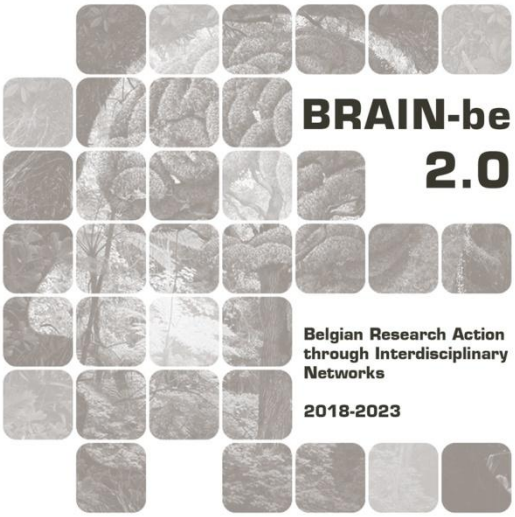
DEEPSUN

Interconnection and valorisation of long-term solar datasets via deep learning

VERONIQUE DELOUILLE (Royal Observatory of Belgium) - CHRISTOPHE DE VLEESCHOUWER (UCLouvain) - NIELS SAYEZ (UCLouvain) – SABRINA BECHET (Royal Observatory of Belgium) – LAURE LEFEVRE (Royal Observatory of Belgium) – OLIVIER LEMAITRE (Royal Observatory of Belgium)

Pillar 2: Heritage science

belspo .be



NETWORK PROJECT

DEEPSUN

Interconnection and valorisation of long-term solar datasets via deep learning

Contract - B2/191/P2/DeepSun

FINAL REPORT

PROMOTORS: VERONIQUE DELOUILLE (Royal Observatory of Belgium)
CHRISTOPHE DE VLEESCHOUWER (UCLouvain)

AUTHORS: NIELS SAYEZ (UCLouvain)
SABRINA BECHET (Royal Observatory of Belgium)
LAURE LEFEVRE (Royal Observatory of Belgium)
OLIVIER LEMAITRE (Royal Observatory of Belgium)

Published in 2025 by the Belgian Science Policy Office
WTCIII
Simon Bolivarlaan 30 bus 7
Boulevard Simon Bolivar 30 bte 7
B-1000 Brussels
Belgium
Tel: +32 (0)2 238 34 11
http://www.belspo.be
http://www.belspo.be/brain-be

Contact person: Georges JAMART
Tel: +32 (0)2 238 36 90

V. Delouille, C. De Vleeschouwer, N. Sayez, S. Bechet, Laure Lefèvre, Olivier Lemaître. '*DeepSun*: *Interconnection and valorisation of long-term solar datasets via deep learning*', Final Report. Brussels: Belgian Science Policy Office 2025 – 41 p. (BRAIN-be 2.0 - (Belgian Research Action through Interdisciplinary Networks))

TABLE OF CONTENTS

ABSTRACT

Context

Sunspots are dark spots appearing in groups on the surface of the Sun as a manifestation of solar magnetism. The magnetic field embedded in sunspots is the driving force behind the solar variability that influences the Earth space environment on a day-to-day basis. The Royal Observatory of Belgium (ROB) is a key player in sunspot observations: In 1939, the ROB Solar station (called 'USET', for Uccle Solar Equatorial Table) started up a solar observing program in collaboration with the Zürich Observatory consisting of daily drawings of the sunspot configuration. This program is still running today. In addition, the USET facility produces white light images, taken since 2002, and CaIIK images, taken since 2012, which will be used in this project. The co-temporal and co-spatial acquisition of drawings, white light, and CaIIK images makes it favourable to interconnect these datasets using novel image processing techniques.

Objectives

The over-arching goal of this project is to produce high-level data products for science data exploitation and dissemination from USET datasets. First, we want to provide a pipeline for the automatic detection, grouping, and classification of sunspot groups from ground-based white light USET images alone. Second, we want to see if it is possible to fill in gaps in historical records by exploiting observations of another correlated modality. Third, we want our results to be used by the community, and also participate in citizen science imitative.

- Objective 1: Automated detection, grouping, and classification of sunspot groups on WL images
- Objective 2: Connection between photospheric (WL) and chromospheric (CaIIK) observations
- Objective 3: Provision of open access of data products produced in the project, dissemination and valorization of the results

Conclusions

For the first objective, we developed SunSCC, an integrated framework for Segmentation, Clustering, and Classification of sunspots from ground-based white light observations. SunSCC uses deep learning models for sunspot segmentation and group classification, connected by a dedicated algorithm for clustering individual sunspots into groups. The framework employs an innovative supervision strategy that alternates between multiple automatically-generated pseudolabels per training sample. This approach shows robustness against atmospheric artifacts, detects typically-missed small sunspots, and minimizes manual annotation to the test set only. The developed clustering methodology leverages sunspot group extents and individual sizes to successfully replicate USET catalog patterns. The proposed CNN architecture for McIntosh classification outperforms existing solutions relying on satellite-based imagery, despite using ground-based observations and the inherent limitations of the subjective McIntosh scheme.

For the second objective, we first investigated solutions based on generative adversarial networks (GANs) for their ability to perform image to image translation, and hence in our case to generate realistic CaIIK images from WL images. We found however that this method introduces physically unfounded artifacts, especially when source and target modalities display substantially distinct characteristics. This led us to propose I2IwFILM, a novel model architecture for solar physics. By combining advanced feature extraction with feature map modulation, this approach enables enriched reconstruction of the target modality through explicitly extracted source features. Evaluations demonstrated that I2IwFILM generates physically consistent predictions with fewer hallucinations compared to GAN-based approaches.

For the third objective, we used virtual observatory approach to distribute our data products. We published 4 papers and attend several conferences. Finally, we devised a citizen science project to provide sunspot classification.

Keywords

Sunspot detection, sunspot grouping, sunspot classification, image to image translation, deep learning, convolutional neural network.

1. INTRODUCTION

Studying sunspots evolution on a long-term basis is a keystone to several areas of Solar Physics, from helioseismology to irradiance modelling and the prediction of space weather. Over the last decades, the **Royal Observatory of Belgium** (ROB) has played an essential role in sunspot observations. In 1939, the ROB Uccle Solar Equatorial Table (USET) station started up a solar observing program consisting of daily drawings of the Sun with the sunspot configuration, a program still running today which constitutes one of the longest and stable series of observations of the Sun. Besides the telescope used to obtain drawings, the USET station also hosts three telescopes equipped with CCD cameras: a telescope for white light (WL) imaging to observe the surface of the Sun (photosphere), a telescope to provide images of the Sun's chromosphere in the Calcium line (denoted 'CaIIK'), and a telescope to observe the Hα line. In this project, we focus on drawings, WL, and CaIIK images. WL images have been taken since 2002, first with a 1024x1024 pixel CCD camera, and in 2012 with an upgraded 048x2048 pixel camera. In 2012, a 2048x2048 CCD camera for the CaIIK line was acquired as well, see Table 1.

The goal of DeepSun was to **exploit and valorize the drawings, WL, and CaIIK images** by leveraging on recent progress made in signal processing, and in particular deep learning.

Type of images	Available since	Drawings since availability	observation days since availability	Drawings, images [2002-2007]	Drawings, images [2012-2019]
Drawings	Mars 1940	23831		1987	2080
WL	February 2002	5135	3895	3465	14680
CaIIK	Augustus 2012	1879	1766	0	16034

Table 1: USET data availability as of July 1, 2019. For CCD cameras, multiple images may be recorded on one day. No images are indicated during [2008-2011] because it was a solar minimum period with few sunspots.

Indeed, there is currently a strong drive from the solar physics community to improve on existing solar image processing methods by taking advantage of progress made in machine and deep learning methods. Most of the time however, existing or preliminary works are dealing with satellite images having a high signal to noise ratio (SNR), making it relatively easy to transfer methods from man-made to solar images. Few works look at long-range data series.

The DeepSun project deals with ground-based images which might have a low SNR and be of inhomogeneous quality. We leverage on advancement happening in the signal processing field and devised the most appropriate deep learning architectures that are adapted to our situation, where adequate pre-processing is essential, and where the chosen algorithms must be robust to noise.

2. STATE OF THE ART AND OBJECTIVES

2.1 Detection, grouping, and classification of sunspots

2.1.1. State of the Art

The last two decades have seen the development of solar information processing methods for the detection, tracking, and classification of solar features (Martens et al. 2012). These methods are typically tailored to the feature of interest, and use manually engineered representation or descriptors of the data.

Accurate **detection of sunspots** is fundamental for the construction of solar activity indices such as the International Sunspot Number (Clette et al., 2007), and constitutes the first step towards sunspot classification. The delineation and classification of sunspots is however partly subjective as two distinct observers might differ in their decisions. Moreover, sunspots are irregular in shape, and have variable intensity and contrast with their surroundings. Ground-based images typically also have different levels of noise and distortion. Detection of a solar feature such as a sunspot relies on the segmentation of an image, which comes in two flavours: region-based or edge-based.

Region-based methods encompass on one hand histogram-based segmentation, where pixels are classified according to their intensities (Steinegger et al., 1998, Pettauer & Brandt, 1997 ; Zharkova et al., 2005) possibly combining this with a Bayesian approach (Turmon et al., 2002), and on the other hand region-growing procedures, where the connectivity of individual pixels is used to incorporate information about the local neighborhood (Muraközy, 2022).

Edge-based methods focus on discontinuities and thus on locating region boundaries. Zharkov et al., 2005 used edge detection, followed by mathematical morphology on continuum image from the Michelson Doppler Imager (MDI, Scherrer et al., 1995) space instrument. They compared their results with NOAA synoptic maps produced manually during one year, and found good correlation. Later on, Curto et al., 2008 as well as Watson et al., 2009 used mathematical morphology on WL images, before applying thresholding to isolate the sunspots. Curto et al., 2008 applied his algorithm on WL images from a ground-based station with the goal to reproduce the Wolf number and to compare it against value obtained manually by operators of the same station. Watson et al., 2009 applied his method on a 1997-2003 MDI continuum dataset to investigate the effect of center to limb variation on sunspot appearance and disappearance.

Carvalho et al., 2020 provided a comparison between intensity-based thresholding and mathematical morphology methods. It showed that mathematical morphology-based approaches perform better than purely intensity-based methods. They also require less pre-processing to remove atmospheric effects in ground-based observations and to correct for limb darkening. These methods are however still very much dependent on the choice of parameters such as the size of the morphological operators. It is therefore of interest to develop a method which is robust against heuristics or parameter's choice.

Once sunspots have been detected, it is necessary to **group the individual sunspots into sunspot groups.** Indeed, sunspots are formed in groups that share physical properties such as belonging to the same magnetic flux loops. Considering a binary mask of sunspots (where '0' means a background pixel

and '1' a sunspot foreground pixel), sunspot groups correspond to local maxima in the spatial distribution of foreground pixels inside this binary mask. Several attempts were made in previous works to find such local maxima. In their work, Curto et al., 2008 used a region growing procedure and predefined criteria of neighborhood to group individual sunspots into a sunspot group. Nguyen et al., 2004 tested various hierarchical algorithms and a simple k-means algorithm.
Since the number of sunspot groups on a given day is not known, the number of clusters *k* to request from the *k*-means algorithm cannot be defined, making this method impractical. Similarly, finding an appropriate termination condition for the hierarchical algorithms proved to be difficult. The same authors then compared in Nguyen et al., 2006 the performance of several hierarchical algorithms with that of a density-based clustering algorithm known as DBSCAN (introduced in Ester et al., 1996), and found a slightly better performance for the latter

Once a sunspot group is detected in successive frames, the problem of **tracking** them reduces to a data association problem, which is generally handled by graph-based solutions. There, a node corresponds to a detection, and an edge represents the link between two regions observed at different times. Verbeeck et al. (2014) propose a simple solution adapted to solar features. More complex methods exist in the computer vision literature, where multi-object tracking is formulated as the problem of partitioning a graph into disjoint sets of nodes (Zamir et al 2012).

Several **sunspot classification schemes** have been developed over the years, starting at the end of the 19th century. The most recent scheme, the McIntosh classification [McIntosh, 1990], is based on the combination of 3 letters *(Z, p,c)* , for a total of 56 admissible McIntosh classes. The number of solar eruptions that occur in a given class is at the basis of several space weather forecasting methods (McIntosh 1990, Leka et al 2019), including the methods used by forecasters from the Regional Warning Center (RWC) at ROB. Attributing a class to a sunspot group is thus fundamental to predict the impact of the Sun on our planet. Most of the time however, it is still being done manually.

Colak & Qahwaji, 2008 combined thresholding and 1-layer artificial neural networks to detect, cluster, and classify sunspot groups according to the McIntosh system. Their work, which represents still a benchmark today, is using both continuum and magnetogram data from the MDI instrument. Abd et al., 2010 used Support Vector Machines on extracted features from MDI continuum images to predict the modified Zurich classification using a one-against all method. They claimed a score of average 80% precision but the method was applied to a restricted dataset (20 samples per class).

The advent of Deep Learning methods allowed computational models that are composed of multiple processing layers to learn representations of data with multiple and hierarchical levels of abstraction (LeCun et al., 2015; Goodfellow et al., 2016). Convolutional Neural Networks (CNNs) have proven superior to previous methods for image classification (Krizhevsky et al., 2012). In sunspot classification, (Knyazeva et al., 2020) used a pre-trained deep neural network and a transfer learning approach to provide an automated McIntosh classification using MDI as well as HMI (Helioseismic and Magnetic Imager, Scherrer et al., 2012) magnetograms. They achieved an average 50% precision score, and significant confusion can be seen between the classes D, E, and F of the modified Zurich class.

Palladino et al., 2022 address sunspot detection and classification using continuum and magnetogram images as well as the Solar Region Summary (SRS) tables of the US Air Force/SOON catalog (SOON ,

2008) as a source of information about the observed sunspot groups. These tables report the location, composition and McIntosh classification of the visible sunspot groups daily. A Faster R-CNN model is used to detect sunspots, and Inception v3 for classification. The precision score varies between 30% and 90% amongst classes, and similarly to Knyazeva et al., 2020, confusion between classes D, E, and F of component Z appears clearly.

2.1.2 Objectives

Our objective is to develop a fully automated system to detect, aggregate, and classify sunspot groups according to the McIntosh scheme using ground-based white light (WL) observations from the USET facility located at the Royal Observatory of Belgium.
We aim at developing a detection method which is robust against particular heuristics or parameter's choice, such as a thresholding level. Regarding the grouping of individual sunspots, we are looking for a statistically sound method. Finally, for the classification of sunspots, we aim at producing an automated McIntosh classification based solely on WL information, where confusion between classes is minimized.

2.2. Connection between photospheric (WL) and chromospheric (CaIIK) observations

2.2.1. State of the art

Deep learning has opened perspectives for new applications in solar physics, in particular in the field of image-to-image translation. Image to image translation have been used in solar physics to mitigate the data scarcity problem, defined as the poor coverage over time, in terms of number of vantage points, or in the observed modality.

One possible way to tackle image-to-image translation is through the adversarial training paradigm. Adversarial training consists in jointly training a pair of CNN networks in competition with each other. These generative models, called 'Generative Adversarial networks' (GANs), learn to capture the statistical distribution of data, and allow to synthesize samples from the learned distribution. We will use conditional GANs (cGANs) which learn a conditional generative model of the data [Isola et al. 2017]. cGANs are suitable for image-to-image translation tasks, where we condition on an input image (e.g. sunspot drawings) and generate a corresponding output image (e.g. sunspot WL image).

In solar physics, cGANs were first used to generate magnetograms out of EUV images, in order to tackle the lack of solar magnetogram on the STEREO mission. Kim et al. (2019) were the first to adopt the Pix2Pix algorithm to produce synthetic magnetograms images from STEREO/EUVI 304 Å data. Their results were limited to a maximum magnetic field strength of ±100 Gauss in the reconstruction and showed low correlations in solar quiet regions. Sun et al. (2023) builds on Pix2Pix, but modifies the U-Net architecture of the generator by adding a self-attention block between the encoder and decoder parts. This allows to put more weight on active regions, and better results were reported. Wang et al. (2018) proposed Pix2PixHD, a modification of Pix2Pix capable of handling a higher dynamic range in the data via a novel adversarial loss and a multi-scale network architecture. Jeong et al. (2020) improved Pix2PixHD to handle ± 3000 Gauss dynamic range in the EUV-to-magnetogram translation, and considered multi-channel input. Jeong et al. (2022) further considered adding a correlation

coefficient (CC) component to the overall objective loss function of the Pix2PixHD model, resulting in a model called Pix2PixCC, whose goal is to provide more realistic backside magnetograms to improve solar coronal magnetic field extrapolations.

Besides the EUV-to-magnetogram cross-modality translation, cGANs were applied in other I2I translation context, such as: the generation of white light (Lawrance et al. 2022) or He I 1083 nm images (Son et al. 2021) from EUV images; the generation of solar magnetograms and EUV images from Galileo sunspot drawings (Lee et al. 2021), and of pseudo-magnetograms from Hα images (Gao et al. 2023) or from CaII-K images (Shin et al. 2020).

Next to adversarial training strategies, diffusion models (DM) have recently emerged as deep generative models that produce new samples by simulating a Markov chain. The Markov chain in DMs starts with pure noise and gradually refines it into a sample that matches the target distribution (Ho et al. 2020). During training of DMs, samples drawn from the target distribution are degraded with noise in what is called a forward process and a denoising model is trained to reverse the degradation in a backward process. DMs have been used in many applications to generate realistic images and have shown promising results in various image synthesis tasks.
In some applications, the backward process of DMs can be guided by integrating the encoded feature of the reference image as input to the U-Net in the reverse process, leading to conditional diffusion models (Choi et al. 2021; Rombach et al. 2022; Saharia et al. 2022).
These methods are well suited to applications where the conditional input has high similarity with the output, such as for inpainting and super-resolution end, but tend to perform poorly in cases where the input and output domains significantly differ, as encountered in solar and heliophysics research.

Although generative deep learning methods have become popular in solar physics, their scientific validity has been so far largely overlooked. In particular, it is known that they may suffer from so-called hallucinations, that is, the generation of features that are not present in the original data.

2.2.2. Objective

Our goal is here is to obtain chromospheric information from photospheric input. We aim at developing an image-to-image translation technique that suffers less from hallucinations than cGANS, and that provides consist3.ent results even when the training dataset is of relatively small size.

3. METHODOLOGY

The DeepSun project was divided into 5 work packages; with the breakdown structure is provided in Figure 1.

- WP1: Management
- WP2: Preparation of USET dataset
- WP3: Static photospheric data exploitation, where we devised an algorithm for the segmentation, clustering and classification of sunspots
- WP4: Dynamic photospheric and chromospheric data exploitation, where we provide cross-modality translation between photospheric and chromospheric images
- WP5: Data access, dissemination, and valorization, whose achievements are described in Section 5 of this report.

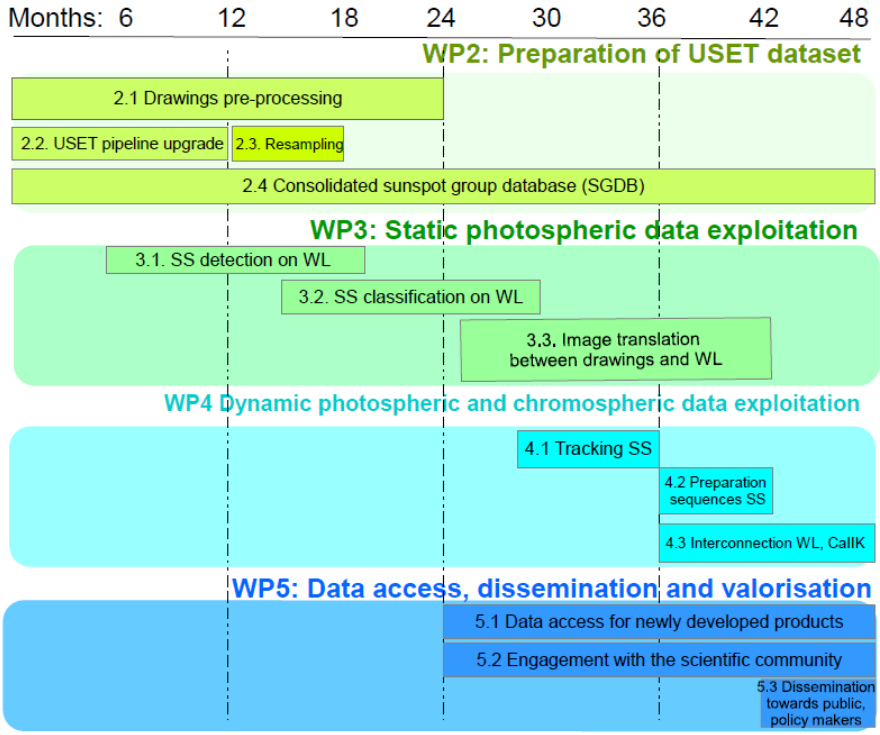


Figure 1 Work breakdown structure and overall planning of WP2 to WP5

3.1. WP1 Management

3.1.1. Task 1.1. Interfacing with BELSPO

The present report constitutes the fourth deliverable of Task 1.1.

3.1.2. Task 1.2. Coordination of the work

Minutes of each meeting of the consortium are recorded on the Gitlab platform (https://gitlab-as.oma.be/SIDC/SILSO_USET/deepsun/). The DeepSun consortium met around every three weeks for progress meetings, and more often when necessary, e.g. when preparing a paper submission.

3.2. WP 2 Preparation of USET data

3.2.1. Task 2.1 Drawing pre-processing: Removing annotations on sunspot drawings

The USET station is equipped with a Visual White light telescope, that projects the image of the Sun on a paper and is used to produce sunspot drawings. Since the 1940s, human operators have made annotations on these drawings regarding the sunspot groups classification and total sunspot counts. These manual annotations need to be removed to make further automated processing possible. We hired job students to do this task with the GIMP editor, freely available at *https://www.gimp.org/*. Drawings from the solar archive recorded between 1954 and 2021 were processed and auxiliary annotations and spurious effects were removed. **(Deliverable of Task 2.1)**

The goal of Task 3.3. was to perform image translation from drawings to WL observations. As explained below, we encounter some serious challenges to achieve this task. Nevertheless, the dataset of pairs of (annotated drawings, non-annotated drawings) produced in Task 2.1 is of high value: it could be used in future projects where a machine learning algorithm would learn how to remove annotations automatically on drawings. This would constitute a great help in the exploitation of all sunspot drawing collections around the world.

3.2.2. Task 2.2 USET pipeline upgrade

We pre-processed the USET White Light (WL) images to accurately center the solar disk within each image, achieving a precision of one pixel. The images were then translated so that the solar disk center aligns with the center of the image frame. This pre-processed dataset is referred to as **L1c**, short for "Level 1 centered." It is openly accessible to registered users via the SOLARNET Virtual Observatory (Mampaey et al., 2025). **(Deliverable 1 of Task 2.2)**

To meet the requirements of WP3 and WP4, we standardized all USET WL images to a uniform resolution of 2048 × 2048 pixels. This adjustment was necessary due to a camera upgrade in June 2008, which resulted in a change in image dimensions—from 1024 × 1024 pixels before the upgrade to 2048 × 2048 pixels afterward. To ensure consistency across the dataset and eliminate issues caused by varying resolutions, we upsampled the earlier images using nearest neighbor interpolation.

We also computed several quality indices for USET images:

- To characterize the diffusion due to high altitude clouds, we compute the median of intensity on and outside of the solar disk.
- To assess the level of atmospheric turbulence we calculate the median filter-gradient similarity (Deng, 2015), which compares the gradient similarity between the raw image and the corresponding filtered image.
- To estimate the presence of clouds we use the cloud cover index (Feng, 2014), we compute an index based on the cross-correlation coefficients between pair of intensity radial profiles for various quadrants of the solar disk. If the Sun is not covered by clouds, the cross-correlation coefficients between each pair are close to one. In presence of clouds, some of the radial profiles are degraded and the cross-correlation coefficients are low.

3.2.3. Task 2.3 Resampling scheme

Some functions now exist in SunPy to resample an image into a coordinate system where the solar rotation is corrected for, and we use these functions in DeepSun.

3.2.4. Task 2.4. Consolidated sunspot group database (SGDB)

From the previous BRAIN project Val-U-Sun, information about drawings have been included into two linked databases:

1. the Drawing Database (DDB), which stores global metadata such as the observer's acronym, total number of spots, sunspot groups, and Wolf number pertaining to one solar drawing,
2. the Sunspot Group Database (SGDB), whose entries contain specific information about a sunspot group, such as its heliographic coordinates, area, or manually made classification in the Zurich and McIntosh systems.

The goal of this Task was to add supplementary information to the SGDB.

- We computed information about the Bounding box (BBox) for each sunspot group from the information about the McIntosh class contained in the SGDB and the value of the solar radius retrieved from the Visual WL observations.
- The SunSCC pipeline developed in this project provides binary maps of Sunspot group, as well as a simplified McIntosh classification. It runs on ROB IT infrastructure.
- The SGDB is integrated into the SIDC Event Database, where a sunspot tracking module has been integrated.

3.3. WP3: detection, clustering, and segmentation of sunspots

We describe methodology adopted to accomplish **Task 3.1 (Detection of sunspot groups)** and **Task 3.2 (Segmentation of sunspot)** as a whole, since we gather these two tasks, together with the clustering of sunspots, in a common pipeline.

The SunSCC pipeline depicted in Figure 2 takes as input images obtained from the USET White Light telescope and prepared according to Task 2.2. It returns as output segmentation masks of individual sunspots, as well as the sunspot groups and their McIntosh classification, with a reliability score. It is divided in three independent blocks as follows. The first block, dealing with the sunspot segmentation, is addressed by a CNN in the form of Unet architecture (Ronneberger et al., 2015) composed of an auto-encoder with skip connections. The second block receives as input the segmentation from the first block, and aggregates the detected sunspots into sunspot groups using an original clustering method inspired by the mean-shift algorithm. Each identified sunspot group is provided along with an angular distance map to the third block. The latter constitutes a classification network designed to predict the McIntosh 3-{components} classification. It is composed of a ResNet34 (He et al., 2016) iimage encoder and three Multi-Layer Perceptrons (MLP) organized in a hierachical way, each having 4 hidden layers and Rectified Linear Units (ReLU) activation functions (Agarap, 2019). We now detail the construction of each block.

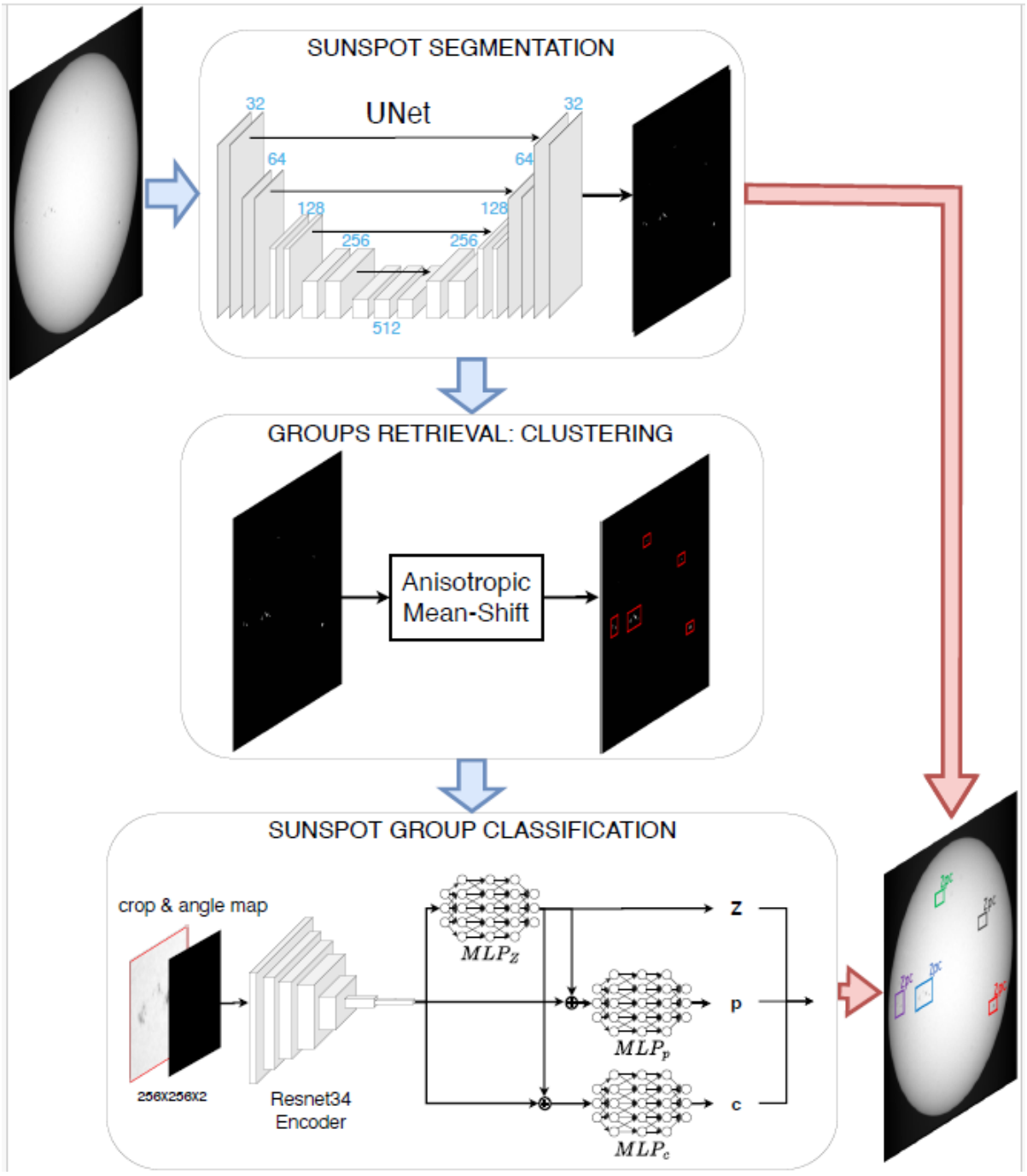


Figure 2: SunSCC pipeline for sunspot segmentation, clustering, and classification. Full-disc images 2048x2048 images are subdivided into a 4x4 grid, each containing segments of dimensions 512x512. A U-net segmentation network predicts masks for each segment that are reassembled into a 2048x2048 segmentation mask. The detected sunspots are aggregated into sunspot groups by a modified mean-shift algorithm. Each identified group is provided along with an angular distance map to a classification network composed of a Resnet34 image encoder and three Multi-Layer Perceptron's with 4 hidden layers and ReLU activation function. Each MLP is specialized in the classification of one component in the McIntosh system.

3.3.1. Task 3.1. Detection and segmentation of sunspots on WL images

We choose to first build a set of unsupervised but error-prone segmentation masks, using conventional image processing techniques, and next used those masks as pseudo-labels to train a CNN model.

To generate unsupervised segmentation masks, also called pseudo-label', we first correct the image for limb darkening. Next, mathematical morphology operations, in the form of a black top-hat transform is applied, similarly to (Watson et al., 2009). This allows to find dark areas surrounded by a bright environment corresponding to sunspots. Finally, a thresholding is applied to obtain a binary mask.

We then train the CNN with stochastic pseudo-labels as follows: We adopt a U-Net architecture (He et al, 2016) for our CNN segmentation model since it has proved its accuracy in delineating high resolution segmentation masks in various fields such as medical imaging and natural images analysis.

Since the smallest sunspots represent only a few pixels and thus may disappear when down-sampling large images, we feed the CNN at full original resolution but use patches of the full image (called \lq segment') of size 512 x 512 pixels in our segmentation models. Note that unlike during the generation of the segmentation masks by thresholding methods in previous section, the limb darkening effect is not corrected on the images given as input to the CNN. Instead, raw white light images are used, and we count on the regularization properties of the CNN to be robust to limb darkening and the presence of clouds.

Finding a single threshold value during the mask generation process is a non-trivial task as the level of illumination may differ not only locally due to the presence of clouds and other artifact source such as atmospheric seeing, but it may also differ from image to image. Moreover, CNNs trained with a systematic labelling error are known to produce predictions with the same error patterns. To circumvent these issues, we consider multiple thresholds per WL images, each of them generating a specific mask or pseudo-label, and we train our model using these various pseudo-labels. This is expected to increase the error patterns variability, thereby preventing the CNN to learn them. Our experimental results demonstrate that this original strategy is effective in mitigating the bias introduced by pseudo-labels in the trained model prediction behaviour.

Clustering of sunspots (added Task)

Once we have a binary segmentation at the pixel level, we need to group together pixels belonging to a same sunspot group, before being able to classify this group.

We devised an original method to find sunspot groups in a WL image, given the sunspot masks provided by our CNN segmentation model. Our approach adapts to the context of sunspot group retrieval the mean-shift clustering algorithm introduced by (Fukunaga & Hostetler, 1975) and popularized in computer vision by (Comaniciu & Meer, 2002). The mean-shift clustering algorithm exploits local gradients of a density function to locate its maxima, or modes, given discrete data samples from that function. In the sunspot clustering context, the procedure to discover the modes, and to associate the observed sunspot to those modes, works as follows. Starting from the center of an observed sunspot, it iteratively shifts according to the gradient of the density function obtained by placing at the center of each observed sunspot a kernel shape weighted by the area of the sunspot. This is repeated until convergence to a local maximum, defining a mode. The convergence procedure is repeated for all the sunspots, which ends up in defining a discrete number of modes, to which sunspots are associated.

The rationale for choosing this method is that it adopts a conventional linear kernel, but takes the shape of sunspot group into account, as well as the aspect of individual sunspots when defining the kernel spatial support in longitudinal and latitudinal directions.

3.2.2. Task 3.2 Classification of sunspots

Input to the classification task are the sunspot groups retrieved by the clustering algorithm. However, sunspot groups that are close to one another deserve a specific treatment, to avoid their mixed patterns misleading the classification network during training and validation. Such cases are identified as groups whose bounding box image contains sunspots belonging to other group(s).
For each of these groups, before presenting the corresponding image to the classifier, we erase the sunspots of other close group(s) by setting their pixel intensity values to the mean value of the background pixels around their contour. In that way, only one group is available to the classifier at a time.

In the training dataset used in this supervised classification task, each of the samples gets an associated label that corresponds to the McIntosh class of the matched sunspot group in the Sunspot group database (SGDB).

The McIntosh classification scheme contains as much as 56 admissible classes, and therefore some of them are poorly observed in our training dataset. On the other hand, some classes have similar morphology and are differentiated only by their sizes. We decided to use a reduced set of classes, as depicted in Figure 3, to tackle imbalance in class sample size. Below we explain our CNN-based classification, and how we used ensemble classifier to get a measure of confidence in the classification prediction.

CNN-based classification

As shown in Figure 2, our classifier relies on a ResNet34 convolutional backbone to encode the visual information associated to the sunspot group to be classified. The network is fed with standard-sized 256x 256 images for all sunspot groups. The cropped images are composed of two channels: the first channel contains the square segment of the WL image centered at center location of the sunspot group formed during clustering, and the second channel contains the corresponding segment of an angular distance map aligned with the WL image. This second channel provide information about the location of the sunspot group on the Sun so that knowledge about distortion of sunspots shape approaching the limb can be learned.

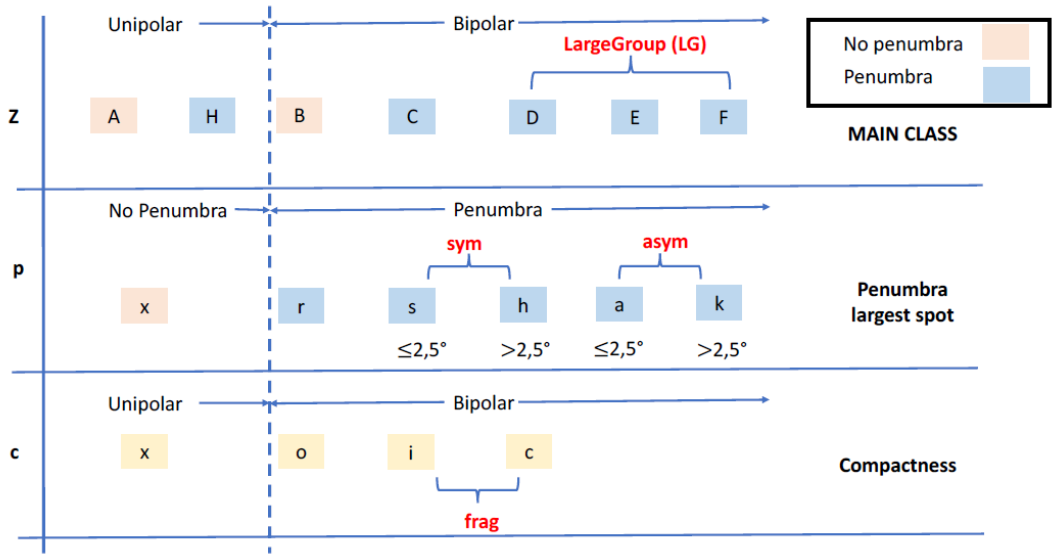


Figure 3: Reduced McIntosh classification scheme used in SunSCC. Red fonts indicate how some of the initial McIntosh classes were merged in a single class.

The embedding representation of the images provided to the ResNet34 backbone is given as input to three Multi-layer perceptrons (MLPs) named MLP_Z, MLP_p and MLP_c, respectively specialized in the classification of each of the Z, p and c component of the McIntosh system. The MLPs are organized hierarchically as shown in Figure Full Pipeline. The objective of this architecture is to mimic the dependency of the p- and c-component classifications on Z.

The classification network was trained using the cross-entropy loss and Adam optimization algorithm (Paske et al, 2019).

From hard to soft predictions using ensemble of classifiers

We applied the precepts of *ensemble learning* (Dietterich, 2000; Peake et al., 2020) to soften the prediction of a sample by a single classification network into a confidence measure of belonging to a particular class. Indeed, the biases of a single classifier are likely to be compensated by biases of other classifiers when combining their predictions, and hence the ensemble prediction accuracy is expected to surpass that of a single classifier (see Dietterich, 2000 for a review).

In our experiments, we initialized with different seeds multiple instances of the classification network, and each of these instances were trained independently. This produced multiple classifiers, hence for each sunspot group, we obtain a pool of predictions. These predictions are then combined into one single output via a majority vote.

Our results show that the class prediction inconsistency across the ensemble of classifiers is a valuable clue to identify the cases for which the majority vote prediction is likely wrong (see Section 4).

Deliverables: The SunSCC algorithm and its results on USET dataset has been published in Sayez et al 2023; the codes and scripts have been adapted to run on the IT infrastructure at ROB.

3.3.3. Task 3.3. Image translation between drawings and WL images

Translation algorithms perform best when the coordinates of the input and target images are aligned—what is known as pixel-to-pixel translation. To achieve this, sunspot drawings need to be registered to the corresponding White Light (WL) images.

As a first step, we examined the uncertainties in sunspot group positions within the drawings. We observed distortions in the shapes of the sunspots, with these deformations becoming more pronounced toward the solar limbs. Additional sources of misalignment—such as those introduced by variations in optical systems—further complicate the process. These factors make the registration non-rigid and inherently challenging. Due to these complexities, we chose to focus instead on translations between CCD photospheric images and CCD chromospheric images, where alignment is more straightforward. (cfr Task 4.3).

3.4. WP4: Dynamic photospheric and chromospheric data exploitation

3.4.1. Task 4.1 Tracking of WL sunspot group

A sunspot tracking module has been integrated into the SIDC Event Database, based on a graph approach.

3.4.2. Task 4.2 Preparation of sequence of sunspots

Various quality indices have been computed on USET dataset, as explained in Task 2.2. For now, these indices are not combined into a single index, but it is possible to select a dataset based on the high value of quality index provided by the USET operator, which we did. In Task 4.3, we will make the interconnection between pairs of WL and CaIIK images, but not between sequences of WL and CaIIK images. Integrating time information into image-to-image translation is far from trivial and will be considered in future work.

3.4.2. Task 4.3. Interconnection between WL and CAII K images

We explain here the pre-processing step, and the methods we used translate WL images into CaII K images.

3.4.2.1. Dataset preparation

We consider the USET dataset that simultaneously monitors the photosphere in WL and the lower chromosphere in Ca II K since 2012. We selected daily pairs of (WL, Ca II K) images, which are of size 2048 × 2048, through a careful process: highest-quality daily WL images are identified using the quality index provided by the USET operator when making the sunspot drawing, then images are paired with the best available Ca II K images, minimizing the time difference between them. A maximum temporal difference of four hours is considered to avoid discrepancies due to too large displacements. For each pair, de-rotation is applied to co-align WL and CaIIK images using the SunPy module [The Sunpy Community, 2020].

The (WL, Ca II K) image pairs then undergo further processing as follows: sub-image of size 512 × 512, called ‘segments’, are extracted from the original full disc images. These segments are centered at sunspot locations, obtained from the USET sunspot group catalog Table Access Protocol (TAP) service (Bechet & Clette, 2024). To reduce computational cost, the segments are resized to 256 × 256 pixels. Figure 4 shows an example of (WL, CaII-K) segment pair.

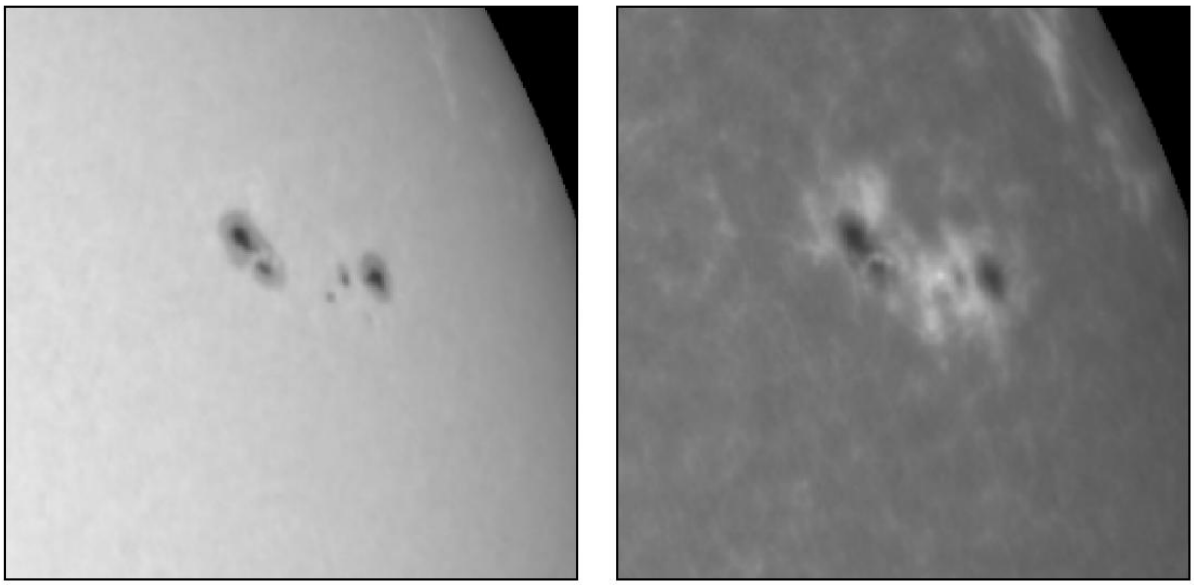


Figure 4. Sample pair of WL and CaII-K segments, which have been spatially aligned and centered near an active region.

3.4.2.1. Translation from WL to Ca II K images

We considered two approaches to translate images: one based on conditional adversarial network , and one based on non-adversarial strategies. We explain these methods and next we introduce our train-test data splitting strategy.

Conditional adversarial network approach

We first adopted the conditional generative adversarial network approach (cGAN) proposed in Jeong et al. (2022) and called Pix2pixCC. It is composed of two sub-networks trained simultaneously: a generator *G* and a discriminator *D*. The generator is a CNN shaped according to the U-Net architecture (Ronneberger et al., 2015) while the discriminator is implemented as a patchGAN (Isola et al., 2017) which manipulates pairs of patches (cropped in the pair of full images) instead of the image pair itself. Figure 4 depicts Pix2pixCC in the context of WL-to-CaIIK translation.

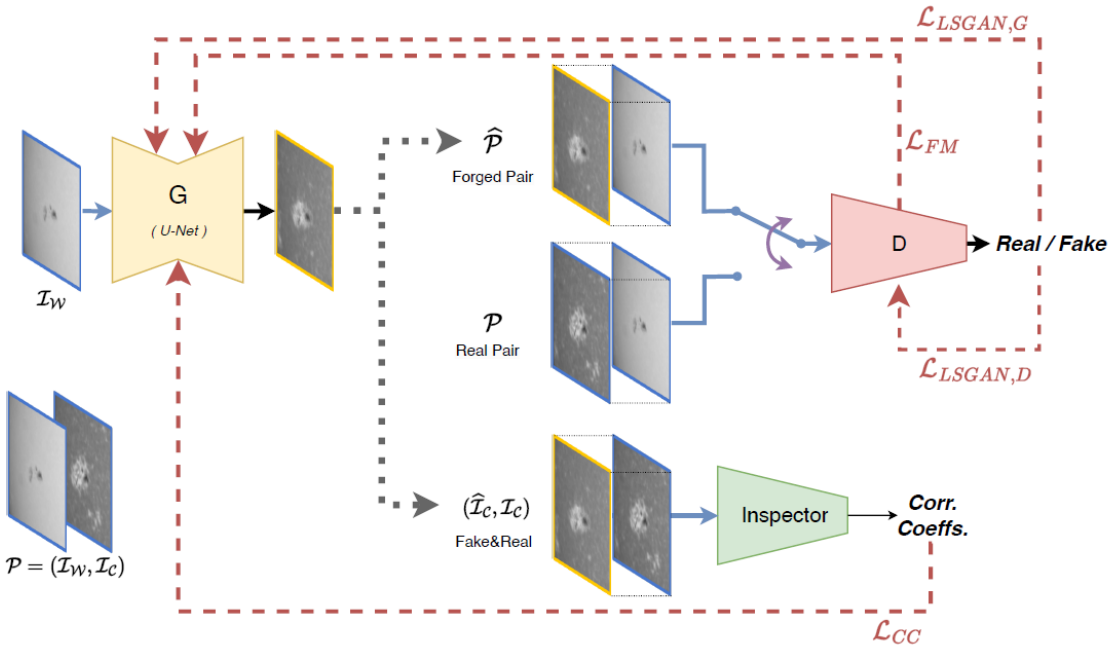


Figure 5: Diagram of Pix2PixCC model and adversarial training for WL-to-CaIIK translation. The generator G converts a WL image to the CaIIK modality, the discriminator D distinguishes original WL-CaIIK pairs from forged pairs. The 'Inspector' module contributes to the update of the weights of G by computing correlation coefficients between original CaIIK image and forged CaIIK image, at multiple scales.

The network receives as input a white light view and outputs an estimation of the corresponding calcium view. The generator is trained to make it close to the original calcium image, when available. Besides, the discriminator *D* is given a pair of images that can either be the original pair *P* or the forged pair $\hat{P}$. The role of *D* is to distinguish real pairs from the forged ones.

During training, *G* is updated to improve the ability to fool *D* using a weighted sum of three loss functions:

a) the least-squares adversarial loss function for generators (Mao et al. 2017)
b) a feature-matching (FM) loss
c) a correlation coefficient based loss based on the computation from the Inspector.

Our experiments reveal that this model results in a significant amount of hallucinations, and hence in a second stage we also looked at non-adversarial strategies.

Non-adversarial strategies

As the adversarial setup of GAN models often generates outputs with hallucinations, we tested several approaches to mitigate this issue in our physics-sensitive application. First, we considered a variant of Pix2PixCC trained without adversarial loss or feature modulation and utilizing an L1 loss between the reconstruction and the target. We denote it $\text{Pix2Pix}_{\text{w/o-Adv}}$.

Second, to mitigate unwanted artifacts in our physics-sensitive application, we propose an alternative approach that utilizes an I2I translation model with enhanced capacity to adapt the output reconstruction to the input content. This algorithm modulates the model's internal feature maps based on **Feature-wise Linear Modulation** (FiLM, Perez et al.), thereby offering the network the ability to inject input-related information through the image translation process.

Formally, image-to-image translation with feature-wise linear modulation (I2IwFiLM) is implemented as depicted in Figure 6. To increase the chance that the guidance vector captures information that is relevant to the output, in a preliminary training stage, the guidance vector $\mathcal{E}^P$ is defined based on the complete WL and CaIIK pair *P*. Specifically, an encoding convolutional neural network denoted GVP_P, and named Guidance Vector Prediction network is trained to extract a 256-dimensional guidance vector $\mathcal{E}^P$ from the concatenated input and output signals. The guidance vector $\mathcal{E}^P$ modulates the output of several convolutional layers in the main U-Net through dedicated linear layers. Each modulated convolutional layer is paired with its own linear layer which takes $\mathcal{E}^P$ as input and produces an additive modulation signal that is combined with the convolutional layer's output. Then, in a second step, another CNN is trained to predict the guidance vector $\mathcal{E}^P$ from the input (WL) signal only.

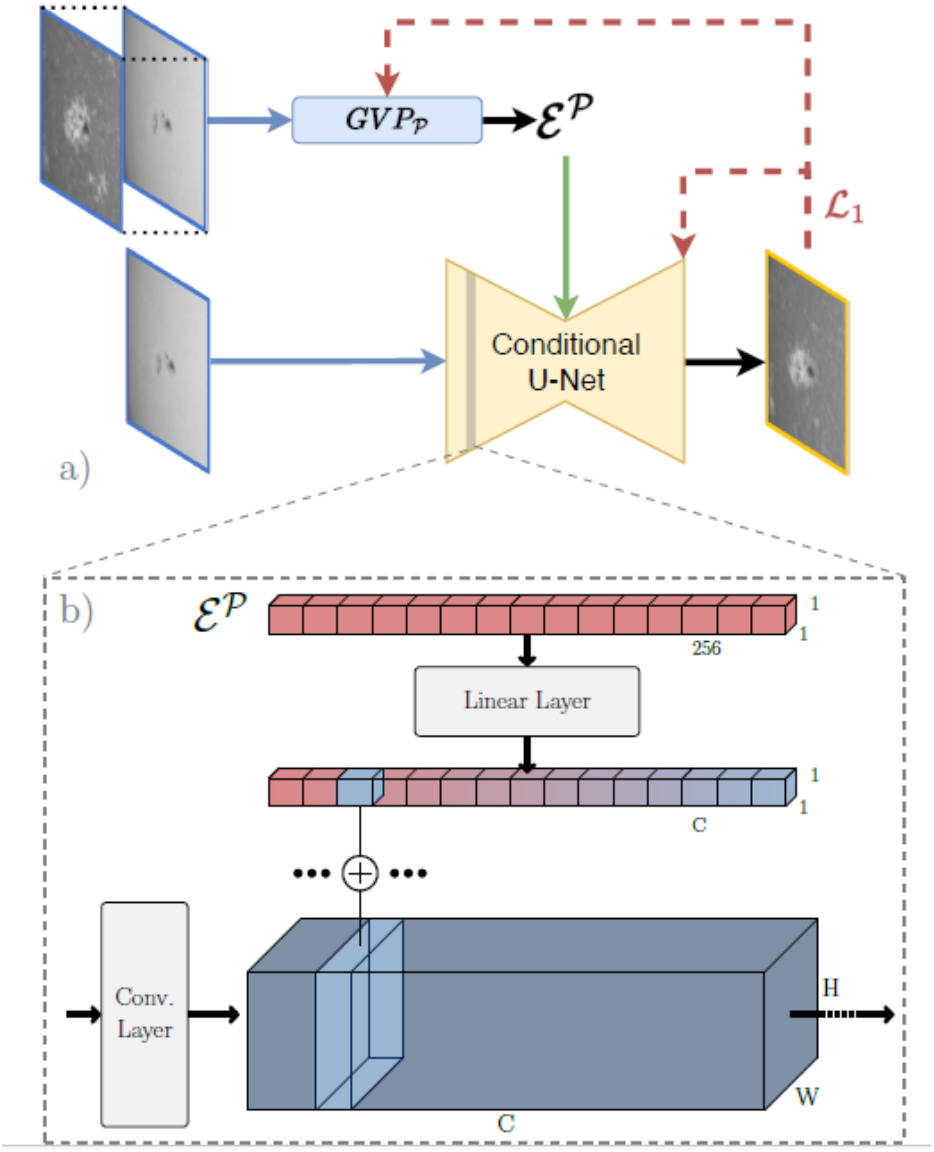


Figure 6: First training stage of I2IwFiLM: a) Main U-net model predicts Calcium image from input WL image and from $\mathcal{E}^P$, the guidance vector extracted from the pairs of WL and CaII K images using a Guidance Vector Prediction network denoted GVPp. b) Details on the modulation of an intermediate feature map of the U-net model by $\mathcal{E}^P$, using a FiLM layer. The height, width and number of channels (depth) of the intermediate feature map are denoted W, H and C, respectively.

Train-test data splitting

To create the test sets of our translation task, two months were selected from each year for both datasets: one winter month and one summer month, to ensure sample diversity. All samples corresponding to these selected months were allocated to the respective test sets. All other samples were selected for training, except the ones falling within a 12-day window on either side of the test samples. The window size was chosen as half of the Sun's rotation period. This results in a WL-to-CaIIK translation dataset including 6,387 samples for training and 1,410 for testing.

Deliverables: This work has been submitted to *Astronomy & Astrophysics*, see Sayez et al.,(2025) and is currently under revision.

4. SCIENTIFIC RESULTS AND RECOMMENDATIONS

4.1 Segmentation, clustering, and classification of sunspots

4.1.1. Segmentation

We used a manually annotated dataset as ground truth to evaluate the quality of the segmentation. A classical metric is the F1-score, defined as: *F1-score = 2TP / (2TP+FN+FP)*, where TP('True Positive') is the number of correct sunspot detections, FN ('False negative') is the number of missed detections and FP('False Positive') is the number of undesired detections. The F1-score thus decreases when there are undetected sunspots (FN) or erroneous sunspot detection (FP).

We compared the performance of three methods:

1. Unsupervised threshold-based segementation
2. CNN-based segmentation that is trained with one pseudo-label corresponding to one threshold value
3. CNN-based segmentation trained with three pseudo-labels per WL images, each pseudo-label being obtained with a different threshold, and being selected alternatively with equal chance of training.

As an illustration of the relative performance of these three methods, we show in Figure 7 the trade-off between the number of non-detected sunspots and the number of merged sunspots - i.e. fusions - for each threshold-based segmenter and CNN-based segmenter. One can observe that CNN-based segmentation regularises sunspot detections since the number of misdetections is significantly reduced when compared with thresholding. The segmentation using three alternating pseudo-labels per WL image during training achieved performance comparable to that of the best CNN-segmenter using a single pseudo-label, both in terms of F1-score and in terms of fusion versus non-detected sunspots trade-off.

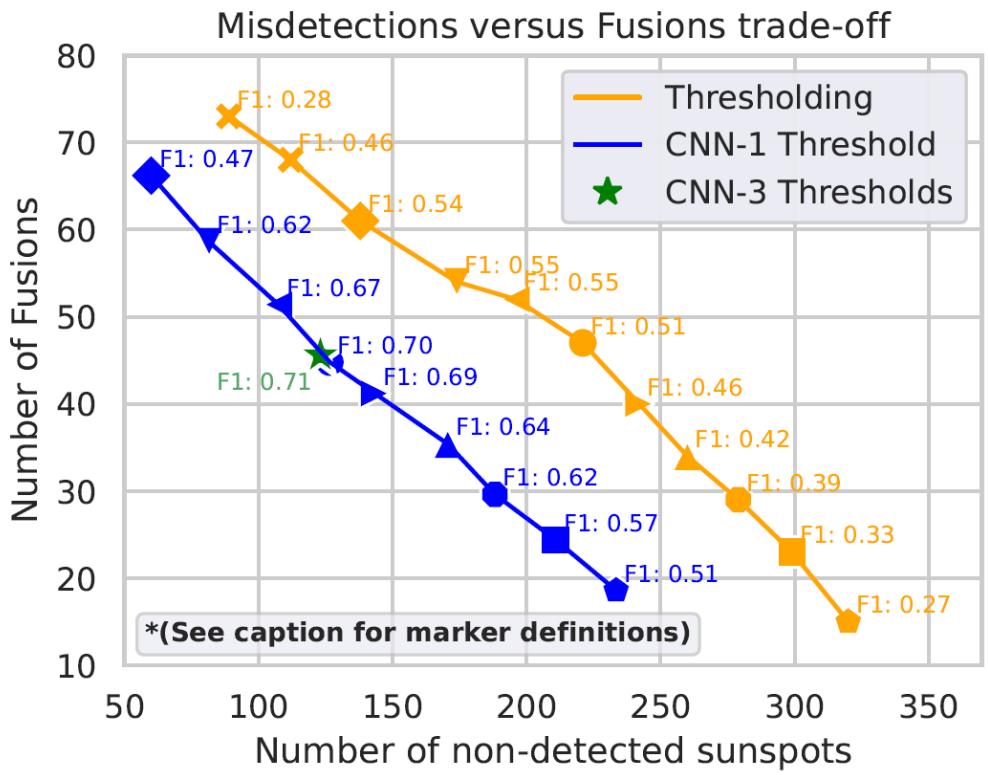


Figure 7 R Trade-off between sunspots remaining undetected and sunspots being merged in a single entity. The thresholding method (orange) labels as sunspot the pixels lying above a threshold value in the Black-TopHat transformed image. CNN-1 (blue) corresponds to the models whose training has been supervised with pseudo-labels obtained with a single and fixed threshold. CNN-3 is trained with 3 distinct pseudo-label images per WL image (green). Markers indicate the threshold used (diamond: τ = 300, down triangle: τ = 325, left triangle: τ = 350, circle: τ = 375, right triangle: τ = 400, up triangle: τ = 425, hexagon: τ = 450, square: τ = 475, pentagon: τ = 500, star: τ1 = 325 τ2 = 375 τ3 = 425). The F1 scores indicate performance on the smallest sunspots (size boundaries < 50 pixels), which are the most challenging to detect.

4.1.2. Clustering

To retrieve sunspot groups, we use the masks produced with the best performing CNN classifier in our segmentation experiments.

The performance of the proposed clustering method is done by measuring the concordance between clusters found by the automatic method, and the SGDB entries. To quantify this concordance, bounding boxes computed from SGDB entries are compared with bounding boxes of groups found by clustering. We performed a grid search to find appropriate values for the latitude bandwidth as well as maximum longitude bandwidth to define the kernels used for clustering. This reveals that our method achieves 80% accuracy when appropriate mean-shift parameters are adopted.

4.1.3. Classification

In our experiments, we used 85% of the available samples for training, and 15% for testing. Each test sample was provided to every CNN such that a majority vote could be performed on the predictions of each component Z, p and c.

Figure 8 shows confusion matrices presenting the performances of the classification of the Z, p and c components. Formally, each entry $\boldsymbol{M}(i, j)$ of the confusion matrix $\boldsymbol{M}$ denotes the rate at which the ground-truth class i is assigned to class j. The diagonal elements represent the per-class recall values, defined as TP/(TP+FN), where in this context TP (resp. FN) indicates a correctly (resp. uncorrectly)

predicted class. On these matrices, one can observe that the majority votes lead to a significant number of correct predictions. Indeed, the diagonal entries have high values and the off-diagonal entries are low or show consistency with possible confusion between 'adjacent' classes in the lifecycle of sunspot groups [McIntosh, 1990]. For instance, confusion between class B and class C and between C and LG for the Z component is expected due to the intermediate status of class C between class B and the classes grouped under LG. The same observation can be done regarding the confusion between class r and others for the p component. Such confusions could be done by a human annotator as well, depending on the meticulousness of their observation of both poles, or their interpretation of a 'rudimentary' penumbra. Therefore, due to the definition of the McIntosh classes themselves, any classification method will present minor confusion cases, up to some extent.

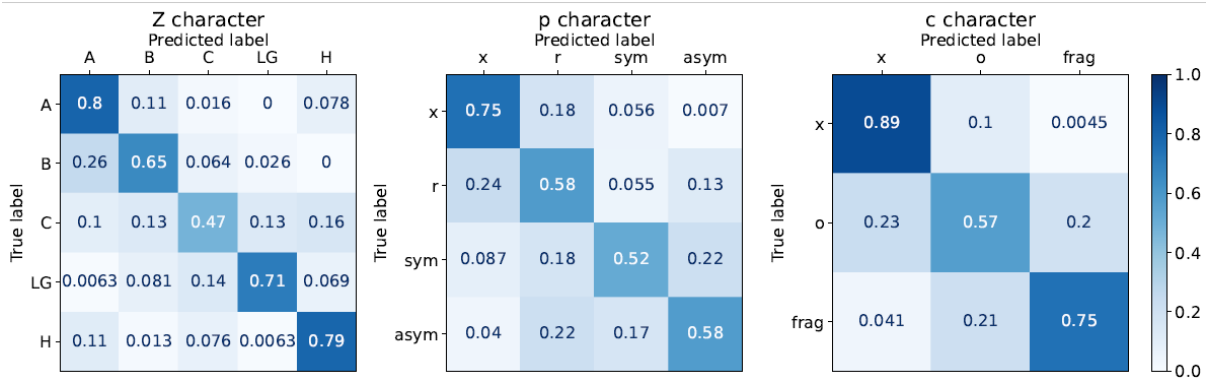


Figure 8 Test set confusion matrices of classification with majority vote for the classification of the Z (left), p (center), c (right) components. In each matrix, the rows represent the true label of test samples while columns represent the label predicted by majority vote. The matrices are row-normalized such that per-class recall values are shown on the diagonals.

Finally, our experiment also showed that:

- The performance of CNN-classifier decreases near the limb, even if some information about limb is provided to the algorithm,
- The CNN-classifier is able to find annotation errors in the SGDB,
- The ensemble classifier allows to identify ambiguous cases in sunspot classification.

4.2 Interconnection between White light and Ca II K images

To evaluate the image-to-image translation model, we used both generic metrics and application-specific metrics.

4.2.1. Generic metrics

Table 2 shows the reconstruction performances using the *root mean squared error* (RMSE), peak *signal-to-noise ratio* (PSNR), structure *similarity index measure* (SSIM) and a *multi-scale* extension of SSIM (MS-SSIM). All these metrics are common in the context of image reconstruction.

As we can see in Table 2, removing adversarial training indeed improves reconstruction performance: $\text{Pix2PixCC}_{\text{w/o-Adv}}$ outperforms Pix2PixCC across all generic metrics. Furthermore, incorporating modulation based on FiLM principles further enhances translation quality.

Task	Model	RMSE ↓	PSNR [dB] ↑	SSIM ↑	MS-SSIM ↑
WL-to-CaIIK	Pix2PixCC	16.77 ± 8.01	24.51 ± 3.83	0.72 ± 0.09	0.75 ± 0.12
	$\text{Pix2PixCC}_{\text{w/o-Adv}}$	15.07 ± 7.74	25.56 ± 4.10	0.75 ± 0.07	0.80 ± 0.10
	I2IwFiLM (Ours)	14.44 ± 7.07	25.90 ± 4.12	0.80 ± 0.09	0.80 ± 0.11

Table 2: Quantitative results of generic image reconstruction metrics for **white light-to-calcium** translation tasks. GAN-based model Pix2PixCC is compared with non-adversarial models: $\text{Pix2PixCC}_{\text{w/o-Adv}}$ and our proposed I2IwFiLM including input-dependent conditioning.

4.2.2. Application-specific performance metrics

Figure 9 shows examples of WL to CaIIK modality translation of high quality, when the sunspot was situated close to the limb. It shows typical model behaviors: Pix2PixCC, trained with adversarial setup, produces visually appealing results with high-frequency features but exhibits a tendency to generate plage regions where there are none, a behaviour termed as 'hallucination'.

In contrast, I2IwFiLM generates outputs that appear blurrier but achieve more physically accurate reconstructions without hallucinations.

In this application, we also observe a variation in translation quality with respect to the angular $\sin(\theta)$ distance between sun centre and sunspot location: regions close to the center show poor translation performance, while regions with $\sin(\theta) > 0.6$ exhibit increasingly accurate reconstructions as they approach the limb. This is explained by the fact that the contrast between features belonging increases with $\sin(\theta)$. Hence faculae, which are visible in WL images and have a physical correspondence with the plage and extended network in Ca II K images, are more visible close to the limb, and are barely visible near the sun center.

We computed application-specific, physically relevant, metrics to compare the performances of the methods, using the plage segmentation masks computed from the original CaIIK and from the AI-generated CaIIK images. First, we used the *Plage Pixel Ratio* (PPR), defined as the ratio of plage pixels in generated and original images and second, the *Intersection over Union* (IoU) was computed as the ratio between the number of pixels that belong to both masks and the number of pixels that belong to at least one mask.

Figure 10 represents histograms of PPR and IoU metrics on test set samples for 3 ranges of $\sin(\theta)$: (a) the center of the Sun ($\sin(\theta) \leq 0.5$) where faculae are indistinguishable from the background in WL images, (b) the transitional region ($0.5 \leq \sin(\theta) \leq 0.75$), and (c) the region nearing the limb ($0.75 \leq \sin(\theta) \leq 1$) with high contrast between faculae and background.

Both Pix2PixCC and I2IwFiLM models demonstrated improved reconstruction performance as $\sin(\theta)$ increases, with higher mean PPR and IoU values and the reduction in their standard deviations. However, both models appear to be limited by the visibility of faculae, resulting in the production of plage regions smaller than the actual ones in the targets. This limitation is revealed by both PPR and IoU achieving their best values in the regions where faculae are most easily discernible.

Examining the models individually, I2IwFiLM exhibits low average PPR and IoU with large variance for $ $\sin(\theta) \leq 0.5$$, indicating that only a fraction of plages in these active regions are consistently reconstructed. In contrast, Pix2PixCC achieves a respectable PPR but poor IoU for these regions, suggesting consistent incorrect reconstruction of plages due to hallucinations.

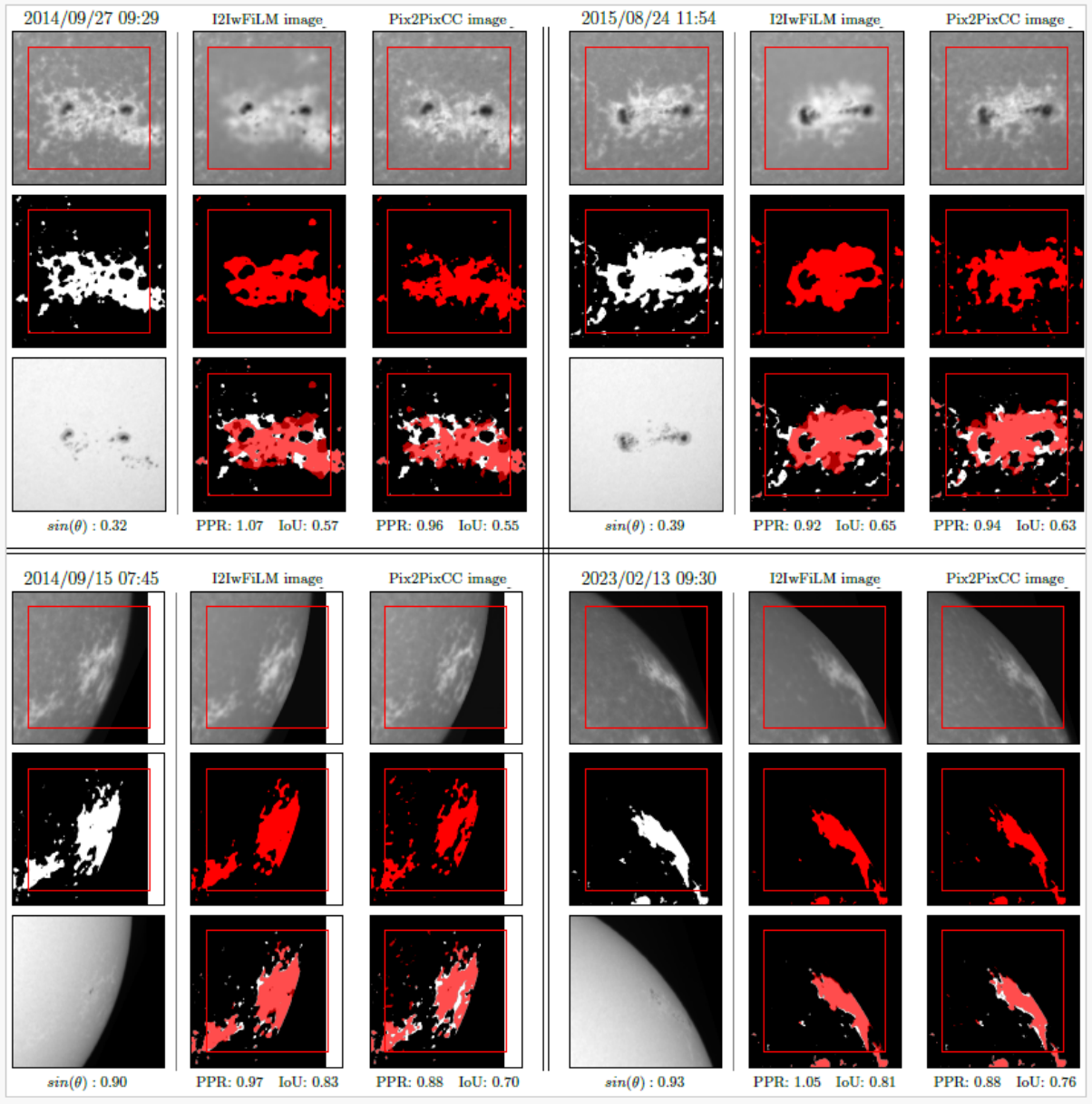


Figure 9 Ca II K reconstructions of plages near the center of the solar disc (Top grids) and near the limb (Bottom grids). In each 3x3 grid: the top row shows the original CaIIK image (left) and its reconstructions by I2IwFiLM (middle) and Pix2PixCC (right). The middle row shows segmentation masks of plages visible in the original CaIIK image (left) and the corresponding reconstructions by I2IwFiLM (middle) and Pix2PixCC (right). Bottom row shows input white light (WL) image (left) and overlays of the target plage mask with the model-generated plage masks from I2IwFiLM (middle) and Pix2PixCC (right), facilitating a direct comparison. The red square corresponds to the sub-image provided as input to the models; a slightly larger image is shown to indicate the consistency with the surroundings.

When comparing the relative performance of the two models, it is noteworthy that the mean PPR of I2IwFiLM initially lags behind that of Pix2PixCC for $\sin(\theta) \leq 0.5$ but surpasses it for $\sin(\theta) > 0.75$, accompanied by a substantial reduction in its standard deviation. Given these results, I2IwFiLM appears to be more suitable for white light-to-calcium translation compared to Pix2PixCC.

These results illustrate the possibility of reconstructing CaIIK plages from faculae observed in WL when these are located at an angular distances of about 40 degrees or more from the disk center.

In future work, these generated images could be compared to observations from data archive, and help to remove inconsistencies these historical archive might have (Chatzistergos et al. 2024).

To reconstruct solar irradiance at CaIIK line though, one would need to obtain information also at smaller angular distances. This would requires to process not only a single image, but a time series of observations, to allow transferring of information from one timestamp to another.

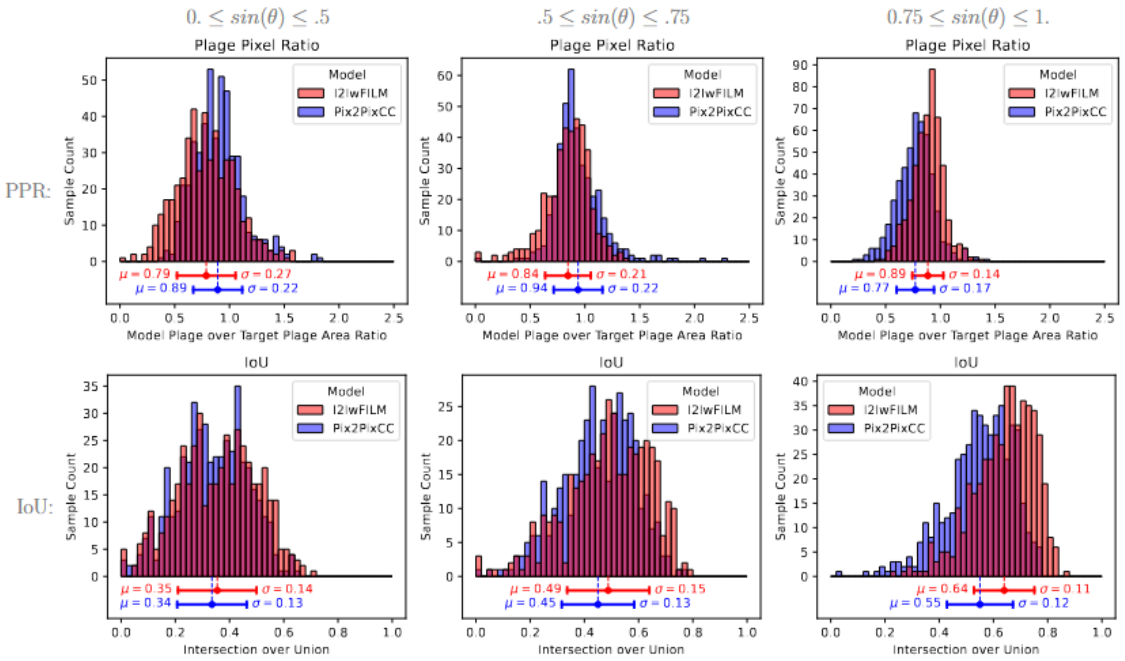


Figure 10 Plage reconstruction performance of Pix2PixCC and I2IwFiLM models for active regions located near the centre of the sun (left column), at intermediate distance to the centre(middle column) and near the limb (right column). Top row shows PPR metric histograms, bottom row shows intersection over union (IoU) histograms.

4.3. Additional results obtained during the project

4.3.1. Distinction between umbrae and penumbrae of a sunspot

We implemented an adaptive thresholding method to discriminate the umbra pixels and penumbra pixels of a sunspot see Figures 11 and 12. This method iteratively sweeps an intensity threshold value such that the initial unique region is divided into multiple regions hopefully corresponding to umbra regions.

At each iteration, the contour of each region is compared with edges found using the Canny Edge Detection algorithm to get a fitting score. In the final sunspot mask, each umbra region is represented with its mask obtained by thresholding with the best fitting of the canny edges during the threshold sweep. The other pixels in the initial sunspot mask are considered penumbra. The addition of this information in the automated McIntosh classification of a sunspot group were however not conclusive, hence it our SunSCC algorithm, it was not included. In a refined version of SunSCC, the includion of this algorithm could however be considered.

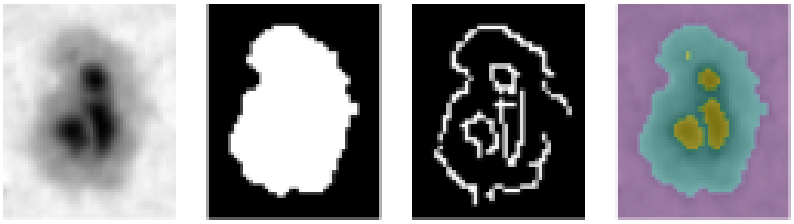


Figure 11 Example of umbra/penumbra discrimination: (A) original image (B) Sunspot mask (C) Canny edges (D) final mask.

Figure 12 Thresholding History: at each step, the contour of each yellow region is compared with Canny edges. The best score gives the optimal threshold value for that region as well as the mask of the associated umbra.

4.3.2. Image to image translation: EUV to magnetogram cross-modality translation

Another application where I2I translation could help is the 'old magnetogram' problem faced in space weather forecast. Indeed, for most magnetohydrodynamics (MHD) models, magnetograms are the primary input (Pomoell and Poedts, 2018) . But as magnetograms are primarily observing the frontside of the Sun only, either through ground-based or space-based facilities such as the SDO/HMI, some of the information inputted to MHD models through magnetogram synoptic maps were acquired up to two weeks earlier. Since the Sun is constantly evolving, such delay causes inaccuracies in the MHD model's output.

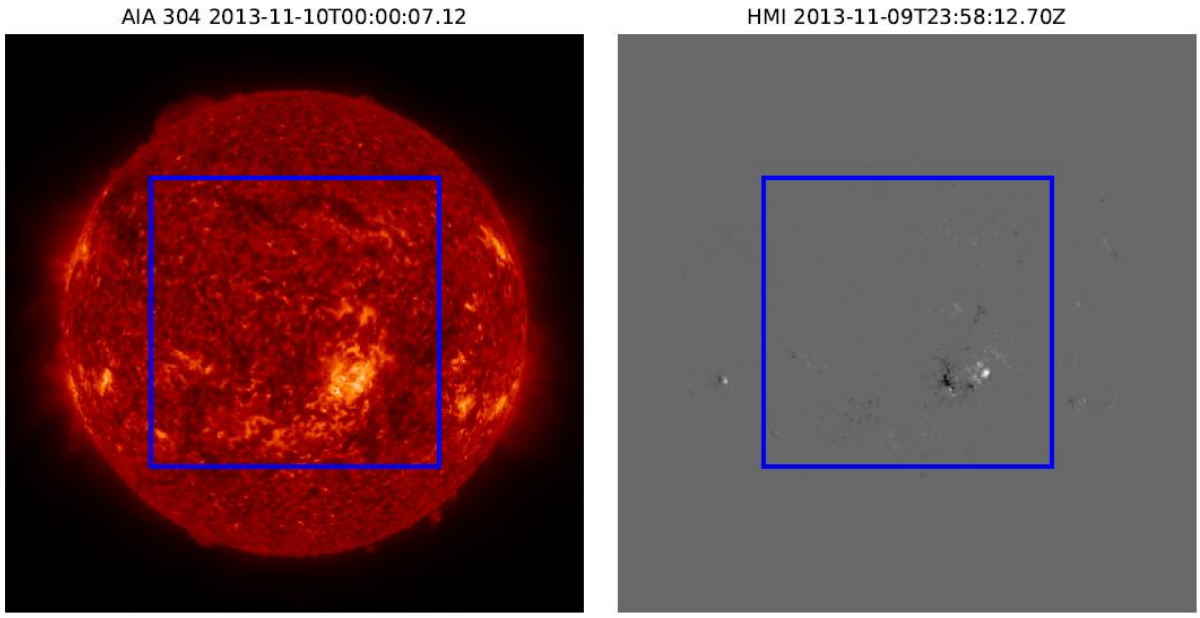


Figure 13 Sample pair of AIA-30.4nm and HMI images used for the EUV-to-magnetogram translation task in Sayez et al 2025. The blue square indicates the region considered in our study.

Extreme Ultraviolet (EUV) telescopes on the contrary are less complex instruments than magnetograms, and as such are available on spacecraft which observe the Sun from other vantage points. This is the case of the Solar Terrestrial Relations Observatory (STEREO)/Extreme UltraViolet

Imager (EUVI, Howard et al, 2008) and the Solar Orbiter (SO)/Extreme Ultraviolet Imager (EUI, Rochus et al. 2020). Since magnetograms and EUV images are both giving information about the same magnetic field, it is relevant to study how well an I2I translation method may provide information about magnetograms using EUV observations as input.

In Sayez et al. (2025), we considered the task of generating SDO/HMI magnetogram from EUV images recorded by the Atmospheric Imaging Assembly (AIA) on board SDO (Lemen et al. 2012). Figure 13 illustrates a sample pair (AIA 30.4nm, HMI) used in our application. When comparing pix2pixCC with I2IwFILM, the same conclusion as for the WL to Ca II K application holds: generic performance metrics were better, and the I2IwFILM would produce less hallucinations.

Prior to the publication in Sayez et al. (2025), V. Delouille also participated in two studies regarding this specific application:

• In Li et al. (2024), we used pix2pixCC along with transfer learning to generate, not full disk magnetogram, but synotpic magnetogram maps that were directly inputted into solar wind models

• In Dannehl et al. (2024) we made an extensive study of the roles of hyperparameters in Pix2pixCC model, and showed in particular the impact of the weights attributed to the various terms in the cGAN loss functions

4.4. Discussion and future prospects

The results from DeepSun have opened new perspectives in various domains, as discussed below.

The SunSCC pipeline could be ported to other stations observing the full Sun in visible light. Since the images will be like the ones of USET, one does not need to re-train the whole network; only fine-tuning the last layers is sufficient. There exisist such observing stations e.g. in Kanzelhöhe (Austria) and in Catania (Italy). By constituting a network of stations, also from regions located more eastwards, it would be possible to have a better coverage, possibly earlier space weather warnings, and a more robust classification obtain as an ‘ensemble’ of individual classifications.

In the image-to-image translation task, we also considered several Diffusion Models, but on our limited dataset, we notice either serious artifact and weak correlation between original and reconstructed signals. Most of these problems could be attributed to the small size dataset, and one avenue to counter this would be to use pre-trained filters.

Another improvement of I2IwFILM could come from using multimodal inputs, as this would allow to better constrain the translation model. The methodology itself could also be utilized in other cross-modality applications. One interesting avenue would be to look at EUV off-disc signals and use multiple EUV wavelengths to reconstruct the visible light observed off-disk, but close to the solar limb. Indeed, this region between 1.1. solar radii and 2.5 solar radii is difficult to observe by space coronagraph due to stray light effect, and since 2022 this region is no more observed routinely by ground-based coronagraphs either. But WL data prior to 2022 could be used to train a model, which would then be used to predict WL coronographic data after 2022.

5. DISSEMINATION AND VALORISATION (WP5)

5.1 Engagement with the scientific community

Oral presentation:

- Niels Sayez, Christophe De Vleeschouwer, Véronique Delouille, Sabrina Bechet, Laure Lefèvre *Mitigating hallucination with non-adversarial strategies for image-to-image translation in solar physics*. International workshop on Machine Learning and Computer Vision in Heliophysics, Sofia, Bulgaria, April 8, 2025
- Niels Sayez, *SunSCC: Segmentation, clustering and classification of sunspots from ground-based observations using deep learning methods*, STCE seminar 14 May 2024, Royal Observatory of Belgium
- Xiaoyue Li, Mark D. Butala, Senthamizh Pavai Valliappan, Jasmina Magdalenic, Véronique Delouille, Luciano Rodriguez and Daria Shukhobodskaia, *Transfer-Solar-GAN: Generation of Input Sources for Solar Wind Models with Deep Learning*, ML-Helio 2022 conference, Boulder, USA, March 24, 2022

Poster presentation:

- Niels Sayez, Christophe De Vleeschouwer, Véronique Delouille, Sabrina Bechet, Laure Lefèvre, *SunSCC: segmenting, grouping and classifying sunspots from ground-based observations using deep learning*, International workshop on Machine Learning and Computer vision in heliophysics, Sofia, Bulgaria, April 20, 2023

5.2. Open access of data

Catalogs pertaining from the Drawing data base (DDB), and from the sunspot group database (SGDB) have been included in two Table Access Protocol (TAP) services, an international standard protocol for data sharing widely used in astrophysics, see https://vo-tap.oma.be/

The USET L1c data set is available from the SOLARNET Virtual Observatory, see https://solarnet.oma.be/ and Mampaey et al. (2025).

5.3. Citizen science project

When working on Task 3.1 and 3.2, we constituted a dataset of about **5000 sub-images representing sunspots groups** observed by the USET WL telescope. To each sub-image corresponds in a database the McIntosh classification made by the USET Operator, as well as the simplified McIntosh class obtained with the SunSCC algorithm. We consider here the simplified McIntosh classification depicted in Figure 3, and the goal of this Citizen Science project is to have a better view on the possible errors in classifying sunspot, by comparing the errors made by a professional (the USET Operator), an automated algorithm, and finally citizens familiarised with the concept without being professionals.

The citizen project is available on the DeeSsun website (https://www.sidc.be/deepsun/), see a print screen of the home page of the project in Figure 14.

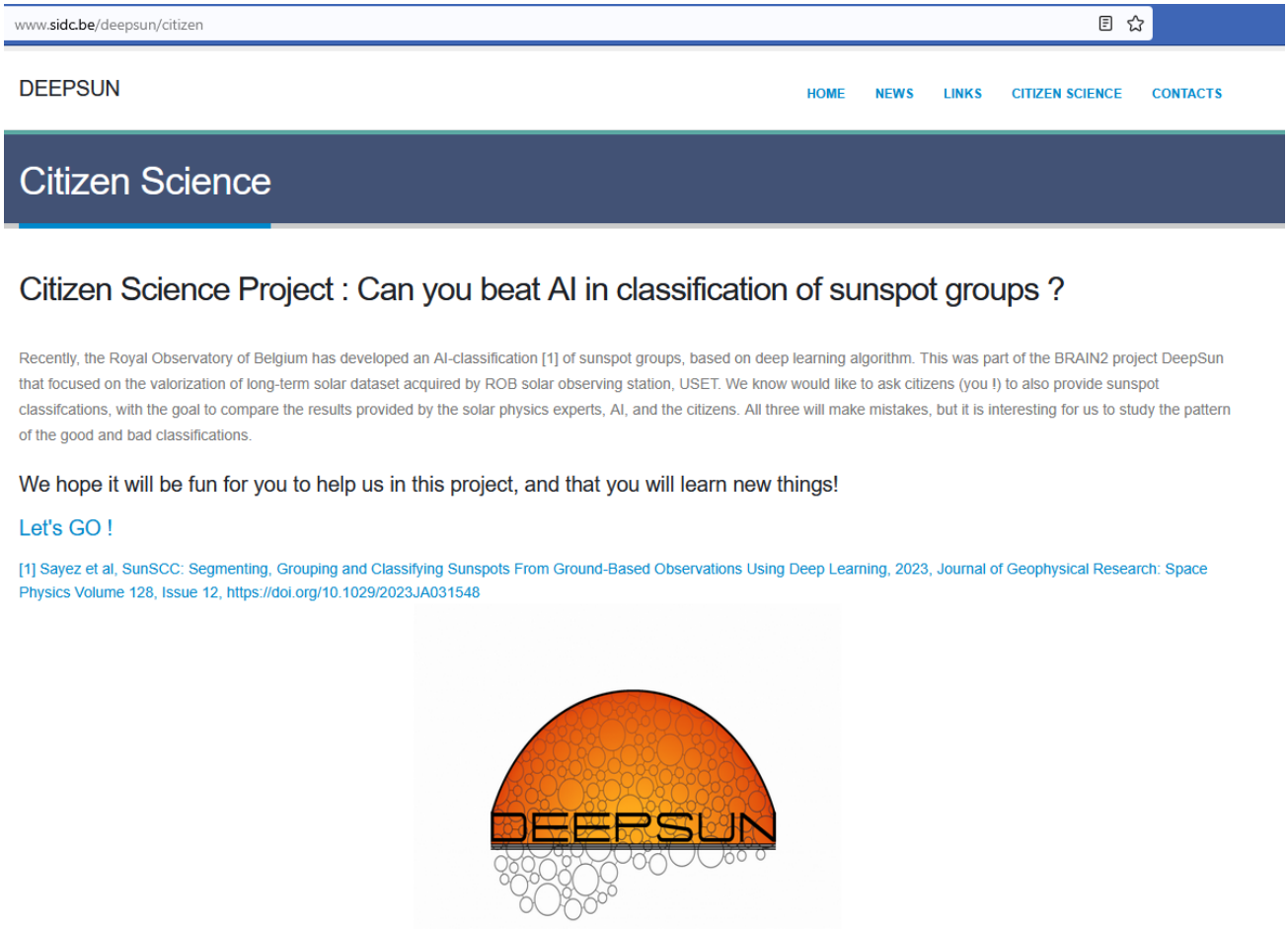


Figure 14: Homepage of the DeepSun Citizen Science project

We provide explanations and examples on the concepts of: Polarity, Penumbra, Symmetry and Distribution of spots. For each of these concepts, the person will have to determine its characteristics; for example, the person must determine if the polarity is 'Unipolar' or 'Bipolar'.

Cut-outs of sunspots groups in WL are submitted for classification. The classification goes step by step, as illustrated in Figure 15: the volunteer must find the characteristics of the sunspot group by checking the appropriate box. The final classification (using simplified McIntosh scheme) is then automatically provided.

We foresee to present this tool at the next ROB Open doors, to initiate the general public to this important activity carried out in our Federal Scientific Institute.

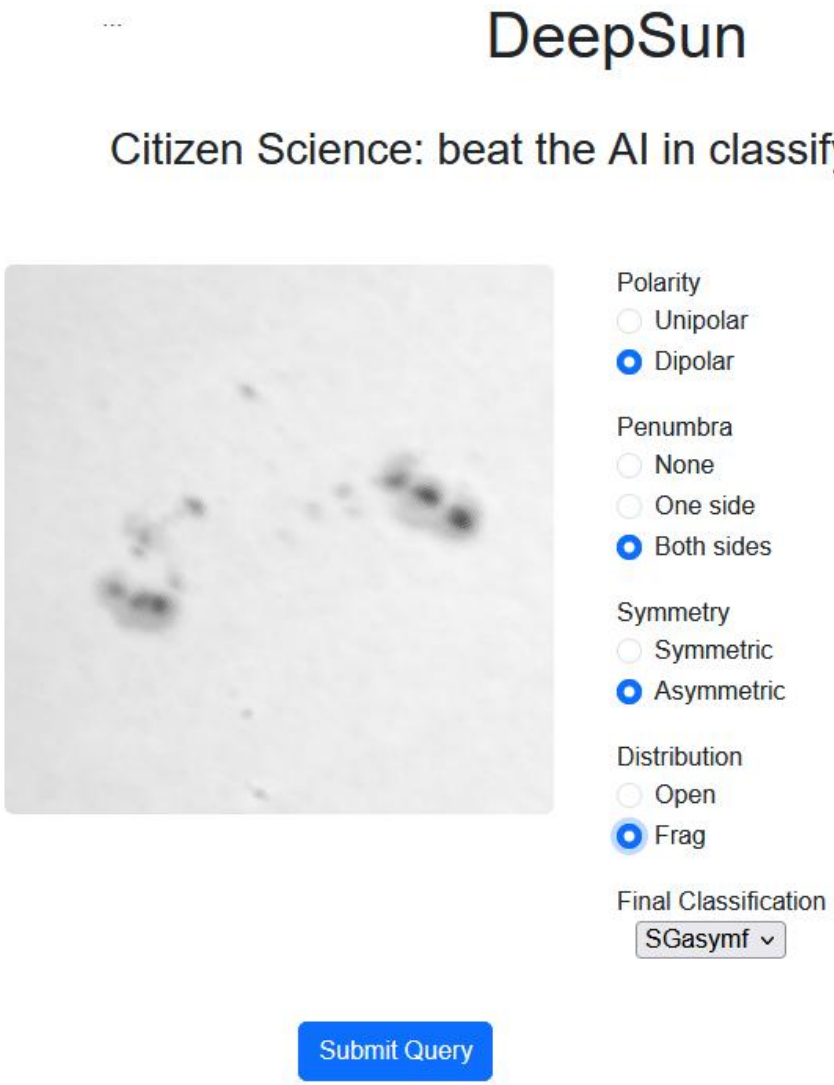


Figure 15: Example of a sunspot group to be classified manually. We use the simplified McIntosh scheme devised in the SunSCC algorithm.

6. PUBLICATIONS

- Sayez, N . Convolutional Neural Networks for the identification, characterization and modality translation of solar active regions from gound-based observations, PhD Thesis, UCLouvain, Mai 2025, available on https://dial.uclouvain.be/pr/boreal/
- Sayez, N., De Vleeschouwer, C., Delouille, V., Bechet, S., Lefèvre L. (2025) Mitigating hallucination with non-adversarial strategies for image-to-image translation in solar physics. *Astronomy & Astrophysics*, under review (minor corrections).
- Li, X., Senthamizh Pavai V. , Shukhobodskaia, D. , Butala, M., Luciano, R., Magdalenic, J., Delouille, V. (2024) A Transfer Learning Method to Generate Synthetic Synoptic Magnetograms. *Space Weather* 22. doi: 10.1029/2023SW003499
- Dannehl, M., Delouille, V., Barra, V. (2024) An Experimental Study on EUV-To-Magnetogram Image Translation Using Conditional Generative Adversarial Networks. Earth and Space Science 11. doi:10.1029/2023EA002974
- Sayez, N., De Vleeschouwer, C., Delouille, V., Bechet, S., Lefèvre L. (2023) SunSCC: Segmenting, Grouping and Classifying Sunspots From Ground-Based Observations Using Deep Learning*, Journal of Geophysical Research: Space Physics* 128(12) doi: 10.1029/2023JA031548

7. ACKNOWLEDGEMENTS

Beside the funding provided by the Belgian Federal Science Policy Office (BELSPO) through this BRAIN DeepSun project, part of the work was sponsored by the Solar-Terrestrial Centre of Excellence (STCE), a collaboration between ROB, the Royal Meteorological Institute, and the Royal Belgian Institute for Space Aeronomy funded by BELSPO.

Computational resources have been provided both by the supercomputing facilities of the UCLouvain (CISM/UCL) and the Consortium des Equipements de Calcul Intensif en Fedération Wallonie Bruxelles (CECI) funded by the Fond de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under convention 2.5020.11 and by the Walloon Region.

We thank the two members of the follow-up committee, Prof. Vincent Barra, and Dr Theodosios Chatzistergos for their time dedicated to this project, and their insightful advices and exchanges. We also thank Prof. Barra and Prof Mark Butala, from University of Glasgow, for having accepted to be part of the jury members for the PhD defense of Dr Niels Sayez. We are grateful to Dr Hun-Jin Jeong for making the Pix2PixCC code available and to Galvez et al. for their well-prepared machine learning dataset. We also acknowledge efforts supporting open-source solar data analysis Python packages we utilized in this work: NumPy, Matplotlib, PyTorch, SunPy, and Astropy.

ANNEXES

References

[SOON, 2008] SOON. https://www.swpc.noaa.gov/products/solar-region-summary, 2008.

[SOLARNET, 2021] SOLARNET, 2021. URL https://solarnet.oma.be.

[USET, 2022] USET Database, 2022. URL https://vo-tap.oma.be/.

[Abd et al, 2010] Mehmmood A. Abd, Sarab F. Majed, and V. Zharkova. Automated classification of sunspot groups with support vector machines. In Khaled Elleithy, Tarek Sobh, Magued Iskander, Vikram Kapila, Mohammad A. Karim, and Ausif Mahmood, editors, *Technological Developments in Networking, Education and Automation*, pages 321–325, Dordrecht, 2010. Springer Netherlands. ISBN 978-90-481-9151-2.

[Agarap 2019] Abien Fred Agarap. Deep learning using rectified linear units (RELU), 2019.

[Amar and Ben-Sharhar 2024] Elad Amar and Ohad Ben-Shahar. Image Synthesis for Solar Flare Prediction. *The Astrophysical Journal Supplement Series* , 271(1):29, March 2024. doi: 10.3847/1538-4365/ad1dd4 10.3847/1538-4365/ad1dd4

[Bechet & Clette, 2024] Sabrina Bechet and Frédéric Clette. USET Sunspot Group catalog. Table Access Protocol service available at http://vo-tap.oma.be/, 2024.

[Carvalho et al. 2020] S. Carvalho, S. Gomes, T. Barata, A. Louren¸co, and N. Peixinho. Comparison of automatic methods to detect sunspots in the Coimbra Observatory spectroheliograms. *Astronomy and Computing,* 32:100385, July 2020. doi: 10.1016/j.ascom.2020.100385.

[Chatzistergos et al. 2024] Theodosios Chatzistergos, Natalie A. Krivova, and Ilaria Ermolli. Understanding the secular variability of solar irradiance: the potential of Ca II K observations. J*ournal of Space Weather and Space Climate,* 14:9, March 2024. doi: 10.1051/swsc/2024006.

[Choi et al 2021] Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. ILVR: Conditioning Method for Denoising Diffusion Probabilistic Models . In 2021 *IEEE/CVF International Conference on Computer Vision (ICCV),* pages 14347–14356, Los Alamitos, CA, USA, October 2021. IEEE Computer Society. doi: 10.1109/ICCV48922.2021.01410.

[Clette et al. 2007] Frédéric Clette, David Berghmans, Petra Vanlommel, Ronald A.M. Van der Linden, André Koeckelenbergh, and Laurence Wauters. From the Wolf number to the international sunspot index: 25 years of SIDC. *Advances in Space Research,* 40(7):919–928, 2007. doi: 10.1016/j.asr.2006.12.045.

[Colak and Qahwaji 2008] T. Colak and R. Qahwaji. Automated McIntosh-Based Classification of Sunspot Groups Using MDI Images. *Solar Physics,* 248(2):277–296, April 2008. doi: 10.1007/s11207-007-9094-3.

[Comaniciu and Meer 2002] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002. doi: 10.1109/34.1000236.

[Creswell et al. 2018] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A. Bharath. Generative Adversarial Networks: An Overview. *IEEE Signal Processing Magazine*, 35(1):53–65, January 2018. doi: 10.1109/MSP.2017.2765202.

[Curto et al. 2008] J. J. Curto, M. Blanca, and E. Martinez. Automatic Sunspots Detection on Full-Disk Solar Images using Mathematical Morphology. *Solar Physics,* 250(2):411–429, August 2008. doi: 10.1007/s11207-008-9224-6.

[Dannehl et al. 2024] Markus Dannehl, Véronique Delouille, and Vincent Barra. An Experimental Study on EUV-To-Magnetogram Image Translation Using Conditional Generative Adversarial Networks. *Earth and Space Science,* 11(4):e2023EA002974, April 2024. doi: 10.1029/2023EA002974.

[Deng 2015] H. Deng, "Objective Image-Quality Assessment for High-Resolution Photospheric Images by Median Filter-Gradient Similarity", *Solar Physics*, vol. 290, no. 5, pp. 1479–1489, 2015. doi:10.1007/s11207-015-0676-1.

[Dietterich, 2000] Thomas G. Dietterich. Ensemble Methods in Machine Learning. In Multiple Classifier Systems, *Lecture Notes in Computer Science,* pages 1–15, Berlin, Heidelberg, 2000. Springer. ISBN 978-3-540-45014-6. doi: 10.1007/3-540-45014-9 1.

[Ester et al. 1996] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining,* KDD'96, page 226231. AAAI Press, 1996.

[Feng et al. 2014] S. Feng, J. Lin, Y. Yang, H. Zhu, F. Wang, and K. Ji, "Automated detecting and removing cloud shadows in full-disk solar images", *New Astronomy*, vol. 32, pp. 24–30, 2014. doi:10.1016/j.newast.2014.03.006.

[Fukunaga and L. Hostetler 1975] K. Fukunaga and L. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, 21(1):32–40, 1975. doi:10.1109/TIT.1975.1055330.

[Gao et al. 2023] Fei Gao, Tie Liu, WenQing Sun, and Long Xu. Generating Space-based SDO/HMI-like Solar Magnetograms from Ground-based Hα Images by Deep Learning. *The Astrophysical Journal Supplement Series,* 266(2):19, June 2023. doi: 10.3847/1538-4365/accbb9.

[Goodfellow et al 2016] Ian J. Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press, Cambridge, MA,USA, 2016. http://www.deeplearningbook.org.
[Hu et al 206] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. doi: 10.1109/CVPR.2016.90.

[Ho et al 2020] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf.

[Howard et al 2008] R. A. Howard, and the SECCHO consortium: Sun Earth Connection Coronal and Heliospheric Investigation (SECCHI). *Space Science Review*, 136(1-4): 67–115, April 2008. doi: 10.1007/s11214-008-9341-4.

[Isola et al. 2017] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-Image Translation with Conditional Adversarial Networks . In 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976, Los Alamitos, CA, USA, July 2017. IEEE Computer Society. doi:10.1109/CVPR.2017.632.

[Jeong et al 2020] Hyun-Jin Jeong, Yong-Jae Moon, Eunsu Park, and Harim Lee. Solar Coronal Magnetic Field Extrapolation from Synchronic Data with AI-generated Farside. *The Astrophysical Journal Letters,* 903(2):L25, November 2020. Doi: 10.3847/2041-8213/abc255

[Jeong et al 2022] Hyun-Jin Jeong, Yong-Jae Moon, Eunsu Park, Harim Lee, and Ji-Hye Baek. Improved AI-generated Solar Farside Magnetograms by STEREO and SDO Data Sets and Their Release. *The Astrophysical Journal Supplement Series* , 262(2):50, October 2022. doi: 10.3847/1538-4365/ac8d66

[Kim et al. 2019] Taeyoung Kim, Eunsu Park, Harim Lee, Yong-Jae Moon, Sung-Ho Bae, Daye Lim, Soojeong Jang, Lokwon Kim, Il-Hyun Cho, Myungjin Choi, and Kyung-Suk Cho. Solar farside magnetograms from deep learning analysis of STEREO/EUVI data. *Nature Astronomy*, 3:397–400, March 2019.

[Knyazeva et al. 2020] Irina Knyazeva, Andrey Rybintsev, Timur Ohinko, and Nikolay Makarenko. Deep-Learning Approach for McIntosh-Based Classification Of Solar Active Regions Using HMI and MDI Images. *In Advances in Neural Computation, Machine Learning, and Cognitive Research III*, pages 239–245, January 2020. ISBN 978-3-030-30424-9. doi: 10.1007/978-3-030-30425-6 28.

[Krizhevsky et al. 2012] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[Lawrance et al., 2022] Bendict Lawrance, Harim Lee, Eunsu Park, Il-Hyun Cho, Yong-Jae Moon, Jin-Yi Lee, A. Shanmugaraju, and Sumiaya Rahman. Generation of Solar Coronal White-light Images from SDO/AIA EUV Images by Deep Learning. *The Astrophysical Journal ,* 937(2):111, October 2022. doi: 10.3847/1538-4357/ac8c24.

[LeCun et al. 2015] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature,* 521:436, 521 2015. doi:10.1038/nature14539.

[Lee et al. 2021] Harim Lee, Eunsu Park, and Yong-Jae Moon. Generation of Modern Satellite Data from Galileo Sunspot Drawings in 1612 by Deep Learning. *The Astrophysical Journal*, 907(2):118, February 2021. doi: 10.3847/1538-4357/abce5f.

[Leka et al. 2019] K. D. Leka, Sung-Hong Park, Kanya Kusano, Jesse Andries, Graham Barnes, Suzy Bingham, D. Shaun Bloomfield, Aoife E. McCloskey, Veronique Delouille, David Falconer, Peter T. Gallagher, Manolis K. Georgoulis, Yuki Kubo, Kangjin Lee, Sangwoo Lee, Vasily Lobzin, JunChul Mun, Sophie A. Murray, Tarek A. M. Hamad Nageem, Rami Qahwaji, Michael Sharpe, Robert A. Steenburgh, Graham Steward, and Michael Terkildsen. A Comparison of Flare Forecasting Methods. II. Benchmarks, Metrics, and Performance Results for Operational Solar Flare Forecasting Systems. *The Astrophysical Journal Supplement Series* , 243(2):36, August 2019. doi: 10.3847/1538-4365/ab2e12

[Lemen et al. 2012] James R. Lemen, and the AIA consortium: The Atmospheric Imaging Assembly (AIA) on the Solar Dynamics Observatory (SDO). *Solar Physics*, 275(1-2):17–40, January 2012. doi: 10.1007/s11207-011-9776-8.

[Li et al. 2024] Xiaoyue Li, Senthamizh Pavai Valliappan, Daria Shukhobodskaia, Mark D. Butala, Luciano Rodriguez, Jasmina Magdalenic, and Véronique Delouille. A Transfer Learning Method to Generate Synthetic Synoptic Magnetograms. *Space Weather*, 22(1):e2023SW003499, January 2024. doi:10.1029/2023SW003499.

[Mampaey et al. 2025] Benjamin Mampaey, Véronique Delouille, and Robbe Vansintjan. The SOLARNET Virtual Observatory: An Effective Way to Search Solar Datasets and Cross-Search with Solar Events. *Solar physics*, 300(2):16, February 2025. doi: 10.1007/s11207-025-02424-0.

[Mao et al. 2017] Xudong Mao, Qing Li, Haoran Xie, Raymond Y.K. Lau, Zhen Wang, and Stephen Paul Smolley. Least Squares Generative Adversarial Networks . In 2017 *IEEE International Conference on Computer Vision (ICCV),* pages 2813–2821, Los Alamitos, CA, USA, October 2017. IEEE Computer Society. doi:10.1109/ICCV.2017.304.

[Martens et al. 2012] P. C. H. Martens, G. D. R. Attrill, A. R. Davey, A. Engell, S. Farid, P. C. Grigis, J. Kasper, K. Korreck, S. H. Saar, A. Savcheva, Y. Su, P. Testa, M. Wills-Davey, P. N. Bernasconi, N. E. Raouafi, V. A. Delouille, J. F. Hochedez, J. W. Cirtain, C. E. DeForest, R. A. Angryk, I. De Moortel, T. Wiegelmann, M. K. Georgoulis, R. T. J. McAteer, and R. P. Timmons. Computer Vision for the Solar Dynamics Observatory (SDO). *Solar Physics,* 275(1-2):79–113, January 2012. doi: 10.1007/s11207-010-9697-y.

[McIntosh, 1990] Patrick S. McIntosh. The Classification of Sunspot Groups. *Solar Physics*, 125(2):251–267, September 1990. doi: 10.1007/BF00158405.

[Muraközy, 2022] Judit Muraközy. Variations of the Internal Asymmetries of Sunspot Groups during Their Decay. *The Astrophysical Journal*, 925 (1):87, January 2022. doi: 10.3847/1538-4357/ac3de6.

[Nguyen et al. 2004] Trung Thanh Nguyen, Claire P. Willis, Derek J. Paddon, and Hung Son Nguyen. On learning of sunspot classification. In Mieczyslaw A. Klopotek, Slawomir T. Wierzchon, and Krzysztof Trojanowski, editors, *Intelligent Information Processing and Web Mining,* pages 59–68, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg. ISBN 978-3-540-39985-8.

[Nguyen et al, 2006] Trung Thanh Nguyen, Claire P. Willis, Derek J. Paddon, and Hung Son Nguyen. A hybrid system for learning sunspot recognition and classification. In 2006 *International Conference on Hybrid Information Technology,* volume 2, pages 257–264, 2006. doi: 10.1109/ICHIT.2006.253620.

[Palladino et al. 2022] Luigi Palladino, Evridiki Ntagiou, Johannes Klug, Judit Palacios, and Ralf Keil. Sunspot Groups Detection and Classification on SDO/HMI Images using Deep Learning Techniques. In 2022 *IEEE Aerospace Conference (AERO),* pages 1–10, March 2022. doi: 10.1109/AERO53065.2022.9843222. ISSN: 1095-323X.

[Paszke et al. 2019] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor
Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An imperative style, high-performance deep learning library, 2019.

[Peake et al., 2020] Edward J Peake, Raphael Chevasson, Stefan Pszczolkowski, Dorothee P Auer, and Christoph Arthofer. Ensemble learning for robust knee cartilage segmentation: data from the osteoarthritis initiative. *BioRxiv*, pages 2020–09, 2020.

[Perez et al. 2017] Ethan Perez, Florian Strub, Harm Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer. *Proceedings of the AAAI Conference on Artificial Intelligence,* 32, 09 2017. doi: 10.1609/aaai.v32i1.11671.

[Pettauer and Brandt, 1997] T. Pettauer and P. N. Brandt. On Novel Methods to Determine Areas of Sunspots from Photoheliograms. *Solar Physics*, 175(1):197–203, September 1997. doi: 10.1023/A:1004903201224.

[Pomoell and Poedts, 2018.] Jens Pomoell and S. Poedts. EUHFORIA: European heliospheric forecasting information asset. *Journal of Space Weather and Space Climate*, 8:A35, June 2018.. doi: 10.1051/swsc/2018020.

[Ramunno et al. 2024] Francesco Pio Ramunno, Hyun-Jin Jeong, Stefan Hackstein, André Csillaghy, Svyatoslav Voloshynovskiy, and Manolis K. Georgoulis. Magnetogram-to-Magnetogram: Generative Forecasting of Solar Evolution. In Dominik Dold, Alexander Hadjiivanov, and Dario Izzo, editors, *Proceedings of SPAICE2024: The First Joint European Space Agency / IAA Conference on AI in and for Space,* pages 75–80, October 2024. doi: 10.5281/zenodo.13885515.

[Rochus et al. 2020] P. Rochus and the EIT Consortium, The Solar Orbiter EUI instrument: The Extreme Ultraviolet Imager. *Astronomy & Astrophysics*, 642:A8, October 2020. doi:10.1051/0004-6361/201936663.

[Rombach et al. 2022] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bjorn Ommer. High-Resolution Image Synthesis with Latent Diffusion Models . In 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR*), pages 10674–10685, Los Alamitos, CA, USA, June 2022. IEEE Computer Society. doi: 10.1109/CVPR52688.2022.01042.

[Ronneberger et al. 2015] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI* 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pages 234–241. Springer, 2015.

[Saharia et al. 2022] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings, SIGGRAPH '22,* New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450393379. doi: 10.1145/3528233.3530757.

[Sayez et al. 2023] Niels Sayez, Christophe De Vleeschouwer, Véronique Delouille, Sabrina Bechet, and Laure Lefèvre. SunSCC: Segmenting, Grouping and Classifying Sunspots From Ground-Based Observations Using Deep Learning. *Journal of Geophysical Research (Space Physics)*, 128(12):e2023JA031548, December 2023. doi: 10.1029/2023JA031548.

[Sayez et al 2025] Niels Sayez, Christophe De Vleeschouwer, Véronique Delouille, Sabrina Bechet, and Laure Lefèvre. Mitigating hallucination with non-adversarial strategies for image-to-image translation in solar physics. *Astronomy & Astrophysics*, 2025. (Submitted)

[Scherrer et al 1995] P. H. Scherrer, R. S. Bogart, R. I. Bush, J. T. Hoeksema, A. G. Kosovichev, J. Schou, W. Rosenberg, L. Springer, T. D. Tarbell, A. Title, C. J. Wolfson, I. Zayer, and MDI Engineering Team. The Solar Oscillations Investigation - Michelson Doppler Imager. *Solar Physics*, 162(1-2):129–188, December 1995. doi: 10.1007/BF00733429.

[Scherrer et al. 2012] P. H. Scherrer, J. Schou, R. I. Bush, A. G. Kosovichev, R. S. Bogart, J. T. Hoeksema, Y. Liu, T. L. Duvall, J. Zhao, A. M. Title, C. J. Schrijver, T. D. Tarbell, and S. Tomczyk. The Helioseismic and Magnetic Imager (HMI) Investigation for the Solar Dynamics Observatory (SDO). *Solar Physics,* 275(1-2):207–227, January 2012. doi: 10.1007/s11207-011-9834-2.

[Shin et al. 2020] Gyungin Shin, Yong-Jae Moon, Eunsu Park, Hyunjin Jeong, Harim Lee, and Sung-Ho Bae. Generation of High-resolution Solar Pseudo-magnetograms from Ca II K Images by Deep Learning. *The Astrophysical Journal Letters*, 895(1):L16, May 2020. doi: 10.3847/2041-8213/ab9085.

[Son et al. 2021] Jihyeon Son, Junghun Cha, Yong-Jae Moon, Harim Lee, Eunsu Park, Gyungin Shin, and Hyun-Jin Jeong. Generation of He I 1083 nm Images from SDO AIA Images by Deep Learning. *The Astrophysical Journal* , 920(2):101, October 2021. doi: 10.3847/1538-4357/ac16dd

[Song et al. 2024] Wei Song, Ying Ma, Haoying Sun, Xiaobing Zhao, and Ganghua Lin. Improving the spatial resolution of solar images using super-resolution diffusion generative adversarial networks, *Astronomy and Astrophysics,* 686:A272, June 2024. doi: 10.1051/0004-6361/202349100.

[Steinegger et al. 1998] Michael Steinegger, Jose A. Bonet, Manuel Vázquez, and Antonio Jim´enez. On the Intensity Thresholds of the Network and Plage Regions. *Solar Physics*, 177(1-2):279–286, January 1998. doi: 10.1023/A:1004920125173.

[Sun et al 2023] Wenqing Sun, Long Xu, Yin Zhang, Dong Zhao, and Fengzhen Zhang. Solar Active Region Magnetogram Generation by Attention Generative Adversarial Networks. *Research in Astronomy and Astrophysics*, 23(2):025003, February 2023. doi: 10.1088/1674-4527/acaa92.

[SunPy 2020] The SunPy Community, Will T. Barnes, Monica G. Bobra, Steven D. Christe, Nabil Freij, Laura A. Hayes, Jack Ireland, Stuart Mumford, David Perez-Suarez, Daniel F. Ryan, Albert Y. Shih, Prateek Chanda, Kolja Glogowski, Russell Hewett, V. Keith Hughitt, Andrew Hill, Kaustubh Hiware, Andrew Inglis, Michael S. F. Kirk, Sudarshan Konge, James Paul Mason, Shane Anthony Maloney, Sophie A. Murray, Asish Panda, Jongyeob Park, Tiago M. D. Pereira, Kevin Reardon, Sabrina Savage, Brigitta M. Sipocz, David Stansby, Yash Jain, Garrison Taylor, Tannmay Yadav, Rajul, and Trung Kien Dang. The sunpy project: Open source development and status of the version 1.0 core package. *The Astrophysical Journal,* 890:68–, 2020. doi: 10.3847/1538-4357/ab4f7a.
[Turmon et al. 2002] M. Turmon, J. M. Pap, and S. Mukhtar. Statistical Pattern Recognition for Labeling Solar Active Regions: Application to SOHO/MDI Imagery. *The Astrophysical Journal*, 568(1):396–407, March 2002. doi: 10.1086/338681.

[Verbeeck et al. 2014] C. Verbeeck, V. Delouille, B. Mampaey, and R. De Visscher. The SPoCA-suite: Software for extraction, characterization, and tracking of active regions and coronal holes on EUV images. *Astronomy & Astrophysics*, 561:A29, January 2014. doi: 10.1051/0004-6361/201321243.

[Wang et al., 2018] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs . In 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8798–8807, Los Alamitos, CA, USA, June 2018. IEEE Computer Society. doi: 10.1109/CVPR.2018.00917.

[Watson et al. 2009] F. Watson, L. Fletcher, S. Dalla, and S. Marshall. Modelling the Longitudinal Asymmetry in Sunspot Emergence: The Role of the Wilson Depression. Solar Physics, 260(1):5–19, November 2009. doi: 10.1007/s11207-009-9420-z

[Zamir 2012] A. Roshan Zamir, A. Dehghan, and M. Shah. GMCP-tracker: Global multi-object tracking using generalized minimum clique graphs. In *Proceedings of the European Conference on Computer Vision (ECCV), 2012.*

[Zharkov et al. 2005] S. Zharkov, V. Zharkova, S. Ipson, and A. Benkhalil. Technique for Automated Recognition of Sunspots on Full-Disk Solar Images. *EURASIP Journal on Applied Signal Processing,* 2005:318462, December 2005. doi: 10.1155/ASP.2005.2573

[Zharkova, et al. 2005] V. V. Zharkova, J. Aboudarham, S. Zharkov, S. S. Ipson, A. K. Benkhalil, and N. Fuller. Solar Feature
Catalogues In EGSO. *Solar Physics*, 228(1-2):361–375, May 2005. doi: 10.1007/s11207-005-5623-0.