Adaptive Context-Aware Generative Adversarial Network For Low-quality Image Enhancement

Xingyu Pan[1*] and Fengling Chen[2]

[1] School of Electronic and Electrical Engineering, Zhengzhou University of Science and Technology, Zhengzhou, 450064, China
[2] Zhengzhou Electric Power College, Zhengzhou, 450003, China
[*] Corresponding author. E-mail: xingyupan_zst@163.com

Received: Oct. 9, 2024; Accepted: Apr. 19, 2025

Low-quality image enhancement methods can effectively improve image quality and details, which have attracted great attention in various fields. However, current methods still face with two issues: (1) They commonly earn a deterministic generation mapping between low-quality and normal images via relying on pixel-level reconstruction, leading to improper brightness and noise in the enhancing process. (2) They use only one type of generative model, either explicit or implicit, which limits flexibility and efficiency of models. To this end, a novel flow-based generative adversarial network with dual attention (FGAN-DA) is devised for data generation. Specifically, FGAN-DA constructs a hybrid generative model via combining explicit and implicit components within the GAN architecture, which effectively alleviates detail blurred and singularity caused by sole generation modeling. FGAN-DA comprises the dual attention feature extraction, invertible flow generation network, the Markov discriminant network. The three modules seamlessly collaborate in enhancing images with good perceptual quality, which effectively boosts the performance of FGAN-DA. Finally, quantitative metrics and visual quality evaluations demonstrate that FGAN-DA sets a new baseline in can generate images with good perceptual quality.

Keywords: Data generation; dual attention; flow generative network

http://dx.doi.org/10.6180/jase.202601_29(1).0012

1. Introduction

Low-quality images, characterized by blurred details, low contrast, color distortion, and noticeable noise, typically result from the limitations of imaging devices or challenging shooting conditions [1–3]. These quality issues not only diminish the overall human visual experience but also pose considerable challenges for various computer vision tasks. For instance, object detection algorithms may struggle to accurately locate and identify objects due to the lack of clear boundaries and details. Clustering tasks are similarly affected, as the reduced clarity and altered colors can hinder the model ability to correctly categorize images. Addressing these challenges requires advanced techniques to enhance images and boost the performance of various systems in dealing with low-quality images [4–6].

Deep learning-based low-quality image enhancement (LIE) approaches have significant advancement over traditional spatial and transform domain-based approaches, which are mainly divided into two branches: supervised-based approaches and unsupervised-based approaches [7, 8]. Supervised-based approaches aim to employ labeled data and a supervised training approach to train network parameters for LIE. For instance, Shen et al. [9] introduce MSR-Net via simulating the traditional MSR algorithm. MSR-Net primarily learns the mapping relationship between the dark and the bright views, using convolutional layers to replace Gaussian surround functions, and takes low/normal-quality images as network input and output. Wei et al. [10] propose the deep Retinex-Net via jointly

learning decomposition and enhancement, and train the model using a data-driven approach. Unsupervised-based methods rely on the constraints of the losses and the quality judgment for training, rather than paired labeled data [11]. For example, A typical example is the GAN-based method. Jiang et al. [12] propose the EnlightGan network to enhance low-light images, utilizing information learned from low-quality samples to regularize unpaired image training, but this approach easily produces localized color artifacts, damaging the visual quality. Zhang et al. [13] devise a self-supervised LIE which only uses low-light images for training. Rooted in the Retinex theory, they suppose the maximum value of the channel and the entropy of the information in the reflection and illumination modules are consistent, using the reflection module as the enhancement image.

Current methods treat the low-quality image enhancement problem as a generation task, for improving image quality [14–16]. However, they rely on either an explicit or implicit generative model to enhance images, which limits its flexibility and efficiency of models. Moreover, these methods typically learn a clear and definite correspondence mechanism between low-quality and normal images, meaning they generate a single high-quality result from the low-quality images instead of modeling the complex distribution of possible outcomes. The reason is that deep-learning models depend on a reference ground truth image, requiring generators to produce an output that closely resembles this reference via reconstruction losses (L1 or L2). Such losses focus primarily on pixel-wise matching rather than pixel distribution matching, which leads to generation images that lack high-frequency details and tend to predict the average outcome [17, 18]. In other words, the low-quality image enhancement should be viewed as a task of selecting the most suitable solution from all possible outputs, while most existing networks only provide a single correct output relative to the reference ground truth images [19, 20].

To this end, a novel flow-based generative adversarial network with dual attention (FGAN-DA) is devised for low-quality image enhancement. Specifically, FGAN-DA constructs a hybrid generative model via jointly optimizing explicit and implicit generative strategies, which comprises the dual attention feature extraction (DAF), invertible flow generation network (IFG), the Markov discriminant network (MD). DAF utilizes channel and spatial attention within the convolutional architecture to extract representations of low-quality samples, which contributes to capturing key information and the precise localization of important regions in images. Next, IFG takes representations as conditional priors to learn a bidirectional mapping that reconfigures normal images to conform to a Gaussian distribution. Such a manner allows the model to simulate the conditional distribution of normal images, enabling it to sample multiple plausible normal results and generate diverse images. Meanwhile, MD uses the adversarial game to provide constraints for improving the detail information of enhancement images. Finally, quantitative and visual results demonstrate that FGAN-DA endows enhancement images with good perceptual quality.

Below is summary of the main contributions: (1) FGAN-DA introduces a hybrid generative model that combines explicit and implicit generation strategies. This approach effectively addresses the limitations of single generative models, providing greater flexibility and efficiency in low-quality image enhancement tasks. (2) FGAN-DA employs an invertible flow generation network to model the conditional distribution of normal images. This network not only increases model flexibility but also enables the generation of diverse, high-quality images with varying exposure levels during inference. (3) Extensive experiments on the LoL and MIT-Adobe FiveK datasets demonstrate that FGAN-DA achieves superior performance in terms of PSNR, SSIM, and LPIPS metrics, outperforming various existing low-quality image enhancement methods.

FGAN-DA is outlined: Section II represents comprehensive methodologies in low-quality image enhancement. Section III details the fundamental components underpinning FGAN-DA. Section IV presents an in-depth analysis of results. Lastly, Section V offers the concluding insights.

2. Flow-based generative adversarial network with dual attention for low-quality image enhancement

A novel flow-based generative adversarial network with dual attention is proposed for low-quality image enhancement (FGAN-DA), which contains the dual attention feature extraction, invertible flow generation network, the Markov discriminant network, as shown in Fig. 1. FGAN-DA, compared with traditional pixel-level paired matching reconstruction loss, achieves the low-quality image enhancement task from a distribution matching view via using an explicit and implicit hybrid generative model. In FGAN-DA, the generator acts as an explicit probability density estimator, simulating the distribution of normal images and producing outputs with varying exposure levels during inference. At the same time, the discriminator in adversarial training serves as an implicit sampler, boosting the detail image quality generated by the generator, thereby achieving high-quality image enhancement

through implicit generation.

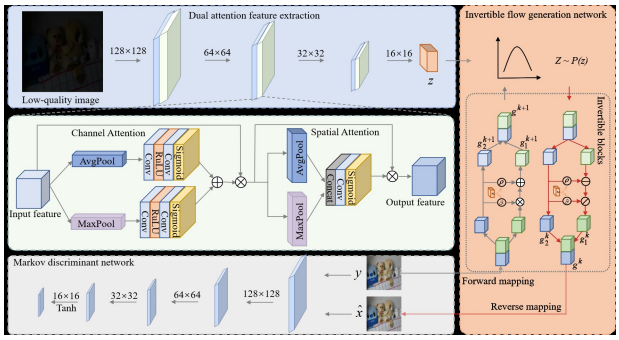

Fig. 1. The illustration of FGAN-DA

Taking a low-quality image as the input, FGAN-DA utilizes the dual feature extraction with three convolutional units to extract representations z. Next, FGAN-DA takes representations as conditional priors, i.e., $z \sim p(z)$, in the invertible flow generation network to learn a bidirectional mapping, i.e., $(z; x) \rightarrow y$ and $(z; y) \rightarrow x$, for generating the high-quality image $\hat{x}$. Meanwhile, FGAN-DA treats the Markov discriminant network as an implicit sampler via the adversarial game to improve the detail information of enhancement images, i.e., $D(y) \rightarrow 1$ and $D(\hat{x}) \rightarrow 0$.

2.1. Dual attention feature extraction

DAF contains three convolutional units, each of which includes the depth separable convolution, the channel attention, and the spatial attention modules. Specifically, given a low-quality image $x \in [H, W, C]$, where H, W, and C stand for the height, the width, and the channel, respectively, DAF applies the depth separable convolution to extract features, formulated as:

$$\bar{z}^{i+1} = f^{i+1}\left(z^i, k^i, s^i\right), i = 1, 2, 3 \tag{1}$$

where k^i and s^i denote the kernel size and the stride in the i-th depth separable convolution layer. $\bar{z}^{i+1}$ and z^i denote outputs of the $i+1$-th depth separable convolution layer and the i-th convolutional unit. Then $\bar{z}^{i+1}$ is normalized via the batch normalization (BN) and ReLU activation layers:

$$z^{i+1} = \text{ReLU}\left(\text{BN}\left(\bar{z}^{i+1}\right)\right) \tag{2}$$

The normalized feature map is then passed through the channel attention and the spatial attention. The former computes channel-wise importance using global average pooling gpool$(\cdot)$, followed by MLP and a sigmoid activation function $\sigma(\cdot)$:

$$\tilde{z}_g^{i+1} = \sigma\left(MLP\left(\text{gpool}\left(\bar{z}^{i+1}\right)\right)\right) \tag{3}$$

For spatial attention, a 7×7 convolution is applied to capture spatial dependencies:

$$\tilde{z}_s^{i+1} = \sigma\left(\text{Conv}\left(\tilde{z}_g^{i+1}\right)\right) \tag{4}$$

Finally, the output of the i-th convolutional unit is computed:

$$z^{i+1} = w_1 \tilde{Z}_g^{i+1} + w_2 \tilde{Z}_s^{i+1} + w_3 \tilde{Z}^{i+1} \tag{5}$$

where w_1, w_2, w_3 are learnable parameters that denote feature importance in fusing process. DAF improves ability to focus on severely low-quality areas of images, resulting in better performance for image enhancement tasks.

2.2. Invertible flow generation network

In FGAN-DA, the objective of the invertible flow generation network is to model the conditional distribution of normal images under the condition of extracted features of low-quality images. Specifically, the network transforms feature maps using B invertible blocks, each consisting of three invertible layers: invertible 1×1 convolution, conditional coupling layer, and injection layer.

During the forward pass, invertible 1×1 convolution establishes information transfer across the channel dimension. The conditional coupling layer splits the channels into two parts. The first part of the features, along with the low-quality image features extracted by the DAF, is fed into a 3-layer convolutional mapping network to establish a relationship between the low-quality and normally images. Simultaneously, the convolutional mapping network predicts the mapping for the second part of the features with respect to the normally image. This layer adds non-linear expressive power and spatial dependency to invertible flow generation network while ensuring invertibility and computationally tractable log-determinant. The conditional coupling layer operation is defined as:

$$\begin{aligned} g_1^{k+1} &= g_1^k \\ g_2^{k+1} &= \exp\left(v^k\left(\left[g_1^k, z\right]\right) \times g_2^k\right) + v^k\left(\left[g_1^k, z\right]\right) \\ k &= 1, 2, \ldots B \end{aligned} \tag{6}$$

where $g^k = \left[g_1^k, g_2^k\right]$, g_1^k and g_2^k denote two parts of g^k, respectively. $[,]$ denotes the concatenation. $v^k(\cdot)$ denotes the convolutional mapping network.

Finally, the injection layer is applied to transform all channels under the condition of z. Unlike the coupling layer, the injection layer utilizes all channels of g^k for feature fusion. This allows all channels and spatial positions to be influenced:

$$g^{k+1} = \exp\left(v^k([z]) \times g^k\right) + v^k([z]) \tag{7}$$

The backpropagation calculation operation is achieved by inverting the addition and multiplication of forward propagation. The reverse mapping of the conditional coupling layer:

$$\begin{aligned} g_1^k &= g_1^{k+1} \\ g_2^k &= g_2^{k+1} - v^k\left(\left[g_1^k, z\right] / \exp\left(v^k\left(\left[g_1^k, z\right]\right)\right)+\right) \\ k &= 1, 2, \ldots B \end{aligned} \tag{8}$$

And the reverse mapping of the injection layer:

$$g^k = g^{k+1} - v^k(z) / \exp\left(v^k(z)\right) \tag{9}$$

2.3. Markov discriminant network

Unlike conventional discriminators, the Markov discriminant network takes as input both the normal image y belonging to the low-quality image x and the generation image $\hat{x}$ obtained by IFG. The output of the Markov discriminant network is a 16×16 feature matrix, which divides the image into small patches before classification. This approach enhances the discriminator's accuracy and allows the network to capture more detailed information during training.

Specifically, first four layers of the Markov discriminant network use 3×3 convolutional kernels with downsampling, BN layer, and Leaky ReLU:

$$\hat{x}_i = \mathrm{ReLU}\left(\mathrm{BN}\left(D_i\left(\hat{x}_{i-1}\right)\right)\right), i = 1, 2, 3, 4 \tag{10}$$

A Tanh function is added on the fifth layer:

$$\hat{x}_5 = \frac{\exp\left(\hat{X}_4\right) - \exp\left(-\hat{x}_4\right)}{\exp\left(\hat{x}_4\right) + \exp\left(-\hat{x}_4\right)} \tag{11}$$

The output matrix $\hat{X}_5$ is of size 16×16, where each element represents a receptive field of the input images. This receptive field captures local features such as textures and fine details, making the discriminator more sensitive to local variations in the images.

By using this structure, the Markov discriminant network is able to effectively differentiate between real and generated images at a patch level, enhancing its precision and robustness.

2.4. The overall loss

To effectively restore the visual quality of images while preserving detailed features, the loss of FGAN-DA consists of adversarial loss $L_{\mathrm{WGAN}}(G, D)$, invertible flow generation loss L_g, and semantics loss L_s :

$$L = \min_G \max_D L_{\mathrm{WGAN}}(G, D) + \lambda_1 L_g + \lambda_2 L_s \tag{12}$$

where λ_1 and λ_2 are trade-off parameters for losses. G denotes the invertible flow generation network. D denotes the Markov discriminant network.

Adversarial loss: The traditional GAN loss function is optimized based on divergence, but this often leads to unstable training, causing issues such as vanishing gradients, which may result in training collapse. Therefore, WGAN with gradient penalty is employed as follows:

$$\begin{aligned} L_{\mathrm{WGAN}}(G, D) =& E[D(y)] - E[D(G(z))] \\ &+ \alpha_{GP} E\left[\left(\left\|\nabla_{x'} D\left(x'\right)\right\|_2 - 1\right)^2\right] \end{aligned} \tag{13}$$

where x' denotes samples along the straight line between corresponding points of the $\hat{x}$ and y. α_{GP} is weight of the gradient penalty and is set as 10.

Invertible flow generation loss: During training, given the conditional input x, the network processes a high-dimensional random variable $y \sim p(y)$ (normal image) and a latent variable $z \sim p(z)$ with a simple and tractable distribution (e.g., Gaussian distribution). The invertible flow generation network applies a series of invertible transformations to map y to $z : (y; x) \rightarrow z$. Conversely, y can be recovered by the inverse mapping $(z; x) \rightarrow y$ by sampling z from the simple distribution, i.e., $z \sim p(z)$. By increasing the depth and complexity of the invertible flow generation network, a more flexible conditional density can be achieved. According to the chain rule, the log probability density of $\log P_{y|x}(y \mid x)$ can be derived as:

$$\log P_{y|x}(y \mid x) = \log P_z(G(y, x)) + \sum_{k=0}^{B-1} \log \left| \det \frac{\partial G\left(g^k, z\right)}{\partial g^k} \right| \tag{14}$$

Then, the objective of the invertible flow generation network is in the following:

$$L_g = -\log P_{y|x}(y \mid x) \tag{15}$$

During the inference phase, the invertible flow generation network generates normal images by taking the input low- quality image and randomly sampling the latent variables z from $\log P_{y|x}(y \mid x)$, running the process in reverse. To accelerate inference, in the experiments, the network directly uses the mean of the randomly sampled z as the latent feature z.

The semantics loss: To make the enhancement images more realistic and visual, the semantics loss is added where perceptual distance is used as the form of the loss. A pre-trained VGG19 architecture is introduced to construct the semantics loss by extracting high-level features from the output of the third convolutional layer before the fourth max-pooling operation.

$$L_s = E\left[\|\psi(y) - \psi(G(z))\|_2^2\right] \tag{16}$$

where $\psi(\cdot)$ represents the semantic extraction network. The detailed calculation process of FGAN-DA is shown in the algorithm 1.

The algorithm 1. Flow-based generative adversarial network with dual attention for low-quality image enhancement.

Input: The low-quality image datasets $X = \{x_i, y_i\}_{i=1}^{n}$, y denotes the corresponding normal images, the epoch number M, the trade-off parameters λ_1 and λ_2.
Output: The image enhancement datasets $\hat{X} = \{\hat{y}_i\}_{i=1}^{n}$, and the parameters of FGAN-DA.

---

1. Initialize model parameters;
2. When $Iter \leq M$:
3. When $i \leq n$:
4. Extracting representations z_i of x_i in the dual attention feature extraction via Eq. (5).
5. Learning a bidirectional mapping to generate enhancement image $\hat{x}_i$, i.e., $(z_i; x_i) \rightarrow y_i$ and $(z_i; y_i) \rightarrow \hat{x}_i$ in the invertible flow generation network, via Eqs. (7) and (9).
6. Generating output matrix of $\hat{X}_i$ in the Markov discriminant network via Eqs. (10) and (11).
7. Computing the adversarial loss $L_{\text{WGAN}}(G, D)$, the invertible flow generation loss L_g, and semantics loss L_s, via Eqs. (12), (15) and (16), respectively.
8. Optimize the aforementioned loss in an end-to-end manner via using the stochastic gradient descent method.
9. **Return:** The image enhancement datasets $\hat{X} = \{\hat{X}_i\}_{i=1}^{n}$, and the parameters of FGAN-DA.

---

3. Results

3.1. Setup

Dataset: Two common datasets i.e., LoL and MIT-Adobe FiveK, in the low-quality image enhancement task, are utilized to assess the performance of FGAN-DA. LoL includes 500 pairs of low-quality and normal images MIT-Adobe FiveK includes 5,000 pairs of low-quality and normal images.

Implementation Details: FGAN-DA is optimizated with the Adam optimizer. The total number of epochs is 150 with 8 batch size. The learning rate is 0.01 with the learning rate decaying by a factor of 0.5 every 30 epochs.

Evaluation Metrics: The image enhancement performance of FGAN-DA is assessed by RSNR, SSIM, and LPIPS, where PSNR measures the average pixel similarity between enhancement images and corresponding normal images. SSIM assesses the structural similarity between enhancement images and corresponding normal images. LPIPS evaluates the perceptual differences between enhancement images and corresponding normal images.

3.2. Comparison with baselines

Comparison methods: Eight baselines of low-quality image enhancement are selected including RetinexGAN [15], DualGAN [8], EnlightGAN [12], HWA-net [16], LLFlow [17], Bread-ME [18], NNPriors [19], and RetinexDIP [20].

Comparison results: Table 1 shows experiment outcomes regarding LOL and MIT-Adobe FiveK in PSNR, SSIM, and LPIPS. FGAN-DA achieves the best performance across both datasets in all three metrics. For example, For LOL dataset, FGAN-DA records a PSNR of 24.95, outperforming all other methods. Similarly, For MIT-Adobe FiveK dataset, FGAN-DA reaches a SSIM of 0.90 , reflecting superior structural similarity compared to the other methods. Three key factors contribute to this success: (1) Dual attention feature extraction effectively extracts features from low-quality images using both channel and spatial attention mechanisms, which allows FGAN-DA to focus on severely degraded areas, leading to superior performance in LIE. (2) Invertible flow generation network effectively models the conditional distribution of normal images through invertible transformations. It not only increases the model's flexibility but also enables it to generate outputs with varying exposure levels during inference, resulting in high-quality outcomes in image enhancement tasks. (3) Markov discriminant network, unlike traditional discriminators, enhances classification accuracy by dividing the image into small regions for analysis. This approach allows the network to capture more detailed information during training, thereby improving the quality of the fine details in the enhancement images.

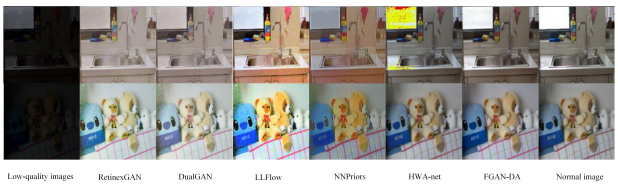

Fig. 2. The visual comparison on the test images from the LOL dataset

In addition, Fig. 2 provides the visual comparison on the test images from the LOL dataset. Due to the LOL dataset being captured under extremely low-quality conditions, it has a very low signal-to-noise ratio with severe noise. Achieving high-contrast restoration without amplifying the noise presents a significant challenge. In the enhancement results of other methods, varying degrees of degradation are observed. For example, RetinexGAN, DualGAN, and LLFlow can restore vibrant colors but of-

Table 1. Comparison results with baselines

Method	LOL			MIT-Adobe FiveK		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
RetinexGAN	14.75	0.68	0.38	18.08	0.75	0.33
DualGAN	16.42	0.56	0.41	20.00	0.77	0.35
EnlightGAN	16.75	0.56	0.35	19.51	0.77	0.31
HWA-net	24.14	0.83	0.13	24.44	0.87	0.09
LLFlow	17.48	0.65	0.32	20.01	0.79	0.26
Bread-ME	18.23	0.72	0.35	21.69	0.81	0.22
NNPriors	15.49	0.65	0.38	17.31	0.72	0.39
RetinexDIP	20.73	0.81	0.17	22.76	0.85	0.18
FGAN-DA	**24.95**	**0.85**	**0.10**	**25.62**	**0.90**	**0.08**

ten exhibit significant noise. In contrast, FGAN-DA boosts the brightness and preserves rich detail and color information. Artifacts are rarely observed, resulting in superior perceptual quality.

3.3. Ablation analysis

In the ablation analysis of the FGAN-DA, the impact of the adversarial loss, the invertible flow generation loss, and the semantic loss are examined on low-quality image enhancement performance. The results in Table 2 show that using only adversarial loss yields relatively low PSNR, SSIM, and LPIPS scores, indicating that adversarial training alone is insufficient for generating high-quality images. When the invertible flow generation loss is introduced, there is a significant improvement in model performance, with a substantial increase in PSNR and SSIM scores and a notable decrease in LPIPS score. This underscores the importance of the invertible flow in enhancing image details and overall structure. Adding the semantic loss further helps in preserving image semantic information, though it does not achieve the best performance. This suggests that semantic information needs to be combined with other loss functions to achieve optimal results. Ultimately, when all three loss functions are used together, the model demonstrates outstanding performance across all evaluation metrics, with a PSNR of 24.95, an SSIM of 0.85, and an LPIPS of only 0.10. These findings validate the advantages of combining multiple loss functions in improving image quality and highlight the importance of considering different loss functions collectively to achieve better enhancement results in low-quality image tasks.

3.4. Effect of different numbers of invertible block

The number impact of invertible blocks on the enhancement performance of FGAN-DA is analyzed, as illustrated in Fig. 3. The results indicate that the number of invertible blocks significantly influences the model performance. Initially, as the number of blocks increases, both SSIM and LPIPS metrics improve, demonstrating better structural similarity and perceptual quality. This suggests that a greater number of blocks helps the model capture more complex dependencies and enhance the representation of LIE. However, after a certain point, the performance starts to degrade as indicated by a decline in SSIM and an increase in LPIPS. This diminishing return might be due to overfitting or increased model complexity, which makes it harder to generalize from the training data. Thus, an optimal balance must be struck between the number of invertible blocks and the overall enhancement performance. This analysis highlights the importance of choosing the right number of invertible blocks to achieve the best image enhancement results.

Table 2. Ablation experiments of each component in FGAN-DA(LOL)

L_{WGAV}	L_g	L_s	PSNR	SSIM	LPIPS
✓			10.25	0.52	0.45
✓	✓		20.96	0.75	0.18
✓		✓	18.22	0.69	0.26
✓	✓	✓	24.95	0.85	0.10

3.5. Parameter analysis

A thorough analysis of two critical parameters, i.e., λ_1 and λ_2, is conducted on LOL and MIT-adobe-FiveK datasets with three metrics. Specifically, λ_1 and λ_2 are constrained in $[1, 0.1, 0.01, 0.001]$, and by varying the values of λ_1 and λ_2, the changes of RSNR,

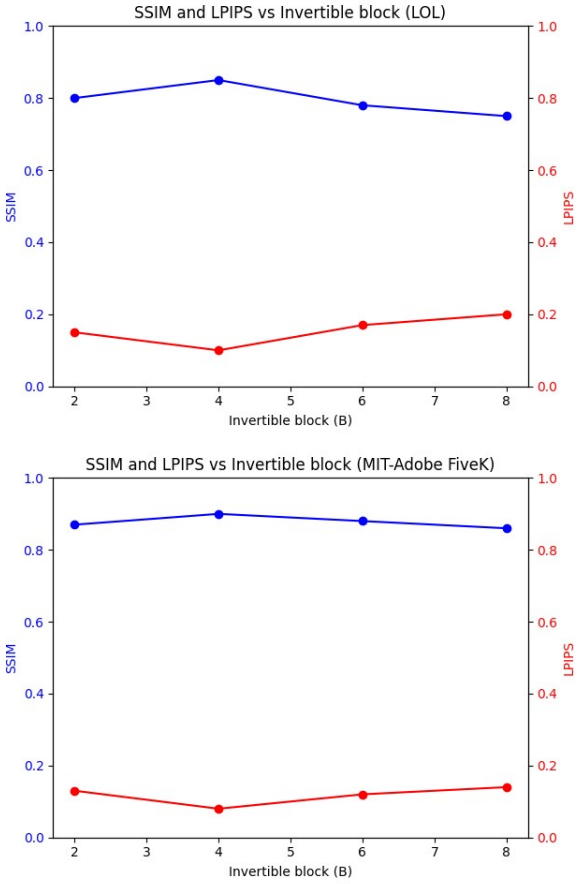

Fig. 3. Effect of different numbers of invertible block on enhancement performance on the LOL and MIT-adobe FiveK datasets

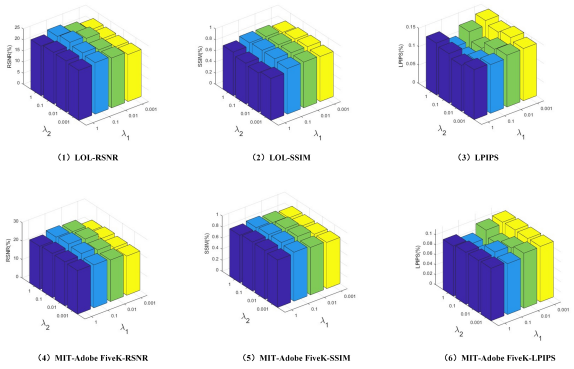

Fig. 4. Varying of parameters on enhancement performance on LOL and MIT-adobe FiveK datasets

SSIM, and LPIPS are recorded in Fig. 4. The results show different combinations of λ_1 and λ_2 significantly impacted the model performance. This indicates that by carefully tuning λ_1 and λ_2, an effective trade-off can be achieved among different metrics, thereby optimizing the overall performance of image enhancement. In the experiments, $\lambda_1 = 0.1$ and $\lambda_2 = 0.01$ are used on LOL and MIT-adobe-FiveK datasets.

4. Conclusions

This paper devises a novel flow-based generative adversarial network (FGAN-DA) for low-quality image enhancement. FGAN-DA effectively addresses the issues of blurred details and single-peak responses often encountered in image enhancement, by combining explicit and implicit generative models. FGAN-DA includes a dual-attention feature extraction, an invertible flow generation network, and a Markov discriminator network. These elements work together to achieve efficient transformation from low-quality images to normal images. Extensive experiments on the LoL and MIT-Adobe FiveK datasets validate that FGAN-DA achieves significant performance improvements across multiple evaluation metrics, including PSNR, SSIM, and LPIPS. Compared to various existing low-quality image enhancement methods, FGAN-DA generates images with superior perceptual quality, higher structural similarity, and closer visual fidelity to real scenes. Future work will explore further optimizations of the network architecture and the potential application of FGAN-DA to other image enhancement tasks and practical scenarios. We also plan to explore the application of FGAN-DA to other image enhancement tasks, such as medical image enhancement and underwater image enhancement, where maintaining perceptual quality and structural similarity is crucial.

References

[1] 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: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, 12140–12149.

[2] Y. Gao, W. Zhang, H. He, L. Cao, Y. Zhang, Z. Huang, and X. Zhao, (2024) *"TSSFN: Transformer-based self-supervised fusion network for low-quality fundus image enhancement"* **Biomedical Signal Processing and Control *89***: 105768. DOI: 10.1016/j.bspc.2023.105768.

[3] P. Li, Z. Chen, L. T. Yang, Q. Zhang, and M. J. Deen, (2017) *"Deep convolutional computation model for feature learning on big data in internet of things"* **IEEE Transactions on Industrial Informatics *14***(2): 790–798. DOI: 10.1109/TII.2017.2739340.

[4] P. Li, A. A. Laghari, M. Rashid, J. Gao, T. R. Gadekallu, A. R. Javed, and S. Yin, (2022) *"A deep multimodal adversarial cycle-consistent network for smart enterprise system"* **IEEE Transactions on Industrial Informatics *19***(1): 693–702. DOI: 10.1109/TII.2022.3197201.

[5] P. Li, J. Gao, J. Zhang, S. Jin, and Z. Chen, (2022) *"Deep Reinforcement Clustering"* **IEEE Transactions on Multimedia**: DOI: 10.1109/TMM.2022.3233249.

[6] K. G. Lore, A. Akintayo, and S. Sarkar, (2017) *"LLNet: A deep autoencoder approach to natural low-light image enhancement"* **Pattern Recognition *61***: 650–662. DOI: 10.1016/j.patcog.2016.06.008.

[7] X. Cheng, J. Zhou, J. Song, and X. Zhao, (2023) *"A highway traffic image enhancement algorithm based on improved GAN in complex weather conditions"* **IEEE Transactions on Intelligent Transportation Systems *24***(8): 8716–8726. DOI: 10.1109/TITS.2023.3258063.

[8] R. Cong, W. Yang, W. Zhang, C. Li, C.-L. Guo, Q. Huang, and S. Kwong, (2023) *"Pugan: Physical model-guided underwater image enhancement using gan with dual-discriminators"* **IEEE Transactions on Image Processing *32***: 4472–4485. DOI: 10.1109/TIP.2023.3286263.

[9] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, and J. Ma, (2017) *"Msr-net: Low-light image enhancement using deep convolutional network"* **arXiv preprint arXiv:1711.02488**: DOI: 10.48550/arXiv.1711.02488.

[10] C. Wei, W. Wang, W. Yang, and J. Liu, (2018) *"Deep retinex decomposition for low-light enhancement"* **arXiv preprint arXiv:1808.04560**: DOI: 10.48550/arXiv.1808.04560.

[11] 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*. 2019, 1632–1640. DOI: DOI:10.1145/3343031.335092.

[12] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. W. Enlightengan, (2021) *"Deep light enhancement without paired supervision., 2021, 30"* **DOI: https://doi. org/10.1109/TIP**: 2340–2349. DOI: 10.1109/TIP.2021.3051462.

[13] Y. Zhang, X. Di, B. Zhang, and C. Wang, (2020) *"Self-supervised image enhancement network: Training with low light images only"* **arXiv preprint arXiv:2002.11300**: DOI: 10.48550/arXiv.2002.11300.

[14] 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*. 2020, 1780–1789. DOI: 10.1109/CVPR42600.2020.00185.

[15] T. Ma, M. Guo, Z. Yu, Y. Chen, X. Ren, R. Xi, Y. Li, and X. Zhou, (2021) *"RetinexGAN: Unsupervised low-light enhancement with two-layer convolutional decomposition networks"* **IEEE Access *9***: 56539–56550. DOI: 10.1109/ACCESS.2021.3072331.

[16] C.-M. Fan, T.-J. Liu, and K.-H. Liu. "Half wavelet attention on m-net+ for low-light image enhancement". In: *2022 IEEE International Conference on Image Processing (ICIP)*. 2022, 3878–3882. DOI: 10.48550/arXiv.2203.01296.

[17] Y. Wang, R. Wan, W. Yang, H. Li, L.-P. Chau, and A. Kot. "Low-light image enhancement with normalizing flow". In: *Proceedings of the AAAI conference on artificial intelligence*. ***36***. 3. 2022, 2604–2612. DOI: 10.1609/aaai.v36i3.20162.

[18] X. Guo and Q. Hu, (2023) *"Low-light image enhancement via breaking down the darkness"* **International Journal of Computer Vision *131***(1): 48–66. DOI: 10.1007/s11263-022-01667-9.

[19] J. Liang, Y. Xu, Y. Quan, B. Shi, and H. Ji, (2022) *"Self-supervised low-light image enhancement using discrepant untrained network priors"* **IEEE Transactions on Circuits and Systems for Video Technology *32***(11): 7332–7345. DOI: 10.1109/tcsvt.2022.3181781..

[20] Z. Zhao, B. Xiong, L. Wang, Q. Ou, L. Yu, and F. Kuang, (2021) *"RetinexDIP: A unified deep framework for low-light image enhancement"* **IEEE Transactions on Circuits and Systems for Video Technology *32***(3): 1076–1088. DOI: 10.1109/tcsvt.2021.3073371.