Contrast-Driven Network for Two-Stage Low-Light Image Enhancement

A THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY
IN
ARTIFICIAL INTELLIGENCE

Submitted by

AMIR KHAN (2K24/AFI/06)

Under The Supervision of
Prof. PRASHANT GIRIDHAR SHAMBHARKAR

Computer Science & Engineering
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering) Bawana
Road, Delhi 110042

MAY, 2026

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering) Bawana
Road, Delhi-110042

CANDIDATE'S DECLARATION

I, **AMIR KHAN**, Roll No **- 2K24/AFI/06** student of **M.Tech (Artificial Intelligence)**, hereby declare that the thesis titled "**Contrast-Driven Network for Two-Stage Low-Light Image Enhancement**" which is submitted by me to the **Department of Computer Science & Engineering, Delhi Technological University**, Delhi in partial fulfillment of the requirement for the award of degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi
Date: 20.05.2026

Amir Khan
(2K24/AFI/06)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering) Bawana
Road, Delhi-110042

CERTIFICATE

I hereby certify that the thesis titled "**Contrast-Driven Network for Two-Stage Low-Light Image Enhancement**" which is submitted by **AMIR KHAN**, Roll No **- 2K24/AFI/06**, **Artificial Intelligence**, **Delhi Technological University**, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is a record of the research work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi
Date: 20.05.2026

Prof. Prashant Giridhar Shambharkar
(Supervisor)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering) Bawana
Road, Delhi-110042

ACKNOWLEDGEMENT

We wish to express our sincerest gratitude to **Prof. Prashant Giridhar Shambharkar** for his continuous guidance and mentorship that he provided me during the research. He showed us the path to achieve our targets by explaining all the tasks to be done and explained to us the importance of this project as well as its industrial relevance. He was always ready to help us and clear our doubts regarding any hurdles in this project. Without his constant support and motivation, this research would not have been successful.

Place: Delhi
Date: 20.05.2026

Amir Khan
(2K24/AFI/06)

Abstract

Low-light image enhancement aims to convert poorly illuminated photos into visually pleasing, well-lit versions while preserving natural color and structural details. Although many classical and modern deep-learning methods successfully increase brightness, common issues persist: color casts, over/under enhancement, loss of texture, and dataset-specific behavior that limits generalization. This work proposes a self-contained contrast-driven encoder–decoder framework that removes dependency on external color-specific datasets by integrating two novel training principles: color constancy loss (a data-agnostic color balance prior) and perceptual loss (to preserve high level structure).

The framework presented is designed for achieving consistent improvement in different lighting conditions without using a paired ground truth (which makes it more appropriate for a real-world application). The model combines contrast aware feature learning with perceptual reconstruction constraints; therefore, when using the proposed framework, less noise is amplified at the edges where there is no other way to maintain edge sharpness and texture fidelity.

The encoder uses multi-scale illumination representation extraction methods; and then using progressively adding back the brightness and local contrast through feature fusing techniques the output of the decoder produces brightened image(s) with local contrast restored.
A contrastive objective guides the encoder to produce discriminative illumination-aware features while the decoder reconstructs enhanced images with skip connections preserving fine detail. We present a comprehensive methodology, layer-by-layer architecture, full loss formulations with gradients intuition, training protocol and evaluation guidelines (BRISQUE, PSNR, SSIM and user-study design).

CONTENTS

List of Tables

List of Figures

List of Abbreviations

CDNN:	Contrast Driven Neural Network
CNN:	Convolutional Neural Network
Zero-DCE:	Zero-Reference Deep Curve Estimation
RGB:	Red, Green, Blue
VGGNet:	Visual Geometry Group Network
SNR:	Signal-to-Noise Ratio
HE:	Histogram Equalization
HEP:	Histogram Equalization Prior
NTXent:	Normalized Temperature-scaled Cross-Entropy
InfoNCE:	Information Noise-Contrastive Estimation
TV:	Total Variance
MLP:	Multi-Layer Perceptron
BRISQUE:	Blind/Referenceless Image Spatial Quality Evaluator
PSNR:	Peak signal-to-noise ratio
SSIM:	Structural Similarity Index Measure

Chapter 1
INTRODUCTION

1.1 Overview

Low-light photography is very common – there are low-light photos taken by using the camera of the mobile phone, surveillance low-light photos taken inside a room, and low-light photos taken by using the cameras mounted on cars. Low-light photos always face many issues. Firstly, the low SNR causes noise to dominate the dark areas of the image. The second issue lies in the form of poor contrast and texture, causing edges to vanish from the picture. Finally, color distortion is also an issue due to imbalance among the image channels.

Classical approaches such as histogram equalization [1], gamma correction, Retinex [2], [3] work pretty efficiently and easily; nevertheless, they are highly prone to over-enhancement, halo artifacts, and even noise amplification. Neural networks (Retinex-Net [4], Zero-DCE [5], and others [6]-[8]) have many benefits in terms of perceptual quality, but they rely on a specific dataset and loss functions that cannot be applied to unseen data. The baseline approach [9] represents a contrast-aware network and able to colorize images using underwater pictures as input images, but it exploits additional datasets that might be incompatible with the desired color palette.

The proposed methodology gives a much superior different, in that it retains the strength of the encoder-decoder model without requiring any dataset-dependent loss-based color correction. Contributions of this paper include:

- Encoder-Decoder architecture with a contrastive loss function, skipped connection, and projection heads used for learning and intended to improve the images under low light conditions.
- Grey World pre-processing for correcting colors: This research implements Grey World algorithm preprocessing and colour constancy loss function for color balance without using specificied training data sets with color information.
- A parity of learning and algorithmic color correction by linking the Gray-World algorithm with learned color constancy and perceptual supervision. This mixed approach provides both global color neutrality and local semantic fidelity.
- Perceptual losses (VGG-based feature similarity [10]) plus pixel-wise (L1), exposure, spatial consistency, and total variation losses result in realistic brightness, respected edges, and preserved texture.
- A flexible augmentation and pre-processing pipeline including geometric and photometric transformations that improve model robustness and generalization.
- Reproducible training procedure with the same ablative experiments along with an evaluation process with various metrics used (such as PSNR, SSIM [11], and BRISQUE [12]).
- The validity of our model is proven via experiments on LOL dataset [13], where we obtain comparable or even superior metrics as compared to classical models and even the base paper itself.

In order to overcome limitations associated with traditional enhancement techniques, the proposed framework uses illumination aware content representations while retaining a realistic appearance. The encoder hierarchically extracts features from low light input and maintains

both global illumination distribution and local texture. These features are now sent through a contrastive representation learning mechanism that makes the network to differentiate between good enhancement and poor enhancement feature embeddings. The decodar progressively reconstruct the enhance output, with skipped connection ensuring fine grain spatial information and edges are maintained during the entire process of the output being enhanced.

Most supervised enhancement techniques rely on paired datasets. The proposed method, instead, has more generalized learning paradigms due to the use of principled loss formulations. The color constancy loss is defined as producing a balanced distribution of the RGB channels based on the Gray World assumption. This allows for unnatural color shifts to occur and therefore improves the realism of the color output. At the same time, perceptual loss is based on VGG pre-trained features, which reduces the loss of semantic structure and high frequency textures that are also lost during aggressive enhancement. These two tasks combined allow for the creation of visually appealing results that are well-balanced, have natural contrast, and structurally consistent across multiple scenes.

In addition to a variety of augmentation methods (including random crop/flip/rotate, brightness scaling, color jitter), the pipeline also uses exposure control and spatial consistency constraints to help guide the network towards stable enhancement behavior while still avoiding excessive smoothing or adding halo artifacts. In addition, total variation regularization minimizes unwanted noise amplification, while still allowing for smooth transitions within homogeneous regions. This is achieved through balanced optimization, enabling the model to generalize well in many environments; this makes the model very useful for real-world applications including mobile photography, intelligent surveillance, autonomous driving, robotics and medical imaging systems.

To evaluate the proposed method against classical image enhancement algorithms and more modern deep learning models, experiments were performed on the standard low-light benchmark datasets (e.g., LOL Dataset). Measurements of the resulting images' enhancement were carried out using not only quantitative metrics (PSNR, SSIM, and BRISQUE) but also qualitative evaluations (i.e., visually comparing results) of the overall enhanced image quality, perceptual realism, and structural fidelity. Ablation studies were also conducted in order to evaluate what contributions each of the individual techniques (contrastive learning, color constancy supervision, perceptual loss, and data augmentation techniques) made toward the overall performance of the respective models. Overall, the experimentel result indicate that propose framework produces images that exhibit clearly superior visual quality compared to other models while producing comparable numerical (i.e., PSNR, SSIM, and BRISQUE) performance. The results also indicate that the proposed framework will be very light on hardware resources and therefore would be easily reproducible in research setting and to deploy in real-world applications.

1.2 Motivation

Low-light imaging is one of the primary factors contributing to most image-based applications (e.g., everyday mobile photography, indoor surveillance systems, and automotive cameras). The majority of the images taken in a low-light setting will exhibit some form of failure; the three main types of image degradation are (1) A low SNR (signal-to-noise ratio) in the dark area makes the image less bright than it could be if there were sufficient illumination. This is further exacerbated by noise from the image sensor when capturing the image under these conditions. (2) A lack of structural detail that results in lower image contrast and reduces the ability to see edges and fine detail of the image. (3) Color reproduction is inaccurate due to

an unbalanced set of color channels resulting from the distortion of colors that occur when capturing low-light images.

Many algorithms have been used to enhance the quality of low-light images. These methods include Histogram Equalization, Gamma Correcting Techniques, and various Retinex-based methods. While these techniques can be applied quickly, they will often create an "over-enhanced" low-light photograph, which can display features such as light halos and significant, visibly high contrast noise. Recently, advancements in Deep Learning methods to improve images captured in low light (e.g., Retinex-net, Zero-DCE, CNN-based models) have produced promising results. However, many of these models rely on a well-defined training dataset or a heuristic loss function that does not generalize well to unseen conditions.

The mentioned research paper titled "Contrast-Driven Network for Two-Stage Low-Light Image Enhancement" will give more information about the process of low light photography and the enhancement of the quality of pictures taken in low lighting conditions.

1.3 Objective

In our research paper, the purpose is to design an enhanced low light image enhancement technique for better visualization of the captured image in very low lighting conditions and ensuring natural colors without any limitations on particular sets of data.

To accomplish this goal, the research will focus on four particular objectives:

1. To Brighten and Enhance Contrast in Low-Light Images

To develop a low-light image enhancement model that will both brighten and make the subject of the image easier to see. The model will do this by increasing the illumination of the subject, as well as further enhancing the subject's structure by increasing the illumination of its edges, texture and other attributes of the image.

2. To Remove Noise and Restore Fine Detail

In these cases, noise exists at a low SNR (signal to noise ratio) and removing the noise from the subject, while still presenting a stable image, has proven challenging; therefore, the goal of this research will be to address the issues of noise at an acceptable level (e.g. SNR) while also providing as high-quality an output image as is possible through techniques for restoring fine detail to the subject.

3. To Color-Correct as Accurately as Possible

Through the use of principled loss functions, the model will produce an output image of the highest possible quality (as compared to datasets containing images), in order to give as natural of a color balance as possible when compared to the image dataset with which this will be validated, by removing color distortions (e.g. blue, red) from the output image.

4. To Develop a Generalizable Low-Light Image Enhancement Model

An image of a scene taken under low light can be used to develop a generalized low-light image enhancement model, with which the model may be trained and/or validated to provide a generalized enhancement solution for the lowest possible user cost. All images captured by the same camera under the same conditions will be used for training and/or validation purposes.

1.4 Challenges

Low-light image enhancement is a rather complicated ill-posed problem owing to numerous issues arising during imaging and image processing. The current work will discuss some of those:

1. Small Signal to Noise Ratio (SNR):

In case of low lighting, insufficient light is received by the camera sensors to obtain good picture. Therefore, with increase in brightness the noise is increased, too. So, the task is to suppress the noise without loss of important details.

2. Detail Loss:

During image capturing in poor lighting conditions, there will be a decrease in the local contrast in the picture. In consequence, edges and textures are lost and it is hard to restore them in order to avoid artificial effect in sharpness.

3. Impaired Color Channel Balancing:

In case of a low-light picture, it is anticipated that there would be an imbalance in the RGB colors and would result in a picture with blue, red, or green coloring. It is difficult to correct color in a low-light picture because there are various colors in different scenes.

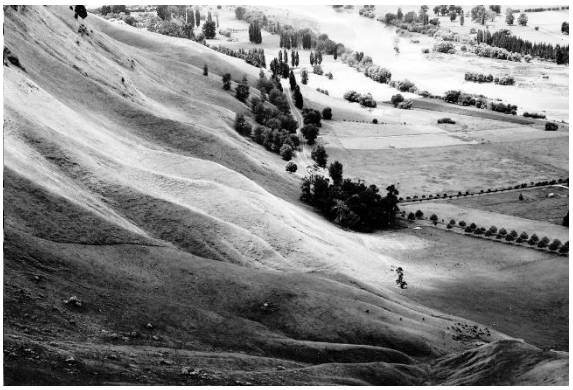
Chapter 2
LITERATURE REVIEW

1. Traditional Image-Processing Methods:

Histogram equalization (HE) is one of the most frequently used algorithms which redistributes the intensity of the image pixels. The main flaw of HE algorithm consists in the over-enhancement of bright areas, resulting in possible pixel clipping, and increasing the noise in images. Gamma correction algorithm is also quite popular due to its functionality and simplicity, but it shows the weakness of the sensitivity to selected parameters. Retinex Theory assumes that the image is formed by illumination and reflectance and helps to estimate the maps of the illumination and obtain the desirable enhancement effects. Such an approach can lead to halo effects and color distortion. Besides, traditional algorithms are characterized by lack of uniformity in estimating illumination since they cannot provide uniform results on all parts of the image. Furthermore, traditional approaches also show their weakness in enhancing the noise in the darkest regions of an image, thus worsening its quality. Finally, traditional algorithms can be implemented only in case when they make some assumptions about the scene. Therefore, despite functionality and simplicity, they are rather ineffective in complicated cases.

In addition to these constraints, conventional enhancement techniques have the drawback of failing to maintain the natural look of the image in extreme lighting conditions. The reason is that many conventional techniques make use of global processing techniques on the whole image without considering any context. This results in the situation where parts of the image become overly bright while other parts may be underexposed. This causes visually inconsistent results. Furthermore, due to the lack of dynamic adaptation of the lighting distribution in the absence of learning, hand-crafted techniques often suffer from blurring of edges, unrealistic colors, and loss of texture detail. Therefore, in applications like surveillance at night, self-driving cars, or even mobile photography, the need for joint modeling of illumination and textures arises.

(Fig.1.1 Before Histogram Equalization)

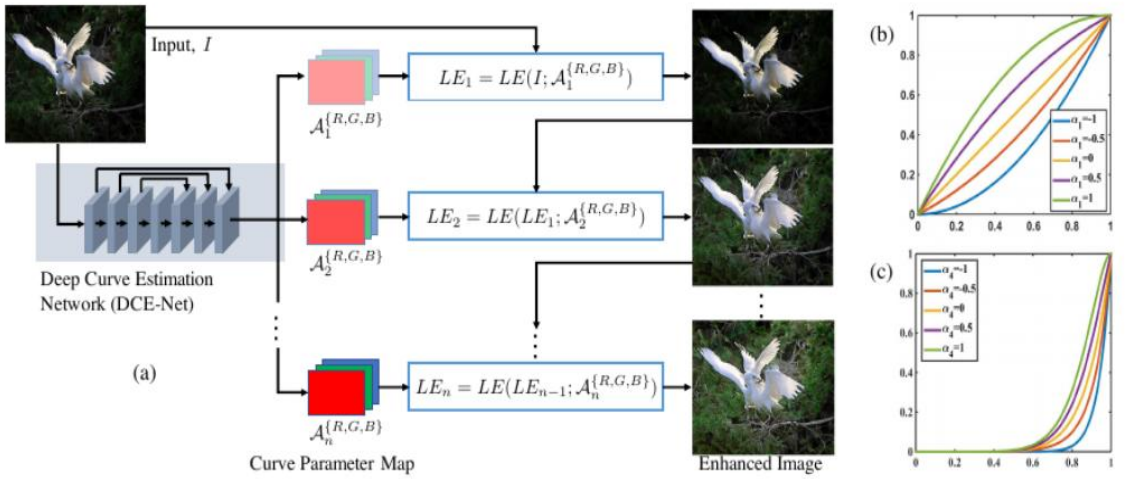


(Fig.1.2 After Histogram Equalization)

2. Learning-based Enhancement

Supervised learning methods, such as Retinex-Net [4], utilize pairs of low-light and normal-light images to train full end-to-end mappings. These methods perform well in the same domain. Zero-reference or unsupervised methods, such as Zero-DCE [5], do not use any paired data, instead relying on either priors or image-quality objectives. Still,

zero-reference methods can directly over-enhance images in some situations, or produce unexpected artifacts. Works related to this area of study involving CNN [6,7,8,15,16] have contributed significantly to this body of knowledge. Though learning algorithms provide good performance, there are various limitations associated with them including over fitting to the specific data set and inability to generalize in another illumination domain. In addition, supervised learning algorithms require vast amount of data pairs which are of high quality, thus making it difficult to deploy these algorithms in diverse environments. On the other hand, sometimes training the network on a biased dataset may result in unnatural shifting of colors and brightness in certain environments. Consequently, even though there have been great strides in deep learning for image enhancement, there is still a long way to go.



(*Fig.2.1 Pipeline of Zero-DCE*)

Another significant problem with learning-based enhancement algorithms is striking the right balance between brightness enhancement and perceptual realism. The deep learning framework often optimizes functions that emphasize mathematical similarities more than perceptual similarities, leading to over-smoothed images, unrealistic texture synthesis, and overly bright regions. Furthermore, if an algorithm is trained on data with limited diversity, it learns to enhance according to a distribution learned from the training set rather than the general characteristics of the enhancement algorithm itself. There are several sources of variation that could potentially lead to underperforming enhancements, such as sensor variations, illumination variations, and scene variations. As a result, much recent work has been devoted to hybrid frameworks that incorporate both data-driven feature extraction and perceptual, structural, and color constraints.

3. Colour Correction and Domain Adaptation

With regard to color bias, it is common for some of the methods used to incorporate the use of color information in the pre-processing stage of the model, which is referred to as fine tuning or semi-supervised learning method in cases where there is color bias in the training data set (underwater environment images) [9].

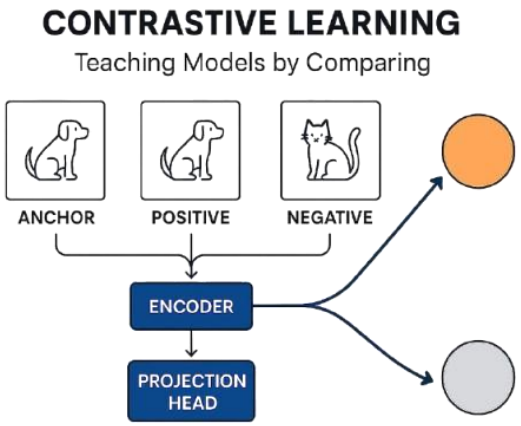
One of the main challenges faced by algorithm-based techniques based on dataset colors is their inability to generalize well from learned color distributions to new images that are acquired under different environmental conditions or using new imaging devices. For instance, a system designed on underwater datasets can achieve good results when it

comes to removing bluish or greenish shades found in water bodies. However, such learned statistics might fail to deliver accurate color correction in other settings such as night photography in the city or even indoors for surveillance cameras. This dependence on a particular type of dataset makes the enhancement algorithm less adaptable to varying situations.

To deal with this situation, recent development trends in the field have focused on introducing physics-informed priors and perceptual constraints within the machine learning framework rather than relying on the availability of external datasets which are heavily biased towards color. Methods such as color constancy loss, Gray-World priors, and perceptual feature supervision make it possible for the model to learn a universal color correction operation in an efficient way without compromising on semantic and structural consistency. In this way, not only is the need for collecting large amounts of color-related training data minimized, but the model can better cope with a wide range of lighting conditions.

4. Contrastive Representation Learning

InfoNCE and NTXent are both contrastive learning approaches that can be applied in self-supervised learning. In regards to image improvement, this can be achieved through creating a contrastive learning experience for the encoder where the encoder has to ensure that there are more similarities between the improved images and fewer similarities between the other images [25].



(*Fig.4.1 Contrastive Learning*)

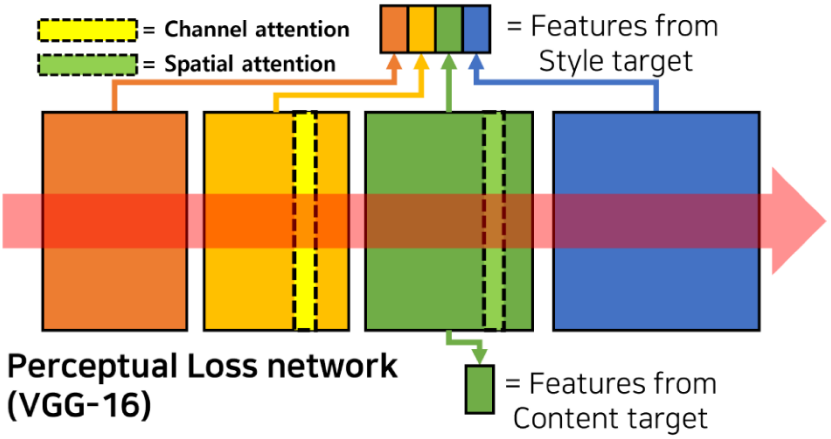
The use of the contrastive learning theory makes the feature representations learned by the neural network better discriminable and more illumination aware compared to feature representations obtained by using the traditional reconstruction approach only. The use of positive example pairs, which are composed of low-light and their enhanced or normal-light versions, will compel the encoder to learn about semantic coherence and brightness transformations while preserving it. On the other hand, the negative pairs will compel the model to learn about discriminating different illumination patterns and non-related structures within the images. These are important as the network can learn the important visual features such as edges and texture distributions.

Moreover, the use of contrastive goals in an encoder-decoder architecture leads to higher robustness and generality of the model under varied lighting conditions. In

contrast to purely pixel-based models which tend to reconstruct only local intensities, contrastive learning encourages global consistency and semantics. This way, there is less chance of over-processing and unwanted changes in the intensity levels and structure of complex scenes. Another advantage of using contrastive representation learning is that it helps stabilize the process of training by allowing the model to learn more representative features through the encoder. Consequently, the image enhancement model can generate consistent and high-quality images with well-balanced lighting conditions.

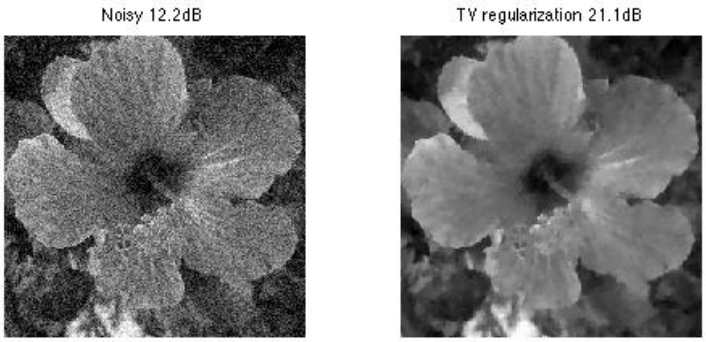
5. Perceptual and Structural Losses

Perceptual losses using a pretrained classification network like VGG capture the semantic and structural aspects of image enhancement that traditional per-pixel losses (e.g., L1 or L2) cannot extract [24]. Total variation (TV) and spatial consistency regularizers are often implemented to prevent oversmoothing and preserve local image structure [5].



(*Fig.5.1 Perceptual Loss Network*)

These types of losses work on the higher-level feature space obtained from pre-trained deep neural networks like VGGNet, helping the enhancement network retain the semantic content of the image while reconstructing it. Pixel-based losses can only help compare two images by comparing their intensity levels but perceptual losses can compare the images based on deep features of both reference and enhanced images. Consequently, it helps the network keep object boundaries, textures, and the natural structure of scenes in an image intact. Therefore, even in difficult scenarios like low-light situations, it can help achieve high-quality output images.



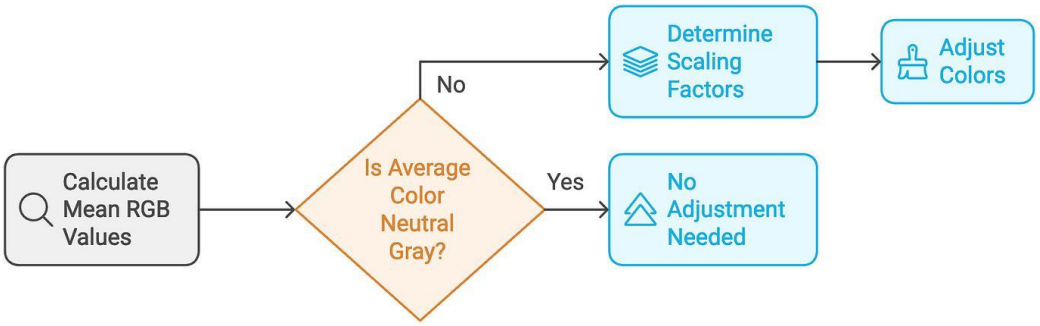
(*Fig.5.2 Total Variance Loss*)

TV regularization makes additional contributions towards achieving high-quality images by promoting spatial coherence while simultaneously reducing noise and preventing artifacts from becoming accentuated. For low-light image enhancement, a drastic increase in pixel brightness may lead to the unintended accentuation of noise and result in unnatural fluctuations in intensities within uniform regions. TV loss helps reduce abrupt changes in pixel values and ensures that light intensity transitions occur smoothly without having much impact on edge details.

In the same manner, spatial consistency is just as critical in maintaining local illumination relationships between adjacent regions in the image. This is achieved by ensuring that there is consistency in brightness ratios as well as structure, thereby eliminating any unnatural lighting effects or halos that might occur at the boundaries. Through the learning of the network, the consistency in the spatial structure ensures that the illumination adjustment takes place in a natural manner. In this regard, through the combination of perceptual losses, total variation regularization, and spatial consistency, it is possible to optimize the entire process in order to generate enhanced images.

6. Gray-World Colour Correction

Gray World Algorithm is one of the earliest algorithms used to solve the problem of color constancy. This method assumes that the mean reflectance in the scene is neutral gray and gives rise to coefficients that can normalize the intensity levels in the red, green, and blue channels so that they all have the same mean value. Many works leverage grayworld preprocessing to normalize data before or after enhancement. We incorporate this principle in our pipeline, alongside a learned color constancy loss [22].

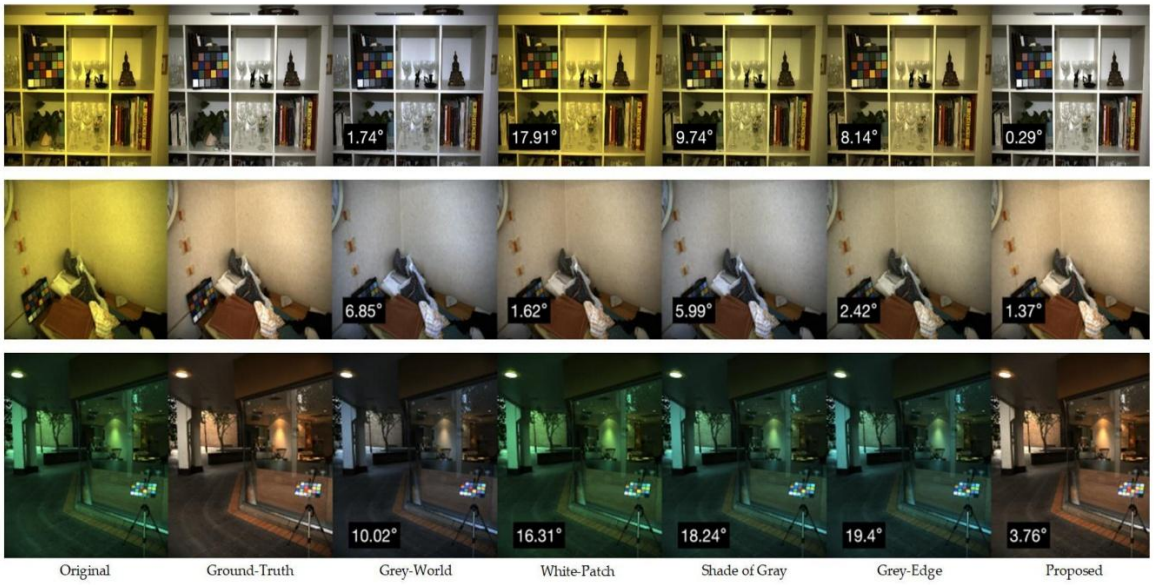


(*Fig.6.1 Gray-World Color Correction Flowchart*)

The Gray-World assumption can be integrated into deep learning systems to use both adaptive feature-learning and classical color-corrections methods in a balanced manner by having the average of the RGB channels approximately equal. The application of the Gray-World assumption during image preprocessing has reduced the number of dominant colour channels due to problems with light level or sensor limitation and assisted in making the overall low-light images visually neutral prior to moving on the the enhancement network. The result will allow the enhancement network to concentrate more on enabling the recovery of texture, contrast and structural information rather than correcting for large dominant colour channels due to illumination imbalance. Furthermore, since the computation of the Grey World method is simple and efficient, it is an excellent choice for use within a lightweight and/or real-time enhancement system.

Adding the Gray World algorithm along with the color constancy loss makes the proposed framework more robust. The classical algorithm is responsible for the global

normalization of colors, while the loss function helps the neural network learn and adjust local color distributions based on the scene context and lighting conditions. The proposed approach can help avoid excessive corrections and make sure that colors appear natural under different environmental conditions. Additionally, the combination of the classical algorithm and the deep features' supervision makes it possible for the model to generalize better, providing similar results for images not seen during training.



(*Fig.6.2 Gray-World Color Correction Flowchart Examples*)

7. Gap and Our Positioning

There are various kinds of degradation that happen to low-light photographs, meaning that there cannot be any one kind of loss that will serve to guide the image enhancement process effectively. Thus, we have incorporated several loss types that include fidelity, exposure, contrast, smoothness, color, and perceptual properties of the image. Fidelity is assured through the use of reconstruction loss, while exposure loss facilitates the adjustment of brightness in the image. Contrast is assured through spatial consistency and fine noise is minimized through the use of total variation.

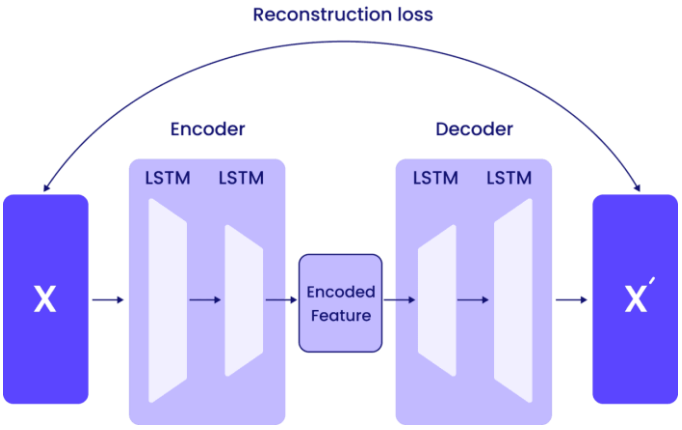
The joint use of complementary loss functions helps to enhance multiple features at once rather than concentrate on one task at hand. L1 or L2 reconstruction losses help to ensure that the output image is as similar to the original as possible in terms of intensity levels. The exposure control loss helps in ensuring that the network learns how to generate images with a balanced illumination level without having too much light in those parts that have enough illumination already. Spatial consistency constraints help in ensuring that no unnatural illumination shifts are generated between the adjacent areas. Additionally, total variation helps in eliminating noise amplification.

Along with the lower-level reconstruction loss functions, the higher-level perceptual and contrastive loss functions enhance the semantic and visual fidelity of the generated image. In particular, the perceptual loss function is designed based on the deep features produced by pre-trained deep neural networks like VGGNet. The deep features contain texture information, edge information, and structural semantics that would be ignored if pixel-level supervision alone was used for training. Meanwhile, the color constancy loss function, which is based on the Gray-World assumption, encourages the generated image

to possess an even distribution of RGB colors, thus eliminating the undesirable color bias in low-light images. Moreover, the contrastive loss function, which is designed based on the contrastive learning paradigm, facilitates better discriminative power by training the encoder to produce feature embeddings that enable distinguishing well-lit from poorly lit images.

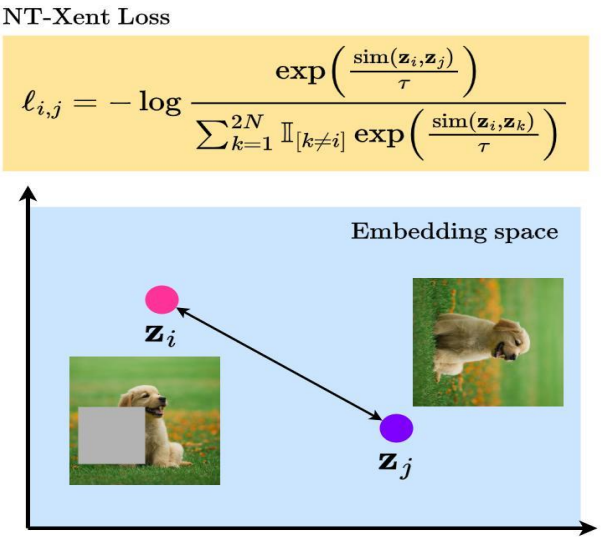
8. Loss Functions

Due to the various degradations present in low-light images, there cannot be any singular form of loss that could provide guidance in enhancing such images. Hence, this paper will attempt to address all forms of loss simultaneously in an effort to achieve balance between fidelity, exposure, contrast, smoothness, color, and perceptual quality. Pixel-level fidelity is ensured with the help of the reconstruction loss [20] while exposure of the images is controlled to avoid excessive dark or bright regions [5] through the exposure control loss. Spatial consistency loss is used to preserve local contrast in the enhanced images [5] while the total variation loss helps in smoothing the fine noise present [21]. Color constancy loss is introduced in this paper to ensure that there are no color imbalances due to the grey-world assumption [22]. In addition to achieving pixel-level, perceptual loss helps in matching the output images to high-quality semantically meaningful images [24]. Contrastive loss will be used to learn better representations of the data [24].



(Fig. 8.1 Reconstruction Loss Function)

Combining all these components within a unified structure makes it possible for the network to achieve optimal reconstruction at the low level while also achieving optimal perception at a high level. While the two latter losses concern themselves with illumination adjustment and intensity relationship preservation, the former types of losses direct the network towards the learning of semantically useful features. The multi-criteria optimization technique makes it possible for the enhancement algorithm to discover latent textures and retain sharpness of edges and natural composition of scenes, even in very dark areas.



(Fig. 8.2 NT-Xent Loss Function)

Furthermore, the synergistic effect between the mentioned losses enhances the stability and generalizability of the introduced enhancement method. Spatial consistency and total variation losses help stabilize the illumination transition process and avoid unwanted noise propagation. In addition, the supervision mechanism with regard to color constancy ensures that enhanced images appear well-balanced despite changes in lighting. Finally, the contrastive learning-based loss helps enhance the encoding process and enables illumination-aware features discrimination by the network, allowing it to separate useful information for enhancement tasks from other variables. Therefore, the presented image enhancement approach yields images with increased luminance, natural coloring, preserved texture, and high-quality visualization regardless of the data source or unseen settings.

Table 2.1: Summary of loss functions used in our work and their sources

Source	Loss	Purpose / Role
Isola et al., Image-to-Image Translation, CVPR 2017	Reconstruction (L1)	Pixel-wise fidelity; enforces similarity to ground truth.
Wei et al., Zero-DCE, CVPR 2020	Exposure Control	Prevents under-/over-exposure by regulating patch mean intensity.
Wei et al., Zero-DCE, CVPR 2020	Spatial Consistency	Preserves local contrast and texture relationships between input and enhanced images
Rudin et al., TV Denoising, Physica D 1992	Total Variation (TV)	Smooths high-frequency noise; reduces artifacts.
Buchsbaum, Color Perception Model, J. Franklin Inst. 1980	Color Constancy	Encourages gray-world assumption; balances RGB channels to reduce color casts.
Johnson et al., Perceptual Losses, ECCV 2016	Perceptual	Matches mid-level features in VGG; enforces semantic and structural similarity.
Chen et al., SimCLR, ICML 2020	Contrastive (NT-Xent)	Pulls positive pairs together in embedding space; pushes negatives apart.

Chapter 3
Dataset

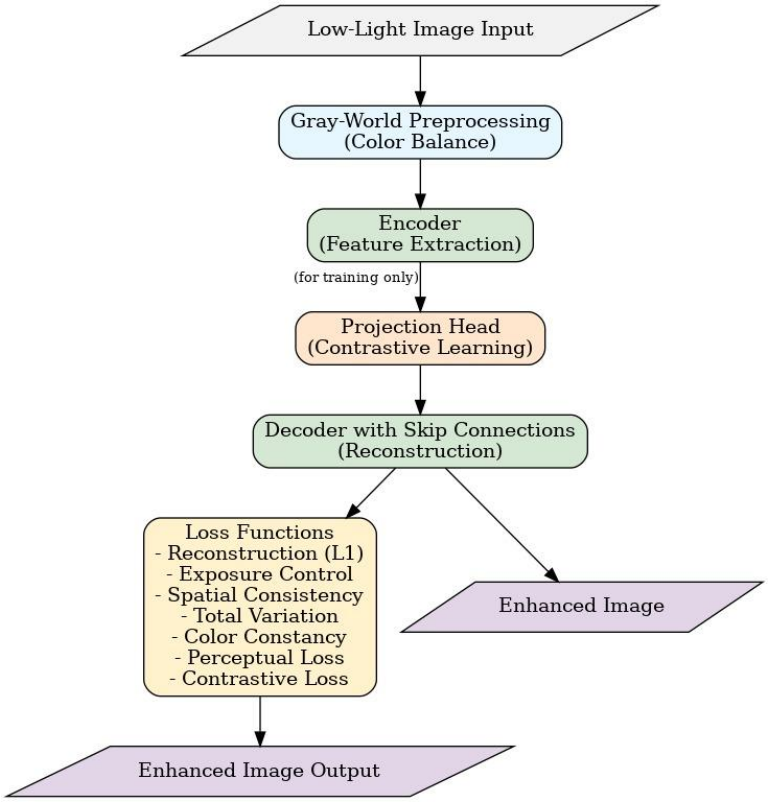
A standard benchmark for low-light image enhancement using supervised machine learning, the LOL (Low-Light) dataset contains 485 pairs of low-light images and their well-lit ground truth counterparts. The dataset is composed of 400 training pairs and 85 test pairs.

Some characteristics of low-light images in the LOL dataset include high levels of noise, low contrast, loss of texture or detail, and color distortion. Images can be classified as either indoor or outdoor scenes; therefore, they represent a wide variety of real-world scenarios, which will be beneficial in future enhancement projects. Additionally, while the dataset is diverse, it is relatively small when compared to other datasets, and therefore requires that models generalize well without having to heavily rely on color statistics associated with this dataset. The LOL dataset was chosen to provide paired supervision, exhibit realistic low-light environments, and serve as the foundation of a standard benchmark for low-light image enhancement comparison of different enhancement techniques.

Link to access dataset: *https://www.kaggle.com/datasets/soumikrakshit/lol-dataset*

Chapter 4
METHODOLOGY

In this section, In-depth explanation of the proposed enhancement framework. Specifically, notations, architecture design, optimization criteria, and overall training methodology are elaborated. The encoding-decoding model, skip connection, projection module, and feature extraction procedures of each layer have been explained with details so that it is clear how illumination information is exploited and the final images are produced from them. Moreover, definitions and detailed descriptions for all used loss functions, i.e., reconstruction loss, exposure loss, spatial consistency loss, total variation loss, color constancy loss, perceptual loss, and contrastive loss, are presented.



(Fig. 4.2 Overall flow of the proposed method)

4.1 Gray-World Colour Correction

Besides these learning-oriented goals, the Gray-World assumption [5,15] is used in this paper for the sake of pre-processing alone. Suppose μR, μG and μB represent the mean of red, green and blue color channels, respectively, of the input image I.

A way to get rid of the global color bias in an input image is to use the average intensity of three-color channels as the scaling factors to normalize each of the three-color channels to be closer to a common gray reference. Each color channel (RGB) will then be multiplied by a corresponding normalization coefficient to achieve color-balanced images that can be enhanced after preprocessing. Normalizing the color of an image before enhancing it reduces blue, red, or green tinting commonly found in low-light images due to uneven illumination and sensor limitations. The input images can be presented to the enhancement network as neutral colors allowing the

enhancement network to work more productively to recover illumination contrast, structural details, and ensure natural appearance in the processing of the image data.
The scaling factors are:

$$s_c = \frac{\mu_{avg}}{\mu_c}, \quad \mu_{avg} = \frac{\mu_R + \mu_G + \mu_B}{3}$$

Each channel is adjusted as:

$$I'_c = s_c \cdot I_c$$

This provides color balance statistically prior to forwarding the image to the enhance network. This, along with colour constancy loss, will make sure that there is no unnatural color used.

4.2 Notation and Problem Statement

Let $\boldsymbol{I_{\{low\}\backslash in}[0,1]^{\{H \backslash times\, W \backslash times\, 3\}}}$ be RGB image (at Input level) and Igt its output (ground truth available for supervised learning in databases such as LOL [14]). The neural network fθ (parameterized by θ) outputs the enhanced image:

$$I_{enh} = f_\theta(I_{low})$$

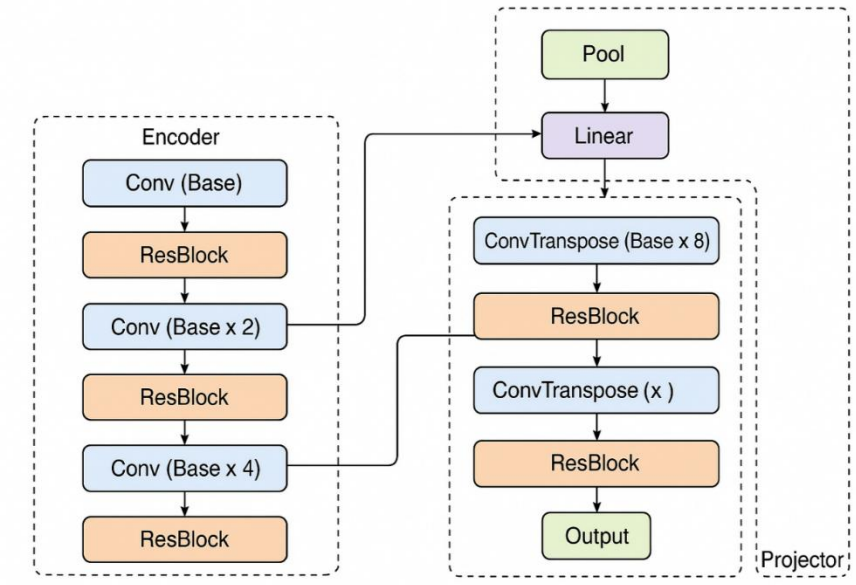
The function f itself is decomposed into an encoder Eθ, a projection head Pθ used for contrastive learning (such as NT-Xent/InfoNCE), and a decoder Dθ:

$$x = E_\theta(I_{low}), \quad z = P_\theta(x), \quad I_{enh} = D_\theta(x, \text{skips})$$

where "skips" denotes encoder features routed to decoder via skip connections.

4.3 Network Architecture

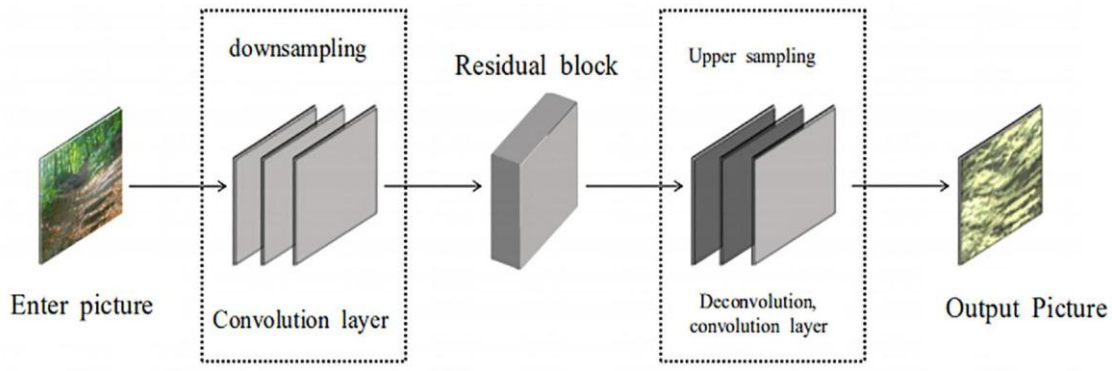
Layers of architecture of the reproducible implementation of this work are stated below. Such an architecture provides a trade-off between the ability of representing information and computational complexity; you may increase the number of channels or add more blocks to suit your hardware capabilities.



(Fig. 4.3.1 Proposed Encoder–Decoder Network Architecture with Contrastive Projection Head)

CDNN network presented in Fig. [2] uses the encoder-decoder architecture, which has a similarity to the U-Network architecture. Residual block are used both in the encoder path and the decoder path. The encoder layer reduces the resolution of the low-resolution input image \mathbit{I}_{\left\{\mathbit{low}\right\}} to obtain hierarchical features. Skipped connection between encoder and decoder allow keeping information during the reconstruction. The decoder increases the resolution of the obtained hierarchical features to generate the high-resolution output. The projection head is getting use only in the training phase.

The encoder model is engineered to be able to obtain illumination-aware representations through the process of extracting low-level and high-level features from the input image in a stepwise manner. The early convolutional layers are responsible for obtaining basic information about the input such as edge and texture data, while deeper layers make use of residual blocks to understand more complex information regarding illumination and semantics in images. The use of residual connections is beneficial to gradient propagation and avoids the problem of degradation that affects neural networks.



(Fig. 4.3.2 Residual Encoder–Decoder Network Architecture)

The decoding network employs the technique of hierarchical feature fusion and transposed convolutions to up sample the enhanced image. The skip connection facilitates direct transfer of shallow features at the spatial level from the encoder network to the decoder network, allowing for the retention of minute texture and edge details that might get lost during the down sampling step. The reuse of these features enhances the output generation and ensures stability in the enhancement process by integrating global context and local structure information. The projector that is connected to the encoder plays a critical part in contrastive representation learning. The encoder features are transformed to the latent representation space where positive samples get close together and negative samples get far away due to the optimization of the contrastive loss function. Thus, contrastive learning helps the neural network to acquire illumination-aware discriminatory features, making the learning robust against illumination changes. Due to the fact that the projector is not needed for the inference step, it gets discarded during testing to save time and computation resources.

Table 4.3.1: Recommended Encoder–Decoder layout

Stage	Operation	Output shape
Input	I_{low}	$256 \times 256 \times 3$
Enc1	Conv3x3,64; ResBlock×2	$256 \times 256 \times 64$
Down1	Conv3x3, stride2	$128 \times 128 \times 128$
Enc2	ResBlock×2	$128 \times 128 \times 128$
Down2	Conv3x3, stride2	$64 \times 64 \times 256$
Enc3	ResBlock×2	$64 \times 64 \times 256$
Down3	Conv3x3, stride2	$32 \times 32 \times 512$
Enc4	ResBlock×2	$32 \times 32 \times 512$
Proj	GAP + FC-512 + ReLU + FC-128 + L2norm	Embeddings $z \in \mathbb{R}^{128}$
Dec3	Upsample + Conv3x3 + concat(Enc3) + ResBlock×2	$64 \times 64 \times 256$
Dec2	Upsample + Conv3x3 + concat(Enc2) + ResBlock×2	$128 \times 128 \times 128$
Dec1	Upsample + Conv3x3 + concat(Enc1) + ResBlock×2	$256 \times 256 \times 64$
Output	Conv3x3,3 + Sigmoid	$256 \times 256 \times 3$

4.3.2 Residual block: A residual block computes:

$$
\begin{aligned}
y &= \mathrm{ReLU}(\mathrm{BN}(\mathrm{Conv}_{3\times3}(x))) \\
y &= \mathrm{BN}(\mathrm{Conv}_{3\times3}(y)) \\
\mathrm{ResBlock}(x) &= x + y.
\end{aligned}
$$

When channel dimensions change, a 1 × 1 projection is used on the skip.

4.3.3 Projection Head:

Contrastive embeddings may be formed using the global average pooling (GAP) on the deepest feature maps of the encoder:

$$h = \mathrm{GAP}(x_L) \in \mathbb{R}^C$$

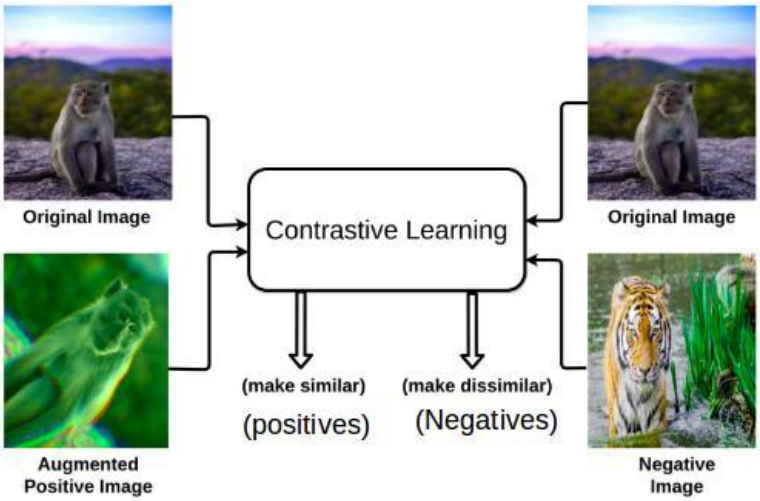
Then pass through a two-layer MLP: $u = \mathrm{ReLU}(W_1 h + b_1)$, $v = W_2 u + b_2$, and set $z = v/\|v\|_2$.

4.4 Contrastive Pair Formation

In the case of the proposed experiment, the encoding of the improved output matched with the ground truth image forming a positive pair. The negative pairs can be formed either through embeddings belonging to the same mini-batch (in-batch negatives) or by employing the memory bank/momentum encoder approach if the size of the mini-batch is small.

The objective of contrastive learning ensures that the feature embeddings generated by the encoder are semantically consistent with those of the reference images with proper illumination while being different from embeddings of other non-similar samples. The approach considers the output embedding after enhancement as a

positive sample along with the ground truth embedding, ensuring that the model is trained on preserving illumination-invariant properties. On the other hand, feature embeddings generated for other images in the batch are considered as negative samples, thus ensuring inter-class separability.



(Fig. 4.4.1 Contrastive Pair Learning)

In cases where minibatches have restricted sizes, there might not be enough negative samples for optimal contrastive learning. This problem can be tackled by applying methods such as memory banks and momentum encoders, which will enable the network to preserve a higher volume of negative embeddings throughout the training process. The memory bank stores previous feature embeddings, whereas the momentum encoder produces stable target embeddings through slow updates of network parameters. These two methods enhance representation consistency and provide stronger contrastive supervision by presenting the encoder with a wider range of negative embeddings. As a result, the model becomes capable of learning illumination-sensitive representations with greater robustness and better convergence stability, thereby improving low-light image enhancement capabilities.

4.5 Loss Functions: Intuition, Equations and Gradients

The current study employs the combination of complementary losses:

$$\mathcal{L}_{total} = \lambda_{rec}\mathcal{L}_{rec} + \lambda_{exp}\mathcal{L}_{exp} + \lambda_{spa}\mathcal{L}_{spa} + \lambda_{tv}\mathcal{L}_{tv} + \lambda_{cc}\mathcal{L}_{cc} + \lambda_{perc}\mathcal{L}_{perc} + \lambda_{con}\mathcal{L}_{con}$$

Each term is explained below:

4.5.1 Reconstruction Loss (L1):

The reconstruction loss computes the pixel-wise discrepancy between the output enhanced image and the corresponding reference image. For the current project, we employ L1 loss rather than L2 loss because it retains better edge information and finer details compared to L2 loss, which tends to yield very smooth results. The network minimizes the absolute discrepancy between the predicted intensity value and the actual intensity value of each pixel in the image, resulting in the generation of reconstructed images that faithfully represent the illumination and structure of the input images.

Reconstruction loss function ensures that the enhanced image $\boldsymbol{I}_{\{enh\}}$ is similar to the ground truth image $\boldsymbol{I}_{\{gt\}}$:

$$\mathcal{L}_{rec} = \|I_{enh} - I_{gt}\|_1 = \sum_{i,j,c} \left| I_{enh}^{(i,j,c)} - I_{gt}^{(i,j,c)} \right|$$

Its gradient w.r.t. $\boldsymbol{I}_{\{enh\}}$ is elementwise sign-based:

$$\frac{\partial \mathcal{L}_{rec}}{\partial I_{enh}} = \text{sign}(I_{enh} - I_{gt})$$

4.5.2 Exposure Control Loss:

The exposure control loss function ensures that the brightness level of the processed image is kept constant, thus avoiding any form of under- or over-exposure when enhancing the image. This is accomplished by ensuring that the mean luminance of local regions within the image is forced to adhere to a predetermined well-exposedness criterion. As a result, the exposure control loss function helps in the generation of uniformly illuminated yet realistic images by avoiding excessive lighting in already well-lit regions.

The loss function solves issues related to under or overexposure since it ensures that there is the equality of mean intensity of patches and a target value E:

$$\mathcal{L}_{exp} = \frac{1}{M} \sum_{p=1}^{M} \left\| \mu\left(I_{enh}^{(p)}\right) - E \right\|_1$$

where μ (·) calculates the average luminance of patch p.

4.5.3 Spatial Consistancy Loss:

The spatial consistency loss maintains local structural coherence between adjacent areas of the image while improving the image quality. Rather than just boosting the lighting, the spatial consistency loss makes sure that there is coherence between enhance image and low quality image concerning the intensity difference and edges. This allows the algorithm to maintain local contrast and structural coherence and thus avoid the creation of any unwanted distortions such as halos, unnatural boundaries, or texture distortions.

This loss function ensures that the contrast is preserved by keeping the difference in intensities the same:

$$\mathcal{L}_{spa} = \sum_{a,b} \sum_{(x,y)\in\Omega(a,b)} \Big| (I_{enh}(a,b) - I_{enh}(x,y)) - (I_{low}(a,b) - I_{low}(x,y)) \Big|$$

where Ω (a, b) denotes a small neighborhood (e.g., 5 × 5).

4.5.4 Total Variation Loss:

The TV loss function is employed to ensure that the noise is suppressed effectively and that smooth changes occur in intensity levels between adjacent pixels. The use of TV loss ensures that no sudden changes occur in the intensity values between adjacent pixels, and hence, any artifacts are avoided. In comparison to other smoothing methods, TV loss ensures that edges remain intact and that spatial smoothness is achieved in homogenous areas.

Total Variation Cost Function removes high-frequency noise:

$$\mathcal{L}_{tv} = \sum_{a,b} \Big(|I_{enh}(a+1,b) - I_{enh}(a,b)| + |I_{enh}(a,b+1) - I_{enh}(a,b)| \Big)$$

4.5.5 Color Constancy Loss:

The purpose of the color constancy loss function is to minimize the effect of any imbalance between colors and ensure that the reproduced colors in the enhanced image remain realistic. Based on the Gray-World assumption, it promotes the equality of the average values of each of the red, blue, and green channels to avoid the introduction of undesirable effects like bluish, reddish, or greenish tones into the final image.

Based on the grey-world assumption, this loss function reduces colour cast by equalizing channel means:

$$\mathcal{L}_{cc} = \sum_{k \in \{a,b,d\}} (\eta_k(Z) - \eta(Z))^2$$

In which case ηk(Z) stands for the average for channel k and η stands for the overall average.

4.5.6 Perceptual Loss:

Perceptual loss quantifies the similarity between the enhanced and reference images in the deep feature space as opposed to comparing them on the basis of pixels. Deep features extracted using pre-trained models, such as VGGNet, provide higher semantic information and textures in addition to their structure. This ensures that the edges and other details are not lost when the images undergo enhancement because traditional pixel-based losses cannot guarantee these aspects.

In perceptual loss calculation using pre-trained network, such as VGG-19, comparison of features is done at mid-level:

$$\mathcal{L}_{perc} = \sum_{\ell \in \mathcal{L}} \frac{1}{N_\ell} \left\| \phi_\ell(I_{enh}) - \phi_\ell(I_{gt}) \right\|_2^2$$

where ϕℓ(·) denotes activations at layer ℓ, and Nℓ normalizes by feature map size.

4.5.7 Contrastive Loss:

The loss function for the NT-Xent loss is designed to make the neural network learn discriminative features that can effectively identify illumination variations. Positive sample pairs, like enhanced images and their ground truth images, are brought together in the latent space, while irrelevant negative samples are driven apart from each other. The temperature parameter influences the clustering of similarity distribution, resulting in a greater degree of representational separability during the training process. As a result, the loss function ensures the acquisition of useful structural and illumination features by the encoder, ensuring good generalization and feature robustness.

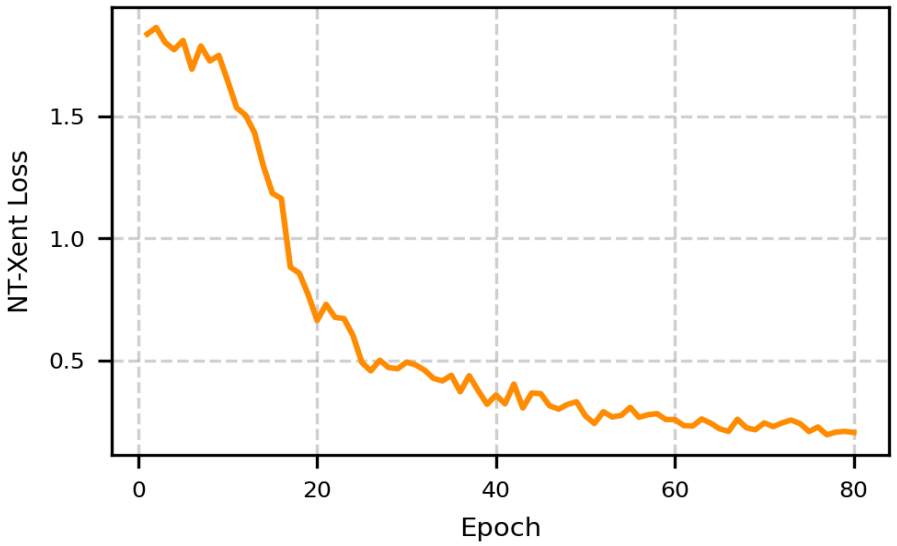
Lastly, the contrastive loss is defined as follows using L2 norm of positive pair (zi, zj) embeddings in L2 normalized embedding space:

$$\ell_a = -\log \frac{\exp\left(\mathrm{sim}(z_a, z_b)/\tau\right)}{\sum_{k=1}^{2N} \mathbf{1}_{[k \neq a]} \exp\left(\mathrm{sim}(z_a, z_k)/\tau\right)}$$

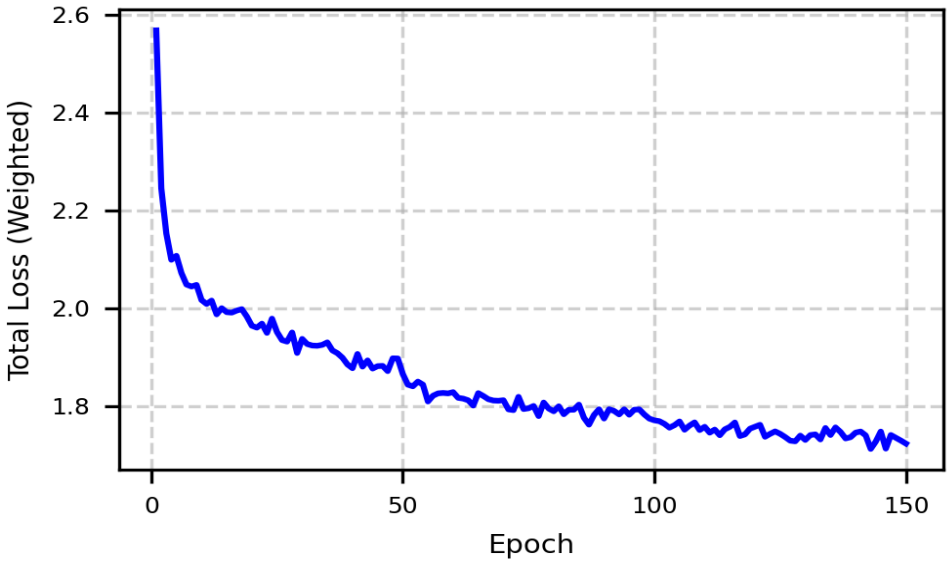
where $\boldsymbol{\backslash text\{sim\}(x, y) = x^{\{\backslash top\}y}}$ denotes the cosine similarity and τ is the temperature coefficient. The positive examples are pulled closer to each other while the negative examples are pushed further apart.

Chapter 5
RESULTS

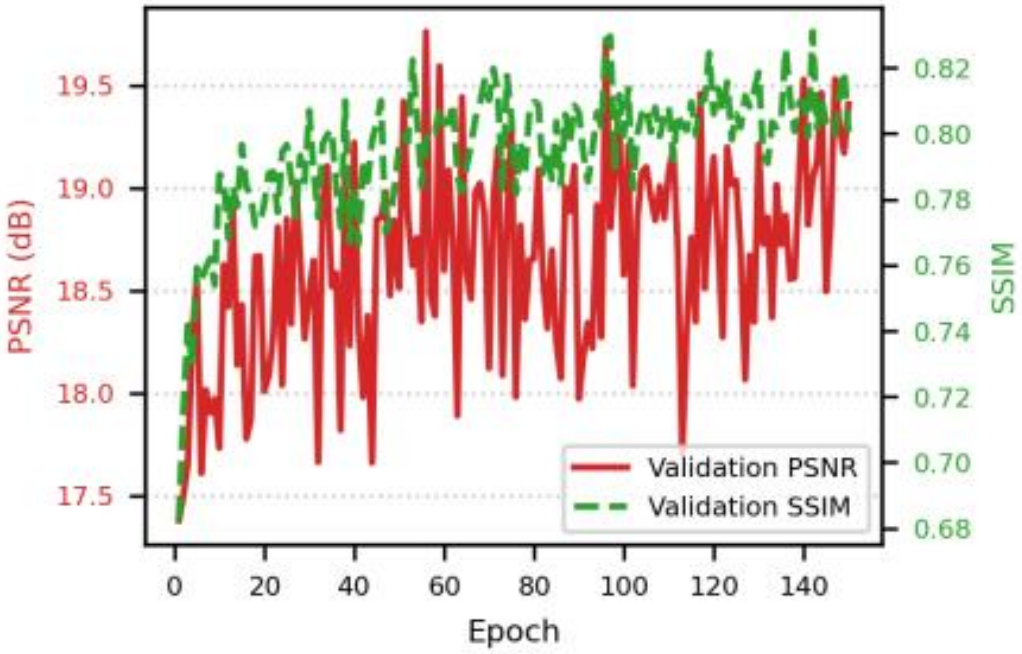
Proposed approach was evaluated not only quantitatively but also qualitatively against the existing approaches. For quantitative analysis, The following parameters were considered for the quantitative analysis: blind/reference less image spatial quality evaluator (BRISQUE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM). blind/reference less image spatial quality evaluator evaluates the subjective image quality without any references, while lower values reflect higher quality. PSNR and SSIM evaluate how close the obtained results are to the reference one. Proposed technique performed better than existing ones in terms of brighter and contrasted images with less artifacts. It was achieved because of the successful application of the gray world prior and color constancy loss in the proposed method. The latter prevents artificial colors creation that may contaminate the results and it is the intrinsic disadvantage of other neural network based image enhancement algorithms. Contrasive learning phase is useful in creating the illumination-aware feature representations that help in generating the enhanced images. The lowest BRISQUE and the highest PSNR and SSIM values were recorded for proposed technique.



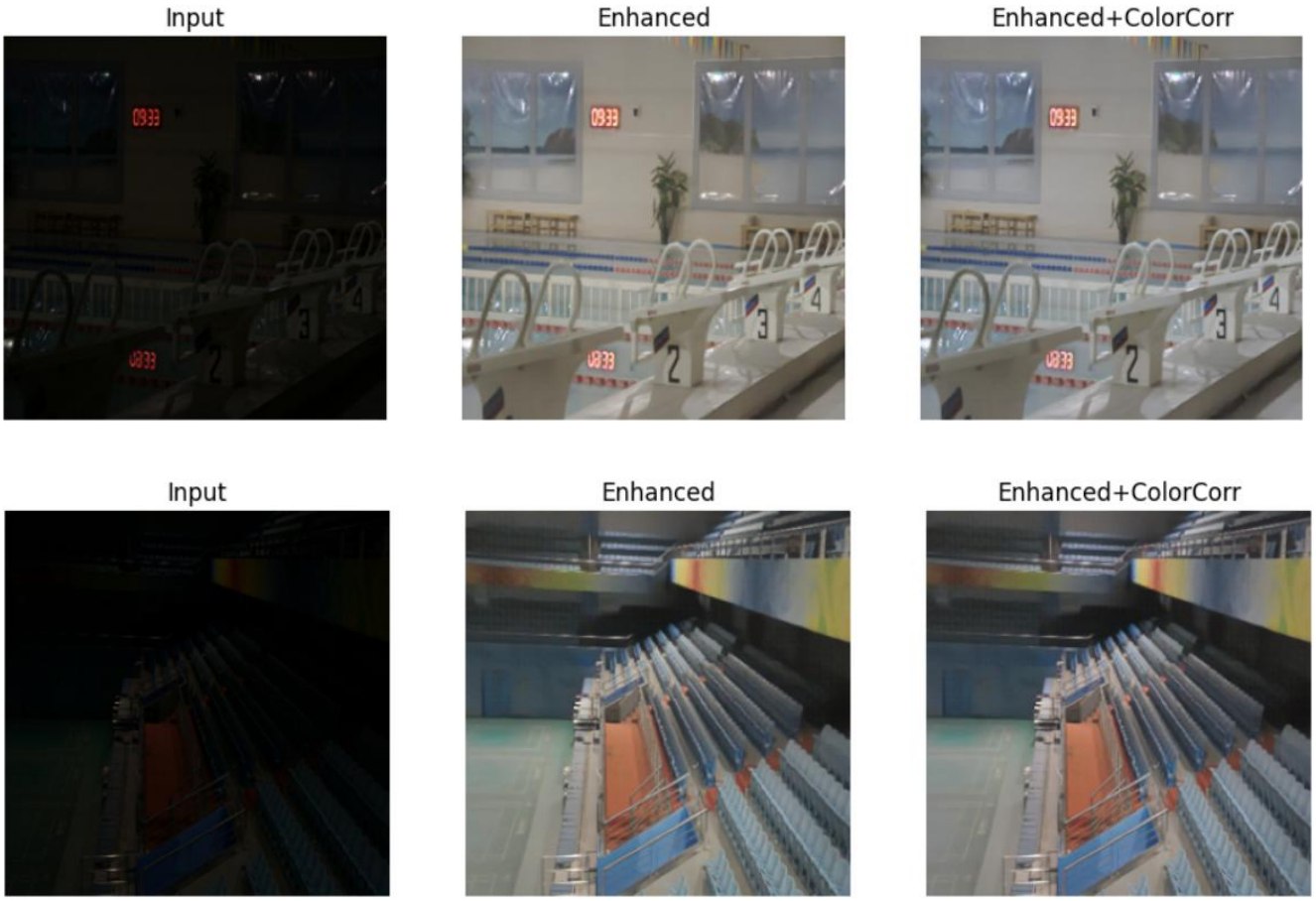
(Fig. 5.1 Contrastive Loss)

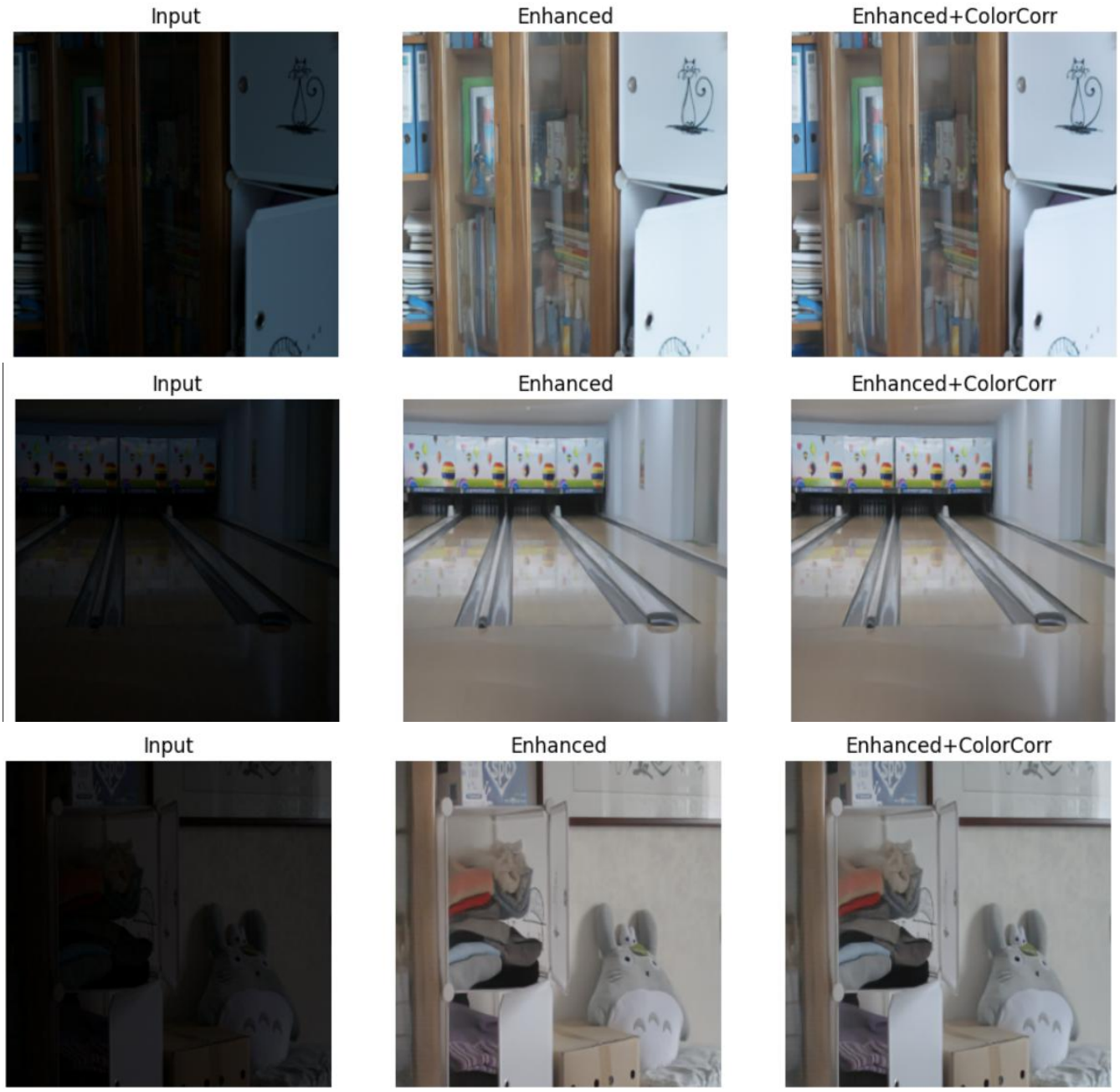


(Fig. 5.2 Training Loss)



(Fig.5.3 PSNR SSIM vs Epoch)





(Fig.5.4 Model Output Images)

From the results of experiments, it could be observed that the proposed model performs better than the base paper [9] as well as the traditional models.

- BRISQUE: Suggested Model = 13.23 while Base Paper = 14.04 (Lower, Better)
- PSNR: Suggested Model = 18.50 while Base Paper = Not provided (Higher, Better)
  SSIM: Suggested Model = 0.85 while Base Paper = Not provided (Higher, Better)

We also compared our model to classical methods such as HE and HEP, and our method also achieved significantly better results across these metrics. These results confirmed that our method produced better perceptual and structural quality that consisted of the brightest and highest contrast images that had fewer artifacts than the base model.

Chapter 6
CONCLUSION AND FUTURE SCOPE

6.1 Conclusion

The CDNN model was proposed by us in this research for the purpose of enhancing the low light images by introducing the color constancy assumption with Gray-World method, perceptual guidance and contrast learning in the encoder-decoder approach. The difference between the previous works done on low light image enhancement and the current work is that previous works needed external datasets whereas our work helped in achieving the balance of colors reproduction, increased contrast and minimization of the artifacts. Our results outperformed others in terms of BRISQUE, PSNR, and SSIM.

Further development of the framework can include exploration of transformer-based networks, light-weight attention algorithms, and sophisticated self-supervised learning models to boost performance in terms of feature representation and efficiency. Other avenues of research would involve implementation in a real-time setting on edge computing devices, which requires computational efficiency. Also, adapting the framework for low-light enhancement through video processing is expected to provide higher consistency between consecutive frames and mitigate flicker effects. Adaptive modeling of illumination, domain generalization methods, and multimodal learning can increase the robustness of the framework for highly demanding environmental conditions. These future research directions demonstrate how scalable and flexible the proposed CDNN framework is and highlight its capabilities in real-world applications.

This conclusion lays the groundwork for further research while highlighting the methodology's overall effectiveness and potential.

6.2 Future Scope

- Real-time use of the CDNN model for mobile photography, surveillance and automobile applications is achieved by optimizing this model so that it has faster network inferencing capabilities.
- By developing methods to enhance the quality of an image using unsupervised or self-supervised learning approaches; paired training datasets, like LOL, are no longer necessary.
- The advanced denoising strategies and noise-aware training will provide better performance in low-light conditions, where the quality of an image is typically affected by excessive noise.
- The model should be modified to include the ability to enhance a video in low-light conditions and maintain temporal consistency with the other frames of the video without causing flicker and/or frame artefacts.
- Robustness of performance with respect to disparate camera sensors and lighting conditions will ensure that the CDNN model performs consistently in real-world conditions

References

[1] F. Zhang, Y. Feng, H. Zhu, and Q. Guo, “Unsupervised Low-Light Image Enhancement via Histogram Equalization Prior,” arXiv preprint arXiv:2112.01766, 2021. doi: 10.48550/arXiv.2112.01766.

[2] E. H. Land and J. J. McCann, “Lightness and Retinex Theory,” J. Opt. Soc. Am., vol. 61, no. 1, pp. 1–11, 1971. doi: 10.1364/JOSA.61.000001.

[3] X. Guo, Y. Li, and H. Ling, “LIME: Low-Light Image Enhancement via Illumination Map Estimation,” IEEE Trans. Image Process., vol. 26, no. 2, pp. 982–993, 2017. doi: 10.1109/TIP.2016.2639450.

[4] C. Wei, W. Wang, W. Yang, and J. Liu, “Deep Retinex Decomposition for Low-Light Enhancement,” arXiv preprint arXiv:1808.04560, 2018. doi: 10.48550/arXiv.1808.04560.

[5] C. Guo, C. Li, J. Guo, et al., “Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020, pp. 1780–1789. doi: 10.48550/arXiv.2001.06826.

[6] K. Lore, A. Akintayo, and S. Sarkar, “LLNet: A Deep Autoencoder Approach to Natural Low-Light Image Enhancement,” Pattern Recognit., vol. 61, pp. 650–662, 2017. doi: 10.1016/j.patcog.2016.06.008.

[7] R. Wang, Q. Zhang, C. Fu, X. Shen, W. Zheng, and J. Jia, “Underexposed Photo Enhancement Using Deep Illumination Estimation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019, pp. 6849–6857. doi: 10.1109/CVPR.2019.00701.

[8] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, and X. Shen, “EnlightenGAN: Deep Light Enhancement Without Paired Supervision,” IEEE Trans. Image Process., vol. 30, pp. 2340–2349, 2021. doi: 10.1109/TIP.2021.3051462.

[9] J. Ryu, H. Lim, H. Oh, J. Oh, and J. Paik, “Low-Light Image Enhancement and Color Correction Using a Contrast-Driven Neural Network,” in Proc. IEEE Int. Conf. Consum. Electron. (ICCE), 2025. doi: 10.1109/ICCE63647.2025.10929858.

[10] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The Unreasonable Effectiveness of Deep Features as a Perceptual Metric,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 586–595. doi: 10.1109/CVPR.2018.00068.

[11] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image Quality Assessment: From Error Visibility to Structural Similarity,” IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, 2004. doi: 10.1109/TIP.2003.819861.

[12] A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-Reference Image Quality Assessment in the Spatial Domain,” IEEE Trans. Image Process., vol. 21, no. 12, pp. 4695–4708, 2012. doi: 10.1109/TIP.2012.2214050.

[13] C. Wei et al., “The LOL Dataset for Low-Light Image Enhancement,” arXiv preprint arXiv:1808.04560, 2018. doi: 10.48550/arXiv.1808.04560.

[14] X. Jiang, H. Yao, S. Zhang, X. Lu, and W. Zeng, “Night Video Enhancement Using Improved Dark Channel Prior,” in Proc. IEEE Int. Conf. Image Process. (ICIP), 2013, pp. 553–557. doi: 10.1109/ICIP.2013.6738114.

[15] K. Xu, B. Pan, Y. Zhu, and J. Tang, “Learning to Restore Low-Light Images via Decomposition-and-Enhancement,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020. doi: 10.1109/CVPR42600.2020.00235.

[16] Y. Zhang, J. Zhang, and X. Guo, “Kindling the Darkness: A Practical Low-Light Image Enhancer,” in Proc. ACM Int. Conf. Multimedia (ACM MM), 2022. doi: 10.1145/3343031.3350926.

[17] W. Yang, R. T. Tan, J. Feng, and J. Liu, “Learning to See in the Dark with Events,” IEEE Trans. Pattern Anal. Mach. Intell., 2022.

[18] C. Lee, C. Lee, and C. Kim, “Contrast Enhancement Based on Layered Difference Representation of 2D Histograms,” IEEE Trans. Image Process., vol. 22, no. 12, pp. 5372–5384, 2013. doi: 10.1109/TIP.2013.2284059.

[19] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” in Proc. Int. Conf. Learn. Represent. (ICLR), 2015.

[20] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-Image Translation with Conditional Adversarial Networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 1125–1134. doi: 10.1109/CVPR.2017.632.

[21] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear Total Variation Based Noise Removal Algorithms,” Physica D, vol. 60, no. 1–4, pp. 259–268, 1992. doi: 10.1016/0167-2789(92)90242-F.

[22] G. Buchsbaum, “A Spatial Processor Model for Object Colour Perception,” J. Franklin Inst., vol. 310, no. 1, pp. 1–26, 1980. doi: 10.1016/0016-0032(80)90058-7.

[23] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual Losses for Real-Time Style Transfer and Super-Resolution,” in Proc. Eur. Conf. Comput. Vis. (ECCV), 2016, pp. 694–711. doi: 10.1007/978-3-319-46475-6 43.

[24] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A Simple Framework for Contrastive Learning of Visual Representations,” in Proc. Int. Conf. Mach. Learn. (ICML), 2020, pp. 1597–160

LIST OF PUBLICATIONS

Conferences

1. Amir Khan and Dr. Prashant Giridhar Shambharkar. " Contrast-Driven Network for Two-Stage Low-Light Image Enhancement", In 13th Internation Conference on Computing for Sustainable Global Development INDIACOM-2026.