Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

A ground-based dataset and a diffusion model for on-orbit low-light image enhancement*

Yiman ZHU[1], Lu WANG‡[1], Jingyi YUAN[1], Yu GUO[1]

[1]*School of Automation, Nanjing University of Science and Technology, Nanjing 210000, China*

E-mail: yiman@njust.edu.cn; wanglu21@njust.edu.cn; jingyi@njust.edu.cn; guoyu@njust.edu.cn

Received Apr. 07, 2024; Revision accepted Sept. 10, 2024; Crosschecked

Abstract: On-orbit service is important for maintaining the sustainability of space environment. A space-based visible camera is an economical and lightweight sensor for situational awareness during on-orbit service. However, it can be easily affected by the low illumination environment. Recently, deep learning (DL) has achieved remarkable success in image enhancement of natural images, but is seldom applied in space due to the data bottleneck. In this article, we first propose a dataset of the Beidou Navigation Satellite for on-orbit low-light image enhancement (LLIE). In the automatic data collection scheme, we focused on reducing the domain gap and improving the diversity of the dataset. We collected hardware in-the-loop images based on a robotic simulation testbed imitating space lighting conditions. To evenly sample poses of different orientation and distance without collision, a collision-free working space and pose-stratified sampling is proposed. Then, a novel diffusion model is proposed. To enhance the image contrast without over-exposure and blurred details, we designed fused attention guidance (FAG) to highlight the structure and dark region. Finally, a comparison of our method with previous methods indicates that our method has better on-orbit LLIE performance.

Unedited

Key words: Satellite capture; Low-light image enhancement; Data collection; Diffusion model; Fused attention
https://doi.org/10.1631/FITEE.2400261 **CLC number:** TP

1 Introduction

Due to the high frequency (HF) of space activities, the space environment has been seriously degraded as a by-product of these activities. As of February 2022, more than 25,000 space objects have been identified, including retired satellites, spacecraft, rocket bodies, and debris (Cowardin and Miller, 2022), which gravely threaten both the functioning and newly launched spacecraft (Ledkov and Aslanov, 2022). Hence, to ensure the long-term sustainability of the space environment, clean up existing space debris, and extend the lifespan of operational spacecraft, active debris removal (ADR) and on-orbit servicing (OOS) have emerged as popular research areas (Poozhiyil et al., 2023). The space robotic arm plays a vital role in performing various tasks associated with ADR and OOS, including capturing, docking, repairing, and refurbishment, as demonstrated in Fig. 1(a). To realize safe and efficient robotic control, reliable sensors and data analysis are required to provide better space situational awareness (SSA), especially when manipulating unknown targets (Harris et al., 2021; Civardi et al., 2024). A space-based visible camera is one of the most essential sensors because of its advantages: it is lightweight in mass, compact in size, economical in power and informative in data (Diao et al., 2011). However, the variable illumination in space can seriously affect the quality of the captured images, es-

‡ Corresponding author

* Project supported by the Postgraduate Research & Practice Innovation Program of Jiangsu Province (No. KYCX23_0481), the National Natural Science Foundation of China (No. 62403242), the Fundamental Research Funds for the Central Universities (No. 30924010932), and the Natural Science Foundation of Jiangsu Province (No. BK20241459)

ORCID: Yiman ZHU, http://orcid.org/0000-0002-7421-0188

© Zhejiang University Press 2024

(a) (b) (c)

Fig. 1 Illustration of on-orbit service and low-light images: (a) demonstrates service star refueling the target star; (b) and (c) are the real images of low-light and normal-light captured by a visible camera during the PRISMA mission (Capuano et al., 2019).

pecially when the satellite is in the earth's shadow, as shown in Figs. 1(b) and 1(c). The captured images can be extremely invisible, making it difficult to extract key features. To recover the buried details and improve data usability for downstream tasks and surveillance efficiency of SSA, we propose a novel framework to solve the problem of on-orbit low-light image enhancement (LLIE).

Image enhancement methods based on classic theory (Rahman et al., 1996; Guo et al., 2017; Ying et al., 2017; Li et al., 2018; Ren et al., 2020) have been widely used to improve the image contrast. Unfortunately, these methods are only suitable for situations where images already contain a good representation of the scene content. For extremely dark and noised images, they are likely to cause severe noisy amplification after enhancement.

Deep learning (DL) methods have been applied to solve problems in various industries, demonstrating significant advancements (Li et al., 2022; Xu et al., 2024, 2023). DL-based methods for LLIE can be classified as one-way neural networks (Shi et al., 2024; Wei et al., 2018; Li and Zhang, 2019; Xu et al., 2021; Dang et al., 2023; Huang et al., 2023; Yan et al., 2024) and generative adversarial networks (GANs) (Chen et al., 2018; Rao et al., 2021; Jiang et al., 2021; Triantafyllidou et al., 2020; Fu et al., 2022). One-way neural networks refer to networks that have only one component. These methods are more dependent on training data and the loss function. GANs can produce accurate and realistic images on more complex tasks owing to the adversarial training mechanism, in which a generator and a discriminator compete with each other until reaching binary equilibrium. However, this mechanism makes the networks collapse more easily in the early stages of training.

Recently, diffusion models have brought the generative models to a new level (Ho et al., 2020; Nichol and Dhariwal, 2021; Chen et al., 2024). In contrast to GANs, which lean on inner-loop maximization, diffusion models are guided on a concise and well-constructed loss, making convergence easier. The denoising diffusion probabilistic model (DDPM) represents a category of deep generative models that have (i) a forward diffusion stage, in which the input data is gradually perturbed over several steps by adding Gaussian noise, and (ii) a reverse denoising stage, in which a generative model is tasked with recovering the original input data from the diffused data by learning to gradually reverse the diffusion process (Croitoru et al., 2023). The generated images exhibit very few artifacts and a high level of details. To date, diffusion models have been successfully applied to many image restoration tasks like image super-resolution (Saharia et al., 2023), image inpainting (Saharia et al., 2022), and image enhancement (Zhou et al., 2023; Hou et al., 2024; Wang et al., 2023).

However, most of these methods are based on natural public images. When brightening on-orbit low-light images, problems still remain, such as the insufficient enhancement, amplified noise, and artificial features. As is commonly known, training data is imperative to the performance of DL models, which brings the demand for datasets in space research to the forefront. However, to the best of our knowledge, there is neither a real-world nor synthetic dataset for on-orbit LLIE tasks because of the difficulty of capturing images in space. In addition, tasks in the computer vision community pay more attention to color fidelity and human-eye habits. For on-orbit missions, it is more important to save computing resources and provide accurate data for downstream tasks like visual measurement and estimation. Images of satellites in space also have different distribution because of their geometric shapes and metal surfaces. Therefore, it is more difficult to precisely reveal the structure and texture details from the dark region without generating more noise.

To solve these problems, in this article, we focus on a spacecraft with a solar panel, using the Beidou Navigation Satellite as an example. An automatic dataset collection scheme based on a 6-DoF robot is devised to tackle the data bottleneck for LLIE in

Unedited

the space environment. Based on the dataset, we propose a novel attention-guided diffusion model to achieve enhancement of on-orbit low-light satellite images, which is proved to be more effective than previous methods. There are four main-contributions of this paper:

- Unlike generating synthetic images as in previous methods, we first construct a dataset for on-orbit LLIE on a hardware-in-the-loop (HIL) testbed, which narrows the domain gap between the real world and training set.
- To reduce manpower and ensure the safety and quality of data collection, we designed a pose-stratified sampling based on the collision-free working space obtained with a physics engine.
- To the best of our knowledge, this is the first novel diffusion model proposed for LLIE in the space environment. Compared with the classic methods and other state-of-the-art DL methods, our method shows better performance.
- Distinguished from the existing diffusion model for image restoration, fused attention guidance (FAG) is devised to extract illumination and structural distribution information during the downsampling stage of training to guide light enhancement and detail preservation.

2 Related Work

2.1 Space dataset

Images in the space environment are scarce and hard to obtain, which hinders the application of intelligent methods. Some efforts have been made to change this situation. The methods of acquiring datasets can be divided into two classes: capturing real images with different exposure and rendering images with a physical engine. Yang et al. (2022) proposed the BUAA-SID1.0 database, which used 3ds Max software to render images from a full viewpoint based on a CAD model of 56 satellites. SPEED is the first publicly available machine learning dataset for spacecraft pose estimation, which consists of synthetic images generated with non-physically based rendering and a few real pictures taken in space missions (Kisantal et al., 2020). Proença and Gao (2020) proposed the realistic rendering dataset URSO, containing labeled images of spacecraft orbiting the earth. Hu et al. (2021) introduced the SwissCube dataset created via physically based rendering to reflect the illumination in space. Musallam et al. (2021) provided the synthetic dataset SPARK, composed of 150,000 annotated multi-modal images aimed at space target recognition and detection. For generalized enhancement learning, Xu et al. (2021) introduced mapping functions to generate images of different exposure. Most datasets are synthetic because real images are extremely hard to obtain. The domain gap between synthetic images and real-scene images may lead to relatively poor performance of the same architectures in real situations (Proença and Gao, 2020; Park et al., 2023). Although some methods have used physically based materials, the domain gap still exists. In addition, there is no public dataset for space object LLIE. In this article, we build a ground-test platform and an automatic collection scheme to make a HIL image dataset for LLIE, which practically addresses this problem.

2.2 LLIE

The early works on LLIE can be typically categorized as histogram equalization-based methods and Retinex model-based methods. Histogram equalization-based methods improve the contrast of an image by balancing the histogram of the entire image. Retinex model-based methods (Rahman et al., 1996; Guo et al., 2017; Li et al., 2018; Ying et al., 2017; Ren et al., 2020) are based on the assumption that a color image can be decomposed into reflectance and illumination. Li et al. (2018) and Ren et al. (2020) improved the noise term in the Retinex model. Guo et al. (2017) transferred the illumination estimation into an optimization problem.

With advancement of low-light data collection, DL-based methods have been proposed. One-way neural networks for LLIE involve only one component during training. Lore et al. (2017) designed a deep autoencoder to learn features by training with synthetically modified images from an Internet database. To achieve good enhancement results on multiple metrics, Shi et al. (2024) designed a UNet++ network trained with a hybrid loss function that emphasized multiple attributes of an image. Some scholars take the interpretability of DL into consideration and integrate the principles of image decomposition with DL-based methods. Based

Unedited

on the Retinex model, Wei et al. (2018) and Wang et al. (2019) used two convolutional neural networks (CNNs) modeling the illumination map and the reflectance map to reconstruct the normal images. Yan et al. (2024) proposed a trainable color space horizontal/vertical/intensity (HVI), then designed a two-branch network to decouple color and intensity. Apart from CNNs, other network architectures have been introduced in LLIE to improve the network ability of feature extraction. Dang et al. (2023) constructed the Wave Transform Block, which represented pixels in a dynamic wave form with signal function to enhance the feature representation of degraded images. Cui et al. (2022) applied an encoder composed of transformers to realize global feature extraction and expand the receptive field in the feature extraction stage. Combining the advantages of both the decomposition principle and new architecture, Cai et al. (2023) formulated a one-stage Retinex-based transformer to model the long-range dependences under the guidance of illumination information. Brateanu et al. (2024) utilized the YUV color space along with transformers to disentangle light and color information in images. However, such improvement on feature extraction is forced at the expense of computational cost. In addition, oneway neural networks count more on training data and supervised loss function. GANs introduced a different learning mechanism, which adaptively adjusts the generator to learn the image distribution better through competition with the discriminator. To realize image enlightenment with adversarial learning, Jiang et al. (2021) proposed EnlightenGAN, composed of a global-local discriminator structure and an attention-guided generator to handle the uneven light conditions. Fu et al. (2022) proposed LE-GAN with an illumination-aware attention module to enhance feature extraction. Rao et al. (2021) extended the Retinex network into the GAN scheme. The image was first decomposed into two components and then enhanced by two generators. Although GANs achieve good performance in LLIE, it is difficult to train GANs and determine the network hyperparameters because of lack of theoretical guidance.

2.3 Diffusion-based image restoration

Recently, diffusion-based generative models have shown outstanding performance and are widely used in image restoration tasks (Yang et al., 2024). Saharia et al. (2023) proposed SR3 and adapted DDPM to conditional image generation, achieving a higher fidelity than GANs. Saharia et al. (2022) found the potential of diffusion models in other tasks and proposed a unified framework Palette for image-to-image translation tasks such as image inpainting and image colorization. Hou et al. (2024) proposed a global structure-aware diffusion process with a curvature regularization term anchored in the intrinsic non-local structures of image data. Wang et al. (2023) proposed a physics-based exposure model to start from a noisy image instead of pure noise in denoising process. Zhou et al. (2023) used a novel pyramid diffusion method to perform sampling in a pyramid resolution style to address the global degradation. Li et al. (2024) proposed FusionDiff, fusing several locally focused source images to obtain globally clear images with DDPM. Lu et al. (2024) proposed a real-time underwater image enhancement (UIE) by applying a novel sampling inference strategy on the DDPM. Guan et al. (2024) designed a DDPM with color compensation as a conditional guide to address UIE. Nathan et al. (2024) trained an unconditional diffusion model prior on the joint space of color and depth to restore underwater images. Although many diffusion-based models have emerged, there is no effective solution for on-orbit LLIE.

3 Data Acquisition

Image datasets in space are difficult to collect, especially those for LLIE training, because it is impossible to capture the real picture of different illuminance with the same state of targets and surroundings when the sensor is mounted on the moving chaser. In this article, a 6-DoF robot will carry the target model simulating spin motion and a camera will collect data of different exposures. To realize automatic collection while avoiding collision and ensuring lack of bias, we first construct a simulation environment in PyBullet 3.2.5 to select suitable poses for the robot. Then the pair-image data will be collected on the ground-test platform. To achieve a data collection procedure without human intervention, the robot arm will decide which pose to perform and plan a trajectory. A collision-free working space and pose stratified sampling method are proposed to ensure diversity and lack of bias, as well as robot collision avoidance with the satellite model and itself.

Unedited

3.1 The ground-test platform

The platform consists of a 6-DoF Universal Robot UR3 and reduced-scale simulate metal models of the Beidou Navigation Satellite. The robot arm carries the model to simulate random motion in space. The metal surface of the model generates strong reflection similar to the real one in space. To avoid collision and singular posture if possible, the motion is set to simple spinning during collection, which will not affect data diversity. To create a space luminous environment of high-fidelity, the room is surrounded by black light-absorbing materials. Three LED light-boxes are placed to provide the Earth's diffuse light at different angles and intensity. The overall settings are shown in Fig. 2.

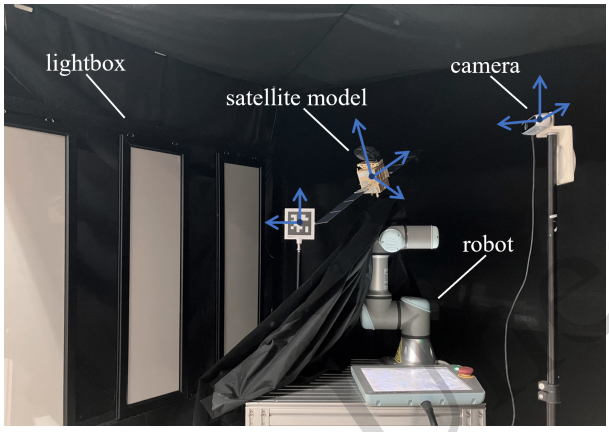
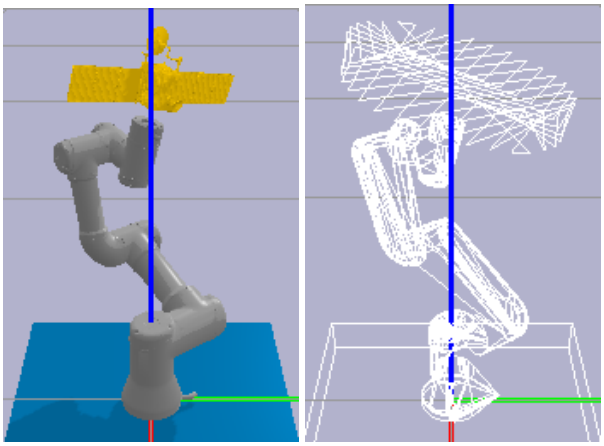


Fig. 2 The layout of the ground-test platform. It is composed of a 6-DoF robot carrying a spinning satellite model and a calibrated camera.

3.2 Collision-free working space

The robot working space defines the space range of task execution. The robot primary and secondary working spaces have been proposed to describe the space that can be reached in one direction and multiple directions (Gupta, 1986). However, these working spaces have not considered the next motion after reaching the target point. In this paper, we propose a collision-free working space, which enables the robot arm to move from initial pose to target pose and simulate satellite spinning without collision. The procedure is as follows:

(1) Set collision geometry. We apply the Bullet physics engine for Python and modify the UR3 model file. The satellite model is added as the last link with a fixed joint. The collision geometry of the satellite is specifically set to cylindrical to avoid collision during spinning. The virtual environment is shown in Figs. 3(a) and 3(b).



(a) The robot model (b) The collision mesh

Fig. 3 The virtual environment in PyBullet. On the left is the visual rendered model, on the right is the collision mesh of the model. (a) The robot model and (b) The collision mesh.

(2) Generate candidate poses. Li et al. (2016) calculated the robot space using the Monte Carlo method, which tries to understand a system by generating a large number of random samples. However, it can be time-consuming to reach convergence. To speed up the sampling, we applied the Quasi Monte Carlo method based on a Halton sequence to generate candidate poses.

(3) Check collision and arrival. According to the candidate pose, joint-space trajectory planning is adopted and during the execution, we check whether the robot collides with the table and the satellite through the Bullet collection detection library.

(4) Finish the loop. If no collision occurs and the robot reaches target poses with a small error tolerance, the poses are added to the collision-free working space. The trajectory is recorded and the robot moves along the trajectory to the initial pose.

The constructed space is shown in Fig. 4; any points in this space can promise secure autonomous data collection.

3.3 Pose-stratified sampling

After constructing the collision-free working space, we selected poses from the space. In an ideal situation, the more poses are sampled, the more uniform and extensive the dataset is. However, con-

Uncorrected

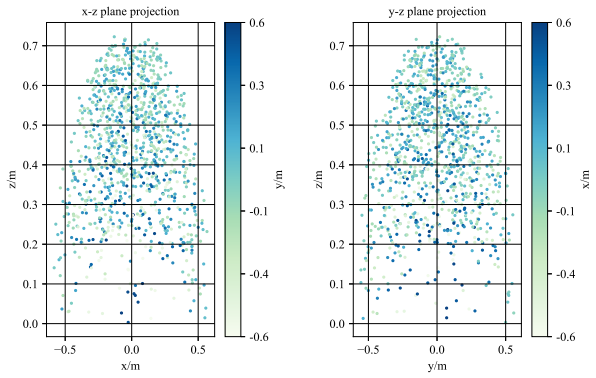


Fig. 4 The $x-z$ and $y-z$ sectional view of the collision-free working space.

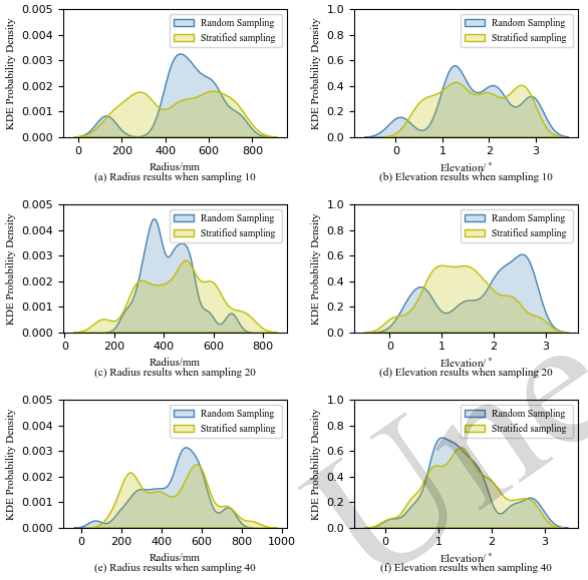


Fig. 5 The probability density comparison of the radius and elevation of different sampling times. The first column is the radius of the camera in the satellite spherical coordinates after different sampling times. The second column is the elevation of the camera in the satellite spherical coordinates after different sampling times.

sidering the limited computing resource, we tend to improve the representativeness of samples as much as possible with a limit size of dataset. With random sampling, the sampled poses may cluster in a certain area because it requires more sample times to approximate the entire distribution. Therefore, we adopt stratified sampling, which converges faster than random sampling.

First, the working space is divided by radius, azimuth angle, and elevation angle in spherical coordinates. Then, the poses are sent to the robot controller via socket. Suppose the initial angle of

Table 1 Composition comparison with other space datasets.

	BUAA 1.5		SPEED		Ours	
	HIL	total	HIL	total	HIL	total
Train	0	1842	5	12005	700	700
Val	0	100	-	-	-	-
Test	0	100	300	3298	20	20

* HIL, hardware-in-the-loop.

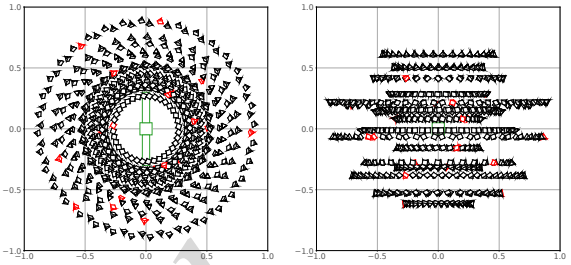


Fig. 6 Visualization of camera poses for images of our dataset in 3D space. The black camera pyramids stand for the training set and red camera pyramids stand for the test set.

the terminal joint is denoted as q_0. The robot carries the satellite around the z-axis of the terminal joint, with a camera capturing photos at 10° intervals. Then we calculate the spherical coordinates of the calibrated camera optical center relative to the satellite. The comparison of random sampling and stratified sampling with 10, 20, and 40 sampling are shown in Fig. 5. The probability density is computed by kernel density estimation (KDE). As illustrated, when the sampling time is small, the random sampling results have a greater probability of clustering in a certain interval, but the distribution of stratified sampling results is robust to the sampling times. Finally, we select 20 pose sampling instances and get 720 pairs of satellite images.

To prove the superiority of our dataset, the number of HIL images and total images is compared with other datasets in the space background, namely BUAA 1.5 (Zhang and Jiang, 2014) for space object recognition and SPEED (Kisantal et al., 2020) for satellite pose estimation. The comparison is illustrated in Table 1. It is shown that our dataset has the largest proportion of HIL images. We also visualized the 3D distribution of the collected images. Fig. 6 indicates that in our dataset, images surrounding the satellite are fully sampled from different viewpoints within a distance of 1 m, whereas in the SPEED dataset, the image poses are restricted.

Uncorrected

3.4 Data preprocessing

To capture high and low illumination images of the same scene, the camera exposure time is set as 156 μs and 1,248 μs, respectively. The pictures are captured with a RealSense D435 camera in the Bayer array format to avoid data compression, especially in low illumination. The Bayer arrays are converted to RGB images with bilinear interpolation. A value in the Bayer array is designated as $I_{i,j}$ and the corresponding pixel value in the RGB image is $T_{i,j}=(R,G,B)$. When $I_{i,j}$ represents red or blue channel (denoted as Case 1 and Case 2, respectively), G and B are calculated from the green and blue channels in the neighboring areas of $I_{i,j}$. The calculation formula is as follows:

$$R=\begin{cases} I_{i,j}, & \text{Case 1} \\ \frac{I_{i-1,j-1}+I_{i+1,j-1}+I_{i-1,j+1}+I_{i+1,j+1}}{4}, & \text{Case 2} \end{cases}$$

$$G=\begin{cases} \frac{I_{i-1,j}+I_{i+1,j}+I_{i,j-1}+I_{i,j+1}}{4}, & \text{Case 1} \\ \frac{I_{i,j-1}+I_{i,j+1}+I_{i-1,j}+I_{i+1,j}}{4}, & \text{Case 2} \end{cases} \quad (1)$$

$$B=\begin{cases} \frac{I_{i-1,j-1}+I_{i+1,j+1}+I_{i-1,j+1}+I_{i+1,j-1}}{4}, & \text{Case 1} \\ I_{i,j}, & \text{Case 2.} \end{cases}$$

When $I_{i,j}$ represents the green channel, $I_{i,j-1}$ represents a red or blue channel, denoted as Case 3 and Case 4. The calculation formula is as follows:

$$R=\begin{cases} \frac{I_{i,j-1}+I_{i,j+1}}{2}, & \text{Case 3} \\ \frac{I_{i-1,j}+I_{i+1,j}}{2}, & \text{Case 4} \end{cases}$$

$$G=\begin{cases} I_{i,j}, & \text{Case3} \\ I_{i,j}, & \text{Case4} \end{cases} \quad (2)$$

$$B=\begin{cases} \frac{I_{i-1,j}+I_{i+1,j}}{2}, & \text{Case 3} \\ \frac{I_{i,j-1}+I_{i,j+1}}{2}, & \text{Case 4.} \end{cases}$$

After converting the Bayer array to RGB format, a center cropping is adopted to reshape the image to 640 × 640. Samples of the dataset are displayed in Fig. 7.

4 LLIE for Satellite Image

To realize on-orbit LLIE in extreme dark illuminance while ensuring the quality of enhanced images, an image-to-image diffusion model with a guided downsampling feature embedding is designed.

Fig. 7 Images in our dataset. In the first and third rows are the normal-light images. In the second and fourth rows are the low-light images.

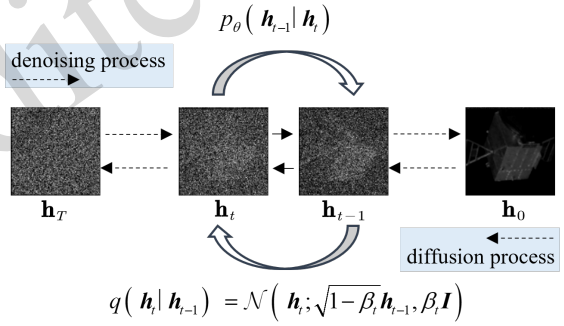


Fig. 8 The diffusion model process. In the forward diffusion process, random Gaussian noise ϵ controlled by the timestep is gradually added to the GT image until the image is completely turned into noise h_T. At the same time, a neural network is trained to predict the Gaussian noise. In the reverse process, ϵ at each step is estimated to recover h_T from the noise. GT, ground truth.

4.1 The diffusion model

Diffusion models comprise a forward diffusion process and a reverse denoising process, as illustrated in Fig. 8. The formulation of the diffusion model-based LLIE problem is demonstrated. Define a low-light image and high-light image $(\boldsymbol{l},\boldsymbol{h})$ pair. Noise is added to $\boldsymbol{h}$ for T times until turning it into an isotropic two-dimensional Gaussian noise. The noise $\boldsymbol{h}$ at step t is designated as $\boldsymbol{h}_t$ $(\boldsymbol{h}_0=\boldsymbol{h})$, and the noise ϵ at step t is scheduled by hyper-parameters β_t, $(0<\beta_1<\beta_2<\ldots<\beta_T\ll 1)$. The Markovian

Unedited

chain is described as follows:

$$\begin{aligned} q\left(\boldsymbol{h}_t \mid \boldsymbol{h}_{t-1}\right) &= \mathcal{N}\left(\boldsymbol{h}_t; \sqrt{1-\beta_t}\boldsymbol{h}_{t-1}, \beta_t \boldsymbol{I}\right) \\ q\left(\boldsymbol{h}_{1:T} \mid \boldsymbol{h}_0\right) &= \Pi_{t=1}^{T} q\left(\boldsymbol{h}_t \mid \boldsymbol{h}_{t-1}\right). \end{aligned} \tag{3}$$

Based on the above property, a closed form of $\boldsymbol{h}_t$, for any intermediate t is obtained

$$q\left(\boldsymbol{h}_t \mid \boldsymbol{h}_0\right) = \mathcal{N}\left(\boldsymbol{h}_t; \sqrt{\gamma_t}\boldsymbol{h}_0, \left(1-\gamma_t\right)\boldsymbol{I}\right), \tag{4}$$

where $\alpha_t = 1-\beta_t$ and $\gamma_t = \prod_{s=1}^{t}\alpha_s$. Because β_t is much smaller than 1, the reverse process can also be considered as a Markovian chain. To recover $\hat{\boldsymbol{h}}_0$ from noise, the posterior probability $p\left(\boldsymbol{h}_{t-1} \mid \boldsymbol{h}_t\right)$ must be achieved, which is difficult to calculate. Therefore, we attempt to use a network to estimate it.

$$p_\theta\left(\boldsymbol{h}_{t-1} \mid \boldsymbol{h}_t\right) = \mathcal{N}\left(\boldsymbol{h}_{t-1}; \mu_\theta\left(\boldsymbol{l}, \boldsymbol{h}_t, t\right), \Sigma_\theta\left(\boldsymbol{l}, \boldsymbol{h}_t, t\right)\right). \tag{5}$$

Following the DDPM (Ho et al., 2020), a similar variational distribution $q\left(\boldsymbol{h}_{1:T} \mid \boldsymbol{h}_0\right)$ is introduced. By deriving a logarithmic likelihood function and variation analysis, the problem of calculating $p\left(\boldsymbol{h}_{t-1} \mid \boldsymbol{h}_t\right)$ is transformed into approximating the posterior probability distribution with $q\left(\boldsymbol{h}_{t-1} \mid \boldsymbol{h}_t, \boldsymbol{h}_0\right)$, which is tractable:

$$q\left(\boldsymbol{h}_{t-1} \mid \boldsymbol{h}_t, \boldsymbol{h}_0\right) = \mathcal{N}\left(\boldsymbol{h}_{t-1}; \tilde{\mu}_t\left(\boldsymbol{l}, \boldsymbol{h}_t, \boldsymbol{h}_0\right), \tilde{\beta}_t \boldsymbol{I}\right), \tag{6}$$

where

$$\begin{aligned} \tilde{\mu}_t &= \frac{\sqrt{\alpha_t}\left(1-\gamma_{t-1}\right)}{1-\gamma_t}\boldsymbol{h}_t + \frac{\sqrt{\gamma_{t-1}}\left(1-\alpha_t\right)}{1-\gamma_t}\boldsymbol{h}_0, \\ \tilde{\beta}_t &= \frac{\left(1-\gamma_{t-1}\right)\left(1-\alpha_t\right)}{1-\gamma_t}. \end{aligned} \tag{7}$$

$\tilde{\beta}_t$ is already known. $\boldsymbol{h}_0$ has a relationship with $\boldsymbol{h}_t$ as described in Eq. 4, so $\tilde{\mu}_t$ is formulated as

$$\tilde{\mu}_t = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{h}_t - \frac{1-\alpha_t}{\sqrt{1-\gamma_t}}\epsilon\right). \tag{8}$$

Use network f_θ to estimate the noise and achieve

$$\mu_\theta\left(\boldsymbol{l}, \tilde{\boldsymbol{h}}_t, \gamma_t\right) = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{h}_t - \frac{1-\alpha_t}{\sqrt{1-\gamma_t}} f_\theta\left(\boldsymbol{l}, \tilde{\boldsymbol{h}}_t, \gamma_t\right)\right), \tag{9}$$

where $f_\theta\left(\boldsymbol{l}, \tilde{\boldsymbol{h}}_t, \gamma_t\right)$ conditions on input low-light image $\boldsymbol{l}$, noisy image $\tilde{\boldsymbol{h}}$, and current noise level γ_t. The GT of noise is recorded in the diffusion process. Therefore, the network is trained to predict noise ϵ by optimizing

$$\mathbb{E}_{(l,h)}\mathbb{E}_{\epsilon,\gamma}\left\|f_\theta(l, \tilde{\boldsymbol{h}}, \gamma) - \epsilon\right\|_p^p. \tag{10}$$

4.2 FAG

Because prediction of RGB channel results is uneven in a color shift (Saharia et al., 2022; Zhou et al., 2023), to avoid error accumulation and save calculation, we directly predict the grayscale image of the enhanced results. Images of satellites in space have a severely uneven illuminance distribution because of the geometric shapes and special surface materials. To improve the enhancement, we propose FAG to highlight the detailed information such as edges and textures of images and guide the network focus on relighting the dark region of images. The guidance map of the input low-light image $\boldsymbol{l}(u, v)$ is calculated as

$$\begin{aligned} \mathrm{FAG} = &[1 - \frac{\lambda}{\sqrt{3}}(R+G+B)] \\ &\oplus \mathcal{FFT}^{-1}\{H(u,v)\cdot\mathcal{FFT}(u,v)\} \end{aligned}. \tag{11}$$

where $\oplus$ means pixel-wise addition. The first part is the reverse of the reversed intensity (RI) channel after converting low-light images into the HSI color space, which guides the network to enhance the dark aeras. λ is a parameter controlled by the degree of image darkness. When the image is extremely dark, the intensity channel is very weak, and λ should be used to highlight the information in the RI channel. The second part corresponds to the HF component of low-light images obtained by a high-pass filter after fast Fourier transform. The FAG illustration is shown in Fig. 9. The value of λ is empirically set to be 20, and the texture areas are strengthened while the high reflective areas are suppressed.

4.3 Network structure

The network is comprised of an encoder and a decoder with residual blocks as the core building block. The skip connections between the encoder and decoder are designed to ensure the reuse rate of features with the same dimension. We find that using pooling layers for downsampling and deconvolution layers instead of the nearest neighbor interpolation for upsampling achieves a better result. We rescale the map with four convolutional layers and obtain fused attention of different scale to improve feature aggregation, as shown in the upper left of Fig. 9. To guide the network with attention, a residual block for FAG is designed, as displayed in the upper right of Fig. 9. The feature map is integrated with FAG and

Unedited

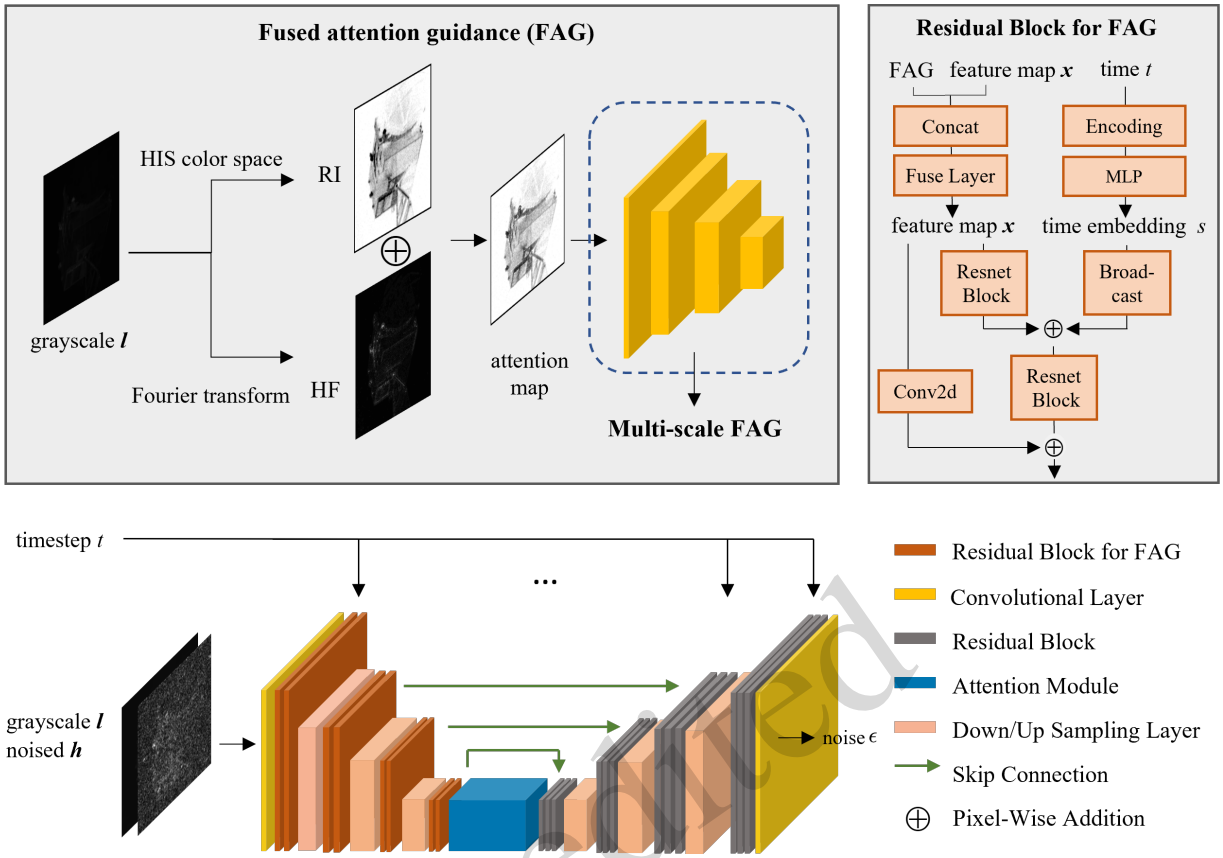


Fig. 9 The network structure. The network takes the grayscale l, h_t, and t as input. l is the low-light image to be lightened. h_t is the noised normal-light image at timestep t in the forward process. The network follows a U-Net structure. Timestep t is fed into a sinusoidal position encoding and multi-layer perceptrons to achieve a time embedding s. Then s is broadcast and fed into each residual block. The FAG of l is reshaped by convolution layers to different scales and integrated into the network with the residual block for FAG

. The network outputs noise ϵ supervised by the noise sampled in the forward process. FAG, fused attention guidance.

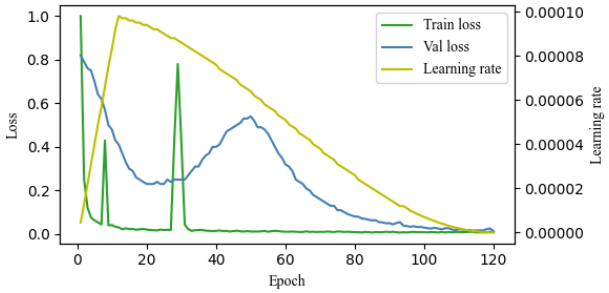


Fig. 10 The curves of the learning rate, training loss, and validation loss.

fed through a fuse layer to obtain the fused feature map. In addition to retaining feature information, the residual block embeds the time information t. Timestep t is exposed to sinusoidal position encoding and multi-layer perceptrons to achieve a time embedding s. Then s is broadcast to add with the fused feature map and fed into a basic ResNet block, which is composed of group normalization, dropout, and a Conv2d layer. The time information is sent to each residual block to make the network conditional on t, which is a significant factor for predicting the noise in different diffusion stages.

5 Experiments

5.1 Implementation details

The entire framework was programmed with PyTorch 1.12.0 (Paszke et al.). Experiments were run on an NVIDIA TITAN RTX GPU with an Intel Xeon E5-2678 CPU. The dataset includes a total of 720 pairs of images, with 700 pairs for training and 20 pairs for testing. The input images were rescaled and

Unedited

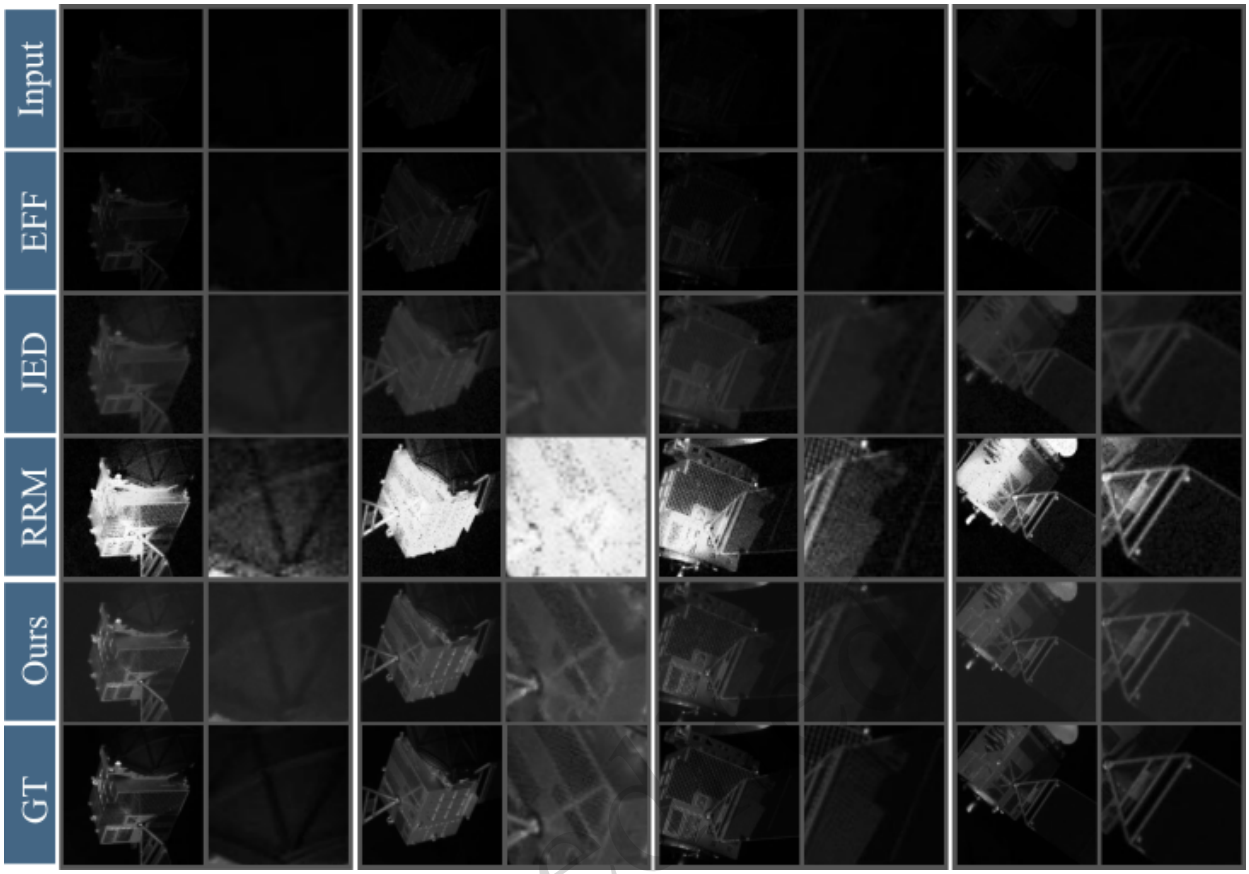


Fig. 11 Visual comparison of the enhanced results and enlarged local details with different traditional methods EFF (Ying et al., 2017), JED (Ren et al., 2020), and RRM (Li et al., 2018).

cropped to 256×256 and augmented with flipping and rotation. During our training, the total timestep T was 2×10^3. The noise weight β_t was arranged by cosine schedule with an offset of 8×10^{-3} to provide a linear drop-off of weight in the middle of the process, while changing very little near both ends to prevent abrupt changes in noise level (Nichol and Dhariwal, 2021). The network was trained for 100 epochs and aimed at reducing a simple l_2 loss. To avoid early over-fitting, we applied a dynamic learning rate lr_i changing over iteration step i scheduled by a linear warm-up (Ma and Yarats, 2021) and decays following cosine annealing (Loshchilov and Hutter, 2017) after reaching $lr = 1 \times 10^{-4}$. The dynamics of the training process are shown in Fig. 10. In the early stages of training, loss oscillations and overfitting exist; as the learning rate decreases and dropout takes effect, the network keeps learning steadily and gradually converges.

5.2 On-orbit LLIE effect analysis

To test the performance of our method, we compared the results with different methods, including the traditional methods EFF (Ying et al., 2017), JED (Ren et al., 2020), and RRM (Li et al., 2018); DL methods RetinexNet (Wei et al., 2018), WaveNet (Dang et al., 2023), RetinexFormer (Cai et al., 2023), LYTNet (Brateanu et al., 2024), PyramidsDiff (Zhou et al., 2023), and GlobalDiff (Hou et al., 2024). The methods were evaluated with image quality assessments (IQAs) PSNR, SSIM, FSIM, and LPIPS. These metrics are described as follows:

PSNR is a mathematical IQA measurement based on the pixel difference between two images (Sara et al., 2019).

SSIM is a perception-based model that considers image degradation as perceived change in structural information, luminance masking, and contrast masking terms (Sara et al., 2019).

FSIM uses the gradient magnitude and phase con-

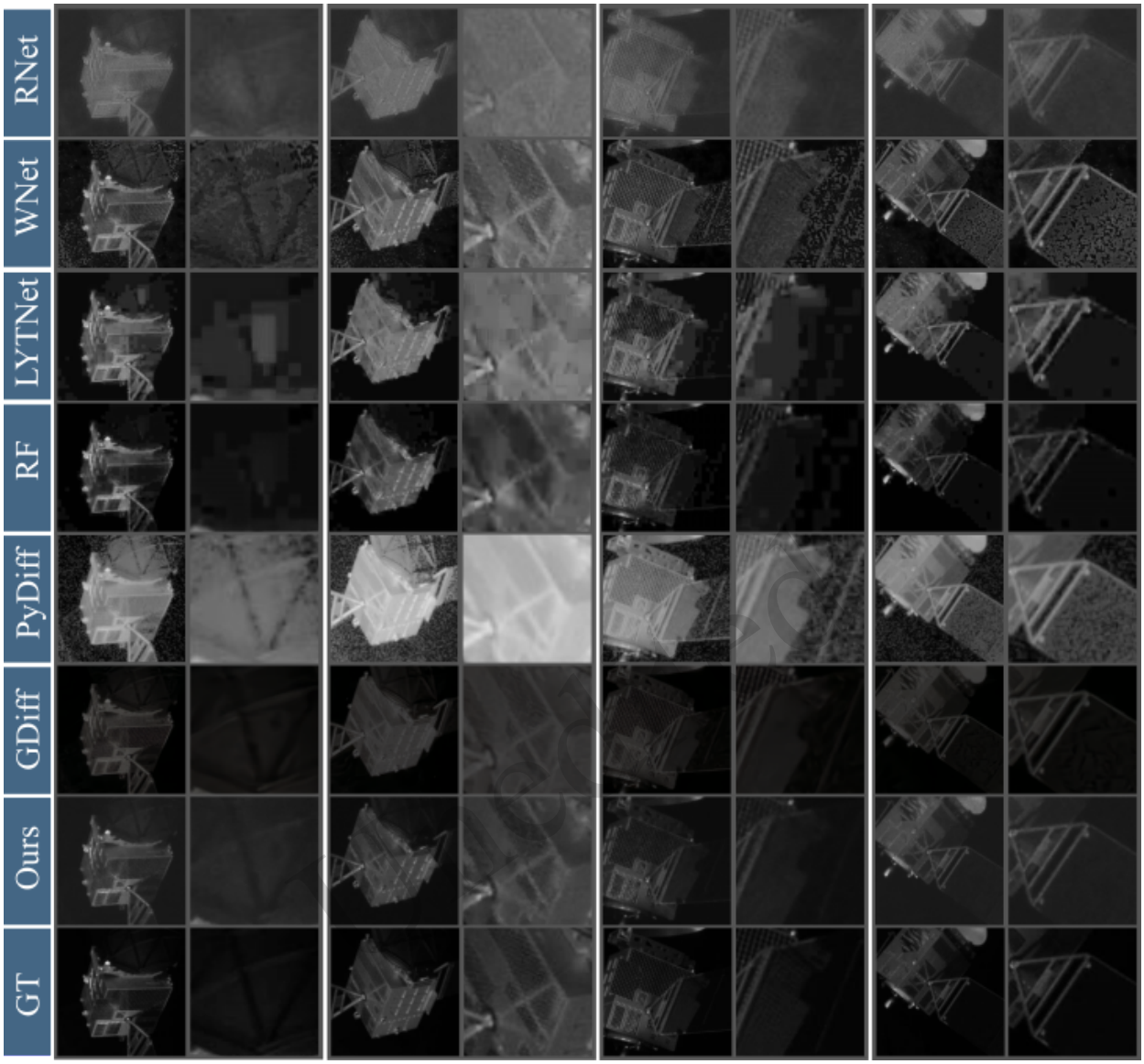


Fig. 12 Visual comparison of the enhanced results and enlarged local details with different DL methods, including RetinexNet (Wei et al., 2018), WaveNet (Dang et al., 2023), RetinexFormer (Cai et al., 2023), LYTNet (Brateanu et al., 2024), PyramidsDiff (Zhou et al., 2023), and GlobalDiff (Hou et al., 2024). DL, deep learning.

gruency for identifying the texture features (Sara et al., 2019).

LPIPS uses a pre-trained network to calculate feature similarity based on semantic information and vision features of higher level (Zhang et al., 2018). The comparison results are illustrated in Table 2. It can be seen that our method achieves the best results on PSNR, FSIM, and LPIPS, which proves its ability in modeling the global pixel value, texture, and deep semantic features of enhanced images. Although RetinexFormer achieves a better score than our method on SSIM, the Retinex-based method has poor visual performance in extremely dark conditions when the decomposition assumption is invalid. Therefore, our method is competitive with SOTA comprehensively. Fig. 11 demonstrates the comparison with classic methods. The EFF and RRM methods are insufficient and over-enhanced. JED enlarges the noise in the dark region and blurs details while increasing the illumination. Fig. 12 illustrates the effect of DL methods. RetinexNet (Rnet) and WaveNet (WNet) cause varying degrees of detail

Table 2 Comparison with previous methods.

Method	Traditional method			DL method						
	EFF ICCVW'17	**JED** ISCAS'18	**RRM** TIP'18	**RNet** BMVC'18	**PyDiff** IJCAI'23	**WNet** PG'23	**GDiff** ICCV'23	**RF** NIPS'23	**LYTNet** arXiv'24	**Ours**
PSNR↑	19.21	<u>24.25</u>	9.284	15.05	10.51	18.03	17.47	23.83	18.31	**25.14**
SSIM↑	0.5499	0.5958	0.3697	0.3368	0.2601	0.4413	0.3487	**0.7602**	0.5503	<u>0.6107</u>
FSIM↑	0.8263	0.8043	0.6231	0.8292	0.6689	0.7677	0.6610	<u>0.8899</u>	0.8144	**0.9102**
LPIPS↓	0.2159	0.3072	0.5062	0.3278	0.5192	0.3578	0.4987	<u>0.2043</u>	0.3496	**0.0951**

* The optimal values are in bold and the suboptimal values are underlined. DL, deep learning.

Table 3 The FAG study.

Method	w/o FAG	w/ FAG
PSNR↑	23.81	25.14
SSIM↑	0.5937	0.6107
FSIM↑	0.8994	0.9102
LPIPS↓	0.1206	0.0951

* FAG, fused attention guidance.

blurring and image distortion. LYTNet and Retinex-Former (RF) have are successful in learning the distribution of the target domain, but the image resolution is degraded. Compared with other newly proposed diffusion-based methods, our method generates less noise and fewer artifacts while brightening the low-light images.

5.3 The ablation study

We performed an ablation study to verify the effectiveness of the FAG. The lightened images with and without FAG are compared in Fig. 13. In the first column are the input low-light images. In the second column are the results without FAG, which have unreal texture artifacts and fail to lighten extremely dark areas. The results in the third column were generated with FAG. The ablation study evaluation metrics are displayed in Table 3, from which we can conclude that FAG improves the quality of feature extraction during downsampling. The visual results and metrics indicate that our method has better performance in low-light enhancement and fidelity maintenance.

Table 4 The number of feasible SIFT features of enhanced images of two continuous frames.

	A	B	C	D	Average
Raw	1	0	0	0	0.25
Enhanced	140	130	159	83	128

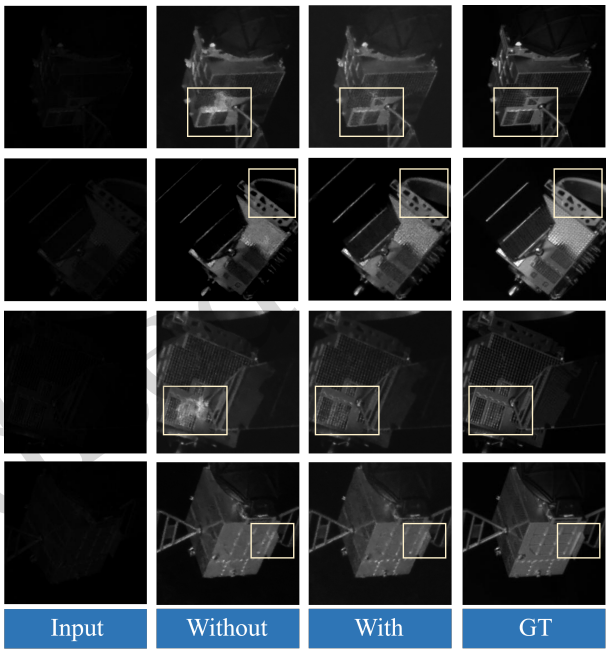


Fig. 13 Visual comparison of the enhanced results with and without FAG. The yellow boxes highlight areas with significant improvement. FAG, fused attention guidance.

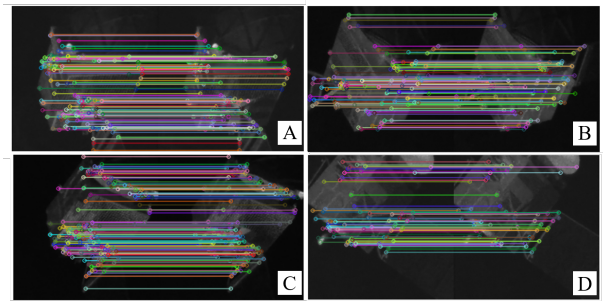


Fig. 14 The SIFT matching result of enhanced images of two continuous frames.

5.4 Engineering applicability analysis

After contrast enhancement, the low-light images can provide more information in high-level visual tasks, which is vital for the success of on-orbit

tasks. To prove that the features of the enhanced images become richer and preserve consistency, we performed SIFT feature point matching on the enhanced images of continuous frames, as shown in Fig. 14. Table 4 shows the SIFT matching results of enhanced images compared with raw images. In low-light images, features can barely be extracted; after our enhancement, more features are extracted.

6 Conclusions and Future Work

This article explores the on-orbit LLIE problem of a Beidou Navigation Satellite under extremely dark conditions. An automatic scheme to create a diverse and high-quality satellite image dataset is proposed, reducing the domain gap between the dataset and real world, which is essential for data-driven methods. Based on the dataset, a novel diffusion model is proposed, trained with guidance based on fused features to enhance image contrast without over-exposure or blurred details. Experiments indicate that our approach outperforms SOTA on-orbit LLIE. In addition, this method can also be transferred to observation missions of asteroids and other target satellites with different configurations. However, the reverse process of the diffusion model is time-consuming. Future work will need to study diffusion accelerators to adapt to onboard computing resources. We will also explore few-shot learning for multiple satellite targets and address other image restoration challenges, such as motion blur and over-exposure, with a multi-task diffusion model.

Contributors (refer to https://www.casrai.org/credit.html)

Yiman ZHU designed the method and drafted the manuscript. Jingyi YUAN helped with experiments. Lu Wang and Yu GUO helped revise the manuscript.

Conflict of interest

All the authors declare that they have no conflict of interest.

References

Brateanu A, Balmez R, Avram A, et al., 2024. LYT-Net: lightweight yuv transformer-based network for low-light image enhancement. https://arxiv.org/abs/2401.15204

Cai YH, Bian H, Lin J, et al., 2023. Retinexformer: one-stage retinex-based transformer for low-light image enhancement. Proc IEEE/CVF Int Conf on Computer Vision, p.12504-12513. https://doi.org/10.1109/ICCV51070.2023.01149

Capuano V, Alimo SR, Ho AQ, et al., 2019. Robust features extraction for on-board monocular-based spacecraft pose acquisition. Proc AIAA Scitech Forum, p.1-15. https://doi.org/10.2514/6.2019-2005

Chen XL, Liu Z, Xie SN, et al., 2024. Deconstructing denoising diffusion models for self-supervised learning. https://arxiv.org/abs/2401.14404

Chen YS, Wang YC, Kao MH, et al., 2018. Deep photo enhancer: unpaired learning for image enhancement from photographs with gans. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.6306-6314. https://doi.org/10.1109/CVPR.2018.00660

Civardi GL, Bechini M, Quirino M, et al., 2024. Generation of fused visible and thermal-infrared images for uncooperative spacecraft proximity navigation. *Adv Space Res*, 73(11):5501-5520. https://doi.org/10.1016/j.asr.2023.03.022

Cowardin H, Miller R, 2022. Orbital debris quarterly news: the intentional destruction of cosmos 1408. https://original-ufdc.uflib.ufl.edu/AA00058853/00096

Croitoru FA, Hondru V, Ionescu RT, et al., 2023. Diffusion models in vision: a survey. *IEEE Trans Pattern Anal Mach Intell*, 45(9):10850-10869. https://doi.org/10.1109/TPAMI.2023.3261988

Cui HS, Li JJ, Hua Z, et al., 2022. TPET: two-stage perceptual enhancement transformer network for low-light image enhancement. *Eng Appl Artif Intell*, 116:105411. https://doi.org/10.1016/j.engappai.2022.105411

Dang JC, Li ZH, Zhong Y, et al., 2023. WaveNet: wave-aware image enhancement. In: Chaine R, Deng Z, Kim MH (Eds), Pacific Graphics, p.1-9. https://doi.org/10.2312/pg.20231267

Diao HF, Li Z, Ma ZH, 2011. Simulation of space-based visible surveillance images for space surveillance. Proc SPIE 8196, Int Symp on Photoelectronic Detection and Imaging 2011: Space Exploration Technologies and Applications, p.94-102. https://doi.org/10.1117/12.899120

Fu Y, Hong Y, Chen LW, et al., 2022. LE-GAN: unsupervised low-light image enhancement network using attention module and identity invariant loss. *Knowl-Based Syst*, 240:108010. https://doi.org/10.1016/j.knosys.2021.108010

Guan MS, Xu HY, Jiang GY, et al., 2024. DiffWater: underwater image enhancement based on conditional denoising diffusion probabilistic model. *IEEE J Sel Top Appl Earth Obs Remote Sens*, 17:2319-2335. https://doi.org/10.1109/JSTARS.2023.3344453

Guo XJ, Li Y, Ling HB, 2017. LIME: low-light image enhancement via illumination map estimation. *IEEE Trans Image Process*, 26(2):982-993. https://doi.org/10.1109/TIP.2016.2639450

Gupta KC, 1986. On the nature of robot workspace. *Int J Robot Res*, 5(2):112-121. https://doi.org/10.1177/027836498600500212

Harris C, Thomas DJ, Kadan J, et al., 2021. Expanding the space surveillance network with spacebased sensors using metaheuristic optimization techniques. Proc Advanced Maui Optical and Space Surveillance Technologies, p.1-13.

Unedited

Ho J, Jain A, Abbeel P, 2020. Denoising diffusion probabilistic models. *Proc 34th Int Conf on Neural Information Processing Systems*, article 574.

Hou JH, Zhu ZY, Hou JH, et al., 2024. Global structure-aware diffusion process for low-light image enhancement. *Proc 37th Int Conf on Neural Information Processing Systems*, article 3490.

Hu YL, Speierer S, Jakob W, et al., 2021. Wide-depth-range 6D object pose estimation in space. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.15870-15879.
https://doi.org/10.48550/arXiv.2104.00337

Huang ZX, Li JJ, Hua Z, et al., 2023. Filter-cluster attention based recursive network for low-light enhancement. *Front Inform Technol Electron Eng*, 24(7):1028-1044.
https://doi.org/10.1631/FITEE.2200344

Jiang YF, Gong XY, Liu D, et al., 2021. Enlightengan: deep light enhancement without paired supervision. *IEEE Trans Image Process*, 30:2340-2349.
https://doi.org/10.1109/TIP.2021.3051462

Kisantal M, Sharma S, Park TH, et al., 2020. Satellite pose estimation challenge: dataset, competition design, and results. *IEEE Trans Aerosp Electron Syst*, 56(5):4083-4098.
https://doi.org/10.1109/TAES.2020.2989063

Ledkov A, Aslanov V, 2022. Review of contact and contactless active space debris removal approaches. *Prog Aerosp Sci*, 134:100858.
https://doi.org/10.1016/j.paerosci.2022.100858

Li CY, Guo CL, Han LH, et al., 2022. Low-light image and video enhancement using deep learning: a survey. *IEEE Trans Pattern Anal Mach Intell*, 44(12):9396-9416.
https://doi.org/10.1109/TPAMI.2021.3126387

Li J, Zhao F, Li XC, et al., 2016. Analysis of robotic workspace based on monte carlo method and the posture matrix. Proc IEEE Int Conf on Control and Robotics Engineering, p.1-5.
https://doi.org/10.1109/ICCRE.2016.7476145

Li MD, Liu JY, Yang WH, et al., 2018. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Trans Image Process*, 27(6):2828-2841.
https://doi.org/10.1109/TIP.2018.2810539

Li MN, Pei RH, Zheng TY, et al., 2024. FusionDiff: multi-focus image fusion using denoising diffusion probabilistic models. *Expert Syst Appl*, 238:121664.
https://doi.org/10.1016/j.eswa.2023.121664

Li N, Zhang J, 2019. Automatic image enhancement by learning adaptive patch selection. *Front Inform Technol Electron Eng*, 20(2):206-221.
https://doi.org/10.1631/FITEE.1700125

Lore KG, Akintayo A, Sarkar S, 2017. LLNet: a deep autoencoder approach to natural low-light image enhancement. *Pattern Recognit*, 61:650-662.
https://doi.org/10.1016/j.patcog.2016.06.008

Loshchilov I, Hutter F, 2017. SGDR: stochastic gradient descent with warm restarts. Proc 5th Int Conf on Learning Representations, p.1-16.
https://doi.org/10.48550/arXiv.1608.03983

Lu SQ, Guan FX, Zhang HY, et al., 2024. Speed-up ddpm for real-time underwater image enhancement. *IEEE Trans Circuits Syst Video Technol*, 34(5):3576-3588.
https://doi.org/10.1109/TCSVT.2023.3314767

Ma J, Yarats D, 2021. On the adequacy of untuned warmup for adaptive optimization. Proc 35th AAAI Conf Artificial Intelligence, p.8828-8836.
https://doi.org/10.48550/arXiv.1910.04209

Musallam MA, Gaudilliere V, Ghorbel E, et al., 2021. Spacecraft recognition leveraging knowledge of space environment: simulator, dataset, competition design and analysis. Proc IEEE Int Conf on Image Processing Challenges, p.11-15.
https://doi.org/10.1109/ICIPC53495.2021.9620184

Nathan OB, Levy D, Treibitz T, et al., 2024. Osmosis: RGBD diffusion prior for underwater image restoration.
https://arxiv.org/abs/2403.14837

Nichol AQ, Dhariwal P, 2021. Improved denoising diffusion probabilistic models. Proc 38th Int Conf on Machine Learning, p.8162-8171.
https://doi.org/10.48550/arXiv.2102.09672

Park TH, Märtens M, Jawaid M, et al., 2023. Satellite pose estimation competition 2021: Results and analyses. *Acta Astronaut*, 204:640-665.
https://doi.org/10.1016/j.actaastro.2023.01.002

Paszke A, Gross S, Massa F, et al. PyTorch: an imperative style, high-performance deep learning library. *Proc 33rd Int Conf on Neural Information Processing Systems*, article 721.
https://doi.org/10.48550/arXiv.1912.01703

Poozhiyil M, Nair MH, Mini-C R, et al., 2023. Active debris removal: a review and case study on leopard phase 0-a mission. *Adv Space Res*, 72(8):3386-3413.
https://doi.org/10.1016/j.asr.2023.06.015

Proença PF, Gao Y, 2020. Deep learning for spacecraft pose estimation from photorealistic rendering. Proc IEEE Int Conf on Robotics and Automation, p.6007-6013.
https://doi.org/10.1109/ICRA40945.2020.9197244

Rahman Z, Jobson DJ, Woodell GA, 1996. Multi-scale retinex for color image enhancement. Proc 3rd IEEE Int Conf on Image Processing, p.1003-1006.
https://doi.org/10.1109/ICIP.1996.560995

Rao N, Lu T, Zhou Q, et al., 2021. Seeing in the dark by component-gan. *IEEE Signal Proc Lett*, 28:1250-1254.
https://doi.org/10.1109/LSP.2021.3079848

Ren XT, Li MD, Cheng WH, et al., 2020. Joint enhancement and denoising method via sequential decomposition. Proc IEEE Int Symp on Circuits and Systems, p.1-5.
https://doi.org/10.1109/ISCAS.2018.8351427

Saharia C, Chan W, Chang HW, et al., 2022. Palette: image-to-image diffusion models. *Proc ACM SIGGRAPH Conf Proc*, article 15.
https://doi.org/10.1145/3528233.3530757

Saharia C, Ho J, Chan W, et al., 2023. Image super-resolution via iterative refinement. *IEEE Trans Pattern Anal Mach Intell*, 45(4):4713-4726.
https://doi.org/10.1109/TPAMI.2022.3204461

Sara U, Akter M, Uddin MS, 2019. Image quality assessment through fsim, ssim, mse and psnr-a comparative study. *J Comput Commun*, 7(3):8-18.
https://doi.org/10.4236/jcc.2019.73002

Shi PF, Xu XW, Fan XN, et al., 2024. Ll-unet++:unet++ based nested skip connections network for low-light image enhancement. *IEEE Trans Comput Imaging*,

Unedited

10:510-521.
https://doi.org/10.1109/TCI.2024.3378091

Triantafyllidou D, Moran S, McDonagh S, et al., 2020. Low light video enhancement using synthetic data produced with an intermediate domain mapping. Proc 16th European Conf, p.103-119.
https://doi.org/10.48550/arXiv.2007.09187

Wang Y, Cao Y, Zha ZJ, et al., 2019. Progressive retinex: mutually reinforced illumination-noise perception network for low-light image enhancement. Proc 27th ACM Int Conf on Multimedia, p.2015-2023.
https://doi.org/10.1145/3343031.3350983

Wang YF, Yu Y, Yang WH, et al., 2023. ExposureDiffusion: learning to expose for low-light image enhancement. Proc IEEE/CVF Int Conf on Computer Vision, p.12404-12414.
https://doi.org/10.1109/ICCV51070.2023.01143

Wei C, Wang WJ, Yang WH, et al., 2018. Deep retinex decomposition for low-light enhancement. *Proc British Machine Vision Conf*, article 155.
https://doi.org/10.48550/arXiv.1808.04560

Xu YD, Yang C, Sun BB, et al., 2021. A novel multi-scale fusion framework for detail-preserving low-light image enhancement. *Inf Sci*, 548:378-397.
https://doi.org/10.1016/j.ins.2020.09.066

Xu YD, Feng K, Yan XA, et al., 2023. CFCNN: a novel convolutional fusion framework for collaborative fault identification of rotating machinery. *Inf Fusion*, 95:1-16.
https://doi.org/10.1016/j.inffus.2023.02.012

Xu YD, Yan XA, Sun BB, et al., 2024. Online knowledge distillation based multiscale threshold denoising networks for fault diagnosis of transmission systems. *IEEE Transactions on Transportation Electrification*, 10(2):4421-4431.
https://doi.org/10.1109/TTE.2023.3313986

Yan QS, Feng YX, Zhang C, et al., 2024. You only need one color space: an efficient network for low-light image enhancement.
https://arxiv.org/abs/2402.05809

Yang L, Zhang ZL, Song Y, et al., 2024. Diffusion models: a comprehensive survey of methods and applications. *ACM Comput Surv*, 56(4):105.
https://doi.org/10.1145/3626235

Yang X, Wang XQ, Wang NN, et al., 2022. SRDN: a unified super-resolution and motion deblurring network for space image restoration. *IEEE Trans Geosci Remote Sens*, 60:5614411.
https://doi.org/10.1109/TGRS.2021.3131264

Ying ZQ, Li G, Ren YR, et al., 2017. A new low-light image enhancement algorithm using camera response model. Proc IEEE Int Conf on Computer Vision Workshops, p.1-8.
https://doi.org/10.1109/ICCVW.2017.356

Zhang HP, Jiang ZG, 2014. Multi-view space object recognition and pose estimation based on kernel regression. *Chin J Aeronaut*, 27(5):1233-1241.
https://doi.org/10.1016/j.cja.2014.03.021

Zhang R, Isola P, Efros AA, et al., 2018. The unreasonable effectiveness of deep features as a perceptual metric. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.586-595.
https://doi.org/10.48550/arXiv.1801.03924

Zhou DW, Yang ZX, Yang Y, 2023. Pyramid diffusion models for low-light image enhancement. *Proc 32nd Int Joint Conf on Artificial Intelligence*, article 199.
https://doi.org/10.48550/arXiv.2305.10028

Unedited