



HAL
open science

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

Bernard Benet, Anders Mårell, Yves Boscardin

► **To cite this version:**

Bernard Benet, Anders Mårell, Yves Boscardin. Utilisation de techniques d'intelligence artificielle pour la détection d'animaux en milieu forestier. APIA2021, Jul 2021, Bordeaux, France. hal-03874232

HAL Id: hal-03874232

<https://hal.science/hal-03874232>

Submitted on 27 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

B. Benet¹, A. Marell², Y. Boscardin²

¹Université Clermont-Auvergne, INRAE, UR TSCF, 63178 Aubière, France

² INRAE, UR EFNO, 45290, Nogent-sur-Vernisson, France

bernard.benet@inrae.fr

Résumé

Un logiciel de vision artificielle qui comprend des opérations de traitement d'images et des méthodes d'intelligence artificielle (deux réseaux de neurones intégrés développés avec les outils Tensorflow et Caffe), a été développé pour détecter et identifier des espèces animales en milieu forestier sur des vidéos enregistrées, en prenant en compte des fortes contraintes au niveau de l'environnement, telles que les conditions d'éclairage très variables dans le sous-bois, les variations de colorimétrie et de morphologie des animaux, qui peuvent être partiellement cachés par des objets naturels (arbres, herbes) et des conditions météorologiques variables (pluie, neige, vent).

Mots-clés

Traitement d'image, intelligence artificielle, forêt, biodiversité, animaux

Abstract

An artificial vision software which includes image processing operations and deep learning methods (two integrated neural networks developed with Tensorflow and Caffe tools), has been developed to detect and identify animal species in forest environments on recorded videos, taking into account strong environmental constraints, such as highly variable lighting conditions in the undergrowth, variations in colorimetry and animal morphology, which may be partially hidden by natural objects (trees, grasses) and variable weather conditions (rain, snow, wind).

Keywords

Image processing, deep learning, forest, biodiversity, animals

1 Introduction

Les scientifiques et les naturalistes sont nombreux à utiliser des dispositifs de capture d'images, appelés aussi « pièges photographiques » pour inventorier et suivre les populations animales et la biodiversité de la faune sauvage [1][2][3][4]. 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 de recherche OPTMix, mené par l'INRAE sur le site 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. Elles ont été compilées dans des vidéos dont l'exploitation était réalisée, il y a quelque temps, manuellement, par des agents de l'INRAE qui étaient amenés à parcourir visuellement ces vidéos, 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'agissait d'un travail fastidieux et contraignant à réaliser, pouvant conduire à des erreurs dans la détection des animaux. Ces contraintes de traitement manuel limitaient les possibilités de suivi à des fins scientifiques mais aussi son application à grande échelle par les gestionnaires.

L'objectif était donc de développer un système de traitement automatique des vidéos afin de compter les animaux et identifier les espèces dans des programmes de suivi de la biodiversité pour faciliter et améliorer l'exploitation de ces données et rendre la méthodologie accessible à un plus grand nombre d'utilisateurs. Un logiciel de vision artificielle, qui comprend des opérations de traitement d'images (fonctions de la librairie OpenCV et fonctions personnalisées) qui ont été complétées par des méthodes d'intelligence artificielle [5] en utilisant des outils tels que TensorFlow et Caffe, a été développé pour détecter et identifier automatiquement des

animaux sur les vidéos enregistrées. L'innovation technologique réside dans la combinaison de ces deux techniques pour pouvoir développer un système de détection des animaux, en temps réel, en prenant en compte des contraintes qui ont une influence importante, aussi bien pour la partie traitement d'image que pour les opérations d'apprentissage et de prédiction de la partie intelligence artificielle du logiciel : les conditions de luminosité sont très variables dans le sous-bois selon les lieux au cours de la journée, avec des problèmes d'ombres relativement important, les conditions météorologiques sont variables (pluie, neige, vent), les variations de colorimétrie et de morphologie des objets naturels (arbres, herbes, animaux) sont importantes, les animaux recherchés peuvent apparaître dans les différents plans de l'image, et peuvent être cachés partiellement par des objets naturels de type arbres, végétation, dans l'environnement forestier. 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 pour un ensemble de vidéos traitées.

2 Matériel et méthode

2.1 Le dispositif d'acquisition vidéo

Des dispositifs d'acquisition d'image ont été positionnés sur des arbres dans différentes parcelles forestières (Figures 1 et 2), dans la forêt domaniale d'Orléans. Ces systèmes sont des appareils photo numérique nommés « GameSpy modèle M-80XT » de la marque Moultrie.



Figure 1: Le dispositif d'acquisition d'image



Figure 2: Les parcelles forestières

Dans les travaux menés, 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 à utiliser 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 d'éviter que ces derniers détériorent les caméras. La caméra acquiert des images toutes les minutes pendant 4 heures à partir du lever et 4 heures jusqu'au coucher du soleil, soit 240 photographies par vidéo. Ces photos sont compilées bout à bout dans une séquence vidéo.

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

2.2 Le logiciel de vision et d'intelligence artificielle

Deux 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 l'utilisation de techniques d'intelligence artificielle, avec le développement de réseaux de neurones avec des outils tels que TensorFlow et Caffe, composent ce logiciel. Le premier objectif du travail réalisé, était de détecter et d'identifier les animaux forestiers, sur les images extraites des vidéos, en travaillant, avec une fréquence de 1Hz. A terme, l'objectif final sera d'implémenter le logiciel développé sur des systèmes embarqués temps réel, en milieu forestier, disposés sur différents arbres, qui pourront être composés d'ordinateur monocarte de type Raspberry pi ou Jetson Nano, et de caméras de type webcam.

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 acquises aux instants t et $t+1$ 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. Le logiciel cherchant à détecter des animaux sur l'image acquise à l'instant $t+1$, on va appliquer, en complément de cette phase de différenciation d'images, un seuillage colorimétrique, dans l'espace couleur HSV (teinte, saturation, intensité lumineuse) sur l'image à l'instant $t+1$, pour éliminer les pixels qui ne vérifient pas des conditions colorimétriques désirées pour les animaux à détecter. Par exemple, les pixels de teinte verte, jaune, rouge sont éliminés. On obtient alors, à l'issue de ces opérations de traitements d'image (différenciation entre deux images successives (t et $t+1$) et élimination des pixels ne vérifiant pas la couleur recherchée des animaux sur l'image à l'instant $t+1$), des groupements de points séparés (labélisés) 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 pour la détection des animaux. 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 positif". La quantité d'imassettes obtenues sur une image donnée, qui dépend de l'environnement (quantité d'herbe) et des conditions climatiques (vent, variation de luminosité, ombres portés), varie approximativement entre 5 et 80. La partie intelligence artificielle 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). La figure 3 ci-dessous présente un exemple avec l'obtention d'imassettes, de différentes tailles, obtenues avec les opérations de traitement d'image, et la figure 4 montre le résultat de détection d'un animal obtenu, avec les opérations d'intelligence artificielle.

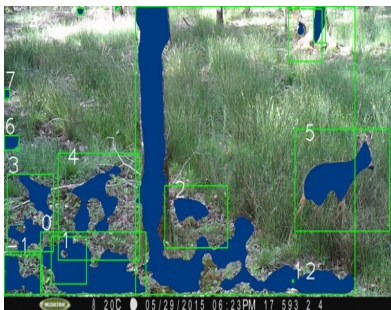


Figure 3: Détection d'imassettes par traitement d'image



Figure 4: Détection d'un animal

2.3 Les opérations d'intelligence artificielle

La figure 5 ci-dessous présente un exemple de type de réseau de neurones utilisé, qui est constitué d'une couche d'entrée, dans laquelle on va disposer nos données (imassettes couleurs RGB de différentes dimensions, pouvant contenir un animal (ou partie animalière) ou du bruit), d'un ensemble de couches cachées et d'une couche de sortie, qui présente le résultat de prédiction obtenu : type d'animal détecté (ou bruit), avec un score d'appartenance entre 0 et 1.

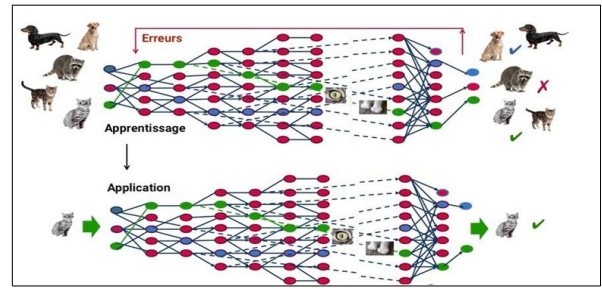


Figure 5 : Type de réseau de neurones utilisé

Une méthode de rétro-propagation de gradient est effectuée pour créer le réseau, dans la phase d'apprentissage, c'est-à-dire pour déterminer les valeurs des poids des liens entre les différents neurones du réseau qui communiquent entre eux. A partir du réseau créé, l'opération de prédiction va consister à envoyer en entrée du réseau une « imasette » et à obtenir à la sortie des types d'objets détectés, avec des scores d'appartenance. Par exemple dans notre application, 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,...

Deux opérations d'intelligence artificielle, avec deux réseaux de neurones utilisés, sont appliquées pour obtenir les résultats de détection/identification d'animaux, avec l'objectif de prendre en compte la qualité de détection mais aussi l'aspect temporel, pour pouvoir à terme, développer un système de détection en temps réel et transmettre les informations de détection en « live ».

Les imassettes obtenues par traitement d'image, dans la première étape du logiciel, sont envoyées dans un premier temps, sur un réseau de neurone de type GoogleNet (réseau de neurones à convolution avec 22 couches qui est une variante du réseau Inception), entraîné sur une base d'image d'apprentissage contenant une importante base d'objets de différentes natures dont 500 classes d'animaux), utilisé avec l'outil Caffe, pour effectuer un filtrage rapide des imassettes. Cette opération permet d'éliminer une grande partie des imassettes (environ 60%), qui contiennent du bruit (arbres, végétation,...) et de garder en mémoire des imassettes pouvant contenir un animal. On ne recherche pas, à ce stade à identifier les animaux trouvés. Ces imassettes restantes sont alors envoyées sur un deuxième réseau de neurone de type Inception V3, développé et utilisé avec l'outil tensorflow, qui a été réentraîné sur la dernière couche, avec une base d'images d'apprentissage personnalisée contenant des animaux forestiers (chevreuil, biche, pigeon, cerf elaphe, sanglier,...), et des éléments de l'environnement forestier (arbres, feuilles, branches, herbes,...). Les figures 6 et 7 présentent des exemples d'images d'apprentissage. 3000 images d'animaux forestiers et de bruit, ont été introduites dans la base d'apprentissage pour obtenir ce deuxième réseau de neurones.

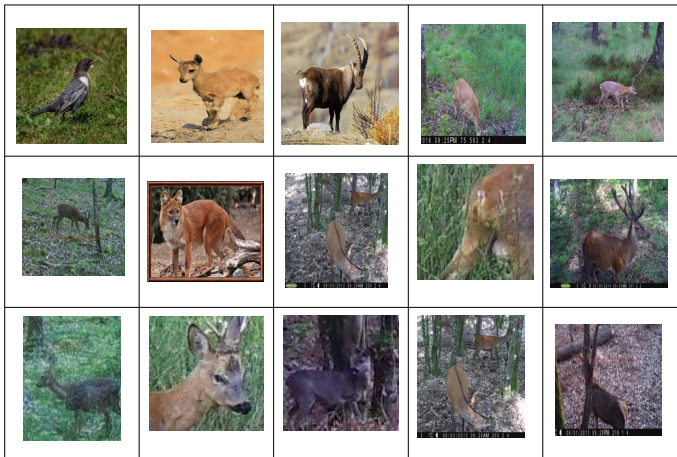


Figure 6: Exemples d'images d'animaux



Figure 7: Exemples d'images de l'environnement forestier

Etant donné que des animaux peuvent être visibles soit en totalité soit partiellement sur les images, car il peuvent être masqués par des arbres, par des branches, ou par de la végétation, il était nécessaire de considérer pour la base d'apprentissage animale, des animaux complets, en considérant différentes faces, orientations, postures et dimensions car les animaux peuvent être présents dans les différents plans de l'image, mais aussi des parties d'animaux (tête, pieds, corps,...). Différents types de réseaux de neurones ont été testés tels que Inception V1, V2 et V3, et des réseaux Mobilenet, pour obtenir un réseau « fiable », au niveau de la qualité des résultats obtenus, en considérant différents jeux de données pour les phases de validation et de test des réseaux, et en faisant varier le nombre d'étapes d'apprentissage.

3 Résultats

Le logiciel développé a été appliqué sur 5000 vidéos acquises entre les années 2013 et 2018, dans des parcelles de la forêt domaniale d'Orléans, acquises à différents moments de la journée, dans différentes périodes de l'année, pour prendre en compte le côté saisonnier dans le suivi de la grande faune. 240 images ont été extraites de chaque vidéo. La figure 8 présente un exemple de résultats des opérations de traitement d'image du logiciel : dans un premier temps, comparaison d'images successives dans l'espace couleur RGB, acquises aux instant t et $t+1$, pour détecter des différences au niveau de tous les pixels des deux images, prise en compte des couleurs des pixels dans l'image à l'instant $t+1$,

opérations de filtrage et de morphologie pour éliminer du bruit, labellisation pour regrouper des pixels sur l'image et obtention des imagettes rectangulaires qui encadrent chaque groupe de points.

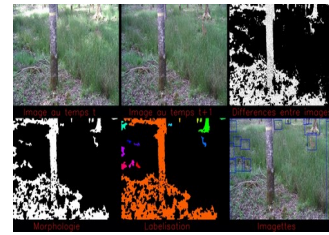


Figure 8: Opérations de traitement d'image

La figure 9 présente le résultat de détection d'un animal (gazelle), obtenue avec les deux réseaux de neurones, dans la phase d'intelligence artificielle.



Figure 9: Détection d'un animal

La figure 10 montre des exemples de résultats de détection obtenus.



Figure 10: Résultats de détection

Concernant la première opération d'intelligence artificielle, menée avec la base d'images GoogleNet, on notera que les animaux que nous rencontrons dans la forêt de Nogent et qui ont été détectés par le logiciel, ont été classés, dans la phase de prédiction, dans des animaux de type sanglier et renard, qui sont présents sur le site, mais aussi de type bouquetin, poule d'eau, dhole qui ne sont pas présents sur le site. Le deuxième réseau de neurones utilisé qui est le réseau Inception V3 qui a été pré-entraîné sur la dernière couche, à partir d'une base d'images d'animaux et de bruits dans l'environnement forestier, a permis de répondre à notre besoin, en matière de qualité de détection et d'identification des animaux forestiers.

Un travail portant sur 1140 vidéos (Nombre d'images traitées = 273600) a été réalisé, pour comparer les résultats obtenus par une analyse visuelle (expertise) avec ceux obtenus avec le logiciel développé et ainsi 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 aux deux opérations d'intelligence artificielle.

Une matrice de confusion a été établie avec le comptage des quatre éléments (Vrai Positif, Faux Positif, Vrai Négatif et Faux Négatif) sur l'ensemble des vidéos traitées, et a permis de déterminer les valeurs de sensibilité, de spécificité et de précision du logiciel (traitement d'image et intelligence artificielle). La définition de ces différents éléments est présentée ci-dessous :

- TP ('True positives' - Vrais positifs) : le nombre total d'images avec animaux (ongulés sauvages, cerf élaphe, chevreuil, sanglier) détectés par l'expertise et par le logiciel

- FP ('False positives' - Faux positifs) : le nombre total d'images, sur lesquelles le logiciel a détecté un animal mais qui ne contiennent aucun animal

- FN ('False negatives' - Faux négatifs) : le nombre total d'images contenant des animaux mais qui n'ont pas été détectés par le logiciel

- TN ('True negatives' - Vrais négatifs) : le nombre total d'images sans animaux (images identifiées sans animaux à la fois par l'expertise et par le logiciel).

Les paramètres de sensibilité, de spécificité et de précision sont définis par les formules suivantes :

-Sensibilité = $TP / (TP + FN)$

-Spécificité = $TN / (TN + FP)$

-Précision = $(TP + TN) / N$ (N est le nombre d'image traitées)

Le tableau 1 ci-dessous présente les données obtenus de la matrice de confusion et les valeurs caractérisant la qualité du logiciel.

		Détection manuelle	
		Positif	Négatif
Détection avec le logiciel	Positif	452	4367
	Négatif	216	272615
Sensibilité		67%	
Spécificité		98%	
Précision		99%	

Tableau 1. Matrice de confusion et qualité du logiciel

La spécificité du logiciel, qui mesure la capacité de l'analyse à donner un résultat négatif lorsque l'image est vide, s'avère élevée avec une valeur de 98 %. Ce qui veut dire que l'analyse permet d'éliminer les images sans animaux. La précision est aussi élevée 0.99. C'est à dire que l'analyse arrive à identifier les images avec des animaux et exclure les images sans animaux.

En revanche, la sensibilité, qui mesure la capacité de l'analyse à donner un résultat positif lorsque l'image contient des animaux, est de 67 %. Ce résultat de non détection de certains animaux sur des images peut s'expliquer par les points suivants :

- des animaux de petites tailles peuvent apparaître dans la scène, avec des couleurs sombres devant des objets naturels de couleurs sombres, par exemple des arbres, ce qui ne permet pas de les distinguer facilement dans l'environnement

- des animaux peuvent être partiellement cachés dans l'environnement forestier, par exemple derrière des arbres, des branches ou de la végétation, et leur visibilité est ainsi limitée sur les images

- des animaux peuvent présenter un profil/une face semblable à un élément de type arbre ou branche, par exemple. Dans ce cas, le logiciel, dans la partie intelligence artificielle, donne des résultats de prédiction de mauvaise qualité : détection d'un objet de type animal avec une probabilité inférieure à 5 %.

Concernant les faux positifs détectés, ils correspondent en général à des arbres ou à des branches, qui ont des couleurs (marron/noir) comme celle de la majorité des animaux forestiers, qui peuvent apparaître dans les images obtenues dans la partie traitement d'image du logiciel, à cause des variations de luminosité entre des images successives.

Sur le plan temporel, la durée moyenne de traitement des 240 images pour une vidéo donnée est de 4mn, sur un PC portable avec 4 cœurs, avec un processeur intel core i7 de 3,8 Ghz, et une carte graphique classique (non CUDA). Ce logiciel enregistre dans un dossier toutes les images détectées, sur un ensemble de vidéos, ce qui permet alors de 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, et d'étudier la biodiversité concernant la grande faune forestière.

4 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, environnement complexe), 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 a donc dû être ajoutée, en complément de la méthode de traitement d'image, pour éliminer une quantité importante de faux positifs et détecter les animaux sur les images extraites des vidéos. Les résultats de détection d'animaux obtenus avec ce logiciel de vision vont permettre aux agents de l'INRAE de Nogent de l'utiliser de façon permanente, pour les animaux forestiers et d'approfondir aussi le côté identification, pour des

animaux détectés qui apparaissent sur les images de façon complète (une grande partie de leurs corps est visible).

Des modifications devront être apportés dans le logiciel développé, pour améliorer la sensibilité de détection, pour pouvoir détecter, le mieux possible, des animaux non détecté, actuellement, en jouant sur des paramètres et des seuils de détection au niveau des deux réseaux de neurones utilisés, tout en limitant la détection de faux positifs.

Le principe général du logiciel développé qui combine la vision artificielle et l'intelligence artificielle, qui peut fonctionner soit en utilisant des réseaux de neurones préentraînés sur des bases de type ImageNet ou GoogleNet, soit en créant des réseaux de neurones personnalisés, entraînés sur des bases d'apprentissage personnalisées contenant des images d'objets que l'on cherche à détecter et identifier, soit en combinant ces deux méthodes, 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.

Dans l'application présentée dans cet article, le logiciel travaille en mode différé, en traitant des vidéos enregistrées, en milieu forestier, mais l'objectif est de développer un système de traitement en temps réel, en disposant des caméras et des ordinateurs de type monocartes, sur des arbres, dans différentes parcelles forestières et de communiquer les résultats obtenus, sur un ordinateur central, pour enregistrer les détections des animaux et prendre éventuellement des décisions en temps réel, de type alerte.

5 Biblio

[1] 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>

[2] 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>

[3] 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>

[4] 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>

[5] 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>