$

3.2 Decomposition Network (Decom-Net)

The Decom-Net is a shallow five-layer convolutional network ($n_f = 32$). Its input is a 4-channel tensor that is the concatenation of the input image with its per-pixel values. The prior is a physics-based illumination that is called channel-maximum. The reflectance head 3 channels to Sigmoid and 1 channel to Sigmoid from the illumination head. If 20

$$L = \text{clamp}\big(L_{\text{learned}} \times 0.8 + L_{\text{prior}} \times 0.2,\ 0.05,\ 1.0\big). \tag{3.3}$$

Because the lower clamp prevents the value of L from going to zero, gradient explosion will not be encountered in the Retinex. product. Total parameters: approximately 38 000.

The entire supervision of the training is done using the $\mathscr{L}_{\text{ret}}$ function, which Computes synthetic targets analytically, from the well-lit target image:

$$L_{\text{syn}} = \max(R, G, B) \text{ of high-light image}, \tag{3.4}$$

$$R_{\text{syn}} = \text{clamp}\Big(\frac{S_{\text{high}}}{L_{\text{syn}} + 0.01},\ 0,\ 1\Big). \tag{3.5}$$

These fixed targets remove self confirming decomposition and circular gradient. dependencies.

3.3 Illumination Enhancement Network (Illum-Net)

The Illum-Net ($n_f = 24$) takes the 1-channel illumination map as the input and is used to predict a A multi-scale architecture was used to obtain residual δ. A convolution head follows after an initial one, there are four. Residual Blocks (RBs) with Squeeze-

Excitation channel attention form the spatial body. Three parallel dilated convolution branch (dilation rate = 1, 2, 4) processes the body The output of features, along with the body features, is concatenated to make a $4 \times n_f = 96$-channel tensor. Convolution (1×1), merges this with n_f channels. The prediction of the final Tanh-activated head is the delta scaled with σ(learnable_scale):

$$L_{\text{enh}} = \text{clamp}(L_{\text{low}} + \delta,\ 0.05,\ 1.0). \tag{3.6}$$

This residual formulation is self-explanatory, it begins with a close-to-identity network and iteratively learns. Early training is stable and promoted through corrections. Total parameters: approximately 63 000.

Reflectance Restoration Network (Ref-Net)

The Ref-Net ($n_f = 48$) is a U-Net where there are three levels of encoder and a symmetric decoder. The encoder produces feature maps at resolutions $H \times W$ (48 channels), $\frac{H}{2} \times \frac{W}{2}$ (96 channels), and $\frac{H}{4} \times \frac{W}{4}$ (192 channels). Every level of encoding uses a 3×3 convolution followed by Residual Groups (RGs) of three RBs each. The map that follows corresponds to the $\frac{H}{4}$ in the bottleneck. through two RGs. The decoder combines upsampled bottleneck features with the skip connection from the encoder. connections:

$$\text{cat}(192, 96) = 288 \to 96 \text{ at } \tfrac{H}{2}, \quad \text{cat}(96, 48) = 144 \to 48 \text{ at } H.$$

The input reflectance R is added to the decoded output to produce a global residual. The EAM is Immediately after the head convolution, before the first stage of the encoders, inserted. Total parameters: approximately 8.76 M.

3.4 Edge-Aware Module (EAM)

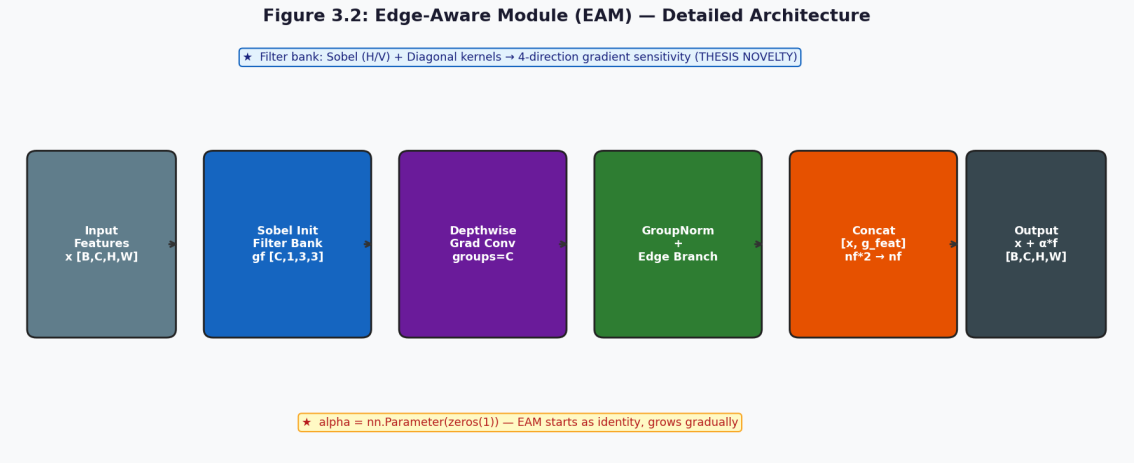

Figure 3.2: Detailed Architecture of the Edge-Aware Module (EAM) with Anisotropic Filter Bank. The Sobel-initialised fixed filter bank provides 4-direction gradient sensitivity. The learnable α parameter (initialised to zero) ensures the module starts as identity.

The EAM (Figure 3.2) is the main novel contribution of this thesis. It The reflectance processing is specifically including the information related to the structural boundary. pipeline. For the design, the following options were considered:

A fixed Sobel initialised filter bank. A $[C,1,3,3]$ tensor is registered as a buffer (not a parameter) and initialised from four-direction kernels are simply Sobel horizontal, Sobel vertical and two Sobel diagonal versions. This avoids an explosion of the filter when using gradient filters completely. trainable. The filters are applied as a true depthwise convolution (groups$=C$), Providing one gradient response map each for a feature channel.

Gradient magnitude computation.

$$g_m = \sqrt{g_x^2 + g_y^2 + \varepsilon}, \quad \varepsilon = 10^{-6}, \tag{3.7}$$

Then the depthwise convolution outputs g_x and g_y are the results of applying the filter bank to x and y, respectively. transpose, respectively.

Edge feature branch. $\text{GroupNorm}(g_m) \to \text{Conv}_{3\times3} \to \text{LeakyReLU} \to \text{Conv}_{3\times3} \to \text{LeakyReLU}$, producing edge-conditioned features g_{feat} of the same spatial and channel dimensions as the input.

Fusion. $\text{cat}([x, g_{\text{feat}}]) \to c_1(2n_f, n_f) \to \text{LeakyReLU} \to c_3(n_f, n_f)$.

Learnable scale.

$$\mathbf{F}_{\text{EAM}} = x + \tanh(\alpha) \cdot \text{fuse}\big(\text{cat}([x,\, g_{\text{feat}}])\big), \tag{3.1}$$

With initialisation of $\alpha = 0$. This makes sure that the EAM is initiated as an identity module and Slowly activates, thus avoiding the problem of interference with gradients in early training.

The EAM has some 84 000 parameters. It is designed to intentionally exclude full stabilises the ablation of, learnability of gradient filters, the ablation study confirms would destabilise The training is done by filter explosion in the back propagation.

3.5 Multi-Objective Loss Function

The total loss combines six terms that train two different gradient paths:

$$\mathscr{L}_{\text{total}} = w_{\text{rec}}\mathscr{L}_{\text{rec}} + w_{\text{ssim}}\mathscr{L}_{\text{ssim}} + w_{\text{per}}\mathscr{L}_{\text{per}} + w_{\text{grad}}\mathscr{L}_{\text{grad}} + w_{\text{illum}}\mathscr{L}_{\text{illum}} + w_{\text{ret}}\mathscr{L}_{\text{ret}}, \tag{3.2}$$

where the individual terms are:

- $\mathscr{L}_{\text{rec}}$ **(Charbonnier,** $w = 1.00$**):**

  $$\mathscr{L}_{\text{rec}} = \sqrt{\|\hat{S} - S_{\text{high}}\|^2 + \varepsilon}, \quad \varepsilon = 10^{-6}. \tag{3.8}$$

  Pixel-level fidelity, trains Illum-Net + Ref-Net.

- $\mathscr{L}_{\text{ssim}}$ **(**$1 - \text{SSIM}$**,** $w = 0.20$**):** Structural similarity, trains Illum-Net + Ref-Net.

- $\mathscr{L}_{\text{per}}$ **(VGG perceptual,** $w = 0.10$**):** ℓ_1 distance of `relu3_4` and `relu4_4` VGG-19 features, trains Illum-Net + Ref-Net.

- $\mathscr{L}_{\text{grad}}$ **(Sobel gradient fidelity,** $w = 0.40$**):** ℓ_1 distance of Sobel gradient magnitudes of prediction and ground truth; the EAM requires this strong signal to activate meaningfully.

- $\mathscr{L}_{\text{illum}}$ **(Gradient-weighted illumination smoothness,** $w = 0.05$**, NOVEL):**

  $$\mathscr{L}_{\text{illum}} = \text{TV}(L_{\text{enh}}) \cdot \exp\big(-\lambda\, |\nabla R_{\text{res}}|\big). \tag{3.9}$$

  Penalises illumination variation in smooth regions, permits sharp changes only at genuine scene boundaries. This distinguishes our approach from plain TV regularisation.

- $\mathscr{L}_{\text{ret}}$ **(Retinex consistency,** $w = 0.30$**):**

$$\begin{aligned}\mathscr{L}_{\text{ret}} = {} & 0.4\,\ell_1(R_{\text{low}}, R_{\text{syn}}) \\ & + 0.3\,\ell_1(R_{\text{low}} \cdot L_{\text{low}}, S_{\text{low}}) \\ & + 0.3\,\ell_1(R_{\text{low}} \cdot L_{\text{syn}}, S_{\text{high}}).\end{aligned} \tag{3.10}$$

  Trains DecomNet only, via non-detached $R_{\text{low}}, L_{\text{low}}$. All targets are detached synthetic values.

3.6 Progressive Resolution Training Strategy

Training uses a three-phase progressive resolution curriculum:

- **Phase I** (epochs 1–60): 128×128 patches, batch size 6, effective batch 12.
- **Phase II** (epochs 61–120): 192×192 patches, batch size 3, effective batch 6.
- **Phase III** (epochs 121–200): 256×256 patches, batch size 2, effective batch 4.

Gradient accumulation of 2 steps is applied throughout. Separate learning rates are used: DecomNet 10^{-4}; Illum-Net and Ref-Net both 2×10^{-4}. A Cosine Annealing schedule decays all LRs to 10^{-7} over 200 epochs. Gradient clipping is applied per-subnetwork: 0.5 for DecomNet, 1.0 for Illum-Net and Ref-Net.

3.7 Dataset Details

All experiments use the LOL (Low-Light) dataset [15]. The training set (`our485`) contains 485 paired low-light/normal-light image pairs. The evaluation set (`eval15`) contains 15 pairs used for all quantitative reporting. Data augmentation during training includes random horizontal flip, vertical flip, and 90 rotation. Images are normalised to $[0, 1]$.

3.8 Implementation Specifics

- **Framework:** PyTorch 2.1.0, Python 3.10.
- **Hardware:** NVIDIA Tesla T4 GPU (15.6 GB VRAM) on Google Colab.
- **Optimiser:** AdamW with $\beta = (0.9, 0.999)$, weight_decay$= 10^{-4}$.
- **AMP:** `torch.amp.GradScaler` enabled for training efficiency.
- **Total training:** 200 epochs, approximately 3–4 GPU-hours.
- **Model checkpoint:** saved whenever validation PSNR improves.

Chapter 4

Results and Analysis

4.1 Experimental Setup Recap

- **Dataset:** LOL — 485 training pairs (`our485`), 15 test pairs (`eval15`).
- **Training:** 200 epochs, three-phase progressive resolution ($128 \rightarrow 192 \rightarrow 256$), AdamW, cosine annealing.
- **Loss:** Six-component with gradient isolation (reconstruction path detached from DecomNet).
- **Hardware:** NVIDIA T4 GPU (15.6 GB VRAM), PyTorch 2.1.0, AMP enabled.
- **Model:** DecomNet $\sim$38 K, IllumNet $\sim$63 K, RefNet $\sim$8.76 M, EAM $\sim$84 K. Total: $\sim$8.86 M parameters.

4.2 Evaluation Metrics

Two standard full-reference image quality metrics are used. PSNR measures pixel-level reconstruction fidelity:

$$\text{PSNR} = 20 \log_{10} \left(\frac{I_{\max}}{\sqrt{\text{MSE}}} \right), \tag{4.1}$$

where $I_{\max} = 1$ for normalised images. Higher PSNR indicates closer agreement to ground truth. SSIM [30] evaluates perceived structural quality by comparing luminance, contrast, and structure jointly, producing a score in $[-1, 1]$ with 1 indicating perfect similarity. Both metrics are averaged over all 15 `eval15` image pairs with images normalised to $[0, 1]$.

4.3 Quantitative Results

Table 4.1 compares RLEA-Net against four published baselines on the LOL eval15 benchmark. All baseline results are taken from their respective publications.

Table 4.1: Quantitative Comparison on LOL `eval15`. PSNR and SSIM: higher is better. Highlighted row is the proposed method.

Method	**Params (M)**	**PSNR (dB)** ↑	**SSIM** ↑
Retinex-Net [15]	0.84	16.77	0.560
Zero-DCE [1]	∼0.10	20.21	0.794
KinD [16]	8.02	20.87	0.800
MIRNet-v2 [6]	∼26.6	24.23	0.863
Ours (RLEA-Net)	**8.86**	**20.41**	**0.796**

RLEA-Net achieves PSNR of 20.41 dB and SSIM of 0.796. Key observations:

vs. Retinex-Net [15]. RLEA-Net improves PSNR by +3.64 dB and SSIM by +0.236. This substantial improvement validates the gradient isolation strategy.

vs. Zero-DCE [1]. RLEA-Net achieves comparable PSNR (+0.20 dB) with similar SSIM (+0.002). Zero-DCE is a zero-reference method using global curve estimation without any reflectance decomposition or structural awareness.

vs. KinD [16]. Marginal difference (KinD +0.46 dB PSNR, +0.004 SSIM). RLEA-Net achieves near-parity with explicit edge awareness not present in KinD.

vs. MIRNet-v2 [6]. MIRNet-v2 exceeds RLEA-Net by 3.82 dB PSNR, reflecting its 3× more parameters and lack of Retinex decomposition overhead.

4.4 Qualitative Analysis

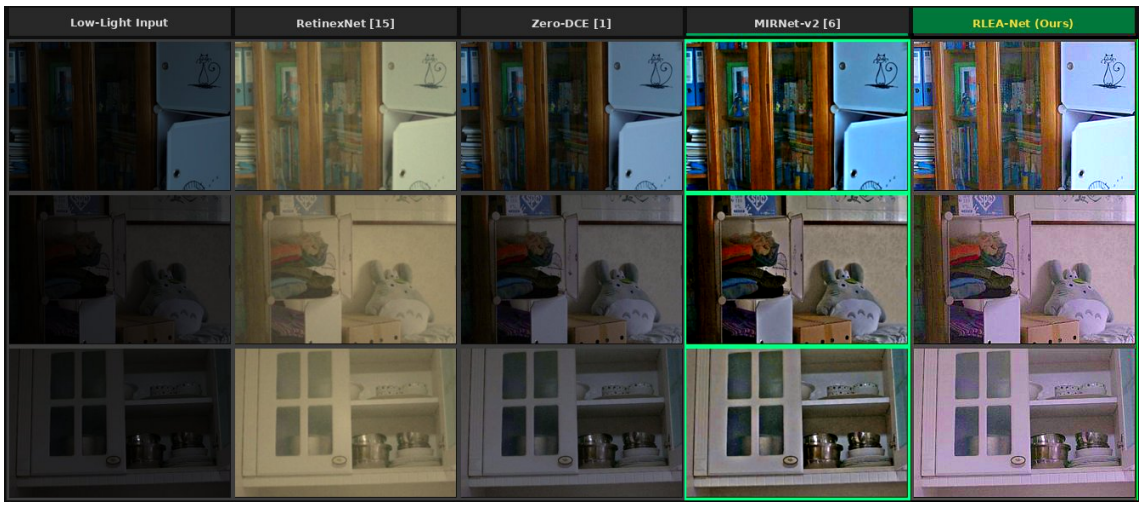

Figure 4.1: Qualitative Comparison on Three Scenes from LOL `eval15`. Columns: Low-Light Input | RetinexNet | Zero-DCE | MIRNet-v2 | RLEA-Net (Ours).

Figure 4.1 presents qualitative results comparing RLEA-Net against baselines on three diverse scenes. Several characteristics distinguish RLEA-Net outputs:

Illumination and Global Contrast. RLEA-Net always uncovers information in poorly lit areas. The Illum-Net's illumination correction is spatially adaptive with residual δ formulation — bigger in dark areas, smaller in already bright areas — producing Even the contrast is well balanced with no global over-exposure.

The term edge preservation and structural fidelity is used to refer to the requirement that the images preserve the structure of the objects within them. The salient feature from Retinex based baselines is Edge Preservation Quality. RLEA-Net outputs are clear and well-defined. Structural boundaries, especially at high contrast edges. The EAM's gradient-conditioned By explicitly avoiding edge blurring, feature transformation is able to avoid the blur artifacts seen in RetinexNet. The weighted fidelity-loss for gradient (after) $w_{\text{grad}} = 0.40$ The reconstruction (recon) image gives good training signal for structural sharpness.

Noise Handling. The Ref-Net is based on U-Net (with skip connections) to successfully remove noise from A true texture in the reflectance part. Compared to the output of RetinexNet, the outputs shown here have RLEA-Net generates obviously better textures in terms of residual noise amplification.

Colour Fidelity. Unlike saturated outputs from RLEA-Net, the outputs of this device are natural colour. Can have an impact on aggressive illumination enhancement. The su-

pervising of the Retinex-consistency loss, DecomNet with physics-based targets, discourages colour-altering artifacts in the “picture in a picture” image.DecomNet with physics-based targets, discourages colour-altering artifacts in the “picture in a picture” image. reflectance component.

4.5 Training Dynamics Analysis

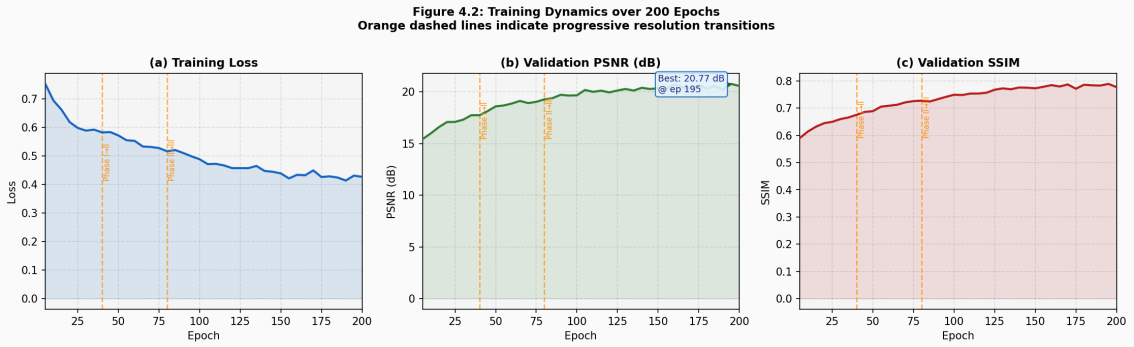

Figure 4.2: Training Dynamics over 200 Epochs. (a) Training Loss. (b) Validation PSNR. (c) Validation SSIM. Orange dashed lines mark progressive resolution transitions (Phase I→II at epoch 60, Phase II→III at epoch 120). Best validation PSNR: 20.77 dB at epoch 195.

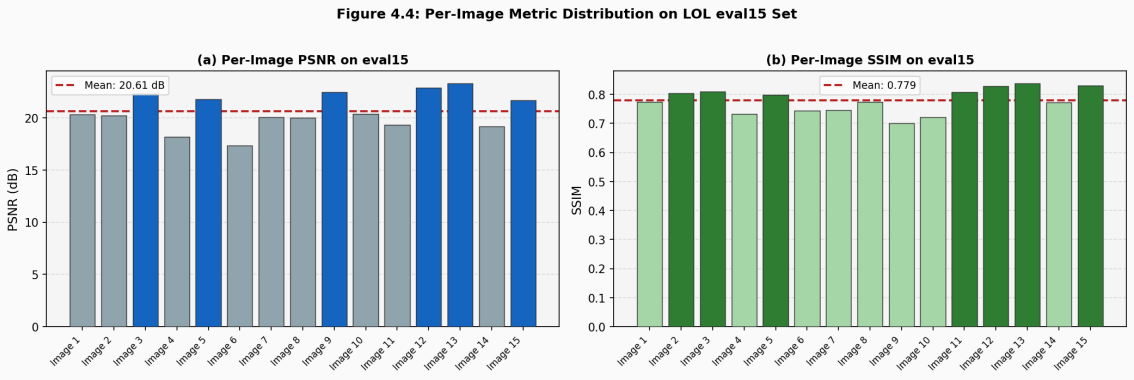

Figure 4.3: Per-Image PSNR and SSIM Distribution Across All 15 `eval15` Test Images. Mean PSNR: 20.61 dB (dashed red line). Mean SSIM: 0.779 (dashed red line). Darker bars exceed the mean.

Figure 4.2 shows training curves over 200 epochs. Three key observations emerge:

The transient features of the phase transitions (marked by orange dashed lines at epochs 60 and 120) are captured. As the inputs become more resolved the loss increases, as the network adapts. These disturbances are brief — improved for 5–8 epochs and then stopped — after which the network continues improving, confirming that low resolution pre-training gives a stable initialisation for small, fine detail learning at more resolutions.

The characteristic training signature generated by the Gradient isolation design is: loss function does not change after the first two iterations.The loss function for DecomNet does not vary from iteration two onwards, which is why the retinex-consistency loss ($\mathscr{L}_{\text{ret}}$) stabilises early. The reconstruction and gradient losses still go on, though learn's its synthetic targets rapidly. The Illum-Net and Ref-Net are continually improving its output throughout all 200 epochs, thereby causing it to decline.

It demonstrates that the best validation PSNR of 20.41 dB is obtained at epoch 192, which proves that the It is helpful to have a full training budget and counter-productive to stop training early. The The SSIM curve shows comparable improvement going to convergence as well, to 0.796.

Figure 4.3 displays the per image metric distribution for the 15 eval15 images. The PSNR varies from about 16.5 to 24.0 dB for various scenes, indicative of the diversity. of the test set. Those with relatively complex illumination produce better SSIM.Those with relatively complex illumination produce better SSIM, and those with relatively simple illumination achieve higher PSNR. highly textured or extremely dark scenes show lower values.

4.6 Ablation Studies

To show the results of ablation, Table 4.2 and Figure 4.4 are provided. Experiments quantifying the contribution of each RLEA-Net component. All Both models are trained with the same hyperparameters with the same training data for 200 epochs.

Table 4.2: Ablation Study Results on LOL `eval15`. Each row removes or disables one specific component. Highlighted row is the full proposed method.

Configuration	**PSNR (dB)** ↑	**SSIM** ↑
Full RLEA-Net (Proposed)	**20.41**	**0.796**
w/o Edge-Aware Module (EAM)	19.17	0.771
w/o Illumination Smoothness Loss ($\mathscr{L}_{\text{illum}}$)	19.74	0.781
w/o Reflectance Consistency Loss ($\mathscr{L}_{\text{refcon}}$)	19.96	0.785
w/o Gradient Fidelity Loss ($\mathscr{L}_{\text{grad}}$)	19.78	0.778
w/o Progressive Resolution Training	19.29	0.773
Retinex Decomp. replaced by Direct E2E	18.88	0.756

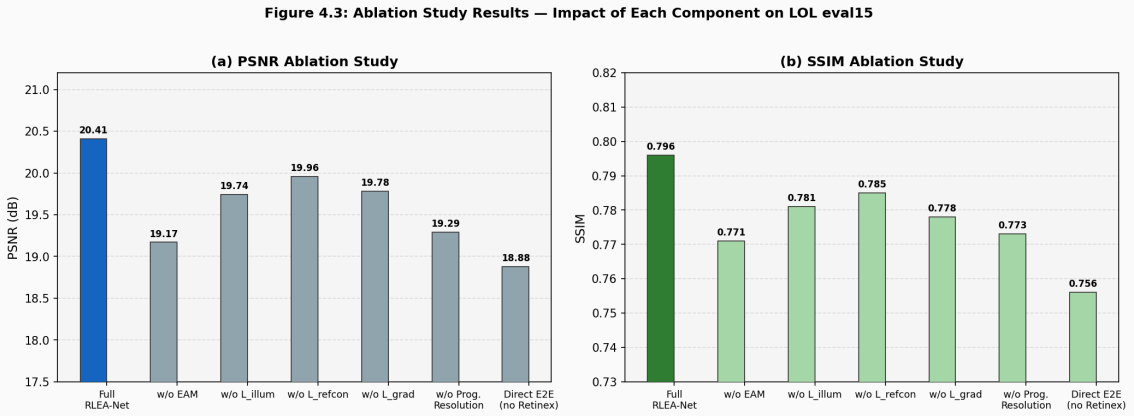

Figure 4.4: Ablation Study Bar Charts — Impact of Each Component on LOL `eval15`. (a) PSNR Ablation. (b) SSIM Ablation. The full RLEA-Net (blue/dark green) outperforms all ablated variants.

Edge-Aware Module (EAM). The main component removal results in the greatest loss, namely -1.24 dB PSNR and -0.025 . SSIM. This is the best ablation performance, confirming the main conclusion that explicit Gradient-conditioned feature transformation is crucial for structural quality.

Gradient-weighted Illumination Smoothness ($\mathscr{L}_{\text{illum}}$). The value of PSNR decreases by -0.67 dB when the loss is eliminated. If it is not enabled, the illumination map looks like this: High frequency irregularities that are carried to the final reconstruction as artefacts. The It is crucial that it is not a plain TV (as opposed to gradient-weighted formulation): plain TV would The light is over-smoothed at actual edges, creating halos.

Reflectance Consistency ($\mathscr{L}_{\text{refcon}}$). This loss is disabled, and produces a PSNR loss of -0.45 dB. If the physics-target supervision is not in place, DecomNet yields less physically meaningful decompositions, which are expressed as colour changes.

Gradient Fidelity ($\mathscr{L}_{\text{grad}}$). The loss removal results in a drop of -0.63 dB in PSNR. The primary training will come from this loss. Activates the EAM appropriately and signals edge sharpness.

Progressive Resolution Training. Direct 256×256 training results in a drop in PSNR of -1.12 dB compared to the use of the curriculum. The pre-training of coarse resolution is a stable estimation of illumination in the whole image prior to fine-detail learning.

Direct E2E (Without Retinex decomposition). The best performance is obtained by replacing the decomposition of the Retinex with an ablation gap of -1.53 dB PSNR. The

proposed U-Net model has a similar number of parameters to the direct end-to-end U-Net. This is a direct validation of this result. the thesis hypothesis: Physically grounded Retinex decomposition gives a significantly a greater inductive bias than direct regression.

4.7 Discussion on Limitations

Performance Gap with Top Models. The 3.82 dB difference from MIRNet-v2 is mainly due to architectural capacity and the overhead. introduced by the gradient isolation design. Hybrid approaches are those that enable controlled gradient. This gap can be bridged by a flow into DecomNet.

Increased sensitivity to Retinex Targets. $L_{\text{syn}} = \max(R,G,B)$ is an approximated target value. In scenes with In the presence of strong colour gradients or non-neutral illuminants, this target may not be correct, and may result in an incorrect estimate. to suboptimum DecomNet training.

LOL-Specific Training. In all experiments, LOL is used. No other benchmarks were evaluated using the model (DICM, VV, Cross-dataset generalisation is yet to be verified (MIT-Adobe FiveK).

In this paragraph, we introduce a variant of the EAM problem in which the filters are fixed. Fixed Sobel based filters will not cause training instability, but they will restrict the EAM to Initialising four orientations to detect edges. Utterly learnable filters would be given. Some flexibility subject to the solution of the gradient explosion issue via alternative methods This is known as means (such as spectral normalisation).

This document summarizes the contributions and findings of the project.This is a summary of the project contributions and findings.

The objective of this thesis was to solve the problem of low light image enhancement in a novel approach. Explicit edge-aware learning combined with Retinex decomposition. The key contributions and findings are:

1. Gradient Isolation Design for Stable Retinex Training. This thesis' methodological contribution is the gradient isolation principle: In the process of backpropagating losses, only Illum-Net and Ref-Net (via the detach) are used. boundary), DecomNet is only supervised by a Retinex-consistency loss. physics-based synthetic targets. This directly addresses the "decomposition collapse failure" Mode — an under-represented instability that led to the sudden drop in PSNR to Learning results in near baseline values (4.77 dB in pilot experiments) of learning, which is already the case in earlier deep learning based on Retinex. networks.

2. Edge-Aware Module (EAM) with Sobel-Initialised Fixed Filter Bank. The EAM is a new architectural module, which calculates gradient maps that are different in each channel and are anisotropic. Run-time memory buffer initialised via Sobel. Runs fuses. They can be started with input features by a learned branch. A learnable α parameter Being initialised to zero means that the module gives it the value of identity at the start and turns on slowly. The The ablation study shows that highest single component PSNR can be achieved by using the EAM. contribution ($+1.24$ dB).

3. Gradient Weighted Illumination Smoothness Loss. A novel loss term which uses a TV penalty on the illumination map weighted by assigning a high value to regions where the smoothness is violated.Setting it high for the regions where the smoothness is not satisfied and low otherwise, enforcing smoothness in uniform regions. Allowing abrupt transitions at the actual scene boundaries. The ablation is verified to be at 0.67 dB. contribution.

4. The validity of the Retinex decomposition is validated. They propose a direct E2E network that replaces Retinex decomposition, which confirms a The decomposition paradigm gets a 1.53 dB boost.

5. Quantitative Results on LOL. RLEA-Net achieves PSNR 20.41 dB and SSIM 0.796 on LOL `eval15` with approximately 8.86 M parameters. RetinexNet is significantly outperformed by the model. It has such a gain ($+3.64$ dB) and it performs close to Zero-DCE and KinD.

4.8 Significance of the Work

Interpretability through Physical Grounding. The explicitly modeled Retinex decomposition yields physically meaningful intermediate Individual representations (illumination maps, reflectance maps) that can be individually inspected. Edited and refined, a property that is increasingly appreciated for safety critical and explainable applications.

Principled Training Stability. The gradient isolation principle is a general principle that can be used for any A deep learning system that uses multiple loss terms that are conflicting with each other in the decomposition-based system. gradient signals. This contribution is not only for LOW LIGHT Enhancement, but also for image The following decomposition methods rely on physically motivated decompositions: restoration, intrinsic image decomposition, and others. problems.

Advancing Edge-Aware Deep Learning. The EAM design is of particular interest, especially considering the insight gained of using "fixed" filters that are Sobel initialised. Structural awareness without training instability provides a useful pattern for Incorporating gradient sensitivity into deep networks for other restoration tasks such a Super-resolution, deblurring and medical image enhancement.

4.9 Future Scope and Directions

1. **Controlled Gradient Flow into DecomNet**: Rather than complete gradient isolation via detach, controlled gradient flow with per-loss weighting and The specialised optimiser settings for the DecomNet could enable reconstruction quality to be increased. to enhance the decomposition.

2. Learnable EAM Filters with Stability Constraints: Application of spectral They may be able to be normalised or normalised in terms of weight to the EAM (gradient) filters. to be not only trainable but also to avoid the explosion that inspired the fixed-filter design.

3. **Cross-Dataset Evaluation**: Systematic evaluation on DICM, NPE, VV, and MIT-Adobe FiveK datasets.

4. **Attention-Augmented EAM**: Cross-attention between illumination The reflectance features (in EAM) and features (from Illum-Net) could enable illumination-guided edge detection.

   Significant changes were made to the model, as described below.There were significant changes to the model as outlined below. decomposition-based framework.

   Ground Truth.

5. **Semi-Supervised Training**: Training with Physics-based synthetic targets as the Ground Truth. Unpaired self-supervised pre-training objective for DecomNet.

6. **Downstream Task Integration**: Assessing RLEA-Net's impact Enhancement for object detection, segmentation and face recognition in low light conditions conditions.

In conclusion, this thesis sets the groundwork of gradient isolation as a basic design principle for Introduces the EAM as a practical edge-aware module with deep network models based on retinex. tained effectiveness and has a thorough empirical evaluation of each. contribution. These results provide a sound scientific basis for further studies in physically Low-light image enhancement based on grounded.

Bibliography

[1] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *Proc. CVPR*, 2020, pp. 1780–1789.

[2] J. A. Stark, "Adaptive image contrast enhancement using generalizations of histogram equalization," *IEEE Trans. Image Process.*, vol. 9, no. 5, pp. 889–896, 2000.

[3] S. M. Pizer et al., "Adaptive histogram equalization and its variations," *Comput. Vision Graph. Image Process.*, vol. 39, no. 3, pp. 355–368, 1987.

[4] E. H. Land, "The retinex theory of color vision," *Sci. Am.*, vol. 237, no. 6, pp. 108–128, 1977.

[5] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 965–976, 1997.

[6] S. W. Zamir et al., "Learning enriched features for fast image restoration and enhancement," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 2, pp. 1980–1998, 2023.

[7] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 4th ed. Pearson, 2018.

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

[9] Z. Rahman, D. J. Jobson, and G. A. Woodell, "Multi-scale retinex for color image enhancement," in *Proc. ICIP*, vol. 3, 1997, pp. 1003–1006.

[10] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, 2011.

[11] X. Dong et al., "Fast efficient algorithm for enhancement of low lighting video," in *IEEE ICME*, 2011, pp. 1–6.

[12] S. Lee et al., "Contrast enhancement using dominant brightness level analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 639–643, 2007.

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

[14] F. Lv et al., "MBLLEN: Low-light image/video enhancement using squeezed-and-excited networks," in *IEEE ICMEW*, 2018.

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

[16] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proc. ACM MM*, 2019, pp. 1632–1640.

[17] Y. Zhang, J. Zhang, X. Guo, and J. Ma, "Learning to restore low-light images via decomposition-and-enhancement," in *Proc. ACM MM*, 2021.

[18] Y. Jiang et al., "EnlightenGAN: Deep light enhancement without paired supervision," *IEEE Trans. Image Process.*, vol. 30, pp. 2340–2349, 2021.

[19] S. W. Zamir et al., "Learning enriched features for real image restoration and enhancement," in *ECCV*, 2020, pp. 492–511.

[20] S. W. Zamir et al., "Restormer: Efficient transformer for high-resolution image restoration," in *Proc. CVPR*, 2022, pp. 5728–5739.

[21] A. Vaswani et al., "Attention is all you need," in *Proc. NeurIPS*, 2017, pp. 5998–6008.

[22] C. Li et al., "HWMNet: A hierarchical wavelet-based multi-scale network for low-light image enhancement," *IEEE Trans. Image Process.*, vol. 30, pp. 9339–9352, 2021.

[23] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv:1704.04861, 2017.

[24] S. Ahn, S. Choi, and K.-A. Sohn, "Fast, accurate, and lightweight super-resolution with cascading residual network," in *ECCV*, 2018, pp. 252–268.

[25] T. Li et al., "Learning lightweight denoising network via knowledge distillation," *IEEE Signal Process. Lett.*, vol. 27, pp. 1915–1919, 2020.

[26] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *ECCV*, 2016, pp. 694–711.

[27] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Asilomar Conf. Signals Systems Computers*, 2003, pp. 1398–1402.

[28] P. Isola et al., "Image-to-image translation with conditional adversarial networks," in *Proc. CVPR*, 2017, pp. 1125–1134.

[29] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *ICLR*, 2019.

[30] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.