

Utilisation de techniques d'intelligence artificielle pour la détection d'animaux en forêt

Bernard BENET¹, Anders MÅRELL², Yves BOSCARDIN²

Bernard BENET



Bernard Benet est Ingénieur/Chercheur en Automatique/Robotique. Depuis 1991 au Cemagref, puis à l'Irstea et actuellement à l'INRAE, dans l'équipe Robotique ROMEA du Centre de Clermont-Fd, il mène des activités de recherche qui portent sur la conception et le développement, sur le plan matériel et logiciel, de systèmes de perception dans l'environnement naturel (capteurs de vision, lasers,...) à embarquer sur des machines agricoles ou des plateformes robotiques, pour des applications d'agriculture de précision ou d'agroalimentaire. Il est notamment spécialisé dans le développement de logiciels de traitement d'images, pour diverses applications de vision artificielle (mesures géométriques et colorimétriques, tâches de détection et d'identification) portant sur différents types d'objets naturels (cultures agricoles diverses, plantes, animaux,...). Il s'intéresse depuis quelques années aux outils et méthodes d'intelligence artificielle, qui apportent beaucoup d'améliorations dans tous les travaux d'analyse d'image menés, en particulier dans les applications de détection et d'identification d'objets naturels (animaux en forêt, ...) et dans les mesures réalisées dans diverses cultures agricoles.

Anders MÅRELL



Anders Mårell est passé de l'étude des effets du changement climatique sur le comportement alimentaire des rennes dans le subarctique suédois à évaluer les effets des ongulés sauvages sur le fonctionnement et la biodiversité des écosystèmes des forêts de plaine de France. Il a soutenu sa thèse de doctorat en 2006 à l'Université suédoise des sciences agricoles, Umeå, Suède. À son arrivée en France, il travaillait pour un réseau européen sur la recherche à long terme des écosystèmes forestiers et des paysages au GIP ECOFOR à Paris, en France. Il a également étudié les feux de forêt en région Méditerranéenne lorsqu'il travaillait à l'URFM, INRA, Avignon. Il dirige actuellement une équipe de recherche au sein de l'unité de recherche EFNO (INRAE) à Nogent-sur-Vernisson, France, où il investit son temps dans la gestion adaptative de la faune sauvage. Plus particulièrement, il mène des recherches et développe des procédures pour mieux intégrer les bénéfices et les coûts socio-écologiques des ongulés sauvages dans la gestion courante.



Yves Boscardin de formation forestière est assistant ingénieur à l'INRAE où il participe aux travaux de l'équipe FONA spécialisée dans les interactions entre la Forêt, les ONgulés et les Activités humaines. Depuis 1991 au Cemagref puis à l'Irstea, il a réalisé de nombreux suivis des populations de cervidés pour mettre au point des indicateurs de changement écologique utilisés dans la gestion. Passionné par le pistage des animaux, que ce soit des indices de présences, des mues ou des abrouissements, ses connaissances du terrain permettent de finaliser les protocoles et les procédures de recherche à la réalité du terrain. Dans le cadre du projet OPTMix en forêt domaniale d'Orléans, il est le

1 Université Clermont Auvergne, INRAE, UR TSCF, 63178, Aubière, France ; bernard.benet@inrae.fr

2 INRAE, UR EFNO, 45290, Nogent-sur-Vernisson, France ; anders.marell@inrae.fr; yves.boscardin@inrae.fr

coordinateur du suivi des rongeurs et de la grande faune. Pour cette dernière, il utilise des pièges photographiques depuis 2013 pour quantifier la fréquentation des différents milieux forestiers.

Résumé. Un logiciel de vision artificielle contenant des fonctions de traitement d'image et d'intelligence artificielle a été développé pour détecter des animaux sur des vidéos enregistrées par des pièges photographiques en milieux forestiers. Le suivi par pièges photographiques est utilisé sur le dispositif expérimental OPTMix en forêt d'Orléans pour estimer la fréquentation des placettes de mesure par les ongulés sauvages (chevreuil, cerf élaphe et sanglier). Les scientifiques et les naturalistes utilisent le suivi par pièges photographiques pour inventorier et suivre les populations animales et la biodiversité de la faune sauvage. Le traitement automatique des images afin de compter les animaux et identifier les espèces dans des programmes de suivi de la biodiversité pourrait faciliter et améliorer l'exploitation de ces données et rendre la méthodologie accessible à un plus grand nombre d'utilisateurs.

Mots clés : Vision artificielle, Intelligence artificielle, forêt, biodiversité, animaux

Abstract. An artificial vision software that contain image processing and deep learning functions has been developed to detect animals on videos recorded by camera traps in forests. Monitoring by camera traps is used on the OPTMix experimental facility in the Orléans forest to estimate wild ungulate pressure (roe deer, red deer and wild boar) on study plots. Scientists and conservation biologists use camera traps to monitor wildlife populations and biodiversity. Automatic image processing that count animals and identify species could facilitate and improve the use of camera traps in biodiversity monitoring programs and make the methodology accessible to a greater number of end users.

Keywords: Artificial vision, Deep learning, forest, biodiversity, animals

Introduction

Les scientifiques et les naturalistes sont nombreux à utiliser des dispositifs de capture d'images, appelés aussi « pièges photographiques »³, pour inventorier et suivre les populations animales et la biodiversité de la faune sauvage. Ces systèmes sont en particulier utilisés dans les régions éloignées et dans les milieux où il est difficile d'observer les animaux comme par exemple dans les vastes étendues de la forêt tropicale. Ces pièges permettent également d'étudier des animaux rares et cryptiques tels que les grands prédateurs (tigre, panthère). En France, le suivi par les pièges photographiques est employé par les agents de l'OFB (Office Français de la Biodiversité) pour étudier et suivre les populations de loup, de lynx et d'ours brun. Ils sont également utilisés par les chasseurs pour suivre les populations de grand gibier (chevreuil, cerf élaphe, sanglier), en milieu forestier.

Dans le cadre du projet OPTMix⁴, mené par l'UR EFNO de Nogent-sur-Vernisson, des systèmes de capture d'image ont été disposés dans les parcelles de la forêt domaniale d'Orléans, pour enregistrer des vidéos ou des photos, tout au long de l'année, pour détecter la présence d'animaux et étudier leurs impacts sur la forêt. Ces dispositifs de capture d'images produisent de grande quantité d'images qui sont compilées dans des vidéos dont l'exploitation était réalisée manuellement par des agents INRAE du site de Nogent-sur-Vernisson. Ces derniers étaient amenés à parcourir visuellement les vidéos (compilation de séquence d'images) dans leur intégralité, en analysant les différentes images, et à enregistrer manuellement des informations telles que les numéros des parcelles forestières sur lesquelles sont effectuée la détection d'animaux, la date de chaque vidéo, les plages temporelles de présence des animaux, l'espèce et la catégorie d'animal, leurs comportements dans une parcelle donnée (position fixe ou mobile). Il s'agit d'un travail fastidieux et contraignant à réaliser, pouvant conduire à des erreurs dans la détection des animaux. Jusqu'à présent, ces contraintes limitent les possibilités de suivi à des fins scientifiques mais aussi son application à grande échelle par les gestionnaires.

Un système de traitement automatique des vidéos afin de compter les animaux et identifier les espèces dans des programmes de suivi de la biodiversité pourrait faciliter et améliorer l'exploitation de ces données et rendre la méthodologie accessible à un plus grand nombre d'utilisateur. Pour répondre à ce besoin, un logiciel de vision artificielle, qui comprend des opérations de traitement d'images (fonctions de la librairie OpenCV et fonctions personnalisées) qui ont dû être complétées par une approche d'intelligence artificielle (Deep Learning) en utilisant un outil tel que TensorFlow, a été développé pour détecter automatiquement des animaux sur les vidéos enregistrées. L'innovation technologique réside dans la combinaison de ces deux techniques pour pouvoir prendre en compte des contraintes qui ont une influence importante sur les opérations de traitement d'image, telles que les conditions de luminosité très variables du sous-bois selon les lieux au cours de la journée qui entraînent des fortes variations au niveau de la couleur des objets naturels (arbres, herbes, animaux), ou encore les mouvements des herbes dus au vent qui rajoutent du bruit dans les images issues des vidéos.

Les intérêts de ce logiciel sont multiples et évitent des opérations manuelles et visuelles fastidieuses : la détection des animaux est réalisée de façon automatique, le traitement des vidéos permet de conserver en mémoire uniquement des images contenant des animaux, au lieu de garder beaucoup de fichiers volumineux en mémoire, et l'enregistrement des résultats de détection d'animaux est effectué automatiquement dans un fichier « résultats de détection d'animaux » pour un ensemble de vidéos traitées.

3 Traduction du mot anglais « camera trap »

4 <https://optmix.inrae.fr/>; OPTMix (Oak Pine Tree Mixture) est un dispositif expérimental constitué d'un réseau de 33 placettes instrumentées en forêt domaniale d'Orléans sur une surface totale de 40 ha. L'objectif est d'étudier les conséquences de la composition (chêne sessile, pin sylvestre, mélange), de la densité (nombre d'arbres/ha), et de la présence d'ongulés sauvages (sangliers, cerfs, chevreuils), dans les peuplements forestiers sur le fonctionnement de la forêt (bilan hydrique, régénération, biodiversité, croissance du peuplement).

Matériel et méthode

Le dispositif d'acquisition de vidéos

Des caméras ont été disposées sur des arbres dans différentes parcelles forestières (Figure 1 et 2), dans la forêt domaniale d'Orléans à proximité de l'unité de recherche EFNO d'INRAE à Nogent-sur-Vernisson.



Figure 1. La caméra GameSpy modèle M-80XT de la marque Moultrie disposée dans son boîtier de protection sur un arbre à 2 mètres de hauteur.



Figure 2. Les parcelles forestières

Le suivi est réalisé sur 12 placettes de forme rectangulaire d'une superficie de 0.5 ha au cours de trois périodes de l'année : février-mars, mai-juin et novembre-décembre. Pendant ces périodes, les données enregistrées par les caméras sur une carte mémoire sont relevées par un opérateur une fois par semaine. L'acquisition automatisée des données a lieu pendant six jours sur les sept jours de la semaine, le 1^{er} jour étant nécessaire à l'appareil pour s'étalonner au lever et coucher du soleil. Pour chaque placette, le dispositif de mesure consiste en quatre caméras vidéo positionnées de façon à ce que la prise de vue soit réalisée depuis ses quatre coins en direction de son centre. Pour cela, chacune des caméras est fixée sur l'arbre le plus proche de chacun des coins, et orientée après avoir matérialisé les diagonales de la parcelle. Elles sont placées à 2 mètres de hauteur, avec une orientation vers le bas de quelques degrés, qui permet d'obtenir une bonne visualisation des parcelles pour la détection des animaux. Cette hauteur "inaccessible" par les animaux, permet également d'éviter que ces derniers détériorent les caméras et de rendre le vandalisme plus difficile (par exemple le vol d'appareils).

Lors de la mise en place, l'étalonnage de la profondeur et de l'orientation du champ de l'appareil est vérifié à chaque pose grâce à un jalon porte-étiquette jaune de 65 cm² et de 30 cm de hauteur posé à 27 mètres de l'arbre porteur (utilisation de la visionneuse portable afin d'affiner le réglage).

La caméra est réglée en mode hybride soit en 2 modes simultanément :

- mode automatique toutes les minutes pendant 4 heures à partir du lever et 4 heures jusqu'au coucher du soleil, soit 240 photographies par périodes de 4 heures de jour;
- mode détection de la présence d'un animal (courte distance de détection 13.7 m +/- 1.5 m donnée constructeur) pendant les 8 heures du mode automatique et pendant la nuit.

En mode détection, l'appareil est réglé pour prendre deux photos successives dans un délai de 3-4 secondes, puis dans un délai de 5 secondes s'il détecte à nouveau un mouvement. Les photos issues du mode automatique sont compilées bout à bout dans une séquence vidéo.

Les images traitées par le logiciel de vision artificielle, extraites des vidéos, sont dans l'espace couleur visible classique (Rouge, Vert, Bleu) et ont une résolution de 1280 x 720 pixels.

Les équipements suivants sont utilisés pour effectuer une campagne de suivi :

- 12 appareils photo numérique GameSpy modèle M-80XT de la marque Moultrie ;
- 1 appareil photo supplémentaire pour effectuer les réglages ;
- 12 caissons de sécurité avec fixation inclinable et anti-vol ;
- 48 fixations sanglées sur les arbres porteurs (12 placettes X 4) ;
- 24 cartes mémoires sd 8Gb (12 appareils X 2) ;
- piles rechargeables AA 4 2450mAh (12 X 8 par appareils X 2, + 8) ;
- 3 chargeurs de capacité 8 piles en simultané ;
- 1 visionneuse de photographies avec sa carte sd.

Le logiciel de vision et d'intelligence artificielle

Deux grandes étapes sont utilisées dans le logiciel développé en langage Python. Des fonctions de traitement d'images (fonctions de la librairie OpenCV et fonctions personnalisées) et utilisation d'outils d'intelligence artificielle (Deep Learning / TensorFlow).

La partie traitement d'image qui comprend diverses fonctions de filtrage, de seuillage, de morphologie mathématique, dans l'espace couleur RGB, permet de comparer sur le plan colorimétrique, des images successives sur les vidéos et ainsi de détecter des variations de couleurs sur tous les pixels des images. Ces différences de couleurs peuvent être soit des éléments qui apparaissent dans la scène (détection d'un animal qui apparaît ou qui est en mouvement), soit des mouvements d'herbes, soit des changements de couleurs des objets dues aux variations de luminosité. On obtient alors, à l'issue de ces opérations de traitements d'image, des groupements de points séparés sur les images. Pour chacun d'entre eux, une zone rectangulaire va être définie. On obtient alors des images de tailles réduites, de différentes dimensions appelées "imassettes", sur lesquelles on va travailler dans une phase d'intelligence artificielle (Deep Learning). A l'intérieur de celles-ci, on peut trouver soit un animal complet (ou une partie physique d'un animal (la tête, le corps, les pattes,...)), soit un élément de l'environnement forestier (arbre, branche, herbe,...) qui correspond à du bruit, que l'on appelle "faux positifs". La quantité d'imassettes obtenues correspondant à des "faux positifs", sur une image donnée, qui dépend de l'environnement (quantité d'herbe) et des conditions climatiques (vent, variation de luminosité), varie approximativement entre 5 et 80.

La partie Deep Learning (outil TensorFlow) qui prend en compte une base d'images d'apprentissage importante contenant des animaux (chevreuil, cerf élaphe, sanglier, pigeon,...) et d'autres types d'objets (bruits) de différentes natures (arbres, herbes, végétaux,...) va permettre d'identifier chacune des "imassettes", en indiquant si un animal y est présent ou non, pour détecter les animaux et d'éliminer les "faux positifs". On obtient, pour chaque imasette, un résultat de type: objet détecté (animal ou autre type) avec un score (probabilité d'appartenance).

A titre d'exemple, la figure 3 ci-dessous présente un exemple de réseau de neurones, qui est constitué d'une couche d'entrée (input layer), qui peut contenir des données de diverses natures (image, signature sonore, données physiques, données de météorologie,...), d'un ensemble de couches cachées (hidden layers), et d'une couche de sortie (output layer) qui présente le résultat de prédiction du réseau de neurones, par exemple un couple de données (type d'animal détecté (chien, chat, chevreuil,...), avec un score d'appartenance entre 0 et 1).

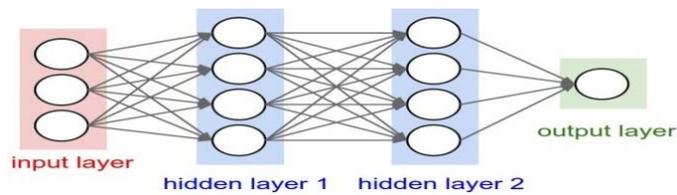


Figure 3. Exemple de réseau de neurones

La figure 4 montre un exemple de résultat : imageries obtenues par vision artificielle (rectangles verts) (a) et résultat de détection d'un animal obtenu avec la technique d'intelligence artificielle (b).



a) Détection d'objets entre 2 images successives

b) Détection d'un animal parmi les objets

Figure 4. Exemple de résultat de détection d'un animal

Description détaillée de la méthode

La figure 5 ci-dessous présente deux images successives, extraites d'une vidéo.

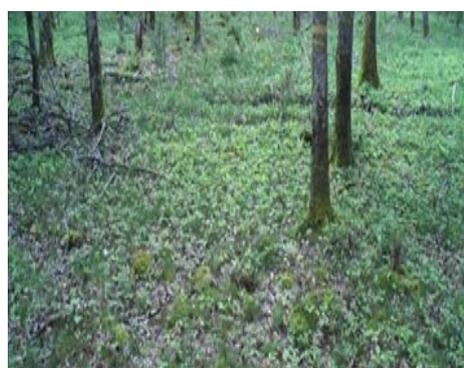
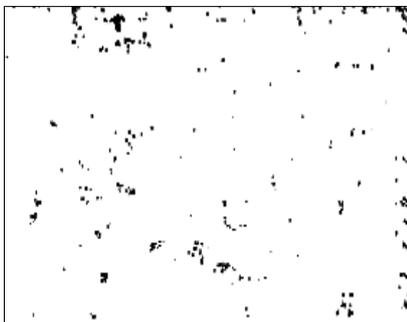


Figure 5. Deux images successives extraites d'une vidéo (images aux instants t et t+1)

Les opérations de traitement d'image effectuées, pour obtenir à la fin des imageries pouvant contenir des "parties d'animaux" (animaux complets, tête, queue, pattes, pieds,...) qui sont en général de "couleur sombre" (marron, noir) recherchés en milieu forestier" sont les suivantes :

- 1) Séparation des plans R, G et B pour les deux images : fonction Split (Red,Green,Blue)
- 2) Soustraction de tous les pixels (i,j) dans les trois plans, pour obtenir trois images :
 $ImageDifB(i,j)=subtract (B_{t+1}, B_t)$ $ImageDifG(i,j)=subtract (G_{t+1}, G_t)$ $ImageDifR(i,j)=subtract (R_{t+1}, R_t)$
- 3) Critère de détection des zones de l'image pouvant contenir un animal sur l'image t+1 (pour tous les pixels (i,j) de l'image) avec la définition de seuils (S1, S2, S3, S4 et S5) ;
Un point P(i,j) est considéré comme un point "animal" s'il vérifie le critère ci-dessous :
 $((ImageDifB(i,j) > S1 \text{ et } ImageDifG(i,j) > S2 \text{ et } ImageDifR(i,j) > S3 \text{ et } (R_{t+1}(i,j) > G_{t+1}(i,j)+S4) \text{ et } (R_{t+1}(i,j) > B_{t+1}(i,j)+S5))$
ou $(B_{t+1}(i,j) > R_{t+1}(i,j) \text{ et } B_{t+1}(i,j) > G_{t+1}(i,j))$
- 4) Opérations de traitement d'image : morphologie, labélisation, composants connectés, contours sur l'image t+1 , pour éliminer le bruit sur les images et regrouper les ensembles de points appartenant aux mêmes objets, en prenant en compte automatiquement, la luminosité de l'environnement et la présence d'animaux dans différentes zones de l'image (premier plan, plan moyen, arrière-plan).

La figure 6 présente le résultat de traitement d'images obtenu à partir des deux images précédentes acquises aux instant t et t+1 (figure 5), pour obtenir des imagerie, de différentes dimensions.



a) Opérations de traitement d'image



b) Obtention d'imagerie

Figure 6. Résultat de traitement d'image

La méthode d'intelligence artificielle

La méthode d'intelligence artificielle utilisée pour détecter et identifier des animaux sur des images, consiste à s'intéresser à deux opérations: la base d'apprentissage de données et le choix du type de réseau de neurones.

La base d'apprentissage doit contenir des images d'animaux (objets recherchés), mais aussi des images de bruits dans l'environnement forestier (branches, arbres, feuilles, ciel,...). Le logiciel va donc pour une imagerie donnée, dire si l'objet détecté est de type animal ou bruit. Etant donné que des animaux peuvent être visibles que partiellement sur des images, il est nécessaire de considérer, pour la base d'images d'apprentissage, des animaux complets, mais aussi des parties d'animaux (tête, pieds, corps,...) en considérant aussi différentes orientations des animaux dans l'environnement.

On peut soit utiliser des bases d'animaux existantes, notamment la base ImageNet de FaceBook qui contient un nombre très important d'images de différentes natures (animaux, voitures, objets urbains, naturels,...), soit créer sa propre base d'images d'apprentissage, en créant des dossiers contenant des images d'animaux et de bruits que l'on peut trouver dans les vidéos traitées (images d'arbres, de feuillage, de cerf, de pigeon, de sanglier, de chevreuil,...), soit faire un mélange des deux types de bases. C'est ce qui a été réalisé dans notre application de détection d'animaux en forêt, pour obtenir un nombre important d'images (animaux forestiers et bruits) permettant de créer un réseau de neurones fiable pour notre application.

Deux types de bases d'apprentissage ont été réalisés : une base avec deux classes (animal et bruit) qui permet de façon binaire de dire si une imagerie contient ou non un animal, et une base avec 5 classes (Arbre, Feuillage, Chevreuil, Cerf élaphe, Sanglier).

Concernant le réseau de neurones, il existe différents types de réseaux : notamment la série Inception V1, V2, V3, la série Resnet ou la série MobileNet avec différents paramètres à régler notamment pour définir le format des images et le nombre de couches cachées désirées. La figure 7 ci-dessous présente un exemple de réseau de neurones. La création du réseau est réalisée avec une opération d'apprentissage qui consiste à disposer dans la couche d'entrée, des images correspondant aux différentes bases d'images et en indiquant en sortie, le type d'objet associé à chaque image d'entrée. Une méthode de rétropropagation de gradient va être effectuée pour créer le réseau, c'est-à-dire pour déterminer les valeurs des poids (weights) (notés W_{mn} en général), entre les différents neurones m et n qui communiquent entre eux. A partir du réseau créé, l'opération de prédiction va alors consister à envoyer en entrée du réseau une imagerie et à obtenir à la sortie des types d'objets détectés, avec des scores. Par exemple le réseau trouvera pour une image d'entrée, différents objets résultats (en fonction des classes d'apprentissage) : Chevreuil avec un score de 0.51, Arbre avec un score de 0,2, Sanglier avec un score de 0.09.

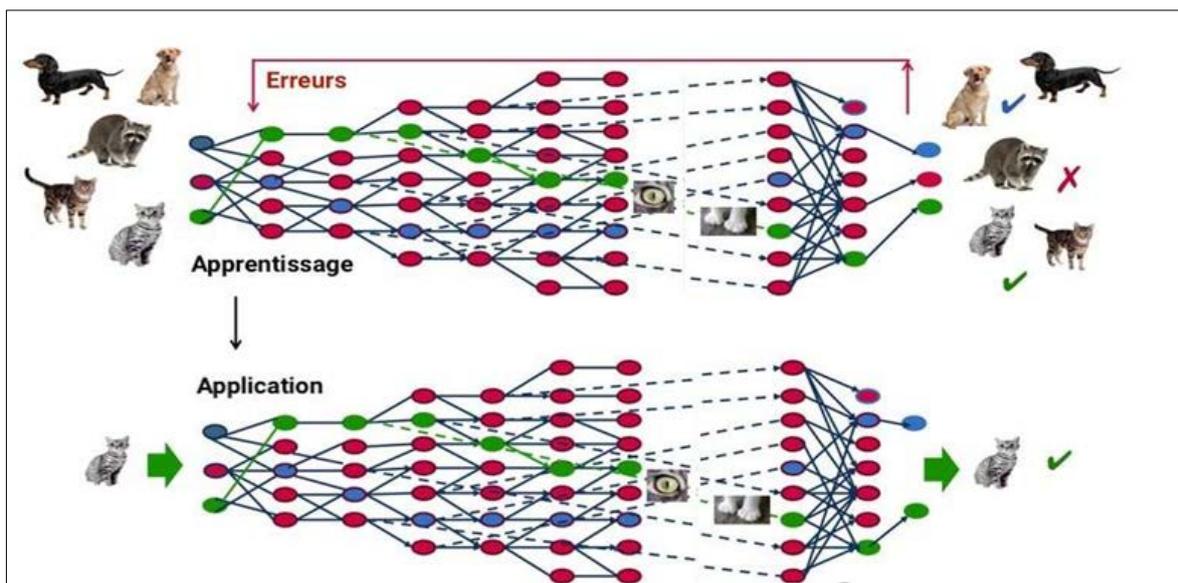


Figure 7. Principe de création d'un réseau de neurones (Apprentissage et Prédiction)

La figure 8 montre le résultat final de détection d'un animal : les quatre imageries de l'image (a) ont été envoyées dans la fonction d'intelligence artificielle (TensorFlow) et parmi ces imageries, l'une d'entre elle contenait un animal de type "ibex" (bouquetin) (détecté avec un score de 0,28) (image (b)).



(a) Imageries



(b) Détection d'un animal

Figure 8. Résultat de détection d'un animal

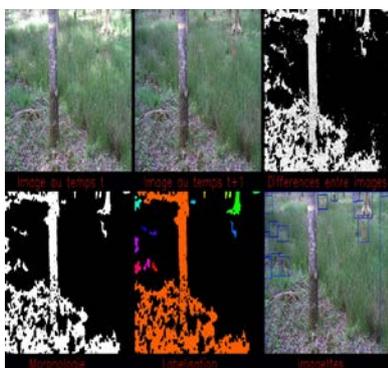
Résultats

Le logiciel développé a été appliqué sur un ensemble de vidéos acquises dans des parcelles de la forêt domaniale d'Orléans, acquises à différents moments de la journée, dans différentes périodes de l'année et au cours de différentes

années. Concernant l'opération d'intelligence artificielle, neuf types de réseaux de neurones et deux types de bases d'apprentissage ont été testés, dans le but d'obtenir le réseau le plus optimal en terme de détection et d'identification des animaux. La figure 9 ci-dessous présente des exemples de résultats de détection d'animaux, avec les différentes étapes (traitement d'image, obtention des imagerie, détection par Deep Learning (type d'animal et score d'appartenance)), en utilisant nos réseaux personnalisés. La figure 10 montre d'autres exemples de résultats de détection, obtenues en utilisant le réseau de neurones Inception V3, préentraîné sur la base d'images ImageNet de facebook (une base internationale d'images qui contient beaucoup d'espèces d'animaux, mais pas nécessairement les espèces animales observables en France). Concernant le travail avec cette base d'images ImageNet, on notera que les animaux que nous rencontrons dans la forêt de Nogent et qui ont été détectés, ont été classés dans des animaux de type sanglier et de renard, qui sont présents sur le site, mais aussi de type bouquetin, poule d'eau, dhole qui ne sont pas présents sur le site. Ce réseau Inception V3 préentraîné donne de très bons résultats de détection d'animaux (mais pas d'identification d'espèce). Cependant, il est préférable de travailler avec son propre réseau et sa propre base d'images d'apprentissage, pour pouvoir adapter le mieux possible le réseau de neurones créé, à sa problématique de détection (perfectionner la détection d'animaux et rendre possible l'identification des espèces). Par exemple, si l'on souhaite détecter et identifier avec précision des types d'oiseaux, d'insectes, de poissons, ... il est conseillé de restreindre sa base d'images à des types d'animaux bien précis : seulement des oiseaux classés dans différentes catégories, ou bien seulement des insectes, ou encore seulement des poissons. Si les images sont de très bonne qualité, on pourrait également envisager pour certaines espèces, de classer les animaux en fonction du sexe et de l'âge.

Un travail comparatif entre l'analyse visuelle des vidéos et les résultats obtenus avec le logiciel développé a été réalisé pour évaluer la qualité de détection des animaux. Sur le plan général, le logiciel détecte correctement les animaux, les différents bruits dans l'environnement forestier (mouvements d'herbes, variations de luminosité,...) sont bien éliminés grâce à l'opération d'intelligence artificielle. Le taux de réussite obtenu avec ce logiciel de détection et d'identification des animaux, pour l'ensemble des vidéos testées (environ 200) est de 90 %. Parmi les échecs, on notera que des animaux qui apparaissaient dans la scène, dans l'arrière-plan, soit de petite taille, soit de façon incomplète (seule une partie du corps était visible), relativement difficile à détecter par la vision humaine, étaient difficilement détectables, par le logiciel développé, malgré la définition de seuil de surface animalière, pour le premier plan, le plan moyen et l'arrière-plan. D'autre part, quelques imagerie contenant du bruit "faux positifs", ont été détectés comme des animaux.

Par exemple, des arbres qui peuvent apparaître dans la partie traitement d'image du logiciel, à cause des variations de luminosité entre des images successives, pouvaient être détectés comme des animaux, car leur couleur est sombre (marron/noir), comme nos animaux recherchés. Dans le logiciel, l'objectif recherché est de privilégier la détection d'animaux, dans les vidéos, même si l'on détecte du bruit (faux positifs) comme des animaux. Les agents utilisateurs du logiciel sont amenés, en parcourant les résultats d'images détectées, à supprimer les images qui ne contiennent pas d'animal.



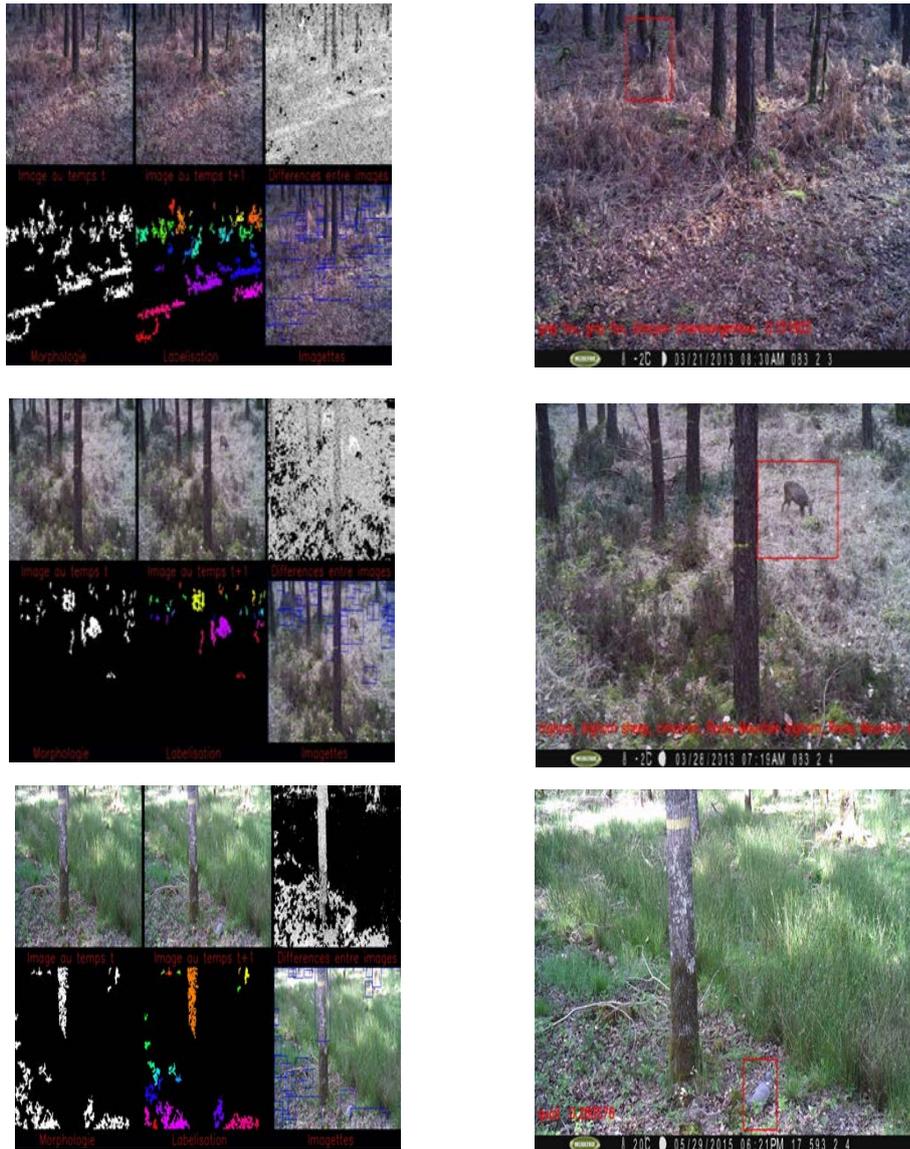


Figure 9. Résultats détaillés de détection d'animaux



Type : sanglier score = 0.61



type : dhole score = 0.8



type : bouquetin score = 0.82



Type : renard score = 0.74

type : poule d'eau score = 0.61

type : dhole score = 0.32

Figure 10. Résultats de détection d'animaux

Sur le plan temporel, une vidéo donnée contient environ 240 images, qui couvrent une période de 4 heures. La caméra acquiert une image chaque minute. Le logiciel développé, en langage Python, sur un PC portable standard (4 cœurs), avec un processeur intel core i7 de 3,8 Ghz, et une carte graphique classique, met environ 5 minutes pour traiter une vidéo.

Les images d'animaux enregistrées sont conservées avec un archivage bien défini. Elles ont le format suivant : **nomdossier_nomvideo_numimage_typeanimal_score.jpg**. D'autre part, un fichier texte enregistre automatiquement tous les résultats de détection, sur l'ensemble des vidéos, pour les différentes parcelles traitées, au cours des différentes périodes de l'année. A partir de ces éléments, les agents du site INRAE de Nogent peuvent réaliser des comparaisons entre les parcelles, entre les périodes de l'année, pour étudier notamment, de façon approfondie, l'influence des animaux sur la dégradation des jeunes pousses en milieu forestier.

Conclusion et Perspectives

Une méthode de vision artificielle a été développée pour détecter des animaux qui apparaissent sur des vidéos. Compte tenu du bruit plus ou moins important sur les images (variation de luminosité, mouvements d'herbes), l'algorithme de traitement d'image développé, basé sur la différenciation d'images successives, n'était pas suffisant pour détecter les animaux (beaucoup de faux positifs étaient détectés). Une méthode d'intelligence artificielle (Deep Learning / TensorFlow) a donc dû être ajoutée, en complément de la méthode de vision artificielle, pour éliminer les faux positifs et ne détecter que les animaux sur les images extraites des vidéos. Les très bons résultats de détection d'animaux obtenus avec ce logiciel de vision vont permettre aux agents de Nogent de l'utiliser de façon permanente, pour les animaux forestiers et d'envisager d'autres types d'utilisations : par exemple, la détection et l'identification d'autres espèces animales, en milieu forestier, ou la comparaison/différenciation de diverses cultures,...

Le type de logiciel développé qui combine la vision artificielle et l'intelligence artificielle, qui peut fonctionner soit en utilisant, par exemple, un réseau de neurones de type Inception V3, préentraîné sur une base de type ImageNet de facebook, soit en créant un réseau de neurones personnalisé, entraîné sur une base d'apprentissage personnalisée contenant des images d'animaux que l'on cherche à détecter et identifier (cerf élaphe, chevreuil, sanglier,...) en milieu forestier, pourrait être appliqué pour détecter d'autres types d'animaux (oiseaux, poissons, insectes,...) pour différentes applications de détection ou de suivi d'animaux, dans des milieux forestiers, agricoles, aquatiques,... en s'intéressant au côté visuel (images), mais aussi sonore (identification des animaux à partir de leur signature sonore), en créant un réseau de neurone avec des données sonores.

Pour des applications générales de détection d'objets mobiles dans un environnement donné, sur le plan matériel et logiciel, tout type de caméra/camescope peut être utilisé pour acquérir des bandes vidéos ou des ensembles d'images. Un ordinateur standard sur lequel est installé le module de vision de type OpenCV et l'outil d'intelligence artificielle de type

TensorFlow, va pouvoir être utilisé pour lancer le logiciel de détection d'objets développés (animaux...), à partir des vidéos/images enregistrées. Dans l'application présentée dans cet article, le logiciel travaille en mode différé, en traitant des vidéos enregistrées, en milieu forestier, il y a quelques temps. Mais ce logiciel pourrait aussi être utilisé en mode temps réel : acquisition d'images avec une caméra et détection d'animaux qui apparaissent dans la scène, en mode "live". Dans ce cas, des dispositifs complets (systèmes d'acquisition d'images + ordinateurs) devraient être installés dans des parcelles forestières.

Cet article est publié sous la licence Creative Commons (CC BY-SA).



<https://creativecommons.org/licenses/by-sa/4.0/>

Pour la citation et la reproduction de cet article, mentionner obligatoirement le titre de l'article, le nom de tous les auteurs, la mention de sa publication dans la revue « Le Cahier des Techniques de l'Inra », la date de sa publication et son URL).

Bibliographie (en savoir plus)

- Norouzzadeh, M.S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M.S., Packer, C., Clune, J. 2018. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences* 115, E5716-E5725. <https://doi.org/10.1073/pnas.1719367115>
- Tabak, M.A., Norouzzadeh, M.S., Wolfson, D.W., Sweeney, S.J., Vercauteren, K.C., Snow, N.P., Halseth, J.M., Di Salvo, P.A., Lewis, J.S., White, M.D., Teton, B., Beasley, J.C., Schlichting, P.E., Boughton, R.K., Wight, B., Newkirk, E.S., Ivan, J.S., Odell, E.A., Brook, R.K., Lukacs, P.M., Moeller, A.K., Mandeville, E.G., Clune, J., Miller, R.S., 2019. Machine learning to classify animal species in camera trap images: Applications in ecology. *Methods Ecol. Evol.* 10, 585-590. <https://doi.org/10.1111/2041-210x.13120>
- Wäldchen, J., Mäder, P., 2018. Machine learning for image based species identification. *Methods Ecol. Evol.* 9, 2216-2225. <https://doi.org/10.1111/2041-210x.13075>
- Yousif, H., Yuan, J., Kays, R., He, Z., 2019. Animal Scanner: Software for classifying humans, animals, and empty frames in camera trap images. *Ecology and Evolution* 9, 1578-1589. <https://doi.org/10.1002/ece3.4747>
- Yu, X., Wang, J., Kays, R., Jansen, P.A., Wang, T., Huang, T., 2013. Automated identification of animal species in camera trap images. *EURASIP Journal on Image and Video Processing* 2013, 52. <https://doi.org/10.1186/1687-5281-2013-52>