

Magnetogram-to-Magnetogram: Generative Forecasting of Solar Evolution

Francesco Pio Ramunno^{*1, 2}, Hyun-Jin Jeong³, Stefan Hackstein¹, André Csillaghy¹, Svyatoslav Voloshynovskiy², Manolis K. Georgoulis^{4, 5}

¹*Institute for Data Science, University of Applied Sciences North Western Switzerland (FHNW), 5210 Windisch, Switzerland*

²*Department of Computer Science, University of Geneva, 1211 Geneva, Switzerland*

³*Department of Astronomy and Space Science, Kyung Hee University, Yongin, 17104, Republic of Korea*

⁴*Johns Hopkins University Applied Physics Laboratory Laurel, MD 20375, USA*

⁵*Research Center for Astronomy and Applied Mathematics of the Academy of Athens, 11527 Athens, Greece*

Investigating the solar magnetic field is crucial to understand the physical processes in the solar interior as well as their effects on the interplanetary environment. We introduce a novel method to predict the evolution of the solar line-of-sight (LoS) magnetogram using image-to-image translation with Denoising Diffusion Probabilistic Models (DDPMs). Our approach combines "computer science metrics" for image quality and "physics metrics" for physical accuracy to evaluate model performance. The results indicate that DDPMs are effective in maintaining the structural integrity, the dynamic range of solar magnetic fields, the magnetic flux and other physical features such as the size of the active regions, surpassing traditional persistence models, also in flaring situation. We aim to use deep learning not only for visualisation but as an integrative and interactive tool for telescopes, enhancing our understanding of unexpected physical events like solar flares. Future studies will aim to integrate more diverse solar data to refine the accuracy and applicability of our generative model. Visit this <https://github.com/fpramunno/MAG2MAG> for the code and <https://huggingface.co/spaces/fpramunno/mag2mag> for the interactive tool.

1 Introduction

The Sun's magnetic field is essential for understanding both the solar interior and the interplanetary environment, which includes space weather events close to Earth [1–6]. This magnetic field originates from within the solar interior and is most accurately measured at heights corresponding to the Sun's visible surface (the photosphere) [7]. Line-of-sight (LoS) magnetograms show the component of the magnetic field vector along the line of sight from the observer to the source and they map the solar photosphere. Solar magnetograms are essential for studying the mag-

netic field of the Sun and instabilities, including solar flares—rapid, intense bursts in the solar atmosphere that usually last a few minutes but can extend for several hours. Using the high-cadence Helioseismic and Magnetic Imager (HMI) [8] on the Solar Dynamics Observatory (SDO) [9], it is possible to study solar flare development and prediction for space weather applications [4, 10], as well as the temporal evolution of magnetic surface properties, including magnetic field flux, the size of active regions, their magnetic polarity inversion lines (PILs), etc. [11, 12].

In the last decade, artificial intelligence had a substantial boost in the research field [13], and the most successful machine learning methods in image analysis are built upon deep non-linear neural networks within the field of generative modelling [14–16]. These networks have been widely applied for various tasks such as image generation [17], super-resolution [18] and image-to-image translation [19]. They have also been applied in the field of heliophysics [4, 20, 21].

Image-to-Image translation aims to translate the domain knowledge of an image to another image. In astronomy, it can be used to transfer the knowledge of a telescope to another telescope [22, 23], allowing us to study objects in another domain, and it can be used to build an "artificial telescope" that analyses an astronomical object from an unseen angle [12, 24]. The study by Jeong, Moon, Park, Lee & Baek (2022) demonstrates how STEREO and SDO data can predict solar farside magnetograms as seen by HMI, enhancing its coverage to provide a 360-degree view of the solar photosphere. Thus, image-to-image translation offers promising applications in astronomy if used correctly.

In this work, we use image-to-image translation to predict the full disc LoS magnetogram from HMI using maps from 24 hours before (see Fig 1). We then

*Corresponding author. E-Mail: francesco.ramunno@fhnw.ch

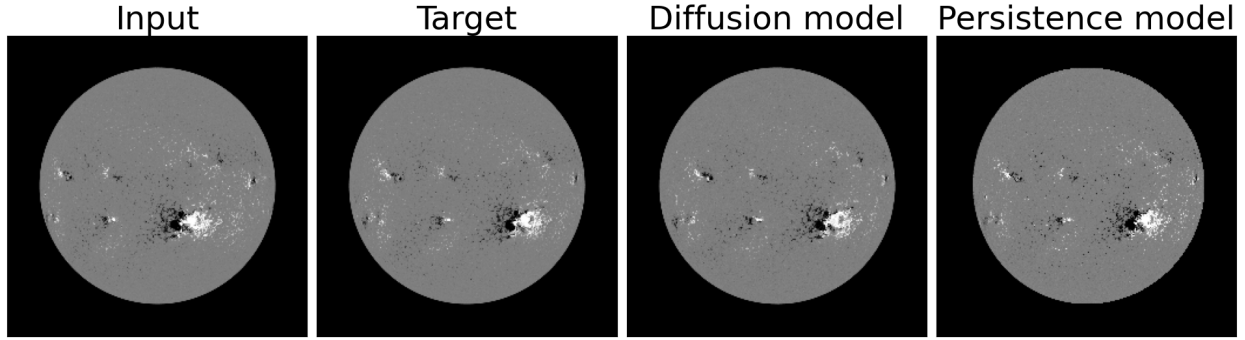


Figure 1: Example of input, target, and a comparison of the model’s prediction with the baseline persistence model. Input image from 25 October 2014 at 17:08 UTC.

evaluate the model’s accuracy by measuring physical parameters comparing them with a persistence model that accounts only for the Sun’s differential rotation.

This paper is organised as follows: In Sect. 2 we present our algorithm implementation and its result. In Sect. 3 we discuss our results and draw the conclusion.

2 Implementation and Results

The three pillars in generative modelling are the Variational Autoencoders (VAEs) [14], Generative Adversarial Networks (GANs) [15] and Denoising Diffusion Probabilistic Models (DDPMs) [16]. The priority criterion for this work is image quality because we intend to test if physics and pixel-level features are respected. For this reason, we do not consider the VAEs despite their speed of inference since they are known to generate blur images. GANs, instead, provide an optimal image quality while being very fast in sampling. However, GANs are notoriously difficult to train and can suffer from limited mode coverage and ambiguous convergence markers. For these reasons, we select the DPPMs due to their high fidelity sample quality and diversity, that can outperform both GANs and VAEs [25].

We use SDO/HMI magnetograms for our purposes [8, 9]. HMI records full-disc images of the solar photosphere at 6173 Å. In this work, we use LoS magnetograms available in compressed JP2 format from the Helioviewer [26] website ¹. The final dataset comprises 43912 paired images from 2013 to 2020 selected with the catalogue by Plutino, Berrilli, Del Moro & Giovannelli (2023). The pairs comprise the nearest LoS magnetogram to the flaring peak with a maximum distance of 1 min and its corresponding version 24 h before. Additionally, for computational efficiency, we decreased the image resolution to 256x256.

¹<https://helioviewer.ias.u-psud.fr/jp2/HMI/>

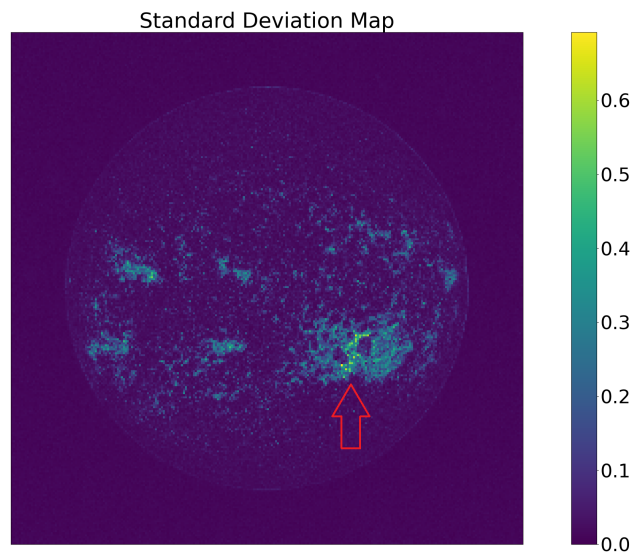


Figure 2: The image shows a standard deviation map derived from 8 repeated (semi-stochastic) model predictions. As expected, the variance shows that the model finds it most difficult to predict the PIL of the AR 12192 (indicated by the arrow), which physically undergoes the most significant changes during a flare. Input image from 24 October 2014 at 17:08 UTC. The standard deviation has been computed on images normalised between -1 and 1, thus the unit of measure is in pixel value.

Our algorithm follows the paper by Saharia *et al.* (2022). Starting from pure Gaussian noise and the input LoS magnetogram, it tries with 1000 applications of the model to predict how the LoS magnetogram will appear in 24 hours. To better evaluate the performance of our model, we employ a set of metrics from the computers science domain that are widely used to assess image quality, such as the Peak-Signal-to-Noise-Ratio (PSNR), the Structural-Similarity-Index-Measure (SSIM) [28] and the Learned Perceptual Image Patch Similarity (LPIPS) distance as a perceptual metric [29]. However, with only those metrics, it is im-

Model	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	FD TOT Flux	FD NET Flux	AR TOT Flux	AR NET Flux	Size AR	PIL length	Jacc Idx \uparrow
DDPM	31.2 \pm	0.7 \pm	0.03 \pm	0.37 \pm	1.25 \pm	5.76 \pm	13.96 \pm	2.9 \pm	3.34 \pm	0.65 \pm
Persistence	0.03	0.0004	0.0002	0.004	0.0	0.14	11.15	0.51	1.37	0.002
	15	0.78	0.10	2.01	1.25	14.44	42.37	8.09	5.65	0.55

Table 1: Results of the experiments based on the metrics explained in Section 2. The symbol \uparrow indicates that a higher value is preferable, while the symbol \downarrow indicates that a lower value is preferable. The SSIM and the Jaccard Idx are designed such that their maximum value is 1. FD stands for full-disc, TOT means total unsigned magnetic, NET refers to the net value, namely the value with the sign. Except for PSNR, SSIM, LPIPS, and Jaccard index, the metrics measure percentage differences between model predictions and true map values, indicating how far off the predictions are in percentage terms. These metrics have been computed over 10 sets of 250 samples each and the metric value and its associated standard deviation are percentage variations.

possible to understand if, while predicting the evolution of the LoS magnetogram, the physics of the processes is also preserved. Thus, we define a "physics metrics" set to judge if the model correctly emulates the physical processes at work. We divide those into the global (full-disc) category and the local (active-region) category. To track all the ARs and compute the metrics, we use the SunPy library [30] to locate all the ARs present at a specific time with their centre through the NOAA (National Oceanic and Atmospheric Administration) AR number. The first physics metrics that we implement are the total magnetic flux for the whole disc and all the ARs, and the net magnetic field flux for the entire disc and all the ARs [7] (See section 3 for the flux calculation from Helioviewer’s JP2 images). Additionally, we compute the AR size and the PIL length, which are two essential parameters to discriminate if the AR has the potential to flare or not [31, 32]. To extract the PIL length from the LoS magnetograms, we follow the algorithm proposed in Cai, Aydin, Ji, Georgoulis & Angryk (2020). We also use the Jaccard Index [34] to approximate the morphology of ARs. To evaluate the results, we compute the percentage variation (PV) between the metric value obtained on the true LoS magnetogram and the value obtained on the predicted LoS magnetogram. The results obtained with the DDPM are compared with the persistence model. **The persistence model is not a machine learning model, and it considers only the Sun’s differential rotation.** Hence, the persistence model does not predict the evolution of the AR and their physics features but it carries forward to see how the Sun will look after a given amount of time. We perform this with the SunPy library. To evaluate our model, we test on 250 cases randomly selected by the validation set. We apply the model 10 times on the 250 data, obtaining ten sets of 250 predictions of the input data. We compute the metrics for all ten sets and

aggregate those results using the mean and the standard deviation. We repeat this analysis because the DDPM is a probabilistic model [16], thus given the same input we will always obtain slightly different results and so we can retrieve an uncertainty estimation. The results are presented in Table 1.

3 Discussion

As stated in Section 2 we have two sets of metrics: the computer science metrics, which include the PSNR, the SSIM, the LPIPS and the physics metrics. Computer science metrics evaluate the quality of the output regardless of its physical meaning. The PSNR and the SSIM are pixel-based similarity scores. The PSNR measures the ratio of the maximum possible signal power to the power of distorting noise affecting its fidelity. As we can see from Table 1 the DDPM beats the Persistence model, which means that the DDPM is better in predicting the actual dynamic range of the image. The SSIM considers the luminance, contrast, and structure; in this case, the Persistence model is better. However, pixel-level metrics are not always reliable for sample quality [18], and it has been observed that PSNR and SSIM tend to prefer blurry outputs [35]. Thus, to better evaluate the performance of the model, we use the LPIPS distance, an L2 distance in the latent space of a pre-trained Alexnet model [36], and in this case, the DDPM beats the Persistence model. In conclusion, regarding the computer science metrics, the DDPM performs better overall than the Persistence model in terms of dynamic range and overall structure of the predicted image.

Nevertheless, we are interested in more than just the aesthetic of the image: we wish to understand if the model can replicate the nature of some of the physics processes. For this reason, we define a set

of metrics that characterise those processes like the magnetic flux, since its changes correlate with the structural complexity of the field, which is fundamental for understanding and predicting events like flares and coronal mass ejections (CMEs) [7]. To assess those measurements, we must work on a physics space, which is not the space between 0 and 1 that Heliowiewer [26] shows. To go back into the physics space, we revert the linear transformation used by the Heliowiewer project ² which goes back to an image whose magnetic field values vary in the range [-250, 250] G. We compute the total and the net magnetic fluxes, both at the full-disc level and at the ARs level. As shown in Table 1, the magnetic flux with a full-disc precision has a lower PV than the ARs precision. The magnetic flux is the product of the magnetic field and the area. Thus, the full-disc measure seems more accurate because we are averaging over a more extensive area that contains lower magnetic field values. In this case, both the magnitude and the net magnetic field flux are better estimated by the DDPM, but to truly understand the model capabilities we must go to the AR level. The PV for the AR is higher than the full-disc, as we can see from Table 1, because in this case we compute the metric over a smaller and more complex area, which is harder to predict. In both cases, the magnitude and the net magnetic field flux, the DDPM is better than the persistence model. Still, there is a difference between the magnitude and the net flux. This difference is shown in the average value and the standard deviation, which determines how uncertain the model is. The magnitude is more straightforward to determine because it represents the overall flux with no interest in the direction of the field. For this reason, the model encounters more difficulties in its prediction because, given a static input (LoS magnetogram 24 hours before), we are predicting the dynamics of a complex field that changes on a much shorter time scale. In addition to this, the stochasticity of the model has a more significant impact, given that the uncertainty of the prediction is equal to 80 % of the average predicted value. In future work, we are interested in better exploring the ability to predict dynamic features, including inputs, that take into account different layers of the Sun (e.g. AIA input images), but also to encode features that are correlated with the magnetic field (e.g. velocity fields [7]).

To test the limit of the DDPM and demonstrate how it behaves on a finer-grained scale, we use as evaluation parameters also the size of the ARs and PIL length, which widely correlate with events such as flares or CMEs, and the Jaccard index which we

use as a proxy for the morphology of the ARs. However, these metrics requires the use of external libraries such as OpenCV [37] and so additional algorithms, differently from the magnetic flux measurements leading to higher error propagation. To prevent this, we use the same hyper-parameters of the algorithms for all the images, even though to be more precise each image should be characterised by its parameters. Nonetheless, we see from Table 1 that for all of these metrics, the DDPM prediction is better than the Persistence model. Consequently, the output given by the DDPM can be used not only as a visualisation technique but also as a scientific result that can be studied and analysed. Moreover, the training and validation datasets consist of image pairs captured near the flare peak and 24 hours earlier. The model's accurate predictions of these images demonstrate its ability to forecast the Sun's behaviour during flare events. Since the length of the PIL and the AR area are reliable indicators of potential flares, accurately predicting these values suggests that the model holds essential information about flare events 24 hours in advance. Nevertheless, flare prediction is not the topic of this project but in future work we aim on using explainable AI to understand the model behaviour [38] during these events. Moreover, due to the probabilistic nature of the DDPM, it is interesting to check what is the most uncertain part of the LoS magnetogram to predict. We input the image 24 hours before the X1.0 flare happened on 25 October 2014 from the AR 12192, and we predict it eight times. To visualise the uncertainty of the model, we concatenate the eight predictions along the channel dimension and then compute the standard deviation per pixel. This process outputs an image where the higher the value is, the bigger the standard deviation. Thus, the model is more uncertain on that pixel prediction. As shown in Figure 2, the values are higher in correspondence with the AR 12192, which is the flaring region, and most importantly, it is more uncertain in the region of polarity inversion (RoPI). This result is consistent with our physical knowledge as shown by Kleint (2017) that photospheric changes during an event like a flare are predominantly located near the PIL. Considering the computational constraints (e.g. size reduction to 256x256 pixels) and the initial design of the model (e.g. using only one LoS magnetogram), the performance on the physics metrics of the DDPM is encouraging. Thus, we are interested in expanding this in future work to test if injecting different types of information (e.g. velocity fields, AIA channels, LoS magnetograms of multiple days) helps the model to understand the evolution of these dynamic processes better and thus explore which processes cause the magnetic field in the lower solar atmosphere to change, which

²https://github.com/Heliowiewer-Project/jp2gen/blob/master/idl/sdo/hmi/hvs_default_hmi.pro

is not fully understood yet.

Acknowledgements: This research was partially funded by the SNF Sinergia project (CRSII5-193716): Robust Deep Density Models for High-Energy Particle Physics and Solar Flare Analysis (RODEM).

References

1. Leka, K. D. & Barnes, G. Photospheric Magnetic Field Properties of Flaring versus Flare-quiet Active Regions. I. Data, General Approach, and Sample Results. *apj* **595**, 1277–1295 (Oct. 2003).
2. Leka, K. D. & Barnes, G. Photospheric Magnetic Field Properties of Flaring versus Flare-quiet Active Regions. II. Discriminant Analysis. *apj* **595**, 1296–1306 (Oct. 2003).
3. Barnes, G. & Leka, K. D. Photospheric Magnetic Field Properties of Flaring versus Flare-quiet Active Regions. III. Magnetic Charge Topology Models. *apj* **646**, 1303–1318 (Aug. 2006).
4. Pandey, C., Angryk, R. A., Georgoulis, M. K. & Aydin, B. in *Lecture Notes in Computer Science* 567–581 (Springer Nature Switzerland, 2023). ISBN: 9783031452758. http://dx.doi.org/10.1007/978-3-031-45275-8_38.
5. Yi, K., Moon, Y.-J. & Jeong, H.-J. Application of Deep Reinforcement Learning to Major Solar Flare Forecasting. *apjs* **265**, 34 (Apr. 2023).
6. Georgoulis, M. K. *et al.* Prediction of solar energetic events impacting space weather conditions. *Advances in Space Research*. ISSN: 0273-1177. <https://www.sciencedirect.com/science/article/pii/S027311772400173X> (2024).
7. Wiegmann, T., Thalmann, J. & Solanki, S. The Magnetic Field in the Solar Atmosphere. *Astronomy and Astrophysics Review* **22**, 78. <https://doi.org/10.1007/s00159-014-0078-7> (2014).
8. Schou, J. *et al.* Design and Ground Calibration of the Helioseismic and Magnetic Imager (HMI) Instrument on the Solar Dynamics Observatory (SDO). *solphys* **275**, 229–259 (Jan. 2012).
9. Pesnell, W. D., Thompson, B. J. & Chamberlin, P. C. The Solar Dynamics Observatory (SDO). *solphys* **275**, 3–15 (Jan. 2012).
10. Son, J., Sung, S.-K., Moon, Y.-J., Lee, H. & Jeong, H.-J. Three-day Forecasting of Solar Wind Speed Using SDO/AIA Extreme-ultraviolet Images by a Deep-learning Model. *The Astrophysical Journal Supplement Series* **267**, 45 (2023).
11. Jeong, H.-J., Moon, Y.-J., Park, E. & Lee, H. Solar coronal magnetic field extrapolation from synchronic data with AI-generated farside. *The Astrophysical Journal Letters* **903**, L25 (2020).
12. Jeong, H.-J., Moon, Y.-J., Park, E., Lee, H. & Baek, J.-H. Improved AI-generated Solar Farside Magnetograms by STEREO and SDO Data Sets and Their Release. *The Astrophysical Journal Supplement Series* **262**, 50 (2022).
13. Collins, C., Dennehy, D., Conboy, K. & Mikalef, P. Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management* **60**, 102383. ISSN: 0268-4012. <https://www.sciencedirect.com/science/article/pii/S0268401221000761> (2021).
14. Kingma, D. P. & Welling, M. *Auto-Encoding Variational Bayes* 2022. arXiv: [1312.6114](https://arxiv.org/abs/1312.6114) [stat.ML].
15. Goodfellow, I. J. *et al.* *Generative Adversarial Networks* 2014. arXiv: [1406.2661](https://arxiv.org/abs/1406.2661) [stat.ML].
16. Ho, J., Jain, A. & Abbeel, P. *Denosing Diffusion Probabilistic Models* 2020. arXiv: [2006.11239](https://arxiv.org/abs/2006.11239) [cs.LG].
17. Rombach, R., Blattmann, A., Lorenz, D., Esser, P. & Ommer, B. *High-Resolution Image Synthesis with Latent Diffusion Models* 2022. arXiv: [2112.10752](https://arxiv.org/abs/2112.10752) [cs.CV].
18. Saharia, C. *et al.* *Palette: Image-to-Image Diffusion Models* 2022. arXiv: [2111.05826](https://arxiv.org/abs/2111.05826) [cs.CV].
19. Li, H. *et al.* *SRDiff: Single Image Super-Resolution with Diffusion Probabilistic Models* 2021. arXiv: [2104.14951](https://arxiv.org/abs/2104.14951) [cs.CV].
20. Ramunno, F. P. *et al.* *Solar synthetic imaging: Introducing denoising diffusion probabilistic models on SDO/AIA data* 2024. arXiv: [2404.02552](https://arxiv.org/abs/2404.02552) [astro-ph.SR].
21. Sun, Z. *et al.* Predicting Solar Flares Using CNN and LSTM on Two Solar Cycles of Active Region Data. *The Astrophysical Journal* **931**, 163. ISSN: 1538-4357. <http://dx.doi.org/10.3847/1538-4357/ac64a6> (June 2022).
22. Kinakh, V. *et al.* Hubble Meets Webb: Image-to-Image Translation in Astronomy. *Sensors* **24**. ISSN: 1424-8220. <https://www.mdpi.com/1424-8220/24/4/1151> (2024).
23. Drozdova, M. *et al.* Radio-astronomical image reconstruction with a conditional denoising diffusion model*. *AandA* **683**, A105. <https://doi.org/10.1051/0004-6361/202347948> (2024).
24. Jarolim, R. *et al.* SuNeRF: 3D Reconstruction of the Solar EUV Corona Using Neural Radiance Fields. *The Astrophysical Journal Letters* **961**, L31 (Jan. 2024).
25. Dhariwal, P. & Nichol, A. *Diffusion Models Beat GANs on Image Synthesis* 2021. arXiv: [2105.05233](https://arxiv.org/abs/2105.05233) [cs.LG].
26. Ireland, J. *et al.* The Helioviewer Project: Discovery For Everyone Everywhere. **40** (Apr. 2009).
27. Plutino, N., Berrilli, F., Del Moro, D. & Giovannelli, L. A new catalogue of solar flare events from soft X-ray GOES signal in the period 1986–2020. *Advances in Space Research* **71**, 2048–2058. ISSN: 0273-1177. <http://dx.doi.org/10.1016/j.asr.2022.11.020> (Feb. 2023).
28. Wang, Z., Bovik, A., Sheikh, H. & Simoncelli, E. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* **13**, 600–612 (2004).
29. Zhang, R., Isola, P., Efros, A. A., Shechtman, E. & Wang, O. *The Unreasonable Effectiveness of Deep Features as a Perceptual Metric* 2018. arXiv: [1801.03924](https://arxiv.org/abs/1801.03924) [cs.CV].
30. 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–. <https://iopscience.iop.org/article/10.3847/1538-4357/ab4f7a> (1 2020).
31. Sadykov, V. M. & Kosovichev, A. G. Relationships between Characteristics of the Line-of-sight Magnetic Field and Solar Flare Forecasts. *The Astrophysical Journal* **849**, 148. ISSN: 1538-4357. <http://dx.doi.org/10.3847/1538-4357/aa9119> (Nov. 2017).
32. Wang, J. *et al.* Solar Flare Predictive Features Derived from Polarity Inversion Line Masks in Active Regions Using an Unsupervised Machine Learning Algorithm. *apj* **892**, 140 (Apr. 2020).
33. Cai, X., Aydin, B., Ji, A., Georgoulis, M. & Angryk, R. *A Framework for Detecting Polarity Inversion Lines from Longitudinal Magnetograms* in (Dec. 2020), 4175–4183.
34. Jaccard, P. Etude de la distribution florale dans une portion des Alpes et du Jura. *Bulletin de la Societe Vaudoise des Sciences Naturelles* **37**, 547–579 (Jan. 1901).

35. Menon, S., Damian, A., Hu, S., Ravi, N. & Rudin, C. *PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models* 2020. arXiv: [2003.03808](https://arxiv.org/abs/2003.03808) [[cs.CV](#)].
36. Krizhevsky, A., Sutskever, I. & Hinton, G. E. *ImageNet Classification with Deep Convolutional Neural Networks* in *Advances in Neural Information Processing Systems* (eds Pereira, F., Burges, C., Bottou, L. & Weinberger, K.) **25** (Curran Associates, Inc., 2012). https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.
37. Bradski, G. The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2000).
38. Panos, B., Kleint, L. & Zbinden, J. Identifying preflare spectral features using explainable artificial intelligence. *aap* **671**, A73. arXiv: [2301.01560](https://arxiv.org/abs/2301.01560) [[astro-ph.SR](#)] (Mar. 2023).
39. Kleint, L. First Detection of Chromospheric Magnetic Field Changes during an X1-Flare. *apj* **834**, 26. arXiv: [1608.02552](https://arxiv.org/abs/1608.02552) [[astro-ph.SR](#)] (Jan. 2017).