Construction de la base de données d’images

Nous utilisons 2 façons différentes pour monter notre database :

1. En utilisant OidV4 pour récupérer des ensembles d’images de OpenimageV6 si les classes sont répertoriées
2. En construisant une base de données personnalisée pour les classes manquantes ou pour ajouter plus d’images.

**Utiliser OidV4 pour récupérer des ensembles d’images de OpenimageV6.**

Vérifier si la classe existe : <https://storage.googleapis.com/openimages/web/visualizer/index.html?set=train&type=detection&c=%2Fm%2F03hlz0c>

Et changer Type -> detection

Récupérer les images et leurs labels

1. Cloner OidV4 : https://github.com/Kozehh/OIDv4\_YOLOv5.git
2. Se placer dans le dossier OIDv4\_ToolKit et exécuter la comande : pip install -r requirements.txt
3. \*\* Conseil -> Rouler la commande qui suit dans un terminal Anaconda pour éviter des problèmes \*\*  
   Télécharger les images et labels avec une commande du genre:   
   python main.py downloader --classes Classe\_1 Classe\_2 Classe\_k --type\_csv train --limit 1000 --n\_threads 5 --multiclasses 1 –y
   1. Si le nom de la classe contient un espace, remplacer l’espace par « \_ ». (ex : Frying\_pan)
   2. Limit : nombre maximal d’images à télécharger par classe
   3. n\_threads peut être ajusté en fonction des ressources de l’ordinateur
4. S’il y a des problèmes spécifiant des fichiers csv, faire l’étape 4.2.
   1. Télécharger les fichiers csv des « Boxes » pour Train, Validation et Test :  
      <https://storage.googleapis.com/openimages/web/download_v4.html>  
      et les mettre dans OIDv4\_ToolKit/OID/csv\_folder/
5. Refaire la commande, mais en changeant --type\_csv train par --type\_csv validation et réduire le nombre d’image (--limit) à environ 70% du nombre d’images de « train ».
6. Refaire la commande, mais en changeant --type\_csv validation par --type\_csv test et réduire le nombre d’image (--limit) à environ 10% du nombre d’images de « train ».

Transformer les labels en format YOLO

1. Créer/modifier classes.txt et ajouter nom des classes (1 classe par ligne)
2. Rouler la commande : python convert\_annotations.py

RÉSULTATS :  
- Dossier contenant les images pour Train, Validation et Test.

- Dossier contenant les labels pour les images de Train, Validation et Test.