

# PROJECT 7 : Développez une preuve de concept



## Plan de travail prévisionnel



Étudiant : Zeruk Viktoriya

Mentor : Louis Willems



## PROBLEMATIQUE:

### *Contexte*

Lors d'un précédent projet (Project 6 : Classez des images à l'aide d'algorithmes de Deep Learning), nous avons utilisé le réseau de neurones convolutif.

Le meilleur modèle du projet précédent : Xception

### *Les données*

Stanford Dogs Dataset

### *Mission*

Tester une nouvelle méthode ViT (Vision Transformer) concurrente au méthode CNN.

L'idée est donc de comparer, en termes de précision et temps de calcul, un Xception et un ViT.

### *Ressources de calcul*

L'entraînement (même partiel) d'un réseau de neurones convolutionnels est très gourmand en ressources. Solutions :

- Limiter le jeu de données, en ne sélectionnant que quelques classes (races de chiens), ce qui permettra déjà de tester la démarche et la conception des modèles, avant une éventuelle généralisation.
- Utiliser la carte graphique de l'ordinateur en tant que GPU (l'installation est un peu fastidieuse, et l'ordinateur est inutilisable le temps du calcul).

## MÉTHODOLOGIE:

- Nous travaillerons avec le data set Stanford Dogs Dataset (Stanford Dogs dataset for FineGrained Visual Categorization)
- Nous utiliserons le modèle Xception comme méthode de référence
- Nous entraînerons le modèle ViT-B/16 en utilisant une méthode de fine-tuning
- La comparaison des modèles se fera à l'aide du temps de calcul et du score de précision (Accuracy)

## SOURCES BIBLIOGRAPHIQUES:

### ARTICLE DE RECHERCHE :

2010.11929v2.pdf (arxiv.org)

<https://openreview.net/pdf?id=LtKcMgGOeLt>

### ARTICLE DE VULGARISATION :

Vision Transformers (ViT) in Image Recognition - 2022 Guide - viso.ai

### CODE SOURCE :

[google-research/vision\\_transformer](https://github.com/google-research/vision_transformer) (github.com)

### TUTORIEL :

[Fine-Tune ViT for Image Classification with Transformers](https://huggingface.co/docs/transformers/training) (huggingface.co)

[Complete Guide to T2T-ViT: Training Vision Transformers Efficiently with Minimal Data](#)