**Projet 8 – Participer à une compétition Kaggle**

**Compétition choisie : TensorFlow – Help Protect the Great Barrier Reef**

**Description de la compétition par Kaggle**

But de la compétition

L'objectif de ce concours est **d'identifier avec précision les étoiles de mer en temps réel** en construisant un **modèle de détection d'objets** entraîné sur des vidéos sous-marines de récifs coralliens.

Votre travail aidera les chercheurs à identifier les espèces qui menacent la grande barrière de corail australienne et à prendre des mesures éclairées pour protéger le récif pour les générations futures.

Le contexte

La magnifique grande barrière de corail australienne est le plus grand récif de corail du monde et abrite 1 500 espèces de poissons, 400 espèces de coraux, 130 espèces de requins, de raies et une grande variété d'autres espèces marines.

Malheureusement, le récif est menacé, en partie à cause de la surpopulation d'une étoile de mer en particulier - l'étoile de mer à couronne d'épines mangeuse de corail (ou **COTS** en abrégé). Des scientifiques, des opérateurs touristiques et des gestionnaires de récifs ont mis en place un programme d'intervention à grande échelle pour contrôler les épidémies de COTS à des niveaux écologiquement durables.

[](https://www.youtube.com/watch?v=UT2noVDFoaA&ab_channel=TensorFlow)

Pour savoir où se trouvent les COTS, une méthode traditionnelle d'étude des récifs, appelée « Manta Tow », est réalisée par un plongeur tuba. Remorqués par un bateau, ils évaluent visuellement le récif, s'arrêtant pour enregistrer les variables observées tous les 200 m. Bien que généralement efficace, cette méthode est confrontée à des limites claires, notamment l'évolutivité opérationnelle, la résolution des données, la fiabilité et la traçabilité.

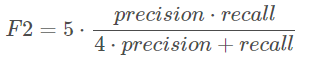
La Great Barrier Reef Foundation a mis en place un programme d'innovation pour développer de nouvelles méthodes d'enquête et d'intervention afin de fournir un changement radical dans le contrôle des COTS. Les **caméras sous-marines collecteront des milliers d'images** de récifs et la technologie de l'IA pourrait considérablement améliorer l'efficacité et l'échelle à laquelle les gestionnaires de récifs détectent et contrôlent les épidémies de COTS.

Pour faire évoluer les systèmes d'arpentage vidéo, l'agence scientifique nationale australienne, CSIRO s'est associée à Google pour développer une technologie innovante d'apprentissage automatique capable d'**analyser de grands ensembles de données d'images avec précision, efficacité et en temps quasi réel.**

**Evaluation des soumissions**

Cette compétition est évaluée sur le **score F2** à **différents seuils d'intersection sur union (IoU).** Le score F2 accorde plus d’importance au recall qu’à la précision, car dans ce cas, il est logique de tolérer certains faux positifs afin de s'assurer que très peu d'étoiles de mer sont manquées.

Formule du F2 score :



La métrique balaie les seuils d'IoU dans la plage de 0,3 à 0,8 avec un pas de 0,05, calculant un score F2 à chaque seuil. Par exemple, à un seuil de 0,5, un objet prédit est considéré comme un "hit" si son IoU avec un objet True est d'au moins 0,5.

Un vrai positif est la première soumission dans un échantillon avec une IoU supérieure au seuil par rapport à une bounding box qui ne match pas.

Une fois que toutes les bounding box de soumission ont été évaluées, toutes les bounding box de soumission sans correspondance sont des faux positifs ; toutes les bounding box de solution non appariées sont des faux négatifs.

Le score F2 final est calculé comme la moyenne des scores F2 à chaque seuil IoU. À l'intérieur de chaque seuil d'IoU, la métrique de concurrence utilise une micro moyenne ; chaque vrai positif, faux positif et faux négatif a le même poids les uns par rapport aux autres vrai positif, faux positif et faux négatif.

Dans votre soumission, vous êtes également invité à fournir un niveau de confiance pour chaque bounding box. Les bounding box sont évalués en fonction de leurs niveaux de confiance. Cela signifie que les bounding box avec une confiance plus élevée seront vérifiées en premier pour les correspondances avec les solutions, ce qui détermine quelles boîtes sont considérées comme des vrais et des faux positifs.

Vous devez vous soumettre à ce concours à l'aide de l'API de séries temporelles python fournie, qui garantit que les modèles n'apparaissent pas dans le temps. Pour utiliser l'API, suivez ce modèle dans Kaggle Notebooks :

**Format du livrable (submission format)**

import greatbarrierreef

env = greatbarrierreef.make\_env() # initialize the environment

iter\_test = env.iter\_test() # an iterator which loops over the test set and sample submission

for (pixel\_array, sample\_prediction\_df) in iter\_test:

sample\_prediction\_df['annotations'] = '0.5 0 0 100 100' # make your predictions here

env.predict(sample\_prediction\_df) # register your predictions

The submission format requires a space delimited set of bounding boxes. For example:

0.5 0 0 100 100

indicates that the image has a bounding box with a confidence of 0.5, at x == 0 and y == 0, with a width and height of 100.

0.3 0 0 50 50 0.5 10 10 30 30

would predict two bounding boxes in the image. Each prediction row needs to include all bounding boxes for the image.