



Rapport Master M1

Master Calcul Haute Performance Simulation (CHPS)

# Reconnaissance des Plantes à l'aide d'un Réseau de Neurones Convolutif

Réalisé par: Aicha Maaoui, Abdeljalil, Mohamed, Farhat, Lydia

Encadré par: Prof. Hugo Bolloré

Janvier 2022

Institut des Sciences et Techniques des Yvelines (ISTY)

# Abstract

La reconnaissance des plantes avec un réseau de neurones, à partir des images données, n'est pas -toujours- une tâche évidente. Ceci est dû au fait qu' :

- Il y a une apparence diversifiée de structures complexes des plantes,
- Le problème de classification est réalisé, dans les majorités des cas, avec un nombre de classes assez élevé.

Le problème de classification des images de plantes est ainsi un problème bien posé, où les classe sont délimitées. Nous avons besoin dans ce cas d'informations sur les l'images pour représenter et identifier l'objet qu'elle contient.

L'objectif de ce projet consiste à la reconnaissance des images de plantes grâce à un réseau de neurones convolutif: On utilise l'algorithme **CNN (Convolutional Neural Network)**.

Initialement, on dispose d'un jeu de données de 3655 images de fleurs, formants 5 classes différentes. Ceci nous permet de faire l'apprentissage d'un classifieur et générer à la fin un model capable de classifier nos images.

Afin de réaliser ce projet, nous avons procédé comme suit:

- Comprendre le fonctionnement des réseaux de neurones en informatique (IA), ainsi que la base de fonctionnement d' un algorithme CNN,
- Travailler sur l'exploration des données afin de déterminer les couches à utiliser dans notre réseau de neurones,
- Etat de l'art des algorithmes de traitement d'image existants dans le domaine de la reconnaissance de plantes,
- Pogrammation en utilisant le langage de programmation `c++` pour le classifieur, ainsi

que le langage *Python* pour la structuration des jeux de données.

Les étapes du projet sont structurés à l'aide de *github* dans le dépôt crée avec des *issues*, qui indiquent l'état d'avancement du projet comme illustré dans la figure (1).

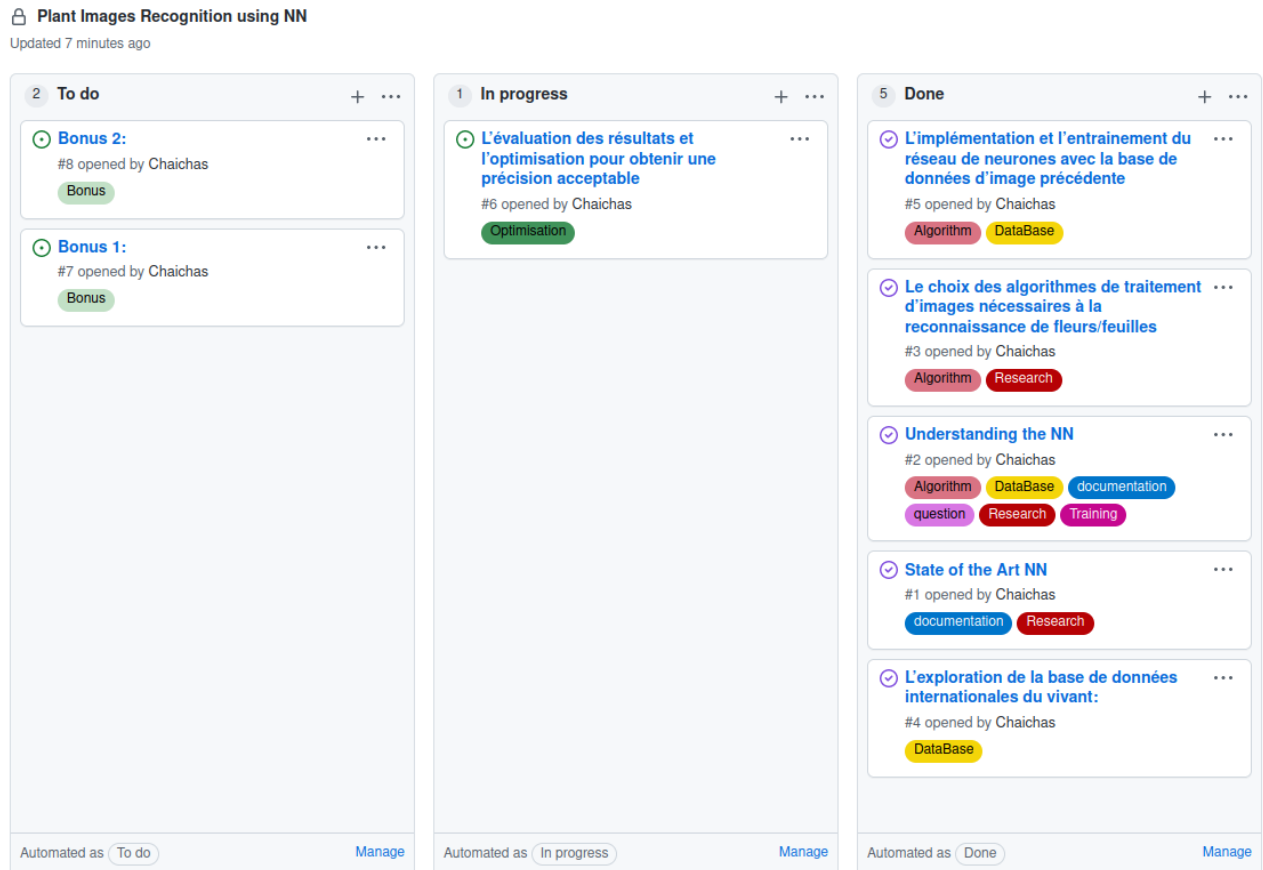


Figure 1: Description de l'état d'avancement du Project CNN.

# Contents

<b>1</b>	<b>Implémentation du CNN</b>	<b>1</b>
1.1	Introduction . . . . .	1
1.2	Description du Réseau de Neurones du Projet . . . . .	3
1.2.1	Description de la base de données . . . . .	3
1.2.2	Description des classes . . . . .	5
	<b>References</b>	<b>6</b>
<b>A</b>	<b>Appendix Chapter</b>	<b>7</b>

# List of Tables

# Listings

# Chapter 1

## Implémentation du CNN

### 1.1 Introduction

Un réseau de neurones est une interprétation machine du cerveau humain, contenant des millions de neurones. Ces derniers transmettent des informations sous forme d'impulsions électriques, *"Deep Neural Network: Qu'est-ce qu'un réseau de neurones profond ?"* 2021.

Ils sont utilisés pour résoudre aux problèmes complexes qui nécessitent des calculs analytiques similaires à ceux du cerveau humain, *"Les Réseaux de Neurones artificiels"* 2018.

On peut citer comme applications les plus courantes des réseaux de neurones, *"Les Réseaux de Neurones artificiels"* 2018:

- **La classification:** C'est la distribution des données par paramètres. Par exemple, on dispose d'un ensemble de personnes à l'entrée. Il faut décider à laquelle d'entre elles on accorde un prêt. Ce processus peut être effectué par un réseau de neurones, en analysant des informations telles que l'âge, la solvabilité, Les antécédents de crédit, etc,
- **La prédiction:** C'est la capacité de prédire la prochaine étape. Par exemple, la hausse ou la baisse d'une action en fonction de la situation du marché boursier,
- **Reconnaissance:** Elle représente actuellement l'utilisation la plus répandue des réseaux de neurones. On cite comme titre d'exemple "Google" lorsqu'on cherche une photo ou dans les appareils photo des téléphones lors de la détection de la position du visage, etc.

### Fonctionnement d'un réseau de neurones:

Un neurone est une unité de calcul qui reçoit des informations, effectue des calculs simples pour les transmettre plus loin. Ils sont divisés en trois types principaux, comme illustré dans la figure (1.1), "Réseaux de Neurones" 2013:

- Entrée ( $X_i$ ), avec les paramètres  $W_i$  (poids) et  $b_j$  (bias),
- Caché ( $F$ ).
- Sortie ( $Y_i$ ).

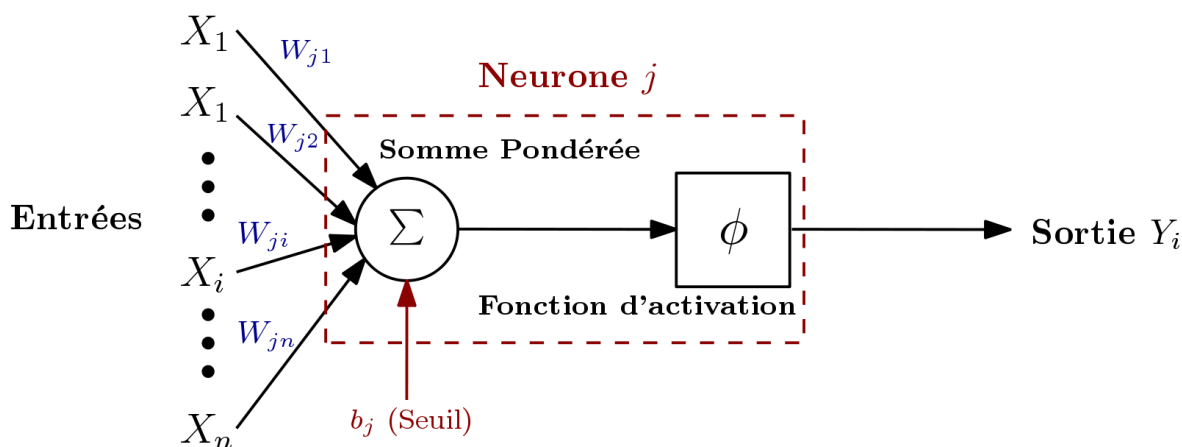


Figure 1.1: Éléments d'un réseau de neurones.

Dans le cas où un réseau de neurones est constitué d'un grand nombre de neurones, on introduit le terme *couche*. Par conséquent, il existe:

- Une couche d'entrée qui reçoit les informations,
- $n$  couches cachées (généralement un max de 3) qui les traitent,
- une couche de sortie qui génère le résultat.

Chacun des neurones a 2 paramètres principaux: (i) les données d'entrée et (ii) les données de sortie.

Lors de l'initialisation du réseau de neurones, les poids  $W_i$  sont attribués d'une manière aléatoire.

La fonction d'activation est un moyen de normaliser les données d'entrée, "Fonction d'activation" 2021. Il existe beaucoup de fonctions d'activation. Cependant, on considère les fonctions basiques : *Linéaire*, *Softmax* et *Sigmoïde (Logistique)*. La principale différence entre ces derniers consiste à la plage de valeurs fournis à la sortie.



## 1.2 Description du Réseau de Neurones du Projet

### 1.2.1 Description de la base de données

Dans cette partie, on se propose de faire l'analyse du travail réalisé sur la partie *Dataset* (Base de données).

#### Collecte des données

Dans un premier temps, on a commencé la collecte des données depuis "*Plantae*" 2020. Cependant, on a rencontré des difficultés, se représentant en:

- Nombres insuffisants d'image pour chaque classe de plantes,
- Complexité des images (beaucoup de détails).

Ce qui ne permet pas de générer un model parfait de reconnaissance d'image. Pour cela, on a considéré les différents jeux de données partagés par des éditeurs dans "*PlantVillage Dataset*" 2020, plus facile à manipuler.

#### Exploration des données

Notre jeux de données se compose des images de fleurs de 5 classes, nommées comme suit:

- Daisy,
- Dandelion,
- Rose,
- Sunflower,
- Tulip.

Pour les images de classes choisies pour les entraîner, il existe des informations perturbatrices, i.e., il y a quelques images qui présentes des insectes perturbants l'unicité des fleurs et ces photos ne représentent qu'une partie du jeux donnés, comme illustré dans la figure (1.2).



Figure 1.2: Présence d'un insecte, qui perturbe l'unicité de la fleur.

Sinon, le jeu de données choisit contient des images adéquates pour générer un model de classification d'image.

### Modifications apportées

Dans le jeux de données collecté depuis le site Kaggle, on trouve l'existence de classes volumineuses, i.e., 984 le maximum d'images de la classe Tulip, et 733 pour le classe de Sunflower.

De plus, les photos ont des tailles très variés. La majorité ont comme dimensions  $200 \times 200$ .

Il était donc important d'ajuster la Dataset utilisée, afin d'avoir un standard avec lequel on peut travailler et générer un model de meilleur performance.

Dans un premier temps, on a redimensionné toutes les photos des classes à une dimension de  $50 \times 50$ , car dans ce cas on aura:

- une conservation de la bonne visualisation (qualité) des images, lues sous forme de matrices carrées réduites,
- Réduction du temps de calcul.

Dans un second temps, on a défini le même nombre d'images associées à chaque classe, pour avoir un jeu de données uniforme.

Pour se faire, on a généré le fichier *resize.py* (en python), qui permet de modifier le jeu de données utilisé pour générer le model.

### 1.2.2 Description des classes

# References

*"Deep Neural Network: Qu'est-ce qu'un réseau de neurones profond ?"* (2021). Data Scientist. URL: <https://datascientest.com/deep-neural-network>.

*"Fonction d'activation"* (2021). Inside Machine Learning. URL: <https://inside-machinelearning.com/fonction-dactivation-comment-ca-marche-une-explication-simple/>.

*"Les Réseaux de Neurones artificiels"* (2018). Juri' Predis. URL: <https://www.juripredis.com/fr/blog/id-19-demystifier-le-machine-learning-partie-2-les-reseaux-de-neurones-artificiels>.

*"Plantae"* (2020). gbif. URL: <https://www.gbif.org/species/6>.

*"PlantVillage Dataset"* (2020). Kaggle. URL: <https://www.kaggle.com/abdallahalidev/plantvillage-dataset>.

*"Réseaux de Neurones"* (2013). Statistic. URL: <http://www.statsoft.fr/concepts-statistiques/reseaux-de-neurones-automatisees/reseaux-de-neurones-automatisees.htm#.YdV2udso9H>.

## Appendix A

# Appendix Chapter

Le Projet est déposé dans le dépôt git "**Reconnaissance-de-Plantes-avec-NN**". Le code SSH de ce dépôt est le suivant:

**git@github.com:Chaichas/Reconnaissance-de-Plantes-avec-NN.git**