IE643: Deep Learning - Theory and Practice **July-Nov 2024**

Project Report: Unsupervised reconstruction of high-quality images from low light images

Team Name: DeepCrafters *Team Member(s): 22b2505*

Abstract

This project aims to enhance low-light, low-resolution images into high-quality, well-lit images using unsupervised learning. A two-stage approach was implemented: Zero-Reference Deep Curve Estimation (Zero-DCE) [3] was used to improve lighting, while Enhanced Super-Resolution GAN (ESRGAN) [8] was applied to increase resolution. The model operates without labeled data and provides real-time results through a Streamlit interface, which displays quantitative metrics evaluating the improvement in lighting and resolution. Results demonstrate effective image enhancement, producing clear, high-resolution images from low-light inputs.
The codes are available at https://github.com/Afnnnan/IE643_CourseProject.

1 Introduction

In the realm of computer vision, enhancing low-light images poses a significant challenge due to the inherent limitations of traditional imaging systems, which often struggle to capture adequate details in suboptimal lighting conditions. The relevance of this task is underscored by its wide-ranging applications, from improving the aesthetic quality of photographs to enhancing critical functionalities such as object recognition in surveillance systems and medical imaging. The inability to properly illuminate subjects can lead to compromised visual quality, miscommunication of information, and hindered performance of automated systems reliant on image analysis.

Recent advancements in deep learning techniques have opened new avenues for addressing these challenges. Specifically, Single Image Super-Resolution (SISR) techniques, which focus on reconstructing high-resolution images from low-resolution inputs, have gained traction within the research community. The success of convolutional neural networks (CNNs), beginning with models like SRCNN [1], has transformed SISR by leveraging architectural innovations and training strategies that optimize performance metrics such as Peak Signal-to-Noise Ratio (PSNR). However, these metrics often do not align with human perception of visual quality, leading to over-smoothed results lacking in essential high-frequency details.

To combat this, perceptual-driven methods have emerged, utilizing techniques such as Generative Adversarial Networks (GANs) [2] to produce results that align more closely with natural image aesthetics. In particular, methods like ESRGAN have significantly improved visual quality, but a gap remains between generated outputs and true high-resolution images. Simultaneously, approaches like Zero-Reference Deep Curve Estimation (Zero-DCE) offer innovative solutions for low-light image enhancement without the need for paired training data, achieving competitive results even under diverse and challenging lighting conditions.

This project leverages these powerful deep learning techniques to propose a robust solution for transforming low-light images into high-quality outputs. Our approach integrates Zero-DCE for lighting enhancement and ESRGAN for super-resolution, enabling us to create visually appealing and detailed images without requiring extensive labeled datasets.

Figure 1: Visual comparison: the original low-light image, the generated image using the proposed approach, and the ground truth. Note that the enhancement is performed on a resized 300x200 image.

The structure of this report is organized as follows: We provide a survey of existing literature in Section 3, followed by a description of our project proposal in Section 4. Experimental details are outlined in Section 6, while future work is discussed in Section 8. Finally, we conclude with a summary and future directions in Section 9.

2 Project Workflow

The workflow of this project is structured to address the challenges associated with low-light image enhancement and high-resolution reconstruction. The following steps outline our comprehensive approach:

- Problem Statement: Identifying and articulating the challenges in low-light image enhancement and the need for high-resolution reconstruction techniques to improve visual quality.
- Literature Review: Conducted a thorough review of existing models and methodologies, including EnlightenGAN [4], SCI [7], Zero-DCE for low-light image enhancement, and SRGAN [6], ESRGAN for high-resolution reconstruction.
- Dataset Collection: Utilized the LoL dataset to provide a diverse set of low-light images for training and evaluation.
- Model Training: Trained the Zero-DCE model for effective low-light image enhancement. Leveraged a pretrained ESRGAN model to generate high-resolution outputs, focusing on enhancing texture and detail.
- Metric Development: Developed and refined evaluation metrics to quantitatively assess the performance of the low-light enhancement and super-resolution models.
- Streamlit Application: Designed and implemented an interactive application using Streamlit for model demonstration and real-time visualization of enhancement results.
- Exploration of Few-Shot Learning Approaches: Investigated few-shot learning techniques, including Dimma [5], to enhance model adaptability and performance with limited training data.

3 Literature Survey

This section reviews the significant works and methodologies relevant to low-light image enhancement and super-resolution, which have informed the development of our approach.

3.1 Low-Light Image Enhancement

Several methods have been proposed for enhancing low-light images. Zero-DCE stands out for its ability to enhance images without relying on paired data. It utilizes a data-driven approach to estimate illumination without introducing unrealistic artifacts, making it suitable for our project. Although EnlightenGAN offers promising results, it was not utilized due to its reliance on partially supervised methods, which do not align with our goal of fully unpaired training. Additionally, the SCI model was explored for fine-tuning with low-resolution images, but the outputs were inferior to those achieved with Zero-DCE.

3.2 Super-Resolution Techniques

In the realm of super-resolution, several models, such as SRGAN and ESRGAN, have demonstrated effective performance. SRGAN uses adversarial training to improve the perceptual quality of images, while ESRGAN enhances the residual learning capabilities, resulting in superior high-resolution outputs. For our project, we leveraged pretrained ESRGAN to transform the low-light enhanced images produced by Zero-DCE into high-resolution formats, thereby optimizing the overall enhancement process.

3.3 Evaluation Metrics

To assess the effectiveness of our approach, we employed several metrics that measure both low-light enhancement and super-resolution quality. Specifically:

- **Structural Similarity Index (SSIM):** Used to evaluate the similarity between the enhanced images and high-quality reference images, SSIM captures changes in structural information, luminance, and contrast, making it suitable for assessing both resolution and lighting improvements.

- **Peak Signal-to-Noise Ratio (PSNR):** PSNR provides a measure of fidelity based on pixel differences, useful for quantifying the resolution quality of the enhanced images.

- **Mean Absolute Error (MAE):** MAE was utilized to capture the overall deviation of pixel values, providing insight into the accuracy of lighting and contrast adjustments in low-light enhancement.

In addition to these, specific attributes such as colorfulness and sharpness were assessed. Colorfulness serves as an indicator of lighting effectiveness, ensuring that the enhanced images maintain natural color balance, while sharpness reflects the resolution quality and clarity, essential for both high-resolution outputs and improved low-light performance.

4 Proposed Approach

We propose a dual-step approach that first enhances low-light images using Zero-DCE and then applies ESRGAN for super-resolution, ensuring high-quality inputs for the latter process. This method differs from existing work by effectively combining enhancement and resolution recovery to restore details lost in low-light, low-resolution images.

4.1 Work done before prep-presentation review

In this section, we describe the initial methods tried for low-light image enhancement, specifically EnlightenGAN and SCI.

4.1.1 EnlightenGAN

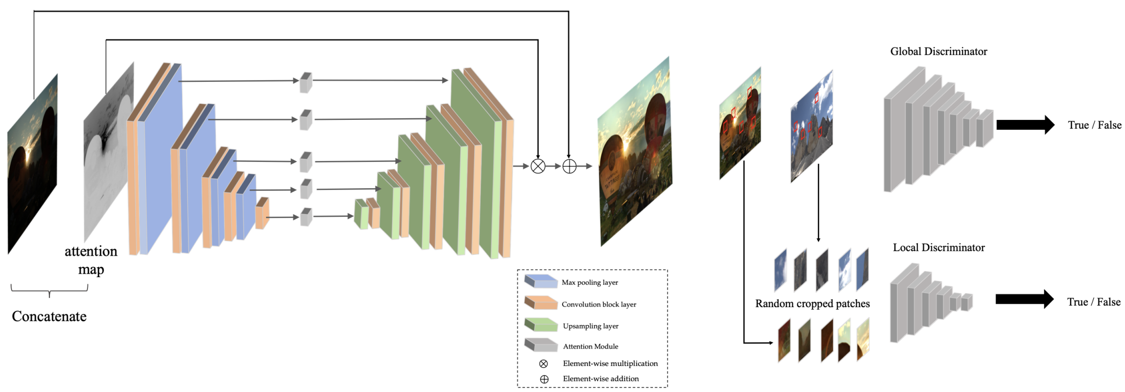

Figure 2: The overall architecture of EnlightenGAN. In the generator, each convolutional block consists of two 3 ×3 convolutional layers followed by batch normalization and LeakyRelu. Each attention module has the feature map multiply with a (resized) attention map.

The approach for EnlightenGAN employs an attention-guided U-Net as the generator and a dual-discriminator structure. The architecture includes:

- **Attention-Guided U-Net Generator:** Utilizes self-regularized attention to focus on darker regions for enhancement.
- **Global-Local Discriminators:** A dual-discriminator structure that includes a global discriminator for the entire image and a local discriminator focusing on random patches, ensuring spatially consistent enhancements.
- **Self-Feature Preserving Loss:** A loss based on VGG feature space to maintain structural consistency between input and enhanced images.

4.1.2 Self-Calibrated Illumination (SCI)

The SCI model is built on lightweight convolutional networks, featuring:

- **Cascaded Structure:** A series of cascading enhancement modules to iteratively improve image brightness and contrast.
- **Self-Calibrated Convolutions:** Utilizes self-calibrated convolution layers for feature refinement across multiple scales, enhancing low-light regions effectively.

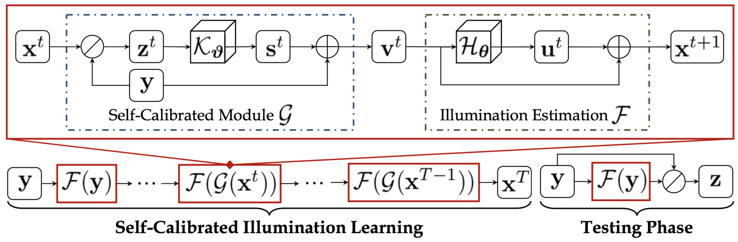

Figure 3: The entire framework of SCI. In the training phase, the SCI consists of the illumination estimation and self-calibrated module. The output of the self-calibrated module is added to the original low-light input to serve as the input for the illumination estimation at the next stage. It is important to note that these two modules share parameters throughout the entire training procedure. In the testing phase, only a single illumination estimation module is utilized.

4.2 Work done after prep-presentation review

4.2.1 Zero-DCE Framework

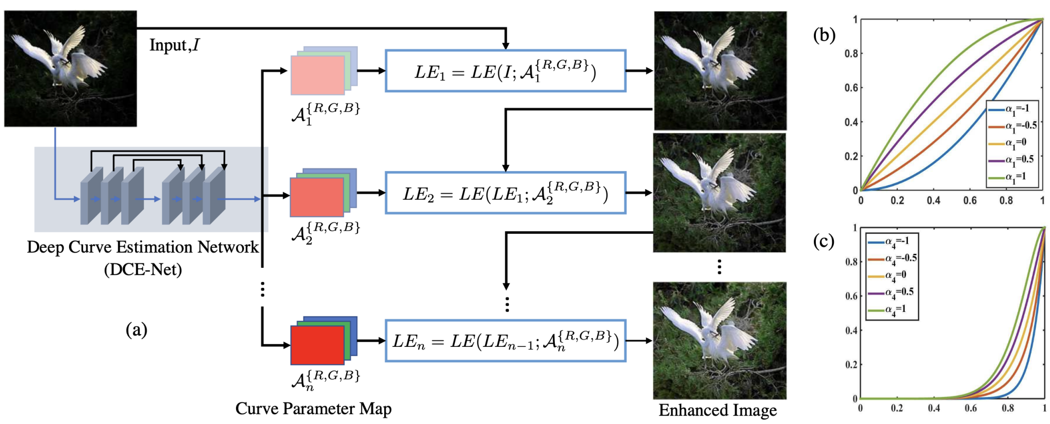

Figure 4: (a) The framework of Zero-DCE. A DCE-Net is devised to estimate a set of best-fitting Light-Enhancement curves (LE-curves) that iteratively enhance a given input image. (b, c) LE-curves with different adjustment parameters α and numbers of iterations n. In (c), α_1, α_2, and α_3 are equal to -1 while n is equal to 4. In each subfigure, the horizontal axis represents the input pixel values while the vertical axis represents the output pixel values.

The framework of Zero-DCE is presented in Fig. 4. A Deep Curve Estimation Network (DCE-Net) is devised to estimate a set of best-fitting Light-Enhancement curves (LE-curves) given an input image. The framework then maps all pixels of the input's RGB channels by applying the curves iteratively to obtain the final enhanced image. The key components of Zero-DCE include the LE-curve, DCE-Net, and non-reference loss functions, which are detailed in the following sections.

4.2.2 Light-Enhancement Curve (LE-curve)

Inspired by the curve adjustments used in photo editing software, a curve is designed to map a low-light image to its enhanced version automatically, where the self-adaptive curve parameters are solely dependent on the input image. The design of the curve has three objectives: 1) Each pixel value of the enhanced image should be within the normalized range of [0,1] to avoid information loss due to overflow truncation; 2) The curve should be monotonous to preserve the differences (contrast) among neighboring pixels; 3) The curve's form should be as simple as possible and differentiable for gradient backpropagation.

To achieve these objectives, a quadratic curve is designed, expressed as:

$$LE(I(x); \alpha) = I(x) + \alpha I(x)(1 - I(x)), \tag{1}$$

where x denotes pixel coordinates, $LE(I(x); \alpha)$ is the enhanced version of the input $I(x)$, and $\alpha \in [-1, 1]$ is the trainable curve parameter that adjusts the magnitude of the LE-curve and controls the exposure level. Each pixel is normalized to [0, 1], and all operations are pixel-wise. The LE-curve is applied separately to the three RGB channels rather than solely on the illumination channel to better preserve inherent color and reduce the risk of over-saturation.

An illustration of LE-curves with different adjustment parameters α is shown in Fig 4(b). The LE-curve complies with the three aforementioned objectives and enables the increase or decrease of the dynamic range of an input image, enhancing low-light regions and removing over-exposure artifacts.

4.2.3 Higher-Order Curve

The LE-curve defined in Eq. (1) can be applied iteratively to enable versatile adjustment for challenging low-light conditions. Specifically,

$$LE_n(x) = LE_{n-1}(x) + \alpha_n LE_{n-1}(x)(1 - LE_{n-1}(x)), \tag{2}$$

where n is the iteration number that controls the curvature. In this work, n is set to 8 to effectively deal with most cases.

4.2.4 Pixel-Wise Curve

A higher-order curve can adjust an image within a wider dynamic range, but since α is global, local regions might be over-/under-enhanced. To address this, α is formulated as a pixel-wise parameter, with each pixel of the input image having a corresponding curve with the best-fitting α. Thus, Eq. (2) can be reformulated as:

$$LE_n(x) = LE_{n-1}(x) + A_n(x) LE_{n-1}(x)(1 - LE_{n-1}(x)), \tag{3}$$

where A is a parameter map of the same size as the given image, preserving monotonous relations among neighboring pixels.

4.2.5 DCE-Net

To learn the mapping between an input image and its best-fitting curve parameter maps, a Deep Curve Estimation Network (DCE-Net) is proposed. The input to the DCE-Net is a low-light image, and the outputs are pixel-wise curve parameter maps for corresponding higher-order curves. The architecture consists of seven convolutional layers with symmetrical concatenation, where each layer contains 32 convolutional kernels of size 3×3 and stride 1, followed by the ReLU activation function. The last convolutional layer employs the Tanh activation function, producing 24 parameter maps for 8 iterations.

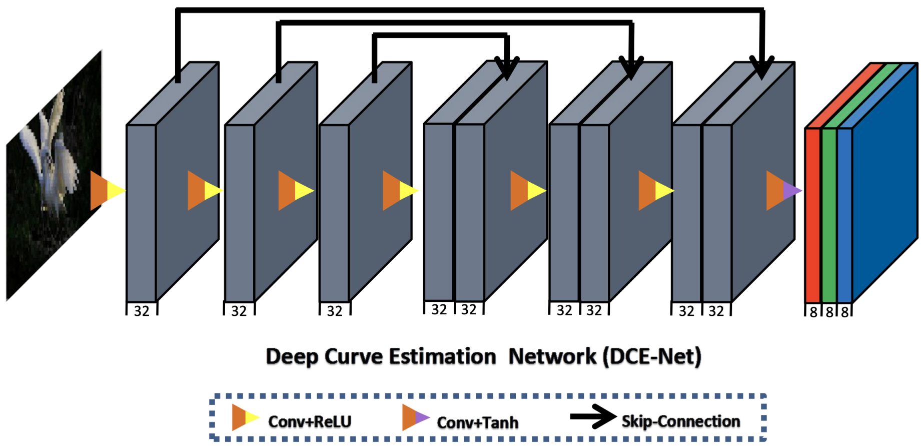

Figure 5: The architecture of Deep Curve Estimation Network (DCE-Net).

4.2.6 Non-Reference Loss Functions

To enable zero-reference learning in DCE-Net, a set of differentiable non-reference losses are proposed, allowing evaluation of the quality of enhanced images. The losses include: - Spatial Consistency Loss - Exposure Control Loss - Color Constancy Loss - Illumination Smoothness Loss

The total loss can be expressed as:

$$L_{\text{total}} = L_{\text{spa}} + L_{\text{exp}} + W_{\text{col}}L_{\text{col}} + W_{\text{tvA}}L_{\text{tvA}}. \tag{4}$$

4.2.7 Enhanced Super-Resolution Generative Adversarial Network (ESRGAN)

In the continuation of our work on low-light image enhancement using Zero-DCE, we explored the application of Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) to further improve the visual quality of images.

4.2.8 Overview of ESRGAN

ESRGAN is an advanced variant of GAN-based super-resolution techniques, designed to generate high-resolution images with finer details and textures. Unlike traditional methods, ESRGAN focuses on perceptual quality and aims to minimize the perceptual difference between the generated images and the high-resolution ground truth. This is achieved through the use of a Relativistic Discriminator that evaluates how realistic a generated image appears compared to real images. Notably, ESRGAN increases the image resolution by four times, enhancing the visual fidelity and detail in the output images.

4.2.9 Network Architecture

The architecture of ESRGAN consists of the following key components:

- **Generator:** The generator network employs Residual-in-Residual Dense Blocks (RRDB) that promote effective information flow and feature extraction, enabling the network to capture intricate de-

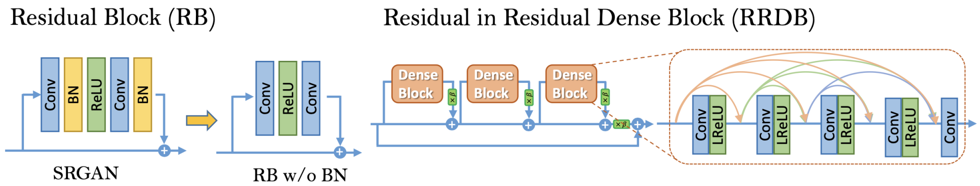

Figure 6: Comparison of Network Architectures for Image Super-Resolution. Left: The architecture of the original SRGAN, where Batch Normalization (BN) layers are removed from the residual blocks. Right: The Residual-in-Residual Dense Block (RRDB) utilized in ESRGAN model, where β represents the residual scaling parameter to optimize feature propagation and improve image quality.

tails.

- **Discriminator:** The Relativistic Discriminator compares real and generated images in a more dynamic way, enhancing the adversarial training process and leading to higher-quality outputs.
- **Perceptual Loss:** ESRGAN utilizes a perceptual loss function that combines content loss and adversarial loss, ensuring the generated images are not only high in pixel fidelity but also visually appealing.

4.2.10 Interface Development

The interface for our image enhancement project was implemented using Streamlit. The application allows users to upload low-light images and optional ground truth images. Upon file upload, the low-light image is processed using the Zero-DCE model, which is defined within the application.

Following the enhancement, the output is fed into the ESRGAN model for super-resolution. The interface displays the original, enhanced, and high-resolution images sequentially. Additionally, it provides functionality to compute and display relevant image quality metrics upon user request.

The code structure includes functions for loading models, preprocessing images, performing inference, and calculating metrics, ensuring an efficient workflow within the Streamlit framework.

5 Dataset Details

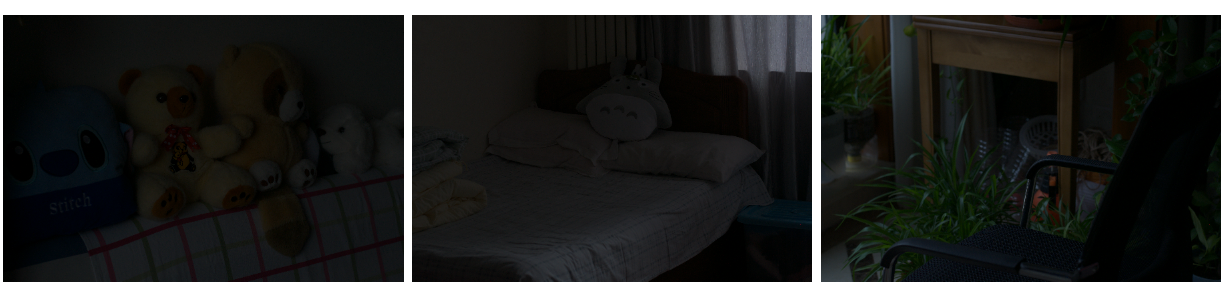

Figure 7: Images from LoL dataset

In this project, we utilized the LoL (Low-Light) dataset, which comprises 480 low-light images for training. Each image is resized to 256 × 256 pixels to ensure uniformity in input dimensions. The dataset includes essential images for tasks involving enhancement and super-resolution. Preprocessing techniques applied include loading the images, decoding them into RGB format, resizing, and normalizing pixel values to the range [0, 1].

The training set consists of 400 images, with additional images reserved for validation and testing. The training dataset is used to train the model, while the validation set monitors performance during training, and the test set is reserved for final evaluation.

6 Experiments

In this section, we describe the experiments conducted, including both positive and negative results, to illustrate the iterative process of model selection and optimization for low-light image enhancement. Details of the training procedure, optimization settings, and hardware configuration are provided.

6.1 Model Selection and Initial Experiments

Figure 8: Progression of test image enhancement: the original low-light test image, the output after enhancement with EnlightenGAN, and the final result after applying super-resolution with ESRGAN.

Figure 9: Test image progression: Original low-light test image, followed by super-resolution using ESRGAN, and final enhancement with EnlightenGAN for lighting correction.

Initially, the EnlightenGAN model was selected for low-light enhancement. An input image, resized to 200x300, was fed into EnlightenGAN, which outputs a preliminary enhanced image. This output was then processed by the ESRGAN model for super-resolution. The resulting images, as shown in Figure 8, contained noticeable artifacts, suggesting a need for fine-tuning to better handle the low-resolution inputs. However, EnlightenGAN is designed for semi supervised enhancement, requiring unpaired low-light and normal-light images. Additionally, EnlightenGAN's older library dependencies posed a secondary challenge.

An alternate approach was also attempted, where resolution was first increased using ESRGAN, followed by low-light enhancement through EnlightenGAN. This method, however, yielded a dimmer result, as shown in Figure 9, indicating that resolution enhancement prior to lighting correction was less effective for our dataset.

6.2 Switch to SCI Model

Figure 10: Test image progression: Enhanced image using SCI, super-resolved image after ESRGAN, and failed fine-tuning attempt showing a vague output.

Given the limitations of EnlightenGAN, we experimented with the Self-Calibrated Illumination (SCI) model. After initial inference, results from SCI demonstrated inferior low-light enhancement compared to EnlightenGAN; however, artifacts and vague output persisted even after fine-tuning, as depicted in Figure 10. This indicated that SCI might not effectively generalize to our specific low-light dataset or benefit sufficiently from the fine-tuning performed.

6.3 Final Model Choice: Zero-DCE

Figure 11: Progression of test image enhancement: the original low-light test image, the output after enhancement with Zero-DCE, and the final result after applying super-resolution with ESRGAN.

Following the experiments with EnlightenGAN and SCI, the Zero-DCE model was selected for its ability to enhance low-light images without requiring paired normal-light images, as it uses a fully unsupervised approach tailored for low-light conditions. The model was trained on the resized 256×256 images, providing visually enhanced results with fewer artifacts, as shown in Figure 11.

6.4 Training and Optimization Settings

The chosen models were trained using the Adam optimizer with a learning rate of 0.0001 and a batch size of 16. Regularization techniques such as weight decay were applied to improve generalization. The ESR-GAN model was initialized using pre-trained weights, originally trained on the Bicubically Downsampled DIV2K Dataset, ensuring a solid foundation for enhancing low-light images. The training process took only 8 minutes on a Colab T4 GPU, showcasing the efficiency of the model architecture and implementation.

The dataset utilized for training consists of low-light images from the LoL dataset, which were processed and organized for efficient training. The training and validation datasets were constructed from images within the LoL dataset, ensuring a diverse range of examples.

The training was conducted in TensorFlow, leveraging its robust capabilities for deep learning.

To evaluate model performance, various loss functions were implemented:

- **Spatial Consistency Loss:** The spatial consistency loss L_{spa} encourages spatial coherence of the enhanced image through preserving the difference of neighboring regions between the input image and its enhanced version:

$$L_{\text{spa}} = \frac{1}{K}\sum_{i=1}^{K}\sum_{j\in\Omega(i)}\left(|(Y_i - Y_j)| - |(I_i - I_j)|\right)^2$$

  where K is the number of local regions, and $\Omega(i)$ is the set of neighboring regions (top, down, left, right) centered at region i. We denote Y and I as the average intensity value of the local region in the enhanced version and input image, respectively. We empirically set the size of the local region to 4×4.

- **Exposure Control Loss:** To restrain under-/over-exposed regions, we use an exposure control loss L_{exp} to control the exposure level. The exposure control loss measures the distance between the average intensity value of a local region to the well-exposedness level E:

$$L_{\text{exp}} = \frac{1}{M}\sum_{k=1}^{M}|Y_k - E|$$

  where M represents the number of non-overlapping local regions of size 16×16, and Y is the average intensity value of a local region in the enhanced image. We set E to 0.6 based on existing practices.

- **Color Constancy Loss:** Following the Gray-World color constancy hypothesis, we use a color constancy loss L_{col} to correct the potential color deviations in the enhanced image:

$$L_{\text{col}} = \sum_{\forall(p,q)\in\epsilon}(J_p - J_q)^2, \quad \epsilon = \{(R,G),(R,B),(G,B)\}$$

  where J_p denotes the average intensity value of p channel in the enhanced image, (p,q) represents a pair of channels.

- **Illumination Smoothness Loss:** To preserve the monotonicity relations between neighboring pixels, we add an illumination smoothness loss L_{tvA}:

$$L_{\text{tvA}} = \frac{1}{N}\sum_{n=1}^{N}\sum_{c\in\xi}\left(|\nabla_x A_n^c| + |\nabla_y A_n^c|\right)^2, \quad \xi = \{R,G,B\}$$

where N is the number of iterations, and ∇_x and ∇_y represent the horizontal and vertical gradient operations, respectively.

- **Total Loss:** The total loss can be expressed as:

$$L_{\text{total}} = L_{\text{spa}} + L_{\text{exp}} + W_{\text{col}} L_{\text{col}} + W_{\text{tvA}} L_{\text{tvA}}$$

where W_{col} and W_{tvA} are the weights of the losses.

These loss functions were combined into a single objective function for optimization, allowing the model to effectively learn from low-light images while considering multiple aspects of image quality.

Throughout the training process, the model's performance was monitored using validation datasets, with adjustments made to the learning rate and other hyperparameters to ensure optimal convergence. The training utilized efficient data loading mechanisms and batch processing to maximize resource utilization.

In conclusion, the training regimen was structured to leverage advanced optimization techniques and comprehensive loss functions, ensuring robust model performance in enhancing low-light images.

6.5 Hardware Configuration

All experiments for the SCI and EnlightenGAN + ESRGAN models were conducted on a 2021 MacBook Pro equipped with an Apple M1 Pro chip, featuring an 8-core CPU with 6 performance cores and 2 efficiency cores, as well as a 14-core GPU. These models required older dependencies.

For the Zero DCE + ESRGAN models, experiments were conducted using Google Colab with a T4 GPU, utilizing the latest (2024) dependencies available in TensorFlow.

6.6 Exploring Few-Shot Learning Approach: Dimma

Dimma is a semi-supervised learning method that enhances low-light images by utilizing a small set of image pairs to replicate scenes captured under challenging lighting conditions specific to a camera. This approach incorporates a convolutional mixture density network to model color distortions based on illumination differences and a conditional UNet architecture that allows users to adjust brightness levels during the enhancement process.

As illustrated in Figure 12, the first phase involves training the dimming module to mimic the color distortions created by a specific camera, followed by training the UNet in an unsupervised manner to restore the original image. The effectiveness of Dimma is further demonstrated in Figure 13, which presents a visual comparison of Dimma itself trained on 3 pairs, 5 pairs, 8 pairs, and the full LOL dataset, highlighting how the model performs with varying amounts of training data.

7 Results

To quantify the performance of the various methods, we employed Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) as metrics. Both metrics provide insights into the quality of light enhancement and super-resolution achieved by the models. Higher PSNR values indicate better reconstruction quality, while higher SSIM values reflect better structural similarity to the ground truth images.

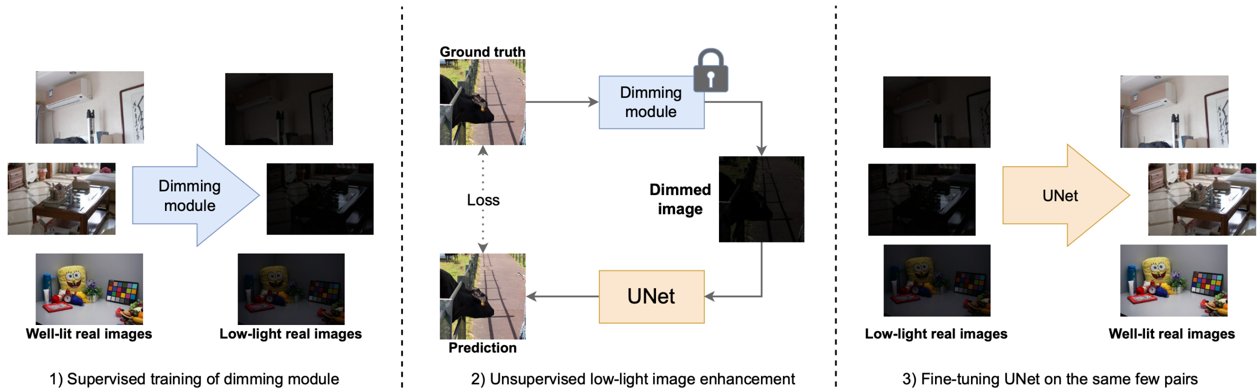

Figure 12: High level picture of Dimma approach

Figure 13: Comparison of Dimma trained on 3, 5, and 8 pairs, as well as the full LOL dataset, against both the ground truth and the image produced by our proposed approach.

Figure 1 shows the input low-light image, the generated image produced by our proposed approach, and the corresponding ground truth image. The visual comparison illustrates the significant improvement in quality achieved through our method, effectively restoring details and colors that are often lost in low-light conditions.

Table 1: Comparison of PSNR and SSIM values for different low-light image enhancement methods.

Method	PSNR (dB)	SSIM
EnlightenGAN + ESRGAN	19.08	0.5056
SCI + ESRGAN	15.64	0.2768
Proposed Approach	**19.75**	**0.5223**

8 Plan for Novelty Assessment

Throughout the project, noise emerged as a significant challenge in the generated images, adversely affecting their overall quality and clarity. To address this issue, it is essential to integrate effective noise removal models that can enhance the denoising process for low-light images.

One promising approach is to utilize convolutional neural networks specifically designed for denoising tasks. These models can be trained on datasets containing pairs of noisy and clean images, enabling them to learn the underlying patterns of noise and effectively remove it while preserving important image details. For instance, a denoising autoencoder can be employed, where the model learns to reconstruct clean images from their noisy counterparts. This method leverages the ability of deep learning to capture complex relationships in the data, making it well-suited for challenging low-light conditions.

Additionally, incorporating traditional denoising techniques, such as the Non-Local Means filter or bilateral filtering, can further improve the results. These methods work by analyzing the spatial relationships among pixels, allowing for effective noise reduction without compromising edge integrity.

By implementing these noise removal strategies, we aim to significantly enhance the quality of the generated images, leading to clearer and more visually appealing outputs. Future experiments will focus on assessing the performance of these models in conjunction with our existing enhancement approaches, with particular attention to the impact on PSNR and SSIM metrics.

9 Conclusion

This project addressed the critical challenge of enhancing low-light images, which often suffer from poor visibility and noise due to inadequate lighting conditions. The inherent difficulties include camera processing variations and a limited number of paired datasets with ground-truth lighting information, complicating the application of supervised learning techniques.

To tackle these issues, we employed a variety of methods, including Zero-DCE and models such as EnlightenGAN and SCI, enhanced by ESRGAN for super-resolution capabilities. Our experiments demonstrated that these approaches effectively improve the quality of low-light images while maintaining natural colors. Notably, our proposed method outperformed existing unsupervised techniques, achieving superior Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) values, indicative of enhanced image quality and structural fidelity to the original images.

The significance of these results lies in their potential applications across various fields, including photography, surveillance, and computer vision, where low-light conditions are prevalent. By successfully enhancing low-light images, we contribute to improved visual interpretation and analysis.

Looking forward, future work will focus on integrating noise removal models to mitigate the noise that persists in the generated images, further refining the quality of our enhancements. Additionally, exploring few-shot learning techniques could optimize our model's performance while requiring fewer training data, ultimately enhancing the robustness and versatility of low-light image enhancement solutions.

References

[1] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. 2014. Available at `https://doi.org/10.48550/arXiv.1501.00092`.

[2] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. 2014. Available at `https://doi.org/10.48550/arXiv.1406.2661`.

[3] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. 2020. Available at `https://doi.org/10.48550/arXiv.2001.06826`.

[4] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. 2019. Available at https://doi.org/10.48550/arXiv.1906.06972.

[5] Wojciech Kozłowski, Michał Szachniewicz, Michał Stypułkowski, and Maciej Zieba. Dimma: Semi-supervised low light image enhancement with adaptive dimming. 2023. Available at https://doi.org/10.48550/arXiv.2310.09633.

[6] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. 2016. Available at https://doi.org/10.48550/arXiv.1609.04802.

[7] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. 2022. Available at https://doi.org/10.48550/arXiv.2204.10137.

[8] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, Yu Qiao, and Xiaoou Tang. Esrgan: Enhanced super-resolution generative adversarial networks. 2018. Available at https://doi.org/10.48550/arXiv.1809.00219.