

Quelle vérité terrain pour les réseaux de neurones en imagerie drone ? Application à la détection de palmier *Raphia* en forêts au Gabon

G.Viennois¹ F. Borne² M.Jaeger² P. Borianne²

¹ AMAP, Univ Montpellier, CNRS, CIRAD, INRA, IRD, Montpellier, France

² CIRAD, UMR AMAP, F-34398 Montpellier, France.

AMAP, Univ Montpellier, CIRAD, CNRS, INRA, IRD, Montpellier, France

g.viennois@cirad.fr

1 Introduction

Les palmiers sont une des familles de plantes les plus utiles, fournissant sous les tropiques, de nombreux avantages économiques, écologiques et sociétaux [1] [2]. L'espèce du genre *Raphia* est la plus largement utilisée et importante d'Afrique. Cependant, la pression anthropique a des impacts négatifs au niveau local et régional sur les *Raphias* et aucune estimation des paramètres de base (nombre d'individus, biomasse ...) n'a été entreprise. Le *Raphia* est une monocotylédone de la famille des *Arecaceae* pouvant atteindre 25 m de hauteur présente dans les milieux marécageux et le long des fleuves. Il n'a pas de stipe et son architecture est caractérisée par une tige monoaxiale constituée d'un bouquet de palmes mortes et vivantes [3]. Les feuilles sont disposées selon une symétrie radiale avec une phyllotaxie en hélice régulière autour de l'axe central. Sur une vue aérienne, un palmier adulte isolé apparaît comme une « étoile » constituée de feuilles disposées radialement autour d'un axe central avec des inclinaisons, des flexions et des torsions variables [4]. En condition naturelle, ces *Raphias* se regroupent en forêts marécageuses appelées Raphiales où les couronnes sont imbriquées. Il est donc très difficile de les individualiser et notamment en vue aérienne. A ce jour, aucune étude n'a été faite sur le dénombrement des *Raphias* en forêt. Les études récentes de comptage de palmiers [5] [6] ne concernent que des plantations de palmiers à huile. Elles utilisent des approches « Deep Learning », et plus précisément de « Réseaux Neuronaux Convolutionnels » [7] [8]. En imagerie satellite, des études [9] [10] montrent le potentiel de ces méthodes pour la reconnaissance d'objets, mais très peu de travaux sur les drones ont été conduits malgré le potentiel de ces vecteurs [11]. Dans cette étude, nous employons le réseau Faster Region-based CNN [12] dédié à la détection d'objets vus du sol sur des images RVB pour la détection de *Raphias* à partir d'images drones en milieu naturel.

2 Matériels et Méthodes

2.1 Images

Le jeu d'images est constitué de 5 ortho-mosaïques (12000 x 19000 pixels) acquises par drone (DJI Mavic Pro) dans la région du Bas Ogooué au Gabon. Les vues ont été acquises entre 100 et 200 mètres d'altitude via une caméra RVB (CMOS 1/23'' 12.35 mégapixels). Les ortho-mosaïques ont été produites avec le logiciel Pix4D. Les résolutions des 5 ortho-images varient de 3,7 cm à 5,9 cm ; 3 sont utilisées comme jeu d'entraînement et 2 comme jeu de validation.

2.2 Méthodologie

Nous avons déployé le réseau de neurones Faster-RCNN [12] sous l'environnement *Caffe* [13] et l'avons spécialisé pour la détection de *Raphias* sur des images drone. Plus précisément, le réseau Faster R-CNN a été couplé au modèle architectural VGG-16 [14] dédié à la reconnaissance d'images à large échelle ; le tout a été pré-entraîné sur la base de données annotées PASCAL VOC 2007 [15]. Nous nous sommes restreints à deux classes correspondant respectivement aux labels « palmier » et « autre ». Le ré-apprentissage a consisté à spécialiser le réseau pré-entraîné à reconnaître des *Raphias*. La dernière couche a donc été renommée et sa pondération évaluée par la spécialisation du réseau sur un jeu d'entraînement. Les ortho-mosaïques d'images sont préalablement découpées en tuiles de 500 x 500 pixels, puis annotées sous ImageJ : l'annotation consiste à faire délimiter visuellement par un expert chaque palmier distinguable, par une boîte encadrant sa couronne. L'apprentissage réalisé, le réseau peut alors être utilisé pour détecter les *Raphias* sur les deux jeux d'images. L'entraînement du réseau est réalisé pour différentes valeurs d'itérations, i.e. nombre de fois que le réseau "considèrera" les données dans ses fonctions itératives de minimisation, afin d'optimiser le réglage pour la plus faible erreur globale de prédiction E. Enfin, les annotations expertes sont comparées avec les annotations prédites par le réseau dans les 2 conditions : entraînement et validation. Pour cela, les boîtes des expertes sont bijectivement appariées à celles du réseau sur les jeux d'entraînement et de validation. La précision (P), le rappel (R) ou *sensibilité* et la F-mesure (F) [16] sont calculés comme suit pour qualifier les prédictions du réseau :

$$E = |FP| + |FN|, \quad P = \frac{|VP|}{|VP| + |FP|}, \quad R = \frac{|VP|}{|VP| + |FN|}, \quad F = \frac{2 \times P \times R}{P + R}$$

où |VP| est le nombre de « Vrai Positif » i.e. des objets identifiés par l'expert et par le réseau comme des palmiers, |FN| le nombre de « Faux Négatif » i.e. des objets identifiés uniquement par l'expert comme des palmiers, et |FP| le nombre de « Faux Positif » i.e. ceux identifiés uniquement par le réseau comme des palmiers.

3 Résultats

La figure 1 illustre la sensibilité du réseau en fonction du nombre d'itérations. L'erreur est stabilisée dès 5000 itérations sur les 2 jeux de données. Dans le cas du jeu d'entraînement, le réseau détecte 850 *Raphias* et commet 1,7 % d'erreurs. Sa F-mesure est de 0,99, indiquant que le réseau a « bien appris ». Sur le jeu d'images de validation (non utilisé pour l'entraînement du réseau), le réseau détecte 360 *Raphias*. Le taux d'erreur est de 54 % dont 20 % de Faux Positifs et 34 % de Faux Négatifs. La F-mesure n'est plus que de 0,63 illustrant la difficulté du réseau à détecter des *Raphias*, en particulier lorsque la densité des individus est importante dans les Raphiales. Ces résultats doivent être modérés du fait que les annotations expertes sont partielles sur les 2 jeux de données.

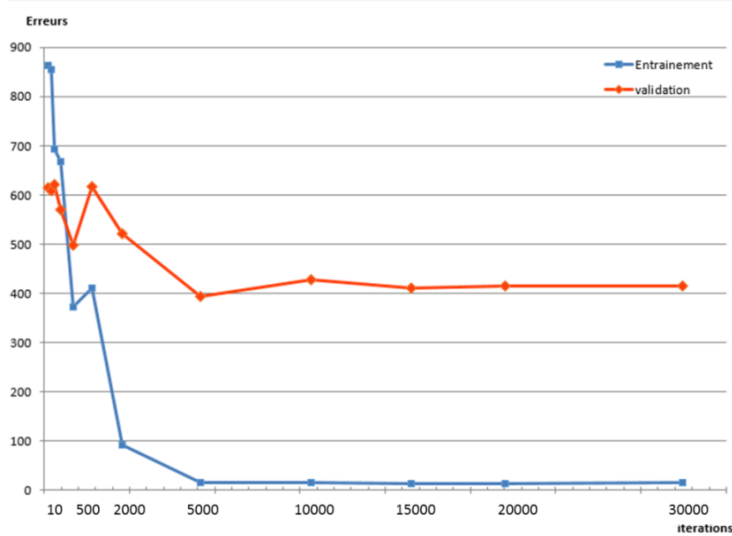


Figure 1: variation de l'erreur durant l'apprentissage

La figure 2 rend compte de la complexité du dénombrement due à l'imbrication des couronnes dans une Raphiale et de la difficulté pour l'expert à correctement annoter l'intégralité des images du jeu d'apprentissage. Le jeu d'annotations créé par l'expert n'est donc pas optimal et impacte la performance de prédiction du réseau.

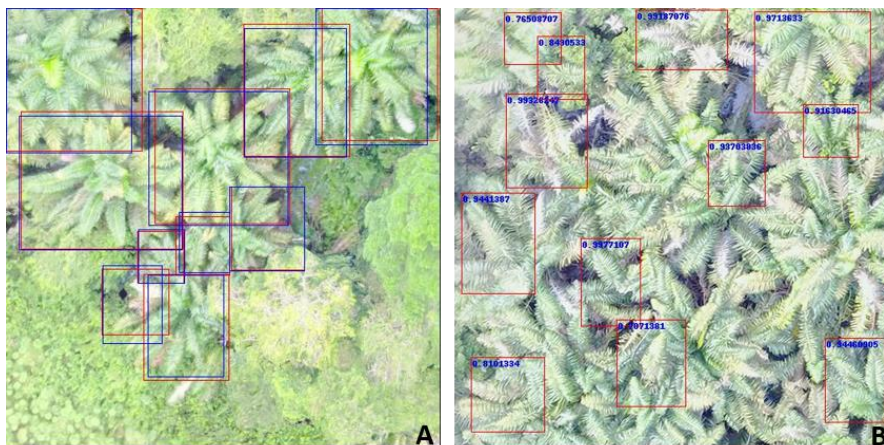


Figure 2 : Annotations et Prédictions. (A) Tuile du jeu d'entraînement avec annotations expertes en bleu et prédictions du réseau en rouge. (B) Prédictions du réseau sur une tuile du jeu de validation : les valeurs sont les probabilités d'appartenance à la classe palmier.

4 Conclusions et perspectives

Quand les conditions de reconnaissance des *Raphias* sont simples (peu d'imbrication), le réseau est plutôt performant. Par contre, dès que la densité des *Raphias* augmente, les performances semblent se dégrader. Toutefois, le réseau est capable d'identifier des *Raphias* non détectés par l'expert. Se pose alors le problème de la validation des résultats. Combien y a-t-il de *Raphias* dans une forêt marécageuse ? L'expert n'est pas en capacité de le déterminer, ce qui impacte l'annotation des jeux de données et questionne sur la pertinence de l'entraînement et de la validation. Une première solution envisagée serait de travailler avec une méthode d'annotation « semi-quantitative », i.e. caractériser un polygone par un intervalle d'effectifs plutôt que par un effectif strict établi individu par individu. Il s'ensuit que la F-mesure n'est plus applicable en l'état. Il nous faudrait donc trouver un indicateur plus adapté à ce cas de figure, peut être à rechercher dans le domaine de la logique floue. Une autre piste complémentaire serait de générer des *Raphias* virtuelles 3D (travail déjà effectué sur d'autres espèces de palmiers [4] [17]) en faisant varier plusieurs paramètres (hauteur, la densité des individus). Le jeu d'annotations serait plus facile à constituer et peut être plus pertinent que l'expertise visuelle. Les jeux d'images usuels (PascalVOC) pour les réseaux de neurones portent sur des objets aux caractéristiques généralement bien définies (contour, couleur, texture), ce qui leur permet d'atteindre des performances excellentes (supérieures à 95 %). Dans notre cas, les caractéristiques sont moins bien définies car il s'agit d'objets naturels, en croissance, multi-imbriqués. Par exemple, les palmes apparaissent comme une flèche gainée verticale, avant de s'ouvrir et s'incliner, produisant ainsi des variations d'aspect importantes d'une couronne à une autre. Cette plasticité limite la performance ; il est nécessaire d'avoir des annotations mieux adaptées à notre cas d'étude, en particulier en simulant la complexité de la réalité pour localiser les individus.

Remerciements

Les auteurs remercient N. Barbier (IRD, Amap) pour l'acquisition des images drones et H. Rey (CIRAD, Amap) pour ses annotations expertes.

Bibliographie

- [1] F. I. Obahiagbon, « A review of the origin, morphology, cultivation, economic products, health and physiological implications of raphia palm », *African journal of food science*, vol. 3, n° 13, p. 447–453, 2009.
- [2] W. L. Eiserhardt, J.-C. Svenning, W. D. Kissling, et H. Balslev, « Geographical ecology of the palms (Arecaceae): determinants of diversity and distributions across spatial scales », *Annals of Botany*, vol. 108, n° 8, p. 1391–1416, 2011.
- [3] M. Hoff et J. Florence, « Note sur une Raphiale du Gabon », *Bul. société d'histoire naturelle de Colmar*, 1977.
- [4] P. de Reffye, R. Lecoustre, J. Dauzat, S. Ouattara, A. Flori, et Y. P. N'cho, « Modelling plant architecture. Application to tropical agronomic perennial plants. The particular case of the Palmaceae. », *Oléagineux (Paris)*, vol. 44, n° 11, p. 537–546, 1989.
- [5] E. K. Cheang, T. K. Cheang, et Y. H. Tay, « Using Convolutional Neural Networks to Count Palm Trees in Satellite Images », p. 3.
- [6] W. Li, H. Fu, L. Yu, et A. Cracknell, « Deep Learning Based Oil Palm Tree Detection and Counting for High-Resolution Remote Sensing Images », *Remote Sensing*, vol. 9, n° 12, p. 22, déc. 2016.
- [7] Y. LeCun, Y. Bengio, et G. Hinton, « Deep learning », *nature*, vol. 521, n° 7553, p. 436, 2015.
- [8] A. Krizhevsky, I. Sutskever, et G. E. Hinton, « Imagenet classification with deep convolutional neural networks », in *Advances in neural information processing systems*, 2012, p. 1097–1105.
- [9] D. Inco, R. Gaetano, C. Dupaquier, et P. Maurel, « Land Cover Classification via Multi-temporal Spatial Data by Recurrent Neural Networks », p. 7.
- [10] D. H. T. Minh *et al.*, « Deep Recurrent Neural Networks for mapping winter vegetation quality coverage via multi-temporal SAR Sentinel- », p. 7.
- [11] S. A. Wich, « Drones and conservation », *Drones and Aerial Observation: New Technologies for Property Rights, Human Rights, and Global Development*, 2015.
- [12] I. Sa, Z. Ge, F. Dayoub, B. Uproft, T. Perez, et C. McCool, « Deepfruits: A fruit detection system using deep neural networks », *Sensors*, vol. 16, n° 8, p. 1222, 2016.
- [13] Y. Jia *et al.*, « Caffe: Convolutional architecture for fast feature embedding », in *Proceedings of the 22nd ACM international conference on Multimedia*, 2014, p. 675–678.
- [14] K. Simonyan et A. Zisserman, « Very deep convolutional networks for large-scale image recognition », *arXiv preprint arXiv:1409.1556*, 2014.
- [15] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, et A. Zisserman, « The pascal visual object classes (voc) challenge », *International journal of computer vision*, vol. 88, n° 2, p. 303–338, 2010.
- [16] D. M. Powers, « Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation », 2011.
- [17] S. Griffon et F. De Coligny, « AMAPstudio: an editing and simulation software suite for plants architecture modelling », *Ecological Modelling*, vol. 290, p. 3–10, 2014.