

Article

RadiantVisions: Illuminating Low-Light Imagery with a Multi-Scale Branch Network

Yu Zhang ^{1,2}, Shan Jiang ^{1,2,*} and Xiangyun Tang ^{1,2} 

¹ Key Laboratory of Ethnic Language Intelligent Analysis and Security Governance of MOE, Minzu University of China, Beijing 100081, China; 22302042@muc.edu.cn (Y.Z.); xiangyunt@muc.edu.cn (X.T.)

² School of Information Engineering, Minzu University of China, Beijing 100081, China

* Correspondence: jshan@muc.edu.cn

Abstract: In the realms of the Internet of Things (IoT) and artificial intelligence (AI) security, ensuring the integrity and quality of visual data becomes paramount, especially under low-light conditions, where low-light image enhancement emerges as a crucial technology. However, the current methods for enhancing images under low-light conditions still face some challenging issues, including the inability to effectively handle uneven illumination distribution, suboptimal denoising performance, and insufficient correlation among a branch network. Addressing these issues, the Multi-Scale Branch Network is proposed. It utilizes multi-scale feature extraction to handle uneven illumination distribution, introduces denoising functions to mitigate noise issues arising from image enhancement, and establishes correlations between network branches to enhance information exchange. Additionally, our approach incorporates a vision transformer to enhance feature extraction and context understanding. The process begins with capturing raw RGB data, which are then optimized through sophisticated image signal processor (ISP) techniques, resulting in a refined visual output. This method significantly improves image brightness and reduces noise, achieving remarkable improvements in low-light image enhancement compared to similar methods. Using the LOL-V2-real dataset, we achieved improvements of 0.255 in PSNR and 0.23 in SSIM, with decreases of 0.003 in MAE and 0.009 in LPIPS, compared to the state-of-the-art methods. Rigorous experimentation confirmed the reliability of this approach in enhancing image quality under low-light conditions.



Citation: Zhang, Y.; Jiang, S.; Tang, X. RadiantVisions: Illuminating Low-Light Imagery with a Multi-Scale Branch Network. *Electronics* **2024**, *13*, 788. <https://doi.org/10.3390/electronics13040788>

Academic Editor: Maria Evelina Fantacci

Received: 13 January 2024

Revised: 9 February 2024

Accepted: 12 February 2024

Published: 17 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: low-light enhancement; multi-scale; noise reduction; image analysis

1. Introduction

With the proliferation of IoT devices, the integration of AI to ensure the security of these interconnected systems becomes imperative [1–4]. One significant aspect of this intersection lies in the domain of visual data capture and transmission. In the field of AI security, especially in the construction of image databases, low-light image enhancement is a common image processing technique. In contemporary interconnected environments, security surveillance, autonomous driving, and various other IoT technologies seamlessly permeate daily lives. Ai and Kwon [5] introduced a novel convolutional network designed to address security concerns in surveillance cameras within smart city environments, leveraging primary support from deep learning and demonstrating efficacy even under extremely low-light conditions. The emergence of these technologies has sparked an increasing demand for high-quality imagery, especially since these devices function under low-light conditions. In the construction or utilization of databases related to AI security, enhancing the quality of certain low-light images is crucial to augment the data. Whether facing low-light environments or challenging lighting scenarios, the robust capture and processing of visual data are essential for maintaining the integrity and reliability of AI-driven security systems. In this context, when delving into the challenges and advancements in AI security, especially concerning visual data from IoT devices, a crucial aspect to explore is the enhancement of

low-light images in the realm of security cameras. Our focus revolves around addressing the task of improving the quality of images captured under low-light conditions.

The realm of low-light image enhancement has witnessed significant progress within the domain of deep learning, with the widespread adoption of convolutional neural networks (CNNs) [6] and transformers [7]. CNNs, renowned for their ability to capture intricate features, are extensively applied to address the challenges presented under low-light conditions. The utilization of transformer models, originally designed for sequential data processing, has emerged as a promising avenue for capturing global contextual information, thereby enhancing the representation of low-light images. Building upon these advancements, various network architectures, including both single-branch and multi-branch structures, have been explored in the realm of low-light image enhancement. Single-branch architectures, such as CNNs, have gained prominence due to their adeptness at capturing intricate features, making them well-suited for addressing challenges posed by low-light conditions. In contrast, multi-branch architectures have also been investigated. However, a notable limitation is the lack of an inherent correlation between the branches, hindering a seamless information exchange. Our approach aims to overcome the inherent limitations of independent multi-branch structures, enhancing their effectiveness in capturing both local and global features for improved low-light image enhancement.

Multi-scale feature extraction has consistently been a focal point in the field of computer vision. In various image enhancement tasks, multi-scale feature extraction has proven to be highly beneficial. Researchers have endeavored to enhance models' sensitivity to different scales through traditional approaches such as pyramid structures and spatial pyramid pooling, as well as emerging techniques in deep learning, like multi-scale convolutions and pyramid convolutions. In the context of prior low-light image enhancement models, including RetiNexNet [8], MBLEN [9], 3D-LUT [10], and IAT [11], a common attribute is observed: the absence of multi-scale feature extraction in their architectures. Due to the inability to capture information at different scales, these models may encounter challenges in effectively enhancing dimly lit images, particularly when variations in illumination exist across different regions of the image. To alleviate this limitation, multi-scale feature extraction has been introduced into the model architecture, enabling a more comprehensive capture of both local details and global structures in an image. Analyzing an image at various scales allows the model to adapt more adeptly to diverse lighting conditions, thereby improving its performance and robustness. This strategic integration aims to ensure that the enhancement model excels particularly in handling images with non-uniform lighting, showcasing superior performance under such circumstances.

Image denoising [12] plays a crucial role in enhancing visual quality by mitigating the presence of noise artifacts, such as graininess and the loss of fine details prevalent in low-light photography. In the context of low-light image enhancement tasks, the issue of pronounced noise in post-enhancement becomes particularly evident. Recognizing this challenge, our approach employs specialized denoising functions to alleviate the noise issues arising from the enhancement of low-light images. This strategic integration of denoising functions stands as a crucial step in ensuring that the enhancement process not only addresses low-light conditions but also maintains high visual quality by mitigating unwanted noise artifacts.

In recent years, researchers have made significant progress in low-light image enhancement. However, under the challenging conditions of complex low-light scenarios, image enhancement methods still face some challenging issues. Restoring low-light images with uneven illumination distribution proves difficult, and low-light images often contain substantial noise. Additionally, the lack of inherent correlations between branch networks hampers effective information exchange. To address these issues, we propose a novel model named MSBN (Multi-Scale Branch Network) for low-light image enhancement. This model integrates multi-scale feature extraction, enabling a more comprehensive capture of both local details and global structures in images. Analyzing images at different scales allows the model to adapt to varying lighting conditions, thereby improving its performance and robustness. The model introduces denoising functions to alleviate noise issues and establishes

correlations between network branches to overcome the inherent limitations of independent multi-branch structures, enhancing effectiveness in capturing both local and global features. This comprehensive method facilitates the nuanced processing of low-light images, effectively balancing the enhancement of image quality with the preservation of natural details and the reduction of noise artifacts. The proposed general model can be applied to various safety domains, including night-time surveillance, autonomous driving, urban lighting, and other safety-related areas. Our contributions are manifold and can be summarized as follows:

- We integrate multi-scale feature extraction into MSBN. This integration enables the model to enhance images with non-uniform lighting conditions, preserving the original uneven illumination and retaining details across different scales after enhancement.
- We introduce a custom denoising loss function tailored specifically for low-light conditions. This feature effectively alleviates the noise issues introduced in low-light images after enhancement, ensuring the clarity of the images.
- Our model combines inter-branch correlations, employing weighted feature fusion to enhance the extraction and integration of prominent features. It strengthens the correlation between color, brightness, and the image itself, resulting in a more realistic effect in the enhanced images.

2. Related Work

2.1. Image Signal Processor

ISP is a pivotal component in computer vision and digital imaging, significantly contributing to image quality enhancement and overall performance. The research landscape surrounding ISP is rich and diverse, with several noteworthy contributions. Liang [13] introduced a novel deep neural network addressing challenges like low-light image generation and RAW-to-RGB conversion, which outperformed existing methods with an efficient design. Park [14] proposed an ISP-focused pre-processing method for optimizing brightness and contrast, particularly benefiting edge detection in applications such as autonomous vehicles and defense. Wang [15] tackled exposure issues with a method leveraging local color distributions, introducing an LCDE module and a dual-illumination learning mechanism, showcasing superior performance on a newly constructed dataset. These studies collectively underscore the pivotal role of ISPs in advancing image processing through innovative algorithmic techniques, architectural designs, and diverse applications.

2.2. Multi-Scale Feature Extraction

Multi-scale feature extraction has been a pivotal aspect of computer vision and image processing, addressing the necessity to capture information at varying levels of granularity. Diverse methods in the literature have approached this challenge through various strategies. In their work, Wang [16] introduced a novel multi-scale feature extraction and normalized attention neural network for image denoising, surpassing state-of-the-art methods by achieving higher PSNR, SSIM values, and overall visual quality in restored images. Liu [17] proposed a deep network for infrared and visible image fusion, incorporating a unique feature learning module and edge-guided attention mechanism. This method outperformed existing approaches across benchmarks and demonstrated robustness under challenging conditions. Qi [18] presented a novel underwater image enhancement network, leveraging semantic region-wise enhancement modules for multi-scale perception. This approach effectively improved color distortion and blurred details in underwater images. Collectively, these studies underscore the significance of multi-scale feature extraction in enhancing the robustness and discriminative power of computer vision systems.

2.3. Image Denoising

Image denoising, aimed at reducing noise while preserving details, has evolved significantly within image processing. Over the past years, a myriad of image priors have been proposed to restore clarity to noisy images, including sparsity, low rank, and self-similarity. Notable methods based on image priors, such as BM3D [19] and WNNM [20], have achieved

remarkable progress in the domain of image denoising. With the advent of deep learning, researchers have increasingly turned to leveraging deep neural networks for image denoising. For instance, DnCNN [21] employed a deep residual network and integrated batch normalization layers to accelerate the training process. In a different approach, CBDNet [22] addressed noise comprehensively throughout the imaging process, utilizing the U-net architecture along with a sub-network to estimate noise levels, thus enhancing denoising performance.

3. Methodology

3.1. Model Structure

Enhancing images under low-light conditions at the sensor level poses a multifaceted challenge in the field of digital image processing. The conventional ISP approach encompasses various subprocesses, such as white balance calibration, demosaicking algorithms, and denoising techniques. However, these traditional methods often amplify noise artifacts and reduce color fidelity, significantly hindering the attainment of optimal image quality.

During the image acquisition process with a camera, an RGB image is influenced by the ambient lighting condition L_i . It undergoes processing via the ISP and is transformed into an sRGB image I_i . This conversion typically results in a substantial loss of information related to the original lighting and color. Our method aims to reverse the ISP workflow [23] by extracting the original RGB values from the sRGB images and adjusting them through the introduction of new lighting conditions L_t [24]. The objective is to generate an sRGB image I_t that closely matches the one under the target lighting conditions. The network encoder f [25] is utilized to represent the inverse operation of ISP, and several individual decoders g_t are added to the encoder f . The function maps $f(I_i)$ to the target I_t under the following lighting conditions:

$$g_t(f(I_i)) = I_t \quad (1)$$

In this process, the network encoder f is utilized to map I_i to the corresponding original RGB data. Additionally, multiple independent decoders g_t are superimposed on the encoder f to generate the target image I_t . For $f(I_i) = I_i M + A$; multiplication for predicting mapping M and addition for mapping A are generated, achieving pixel-level adjustments [11]. The equation of our MSBN model was formulated as follows:

$$g_t(\cdot) = \left(\max \left(c \sum_j W_{ci,cj} (I_i M + A), \varepsilon \right) \right)^\gamma \quad (2)$$

where $W_{ci,cj}$ denotes a 3×3 joint color transformation matrix, amalgamating the influences of white balance and color transformation. It is applied to each pixel of the input image to achieve white balance and color correction. Nine queries are employed to control the parameters of $W_{ci,cj}$, with each query potentially corresponding to an element or a set of elements in the matrix. This allows for the fine-tuning of the color transformation matrix to achieve the desired effects.

The symbol γ denotes the gamma correction parameter, adjusting image brightness and contrast for a more natural appearance. Modifying γ influences the overall brightness of the output image. Additionally, ε is a small value preventing numerical instability, ensuring formula stability by averting division by zero or other numerical issues during computation.

Building upon the ISP theory outlined above, the MSBN model is proposed, emphasizing its multi-scale module, branch correlation module, and denoising loss module.

The MSBN model is illustrated in Figure 1; it undergoes multi-scale feature extraction as its initial step. By comprehensively utilizing feature information at different scales, our approach facilitates a comprehensive understanding of image content. Features at smaller scales excel at capturing intricate details and textures, while those at larger scales contribute to grasping global structures and contextual relationships. This integrated processing across multiple scales enhances the accuracy and robustness of the algorithm, effectively mitigating the impact of noise on image processing results.

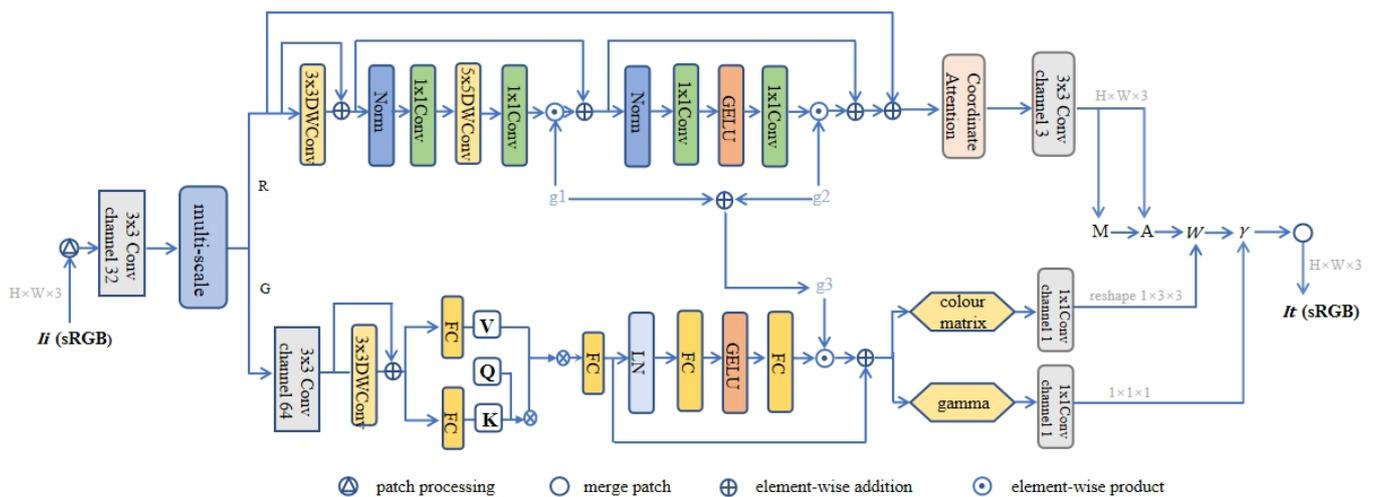


Figure 1. Diagram of the structure of our branch network MSBN. Different colors are used to distinguish various network processes.

After undergoing the steps outlined above, the network is split into two branches [26]: the R branch and the G branch. The R branch is designed to generate M and A components, inversely transforming the input sRGB image I_i to an RGB image. The G branch is designated for querying, decoding, and predicting W and γ , subsequently producing the sRGB image I_t .

The R branch begins by expanding the channels through a 3×3 convolution for initial processing. Subsequently, 1×1 convolution and 5×5 depthwise separable convolution are employed for further feature extraction and integration. The output undergoes optimization via an additional normalization and 1×1 convolution layer. Feature capturing is enhanced through GELU activation and a coordinate attention layer. Ultimately, the channels are reduced to three using a 3×3 convolution, and M and A components are generated using ReLU activation. Throughout this process, four skip connections [27] and the layer scale [28] method (multiplied by factors g_1/g_2) ensure rapid and stable convergence.

In the G branch, two convolutional layers serve as encoders to process and encode features. The encoded features are then forwarded to the prediction module. In this module, the component query (Q) is initialized to zero, and no additional multi-head self-attention mechanism is employed. Q is a learnable embedding associated with the keys (K) and values (V) generated from the encoded features. The position encoding of K and V is produced through deep convolution. After passing through a feed-forward network, which consists of two linear layers, the model incorporates two parameters with specific initialization. These parameters are utilized to output the color matrix and gamma values.

1. Multi-scale module

Integrating a multi-scale module into MSBN incorporates the Res2Net [29] structure and SE block [30], as illustrated in Figure 2. The Res2Net structure leverages grouped input feature maps in which each set of filters extracts features from a corresponding input group. This iterative process continues until all feature maps are processed. The resulting features are then concatenated and fused using 1×1 filters. As input features transform into output features, the equivalent receptive field expands, creating multiple feature scales through the combined effects of 3×3 filters. Simultaneously, the SE block enhances the model’s attention mechanism and adaptively re-calibrates channel-wise feature responses by explicitly modeling inter-dependencies among channels. This synergistic approach enhances the model’s capacity for sophisticated multi-scale feature extraction, contributing to its academic robustness.

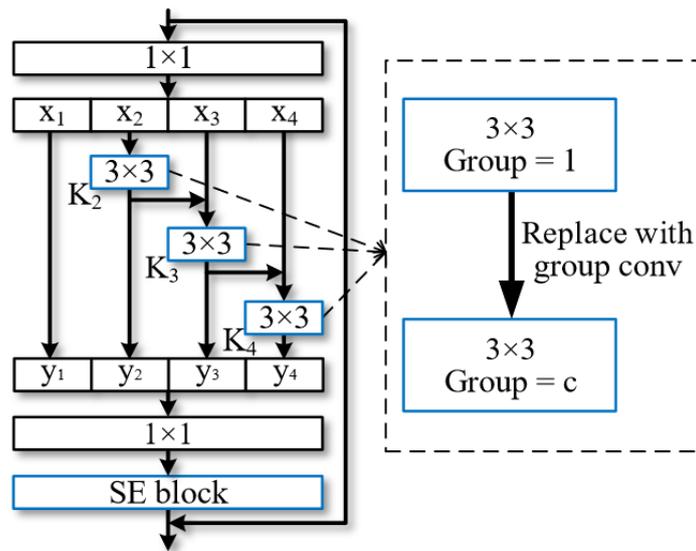


Figure 2. Network architecture with multi-scale feature extraction. Blue is used to differentiate from the 1×1 filters.

2. Branch correlation module

Introducing the branch correlation adjustment parameter g_3 in the low-light image enhancement network improves the performance and adaptability of the network. As shown in Equation (4), the value of this parameter is calculated through tensors C_1 and C_2 , which are correlated with g_1 and g_2 in branch R. g_1 and g_2 denote the neural network parameters involved in the computation of M and A in Equation (2). The incorporation of g_3 is introduced to enhance the precision of calculating W and γ in Equation (2).

$$g_3 = g_1 \cdot C_1 + g_2 \cdot C_2 \tag{3}$$

The effect of introducing g_3 is manifested in two main aspects. Firstly, it allows us to introduce a certain degree of correlation between branch R and branch G, aiding in coordinating the learning processes of the two branches. This coordination enhances the network’s ability to extract and understand image features under low-light conditions. Secondly, by introducing the learnable parameter g_3 , the network becomes more flexible, enabling it to adjust the relationship between branches based on specific input data and task requirements. This flexibility is crucial for improving the network’s generalization performance, ensuring excellent performance in various scenarios and lighting conditions.

3.2. Loss Function

In terms of loss function design, three main components are utilized for comprehensive optimization. The first part, smooth L1 loss, focuses on minimizing the pixel-level differences between the enhanced and high-quality images. The second part, perceptual loss, ensures that the enhanced images maintain similar feature representations with the original high-quality images in the high-dimensional feature space using a pre-trained VGG model [31]. The pre-trained VGG model possesses powerful feature extraction capabilities, facilitating the preservation of semantic information in images. Simultaneously, its training on large-scale datasets enables it to capture features relevant to human perceptual similarity, thereby better maintaining consistency in image perception. Through transfer learning, the model can adapt more quickly to tasks under low-light conditions.

To further enhance image quality, we introduced a denoising loss function $\mathcal{L}_{\text{denoising}}$ [32]. The primary objective of this loss function is to minimize the noise discrepancies between the generated image and the original image, allowing the model to better preserve key features of the image. By considering subtle differences between the denoised result and the

source, as well as target images, $\mathcal{L}_{\text{denoising}}$ effectively guides the training process, ensuring that the generated images are more realistic and clearer while maintaining the structural details of the original image. The formula is as follows:

$$Loss_{\text{denoising}} = \frac{1}{2} \left((D_1 - S_2)^2 + (D_2 - S_1)^2 \right) + \frac{1}{2} \left((D_1 - T_1)^2 + (D_2 - T_2)^2 \right) \quad (4)$$

This loss function is divided into two components. The first part is the residual loss, which quantifies the residuals between the generated image and the original image at different scales. The second part is the consistency loss, which measures the consistency between the generated image and the target image at different scales. D and T denote I_t in Equations (1) and (2). The terms D_1 and D_2 denote the downsampled results of the enhanced image. Similarly, S_1 and S_2 denote the downsampled results of the normal image. T_1 and T_2 denote the twice-downsampled results of the enhanced image.

The entire loss function incorporates a weighted combination of three types of losses, as defined in Equation (5), resulting in the total loss ($Loss_{\text{total}}$) as follows:

$$Loss_{\text{total}} = Loss_1 + \lambda_1 \cdot Loss_2 + \lambda_2 \cdot Loss_{\text{denoising}} \quad (5)$$

where smooth L1 loss is referred to as $Loss_1$, indicating the utilization of high-dimensional features extracted via VGG as $Loss_2$.

4. Experimental Results

4.1. Experimental Settings

We used the PyTorch framework on an NVIDIA 4090 GPU to train our model for 250 epochs on the LOL-V2-real dataset [33]. LOL-V2-real contains 689 low-/normal-light image pairs for training and 100 pairs for testing. This dataset comprised images from both indoor and outdoor scenes, each with a resolution of 600×400 . To optimize MSBN, the Adam optimizer was chosen with an initial learning rate set to 2×10^{-4} , accompanied by a weight decay of 5×10^{-4} . To enhance the model's generalization and reduce the risk of overfitting, a data augmentation strategy was employed. In addition to the original images, the model was fed with horizontally and vertically flipped versions of the original images, enriching the dataset. This strategy significantly boosted the model's performance during both the validation and testing phases. Through experiments, remarkable improvements were observed in image quality and restoration accuracy for the low-light enhancement task. As depicted in Figure 3, the visual results demonstrate the effectiveness of our method in enhancing image details and overall image quality under low-light conditions.

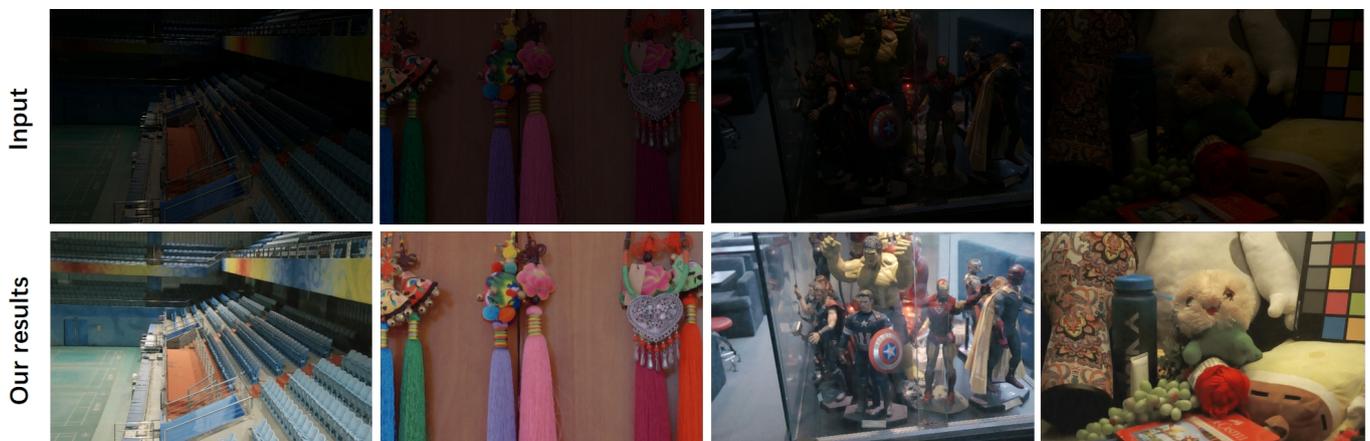


Figure 3. The results obtained from our method for low-light enhancement are shown in the figures.

In this experiment, SSIM (structural similarity index measure), PSNR (peak signal-to-noise ratio), MAE (mean absolute error), and LPIPS (learned perceptual image patch

similarity) were employed as metrics to assess the image quality of the enhanced images. SSIM measures the structural similarity between two images, considering factors such as brightness, contrast, and structure, and it produces a score in the $[-1, 1]$ range, where 1 indicates perfect similarity. On the other hand, PSNR measures the quality of an image by evaluating the peak signal-to-noise ratio, with higher values indicating better image fidelity. Additionally, MAE is used to quantify the average absolute differences between corresponding pixel values in the original and enhanced images. A lower MAE value signifies better image accuracy. Furthermore, LPIPS is employed to evaluate perceptual similarity, measuring the perceptual distance between images, considering human visual perception, and it is often used to assess the perceived quality of images. Together, these metrics provide a comprehensive evaluation of the enhanced image quality, considering both structural and perceptual aspects.

4.2. Visual and Perceptual Comparisons

We compared the enhancement effects under low-light conditions of the MSBN method with several other methods [8–11,34–37]. As depicted in Figure 4, it became apparent that Zero-DCE [37] faces significant challenges when operating in low-light environments, as evidenced by its performance. The image processing results frequently yielded a perceptually darker visual representation of the images. In contrast, RetiNexNet [8] exhibits a notable issue primarily related to color restoration. It shows pronounced color discrepancies that may distort the authentic colors of objects in the image, thereby compromising image quality and fidelity. The IAT [11] exhibits a performance that approximates the ground truth to a greater extent than its predecessors; however, it is plagued by pronounced noise artifacts. In contrast, our method not only surpasses the IAT in terms of reduced noise but also yields clearer and more visually distinct images. Through a comparative analysis of the output images from each model, our method outperforms other models in processing low-light images. It not only significantly improves the brightness and clarity of the pictures but also makes the enhanced images more akin to real scenes.

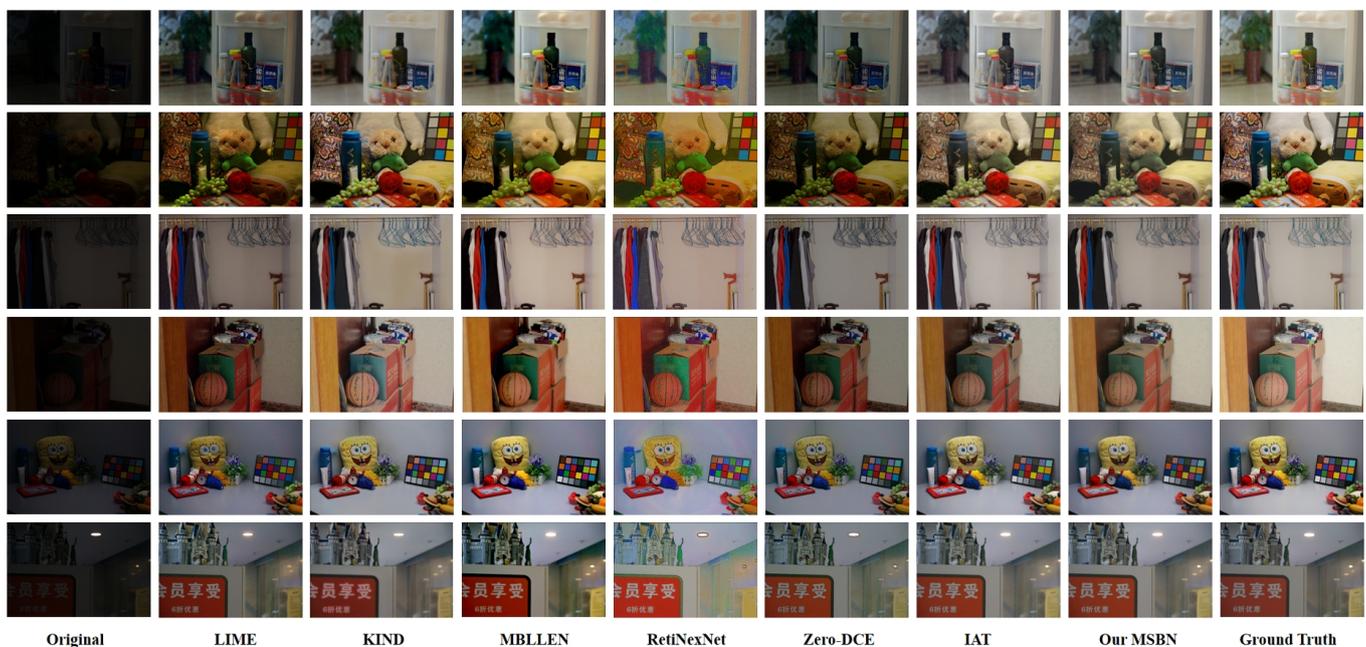


Figure 4. A performance comparison between our MSBN and other methods, including LIME (Low-Light Image Enhancement), KIND (Kindling the Darkness), MBLLEN (Multi-Branch Low-Light Enhancement Network), RetiNexNet (Deep Retinex Decomposition for Low-Light Enhancement), Zero-DCE (Zero-Reference Deep Curve Estimation), and IAT (Illumination Adaptive Transformer).

As illustrated in Figure 5, to effectively address the issue of uneven illumination distribution in low-light image enhancement, MSBN introduces multi-scale feature extraction and integrates denoising functions to reduce noise. In the ground truth image, the red box, when contrasted with the area above, revealed an uneven illumination distribution. In contrast to our MSBN, other methods did not exhibit this uneven illumination distribution. Simultaneously, the green box illustrates that, compared to other methods, MSBN exhibits less image noise. These results demonstrate the significant efficacy of MSBN in tackling these issues.

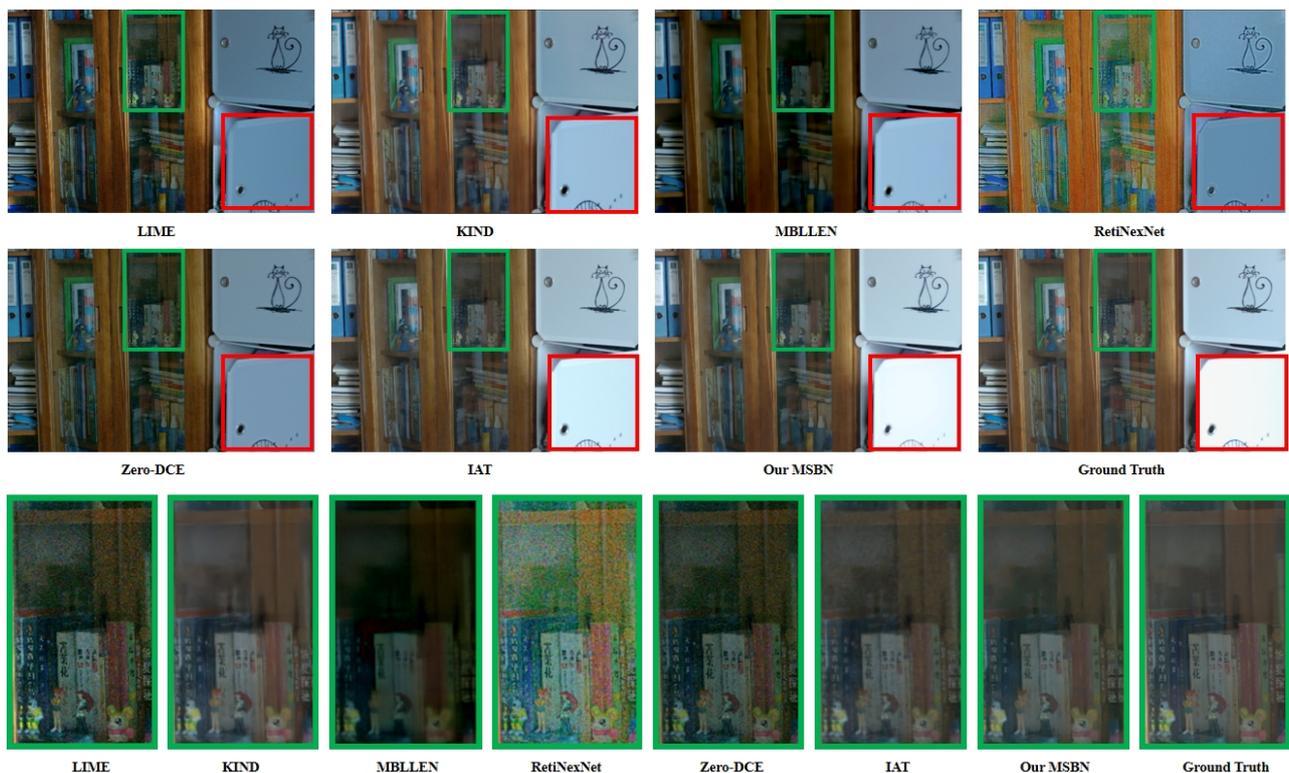


Figure 5. A comparison of MSBN and other methods in detail. Red is used to compare uneven illumination distribution, and green is used to compare image noise.

Table 1 presents the performance of various methods in the low-light image enhancement task on the LOL-V2-real dataset. Four metrics, MAE, LPIPS, PSNR, and SSIM, were employed to assess the efficacy of each method. Notably, conventional methods like LIME and Zero-DCE exhibited lower PSNR and SSIM values while having relatively higher MAE and LPIPS values, indicating their limited effectiveness in low-light image enhancement. In contrast, advanced models such as UFormer, KIND, IAT, and MSBN demonstrated significant improvements, with MSBN achieving the highest PSNR and SSIM scores, as well as the lowest MAE and LPIPS scores. It outperformed the second-best IAT by 0.255 and 0.23 in PSNR and SSIM, and it was 0.003 and 0.009 lower in MAE and LPIPS, respectively, showcasing its superior capability in enhancing low-light images using the LOL-V2-real dataset. In terms of parameters, MSBN ranked third with only 0.16M, highlighting its effectiveness in achieving both quantitative performance and model efficiency.

Table 1. Quantitative evaluation of various methods on LOL-V2-real dataset. ↓ indicates that smaller values are better, and ↑ indicates that larger values are better.

Methods	MAE ↓	LPIPS ↓	PSNR ↑	SSIM ↑	Params (M) ↓
LIME	0.155	0.397	15.481	0.480	-
Zero-DCE	0.149	0.412	14.893	0.517	0.08
RetiNexNet	0.115	0.196	18.554	0.733	0.84
MBLEN	0.122	0.271	18.115	0.717	20.47
3D-LUT	0.130	0.276	17.653	0.719	0.6
UFormer	0.114	0.144	18.764	0.774	5.29
KIND	0.098	0.121	19.621	0.770	8.16
IAT	0.065	0.097	23.487	0.828	0.09
MSBN	0.062	0.088	23.742	0.851	0.16

4.3. Hyperparameter Details

Table 2 provides detailed information on the hyperparameter search and optimal configuration during the low-light image enhancement experiments on the LOL-V2-real dataset. We meticulously examined various hyperparameters, including batch size, display iteration, learning rate, number of epochs, and weight decay. The search range for each hyperparameter was specified, emphasizing the optimal configuration that yielded the best performance. We paid special attention to understanding the relationship between experimental results and different hyperparameter values throughout this tuning process to determine the optimal configuration. Ultimately, we found that setting the batch size to 8, the display iteration to 10, the learning rate to 2×10^{-4} , the number of epochs to 250, and the weight decay to 5×10^{-4} resulted in the best performance of our model using the LOL-V2-real dataset. The selection of this optimal configuration underwent thorough experimental validation, ensuring the robustness and effectiveness of our proposed low-light image enhancement method in various aspects.

Table 2. Hyperparameter search and optimal configuration.

Hyperparameter	Search Space	Optimal Configuration
Batch size	4–32	8
Display iteration	5–20	10
Learning rate	2×10^{-3} – 2×10^{-5}	2×10^{-4}
Number of epochs	50–300	250
Weight decay	1×10^{-3} – 1×10^{-5}	5×10^{-4}

4.4. Ablation Experiment

According to the results of the ablation experiment concerning the low-light image enhancement framework in Table 3, a noticeable improvement in performance was observed upon the incorporation of individual modules. Upon the simultaneous integration of the three modules, there were decreases of 0.026 and 0.024 in MAE and LPIPS, respectively, while PSNR and SSIM showed increases of 1.791 and 0.036, respectively.

These results indicate that these three modules showed significant improvements in addressing uneven illumination distribution, suboptimal denoising performance, and insufficient correlation among branch network issues in low-light image enhancement.

Table 3. The results of the ablation experiments were obtained by adding different modules, with the baseline network represented as O, the multi-scale module as A, the denoising loss module as B, and the branch correlation module as C. ↓ indicates that smaller values are better, and ↑ indicates that larger values are better.

	MAE ↓	LPIPS ↓	PSNR ↑	SSIM ↑
O	0.088	0.112	21.951	0.815
O + A	0.075	0.103	23.173	0.833
O + B	0.081	0.117	22.941	0.828
O + C	0.094	0.105	22.031	0.819
O + A + B + C	0.062	0.088	23.742	0.851

5. Conclusions

We have proposed the MSBN model specifically designed for low-light image enhancement tasks. Integrating multi-scale feature extraction, branch correlation, and advanced de-noising techniques, MSBN effectively addresses uneven illumination distribution, suboptimal denoising performance, and insufficient correlation among branch network issues. With the incorporation of a vision transformer for enhanced feature extraction, the optimization of raw RGB data through image signal processor techniques results in refined visual output. The application of a composite loss function enhances robustness, showcasing significant advancements in luminance and noise reduction compared to traditional methods. Our method shows notable improvements on the LOL-V2-real dataset, demonstrating its effectiveness in addressing low-light challenges. The proposed general approach is applicable to and appropriate for many aspects of AI security, such as enhancing the performance of intrusion detection systems and improving the accuracy of facial recognition under low-light conditions.

In subsequent stages, we will explore avenues to refine the MSBN model in order to comprehensively tackle a broader range of low-light scenarios. Efforts will be directed toward leveraging transfer learning from related domains to boost the model's generalization capabilities. These endeavors are poised to contribute further to the model's efficacy, ensuring its applicability in diverse real-world low-light imaging scenarios.

Author Contributions: Conceptualization, Y.Z. and S.J.; methodology, Y.Z. and S.J.; software, Y.Z.; validation, S.J. and X.T.; investigation, Y.Z.; writing—original draft preparation, Y.Z.; writing—review and editing, Y.Z., S.J. and X.T.; funding acquisition, S.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (No. 52105167 and No. 62302539).

Data Availability Statement: The data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Hu, C.; Zhang, C.; Lei, D.; Wu, T.; Liu, X.; Zhu, L. Achieving Privacy-Preserving and Verifiable Support Vector Machine Training in the Cloud. *IEEE Trans. Inf. Forensics Secur.* **2023**, *18*, 3476–4291. [[CrossRef](#)]
- Zhang, C.; Hu, C.; Wu, T.; Zhu, L.; Liu, X. Achieving Efficient and Privacy-Preserving Neural Network Training and Prediction in Cloud Environments. *IEEE Trans. Dependable Secur. Comput.* **2023**, *20*, 4245–4257. [[CrossRef](#)]
- Zhang, C.; Luo, X.; Liang, J.; Liu, X.; Zhu, L.; Guo, S. POTA: Privacy-Preserving Online Multi-Task Assignment with Path Planning. *IEEE Trans. Mob. Comput.* **2023**, *in press*. [[CrossRef](#)]
- Zhang, C.; Zhao, M.; Liang, J.; Fan, Q.; Zhu, L.; Guo, S. NANO: Cryptographic Enforcement of Readability and Editability Governance in Blockchain Database. *IEEE Trans. Dependable Secur. Comput.* **2023**, *in press*. [[CrossRef](#)]
- Ai, S.; Kwon, J. Extreme low-light image enhancement for surveillance cameras using attention U-Net. *Sensors* **2020**, *20*, 495. [[CrossRef](#)] [[PubMed](#)]

6. Panwar, M.; Gaur, S.B. Inception-based CNN for low-light image enhancement. In *Computational Vision and Bio-Inspired Computing, Proceedings of the 5th International Conference on Computational Vision and Bio Inspired Computing (ICCVBIC 2021), Coimbatore, India, 25–26 November 2021*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 533–545.
7. Wang, T.; Zhang, K.; Shen, T.; Luo, W.; Stenger, B.; Lu, T. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023*; Volume 37, pp. 2654–2662.
8. Wei, C.; Wang, W.; Yang, W.; Liu, J. Deep retinex decomposition for low-light enhancement. *arXiv* **2018**, arXiv:1808.04560.
9. Lv, F.; Lu, F.; Wu, J.; Lim, C. MBLLEN: Low-Light Image/Video Enhancement Using CNNs. In *Proceedings of the British Machine Vision Conference (BMVC 2018), Newcastle, UK, 3–6 September 2018*; Volume 220, p. 4.
10. Zeng, H.; Cai, J.; Li, L.; Cao, Z.; Zhang, L. Learning Image-adaptive 3D Lookup Tables for High Performance Photo Enhancement in Real-time. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 2058–2073. [[CrossRef](#)] [[PubMed](#)]
11. Cui, Z.; Li, K.; Gu, L.; Su, S.; Gao, P.; Jiang, Z.; Qiao, Y.; Harada, T. You Only Need 90K Parameters to Adapt Light: A Light Weight Transformer for Image Enhancement and Exposure Correction. In *Proceedings of the 33rd British Machine Vision Conference (BMVC 2022), London, UK, 21–24 November 2022*; p. 238.
12. Tian, C.; Fei, L.; Zheng, W.; Xu, Y.; Zuo, W.; Lin, C.W. Deep learning on image denoising: An overview. *Neural Netw.* **2020**, *131*, 251–275. [[CrossRef](#)] [[PubMed](#)]
13. Liang, L.; Zharkov, I.; Amjadi, F.; Joze, H.R.V.; Pradeep, V. Guidance network with staged learning for image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021*; pp. 836–845.
14. Park, K.; Chae, M.; Cho, J.H. Image pre-processing method of machine learning for edge detection with image signal processor enhancement. *Micromachines* **2021**, *12*, 73. [[CrossRef](#)] [[PubMed](#)]
15. Wang, H.; Xu, K.; Lau, R.W. Local color distributions prior for image enhancement. In *Proceedings of the European Conference on Computer Vision (ECCV 2022), Tel Aviv, Israel, 23–27 October 2022*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 343–359.
16. Wang, Y.; Song, X.; Gong, G.; Li, N. A multi-scale feature extraction-based normalized attention neural network for image denoising. *Electronics* **2021**, *10*, 319. [[CrossRef](#)]
17. Liu, J.; Fan, X.; Jiang, J.; Liu, R.; Luo, Z. Learning a deep multi-scale feature ensemble and an edge-attention guidance for image fusion. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *32*, 105–119. [[CrossRef](#)]
18. Qi, Q.; Li, K.; Zheng, H.; Gao, X.; Hou, G.; Sun, K. SGUIE-Net: Semantic attention guided underwater image enhancement with multi-scale perception. *IEEE Trans. Image Process.* **2022**, *31*, 6816–6830. [[CrossRef](#)] [[PubMed](#)]
19. Yahya, A.A.; Tan, J.; Su, B.; Hu, M.; Wang, Y.; Liu, K.; Hadi, A.N. BM3D image denoising algorithm based on an adaptive filtering. *Multimed. Tools Appl.* **2020**, *79*, 20391–20427. [[CrossRef](#)]
20. Ou, Y.; He, B.; Luo, J.; Li, B. Improving the denoising of WNNM-based imagery: three different strategies. *Remote Sens. Lett.* **2021**, *12*, 307–314. [[CrossRef](#)]
21. Murali, V.; Sudeep, P. Image denoising using DnCNN: An exploration study. In *Advances in Communication Systems and Networks*; Springer: Singapore, 2020; pp. 847–859.
22. Bled, C.; Pitie, F. Assessing Advances in Real Noise Image Denoisers. In *Proceedings of the 19th ACM SIGGRAPH European Conference on Visual Media Production, Austin, TX, USA, 2–4 October 2022*; pp. 1–9.
23. Brooks, T.; Mildenhall, B.; Xue, T.; Chen, J.; Sharlet, D.; Barron, J.T. Unprocessing Images for Learned Raw Denoising. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2019), Long Beach, CA, USA, 15–20 June 2019*.
24. Lore, K.G.; Akitayo, A.; Sarkar, S. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognit.* **2017**, *61*, 650–662. [[CrossRef](#)]
25. Afifi, M.; Brown, M.S. Deep white-balance editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020*; pp. 1397–1406.
26. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-end object detection with transformers. In *Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 213–229.
27. Wu, D.; Wang, Y.; Xia, S.T.; Bailey, J.; Ma, X. Skip connections matter: On the transferability of adversarial examples generated with resnets. *arXiv* **2020**, arXiv:2002.05990 .
28. Touvron, H.; Cord, M.; Sablayrolles, A.; Synnaeve, G.; Jégou, H. Going deeper with image transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision, Nashville, TN, USA, 19–25 June 2021*; pp. 32–42.
29. Gao, S.H.; Cheng, M.M.; Zhao, K.; Zhang, X.Y.; Yang, M.H.; Torr, P. Res2Net: A New Multi-scale Backbone Architecture. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 652–662. [[CrossRef](#)] [[PubMed](#)]
30. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018*; pp. 7132–7141.
31. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In *Proceedings of the Computer Vision (ECCV 2016): 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016*; *Proceedings Part II 14*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 694–711.
32. Mansour, Y.; Heckel, R. Zero-Shot Noise2Noise: Efficient Image Denoising without any Data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023*; pp. 14018–14027.

33. Yang, W.; Wang, S.; Fang, Y.; Wang, Y.; Liu, J. From Fidelity to Perceptual Quality: A Semi-Supervised Approach for Low-Light Image Enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
34. Zhang, Y.; Zhang, J.; Guo, X. Kindling the darkness: A practical low-light image enhancer. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 1632–1640.
35. Guo, X.; Li, Y.; Ling, H. LIME: Low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.* **2016**, *26*, 982–993. [[CrossRef](#)] [[PubMed](#)]
36. Wang, Z.; Cun, X.; Bao, J.; Zhou, W.; Liu, J.; Li, H. Uformer: A general u-shaped transformer for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022; pp. 17683–17693.
37. Guo, C.; Li, C.; Guo, J.; Loy, C.C.; Hou, J.; Kwong, S.; Cong, R. Zero-reference deep curve estimation for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1780–1789.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.