DEEPSUN

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

Contract - B2/191/P2/DeepSun

RÉSUMÉ

CONTEXTE

Les taches solaires sont des manifestations du magnétisme solaire qui apparaissent à la surface du Soleil comme des taches sombres groupées. L'étude à long terme de l'évolution des taches solaires est fondamentale en physique solaire, que ce soit pour l'hélioseismologie, la modélisation de l'irradiance solaire ou encore pour la prévision de la météo spatiale.

L'Observatoire royal de Belgique (ORB) joue un rôle majeur dans l'observation des taches solaires: en 1939, la station solaire de l'ORB (appelée « USET », pour Uccle Solar Equatorial Table) a lancé un programme d'observation en collaboration avec l'Observatoire de Zurich. Ce programme consiste à réaliser au moins un dessin de la configuration des taches solaires par jour; ce programme est toujours en cours aujourd'hui. De plus, la station USET acquière depuis 2002 des images des images en lumière blanche (LB) et depuis 2012 des images dans la raie du calcium (CaIIK). L'acquisition au même endroit (co-spatiale) et au même moment (co-temporelle) de ces images nous permet d'établir des liens entre ces données à l'aide de techniques de traitement d'images.

OBJECTIFS ET MÉTHODOLOGIE

La raison d'être de ce projet est de produire des données à haute valeur ajoutée à partir des différents ensembles de données USET afin de faciliter leur exploitation à des fins scientifiques et leur diffusion.

Notre premier objectif concerne la détection, le regroupement et la classation automatique des groupes de taches solaires, et ce à partir des seules images en lumière blanche acquises par USET. Une telle classification servira d'aide aux prévisionnistes de la météorologie spatiale et permettra également des études systématiques des taches solaires. Pour atteindre cet objectif, nous avons développé le pipeline 'SunSCC' représenté à la Figure 1. Celui-ci prend en entrée des images USET en lumière blanche et fournit en sortie: des masques de segmentation des taches solaires individuelles, des rectangles englobant les groupes de taches solaires et enfin une classification de McIntosh de ces groupes, accompagnée d'un score de fiabilité. Le pipeline est divisé en trois blocs indépendants:

1) Un bloc de segmentation des taches solaires, basé sur une architecture U-net comprenant des réseaux de neurones convolutifs (CNN) et qui s'est révélée efficace pour les tâches de segmentation
2) un bloc d'agrégation permettant de regrouper les taches solaires individuelles en un groupe de taches solaires. Ce bloc est basé sur un algorithme de mean-shift modifié ;
3) un bloc de réseau de classification, construit autour de perceptrons multicouches (MLP), conçu pour prédire la classification de McIntosh, composée de trois composantes, du groupe de taches solaires.

Notre second objectif est de combler les lacunes dans les archives historiques en exploitant les corrélations physiques entre différentes modalités d'observation. En partant des observations en lumière blanche des taches solaires, disponibles depuis longtemps, notre but est de reconstituer les plages correspondantes dans les images Ca IIK (voir Figure 2). Pour ce faire, nous avons d'abord

exploré des solutions basées sur les réseaux antagonistes générateurs (Generative Adversarial Network ou GANs), en raison de leur capacité à effectuer un transfert de modalités, càd dans notre cas, de générer des images Ca IIK réalistes à partir d'images en lumière blanche. Nous avons cependant constaté que cette méthode introduit des artefacts dépourvus de fondement physique. Cela nous a conduits à proposer I2IwFILM, une nouvelle architecture pour l'exploitation des données en physique solaire. En combinant une extraction avancée de caractéristiques à une modulation des cartes de caractéristiques (feature map), cette approche permet une reconstruction enrichie de la modalité cible (ici Ca IIK) à partir de caractéristiques sources extraites explicitement.

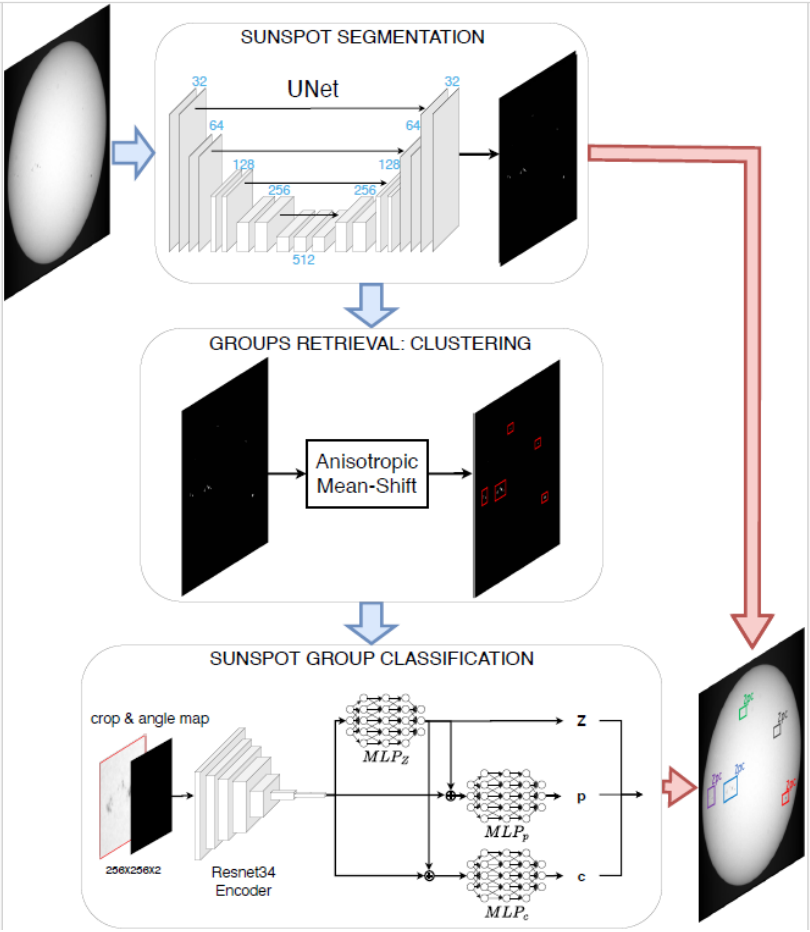

Figure 1. Pipeline SunSCC pour la segmentation, le regroupement et la classification des taches solaires. Les images du disque solaire complet, de résolution 2048x2048, sont subdivisées en 16 cellules, chaque cellule contenant un segment de 512x512 pixels. Un réseau de segmentation U-net prédit des masques pour chaque segment, qui sont ensuite réassemblés en un masque de segmentation complet de 2048x2048. Les taches solaires détectées sont regroupées en groupes de taches solaires à l'aide d'un algorithme de mean-shift modifié. Chaque groupe identifié est associé à une carte de distances angulaires transmise à un réseau de classification composé d'un encodeur d'images ResNet34 et de trois MLP, chacun étant spécialisé dans la classification d'une des composantes du système de McIntosh.

Enfin, nous voulons que nos produits de données soient mises à la disposition de la communauté scientifque, et nous voulons encourager le grand public à participer à nos efforts de recherche via une initiative citoyenne.

CONCLUSIONS

Concernant notre premier objectif, la partie segmentation du pipeline SunSCC fournit des masques de taches solaires fiables, même lorsque des artefacts atmosphériques dégradent les images USET. La méthodologie développée pour délimiter les groupes de sunspots exploite l'étendue des groupes de taches solaires ainsi que la taille des taches individuelles pour reproduire avec succès les formes et tailles de groupes observés dans le catalogue USET. Enfin, l'architecture CNN proposée pour la classification de McIntosh surpasse les solutions existantes basées sur des images satellites, malgré l'utilisation exclusive d'observations prises à partir du sol.

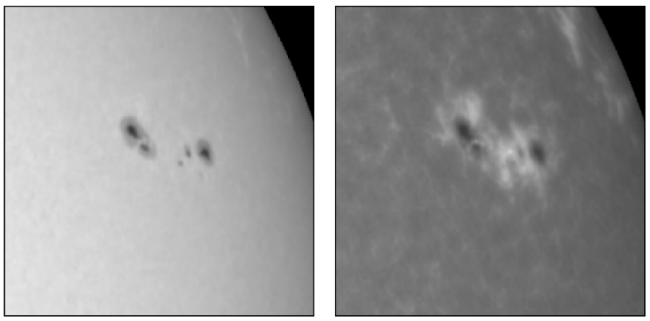

Figure 2. Exemple de paire de sous-images LB et Ca II-K, spatialement alignées et centrées autour d'une région active. La sous-image LB montre des taches solaires ainsi que des régions légèrement plus brillantes que l'arrière-plan autour des taches, appelées facules. Ces facules correspondent physiquement aux plages brillantes observées dans la sous-image Ca II-K autour des taches solaires sombres.

Pour notre second objectif, les évaluations ont démontré que le schéma I2IwFILM génère des prédictions physiquement cohérentes avec moins d'hallucinations que les approches basées sur des réseaux antagonistes. Un exemple de ce type de prédiction est présenté à la Figure 3. Ces images générées pourraient être comparées aux observations Ca II-K issues d'archives historiques, et ainsi contribuer à corriger d'éventuelles incohérences présentes dans ces archives.

Enfin, nous avons utilisé des observatoires virtuels standardisés pour diffuser les produits de données USET auprès de la communauté scientifique. Afin valoriser au mieux les résultats du projet DeepSun, nous avons également proposé un projet de science participative, dans lequel la tâche consiste à fournir une classification des taches solaires, dans le but de comparer les erreurs des méthodes de classification automatisée et manuelle.

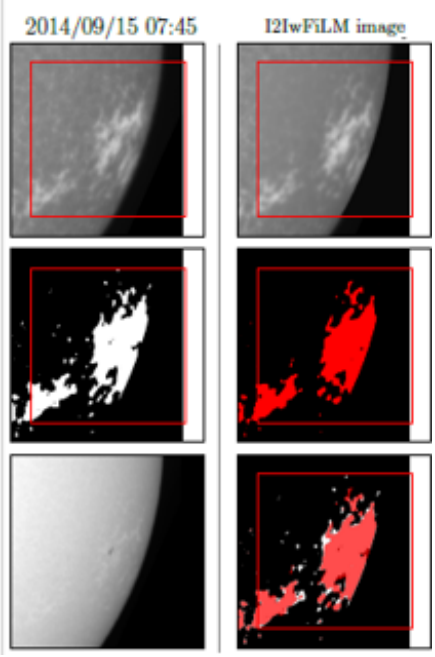

Figure 3 Reconstructions en Ca II-K des plages proches du limbe à partir d'observations réalisées le 15 septembre 2014 à 07h15, à l'aide de notre algorithme I2IwFiLM. La rangée du haut montre l'image Ca II-K originale (à gauche) et sa reconstruction par I2IwFiLM (à droite). La rangée du milieu présente les masques de segmentation des plages visibles dans l'image Ca II-K originale (à gauche) et les reconstructions correspondantes générées par I2IwFiLM (à droite). La rangée du bas montre l'image d'entrée en lumière blanche (LB) (à gauche) et une superposition du masque cible des plages avec les masques de plages générés par I2IwFiLM (à droite), facilitant une comparaison directe. Le carré rouge correspond à la sous-image utilisée en entrée des modèles ; une image légèrement plus grande est affichée pour montrer la cohérence avec les régions environnantes.

MOTS CLÉS

Segmentation de taches solaires, groupement de taches solaires, classification de taches solaires, translation d'un type d'image à un autre , apprentissage en couches profonde