Aberystwyth University

Reliability of AI-generated magnetograms from only EUV images

Liu, Jiajia; Wang, Yimin; Huang, Xin; Korsós, Marianna B.; Jiang, Ye; Wang, Yuming; Erdélyi, Robert

Published in:
Nature Astronomy

DOI:
[10.1038/s41550-021-01310-6](https://doi.org/10.1038/s41550-021-01310-6)

Publication date:
2021

Citation for published version (APA):

Liu, J., Wang, Y., Huang, X., Korsós, M. B., Jiang, Y., Wang, Y., & Erdélyi, R. (2021). Reliability of AI-generated magnetograms from only EUV images. *Nature Astronomy*, 5(2), 108-110. <https://doi.org/10.1038/s41550-021-01310-6>

Document License CC BY-NC

General rights

Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400
email: is@aber.ac.uk

Supplementary information

**Reliability of AI-generated magnetograms
from only EUV images**

In the format provided by the
authors and unedited

Supplementary Information:

Reliability of AI-generated magnetograms from only EUV images

Jiajia Liu^{1,2,*}, Yimin Wang^{1,3,*}, Xin Huang⁴, Marianna B. Korsós⁵, Ye Jiang⁶, Yuming Wang⁷, Robert Erdélyi^{1,8,9,+}

¹*Solar Physics and Space Plasma Research Centre (SP2RC), School of Mathematics and Statistics, The University of Sheffield, Sheffield, S3 7RH, UK*

²*Astrophysics Research Centre, School of Mathematics and Physics, Queen's University, Belfast, BT7 1NN*

³*School of Electrical Engineering, University of Jinan, Jinan 250022, China*

⁴*Key Laboratory of Solar Activity, National Astronomical Observatories of Chinese Academy of Sciences, Beijing, China*

⁵*Department of Physics, Aberystwyth University, Penglais, Aberystwyth SY23 3BZ, UK*

⁶*Department of Computer Science, The University of Sheffield, Sheffield S1 4DP, UK*

⁷*CAS Key Laboratory of Geospace Environment, Department of Geophysics and Planetary Sciences, University of Science and Technology of China, Hefei, Anhui 230026, China*

⁸*Department of Astronomy, Eötvös Loránd University, Budapest, Pázmány P. sétány 1/A, H-1117, Hungary*

⁹*Gyula Bay Zoltán Solar Observatory (GSO), Hungarian Solar Physics Foundation (HSPF), Petőfi tér 3., Gyula, H-5700, Hungary*

* *These authors contribute equally to this Matters Arising*

+ *Corresponding Author, email: robertus@sheffield.ac.uk*

Data and Method

In this section, we comprehensively recap the method used¹. Neural networks, especially convolutional neural networks (CNNs), have proved their high potential in computer vision in various fields such as autonomous vehicles, medical imaging and remote sensing. Goodfellow et al. 2014 introduced the generative adversarial networks (GANs) for generating high quality images with CNNs as the workhorse behind it². The structure of GANs is composed of two adversarial models: a generator (G) that captures the data distribution and aims to produce more and more realistic images, and a discriminator (D) that distinguishes whether an image is real or generated by G . The training process is performed through an adversarial process, in which G and D are trained in turn.

In contrast to GANs, where new images are generated from noise, Isola et al. 2017 extended GANs to a conditional model, called *pix2pix*, where the authors conditioned on an input image for

generating a corresponding output image³. KPL19 applied the above method *pix2pix* for the image translation of photospheric magnetograms from SDO/AIA 304 Å images, to build a model in order to predict the farside magnetograms of the Sun. More details of the method could be found in the above mentioned literature.

To independently validate the results in KPL19, we followed the steps described therein. Full-disk EUV observations of the Sun from the SDO/AIA 304 Å passband observed from 1 January 2011 to 31 December 2017 were downloaded with a cadence of 12 hours (at 00 UT and 12 UT each day). These data were then processed with the standard *aia_prep.pro* function included in the SolarSoft packages (*SSW*) and converted to 8-bit images with sizes of 1024×1024 pix². The natural logarithms of the original data have also been normalised to -1 to 1 with saturation values of $\ln 15$ (lower limit) and $\ln 600$ (upper limit) before the conversion. Similar processes have been applied to the simultaneous SDO/HMI full-disk line-of-sight (LOS) magnetograms, with the lower and upper saturation limits of ± 100 G for the rescaling and normalisation, i.e. pixels with magnetic field strength above (below) 100 G (-100 G) are set to be 100 G (-100 G) before the data being normalised to the range of -1 to 1.

In total, a dataset containing 4970 pairs of SDO/AIA 304 Å and SDO/HMI photospheric magnetic field images were built (full list of data used available at https://github.com/yiminking/pix2pix_EUV2HMI_datasets). We note that this dataset contains 2 pairs of instances less than those of KPL19, because HMI observations several hours around 2014-12-29 00 UT and AIA 304 Å observations several hours around 2017-04-30 12 UT were corrupted. These images were then used as the input to the code <https://github.com/tykimos/SolarMagGAN> provided by KPL19 to independently validate the model described therein. The same as KPL19, we used images observed in September and October in each year (825 pairs in total) for the test purpose and the rest (4145 pairs) for the training purpose.

KPL19 trained the generator network and the discriminator network for 500,000 iterations (~ 120 epochs) due to the limitation of their computational capability. They were not able to demonstrate whether the model would perform better if more iterations had been run. Instead, we trained the networks for a much larger number of iterations (1,500,000) and found the best model was built at 505,000 iterations, corresponding to ~ 122 epochs.

Full-disk Parameters

KPL19 evaluated the correlation of the full-disk total unsigned magnetic flux (TUMF) between the rescaled (i.e. original but after rescaling with saturation limits of ± 100 G) and the generated magnetograms and found the correlation was as high as 0.97. The pixel-to-pixel (with 8×8 bins) cross correlation (P2P CC) was 0.77 and the normalised mean square error (NMSE) was found to be 0.053. We also evaluated the above parameters. The correlations of the TUMFs between the rescaled and the generated magnetograms are found to be 0.65, which is not as good as that claimed

in KPL19, but still showing a strong evidence that the model was successful in finding the TUMFs using observations from the SDO/AIA 304 Å passband. We note here that, pixels with absolute magnetic field strength less than 10 G were set to be 0 in the evaluations, following KPL19.

We evaluate the original P2P CC (with no bins), as well as the CCs with 8×8 bins using the average, maximum and median values of each bin. The resulted P2P CC is then found to be 0.09 ± 0.05 . The CCs with 8×8 bins using the average, maximum and median values are found to be 0.20, 0.10 and 0.16, respectively. These CCs are all considerably lower than the one in KPL19 (0.77 using the average value of each bin) and indicate a poor performance of the model in terms of comparing the generated and rescaled magnetograms pixel by pixel. Further, we have found the NMSE as 1.97 ± 0.33 , indicating that the difference of the generated and the rescaled magnetic field strengths is on average almost twice of the rescaled values.

KPL19 did not evaluate the correlation of the net magnetic flux (NMF) between the rescaled and the generated magnetograms. However, the NMF has been widely suggested as one of the key characteristics of solar active regions⁴ that could, for example, measure whether the magnetic field is mostly open or not⁵. The NMF has also been suggested to be potentially important in solar eruptive events including CMEs and flares^{6–8}. We have found the correlation of the NMF with an average value of 0.20. This, again, confirms our observation in the main text that the model fails in reconstructing the relative relations between the positive and negative polarities.

Active Region Parameters

The automatic detection system^{9,10} is based on a region-growing method and contains four functional modules. The first module is to find kernel pixels of the active regions by applying the intensity segmentation method on the magnetograms. The intensity threshold level in this module has been chosen as ± 80 G considering the saturation limits were set as ± 100 G during the image processing. The second module is designed to remove non-AR kernel pixels. The region-growing operation is performed from the detected kernel pixels in the third function. The lower limit for the region-growing being stopped was set to be ± 30 G, which is 3 times the noise level of SDO/HMI observations¹¹. The last module performs the morphological closing operations to merge neighbouring active regions.

Leaving out extra active regions in the generated magnetograms, which are not in the real observations, we define the detection rate r_d as N_g/N_r . Here, N_g is the number of active regions in each generated magnetogram which are also present in the observations, and N_r is the total number of active regions in each observed (rescaled) magnetogram. Supplementary Figure 3(a) shows the distribution of the detection rate. It is shown that, on average, the model reproduces less than half (0.46 ± 0.28 , dashed vertical line) of the active regions in each observed magnetograms. In about 61% cases, less than half are reproduced. Let us now define the effective radius of an active region R_e as the radius of a circle that has the same area as the active region, D_a as the

distance between the area-weighted centres and D_f as the distance between the magnetic flux-weighted centres of two corresponding active regions with one in the generated and the other in the rescaled magnetogram. Supplementary Figure 3(b) and (c) are the distributions of D_a/R_e and D_f/R_e , respectively. On average, an active region in the generated magnetograms is offset by a factor of 0.21 ($\sim 1.3^\circ$ in heliographic coordinates) away from the real location of the active region in the rescaled magnetograms.

Comparisons of some key parameters of the detected active regions are presented in Supplementary Figure 4. Blue crosses (black dots) are the areas of the positive (negative) polarities of the active regions in the rescaled (x -axis) and the generated (y -axis) magnetograms, respectively. The blue (black) line is the linear fit of the blue crosses (black dots) with the intercept forced to be zero. The slopes of the lines are 0.75 and 0.91, respectively. The corresponding CCs are 0.79 and 0.77 respectively for the positive and negative polarities. However, the dots are quite scattered. The R2 score is defined by

$$R2 = 1 - \Sigma_i(y_i - \bar{y})^2 / \Sigma_i(y_i - f_i)^2, \quad (1)$$

where y_i is the i -th observed value, $\bar{y}$ the average of all the observed values and f_i the i -th predicted (generated) value. The R2 scores for the positive and negative polarities in active regions are found to be 0.59 and 0.58, respectively. The above values indicate a fairly good performance of the model in predicting the areas of the active regions.

The CC, R2 score and slope of the linear fit between the TUMF of the rescaled and generated active regions (Supplementary Figure 4b) are 0.81, 0.63 and 0.78, respectively. This, again, suggests a good performance, though the model, on average, underestimates the TUMF of the active regions. The CC, R2 score and slope of the linear fit between the NMF of the observed and generated active regions (Supplementary Figure 4c) are 0.50, -0.09 and 1.08, respectively. The percentage of instances where the model yields an opposite sign (OS) of the NMF to the true ones from the rescaled magnetograms is estimated to be as high as 32.6%. This means that the model fails in reproducing the NMF of the active regions. It has also been found that the model performance is not good in reproducing the total number of the PILs (panel d) with an R2 score of 0.36, the total length of the PILs (panel e) with an R2 score of 0.32 and the average magnetic gradient across the PILs (panel f) with an R2 score of 0.22.

Potential Data Leakage

Following KPL19, we used images observed in September and October in each year as the test set and the rest as the training set. However, the Sun rotates at a period of slightly less than one month (~ 27.3 days). This means that some information from images observed in September (October)

could be contained in observations in August (November). For example, both of the SDO/AIA 304 Å and HMI photospheric magnetic field observations of active region NOAA 12685 on 27 October 2017 appear to be very similar to the active region located at almost the same position (with no NOAA number) on 23 November 2017. Thus, there is a concern that there might be some potential data leakage in the process proposed by KPL19.

To examine the above conjecture, we have built another model using the same process but with images observed in October and November as the test set, images observed in September and December omitted and the rest as the training set. We demonstrate that this new model suffers minimal data leakage, considering that only very few active regions exist more than two Carrington rotations^{12, 13}. Since little or no information is now shared between the training and test sets, the new model should perform worse than the previous model. In the new model, 3324 and 835 pairs of observations have been used as the training and test sets, respectively. The CC of the full-disk TUMF drops to 0.03 (compared to 0.69 in the previous model). While the CC of the full-disk NMF drops to -0.09 (compared to 0.15 in the previous model). The NMSE rises to 2.54 from 1.97. The average detection rate of active regions drops to 0.41, compared 0.46 in the previous model. The performance of these two models is comparable on the evaluated active region parameters in the previous section, except that the CC and R2 score of the active region NMF drop to 0.23 and -0.51, compared to 0.50 and -0.09 in the previous model. All the above indicators and measures suggest a potential data leakage problem in the model proposed by KPL19.

Supplementary Figure 1: **Comparison of the total unsigned magnetic flux (TUMF) and the net magnetic flux (NMF)**. Panels (a) and (b) are the relationships between the TUMF and NMF respectively of the original (x-axes) and rescaled (y-axes) magnetic magnetograms. Black and blue dots are with saturation limits of ± 100 G and ± 625 G. Blue and black lines are the corresponding linear fits. CC stands for cross correlation, R2 is the R-square score (Supplementary Eq. 1) and OS is the abbreviation for the percentage of instances where the model yields an opposite sign of the NMF to the true ones.

Supplementary Figure 2: **Active regions detected from the rescaled and AI-generated magnetograms.** The original photospheric magnetic field observed by SDO/HMI is shown in panel (a) and the AI-generated magnetic field in panel (b) at 00:00 UT on September 28 2011. Blue curves are the edges of the active regions and vermilion curves are their corresponding polarity inversion lines.

Supplementary Figure 3: **Distributions of the detection rate and distances of centres of active regions in the rescaled and generated magnetograms.** (a) is the distribution of the active region detection rate. The vertical dashed line marks a detection rate of 0.5, The orange label is the average detection rate with standard deviation. The black label on the left is the percentage of the total cases in the test set where less than half of the active regions were reconstructed in the model. (b) and (c) are the distributions of the distance between the area-weighted centres and the magnetic flux-weighted centres. Coloured labels are their corresponding average values and standard deviations.

Supplementary Figure 4: **Correlations of active region parameters.** X- and Y-axes in each panel are for the rescaled and AI-generated magnetograms, respectively. Blue crosses (black dots) in panel (a) are the areas of the positive (negative) polarities in the generated and observed active regions. The blue and black lines are the corresponding linear fits. Black dots in panels (b) to (f) are the TUMF, NMF, the number of PILs, the total length of PILs and the average gradient across the PILs in the active regions detected from the rescaled and generated magnetograms. Purple lines are their corresponding linear fits. CC stands for cross correlation, R2 is the R-square score (Supplementary Eq. 1) and OS is the abbreviation for the percentage of instances where the model yields an opposite sign of the NMF to the true ones.

Supplementary References

1. Kim, T. *et al.* Solar farside magnetograms from deep learning analysis of STEREO/EUVI data. *Nat. Astron.* **3**, 397–400 (2019).
2. Goodfellow, I. *et al.* Generative adversarial nets. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D. & Weinberger, K. Q. (eds.) *Adv. in Neu. Inf. Proc. Sys.* 27, 2672–2680 (Curran Associates, Inc., 2014).
3. Isola, P., Zhu, J., Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 5967–5976 (2017).
4. Leka, K. D. & Barnes, G. Photospheric Magnetic Field Properties of Flaring versus Flare-quiet Active Regions. I. Data, General Approach, and Sample Results. *Astrophys. J.* **595**, 1277–1295 (2003).
5. Sun, X. *et al.* Why Is the Great Solar Active Region 12192 Flare-rich but CME-poor? *Astrophys. J. Lett.* **804**, L28–L33 (2015).
6. Zhang, J. & Wang, J. Are Homologous Flare-Coronal Mass Ejection Events Triggered by Moving Magnetic Features? *Aastrophys. J. Lett.* **566**, L117–L120 (2002).
7. Tian, L. R. & Liu, Y. Relationship between decrease of net magnetic flux and solar major events. *Astron. Astrophys.* **406**, 337–344 (2003).
8. Wang, W. *et al.* Buildup of a highly twisted magnetic flux rope during a solar eruption. *Nat. Comm.* **8**, 1330–1339 (2017).
9. Wang, Y. & Zhang, J. A Statistical Study of Solar Active Regions That Produce Extremely Fast Coronal Mass Ejections. *Astrophys. J.* **680**, 1516–1522 (2008).
10. Zhang, J., Wang, Y. & Liu, Y. Statistical Properties of Solar Active Regions Obtained from an Automatic Detection System and the Computational Biases. *Astrophys. J.* **723**, 1006–1018 (2010).
11. Liu, Y. *et al.* Comparison of Line-of-Sight Magnetograms Taken by the Solar Dynamics Observatory/Heliioseismic and Magnetic Imager and Solar and Heliospheric Observatory/Michelson Doppler Imager. *Sol. Phys.* **279**, 295–316 (2012).
12. Schrijver, C. J. & Zwaan, C. *Solar and Stellar Magnetic Activity*. Cambridge Astrophysics (Cambridge University Press, 2000).
13. van Driel-Gesztelyi, L. & Green, L. M. Evolution of Active Regions. *Living Rev. in Sol. Phys.* **12**, 1–98 (2015).