nature astronomy

Article

https://doi.org/10.1038/s41550-025-02484-z

Astronomical image denoising by self-supervised deep learning and restoration processes

Received: 13 January 2023

Accepted: 16 January 2025

Tie Liu[1,2,3,4] ✉, Yuhui Quan[5,6] ✉, Yingna Su[3,4] ✉, Yang Guo[1,2] ✉, Shu Liu[7], Haisheng Ji[3,4], Qi Hao[1,2], Yulong Gao[1,2], Yuxia Liu[8,9], Yikang Wang[1,2], Wenqing Sun[10] & Mingde Ding[1,2]

Image denoising based on deep learning has undergone significant advances in recent years. However, existing deep learning methods lack quantitative control of the deviation or error of denoised images. The neural network Self2Self was designed to denoise single images. It is trained on single images and then denoises them, although training is costly. In this work, we explore training Self2Self on an astronomical image and denoising other images of the same kind, a process that is also suitable for quickly denoising immense images in astronomy. To address the deviation issue, the abnormal pixels whose deviation exceeds a predefined threshold are restored to their initial values. The noise reduction is due to training, denoising and restoring and is, therefore, named the TDR method. With the TDR method, the noise level of solar magnetograms improved from about 8 to 2 G. Furthermore, the TDR method was applied to galaxy images from the Hubble Space Telescope, making weak galaxy structures much clearer. This capability of enhancing weak signals makes the TDR method applicable to various disciplines.

Astronomy has entered the big data era due to advances in ground-based and space-borne telescopes. Immense digital images are being collected, so that denoising these images is becoming increasingly important. Taking solar physics as an example, the Solar Dynamics Observatory[1] is a space-borne telescope that plays an important role in collecting data, with its higher-level derivatives exceeding 7 PB in size[2]. Additionally, ground-based telescopes, such as the Goode Solar Telescope[3] and the New Vacuum Solar Telescope[4], are also contributing notably to the wealth of solar images. In the field of galaxies and cosmology, the Hubble Space Telescope (HST) is renowned for providing high-quality images. Since its launch in 1990, the HST has conducted more than 1.5 million observations.

The aforementioned astronomical instruments can be simplified into two parts: a light-focusing system and a charge-coupled device (CCD) camera. All astronomical images obtained by these instruments inevitably contain noise, mainly from imperfections in these two parts. Moreover, additive white Gaussian noise, impulse noise (salt-and-pepper noise), quantization noise and Poisson noise have

---

[1]School of Astronomy and Space Science, Nanjing University, Nanjing, People's Republic of China. [2]Key Laboratory of Modern Astronomy and Astrophysics (Nanjing University), Ministry of Education, Nanjing, People's Republic of China. [3]Key Laboratory of Dark Matter and Space Astronomy, Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing, People's Republic of China. [4]School of Astronomy and Space Science, University of Science and Technology of China, Hefei, People's Republic of China. [5]School of Computer Science and Engineering, South China University of Technology, Guangzhou, People's Republic of China. [6]Pazhou Lab, Guangzhou, People's Republic of China. [7]Department of Rheumatology, Nanjing Drum Tower Hospital, The Affiliated Hospital of Nanjing University, Nanjing, People's Republic of China. [8]College of Engineering, Peking University, Beijing, People's Republic of China. [9]State Key Laboratory of Low-Carbon Smart Coal-Fired Power Generation and Ultra-Clean Emission, China Energy Science and Technology Research Institute Co., Ltd, Nanjing, People's Republic of China. [10]Faculty of Electrical Engineering and Computer Science, Ningbo University, Zhejiang, People's Republic of China. ✉e-mail: liu_tie_zimu@nju.edu.cn; csyhquan@scut.edu.cn; ynsu@pmo.ac.cn; guoyang@nju.edu.cn

frequently been studied[5]. Additive white Gaussian noise primarily originates from analogue circuitry. Quantization noise, impulse noise and Poisson noise mainly result from manufacturing defects, bit errors and insufficient photon counts during image acquisition and transmission[6].

Researchers have made continuous efforts to develop denoising methods. The most common approaches are spatial domain and transform domain methods. Spatial domain methods exploit the similarities between pixels or patches to perform denoising. Gaussian filters, for example, are widely used to reduce high-frequency signals[7]. Transform domain methods involve transforming images into the frequency domain and analysing the similarities and dissimilarities of the transformed coefficients. The Fourier transform is also a widely used technique. Moreover, based on different image features and task requirements, other denoising techniques, such as anisotropic diffusion methods, hybrid methods and morphological analysis, are also frequently used[6,8,9].

Recently, deep learning methods have emerged and shown notable denoising performance that surpasses classic methods. These methods can be categorized into two types: supervised learning (for example, refs. 10–13) and unsupervised learning (for example, refs. 14,15). The former learns the denoiser on pairs of clean and noisy images whereas the latter does not require clean images. Representative noise reduction works based on supervised learning include U-net[16] for image enhancement and denoising[17–19], generative adversarial networks[20] to recover features of degraded images and denoising[13,21,22]. One key factor that contributes to the performance of these supervised deep learning methods is the availability of large numbers of clean and noisy image pairs for training. However, collecting a substantial number of such pairs can be challenging and expensive.

Further research led to unsupervised learning methods such as the neural network (NN) Noise2Noise, which first showed the possibility of restoring images after being training on corrupted images alone[14]. Similar approaches, such as Noise2Void[23] and Noise2Self[24], have further explored learning denoisers on noisy images. The application[15] of Noise2Noise to images from the Swedish 1 m Solar Telescope has revealed its potential to recover weak signals. To achieve good performance, the aforementioned unsupervised learning methods typically require the NN to be trained on a sufficient number of external images that are closely related to the image being processed. However, collecting such images is challenging. To address this challenge, the NN Self2Self[25] was designed. Unlike other unsupervised methods, when Self2Self is trained on a single image, it achieves good denoising results, which makes Self2Self suitable for scenarios where limited images are available. Because the training data contain only noisy images, denoising by Self2Self is a self-supervised learning method and a form of unsupervised deep learning. However, unsupervised learning methods are susceptible to producing abnormal results due to the lack of supervision.

The noise reduction of astronomical images not only aims to make the images cleaner but also preserve the accuracy of each pixel. To ensure image accuracy, the change to each pixel in a denoised image should not exceed the initial noise level. Based on Self2Self, we proposed a general denoising method called the TDR method, which consists of three operations: (1) training (T) on a single image, (2) denoising (D) other images of the same kind and (3) restoring (R) abnormal pixels. The TDR method controls the accuracy of the denoised images and is suitable for denoising scientific data from various disciplines and fields. Astronomical images were used to verify the denoising of this method.

Results

Denoising and evaluation of HMI magnetograms

We collected ten line-of-sight (LOS) magnetograms from the helioseismic and magnetic imager (HMI)[26], which we used to train and test the NN Self2Self. First, the ten magnetograms were divided into the training group (containing one magnetogram) and the denoising group (containing the other nine magnetograms). Second, considering that each of the ten magnetograms could serve as the training image, we conducted ten different training and denoising experiments so that each of the ten magnetograms was used as the training image and the other nine as the denoising images. We found that the denoising results of the ten experiments were almost the same, which proves that denoising by Self2Self is robust.

Figure 1 compares the initial and denoised HMI LOS magnetograms. The denoised images come from the first training strategy, with the first image used as the training sample. The first row shows that the primary denoised magnetogram (titled 'Denoised HMI LOS') is much cleaner and is highly consistent with the initial magnetogram. The residual images represent the absolute error and show that the noise is randomly distributed. The amplitude of the noise is correlated with the signal strength. However, in regions with a strong magnetic field, the positive field is underestimated (red) and the negative field is overestimated (blue). Considering the absolute values, the strong positive and negative magnetic fields were both underestimated. The average value of the underestimation is about 10 G, which is on the same scale as the noise level[27]. The underestimations can have an impact on quantitative analyses, such as calculations of the magnetic flux and oscillations[28]. The last image in the first row of Fig. 1 shows that the signal-to-noise ratio (SNR) is high in regions with a strong magnetic field, whereas the SNR in other regions is relatively unsatisfactory, as the weak magnetic field is almost equivalent to the noise.

It is difficult to prevent underestimation by adjusting the NNs and training strategies because unsupervised learning methods are difficult to control. In a different train of thought, we propose a simpler solution. First, we find the abnormal pixels whose absolute value changes by more than a threshold R after denoising. Second, we restore the abnormal pixels to their initial values. The senior denoised images were obtained with this restoring operation and are titled '(Denoised+R10) HMI LOS' for $R = 10$ G and '(Denoised+R5) HMI LOS' for $R = 5$ G (second and third rows of Fig. 1). The deviation between the value of every pixel in the magnetogram before and after the noise reduction did not exceed 10 G for '(Denoised+R10) HMI LOS' and 5 G for '(Denoised+R5) HMI LOS'. The corresponding residual images show that the underestimation was effectively controlled. The proportions of the restored pixels are 13.98% for the '(Denoised+R10) HMI LOS' magnetogram and 44.45% for the '(Denoised+R5) HMI LOS' magnetogram. To ensure the accuracy of the denoised images, R should be less than the noise level. So $R = 5$ G is acceptable and suitable for the HMI LOS magnetograms.

We plot the pixel values of the initial and denoised HMI LOS magnetograms shown in Fig. 1 along the horizontal and vertical axes in Fig. 2. We fitted the scattered points with linear functions. The points in Fig. 2 show a high correlation, with linear coefficients (LCs) of 0.93, 0.99 and 1.00, respectively. Note that the first scatter plot also shows the underestimation, which is indicated by the slight deviation of the two ends from the red line in the first panel. By contrast, the second and third panels show that the underestimation is well reduced for the '(Denoised+R10) HMI LOS' and '(Denoised+R5) HMI LOS' magnetograms. The absolute error (AE), mean squared error (MSE), pixel-to-pixel Pearson correlation coefficient (PCC), LC, average flux of residual images and average absolute flux of residual images were calculated. The calculation methods are described in the Supplementary Information. We list the mean values and standard deviations of AE, MSE, PCC, LC, average flux of residual images and average absolute flux of residual images for the nine denoised magnetograms in the Supplementary Information. These quantitative evaluations with their small standard deviations indicate that the TDR method is robust and reliable.

Figure 3 shows histograms of pixel values in the range [−20 G, 20 G] for the magnetograms of Fig. 1 and the Gaussian fitting curves. The standard deviation of a Gaussian fit represents the noise level of

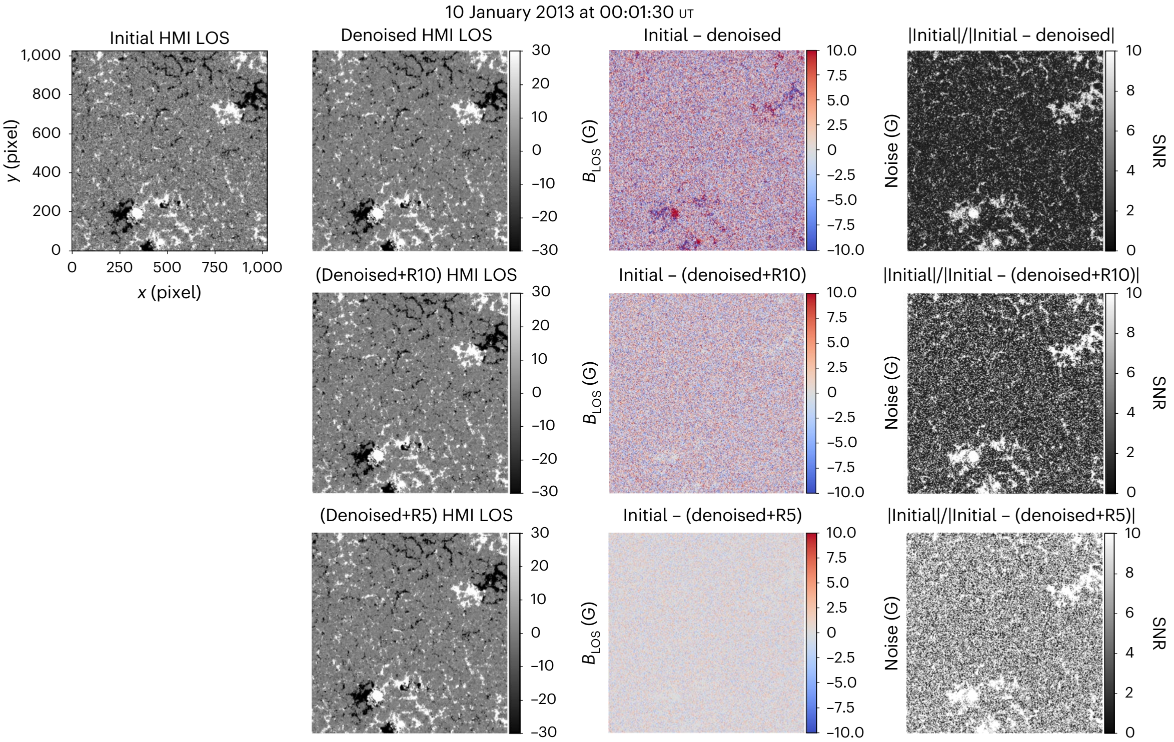

Fig. 1 | Comparison of the initial and denoised HMI LOS magnetograms. The initial magnetogram observed at 00:01:30 UT on 10 January 2013 is in the first column. The primary image denoised by Self2Self is titled 'Denoised HMI LOS'. Senior denoised images with restoring operations are titled '(Denoised+R10) HMI LOS' with $R = 10$ G and '(Denoised+R5) HMI LOS' with $R = 5$ G. The third column shows the residual images (difference maps obtained by subtracting the denoised image from the initial image). SNR images are displayed in the last column. Note that pixels where the initial value equals the denoised value are extremely rare, and these were excluded from the SNR images. B_{LOS} denotes the magnetic field intensity of the LOS magnetograms.

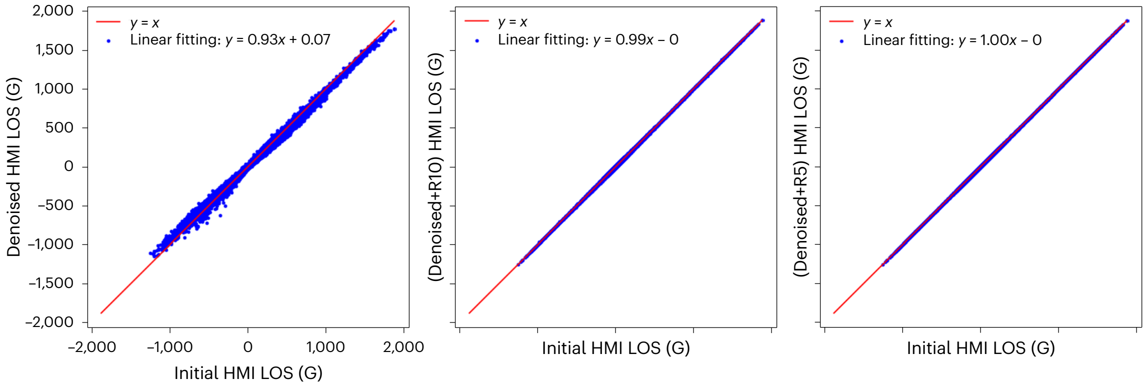

Fig. 2 | Scatter plots for the initial and denoised magnetograms. The pixel values of the initial and denoised magnetograms are scattered along the horizontal and vertical axes, respectively. A line ($y = x$) is displayed for reference. We fitted the dots with a linear function ($y = ax + b$). The parameters a and b are listed at the top left corner.

the magnetogram[27]. The noise levels were reduced from 8.09 to 2.07 G for the 'Denoised' magnetogram, to 2.04 G for the '(Denoised+R10)' magnetogram and to 2.12 G for the '(Denoised+R5)' magnetogram, which are better than the average noise level of 3.21 G achieved by a supervised learning method[29]. The mean value and standard deviation of the noise level for the nine magnetograms are listed in the Supplementary Information. We found that the restoring operation did not excessively increase the noise level although it was based on the premise of maintaining the accuracy of the pixel value. Moreover, we noticed that both sides of the red histograms became convex with $R = 10$ or 5 G. The convex parts were caused by the restored pixels, which are inside the Gaussian fitting of the initial magnetogram and outside the Gaussian fitting of the denoised magnetogram. This indicates that the restoring operation makes a portion of the weak magnetic signals between the initial noise level (~8 G) and the new noise level (~2 G) become distinguishable.

Figure 4 shows that, in the power spectrum, the difference between the initial and denoised magnetograms starts at a frequency of about 0.10 pix^{-1} for the 'Denoised' magnetogram, 0.28 pix^{-1} for the '(Denoised+R10)' magnetogram and 0.50 pix^{-1} for the '(Denoised+R5)' magnetogram. These results indicate that the denoising works best at scales smaller than 10 × 10 pixels, whereas in scales larger than 10 × 10 pixels, the denoised magnetograms are highly consistent with the initial magnetograms. However, magnetic structures smaller than

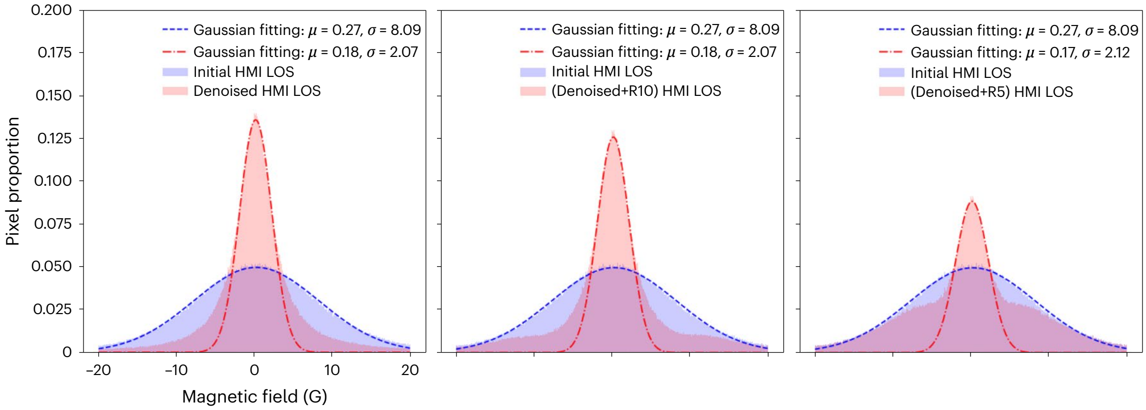

Fig. 3 | Gaussian fitting of the initial and denoised magnetograms. The blue and red regions show the distributions of the pixel values in the range [−20 G, 20 G]. The corresponding Gaussian fits are marked by the blue dashed curve and red dashed-dotted curve, respectively. The mean values μ and standard deviations σ are listed at the top.

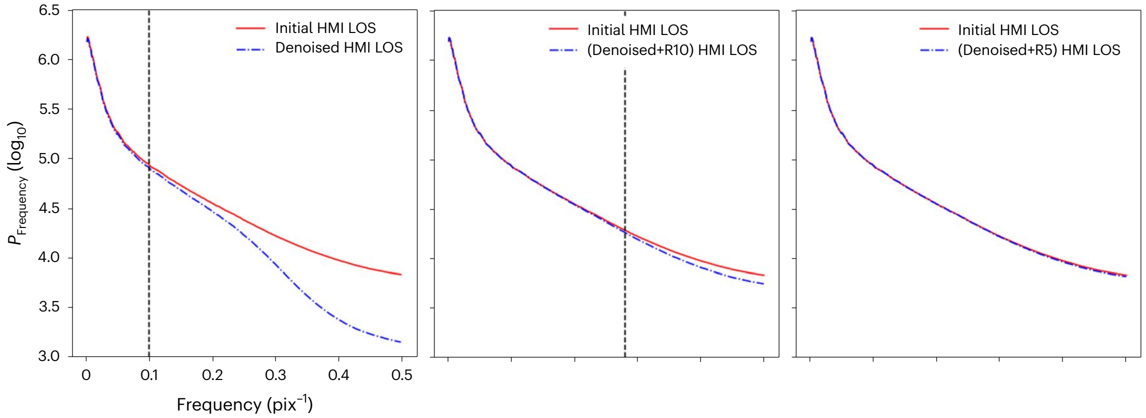

Fig. 4 | Spatial Fourier power spectra of the initial and denoised magnetograms. The vertical dashed line marks the location x where the two curves start to deviate from each other. $x = 0.10$ for the first panel, and $x = 0.28$ for the second panel. $P_{\text{Frequency}}$ denotes the power spectral intensity of the spatial Fourier power spectra.

10 × 10 pixels may exist. By incorporating the restoring operation, we can preserve structures at notably smaller scales, approximately 4 × 4 pixels for the '(Denoised+R10)' magnetograms and 2 × 2 pixels for the '(Denoised+R5)' magnetograms.

Scientific application to Hubble images

In astronomy, it is difficult to detect the faint structures of extragalactic galaxies within a limited observation time, even when using space telescopes such as HST. Our TDR method can denoise extragalactic images with limited exposure time and help to identify faint objects and structures. Here, we show an application of the TDR method in the observation of the nearby galaxy NGC 1365 ($z = 0.0054$ at a distance of 21.2 Mpc, where 1" ≈ 90 pc; for example, see refs. 30–33). Figures 5 and 6 are images of the central region (30" × 30") of NGC 1365 as observed by HST with the F555W and F814W filters. As shown in the first panel of Fig. 5, because of the notable noise in the image, which had an exposure time of 10 s, the galactic structures are indistinctive. We used the TDR method with $R = \infty$, 0.1, 0.2 or 0.3 (the unit is counts per second) to denoise the 10-s image. The denoised images are in the middle panels. Note that the primary denoised image ($R = \infty$) became cleaner, but some abnormal pixels marked by the red circle were generated. These abnormal pixels are more obvious and marked by the red ellipse in the corresponding density-scatter plot. Figure 6 indicates that the abnormal pixels occurred more frequently with the F814W filter. Observe in the residual maps that some pixels initially exhibited abnormalities in the 10-s images compared to the corresponding 100-s images. Furthermore, these abnormal pixels were accentuated during the denoising process.

To remove the enhanced abnormal pixels, we applied the restoring operation. The senior denoised images with $R = 0.1$, 0.2 or 0.3 are shown in the third, fourth and fifth panels of Figs. 5 and 6. By comparing the corresponding density-scatter plots, we found that the best denoised images are for $R = 0.3$. The Gaussian fits in the last panel of the second row in Figs. 5 and 6 indicate that the noise level of the HST images is about 0.4 or 0.5 counts per second. To meet a reasonable accuracy requirement, R needs to be less than the noise level. So $R = 0.3$ is acceptable and appropriate for the HST images.

Comparing the 100-s HST images in the last panel of the upper row of Figs. 5 and 6, we found that the senior denoised images with $R = 0.3$ are the best, The identifications of fainter star clusters and dust lanes are reliable. The residual maps in Figs. 5 and 6 show that the restoring operation can suppress abnormal pixels generated during the denoising process, but the initially existing abnormal pixels are retained as signals. Correcting these initially existing abnormal pixels relies on a comparison with long-exposure images, which is beyond the scope of noise reduction and may be adopted as a future improvement of the TDR method.

Summary and discussion

In this study, we investigate the denoising of astronomical images based on the unsupervised NN Self2Self. The denoising process consists of three operations: (1) training (T) Self2Self on a single image, (2) denoising (D) other images of the same kind and (3) restoring (R) abnormal pixels according to the accuracy requirements. We first tested this TDR method on HMI LOS magnetograms and found that the noise level

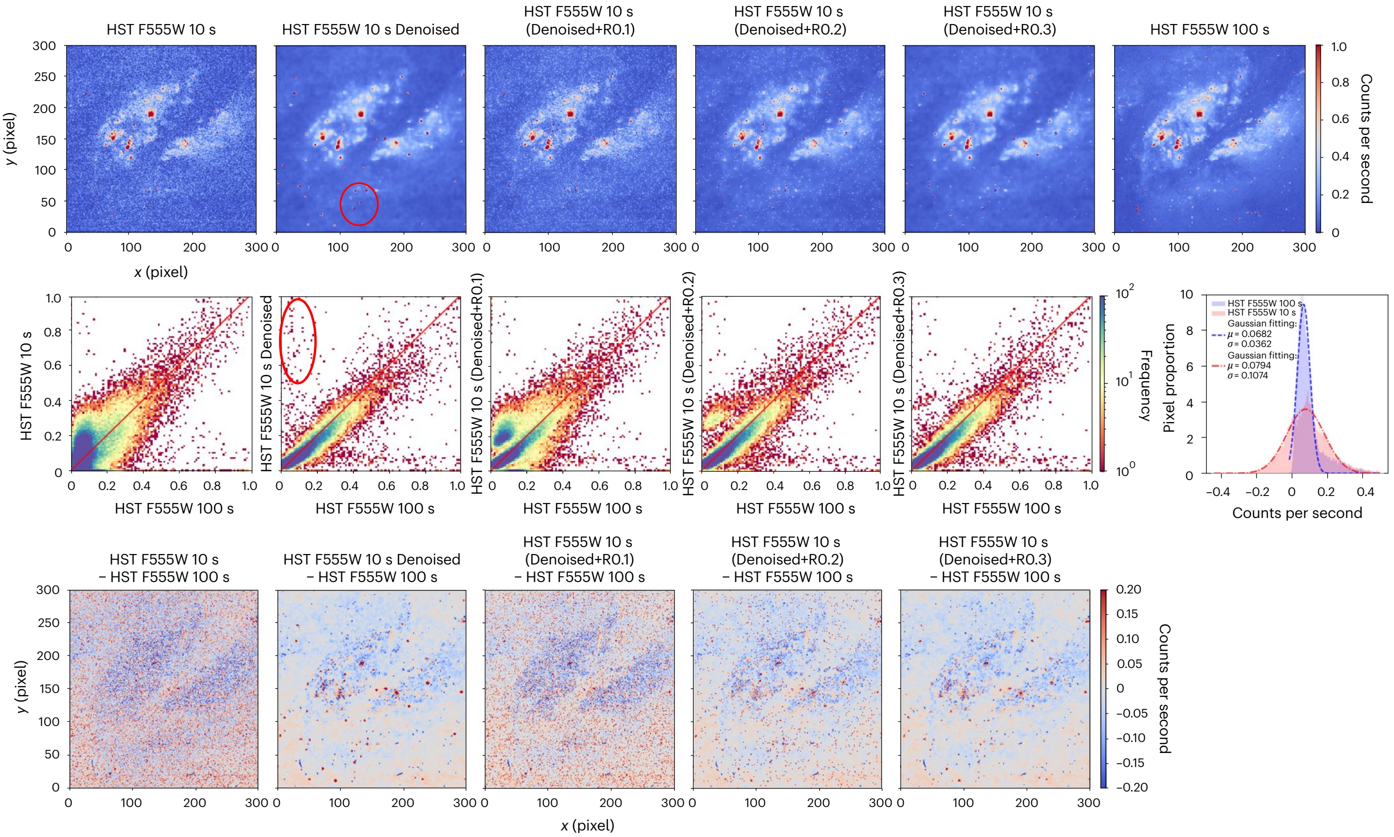

Fig. 5 | Examples of denoised images of the galactic centre of NGC 1365 observed by HST with the F555W filter. The first and last panels in the upper row show images with exposure times of 10 or 100 s. The middle four images are the corresponding denoised images with $R = \infty$ (equivalent to the restoring operation not working), 0.1, 0.2 and 0.3, respectively. Density-scatter plots of the first five images versus the last image in the upper row are displayed in the first five panels of the second row. Histograms and Gaussian fits of the HST images with exposure times of 10 or 100 s are displayed in the last panel of the second row. The residual maps are displayed in the bottom row. The unit for the pixel values is counts per second.

improved from about 8 to 2 G. In a detailed comparison of the initial and denoised magnetograms, we found that the denoised magnetograms were cleaner and smoother without sacrificing the temporal or spatial resolution. After the denoising process, we noticed an underestimation of the strong magnetic field in the primary denoised magnetograms. The underestimate exceeded the noise level. To address this issue, we proposed a two-step solution: (1) identify abnormal pixels whose change values exceed a threshold R and (2) restore those pixels to their initial values. Generally, R should be lower than the noise level of the image. When $R = 0$, the denoised images are restored to the initial images. When $R = \infty$, the restoring operation does not work. By controlling the value of R, we can obtain senior denoised images that meet a reasonable accuracy requirement.

The denoising performance of the TDR method is supported by quantitative evaluations using AE, MSE, PCC, LC and the average flux of residual images, which illustrate the robustness and reliability of the method. The denoising of HST images demonstrates the effectiveness of the TDR method in improving the clarity and detail of galaxy structures. The denoised images with a 10-s exposure time exhibit enhanced consistency compared with the corresponding images with a 100-s exposure time. This implies that the TDR method can significantly reduce the noise level of 10-s HST images. The restoration process can complement the deep learning results, which indicates that just attaining good scores for AE, MSE, PCC and so on is not sufficient. The pixel-level accuracy is more important for many scientific fields. Restoring the abnormal pixels to their initial values is the simplest restoration process. We recommend its use in other restoration processes, such as restoring to the mean values of the ambient pixels and other reliable values.

Regarding the accuracy of the denoising results, different options were adopted:

(1) Clean target images are not available for observational images. If images with higher exposure times are also not available, the denoising accuracy is estimated by comparing the denoised images with the original images. The goal is to control the changes in pixel values before and after denoising so that they are less than the noise level and make the denoised images meet a reasonable accuracy requirement, as illustrated in 'Denoising and evaluation of HMI magnetograms'.
(2) If images with higher exposure times are available, the denoising accuracy is estimated by comparing the denoised images with those images, as discussed in 'Scientific application to Hubble images'.
(3) If clean target images are available, we define the denoising accuracy by comparing the denoised images with the clean target images. The Supplementary Information gives examples of denoising synthetic images produced by the code RAMENS[34]. These have known noise.

The latter two scenarios are more suitable for supervised learning methods. In astronomy, the first scenario is common, and the TDR method could play a useful role.

For astronomical observations, both high temporal resolution and high SNR are necessary and important but they often conflict with each other. Considering different goals, to achieve a higher SNR, longer integration times are needed, whereas to achieve a higher temporal resolution, the SNR has to be sacrificed. Taking the HST as an example, ultra-deep images are usually obtained with long exposure times, for

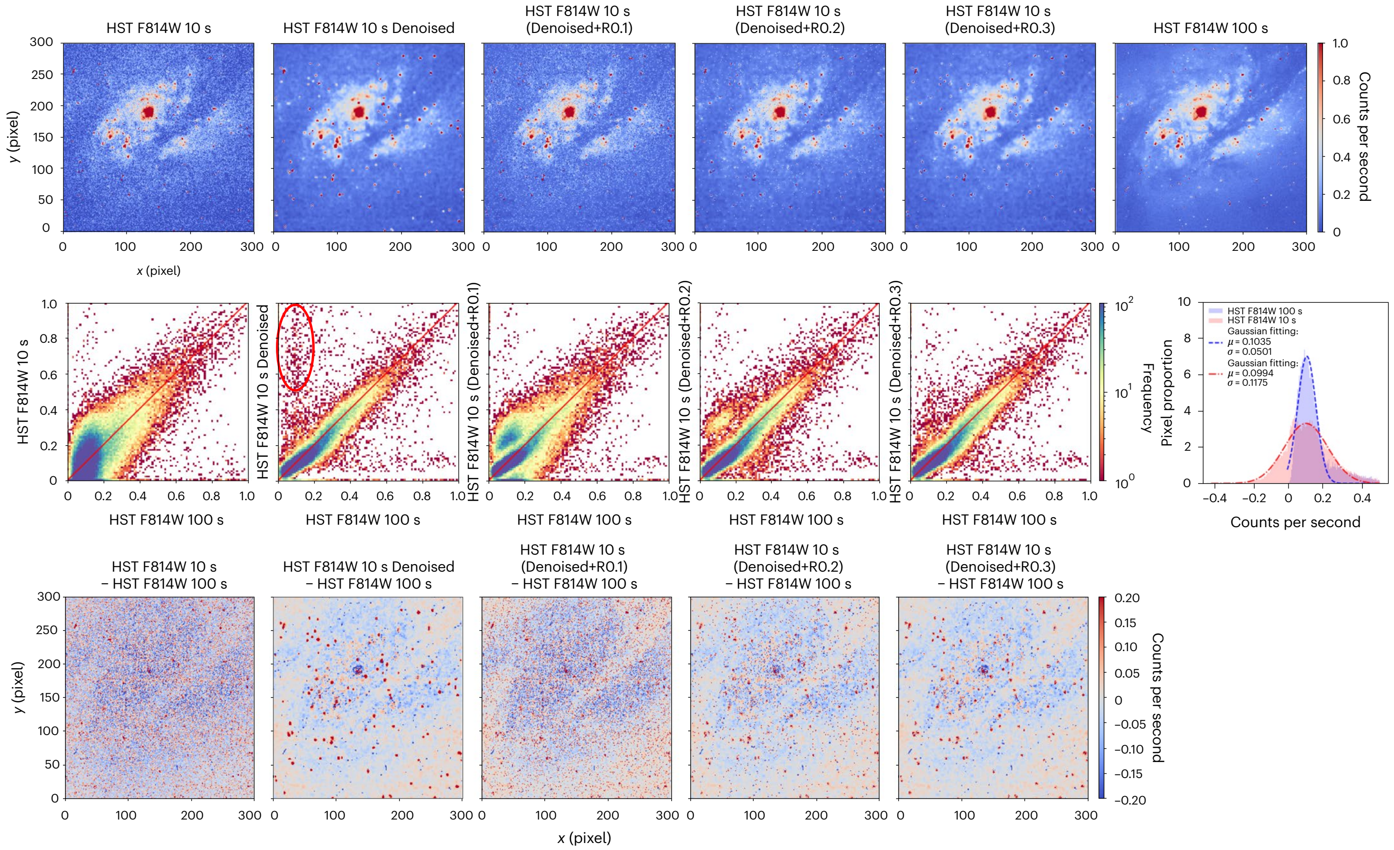

Fig. 6 | Examples of denoised images of the galactic centre of NGC 1365 observed by HST with the F814W filter. The layout and meanings of each panel are the same as those in Fig. 5 but for the F814W filter.

example, several days[35]. Taking advantage of the TDR method, we propose a two-step strategy to achieve both high temporal resolution and high SNR: (1) observe the objects with a high temporal resolution to get images with a low SNR and (2) improve the SNR with the TDR method. This is especially beneficial for astronomical sky survey projects (for example, refs. 36,37) that require abundant observation times. The TDR method holds promising potential for denoising and improving imaging efficiency in diverse disciplines.

Methods

Basic ideas

Image denoising aims to recover a clear version of a corrupted image. It is widely used in many applications. In general, image denoising estimates a clear image $\mathbf{x}$ from a corrupted image $\mathbf{y}$ generated by the following process: $\mathbf{y} = \mathbf{x} + \mathbf{n}$, where $\mathbf{n}$ denotes the noise distribution.

Inspired by the effectiveness of blind-spot methods in multi-image unsupervised denoising[23,24], we investigated a self-supervised NN, Self2Self, with the following basic form:

$$\min_{\boldsymbol{\theta}} \ell(\mathcal{F}_{\boldsymbol{\theta}}(\hat{\mathbf{y}}), \hat{\mathbf{y}}), \qquad (1)$$

where $\hat{\mathbf{y}}$ denotes counterpart images that are generated from the corrupted image $\mathbf{y}$, and $\mathcal{F}_{\boldsymbol{\theta}}(\cdot)$ represents the NNs, with $\boldsymbol{\theta}$ denoting the parameter vector; ℓ measures the distance between two images.

Compared to multi-image supervised and unsupervised methods, an effective single-image self-supervised recovery approach is much more challenging to develop. The model learned from single-image supervision will probably be overfitted and biased, as the amount of training data is very limited. Although a single model may be biased, an ensemble of several NNs with different model configurations can cancel out the bias. Nevertheless, there are still three key problems that need to be solved when using an NN ensemble for this task: (P1) How can we train and maintain several models economically? Separately training a number of NNs is impractical. (P2) How can we improve the effectiveness of each learned model? A model supervised by a single corrupted image may just memorize the target or learn identity mapping. (P3) How can we ensure the diversity of the learned models so that their ensemble can make a noticeable improvement?

In this paper, we solve these problems which dropout[38], which refers to randomly dropping out nodes in an NN. We construct an NN with dropout layers and use dropout during training and testing of the NN:

- During training, the dropout layers enable a single NN to approximate a large number of different NN architectures by randomly dropping out nodes. Thus, this is a very computationally cheap method for training several models (answer to P1).
- Dropout naturally prevents the NN from being overfitted when there is a limited amount of data. It reduces the possibility of purely memorizing the single target and learning trivial mappings. Dropout is also applied to the input layer, which is equivalent to randomly dropping pixels from the input image to prevent the model from learning an identity mapping. Together with partial convolutions[39] that ignore the effect of masked pixels during restoration, our dropout NN model can learn single-image recovery effectively (answer to P2).
- When testing, we use dropout again on the trained NN to generate several models for image recovery. It has been shown that a model trained using dropout has high model uncertainty and that the dropout models generated from it during testing have sufficient diversity (answer to P3).

Training

In training, a series of counterpart images $\hat{\mathbf{y}}_1, \dots, \hat{\mathbf{y}}_M$ were generated from the given corrupted image $\mathbf{y}$ as follows:

$$\hat{\mathbf{y}}_m = B_m \mathbf{y}, \quad \text{for all } m, \tag{2}$$

where B_m is a random binary diagonal matrix with its diagonal entry $B_m(i, i)$ drawn from a Bernoulli distribution with probability p. Then, we trained our NN model $\mathcal{F}_{\boldsymbol{\theta}}(\cdot)$ to map each $\hat{\mathbf{y}}_m$ to $\mathbf{y}$, with the dropout and loss functions:

$$\ell(\boldsymbol{\theta}) = \sum_{m=1}^{M} \|(I - B_m)(\mathcal{F}_{\boldsymbol{\theta}}(\hat{\mathbf{y}}_m) - \mathbf{y})\|_2^2. \tag{3}$$

Note that $I - B$ implies that the loss is defined on the dropped pixels. As an improvement, data augmentation was used by flipping the given image horizontally, vertically and diagonally. For computational efficiency, when noisy images from the same dataset are strongly relevant to each other, we first trained the base NN parameters for this dataset, then we separately fine-tuned each image by updating the parameters from the base one.

Testing

In the test stage, several new NNs ($\mathcal{F}_{\boldsymbol{\theta}_1}, \dots, \mathcal{F}_{\boldsymbol{\theta}_N}$) are formed by randomly running dropout on the configured layers of the trained NN $\mathcal{F}_{\boldsymbol{\theta}^*}$. Then, several recovered images $\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_N$ are generated by feeding a randomly re-corrupted version of $\mathbf{y}$ to each of the newly formed NNs. The recovered images are then averaged to obtain the final result $\mathbf{x}^*$:

$$\mathbf{x}^* = \sum_{n=1}^{N} \hat{\mathbf{x}}_i = \sum_{n=1}^{N} \mathcal{F}_{\boldsymbol{\theta}_n}(B_{M+n}\mathbf{y}). \tag{4}$$

NN architecture

The NN architecture was borrowed from U-net[16] with some modifications. See the Supplementary Information for an illustration. There are two main differences between a U-net network and our network. First, we introduced dropout to the Conv layers. For each weight entry in a layer with dropout, there is a probability that it will be set to zero. The remaining entries are then scaled to maintain the energy level. The dropout layers are used in both training and testing, which regularizes the NN, prevents it from learning trivial mappings and allows several diverse NNs to be learned simultaneously and economically. Second, we used partial convolutions instead of standard convolutions in the encoder. This can eliminate the effect of the dropped pixels during processing, with both increased effectiveness and efficiency being observed in practice.

Data availability

The HMI LOS magnetograms and HST galaxy images in this paper were downloaded from the Joint Science Operations Center (http://jsoc.stanford.edu/) and Mikulski Archive for Space Telescopes (https://archive.stsci.edu/). Interested parties can also download the corresponding data from the project page (https://github.com/zimugh/Denoising-TDR/).

Code availability

The codes for the TDR method are publicly available. We provide a Python notebook to show the denoising examples at https://github.com/zimugh/Denoising-TDR/.

References

1. Pesnell, W. D., Thompson, B. J. & Chamberlin, P. C. The Solar Dynamics Observatory (SDO). *Sol. Phys.* **275**, 3–15 (2012).
2. Schrijver, C. J. et al. Division E Commission 10: solar activity. *Proc. Int. Astron. Union* **29A**, 245–277 (2016).
3. Cao, W. et al. Scientific instrumentation for the 1.6 m New Solar Telescope in Big Bear. *Astron. Nachrichten* **331**, 636 (2010).
4. Liu, Z. et al. New vacuum solar telescope and observations with high resolution. *Res. Astron. Astrophys.* **14**, 705–718 (2014).
5. Buades, A., Coll, B. & Morel, J. M. A review of image denoising algorithms, with a new one. *Multiscale Model. Simul.* **4**, 490–530 (2005).
6. Goyal, B., Dogra, A., Agrawal, S., Sohi, B. & Sharma, A. Image denoising review: from classical to state-of-the-art approaches. *Inf. Fusion* **55**, 220–244 (2020).
7. Seddik, H. & Ben Braiek, E. Efficient noise removing based optimized smart dynamic Gaussian filter. *Int. J. Comput. Appl.* **51**, 1–13 (2012).
8. Elad, M. & Aharon, M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.* **15**, 3736–3745 (2006).
9. Yang, H.-Y., Wang, X.-Y., Niu, P.-P. & Liu, Y.-C. Image denoising using nonsubsampled shearlet transform and twin support vector machines. *Neural Netw.* **57**, 152–165 (2014).
10. Burger, H. C., Schuler, C. J. & Harmeling, S. Image denoising: can plain neural networks compete with BM3D? In *Proc. 2012 IEEE Conference on Computer Vision and Pattern Recognition* 2392–2399 (IEEE, 2012).
11. Vemulapalli, R., Tuzel, O. & Liu, M.-Y. Deep Gaussian conditional random field network: a model-based deep network for discriminative denoising. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 4801–4809 (IEEE, 2016).
12. 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.* **26**, 3142–3155 (2017).
13. Park, E., Moon, Y.-J., Lim, D. & Lee, H. De-noising SDO/HMI solar magnetograms by image translation method based on deep learning. *Astrophys. J.* **891**, L4 (2020).
14. Lehtinen, J. et al. Noise2Noise: learning image restoration without clean data. In *Proc. 35th International Conference on Machine Learning, ICML 2018* (eds Dy, J. & Krause, A.) 4620–4631 (International Machine Learning Society, 2018).
15. Díaz Baso, C. J., de la Cruz Rodríguez, J. & Danilovic, S. Solar image denoising with convolutional neural networks. *Astron. Astrophys.* **629**, A99 (2019).
16. Ronneberger, O., Fischer, P. & Brox, T. U-net: convolutional networks for biomedical image segmentation. In *Proc. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (eds Navab, N. et al.) 234–241 (Springer International Publishing, 2015).
17. Chen, C., Chen, Q., Xu, J. & Koltun, V. Learning to see in the dark. In *Proc. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* 3291–3300 (IEEE, 2018).
18. Vojtekova, A. et al. Learning to denoise astronomical images with U-nets. *Mon. Not. R. Astron. Soc.* **503**, 3204–3215 (2020).
19. Qi, J. et al. An improved U-net model for astronomical images denoising. In *Proc. 2022 China Automation Congress (CAC)* 1901–1905 (IEEE, 2022).
20. Goodfellow, I. et al. Generative adversarial networks. *Commun. ACM* **63**, 139–144 (2020).
21. Schawinski, K., Zhang, C., Zhang, H., Fowler, L. & Santhanam, G. K. Generative adversarial networks recover features in astrophysical images of galaxies beyond the deconvolution limit. *Mon. Not. R. Astron. Soc.* **467**, L110–L114 (2017).
22. Chen, J., Chen, J., Chao, H. & Yang, M. Image blind denoising with generative adversarial network based noise modeling. In *Proc. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* 3155–3164 (IEEE, 2018).
23. Krull, A., Buchholz, T.-O. & Jug, F. Noise2Void - learning denoising from single noisy images. In *Proc. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 2124–2132 (IEEE, 2018).

24. Batson, J. & Royer, L. Noise2Self: Blind denoising by self-supervision. In *Proc. 36th International Conference on Machine Learning* (eds Chaudhuri, K. & Salakhutdinov, R.) 524–533 (PMLR, 2019).
25. Quan, Y., Chen, M., Pang, T. & Ji, H. Self2Self with dropout: learning self-supervised denoising from single image. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2020).
26. Schou, J. et al. Design and ground calibration of the helioseismic and magnetic imager (HMI) instrument on the Solar Dynamics Observatory (SDO). *Sol. Phys.* **275**, 229–259 (2012).
27. Liu, Y. et al. Comparison of line-of-sight magnetograms taken by the Solar Dynamics Observatory/helioseismic and magnetic imager and Solar and Heliospheric Observatory/Michelson Doppler imager. *Sol. Phys.* **279**, 295–316 (2012).
28. Ji, H. et al. Magneto-acoustic oscillations observed in a solar plage region. *Res. Astron. Astrophys.* **21**, 179 (2021).
29. Park, E., Moon, Y.-J., Lim, D. & Lee, H. De-noising SDO/HMI solar magnetograms by image translation method based on deep learning. *Astrophys. J. Lett.* **891**, L4 (2020).
30. Lindblad, P. O. NGC 1365. *Astron. Astrophys. Rev.* **9**, 221–271 (1999).
31. Galliano, E., Alloin, D., Pantin, E., Lagage, P. O. & Marco, O. Mid-infrared imaging of active galaxies, active nuclei and embedded star clusters. *Astron. Astrophys.* **438**, 803–820 (2005).
32. Wang, J. et al. Imaging the circumnuclear region of NGC 1365 with Chandra. *Astrophys. J.* **694**, 718–733 (2009).
33. Gao, Y. et al. The nuclear region of NGC 1365: star formation, negative feedback, and outflow structure. *Astrophys. J.* **913**, 139 (2021).
34. Iijima, H. & Yokoyama, T. A three-dimensional magnetohydrodynamic simulation of the formation of solar chromospheric jets with twisted magnetic field lines. *Astrophys. J.* **848**, 38 (2017).
35. Beckwith, S. V. W. et al. The Hubble ultra deep field. *Astron. J.* **132**, 1729–1755 (2006).
36. Su, Y. et al. The Milky Way Imaging Scroll Painting (MWISP): project details and initial results from the Galactic longitudes of 25.°8–49.°7. *Astrophys. J. Suppl. Ser.* **240**, 9 (2019).
37. Norris, R. P. et al. The evolutionary map of the Universe pilot survey. *Publ. Astron. Soc. Aust.* **38**, e046 (2021).
38. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
39. Liu, G. et al. Image inpainting for irregular holes using partial convolutions. In *Proc. Computer Vision – ECCV 2018* (eds Ferrari, V. et al.) 89–105 (Springer International Publishing, 2018).

Acknowledgements

The magnetograms were observed by the HMI onboard the Solar Dynamics Observatory, which is a mission of the Living with a Star Program from NASA. The galaxy images are based on observations made by the NASA/ESA HST obtained from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. These galaxy observations are associated with programme 5222. The NN Self2Self was constructed with TensorFlow. This work is supported by the National Key R&D Program of China (Grant Nos. 2022YFF0503004 to T.L., Y.G. and Y.W., 2021YFA1600502 and 2022YFF0503001 to Y.S. and 2020YFC2201201 to Y.G.), by the National Natural Science Foundation of China (Grant Nos. 12203023 to T.L., 62372186 to Y.Q., 12333009 to Y.G. and 12173092 to Y.S.), by the Natural Science Foundation of Guangdong Province (Grant Nos. 2022A1515011755 and 2023A1515012841 to Y.Q.), by the Fundamental Research Funds for the Central Universities (Grant No. 2023ZYGXZR022 to Y.Q.) and by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB0560000 to Y.S. and H.J.).

Author contributions

T.L., Y.Q., Y.S. and Y.G. led the study, developed the method and wrote the manuscript. S.L., H.J. and Q.H. contributed to providing the scientific data and the analysis of the denoising of the magnetograms. Y.L.G. analysed the application to Hubble images. Y.L., Y.W., W.S. and M.D. ran the tests on synthetic images, improved the definition of the accuracy and polished the manuscript. All authors contributed to the discussion and to improving the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41550-025-02484-z.

Correspondence and requests for materials should be addressed to Tie Liu, Yuhui Quan, Yingna Su or Yang Guo.

Peer review information *Nature Astronomy* thanks the anonymous reviewers for their contribution to the peer review of this work.