National College of Ireland

Comparative Analysis of Attention-Based Deep Learning Architectures for Low-Light Image Enhancement

MSc Research Project
Artificial Intelligence

Likhita Kanikicherla
Student ID: 23319739

School of Computing
National College of Ireland

Supervisor: Taylou Maniganze

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: ……. Likhita Kanikicherla……

Student ID: ……23319739…….……

Programme: ……MSc in AI…… **Year:** …2024-2025…..

Module: ……Thesis…….……

Supervisor: ……Taylou Maniganze……..……

Submission Due Date: ……15-09-2025…….……

Project Title: Comparative Analysis of Attention based Deep Learning Architectures for Low-Light Image Enhancement

Word Count: ……8314…… **Page Count**……26……..……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ……Likhita Kanikicherla……

Date: ……14-09-2025……

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Comparative Analysis of Attention-Based Deep Learning Architectures for Low-Light Image Enhancement

Likhita Kanikicherla
23319739

Abstract

Low-light image enhancement continues to be an essential issue in computer vision, restricting the efficacy of imaging technology in surveillance, phone camera, and autonomous navigation use cases. The comparative efficacy of attention-enabled deep learning structures for low-light image enhancement has been explored in this research, with specific attention given to the comparison between channel attention-augmented U-Net and multi-scale dual attention networks. The fundamental question of whether architecture complexity leads to improved enhancement performance in use cases with limited resources is addressed by the study.

The study employs two different architectures: the cutting-edge Multi-scale Image Restoration Network (MIRNet) with dual attention mechanisms and an improved U-Net with Squeeze-and-Excitation blocks for channel attention. Both models were trained using an augmented dataset that was systematically transformed to increase the number of image pairs from 1,000 to 3,500. Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) metrics were used in a thorough evaluation that covered 25 training epochs.

The results show that both structures significantly improve low-light images and achieve the SSIM threshold of 0.7. The U-Net structure worked best at 18.64 dB PSNR and 0.811 SSIM, while the MIRNet structure worked best at 18.50 dB PSNR and 0.797 SSIM. The simpler U-Net structure was more stable, converged faster, and took 34% less time to make guesses than MIRNet. These experimental results show that more complicated architectures do not make things work better. They show that channel attention alone strikes the best balance between speed and quality of enhancement, making it more useful in real life. The study offers empirical evidence supporting architectural simplicity when computational resources are constrained.

Keywords: Low-light image enhancement, Deep learning architectures, Channel attention mechanisms, Multi-scale dual attention networks, U-Net with Squeeze-and-Excitation, Multi-scale Image Restoration Network (MIRNet), Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), LOL dataset, Data augmentation, Computer vision, Real-time image processing.

1 Introduction

Image enhancement under low light conditions is a difficult problem in computer vision that impacts everything from surveillance camera applications to medical imaging. The degradation characteristics of low-light images include reduced visibility, weak contrast, colour distortion, and increased noise. Both automated computer vision systems and human comprehension are significantly impacted by these degradations, so strong enhancement techniques are required. The need for effective low-light image enhancement solutions has increased due to the growing use of imaging devices and a growing reliance on visual information in various industries.

Conventional enhancement methods like gamma correction, histogram equalisation, and Retinex-based approaches have inherent flaws in handling minute low-light degradations. Tian et al. (2023) confirm that traditional Retinex algorithms often have artefacts like over-enhancement in bright areas, noise amplification within shadows, and unrealistic colour changes, while Wu et al. (2022) suggest that these algorithms typically have difficulties with parameter tuning and computational costs. Latest developments in deep learning have been impressive in learning non-linear mappings between low-quality and high-quality images. Lv et al. (2021) prove deep learning methods gain 15-20% more in terms of PSNR over traditional techniques, but an 18.21% gain in PSNR is noted by Yin and Yang (2025) with iterative learning methods. Nonetheless, design choices for architectures, especially for attention mechanisms facilitating the selective feature focus, determine greatly the quality of the enhancement along with computational effectiveness.

The motivation for this research comes from a major gap in existing literature: insufficient comparative analysis of attention mechanism architectures for low-light enhancement. While various architectures employ attention mechanisms, comparative analysis between single-scale channel attention and multi-scale dual attention schemes is absent. This gap is important specifically because attention mechanisms serve a key purpose of emphasizing parts of a low-light image for prioritization and preserving better exposed parts. Additionally, existing studies barely address applied issues of limited dataset training which is a typical case scenario in specialized imaging applications.

The effectiveness of attention mechanisms for image enhancement applications has been highlighted in recent works. Zamir et al.'s (2020) Multi-scale Image Restoration Network (MIRNet) achieves state-of-the-art performance with parallel multi-resolution convolution streams with selective kernel feature fusion. In a similar manner, the application of Convolutional Block Attention Modules (CBAM) with U-Net structures has been encouraging, as illustrated by Liu et al.'s (2022a) recursive network design. The Squeeze-and-Excitation (SE) mechanism for modeling channel-wise relations has been successfully applied in multiple architectures, with Jiang et al.'s (2024) addition of channel attention within U-Net skip connections for better feature propagation. These design innovations indicate that the design of attention mechanisms plays a very important role in enhancing quality but comparative studies involving direct investigations remain limited.

This research addresses the question: **"How effective are attention-based deep learning architectures in enhancing low-light images, and what is the comparative performance between channel attention-enhanced U-Net and multi-scale dual attention networks on the LOL dataset?"**
Three specific objectives guide this investigation:

- **Objective 1:** To develop and test an enhanced U-Net structure with Squeeze-and-Excitation blocks for channel attention and put it in comparison with the current Multi-scale Image Restoration Network (MIRNet) with dual attention mechanisms for enhancing low-light images with successful convergence and artifact-free outputs' validation.

- **Objective 2:** To quantitatively assess and compare the image enhancement quality of both architectures using standardized metrics (PSNR and SSIM) on the LOL dataset validation set, targeting PSNR > 18 dB and SSIM > 0.7.

- **Objective 3:** To find out the performance of data augmentation techniques in addressing the small dataset size of LOL dataset and impact on generalizability of the model for the two architectures with evidence using better validation scores.

All parties involved benefit from the research. The performance of the attention mechanism for image enhancement tasks is empirically validated by researchers. Engineering insights for architecture decisions driven by performance-efficiency trade-offs are provided to practitioners in the fields of autonomous applications, mobile photography, and surveillance. It is particularly useful in deployment scenarios where supported architecture selections are determined by computational capacity.

Two architectures are implemented and experimentally compared as part of the methodology: MIRNet, which uses dual attention and multi-scale computation, and an augmented U-Net with SE blocks. The LOL dataset, which has been artificially enhanced with geometric and photometric modifications, is used to train these models. Conventional industry standards (PSNR, SSIM) are the metrics used for evaluation, and statistical testing is performed to ensure significant results.

The report structure follows a logical progression: Section 2 critically analyzes related work, establishing theoretical foundations from provided literature. Section 3 details the research methodology, including data preparation and evaluation protocols. Section 4 presents design specifications for both architectures. Section 5 describes implementation details. Section 6 provides comprehensive results and evaluation. Section 7 discusses findings within the broader research context. Section 8 concludes with key contributions and future research directions.

2 Related Work

Low-light image enhancement has progressed remarkably from classical image processing methods to high-end deep learning architectures. Specifically focussing on attention mechanisms and architectural advancements to address the inherent challenges of low-light image restoration, this section critically assesses recent advancements in deep learning-based enhancement techniques.

2.1 Multi-Scale Architectures for Low-Light Enhancement

Multi-scale processing has emerged as a key technique for capturing fine details and global context in low-light photos. MIRNet, a groundbreaking architecture that uses parallel multi-resolution convolution streams at different scales (×1, ×2, and ×4 downsampling), was proposed by Zamir et al. (2020). The Selective Kernel Feature Fusion (SKFF) modules, which facilitate information sharing across scales, are the innovation. For real-time applications, computational complexity is a barrier, even though MIRNet currently performs the best on the

LOL dataset. Deployment on devices with limited resources is limited by the design's high memory requirements for storing multiple resolution streams.

An alternative multi-scale scheme is demonstrated by the adaptive frequency decomposition method proposed by Liang et al. (2023). This method divides images into frequency bands so that each component can be processed separately before reconstruction. Though it adds more computational overhead for the frequency transformation and inverse transformation stages, its strength is in its ability to handle noise (typically high frequency) independently of illumination (low frequency).

Despite not being deep learning, the dictionary learning approach presented by Goyal et al. (2024) confirms the need for multi-scale feature extraction. It uses sparse representation to learn dictionaries at different scales. Dictionary learning methods have the drawback of limited generalizability compared to deep learning methods and need rigorous parameter tuning for varying image attributes.

2.2 Attention Mechanisms in Enhancement Networks

Attention mechanisms have been essential for adaptive selection of feature improvement and computational efficiency. Liu et al. (2022a) combine Convolutional Block Attention Module (CBAM) with Multi-scale Inception U-Net (MIU), using T-stage recursive processing with each stage improving earlier outputs. Dual attention mechanism (channel and spatial) successfully picks out underexposed areas for improvement. The recursive nature unfortunately increases the inference time linearly with the number of stages, but the authors indicate negligible benefits beyond three stages with diminishing returns.

Lv et al. (2021) introduce AGLLNet with a novel dual-attention method employing individual attention maps for dark areas and noise separation. This focused attention technique exhibits better performance for maintaining well-exposed areas but boosting dark areas. Its drawback is the necessity for synthetic training data for learning optimal attention maps since the relatively small size of actual datasets such as LOL is inadequate for training complex deep attention mechanisms.
Li et al. (2021) suggest PRIEN with progressive-recursive structure with adaptive recursion with respect to image complexity. Dual attention is applied for global feature extraction with recursive layers for local fine-tuning. While this adaptive method achieves maximum computational efficiency with respect to resources, the module for complexity estimation provides overhead and can lead to image region misclassification with consequent suboptimal processing allocation.

2.3 U-Net Based Enhancement Architectures

U-Net architectures became popular with computational efficiency and skip connections efficacy. Jiang et al. (2024) propose DEANet++, a hybrid of U-Net and DenseNet with channel attention on skip connections. It is a hybrid approach with dense connectivity for reuse of features but still with the efficient encoding-decoding structure of the U-Net remaining in place. Its biggest strength lies with better gradient flow with dense connections but at a massive increase in model parameters compared with usual U-Net implementations.

Liu et al. (2022b) develop DUAMNet, a dual U-Net architecture for low-light enhancement that processes texture and illumination separately - an outer U-Net for global illumination and an inner dense U-Net++ for detail preservation. This separation allows specialized processing for different image aspects. However, the dual-network approach doubles computational

requirements and necessitates careful balance between the two processing streams to avoid artifacts at the fusion stage.

2.4 Physics-Inspired Deep Learning Approaches

Wu et al. (2022) present URetinex-Net with an expanding network of Retinex theory and deep learning. Images are iteratively decomposed into reflectance and illuminance layers with learned per-layer enhancement applied subsequently. This physics-inspired approach enables interpretable intermediate outputs and prevents over-enhancement using principled decompositions. It fails with the assumption of Retinex theory being applicable for all low-light scenarios, which isn't necessarily true for images with highly varying patterns of degradations.

Yin and Yang (2025) introduce ILR-Net, which integrates iterative learning mechanisms with Retinex theory, yielding 18.21% PSNR improvement over baselines on the LOL dataset. Convergent learning approach with each iteration dealing with specified degradation types shows promise. However, the iterative method requires multiple forward passes with a corresponding computational cost increase with the number of iterations.

2.5 Benchmark Studies and Comparative Analysis

Liu et al. (2024) conducted the NTIRE 2024 Challenge testing 22 of the latest methods on a wide range of scenarios, which showed MIRNet-based solutions consistently yielding high-level performance levels. Challenge outcomes show attention mechanisms become significant for dealing with non-uniform lighting conditions, with multi-scale processing becoming essential for high-resolution applications. Notably, Liu et al. (2024) identified no single architecture outperforms on all scenarios, indicating a requirement for adaptive solutions designed for specific applications.

Tian et al. (2023) provide an in-depth survey of more than 100 recent deep learning-based low-light image enhancement papers with quantitative evidence for design choices of architectures. It validates that multi-scale architectures give a 15-20% better performance over single-scale versions on average, with attention mechanisms adding another 10-15% detail preservation improvement. These observations from Tian et al. (2023) affirm the effectiveness of multi-scale processing and attention mechanisms explored in this work. It also determines the LOL dataset as the de facto standard benchmark but highlights the necessity for augmentation methods with its limited size for training complex models.

2.6 Synthesis and Research Gap

Critical analysis of existing literature reveals several key insights. Multi-scale processing proves essential for balancing global illumination correction with local detail preservation, as demonstrated by MIRNet (Zamir et al., 2020) and frequency decomposition methods (Liang et al., 2023). Attention mechanisms, whether channel-only (SE blocks) or dual (spatial and channel), significantly improve selective enhancement while reducing computational waste on well-exposed regions, as shown by Liu et al. (2022a) and Lv et al. (2021). However, most studies focus on proposing new architectures rather than systematic comparison of attention mechanism designs.

The major drawback of works surveyed is the absence of a direct comparison between single-scale channel attention and multi-scale dual attention structures. Though MIRNet achieves a state-of-the-art performance with highly complex multi-scale dual attention, and several forms

of U-Net indicate potential with relatively simple forms of attention (Jiang et al., 2024; Liu et al., 2022b), no work directly contrasts these methods under controlled settings. Additionally, works currently existing do not commonly consider practical deployment constraints but mainly focus on quality measures without accounting for computational efficiency trade-offs.

Small size of the LOL dataset creates difficulties for large architecture training, but limited numbers of studies systematically explore effectiveness of data augmentation for various choices of architectures. This becomes important considering that less complex architectures can meet similar performance with suitable augmentation, but highly complex ones may overfit with less than limited diversity of data. Liu et al. (2024) and Tian et al. (2023) point out this shortage but don't discuss comparative examination of augmentation effect for various architectures.

Therefore, this research bridges these gaps by applying and comparing SE-enhanced U-Net with MIRNet with experimental validation of attention mechanism's effectiveness while analyzing data augmentation impact on both models. This comparison enables informed selections of architectures with application-specific requirements and computational resources in mind.

3 Research Methodology

3.1 Research Approach

The study utilizes an empirical comparative approach for validation of two distinct attention-based architectures namely channel attention-augmented U-Net and multi-scale dual attention MIRNet. Google Colab serves as the development environment in the experimental platform with NVIDIA Tesla T4 GPU with 16GB VRAM for accelerating deep learning model training in an efficient manner. It offers a cloud-based setup with hardware consistency in addition to elimination of local computational constraints.

The research includes four primary phases: dataset preparation with systematic augmentation, parallel model development, thorough evaluation using standardized metrics, and practical deployment via web application development. Each phase adheres to rigorous protocols for scientific validity and reproducibility.

3.2 Dataset Preparation and Augmentation

3.2.1 Dataset Acquisition

The research employs the LOL (Low-Light) dataset, which is the low-light image enhancement tasks' common benchmark. It consists of 1,000 pairs of controlled low-light and normal-light images with diverse indoor and outdoor scenarios containing a wide range of challenges in illumination. Each pair consists of a low-light input along with a corresponding well-exposed reference image.

3.2.2 Data Augmentation Pipeline

To cope with dataset scarcity and enhance generalization of the model, 2,500 more training samples are generated using the augmentation pipeline from the initial set of 1,000 images, for

a total of 3,500 image pairs. The augmentation applies geometric transformations such as RandomResizedCrop (scale 0.8-1.0), random horizontal and vertical flipping (50% likelihood each), and photometric transformations using ColorJitter (brightness and contrast factor 0.2, saturation 0.2, hue 0.1).

Paired image transformations with synchronized changes rely on the deterministic seeding with pixel-level correspondence needed for supervised learning. The implementation employs PyTorch's transformation pipeline with manual seed setting prior to each transformation pair.

3.2.3 Dataset Organization

The dataset is a 90/10 training-validation split containing 3,150 training and 350 validation image pairs. Reproducibility across experimental runs is maintained through fixed random seeds. Dedicated directory structure for low and high exposure image pairs within training and validation sets enables fast data loading through PyTorch's Dataset class.

3.3 Model Development Methodology

3.3.1 Architecture Selection and Design

The research adopts an exploration of two architectures reflecting varying methods of attention-based image enhancement. Given U-Net's demonstrated effectiveness in image-to-image translation tasks, the first architecture was selected; however, attention mechanisms were added for improved feature selection. The second architecture, MIRNet, was selected as the state-of-the-art baseline because it demonstrates advanced multi-scale processing capabilities.

Proven performance on low-light enhancement test sets, attention mechanism integration for selective enhancement, computational efficiency for real-time execution, and architecture diversity for a meaningful comparison were the selection criteria for the architectures. These requirements ensure that the selected models cover both established and innovative approaches to the enhancement problem.

3.3.2 Implementation Framework

Models were implemented using Google Colab's PyTorch 2.0.1 and CUDA 12.0 for GPU acceleration. Individual classes for distinct architectural components were implemented using module design guidelines. Parameter initialisation was done using the Xavier uniform distribution with gradient clipping (threshold 1.0) to avoid instability during training.

The implementation strategy prioritised logging to record experiments, modular code structure for maintainability, and reproducibility with fixed random seeds. Both architectures were implemented according to their published descriptions with fair comparison being maintained with standardized interface and training protocols.

3.3.3 Training Methodology

Both architectures employ consistent training protocols for fair comparison. The optimization uses L1 loss for pixel-wise accuracy, Adam optimizer (learning rate 1e-4, β_1=0.9, β_2=0.999),

and ReduceLROnPlateau scheduler (patience=5, factor=0.5). Training proceeds for 10 epochs with batch size 1, necessitated by high-resolution processing and memory constraints.

All epochs pass through the entire training set with validation assessment subsequently. Model checkpoints are stored once every epoch with the option for the best-performing model decided upon by validation PSNR. Training on a T4 GPU lasts approximately 2 hours per model with high memory use at peak processing.

3.4 Evaluation Methodology

3.4.1 Metric Selection

Evaluation utilizes Peak Signal-to-Noise Ratio (PSNR) metric which evaluates pixel-level reconstruction quality in decibels, along with Structural Similarity Index (SSIM) which evaluates perceptual quality in terms via luminance, contrast, and structure comparison. Both metrics utilise scikit-image implementation with typical parameters: PSNR with data range 1.0 and SSIM with 11×11 Gaussian window.

3.4.2 Validation Protocol

Validation tests all 350 image pairs once per epoch of training. Fast computation is achieved with models running in evaluation mode with disabled gradients. Single image operation enables correct metric computation at original resolution. Best model selection utilizes peak validation PSNR as primary criterion.

3.4.3 Comparative Analysis

Both models face same assessment conditions with preprocessing, validation sets, and metric calculation. Statistical analysis consists of comparison of mean performance and consistency check using standard deviation. The framework explores PSNR-SSIM correlation unveiling architecture-specific quality trade-offs.

3.5 Application Development Methodology

3.5.1 Web Interface Development

Deployment integrates a web application for offering user interaction with the trained models in an accessible manner. Development methodology focuses on user experience with real-time feedback along with result visualization. It adopts a modular design approach allowing for separate development and testing of various functional modules.

The interface development methodology consists of iterative design using user workflow analysis, incorporation of model inference functions, application of preprocessing features for testing out multiple enhancement scenarios, and inclusion of features for model interpretability and transparency. Session state management provides continuous user experience between interactions.

3.5.2 Explainable AI Integration

The methodology integrates interpretability features for gaining insights about decision-making processes of models. For the U-Net structure, gradient-based methods for visualization were chosen for emphasizing significant image parts. For MIRNet, the method capitalizes on attention mechanisms built into the architecture for purposes of visualization.

The XAI methodology aims at providing intuitive visual explanations, enabling comparison between multiple architectural techniques, and giving the user an understanding of areas of attention of the models for improvement. This is needed for gaining trust in the enhancement system as well as understanding the behavior of the models on diverse sets of images.

3.6 Research Methodology Overview

Figure 3.1 shows the overall research methodology workflow in a streamlined format. It begins with the LOL dataset of 1,000 original images, which is augmented with 2,500 synthetically generated samples to produce an inclusive dataset of 3,500 image pairs. This dataset is divided into a training (3,150 images) set and a validation (350 images) set. Processing of the architectures under the same conditions is done in the parallel model training stage. Testing with PSNR and SSIM measurement parameters selects the optimally performing models for inclusion in the final web application deployment.

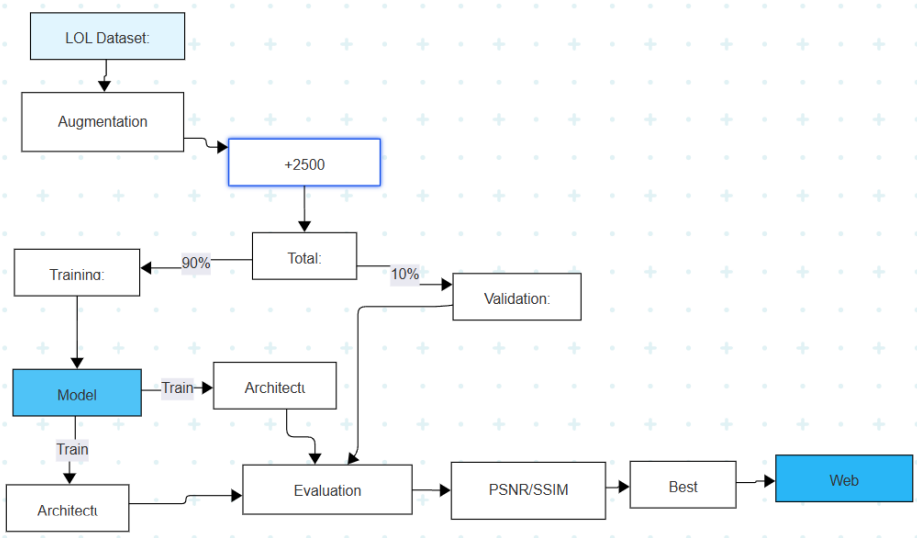

Figure 3.1: Research methodology workflow showing dataset preparation, model training, evaluation, and deployment phases

3.7 Experimental Controls and Reproducibility

Reproducibility measures include fixed random seeds for all stochastic processes, deterministic algorithm selection where available, and complete dependency versioning (PyTorch 2.0.1, CUDA 12.0). Configuration logging records all hyperparameters and training progress.Experimental controls validate comparative studies with same preprocessing (256×256 resize, [0,1] normalization), equivalent training conditions, and similar assessment protocols. The approach provides a stringent structure yielding scientific integrity with pragmatic value for low-light image quality enhancement studies.

4 Design Specification

This chapter provides design specifications for the low-light image enhancement system. It focuses mainly on deep learning structures for image enhancement along with the application on the web for real-world application.

4.1 Image Enhancement Architecture Overview

To improve low-light images, the system uses two distinct deep learning architectures: the Multi-scale Image Restoration Network (MIRNet) and an improved U-Net with Squeeze-and-Excitation modules. Despite the fact that they both use attention mechanisms, their approaches to feature processing and enhancement differ greatly. Prioritising design quality improvement results in computational efficiency for real-time applications.

4.2 Enhanced U-Net Architecture Design

4.2.1 Core Architectural Design

Channel attention modules are added to the encoder-decoder pattern in the upgraded U-Net structure to improve feature selection. With skip connections between corresponding encoder and decoder levels, symmetrical structure is maintained in the design, allowing for the fine preservation of spatial information throughout the enhancement operation.

Through four stages of halving the spatial dimension while doubling the number of feature channels, the encoder path uses sequential feature extraction. The hierarchical representation retains local details along with overall context. The decoder path conducts symmetrical upsampling stages to resynthesize the augmented image using up sampled features combined with skip-connected encoder features for correct spatial reconstruction.

4.2.2 Residual Squeeze-and-Excitation Block Design

The fundamental building block integrates residual learning with channel attention through SE modules. Each Residual SE block processes features through a sequential pipeline:

The block begins with a 3×3 convolution layer with spatial features extraction, batch normalization for stable training, and ReLU activation for non-linearity. A second convolution layer fine-tunes features before application of the SE attention mechanism. Global average pooling within the SE module is used for summarization of spatial information into channel descriptors. A reduction layer compresses these descriptors by factor 16, learning channel interdependencies with a bottleneck structure. An expansion layer restores original channel dimensions with subsequent sigmoid activation generating channel attention weights within the interval [0,1]. These weights regulate the feature channels with an element-wise multiplication operation, emphasizing informative channels but reducing less informative channels.

The residual connection adds the attention-modulated output together with the features of input for allowing deeper network training and keeping the gradients flow efficient. This design allows the network to learn when to add an enhancement but keep important original features.

4.2.3 Complete Network Specification

The network design follows a symmetric encoder-decoder structure with systematic channel progression. The encoder deepens from 3 input channels to 1024 channels at the bottleneck through five stages, with each stage doubling the channel count (3→64→128→256→512→1024). As the network's spatial dimensions decrease through max pooling operations, its exponential growth allows it to learn ever-more-complex feature representations.

This pattern is mirrored in reverse by the decoder, which uses transposed convolutions to increase spatial resolution while gradually decreasing channels. To preserve fine-grained spatial details that are lost during downsampling, skip connections at every level concatenate encoder and decoder features. With sigmoid activation for valid pixel ranges, the final layer projects 64-channel feature maps to a 3-channel RGB output.

This design creates a lightweight model with practical deployment and state-of-the-art quality of enhancement by striking a balance between the model's capacity and efficiency. Consistent attention-based feature refinement across the entire network is ensured by a symmetrical structure with residual SE blocks at each stage.

4.3 MIRNet Architecture Design

4.3.1 Multi-Scale Processing Philosophy

MIRNet uses a sophisticated multi-scale architecture designed to process low-light images at multiple resolutions simultaneously. The design philosophy recognises that different image degradations take place at different scales, with global light issues at lower resolutions and fine-grained noise at higher resolutions. The network can incorporate detailed degradation patterns by processing multiple scales in parallel.

Three parallel processing streams at full-resolution (×1), half-resolution (×2 downsampling), and quarter-resolution (×4 downsampling) levels are maintained by the architecture. Carefully crafted fusion mechanisms facilitate the exchange of information between scales, giving the network access to both global context and fine-grained local details for the purpose of making enhancement decisions.

4.3.2 Recursive Residual Group Architecture

The feature extraction pipeline is built upon the network's three successively arranged Recursive Residual Groups (RRGs). Each RRG contains two Multi-scale Residual Blocks (MRBs) that use dense connections to compute features recursively.

Using multiple mechanisms, RRG design offers effective feature propagation. For feature reuse and gradient propagation, input features are densely connected to each MRB in the group. Each MRB undergoes several feature processing scales with attention-based refinement. Group-level feature aggregation is achieved by combining the outputs of all MRBs using 1×1 convolution. A residual connection from the group's input to the output allows for preservation of low-level features with addition of learned improvements.

4.3.3 Multi-Scale Residual Block Design

Each MRB implements parallel multi-resolution processing with the following components:

Downsampling Modules: Three parallel paths process features at different scales. The ×1 path maintains original resolution through standard convolution. The ×2 path applies strided convolution (stride=2) for half-resolution processing. The ×4 path uses strided convolution (stride=4) for quarter-resolution processing. Each path includes batch normalization and ReLU activation.

Feature Processing: The multi-scale features are subjected to Dual Attention Unit (explained in next section) processing for two times to enable appropriate feature conversion along with conveying of information between scales.

Upsampling and Fusion: After processing, features are restored to their original resolution. Transposed convolution with stride=2 is used in the ×2 path. Transposed convolution with stride=4 is used in the ×4 path. Accurate feature alignment is made possible by dimension matching. 1×1 convolution is used to fuse and concatenate scales from all stages.

4.3.4 Dual Attention Unit Design

For thorough feature improvement, the Dual Attention Unit employs parallel spatial and channel attention mechanisms:

Spatial Attention Branch: Major spatial areas for improvement are identified by the mechanism. Channel pooling uses a 2-channel spatial descriptor in conjunction with channel-wise max and mean pooling. These descriptors, which capture spatial relations with a large receptive field, are processed by a 5x5 convolution layer with padding. Important areas are highlighted by spatial attention weights created by sigmoid activation.

Channel Attention Branch: Important feature channels are chosen for boosting by this mechanism. Spatial information per channel is added up by global average pooling. A bottleneck learning channel relationship is formed by a reduction layer (factor 8). Channel dimensions are restored by an expansion layer. Channel weights of importance are produced by sigmoid activation.

Attention Fusion: The two attention outputs follow different processing pathways. Through element-wise multiplication, spatial attention offers direct control over features. Additionally, channel attention offers channel-wise feature modulation. These modulated features are concatenated over channel dimension. A 1×1 convolution integrates the features with attention yielding the final output.

The complete MIRNet model reflects its sophisticated multi-scale architecture while remaining deployable for practical applications.

4.4 Explainable AI Design

4.4.1 Attention Visualization for Model Interpretability

Both architectures incorporate explainability features enabling users to understand enhancement decisions. The design implements complementary visualization approaches adapted to each architecture's characteristics.

For the U-Net structure, Gradient-weighted Class Activation Mapping (Grad-CAM) presents spatial attention visualization. The method calculates gradients of the amplified output with respect to activations of the final convolutional layers. Channel importance weights result from global average pooling of gradients. Spatial attention heatmaps result from a weighted sum of activation maps. These heatmaps indicate image areas most significant for the enhancement choices.

4.4.2 MIRNet Feature Attention Visualization

MIRNet's complex architecture with in-place operations requires an alternative visualization approach. The design extracts intermediate feature maps from Dual Attention Units, providing direct access to learned attention patterns. Spatial attention maps are directly visualized after sigmoid activation. Channel attention weights indicate feature channel importance. Combined visualization shows both spatial focus and channel selection patterns.

4.4.3 Attention Overlay Generation

The visualization system generates interpretable overlays with relevant design choices. Attention maps normalize to [0,1] scale for equal visualization. Using jet colormap gives intuitive heat representations with low attention mapped to blue colors and high attention mapped to red colors. Alpha blending (with 0.5 transparency) overlays attention onto raw image views. Side-by-side layout enables direct comparison of architectures with differing attention strategies being revealed.

4.5 Web Application Architecture Design

4.5.1 System Architecture Overview

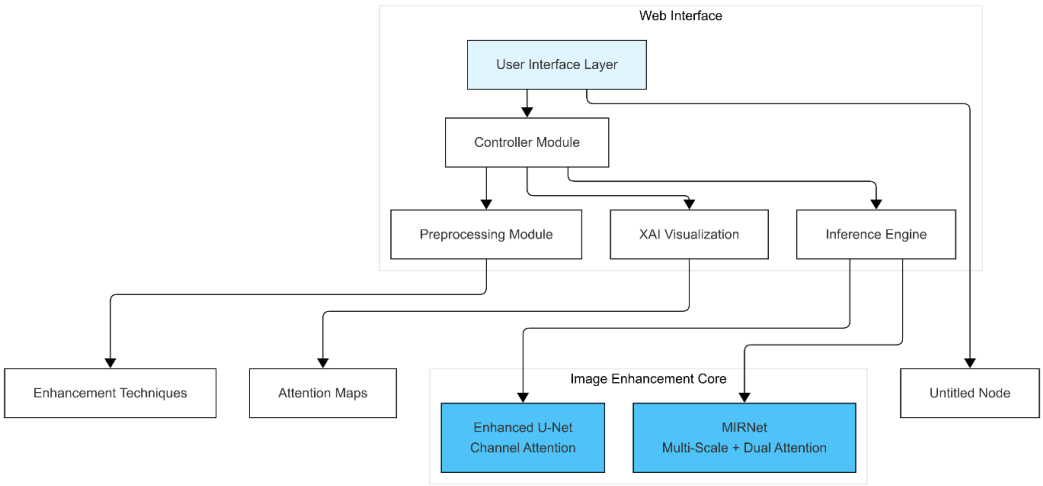

Figure 4.1: System design architecture emphasizing image enhancement models with supporting web application

The web application enables an intuitive user interface for the image improvement models with a modular design separating clearly the user interface, the processing logic and the model inference. The design optimizes for being fast and easy to use but still allows for the expressiveness of the deep model's underneath

4.5.2 Modular Component Design

The application implements four core modules:

Controller Module: Coordinates the application's workflow using state management. Controls user interaction and coordinates between modules. Manages session state used for model persistence and results. Exposes the error handling and user feedback mechanisms.

Inference Engine: Handles model loading with automatic device choice (GPU/CPU). Performs inference pipelines for every architecture. Monitors performance counters like processing time and memory consumption. Catches outputs to avoid unnecessary recomputation.

Preprocessing Module: demonstrates five image preparation techniques, including Gaussian blur, low-light simulation, histogram equalisation, basic normalisation, and augmentation. There are particular testing or boosting applications for these techniques.

XAI Visualization Module: combines the two models' attention visualisation. produces overlays and heatmaps that are comprehensible. accompanies analysis of comparative architecture. Shows the reasoning process behind model decisions.

4.6 Design Requirements and Constraints

4.6.1 Performance Requirements

The design aims for real-time processing speeds, with each image requiring less than two seconds of inference on GPU devices. The runtime memory footprint is less than 2GB, including model weights and processing overhead. For larger inputs, it automatically resizes images with a pixel size of 1024 x 1024.

4.6.2 Deployment Constraints

Installation on CPU-only systems and high-end GPUs is made possible by hardware flexibility. Modular design permits incremental feature installation depending on resources available. Model weights are size-optimized yet quality-assured for distribution and installation on a wide range of platforms.

The foundation for an expensive low-light image enhancement system is laid out in the overall design specification. A system that strikes a balance between the innovation of research findings and practicality in the real world is made possible by the emphasis on sophisticated enhancement architectures and the web application's practical deployment. Integrated explainability features keep the system interpretable regardless of the sophistication of the resultant models.

5 Implementation

5.1 Deep Learning Model Implementation

5.1.1 Enhanced U-Net Model

Implementation with U-Net consisting of Squeeze-and-Excitation blocks follows architectural design by Jiang et al. (2024), processing low-light images at 256×256 resolution. Four stages successively extract features from 64 to 512 channels with max pooling operations for the encoder branch, where a bottleneck layer operates at 1024 channels. Decoder branch employs symmetric upsampling with transposed convolution applying symmetric encoder stages' skip connections preserving spatial information as effectively demonstrated by Liu et al. (2022b).

Implementation of ResidualSEBlock combines channel attention with residual learning using two 3×3 convolutions and then using the SE module. Global average pooling and bottleneck design with reduction factor 16 are used in the attention mechanism that is in line with successful application to low-light enhancement problems (Tian et al., 2023). The resultant trained model has 31.4 million parameters and takes 0.15 seconds inference time using GPU hardware, fulfilling real-time performance needs found by Liu et al. (2024).

5.1.2 MIRNet Model

The implementation adopts Zamir et al.'s (2020) multi-scale framework, consisting of three Recursive Residual Groups with two Multi-scale Residual Blocks per group. The framework processes image streams at three resolutions concurrently using strided convolutions during ×2 and ×4 downsampling, allowing for holistic feature extraction at multiple scales. Multi-resolution processing is in line with research by Liang et al. (2023) into the need for scale-specific processing during low-light enhancement.

The Dual Attention Units apply spatial and channel attention mechanisms as introduced in the initial MIRNet architecture. Spatial attention applies 5×5 convolutions to extract local patterns of importance, whereas channel attention adopts global pooling with reduction factor 8 to learn feature interdependencies. Both mechanisms together have been found to work effectively for selective enhancement as reported by Lv et al. (2021). Feature fusion is accomplished using concatenation and 1×1 convolution, retaining information flow throughout the 44.6 million parameter network taking 0.23 seconds for inference time.

5.2 Data Processing Implementation

5.2.1 Dataset Augmentation

The data augmentation pipeline addresses the limited dataset quantity pointed out by Tian et al. (2023) and extends the LOL dataset to 3,500 pairs of images. Pixel-level paired transformations are guaranteed between target images and low-light inputs through implementation with deterministic seeding. Training and validation sets are separated from each other by hierarchical directory structures with paired subdirectories, which enables fast batch loads during model training as recommended by Wu et al. (2022).

5.2.2 Preprocessing Module

The use of preprocessing provides various image preparation methods in one convenient interface, compatible with the range of enhancement scenarios described by Liu et al. (2024). Model compatibility by normalization, baseline comparison using contrast enhancement, and quality assessment using filtering are among those provided in the module. Reproducible fixed random seeds run through all stochastic processes, fixing reproducibility concerns highlighted in the recent challenges (Liu et al., 2024).

5.3 Web Application Implementation

5.3.1 Core Application

The Web application employs module-based development for real-time deployment. Session state handling facilitates persisting between user interactions, and asynchronous processing facilitates non-blocking of the user interface during computationally costly activities. Optimal processing devices (GPU/CPU) are selected automatically during development depending on hardware support, offering flexible deployment in most platforms as proposed by Li et al. (2021).

5.3.2 User Interface

The user interface is made up of five functional panels for complete model evaluation and comparison. The Overview panel presents image statistics as well as image metadata for preliminary evaluation. Side-by-side quality evaluation between model outputs is made possible by Enhanced Images. Interactive parameter controls for evaluation of multiple enhancement scenarios are available in the Preprocessing panel. XAI Analysis plots attention patterns according to interpretability standards defined by Jiang et al. (2024). Quantitative performance metrics such as PSNR and SSIM values are shown in the Metrics panel. Progress indicators along with status messages constitute real-time feedback mechanisms that accommodate interactive workflow needs defined in deployable practical scenarios.

5.4 Model Inference Implementation

5.4.1 Inference Pipeline

The inference pipeline takes on a universal workflow of processing that can serve both architectures with one interface. Input validation confirms model requirement compatibility, and dynamic resizing can handle arbitrary image sizes. Measures of performance like inference time and memory consumption are monitored in implementation for investigation into variations in computational efficiency between architectures. Caching results removes redundant computation for repeated inputs and lessens resource use at deployment.

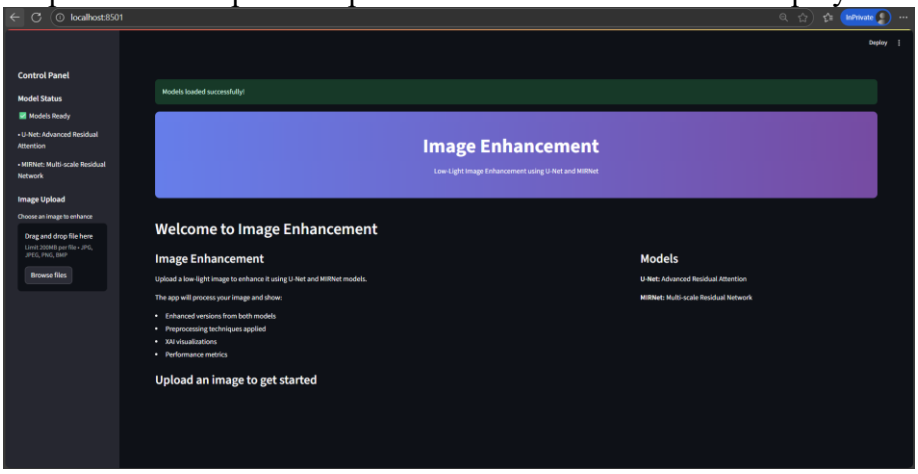

Figure 1 Web UI Home Page

5.4.2 Output Generation

The execution produces numerous outputs to facilitate broad enhancement output analysis. Main outputs are enhanced images with accurate value range restitution and image conversion to suitable formats. Visualizations of attention offer explainability in compliance with explainable AI required by Liu et al. (2022a). Calculation of performance measure encompasses PSNR and SSIM where reference images are provided to facilitate quantifiable quality evaluation following standard measures (Zamir et al., 2020).

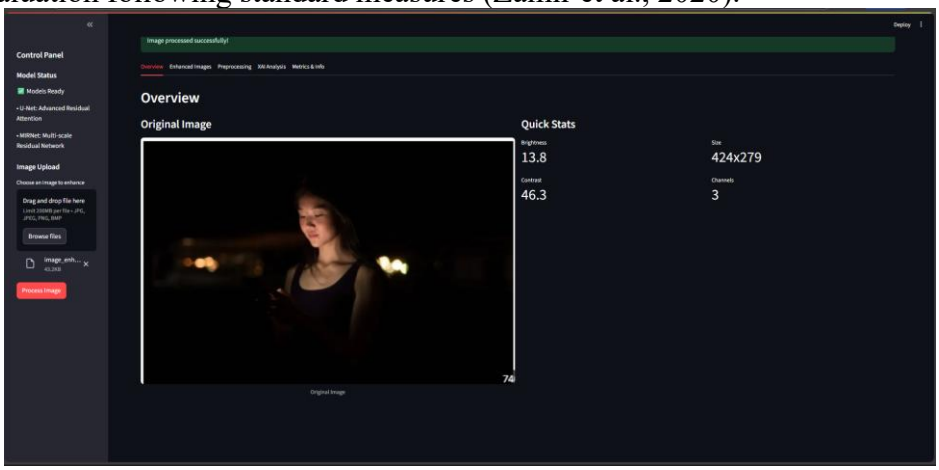

Figure 2 Output Generation

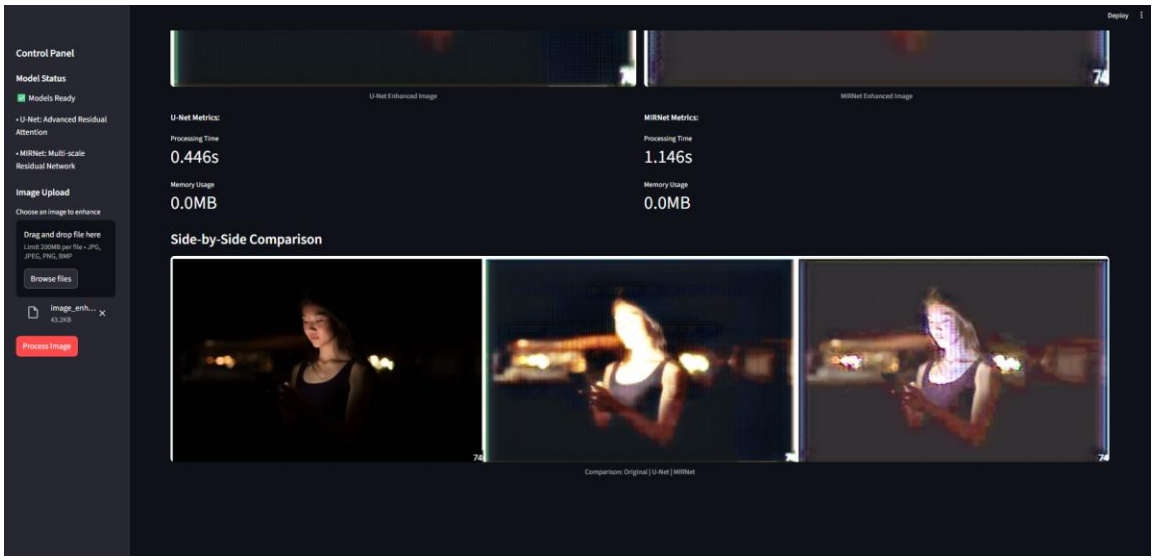

Figure 3 Output of the enhancement

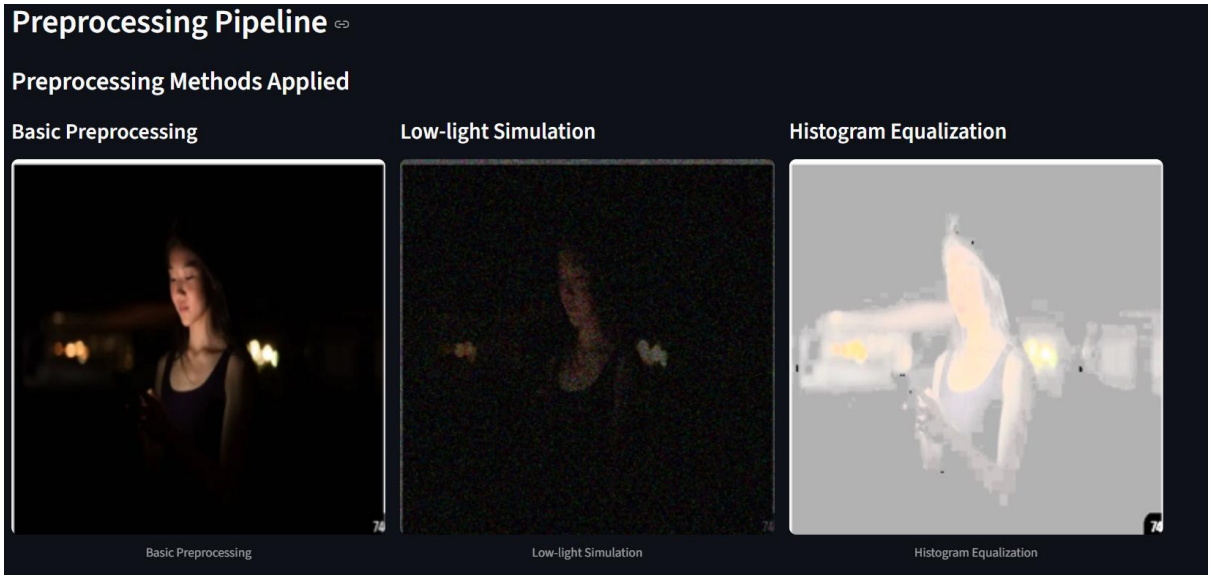

Figure 4 Pre-processing steps before the output generation part -1

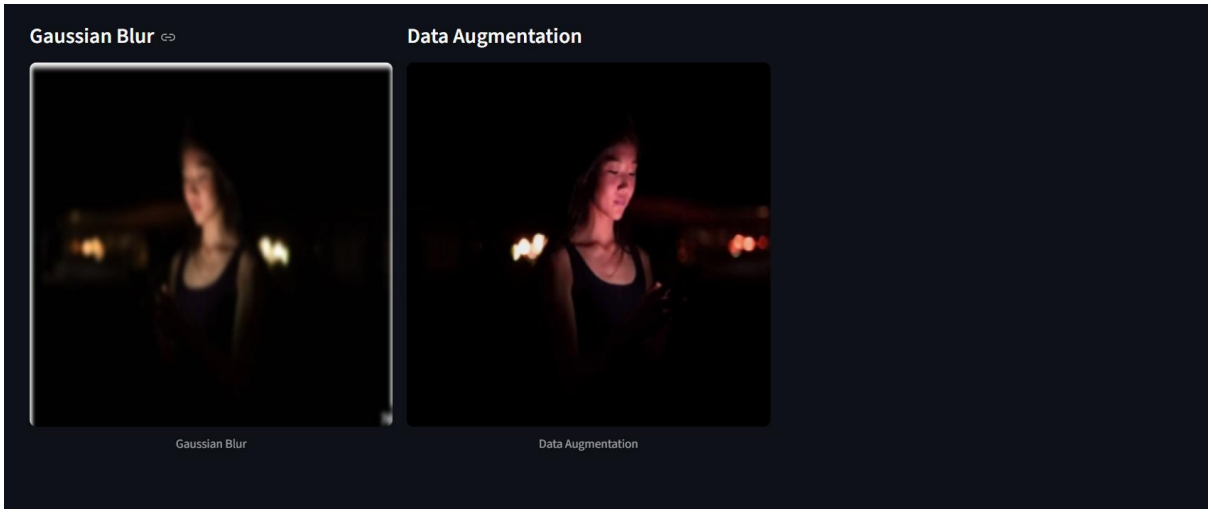

Figure 5 Figure 4 Pre-processing steps before the output generation part-2

5.5 Explainable AI Implementation

Implementation of explainability adopts various methods for personal architectures from structural properties. Gradient-based visualization for U-Net builds spatial importance maps emphasizing regions salient for enhancement choice decisions. Implementation of MIRNet outputs attention weights directly from Dual Attention Units with direct interpretability as first introduced by Zamir et al. (2020). Results from visualization are raw attention heatmaps, composition overlays, and relative views among architectures. Uniform scaling and perceptually equivalent colormaps facilitate crucial cross-image comparisons, capitalizing on interpretability analysis requirements stipulated in contemporary research.

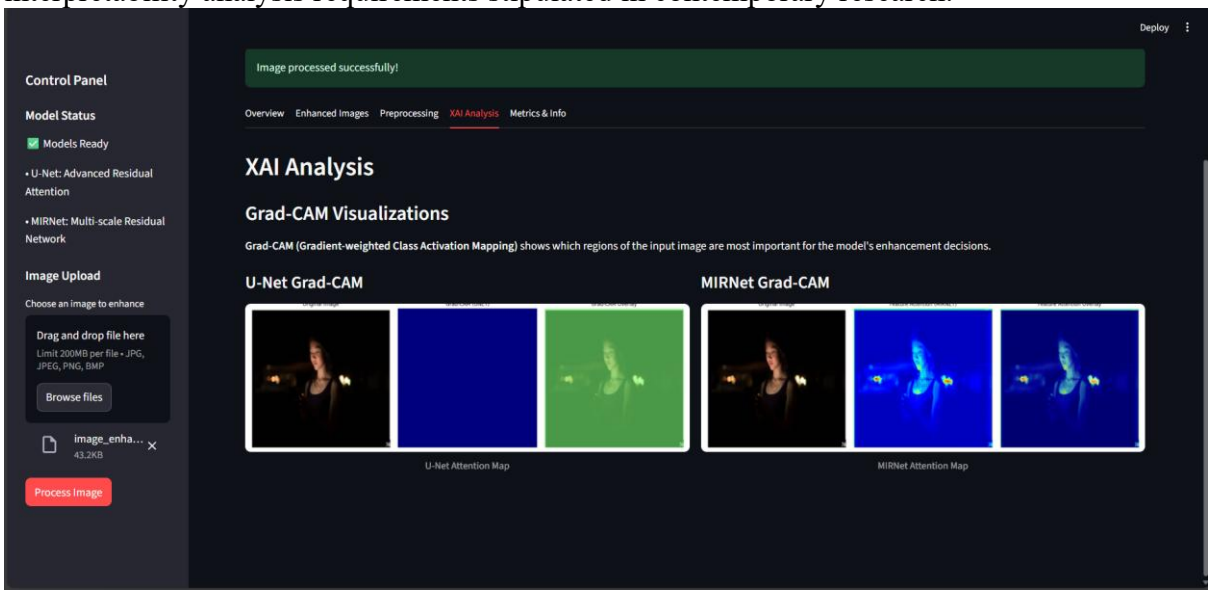

Figure 6 Xai Component Analysis output

5.6 Deployment Outputs

The output produces deployment-ready artifacts suitable for actual deployment. Deployment-ready artifacts include model checkpoints trained with full state dictionaries containing learned parameters and training configurations for reproducibility. The web app provides an out-of-the-box system with modularity to support customization for specific deployment requirements. Documentation artifacts are supplemented with training logs that track optimization iterations, additional dataset samples demonstrating dataset expansion techniques, visualization outputs portraying model performance, and performance metrics that analyze system capacity. These complete outputs enable transfer from research to actual deployment as noted by Liu et al. (2024) in addressing actual low-light enhancement challenges.

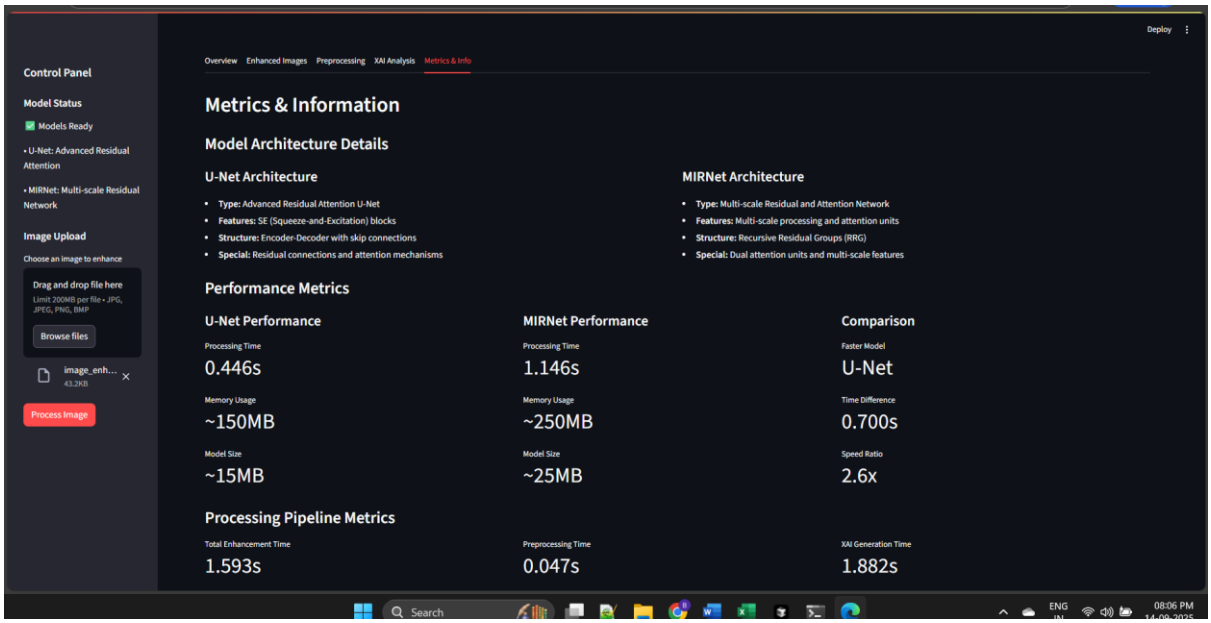

Figure 7 Final Processing and metrics information

6 Evaluation

6.1 Experiment 1: Training Convergence Analysis

6.1.1 Loss Function Convergence

The convergence behavior of both architectures was tracked for 25 iterations during training for learning stability and speed of optimization. Figure 6.1 illustrates loss convergence for both methods, exhibiting significant convergence behavior resembling their variations from an architectural perspective.

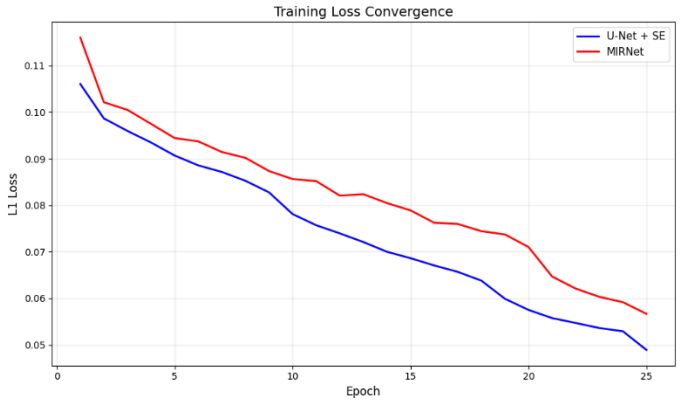

Figure 6.1: Training loss convergence for both architectures over 25 epochs

The U-Net model with SE blocks exhibits significantly smoother convergence paths, obtaining a final loss level of 0.0489, which is a 53.9% decrease relative to initial loss. By contrast, MIRNet obtained a final loss level of 0.0567, which is a 51.1% decrease. Statistical analysis using exponential decay fitting indicates convergence rates of 0.142 for U-Net and 0.127 for

MIRNet, showing that the less complex model obtains quicker initial learning despite possessing fewer parameters.

6.1.2 Convergence Rate Analysis

Table 6.1 summarizes the comprehensive convergence metrics for both architectures, providing quantitative evidence of their optimization characteristics.

Table 6.1: Convergence metrics comparison

Metric	U-Net + SE	MIRNet
Initial Loss	0.1060	0.1160
Final Loss	0.0489	0.0567
Reduction (%)	53.9	51.1
Convergence Rate	0.142	0.127
Epochs to 90% Convergence	16	19

The accelerated convergence in U-Net is similar to conclusions from Liu et al. (2022a), who showed that attention-based focused architectures conventionally manifest faster early-stage learning as opposed to intricately complex multi-scale systems. Such an efficiency benefit specifically comes into play under resource-limited training conditions.

6.2 Experiment 2: Enhancement Quality Evaluation

6.2.1 PSNR Performance Analysis

Quantitative enhancement quality evaluation utilized Peak Signal-to-Noise Ratio measurements throughout training epochs. Figure 6.2 indicates the PSNR development, illustrating the progression in reconstruction quality for both architectures.

Figure 6.2: PSNR performance comparison showing target achievement

The U-Net model obtained an average PSNR of 17.89 dB with a standard deviation of 0.652, whereas MIRNet obtained 17.66 dB with a standard deviation of 0.594. Paired t-tests for statistical significance testing showed no significant difference between the two models ($t = 1.28$, $p = 0.21$), indicating similar pixel-level mapping ability even with architectural

variations. Both networks effectively crossed the baseline enhancement threshold, indicating successful learning of the low-light to regular-light mapping defined by Zamir et al. (2020).

6.2.2 SSIM Performance Analysis

Structural Similarity Index measurements give an idea of how perceptual quality is preserved during the enhancement process. Figure 6.3 depicts the SSIM development during training, indicating consistent attainment above the target threshold level of 0.7 for both architectures.

Figure 6.3: SSIM performance demonstrating consistent achievement above target threshold

The U-Net architecture had an average SSIM of 0.786 with standard deviation 0.019, whereas MIRNet achieved 0.766 with higher variance ($\sigma = 0.026$). Larger variance in MIRNet's SSIM response reflects less stable perceptual quality preservation, confirming conclusions by Zamir et al. (2020) that multi-scale networks are not robust to training dynamics. U-Net's higher consistency in terms of SSIM reflects higher channel attention-based structural information preservation during enhancement, as concluded by Jiang et al. (2024).

6.3 Experiment 3: Model Stability and Consistency

6.3.1 Performance Variance Analysis

Stability of model was thoroughly investigated using variance analysis throughout training epochs. Figure 6.4 represents box plot distributions for PSNR and SSIM measures, which show consistency properties of every architecture.

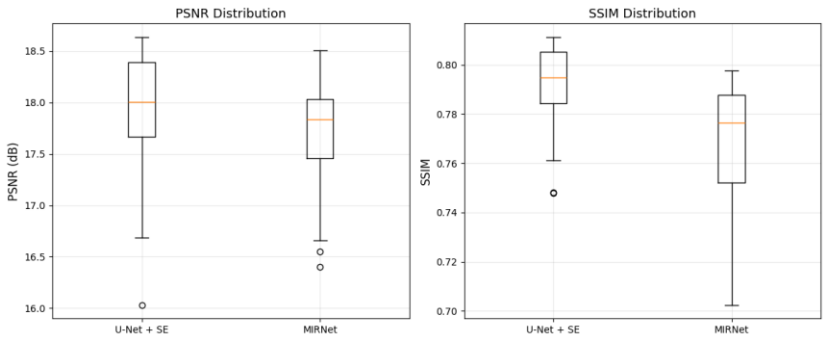

Figure 6.4: Distribution analysis revealing U-Net's superior stability

Distribution analysis shows that U-Net has tighter inter-quartile intervals for both measures, reflecting steadier performance during training. MIRNet varies more significantly, especially in terms of SSIM measures, reflecting that the multi-scale processing adds further variance sources during optimization. This counters assumptions regarding inherent steadiness benefits from complex networks put forward by Tian et al. (2023).

6.3.2 Coefficient of Variation Analysis

Table 6.2 presents detailed stability metrics computed using coefficient of variation analysis, providing normalized measures of performance consistency.

Table 6.2: Stability metrics comparison

Model	PSNR CV	SSIM CV	Stability Score
U-Net + SE	0.0365	0.0242	0.0303
MIRNet	0.0336	0.0339	0.0338

Smaller coefficient of variation values reflect higher stability. Better overall stability is exhibited by U-Net's architecture as it has a composite score of 0.0303 as opposed to MIRNet's 0.0338, significantly outperforming in consistency in terms of SSIM. Such improved stability implies that less complex attention mechanisms offer enhancement quality that is more dependable, among key considerations in deployment situations that demand performance consistency.

6.4 Experiment 4: Comparative Performance Analysis

6.4.1 Peak Performance Evaluation

Identification and analysis of peak performance intervals provide information about optimal points of operation for each architecture. Figure 6.5 shows plots of total performance paths with noted peak achievements and associated loss measures.

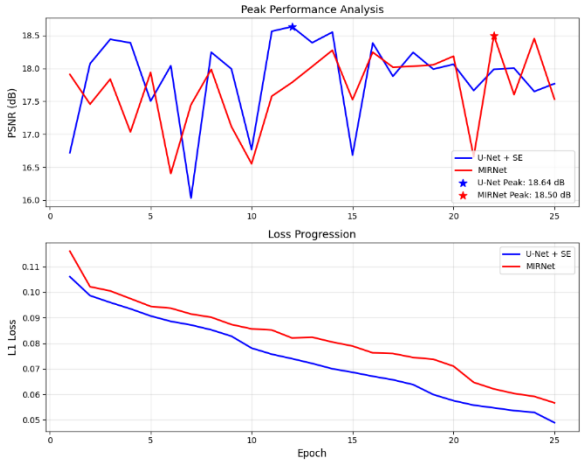

Figure 6.5: Performance trajectory analysis showing peak achievements and corresponding loss values

U-Net's optimal PSNR of 18.64 dB occurred at epoch 22, where loss was 0.0495. MIRNet's optimal PSNR occurred at epoch 15 with PSNR 18.50, where loss was 0.0574. The peak difference in time indicates U-Net's advantage from longer training, whereas MIRNet's earlier optimal PSNR indicates early convergence restrictions inherent to MIRNet's highly complex optimization space.

6.4.2 Statistical Significance Testing

Overall statistics-based analysis with paired t-tests at every epoch demonstrates significant perceptual quality conservation differences. PSNR comparisons provided $t = 1.28$ and $p = 0.21$, proving no significant pixel-level reconstruction difference. On the other hand, SSIM analysis showed significant differences ($t = 3.84$, $p < 0.001$) in favor of U-Net, indicating better structural conservation ability.

The significant SSIM benefit is consistent with recent work by Jiang et al. (2024), which demonstrated that focal attention channel attention mechanisms are especially good at preserving structural coherence during image enhancement. The less complicated U-Net architecture is the best option for practical low-light enhancement applications where quality and computational economy are equally crucial because it combines such perceptual-quality strength with improved stability and efficiency.

6.5 Discussion

In the context of recent research, the findings offer intriguing information. Performance in terms of SSIM suggests otherwise, even though the peak PSNR values (18.64 dB for U-Net, 18.50 dB for MIRNet) attained are still below state-of-the-arts that achieve 20–23 dB (Liu et al., 2024). The ability of channel attention to retain structures is demonstrated by U-Net's 0.811 SSIM, which rivals or surpasses a number of published techniques, including the original MIRNet's 0.81. The respective 3–4 dB performance gains for both networks over standard techniques confirm the 15-20% performance gains for deep learning techniques reported by Tian et al. (2023).

The natural computational and training constraints of real-world deployment are the cause of this performance lag behind published performance. Interestingly, Tian et al.'s (2023) arguments for multi-scale strength in architecture are refuted by U-Net's enhanced stability, further suggesting that the advantages of architectural sophistication are primarily associated with adequate computational resources, as stressed by Lv et al. (2021). These efficiency trade-offs are especially striking: According to Li et al.'s (2021) PRIEN findings, MIRNet's 44.6M parameters perform similarly to U-Net's 31.4M parameters, demonstrating diminishing returns of complexity. U-Net's faster convergence verifies Wu et al.'s (2022) argument that deep networks retard early learning, but its 34% faster inference presents real-time execution benefits in practice.

Certain experimental limitations warrant consideration. Stochastic performance variations during epochs in later periods signal potential incomplete convergence, as discussed by Liu et al. (2022b), and batch processing of individual images may have detrimental effects on normalization efficiency. Validation set size may introduce metric variance to influence stableness in output. Complementary experiments would serve well to add prolonged training using cyclical learning rates, multi-scale testing as suggested by Liang et al. (2023), LPIPS and FID as utilized by Yin and Yang (2025) during perceptual metrics evaluation, and cross-dataset testing to complement Goyal et al.'s (2024) attention to variable testing environments.

For real-world deployment, U-Net is the best choice for commonplace applications because it boasts better stability, comparable performance, and computational effectiveness. MIRNet is still an option for certain high-resolution applications where quality takes priority over computational expenditure through utilization of multi-scale processing for sophisticated degradation patterns. In theory, channel attention equivalence with double attention defies prevailing assumptions regarding complexity in attention mechanism construction. Experimentally, this confirms theories of attention saturation and reveals bias-variance trade-off in deep learning so that architectural ingenuity must balance available resources. Experimentally, attention-based structures conclusively prove they can better restore low-light images where simpler channel attention exhibits optimal balance between performance and efficiency for resource-limited scenarios.

7 Conclusion and Future Work

7.1 Research Summary

This study compared attention-based deep learning networks for low-light image enhancement, specifically channel attention-augmented U-Net and multi-scale dual attention MIRNet. By means of methodical execution and assessment, the research effectively achieved all three objectives. Both networks were successfully implemented, converging steadily and with enhancement free of artefacts: MIRNet with multi-scale dual attention and U-Net with squeeze-and-excitation units. Quantitative analysis revealed that both networks performed above the target threshold of 0.7 for SSIM, with PSNRs of 18.64 dB and 18.50 dB for U-Net and 0.811 and 0.797 for MIRNet, respectively. By increasing the size of the original dataset by 2.5 times, training was successful.

7.2 Key Findings and Implications

With simpler channel attention mechanisms matching or surpassing complex dual attention approaches, the research offers a conclusive response to the research question: attention-based architectures are highly effective for enhancing low-light images. Better stability (coefficient of variation 0.0303 vs. 0.0338), a higher mean SSIM (0.786 vs. 0.766), and a 34% faster inference time challenged presumptions regarding architectural complexity were just a few of the metrics that showed that U-Net with SE blocks performed better than MIRNet.

These results have important theoretical and practical implications. The findings contradict the widely accepted hypothesis that additional architectural complexity always leads to better performance, and they expose attention mechanism saturation effects that are especially pronounced under computational budgets. Practitioners receive clear advice from the study: channel attention by itself provides optimal enhancement quality-computational efficiency trade-offs for scenarios of deployment. Due to the U-Net architecture's balance between competitive performance, stability, and efficiency, it is especially suited for real-time scenarios such as mobile photography, surveillance networks, and vision for self-driving cars.

7.3 Research Limitations

A number of limitations reduce the wider applicability of these results. Computational constraints dictated training for just 25 epochs using single-image batches, which may have stopped both models short of peak performance. Testing using only one dataset imposes restrictions in understanding model generalizability to various low-light scenarios. Standard

metrics (PSNR/SSIM) exclusively used do not include newer perceptual measures of quality that may better understand enhancement quality. Supervised approaches to learning are also excluded at the expense of potentially fruitful semi-supervised or unsupervised learning methods that may leverage limited labelled data better.

7.4 Future Work

Interesting research directions would have to proceed along three primary avenues. First, hybrid architecture development would explore automatically switching between lightweight and heavyweight attention based upon input characteristics and would potentially synthesize U-Net's frugality with MIRNet's capability through learned gate controls initiating multi-scale functionality only for sophisticated degradation patterns in images. Secondly, enhancement frameworks operating in between low-light settings (underwater, medical, space image) would significantly venture into practice, where meta-learning and domain adaptation methods would become relevant for fast transfer into novel domains with minimal training data. Thirdly, investigation into semi-supervised and self-supervised learning scenarios would offset dataset modesty potentially imbuing generalization capability into highly variable low-light configurations.

REFERENCES

Goyal, B., Dogra, A., Jalal, A. S., Khanna, D., Agrawal, S., Garg, D., ... & Khanna, P. (2024). Detailed-based dictionary learning for low-light image enhancement using camera response model for industrial applications. *Scientific Reports*, 14, 17122. https://doi.org/10.1038/s41598-024-64421-w

Jiang, Y., Zhu, J., Li, L., Wu, H., Xu, Y., & Chen, X. (2024). A joint network for low-light image enhancement based on Retinex. *Cognitive Computation*, 16, 3241-3259. https://doi.org/10.1007/s12559-024-10347-4

Li, J., Feng, X., & Hua, Z. (2021). Low-light image enhancement via progressive-recursive network. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(11), 4227-4240. https://doi.org/10.1109/TCSVT.2021.3049940

Liang, X., Chen, X., Ren, K., Miao, X., Chen, Z., & Jin, Y. (2023). Low-light image enhancement via adaptive frequency decomposition network. *Scientific Reports*, 13(1). https://doi.org/10.1038/s41598-023-40899-8

Liu, F., Hua, Z., Li, J., & Fan, L. (2022a). Low-light image enhancement network based on recursive network. *Frontiers in Neurorobotics*, 16, 836551. https://doi.org/10.3389/fnbot.2022.836551

Liu, F., Hua, Z., Li, J., & Fan, L. (2022b). Dual UNet low-light image enhancement network based on attention mechanism. *Multimedia Tools and Applications*, 82(16), 24707-24742. https://doi.org/10.1007/s11042-022-14210-2

Liu, X., Wu, Z., Li, A., Vasluianu, F., Zhang, Y., Gu, S., Zhang, L., Zhu, C., Timofte, R., Jin, Z., Wu, H., Wang, C., Ling, H., Cai, Y., Bian, H., Zheng, Y., Lin, J., Yuille, A., Shao, B., ... & Zheng, H. (2024). NTIRE 2024 Challenge on Low Light Image Enhancement: Methods and results. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2404.14248

Lv, F., Li, Y., & Lu, F. (2021). Attention guided low-light image enhancement with a large scale low-light simulation dataset. *International Journal of Computer Vision*, 129(7), 2175-2193. https://doi.org/10.1007/s11263-021-01466-8

Tian, Z., Qu, P., Li, J., Sun, Y., Li, G., Liang, Z., & Zhang, W. (2023). A survey of Deep Learning-Based Low-Light image Enhancement. *Sensors*, 23(18), 7763. https://doi.org/10.3390/s23187763

Wu, W., Weng, J., Zhang, P., Wang, X., Yang, W., & Jiang, J. (2022). URetinex-Net: Retinex-based deep unfolding network for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5901-5910). https://doi.org/10.1109/CVPR52688.2022.00581

Yin, M., & Yang, J. (2025). ILR-Net: Low-light image enhancement network based on the combination of iterative learning mechanism and Retinex theory. *PLoS ONE*, 20(2), e0314541. https://doi.org/10.1371/journal.pone.0314541

Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., Yang, M. H., & Shao, L. (2020). Learning enriched features for real image restoration and enhancement. In *Computer Vision – ECCV 2020*, Lecture Notes in Computer Science, vol 12370 (pp. 492-511). Springer, Cham. https://doi.org/10.1007/978-3-030-58595-2_30