Analysis of Solar Images at Different Wavelengths to Identify Active Regions in the Sun

Key-Words: Space Weather, Image Classification, Solar Flares

Authors:
Ana Luísa Fogarin de Sousa Lima [Unicamp]
Prof. Dr. André Leon Sampaio Gradvohl [Unicamp]

INTRODUCTION

The Sun is in constant activity. Some events that occur in the Sun, such as solar explosions or coronal mass ejections, release energy and particles that can hit Earth and cause several damages in satellites or even in the electrical power transmission, as observed during the solar storm in March of 1989 (BOTELER, 2019). However, the prediction of these phenomena can prevent severe consequences.

One of the mechanisms that can help predict solar flares is the automatic identification of active regions in the solar chromosphere. An active region on the Sun is an area with a strong magnetic field. Solar flares and coronal mass ejections blast forth from active regions.

The study reported in this paper used images of the Sun from three different instruments: the Helioseismic and Magnetic Imager Continuum (HMI Continuum), which shows the visible light of the surface of the Sun; the AIA1600 Å, which presents the so-called region of transition of the Sun with temperatures around 6000 Kelvin; AIA1700 Å that also represents the region of transition of the Sun, but with temperatures around of 4500 Kelvin. Figure 1 illustrates the different types of solar images at different wavelengths.

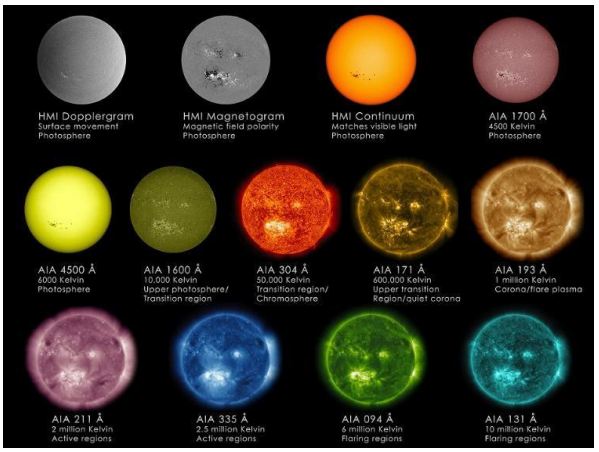

Figure 1: Different images of the Sun in various wavelengths. Fonte: https://www.nasa.gov/mission_pages/sunearth/science/Sun-Wavelength-Chart.html

We compiled a dataset of images obtained from these instruments at three wavelengths (HMI Continuum, AIA 1700, and AIA 1600). The dataset is composed, in total, of 4698 images. We also developed a script in Python to automatize the download

process and convert the images from Flexible Image Transport System (FITS) to Portable Network Graphics (PNG). This procedure is important for preparing and standardizing data for further processing.

Finally, using the DIGITS software, we trained image classification models to classify the solar flares according to the intensity of the X-ray emission of the solar flare in that active region. The main goal is to improve the automatic detection of active regions and help identify possible solar flares.

METHODOLOGY:

We obtained the information about the solar flares from a database with more than 2.257 registers with the most intense solar flares since 1997 (GRADVOHL; FERNANDES, 2017). To make processing such a huge amount of data feasible, we developed a script in Python to filter and process the data from the file with the information about the solar flares.

After that, we downloaded the images from the Joint Science Operations Center (JSOC). Therefore, the first step of the script was to filter the events that happened after 2011 since the repository only keeps registers after that date. In the download process, we used the Data Record Management System (DRMS) application programming interface (GLOGOWSKI et al., 2019).

At first, we separated the images only according to their wavelength. However, after running some tests, we consider the necessity of dividing them according to the type of solar flares that showed up. Then, we implemented the script to do all these processes automatically.

To run the script on download mode, the user should use the following command line:

Python download_images.py <solar_explosion_info_file.csv> 1

The input file must contain the date, time, and the types of the explosions. The script will try to find the exact date and time, but if there is no registry available, it will search for registries between a half-hour (for less or more) of the provided time. Unfortunately, in some cases, images cannot be found even in the range time, making the download impossible. The script stores the information about flares that cannot be downloaded in a bin file.

Table 1 shows the number of images from each type of solar flares the script obtained. We considered four classes (types) of flares. X-class flares are the strongest ones, followed by M-class flares, C-class flares, and B-class flares, in that order.

Relation of images of the data base					
Wave lenght	Total of images	Class X	Class M	Class C	Class B
AIA 1600	1568	37	232	614	685
AIA 1700	1537	37	228	594	678
HMI Continuum	1593	37	232	627	697

Table 1 - Relation of images obtained.

Initially, we downloaded the images in FITS format. The, we converted the ones in AIA 1600 Å to PNG using ALP's Labeling Tool (ALPER, 2017). After the conversion, we did a manual process, and we discarded approximately 450 images due to errors during the conversion.

Therefore, we implemented a new feature on the script: the automatic conversion of FITS files to PNG, based on a

code available on GitHub (HENNEY, 2015). Then, with some params of luminosity, defined according to every wavelength, the script converted and stored the new images in a new folder. The user should use the following command to use this functionality:

Python download_images.py <solar_explosion_info_file.csv> 2

As mentioned before, we built new databases (GRADVOHL; LIMA; LIU, 2021a, b, c), for tests and we used the latter on the real models. The distribution of training images, validation and tests was 70%, 20%, and 10%, respectively.

RESULTS AND DISCUSSION:

For the models using HMI Continuum images, four were successful as described in Table 2. It is interesting to observe that models 1 and 2 used the same dataset but different solvers: the first one used Adam (DIEDERIK P. KINGMA, 2014) and the second used Stochastic Gradient Descent (SGD) (BOTTOU, 2012). Still, the difference between their results is huge: the first one had a maximum accuracy of 50%, while the second hit 93%.

	Contrinuum 1	Contrinuum 2	Contrinuum 6	Contrinuum 8
Dataset	Continuum	Contrinuum	Contrinuum 3	Contrinuum 3
Training epochs	15	30	40	20
Solver Type	Adam	SGD	SGD	SGD
Learning Rate	0.01	0.01	0.005	0.005
Result	Done	Done	Done	Done
Time	30 min	1 hour	5 hours	2 hours

Table 2: Models using Continuum Images.

The results were similar on models 2, 6, and 8. We achieved accuracy between 80% and 92%. However, model 6 took more the twice the time to complete. In general, model 8 had the best performance because of its high accuracy of training and validation (first one around 85% and second around 89%).

However, Model 1 presents overfitting, which happens when the model adapts too well to the dataset and ends up memorizing the items, as illustrated in Figure 2.

Figure 2: Model one with overfitting.

Of the seven models using AIA 1600 Å images, only two were successful, described in Table 3. We canceled some models because they were using the Adam solver, which did not have promising results using HMI Continuum images, but also because of the long amount of time they would complete.

	AIA 1600 5	AIA 1600 6
Dataset	AIA 1600 2	AIA 1600 2
Training epochs	15	20
Solver Type	SGD	SGD
Learning Rate	0.01	0.002
Result	Done	Done
Time	1 hour 52 min	2 hours 30 min

Table 3: Models using AIA 1600 Å Images.

Both models used the same dataset and SGD solver but different epochs (complete passages on the data) and learning rates. The accuracy is around 80% and 90%, but the first model had a slightly better performance since the loss value is lower in it.

Concerning the images in AIA 1700 Å, we built five models. Three of them had success as described in Table 4, and two resulted in errors. Models 2 and 5 used the same dataset and solver, differing on the learning rate (0.001 for the first and 0.002 for the second). Although these small differences, model 5 had better results, hitting a maximum accuracy of 90% and a loss value of only 0.42, when Model 2's accuracy was not more than 86%, and loss was around 0.49.

	AIA 1700 2	AIA 1700 4	AIA 1700 5
Dataset	AIA 1700	AIA 1700	AIA 1700 2
Training epochs	20	20	50
Solver Type	SGD	SGD	SGD
Learning Rate	0.001	0.002	0.002
Result	Done	Done	Done
Time	2 hours	2 hours 20 min	1 hour 20 min

Table 4: Models using AIA 1700 Å Images

Table 5 presents an overview of the results of the models to better analysis of the results.

	Contrinuum 1	Contrinuum 2	Contrinuum 6	Contrinuum 8	AIA 1600 5	AIA 1600 6	AIA 1700 2	AIA 1700 4	AIA 1700 5
Loss (train)	0	0.40	0.25	0.44	0.33	0.50	0.49	0.42	0.25
Accuracy (train)	28% - 49%	80% - 93%	84% - 92%	80% - 90%	80% - 90%	75% - 86%	75% - 86%	80% - 90%	85% - 93%
Loss (validation)	0	0.59	0.75	0.76	0.5	0.5	0.67	0.62	0.47
Accuracy (validat	45%	76%	75%	76%	86%	81%	75%	76%	85%

Table 5: Overview of the mode

CONCLUSION:

This study used 4698 images in three different wavelengths (HMI Continuum, AIA 1600 Å, and AIA 1700 Å). We used 70% of the dataset for training, 20% for validation, and 10% for testing. Thus, we built different datasets, with different parameters and used for training the classification models.

We observed that solver SGD had better results during those trainings, once models that used solver Adam had errors and problems, such as overfitting (JABBAR; KHAN, 2014). Learning rates lower than 0.01 also presented better results. Many models had accuracy higher than 80%, and the maximum value was 95%.

We consider those values to be high, just as other similar studies in the same research field do, but that used HMI magnetograms, had accuracy of around 50% (BOBRA; COUVIDAT, 2015).

Besides, the script we developed makes the download process easier. The conversion of Sun images in different wavelengths can be a huge facilitator to following space weather research, especially under the processing of solar images, since a huge number of images can be downloaded automatically.

BIBLIOGRAPHY

ALPER. Alp's Labeling Tool. 2017. Available at: <https://alpslabel.wordpress.com/2017/01/26/alt/>. Access on jan 23, 2021.

BOBRA, M. G.; COUVIDAT, S. **Solar Flare Prediction Using SDO/HMI Vector Magnetic Field Data with a Machine-Learning Algorithm**. The Astrophysical Journal, v. 798, n. 2, p. 135, 8

jan. 2015.DOI: 10.1088/0004-637X/798/2/135.

BOTELER, D. H. **A 21st Century View of the March 1989 Magnetic Storm**. *Space Weather*, 17, 1427– 1441. https://doi.org/10.1029/2019SW002278

BOTTOU, L. **Stochastic Gradient Tricks**. In: MONTAVON, G.; ORR, G. B.; MÜLLER, K.-R. (Ed). **Neural Networks, Tricks of the Trade, Reloaded**. [S.l.]: Springer, 2012. (Lecture Notes in Computer Science (LNCS 7700)). p. 430–445. Available at: <http://leon.bottou.org/papers/bottou-tricks-2012>.

DIEDERIK P. KINGMA, J. B. Adam: **A Method for Stochastic Optimization**. arXiv, 2014. Available at: <https://arxiv.org/abs/1412.6980>. Access on jan 23, 2021.

GLOGOWSKI, K. et al. **drms**: **A Python package for accessing HMI and AIA data.** Journal of Open Source Software, The Open Journal, v. 4, n. 40, p. 1614, 2019. DOI: 10.21105/joss.01614.

GRADVOHL, A. L. S.; LIMA, A. L. F. S.; LIU, J. **Images of solar flares in 1600 Angstrom wavelenght**. Versão 1.1. [S.l.]: Zenodo, jan. 2021. DOI: 10.5281/zenodo.4435218.

GRADVOHL, A. L. S.; LIMA, A. L. F. S.; LIU, J. **Images of solar flares in 1700 Angstrom wavelength**. Versão 1.0. [S.l.]: Zenodo, fev. 2021. DOI: 10.5281/zenodo.4484104.

GRADVOHL, A. L. S.; LIMA, A. L. F. S.; LIU, J. **Images of solar flares in HMI Continuum.**

Versão 1.0. [S.l.]: Zenodo, fev. 2021. DOI: 10.5281/zenodo.4484112.

GRADVOHL, A. L. S.; FERNANDES, M. E. R. **Samples of solar flares classes, active regions and time of occurrence**. [S.l.: s.n.], 2017. DOI: 10.5281/zenodo.1048995. Access on jan 27, 2021.

HENNEY, W. **Convert a FITS file to an image file**. 2015. Available at: <https://github.com/will-henney/fits2image>. Access on jan 16, 2021.

JABBAR, H. K.; KHAN, R. Z. **Methods to Avoid Over-ftting and Under-ftting in Supervised Machine Learning** (Comparative study), 2014. Available at: <https://www.researchgate.net/profile/Haider_Allamy/publication/295198699_METHODS_TO_AVOID_OVER-FITTING_AND_UNDER-FITTING_IN_SUPERVISED_MACHINE_LEARNING_COMPARATIVE_STUDY/links/56c8253f08aee3cee53a3707.pdf>. Access on jan 23, 2021.