DEEPSUN

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

Contract - B2/191/P2/DeepSun

SUMMARY

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. 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.

The Royal Observatory of Belgium (ROB) is a key player in sunspot observations: In 1939, the ROB ground-based 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 still runs today. In addition, the USET facility produces since 2002 white light (WL) images of the photosphere and since 2012 images of the chromosphere in the CaIIK bandpass. The co-temporal and co-spatial acquisition of white light and CaIIK images makes it favourable to interconnect these datasets using novel image processing techniques.

OBJECTIVE AND METHODOLOGY

The over-arching goal of DeepSun is to exploit USET datasets and produce high-level data products to be used for science advancement and dataset valorization.

Our first objective is to automatically detect, group, and classify sunspot groups from white light USET images alone. Such classification will provide an aid to space weather forecasters as well as allowing systematic studies of sunspots. To reach this objective, we developed the SunSCC pipeline depicted in Figure 1. It takes as input USET white light images and returns as output segmentation masks of individual sunspots, as well as a bounding for the corresponding sunspot groups and their McIntosh classification, with a reliability score. It is divided in three independent blocks:

1) A sunspot segmentation block, based on a U-net architecture, composed of Convolutional Neural Networks (CNN) and that has proved effective in segmentation tasks
2) An aggregation block to cluster individual sunspots into a sunspot group based on a modified mean-shift algorithm,
3) A classification network block built around Multi-Layer Perceptrons (MLP) and designed to predict the McIntosh classification, composed of three components, of the sunspot group.

Our second objective is to fill in gaps in historical records by exploiting physical correlation between different modality of observations. Starting from long-time observed white light observations of sunpots, our aim is to recover the corresponding plages in Ca IIK images, see Figure 2. To do so, 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 white light images. We found however that this method introduces physically unfounded artifacts. This led us to propose *I2IwFILM*, a novel model architecture for solar physics data exploitation. By combining advanced feature extraction with feature map modulation, this approach enables enriched reconstruction of the target modality (here CaIIK) through explicitly extracted source features.

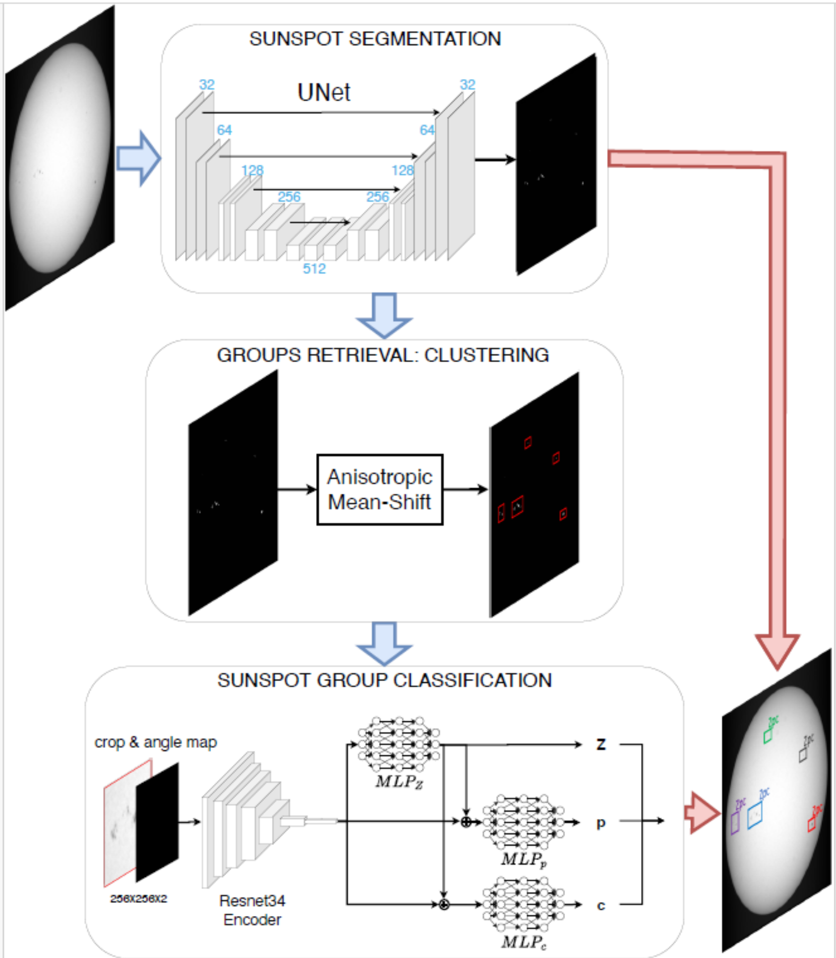

Figure 1: SunSCC pipeline for sunspot segmentation, clustering, and classification. Full-disc images 2048x2048 images are subdivided into a 4x4 grid, each containing segments of dimensions 512x512. A U-net segmentation network predicts masks for each segment that are reassembled into a 2048x2048 segmentation 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 MLPs, each MLP being specialized in the classification of one component in the McIntosh system.

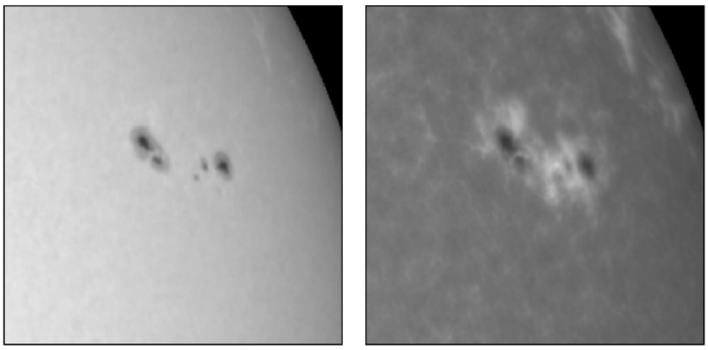

Figure 2. Sample pair of WL and CaII-K sub-images, which have been spatially aligned and centered near an active region. The WL sub-image shows sunspots as well as regions around the sunspots that are slightly brighter than the background and called faculae. These faculae correspond physically to the bright 'plages' observed in the CaII-K sub-image around the dark sunspots.

Finally, we want our results to be easily accessible by the scientific community and include citizens in the further exploitations of our results.

CONCLUSIONS

Concerning our first objective, the segmentation part of the SunSCC pipeline provides reliable sunspot masks even when atmospheric artifacts degrade USET images. The developed clustering methodology leverages sunspot group extents and individual sunspot sizes to successfully replicate USET catalog patterns. Finally, the proposed CNN architecture for McIntosh classification outperforms existing solutions relying on satellite-based imagery, despite using only ground-based observations.

For our second objective, evaluations demonstrated that our proposed I2IwFILM scheme generates physically consistent predictions with fewer hallucinations compared to adversarial-based approaches. An example of such translations is displayed in Figure 3. Such generated images could be compared to CaIIK observations from historical data archives and help to remove possible inconsistencies in such archives.

Finally, we distributed USET data products to the scientific community via standard Virtual Observatories. To further disseminate and valorize the results from DeepSun, we also proposed a citizen science project where the task is to provide a sunspot classification, in view of comparing errors made by automated and manual classification methods.

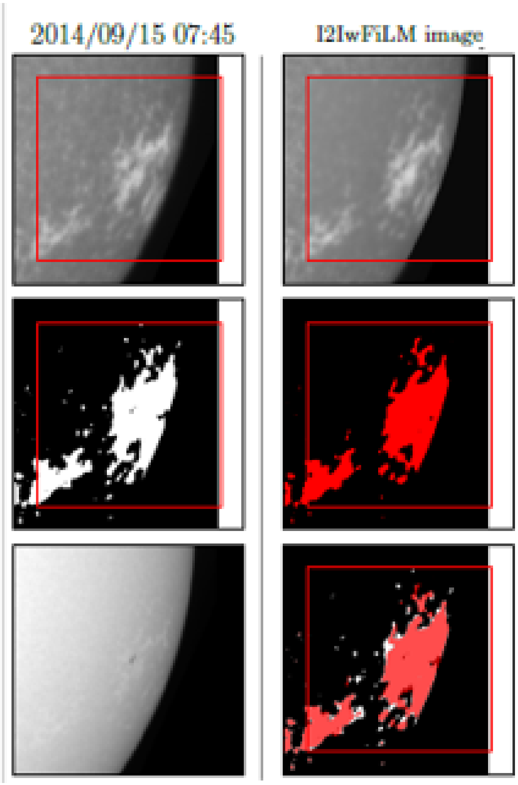

Figure 3 Ca II K reconstructions of plages near the limb from observations made on 15 September 2014 at 07:15, using our proposed I2IwFILM algorithm. The top row shows the original CaIIK image (left) and its reconstructions by I2IwFiLM (right) The middle row shows segmentation masks of plages visible in the original CaIIK image (left) and the corresponding reconstructions by I2IwFiLM (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 (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.

KEYWORDS

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