MSFE-GAN: Multi-Scale Feature Extraction GAN for Perceptually Enhanced Low-Light Images

1st Pansilu Wijesiri
Department of Computer Science and Engineering
University of Westminster
London, UK
pansiluwijesiri@gmail.com

2nd Guhanathan Poravi
School of Computing
Informatics Institute of Technology
Colombo, Sri Lanka
guhanathan.p@iit.ac.lk

***Keywords*—Low-Light Enhancement, Multi-Scale GAN, Frequency Domain Processing, Image Restoration, Adversarial Learning.**

I. Introduction

Low-light image enhancement plays a crucial role in downstream computer vision applications such as autonomous driving, semantic segmentation, and security surveillance. Conventional enhancement methods often overlook non-uniform illumination handling, which leads to overexposure, detail, and texture loss within the brightness distribution. As to address these limitations, MSFE-GAN (Multi-Scale Feature Extraction GAN) is proposed with a novel generative adverserial network (GAN) that incoporates spatial and frequency-domain processing as to achieve the overlooked perceptual quality and structural integrity in the enhanced low-light images.

Unlike the traditional methods that apply uniform brightness adjustments, MSFE-GAN introduces a U-Net-based generator for multi-scale feature extraction and a Fourier-based refinement module for high-frequency detail and texture preservation. A dual-Markovian discriminator is employed within the GAN network to ensure global consistency and local texture fidelity, which produces a high-quality, visually coherent image enhancement result.

II. Literature Review

Recent deep learning-based methods have shown significant advancements in the low-light image enhancement space. Traditional enhancement methods, such as histogram equalization and Retinex-based methods, tend to struggle with non-uniform lighting conditions, which often result in unnatural detail and structural incoherency [2]. Deep learning models such as LLNet, SIDNet, and Retinex-Net enhance brightness and suppress noise, but their reliance on spatial-domain processing can produce over-smoothing and details loss [3]. GAN-based methods such as EnlightenGAN and Zero-DCE, uses perceptual enhancement that still struggles with fine-texture preservation and enhancement of intricate details within the image [1]. These domain-specific feature information in fusion resolves the overlooked issues in LLIE.

III. Proposed GAN-Based Enhancement Method

MSFE-GAN acts on spatial and frequency-domain features.

A. Multi-Scale Feature Extraction (MSFE)

A U-Net-based generator extracts the spatial domain features for adaptive brightness and color correction while preserving structural integrity and resolving overexposure effects.

B. Frequency-Based Texture Refinement

To preserve the fine details and prevent oversmoothing, MSFE-GAN integrated Fast Fourier Transform (FFT) to extract the high-frequency information alongside the spatial feature distributions. Here, 2D Discrete Fourier (DFT) is applied for frequency featurs, targetting detail and texture preservation.

$$F(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) e^{-j2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)} \tag{1}$$

where $f(x,y)$ is the spatial image pixels, and $e^{-j2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)}$ is the complex exponential basis function. Then the features are reconstructed with Inverse Fourier Transform (IFT) [5].

C. Dual Markovian Discriminator

A dual Markovian discriminator in low-light image enhancement enables enhancement of both global (entire image, as similar to a standard discriminator) and additionally local (patch-based) image attributes, thereby enforcing global brightness, contrast consistency, and preserved intricacy [4].

IV. Experimental Evaluation

MSFE-GAN was implemented using PyTorch, utilizing an NVIDIA Tesla T4 GPU. The model was trained for 600 epochs using the Adam optimizer ($\beta_1 = 0.5, \beta_2 = 0.999$) with a learning rate of 2×10^{-3} with the cosine annealing schedule. Experiments were conducted on LSRW, LOLv1, LOLv2-Real, and LOLv2-Synthetic paired low-light datasets. Performance was evaluated with standard metrics, PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index Measure) scores targeting visual and structural coherency. The model maintains a compact size of 31.5MB, ensuring efficiency.

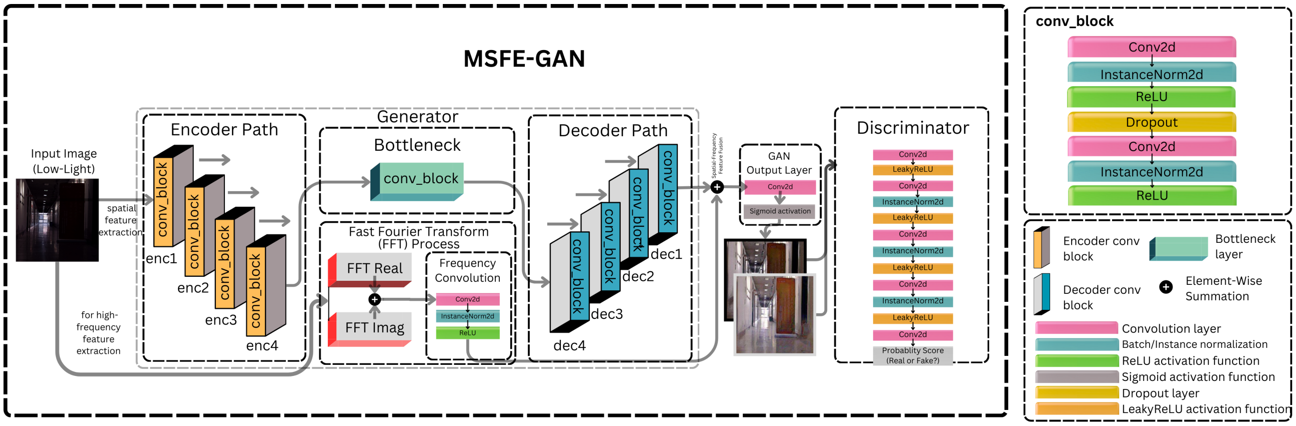

Fig. 1. High-level architecture of MSFE-GAN. The generator follows an encoder-decoder structure with Fourier-based (FFT) refinement, while the discriminator assesses spatial and frequency domain consistency with the dual-Markov discriminator.

TABLE I
PERFORMANCE COMPARISON ON LSRW AND LOLv2-REAL DATASETS

Method	**LSRW**		**LOLv2-Real**	
	PSNR (dB) ↑	**SSIM ↓**	**PSNR (dB) ↑**	**SSIM ↓**
RetinexNet	15.609	0.414	15.47	0.567
MIRNet	16.470	0.477	20.02	0.820
RetinexFormer	19.570	0.578	22.80	0.840
DPEC	19.643	0.576	22.89	0.863
MSFE-GAN	**20.12**	**0.580**	**23.12**	**0.875**

A. Qualitative Results

Fig. 2 demonstrates the visual enhancement results comparing MSFE-GAN with baseline methods. While RUAS still underperforms and KinD produces a weak overall enhancement effect, MirNet and MSFE-GAN perform well in preserving a balance in enhancement. Yet, MSFE-GAN takes over the texture and detail retention, to maintain structural integrity.

B. Quantitative Results

As in Table I, for LSRW, MSFE-GAN achieves the highest PSNR (20.12 dB) and SSIM (0.580), surpassing RetinexNext, MIRNet, RetinexFormer, and DPEC. Similarly, for LOLv2-Real, MSFE-GAN outperforms all methods with a PSNR of 23.12 dB and SSIM of 0.875, which demonstrates superior perceptual quality and structural integrity. The proposed methods are able to achieve such high fidelity to the ground truth while maintaining superiority in efficiency in comparison to state-of-the-art (SOTA) methods.

V. CONCLUSION

This paper presents MSFE-GAN, a novel GAN-based method for low-light image enhancement. Through leveraging multi-scale features in the spatial and frequency domains, the method achieves the highest perceptual quality, improved texture retention with a balanced illumination. Quantitative results on the LSRW and LOLv2-Real datasets demonstrated SOTA performance, with the method outperforming the existing methods in both PSNR and SSIM. Additionally, being compact (31.5MB) in model size, it also ensures efficiency. Future works include further optimizing for edge-device integration.

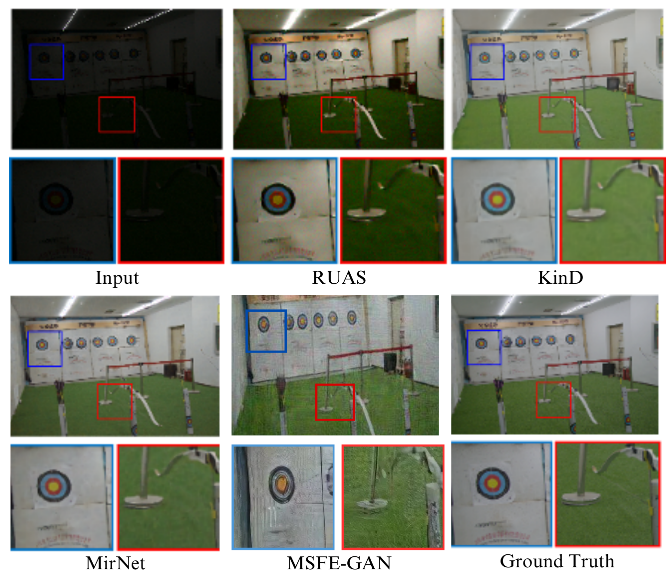

Fig. 2. Qualitative comparison of MSFE-GAN against existing methods on the LOLv1 dataset. Proposed method produces more natural brightness balance and improved texture details in comparison to other methods.

REFERENCES

[1] Y. Fu, Y. Hong, L. Chen, and S. You, "LE-GAN: Unsupervised low-light image enhancement network using attention module and identity invariant loss," *Knowl.-Based Syst.*, vol. 240, p. 108010, 2022.

[2] S. Hu, J. Yan, and D. Deng, "Contextual Information Aided Generative Adversarial Network for Low-Light Image Enhancement," *Electronics*, vol. 11, no. 1, p. 32, 2022.

[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, 2023.

[4] L. Wang, L. Zhao, T. Zhong, and C. Wu, "Low-light image enhancement using generative adversarial networks," *Sci. Rep.*, vol. 14, no. 1, p. 18489, 2024.

[5] K. Shang, M. Shao, Y. Qiao, and H. Liu, "Frequency-aware network for low-light image enhancement," *Comput. Graph.*, vol.118, pp.210, 2024.