Compte rendu entraînement sur le MBB

Les images du Dataset ont été aléatoirement séparées en 3 ensembles : Train (avec probabilité 0.8), Val (avec probabilité 0.1) et Test (avec probabilité 0.1).

Les ensembles obtenus ont comme taille :

* Train : 11724 images
* Val : 1427 images
* Test : 1487 images

L’entraînement initialement lancé sur 600 épochs s’est terminé à la 390ème époch en early stopping (pas d’amélioration des performances sur l’ensemble de validation pendant 100 épochs).

Les meilleures performances ont été observées à la 290ème époch.

Il aura fallu 15.688 heures (architecture large).

En validation les performances sont :

Précision | Rappel | mAP@.5 | mAP@.95

0.995 | 0.99 | 0.995 | 0.792

Avec les paramètres conf-thres=0.5 et iou\_thres=0.2 pour le test.

En test les performances sont :

Précision | Rappel | mAP@.5 | mAP@.95

0.992 | 0.988 | 0.993 | 0.796

La matrice de confusion montre ce que l’on sait déjà :

* 99% des hippocampes sont bien détectés (Rappel)
* 1% des hippocampes ne sont pas « vu » (1-Rappel)
* 100% des mauvaises détections d’hippocampes sont du fond

Les courbes de loss montrent le surapprentissage du modèle pour un nombre trop grand d’épochs.

Les exemples sur les batch de validation montre que les hippocampes bien visibles (tête en entière et/ou une grande partie du corps de profil) sont détectés avec des scores très élevés (>=0.8) et que, lorsque l’hippocampe est moins visible, le score diminue ou celui-ci n’est pas détecté.

Avec les paramètres conf-thres=0.8 et iou\_thres=0.2 pour le test.

En test les performances sont :

Précision | Rappel | mAP@.5 | mAP@.95

0.995 | 0.976 | 0.987 | 0.794

Avec les paramètres conf-thres=0.5 et iou\_thres=0.5 pour le test.

En test les performances sont :

Précision | Rappel | mAP@.5 | mAP@.95

0.991 | 0.991 | 0.994 | 0.797

On pourrait faire petit script pour chercher la meilleure combinaison de threshold mais pour l’instant 0.5/0.5 semble être un bon compromis.