UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

GABRIEL TADIELLO MORAES

Controlling Low-Light Image Enhancement with Restoration Level Estimator

Thesis presented in partial fulfillment of the requirements for the degree of Master of Computer Science

Advisor: Prof. Dr. Manuel Menezes de Oliveira Neto

Porto Alegre
May 2025

CIP — CATALOGING-IN-PUBLICATION

Moraes, Gabriel Tadiello

Controlling Low-Light Image Enhancement with Restoration Level Estimator / Gabriel Tadiello Moraes. – Porto Alegre: PPGC da UFRGS, 2025.

48 f.: il.

Thesis (Master) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR–RS, 2025. Advisor: Manuel Menezes de Oliveira Neto.

1. Low-Light Image Enhancement. 2. Controllable Image Enhancement. 3. Image Restoration. 4. Computational Photography. 5. Computer Vision. 6. Deep Learning. I. Oliveira Neto, Manuel Menezes de. II. Título.

ABSTRACT

Low-light image enhancement (LLIE) is an important task in computer vision, addressing the need to improve the visual quality of images captured in suboptimal lighting conditions. Enhanced images are not only more visually appealing but also more effective for downstream tasks such as object detection and classification. However, while the desired level of enhancement is often subjective and varies across users and applications, previous LLIE techniques do not allow users to control the desired level of enhancement intensity. We introduce the Restoration Level Estimator (RLE) block, a novel component designed to provide control over the enhancement level in existing LLIE models. The RLE block can be seamlessly integrated into convolutional neural networks, adding a new channel that allows users to adjust the level of enhancement applied to input images. Furthermore, our experiments show that, in addition to offering control, the RLE block can improve the overall performance of LLIE models as measured by PSNR, SSIM, and LPIPS. We demonstrate the flexibility of our approach across multiple LLIE models, highlighting its potential to improve both user experience and model performance.

Keywords: Low-Light Image Enhancement. Controllable Image Enhancement. Image Restoration. Computational Photography. Computer Vision. Deep Learning.

Controlando a Realce de Imagens em Baixa Iluminação com Estimador de Nível de Restauração

RESUMO

O realce de imagens em baixa iluminação (Low-Light Image Enhancement - LLIE) é uma tarefa importante em visão computacional, visando melhorar a qualidade visual de imagens capturadas em condições de iluminação subótimas. Imagens realçadas não apenas se tornam mais agradáveis visualmente, mas também mais eficazes para tarefas subsequentes, como detecção e classificação de objetos. No entanto, o nível desejado de realce é frequentemente subjetivo e varia entre usuários e aplicações, o que motiva a necessidade de métodos de LLIE controláveis, que permitam ajustar a intensidade do realce.
Neste trabalho, introduzimos o bloco Estimador de Nível de Restauração (Restoration Level Estimator - RLE), um componente projetado para proporcionar controle sobre o nível de realce em modelos existentes de LLIE. O bloco RLE pode ser integrado em redes neurais convolucionais, adicionando um novo canal que permite aos usuários ajustar o nível de realce aplicado às imagens de entrada. Além disso, nossos experimentos mostram que, além de oferecer controle, o bloco RLE pode, em alguns casos, melhorar o desempenho geral dos modelos de LLIE. Demonstramos a flexibilidade da nossa abordagem em diversos modelos de LLIE, destacando seu potencial para aprimorar tanto a experiência do usuário quanto o desempenho do modelo.

Palavras-chave: Realce de Imagens em Baixa Iluminação. Realce de Imagens Controlável. Restauração de Imagens. Fotografia Computacional. Visão Computacional. Aprendizado Profundo.

LIST OF ABBREVIATIONS AND ACRONYMS

CNN	Convolutional Neural Network
LLIE	Low-Light Image Enhancement
LPIPS	Learned Perceptual Image Patch Similarity
MEF	Multi-Exposure Image Fusion
PSNR	Peak Signal-to-Noise Ratio
RLE	Restoration Level Estimator
SAM	Segment Anything Model
SKF	Semantic-aware Knowledge-guided Framework
SSIM	Structural Similarity Index Measure

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 INTRODUCTION

Low-light image enhancement is a challenging task influenced by various factors such as environmental lighting conditions and camera settings. The goal of enhancing low-light images is not only to improve their visual appeal but also to make them more suitable for downstream tasks, such as object detection and segmentation (Diamond et al., 2021). Traditional methods often aim for a universal solution independent of the lighting conditions of the input images, which may not address individual user preferences or specific requirements of different regions within an image. Therefore, providing user control over the enhancement process is highly desirable.

Recent advancements in deep learning have led to significant improvements in low-light image enhancement techniques. However, most previous approaches enhance images uniformly, which can be problematic in scenarios where different regions of the image require different levels of enhancement. Additionally, the subjective nature of what constitutes a well-lit image varies among users, making it essential to offer a controllable enhancement mechanism.

We present a method that addresses those challenges by introducing a Restoration Level Estimator (RLE) block that can be added to existing CNN-based low-light image enhancement (LLIE) networks. The RLE block allows users to adaptively control the amount of enhancement applied to an image (*i.e.*, it provides both global and local control).

Our method enables users to adjust the enhancement level according to their preferences. This control is achieved by manipulating the output of the RLE block, which corresponds to the optimal level of enhancement predicted for the input image. Additionally, our method supports local control, allowing users to specify multiple regions within an image and control their enhancement levels individually.

We evaluate the proposed method both quantitatively and qualitatively. The impact of adding the proposed RLE block to existing networks is assessed using standard metrics such as PSNR and SSIM. We demonstrate our method's ability to provide control over the image's enhancement level, both globally and locally.

Figure 1.1 provides an overview of our method, illustrating how the RLE module can be incorporated into existing CNN-based LLIE models to introduce user control over the enhancement process. The figure displays the results of applying different enhancement levels to an input image with a high dynamic range. It also shows how we use the

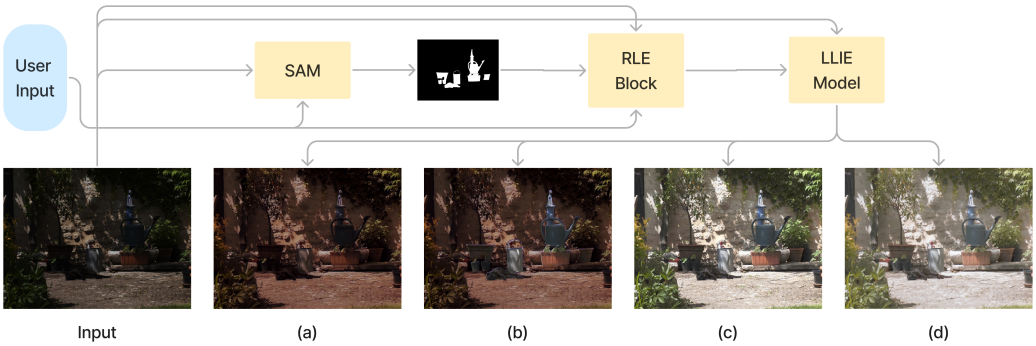


Figure 1.1 – Overview of our method and demonstration of its low-light image enhancement control capabilities. The diagram illustrates the processing pipeline: first, the user can optionally specify regions of interest using the Segment Anything Model (Kirillov et al., 2023). The RLE block then estimates an optimal enhancement level to be added as an additional input to the LLIE model (LLFlow + SKF in this example), which can be further adjusted by increasing or decreasing the estimated value. The bottom row showcases the results: the enhanced image obtained using the optimal level estimated by the RLE block is shown in (c); (a) image with lower enhancement obtained by decreasing the value estimated by the RLE block; (b) similar to (a), but with selected regions corresponding to the SAM mask brightened; (d) image with bigger enhancement obtained by increasing the value estimated by the RLE block, resulting in a brighter image.

Segment Anything Model (SAM) (Kirillov et al., 2023) to enable users to specify regions of interest for local adjustments.

The **contributions** of this work include:

- A Restoration Level Estimator block that can be added to existing low-light image enhancement methods, providing both global and local control over the enhancement process;
- A method for controlling the enhancement of multiple regions within an image, enabling users to adjust the enhancement levels individually for different areas;
- Experiments to evaluate the impact of adding the RLE block to existing methods and to showcase the control of low-light image enhancement.

This dissertation is structured as follows. Section 1.1 defines the problem of low-light image enhancement and introduces key concepts relevant to this work. Chapter 2 presents a review of related work, covering both low-light image enhancement methods and techniques for controllable image restoration. Chapter 3 describes our proposed method for introducing enhancement control to existing LLIE networks. Chapter 4 presents the experimental results and analysis of our method. Finally, Chapter 5 provides our conclusions and outlines possible directions for future work. Additionally, Ap-

pendix A describes the smartphone application we developed to capture burst sequences with varying exposure settings, as well as the dataset collected through this process.

1.1 Problem Definition

Low-light image enhancement (LLIE) aims to improve the perceptual quality and visibility of images captured under poor lighting conditions. Given an input image $I \in \mathbb{R}^{W \times H \times C}$, where W, H and C are respectively the width, height, and number of channels of image I. Typically, I is a color (RGB) image, with $C = 3$. The enhancement process can be formulated as:

a)Back lit b)Uneven light c)Dim light

d)Extremely low e)Colored light f)Boosted noise

Figure 1.2 – Examples of low-light image degradation observed in natural imaging scenarios, highlighting the diversity and complexity of challenges faced by LLIE methods.
Source: (Li et al., 2022)

$$\hat{R} = \mathcal{F}(I; \theta), \tag{1.1}$$

where $\mathcal{F}$ is a deep neural network with trainable parameters θ, and $\hat{R} \in \mathbb{R}^{W \times H \times C}$ is the enhanced output. The network is trained to minimize the difference between the enhanced result and the corresponding well-exposed ground truth image R, by solving:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, \mathcal{L}(\hat{R}, R), \tag{1.2}$$

where $R \in \mathbb{R}^{W \times H \times C}$ is the ground truth, and $\mathcal{L}(\hat{R}, R)$ is a loss function designed to guide the optimization.

While this formulation defines the problem from a learning perspective, low-light images encountered in real-world scenarios are affected by different types of degradation. Figure 1.2 illustrates examples of such images, which suffer from issues such as reduced contrast, noise amplification, and color distortion due to suboptimal lighting. In this work, we focus on extremely low-light images, where the available illumination is minimal, affecting the visibility of elements in the scene.

Furthermore, we refer to the enhancement level as the degree to which an LLIE model modifies an image to restore its visibility and perceptual quality. It reflects how strongly the enhancement process alters the input image to resemble a well-exposed version. Most learning-based LLIE methods apply a fixed enhancement level, which may not be ideal for all scenes or user preferences. In this work, we introduce a mechanism for estimating an optimal enhancement level of an image, while also allowing user adjustments.

2 RELATED WORK

This section discusses works related to low-light image enhancement (LLIE) and controllable image restoration.

2.1 Low-Light Image Enhancement

Low-light image enhancement has been an extensively studied subject, with traditional approaches often relying on histogram equalization (Abdullah-Al-Wadud et al., 2007) or the Retinex theory (Guo; Li; Ling, 2017). Histogram equalization improves contrast by redistributing pixel intensities. The Retinex theory assumes that an image can be decomposed into reflectance and illumination, and enhancement is achieved by adjusting the illumination component (Land; McCann, 1971).

These methods have been surpassed in recent years by learning-based methods, which achieve significantly improved results in terms of both detail preservation and noise reduction (Yu; Li; Yang, 2022)(Jiang et al., 2024)(Singh; Parihar, 2023)(Pan et al., 2024)(Zhang et al., 2024). As such, we restrict the discussion to learning-based techniques. For a discussion of traditional methods for LLIE, we refer the reader to the survey by Wang et al. (Wang et al., 2020).

LLNet (Lore; Akintayo; Sarkar, 2017) was the first technique to use a deep-learning-based approach for low-light image enhancement. It uses an autoencoder-based architecture and was trained on a synthetic dataset created by darkening and adding noise to well-exposed images. Expanding on the Retinex theory, Retinex-Net (Wei et al., 2018) proposed the first deep-learning approach inspired by Retinex decomposition. They also introduced the first paired dataset of real low-light and well-exposed images, enhancing the training process with more realistic data.

Several other Retinex-based methods have been proposed to refine this idea. KinD (Zhang; Zhang; Guo, 2019) employs three subnetworks for layer decomposition, reflectance restoration, and illumination adjustment. KinD++ (Zhang et al., 2021) improves KinD with a Multi-Scale Illumination Attention (MSIA) module to reduce artifacts and enhance illumination consistency. DeepUPE (Wang et al., 2019a) is another Retinex-based model that estimates an illumination map rather than directly enhancing images, applying illumination-aware constraints to improve exposure and color balance.

Beyond Retinex-based models, other supervised learning approaches have ex-

plored different enhancement strategies. SNR (Xu et al., 2022c) enhances images by dynamically processing different regions based on their signal-to-noise ratio, balancing noise reduction and detail preservation. LLFlow (Wang et al., 2022) introduces normalizing flows to model the distribution of low-light and enhanced images. WaveNet (Dang et al., 2023) enhances low-light images by modeling enhancement as a wave modulation process, using wave-based representations to improve both efficiency and detail preservation.

All previous methods are based on supervised learning and thus rely on paired datasets, which are difficult to collect for low-light image enhancement. To remove this dependency, unsupervised and zero-reference learning techniques have been explored. EnlightenGAN (Jiang et al., 2021) was the first unsupervised LLIE model, using a generative adversarial network (GAN) to enhance low-light images without requiring paired training data. Zero-DCE (Guo et al., 2020) proposed the first zero-reference approach for LLIE by introducing DCE-Net, a lightweight neural network that estimates pixel-wise enhancement curves for dynamic range adjustment. DRBN (Qiao et al., 2021) introduced a hybrid approach, combining supervised recursive band learning and unsupervised adversarial learning to balance perceptual quality and signal fidelity.

More recently, transformer-based methods have been introduced to capture long-range dependencies for LLIE. LYT-Net (Brateanu et al., 2024) is a lightweight transformer-based LLIE method that processes luminance and chrominance separately in the YUV color space. Retinexformer (Cai et al., 2023) is a one-stage transformer-based LLIE method that integrates illumination-guided self-attention to refine Retinex decomposition and improve enhancement quality.

Diffusion models have also been explored for LLIE. GlobalDiff (Hou et al., 2023) is a diffusion-based LLIE model that introduces global structure-aware and uncertainty-guided regularization to refine the diffusion process. PyDiff (Zhou; Yang; Yang, 2023) proposes a pyramid diffusion sampling strategy, which progressively increases resolution during enhancement, improving both processing speed and stability while mitigating global degradation effects.

Unlike standalone LLIE models, SKF (Wu et al., 2023) proposes a framework to enhance existing LLIE methods by improving their performance with semantic priors from a pre-trained segmentation network. It refines enhancement quality through a semantic-aware embedding (SE) module, which integrates scene semantics into feature representations. Additionally, semantic-guided color histogram (SCH) loss and semantic-

guided adversarial (SA) loss help maintain color consistency and texture realism.

While previous methods based on deep learning have made significant advancements in low-light image enhancement, they produce outputs with a fixed level of enhancement determined solely by the network during training. This fixed enhancement may not suit the varying preferences or needs of different users. In contrast, our approach automatically estimates an optimal enhancement level while also providing the flexibility of user control. This allows users to adapt the intensity of the enhancement as needed, offering a more flexible and customizable solution for low-light image enhancement. Our method can be used with CNN-based LLIE techniques, complementing them and strengthening their potential.

2.2 Controllable Image Restoration

Ni et al. (Ni et al., 2021) introduced a deraining method with bidirectional rain intensity control from a single input image, allowing users to control rain removal and rendering. Jiang et al. (Jiang; Zhang; Timofte, 2021) presented a model for JPEG artifact removal with adjustable control. It uses a decoupler module to separate the quality factor, which is then used by the reconstructor module to balance artifact removal and detail preservation. Yao et al. (Yao et al., 2023) presented a method for image denoising that allows user control over the intensity of denoising, providing the ability to preserve more details or to reduce noise further.

He et al. (He; Dong; Qiao, 2020) proposed a method based on residual learning for image restoration with multi-dimensional control to enable adjustment over various types and levels of image degradation. They propose a layer that modifies channel-wise features with depthwise convolution filters to allow control over restoration levels. Wang et al. (Wang et al., 2019b) introduced CFSNet for controllable image restoration. It features a dual-branch structure with coupling modules to adjust features for either low distortion or high perceptual quality, providing interactive control over the restoration process. Dynamic-Net (Shoshan; Mechrez; Zelnik-Manor, 2019) also allows controllable image restoration by first training the network with one objective and then adding tuning blocks trained with a different objective. Restoration level control is provided by a scalar parameter that adjusts the influence of these tuning blocks. Deep Network Interpolation (DNI) (Wang et al., 2019) enables smooth transitions in image restoration by applying linear interpolation in the parameter space of neural networks pre-trained for different effects.

All techniques mentioned so far were applied to various image restoration problems, such as denoising, JPEG deblocking, and deraining.

Specifically for low-light image enhancement, Xu et al. introduced ReCoRo (Xu et al., 2022a), a framework that allows users to enhance specific regions of low-light images using imprecise masks. ReCoRo uses SPADE blocks and a dual-discriminator setup to ensure realistic enhancements based on user input. It incorporates domain-specific data augmentations to handle mask imprecision, making it ideal for mobile applications where users provide rough enhancement guidance. Yin et al. (Yin et al., 2023) introduced CLE, a diffusion-based framework for low-light image enhancement with high user control. It allows users to specify brightness levels and enhance specific regions using the Segment Anything Model (SAM) (Kirillov et al., 2023). CLE Diffusion handles imprecise inputs and ensures consistent brightness with domain-specific augmentations and a novel encoding technique. It also minimizes color distortions and focuses on low SNR regions, providing robust tools for precise brightness and local adjustments.

Similarly to ReCoRo (Xu et al., 2022a) and CLE (Yin et al., 2023), our method supports user-guided control and local adjustments for low-light image enhancement. However, ours differs as it is a flexible module designed to integrate seamlessly into existing LLIE networks. This modularity enables our approach to complement a variety of networks while preserving their original strengths.

3 CONTROLLING LOW-LIGHT IMAGE ENHANCEMENT

Our method improves existing low-light image enhancement networks by introducing a Restoration Level Estimator (RLE) block. This section outlines the structure and training process of the RLE block and its integration with existing networks. Figure 1.1 shows an overview of our method.

3.1 Restoration Level Estimator

The architecture of the RLE block is shown in Figure 3.1. It consists of four blocks, each composed of two convolution layers (represented by yellow blocks) and one 2×2 max polling layer (represented by red blocks). The blocks are followed by a global average pooling layer, shown in green. All convolution layers use 3×3 kernels with no padding. The convolution layers in the same block have the same number of filters (channels). In the first four blocks, the number of filters is 24, 48, 96, and 192, respectively. The input to the RLE block is the original low-light RGB image. Its output is a scalar value, which is replicated into a tensor matching the input image's width and height and concatenated with the original image as an extra channel (represented in light green in Figure 3.1) to form the input for the main enhancement network (shown as the large gray block). A minor change in the input layer of the main network is required to accommodate the extra channel. Assuming the first layer is convolutional, this means increasing its number of input filters by one. As a result, the main network must be trained jointly with the RLE block, and its original pre-trained weights cannot be reused.

The RLE block's output drives the combined network to obtain the optimal enhancement level estimated by the model, typically resulting in a well-exposed image. Through a provided user interface, one can interactively modify the RLE block's output to adjust the image's enhancement level. Increasing the RLE block's predicted value produces brighter images while decreasing it results in darker images (Figure 1.1). This mechanism enables both automatic enhancement and user control over the final image brightness.

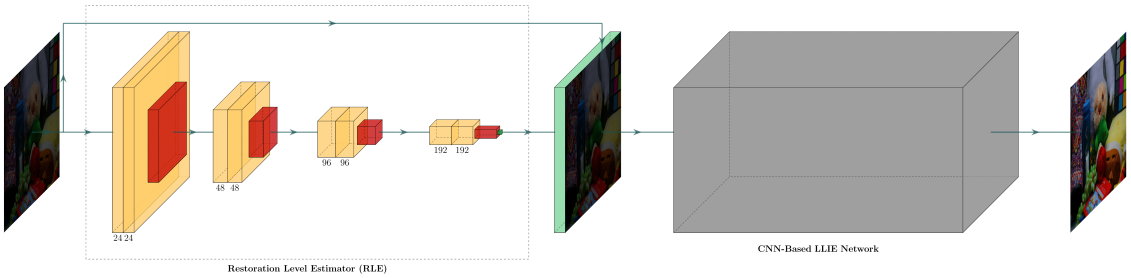


Figure 3.1 – Architecture of the Restoration Level Estimator (RLE) block, highlighted by the dashed rectangle. The RLE block takes a low-light image as input and processes it through four blocks, each consisting of two convolutional layers (yellow boxes) followed by a max pooling layer (red boxes), doubling the number of filters at each step. After the last block, there is a global average pooling layer, producing a scalar that is reshaped into a tensor with the same width and height as the input image. This tensor is concatenated with the original low-light image to form the input for an arbitrary CNN-based LLIE network (gray box).

3.2 Local Enhancement Control

Inspired by ReCoRo (Xu et al., 2022a) and CLE Diffusion (Yin et al., 2023), we introduce local control to our low-light image enhancement module. This feature allows users to specify multiple regions within an image and control their enhancement level individually. We utilize SAM (Kirillov et al., 2023) to generate masks for these regions.

By default, the output of the RLE block is replicated into a fourth channel and concatenated with the original input RGB image. Each SAM-generated mask allows one to interactively change (increase or decrease) the value of the RLE block's prediction. These masks are applied after the RLE estimate is computed, modifying the replicated output locally. Additionally, a global adjustment factor can be applied uniformly to all pixels. The final RLE channel is obtained by summing the replicated RLE output, the scaled masks, and the global adjustment.

3.3 Training Process

The RLE block is trained jointly with the primary LLIE network, but we use two distinct loss functions: one for the primary network alone and another for the RLE block alone. The primary network loss function estimates the difference between the predicted and the ground truth images.

To ensure that the RLE block accurately represents the enhancement level, we

estimate the optimal enhancement level as:

$$rle\_gt(ll, gt) = avg(gt) - avg(ll), \tag{3.1}$$

where ll is the low-light image, gt is the ground truth, and $avg(I)$ calculates the average intensity of the image I. Essentially, this is equivalent to the difference in brightness between the ground truth image and the low-light image. Although this estimate serves as the ground truth for the RLE block output, it is important to note that the enhancement performed by the combined network is more complex than a brightness modification. We selected the L1 loss function used to train the block. Therefore, the overall loss function can be represented as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{main}}(main(ll), gt) + \lambda \cdot |rle(ll) - rle\_gt(ll, gt)|, \tag{3.2}$$

where $main(I)$ is the output of the primary network, $rle(I)$ is the output of the RLE, and λ is a weight factor for the RLE loss. Our training dataset includes pairs of (low-light, well-exposed) images as well as pairs of (well-exposed, well-exposed) images. For the latter, the ground truth of the RLE block is set to zero, indicating that no enhancement should be applied. Empirical results demonstrate that including well-exposed image pairs in the training process significantly improves the model's ability to control the enhancement level. Without well-exposed pairs, applying extreme adjustments to the enhancement level often leads to noticeable color distortions. In contrast, incorporating well-exposed pairs allows the model to achieve a broader range of adjustments without introducing color artifacts, as illustrated in Figure 3.2.

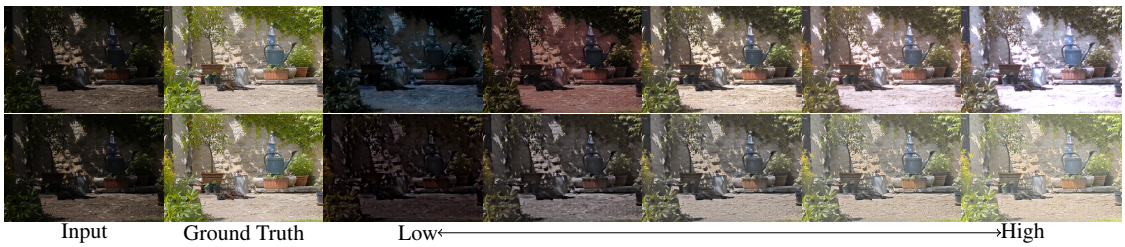
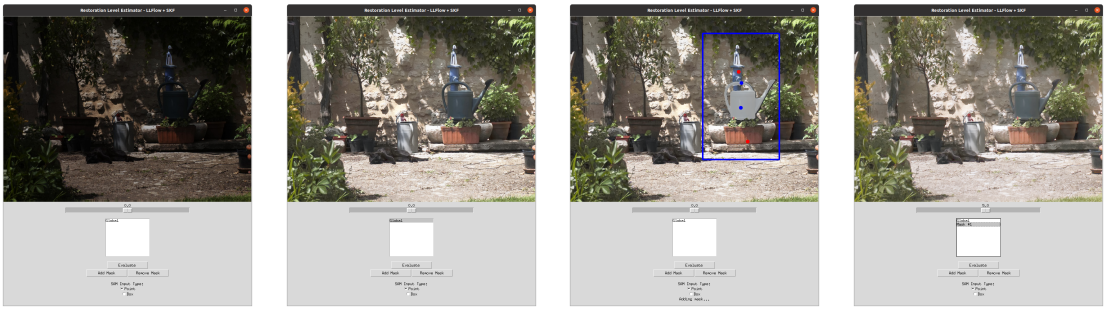


Figure 3.2 – Demonstration of enhancement control using the RLE block with the LLFlow + SKF model. The first row shows results from the model trained without well-exposed pairs, exhibiting color distortions when the enhancement level is drastically increased or decreased. The second row shows results from the model incorporating well-exposed pairs to the training data, avoiding color distortions even at extreme enhancement levels.

3.4 Interactive Enhancement Control

By default, the enhancement level in networks modified with the RLE block corresponds to the optimal value estimated by the block. To provide users with the ability to manually adjust the enhancement level, we developed an application that enables interactive control of the modified networks. The application provides a slider to adjust the enhancement level, letting users see how increasing or decreasing it affects the result.

Additionally, the application supports local enhancement control, as users can specify regions of interest within the image and adjust the enhancement level independently for each region. This feature is useful in scenarios where different parts of the image require varying levels of enhancement. Figure 3.3 shows the application interface, illustrating how users can interactively modify the enhancement level globally or for selected regions.



a) b) c) d)

Figure 3.3 – Screenshots of our user interface for enhancement control of LLFlow + SKF with the RLE block, where: (a) shows the low-light input image; (b) shows the result of LLFlow + SKF with the optimal enhancement level estimated by the RLE block; (c) illustrates the process of creating a mask, represented by the light gray region; When creating a mask, users can specify the region of interest using the Segment Anything Model interface. (d) shows the result after increasing the enhancement level both globally and locally for the region highlighted in (b); The enhancement level can be adjusted globally or per mask by selecting the desired item from the list and moving the slider.

4 RESULTS

One key feature of adding the RLE block to an existing model is the ability to control the enhancement level applied to images. This section illustrates this fact. The resulting network incorporates the RLE block's layers and the input corresponding to the main network module includes an additional channel. To evaluate the potential impact of these changes, we compare the performance of existing models with and without the RLE block.

4.1 Low-Light Image Enhancement Dataset

Several LLIE methods rely on synthetic data generated by darkening well-exposed images and adding noise, such as in LLNet (Lore; Akintayo; Sarkar, 2017), due to the difficulty of acquiring large-scale paired datasets. Synthetic approaches allow controlled data generation and scalable training. However, models trained exclusively on synthetic data often fail to generalize well to real-world low-light conditions, which involve complex noise patterns, color distortions, and lighting inconsistencies. For this reason, real datasets with paired low-light and well-exposed images, such as LOL (Wei et al., 2018) and SICE (Cai; Gu; Zhang, 2018), tend to yield better performance and more visually pleasing results.

Inspired by the SICE dataset (Cai; Gu; Zhang, 2018), which uses multi-exposure image fusion (MEF) to generate high-quality reference images, we explored a similar approach to construct a low-light image enhancement dataset of pictures taken with smartphones. Our goal was to collect burst sequences of images taken at different exposure settings and generate their corresponding ground truth using Multi-Exposure Image Fusion (MEF) techniques. These methods process each sequence to produce a well-exposed result, preserving details in both bright and dark regions.

To capture these images, we developed a smartphone app capable of recording burst sequences with adjustable exposure settings. However, due to delays between consecutive shots, misalignments occurred in dynamic scenes (e.g., moving leaves, shadows, and people), leading to artifacts in the MEF-generated outputs. Further details on the app and dataset collection are provided in Appendix A.

Given this limitation, we considered our dataset unsuitable for training deep-learning models. Instead, we used the LOL dataset (Wei et al., 2018) to evaluate our

method. This dataset is widely used as a benchmark for low-light image enhancement, providing paired low-light and well-exposed images for both training and evaluation.

	PSNR	**SSIM**	**LPIPS**
RetinexNet	18.17	0.6956	0.4041
RetinexNet w/ RLE	19.18	0.7242	0.3472
UNet	19.85	0.8195	0.1281
UNet w/ RLE	20.11	0.8025	0.1255
LLFlow + SKF	25.91	0.8657	0.1933
LLFlow + SKF w/ RLE	25.81	0.8665	0.1249
WaveNet	22.53	0.8419	0.0772
WaveNet w/ RLE	22.49	0.8628	0.0639

Table 4.1 – Performance comparison between the original models and adding the RLE block.

	PSNR	**SSIM**	**LPIPS**
RetinexNet (Original)	16.77	0.462	0.474
RetinexNet (Retrained)	18.17	0.724	0.347
LLFlow + SKF (Original)	25.94	0.865	0.125
LLFlow + SKF (Retrained)	25.91	0.866	0.193
WaveNet (Original)	24.54	0.856	-
WaveNet (Retrained)	22.53	0.8419	0.0772

Table 4.2 – Performance comparison of the original models as reported on their respective papers and their results when retraining them locally.

4.2 Evaluated Models

For our experiments, we selected five CNN-based models and assessed their performance on the LOL dataset with and without the RLE block. The selected models were LLFlow (Wang et al., 2022), LLFlow with SKF (Wu et al., 2023), WaveNet (Dang et al., 2023), RetinexNet (Wei et al., 2018), and UNet (Ronneberger; Fischer; Brox, 2015). They were selected to include both recent (LLFlow, LLFlow with SKF and WaveNet) and classic ones (RetinexNet and UNet). For the LLFlow, LLFlow + SKF, and WaveNet models, we adapted the original code provided by the authors to include the RLE block. The RetinexNet and the UNet models were reimplemented in Keras (Chollet et al., 2015), each with an integrated RLE block. Since the LLFlow + SKF outperforms the LLFlow, here we only show the results for LLFlow + SKF.

4.3 Performance Metrics

To compare performance, we used three commonly adopted evaluation metrics: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS). The PSNR measures the fidelity between a predicted image $\hat{I}$ and a reference image I based on the mean squared error (MSE):

$$PSNR = 10 \cdot \log_{10}\left(\frac{L^2}{MSE}\right), \text{where } MSE = \frac{1}{N}\sum_{i=1}^{N}\left(I_i - \hat{I}_i\right). \tag{4.1}$$

Here, L is the maximum possible pixel value, and N is the total number of pixels. The SSIM evaluates image similarity by considering luminance, contrast, and structure. For image x and y, the SSIM is defined as:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)}, \tag{4.2}$$

where μ_x^2, μ_y^2 are the means, σ_x^2, σ_y^2 the covariance, and C_1, C_2 are small constants to stabilize the division. Finally, the LPIPS metric (Zhang et al., 2018) measures perceptual similarity based on deep feature embeddings from pre-trained networks. It compares activations from multiple layers of a deep neural network (e.g., VGG) and correlates well with human visual perception. Lower LPIPS values indicate higher perceptual similarity.

	PSNR	**SSIM**	**LPIPS**
RetinexNet + SKF	20.42	0.711	0.216
RetinexNet w/ RLE	19.18	0.724	0.347
LLFlow + SKF	25.94	0.865	0.125
LLFlow + SKF w/ RLE	25.81	0.867	0.125

Table 4.3 – Performance comparison of the RetinexNet and LLFlow models adding the SKF and/or the RLE block. We were not able to reproduce the results of LLFlow without SKF to evaluate adding RLE to only LLFlow, and the authors did not provide the pre-trained models nor the code for the RetinexNet with the SKF framework.

4.4 Performance Evaluation

We evaluated the potential impact of the RLE block on the performance of image enhancement networks. Table 4.1 summarizes this evaluation. In the case of RetinexNet and UNet, the addition of the RLE block resulted in a slight improvement in image quality,

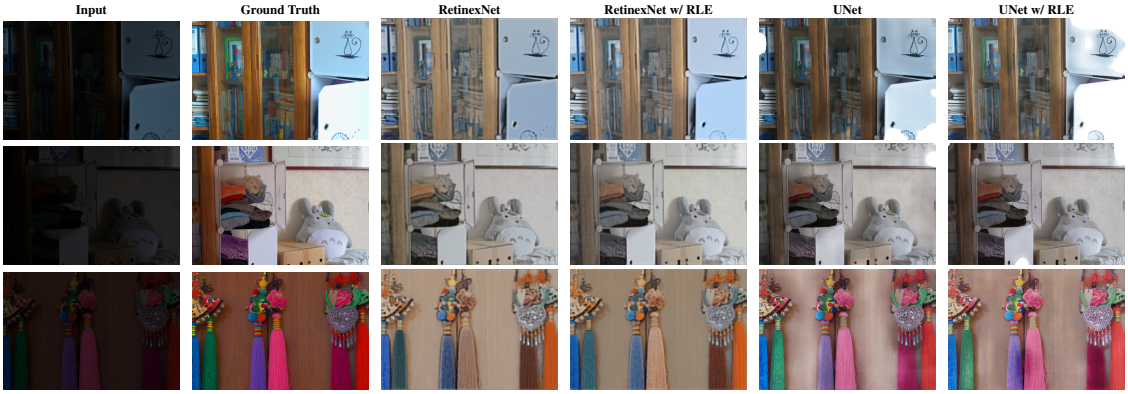


Figure 4.1 – Comparison of RetinexNet and UNet results without and with the RLE block.

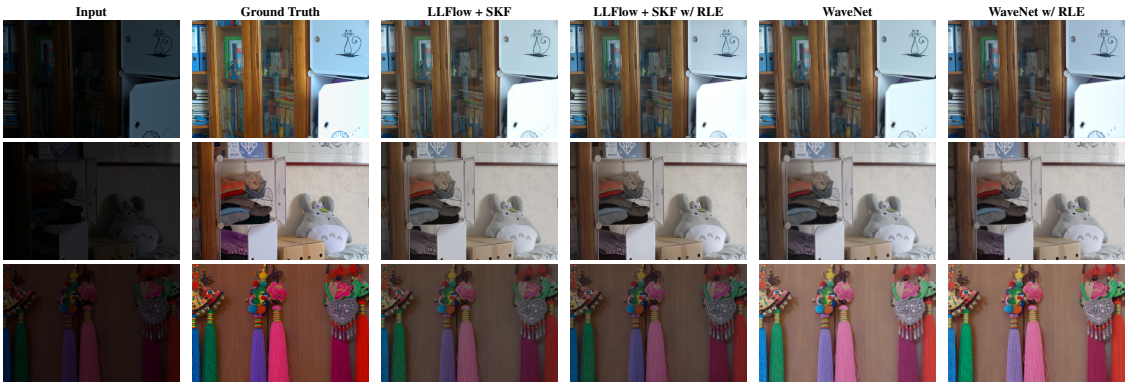


Figure 4.2 – Comparison of LLFlow + SKF and WaveNet results without and with the RLE block.

as supported by the corresponding increases in PSRN, SSIM, and decrease in LPIPS, as shown in Table 4.1. For the LLFlow + SKF and the WaveNet models, the RLE block had minimal impact on the PSNR but showed improvements in SSIM and LPIPS, indicating a potential enhancement in perceptual quality. Besides, the models also benefit from the ability to control their levels of enhancement. Figures 4.3, 4.4, 4.5, and 4.6 compare results from RetinexNet, UNet, LLFlow + SKF, and WaveNet with and without the use of the RLE block. The perceived improvement in visual quality resulting from incorporating the RLE block is more evident in the classic networks (RetinexNet, UNet), though some pre-existing artifacts in the base network outputs remain unchanged. For the more recent networks (LLFlow + SKF, WaveNet), the outputs with and without the RLE block are visually similar, with the primary benefit being the added enhancement control.

Figure 4.7 presents the results of the four LLIE models with the RLE block, highlighting its versatility in providing controllable low-light image enhancement. Figure 4.8 provides additional results demonstrating the enhancement control with the LLFlow + SKF model. By adjusting the scaling value, users can achieve fine control over the level of enhancement applied to the input images: decreasing the value results in less intense enhancement, while increasing it produces a brighter output. This flexibility allows users

Figure 4.3 – Comparison of RetinexNet results without and with the RLE block. The results with the RLE block show a slight improvement, particularly in color saturation.

to intuitively adjust the level of enhancement to their specific needs. Figure 4.9 compares our method with the CLE Diffusion and with adjusting the input's contrast. Unlike contrast modification, which is able to darken the image but fails to brighten it, both CLE Diffusion and our approach offer smoother and more realistic control over the enhancement level.

It is also possible to use the RLE block to perform local enhancement control independently over multiple regions, leaving the brightness levels of surrounding areas unaffected. Each targeted region is specified using a mask, allowing for individual adjustments. Increasing a mask's scaling value progressively brightens the corresponding region, while decreasing it darkens the region. This is illustrated in Figures 4.10 and 4.11. One should note that, when performing local enhancement control, the boundary of the selected regions might become overly bright as the enhancement level increases beyond a certain limit.

4.5 Reproducibility Challenges

Reproducing results from deep learning-based models can often be challenging due to various factors, such as the lack of precise implementation details or differences in hardware and software environments. Table 4.2 presents a comparison between the

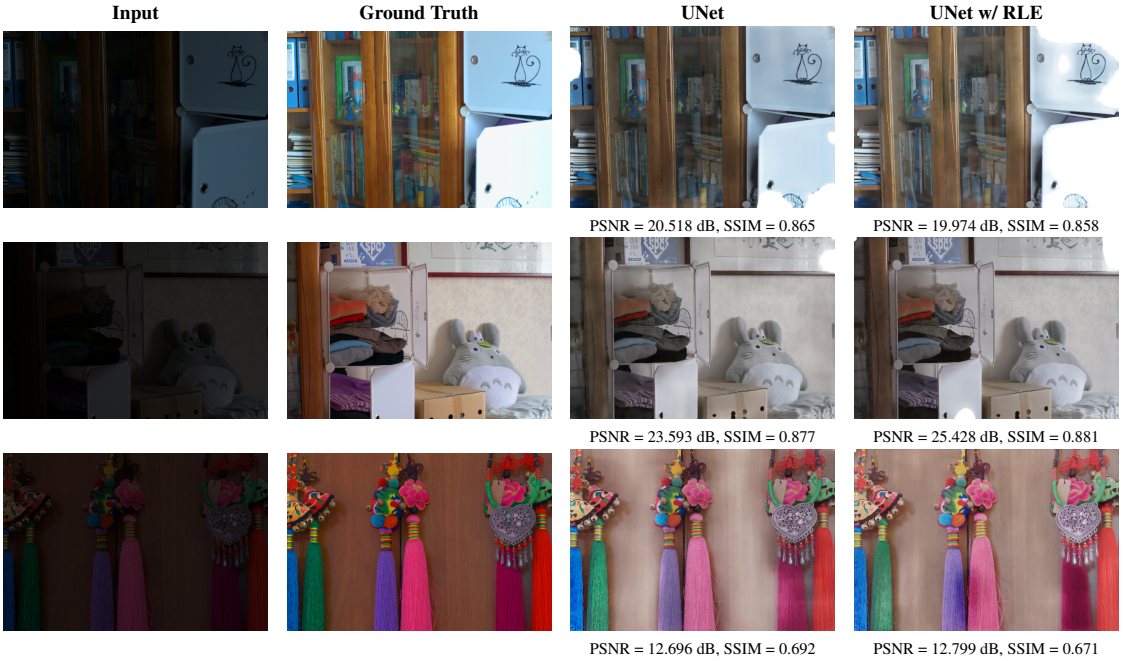


Figure 4.4 – Comparison of UNet results without and with the RLE block. The results with the RLE block show a slight improvement in color saturation. UNet pre-existing artifacts at the bottom right of the image in the first row are minimized although they are still visible around the cat drawing in the top right.

original results reported by the authors for the LOL dataset and the results we obtained by retraining these models using the parameters and number of epochs reported in their respective papers. In some cases, such as with RetinexNet, we observed significant improvements in PSNR and SSIM upon retraining. For LLFlow + SKF, the retrained model produced similar results to the original, although with a slight increase in LPIPS, suggesting minor variation in perceptual quality. However, for WaveNet, the retrained model performed worse than the reported in terms of PSNR and SSIM. We found it hard to reproduce the results of LLFlow without the SKF framework, as we were unable to successfully retrain the model despite following the authors's guidelines.

4.6 Comparison with SKF

Although a direct comparison between RLE and SKF is challenging due to reproducibility issues, Table 4.3 provides a comparison of the available results. For RetinexNet, we were able to evaluate its performance with the RLE block and compare it to the performance of RetinexNet with the SKF framework using the results provided by the authors. However, we could not evaluate RetinexNet with both RLE and SKF simultaneously because the code was not made available. Similarly, while we successfully evaluated

Figure 4.5 – Comparison of LLFlow + SKF results without and with the RLE block. The outputs are visually similar, demonstrating that the RLE block preserves the original enhancement quality while providing additional user control.

LLFlow with both SKF and RLE, we could not integrate RLE into LLFlow alone.

From the available results, we noticed that adding SKF to models like RetinexNet and LLFlow has a more significant impact on performance metrics, especially in terms of PSNR and SSIM. This suggests that SKF improves image quality more effectively than RLE in terms of traditional metrics. However, the RLE block introduces a unique advantage by enabling users to control the level of enhancement, offering flexibility that SKF does not provide. Thus, while SKF tends to boost performance more significantly, RLE offers the benefit of controllability without severely impacting model's performance.

As demonstrated in the case of LLFlow, RLE and SKF can be combined, yielding the performance improvements associated with SKF alongside the enhancement control offered by RLE. However, we were unable to test this combination with other models due to reproducibility issues with the provided SKF implementations and the challenges of adapting SKF to other models.

Figure 4.6 – Visual comparison of WaveNet results without and with the RLE block. As in Figure 4.5, the outputs are visually similar, demonstrating that the RLE block preserves the original enhancement quality while enabling additional user control.

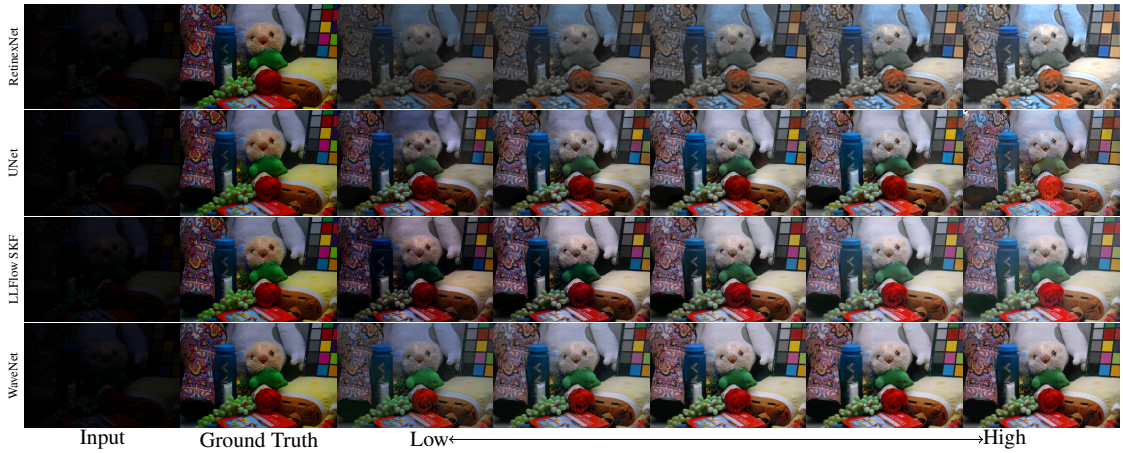


Figure 4.7 – Results of RetinexNet, UNet, LLFlow + SKF, and WaveNet models as the estimated value of the RLE block is adjusted. Increasing the value produces a brighter image, while decreasing it results in a darker image. The enhancement level range varies across models, and for this example, the values were selected to produce results with comparable overall illumination in each column.

Figure 4.8 – Additional results demonstrating the enhancement control of the LLFlow + SKF model after incorporating the RLE block, with varying RLE values shown across different images.

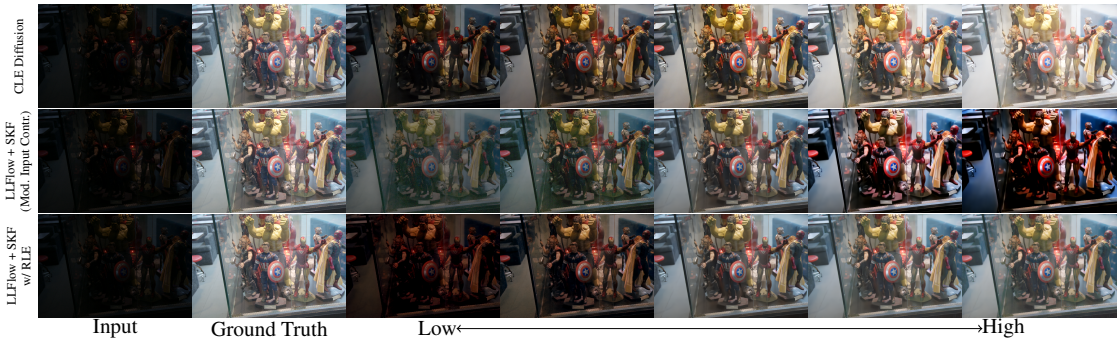


Figure 4.9 – Comparison of enhancement control using CLE Diffusion, LLFlow + SKF with input contrast modification, and LLFlow + SKF with our proposed RLE block. While reducing contrast in the input image leads to a darker output, increasing it does not effectively brighten the image. Both CLE Diffusion and our RLE-based approach produce more natural and perceptually consistent enhancements across varying levels, unlike simple contrast manipulation.

Figure 4.10 – Local enhancement control with the RLE block. The (highlighted) mask in the first image indicates the region targeted for adjustable enhancement. In the subsequent images, increasing the mask's scaling value progressively brightens the selected region, while decreasing it darkens the region, leaving the brightness levels of surrounding areas unaffected. However, the boundary of the selected region tends to become overly bright as the enhancement level increases.

a) b) c) d) e)

Figure 4.11 – Independent enhancement of multiple image regions using the RLE block. (a) Reference image. (b) The two highlighted masks indicate the regions targeted for enhancement. (c) Increased enhancement of the left mask's region; (d) Increased enhancement of the right mask's region. (e) Increased enhancement of both masks' regions.

5 CONCLUSION AND FUTURE WORK

This work presented a novel approach for low-light image enhancement control by introducing the Restoration Level Estimator block. Our method allows for both global and local control over the enhancement level, adding flexibility and user control to existing low-light image enhancement networks. The RLE block can be integrated into CNN-based models, providing users the ability to fine-tune the level of enhancement applied to input images.

We evaluated the impact of adding the RLE block to some recent (LLFlow + SKF, WaveNet) as well as classic (RetinexNet, UNet) LLIE models. Although the impact on performance varied across models, we demonstrated that interactively modifying the value predicted by the RLE block directly adjusts the enhancement level of the input image.

In addition to global control, our approach offers local control, allowing users to adjust the enhancement levels of specific areas within an image. By using masks generated with the Segment Anything Model, users can adapt the enhancement level for multiple regions independently. Since the RLE block does not correct pre-existing artifacts from the base model, any distortion present in the original enhancement network is likely to persist.

Overall, the RLE block represents an important step toward making LLIE methods more flexible and adaptable to user preferences. As future work, we would like to extend this approach to include diffusion-based and transformer-based models, as well as investigate its application in low-light video enhancement. In particular, we are interested in exploring the use of the DID dataset (Fu et al., 2023) and potentially integrating the RLE block with recent video-based methods such as LAN (Fu et al., 2023). We also intend to address the overly-bright boundaries that can occur around masked regions during local enhancement.

REFERENCES

ABDULLAH-AL-WADUD, M. et al. A dynamic histogram equalization for image contrast enhancement. **IEEE Transactions on Consumer Electronics**, v. 53, n. 2, p. 593–600, 2007.

BRATEANU, A. et al. Lyt-net: Lightweight yuv transformer-based network for low-light image enhancement. **arXiv preprint arXiv:2401.15204**, 2024.

CAI, J.; GU, S.; ZHANG, L. Learning a deep single image contrast enhancer from multi-exposure images. **IEEE Transactions on Image Processing**, v. 27, n. 4, p. 2049–2062, 2018.

CAI, Y. et al. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In: **2023 IEEE/CVF International Conference on Computer Vision (ICCV)**. [S.l.: s.n.], 2023. p. 12470–12479.

CHOLLET, F. et al. **Keras**. 2015. <https://keras.io>.

DANG, J. et al. WaveNet: Wave-Aware Image Enhancement. p. 9 pages, 2023. Available from Internet: <https://diglib.eg.org/handle/10.2312/pg20231267>.

DIAMOND, S. et al. Dirty pixels: Towards end-to-end image processing and perception. **ACM Transactions on Graphics (SIGGRAPH)**, ACM, 2021.

FU, H. et al. Dancing in the dark: A benchmark towards general low-light video enhancement. In: **2023 IEEE/CVF International Conference on Computer Vision (ICCV)**. [S.l.: s.n.], 2023. p. 12831–12840.

GUO, C. et al. Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. In: **2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**. Seattle, WA, USA: IEEE, 2020. p. 1777–1786. ISBN 9781728171685. Available from Internet: <https://ieeexplore.ieee.org/document/9157813/>.

GUO, X.; LI, Y.; LING, H. Lime: Low-light image enhancement via illumination map estimation. **IEEE Transactions on Image Processing**, v. 26, n. 2, p. 982–993, 2017.

HE, J.; DONG, C.; QIAO, Y. Interactive Multi-dimension Modulation with Dynamic Controllable Residual Learning for Image Restoration. In: VEDALDI, A. et al. (Ed.). **Computer Vision – ECCV 2020**. Cham: Springer International Publishing, 2020. p. 53–68. ISBN 9783030585655.

HOU, J. et al. Global structure-aware diffusion process for low-light image enhancement. **Advances in Neural Information Processing Systems**, 2023.

JIANG, B. et al. FBGAN: multi-scale feature aggregation combined with boosting strategy for low-light image enhancement. **Vis. Comput.**, Springer Science and Business Media LLC, v. 40, n. 3, p. 1745–1756, mar. 2024.

JIANG, J.; ZHANG, K.; TIMOFTE, R. Towards Flexible Blind JPEG Artifacts Removal. In: **2021 IEEE/CVF International Conference on Computer Vision (ICCV)**. [s.n.], 2021. p. 4977–4986. ISSN: 2380-7504. Available from Internet: <https://ieeexplore.ieee.org/document/9711459>.

JIANG, Y. et al. EnlightenGAN: Deep Light Enhancement Without Paired Supervision. **IEEE Transactions on Image Processing**, v. 30, p. 2340–2349, 2021. ISSN 1057-7149, 1941-0042. Available from Internet: <https://ieeexplore.ieee.org/document/9334429/>.

KIRILLOV, A. et al. **Segment Anything**. 2023. Available from Internet: <https://arxiv.org/abs/2304.02643>.

LAND, E. H.; MCCANN, J. J. Lightness and retinex theory. **J. Opt. Soc. Am.**, Optica Publishing Group, v. 61, n. 1, p. 1–11, Jan 1971. Available from Internet: <https://opg.optica.org/abstract.cfm?URI=josa-61-1-1>.

LI, C. et al. Low-light image and video enhancement using deep learning: A survey. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 44, n. 12, p. 9396–9416, 2022.

LI, H. et al. Detail-preserving multi-exposure fusion with edge-preserving structural patch decomposition. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 31, n. 11, p. 4293–4304, 2021.

LORE, K. G.; AKINTAYO, A.; SARKAR, S. LLNet: A deep autoencoder approach to natural low-light image enhancement. **Pattern Recognition**, v. 61, p. 650–662, jan. 2017. ISSN 00313203. Available from Internet: <https://linkinghub.elsevier.com/retrieve/pii/S003132031630125X>.

NI, S. et al. Controlling the Rain: from Removal to Rendering. In: **2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**. [s.n.], 2021. p. 6324–6333. ISSN: 2575-7075. Available from Internet: <https://ieeexplore.ieee.org/document/9578173>.

PAN, H. et al. DICNet: achieve low-light image enhancement with image decomposition, illumination enhancement, and color restoration. **Vis. Comput.**, Springer Science and Business Media LLC, feb. 2024.

QIAO, Z. et al. Deep semi-supervised learning for low-light image enhancement. In: **2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)**. [S.l.: s.n.], 2021. p. 1–6.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: NAVAB, N. et al. (Ed.). **Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015**. Cham: Springer International Publishing, 2015. p. 234–241. ISBN 978-3-319-24574-4.

SHOSHAN, A.; MECHREZ, R.; ZELNIK-MANOR, L. Dynamic-Net: Tuning the Objective Without Re-Training for Synthesis Tasks. In: **2019 IEEE/CVF International Conference on Computer Vision (ICCV)**. Seoul, Korea (South): IEEE, 2019. p. 3214–3222. ISBN 9781728148038. Available from Internet: <https://ieeexplore.ieee.org/document/9008765/>.

SINGH, K.; PARIHAR, A. S. Illumination estimation for nature preserving low-light image enhancement. **Vis. Comput.**, Springer Science and Business Media LLC, jan. 2023.

WANG, R. et al. Underexposed photo enhancement using deep illumination estimation. In: **2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2019. p. 6842–6850.

WANG, W. et al. CFSNet: Toward a Controllable Feature Space for Image Restoration. In: **2019 IEEE/CVF International Conference on Computer Vision (ICCV)**. [s.n.], 2019. p. 4139–4148. ISSN: 2380-7504. Available from Internet: <https://ieeexplore.ieee.org/document/9009077>.

WANG, W. et al. An experiment-based review of low-light image enhancement methods. **IEEE Access**, v. 8, p. 87884–87917, 2020.

WANG, X. et al. Deep Network Interpolation for Continuous Imagery Effect Transition. In: **2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**. [s.n.], 2019. p. 1692–1701. ISSN: 2575-7075. Available from Internet: <https://ieeexplore.ieee.org/document/8953362>.

WANG, Y. et al. Low-Light Image Enhancement with Normalizing Flow. **Proceedings of the AAAI Conference on Artificial Intelligence**, v. 36, n. 3, p. 2604–2612, jun. 2022. ISSN 2374-3468, 2159-5399. Available from Internet: <https://ojs.aaai.org/index.php/AAAI/article/view/20162>.

WEI, C. et al. Deep Retinex Decomposition for Low-Light Enhancement. 2018. Available from Internet: <https://arxiv.org/abs/1808.04560>.

WU, Y. et al. Learning Semantic-Aware Knowledge Guidance for Low-Light Image Enhancement. In: **2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**. Vancouver, BC, Canada: IEEE, 2023. p. 1662–1671. ISBN 9798350301298. Available from Internet: <https://ieeexplore.ieee.org/document/10205052/>.

XU, D. et al. ReCoRo: Re gion- Co ntrollable Ro bust Light Enhancement with User-Specified Imprecise Masks. In: **Proceedings of the 30th ACM International Conference on Multimedia**. Lisboa Portugal: ACM, 2022. p. 1376–1386. ISBN 9781450392037. Available from Internet: <https://dl.acm.org/doi/10.1145/3503161.3547813>.

XU, F. et al. Multi-exposure image fusion techniques: A comprehensive review. **Remote Sensing**, v. 14, n. 3, 2022. ISSN 2072-4292. Available from Internet: <https://www.mdpi.com/2072-4292/14/3/771>.

XU, X. et al. SNR-Aware Low-light Image Enhancement. In: **2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**. New Orleans, LA, USA: IEEE, 2022. p. 17693–17703. ISBN 9781665469463. Available from Internet: <https://ieeexplore.ieee.org/document/9878461/>.

YAO, M. et al. Toward Interactive Self-Supervised Denoising. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 33, n. 10, p. 5360–5374, oct. 2023. ISSN 1558-2205. Available from Internet: <https://ieeexplore.ieee.org/abstract/document/10059001>.

YIN, Y. et al. CLE Diffusion: Controllable Light Enhancement Diffusion Model. In: **Proceedings of the 31st ACM International Conference on Multimedia**. Ottawa ON Canada: ACM, 2023. p. 8145–8156. ISBN 9798400701085. Available from Internet: <https://dl.acm.org/doi/10.1145/3581783.3612145>.

YU, X.; LI, H.; YANG, H. Two-stage image decomposition and color regulator for low-light image enhancement. **Vis. Comput.**, Springer Science and Business Media LLC, jul. 2022.

ZHANG, J. et al. Invertible network for unpaired low-light image enhancement. **Vis. Comput.**, Springer Science and Business Media LLC, v. 40, n. 1, p. 109–120, jan. 2024.

ZHANG, R. et al. The unreasonable effectiveness of deep features as a perceptual metric. In: **2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2018. p. 586–595.

ZHANG, Y. et al. Beyond brightening low-light images. **Int. J. Comput. Vis.**, Springer Science and Business Media LLC, v. 129, n. 4, p. 1013–1037, abr. 2021.

ZHANG, Y.; ZHANG, J.; GUO, X. Kindling the darkness: A practical low-light image enhancer. In: **Proceedings of the 27th ACM International Conference on Multimedia**. New York, NY, USA: ACM, 2019. (MM '19), p. 1632–1640. ISBN 978-1-4503-6889-6. Available from Internet: <http://doi.acm.org/10.1145/3343031.3350926>.

ZHOU, D.; YANG, Z.; YANG, Y. Pyramid diffusion models for low-light image enhancement. In: ELKIND, E. (Ed.). **Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23**. International Joint Conferences on Artificial Intelligence Organization, 2023. p. 1795–1803. Main Track. Available from Internet: <https://doi.org/10.24963/ijcai.2023/199>.

APPENDIX A — CONTROLLABLE BURST PHOTOGRAPHY ON SMARTPHONES FOR MULTI-EXPOSURE IMAGE FUSION

This appendix details the development and evaluation of a smartphone app designed for capturing burst sequences of images with user-specified ISO and exposure time settings. The primary goal of this app is to enable the acquisition of multi-exposure image sequences for use in multi-exposure image fusion (MEF) techniques (Xu et al., 2022b), facilitating the creation of well-exposed ground-truth images from a range of lighting conditions.

Inspired by the SICE dataset (Cai; Gu; Zhang, 2018), this project sought to extend the idea of multi-exposure imaging to smartphone cameras, focusing on the Xiaomi Redmi Note 11S, a widely available device. Unlike the SICE dataset, which was collected using a combination of DSLRs and other consumer-grade cameras, this work emphasizes data collected exclusively with smartphones. Smartphones are increasingly the primary imaging devices for most users, making datasets derived from them especially relevant for applications such as low-light image enhancement.

In the following sections, we describe the implementation of the app, present examples of sequences captured with it, demonstrate MEF results using these sequences, and discuss the limitations of the current approach.

A.1 Controllable Burst Photography App

To facilitate the collection of burst sequences with varied lighting settings, we developed a custom smartphone application using Kotlin and Android Studio. The app uses the Camera2 API to precisely control camera parameters, enabling users to adjust ISO and exposure time for each capture. Other settings, such as focus and white balance, were left in automatic mode and configured for optimal image quality.

The app's main interface allows users to define and manage a sequence of capture settings. It displays a list of user-specified parameters that can be easily edited or removed. Additionally, the app includes a live camera preview feature, enabling users to see how the current settings affect the image. This functionality is particularly useful for adjusting parameters to match the lighting conditions of the scene being captured.

A.1.1 Capturing Burst Sequences

The burst sequences of images were captured using the app with the smartphone mounted on a tripod to ensure stability throughout the process. To further minimize camera movement, the burst was triggered using a Bluetooth controller, capturing all images sequentially. All captures were taken in outdoor environments, chosen for their more challenging lighting conditions.

Figure A.1 – Two examples of multi-exposure image sequences captured using our custom app. In each sequence, the exposure time increases from left to right. Images with lower exposure times preserve details in brighter regions, while those with higher exposure times exhibit saturation in bright areas but reveal more details in darker regions.

Figure A.1 presents examples of sequences captured with the app. One notable limitation of our approach is the significant delay between each image capture within a burst. This delay makes it difficult to obtain perfectly aligned sequences when there is movement within the scene. In such cases, the resulting images of the sequence are not perfectly aligned, as illustrated in Figure A.2.

A.2 Multi-Exposure Image Fusion

Traditional digital imaging systems often lose structural details in high dynamic range scenes due to their limited brightness capture capabilities. Multi-exposure fusion (MEF) techniques address this by combining multiple images in the non-linear brightness domain to produce high-visibility results (Xu et al., 2022b). For generating well-exposed ground-truth images from our dataset's burst sequences, we employed the MESPD-MEF method (Li et al., 2021). Its edge-preserving approach effectively reduces halo effects while enhancing detail retention in both bright and dark regions, offering a reliable solution for producing high-quality fused images from exposure-varied sequences.

The burst sequences were processed using the source code provided by the authors

Figure A.2 – A subset of images from a burst sequence captured with our custom app, illustrating misalignment issues in dynamic scenes, here caused by wind. The first row displays the burst images, while the second row shows magnified views of the highlighted region. The differences between the images are particularly noticeable in the branches of the tree located in the top-right corner of the highlighted areas.

a) b) c)

Figure A.3 – Results of the multi-exposure image fusion technique applied to our burst sequences. The fused images in (a) and (b) demonstrate a higher dynamic range than any single image in their respective input sequences, successfully preserving details in both bright and dark areas. However, in (c), some regions, such as the tree branches, appear blurry due to misalignment caused by movement from wind during capture.

of the MESPD-MEF method, running on a desktop computer. Figure A.3 showcases the results of this method applied to the previously shown image sequences. For the first two sequences, the method produced high-quality results with a wider dynamic range than any individual input image, effectively preserving structural details and enhancing visibility. However, for the final sequence, despite the improved dynamic range, certain regions, particularly the leaves on the trees, appeared blurry. This was caused by wind during capture, leading to misalignment between images in the sequence. Due to this limitation, we determined that collecting a dataset with approach was unsuitable for training machine-learning models.

APPENDIX B — RESUMO EXPANDIDO

O aprimoramento de imagens em baixa luminosidade é uma tarefa desafiadora, influenciada por fatores como iluminação e configurações da câmera. Além de melhorar o aspecto visual, busca-se tornar as imagens mais adequadas para tarefas como detecção de objetos. No entanto, soluções tradicionais muitas vezes ignoram as preferências do usuário ou necessidades regionais da imagem, tornando desejável oferecer controle sobre o processo de aprimoramento.

Com os avanços do aprendizado profundo, novas técnicas têm melhorado significativamente o aprimoramento de imagens. No entanto, a maioria aplica ajustes de forma uniforme, o que pode ser inadequado para imagens com regiões que requerem níveis distintos de aprimoramento. A percepção subjetiva do que constitui uma boa iluminação reforça a necessidade de oferecer controle ao usuário.

Neste trabalho, propomos o Estimador de Nível de Restauração (*Restoration Level Estimator* - RLE), um módulo que pode ser integrado a redes de aprimoramento existentes para permitir controle adaptativo, tanto global quanto local. Usuários podem ajustar o nível de aprimoramento para a imagem inteira ou para regiões específicas, definidas com auxílio do *Segment Anything Model* (SAM).

Avaliações quantitativas e qualitativas demonstram que o RLE mantém a qualidade do aprimoramento original, medido por métricas como PSNR e SSIM, enquanto adiciona maior controle ao usuário. A Figura B.1 mostra uma visão geral do método, destacando como o módulo RLE permite controle global e local sobre o aprimoramento.

As **contribuições** desse trabalho incluem:

- Um módulo que adiciona controle de nível de aprimoramento a métodos existentes;
- Um mecanismo para ajustar níveis de aprimoramento em regiões específicas da imagem;
- Experimentos que demonstram a o controle do nível de aprimoramento com o RLE e seu impacto em redes existentes.

B.1 Trabalhos Relacionados

Esta seção apresenta os principais trabalhos relacionados ao aprimoramento de imagens em baixa luminosidade (*Low-Light Image Enhancement* - LLIE).

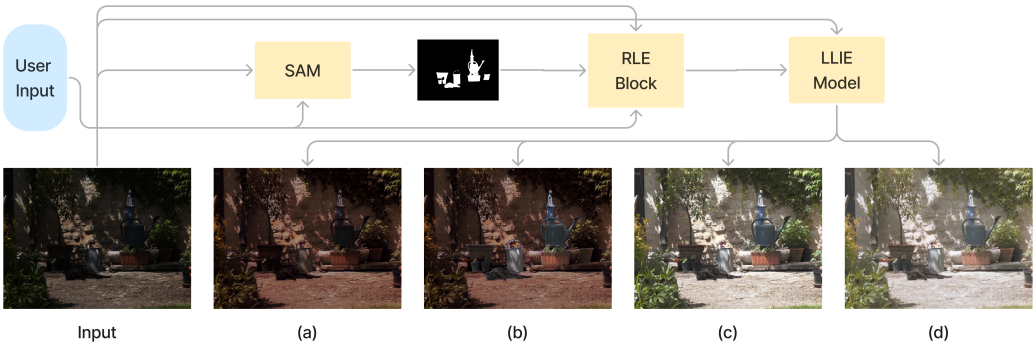


Figure B.1 – Visão geral do nosso método e demonstração de suas capacidades de controle para aprimoramento de imagens em baixa luminosidade. O diagrama mostra o pipeline de processamento: primeiro o usuário define regiões de interesse e o bloco RLE estima o nível ideal de aprimoramento, que é usado como entrada adicional para o modelo LLIE (LLFlow + SKF, neste exemplo). O usuário pode ajustar o nível estimado modificando seu valor tanto globalmente quanto localmente. A linha inferior exibe os resultados: (a) o nível de aprimoramento é reduzido globalmente; (b) o aprimoramento global permanece inalterado, mas regiões selecionadas são iluminadas; (c) a imagem é aprimorada com o nível ideal estimado pelo RLE; (d) o nível de aprimoramento é aumentado globalmente, resultando em uma imagem mais clara.

B.1.1 Aprimoramento de Imagens em Baixa Luminosidade

O aprimoramento de imagens em baixa luminosidade é um tema amplamente estudado, com abordagens tradicionais baseadas em equalização de histograma (Abdullah-Al-Wadud et al., 2007) e na teoria de Retinex (Guo; Li; Ling, 2017). Contudo, métodos baseados em aprendizado profundo têm alcançado resultados superiores, com melhor preservação de detalhes e redução de ruído. Nesse contexto, diferentes estratégias têm sido exploradas, incluindo *autoencoders* (Lore; Akintayo; Sarkar, 2017), variações baseadas no modelo Retinex (Wei et al., 2018), redes adversariais generativas (Jiang et al., 2021), e *normalizing flows* (Wang et al., 2022). Com o objetivo de aprimorar os resultados de métodos existentes, Wu et al. (Wu et al., 2023) propuseram o SKF, um *framework* guiado por semântica que pode ser incorporado a outras redes, garantindo consistência de cores e texturas naturais.

Embora esses métodos representem avanços significativos, eles produzem resultados com níveis de aprimoramento fixos, limitados às configurações aprendidas durante o treinamento. Nosso método, ao contrário, combina estimativas automáticas de aprimoramento com controle ajustável, permitindo adaptações globais e regionais conforme as preferências do usuário.

B.1.2 Restauração de Imagens com Controle Ajustável

Trabalhos recentes têm explorado formas de restaurar imagens com controle ajustável. Métodos como o de He et al. (He; Dong; Qiao, 2020) e Wang et al. (Wang et al., 2019b) introduziram redes que permitem ajustar níveis de restauração, equilibrando fidelidade ao sinal e qualidade visual. Técnicas como Dynamic-Net (Shoshan; Mechrez; Zelnik-Manor, 2019) e Deep Network Interpolation (DNI) (Wang et al., 2019) permitem transições suaves entre diferentes objetivos de restauração, controlados por parâmetros ajustáveis. Essas técnicas foram aplicadas para diferentes problemas de restauração de imagens, como remoção de ruído, JPEG *deblocking*, e *deraining*.

Especificamente para LLIE, ReCoRo (Xu et al., 2022a) introduziu um *framework* que permite o controle do nível de aprimoramento e usa máscaras imprecisas para guiar ajustes regionais. CLE Diffusion (Yin et al., 2023) propôs uma técnica para controle de aprimoramento baseada em difusão, utilizando o *Segment Anything Model* (Kirillov et al., 2023) para especificar as máscaras para controle local.

Nosso método compartilha o objetivo de oferecer controle ao usuário, como ReCoRo e CLE, mas diferencia-se por ser um módulo versátil que pode ser integrado a redes LLIE já existentes. Essa modularidade permite complementar diferentes arquiteturas, preservando suas vantagens originais.

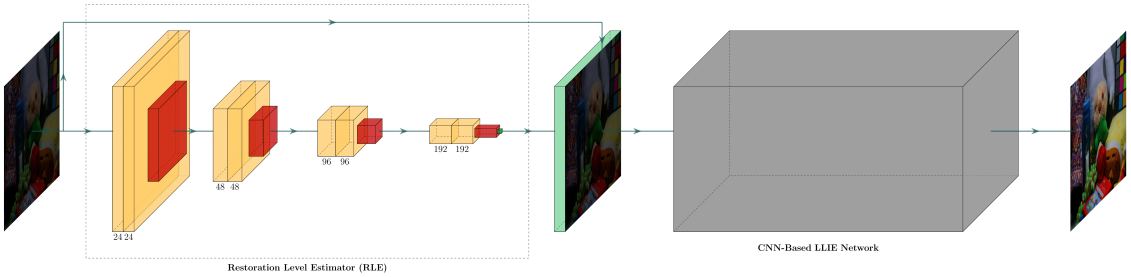


Figure B.2 – Arquitetura do bloco RLE, destacado pelo retângulo tracejado. O bloco RLE recebe uma imagem com baixa luminosidade como entrada e a processa através de blocos de camadas convolucionais seguidas de *max pooling*. Após o último bloco, uma camada de *global average pooling* gera um escalar, que é transformado em um tensor com as mesmas dimensões espaciais da imagem de entrada. Esse tensor é concatenado com a imagem original para formar a entrada de uma rede LLIE baseada em redes neurais convolucionais (caixa cinza).

B.2 Controlando o Aprimoramento de Imagens em Baixa Luminosidade

Este trabalho propõe um método que aprimora redes existentes de melhoria de imagens em baixa luminosidade, introduzindo um bloco chamado Estimador de Nível de Restauração (Restoration Level Estimator - RLE). O RLE combina uma estimativa automática de nível de aprimoramento com controle interativo, gerando resultados bem iluminados sem necessidade de interação mas ao mesmo tempo possibilitando ajustes conforme as preferências do usuário. Além disso, o nosso método integra um controle local inspirado em trabalhos recentes, permitindo que diferentes regiões de uma mesma imagem sejam tratadas de forma independente.

B.2.1 Estimador de Nível de Restauração

O bloco RLE é um módulo compacto adicionado à rede de aprimoramento de baixa luminosidade, com o objetivo de estimar automaticamente o nível ideal de aprimoramento para uma imagem. A sua arquitetura está ilustrada na Figura B.2, e consiste de blocos consecutivos de camadas convolucionais seguidas por max pooling, com uma camada de global average pooling ao final para gerar um valor escalar que representa o nível de aprimoramento a ser aplicado na imagem. Esse valor é replicado para formar um tensor com as mesmas dimensões espaciais da imagem original, que é então concatenado à imagem original como um canal adicional, servindo de entrada para a rede principal de aprimoramento.

Como o RLE apenas estima o nível de aprimoramento, ele deve ser integrado a uma rede principal responsável pelo aprimoramento, e os dois módulos são treinados simultaneamente. A função de loss da rede principal é calculada com base na diferença entre a imagem gerada e a imagem de referência. Já a função de loss do RLE considera a diferença absoluta entre o nível de aprimoramento estimado pelo RLE e um nível ideal calculado como a diferença média de intensidade entre a imagem de entrada e a de referência. Este treinamento conjunto permite ao RLE capturar informações sobre a luminosidade ideal para diferentes condições de iluminação, enquanto a rede principal aprende a utilizar a saída do RLE para guiar o nível de aprimoramento.

B.2.2 Controle Interativo de Aprimoramento

Além do aprimoramento automático, o RLE possibilita o controle interativo da intensidade do aprimoramento. A interface desenvolvida permite que o usuário facilmente controle o nível de aprimoramento global atrvés de um slider. Ao modificar o slider, o seu valor é somado à saída do RLE, gerando imagens mais claras ou mais escuras de acordo com suas preferências.

Além de controle global também adicionamos um controle local inspirado em abordagens recentes, como ReCoRo (Xu et al., 2022a) e CLE Diffusion (Yin et al., 2023). No caso do RLE, nós alcançamos controle global utilizando o modelo Segment Anything (SAM) (Kirillov et al., 2023) para facilitar a geração de máscaras binárias para as regiões de interesse. A cada máscara gerada, o valor estimado pelo RLE pode ser alterado de forma independente, aumentando ou diminuindo o nível de aprimoramento na região correspondente. Para isso, basta selecionar a máscara desejada e modificar o slider. Esse mecanismo permite tratar detalhes locais da imagem de maneira flexível e adaptada às necessidades de cada usuário, como dar mais destaque a algum elemento da imagem, ou iluminar áreas subexpostas sem afetar regiões já bem iluminadas.

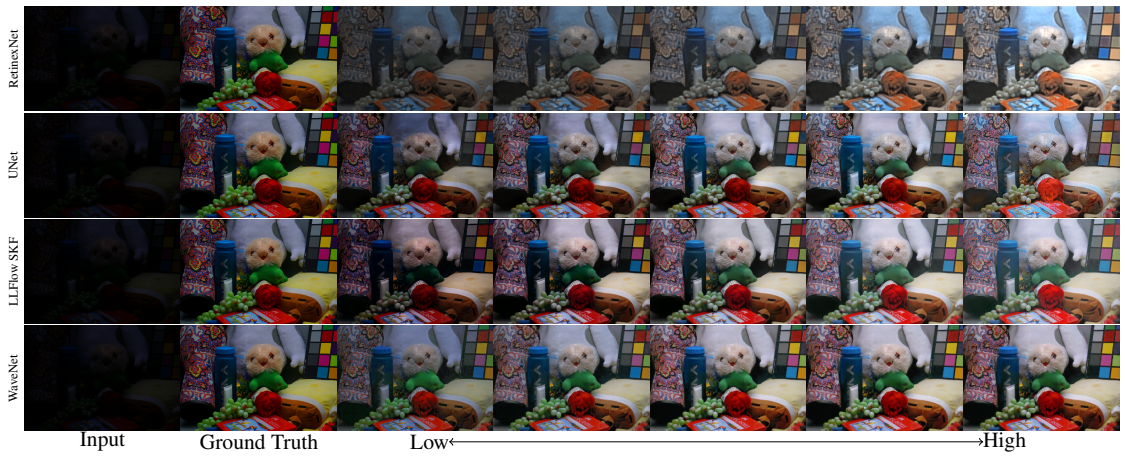


Figure B.3 – Resultados dos modelos RetinexNet, UNet, LLFlow + SKF e WaveNet conforme o valor estimado pelo bloco RLE é ajustado. O aumento do valor gera uma imagem mais clara, enquanto a redução resulta em uma imagem mais escura.

B.3 Resultados

Nessa Seção discutiremos os experimentos realizados para avaliar o nosso módulo tanto em termos mais quantitativos, com as métricas PSNR, SSIM e LPIPS, quanto sub-

Figure B.4 – Controle local de aprimoramento com o bloco RLE. A máscara destacada em azul na primeira imagem indica a região alvo para o aprimoramento ajustável. Nas imagens subsequentes, o aumento do valor de escala da máscara torna a região selecionada progressivamente mais brilhante, enquanto a redução do valor a escurece, sem alterar os níveis de brilho das áreas ao redor.

jetivamente, ao demonstrar o controle de nível de aprimoramento obtido ao adicionar o nosso bloco a uma rede existente. Uma característica central da adição do bloco RLE a um modelo existente é a capacidade de controlar o nível de aprimoramento aplicado às imagens. Esta seção ilustra esse fato, avaliando o impacto do bloco RLE em redes de aprimoramento de imagens e comparando seu desempenho com outros métodos.

Para avaliar as mudanças estruturais introduzidas pelo RLE, foram selecionados cinco modelos baseados em CNN, avaliados no conjunto de dados LOL (Wei et al., 2018): LLFlow (Wang et al., 2022), LLFlow com SKF (Wu et al., 2023), WaveNet (Dang et al., 2023), RetinexNet (Wei et al., 2018) e UNet (Ronneberger; Fischer; Brox, 2015). Esses modelos incluem redes recentes (LLFlow, LLFlow com SKF e WaveNet) e clássicas (RetinexNet e UNet). Para os modelos LLFlow, LLFlow + SKF e WaveNet, o código original foi adaptado para incluir o bloco RLE. Os modelos RetinexNet e UNet foram reimplementados em Keras (Chollet et al., 2015), com a integração do RLE.

Os resultados, resumidos na Tabela B.1, indicam que a adição do bloco RLE teve impacto mínimo no PSNR para LLFlow + SKF e WaveNet, mas trouxe melhorias nos índices SSIM e LPIPS, sugerindo potencial ganho na qualidade perceptual. Para RetinexNet e UNet, a inclusão do RLE resultou em ligeira melhora na qualidade da imagem, evidenciada pelo aumento nos índices PSNR, SSIM e pela redução do LPIPS. Além disso, todos os modelos passaram a oferecer a capacidade de controle sobre os níveis de aprimoramento.

A Figura B.3 ilustra a versatilidade do RLE ao permitir aprimoramento de imagens controlável em quatro modelos de LLIE. Ajustando o valor estimado pelo RLE, é possível reduzir ou intensificar o aprimoramento das imagens de maneira intuitiva. A Figura B.4 destaca o controle em regiões específicas da imagem, mostrando a flexibilidade oferecida pelo bloco RLE.

Um desafio encontrado foi a reprodução dos resultados relatados por modelos

baseados em aprendizado profundo, devido a diferenças em hardware, ambientes de software ou à falta de detalhes precisos de implementação. A Tabela B.2 compara os resultados originais reportados pelos autores com aqueles obtidos ao retrainar os modelos usando os parâmetros e datasets descritos nos artigos. Em alguns casos, como o RetinexNet, houve melhorias significativas nos índices PSNR e SSIM após a retrainagem. Para LLFlow + SKF, os resultados retrainados foram similares aos originais, com leve aumento no LPIPS. Já para o WaveNet, os resultados foram inferiores em PSNR e SSIM. Não foi possível reproduzir os resultados do LLFlow sem o SKF, apesar de seguir as orientações fornecidas pelos autores.

Embora a comparação direta entre RLE e SKF seja limitada pelos desafios de reprodução, a Tabela B.3 apresenta os resultados disponíveis. Observa-se que o SKF tende a melhorar métricas de desempenho, como PSNR e SSIM, de maneira mais significativa. No entanto, o RLE oferece a vantagem única de controle sobre o nível de aprimoramento, permitindo maior flexibilidade sem prejudicar drasticamente o desempenho. No caso do LLFlow + SKF, os dois métodos combinados oferecem tanto a melhora nas métricas associada ao SKF e a flexibilidade de controle trazida pelo RLE.

	PSNR	**SSIM**	**LPIPS**
RetinexNet	18.17	0.6956	0.4041
RetinexNet w/ RLE	19.18	0.7242	0.3472
UNet	19.85	0.8195	0.1281
UNet w/ RLE	20.11	0.8025	0.1255
LLFlow + SKF	25.91	0.8657	0.1933
LLFlow + SKF w/ RLE	25.81	0.8665	0.1249
WaveNet	22.53	0.8419	0.0772
WaveNet w/ RLE	22.49	0.8628	0.0639

Table B.1 – Comparação de desempenho entre os modelos originais e as versões estendidas com o bloco RLE.

B.4 Conclusão

Este trabalho apresentou uma abordagem inovadora para o controle de aprimoramento de imagens em baixa luminosidade por meio do bloco RLE. Nosso método automaticamente gera imagens bem expostas ao mesmo tempo que permite controle do nível de aprimoramento, adicionando flexibilidade e controle do usuário a redes existentes de LLIE. O bloco RLE pode ser integrado a modelos convolucionais, que são amplamente

utilizados para esse resolver esse problema.

Avaliamos o impacto da adição do bloco RLE a alguns modelos LLIE existentes. Embora o impacto no desempenho tenha variado entre os modelos, demonstramos que a modificação interativa do valor previsto pelo bloco RLE ajusta diretamente o nível de iluminação da imagem aprimorada, permitindo um controle eficaz do aprimoramento.

Além do controle global, nossa abordagem oferece controle local, possibilitando que os usuários ajustem os níveis de aprimoramento de áreas específicas dentro de uma imagem. Utilizando máscaras geradas pelo Segment Anything Model, é possível adaptar o nível de aprimoramento de múltiplas regiões de forma independente.

De forma geral, o bloco RLE representa um passo importante para tornar os métodos de LLIE mais flexíveis e adaptáveis às preferências dos usuários. Trabalhos futuros buscam estender esta abordagem para incluir modelos baseados em difusão e transformers, além de investigar sua aplicação no controle de aprimoramento de vídeos em baixa luminosidade.

	PSNR	**SSIM**	**LPIPS**
RetinexNet (Original)	16.77	0.462	0.474
RetinexNet (Retrained)	18.17	0.724	0.347
LLFlow + SKF (Original)	25.94	0.865	0.125
LLFlow + SKF (Retrained)	25.91	0.866	0.193
WaveNet (Original)	24.54	0.856	-
WaveNet (Retrained)	22.53	0.8419	0.0772

Table B.2 – Desempenho dos modelos originais conforme relatado pelos autores e os resultados do nosso re-treinamento no mesmo conjunto de dados.

	PSNR	**SSIM**	**LPIPS**
RetinexNet + SKF	20.42	0.711	0.216
RetinexNet w/ RLE	19.18	0.724	0.347
LLFlow + SKF	25.94	0.865	0.125
LLFlow + SKF w/ RLE	25.81	0.867	0.125

Table B.3 – Performance comparison of the RetinexNet and LLFlow models adding the SKF and/or the RLE block. We were not able to reproduce the results of LLFlow without SKF to evaluate adding RLE to only LLFlow, and the authors did not provide the pre-trained models nor the code for the RetinexNet with the SKF framework.