

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique Université des Sciences et de la Technologie Houari Boumediene Faculté d'Électronique et d'Informatique

Département Informatique

Mémoire de Master

Spécialité : Informatique visuelle

Détection et classification automatique des taches solaires par des méthodes de deep learning

Sujet proposé par : DAIFFALLAH khalil *Réalisé par :* ATCHI Imene BOUACHAT Anfel

Binôme n° : IV010/2023

22 juin 2023

9

11

11

11

11

12

12

14

14

15

15

15

15

15

17

17

17

17

17

18

18

19

19

19

20

21

22

23

23

23

26

27

27

27

27

32

32

34

Table des matières

INTRODUCTION GÉNÉRALE Chapitre : Environnement d'études 0.10.2Présentation du CRAAG Historique du CRAAG 0.30.4Missions du CRAAG Structure du CRAAG 0.5La recherche sur les taches solaire au CRAAG 0.60.7Conclusion Chapitre 2 : Etat de l'art 0.8 Le Soleil 0.9La Structure interne du Soleil 0.9.1La photosphère 0.9.2La chromosphère 0.9.30.9.4La couronne 0.10 L'activité solaire 0.10.1 Les taches solaires 0.12 La classification magnétique des taches solaires 0.12.1 La classification magnétique du mount Wilson-Mcintosh : . . 0.13.1 Les réseaux de neurones 0.13.3 Jeu de données d'un réseau de neurones convolutifs 0.14 Synthèse de quelque approches de détection et classification des taches Chapitre 3 : Le traitement d'images et calcul du nombre de Wolf 0.15.2 Jeu de données

0.15.6 L'évaluation des résultats et test sur les données fournies par

0.15.7 Conclusion \ldots	36
Chapitre 4 : Conception d'un Système de classification des taches so-	
laires	37
$0.15.8 Introduction \ldots \ldots$	37
0.15.9 La collect de jeu de données	37
0.15.10 L'architecture globale de notre système	38
0.15.11 Phase d'apprentissage basé sur un modèle RNC	40
$0.15.12$ Conclusion \ldots	41
Chapitre 5 : Implémentation et évaluation des résultats de	44
$0.15.13$ Introduction \ldots	44
0.15.14 Les critères d'évaluation des résultats	44
0.15.15 L'évaluation du modèle de classification	45
0.15.16 L'environnement de développement	46
0.15.17 La présentation de l'interface	49
0.15.18 L'évaluation du modèle de classification sur le jeu de test et	
discussion	51
$0.15.19$ Conclusion \ldots	51
CONCLUSION GÉNÉRALE	53

Table des figures

1	Logo du CRAAG
2	Organisation du CRAAG
3	Structure interne du Soleil
4	Image à haute résolution de la surface du Soleil
5	Progression du cycle solaire
6	Aperçu d'un neurone artificiel ayant plusieurs entrées
7	Aperçu d'un réseau de neurone ayant une couche cachée
8	Aperçu d'un réseau de neurones convolutif
9	Images des contours des taches
10	Image binarisée
11	Taches solaires ségmentés
12	Progression du cycle solaire
13	Détection et segmentation des taches solaires sur l'image originale
	capturée par la lunette du GRAAG le 27 mars 2023
14	Image magnétogramme d'une tache solaire de type Beta 38
15	Architecture globale de notre système de classification
16	Image en niveaux de gris
17	Conversion de l'image en matrice numpy 40
18	Architecture RNC utilisée pour la classification des images magnétogrammes
	des tache solaire Fig 1 \ldots 41
19	Architecture utilisée pour la classification des images magnétogrammes
	des tache solaire Fig 2. \ldots 42
20	Graph de précision
21	loss graph
22	Page d'accueil 1
23	Page d'accueil 2
24	Page de classification
25	Matrice de confusion sur le jeu de test

Liste des tableaux

1	Synthèse de quelques travaux sur la détection et classification des	
	taches solaires partie 1	24
2	Synthèse de quelques travaux sur la détection et classification des	
	taches solaires partie 2	25

Liste des Algorithmes

1	Algorithme général du calcul de nombre de Wolf	28
2	Algorithme de DBSCAN	33

Remerciement

I would like to express my deepest gratitude to God for His blessings, guidance, and strength throughout this journey. His unwavering support has been instrumental in my success, Al Hamduli Allah.

I am incredibly thankful to my family and friends Nina, Meriem and Feriel and all the names I can't mention. for their continuous love, encouragement, and belief in my abilities. their support has been a constant source of motivation for me.

I extend my heartfelt appreciation to my supervisor Mr DAIFALLAH Khalil for his valuable guidance, expertise, and continuous encouragement. his insightful feedback and dedication have been crucial in shaping this research project.

I would also like to thank the entire team at CRAAG. Their collective expertise and contributions have significantly enriched my understanding in the field.

A special word of thanks goes to my colleague, Imene ATCHI, for her collaboration, insightful discussions, and assistance throughout this project. Her dedication and enthusiasm have been truly inspiring.

Lastly, I would like to express my sincere gratitude to Professor Insaf SETITRA, for igniting my passion for the field of computer vision. Her teachings and guidance have played a pivotal role in my interest and growth in this domain. I am immensely grateful to all the individuals who have played a part, directly or indirectly, in completing this project.

BOUACHAT ANFEL

Remerciement

Je tiens à exprimer ma gratitude et mes remerciements à Dieu pour m'avoir accordé la force et la santé nécessaires pour aller jusqu'au bout de ce projet. Je suis également reconnaissant envers mes parents, mes piliers et sources d'inspiration, ainsi que mes sœurs Yasmine, Serine, Khadidja et mon frère Loukmane, qui ont toujours été présents pour me soutenir et encourager mes ambitions. Sans oublier ma petite nièce Éline, ses sourires et ses câlins ont apporté une dose de douceur et de motivation supplémentaire.

Je souhaite également exprimer ma profonde reconnaissance envers mes superviseurs, Mr Khalil DAIFALLAH et Mr Djounai BABA AISSA, pour leurs guidance, leurs soutien et leurs précieux conseils tout au long de la réalisation de ce mémoire. Leur expertise et leur disponibilité ont grandement contribué à la réussite de ce travail.

Un merci particulier à toute l'équipe du GRAAG qui a été présente à chaque étape de ce projet. Votre collaboration, votre soutien et votre camaraderie ont été essentiels pour mener ce projet à bien.

Enfin, un grand merci à vous tous qui avez contribué de près ou de loin à la réalisation de ce mémoire.

ATCHI IMENE

Abstract

This Master's thesis project aims to assist astronomers in the detection and classification of solar sunspots using magnetogram images. The analysis of sunspots is of great importance in astronomy as it allows for a better understanding of the behaviour and evolution of the Sun, the identification of relationships with other solar phenomena, and the prediction of solar eruptions. Magnetogram images play a significant role in the classification of sunspots as they provide information about the associated magnetic field. The project involves collecting magnetogram images of sunspots to build a dataset, which was divided into three parts (training, validation, and testing). We then performed preprocessing on the training and validation datasets before training a neural network on this data. The preprocessing involved transforming the sunspot images into grayscale and resizing them. The training of the dataset was conducted using a 2D convolutional neural network. Finally, we tested the neural network model's results on the test dataset and achieved an accuracy of 99.8%. To implement the designed system, a graphical user interface was developed to enable astronomers to use this automatic diagnostic system.

Keywords- Deep learning, Sunspot classification, Astronomy, space weather

Résumé

Ce projet de fin d'études de cycle Master vise le recours aux images magnétogrammes, pour porter assistance aux astronomes dans la détection et la classification des taches solaires. L'analyse des taches solaires est d'une grande importance en astronomie, car elle permet de mieux comprendre le comportement et l'évolution du Soleil, identifier les relations avec d'autres phénomènes solaires et de prédire les éruptions solaires.

Les images magnétogrammes sont importantes dans la classification des taches solaires. Elles fournissent des informations sur le champ magnétique associé aux taches solaires. Dans le cadre de ce projet, nous avons d'abord collecter des images magnétogrammes des taches solaire pour construire un jeu de données qui est découpé en 3 parties (entraînement, validation et test). Puis, nous avons procédé à un prétraitement du jeu d'entraînement et de validation avant d'entraîner un réseau de neurones sur ce jeu de données. Ce prétraitement des données inclut la transformation des images des taches solaires en niveaux de gris et le redimensionnement. L'entraînement du jeu de données a été effectué sur un réseau de neurones convolutif 2D. Enfin, nous avons tester les résultats du modèle de réseau de neurones su le jeu de test et avons obtenu une précision de 99.8%. Pour mettre en place le système conçu, une interface graphique a été implémentée afin de permettre l'utilisation de ce système de diagnostic automatique par des astronomes.

Keywords— Deep learning, Sunspot classification, Astronomy, space weather

INTRODUCTION GÉNÉRALE

L'analyse des taches solaires revêt une grande importance dans le domaine de l'astronomie car elle permet aux astronomes de mieux comprendre le comportement et l'évolution du Soleil. Les astronomes chinois furent les premier à observer les taches solaires et les mentionner dans les manuscrit, mais ces caractéristiques n'ont pas été systématiquement étudiées jusqu'à l'invention de la lunette astronomique en 1609 [1] par l'astronome italien Galilée.

L'emplacement, la superficie et le mouvement des taches solaires au niveau de la surface sont des paramètres importants pour l'étude à long terme du Soleil[2].

Les taches solaires sont souvent des précurseurs des éruptions solaires. Elles sont importantes pour comprendre le cycle solaire, la dynamo solaire et les interactions Soleil-Terre. Les perturbations magnétiques qui arrivent sur Terre (orages géomagnétiques) sont véhiculées par les éjections de masse coronale (CMEs) sont également associées à des taches solaires instables. La surveillance des taches solaires est donc cruciale pour l'activité solaire et son effet sur la Terre et son environnement. En raison de la variabilité des taches solaires, de la présence de bruit et d'artefacts dans les images, ainsi que de l'absence de critères de caractérisation standardisés, leur détection et leur classification représentent des défis majeurs.

La classification des taches solaires est un processus manuel qui prend du temps car chaque tache doit être reconnue, marquée et classée [3]. La classification des groupes de taches solaires est importante pour prédire les éruptions.

L'utilisation de l'intelligence Artificielle dans la classification des taches solaires peut aider à surmonter les problèmes de classification manuelle en permettant une analyse plus rapide et plus précise de vastes ensembles de données. Les algorithmes d'IA peuvent détecter et classer automatiquement les taches solaires dans des ensembles d'images gigantesques, ce qui pourrait aider les chercheurs à mieux comprendre le comportement du Soleil et son effet sur le climat spatial [4]. La classification des taches solaires est donc un domaine très actif qui pourrait révolutionner la façon dont nous étudions notre astre du jour et ses effets sur notre planète. Dans cette recherche, nous proposons une nouvelle méthode de détection et de classification des taches solaires basées sur des techniques de traitement d'images et d'apprentissage automatique. Nous utilisons des techniques de traitement d'images, et des algorithmes de clustering pour le calcul des groupes des taches ainsi que le nombre de Wolf. Nous utiliserons également une architecture d'un réseau neuronal, pour classer les taches solaires en différents classes en fonction de la polarité de leur champs magnétique de et de leur évolution. Nous évaluerons nos méthodes sur différents données annotées par des experts, et nous les comparons aux méthodes existantes. Nous montrerons que nos méthodes atteignent une précision et une robustesse dans la détection et la classification des taches solaires, et qu'elles peuvent fournir des informations précieuses et importantes pour la recherche en physique solaire. Afin de présenter notre travail, ce document est divisé en 5 chapitres structurés de la manière suivante :

- Dans le chapitre 1, nous allons introduire quelques informations sur l'établissement d'étude CRAAG¹ ainsi que sa structure organisationnelle et ses missions.
- Dans le chapitre 2, nous introduirons le Soleil , l'activité solaire et le nombre de

^{1.} Centre de Recherche en Astronomie Astrophysique et Géophysique

Wolf, par la suite nous allons détailler l'apprentissage profond qui est exploité dans plusieurs travaux de recherche afin de faciliterla tâche aux astronomes afin. Enfin ,nous présenterons une synthèse de quelques travaux de détection et classification des taches solaires.

- Dans le chapitre 3, nous présenterons la première partie du projet qui traite le calcul de nombre de Wolf avec des méthodes de traitement d'images et le clustering.
- Dans le chapitre 4, nous allons aborder notre conception de la solution que nous avons proposé pour le problème de classification des taches solaire.
- Dans le chapitre 5, nous présenterons l'implémentation de notre système, qui va incluer les différents outils utilisés pour la réalisation de notre système, l'interface graphique mais également une évaluation une évaluation et une discussion sur les résultats obtenus et notamment l'évaluation de celui-ci.

Chapitre 1 : Environnement d'études

0.1 Introduction

Dans ce chapitre nous allons introduire quelque information sur l'établissement d'étude CRAAG. ainsi que sa structure organisationnelle comprend un directeur du centre, un directeur adjoint, un secrétaire général et plusieurs services administratifs, et ses missions.

0.2 Présentation du CRAAG

Le Centre de Recherche en Astronomie Astrophysique et Géophysique (CRAAG) est un Établissement Public à Caractère Scientifique et Technologique (EPST) en Algérie [5]. Le CRAAG est un organisme de recherche scientifique et technique spécialisé dans les domaines de l'astronomie, de l'astrophysique et de la géophysique. Il mène des recherches, des observations et des études dans ces domaines, contribuant ainsi à l'avancement des connaissances dans le domaine de l'espace et de la Terre. Il a été créé en 1985 par décret 83-521, mais ses origines remontent à la création de l'Observatoire d'Alger en 1890 et de l'Institut de Météorologie, de Physique du Globe et d'Astronomie (IMPGA) en 1931. La figure 1 représente le LOGO du centre.



FIGURE 1 – Logo du CRAAG

0.3 Historique du CRAAG

Le CRAAG est le résultat de fusion des deux organismes : l'Observatoire d'Alger, créé en 1890, et l'Institut de Physique du Globe d'Alger (IMPGA), créé en 1931.

L'Observatoire d'Alger a connu une intense activité scientifique durant cette période. Il a notamment participé au projet international Carte du Ciel, un projet de cartographie astronomique couvrant l'ensemble du ciel. Cette participation eut lieu entre 1909 et 1925, l'Observatoire d'Alger a contribué à la cartographie et à l'observation d'un grand nombre d'étoiles. De plus, l'Observatoire d'Alger a également fait d'importantes découvertes, notamment la découverte d'astéroïdes. L'une d'elles porte le nom de "Bouzaréah", en référence à l'endroit proche de l'observatoire.

0.4 Missions du CRAAG

Les missions du CRAAG sont [5] :

- Entreprendre des recherches et des études en astronomie, astrophysique et géophysique et d'exploiter les divers résultats qui en découlent.
- Assurer la surveillance sismique du territoire national de façon permanente et d'établir les liaisons nécessaires avec les autorités compétentes et les structures opérationnelles concernées.
- étendre et de densifier le réseau sismologique à l'ensemble du territoire national et d'en assurer l'exploitation et la maintenance.
- établir les catalogue et carte sismique, de recenser et de regrouper les documents anciens, y compris ceux existant à l'étranger, pour approfondir et enrichir l'histoire de la sismicité algérienne et constituer la banque des données sismiques.
- évaluer l'aléa sismique du territoire national.
- réaliser des études sur les risques géologiques et de participer la micozonation des centres urbains.
- participer au aux actions de réduction des risques sismiques.
- assurer le suivi permanent des phénomènes géophysique et astronomique naturels.
- développer les observations et les réseaux géophysique et astronomique.
- développer et d'étendre, a travers le territoire national, les stations et réseaux géomagnétiques astronomiques et services de l'heure ainsi que le réseaux gravimétrique et l'exception de réseaux fondamentaux.
- exploiter les données des observations astronomiques et géophysiques terrestres, aériennes et satéllitaires; D'entreprendre des études de la forme et des mouvement de la terre.
- entreprendre des études en astronomie fondamentales, solaire, satellitaire et des études liées à l'évolution de l'univers.
- mener des études sur les milieux interplanétaires et interstellaires.
- définir et conserver l'étalon du temps national et des fréquences et de contribuer à leur diffusion
- concevoir et de réaliser des instruments spécifiques aux travaux de recherche en astronomique et géophysique

0.5 Structure du CRAAG

La figure 2 montre l'organisation du Centre de Recherche en Astronomie Astrophysique et Géophysique **CRAAG**.



FIGURE 2 – Organisation du CRAAG

0.6 La recherche sur les taches solaire au CRAAG

La recherche sur la classification des taches solaires est essentielle pour faire progresser notre connaissance du Soleil, améliorer les capacités de prédiction des conditions météorologiques spatiales et approfondir notre compréhension de la relation Soleil-Terre. Le Centre de Recherche en Astronomie Astrophysique et Géophysique mènent des études sur la classification des taches solaires pour plusieurs raisons, telles que la compréhension de l'activité solaire, la prédiction des conditions météorologiques spatiales, Suivre le cycle solaire, Faire avancer la physique solaire.

0.7 Conclusion

Après avoir introduit l'établissement d'étude, sa structure organisationnelle et ses missions, dans le chapitre suivant, nous allons introduire le Soleil , l'activité solaire et le nombre de Wolf, par la suite nous allons détailler l'apprentissage profond qui est exploité dans plusieurs travaux de recherche afin d'aider les astronomes. Enfin nous présenterons une synthèse de quelques travaux de détection et classification des taches solaire.

Chapitre 2 : Etat de l'art

0.8 Introduction

Ce chapitre est consacré aux concepts de base liés au Soleil, taches solaires, la météorologie spatiale et à l'apprentissage profond. Dans les sections qui suivent, nous allons introduire les différentes couches du Soleil, l'activité solaire, et le nombre de Wolf. Puis nous allons détailler l'apprentissage profond qui est exploité actuellement dans plusieurs travaux de recherche dans le but d'aider les astronomes à calculer le nombre de taches solaire et classifier ces derniers. Enfin nous présenterons une synthèse de quelques travaux de détection et la classification de ces taches, basés sur l'apprentissage profond.

0.9 Le Soleil

Le Soleil est une étoile sous forme d'une gigantesque boule de gaz d'un volume 110 fois la taille de la Terre, un diamètre de 1,392 million de kilomètres et d'une masse d'environ 2x 1030 kilogrammes, et du moment qu'il est suffisament proche de la Terre, cela nous permet d'étudier sa surface et son environnement proche. Plus de 75% de le masse du Soleil est composé d'hydrogène, 25% d'hélium et 0.1% d'éléments plus lourds [6].

0.9.1 La Structure interne du Soleil

La partie interne du Soleil est elle même composée de 3 couches :

- Le noyau représente la partie dans laquelle l'énergie du Soleil est générée grâce à des réactions de fusion thermonucléaire où la température peut atteindre environ 15 millions degrés. Cette couche représente 25% du diamètre solaire et 60% de la masse du Soleil, ce qui lui attribue une grande densité.
- La zone radiative est une couche qui englobe le noyau, elle représente 45% du rayon du Soleil. L'énergie créée dans le noyau est transportée vers cette région par des photons.
- La zone convective correspond à la couche extérieure couvrant 30% du diamètre du Soleil . La température dans cette région descend sous le million degrés. Le transport d'énergie se fait par convection, ce qui signifie que le gaz chaud remonte vers la surface, libère l'énergie en se refroidissant puis retombe vers l'intérieur, et ainsi de suite.

0.9.2 La photosphère

La photosphère ou la surface du Soleil est une région de plusieurs centaines de kilomètres d'épaisseur qui est la dernière couche opaque et brillante que nous voyons en l'observant avec un instrument adéquat. Une légère baisse de température dans cette région provoque une diminution rapide de la densité, de sorte que le disque solaire a un contour bien défini plutôt que des limites floues. Les images de haute résolution montrent que la surface du Soleil a un aspect granuleux comme dans la figure 4, ces granules sont le résultat de la convection solaire.

Les granules ont une forme polygonale de quelques centaines de kilomètres pour les plus petite, jusq'à environ 2000km. La moyenne étant de 1000km. Chaque granule ne vit que



FIGURE 3 – Structure interne du Soleil



FIGURE 4 – Image à haute résolution de la surface du Soleil .

quelques minutes alors l'aspect de sa surface évolue et change. Le gaz qui sort du centre de grain a une température 300 degré plus élevée que celui qui replonge aux bords.

0.9.3 La chromosphère

C'est une couche qui est au-dessus de la photosphère. Elle a une épaisseur de quelques milliers de kilomètres. La température dans cette couche remonte de 4000 à 10.000 degrés, due à sa faible densité. La chromosphère est quasiment transparente, donc invisible en plein jour, par contre, elle est observable durant les éclipses solaires où elle apparaît comme un anneau rougeâtre mince qui entoure le bord solaire.

0.9.4 La couronne

C'est la couche la plus externe du Soleil, elle est aussi visible lors des éclipses totales du Soleil. La température dans cette couche s'élève à plus de un million degrés, ce qui est beaucoup plus élevée que la surface visible du Soleil.

0.10 L'activité solaire

0.10.1 Les taches solaires

Les taches solaires sont des régions sombres de la surface du Soleil dont le diamètre varie entre des milliers et des centaines de milliers de kilomètre et durent entre quelque jours et plusieurs mois. La photosphère est de l'ordre de 6000 degrés au niveau de taches solaires 3000 degré. Le champ magnétique dans ces régions est environ 2500 fois plus fort que celui de la Terre, bien plus élevé que partout ailleurs sur le Soleil. Le déplacement de ces champs magnétiques sont si puissants qu'ils empêchent une partie de la chaleur du Soleil d'atteindre la surface, c'est pourquoi ces régions apparaissent plus sombres que les autres parties de la surface du Soleil.

Structure des taches solaires

- L'ombre : La partie la plus sombre où le vecteur champ magnétique est pratiquement normal à la surface.
- La pénombre : La partie la moins sombre qui entour l'ombre où le vecteur champ magnétique est incliné.

0.10.2 La météorologie spatiale (space weather) :

La météorologie spatiale ou le space weather désigne les conditions environnementales qui prévalent dans l'espace interplanétaire qui entoure notre planète. L'activité du Soleil est son impact sur la Terre fait partie du climat spatiale. Le vent solaire est un élément important de l'activité solaire. Il s'agit d'un flux continu de particules chargées émis par le Soleil et se propage à travers tout le système solaire.

0.10.2.1 Éruptions solaires :

0.10.2.2 Coronal Mass Ejections (CMEs) : l'éjection de masse coronale (CME) est une vaste éruption de plasma et de champ magnétique provenant de la couronne du

Soleil . Les CME peuvent se déplacer à des vitesses allant jusqu'à 3000 km/s et transporter des milliards de tonnes de matière. Lorsqu'une CME atteint la Terre, elle peut interagir avec le champ magnétique terrestre et provoquer des tempêtes géomagnétiques, qui peuvent avoir des effets néfastes sur les satellites, les réseaux électriques, les communications, la navigation et la santé humaine.

0.10.3 Le cycle de l'activité solaire :

Le cycle solaire est un phénomène cyclique qui se répète approximativement tous les 11 ans, bien que la durée des cycles puisse varier. La durée moyenne d'un cycle solaire est de 11,2 ans. L'activité solaire est calculée en comptant le nombre de taches observées à la surface du Soleil . Lorsque le nombre de taches est élevé, cela signifie que l'activité solaire est élevée. A l'inverse, lorsque le nombre de taches solaires est faible, cela signifie une faible activité solaire. La figure desous montre le nombre de taches solaires en fonction de temps, on nous montre aussi la prédiction du 25ème cycle [7].



Progression du cycle solaire - Nombre de taches solaires

FIGURE 5 – Progression du cycle solaire

0.11 Le nombre de Wolf

Le nombre de Wolf est un indicateur quantitatif reflétant l'activité globale sur le Soleil à l'aide du nombre de taches solaires et du nombre de groupes de taches solaires présents à la surface du Soleil . Le nombre de Wolf est une mesure utilisée pour étudier l'activité du Soleil . Ce nombre de taches solaires nous fournit des informations précieuses sur le comportement du Soleil et nous aide à prédire et comprendre différents effets sur la Terre. Une augmentation du nombre de taches solaires est généralement associée à une activité solaire intense, ce qui accroît la probabilité d'éruptions solaires et d'éjections de masse coronale. En surveillant attentivement le nombre de taches solaires, les scientifiques peuvent

évaluer le potentiel de ces événements et émettre des avertissements concernant leur impact sur la magnétosphère terrestre, les satellites, les réseaux électriques et les systèmes de communication.

Le nombre relatif de taches solaires R est calculé à l'aide de la formule :

$$R = k(10g + f)$$

 ${\bf f}$: est le nombre de taches

g : le nombre de group

0.12 La classification magnétique des taches solaires

La classification magnétique des taches solaires est une méthode qui permet de décrire la complexité et le potentiel éruptif des régions actives du Soleil . Selon la distribution et l'orientation des polarités magnétiques au sein d'un groupe de taches solaires, on distingue plusieurs types de classifications : alpha, bêta, gamma, delta, ou des combinaisons de ces lettres. Plus la classification est complexe, plus le risque d'éruption solaire est élevé.Les taches solaires de type delta sont les plus dangereuses, car elles contiennent des ombres de polarités opposées dans une même pénombre, ce qui crée des conditions favorables à la reconnexion magnétique et à la libération d'énergie. Les taches solaires sont également liées au cycle solaire, qui correspond à la variation périodique de l'activité solaire sur une durée d'environ 11 ans. Le nombre et la taille des taches solaires varient en fonction du cycle solaire, ainsi que l'intensité du rayonnement solaire qui atteint la Terre. Les taches solaires sont donc un indicateur important de l'état du Soleil et de son influence sur le climat terrestre.

0.12.1 La classification magnétique du mount Wilson-Mcintosh :

La classification magnétique du mont Wilson-McIntosh est un système utilisé pour classer les groupes de taches solaires en fonction de leurs propriétés magnétiques. Le système de classification a été développé par *Patrick McIntosh* et est basé sur la forme générale du groupe de taches solaires et sa configuration magnétique. La classification magnétique des groupes de taches solaires du Mont Wilson est considérée comme utile pour prévoir les éruptions solaires. Le système de classification est basé sur trois éléments : la forme générale du groupe de taches solaires, la polarité principale du groupe de taches solaires et la polarité suivante du groupe de taches solaires.

Le système de classification du Mont Wilson comporte quatre classes importante, alpha, bêta, gamma et delta.

0.13 L'apprentissage profond

L'apprentissage profond est un ensemble de méthodes d'apprentissage automatique qui permet à l'ordinateur de construire des concepts complexes à partir de concepts plus simples. L'idée d'apprendre la bonne représentation des données fournit une perspective sur l'apprentissage en profondeur. La profondeur permet à l'ordinateur d'apprendre un programme informatique en plusieurs étapes [8]. Au cours des dernières années, l'apprentissage profond a eu un impact important sur la science et la technologie. Il a automatisé la reconnaissance vocale, textuelle ou encore celle des images. La reconnaissance des images a notamment apporté du progrès dans le domaine de l'astronomie. L'apprentissage profond fournit une analyse pertinente des données.

0.13.1 Les réseaux de neurones

Un réseau de neurones est un modèle d'apprentissage informatique dont la structure est similaire à celle du réseau de neurones du cerveau humain. Un noeud de neurone reçoit une ou plusieurs entrées, applique des poids appropriés aux signaux, et passe les entrées pondérées au prochain neurone [9]. La Figure 6 illustre le schéma d'un neurone artificiel qui reçoit plusieurs entrées. Le regroupement de plusieurs neurones forme un réseau de neurones. Un réseau de neurones est structuré en couches tel que chacune des couches prend en entrée les neurones de la couche précédente. Les réseaux de neurones ayant 2 couches, une couche en entrée et une en sortie est dit simple ou **Perceptron**.



FIGURE 6 – Aperçu d'un neurone artificiel ayant plusieurs entrées

Le réseau de neurones ayant en plus des couches d'entrée et de sortie des couches supplémentaires est dit **perceptron multicouche**. Il dispose entre la couche en entrée et la couche en sortie d'une ou de plusieurs couches dites « cachées ». La figure 7 illustre la structure d'un réseau de neurones perceptron ayant une couche cachée.



FIGURE 7 – Aperçu d'un réseau de neurone ayant une couche cachée

Les neurones d'entrée sont chargées de recevoir des données, de les traiter puis les transmettre aux neurones de la couche suivante. Les neurones sont caractérisés par des poids et biais qui sont des valeurs numériques et elles sont mises à jour constamment par les réseaux pour obtenir de meilleurs résultats en sortie. Plus le réseau de neurones contient un nombre important de couches cachées, plus il fait des traitements pour mieux ajuster ses poids. Il devient donc un réseau de neurone profond. Les réseaux de neurones sont aujourd'hui la solution à la problématique d'apprentissage de la machine car ils peuvent recevoir et traiter des données complexes puis apprendre d'eux même grâce à des algorithmes d'apprentissage leurs permettant de mettre à jour leurs poids. Il existe de différents types de réseaux de neurones : les réseaux de neurones récurrents et réseaux de neurones convolutifs. Les réseaux de neurones convolutifs sont principalement utilisés pour la reconnaissance et la classification d'image.

0.13.2 Les reseaux de neurones convolutifs

Les réseaux de neurones convolutifs dit RNC sont une classe de réseaux de neurones profonds utilisée pour l'analyse des images. Ils sont composés d'une succession de blocs de traitement pour extraire les caractéristiques discriminant la classe d'appartenance de l'image des autres classes [10]. La figure 8 illustre la structure d'un réseau de neurones convolutif.

Les couches principales des RNCs pour le traitement de l'image et extraction des caractéristiques en entrée sont [11] :

— Couche de convolution (CONV) : Couche d'un RNC chargée d'extraire et de retourner les caractéristiques (features) d'une image, et ce en calculant la convolution²

^{2.} La convolution est une opération mathématique qui prend 2 arguments. Dans la terminologie des RNCs, l'image et le filtre qui lui est appliqué représentent respectivement le premier et le second argument de la convolution



FIGURE 8 – Aperçu d'un réseau de neurones convolutif

entre l'image et le filtre.

- Couche de correction (ReLU) : Couche d'un RNC, elle est aussi appelée "ReLu " en référence à la fonction d'activation (Unité de rectification linéaire) qui très utilisée en apprentissage profond. Elle est appliquée en sortie d'une couche de convolutions pour briser partiellement la linéarité car la linéarité d'une image n'est pas importante. Cette fonction permet de creuser l'écart entre deux caractéristiques.
- Couche de pooling (POOL) : Elle permet de réduire la résolution des cartes d'activation issues de la couche de convolution tout en gardant les caractéristiques les plus importantes afin de minimiser les paramètres et les calculs au sein du modèle.
 Max pooling et Average pooling sont des opérations de pooling où le maximum et la moyenne de la carte d'activation sont respectivement pris.
 Les couches de classification dans les RNCs sont :
- Couche de mise a plat (Flatten) : Elle consiste à mettre bout à bout toutes les images (matrices) et en faire un (long) vecteur. Cette étape est la dernière de la partie extraction des informations.
- Couche « entièrement connectée » (Fully Connected) : Couche de type Perceptron qui permet la classification de l'image à partir des caractéristiques extraites par la succession de blocs de traitement entièrement connectée. Toutes les entrées de la couche sont connectées aux neurones de sorties de celle-ci car chaque neurone attribue à l'image une valeur de probabilité d'appartenance à la classe i parmi les c classes possibles.
- couche de perte (LOSS) qui est la dernière couche du réseau. Elle calculée sur base de l'erreur entre la prédiction du réseau et la valeur réelle. Parmi les fonctions de perte les plus connues nous pouvons citer la fonction Softmax et la fonction Minimum Squared Error (MSE)

0.13.3 Jeu de données d'un réseau de neurones convolutifs

Le processus d'entraînement dans un réseau de neurone consiste à ajuster les poids nécessaires des nœuds en fonction de la véracité du résultat prédit par le réseau et la vérité terrain qui est une sortie connue par le modèle. Ainsi, le modèle sera entraîné à reconnaître des tendances, classer des données et prévoir des événements à venir. Le jeu d'entraînement est l'échantillon de données biaisé qui sera présenté en entrée au réseau de neurones pour s'entraîner à apprendre les différentes caractéristiques (features) de ces données.

0.13.4 Concept d'apprentissage profond

L'apprentissage profond (ou deep learning) est un sous-domaine de l'intelligence artificielle qui utilise des réseaux de neurones artificiels multicouches pour apprendre et découvrir des idées à partir des données. Les modèles d'apprentissage profond sont capables d'apprendre par eux-mêmes et sont utilisés dans de nombreux domaines pour résoudre des problèmes complexes.

0.13.5 Pré-traitement sur les données

Le pré-traitement est une série de transformations appliquées à une image initiale pour améliorer la qualité de l'image et permettre une analyse statistique comparable. Le prétraitement des données comprend entre autre la segmentation, la réduction de bruit et la normalisation.

0.14 Synthèse de quelque approches de détection et classification des taches solaires

Les tableaux 1 et 2 illustre une synthèse de notre étude de quelques travaux sur les differentes approches de detection et classification des taches solaires Ces travaux sont décrits selon les 3 critères suivants :

- Méthode, Architecture et paramètres du modèl : spécifie la technique utilisé (apprentissage par transfert, clustering..etc)
- La classification du modèle : elle peut étre une classification zurich ou mount wilson ou même une classification binaire comme dans l'étude 1 (quiet sun, sunspot)
- Le jeu de données utilisé : elle dépend de la classification du modèle ou la méthode utilisé, des images magnetogram intensitygram, continuum, ou two channel pictures.

Bien que de nombreux travaux aient été réalisés sur la reconnaissance automatique des taches solaires, il existe peu d'études sur la classification automatique des types magnétiques des taches solaires par des méthodes d'apprentissage automatique. Channabasav et al [4], Ils ont appliqué un réseau de neurones convolutionnel (CNN) basé sur AlexNet ainsi qu'une machine à vecteurs de support (SVM) pour la classification et ont obtenu une précision de 99,72%. Nguyen et al [12], et Sinh Hoa Nguyen et al [13] ont proposé des systèmes de classification sur la base de la classification zurich, Nguyen et al. ont utilisé une méthode hiérarchique basée sur un système approximatif, tandis que Sinh Hoa Nguyen et al. ont utilisé un réseau neuronal complètement connecté pour leur approche de classification. Yuanhui Fang et al [14] ont utilisé 3 modéle RNC pour 3 types d'images (magnetogramme, continuum, two channel pictures), la classification était faite sur 3 classes alpha, beta et beta-X. Contrairement aux travaux étudiés, nous allons exploiter dans ce travail une architecture RNC 2D. en utilisant le système de classification de Mount Wilson, les classes magnétiques des régions actives sont étiquetés comme étant l'un des groupes unipolaires Alpha, le groupe bipolaire Beta, ou d'autres groupes complexes multipolaires, Beta-x. Ces trois types magnétiques sont identifiés automatiquement en utilisant la méthode du réseau neuronal convolutif RNC, avec des images magnétogrammes collectées à partir du site web Space Weather Live, prises durant la période 2013-2018.

Référence	Méthode, Architecture et		
	paramètres du modèle	Classification du modèle	Jeu de données
Channabasav			
et al.[4]	Alexnet based CNN and SVM for classification 99.72% accuracy	2 classes sunspot, quiet sun	1000 MDI sunspot images, 1000 MDI quiet sun images, 1000 HMI sunspot images,
			1000 IIMI duter suit itilages.
Nguyen et al.[12]	hybrid système for detection clustering and classification of sunspots.	7 classes (modified zurich scheme)	Non disponibles
Sinh Hoa Nguyen			
	méthode d'apprentissage	classification Zurich	NASA/SOHO [15]
et al. (2005) [13]	hiérarchique basée sur des	représenté par	ARMaps
	ensembles approximatifs	l'arbre de décision.	Hawaii University [16]
Colak			
et al.[17]	Preprocessing (sunspots from intensitygrams, AR ³ from magnetogram) fully connected neural networks	systèm de classification McIntosh	SOHO/MDI magnetogram and intensitygram images.[18]
TADIE 1	Cunthèse de anolanes transmus de la	détection et cleasification des	tachas coleiras neutio 1

_
artie .
рa
solaires
taches
des
classification 4
et
tection e
Ę,
la
sur
travaux
quelques
de
– Synthèse (
\mathbf{TABLE}

3. Active Regions

Yuanhui FangDaramètres du modèleJeu de deYuanhui Fang3CNN models (A, B et C)Classification du modèleJeu de dea 3CNN models (A, B et C)Mount WilsonMount WilsonImages and two-channel picturesImages and two-channel picturesImages and two-channel picturesImages and two-channel picturesImages alphaSDO/HMIet al.[14]images and two-channel picturesImages and two-channel picturesImages alphaSDO/HMIon magnetogram, continuum imagesclassification, classe betaXImput layer (227, 227, 1)Imput layer (227, 227, 1)Imput layer (227, 227, 1)Our approachConv2d Batch norm max pooling)x2Mount WilsonConv2d Batch norm max pooling)x2Mount WilsonSpace weaOur approachFlatten, dense with softmaxclassification, classe letaXscrapped oscrapped oactivationlearning rate of 0.0001and batch size of 32with 200 epochsscrapped o	Référence	Méthode, Architecture et		
Yuanhui Fang 3CNN models (A, B et C)Mount Wilsonat al.[14]3CNN models (A, B et C)Mount Wilsonon magnetogram, continuum images and two-channel pictures respectively, resulting 95% accuracy on continuum imagesMount Wilsonclassification, classe beta, classe beta on continuum imagesSDO/HMIAlexnet based architecture, Input layer (227, 227, 1) (Conv2d Batch norm max pooling)x2 (Conv2d Batch norm max pooling)x2 (Conv2d Batch norm max pooling)x2 (Conv2d Batch norm max pooling)x2 activationMount Wilson classe beta, classe alpha scraphed classification, classe alpha scraphed classification classe beta, classe beta, classe beta, classe alpha scraphed classification dates beta, classe beta, classe beta, classe beta, classe beta, classe beta, classe beta scraphed classe beta, classe beta, classe beta, classe beta scraphed classe beta, classe beta, classe beta, classe beta scraphed classe beta, classe beta, classe beta		paramètres du modèle	Classification du modèle	Jeu de données
at al. [14]3CNN models (A, B et C) on magnetogram, continuum images and two-channel pictures respectively, resulting 95% accuracy on continuum imagesMount Wilson classification, classe alpha classe beta, classe beta classe beta, classe beta classe beta, classe beta classe beta, classe alpha classe beta, classe alpha beta drate Input layer (227, 227, 1) Conv2d Batch norm max pooling)x2 Conv2d Batch norm max pooling)x2 Conv2d Batch norm max pooling)x2 Conv2d Batch norm max pooling)Mount Wilson classe beta, classe alpha scrapsed classe beta, classe alpha scrapped classe beta, classe beta scrapped classe beta, classe beta scrapped classe beta, classe beta	Yuanhui Fang			
et al.[14]on magnetogram, continuum images and two-channel pictures respectively, resulting 95% accuracy on continuum imagesMount Wilson classification, classe alpha classe beta, classe betaXSDO/HMIAlexnet based architecture, Input layer (227, 227, 1) (Conv2d Batch norm max pooling)x2 (Conv2d Batch norm max pooling)x2 (Conv2d Batch norm max pooling)x2 (Conv2d Batch norm max pooling) (Conv2d Batch norm max pooling) activationMount Wilson classe beta, classe alpha scraphed c classe beta, classe betaXOur approach ad match size of 32 with 200 epochsElastification, classe alpha classe beta, classe betaX		3CNN models (A, B et C)		
et al.[14]images and two-channel pictures respectively, resulting 95% accuracy on continuum imagesclassification, classe betaX classe beta, classe betaXSDO/HMI SDO/HMInew continuum images on continuum imagesAlexnet based architecture, Input layer (227, 227, 1) (Conv2d Batch norm max pooling)x2 (Conv2d Batch norm)x2 (Conv2d Batch norm)x2 (Conv2d Batch norm)x2 (Conv2d Batch norm)x2 mount Wilson flatten, dense with softmax activationMount Wilson classification, classe alpha scraphed classification, classe beta, classe betaX		on magnetogram, continuum	Mount Wilson	
respectively, resulting 95% accuracy on continuum imagesclasse beta, classe betaXAlexnet based architecture, Input layer (227, 227, 1) (Conv2d Batch norm max pooling)x2 (Conv2d Batch norm max pooling)x2 (Conv2d Batch norm max pooling)x2 (Conv2d Batch norm max pooling)Mount Wilson classification, classe alpha scrapped c classe beta, classe betaXOur approach approach activationFlatten, dense with softmax classe beta, classe beta, classe betaXSpace wea scrapped c classification, classe betaX	et al.[14]	images and two-channel pictures	classification, classe alpha	SDO/HMI SHARP
on continuum imageson continuum imagesAlexnet based architecture, Input layer (227, 227, 1) (Conv2d Batch norm max pooling)x2 (Conv2d Batch norm max pooling)x2 (Conv2d Batch norm max pooling) (Conv2d Batch norm max pool		respectively, resulting 95% accuracy	classe beta, classe betaX	
Alexnet based architecture, Input layer (227, 227, 1)Alexnet based architecture, Input layer (227, 227, 1)Our approach Conv2d Batch norm max pooling) (Conv2d Batch norm max pooling) (Conv2d Batch norm max pooling) Flatten, dense with softmax activationMount Wilson classification, classe alpha classification, classe beta, classe beta, classe beta scrapped c dastivation		on continuum images		
Input layer (227, 227, 1)Conv2d Batch norm max pooling)x2(Conv2d Batch norm max pooling)(Conv2d Batch norm)x2(Conv2d Batch norm)x2(Conv2d Batch norm max pooling)(Conv2d Batch norm max pooling)<		Alexnet based architecture,		
Our approach(Conv2d Batch norm max pooling)x2 (Conv2d Batch norm max pooling) (Conv2d Batch norm max pooling) Flatten, dense with softmax activationMount Wilson classification, classe alpha classification, classe beta, classe beta scrapped c and batch size of 32 with 200 epochsSpace wea scrapped c		Input layer $(227, 227, 1)$		
Our approach(Conv2d Batch norm)x2 (Conv2d Batch norm max pooling)Mount Wilson classification, classe alpha classificationSpace wea scrapped c ascrapped c i and batch size of 32 with 200 epochs		(Conv2d Batch norm max pooling)x2		
Our approach (Conv2d Batch norm max pooling) MOULD WILSOIL Space wea Flatten, dense with softmax classification, classe alpha Space wea activation activation classe beta, classe betaX scrapped classe add batch size of 32 with 200 epochs scrapped classe scrapped classe		(Conv2d Batch norm)x2	M_{\odot}	
Our approact Flatten, dense with softmax classification, classe alpua scrapped classe activation activation classe beta, classe betaX scrapped classe activation and batch size of 32 with 200 epochs scrapped classe	denominan un	(Conv2d Batch norm max pooling)	IVIOUIIU VVIISOUL	Space weather
activation learning rate of 0.0001 and batch size of 32 with 200 epochs	our approach	Flatten, dense with softmax	Classification, classe alpita	scrapped data[19]
learning rate of 0.0001 and batch size of 32 with 200 epochs		activation	classe deta, classe deta Λ	
and batch size of 32 with 200 epochs		learning rate of 0.0001		
with 200 epochs		and batch size of 32		
		with 200 epochs		

TABLE 2 – Synthèse de quelques travaux sur la détection et classification des taches solaires partie 2

0.15 Conclusion

Après avoir abordé le Soleil et ses caractéristiques, nous avons détaillé l'apprentissage profond et présenté une synthèse de quelques recherche relative à la détection et classification des taches solaire, sur des images magnétogrammes et d'autres types d'images. Dans le chapitre suivant, nous allons abordé la première partie de notre étude qui est le traitement des images et le calcul de nombre de Wolf, ainsi que la collecte des données utilisé dans cette partie.

Chapitre 3 : Traitement d'images et calcul de nombre du Wolf

0.15.1 Introduction

Après avoir abordé les différentes disciplines de l'apprentissage profond et avoir présenté quelques travaux sur la détection et la classification des taches solaires, ce chapitre sera consacré au traitement d'images et le calcul de nombre de Wolf. il sera présenté la méthode utilisée pour la collecte du jeu de données utilisé pour cette partie. Puis, seront détaillées les différentes phases du traitement d'images apporté à ces données. Ensuite, le calcul du nombre de Wolf qui est une tache importante dans notre étude pour faire la prédiction du cycle solaire suivant. Et enfin une explication sur l'utilisation du système avant la conclusion.

0.15.2 Jeu de données

L'approche que nous proposons utilise des images complètes du disque solaire capturées par l'instrument SDO HMI⁴. Ces images ont été utilisées pour développer et évaluer une nouvelle méthode de calcul du nombre de Wolf. Le jeu de données utilisé dans cette étude comprend des images collectées du 1er janvier 2013 au 1er juin 2023. En collectant les données à l'aide d'un script du web scraping, nous avons rassemblé les informations nécessaires pour cette étude. Les images ont été extraites de l'archive des images SDO disponibles sur le site web de la NASA [20]. Quant aux nombres de Wolf correspondants, ils ont été collectés à partir du Royal Observatory of Belgium, une source réputée et fiable dans ce domaine [21]. Cette approche nous permet d'analyser et d'évaluer notre nouvelle méthode de calcul du nombre de Wolf basée sur le jeu de données collecté.

0.15.3 Traitement d'images

Le processus de calcul du nombre de Wolf à partir d'une image solaire comprend plusieurs étapes de traitement d'images essentielles pour obtenir des résultats précis et fiables. L'algorithme 1 résume toute les étapes réalisé dans la partie du calcul de nombre de Wolf :

^{4.} Solar Dynamics Observatory - Helioseismic and Magnetic Imager

Algorithme 1: Algorithme général du calcul de nombre de Wolf

```
Require: Image, \epsilon, threshold
```

Ensure: Wolf number

```
Image \leftarrow Lire_Image();
Max\_contour \leftarrow trouver\_Max\_contour(Image);
x, y, height, width \leftarrow Features_Max_contour(Max_contour);
cropped_image \leftarrow Image[y : y + height, x : x + width];
blurred_image \leftarrow Gaussian_blur(Image);
Sobel_x \leftarrow sobel_xaxis(blurred_image);
Sobel_v \leftarrow sobel_vaxis(blurred_image);
gradient_image \leftarrow square(Sobel_x<sup>2</sup> + Sobel_y<sup>2</sup>);
Normalized_gradient \leftarrow normalize_gradient(gradient_image);
Image_binaire \leftarrow threshold(Normalized_gradient, threshold_value);
Liste_contours \leftarrow Trouver_contours(Image_binaire);
Liste\_features \leftarrow Liste();
foreach contour in Liste_contours do
      Surface\_contour \leftarrow trouver\_surface(contour);
      Pos_x, Pos_y \leftarrow position\_contour(contour);
      if Surface\_contour < 1000 ou Surface\_contour > 2 then
          Liste_features.insere([Pos_x, Pos_y, Surface_contour]);
      end
     end
Min_samples \leftarrow 1;
labels \leftarrow DBSCAN(Liste_features, eps, Min_samples);
nbr_cluster \leftarrow trouver_nombre_clusters(labels);
Wolf_number \leftarrow labels.size() + nbr_cluster * 10;
return Wolf_number;
```

0.15.3.1 Recadrage de l'image : Le recadrage de l'image est une étape importante du processus de calcul du nombre de Wolf à partir d'une image solaire. Son objectif principal est d'isoler le disque solaire complet en éliminant les éléments indésirables qui pourraient interférer avec la détection des taches solaires. La figure **??** est l'image du disque solaire complet.

0.15.3.2 La réduction du bruit Avant de détecter les contours de l'image, un filtre gaussien a été appliqué pour prétraiter l'image et réduire le bruit. Le filtre Gaussien est un filtre de lissage qui atténue les variations brutales d'intensité de l'image. Son équation mathématique est la suivante :

$$G(x,y) = \frac{1}{2\pi\delta^2} * e^{-\frac{x^2 + y^2}{2\delta^2}}$$
(1)



En appliquant ce filtre à une image, les valeurs des pixels sont modifiées en fonction de leur proximité avec les pixels voisins. Cela permet de réduire le bruit et de lisser les variations de l'intensité des pixels, ce qui peut rendre la détection des sunspots plus précise en éliminant les détails indésirables tels que les granules du soleil.

0.15.3.3 Utilisation de l'opérateur de Sobel : L'opérateur de Sobel, également connu sous le nom de filtre de Sobel, est largement utilisé dans le traitement d'images et la vision par ordinateur, en particulier dans les algorithmes de détection des contours. Il a été développé par I. Sobel et G. Feldman en 1968. Cet opérateur utilise deux matrices de convolution, l'une pour détecter les changements horizontaux et l'autre pour détecter les changements verticaux dans une image. Ces matrices sont appliquées en effectuant une convolution avec l'image originale, ce qui permet de calculer des approximations des dérivées de l'image par rapport aux axes horizontal et vertical. La figure 13 représente l'image résultante après l'application du filtre de Sobel pour détecter les contours des taches solaires.



FIGURE 9 – Images des contours des taches

En combinant les résultats de ces deux convolutions, l'opérateur de Sobel crée une nouvelle image qui met en évidence les contours et les bords de l'image originale. Les valeurs des pixels dans cette image résultante représentent l'intensité des gradients dans les directions horizontale et verticale, fournissant ainsi des informations sur les variations rapides de luminosité et de couleur qui indiquent la présence de contours dans l'image. Pour une Image I, Gx et Gy sont deux images qui, à chaque point, contiennent respectivement les approximations des dérivées horizontales et verticales. La figure 10(a) et 10(b) représentent respectivement les dérivées horizontale et verticale d'une image de dique solaire.



(a) image dérivée horizontale



(b) image dérivée verticale

Après avoir calculé les gradients Gx et Gy, nous avons calculé le gradient total en utilisant la formule suivante :

$$G_T = \sqrt{G_x^2 + G_y^2} \tag{2}$$

Le calcul du gradient total nous permet de quantifier la magnitude du changement d'intensité des pixels dans différentes directions. Ainsi, nous pouvons identifier les zones de l'image où les contours sont plus prononcés. La figure 10(c) est la figure initiale du disque solaire dont nous avons appliqué le filtre sobel précédement, et la figure 10(d) représente les deux images dérivé combiner ensemble par le calcul de gradient.



(c) image originale



(d) image combinée

0.15.3.4 Binarisation d'image Après la détection des contours avec l'opérateur de Sobel et le calcul du gradient total, nous appliquons un seuillage pour obtenir une image binaire. Les pixels correspondant à un sunspot sont définis comme 1, tandis que les autres pixels sont définis comme 0. Cela permet de segmenter l'image en deux régions distinctes : les sunspots et le reste de l'image. Cette étape facilite l'analyse ultérieure des sunspots dans l'image (Voir Figure 10).



FIGURE 10 – Image binarisée

Une fois les taches solaires identifiées dans l'image, il est essentiel de trouver leurs groupes correspondants afin de pouvoir les analyser et calculer le nombre de Wolf. Pour réaliser cette tâche, nous avons utilisé le clustering.

0.15.4 Clustering

Le clustering est une méthode d'apprentissage non supervisé qui regroupe des objets similaires en fonction de leurs caractéristiques communes. Dans notre étude, nous avons utilisé le clustering pour regrouper les taches solaires en se basant sur leur taille et leur emplacement spatial.

La taille des taches solaires a été utilisée comme critère de distinction pour les différencier des autres éléments présents dans l'image. Si une tache solaire dépasse un certain seuil de taille, elle est considérée comme non pertinente. De même, si elle est trop petite, elle est également exclue de l'analyse car elle ne correspond pas aux caractéristiques typiques d'une tache solaire.

En plus de la taille, nous avons également pris en compte l'emplacement spatial des taches solaires. Cela nous a permis de comprendre la structure et la distribution des taches solaires sur la surface du soleil. En regroupant les taches solaires qui partagent des positions spatiales similaires.

Une fois les caractéristiques des contours détectées, nous avons appliqué l'algorithme DBS-CAN [22] pour effectuer une segmentation des contours en groupes cohérents.

0.15.5 Algorithme utilisé

L'algorithme DBSCAN [22] a été choisi pour la détection des groupes de sunspots en raison de ses avantages spécifiques. Contrairement à l'algorithme k-means qui nécessite la spécification préalable d'un paramètre k pour déterminer le nombre de clusters, DBSCAN ne nécessite pas cette information. Cela le rend plus flexible et adapté à des situations où le nombre de clusters n'est pas connu à l'avance ou peut varier. De plus, DBSCAN est réputé pour préserver la forme des groupes, ce qui signifie qu'il peut identifier des clusters de formes arbitraires, y compris des formes complexes et non linéaires, ce qui est particulièrement important pour la détection des sunspots. Ainsi, l'utilisation de DBS-CAN nous a permis de trouver efficacement les groupes de sunspots tout en évitant les contraintes de spécification du nombre de clusters et en préservant la forme des groupes.

Algorithme 2: Algorithme de DBSCAN Input Dataset D, Epsilon ϵ , Minimum points MinPtsOutput Clusters C $C \leftarrow \emptyset$: // Set of clusters visited $\leftarrow \emptyset$; // Set of visited points foreach point p in D do if p is not visited then mark p as visited; $NeighborPts \leftarrow regionQuery(p, \epsilon);$ if size(NeighborPts) < MinPts then mark p as noise; end else $C_{new} \leftarrow \text{next cluster};$ expandCluster(p, NeighborPts, C_{new} , ϵ , MinPts); add C_{new} to C; end end end



FIGURE 11 – Taches solaires ségmentés

0.15.6 L'évaluation des résultats et test sur les données fournies par le CRAAG

Pour évaluer l'efficacité de notre programme, nous avons tracé la courbe du cycle solaire sur une période allant de mai 2010 à mai 2023. Nous avons ajusté les paramètres du système, notamment le filtre du bruit, afin d'obtenir une courbe plus précise du cycle solaire. Ensuite, nous avons comparé notre courbe avec la courbe de la Moyenne mensuelle des taches solaires du Royal Observatory of Belgium.

La figure 12 présente trois graphiques représentant la courbe du cycle solaire. Le premier graphique a été tracé en appliquant un filtre gaussien d'une seule itération avec un noyau de (5,5). Le deuxième graphique correspond à la courbe obtenue après l'application d'un filtre de délétion et d'érosion à l'image, en plus du filtre de bruit. Le troisième graphique présente la courbe obtenue après l'application du filtre de bruit avec deux itérations.

En analysant les graphiques, on observe que le premier graphique présente la meilleure représentation du cycle solaire. Il met en évidence la présence d'un cycle solaire avec un maximum autour des années 2013-2014, suivi d'une décroissance jusqu'aux années 2017-2019.

En comparant la courbe obtenue avec celle du Royal Observatory of Belgium, on constate une marge de différence d'environ 23.34. Cette valeur représente la moyenne des écarts entre notre courbe et celle du Royal Observatory of Belgium.

Il est important de noter que la différence observée entre notre courbe et celle du Royal Observatory of Belgium, notamment pendant les années d'activité solaire intense, peut être attribuée au coefficient K du nombre de Wolf. Le coefficient K est utilisé pour corriger les mesures effectuées par différents observatoires afin de les rendre comparables.

Dans notre programme, le coefficient K a été fixé à 1, ce qui signifie que nous n'avons pas appliqué de correction aux mesures de taches solaires. En revanche, le Royal Observatory of Belgium utilise les résultats du nombre de Wolf provenant de plusieurs stations, mais les détails précis sur le calcul du coefficient K ne sont pas clairement spécifiés.

Il est important de prendre en compte cette différence dans l'interprétation des résultats et de reconnaître que notre courbe du cycle solaire peut présenter des variations par rapport à celle du Royal Observatory of Belgium en raison de cette différence dans l'utilisation du coefficient K.

Nous avons également testé l'algorithme sur une image capturée par la lunette du GRAAG le 27 mars 2023. Malgré la qualité inférieure de l'image par rapport aux images provenant du satellite SDO, nous avons obtenu de bons résultats. Le nombre de Wolf calculé pour cette image était de 107, ce qui indique une activité solaire significative. De plus, nous avons identifié 9 groupes de sunspots dans l'image, ce qui suggère une distribution spatiale intéressante des taches solaires.

En comparant notre résultat avec l'estimation du nombre de Wolf fournie par le NOAA Space Weather Prediction Center de la National Oceanic and Atmospheric Administration [23] pour cette même journée, qui était de 125, nous constatons une concordance satisfaisante. Malgré la différence de qualité d'image et de méthode d'estimation, notre algorithme parvient à fournir des résultats cohérents et précis.



FIGURE 12 – Progression du cycle solaire





FIGURE 13 – Détection et segmentation des taches solaires sur l'image originale capturée par la lunette du GRAAG le 27 mars 2023

0.15.7 Conclusion

Nous avons présenté dans ce chapitre les différentes étapes et traitements effectuées par lesquels passent les images collectées à partir du web avant de passer par un algorithme de clustering automatique. Dans le chapitre qui suit, nous présenterons la deuxième partie cruciale de notre projet qui consiste à la classification des images de taches solaires.

Chapitre 4 : Conception d'un système de classification des taches solaires

0.15.8 Introduction

Après avoir abordé les différentes disciplines de l'apprentissage profond et avoir présenté quelques travaux sur la détection et classification des taches solaires dans le premier chapitre, Dans ce chapitre, nous allons nous concentrer sur l'étude conceptuelle de notre système de classification des taches solaires. Nous commencerons par présenter l'ensemble de données utilisé pour cette étude, ainsi que la méthode de collecte des données. Ensuite, nous aborderons l'architecture globale du système de classification des taches solaires. Cela comprendra une description des différents modules et composants du système. Par la suite, nous détaillerons les différentes phases de pré-traitements appliquées aux données d'entrée. Les pré-traitements sont des étapes importantes qui permettent de préparer les données avant de les utiliser dans le processus de classification. Nous aborderons des techniques telles que le redimensionnement des images, la normalisation des valeurs, etc. L'objectif de cette partie conceptuelle est de fournir une compréhension approfondie du système de classification des taches solaires, en mettant l'accent sur les aspects liés aux données et aux pré-traitements. Cela nous permettra de poser les bases nécessaires pour la mise en œuvre pratique du système dans les chapitres suivants.

0.15.9 La collect de jeu de données

La collecte de données est une phase essentielle dans l'entraînement d'un modèle de classification, en particulier dans les cas où les données annotées sont limitées. Pour remédier à cette limitation, nous avons décidé de créer notre propre base de données. Space Weather Live [24] est un site web qui fournit des données visualisées, y compris des images annotées. Nous avons donc choisi d'utiliser des techniques de "web scraping" pour collecter les données nécessaires à partir de ce site web. L'objectif du web scraping est d'extraire automatiquement des informations de pages web en analysant leur structure HTML ou en utilisant les API disponibles. En utilisant le web scraping, nous avons pu collecter des images annotées de Space Weather Live. Cela nous a permis de créer notre propre base de données avec des données annotées. Avec cet ensemble de données collectées, nous pouvons procéder à l'entraînement d'un modèle de classification basé sur un réseau neuronal convolutif. Le modèle entraîné sera alors capable de reconnaître et de classer les caractéristiques solaires telles que les taches solaires.

Nous avons commencé à collecter des données du 1er mai 2010 jusqu'en 31 mai 2023. Cette période a été choisie parce qu'elle coïncidait avec un cycle solaire élevé, au cours duquel un nombre important de taches solaires ont été observées. En nous concentrant sur cette période, nous avons réussi à télécharger un total de 6991 images de taches solaires, qui sont réparties en 3 classes : alpha, bêta et bêta-x. Dans notre approche, nous avons pris la décision de regrouper les autres classes de la classification de Mount Wilson en une seule classe appelée "bêta-x". Cette décision a été motivée par le fait que ces types de taches solaires apparaissent rarement dans notre jeu de données.

En regroupant ces classes moins fréquentes en une seule classe (bêta-x), nous simplifions la tache de classification en réduisant le nombre de classes distinctes à prédire. Cela permet

de mieux équilibrer la distribution des données et d'améliorer les performances de notre modèle de classification.

0.15.10 L'architecture globale de notre système

Cette partie représente la vue globale de notre solution où sont décrits en premier les différentes phases permettant une analyse des images magnétogrammes (Voir figure 14). puis l'acteur principal qui utilisera ce système pour analyser de nouvelles images.la figure 15 résume l'architecture globale utilisé pour notre approche.



FIGURE 14 – Image magnétogramme d'une tache solaire de type Beta



FIGURE 15 – Architecture globale de notre système de classification.

Notre approche de détection et classification des taches solaires inclut principalement deux phases : 1) Une phase de prétraitement des images et 2) une phase d'apprentissage par

un modèle RNC La phase de pré traitement prend en entrée des images de taches solaires SHARP 5 annotées sur lequelles un ensemble de traitement est effectué afin de retourner des images qui soient conformes à l'entrée du modèle de classification. La phase de d'apprentissage prend les images retournées de la phase de prétraitement pour apprendre les caractéristiques des taches solaires afin de les classer selon ces caractéristiques.

la phase de pré-traitement inclut 2 étapes : transformation du jeu de données, découpage du jeu de données.

0.15.10.1 Étape de transformation La phase de transformation des données est une phase cruciale de processus d'analyse et préparation des données, elle vise a préparer les données pour pouvoir les utiliser dans un modél d'apprentissage. conversion des images en niveaux de gris La conversion des images en niveau de gris est une étape importante pour simplifier l'analyse des images en éliminant les informations de couleur. L'utilisation d'images en niveaux de gris présente plusieurs avantages, notamment en termes de réduction de la taille des données, de réduction des coûts de stockage et de calcul, ainsi que de simplification des opérations de traitement d'images. La figure 16 est une représentation exemplaire d'une image en niveau de gris de la tache dans la figure 14. La figure 17 montre la représentation matricielle des images qui vont être les entrées de notre modèl.



FIGURE 16 – Image en niveaux de gris

^{5.} Solar HMI Active Regions Patches



FIGURE 17 – Conversion de l'image en matrice numpy

Redimensionnement Le redimensionnement des images est une étape de prétraitement courante dans l'apprentissage automatique lorsqu'on travaille avec des données d'images. Il s'agit de modifier les dimensions (largeur et hauteur) des images pour obtenir la taille souhaitée. Dans notre cas, nous avons choisit une dimension de (227, 227).

0.15.10.2 Étape de découpage La phase de découpage consiste à répartir le jeu de données, dans ce cas des images magnétogramme des taches solaire en trois groupes.Le premier groupe est utilisé pour l'entraînement du modèle de réseau neuronal convolutif (RNC), le deuxième groupe pour la validation des données et le troisième groupe pour réaliser des tests, et pour interpréter les résultats.Plusieurs itérations ont été effectuées pour la répartition des nombres d'images dans chaque groupe jusqu'à obtention d'une précision élevée. Dans notre cas pour une précision de 99.9%, sur un dataset de 6991 images, la répartition est de : 80% de la dataset utilisé pour l'entraînement, et 20% utilisé pour la validation.

0.15.11 Phase d'apprentissage basé sur un modèle RNC

Le modèle choisi pour la classification est un modèle de réseau neuronal convolutif (RNC) basé sur l'architecture pré-entraînée d'AlexNet. La Figure 18 et 19 présente l'architecture du modèle sélectionné.

Ce modèle est composé de 17 couches, dont 5 couches convolutives (Conv2). La première couche convolutive utilise un noyau (ou filtre) de taille 11x11, tandis que la deuxième



FIGURE 18 – Architecture RNC utilisée pour la classification des images magnétogrammes des tache solaire Fig1

couche convolutive utilise un noyau de taille 5x5. Les couches convolutives suivantes utilisent un noyau de taille 3x3. La première couche convolutive est configurée avec 96 filtres, la deuxième couche en a 256, tandis que les troisième et quatrième couches en comprennent 384, et la cinquième couche compte 256 filtres. Ces filtres permettent au modèle d'apprendre des caractéristiques spécifiques à partir des images d'entrée et de les utiliser pour la classification des taches solaires.

Après chaque couche convolutive, une couche de normalisation par lots (batch normalization) est ajoutée. Cela permet de normaliser les activations des neurones pour accélérer l'apprentissage et améliorer la stabilité du modèle. De plus, après chaque couche de normalisation par lots, à l'exception de la troisième et de la quatrième couche, une couche de maxpooling (maxpool) est appliquée avec un pool de taille 3. Cette opération réduit la dimension spatiale des caractéristiques extraites, tout en préservant les caractéristiques les plus importantes. Ainsi, ces couches de normalisation par lots et de maxpooling sont utilisées pour renforcer les performances du modèle en termes d'apprentissage et de représentation des informations des images de taches solaires.

Après les étapes de convolution, de normalisation par lots et de maxpooling, les caractéristiques extraites des images sont aplaties (Flatten) pour être converties en un vecteur unidimensionnel. Ce vecteur est ensuite transmis à une couche entièrement connectée (Dense) avec 4096 neurones. Cette couche permet de capturer des combinaisons linéaires des caractéristiques extraites et d'apprendre des représentations plus abstraites des données. Pour réduire le risque de surapprentissage, un taux de dropout de 50% est appliqué à cette couche, ce qui signifie que la moitié des neurones sont désactivés aléatoirement lors de l'entraînement pour éviter une dépendance excessive entre les neurones.

Enfin, la sortie de la couche Dense avec 4096 neurones est acheminée vers une dernière couche Dense de 3 neurones. Cette couche utilise la fonction d'activation Softmax, qui attribue des probabilités à chaque classe de la classification multi-classes (alpha, beta, betaX). La classe avec la plus grande probabilité est choisie comme prédiction finale du modèle pour chaque image de tache solaire.

0.15.12 Conclusion

Nous avons présenté dans ce chapitre les différentes étapes et traitements effectuées par lesquels passent une images de tache solaire avant de passer par un modèle de classification automatique, ainsi que la stratégie d'entraînement du modèle adoptée. Dans le chapitre qui suit, nous présenterons les résultats qui ont été obtenues sur les données de test et



FIGURE 19 – Architecture utilisée pour la classification des images magnétogrammes des tache solaire Fig2.

l'interface graphique.

Chapitre 5 : Implémentation et évaluation des résultats du modèle

0.15.13 Introduction

Ce chapitre présente l'implémentation des différentes étapes de conception du système de classification des taches solaire, en analysant des images magnétogrammes des ces dernieres. Nous allons décrire en détail les résultats d'apprentissage du modèle, les tests effectués sur de nouvelles données et enfin présenter les différents outils utilisées avec un aperçu de l'interface Web réalisée.

0.15.14 Les critères d'évaluation des résultats

Afin d'évaluer les performances d'un modèle de réseau de neurones, plusieurs mesures statistiques sont utilisées tel que la courbe Précision-Rappel (Precision-Recall) et le coefficient F-measure. Ces mesures sont essentiellement utilisées en apprentissage machine. Elles sont calculées à partir des résultats obtenus par le modèle sur un jeu de données test que le modèle n'a jamais vu. Dans un contexte de multi-classification, pour la sortie du réseau de neurones, nous pouvons évaluer les performances du modèle par les mesures statistiques suivantes :

0.15.14.1 Précision La précision est une métrique qui nous renseigne sur le pourcentage des prédictions positives correctes. Elle est calculée comme : le nombre d'instances correctement classées pour une classe donnée divisé par le nombre total d'instances prédites comme cette classe.

$$precision = \frac{TP}{TP + FP} \tag{3}$$

0.15.14.2 Rappel Le rappel est une métrique qui nous renseigne sur la qualité de tous les positifs réels, elle nous permet de savoir quel pourcentage d'échantillons positifs réels a été correctement classé. Il est calculé comme le nombre d'instances correctement classées pour une classe donnée divisé par le nombre total d'instances réelles de cette classe.

$$Rappel = \frac{TP}{TP + FN} \tag{4}$$

0.15.14.3 Matrice de confusion Il s'agit d'une représentation tabulaire montrant le nombre d'instances correctement et mal classées pour chaque classe prédite par le modèle. Les matrices de confusion fournissent un aperçu des performances du modèle pour chaque classe et identifient des erreurs de classification spécifiques.

0.15.14.4 F1 score La métrique F1 score ou F measure ou encore Precision-Recall est est une métrique qui montre le compromis entre les métrique de précision et rappel car elle prend en compte à la fois les faux positifs et les faux négatifs. Elle est calculée comme suit :

$$F1score = \frac{2 \times precision \times rappel}{precision + rappel}$$
(5)

0.15.15 L'évaluation du modèle de classification

0.15.15.1 Métrique de précision dans la phase d'apprentissage : Le graphe 20 représente la métrique 'précision', sur le jeu de validation et sur le jeu d'entraînement en fonction du nombre des époques. La précision représente ici le taux de des prédictions correctement faite par le modèle. Les époques sont le nombre d'itérations pour lesquels le modèle fait le tour du jeu d'entraînement.

La précision sur le jeu d'entraînement commence avec une valeur de 0.5 et commence à 0.32 pour le jeu de validation. Au fur et à mesure des époques, la précision sur les jeux d'entraînement et validation croit et décroît, car le modèle est en cours d'apprentissage, il est à la recherche d'un maximum local. A la fin de son apprentissage, le modèle atteint une précision de 99% sur le jeu d'entraînement et atteint une précision de 99% sur le jeu d'entraînement et atteint une précision de 99% sur le jeu d'entraînement et atteint une précision de 99% sur le jeu d'entraînement et atteint une précision de 99% sur le jeu d'entraînement et atteint une précision de 99% sur le jeu d'entraînement et atteint une précision de 99% sur le jeu d'entraînement et atteint une précision de 99% sur le jeu d'entraînement et atteint une précision de 99% sur le jeu d'entraînement et atteint une précision de 99% sur le jeu d'entraînement et atteint une précision de 99% sur le jeu d'entraînement et atteint une précision de 99% sur le jeu d'entraînement et atteint une précision de 99% sur le jeu d'entraînement et atteint une précision de 99% sur le jeu d'entraînement et atteint une précision de 99% sur le jeu d'entraînement et atteint une précision de 99% sur le jeu de validation ce qui reflète une bonne mémorisation des données biaisées.



FIGURE 20 – Graph de précision

0.15.15.2 Métrique de perte dans la phase d'apprentissage : Le graphe 21 représente la métrique perte sur le jeu de validation et sur le jeu d'entraînement en fonction du nombre des époques. La métrique perte nous renseigne sur les mauvaises prédictions faite par le modèle, et c'est la moyenne des différences au carré entre les valeurs vérité terrain et les valeurs et prédites. La perte sur le jeu d'entraînement commence avec une valeur 1 sur le jeu de validation et 2.5 sur le jeu d'entraînement. À la fin des époques, on remarque que la perte a été de 0.1 sur le jeu d'entraînement et a été de 0.1 sur le jeu de validation ce qui reflète beaucoup de prédictions correctes.



FIGURE 21 – loss graph

0.15.16 L'environnement de développement

Langage de programmation Python : langage de programmation haut niveau et polyvalent. Il est muni de plusieurs librairies d'algèbre linéaire, de traitement d'images, d'apprentissage automatique et d'apprentissage profond.

0.15.16.1 Partie backend

Kaggle : Kaggle est une plateforme gratuite proposée par Google qui permet aux utilisateurs d'écrire et d'exécuter du code Python directement depuis un navigateur web. Elle offre un accès à des ressources matérielles dans le cloud, telles que du stockage, de la mémoire vive, ainsi que des GPU et des TPU.

Ce qui distingue Kaggle, c'est sa vaste collection de bases de données disponibles, qui peuvent être explorées, utilisées et partagées entre les utilisateurs. De plus, la plateforme permet également de partager des scripts et des codes avec d'autres utilisateurs. Kaggle organise également des compétitions et des concours entre les spécialistes des données scientifiques et de l'apprentissage automatique, favorisant ainsi la culture de l'open source et l'échange de connaissances.



 Keras : est une librairie encapsulant Tensorflow. c'est une interface API Python qui facilite la manipulation et l'entraînement de différentes architectures de réseaux neuronaux.



— **Tensorflow :** C'est une bibliothèque, développée par Google dans le domaine de l'apprentissage profond, elle offre une approche open source pour créer des data-flows 6 .



— Selenium : Selenium est un framework open source utilisé pour automatiser les navigateurs web. Il permet d'interagir avec les éléments d'une page web, d'effectuer des actions telles que cliquer sur des boutons et remplir des formulaires, et d'extraire des données des pages web. Selenium prend en charge de nombreux langages de programmation, dont Python, ce qui le rend polyvalent et largement utilisé pour le web scraping.



— OpenCV : est une librairie libre qui offre une large gamme de fonctionnalités pour le traitement d'images et de vidéos. Elle est largement utilisée dans des domaines tels que la vision par ordinateur, la réalité augmentée, la robotique et bien d'autres applications nécessitant des opérations de traitement d'images efficaces et rapides.



0.15.16.2 Partie frontend

- Django Framework : C'est un framework web en Python qui offre la possibilité de développer rapidement des sites web sécurisés et faciles à maintenir. En plus de sa fonctionnalité principale en tant que framework web, Django est également utilisé pour le déploiement de modèles de réseaux neuronaux.
- Hyper Text Markup Language (HTML) : est le code utilisé pour structurer une page Web et son contenu.

^{6.} structures qui décrivent la façon dont les données se déplacent sur un graphique ou une



— Cascading Style Sheets (CSS) : est un langage de feuille de style utilisé pour décrire la présentation d'un document écrit en HTML. Il permet de décrire la façon dont les éléments doivent être affichés à l'écran.



série de noeuds de traitement ou un nœud représente une opération mathématique et et chaque connexion entre ces noeuds est une flèche de données multidimensionnelle (tensor)

— JavaScript : JavaScript est un langage de programmation utilisé pour créer des fonctionnalités interactives sur les pages web. Il permet d'ajouter des éléments dynamiques et complexes à une page. il est largement utilisé dans le développement web pour améliorer l'expérience utilisateur et rendre les sites web plus interactifs et réactifs.



0.15.17 La présentation de l'interface

Afin de faciliter l'interaction entre le modèle de classification et les professionnel de l'astronomie, nous avons réalisé une interface graphique Web en Python. Cette interface permet au astronomes de calculer le nombre de taches solaire (nombre de Wolf), classifier les taches solaire en 3 classes, et visualiser les résultats. La figure 22 représente la première fenêtre d'accueil, la figure 23 représente la page de résultat de calcul de nombre de Wolf, et la figure 24 représente l'interface de classification des taches solaire.

	Classification Page About Sunspoty How to use me?
Algorithme Drag & Drop to Upload File OR BROWSE FILE Choisir un fichier choisi Use Auron fichier choisi	Algorithme DBSCAN Sobel Kernel Must be an odd number sensitivity must be an even number DBSCAN Kernel type kernel size SUBMIT

FIGURE 22 – Page d'accueil 1

Détection et classification automatique des taches solaires par des méthodes de deep learning



FIGURE 23 – Page d'accueil 2



FIGURE 24 – Page de classification

0.15.18 L'évaluation du modèle de classification sur le jeu de test et discussion

0.15.18.1 Evaluation des résultats selon la métrique de précision : Nous avons obtenu un résultat de 0.66. On peut dire que notre modèle a donc correctement prédit 66% de notre jeu de test positif, Cela met en évidence que la métrique de précison seule n'est pas une bonne mesure, nous devons considérer aussi la valeur de rappel.

0.15.18.2 Évaluation des résultats selon la métrique de rappel : Nous avons obtenu un résultat de 0.55. On peut dire alors que notre modèle a su classifier la moitier ne notre test set. En revanche, cette mesure n'est pas suffisante pour dire que le modèle est parfait car le modèle peut faire parfois e fausse alerte. Par conséquent, le rappel indépendant n'est pas une bonne mesure.

0.15.18.3 Évaluation des résultats selon la métrique de F-score : Nous avons obtenu un résultat de 0.54. Ce score représente le compromis entre la valeur de précision et le rappel car on ne peut pas déterminer si un algorithme est meilleur qu'un autre. Le score F1 favorise les classificateurs qui ont une précision et un rappel similaires. Ainsi, le score F1 est une meilleure mesure à utiliser si on cherche a avoir un équilibre entre Précision et Rappel.

0.15.18.4 Evaluation des résultat selon la matrice de confusion : La figure 25 est une représentation schématique des classes prédites en comparaison avec les classes correctes.

Classe Alpha :

- Prédictions correctes : 65 images
- Erreurs de prédiction : Confusion entre Alpha et Beta

Classe Beta :

- Prédictions correctes : 68 images
- Erreurs de prédiction : Confusion entre Beta et Alpha

Classe BetaX :

- Prédictions correctes : Un faible nombre d'images 35% (en raison du manque de données)
- Erreurs de prédiction : Confusion avec la classe Beta

Cette analyse révèle que la classification des taches solaires de type Alpha et Beta est relativement bonne, avec une précision de 66% pour chaque classe. Cependant, la classe BetaX présente des difficultés de classification en raison du manque de données spécifiques à cette classe. Cela conduit à une confusion entre la classe Beta et la classe BetaX dans la plupart des prédictions. Pour améliorer la précision de la classification, il serait nécessaire de collecter davantage de données spécifiques à la classe BetaX.

0.15.19 Conclusion

Nous avons vu dans ce chapitre, l'application fonctionnelle réalisée depuis la conception proposée dans le chapitre 3 afin de résoudre la problématique de notre projet. Nous avons commencé par décrire les critères d'évaluation d'un classificateur, puis nous avons évalué les performances de notre modèle de classification en fonction de ces critères. Ensuite, nous avons présenté les différentes bibliothèques que nous avons utilisées tant au niveau



FIGURE 25 – Matrice de confusion sur le jeu de test

du back-end que du front-end pour la réalisation de notre système. Enfin, nous avons inclus des captures d'écran de l'interface utilisateur que nous avons développée. Nous avons clôturer ce chapitre par une interprétation de test réel de notre système de classification en comparant les résultat de notre modèle avec la classification des experts montré dans le site space weather live. Cette comparaison nous a permis de tirer des conclusions sur l'efficacité de notre modèle et de discuter des performances obtenues.

CONCLUSION GÉNÉRALE

A l'issue de ce projet, nous avons développé une application permettant la détection des taches solaire, le calcul de nombre de Wolf sur des images white light et la classifications de ces taches solaires sur des images magnétogrammes selon la classification mount wilson (alpha, beta, betaX).

L'application est destinée à assister les astrophysiciens et les astronomes à classifier les taches solaire et calculer le nombre de Wolf pour pouvoir tracer la courbe du cycle solaire. Pour réaliser cette application, nous avons d'abord étudié les solutions existantes utilisant les téchniques de l'apprentissage profond et le traitement d'images. Contrairement à ces solutions, qui se basent sur des méthodes ancienne avec des classifications de zurich. Nous avons exploité un réseau de neurones convolutifs pour implémenter notre système de classification et des techniques de traitement d'images pour la détection et le calcul de nombre de Wolf.

Dans la phase de pré-traitement, nous avons effectué 2 opérations : la transformation de jeu de données, puis son découpage en jeu d'entrainement et jeu de validation.

La transformation des images magnétogrammes implique tout d'abord la conversion des images RVB en niveaux de gris, ainsi que leur redimensionnement. En ce qui concerne le découpage du jeu de données, il a été divisé en trois parties : un jeu d'entraînement représentant 80%, un jeu de validation représentant 20%, et un jeu de test. Toutes ces images ont été collectées à partir du site web Space Weather Live en utilisant la méthode du web scraping. La collecte de ces images a nécessité beaucoup de temps, ainsi que les ressources matériel.

Suite à la phase de pré-traitement, nos données deviennent conformes aux entrées du modèle RNC 2D et seront utilisées pour l'apprentissage. Le modèle RNC 2D est considéré comme le plus adapté à une bonne exploitation des images magnétogrammes et à l'éxtraction de ses caractéristiques.

Les résultats de test ont montré une précision de 99.8%. Cependant, nous avons besoin d'évaluer notre solution sur un jeu de test plus important afin de valider ce système de classification.

Durant toutes les phases, de l'état de l'art jusqu'à l'implémentation, ce projet nous a permis de découvrir le domaine de l'apprentissage profond, mettre en pratique les connaissances acquises pour répondre à un éventuel besoin réel dont font face nos astronomes. Malgré le manque des données annotées et les difficultés rencontrées lors de la collecte des données, ainsi que la nouveauté de ce domaine de recherche et le Le délai accordé pour terminer ce projet, nous avons réussi à surmonter ces obstacles et poursuivre notre projet de recherche. Nous avons utilisé des téchniques d'extractions des données de site web (Harvesting ou le web scrapping) pour créer notre propre base de données et nous avons explorer des méthode d'apprentissages profond pour la classification de ce jeu de données. Bien que cela ait été un défit, nous sommes convaincu que nos résultats et nos découvertes contribueront à l'avancement de la classification des taches solaires et à notre compréhension du Soleil.

Références

- [1] Sun Solar Flares, Wind, Spectroheliograph, Coronal Temperature | Britannica. https://www.britannica.com/place/Sun/History-of-observation.
- [2] J.M. Vaquero. Historical sunspot observations : A review. Advances in Space Research, 40(7) :929–941, January 2007.
- [3] D Stenning, V Kashyap, T C M Lee, D A van Dyk, and C A Young. Morphological Image Analysis and Its Application to Sunspot Classi [U+FB01] cation.
- [4] Channabasava Chola and J V Biabl Benifa. Detection and classification of sunspots via deep convolutional neural network. *Global Transitions Proceedings*, 3(1):177–182, June 2022.
- [5] Centre de Recherche en Astronomie, Astrophysique et Géophysique (CRAAG). https://www.craag.dz/pesentation.php.
- [6] Olivier Esslinger. La surface du Soleil, les taches et le magnétisme. https: //astronomes.com/soleil-etoile/soleil-surface-tache-magnetisme/.
- [7] Progression du cycle solaire activité solaire. https://spaceweatherlive.com/ fr/activite-solaire/cycle-solaire.html.
- [8] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press, 2016. http://www.deeplearningbook.org.
- [9] B. L. Yoon. Artificial neural network technology. SIGSMALL/PC Notes, 15(3):3–16, August 1989.
- [10] Yann Lecun and Yoshua Bengio. Convolutional networks for images, speech, and time-series. MIT Press, 1995.
- [11] Afshine Amidi and Shervine Amidi. Convolutional neural networks cheatsheet. https://stanford.edu/~shervine/teaching/cs-230/ cheatsheet-convolutional-neural-networks. Accessed : 2021-04-30.
- [12] Trung Nguyen, Claire Willis, Derek Paddon, and Hung Son Nguyen. A hybrid system for learning sunspot recognition and classification. *Hybrid Information Technology*, *International Conference on*, 2:257–264, 11 2006.
- [13] Sreejith Padinhatteeri, Paul A. Higgins, D. Shaun Bloomfield, and Peter T. Gallagher. Automatic detection of magnetic delta in sunspot groups. *Solar Physics*, 291(1):41– 53, January 2016.
- [14] Yuanhui Fang, Yanmei Cui, and Xianzhi Ao. Deep learning for automatic recognition of magnetic type in sunspot groups. *Advances in Astronomy*, 2019 :Article ID 9196234, 10 pages, 2019.
- [15] SOHO Sunspots and Earth. https://soho.nascom.nasa.gov/data/synoptic/ sunspots_earth/.

- [16] University of Hawaii at Manoa. https://manoa.hawaii.edu/.
- [17] T. Colak and R. Qahwaji. Automated mcintosh-based classification of sunspot groups using mdi images. Solar Physics, 248(2) :277–296, April 2008.
- [18] Stanford Solar Observatories. http://soi.stanford.edu/.
- [19] Magnetogram-2 Kaggle Dataset. https://www.kaggle.com/datasets/ anfelsarabouachat/magnetogram-2.
- [20] Site web de la NASA. https://sdo.gsfc.nasa.gov/.
- [21] Royal Observatory of Belgium. http://www.sidc.be/.
- [22] 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 2nd International Conference on Knowledge Discovery and Data Mining, 1996.
- [23] NOAA Space Weather Prediction Center de la National Oceanic and Atmospheric Administration. https://www.noaa.gov/.
- [24] SpaceWeatherLive.com. https://www.spaceweatherlive.com/.