Low Light Image Enhancement

Khalid Bin Mahamud Bhuiya, Md. Ashfaq Bin Hoque, Md. Rifat Al Mostakim, Tafsirul Islam Shafin

Computer Science & Engineering, American International University-Bangladesh, Dhaka, Bangladesh

Abstract

Low-light image enhancement presents significant challenges in image processing, particularly in preserving visibility, details, and color accuracy. This paper introduces an advanced Generative Adversarial Network (GAN) model that addresses these limitations through innovative architectural design, including enhanced Residual block and Attention Module. The proposed U-Net based generator with a novel PatchGAN discriminator Structure leverages enhanced residual blocks and channel-spatial attention mechanisms to reconstruct high-quality images. Experimental validation on multiple datasets demonstrates superior performance, achieving a Peak Signal-to-Noise Ratio of 32.21 dB, a Structural Similarity Index of 0.9678, and a low Learned Perceptual Image Patch Similarity of 0.0172, thereby outperforming current state-of-the-art low-light image enhancement techniques.

I. Introduction

Low-light image enhancement is a crucial field in image processing and computer vision, addressing challenges such as poor visibility, loss of details, and color distortion. Traditional methods such as histogram equalization and Retinex-based techniques have long provided the foundation for enhancing images captured under low-light conditions. For instance, Pan et al. (2022) introduced a Retinex-based method that employs a neighborhood-weighted illumination map estimation and fusion strategy to effectively balance brightness and contrast [1]. Similarly, Han et al. (2023) demonstrated that combining Retinex theory with histogram equalization using frequency-domain fusion can significantly improve the quality of road images captured under low illumination [2]. Despite these advances, traditional approaches often struggle with noise sensitivity, over-enhancement, or loss of local detail. In recent years, deep learning particularly Generative Adversarial Networks (GANs) has emerged as a powerful alternative for overcoming these limitations.

II. Related Work

A. Traditional Approaches to Low-Light Image Enhancement

Traditional methods for low-light image enhancement include both histogram-based and Retinex-based approaches. Histogram equalization improves overall image contrast by redistributing pixel values but can lead to local contrast imbalances. Retinex-based methods decompose images into reflectance and illumination components to restore natural lighting. For example, Pan et al. (2022) proposed an enhancement framework that first estimates an illumination map via a neighborhood-weighted approach, then uses a combination of Gamma Correction and a Sine function to adaptively enhance brightness while preserving details [1]. In parallel, Han et al. (2023) addressed low-illumination challenges in road images by fusing the strengths of Retinex-based methods (for edge and texture preservation) with histogram equalization (for improved illumination), leveraging a frequency-domain fusion technique based on the Discrete Cosine Transform (DCT) [2].

B. Deep Learning-Based Methods

With the advent of deep learning, convolutional neural networks (CNNs) have set new performance benchmarks in low-light image enhancement. Recent works have introduced end-to-end Retinex networks and illumination estimation techniques that learn to adjust and correct image brightness from data. However, even these methods sometimes encounter issues such as local overexposure and the loss of fine texture details.

C. Generative Adversarial Networks for Low-Light Image Enhancement

Generative Adversarial Networks (GANs) have emerged as a robust tool for enhancing low-light images by leveraging adversarial training to produce realistic, high-quality outputs. Several GAN-based approaches have been proposed in recent years. For instance, Zero-DCE and Zero-DCE++ utilize deep curve estimation to dynamically adjust brightness levels, although they can sometimes introduce color distortion. CycleGAN and Cycle-Consistent GANs, designed for unpaired learning, enable image enhancement without strict dataset pairing, but may exhibit artifacts and overexposure in certain areas. Retinex-GANs integrate Retinex theory with GAN architectures to better preserve illumination details; however, these models can still face challenges in enhancing shadow regions. Progressive GANs, such as ProGAN, decompose reflections and utilize transmission networks to enhance nighttime images, though they occasionally produce undesirable color shifts. More recent hybrid networks fuse GANs with attention mechanisms and multi-scale feature extraction—building on the insights of traditional methods refine both global illumination and local details, marking a significant advancement in low-light image enhancement techniques [1] [2].

D. The Proposed Method

In this work, we present an advanced GAN-based model tailored for low-light image enhancement, incorporating several innovative components to significantly improve image quality. The model features a U-Net generator architecture, enhanced with an Efficient Residual Block structure, which facilitates robust feature extraction and propagation. The downsampling and upsampling processes are guided by the Enhanced Channel Spatial Attention module, which integrates both channel and spatial attention mechanisms, ensuring more focused learning of critical image features. Additionally, a Dual Attention Mechanism is applied through the integration of the Efficient Residual Blocks and the SE block, which optimizes both spatial and channel-wise features. To further refine the model's performance, the generator utilizes attention-guided skip connections, enabling the preservation of fine-grained details while promoting smooth image enhancement. The adversarial learning framework is supported by a PatchGAN discriminator that ensures high-quality image generation, distinguishing between real and enhanced low-light images. The combination of these advanced components leads to a robust model capable of enhancing low-light images with improved clarity, contrast, and detail.

E. Comparative Analysis and Performance Evaluation

Our GAN-based approach outperforms current state-of-the-art methods on both synthetic and real low-light image datasets. Specifically, it achieves higher Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM), which are indicative of improved image quality, while also demonstrating lower Learned Perceptual Image Patch Similarity (LPIPS), thereby reducing perceptual distortion.

III. Proposed Model

A. Proposed Generator Model

As illustrated in Fig. 1, our proposed model is based on a U-Net architecture, which consists of an encoder, a bottleneck, and a decoder. The encoder is responsible for extracting hierarchical features from the input image, while the decoder reconstructs the output image from these features. The bottleneck, located between the encoder and decoder, refines the extracted features using Efficient Residual Blocks with channel-spatial-sqeeze and excitation attention module [Fig. 3].

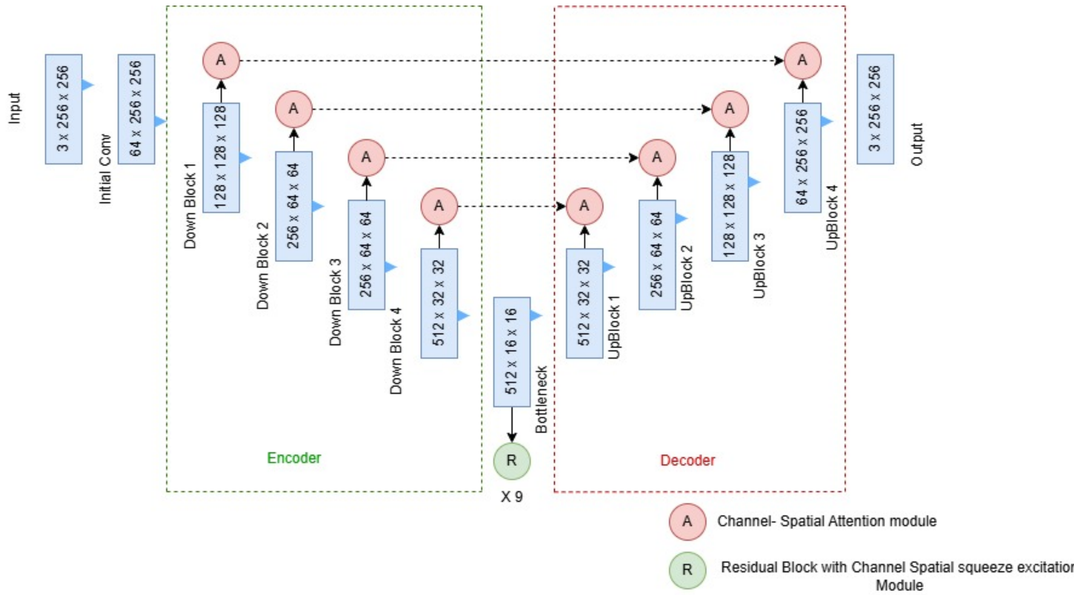

Fig. 1: Proposed Generator Model.

The encoder comprises four downsampling blocks, each of which reduces the spatial dimensions of the feature maps while increasing the number of channels. Each downsampling block includes a 3x3 convolutional layer with a stride of 2, followed by an instance normalization layer and a ReLU activation. Additionally, an Enhanced Channel-Spatial Attention Module (see Fig. 3) is applied to dynamically adjust the importance of different channels and spatial regions, allowing the network to focus on more informative features. The decoder consists of four upsampling blocks, each of which increases the spatial dimensions of the feature maps while reducing the number of channels. Each upsampling block includes a nearest-neighbor upsampling layer, a 3x3 convolutional layer, an instance normalization layer, and a ReLU activation. Skip connections between the encoder and decoder ensure that spatial and contextual information is preserved, enabling high-quality image reconstruction. The final output is produced by a 7x7 convolutional layer with a tanh activation function.

B. Proposed Residual Block

As shown in Fig. 2, the Efficient Residual Block is a key component of the bottleneck in our proposed model. It is designed to refine the feature maps extracted by the encoder while minimizing computational complexity. Each residual block consists of two main branches: The first branch includes depthwise separable convolutions, which reduce the number of parameters and computations while maintaining feature richness. This is followed by a 1x1 convolutional layer to combine the features. The second branch is a skip connection that adds the input feature map to the output of the first branch, ensuring that important information is preserved.

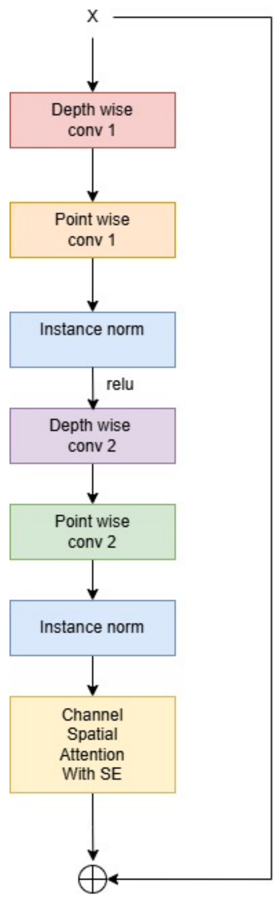

Fig. 2: Proposed Residual Block.

Additionally, each residual block incorporates an Enhanced Channel-Spatial Attention with Squeeze-and-Excitation (SE) Module (see Fig. 3) to further enhance feature representation. This module combines channel-wise and spatial attention mechanisms with squeeze-and-excitation operations, allowing the network to focus on the most relevant features.The use of residual blocks enables the model to capture long-range dependencies and refine the feature maps effectively, improving the overall performance of the network. The bottleneck consists of nine such residual blocks, which enable the model to capture long-range dependencies and refine the feature maps effectively, improving the overall performance of the network.

C. Proposed Attention Module

As depicted in Fig. 3, This module computes channel-wise attention weights using both average pooling and max pooling. The pooled features are passed through a fully connected layer with a reduction ratio of 16, and the resulting weights are applied to the input feature maps to emphasize important channels. Enhanced Spatial Attention module computes spatial attention weights by concatenating the average and max pooled features along the channel dimension, followed by a 7x7 convolutional layer. The resulting weights are applied to the input feature maps to highlight important spatial regions. Squeeze-and-Excitation (SE) Block enhances channel attention by explicitly modeling interdependencies between channels. It uses global average pooling to squeeze spatial dimensions and a fully connected layer to excite important channels.

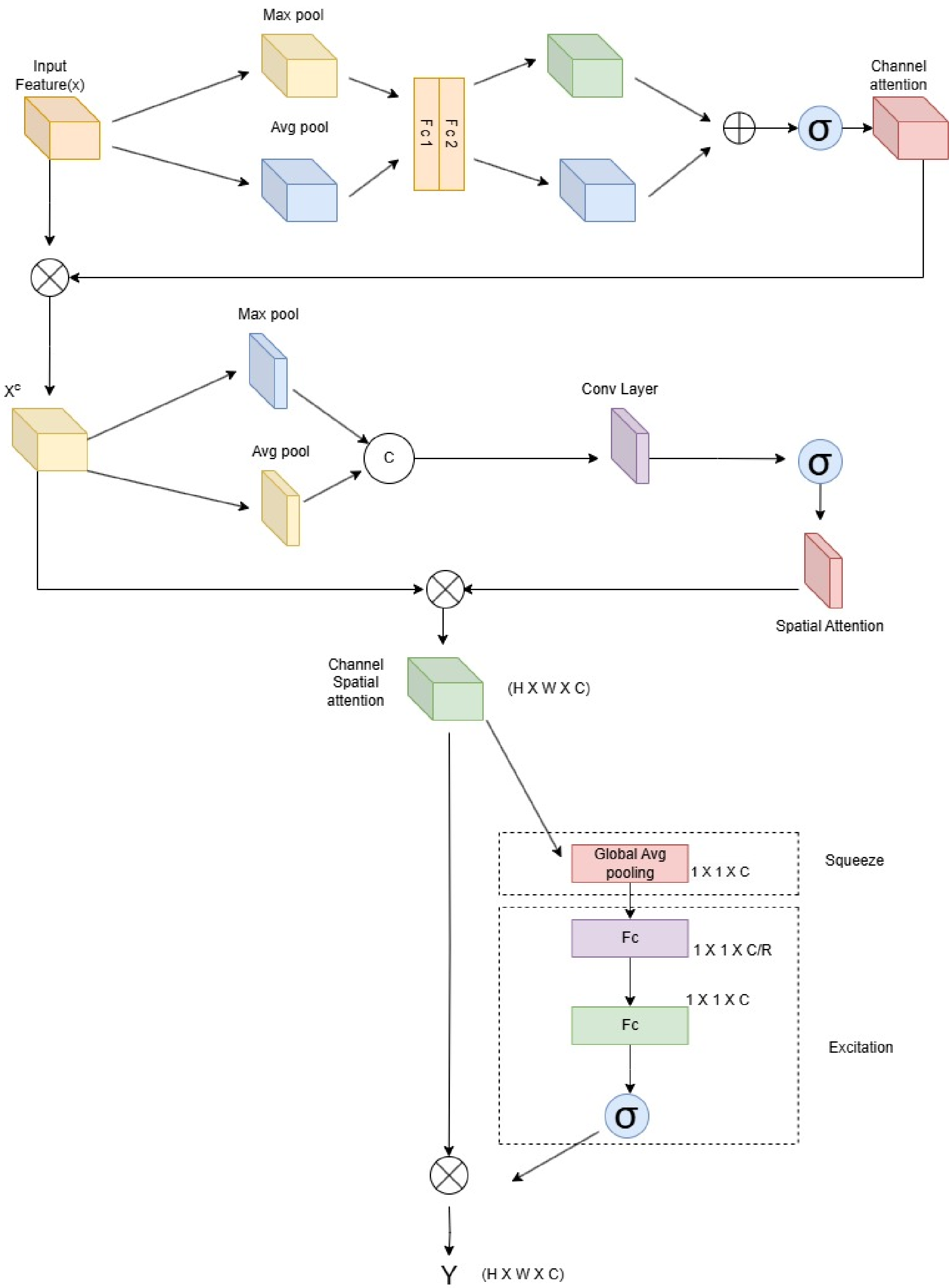

Fig. 3: Proposed Attention Module.

The outputs of the channel attention and spatial attention modules are combined through element-wise addition, and the result is further refined by the SE block. This combined attention mechanism allows the network to focus on both channel-wise and spatial-wise important features, improving its ability to handle complex image transformations.

D. Proposed Discriminator Model

As illustrated in Fig. 4, the proposed discriminator is based on a PatchGAN architecture, which classifies local image patches as real or fake rather than classifying the entire image. This patch-level approach allows the discriminator to focus on high-frequency details and local texture patterns, making it particularly effective for image-to-image translation tasks. The discriminator consists of several convolutional layers equipped with spectral normalization and instance normalization, which help stabilize training and improve generalization. Unlike traditional PatchGAN setups where both real and fake images are provided, our discriminator receives a single input image, simplifying the architecture while still enabling the discriminator to provide detailed feedback to the generator.

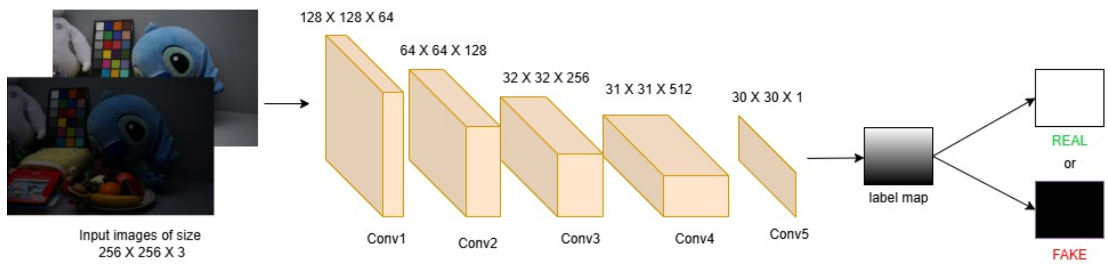

Fig. 4: Proposed Discriminator Model.

The first layer of the discriminator is a 4x4 convolutional layer with a stride of 2 and padding of 1, reducing the spatial dimensions of the input image by half while increasing the number of channels to 64. A LeakyReLU activation with a negative slope of 0.2 is applied to introduce non-linearity. The second layer is another 4x4 convolutional layer with a stride of 2 and padding of 1, increasing the number of channels to 128. An instance normalization layer is applied to stabilize the training process, followed by a LeakyReLU activation. The third layer is a 4x4 convolutional layer with a stride of 2 and padding of 1, increasing the number of channels to 256. Instance normalization and a LeakyReLU activation are applied to further refine the features. The fourth layer is a 4x4 convolutional layer with a stride of 1 and padding of 1, increasing the number of channels to 512. Instance normalization and LeakyReLU are again applied. The final layer is a 4x4 convolutional layer with a stride of 1 and padding of 1, reducing the number of channels to 1 and producing a single-channel output map. Each pixel in this output map corresponds to the classification of a local patch in the input image. A dropout layer with a dropout rate of 0.3 is applied to prevent overfitting.

All convolutional layers in the discriminator are equipped with spectral normalization, which constrains the Lipschitz constant of the discriminator. This ensures stable training and prevents the discriminator from becoming too powerful, which could otherwise hinder the generator's learning process. The use of spectral normalization, along with the progressively refined feature extraction through each layer, helps the discriminator provide detailed, local feedback to the generator, guiding it to produce more realistic images.

IV. Experiment

A. Dataset

Experiments were conducted using the LoLI-Street dataset (Islam et al., 2024) for low-light image enhancement (LLIE). The dataset comprises 30,000 training images and 3,000 validation images. Both the training and validation sets contain low-light street images and their corresponding ground truth high-light images as shown in 5. For our experiments, we selected a subset of 5,000 images for training. The validation dataset contains 3 levels of light intensity: light, moderate, and dense. For our experiments, we specifically worked with the dense dataset for validation. The number of dense images in the validation dataset is 500.

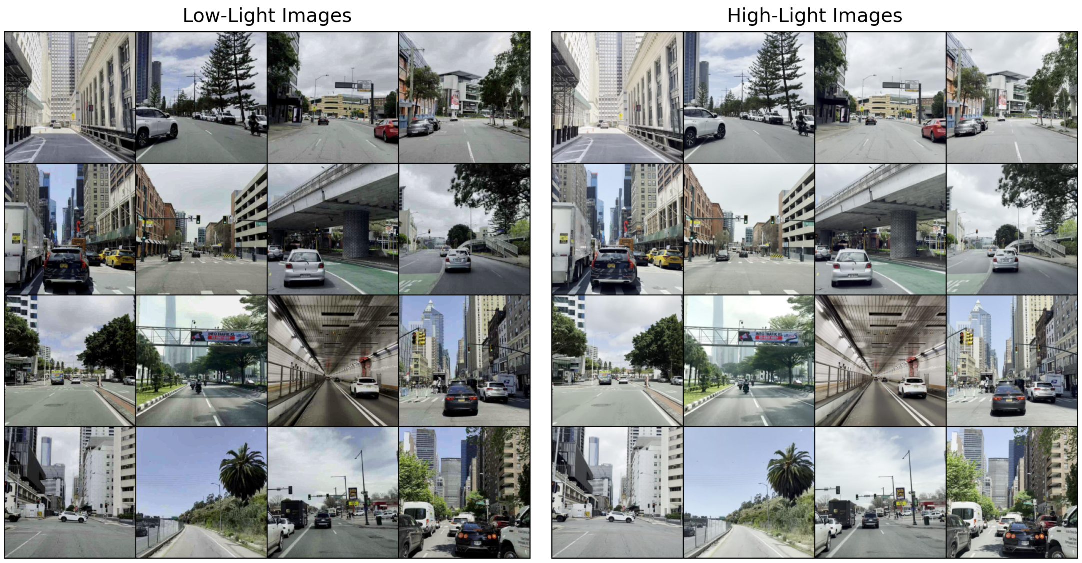

Fig. 5: LoLI Dataset Samples.

B. Metrics

We evaluated the performance of our low-light image enhancement model using three widely recognized metrics: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS). These metrics provide a comprehensive assessment of image quality across our selected LoLI-Street dataset. PSNR and SSIM offer pixel-level accuracy measurements, while LPIPS captures perceptual quality differences, enabling a nuanced evaluation of the model's image enhancement capabilities.

C. Settings

All models are trained using the Adam optimizer with initial learning rates of 2×10^{-4} for the generator and 1×10^{-4} for the discriminator, using $\beta_1 = 0.5$ and $\beta_2 = 0.999$. Training is performed with mixed precision using gradient accumulation over 4 steps. The generator loss combines adversarial loss and L_1 loss with a weighting factor of 100. The discriminator is optimized using BCEWithLogitsLoss. Models are trained for **500** epochs, saving checkpoints every 5 epochs. Unless otherwise stated, training is conducted on a batch size of **32**, and multi-GPU training is enabled using DataParallel when available.

V. Results

In the results section, we present the performance of our low-light image enhancement model evaluated on the selected datasets. The average Peak Signal-to-Noise Ratio (PSNR) achieved by our model is **32.21** dB, indicating a strong ability to preserve image details and minimize noise. The Structural Similarity Index (SSIM) score of **0.9678** demonstrates excellent preservation of structural similarity between the enhanced and ground truth images, reflecting high perceptual quality. Additionally, the model achieves an average Learned Perceptual Image Patch Similarity (LPIPS) score of **0.0172**, further confirming the perceptual enhancement capabilities of the model. These results highlight the effectiveness of our model in enhancing low-light images while maintaining both pixel-level accuracy and perceptual quality across various datasets.

Fig. 6: Simulated High Light Image

VI. Discussion

Table I presents comparative analysis of SOTA models on the LoLI dataset. The best results are in **bold**, and the second-best are underlined.

TABLE I: Performance Comparison of SOTA Models on LoLI Street Dataset

Method	PSNR↑	SSIM↑	LPIPS↓
RetinexF. [3]	27.87	0.3398	0.3037
RQ-LLIE [4]	29.03	0.9167	0.0326
CUE [5]	30.58	0.9100	0.0201
LLFormer [6]	31.62	<u>0.9274</u>	**0.0131**
DiffLL [7]	31.04	0.9165	0.0273
PairLIE [8]	27.66	0.8702	0.0357
FourLLIE [9]	28.07	0.8828	0.1188
SCI [10]	27.79	0.3394	0.3048
TriFuse [11]	<u>31.67</u>	0.9214	0.0201
Ours	**32.21**	**0.9678**	<u>0.0172</u>

For the LLIE tasks, we present quantitative comparisons with state-of-the-art methods on the LoLI-Street dataset, as shown in Table I. Our proposed model achieves the highest performance across multiple evaluation metrics, surpassing previous methods. Specifically, it yields the best PSNR of **32.21** dB and SSIM of **0.9678**, demonstrating superior pixel-level accuracy and structural similarity. Furthermore, our model achieves a highly competitive perceptual quality with an LPIPS score of <u>0.0172</u>, outperforming most prior methods except LLFormer. These results underscore the effectiveness of our approach in enhancing low-light images while maintaining perceptual fidelity.

References

[1] X. Pan, C. Li, Z. Pan, J. Yan, S. Tang, and X. Yin, "Low-light image enhancement method based on Retinex theory by improving illumination map," *Applied Sciences*, vol. 12, no. 5257, 2022. [Online]. Available: https://doi.org/10.3390/app12105257

[2] Y. Han, X. Chen, Y. Zhong, Y. Huang, Z. Li, P. Han, Q. Li, and Z. Yuan, "Low-illumination road image enhancement by fusing Retinex theory and histogram equalization," *Electronics*, vol. 12, no. 990, 2023. [Online]. Available: https://doi.org/10.3390/electronics12040990

[3] Y. Cai, H. Bian, J. Lin, H. Wang, R. Timofte, and Y. Zhang, "Retinexformer: One-stage retinex-based transformer for low-light image enhancement," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 12504–12513.

[4] Y. Liu, T. Huang, W. Dong, F. Wu, X. Li, and G. Shi, "Low-light image enhancement with multi-stage residue quantization and brightness-aware attention," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 12140–12149.

[5] N. Zheng, M. Zhou, Y. Dong, X. Rui, J. Huang, C. Li, and F. Zhao, "Empowering low-light image enhancer through customized learnable priors," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 12559–12569.

[6] T. Wang, K. Zhang, T. Shen, W. Luo, B. Stenger, and T. Lu, "Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method," in *Proc. AAAI Conf. Artif. Intell.*, 2023, pp. 2654–2662.

[7] H. Jiang, A. Luo, H. Fan, S. Han, and S. Liu, "Low-light image enhancement with wavelet-based diffusion models," *ACM Trans. Graph.*, vol. 42, no. 6, pp. 1–14, 2023.

[8] Z. Fu, Y. Yang, X. Tu, Y. Huang, X. Ding, and K. K. Ma, "Learning a simple low-light image enhancer from paired low-light instances," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 22252–22261.

[9] C. Wang, H. Wu, and Z. Jin, "Fourllie: Boosting low-light image enhancement by fourier frequency information," in *Proc. 31st ACM Int. Conf. Multimedia*, 2023, pp. 7459–7469.

[10] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo, "Toward fast, flexible, and robust low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 5637–5646.

[11] M. T. Islam, I. Alam, S. S. Woo, S. A. I. H. Lee, and K. Muhammad, "LoLI-Street: Benchmarking low-light image enhancement and beyond," in *Proc. Asian Conf. Comput. Vis.*, 2024, pp. 1250–1267.