

GM-MoE: Low-Light Enhancement with Gated-Mechanism Mixture-of-Experts

Minwen Liao^{1*} Haobo Dong^{2*} Xinyi Wang^{3*} Kurban Ubul^{1†}
 Yihua Shao^{4†} Ziyang Yan^{5†}

¹ Xinjiang University ² Harbin University of Commerce

³ Changchun University of Science and Technology

⁴ University of Science and Technology Beijing ⁵ University of Trento

Abstract

Low-light enhancement has wide applications in autonomous driving, 3D reconstruction, remote sensing, surveillance, and so on, which can significantly improve information utilization. However, most existing methods lack generalization and are limited to specific tasks such as image recovery. To address these issues, we propose **Gated-Mechanism Mixture-of-Experts (GM-MoE)**, the first framework to introduce a mixture-of-experts network for low-light image enhancement. GM-MoE comprises a dynamic gated weight conditioning network and three sub-expert networks, each specializing in a distinct enhancement task. Combining a self-designed gated mechanism that dynamically adjusts the weights of the sub-expert networks for different data domains. Additionally, we integrate local and global feature fusion within sub-expert networks to enhance image quality by capturing multi-scale features. Experimental results demonstrate that the GM-MoE achieves superior generalization compared to over 20 existing approaches, reaching state-of-the-art performance on PSNR on 5 benchmarks and SSIM on 4 benchmarks, respectively. Code is available at: <https://github.com/Sameenok/gm-moe-lowlight-enhancement.git>

1. Introduction

Low-light image enhancement (LLIE) is a crucial research area with diverse applications, including autonomous driving [54], low-light scene reconstruction [35, 55], remote sensing [35, 42, 59], and image/video analysis [39–41], since it enhances visibility and preserves

*These three authors contributed equally to this work and are considered co-first authors.

†Corresponding authors.

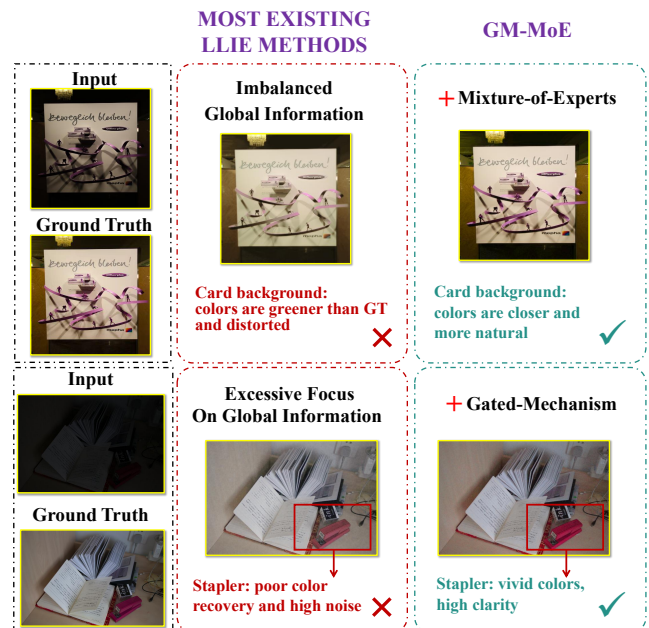


Figure 1. Given a low-light image, our GE-MoE achieves better performance (for both object and whole scene) compared with LightenDiffusion [18]

fine details, enabling more reliable scene understanding in challenging lighting conditions. Although recent advancements [22, 25] have improved LLIE performance, most existing methods focus on addressing specific challenges, such as noise suppression or detail restoration, rather than providing a comprehensive solution for diverse low-light scenarios. First, many existing algorithms have the problem of unbalanced global information due to local enhancement. Traditional algorithms such as histogram equalization enhance the image through a single strategy, which often over-enhances local areas, resulting in loss of image details or

overexposure [9]. Convolutional Neural Networks(CNN)-based methods use a network with multiple layers of convolutions [3, 32], which makes it difficult to learn the global distribution of illumination and restore it [14].

The transformer will ignore local color continuity due to an excessive focus on global information [61], which leads to the problem of color distortion. At the same time, there is a problem of insufficient cross-domain generalization ability. Existing methods, such as SurroundNet [67], are usually trained on specific datasets, and the model design lacks consideration of photos from different data domains. This results in a sharp decline in performance under unknown lighting conditions, making it difficult to achieve robust image enhancement and poor generalization ability. Meanwhile, because the problems of noise, color distortion, and blurred details in low-light images are coupled with each other, it is difficult for a single model to be optimized collaboratively. For example, suppressing noise may result in a sacrifice of details, and increasing the brightness of low-light areas may amplify color distortion. Therefore, it is a difficult problem to solve and balance the effect of image recovery. These problems limit the application of LLIE technology in complex scenes, and there is an urgent need for a unified framework that can not only enhance multiple tasks but also dynamically adapt to different lighting scenes.

To address these issues, we propose an innovative **Gated-Mechanism Mixture-of-Experts (GM-MoE)** system for low-light image enhancement. The method is based on an improved U-Net [38] architecture, incorporating a gated mechanism expert network with dynamic weight adjustment to adapt to photo inputs from different data domains. GM-MoE consists of three sub-expert networks, each of which is used to solve different image enhancement tasks, namely color correction, detail recovery, and problems. The gating mechanism assigns appropriate weights to each sub-expert based on the different lighting and scene conditions of the image, achieving a balance between the image enhancement problems and thus achieving the best image enhancement under different lighting conditions and scenes. As shown in Fig. 2, our framework achieves a higher PSNR and SSIM in multiple data sets compared to other state-of-the-art methods discussed in the literature.

In summary, our contributions are as follows:

- We are the first to apply GM-MoE to low-light image enhancement, proposing a method that combines a dynamic gating weight adjustment network with a multi-expert module to achieve effective generalization across different data domains.
- We propose a dynamic gating mechanism that adaptively adjusts the MoE weights according to varying lighting conditions, thereby optimizing image enhancement.
- Our model achieves superior performance across multiple datasets and downstream tasks. Extensive experiments

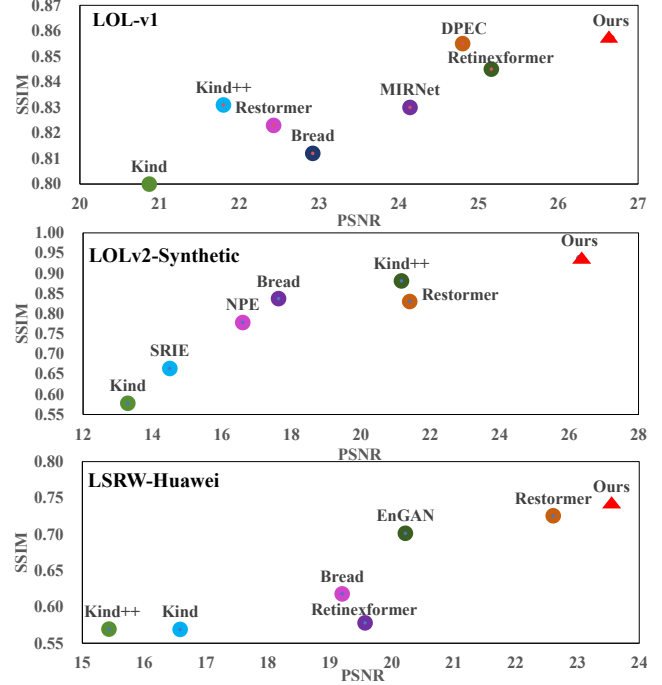


Figure 2. The comparison results among GM-MoE and the SOTA low-light image enhancement methods on the LOL-v1 , LOLv2-Synthetic and LSRW-Huawei benchmarks. GM-MoE outperforms all of compared approaches on both PSNR and SSIM metrics.

demonstrate that GM-MoE surpasses SOTA methods on different metrics across five datasets while maintaining strong generalization ability.

2. Related Work

2.1. Low-light Image Enhancement

Intensity Transformation Techniques. Traditional low-light enhancement methods improve low-light images by directly processing pixel intensity values, including histogram equalization (HE) [9, 36] and Gamma Correction (GC) [17, 31, 34]. HE enhances the contrast by redistributing the intensity histogram of the image, but it tends to over-enhance and, therefore, often amplifies noise. On the other hand, GC adjusts brightness using a nonlinear transformation, but it does not adapt well to complex lighting conditions, which leads to unnatural visual effects. In addition, adaptive contrast enhancement methods [15, 68] modify contrast based on local pixel statistics to enhance details, but they may inadvertently introduce noise. These methods often fail to take into account the full complexity of lighting, leading to perceptual differences compared to images captured under typical lighting conditions.

Perception-Based Models. To compensate for these deficiencies, some methods simulate the human visual mech-

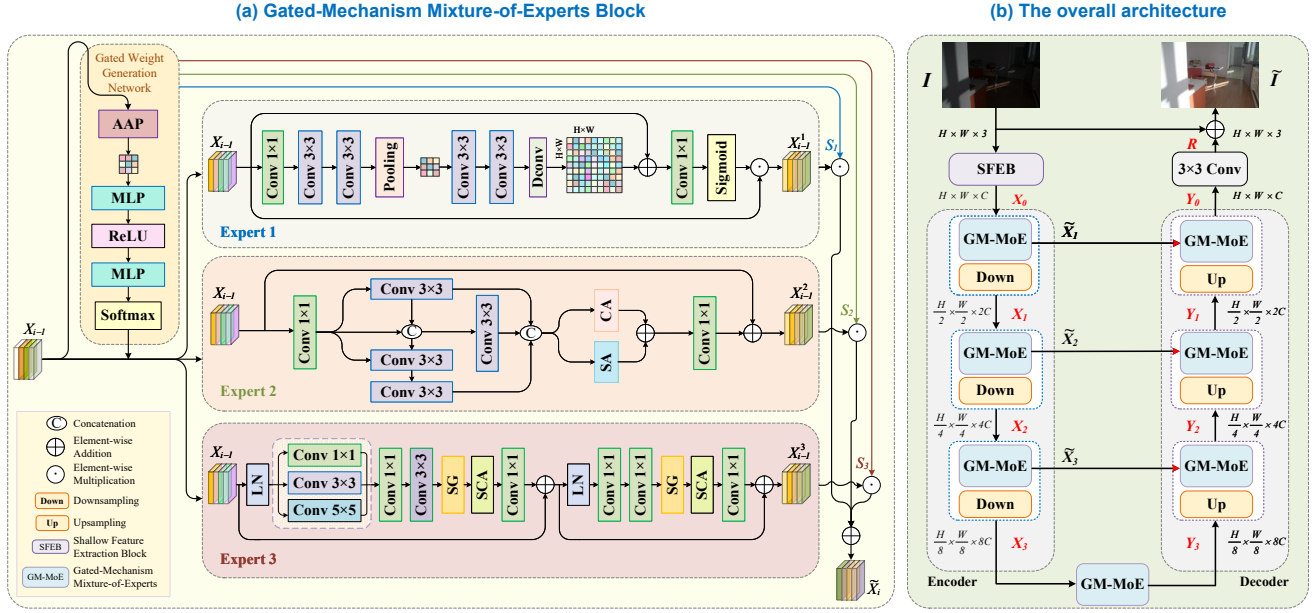


Figure 3. **Overview of the proposed GM-MoE.** (a) The GM-MoE module comprises a gated weight generation network and three specialized sub-expert networks. (b) The overall network adopts a U-Net-like encoder-decoder architecture. Given an input image, it first undergoes processing through the Shallow Feature Extraction Block (SFEB). Then, the GM-MoE module facilitates multi-scale feature fusion via multiple downsampling and upsampling operations, ultimately generating the enhanced output image.

anism, such as the Retinex theory, which decomposes an image into reflection and illumination components. Multi-scale Retinex (MSR) [21, 30] enhances contrast at different scales but may lead to color differences due to inaccurate light estimation. The dark channel prior (DCP) [14, 51], originally used for defogging, was later adapted for low-light enhancement but is prone to color oversaturation in complex scenes. Recently, the global brightness ranking regularization method proposed by Li et al., [26] has improved visual effects, though it can still introduce issues during enhancement.

Deep Learning-Based Approaches. In recent years, deep learning has promoted the application of convolutional neural networks in low-light image enhancement [10, 12, 27, 29, 56, 61, 63, 66, 67]. For example, Chen et al., [3] proposed the SID model to directly convert low-light images to normal-light images; Guo et al., [36] used a lightweight network to implement pixel-level curve estimation (ZeroDCE); Jiang et al., [19] employed GANs for unsupervised learning in EnlightenGAN. In addition, Wei et al. proposed Retinex-Net [50], Zhang et al., introduced KinD [64], and Liu et al., improved the RUAS model [27], all of which significantly enhanced image restoration performance. In the field of Transformers, Liang et al., [27] proposed SwinIR, and Zamir et al., [61] developed Restormer, both of which achieve image restoration by capturing global features. Meanwhile, Wang et al., [67] introduced LLFlow, and Xu et al., [53] proposed the SNR-Aware method, uti-

lizing normalizing flow and signal-to-noise ratio optimization, respectively, to enhance image details. Although these methods have made significant progress in low-light image enhancement, most of them adopt a single neural network structure. When dealing with different lighting conditions and complex scenes, their generalization ability remains limited, making it difficult to simultaneously perform multiple tasks such as noise suppression, detail restoration, and color correction across different datasets.

2.2. Multi-Expert Systems

MoE was originally proposed by Jacobs et al., [16] and aims to build a system consisting of multiple independent networks (experts), each responsible for processing a specific subset of data. The approach emphasizes that combining the expertise of different models can significantly improve the overall performance when dealing with complex tasks. In recent years, MoE techniques have demonstrated excellence in several domains, including image recognition [25], machine translation [24], scene parsing [28], speech recognition [5], and recommender systems [1].

In the field of low-light image enhancement, there is a lack of MoE applications. The expert network module possesses the capability to simultaneously address multiple low-light image enhancement challenges, a characteristic that grants it unique advantages in the field of low-light image enhancement. Moreover, existing methods exhibit

significant limitations: most low-light enhancement models can only achieve satisfactory restoration results for a single problem and lack the ability to adaptively adjust processing strategies based on the illumination conditions of the input image.

To address the limitations of existing methods and leverage the unique strengths of expert networks, this paper proposes a novel Multi-Expert Low-Light Enhancement Network. As illustrated in Fig. 3, we design a dynamically weighted multi-expert network based on the U-Net architecture, where three specialized sub-expert modules are optimized for color distortion, detail loss, and low-contrast issues, respectively. To overcome the insufficient generalization ability of existing models, we innovatively introduce a Dynamic Weight Adjustment Network, which automatically adjusts the weight allocation of the three sub-networks based on input image features, thereby achieving adaptive enhancement for images from different data domains.

3. Method

In the model we designed, the overall architecture is as follows: The input dark image $I \in \mathbb{R}^{H \times W \times 3}$ is first processed by SFEB to obtain low-level features $X_0 \in \mathbb{R}^{H \times W \times C}$, where H and W denote the spatial dimensions of the image and C denotes the number of channels. These features are then input into a module similar to the U-Net [38] network architecture, which contains encoder layers to further extract deeper features $F_d \in \mathbb{R}^{H \times W \times 2C}$. The GM-MOE module is introduced in the encoder and decoder of each layer. During encoding, the encoder of each layer compresses the image features by gradually reducing the spatial dimension and increasing the channel capacity, while the decoder gradually restores the image resolution by upsampling the low-resolution feature map $F_l \in \mathbb{R}^{\frac{H}{8} \times \frac{W}{8} \times 8C}$ to progressively recover the image resolution. To optimize feature recovery, pixel-shuffle techniques are introduced to improve the effects of upsampling. To aid low-light image feature recovery, the initial features are preserved and fused between the encoder and decoder via skip connections. Each layer of the GM-MOE module is responsible for fusing the lower-level features of the encoder with the higher-level features of the decoder, thereby enriching the structure and texture details of the image. In the final stage, the deep features F_d further enhances the detailed features at the spatial resolution, and the residual image generated by the convolution operation R is added to the input image I to obtain the final enhanced image $\hat{I} = I + R$.

3.1. Gated-Mechanism Mixture-of-Experts

GM-MOE block consists of a Gated Weight Generation Network and an expert network module, where the expert network includes a color restoration submodule, a de-

tail enhancement submodule, and an advanced feature enhancement network module. The following sections describe these modules in turn.

In order to achieve adaptive feature extraction in different data domains, we propose a GM-MoE network. First, the input image is passed through adaptive average pooling to convert the image features into a feature vector. Then, this feature vector passes through a fully connected layer with an activation function, and then through another fully connected layer to project onto three expert networks to generate the weights s_1, s_2, s_3 . These weights enable the network to dynamically adjust its parameters based on photos from different data domains (i.e., different scenes and different lighting characteristics), ensuring that the sum of the weights is 1.

$$S = [s_1, s_2, s_3], s_1 + s_2 + s_3 = 1. \quad (1)$$

Each expert network Net_i processes the input feature X_{i-1} and generates the corresponding output feature X_{i-1}^i , where $i \in \{1, 2, 3\}$:

$$X_{i-1}^i = Net_i(X_{i-1}). \quad (2)$$

The final output feature \tilde{X}_i is obtained by summing the outputs of all expert networks weighted by their weights:

$$\tilde{X}_i = s_1 X_{i-1}^1 + s_2 X_{i-1}^2 + s_3 X_{i-1}^3. \quad (3)$$

This adaptive weighting mechanism combined with multiple expert networks enables the gated weight generation network to effectively capture domain-specific features after feature extraction, thereby improving the robustness and adaptability of the model to photos in different data domains.

As shown in Figure 3, the color restoration expert network (Expert1, also named Net1) is used to restore the color information of images under low light conditions. The color restoration subnetwork we designed employs pooling operations to focus on key color features during downsampling while simultaneously learning image information. Additionally, deconvolution operations are utilized to precisely restore image details, with nonlinear interpolation adopted to ensure smooth and natural color transitions:

$$Y_1(i, j) = \sum_{m, n} w_{mn} X(i + m, j + n) \quad (4)$$

where $Y_1(i, j)$ denotes the output at position (i, j) , $X(i + m, j + n)$ is the input feature at the adjacent position, and w_{mn} is the interpolation weight, which satisfies: $w_{mn} = 1$ to ensure luminance consistency. To preserve the original image characteristics, the processed tensor is connected to the input through a residual connection. Finally, a Sigmoid activation function is used to limit the color output to the

interval $[0, 1]$, reducing color anomalies and oversaturation problems and ensuring that the enhanced image colors are natural and realistic.

As shown in Fig. 3, the detail enhancement subnetwork (Expert2, also named Net2) uses convolutions and attention mechanisms to enhance image details. The network uses different attention mechanisms in combination for feature extraction. Among them, important channel features are extracted through the channel attention mechanism. At the same time, the spatial attention mechanism is used, which combines Max Pooling and Avg Pooling, and then processed by convolution. To fuse the characteristics of different attention mechanisms, we concatenate different attentions, where max pooling and average pooling are used to focus on key spatial positions in the image. Finally, the outputs of channel attention and spatial attention are combined with the original input image through a residual connection to preserve the original features and enhance the detail recovery ability. This structure improves the detail recovery ability of the image.

As shown in Fig. 3, the advanced feature enhancement subnetwork (Expert3, also named Net3) improves image quality through convolution, multi-scale feature extraction, a gating mechanism, and an attention mechanism. The input image is passed through a multi-scale convolution to extract and fuse features. These fused features are then processed further by a gating network (SG) and a channel attention mechanism (SCA). Finally, the enhanced features are added back to the input image via a residual connection to preserve the original details.

This method can effectively adapt to low-light scenes and improve image quality by dynamically adjusting the weights of the expert network.

3.2. Shallow Feature Extraction Module

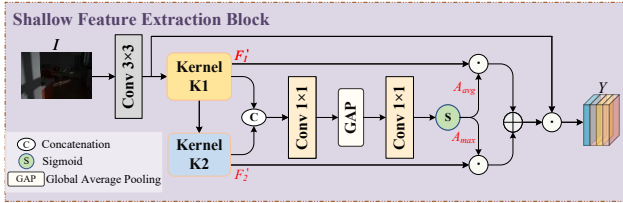


Figure 4. **Shallow Feature Extraction Block.** The architecture of SFEB uses parallel convolutional and GAP for multi-scale feature capture.

To improve the effectiveness of image feature extraction and suppress invalid features, we design a multi-scale feature enhancement module (SFEB, Shallow Feature Enhancement Block) to process the input feature map $X \in \mathbb{R}^{C \times H \times W}$. The SFEB generates a feature map $F_1 \in \mathbb{R}^{C \times H \times W}$ through a 3×3 depth separable convolution, as shown in Fig. 4. In addition, SFEB also obtains

a feature map $F_2 \in \mathbb{R}^{C \times H \times W}$ using different sizes of convolution kernels through hole convolution (dilation rates) to capture multi-scale spatial information.

SFEB uses two convolutions to compress the channel numbers of F_1 and F_2 to form the fused feature map $F_e \in \mathbb{R}^{C' \times H \times W}$, which fuses different feature information.

To introduce the attention mechanism, SFEB performs global pooling on the fused feature map F_e to obtain the channel-weighted features A_{avg} and A_{max} . Then, the attention map $F_w \in \mathbb{R}^{C' \times H \times W}$ is generated by channel concatenation and a 7×7 convolution to enhance the features of key regions:

$$F_w = F_1' \odot A_{avg} + F_2' \odot A_{max} \quad (5)$$

Finally, the output feature map Y is obtained by element-wise multiplication of the input feature X and the attention map F_w :

$$Y = X \odot F_w \quad (6)$$

This design, which combines a multi-scale convolution with an attention mechanism, gives SFEB stronger feature extraction capabilities, thereby improving the brightness and detail recovery of the image.

3.3. Loss Function

To ensure that the network-generated image \hat{I} is as close as possible to the reference image I_{gt} , we introduce the peak signal-to-noise ratio loss (PSNRLoss) as a loss function in the training to measure and maximize the quality of the output image. We first define the mean squared error (MSE) as:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N \left(\hat{I}(i) - I_{gt}(i) \right)^2 \quad (7)$$

and then define the PSNR loss based on the MSE:

$$\text{PSNR loss} = -\frac{10}{\log(10)} \cdot \log(\text{MSE} + \epsilon) \quad (8)$$

where N is the total number of pixels in the image, $\hat{I}(i)$ and $I_{gt}(i)$ are the predicted and true values at pixel position i , respectively, and ϵ is a small positive number used to prevent the denominator from becoming zero. During training, the network weights are updated by minimizing the PSNR Loss to improve the network's image restoration ability.

4. Experiment

4.1. Datasets and Implementation Details

To evaluate the effectiveness of GM-MoE system, five prominent LLIE datasets were employed: LOLv1 [50], LOLv2-Real [58], LOLv2-Synthetic [58], LSRW-Huawei [13], and LSRW-Nikon [13]. Specifically, LOLv1 contains 485 training pairs and 15 test pairs captured

Methods	LOL-v1 [50]		LOLv2-Real [58]		LOLv2-Synthetic [58]		#Param (M)
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	
SID [4]	14.35	0.436	13.24	0.442	15.04	0.610	7.76
RF [23]	15.23	0.452	14.05	0.458	15.97	0.632	21.54
UFormer [49]	16.36	0.771	18.82	0.771	19.66	0.871	5.29
EnGAN [20]	17.48	0.620	18.23	0.617	16.57	0.734	114.35
Restormer [61]	22.43	0.823	19.94	0.827	21.41	0.830	26.13
Retinexformer [2]	25.16	0.845	22.80	0.840	25.67	0.930	1.61
DeepUPE [43]	14.38	0.446	13.27	0.452	15.08	0.623	1.02
LIME [12]	16.76	0.560	15.24	0.419	16.88	0.757	-
MF [6]	18.79	0.640	18.72	0.508	17.50	0.773	-
NPE [44]	16.97	0.589	17.33	0.452	16.60	0.778	-
SRIE [7]	11.86	0.500	14.45	0.524	14.50	0.664	0.86
RetinexNet [50]	16.77	0.560	15.47	0.567	17.13	0.798	0.84
Kind [64]	20.86	0.790	14.74	0.641	13.29	0.578	8.02
Kind++ [65]	21.80	0.831	20.59	0.829	21.17	0.881	8.27
MIRNet [60]	24.14	0.830	20.02	0.820	21.94	0.876	31.76
SNR-Net [52]	24.61	0.842	21.48	0.849	24.14	0.928	39.12
Bread [11]	22.92	0.812	20.83	0.821	17.63	0.837	2.12
DPEC [45]	24.80	0.855	22.89	0.863	26.19	0.939	2.58
PairLIE [8]	23.53	0.755	19.89	0.778	19.07	0.794	0.33
LLFormer [46]	25.76	0.823	20.06	0.792	24.04	0.909	24.55
QuadPrior [47]	22.85	0.800	20.59	0.811	16.11	0.758	1252.75
3DLUT[62]	14.35	0.445	17.59	0.721	18.04	0.800	0.59
Sparse[58]	17.20	0.640	20.06	0.816	22.05	0.905	1.08
RUAS[37]	18.23	0.720	18.37	0.723	16.55	0.652	1.03
DRBN [57]	20.13	0.830	20.29	0.831	23.22	0.927	1.83
Ours	26.66	0.857	23.65	0.806	26.30	0.937	19.99

Table 1. **Quantitative Comparison on LOL-v1, LOLv2-Real, and LOLv2-Synthetic Datasets.** The best results are highlighted in bold red, and the second-best results are highlighted in bold blue.

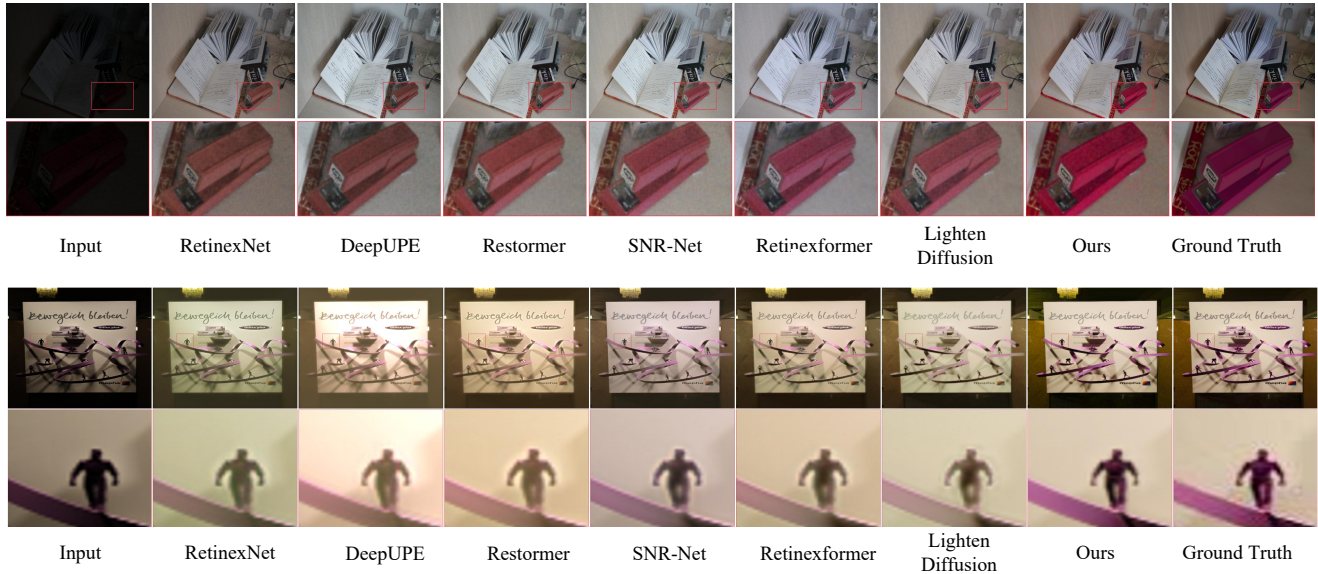


Figure 5. **Qualitative comparison on LOLv1 (first row) and LOLv2-Synthetic(second row) .** It can be seen that the proposed method significantly improves image clarity, and the colours are closer to reality.

from real scenes under different exposure times. LOLv2-Real, which includes 689 training pairs and 100 test pairs collected by adjusting exposure time and ISO, and LOLv2-Synthetic, which is generated by analyzing the lighting distribution of low-light images and contains 900 training pairs and 100 test pairs. The LSRW-Huawei and LSRW-Nikon

datasets each contain several real low-light images captured by devices in real-world scenes.

Implementation Details: The GM-MoE was developed using the PyTorch framework and trained on an NVIDIA 4090 GPU. The training process began with an initial learning rate of 1.0×10^{-3} , which was managed using a multi-step

Methods	LSRW-Huawei [13]		LSRW-Nikon [13]	
	PSNR	SSIM	PSNR	SSIM
SID [4]	17.47	0.652	16.33	0.613
RF [23]	19.05	0.637	18.77	0.630
UFormer [49]	19.77	0.643	19.77	0.643
EnGAN [20]	20.22	0.701	20.71	0.659
Restormer [61]	22.61	0.725	21.20	0.677
Retinexformer [2]	19.57	0.578	-	-
LIME [12]	17.00	0.382	13.53	0.332
MF [6]	18.26	0.428	15.44	0.400
NPE [44]	17.08	0.391	14.86	0.374
SRIE [7]	13.42	0.428	13.26	0.140
RetinexNet [50]	19.98	0.688	19.86	0.650
Kind [64]	16.58	0.569	11.52	0.383
Kind++ [65]	15.43	0.570	14.79	0.475
MIRNet [60]	19.98	0.609	17.10	0.502
SNR-Net [52]	20.67	0.591	17.54	0.482
Bread [11]	19.20	0.618	14.70	0.487
LightenDiffusion [18]	18.56	0.539	-	-
3DLUT[62]	18.12	0.659	17.81	0.629
Sparse[58]	20.33	0.699	20.19	0.657
RUAS[37]	20.46	0.704	20.88	0.664
DRBN [57]	20.61	0.710	21.07	0.670
Ours	23.55	0.741	22.62	0.700

Table 2. **Quantitative comparison of the LSRW-Huawei and LSRW-Nikon datasets.** The best results are highlighted in bold red, and the second-best results are highlighted in bold blue.

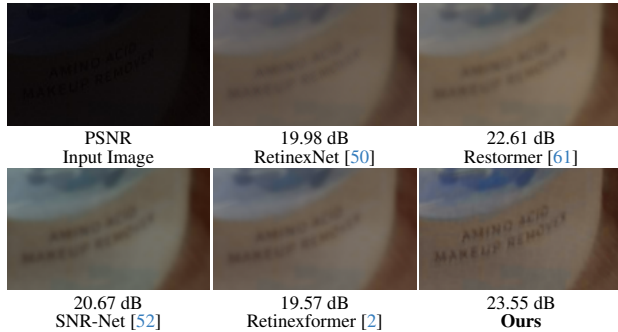


Figure 6. **Image enhancement example.** Qualitative comparison on LSRW-Huawei. Our network clearly restores the fine details of the mineral water text in the image.

scheduler. The Adam [33] optimizer, configured with a momentum parameter of 0.9, was used for optimization. During training, input images were resized to 256×256 pixels and subjected to data augmentation techniques, including random rotations and flips, to enhance model generalization. A batch size of 4 was maintained, and the training regimen consisted of a total of 2.0×10^6 iterations. Performance was evaluated using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) [48] as the primary metrics.

4.2. Low-light Image Enhancement

Quantitative Results. Among the deep learning methods for low-light image enhancement are SID [4], RF [23],

UFormer [49], EnGAN [20], Restormer [61], Retinexformer [2], DeepUPE [43], RetinexNet [50], Kind [64], Kind++ [65], MIRNet [60], SNR-Net [52], Bread [11], DPEC [45], as well as traditional methods such as, LIME [12], MF [6], NPE [44], and SRIE [7], GM-MoE achieved an overall performance improvement on all datasets. Tab. 1 shows the quantitative comparison results of GM-MoE with a variety of SOTA image enhancement algorithms. GM-MoE achieved PSNR improvements of 1.5, 1.24, and 0.11 dB on the LOL-v1 [50], LOLv2-Real [58], and LOLv2-Synthetic [58] datasets, respectively, compared to the second-ranked model in each dataset, significantly improving image quality. This showed that GM-MoE performed well on these classic datasets, consistently outperforming other methods in terms of detail recovery and color enhancement, both in synthetic and real low-light conditions. In addition, as shown in Tab. 2, GM-MoE also achieved significant improvements over other SOTA methods on the LSRW-Huawei [13] and LSRW-Nikon [13] datasets, achieving PSNR improvements of 0.94 dB and 1.42 dB over the next best method, Restormer [61], respectively. These two datasets contained a large amount of noise and artifacts, and GM-MoE was able to effectively reduce artifacts and recover image details and features in high-noise environments. For more results please refer to supplementary material.



Figure 7. A visual comparison of the object detection task in low-light scenes (left) and scenes enhanced by GM-MoE (right).

Qualitative Results. A visual comparison of the GM-MoE and the other algorithms is shown in Fig. 5, and 6 (Zoom in for better visualization.). Previous methods exhibited poor edge detail processing, with some blurring effects and noise, as shown in Fig. 6 for RetinexNet [50], Retinexformer [2], and SNR-Net [52]. Moreover, multiple networks had issues with color distortion, as shown in Fig. 5 for RetinexNet [50] and Fig. 6 for SNR-Net [52]. In addition, there were cases of underexposure or overexposure, as seen in DeepUPE [43] in Fig. 5. In contrast, our work effectively restored colors, efficiently restored details, extracted shallow features, significantly reduced noise, and reliably preserved colors. As can be seen, our method outperformed other supervised and unsupervised algorithms across various scenarios and excelled in multiple metrics. The left and right parts of Fig. 7 show the performance of object detection in low-light scenes (left) and photos enhanced by GM-MoE, respectively.

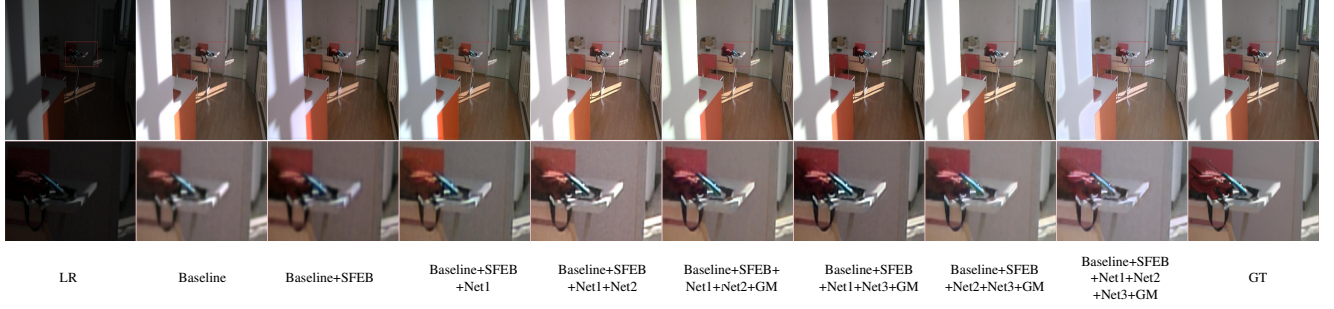


Figure 8. **Qualitative comparison of different ablation settings.** It can be observed that the final configuration (Baseline+SFEB+Net1+Net2+Net3+GM) produces the best results.

4.3. Ablation Study.

We incrementally added modules to the baseline model to assess their contributions and to conduct ablation studies on the LOL-v2-real [58] and LOL-v2-syn [58] datasets. The results are reported in Tab. 3.

The Effectiveness of the Shallow Feature Extraction Network. To verify the effectiveness of the shallow feature extraction network, we first introduced SFEB into the baseline model. Comparing the results of Experiment 1 (baseline model) with those of Experiment 2 (baseline + SFEB), we found that, on the LOL-v2-syn dataset, PSNR improved by **3.09 dB** and SSIM improved by 0.0215. This shows that SFEB can effectively extract the shallow features of an image, providing better feature input for the subsequent GM-MoE module.

Is the Effectiveness Among Multiple Experts and the Performance Complementary? To explore the role of each expert module (Expert1, Expert2, Expert3) and its complementary role in image enhancement, we gradually added each subnetwork to the model and performed ablation experiments. After adding **Expert1**, compared with Experiment 2, PSNR increased by **1.08 dB** (LOL-v2-real). After adding **Expert2**, compared with Experiment 3, PSNR increased by 0.76 dB (LOL-v2-real) and 0.79 dB (LOL-v2-syn), and SSIM increased by 0.0575 and 0.0891, respectively. Subsequently, we removed Expert1, Expert2, and Expert3 respectively through ablation experiments (Experiments 5–7), and a decrease in performance was observed in all cases. This shows that the expert modules can work together in synergy after being integrated into a module to solve the image enhancement problems they are designed for, improving the overall image restoration effect.

Does the Gated Weight Generation Network Improve Generalization Ability? In the complete model, we further introduce a gated weight generation network. Compared to models without this mechanism, it dynamically adjusts the weights of individual experts based on images from different data domains, thereby enhancing cross-domain generalization capability.

Qualitative results. As shown in the ablation study results

ID	Baseline	SFEB	Net1	Net2	Net3	GM	LOL-v2-real [58]		LOL-v2-syn [58]	
							PSNR	SSIM	PSNR	SSIM
1	✓						19.45	0.7079	20.35	0.7431
2	✓	✓					20.27	0.7236	23.44	0.7646
3	✓	✓	✓				21.35	0.7446	24.35	0.8436
4	✓	✓	✓	✓			22.11	0.8021	25.14	0.9327
5	✓	✓	✓	✓		✓	23.23	0.8045	26.08	0.9351
6	✓	✓	✓	✓	✓	✓	23.31	0.8054	26.12	0.9362
7	✓	✓	✓	✓	✓	✓	23.35	0.8055	26.15	0.9366
8	✓	✓	✓	✓	✓	✓	23.65	0.8060	26.29	0.9371

Table 3. **Ablation study results on LOL-v2-real and LOL-v2-syn datasets.** The best results are highlighted in bold red.

in Fig. 8, images generated using only the baseline model exhibit noticeable blurring and loss of detail. The sequential integration of SFEB, Net1, Net2, Net3, and the gated weight generation module (GM) progressively enhances the model’s ability to recover low-light images. Each component contributes uniquely to the overall performance, and the complete model achieves the best results.

5. Conclusion

This paper proposes the GM-MoE framework, which dynamically balances three expert subnetworks based on input image features, simultaneously addressing color bias, detail loss, and insufficient illumination in low-light images while ensuring strong generalization across diverse data domains. Extensive quantitative and qualitative evaluations on five benchmark datasets demonstrate that GM-MoE outperforms existing methods in both PSNR and SSIM metrics. Future work will focus on real-time enhancement and adaptive optimization.

Acknowledgments

This work was supported by the Xinjiang Uygur Autonomous Region Tianshan Ying Talents Leading Talents Program for Scientific and Technological Innovation (2023TSYCLJ0025), the National Natural Science Foundation of China (No. 62266044, 61563052, 61862061), and the National Key Research and Development Program of China (2021YFB2802100).

References

- [1] Shuqing Bian, Xingyu Pan, Wayne Xin Zhao, Jinpeng Wang, Chuyuan Wang, and Ji-Rong Wen. Multi-modal mixture of experts representation learning for sequential recommendation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 110–119, 2023. 3
- [2] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12504–12513, 2023. 6, 7
- [3] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3291–3300, 2018. 2, 3
- [4] Chen Chen, Qifeng Chen, Minh N. Do, and Vladlen Koltun. Seeing motion in the dark. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3184–3193, 2019. 6, 7
- [5] William Fedus, Barret Zoph, and Noam Shazeer. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *Journal of Machine Learning Research*, 23(120):1–39, 2022. 3
- [6] Xueyang Fu, Delu Zeng, Yue Huang, Yinghao Liao, Xinghao Ding, and John Paisley. A fusion-based enhancing method for weakly illuminated images. *Signal Processing*, 129:82–96, 2016. 6, 7
- [7] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2782–2790, 2016. 6, 7
- [8] Zhenqi Fu, Yan Yang, Xiaotong Tu, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Learning a simple low-light image enhancer from paired low-light instances. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22252–22261, 2023. 6
- [9] Rafael C Gonzalez and Richard E Woods. *Digital image processing*. Prentice Hall Press, 2008. 2
- [10] Xiaojie Guo. Lime: A method for low-light image enhancement. In *Proceedings of the 24th ACM international conference on Multimedia*, pages 87–91, 2016. 3
- [11] Xiaojie Guo and Qiming Hu. Low-light image enhancement via breaking down the darkness. *International Journal of Computer Vision*, 131(1):48–66, 2023. 6, 7
- [12] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 26(2):982–993, 2017. 3, 6, 7
- [13] Jiang Hai, Zhu Xuan, Ren Yang, Yutong Hao, Fengzhu Zou, Fang Lin, and Songchen Han. R2rnet: Low-light image enhancement via real-low to real-normal network. *Journal of Visual Communication and Image Representation*, 90: 103712, 2023. 5, 7
- [14] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010. 2, 3
- [15] Po-Wen Hsieh, Pei-Chiang Shao, and Suh-Yuh Yang. Adaptive variational model for contrast enhancement of low-light images. *SIAM Journal on Imaging Sciences*, 13(1):1–28, 2020. 2
- [16] Robert A. Jacobs, Michael I. Jordan, Steven J. Nowlan, and Geoffrey E. Hinton. Adaptive mixtures of local experts. *Neural Computation*, 3(1):79–87, 1991. 3
- [17] Inho Jeong and Chul Lee. An optimization-based approach to gamma correction parameter estimation for low-light image enhancement. *Multimedia Tools and Applications*, 80: 18027–18042, 2021. 2
- [18] Hai Jiang, Ao Luo, Xiaohong Liu, Songchen Han, and Shuaicheng Liu. Lightendiffusion: Unsupervised low-light image enhancement with latent-retinex diffusion models. In *European Conference on Computer Vision*, pages 161–179. Springer, 2025. 1, 7
- [19] J. Jiang, X. Zheng, S. Luo, and H. Fu. Enlightengan: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing*, 30:2340–2349, 2019. 3
- [20] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing*, 30:2340–2349, 2019. 6, 7
- [21] D. J. Jobson, Z. Rahman, and G. A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing*, 6(7):965–976, 1997. 3
- [22] Wonjun Kim. Low-light image enhancement: A comparative review and prospects. *IEEE Access*, 10:84535–84557, 2022. 1
- [23] Satoshi Kosugi and Toshihiko Yamasaki. Unpaired image enhancement featuring reinforcement-learning-controlled image editing software. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(07):11296–11303, 2020. 6, 7
- [24] Dmitry Lepikhin, HyounJoong Lee, Yuanzhong Xu, Dehao Chen, Orhan Firat, Yanping Huang, Maxim Krikun, Noam Shazeer, and Zhifeng Chen. Gshard: Scaling giant models with conditional computation and automatic sharding. *arXiv preprint arXiv:2006.16668*, 2020. 3
- [25] Chongyi Li, Chunle Guo, Linghao Han, Jun Jiang, Ming-Ming Cheng, Jinwei Gu, and Chen Change Loy. Low-light image and video enhancement using deep learning: A sur-

- vey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):9396–9416, 2022. 1, 3
- [26] R. Li, W. Cai, Y. Zhang, J. Zhang, and J. Liu. Low-light image enhancement via progressive-recursive network. *IEEE Transactions on Image Processing*, 30:6459–6472, 2021. 3
- [27] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 1833–1844, 2021. 3
- [28] Tingting Liang, Xiaojie Chu, Yudong Liu, Yongtao Wang, Zhi Tang, Wei Chu, Jingdong Chen, and Haibin Ling. Cbnet: A composite backbone network architecture for object detection. *IEEE Transactions on Image Processing*, 31:6893–6906, 2022. 3
- [29] Choon Chen Lim, Yuen Peng Loh, and Lai-Kuan Wong. Lau-net: A low light image enhancer with attention and re-sizing mechanisms. *Signal Processing: Image Communication*, 115:116971, 2023. 3
- [30] Haoning Lin and Zhenwei Shi. Multi-scale retinex improvement for nighttime image enhancement. *optik*, 125(24):7143–7148, 2014. 3
- [31] Shouxin Liu, Wei Long, Yanyan Li, and Hong Cheng. Low-light image enhancement based on membership function and gamma correction. *Multimedia Tools and Applications*, 81(16):22087–22109, 2022. 2
- [32] Kin Gwn Lore, Adebayo Akintayo, and Soumik Sarkar. Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017. 2
- [33] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *ICLR*, 2017. 7
- [34] Shanto Rahman, Md Mostafijur Rahman, M. Abdullah-Al-Wadud, Golam Dastagir Al-Quaderi, and Mohammad Shoy-aib. An adaptive gamma correction for image enhancement. *Eurasip Journal on Image & Video Processing*, 2016(1):35, 2016. 2
- [35] Fabio Remondino, Ali Karami, Ziyang Yan, Gabriele Mazzacca, Simone Rigon, and Rongjun Qin. A critical analysis of nerf-based 3d reconstruction. *Remote Sensing*, 15(14):3585, 2023. 1
- [36] Wenqi Ren, Sifei Liu, Lin Ma, Qianqian Xu, Xiangyu Xu, Xiaochun Cao, Junping Du, and Ming-Hsuan Yang. Low-light image enhancement via a deep hybrid network. *IEEE Transactions on Image Processing*, 28(9):4364–4375, 2019. 2, 3
- [37] Liu Risheng, Ma Long, Zhang Jiaao, Fan Xin, and Luo Zhongxuan. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 6, 7
- [38] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015. 2, 4
- [39] Yihua Shao, Siyu Liang, Zijian Ling, Minxi Yan, Haiyang Liu, Siyu Chen, Ziyang Yan, Chenyu Zhang, Haotong Qin, Michele Magno, et al. Gwq: Gradient-aware weight quantization for large language models. *arXiv preprint arXiv:2411.00850*, 2024. 1
- [40] Yihua Shao, Yeling Xu, Xinwei Long, Siyu Chen, Ziyang Yan, Yang Yang, Haoting Liu, Yan Wang, Hao Tang, and Zhen Lei. Accidentblip: Agent of accident warning based on ma-former. *arXiv preprint arXiv:2404.12149*, 2024.
- [41] Yihua Shao, Minxi Yan, Yang Liu, Siyu Chen, Wenjie Chen, Xinwei Long, Ziyang Yan, Lei Li, Chenyu Zhang, Nicu Sebe, et al. In-context meta lora generation. *arXiv preprint arXiv:2501.17635*, 2025. 1
- [42] Anirudh Singh, Amit Chougule, Pratik Narang, Vinay Chamola, and F Richard Yu. Low-light image enhancement for uavs with multi-feature fusion deep neural networks. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022. 1
- [43] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–10, 2019. 6, 7
- [44] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, 22(9):3538–3548, 2013. 6, 7
- [45] Shuang Wang, Qianwen Lu, Boxing Peng, Yihe Nie, and Qingchuan Tao. Dpec: Dual-path error compensation method for enhanced low-light image clarity, 2024. 6, 7
- [46] Tao Wang, Kaihao Zhang, Tianrun Shen, Wenhan Luo, Bjorn Stenger, and Tong Lu. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2654–2662, 2023. 6
- [47] Wenjing Wang, Huan Yang, Jianlong Fu, and Jiaying Liu. Zero-reference low-light enhancement via physical quadruple priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 26057–26066, 2024. 6
- [48] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 7
- [49] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17683–17693, 2022. 6, 7
- [50] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *British Machine Vision Conference*, pages xx–xx, 2018. 3, 5, 6, 7
- [51] Haoran Xu, Jianming Guo, Qing Liu, and Lingli Ye. Fast image dehazing using improved dark channel prior. In *2012 IEEE international conference on information science and technology*, pages 663–667. IEEE, 2012. 3
- [52] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *Proceedings of*

- the *IEEE/CVF conference on computer vision and pattern recognition*, pages 17714–17724, 2022. 6, 7
- [53] Y. Xu and et al. Snr-aware: Signal-to-noise ratio aware detail enhancement for low-light images. *IEEE Transactions on Image Processing*, 2023. 3
- [54] Ziyang Yan, Wenzhen Dong, Yihua Shao, Yuhang Lu, Liu Haiyang, Jingwen Liu, Haozhe Wang, Zhe Wang, Yan Wang, Fabio Remondino, et al. Renderworld: World model with self-supervised 3d label. *arXiv preprint arXiv:2409.11356*, 2024. 1
- [55] Ziyang Yan, Lei Li, Yihua Shao, Siyu Chen, Wuzong Kai, Jenq-Neng Hwang, Hao Zhao, and Fabio Remondino. 3dsce-needitor: Controllable 3d scene editing with gaussian splatting. *arXiv preprint arXiv:2412.01583*, 2024. 1
- [56] Shaoliang Yang, Dongming Zhou, Jinde Cao, and Yanbu Guo. Lightingnet: An integrated learning method for low-light image enhancement. *IEEE Transactions on Computational Imaging*, 9:29–42, 2023. 3
- [57] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 6, 7
- [58] Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing*, 30:2072–2086, 2021. 5, 6, 7, 8
- [59] Zishu Yao, Guodong Fan, Jinfu Fan, Min Gan, and CL Philip Chen. Spatial-frequency dual-domain feature fusion network for low-light remote sensing image enhancement. *IEEE Transactions on Geoscience and Remote Sensing*, 2024. 1
- [60] Hu Yu, Naishan Zheng, Man Zhou, Jie Huang, Zeyu Xiao, and Feng Zhao. Frequency and spatial dual guidance for image dehazing. In *European Conference on Computer Vision*, pages 181–198. Springer, 2022. 6, 7
- [61] Syed Waqas Zamir, A R Rajeev Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5728–5739, 2022. 2, 3, 6, 7
- [62] Hui Zeng, Jianrui Cai, Lida Li, Zisheng Cao, and Lei Zhang. Learning image-adaptive 3d lookup tables for high performance photo enhancement in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 6, 7
- [63] X. Zhang, Y. Fan, S. Deng, Y. Huang, and Y. Zhang. Ll-former: A lightweight transformer for low-light image enhancement with information distillation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 3233–3241, 2022. 3
- [64] Y. Zhang, J. Zhang, X. Guo, and C. Chen. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 1632–1640, 2019. 3, 6, 7
- [65] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *International Journal of Computer Vision*, 129(4):1013–1037, 2021. 6, 7
- [66] Zunjin Zhao, Bangshu Xiong, Lei Wang, Qiaofeng Ou, Lei Yu, and Fa Kuang. Retinexdip: A unified deep framework for low-light image enhancement. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(3):1076–1088, 2021. 3
- [67] Fei Zhou, Xin Sun, Junyu Dong, and Xiao Xiang Zhu. Surroundnet: Towards effective low-light image enhancement. *Pattern Recognition*, 141:109602, 2023. 2, 3
- [68] Zhaorun Zhou, Zhenghao Shi, and Wenqi Ren. Linear contrast enhancement network for low-illumination image enhancement. *IEEE Transactions on Instrumentation and Measurement*, 72:1–16, 2022. 2