Reçu : 11 avril 2022 | Accepté : 22 mars 2023

DOI : 10.1111/1365-2656.13932

Journal of Animal Ecology BRITISH ECOLOGICAL SOCIETY

RESPONSABILITÉ DE L'ARTICLE

Méthode contemporaine pour la recherche sur la société animale dans le Wild

# DeepWild : Application de l'outil d'estimation de la pose DeepLabCut pour le suivi du comportement des chimpanzés et des bonobos sauvages

**Charlotte Wiltshire**[1] | **James Lewis-Cheetham**[1] | **Viola Komedová**[1] | **Tetsuro Matsuzawa**[2,3] | **Kirsty E. Graham**[1] | **Catherine Hobaiter**[1]

[1]Wild Minds Lab, École de psychologie et de neurosciences, Université d'Amsterdam St Andrews, St Andrews, UK

[2]Département de pédagogie, Université Chubu Gakuin, Gifu, Japon

[3]Division des sciences humaines et sociales, Institut californien des sciences humaines et sociales Technology, Pasadena, Californie, États-Unis

**Correspondance**
Catherine Hobaiter
Courriel : clh42@st-andrews.ac.uk

**Informations sur le financement**
Programme-cadre Horizon 2020, numéro de subvention/attribution : 802719 ; Programme de financement du redémarrage de la recherche de St Andrews

**Rédacteur en chef :** Thibaud Gruber

## Résumé

1. L'étude du comportement animal nous permet de comprendre comment différentes espèces et individus naviguent dans leur monde physique et social. Le codage vidéo du comportement est considéré comme un étalon-or : il permet aux chercheurs d'extraire des ensembles de données comportementales riches et nuancées, de valider leur fiabilité et de reproduire les recherches. Toutefois, dans la pratique, les vidéos ne sont utiles que si les données peuvent être extraites efficacement. La localisation manuelle de séquences pertinentes parmi 10 000 heures prend énormément de temps, tout comme le codage manuel du comportement animal, qui nécessite une formation approfondie pour être fiable.
2. Les approches d'apprentissage automatique sont utilisées pour automatiser la reconnaissance de modèles dans les données, ce qui réduit considérablement le temps nécessaire à l'extraction des données et améliore la fiabilité. Cependant, le suivi des informations visuelles pour reconnaître un comportement nuancé est un problème difficile et, à ce jour, les outils de suivi et d'estimation de la pose utilisés pour détecter le comportement sont généralement appliqués lorsque l'environnement visuel est hautement contrôlé.
3. Les chercheurs en comportement animal sont intéressés par l'application de ces outils à l'étude des animaux sauvages, mais il n'est pas clair dans quelle mesure cela est actuellement possible, ni quels outils sont les mieux adaptés à des problèmes particuliers. Pour combler cette lacune, nous décrivons les nouveaux outils disponibles dans ce domaine en évolution rapide, nous suggérons des conseils pour le choix des outils, nous fournissons une démonstration pratique de l'utilisation de l'apprentissage automatique pour suivre les mouvements dans les données vidéo des singes sauvages, et nous mettons nos modèles de base à disposition pour utilisation.
4. Nous utilisons un outil d'estimation de la pose, DeepLabCut, pour démontrer le succès de l'entraînement de deux modèles pilotes d'un problème extrêmement difficile d'estimation de la pose et de suivi : les chimpanzés et les bonobos sauvages vivant dans la forêt, multi-animaux, dans des contextes comportementaux, à partir de séquences vidéo tenues à la main.
5. Avec DeepWild, nous montrons que, sans nécessiter d'expertise spécifique en apprentissage automatique, l'estimation de la pose et le suivi des mouvements de primates sauvages vivant en liberté dans des environnements visuellement complexes est un objectif réalisable pour les chercheurs en comportement.

**KEYWORDS**
intelligence artificielle, automatisation, comportement, apprentissage profond, apprentissage automatique, primate

## 1 | INTRODUCTION

L'étude du comportement animal nous permet de comprendre comment différents individus naviguent dans leur monde physique et social, et les comparaisons entre espèces peuvent donner un aperçu de la trajectoire évolutive des capacités comportementales. Les enregistrements vidéo constituent une collecte de données particulièrement abondante et robuste, permettant d'extraire de nombreux types de comportement : de l'organisation sociale à la communication, en passant par les mouvements, etc. Contrairement à l'observation directe, il est possible de revenir à plusieurs reprises sur les mêmes événements, ce qui permet aux chercheurs d'explorer de nouvelles questions et d'améliorer ou de valider la collecte de données sur les questions existantes. Par conséquent, le codage vidéo est considéré comme un étalon-or : il permet d'obtenir des ensembles de données comportementales riches et nuancées, sur lesquelles des recherches peuvent être menées et reproduites par d'autres, aujourd'hui et à l'avenir. Outre la collecte ciblée de nouvelles données vidéo, de nombreux groupes de recherche ont créé de vastes archives vidéo à partir desquelles nous pouvons extraire des données (par exemple, Arandjelovic et al., 2016 ; Bain et al., 2021 ; Burton et al., 2015 ; Hobaiter et al., 2021 ; Schofield et al., 2019). Ces archives représentent des arcs de données : des ressources stables et à long terme qui nous aident à continuer à aborder des questions scientifiques sur des taxons, tels que les primates, qui connaissent un déclin catastrophique de leur population (Estrada et al., 2017). En outre, ces ressources numériques contribuent à lever les obstacles financiers et physiques systémiques liés à la collecte de données comportementales dans la nature, ouvrant ainsi la recherche scientifique à un groupe de chercheurs plus diversifié.

Toutefois, dans la pratique, ces vidéos ne sont utiles que si des données peuvent en être extraites efficacement. La localisation manuelle des séquences pertinentes dans des centaines ou des milliers d'heures d'archives prend énormément de temps, tout comme le codage manuel ultérieur du comportement des animaux, qui nécessite une formation approfondie pour assurer la fiabilité et limiter les erreurs et les biais des codeurs (Munch et al., 2019 ; Pathak et al., 2003). Lorsque la charge de travail et les coûts financiers associés dépassent ceux de la collecte de nouvelles données, ces archives potentiellement inestimables restent inutilisées.

Les approches d'apprentissage automatique sont utilisées pour automatiser la reconnaissance de modèles dans les données (Hastie et al., 2001) et peuvent considérablement réduire le temps nécessaire à l'extraction des données, tout en améliorant la fiabilité des résultats (Schofield et al., 2019). Ils ont été utilisés avec succès dans divers ensembles de données comportementales, de l'acoustique (Bianco et al., 2019) à la taxonomie (Wäldchen & Mäder, 2018), en passant par le mouvement (par exemple, les mouches : Günel et al., 2019 ; robots et humains : Islam et al., 2021 ; poissons : Mei et al., 2021 ; souris : Sheppard et al., 2022). Plus récemment, l'extension de ces méthodes aux données visuelles a connu un succès considérable, avec une explosion soudaine d'outils permettant d'automatiser la reconnaissance des formes dans les données photographiques et vidéo. Alors que les premiers algorithmes se sont concentrés sur les données photographiques (pour un examen approfondi, voir Weinstein, 2017), leur extension aux données vidéo est particulièrement pertinente pour la recherche comportementale, car elle permet de capturer des informations au fil du temps.

L'une des utilisations les plus répandues de l'apprentissage automatique avec les données vidéo est la reconnaissance des espèces, par exemple à partir des données des pièges photographiques (voir le tableau S1 pour des exemples). Les pièges photographiques peuvent être déployés en grand nombre, sur de vastes zones, et laissés en place pour capturer des données 24 heures par jour. Avec les considérations appropriées (Swann et al., 2011), ils permettent la surveillance d'espèces et d'individus qui ne sont généralement pas faciles à observer.

observables en personne : les populations qui ne sont pas habituées à l'observation directe, ou qui sont clairsemées ou nocturnes. Les pièges photographiques sont un moyen efficace d'aborder les questions relatives à la présence, à l'abondance et à la diversité des espèces, ainsi que de surveiller la distribution et la densité dans le temps à l'intérieur et entre les lieux (Steenweg et al., 2016). Cependant, lorsqu'ils sont utilisés à grande échelle, ils créent de vastes quantités de données vidéo dont le décodage peut prendre énormément de temps. Dans un exemple, le tri manuel des données des pièges photographiques sur la surveillance des loups avait un décalage d'environ 5 ans (Tuia et al., 2022). Avec l'utilisation de l'apprentissage automatique (Microsoft AI4Earth MegaDetector : Beery et al., 2019), toutes les données ont été étiquetées dans les 12 mois, ce qui a permis d'examiner les données avant le début de la prochaine saison de surveillance (Tuia et al., 2022). Dans un autre modèle, l'identification des primates individuels entre les espèces a pu être traitée avec une précision de 94 % à plus de 30 images par seconde (Guo et al., 2020).

Les outils automatisés d'identification des espèces divisent les vidéos en plusieurs images et examinent chacune d'entre elles pour effectuer un tri initial en filtrant les images vierges (AIDE : Kellenberger et al., 2020 ; Wildlife Insights : Ahumada et al., 2019 ; Microsoft AI4Earth MegaDetector : Beery et al., 2019), suivie de l'identification des espèces (par exemple Narouzzadeh et al., 2018 ; Willi et al., 2018 ; Yu et al., 2013 ; AIDE : Kellenberger et al., 2020 ; Wildlife Insights : Ahumada et al., 2019 ; Whytock et al., 2021) et même des individus, ce que l'on appelle la "réidentification individuelle" (Wildbook : Berger-Wolf et al., 2017 ; Guo et al., 2020 ; Schofield et al., 2019) qui ont été capturés. Les outils automatisés actuels effectuent ce travail si rapidement qu'ils sont utilisés pour envoyer des alertes en temps réel en cas de présence inattendue d'humains ou de véhicules inconnus dans les zones protégées, offrant ainsi des possibilités de réponse rapide aux équipes de conservation sur le terrain (wpsWatch : Tuia et al., 2022). Des systèmes similaires informent les communautés locales de l'approche d'animaux sauvages potentiellement dangereux, tels que les éléphants (Premarathna et al., 2020). L'identification des individus peut représenter un problème particulièrement difficile pour les codeurs humains. Dans une étude portant sur 23 chimpanzés et contenant environ un million d'images, les annotateurs humains ayant bénéficié d'une exposition de 1 à 2 heures n'ont atteint qu'une précision de 20 % (novices) et de 42 % (experts). Un modèle entraîné sur les mêmes données n'a pris que quelques secondes pour atteindre une précision de 84 % (Schofield et al., 2019), sans parler des économies réalisées par la suite dans le traitement du vaste ensemble de données. Des modèles d'identification individuelle aussi précis sont désormais disponibles pour un nombre croissant de taxons (tigres : Li et al., 2020, éléphants : Körschens & Denzler, 2019, bovins : Bergamini et al., 2018, primates : Guo et al., 2020 ; pour une vue d'ensemble, voir : Schneider et al., 2018). Néanmoins, un investissement initial important est généralement nécessaire pour développer des ensembles d'entraînement, et les outils ultérieurs sont généralement spécifiques à une population, voire à un site, ce qui limite leur généralisation.

Si l'analyse de photographies (ou d'images discrètes d'une vidéo) constitue déjà une approche puissante, l'étude de nombreux comportements nécessite l'extraction de données sur l'ensemble des images vidéo, c'est-à-dire la restitution de la composante temporelle. Une application récente de cette méthode chez les chimpanzés sauvages utilise un pipeline qui passe de la détection et du suivi des chimpanzés à l'identification individuelle et à l'identification de catégories comportementales, telles que l'alimentation (Bain et al., 2021). Cependant,

Pour ce faire, il faut choisir à l'avance le comportement qui nous intéresse et, à ce jour, cela n'a été démontré que pour deux comportements ayant des signatures auditives et visuelles distinctes (le cassage de noix et le tambourinage). Une autre approche de la catégorisation des comportements est l'estimation de la pose et le suivi des mouvements : dans ce cas, des points individuels sont marqués sur le corps et leur position relative est suivie d'une image à l'autre. Le même modèle de base peut toutefois être utilisé pour générer des données de coordonnées en vue d'analyses cinématiques des mouvements d'utilisation d'outils et des actions gestuelles, bien que cela nécessite probablement des outils de segmentation comportementale distincts lors d'une étape ultérieure. Une approche potentiellement puissante consiste à les combiner en utilisant des réseaux neuronaux convolutifs (CNN) d'action spatiotemporelle (Achour et al., 2020), qui conservent certaines informations sur le contexte visuel plus large dans lequel se situe le comportement, avec des approches d'estimation de la pose qui fournissent une analyse cinématique affinée d'ac- tions particulières. Une liste complète des outils de suivi basés sur l'apprentissage automatique, avec des informations sur leur utilisation et leur fonctionnalité, est disponible dans le tableau S2.

Dans certains cas, ces outils permettent de suivre la position d'animaux individuels les uns par rapport aux autres et par rapport à leur environnement, ce qui permet, par exemple, d'étudier en détail les mouvements de groupe de centaines d'individus en synchronisation (Walter & Couzin, 2021). La génération la plus récente d'outils de suivi permet d'estimer la pose en suivant plusieurs points sur un individu (pour une liste complète des outils d'estimation de la pose, de leur facilité d'utilisation et de leur fonctionnalité, voir le tableau S3). Cette méthode offre une certaine souplesse dans le choix des comportements suivis et la possibilité d'analyser en détail les mouvements au sein d'un comportement (voir Panadeiro et al., 2021 pour un résumé approfondi). Par exemple, l'étude des expressions faciales (Wang & Lien, 2009) ou l'analyse de la démarche (Rohan et al., 2020) chez l'homme (Khan & Wan, 2018 ; Sarafianos et al., 2016). Mais avec l'arrivée récente de logiciels "prêts à l'emploi" qui intègrent des interfaces conviviales basées sur le non-codage, l'intérêt pour le domaine plus large du comportement animal a explosé (Panadeiro et al., 2021 ; Tuia et al., 2022).

Il y a des raisons évidentes pour lesquelles - bien que l'utilisation de l'extraction de données vidéo soit une méthode puissante pour les études robustes du comportement animal, le codage manuel prend énormément de temps et - même avec des périodes de formation substantielles - les chercheurs expérimentés sont toujours sujets à une certaine erreur humaine. Même les problèmes relativement "simples", tels que le marquage de deux points (par exemple dans le cas du lip-smacking ; Pereira, Kavanagh, et al., 2020), exigent que ces points soient marqués manuellement sur chaque image, et avec des fréquences d'images typiques de 25 images par seconde et des comportements mesurés en minutes, cela représente un investissement en temps considérable - souvent des mois de travail. Avec un modèle approprié, les outils d'apprentissage automatique peuvent extraire les mêmes données en quelques minutes ou secondes. Il y a bien sûr d'importantes mises en garde : les modèles appropriés sont rarement disponibles "sur étagère" (cf. le "zoo de modèles" de DeepLabCut, Kane et al., 2020). Et, comme dans le cas du codage manuel, ces modèles vous fournissent généralement des données brutes (coordonnées x-y pour chaque point marqué dans le cadre) qui nécessitent un traitement supplémentaire substantiel pour être traduites en catégories ou mesures comportementales. Par exemple : effectuer une analyse de la marche sur les coordonnées pour extraire le rythme de la marche (Prakash et al., 2018). Cependant, les

Les outils d'apprentissage automatique qui classent les comportements à partir des coordonnées émergent des travaux de laboratoire (par exemple Hsu & Yttri, 2021), et pourraient bientôt être étendus aux données sauvages.

Le suivi des informations visuelles de manière aussi détaillée est un problème difficile et, à ce jour, les outils algorithmiques de suivi sont généralement appliqués dans le cadre d'études en laboratoire où l'environnement est fixe et/ou contrôlé, et ont tendance à avoir été développés pour des espèces animales modèles largement utilisées telles que les souris (par exemple, drosophile : Yu et al., 2011 ; rongeurs : Geuther et al., 2019 ; fourmis : Gal et al., 2020 ; vers : Kiel et al., 2018 ; poissons : Xu & Cheng, 2017 ; pour un résumé des outils logiciels actuellement disponibles, voir le tableau S2). Les avancées récentes incluent des descriptions tridimensionnelles des mouvements d'un individu dans son environnement. Pour ce faire, il faut au moins deux angles de caméra statiques qui peuvent être utilisés pour fournir l'estimation de la profondeur nécessaire pour recréer des distances objectives (sans cela, la distance entre deux points dans une seule image est arbitraire ; sont-ils petits ou sont-ils éloignés ? Bien que, cf. Haucke et al., 2021). L'éventail des espèces a également commencé à s'élargir, passant des espèces modèles de laboratoire, par exemple les souris (Gosztolai et al., 2021 ; Karashchuk et al., 2021), les mouches (Gosztolai et al., 2021 ; Günel et al., 2019 ; Karashchuk et al., 2021), aux primates et aux mammifères de plus grande taille (macaques : Bala et al., 2020 ; Gosztolai et al., 2021 ; Marks et al., 2022 ; guépards : Nath et al., 2019).

L'étude du comportement social nécessite le suivi de plus d'u n individu, ce qui exige plus qu'une simple extension de la méthode de l'individu unique. Le modèle doit être capable non seulement de suivre les parties du corps, mais aussi de savoir à qui ces parties appartiennent (c'est-à-dire que le coude A appartient à l'individu A, même s'il échange sa place avec l'individu B ou C). Par conséquent, des investissements supplémentaires en temps sont nécessaires pour la formation afin de corriger manuellement les échanges accidentels de parties du corps (Mathis et al., 2018 ; Pereira et al., 2019 ; Pereira, Tabris, et al., 2020), mais ces investissements peuvent être plus que compensés par la capacité ultérieure d'automatiser la génération rapide de données sur de nombreux individus. Certains outils, comme TRex (Walter & Couzin, 2021), se concentrent sur le suivi d'un très grand nombre d'individus, par exemple pour étudier les mouvements d'un troupeau ou d'une école ; d'autres, comme les options de suivi multi-animal de SLEAP, peuvent suivre des parties distinctes du corps sur un nombre modéré d'individus (c'est-à-dire <100 ; Pereira et al., 2019 ; Pereira, Tabris, et al., 2020). Jusqu'à récemment, les logiciels d'estimation de la pose étaient limités aux laboratoires et, de plus en plus, aux animaux domestiques et aux animaux de compagnie (Kane et al., 2020), ainsi qu'aux environnements captifs tels que les zoos (Hayden et al., 2021 ; Marks et al., 2022). Dans ces environnements, le "bruit visuel" est à la fois relativement faible et stable d'une vidéo à l'autre. Comme pour les humains, il est beaucoup plus facile pour les outils d'apprentissage automatique de détecter un animal en mouvement lorsque rien d'autre ne bouge dans le cadre, ou un animal sur un f o n d uni avec un bon éclairage. Toutefois, grâce à des logiciels de plus en plus sophistiqués capables d'apprendre à partir de plusieurs individus et dans des conditions plus variables, les outils d'estimation de la pose pourraient enfin être étendus aux populations sauvages. Les chercheurs qui étudient les variations comportementales dans un large éventail de disciplines peuvent ainsi bénéficier d'une puissance considérable.

plines : de l'écologie à la cognition, de la conservation à la culture.

Bien qu'il semble y avoir un intérêt significatif à essayer de le faire, avec autant d'outils d'apprentissage automatique disponibles, il peut être difficile de savoir lesquels sont adaptés aux différents types de données et de questions. Des résumés récents sont disponibles pour le laboratoire

(Panadeiro et al., 2021) et de conservation (Tuia et al., 2022), mais moins d'informations sont disponibles pour les spécialistes du comportement qui travaillent avec des populations sauvages. Le choix de l'outil à utiliser peut être abordé en considérant quelques questions clés (figure 1 ; pour une liste actualisée des logiciels disponibles, voir les tableaux S2 et S3).

Dans cet article, nous prenons l'un des principaux outils actuellement disponibles, DeepLabCut (Mathis et al., 2018), et fournissons un exemple travaillé de sa fonctionnalité avec un ensemble de données particulièrement difficile : celui des vidéos de chimpanzés sauvages et de bonobos. Nous le faisons du point de vue d'un groupe de chercheurs en comportement animal, avec une expérience substantielle dans le travail avec le codage manuel de l'extraction de comportements nuancés à partir de la vidéo, mais seulement des compétences de base en apprentissage automatique.

Initialement développé pour le suivi comportemental de la souris et de la drosophile (Mathis et al., 2018), DeepLabCut a depuis été appliqué à un large éventail d'autres espèces (rats, Clemensson et al., 2020 ; poissons, Habe et al., 2021 ; guépards, Joska et al., 2021 ; chevaux, Tsuruo et al., 2020). DeepLabCut propose une estimation de la pose de plusieurs animaux, une interface graphique conviviale et des exemples de vidéos de suivi. L'extraction de données visuelles à partir de vidéos de singes sauvages vivant dans la forêt est peut-être l'une des tâches les plus difficiles pour l'apprentissage automatique : les singes se déplacent librement dans les trois dimensions de leur environnement, nous nous déplaçons en suivant les singes, nos caméras portatives se déplacent, l'éclairage est souvent sombre mais peut inclure des contrastes spectaculaires - avec des singes sombres, dans une forêt sombre, éclairés à contre-jour par des ciels lumineux. Enfin, les forêts elles-mêmes sont visuellement denses, avec de nombreux obstacles visuels (branches, arbres, feuilles, autres singes) qui se déplacent eux-mêmes. En outre, nous formons ici un modèle qui inclut des individus de deux espèces étroitement apparentées mais physiquement distinctes : les bonobos et les chimpanzés, y compris deux singes.

sous-espèces de chimpanzés (Afrique de l'Est et de l'Ouest) et des individus de toutes les classes d'âge et de sexe, ainsi que des populations vivant dans différents types d'habitats. Une décision typique que les chercheurs doivent prendre est de savoir s'il faut augmenter la taille de l'ensemble de formation dans un laps de temps donné en demandant à plusieurs personnes de marquer les cadres. Bien que cela augmente la taille de l'ensemble de formation, cela introduit un nouvel aspect du bruit dans les données : la variation entre les marqueurs. Nous fournissons un exemple de base de ce compromis en entraînant un deuxième modèle. Le modèle 2 reproduit le modèle 1, mais inclut des images supplémentaires dans l'ensemble d'entraînement (augmentation de 27 %) marquées par un deuxième marqueur. Nous avons fourni ce que l'on pourrait considérer comme des ensembles d'entraînement minimalistes (<2000 images ; cf., par exemple, 195 228 images utilisées pour c r é e r OpenMonkeyStudio, un estimateur de pose pour les primates en captivité ; Bala et al., 2020), ce qui représente ~100-140 heures de codage humain à produire. Par conséquent, nos données et nos résultats représentent probablement une valeur aberrante en termes de difficulté de la tâche : en substance, si notre modèle, entraîné sur un ensemble minimaliste d'images, peut effectuer un suivi de base malgré le niveau élevé de diverses formes de bruit visuel dans ces données, cela suggère que des modèles similaires pourraient fonctionner pour la plupart des autres ensembles de données vidéo comportementales de primates.

## 2 | MATÉRIEL ET MÉTHODES

### 2.1 | Utilisation de DeepLabCut

De nombreux guides d'utilisation sont disponibles pour DeepLabCut, y compris ceux des développeurs (voir : DeepLabCut Github, 2021a, 2021b), ainsi que ceux des utilisateurs (par exemple Gadea, 2021). Téléchargement et installation de DeepLabCut et utilisation initiale de l'interface utilisateur graphique (GUI)

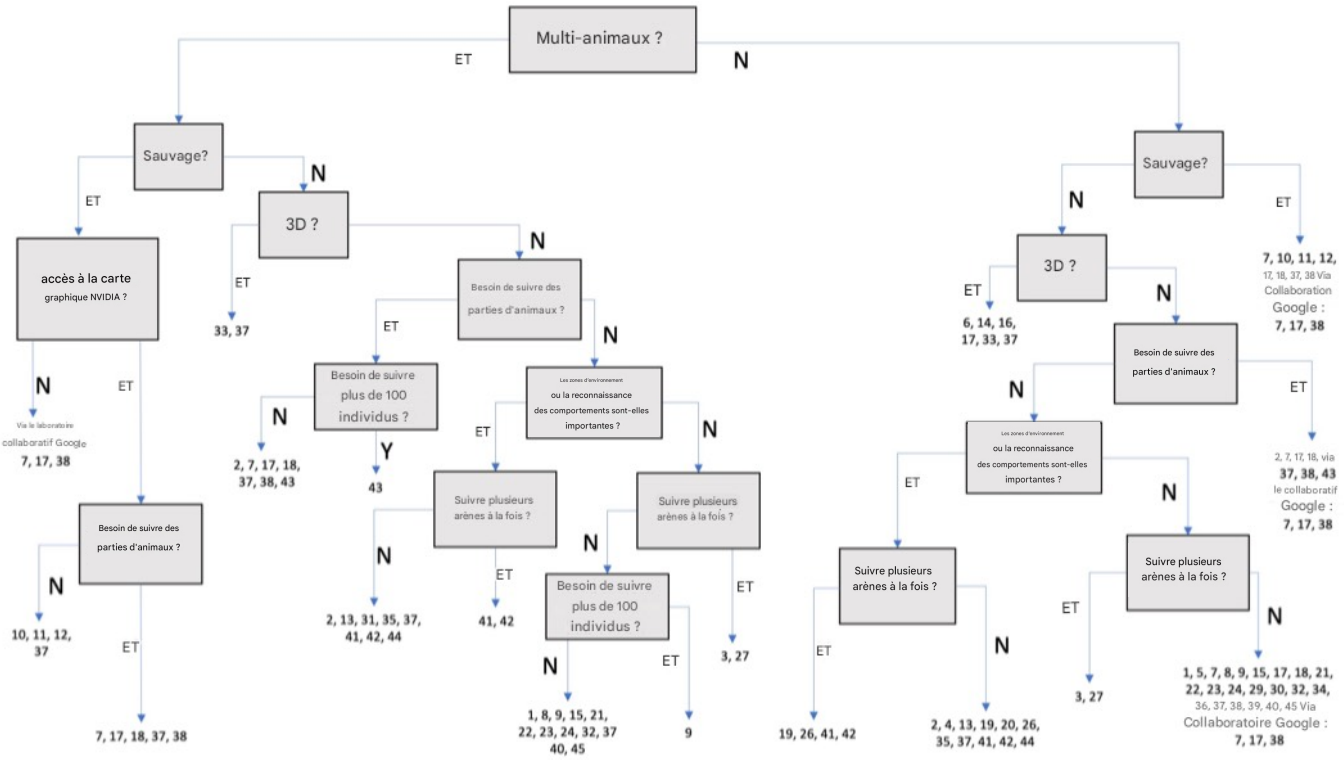


**FIGURE 1** Arbre de décision pour la sélection des logiciels. Les logiciels sont numérotés et reliés au tableau S2, qui fournit une description de chaque outil, de ses utilisations antérieures et de ses fonctionnalités. Une évaluation plus détaillée du sous-ensemble d'outils de suivi qui fournissent une estimation de la pose est présentée dans le tableau S3.

exige que les utilisateurs installent d'abord Python. La distribution Anaconda de Python est recommandée car elle comprend des packs préinstallés utiles. Bien que les développeurs de DeepLabCut fournissent des instructions pour installer Python et DeepLabCut à partir de zéro (DeepLabCut Github, 2021c), il est utile d'avoir une compréhension de base de Python ou du terminal de commande du système d'exploitation choisi.

DeepLabCut peut être utilisé avec ou sans matériel spécialisé. Une unité de traitement graphique (GPU) est recommandée et réduit le temps de formation. Cependant, du matériel informatique moderne standard peut être utilisé. Il est également possible d'utiliser Google Colaboratory pour accéder à un GPU gratuit hébergé dans le nuage. L'interface graphique de DeepLabCut n'est pas disponible si l'on utilise Google Colaboratory, et les utilisateurs devront avoir une connaissance plus approfondie de Python, mais des exemples de manuels et de tutoriels sont disponibles sur le site Youtube de DeepLabCut (https://www.youtube.com/ channel/UC2HEbWpC_1v6i9RnDMy-dfA).

Une fois installée et ouverte, l'interface graphique utilise des onglets pour guider les utilisateurs tout au long du processus de création de nouveaux projets ou d'ouverture de projets existants. Elle n'offre pas encore de fonctionnalités telles qu'un "graphique des pertes" qui permet aux utilisateurs de suivre la progression de la formation au modèle (cf. SLEAP, Pereira, Tabris, et al., 2020). Mais une version simple sur mesure peut être facilement générée pour évaluer les pertes une fois l'entraînement terminé (exemple de code disponible ici : https://github.com/Wild-Minds/DeepWild). Les graphiques de perte aident les utilisateurs à comprendre quand mettre fin à l'entraînement du modèle, car le surentraînement peut conduire à un surajustement, ce qui réduit les performances du modèle. DeepLabCut recommande de mettre fin à l'entraînement lorsque le graphique de perte est trop élevé, d'où l'utilité de l'aide visuelle d'un graphique.

## 2.2 | Données et sujets d'étude

Nous avons extrait les données vidéo de la base de données du dictionnaire des grands singes (Hobaiter et al., 2021). Les vidéos ont été enregistrées entre 2013 et 2020, et ont toutes été enregistrées à l'origine en haute définition ou en 4 K à l'aide de caméscopes Panasonic de poche avec une fréquence d'images de 25 images par seconde (par exemple HCV770 ou HCVX-F1). Les données vidéo originales ont été collectées auprès d'un bonobo *Pan paniscus* et de quatre communautés de chimpanzés *Pan troglodytes* appartenant à deux sous-espèces (chimpanzés d'Afrique de l'Est : *Pan troglodytes schweinfurthi*, chimpanzés d'Afrique de l'Ouest : *Pan troglodytes verus*). Bien que très proches, les différentes espèces de *Pan* présentent néanmoins des différences caractéristiques dans la morphologie et les mouvements (Doran, 1993 ; Jungers & Susman, 1984).

La population de bonobos incluse était celle de Wamba dans la Réserve Scientifique de Luo en République Démocratique du Congo, à partir de laquelle nous avons inclus deux groupes voisins de bonobos, le groupe E1 et le groupe P, dont les aires de répartition se chevauchent et qui se rencontrent fréquemment. L'habitat des communautés de Wamba est caractérisé par une forêt primaire et secondaire sèche (Hashimoto et al., 1998 ; Terada et al., 2015) au sein d'un habitat anthropique (Terada et al., 2015). Trois des quatre communautés de chimpanzés étaient des communautés de chimpanzés d'Afrique de l'Est : Sonso et Waibira se trouvent toutes deux dans la réserve forestière de Budongo, en Ouganda, et le groupe M dans la réserve forestière centrale de Kalinzu, en Ouganda. Leurs habitats sont caractérisés par une forêt secondaire dense de moyenne altitude, à feuilles semi-décidues (Eggeling, 1947).

La quatrième communauté de chimpanzés est celle de Bossou, en Guinée, une communauté de chimpanzés d'Afrique de l'Ouest qui vit dans des fragments de forêt au sein d'un habitat anthropogénique et qui est filmée à un nettoyage ouvert qu'elle visite régulièrement pour casser des noix (Matsuzawa et al., 2011).

## 2.3 | Éthique

L'approbation éthique pour la collecte des données originales et l'utilisation de la base de données du dictionnaire des grands singes (Hobaiter et al., 2021) a été fournie par l e Comité d'éthique et de bien-être des animaux de l'Université de St Andrews (code d'approbation : PS15842). L'approbation éthique a été fournie par l'Uganda Wildlife Authority et le Conseil national ougandais pour la science et la technologie (NS179) pour la collecte originale de données vidéo sur les chimpanzés en Ouganda, par le Ministère de la Recherche Scientifique et Technologie, pour la collecte de données originales de vidéos de bonobos en République démocratique du Congo, et par le Ministre de l'Enseignement Supérieur et de la Recherche Scientifique, et la Direction Générale de la Recherche Scientifique et de l'Innovation Technologique pour la collecte de données originales de vidéos de chimpanzés en Guinée.

## 2.4 | Sélection vidéo

Les vidéos ont été choisies de manière à inclure le plus de "bruit" visuel possible. Le "bruit" fait référence à des variations qui augmentent la difficulté de discerner les données visuelles disponibles pour l'apprentissage. Par exemple, le bruit est généré par des variations de comportement, ainsi que par l'espèce, l'âge et le sexe des individus. Le bruit est également généré par des variations telles que : un éclairage inégal, un mauvais éclairage, un fort contraste, des ombres, une similitude de couleur ou de texture entre l'individu concerné et l'environnement, le chevauchement d'individus, des parties du corps masquées, le mouvement de la caméra, le mouvement de l'individu, le mouvement de l'environnement. Tous ces facteurs augmentent les difficultés de reconnaissance et de suivi des parties du corps pour l'estimation de la pose. Étant donné que nos données sont soumises à tous ces facteurs, souvent plusieurs à la fois, nous avons entraîné notre modèle à incorporer une variation représentative dans notre ensemble d'entraînement.

Un problème typique auquel les chercheurs sont confrontés est le nombre d'images d'entraînement nécessaires. Dans un environnement de laboratoire contrôlé, DeepLabCut peut commencer le suivi avec seulement quelques centaines d'images marquées (Lauer et al., 2021 ; Mathis et al., 2018), les modèles réussis étant créés sur quelques milliers d'images pour les vidéos de laboratoire (par exemple, 1080 images pour des souris sombres sur un fond blanc uni, Mathis et al., 2018). Cependant, le nombre d'images d'entraînement nécessaires reflète largement la quantité de bruit visuel dans vos données. Par conséquent, un grand nombre d'images est nécessaire pour les données visuellement plus bruyantes (par exemple, 1080 images pour des souris sombres sur un fond blanc uni, Mathis et al.
>13 000 images pour les macaques dans un environnement ouvert de type zoologique : Labuguen et al., 2021 ; 7600 images pour des ouistitis hébergés dans un zoo avec plusieurs animaux : Lauer et al., 2021 ; 7588 images pour les guépards dans une savane ouverte, Joska et al., 2021). Cependant, le marquage manuel des images nécessite un investissement initial substantiel pour développer les ensembles d'entraînement, et il est probable qu'après un certain temps, les rendements diminuent dans la mesure où l'on ne dispose pas d'un système de marquage manuel.

Il s'agit d'un compromis entre le temps investi et l'augmentation de la précision du modèle. Une deuxième question est de savoir s'il faut utiliser plusieurs codeurs humains pour établir l'ensemble d'entraînement - cela peut réduire considérablement le délai nécessaire pour développer l'ensemble d'entraînement, mais peut introduire un bruit supplémentaire en termes d e différences entre les codeurs (même les codeurs formés atteignent rarement une fiabilité inter-observateurs parfaite, par exemple la variabilité humaine dans le RMSE des pixels sur le marquage des images DeepLabCut chez les souris était de 3 à 4 ; Mathis et al., 2018). Ici, nous fournissons aux modèles un ensemble d'entraînement minimaliste (<2000 images) et entraînons deux modèles pour étudier si l'augmentation de la taille de l'ensemble d'entraînement est compensée par la variabilité multi-codeurs (voir le tableau 1 pour le résumé).

Le modèle 1 contenait 1 375 images d'entraînement provenant de 55 vidéos. Celles-ci incluaient deux espèces (les bonobos et les chimpanzés) pour un total de
5 communautés de singes (Wamba-E1, Wamba-P, Sonso, Waibira et Bossou), et toutes les images d'entraînement ont été marquées par un seul chercheur. Le modèle 2 était une extension du modèle 1, avec 825 images supplémentaires provenant de 55 nouvelles vidéos, y compris le marquage par un deuxième codeur et une communauté supplémentaire de chimpanzés d'Afrique de l'Est, le groupe Kalinzu M (ensemble d'entraînement total : 2200 images, 110 vidéos, 6 communautés de singes).

## 2.5 | Préparation de la vidéo

Toutes les vidéos ont été limitées à un maximum de 90 s afin de réduire tout effet de la longueur de la vidéo sur l'analyse (pour les cadres de test, le marquage de n cadres sur un total de 1 000 donne un ratio marqué/nouveau plus élevé dans les vidéos que le marquage de *n* cadres sur un total de 10 000). Les vidéos durent de 6 à 88 s (moyenne = 45 s ; écart-type = 22 s). Les vidéos ont été exclues si plus de 7 individus étaient présents afin de limiter le temps investi dans le marquage manuel (il convient de noter que même s'ils sont formés sur des vidéos limitées à un maximum de cinq individus, les modèles DeepLabCut peuvent ensuite suivre jusqu'à 100 individus dans des vidéos inédites).

## 2.6 | Détails du modèle

Les images ont été marquées à l'aide de 18 points clés (figure 2), ce qui a nécessité en moyenne 2 heures par vidéo (10 ou 25 images), bien que ce temps varie considérablement en fonction des niveaux de bruit visuel et du nombre d'individus présents dans l'image. Si un point clé n'était pas visible de la caméra, il n'était pas marqué pour cette image.

La formation a été réalisée sur un ZBook Create G7 équipé d'un Intel® Core™ i7-10750H (fréquence de base de 2,6 GHz, jusqu'à 5,0 GHz avec la technologie Intel® Turbo Boost, 12 Mo de cache L3, 6 cœurs) et de 32 Go de RAM DDR4-3200 MHz. Nous n'avons pas dérogé aux options par défaut suggéré pour l'entraînement de modèles multi-animaux. Comme étape supplémentaire, nous avons entraîné une version du modèle 1 sur une seule carte Nvidia Tesla V100 sur des nœuds avec un Intel(R) Xeon(R) Gold 6130 CPU @ 2.10GHz pour comparer le temps d'entraînement, compte tenu de la puissance de calcul supplémentaire. Les modèles entraînés de cet article sont accessibles au public sur notre GitHub et archivés dans Zenodo, voir la déclaration de disponibilité des données pour plus de détails.

## 2.7 | Performance de l'entreprise

Nous avons utilisé la distance euclidienne absolue moyenne pour comparer les points générés par le modèle et les points étiquetés par l'homme. Ces valeurs sont obtenues en calculant la distance euclidienne entre les points générés par le modèle et les points étiquetés par l'homme, pour chaque détection. Pour les performances sur les images de test, ces valeurs sont rapportées par DeepLabCut et ne sont exécutées que lorsque les points ont été prédits avec une probabilité supérieure au seuil p de 0,6. Pour les performances sur les nouvelles images vidéo, ces valeurs sont rapportées pour toutes les détections effectuées par le modèle.

## 2.8 | Utilisation 1 : performances sur les images "test", modèle DLC et second codeur humain

DeepLabCut conserve 5 % des images marquées manuellement que les utilisateurs fournissent en tant qu'ensemble "test". Ces images ne sont pas incluses dans l'apprentissage du modèle et sont utilisées pour l'évaluation des performances du modèle. Ici, la performance est une comparaison entre les points dérivés du modèle et ceux étiquetés par le marqueur humain (dans ce cas, tous les marqueurs 1). En plus de cette comparaison, nous avons comparé les performances d'un deuxième humain en calculant la distance euclidienne absolue moyenne entre les points marqués par le premier humain et les points marqués par le deuxième humain sur le même ensemble de test.

Il est à noter que les images sont tirées de certaines des mêmes vidéos que celles utilisées pour l'entraînement du modèle, ce qui signifie que le modèle a l'expérience du type d'informations visuelles sur lequel il est testé. Toutefois, le fait de fournir à un modèle un ensemble d'entraînement comprenant un sous-ensemble d'images provenant de toutes les vidéos qu'un chercheur souhaite coder représente toujours un gain de temps substantiel dans la pratique (les utilisateurs marquent un maximum de 25 images par vidéo lors de la configuration des ensembles d'entraînement, ce qui équivaut à marquer ~1 s de vidéo pour chaque vidéo).

## 2.9 | Utilisation 2 : performances sur des vidéos "inédites

Les nouvelles vidéos, dans lesquelles aucune image de la vidéo n'a été incluse dans la formation, représentent une tâche plus difficile. Les vidéos peuvent inclure

**TABLE 1** Résumé des ensembles d'entraînement utilisés dans le modèle 1 et le modèle 2. Nous indiquons le nombre d'annotateurs, le nombre de vidéos dont les images d'entraînement ont été extraites, le nombre total d'images marquées pour l'entraînement, le nombre d'espèces sur lesquelles le modèle a été entraîné et la répartition test/entraînement utilisée lors de l'entraînement du modèle.

| | # Annotateurs | # Vidéos | # Cadres | # Espèces | # Communautés | Séparation de la formation et de l'essai |
|---|---|---|---|---|---|---|
| Modèle 1 | 2 | 55 | 1375 | 2 | 5 | 95/5 |
| Modèle 2 | 1 | 110 | 2200 | 2 | 6 | 95/5 |

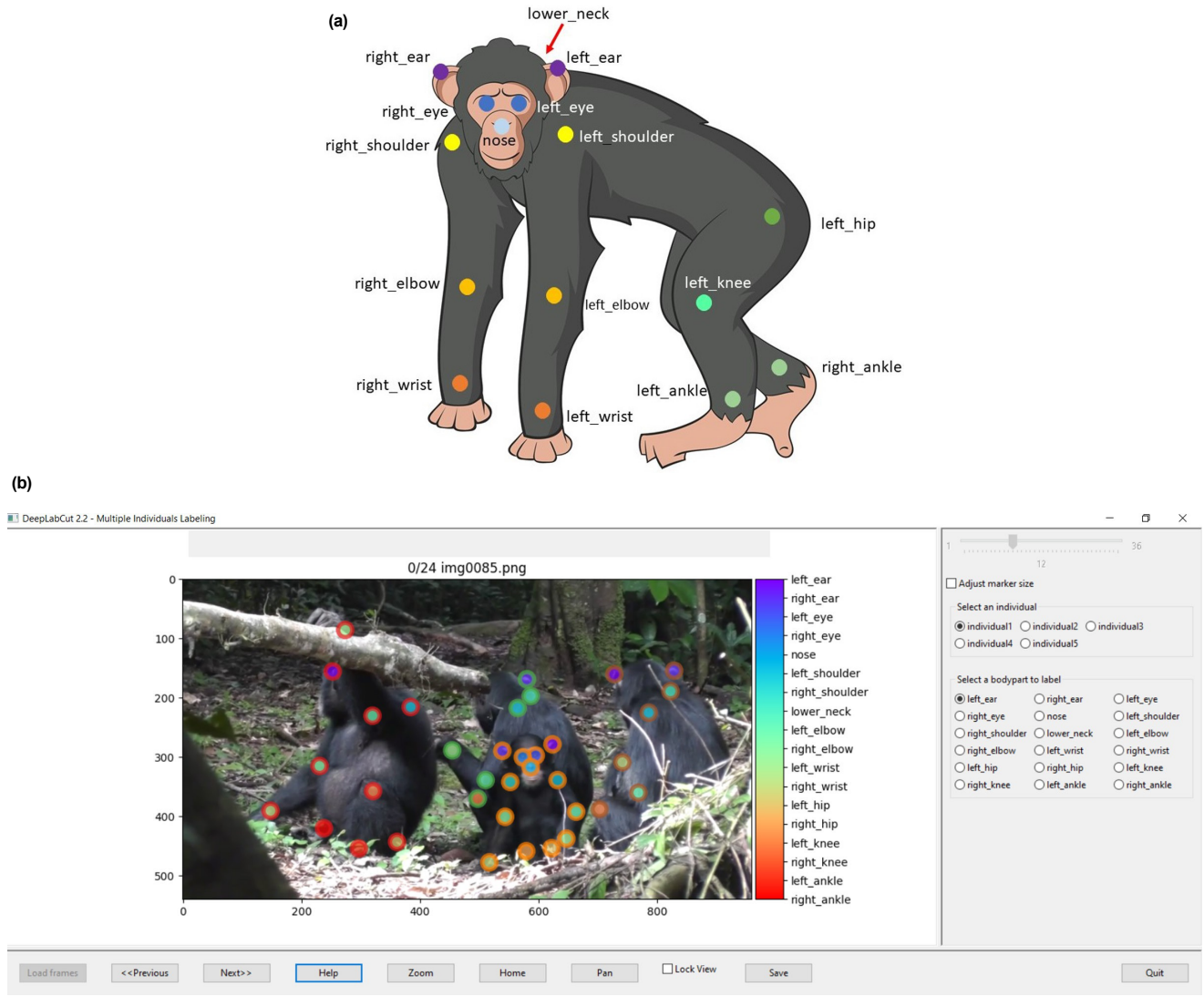


**FIGURE 2** (a) Les 18 points clés marqués. Chacun de ces points clés est marqué, lorsqu'il est visible, sur chaque individu du cadre. (b) Exemple d'image de l'interface utilisateur graphique de DeepLabCut. Ici, nous montrons 36 points clés marqués sur quatre chimpanzés d'Afrique de l'Est, trois adultes et un bébé.

L'éclairage, les angles, les mouvements, les distances, les environnements et les individus que le modèle n'a pas rencontrés lors de l'entraînement. Pour tester les performances sur de nouvelles vidéos, 25 images de 9 vidéos (total = 250 images) ont été marquées manuellement par un codeur expérimenté (CW). Cela a permis d'obtenir les coordonnées x-y de chaque partie du corps de chaque individu pour chaque image marquée. Nous avons ensuite introduit ces vidéos dans le modèle pour générer l e s coordonnées prédites par le modèle et nous avons comparé les performances avec l e s images marquées manuellement. Les nouvelles vidéos ont été classées en trois catégories : facile, moyen et difficile, en fonction de la quantité de bruit visuel présent dans la vidéo (pour plus de détails, voir le tableau S4).

## 3 | RÉSULTATS

Le modèle 1 a pris 28 heures pour s'entraîner jusqu'à 200 000 itérations, moment où l'optimiseur a signalé une perte de 0,001 sur le ZBook. L'erreur d'entraînement a été signalée à 5,96 pixels, l'erreur de test à 18,46 pixels (où une coupure p de 0,6 a été appliquée : erreur d'apprentissage : 4,38 pixels, erreur de test 10,12 pixels). Une version correspondante du modèle 1 a été entraînée sur la Tesla V100, plus puissante sur le plan informatique, sur des nœuds dotés d'un processeur Intel(R) Xeon(R) Gold 6130 à 2,10 GHz. L'entraînement a duré le même temps (26,5 heures) pour atteindre une perte de 0,002, ce qui s'est produit à 100 000 itérations.

Le modèle 2 a nécessité 26 heures d'entraînement à 200 000 itérations et une perte de 0,001 sur le ZBook. L'erreur d'entraînement était de 7,31 pixels, l'erreur de test de 18,63 pixels (avec un seuil p de 0,6 : erreur d'entraînement de 4,6 pixels, erreur de test de 9,64 pixels) : 4,6 pixels, erreur de test 9,64 pixels).

### 3.1 | Performance humaine

La distance euclidienne absolue moyenne du second marqueur humain (par rapport au marqueur humain original) était de 26,09 (écart-type = 14,31) sur l'ensemble des points. Comme pour les performances du modèle, cette distance varie en fonction de la partie du corps (tableau 2).

**TABLE 2** Distance euclidienne absolue moyenne entre les codeurs humains par partie du corps dans $n = 570$ images. Il convient de noter que toutes les parties du corps ne sont pas visibles dans toutes les images, ce qui explique le nombre de points par partie du corps.
varie et est indiqué par *N*.

| Partie du corps | Moyenne absolue Euclidienne distance (SD) | *N* |
|---|---|---|
| Cheville | 33.44 (27.65) | 41 |
| Oreille | 13.36 (9.43) | 63 |
| Coude | 43.06 (35.23) | 85 |
| L'œil | 7.05 (9.06) | 40 |
| Hanche | 48.83 (27.05) | 48 |
| Genou | 23.95 (25.89) | 88 |
| Cou | 27.12 (13.21) | 43 |
| Nez | 5.31 (2.97) | 43 |
| Epaule | 29.81 (16.54) | 66 |
| Poignet | 29.06 (22.44) | 53 |
| tous | 26.10 (14.31) | 570 |

### 3.2 | Performances du modèle sur les cadres d'essai

Malgré une variabilité supplémentaire dans l'entrée des codeurs (2 codeurs) et l'ajout d'une nouvelle population (Kalinzu), le modèle 2 ($n = 2200$ images d'entraînement, figure 3a) a surpassé le modèle 1 (n = 1375 images d'entraînement) pour toutes les parties du corps (figure 3b). Des exemples de suivi des modèles 1 et 2 sur des vidéos de test sont présentés dans les vidéos S1 et S2. La variation par rapport aux images d'entraînement originales marquées par l'homme était considérablement plus faible dans les deux modèles que celle du second marqueur humain, avec des valeurs absolues moyennes de distance euclidienne jusqu'à ~10 fois plus petites (par exemple, hanche).

### 3.3 | Performances du modèle sur des vidéos inédites

Des exemples de suivi des modèles 1 et 2 sur de nouvelles vidéos sont présentés dans les vidéos S5-S10 (les vidéos S5-S8 représentent une bonne performance, les vidéos S9 et S10 une performance médiocre). Le modèle 2 a obtenu u n plus grand nombre de détections sur toutes les vidéos que le modèle 1 (tableau 3), mais n'a obtenu une distance euclidienne absolue moyenne inférieure que dans huit des 17 vidéos, bien que ces valeurs aient été généralement générées à partir d'un plus grand nombre de détections (tableau 3).

Les deux modèles ont éprouvé des difficultés à assembler les détections (par exemple, détecter un coude) en assemblages (par exemple, le détecter comme étant le coude de l'individu A). détection d'un coude) en assemblages (par exemple, détection du coude de l'indi- vidu A), entraînant un manque de suivi sur neuf vidéos pour le modèle 1 (kalinzu10, kalinzu18, sonso17, sonso3, sonso4, waibira1, waibira17, wamba11 et pas d'assemblage pour waibira15) et huit vidéos pour le modèle 2 (kalinzu10, kalinzu18, kalinzu20, sonso3, sonso4, waibira1, waibira17, wamba11).

En revanche, si l'on considère des parties spécifiques du corps, le modèle 2 est plus performant que le modèle 1, à l'exception de l'oreille et de l'épaule (8 parties du corps sur 10, tableau 4).

## 4 | DISCUSSION

En utilisant DeepLabCut, nous avons entraîné avec succès deux modèles sur un problème d'estimation de pose extrêmement difficile : des chimpanzés et des bonobos sauvages vivant dans la forêt, multi-animaux, dans des contextes comportementaux, à partir de séquences vidéo tenues à la main. Nos modèles sont robustes pour les deux espèces de *Pan* étroitement apparentées, pour des individus d'âges et de sexes différents, et pour un large éventail d'environnements socio-écologiques - y compris des clairières ouvertes aux forêts denses, et des comportements statiques tels que le toilettage aux comportements hautement dynamiques tels que le jeu.

Les performances de suivi sur les vidéos de test, des vidéos dont certaines images avaient été utilisées pour l'entraînement, étaient nettement meilleures que les variations entre codeurs humains sur des images vidéo similaires de chimpanzés sauvages. La comparaison directe des performances sur les vidéos de test et les nouvelles vidéos est difficile car les vidéos sont elles-mêmes très variables - ainsi, le fait qu'une partie du corps soit visible (et donc potentiellement détectable) ou occultée varie également. Les performances sur des vidéos entièrement nouvelles étaient plus faibles, mais le suivi était encore souvent réussi, avec une précision dans les parties du corps les plus faciles (oreilles, yeux, nez) atteignant des niveaux similaires à ceux de la variation des marqueurs interhumains sur les parties du corps plus difficiles (hanches, épaules).

Le modèle 2 a montré une amélioration de la détection des points du corps par rapport au modèle 1 (environ 10 % de détections en plus). La précision du modèle 2 dans cet ensemble plus large de détections a montré une amélioration constante pour 8 des 10 parties du corps, avec une distance euclidienne absolue moyenne généralement comprise entre la moitié et les trois quarts de celle du modèle 1. Outre un ensemble d'entraînement plus important, le modèle 2 comprenait des images d'entraînement marquées par un deuxième codeur humain et une communauté de chimpanzés supplémentaire. DeepLabCut donne la priorité à la précision (degré d'exactitude de la détection des points) et exige par conséquent un niveau de confiance relativement élevé pour indiquer un point, ce qui peut entraîner un rappel plus faible (nombre de points détectés avec succès) dans les modèles dotés d'ensembles d'entraînement homogènes. L'intégration de la diversité dans les ensembles d'entraînement est une étape essentielle dans le développement des modèles de buste et nécessite une sélection minutieuse du matériel vidéo et une compréhension du contenu des ensembles vidéo auxquels le modèle sera appliqué.

Bien qu'il s'agisse d'une première étape importante, l'utilisation du suivi automatisé de la pose et du comportement des primates sauvages se heurte encore à plusieurs limites. La plus importante est peut-être l'investissement en temps nécessaire au développement du modèle - nos modèles représentent une première étape, mais nécessitent encore des développements supplémentaires avant d'être suffisamment robustes pour ne plus nécessiter de correction humaine post-hoc par le codeur. Cependant, des ensembles d'entraînement plus importants nécessitent un investissement en temps plus substantiel. Les deux modèles actuels de DeepWild ont utilisé des ensembles d'entraînement minimaux (<2000 images), représentant environ 110-200 heures-personnes d'investissement à produire. L'entraînement du modèle a nécessité 26 à 28 heures supplémentaires, le temps supplémentaire étant investi dans d'autres travaux (il est intéressant de noter qu'il n'y a pas eu de gain en temps d'entraînement par rapport à une perte similaire dans l'utilisation d'une plus grande puissance de calcul). Compte tenu d'un investissement estimé à 200 heures et d'un taux de majoration humain d'environ 2 heures pour 25 images, l'utilisation du modèle 2

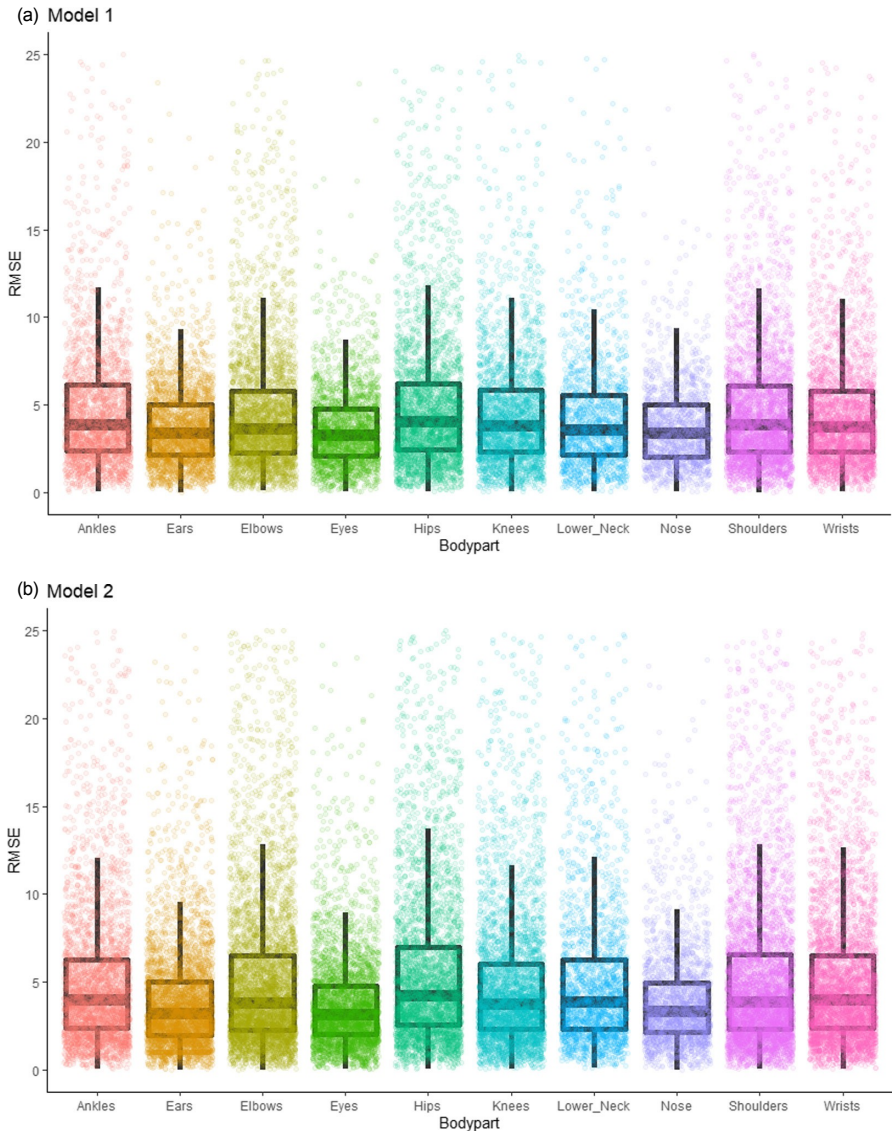


**FIGURE 3** Performances des modèles sur les différentes parties du corps. Valeurs absolues moyennes de la distance euclidienne pour le modèle 1 (a) et le modèle 2 (b) sur l'ensemble des parties du corps (étiquetées comme RMSE par DeepLabCut). La figure montre des valeurs comprises entre 0 et 25, la plage complète étant plus large.

L'investissement en temps serait rentabilisé après seulement 40 à 50 minutes de vidéo, soit une fraction du temps codé dans de nombreuses études sur le comportement des animaux. Néanmoins, les coûts de développement des ensembles d'entraînement peuvent encore représenter un obstacle à l'accès initial. Étant donné les bonnes performances du modèle sur les images de test, une approche pour les chercheurs en comportement qui souhaitent commencer à utiliser le suivi dans de très grands ensembles de données vidéo consiste à marquer manuellement un petit sous-ensemble d'images pour chaque vidéo qu'ils ont l'intention de suivre. Cette méthode constitue un point d'entrée relativement peu coûteux pour l'utilisation d'outils d'apprentissage automatique permettant de générer rapidement des données très précises sur l'estimation de la pose et le suivi du comportement. La flexibilité accrue nécessaire pour suivre des vidéos entièrement nouvelles requiert un investissement dans un ensemble d'entraînement plus substantiel. Une approche consiste à examiner attentivement les parties du corps susceptibles de présenter un intérêt pour les différents projets, puis à marquer un ensemble complet, même si seules quelques parties sont nécessaires pour un projet donné. Bien que cela augmente le temps de marquage pour l'élaboration d'un ensemble de formation individuel, les ensembles de formation suivants peuvent être assemblés avec les ensembles existants pour produire des modèles de plus en plus puissants. À un certain stade, les performances du modèle seront probablement telles qu'il ne sera plus nécessaire d'utiliser d'autres cadres de formation. Un autre avantage de cette méthode d'apprentissage multi-ensembles est qu'elle permet d'obtenir des modèles de plus en plus performants.

L'approche de la construction multi-ensembles est que les ensembles d'entraînement peuvent être combinés de différentes manières pour adapter un modèle particulier à un besoin spécifique - par exemple, des individus, des espèces ou des contextes socio-écologiques particuliers. La construction d'ensembles multiples peut être particulièrement efficace si elle peut être adoptée en collaboration par des groupes de recherche, par exemple dans le cadre d'un consortium de recherche tel que ManyPrimates (Altschul et al., 2019), ce qui permet d'atténuer le coût pour un chercheur ou un groupe particulier, tout en produisant rapidement de grands ensembles d'entraînement et des modèles très flexibles. Une approche similaire est adoptée avec le zoo de modèles de DeepLabCut (Kane et al., 2020), où les modèles de base peuvent être fournis et téléchargés par les utilisateurs, qui peuvent ensuite les affiner en fonction de leurs besoins spécifiques. Un autre moyen de compenser les coûts de développement est d'employer une approche scientifique communautaire - où des interfaces graphiques conviviales en ligne permettent aux membres du public de consacrer du temps à des projets de recherche. Déjà utilisées intensivement dans le cadre de travaux de piégeage par caméra (par exemple Chimp&See ; Arandjelovic et al., 2016) pour l'identification des espèces et la classification des comportements, des plateformes telles que Zooniverse (www.zooniverse.org) offrent aux scientifiques un moyen facile d'héberger des projets scientifiques communautaires en ligne, y compris des outils intégrés permettant d'évaluer l'impact des projets sur l'environnement.

**TABLE 3** Performance du modèle sur les nouvelles vidéos. Le modèle 1 contenait 1 375 images provenant de 55 vidéos de 5 communautés *Pan* étiquetées par un seul codeur, et le modèle 2 contenait 2 200 images provenant de 110 vidéos de 6 communautés *Pan* étiquetées par deux codeurs. Les vidéos ont été classées en fonction de leur difficulté sur la base des facteurs de bruit visuel présents. MAED = distance euclidienne absolue moyenne.

| | | Modèle 1 | | Modèle 2 | |
|---|---|---|---|---|---|
| Vidéo | Difficulté | MAED (SD) | n détections | MAED (SD) | *n* détections |
| Bossou7 | Facile | 109.3 (212.1) | 470 | 27.4 (45.0) | 541 |
| Bossou8 | Facile | 44.6 (74.8) | 175 | 31.9 (55.4) | 279 |
| Kalinzu19 | Facile | 80.3 (71.3) | 148 | 78.8 (71.3) | 214 |
| Sonso17 | Facile | 19.1 (21.2) | 59 | 15.6 (13.1) | 65 |
| Waibira7 | Facile | 26.3 (40.6) | 55 | 49.9 (123.7) | 83 |
| Wamba16 | Facile | 27.9 (42.3) | 136 | 69.9 (76.6) | 93 |
| Kalinzu18 | Moyen | 94.2 (13.1) | 3 | 161.7 (108.7) | 4 |
| Kalinzu20 | Moyen | 56.5 (62.8) | 62 | 48.7 (60.4) | 63 |
| Sonso3 | Moyen | 19.4 (23.0) | 33 | 172.3 (258.8) | 17 |
| Sonso6 | Moyen | 23.2 (88.3) | 227 | 85.5 (172.6) | 189 |
| Waibira17 | Moyen | 16.5 (22.0) | 152 | 43.4 (130.3) | 154 |
| Kalinzu10 | Dur | 62.6 (47.6) | 4 | 31.7 (57.4) | 7 |
| Sonso4 | Dur | 11.1 (14.4) | 24 | 26.6 (30.1) | 9 |
| Sonso9 | Dur | 31.3 (80.3) | 109 | 88.5 (93.6) | 64 |
| Waibira1 | Dur | 134.7 (246.1) | 15 | 131.4 (143.3) | 15 |
| Waibira18 | Dur | 74.4 (142.0) | 91 | 101.7 (165.5) | 159 |
| Wamba11 | Dur | 170.8 (408.0) | 38 | 82.7 (126.1) | 24 |
| Tous | | 60.4 (144.8) | 1801 | 54.5 (104.8) | 1980 |

**TABLE 4** Performance du modèle pour les parties du corps dans les nouvelles vidéos. Le modèle 1 contenait 1 375 images provenant de 55 vidéos de 5 communautés *Pan* étiquetées par un seul codeur, et le modèle 2 contenait 2 200 images provenant de 110 vidéos de 6 communautés *Pan* étiquetées par deux codeurs. Les vidéos ont été classées en fonction de leur difficulté sur la base des facteurs de bruit visuel présents. MAED = distance euclidienne absolue moyenne.

| | Modèle | | 1Modèle 2 | |
|---|---|---|---|---|
| Partie du corps | MAED (SD) | *n* détections | MAED (SD) | *n* détections |
| Cheville | 86.9 (131.6) | 66 | 53.8 (99.7) | 51 |
| Oreille | 40.9 (117.1) | 219 | 58.9 (118.5) | 357 |
| Coude | 109.5 (199.4) | 143 | 53.0 (96.6) | 247 |
| L'œil | 32.8 (128.8) | 389 | 26.9 (84.1) | 247 |
| Hanche | 85.4 (89.5) | 38 | 67.0 (81.7) | 105 |
| Genou | 87.7 (147.2) | 103 | 69.8 (118.8) | 77 |
| Cou | 78.2 (162.8) | 154 | 53.6 (84.4) | 170 |
| Nez | 48.7 (166.5) | 222 | 37.4 (103.5) | 140 |
| Epaule | 47.4 (80.7) | 345 | 70.0 (127.9) | 362 |
| Poignet | 116.7 (207.2) | 122 | 55.3 (81.6) | 224 |
| Tous | 60.4 (144.7) | 1801 | 54.5 (104.8) | 1980 |

la fiabilité des données fournies. Pour contribuer à ces efforts, nous avons mis les modèles de base utilisés dans cet article en libre accès et avons collaboré avec les développeurs de DeepLabCut pour permettre le marquage en ligne en libre accès des images de notre ensemble de données, qui seront régulièrement ajoutées au modèle de base afin d'améliorer les performances (voir la déclaration de disponibilité des données pour plus de détails).

L'utilisation d'outils d'estimation de la pose pour suivre les mouvements dans le comportement ani- maux ne représente qu'une première étape de l'analyse, générant de grandes quantités de données de position qui nécessitent ensuite une analyse plus approfondie pour détecter des schémas de mouvement particuliers, par exemple : un geste d'extension ou un geste d'ouverture.

trempage d'un outil à eau. Plusieurs outils de suivi et d'estimation de la pose proposent désormais des options d'analyses comportementales simples (voir tableaux S2 et S3). Là encore, les options préexistantes ne sont généralement disponibles que pour les comportements fréquemment utilisés dans les espèces de laboratoire modèle (par exemple, l'analyse de la démarche chez les rongeurs ; Adonias et al., 2019), mais certains outils intègrent désormais l'étiquetage du comportement lors du marquage des points clés des ensembles d'entraînement afin de permettre des analyses comportementales plus personnalisées (par exemple, Junior et al., 2012).

Le codage automatisé de la pose et du mouvement offre des moyens plus rapides, plus précis et plus robustes pour soutenir les approches actuelles de codage du comportement chez les primates sauvages. En outre, la génération de "big data

L'analyse des mouvements rythmiques sur des échelles de temps auparavant inaccessibles, ainsi que la disponibilité d e données vidéo collaboratives à grande échelle sur le comportement des primates (par exemple Arandjelovic et al., 2016 ; ou Hobaiter et al., 2021) nous permettent de poser de nouvelles questions et de modéliser de nouveaux processus, par exemple en explorant la variation à la fois dans l'espace, entre les populations et les espèces, et dans le temps, d'une génération à l'autre. Par exemple, les analyses de mouvements rythmiques, tels que la marche, le claquement de lèvres ou le tambour (cf. Eleuteri et al., 2022 ; Pereira, Kavanagh, et al., 2020 ; Schweinfurth et al., 2022) ; les analyses de signaux gestuels bénéficieraient de descriptions systématiques de la variation des modèles de mouvement au sein et entre les types d'action, ou de caractéristiques telles que l'accentuation et l'excitation (cf. Graham et al.., 2022 ; Grund et al., 2023) ; et les analyses de la variation des mouvements et de l'efficacité des trajectoires de mouvement pourraient être appliquées à des questions sur le développement ontogénétique et l'acquisition de l'utilisation d'outils. Nous décrivons les nouveaux outils disponibles dans ce paysage en évolution rapide et proposons des conseils pour la sélection des outils. Avec DeepWild, nous montrons que, sans nécessiter d'expertise spécifique en apprentissage automatique, l'estimation de la pose et le suivi des mouvements des primates sauvages vivant en liberté dans des environnements visuellement complexes est désormais un objectif réalisable pour les chercheurs en comportement.

## CONTRIBUTIONS DES AUTEURS

Catherine Hobaiter et Charlotte Wiltshire ont conçu les idées et la méthodologie ; Catherine Hobaiter, Tetsuro Matsuzawa et Kirsty E Graham ont recueilli les données vidéo originales ; Charlotte Wiltshire et Viola Komedová ont codé les données ; Charlotte Wiltshire, James Lewis-Cheetham et Catherine Hobaiter ont analysé les données ; Catherine Hobaiter et Charlotte Wiltshire ont dirigé la rédaction du manuscrit. Catherine Hobaiter et Charlotte Wiltshire ont dirigé la rédaction du manuscrit. Tous les auteurs ont apporté une contribution critique aux versions préliminaires et ont donné leur accord final pour la publication.

## REMERCIEMENTS

Nous remercions les éditeurs invités, le Dr Thibaud Gruber et le Dr Erica van de Waal, pour leur invitation à contribuer à cet article, et nous remercions le rédacteur en chef du journal et deux évaluateurs anonymes pour leurs conseils constructifs sur la façon de l'améliorer. Nous remercions le personnel et les communautés des stations de terrain dans lesquelles nous avons recueilli nos données en Ouganda, en République démocratique du Congo et en Guinée. Nous remercions en particulier le personnel de la Budongo Conservation Field Station pour le soutien qu'il nous a apporté pendant des années et qui a facilité la collecte de données vidéo dans le cadre de projets entre 2004 et 2022. Nous remercions les professeurs Hashimoto et Furuichi pour leur autorisation de collecter des données vidéo sur les sites de Kalinzu et de Wamba. Nous remercions le Dr. Aly Garpard Soumah, de l'Institut de Recherche Environnementale de Bossou (IREB), pour l'autorisation de collecter des données sur le site de Bossou, qui a fonctionné en continu grâce à la collaboration entre les chercheurs de l'Institut de Recherche sur les Primates de l'Université de Kyoto, dirigé par Yukimaru Sugiyama, et les chercheurs guinéens, y compris Jérémie Koman, Soh Pagh, Kohl et Kohl, pour la collecte de données sur le site de Bossou : Jeremie Koman, Soh Pletah Bonimy, Bakary Coulibary, Tamba Tagbino, Makan Kourouma, Mamadou Diakite, Cécé Kolié, Iba Conde et Sekou Moussa Keita. Nous remercions également les autorités guinéennes qui ont donné leur autorisation au chercheur à long terme : Ministre de l'Enseignement Supérieur et de la Recherche Scientifique, et Direction Générale de la Recherche Scientifique et de l'Innovation Technologique. Nous remercions Alexander Mielke pour son aide dans la réalisation de l'étude.

R-code pour les figures. Tous les projets de recherche en Ouganda ont été menés avec l'autorisation de l'Uganda Wildlife Authority et du Conseil national ougandais pour la science et la technologie. Tous les projets de recherche ont été menés avec l'autorisation éthique du Comité d'éthique et de bien-être animal de l'Université de St Andrews. Nous remercions les développeurs des programmes que nous avons évalués et les communautés d'apprentissage automatique pour leur travail et la patience dont ils ont fait preuve en répondant à nombre de nos questions, ainsi que notre groupe de laboratoire pour ses discussions constructives. Ce projet a été financé par le 8e programme-cadre de l'Union européenne, Horizon 2020 (numéro de convention de subvention : 802719) et le St Andrews Restarting Research Funding Scheme (2020).

## DÉCLARATION DE CONFLIT D'INTÉRÊTS

Les auteurs déclarent n'avoir aucun conflit d'intérêt.

## DÉCLARATION DE DISPONIBILITÉ DES DONNÉES

Les modèles DeepWild ainsi que toutes les données et le code utilisés dans cet article peuvent être téléchargés dans notre dépôt Github https://github.com/Wild-Minds/DeepWild, qui est archivé dans Zenodo à https://doi.org/10.5281/zenodo.7414432 (Wiltshire et al., 2022). Des informations sur l'utilisation des données de la Great Ape Video Database sont disponibles à https://doi.org/10.5281/zenodo.5600472 (Hobaiter et al., 2021). Une interface en ligne à accès libre pour marquer des images supplémentaires est disponible ici : https://contrib.deeplabcut.org/label Les images reçues par ce biais seront utilisées pour mettre à jour régulièrement le modèle de base dans notre Github et partagées avec le zoo modèle DeepLabCut.

## ORCID

*Viola Komedová* https://orcid.org/0000-0001-5554-7271 *Tetsuro Matsuzawa* https://orcid.org/0000-0002-8147-2725 *Kirsty E. Graham* https://orcid.org/0000-0002-7422-7676 *Catherine Hobaiter* https://orcid.org/0000-0002-3893-0524

## R EFER EN CE S

Achour, B., Belkadi, B., Filali, I., Laghrouche, M. et Lahdir, M. (2020). Analyse d'images pour l'identification individuelle et la surveillance du comportement alimentaire des vaches laitières basée sur des réseaux de neurones convolutionnels (CNN). *Biosystems Engineering*, *198*(1), 31-49. https://doi.org/10.1016/j.biosystemseng.2020.07.019

Adonias, A. F., Ferreira-Gomes, F., Allonso, R., Neto, F. et Cardoso, J. S. (2019). Vers l'analyse automatique de la marche du rat dans des conditions d'illumination sous-optimales. *Reconnaissance des formes et analyse d'images*, 247-259. https://doi.org/10.1007/978-3-030-31321-0_22

Ahumada, J. A., Fegraus, E., Birch, T., Flores, N., Kays, R., O'Brien, T. G., Palmer, J., Schuttler, S., Zhao, J. Y., & Jetz, W. (2019). Wildlife insights : Une plateforme pour maximiser le potentiel des données de pièges photographiques et d'autres capteurs pas- sifs sur la faune pour la planète. *Environmental Conservation*, *47*(1), 1-6. https://doi.org/10.1017/S0376892919000298

Altschul, D. M., Beran, M. J., Bohn, M., Call, J., De Troy, S., Duguid, S. J., Egelkamp, C. L., Fichtel, C., Fischer, J., Flessert, M., Hanus, D., Haun, D. B. M., Haux, L. M., Hernandez-Aguilar, R. A., Herrmann, E., Hopper, L. M., Joly, M., Kano, F., Keupp, S., ... Watzek, J. (2019). Établir une infrastructure pour la collaboration dans la recherche sur la cognition des primates. *PLoS ONE*, *14*, e0223675. https://doi.org/10.1371/journal. pone.0223675

Arandjelovic, M., Stevens, C. R., McCarthy, M. S., Dieguez, P., Kalan, A. K., Maldonado, N., Boesch, C. et Kuehl, H. S. (2016). Chimp&See : Une plateforme de science citoyenne en ligne pour l'annotation du comportement, de la démographie et de l'identification individuelle des chimpanzés à grande échelle et à distance à l'aide de pièges à caméra. *PeerJ Preprints*. https://doi.org/10.7287/ peerj.preprints.1792v1

Bain, M., Nagrani, A., Schofield, D., Berdugo, S., Bessa, J., Owen, J., Hockings, K. J., Matsuzawa, T., Hayashi, M., Biro, D., & Carvalho, S. (2021). Automated audiovisual behaviour recognition in wild pri- mates. *Science Advances*, *7*(46), eabi4883.

Bala, P. C., Eisenreich, B. R., Yoo, S. B. M., Hayden, B. Y., Park, H. S. et Zimmerman, J. (2020). Automated markerless pose estimation in freely moving macaques with OpenMonkeyStudio (Estimation automatisée de la pose sans marqueur chez des macaques en mouvement libre avec OpenMonkeyStudio). *Nature Communications*, *11*, 4560. https://doi.org/10.1038/s41467-020-18441-

Beery, S., Morris, D. et Yang, S. (2019). Pipeline efficace pour l'examen des images de pièges à caméra. *arXiv*, 1907.06772. https://doi.org/10.48550/arXiv.1907.06772.

Bergamini, L., Porrello, A., Capobianco Dondona, A., Del Negro, E., Mattioli, M., D'Alterio, A., & Calderara, S. (2018). Em- litation multi-vues pour la ré-identification des bovins. In *4th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. https://doi.org/10.1109/SITIS.2018.00036

Berger-Wolf, T. Y., Rubenstein, D. I., Stewart, C. V., Holmberg, J. A., Parham, J., Menon, S., Crall, J., Van Oast, J., Kiciman, E., & Joppa, L. (2017). Wildbook : Crowdsourcing, computer vision, and data science for conservation. *arXiv*. https://doi.org/10.48550/arXiv.1710.08880

Bianco, M. J., Gerstoft, P., Traer, J., Ozanich, E., Roch, M. A., Gannot, S., & Deledalle, C.-A. (2019). Apprentissage automatique en acoustique : Théorie et applications. *The Journal of the Acoustical Society of America*, *146*(1), 3590-3628. https://doi.org/10.1121/1.5133944

Burton, A. C., Neilson, E., Moreira, D., Ladle, A., Steeweg, R., Fisher, J. T., Bayne, E., & Boutin, S. (2015). Le piégeage photographique de la faune : A review and recommendations for linking surveys to ecological processes. *The Journal of Applied Ecology*, *52*(3), 675-685.

Clemensson, E., Abbaszadeh, M., Fanni, S., Espa, E., & Cenci, M. A. (2020). Suivi de rats dans des chambres de conditionnement opérant à l'aide d'une caméra vidéo maison polyvalente et de DeepLabCut. *Journal of Visualized Experiments*, *160*, e61409. https://doi. org/10.3791/61409

DeepLabCut Github. (2021a). https://github.com/DeepLabCut/DeepLabCut/blob/master/docs/UseOverviewGuide.md

DeepLabCut Github. (2021b). https://deeplabcut.github.io/DeepLabCut/docs/intro.html

DeepLabCut Github. (2021c). https://github.com/DeepLabCut/DeepLabCut/blob/master/docs/installation.md

Doran, D. M. (1993). Comparative locomotor behavior of chimpan- zees and bonobos : The influence of morphology on locomotion. *American Journal of Physical Anthropology*, *91*(1), 83-98.

Eggeling, W. J. (1947). Observations on the ecology of the Budongo rain forest, Uganda. *The Journal of Ecology*, *34*(1), 20-87.

Eleuteri, V., Henderson, M., Soldati, A., Badihi, B., Zuberbühler, K. et Hobaiter, C. (2022). The form and function of chimpanzee but- tress drumming. *Animal Behaviour*, *192*, 189-205. https://doi. org/10.1016/j.anbehav.2022.07.013

Estrada, A., Garber, P. A., Rylands, A. B., Roos, C., Fernandez-Duque, E., Di Fiore, A., Nekaris, K. A. I., Nijman, V., Heymann, E. W., Lambert, J. E., Rovero, F., Barelli, C., Setchell, J. M., Gilllespie, T. R., Mittermeier, R. A., Arregoitia, L. V., de Guinea, M., Gouveia, S., Dobrovolski, R., ... Li, B. (2017). Crise d'extinction imminente des primates du monde : Pourquoi les primates sont importants. *Science Advances*, *3*(1), e1600946. https://doi.org/10.1126/sciadv.1600946

Gadea, G. H. (2021). https://guillermohidalgogadea.com/openlabnote book/

Gal, A., Saragosti, J., & Kronauer, D. J. C. (2020). anTraX, a software pack- age for high-throughput video tracking of color-tagged insects. *eLife*, e58145. https://doi.org/10.7554/eLife.58145

Geuther, B. Q., Deats, S. P., Fox, K. J., Murray, S. A., Braun, R. E., White, J. K., Chesler, E. J., Lutz, C. M. et Kumar, V. (2019). Suivi robuste de la souris dans des environnements complexes à l'aide de réseaux neuronaux. *Communications Biology*, *2*, 124. https://doi.org/10.1038/ s42003-019-0362-1

Gosztolai, A., Günel, S., Lobato-Ríos, V., Abrate, M. P., Morales, D., Rhodin, H., Fua, P. et Ramdya, P. (2021). LiftPose3D, une approche basée sur l'apprentissage profond pour transformer les poses bidimensionnelles en poses tridimensionnelles chez les animaux de laboratoire. *Nature Methods*, *18*, 975-981. https:// doi.org/10.1038/s41592-021-01226-z

Graham, K. E., Badihi, G., Safryghin, A., Grund, C. et Hobaiter, C. (2022). A socio-ecological perspective on the gestural communication of great ape species, individuals, and social units (Une perspective socio-écologique sur la communication gestuelle des espèces de grands singes, des individus et des unités sociales). *Ethology, Ecology, & Evolution*, *34*(2), 235-259.

Grund, C., Badihi, G., Graham, K. E., Safryghin, A. et Hobaiter, C. (2023). GesturalOrigins : A bottom-up framework for establishing system- atic gesture data across ape species. *Behaviour Research Methods*. https://doi.org/10.3758/s13428-023-02082-9

Günel, S., Rhodin, H., Morales, D., Campagnolo, J., Ramdya, P., & Fua, R. (2019). DeepFly3D, une approche basée sur l'apprentissage profond pour le suivi 3D des membres et des appendices chez la *drosophile* adulte attachée. *eLife*. *8*. e48571. https://doi.org/10.7554/eLife.48571

Guo, S., Xu, P., Miao, Q., Shao, G., Chapman, C. A., Chen, X., He, G., Fang, D., Zhang, H., Sun, Y., Shi, Z., & Li, B. (2020). Automatic identification of individual primates with deep learning techniques. *iScience*, *23*, e101412. https://doi.org/10.1016/j.isci.2020.101412

Habe, H., Takeuchi, Y., Terayama, K., & Sakagami, M. (2021). Pose estimation of swimming fish using NACA airfoil model for collective behavior analysis. *Journal of Robotics and Mechatronics*, *33*(3), 547–555. https://doi.org/10.20965/jrm.2021.p0547

Hashimoto, C., Tashiro, Y., Kimura, D., Enomoto, T., Ingmanson, E. J., Idani, G., & Furuichi, T. (1998). Habitat use and ranging of wild bonobos (*Pan paniscus*) at Wamba. *International Journal of Primatology*, *19*(1), 1045–1060.

Hastie, T., Tibshirani, R. et Friedman, J. H. (2001). *The elements of statisti- cal learning : Data mining, inference, and prediction*. Springer.

Haucke, T., Kühl, H. S., Hoyer, J. et Steinhage, V. (2021). Overcoming the distance estimation bottleneck in estimating animal abundance with camera traps. *arXiv*. https://doi.org/10.48550/arXiv.2105.04244

Hayden, B. Y., Park, H. S. et Zimmermann, J. (2021). Automated pose estimation in primates. *American Journal of Primatology*, *84*(10), e23348. https://doi.org/10.1002/ajp.23348

Hobaiter, C., Gal Badihi, G., de Melo Daly, G. B., Eleuteri, V., Graham, K. E., Grund, C., Henderson, M., Rodrigues, E. D., Safryghin, A., Soldati, A. et Wiltshire, C. (2021). Base de données vidéo du dictionnaire des grands singes. *Zenodo*. https://doi.org/10.5281/zenodo.5600471

Hsu, A. I. et Yttri, E. A. (2021). B-SOiD, an open-source unsupervised al- gorithm for identification and fast prediction of behaviors (B-SOiD, un algorithme non supervisé à code source ouvert pour l'identification et la prédiction rapide des comportements). *Nature*, *12*, 5188. https://doi.org/10.1038/s41467-021-25420-x

Islam, M. D., Mo, J. et Sattar, J. (2021). Robot-to-robot relative pose es- timation using humans as markers. *Autonomous Robots*, *45*(1), 579-593. https://doi.org/10.1007/s10514-021-09985-6

Joska, D., Clark, L., Muramatsu, N., Jericevich, R., Nicholls, F., Mathis, A., Mathis, M. W., & Patel, A. (2021). AcinoSet : A 3D pose estima- tion dataset and baseline models for cheetahs in the wild. In *IEEE International Conference on Robotics and Automation (ICRA)* (pp. 13901-13908). https://doi.org/10.1109/ICRA48506.2021.9561338

Jungers, W. L. et Susman, R. L. (1984). Body size and skeletal allometry in African apes. In R. L. Sussman (Ed.), *The pygmy chimpanzees* (pp. 137-177). Springer.

Junior, C. F. C., Pederiva, C. N., Bose, R. C., Garcia, V. A., Lino-de-Oliveira, C., & Marino-Neto, J. (2012). ETHOWATCHER : Validation d'un outil d'analyse comportementale et de suivi vidéo chez les animaux de laboratoire. *Computers in Biology and Medicine*, *42*(2), 257-264.

Kane, G. A., Lopes, G., Saunders, J. L., Mathis, A. et Mathis, W. M. (2020). Rétroaction en boucle fermée, en temps réel et à faible latence, à l'aide d'un système d'information sur la santé.

suivi de la posture sans marqueur. *eLife*, *9*, e61909. https://doi.org/10.7554/eLife.61909

Karashchuk, P., Rupp, K. L., Dickinson, E. S., Walling-Bell, S., Sanders, E., Bingni, E. A., Brunton, W. et Tuthill, J. C. (2021). Anipose : A tool- kit for robust markerless 3D pose estimation. *Cell Reports*, *36*, 13. https://doi.org/10.1016/j.celrep.2021.109730

Kellenberger, B., Tuia, D. et Morris, D. (2020). AIDE : Accelerating image-based ecological surveys with interactive machine learning (Accélérer les enquêtes écologiques basées sur l'image avec l'apprentissage automatique interactif). *Methods in Ecology and Evolution*, *11*(12), 1716-1727. https://doi.org/10.1111/2041-210X.13489

Khan, N. U. et Wan, W. (2018). Une revue de l'estimation de la pose humaine à partir d'une seule image. In *2018 International Conference on Audio, Language and Image Processing (ICALIP)*. https://doi.org/10.1109/ICALIP.2018.8455796

Kiel, M., Berh, D., Daniel, J., Otto, N., Steege, A. T., Jiang, X., Liebau, E., & Risse, B. (2018). Un traqueur de vers polyvalent basé sur la FIM. bioRxiv. https://doi.org/10.1101/352948

Körschens, M. et Denzler, J. (2019). ELPephants : A fine-grained data- set for elephant Re-identification. In *2019 IEEE/CVF International Conference on Computer Vision Workshop, (ICCVW)*. https://doi.org/10.1109/ICCVW.2019.00035

Labuguen, R., Matsumoto, J., Negrete, S. B., Nishimaru, H., Nishijo, H., Takada, M., Go, Y., Inoue, K. et Shibata, T. (2021). MacaquePose : A novel "in the wild" macaque monkey pose dataset for markerless motion capture. *Frontiers in Behavioural Neuroscience*, *14*, 581154. https://doi.org/10.3389/fnbeh.2020.581154

Lauer, J., Zhou, M., Ye, S., Menegas, W., Nath, T., Rahman, M. M., Santo, V. D., Soberanes, D., Feng, G., Murthy, V. N. et Lauder, G. (2021). Multi-animal pose estimation and tracking with DeepLabCut. bioRxiv. https://doi.org/10.1101/2021.04.30.442096

Li, S., Li, J., Tang, H., Qian, R. et Lin, W. (2020). ATRW : A benchmark for Amur tiger re-identification in the wild. *arXiv*. https://doi.org/10.1145/3394171.3413569

Marks, M., Qiuhan, J., Sturman, O., von Ziegler, L., Kollmorgen, S., von der Behrens, W., Mante, V., Bohacek, J. et Yanik, M. F. (2022). Deep- learning based identification, tracking, pose estimation, and behavior classification of interacting primates and mice in complex environnements. *bioRxiv*. https://doi.org/10.1101/2020.10.26.355115

Mathis, A., Mamidanna, P., Curry, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut : Esti- mation de pose sans marqueur de parties du corps définies par l'utilisateur avec l'apprentissage profond. *Nature Neuroscience*, *21*(1), 1281-1289.

Matsuzawa, T., Humle, T. et Sugiyama, Y. (2011). *Les chimpanzés du Bossou et du Nimba*. Springer Nature.

Mei, J., Hwang, J.-N., Romain, S., Rose, C., Moore, B. et Magrane, K. (2021). Absolute 3d pose estimation and length measurement of severely deformed fish from monocular videos in longline fishing. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. https://doi.org/10.1109/ ICASSP39728.2021.9414803

Munch, K. L., Wapstra, E., Thomas, S., Fisher, M. et Sinn, D. L. (2019). Qu'est-ce que nous mesurons ? Les novices sont d'accord entre eux (mais pas toujours avec les experts) dans leur évaluation du comportement des chiens. *Ethology*, *125*(4), 203-211.

Narouzzadeh, M. S., Nguyen, A., Kosala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Identifier, compter et décrire automatiquement les animaux sauvages dans les images de pièges photographiques avec l'apprentissage profond. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(25), E5716-E5725. https://doi.org/10.1073/pnas.1719367115

Nath, T., Mathis, A., Chen, A. C., Patel, A., Bethge, M., & Mathis, M. W. (2019). Utilisation de DeepLabCut pour l'estimation de la pose sans marqueur 3D à travers les espèces et les comportements. *Nature Protocols*, *14*, 2152-2176. https://doi.org/10.1038/s41596-019-0176-

Panadeiro, V., Rodriguez, A., Henry, J., Wlodkowic, D. et Andersson, M. (2021). Examen de 28 logiciels gratuits de suivi des animaux : Caractéristiques et limites actuelles. *Nature*, *50*(1), 246-254. https://doi.org/10.1038/s41684-021-00811-1

Pathak, S. D., Ng, L., Wyman, B., Fogarasi, S., Racki, S., Oelund, J. C., Spark, B. et Chalana, V. (2003). Quantitative image analysis : Software systems in drug development trails. *Drug Discovery Today*, *8*(10), 451-458.

Pereira, A. S., Kavanagh, E., Hobaiter, C., Slocombe, K. E., & Lameira, A. R. (2020). Les claquements de lèvres des chimpanzés confirment la continuité primate pour l'évolution du rythme de la parole. *Biology Letters*, *16*, 20200232. https:// doi.org/10.1098/rsbl.2020.0232

Pereira, T. D., Aldarondo, D. E., Willmore, L., Kislin, M., Wang, S. S.-H., Murthy, M. et Shaevitz, J. W. (2019). Estimation rapide du poste animal à l'aide de réseaux neuronaux profonds. *Nature Methods*, *16*, 117-125. https://doi.org/10.1038/s41592-018-0234-5

Pereira, T. D., Tabris, N., Li, J., Ravindranath, S., Papdoyannis, E. S., Wang, Z. Y., Turner, D. M., McKenzie-Smith, G., Kocker, S. D., Falkner, A. L., Shaevitz, J. W. et Murthy, M. (2020). SLEAP : Multi-animal pose tracking. *bioRxiv*. https://doi.org/10.1101/2020.08.31.276246

Prakash, C., Kumar, R., Mittal, N. et Raj, G. (2018). Vision based identi- fication of joint coordinates for marker-less gait analysis. *Procedia Computer Science*, *132*(1), 68-75.

Premarathna, K. S. P., Rathnayaka, R. M. K. T. et Charles, J. (2020). An elephant detection system to prevent human-elephant conflict and tracking of elephant using deep learning. In *5th International Conference on Information Technology Research (ICITR)*. https://doi. org/10.1109/ICITR51448.2020.9310798

Rohan, A., Rabah, M., Hosny, T. et Kim, S.-H. (2020). Human pose estimation-based real-time gait analysis using convolutional neural network. *IEEE Access*, *8*(1), 191542-191550. https://doi. org/10.1109/ACCESS.2020.3030086

Sarafianos, N., Boteanu, B., Ionescu, B. et Kakadiaris, I. A. (2016). 3D human pose estimation : A review of the literature and analysis of covariates. *Computer Vision and Image Understanding*, *152*(1), 1-20. https://doi.org/10.1016/j.cviu.2016.09.002

Schneider, S., Taylor, G. W., Linquist, S., & Kremer, S. C. (2018). Past, present and future approaches using computer vision for animal re-identificationfromcameratrapdata. *Methodsin Ecologyand Evolution*, *10*(4), 461-470. https://doi.org/10.1111/2041-210X.13133

Schofield, D., Nagrani, A., Zisserman, A., Hayashi, M., Matsuzawa, T., Biro, D., & Carvalho, S. (2019). Reconnaissance de visage de chimpanzé à partir de vidéos dans la nature à l'aide de l'apprentissage profond. *Science Advances*, *5*(9), eaaw0736. https://doi.org/10.1126/sciadv.aaw0736

Schweinfurth, M. K., Baldridge, D. B., Finnerty, K., Call, J., & Knoblich, G. K. (2022). Coordination interindividuelle chez les chimpanzés marcheurs. *Current Biology*, *32*(23), 5138-5143.

Sheppard, K., Gardin, J., Sabnis, G. S., Peer, A., Darell, M., Deats, S., Geuther, B., Lutz, C. M. et Kumar, V. (2022). Stride-level analy- sis of mouse open field behavior using deep-learning-based pose estimation. *Cell Reports*, *38*(2), 110231. https://doi.org/10.1016/ j.celrep.2021.110231

Steenweg, R., Hebblewhite, M., Kays, R., Ahumada, J., Fisher, J. T., Burton, C., Townsend, S. E., Carbone, C., Rowcliffe, J. M., Whittington, J., Brodie, J., Royle, J. A., Switalski, A., Clevenger, A. P., Heim, N., & Rich, L. N. (2016). Scaling-up camera traps : Surveillance de la biodiversité de la planète avec des réseaux de capteurs à distance. *Frontiers in Ecology and Environment*, *15*(1), 26-34. https://doi.org/10.1002/fee.1448

Swann, D. E., Kawanishi, K., & Palmer, J. (2011). Evaluating types and features of camera traps in ecological studies : A guide for research- ers. Dans A. F. O'Connell, J. D. Nichols, & K. U. Karanth (Eds.), *Camera traps in animal ecology*. Springer. https://doi.org/10.1007/978-4- 431-99495-4_3

Terada, S., Nackoney, J., Sakamaki, T., Mulavwa, M. N., Yumoto, T., & Furuichi, T. (2015). Habitat use of bonobos (*Pan paniscus*) at Wamba : Sélection des types de végétation pour l'habitat, l'alimentation et le sommeil nocturne. *American Journal of Primatology*, *77*(6), 701-713. https://doi.org/10.1002/ajp.22392

Tsuruo, A., Ringhofer, M., Yamamoto, S. et Ikeda, K. (2020). Modèle mathématique de l'interaction entre le cheval et le cavalier pendant le saut à cheval. *2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 939-943.

Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., Mathis, A., Mathis, M. W., van Langevelde, F., Burghardt, T., Kays, R., Klinck, H., Wikelski, M., Couzin, I. D., van Horn, G., Crofoot, M. C., Stewar, C. V., & Berger-Wolf, T. (2022). Perspectives in machine learning for wildlife conservation (Perspectives de l'apprentissage automatique pour la conservation de la faune). *Nature Communications*, *13*, 792. https://doi.org/10.1038/s41467-022-27980-y

Wäldchen, J. et Mäder, P. (2018). Machine learning for image based spe- cies identification. *Methods in Ecology and Evolution*, *9*(11), 2216- 2225. https://doi.org/10.1111/2041-210X.13075

Walter, T. et Couzin, I. D. (2021). TRex, a fast multi-animal tracking system with markerless identification, and 2D estimation of pos- ture and visual fields. *eLife*, *10*, e64000. https://doi.org/10.7554/ eLife.640000

Wang, T.-H. et Lien, J.-J. (2009). Facial expression recognition system based on rigid and non-rigid motion separation and 3D pose estima- tion. *Pattern Recognition*, *42*(5), 962-977. https://doi.org/10.1016/ j.patcog.2008.09.035

Weinstein, B. G. (2017). Une vision informatique pour l'écologie animale. *Journal of Animal Ecology*, *87*(3), 533-545. https://doi.org/10.1111/1365-2656.12780

Whytock, R. C., Świeżewski, J., Zwerts, J. A., Bara-Słupski, T., Pambo, A. F. K., Rogala, M., Bahaa-el-din, L., Boekee, K., Brittain, S., Cardoso, A. W., Henschel, P., Lehmann, D., Momboua, B., Opepa, C. K., Orbell, C., Pitman, R. T., Robinson, H. S. et Abernethy, K. A. (2021). Robust ecological analysis of camera trap data labelled by a machine learning model (Analyse écologique robuste de données de pièges photographiques étiquetées par un modèle d'apprentissage automatique). *Methods in Ecology and Evolution*, *12*(6), 1080-1092. https://doi.org/10.1111/2041-210X.13576

Willi, M., Pitman, R. T., Cardoso, A. W., Locke, C., Swanson, A., Boyer, A., Veldthuis, M., & Fortson, L. (2018). Identifier les spe- cies animales dans les images de pièges photographiques en utilisant l'apprentissage profond et la sci- ence citoyenne. *Methods in Ecology and Evolution*, *10*(1), 80-91. https://doi. org/10.1111/2041-210X.13099

Wiltshire, C., Lewis-Cheetham, J., Komedová, V., Matsuzawa, T., Graham, K. E. et Hobaiter, C. (2022). WildMinds/DeepWild : Deep Wild. Zenodo. https://doi.org/10.5281/zenodo.7414432

Xu, Z. et Cheng, X. E. (2017). Suivi du poisson zèbre à l'aide de réseaux neuronaux convolutifs. *Scientific Reports*, *7*(1), 42815. https://doi. org/10.1038/srep42815

Yu, X., Wang, J., Kays, R., Jansen, P. A., Wang, T., & Huang, T. (2013). Identification automatisée des espèces animales dans les images de pièges photographiques. *EURASIP Journal on Image and Video Processing*, *52*(1). https://doi. org/10.1186/1687-5281-2013-52

Yu, X., Zhou, H., Wu, L. et Liu, Q. (2011). High-performance drosophila movement tracking. In *2011 Third Chinese Conference on Intelligent Visual Surveillance.* https://doi.org/10.1109/IVSurv.2011.6157027

## INFORMATIONS COMPLÉMENTAIRES

Des informations complémentaires sont disponibles en ligne dans la section Informations complémentaires à la fin de cet article.

**Tableau S1.** Exemples d'outils d'apprentissage automatique pour l'identification automatisée des espèces et des individus. Pour douze outils couramment utilisés, nous décrivons les espèces pour lesquelles ils ont été utilisés, si une formation supplémentaire de l'utilisateur est nécessaire, et s'ils peuvent ré-identifier les individus.**Tableau S2.** Liste complète des outils d'apprentissage automatique actuellement disponibles pour automatiser le suivi des animaux. Nous indiquons s'ils sont capables de suivre plusieurs animaux, s'ils ont été utilisés pour des données "sauvages", c'est-à-dire pour des animaux vivant en liberté dans des environnements non contrôlés, y compris des environnements naturels à grande échelle, et s'ils sont capables d'identifier des individus.

Nous avons répertorié les espèces pour lesquelles l'outil est actuellement utilisé, le style de suivi et les commentaires sur les exigences en matière de fonctionnalité. Nous énumérons les espèces pour lesquelles l'outil a été utilisé, le style de suivi et des commentaires supplémentaires sur la fonctionnalité ou les exigences. Cette liste est conçue pour être utilisée en conjonction avec la figure 1, qui fournit des orientations décisionnelles sur le ou les outils susceptibles de convenir à une utilisation particulière.**Tableau S3.** Description détaillée de la facilité d'utilisation et de la fonctionnalité des outils actuellement disponibles pour l'estimation automatisée de la pose des animaux, toutes espèces confondues. Nous prenons en compte la fonction spécifique de l'outil, s'il fournit une interface utilisateur graphique (GUI), si les utilisateurs doivent comprendre des langages de codage tels que Python, si le suivi multi-animal (MA) est disponible, le format de sortie des données, si l'utilisateur a spécifié les caractéristiques à suivre, s'il nécessite l'accès à une carte graphique spécifique, et si, en l'absence d'une telle carte, l'outil ne peut pas être utilisé ; s'il est compatible avec Google Colaboratory (colab) ; s'il a déjà été utilisé avec des individus "sauvages" ou en liberté dans des environnements visuellement complexes ; quelle documentation et quels didacticiels sont actuellement disponibles ; et tout autre avantage ou inconvénient particulier associé à la fonctionnalité actuelle. Il convient de noter que nous n'avons pas inclus trois outils supplémentaires qui ne sont actuellement adaptés qu'à l'estimation de la pose avec des espèces spécifiques : Open Monkey Studio (macaques), AlphaTracker (souris) et DeepFly3d (mouches).**Tableau S4.** Classification des nouvelles vidéos en fonction du bruit visuel.**Vidéo S1.** Modèle 1 de suivi de la vidéo test Sonso10. Dans cette vidéo, quatre chimpanzés d'Afrique de l'Est sont assis dans le sous-bois, les deux individus immatures se battent et jouent. Il y a des mouvements de la caméra et du sous-bois, et les individus se déplacent d'avant en arrière l'un par rapport à l'autre. Certaines images de cette vidéo ont été incluses dans l'ensemble d'entraînement, et le suivi est très bon tout au long du processus. La vidéo est disponible ici : https://tinyurl.com/DeepWildvideos.

**Vidéo S2.** Modèle 2 de suivi de la vidéo test Sonso10. Dans cette vidéo, quatre chimpanzés d'Afrique de l'Est sont assis dans le sous-bois, les deux individus immatures se battent et jouent. Il y a des mouvements de la caméra et du sous-bois, et les individus se déplacent d'avant en arrière l'un par rapport à l'autre. Certaines images de cette vidéo ont été incluses dans l'ensemble d'entraînement. Le suivi est excellent tout au long du processus, avec une stabilité accrue des points par rapport au modèle 1. La vidéo est disponible ici : https://tinyurl.com/ DeepWildvideos.

**Vidéo S3.** Modèle 1 de suivi de la vidéo test Wamba10. Dans cette vidéo, trois bonobos sont assis dans un sous-bois dense. La vidéo a été classée comme "difficile" et certaines images de cette vidéo ont été incluses dans l'ensemble d'apprentissage. Les principaux points clés sont bien suivis tout au long de la vidéo, mais certaines parties du corps sont parfois manquées ou perdues, et le modèle confond certaines parties de l'environnement avec les bonobos, ajoutant des points clés hors de propos qui nécessiteraient un nettoyage manuel. La vidéo est disponible ici : https://tinyurl.com/DeepWildvideos.

**Vidéo S4.** Modèle 2 de suivi de la vidéo test Wamba10. Comme dans la vidéo S3, trois bonobos sont assis dans un sous-bois dense. La vidéo a été classée comme "difficile" et certaines images de cette vidéo ont été incluses dans l'ensemble d'apprentissage. Bien qu'il subsiste quelques problèmes de suivi, par exemple des points clés mal placés, les performances de suivi sont nettement supérieures à celles du modèle 1. La vidéo est disponible ici : https://tinyurl.com/ DeepWildvideos.

**Vidéo S5.** Suivi par le modèle 1 de la nouvelle vidéo Waibira17. Deux chimpanzés d'Afrique de l'Est marchent près de la caméra. La vidéo est courte et entièrement nouvelle pour le modèle, le suivi est très bon tout au long de la vidéo. La vidéo est disponible ici : https://tinyurl.com/DeepWildvideos

**Vidéo S6.** Suivi par le modèle 2 de la nouvelle vidéo Waibira17. Comme dans la vidéo S5, deux chimpanzés d'Afrique de l'Est s'approchent de la caméra. La vidéo est courte et entièrement nouvelle pour le modèle. Le suivi est excellent tout au long de la vidéo, avec une amélioration du suivi du deuxième individu partiellement obscurci par rapport au modèle 1. La vidéo est disponible ici : https:// tinyurl.com/DeepWildvideos.

**Vidéo S7.** Modèle 1 de suivi d'une nouvelle vidéo Sonso6. Trois chimpanzés d'Afrique de l'Est se trouvent dans une zone ouverte près de la route. L'un d'entre eux traverse la route en courant, la caméra fait un panoramique et les deux individus sont vus l'un à côté de l'autre dans la forêt. La vidéo est entièrement nouvelle pour le modèle et le suivi est très bon tout au long de la vidéo. La vidéo est disponible ici : https://tinyurl.com/DeepWildvideos.

**Vidéo S8.** Modèle 2 de suivi de la nouvelle vidéo Sonso6. Comme dans la vidéo S7, trois chimpanzés d'Afrique de l'Est se trouvent dans une zone ouverte près de la route. L'un d'entre eux traverse la route en courant, la caméra fait un panoramique et les deux individus sont vus l'un à côté de l'autre dans la forêt. La vidéo était entièrement nouvelle pour le modèle et le suivi est excellent tout au long, avec une stabilité accrue des détections par rapport au modèle 1. La vidéo est disponible ici : https://tinyurl.com/DeepWildvideos.

**Vidéo S9.** Modèle 1 de suivi de la nouvelle vidéo Sonso4. Comme dans la vidéo S7, trois chimpanzés d'Afrique de l'Est se trouvent sur une branche dans la canopée. Les individus sont rétroéclairés, il y a des obstacles à la vue, des individus qui se chevauchent, des changements d'angle de caméra et des zooms. La vidéo a été classée comme "difficile" et était entièrement nouvelle pour le modèle. Les performances de suivi varient d'un individu à l'autre, mais sont médiocres pour les deux individus de droite. Il y a très peu de points clés mal placés, mais certaines parties du corps sont systématiquement non reconnues ou mal identifiées. Vidéo est disponible ici : https://tinyurl.com/DeepWildvideos.

**Vidéo S10.** Modèle 2 de suivi de la nouvelle vidéo Sonso4. Comme dans la vidéo S7, trois chimpanzés d'Afrique de l'Est se trouvent sur une branche dans la canopée. Les individus sont rétroéclairés, il y a des obstacles à la vue, des individus qui se chevauchent, des changements d'angle de caméra et des zooms. La vidéo a été classée comme "difficile" et était entièrement nouvelle pour le modèle. Les performances de suivi varient d'un individu à l'autre, mais restent médiocres pour les deux individus de droite. Toutefois, on constate une nette amélioration par rapport au modèle 1, avec moins d'erreurs d'identification et un suivi plus stable des points clés pour les individus les plus difficiles. La vidéo est disponible ici : https://tinyurl. com/DeepWildvideos

**Vidéo S11.** Modèle 1 de suivi de la vidéo test Sonso5. Deux chimpanzés d'Afrique de l'Est traversent une clairière dans la forêt. La zone est bien éclairée et il y a peu d'obstacles. Certaines images de cette vidéo ont été incluses dans l'ensemble d'entraînement, le suivi est satisfaisant dans l'ensemble. La vidéo est disponible ici : https://tinyurl.com/DeepWildvideos

**Vidéo S12.** Modèle 2 de suivi de la vidéo test Sonso5. Deux chimpanzés d'Afrique de l'Est traversent une clairière dans la forêt. La zone est bien éclairée et il y a peu d'obstacles. Certaines images de cette vidéo ont été incluses dans l'ensemble d'entraînement, le suivi est satisfaisant dans l'ensemble. La vidéo est disponible ici : https://tinyurl.com/DeepWildvideos.

**Comment citer cet article :** Wiltshire, C., Lewis-Cheetham, J., Komedová, V., Matsuzawa, T., Graham, K. E., & Hobaiter, C. (2023). DeepWild : Application de l'outil d'estimation de la pose DeepLabCut pour le suivi du comportement des chimpanzés et des bonobos sauvages. *Journal of Animal Ecology*, *92*, 1560-1574. https://doi.org/10.1111/1365-2656.13932