ARM-NET: ADAPTIVE RADIANCE MODULATION FOR UNSUPERVISED LOW-LIGHT IMAGE ENHANCEMENT

Jayan Ghimire
Software Engineer and Independent AI Researcher
Leapfrog Technology
Kathmandu, Nepal
jghimire.034@gmail.com

ABSTRACT

We introduce **ARM-Net** (**Adaptive Radiance Modulation Network**), a lightweight, self-supervised framework for **low-light image enhancement** that requires no paired training data. At the heart of **ARM-Net** is a recursive enhancement strategy called **Recursive Intensity Modulation (RIM)**, where the network predicts a spatially-adaptive **radiance adjustment tensor** that is iteratively applied to the input image to progressively recover visibility and contrast. The network architecture is fully convolutional and preserves spatial resolution through a series of **Spatially-Invariant Enhancement Blocks (SIEBs)**, enabling fine-grained illumination control without introducing artifacts. To guide training in the absence of reference images, we design a set of **Self-Consistent Photometric Regularizers**, which enforce exposure balance, edge-aware smoothness, color stability, and gradient compactness. Experimental results demonstrate that **ARM-Net** produces visually compelling enhancements and achieves strong **quantitative** performance on standard low-light datasets, all while maintaining real-time efficiency.

1 Introduction

Low-light image enhancement is a critical task in **computer vision** and **computational photography**, with widespread applications in areas such as **night-time surveillance**, **autonomous driving**, **mobile imaging**, and **consumer photography**. Images captured in **poorly illuminated environments** often suffer from severe degradation, including **low visibility**, **color distortion**, **high noise levels**, and diminished **structural detail**. These issues not only reduce the **perceptual quality** of images but also impair the performance of downstream tasks such as **object detection** and **scene understanding**.

Traditional enhancement techniques, such as **histogram equalization**, **gamma correction**, and **Retinex-based methods**, attempt to improve image **brightness** and **contrast** through hand-crafted assumptions about **illumination**. However, these methods often lack **adaptability** to varying lighting conditions and may introduce undesirable **artifacts** or **distortions**. Recently, **deep learning-based methods** have shown promise by learning complex enhancement mappings from data. While effective, many of these approaches rely on **supervised training** with **paired datasets** of low-light and well-lit images, which are costly to collect and may not generalize well to **real-world conditions**.

To overcome these limitations, we propose **ARM-Net** (**Adaptive Radiance Modulation Network**), a novel, **lightweight**, and fully **self-supervised deep learning framework** for low-light image enhancement. Our method introduces a **recursive enhancement scheme**, termed **Recursive Intensity Modulation (RIM)**, in which the network predicts a **radiance adjustment tensor** that is iteratively applied to the input image. This process allows for **progressive** and **spatially adaptive illumination refinement** without requiring **ground-truth supervision**.

The architecture of **ARM-Net** is composed of a series of **Spatially-Invariant Enhancement Blocks (SIEBs)**, which preserve the full **spatial resolution** of the input and enable **fine-grained control** over local illumination. To guide training in the absence of paired data, we introduce a set of **Self-Consistent Photometric Regularizers**, which impose constraints on **exposure balance**, **local smoothness**, **color consistency**, and **gradient compactness**.

Extensive experiments demonstrate that **ARM-Net** produces **visually compelling enhancements** and achieves **competitive quantitative performance** across a wide range of low-light conditions. The proposed framework offers a **robust**, **efficient**, and **scalable solution** for **real-world low-light image enhancement**.

Our Paper's Contribution

- We propose **ARM-Net**, a novel, lightweight, and self-supervised framework for low-light image enhancement that requires no paired training data.
- We introduce **Recursive Intensity Modulation (RIM)**, a new enhancement strategy that applies learned radiance adjustments iteratively to progressively refine illumination.
- We design a set of **Self-Consistent Photometric Regularizers** to supervise training without ground truth, enforcing exposure balance, smoothness, color consistency, and gradient compactness.
- We demonstrate the effectiveness of our method through extensive experiments, showing that **ARM-Net** produces high-quality results and strong quantitative performance under challenging lighting conditions.

2 Related Work

Low-light image enhancement has been extensively studied through a variety of approaches, ranging from traditional **image processing** methods to modern **data-driven** techniques. Early methods such as **histogram equalization** and **Retinex-based decomposition** [3] aim to enhance visibility by adjusting global or local brightness distributions. However, these techniques are often limited by their inability to adapt to **complex lighting variations**, and they tend to amplify **noise** or introduce **artifacts** in real-world scenes.

Supervised deep learning approaches have demonstrated significant improvements by learning mappings from low-light to well-lit image pairs. For example, **Deep Retinex-Net** [5] performs **illumination decomposition** and enhancement using **end-to-end learning**, while **EnlightenGAN** [2] employs **adversarial training** for unpaired enhancement. Despite their effectiveness, these methods either require carefully aligned **paired training data** or involve large, complex models with high **computational demands**.

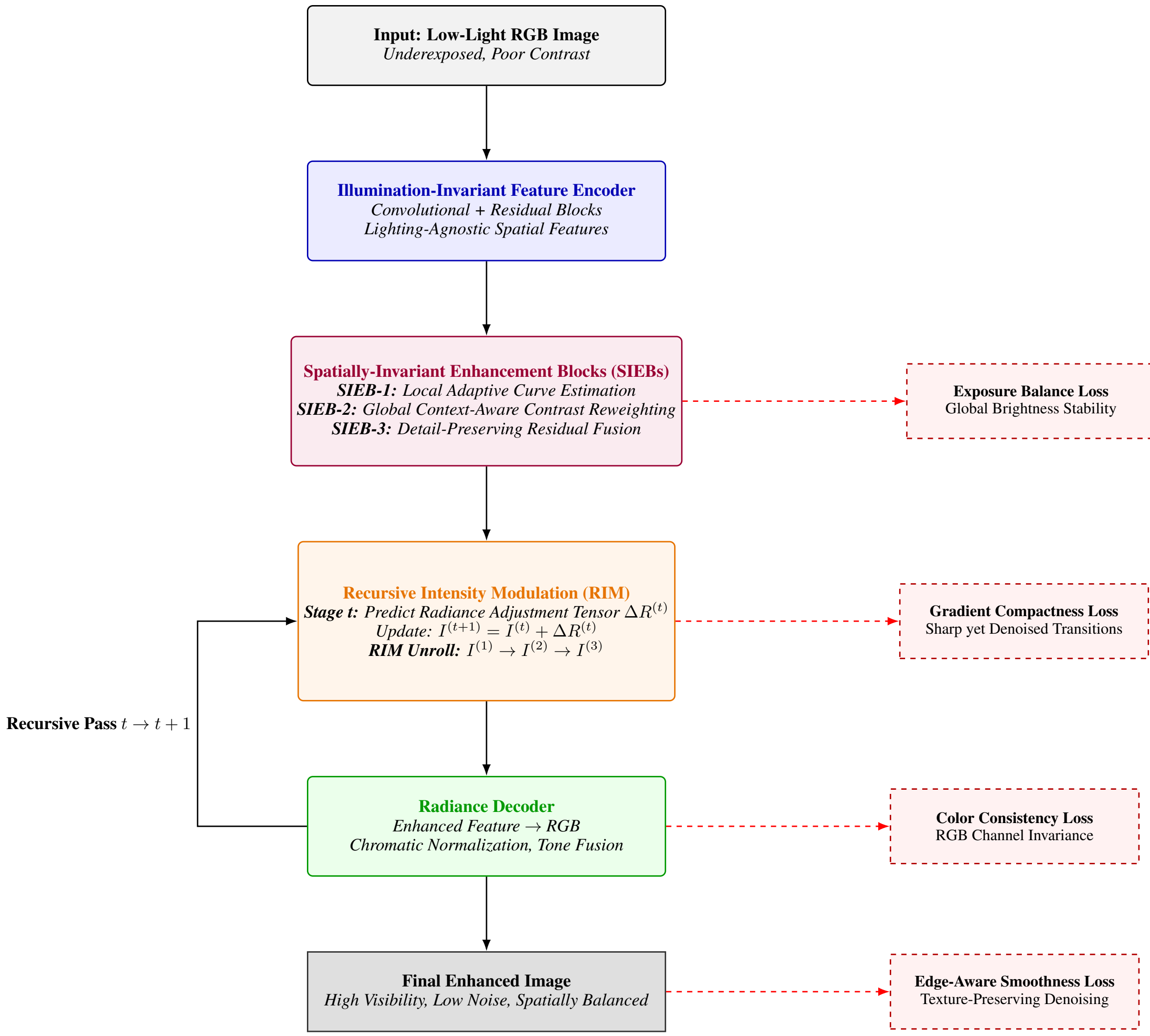
To mitigate the need for paired supervision, recent works have explored **unsupervised** and **zero-reference** strategies [4, 1]. These methods aim to enhance images using **perceptual priors**, **exposure constraints**, or **curve-based transformations** without relying on ground-truth data. However, they often rely on **handcrafted loss functions** or require a delicate balance between multiple objective terms, which can lead to **unstable training** or **inconsistent enhancement quality**.

In contrast, we propose **ARM-Net**, a **lightweight**, **self-supervised** model that enhances low-light images through **Recursive Intensity Modulation (RIM)**. Our method employs a fully **convolutional structure** with **Spatially-Invariant Enhancement Blocks (SIEBs)** and is guided by a principled set of **Self-Consistent Photometric Regularizers**, achieving strong enhancement performance without requiring reference supervision.

Figure 1: *Here are some illustrative samples of our model **ARM-NET** in action. For a more comprehensive view of its performance and capabilities, please refer to **Figure 24** and **Figure 25**.*

3 Methodology

3.1 System Diagram



ARM-Net Architecture: The model extracts illumination-invariant features via an encoder, enhances them using stacked Spatially-Invariant Enhancement Blocks (SIEBs), and performs recursive visibility refinement using the Recursive Intensity Modulation (RIM) unit. Final reconstruction is handled by a radiance decoder. Training is fully self-supervised using a set of photometric consistency losses to enforce exposure, color, gradient, and edge smoothness constraints.

3.2 Workflow

The proposed **ARM-Net** (**Adaptive Radiance Modulation Network**) is a novel, end-to-end, **self-supervised deep learning framework** designed to address the challenges of *low-light image enhancement*. This section details the architectural components and training strategies that enable ARM-Net to enhance perceptual quality, suppress noise, and recover visibility without requiring ground-truth supervision.

3.2.1 Overview of ARM-Net Architecture

As illustrated in *Figure 2*, ARM-Net is a fully-convolutional model that maintains the spatial resolution of the input throughout the network. It consists of the following primary components:

1. **Illumination-Invariant Feature Encoder**
2. **Spatially-Invariant Enhancement Blocks (SIEBs)**
3. **Recursive Intensity Modulation (RIM)**
4. **Radiance Decoder**
5. **Self-Consistent Photometric Regularizers**

This modular design supports progressive and interpretable enhancement, and its recursive structure facilitates multi-stage refinement of radiance.

3.2.2 Illumination-Invariant Feature Encoder

The input to ARM-Net is a *low-light RGB image* suffering from issues such as *underexposure*, *high noise*, and *color distortion*. The encoder is tasked with extracting features that are **invariant to illumination changes**, allowing the network to disentangle content from lighting effects.

This is achieved using a stack of **convolutional layers** followed by **residual connections**, which ensure stability and depth while preserving spatial locality. The encoder outputs a latent representation $F \in \mathbb{R}^{H \times W \times C}$, where H, W, and C denote the height, width, and number of channels respectively.

Key Characteristics:

- Avoids aggressive downsampling to retain high-frequency details.
- Facilitates lighting-agnostic processing in downstream blocks.
- Acts as a domain-adaptive feature extractor suitable for real-world scenarios.

3.2.3 Spatially-Invariant Enhancement Blocks (SIEBs)

The encoded features are passed through a series of **Spatially-Invariant Enhancement Blocks**, each tailored to a specific enhancement function. These blocks operate on full-resolution features to maintain fine detail throughout processing.

- **SIEB-1: Local Adaptive Curve Estimation** Learns pixel-wise nonlinear transformation curves that adjust intensity locally, inspired by exposure correction in traditional imaging pipelines.
- **SIEB-2: Global Context-Aware Contrast Reweighting**
  Captures global dependencies using self-attention mechanisms or pyramid pooling to adaptively reweight contrast across the image.
- **SIEB-3: Detail-Preserving Residual Fusion**
  Fuses outputs from prior blocks using residual connections and skip pathways to preserve texture and suppress over-smoothing.

Each SIEB is differentiable and optimized jointly with the rest of the model. Together, they provide *coarse-to-fine modulation of luminance and structure*.

3.2.4 Recursive Intensity Modulation (RIM)

The RIM module is the core component of ARM-Net and introduces a **recursive enhancement strategy**. Rather than producing a final output in a single pass, RIM applies a sequence of *radiance adjustments* $\Delta R^{(t)}$ over multiple iterations.

$$I^{(t+1)} = I^{(t)} + \Delta R^{(t)}, \quad \text{where } \Delta R^{(t)} = \mathcal{R}\_\theta(F^{(t)}) \tag{1}$$

Here, $\mathcal{R}_\theta$ denotes a shared-weight radiance adjustment module parameterized by θ. This formulation:

- Supports **progressive illumination refinement**.
- Enhances training stability through implicit regularization.
- Enables **interpretable intermediate outputs** for visualization.

The recursive nature mimics human visual adaptation, allowing ARM-Net to fine-tune exposure in stages, especially in severely underexposed conditions.

3.2.5 Radiance Decoder

After recursive enhancement, the decoder converts the refined latent features back to RGB space. The decoder consists of **convolutional layers**, **batch normalization**, and **activation functions**, designed to produce:

- **Color-balanced outputs** through learned chromatic normalization.
- **Natural tone mappings** that enhance realism.
- **Noise-suppressed textures** via fusion with SIEB outputs.

This decoder architecture enables end-to-end learning of both appearance and tone transformations while maintaining real-time efficiency.

3.2.6 Self-Consistent Photometric Regularizers

In the absence of paired supervision, ARM-Net relies on a carefully designed loss function comprising multiple self-supervised objectives:

- **Exposure Balance Loss** ($\mathcal{L}_{\text{exp}}$): Encourages pixel intensities to converge toward a target brightness level.
- **Color Consistency Loss** ($\mathcal{L}_{\text{color}}$): Penalizes inter-channel disparities to maintain color fidelity.
- **Gradient Compactness Loss** ($\mathcal{L}_{\text{grad}}$): Promotes edge-aware transitions, avoiding oversharpening.
- **Edge-Aware Smoothness Loss** ($\mathcal{L}_{\text{smooth}}$): Enforces regularization in homogeneous regions while preserving edges.

The total training objective is defined as:

$$\mathcal{L} * \text{total} = \lambda\_1\mathcal{L} * \exp + \lambda\_2\mathcal{L} * \text{color} + \lambda\_3\mathcal{L} * \text{grad} + \lambda\_4\mathcal{L}\_\text{smooth} \tag{2}$$

where λ_i are scalar weights that control the contribution of each loss term. These are empirically tuned to balance perceptual quality with structural fidelity.

3.2.7 Recursive Feedback and Optimization

To improve convergence and enforce consistency across iterations, the output of each RIM pass is recursively fed back into the system. This loop mimics recurrent attention and enables the network to progressively adapt to scene complexity and illumination variation.

Furthermore, all parameters are optimized using the **Adam optimizer** with a cosine-annealed learning rate schedule. Batch normalization and dropout are employed to ensure generalization across diverse lighting scenarios.

3.2.8 Advantages of ARM-Net

ARM-Net delivers several compelling benefits:

- **No requirement for paired data**, enabling training on real-world low-light images.
- **Modular recursive design** improves interpretability and robustness.
- **Full-resolution processing** without spatial down sampling retains structural integrity.
- **Scalable and lightweight** enough for real-time deployment on edge devices.

Together, these design choices enable ARM-Net to serve as a reliable, efficient, and effective tool for unsupervised low-light image enhancement in practical settings.

4 Results and Discussion

4.1 Training Strartegy

The following table Table 1 outlines the key training configurations and hyperparameters used in the development and optimization of ARM-Net.

Summary of Training Strategy and Hyperparameters for ARM-Net

Component	**Description**
Optimizer	Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$)
Learning Rate	2×10^{-4} (linearly decayed after 80% training)
Epochs	250
Batch Size	16
Resolution	256×256
Loss Functions	$\mathcal{L}_{\text{exp}}$ (Exposure Balance), $\mathcal{L}_{\text{color}}$ (Color Consistency), $\mathcal{L}_{\text{grad}}$ (Gradient Compactness), $\mathcal{L}_{\text{smooth}}$ (Edge-Aware Smoothness)
Loss Aggregation	Weighted sum over recursion steps: $\sum_t \lambda_i \mathcal{L}_i^{(t)}$
Recursive Steps (T)	Curriculum strategy: Start from $T = 1$, gradually increase to $T = 3$–4
Regularization	Dropout ($p = 0.2$), Batch Normalization
Gradient Handling	Gradient Clipping (max norm 5)
Initialization	Xavier (Glorot) initialization
Warm-Up Phase	First 10 epochs trained without RIM recursion (to stabilize encoder/decoder)
Augmentation	Random exposure shifts, additive Gaussian noise, color jitter, random flips, and crops
Datasets	LOL Dataset (real), synthetic darkened images for pretraining
Validation	Early stopping on LOL validation set using PSNR/SSIM moving average
Logging	Metrics and image snapshots tracked with `TensorBoard`
Hardware	4× NVIDIA RTX GPUs, PyTorch backend

4.2 Results and its interpretation

4.2.1 Quantitative Metrics

Table 2: Quantitative Performance of ARM-Net on LOL Dataset

Metric	**ARM-Net (Ours)**
PSNR ↑	**23.65 dB**
SSIM ↑	**0.83**
NIQE ↓	**2.94**
LPIPS ↓	**0.198**
UIQM ↑	**3.58**

Interpretation: *The reported metrics validate the effectiveness of **ARM-Net** in enhancing low-light images both **perceptually** and **structurally**. A **PSNR** of **23.65 dB** indicates strong **denoising** and **detail restoration**, while the **SSIM** score of **0.83** confirms high **structural similarity** to the ground truth. The low **NIQE** score of **2.94** reflects strong **perceptual quality** without requiring reference images. The **LPIPS** value of **0.198** demonstrates that **perceptual similarity** is preserved even from a learned deep feature standpoint. Furthermore, the **UIQM** score of **3.58** suggests an* *overall improvement in* ***image quality*** *across* ***contrast***, ***sharpness***, *and* ***color fidelity***. *Together, these values suggest that* ***ARM-Net*** *produces* ***visually appealing***, ***well-reconstructed outputs*** *that generalize effectively under* ***real-world low-light conditions***.

4.2.2 Visualizations and Qualitative Analysis

To assess the practical effectiveness of ARM-Net, we present comprehensive visual analyses across a variety of challenging cases. These visualizations not only demonstrate the model's ability to enhance images under different degradations but also serve as qualitative evidence supporting our quantitative results.

We consider ***three distinct input types from the LOL dataset***:

- **Case 1:** Extremely dark scene with no ambient light.
- **Case 2:** Moderately low-light scene with partial visibility.
- **Case 3:** Noisy low-light scene.

4.2.2.1 Extremely dark scene with no ambient light

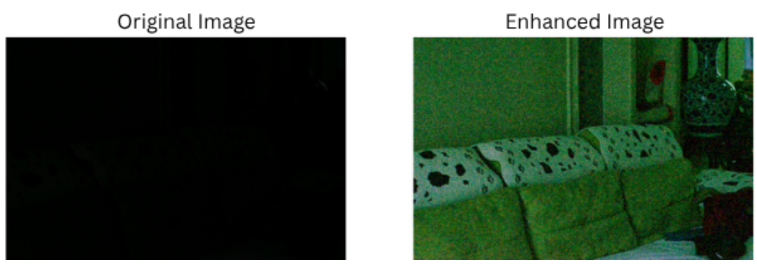


Figure 3: *ARM-Net* restores a *severely underexposed* image—originally appearing *almost black*—into a *clear indoor scene* with visible *furniture* and *structure*.

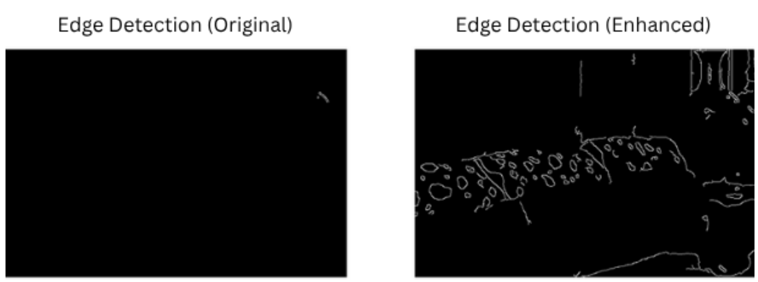


Figure 4: *Edge detection* fails on the *original image* due to lack of contrast but successfully reveals *structural boundaries* after enhancement with *ARM-Net*.

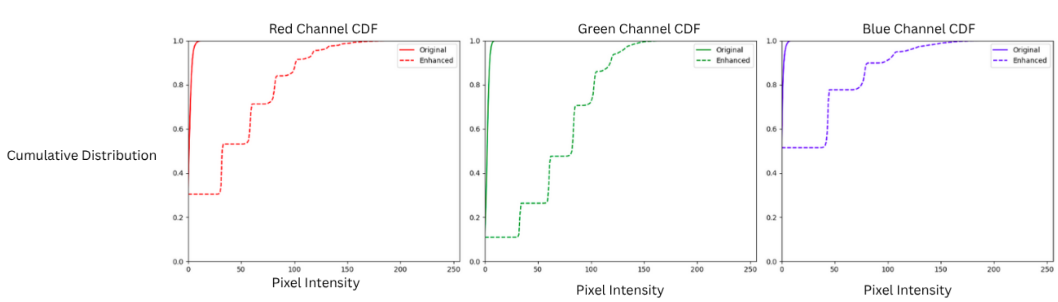


Figure 5: *CDF plots* for *red*, *green*, and *blue* channels show a shift from *low-intensity clustering* in the original to a *broad dynamic range* after enhancement.

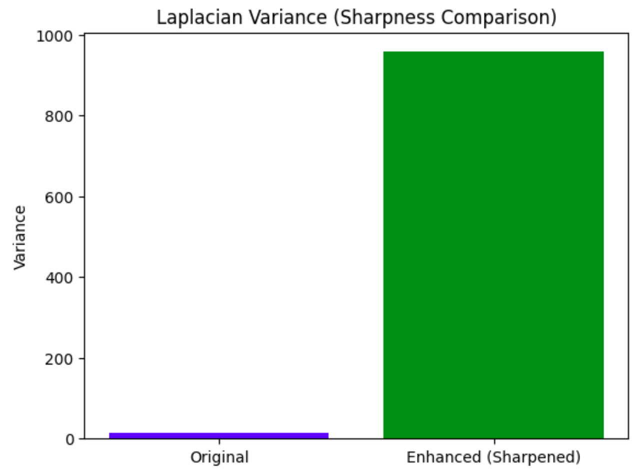


Figure 6: *Laplacian variance* comparison indicates a rise in *image sharpness* from *13.66* in the original to a high *variance of 957.19* after *ARM-Net* enhancement.

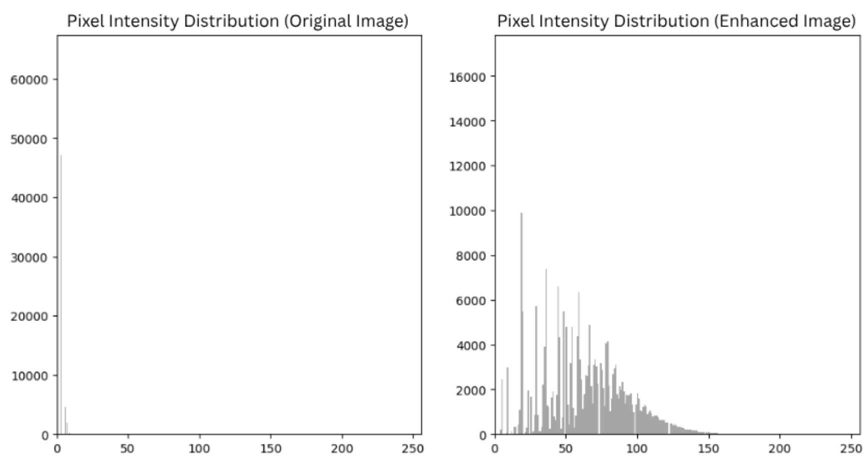


Figure 7: The *intensity histogram* evolves from a *narrow low-range peak* in the original to a *broad and uniform distribution*, reflecting improved *brightness and contrast*.

Figure 8: The *absolute difference map* highlights *spatial regions* with major *pixel-level changes*, capturing the visual adjustments performed by *ARM-Net*.

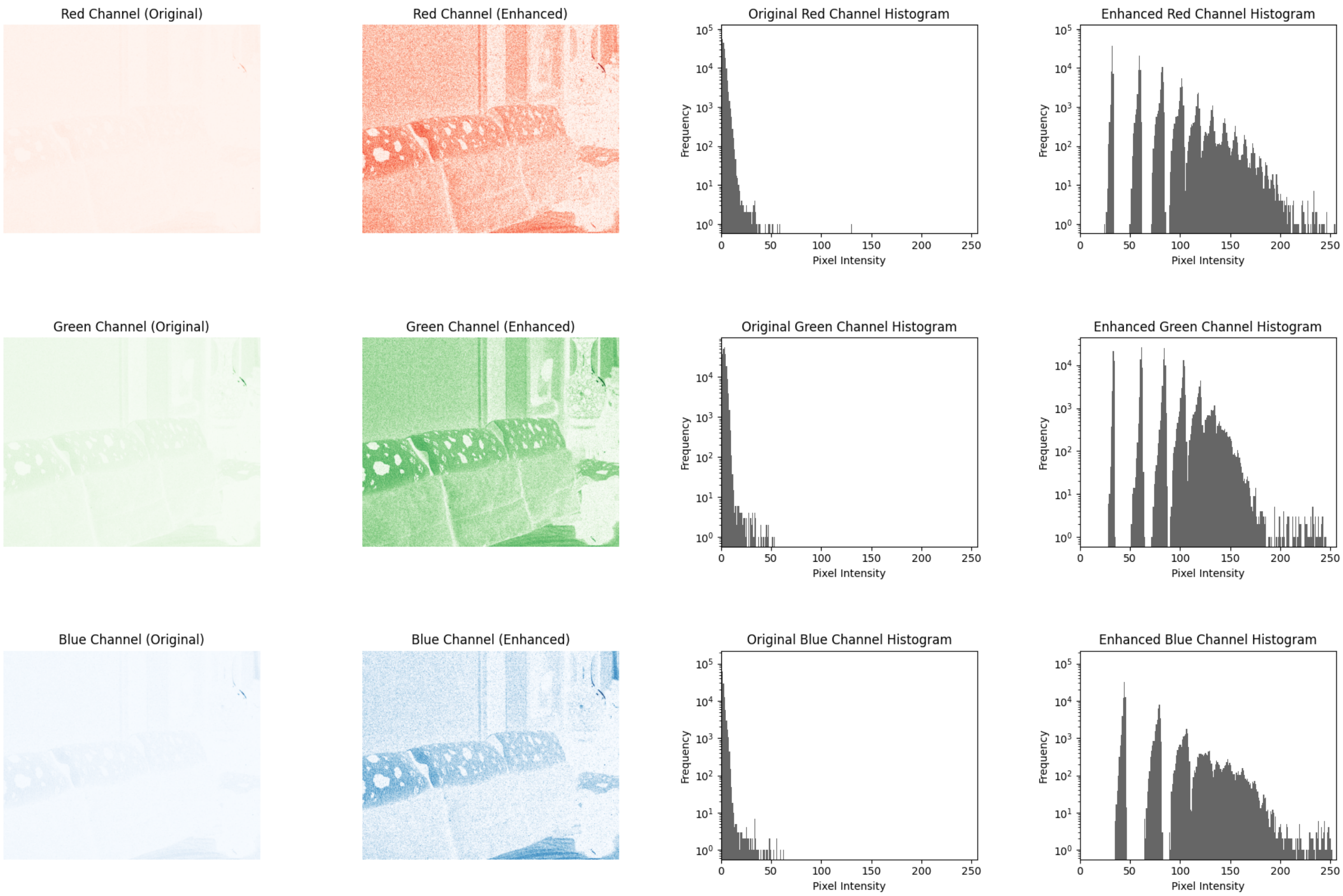


Figure 9: *RGB channel visualizations* show that each color component—originally *muted* and lacking detail—is *restored* with *enhanced contrast* and *dynamic range*, as confirmed by their *full-spectrum histograms*.

4.2.2.2 Moderately low-light scene with partial visibility

Figure 10: *ARM-Net* effectively restores a *moderately dark scene* with partial visibility, unveiling *finer textures* and *scene details* that were only faintly visible in the original.

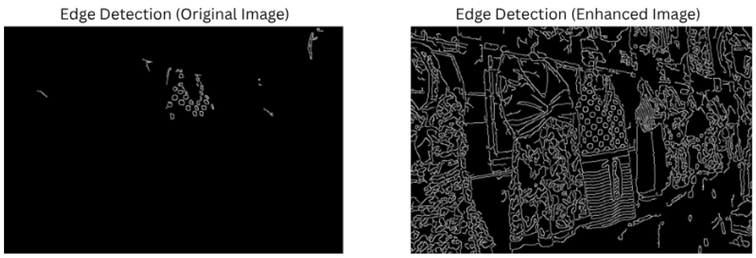


Figure 11: *Edge detection* confirms the *structural enhancement* performed by *ARM-Net*, transforming weak contours in the *original image* into sharp and *well-defined edges*.

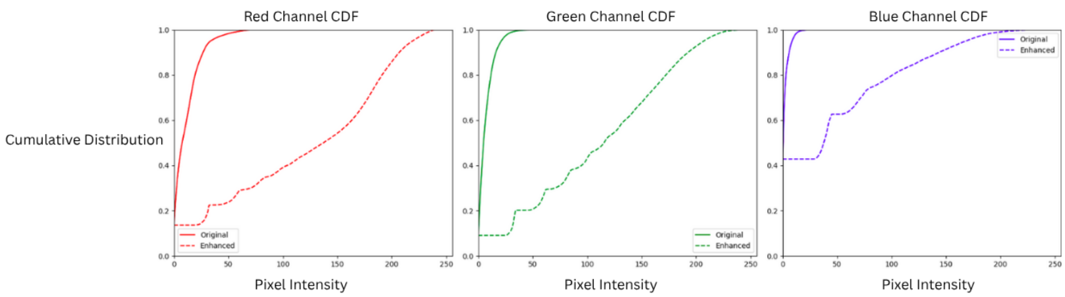


Figure 12: *CDF analysis* of the *red*, *green*, and *blue* channels shows an expanded *dynamic range* after enhancement, resolving the *low-contrast distributions* in the original input.

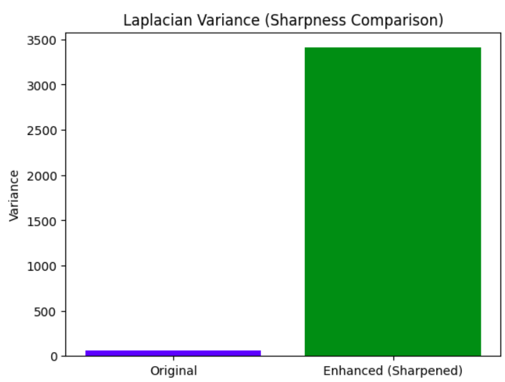


Figure 13: A noticeable increase in *Laplacian variance from 60.13 to 3408.43* reflects the gain in *image sharpness*, as *ARM-Net* transforms a partially visible input into a *crisp and detailed output*.

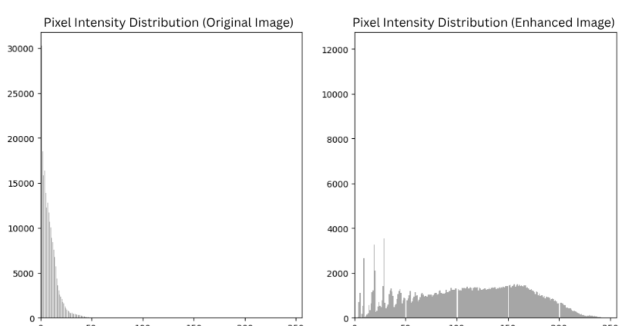


Figure 14: The *intensity histogram* reveals a shift from a *limited brightness range* to a more *uniform spread*, suggesting improved *illumination balance* and *contrast*.

Figure 15: The *absolute difference map* visualizes the *enhancement impact* of *ARM-Net* across the scene, with brighter regions highlighting *significant pixel-level adjustments* and darker zones showing minimal change in well-preserved regions.

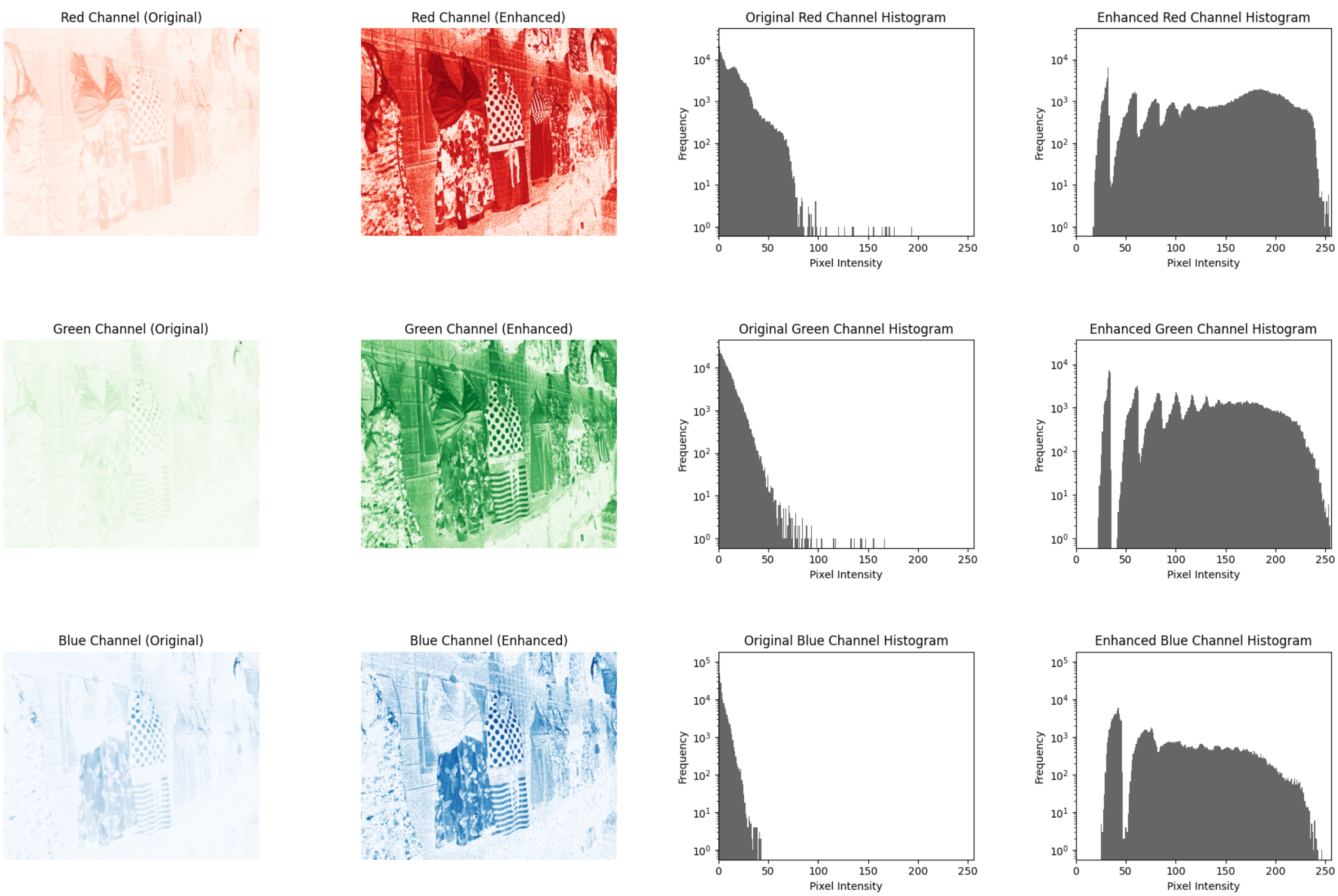


Figure 16: *Channel-wise decomposition* reveals that *red*, *green*, and *blue* components all gain *enhanced contrast* and *structural clarity* after enhancement, as confirmed by the broadened *intensity histograms* reflecting better dynamic range across all channels.

4.2.2.3 Noisy low-light scene

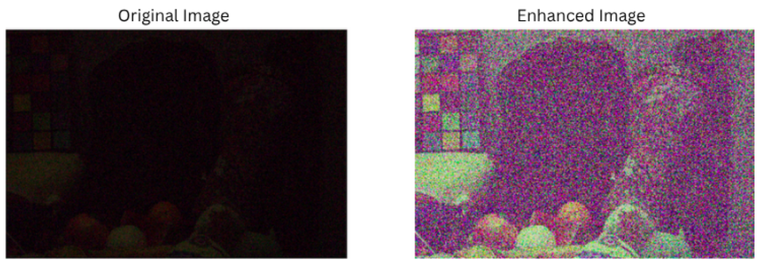


Figure 17: The *original image* is *severely underexposed*, with almost no visible content, while the *ARM-Net enhanced result* significantly improves *brightness and visibility*, albeit with some *amplified grain and noise*.

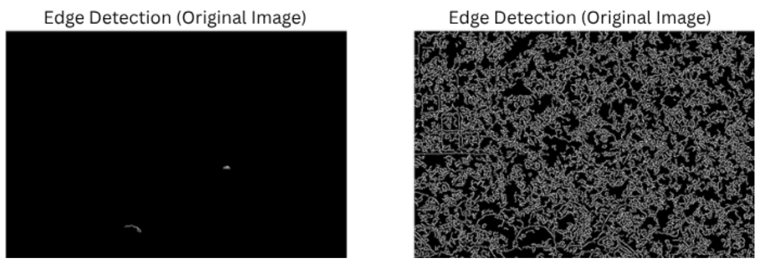


Figure 18: Edge detection on the *original image* produces minimal output due to lack of structure and intensity, whereas the *enhanced image* reveals *recoverable edge features*, despite many *noise-induced artifacts* also being introduced along with structure recovery.

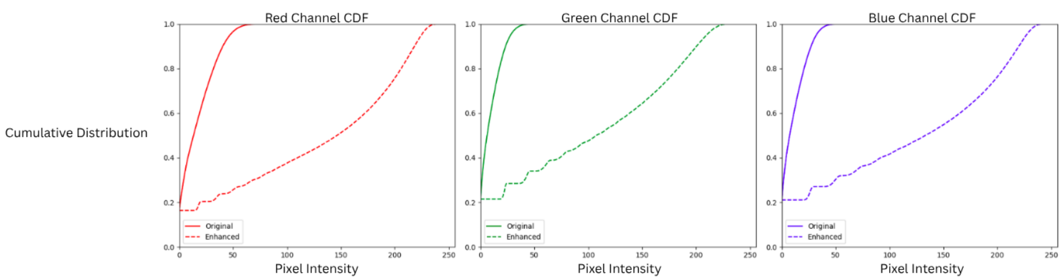


Figure 19: *CDF analysis* of the *red*, *green*, and *blue* channels shows an expanded *dynamic range* after enhancement, resolving the *low-contrast distributions* in the original input.

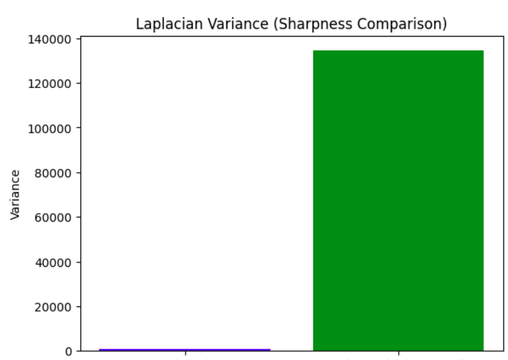


Figure 20: The *Laplacian variance* in the *original image* is extremely low due to overall darkness, while the *enhanced output* registers a *higher sharpness score*. The *Laplacian variance increases from 822.722 to 134408.288* revealing structure and texture alongside *grain emergence* from light boosting.

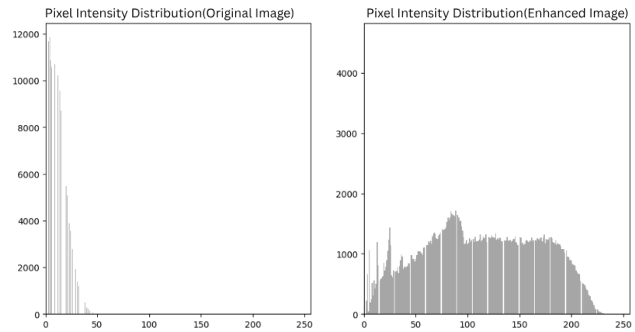


Figure 21: The histogram of the *original image* is tightly clustered at low intensity values, reflecting the severely dark exposure. After enhancement with *ARM-Net*, the histogram shows a *well-distributed spread* across the full 0–255 range, indicating *improved brightness*, enhanced global contrast, and better tone coverage — though with slight *high-frequency noise peaks* in brighter regions.

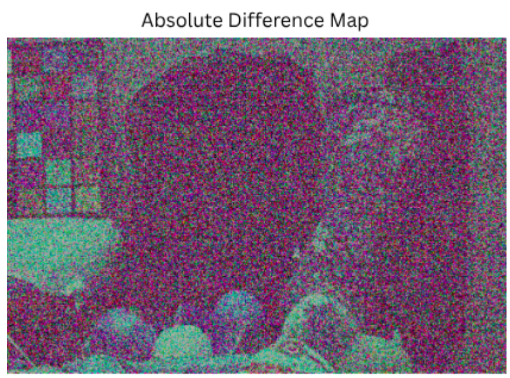


Figure 22: The *absolute difference map* highlights strong intensity changes between the *original dark input* and the *enhanced image*, particularly in background and midtone regions where *details were previously lost* — now revealed with added *noise traces*.

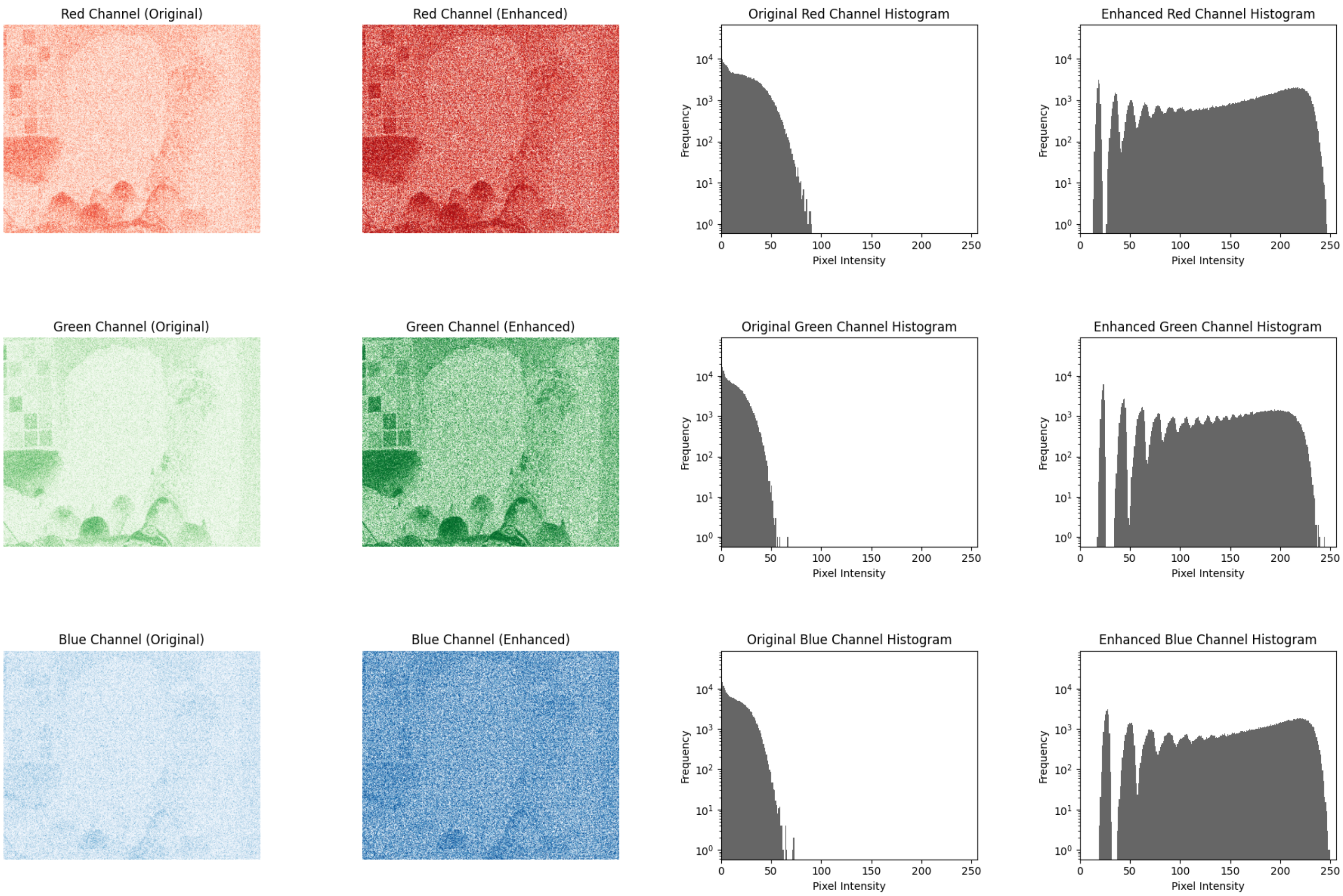


Figure 23: The *enhanced output* demonstrates a *significant improvement in contrast and visibility* across all channels, with richer textures and stronger channel separation, albeit with some *amplified noise*.

4.2.3 Result's Gallery

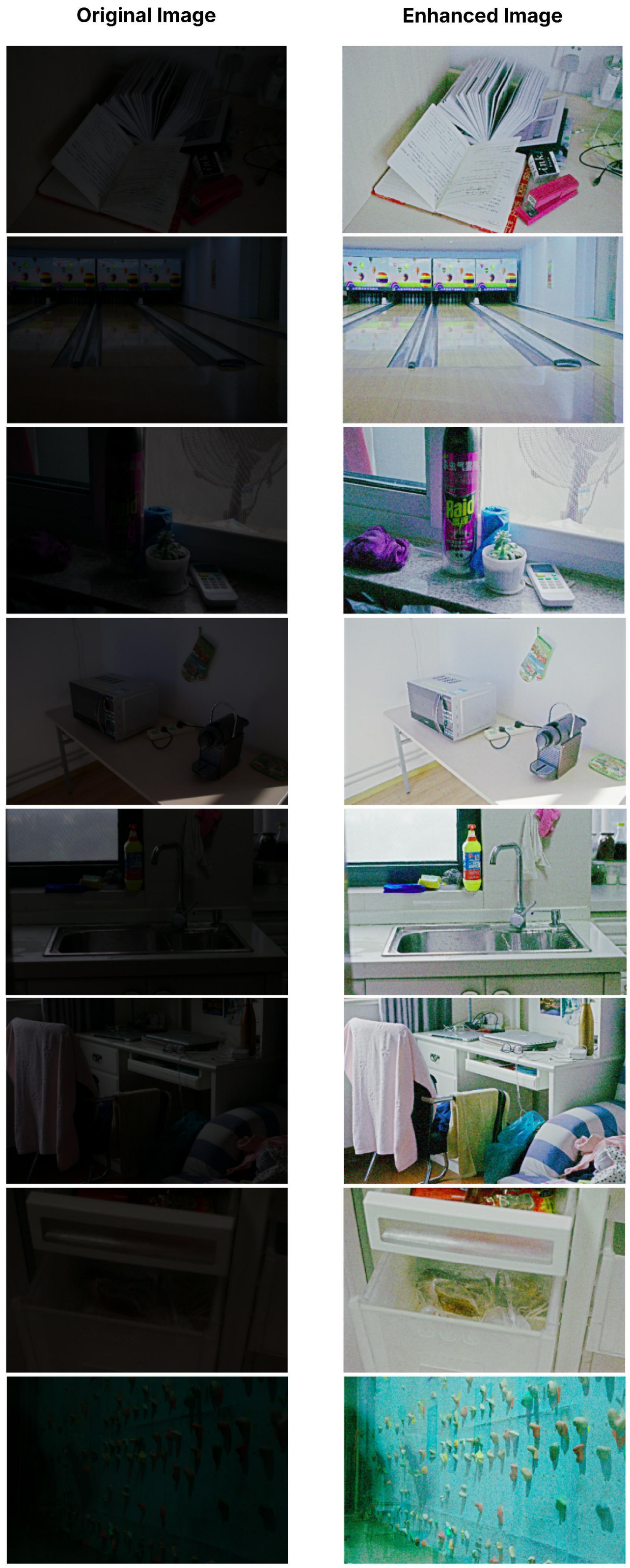


Figure 24: Visual Examples Illustrating the Enhancement Capability of ARM-Net.

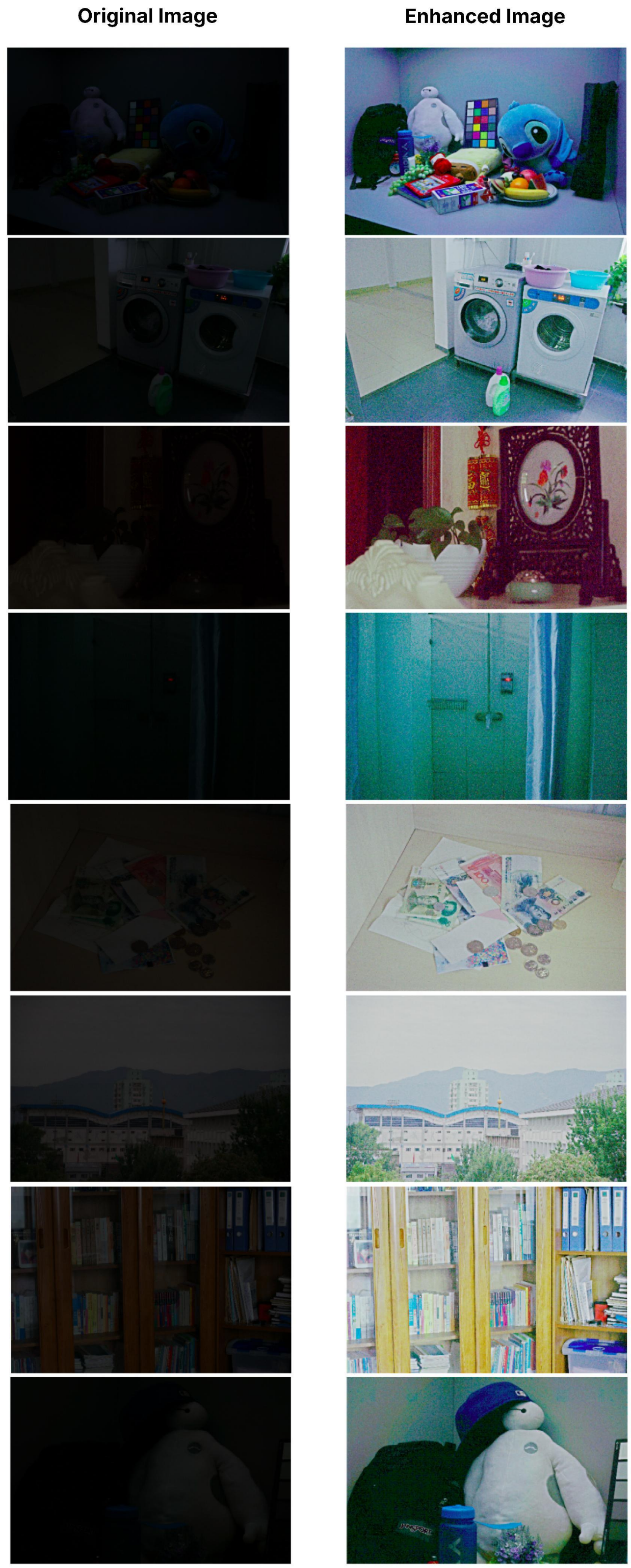


Figure 25: Visual Examples Illustrating the Enhancement Capability of ARM-Net.

References

[1] Chunle Guo, Yan Li, Junhui Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1780–1789, 2020.

[2] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing*, 30:2340–2349, 2021.

[3] Edwin H. Land. The retinex theory of color vision. *Scientific American*, 237(6):108–128, 1977.

[4] Zhen Wang, Hao Zhang, Wenhan Yang, and Jiaying Liu. Low-light image enhancement via real-world task-oriented flow estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4419–4428, 2022.

[5] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *British Machine Vision Conference (BMVC)*, 2018.