



Projet pyStill

Plant Expert at your finger

Loïc AXILAI
Noël MAURICE
Mathis RALAIVAO

- Présentation du projet
- Exploration des données
- Classification
- Segmentation
- Bounding Box
- Bilan et conclusion
- Démo Streamlit



Présentation du projet

L'objectif de ce projet est de localiser et classifier l'espèce d'une plante dans une image.

Une fois la classification faite, le système doit retourner à l'utilisateur une description de celle-ci et identifier les possibles maladies.

L'application est capable à partir d'une image prise par l'appareil photo de donner une succession d'informations à l'utilisateur.

- Ce projet est en premier lieu un problème de classification avec reconnaissance d'images.
- Nous avons complété l'application avec un modèle de segmentation de la feuille (détourage) dans l'image ainsi qu'une « bounding box ».
- Le système collecte des informations depuis un site spécialisé sur les maladies des plantes et les restitue à l'utilisateur après identification.
- Les réalisations du projet sont illustrés à travers un site développé avec Streamlit et font l'objet d'une démonstration à la fin de la présentation.

Exploration des données

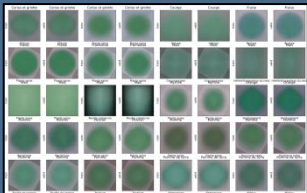
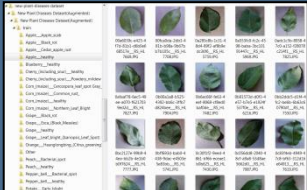
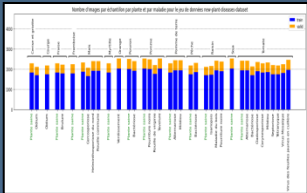
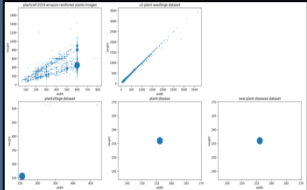
Etude des différentes sources de données disponibles (COCO, Google, Kaggle)

- Découverte des jeux de données (nombre de fichiers, type de données, nature des informations, ...)
- Analyse de la distribution des tailles d'image
- Analyse de la distribution des images par classe
- Structuration des fichiers (organisation selon le pattern Keras)
- Image moyenne par classe

Jeux de données retenus pour le Projet

- Kaggle new-plant-diseases-dataset pour le modèle de classification
- Google Open Image Dataset pour l'ajout d'une classe hors périmètre
- Kaggle plantvillage-dataset pour le modèle de segmentation

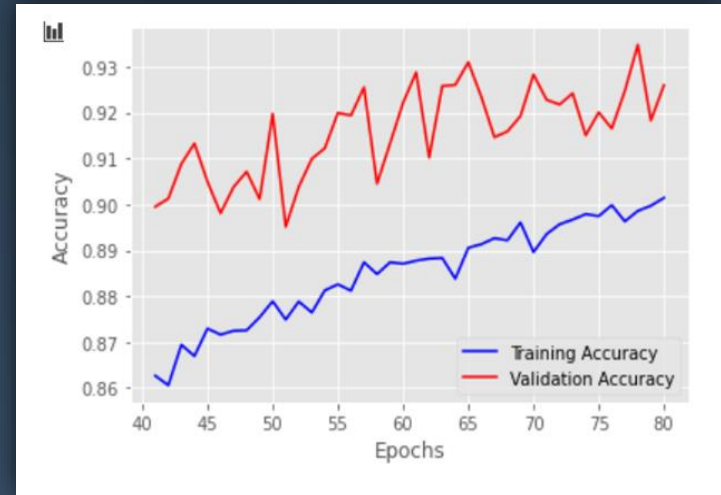
| dataset | ext | size |
|--|-----|--------|
| new-plant-diseases-dataset | JPG | 87988 |
| plant-disease | JPG | 186688 |
| plant-disease | PNG | 2 |
| plantleaf-2019-amazon-rainforest-plants-images | CSV | 1 |
| plantleaf-2019-amazon-rainforest-plants-images | JPG | 3142 |
| plantleaf-2019-amazon-rainforest-plants-images | XML | 3133 |
| plantvillage-dataset | JPG | 162914 |
| plantvillage-dataset | PNG | 2 |
| v2-plant-seedlings-dataset | PNG | 11078 |



Classification

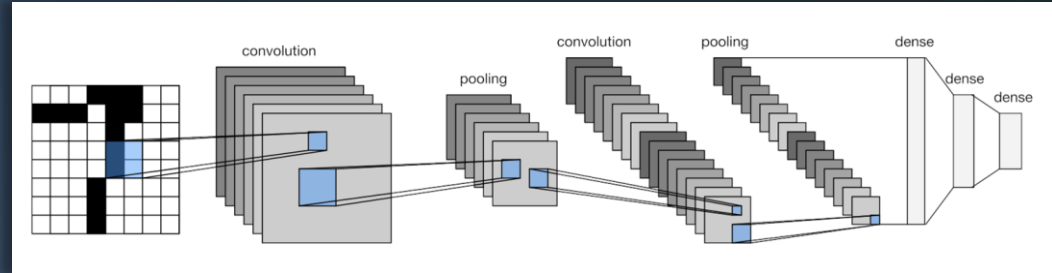
Convolutional Neural Network ou « CNN »

- Mode exploratoire
- Baseline du projet
- Premières conclusions
 - Apprentissage long
 - Ressources systèmes très sollicitées
 - Performance améliorable : > 80%
 - Surapprentissage

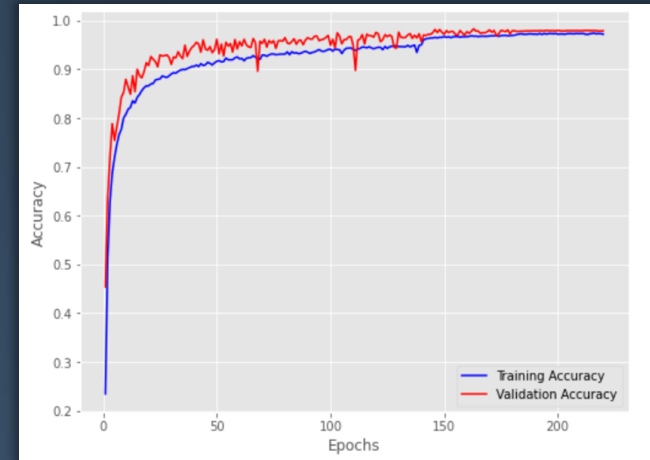


Classification

Architecture « LeNet »



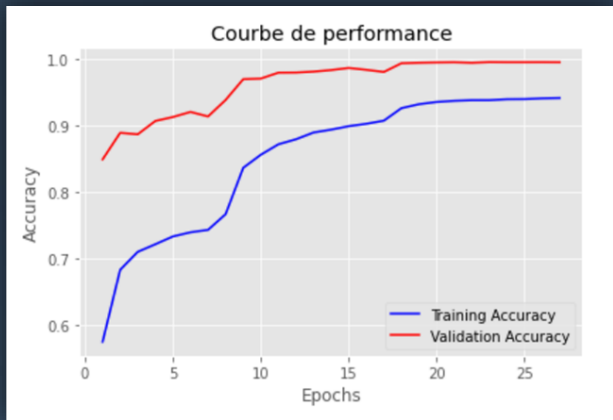
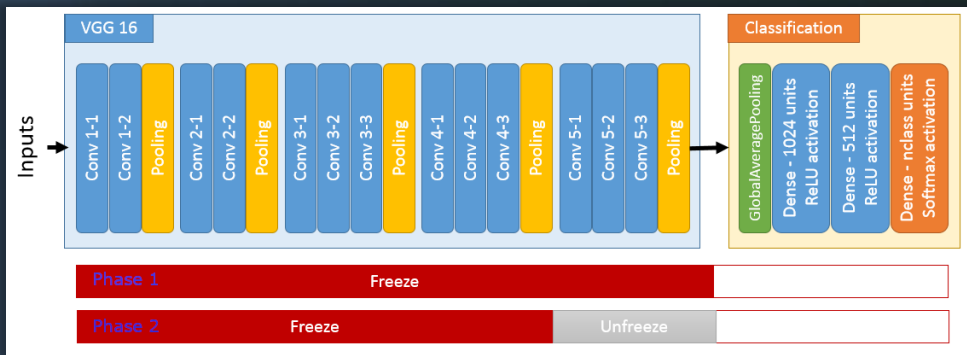
- Modèle plus adapté au traitement des images du projet
- Ressources systèmes insuffisantes : portage sur Google Colab
- Suppression du surapprentissage du modèle
- Optimisation de l'apprentissage
- Meilleurs scores que modèle précédent : > 97%



Classification

Architecture « VGG16 » Transfer Learning

- Mode exploratoire
- Entrainement en 2 phases
- Entrainement bien plus rapide
- Images de plus grandes dimensions
- Très bonnes performances > 99%





Classification

Amélioration des prédictions

- Eviter détection images hors périmètre
- Ajout d'une classe supplémentaire
- Entrainement
- Prédictions correctes

```
Fichier :          TomatoYellowCurlVirus4.JPG
Classe prédite :   Tomato__Tomato_Yellow_Leaf_Curl_Virus
Indice de confiance : 100.0%

Fichier :          TomatoYellowCurlVirus5.JPG
Classe prédite :   Tomato__Tomato_Yellow_Leaf_Curl_Virus
Indice de confiance : 100.0%

Fichier :          TomatoYellowCurlVirus6.JPG
Classe prédite :   Tomato__Tomato_Yellow_Leaf_Curl_Virus
Indice de confiance : 100.0%

Fichier :          _chat.jpg
Classe prédite :   Other
Indice de confiance : 100.0%

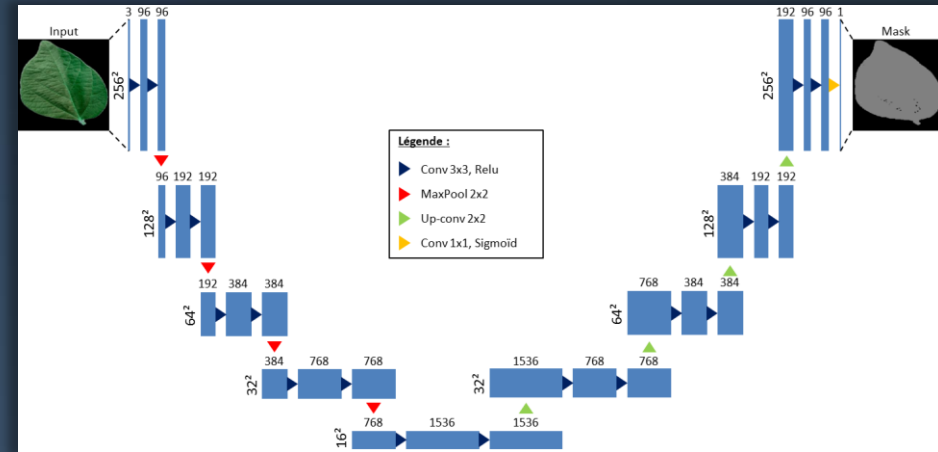
Fichier :          _visage_femme.jpg
Classe prédite :   Other
Indice de confiance : 100.0%
```


Segmentation

Modèle de type « U-Net »

- Objectif : détecter la feuille dans la photo et séparer de l'arrière plan
- Instanciation d'un modèle avec l'architecture ci-contre
- Entrainement du modèle avec fonction de perte coefficient de Dice :

$$S = \frac{2 |X \cap Y|}{|X| + |Y|}$$



Segmentation

Modification jeu de données

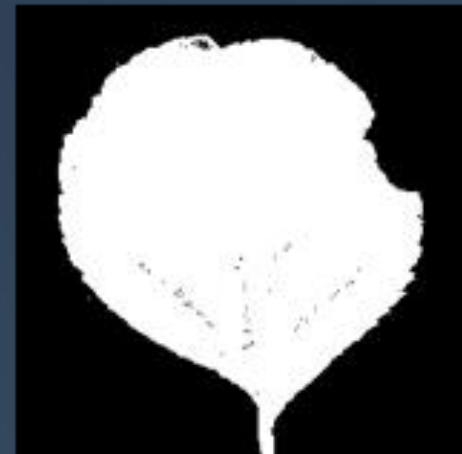
Image d'origine



Image segmentée



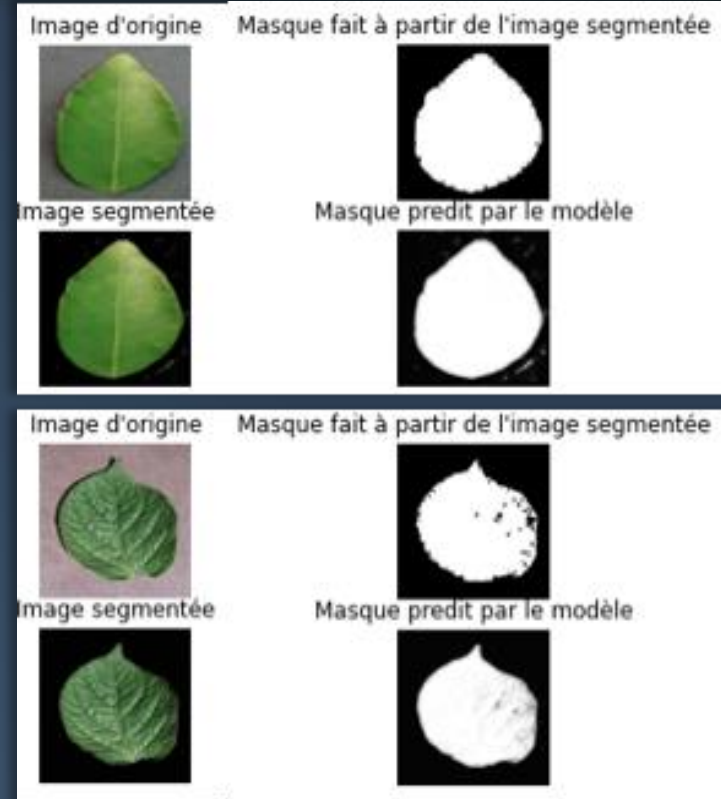
Masque généré



Segmentation

Efficacité du modèle

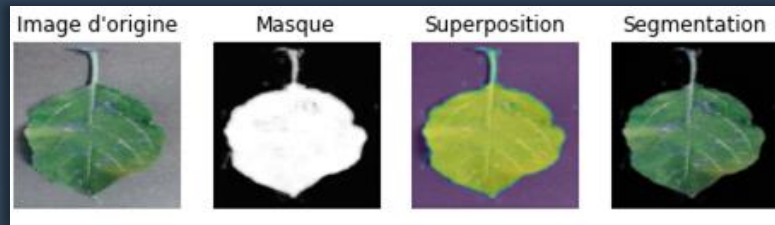
- Visuellement de bonnes prédictions
- Une précision de 95%
- Segmentation réalisable avec les masques prédits



Bounding box

Génération de « Bounding box »

- Extraction de la composante graphique principale de l'image :
 - « Bounding box » à partir des masques générés par le modèle type U-Net
 - Utilisation de la détection de contours avec le module OpenCV
- Limite de la génération : qualité du modèle de segmentation





Bilan & Résultats

Bilan

- Scientifique
- Technique et professionnel
- Humain

Résultats

- Identification des plantes saines et malades
- Informations complémentaires à l'utilisateur
- Extraction de la composante principale des images
- Valorisation des résultats
 - Rapport Technique d'Evaluation
 - Interface graphique
 - Github

A close-up photograph of a single green leaf with water droplets on its surface, resting on a dark, reflective surface. The leaf is positioned on the left side of the slide, with its reflection visible below it. The background is dark and out of focus.

Améliorations & Conclusion

Axes d'améliorations

- **Modèle de classification avec plus de robustesse**
Indice de confiance plus significatif pour les plantes hors périmètre
- **Enrichissement du domaine de prédictions**
Augmenter le nombres de classes prédites
- **Offre de services**
Proposer traitement curatif ou préventif, coût traitement à l'hectare, nombre de passages, ressources nécessaires...
Producteurs peuvent vérifier rentabilité économique du traitement et mesurer impact écologique.

Conclusion

Site Streamlit Démonstration




Projet pyStill

Menu

- ☒ Présentation
- ☐ Jeux de données
- ☐ Exploration
- ☐ PCA
- ☐ Méthodologie
- ☐ Modélisation
- ☐ Segmentation (démon)
- ☐ Identification (démon)
- ☐ Conclusion

Promotion Data Scientist
Bootcamp Octobre 2020

Participants :

 AXILAIIS Loïc
 MAURICE Noël
 RALAIVAO Mathis

Plant Expert at your finger

Présentation



La production végétale est à la base de toute la chaîne d'alimentation des humains et des animaux. La reconnaissance des maladies sur les plantes est donc un enjeu majeur. C'est la première étape avant de pouvoir lancer un traitement sanitaire sur une culture.

Le projet **pyStill** propose de détecter d'éventuelles maladies sur les plantes. Nous nous sommes concentrés sur les plantes suivantes : Cerise et griotte, Courge, Fraise, Framboise, Maïs, Myrtille, Orange, Pêche, Poivron, Pomme, Pomme de terre, Raisin, Soja, Tomate.

Le principe est simple, il suffit de fournir la photographie d'une feuille de la plante et le système indique si la plante est saine. Dans le cas contraire, le système indique la maladie détectée.

Lorsqu'une maladie est détectée, le système propose quelques informations et conseils pour le traitement.

Ce projet s'appuie sur des algorithmes de Deep Learning entraînés sur quelques dizaines de milliers d'images.