

Étude de la biodiversité Comptage et identification des espèces



Le porteur

TotalEnergies est une compagnie multi-énergies intégrée mondiale. TotalEnergies est producteur et fournisseur d'énergies : pétrole et biocarburants, gaz naturel et gaz verts, renouvelables et électricité. Nos plus de 100 000 collaborateurs s'engagent pour une énergie meilleure, plus abordable, plus sûre, plus durable et accessible au plus grand nombre. Présent dans plus de 120 pays, notre ambition est de devenir la major de l'énergie responsable.

Contexte

Dans le contexte actuel de transition énergétique, TotalEnergies développe continuellement ces activités liées aux énergies renouvelables et cherche à renforcer ces ambitions et engagements envers la préservation de la biodiversité. Le Groupe souhaite ainsi disposer de connaissances sur la biodiversité présente sur chaque site dans le but de réduire son impact sur l'environnement.

L'équipe R&D Biodiversité travaille sur des projets de recherche afin d'améliorer ces pratiques au regard de la conservation et de la préservation de la biodiversité notamment l'amélioration des outils de suivi, dont les pièges photos ou « *phototrap* ». Ces dispositifs non-invasifs permettant de photographier la faune afin de réaliser un inventaire. TotalEnergies s'intéresse à ces techniques pour les appliquer au suivi de la biodiversité sur ses sites. A ce jour, les méthodes utilisées impliquent une étape de traitement et d'identification relativement coûteuse en temps au vu du nombre important de photos générées par ces dispositifs. Par conséquent, l'équipe cherche à développer un processus interne de traitement de ces photos visant à disposer d'un algorithme d'identification automatique des espèces qui permettraient d'effectuer un monitoring continu de la biodiversité.

Exemple de projet actuellement en cours: [Ouganda : Programme pour la biodiversité du projet Tilenga](#)

Objectifs

L'objectif de ce **Data Challenge** est de développer une méthodologie capable de traiter les images issues des **pièges photos** disposées sur des sites de TotalEnergies à travers le monde.

1. **Détection** : Identifier la présence d'animaux dans les images.
2. **Classification** : Classer les animaux détectés par espèce.
3. **Comptage** : Compter le nombre d'individus d'une espèce présente sur chaque image.



Fig. 1 – Exemple de détection/classification.



Fig. 2 – Exemple de détection/classification avec plusieurs individus.

Défi : Généralisation à des espèces non annotées

La généralisation à d'autres espèces dans le cadre de ce **Data Challenge** représente le défi majeur, car elle implique de créer un modèle capable de détecter et de classer des animaux qui n'ont pas été annotés dans le jeu de données d'origine. Le but est d'étendre le modèle au-delà des classes initiales et d'adapter sa capacité à identifier une large variété d'espèces animales dans des environnements divers. L'objectif est donc de créer une méthodologie flexible, capable de s'adapter à des environnements variés, où des espèces très différentes peuvent être présentes. La technologie doit être facilement évolutive pour permettre d'introduire de nouvelles données (nouvelle espèce, nouveau nom de classe, base de données sur une nouvelle zone) et d'assurer une amélioration continue. Cela implique d'intégrer une forte capacité de **généralisation**.

Descriptif des jeux de données

Notes :

- Les données sont accompagnées d'un Notebook en Python (visualisation des données).
- Les données représentent un petit sous-ensemble des données complètes.

Données annotées

Le jeu de données comprend 9 861 images au format JPEG. La labélisation des photos a été effectuée manuellement.

Liste des 8 classes (espèces) disponibles pour le data challenge :

- Ardea-cinerea (<https://fr.wikipedia.org/wiki/Ardea>)
- Canis-lupus-familiaris (<https://fr.wikipedia.org/wiki/Chien>)
- Capreolus-capreolus (https://fr.wikipedia.org/wiki/Chevreuil_d%27Europe)
- Genetta-genetta (https://fr.wikipedia.org/wiki/Genette_commune)
- Martes-martes (https://fr.wikipedia.org/wiki/Martre_des_pins)
- Meles-meles (https://fr.wikipedia.org/wiki/Blaireau_europ%C3%A9en)
- Sus-scrofa (<https://fr.wikipedia.org/wiki/Sanglier>)
- Vulpes-vulpes (<https://fr.wikipedia.org/wiki/Vulpes>)

Étant donné la diversité des environnements (lumière, météo, saisons, etc.), des techniques d'augmentation des données peuvent être utilisées pour enrichir les images disponibles, notamment par des changements de luminosité, ou transformations géométriques, etc. Ce prétraitement aidera à renforcer la robustesse des modèles face à des images variées et complexes.

Données non annotées

Vous disposez de 541 images non labélisées.

L'objectif est de tester et d'améliorer la capacité des algorithmes/modèles à généraliser, c'est-à-dire à identifier et compter la faune sur de nouvelles images qui n'ont pas été préalablement annotées.

Pour simplifier la démarche dans le cadre du Data Challenge, les images non annotées proviennent de régions géographiques similaires à celles des images annotées et vous avez le nom de l'espèce qui est indiqué dans le nom du fichier. Dans un contexte réel la méthodologie doit être capable de s'appliquer à des images issues de toutes les régions du monde (Afrique, Amérique, Asie, etc.) avec leurs spécificités géographiques et environnementales. L'objectif est donc de rendre la méthodologie suffisamment flexible et précise pour être utilisée à l'échelle mondiale, contribuant ainsi à la surveillance et à la préservation de la biodiversité dans divers écosystèmes.

Bon travail et bonne chance à tous !