Report on

Facial Emotion Recognition and Object Detection in Low-Light Conditions using Deep Learning

By

Roland Singh (202300279)

In partial fulfillment of requirements for the award of degree in

Bachelor of Technology in Computer Science and Engineering
(2025)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SIKKIM MANIPAL INSTITUTE OF TECHNOLOGY
(A constituent college of Sikkim Manipal University)
MAJITAR, RANGPO, EAST SIKKIM – 737136

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

Department of Mathematics

CERTIFICATE

This is to certify that ***Mr. Roland Singh*** of ***Sikkim Manipal Institute of Technology*** participated and successfully completed the Summer Internship Programme-2025 on ***"Real-Time Facial Emotion Recognition and Object Detection in Low-Light Areas using Deep Learning"*** in the Department of Mathematics, IIT Guwahati during May 22 – July 10, 2025.

Dr. S. Natesan **(Supervisor)**	**Dr. Bikash Bhattacharjya** **(Coordinator, Summer Internship Programme)**	**Dr. S. Natesan** **(Head of the Department)**

Acknowledgement

My internship was conducted from 22nd May 2025 to 9th July 2025. During this time, I had the enriching opportunity to interact and collaborate with students at IIT Guwahati. Engaging in academic and informal discussions with them allowed me to gain a deeper understanding of the competitive and intellectually stimulating environment of the institute. Overall, this experience was unforgettable and truly meaningful, marking a significant milestone as my first internship.

I would like to express my sincere gratitude to Prof. Natesan Srinivasan, Head of the Department of Mathematics, Indian Institute of Technology Guwahati, for his invaluable guidance, encouragement, and constant support throughout the course of this internship. His insights and mentorship were instrumental in shaping both the direction and outcome of this project.

I am also thankful to the Department of Computer Science and Engineering, Sikkim Manipal Institute of Technology, for providing me with the opportunity and necessary resources to undertake this summer research internship.

Special thanks to my peers and lab colleagues for their insightful discussions and constructive suggestions, which significantly enhanced the quality of this work. I would also like to acknowledge the open-source platforms and datasets—RoboFlow for facial emotion data and ExDark for object detection data—which were essential for the practical implementation of this research.

Abstract

Images captured by computer vision systems under low-light conditions often suffer from several challenges such as high noise levels, poor illumination, excessive reflectance, and low contrast. These factors make object detection particularly difficult. Significant research has been conducted on enhancing such images using both traditional pixel manipulation techniques and deep learning approaches some focusing on improving illumination, while others aim to reduce noise.

In our work, we address this problem in two distinct phases:

1. Image Enhancement Phase: We investigate which image enhancement algorithms are best suited for downstream tasks like object and facial emotion detection—where preserving meaningful features is more critical than merely improving visual quality. Specifically, we compare basic histogram-based techniques such as Adaptive Histogram Equalisation, CLAHE (Contrast Limited Adaptive Histogram Equalization) with more advanced generative methods like Diffusion Models and deep learning-based approaches such as Zero-DCE++.

2. Object and Facial Emotion Detection Phase: We apply a range of object and facial emotion detection models to the enhanced images, evaluating their effectiveness. This includes both pretrained networks and custom CNN architectures.

Our evaluation showed that the facial emotion recognition model achieved good performance with a validation accuracy of around 81%, supported by steady convergence trends in training curves and consistent performance across unseen test data.

To further enhance low-light scenes, we integrate the Zero-DCE++ framework, which includes DCENet for estimating light enhancement curves in a zero-reference, unsupervised manner. This significantly boosts detection performance in poorly illuminated environments.

The object detection model achieved a precision of approximately 0.75, and a recall close to 0.78, highlighting its ability to correctly detect objects even in low-light conditions. Furthermore, the mean Average Precision (mAP) at IoU threshold 0.5 reached about 0.7, demonstrating the model's effectiveness across a range of localization thresholds.

Contents

List of Figures

1 Introduction

Facial Emotion Recognition (FER) and Object Detection are pivotal applications at the intersection of computer vision and real-world AI systems. FER enables machines to interpret and respond to human emotions by analyzing facial expressions, while object detection facilitates the identification and localization of various entities in an image. Both tasks are foundational for systems in surveillance, human-computer interaction, smart environments, and safety-critical applications.

However, traditional FER and object detection systems often struggle under low-light conditions, which are prevalent in real-world scenarios such as night time surveillance, low-lit indoor environments, or poor camera exposure during online meetings. Low illumination significantly degrades image quality, reduces contrast, and obscures critical features, thereby impacting model performance.

This project proposes a robust, real-time FER and object detection framework that integrates Convolutional Neural Networks (CNNs) with advanced low-light image enhancement techniques, such as Contrast Limited Adaptive Histogram Equalization, Diffusion Models and so on.

2 Facial Emotion Recognition

2.1 Introduction

Facial expressions are one of the most powerful and intuitive modes of non-verbal communication used by humans to convey emotions, intentions, and social signals. The ability to automatically detect and interpret these facial emotions plays a significant role in various fields including psychology, human-computer interaction (HCI), security systems, healthcare, and entertainment. As society moves towards more intelligent and responsive systems, facial emotion detection has become an essential component for developing empathetic and adaptive technologies.

2.1.1 Model Problem

Here we deal with the problem of low-light image enhancement, which is a crucial pre-processing step for downstream tasks like Facial Emotion Recognition (FER). The goal is to restore perceptual quality and structural detail in poorly illuminated images.

Let the observed low-light image be denoted by $I_{low}(x, y)$. The general formulation of an image enhancement process can be described as:

$$I_{enh}(x, y) = \mathcal{F}(I_{low}(x, y); \theta),$$

where:

- $\mathcal{F}$ is the enhancement function or model (e.g., CLAHE, Retinex, or a learned neural model like LL-Dif),
- θ represents the parameters of the enhancement technique.
- $I_{enh}(x, y)$ is the enhanced output image.

The LLDiffusion model was explored out of curiosity to understand its potential in enhancing low-light images. Preliminary visual evaluations revealed that LLDiffusion is highly effective at restoring illumination and contrast in extremely dark scenes while preserving fine details and textures. It produces significantly brighter and more natural-looking outputs compared to traditional enhancement methods.

2.1.2 Background/Literature Survey

Recent advances in Facial Emotion Recognition (FER) leverage deep learning, particularly Convolutional Neural Networks (CNNs), to significantly outperform traditional machine learning approaches in accuracy and robustness. Benchmark datasets such as FER2013 and RAF-DB have become standard in evaluating these models, offering diverse facial expressions captured under controlled conditions. However, a major limitation of these datasets is their reliance on well-lit facial images, making them less effective in real-world, low-light scenarios.

Recent studies have explored various image enhancement techniques to address this issue. Methods like CLAHE have been shown to improve visibility by enhancing local contrast and restoring features otherwise lost in poor lighting. More recently, models such as LL-Dif (2024) have introduced diffusion-based enhancement methods specifically aimed at improving FER

performance in low-light environments. While promising, many of these approaches are computationally intensive and often lack real-time adaptability, limiting their deployment in practical applications such as surveillance, driver monitoring, or mobile devices.

2.1.3 Objectives of the Work

1. Dataset Collection and Annotation

- Collect low-light images for facial emotion detection tasks from RoboFlow.
- Perform manual labelling of the collected images.

2. Data Augmentation

- Apply image augmentation techniques (rotation, flipping, brightness adjustment, noise addition) to increase the dataset size and improve model performance.

3. Image Enhancement

- Enhance images using traditional methods like Histogram Equalisation (HE) and Contrast Limiited Adaptive Histogram Equalisation (CLAHE).
- Further improve quality using Diffusion Models for low-light image restoration.

4. Model Development and Training

- Build a custom CNN model to detect facial emotions and objects from enhanced images.

5. Benchmarking with Pretrained Models

- Compare the custom CNN with VGGNet, ResNet-50 and other models on the same dataset using training/validation accuracy, loss.

6. Testing on Unseen Data

- Evaluate the system on new low-light images not seen during training, after enhancement, to verify real-world performance.

7. Result Analysis

- Use metrics like confusion matrix, classification report to assess model effectiveness and propose directions for improvement.

2.1.4 Theory of Proposed Technique

Low-light is a challenging environment for image processing and computer vision tasks, for better visibility and quality a natural idea is to employ image enhancement as pre-process stage before proceeding to high-level vision.

a. Adaptive Histogram Equalisation

Histogram Equalization which is a computer image processing technique used to improve contrast in images. It accomplishes this by effectively spreading out the most frequent intensity values, i.e., stretching out the intensity range of the image. This method usually increases the global contrast of images when it's usable data is represented by close contrast values. This allows for areas of lower local contrast to gain a higher contrast.

Algorithm 1 Adaptive Histogram Equalization

Input: Low-light image I
Output: Enhanced image $Ienhanced$
Convert I to grayscale: $Igray = \text{rgb2gray}(I)$
Divide $Igray$ into overlapping tiles
for each tile T in $Igray$ **do**
    Compute histogram H_T of tile T
    Apply contrast limiting to H_T
    Compute cumulative distribution function CDF_T
    Apply $CDFT$ to pixels in T
end for
Interpolate pixel values in overlapping regions
$I_{\text{enhanced}} = \text{apply\_values}(I)$
return $Ienhanced$

b. Contrast Limited Adaptive Histogram Equalization

CLAHE (Contrast Limited Adaptive Histogram Equalization) is an image enhancement technique used to improve the contrast of images, especially in low-light conditions. Unlike global histogram equalization, CLAHE operates on small regions (called tiles) of the image, enhancing the local contrast while limiting noise amplification through a clip limit. This makes CLAHE particularly effective for highlighting facial features or details in medical and low-light imagery without over-enhancing noise.

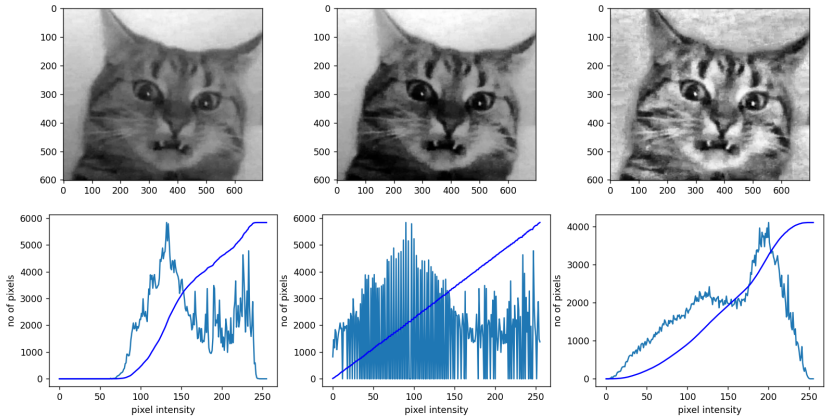


Figure 1: Result after applying Histogram Equalisation and CLAHE

c. Diffusion Model

Low-Light Diffusion Models are generative models used to enhance images captured in poor lighting conditions. They work by learning the reverse process of noise addition and gradually recovering a clean (enhanced) image from a noisy one.

Forward Diffusion Process

In the forward process, Gaussian noise is gradually added to a clean image x_0 over T steps, forming a Markov chain. The process is defined as:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t} \cdot x_{t-1}, \beta_t \cdot \mathbf{I}) \tag{1}$$

where:

- x_0 is the original clean image.
- x_t is the noisy image at time step t.
- β_t is a small positive variance value controlling noise level at each step.
- $\mathcal{N}$ denotes the normal (Gaussian) distribution.

Reverse Denoising Process

The reverse process aims to recover the clean image from the noisy image by learning the mean and variance of the reverse distribution:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \tag{2}$$

where:

- μ_θ is the mean predicted by a neural network.
- Σ_θ is the variance, which can be fixed or learned.
- θ represents the parameters of the neural network.

Training Objective

The most commonly used loss function is the simplified mean squared error (MSE) loss, which compares the true noise ϵ with the predicted noise ϵ_θ:

$$\mathcal{L}_{simple} = E_{x_0,\epsilon,t}\left[\|\epsilon - \epsilon_\theta(x_t, t)\|^2\right] \tag{3}$$

where:

- $\epsilon \sim \mathcal{N}(0, \mathbf{I})$ is the Gaussian noise added in the forward process.
- ϵ_θ is the predicted noise by the model at time step t.

Figure 2: Images before and after applying LL-Diff

2.2 Methodology

In our project, we chose to work with the low-light facial emotion database from RoboFlow because it contains low-light images. Edges, blobs, hues, and last layers are examples of low-level features. In our case, there are 7 Basic Emotions namely Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise.

The dataset contains a total of 10,807 images, which are divided into three parts: training, testing, and validation. 80 percent of the images, which equals 8,645, are used for training the model, helping it learn patterns and features from the data. 10 percent of the dataset, or 1,081 images, is allocated for testing to evaluate the model's performance after training.

For the training set, data augmentation techniques are applied to improve the model's generalization. These include rescaling pixel values to the [0, 1] range, randomly rotating images up to 20 degrees, applying a zoom transformation of up to 20 percent, and randomly flipping images horizontally. The validation data is only rescaled without any augmentation to ensure an unbiased evaluation.

2.2.1 Model Architecture

In this work, a custom CNN was designed to perform image classification on grayscale images of size 128×128. The model begins with a convolutional layer consisting of 32 filters of size 3×3 with ReLU activation and same padding to preserve spatial dimensions. This is followed by another convolutional layer with 64 filters, again using 3×3 kernels and ReLU activation, enabling the model to capture low-level edge and texture features. Batch normalization is applied subsequently to stabilize and accelerate the training process. A max pooling layer with a 2×2 window is then used to reduce spatial resolution, followed by a dropout layer with a rate of 0.3 to prevent overfitting.

To capture more complex and abstract features, a deeper convolutional layer with 128 filters, followed by batch normalization, max pooling, and another dropout layer. This multi-layered convolutional block allows the network to learn a hierarchy of features from fine to coarse. After the feature extraction stages, the output is flattened into a one-dimensional vector, which is passed through a fully connected layer with 128 neurons, 64 neurons and ReLU activation. We have also used optimiser Adam, one of the most widely used optimization algorithms in deep learning due to it's adaptive learning rate.

The initial accuracy after training for 50 epochs, achieved a training accuracy of approximately 73.3%, which indicated that the network was struggling to effectively learn discriminative features for facial emotion classification. Due to this limited performance, transfer learning was employed to enhance accuracy and generalization. Since VGG16 was trained on RGB images, while our dataset consists of grayscale images of size 128×128, we adjusted the input layer to accept 3-channel input by replicating the single grayscale channel three times. After training the top custom layers for a few epochs, we fine-tuned the deeper layers of VGG16 by unfreezing the last few convolutional blocks.

2.3 Results

We conducted our experiments on Google Colab using GPU acceleration. Regarding hardware specifications, Google Colab offers access to NVIDIA TeslaK80, T4, P4, P100, and V100GPUs, depending on availability.

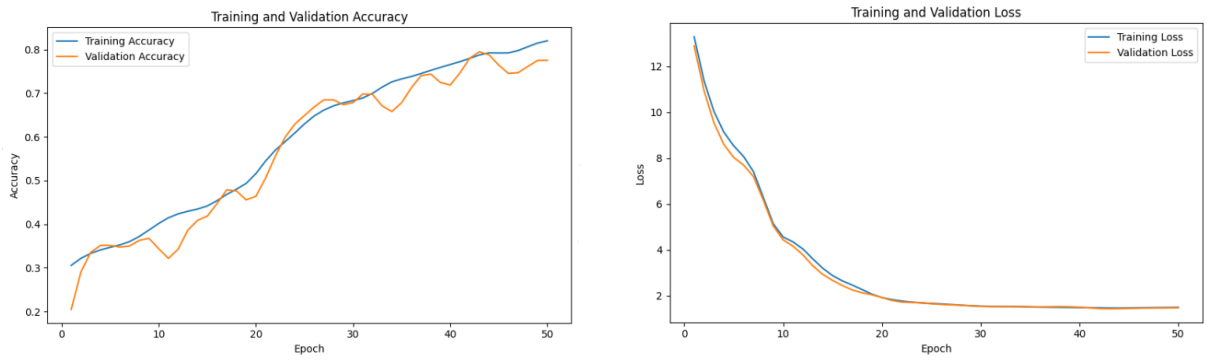
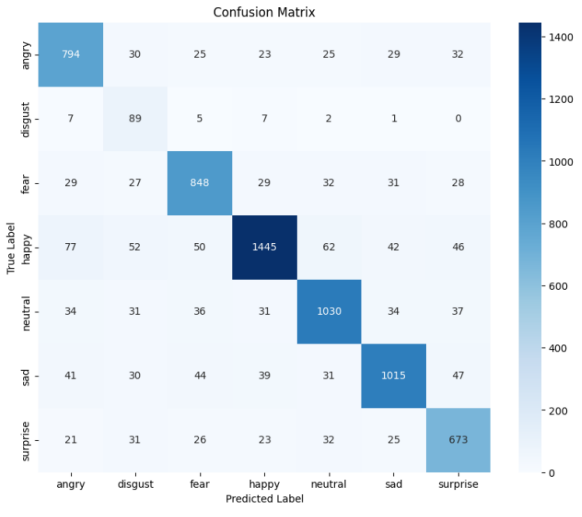


Figure 3: a)Training and Validation Accuracy b)Training and Validation Loss

As seen on Fig 2. both training and validation accuracy show a steady increase over the epochs. The validation accuracy closely follows the training accuracy throughout the training period, with minor fluctuations. Both curves converge to around 0.78–0.80 by the end, suggesting good generalization and no overfitting.

Similarly, Initial loss values are high, but they sharply drop and plateau around epoch 20, indicating fast learning initially. After epoch 20, the loss continues to decrease more slowly and stabilizes around 1.5–2.



Classification Report:

	precision	recall	f1-score	support
angry	0.79	0.83	0.81	958
disgust	0.31	0.80	0.44	111
fear	0.82	0.83	0.82	1024
happy	0.90	0.81	0.86	1774
neutral	0.85	0.84	0.84	1233
sad	0.86	0.81	0.84	1247
surprise	0.78	0.81	0.79	831
accuracy			0.82	7178
macro avg	0.76	0.82	0.77	7178
weighted avg	0.84	0.82	0.83	7178

Figure 4: Confusion Matrix and Classification Report

As seen on Fig 3. out model achieves an overall accuracy of 81% on the validation set. Among the classes, happy stands out with the highest precision (0.90) and F1-score (0.86), followed closely by sad and neutral, both with F1-scores of 0.84, showing that the model is particularly effective in recognizing these emotions. In contrast, the disgust class exhibits the weakest performance, with a precision of just 0.31 and an F1-score of 0.44.

Model	Accuracy
VGG16	80.85%
ResNet50	82.53%
MobileNetV2	80.68%
EfficientNetB0	84.04%

Table 1: Performance Comparison Between Pretrained Models on Facial Emotion Recognition

Figure 5: Emotion detection in low-light conditions a) Happy b) Sad c) Neutral d) Surpise e) Angry f) Disgust

3 Object Detection

3.1 Introduction

Object detection is a cornerstone task in computer vision that involves identifying and localizing objects within images or video frames. It serves as the foundation for numerous real-world AI applications, including surveillance, autonomous vehicles, industrial automation, robotics, and smart city infrastructure. Accurate object detection allows systems to perceive and interact with their environment intelligently.

However, one of the significant challenges faced by object detection models is their performance degradation in low-light conditions. Scenarios such as night-time surveillance, dimly lit indoor environments, or poorly exposed camera feeds often result in images with low contrast, noise, and diminished visual details. These impairments obscure object boundaries and features, making detection tasks considerably more difficult for both traditional and deep learning-based approaches.

3.1.1 Model Problem

Here we deal with the problem of low-light image enhancement, which is a crucial pre-processing step for downstream tasks like Object Detection. The goal is to restore perceptual quality and structural detail in poorly illuminated images.

The enhanced image I_{out} is obtained iteratively by applying the following update rule:

$$I_{t+1} = I_t + A_t \odot I_t \odot (1 - I_t)$$

where

- I_t is the enhanced image at iteration t, with $I_0 = I_{in}$ as the input low-light image,
- $A_t = f_\theta(I_t)$ is the enhancement curve map predicted by the neural network at iteration t,
- $\odot$ denotes element-wise multiplication,
- $t = 0, 1, 2, \ldots, T-1$, where T is the total number of enhancement iterations.

Thus, the final enhanced image is

$$I_{out} = I_T.$$

To boost detection efficacy, the work integrates the Zero-DCE++ enhancement approach with YOLOv5. The purpose of choosing Zero-DCE++ over GAN and CNN-based approaches is the advantage of zero-reference as proposed by Chongyi et al., i.e., the image dataset need not contain paired or unpaired image data to supplement its training process. Additionally, DCE-net makes it computationally efficient due to the lower number of parameters, which makes it an ideal candidate for utilization in real-time environments.

3.1.2 Background/Literature Survey

Object detection in natural and controlled environments has seen significant advancements due to the proliferation of deep learning techniques, particularly Convolutional Neural Networks (CNNs). Models such as YOLO (You Only Look Once), SSD (Single Shot MultiBox Detector), and Faster R-CNN have achieved impressive accuracy and real-time performance under standard lighting conditions. However, their performance often deteriorates when applied to images captured in low-light scenarios, primarily due to poor visibility, noise, and loss of critical features.

The enhancement of video object detection in low-light environments has been a growing area of interest in the recent past, owing to its numerous practical uses in areas such as surveillance, autonomous driving, and medical imaging. Several works have been presented to overcome the problems caused by low-light conditions, and each of these works presents new techniques for image processing and detection. These sections present a brief state-of-the-art study on low light image enhancement and object detection, analyze the potential and drawbacks of existing approaches, and establish the basis for the proposed Zero-Difference Deep Curve Estimation model. This survey covers a broad range of methods, such as dual-illumination estimation, dedicated deep learning libraries, and sophisticated neural network structures, which show that many approaches can be used to address the challenges of the nocturnal environment.

3.1.3 Objectives of the Work

1. Dataset Collection and Annotation

- Utilize publicly available low-light image datasets such as NOD and ExDARK for training and evaluation.
- Perform necessary annotation conversion (e.g., to YOLO format) and resize images for compatibility with the detection pipeline.

2. Data Augmentation

- Apply image augmentation techniques (rotation, flipping, brightness adjustment, noise addition) to increase the dataset size and improve model robustness.

3. Image Enhancement

- Integrate the Zero-DCE++ framework as a preprocessing module to enhance low-light images by estimating pixel-wise adjustment curves without the need for reference ground-truth images.

4. Model Development and Training

- Integrate the enhanced images into a modified YOLOv8s object detection pipeline.

5. Testing on Unseen Data

- Evaluate the system on new low-light images not seen during training, after enhancement, to verify real-world performance.

6. Result Analysis

- Use metrics like confusion matrix, classification report, and mAP to assess model effectiveness and propose directions for improvement.

3.2 Theory of Proposed Model

The Zero-DCE++ approach considers a dimly lit input image and learns the mapping curve to provide a brightly illuminated image via a Deep Curve Estimation network (DCE-Net). Further, the mapping curve is utilized for adjusting the dynamic pixels' range of the original image Red, Green, and Blue (RGB) channels iteratively to obtain an enhanced final version of the image [50]. The improved image's dynamic range and the surrounding pixels' contrast are preserved while best-fitting curves are approximated (curve parameter maps).

Light Enhancement Curves In this approach, the parameters of the Light Enhancement (LE) curve depend entirely upon the input image to understand the mapping between the poorly lit image and its enhanced version. Every individual pixel of the original image will receive its corresponding enhancement curve. Eq. (1) highlights the quadratic curve expression to achieve the designed image enhancement.

$$LE^{(n)}(s) = LE^{(n-1)}(s) + P(s) \cdot LE^{(n-1)}(s) \cdot \left(1 - LE^{(n-1)}(s)\right) \tag{4}$$

Start: Begin the algorithm.
Step 1: Input low-light image (I_{in}): Take the low-light image I_{in} as input.
Step 2: Pre-process the Input Image: Resize to 256 × 256.
Normalize pixel values to [0,1].
I_{pre} = preprocess_image (I_{in}).
Step 3: Initialize DCE-Net Model: Initialize the DCE-Net model.
Step 4: Estimate Light-Enhancement Curves: Use the DCE-Net model to estimate light-enhancement curves.
Step 5: Apply Light-Enhancement Curves:
Step 6: Apply LE-curves to each RGB channel.
Adjust and clip pixel values to [0,1].
$I_{\text{enhanced}} = apply\_le\_curves(I_{pre}, LE\_curves)$.
Step 7: Post-process the Enhanced Image: Convert the enhanced image pixel values back to the range [0,255].
Step 8: Return enhanced image (I_{out}): Output the enhanced image I_{out}.
Stop: End the algorithm.

Figure 6: Algorithm for DCE-Net

Figure 7: Images before and after applying Zero-DCE++ framework in ExDark dataset

3.3 Methodology

3.3.1 Model Architecture

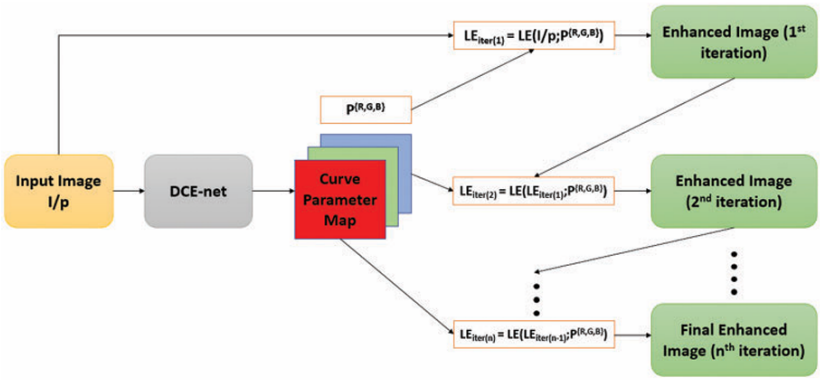


Figure 8: Diagrammatic representation of Zero-DCE++ enhancement framewor

The architecture of the Enhanced Zero DCE model mainly consists of a DCE-Net, which is a simple Convolutional Neural Network (CNN) that has the following functionalities.

3.3.2 Input Image Processing

I_{in} is the input image, which is a poor-light image that undergoes some preprocessing and is resized to a standard size for processing in the network.

Convolutional Layers DCE-Net has seven convolutional layers. All the layers have 32 convolutional filters of size 3×3 with a stride size of 1 with the exception of the last layer, and the activations apply a Rectified Linear Unit (ReLU) to the output of each convolutional layer. These layers aim to define the feature extraction of the given input image and the learning of mapping to the enhancement curves.

- **Mathematical Model:** Let $I^{(l)}$ represent the output feature map of the l-th convolutional layer. The operation can be defined as:

$$I^{(l)} = ReLU\left(W^{(l)} * I^{(l-1)} + b^{(l)}\right) \tag{5}$$

  Where $W^{(l)}$ are the weights, $b^{(l)}$ are the biases of the l-th layer and $*$ denotes the convolution operation.

Final Convolutional Layer The last convolutional layer employs the hyperbolic tangent activation function to generate 24 feature maps that are equivalent to the high-order tonal curves across eight iterations. This activation helps to keep the output values within the range of [-1, 1], which is useful in dynamic range control.

- **Mathematical Model:** The final output I_{out} is computed as:

$$I_{out} = Tanh\left(W^{(7)} * I^{(6)} + b^{(7)}\right) \tag{6}$$

  Where $I^{(6)}$ is the output from the sixth convolutional layer.

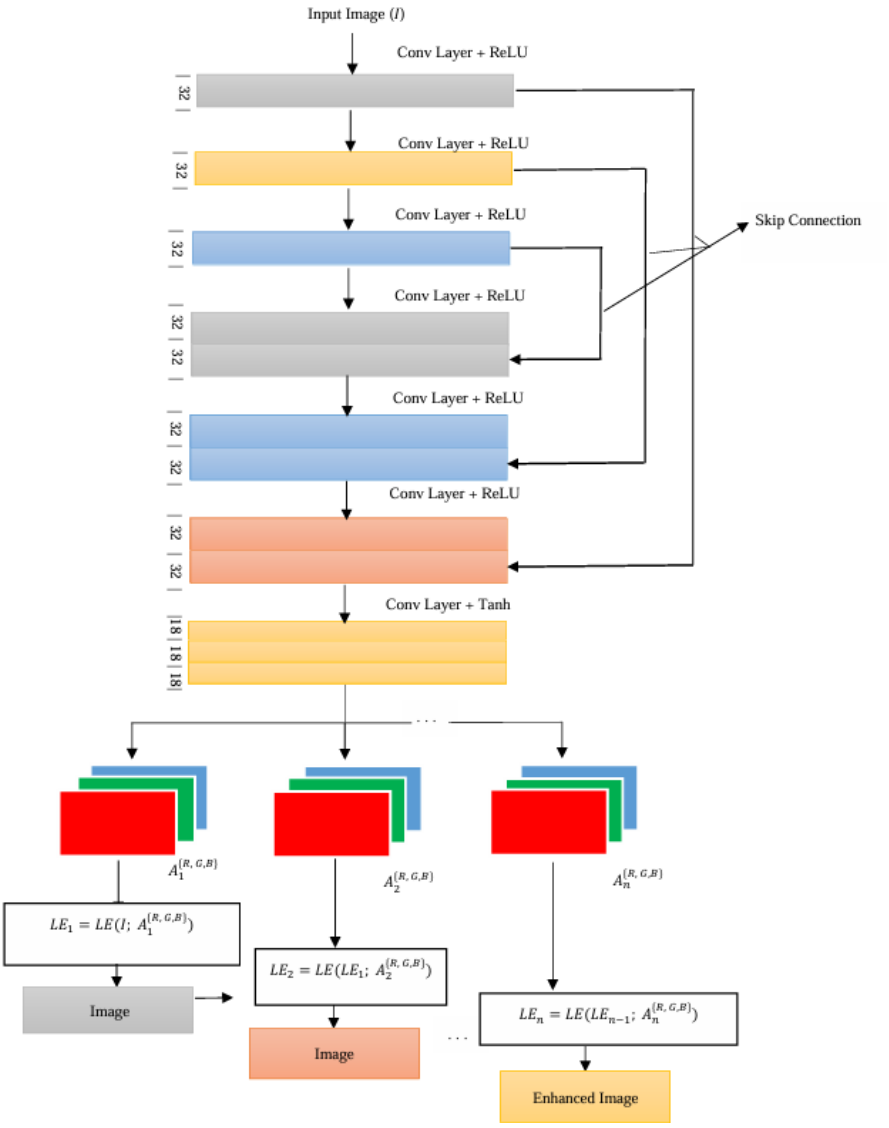


Figure 9: DCE-Net Architecture

In this project, we focused on detecting objects in low-light environments using the ExDark (Exclusively Dark) dataset, which consists of 8,363 images, all captured under challenging nighttime or poorly lit conditions. These conditions mimic real-world surveillance or nighttime scenes, making the dataset ideal for evaluating object detection models under such constraints.

To achieve efficient and accurate detection, we employed the YOLOv8 (You Only Look Once, Version 8) model. YOLOv8 was selected due to its balance between speed and accuracy, lightweight architecture, and ability to perform real-time, multi-scale object detection. It is particularly well-suited for edge deployment and constrained environments where rapid inference is crucial.

The dataset was preprocessed to enhance image quality and improve detection performance. The dataset was then split into training, validation, and test sets in an 80:10:10 ratio. We used pretrained YOLOv8 weights from the COCO dataset and fine-tuned the model on our enhanced ExDark dataset.

3.4 Results

As illustrated in Figure 9, the model exhibited a consistent decrease in training and validation loss, while precision and recall steadily improved over the epochs, indicating that the pre-processing techniques and fine-tuning strategies were effective in enhancing the YOLOv8 model's ability to detect objects under low-light conditions.

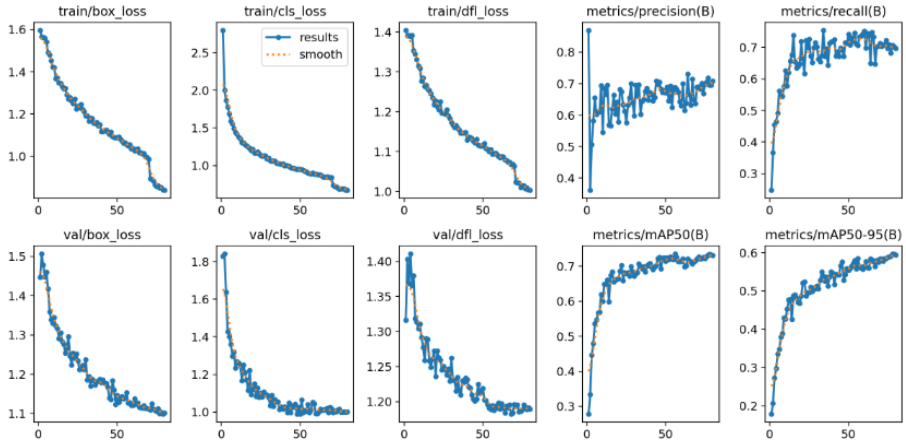


Figure 10: Performance Metrics

The model achieved a precision of approximately 0.75, and a recall close to 0.78, highlighting its ability to correctly detect objects even in low-light conditions. Furthermore, the mean Average Precision (mAP) at IoU threshold 0.5 reached about 0.7, demonstrating the model's effectiveness across a range of localization thresholds.

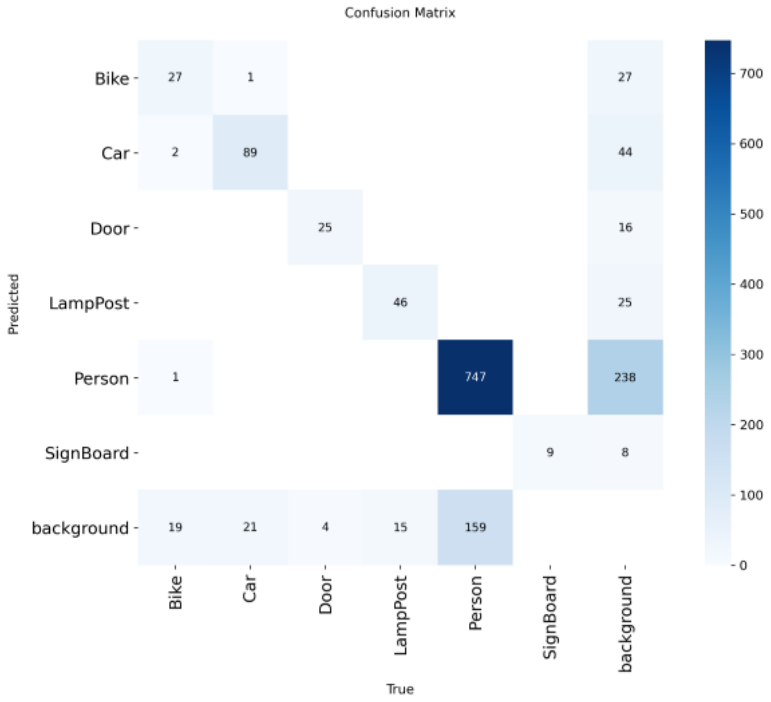


Figure 11: Confusion Matrix

Figure 12: Results for identification of target objects

Figure 13: Some of the other examples of Object detection in Low-light.

4 Conclusion and Future Work

4.1 Summary of Achievement

During the course of this project, I explored a wide range of image enhancement and detection techniques, particularly in the context of low-light conditions and facial emotion/object detection tasks. The key achievements of this work are as follows:

1. **Implementation of Traditional Enhancement Methods** I began by implementing classical low-light image enhancement techniques such as CLAHE (Contrast Limited Adaptive Histogram Equalization) and Histogram Equalization. These methods, while simple and computationally efficient, helped in understanding fundamental image processing operations and their limitations when applied to complex, real-world datasets.

2. **Application of Deep Learning–Based Enhancement Using Diffusion Models** I then transitioned to advanced generative approaches and successfully applied a low-light image enhancement diffusion model (LL-Diff). This model offered a strong ability to generate high-quality, noise-free images without requiring paired ground-truth data. The results demonstrated significant improvements over traditional methods, particularly in texture preservation and detail enhancement.

3. **Exploration of State-of-the-Art Models like DCENet and Zero-DCE++** I studied and compared deep learning architectures specifically designed for low-light enhancement:

   - DCENet (Deep Curve Estimation Network) provided an efficient, end-to-end solution for image enhancement using curve estimation.
   - Zero-DCE++ stood out as a lightweight, unsupervised model that enhances images without reference ground-truth. Its ability to learn enhancement mappings directly from low-light images is a major leap in the field.

4. **Integration of Emotion and Object Detection Pipelines** After enhancement, I successfully integrated emotion recognition and object detection modules to assess the impact of enhanced image quality on downstream vision tasks. The analysis revealed that improved illumination significantly boosts detection accuracy.

5.**Insight into Academic and Research Culture at IIT Guwahati (Mathematics Dept.)** Through interactions and indirect exposure to the academic workflow of the Mathematics Department at IIT Guwahati, I gained valuable insights into the professional and research environment. The department maintains a structured yet flexible culture, encouraging independent exploration, regular discussions, and rigorous academic standards. Faculty members are deeply involved in research, often blending applied mathematics with computational techniques, which motivated me to further delve into technical problem-solving with a scientific mindset.

4.2 Main Diffuiculties Encountered

The main difficulty encountered in the project on facial emotion and object detection in low-light images is:

Degraded Image Quality in Low-Light Conditions

Low-light images suffer from the following issues, which directly degrade the performance of detection and recognition models:

1. High Noise Levels Low-light sensors amplify signals, introducing significant random noise. This noise masks facial features and object boundaries, making detection inaccurate.

2. Poor Contrast and Illumination Essential features (e.g., facial landmarks, object edges) become hard to distinguish. Emotion classifiers rely on subtle features like eye creases, lip curvature, etc.—which become indistinct in dim lighting.

3. Color Distortion Low-light images often have unnatural color hues, confusing models trained on well-lit images.

4. Blurring and Motion Artifacts Longer exposure time or camera shake leads to blurry images. Blur further weakens keypoint detection for both faces and objects.

4.3 Summary/Conclusion

In this project, we addressed the challenges of facial emotion and objects in low-light conditions by integrating traditional image enhancement techniques with modern deep learning approaches. Specifically, we applied Contrast Limited Adaptive Histogram Equalization (CLAHE) and Histogram Equalization (HE) and also LL-Diff to enhance local contrast and recover visual features from underexposed images. These enhancements significantly improved the visibility and interpretability of facial features, which are critical for accurate classification.

Our evaluation showed that the facial emotion recognition model achieved good performance with a validation accuracy of around 81%, supported by steady convergence trends in training curves and consistent performance across unseen test data. The confusion matrix and classification report further confirmed the model's ability to differentiate between various emotions with minimal misclassification.

The trained object detection model demonstrated strong object detection capabilities in low-light scenarios, achieving a precision of around 75% and a recall nearing 78%, which reflects its reliability in identifying true positives under challenging lighting conditions.

Overall, our results demonstrate that combining classical enhancement techniques with deep CNN-based detection significantly boosts performance in low-light facial emotion recognition scenarios, with applications in real-world domains like surveillance, automotive safety, and mobile interaction systems.

4.4 Future Work

While the current system demonstrates promising results in object and facial emotion detection under low-light conditions, several avenues remain open for future enhancement. Combining traditional (CLAHE/HE) and deep learning-based techniques (like LL-Diff or Zero-DCE++) adaptively, depending on illumination level or scene type, could yield more robust results across diverse conditions. In other words, develop or adopt models that can estimate scene illumination levels and automatically choose the most suitable enhancement method (HE/CLAHE/LL-Diff) dynamically at inference time.

Additionally, future work could explore model compression techniques such as pruning and quantization to reduce computational complexity and enable real-time performance on edge devices.

References

[1] Tan, X., Triggs, B. (2010). Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE transactions on image processing, 19(6), 1635-1650.

[2] W. Zhiqiang and L. Jun, "A review of object detection based on convolutional neural network," in Proc. 36th Chin. Control Conf. (CCC), Jul. 2017, pp. 11104–11109,

[3] S. Agarwal, J. O. D. Terrail, and F. Jurie, "Recent advances in object detection in the age of deep convolutional neural networks," 2018, arXiv:1809.03193.

[4] H.K.Leung, X.Z.Chen, C.W.Yu, Liang, J.Y.Wuet al., "A deep-learning-based vehicle detection approach for insufficient and night-time illumination conditions,"Applied Sciences, vol. 9, no. 22, pp. 4769, 2019

[5] A. B. Amjoud and M. Amrouch, "Convolutional neural networks backbones for object detection," in Image and Signal Processing, A. El Moataz, D. Mammass, A. Mansouri, and F. Nouboud, Eds. Cham, Switzerland: Springer, 2020, pp. 282–289,

[6] Fuad, M. T. H., Fime, A. A., Sikder, D., Iftee, M. A. R., Rabbi, J., Al-Rakhami, M. S., ... Islam, M. N. (2021). Recent advances in deep learning techniques for face recognition. IEEE Access, 9, 99112-99142.

[7] I. Morawski, Y. A. Chen, Y. S. Lin and W. H. Hsu, "Nod: Taking a closer look at detection under extreme low-light conditions with night object detection dataset," arXiv preprint arXiv:2110.10364, 2021.

[8] Chen, Y., Li, J., Yang, W., Xu, Y., Li, Z. (2023). DiffLL: Low-Light Image Enhancement via Wavelet-Based Diffusion Models. arXiv preprint arXiv:2307.14659.

[9] J. Wang, P. Yang, Y. Liu, D. Shang, X. Hui et al., "Research on improved YOLOv5 for low-light environment object detection," Electronics, vol. 12, no. 1, pp. 3089–3111, 2023

[10] Turab, M. (2025). A comprehensive survey on image signal processing approaches for low-illumination image enhancement. arXiv preprint arXiv.2502.05995.