

Proposition de stage de Master 2 Modélisation en écologie

Nom de l'entreprise ou du laboratoire: **LISAH ou IGEPP**

Adresse où se déroulera le stage:

LISAH	IGEPP
2 place Pierre Viala 34060 Montpellier	Domaine de la Motte BP 35327 35653 Le Rheu cedex

Identification d'espèces végétales d'intérêt écohydrologique dans les paysages agricoles : apports du Deep Learning et de l'Intelligence Artificielle

Contexte scientifique et résumé du projet de stage:

Pouvoir identifier des espèces végétales sur le terrain est particulièrement important en agriculture afin de détecter précocement les mauvaises herbes (Olsen et al 2019, Pearline et al 2019), en écologie pour analyser la richesse en biodiversité d'un paysage, suivre les populations d'espèces menacées ou invasives et analyser les impacts du changement climatique sur la distribution des espèces (Waldchen and Mader 2018), en hydrologie pour caractériser les propriétés physiques des couverts végétaux vis-à-vis des écoulements (Rubol 2018). Cette activité a été longtemps réalisée par des botanistes via des relevés visuels, limitant de ce fait l'exhaustivité et/ou la taille des zones couvertes. Pouvoir automatiser cette activité d'identification des espèces pour l'appliquer à l'échelle d'un paysage est un véritable défi méthodologique.

L'automatisation offre plusieurs avantages : identifier des espèces devient accessible aux non-initiés et cela permet de répondre aux défis actuels et à venir sur le déclin de la biodiversité mais aussi des botanistes. L'automatisation permet également une plus grande exhaustivité dans la caractérisation des couverts végétaux lorsqu'il s'agit d'estimer leur propriété fonctionnelle vis-à-vis d'une fonction écosystémique (par exemple leur perméabilité vis-à-vis d'un flux physique). Grâce à l'essor des nouvelles sources d'acquisition de données (proxidtection, capteurs intelligents sur smartphones) et de traitement et classification des images obtenues (deep learning, machine learning), il est aujourd'hui possible d'identifier automatiquement des espèces et de les cartographier (Olsen et al 2019, Alvarez-Vanhard 2020, Sun et al 2017). Son application en agriculture, et en particulier aux linéaires agricoles (fossés, talus, haies) très riches en biodiversité et canalisant de nombreux flux abiotiques (eau, sédiments, polluants), est encore insuffisamment développée à l'échelle du paysage.

Dans ce contexte, un jeu de données préliminaire a été collecté sur les espèces végétales spontanées trouvées dans un réseau hydrographique (ici des fossés agricoles) dans un paysage agricole. Le jeu de données est constitué d'un relevé exhaustif botanique cartographié et d'une campagne de mesure par proxidtection où environ 4000 photos ont été prises selon des angles de vue différents (appareil photo sur perche) le long du linéaire. Un premier traitement photogrammétrique des images a permis la création d'une orthoimage à résolution sub-centimétrique ainsi qu'un positionnement de l'emprise de chacune des photos sur l'orthoimage associée aux propriétés de prise de vue : angle, inclinaison et résolution au sol.

L'objectif principal du stage est d'estimer la plus-value à utiliser les images brutes, multi-angle, dans la classification (supervisée ou non) d'une orthoimage. On se basera sur les méthodes de classification classiques (Sommer et al, 2011) mais aussi des approches de type Deep Learning pour effectuer la classification (Heipke & Rottensteiner, 2020). Il s'agira également d'étudier les possibilités d'extrapolation de la méthode à l'échelle du paysage.

Les résultats du stage permettront d'améliorer notre connaissance de la biodiversité végétale dans les milieux agricoles ou semi-naturels, en particulier les milieux interstitiels entre parcelles. Ils participeront à la collecte de jeux de données permettant la calibration de modèles de distribution des espèces, et pourront être utilisés pour la détection précoce d'espèces invasives dans les paysages.

Déroulé du stage et profil recherché

Le stage débutera par une phase de prise en main des données, des logiciels et de la bibliographie (1 mois). Une phase de classification de l'orthoimage sera réalisée (1 mois), où il sera possible d'appliquer des méthodes de classification déjà éprouvées, tant en machine learning qu'en deep learning pour le traitement d'images à haute résolution (Sommer et al, 2011 ; Hu et al, 2018). Cela offrira des premiers classifieurs dont les performances pourront être évaluées et éventuellement combinées (approches ensemblistes). La majeure partie du stage (3 mois) sera consacrée à l'utilisation d'information multimodales (Gaetano et al 2018) afin d'intégrer des données photogrammétriques brutes au processus de classification. Enfin, la dernière partie du stage (1 mois) sera consacrée à la rédaction du rapport et la préparation de soutenance du/de la stagiaire. NB : une participation à la collecte d'échantillons de terrain pour compléter les bases de données botaniques sur les espèces végétales des fossés agricoles pourra être envisagée selon les intérêts du stagiaire.

Profil recherché:

Etudiant intéressé par le traitement d'image, les statistiques et l'analyse de données spatialisées. La maîtrise d'un langage de programmation adapté à ces questions scientifiques, comme Python, est nécessaire. Une connaissance préalable des algorithmes et bibliothèques logiciels adaptés au sujet du stage (scikit-learn, scikit-image, TensorFlow, Keras) sera bienvenue.

Montant des indemnités de stage: environ 500 euros/mois.

Responsable du stage (personne à contacter): Fabrice Vinatier (CR) et Nicolas Parisey (IR) (fabrice.vinatier@inrae.fr et nicolas.parisey@inrae.fr)

References :

- Olsen, A., Konovalov, D. A., Philippa, B., Ridd, P., Wood, J. C., Johns, J., Banks, W., Girgenti, B., Kenny, O., Whinney, J., Calvert, B., Azghadi, M. R., & White, R. D. (2019). DeepWeeds: A Multiclass Weed Species Image Dataset for Deep Learning. *Scientific Reports*, 9(1).
- Christian Heipke & Franz Rottensteiner (2020) Deep learning for geometric and semantic tasks in photogrammetry and remote sensing, *Geo-spatial Information Science*, 23:1, 10-19
- Gaetano, R.; Ienco, D.; Ose, K.; Cresson, R. A Two-Branch CNN Architecture for Land Cover Classification of PAN and MS Imagery. *Remote Sens.* 2018, 10, 1746.
- Heipke, C., & Rottensteiner, F. (2020). Deep learning for geometric and semantic tasks in photogrammetry and remote sensing. *Geo-Spatial Information Science*, 23(1), 10-19.
- Hu, Y.; Zhang, Q.; Zhang, Y.; Yan, H. A Deep Convolution Neural Network Method for Land Cover Mapping: A Case Study of Qinhuangdao, China. *Remote Sens.* 2018, 10, 2053.
- Olsen, A., Konovalov, D. A., Philippa, B., Ridd, P., Wood, J. C., Johns, J., Banks, W., Girgenti, B., Kenny, O., Whinney, J., Calvert, B., Azghadi, M. R., & White, R. D. (2019). DeepWeeds: A Multiclass Weed Species Image Dataset for Deep Learning. *Scientific Reports*, 9(1).
- Rubol, S., Ling, B., & Battiato, I. (2018). Universal scaling-law for flow resistance over canopies with complex morphology. *Scientific Reports*, 8(1), 4430.
- Sommer C, Straehle C, Köthe U, Hamprecht FA. 2011. Ilastik: Interactive learning and segmentation toolkit. 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. IEEE: Piscataway, NJ, USA, 230-3
- Sun, Y., Liu, Y., Wang, G., & Zhang, H. (2017). Deep Learning for Plant Identification in Natural Environment. *Computational Intelligence and Neuroscience*, 2017, 1-6.