

Apprentissage de représentations d'images multimodales : texte / image via CLIP

Dans le cadre de ce TP, nous partirons et nous nous inspirerons du projet suivant :

[https://github.com/ckurtz/teaching_UPC_UFRMI/blob/main/
CLIP_tutorial_\(with_medical_images_ROCO\).ipynb](https://github.com/ckurtz/teaching_UPC_UFRMI/blob/main/CLIP_tutorial_(with_medical_images_ROCO).ipynb)

Nous utiliserons les données fournies à l'url suivante :

[https://drive.google.com/drive/folders/
1Nr3S82aY4R8tH7M2tps6cPYFAB9mP_A6?usp=share_link](https://drive.google.com/drive/folders/1Nr3S82aY4R8tH7M2tps6cPYFAB9mP_A6?usp=share_link)



Pour accéder aux données depuis votre notebook (via Google colab), il vous faudra posséder le dossier *data* à la racine de votre répertoire Google Drive.

Plutôt que de copier les données (très long), vous pouvez normalement créer un raccourci / shortcut depuis votre drive pour que ce dossier apparaisse comme « Partagé avec moi ».

Consignes :

1. Commencer par vous familiariser avec le code source proposé puis exécuter chaque portion de manière à observer les résultats offerts pour chaque démo
2. Quels sont les différents modèles de CLIP disponibles ? Proposez une évaluation quantitative et une confrontation de ces différents modèles
3. Montrer les résultats de CLIP dans le domaine médical sur d'autres exemples du jeu de données
4. Calculer les performances de CLIP sur une tâche de classification d'images naturelles, sur un autre jeu de données que CIFAR100

Modalités / évaluation :

- Travail individuel - Déposer sur Moodle un compte rendu de TP (2 à 3 pages), incluant un lien vers vos codes sources