

Article

Not peer-reviewed version

A Deep Learning-Based Method for Enhancing the Signal-to-Noise Ratio of Star Sensor Images

[Jian Guan](#) , Hanye Yu , [Yanpeng Wu](#) * , Xiaofeng Li , Rongzheng Cao

Posted Date: 28 May 2026

doi: 10.20944/preprints202605.1936.v1

Keywords: star sensor; detector readout noise; spatial stray-light interference; deep learning; frame-by-frame spatial deformable convolution; frequency-domain loss constraint



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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

Article

A Deep Learning-Based Method for Enhancing the Signal-to-Noise Ratio of Star Sensor Images

Jian Guan ^{1,2}, Hanye Yu ², Yanpeng Wu ^{2,*}, Xiaofeng Li ² and Rongzheng Cao ²

¹ School of Computer Science and Technology, Xidian University, Xi'an 710126, China

² Beijing Institute of Control Engineering, Beijing 100190, China

* Correspondence: 20033110168@stu.xidian.edu.cn

Abstract

In window tracking mode, stray light and detector readout noise can submerge star-spot signals in star sensor images. The resulting degradation reduces centroid extraction accuracy and may even cause extraction failure, thereby preventing precise attitude determination. This study uses the self-supervised spatiotemporal denoising model ASTERIS as the baseline. ASTERIS integrates 3D spatiotemporal inputs with a global attention mechanism for joint noise modeling, thereby providing stronger denoising and restoration capability than conventional methods such as multi-frame stacking. However, ASTERIS lacks adaptive compensation for subpixel jitter in on-orbit star images and has difficulty preserving the high-frequency morphology of star spots, affecting denoising performance and centroiding accuracy. To address these limitations, this study introduces two improvements. First, frame-by-frame spatial deformable convolution is incorporated into the decoder upsampling stage to adaptively compensate for subpixel offsets, actively suppress background noise, and lower the parameter count. Second, a complex-valued frequency-domain loss with a high-frequency weighted mask is designed to jointly constrain the amplitude and phase spectra, thereby preserving high-frequency star-spot details. Experimental results show that, for star images with extremely low signal-to-noise ratios, the proposed method improves the peak signal-to-noise ratio by approximately 60-fold and reduces the centroid localization error to approximately 0.1 pixels. This performance is substantially better than that of the original ASTERIS model, which improves the peak signal-to-noise ratio by approximately 9-fold and yields an error of approximately 0.4 pixels, and the multi-frame stacking method, which improves the peak signal-to-noise ratio by approximately 4-fold and yields an error of approximately 0.5 pixels. The deep learning method presented in this paper provides a novel solution for centroid extraction of star sensors under strong noise interference in orbit and achieves satisfactory results. Future work will focus on lightweight network design to enable on-orbit engineering applications.

Keywords: star sensor; detector readout noise; spatial stray-light interference; deep learning; frame-by-frame spatial deformable convolution; frequency-domain loss constraint

1. Introduction

Star sensors are core optical sensors for spacecraft attitude measurement. The attitude information provided by star sensors directly affects the geocoding accuracy of remote sensing images and therefore influences downstream remote sensing applications. During operation, star sensors capture stellar images, extract the centroid coordinates of star spots, and determine spacecraft attitude or detect space targets through star-map recognition [1–3]. To improve measurement accuracy and data update rates, star sensors commonly operate in window tracking mode during on-orbit observation. In this mode, the positions of known stars in the current image are predicted from the attitude information at the previous moment, and small windows centered on these predicted

positions are cropped. A field-programmable gate array (FPGA) then processes each window in parallel and independently calculates the centroid of the corresponding star spot [4–6]. Under ideal conditions, each window contains a single star, with its centroid located at the center of the window, and the grayscale distribution follows a two-dimensional Gaussian point spread function (PSF) [7,8]. Figure 1 shows the small windows cropped by the FPGA from the original star map.

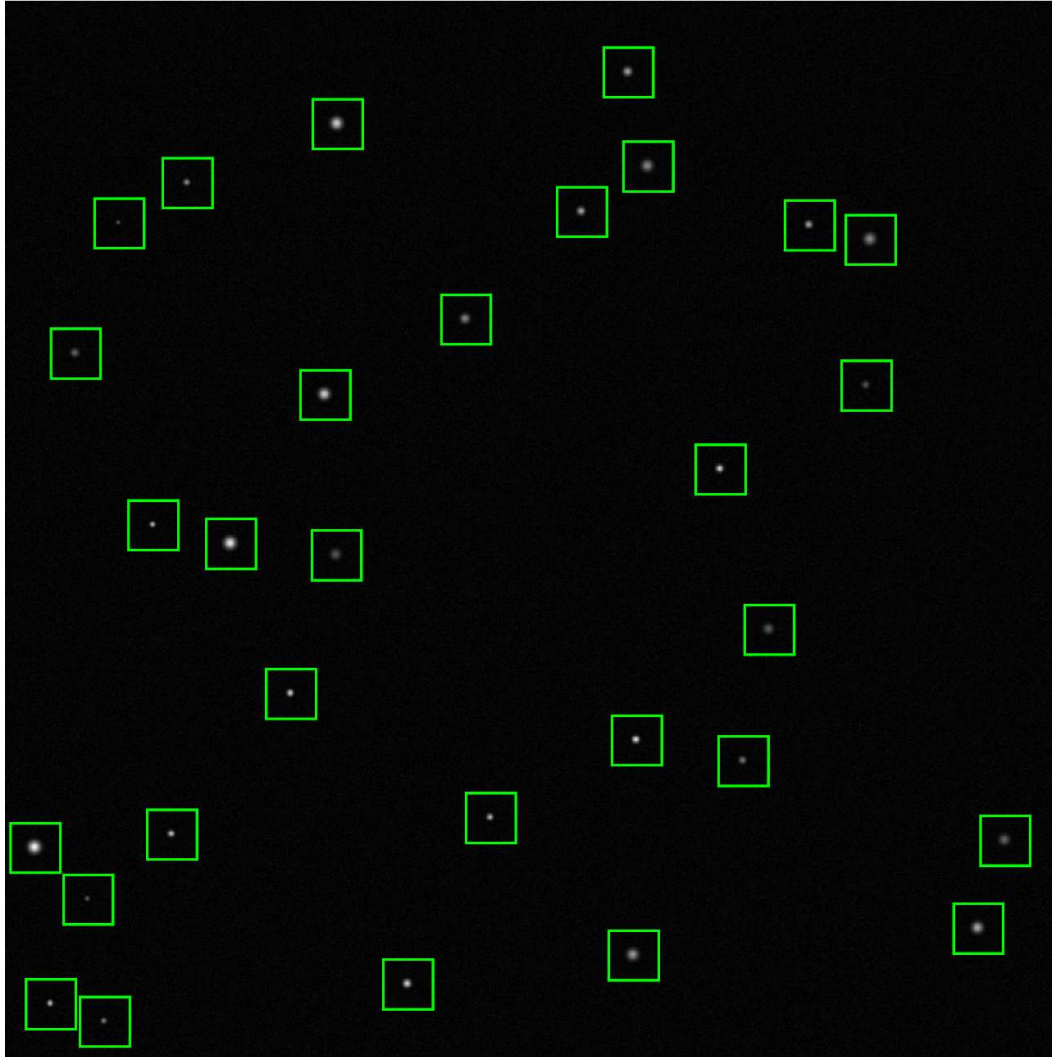


Figure 1. Schematic diagram of small windows cropped by the FPGA from the original star map.

In actual on-orbit environments, however, the quality of window images can be strongly degraded by stray-light interference from Earth airglow, direct sunlight, and internal scattering in the optical system, together with detector readout noise and dark current. Under such interference, star-spot signals are readily submerged in background noise. The resulting degradation reduces centroid extraction accuracy and may ultimately cause attitude determination failure [9,10].

Among existing stray-light suppression methods, multi-frame stacking remains the most widely used technical route in engineering applications [11,12]. Common implementations include mean stacking, median stacking, and weighted stacking [13]. Mean stacking can suppress Gaussian noise effectively, but this strategy is sensitive to outlier frames, namely abnormal observations caused by sudden and short-duration interference in a continuously acquired image sequence. Median stacking is robust to transient noise such as cosmic rays, but this strategy cannot remove spatially continuous background gradients. Weighted stacking assigns weights according to the inter-frame signal-to-noise ratio, yet the weight-estimation process is itself easily disturbed by noise [14,15]. Overall, multi-

frame stacking methods share three limitations: the statistical assumptions for stray light and noise are overly idealized, adaptive compensation for subpixel jitter of star spots is limited, and high-frequency star-spot details are often weakened during background suppression. These limitations may lead to missed detection of dim stars or photometric distortion [16,17].

Methods beyond multi-frame stacking also have clear constraints. Wavelet-threshold denoising [9] suppresses noise by exploiting sparsity in the wavelet domain, but threshold selection is empirical and the method is prone to pseudo-Gibbs artifacts. Non-local means (NLM) filtering [11] performs weighted averaging based on image self-similarity and is effective for uniform-background denoising, although the computational complexity is high and texture preservation remains limited. Three-dimensional block matching (BM3D) and the volumetric extension BM4D [14,15] achieve strong denoising performance through collaborative filtering of grouped blocks. However, these methods require accurate block matching, are prone to failure in star maps affected by severe stray-light interference, and are difficult to deploy in real-time star-sensor systems. In general, these traditional methods depend on manually designed priors, provide limited adaptive modeling of spatially non-stationary stray light and spectral aliasing between stray light and star-spot signals, and often incur high computational cost. Therefore, this study uses multi-frame stacking, the most common and computationally simple engineering method, as the representative traditional baseline to evaluate the improvement provided by deep learning.

Deep learning has recently provided a promising route for star-map restoration. Rather than relying on manually specified priors, deep learning methods can learn restoration priors directly from image pixels and are therefore less constrained by the assumptions that limit traditional algorithms. Convolutional neural network (CNN)-based methods have shown good performance in star-spot detection and centroid positioning [18,19]. Transformer architectures have further improved image-restoration capability. Restormer [20] uses Multi-Dconv Head Transposed Attention (MDTA) to perform global interaction along the channel dimension and combines MDTA with a Gated-Dconv Feed-Forward Network (GDFN), thereby improving restoration quality while maintaining computational efficiency. The cutting-edge achievement model ASTERIS [21] extends Restormer to 3D spatiotemporal sequences and uses self-supervised learning to improve the detection sensitivity of the James Webb Space Telescope by one magnitude.

Despite these advances, directly applying existing deep learning methods to star-sensor window tracking remains challenging, particularly under severe stray-light interference that submerges star spots in noise. Two limitations are especially important. First, existing methods provide insufficient adaptive compensation for subpixel-level jitter of star spots. In orbit, platform jitter and optical distortion cause small inter-frame shifts of star spots, preventing accurate registration and alignment of multi-frame star images and consequently degrading the denoising performance of models such as ASTERIS [21]. Ordinary convolution responds weakly to such displacement, leading to feature misalignment. Second, existing methods lack an explicit mechanism for preserving the high-frequency energy of star spots. The star spot signal appears as a medium-to-high frequency component in the window image [22]. During training, a network may weaken star-spot sharpness to obtain a smoother background, which appears as reduced peak intensity and an increased PSF diffusion radius. To address these limitations, this study introduces two improvements.

First, frame-by-frame spatial deformable convolution, hereafter referred to as deformable convolution, is incorporated into the decoder to replace standard 3D convolution. In the decoder upsampling stage, a spatial offset field is predicted independently for each frame and then used for two-dimensional deformable convolution. This design enables the sampling positions of the convolution kernel to adaptively follow the actual locations of star spots. The strategy is inspired by deformable convolution [23] and its lightweight improvement [24]. Placing deformable convolution in the upsampling stage is advantageous because high-resolution feature maps support more accurate offset estimation, and deformable convolution close to the output end can more directly correct positional errors accumulated by the encoder. The operation compensates only for

displacement in the spatial dimensions and does not alter the temporal axis, which is consistent with the physical fact that star-spot displacement occurs within the image plane.

Second, a frequency-domain constraint is introduced into the loss function. A complex-domain frequency-domain loss and a high-frequency weighted mask, hereafter referred to as the frequency-domain loss constraint, are used to explicitly constrain the high-frequency components of star spots. This design draws on the structural similarity index [25] and joint wavelet-frequency-domain loss [26]. Strong denoising performance is obtained when the magnitude of this loss is comparable to that of the spatial-domain loss. Recent methods based on diffusion models [27], implicit neural representations [28], and joint loss functions integrating frequency-domain and perceptual losses [29] also support the effectiveness of frequency-domain constraints.

Because noise-free ground-truth star maps are difficult to obtain during on-orbit operation, self-supervised learning provides a practical training strategy. Self-supervised frameworks such as Noise2Noise [30], Noise2Void [31], and Noise2Self [32] establish the theoretical basis for training without clean targets. CNN-based denoising methods such as DnCNN [33] and FFDNet [34s] have also informed subsequent research. Both convolution-based 3D Res U-Net [19] and Transformer-based Restormer [20] can be trained within a self-supervised framework. ASTERIS adopts this self-supervised spatiotemporal denoising strategy to enhance deep-space images without ground truth. With ASTERIS as the baseline, this study also uses a self-supervised strategy and trains the model with multi-frame noisy sequences acquired in the window tracking mode of star sensors.

Interference in actual on-orbit environments can be broadly divided into two categories [9,10]:

The first category is detector readout noise, which refers to random fluctuations generated by electronic circuits during image-sensor signal readout, including reset noise, amplification noise, and quantization noise. This type of noise approximately satisfies a zero-mean independent and identically distributed assumption in the temporal domain and appears spatially as an isotropic random shot-like pattern.

The second category is spatially non-stationary stray light, which refers to background radiation entering the optical system from the external spacecraft environment, including Earth airglow, moonlight, and internal scattering in the optical system. This interference forms a spatially nonuniform but temporally slowly varying or stable distribution on the image plane. Its typical characteristics include spatial gradients or local highlights, approximate constancy within a short-time window, and spectral concentration in the low-frequency region.

Traditional multi-frame stacking methods suppress zero-mean random noise by exploiting temporal redundancy. For spatially non-stationary stray light, however, stacking can only preserve the mean background and cannot improve spatial nonuniformity, even when the interference is independent between frames [11–15]. The effectiveness of self-supervised deep learning methods also depends on the assumption that inter-frame noise is independent and identically distributed. For interference with a nonzero mean, such as dark-field noise or fixed-pattern noise, any method based only on temporal independence cannot remove the mean component, because both stacking and self-supervised learning preserve this component.

The simulation dataset used in this study mainly contains detector readout noise and spatial random stray-light interference, both of which are independently generated for each frame and have an approximately zero statistical expectation. Under these conditions, deep learning methods can fully exploit their advantages. The remainder of this paper is organized as follows: Section 2 reviews related work, Section 3 describes the proposed method, Section 4 presents the experimental design and results, and Section 5 summarizes the study and discusses future directions.

2. Related Work

2.1. Traditional Star-Image Denoising Methods

Traditional star-image denoising methods can be divided into spatial-domain filtering, transform-domain filtering, and multi-frame stacking. Spatial-domain filtering methods, including

Gaussian, median, and bilateral filtering, are simple to implement but have clear limitations. Gaussian filtering blurs edges, median filtering cannot handle gradually varying backgrounds, and bilateral filtering requires careful parameter tuning [11,12]. Transform-domain filtering methods, such as DCT and wavelet-threshold denoising [9], depend strongly on threshold selection and basis-function matching and are prone to distortion when spectral overlap occurs [13,14]. Multi-frame stacking is the most widely used approach. Mean stacking is susceptible to outlier frames, median stacking cannot remove spatially continuous backgrounds, and improved methods based on image autocorrelation [16] depend on the success rate of initial star-spot extraction. Overall, traditional methods rely on overly idealized statistical assumptions about stray light. This dependence makes it difficult to preserve the PSF morphology of star spots while suppressing the background and can lead to missed detection of dim stars or grayscale distortion [17].

2.2. Deep Learning-Based Methods

Self-supervised learning: Noise2Noise demonstrates that, under zero-mean inter-frame independent noise, training can be performed using only pairs of noisy images. The theoretical basis is as follows: for a noisy pair in the form $x_1=y+n_1$ and $x_2=y+n_2$, where y denotes the ground truth and n_1 and n_2 denote noise interference, minimizing the difference between x_1 and x_2 is equivalent to learning a mapping from noisy images to the clean signal y . Noise2Void and Noise2Self further enable training from a single noisy image by constructing self-supervised signals through blind-spot networks or pixel-masking strategies. Noise2Noise does not use temporal-axis information in multi-frame sequences, whereas 3D models such as ASTERIS stack sequences into 3D tensors and use spatiotemporal self-attention to jointly model spatial details and temporal dependencies. This strategy provides stronger performance under low signal-to-noise ratios. The method can also reduce the influence of outlier frames in multi-frame stacking, because the grayscale distribution of an outlier frame differs substantially from that of other frames in the sequence. To minimize fitting error, the network tends to treat such frames as noise and suppress their contribution.

3D Res U-Net architecture, abbreviated as 3DRU-Net: This architecture introduces a 3D encoder-decoder structure with residual connections [19], which alleviates gradient vanishing and has been widely used in medical image segmentation and denoising. However, its convolution operations are local operators, and long-range spatiotemporal dependencies must be transmitted indirectly through repeated downsampling and upsampling. Under the operating conditions considered in this study, this limitation can cause loss of high-frequency star-spot details. Static convolution kernels also assign the same spatial weights to all inputs, which makes adaptive discrimination between star spots and stray-light backgrounds difficult.

Transformer architecture: Restormer [20] uses MDTA to perform global interaction along the channel dimension and combines MDTA with GDFN to improve restoration quality efficiently. MDTA transposes self-attention computation from the spatial dimension to the channel dimension, reducing the computational complexity from $O(N^2)$ to $O(C^2)$, where N is the number of pixels and C is the number of channels, with C much smaller than N . ASTERIS [21] extends Restormer to 3D spatiotemporal sequences and adopts self-supervised spatiotemporal denoising. Unlike 3DRU-Net, which relies on local convolution and indirect transmission of long-range dependencies, the Restormer-based 3D Transformer enables each spatial position to interact directly with all positions in the image through MDTA, thereby providing a global receptive field. Dynamic attention weights can also focus adaptively on star-spot regions according to the input signal. Experiments show that, under low signal-to-noise ratios, ASTERIS achieves a substantially greater PSNR improvement than 3DRU-Net [21].

Deformable convolution: Deformable convolution [23] enables convolution kernels to adaptively follow target deformation by learning sampling-point offsets and can compensate for subpixel displacement of star spots. These offsets are predicted from the input feature map by an additional convolutional layer. Directly using 3D deformable convolution, however, introduces unnecessary offsets in the temporal dimension and entails high computational overhead. Recent

research [24] has explored the placement of deformable convolution in the decoder stage. Inspired by this strategy, this study applies frame-by-frame spatial deformable 2D convolution in the upsampling stage. The experiments show that this placement is superior to applying deformable convolution in the downsampling stage, although the downsampling-stage comparison is not discussed in detail.

3. Proposed Method

3.1. Problem Formulation

Given an input sequence of window images with a window size of $H \times W$, the ideal grayscale distribution of the star spot follows the PSF [7], whereas the observed image is degraded by stray light and noise interference. The objective is to learn the mapping between the degraded observation and the ideal star-spot signal so that a clean image can be restored.

Each noisy image frame I_t can therefore be regarded as the superposition of three types of signals as follows:

$$I_t = S_t + B_t + N_t$$

where S_t denotes the clean star-spot signal, whose grayscale distribution approximately follows the PSF; B_t denotes spatial random stray light, which is independent between frames and has an approximately zero mean; and N_t denotes detector readout noise, which approximately satisfies a zero-mean Gaussian independent and identically distributed assumption in the temporal sequence as follows:

$$N_t \sim N(0, \sigma_N^2)$$

The objective of this study is to design a trainable mapping function that takes a multi-frame sequence I_t as input and outputs the corresponding restored multi-frame image S_t .

3.2. Baseline ASTERIS Model

This study adopts ASTERIS as the backbone network [21]. The core of ASTERIS is a 3D Transformer encoder-decoder structure. The input sequence is stacked into a 3D tensor $X (T \times H \times W)$, processed by a 3D convolutional layer, and then fed into the MDTA module. MDTA performs self-attention along the channel dimension, enabling each channel to interact with all spatiotemporal positions and thereby providing a global receptive field. GDFN enhances nonlinear representation through a gating mechanism [20]. By stacking multiple Transformer blocks, the network progressively learns the mapping from noisy observations to the clean signal.

ASTERIS uses a self-supervised learning strategy in which adjacent-frame noisy image pairs serve as supervisory signals. When the noise satisfies zero-mean and inter-frame independence assumptions, learning inter-frame mappings enables the network to converge toward the denoised clean signal. The loss function jointly optimizes two spatial-domain losses. $L1_{stack}$ is a smooth L1 loss that constrains pixel-level differences between the single-frame output and the target, thereby preserving high-frequency details and star-spot edges. $L2_{mean}$ is a mean squared error loss for the sequence-mean image, which suppresses the accumulation of inter-frame random noise and improves the signal-to-noise ratio. These two losses provide complementary constraints.

3.3. Theoretical Advantages over Multi-Frame Stacking

Traditional multi-frame stacking linearly combines pixel values along the temporal axis, and its denoising capability is derived from temporal redundancy. This strategy has three inherent limitations. First, for spatially non-stationary backgrounds, stacking can only preserve the mean

distribution and cannot improve spatial nonuniformity. Second, stacking cannot compensate for subpixel jitter of star spots, so direct stacking increases the energy diffusion radius. Third, stacking is essentially a low-pass filtering operation that weakens the high-frequency morphology of star spots. The improved ASTERIS-based deep learning method used in this study addresses these limitations through four aspects.

First, spatial information helps suppress non-stationary backgrounds. Through 3D convolution and attention mechanisms, the deep learning model learns the spatial distribution of the background within a single frame and uses a large receptive field to smooth the background adaptively. Even when the background mean is nonzero, the network can use spatial context to attenuate the visual impact of the background. By contrast, multi-frame stacking fully preserves the mean background and is therefore less effective at improving the signal-to-noise ratio.

Second, deformable convolution provides adaptive compensation for subpixel jitter. By learning spatial offsets for each sampling point, deformable convolution enables the sampling grid of the convolution kernel to align actively with the center of the star spot and supports precise subpixel-level fusion.

Third, the frequency-domain loss explicitly preserves the high-frequency morphology of star spots. This loss constrains the amplitude and phase spectra simultaneously and strengthens the mid- and high-frequency components through a high-frequency weighted mask. This design helps avoid excessive smoothing, whereas the low-pass characteristics of multi-frame stacking tend to weaken star-spot edges.

Fourth, MDTA provides joint spatiotemporal modeling. This mechanism enables each spatial position to perceive information from all temporal frames simultaneously, allowing the network to distinguish temporally varying star spots from temporally stable backgrounds and to suppress the background spatially. The experiments in Section 4 verify these advantages.

3.4. Frame-by-Frame Spatial Deformable Convolution

3.4.1. Design Motivation

Deformable convolution is applied in the upsampling stage for three reasons. First, the feature maps in the upsampling stage have higher resolution, which supports accurate prediction of subpixel offsets. Second, the decoder fuses details lost during downsampling, and deformable convolution can directly correct accumulated positional errors. Third, gradient propagation is more direct near the output end, which improves training stability. Platform vibration, optical distortion, and other factors can also cause residual subpixel inter-frame registration errors, usually within 0.2 pixels. Deformable convolution compensates for these errors adaptively by learning spatial offsets for each sampling point, thereby improving cross-frame fusion accuracy. Replacing standard 3D convolution in the decoder with frame-by-frame 2D deformable convolution does not affect temporal modeling, because temporal dependencies are mainly captured by encoder MDTA and transmitted to the decoder through residual skip connections [21]. This design also reduces the number of parameters by approximately one quarter without increasing time overhead. Therefore, deformable convolution is integrated into the upsampling stage of the ASTERIS decoder.

Deformable convolution can also suppress background noise. Star-spot energy is locally concentrated, whereas background noise is spatially diffuse. Standard convolution samples background noise indiscriminately on a regular grid. By contrast, deformable convolution learns spatial offsets that guide sampling points toward star-spot regions and away from the background. During training, the offset-prediction network shifts sampling points near star spots toward the star-spot center, whereas sampling points in background regions are guided toward informative star-spot regions or kept at zero offset. The contribution of the background is therefore relatively suppressed. In essence, this operation performs adaptive spatial resampling. By minimizing the reconstruction loss, the sampling grid gradually aligns with the star-spot region and focuses on star-spot features. Unlike traditional denoising methods such as mean filtering, which smooth the entire image

indiscriminately, deformable convolution provides an active selection mechanism that bypasses the background adaptively and collects information from the vicinity of star spots. This mechanism suppresses background noise without weakening star-spot energy.

3.4.2. Implementation Principle

Let the input feature map be X ($N \times C \times T \times H \times W$). The procedure is described as follows:

1. Frame-by-frame splitting: the feature map is split along the temporal dimension into T two-dimensional frames X_t ($N \times C \times H \times W$).

2. Offset prediction: for each frame, a two-dimensional convolution with shared weights, Convoffset , predicts the spatial offset O_t . The number of output channels is $2G \times k^2$, where k is the convolution kernel size and G is the number of groups, which is usually set to 1. Each offset is a floating-point value that represents the subpixel displacement of a sampling point in the x and y directions. The offset-prediction formula can be written as $O_t = \text{Convoffset}(X_t)$, and the shape of O_t is $(N, 2G \times k^2, H, W)$. All T frames share the same Convoffset , thereby reducing the number of parameters.

3. Two-dimensional deformable convolution: the `torchvision.ops.deformconv2d` function provided by Python is used to perform deformable convolution on each frame. The function takes the input feature map X_t , the offset O_t , and the weight W_{dc} and the bias b_{dc} of Convoffset as inputs. The output Y_t has the shape (N, C', H', W') . All T frames share W_{dc} and b_{dc} .

4. Temporal stacking: Y_t is concatenated along the temporal dimension to obtain Y (N, C', T, H', W').

5. Offset initialization: the offsets are initialized to 0; that is, W_{dc} and b_{dc} are initially set to 0. During training, the offsets are learned progressively from small to large values. This process is consistent with the natural optimization trajectory, facilitates convergence to a favorable local optimum, and improves training stability. Figure 2 shows the improved ASTERIS model.

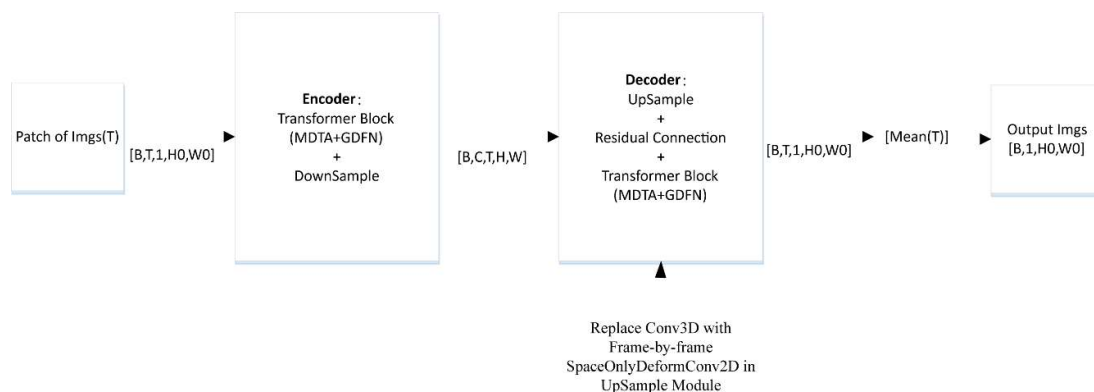


Figure 2. Schematic diagram of the improved ASTERIS model.

3.5. Frequency-Domain Loss Constraint

3.5.1. Complex-Domain Loss

Let the image size be $H \times W$. The predicted image and the target image are transformed from the spatial domain to the frequency domain by a two-dimensional Fourier transform F . In this representation, low-frequency components mainly describe smooth background variations, including interferences such as noise and stray light, whereas high-frequency components encode star-spot edges and fine details. The amplitude spectrum $|F|$ characterizes the energy of each frequency component, whereas the phase spectrum encodes positional information. For star-spot centroiding, the phase spectrum is particularly important because phase determines the position of

the star spot and the symmetry of the PSF. A constraint imposed only on the amplitude spectrum, namely $|\hat{F}|$, can preserve high-frequency energy but cannot ensure phase consistency, which may introduce positional shifts or shape distortion. Therefore, the complex-domain loss constrains amplitude and phase simultaneously as follows:

$$L_{\text{complex}} = \frac{1}{H \cdot W} \sum_{u=0}^{H-1} \sum_{v=0}^{W-1} |\hat{F}(u,v) - F_{\text{gt}}(u,v)|$$

where (u,v) denotes the frequency-domain coordinates. This loss penalizes amplitude and phase discrepancies simultaneously. Specifically, the squared modulus of the complex difference can be expanded as follows, where $(\theta_{\wedge} - \theta_{\text{gt}})$ denotes the phase difference:

$$|\hat{F} - F_{\text{gt}}|^2 = |\hat{F}|^2 + |F_{\text{gt}}|^2 - 2|\hat{F}||F_{\text{gt}}|\cos(\theta_{\wedge} - \theta_{\text{gt}})$$

3.5.2. High-Frequency Weighted Mask

The useful high-frequency energy of a star spot is mainly distributed near the boundary frequency of the PSF, whereas low-frequency components are largely associated with background variation and stray-light interference. To ensure that the frequency constraint focuses on star-spot morphology rather than smooth background components, the normalized frequency radius is defined as follows:

$$r(u,v) = \sqrt{\left(\frac{u}{H} - 0.5\right)^2 + \left(\frac{v}{W} - 0.5\right)^2}$$

The radius r lies in $[0, 0.707]$. Given a threshold r_0 , the mask is set to $M(u,v) = 1$ when $r > r_0$ and $M(u,v) = 0$ otherwise. The weighted frequency-domain loss is defined as follows:

$$L_{\text{freq}} = \frac{1}{H \cdot W} \sum_{u=0}^{H-1} \sum_{v=0}^{W-1} M(u,v) |\hat{F}(u,v) - F_{\text{gt}}(u,v)|$$

The mask assigns greater emphasis to the mid- and high-frequency components that contain star-spot boundary and shape information. At the same time, the mask suppresses low-frequency background interference and reduces unnecessary computation. The experiments in this study indicate that $r_0 = 0.2$ provides the best restoration of star-spot details. The detailed threshold-selection procedure is not discussed further.

3.5.3. Loss-Balance Coefficient

The frequency-domain loss L_{freq} is calculated in a manner consistent with $L_{1\text{stack}}$ in ASTERIS [21]. Because the spatial-domain losses $L_{1\text{stack}}$ and $L_{2\text{mean}}$ are multiplied by coefficients on the order of $1e6$ in the baseline model, a balance coefficient λ_{freq} is introduced for L_{freq} . The total loss is expressed as follows:

$$L_{\text{total}} = 0.125e6 \times L_{1\text{stack}} + 1e6 \times L_{2\text{mean}} + \lambda_{\text{freq}} \times L_{\text{freq}}$$

A grid search over $\lambda_{\text{freq}} \in \{0, 100, 500, 1000, 2000, 5000\}$ identifies 1000 as the optimal value.

3.6. Network Training

Model training uses the Adam optimizer with an initial learning rate of $1e-4$, which is annealed to $1e-5$ through a cosine schedule. The batch size is set to 3, and the network is trained for 20 epochs. The learning rate of the deformable-convolution offsets is set to $1e-5$ to avoid unstable offset updates and reduce the risk of convergence to a noise-induced local optimum.

4. Experimental Validation

4.1. Dataset and Evaluation Metrics

Dataset construction method: Real window images obtained through ground calibration are used, and the window size is 48×48. The star-spot energy and diffusion radius are randomly set. For each star spot, 4000 windows are calibrated, and the calibrated centroid is allowed to vary within 0.2 pixels around the window center, with coordinates (25,25). Setting all centroids at the window center simulates ideal registration and alignment of all frame window images, whereas centroid fluctuation within 0.2 pixels represents the alignment result under real conditions. Detector readout noise and spatial random stray light, which are uncorrelated between frames and have an approximately zero expected value, are then added to each window image through a simulation program to simulate real on-orbit operating conditions. The dataset does not include inter-frame gradually varying stray-light interference with a nonzero mean caused by solar illumination conditions. The peak signal-to-noise ratio (PSNR) of images containing stray light and readout noise ranges from -6 to 15 dB, with an average of 8 dB. Among these images, 800 windows are randomly selected for testing, with every 8 images forming one group for denoising and restoration, including multi-frame stacking. This setting yields 100 groups of test conditions, whereas the remaining images are used for training. A dataset of 30 star spots is generated in total, namely 30×4000 groups of images that cover star spots with different brightness levels and morphologies. The dataset for each star spot is divided according to the 4:1 ratio described above, and the training datasets of all star spots are mixed for model training to improve generalization across different star spots. The test dataset contains 30×100 groups of test data. The calculation formula for the peak signal-to-noise ratio is as follows:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right)$$

$$\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2$$

where MAX denotes the maximum pixel value of the star spot, $m \times n$ denotes the image size, I denotes the original image before noise addition, and K denotes the noisy image to be evaluated.

Evaluation metrics: The evaluation metrics include peak signal-to-noise ratio, centroiding-error standard deviation, and noise standard deviation (σ_n) [35]. The centroiding-error standard deviation is defined as the standard deviation of the deviations between the centroid obtained from the denoised image and the window center, in pixels. The “Local-window Centroiding Method” is used for centroid extraction. The noise standard deviation is calculated as follows, where a smaller value indicates a smoother denoised image:

$$\sigma_n = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} (K_i - I_i)^2}$$

where N denotes the total number of pixels, I denotes the original image before noise addition, and K denotes the noisy image to be evaluated.

4.2. Experimental Results and Comparison with Multi-Frame Stacking (Mean) and the Baseline Model

Each star spot in the test set is processed using multi-frame stacking, the baseline model, and the improved baseline model, namely deformable convolution combined with the frequency-domain loss constraint. Table 1 summarizes the statistical results for the signal-to-noise ratio of the window images and the centroiding-error standard deviation.

Table 1. Denoising results for the window-image signal-to-noise ratio and centroiding-error standard deviation of 30 star spots.

Method	PSNR range (dB)	Average PSNR (dB)	Centroiding error range	Centroiding-error std.	Average noise std. (dB)
Multi-frame stacking	[9.75,17.42]	14.08	[0.004,1.402]	0.489	0.61
Baseline model	[9.58,27.12]	17.55	[0.058,1.192]	0.421	0.29
Improved baseline model	[8.90,37.36]	25.80	[0.013,0.451]	0.119	0.21

The experimental comparison shows that the improved baseline model achieves much higher signal-to-noise ratio improvement and higher centroiding accuracy than the other two methods. This result verifies that the two proposed improvements further enhance the denoising capability of the baseline model. The results also support the theoretical analysis in Section 3 regarding the advantages of the proposed deep learning method over multi-frame stacking. The noise standard deviation indicates that multi-frame stacking leaves more residual noise after denoising and produces poorer background smoothness.

4.3. Ablation Study

Ablation experiments are conducted separately on the two proposed improvements, and the results are shown in Tables 2 and 3. When both improvements are ablated, the processing results are the same as those of the baseline model in Table 1.

Table 2. Denoising results for 30 star spots when deformable convolution is ablated.

Average PSNR (dB)	Centroiding-error std.	Average noise std. (dB)
18.12	0.298	0.26

Table 3. Denoising results for 30 star spots when the frequency-domain loss constraint is ablated.

Average PSNR (dB)	Centroiding-error std.	Average noise std. (dB)
18.83	0.310	0.22

Tables 2 and 3 show that ablating deformable convolution has a greater impact on the signal-to-noise ratio and noise standard deviation. This result is consistent with the theoretical analysis that deformable convolution effectively suppresses background noise interference. By contrast, ablating the frequency-domain loss constraint has a greater impact on centroiding accuracy. This result is also consistent with the analysis that this loss encourages the model to focus on high-frequency image information and is therefore conducive to restoring the PSF morphology of star spots. Together, the two proposed improvements are complementary, and their combined use enhances the denoising capability of the baseline model.

For one star spot in the test set with relatively weak energy, restoration is difficult; this subset contains 100 groups \times 8 noisy frames. Figures 3 and 4 show the statistical curves of the window-image signal-to-noise ratio and centroiding-error standard deviation after processing by the five methods. In these figures, the curves labeled Multiframe Method, DFConv+Freq, DFConv, Freq, and Org denote multi-frame stacking, the improved baseline, the baseline with only deformable convolution, the baseline with only the frequency-domain loss constraint, and the baseline model, respectively. The corresponding data are listed in Table 4.

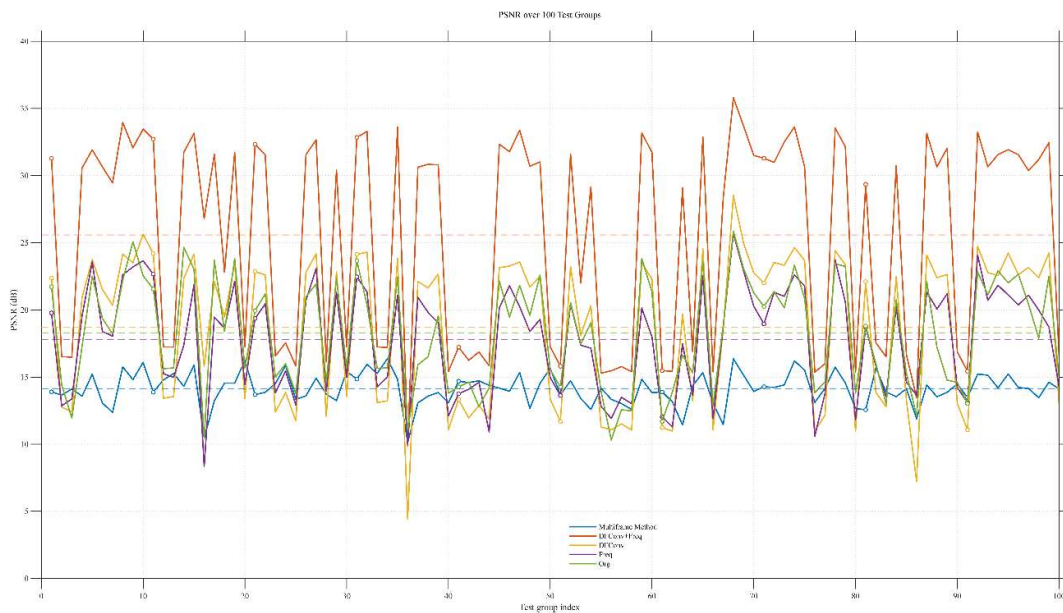


Figure 3. Statistical curve of the window-image PSNR for one star spot after processing by the five methods.

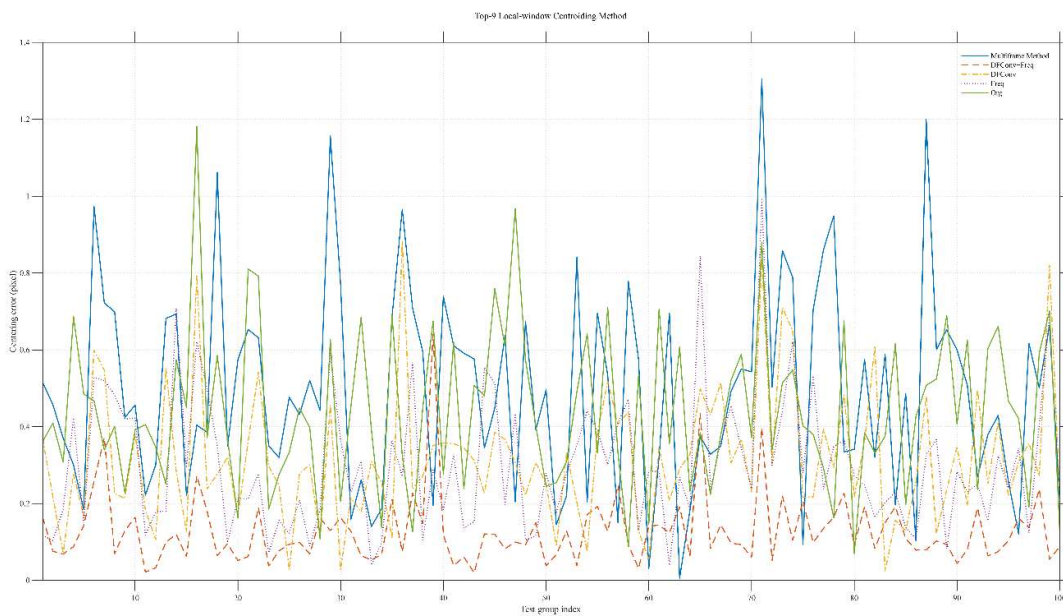


Figure 4. Statistical curve of the window-image centroiding-error standard deviation for one star spot after processing by the five methods.

Table 4. Denoising results for the window-image signal-to-noise ratio and centroiding-error standard deviation of one star spot.

Method	PSNR range (dB)	Average PSNR (dB)	Centroiding error range	Centroiding-error std.	Average noise std. (dB)
Multi-frame stacking	[10.13,16.38]	14.16	[0.005,1.305]	0.494	0.63
Baseline model	[10.31,25.86]	17.80	[0.069,1.181]	0.441	0.31
Improved baseline model	[9.87,35.81]	25.58	[0.019,0.435]	0.128	0.22
Baseline with deformable convolution only	[4.44,28.52]	18.70	[0.028,0.882]	0.318	0.23
Baseline with frequency-domain loss only	[8.34,24.08]	18.30	[0.039,0.994]	0.300	0.27

Figure 5 shows a group of eight noisy test images for this star spot. The star spot is almost submerged by noise, making centroid extraction impossible.

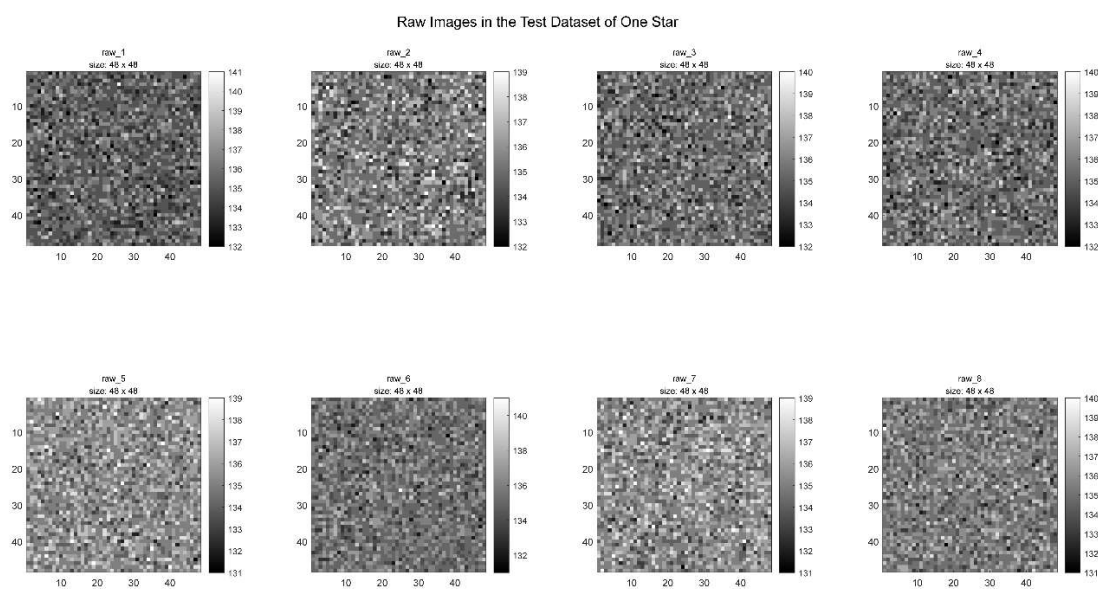


Figure 5. A group of eight noisy images for one star spot.

Figure 6 shows the denoised images obtained after processing the group of noisy images above by the five methods, together with the original noise-free star-spot image.

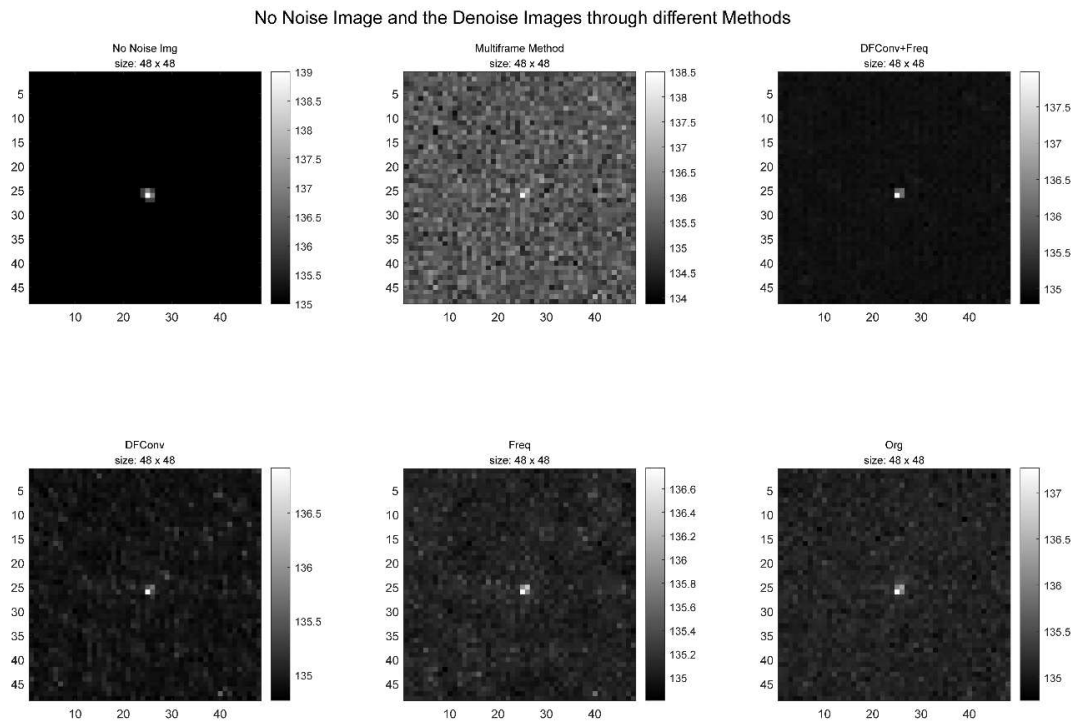


Figure 6. Comparison of the original noise-free image of one star spot and the results obtained by processing the group of noisy images above using five methods.

As shown in Figure 6, the proposed method achieves a substantially higher signal-to-noise ratio than the other methods after denoising, and the star spot is restored more accurately with clearer edges. By contrast, multi-frame stacking yields a markedly lower signal-to-noise ratio and poorer background smoothness.

5. Discussion and Future Work

This study proposes a deep learning method based on frame-by-frame spatial deformable convolution and complex-domain frequency-domain constraints for star-sensor window images degraded by detector readout noise, which follows a Gaussian distribution, spatially random stray-light interference, which is independent between frames and has an approximately zero mean. With ASTERIS as the baseline, the proposed method retains the advantages of the baseline model while introducing two improvements: frame-by-frame spatial deformable convolution in the decoder upsampling stage and a complex-domain frequency-domain loss constraint. These improvements compensate for subpixel jitter of star spots, enhance denoising capability, preserve the high-frequency morphology of star spots, and reduce the number of model parameters by approximately one quarter. In the experiments, the proposed method achieves substantially better denoising performance than multi-frame stacking and the baseline model. The proposed method improves PSNR by approximately 60-fold, and the centroiding error reaches approximately 0.1 pixels, which is smaller than the artificially added centroid-jitter range in the test set, approximately 0.2 pixels.

For interference with a nonzero mean, any method that relies only on temporal independence, including multi-frame stacking and self-supervised learning, cannot remove the mean component. Although the proposed method theoretically cannot eliminate this component, its spatially adaptive filtering capability enables better suppression of background spatial nonuniformity than multi-frame stacking and can reduce star-spot centroiding errors.

Future work will introduce a background estimation branch into the model, allowing the static background to be subtracted at the input end through a learnable background-separation network. This extension is expected to improve the adaptability of the proposed method to complex on-orbit environments and diverse remote sensing tasks. Lightweight network design and dynamically adaptive frequency-domain loss weighting are also important directions for future research. Real on-orbit star image data will be acquired from various on-orbit missions to construct more diverse test benchmarks, to verify and improve the generalization capability against unknown noise distributions and stray light patterns of the model, thereby laying a solid foundation for eventual engineering deployment.

Author Contributions: J.G.: Conceptualization, Methodology, Validation, Visualization, Investigation, Writing original draft, Review & editing. H.Y.: Data curation, Resources, Methodology, Investigation, Writing original draft, Review & editing. Y.W.: Conceptualization, Resources, Investigation, Review & editing. X.L.: Conceptualization, Review & editing. R.C.: Data curation, Resources, Investigation, Review & editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by China Civil Aerospace Space Debris Special Project, grant number D020101.

Data Availability Statement: The simulated windowed image datasets generated and analyzed in this study have not been publicly archived at present but are available from the corresponding author upon reasonable request. The source code is available from the corresponding author (Y. Wu, 20033110168@stu.xidian.edu.cn) for reproducing the results or extending the reported research. All underlying data required to understand, evaluate, and build upon the reported findings are provided in this article.

Acknowledgments: The authors gratefully acknowledge Xiaogang Dong, Cong Tian, and Ruiming Zhong for their valuable guidance and suggestions throughout this research.

Conflicts of Interest: The authors declare there are no conflicts of interest for this manuscript.

References

1. Liebe, C. Star trackers for attitude determination. *IEEE Aerosp. Electron. Syst. Mag.* **1995**, *10*, 10–16. <https://doi.org/10.1109/62.387971>.
2. Eisenman, A.R.; Liebe, C.C.; Joergensen, J.L. New generation of autonomous star trackers. In Proceedings of the SPIE's 1997 International Symposium on Optical Science, Engineering and Instrumentation, San Diego, CA, USA, 27 July-1 August 1997; Volume 3219, pp. 210-224. <https://doi.org/10.1117/12.298121>.
3. Wang, H.; Jiang, J.; Zhang, G. A high-precision star tracker for spacecraft attitude determination. *IEEE Trans. Aerosp. Electron. Syst.* **2019**, *55*, 2475-2485. <https://doi.org/10.1109/taes.2019.2896506>.
4. Mehta, D.S.; Parsa, K. A star pattern recognition technique for star trackers. *Adv. Space Res.* **2019**, *64*, 2055-2068. <https://doi.org/10.1016/j.asr.2019.07.042>
5. Rufino, G.; Accardo, D. Enhancement of the centroiding algorithm for star tracker measure refinement. *Acta Astronaut.* **2003**, *53*, 135-147. [https://doi.org/10.1016/s0094-5765\(02\)00199-6](https://doi.org/10.1016/s0094-5765(02)00199-6).
6. Marcelino, G.M.; Schulz, V.H.; Seman, L.O.; Bezerra, E.A. Centroid determination hardware algorithm for star trackers. *Int. J. Sensor Networks* **2020**, *32*, 1-14. <https://doi.org/10.1504/ijnsnet.2020.104458>.
7. Hou, Y.; Zhao, R.; Ma, Y.; He, L.; Zhu, Z. An on-orbit correction method for high dynamic APS star tracker based on adaptive filtering. *Acta Photonica Sin.* **2021**, *50*, 155. <https://doi.org/10.3788/gzxb20215002.0212004>.
8. Tomasi, C.; Manduchi, R. Bilateral filtering for gray and color images. In Proceedings of the IEEE International Conference on Computer Vision, Bombay, India, 7 June 1998; pp. 839-846. <https://doi.org/10.1109/ICCV.1998.710815>.
9. Donoho, D. De-noising by soft-thresholding. *IEEE Trans. Inf. Theory* **1995**, *41*, 613-627. <https://doi.org/10.1109/18.382009>.
10. Portilla, J.; Strela, V.; Wainwright, M.; Simoncelli, E. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Trans. Image Process.* **2003**, *12*, 1338-1351. <https://doi.org/10.1109/tip.2003.818640>.

11. Buades, A.; Coll, B.; Morel, J.M. A review of image denoising algorithms, with a new one. *Multiscale Model. Simul.* **2005**, *4*, 490-530. <https://doi.org/10.1137/040616024>.
12. Lebrun, M.; Buades, A.; Morel, J.M. A nonlocal Bayesian image denoising algorithm. *SIAM J. Imaging Sci.* **2013**, *6*, 1665-1688. <https://doi.org/10.1137/120874989>.
13. Gu, S.; Zhang, L.; Zuo, W.; Feng, X. Weighted nuclear norm minimization with application to image denoising. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, Ohio, 24-27 June 2014; pp. 2862-2869. <https://doi.org/10.1109/CVPR.2014.366>.
14. Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process.* **2007**, *16*, 2080-2095. <https://doi.org/10.1109/tip.2007.901238>.
15. Maggioni, M.; Katkovnik, V.; Egiazarian, K.; Foi, A. Nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE Trans. Image Process.* **2013**, *22*, 119-133. <https://doi.org/10.1109/TIP.2012.2210725>.
16. Wang, X.; Wang, X. Multiple targets sparse matching for binocular vision positioning system with large field of view. *Infrared Laser Eng.* **2018**, *47*, 0726001. <https://doi.org/10.3788/irla201847.0726001>.
17. Zhao, H.; Lembeck, M.F.; Zhuang, A.; Shah, R.; Wei, J. Real-time convolutional-neural-network-based star detection and centroiding method for CubeSat star tracker. *IEEE Trans. Aerosp. Electron. Syst.* **2025**, *61*, 8172-8184. <https://doi.org/10.1109/taes.2025.3542744>.
18. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5-9 October 2015; pp. 234-241. https://doi.org/10.1007/978-3-319-24574-4_28.
19. Içek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger, O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Athens, Greece, 17-21 October 2016; pp. 424-432. https://doi.org/10.1007/978-3-319-46723-8_49.
20. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M. Restormer: efficient transformer for high-resolution image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18-24 June 2022; pp. 5728-5739. <https://doi.org/10.1109/CVPR52688.2022.01167>.
21. Guo, Y.; Zhang, H.; Li, M.; Yu, F.; Wu, Y.; Hao, Y.; Huang, S.; Liang, Y.; Lin, X.; Li, X.; et al. Deeper detection limits in astronomical imaging using self-supervised spatiotemporal denoising. *Science* **2026**, *392*, eady9404. <https://doi.org/10.1126/science.ady9404>.
22. Fang, B.-L.; Wang, J.-G.; Feng, G.-B. Calculation of spot centroid based on physical informed neural networks. *Acta Phys. Sin.* **2022**, *71*, 200601. <https://doi.org/10.7498/aps.71.20220670>.
23. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22 October 2017; pp. 764-773. <https://doi.org/10.1109/ICCV.2017.89>.
24. Peng, Y.; Xie, W.; Li, J.; Wu, S.; Wang, Z.; Hu, B.; Yao, J. LDG: Lightweight Deformable 3D Gaussians for Single-View Dynamic Scene Reconstruction. In Proceedings of the 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hyderabad, India, 6-11 April 2025; pp. 1-5. <https://doi.org/10.1109/ICASSP49660.2025.10890036>.
25. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600-612. <https://doi.org/10.1109/tip.2003.819861>.
26. Čelić, J.; Lang, R.G.; Steinmassl, S.; Hinton, J.; Funk, S. A novel approach to optimizing the image cleaning performance of Imaging Atmospheric Cherenkov Telescopes. *Astron. Astrophys.* **2025**, *699*, A96. <https://doi.org/10.1051/0004-6361/202554824>.
27. Gong, X.; Li, T.; Wang, R.; Hu, S.; Yuan, S. Beyond the Remote Sensing Ecological Index: A Comprehensive Ecological Quality Evaluation Using a Deep-Learning-Based Remote Sensing Ecological Index. *Remote Sens.* **2026**, *17*, 558. <https://doi.org/10.3390/rs17030558>.

28. Zhu, C.; Deng, S.; Song, X.; Li, Y.; Wang, Q. Mamba Collaborative Implicit Neural Representation for Hyperspectral and Multispectral Remote Sensing Image Fusion. *IEEE Trans. Geosci. Remote. Sens.* **2025**, *63*, 1-15. <https://doi.org/10.1109/tgrs.2025.3537638>.
29. Shah, Z.H.; Müller, M.; Hübner, W.; Ortkrass, H.; Hammer, B.; Huser, T.; Schenck, W. Image restoration in frequency space using complex-valued CNNs. *Front. Artif. Intell.* **2024**, *7*, 1353873. <https://doi.org/10.3389/frai.2024.1353873>.
30. Lehtinen, J.; Munkberg, J.; Hasselgren, J.; Laine, S.; Karras, T.; Aittala, M.; Aila, T. Noise2Noise: learning image restoration without clean data. In Proceedings of the International Conference on Machine Learning, Stockholm, Sweden, 10-15 July 2018; pp. 2965-2974.
31. Krull, A.; Buchholz, T.-O.; Jug, F. Noise2Void: learning denoising from single noisy images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16-20 June 2019; pp. 2129-2137. <https://doi.org/10.1109/CVPR.2019.00223>.
32. Batson, J.; Royer, L. Noise2Self: blind denoising by self-supervision. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9-15 June 2019; pp. 524-533. <https://doi.org/10.5281/zenodo.11343973>.
33. Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. *IEEE Trans. Image Process.* **2017**, *26*, 3142-3155. <https://doi.org/10.1109/tip.2017.2662206>.
34. Zhang, K.; Zuo, W.; Zhang, L. FFDNet: toward a fast and flexible solution for CNN-based image denoising. *IEEE Trans. Image Process.* **2018**, *27*, 4608-4622. <https://doi.org/10.1109/tip.2018.2839891>.
35. Liu, J.G.; Li, J.; Hao, Z.H. Study on detection sensitivity of APS star tracker. *Opt. Precis. Eng.* **2006**, *14*, 553-557. <https://doi.org/10.3321/j.issn:1004-924X.2006.04.005>.

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