



Article Real-Time Low-Light Imaging in Space Based on the Fusion of Spatial and Frequency Domains

Jiaxin Wu ^{1,2,3}, Haifeng Zhang ^{1,3,*}, Biao Li ^{1,3}, Jiaxin Duan ^{1,3,4}, Qianxi Li ^{1,2,3}, Zeyu He ^{1,2,3}, Jianzhong Cao ^{1,3} and Hao Wang ^{1,3}

- ¹ Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China; wujjaxin221@mails.ucas.ac.cn (J.W.); hezeyu23@mails.ucas.ac.cn (Z.H.)
- ² University of Chinese Academy of Sciences, Beijing 100049, China
- ³ Xi'an Key Laboratory of Spacecraft Optical Imaging and Measurement Technology, Xi'an 710119, China
- ⁴ School of Opto-Electronical Engineering, Xi'an Technological University, Xi'an 710021, China
- Correspondence: zhanghf@opt.ac.cn

Abstract: Due to the low photon count in space imaging and the performance bottlenecks of edge computing devices, there is a need for a practical low-light imaging solution that maintains satisfactory recovery while offering lower network latency, reduced memory usage, fewer model parameters, and fewer operation counts. Therefore, we propose a real-time deep learning framework for low-light imaging. Leveraging the parallel processing capabilities of the hardware, we perform the parallel processing of the image data from the original sensor across branches with different dimensionalities. The high-dimensional branch conducts high-dimensional feature learning in the spatial domain, while the mid-dimensional and low-dimensional branches perform pixel-level and global feature learning through the fusion of the spatial and frequency domains. This approach ensures a lightweight network model while significantly improving the quality and speed of image recovery. To adaptively adjust the image based on brightness and avoid the loss of detailed pixel feature information, we introduce an adaptive balancing module, thereby greatly enhancing the effectiveness of the model. Finally, through validation on the SID dataset and our own low-light satellite dataset, we demonstrate that this method can significantly improve image recovery speed while ensuring image recovery quality.

Keywords: low-light imaging; real time; frequency domain

1. Introduction

With the rapid development of space technology, the key technology for the cleaning and recovery of space targets relying on machine vision has attracted extensive attention and research due to its low cost and rich variety of algorithms. However, most image processing techniques rely on high image quality, while the actual space environment is often severely limited in terms of illumination, significantly degrading the performance of these algorithms. To improve the practical application of these algorithms, a practical low-light high-quality imaging solution is essential. However, traditional methods such as histogram equalization and gamma correction directly enhance the brightness and contrast of the entire image globally, without considering the uneven lighting conditions, resulting in enhanced images that deviate significantly from reality. Therefore, by leveraging the powerful fitting capability of deep learning frameworks and incorporating local and global feature learning, low-light image restoration provides us with a novel solution.

A practical solution must possess a lower network latency, reduced memory usage, fewer model parameters, and a lower operation count while maintaining satisfactory restoration. However, these characteristics are contradictory. Currently, most solutions sacrifice some model speed and computational efficiency to improve image restoration quality [1–5]. These approaches overlook the performance bottlenecks of edge computing



Citation: Wu, J.; Zhang, H.; Li, B.; Duan, J.; Li, Q.; He, Z.; Cao, J.; Wang, H. Real-Time Low-Light Imaging in Space Based on the Fusion of Spatial and Frequency Domains. *Electronics* 2023, *12*, 5022. https://doi.org/ 10.3390/electronics12245022

Academic Editor: Hideaki IIduka

Received: 22 November 2023 Revised: 10 December 2023 Accepted: 12 December 2023 Published: 15 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). devices in the actual deployment process. For instance, the SID [6] and SGN [2] methods require 562 GMAC and 2474 GMAC floating-point operations, respectively, to restore a single 4K resolution raw image. Providing such computational power is extremely challenging for edge devices in space missions. Therefore, we are committed to building a lightweight model that guarantees restoration quality.

Due to the sequential nature of most network architectures, where the output of one module serves as the input for the next module, the network inference speed is generally improved through hardware acceleration or by reducing the complexity of the network model at the cost of compromising the restoration quality. However, the hardware advantages have still not been fully exploited. To leverage the parallel computing characteristics of the hardware, we constructed a multi-scale parallel network that simultaneously processes different spatial scales after the image input. Additionally, we extensively employed 3×3 convolutional kernels to significantly enhance the network's inference speed while ensuring image quality.

CNN is a layered approach to data representation, where higher-level feature representations rely on lower-level feature representations. It progressively abstracts features in a hierarchical manner, extracting more advanced semantic information. However, once instantiated, the receptive field of a CNN is limited by the size of the convolutional kernel, making it challenging to comprehend global information effectively. Currently, obtaining global information often involves using transformer structures with a large number of parameters, which does not align with the requirements of lightweight network designs. Therefore, we propose utilizing frequency-domain information to enhance the network's ability to learn global features. By combining the frequency and spatial domains, we construct an end-to-end network that ensures a lightweight design while capturing both long-term and short-term information. This approach significantly improves the ability to restore low-light images.

In recent years, researchers have made significant progress in low-light image enhancement, deblurring, denoising, and other related areas [1–3,6–9]. These achievements typically assume that images are captured in somewhat dim environments with relatively moderate noise simulation. However, the lighting environment in space is complex and uneven, resulting in an uneven distribution of pixel intensity values in low-light images. A significant number of pixels have low intensity values, which leads to the loss of detailed pixel information in the feature data extracted by the deep learning network during training and inference. This occurs because the intensity values of the image pixels are too small, resulting in network feature data being rounded to zero during parameter propagation and thus the loss of detailed pixel information. To ensure that network architectures incorporating frequency-domain information can better improve the image restoration effects, we drew inspiration from the idea of A-rate thirteen-fold cross-validation and introduced an adaptive balancing mechanism to simulate the camera's ISO setting. Before feeding the image data into the deep learning network, we used an adaptive balance module to enhance the quality of the restored image. This approach significantly improved upon the limitations of solely amplifying the original image, as in SID [6], resulting in a greater enhancement in image quality.

In summary, our contributions are as follows:

- 1. We constructed a lightweight end-to-end network that processes image data obtained from the original sensor through multi-scale parallel processing, greatly reducing the image acquisition time;
- 2. The multi-scale image processing branch incorporates both spatial- and frequencydomain fusion networks, reducing the network parameters and computational complexity while ensuring image restoration quality;
- 3. We introduced an adaptive balancing mechanism as a pre-processing step to enhance the effectiveness of lightweight application models.

2. Related Work

Low-light imaging has received extensive research attention in recent years, and we will briefly introduce three aspects of this research: lightweight models, low-light image enhancement, and the frequency-domain processing of images.

2.1. Lightweight Models

The goal of lightweight networks is to reduce model parameterization and complexity while preserving model accuracy. They can be broadly categorized into model compression and lightweight network structure design. Model compression techniques such as pruning, quantization, and decomposition are used to decrease the size of the model. Network structure design involves designing more lightweight network architectures like MobileNet, ShuffleNet, and EfficientNet. The latest lightweight low-light image enhancement networks include DCE [3], LLPackNet [4], and RETNet [10]. DCE reduces the number of convolutional layers, but the increased processing at low scales still results in significant network latency. LLPackNet operates at higher scale spaces but yields subpar results. RETNet achieves a balance between restoration quality and inference speed by utilizing parallel inference with multiple-scale networks. However, its heavy reliance on CNNs neglects the extraction of global structural features. Therefore, by considering the pros and cons of these three network approaches, we adopted a multi-scale processing model suitable for hardware acceleration. The aim was to accelerate inference speed, reduce model parameters and complexity, and better adapt to spatial environments [11–15].

2.2. Low-Light Image Enhancement

In recent years, many excellent algorithms have emerged for low-light image enhancement. Traditional algorithms include methods such as histogram equalization [16] and gamma correction that simultaneously compress bright pixels and increase the brightness of dark regions. However, these methods often assume that the image already contains a good representation of the scene content and do not explicitly model image noise. Methods based on the Retinex model [17] decompose low-light images into reflection and illumination components, which are composed of priors or regularization. The estimated reflection component is considered as the enhanced result. However, inaccurate prior information may lead to the appearance of artifacts and color deviations in the enhancement results [18].

Deep learning methods have become popular for low-light image enhancement due to their enhanced accuracy, robustness, and speed. These methods include supervised learning (SL) approaches like LLNet [19], MBLLEN [20], Retinex-Net [17], LightenNet [21], and SICE [22]; reinforcement learning (RL) approaches like DeepExposure [23]; unsupervised learning (UL) approaches like Enlightenment [24]; and semi-supervised learning (SSL) approaches like DRBN [25]. These methods have opened up new avenues for low-light image enhancement through end-to-end deep networks. However, when considering extremely low-light imaging scenarios, such as space environments, where there is significant noise and color distortion and network inference speed and robustness are crucial, the aforementioned networks may not meet the practical deployment requirements.

Chen [6] created the SID dataset and proposed an end-to-end network for restoring high-definition RAW images from extremely low-light conditions. This work prompted several studies on recovering extremely low-light images [4,5], with most of them using U-Net as a reference network for restoration, achieving good restoration results. However, these methods require substantial processing time and memory usage. Wei [26] proposed a noise formation framework for synthesizing dark RAW images for CMOS sensors. Taking inspiration from this, we developed a lightweight low-light imaging system based on the original sensor's RAW image data.

2.3. Frequency-Domain Processing of Images

Frequency-domain processing is an image processing technique that involves transforming an image from the spatial domain to the frequency domain for operations such as filtering, enhancement, and compression. The frequency domain reflects the degree of spatial variations in an image. Common frequency-domain processing methods include Fourier transform and wavelet transform. In recent years, many computer vision tasks have employed frequency-domain-based methods. FDA [27] swaps low-frequency spectra between images to mitigate the impact of style variations on image segmentation. LaMa [28] applies the structure of fast Fourier convolution to image rendering. GFNet [29] learns long-range spatial dependencies in the frequency domain for image classification. SDWNet [30] introduces wavelet transform into deep networks. Considering the unique characteristics of the spatial environment, we propose a deep neural network that merges the frequency domain and spatial domain of the image. This approach aims to introduce global features, capture long-term and short-term information, and enhance the quality of the restored images while minimizing the number of model parameters and computational requirements.

3. Methods

3.1. Network Architecture

A low-light imaging system needs to ensure a reliable restoration quality, higher memory efficiency, faster computational speed, and deployability on edge devices. Currently, most network models enhance low-light images by applying low-light image enhancement techniques to RGB images processed by cameras. However, the camera algorithm processing of the original data can result in the loss of certain information. Therefore, we developed the end-to-end solution depicted in Figure 1, where the original image data from the imaging system are used as the input to generate the corresponding RGB-format image.



Figure 1. Network backbone structure. The network employs an end-to-end solution with multi-scale parallel processing, taking the input of raw image data from the imaging system and generating corresponding images in RGB format.

As shown in Figure 2, due to the 2×2 Bayer pattern used for stitching the raw images from the camera, specific color information is included. To decouple the four color channels and perform downsampling on the matrix data separately, we utilized Pixel-Shuffle. The



downsampling factors were 2, 8, and 32, and the three branches were processed in parallel, as illustrated in Figure 1.

Figure 2. The camera's raw image was stitched using a 2×2 Bayer pattern, which contained specific color information. We utilize Pixel-Shuffle to separate the four color channels.

As shown in Figure 1, in the low-dimensional branching (LDB) branch, the data dimension was reduced to 4 and the image size was (H/2, W/2) after downsampling with a factor of 2. Due to the high image resolution, traditional CNNs have weaker global feature learning abilities. Therefore, we introduced frequency-domain processing inspired by Res FFT-Conv [31] and constructed the ResFFT module. The details of ResFFT will be discussed in Section 3.2. Similarly, in the medium-dimensional branching (MDB) branch (Figure 3), the data dimension was 64 and the image size was (H/4, W/4) after downsampling with a factor of 4. The ResFFT module was also introduced to enhance the network's global feature learning abilities with frequency-domain information. By combining the frequency domain and spatial domain, the network can capture long-term and short-term information while remaining lightweight.

As shown in Figure 1, in the high-dimensional branching (HDB) branch, after downsampling with a factor of 32, the data dimension became 1024, and the image size was (H/32, W/32). Since the resolution in the high-dimensional branch was the lowest, adding frequency-domain processing would result in excessive computation with minimal effects. Therefore, the focus was on the high-dimensional features of the image. We performed initial data processing using Conv-block and then adopted the Multi-RDB module, as shown in Figure 3. The Multi-RDB module was constructed by densely connecting several RDB (residual dense block) modules. It comprehensively learned the high-dimensional feature information of the original data, enhanced the utilization of information in the original data, and thus improved the model's performance.

After the HDB branch, the data were upsampled four times using Pixel-Shuffle and concatenated with the data from the MDB branch (Figure 4). The concatenated data were then combined again with the data from the LDB branch to obtain a 24-dimensional matrix. After adjusting the dimensions using convolutional kernels and upsampling with a factor of 2, the low-light enhanced RGB image was obtained. In the overall network architecture, we followed the principles outlined in RepVGG [32], on a GPU, the computational density of a 3×3 convolution is four times higher (theoretical computation divided by the time taken) compared to 1×1 and 5×5 convolutions [33]. Therefore, the entire network was constructed using 3×3 convolutional kernels. In the HDB branch, to deal with the higher parameter dimensionality, we employed grouped convolution to reduce the computational complexity.



Figure 3. Multi-RDB module, formed by dense residual connections with several RDB (residual dense block) modules.



Figure 4. Res-FFT module, formed by the addition of the Conv-block, frequency-domain branch, and original image branch.

3.2. Frequency-Domain Processing

Discrete Fourier transform (DFT) is the cornerstone of modern digital signal processing, with its one-dimensional mode shown in Equation (1):

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j\frac{2\pi}{N}kn}$$
(1)

x[n] is a complex sequence, where X[k] represents the spectrum at frequency $\omega_k = \frac{2\pi}{N}kn$ and *j* represents the imaginary unit. It is evident that the spectrum at any frequency carries global information. This introduces the two-dimensional Fourier transform, which is applicable to images. Its formula is shown in Equations (2) and (3):

$$F(u,v) = \sum_{x=0}^{M-1} \left(\sum_{y=0}^{N-1} f(x,y) e^{-j2\pi \frac{ux}{M}}\right) e^{-j2\pi \frac{vy}{N}}$$
(2)

$$f(x,y) = \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u,v) e^{j2\pi(\frac{ux}{M} + \frac{vy}{N})}$$
(3)

We can consider the two-dimensional discrete Fourier transform (DFT) as two consecutive one-dimensional DFTs, as in Equation (2). Based on the fast computation algorithm FFT (fast Fourier transform) for the one-dimensional DFT, the two-dimensional Fourier transform can be converted into two successive one-dimensional FFTs, thereby enabling fast conversion between the spatial domain and the frequency domain.

As shown in Figure 5, in the same scene, high-exposure images contain more frequencydomain information compared to low-exposure images. Inspired by Res FFT-Conv [31], we introduced ResFFT into our network for frequency-domain processing, capturing both the long-term and short-term interactions and learning the global structural information in the image. As shown in Figure 4, ResFFT is composed of a Conv-block, a frequency-domain branch, and the original image branch. By incorporating the LeakyReLU function in the branches to adjust the thresholds, the information from both the spatial and frequency domains is learned within the same module to some extent, avoiding the deficiency of global feature learning in CNNs.



Figure 5. Comparison of high-illumination images and low-illumination images in the frequency domain.

3.3. Adaptive Balance

Due to the low photon count and low signal-to-noise ratio in low-exposure images, the intensity values of the original image data are typically low. This can lead to information loss during the training and inference process of neural networks. SID [6] proposed a front-end amplifier that multiplies all pixel values in the image by a factor of Amp, as in Equation (4):

$$Amp = m \cdot \left(\frac{\sum_{i,j} x_{i,j}}{H \cdot W}\right)^{-1} \tag{4}$$

where $x_{i,j}$ represents the pixel intensity and m is a hyper-parameter that constrains the amplification factor to within [0, 1]. However, low-light images based on the space environment differ from this scenario. Low-light images in space often have saturated highlights, which can introduce halo artifacts. The low-photon count is uniform, resulting in an overall low image intensity. This amplification factor is not suitable for these conditions. One needs a higher amplification factor for low intensity values compared to high intensity values, making the amplification curve approximate a logarithmic function. Drawing inspiration from the A-law compression used in PCM encoding in communication, we approximated the logarithmic function using a segmented linear approach. Smaller values are assigned larger weights, while larger intensity values are assigned smaller weights, as shown in Figure 6.



Figure 6. Weight allocation demonstration.

According to the principle of dividing into thirteen parts with the A rate, we divide the interval [0, 1] into eight equal parts, and the interval of each part is:

$$\left(\frac{2^5 \cdot (n-1)}{2^8}, \frac{2^5 \cdot n}{2^8}\right) n \in [0, 8]$$
(5)

By extending the interval [0, 1] to [0, 255], we obtain the following:

$$(2^5 \cdot (n-1), 2^5 \cdot n) \ n \in [0, 8]$$
 (6)

We set b_n as an intermediate variable:

$$b_n = 2^5 \cdot n, n \in [0, 8] \tag{7}$$

$$w_{i,j} = 2^{5-n}, b_n < x_{i,j} < b_{n+1}, if \quad n = 0, w_{i,j} = 16$$
(8)

$$b_{i,j} = b_{i,n-1} = 2^5 \cdot (n-1), b_n < x_{i,j} < b_{n+1}, if \quad n = 0, b_{i,j} = 0$$
(9)

Due to the common use of 8-bit encoding in cameras, we can derive Equations (7) and (8), where $w_{i,i}$ is the weighting parameter, and $b_{i,i}$ is the offset. The pixel intensity values of the image after adaptive balance are calculated as in Equation (10), where m is a hyperparameter that controls the overall brightness:

$$X' = m \cdot (W \cdot X + B_{n-1}) = m \cdot \begin{pmatrix} w_{1,1}x_{1,1} + b_{1,1} & w_{1,2}x_{1,2} + b_{1,2} & \cdots & w_{1,j}x_{1,j} + b_{1,j} \\ w_{2,1}x_{2,1} + b_{2,1} & w_{2,2}x_{2,2} + b_{2,2} & \cdots & w_{2,j}x_{2,j} + b_{2,j} \\ \vdots & \vdots & \ddots & \vdots \\ w_{i,1}x_{i,1} + b_{i,1} & w_{i,2}x_{i,2} + b_{i,2} & \cdots & w_{i,j}x_{i,j} + b_{i,j} \end{pmatrix}$$
(10)

. 1

4. Experimental Results

4.1. Experimental Settings

We used the PyTorch framework to train and test our network on an AMD R7-6800H and RTX3060. We employed the Adam optimizer for training, conducting 1 million iterations. The learning rate was set to 0.001 for the first 250,000 iterations and then reduced by a factor of 10 every 250,000 iterations. During the training process, we utilized the random cropping of 512 \times 512 patches and weight normalization to constrain all convolutional layers. During the training process, we employed a weighted combination of three types of losses, as defined in Equation (11), to form the *Loss*_{total}:

$$Loss_{total} = \lambda_1 \cdot Loss_1 + \lambda_2 \cdot Loss_2 + \lambda_3 \cdot Loss_3 \tag{11}$$

where L1 loss is referred to as *Loss*₁, SSIMLoss is referred to as *Loss*₂, and we indicate the utilization of high-dimensional features extracted by VGG as Loss₃.

We used the low-light SID dataset [6], which consists of outdoor scenes where the illuminance range of the camera is typically between 0.2 lux and 5 lux. In indoor scenes, the camera illuminance generally falls between 0.03 lux and 0.3 lux. The dataset includes short-exposure raw images and corresponding long-exposure ground-truth (GT) images for benchmark testing. It comprises two sub-datasets, with the Sony dataset having an image resolution of 4256×2848 pixels and the Fuji dataset having an image resolution of 6032×4032 pixels. Our main focus was on testing via the Sony dataset.

We compared three publicly available code solutions for low-light image enhancement: SID, LDC, and SGN. In Table 1, we used multiple evaluation metrics related to practical deployability: the number of model parameters, number of MAC operations, inference speed, and restoration quality. Since most reported inference speed metrics are based on high-performance GPUs, and our objective was reliable deployability on resourceconstrained edge devices with limited computational power, we tested the CPU inference speed and the number of GMAC operations. Therefore, we provide a comparison of both the CPU and GPU inference speeds in the report.

Table 1. Compared to networks with large parameter and computational requirements, we made progress in terms of inference speed and computational efficiency. Moreover, the image restoration quality reached a satisfactory level. The bold font signifies that we have achieved the optimal level in the corresponding indicator.

		SID	LDC	SGN	Ours
Parameters (million)		7.78	8.6	3.5	0.891
GMACs	Sony	562.06	>2000	>2000	60.5
	Fuji	1273.64	>2000	>2000	121.07
CPU inference times (s)	Sony	4.21	>50	20.52	0.96
	Fuji	8.11	>50	>50	1.78
GPU inference times (ms)	Sony	197.66	>1500	1113.85	57.92
	Fuji	384.87	>1500	>1500	106.26
PSNR (dB)/SSIM	Sony	28.88/0.787	29.56/0.799	28.91/0.789	28.85/0.794
	Fuji	26.61/0.680	26.70/0.681	26.90/0.683	26.62/0.681

Based on the data in Table 1, our network showed significant improvements across all five metrics. The SID, LDC, and SGN networks had 4–10 times more parameters and significantly more GMAC operations than our network, but we could maintain comparable restoration quality. Furthermore, our inference speed was 4–10 times faster than the aforementioned three networks and was acceptable for edge devices in space environments.

Based on the data comparison in Table 2, our lightweight low-light enhancement network outperformed LLPackNet in all five metrics. Taking the Sony dataset as an example, we achieved improvements of 1.02 dB/0.044 in the PSNR/SSIM with fewer parameters than LLPackNet. Compared to the RETNet network, with an increase of 0.11 M parameters and minimal improvement in the floating-point operation metric GMACs, we observed improvements of 0.19 dB/0.004 in the PSNR/SSIM.

Table 2. Comparison to the two currently known best-performing lightweight networks. The bold font signifies that we have achieved the optimal level in the corresponding indicator.

		LLPackNet	RETNet	Ours
Parameters (million)		1.16	0.785	0.891
GMACs	Sony	83.46	59.8	60.5
	Fuji	166.12	119.66	121.07
CPU inference times (s)	Sony	1.73	0.76	0.96
	Fuji	3.25	1.4	1.78
GPU inference times (ms)	Sony	70.96	46.24	57.92
	Fuji	138.34	93.83	106.26
PSNR (dB)/SSIM	Sony	27.83/0.75	28.66/0.790	28.85/0.794
	Fuji	24.13/0.59	26.60/0.682	26.62/0.681

To enhance the effectiveness of the entire network in the space environment, we constructed a dataset consisting of 1574 low-light satellite images with a resolution of 4256×2848 , using a methodology similar to that described for building the datasets in the SID paper. In comparison to models with a larger number of parameters like SID, LDC, and SGN, our network maintained good restoration quality even when achieving a 4–10 times faster inference speed. Furthermore, compared to LLPackNet, we significantly improved both the inference speed and restoration quality. In the case of the RETNet network,

there were minimal differences in the inference time, but we observed improvements of 0.59/0.021 in the PSNR/SSIM. Our constructed network exhibited superior global structural learning capabilities due to the more regular satellite structures in our low-light satellite dataset, even surpassing the SID network in the corresponding indicators.

As shown in Table 3, in order to better evaluate the color reconstruction performance of our network framework, we adopted a dataset construction method similar to that described in the SID paper and captured 30 sets of low-light images of the Macbeth color checker under different illuminations, along with their corresponding high-light images. By conducting inference using our network trained on both the SID and our own satellite datasets, the restored images achieved an average PSNR of 28.79 dB and SSIM of 0.765, surpassing the performance metrics of both the SID dataset and our self-constructed satellite datasets. The specific results are shown in Figure 7, in which only the color block information of the Macbeth color checker was retained for better comparison.

Table 3. The comparison between CPU inference speed and restoration quality in our constructed satellite dataset and 30 groups of images of the Macbeth color checker captured under different illuminations.

		SID	LDC	SGN	LLPackNet	RETNet	Ours
CPU inference	times (s)	4.21	>50	20.52	1.73	0.76	0.96
PSNR (dB)/SSIM	Satellite Macbeth	26.53/0.677 28.48/0.736	26.95/0.679 28.97/0.763	26.75/0.672 28.78/0.738	25.35/0.632 27.12/0.657	26.13/0.653 28.13/0.729	26.72/0.674 28.79/0.765



Figure 7. Comparison of low-light image restoration. The **first row** shows the ground-truth (GT) images, the **second row** displays the low-light images, and the **third row** exhibits the restored images. The first four columns represent the restoration results of the SID dataset, the fifth column represents the restoration results of the satellite dataset, and the sixth column represents the restoration results of the Macbeth color checker constructed by our team under low illumination.

It is evident that our network achieved excellent restoration results on low-light images. The excellent performance of our proposed lightweight framework is attributed to the multi-scale parallel processing network. In the LDB and MDB branches, we constructed the ResFFT module to fuse the spatial and frequency domains and learn the global and local features of the image, capturing both long-term and short-term information. In the HDB branch, we adopted the multi-RDB module to focus on the high-dimensional spatial features of the image, achieving effects similar to methods like SID and SGN.

Additionally, due to the non-uniform illumination in low-light images and the predominantly low pixel values, we introduced an adaptive balance pre-processing module. This module segmented and linearly amplified the pixel values to avoid the loss of detailed feature information caused by the gradient operations becoming zero during network inference. It greatly enhanced the stability and effectiveness of the network. Furthermore, the multi-scale parallel processing and significant reduction in parameter quantity resulted in a substantial increase in inference speed. Our model achieved a balance between inference speed and restoration quality, making it more suitable for deployment on hardware devices in space.

4.2. Ablation Experiment

In our network configuration, the ResFFT module and adaptive equalization module were implemented to verify their effectiveness. We conducted experiments by removing the frequency-domain branches of the ResFFT module and the adaptive equalization module separately, and the results are shown in Table 4.

Table 4. The results of the ablation experiments were obtained by separately removing the frequencydomain branch and the adaptive equalization module.

	PSNR (dB)/SSIM	Decreasing Value	
Original network architecture	28.85/0.794	0/0	
Removing the frequency-domain modules	28.13/0.772	0.72/0.22	
Removing the adaptive balancing module	28.31/0.785	0.54/0.09	

When we disabled all frequency-domain branches and trained the network on the SID Sony dataset, the PSNR/SSIM decreased from 28.85 dB/0.794 to 28.13 dB/0.772, resulting in a decrease of 0.72 dB/0.22.

Similarly, when we excluded the adaptive equalization module during the network training process, the PSNR/SSIM dropped from 28.85 dB/0.794 to 28.31 dB/0.785, with a decrease of 0.54 dB/0.09.

These results indicate the effectiveness of both the frequency-domain branches of the ResFFT module and the adaptive equalization module in maintaining the network performance.

5. Conclusions

In this paper, we developed a fast and lightweight model for low-light imaging. The original image data were fed into a parallel multi-scale network, allowing for innovative pixel-level and global feature learning in both the spatial and frequency domains. By introducing a pre-adaptive balancing mechanism, we enhanced the effectiveness of the model in adapting to different light levels. When compared to models with parameter quantities 4–10 times larger than ours, our model achieved 4–20 times faster inference speeds on the CPU and GPU while preserving recovery quality on the SID dataset and our custom low-light satellite image dataset. Furthermore, our model achieved superior recovery results with only a marginal increase of 0.7 in the GMAC operations compared to existing lightweight models. This fulfills the requirements of real-time performance and deployability for space missions. Ablation experiments confirmed the effectiveness of our proposed network model compared to solely using CNNs for spatial-domain processing and highlighted the importance of adaptive balancing for the entire system.

This paper presents a pre-processing step for space missions, providing improved generalization for subsequent tasks such as target tracking and non-cooperative target pose estimation algorithms in space applications.

Author Contributions: Conceptualization, H.Z.; methodology, J.W. and B.L.; software, J.W., Q.L. and J.D.; validation, J.W. and Z.H.; investigation, H.Z.; writing—original draft preparation, J.W.; writing—review and editing, H.Z., J.C. and H.W.; funding acquisition, H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Shaanxi provincial fund 2023-YBGY-234.

Data Availability Statement: Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Zhang, X.; Shen, P.; Luo, L.; Zhang, L.; Song, J. Enhancement and noise reduction of very low light level images. In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), Tsukuba, Japan, 11–15 November 2012; pp. 2034–2037.
- 2. Gu, S.; Li, Y.; Van Gool, L.; Timofte, R. Self-Guided Network for Fast Image Denoising. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27–28 October 2019; pp. 2511–2520. [CrossRef]
- Xu, K.; Yang, X.; Yin, B.; Lau, R.W. Learning to Restore Low-Light Images via Decomposition-and-Enhancement. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 2278–2287. [CrossRef]
- Atoum, Y.; Ye, M.; Ren, L.; Tai, Y.; Liu, X. Color-wise Attention Network for Low-light Image Enhancement. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; pp. 2130–2139. [CrossRef]
- Ai, S.; Kwon, J. Extreme Low-Light Image Enhancement for Surveillance Cameras Using Attention U-Net. Sensors 2020, 20, 495. [CrossRef] [PubMed]
- 6. Chen, C.; Chen, Q.; Xu, J.; Koltun, V. Learning to See in the Dark. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3291–3300. [CrossRef]
- 7. Remez, T.; Litany, O.; Giryes, R.; Bronstein, A.M. Deep Convolutional Denoising of Low-Light Images. *arXiv* 2017, arXiv:1701.01687.
- Gu, S.; Zhang, L.; Zuo, W.; Feng, X. Weighted Nuclear Norm Minimization with Application to Image Denoising. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 2862–2869. [CrossRef]
- Maharjan, P.; Li, L.; Li, Z.; Xu, N.; Ma, C.; Li, Y. Improving Extreme Low-Light Image Denoising via Residual Learning. In Proceedings of the IEEE International Conference on Multimedia and Expo, ICME 2019, Shanghai, China, 8–12 July 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 916–921. [CrossRef]
- 10. Lamba, M.; Mitra, K. Restoring Extremely Dark Images in Real Time. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 3486–3496. [CrossRef]
- 11. Huang, Q. Towards Indoor Suctionable Object Classification and Recycling: Developing a Lightweight AI Model for Robot Vacuum Cleaners. *Appl. Sci.* 2023, *13*, 10031. [CrossRef]
- 12. Hsia, C.H.; Lee, Y.H.; Lai, C.F. An Explainable and Lightweight Deep Convolutional Neural Network for Quality Detection of Green Coffee Beans. *Appl. Sci.* **2022**, *12*, 10966. [CrossRef]
- 13. Huang, Q. Weight-Quantized SqueezeNet for Resource-Constrained Robot Vacuums for Indoor Obstacle Classification. *AI* 2022, 3, 180–193. [CrossRef]
- 14. Tang, Z.; Luo, L.; Xie, B.; Zhu, Y.; Zhao, R.; Bi, L.; Lu, C. Automatic Sparse Connectivity Learning for Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* 2023, 34, 7350–7364. [CrossRef] [PubMed]
- 15. Hu, W.; Che, Z.; Liu, N.; Li, M.; Tang, J.; Zhang, C.; Wang, J. CATRO: Channel Pruning via Class-Aware Trace Ratio Optimization. *IEEE Trans. Neural Netw. Learn. Syst.* 2023, 1–13. [CrossRef] [PubMed]
- 16. Ibrahim, H.; Pik Kong, N.S. Brightness Preserving Dynamic Histogram Equalization for Image Contrast Enhancement. *IEEE Trans. Consum. Electron.* **2007**, *53*, 1752–1758. [CrossRef]
- 17. Wei, C.; Wang, W.; Yang, W.; Liu, J. Deep Retinex Decomposition for Low-Light Enhancement. arXiv 2018, arXiv:1808.04560.
- 18. Li, C.; Guo, C.; Han, L.; Jiang, J.; Cheng, M.M.; Gu, J.; Loy, C.C. Low-Light Image and Video Enhancement Using Deep Learning: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 9396–9416. [CrossRef]
- 19. Lore, K.G.; Akintayo, A.; Sarkar, S. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognit.* **2017**, *61*, 650–662. [CrossRef]
- Lv, F.; Lu, F.; Wu, J.; Lim, C. MBLLEN: Low-light Image/Video Enhancement Using CNNs. In Proceedings of the British Machine Vision Conference, Newcastle, UK, 3–6 September 2018.
- 21. Li, C.; Guo, J.; Porikli, F.; Pang, Y. LightenNet. Pattern Recogn. Lett. 2018, 104, 15–22. [CrossRef]
- 22. Cai, J.; Gu, S.; Zhang, L. Learning a Deep Single Image Contrast Enhancer from Multi-Exposure Images. *IEEE Trans. Image Process.* 2018, 27, 2049–2062. [CrossRef] [PubMed]
- Yu, R.; Liu, W.; Zhang, Y.; Qu, Z.; Zhao, D.; Zhang, B. DeepExposure: Learning to Expose Photos with Asynchronously Reinforced Adversarial Learning. In Proceedings of the 32nd International Conference on Neural Information Processing Systems, NIPS'18, Montreal, QC, Canada, 3–8 December 2018; Curran Associates Inc.: Red Hook, NY, USA, 2018; pp. 2153–2163.
- Jiang, Y.; Gong, X.; Liu, D.; Cheng, Y.; Fang, C.; Shen, X.; Yang, J.; Zhou, P.; Wang, Z. EnlightenGAN: Deep Light Enhancement Without Paired Supervision. *IEEE Trans. Image Process.* 2021, 30, 2340–2349. [CrossRef] [PubMed]
- Chi, L.; Tian, G.; Mu, Y.; Xie, L.; Tian, Q. Fast Non-Local Neural Networks with Spectral Residual Learning. In Proceedings of the 27th ACM International Conference on Multimedia, MM '19, Nice, France, 21–25 October 2019; ACM: New York, NY, USA, 2019; pp. 2142–2151. [CrossRef]
- Wei, K.; Fu, Y.; Yang, J.; Huang, H. A Physics-Based Noise Formation Model for Extreme Low-Light Raw Denoising. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 2755–2764. [CrossRef]

- Yang, Y.; Soatto, S. FDA: Fourier Domain Adaptation for Semantic Segmentation. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 4084–4094. [CrossRef]
- Suvorov, R.; Logacheva, E.; Mashikhin, A.; Remizova, A.; Ashukha, A.; Silvestrov, A.; Kong, N.; Goka, H.; Park, K.; Lempitsky, V. Resolution-robust Large Mask Inpainting with Fourier Convolutions. In Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–8 January 2022; pp. 3172–3182. [CrossRef]
- Rao, Y.; Zhao, W.; Zhu, Z.; Lu, J.; Zhou, J. Global Filter Networks for Image Classification. In *Advances in Neural Information Processing Systems*; Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W., Eds.; Curran Associates, Inc.: Brooklyn, NY, USA, 2021; Volume 34, pp. 980–993.
- Zou, W.; Jiang, M.; Zhang, Y.; Chen, L.; Lu, Z.; Wu, Y. SDWNet: A Straight Dilated Network with Wavelet Transformation for image Deblurring. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 11–17 October 2021; pp. 1895–1904. [CrossRef]
- 31. Mao, X.; Liu, Y.; Shen, W.; Li, Q.; Wang, Y. Deep Residual Fourier Transformation for Single Image Deblurring. *arXiv* 2021, arXiv:2111.11745.
- Ding, X.; Zhang, X.; Ma, N.; Han, J.; Ding, G.; Sun, J. RepVGG: Making VGG-style ConvNets Great again. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 13728–13737. [CrossRef]
- Cuda c++ Best Practices Guide. Available online: https://docs.nvidia.com/cuda/pdf/CUDA_C_Best_Practices_Guide.pdf (accessed on 30 November 2020).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.