

Développez une preuve de concept

Projet 7 - Parcours Ingénieur Machine Learning

Plan de travail prévisionnel

Thématique choisie & sources bibliographiques :

L'objectif de ce projet étant la réalisation d'un travail de veille technologique, on a ici choisi de s'intéresser à un type d'algorithme relativement innovant pour les applications de vision par ordinateur : les "Vision Transformers".

Les Vision Transformers, ou "ViT" ont été illustrés dans un premier temps par une équipe de recherche de Google en 2021 à la conférence International Conference on Learning Representations¹ et ont depuis fait l'objet d'un certain nombre d'études récentes sur lesquelles nous pourrions également nous appuyer.²³

Exploration technique envisagée :

Dans un premier temps, il s'agira de bien comprendre le principe de ce type de modèles et notamment des mécanismes d'attention mis en œuvre et de leur application sur des images.

Dans un second temps, on analysera les avantages et inconvénients attendus sur ce type de modèles par rapport à l'état de l'art et en particulier des modèles de réseaux de neurones exploitant le principe de la convolution, qui n'est pas mis en œuvre dans les Vision Transformers.

Les expérimentations envisagées sont les suivantes :

- Essai d'un modèle pré-entraîné sur le dataset du Projet 6 du parcours Ingénieur Machine Learning "Classifiez des images à l'aide d'algorithmes de Deep Learning", sur une tâche de classification de races de chiens.
- Comparaison des performances obtenues et des capacités d'entraînement de ces 2 modèles ViT par rapport à une baseline performante créée et entraînée lors du projet 6, un modèle de type réseau de neurones convolutif : ResNet-50.

¹ Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby. *An image is worth 16x16 words: transformers for image recognition at scale*. In ICLR 2021.

² Maithra Raghu, Thomas Unterthiner, Simon Kornblith, Chiyuan Zhang, Alexey Dosovitskiy. *Do Vision Transformers See Like Convolutional Neural Networks ?* In arXiv:2108.08810

³ Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, *Attention Is All You Need*, In arXiv:1706.03762