Mr. T. Sreenivasu, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science,
Engineering and Technology
An Open Access Journal

Attention-Driven Low Light Image Enhancement Using lightweight CNN

Mr. T. Sreenivasu. M.Tech (Ph. D.) Sr. Assistant Professor
L. Vishnu Vardhan, Sk.Sayyad Baji, P.Yesuratnam, CH.Bhaskar, V.Prasanna Kumar

Abstract- Low-light image enhancement is an important task in computer vision that aims to improve visibility and preserve critical details in images captured under poor lighting conditions. In the base paper, a CNN-based method using a simple encoder–decoder architecture is employed for image enhancement. Although this approach effectively increases overall brightness, it presents several limitations. The model treats all pixels equally without prioritizing important regions, lacks an attention mechanism to focus on semantically significant features such as faces or text, and struggles to generalize effectively across diverse real-world lighting conditions. To address these limitations, this work proposes an enhanced Low-Light Image Enhancement (LLIE) model that integrates the Convolutional Block Attention Module (CBAM) into the encoder–decoder network. CBAM introduces both Channel Attention, which prioritizes important feature channels, and Spatial Attention, which focuses enhancement on key regions of the image.By incorporating these attention mechanisms, the proposed model improves brightness and clarity while preserving important structural details and features. As a result, the enhanced system produces higher visual quality images and becomes more suitable for practical applications such as surveillance systems, digital photography, and mobile vision applications.

Keywords: Low-Light Image Enhancement, CBAM, Channel Attention, Spatial Attention.

I. INTRODUCTION

1.1 Background of Low-Light Image Enhancement

Images captured under low-illumination conditions frequently suffer from poor visibility, reduced contrast, color distortion, and loss of important details. These limitations negatively affect the performance of computer vision systems used in applications such as surveillance, autonomous driving, smartphone photography, and security monitoring. Therefore, improving low-light images is an essential preprocessing step for many vision-based technologies.

Traditional image enhancement techniques typically rely on methods such as histogram equalization, gamma correction, and Retinex-based algorithms. Although these techniques can partially improve brightness and contrast, they often introduce problems such as noise amplification, excessive enhancement, and color inconsistency. These drawbacks limit their effectiveness in real-world environments where illumination conditions vary significantly.

With the rapid progress of deep learning, Convolutional Neural Networks (CNNs) have become powerful tools for image enhancement tasks. CNN-based [1]models can automatically learn complex visual features from training data and generally provide better performance than conventional approaches. Among these models, encoder–decoder architectures are widely adopted because they are capable of capturing both global contextual information and fine local details.

However, many existing CNN-based enhancement models process all image regions uniformly and do not prioritize important visual features such as edges, textures, faces, or textual information. As a result, even though brightness may improve, important semantic information may not be effectively preserved.

1.2 Challenges in Low-Light Imaging

Low-light images[2], [3], [4], [5] introduce multiple challenges for image processing systems. Poor illumination [6]often leads to reduced visibility, weak contrast, and increased image noise. In such conditions, significant visual details can become obscured or distorted due to insufficient lighting.

Furthermore, many enhancement techniques increase brightness at the cost of amplifying noise, which further degrades the visual quality of the image. These challenges make it difficult for computer vision algorithms to perform tasks such as object detection, facial recognition, and scene interpretation effectively.

1.3 Motivation

The primary motivation of this research is to design an intelligent enhancement model capable of concentrating on the most informative regions and features within an image. Attention mechanisms in deep learning have demonstrated strong potential in improving feature representation by highlighting relevant information while suppressing less useful data.

By incorporating an attention mechanism into the image enhancement framework, the model can selectively enhance important image features while minimizing unnecessary processing of irrelevant areas.

1.4 Objectives of the Proposed Work

The main goal of this study is to develop an improved low-light image enhancement model based on a CNN encoder–decoder architecture combined with the Convolutional Block Attention Module. The proposed system aims to:

- Increase image brightness
- Improve contrast levels
- Preserve essential image details
- Produce visually pleasing outputs suitable for real-world applications

II. RELATED WORK

Low-light image enhancement [7] has been extensively explored in the fields of image processing and computer vision. Earlier approaches mainly relied on conventional image processing techniques such as histogram equalization[8], [9], gamma correction, and Retinex-based models[10], [11]. These methods attempt to modify brightness and contrast but often struggle to handle complex illumination variations.

Histogram equalization enhances contrast by redistributing the intensity values of pixels across the image. However, this technique can sometimes produce over-enhanced images and unnatural color appearances. Retinex-based[5], [9], [10], [11], [12] methods attempt to imitate the human visual system by separating illumination from reflectance components in an image. Although these approaches generally produce better results than basic techniques, they may still introduce artifacts and noise.

Recent progress in deep learning [5], [10], [13], [14]has led to the development of CNN-based image enhancement models[15]. Encoder–decoder networks are commonly used to learn the transformation from low-light images to well-illuminated images. These networks extract high-level features from the input image and reconstruct an improved output image. Despite their advantages, many CNN-based methods still lack mechanisms that allow them to focus on the most important areas of an image.

Attention mechanisms have recently been incorporated into deep learning models to address this limitation. The Convolutional Block Attention Module (CBAM) is an effective attention module that sequentially applies channel attention and spatial attention [16], [17], [18]to refine feature maps. Integrating CBAM into enhancement networks enables better feature representation and allows the model to concentrate on important image regions.

III. SYSTEM ARCHITECTURE

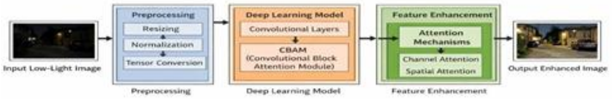

Fig 1: System Flow

The proposed Low-Light Image Enhancement (LLIE) system aims to improve the visibility and quality of images captured under poor lighting conditions. The architecture is built upon a CNN-based encoder–decoder [19], [20]network integrated with the Convolutional Block Attention Module (CBAM). The encoder extracts hierarchical feature representations from the input image, while the decoder reconstructs the enhanced image with improved illumination and clarity.

Unlike conventional CNN-based enhancement models [21] that treat all pixels equally, the proposed framework employs attention mechanisms to emphasize important features and regions within the image. The CBAM module contains two sequential components: Channel Attention and Spatial Attention.

Channel Attention focuses on identifying the most informative feature channels, while Spatial Attention highlights the most relevant spatial regions of the image. Through these mechanisms, the model can prioritize areas such as faces, text, and objects while reducing the influence of irrelevant noise. This selective enhancement leads to improved contrast, preserved structural information, and better visual quality.

3.1 Encoder–Decoder Network

The encoder–decoder structure serves as the core component of the proposed LLIE model. The encoder extracts feature representations from the input low-light image using multiple convolutional layers. Each layer captures different types of information such as edges, textures, and semantic patterns.

As the image moves through the encoder, the spatial dimensions gradually decrease while the number of feature channels increases[19]. This process enables the network to capture high-level contextual information required for accurate enhancement.

The decoder reconstructs the enhanced image using up-sampling layers[5], [12], [22] combined with convolution operations. These layers gradually restore the spatial resolution of the image. Skip connections between the encoder and decoder are also employed to preserve important low-level features that contribute to fine details in the output image.

This architecture allows efficient feature extraction and reconstruction, making it suitable for real-time enhancement applications.

3.2 CBAM Attention Mechanism

To improve the performance of the basic encoder–decoder network, the Convolutional Block Attention Module (CBAM) [7], [16], [20], [21]is integrated into the architecture. CBAM introduces attention mechanisms that allow the network to focus on the most informative features during the enhancement process.

The CBAM module consists of two components: Channel Attention and Spatial Attention. Channel Attention works by analyzing the relationships between different feature channels. It assigns higher weights to the channels that contain more useful information for image enhancement. This helps the model prioritize important features while reducing the influence of less relevant ones.

Spatial Attention focuses on identifying important regions within the image. It generates an attention map that highlights areas that require more enhancement, such as objects, faces, and text. By applying spatial attention, the model ensures that enhancement is concentrated on meaningful parts of the image rather than uniformly across all pixels.
The combination of channel and spatial attention significantly improves feature representation, leading to better brightness adjustment, contrast

enhancement, and detail preservation in low-light images.

IV. IMPLEMENTATION DETAILS

The proposed Low-Light Image Enhancement model [23] is implemented using a deep learning framework based on convolutional neural networks integrated with the Convolutional Block Attention Module (CBAM)[13]. The system processes low-light images[3], [8], [12], [18], [22], [24] and generates enhanced outputs with improved brightness and preserved structural details.

The implementation involves dataset preparation, preprocessing, model training, and optimization procedures.

Hyperparameter	Value	Description
Batch size	8	Number of images processed in one training iteration
Learning Rate	0.0001	Controls how quickly model parameters are updated
Epochs	50	Number of complete training cycles.
Optimizer	Adam	Used to minimize the loss function
Image Resolution	256x256	Size for input images used for training

Table 1:Parameter and its values.

4.1 Dataset Preparation

The performance of the proposed Low-Light Image Enhancement[25] model depends on the quality of the training dataset. In this work, the LOL (Low-Light) Dataset is used for training and evaluation[11]. The LOL dataset contains paired images consisting of low-light input images and their corresponding normal-light reference images. These image pairs enable the neural network to learn the mapping between poorly illuminated images and their enhanced versions.

Dataset Name	Training Images	Testing Images	Image Type	Description
LOL Dataset	485	15	Paired Images	Contains low light and ground truth images.+

Table 2: Dataset description

Before feeding the images into the network, all images are resized to a fixed spatial resolution to maintain uniformity in input size. This preprocessing step helps in reducing computational complexity and improves training efficiency.

Let the input low-light image be represented as:

$$I_{\{low\}} \in \backslash mathbb\{R\}^{\{H \times W \times C\}}$$

where
H represents the height of the image,
W represents the width of the image, and
C represents the number of color channels (C = 3 for RGB images).

The objective of the enhancement network is to transform the low-light image into an enhanced output image with improved brightness and clarity. This transformation can be represented as:

$$I_{\{enhanced\}} = F(I_{\{low\}};\ \theta)$$

where
F represents the enhancement function implemented by the CNN encoder–decoder architecture integrated with the CBAM attention module, and θ represents the learnable parameters of the network.

During the training process, the network learns the optimal parameters by minimizing the difference between the predicted enhanced image and the corresponding ground truth image. This allows the model to effectively restore illumination details and improve the overall visual quality of low-light images.

4.2 Image Preprocessing

Before feeding the images into the enhancement network[11], [17], [19], [21], several preprocessing steps are applied to improve training stability and model performance. Image preprocessing ensures that the input data is consistent and suitable for the deep learning model.

The first step is image normalization, where pixel intensity values are scaled to a smaller range. Normally, image pixel values range from 0 to 255. These values are converted into the range 0 to 1, which helps the neural network learn more efficiently and speeds up the convergence during training.

The normalization process can be represented as:

$$I_{norm} = \frac{I}{255}$$

where
I = original image pixel value I_norm = normalized pixel value.

In addition to normalization, images are resized to a fixed dimension before being used as input to the neural network. This ensures that all images have the same spatial resolution during training.

Let the resized image be represented as:

$$I_{resized} \in \mathbb{R}^{H \times W \times C}$$

where
H = image height W = image width
C = number of color channels.
These preprocessing steps help improve the learning capability of the model and ensure consistent input representation for the encoder–decoder network with CBAM attention modules.

Step	Operation	Purpose
1.	Image Resizing	Convert all images to the same resolution
2.	Normalization	Scale pixel values between 0 and 1
3.	Tensor Conversion	Convert images into tensor format for neural network processing

Table 3:Steps in Preprocessing.

4.3 Loss Function

The loss function is used to measure the difference between the enhanced image produced by the network and the corresponding ground truth image. During training, the model adjusts its parameters in order to minimize this difference and produce more accurate enhancement results.

In this work, the Mean Squared Error (MSE) loss function is used to evaluate the reconstruction quality of the enhanced image. The MSE loss calculates the average squared difference between the predicted pixel values and the ground truth pixel values.

The loss function can be expressed as:

$$L_{MSE} = \frac{1}{N} \sum_{i=1}^{N} (I_{enhanced} - I_{gt})^2$$

Where

L_MSE represents the Mean Squared Error loss,
N represents the total number of pixels in the image, I_enhanced represents the enhanced image generated by the model, and I_gt represents the corresponding ground truth image.

V. RESULTS

Fig2:Outputs

5.1 Evaluation Metrics

To evaluate the performance of the proposed low-light image enhancement model, commonly used image quality metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) are used. These metrics help measure the similarity between the enhanced image and the corresponding ground truth image.

PSNR measures the reconstruction quality of the enhanced image, while SSIM evaluates the structural similarity between two images

5.2 PSNR Analysis

Peak Signal-to-Noise Ratio (PSNR) is used to evaluate the quality of the enhanced image by comparing it with the reference image. A higher PSNR value indicates better image reconstruction and enhancement quality.

The PSNR value is calculated based on the Mean Squared Error between the enhanced image and the ground truth image. Higher PSNR values indicate that the enhanced image is closer to the original reference image

5.3 SSIM Analysis

Structural Similarity Index (SSIM) is used to measure the similarity between the enhanced image and the reference image in terms of structure, brightness, and contrast.

SSIM values range between 0 and 1. A value closer to 1 indicates that the enhanced image has high structural similarity with the ground truth image.

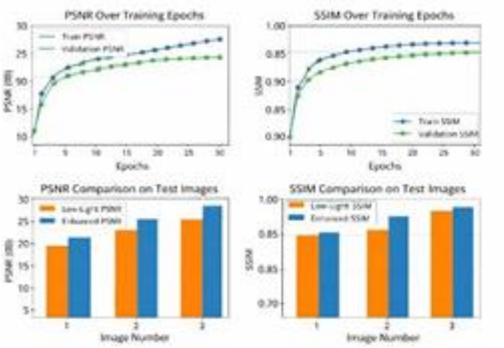

Fig 3: Graphical Representation of Metric Values

The proposed CBAM-based enhancement model produces higher SSIM values compared to conventional CNN-based methods, indicating better preservation of image structures.

PSNR Value	SSIM Value		Enhancement Rate%
22.2343	0.8865		255.09%

Table 4 :Metric values.

VI. DISCUSSION

The experimental results demonstrate that integrating the Convolutional Block Attention Module with the CNN encoder–decoder architecture improves the performance of low-light image enhancement.

The channel attention mechanism helps the model focus on important feature channels, while the spatial attention mechanism highlights important regions in the image. This selective attention improves brightness correction and detail preservation in low-light images.

The proposed model produces enhanced images with better visibility, reduced noise, and improved contrast. These improvements make the model suitable for real-world applications such as surveillance systems, mobile photography, and computer vision tasks.

VII. CONCLUSION

In this paper, a CBAM-based Low-Light Image Enhancement model was proposed to improve the visibility of images captured under poor lighting conditions. The proposed method integrates channel attention and spatial attention mechanisms into a CNN encoder–decoder architecture to enhance important image features.

Experimental evaluation shows that the proposed method improves brightness, contrast, and structural details in low-light images. The use of attention mechanisms helps the model focus on important regions of the image, resulting in better visual quality.

The proposed approach can be applied in various real-world applications such as surveillance, photography, and mobile vision systems. Future work may focus on improving the model's efficiency and exploring advanced attention mechanisms for further enhancement.

REFERENCES

1. M. M. Ghazi and H. K. Ekenel, "A Comprehensive Analysis of Deep Learning Based Representation for Face Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 102–109. doi: 10.1109/CVPRW.2016.20.
2. M. J. C. S. Reis, "Low-Light Image Enhancement Using Deep Learning: A Lightweight Network with Synthetic and Benchmark Dataset Evaluation," Appl. Sci., vol. 15, no. 11, p. 6330, Jun. 2025, doi: 10.3390/app15116330.
3. H. Tang, H. Zhu, L. Fei, T. Wang, Y. Cao, and C. Xie, "Low-Illumination Image Enhancement Based on Deep Learning Techniques: A Brief Review," Photonics, vol. 10, no. 2, p. 198, Feb. 2023, doi: 10.3390/photonics10020198.
4. C. Guo et al., "Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA: IEEE, Jun. 2020, pp. 1777–1786. doi: 10.1109/CVPR42600.2020.00185.
5. K. G. Lore, A. Akintayo, and S. Sarkar, "LLNet: A Deep Autoencoder Approach to Natural Low-light Image Enhancement," Apr. 15, 2016, arXiv: arXiv:1511.03995. doi: 10.48550/arXiv.1511.03995.
6. C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to See in the Dark," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT: IEEE, Jun. 2018, pp. 3291–3300. doi: 10.1109/CVPR.2018.00347.
7. Y. Jiang, J. Zhu, L. Li, and H. Ma, "A Joint Network for Low-Light Image Enhancement Based on Retinex," Cogn. Comput., vol. 16, no. 6, pp. 3241–3259, Nov. 2024, doi: 10.1007/s12559-024-10347-4.
8. A. M. Reza, "Realization of the Contrast Limited Adaptive Histogram Equalization (CLAHE) for Real-Time Image Enhancement," J. VLSI Signal Process. Syst. Signal Image Video Technol., vol. 38, no. 1, pp. 35–44, Aug. 2004, doi: 10.1023/B:VLSI.0000028532.53893.82.
9. X. Ren, W. Yang, W.-H. Cheng, and J. Liu, "LR3M: Robust Low-Light Enhancement via Low-Rank Regularized Retinex Model," IEEE Trans. Image Process., vol. 29, pp. 5862–5876, 2020, doi: 10.1109/TIP.2020.2984098.
10. M.-T. Duong, S. Lee, and M.-C. Hong, "DMT-Net: Deep Multiple Networks for Low-Light Image Enhancement Based on Retinex Model," IEEE Access, vol. 11, pp. 132147–132161, 2023, doi: 10.1109/ACCESS.2023.3336411.
11. C. Wei, W. Wang, W. Yang, and J. Liu, "Deep Retinex Decomposition for Low-Light Enhancement," Aug. 14, 2018, arXiv: arXiv:1808.04560. doi: 10.48550/arXiv.1808.04560.
12. M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, "Structure-Revealing Low-Light Image Enhancement Via Robust Retinex Model," IEEE Trans. Image Process., vol. 27, no. 6, pp. 2828–2841, Jun. 2018, doi: 10.1109/TIP.2018.2810539.
13. Z. H. Qaisar, R. Khan, A. Mehmood, A. Khan, and M. M. Ibrahim, "A deep adaptive framework for low-light image enhancement in adverse lighting conditions," Vis. Comput., vol. 41, no. 14, pp.

12109–12124, Nov. 2025, doi: 10.1007/s00371-025-04147-6.

14. T. Matsui and M. Ikehara, "Low-Light Image Enhancement Using a Simple Network Structure," IEEE Access, vol. 11, pp. 65507–65516, 2023, doi: 10.1109/ACCESS.2023.3290490.
15. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," May 18, 2015, arXiv: arXiv:1505.04597. doi: 10.48550/arXiv.1505.04597.
16. S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in Computer Vision – ECCV 2018, vol. 11211, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., in Lecture Notes in Computer Science, vol. 11211. , Cham: Springer International Publishing, 2018, pp. 3–19. doi: 10.1007/978-3-030-01234-2_1.
17. T. Celik and T. Tjahjadi, "Contextual and Variational Contrast Enhancement," IEEE Trans. Image Process., vol. 20, no. 12, pp. 3431–3441, Dec. 2011, doi: 10.1109/TIP.2011.2157513.
18. X. Guo, "LIME: A Method for Low-light IMage Enhancement," in Proceedings of the 24th ACM international conference on Multimedia, Amsterdam The Netherlands: ACM, Oct. 2016, pp. 87–91. doi: 10.1145/2964284.2967188.
19. R. Yasrab, N. Gu, and X. Zhang, "An Encoder-Decoder Based Convolution Neural Network (CNN) for Future Advanced Driver Assistance System (ADAS)," Appl. Sci., vol. 7, no. 4, p. 312, Mar. 2017, doi: 10.3390/app7040312.
20. Y. Ji, H. Zhang, Z. Zhang, and M. Liu, "CNN-based encoder-decoder networks for salient object detection: A comprehensive review and recent advances," Inf. Sci., vol. 546, pp. 835–857, Feb. 2021, doi: 10.1016/j.ins.2020.09.003.
21. J. Wang and Y. Hu, "An Improved Enhancement Algorithm Based on CNN Applicable for Weak Contrast Images," IEEE Access, vol. 8, pp. 8459–8476, 2020, doi: 10.1109/ACCESS.2019.2963478.
22. S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness Preserved Enhancement Algorithm for Non-Uniform Illumination Images," IEEE Trans. Image Process., vol. 22, no. 9, pp. 3538–3548, Sep. 2013, doi: 10.1109/TIP.2013.2261309.
23. W. Wang, X. Wu, X. Yuan, and Z. Gao, "An Experiment-Based Review of Low-Light Image Enhancement Methods," IEEE Access, vol. 8, pp. 87884–87917, 2020, doi: 10.1109/ACCESS.2020.2992749.
24. Z. Cui, W. Wang, Z. Xu, and C. Wang, "A Low-light Image Enhancement Method Balancing Brightness Boosting and Noise Smoothing," in Proceedings of the 2022 6th International Conference on Computer Science and Artificial Intelligence, Beijing China: ACM, Dec. 2022, pp. 49–54. doi: 10.1145/3577530.3577538.
25. Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, Pacific Grove, CA, USA: IEEE, 2003, pp. 1398–1402. doi: 10.1109/ACSSC.2003.1292216.