The DeepFaune initiative: a collaborative effort towards the automatic identification of French fauna in camera-trap images

Noa Rigoudy[1,2], Abdelbaki Benyoub[3], Aurélien Besnard[1], Carole Birck[4], Yoann Bollet[5], Yoann Bunz[3], Jules Chiffard Carriburu[6], Gérard Caussimont[7], Elias Chetouane[8,9], Pierre Cornette[10], Anne Delestrade[11], Nina De Backer[12], Lucie Dispan[13], Maden Le Barh[13], Jeanne Duhayer[2,6], Jean-François Elder[14], Jean-Baptiste Fanjul[5,15], Jocelyn Fonderflick[16], Nicolas Froustey[17], Mathieu Garel[18], William Gaudry[18], Agathe Gérard[19], Olivier Gimenez[1], Arzhela Hemery[1], Audrey Hemon[20], Jean-Michel Jullien[21], Isabelle Malafosse[16], Mircea Marginean[22], Vincent Prunet[23], Julien Rabault[24], Malory Randon[25], Antoine Régnier[10], Romain Ribière[26], Jean-Claude Ricci[17], Sandrine Ruette[18], Jérôme Sentilles[27], Nathalie Siefert[23], Bethany Smith[28], Guillaume Terpereau[8,9], Wilfried Thuiller[8,9], Antonio Uzal[28], Valentin Vautrain[12], Gaspard Dussert[2], Bruno Spataro[2], Vincent Miele[2,*] and Simon Chamaillé-Jammes[1,*]

*Corresponding authors: vincent.miele@univ-lyon1.fr; simon.chamaille@cefe.cnrs.fr

1 CEFE, Université de Montpellier, CNRS, EPHE, IRD, Montpellier, France

2 Université Lyon 1, CNRS, UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, Villeurbanne, France

3 Parc National des Ecrins, Domaine de Charance, Gap, France

4 Asters, Conservatoire d'espaces naturels de Haute-Savoie, 84 route du Viéran, PAE de Pré Mairy, Pringy, France

5 Fédération Départementale des Chasseurs de l'Ain, 19 Rue du 4 Septembre, Bourg-en-Bresse, France

6 Office Français de la Biodiversité, Direction de la Recherche et Appui Scientifique, Service "Santé de la faune et fonctionnement des écosystèmes agricoles", France

7 Fond d'Intervention Eco-Pastoral (FIEP) Groupe Ours Pyrénées, 1 rue Boyrie, Pau, France

8 Université Grenoble Alpes, CNRS, Université Savoie Mont Blanc, LECA, Laboratoire d'Ecologie Alpine, Grenoble, France

9 Observatoire spatio-temporel de la biodiversité et du fonctionnement des socio-écosystèmes de montagne (ORCHAMP), France

10 Institut du Développement et des Ressources en Informatique Scientifique (IDRIS), IDRIS-CNRS Campus universitaire d'Orsay, rue John Von Neumann, Orsay, France

11 Centre de Recherches sur les Ecosystèmes d'Altitude (CREA Mont-Blanc), Observatoire du Mont-Blanc, Chamonix, France

12 Parc naturel régional du Perche, Courboyer, Perche-en-Nocé, France

13 CERFE - URCA, 5 rue de la Héronnière, Boult-aux-Bois, France

14 Parc naturel régional des Marais du Cotentin et du Bessin, 3 village Ponts d'Ouve, Saint-Côme-du-Mont, Carentan-les-Marais, France

15 Fédération Départementale des Chasseurs du Jura, Route de la Fontaine Salée, Arlay, France

16 Parc national des Cévennes, 6 bis, Place du palais, 48400 Florac-Trois-Rivières, France

17 Institut Méditerranéen du Patrimoine Cynégétique et Faunistique, Domaine expérimental agri-environnement, Villa "Les Bouillens", Vergèze, France

18 Office Français de la Biodiversité, Direction de la Recherche et de l'Appui Scientifique, Montfort 01330 Birieux & 5 allée Bethleem Gières, France
19 Parc naturel régional des Ballons des Vosges, gestionnaire de la RNN de la Tourbière de Mâchais, 1 place des verriers, Wildenstein, France
20 Etablissement public du Mont-Saint-Michel, 16 route de la Caserne, Beauvoir, France
21 Chemin de Pré Rond-Les Mariages, Le Châtelard, France
22 Fauna & Flora International, The David Attenborough Building, Pembroke Street, Cambridge, CB2 3QZ, United Kingdom
23 Parc national du Mercantour, 23 rue d'Italie CS 51316 06006 Nice, France
24 Institut de Recherche en Informatique de Toulouse, DPD - Université Toulouse III - Paul Sabatier, 118 Route de Narbonne, Toulouse, France
25 Fédération Départementale des Chasseurs de la Drôme, 3132 Route des Sétérées, Crest, France
26 Parc naturel régional de l'Aubrac, Place d'Aubrac, Aubrac, France
27 Office Français de la Biodiversité, Direction de la Recherche et Appui Scientifique, Service "Conservation et Gestion des Espèces à Enjeux", France
28 School of Animal, Rural and Environmental Sciences, Nottingham Trent University, NG25 0QF, United Kingdom

Abstract

Camera traps have revolutionized how ecologists monitor wildlife, but their full potential is realized only when the hundreds of thousands of collected images can be readily classified with minimal human intervention. Deep-learning classification models have allowed extraordinary progress towards this end, but trained models remain rare and are only now emerging for European fauna. We report on the first milestone of the DeepFaune initiative (https://www.deepfaune.cnrs.fr), a large-scale collaboration between more than 40 partners involved in wildlife research, conservation and management in France. We developed a two-step pipeline built upon the MegaDetectorV5 algorithm for the detection of animals, humans or vehicles, and on a classification model trained to recognize 22 species or higher-level taxa. The classification model achieved 0.97 validation accuracy and often >0.92 precision and recall for all classes, and these performances were overall higher than 0.90 when tested on an independent out-of-sample dataset. We implemented this pipeline in a software to classify images stored locally on a personal computer, so as to provide a free, user-friendly and high-performance tool for wildlife practitioners to automatically classify camera-trap images.

Keywords: camera traps; classification; deep-learning; Europe; France; software

1. Introduction

Camera traps have revolutionized the way ecologists monitor biodiversity and population abundances (O'Connell et al. 2011; Howe et al. 2017). Camera traps enable to scale-up monitoring efforts dramatically, both in space and time, as they are relatively cheap, easy to deploy and autonomous tools (Steenweg et al. 2017). The continuous monitoring capacity they provide also facilitates the detection of rare species. Therefore, and unsurprisingly, it has become common practice to deploy tens to hundreds of camera traps in monitoring programs.

The full potential of camera traps is however only realized when the hundreds of thousands of images, many being empty from spurious detections, can be rapidly classified with minimal human intervention (Chen et al. 2014; Schneider et al. 2018; Wearn et al. 2019; Tuia et al. 2022) Since the beginning, machine learning approaches, and in particular deep learning models, have held the promise to solve this issue. Recent works have confirmed their power: for instance, deep-learning models developed by different research teams (Willi et al. 2019; Tabak et al. 2019; Whytock et al. 2021) have obtained > 0.90 recall and precision for a number of mammal species in North American ecosystems, African savannas and tropical forests.

The number of image classification models available to ecologists currently remains low and their taxonomic coverage is still limited. So far, most of the image classification models developed on camera-trap data have been trained, even when based on millions of pictures, using images collected by a few partners in one or two sites. For example, Norouzzadeh and colleagues (2018)

trained their model on 3.2 million images taken from the Serengeti Snapshot project (Swanson et al. 2015) to identify 48 African mammals, yet all the data was collected from one large study area in the Serengeti National Park as part of the same monitoring project. Although this can lead to highly accurate models when tested on data collected in the same sites or applied to similar systems, such models may fall short from generalizing to other fauna or contexts (Beery et al. 2018). For instance, we are not aware of any European research group that has either collected such a large camera-trap image dataset or trained an equally-powerful image classification model on European fauna up to this day. We believe there is potential in aggregating multiple small datasets from a large number of partners to build sufficiently large training, validation and test datasets to develop efficient classification models.

The pros and cons of a cross-partner image aggregation strategy, and its ability to lead to successful models, have yet to be investigated. Recent initiatives such as Wildlife Insights, an online platform to manage and classify camera-trap images (https://www.wildlifeinsights.org), have embraced this strategy. Naturally, each initiative will have its own terms of participation, which may not be acceptable for certain institutional partners. For example, online platforms require the users to upload their images, and sometimes have non-optional data-sharing policies. This approach may not be suitable for certain institutional partners which have legally-bounding contracts on data collection or privacy concerns.

We report here on the DeepFaune initiative (https://www.deepfaune.cnrs.fr). This initiative aims at (1) aggregating camera-trap images from many French institutions or research groups to create a collective and large-scale dataset, (2) developing a deep-learning-based species identification pipeline, composed of a detection and a classification model, that can separate empty pictures from pictures containing humans, vehicles or animals, identifying the species or higher-level taxa in the latter case, and (3) providing a free, user-friendly and efficient software for practitioners to run the identification pipeline on self-hosted images with standard personal computers.

2. Methods

2.1 Partners and data collection

The initiative brings together over 40 partners (see the complete list at https://www.deepfaune.cnrs.fr) representing institutions managing protected areas, hunting federations or academic research groups across European France. Partners provided camera-trap images or videos, originating from various monitoring projects (e.g. general monitoring of Alpine fauna, monitoring of the burrows of badgers) or opportunistic surveillance (e.g. to assess the presence of wolves). One additional partner joined the project, after initial data collection, providing annotated pictures to use as an out-of-sample dataset for a stringent test of the detection and classification models' accuracies (see below).

2.2 Original dataset

We gathered and sorted 641,842 annotated pictures and 4,297 annotated videos, after removal of corrupted or other problematic files. These pictures and videos included either one of the 134 animal species originally identified by the partners, a human, a vehicle, or were empty (3% of the total), i.e. had no animal, human or vehicle visible. In some instances, annotations had been made at a higher taxonomic level than the species level (e.g. bird, rodent). The amount of data provided by each partner varied a lot, ranging from 20 videos to more than 260,000 images. Many partners shared data from camera traps set to take a series of images for a single detection event . We thereafter defined these consecutive images as "sequences", considering that two images taken within 20s, at the same site, belonged to the same sequence. Finally, as we wanted to train our model to classify camera-trap images, we converted videos into images by retaining one frame per second out of the first 4 seconds of each video.

2.3 The DeepFaune identification pipeline

In this section, we give an overview of the DeepFaune identification pipeline. The pipeline consists in two steps: (step 1) detecting animals, humans and vehicles and filtering out empty images using a robust detection model, and (step 2) using a CNN classification model to identify the species in the images where an animal was detected. This approach has been the recurring leading approach in recent iWildcam competitions (Beery et al. 2021). We detail the two steps hereafter.

In step 1, we filtered out images that did not contain an animal (i.e. empty or containing a human or a vehicle) with MegaDetectorV5 (model v5a; https://github.com/microsoft/CameraTraps/; Beery et al. 2019). The images that did not contain an animal were directly attributed the corresponding classes: empty, human or vehicle according to the prediction made by MegaDetectorV5. MegaDetector v5a is based on the image detection model YOLOv5 (Redmon et al. 2016) and allowed us to produce one cropped image per animal that was detected (there were potentially multiple individuals in a given image) or per human/vehicle. We retained cropped images that had a confidence score higher than 0.8 to decrease the risk of false detection.

In step 2, we used a trained CNN model to classify the images where an animal was detected amongst 22 animal species or higher-order taxonomical groups that were common enough in our original dataset: European badger (*Meles meles)*, bird, chamois (alpine, *Rupicapra rupicapra,* and Pyrenean, *Rupicapra pyrenaica*), cat (domestic *Felis catus* and wild), cow (*Bos Taurus)*, dog (*Canis familiaris)*, equidae, red fox (*Vulpes vulpes*), ibex (*Capra ibex*, alpine only), lagomorph, lynx (*Felix lynx)*, marmot (*Marmota marmota)*, micromammal, mustelidae, mouflon (*Ovis gmelini)*, red deer (*Cervus elaphus)*, roe deer (*Capreolus capreolus)*, sheep (*Ovis aries*) squirrel (*Sciurus vulgaris)*, bear

(Ursus arctos), wild boar (*Sus scrofa)* and wolf (*Canis lupus lupus*). For each processed image, the model computes a prediction score for each class and returns the class with the highest score. The prediction score expresses the certainty on the attribution of the image to a given class. A prediction score of 1 for one class indicates that the model is certain that an object of this class is present in the image. Conversely, a score of 0 for one class indicates that the model is certain that an object of this class is not present in the image. As we used a trained and readily available detection model, we will only detail the training and validation phases for the classification model (step 2) in the next sections.

2.4 Training and validation datasets for the classifier model

We prepared the training and validation datasets by using MegaDetectorV5 (Beery et al. 2019) on our original dataset to filter out empty images and images containing humans and/or vehicles. This enabled us to avoid shortcut learning (Geirhos et al. 2020) as we trained and validated our CNN classification model on annotated images containing an animal with little to no background. Since the images in the original dataset were annotated, it was possible to annotate any cropped image in the training and validation datasets using the annotation of its parent image. For instance, if MegaDetectorV5 detected two animals in an image originally annotated as 'wild boar', we considered that the two resulting cropped images contained a wild boar to train the model. After this step, we remained with a dataset of 418,726 cropped animal images. However, we observed that some images were annotated as containing an animal in the original dataset but that the animal was very hard to distinguish. This specifically concerned small animals (e.g. rodents, birds), which were detected far from the camera trap but were only visible to an expert human eye. This was also the case when images were annotated by humans using consecutive images taken in a small interval. A single image containing a small part of an animal can then be annotated by a human observer given the images preceding or following the focal image. As a consequence, MegaDetectorV5 was sometimes unable to find an animal on all the images of a given sequence.

CNN models require independent training and validation datasets to learn successfully (Kapoor and Narayanan 2022) , and as commonly done, we split the whole image dataset into a training dataset containing 90 % of the images, and a validation dataset containing the remaining 10%. One issue that could arise during this split is that pictures collected within the same sequence are used in both the training and validation datasets, which would threaten their independence. As our original dataset contained pictures and videos collected with many different camera-trap models, there was no consistency in how the date and time stamps were stored. We thus chose to identify sequences through image filenames instead. We used a heuristic based on text mining in the R package *stringdist* (van der Loo 2014) to compare filenames to each other. We considered that two files were not independent

if their filenames were too similar, i.e. if their similarity was above 0.9. This conservative threshold insured there was no image from the same sequence or video that was present in both the training and the validation datasets.

Class imbalance, that is the fact that the number of images per class differs, is also known to affect both training and validation (Johnson and Khoshgoftaar 2019). Here, we propose a novel approach that combines downsampling and upsampling during the training and validation phases. At each epoch of the training phase, we downsampled the class which was overrepresented down to a multiple of the rarest class. We chose to have a ratio of 5 between the number of images of the most common class and the number of images of the rarest class. This was estimated for every epoch, such that every image is seen by the model if we fit the model for enough epochs. As a consequence, images of the rarest classes were used in many epochs and can thus be considered upsampled whereas those of the most common classes are downsampled. This way, the optimization problem was different at every epoch, but more balanced. We considered that the stochasticity induced by the sampling process at each epoch had a positive impact on the model fit procedure and we observed a continuous reduction in model loss through epochs. We used the same idea to handle imbalance during validation. This enabled us to compute what we called a balanced validation accuracy.

2.5 Training and validation phases for the classifier model

We used transfer learning starting from an EfficientNetB3 model (Tan and Le 2020) that was pre-trained on Imagenet, with a resolution of 300 × 300, using TensorFlow-Keras. We performed image augmentation using the *imgaug* Python library (https://github.com/aleju/imgaug) during the training phase. Each image in each batch was modified with a random set of transformations such as horizontal flip, affine transformations, gray scale transformation, blurring, brightness or contrast changes. We used the softmax output as prediction score. We then estimated the model for a maximum of 120 epochs, monitored the overall balanced validation accuracy and stopped the estimation when it did not increase with a patience of 10 epochs.

To estimate the quality of the model, we computed global accuracy (i.e. the probability that a prediction is correct), precision (i.e. the probability that an image classified as showing species i actually has species i present) and recall (the true positive rate, i.e. the probability that an image having species i present is actually classified as showing species i, i.e. the true positive rate or sensitivity).

2.6 Testing our pipeline performance on out-of-sample data using sequences

We investigated the ability of our two-step identification pipeline to perform accurate identification of species in an out-of-sample dataset, i.e. in images taken in contexts that have never been seen

during the training stage (Schneider et al. 2020). Our out-of-sample dataset originated from 27 camera traps installed in 11 different locations distributed along two altitudinal gradients in the Alps, France. It was made of 25,836 annotated pictures, containing images of humans (but no vehicles) and 16 identified animal species (cat, cow, lynx, bear, equidae and mouflon were lacking), and 34.5% of empty images.

We first assessed the detection quality of the MegadetectorV5-based detection step of our pipeline using the rate of false negative (FN) detections. A FN detection occurs when MegaDetectorV5 fails in detecting an animal, a person or a vehicle in an image where an annotator detected it. A high FN rate indicates that MegadetectorV5 commonly fails to detect the animal or human present in the picture. For empty images, a FN detection is when MegaDetectorV5 erroneously detects an element in an image that is known to be empty. We computed the FN rate at the image and sequence level for each class. Sequences are considered as FN detections when all the images in the sequence are false negatives. When MegaDetectorV5 detected more than one animal or human, we kept the one with the highest score and created a single cropped image per original image for the next step.

We then explored the classification accuracy of the CNN-based classification step of the pipeline. We predicted the species present at the level of the sequence and not at the level of the individual image. We chose to predict on the sequence level for two reasons: camera traps are now often set to take images in bursts and animals commonly stay several seconds near the camera, triggering it multiple times in a row. To do so, we extracted the date and time of each image of the out-of-sample dataset using the EXIF information. We obtained the prediction score at the sequence level by first computing the prediction score for each image of the sequence that was not predicted as empty. We then summed these scores and considered that the class with the highest total prediction score was the predicted class for the sequence. Images that were not in a sequence were classified using their own prediction score.

2.7 The DeepFaune software

We found that MegaDetectorV5 created a bottleneck in our pipeline speed when used on a CPU (about 2 to 3 seconds per image in our experiments on various personal computers) due to its high image resolution (1280) and large model size. We therefore developed a much faster alternative using YOLOv5, which is a high-performance medium-sized detection model with a resolution of 640. We trained the YOLOv5-based detector using the cropped images previously used in this study. We then implemented this alternative detector (about 0.3 second per image) and our CNN classifier inside a free, user-friendly and independent software. This enables users to input their images, stored locally, directly in the software, and classify these images either as empty, human, vehicle or one of the 22

animal classes on which the CNN classifier was trained. Users can decide of a threshold on the prediction score below which the images will not be classified and will be set aside in a 'undefined' class for visual inspection. The software only requires the installation of Python v3 and its PyTorch module. To overcome the common difficulties associated with the installation of Python libraries on Windows operating systems, we created a standalone and ready-to-use .exe program for these systems. The software is available on the project's website: https://www.deepfaune.cnrs.fr.

3. Results

3.1 Performance of the CNN model on the validation dataset

We obtained an overall balanced validation accuracy of 0.967. Recall and precision were >0.92 for all classes (Table 1).

Table 1. **Classification performance metrics of the CNN classifier model computed on the validation dataset.** Nb Training, respectively Nb Validation, refers to the number of images of a given animal class within the training, respectively validation, dataset.

Class	Nb Training	Nb Validation	Precision	Recall
badger	6566	805	0.98	0.98
ibex	976	270	0.92	0.92
red deer	57576	4958	0.93	0.97
chamois	82751	10921	0.95	0.97
cat	4336	556	1.00	0.92
roe deer	48538	5309	0.95	0.95
dog	8559	1041	0.94	0.97
squirrel	3974	604	0.97	0.99
equidae	1056	119	0.95	0.96
lagomorph	4673	624	0.98	0.97
wolf	5231	738	0.96	0.94
lynx	778	123	0.99	0.98
marmot	1699	239	0.99	0.90
micromammal	2763	444	0.98	0.98
mouflon	952	115	0.95	0.92
sheep	38505	8528	0.97	1.00
mustelidae	3386	538	0.95	0.93
bird	24844	3594	0.96	0.99
bear	6625	1415	0.97	0.98
red fox	35139	5387	0.95	0.97
wild boar	22383	2664	0.94	0.96
cow	7193	1220	1.00	0.94

We observed identification mismatches between the different ungulate species, mainly on images where only a part of the animal was present (for instance, the back of a roe deer was recognized as a red deer). Without surprise, the lowest performances among ungulates were observed for ibex and mouflon for which we had less than a thousand images in the training dataset. Interestingly, we obtained high performance scores for lynx regardless of its small sample size (778 images). This may be due to its characteristic coat color and spotted pattern which recognition and classification.

3.2 Performance of the whole pipeline on the out-of-sample dataset

During the detection stage based on MegaDetectorV5, we observed that the false negative (FN) rates on individual images where generally excellent, near or below 10% for the large species, but much higher for smaller species (Figure 1). FN rates were above 15% for micromammals and birds. The FN for empty images was also very low, which reveals that MegaDetectorV5 performed well on filtering out empty images. We note however that the magnitude of the difference between individual and sequence-level FN rates varied among classes. For the majority of classes (11/18), we obtained twice as few FNs in the sequence-based approach as in the individual-based approach, which justified using the sequence level for the detection step in our pipeline.

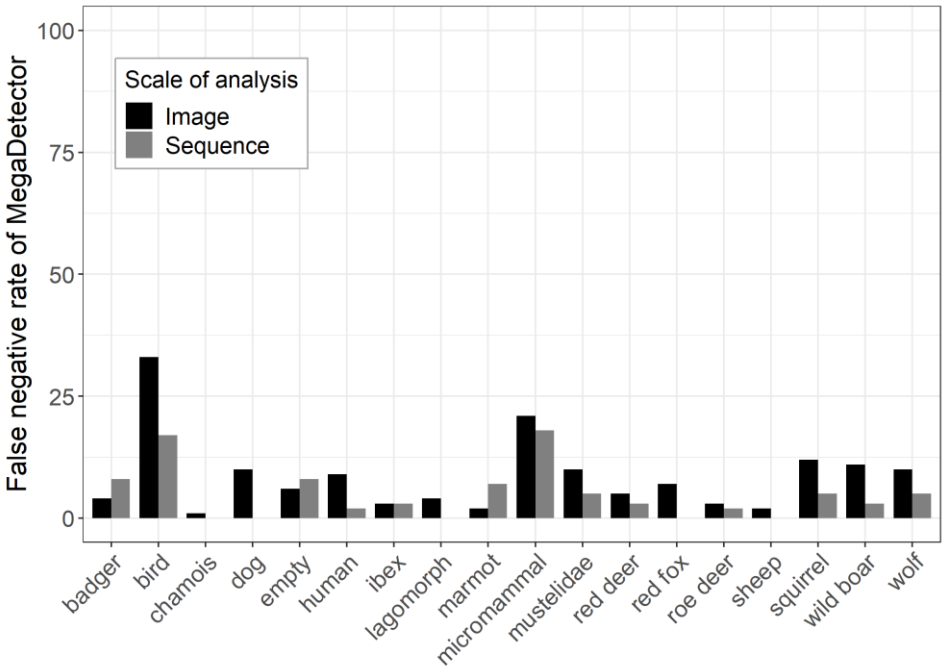

Figure 1. False negative rates obtained with MegaDetectorV5 on the out-of-sample dataset computed at the image and sequence level. A false negative image is when the image containing an animal was predicted as empty by MegaDetectorV5, or when an empty image was predicted as containing an animal. A false negative sequence is a sequence of images for which all the images are false negatives; by construction, the false negative rate for sequences is necessarily lower or equal to that of individual images. For empty images, it is the opposite.

The CNN-based classification step achieved an overall accuracy of 0.883 when using sequences. Results were generally significantly better when identification was done at the sequence level (Table 2, Figure 2). The best performance was achieved for large ungulates and squirrel, with precision and recall being about 0.9 for squirrel, red deer, roe deer, chamois, and to a lesser extent for wild boar and fox (Table 2). Performance remained good for micromammals and birds but were lower on the out-of-sample dataset than for the validation dataset (see Table 1) due to some confusions between the two categories (appendix Figure 4). The performance for the few images containing a wolf in the out-of-sample dataset was moderate. However, manual inspection revealed that vegetation in the foreground could be the source of this issue (see Figure 3). This issue might therefore be specific to this dataset.

Table 2. **Classification performance metrics computed on the out-of-sample dataset, using images or sequences**. Model performance at the image level is expected to be lower than at the sequence level as people who annotated the images had the complete 3-images sequences available and could use previous or subsequent images to clarify uncertainties.

Taxa	**Recall at the sequence (image) level**		**Precision at the sequence (image) level**	
badger	1.00	(0.94)	0.80	(0.76)
ibex	0.83	(0.73)	1.00	(1.00)
red deer	0.89	(0.83)	0.84	(0.64)
chamois	0.91	(0.87)	0.97	(0.96)
roe deer	0.85	(0.80)	0.96	(0.94)
dog	1.00	(0.89)	0.83	(0.28)
squirrel	0.91	(0.82)	0.96	(0.94)
lagomorph	0.94	(0.87)	0.63	(0.49)
wolf	0.83	(0.72)	0.71	(0.52)
marmot	0.96	(0.88)	0.87	(0.89)
micromammal	0.75	(0.69)	0.99	(0.99)
sheep	0.95	(0.84)	0.64	(0.88)
mustelidae	0.95	(0.87)	0.65	(0.48)
bird	0.98	(0.98)	0.74	(0.59)
red fox	0.87	(0.82)	0.88	(0.68)
wild boar	0.95	(0.92)	0.83	(0.79)

Figure 2. Using sequences decreases the false positive rate in classification predictions. A sequence of three images of a chamois, taken over a 3-second interval (and cropped by MegaDetectorV5). The two first images are predicted as chamois (0.99). The third image is misclassified as a sheep with a lower score (0.93), but the score at the sequence-level is highest for chamois. The three images are therefore predicted to show a chamois by the DeepFaune pipeline.

There remained some mismatches between classes (appendix Figure 4). For instance, there were 48 sequences of roe deer that were classified as red deer and we observed confusions between micromammals and birds (25 sequences) as mentioned earlier. We observed no confusion between wolf and dog predictions, which suggested that our model could distinguish between the two. We noted confusions between wolf and red fox for images taken at night (3 sequences).

Figure 3. Examples of images in which a wolf is not correctly identified, possibly due to vegetation in the foreground. A wolf was classified as a bird (prediction score 0.9) in a sequence of 1 image (left), and as a red deer (prediction score 0.77) in a sequence of 1 image (right).

4. Discussion

The DeepFaune initiative represents a successful multi-partner collaboration to build a large-scale camera-trap image dataset and develop one of the first pipelines readily available to automatically classify camera-trap images collected in Europe. Our current model allows for predicting the presence of 24 different classes in camera-trap images (humans, vehicles and 22 animal species or higher order taxa) as well as filtering out empty images. Implementing MegaDetectorV5 in our pipeline allowed for analyzing datasets in which empty images were numerous and detecting humans and vehicles

prior to animal classification. Although we did use the standard approach of transfer learning, we implemented a number of tricks to deal with class imbalance and the independence of training and validation datasets that could be useful to others. Ultimately, our species classification model performed extremely well on the validation dataset and provided robust results on out-of-sample data. Additionally, we built a graphical user interface (GUI) and released our "DeepFaune software" so that practitioners can easily run the model locally on regular personal computers and automatically sort their newly acquired images based on the model predictions.

4.1 Model performance and relevance for ecological studies

There are currently very few species-recognition models available for European fauna and our results could therefore be used to benchmark future studies. The quality of our model appears very good, even by deep-learning classification standards, and compares favorably with results from similar exercises conducted on fauna from other continents. For instance, overall accuracy of 0.97, 0.78 and 0.94 on validation datasets were respectively reported in studies on North American fauna (Tabak et al. 2019), Central African fauna (Whytock et al. 2021) and East African fauna (Norouzzadeh et al. 2018). Our model also improves on the one reported by Carl et al. and colleagues (2020) which is one of the rare existing models focused on European fauna, as we achieved much higher accuracy at the species level.

We managed to obtain good results with a moderately sized dataset (by deep-learning standards), by overcoming a number of obstacles. Firstly, the large imbalance (e.g. two orders of magnitude more images of chamois than ibex) would have biased the classification towards the most common species. We therefore implemented an original approach combining downsampling and upsampling techniques that successfully prevented the emergence of a relationship between identification performance and the number of images per class. Secondly, deep learning models are known to be able to learn 'shortcuts' (Geirhos et al. 2020) and can misclassify images due to spurious similarities (e.g. similar image backgrounds). In our context, this would correspond to learning contextual elements (e.g. snow, peculiar vegetation) that would be associated with the species, rather than learning to recognize the species itself. To avoid this potential issue, we chose to classify cropped images, and not whole images (as opposed to Whytock et al. 2021), with an additional procedure of image augmentation.

Overall, class-specific recall and precision are high enough for the model to be useful for many specific studies, some of which we highlight now: (1) automatized monitoring of large ungulates, a guild of important management interest in Europe. Ungulates are indeed generally very well classified by the model, with recall and precision values above 0.9 for most species (as explained in the Results section, performance for ibex and mouflon are low, but will likely improve as new pictures are

continuously added in the database). These results suggest that our model could be a useful tool to facilitate studies investigating the effects of management practices on locally abundant ungulates (e.g. can wild boar population dynamics be controlled by hunting?) or the dynamics of prey under predation (e.g. do roe deer populations decrease as wolves return?); (2) monitoring large mammalian communities. Although model accuracy levels were generally lower for classes other than ungulates, the model's performance and taxonomic coverage suggest that it could be useful to automatically sort out hundreds of thousands of pictures for a range of taxonomic groups. This could prove especially useful in studies looking at the impact of anthropization on large mammal communities for example. A human intervention would still be required to manually verify the pictures for which prediction scores are low or to identify the species present in images where the model classifies at a higher taxonomic level (e.g. mustelidae), but this effort should be dramatically reduced when using our model; (3) quantifying human disturbance levels, as our pipeline is able to distinguish images containing humans, vehicles and/or domestic animals such as cattle (sheep, cow) and dogs. The current model performances should therefore enable to build reliable metrics estimating human activities along wide gradients.

We emphasize here that even though precision and recall might be generally high, caution should be exerted when using the model to classify them, as relevant pictures might have been discarded at the detection stage by MegaDetectorV5. This is clearly demonstrated by our out-of-sample test on bird pictures, which can be well classified by the model (arguably at this broad taxonomic level) but will sometimes be discarded at the detection stage as the size of the bird is too small for MegaDetectorV5 to identify its presence. Pictures of micromammals are also sometimes mistakenly discarded by MegaDetectorV5 (Beery et al. 2019).

Importantly, the general high quality of the results of our two-step approach was conserved when applied to an out-of-sample dataset. It is a common error to take model results on validation datasets at face value and assume similar accuracies will be observed in new applications. Previous studies (Whytock et al. 2021) have shown that this is not the case and observed dramatic declines in accuracy when tested on out-of-sample datasets, to the point where the usefulness of the model could sometimes be called into question. Our out-of-sample test, based on heterogeneous images from 11 locations spread over two altitudinal gradients in France, suggests that our model performances are robust to new applications. Additionally, this out-of-sample test was designed to test the applicability of our pipeline to real-world scenarios. It revealed that (1) although MegaDetectorV5 is generally very efficient, the rate of false negatives, i.e. when the image is erroneously assumed empty and thus discarded, can be significant (> 15%) for small animals, micromammals and birds in particular. This is a critical point that should be kept in mind when designing studies and interpreting results with our approach and warrants specific attention and, ideally, study-specific quantifications; (2) using the

average prediction score over temporally-close pictures (i.e. sequence) can improve classification results if classification at the sequence-level is sufficient. Sequence-level prediction was sufficient because, over a few pictures, at least one image is likely to be classified very reliably by the model (e.g. prediction score > 0.95), increasing largely the sequence-level prediction score for this class, and successfully identifying the animal present in all images of the same sequence. Also, confusion between species are less likely to occur over several images. Sequences of pictures are commonly used in camera trapping to account for differences in fields of detection, camera views, and generally to increase chances of getting at least one identifiable picture of the animal triggering the camera trap. We demonstrate here that this approach can be leveraged when using classification models. Our work also highlights that, when using sequence-based identification, the performance of classification models trained without sequence information are likely under-estimated in contrast to what can be achieved at the sequence level. In the future, sequence or temporal information could be directly integrated into the training step, as done in context CNN models (Beery et al. 2020; Tuia et al. 2022). We could not use this approach here as many partners provided data without information about the specific trapping sites where the pictures came from nor could we retrieve time-stamp EXIF information for each file.

4.2 Lessons learned from a successful multi-partner initiative

One of the main strengths of the DeepFaune initiative lies in the creation of a nation-wide network among key actors in French biodiversity research, conservation and management. Under the lead of an academic research group originating from two distinct laboratories, over 40 partners have shared camera-trap pictures and videos allowing to build what is likely to be one of the largest databases of camera-trap medias in France, in both number (over 640,000 photos and 4,200 videos) and taxonomic coverage.

This success should not hide the technical challenges of working with such a large number of actors. Building the original database clearly revealed the intricacies of dealing with multiple, high-volume data transfers as well as with the strong heterogeneity in data acquisition and organization among the partners. The devil is in the details, and harmonizing directory, file and species names were all very time-consuming tasks that had to be dealt with by combining automatic and manual interventions. Partner-specific data management appears as one of the strongest barriers to creating efficient large-scale datasets that do not depend on a single monitoring program (as opposed to the Serengeti Snapshot (Swanson et al. (2015) dataset for instance). Interestingly however, some of our individual partners were members of the same institution but worked in different sites. For example, we received data from several national parks that all belong to the same institution (Parc Nationaux de France). Similarly, several teams from the Office Français de la Biodiversité shared data, all having

different formats. In such cases, it would seem beneficial for the master institution to provide detailed data management guidelines that would allow standardized data management internally. In this context, centralized data management platforms that enforce data standardization, developed either within institutions or at national or international levels, might facilitate future works.

Having numerous and diverse partners is key however to ground our work in the reality of end-users, as the DeepFaune initiative was originally conceived to develop a free and easy-to-access tool of sufficient quality for field practitioners. Beyond the data collection stage, a regular communication between the leading academic team and the partners was critical to identify key expectations and potential difficulties in appropriation. Expectations shared by all partners were two-fold: (1) all partners expected a high performance of the classification of empty pictures for the pipeline to be useful, as these usually represent a large share of collected pictures. In the current version of the pipeline, we rely on MegaDetectorV5 and are thus bound to its own performance, which are generally good; (2) the pipeline should be easy to implement and run on a standard personal computer. This is a long-standing issue regarding deep-learning models, as even when training is not required and only predictions are expected, the installation of libraries necessary for computation (e.g. Tensorflow) can be difficult. Uploading pictures on an online platform running its own servers naturally solves this issue. We found that many partners were reluctant to follow this approach either because of the need to upload gigabytes of images online or because the data-sharing policies were too stringent. We thus decided to implement our pipeline in a free Python-based software that can easily be launched on common operating systems (Windows, macOS, Linux). Feedback from the partners will be critical to ensure that our approach is a credible solution. The developments of the software can be followed at https://www.deepfaune.cnrs.fr/.

Expectations of our partners differed regarding the classification model and how to use its prediction to sort out images. What makes an 'optimal' sorting strategy might in fact vary between partners. Some partners are interested in minimizing false negatives over false positives. This can be the case when studying sensitive species, such as wolves, as positive classifications will in any case be verified manually. For other studies, the balance between false positives and false negatives matters less. This is generally the case in occupancy modeling studies of common species (Gimenez et al. 2022). We also foresee that prediction scores will, in a near future, be directly used in the inference process allowing the uncertainty of the classification to be propagated into the estimation of any metric of interest. Irrespective of whether this will be successful or not, in response to this diversity of needs, we decided to let the users decide of a threshold on the prediction score below which the images will not be classified and requires manual inspection. We believe that this approach gives an important level of flexibility and that users will be able to learn what threshold works best for them by trial-and-error.

4.3 Conclusion

In conclusion, our work provides a rather successful classification model of the European fauna in camera-trap images, which can be easily used by practitioners through the Deepfaune software on self-hosted images using a standard personal computer. Such an approach contrasts with some recent developments that favor cloud-computing. Our work however remains a work-in-progress and feedbacks on the use of the model and the associated software, as well as annotations of images for which the model failed, should allow to improve the model's performances in a near future.

5. Contributions and conflict of interest disclosure

N. Rigoudy set up the collaborative network with S. Chamaillé-Jammes, organized data collection, and cleaned-up and managed the dataset with V. Miele. V. Miele developed the model and the graphical-user-interface. S. Chamaillé-Jammes initiated the project with N. Rigoudy and now manages it with V. Miele. B. Spataro co-supervised the use of the LBBE PRABI computing cluster. G.Dussert later joined the initiative and presently works on improving model performances. All five make up the DeepFaune team. Other authors contributed data on their behalf or on behalf of their institutions. The authors declare that they have no financial nor intellectual conflicts of interest with the content of this article.

6. Acknowledgments

This work was performed using the computing facilities of the CC LBBE/PRABI. This work has benefited from discussions initiated within meetings of the GDR 'Ecologie Statistique' of the CNRS. V. Miele would like to thank the LECA laboratory for having hosted him in Chambéry, and L. Humblot for insightful discussions about the software. Also, for many partners, field technicians and students contributed largely to data collection, and they are acknowledged for their critical work. Some partners could not share images at the time of our study but actively participated in the discussions around future model developments: PNR des Ardennes, PNR de Chartreuse, PNR du Haut-Languedoc, PNR du Haut Jura, PNR du massif des Bauges, PNR de Préalpes d'Azur, PN de la Vanoise and PN des Calanques, FNC.

7. Funding

The DeepFaune team is funded by the CNRS, the university Lyon 1, the 'Ecole Normale Supérieure de Lyon' and the 'Programme National de Recherche en Intelligence'. The CNRS provided additional financial support through its 'Mission pour les Initiatives Transverses et Interdisciplinaires'. The following funders supported data collection through various programs (supported partners in parenthesis): Region Auvergne-Rhône-Alpes (FDC Drôme, Ain, Jura; CREA), Réseau de Transport

d'Electricité (CEFE), Fédération Nationale des Chasseurs (FDC Ain, Jura, Office Français de la Biodiversité (through the Ecocontribution, FDC Ain, Jura), Conseils Départementaux de l'Ain et du Jura (FDC Ain, Jura), Fonds européens FEDER/POIA (CREA), Département de la Haute Savoie (CREA), Communauté de Communes Vallée de Chamonix (CREA), DREAL Bourgogne-Franche-Comté (FDC Jura), DREAL Auverge-Rhône-Alpes (FDC Ain), DREAL Normandie (PNR Perche), DREAL Centre-Val-de-Loire (PNR Perche), Fédéraction Départmentale des Chasseurs de la Marne (CERFE), Réseau SNCF (CERFE), Autoroutes SANEF (CERFE), Nottingham Trent University and Fauna & Flora International (B. Smith & A. Uzal & M. Marginean).

8. References

Beery S, Agarwal A, Cole E, Birodkar V (2021) The iWildCam 2021 Competition Dataset. arXiv preprint

Beery S, Morris D, Yang S (2019) Efficient Pipeline for Camera Trap Image Review. arXiv preprint arXiv:190706772 [cs]

Beery S, van Horn G, Perona P (2018) Recognition in Terra Incognita. arXiv preprint arXiv:180704975 [cs, q-bio]

Beery S, Wu G, Rathod V, et al (2020) Context R-CNN: Long Term Temporal Context for Per-Camera Object Detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp 13075–13085

Carl C, Schönfeld F, Profft I, et al (2020) Automated detection of European wild mammal species in camera trap images with an existing and pre-trained computer vision model. Eur J Wildl Res 66:62. https://doi.org/10.1007/s10344-020-01404-y

Chen G, Han TX, He Z, et al (2014) Deep convolutional neural network based species recognition for wild animal monitoring. In: 2014 IEEE International Conference on Image Processing (ICIP). pp 858–862

Geirhos R, Jacobsen J-H, Michaelis C, et al (2020) Shortcut Learning in Deep Neural Networks. Nat Mach Intell 2:665–673. https://doi.org/10.1038/s42256-020-00257-z

Gimenez O, Kervellec M, Fanjul J-B, et al (2022) Trade-off between deep learning for species identification and inference about predator-prey co-occurrence. Computo. https://doi.org/10.57750/yfm2-5f45

Howe EJ, Buckland ST, Després-Einspenner M-L, Kühl HS (2017) Distance sampling with camera traps. Methods in Ecology and Evolution 8:1558–1565. https://doi.org/10.1111/2041-210X.12790

Johnson JM, Khoshgoftaar TM (2019) Survey on deep learning with class imbalance. Journal of Big Data 6:27. https://doi.org/10.1186/s40537-019-0192-5

Kapoor S, Narayanan A (2022) Leakage and the Reproducibility Crisis in ML-based Science. arXiv preprint arXiv:2207.07048 [cs.LG]

Norouzzadeh MS, Nguyen A, Kosmala M, et al (2018) Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proc Natl Acad Sci USA 115:E5716–E5725. https://doi.org/10.1073/pnas.1719367115

O'Connell AF, Nichols JD, Karanth KU (eds) (2011) Camera Traps in Animal Ecology. Springer Japan, Tokyo

Redmon J, Divvala S, Girshick R, Farhadi A (2016) You Only Look Once: Unified, Real-Time Object Detection. arXiv preprint arXiv:1506.02640 [cs.CV]

Schneider S, Greenberg S, Taylor GW, Kremer SC (2020) Three critical factors affecting automated image species recognition performance for camera traps. Ecol Evol 10:3503–3517. https://doi.org/10.1002/ece3.6147

Schneider S, Taylor GW, Kremer SC (2018) Deep Learning Object Detection Methods for Ecological Camera Trap Data. arXiv preprint arXiv:180310842 [cs]

Steenweg R, Hebblewhite M, Kays R, et al (2017) Scaling-up camera traps: monitoring the planet's biodiversity with networks of remote sensors. Frontiers in Ecology and the Environment 15:26–34. https://doi.org/10.1002/fee.1448

Swanson A, Kosmala M, Lintott C, et al (2015) Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. Sci Data 2:150026. https://doi.org/10.1038/sdata.2015.26

Tabak MA, Norouzzadeh MS, Wolfson DW, et al (2019) Machine learning to classify animal species in camera trap images: Applications in ecology. Methods Ecol Evol 10:585–590. https://doi.org/10.1111/2041-210X.13120

Tan M, Le QV (2020) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv preprint arXiv:1905.11946 [cs.LG]

Tuia D, Kellenberger B, Beery S, et al (2022) Perspectives in machine learning for wildlife conservation. Nat Commun 13:792. https://doi.org/10.1038/s41467-022-27980-y

van der Loo MPJ (2014) The stringdist Package for Approximate String Matching. The R Journal 6:111–112

Wearn OR, Freeman R, Jacoby DMP (2019) Responsible AI for conservation. Nat Mach Intell 1:72–73. https://doi.org/10.1038/s42256-019-0022-7

Whytock RC, Świeżewski J, Zwerts JA, et al (2021) Robust ecological analysis of camera trap data labelled by a machine learning model. Methods Ecol Evol 2041–210X.13576. https://doi.org/10.1111/2041-210X.13576

Willi M, Pitman RT, Cardoso AW, et al (2019) Identifying animal species in camera trap images using deep learning and citizen science. Methods Ecol Evol 10:80–91. https://doi.org/10.1111/2041-210X.13099

Appendix

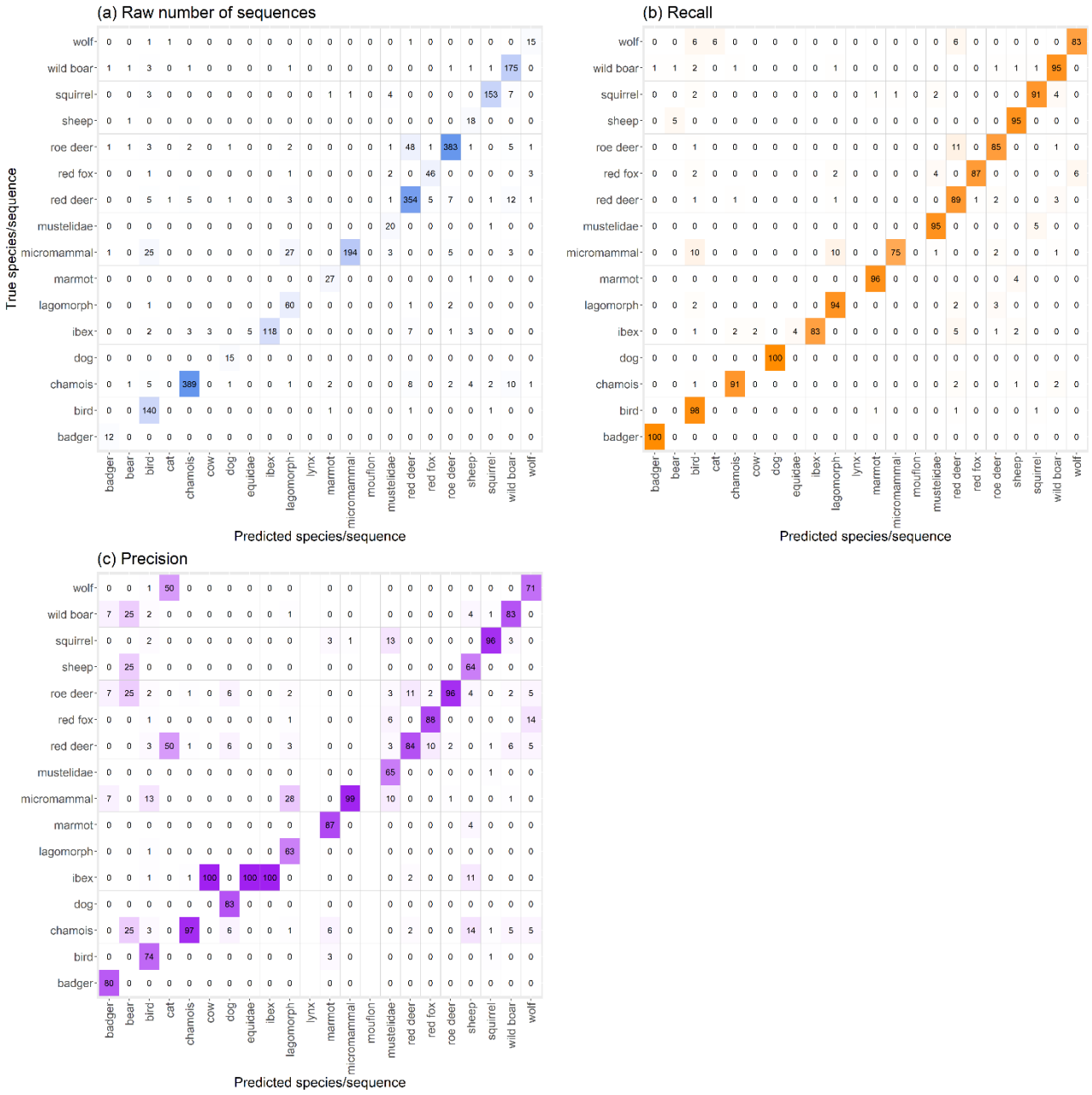

Figure 4. Confusion matrices of the predictions of the CNN-based classifier on the out-of-sample dataset.