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 diffusion model for on-orbit low-light image enhancement*

Yiman ZHU, Lu WANG‡, Jingyi YUAN, Yu GUO

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. 7, 2024; Revision accepted Sept. 10, 2024; Crosschecked June 30, 2025

Abstract: On-orbit service is important for maintaining the sustainability of the 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 has achieved remarkable success in image enhancement of natural images, but it is seldom applied in space due to the data bottleneck. In this study, we first propose a dataset of BeiDou navigation satellites for on-orbit low-light image enhancement (LLIE). In the automatic data collection scheme, we focus on reducing the domain gap and improving the diversity of the dataset. We collect hardware-in-the-loop images based on a robotic simulation testbed imitating space lighting conditions. To evenly sample poses of different orientations and distances without collision, we propose a collision-free workspace and pose-stratified sampling. Subsequently, we develop a novel diffusion model. To enhance the image contrast without over-exposure and blurred details, we design fused attention guidance to highlight the structure and the dark region. Finally, a comparison of our method with previous methods indicates that our method has better on-orbit LLIE performance.

Key words: Satellite capture; Low-light image enhancement (LLIE); Data collection; Diffusion model; Fused attention

https://doi.org/10.1631/FITEE.2400261 **CLC number:** TP391

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 existing 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 refurbishing, as demonstrated in Fig. 1a. 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

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (Nos. 62403242 and 61973167), the Fundamental Research Funds for the Central Universities (No. 30924010932), and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (No. KYCX23_0481)

ORCID: Yiman ZHU, https://orcid.org/0000-0002-7421-0188; Lu WANG, https://orcid.org/0000-0001-8147-1581

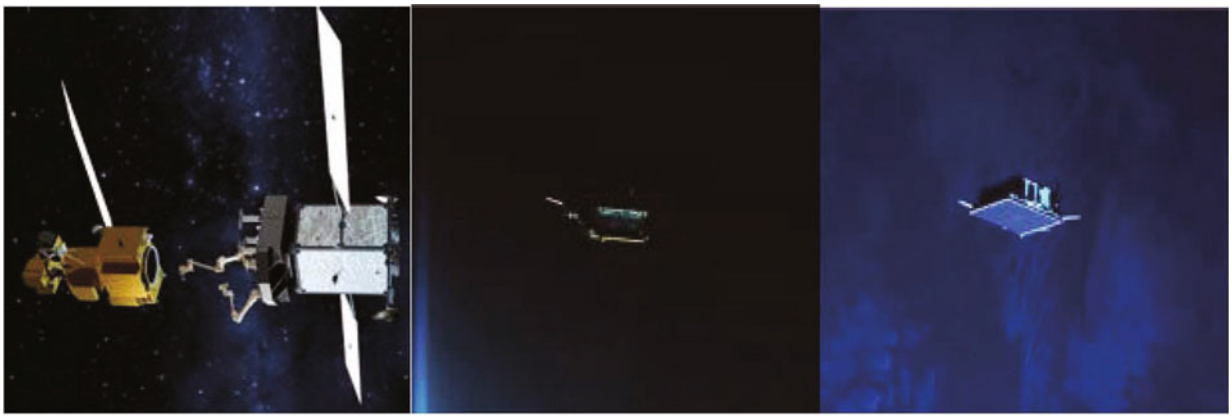

Fig. 1 Illustration of on-orbit service and low-light images: (a) demonstrates a service star refueling the target star; (b) and (c) are the real images of low-light and normal-light captured by a visible camera during the PRISMA mission, respectively (Capuano et al., 2019)

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, especially when the satellite is in the Earth's shadow, as shown in Figs. 1b and 1c. 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 the 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 classical theory (Rahman et al., 1996; Guo et al., 2017; Ying et al., 2017; Li MD et al., 2018; Ren et al., 2018) have been widely used to improve image contrast. Unfortunately, these methods are suitable only for situations in which images already contain a good representation of the scene content. For extremely dark and noisy images, they are likely to cause severe noise amplification after enhancement.

Deep learning (DL) methods have been applied to solve problems in various industries, demonstrating significant advancements (Li CY et al., 2022; Xu et al., 2023, 2024). DL-based methods for LLIE can be classified as one-way neural networks (Wei et al., 2018; Li N and Zhang, 2019; Xu et al., 2021; Dang et al., 2023; Huang et al., 2023; Shi et al., 2024; Yan et al., 2024) and generative adversarial networks (GANs) (Chen YS et al., 2018; Triantafyllidou et al., 2020; Jiang et al., 2021; Rao et al., 2021; Fu et al., 2022). One-way neural networks refer to architectures that consist of a single 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 a state of equilibrium. However, this mechanism makes the networks more prone to collapse 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 XL et al., 2024). In contrast to GANs, which rely on inner-loop maximization, diffusion models are guided by a concise and well-constructed loss function, making them easier to converge. The denoising diffusion probabilistic model (DDPM) represents a category of deep generative models that include two stages: (1) a forward diffusion stage, in which the input data are gradually perturbed over several steps by adding Gaussian noise, and (2) 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 (Wang YF et al., 2023; Zhou et al., 2023; Hou et al., 2024).

However, most of these methods are based on natural public images. When enhancing on-orbit

low-light images, problems remain, such as insufficient enhancement, amplified noise, and artificial features. As is commonly known, training data are 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 a synthetic dataset for on-orbit LLIE tasks because of the difficulty in 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 distributions 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, we focus on a spacecraft with a solar panel, using one of the BeiDou navigation satellites as an example. We devise an automatic dataset collection scheme based on a 6-degree-of-freedom (6-DoF) robot to tackle the data bottleneck for LLIE in 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 proven to be more effective than previous methods. There are four main contributions of this paper:

1. Unlike previous methods that generate synthetic images, we 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 the training set.

2. To reduce manpower and ensure the safety and quality of data collection, we design a pose-stratified sampling based on the collision-free workspace obtained with a physics engine.

3. 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 state-of-the-art (SOTA) DL methods, our method shows better performance.

4. 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 works

2.1 Space dataset

Images captured in space are scarce and hard to obtain, which hinders the application of intelligent methods. Some efforts have been made to change this situation. The methods for acquiring datasets can be divided into two classes: capturing real images with different exposures and rendering images with physical engines. Yang X 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, consisting 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 with different exposures. 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 study, we build a ground-test platform and an automatic collection scheme to make an HIL image dataset for LLIE, which practically addresses this problem.

2.2 LLIE

Early works on LLIE can typically be categorized as histogram equalization-based methods and Retinex model-based methods. Histogram equalization-based methods enhance the image

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

With the 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 across 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 considered the interpretability of DL and integrated the principles of image decomposition with DL-based methods. Based on the Retinex model, Wei et al. (2018) and Wang Y 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), and 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's ability of feature extraction. Dang et al. (2023) constructed the wave transform block, which represented pixels in a dynamic wave form with a signal function to improve 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 the decomposition principle and new architecture, Cai et al. (2023) formulated a one-stage Retinex-based Transformer to model the long-range dependencies under the guidance of illumination information. Brateanu et al. (2024) used the YUV (Y: luminance; UV: chrominance) 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, one-way neural networks count more on training data and a supervised loss function. GANs introduce 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 help with 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 a lack of theoretical guidance.

2.3 Diffusion-based image restoration

Recently, diffusion-based generative models have shown outstanding performance and have been widely used in image restoration tasks (Lin et al., 2024; Yang L et al., 2024). Saharia et al. (2023) proposed SR3 and adapted DDPM to conditional image generation, achieving higher fidelity than GANs. Saharia et al. (2022) found the potential of diffusion models in other tasks and proposed a unified framework called Palette for image-to-image translation tasks, such as image inpainting and 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 YF et al. (2023) proposed a physics-based exposure model to start the denoising process from a noisy image rather than pure noise. Zhou et al. (2023) used a novel pyramid diffusion method to perform sampling in a pyramid resolution style to address the global degradation. Li MN et al. (2024) proposed FusionDiff, fusing multiple locally focused source images to obtain globally clear images using 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 real images of varying illuminance with the same state of targets and surroundings when the sensor is mounted on the moving chaser. In this study, a 6-DoF robot carries a target model that simulates spin motion, and a camera collects data with different exposures. To realize automatic collection while avoiding collision and ensuring a lack of bias, we first construct a simulation environment using PyBullet 3.2.5 to select suitable poses for the robot. Then, paired-image data are collected on a ground-test platform. To achieve a data collection procedure without human intervention, the robot arm decides which pose to perform and plans a trajectory. A collision-free workspace and pose-stratified sampling method are proposed to ensure diversity, lack of bias, and robot collision avoidance with the satellite model.

3.1 Ground-test platform

The platform consists of a 6-DoF universal robot UR3 and a reduced-scale metal model that simulates one of the BeiDou navigation satellites. The robot arm carries the model to simulate random motion in space. The metal surface of the model generates strong reflections that closely resemble those observed in real space environments. To avoid collision and singular posture, the motion is set to simple spinning during collection, which will not affect data diversity. To create a space-like 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 diffusion light at different angles and intensities. The overall settings are shown in Fig. 2.

3.2 Collision-free workspace

The robot workspace defines the space range of task execution. The primary workspace is the set of positions reachable by the robot from a single direction, while the secondary workspace is the set of positions reachable from any direction (Gupta, 1986).

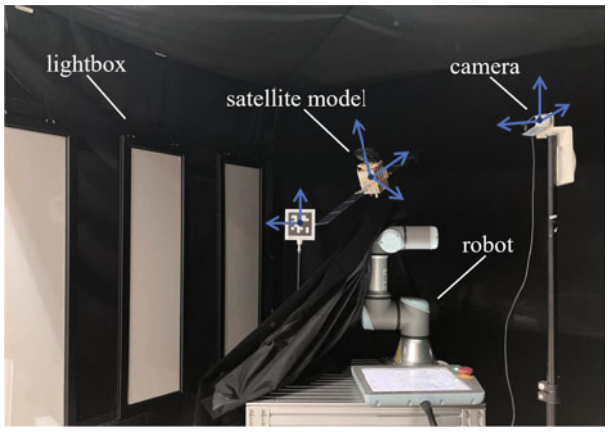

Fig. 2 Layout of the ground-test platform. It is composed of a 6-DoF robot carrying a spinning satellite model and a calibrated camera

However, these workspaces have not considered the robot's next motion after reaching a target point. In this paper, we propose a collision-free workspace, which enables the robot arm to move from an initial pose to a target pose and simulate satellite spinning without any 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 Fig. 3.

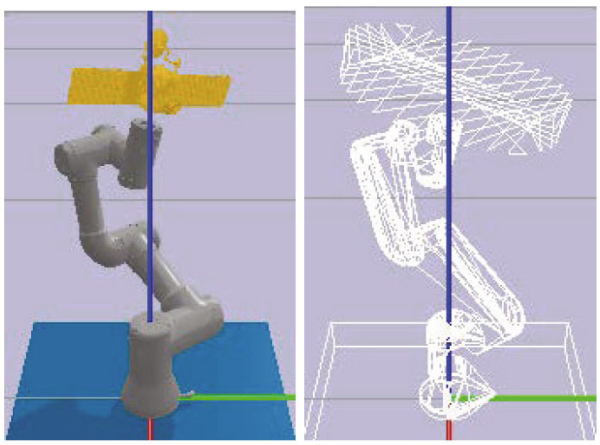

Fig. 3 Virtual environment in PyBullet: visual rendering (left) and collision mesh (right) of the robot model

2. Generate candidate poses. Li J et al. (2016) calculated the robot workspace using the Monte Carlo method, which tries to understand a system by generating a large number of random samples.

However, this process can be time-consuming to reach convergence. To speed up the sampling, we apply 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. During the execution, we use the Bullet collection detection library to check whether the robot collides with the table and the satellite.

4. Finish the loop. If no collision occurs and the robot reaches the target pose within a small error tolerance, the pose is added to the collision-free workspace. Its trajectory is recorded, and the robot returns to the initial pose along the recorded trajectory.

The constructed workspace is shown in Fig. 4. Any point in this workspace ensures secure autonomous data collection.

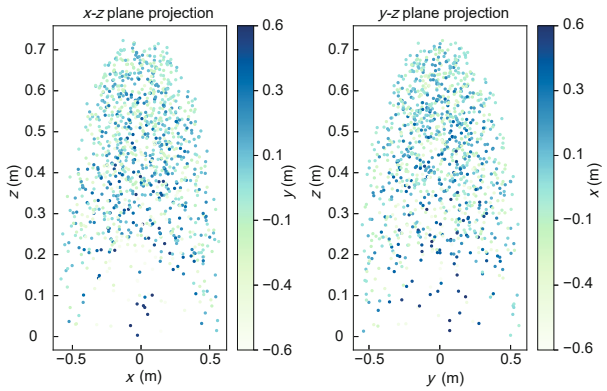

Fig. 4 x-z and y-z sectional views of the collision-free workspace

3.3 Pose-stratified sampling

After constructing the collision-free workspace, we can select poses from the workspace. In an ideal situation, the more poses are sampled, the more uniform and extensive the dataset becomes. However, considering the limited computing resources, we tend to improve the representativeness of samples as much as possible with a limited dataset size. With random sampling, the sampled poses may cluster in a certain area because it requires more samples to approximate the entire distribution. Therefore, we adopt stratified sampling, which converges faster than random sampling.

The workspace is first divided by radius, azimuth angle, and elevation angle in spherical coordinates. Then the poses are sent to the robot controller via socket communication. Suppose that the initial angle of 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° angular intervals. Subsequently, 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 samples is shown in Fig. 5. The probability density is computed by kernel density estimation (KDE). As illustrated, when the number of samples 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 number of samples. Finally, we select 20 pose sampling instances and obtain 720 pairs of satellite images.

To prove the superiority of our dataset, we compare the number of HIL images and the total number of images with those of other datasets in the space background, namely BUAA 1.5 (Zhang HP and Jiang et al., 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 visualize 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.

Table 1 Comparison of the number of HIL images and the total number of images between our dataset and other datasets in the space background

Set	BUAA 1.5		SPEED		Ours	
	HIL*	Total	HIL*	Total	HIL*	Total
Training	0	1842	5	12 005	700	700
Val	0	100	–	–	–	–
Test	0	100	300	3298	20	20

HIL*: the number of hardware-in-the-loop images; Val: the validation set; "–" denotes that relevant data are not available

3.4 Data preprocessing

To capture high and low illumination images of the same scene, the camera exposure time is set to 156 μs and 1248 μs, respectively. The images are captured with a RealSense D435 camera in the Bayer

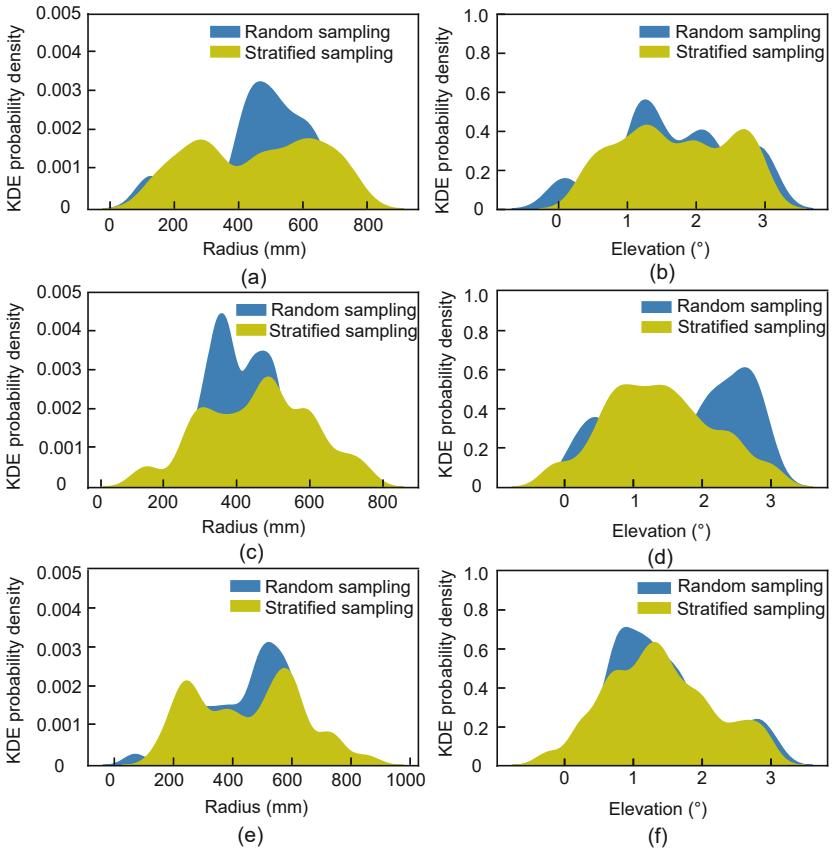

Fig. 5 Probability density comparison of the radius and elevation of different numbers of samples: the radius of camera in the satellite spherical coordinates when sampling 10 (a), 20 (c), and 40 (e) samples, and the elevation of the camera in the satellite spherical coordinates when sampling 10 (b), 20 (d), and 40 (f) samples

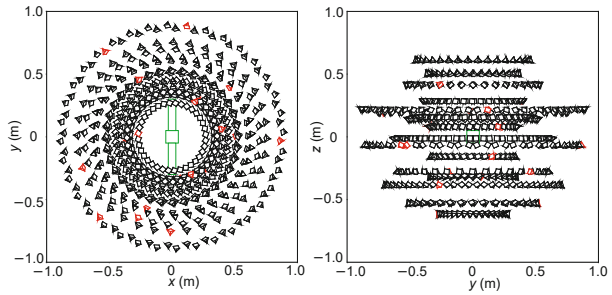

Fig. 6 Visualization of camera poses for images of our dataset in 3D space. The black camera pyramids stand for the training set and red camera pyramids stand for the test set. References to color refer to the online version of this figure

array format to avoid data compression, especially in low illumination. The Bayer arrays are converted to red/green/blue (RGB) images with bilinear interpolation. A pixel in the Bayer array is designated as $I_{i,j}$ ($i = 1, 2, ..., M$ and $j = 1, 2, ..., N$), where M and N represent the array size. The corresponding pixel in the RGB image is $\boldsymbol{T}_{i,j} = (R, G, B)$. When $I_{i,j}$ represents red or blue channel (denoted as Case 1 or Case 2, respectively), G and B are calculated from the green and blue channels in the neighboring areas of $I_{i,j}$. The calculation formulas are as follows:

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

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

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

When $I_{i,j}$ represents the green channel, $I_{i,j-1}$ represents a red or blue channel, denoted as Case 3

or Case 4, the calculation formulas are as follows:

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

$$G=\begin{cases}I_{i,j}, & \text{Case 3},\\ I_{i,j}, & \text{Case 4},\end{cases} \tag{2b}$$

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

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

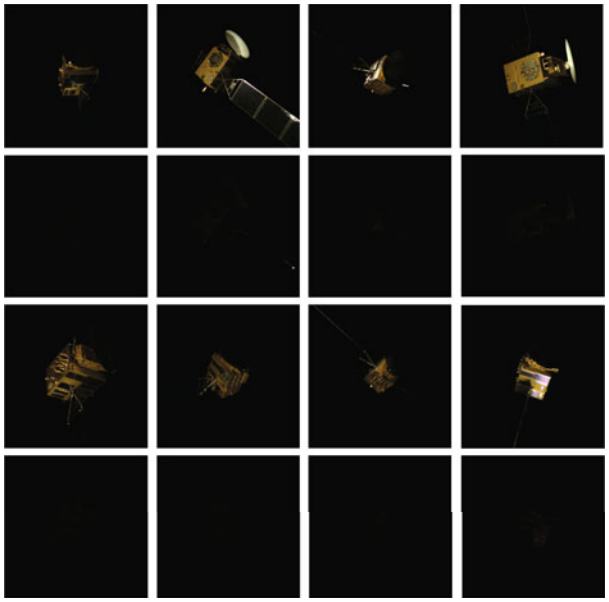

Fig. 7 Images in our dataset. The first and third rows are normal-light images, while the second and fourth rows are low-light images

4 LLIE for satellite images

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.

4.1 Diffusion model

Diffusion models comprise a forward diffusion process and a reverse denoising process, as illustrated in Fig. 8. The formulation of the diffusion

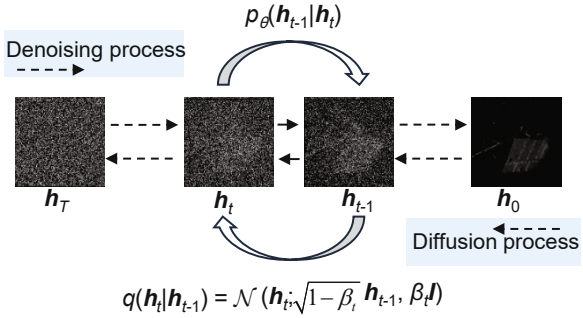

Fig. 8 Diffusion model process. In the forward diffusion process, random Gaussian noise ϵ controlled by the timestep is gradually added to the ground truth (GT) image until the image is completely turned into noise h_T. At the same time, a neural network is trained to predict the Gaussian noise. In the reverse process, ϵ at each step is estimated to recover h_0 from the noise

model-based LLIE problem is demonstrated. A low-light image and high-light image pair $(\boldsymbol{l},\boldsymbol{h})$ is defined. Noise is added to $\boldsymbol{h}$ for T times until turning it into an isotropic 2D Gaussian noise. The noised $\boldsymbol{h}$ at step t is designated as $\boldsymbol{h}_t$ $(\boldsymbol{h}_0=\boldsymbol{h})$, and the noise $\boldsymbol{\epsilon}$ at step t is scheduled by hyper-parameters β_t $(0<\beta_1<\beta_2<\ldots<\beta_T\ll 1)$. The Markovian chain is described as follows:

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

where $\mathcal{N}$ represents the normal distribution.

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

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

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

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

where θ stands for the network parameters, and μ_θ and Σ_θ are the mean and covariance to be estimated, respectively.

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 $p(\boldsymbol{h}_{t-1} \mid \boldsymbol{h}_t)$ is transformed into approximating the posterior probability distribution with $q(\boldsymbol{h}_{t-1} \mid \boldsymbol{h}_t, \boldsymbol{h}_0)$, which is tractable:

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

where

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

Since $\boldsymbol{h}_0$ has a relationship with $\boldsymbol{h}_t$ as described in Eq. (4), $\tilde{\mu}_t$ is formulated as

$$\tilde{\mu}_t(\boldsymbol{h}_t, \boldsymbol{h}_0) = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{h}_t - \frac{1-\alpha_t}{\sqrt{1-\gamma_t}}\boldsymbol{\epsilon}\right). \tag{8}$$

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 ground truth (GT) of noise is recorded in the diffusion process. Therefore, the network is trained to predict noise $\boldsymbol{\epsilon}$ by

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

4.2 FAG

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

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

where u and v are the pixel labels, and $\oplus$ means pixel-wise addition. The first part is the reverse of the reversed intensity (RI) channel after converting low-light images into the hue/saturation/intensity (HSI) color space, which guides the network to lighten the dark regions. λ is a parameter controlled by the degree of image darkness. When the image is extremely dark, the intensity channel becomes 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 $H(\cdot)$ after fast Fourier transform $\mathrm{FFT}(\cdot)$. The FAG illustration is shown in Fig. 9. The value of λ is empirically set to be 20, and the texture details are strengthened while the highly reflective regions are suppressed.

4.3 Network structure

The network consists of an encoder and a decoder with residual blocks as the core building block. 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 better performance. We rescale the map with four convolutional layers and obtain fused attention of different scales 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 passes through a fused layer to obtain the fused feature map. In addition to retaining feature information, the residual block embeds the timestep t, which is exposed to sinusoidal position encoding and multi-layer perceptrons (MLP) 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 Conv2d layers. 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., 2019). Experiments were

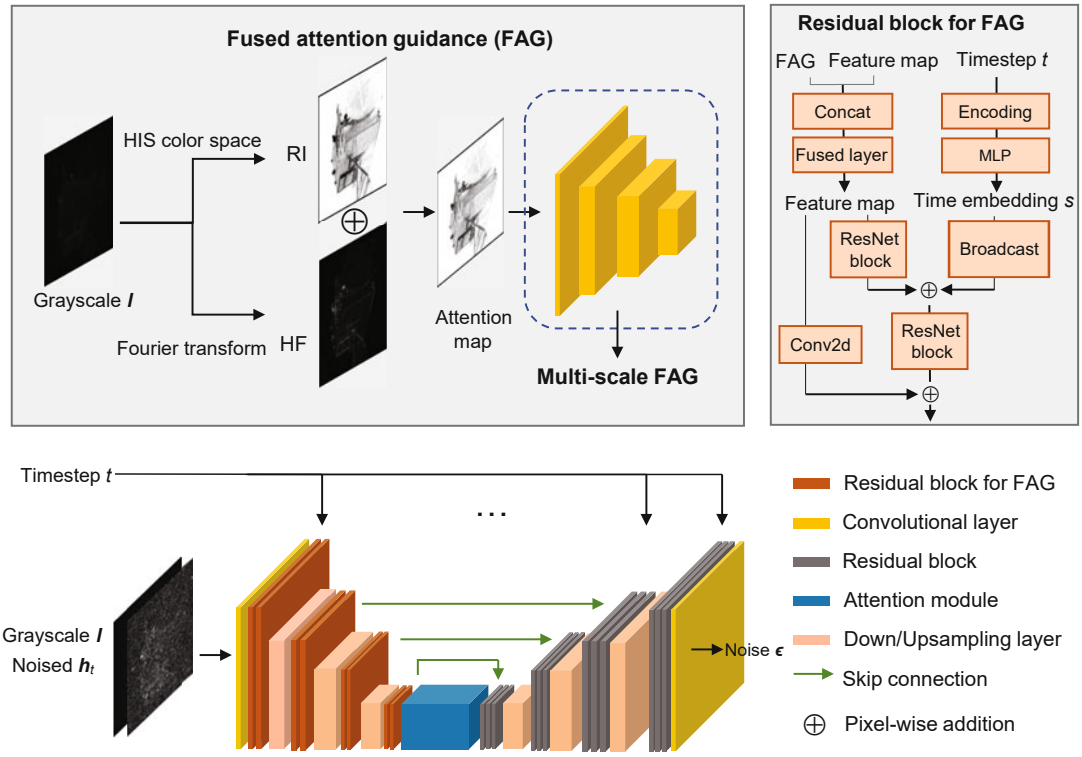

Fig. 9 Network structure. The network takes the grayscale l, h_t, and t as input. l is the low-light image to be lightened. h_t is the noised normal-light image at timestep t in the forward process. The network follows a U-Net structure. Timestep t is fed into a sinusoidal position encoding and MLP to 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. Conv2d: 2D convolutional layer; ResNet: residual network

run on an NVIDIA TITAN RTX graphics processing unit (GPU) with an Intel Xeon E5-2678 central processing unit (CPU). The dataset included a total of 720 pairs of images, with 700 pairs for training and 20 pairs for testing. Input images were rescaled and cropped to 256×256 pixels 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 a cosine schedule with an offset of 8×10^{-3} to provide a linear drop-off of noise weight in the middle stages of the process, while maintaining stability at both ends to avoid abrupt changes in the noise level (Nichol and Dhariwal, 2021). The network was trained for 100 epochs and aimed at reducing a simple loss l_2 (based on the squared l_2 norm). 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 decaying following cosine annealing (Loshchilov and Hutter, 2017) after reaching $\mathrm{lr} = 1 \times 10^{-4}$. The dynamics of the training process is shown in Fig. 10. In the early stages of training, loss oscillations and overfitting existed. However, as the learning rate decreased and dropout took effect, the network kept learning steadily and gradually converged.

5.2 On-orbit LLIE effect analysis

To test the performance of our method, we compared it with several existing methods, including traditional methods such as EFF (Ying et al.,

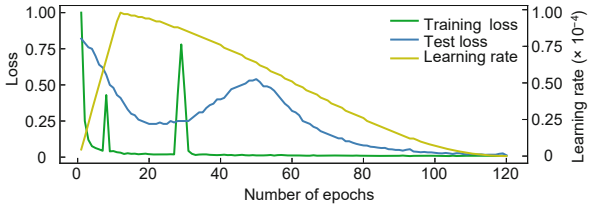

Fig. 10 Curves of the learning rate, training loss, and test loss. References to color refer to the online version of this figure

2017), JED (Ren et al., 2018), and RRM (Li MD et al., 2018), and DL methods such as RetinexNet (Wei et al., 2018), WaveNet (Dang et al., 2023), Retinex-based Transformer (Retinexformer) (Cai et al., 2023), lightweight YUV Transformer-based network (LYTNet) (Brateanu et al., 2024), pyramids diffusion (PyDiff) (Zhou et al., 2023), and global structure-aware diffusion (GDiff) (Hou et al., 2024). All these methods were evaluated with image quality assessment (IQA) metrics like the peak signal-to-noise ratio (PSNR), structure similarity index measure (SSIM), feature similarity index measure (FSIM), and learned perceptual image patch similarity (LPIPS). These metrics are described as follows:

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

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

FSIM uses the gradient magnitude and phase congruency to identify texture features (Sara et al., 2019).

LPIPS uses a pre-trained network to calculate feature similarity based on semantic information and higher-level vision features (Zhang R et al., 2018).

The comparison results are illustrated in Table 2. It can be seen that our method achieved 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 achieved 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 the SOTA methods comprehensively. Fig. 11 demonstrates the visual comparison between our method and traditional methods. EFF showed insufficient contrast enhancement, while RRM tends to overbrighten the image. JED enlarged the noise in the dark region and blurred details while increasing the illumination. Fig. 12 illustrates the visual comparison between our method and other DL methods. RetinexNet and WaveNet resulted in varying degrees of detail blurring and image distortion. LYTNet and Retinexformer were successful in learning the distribution of the target domain, but they degraded the image resolution. Compared with other newly proposed diffusion-based methods, our method enhanced the low-light images with less noise and fewer artifacts.

5.3 Ablation study

We performed an ablation study to verify the effectiveness of FAG. Fig. 13 shows the comparison of enhanced images with and without FAG. The first column shows the input low-light images. The second column displays the results without FAG, which show unreal texture artifacts and fail to lighten extremely dark regions. The third column displays the results generated with FAG. The ablation study of IQA metrics is displayed in Table 3, from which we can conclude that FAG improved the quality of feature extraction during the downsampling process. The visual results and IQA metrics indicated that our method achieved superior performance in low-light image enhancement and fidelity maintenance.

5.4 Engineering applicability analysis

After contrast enhancement, the low-light images can provide more information for high-level

Table 2 Comparison of IQA metrics between our method and previous methods

IQA metric	Traditional method			DL method						Ours
	EFF CAIP'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	
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 upward arrow indicates that a higher value is better, while the downward arrow indicates that a lower value is better. The optimal values are in bold and the suboptimal values are underlined. RNet: RetinexNet; WNet: WaveNet; RF: Retinexformer

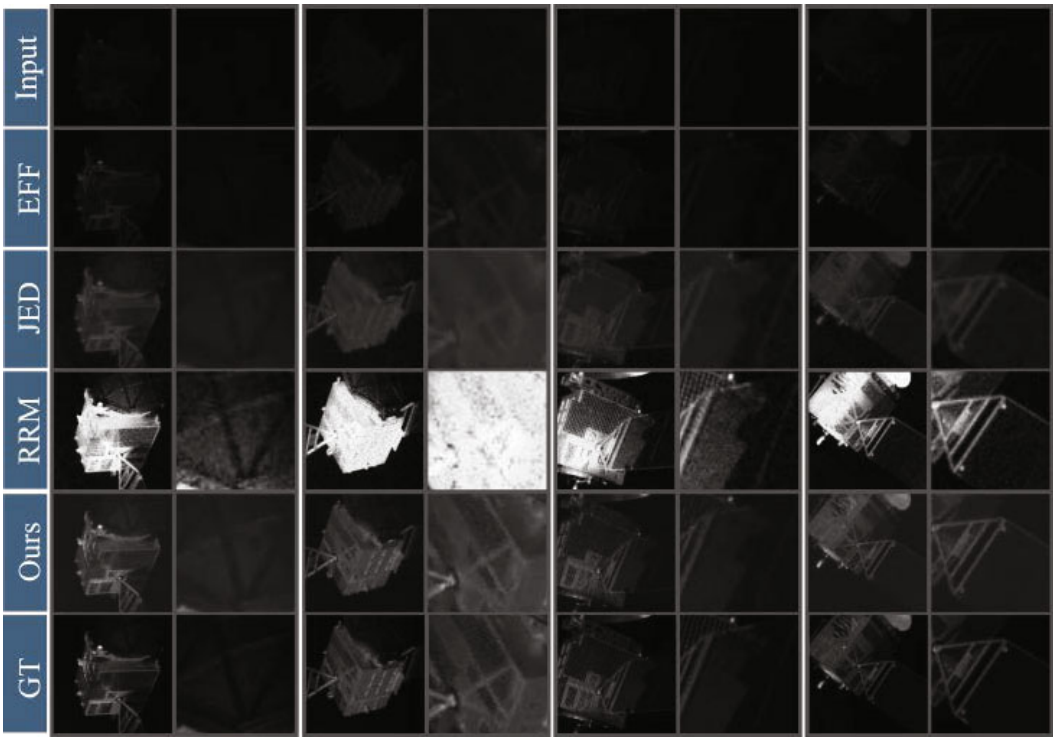

Fig. 11 Visual comparison of the enhanced results and enlarged local details between our method and traditional methods EFF (Ying et al., 2017), JED (Ren et al., 2018), and RRM (Li MD et al., 2018)

Table 3 IQA metrics with and without FAG

IQA metric	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

w/: with; w/o: without. The upward arrow indicates that a higher value is better, while the downward arrow indicates that a lower value is better

Table 4 Number of feasible SIFT features of enhanced images of continuous frames

Image	Number of SIFT features				
	A	B	C	D	Average
Raw	1	0	0	0	0.25
Enhanced	140	130	159	83	128

A, B, C, and D: four pairs of continuous frames

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, and after our enhancement, more features were extracted.

6 Conclusions

This paper explores the on-orbit low-light image enhancement problem of BeiDou navigation satellites 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 the 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 detail blurring. Experiments indicate that our approach outperforms SOTA on-orbit LLIE. In addition, this method can 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 focus on studying 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

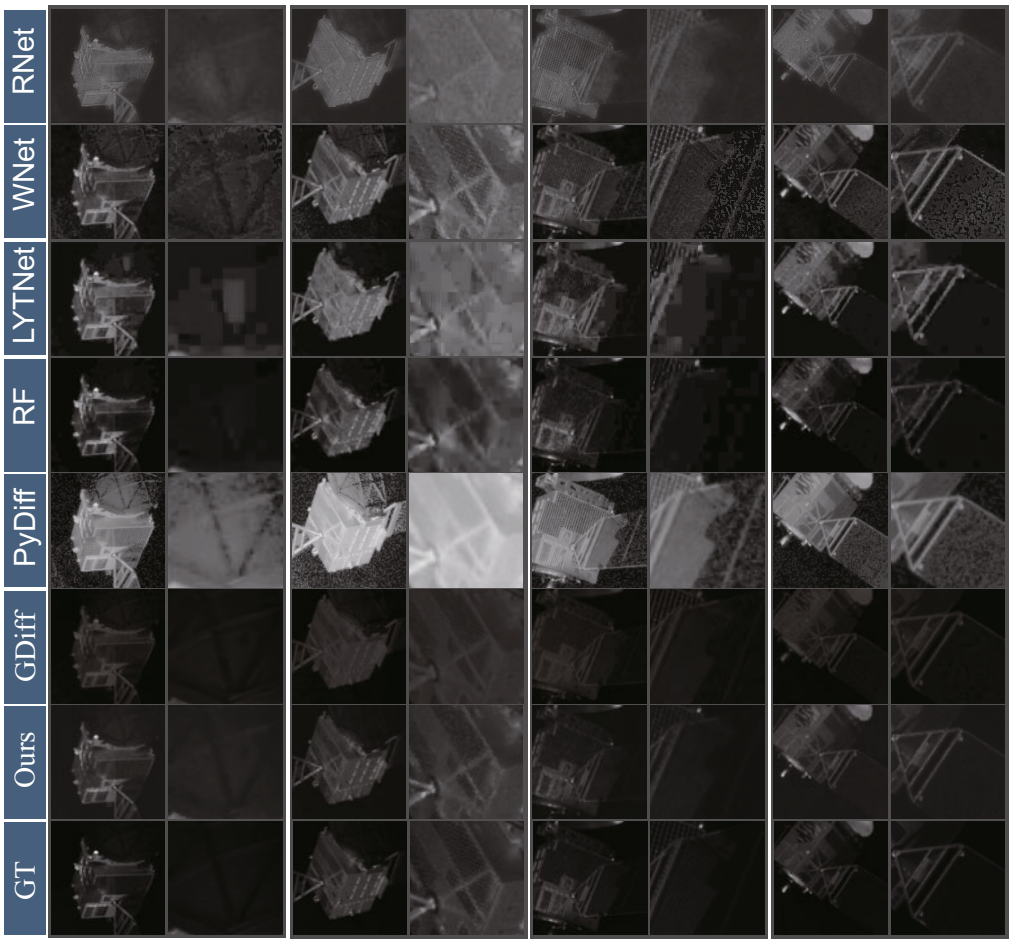

Fig. 12 Visual comparison of the enhanced results and enlarged local details with different DL methods, including RetinexNet (RNet) (Wei et al., 2018), WaveNet (WNet) (Dang et al., 2023), LYTNet (Brateanu et al., 2024), Retinexformer (RF) (Cai et al., 2023), PyDiff (Zhou et al., 2023), and GDiff (Hou et al., 2024)

diffusion model.

Contributors

Yiman ZHU designed the method and drafted the paper. Lu WANG revised the paper. Jingyi YUAN assisted with the experiment. Yu GUO supervised the project and finalized the paper.

Conflict of interest

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

Data availability

The data that support the findings of this study are available from the corresponding author or the first author upon reasonable request.

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.12470-12479. 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*

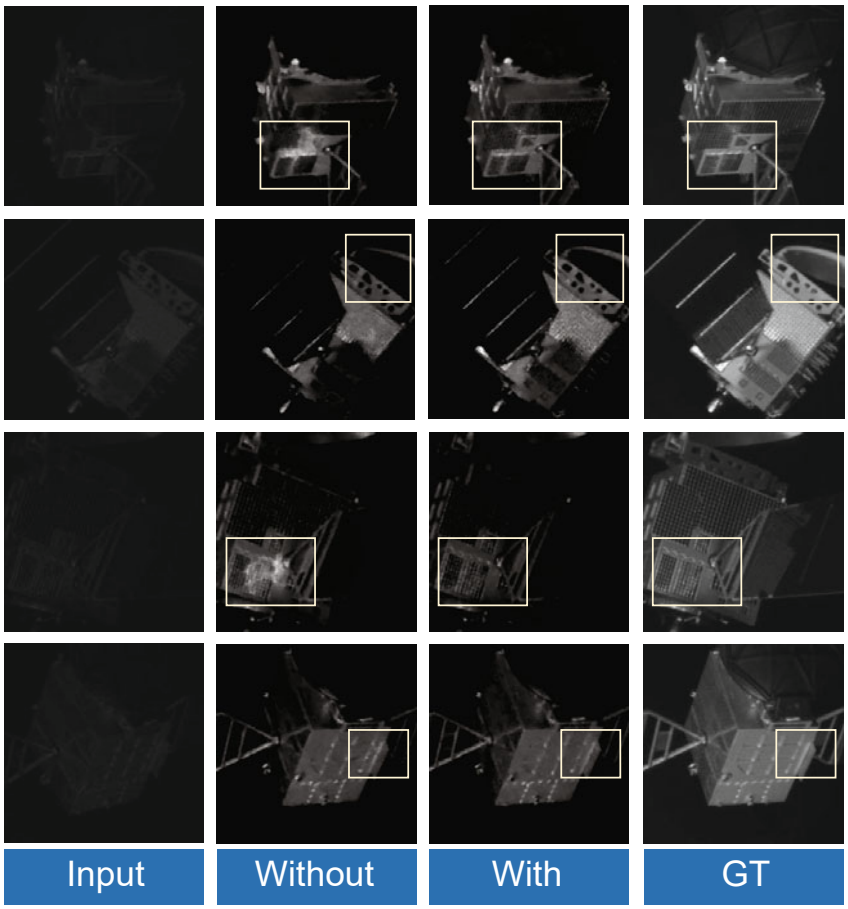

Fig. 13 Visual comparison of the enhanced results with and without FAG. The yellow boxes highlight areas with significant improvement. GT: ground truth. References to color refer to the online version of this figure

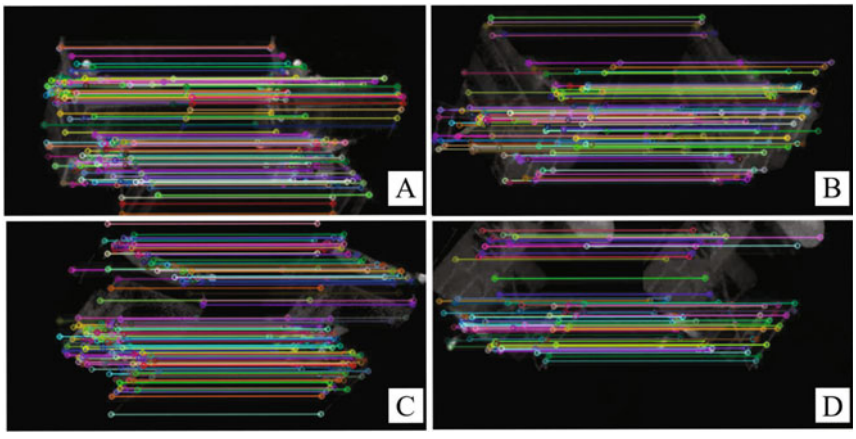

Fig. 14 SIFT matching results of enhanced images of continuous frames. A, B, C, and D: four pairs of continuous frames

Res, 73(11):5501-5520. https://doi.org/10.1016/j.asr.2023.03.022

Cowardin H, Miller R, 2022. The intentional destruction of cosmos 1408. *Orbit Debr Q News*, 26(1):1-5.

Croitoru FA, Hondru V, Ionescu RT, et al., 2023. Diffusion models in vision: a survey. *IEEE Trans Patt 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. Proc 31st Pacific Conf on Computer Graphics and Applications, p.21-29. https://doi.org/10.2312/pg.20231267

Diao HF, Li Z, Ma ZH, 2011. Simulation of space-based visible surveillance images for space surveillance. Int Symp on Photoelectronic Detection and Imaging, 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 space-based sensors using metaheuristic optimization techniques. Proc Advanced Maui Optical and Space Surveillance Technologies, p.1-13.

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.15865-15874. https://doi.org/10.1109/CVPR46437.2021.01561

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 Patt 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

Lin LQ, Li ZK, Li RK, et al., 2024. Diffusion models for time-series applications: a survey. *Front Inform Technol Electron Eng*, 25(1):19-41. https://doi.org/10.1631/FITEE.2300310

Lore KG, Akintayo A, Sarkar S, 2017. LLNet: a deep autoencoder approach to natural low-light image enhancement. *Patt Recogn*, 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.

Lu SQ, Guan FX, Zhang HY, et al., 2024. Speed-up DDPM for real-time underwater image enhancement. *IEEE Trans Circ 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 on Artificial Intelligence, p.8828-8836. https://doi.org/10.1609/aaai.v35i10.17069

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.

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., 2019. PyTorch: an imperative style, high-performance deep learning library. Proc 33rd Int Conf on Neural Information Processing Systems, Article 721.

Poozhiyil M, Nair MH, Rai MC, 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., 2018. 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, 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 Patt 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 Imag*, 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 on Computer Vision, p.103-119. https://doi.org/10.1007/978-3-030-58601-0_7

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. Exposure diffusion: 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.

Xu YD, Yang C, Sun BB, et al., 2021. A novel multi-scale fusion framework for detail-preserving low-light image enhancement. *Inform 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. *Inform Fus*, 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 Trans Transp Electrific*, 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):1-39. 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*, 60:1-11. https://doi.org/10.1109/TGRS.2021.3131264

Ying ZQ, Li G, Ren YR, et al., 2017. A new image contrast enhancement algorithm using exposure fusion framework. Proc Int Conf on Computer Analysis of Images and Patterns, p.36-46. https://doi.org/10.1007/978-3-319-64698-5_4

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/10.1109/CVPR.2018.00068

Zhou DW, Yang ZX, Yang Y, 2023. Pyramid diffusion models for low-light image enhancement. Proc 32nd Int Joint Conf on Artificial Intelligence, p.1795-1803. https://doi.org/10.24963/ijcai.2023/199