GlowNet: A Lightweight Attention based Self-Guided Deep Neural Network for Low-Light Image Enhancement

Abhiroop Chatterjee[1]
abhiroopchat1998@gmail.com
Susmita Ghosh[1]
susmitaghoshju@gmail.com
Ashish Ghosh[2]
ash@isical.ac.in

[1] Jadavpur University
Kolkata, India
[2] Machine Intelligence Unit
Indian Statistical Institute
Kolkata, India

Abstract

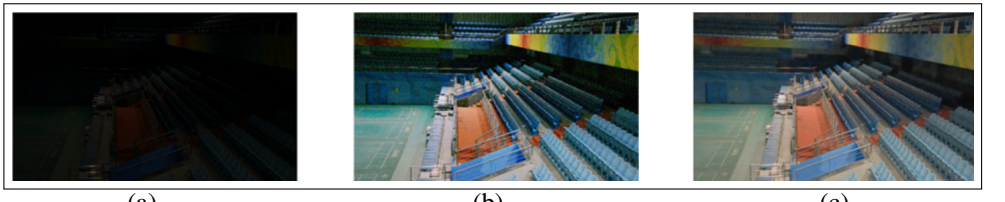
Low-light image enhancement poses significant challenges in machine vision due to poor illumination conditions of the scenes. The present study introduces a novel deep-learning approach specifically designed for enhancing low-light images. The proposed technique exploits a deep neural network architecture integrating channel & multi-scale spatial attention, self-attention, and dilated convolutions. These components are adept at capturing contextual information across channels, scales, and features thereby enabling effective enhancement while minimizing noise and artifacts. For generating the pixel values of the enhanced image, function approximation technique is used that aids in mapping the given input image into an enhanced image by adjusting the dynamic range of low-light images. In doing this, customized loss functions have been utilized to guide training and optimize the model's performance, addressing key aspects such as color constancy, exposure fidelity, illumination smoothness, and spatial consistency between neighboring pixels. Empirical assessments on benchmark datasets show that the proposed approach outperforms existing approaches, both in quantitative metrics such as PSNR, SSIM, and MAE, as well as in qualitative visual outcomes. GlowNet achieved a PSNR of 18.29 dB, SSIM of 0.62, and MAE of 92.51 on the benchmark LOL dataset.

1 Introduction

Enhancing images captured in low-light conditions is a significant challenge in computer vision and image processing [12], with implications for various fields like surveillance, photography, and autonomous driving. Low-light images typically exhibit reduced visibility, increased noise, and loss of detail, making further analysis more difficult. To overcome these issues, we introduce GlowNet, a novel self-guided method featuring an advanced neural network architecture. GlowNet integrates channel attention, spatial attention, self-attention mechanisms, and dilated convolutions for precise polynomial estimation. This design captures contextual information across channels and scales, prioritizing key features for

enhancement while minimizing outliers. GlowNet also employs curve estimation to modulate the dynamic range of images effectively. GlowNet's self-supervised approach, utilizing fewer parameters compared to SOTA approaches like MIRNet-v2 [17], reduces overfitting and avoids the need for paired supervision. As shown in Figure 1, GlowNet achieves impressive image enhancement with restored fine-grained details utilizing only 0.24 million parameters compared to MIRNet-v2's 5.85 million, demonstrating its efficiency as a **lightweight network** by employing a compact architecture that leverages advanced feature extraction techniques and efficient feature fusion modules to deliver competitive performance with significantly fewer computational resources. By leveraging deep neural network [9] and attention mechanisms [4], GlowNet enhances low-light images while preserving details and minimizing artifacts.

This paper is organized in the following manner: Section 2 reviews related research in this field; Section 3 details GlowNet's design; Section 4 covers datasets, experimental settings, and results; Section 5 presents the conclusions and outlines future research directions.



(a) (b) (c)

Figure 1: Visual comparison: (a) A low-light image as input, (b) Enhanced using the proposed GlowNet, (c) Enhanced using the SOTA MIRNet-v2 [17].

2 Related Research

In recent years, low-light image enhancement has advanced significantly with deep learning techniques. MIRNet-v2 [17] concentrates on high-resolution features while leveraging contextual details from low-resolution images. SRIE [2] by Fu et al. estimates reflectance and illumination simultaneously, while LIME [5] refines a coarse illumination map using structural priors. Li et al. [7] introduced a Retinex model which estimates illumination using an optimization problem without logarithmic transformation. CNN-based models, like LLNet [10] and Retinex-Net [15], rely on paired data, which can be costly and impractical. Wang et al. [14] use intermediate illumination and paired data but face similar challenges. Unsupervised GAN-based methods, such as EnlightenGAN [6], avoid the need for paired data by using global-local discriminators and self-regularized losses. Zero-DCE [3] employs a lightweight network to estimate image-specific curves without reference images, while KinD Net [13] decomposes images into illumination and reflectance for flexible adjustments. The STAR model [16] uses local derivatives for regularization, RUAS [8] applies neural architecture search based on the Retinex rule, and SCI [11] presents a cascaded illumination learning process. The Generative Diffusion Prior (GDP) [1] uses a pre-trained diffusion model for unsupervised restoration with a new conditional guidance protocol.

As mentioned, the proposed framework encapsulates attention methods, self-guidance and estimates intricate pixel transformation through minimizing loss functions. The method integrates various attention mechanisms at strategic positions to selectively focus on informative features, dynamically adjusting their importance and enhancing feature representations.

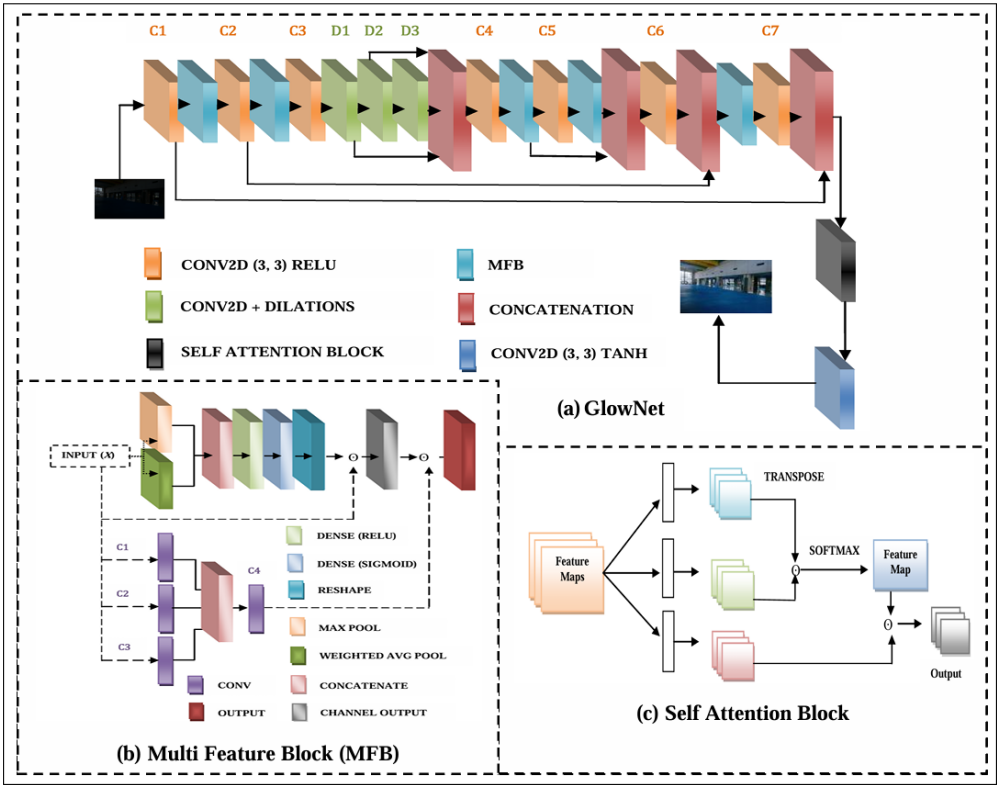


Figure 2: Block diagram of the proposed architecture: (a) GlowNet - core network for low-light enhancement with multi-scale and attention mechanisms; (b) Multi-Scale Feature Block (MFB) - captures fine details with multi-scale convolutions; (c) Self-Attention Block - captures global context for natural enhancement.

3 Methodology

In this work, the illumination enhancement task is formulated as a pixel-to-pixel mapping problem, where the objective is to transform a low-light image $X \in \mathbb{R}^{H\times W\times 3}$ into an enhanced image $X' \in \mathbb{R}^{H\times W\times 3}$, with H and W representing the height and width of the image, and 3 represents the total color channels. The transformation is modeled by a polynomial function $P(X)$ of degree n:

$$X' = P(X) = a_0 + a_1 X + a_2 X^2 + \cdots + a_n X^n, \tag{1}$$

where $a_0, a_1, \ldots, a_n \in \mathbb{R}$ are the polynomial coefficients, and $\mathbb{R}$ represents the set of real numbers. These coefficients govern the nonlinear transformation. The optimal coefficients are learned by minimizing a composite loss function $L(X, X')$, which is optimized using gradient descent.

3.1 Multi-Scale Feature Block (MFB)

The Multi-Scale Feature Block (MFB) (Figure 2) captures multi-scale contextual information using convolutions of varying sizes combined with channel and spatial attention mechanisms. This is given below in detail.

Channel-Wise Attention

The channel-wise attention mechanism emphasizes informative features across channels. It computes global max pooling ($f_{\text{GMP}}(c)$) and global average pooling ($f_{\text{GAP}}(c)$) descriptors:

$$f_{\text{GMP}}(c) = \max_{i,j} F_{(i,j,c)}, \quad f_{\text{GAP}}(c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_{(i,j,c)}, \tag{2}$$

where $F_{(i,j,c)}$ represents the pixel value at location (i, j) in channel c. These descriptors are concatenated and processed by two successive fully connected layers (FC_1 and FC_2) to produce channel attention weights, $\mathbf{w}_c$:

$$\mathbf{w}_c = \sigma(\text{FC}_2(\text{ReLU}(\text{FC}_1([\mathbf{f}_{\text{GMP}} \parallel \mathbf{f}_{\text{GAP}}])))), \tag{3}$$

where $\sigma(\cdot)$ is the sigmoid activation. The refined feature map F' is produced by scaling the original feature map F:

$$F'_{(i,j,c)} = F_{(i,j,c)} \cdot w_c. \tag{4}$$

Multi-Scale Spatial Attention

To capture spatial relationships, convolutions with kernel sizes 3×3, 5×5, and 7×7 are applied to F, producing feature maps F_3, F_5, and F_7. These are concatenated and passed through a 1×1 convolution operation to generate a spatial attention map, $\mathbf{w}_s$:

$$\mathbf{w}_s = \sigma(\text{Conv}_{1\times1}([F_3 \parallel F_5 \parallel F_7])), \tag{5}$$

where $\mathbf{w}_s \in \mathbb{R}^{H \times W \times 1}$. The final output F'' is:

$$F''_{(i,j,c)} = F'_{(i,j,c)} \cdot w_{s(i,j)}. \tag{6}$$

This combination of multi-scale convolutions and attention mechanisms captures both fine details and broader context, improving low-light image enhancement by focusing on significant regions and preserving important structures.

3.2 GlowNet: The Attention-Based Self-Guided Deep Neural Network

As mentioned earlier, GlowNet employs advanced deep learning mechanisms like **channel and spatial attention**, **self-attention**, and **dilated convolutions** to enhance low-light images. Its architecture, illustrated in **Figure 2**, captures both local and global image features. The input layer is designed for RGB images, followed by a sequence of convolutional layers C_i (where $i = 1, 2, \ldots, 7$) for hierarchical feature extraction.

To capture multi-scale features, **Multi-scale Feature Blocks (MFBs)** (**Figure 2(b)**) are placed after layers C_1, C_2, C_4, and C_5 (**Figure 2(a)**). After C_3, three **dilated convolutional**

layers D_i (with dilation rates of 2, 3, and 4) are utilized to broaden the receptive field without adding extra parameters, thereby efficiently capturing multi-scale context.

Feature maps from the dilated convolutions are concatenated to integrate multi-scale information. Concatenation layers (red blocks in **Figure 2(a)**) merge features from earlier and later layers, enhancing representational power. For instance, concatenating feature maps from C_2 with C_6 and C_1 with C_7 promotes smooth gradient flow and fuses features across different abstraction levels.

The **self-attention mechanism** (black block in **Figure 2(c)**) is applied to the concatenated feature maps to capture global context by computing attention scores. It does this by creating query (Q), key (K), and value (V) matrices:

$$Q = W_q Z, \quad K = W_k Z, \quad V = W_v Z, \tag{7}$$

where W_q, W_k, and W_v are learned weights, and Z is the input feature map obtained from the last concatenation layer of the GlowNet. The attention scores are then calculated as:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{Q^T K}{\sqrt{d_k}}\right) V, \tag{8}$$

where d_k is the dimension of the key vectors. The output feature maps are processed through a final convolutional layer with a **hyperbolic tangent (tanh) activation** to produce the enhanced image while preserving key features.

The effectiveness of these components is demonstrated by quantitative results in **Table 1**, showing superior performance in low-light enhancement tasks.

3.3 Loss Functions

In this work, image enhancement is viewed as an optimization problem with several key loss functions: color difference, exposure fidelity, illumination smoothness, and spatial consistency between neighboring pixels. Each guides the model to produce high-quality enhanced images and are described below in details.

Weighted Color Difference Loss

Maintaining consistent color balance across RGB channels is crucial for avoiding artifacts in low-light image enhancement. The **color difference loss** ensures natural color tones by balancing the mean values of each RGB channel in the enhanced image X':

$$\bar{X'}_{\text{RGB}} = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X'_{(i,j)}, \tag{9}$$

where $\bar{X'}_{\text{RGB}} = [\bar{X'}_R, \bar{X'}_G, \bar{X'}_B]$. To balance colors, we minimize the squared differences between mean channel values:

$$d_{RG} = (\bar{X'}_R - \bar{X'}_G)^2, \tag{10}$$

$$d_{RB} = (\bar{X'}_R - \bar{X'}_B)^2, \tag{11}$$

$$d_{GB} = (\bar{X'}_G - \bar{X'}_B)^2. \tag{12}$$

α adjusts the importance of these differences and effectively modulates the weight of each color difference d_{ij} thereby impacting the sensitivity of the loss to color imbalances in the enhanced image. This adjustment is mathematically represented by:

$$\text{weighted\_d}_{ij} = \alpha \cdot d_{ij}, \quad \text{for } (i,j) \in \{(R,G),(R,B),(G,B)\}. \tag{13}$$

The **color difference loss**, L_{color}, is now computed as:

$$L_{\text{color}} = \sqrt{\text{weighted\_d}_{RG}^2 + \text{weighted\_d}_{RB}^2 + \text{weighted\_d}_{GB}^2}. \tag{14}$$

This loss function ensures that the color balance remains consistent and natural across various lighting conditions, thereby reducing the risk of introducing unnatural color shifts and preserving the visual authenticity of the enhanced image.

Deviation Penalizing Exposure Loss

The **spatial attention-based exposure loss** refines exposure levels to achieve the desired brightness while preserving essential details. It dynamically weighs the loss function using spatial attention, which focuses on key regions. Given an input image $X \in \mathbb{R}^{H\times W\times 3}$, the global mean intensity ($\bar{X}$) across all color channels is computed as:

$$\bar{X} = \frac{1}{H\times W} \sum_{c\in\{R,G,B\}} \sum_{i=1}^{H} \sum_{j=1}^{W} X_{(i,j,c)}, \tag{15}$$

To assess regional brightness, we apply average pooling with a non-overlapping window of size $k\times k$. For a given pooled region indexed by (p,q), where p and q refer to the position of the pooling window in the pooling grid, the average intensity is:

$$\bar{X}_{\text{region}}(p,q) = \frac{1}{k^2} \sum_{i=pk}^{(p+1)k-1} \sum_{j=qk}^{(q+1)k-1} X_{(i,j,c)}, \tag{16}$$

where (i,j) are the coordinates within the pooling window, and pk and qk denote the top-left corner of the pooling window.

The **spatial attention weights** $\mathbf{w}_{\text{attention}} \in \mathbb{R}^{H\times W}$ are computed to highlight significant regions. These weights are derived from global average pooling over all color channels:

$$\mathbf{w}_{\text{attention}}(i,j) = \frac{\sum_c X_{(i,j,c)}}{\sum_{i,j,c} X_{(i,j,c)}}, \tag{17}$$

The **weighted mean** for each region is then calculated as:

$$\bar{X}_{\text{weighted}}(p,q) = \bar{X}_{\text{region}}(p,q) \cdot \mathbf{w}_{\text{attention}}(p,q). \tag{18}$$

The **exposure loss,** L_{exposure}, is defined as the squared difference between the weighted mean and a target mean μ, where μ is set to 0.65 in the present experiment:

$$L_{\text{exposure}} = \frac{1}{N} \sum_{p,q} \left(\bar{X}_{\text{weighted}}(p,q) - \mu\right)^2, \tag{19}$$

where N is the number of $k\times k$ pooled regions.

This loss function effectively balances both global and local exposure adjustments, emphasizing crucial regions to ensure natural and visually appealing results.

Illumination Smoothness Loss

The **Illumination Smoothness Loss,** $L_{\text{illumination}}$, promotes smooth transitions in illumination by penalizing abrupt changes in pixel values between neighboring pixels. This loss function combines the total variations in both vertical and horizontal directions to reduce artifacts and improve image quality. The loss is formulated as:

$$L_{\text{illumination}} = \frac{1}{(H-1)\cdot W}\sum_{i=1}^{H-1}\sum_{j=1}^{W}\left(X_{(i+1,j)} - X_{(i,j)}\right)^2 + \frac{1}{H\cdot(W-1)}\sum_{i=1}^{H}\sum_{j=1}^{W-1}\left(X_{(i,j+1)} - X_{(i,j)}\right)^2 \tag{20}$$

This formulation ensures that both vertical and horizontal transitions are penalized appropriately, promoting smooth illumination changes and reducing abrupt intensity variations.

Spatial Consistency Loss

The **Spatial Consistency Loss,** L_{spatial}, ensures local structural consistency between the original and enhanced images. It is calculated by:

$$M_{\text{orig}} = \frac{1}{C}\sum_{c=1}^{C} Y_{\text{true},c}, \quad M_{\text{enh}} = \frac{1}{C}\sum_{c=1}^{C} Y_{\text{pred},c}. \tag{21}$$

where M_{orig} and M_{enh} are the mean channel values for the original and enhanced images, respectively and C is the number of channels. The loss compares gradients of these means in eight directions (left, right, up, down, and four diagonals):

$$L_{\text{spatial}} = \sum_{d}\left(G_{\text{orig},d} - G_{\text{enh},d}\right)^2. \tag{22}$$

where $G_{\text{orig},d}$ and $G_{\text{enh},d}$ represent gradients in the d-th direction for the original and enhanced images, respectively. Specifically, this loss measures deviations in local gradients, reflecting the consistency of features e.g., edges, textures. Minimizing these deviations helps maintain spatial coherence and reduces artifacts, leading to a more natural enhancement.

Total Loss

The Total Loss denoted as, L_{total}, is given by:

$$L_{\text{total}} = \lambda_{\text{illumination}}\cdot L_{\text{illumination}} + \lambda_{\text{spatial}}\cdot L_{\text{spatial}} + \lambda_{\text{color}}\cdot L_{\text{color}} + \lambda_{\text{exposure}}\cdot L_{\text{exposure}}, \tag{23}$$

with weights $\lambda_{\text{illumination}} = 200$, $\lambda_{\text{spatial}} = 1$, $\lambda_{\text{color}} = 5$, and $\lambda_{\text{exposure}} = 10$, ensuring balanced contributions for optimal enhancement.

4 Results

We evaluated GlowNet qualitatively and quantitatively on two datasets: LOL [15] having 500 low-light images, and Dark Face [13] with 6000 low-light facial images. Images were resized to 256x256 patches, trained with a batch size of 16 using the Adam optimizer and a specified learning rate of 1e-4 for 300 epochs, and halved every 100 epochs. Training was performed on an NVIDIA® V100 Tensor Core GPU, processing each batch in 407 ms.

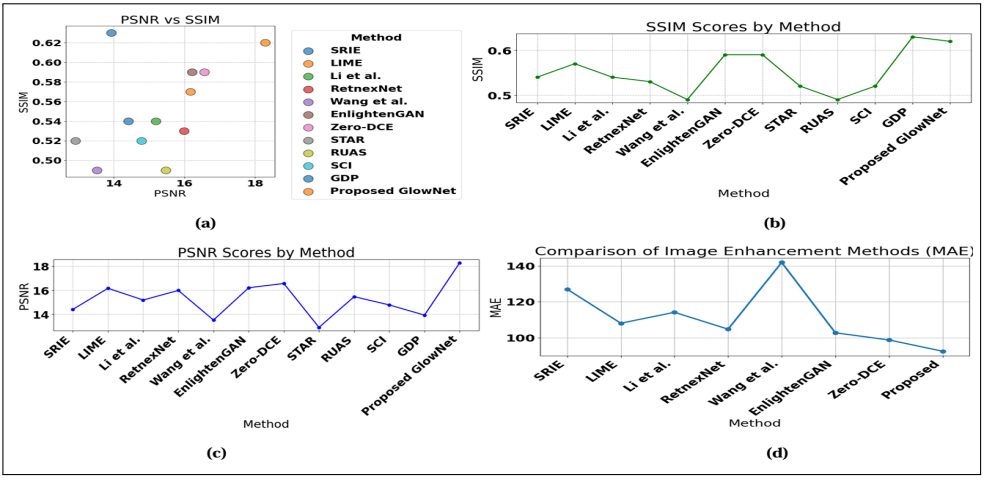


Figure 3: Visual Analysis on LOL [15] dataset: (a) Combined effect of PSNR and SSIM on different methods. Performance comparison of GlowNet against SOTA methods on (b) SSIM, (c) PSNR, and (d) MAE metrics.

4.1 Quantitative and Qualitative Analysis

To assess the effectiveness, GlowNet was compared with eleven SOTA methods using PSNR, SSIM, and MAE metrics (Figure 3). GlowNet achieved the highest PSNR of 18.29 dB, a competitive SSIM of 0.62, and the lowest MAE of 92.51 on the LOL [15] dataset, outperforming all other methods across these metrics in Table 1. In an indoor scene image (Figure 4), GlowNet improves the visibility of darker areas while maintaining the original color of the input image. In contrast, some existing methods over-smooth the details. Figure 4 demonstrates that while Zero-DCE [3] retains good detail, methods like SRIE [2], LIME [5], Wang et al. [14], and EnlightenGAN [6] fail to recover clear facial features and introduce artifacts. In contrast, GlowNet effectively preserves shadow details and maintains natural exposure.

Figure 5 also showcases the qualitative results achieved by GlowNet against other SOTA methods on the LoL [15] dataset. Figure 6 showcases GlowNet's proficiency in improving real world extreme low-light images clicked by us in natural dark conditions presented in the upper row of Figure 6. Glownet was able to enhance the object colors under extreme dark conditions for all three images. The lower row of Figure 6 presents complex facial detection task where GlowNet was able to enhance faces with proper fidelity as depicted in Figure 6 on the Dark Face [13] dataset, effectively revealing previously unseen faces and enhancing overall image clarity.

4.2 Ablation Study

To evaluate GlowNet's key components, an ablation study was performed by removing the Multi-Scale Feature Block (MFB) and Self-Attention (SA), independently and together. As shown in Table 2 and Figure 7, both MFB and SA significantly enhance image illumination, with their combination achieving the best overall performance. Additionally, kernel size (ks) and feature maps (fm) were varied. Optimal results were achieved with ks = 3 and fm = 32.

Method	PSNR ↑	SSIM ↑	MAE ↓	Conference/Journal
SRIE [2]	14.41	0.54	127.08	CVPR, 2016
LIME [5]	16.17	0.57	108.12	IEEE TIP, 2016
Li et al. [7]	15.19	0.54	114.21	IEEE TIP, 2018
Retinex-Net [15]	15.99	0.53	104.81	BMVC, 2018
Wang et al. [14]	13.52	0.49	142.01	CVPR, 2019
EnlightenGAN [6]	16.21	0.59	102.78	IEEE TIP, 2021
Zero-DCE [3]	16.57	0.59	98.78	CVPR, 2020
STAR [16]	12.91	0.52	–	IEEE TIP, 2020
RUAS [8]	15.47	0.49	–	CVPR, 2021
SCI [11]	14.78	0.52	–	CVPR, 2022
GDP [1]	13.93	**0.63**	–	CVPR, 2023
Proposed GlowNet	**18.29**	0.62	**92.51**	—

Table 1: Performance analysis and comparison of GlowNet against other SOTA methods in terms of PSNR, SSIM, and MAE on LOL [15] dataset.

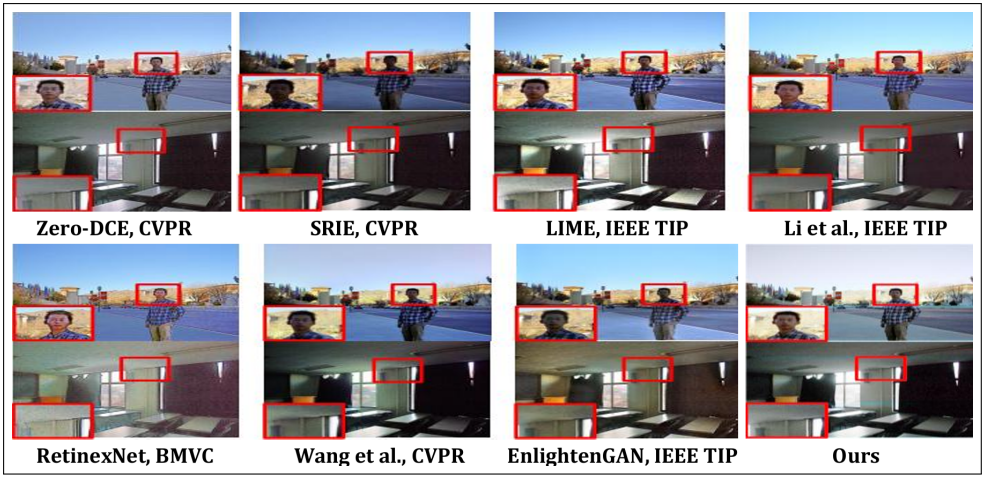


Figure 4: Qualitative comparison of low-light enhancement results on facial images and room interiors. The comparison highlights the effectiveness of the enhancement techniques in preserving fine details, textures, and overall image quality across different types of scenes.

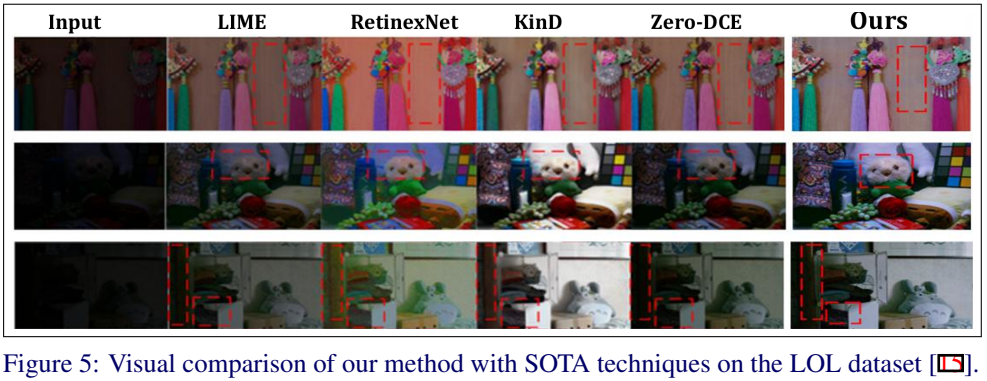


Figure 5: Visual comparison of our method with SOTA techniques on the LOL dataset [15].

Figure 6: Qualitative Analysis: Applying GlowNet on extreme low-light images (Upper row) and Dark Face [13] dataset for facial recognition in dark (Lower row).

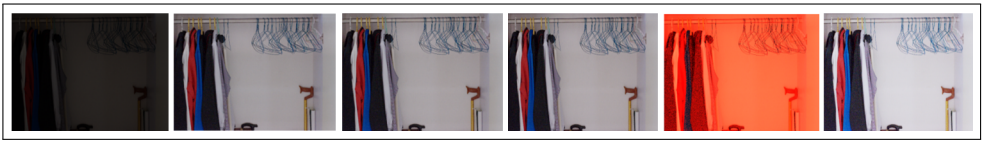


Figure 7: Images with varying epochs and loss functions: From left to right: (1) Input image, (2) After 1 epoch of training, (3) After 10 epochs, (4) After 100 epochs, (5) Without weighted color difference loss, (6) Optimized image after 300 epochs with the total loss.

Components	**PSNR** ↑	**SSIM** ↑	**MAE** ↓
Proposed – (MFB+SA) (ks=3, fm=32)	17.64	0.56	96.32
Proposed – MFB (ks=3, fm=32)	17.84	0.58	93.23
Proposed – SA (ks=3, fm=32)	18.00	0.57	94.55
Proposed with (ks=3, fm=64)	18.00	0.61	92.45
Proposed with (ks=3, fm=128)	18.00	0.60	92.42
Proposed with (ks=5, fm=24)	17.97	0.60	92.38
Proposed with (ks=5, fm=48)	17.99	0.60	92.58
Proposed with (optimal) (ks=3, fm=32)	**18.29**	**0.62**	**92.51**

Table 2: Results of ablation study showing the impact of different components and configurations on model's performance. Kernel size (ks) and feature maps (fm) are specified for each configuration. PSNR, SSIM, and MAE metrics are reported.

5 Conclusions and Future Directions

GlowNet, a lightweight attention-guided deep network for low-light image restoration, employs channel, spatial, and self-attention mechanisms along with dilated convolutions for effective contextual information capture. Guided by four loss functions: color constancy, exposure, illumination smoothness, and spatial consistency—it enhances image quality significantly. Extensive experiments, including an ablation study, demonstrate its superiority. With only 0.24 million parameters, it operates without paired or unpaired supervision, ensuring broad applicability in diverse real-world scenarios such as low-light photography, surveillance video enhancement, and medical imaging. However, the present implementation does exhibit some minor artifacts for extremely dark images, suggesting areas for potential enhancement. Future research could focus on refining the attention mechanisms to better capture intricate details and the addition of structure priors for improved realism.

Acknowledgement

A part of this research work has received support from the IDEAS- Institute of Data Engineering, Analytics, and Science Foundation, The Technology Innovation Hub, Indian Statistical Institute, Kolkata under Project No /ISI/TIH/2022/55/ dated September 13, 2022.

References

[1] B. Fei, Z. Lyu, L. Pan, J. Zhang, W. Yang, T. Luo, B. Zhang, and B. Dai. Generative diffusion prior for unified image restoration and enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9935–9946, 2023.

[2] X. Fu, D. Zeng, Y. Huang, X. P. Zhang, and X. 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.

[3] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1780–1789, 2020.

[4] M. H. Guo, T. X. Xu, J. J. Liu, Z. N. Liu, P. T. Jiang, T. J. Mu, and S. M. Hu. Attention mechanisms in computer vision: A survey. *Computational Visual Media*, 8(3):331–368, 2022.

[5] X. Guo, Y. Li, and H. Ling. LIME: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 26(2):982–993, 2016.

[6] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, and Z. Wang. EnlightenGAN: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing*, 30:2340–2349, 2021.

[7] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo. Structure-revealing low-light image enhancement via robust Retinex model. *IEEE Transactions on Image Processing*, 27 (6):2828–2841, 2018.

[8] R. Liu, L. Ma, J. Zhang, X. Fan, and Z. Luo. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10561–10570, 2021.

[9] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi. A survey of deep neural network architectures and their applications. *Neurocomputing*, 234:11–26, 2017.

[10] K. G. Lore, A. Akintayo, and S. Sarkar. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017.

[11] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5637–5646, 2022.

[12] A. Mondal, S. Ghosh, and A. Ghosh. Partially camouflaged object tracking using modified probabilistic neural network and fuzzy energy-based active contour. *International Journal of Computer Vision*, 122:116–148, 2017.

[13] U. al Qura. Dark face dataset dataset. https://universe.roboflow.com/umm-al-qura-erskc/dark-face-dataset, September 2023. Roboflow Universe.

[14] R. Wang, Q. Zhang, C. W. Fu, X. Shen, W. S. Zheng, and J. Jia. Underexposed photo enhancement using deep illumination estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6849–6857, 2019.

[15] C. Wei, W. Wang, W. Yang, and J. Liu. Deep Retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. Published in British Machine Vision Conference, 2018.

[16] J. Xu, Y. Hou, D. Ren, L. Liu, F. Zhu, M. Yu, and L. Shao. STAR: A structure and texture aware Retinex model. *IEEE Transactions on Image Processing*, 29:5022–5037, 2020.

[17] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M. H. Yang, and L. Shao. Learning enriched features for fast image restoration and enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2):1934–1948, 2022.

[18] Y. Zhang, J. Zhang, and X. Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 1632–1640, October 2019.