School of
Electronics and Communication Engineering

Mini Project Report

on

Enhancement of Videos Captured in Low-light Conditions

By:

1. **Swaroop U A** 01FE20BEC293
2. **Ankit Raichur** 01FE20BCS317
3. **Vinod Patil** 01FE20BEC143
4. **Allabakash G** 01FE20BEC261

Semester: V, 2022-2023

Under the Guidance of

Dr. Uma Mudenagudi
Dr. Ujwala Patil

K.L.E SOCIETY'S
KLE Technological University,
HUBBALLI-580031

2022-2023

SCHOOL OF ELECTRONICS AND COMMUNICATION ENGINEERING

CERTIFICATE

This is to certify that project entitled **" Enhancement of Videos Captured in Low-Light Conditions "** is a bonafide work carried out by the student team of **"Swaroop U A (01FE20BEC293), Ankit Raichur (01FE20BCS317), Vinod Patil (01FE20BEC143), Allabakash G (01FE20BEC261)"**. The project report has been approved as it satisfies the requirements with respect to the mini project work prescribed by the university curriculum for BE (V Semester) in School of Electronics and Communications Engineering of KLE Technological University for the academic year 2022-2023.

Dr. Ujwala Patil **Dr. Uma Mudenagudi** **Guide**	**Dr. Nalini C Iyer** **Head of School**	**Dr. B.S.Anami** **Registrar**

External Viva:

Name of Examiners **Signature with date**

1.
2.

ACKNOWLEDGMENT

The sense of contentment and elation that accompanies the successful completion of our project and its report would be incomplete without mentioning the names of the people who helped us in accomplishing this. We are indebted to our guide Dr. Ujwala Patil and Dr. Uma Mudenagudi who were nothing but a constant source of enthusiasm and whose profound guidance, valuable suggestions, and beneficent direction was mainly responsible for us to complete this project.
We take this opportunity to express our deep sense of gratitude and sincere thanks to our Head of the department Dr. Nalini C Iyer for her tremendous source of inspiration and help in challenging our effort in the right direction. Last but not least we like to thank all the course faculty, teaching and non- teaching staff for helping us during the project.

By:

Project Team

ABSTRACT

Videos captured in low-light conditions frequently have poor brightness, low contrast, a limited range of grayscale, color distortion, jitter, and substantial noise, which negatively affects the subjective visual experience of human eyes and severely restricts the effectiveness of high-level machine vision techniques. there are several architectures proposed in a low light video enhancement, Zero-DCE (DCE-Net) is one of the existing architectures, given an input image, the Deep Curve Estimation Network (DCE-Net) is designed to estimate a set of Light-Enhancement curves (LE-curves) that fit the input image the best. The framework then applies the curves iteratively to map each RGB channel's pixels to create the final improved image. It continuously improves and uses the image with the highest exposure as its ground truth until successful results are reached. Given that Zero-DCE is an unsupervised technique, The noise in the enhanced image prevents it from extracting local and global features. Therefore, we propose an architecture namely Low Light Video Enhancement(LLVE-Net), in which we are using perceptual loss addition to the loss functions used in Zero-DCE. The enhanced image in DCE-Net does not extract local and global features accurately and the enhanced image consists of high-level noise, to extract local and global features accurately we are using a cascading technique where the enhanced image is cascaded back to LLVE-Net so that it extracts those features that the model has not previously learned. So by cascading we can extract nearly all the local and global features and make our model more robust, by doing so our model consistently outperforms Zero-DCE.

Contents

List of Figures

Chapter 1

Introduction

Photos that are taken in less than optimal lighting conditions due to unavoidable environments, often have technical limitations. From 1.1 we can see that these include insufficient and unbalanced lighting in the environment, incorrect object placement against extreme back-light, and under-exposure during image capture. Such low-light photos suffer from poor aesthetic quality and poor information transmission. Both human vision, which favours bright pictures, and the many sophisticated system that rely on computer vision algorithms—like all-day autonomous driving and bio-metric recognition are challenged by these problems. Many techniques, from histogram- or cognition-based ones to learning-based approaches, have been proposed to minimize the deterioration. In this paper, we describe an unique deep learning-based method for enhancing low-light image quality. It can work in a variety of lighting situations, including uneven and bad lighting situations. Our deep learning-based approach has the distinct benefit of zero-reference, which means that it doesn't require any paired or even unpaired data during the training phase. The contributions are summarized as follows. In this paper we implemented perceptual loss in addition to the losses used in zero- DCE. We proposed a cascade approach to our model for it to learn more characteristics both locally and globally. The enhanced image is been predicted more accurately.

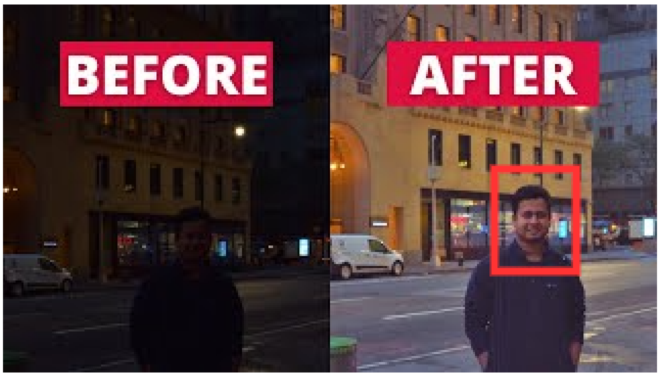

Figure 1.1: Human face detection in low-light

1.1 Motivation

Figure 1.2: Autonomous vehicle's night driving scenario

- Figure 1.2 illustrates how difficult it is for autonomous cars to recognise lanes on the road. This is where our algorithm excels and can enhance the performance of such systems.
- The low dynamic range and high noise levels of videos will impact how well computer vision algorithms work
- To make computer vision algorithms reliable in low-light situations, low-light video enhancement will require time to increase a video's visibility

1.2 Objectives

- To perform extensive literature survey on available SOTA methods for low-light video enhancement
- To develop a learning-based algorithm for enhancement of video captured in low light conditions
- To demonstrate the results of the proposed architecture using benchmark dataset in comparison with SOTA methods

1.3 Literature Survey Towards Low-light Image enhancement

1. A CNN for Lowlight Image Enhancement.[6]

They used a simple convolution Network with a residual Network.

Inferences :

- LLCNN learns to filter low-light pictures using various kernels before integrating multiscale feature maps in order to make improved shots that look to have been taken in well-lit circumstances while preserving the original features and textures

- From 1.3 we can see that one convolutional layer is utilised for pre-processing to provide uniform input, while another is used to generate an improved picture. Many carefully designed convolutional modules are sandwiched between those two layers. They use 64 filters in total, with the exception of the last filter. The number of filters used in the final layer is determined by the number of colour channels employed

- LLCNN, the suggested method, learns to increase picture brightness and contrast adaptively. A distinct module in LLCNN is designed to aid in training and performance enhancement. We also found that SSIM loss performs better for low-light picture enhancement applications

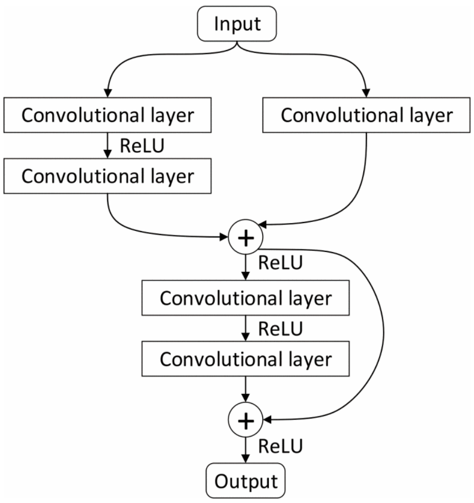

Figure 1.3: Architecture diagram of CNN

2. Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. [2]
Unsupervised learning technique using unpaired data.

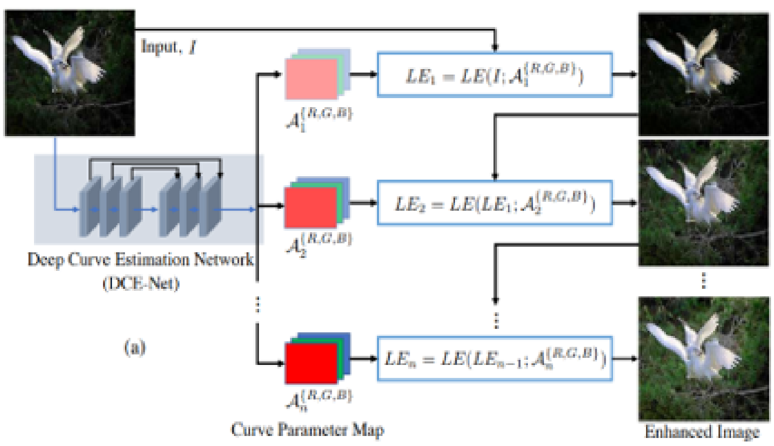

Figure 1.4: Architecture diagram of zero-DCE++

Inferences :

- The first low-light enhancement Network decreases the danger of overfitting due to its independence from paired and unpaired training data as shown in figure 1.4. As a result, their technique generalises well to a wide range of illumination circumstances.

- Their model creates an image-specific curve that can approximate both pixel-level and higher-order curves through iterative application. Such an image-specific curve is capable of executing mapping across a large dynamic range.

- To show how deep image enhancement models may be trained without reference pictures by employing task-specific non-reference loss functions that indirectly quantify the enhancement's quality.

3. From Fidelity to Perceptual Quality: A Semi-Supervised Approach for Low-Light Image Enhancement. [9]

A semi-supervised learning approach using DRBN.

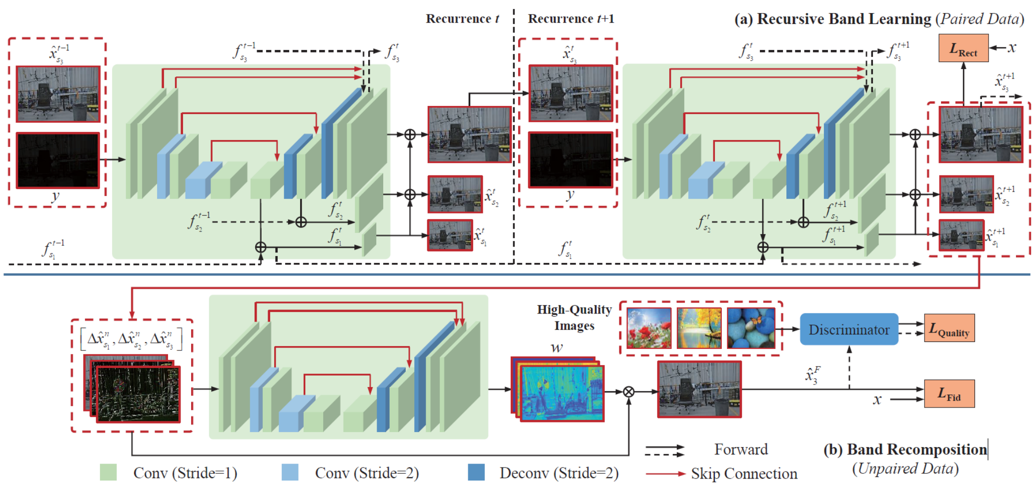

Figure 1.5: Framework of fidelity to perceptual quality

Inferences :

- This is the first attempt to provide a semi-supervised learning framework for low-light picture improvement, with a deep recursive band representation connecting fully-supervised and unsupervised frameworks to incorporate their advantages.

- In the framework shown in figure 1.5 we can see that it is well-suited for obtaining a set of coarse-to-fine band representations. Estimates of these band representations are mutually advantageous due to recursive end-to-end training, capable of reducing noise and fixing details.

- Under the perceptual supervision of quality-guided adversarial learning, deep band representations are recomposed. The discriminator's actual pictures are chosen perceptually based on the mean opinion score (MOS). This is also, as far as we know, the first trial in low-light picture enhancing tasks.

4. Extremely Low-light Image Enhancement with Scene Text Restoration. [3]
Simple U-Net architecture used to restore the text to complete the pipeline

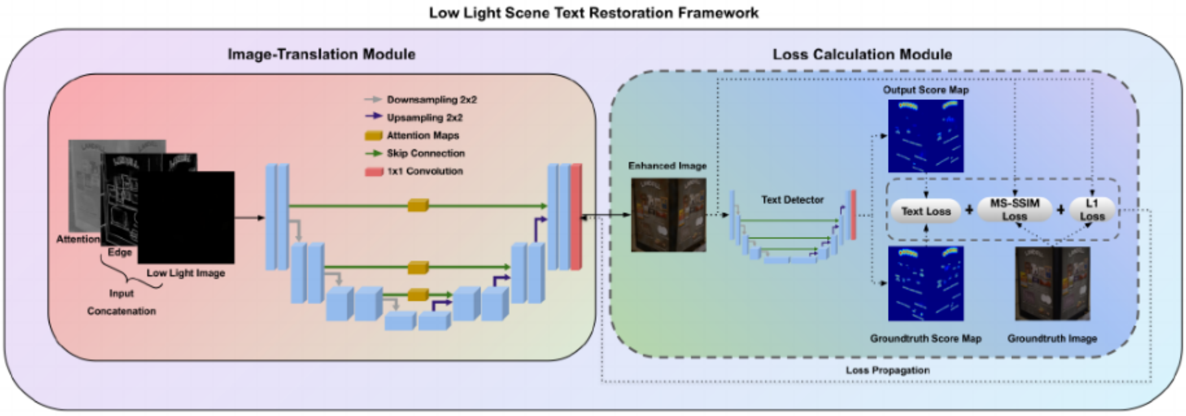

Figure 1.6: Framework of low-light image enhancement with scene text restoration

Inferences :

- Initially, they proposed an image improvement system capable of concurrently increasing both low-light picture quality (in general) and scene text with a revolutionary text detection loss. As we can see in figure 1.6 the model allows the augmented pictures to retain finer low-level information such as character borders and forms without losing overall image quality, resulting in successful text identification.
- In terms of dataset, they have annotated the texts in the real low-light dataset—See In the Dark (SID) Sony and create a synthetic low-light dataset based on the commonly used ICDAR-15 dataset.
- Extensive testing has revealed that their suggested model outperforms other models in terms of text identification and spotting tasks on improved low-light pictures in both the See In the Dark (SID) and ICDAR15 datasets.

1.4 Problem statement

To propose a learning-based technique for Enhancement of videos captured in Low-light conditions.

1.5 Application in Societal Context

Enhancement of videos captured in low-light conditions, is one of the requirements in present societal conditions. If we can achieve more night vision capacity through such learning based models, it will be a useful to many elements/roles of society.

Enhancement of images captured in low-light conditions can have novel societal applications as follows:

1. Night mode captures in mobile phones
2. Autonomous driving in night mode conditions
3. Analyzing low-light camera footage in forensics
4. Help people with Nyctalopia
5. Night vision equipment
6. Espionage
7. Drones and combat planes in target acquisition during night missions
8. Prevents mistargeting
9. Counter insurgency

Chapter 2

System design

In this chapter, we will be looking towards the functional block diagram and the final design that is being implemented.

2.1 Proposed block diagram of LLVE-Net

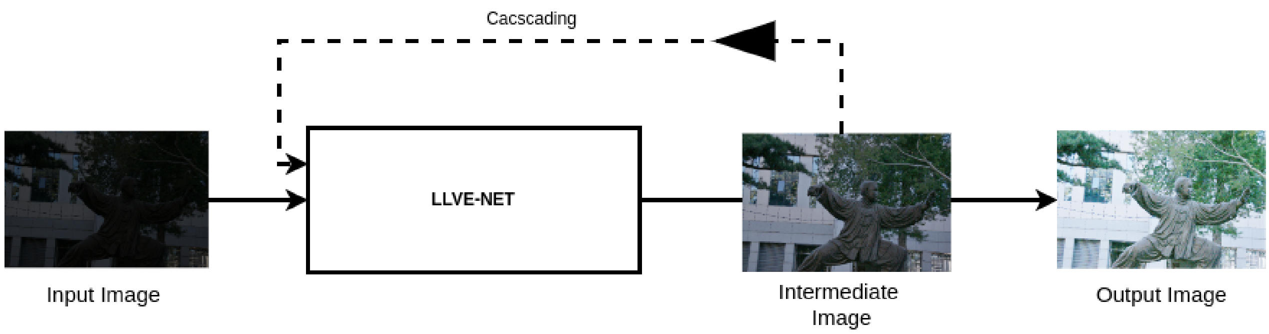

Figure 2.1: Block diagram of proposed architecture (LLVE-Net)

From figure 2.1 we can see that Low-light images are given as input to LLVE Architecture which produces the intermediate image this intermediate image is cascaded back to the Network as the input image, this process in which we are passing the enhanced image back to the Network is called Cascading.

2.1.1 Why cascading?

Here in cascading, we are passing the enhanced image back to the Network again as an input to the LLVE-Net, by cascading the enhanced image back to the network the model learns more local features in order to enhance the image more effectively and make the model robust.

2.1.2 Overview of LLVE-Net architecture

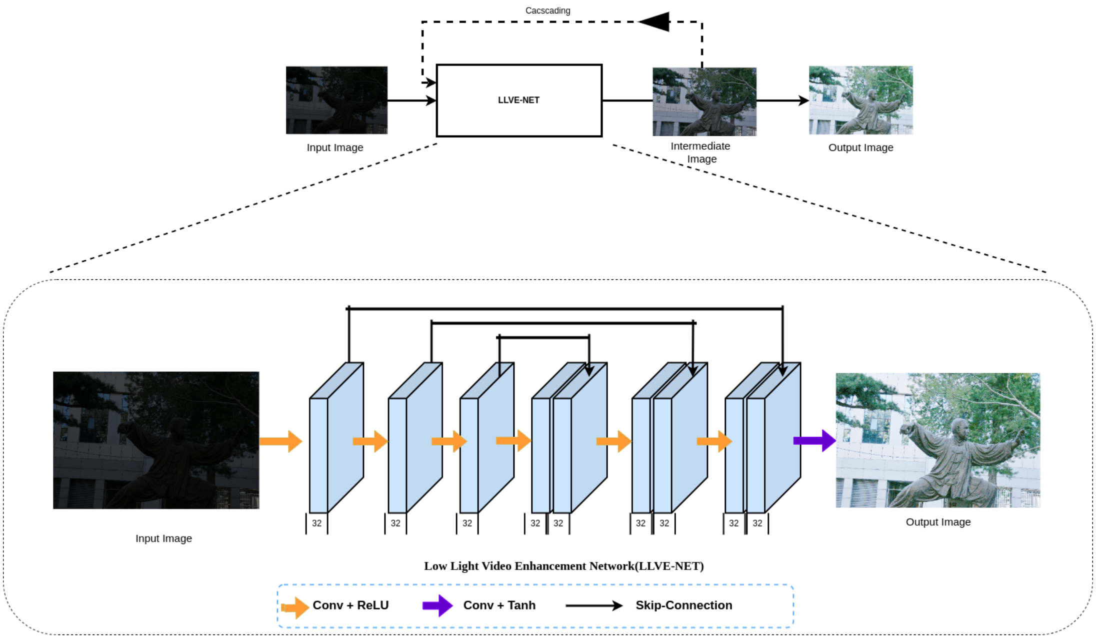

Figure 2.2: LLVE-Net Architecture

A Low Light Video Enhancement Network (LLVE-Net) is proposed to learn additional local and global features as shown in fig 2.2 , along with the mapping between an input image and its best-fitting curve parameter maps. The above figure depicts the complete Network design and parameter settings of LLVE-Net. The LLVE-Net takes a low-light image as input and returns a series of pixel-wise curve parameter mappings for matching higher-order curves.
We use a standard CNN composed of seven convolutional layers concatenated symmetrically. Each layer is made up of 32 convolutional kernels, each measuring 3 by 3, stride 1, and the ReLU activation function. The batch normalization and down sampling layers, which alter the relationships between adjacent pixels, are removed. The Tanh activation comes after the final convolutional layer.
We adopted an approach called cascading, which allows us to send the intermediate image back into the model as input so that it may learn additional local and global features. Experiments have shown that the new method improves the picture even more accurately.

Chapter 3

Implementation details

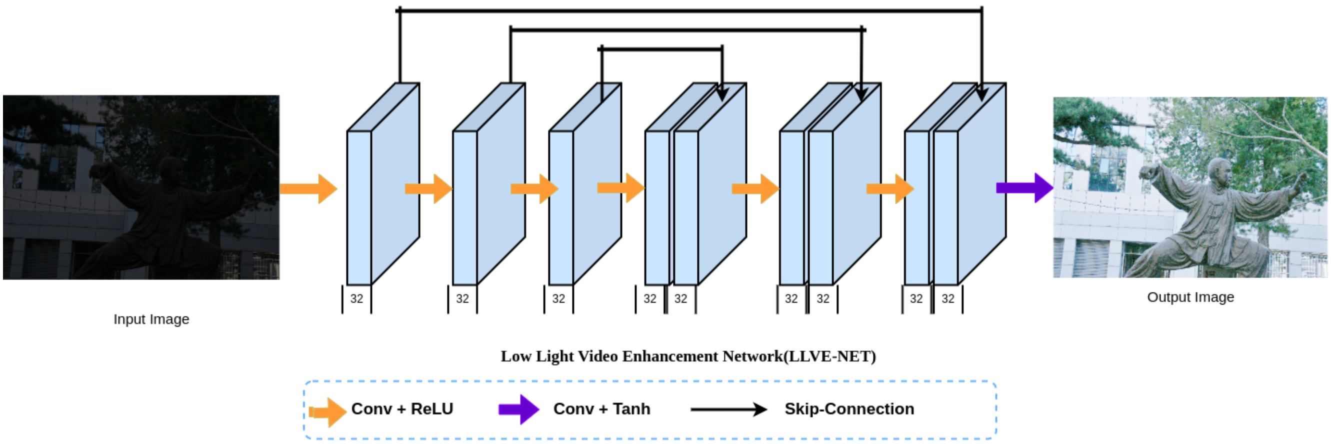

Figure 3.1: LLVE-Net architecture

The DCE-Net model used 360 multi-exposure sequences from Part 1 of the SICE dataset to train their model, which included both low-light and over-exposed images. The LOL dataset was utilised as paired input images for the LLVE-Net as shown in figure 3.1. We randomly divided 500 paired input images into two parts (485 images for training and the rest for validation). The training and testing images are resized to 512×512×3.

A batch size of 8 is used. Each layer's filter weights are set to the conventional zero mean and 0.02 standard deviation Gaussian function. Bias is set to be a constant. For Network optimization, we utilise the ADAM optimizer with default parameters and a fixed learning rate of $1e^{-4}$. To balance the scale of losses, the weights w_1,w_2,w_{col}, and w_{TV_A} are set to 0.2,0.5 0.5, and 20 respectively.

3.1 Loss functions

1)**Spatial Consistency Loss :**
By keeping the distinction between nearby regions in the input image and its enhanced form, Lspa promotes spatial coherence of the enhanced image

$$L_{spa} = \frac{1}{K}\sum_{k=1}^{M}\sum_{j\in\Omega(i)}(|Y_i - Y_j| - |I_i - I_j|)^2, \tag{3.1}$$

where Y is the average intensity value of a local region in the enhanced image and M is the number of 16x16 non-overlapping local regions.

2)**Exposure Control Loss:**
We create an exposure control loss Lexp to regulate the exposure level in order to limit under- and overexposed regions. The difference between the average intensity value of a local area and the well-exposedness level E is known as the exposure control loss. We set E as the RGB colour space's grey level in accordance with conventional wisdom. Although we did not see much of a performance change by setting E inside, we used a value of 0.6 in our trials. Lexp's loss might be written as

$$L_{exp} = \frac{1}{M}\sum_{k=1}^{M}|Y_k - E|, \tag{3.2}$$

where Y is the average intensity value of a local region in the enhanced image and M is the number of 16x16 non-overlapping local regions.

3)**Color Constancy Loss:** We create a colour constancy loss to correct any potential colour deviations in the enhanced image and to establish relationships between the three adjusted channels, all while adhering to the Gray-World colour constancy hypothesis, which states that colour in each sensor channel averages to grey over the entire image. It is possible to express the colour constancy loss L_{col} as:

$$L_{col} = \sum_{\forall(p,q)\in\varepsilon}(J^p - J^q)^2, \varepsilon = \{(R,G),(R,B),(G,B)\}, \tag{3.3}$$

where (p,q) designates a pair of channels and J p stands for the average intensity value of the p channel in the enhanced image.

4)**Illumination Smoothness Loss :** We add an illumination smoothness loss to each curve parameter map A in order to maintain the monotonicity relations between adjacent pixels. The definition of the illumination smoothness loss L_{TV_A} is

$$L_{tv_A} = \frac{1}{N}\sum_{n=1}^{N}\sum_{c\in\xi}(|\bigtriangledown_x A_n^c| + |\bigtriangledown_x A_n^c|)^2, \xi = \{R,G,B\} \tag{3.4}$$

where N is the number of iteration, $\bigtriangledown_x$ and $\bigtriangledown_y$ represent the horizontal and vertical gradient operations, respectively

5)**Perceptual Loss :** The mean squared error or the least absolute error (L1) error for the base loss is used to compare the activations at the same layer for the (target) original image and the generated image. The perceptual loss $L_{VGG/i,j}$ is defined as

$$L_{VGG/i,j} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{x=1}^{H_{i,j}} (\psi_{i,j}(I^{GT})_{x,y} - \psi_{i,j}(G_{\theta_g}(I^{LL}))_{x,y})^2 \tag{3.5}$$

where ,$\psi_j(x)$be the activations of the j^{th} layer of the Network ψ when processing the image x.

6)**Total loss :** In total loss we provide weights to each of the loss , compute them and pass it to back propagation.

$$L_{total} = \omega_1(L_{spa} + L_{exp} + L_{col} + L_{tv_A}) + \omega_2 L_{VGG/i,j} \tag{3.6}$$

where ω_1 andω_2 are weights.

Chapter 4

Results and discussions

4.1 Dataset Description

- SICE Dataset [5]
  - Training Samples: 2000
  - Testing Sample: 64
  - The dataset is captured with multiple exposure levels.

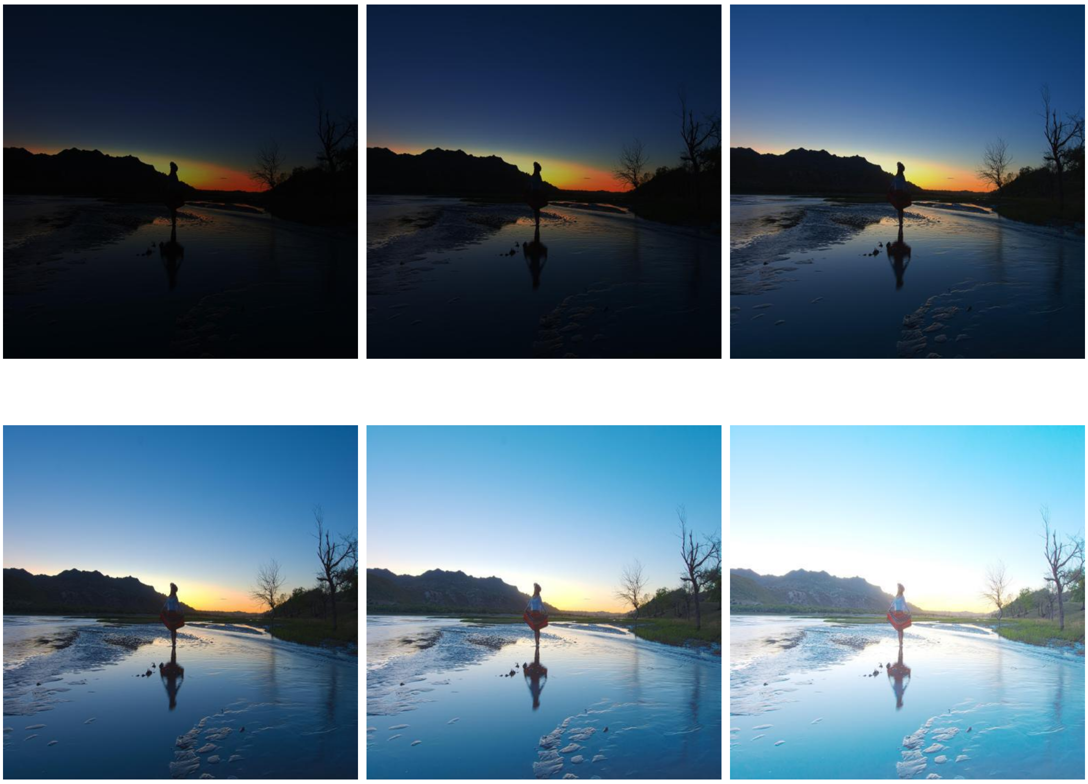

Figure 4.1: The above images depict low-contrast images with different exposure levels.

- LOL dataset [8]
  - Training Images : 485
  - Testing Images : 15

Figure 4.2: First row depicts low lit images, second row depicts ground truth images

- NTIRE (cube++) Dataset [1]
  - Training Images : 1220
  - Testing Images : 50

Figure 4.3: First row depicts NTIRE input images, second row depicts NTIRE ground truth images

4.2 Experimental Results

- Zero-DCE(Results of NTIRE Dataset [1])

Figure 4.4: First row depicts NTIRE input images, the second row depicts output enhanced images, the third row depicts NTIRE ground truth images

- The enhanced image consists of white sheet and the enhanced image is not accurate

4.3 Result Analysis

- Zero-DCE Results

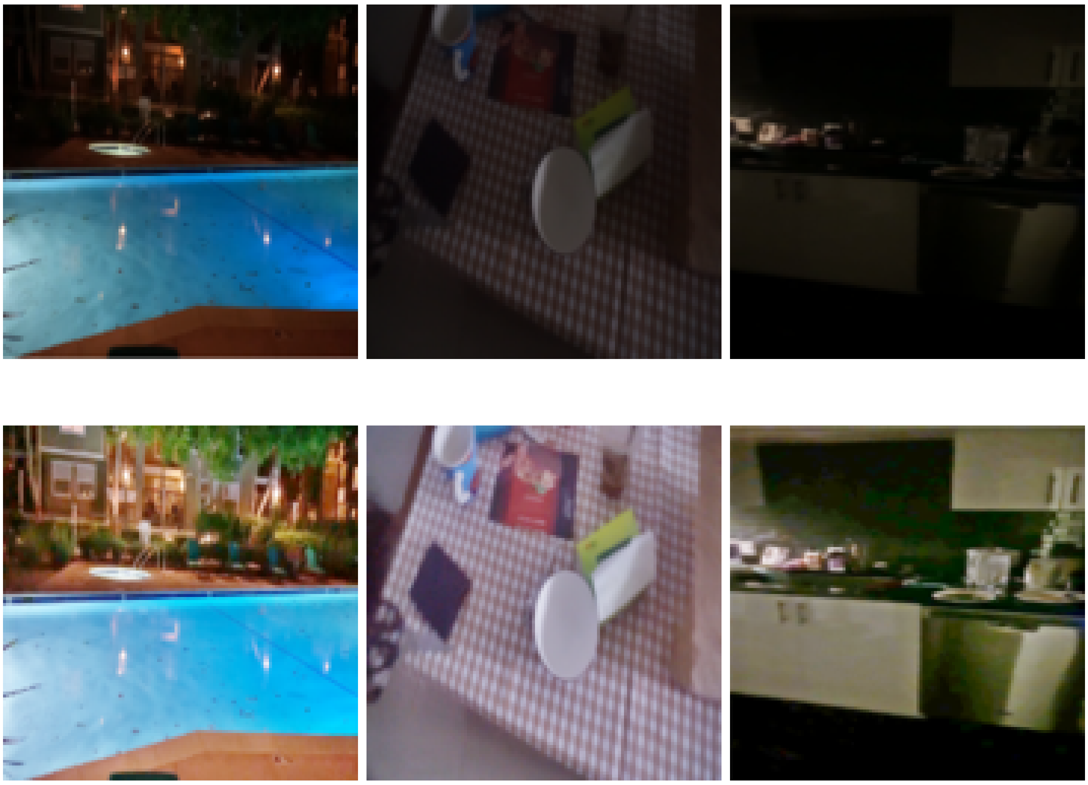

Figure 4.5: First row depicts input images of SICE dataset [5], second row depicts enhanced output images

- From figure 4.5 the enhanced image consists of noise and the enhancement is also not much accurate and cannot be used for various applications.

- LLVE-Net Results

Figure 4.6: First column depicts low lit input images, the second column depicts output enhanced images, the third column depicts ground truth images

- The results above were obtained using the LOL dataset[8].
- From figure 4.6 the Enhanced Image consists of less noise as compared to zero-dce and the enhanced image is more accurate.

4.4 Evaluation metrics

We use the Peak Signal Noise Ratio **PSNR**[4] and Structural Similarity Index Measure (SSIM) evaluation metrics to assess the outcomes produced by the suggested architecture. While **SSIM**[7] is a comprehensive metric that measures image quality loss caused by processing, such as data comprehension or data transfer loss, and is a complete reference metric that requires the ground truth and the processed image, PSNR is a measure of the power of corrupting noise that affects the fidelity of its representation and the ratio between the maximum possible power of the signal. PSNR and SSIM are defined as:

$$PSNR = 10log_{10}(\frac{P_{signal_{max}}}{MSE}) \tag{4.1}$$

where $p_{signal_{max}}$ is the maximum power of the signal

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

where $\mu_x\mu_y\sigma_x\sigma_y and \sigma_{xy}$ are means, standard deviations and covariances for two images x and y

4.4.1 Comparison Table

PSNR and SSIM Comparison		
MODEL	**PSNR**	**SSIM**
Zero-DCE	16.57	0.59
LLVE-Net	**27.4**	**0.7**
Zeo-DCE(NTIRE)	10.22	0.37

Figure 4.7: PSNR and SSIM Comparison of proposed architecture(LLVE-Net) with Zero-DCE

The bold numbers indicated depicts the PSNR and SSIM of our proposed architecture.

Chapter 5

Conclusions and future scope

5.1 Conclusion

- We proposed a deep Network for low-light video enhancement. It can be trained with both zero reference inputs and paired input.
- This is accomplished by redesigning the Zero-DCE Network architecture, reformulating the curve estimation, and managing the sizes of the input picture.
- We have introduced perceptual loss and cascading method to the input image, resulting in a model that is significantly light-weight and faster for real applications.
- Our technique excels at improving both performance and efficiency. Experiments show that our solution outperforms existing light enhancement technologies.
- The obtained PSNR[4] and SSIM[7] of proposed architecture for low-light video enhancement (LLVE-Net) is **27.4** and **0.7** respectively.

5.2 Future scope

- To develop a denoising algorithm as a plugin to LLVE-Net architecture.
- Introducing stopping Criteria for the Cascaded Network.
- To introduce a method to reduce video jitter.
- To improve results of Zero-DCE on NTIRE dataset.

Bibliography

[1] Egor Ershov et al. "The Cube++ Illumination Estimation Dataset". In: *IEEE Access* 8 (2020), pp. 227511–227527.

[2] Chunle Guo et al. "Zero-reference deep curve estimation for low-light image enhancement". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 1780–1789.

[3] Po-Hao Hsu et al. "Extremely low-light image enhancement with scene text restoration". In: *2022 26th International Conference on Pattern Recognition (ICPR)*. IEEE. 2022, pp. 317–323.

[4] Quan Huynh-Thu and Mohammed Ghanbari. "Scope of validity of PSNR in image/video quality assessment". In: *Electronics letters* 44.13 (2008), pp. 800–801.

[5] Shahidul Islam Khan and Abu Sayed Md Latiful Hoque. "SICE: an improved missing data imputation technique". English. In: *Journal of Big Data* 7 (June 2020). Article. ISSN: 21961115. URL: https://link.gale.com/apps/doc/A626539838/AONE?u=anon~94a45a71&sid=googleScholar&xid=eebe0390.

[6] Li Tao et al. "LLCNN: A convolutional neural network for low-light image enhancement". In: Dec. 2017, pp. 1–4. DOI: 10.1109/VCIP.2017.8305143.

[7] Shiqi Wang et al. "SSIM-motivated rate-distortion optimization for video coding". In: *IEEE Transactions on Circuits and Systems for Video Technology* 22.4 (2011), pp. 516–529.

[8] Chen Wei et al. "Deep retinex decomposition for low-light enhancement". In: *arXiv preprint arXiv:1808.04560* (2018).

[9] Wenhan Yang et al. "From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 3063–3072.