Détection et caractérisation non-supervisée d'arbres à partir d'images aériennes

Olivier Schwander

2026

Environnement

Équipes et laboratoires :

- Équipe MLIA, ISIR, Sorbonne Université
- Équipe FEA, IEEE, Sorbonne Université

Encadrement et contacts :

- à l'ISIR, Olivier Schwander `<olivier.schwander@isir.upmc.fr>`
- à l'IEEE, Jacques Gignoux `<jacques.gignoux@upmc.fr>`

Ce stage s'adresse à des étudiant·e·s en apprentissage statistique et intelligence artificielle et lea stagiaire sera localisé à l'ISIR sur le campus Jussieu.

Contexte

Au sein de l'Institut d'écologie et de sciences de l'environnement de Paris, les écologues s'intéressent à la dynamique des populations d'arbres pour comprendre le fonctionnement des écosystèmes à différentes échelles de temps et d'espace. Les méthodes actuelles reposent sur une identification et un suivi manuel des arbres à partir d'images aériennes prises par des drones. Le temps nécessaire pour cet étiquetage limite bien sûr le nombre d'arbres étudiés.

L'objectif de ce projet est d'étudier des méthodes de classifications automatiques pour des vues aériennes d'images, en respectant les contraintes et les objectifs donnés par les écologues. Il s'agira entre autre d'être capable de segmenter finement les populations d'arbres et de caractériser leur état de croissance (voir Figure 1).

On peut voir Figure 2 que les images sont très hétérogènes en qualité et résolution spatiale. Elles ont été en effet acquises par différents moyens (avion, drone et même cerf-volant, avec des caméras différentes et sur des périodes allant de 1963 à nos jours).

Objectifs

Les méthodes supervisées pour la détection d'objets montrent leurs limites quand les données sont très hétérogènes et que la supervision est difficile à obtenir. L'objectif du stage est d'explorer les capacités des modèles de fondation à extraire les informations pertinentes à partir des images aériennes et à s'adapter à différents domaines (notamment les différentes périodes de prise de vue).

Au niveau méthodes, trois axes sont principalement envisagés :

- les modèles permettant la segmentation zero-shot [8, 9] ;
- les modèles exploitant un alignement texte-image pour améliorer la segmentation [2] ;
- les techniques d'adaptation de modèle lors de l'inférence [6].

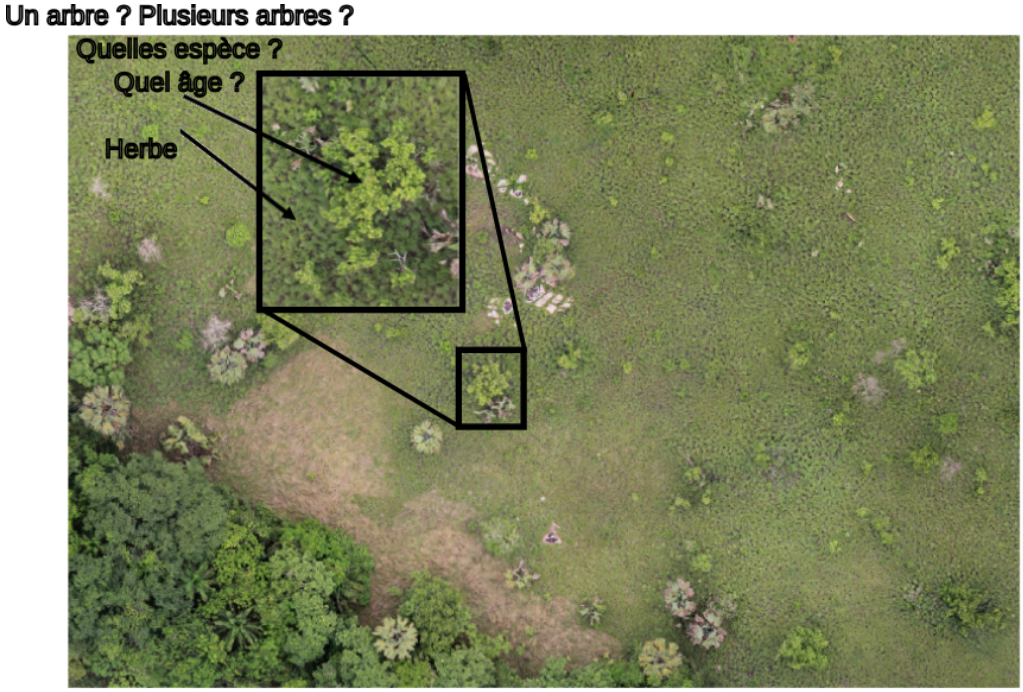

FIGURE 1 : Image de savane (vue de drone 2012). De nombreux arbres sont présents sur cette image, les écologues s'intéressent à leur position et à leur taille. Une des difficultés est ici de distinguer individuellement les abres.

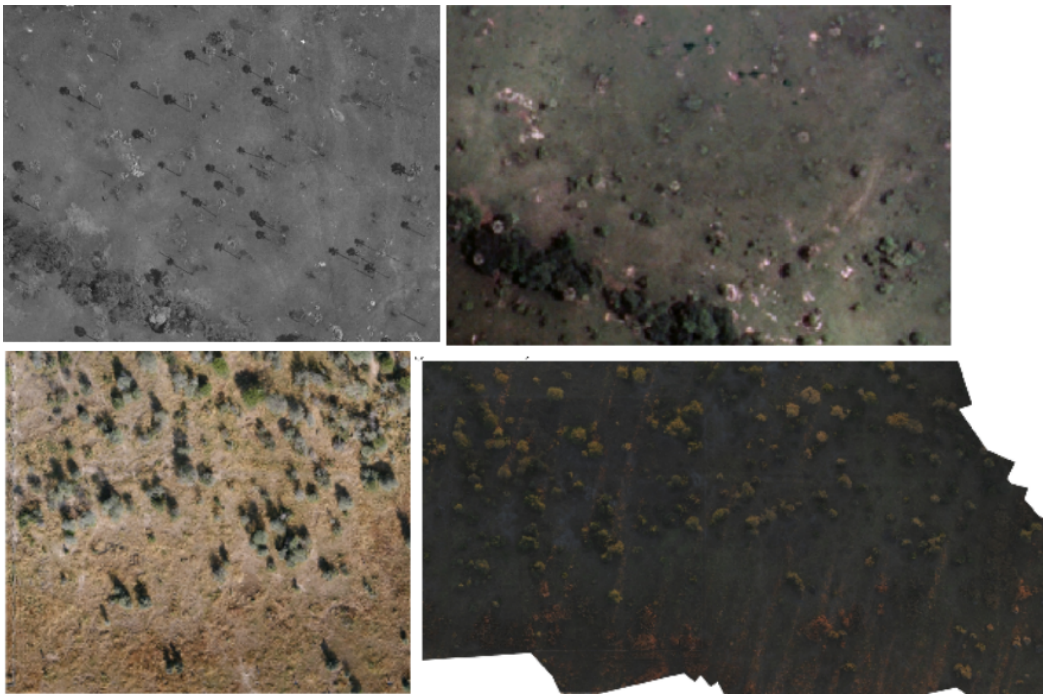

FIGURE 2 : Images issues de différentes campagnes de mesure. En haut à gauche, prise de vue d'avion en 1963, à droite en 1988 ; en bas à gauche, vue depuis un cerf-volant en 2008, à droite, depuis un drone en 2012. On observe de très grosses différences d'angles de vue, de luminosité et de qualité.

Du point de vue applicatif, il sera particulièrement intéressant d'obtenir un modèle unique, capable de traiter les données passées et futures, quel que soit le type de prise de vue et la qualité des données. De plus, l'alignement texte-image permettra aux écologues de préciser en langue naturelle les résultats souhaités.

On veillera en particulier à s'intéresser aux modèles compacts pour limiter le coût calculatoire et surtout le coût environnemental.

Pré-requis

- Apprentissage statistique, intelligence artificielle, apprentissage profond
- Outils logiciels classiques (Python, PyTorch, etc)
- Intérêt pour les applications en écologie, mais aucune connaissance particulière n'est requise

Bibliographie

[1] Moondream/Moondream3-preview · Hugging Face. Retrieved November 20, 2025 from https://huggingface.co/moondream/moondream3-preview

[2] Luca Barsellotti, Lorenzo Bianchi, Nicola Messina, Fabio Carrara, Marcella Cornia, Lorenzo Baraldi, Fabrizio Falchi, and Rita Cucchiara. 2025. Talking to DINO : Bridging Self-Supervised Vision Backbones with Language for Open-Vocabulary Segmentation. https://doi.org/10.48550/arXiv.2411.19331

[3] Hritam Basak and Zhaozheng Yin. 2025. SemiDAViL : Semi-supervised Domain Adaptation with Vision-Language Guidance for Semantic Segmentation. https://doi.org/10.48550/arXiv.2504.06389

[4] Shuo Chen, Jindong Gu, Zhen Han, Yunpu Ma, Philip Torr, and Volker Tresp. 2023. Benchmarking Robustness of Adaptation Methods on Pre-trained Vision-Language Models. https://doi.org/10.48550/arXiv.2306.02080

[5] Matteo Farina, Gianni Franchi, Giovanni Iacca, Massimiliano Mancini, and Elisa Ricci. 2024. Frustratingly Easy Test-Time Adaptation of Vision-Language Models. https://doi.org/10.48550/arXiv.2405.18330

[6] Matteo Farina, Massimiliano Mancini, Giovanni Iacca, and Elisa Ricci. 2025. Rethinking Few-Shot Adaptation of Vision-Language Models in Two Stages. https://doi.org/10.48550/arXiv.2503.11609

[7] Gustavo Adolfo Vargas Hakim, David Osowiechi, Mehrdad Noori, Milad Cheraghalikhani, Ali Bahri, Moslem Yazdanpanah, Ismail Ben Ayed, and Christian Desrosiers. 2024. CLIPArTT : Adaptation of CLIP to New Domains at Test Time. https://doi.org/10.48550/arXiv.2405.00754

[8] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. 2023. Segment Anything. https://doi.org/10.48550/arXiv.2304.02643

[9] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. 2024. DINOv2 : Learning Robust Visual Features without Supervision. https://doi.org/10.48550/arXiv.2304.07193