REPORT FOR LOW-LIGHT IMAGE DENOISING

Soumya Gupta

Enrollment no. 23112101

1. INTRODUCTION

Low light Image enhancement (LLIE), a pivotal yet challenging task in computer vision, aims to improve visibility and contrast across a broad spectrum of low-light scenarios, including uneven illumination, extreme darkness, backlighting, and night. Additionally, LLIE strives to correct imperfections like noise, artifacts, and color distortion. These challenges, arising in darkness or through illumination enhancement, affect both human visual perception and downstream tasks like object detection and scene segmentation

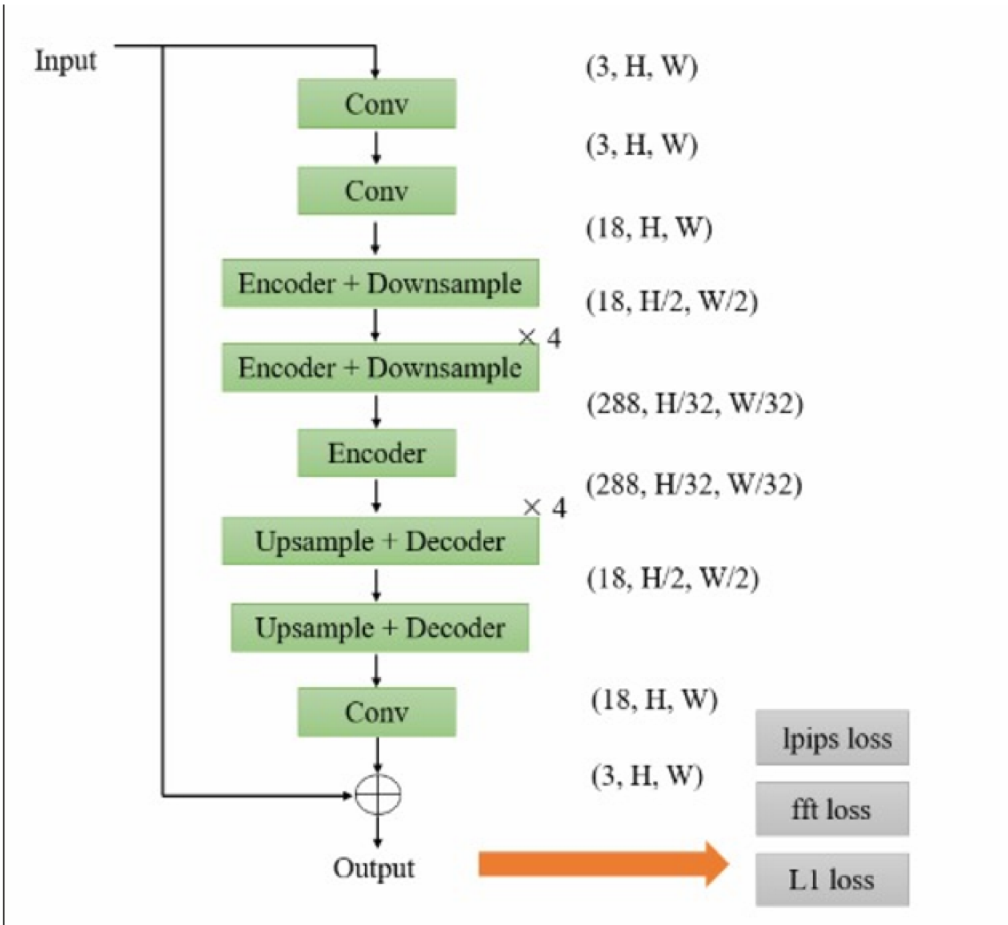

Figure 5. The network architecture of team LVGroup_HFUT.

2. Architecture Overview

For this project, I tried to implement model named “Image Lab” as proposed in the NTIRE Challenge 2024 Paper.(https://arxiv.org/pdf/2404.14248)

a. Initial Convolution Block:

The model starts with an initial convolution block to process the input image. This block uses a double convolutional layer to extract basic features

b. Encoder + Downsample x4:

The image then goes through four consecutive encoder blocks, each consisting of a double convolution block followed by downsampling. The number of channels increases progressively while the spatial dimensions (height and width) are reduced by a factor of 2 at each step. This allows the model to capture features at multiple scales, with the deeper layers focusing on more abstract representations.

c. Bottleneck:

A single bottleneck block without downsampling is applied to further process the features at the reduced dimensions.

d. Upsample + Decoder x4:

Next, the image undergoes four upsampling and decoder blocks. Each block increases the spatial dimensions by a factor of 2 and reduces the channel count while incorporating skip connections from the corresponding encoder layers to restore the original image resolution while refining the learned features.

e. Final Convolution Block:

The final convolutional layer refines the output of the decoder path to produce the enhanced low-light image. It consists of a single convolutional layer:

3. Training Process:

- **Optimizer:** Adam optimizer was used for training the model. The learning rate started at 0.001, indicating that the model made relatively larger updates in the beginning. Over the course of 100 epochs, the learning rate gradually decreased to 0.00001.
-
- **Learning Rate Scheduler:** A CosineAnnealingLR scheduler was applied with T_max set to 10 and eta_min set to 0.000001, allowing smooth decay of the learning rate.

a. Training Objective:

The objective of training the model is to minimize a loss function that measures how well the model's predictions match the target outputs. The loss function used in this training was a combination of three different types of loss:

L1 Loss: This measures the absolute differences between the predicted and target images. It helps ensure that the predicted images are as close as possible to the actual high-quality images.

1. **SSIM Loss:** The SSIM (Structural Similarity Index) loss assesses the visual similarity between two images by considering changes in structural information, such as

luminance, contrast, and texture. This loss encourages the model to preserve structural information in the enhanced images.

2. **FFT Loss:** This measures the difference between the predicted and target images in the frequency domain, encouraging the model to preserve high-frequency details in the enhanced images.

   The model was trained using Kaggle's GPU P100's environment with over 10 epochs, having a loss function L defined by:

$$L_{combined} = 0.5 \cdot (1 - \mathrm{SSIMLoss}(pred, target)) + 0.1 \cdot \mathrm{L1Loss}(pred, target) + 0.4 \cdot \mathrm{FFTLoss}(pred, target)$$

4. Summary:

The architecture leverages multi-scale feature extraction and refinement, along with effective noise suppression techniques, to enhance the quality of low-light images. The combination of convolutional layers, encoder-decoder structure with skip connections, and sophisticated loss functions allows the model to produce visually appealing and high-quality enhanced images from low-light inputs.

The output Enhanced Image got an average PSNR value of 18 dB on Training Set

PSNR: 23.00 dB

PSNR: 19.17 dB

PSNR: 21.26 dB
Average PSNR: 18.08 dB