Quantitative Evaluation of the Applicability and Stability of deepCR for Cosmic Ray Identification in CSST Survey Data Processing: A Postprint

Authors: Lin Zhun, Huang Weirong, Wang Feng, Deng Hui, Mei Ying, Weirong Huang

Date: 2023-10-18T00:00:00+00:00

Abstract

The deepCR cosmic ray identification method is an effective approach for cosmic ray removal from Hubble Space Telescope (HST) data; however, whether this method can satisfy the requirements of the China Space Station Telescope (CSST) has consistently lacked scientific quantitative analysis. Utilizing real observational data from the Hubble Telescope, we conducted an in-depth analysis of the deepCR cosmic ray identification method and performed empirical studies on its stability and usability. The results demonstrate that deepCR exhibits high sensitivity for cosmic ray identification in sky background regions, but its sensitivity decreases progressively closer to the center of stellar images. We analyzed the relationship between cosmic ray density and photometric precision, finding that when cosmic ray density reaches 9%, nearly 100% of stars are contaminated by cosmic rays; when cosmic ray density reaches 14%, anomalous photometric results ranging from 20% to 50% occur for stars with different profile areas. Experimental results indicate that the deepCR cosmic ray identification model possesses relatively good stability, allowing for extended application after a single training session. Nevertheless, it still faces a series of challenges in high-precision photometry applications and other scenarios, necessitating targeted solutions.

Full Text

A Quantitative Assessment of the Usability and Stability of the deepCR Cosmic Ray Identification Method for Survey Data Processing

Lin Zhun, Huang Weirong, Wang Feng*, Deng Hui, Mei Ying

Center for Astrophysics, School of Physics and Materials Science, Guangzhou University, Guangzhou 510006, Guangdong, China

Abstract

The deepCR cosmic ray identification method can effectively remove cosmic rays from Hubble Space Telescope (HST) data. However, whether this method can meet the requirements of the China Space Station Telescope (CSST) has lacked scientific quantitative analysis. This paper analyzes the deepCR cosmic ray method in depth using real observational data from the Hubble Telescope and conducts an empirical study of its stability and usability. Results show that deepCR performs well in identifying cosmic rays in sky background regions, but its sensitivity decreases as it approaches the center of stars. We analyze the correlation between cosmic ray density and photometric accuracy, demonstrating that when cosmic ray density reaches 9%, almost all stars (100%) are contaminated by cosmic rays; when density reaches 14%, there are abnormal photometric results (20%-50%) for stars with different contour areas. The analysis shows that the deepCR cosmic ray identification model is relatively stable and can be applied over a longer period after one training session. However, it still faces a series of problems in application scenarios such as high-precision photometry, which need to be addressed in the near future.

Keywords: CSST; cosmic ray; deepCR; quantitative assessment

1 Introduction

The deepCR cosmic ray identification method can effectively remove cosmic rays from Hubble Space Telescope (HST) images, but whether this method can meet the requirements of the China Space Station Telescope (CSST) has always lacked scientific quantitative analysis. Using real HST observational data, this study conducts an in-depth analysis of the cosmic ray identification method and empirically investigates its stability and usability. Results show that the method has high sensitivity for identifying cosmic rays in sky background regions, but its sensitivity decreases as it approaches the center of stellar images.

Cosmic rays are high-energy charged particles from space [1-2] that can penetrate detectors and leave energy traces. They frequently interfere with normal astronomical observations, particularly in astronomical image processing. To improve data quality and reliability and ensure scientific analysis accuracy, cosmic rays must be correctly identified and removed [3].

The most traditional approach involves taking multiple exposures of the same sky region, aligning these exposures, and calculating a median image to obtain a cosmic-ray-free image. Each exposure is then compared with the median image to identify cosmic rays [4]. This method works well but is not suitable for single-exposure images. For single-exposure astronomical images, various techniques have been proposed. One approach uses a delta function based on the point spread function to construct a spatial filter for convolution with the original image, identifying cosmic rays by setting thresholds based on the noise characteristics of the filtered image. This program requires multiple iterations to better identify cosmic rays composed of multiple pixels, is relatively time-consuming, and requires well-sampled data with full width at half maximum greater than or equal to two pixels. Another method exploits the sharp edges and asymmetry of cosmic rays, subsampling the original image, convolving it with a Laplacian operator, and obtaining a Laplacian image after restoring to the original resolution. A fine structure is constructed using median filtering, and cosmic rays are identified by setting contrast thresholds between the Laplacian image and noise model. While this method achieves good detection results and works well with different image data, it requires long processing times for large images and manual contrast adjustment. A histogram-based statistical method [7-8] divides the image into sub-images, analyzes the histogram distribution of multiple sub-images, and identifies pixels deviating from the distribution as cosmic rays. This method is suitable for processing spectral image data and runs quickly, but its cosmic ray identification effect is not as good as reference [6] and requires continuous iteration until no new cosmic rays are identified.

With the development of deep learning technology, which has attracted great attention due to its high accuracy and efficiency, researchers have begun applying deep learning methods to cosmic ray identification. The deepCR framework includes two independent deep neural networks for marking cosmic rays and restoring images after marking. It has demonstrated higher recall rates and faster processing speeds than Laplacian edge detection algorithms on HST ACS/WFC (F606W) data [10].

The China Space Station Telescope (CSST) [11] is a major scientific project under China's manned space program, primarily tasked with conducting large-scale sky surveys. In survey observations, cosmic ray interference is a serious problem. The CSST's main survey camera consists of multiple detector mosaics. Based on estimates from HST data at similar orbital altitudes, each detector has approximately 2.34 million pixels affected by cosmic rays during a 150-second exposure. Cosmic rays corrupt the readings of these pixels, so their positions must be identified to avoid affecting scientific measurements. In survey mode, since each detector only covers its corresponding sky region once, traditional multi-exposure combination methods cannot be used, and cosmic ray removal must be achieved using single-exposure images.

While deepCR has achieved good results, its real-world effectiveness still lacks quantitative analysis. Whether this method can be applied to future CSST data processing remains a question worthy of study. This paper systematically quantifies the practical application stability and usability of deepCR based on HST observational data.

2 deepCR Cosmic Ray Identification Method

The deepCR cosmic ray identification model consists of two independent deep neural networks: deepCR-mask and deepCR-inpaint. The deepCR-mask network predicts the probability that each pixel is affected by cosmic rays, converting the input image into a binary map where 0 represents non-cosmic-ray pixels and 1 represents cosmic-ray pixels. The deepCR-inpaint network predicts pixel values without cosmic ray interference to restore the image to its uncontaminated state. Both networks are based on the deepCR-mask architecture. The model's training data includes images affected by cosmic rays and corresponding binary label maps, with accurate binary labels created by comparing each exposure with a median image.

3 Test Data Preparation

Each test dataset group consists of multiple exposure scientific images from HST ACS/WFC (F606W) with the same field of view and calibration. Using AstroDrizzle [12], we automatically aligned multiple observation images through sky projection to obtain multiple mapped median images, which were then mapped back to each original observation image. Cosmic rays in the original observations were replaced with pixel values from the mapped median images. We used the data quality array from the original observations as a mask for bad pixels. A saturation mask was created for pixels with values exceeding 70,000 electrons to ensure these anomalous pixels did not participate in subsequent model evaluation. Using a 7$×$7 pixel block, we obtained clean images with cosmic rays removed from each original observation, which served as reference images for subsequent quantitative evaluation experiments.

4 Quantitative Evaluation Based on Deep Learning Metrics

Since cosmic rays account for a small proportion of the entire image, leading to imbalanced positive and negative samples, precision and recall are the most important evaluation metrics. TP represents the number of cosmic-ray pixels correctly identified as cosmic rays, FP represents the number of non-cosmic-ray pixels incorrectly identified as cosmic rays, and FN represents the number of cosmic-ray pixels incorrectly identified as non-cosmic-ray.

5 Experimental Results

5.1 Evaluation of Cosmic Ray Identification Across Full Image Regions

We first evaluated cosmic ray identification across the entire image region for all test data. The results show high recall and precision rates. However, since most cosmic rays fall on the sky background, this may lead to overestimated performance. To obtain more credible precision and recall, we further evaluated star regions.

We calculated the RMS of the clean image background and used thresholds of 3, 5, 10, 20, 40, 80, and 160 times the RMS to extract star connected regions. These connected regions served as masks to evaluate cosmic ray identification performance at different distances from star centers. shows the variation of deepCR performance with distance from star centers.

Table 1 Variation of deepCR Performance with Distance from the Centre of Stars

Cosmic Ray Position	Recall (%)	Precision (%)
Full graph	88.8	95.0
3 RMS	85.5	87.6
5 RMS	83.5	87.7
10 RMS	79.8	88.3
20 RMS	74.2	86.4
40 RMS	66.9	81.1
80 RMS	59.7	74.6
160 RMS	56.2	72.8

The results show that deepCR's sensitivity and accuracy for identifying cosmic rays on stars decrease significantly compared to the full image region, and the closer to the star center, the lower the recall and precision. The deepCR method easily misses cosmic rays in the central regions of stars.

5.2 Quantitative Evaluation Based on Photometry

Although precision and recall are important metrics for evaluating model performance, they cannot reflect the model's effectiveness in practical application scenarios. We further analyzed deepCR's cosmic ray removal effect based on photometric results. The photometry process is shown in [Figure 2: see original paper].

The source extraction criteria were: (1) single pixel value higher than 3 times background noise; (2) pixel count greater than 16; (3) aspect ratio between 0.8 and 1.2 to avoid unreliable photometric results; (4) only point sources were analyzed. The extraction range was set to positions more than 16 pixels from the image edges. Stars in each group were matched, with stars having coordinate distances less than 0.1 considered the same star. To obtain more accurate cosmic ray removal effects, the same photometric center coordinates and aperture were used for each identical star across the same dataset.

We defined stars with flux differences greater than 3 times the flux noise between original and clean images as cosmic-ray-contaminated stars. Statistical analysis of the test data photometric results shows that the proportion of anomalous stars in each dataset ranges from 2% to 5%. Among stars affected by cosmic rays, deepCR's photometric results show…

[Figure 3: see original paper] shows the photometric results. The black line represents the ratio of photometric results between the original image and clean image, while the red line represents the ratio between deepCR-processed images and clean images. If the flux ratio is closer to 1, deepCR's effect is closer to the reference result. While the overall removal effect is good, some stars are still incorrectly removed or have cosmic rays missed, leading to anomalous photometric results.

We examined images of stars with flux ratios significantly less than 1. Each row shows: the first sub-image is the clean image; the second is the original image; the third is the difference between clean and original, showing real cosmic ray positions; the fourth is the difference between deepCR-processed and clean images, showing misidentified or missed cosmic rays. In examples where the flux ratio is significantly less than 1, we found that deepCR may incorrectly identify small-profile entire stars as cosmic rays, significantly affecting photometric accuracy. In other cases, when cosmic rays fall in star center regions, deepCR often fails to identify them. When cosmic rays fall on star edges, identification is good. When cosmic rays on stars are few or weak, they don't significantly affect photometric results.

5.3 Impact of Cosmic Ray Density

Cosmic ray density is a crucial factor affecting photometric accuracy. [Figure 6: see original paper] shows the impact of cosmic ray density on photometric accuracy, where the x-axis represents the cosmic ray proportion in the full image, and the y-axis represents the proportion of stars contaminated by cosmic rays (with contamination defined as original image photometric results outside 3 times the flux error range of clean image results) and the proportion of anomalous photometric results (defined as photometric results outside 3 times the flux error range of clean image results). The red and blue lines represent linear fitting results for stars of different sizes, with Pearson correlation coefficients $r = 0.86$, $p = 7\times10^{-200}$ for 16-50 pixel stars, and $r = 0.95$, $p = 2\times10^{-200}$ for 200-400 pixel stars.

The results show that when cosmic ray density reaches 9%, almost all stars (100%) are contaminated by cosmic rays. When density reaches 14%, 20%-50% of stars show anomalous photometric results. Since cosmic ray density increases with exposure time, we recommend that CSST use shorter exposure times during surveys to control cosmic ray numbers.

6 Model Stability Analysis

The quality and stability of astronomical images are often difficult to guarantee, so a good deep learning model should have high stability. Previous research lacks model stability studies. This paper analyzes model stability through deep learning evaluation metrics and photometric accuracy.

Using HST data from 2010-2020, the F-score based on cosmic rays on stars shows no significant change over time, with values of (85$±2.4)±$2.1)%, indicating the model is relatively stable and can be applied over long periods after one training session.

[Figure 8: see original paper] shows photometric accuracy of data from different observation times. The x-axis spans dates from 2010-2020, the y-axis represents the flux ratio of all stars in each dataset, the red line is the mean, and the green line represents the RMS of the mean. To obtain more accurate RMS, the top and bottom 5% of data were removed before calculation. The RMS of each dataset is basically within the error range, showing that deepCR photometric accuracy has no significant relationship with detector working age.

7 Conclusion

This paper systematically and quantitatively analyzed the deepCR method using HST data from 2010-2020. Compared to the full image region, the sensitivity for identifying cosmic rays on stars decreases closer to the star center, dropping from 88.8% to 56.2%. We studied the relationship between cosmic ray density and the proportion of contaminated stars. When cosmic ray pixel proportion reaches 9%, almost all stars (100%) are contaminated; when density reaches 14%, 20%-50% of stars with different contour areas show anomalous photometric results. The deepCR model is relatively stable, with no significant changes in F-score or photometric accuracy over time. However, in high-precision photometry applications, it still faces challenges that require targeted solutions. We recommend that CSST use shorter exposure times during surveys to control cosmic ray density.

References

[1] LU B Q. Study of cosmic rays and dark matter detection. Nanjing: Nanjing University. [2] DEUSTUA S E, TANCREDI G, MILES N D. I. detection and characterization of cosmic rays. The Astrophysical Journal. 86-113. [3] FENG H X, CHEN J J, DENG J R, et al. Cosmic-ray muons extraction technology in CCD image. Astronomical Research & Technology. 201-209. [4] FRANKLIN B E, NEUSCHAEFER L W. Removing cosmic-ray hits from multi. Publications of the Astronomical Society of the Pacific. 703-710. [5] RHOADS J E. Cosmic-ray rejection by linear filtering of single images. Publications of the Astronomical Society of the Pacific. 796-806. [6] VAN DOKKUM P G. Cosmic-ray rejection by laplacian edge detection. Publications of the Astronomical Society of the Pacific. 1420-1429. [7] LIU T T, PENG Q Y. Comparison of CCD image cosmic-ray removal algorithms. Astronomical Research & Technology. 140-149. [8] FARAGE C L. Publications of the Astronomical Society of Australia. 249-256. [9] PIMBBLET K A. Evaluation of cosmic ray rejection algorithms on single-shot exposures. Publications of the Astronomical Society of the Pacific. 148-153. [10] ZHANG K M, BLOOM J S. DeepCR: cosmic ray rejection with deep learning. Astrophysical Journal. 24-32. [11] ZHAN H. The

wide-field multiband imaging and slitless spectroscopy survey to be carried out by the Survey Space Telescope of China Manned Space Program. Chinese Science Bulletin. 1290-1298. [12] HACK W J, DENCHEVA N, BARBARAY K. SEP: source extractor as a library. Journal of Open Source Software. 393-404. [13] BERTIN E, ARNOUTS S. SExtractor: software for source extraction. Astronomy and Astrophysics Supplement Series. 393-404. [14] AstroDrizzle: more than a new MultiDrizzle. Proceedings of the American Astronomical Society Meeting. 2012.

Note: Figure translations are in progress. See original paper for figures.