
Détection d'animaux camouflés par information temporelle et modèles Vision-Langage

Mounir Ammam

Université de Montréal

mounir.ammam@umontreal.ca

Byungsuk Min

Université de Montréal

byungsuk.min@umontreal.ca

Yanis Chikhar

Université de Montréal

yanis.chikhar@umontreal.ca

1 Description du Projet

La détection d'animaux camouflés dans des environnements naturels est un problème difficile en vision par ordinateur, en particulier lorsque les indices visuels statiques sont faibles. Le jeu de données *Moving Camouflaged Animals* (MoCA) fournit des séquences vidéo d'animaux camouflés annotées avec des boîtes englobantes et des types de mouvement, ce qui en fait un cadre idéal pour étudier ce défi¹.

L'objectif de ce projet est d'évaluer dans quelle mesure l'information temporelle (le mouvement) et les modèles *Vision-Langage* (VLM) peuvent améliorer la détection d'animaux camouflés par rapport à des approches basées uniquement sur des images individuelles.

Dans un premier temps, nous établirons une base de référence à l'aide d'un modèle de détection image par image. Ensuite, nous intégrerons des indices de mouvement (différence d'images, flot optique ou agrégation temporelle simple) afin de mesurer l'apport de l'information temporelle.

Dans un second temps, nous utiliserons un modèle *Vision-Langage* préentraîné (par exemple CLIP) comme composant sémantique complémentaire pour reclasser ou filtrer des régions candidates à l'aide de descriptions textuelles telles que « *animal camouflé dans la végétation* ». Cette approche ne nécessite pas d'entraînement supplémentaire du VLM.

Les différentes méthodes seront comparées à l'aide de métriques standards de détection et d'analyses d'erreurs selon le type de mouvement et le niveau de camouflage. Ce projet adopte une approche exploratoire et vise à mieux comprendre le rôle du mouvement et des connaissances sémantiques dans la détection d'animaux camouflés.

¹<https://www.robots.ox.ac.uk/~vgg/data/MoCA/>