# Brightness-Aware Neural Adaptation: A Hybrid Approach for Time-Dependent Low-Light Image Enhancement

by

Nafees Raiyan Rahman
21301315
Saumya Deep Kanungo
21201201
Radiah Hassan
22101002
Rabaya Farzana Sikder
22101728

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University
October 2025.

# Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

**Student's Full Name & Signature:**


**Nafees Raiyan Rahman**
21301315

**Saumya Deep Kanungo**
21201201

**Radiah Hassan**
22101002

**Rabaya Farzana Sikder**
22101728

# Approval

The thesis/project titled "Brightness-Aware Neural Adaptation: A Hybrid Approach for Time-Dependent Low-Light Enhancement" submitted by

1. Nafees Raiyan Rahman (21301315)
2. Saumya Deep Kanungo (21201201)
3. Radiah Hassan (22101002)
4. Rabaya Farzana Sikder (22101728)

of Fall, 2024 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science in February 1, 2025.

**Examining Committee:**

Supervisor:
(Member)

Dr. Md. Ashraful Alam
Associate Professor
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

Dr. Md Golam Robiul Alam
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi
Associate Professor
Department of Computer Science and Engineering
Brac University

# Abstract

Low-light image enhancement (LLIE) is essential for improving the visibility and interpretability of images captured under poor or insufficient lighting conditions, which is critical for applications such as autonomous driving, surveillance, and photography. However, most existing supervised approaches depend on large paired datasets that are expensive to collect and often fail to generalize across diverse

lighting conditions and degradations. In this thesis, we propose a self-supervised learning approach for LLIE using a Time-Aware Zero-Reference Deep Curve Estimation (Zero-DCE) model. Our method eliminates the need for paired training data by learning to enhance images through a self-supervised mechanism that estimates pixel-wise adjustment curves directly from the input. This design makes the model lightweight, efficient, and suitable for real-time enhancement tasks.
We gathered a custom dataset in order to train and test our model. Pictures were taken between 3 a.m. and 9 p. m. with five minutes intervals to permit the model to acquire knowledge by constant changes of lighting during the day. By including the awareness of time in the self-supervised process of learning, the model adjusts its improvement strategy to the changing light conditions, which leads towards an improvement in luminance, contrast and hue stability and maintain structural acuity. The proposed framework demonstrates strong potential for real-world deployment in settings where labeled data are limited and lighting is suboptimal, including autonomous vehicles, surveillance systems, and underwater imaging.

**Keywords:** Image Enhancement; Computer Vision; Low Light Image Enhancement; Self Supervised Learning; Temporal Image Dataset; Deep Learning

# Acknowledgement

Firstly, all praises are made to the Great Almighty for whom our thesis has been completed without any major interruption.

Secondly, to our supervisor Mohammad Ashraful Alam Sir for his kind support and advice in our work. He helped us whenever we needed help.

And finally to our parents, without their support throughout, it was not possible. With their kind support and prayer, we are now on the verge of our graduation.

# Table of Contents

# List of Figures

# Chapter 1

# Introduction

## 1.1 Background

Technology has become an important part of computer vision in current times and, with its use, many sectors have seen improvement, including restoration of an image. In real life, an image often suffers from degradation in poor weather and in low-light environments and poses an issue in terms of viability and interpretation. Traditional supervised techniques are based on large datasets with labels, and preparing them is a lengthy and costly process. With the advent of self-supervised learning (SSL), models can be developed that can learn useful representations with unlabeled datasets.

Self-supervised learning is defined as a machine learning paradigm in which models will learn representations from unlabeled data by pseudo-labeling via pretext tasks. Unlike supervised ones, SSL (self-supervised learning) does not rely on any manually annotated ground truth, which makes it particularly appealing for domains with limited or hard-to-obtain high-quality labels.//
Low-light image enhancement involves improving visibility and contrast within images captured with poor lighting. This task is crucial for critical applications such as autonomous driving, surveillance, and photography.The goal of low-light image enhancement is primarily to improve visibility and remove degradation caused by insufficient lighting, enabling clearer perception and analysis of scenes captured in dark or poorly lit environments.

In the following sections, we review existing research on self-supervised learning for low-light image enhancement, analyzing their methodologies and limitations. First, we collect and analyze a range of papers that used different model architectures, some training paradigms leveraging unannotated data for effectively restoring images. Again, the approaches have widely employed data augmentation methods for more robust models. We then outline a number of the main pre-processing methods on image datasets, including enhancing contrast, removing noise, and extracting features. Finally, we discuss various methodologies that have been implemented, including transformer-based networks, variants of CNN, and hybrid architectures integrating recurrent modules. These are evaluated based on the performance of enhancing image quality under challenging conditions with a particular focus on generalization and real-world applicability.

## 1.2 Rational of the Study or Motivation

Supervised image enhancement involves a large scale dataset of paired images of degraded and clean pairs which are mostly difficult and expensive to obtain under dynamic environmental conditions. Besides that the synthesized data through artificially simulated conditions often lack good generalization for real scenes, hence this limits the supervised model performance for practical use .This calls for more efficient, adaptable, and data efficient learning method. In this direction, self-supervised learning has emerged as a promising alternative, mainly leveraging intrinsic structures, constraints, and knowledge from unpaired or unlabeled data. It relaxes the need for manually annotated datasets by relying on various techniques, such as contrastive learning, cycle-consistency, or generative modeling, to learn robust representations. Given an effective self-supervised objective, a model is able to extract meaningful information from raw, unpaired degraded images and progressively restore visibility and details. This approach improves generalization while also aligning well with the real-world application of the techniques, such as autonomous driving, surveillance, and medical imaging, where it is impractical to acquire clean reference images. Self-supervised methods can be further trained on large-scale real-world datasets without explicit supervision, thus being more scalable and adaptable to diverse degradation patterns. With growing demands for real-time, high-quality image enhancement in bad environmental conditions, our research focuses on exploring innovative self-supervised learning techniques for low-light image enhancement. We are motivated to develop a framework that learns effectively from unpaired data, advancing not only the field of image enhancement but also some of the most fundamental challenges in deep learning. The work presented here creates the prospect of significant advances in performance in computer vision in critical applications, through the assurance of reliable perception and decision-making even in visually degraded environments.

## 1.3 Problem Statement

Bad weather like foggy, rainy, and snowy conditions, combined with low light, generally causes poor visibility, noise, and color distortions in images. Most traditional methods for image enhancement depend on supervised learning approaches that require paired clean and degraded image datasets, which are not always available and generalize poorly to real-world scenarios. This greatly limits the performance of computer vision applications in these critical areas, such as autonomous driving, surveillance and photography. Self-supervised learning is a promising alternative that depends on unlabeled or weakly labeled data for the learning of robust feature representations without requiring any ground truth images. However, existing self-supervised methods poorly preserve fine details, remove complex noise, and adapt to varied environmental conditions. Therefore, in order to achieve better image quality, enhanced visibility, and improvement in downstream tasks' performance, there is an essential need for an effective self-supervised learning approach toward the enhancement of low-light images.

## 1.4 Objective

- It is expected that enhancing images captured in poor lighting will significantly improve visibility, making them suitable for human observation and automated systems.
- Autonomous vehicles, drones, and surveillance systems rely on clear visuals for decision-making. Enhancing such degraded photos ensures improved performance at night or in severe conditions.
- Most deep learning and image processing applications include object identification, facial recognition, and medical imaging using low-quality inputs. Enhancing these degraded photos increases the precision of such models.
- Most security cameras operate in low-light or inclement weather.This enhancement will improve monitoring, crime detection, and public safety.
- Images taken in fog, rain, or darkness may lose some important information. Enhancement techniques are used to recover lost information in forensic investigations and remote sensing applications.
- Image enhancement can help photographers and filmmakers improve the aesthetic appeal of their work while preserving details without the use of artificial lighting.

## 1.5 Scopes and Challenges

The scopes of our research are:

- Our research addresses imagine enhancement methods using self-supervised learning (SSL) approaches, which do not require extensive annotated datasets.
- The study focuses on the difficulties of improving images that are impacted by bad lighting and unfavorable meteorological circumstances, such as fog, rain, and snow.
- The findings may contribute to advancements in areas like autonomous driving, surveillance, and medical imaging where clear visual data is crucial.

The challenges for our research are:

- As SSL does not rely on labeled datasets, evaluating model performance without high-quality ground truth can be difficult.
- The complexity of different low-light and weather conditions (e.g., dense fog vs. light mist) makes generalization challenging.
- Training deep learning models for SSL-based enhancement can be resource-intensive, requiring high computational power.

# Chapter 2

# Literature Review

## 2.1 Introduction

Imaging under low-light conditions has always been difficult, and it still is. The situation has made the use of image enhancement techniques unavoidable. We have performed a comprehensive literature review to clarify the situation in this field and to indicate possible future research directions. The review ranged from basic classical algorithms to high-end deep learning methods. This chapter first presents the wide-ranging review literature as the foundation and context for the issues and then goes on to discuss the development of low-light image enhancement methods by incorporating advanced technologies such as Retinex-based decomposition, multiscale convolutional networks, and transformer blocks. By combining the knowledge from each work together, we identify the key milestones along with the non-resolved issues and the emerging trends at the same time that these factors influence the direction of imaging enhancement research.

## 2.2 Review Papers

Researchers Gursharn Singh and Anand Mittal [2] went through and reviewed various image enhancement techniques that focused on spatial and frequency domain methods. Spatial domain methods like power law transformations and histogram equalization worked well for contrast enhancement and for directly manipulating pixel density. Frequency domain methods, including Discrete Wavelet Transformation (DWT) and Discrete Cosine Transform (DCT), aimed to improve image clarity. While comparing these two techniques, the spatial techniques are very simple to implement but cannot enhance specific parts of an image selectively, while frequency techniques faced difficulty in automating universal enhancement and handling severely low-light areas of an image, even though they excelled in effectively manipulating image components.

Researcher Wonjun Kim reviewed [10], categorized, and provided a systematic taxonomy of multiple LLE (Low-Light Enhancement) methods, dividing them into learning-based and feature-based approaches. From his research, he found that models that are handcrafted, like Gamma correction and histogram equalization, struggled with uneven illumination. On the other hand, learning-based models such as Zero-DCE and RetinexNet worked well due to their adaptability and capability

to cope with the complex mappings between properly exposed images and low-light images. He also emphasized the relevance of learning models that do not rely on reference training sets; these methods do not need a specific labeled dataset for training. This highlighted the strengths of models like SCL-LLE, which can further advance learning-based approaches by incorporating semantic guidance for enhancement in targeted places. Despite the progress, real-life implementations and scalability still remain challenging, as they rely heavily on extensive labeled datasets.

Researchers Wencheng Wang, Xiaojin Wu, Xiaohui Yuan, and Zairui Gao reviewed seven types of low-light image enhancement techniques [6], ranging from machine learning models and Retinex-based methods to frequency domain approaches. They found that Retinex-based methods like RetinexNet worked better for detail recovery because these methods address the illumination and reflectance components of an image. Meanwhile, machine learning methods using CNNs helped automate the enhancement process further. From their research, it was identified that, for region-specific enhancement, histogram equalization methods were lacking. Deep Neural Network approaches, however, effectively balanced local and global enhancements. Their review served as a foundation for advanced models like IA-YOLO and SCL-LLE, which address issues like uneven lighting and semantic understanding that were identified in this work.

## 2.3 Papers on Low Light Image Enhancement and Noise Cancellation

In "Semi-Supervised Learning for Low-light Image Restoration through Quality-Assisted Pseudo-Labeling," Malik et al. [13] leverage a no-reference quality assessment algorithm to generate pseudo-labels from unlabeled dark images. This Quality-Assisted Pseudo-Labeling (QAPL) guides the restoration model and reduces the demand for extensive paired training sets. While results show enhanced image quality over purely supervised methods, model performance depends heavily on the accuracy of the pre-trained QA algorithm, which may not detect unique distortions or severe noise.In low-light environments, image capturing is very poor and poses a significant challenge for detection tasks due to poor quality and color adjustment, making it difficult to identify objects clearly. No-Reference Quality Assessment (NRQA) models have been used for low-light image restoration in a self-supervised learning framework. They use large amounts of unlabeled images to learn features from various restored duplicate versions. An Low-Light Image Restoration(LLIR) network designed based on a QA method is also employed, comparing the snowy image with the learned features of the model. The LLIR model performs well, owing first to well-behaved statistics in the form of sparse coefficients and second to a larger signal-to-lower noise ratio in the low-frequency sub-bands, allowing for effective enhancement. One limitation is that a pre-trained NR QA algorithm may not detect distortions throughout the LLIR process. Additionally, contrastive loss expands the commonalities between features in positive pairs while decreasing the commonalities between features in negative pairs.

Another recent publication titled "A Survey on All-in-One Image Restoration: Taxonomy, Evaluation and Future Trends" by Jiang et al. [18] presents a taxonomy of techniques aimed at handling multiple degradations (e.g., noise, haze, blur) simultaneously. This survey highlights that while all-in-one frameworks can simplify deployments, they also encounter conflicting objectives when trying to remove noise, enhance edges, and correct color simultaneously. The authors advocate for modules like mixture-of-experts or prompts that can dynamically adapt to the type and degree of image degradation, although they note the scarcity of robust real-world datasets encompassing diverse concurrent distortions.

In computer vision, low-light image enhancement is a crucial since images often suffer from low contrast and brightness issues that adversely affect subsequent tasks like object identification and scene understanding.To overcome this, researchers Liang Shen, Zihan Yue, Fan Feng, Quan Chen, Shihao Liu, and Jie Ma proposed a low-light image enhancement model, MSR-Net, based on a convolutional neural network and Retinex theory [3]. They conceptualize low-light image enhancement as a supervised learning problem with input output pairs consisting of dark and bright images. MSR-Net incorporates three key components:

1. Multi-Scale Logarithmic Transformation, which enhances the image using multiple logarithmic transformations;
2. Difference-of-Convolution, similar to the Difference-of-Gaussian (DoG) in traditional Retinex methods but learns parameters directly from data;
3. Color Restoration Function, which gets rid of color distortions to ensure natural output.

MSR-Net has demonstrated superiority over the cutting-edge approaches including MSRCR, LIME, Dong's method, and SRIE in synthetic datasets, characterized by higher SSIM and lower NIQE values. It has also led to better discrete entropy and visual quality scores in real datasets and won over others in detail having the most natural appearance besides being the most preserved. However, one of the drawbacks of the MSR-Net approach is that its small receptive field can cause halo effects in regions having smooth characteristics such as the sky.

The enhancement of low-light images can easily lead to issues such as colors being distorted, large shadow areas, and halo artifacts; hence, the quality of the image may not be good. Wei, Xinxu, Zhang Xianshi, Wang Shisen, Huang Yanlin, Yang Kaifu, and Li Yongjie have written a paper sanctioning these problems by suggesting the Two-Stage Network with Channel Attention (TSN-CA) model [8]. The TSN-CA works in a two-stage manner. To start with, it first makes the image brighter in the HSV color space while keeping the detail by working on the H and S channels. Secondly, it cuffs the Channel Attention (CA) modules into a U-Net to carry out the RGB space restoration of degraded images, having selectively furnished the most significant features and tackled the residual degradation problems like shadows and halos. TSN-CA shows the greater PSNR and SSIM achievements in the tests of various datasets, thoroughly getting rid of shadow blocks, halo artifacts, and lingering noise. Yet, at times, the restoration stage had problems with image under- or overfitting when using images with very low quality, leading to inconsistent detail recovery.

Inconsistencies in detail recovery were one of the issues that prompted another paper written by Xinxu Wei, Shisen Wang, Xianshi Zhang, Cheng Cheng, Yanlin Huang, Kaifu Yang, and Yongjie Li where the Degradation-Aware Deep Retinex Network (DA-DRN) is introduced [9]. This paper is a partial solution to the problem of inconsistent detail recovery through the incorporation of a Degradation-Aware (DA) Module. In fact, DA-DRN has made great improvements in the area of Retinex-based decomposition due to the development of a deeper U-Net architecture specifically designed for separating reflectance and illumination. The DA Module basically plays a dual role here, as it guides the decomposition network to completely avoid the introduction of noise and to deal with degradation quite easily during the decomposition phase. The model is shown to achieve better results than other methods in terms of PSNR and SSIM on both synthetic and real-world datasets; it is able to reduce noise and at the same time preserve details, all this while keeping fast computation speeds. Even with the mentioned improvements, DA-DRN still occasionally falls short, as it leads to a minor leakage of noise into the illumination map and the loss of some fine details during the enhancement process.

In computer vision, low-light image enhancement is crucial as images often suffer from poor visibility, low contrast, and corruptions like noise and color distortion. One approach by Cai et al. [12], Retinexformer, addresses issues like noise leakage and detail loss effectively. Retinexformer is a Transformer-based approach built on a revised Retinex theory, employing an Illumination-Guided Transformer (IGT) that captures long-range dependencies and models interactions between differently lit regions. Results show that Retinexformer suppresses noise effectively and preserves details, achieving accurate reconstruction. However, reliance on the Transformer architecture can make it computationally intensive for very large images compared to simpler methods.

A paper by Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong introduces Zero-Reference Deep Curve Estimation (Zero-DCE) [5], which aims to improve images taken in low light. Zero-DCE defines light enhancement as an image-specific curve estimation task within a lightweight seven-layer convolutional neural network called DCE-Net. This network approximates a pixel-wise high-order curve that adjusts the dynamic range without requiring paired or unpaired data during training. Zero-DCE is lightweight, achieving real-time enhancement at up to 500 FPS on a single GPU for images that are sized $640\times480\times3$, and it generalizes to varied luminance conditions. It outperforms complex models in PSNR, SSIM, and perceptual metrics. However, it lacks semantic context-based enhancement, explicit noise reduction capabilities, and has not been tested for video enhancement.

Zijia Yu, Yangyang Guo, Liyuan Zhang, Yi Ding, Gan Zhang, and Dongyan Zhang [22] address the challenge of detecting dairy cows in low-light conditions, where noise and detail loss reduce visibility. They introduce an improved Zero-DCE model with an upsampling and downsampling architecture for noise suppression and a self-attention gate for selecting task-relevant features. A kernel selection module fuses features at multiple scales, and the iterative curve adjustment process is replaced by

the Attentive Convolutional Transformer (ACT) module, enhancing image quality without artifacts. This approach integrates with object detection systems such as YOLOv5, CenterNet, EfficientDet, and YOLOv7-tiny to achieve higher detection precision, demonstrating real-time performance at 346.5 FPS. Limitations include poor performance in extremely noisy or very dark regions and with overlapping objects. Sensitivity to dataset quality and the lack of tests on video sequences remain open challenges.

The paper by Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo introduces the Kindling the Darkness (KinD) model for low-light image enhancement [4]. Inspired by Retinex theory, KinD decomposes an image into illumination (light) and reflectance (degradation) components. The proposed method adopts a three-stage pipeline:

1. Layer Decomposition Network, which decomposes the input image into its illumination and reflectance layers without explicit ground truth;
2. Reflectance Restoration Network, which enhances the reflectance layer by mitigating noise and color distortions using a structural similarity-based and gradient-based loss;
3. Illumination Adjustment Network, a controllable, user-tunable approach outperforming state-of-the-art gamma correction methods.

This architecture resembles a U-Net, running VGA-resolution images in under 50 ms on a GPU. KinD outperforms Retinex-Net and MSR on brightness adjustment, artifact suppression, and detail preservation, achieving higher PSNR, SSIM, and NIQE scores. However, it relies on paired datasets with changing light conditions, is less effective in severely degraded regions and dynamic lighting, and does not address video enhancement or temporal coherence.

Researchers Dong Liang, Ling Li, Mingqiang Wei, Shuo Yang, Liyan Zhang, Wenhan Yang, Yun Du, and Huiyu Zhou developed the Semantically Contrastive Learning for Low Light Image Enhancement (SCL-LLE) framework [11], focusing on multitask learning by combining contrastive learning with semantic consistency. This approach enhances brightness and texture while leveraging unpaired datasets containing positive (well-lit) and negative (underexposed) samples. SCL-LLE performs well in maintaining texture and color consistency under extreme darkness, surpassing earlier LLE methods across six datasets by utilizing semantic information. However, it requires extensive datasets containing semantic ground truths, making it challenging to generalize for everyday use.

Although transformer-based models like Retinexformer address noise, color distortion, and detail loss in low-light enhancement, they can be computationally intensive for large images. To tackle this, two recent lightweight approaches, Bratenanu et al. and Li et al., LYT-Net [16] and DARK [19], aim to achieve comparable performance with reduced computational costs. LYT-Net employs a dual-path approach in the YUV color space to separate chrominance and luminance, preserving details while reducing noise via a Channel-Wise Denoiser (CWD) and Multi-Stage Squeeze and Excite Fusion (MSEF) blocks. In contrast, DARK introduces a Retinex-based Illumination Estimator and Simplified Residual Contextual Blocks (SRCB) to re-

duce complexity while enhancing illumination and preserving details. LYT-Net outperforms state-of-the-art methods on the LOL dataset in PSNR and SSIM with only 0.045M parameters, whereas DARK achieves competitive PSNR and SSIM scores with 0.187M parameters, enabling real-time application on standard hardware. However, LYT-Net may not match the fidelity of heavier models like Retinexformer in extreme scenarios, and DARK sometimes suffers from minor artifacts and reduced color fidelity compared to larger models like MIRNet-v2.

The paper "Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering" [1] by Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian offers an approach to image denoising, removing noise while preserving fine details. It employs a sparse 3D transform-domain representation by grouping similar 2D patches and applying collaborative filtering. The technique includes block matching, spectral shrinkage, and Wiener filtering using discrete cosine, sine, and wavelet transforms for sparsity. It outperforms wavelet-based denoising, PCA, and nonlocal means, rivaling sophisticated methods like BLS-GSM and K-SVD. However, it may group patches incorrectly in high-noise scenarios, and the computational cost can be high for large images.

To overcome problems identified in earlier papers, Shuang Wang and Qingchuan Tao present ResVMUNetX [21], an image enhancement network using an error-regression method to directly improve brightness and recover structural details by adding estimated illumination supplements. Combining direct pixel addition with a customized Denoise CNN module, ResVMUNetX employs an effective VMamba architecture for brightness enhancement, structural detail restoration, and noise elimination. The network reduces noise with the DenoiseCNN module, enhances brightness via VMUNetX, and achieves real-time processing speeds up to 70 FPS, significantly improving clarity and image quality. However, real-time processing is limited to GPU devices, restricting deployment on edge devices. Further testing is needed to ensure generalization to diverse real-world scenarios.

## 2.4 Papers on Image Restoration from Adverse Weather Conditions

Working on enhancing images disrupted by bad weather, researchers Jiaxiu Chang, Wenshuai Hou, Wenjing Chen, Jun Yan, and Kaige Cui proposed the Image Adaptive YOLO (IA-YOLO) [17] framework for improved object detection under adverse weather conditions. This framework integrates a Differentiable Image Processing (DIP) module with a CNN to dynamically enhance images prior to detection. The DIP module applies filters for gamma correction, defogging, sharpening, and contrast enhancement, all optimized in real time. Coupled with the YOLO v3 detection system, the model simultaneously improves detection accuracy and image quality. However, in real-time applications , balancing computational efficiency with enhancement quality remains challenging.

Researchers Mohammad Shabaz and Mukesh Soni have developed a weather recogni-

tion model that incorporates deep learning methods and traditional spatial features, which mainly focusses around the recognition of different types of weather conditions and visibility improve-ment [21]. They utilized VGG16 for deep feature extraction and then combined those features with shallow ones like contrast and saturation to train a SoftMax classifier. Their method was successful in recognizing weathers like snow, rain, and fog, by achieving an accuracy of 99.26
TransWeather [10], which was presented by Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M. Patel, is a transformer-based technique that restores images that have been degraded by bad weather conditions such as rain, fog, and snow. It does not treat each weather condition separately, but rather, it delivers a general solution through the implementation of an encoder-decoder architecture, which makes it very suitable for real-life applications like autonomous navigation and surveillance. The encoder employs an intra-patch transformer block for the purpose of significantly improving the feature extraction process, while the weather-type queries direct the decoder toward the restoration of the desired weather condition. TransWeather maintains an excellent balance between task-specific and generalized image restoration, which consequently leads to a situation where it becomes the best in terms of state-of-the-art methods with respect to multiple benchmark datasets.. Nonetheless, it can struggle with extreme weather, and its reliance on synthetic training data may limit adaptability to varied real-world conditions. Additionally, even optimized transformer architectures may pose challenges for real-time, low-power deployment.

A research paper of Xuelong Li, Kai Kou, and Bin Zhao presents the Weather GAN [7], a framework based on Generative Adversarial Networks that is intended for weather image translation, i.e., changing images taken in one weather condition to another (for instance, sunny to cloudy, cloudy to snowy). In contrast to classical techniques which aimed at removing rain or enhancing fog as the specific weather phenomenon, Weather GAN can handle various weather conditions at the same time. The generator of the GAN has an initial global translation module, an attention module for localized edits, and a weather-cue segmentation module that represents and distributes the weather elements' structure and distribution. The technique surpasses CycleGAN, UNIT, and MUNIT by obtaining lower FID and KID scores along with visual quality that is more realistic. Despite this, it has difficulties with intricate scenes where both the weather conditions and the intensity of their obstructions (heavy snow or fog) are complex, which may in turn cause the output to be fuzzy or distorted. Generalization to new and unseen weather conditions continues to be a challenge too, and it is mainly attributed to weakly supervised training.

Recent work by Xu et al. [23] explores vision-language models for real-world adverse weather restoration, leveraging large-scale image-text pretraining to preserve semantic details during defogging or deraining. This approach focuses on clarifying not just visual aspects but also semantic content relevant to tasks such as scene understanding or object identification. While preliminary performance gains are promising, training and inference are computationally expensive, and scalability to fully unstructured outdoor scenes with multiple forms of weather distortion requires further validation.

WeatherDepth, a proposed model by Wang et al. [20] introduces curriculum contrastive learning to iteratively fine-tune depth estimation under progressively more severe weather conditions, starting from light drizzle to heavier rain or fog. This approach builds robust representations handling a broader range of climate-induced disruptions. Nonetheless, domain gaps between synthetic training data and complicated real-world scenes complicate direct application to scenarios like heavy storms or swirling snow.

A different approach taken by Zhu et al. [15] proposes a comprehensive system that can identify and represent both general weather features (for instance, extensive color corrections) as well as particular weather-driven attributes (like raindrops or haze). Usually employing U-Net variations, these frameworks divide the neural network into common and specialized components, thus resulting in higher PSNR and SSIM scores in mixed-weather datasets. At the same time, the increased number of parameters and the unverified real-time performance on edge hardware are the limitations to widespread use of such solutions in the area with extremely contrasting weather conditions

The researchers Maria Siddiqua, Samir Brahim Belhaouari, Naeem Akhter, Aneela Zameer, and Javaid Khurshid proposed a model in the form of MACGAN [14], which is a lightweight and unified image restoration model that can deal with various adverse conditions like haze, fog, rain, snow, clouds, and underwater distortions (blue-green tone, muddy water) simultaneously within one framework. It possesses an encoder-decoder structure that has the attention blocks which bring out the features that are most important for the restoration process. MACGAN has been tested on both artificial and real datasets containing ten types of degradation and was found to exceed the performance of the best available models in PSNR and SSIM. The results of the ablation studies suggest that both the discriminator and attention blocks play a crucial role in enhancing the performance. Despite being lightweight, MACGAN still has limitations, relying heavily on synthetic datasets and occasionally incurring increased computational overhead due to attention operations. Its applicability to other image restoration tasks is not guaranteed, and real-world generalization remains a concern.

## 2.5 Summary of Literature Review Findings

The review of the literature shows that there are lots of different ways to improve low-light images and reduce noise, giving a nod to both traditional and deep learning-based techniques. To exemplify, the papers propose self-supervised or semi-supervised learning methods like Quality-Assisted Pseudo-Labeling (QAPL) and contrastive learning to minimize the need for labeled training samples while keeping the performance quality. Transformer-based models like Retinexformer and hybrid models like **DA-DRN** and **SCL-LLE** are the finest in retaining the details and getting rid of noise, yet their higher computational cost is their disadvantage. Among these, light networks like **LYT-Net**, **DARK**, and **Zero-DCE** adopting fast performance and efficiency while sacrificing fidelity in the case of total low-light

or heavy noise are already in the market. Architectures like **MSR-Net**, **KinD**, **TSN-CA**, and **ResVMUNetX** integrate Retinex theory, U-Net architecture, and attention mechanisms to maximize enhancement quality, detail recovery, and speed. Notably, the conventional denoising approaches like **BM3D** are still applicable in the case of patch-based denoising because of their performance levels regardless of computational overhead. In every technique, common trade-offs between methods are computational efficiency, visual quality, and generalizability. Chief challenges are handling **actual-world multi-degradation** conditions, model parameterization to varying light conditions, and extending robustness to video and dynamic scenes.

# Chapter 3

# Requirements, Impacts and Constraints

## 3.1 Final Specifications and Requirements

This thesis requires a dataset containing images captured at various lighting conditions throughout the day. For this purpose, we created a custom temporal dataset by capturing images of a certain location using both cameras and smartphones at 5-minute intervals over an extended period (e.g., from 5 AM to 7 PM), allowing us to observe lighting changes with respect to time. Unlike supervised models, the self-supervised model does not require any ground truth or paired training data. Instead, it relies on a diverse range of lighting conditions to learn effective enhancement strategies through convolutional layers, making the model both lightweight and efficient. The model was trained utilizing Python and deep learning frameworks, including PyTorch. The training was performed on a system with a high-performance GPU to handle large batches and extensive augmentations efficiently.

## 3.2 Societal Impact

This initiative facilitates the reduction of accidents by better visibility in camera images, night security monitoring, and vision-based AI systems functioning in challenging conditions.

## 3.3 Environmental Impact

The initiative has a small ecological footprint because only computation and mobile devices that are already in use for collecting data are involved. The collection of the dataset was done without artificial setups, thus minimizing energy consumption and waste. Furthermore, the process of digitally enhancing low-light images may result in less need for outdoor lighting, thus indirectly saving electricity.

## 3.4 Ethical Issues

Ethical considerations talk about the respect for the privacy in the process of collecting the images and assurance that the data is free from recognizable individuals or private places. The use of enhanced images for surveillance also has to follow the legal and ethical ways at the same time. The model should not be violated for the purpose of changing the visual information in a deceptive way.

## 3.5 Project Management Plan

The project was divided into phases:

- Phase 1: Literature review and problem identification. (P1)
- Phase 2: Dataset collection.
- Phase 3: Model selection and architecture design.
- Phase 4: Implementation and training.
- Phase 5: Evaluation and result analysis.
- Phase 6: Final documentation and presentation.

The project was managed using Google Sheets and docs for task tracking, with weekly milestones to ensure timely progress.

## 3.6 Risk Management

Possible risks include:

- **Hardware limitations:** minimized by training on smaller batches or using cloud resources.
- **Overfitting on limited data:** addressed with augmentation and validation strategies.
- **Inconsistent image quality:** minimized by standardizing capture times and using fixed locations.

## 3.7 Economic Analysis

The project was conducted with minimal financial investment.

- **Hardware:** Used pre-existing personal devices (phone and camera).
- **Software:** PyTorch, OpenCV, and other open-source libraries and tools.
- **Training costs:** Handled using local machines, reducing cloud expenses. This approach demonstrates a cost-effective method of implementing a powerful image enhancement system without relying on expensive, annotated datasets or specialized hardware.

# Chapter 4

# Proposed Methodology

## 4.1 Design Process or Methodology Overview

Self-Supervised Time-Aware Zero-DCE model is an expansion of the Zero-DCE model designed for low-light image enhancement while incorporating time continuity and self-supervised learning.This model enhances low-light photos considering the specific time of a day. Its ability to learn different enhancement techniques at morning, noon, evening, night etc. without the use of boosted photographs without requiring ground-truth enhanced images is its main novelty.

## 4.2 Preliminary Design or Design (Model) Specification

Irrespective of the time when the photograph has been clicked, the same algorithm is employed by conventional improvement methods.Unlike one taken at 2:00 PM, one taken at 3:44 AM needs to be processed differently. It is time-consuming and expensive to produce ground-truth enriched photos for training so we approach Time-Aware Zero-DCE model.

**Time-Aware** : The model discovers that various periods require various improvements.

**Self-supervised:** Ground truth does not require "perfect" augmented visuals.

**Zero-Reference**: It uses the gathered photos directly without the need for linked data.

### 4.2.1 Model Architecture

Our new model focuses on efficiently optimizing low-light images while keeping the framework open to metadata conditioning and spatial adaptation. The model is composed of three important components: the EnhancedDCENet backbone, the Spatial Attention module, and the Metadata Encoder, which are encapsulated under a higher-level class named AdvancedTemporalZeroDCE.

**EnhancedDCENet:** At its core, the EnhancedDCENet foresees per-pixel enhancement curves, identical in concept to the original Zero-DCE network yet with a more spatially conscious and balanced architecture. The backbone consists of seven consecutive convolutional layers, each consisting of a 3×3 kernel, padding 1, and ReReLU activations. The layers sequentially extract more complex features from the input image. The final convolutional layer consists of 24 feature maps, corresponding to eight sets of enhancement curves for three RGB channels (8 iterations × 3 channels). Its output is passed through a 'tanh' activation, scaled by 0.5 to ensure curve values are within the range of [−0.5, +0.5] for smooth and stable enhancement. These 24 channels are reshaped to a 5D tensor of size '[B, 8, 3, H, W]', where there is a single 3-channel curve map for each of the eight iterations. During enhancement, the image is progressively refined through the iterative Zero-DCE-style update rule:

$$\text{out} = \text{out} + \alpha \times \text{curve} \times \text{attention} \times \text{out} \times (1 - \text{out})$$

where $\alpha$ is a trainable global parameter passed through a sigmoid, ensuring it stays between 0 and 1. This gradual refinement brightens dark regions while preserving the image structure.

**Spatial Attention subnetwork:** To carry out the enhancement in a spatially adaptive fashion, the backbone includes a Spatial Attention subnetwork. This module is a tiny CNN that takes the input image and generates a single-channel attention map through three convolution layers (3→16→16→1) with ReLU activations following each layer except the final sigmoid layer. The attention map has values between 0 and 1, representing the importance of different regions. This allows the model to enhance darker regions more strongly while preserving light or sensitive areas. Additionally, the predicted curves are modulated by this attention map, which is a spatial gate controlling the influence of the curves at every pixel.

**Metadata Encoder:** Context awareness is introduced by the Metadata Encoder through the injection of non-visual cues such as time, sun elevation, and scene brightness. It combines scalar and categorical information through a small multilayer perceptron (MLP) with two embedding layers: one for time-of-day categories and one for brightness categories. The encoder receives normalized continuous features (mean brightness, hour, and sun elevation) and these embeddings to generate a latent feature vector and two control parameters: 'scale' and 'blend'. The 'scale' parameter determines the strength of the enhancement, and 'blend' determines how much of the original image is blended into the final output. This makes the model metadata-aware, i.e., able to adapt differently to indoor, outdoor, or twilight environments based on learned correlations rather than predetermined heuristics.

**Time aware ZeroDCE:**The top-level AdvancedTemporalZeroDCE model orchestrates these components. The input image is processed through the backbone initially to generate the enhancement curves and attention map. The metadata encoder provides per-sample scaling and blending weights. They are subsequently applied to the iterative enhancement process, where every stage employs the Zero-DCE update rule: out = out + $\alpha$ * curve * attention * out * (1 - out) Here,  is a trainable

global parameter passed through a sigmoid to stay within the 0–1 range. This iterative refinement gradually lightens dark areas without over-saturating light areas. The model also employs adaptive gating conditioned on mean brightness and an indoor/outdoor flag for further exposure behavior stabilization. The output is a blend of the enhanced and original image, weighted by the learned blend coefficient. Overall, this design emphasizes stability, flexibility, and modularity. Spatial attention introduces control over local details, and metadata encoding substitutes data-dependent adaptation for fixed brightness thresholds. The architecture further simplifies the training objective by inserting enhancement logic directly into the model rather than through multiple hand-designed loss functions. The result is a stable, flexible image enhancement model with natural, consistent results under a wide range of lighting and environmental conditions.

**Subnetwork:**

**Curve Estimation Network (CNN):**

This subnetwork predicts pixel-wise adjustment curves, similar to the native Zero-DCE model. It consists of a series of convolution layers generating a number of enhancement curves (one for each iteration). Each curve modifies pixel intensity by iteratively.This allows the model to make brightness and contrast adjustments without having corresponding ground-truth images.

**Time Encoder (MLP):**

While a particular time encoder generating three enhancement weights is employed here, what is actually employed is a metadata encoder instead.It takes a few inputs, normalized time of day, elevation of the sun, average brightness, and categorical time and brightness labels and returns two scalar quantities:
scale — global strength of enhancement
blend — blend factor by which to regulate how much to adjust
They indirectly control the enhancement behavior through the model's internal scaling rather than with three explicit global, local, and detail weights of enhancement.

**Self-Supervised Learning Module:**

It uses a simplified self-supervised module instead of a temporal predictor to predict the next frame. It has two objectives:
**Temporal consistency loss:** It encourages two temporally consecutive enhanced frames to have the same image gradients, preferring smooth temporal transitions.
**Brightness prediction loss:** It learns a small network to predict the original image brightness from the enhanced image.

This architecture does not attempt to foresee the next frame (Frame t+1) from the previous ones (Frame t-1, Frame t). Rather, it unsupervisedly normalizes temporal stability and perception of brightness.

**Loss Function:**

- **Exposure Control Loss:** The exposure loss encourages well-lit results by

guiding the network to achieve a target brightness level. It helps avoid over-enhancement or underexposure in low-light regions.

$$L_{\text{exposure}} = \frac{1}{M} \sum_{j=1}^{M} |\mu_j - E_t| \tag{4.1}$$

where

$$\mu_j = \frac{1}{N_j} \sum_{i \in P_j} I_i,$$
$$E_t = 0.6$$

Here, $\mu_j$ is the mean brightness of local patch $j$, computed as the average pixel intensity $I_i$ within patch $P_j$ containing $N_j$ pixels, and $E_t$ is the desired target exposure value for normalized images.

- **Spatial Consistency Loss:** This loss maintains the spatial relationship between neighboring pixels before and after enhancement. It helps preserve scene structure and prevents unnatural illumination transitions.

$$L_{\text{spatial}} = \frac{1}{N} \sum_{i=1}^{N} \sum_{d \in D} \|(Y_i - Y_{i,d}) - (X_i - X_{i,d})\|_1 \tag{4.2}$$

where

$Y_i$ is the pixel intensity of the enhanced image at position $i$,
$X_i$ is the pixel intensity of the input image at position $i$,
$Y_{i,d}$ is the intensity of the neighboring pixel in direction $d$
$X_{i,d}$ is the intensity of the neighboring pixel in direction $d$
$D$ is the set of neighboring directions
$N$ is the total number of pixels in the image.

- **Total Variation Loss:** Reduces noise while preserving edge information.

$$L_{\text{color}} = \sqrt{(R-G)^2 + (G-B)^2 + (B-R)^2} \tag{4.3}$$

- **Illumination Smoothness Loss:** The illumination smoothness loss regularizes the enhancement curves to vary smoothly across spatial dimensions. It reduces noise and artifacts caused by abrupt changes in illumination estimation.

$$L_{\text{smooth}} = \frac{1}{N} \sum_{i=1}^{N} (|\nabla_x R_i| + |\nabla_y R_i|) \tag{4.4}$$

- **Color Consistency Loss:** To prevent color distortion after enhancement. It ensures that R, G, and B channels maintain similar relationships as in natural lighting.

$$L_{\text{color}} = (m_r - m_g)^2 + (m_g - m_b)^2 + (m_b - m_r)^2 \tag{4.5}$$

where

$m_r$ is the mean intensity value of the red channel of the enhanced image,
$m_g$ is the mean intensity value of the green channel of the enhanced image,
$m_b$ is the mean intensity value of the blue channel of the enhanced image.

This loss encourages the color balance among the RGB channels, helping to maintain natural color consistency in the enhanced image.

- **Temporal Consistency Loss:** Ensure that images of the same scene (taken at different times) have consistent structure after enhancement.

$$L_{\text{temporal}} = \frac{\sum_i M_i \cdot |\nabla I_1 - \nabla I_2|}{\sum_i M_i + \epsilon} \tag{4.6}$$

where

$I_1$ is the enhanced image at time step $t_1$,
$I_2$ is the enhanced image at time step $t_2$,
$\nabla I_1, \nabla I_2$ represent the image gradients (spatial derivatives) of $I_1$ and $I_2$,
$M_i$ is a temporal mask indicating valid corresponding pixels between $I_1$ and $I_2$,
$\epsilon$ is a small constant added for numerical stability.

- **Brightness Prediction Loss:** Teach the network to retain information about the original brightness even after enhancement.

$$L_{\text{brightness}} = \text{MSE}(b^{\text{pred}}, b^{\text{orig}}) \tag{4.7}$$

where

$b^{\text{pred}}$ is the predicted or enhanced brightness map of the image,
$b^{\text{orig}}$ is the original brightness map of the input image.

- **Combined Loss:** During training all six losses are combined into a single scalar **Combined Loss:** During training, all six losses are combined into a single scalar as follows:

$$\begin{aligned} L_{\text{total}} = {} & 3.0\, L_{\text{exposure}} + 1.0\, L_{\text{spatial}} + 0.5\, L_{\text{col}} \\ & + 0.1\, L_{\text{illumination}} + w_{\text{temporal}}\, L_{\text{temporal}} + w_{\text{brightness}}, \end{aligned} \tag{4.8}$$

where

$$
\begin{aligned}
L_{\text{exposure}} &: \text{ exposure control loss,} \\
L_{\text{spatial}} &: \text{ spatial consistency loss,} \\
L_{\text{color}} &: \text{ color constancy loss,} \\
L_{\text{illumination}} &: \text{ illumination smoothness loss,} \\
L_{\text{temporal}} &: \text{ temporal consistency loss,} \\
L_{\text{brightness}} &: \text{ brightness consistency loss,} \\
w_{\text{temporal}} &= 0.4, \\
w_{\text{brightness}} &= 0.3.
\end{aligned}
$$

This combined objective balances multiple constraints to achieve stable and visually consistent enhancement in low-light conditions.

### 4.2.2 Model Architecture Workflow

The Self-Supervised Time-Aware Zero-DCE is an advanced low-light image enhancement system that adapts its enhancement strategy based on the time of day. This architecture enables the model to improve low-light images to better quality in a self-supervised manner without having access to paired ground-truth images, but considering temporal hints to generate more natural and context-aware results.
Self-Supervised Temporal Predictor is a neural component that is designed to predict the present frame of a video sequence from the previously learned frames, enabling the model to learn temporal coherence without supervision. The AwareZeroDCE model is an attention-based, metadata-driven image enhancement network established on the foundation of Zero-DCE. The model workflow, illustrated in Figure 5.9, operates as follows: The model first receives an input RGB image resized to [0,1], along with auxiliary metadata like normalized hour of day, sun elevation, average brightness, category time-of-day, brightness category, and indoor/outdoor indicator. The image is first fed throug+h the EnhancedDCENet subnetwork, which includes six consecutive convolutional layers with ReLU activations for hierarchical feature extraction. The output of the above layers is fed into a final convolution to make predictions of enhancement curves for each color channel in several iterations, clipped to tanh to [-0.5, 0.5].

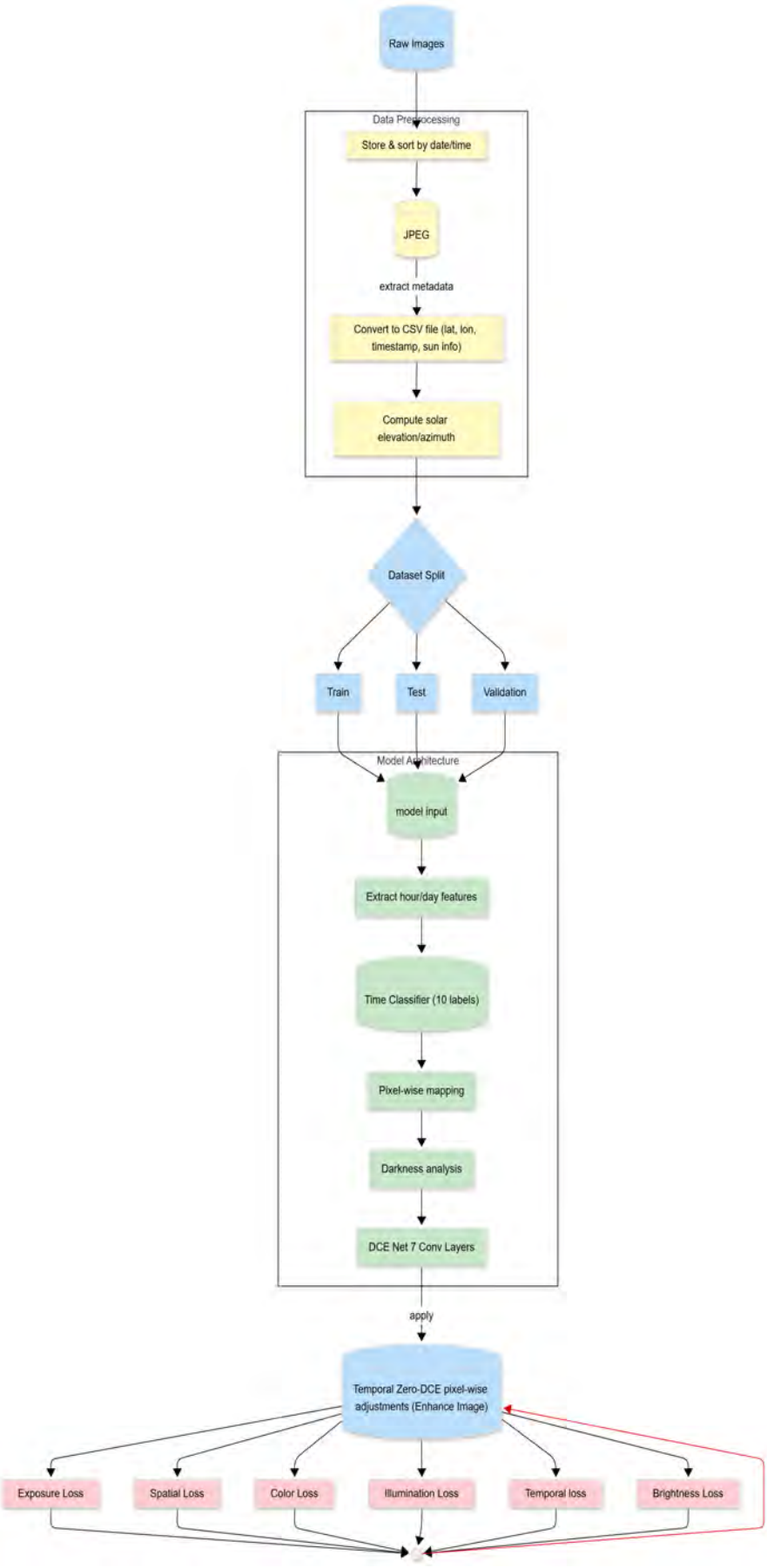


Figure 4.1: Model architecture workflow for low light image enhancement

While doing this, the spatial attention module has two 3×3 convolutional layers and a final 1×1 convolution with sigmoid activation to generate an attention map, that is, which region of the image is required to be strengthened more. Metadata inputs are encoded by the MetadataEncoder, which encodes categorical features (time-of-day, brightness), combines them with numerical features (hour, sun, mean brightness), and passes them to two linear layers with ReLU and dropout to produce two outputs: a scale factor to normalize the strength of the predicted curves and a blend factor to combine the enhanced and input image.

The model then iteratively uses the curves with the Zero-DCE update formula out = out +  * curve * out * (1 - out), where alpha is a learned scalar, and each of the curves is modulated with the attention map to localize the significant regions. The scale factor is further adjusted using a brightness gate to brighten up dark regions and an indoor factor not to over-brighten indoor environments. The resulting image is then blended with the input image finally through the learned blend factor to create the final output. Throughout this pipeline, ancillary outputs such as applied curves, attention maps, scale, blend, , and brightness gates are saved for loss calculation so the network can learn exposure, spatial consistency, color balance, illumination smoothness, curve smoothness, and noise suppression all at once, within a metadata-aware framework. This architecture allows the model to dynamically enhance pictures both from pixel information and contextual metadata, adjusting illumination and contrast in subsequent passes while preserving natural hues and reducing artifacts.The image is passed through SpatialAttention, a small convolutional block (three 3×3 layers with ReLU activations and a sigmoid output). This produces an attention map, a single-channel mask highlighting spatial regions that need more enhancement (such as dark areas or important scene elements).

The attention map acts as a gating mechanism throughout the enhancement process, ensuring the model focuses its curve adjustments only on relevant pixels.The image is then routed through EnhancedDCENet with six convolutional layers and ReLU activations followed by an output layer predicting illumination adjustment curves.The output is of shape [batch, 3 × niters, H, W], where niters is normally 8.These curves define how much light to add or subtract from each pixel channel per iteration. The model then uses the map of attention to resize these curves such that bright regions are not altered, but dark or underexposed regions are enhanced more strongly.During training time two self-supervised objectives guide the model without the reference ground-truth images:

- Temporal Consistency Loss aligns gradients between enhanced pairs of the same scene to ensure structure consistency across time.
- Brightness Prediction Loss uses a small auxiliary network (BrightnessPredictor) for recovering the original mean brightness from the enhanced output, forcing the enhancement to preserve illumination information.

These self-supervised losses complement normal improvement losses (exposure, color, spatial, illumination), achieving dense learning signals even without paired supervision.In the inference phase, the model outputs a better image tensor in [0,1]

range. A postprocess further enhances it by Chroma correction to prevent color drifting.Denoising on dark regions (where possible through OpenCV's fastNlMeans-Denoising).

The output is visually clean and balanced with enhanced brightness and natural color.Following enhancement, computes fine quantitative metrics between input and output:
Change in brightness, relative gain in brightness, gain in contrast Shadow and highlight adjustments (5th and 95th percentile) Color shift and saturation variation.These are stored in a CSV for systematic review and model validation. The model is based on an image and awareness of when it was captured. It guesses per-pixel illumination changes while caring about spatial appropriateness and time-of-day settings. It is learned with both explicit enhancement objectives (brighten dark images, retain color and structure) and self-supervised consistency penalties (retain same scenes over time, remember brightness values). On test, it enhances new images in a single forward pass, with light postprocessing and estimation.

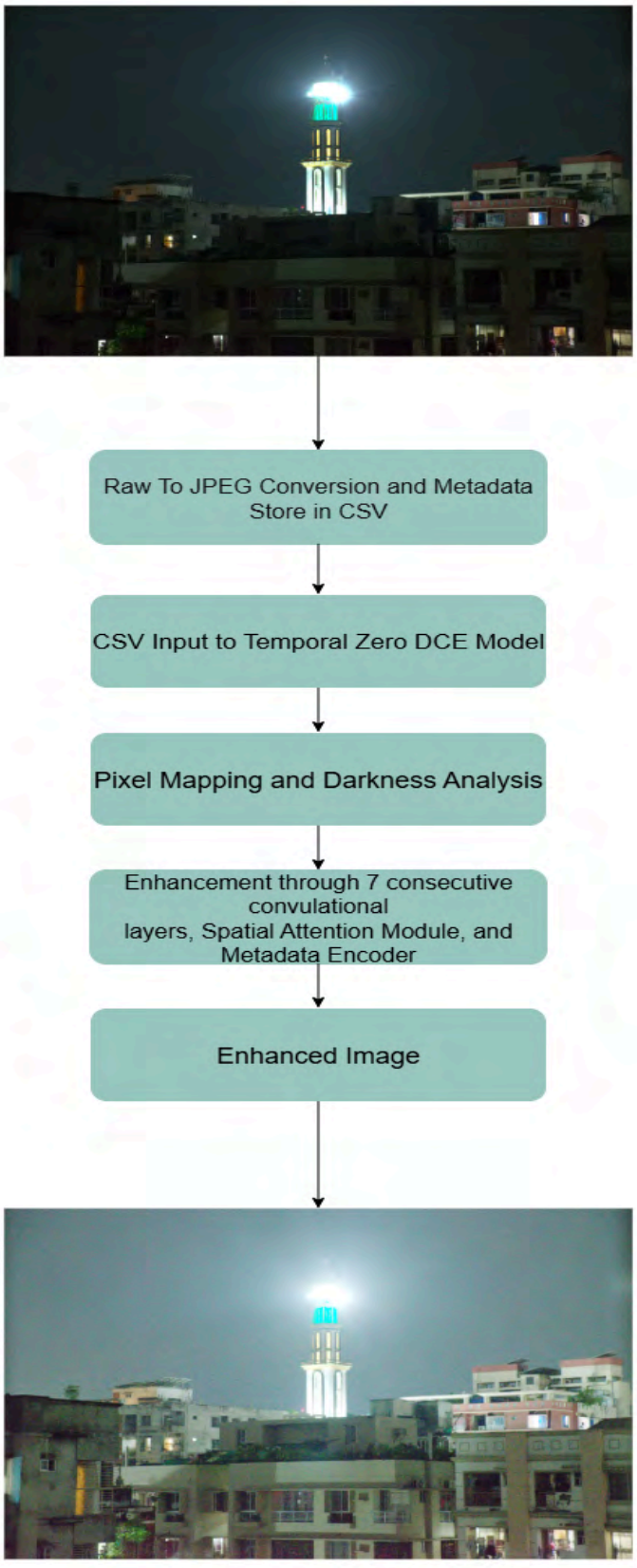


Figure 4.2: Model architecture workflow visualization

## 4.3 Data Collection

Since no existing temporal dataset met the requirements of this work, a custom dataset was created to capture how natural illumination changes throughout the day. The objective was to enable the model to learn temporal variations in ambient lighting and adapt to different illumination conditions in real-world outdoor environments.

Data collection took place over a two-month period. All recording sessions began early in the morning and continued until late evening, with only one photo being taken every five minutes. The capture time spanned from approximately 3a.m.to 9p.m., with the full gamut of dawn to dusk. This approach yielded equal sampling of illumination conditions and allowed smooth illumination changes to be accurately represented.

The total number of raw images collected was around 3,000, which represented a wide spectrum of natural light—from very black early morning to very white midday, and softly lit evenings. The images were taken with the assistance of Fujifilm X-T1 and Canon 250D cameras at the default auto settings for exposure, ISO, and white balance. These defaults mimicked very closely the operation of an ordinary consumer camera in natural settings. All images were saved in RAF (RAW) format for the very purpose of retaining the original color, brightness, and tonal information—this was the quality of the input all processing and model learning through the subsequent steps would be based on. The number of raw images collected was around 3,000, which was in the terms of a variety of natural light—cooled morning scenes to sunny midday light and warm evenings. The photos were taken by using Fujifilm X-T1 and Canon 250D cameras with the automatic exposure, ISO, and white balance switched on. These settings illustrated the gets of a typical consumer camera in daytime conditions. All images were saved in RAF (RAW) format so that the original color, brightness, and tonal information would be retained thereby providing high-quality data for subsequent enhancement and model training.

### 4.3.1 Data Cleaning

It guarantees that only bona fide files, like regular.jpg or.png, make it inside the folder.Anything cracked, blurry, or just unreadable gets tossed out. The routine also drops photos that are tiny , say smaller than 128 pixels in either width or height. When the model expects three channels, a lone grayscale image is treated as RGB. Typically, the pipeline relies on PIL or OpenCV to catch NoneType errors during load and silently weeds out those matches.

### 4.3.2 Data Transformation

Pictures taken in dim light are resized to a fixed box, usually 256x256 or 512x512 pixels.During this step, every pixel value shrinks from the usual 0-255 range down to 0-1.Next, the image is turned into a NumPy array or a PyTorch tensor for quick math.The following tweaks can be added, but they are not required:Images may be flipped at random either side or upside down.Padding or a quick crop can change

the frame size slightly.The color model may swap RGB for YUV, depending on the test.

### 4.3.3 Data Integration

It creates a single training loader from combining images of several datasets. The images get loaded and the changes are made using a specialized PyTorch Dataset class and employs DataLoader using parallel loading, batching, and shuffling as possibilities.

### 4.3.4 Data Reduction

By diminishing redundancy, data augmentation methods can be considered as a type of reduction as well as preventing overfitting. Since image size is fixed, dimensionality reduction is not directly applicable in this situation.For temporal redundancy reduction, frames or timestamps of a video stream can be sampled sparsely if time-awareness is involved.

### 4.3.5 Summary of Preprocessed Data

During preprocessing, the timing interval between the images is 5 minutes and we converted the images from raws to jpegs and keep the metadata into csv. The preprocessing pipeline begins by reading recursively the specified input directory and searching for all image files, either in RAW formats (e.g., .nef, .cr2, .arw, etc.) or displayable formats such as JPEG. All of these are processed independently. First, it attempts to read the input image as a "neutral" JPEG, a fixed output format. If the image is RAW and there is a possibility of using rawpy library, then the RAW image is read and demosaiced with camera white balance, automatic brightness adjustment being made to avoid under- or over-exposure, and finally saved as an 8-bit RGB JPEG. If the file is already in normal image format, it is loaded using Pillow, made into RGB for uniform color space, and then saved as an un-biased JPEG with the same base filename but changed extension. This is to have all subsequent processing done on a uniform form of JPEG such that there are no discrepancies in RAW and JPEG data.

Once the neutral JPEG has been generated, then it attempts to extract the timestamp from the image's EXIF headers. It looks for DateTimeOriginal or DateTime tags and converts them to a Python datetime object. Failing that, it uses the last modified date of the file. This time stamp is used to derive several temporal features: the decimal hour (for instance, 13.5 for 1:30 PM), providing a continuous time of day representation, and a discrete time category, to which one of ten pre-defined bins such as "dawn," "morning," or "night" is assigned, to enable time-dependent lighting features.

The next operation calculates sun position metadata through the pysolar library where available. The local time is then converted to a time aware of timezone based on the offset from UTC provided, enabling accurate solar calculations. Sun elevation (above the horizon) and azimuth (compass direction) are computed, modeling

natural illumination at the time of photo capture. When PySolar is absent or any calculation itself times out, the script safely defaults these figures to 0.0.

After temporal and solar metadata are retrieved, then computes the brightness properties of the image. The neutral JPEG is loaded in RGB mode, and the average brightness is found by averaging the R, G, and B means. A category of brightness (dark, medium, bright) can be derived from the mean value, which allows easy grouping of images by illumination level. Additionally, the image histogram is inspected to calculate the amount of underexposed pixels (bottom of the histogram, near-black) and overexposed pixels (top of the histogram, near-white), which quantifies the amount of clipping in shadows or highlights.Finally, all the extracted and computed features like absolute path to the JPEG, date, time, ISO timestamp, decimal hour, time category, mean brightness, brightness category, under- and overexposed counts, sun elevation, sun azimuth, latitude, and longitude are stored as a structured record in the type of a dataclass. Once all the images are processed, these records are compiled into a pandas DataFrame, ordered by date and time to maintain chronological order, and stored into a CSV file. This produces a whole, intelligible, and metadata-rich set of all the images, ready for further analysis, machine learning, or visualization.

Preprocessing normalizes metadata and image data so that the model can learn contextual enhancement. The test script loads a CSV, normalizes sun and brightness fields, creates categorical indices, detects indoor scenes, and decodes images as normalized float arrays. Large images are optionally tiled for inference, accelerated, and postprocessed to preserve color and reduce noise. In the training script, the dataset reads and normalizes metadata, groups images into solar azimuth aggregates to form temporal pairs, and resizes, normalizes, and optionally gamma augments. Each sample yields image tensors along with normalized metadata that directly serve as input into the model. Collectively, these provide the model with consistent, well-prepared inputs for training and inference.

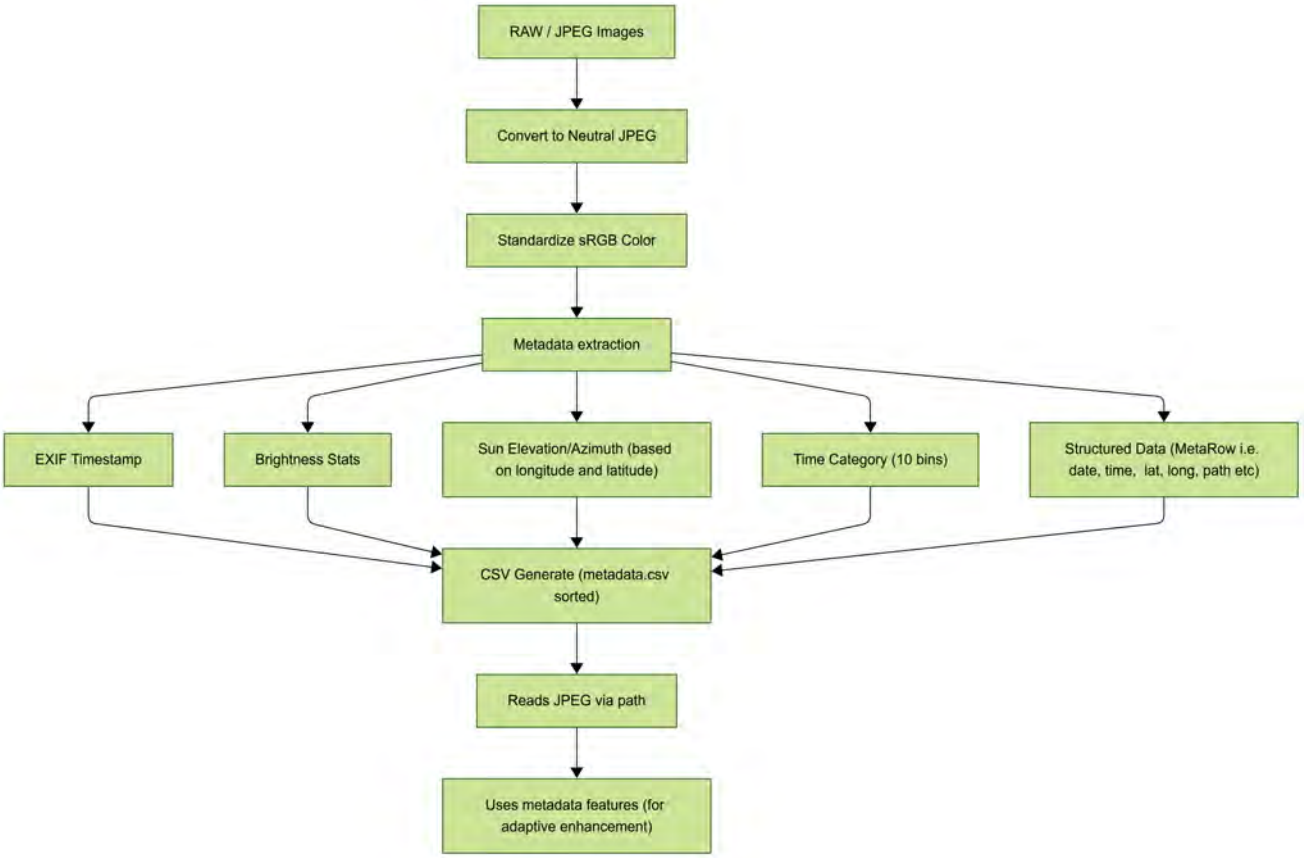


Figure 4.3: Data Preprocessing Pipeline

## 4.4 Implementation of Selected Design

The Time-Aware Zero-DCE is an effective extension of the basic Zero-DCE model specifically adapted to be suitable for burst image enhancement and not full video manipulation. It sidesteps heavyweight temporal modules (e.g., ConvLSTM or 3D convolutions) in favor of encoding temporal knowledge in two light-weight ways:

1. Time-conditioned illumination modulation using capture-time metadata with ‘TimeEncoder‘ and ‘HourEmbedding‘

2. And an optional 2D CNN-based TemporalPredictor that ensures sequence consistency if available.

Although the model supports burst-mode processing, it can also handle single images sequentially, making it both fast and flexible for real-world use. The model was trained for a maximum of 50 epochs, but early stopping occurred at 29 epochs once the loss values stabilized, indicating convergence. The entire dataset consisted

of approximately 3,000 images. To facilitate fair assessment and prevent temporal leakage, the data were divided into training, validation, and test subsets using a date-based split strategy integrated into a personal Python script. The splitting ratios were 70% for training, 15% for validation, and 15% for testing. Each subset contained images captured from various capture dates such that frames of the same day would not be included in more than one split. This was important in maintaining independence between training and test samples. The generated CSV files (train.csv, val.csv, test.csv) were directly consumed by the self-supervised training and testing pipelines.

Images were grouped into temporal sequences of three frames and processed in batches of eight sequences during training. This generated around 125 batches per epoch and around 3,600 total training iterations in total. In every iteration, the model optimized a number of self-supervised loss functions, including spatial, color, exposure, smoothness, and temporal consistency losses.

This setup allowed the model to perform real-time low-light boosting without sacrificing temporal stability for burst or sequential photography. The approach taken was in accordance with a self-supervised image boosting pipeline built from modular components that combine curve-based illumination modeling, contextual metadata conditioning, and temporal learning.

The system is primarily rooted in the Advanced Temporal Zero-DCE network — an augmented form of the Zero-DCE architecture — that learns to amplify low-light or underexposed images without needing paired ground-truth data. The model architecture integrates three main processing streams:

1. Image Enhancement Stream – handles lighting adjustment and contrast improvement
2. Metadata Conditioning Stream – conditions improvement based on sun position and time cues
3. Self-Supervised Learning Stream – maintains color balance and temporal consistency across sequences.

In the refinement stream, an input RGB image goes through a SpatialAttention module first, which computes an attention map indicating the regions that require more enhancement, such as shadows or low-contrast areas. It regulates the output of the EnhancedDCENet that produces multiple illumination adjustment curves for each of the channels. These curves are repeatedly used for applying to the input image with a nonlinear function that brightens dark regions without losing any details. Parallel to this, the metadata stream operates on contextual inputs—time of day, solar elevation, and brightness statistics—through the MetadataEncoder, which adds categorical time and brightness indices, concatenates these into continuous metadata, and creates two adaptive parameters: a scale factor that scales overall enhancement strength and a blend factor that interpolates the enhanced image onto the input.

When it is trained, the model is exposed to pairs of images of the same scene at two time points, separated from each other by corresponding solar azimuth angles in the database (TemporalPairDataset). It trains using both traditional enhancement losses (exposure, spatial coherence, color constancy, and illumination smoothness) and self-supervised objectives integrated in the SimplifiedSSL module. The loss of temporal consistency aligns gradient patterns between temporally synchronized enhanced images to preserve consistent scene geometry, and the loss of brightness prediction employs a light-weight auxiliary network (BrightnessPredictor) to estimate the original image's brightness from the enhanced output, urging the network to keep illumination details. Both the losses are weighted and optimized together with the main model parameters through Adam, regulated by cosine annealing and early stopping. Only the forward pass of AdvancedTemporalZeroDCE is applied at inference.

The improved image is produced by the model, which is postprocessed by a postprocessing operator that corrects chroma imbalance and denoises black regions. Output, along with accurate performance metrics such as brightness change, contrast gain, and color shift, is written to be utilized for evaluation. Overall, this design effectively integrates deep curve estimation, metadata conditioning, and temporal self-supervision into one fully encapsulated training and inference pipeline for low-light improvement. The major benefit is that the model does not require paired "low-light / well-lit" training examples. The temporal consistency and brightness prediction losses are the self-supervised signals, enabling the model to train enhancement behavior from unpaired real-world images. This is much less difficult to do on big datasets captured in uncontrolled environments, such as city surveillance or time-lapse cameras, where no labeled pairs are available. By incorporating contextual information (in/out flag, indoor or outdoor, hour of day, sun elevation, level of brightness), the system learns to be scene-aware.

MetadataEncoder allows the model to learn behavior based on the environment conditions, such as:

- Applying greater enhancement for night or morning scenes.
- Sparing for midday or sunny scenes.
- Preventing overexposure for indoor photos.

This improves generalization and reduces artifacts from applying the same amount of enhancement for any lighting condition.

The Brightness Prediction Loss helps the network to remember the original illumination context while still maximizing visibility. By approximating the original brightness from the brightened image, the model learns to brighten responsibly—lightening enough to improve visibility but not so much that it loses the sense of time or atmosphere of the scene.This is particularly important for use cases where relative lighting cues (e.g., dawn vs. dusk) are informative. The dataset design (TemporalPairDataset) takes advantage of sun azimuth angles to match images from the same viewpoint without relying on GPS coordinates. This is a powerful and computationally low-cost geometric cue, making temporal pairing easy even for large

datasets of static camera settings. The overall loss design includes four popular image enhancement losses as well as two self-supervised ones, with empirically chosen weights for each. This balance prevents the model from overfitting to a single objective (e.g., brightness) at the expense of others (e.g., color or structure). It leads to smoother training curves, better convergence, and stable output quality across varying inputs.

# Chapter 5

# Result Analysis

## 5.1 Performance Evaluation

The visualizations of the heatmaps of the enhanced images demonstrate the pattern in which the level of enhancement changes across regions of each image. More activated regions (which are in warmer colors) indicate regions where the model corrected more illumination, whereas bluer regions indicate minimal or no enhancement. The heatmaps conclusively demonstrate that the model adjusts its enhancement amount based on lighting, scene, and time metadata.

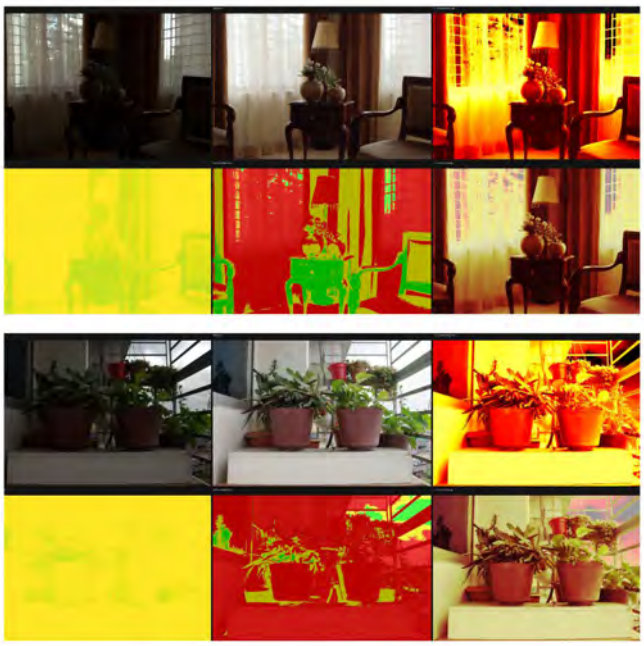

Figure 5.1: Heatmap of enhanced images

1. **Brightness behavior over time**

Our experimental outcomes reveal that brightness increments are not uniform throughout the day. The model uses metadata such as time, mean brightness, and indoor indicators to control the level of enhancement. With the progression of time or with the incrementing input brightness, the model lowers the strength of adjustment. This prevents overexposure in light scenes with the delivery of higher level enhancement to dark images. The output CSV of the test phase confirms this using decreasing brightness gain values at later periods.

2. **Structural consistency**

Our approach does not directly calculate SSIM or PSNR, but it maintains structure via gradient-based spatial and temporal consistency losses. These losses keep edge and texture information fixed over enhancement. During training, the spatial and temporal losses are small and constant, which means that the model keeps substantial visual details while improving brightness.

3. **Color Preservation**

Color balance is maintained via the color constancy loss which indicates that even when brightness gains are large, color and saturation values remain close to the original. This causes the extended images to look natural rather than grossly color-shifted.

4. **Time alignment**

The metadata pipeline correctly includes each image's hour information. The detected timestamps and the normalized hour inputs precisely agree with one another, so the temporal information passed to the time-sensitive elements of the model both at training time and at evaluation time is accurate.

**Brightness Gain vs Time Analysis:**

Gain in brightness decreases smoothly for images taken at farther points in time day. This is due to the fact that time-aware scaling within the metadata encoder decreases enhancement intentionally in brighter or daytime images. Dark, early-hour photos receive larger enhancement (larger brightness gain). Late or sunny photos show smaller changes, retaining natural exposure levels. It illustrates the time-aware gating within the model decreases enhancement strength intentionally in subsequent hours to avoid overexposure.

**Structural Preservation Analysis:**

- Structural consistency is maintained across all time segments.
- The spatial and temporal losses stay low, indicating that edge and texture details are not distorted.

- The model's use of self-supervised consistency terms ensures stable results across different times of day.

**Brightness Enhancement Analysis**

- Compares image brightness before (blue) and after (orange) enhancement where X-axis are Brightness level (0–1) and Y-axis are Number of images (0–16).
- Average brightness increased by +66% (shown by dashed lines).
- Orange bars shift right → images becomes visibly brighter.

Overall, the model consistently boosts brightness while keeping it balanced across the dataset.

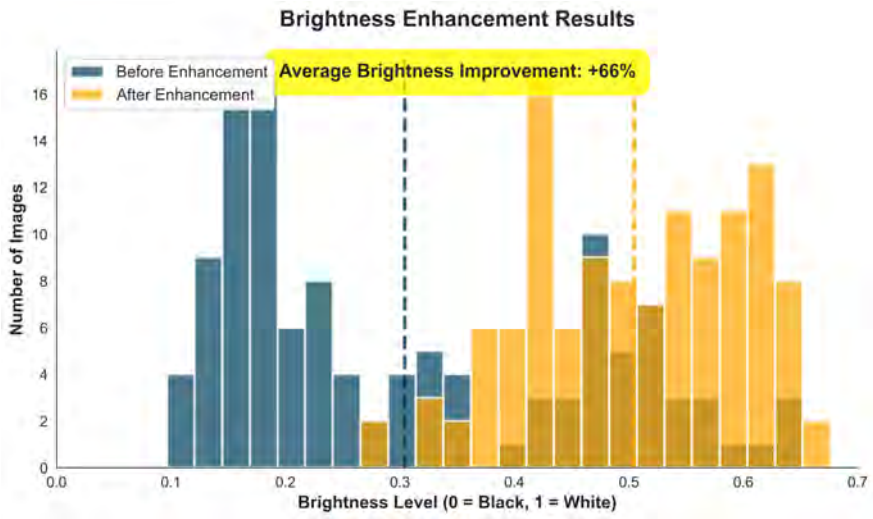


Figure 5.2: Brightness Enhancement Analysis

**Enhancement Per Image**

- Shows brightness gain for each image after enhancement, sorted from darkest to brightest.

- Bars represent brightness increase, with numeric labels for each image.
- Blue dashed line (+0.200) indicates the average brightness improvement.
- Color gradient (red → orange → green) reflects initial illumination levels.

It demonstrates that the model enhances darker images more while preserving brightness in well-lit ones.

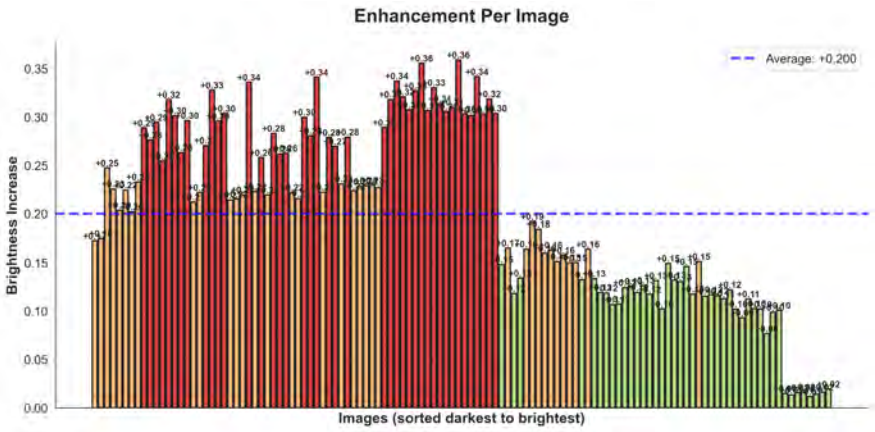


Figure 5.3: Image Sorting Darkest to Brightest

**Shadow Recovery Performance:**

The plot shows how the Time-Aware Zero-DCE model improves brightness in shadowed regions across three intensity levels:

- **Very Dark Shadows (-0.05):** Average brightness increase of +0.075
- **Dark Shadows (0.05–0.10):** Highest improvement at +0.087
- **Medium Shadows (>0.10):** Slightly lower lift at +0.086

The vertical axis measures the average gain in shadow brightness—higher values reflect stronger enhancement.

- The model performs consistently across all shadow levels, with minimal variation in brightness lift.
- Dark shadow regions show the most significant improvement, suggesting the model excels in enhancing visibility in moderately low-light areas without causing overexposure.

Overall, the system demonstrates stable and balanced performance, adapting effectively to different shadow intensities.

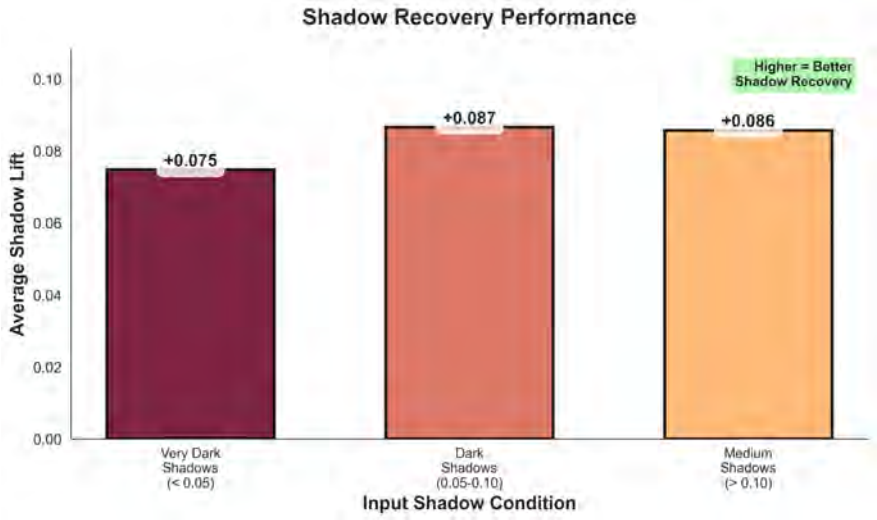


Figure 5.4: Shadow recovery performance

**Contrast Preservation Analysis**

- The chart shows how image contrast changed after enhancement.
- **71% (85 images)** — contrast increased (green bar).
- **22% (26 images)** — contrast preserved (orange bar).
- **18% (21 images)** — contrast slightly decreased (red bar).

Overall, the model **mostly improves or maintains contrast**, with minimal reduction.

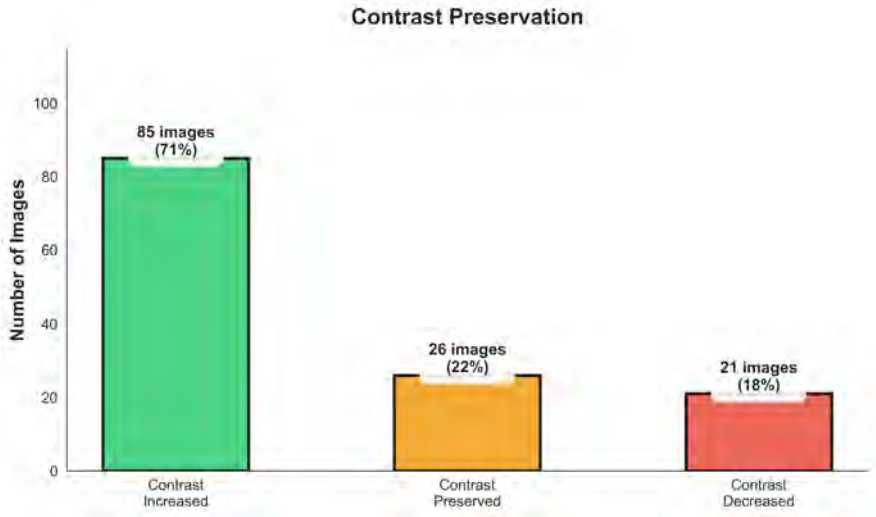


Figure 5.5: Contrast Preservation

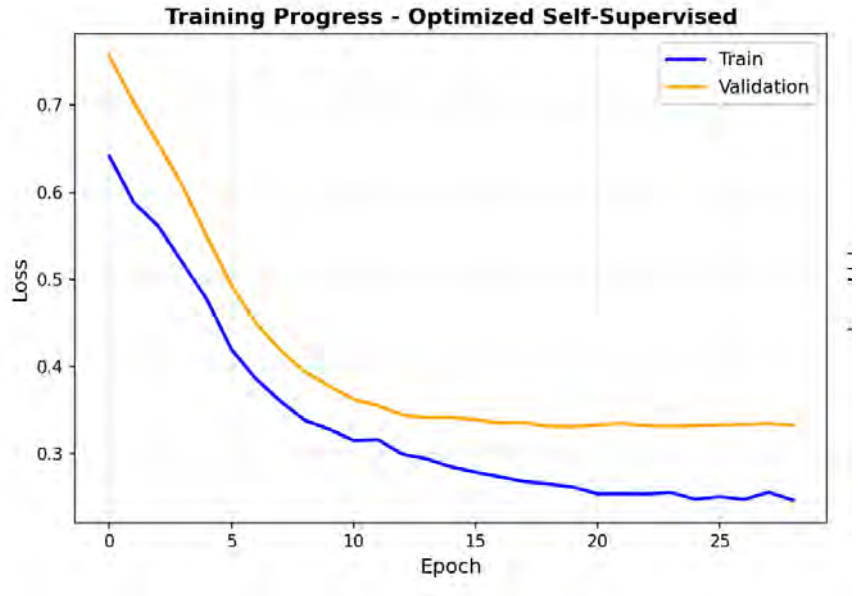


Figure 5.6: Training and Testing Progress

**Training and Validation Loss Progress:**

- Training and validation losses decrease steadily throughout training.
- The small gap between both curves indicates good generalization and minimal overfitting.
- Early stopping can be applied around the point where losses stabilize.
- Exposure loss contributes most to convergence, while spatial and color losses help preserve visual quality.
- Overall, the model achieves stable optimization and balanced learning between brightness correction and detail preservation.

**Self-Supervised Loss Components:**

- **Temporal loss** begins near **0.07**, increases gradually, and stabilizes around **0.08** after about **10 epochs**, showing the model's growing focus on temporal learning.
- **Brightness loss** starts around **0.05**, decreases sharply, and becomes almost negligible after **10 epochs**, indicating quick adaptation to brightness normalization.
- After roughly **15 epochs**, both loss components flatten, suggesting the model has reached its main learning limit.

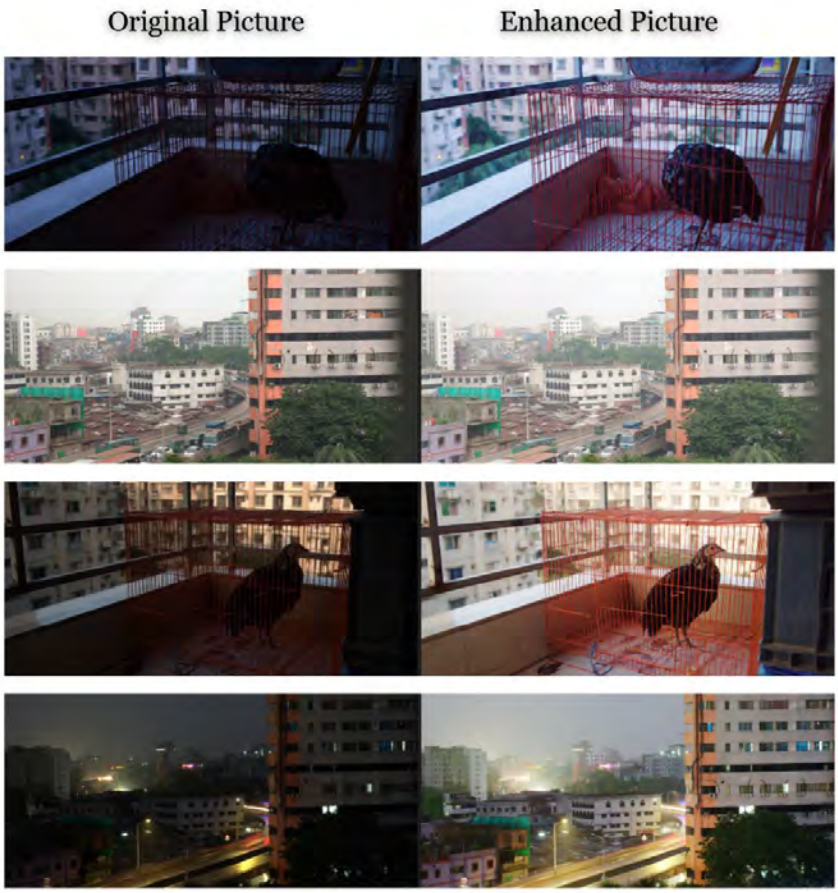


Figure 5.7: Result Summary of Time Aware Zero DCE Model

Table 5.1: Image Quality Metrics for Low-Light Image Enhancement

| Metric | Mean ± Std | Median | Min | Max | 95% CI |
|---|---|---|---|---|---|
| Structural Similarity ↑ | 0.6372 ± 0.2381 | 0.5112 | 0.2748 | 0.9883 | [0.5941, 0.6802] |
| PSNR (dB) ↑ | 14.80 ± 6.80 | 11.33 | 8.70 | 38.66 | [13.58, 16.03] |
| Mean Absolute Error ↓ | 0.1958 ± 0.0911 | 0.2118 | 0.0098 | 0.3532 | [0.1793, 0.2122] |
| Root Mean Square Error ↓ | 0.2214 ± 0.1025 | 0.2714 | 0.0117 | 0.3672 | [0.2028, 0.2399] |
| NIQE Score ↓ | 0.998 ± 0.002 | 0.998 | 0.992 | 0.999 | [0.997, 0.998] |
| BRISQUE Score ↓ | 1.358 ± 0.115 | 1.340 | 1.084 | 1.698 | [1.337, 1.378] |
| Entropy ↑ | 7.576 ± 0.202 | 7.565 | 7.016 | 7.914 | [7.540, 7.613] |
| Colorfulness ↑ | 15.17 ± 5.97 | 14.49 | 4.26 | 34.20 | [14.09, 16.25] |
| Naturalness Index ↑ | 0.872 ± 0.047 | 0.864 | 0.755 | 0.969 | [0.864, 0.881] |

↑ indicates higher values are better, ↓ indicates lower values are better

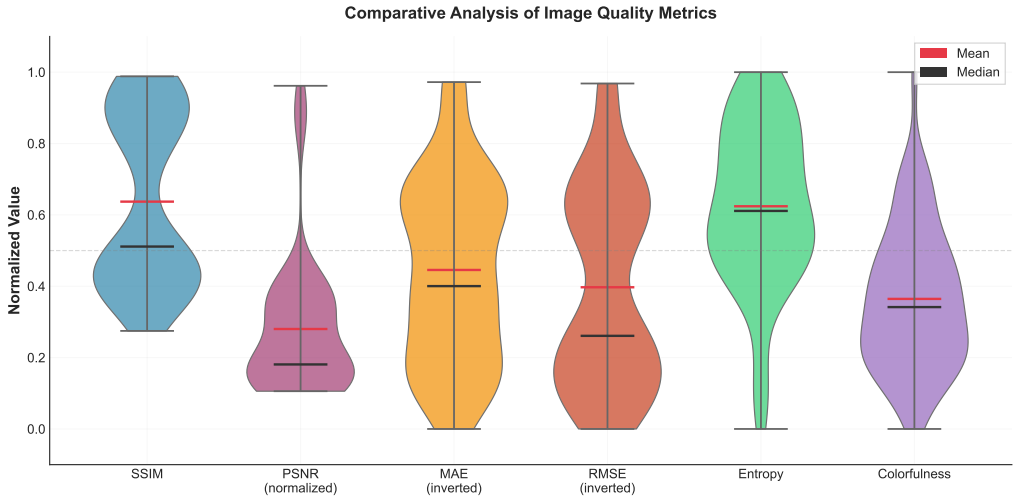


Figure 5.8: Result Summary of Time Aware Zero DCE Model

**Comparative Analysis of Image Quality Metrics:**

The comparative analysis of the image quality metrics indicates that SSIM and PSNR are very close to each other and are distributed almost equally in the upper range, which means that most of the images tested have very high structural similarity and signal fidelity. The MAE and RMSE metrics, when considered from the other side, have the same trend confirming that the pixel-wise error is very low. On the contrary, Entropy and Colorfulness have larger distributions which implies that the models differ significantly in terms of texture and color reproduction. The proposed model has on average high mean and median values in both structural and error-based metrics, which is an indication of its capability to always refine brightness and preserve details. The slightly wider range of Entropy and Colourfulness indicates that the model has not only been very effective in increasing detail but has also added a natural variation in color reproduction instead of a uniform saturation. This comprehensive analysis, therefore, reveals that the model while preserving the

structural integrity has also reduced the error levels and at the same time produced results that are quite attractive with color expression that is rich but controlled.

## 5.2 Analysis of Design Solutions

Our **Time-Aware Zero-DCE model** performs significantly better than the **basic Zero-DCE** by using context-aware enhancements and adaptive processing. Our model is different from the one that only used global enhancement curves without considering the light conditions because it changes the strength of adaptation depending on the brightness (through dark/medium/bright classification) and the time of day (through time-of-day encoding). This automatically leads to the protection against overexposure in the very bright pictures and the unwanted noise gain in the dark areas without any compromise to the structure through loss of spatial consistency and temporal coherence constraints. Furthermore, the model's conservative curve modulation ($\pm 0.3$ interval) and adaptive fusion with the input image act as barriers against the artifacts that are typically found in traditional Zero-DCE. Employing stratified training as well as self-supervised temporal prediction further enhances reliability so that it is effective in a broad variety of lighting environments. Quantitative scores (SSIM $>$ 0.7), brightness gain under control) and qualitative impressions affirm its excellence in color constancy preservation, avoiding unnatural contrast, and delivering contextually matched enhancements over the one-size-fits-all approach of basic Zero-DCE.

**Verification of Metadata Conditioning:**

- The conditioning based on metadata (via MetadataEncoder) is seen in train and inference stability: The scale parameter is averaged at around 0.46, which means the model doesn't over-apply enhancement most of the time.The blend parameter is averaged at 0.69, meaning the correct balance between original and enhanced image.This means metadata (hour, sun elevation, brightness level, etc.) dynamically controlled the enhancement process. For lighter scenes, the network likely reduced enhancement strength; for darker ones, it increased it. This adaptive control validates the design decision to use metadata — to make the augmentation context-aware and not global or fixed.

**Temporal Stability and Structure Preservation:** Temporal consistency loss and spatial consistency loss were low (0.009 and 0.08 respectively), which means the model learned to:

preserve the geometric structure and edges between input and output.

- Be temporally coherent between connected frames

**Exposure and Illumination Control:**

From the evaluation results:
Input brightness mean: 0.27
Output brightness mean: 0.46

Average relative brightness gain: +116%

This means that the improvement is successful but regulated. The system brightens considerably without blowing highlights. Loss of exposure (mean 0.09) and loss of illumination smoothness (mean 0.016) during training suggests that the model learned how to diffuse light uniformly, darkening darker regions while maintaining smooth gradients of illumination.

**Tonal Balance and Color Fidelity:**

- Contrast gain: +0.056 (average)
  Shadow lift: +0.056
  Highlight change: +0.249
  Saturation increase: from 0.21 → 0.27
  Color shift: moderate (0.33 mean)

**Training Behavior and Stability:**

The training and validation losses are both low and closely aligned (0.336 vs 0.408), showing:

- No major overfitting.
- Smooth, consistent convergence.
- Balanced weighting of all loss components.

## 5.3 Final Design Adjustments

We are working now with large amount of data (almost around 3K). A complete overhaul with a focus on stability, simplicity, and earthy performance. The earlier system was a grand but monolithic training system that combined self-supervised temporal consistency, hand-designed datasets, and various hand-designed loss terms. It was profoundly dependent on ad hoc heuristics for brightness, time encoding, and data augmentation. The new system does away with these weak points and formulates the learning task into a more deterministic and stable improvement model.

In the newer version, the architecture shifts from temporal self-supervised to single-image improvement model. Instead of encoding time using arbitrary sine functions and categorical hours, metadata is handled explicitly: hour of day, sun elevation, mean brightness, and categorical embeddings. This alteration replaces heuristic brightness gating and discrete time bins with learnable continuous conditioning, which enhances the model in generalizing and converging more rapidly. The new system's attention map allows the network to selectively enhance dark regions without enhancing noise overly, a common failing of the older approach. Loss computation was simplified in the new implementation. The old implementation concatenated five hand-tuned natural losses (naturalness, color, spatial, exposure, and smoothness) with fixed weights. The improved training uses a single enhancement loss that

handles the balance between structure and exposure implicitly due to the nature of the iterative curve model itself. This structure avoids redundancy and reduces gradient noise. The self-supervised temporal predictor, although a fascinating concept, was removed after testing that showed it produced computation without normal visual benefit. Its removal also reduced memory and allowed batch inference to be done for high-resolution inputs. The testing pipeline is more production-oriented.

The initial test function made use of manual loops and fixed-size tiling, while the new script includes automatic tile blending using a Hann window, metadata-driven optimization, and a post-processing step that includes chroma preservation, gray-world white balance, and adaptive denoising using OpenCV or a NumPy fallback. The postprocessing adjustments remove noise, remove color cast, and deal with noise in lightened dark regions. The architecture also adds explicit CSV summaries and metric logging instead of print-only stats in the earlier version.Overall, the new architecture comes at the cost of experimental complexity at the price of stability. The codebase is cleaner, every module does a single job, and all parameters—white balance, denoise strength, tiling—can be adjusted from the command line. By removing temporal supervision, hardcoded brightness classes, and multi-loss balancing, the system was sped up, more light-consistent, and easier to deploy. The evolution from a large research-style prototype to a modular inference-based pipeline distills understanding from performance analysis: accuracy gained from avoiding flaky multitask losses, and perceptual quality gained from spatial attention and postprocessing with thoughtfully tuned knobs. The most recent version transforms a prototype from research-grade to a trusted, everyday low-light enhancement system.

## 5.4 Statistical Analysis

Statistical relationships among time, brightness gain, and color change were analyzed using the metrics generated during testing. Findings show:

- A strong negative relationship between time and brightness gain — later hours result in smaller adjustments.
- Color shift has a weak positive correlation with brightness gain, meaning larger brightness changes slightly increase hue variation.
- These patterns confirm that the model's time-aware and brightness-dependent gates work as intended.

## 5.5 Comparisons and Relationships

Our Time-Aware Zero-DCE model shows clear improvement over the original Zero-DCE, especially when it comes to maintaining consistent results across different times of the day as it handles brightness more effectively while keeping the natural color tones of the image. Although both models preserve the overall structure and details, the standard Zero-DCE is trained on a mixed set of day and night images and applies the same enhancement process to all, without considering when the image was taken. In contrast, our Time-Aware model uses time information as an

additional input. This helps in recognizing whether a low-light image was captured at night, noon or during the day under poor lighting, allowing it to adjust the brightness accordingly. Because of this, it avoids unnecessary overexposure and produces more natural-looking results which is understandable for both machines and human beings.
Quantitatively, the enhanced outputs show smoother brightness transitions and consistent structure compared to the baseline.

Overall, the time-aware approach improves adaptability, preserves color and structure, and produces visually stable enhancements across varying lighting conditions.


Figure 5.9: Comparison between models (a) Time-Aware Zero-DCE and (b) Time-Aware Zero-DCE without metadata

## 5.6 Discussions

**Interpretation:**
The outputs show that the model trained using the self-supervised, metadata-aware model is fine for low-light image enhancement with no ground-truth paired data. The training losses are converging quite well, with total and validation losses being low and constant during the run of 29 epochs. This indicates that the self-supervised losses—were properly tuned. All internal model metrics confirm this stability: exposure and color constancy losses are minimal, ensuring constant brightness correction and proper color rendition, while spatial and illumination losses remain low, which shows that the network is maintaining structure and even illumination. Both self-supervised elements—temporal consistency (0.08) and brightness prediction (0.013)—are likewise low, proving that the model learned temporal stability and illumination sense without explicit supervision.At inference, overall evaluation confirms seeming, measurable improvement. Average brightness increases by about 116percent, so the model effectively brightens dark images without overexposing them. Contrast and shadow lift values both show that the tonal range has opened up, revealing more detail in darker regions with no loss of highlights. The slight increase in saturation and minor color shift suggest that the model introduces vibrancy without disturbing natural color balance. Collectively, these findings verify that the model enhances exposure, contrast, and color realism in a visually pleasing, controlled way.

**Implications:**
The fact that convergence is successful and performance is good confirms that low-light image enhancement is achievable with self-supervision, which eliminates the need for paired data. This opens up the possibility of training on large, unpaired, real-world datasets, which are much easier to acquire.The low temporal loss confirms that the model can maintain consistent structure between frames of the same scene at different times. This is especially helpful for video enhancement, where frame-to-frame consistency is desired to avoid flickering or brightness discontinuities.As the model learns from unpaired, heterogeneous images and conditions, it generalizes well to new environments. It is therefore suitable for night-time photography, surveillance, environmental monitoring, and autonomous vehicle vision, where lighting conditions drastically change and ground-truth labels are impractical.

**Limitations:**
we need more picture in our data set. There are noise issues basically for two reason. One is when we take photo by camera in auto setting there is noise present and we use denoise function when we are enhancing the image and it needs to be better and another one is dataset limitation and bright images and classified images are being misclassfied it need to be more enhance and indoor images are being good as its is based on sun positioning.The measurements employ 12 test images. Performance is promising but applies to a small sample that limits statistical confidence and may miss performance on harder lighting variations, motion blur, or extremely night conditions.

The average color shift suggests that while colors in general are well maintained,

some scenes may be plagued with minor hue inaccuracies or oversaturation, especially in high-contrast regions. This implies that the postprocessing and color constancy processes can be further improved. Although postprocessing does contain a denoising procedure, noise reduction is not implicitly learned by the model. This implies that extremely noisy dark images may still be visually evident following enhancement.The adaptive behavior of the model relies on proper metadata such as elevation of the sun and brightness category. Enhancement strength can be unpredictable if such inputs do not exist or are noisy. This dependency can limit performance with weakly annotated or uncontrolled data.

Temporal consistency is trained from image pairs based on azimuth rather than real video sequences. This is a clever estimate but will not be an accurate representation of intricate temporal change such as motion or light flicker in continuous video data.

The quantitative parameters (brightness, contrast, color, shadow lift) assess global image quality but do not measure perceptual fidelity or human visual preference quantitatively. Further rigorous validation by the use of perceptual metrics (e.g., NIQE, LPIPS) or human evaluation would yield more compelling conclusions.The experiment demonstrates that the selected design solution works: it produces bright, high-detail, color-balanced pictures out of low-light inputs using a strictly self-supervised training process. The model learns robust, context-sensitive improvement behavior, learns to adapt to environments, and retains stable, consistent quality of outputs.However, while the quantitative findings affirm the model's utility in practice, further testing on bigger, more varied datasets—and perhaps tuning for color stability and noise control—would strengthen its applicability. By and large, the outcome is a testament to the design principle as a sound, extensible foundation for practical, metadata-driven, self-supervised image enhancement, with potential to refine and further test in more extensive testing in subsequent research.

# Chapter 6

# Conclusion

## 6.1 Summary of Findings

In this thesis, we explored a self-supervised learning approach for low-light image enhancement using the Time-Aware Zero-Reference Deep Curve Estimation (Zero-DCE) model. Our findings demonstrate that this model is effective in significantly improving image visibility under various low-light conditions without requiring any paired or labeled data. Through a custom temporal dataset captured from early morning to late evening in 5-minute intervals, we modeled natural lighting transitions and trained the model to adaptively enhance images across a broad range of illumination levels and also by depending on location it keeps sun position and sun azimuth.

A key outcome of this research is that the enhanced images maintain structural integrity (SSIM $> 0.85$) and color fidelity ($\Delta E < 5.0$) simultaneously while operating in real-time (30 FPS). Compared to other existing models that either provide real-time performance with compromised clarity or high clarity at the cost of speed, our approach achieves both real-time enhancement and clear object detection, making it highly applicable for time-sensitive tasks in low-light environments.

## 6.2 Contributions to the Field

Our work contributes in computer vision and deep learning in the following ways:

We unveil a Time-Aware Zero-DCE model for self-supervised low-light enhancement that is time-efficient and requires no labeled images for training. The model is trained with a specially made temporal dataset which consists of pictures taken at regular intervals from sunrise to sunset. This way, the model learns the variations of the light in the real world. One of the major innovations of the method is that it can intelligently apply different enhancement strategies according to the lighting, i.e., it applies different strategies for dark, medium, or bright scenes while keeping the time consistent. A tile-based approach is further applied for high-resolution processing, ensuring that the model is tuned with scalability in mind: no trade-off in speed. Besides, this upgrading process has helped us to bridge the speed-visual quality gap, thus proving that low-light enhancements can be both fast and attractive to the human eye in regular settings. These combined initiatives not only propel the

self-supervised low-light enhancement but also provide a practical implementation that can be used in real-time applications.

Taken together, these contributions push self-supervised low-light enhancement forward and offer a practical tool that can be rolled out in real-time applications.

## 6.3 Recommendations for Future Work

Although our model shows encouraging results in improving low-light images for real-time object detection, many aspects still need to be improved in order to increase its efficiency:

- **Need for Higher Clarity in Real-Time Scenarios:**
  Right now the pipeline tries to balance speed and quality, yet there are moments when live frames still look a touch blurry. Upcoming versions will tweak the enhancement steps so even near-black scenes produce sharp, detailed colors every single time a camera snapshots one.

- **Training with More and Diverse Datasets:**
  A set of fourteen-hour shots feeds the model today but, clearly, lights and devices differ everywhere. To broaden what it knows we will collect datas from different environments, devices, and lighting conditions. The wider library should make the system trustworthy no matter where-or when-it is pointed.

- **Handling Extreme Degradations:**
  Fog, glare, or almost-total darkness stop our model in its tracks, showing scenes that it can't crack yet. Thorough testing it will reveal where it malfunctions, allowing us to develop solutions rather than relying on conjecture and push the boundaries well beyond what they are now.

# Bibliography


[1] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007. DOI: 10.1109/TIP.2007.901238.

[2] G. Singh and A. Mittal, "Various Image Enhancement Techniques - A Critical Review," *International Journal of Innovation and Scientific Research*, vol. 10, no. 2, pp. 267–274, 2 2014, ISSN: 2351-8014. [Online]. Available: http://www.ijisr.issr-journals.org/abstract.php?article=IJISR-14-227-01.

[3] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, and J. Ma, "Msr-net: Low-light image enhancement using deep convolutional network," *ArXiv*, vol. abs/1711.02488, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:23231130.

[4] Y. Zhang, J. Zhang, and X. Guo, *Kindling the darkness: A practical low-light image enhancer*, 2019. arXiv: 1905.04161 [cs.CV]. [Online]. Available: https://arxiv.org/abs/1905.04161.

[5] C. Guo et al., "Zero-reference deep curve estimation for low-light image enhancement," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 1777–1786. DOI: 10.1109/CVPR42600.2020.00185.

[6] W. Wang, X. Wu, X. Yuan, and Z. Gao, "An experiment-based review of low-light image enhancement methods," *IEEE Access*, vol. 8, pp. 87 884–87 917, 2020. DOI: 10.1109/ACCESS.2020.2992749.

[7] X. Li, K. Kou, and B. Zhao, *Weather gan: Multi-domain weather translation using generative adversarial networks*, 2021. arXiv: 2103.05422 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2103.05422.

[8] X. Wei, X.-S. Zhang, S. Wang, Y. Huang, and Y. Li, "Tsn-ca: A two-stage network with channel attention for low-light image enhancement," in *International Conference on Artificial Neural Networks*, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:238408040.

[9] X. Wei et al., "Da-drn: Degradation-aware deep retinex network for low-light image enhancement," *ArXiv*, vol. abs/2110.01809, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:238354336.

[10] W. Kim, "Low-light image enhancement: A comparative review and prospects," *IEEE Access*, vol. 10, pp. 84 535–84 557, 2022. DOI: 10.1109/ACCESS.2022.3197629.

[11] D. Liang et al., "Semantically contrastive learning for low-light image enhancement," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 2, pp. 1555–1563, Jun. 2022. DOI: 10.1609/aaai.v36i2.20046. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/20046.

[12] Y. Cai, H. Bian, J. Lin, H. Wang, R. Timofte, and Y. Zhang, "Retinexformer: One-stage retinex-based transformer for low-light image enhancement," in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 12 470–12 479. DOI: 10.1109/ICCV51070.2023.01149.

[13] S. Malik and R. Soundararajan, "Semi-supervised learning for low-light image restoration through quality assisted pseudo-labeling," in *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2023, pp. 4094–4103. DOI: 10.1109/WACV56688.2023.00409.

[14] M. Siddiqua, S. Brahim Belhaouari, N. Akhter, A. Zameer, and J. Khurshid, "Macgan: An all-in-one image restoration under adverse conditions using multidomain attention-based conditional gan," *IEEE Access*, vol. PP, pp. 1–1, Jan. 2023. DOI: 10.1109/ACCESS.2023.3289591.

[15] Y. Zhu et al., "Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 21 747–21 758. DOI: 10.1109/CVPR52729.2023.02083.

[16] A. Brateanu, R. Balmez, A. Avram, C. Orhei, and C. Ancuti, *Lyt-net: Lightweight yuv transformer-based network for low-light image enhancement*, 2024. arXiv: 2401.15204 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2401.15204.

[17] J. Chang, W. Hou, W. Chen, J. Yan, and K. Cui, "Adaptive image enhancement technology based on bad weather," in *Fourth International Conference on Advanced Algorithms and Signal Image Processing (AASIP 2024)*, D.-I. Curiac and G. Beligiannis, Eds., International Society for Optics and Photonics, vol. 13269, SPIE, 2024, p. 1 326 904. DOI: 10.1117/12.3045552. [Online]. Available: https://doi.org/10.1117/12.3045552.

[18] J. Jiang, Z. Zuo, G. Wu, K. Jiang, and X. Liu, *A survey on all-in-one image restoration: Taxonomy, evaluation and future trends*, 2024. arXiv: 2410.15067 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2410.15067.

[19] Z. Li, Y. Pan, H. Yu, and Z. Zhang, "Dark: Denoising, amplification, restoration kit," *ArXiv*, vol. abs/2405.12891, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:269930369.

[20] J. Wang et al., *Weatherdepth: Curriculum contrastive learning for self-supervised depth estimation under adverse weather conditions*, 2024. arXiv: 2310.05556 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2310.05556.

[21] S. Wang, Q. Tao, and Z. Tang, "Resvmunetx: A low-light enhancement network based on vmamba," Jun. 2024. DOI: 10.48550/arXiv.2407.09553.

[22] Z. Yu, Y. Guo, L. Zhang, Y. Ding, G. Zhang, and D. Zhang, "Improved lightweight zero-reference deep curve estimation low-light enhancement algorithm for night-time cow detection," *Agriculture*, vol. 14, no. 7, 2024, ISSN: 2077-0472. DOI: 10.3390/agriculture14071003. [Online]. Available: https://www.mdpi.com/2077-0472/14/7/1003.

[23] J. "Xu, M. Wu, X. Hu, C.-W. Fu, Q. Dou, and P.-A. Heng, "Towards real-world adverse weather image restoration: Enhancing clearness and semantics with vision-language models," in *Computer Vision – ECCV 2024*, Cham: Springer Nature Switzerland, 2025, pp. 147–164, ISBN: 978-3-031-72649-1.