Indonesian Journal of Science & Technology

Journal homepage: http://ejournal.upi.edu/index.php/ijost/

MRD-UNet: Detail-Preserving Low-Light Image Enhancement with Multiscale Residual Dense Networks and Edge-Texture Guided Loss

I Gede Susrama Mas Diyasa[1,], Kraugusteeliana Kraugusteeliana[2], Riko Okananta[1], Anita Muliawati[2], Hamonangan Kinantan Prabu[2], Sayyidah Humairah[3], Ni Made Ika Marini Mandenni[4], Prisma Aji Riyantoko[5], Deshinta Arrova Dewi[6]*

[1]University of Pembangunan Nasional Veteran Jawa Timur, Surabaya, Indonesia
[2]University of Pembangunan Nasional Veteran Jakarta, Jakarta, Indonesia
[3]University of Patras, Rio Campus Patras, Greece
[4]Udayana University, Denpasar, Bali, Indonesia
[5]Okayama University, Okayama, Japan
[6]INTI International University, Nilai, Malaysia
*Correspondence: E-mail: igsusrama.if@upnjatim.ac.id

ABSTRACT

Low-light image enhancement is an important task in computer vision, particularly for night photography, autonomous driving, and video surveillance. Although recent deep learning methods perform well in reducing global noise, many models still degrade micro-structural details and texture fidelity because of excessive smoothing under extreme low-light conditions. This study proposes MRD-UNet, an encoder–decoder model designed to enhance low-light images while preserving structural integrity and textural sharpness. The model introduces two main components. First, the MRD-Block combines residual connections, dense connections, and multiscale convolutions to capture both global contextual features and fine local details. Second, an Edge-Texture Guided Loss based on Sobel and Laplacian operators is used to guide the learning process, improving edge consistency and high-frequency detail preservation. Experiments on the LOL-V1 benchmark show that MRD-UNet outperforms several state-of-the-art methods, achieving an SSIM of 0.911 and PSNR of 26.09 dB.

ARTICLE INFO

Article History:
Submitted/Received 09 Oct 2025
First Revised 11 Mar 2026
Accepted 04 Apr 2026
First Available Online 28 Apr 2026
Publication Date 30 Apr 2026

Keyword:
Contextual attention module,
Edge-texture guided loss,
Interlevel residual learning,
Low-light image enhancement,
Multiscale residual dense networks.

1. INTRODUCTION

Image acquisition under low-light conditions frequently introduces various artifacts, such as low contrast, loss of structural details, and the emergence of noise, which significantly degrade the quality of the resulting digital images [1]. In practice, manual interventions such as employing external flash units, adjusting ISO sensitivity, or utilizing other built-in camera settings often fail to provide a comprehensive solution. For instance, while increasing the ISO may enhance light sensitivity, it simultaneously amplifies image noise, thereby leading to a low signal-to-noise ratio (SNR), which results in diminished visual quality [2]. Furthermore, the combination of low contrast and extreme underexposure makes structural details difficult to identify or even entirely unrecoverable. In applications that prioritize image visibility over aesthetic quality, such as autonomous driving and video surveillance, structural sharpness is considerably more critical than mere noise reduction. Consequently, this research focuses on addressing the preservation and restoration of structural details in digital low-light images.

Low-light Image Enhancement (LLIE) is an established domain in computer vision, characterized by a rapid evolution of diverse methodologies. Existing approaches can be broadly categorized into traditional techniques and more advanced deep learning frameworks. Traditional methods, such as Histogram Equalization (HE) [3, 4] and its variant, Contrast Limited Adaptive Histogram Equalization (CLAHE) [5], enhance global and local contrast by redistributing pixel intensities. However, these techniques frequently result in over-exposure due to their limited adaptability to lighting variations and often inadvertently amplify sensor noise. Such extreme noise levels not only degrade image quality but also compromise the underlying structural content. Furthermore, more robust approaches based on Retinex Theory have been developed [6,7], which decompose an image into illumination and reflectance components to better emulate human visual perception. Nonetheless, these methods often suffer from oversaturation, rendering the enhanced images unnatural, and face significant challenges in achieving precise image decomposition. Consequently, traditional methods focusing primarily on illumination and contrast components remain inadequate for comprehensive LLIE, as they often overlook the critical interdependency between noise suppression and structural detail preservation.

Driven by the rapid advancement of neural networks, deep learning approaches have become the preferred and more promising alternative due to their robustness in addressing broad and complex challenges. In 2018, Retinex-Net [8] was developed, integrating traditional Retinex Theory into a deep learning framework, but its outputs often suffer from visual artifacts, oversaturation, and blurring. Similar approaches, such as KinD [9] and KinD++ [2], encounter over-exposure issues that cause structural details to degrade under excessive illumination intensity. Tao *et al*. [10] proposed a distinct approach based on Deep-CNNs with residual connections, which effectively produces smooth images without over-exposure. Nonetheless, the excessive use of residual connections without specific auxiliary mechanisms leads to the degradation of local details, resulting in an overly smoothed appearance. Furthermore, although GAN-based architectures incorporating attention modules have been implemented [11], extreme noise remains prominently visible. In recent years, Transformer-based self-attention approaches, such as SNR-Net [12], Restormer [13], and LLFormer [14] have emerged, yet, their primary focus remains on noise reduction, often at a high computational cost due to the self-attention mechanism. Based on this review of deep learning methodologies, there is still a significant lack of LLIE methods that prioritize the maximum preservation of structural details, a gap that this research aims to address.

DOI: https://doi.org/10.17509/ijost.v11i1.90990
p- ISSN 2528-1410 e- ISSN 2527-8045

To address the aforementioned challenges regarding the significance of structural detail, we propose a novel Low-light Image Enhancement model designated as MRD-UNet. Built upon the U-Net Encoder-Decoder architecture [15], which has demonstrated strong versatility across image reconstruction tasks including super-resolution [16], MRD-UNet integrates four core components designed for synergistic performance.

(i) The first component is the Multiscale Residual Dense Networks (MRD-Nets), a convolution block design that incorporates a hierarchical multi-scale receptive field to progressively capture features from global contexts to local details. It further utilizes Residual Connections [17] to ensure gradient stability in deep, complex networks, and Dense Connections [18] to enrich feature representation and facilitate feature-scale mixing.

(ii) The second component, Edge-Texture Guided Loss, employs a weighted combination of Laplacian (texture) and Sobel (edge) operators as a specialized loss function to optimize image sharpness. In practice, this is augmented with L1-loss and Multiscale Structural Similarity Index Measure (MS-SSIM) loss [19] to enhance training stability.

(iii) The third component, Interlevel Residual Learning (IRL), introduces an auxiliary mechanism through shortcut connections between convolution block levels within the encoder and decoder paths. IRL serves as an effective mechanism for inherent noise suppression while simultaneously improving gradient flow stability.

(iv) The final component is the Contextual Attention Module (CAM), which consists of channel attention via global pooling and spatial attention derived from parallel multi-scale Separable Convolutions to enhance global understanding and produce more natural visual nuances.

By integrating these components, we present a new model, MRD-UNet, designed not only to enhance low-light image quality quantitatively (PSNR, SSIM) but also qualitatively by preserving structural details and producing more natural-looking images. For clarity, the remainder of this paper is organized as follows: Section 2 describes the proposed methods and architecture, Section 3 presents experimental results and analysis, and Section 4 provides conclusions.

2. METHODS

2.1. Dataset Preparation

In this study, we employed the LOL (Low-Light Dataset) Benchmark for the development of our model, which includes LOL-V1 real capture [8], LOL-V2 real capture and synthetic [20], in order to evaluate the generalization capability of the model across various low-light conditions. The LOL-V1 real capture dataset consists of 485 pairs of low-light and normal-light images for training and 15 pairs for evaluation, yielding a total of 500 paired images. The LOL-V2 real capture dataset contains 689 training pairs and 100 evaluation pairs, for a total of 789 paired images, with more diverse conditions compared to LOL-V1. Meanwhile, the LOL-V2 Synthetic dataset comprises 900 pairs for training and 100 pairs for evaluation. Each dataset was used to train the model independently, with the expectation that the model could focus more effectively on the specific characteristics of each dataset.

Furthermore, we applied a preprocessing pipeline to all images to ensure stability before feeding them into the model. The preprocessing steps included image resizing and pixel normalization to the range [0, 1], without applying any additional data augmentation. For the LOL-V1 and LOL-V2 real capture datasets, each image was resized to 256 × 384, aiming to improve computational efficiency while preserving the structural integrity of the images. In contrast, the LOL-V2 Synthetic dataset was resized to 256 × 256 with the same objective. For pixel normalization, all datasets were processed consistently by scaling the intensity values to

DOI: https://doi.org/10.17509/ijost.v11i1.90990
p- ISSN 2528-1410 e- ISSN 2527-8045

the range [0, 1], ensuring training stability and facilitating model convergence. Finally, we divided the training data in each Training folder from each dataset into a train set and validation set with a ratio of 90:10, where 90% is for training and the remaining 10% is for validation. The split was performed using a fixed random seed of 42 to ensure that the exact partition is reproducible across runs. This normalization and partition step was applied to the LOL-V1, LOL-V2 real capture, and LOL-V2 synthetic datasets individually, so that the final dataset ready for model training consisted of these three datasets without any mixing or merging processes.

2.2. Proposed Model Architecture

Regarding the model architecture, this study proposes a modified U-Net specifically designed to enhance the quality of low-light images, named MRD-UNet. The model consists of two main paths: the contractive path and the expansive path, similar to the standard U-Net. In addition, skip connections are employed to preserve spatial features that may be lost during downsampling [21], and a bottleneck section is included to contain rich feature representations of the image. The overall architecture of MRD-UNet can be seen in **Figure 1**.

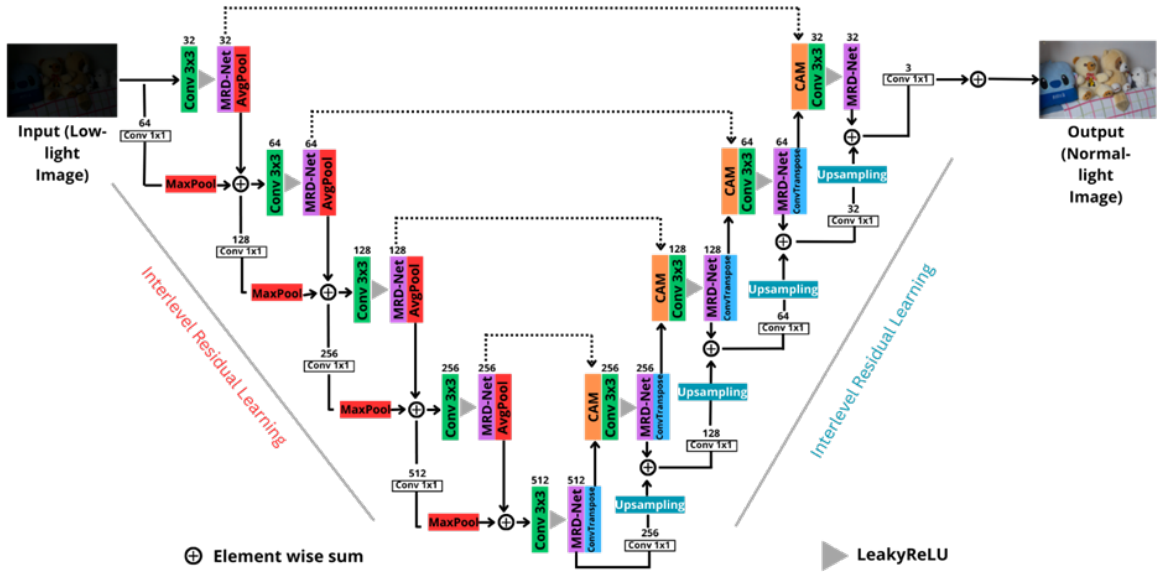


Figure 1. MRD-UNet Architecture

In **Figure 1**, it can be observed that the MRD-UNet model is equipped with a special mechanism called Interlevel Residual Learning (IRL). Unlike conventional residual learning, which is typically applied through shortcut connections after several convolution operations, IRL is implemented across levels at every stage of the U-Net architecture. This mechanism offers several advantages, including maintaining stable gradient flow during backpropagation, preventing excessive loss of image features between levels, and accelerating training convergence. It is worth noting that within the encoder, the main feature path employs 2×2 average pooling for downsampling, while the IRL shortcut branch applies 2×2 max pooling before the element-wise addition. This asymmetric downsampling design is intentional, as average pooling in the main path better preserves background illumination information across levels, while max pooling in the shortcut path retains dominant structural features that would otherwise be suppressed by averaging. Essentially, Interlevel Residual Learning (IRL) is inspired by the original idea of residual learning introduced in *Deep Residual Learning for Image Recognition* by Kaiming He *et al.* [17], but with certain adaptations tailored to the

DOI: https://doi.org/10.17509/ijost.v11i1.90990
p- ISSN 2528-1410 e- ISSN 2527-8045

multilevel feature characteristics of the U-Net architecture. To better understand the difference between conventional residual learning and Interlevel Residual Learning, **Figure 2** provides a comparative illustration.

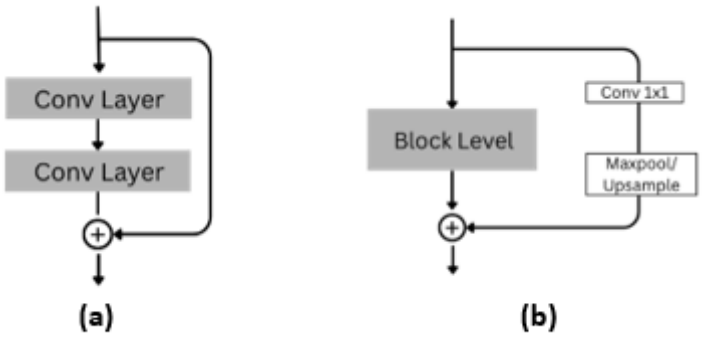


Figure 2. Comparison Between Residual Learning and IRL: (a) Original Residual Learning [17], (b) Interlevel Residual Learning

Figure 2 shows the distinction between the residual learning proposed in [17] and the Interlevel Residual Learning introduced as an innovation in this study. The key difference lies in the fact that conventional residual learning bypasses only convolutional layers without passing through downsampling operations, whereas IRL explicitly spans across downsampling layers. This is the reason it is referred to as Interlevel Residual Learning. In addition to IRL, several other innovations are proposed in this research as detailed the following subsections.

2.2.1. Multiscale Residual Dense Networks (MRD-Net)

The architecture of MRD-Net (Multiscale Residual Dense Network), as illustrated in **Figure 3**, is designed to enhance feature extraction capabilities by leveraging the advantages of multi-scale convolution and residual dense connections. In the encoder and decoder stages, all MRD-Net blocks operate with a dilation rate of 1, preserving fine spatial detail at each hierarchical level. At the bottleneck, however, MRD-Net is configured with a dilation rate of 5 for all convolutional branches, substantially expanding the receptive field at the deepest feature representation level to capture wider contextual dependencies before decoding begins. The design of this architecture is inspired by the Residual Dense Network (RDN) proposed by Zhang *et al*. in their work on image restoration [22]. In this study, RDN has been further modified with the integration of multiscale convolution and asymmetric convolution, as the original RDN employed only 3×3 convolutions. These modifications enable MRD-Net to be more adaptive to both global and local image characteristics, which is particularly critical for the task of low-light image enhancement. The core block of MRD-Net consists of convolutional layers with varying kernel sizes, including Conv 5×5, Conv 3×3, Conv 3×1, and Conv 1×3, allowing the network to capture features at different spatial scales. Specifically, the 5×5 convolution extracts global information with a larger receptive field, while the 3×3 convolution focuses on local details. The combination of Conv 3×1 and Conv 1×3 approximates the effect of a 3×3 convolution but with fewer parameters, while simultaneously improving sensitivity to horizontal and vertical patterns independently. The outputs of these convolutions are integrated through dense connections (represented by blue and green arrows), ensuring that feature information from each layer remains accessible and can be effectively utilized by subsequent layers without degradation.

DOI: https://doi.org/10.17509/ijost.v11i1.90990
p- ISSN 2528-1410 e- ISSN 2527-8045

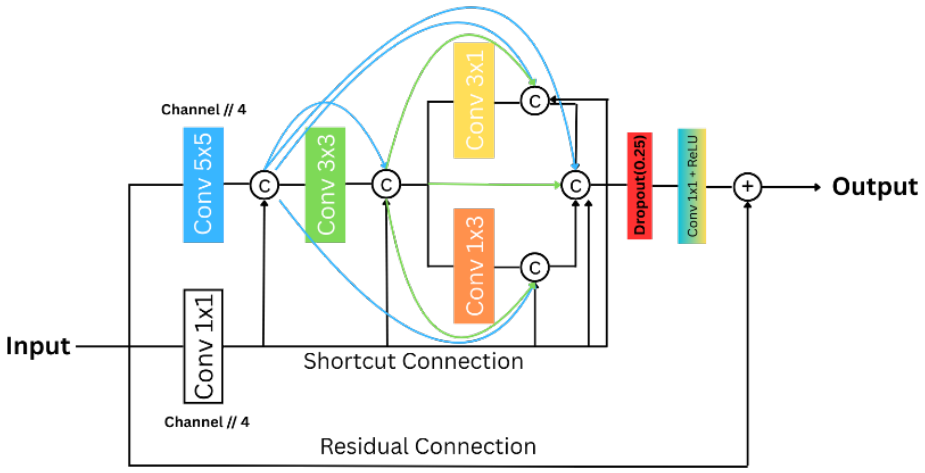


Figure 3. MRD-Net Architecture

The primary strength of MRD-Net lies in its ability to perform multi-scale feature fusion while maintaining information flow stability. Through the integration of residual shortcuts with element-wise summation operation, the network ensures stable gradient propagation during backpropagation, effectively mitigating the vanishing gradient problem commonly encountered in deeper networks. Moreover, the dense connectivity within the block enables each layer to directly utilize features from all preceding layers, thereby enriching the representational capacity and capturing a broader variety of feature patterns. The inclusion of a dropout mechanism (0.25) further reduces the risk of overfitting, enhancing the model's robustness against noise and variations in illumination. Overall, the combination of multi-scale convolutions and residual dense connections empowers MRD-Net to effectively capture both fine-grained details and global structural information in images, making it highly effective for the challenging task of low-light image enhancement.

2.2.2. Contextual Attention Module (CAM)

The Contextual Attention Module (CAM) architecture shown in **Figure 4** is designed to enhance the model's ability to adaptively capture both spatial and channel-wise contextual features, particularly under varying illumination conditions. In several related studies, such as the Illumination Attention Module (IAM) proposed by Wang *et al*. and Global Fusion Attention (GFA) by Yin *et al*., channel-based and spatial-based attention mechanisms were performed sequentially [23, 24]. In contrast, CAM adopts a parallel design for channel attention and spatial attention. A similar parallelization strategy was also employed in *LAU-Net: A Low-Light Image Enhancer with Attention and Resizing Mechanisms* by C.C. Lim *et al*., where Global Average Pooling (GAP) and fully connected networks were used for channel attention, followed by max-pooling and average pooling with local kernels for spatial attention, with the final weights concatenated to form the output [25]. CAM extends this mechanism by introducing separable convolutions, integrating both GAP and Global Max Pooling (GMP), and employing an element-wise summation as the final operation to maintain computational efficiency while enhancing representational richness.

The CAM module begins by merging features from the skip connection and decoder path, followed by a 1×1 convolution to unify channel dimensions. It then splits into two main branches: the first focuses on global contextual understanding through GAP and GMP, which are subsequently processed by a fully connected layer and reshaped back into spatial dimensions. The second branch applies Depthwise Separable Convolutions (DWSC) with 3×3 kernels at three different dilation rates (1, 3, and 7) and average pooling with a 5×5 kernel

DOI: https://doi.org/10.17509/ijost.v11i1.90990
p- ISSN 2528-1410 e- ISSN 2527-8045

(stride 1, padding 2) to obtain global illumination maps. This multi-scale dilated convolution strategy enables the network to efficiently capture fine-grained local details as well as global contextual dependencies, while keeping the parameter count low. Finally, the attention map results from the spatial branch and channel levels are subjected to element-wise multiplication operations to obtain their respective feature maps, and the final result is a combination of pure features, channel attention maps, and spatial attention maps.

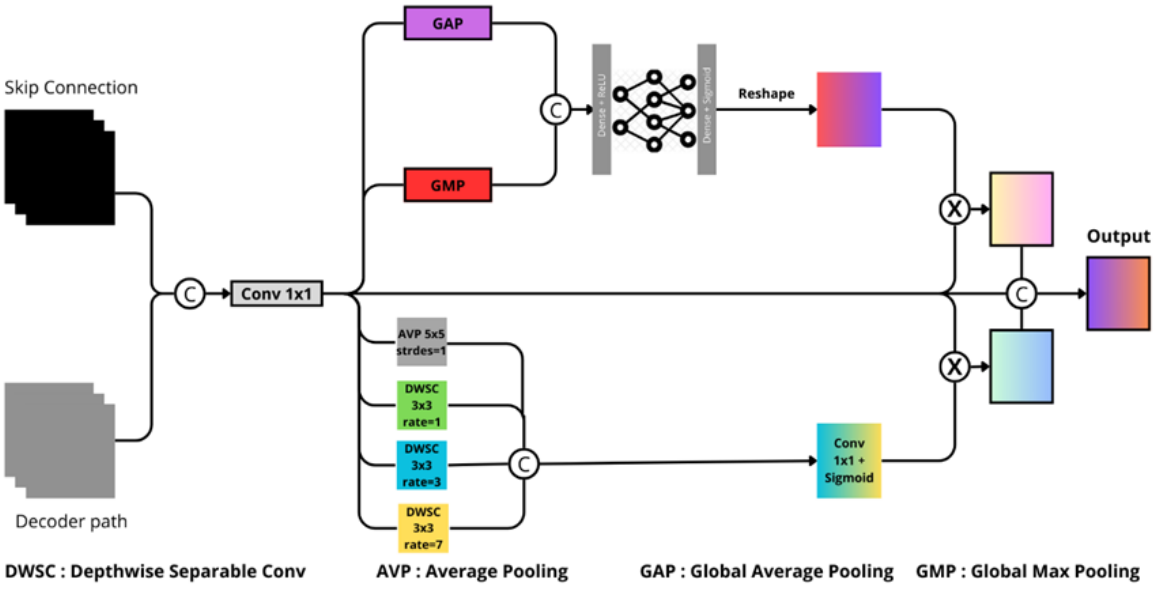


Figure 4. Contextual Attention Module (CAM)

The strength of the CAM module lies in its ability to dynamically adjust the focus of attention under varying lighting conditions. Through the combined use of GAP and GMP, it effectively extracts a global representation that highlights the most relevant channels. Meanwhile, multi-scale DWSC and Average Pooling ensure sensitivity to subtle textures in both bright and dark areas, while also maintaining global spatial awareness. The inclusion of sigmoid activation further normalizes the attention weights, emphasizing features relevant to lighting quality while suppressing less important information. In other words, CAM not only improves the model's robustness to noise and low-light conditions, but also enhances its adaptability in producing more consistent feature representations across different light intensities, making it highly effective for low-light image enhancement and general image quality enhancement tasks.

2.3. Training Technique

The success of deep learning-based research depends not only on sophisticated model architecture and high-quality datasets, but equally on appropriate training procedures that can be reliably reproduced. Hardware and framework for running all experiments in this study were implemented using TensorFlow with the Keras API and executed on the Kaggle Notebook environment with GPU P100 accelerator with 16 GB VRAM. The following subsections discuss the training techniques used for MRD-UNet in detail, divided into two sections: the loss function design and other training configurations applied to MRD-UNet training.

DOI: https://doi.org/10.17509/ijost.v11i1.90990
p- ISSN 2528-1410 e- ISSN 2527-8045

2.3.1. Loss Function

The loss function in this study was specifically designed for the task of Low-Light Image Enhancement and aligned with the evaluation metrics used to assess image quality. According to several recent studies on low-light image enhancement, the implementation of loss functions primarily employs the L1 loss (Mean Absolute Error, MAE) as the core component due to its stability, robustness against outliers, and accuracy in capturing pixel-level differences [26]. In addition to MAE as a standard loss, task-specific losses such as SSIM loss, perceptual loss, and color loss have been widely adopted to further improve the quality of enhanced images [26]. However, perceptual loss often requires substantial computational resources, as it relies on large pre-trained networks such as VGG-Net. In our experiments with MRD-UNet, the inclusion of perceptual loss introduced instability and hindered convergence, making it unsuitable for training MRD-UNet. Regarding color loss, numerous studies in low-light image enhancement have leveraged it for color consistency with different approaches. For example, Jiang *et al.* [27] utilized multiple color spaces to compute differences in color intensities, while Shen *et al.* [28] proposed a more natural approach by comparing color distributions through histograms, which are more sensitive to tonal variations rather than pixel-wise differences. Aharon *et al.* [29] further advanced this concept by employing the Earth Mover's Distance (EMD) between predicted and reference images in the YUV color space. Motivated by the previous studies, we attempted to implement color loss using EMD on RGB color maps, but the results were not significant, leading us to exclude color loss from MRD-UNet.

SSIM loss presents a compelling approach that aligns well with the objective of this study, namely to enhance detail preservation and naturalness in reconstructed images. By incorporating SSIM loss, the model focuses more effectively on luminance, contrast, and structural components, ultimately supporting the generation of sharp and natural images. Tao *et al.* [10] successfully employed SSIM loss alone, without additional loss functions, and achieved strong improvements in both SSIM and PSNR scores. However, the standard global SSIM is insufficient for evaluating local details with higher precision. Multiscale SSIM (MS-SSIM), introduced by Wang *et al.* [30] and popularized by Snell *et al.* [19] as a loss function, provides a more powerful approach to maximizing image naturalness and perceptual quality. Nevertheless, we observed that SSIM or MS-SSIM alone did not fully emphasize texture and fine structural details. Therefore, we introduced an additional loss function designed specifically to evaluate texture and structure more accurately. This was achieved by combining Laplacian feature maps, which are highly sensitive to texture information, with Sobel feature maps, which are effective for edge structure detection, into a new function termed edge-texture loss. This additional loss is expected to significantly improve the evaluation of structural fidelity. Consequently, the final loss function for MRD-UNet training was formulated as in Eq. (1).

$$L_{total} = \alpha \times L_{MAE} + \beta \times L_{MS-SSIM} + \lambda \times L_{edge-texture} \quad (1)$$

As shown in Equation (1), the proposed loss function is defined as the summation of three different loss terms: MAE loss, MS-SSIM loss, and edge-texture loss, each weighted by coefficients *(α, β, λ)* to regulate their contribution to the overall loss. The coefficient α is set to 0.5, giving dominant weight to MAE loss as the fundamental component, meaning that MAE loss contributes 50% to the total loss calculation. The coefficient β is set to 0.3, limiting the contribution of MS-SSIM loss to only 30% of its original value, as higher SSIM weight was found to introduce color distortions and instability in the total loss. Similarly, λ is set to 0.2,

DOI: https://doi.org/10.17509/ijost.v11i1.90990
p- ISSN 2528-1410 e- ISSN 2527-8045

allowing only 20% of the edge-texture loss contribution to the total loss. The detailed mathematical formulation of each individual loss term is discussed in the following subsections.

The MAE loss, serving as the fundamental loss for measuring pixel-level discrepancies, is calculated by computing the absolute difference between the ground truth pixels and the model's predicted pixels. Its mathematical definition is given in Equation (2).

$$L_{MAE}(x,y) = \frac{1}{N}\sum_{i=1}^{N}|x_i - y_i| \quad (2)$$

where x_i denotes the predicted pixel intensity at the i-th pixel and y_i denotes the corresponding ground-truth pixel intensity, while N represents the total number of pixels in the image.

The second component is the MS-SSIM loss. Although SSIM was originally designed as a single-scale metric, Wang *et al*. [30] optimized it into a multi-scale version, and MS-SSIM was first introduced as a loss function by Snell *et al*. [19]. The term *multi-scale* here does not refer to having multiple receptive fields, but rather to a mechanism of repeatedly downsampling the image a predetermined number of times. The more levels that are evaluated, the more detailed the results will be. In the original paper by Wang *et al*. [30], MS-SSIM is performed with five downsampling steps, and the SSIM score is calculated at each level. A scale factor is then applied to control the contribution of each level's SSIM score.

In contrast, the MS-SSIM we implemented uses only four downsampling steps (four levels of feature maps). The SSIM scores at each level were then averaged, without assigning specific weights to individual levels. We employed 2×2 average pooling with stride 2 as the downsampling method, with the aim of better preserving background information at each level, not only structural and textural details. Furthermore, we did not assign explicit weights to each level since the overall average score turned out to be more stable and provided a balanced representation across all levels. Thus, the MS-SSIM we applied can be expressed as shown in equation (3) – (5).

$$SSIM(x,y) = \frac{(2\mu_x\mu_y+C1).(2\sigma_{xy}+C2)}{(\mu_x^2+\mu_y^2+C1).(\sigma_x^2+\sigma_y^2+C2)} \quad (3)$$

$$MS\_SSIM(x,y) = \frac{1}{M}\sum_{i=1}^{M} SSIM\,(x_i,y_i) \quad (4)$$

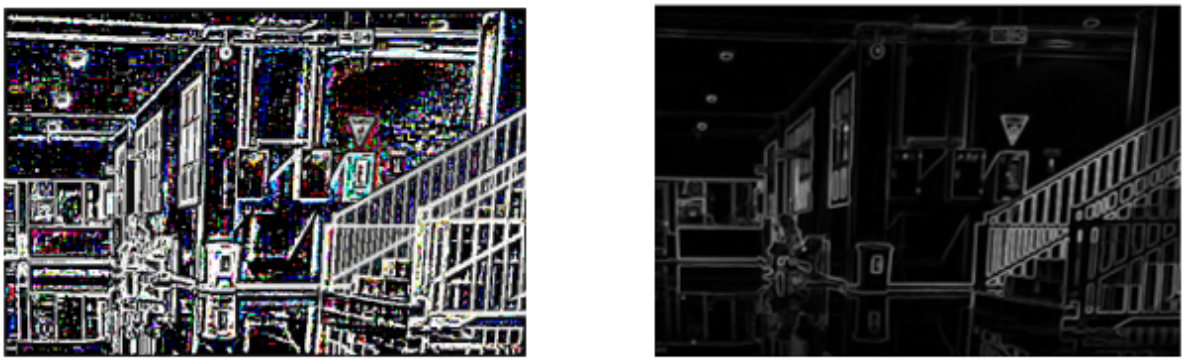
$$L_{MS-SSIM} = 1 - MS\_SSIM \quad (5)$$

Equation (3) represents the concept of single-scale SSIM, which measures luminance, contrast, and structure only on the original image, where *x* denotes the reconstructed image generated by the model and *y* is the reference (ground truth) image, both assumed to lie within the range [0, 1]. The MS-SSIM we designed is presented in Equation (4), where *M* represents the number of desired downsampling steps. Finally, in Equation (5), the dissimilarity is computed by subtracting the obtained value from the maximum value of 1.

The next component is the Edge-Texture Loss, which we specifically designed to achieve more accurate edge-aware and texture-aware learning. This loss is constructed from two fundamental filters: the Laplacian filter and the Sobel filter. The concept of the Laplacian operator originates from partial differential mathematics and was first introduced by Marr *et al*. [31] in 1979 for applications in computer vision and edge detection. The Laplacian filter is well-known for its high sensitivity to micro-textures and high-frequency details, making it highly suitable for accurately extracting texture information from images. However, the Laplacian filter is often not robust to noise, which can result in images that appear sharp but

DOI: https://doi.org/10.17509/ijost.v11i1.90990
p- ISSN 2528-1410 e- ISSN 2527-8045

contain high levels of noise. In contrast, the Sobel filter was first introduced by Irwin Sobel and Gary Feldman, employing a 3×3 convolution with horizontal and vertical differential weights, commonly referred to as the Sobel-Feldman Operator. The Sobel filter is widely recognized for its effectiveness in detecting macro edges (structural information) while maintaining robustness against moderate noise. However, it is less sensitive to fine texture details, focusing instead on the primary contours. The differences between the two filters can be observed in **Figure 5**.



(a) (b)

Figure 5. Differences in Feature Maps between Laplacian and Sobel filter: (a) Laplacian Filter Result, (b) Sobel Filter Result

From **Figure 5 (a)**, we can observe that the Laplacian filter produces feature maps that are sharp and rich in detail, whereas in **Figure 5 (b)**, the Sobel filter generates features that are cleaner from noise and strongly emphasize macro structures. Therefore, we consider these two concepts to be complementary, as they help produce images that are both sharp and reasonably robust to noise. To this end, we combine both filters into a single formulation with weighting factors of γ = 0.6 for the Sobel filter and δ = 0.4 for the Laplacian filter. This combined formulation is then integrated as one of the loss functions in MRD-UNet, which we refer to as the Edge-Texture Loss, with its mathematical representation given in Equation (6) – (15).

Sobel Gradient Magnitude:

$$G_x(I) = I * S_x \tag{6}$$

$$G_y(I) = I * S_y \tag{7}$$

where I is an image, S_x is the horizontal Sobel filter and S_y is the vertical Sobel filter, defined as:

$$S_x = \begin{matrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{matrix} \tag{8}$$

$$S_y = \begin{matrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{matrix} \tag{9}$$

while the Sobel Gradient Magnitude is defined as:

$$Sobel(I) = \sqrt{(G_x(I))^2 + (G_y(I))^2} \tag{10}$$

DOI: https://doi.org/10.17509/ijost.v11i1.90990
p- ISSN 2528-1410 e- ISSN 2527-8045

Laplacian Response:

$$Laplacian(I) = |\, I * L \,| \tag{11}$$

where I is an image and L is Laplacian Filter, defined as:

$$L = \begin{matrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{matrix} \tag{12}$$

Calculate Sobel and Laplacian loss with L1 distance:

$$L_{sobel}(x,y) = \frac{1}{N}\sum_{i=1}^{N} |Sobel(x)_i - Sobel(y)_i| \tag{13}$$

$$L_{laplace}(x,y) = \frac{1}{N}\sum_{i=1}^{N} |Laplacian(x)_i - Laplacian(y)_i| \tag{14}$$

Final Edge-Texture Loss:

$$L_{edge-texture}(x,y) = \gamma \times L_{sobel}(x,y) + \delta \times L_{laplace}(x,y) \tag{15}$$

From the mathematical formulation presented in Equations (6) through (15), it can be observed that the Edge-Texture Loss is solely derived from the L1 distance between the Sobel gradient magnitudes of the reconstructed image and the reference image, as well as the L1 distance between the Laplacian features of the reconstructed image and the reference image. The combination of these two components is controlled by the weighting factors γ and δ, where γ = 0.6 is set to emphasize the Sobel loss for suppressing noise, and δ = 0.4 is assigned to enhance texture sharpness through the Laplacian response.

2.3.2. Training Strategy

The training strategy for MRD-UNet is organized into three distinct phases, each with a fixed batch size of 8 and a maximum of 1,000 epochs. All convolutional and dense layer weights were initialized using Glorot uniform initialization, and all bias terms were initialized to zero. Early stopping was applied in all three phases, monitoring validation loss as the criterion for termination. In the first phase, the model was trained using the Adam optimizer with a learning rate of 0.0001, β_1 = 0.9, and β_2 = 0.99, with an early stopping patience of 20 epochs. After completing the first phase, the model weights were saved and reloaded, and retraining continued in the second phase using the AdamW optimizer with a learning rate of 0.00005 and an early stopping patience of 40 epochs. In the third and final phase, the model was again reloaded and further trained with AdamW at a reduced learning rate of 0.00001, maintaining the early stopping patience at 40 epochs. Through this progressive three-phase scheme, MRD-UNet was able to achieve stable convergence while avoiding premature termination in the later fine-tuning stages. This training strategy was applied for training MRD-UNet in LOL-V1, LOL-V2 Real Capture, and LOL-V2 Synthetic.

3. RESULTS AND DISCUSSION

3.1. Result Comparison with the Other Model

To objectively evaluate the performance of MRD-UNet, it is compared against several representative state-of-the-art methods in low-light image enhancement, including EnlightenGAN [11], KinD [9], KinD++ [2], and SNR-Net [12]. EnlightenGAN was selected for its capacity to produce perceptually natural images, KinD and KinD++ for their strength in maintaining color consistency, and SNR-Net as a modern transformer-based approach. To ensure fairness, all baseline results were obtained by running inference on each original pre-

DOI: https://doi.org/10.17509/ijost.v11i1.90990
p- ISSN 2528-1410 e- ISSN 2527-8045

trained model and Python function wrapper directly on the same test splits used to evaluate MRD-UNet, rather than reproducing reported values from the original papers. All models received identical preprocessed inputs and were evaluated using the same metrics and libraries described in the quantitative comparison protocol above. This section is further divided into two parts: quantitative comparison using several predefined evaluation metrics, and qualitative/visual comparison by presenting example images generated by each model.

In this section, we present the quantitative comparison results, including PSNR for pixel-level intensity accuracy, SSIM [32] for structural detail, contrast, and luminance, LPIPS [33] for perceptual quality, and NIQE [34] for image naturalness. PSNR and SSIM were computed on grayscale images converted from RGB with a data range of 1.0. LPIPS was computed using the VGG-16 backbone [35], and NIQE was computed on RGB outputs clipped to [0, 1]. All metrics were evaluated on the official test splits of each dataset without any post-processing. The quantitative comparison results are presented in **Tables 1 to 3**.

Table 1. Quantitative Comparison in LOL-V1 [8] Dataset.

Methods	PSNR	SSIM	LPIPS	NIQE
EnlightenGAN [11]	18.87	0.800	0.306	4.74
KinD [9]	12.51	0.742	0.288	4.91
KinD++ [2]	18.34	0.820	0.239	5.41
SNR-Net [12]	25.18	0.849	0.229	5.72
MRD-UNet (ours)	26.09	0.911	0.135	4.51

Table 2. Quantitative Comparison in LOL-V2 Real Capture [20] Dataset.

Methods	PSNR	SSIM	LPIPS	NIQE
EnlightenGAN [11]	20.43	0.811	0.303	5.25
KinD [9]	11.03	0.722	0.321	5.27
KinD++ [2]	18.25	0.789	0.240	5.97
SNR-Net [12]	21.16	0.849	0.227	5.34
MRD-UNet (ours)	19.65	0.873	0.159	4.87

Table 3. Quantitative Comparison in LOL-V2 Synthetic [20] Dataset.

Methods	PSNR	SSIM	LPIPS	NIQE
EnlightenGAN [11]	18.87	0.783	0.214	5.78
KinD [9]	14.12	0.765	0.296	6.44
KinD++ [2]	18.00	0.821	0.233	6.56
SNR-Net [12]	23.51	0.925	0.095	5.45
MRD-UNet (ours)	24.43	0.947	0.057	5.63

Based on the quantitative evaluations conducted on three benchmark datasets (LOL-V1, LOL-V2 Real, and LOL-V2 Synthetic) as presented in **Tables 1–3**, the proposed MRD-UNet consistently demonstrates superior performance compared to representative state-of-the-art models, including EnlightenGAN [11], KinD [9], KinD++ [2], and SNR-Net [12], particularly in terms of SSIM and LPIPS.

On the LOL-V1 dataset (**Table 1**), MRD-UNet achieved SSIM = 0.911 and PSNR = 26.09, the highest scores among all models, indicating its strong ability to generate images with better structural fidelity and reduced noise. SNR-Net [12], a modern transformer-based approach, also demonstrated competitive performance with PSNR = 25.18, but still fell short compared to MRD-UNet. In contrast, Retinex-based models such as KinD [9] and KinD++ [2] lagged

DOI: https://doi.org/10.17509/ijost.v11i1.90990
p- ISSN 2528-1410 e- ISSN 2527-8045

significantly in terms of both structural quality and detail sharpness, as reflected in their relatively low PSNR values (approximately 12–18 dB), which aligns with their known tendency to produce overexposed outputs with color distortion when adapted to deep learning frameworks. EnlightenGAN [11], which is widely recognized for producing visually natural images, achieved NIQE = 4.74, but MRD-UNet surpassed it with NIQE = 4.51, yielding more natural-looking results. Furthermore, from the perspective of perceptual quality, MRD-UNet delivered a substantially lower LPIPS = 0.135, whereas all other models reported values above 0.2, highlighting its superior perceptual similarity to the ground truth. In summary, MRD-UNet achieved the strongest overall performance on the LOL-V1, with significantly higher perceptual and structural quality. These results confirm that the baseline models have difficulty in maintaining good and accurate structural details because the SSIM and LPIPS scores are far below those of the MRD-UNet.

For the LOL-V2 Real dataset (**Table 2**), MRD-UNet maintained its superiority in terms of SSIM (0.873), although its PSNR (19.65) was slightly lower than that of SNR-Net [12] (21.16) and EnlightenGAN [11] (20.43). This is consistent with the noted tendency of transformer-based approaches like SNR-Net to trade structural sharpness for noise suppression [12], which favors global luminance recovery over fine texture fidelity. The PSNR gap observed here is therefore expected, as a model designed to prioritize structural fidelity and perceptual similarity will not necessarily lead in a metric that measures pixel-wise intensity accuracy alone. Notably, MRD-UNet achieved a much lower LPIPS (0.159) compared to all other models, which demonstrates that the generated results are perceptually closer to the ground truth from a human visual perspective. In addition, MRD-UNet achieved the best naturalness with NIQE = 4.87, whereas the other models yielded scores above 5, confirming that MRD-UNet produced more natural results. Hence, MRD-UNet remains globally superior in this setting, even though its PSNR did not lead among all compared methods.

On the LOL-V2 Synthetic dataset, MRD-UNet again outperformed nearly all competing methods across most metrics, achieving SSIM = 0.947, PSNR = 24.43, and LPIPS = 0.057, all of which are significantly better than the other models. SNR-Net [12] came close, with SSIM = 0.925 and PSNR = 23.51, but its higher LPIPS (0.095) indicated lower perceptual fidelity compared to MRD-UNet. This pattern reflects the trade-off inherent in noise-suppression-oriented designs, where prioritizing smoothness over structural supervision limits perceptual fidelity as measured by LPIPS. Meanwhile, KinD [9] and KinD++ [2] continued to underperform, particularly under synthetic low-light conditions with high illumination variability, as reflected by their stagnant PSNR values (14–18 dB) and substantially higher LPIPS scores. In terms of naturalness, SNR-Net [12] slightly outperformed MRD-UNet with a NIQE score of 5.45, compared to 5.63 for MRD-UNet. Nevertheless, MRD-UNet remained superior in SSIM, PSNR, and LPIPS, despite being slightly behind in NIQE. The slight NIQE gap is worth noting, as NIQE tends to penalize high-frequency detail as unnatural, meaning models that produce sharper structural outputs may score slightly lower even when their outputs are perceptually closer to the ground truth, as confirmed by the lower LPIPS of MRD-UNet.

Overall, these comparisons demonstrate that MRD-UNet exhibits strong advantages in preserving structure, texture, and fine visual details across both real and synthetic low-light conditions. This outcome confirms that treating low-light enhancement primarily as a denoising or illumination-normalization problem, as most prior methods do, leaves structural detail preservation inadequately addressed. The consistent dominance of MRD-UNet in SSIM and LPIPS confirms that explicitly supervising edge and texture information during training, rather than relying solely on pixel-wise losses, produces measurably stronger structural

DOI: https://doi.org/10.17509/ijost.v11i1.90990
p- ISSN 2528-1410 e- ISSN 2527-8045

fidelity and perceptual naturalness, establishing it as a more adaptive and reliable solution compared to existing state-of-the-art methods.

In this section, we present the visual results produced by each model, alongside the input low-light images and their corresponding ground truth, with the aim of illustrating the extent to which our method is capable of performing low-light image enhancement under various conditions. We also provide specific local patches from the images to give a clearer depiction of local-level quality. The results are demonstrated and discussed across three experimental categories, namely the LOL-V1 dataset [8], the LOL-V2 real-capture dataset, and the LOL-V2 synthetic dataset [20]. More detailed results and explanations can be seen in **Figures 6–8**.

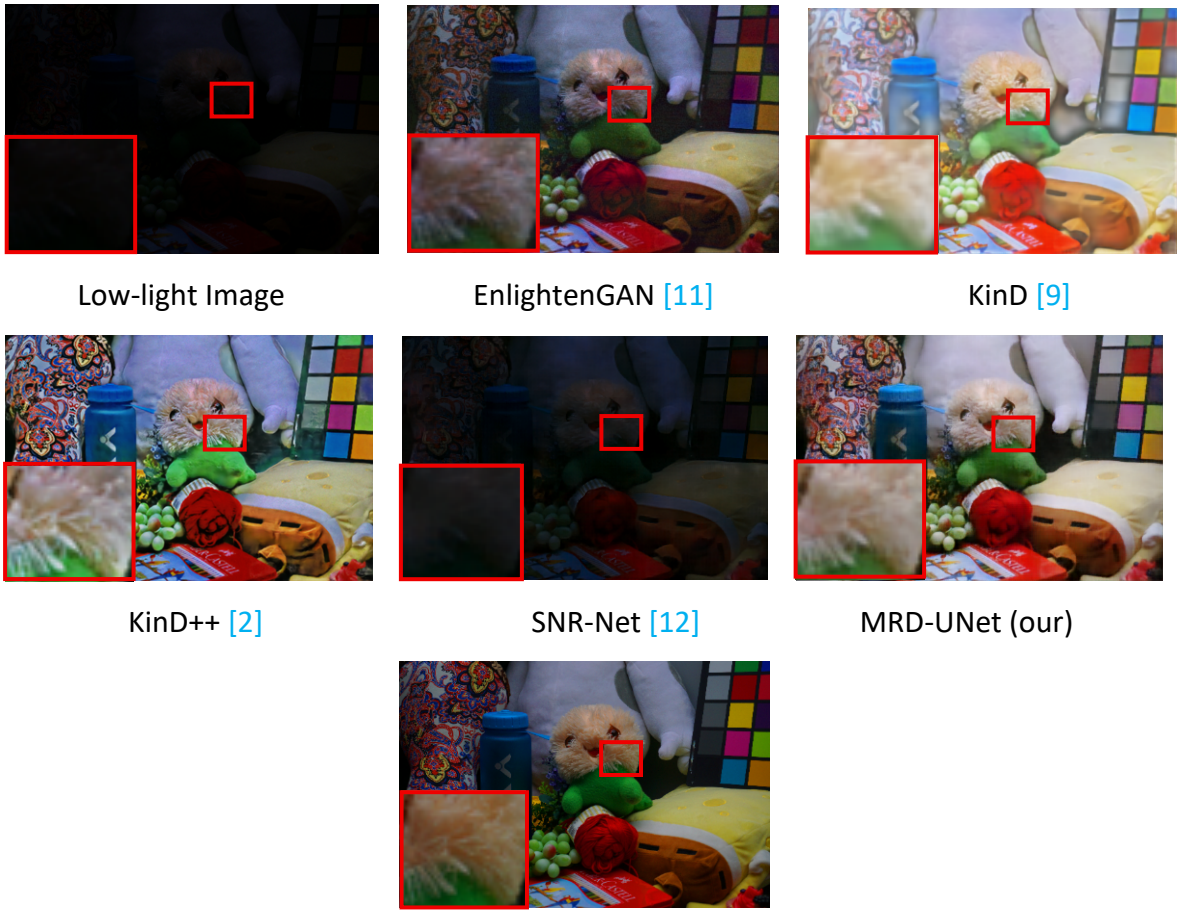


Figure 6. Visual comparison of our method with State-of-the-art methods on LOL-V1 [8] Dataset. Our method produces a more natural image overall and remains sharp in every detail of the image patch.

Figure 6 shows representative samples of the experimental results on the LOL-V1 dataset. It can be observed that the images generated by our model appear highly natural at the global level, with accurate and stable colors (avoiding oversaturation). The fine details shown in the local patches demonstrate that our method is more effective at preserving structural details compared to other evaluated methods. KinD++ [2] also performs relatively well in maintaining structural details, but tends to produce oversaturated and less natural results. EnlightenGAN [11] generates images that appear fairly natural, but they often still contain noticeable noise. KinD [9] produces images with excessively high illumination intensity, resulting in blurred and less sharp fine details, whereas SNR-Net [12] fails to achieve optimal visibility, with outputs

DOI: https://doi.org/10.17509/ijost.v11i1.90990
p- ISSN 2528-1410 e- ISSN 2527-8045

that remain overly dark. Consequently, MRD-UNet can be considered reliable in producing high-quality images on the LOL-V1 dataset.

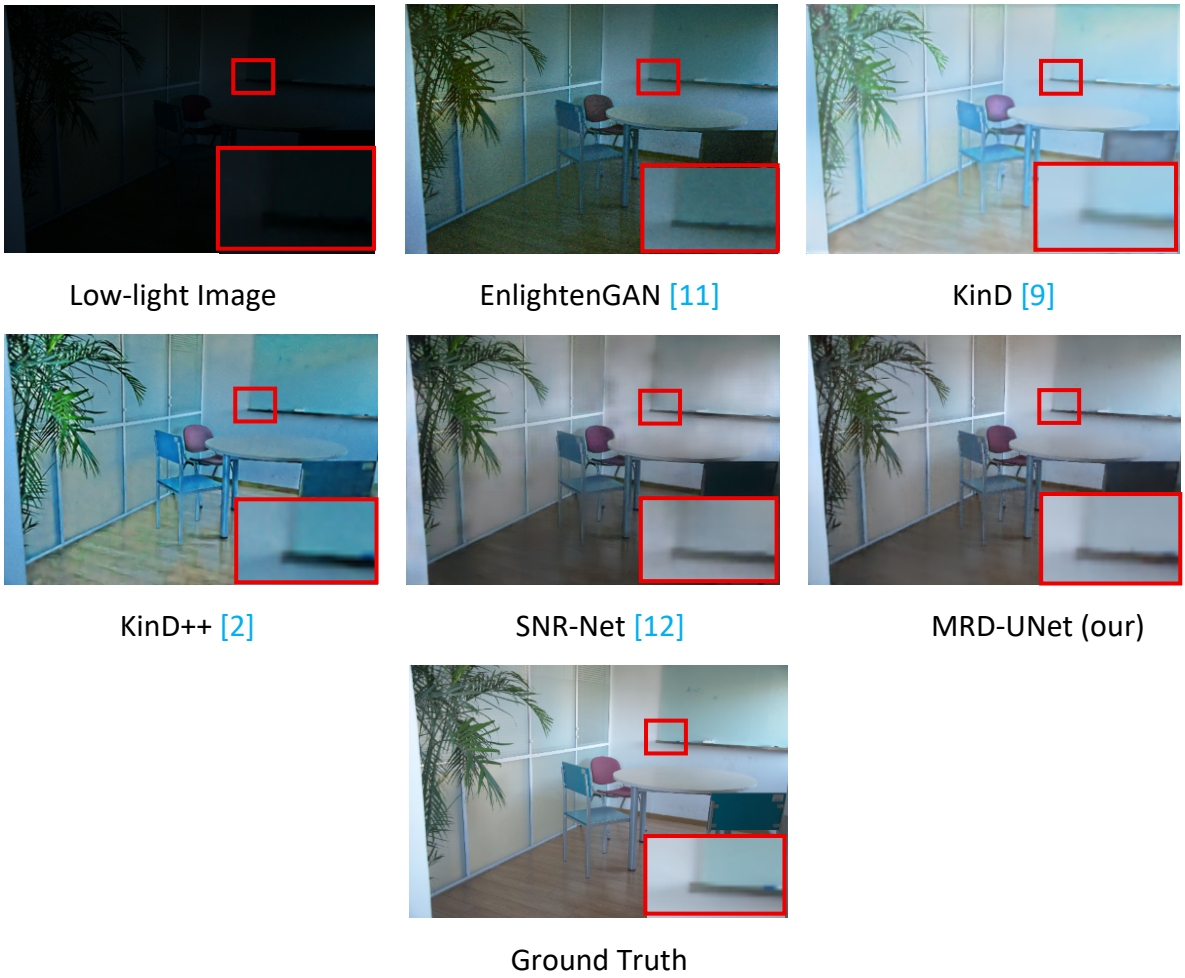


Figure 7. Visual comparison of our method with State-of-the-art methods on LOL-V2 [20] Real Capture Dataset. Our method is cleaner from local disturbances such as noise and more stable in lighting.

Figure 7 presents sample results on the LOL-V2 real-capture dataset, where MRD-UNet again demonstrates superiority in terms of naturalness, global illumination stability, accurate color distribution (without distortion), and freedom from local artifacts or noise. SNR-Net [12] emerges as a competitive method that also generates highly natural-looking images, although random black artifacts remain clearly visible to the human eye. EnlightenGAN [11] also shows promising performance in preserving structural details, but the global colors of the images are distorted, exhibiting a greenish tone that reduces naturalness. Meanwhile, KinD [9] and KinD++ [2] continue to suffer from issues of oversaturation and overbrightness, with KinD++ [2] additionally exhibiting bluish color distortion. Therefore, MRD-UNet can also be regarded as the most suitable option for the LOL-V2 real-capture evaluation.

Finally, **Figure 8** illustrates the results on the LOL-V2 synthetic dataset. MRD-UNet consistently demonstrates superiority across various aspects, producing results highly similar to the reference ground truth. SNR-Net [12] again performs competitively, but instabilities appear in the rendering of haze/smoke, which manifests as unnatural linear streaks of strong contrast. This phenomenon deviates from the reference image, where the haze distribution is smooth and natural. Furthermore, almost all evaluated models fail to achieve accurate and consistent color reproduction, often displaying dull greenish tones. In contrast, MRD-UNet

DOI: https://doi.org/10.17509/ijost.v11i1.90990
p- ISSN 2528-1410 e- ISSN 2527-8045

convincingly demonstrates that the distribution of color intensities in its outputs closely matches the ground truth, while also preserving sharper structural details. Thus, MRD-UNet remains highly promising as a reliable model when confronted with images exhibiting distributions and characteristics similar to those in the LOL-V2 synthetic dataset.

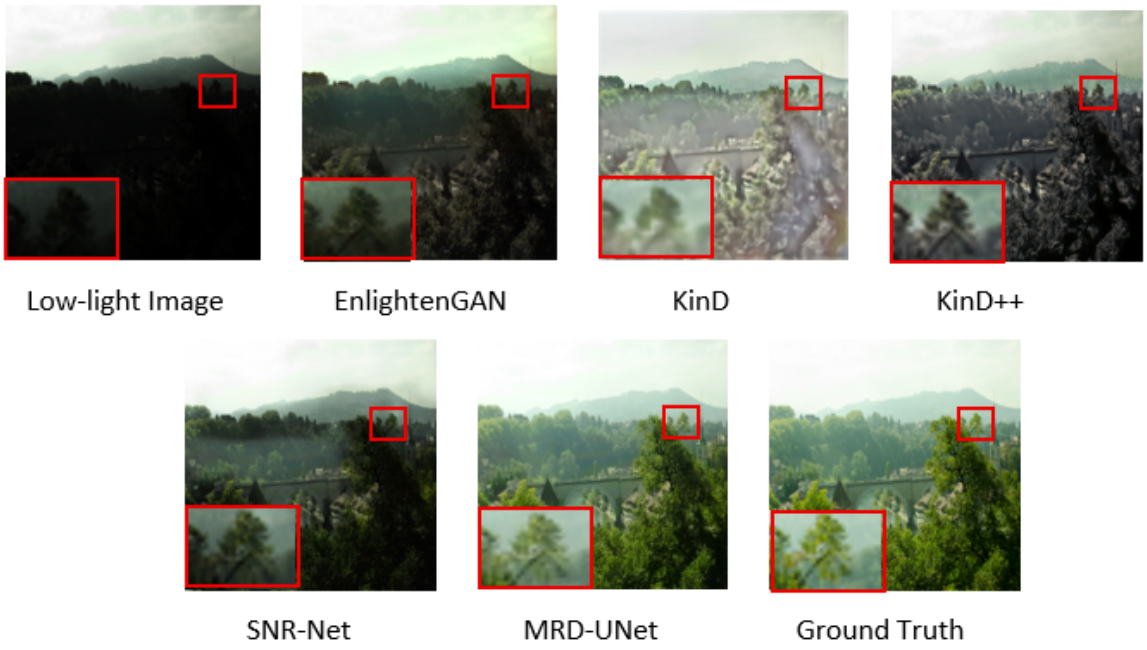


Figure 8. Visual comparison of our method with State-of-the-art methods on LOL-V2 [20] Synthetic Dataset. Our method has higher color accuracy and is visually very similar to the ground truth.

3.2. Ablation Study

To verify the contribution of each component integrated into MRD-UNet, we conducted a dedicated ablation study. This experiment aims to validate how, and to what extent, several modules and mechanisms within MRD-UNet contribute to its overall performance. These components include the Multiscale Residual Dense Networks (MRD-Net), Interlevel Residual Learning (IRL), the Contextual Attention Module (CAM), and the custom loss function consisting of the edge-texture guided loss and multiscale SSIM. This ablation study is designed to isolate the contribution of each component and identify whether their inclusion is individually justified.

The ablation experiments were conducted exclusively on the LOL-V1 real-capture dataset using a shortened two-phase training strategy, with a fixed batch size of 8 and a maximum of 1,000 epochs per phase. Early stopping was applied in both phases, monitoring validation loss as the termination criterion. In the first phase, each ablation variant was trained using the Adam optimizer with a learning rate of 0.0001, $\beta_1 = 0.9$, $\beta_2 = 0.99$, and an early stopping patience of 10 epochs. In the second phase, training was resumed from the saved weights using the AdamW optimizer with a learning rate of 0.00005 and an early stopping patience of 20 epochs. This two-phase scheme differs from the three-phase strategy used for the full benchmark training and was intentionally simplified to reduce computational cost while still highlighting the relative contribution of each component. The results of this experiment are presented in the form of a quantitative comparison to provide a more objective assessment, as shown in **Table 4**.

DOI: https://doi.org/10.17509/ijost.v11i1.90990
p- ISSN 2528-1410 e- ISSN 2527-8045

Table 4. Ablation study of methods in MRD-UNet

Methods	PSNR	SSIM	LPIPS	NIQE
Without Custom Loss Function	24.03	0.864	0.288	4.54
Without MRD-Nets	24.66	0.891	0.235	4.87
Without CAM	24.72	0.906	0.203	4.69
Without IRL	24.25	0.902	0.218	4.72
MRD-UNet	24.95	0.899	0.213	4.51

Table 4 demonstrates that all components have significant contributions, albeit in different aspects of image quality. When MRD-UNet was trained with all components, the model achieved SSIM = 0.899, PSNR = 24.95, LPIPS = 0.213, and NIQE = 4.51. This configuration establishes a balanced combination across all image quality aspects and serves as the baseline for subsequent comparisons.

Without Custom Loss Function. When MRD-UNet was trained without the proposed custom loss and instead used only MAE (Mean Absolute Error), the SSIM score dropped considerably from 0.899 to 0.864. This indicates that the custom loss plays a crucial role in preserving structural quality, contrast, and luminance, which directly impacts the naturalness of the image. This finding is further supported by the NIQE score, which increased from 4.51 to 4.54, where higher NIQE values indicate lower perceived naturalness. Additionally, PSNR decreased sharply from 24.95 to 24.03, confirming that the custom loss also contributes to per-pixel intensity accuracy. Finally, LPIPS increased significantly from 0.213 to 0.288 (a lower score means a more natural image), emphasizing the importance of the custom loss in producing perceptually high-quality images. These variations are theoretically consistent with the design of the custom loss, where MS-SSIM directly optimizes luminance, contrast, and structure, while the Sobel and Laplacian supervision enforces edge and texture alignment, explaining why LPIPS is the most severely affected metric when this component is removed.

Multiscale Residual Dense Networks (MRD-Nets). When MRD-UNet was trained without MRD-Nets at each block level (replaced by a 3×3 convolution + ReLU), all evaluation metrics exhibited performance degradation. Specifically, SSIM dropped to 0.8913, PSNR decreased to 24.66, while LPIPS and NIQE increased to 0.235 and 4.87, respectively. This suggests that MRD-Nets provide benefits similar to those of the custom loss function, though their contribution is less pronounced. The similarity is likely because both mechanisms address feature richness through different means, one architectural and one supervisory, and neither alone is sufficient to fully substitute the other. The consistent degradation across all metrics further confirms that the additional complexity of MRD-Net is justified by the quality gains it produces.

Contextual Attention Module (CAM). When CAM was removed, SSIM slightly improved to 0.906 and LPIPS decreased to 0.203, both surpassing the full-component configuration. This counterintuitive result suggests that CAM may introduce a small degree of feature blending that marginally softens structural details, as its global pooling operations aggregate spatial context across the entire feature map. While this slightly reduces local sharpness metrics, it appears to benefit global illumination consistency, as evidenced by the PSNR and NIQE drops when CAM is removed.

Interlevel Residual Learning (IRL). This mechanism is particularly effective in producing smoother and cleaner images. When IRL was removed, a phenomenon similar to CAM was observed: SSIM increased slightly to 0.902, indicating little impact on structural sharpness. However, PSNR dropped considerably to 24.25, while LPIPS and NIQE increased to 0.218 and 4.72, respectively. This suggests that although IRL may not enhance structural detail

DOI: https://doi.org/10.17509/ijost.v11i1.90990
p- ISSN 2528-1410 e- ISSN 2527-8045

sharpness, it provides substantial benefits by improving smoothness and cleanliness (higher PSNR) and by yielding more natural images (lower LPIPS and NIQE) when included. Overall, the ablation results demonstrate that all components within MRD-UNet are significant, with each playing a distinct role that is not fully captured by any single evaluation metric.

4. CONCLUSION

We proposed MRD-UNet, a U-Net-based architecture that integrates MRD-Net, IRL, CAM, and a custom Edge-Texture Guided loss to address the structural detail preservation and perceptual naturalness limitations prevalent in existing low-light image enhancement methods. Benchmark evaluations on LOL-V1 and LOL-V2 (real capture and synthetic) demonstrate that MRD-UNet improves SSIM and LPIPS over EnlightenGAN, KinD, KinD++, and SNR-Net across all three datasets, and achieves competitive PSNR on LOL-V1 and LOL-V2 synthetic, although a PSNR gap on the LOL-V2 real capture dataset indicates that pixel-level luminance recovery under diverse real-world conditions remains a limitation. The ablation study confirms that each component serves a distinct role, with the custom loss and MRD-Net most strongly contributing to structural sharpness and perceptual quality, and IRL and CAM primarily benefiting pixel-level accuracy and image naturalness. These properties position MRD-UNet as a practically relevant solution for detail-sensitive applications such as surveillance imaging, mobile photography, and embedded vision systems, where perceptual naturalness and structural fidelity are prioritized alongside computational efficiency. Future work may address the identified PSNR limitation by exploring training on more diverse paired real-world datasets and investigating adaptive loss weighting strategies, though such extensions may involve additional complexity trade-offs.

5. ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Faculty of Computer Science, University of Pembangunan Nasional Veteran Jawa Timur, Surabaya, Indonesia, for the institutional support and resources provided throughout this research.

6. AUTHORS' NOTE

The authors declare that there is no conflict of interest regarding the publication of this article. The authors confirmed that the paper was free of plagiarism.

7. REFERENCES

[1] Hao, S., Han, X., Guo, Y., Xu, X., and Wang, M. (2020). Low-light image enhancement with semi-decoupled decomposition. *IEEE Transactions on Multimedia, 22*(12), 3025–3038.

[2] Zhang, Y., Guo, X., Ma, J., Liu, W., and Zhang, J. (2021). Beyond brightening low-light images. *International Journal of Computer Vision, 129*(4), 1013–1037.

[3] Cheng, H. D., and Shi, X. J. (2004). A simple and effective histogram equalization approach to image enhancement. *Digital Signal Processing, 14*(2), 158–170.

[4] Vijayalakshmi, D., Nath, M. K., and Acharya, O. P. (2020). A comprehensive survey on image contrast enhancement techniques in spatial domain. *Sensing and Imaging, 21*(1), 21-40.

[5] Pizer, S. M. (1990). Contrast-limited adaptive histogram equalization: Speed and effectiveness. *Proceedings of the First Conference on Visualization in Biomedical Computing, 1990,* 337-345.

[6] Jobson, D. J., Rahman, Z., and Woodell, G. A. (1997). Properties and performance of a center/surround retinex. *IEEE Transactions on Image Processing, 6*(3), 451–462.

[7] Fu, X., Zeng, D., Huang, Y., Zhang, X.-P., and Ding, X. (2016). A weighted variational model for simultaneous reflectance and illumination estimation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016*, 2782-2790.

[8] Wei, C., Wang, W., Yang, W., and Liu, J. (2018). Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv, 1808, 04560.*

[9] Zhang, Y., Zhang, J., and Guo, X. (2019). Kindling the darkness: A practical low-light image enhancer. *arXiv preprint arXiv, 1905, 04161.*

[10] Tao, L., Zhu, C., Xiang, G., Li, Y., Jia, H., and Xie, X. (2017). LLCNN: A convolutional neural network for low-light image enhancement. *Proceedings of the IEEE Visual Communications and Image Processing (VCIP), 2017*, 1-4.

[11] Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., and Wang, Z. (2021). EnlightenGAN: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing, 30*, 2340-2349.

[12] Xu, X., Wang, R., Fu, C.-W., and Jia, J. (2022). SNR-aware low-light image enhancement. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022,* 17714-17724.

[13] Zamir, S.W., Arora, A., Khan, S.H., Hayat, M., Khan, F.S., and Yang, M. (2022). Restormer: Efficient transformer for high-resolution image restoration. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022*, 5718-5729.

[14] Wang, T., Zhang, K., Shen, T., Luo, W., Stenger, B., and Lu, T. (2022). Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. *arXiv preprint arXiv, 2212, 11548.*

[15] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *arXiv preprint arXiv, 1505, 04597.*

[16] Cao, J., Chen, Z., Cui, H., Ji, X., Wang, X., Liang, Y., and Tian, Y. (2023). Improved wavelet prediction superresolution reconstruction based on U-Net. *IET Image Processing, 17*, 3464–3476.

[17] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016,* 770-778.

[18] Huang, G., Liu, Z., and Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017,* 2261-2269.

[19] Snell, J., Ridgeway, K., Liao, R., Roads, B. D., Mozer, M. C., and Zemel, R. S. (2015). Learning to generate images with perceptual similarity metrics. *arXiv preprint arXiv, 1511, 06409.*

[20] Yang, W., Wang, W., Huang, H., Wang, S., and Liu, J. (2021). Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing, 30*, 2072–2086.

DOI: https://doi.org/10.17509/ijost.v11i1.90990
p- ISSN 2528-1410 e- ISSN 2527-8045

[21] Surono, S., Rivaldi, M., Dewi, D. A., and Irsalinda, N. (2023). New approach to image segmentation: U-Net convolutional network for multiresolution CT image lung segmentation. *Emerging Science Journal, 7*(2), 498–506.

[22] Zhang, Y., Tian, Y., Kong, Y., Zhong, B., and Fu, Y.R. (2021). Residual Dense Network for Image Restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 43*(7), 2480-2495.

[23] Wang, L., Zhao, L., Zhong, T., and Wu, C. (2024). Low-light image enhancement using generative adversarial networks. *Scientific Reports, 14*, 18489.

[24] Yin, M., and Yang, J. (2025). ILR-Net: Low-light image enhancement network based on iterative learning mechanism and Retinex theory. *PLoS ONE, 20*(2), e0314541.

[25] Lim, C. C., Loh, Y. P., and Wong, L.-K. (2023). LAU-Net: A low light image enhancer with attention and resizing mechanisms. *Signal Processing: Image Communication, 115*, 116971.

[26] Jingchun, Z., Su, G. E., and Sunar, M. S. (2024). Low-light image enhancement: A comprehensive review on methods, datasets and evaluation metrics. *Journal of King Saud University - Computer and Information Sciences, 36*(10), 102234.

[27] Jiang, B., Wang, X., Yang, N., Liu, Y., Chen, X., and Wu, Q. (2025). Semantic-aware low-light image enhancement by learning from multiple color spaces. *Applied Sciences, 15*(10), 5556.

[28] Shen, X., Li, H., Li, Y., and Zhang, W. (2025). ColorBoost-LLIE: A multi-loss guided low-light image enhancement algorithm with decoupled color and luminance restoration. *Displays, 87*, 102979.

[29] Avi-Aharon, M., Arbelle, A., and Raviv, T. R. (2020). DeepHist: Differentiable joint and color histogram layers for image-to-image translation. *arXiv preprint arXiv, 2005, 03995.*

[30] Wang, Z., Simoncelli, E. P., and Bovik, A. C. (2003). Multiscale structural similarity for image quality assessment. *Proceedings of the IEEE Asilomar Conference on Signals, Systems and Computers, 2,* 1398–1402.

[31] Marr, D., and Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London B, 207*, 187–217.

[32] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing, 13*(4), 600–612.

[33] Zhang, R., Isola, P., Efros, A.A., Shechtman, E., and Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018,* 586-595.

[34] Mittal, A., Soundararajan, R., and Bovik, A. C. (2013). Making a completely blind image quality analyzer. *IEEE Signal Processing Letters, 20*(3), 209–212.

[35] Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv, 1409, 1556.*

DOI: https://doi.org/10.17509/ijost.v11i1.90990
p- ISSN 2528-1410 e- ISSN 2527-8045