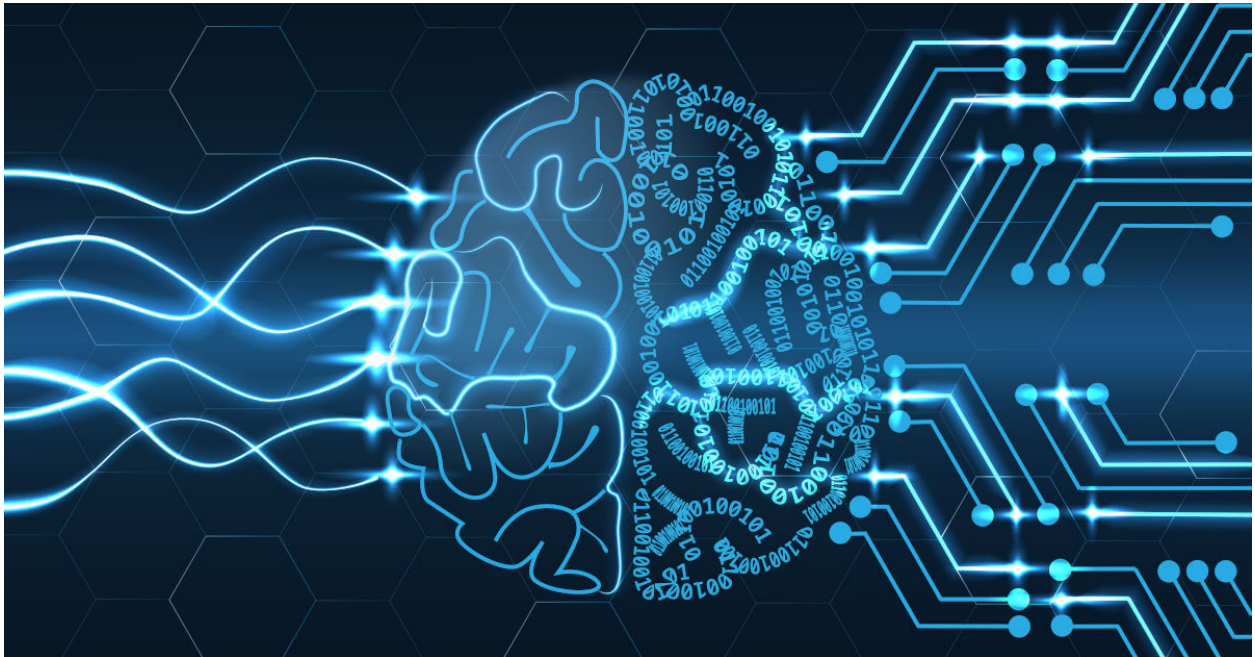


Rapport Projet Deep Learning : Airplanes Image Classification



Faical Toubali Hadaoui
Mehdi Sensali
Ignacio Lucas Oros Campo
Thibaud Merieux
Cédric Martin

20 Mars 2021

Contents

1	Description du sujet	3
2	Méthodologie suivie	4
2.1	Acquérir les données	4
2.2	Annoter les données	5
2.3	Partionnement des images pour l'entraînement, les tests et la validation	5
3	Prognostic pour le reste du projet	5
4	Script de chargement des données	5
5	Liens	5

1 Description du sujet

Dans le cadre de ce projet nous nous proposons de réaliser un classifieur d'images d'avions. Dans une première version nous allons séparer les avions commerciaux et les avions militaires et puis dans une seconde version nous allons distinguer entre 10 différentes classes (5 pour chaque catégorie).

Les modèles d'avion qui constitueront les classes seront les suivants :

Avions militaires



Figure 1: Dassault Rafale



Figure 2: Lockheed Martin F-35 Lightning II



Figure 3: Eurofighter Typhoon



Figure 4: A400M Atlas Airbus



Figure 5: Soukhoi Su-57

Avions commerciaux



Figure 6: Airbus A380



Figure 7: Airbus Beluga



Figure 8: Boeing 747



Figure 9: Boeing 777



Figure 10: Dassault Falcon 8X

2 Méthodologie suivie

2.1 Acquérir les données

Pour acquérir les données nous avons recherché successivement dans le moteur de recherche les différents modèles d'avion constituant les classes de notre sujet. Ensuite nous avons extrait tous les URL chargés

de google images sous la forme d'un fichier texte grâce à un code javascript introduit sur la console du navigateur. Ce fichier texte contenant les URL est à son tour passé en paramètre à un script Python qui télécharge chaque photo dont l'URL est listé et les garde toutes dans un dossier. Finalement il faut aller dans le dossier des images téléchargées et faire un filtrage des images qui sont valides à la main ainsi que supprimer le maximum de doublons possibles manuellement.

2.2 Annoter les données

Nous avons décidé d'utiliser un script écrit en python pour annoter les images téléchargées. En effet même si avec une base de donnée de 4000 images ce travail pourrait être fait à la main, un tel travail serait tédieux et long à faire donc dans l'idée d'optimiser le processus. L'utilisation d'un script est l'option la plus rapide. Une fois l'annotation réalisée, nous vérifions rapidement que le label associé à chaque photo est bien cohérent.

2.3 Partitionnement des images pour l'entraînement, les tests et la validation

La proportion choisie pour partitionner le dataset est la suivante : 70% des images seront dédiées à l'entraînement, 15% des images seront dédiées à la verification et 15 % au test.

3 Prognostic pour le reste du projet

Puisque les principaux traits des avions sont assez similaires, un Boeing 747 étant proche d'un Boeing 777, le problème n'est pas si facile à résoudre. Le réseau devra comprendre quels sont les différences entre ces avions similaires, qui ne seront même pas pris sous le même angle. Cependant, différencier un avion commercial d'un avion militaire est assez facile à l'oeil, alors le réseau de neurones devrait réussir sans accroc cette étape.

4 Script de chargement des données

Pour charger les images que notre base de données contient, nous avons utilisé le script Notebook déposé sur Moodle en l'adaptant à notre base de données d'images **Dataset** qui contient des sous-dossiers correspondant aux types d'avions (**Commercial** / **Militaire**) . Chaque sous-dossier comporte lui meme les modèles d'avions. Par exemple le sous-dossier **Militaire** contient les sous-dossiers correspondant aux différents modèles d'avions :

- Dassault Rafale
- Lockheed Martin F-35 Lightning II
- Eurofighter Typhoon
- A400M d'Airbus
- Soukhoi Su-57

Pour héberger notre base de données nous avons choisi GitHub, ainsi un Git clone sera très rapide depuis Google Colab.

5 Liens

Lien vers les images de la base de données sur github

https://github.com/faicaltoubali/Airplanes_Image_Classification