

Physics informed deep learning to super-resolve and cross-calibrate solar magnetograms

Andres Munoz-Jaramillo (✉ amunozj@boulder.swri.edu)

Southwest Research Institute <https://orcid.org/0000-0002-4716-0840>

Anna Jungbluth

University of Oxford

Xavier Gitiaux

George Mason University

Paul Wright

Wright AI Ltd <https://orcid.org/0000-0001-9021-611X>

Carl Shneider

CWI <https://orcid.org/0000-0002-3689-6959>

Shane Maloney

Trinity College Dublin

Atılım Güneş Baydin

University of Oxford <https://orcid.org/0000-0001-9854-8100>

Yarin Gal

University of Oxford <https://orcid.org/0000-0002-2733-2078>

Michel Deudon

Element AI

Freddie Kalaitzis

University of Oxford <https://orcid.org/0000-0002-1471-646X>

Article

Keywords: super-resolution techniques, deep learning, deep neural networks

Posted Date: October 6th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-713430/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Physics informed deep learning to super-resolve and cross-calibrate solar magnetograms

Andrés Muñoz-Jaramillo^{1,†}, Anna Jungbluth^{2,†}, Xavier Gitiaux^{3,†}, Paul J. Wright^{4,5,†}, Carl Schneider⁶, Shane A. Maloney^{7,10}, Atılım Güneş Baydin², Yarin Gal², Michel Deudon⁹, and Freddie Kalaitzis^{2,8}

[†] *These authors contributed equally to this publication*

¹ *Southwest Research Institute*

² *University of Oxford*

³ *George Mason University*

⁴ *Stanford University*

⁵ *Wright AI Ltd*

⁶ *Dutch National Center for Mathematics and Computer Science (CWI)*

⁷ *Trinity College Dublin*

¹⁰ *Dublin Institute for Advanced Studies*

⁸ *Frontier Development Lab*

⁹ *Element AI*

September 20, 2021

Abstract

Super-resolution techniques aim to increase the resolution of images by adding detail. Compared to upsampling techniques reliant on interpolation, deep learning-based approaches learn features and their relationships across the training data set to leverage prior knowledge on what low resolution patterns look like in higher resolution images. As an added benefit, deep neural networks can learn the systematic properties of the target images (i.e. texture), combining super-resolution with instrument cross-calibration. While the successful use of super-resolution algorithms for natural images is rooted in creating perceptually convincing results, super-resolution applied to scientific data requires careful quantitative evaluation of performances. In this work, we demonstrate that deep learning can increase the resolution and calibrate space- and ground-based imagers belonging to different instrumental generations. In addition, we establish a set of measurements to benchmark the performance of scientific applications of deep learning-based super-resolution and calibration. We super-resolve and calibrate solar magnetic field images taken by the Michelson Doppler Imager (MDI; resolution $\sim 2''/\text{pixel}$; science-grade, space-based) and the Global Oscillation Network Group (GONG; resolution $\sim 2.5''/\text{pixel}$; space weather operations, ground-based) to the pixel resolution of images taken by the Helioseismic and Magnetic Imager (HMI; resolution $\sim 0.5''/\text{pixel}$; last generation, science-grade, space-based).

Super-resolution (SR) is an image processing technique that aims to increase the resolution of an image by adding sub-pixel detail.¹ The information used for adding detail can come from sub-pixel shifts provided by sequences of images (frequency domain), or by a good understanding of the degradation processes, including blurring, that cause the loss of detail (i.e., atmospheric seeing, point spread function, etc.)¹ In the case of applications with sufficient and representative low (LR) and high-resolution (HR) samples, context can provide an additional source of information (i.e., the knowledge that all LR images belong to a specific category). convolutional neural networks (CNNs) are especially suited for this type of application due to their ability to empirically map the underlying connections between an image pixel and those surrounding it.² Furthermore, neural networks can learn features and feature relationships that are inherent

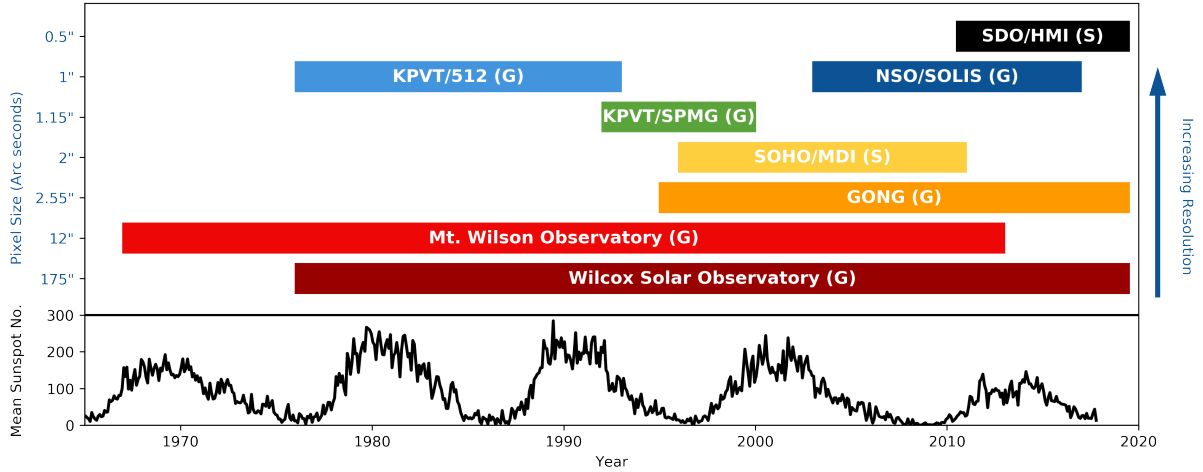


Figure 1: **Overview of ground (G) and space-based (S) observations of the Sun.** The top panel shows the pixel size in arc seconds, and the operation time span of eight different instruments. The bottom panel shows the variation of the mean sunspot number over the last 50 years. To date, the Helioseismic and Magnetic Imager (HMI), on board the Solar Dynamic Orbiter (SDO), provides the highest resolution full disk magnetograms. This is followed by the 512 channel magnetograph of the Kitt Peak Vacuum Tower (KPVT) and the Synoptic Optical Long-term Investigations of the Sun (SOLIS) by the National Solar Observatory (NSO), the CCD spectromagnetograph (SPMG), the Michelson Doppler Imager (MDI) on board the Solar and Heliospheric Observatory (SOHO), the Global Oscillation Network Group (GONG), the Mt. Wilson Observatory, and finally the Wilcox Solar Observatory.

to the data domain as a whole. For example, human faces typically have a nose underneath two eyes and above a mouth; meaning that pixels near the center of a low-resolution image of a human face can be safely assumed to constitute a nose.³ The majority of applications of CNNs to super-resolution involve the super-resolution of natural images (i.e., images with three color channels representing red, green, and blue). These approaches focus on a super-resolution outcome that is tailored for human visual perception. In other words, their objective is the production of images that look right to the human observer.⁴

Deep learning applications of super-resolution to the physical sciences have tremendous potential due to their ability to simultaneously super-resolve (add scientifically accurate detail to images) and cross-calibrate (correct systematic differences between instruments). However, scientific images have a significantly larger dynamic range than natural images (each image pixel can assume real values spanning several orders of magnitude) and the generation of perceptually correct images is not a sufficient outcome because their use typically involves precise numerical calculations that are quantitatively more sensitive than measuring perceptual quality. Our work has three main objectives: (1) to demonstrate that a deep learning approach can leverage the information present in astronomical images with the goal of adding detail to low resolution images while maintaining their scientific accuracy; (2) to show how super-resolving a scientific image via deep learning also homogenizes instrument systematic properties, (3) to establish a set of quantitative performance measurements that can be used to benchmark the performance of different super-resolution algorithms for astronomical images, as well as to benchmark the performance of future applications of super-resolution to the physical sciences.

1 Solar Magnetogram Calibration: an outstanding problem for 50 years

Over the last 50 years, space- and ground-based instruments have mapped the solar surface magnetic field (Figure 1). These images, known as magnetograms, have significantly advanced our understanding of solar magnetism,⁵ understanding of the solar corona,⁶ and prediction of space-weather events.⁷ Magnetograms are constructed from measurements of spectral polarization⁸ and are thus obtained solving an ill-posed problem. Despite the wealth of archival data, differences in resolution, spectral inversion techniques, instrument noise levels, or other instrument properties prevent us from easily combining data across instruments to study small-scale magnetic field structures over multiple solar cycles (Figure 1).⁹ Unlike traditional cross-calibration techniques such as pixel-to-pixel comparison,¹⁰ histogram equalization¹¹ or harmonic scaling,¹² our results indicate that deep learning can leverage all the information and context present in magnetograms. This allows us to encode the structure of the magnetic field in a lower dimension latent space and then map magnetograms from one instrument to the other.

Previous deep learning approaches relied on physics-based models to simulate high-resolution magnetograms to use as ground truth.⁹ Other deep learning approaches super-resolved a down-scaled version of the same instrument.¹³ The novelty in our approach is that we use deep learning to cross-calibrate and super-resolve across different instruments. We cross-calibrate and super-resolve line-of-sight (LOS) magnetograms from the Michelson Doppler Imager (MDI; $\sim 2''/\text{pixel}$; science-grade, space-based) on board the Solar and Heliospheric Observatory (SoHO),¹⁴ as well as LOS magnetograms taken by the National Solar Observatory’s (NSO) Global Oscillation Network Group (GONG; $\sim 2.5''/\text{pixel}$; space weather operations, ground-based)¹⁵ to the $\sim 0.5''/\text{pixel}$ resolution of magnetograms taken by the Helioseismic and Magnetic Imager (HMI; last generation, science-grade, space-based) on board the Solar Dynamics Observatory (SDO).

2 Quantities to Assess Physical Properties of Super-Resolved Magnetograms

Compared to natural images, which often consist of three color channels with integer pixel values, line-of-sight measurements of the magnetic field of the Sun measure spectral signatures from which the magnetic field strength is estimated. As such, each pixel value describes the strength of the magnetic field which is now a signed quantity and not constrained to a maximum value. Given that the Sun is a 3-dimensional object projected onto a 2-dimensional image, measurements around the limb show larger projection effects than those closer to the center of the solar disk.

To measure the performance of any super-resolution or cross-calibration operation of solar magnetograms, it is essential to approach them as scientific measurements rather than standard images. We propose to use the following quantities to compare performances of super-resolution/cross-calibration approaches. Note that these quantities are post-mortem measurements that, we believe, should be reported for any super-resolution/cross-calibration methods of solar magnetograms. However, they can be easily adapted to other astronomical data.

We denote $\hat{B}_{i,j,n}$ as the super-resolved magnetic field at pixel (i, j) and patch n ; and we denote $B_{i,j,n}$ the ground truth target magnetic field value for the corresponding patch and at the same location. Each patch n , unless specified otherwise, refers to an area of 128×128 pixels, corresponding to $1/1024$ of a full disk HMI magnetogram.

- *Correlations:* We follow reference¹⁶ and measure how super-resolved magnetograms are cross-calibrated to their high resolution counterpart by measuring the Pearson correlation

coefficient between $\hat{B} = \{\hat{B}_{i,j,n}\}$ and $B = \{B_{i,j,n}\}$ across all pixel i, j and patches n :

$$\rho = \frac{\sum_{i,j,n} (B_{i,j,n} - \bar{B})(\hat{B}_{i,j,n} - \bar{\hat{B}})}{\sqrt{\sum_{i,j,n} (B_{i,j,n} - \bar{B})^2 \sum_{i,j,n} (\hat{B}_{i,j,n} - \bar{\hat{B}})^2}}, \quad (1)$$

where \bar{B} and $\bar{\hat{B}}$ is the average ground-truth and super-resolved magnetic field across all pixels and patches. ρ takes value between 0 and 1. The larger ρ the better is the cross-calibration of the super-resolved magnetograms to their high-resolution counterpart.

- *Signed fluxes:* Magnetic fields are divergence free: the integration of the radial magnetic field over the solar surface sums to zero. On a full disc, positive or negative biases in signed fluxes would violate the zero divergence of the solar magnetic field. To evaluate how a super-resolution technique conserves the signed flux, we calculate the signed flux of a pixel by converting the line-of-sight field into radial field and correcting for area foreshortening.:

$$E_{i,j,n}^{flux} = \hat{\phi}_{i,j,n} - \phi_{i,j,n} \quad (2)$$

- *Extreme values:* Regions with extreme values of magnetic field occupy areas that are smaller than the area covered by a pixel of a magnetogram with the resolution of MDI or GONG. Therefore, the ability of a super-resolution technique to generate these sub-pixel extreme values is of special interest, particularly for the study of sunspots and active regions. Moreover, extreme values of the magnetic field have low frequency. Therefore, they may be difficult to be captured by a super-resolution technique that learns its predictions from data with limited number of occurrences of extreme values. To measure the ability of a super-resolution technique to reproduce the tail of the distribution of magnetic field, we compute the absolute difference between the minimum/maximum magnetic field over each 128×128 patch n :

$$E_n^{max} = \left| \max B_n - \max \hat{B}_n \right| \quad (3)$$

and

$$E_n^{min} = \left| \min B_n - \min \hat{B}_n \right| \quad (4)$$

- *Gradients:* The filling factor or ratio of the area occupied by the magnetic field to the total area is smaller at high resolution than at low resolution. That is, large magnetic field values occupy smaller area in high resolution magnetograms than in lower resolution magnetograms. Pixel-level gradients of magnetic field values can quantify variations of the magnetic field around a low-resolution pixel. It also helps to evaluate how a super-resolution technique captures polarity inversion and defines boundaries between positive and negative regions. We compute $E_{i,j,n}^g$ as the (i, j) pixel of the image gradient of the difference $\hat{B}_n - B_n$ between the predicted and true magnetograms of patch n :

$$E_{i,j,n}^g = g(\hat{B}_n - B_n)_{i,j}, \quad \text{where} \quad (5)$$

$$g(I)_{i,j} = \sqrt{g_x(I)_{i,j}^2 + g_y(I)_{i,j}^2}, \quad \text{with} \quad (6)$$

$$g_x(I) = G_x * I. \quad (7)$$

is the (i, j) pixel of the output image of the Sobel operator g on image I . G_x and G_y are 3×3 kernels, that convolve an image to produce the smoothed finite difference on the x and y image dimensions respectively.

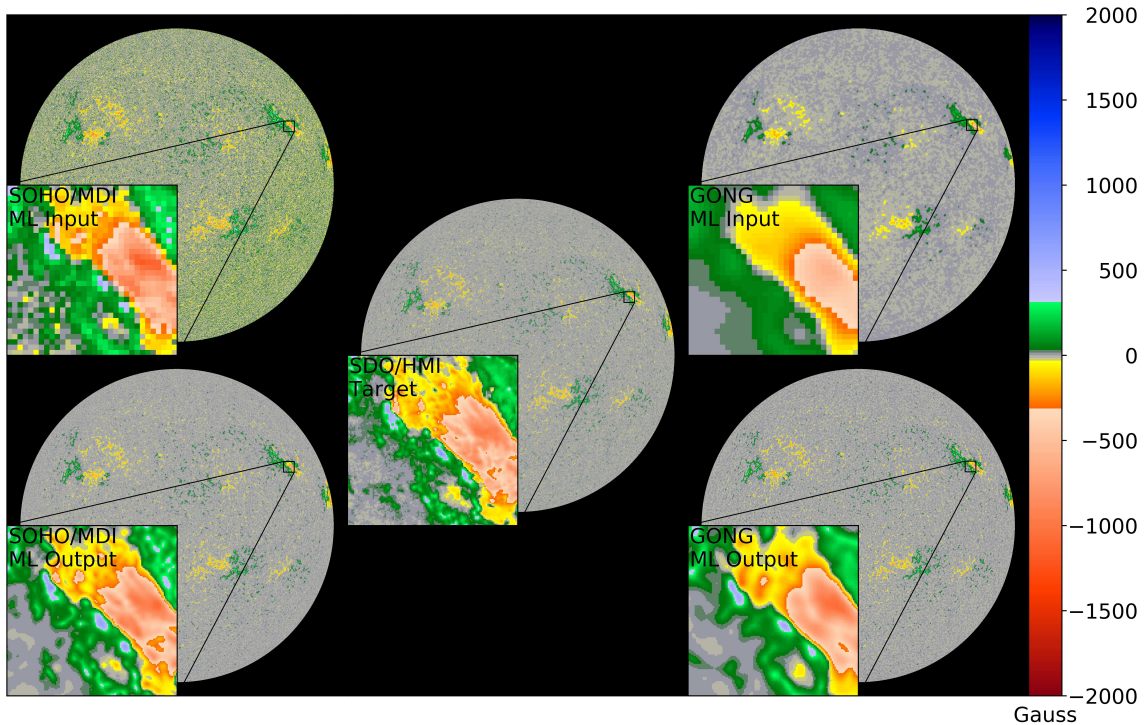


Figure 2: **Example of the resolution differences between an input magnetogram (MDI/GONG), the deep learning output and the target (HMI).** The insets show a 128×128 patch with clear resolution differences between the model inputs (MDI, and GONG) and the model target (HMI). All panels show nearly simultaneous measurements made by GONG, MDI, and HMI, on 2011-March-14.

To measure the performance of super-resolution, we compute the signed flux and extreme values at small spatial scales using patches of size 4×4 , 8×8 , 16×16 and 32×32 pixels. In addition, we also calculate the Pearson correlation as a function of magnetic field strength and location on the surface of the Sun. This allows us to understand how the performance of any super-resolution technique applied to the solar magnetic field depends on spatial scale and strength of the magnetic field.

3 Results

Figure 2 shows full disk images of our input data (top row) and the best results of our deep learning super-resolution network (bottom row). The insets show one of the 1024 patches used to split up the Sun during training. These results were achieved with a loss function that features a combination of differentiable physics-informed terms including the mean square error (MSE), magnetic field gradients, pixel histograms, and self-similarity penalties. More information on the loss function can be found in the Supplementary Information. The super-resolved full disk magnetograms of MDI and GONG have noise levels, texture, and relative magnetic field intensity akin to those of the HMI target. Zooming in closer, the insets show higher-resolution structures for the model’s outputs, which better match those of the HMI target. The improvement is especially significant for GONG, with a striking difference in small-scale structures between the input and output patch.

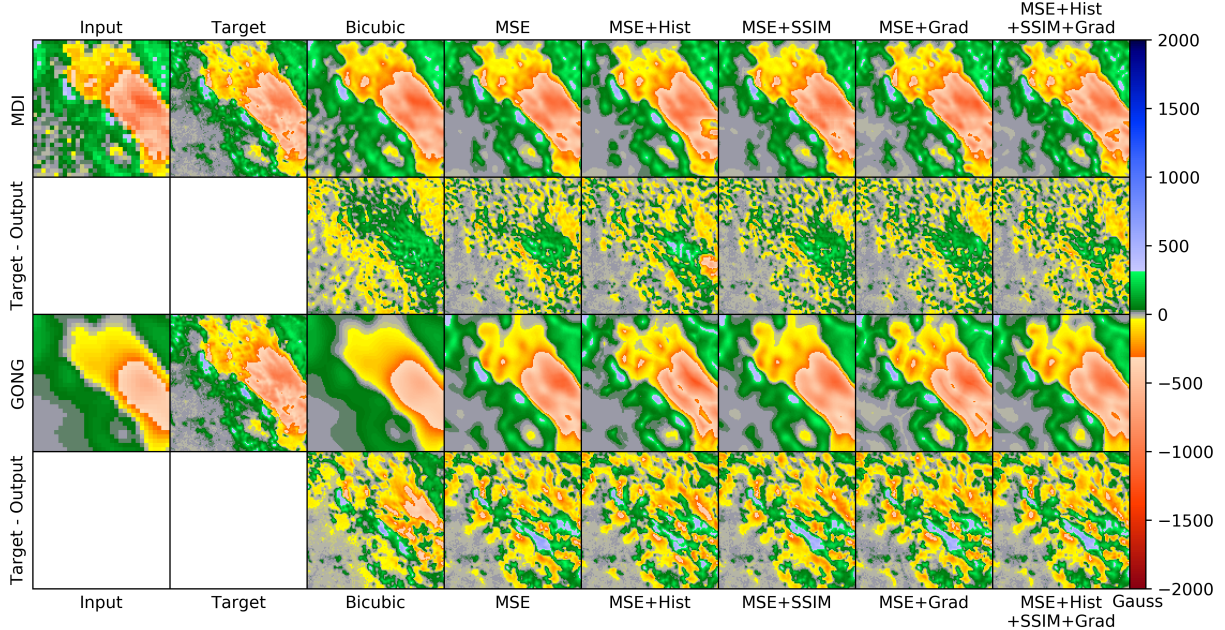


Figure 3: **Comparison of the same area of a magnetogram on 2011-March-14 for the input, target and deep learning output with different loss functions.** The top two rows show MDI magnetograms as input, the bottom two rows show a comparison for conversion of GONG magnetograms.

3.1 Baseline

To benchmark our machine learning approach, we compare against a bicubic upsampling baseline. Bicubic upsampling interpolates only the information contained in the low resolution image, and does not add new information to the higher resolution counterpart.

We follow the method presented in¹⁶ and apply a cross-calibration factor to MDI and GONG. We perform a linear regression of low resolution magnetic fields (MDI or GONG) against high resolution magnetic field (HMI) of the form $\text{MDI/GONG} = a + b \times \text{HMI}$. We find that $b = 1.3^1$ for MDI and $b = 0.7$ for GONG. We construct our baseline by bicubic upsampling and then, scaling of MDI by $1/1.3$ and GONG by $1/0.7$.

In addition to our baseline, we also compare our work to results achieved with the same neural network but employing a loss function that is only based on the mean squared error of the deep learning output and target. This allows us to highlight the need for including physics-informed terms in the loss function when handling scientific data.

3.2 Ablation Study – Optimization Penalty Terms

3.2.1 Visual Comparison

Figure 3 compares super-resolved magnetograms obtained with different loss functions to the input (MDI/GONG), the target (HMI), and our bicubic upsampling baseline. The first row (third row) in Figure 3 shows the same patch of the Sun with MDI (GONG) as input. The second and last row show the calculated difference between the up-sampled magnetograms and the target. The input patches consists of 32×32 pixels, while the target and deep learning output measure 128×128 pixels.

¹⁶ finds a value of $b = 1.4$, but their regression uses uniform weight across all inputs. Instead, we bin the low resolution magnetic field and weight each input by the fraction of points that fall into the bin the input belongs to. A weighted regression balances the impact of low and high magnetic field.

Starting with our baseline, the bicubic up-sampled MDI magnetogram still shows the salt-and-pepper like noise structure that is present in the MDI input in the lower left corner of the magnetogram patch. This is because simple upsampling techniques extrapolate the magnetic field, including its noise, to the higher resolution image. Moreover, bicubic upsampling cannot leverage the information present in the whole dataset of magnetograms. When bicubic upsampling GONG, edges around active regions become sharper, but the large, patch-like features do not increase in detail.

Using our deep learning model with a simple mean squared error loss removes the noise floor of the MDI input image. In addition, we start to recover small-scale features in and around active regions. Adding optimization penalty terms to the MSE loss modifies details in the high resolution reconstructions. It also visibly reduces the characteristic size of the structures in the difference images. We see this as evidence that the additional loss terms allows the CNN to better capture the structure of the target magnetic field. However, purely visual inspection of the images is not enough to find significant differences or distinguish which loss function is best at recovering high resolution features.

3.2.2 Quantitative Comparison

Figure 4 is an ablation study that compares the effect of each component included in the loss function on the reconstruction of the magnetic field. We compare performances by evaluating the post-mortem quantities introduced in section 2 and calculate (a) differences in extreme magnetic field values, (b) the Pearson correlation coefficient, (c) differences in image gradients, and (d) differences in the signed flux of the target and deep learning output magnetograms. All metrics are calculated on a pixel-to-pixel basis across our test set, which contains approximately 1 million patches for MDI and 8 million patches for GONG. Figure 4 shows the results obtained for MDI input magnetograms.

As mentioned above, a simple MSE loss succeeds at creating visually pleasing magnetogram outputs that show a higher level of detail than the input magnetograms (see Figure 3). However, an objective function based exclusively on MSE is unable to reconstruct extreme values of the magnetic field (i.e. the strongest positive and negative magnetic fields in a patch) properly as shown in Figure 4a. Looking at how well extreme values are reconstructed, we observe double peaks centered around ± 100 Gauss in the bicubic baseline, and when using an MSE loss. With MSE alone, the neural network consistently underestimates the magnitude of extreme values, leading to a peak centered around $+100$ Gauss as the difference between maximum field values, and the second peak centered around -100 Gauss as the difference between minimum field values of the target and the deep learning output.

Including a gradient penalty term in the loss function (indicated in the third row of Figure 4 as MSE + Grad) removes the double peaks and centers the distribution around zero (red line). Taking image gradients into account is a measure often used in computer vision to improve edge detection and texture matching.¹⁷ For the application to magnetograms, edge detection aids to define boundaries around active regions, and texture matching helps to recover detailed features in the high resolution image. Despite these improvements, maximum fields are still slightly underestimated, as indicated by the fact that the distribution of extreme values is asymmetrically skewed towards positive values for an MSE + gradient loss function (Figure 4a, third row).

On average, the sum of the magnetic field values on the surface of the Sun is expected to be close to zero. Deviations from zero only occur when the leading part of an active region comes into view of the instrument, and the following cancellation of the magnetic field cannot be viewed yet. Biases in reconstructing positive or negative fields in the super-resolved magnetic field would violate important properties of the magnetic field. The histogram penalty (MSE + Grad + Hist in the fourth row of Figure 4) manages to mostly correct the skewed distribution of extreme values while also slightly shifting the discrepancies in image gradients (Figure 4c)

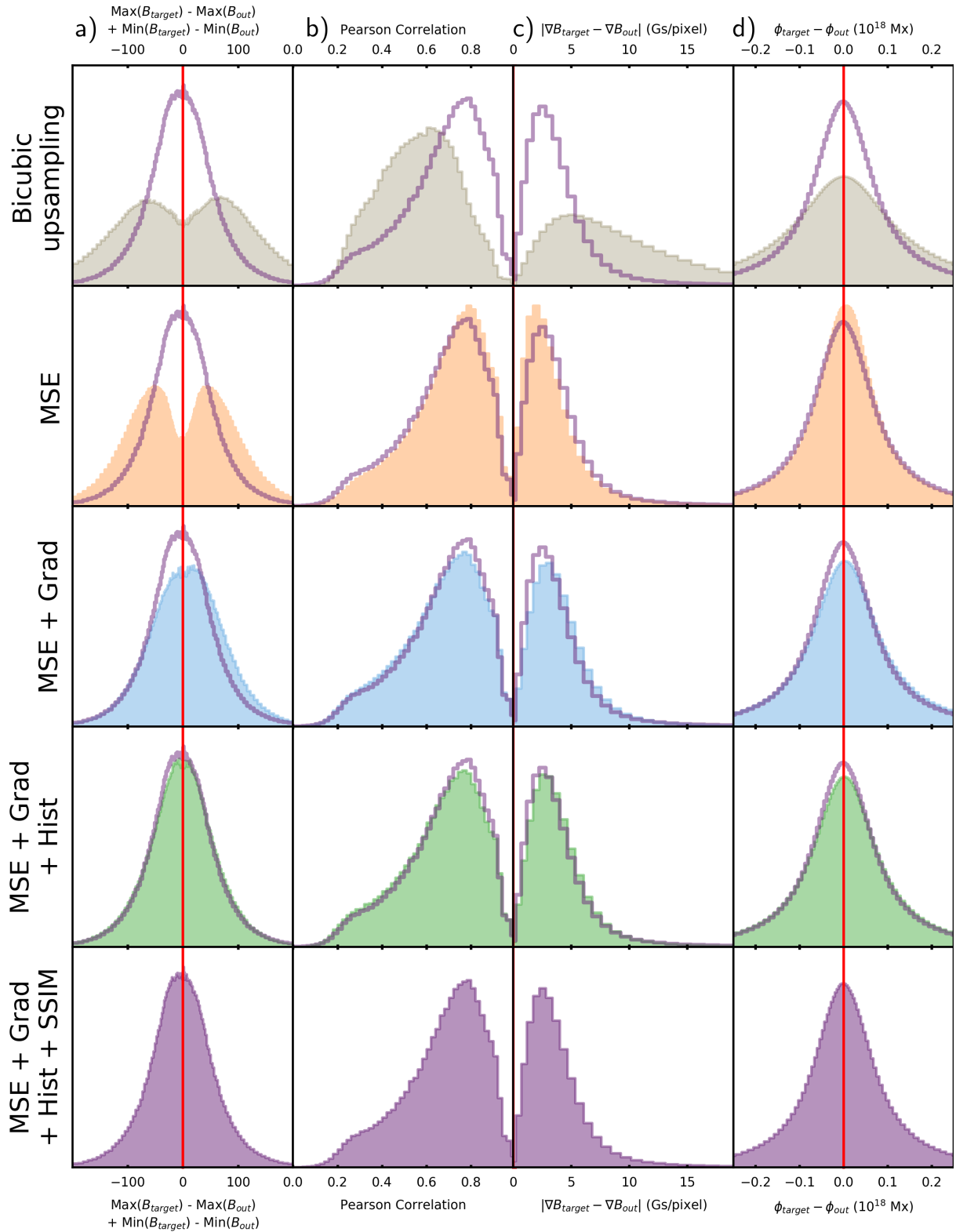


Figure 4: **Quantitative comparison of the performance of different loss functions trained on MDI magnetograms.** All loss functions are based on the mean squared error term, plus up to three additional penalty terms. The shaded histograms correspond to calculations across the test set for patches within 90% of the solar disk radius. A red vertical line indicates the ideal value for the corresponding performance metric. The two leftmost columns show calculations performed across patches of 32×32 pixels. The two rightmost columns show calculations performed per pixel in each patch of the test set. Columns left to right show: 1. added difference between the target and output maxima and minima per patch. 2. Pearson correlation per patch. 3. Magnitude of the target and output gradient difference per pixel. 4. the difference in target and output magnetic flux per pixel. To compare MSE+Grad+Hist+SSIM with the rest, the bottom row is superimposed (non-shaded histogram) on all other rows.

closer to zero.

We improve further performances by adding a similarity penalty term (SSIM, see Supplementary Information) that forces the model to learn spatial structures of the solar magnetic field (MSE + Grad + Hist + SSIM in the fifth row of Figure 4).

4 Discussion

Area	MDI	SR-MDI	GONG	SR-GONG
$0 \leq r \leq 1$	0.89	0.93	0.77	0.88
$0 \leq r \leq 2/3$	0.91	0.94	0.82	0.90
$0 \leq r \leq 1/3$	0.90	0.94	0.80	0.88
$1/3 \leq r \leq 1/2$	0.91	0.94	0.82	0.90
$1/2 \leq r \leq 3/4$	0.91	0.94	0.83	0.90
$10G \leq field$	0.92	0.94	0.82	0.90
$600G \leq field$	0.97	0.98	0.94	0.97
$0 \leq field \leq 600G$	0.82	0.88	0.68	0.81

Table 1: **Comparison of MDI/super-resolved SR-MDI and GONG/super-resolved SR-GONG with HMI.** The radius of the Sun r is normalized to 1 at the limb. The table reports the Pearson correlation coefficient between MDI and HMI; and, GONG and HMI. Higher values are better. To compare with HMI, MDI and GONG images are upsampled as described in section 2.1. The Pearson correlation coefficient is computed as in (1)

Kernel size <i>Pixels</i>	Gradient <i>Gauss</i>				Extreme values <i>Gauss</i>			
	MDI	SR-MDI	GONG	SR-GONG	MDI	SR-MDI	GONG	SR-GONG
2	4.91	3.38	4.06	3.90	8.07	7.04	8.52	7.92
4	4.92	3.38	4.07	3.91	10.70	10.27	13.18	11.45
8	4.93	3.39	4.08	3.91	16.33	16.51	22.37	18.72
16	4.99	3.40	4.09	3.93	30.64	30.29	42.0	35.39
32	5.06	3.43	4.13	3.96	66.69	61.82	87.74	74.02

Table 2: **Quantitative Comparison between MDI/super-resolved SR-MDI and HMI; and between GONG/super-resolved SR-GONG and HMI.** The table reports the average over patch of size 4×4 , 8×8 , 16×16 and 32×32 over the extreme values and gradient metrics reported in section 1. To compare with HMI, MDI and GONG images are upsampled as described in section 2.1.

In this section, we demonstrate the value added of using super-resolved magnetograms over their lower resolution counterparts. Specifically, we investigate small and large scale structures, homogenization properties, and temporal patterns of the super-resolved magnetic fields.

4.1 Homogenization

In Figure 5, the first (third) row shows a pixel-to-pixel correlation plot between target magnetograms and super-resolution output for the entire test set for MDI (GONG). The test set contains ≈ 25 (≈ 125) million pixels for MDI (GONG). The orange lines highlight regression lines between the output and target. To put these results into perspective, Figure 5 also shows

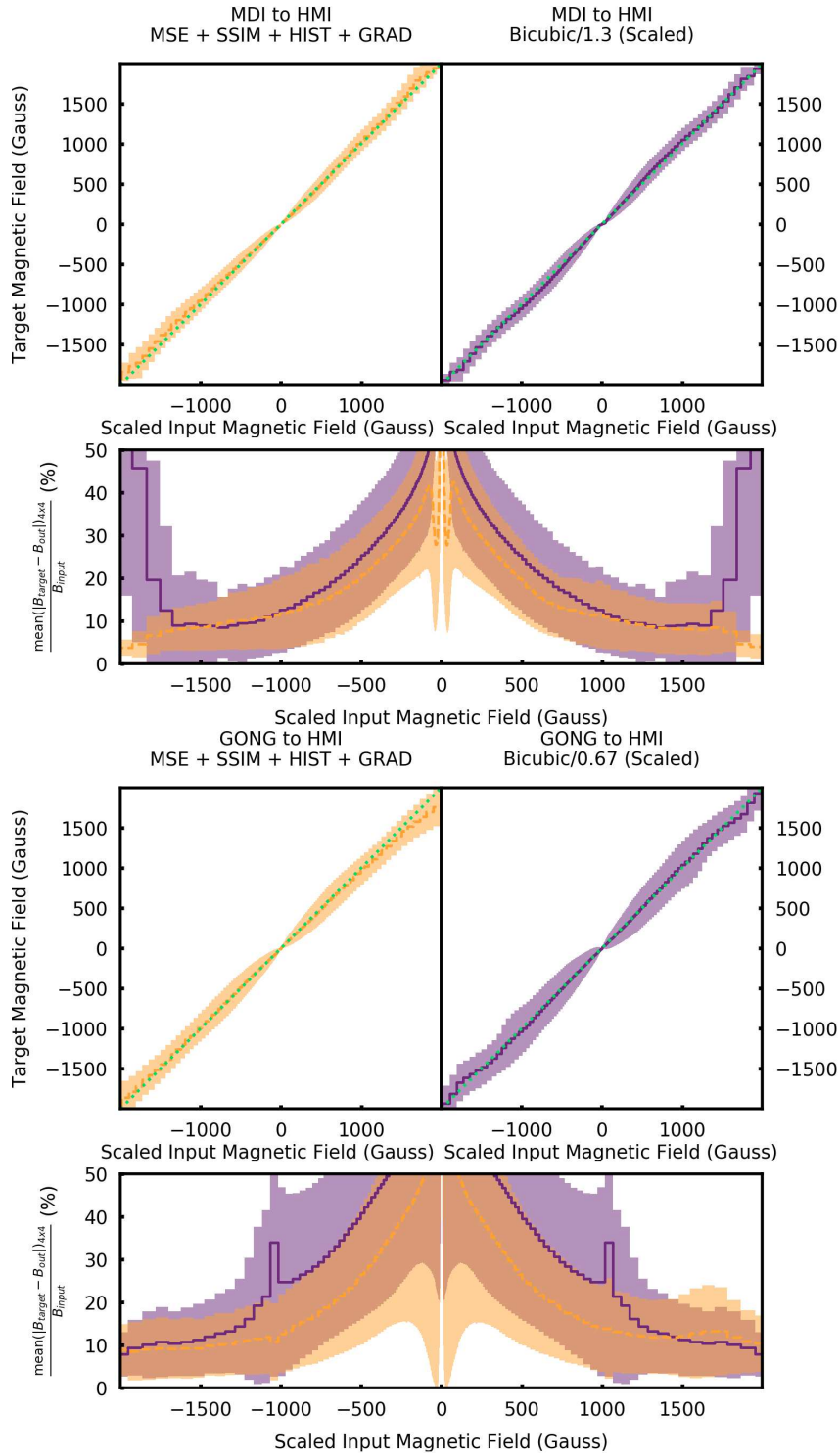


Figure 5: **2D Histogram of pixel-to-pixel comparison for test patches of the MDI → HMI (top two rows) and GONG → HMI (bottom two rows) transformation.** Top row: the dashed line is the average magnetic field output by super-resolutions models, where the average is taken over all values that fall within a given bin of the target magnetic field (x-axis). Bottom row: The orange- and purple-shaded areas are the standard deviations per x-axis value, for the super-resolution output and bicubic upsampling baseline, respectively. Bottom row: Relative residuals and relative variance of super-resolution and bicubic upsampling baseline, for MDI and GONG.

a comparison of the correlation between the bicubic upsampling baseline and the target magnetograms (purple graph on the right). Scatter plots (Figure 5) aligning with the 45° degree diagonal (top) or small residuals (bottom) indicate good cross-calibration. Our deep learning approach centers the correlation plots more on the 45° degree diagonal, thereby improving the cross-calibration between MDI (GONG) and HMI. This is clearly visible for GONG (Figure 5, third row) and less strongly observable for MDI (Figure 5, first row).

Figure 5 suggests great performance of both the bicubic baseline (purple) and our deep learning approach (orange) when only looking at the correlation plots as both comparisons are strongly centered around $x = y$. The second and fourth row in Figure 5 show further quantitative visualizations of both the **cross-calibration** and **super-resolution** showing that deep learning is more effective at performing super-resolution. We measure this improvement by investigating the average deviation of the output and target across a 4×4 pixel area compared to the corresponding low resolution pixel.

These results highlight one of the main challenges of finding suitable quantities to capture super-resolution. When looking simply at averages, it is easy to become overly confident and misrepresent the quality of results. Our work encourages benchmarking super-resolution techniques with a quantitative assessment that measures the reconstruction of small-scale structures of the magnetic field. We specifically want to encourage the presentation of results that are not perfect to accurately communicate the limitations of proposed results.

4.2 Large-scale structures

Table 1 replicates the quantitative assessment in Table 1 and 2 of reference¹⁶ and compares the Pearson correlation coefficient between super-resolved MDI and GONG magnetograms across different radial regions of the Sun and different values of the magnetic field. For both MDI and GONG, the Pearson coefficient is computed on our test set of magnetograms from March 2011. Our results show that our deep learning approach generates magnetograms that contains information present in HMI magnetograms, but not in their low-resolution counterparts. Across all radial regions and field values, the Pearson correlation coefficient between super-resolved and HMI magnetograms increases by 5 – 7% relative to the correlation between lower resolution and HMI magnetograms.

4.3 Small-scale structures

In Table 2, we compare statistics of the magnetic field between HMI and super-resolved MDI/GONG over kernel of various pixel sizes. We benchmark our results against our baseline approach (bicubic upsampling and linear rescaling). Our results show that the difference in gradient between HMI and deep learning output over small kernels (2 to 4 pixels) is 30% (for MDI) and 4% (for GONG) smaller than between HMI and the baseline outputs. Similar improvements are observed for extreme values of the magnetic field within kernel of different sizes. It confirms that our deep learning super-resolution captures details that are averaged out at lower resolution. Remarkably, improvements in small-scale patterns extend to structures of size larger than the upscaling factor (4).

4.4 Time-series of the line-of-sight magnetic field

To understand the quality of homogenization, in Figure 6, we study a $900'' \times 300''$ region (located in the northern hemisphere) during March 2011

We show the input MDI and GONG magnetograms for the last set of data in the test-set (14:24 UT and 21:24 UT respectively). We also plot the time-series of the signed and unsigned magnetic field extracted from the test set, and the respective super-resolution and target magnetograms.

Figure 6 shows that our deep learning approach both increases the resolution of the MDI and GONG data, and cross-calibrates the magnetic field observations with HMI. While the time-series of the *signed* magnetic field show good agreement between HMI and super-resolution MDI, plots of unsigned magnetic field amplify low-level magnetic field values (e.g. noise) that are not learned by a neural network. Prior to generating the super-resolution time-series in this panel, we include a pixel-wise noise term modeled by a Gaussian with a standard deviation of 4.7 Gauss. This is a representative noise-level near disk-center for 720s HMI data.¹⁰

We also show the ratios of HMI to super-resolution MDI (and HMI to super-resolution GONG). Noticeably, the deep learning model does not contain the 24-hour variability present in HMI data (See Methods; § 6.4). This is a known distortion in HMI that is not present in MDI nor GONG. The neural network considers this 24-hour variability as noise and averages it away.

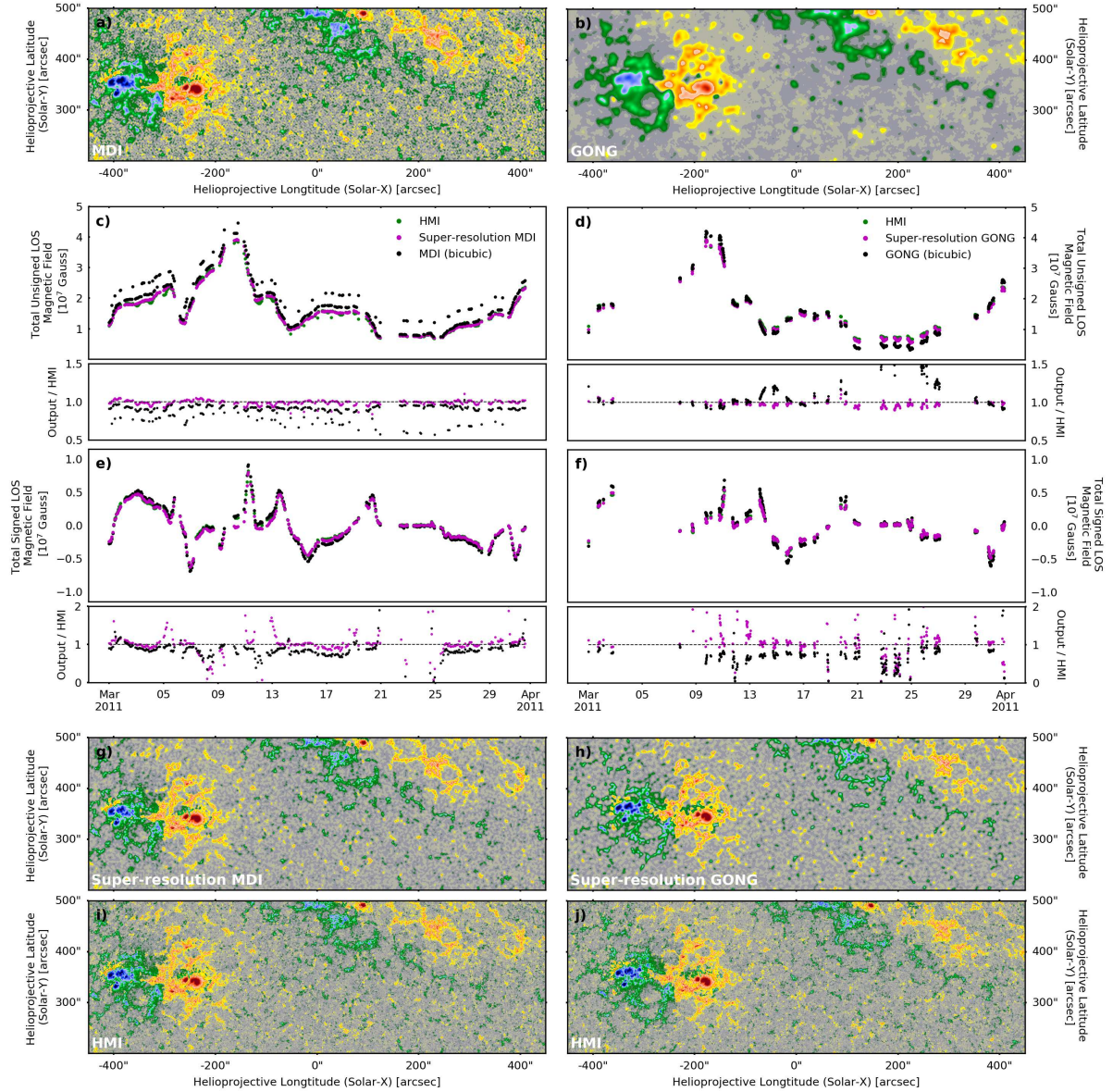


Figure 6: **Time-series of the line-of-sight signed and unsigned magnetic field with accompanying magnetograms (at a single time-step).** (a & b) Input ($900'' \times 300''$) magnetograms shown for *SoHO*/MDI and GONG at the last time-step in the test-set (14:24 UT and 21:24 UT respectively), on 31st March 2011. These magnetograms are super-resolved, with the respective output of the trained neural network shown in (g & h). These can be visually compared to the high-resolution target SDO/HMI images at the same time (i & j). (c & d) The total (summed) unsigned magnetic field calculated as a function of time from panels (a, g, i) and (b, h, j) respectively. Also shown is the ratio of the MDI-to-HMI (black) and super-resolution MDI-to-HMI (magenta) time-series. The expected value for cross-calibrated data is equal to unity; clear periodicity is observed in these time-series, consistent with the 24-hour periodicity seen in HMI data. In these plots, a small constant noise component of $\sigma = 4.7$ Gauss has been added to the unsigned time-series to account for de-noising by the neural network. This value was calculated from a single time-step in the training set by fitting a Gaussian to the histogram of pixel values < 10 Gauss. (e & f): The total signed magnetic field shown similarly to (c & d).

5 Conclusions

This paper shows that deep learning-based super-resolution successfully upsamples and homogenizes solar magnetic field images. We demonstrate the suitability of our approach by upsampling and cross-calibrating MDI (GONG) magnetograms to the characteristics of HMI. We show that a careful design of the loss function of the neural network improves the quality of the super-resolution application, a conclusion that may be applicable to any deep-learning super-resolution application in the physical sciences. In the loss function, we include penalty terms that constrain the distribution and gradients of magnetic field values to better match the ground truth. We further propose a set of quantities to evaluate the quality of (1) cross-calibration, and (2) super-resolution of magnetograms that can also be applied across disciplines.

An important contribution of this work is to offer a benchmark of measurements and methods for performance comparison of future machine learning-based approaches that cross-calibrate/super-resolve solar magnetic field images. We compare moments of the magnetic field at various spatial scales to capture how our technique super-resolves MDI and GONG magnetograms. Establishing benchmarks is necessary for the development and progress of deep learning approaches for solar magnetic field research. Furthermore, we invite these efforts to report the same proposed metrics for the same test month (March 2011), so that the community can transparently assess the state-of-the-art.

Future work will explore including temporal information into the deep learning architecture through multi-frame super-resolution.¹⁸ Moreover, it is essential to investigate how the pre-processing of the solar magnetograms, including feature alignment and re-projection in a common coordinates system affect the performances of a deep learning approach. Lastly, our current deep learning approach does not allow us to quantify how confident the model is about its predictions, particularly for periods where there is no ground truth. It is a promising avenue for future research to implement a probabilistic machine learning approach that would estimate the uncertainty of its super-resolved predictions, all the more as super-resolution is an ill-posed problem with many super-resolved images being consistent with the same low-resolution input.

Acknowledgments

This work was initiated at the 2019 NASA Frontier Development Lab (FDL) 2019. NASA FDL is a public-private partnership between NASA, the SETI Institute and private sector partners including Google Cloud, Intel, IBM, Lockheed Martin, NVIDIA, and Element AI.

The Solar Dynamics Observatory data is courtesy of NASA/SDO and the AIA, EVE, and HMI science teams.

This work utilizes data from the Michelson Doppler Imager (MDI) and the Helioseismic Magnetic Imager (HMI) distributed by the Joint Science Operations Center (JSOC) at Stanford University <http://jsoc.stanford.edu/>. We also use data from the National Solar Observatory's Integrated Synoptic Program, which is operated by the Association of Universities for Research in Astronomy, under a cooperative agreement with the National Science Foundation and with additional financial support from the National Oceanic and Atmospheric Administration, the National Aeronautics and Space Administration, and the United States Air Force. The GONG network of instruments is hosted by the Big Bear Solar Observatory, High Altitude Observatory, Learmonth Solar Observatory, Udaipur Solar Observatory, Instituto de Astrofísica de Canarias, and Cerro Tololo Interamerican Observatory.

This research was partially supported by NASA grants 80NSSC18K0671, 80NSSC19M0165, 80NSSC19K1207, and 80NSSC20K0602. A. Munoz-Jaramillo acknowledges support from IR grant 15-90480 from the Southwest Research Institute. P. J. Wright acknowledges support from NASA Contract NAS5-02139 (HMI) to Stanford University. A. G. Baydin is supported by EPSRC/MURI grant EP/N019474/1 and by Lawrence Berkeley National Lab. We thank

Sairam Sundaresan and Santiago Miret from Intel for their advice on this project. Our CNN training framework was built using PyTorch.¹⁹ Patch alignment during data preparation was performed using Scikit-image.²⁰ All our analysis was performed using the AstroPy,²¹ NumPy,²² SciPy,²³ SunPy,²⁴ and Matplotlib²⁵ packages.

Author contributions

A. Muñoz-Jaramillo defined the project during the 2019 Frontier Development Laboratory, ran the CNN training experiments, helped develop the metrics of performance, data pre-processing framework, visualized the results, and wrote the paper.

A. Jungbluth and X. Gitiaux developed the metrics used in the CNN loss function, helped develop the CNN training and data pre-processing frameworks, analyzed the results, and wrote the paper.

P. J. Wright defined the project during the 2019 Frontier Development Laboratory, performed the comparison between deep learning output and target instruments, analyzed the results, and wrote the paper.

C. Shneider and S. A. Maloney helped develop the CNN training and data pre-processing frameworks.

F. Kalaitzis defined the project during the 2019 Frontier Development Laboratory and planned the CNN training experiments using different quantitative assessments.

A. G. Baydin and Y. Gal defined the project during the 2019 Frontier Development Laboratory. M. Deudon helped develop the CNN training framework.

Data Availability

The code used in this work is available at <https://doi.org/10.5281/zenodo.3750372>. The *SDO/HMI*, and *SoHO/MDI* data are freely available from the Joint Science Operations Center (<http://jsoc.stanford.edu/>).

6 Methodology

6.1 Data

We use solar magnetograms from *NSO/GONG*, *SoHO/MDI*,^{26,27} and *SDO/HMI*.^{14,28,29}

The data is pre-processed according to the following three steps: (i) standardization of the Sun’s orientation by rotating solar north to image north; (ii) standardization of the detector resolution, and the size of the Sun in the image to the observed size at 1 Astronomical Unit (AU; the average Sun-Earth distance); (iii) alignment of features of magnetograms through registration and x-y shifting.

To train our super-resolution architecture, we leverage overlapping observation periods between MDI and HMI (2010 – 2011) and between GONG and HMI (2010 – 2019), which provides us with $\sim 9,000$ ($\sim 19,000$) MDI-HMI (GONG-HMI) magnetogram pairs. We split the data into training/validation/test sets by randomly allocating ten months to the training set, one month to the validation set, and one month to the test set for each overlapping year.

In the case of GONG-HMI, we only use even years (2010, 2012, 2014, 2016, & 2018) for this work to keep the data volume manageable. The test set comprises magnetograms taken at a 96 minute cadence for MDI, and a 10 minute cadence for GONG. Across all experiments we choose June 2010 and March 2011 as our test month. Table 1 and 2 use only March 2011, which is a more conservative setting particularly GONG, since the whole 2011 year is not in the test set.

Each full-disk magnetogram is split into 1024 patches of size 32×32 pixels for the low-resolution input and 128×128 pixels for the high-resolution target. We augment data through

random polarity flips and North-South, East-West reflections. This ensures that our sample comprised magnetograms that would appear in a different solar cycle.

As the structure of the global magnetic field changes on significantly longer time-scales than the small-scale features, by splitting each magnetogram into 1024 patches, we directly expose the neural network to the small-scale magnetic field configuration only. As this evolves over timescales of hours, we do not need to provide a time buffer between the training, validation, and test months. For other tasks, the slow evolution of the global magnetic field could unintentionally leak into the test set if a sufficient buffer isn't present.

Our results indicate that the effectiveness of the neural network to cross-calibrate and super-resolve magnetic field is sensitive to the signal-to-noise ratio within a magnetogram. The signal-to-noise ratio is affected by (1) the strength of the magnetic field itself, and (2) the proximity to the solar limb. HMI's noise level is 15.³⁰ Figure 9 compares the Pearson correlation metric calculated for super-resolved MDI magnetograms as a function of patch location and magnetic field value. The grey shaded histograms were calculated for patches across the full solar disk, while the blue shaded histograms were calculated for patches that lie within 90% of the radius of the solar disk. In addition, we compare the Pearson correlation for all patches (left column), and those that have an average unsigned field larger than 15 Gauss (right column). Looking at all magnetic field values, we can see that the distribution of Pearson correlation coefficients shows two peaks around 0.25 and 0.75 when patches across the entire solar disk are considered

(9, left column, grey histogram). Discarding patches that lie outside of the central 90% of the radius of the solar disk removes the double peak and shifts the distribution to be asymmetrically centered around 0.8 (9, left column, blue histogram). When we also disregard patches that show magnetic field strengths close to HMI's noise level, we see that the distribution of Pearson correlation coefficients becomes substantially narrower and is more symmetrically centered around 0.8. This observation supports our finding that magnetogram patches with field strengths around the noise level are harder to align and less reliable for the neural network to learn. In addition, near the solar limb the magnetic field intensity weakens due to projection effects and a reduction of the effective resolution of the instruments. In the post-mortem evaluation of our results, we therefore focus on patches that lie within 90% of the radius of the solar disk, unless otherwise specified.

	Training/Validation	Testing
MDI → HMI	April 2010 to April 2011, excluding test set	June 2010, March 2011
GONG → HMI	2012, 2014, 2016, 2018, excluding test set	March 2011, April 2012 December 2014, February 2016 November 2018

Table 3: **Data split for model training, validation, and test.**

6.2 Neural Network Architecture

The deep learning model used in this work was adapted from the HighRes-net model¹⁸² (see Figure 7).

The input data consists of a magnetogram and a location channel. The location channel captures the distance from disk center and gives the network information necessary for estimating projection and foreshortening effects at the solar limb. The data is encoded into 64 channels through a series of convolution operations. In the decoding operation, each patch of 32×32 pixels is increased through bilinear upsampling to a patch of 128×128 pixels.

²<https://github.com/ElementAI/HighRes-net>

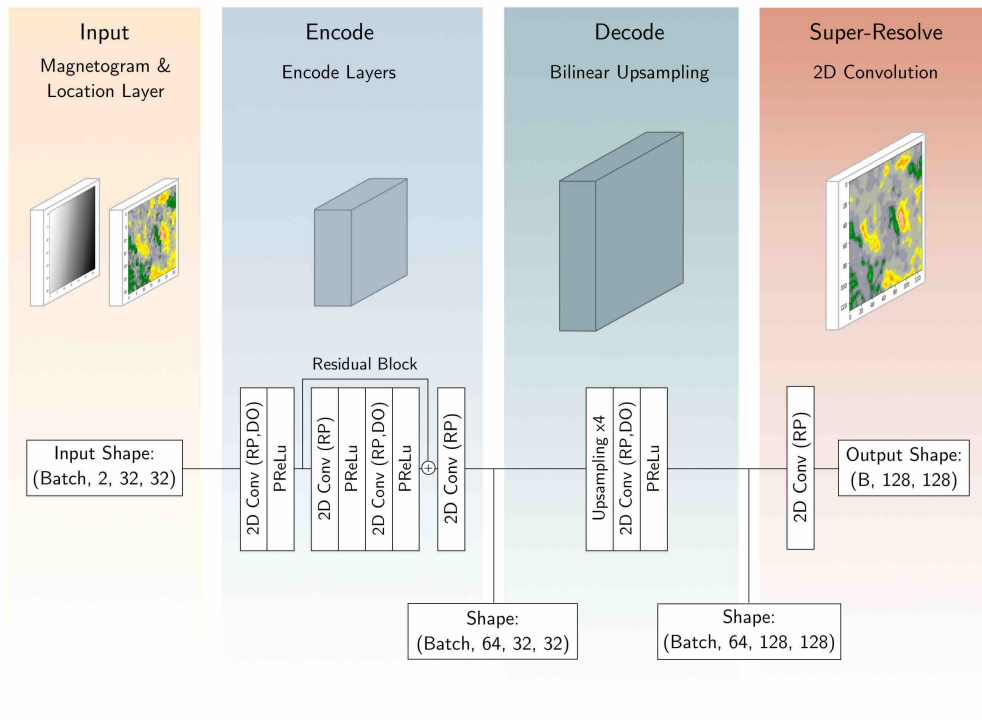


Figure 7: **Diagram of the deep learning model used in this work.** The input data consists of a magnetogram and a location channel, each of size 32×32 pixels. The main operations of the model consist of an encoder and a decoder. Through bi-linear upsampling by a factor of 4, the size of the image is increased from 32×32 to 128×128 pixels per patch. RP and DO denote whether a convolutional layer was trained with reflection padding or Monte-Carlo drop-out respectively.

6.3 Optimization Penalty Terms

Training a neural network involves the minimization of an objective function that quantifies how well the transformation of the input matches the target. This objective function is typically referred to as the *loss* function (\mathcal{L}). As super-resolution is an ill-posed problem (i.e., a one-to-many operation), multiple super-resolution outputs can explain the same low-resolution input. For scientifically useful applications of super-resolution, the model output should capture the physical properties of the target, and cannot just be perceptually convincing. To better respect the physical properties of the target high-resolution magnetograms, we construct a loss function that combines four terms, each of which aims to capture a different aspect of what makes a magnetogram physically plausible:

$$\mathcal{L} = \mathcal{L}_{l_2} + w_{grad}\mathcal{L}_{grad} + w_{hist}\mathcal{L}_{hist} + w_{ssim}\mathcal{L}_{ssim}, \quad (8)$$

- \mathcal{L}_{l_2} penalizes the mean squared error between the super-resolved output and the target, and captures pixel-based differences in signed flux;
- \mathcal{L}_{grad} penalizes the mean squared difference between pixel gradients of the super-resolved output and the target. The gradients are approximated using a Sobel operator³¹. \mathcal{L}_{grad} aims to capture the gradients present at the boundaries between positive and negative polarities;
- \mathcal{L}_{hist} penalizes an approximation of the total variation distance between magnetic field distributions of the output and target magnetograms. For that, we calculate a differentiable pixel histogram using the method described in³²;
- \mathcal{L}_{ssim} measures the structural similarity between regions surrounding each pixel, including similarities in contrast, unsigned flux and variance.

We chose loss-weights w to scale each term’s contribution to the same order of magnitude as the \mathcal{L}_{l_2} term. We refined their values by conducting a partial grid search that finds the values for w_{grad} , w_{hist} and w_{ssim} that minimize $\mathcal{L}_2 + w_{grad}\mathcal{L}_{grad}$, $\mathcal{L}_2 + w_{hist}\mathcal{L}_{hist}$ and $\mathcal{L}_2 + w_{ssim}\mathcal{L}_{ssim}$, respectively. Then, we used the weights resulting from this grid search to minimize the loss function (8). Table 4 shows the optimal loss coefficients for MDI and GONG. These coefficients are different due to the difference in the systematic properties of the GONG and MDI instruments, as well as the most ambitious upscaling target when using GONG as a source.

	MSE	Gradient	Histogram	SSIM
Optimal MDI Run	1	5	1e-5	5e-4
Optimal GONG Run	1	5	1e-6	5e-5

Table 4: **Optimal training loss coefficients for MDI and GONG super-resolution and calibration**

6.4 24 hour variability in HMI data

As an extension to Figure 6, in Figure 8 we show the HMI and Super-resolution MDI time-series along with the ratio of time-series, and the radial velocity of HMI. The observed oscillations arise from a Doppler shift in the spectral line due to the orbital variation of the spacecraft.³³

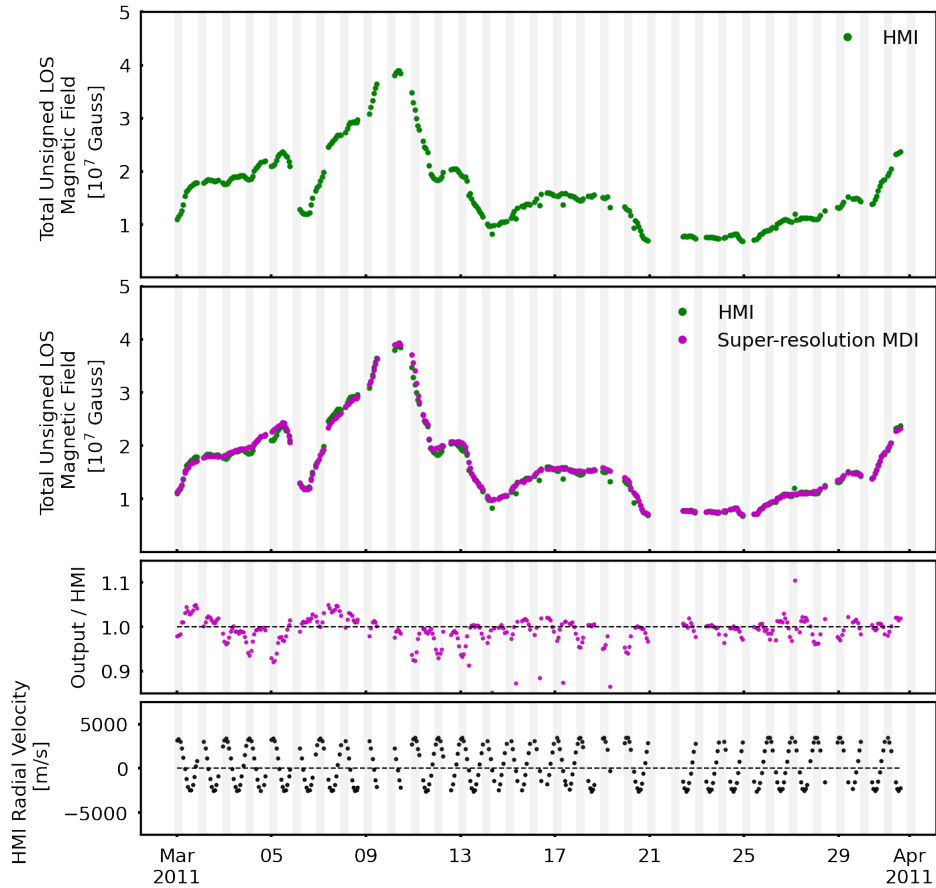


Figure 8: **A demonstration of the HMI radial velocity leaking into HMI LOS magnetic field data.** Top two panels show the total unsigned magnetic field from HMI and the Super-resolution MDI. Bottom two panels show the ratio of Super-resolution output to HMI, and the HMI radial velocity where positive values are away from the Sun (bottom panel). In each panel, vertical gray bars are shown with a 24-hour periodicity starting on 1st March 2011. It is seen that this periodicity leaks into the HMI data, but is not observed in the cross-calibrated and Super-resolution MDI data (as would be expected).

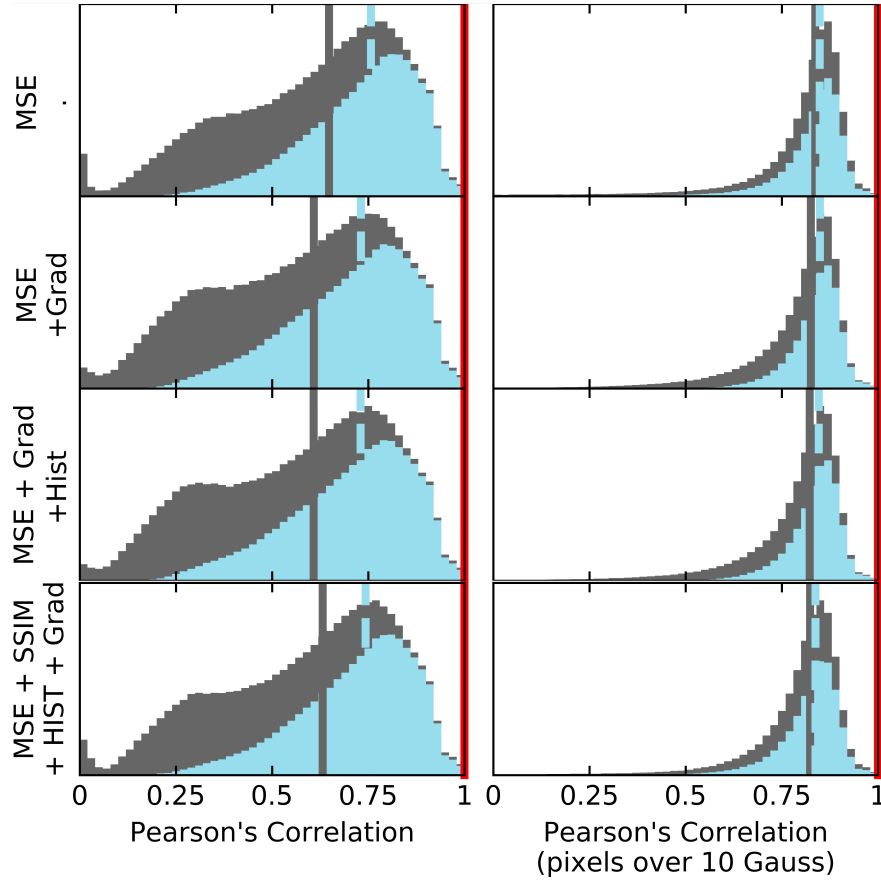


Figure 9: **Quantitative comparison of the performance of different loss functions trained on MDI magnetograms.** All loss functions are based on the mean squared error term, plus up to three additional penalty terms. The grey shaded histograms correspond to calculations performed across the full solar disk. The blue histograms were calculated for patches that fall within 90% of the solar disk radius. The first column shows the Pearson correlation calculated for all magnetic field values, while the second column shows the Pearson correlation calculated for magnetic fields with an absolute value above 10 Gauss. In the ideal case, the Pearson correlation is 1 (indicated by the red line).

References

- [1] Shukla, A., Merugu, S. & Jain, K. *A Technical Review on Image Super-Resolution Techniques*, 543–565 (Springer Singapore, 2020). URL https://doi.org/10.1007/978-981-15-3125-5_54.
- [2] Yang, W. *et al.* Deep learning for single image super-resolution: A brief review. *IEEE Transactions on Multimedia* **21**, 3106–3121 (2019). URL <http://dx.doi.org/10.1109/TMM.2019.2919431>.
- [3] Dahl, R., Norouzi, M. & Shlens, J. Pixel Recursive Super Resolution. *arXiv e-prints* arXiv:1702.00783 (2017). 1702.00783.
- [4] Wang, Z., Chen, J. & Hoi, S. C. H. Deep learning for image super-resolution: A survey (2020). 1902.06068.
- [5] Hathaway, D. H. The Solar Cycle. *Living Reviews in Solar Physics* **7**, 1 (2010).
- [6] Linker, J. A. *et al.* Magnetohydrodynamic modeling of the solar corona during whole sun month. *Journal of Geophysical Research: Space Physics (1978–2012)* **104**, 9809–9830 (1999). URL <https://doi.org/10.1029/1998JA900159>.
- [7] Tóth, G. *et al.* Space weather modeling framework: A new tool for the space science community. *Journal of Geophysical Research: Space Physics (1978–2012)* **110** (2005). URL <https://doi.org/10.1029/2005JA011126>.
- [8] Borrero, J. M. & Ichimoto, K. Magnetic Structure of Sunspots. *Living Reviews in Solar Physics* **8**, 4 (2011). 1109.4412.
- [9] Díaz Baso, C. J. & Asensio Ramos, A. Enhancing SDO/HMI images using deep learning. *Astronomy & Astrophysics* **614**, A5 (2018). 1706.02933.
- [10] 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. *Solar Physics* **279**, 295–316 (2012).
- [11] Riley, P. *et al.* A Multi-Observatory Inter-Comparison of Line-of-Sight Synoptic Solar Magnetograms. *Solar Physics* **289**, 769–792 (2014).
- [12] Virtanen, I. & Mursula, K. Photospheric and coronal magnetic fields in six magnetographs. III. Photospheric and coronal magnetic fields in 1974–2017. *Astronomy and Astrophysics* **626**, A67 (2019).
- [13] Rahman, S. *et al.* Super-resolution of SDO/HMI Magnetograms Using Novel Deep Learning Methods. *Astrophys. J.* **897**, L32 (2020).
- [14] Scherrer, P. H. *et al.* The Helioseismic and Magnetic Imager (HMI) Investigation for the Solar Dynamics Observatory (SDO). *Solar Physics* **275**, 207–227 (2012).
- [15] Harvey, J. W., Hill, F., Kennedy, J. R., Leibacher, J. W. & Livingston, W. C. The Global Oscillation Network Group (GONG). *Advances in Space Research* **8**, 117–120 (1988).
- [16] 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. *Solar Physics* **279**, 295–316 (2012).
- [17] Forsyth, D. A. & Ponce, J. *Computer vision: a modern approach* (Prentice Hall Professional Technical Reference, 2002).

- [18] Deudon, M. *et al.* Highres-net: Recursive fusion for multi-frame super-resolution of satellite imagery (2020). 2002.06460.
- [19] Paszke, A. *et al.* Pytorch: An imperative style, high-performance deep learning library. In Wallach, H. *et al.* (eds.) *Advances in Neural Information Processing Systems 32*, 8024–8035 (Curran Associates, Inc., 2019). URL <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>.
- [20] van der Walt, S. *et al.* scikit-image: image processing in Python. *PeerJ* **2**, e453 (2014). URL <https://doi.org/10.7717/peerj.453>.
- [21] Astropy Collaboration *et al.* The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. *Astronomy J.* **156**, 123 (2018). 1801.02634.
- [22] Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).
- [23] Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17**, 261–272 (2020).
- [24] The SunPy Community *et al.* The sunpy project: Open source development and status of the version 1.0 core package. *The Astrophysical Journal* **890**, 68– (2020). URL <https://iopscience.iop.org/article/10.3847/1538-4357/ab4f7a>.
- [25] Hunter, J. D. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering* **9**, 90–95 (2007).
- [26] Domingo, V., Fleck, B. & Poland, A. I. The SOHO Mission: an Overview. *Solar Physics* **162**, 1–37 (1995).
- [27] Scherrer, P. H. *et al.* The Solar Oscillations Investigation - Michelson Doppler Imager. *Solar Physics* **162**, 129–188 (1995).
- [28] Pesnell, W. D., Thompson, B. J. & Chamberlin, P. C. The Solar Dynamics Observatory (SDO). *Solar Physics* **275**, 3–15 (2012).
- [29] Schou, J. *et al.* Design and Ground Calibration of the Helioseismic and Magnetic Imager (HMI) Instrument on the Solar Dynamics Observatory (SDO). *Solar Physics* **275**, 229–259 (2012).
- [30] Hoeksema, J. T. *et al.* The helioseismic and magnetic imager (hmi) vector magnetic field pipeline: overview and performance. *Solar Physics* **289**, 3483–3530 (2014).
- [31] Sobel, I. An isotropic 3x3 image gradient operator. *Presentation at Stanford A.I. Project* (1968).
- [32] Wang, Z., Li, H., Ouyang, W. & Wang, X. Learnable histogram: Statistical context features for deep neural networks (2018). 1804.09398.
- [33] Couvidat, S. *et al.* Observables Processing for the Helioseismic and Magnetic Imager Instrument on the Solar Dynamics Observatory. *Sol. Phys.* **291**, 1887–1938 (2016). 1606.02368.