Data and models for sunspots detection in solar images captured with smart telescopes

Olivier Parisot

Luxembourg Institute of Science and Technology (LIST)
5 Avenue des Hauts-Fourneaux, 4362 Esch-sur-Alzette, Luxembourg
olivier.parisot@list.lu

Abstract. Observing the sun with Electronically Assisted Astronomy allows understanding of solar phenomena, contributes to scientific research, promotes science education, and provides recreational enjoyment, all while requiring strict adherence to safety measures. In this paper, we present an annotated dataset of solar images captured with smart telescopes, and we show how this dataset allows to train Deep Learning YOLOv7 models for the detection of sunspots. Both data and Deep Learning model can be used by the general public to observe sun with contextual information.

Keywords: Solar images · Sunspots · Smart telescopes

1 Introduction

Sun observation using equipment accessible to amateur and professional astronomers offers a mix of educational, scientific and recreational benefits, and can be done outdoors or in town centres, as long as the visible horizon allows the Sun to be pointed at. This activity improves understanding of solar physics and practical skills while allowing amateurs to provide valuable data for scientific research, particularly in monitoring solar activity. Even if the result will not be as detailed and impressive as the results provided by space agencies services like the Space Weather Prediction Center [1], it fosters a sense of wonder, community engagement and promotes public awareness and education [7].

In particular, solar observation allows to study sunspots, one of the most visible features of solar activity cycles. Appearing as dark regions of varying size, sunspots are manifestations of strong perturbations in the magnetic field of the Sun, and are continuously studied to analyze solar activity cycles. Sunspots are also regularly monitored by passionate amateur astronomers, and in May 2024 they were a warning of a solar storm causing fantastic northern lights observed at unusual latitudes [2].

In practice, always using a dedicated and safe filter, solar observation can be done with the naked eye [17], with a refractor or reflector type instrument [6].

[1] https://www.swpc.noaa.gov

[2] https://en.wikipedia.org/wiki/May_2024_solar_storm

Moreover, it can be carried out via Electronically Assisted Astronomy (EAA), applied by astronomers to easily observe the sky (night and day). Sometimes known as *video astronomy*, EAA consists in collecting images directly from a sensitive camera coupled to an optical system, and then applying lightweight image processing on a computing device to generate enhanced images in near real time of targets directly on screens (tablet, smartphone, even TV) [10].

The recent advent of intelligent telescopes has transformed observation, making it accessible to almost everyone. These telescopes automatically manage phases that can be tedious, such as initialization with sky recognition, tracking, and focusing. These telescopes require no prior knowledge, making it possible to capture and share instant solar images in less than several minutes. We can mention that dedicated portable robotic telescopes were already designed and used to observe Sun for educational purposes [13].

Nowadays, and as is the case in many fields, Artificial Intelligence is increasingly used to support observational astronomy [11]. For instance, AI can also be used to help and encourage solar observation, both for amateurs and professionals, as is being done for STEM outreach [15]. In this paper, we describe SunspotsYoloDataset, a collection of annotated solar images captured by two smart telescopes, and we explain how we trained Deep Learning models to detect sunspots on this kind of images. On the one hand, the dataset shows what it is possible to obtain in terms of solar images with equipment and from observation conditions accessible to amateurs. On the other hand, detection models can serve as a tool to help amateur astronomers gain contextual information about what is being observed.

The present paper is structured as follows. In Section 2, we list and discuss existing techniques for sunspots in astronomical images. In Section 3, we present how we collected and compiled data to generate the SunspotsYoloDataset dataset. We detail different our approach to detect sunspots based on this dataset in Section 4. Finally, we discuss the results in Section 5, and we propose some perspectives in Section 6.

2 Related works

In recent years, numerous automated methods for detecting sunspots have been proposed. Traditional image processing methods rely primarily on the intensity of sunspots, as they tend to appear darker than their surroundings [5,4]. These techniques require a threshold to separate sunspots from the background, which will depend greatly on the images to be analysed. Thus, many Machine Learning techniques were proposed and we can mention the following recent works:

- YOLO-based approaches (You Only Look Once) were applied, in particular YOLOv5 models on images from the Geophysical and Astronomical Observatory of the University of Coimbra [14], and YOLOv5 models to track sunspots from FITS images taken by the Solar Dynamics Observatory [3].

[3] `https://github.com/ChrisToumanian/solar-yolo`

– Analysis of Solar Dynamics Observatory data with two techniques Single Shot MultiBox Detector (SSD) and the Faster Region-based Convolutional Neural Network (R-CNN) for multi-classes detection (sunspots, but also coronal holes and prominences) [2].
– Images classification with Support Vector Machines (SVM) on data coming from the Michelson Doppler Imager [3].
– HybridVR, based on ResNet50 and VGG16 to extract key features of activity and environmental characteristics from observed solar images [19].
– It's not the same task, but we can mention that Deep Learning techniques are also used for flares detection in Solar Dynamics Observatory data [1].

These methods require realistic and annotated training datasets to be effective. Data like that from the Solar Dynamics Observatory is often used by the scientific community. As far as we know, there is not much work based on images captured in conditions and with equipment accessible to both professionals and amateurs, except the Solar Database [4].

Moreover, one of the limit of existing detection models is the potential confusion between potential unexpected observation conditions (clouds, trees, etc.) and real sunspots (Figure 1).

Fig. 1: Solar image captured with a Vespera smart telescope (16/1/2024) and annotated with a YOLOv5 detection model presented in [14]: the models detect trees as sunspots – these are false positives.

[4] http://solardatabase.free.fr

In the following section we explain how we have captured and then annotated a dataset of solar images obtained in conditions that are not ideal, but which then allow us to train models less sensitive to these problems.

3 Data acquisition with smart telescopes

For this work, we collected a large amount of high-resolution solar images with smart telescopes. Between January 2023 and May 2024, The images were gathered in Luxembourg, France and Belgium, in urban environments and with a rather polluted sky, by using two robotic instruments:

- Stellina [5]: its optical part consists of an ED doublet (Extra-low Dispersion) with a focla ratio of f/5 (aperture of 80 mm, focal length of 400 mm) – imaging is carried out using a Sony IMX178 CMOS sensor (resolution of 6.4 million pixels, i.e. 3096 × 2080).
- Vespera [6] (Figure 2): its optical part consists of an apochromatic quadruplet with a focal ratio of f/4 (aperture of 50 mm, focal length of 200 mm) – imaging is carried out using a a Sony IMX462 CMOS sensor (resolution of 2 million pixels, i.e. 1920 × 1080).

Dedicated solar filters were used for both telescopes, allowing only the desired wavelengths to pass through, and to preserve the integrity of the telescope and sensors (Figure 2).

For each solar observation session, the instruments were properly balanced using a spirit level on a stable floor.

We also managed to get images even when conditions weren't ideal, and in particular in two situations: when the sky was cloudy, and when the sun was very low on the horizon (to get images masked by tree branches).

As a result, we collected thousands of high-resolution JPEG images, which were then cropped into patches of 640 × 640 pixels. In order to make this set of patches more heterogeneous in terms of image quality, we applied post-processing operations on them, such as rotations, contrast enhancement and color saturation, by using a genetic algorithm introduced in [12]. We then annotated these post-processed patches using MakeSense [7]: dedicated to annotating data sets for recognition purposes, this interactive web application allows anyone to draw, move, resize and label bounding boxes corresponding to elements present in images. So we've annotated all the sunspots on the images, and we've taken care to ignore any patterns that aren't sunspots (as in Figure 1).

The images and annotations were compiled into the SunspotsYoloDataset, a set of 2198 solar images formatted with the YOLO standard, and splitted as 3 sets (1690 for training, 380 for validation, 128 for test). This means that there are separate files for images and annotations (i.e. text files containing the positions of sunspots), all stored and compressed in a ZIP file. SunspotsYoloDataset is available from an open archive on Zenodo [9].

[5] `https://vaonis.com/stellina`
[6] `https://vaonis.com/vespera`
[7] `https://www.makesense.ai`

Fig. 2: A Stellina (left) and a Vespera (right) smart telescopes, with dedicated solar filters.

4 Approach

Based on SunspotsYoloDataset, we then trained several YOLOv7 models with different architectures (normal, tiny) and various parameters (with or without transfer learning, etc.). To this end, we used the official implementation available on Github [18]. This Python source code can be used to train a YOLOv7 model using standard architecture and transfer learning (based on a pre-trained YOLOv7 model [8]) with the following parameters:

```
python3 train.py --img 640 --weights yolov7.pt
          --data 'data/custom.yaml' --single-cls
          --workers 8 --batch-size 4 --epochs 300
          --cfg cfg/training/yolov7.yaml
          --name sunspots-yolov7
          --hyp data/hyp.scratch.p5.yaml
```

Models training was realized by using Virtual Containers managed via an instance of Portainer [9], used to exploit the following hardware: 40 cores with 128 GB RAM (Intel(R) Xeon(R) Silver 4210 @ 2.20 GHz CPU) and NVIDIA Tesla V100-PCIE-32 GB as GPU.

[8] https://github.com/WongKinYiu/yolov7/releases/download/v0.1/yolov7.pt

[9] https://www.portainer.io/

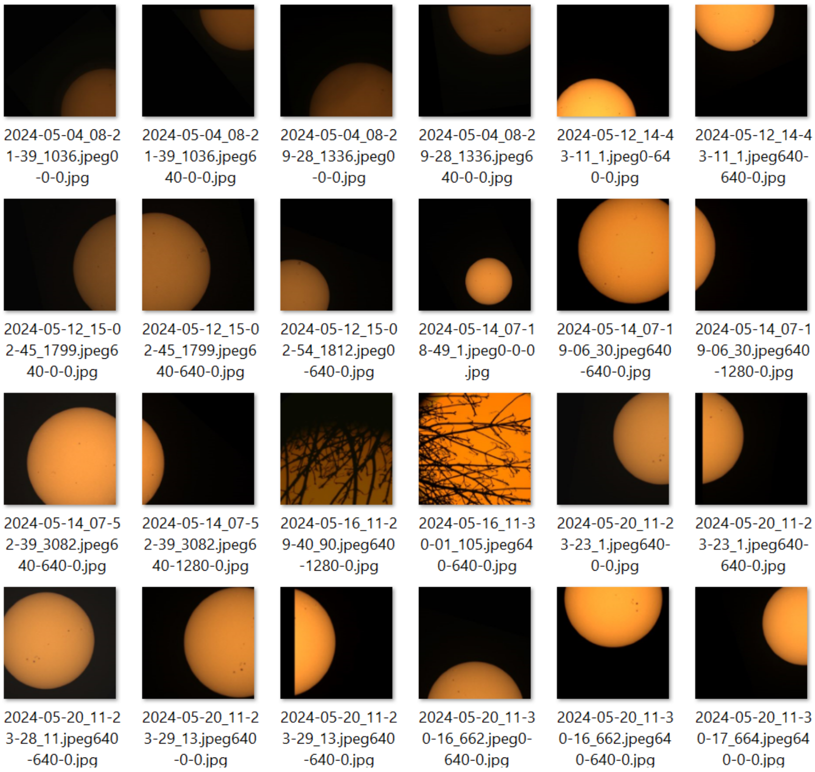

Fig. 3: A snapshot of 24 images present in the test set of SunspotsYoloDataset, without the annotations.

We compared the obtained YOLOv7 models to the YOLOv7 models trained with the dataset called *OGAUC solar images marked for YOLO CNN - cycle 24*, introduced by [14], available from Kaggle [10] and derived from [14]. This annotated dataset was originally used to train YOLOv5 models.

The different training pipelines led to the results shown in Table 1. In this table, we used the following evaluation metrics:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$mAP = \sum_{n} (R_n - R_{n-1}) P_n \tag{3}$$

[10] `https://www.kaggle.com/datasets/aaiisec/ogauc-solar-images-marked-for-yolo-cnn-cycle-24/`

Table 1: Accuracies of different training pipelines based on two datasets, evaluated on the test set of SunspotsYoloDataset, i.e. 128 solar annotated images captured with smart telescopes and with a resolution of 640 x 640 pixels.

Architecture	Training dataset	Precision	Recall	mAP@.5	mAP@.5:.95
YOLOv7	Data introduced by [14]	0.815	0.706	0.792	0.452
YOLOv7-tiny	Data introduced by [14]	0.832	0.747	0.84	0.492
YOLOv7	SunspotsYoloDataset	0.817	0.814	0.88	0.559
YOLOv7-tiny	SunspotsYoloDataset	0.869	0.914	0.915	0.728

From these results, we can make the following observations:

- As expected, the trained YOLOv7 models enable the transformation of a given input solar image into an annotated image with bounding boxes indicating the estimated position of the detected sunspots (Figure 5).
- The best model was obtained with the YOLOv7 tiny architecture (i.e. 607596 parameters), after 300 epochs (Figure 4).
- Despite its larger size, the YOLOv7 model having the classical architecture (i.e. 36481772 parameters) is not better (in terms of Precision, Recall, mAP).
- Unsurprisingly, models trained with SunspotsYoloDataset are more efficient to ignore false positives in solar images disturbed by undesired observation conditions.

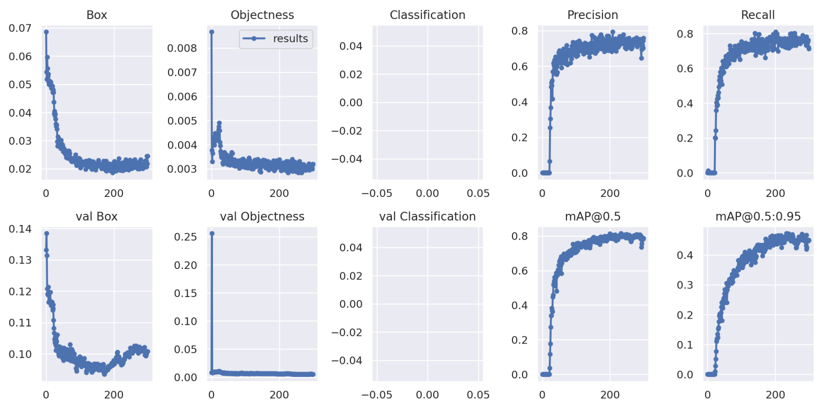

Fig. 4: The evolution of the metrics obtained during the training of the YOLOv7 model on SunspotsYoloDataset.

By using the YoloV7 implementation [18], trained models can by applied on solar images with the following command line (without inference-time data augmentation):

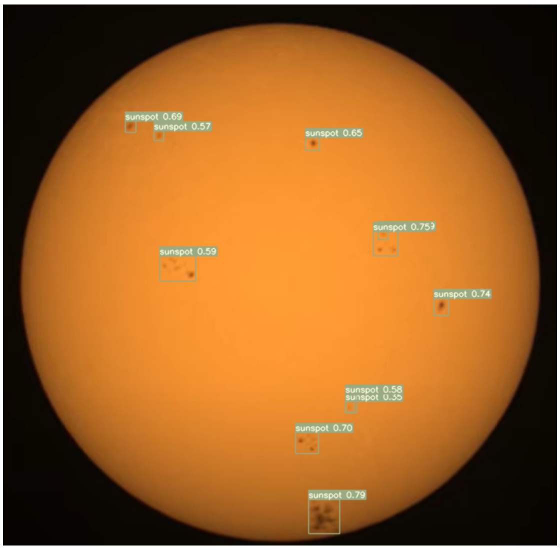

Fig. 5: A solar image captured with a Vespera smart telescope (28/5/2024) and then annotated by the YOLOv7 tiny model trained with SunspotsYoloDataset.

```
python3 detect.py --weights [MODEL PATH]
                  --source [IMAGE PATH OR DIRECTORY]
```

We then applied eXplainable AI techniques to check the robustness of the trained detection models. More precisely, we used Grad-CAM (Gradient-weighted Class Activation Mapping) [16] to visualise the regions of the solar images that contribute most to the sunspots detection. To this end, we ran model inference by using the *pytorch-grad-cam* Python package [11], providing technical support to produce Grad-CAM heatmaps over solar images annotated with YOLOv7 models (Figure 6).

In this way, we were able to check visually when the Yolov7 model correctly detects sunpots, using the right areas of the images, while avoiding false positives by ignoring patterns that are not sunpots (for example, shadows caused by tree branches when the sun is low in the sky) or when images are hazy (due to clouds).

5 Discussion

5.1 Benefits

SunspotsYoloDataset can be used to train detection models directly, as presented in the previous section. Moreover, it can be used as an add-on to and existing other dataset, in order to get more training and/or validation data to obtain better models.

[11] `https://github.com/jacobgil/pytorch-grad-cam`

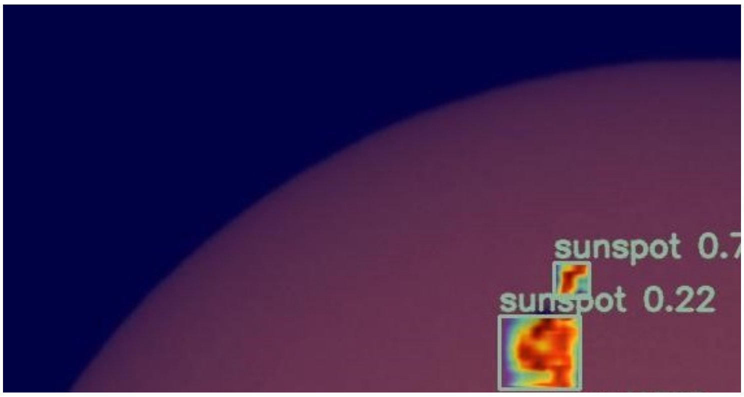

Fig. 6: Heatmap obtained after the processing of a solar image captured with a Vespera smart telescope (27/2/2024), annotated by the YOLOv7 tiny model trained with SunspotsYoloDataset and then analyszed with Grad-CAM.

YOLOv7 models can help to control the acquisition of solar images, with similar observation setups, or with equivalent robotic devices (such as ZWO Asiar [8]). For example, if the YOLO model determines that sunspots are present during a solar observation session, capture can continue to collect more data and notification can be sent to the end user. Conversely, if after a certain period of time no sunspot is detected on the image produced by the instrument, the software controlling the setup can warn the astronomer to stop the acquisition.

5.2 Limitations

SunspotsYoloDataset was obtained with specific equipment (aperture between 50 and 80 mm, focal length between 200 mm and 400 mm, recent CMOS sensors, altazimuth mounts) and imperfect conditions (clouds, obstacles, etc.). Detection models that are obtained from these annotated images can therefore be applied to images obtained with identical equipment or with similar characteristics (i.e. other smart telescopes with similar optical and technical characteristics).

Thus, the application of these detection models to images obtained with instruments of different focal lengths would require the creation of an additional dataset comprising this specific type of data, which would then be used to retrain the models. To take an example, images captured with a significantly longer focal length will have a higher resolution (from an optical point of view), manifested by larger and more detailed sunspots, and the models obtained in this article have not been trained to process such images.

6 Conclusion and perspectives

This paper presents different Deep Learning approaches for sunspot detection on solar data captured with two recent smart telescopes (Stellina and Vespera) and specific solar filters.

Between March 2023 and May 2024, we methodically collected and annotated 2198 solar images with the positions of sunspots that are actually in the images, in order to build a ready-to-train dataset called SunspotsYoloDataset. We trained YOLOv7 models to automatically detect sunspots while ignoring undesired objects like clouds, and we analyzed the results of models with the Grad-CAM technique.

In future research, we aim to build and publish new versions of the dataset after capturing and processing additional solar images with smart telescopes of different characteristics (focal length and sensors), and we plan to work on optimisations to personalise the models presented and reduce their size so that they can be integrated into low-resource devices.

Acknowledgements: This research was funded by the Luxembourg Institute of Science and Technology (LIST), internal grant. Data storage, data processing and detection model training were conducted on the LIST Artificial Intelligence and Data Analytics platform (AIDA), with the assistance of Jean-François Merche and Raynald Jadoul.

Data availability: Annotated solar images can be found in the SunspotsYoloDataset dataset, stored and shared on the Zenodo platform [9]. Additional materials used to support the results of this paper are available from the corresponding author upon request.

References

1. Abed, A.K., Qahwaji, R., Abed, A.: The automated prediction of solar flares from sdo images using deep learning. Advances in Space Research **67**(8), 2544–2557 (2021)
2. Baek, J.H., Kim, S., Choi, S., Park, J., Kim, J., Jo, W., Kim, D.: Solar event detection using deep-learning-based object detection methods. Solar Physics **296**(11), 160 (2021)
3. Chola, C., Benifa, J.B.: Detection and classification of sunspots via deep convolutional neural network. Global Transitions Proceedings **3**(1), 177–182 (2022)
4. Fan, D., Yang, Y., Feng, S., Dai, W., Liang, B., Xiong, J.: Sipnet & sahi: Multiscale sunspot extraction for high-resolution full solar images. Applied Sciences **14**(1), 7 (2023)
5. Hanaoka, Y.: Automated sunspot detection as an alternative to visual observations. Solar Physics **297**(12), 158 (2022)
6. Jenkins, J.L.: The Sun and how to Observe it. Springer (2009)

7. Larsen, K.: Transits, spots, and eclipses: The sun's unique role in outreach. Journal of the American Association of Variable Star Observers (JAAVSO) **49**(1), 90 (2021)
8. O'Brien, M.: A Deep Sky Astrophotography Primer: Creating Stunning Images Is Easier Than You Think! Springer Nature (2023)
9. Parisot, O.: SunspotsYoloDataset: annotated solar images captured with smart telescopes (January 2023 - May 2024) (2024). https://doi.org/10.5281/zenodo.11441091
10. Parisot, O., Bruneau, P., Hitzelberger, P., Krebs, G., Destruel, C.: Improving accessibility for deep sky observation. ERCIM News **2022**(130) (2022)
11. Parisot, O., Jaziri, M.: Assisting observational astronomy for public with smart telescopes and explainable artificial intelligence (2023)
12. Parisot, O., Tamisier, T.: Applying genetic algorithm and image quality assessment for reproducible processing of low-light images. In: Proceedings of the 2nd International Conference on Image Processing and Vision Engineering - Volume 1: IMPROVE,. pp. 189–194. INSTICC, SciTePress (2022)
13. Pérez, J.G., Díaz, S.A., Barrios, E.B., Valencia, J.O., Ortega, J.P., Peña, J.M.: Portable robotic telescope for the teaching of solar astronomy. In: Journal of Physics: Conference Series. vol. 1514, p. 012005. IOP Publishing (2020)
14. Santos, J., Peixinho, N., Barata, T., Pereira, C., Coimbra, A.P., Crisóstomo, M.M., Mendes, M.: Sunspot detection using yolov5 in spectroheliograph h-alpha images. Applied Sciences **13**(10) (2023). https://doi.org/10.3390/app13105833
15. Saville, M.A., Ball, R.C., Harris, G.A., Willenbrink-Sahin, S.S.: Reading in the dark-radar imaging demonstrator for stem outreach of autonomous systems. In: 2022 IEEE Integrated STEM Education Conference (ISEC). pp. 77–83. IEEE (2022)
16. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: IEEE InternationalCconference on Computer Vision. pp. 618–626 (2017)
17. Vaquero, J.M., Vázquez, M., Vaquero, J.M., Vázquez, M.: Naked-eye sunspots. The Sun Recorded Through History: Scientific Data Extracted from Historical Documents pp. 57–102 (2009)
18. Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M.: YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:2207.02696 (2022)
19. Wu, S., Liu, Y., Yang, L., Liu, X., Li, X., Xiang, Y., Gong, Y.: Astroser: Leveraging deep learning for efficient content-based retrieval in massive solar-observation images. Publications of the Astronomical Society of the Pacific **135**(1054), 124501 (2023)