Real-Time Wildlife Monitoring

Machine Learning Analysis of Camera Trap Photos for an Automated Wildlife Alert System in the Romanian Carpathian Mountains

Thomas Ratsakatika

28 June 2024

Word Count: 4,999

Declaration

This report is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text and/or bibliography.

Code and Data

The code that accompanies this research is available at: https://github.com/ratsakatika/camera-traps (v1.0.0).

The dataset used for testing and training the models in this report is proprietary to Fundația Conservation Carpathia and cannot be uploaded onto a public repository. Some example camera trap photos have been made available in the code repository for use in the Example Tutorial Notebook.

Acknowledgements

I would like to thank my co-supervisors, Professor Srinivasan Keshav and Dr Ruben Iosif, for their invaluable guidance, insights and expertise throughout this project. I would also like to thank Fundația Conservation Carpathia's Wildlife and Rapid Intervention Team, including Zsolt Miholcea, Laviniu Terciu and Gabi Dudu, who hosted me in Romania, deployed the camera traps and provided valuable feedback.

Abstract

This research introduces an AI-based alert system to reduce human-wildlife conflicts in the Romanian Carpathian Mountains. Globally, conflicts between people and wildlife are rising due to population growth, shifting land use patterns and climate change. In Romania, mountain communities are impacted by bears and wild boars, which damage livestock, crops and property. These conflicts can undermine conservation efforts and may result in the killing of problematic animals. In collaboration with Fundația Conservation Carpathia, this research supports Rapid Intervention Teams who respond to wildlife activity in mountain villages. Six years of camera trap data are used to train and test AI models to detect and classify European mammals. These models are integrated into an alert system and deployed in three locations. The new pipeline improves on the state-of-the-art for detecting and classifying bears and wild boars. Preliminary results from the field deployment show a positive impact on conservation efforts. This is the first known study to use remote processing of 4G-enabled camera trap images to operate a human-wildlife conflict alert system. It is also the first known study to design and evaluate all stages of an AI-based wildlife alert system, from data collection and model training to field deployment and conservation impact.

Contents

1 Introduction

1.1 Context, motivation and objectives

Human-wildlife conflict (HWC) is one of the most complex risks facing biodiversity conservation [1]. HWC arises when wildlife threatens people's safety, livelihoods or well-being [2]. This can undermine perceptions of conservation efforts and sometimes results in the killing of problematic animals [3].

Romania's Carpathian Mountains are home to one of Europe's largest forest ecosystems and most significant populations of brown bears, wolves and lynxes [4–7]. The habitat also hosts a sizeable wild boar population, providing an important food source for wolves and bears [8, 9]. Since 1990, HWC has increased driven by population growth, shifting land use patterns, supplementary feeding and climate change [10–12]. Bears and wild boar are the primary sources of HWC due to their destruction of livestock, crops, grassland and fences, and the resulting economic and social impact on rural communities [13].

Fundația Conservation Carpathia (FCC), a nature conservation and restoration foundation, works with rural communities in Romania's Carpathian Mountains to promote peaceful coexistence between people and wildlife [14]. FCC implements proactive conflict reduction initiatives to increase local acceptance of bears and wild boars. This includes three Rapid Intervention Teams who, if alerted swiftly, can arrive at the location of an incursion and deter the animals before they cause damage [15]. However, each team covers a large area of *c.* 150 km^2, and as incursions occur primarily at night, they are often notified too late. More timely alerts would optimise the Rapid Intervention Team's nightly patrols, improve community relations and enable FCC to scale its HWC prevention work in the region.

In response to FCC's conservation challenge, this research aims to build and deploy an automated alert system for bears and wild boars near their field office in Rucăr (Fig. 1). The objectives are to: (1) obtain and process six years of camera trap data labelled by FCC and establish a ground truth; (2) build a machine learning pipeline to detect and classify animals in camera trap data, optimising for bears and wild boars; and, (3) deploy an alert system (Fig. 2), and assess inference performance and conservation impact.

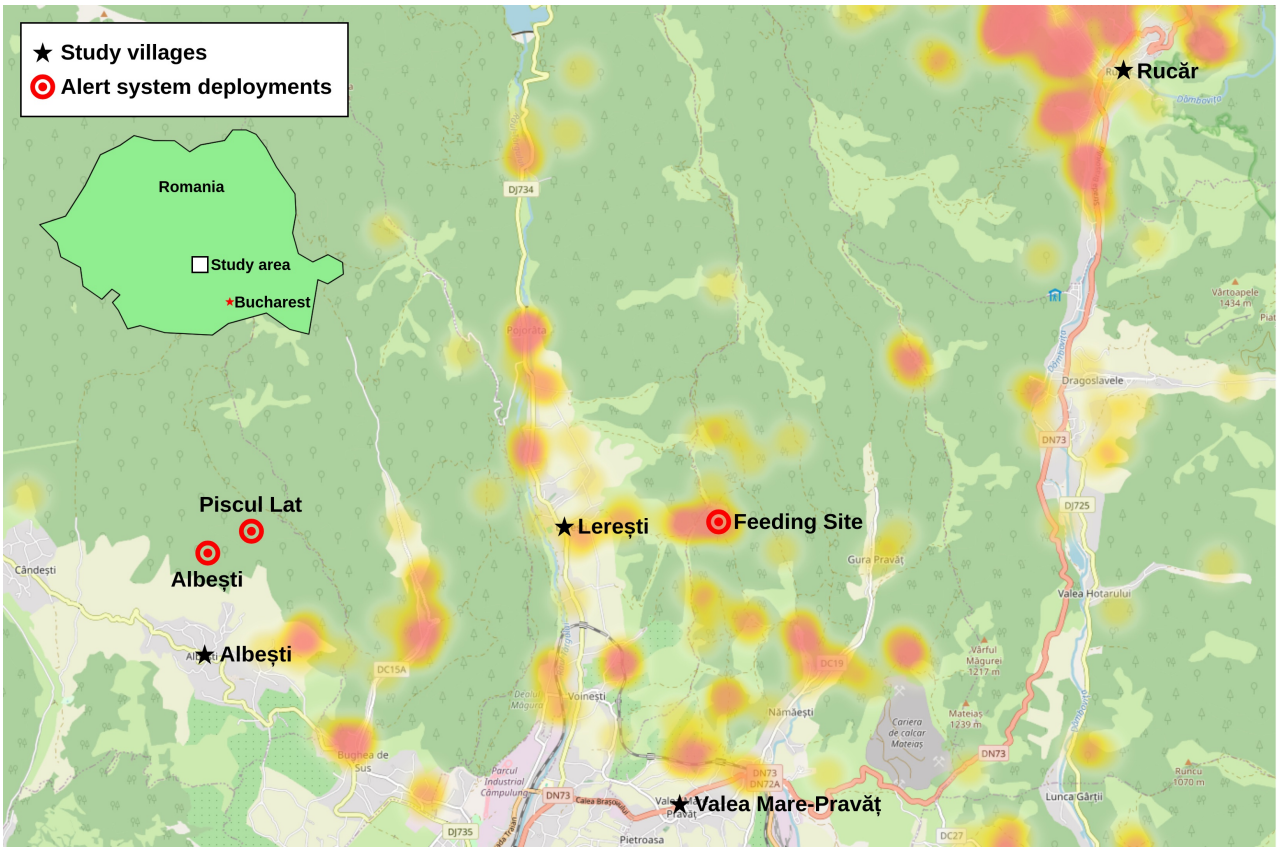

Figure 1: Heat map of human-wildlife conflict occurrences in study area (FCC data 2019-24, unpublished). Study villages and alert system deployments also shown.

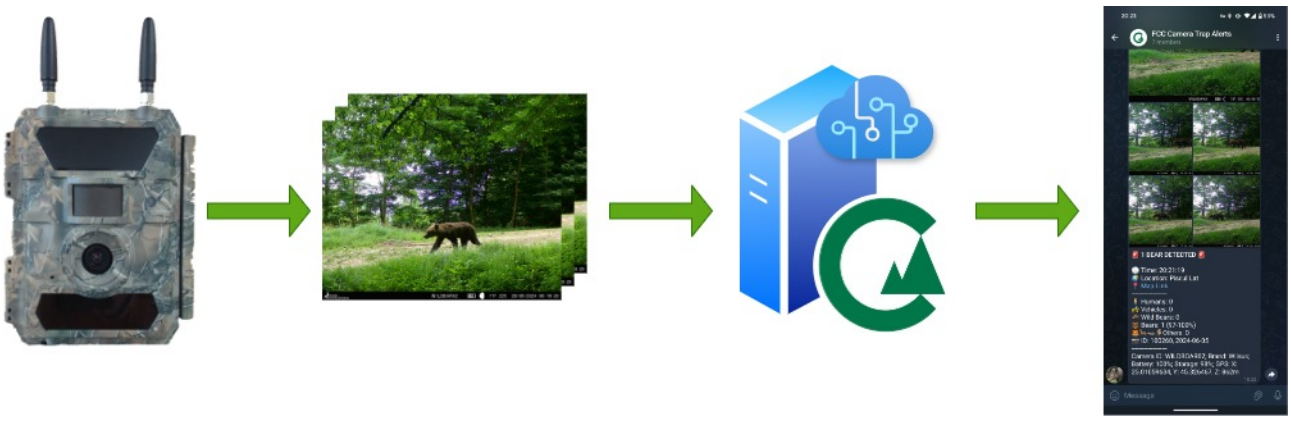

Figure 2: Alert system example. (1) motion triggers camera trap, (2) photo sequence emailed, (3) AI pipeline on FCC's server assigns species label(s), (4) instant message alert sent to Rapid Intervention Team

1.2 Report structure

Section 2 discusses related work. The structure then follows the three research objectives. Section 3 details the collection and processing of FCC's data, and determines a modelling strategy. Section 4 tests and fine-tunes computer vision models. Section 5 details the design and deployment of an alert system and assesses its accuracy and conservation impact. Section 6 discusses the results and concludes.

2 Related work

2.1 Computer vision

Machine learning analysis of camera trap photos has progressed significantly in the past decade. Convolutional Neural Networks (CNN) were first applied to the task in 2014 [16] and deeper variants, specifically Residual Network [17] and Visual Geometry Group [18] architectures, soon became the state-of-the-art (SOTA) underpinned by the release of large labelled datasets from the Serengeti and North America [19, 20]. In the early 2020s, Vision Transformers (ViTs) superseded CNNs by splitting images into equal-sized patches and leveraging the transformer's attention mechanism to learn long-range spatial dependencies [21]. CNNs have since been modernised to incorporate some of the features of ViTs, resulting in ConvNeXts, which, alongside ViTs, are now considered SOTA in computer vision [22].

There are two main approaches to analysing camera trap photos. Earlier models, such as MLWIC2, TrapperAI and ReWilding Europe, perform one-shot species classification on full-size images [23–25]. However, it was found that models trained on one region struggle to generalise to others [26]. In 2019, a more efficient two-stage pipeline was proposed, with a generalisable detector, "MegaDetector", and context-specific classifier [27] (Fig. 3). Based on the YOLOv5 object detection architecture [28] and maintained by Microsoft [29], MegaDetector outputs bounding box, confidence and class labels (animal, person or vehicle). MegaDetector is widely considered the SOTA detection model for camera trap images, and MegaDetector crops have been used to train specialised classifier models [19, 30–34].

Most classification models specialise in African and North American species, likely due to the bias towards these regions in major camera trap image repositories [35, 36]. The DeepFaune Initiative is an important exception, however, and is considered the SOTA classification model in this research. DeepFaune can classify 26 European species with $>$0.90 test-set precision (percentage of true positives) and recall (percentage of true positives correctly identified) and is the only known model to classify both European bears and wild boars [31]. DeepFaune implements a two-step pipeline, with a custom YOLOv8-based detection model [28] and a ViT classifier[1], pre-trained on 142 million images using the DINOv2 self-supervised method [37]. Both models are open source and accessible on GitLab [38].

[1]The original paper describes a ConvNeXt classifier; however, correspondence with the authors confirmed this has since been upgraded to a pre-trained ViT.

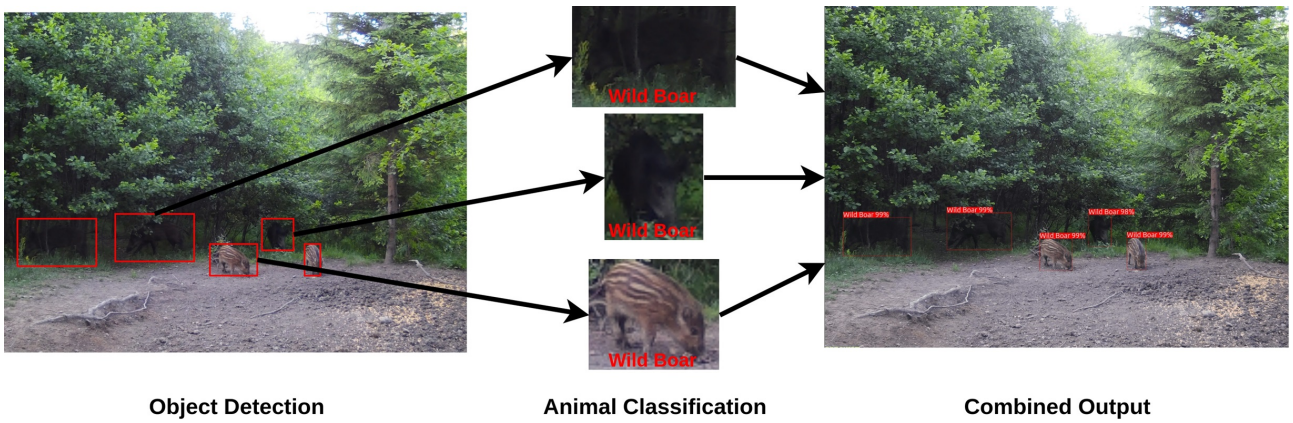

Figure 3: Two-stage pipeline - detection and classification

2.2 Alert systems

The literature on AI-enabled wildlife alert systems is limited. Whytock *et al.* [39] adapted an off-the-shelf camera trap to reduce human-elephant conflict in Gabon. The system ran a TensorFlow lite model on a Raspberry Pi 4B and sent messages via the Iridium satellite network. While the system operated for three months without intervention, an unstable satellite connection resulted in a median alert delay of 7.35 minutes, increasing to five days in some cases. TrailGuard AI, developed by Dertien *et al.* [40] to reduce human-tiger conflict in India, reduced this delay to 30-42 seconds using the 4G network and reported an unverified battery life of 25 months. Zualkernan *et al.* [41] optimise six CNN-based models to run on a Raspberry Pi, Google Coral and NVIDIA Jetson. They found that the Raspberry Pi used the least power but had the slowest processing speed. Ronoh *et al.* and Lathesparan *et al.* also build Raspberry Pi-based systems, but their studies are limited to lab tests [42, 43]. There is no known literature on systems that remotely process images sent from 4G camera traps, likely due to limited network coverage in many HWC locations. However, a blog post from not-for-profit "Hack the Planet" demonstrates this approach for an automated bear deterrent system in Romania [44].

The following section discusses the field visit, data and establishes a baseline performance with DeepFaune's detection and classification models.

3 Data collection and processing

3.1 Field visit

A field visit to FCC's headquarters in Brașov, Romania and several mountain villages was conducted from 13-20 April 2024. The objectives were to (1) transfer the camera trap data, (2) survey sites for installing camera traps, and (3) discuss with FCC the context and specific objectives for the alert system.

After removing duplicates, 2.1 terabytes of data were obtained, consisting of 358,455 images and videos captured over five years to March 2024. This data had previously been used for species occupancy and density studies [6, 7]; however, only 5.3% (19,027 files) had been labelled therefore its full potential was unrealised. FCC stated that the data required for further ecological studies were species, animal count, location, time and date.

The villages of Valea Mare-Pravăț, Rucăr and Lerești were surveyed to inform the field deployment strategy (Fig. 1, 4). All three sites had strong 4G signal; however, none had electricity access. Each faced unique challenges regarding camera placement. Valea Mare-Pravăț had communal grazing land vital for local farmers but frequently damaged by foraging wild boars. The open terrain made wildlife routes unpredictable, requiring multiple cameras for reliable detection. Similarly, in Rucăr, nested in a valley surrounded by dense forest, local rangers noted that while bears used a regular route to enter the village, wild boars approached from all directions. Lerești was ideal for an initial deployment. Fencing created common routes for wild boars and bears to enter the village, simplifying camera trap placement.

Figure 4: (top left) grazing land damage in Valea Mare-Pravăț (top right) crops and orchards in Lerești (bottom left) common wild boar route in Lerești (bottom right) forests in Rucăr increase conflict risk.

FCC's rangers outlined four essential requirements for the alert system. First, a minimum three-month battery life to match their existing camera traps. Second, the ability to handle multiple false triggers as passive infrared motion sensors are difficult to calibrate. Third, robust weatherproofing as minor water ingress quickly results in corrosion. Fourth, the capability to process infrared photos as most wildlife incursions occur at night.

3.2 Establishing ground truth labels

Ground truth species labels are required to train a classification model. FCC provided a database of labels for 19,027 images and videos; however, they were not directly linked to specific files. Prioritising images only, an algorithm was built to match each label with an image file using its metadata. Of the 10,495 image labels provided, 8,457 (80.6%) were matched to a specific file. An advantage of this method was that sequences – multiple photos taken in quick succession – were also labelled. As FCC had not labelled sequences, this augmented the labelled dataset to 19,597 images.

Past research shows that volunteer labellers have a precision of 96.6% [45]; therefore FCC's species labels were compared with labels output by DeepFaune to establish a ground truth. FCC's species classes were mapped to DeepFaune's for direct comparison (Appendix B). Of the 19,597 images, FCC and DeepFaune's labels matched 17,142 (87.5%), and the ground truth was assigned accordingly. Images of bison and otter, for which no DeepFaune equivalents existed, were assigned FCC's label. Images that FCC had labelled "unknown" and DeepFaune had labelled "empty" were assigned as "empty". A tool was built to manually label the remaining 2,066 images (Appendix C).

3.3 Data analysis and modelling strategy

Establishing ground truth labels enabled the assessment of DeepFaune's performance on FCC's dataset and the systematic identification of areas for improvement.

Table 1: DeepFaune's performance versus human baseline

Class	n	FCC			DeepFaune		
		Precision	Recall	F1 Score	Precision	Recall	F1 Score
Badger	422	0.95	0.95	0.95	0.99	0.93	0.96
Bear	2936	0.97	0.98	0.97	1.00	0.96	0.98
Bird	621	0.97	0.99	0.98	1.00	0.89	0.94
Cat	362	0.98	0.98	0.98	1.00	0.97	0.98
Chamois	33	1.00	0.97	0.98	1.00	0.82	0.90
Cow	143	1.00	0.81	0.90	0.68	1.00	0.81
Dog	528	0.88	0.90	0.89	0.98	0.93	0.96
Equid	217	0.73	0.88	0.80	0.97	0.97	0.97
Fox	2256	0.98	0.98	0.98	1.00	0.97	0.98
Goat	150	0.94	0.73	0.82	0.99	0.96	0.98
Human	44	0.53	0.55	0.54	0.38	0.86	0.53
Lagomorph	107	0.98	0.95	0.97	0.98	0.96	0.97
Lynx	1147	0.97	0.99	0.98	1.00	0.87	0.93
Mustelid	167	0.92	0.94	0.93	0.96	0.93	0.94
Red deer	3667	0.98	0.98	0.98	0.99	0.96	0.97
Roe deer	1761	0.98	0.97	0.98	0.98	0.95	0.97
Sheep	84	1.00	0.88	0.94	0.99	0.89	0.94
Squirrel	47	1.00	1.00	1.00	0.92	0.98	0.95
Vehicle	14	0.88	1.00	0.93	0.93	1.00	0.97
Wild boar	3293	0.99	0.98	0.98	1.00	0.95	0.98
Wolf	610	0.96	0.96	0.96	0.99	0.97	0.98

Table 1 shows DeepFaune performs well across all classes, particularly wild boar and bears (precision and recall $>$0.95). While DeepFaune was more precise, FCC had better recall. FCC requested that the alert system prioritise recall. Therefore, a research objective of improving on DeepFaune's recall for bears and wild boars was set. Specifically, the target would be to match FCC's human labels by increasing recall from 95.7% to **97.5% for bears** and from 95.3% to **97.6% for wild boar**. This was considered feasible as similar accuracy had been achieved by DeepFaune's authors [31]. A confusion matrix was analysed to determine how to achieve this objective (Fig. 5). Respectively, 2.1% and 2.3% of mislabelled wild boar images were classified as "empty" (a detection failure) and "unknown" (a classification failure). A similar pattern was observed for bears, thus suggesting two ways forward.

First was to replace DeepFaune's animal detection model with a more accurate pre-trained alternative. While MegaDetector is considered the SOTA detection model, DeepFaune's authors found that it created a processing bottleneck (2-3 seconds/image) for their application of analysing large image libraries on personal computers. Therefore, they trained a custom model, which was faster (0.3 seconds/image) but less accurate (79.6% recall on large animals versus 85.1% for MegaDetector) [31]. Since 2-3 seconds is acceptable for an alert system, switching to MegaDetector would be investigated. Fine-tuning or training a new detection model was not considered as this would require manually annotating a large dataset with bounding boxes and good pre-trained models already existed (i.e. MegaDetector).

Second was to fine-tune DeepFaune's species classification model with FCC's dataset. DeepFaune's ViT classifier was pre-trained on 142 million images and fine-tuned with 787,575 images, mostly from France [31]. FCC's dataset was small in comparison; however, it could add new information to help the model make predictions in the Carpathian Mountains' varied landscape, vegetation and lighting conditions.

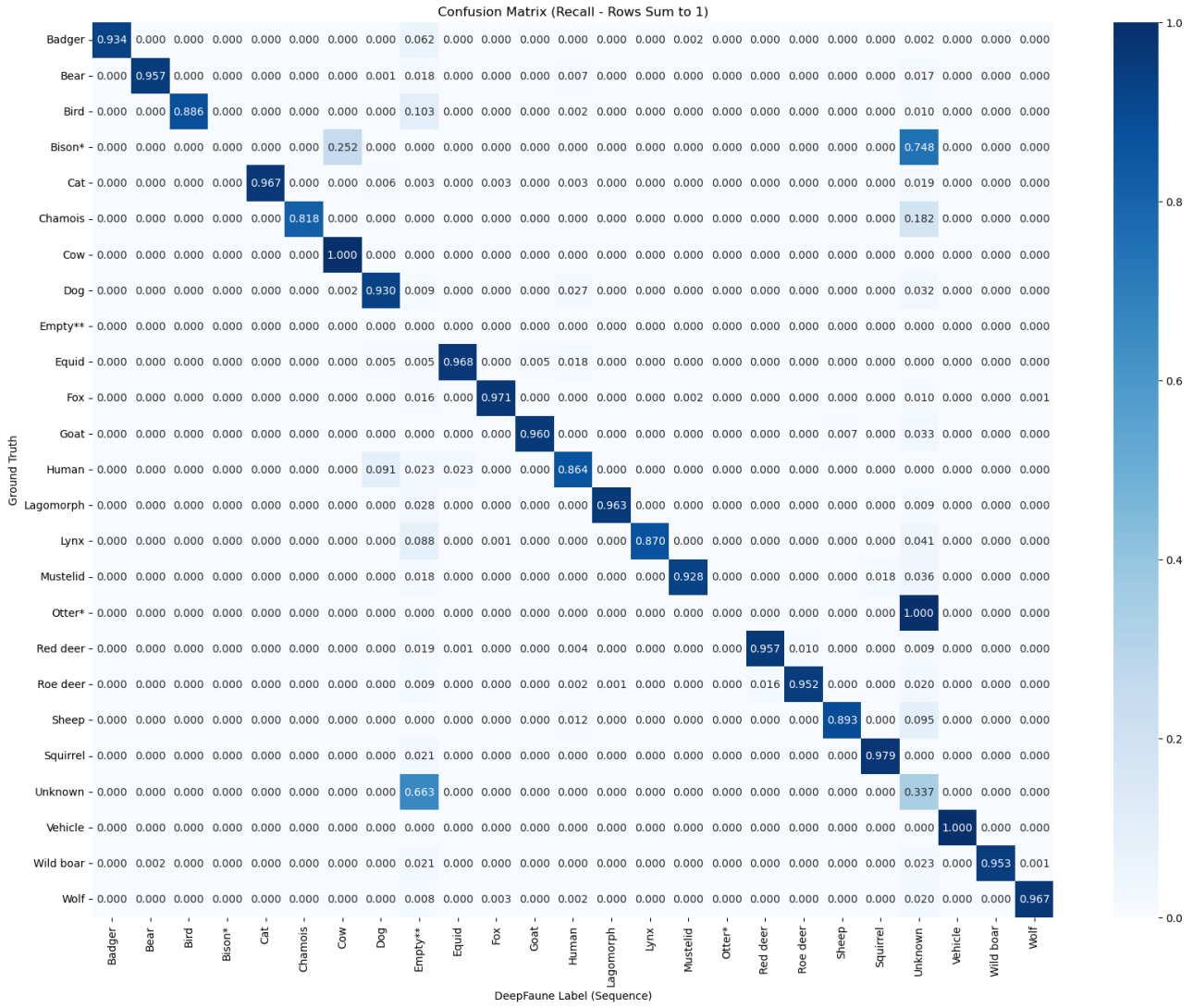

Figure 5: Confusion matrix

The following section details these approaches and their results.

4 Model selection and fine-tuning

4.1 Detection: MegaDetector versus DeepFaune

The MegaDetector and DeepFaune detection models were run on FCC's entire labelled dataset. Since the models were pre-trained and unmodified, no training or validation sets were used. Both models output four classes – empty, human, vehicle and animal – and a confidence level between 0 and 1 (no confidence threshold was used). As humans, vehicles and animals occasionally occurred together, these classes were assigned a single "non-empty" label. FCC's dataset was heavily biased, with 18,927 non-empty images and only 112 empty images. It was therefore augmented with empty camera trap images obtained from the LILA BC repository [36] to create a 50:50 empty/non-empty split. A dataset from a mountainous landscape in Idaho, USA was selected due to the accuracy of its empty labels which, unlike others, did not include people or man-made objects [46]. 18,815 empty images were sampled randomly and processed with MegaDetector and DeepFaune's detection models. These results were combined with the results from FCC's dataset, and the models' performance was calculated for confidence thresholds from 0.01-0.98, where any "non-empty" label with a confidence level below the threshold was re-classified as "empty" (Fig. 6).

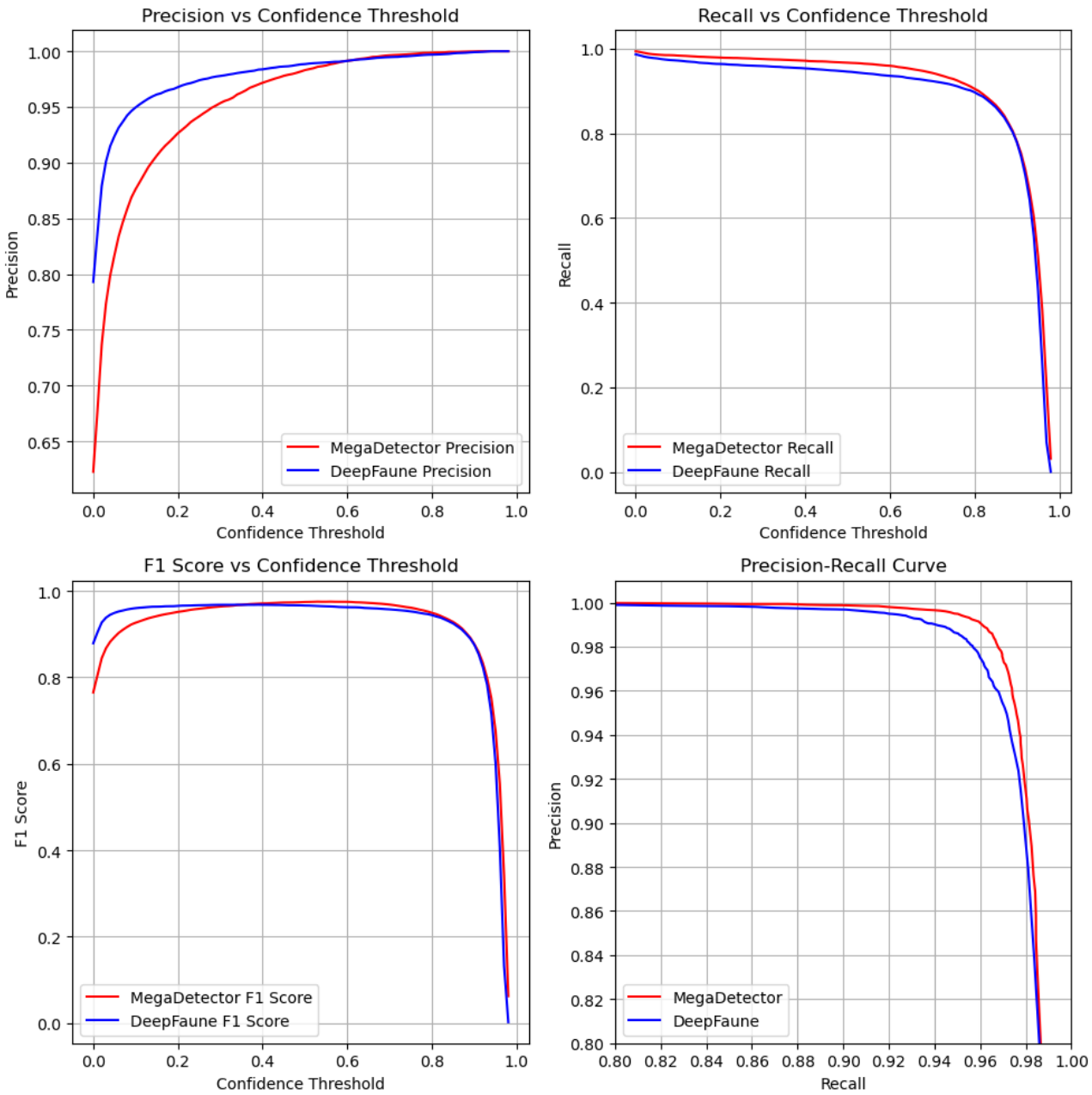

Figure 6: Detection model performance comparison

Table 2: MegaDetector detection performance

Threshold	Precision	Recall	F1 Score
0.10	0.8762	0.9834	0.9267
0.15	0.9063	0.9806	0.9420
0.20	0.9265	0.9787	0.9519
0.25	0.9415	0.9772	0.9590

Table 3: DeepFaune detection performance

Threshold	Precision	Recall	F1 Score
0.01	0.8051	0.8369	0.9044
0.02	0.8790	0.9810	0.9272
0.03	0.9009	0.9791	0.9384
0.04	0.9148	0.9778	0.9452

MegaDetector had better recall and DeepFaune was generally more precise for a given confidence threshold. However, the precision-recall curve shows that MegaDetector managed this trade-off better and is the superior model when optimising for either. Tables 2 and 3 show that for a target recall of 0.98 (aligned with the research objective), MegaDetector required a confidence threshold of 0.15, resulting in 0.91 precision and 0.94 F1 score. This compared favourably with DeepFaune, which required a threshold of 0.02, resulting in 0.88 precision and 0.92 F1 score. It was noted that MegaDetector's precision deteriorated rapidly below

a confidence threshold of 0.20, increasing the risk of false positives. Thus, different confidence thresholds would be tested during deployment to optimise the user experience. A limitation of this comparison was that both models' ability to detect multiple objects in one photo was not evaluated. This was because there was no ground truth for the number of animals/humans/vehicles in each image, only if any were present.

4.2 Classification: Fine-tuning DeepFaune

4.2.1 Train/validation/test set creation

Fine-tuning DeepFaune's classification model required cropped animal images, therefore MegaDetector was run on FCC's dataset. As many photos contained multiple animals, this augmented the dataset from 19,793 full-sized to 26,044 cropped images (Fig. 7). All animals in the same photo were assigned the same species label.

Figure 7: Sample of cropped wild boar photos

The images were sorted into training, validation and test sets. The dataset was first assessed for biases that could impact the training or lead to shortcut learning (i.e. the model learning that a particular camera position is associated with a specific species [47]). The distribution of images across species, day/night, season, site location and image quality was assessed (Appendix D). The key findings were:

- Wild boar and bear were the largest animal classes, excluding red deer. This was accepted as it would bias performance towards these priority species.
- The proportion of day/night photos was balanced for bears but biased towards night (64%) for wild boars. This bias was accepted as wild boars are more active at night.
- The proportion of photos across seasons was balanced for wild boars; however, there was a notable reduction in bear photos during the winter. As this distribution reflected the expected hibernation behaviour, this bias was accepted.
- No one site had more than 10% and 11% of wild boar and bear captures, respectively. While five sites only captured wild boars and three only captured bears, these represented a small proportion of

total captures.

- 88% of wild boar and 85% of bear photos were labelled "medium" or "good" quality, indicating that the dataset was not biased towards "bad" or "excellent" quality photos.
- 77% of photos belonged to sequences and risked leaking information if split across sets.

The dataset was grouped by image sequence to avoid data leakage and split 70:15:15 between the training, validation and training sets. The resulting distribution is shown in Appendix E. Splitting by sequence resulted in a less precise set split at the class level; however, the deviations were considered minor and comparable to other studies.

4.2.2 Fine-tuning

A conservative fine-tuning strategy was adopted as DeepFaune's classification model had been pre-trained on a much larger dataset than FCC's. The first 18 of the ViT's 24 layers were frozen (75%), meaning their parameter remained unchanged. Early ViT layers detect detailed features such as edges and textures, while later layers identify more abstract relationships [48]. Thus, fine-tuning the latter layers aligned with the objective of providing contextual information. After some experimentation, a low learning rate of 10^{-6} was used as higher rates quickly overfitted. In addition to the training and validation loss, the training loop was designed to monitor precision and recall for wild boars, bears and all classes combined. A custom cross-entropy loss function was created, L_{custom}, which gave the option to penalise poor performance on wild boars and bears:

$$L_{\text{custom}} = L_{\text{CE}} + \lambda \cdot L_{\text{penalty}}$$

where L_{CE} was the cross-entropy loss over all classes, λ was the penalty factor (≥ 0), and L_{penalty} was the cross-entropy loss for the wild boars and bears only.

Fine-tuning was performed on an NVIDIA A100 40GB GPU in the JASMIN computing cluster [49] due to the ViT's size[2]. The entire dataset was pre-loaded to the GPU to minimise data-transfer bottlenecks. The maximum viable batch size of 32 was used to increase training stability.

Two models from the experimentation were selected for further discussion: one with a standard loss function and one with the wild boar/bear penalty factor, $\lambda = 10$. Both models were trained for 30 epochs, and the models' state at the best epoch was selected. The best epoch was defined as the epoch with the maximum harmonic mean of wild boar and bear recall (see Appendix E for training statistics).

Table 4: Fine-tuning performance: bears

Model	Validation (n = 352)			Test (n = 526)		
	Precision	Recall	F1	Precision	Recall	F1
DeepFaune	**0.9620**	0.9347	0.9481	**0.9915**	0.8916	0.9389
Fine-tuned (balanced loss function)	0.9499	0.9688	**0.9592**	0.9740	0.9259	0.9493
Fine-tuned (biased loss function)	0.9421	**0.9716**	0.9566	0.9820	**0.9354**	**0.9581**

Table 5: Fine-tuning performance: wild boars

Model	Validation (n = 703)			Test (n = 516)		
	Precision	Recall	F1	Precision	Recall	F1
DeepFaune	**0.9792**	0.9360	0.9571	**0.9567**	0.9419	0.9492
Fine-tuned (balanced loss function)	0.9704	0.9801	**0.9752**	0.9380	0.9671	**0.9523**
Fine-tuned (biased loss function)	0.9402	**0.9844**	0.9618	0.9212	**0.9748**	0.9473

[2]304 million parameters vs 25.6 million for ResNet-50.

Table 6: Fine-tuning performance: all classes (weighted by support)

Model	Validation (n = 3669)			Test (n = 3557)		
	Precision	Recall	F1	Precision	Recall	F1
DeepFaune	0.9355	0.9117	0.9204	0.9343	0.9157	0.9229
Fine-tuned (balanced loss function)	**0.9548**	**0.9531**	**0.9522**	**0.9477**	**0.9452**	**0.9455**
Fine-tuned (biased loss function)	0.9437	0.9409	0.9405	0.9439	0.9407	0.9414

Tables 4-6 detail the models' performance for bears, wild boars and all classes. The results show that the bias loss model maximised recall for bears and wild boars as desired. Recall for bear and wild boars increased by 3.7% and 4.8% respectively on the validation set, and 4.4% and 3.3% on the test set, suggesting the performance is generalisable. The balanced loss model consistently outperformed the others when assessed on all species. DeepFaune outperformed on precision for wild boar and bear, demonstrating the precision-recall trade-off in a training protocol optimised to maximise recall. However, the balanced loss model's outperformance on all classes suggests that there was a universal benefit to fine-tuning DeepFaune on FCC's dataset.

The following section discusses the deployment of these models in an alert system.

5 Field deployment

5.1 Alert system development

5.1.1 Hardware selection

Two hardware strategies were assessed: (1) processing images on-camera and only transmitting an alert when necessary, and (2) transmitting every image for remote processing. Processing on-device with a Raspberry Pi 4B was initially investigated (Fig. 8.a) as this was the dominant approach in the literature. This was quickly ruled out, however. While DeepFaune's models were successfully run on the Raspberry Pi, there were significant limitations regarding power consumption, memory, processing, weatherproofing, scalability and adaptability. A discussion with the developer of the alert system in Gabon [39] verified these limitations and insights from the field visit validated that they would be impractical. Therefore, it was decided that a commercial 4G camera with remote processing would be used (Fig. 8.b-c).

(a) Raspberry Pi-based prototype **(b)** Wilsus camera trap **(c)** UOVision camera trap

Figure 8: Raspberry Pi-based prototype and commercial traps

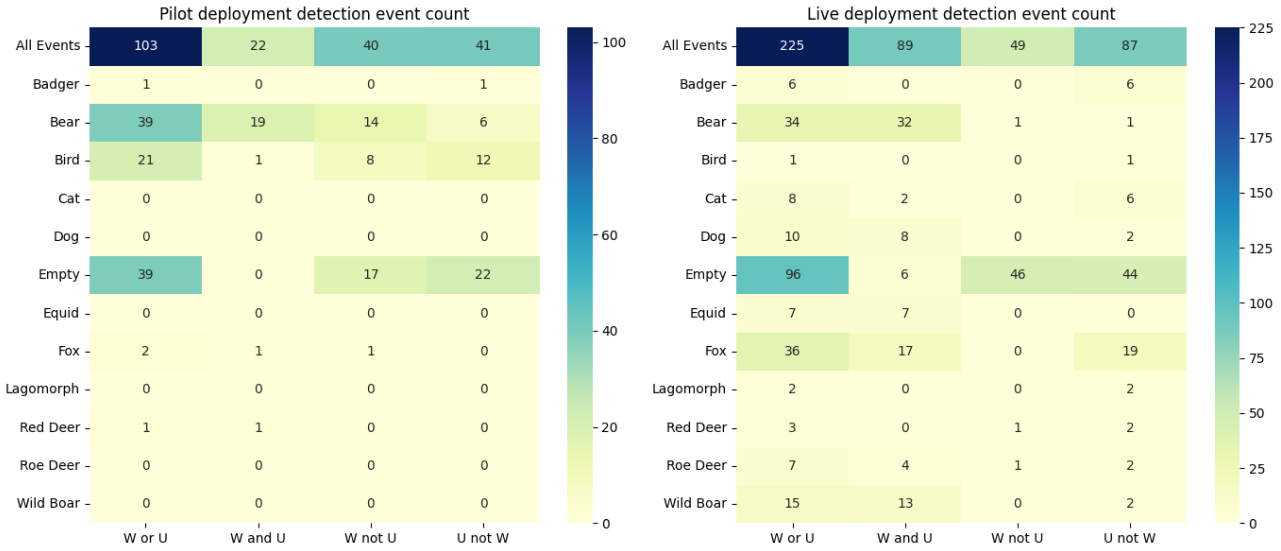

Figure 9: Wilsus (W) and UOVision (U) motion sensitivity comparison. Photos arriving within 60 seconds of each other are grouped into one "detection event".

5.1.2 Pilot deployment

Two 4G camera traps were deployed at a wildlife feeding site 1.5km from Lerești (Fig. 1) on 14 May 2024. This site was chosen to maximise wildlife captures while the alert system was in development. Two camera brands – Wilsus and UOVision – were deployed to compare their motion sensitivity. The cameras were configured to send photos to a dedicated inbox[3]. A basic alert system based on DeepFaune's original models was developed to regularly check this inbox and send an alert to a Telegram group if an animal was detected. This script was deployed on a Raspberry Pi, which acted as an always-on server, and the system's performance was assessed over two weeks from 15-28 May 2024. FCC's lead wildlife researcher joined the Telegram group to provide initial feedback. The alert system ran uninterrupted over the two weeks and achieved precision and recall of 99.4% and 95.3% on the 107 images it *received*. It was observed, however, that the system's *overall* recall was affected by the cameras' sensitivity. Excluding false triggers ("empty"), there were 23 events that the Wilsus camera detected, but UOVision did not, and 19 events that UOVision detected and Wilsus did not (Fig 9). FCC stated that receiving alerts for all wildlife, not only wild boars and bears, was beneficial. This feedback was incorporated into the final design.

5.1.3 Live deployment

Following the pilot, FCC obtained another pair of Wilsus and UOVision cameras and all four were redeployed to two sites 0.5km north of Albești (Fig. 1) on 29 May 2024. These sites were not surveyed during the field visit; however, FCC's rangers requested them due to a recent increase in bear and wild boar incursions nearby. An advanced version of the alert system was deployed on 3 June 2024 based on insights from the pilot. DeepFaune's detection model was replaced with MegaDetector, with a confidence threshold of 0.15 based on the findings in section 4.1. DeepFaune's classifier was retained to minimise potential sources of unexpected behaviour, and the fine-tuned classifiers were assessed retrospectively. A classification threshold of 0.2 was used to maximise recall. Four additional FCC staff, including the Head Ranger, were added to the Telegram group, and the alerts were translated into Romanian. A virtual machine[4] was set up on FCC's local server, providing a permanent hosting solution.

Example alerts are shown in Figure 10 (English) and Figure 11 (Romanian). Figure 12 shows a high-level process flow diagram for this system. The code and detailed instructions are on GitHub [50].

[3] fcccameratraps@gmail.com

[4] 4x2GHz cores, 8GB RAM, 50GB storage, Ubuntu accessed via SSH.

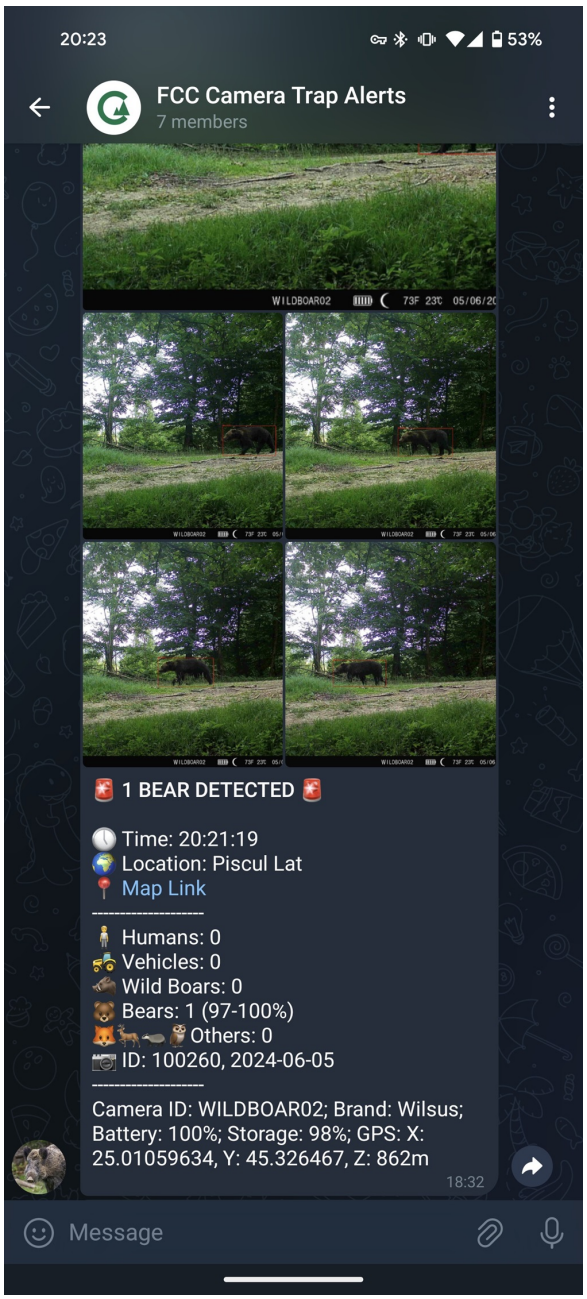

Figure 10: Bear alert (English)

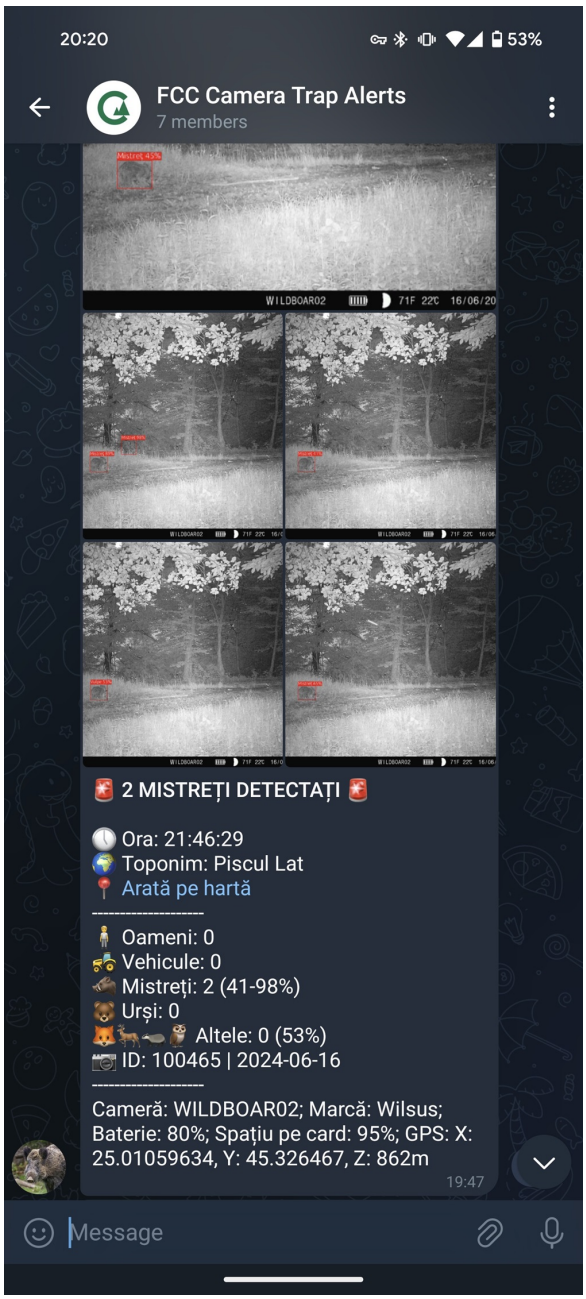

Figure 11: Wild boar alert (Romanian)

The alert messages were designed to provide critical decision-making information while ensuring clarity. A detection event could include five photos, each featuring multiple instances of the same or different species, each with a detection and classification confidence level. This data was processed to generate one of three clear alerts: priority (bear/wild boar), human/vehicle, and non-priority animal (Appendix F). The alert type and species were displayed with actionable data: time, location and a map link. Summarised details were then provided alongside the camera's battery and SD card levels to assist with maintenance.

The system ensured that photos of people were handled ethically. FCC is responsible for monitoring illegal activities, including poaching, and photos of people assist with this. To protect the public's privacy, however, safeguards were integrated to prevent photos of people being sent during daylight hours (06:00-21:00). This approach and additional protocols regarding image storage were confirmed with FCC in writing.

To support wider use of the camera trap data, the system records each detection in a database. Each photo sequence is assigned a Sequence ID and every photo is assigned a unique File ID. Each record is then populated with the detection and classification details, and metadata. The database and photos are stored on FCC's server, and the system emails the latest database to FCC's wildlife team weekly.

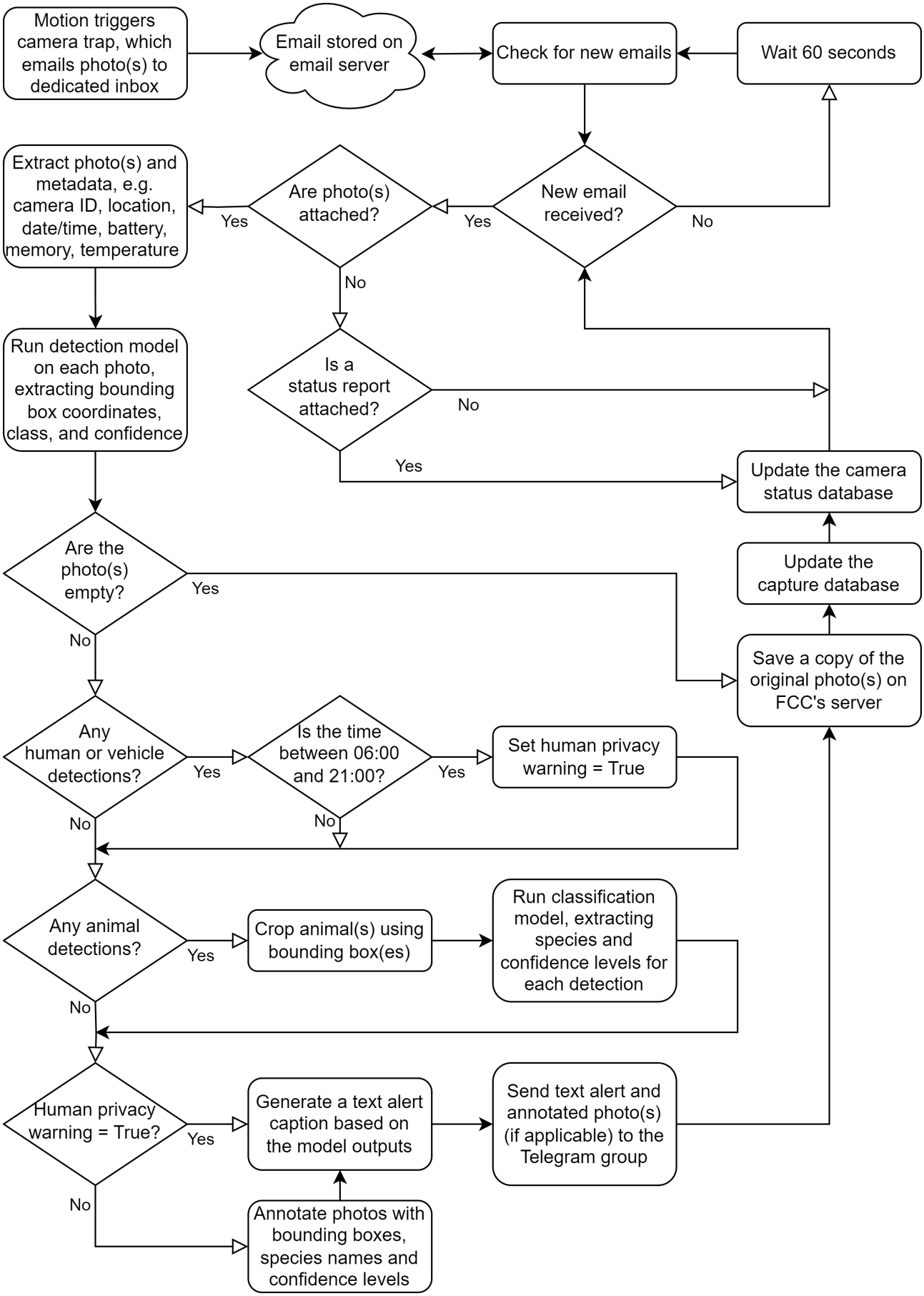

Figure 12: Alert system flow diagram

5.2 Performance

The system's inference performance across both deployments was assessed as of 21 June 2024 (39 days). It had processed 1,211 photos from 377 detection events (328 excluding humans/vehicles). Ground truth labels were manually assigned to every image (Appendix C), excluding 36 where the animal was indistinguishable. The detection and classification pipeline was run retrospectively on the remaining 1,175 photos to assess how the system would have performed with both fine-tuned classifiers. Tables 8-9 show the resulting precision, recall and F1 statistics.

Table 7: Alert system precision

	n	DeepFaune	Balanced Loss	Biased Loss
Badger	7	0.8750	**1.0000**	**1.0000**
Bear	287	**1.0000**	0.9894	0.9859
Bird	29	0.9333	**0.9655**	0.9643
Cat	9	**0.8333**	**0.8333**	**0.8333**
Dog	44	**0.9714**	0.9024	0.8974
Empty	398	**0.9895**	**0.9895**	**0.9895**
Equid	46	0.9773	**1.0000**	**1.0000**
Fox	69	**0.9577**	**0.9577**	0.9189
Human	143	**0.9655**	**0.9655**	**0.9655**
Lagomorph	2	**1.0000**	**1.0000**	**1.0000**
Mustelid	1	**0.3333**	**0.3333**	**0.3333**
Red Deer	7	**0.7143**	0.4000	0.6000
Roe Deer	31	**1.0000**	**1.0000**	**1.0000**
Squirrel	2	**0.1818**	0.0000	0.0000
Vehicle	11	**1.0000**	**1.0000**	**1.0000**
Wild Boar	89	**0.9655**	0.9457	0.9158
Weighted Avg.	1175	**0.9779**	0.9714	0.9670

Table 8: Alert system recall

	n	DeepFaune	Balanced Loss	Biased Loss
Badger	7	**1.0000**	**1.0000**	0.8571
Bear	287	0.9617	**0.9756**	0.9721
Bird	29	**0.9655**	**0.9655**	0.9310
Cat	9	**0.5556**	**0.5556**	**0.5556**
Dog	44	0.7727	**0.8409**	0.7955
Empty	398	**0.9497**	**0.9497**	**0.9497**
Equid	46	**0.9348**	0.9130	0.8696
Fox	69	**0.9855**	**0.9855**	**0.9855**
Human	143	**0.9790**	**0.9790**	**0.9790**
Lagomorph	2	**1.0000**	**1.0000**	**1.0000**
Mustelid	1	**1.0000**	**1.0000**	**1.0000**
Red Deer	7	0.7143	**0.8571**	**0.8571**
Roe Deer	31	**1.0000**	**1.0000**	**1.0000**
Squirrel	2	**1.0000**	0.0000	0.0000
Vehicle	11	**0.9091**	**0.9091**	**0.9091**
Wild Boar	89	0.9438	**0.9775**	**0.9775**
Weighted Avg.	1175	0.9481	**0.9549**	0.9489

Table 9: Alert system F1 score

	n	DeepFaune	Balanced Loss	Biased Loss
Badger	7	0.9333	**1.0000**	0.9231
Bear	287	0.9805	**0.9825**	0.9789
Bird	29	0.9492	**0.9655**	0.9474
Cat	9	**0.6667**	**0.6667**	**0.6667**
Dog	44	0.8608	**0.8706**	0.8434
Empty	398	**0.9692**	**0.9692**	**0.9692**
Equid	46	**0.9556**	0.9545	0.9302
Fox	69	**0.9714**	**0.9714**	0.9510
Human	143	**0.9722**	**0.9722**	**0.9722**
Lagomorph	2	**1.0000**	**1.0000**	**1.0000**
Mustelid	1	**0.5000**	**0.5000**	**0.5000**
Red Deer	7	**0.7143**	0.5455	0.7059
Roe Deer	31	**1.0000**	**1.0000**	**1.0000**
Squirrel	2	**0.3077**	0.0000	0.0000
Vehicle	11	**0.9524**	**0.9524**	**0.9524**
Wild Boar	89	0.9545	**0.9613**	0.9457
Weighted Avg.	1175	0.9614	**0.9620**	0.9568

The results show that the balanced loss model achieved a recall for bears and wild boars of 97.6% and 97.8%, respectively, exceeding the research objective set in section 3.3. The balanced loss model achieved the highest recall and F1 scores on all species except red deer, horses and squirrels. The biased loss model achieved slightly higher recall than DeepFaune on wild boar and bear; however, unlike the balanced loss model, its F1 scores were lower for the two species. As found in section 4.2.2, DeepFaune generally achieved the highest precision. Again, it was noted that the sensitivity of the cameras reduced the overall recall of the system. Wilsus and UOVision achieved respective recalls of 97.1% and 85.3% on bears and 86.7% and 93.3% on wild boars (Appendix G).

To assess the system's conservation impact, an interview was conducted with FCC's Head Ranger on 19 July 2024. He stated that the team had responded to 3-4 encounters due to alerts, including a bear and wild boars. He noted the system's scalability and speed, calling it a significant improvement over directly speaking with farmers. He also noted that Telegram improved data management, as his whole team could access the alerts. Regarding areas for improvement, he noted that the system sometimes makes mistakes, usually when a sequence of photos results in multiple species classifications when there is only one.

The following section revisits the three research objectives and concludes.

6 Discussion and conclusions

6.1 Discussion

6.1.1 Objective 1: Data collection and processing

FCC's data was successfully integrated into a machine learning pipeline. However, its full potential has yet to be realised. Of 19,027 labels, 8,457 (44%) were matched to a specific file. While this resulted in 26,044 labelled images by extracting sequences and cropping individual animals, further augmentation is possible. First, the same matching technique described in section 3.2 could be applied to FCC's 8,532 video labels. Multiple frames could then be sampled and individuals extracted, resulting in a much larger dataset. Second, semi-supervised learning [51] could leverage FCC's entire database of 358,455 images and videos. DeepFaune, fine-tuned with the "true" labelled dataset, could assign "pseudo-labels" to FCC's unlabelled data. While ensuring a minimum number of true labels per batch to reduce confirmation bias [52], the highest confidence pseudo-labels could be used for further fine-tuning. Finally, image transformations,

including flipping, cropping, and contrast adjustments, could synthesise more data [20, 31, 32].

6.1.2 Objective 2: Model selection and fine-tuning

MegaDetector outperformed DeepFaune's detection model when optimising for either precision or recall. This was confirmed on FCC's entire labelled dataset and aligned with the literature [31]; therefore, there was sufficient confidence to directly implement MegaDetector in the live alert system. MegaDetector exceeded expectations during field tests, achieving a recall of 0.99 on individual images. Detection accuracy may be further improved by upgrading to MegaDetector v6, due for release shortly [53].

Fine-tuning DeepFaune's classifier with a balanced loss function increased recall on bears by 3.4% and wild boars by 2.5%. Integrating recall metrics into the training loop improved experimental efficiency, accelerating the discovery of the best model. While validation set recall increased steadily during training (Appendix E), the divergence in training and validation loss suggested that additional data - as discussed in section 6.1.1 - would be beneficial. Nonetheless, despite the divergence, the balanced loss model outperformed on both test and field datasets, aligning with literature on the benefits of domain-specific metrics [54]. The custom loss function successfully increased recall for bear and wild boar by 4.4% and 3.3%, respectively, although precision was impacted. A recommended next step is to conduct additional experiments with hyperparameter optimisation frameworks as Optuna [55].

6.1.3 Objective 3: Field deployment

The detection and classification pipeline achieved the research objective of matching the recall of FCC's manual labels. While additional data and hyperparameter tuning may further improve the models, the field test revealed that important areas for future work lie outside of training more models.

First, a more sophisticated analysis of the model outputs could lead to clearer alerts. For example, analysing the top three species classifications is common practice [19]. This would be useful for detection events containing a group of wild boars, which occasionally resulted in some incorrect classifications and confusing photo annotations. As species rarely mix, assigning all animals in a group the same label based on an aggregate analysis of the top three predictions could generate more realistic alerts. Furthermore, using higher confidence thresholds for difficult-to-detect species (e.g. squirrels) could reduce false positives.

Second, the cameras' sensitivity had a major impact on recall. A rigorous analysis of relative performance was outside the scope of this research; however, initial results suggested that the UOVision camera was more sensitive, especially for smaller animals. A key limitation was that the cameras' response time was not investigated. Wilsus occasionally captured animals when UOVision triggered too late (Appendix G). It is recommended that FCC retain at least one pair of cameras in the same position for further investigation, especially as passive infrared sensors perform differently in winter.

Third, the interview with the Head Ranger highlighted the importance of balancing recall with credibility. FCC stated a strong preference for high recall; however, implementing this is more complicated than reducing confidence thresholds. Some false positives resulted in absurd alerts (e.g. a small dog being classified as an ibex), undermining trust in the system. Further qualitative research into the rangers' perceptions of the alert system is required to ensure its utility as an HWC reduction tool is maximised. A workshop will be organised with FCC this summer will explore these issues further.

6.2 Conclusions

This research achieved three objectives. First, it showed that FCC's camera trap data can be successfully integrated into a machine learning pipeline. Second, it showed that the SOTA classification pipeline for European mammals could be improved by replacing the detector with MegaDetector and fine-tuning the classifier. Third, it showed that these models can be integrated into an alert system that delivers tangible conservation impact. This research is the first known study to use remote processing of 4G camera trap images to operate an HWC alert system. It is also the first known study to design and assess all stages of an AI-based wildlife alert system, from data collection and model training to field deployment and conservation impact.

References

[1] Frank B, Glikman JA, Marchini S, editors. Human–Wildlife Interactions: Turning Conflict into Coexistence. Conservation Biology. Cambridge: Cambridge University Press; 2019. Available from: https://www.cambridge.org/core/books/humanwildlife-interactions/7E526D390A238172CB2719714C5BAFEF.

[2] IUCN. International Union for Conservation of Nature and Natural Resources, Human-wildlife conflict [Resource]; 2022. Available from: https://www.iucn.org/resources/issues-brief/human-wildlife-conflict.

[3] Bhammar H. Human-Wildlife Conflict: Global Policy and Perception Insights World Bank Research [Text/HTML]; 2023. Available from: https://www.worldbank.org/en/programs/global-wildlife-program/brief/human-wildlife-conflict-global-policy-and-perception-insights.

[4] Kucsicsa G, Dumitrică C. Spatial modelling of deforestation in Romanian Carpathian Mountains using GIS and Logistic Regression. Journal of Mountain Science. 2019 May;16(5):1005-22. Available from: https://doi.org/10.1007/s11629-018-5053-8.

[5] Iosif R, Pop MI, Chiriac S, Sandu RM, Berde L, Szabó S, et al. Den structure and selection of denning habitat by brown bears in the Romanian Carpathians. Ursus. 2020 Apr;2020(31e5):1-13. Publisher: International Association for Bear Research and Management. Available from: https://bioone.org/journals/ursus/volume-2020/issue-31e5/URSUS-D-18-00010.1/Den-structure-and-selection-of-denning-habitat-by-brown-bears/10.2192/URSUS-D-18-00010.1.full.

[6] Iosif R, Popescu VD, Ungureanu L, Șerban C, Dyck MA, Promberger-Fürpass B. Eurasian lynx density and habitat use in one of Europe's strongholds, the Romanian Carpathians. Journal of Mammalogy. 2022 Apr;103(2):415-24. Available from: https://doi.org/10.1093/jmammal/gyab157.

[7] Dyck MA, Iosif R, Promberger–Fürpass B, Popescu VD. Dracula's ménagerie: A multispecies occupancy analysis of lynx, wildcat, and wolf in the Romanian Carpathians. Ecology and Evolution. 2022;12(5):e8921. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.8921.

[8] Sin T, Gazzola A, Chiriac S, Rîșnoveanu G. Wolf diet and prey selection in the South-Eastern Carpathian Mountains, Romania. PLOS ONE. 2019 Nov;14(11):e0225424. Publisher: Public Library of Science. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0225424.

[9] Kolchin S. Feeding Associations between the Asiatic Black Bear (Ursus thibetanus) and the Wild Boar (Sus scrofa) in the Sikhote-Alin Mountains. Biology Bulletin. 2018 Jul;45:751-5.

[10] Mustățea M, Pătru-Stupariu I. Using Landscape Change Analysis and Stakeholder Perspective to Identify Driving Forces of Human–Wildlife Interactions. Land. 2021 Feb;10(2):146. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute. Available from: https://www.mdpi.com/2073-445X/10/2/146.

[11] Mihalik B, Frank K, Astuti PK, Szemethy D, Szendrei L, Szemethy L, et al. Population Genetic Structure of the Wild Boar (Sus scrofa) in the Carpathian Basin. Genes. 2020 Oct;11(10):1194. Number: 10 Publisher: Multidisciplinary Digital Publishing Institute. Available from: https://www.mdpi.com/2073-4425/11/10/1194.

[12] Abrahms B, Carter NH, Clark-Wolf TJ, Gaynor KM, Johansson E, McInturff A, et al. Climate change as a global amplifier of human–wildlife conflict. Nature Climate Change. 2023 Mar;13(3):224-34. Publisher: Nature Publishing Group. Available from: https://www.nature.com/articles/s41558-023-01608-5.

[13] Pătru-Stupariu I, Nita A, Mustățea M, Huzui-Stoiculescu A, Fürst C. Using social network methodological approach to better understand human–wildlife interactions. Land Use Policy. 2020 Dec;99:105009. Available from: https://www.sciencedirect.com/science/article/pii/S0264837720309534.

[14] FCC. Fundația Conservation Carpathia | About Us; 2024. Available from: https://www.carpathia.org/about/.

[15] FCC. Fundația Conservation Carpathia | Wildlife; 2024. Available from: https://www.carpathia.org/wildlife/.

[16] Chen G, Han TX, He Z, Kays R, Forrester T. Deep convolutional neural network based species recognition for wild animal monitoring. In: 2014 IEEE International Conference on Image Processing (ICIP); 2014. p. 858-62. ISSN: 2381-8549. Available from: https://ieeexplore.ieee.org/document/7025172.

[17] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. arXiv; 2015. ArXiv:1512.03385 [cs]. Available from: http://arxiv.org/abs/1512.03385.

[18] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv; 2015. ArXiv:1409.1556 [cs]. Available from: http://arxiv.org/abs/1409.1556.

[19] Norouzzadeh MS, Nguyen A, Kosmala M, Swanson A, Palmer MS, Packer C, et al. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proceedings of the National Academy of Sciences. 2018 Jun;115(25):E5716-25. Publisher: Proceedings of the National Academy of Sciences. Available from: https://www.pnas.org/doi/abs/10.1073/pnas.1719367115.

[20] Tabak MA, Norouzzadeh MS, Wolfson DW, Sweeney SJ, Vercauteren KC, Snow NP, et al. Machine learning to classify animal species in camera trap images: Applications in ecology. Methods in Ecology and Evolution. 2019;10(4):585-90. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.13120. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13120.

[21] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al.. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv; 2021. ArXiv:2010.11929 [cs] version: 2. Available from: http://arxiv.org/abs/2010.11929.

[22] Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S. A ConvNet for the 2020s. arXiv; 2022. ArXiv:2201.03545 [cs]. Available from: http://arxiv.org/abs/2201.03545.

[23] Tabak M, Norouzzadeh MS, Wolfson D, Sweeney S, Vercauteren K, Snow N, et al. MLWIC: Machine Learning for Wildlife Image Classification in R; 2018.

[24] Bubnicki JW, Churski M, Kuijper DPJ. trapper: an open source web-based application to manage camera trapping projects. Methods in Ecology and Evolution. 2016 Oct;7(10):1209-16. Publisher: John Wiley & Sons, Ltd. Available from: https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.12571.

[25] Gupta A. ReWilding-Europe-Yolov8 · Hugging Face; 2023. Available from: https://huggingface.co/skylord/ReWilding-Europe-Yolov8.

[26] Beery S, van Horn G, Perona P. Recognition in Terra Incognita. arXiv; 2018. ArXiv:1807.04975 [cs, q-bio]. Available from: http://arxiv.org/abs/1807.04975.

[27] Beery S, Morris D, Yang S. Efficient Pipeline for Camera Trap Image Review. arXiv; 2019. ArXiv:1907.06772 [cs]. Available from: http://arxiv.org/abs/1907.06772.

[28] Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. arXiv; 2016. ArXiv:1506.02640 [cs]. Available from: http://arxiv.org/abs/1506.02640.

[29] Hernandez A, Miao Z, Vargas L, Dodhia R, Lavista J. Pytorch-Wildlife: A Collaborative Deep Learning Framework for Conservation. arXiv; 2024. ArXiv:2405.12930 [cs]. Available from: http://arxiv.org/abs/2405.12930.

[30] Vélez J, McShea W, Shamon H, Castiblanco-Camacho PJ, Tabak MA, Chalmers C, et al. An evaluation of platforms for processing camera-trap data using artificial intelligence. Methods in Ecology and Evolution. 2023;14(2):459-77. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.14044. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.14044.

[31] Rigoudy N, Dussert G, Benyoub A, Besnard A, Birck C, Boyer J, et al. The DeepFaune initiative: a collaborative effort towards the automatic identification of European fauna in camera trap images. European Journal of Wildlife Research. 2023 Oct;69(6):113. Available from: https://doi.org/10.1007/s10344-023-01742-7.

[32] Schneider D, Lindner K, Vogelbacher M, Bellafkir H, Mühling M, Farwig N, et al. Recognizing European mammals and birds in camera trap images using convolutional neural networks. In: 3rd international workshop on camera traps, AI, and ecology; 2023. .

[33] Leorna S, Brinkman T. Human vs. machine: Detecting wildlife in camera trap images. Ecological Informatics. 2022 Dec;72:101876. Available from: https://www.sciencedirect.com/science/article/pii/S1574954122003260.

[34] Bothmann L, Wimmer L, Charrakh O, Weber T, Edelhoff H, Peters W, et al. Automated wildlife image classification: An active learning tool for ecological applications. Ecological Informatics. 2023;77:102231. Available from: https://www.sciencedirect.com/science/article/pii/S1574954123002601.

[35] Morris D. Camera Trap Machine Learning Survey; 2024. Original-date: 2019-04-21T20:27:20Z. Available from: https://github.com/agentmorris/camera-trap-ml-survey.

[36] LILA. LILA BC Datasets; 2024. Available from: https://lila.science/datasets/.

[37] Oquab M, Darcet T, Moutakanni T, Vo H, Szafraniec M, Khalidov V, et al.. DINOv2: Learning Robust Visual Features without Supervision. arXiv; 2024. ArXiv:2304.07193 [cs]. Available from: http://arxiv.org/abs/2304.07193.

[38] Rigoudy N, Dussert G, Benyoub A, Besnard A, Birck C. DeepFaune / DeepFaune Software · GitLab; 2024. Available from: https://plmlab.math.cnrs.fr/deepfaune/software.

[39] Whytock RC, Suijten T, van Deursen T, Świeżewski J, Mermiaghe H, Madamba N, et al. Real-time alerts from AI-enabled camera traps using the Iridium satellite network: A case-study in Gabon, Central Africa. Methods in Ecology and Evolution. 2023;14(3):867-74. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.14036.

[40] Dertien JS, Negi H, Dinerstein E, Krishnamurthy R, Negi HS, Gopal R, et al. Mitigating human–wildlife conflict and monitoring endangered tigers using a real-time camera-based alert system. BioScience. 2023 Oct;73(10):748-57. Available from: https://doi.org/10.1093/biosci/biad076.

[41] Zualkernan I, Dhou S, Judas J, Sajun AR, Gomez BR, Hussain LA. An IoT System Using Deep Learning to Classify Camera Trap Images on the Edge. Computers. 2022 Jan;11(1):13. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute. Available from: https://www.mdpi.com/2073-431X/11/1/13.

[42] Ronoh E, Mirau S, Dida M. Human-Wildlife Conflict Early Warning System Using the Internet of Things and Short Message Service. Engineering, Technology & Applied Science Research. 2022 Apr;12:8273-7.

[43] Lathesparan R, Sharanjah A, Thushanth R, Sathiyavarathan K, Nifras M, Wickramaarachchi W. Real-time Animal Detection and Prevention System for Crop Fields; 2021. .

[44] Broch T. Using AI to keep bears, humans and livestock safe; 2023. Available from: https://engineering.q42.nl/ai-bear-repeller/.

[45] Swanson A, Kosmala M, Lintott C, Simpson R, Smith A, Packer C. Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. Scientific Data. 2015 Jun;2(1):150026. Publisher: Nature Publishing Group. Available from: https://www.nature.com/articles/sdata201526.

[46] Idaho Department of Fish and Game. Idaho Camera Traps. LILA BC; 2021. Version Number: 2021.07.19. Available from: https://lila.science/datasets/idaho-camera-traps/.

[47] Geirhos R, Jacobsen JH, Michaelis C, Zemel R, Brendel W, Bethge M, et al. Shortcut Learning in Deep Neural Networks. Nature Machine Intelligence. 2020 Nov;2(11):665-73. ArXiv:2004.07780 [cs, q-bio]. Available from: http://arxiv.org/abs/2004.07780.

[48] Ghiasi A, Kazemi H, Borgnia E, Reich S, Shu M, Goldblum M, et al.. What do Vision Transformers Learn? A Visual Exploration. arXiv; 2022. ArXiv:2212.06727 [cs]. Available from: http://arxiv.org/abs/2212.06727.

[49] Lawrence BN, Bennett V, Churchill J, Juckes M, Kershaw P, Oliver P, et al.. The JASMIN super-data-cluster. arXiv; 2012. ArXiv:1204.3553 [physics]. Available from: http://arxiv.org/abs/1204.3553.

[50] Ratsakatika T. AI for Wildlife Monitoring. University of Cambridge; 2024. Available from: https://github.com/ratsakatika/camera-traps.

[51] van Engelen JE, Hoos HH. A survey on semi-supervised learning. Machine Learning. 2020 Feb;109(2):373-440. Available from: https://doi.org/10.1007/s10994-019-05855-6.

[52] Arazo E, Ortego D, Albert P, O'Connor NE, McGuinness K. Pseudo-Labeling and Confirmation Bias in Deep Semi-Supervised Learning. arXiv; 2020. ArXiv:1908.02983 [cs]. Available from: http://arxiv.org/abs/1908.02983.

[53] Microsoft. Pytorch Wildlife - MegaDetector. Microsoft; 2024. Original-date: 2018-10-11T18:02:42Z. Available from: https://github.com/microsoft/CameraTraps.

[54] Dash T, Chitlangia S, Ahuja A, Srinivasan A. A review of some techniques for inclusion of domain-knowledge into deep neural networks. Scientific Reports. 2022 Jan;12(1):1040. Publisher: Nature Publishing Group. Available from: https://www.nature.com/articles/s41598-021-04590-0.

[55] Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A Next-generation Hyperparameter Optimization Framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. KDD '19. New York, NY, USA: Association for Computing Machinery; 2019. p. 2623-31. Available from: https://doi.org/10.1145/3292500.3330701.

A Glossary

Term	Definition
Bounding Box	A rectangular box used to define the location of an object in an image.
Camera Trap	A camera designed to capture photographs of wildlife when triggered by movement.
Classification	The process of predicting the class or category of an object in an image.
Confidence Threshold	The minimum confidence score a model must achieve to classify an object.
Convolutional Neural Network (CNN)	A class of deep neural networks commonly used for analysing images.
ConvNeXts	A modern CNN architecture that incorporates features from transformers.
Cross Entropy Loss	A function used to measure the performance of a classification model; lower loss indicates better performance.
Detection	The process of locating objects within an image.
F1 Score	The harmonic mean of precision and recall.
Fine-tuning	Performing additional training on a pre-trained model with new data to adapt the model for a specific task.
Graphics Processing Unit (GPU)	A specialised processor that can efficiently perform machine learning calculations.
Ground Truth	The actual true values, which are used as a reference to train and evaluate models.
Metadata	Additional data embedded in a file, such as the time and date of a camera trap photo.
Precision	The proportion of true positive results among all positive results.
Pre-Trained (Model)	A model that has been previously trained on a large dataset.
Pseudo Label	An estimated label generated by a model for unlabelled data, used in semi-supervised learning.
Raspberry Pi	A small, affordable computer.
Recall	The proportion of true positive results among all actual positive instances in the dataset.
Residual Network (ResNet)	A type of neural network that allows deeper networks by using residual connections to mitigate the vanishing gradient problem.
Self-Supervised Learning	A learning method where the model generates its own labels from the input data.
Semi-Supervised Learning	A learning method that combines a small amount of labelled data with a large amount of unlabelled data during training.
Telegram	A mobile messaging application used for sending and receiving messages.
Test Set	A subset of the dataset used to assess the performance and generalisation of the trained model.
Training	The process of teaching a machine learning model using a dataset.
Training Set	A subset of the dataset used to train a model.
Validation Set	A subset of the dataset used to tune the model's parameters and prevent overfitting during training.
Visual Geometry Group (VGG)	A deep CNN known for its simplicity (consisting of small (3x3) convolution filters) and strong performance in image recognition tasks (notably in the VGG-16 and VGG-19 models).
Vision Transformer (ViT)	A type of neural network architecture that applies transformer models directly to image patches for image classification.
You Only Look Once (YOLO)	A computer vision model specialising in object detection.

B FCC to DeepFaune class mapping

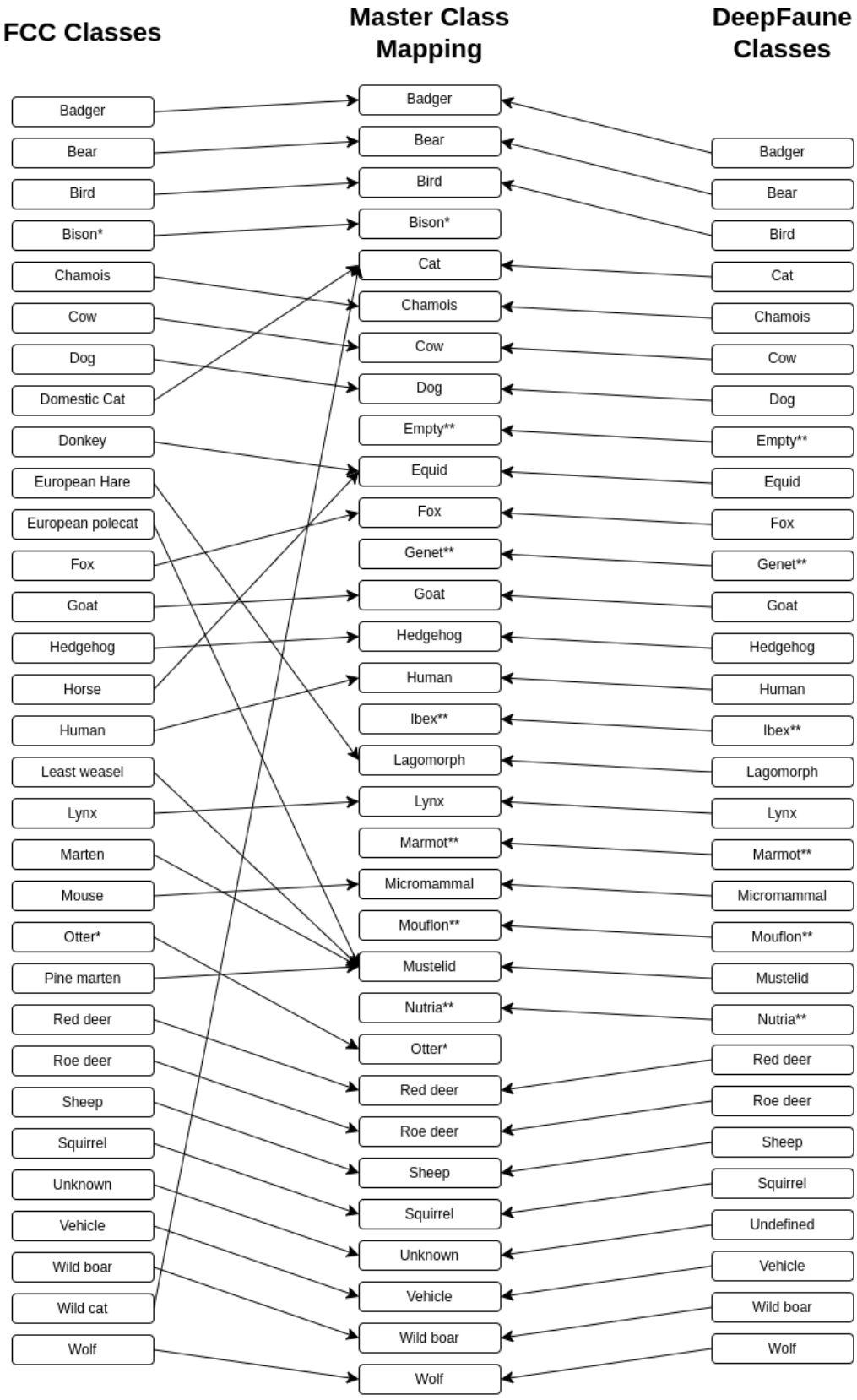

Figure 13: Class mapping from FCC to DeepFaune

C Manual labelling tools

Figure 14: Initial ground truth manual labelling tool

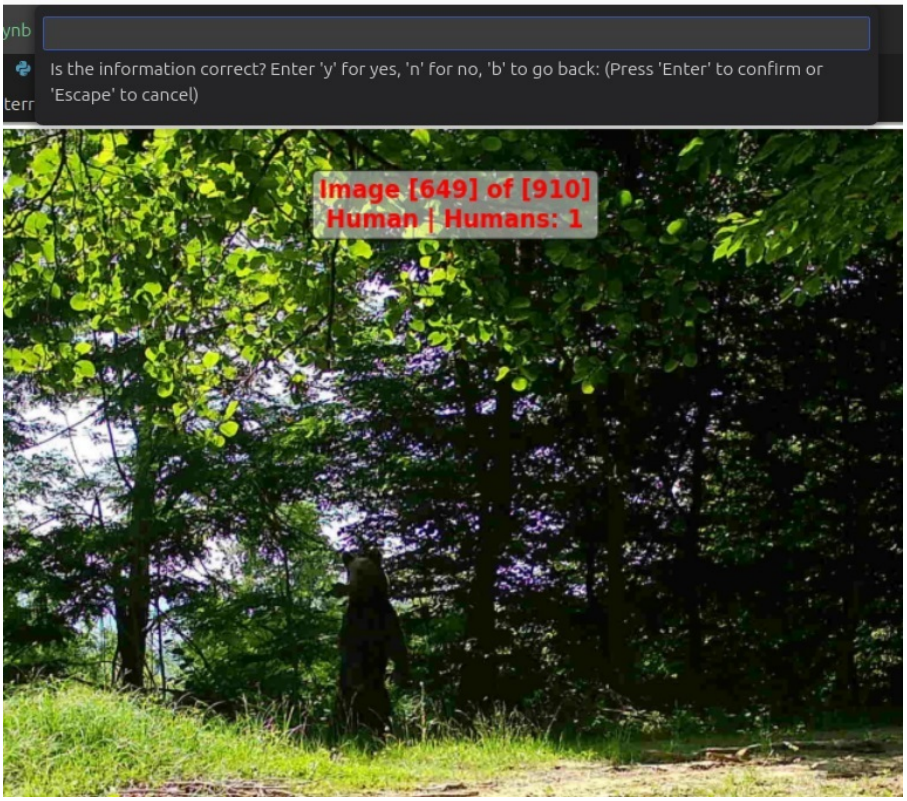

Figure 15: Verification of live field deployment photos. Standing bear misidentified as human.

D FCC dataset bias analysis

Table 10: Species distribution by day/night

Day/Night	Bear	Lynx	Wild boar	Wolf
Day	1475	426	1195	253
Night	1461	721	2098	357
Difference (Night vs Day)	-1%	69%	76%	41%

Table 11: DeepFaune day/night performance for (all species)

Condition	n	Precision	Recall	F1 Score
RGB Cameras (Day)	5893	0.9989	0.9676	0.9830
RGB Cameras (Night)	4130	0.9949	0.9506	0.9722
IR Cameras (Day)	4168	0.9963	0.9813	0.9888
IR Cameras (Night)	5406	0.9954	0.9082	0.9498
All Cameras (Day)	10061	0.9978	0.9734	0.9854
All Cameras (Night)	9536	0.9952	0.9247	0.9587

Table 12: Species distribution by image quality

Image Quality	Bear	Lynx	Wild boar	Wolf
Bad	255	111	391	71
Medium	1521	508	1870	269
Good	1029	445	931	254
Very Good	131	83	101	16

Table 13: Species occurrences by season

Season	Bear	Lynx	Wild boar	Wolf
Autumn	1280	117	1056	112
Spring	992	598	927	172
Summer	427	5	437	30
Winter	237	427	873	296

Table 14: Species occurrences as a percentage of total images by season

Season	Bear (%)	Lynx (%)	Wild boar (%)	Wolf (%)
Autumn	19.43	1.78	16.03	1.70
Spring	17.42	10.50	16.27	3.02
Summer	22.65	0.27	23.18	1.59
Winter	4.37	7.87	16.09	5.45

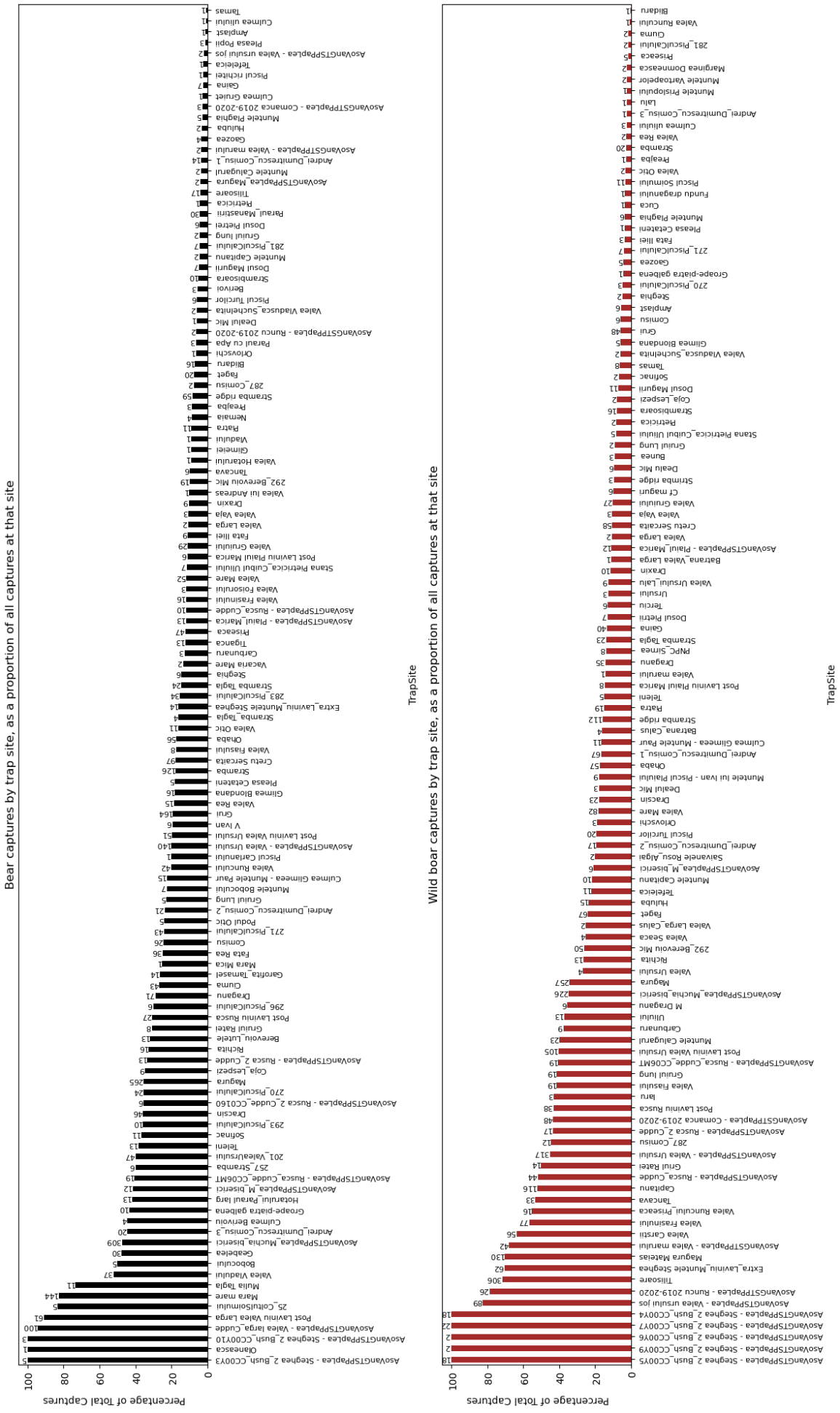

Figure 16: Bear and wild boar captures by site as a percentage of total captures at that site.

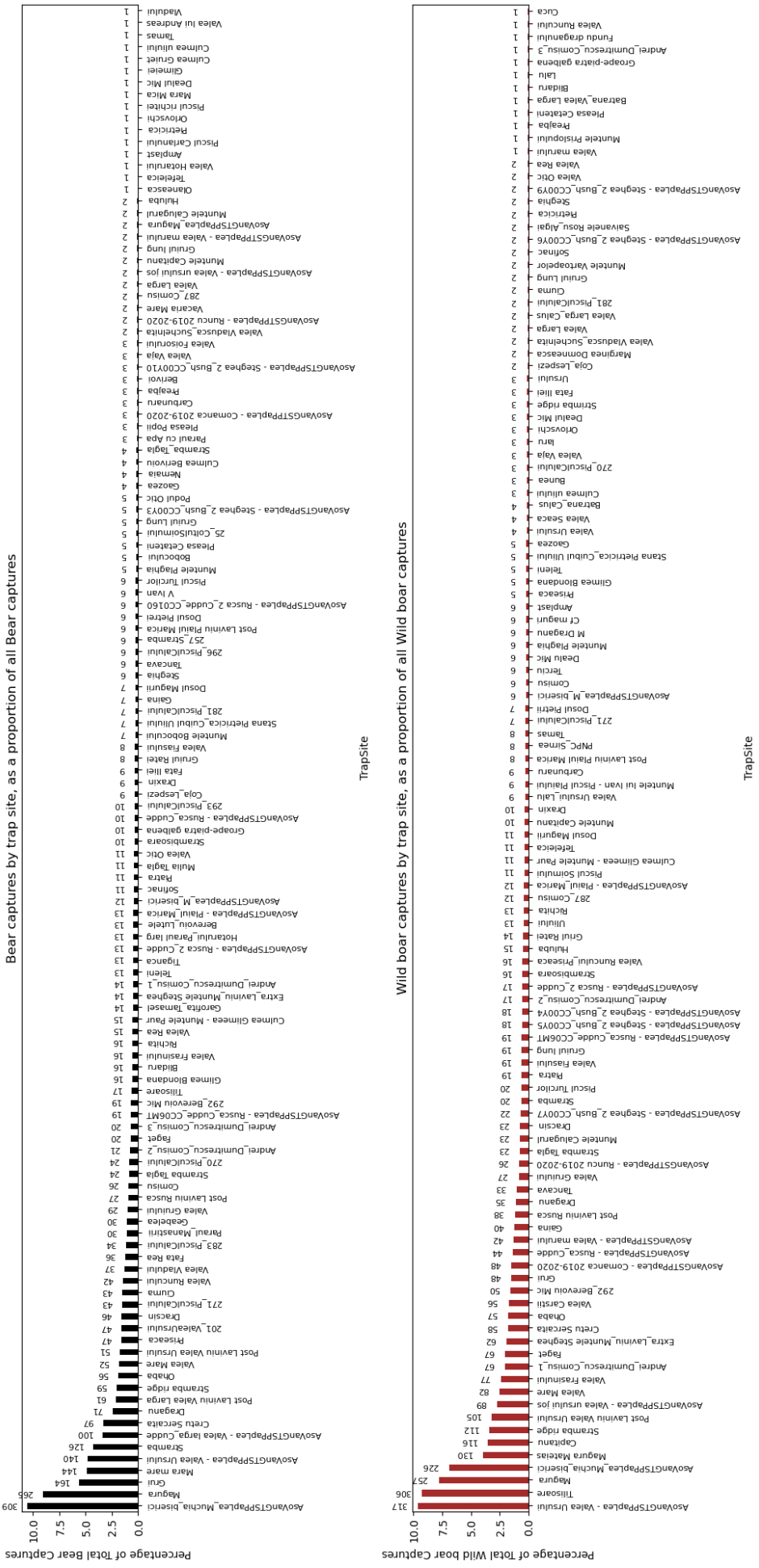

Figure 17: Bear and wild boar captures by site as a percentage of total wild boar captures at all sites.

E Training statistics

Table 15: Train, validation and test split by species

	Total	Train	Validation	Test	Train %	Validation %	Test %
Badger	417	305	59	53	73	14	13
Bear	3207	2329	352	526	73	11	16
Bird	939	695	123	121	74	13	13
Cat	356	276	25	55	78	7	15
Chamois	41	34	4	3	83	10	7
Cow	348	253	65	30	73	19	9
Dog	1110	894	115	101	81	10	9
Equid	446	214	50	182	48	11	41
Fox	2286	1608	328	350	70	14	15
Goat	1593	1350	165	78	85	10	5
Lagomorph	107	81	15	11	76	14	10
Lynx	1172	764	224	184	65	19	16
Mustelid	164	112	31	21	68	19	13
Red deer	4178	2928	632	618	70	15	15
Roe deer	2198	1418	444	336	65	20	15
Sheep	756	275	240	241	36	32	32
Squirrel	48	35	8	5	73	17	10
Wild boar	4486	3267	703	516	73	16	12
Wolf	710	498	86	126	70	12	18
Total	**24562**	**17336**	**3669**	**3557**	**71**	**15**	**14**

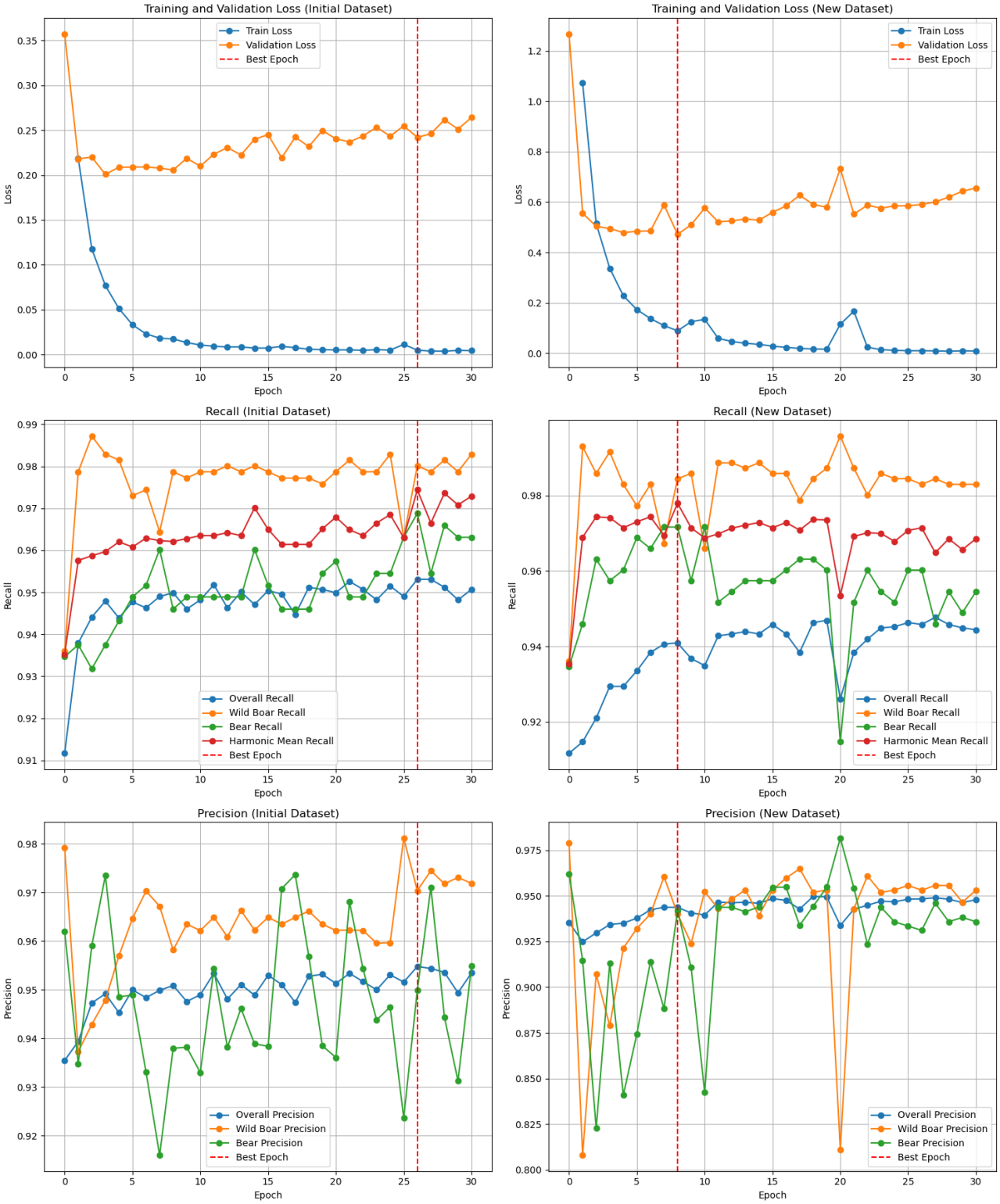

Figure 18: Training statistics. The balanced loss model at epoch 26 and biased loss model at epoch 8 were used.

F Human and non-priority alert examples

Figure 19: Human/vehicle alert, including message warning users not to share photos of people

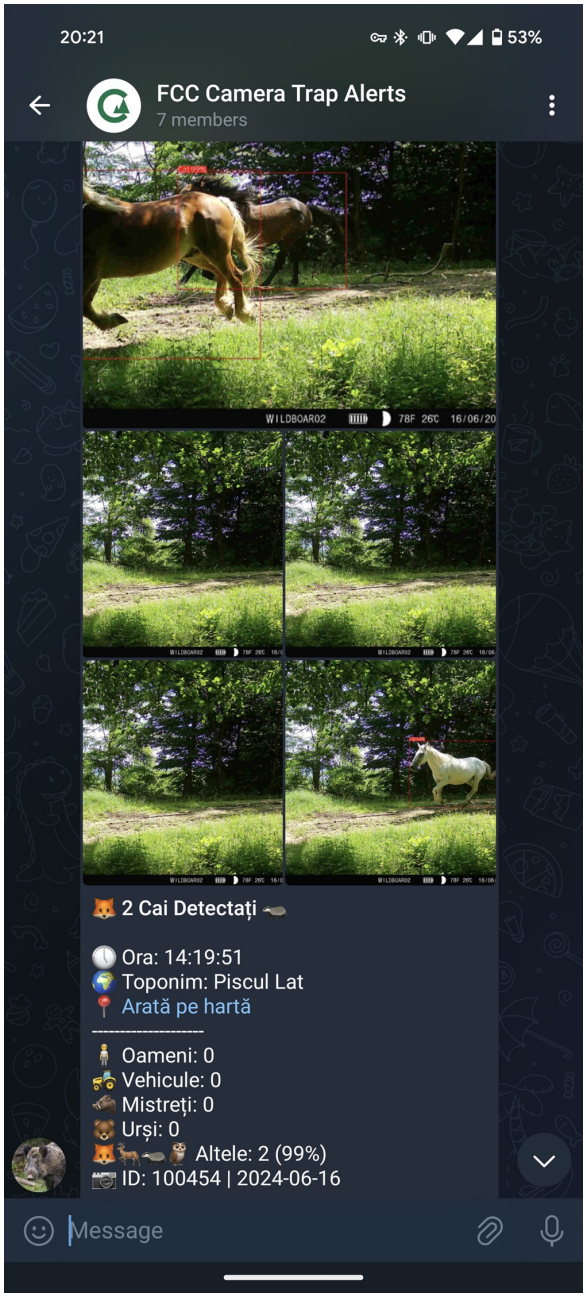

Figure 20: Non-priority animal alert

G Camera recall comparison

Table 16: Pilot (assumes all detection events were captured by either Wilsus or UOVision)

Species	n	Wilsus	UOVision
Badger	1	0.0000	**1.0000**
Bear	39	**0.8462**	0.5641
Bird	21	0.4286	**0.5714**
Fox	2	**0.5000**	**0.5000**
Red Deer	1	**1.0000**	0.0000
Macro Avg.	-	**0.5550**	0.5271
Weighted Avg.	-	**0.6875**	0.5625

Table 17: Live (assumes all detection events were captured by either Wilsus or UOVision)

Species	n	Wilsus	UOVision
Badger	6	0.0000	**1.0000**
Bear	34	**0.9706**	0.8529
Bird	1	0.0000	**1.0000**
Cat	8	0.2500	**0.7500**
Dog	10	**0.8000**	0.7000
Equid	7	**1.0000**	0.8571
Fox	36	0.4722	**0.6111**
Lagomorph	2	0.0000	**1.0000**
Red Deer	3	0.3333	**0.6667**
Roe Deer	7	**0.7143**	**0.7143**
Wild Boar	15	0.8667	**0.9333**
Macro Avg.	-	0.4916	**0.8259**
Weighted Avg.	-	0.6667	**0.7752**

Figure 21: UOVision slow to respond. While the UOVision camera (white banner, bottom left) was triggered by the fox, it responded too late missing the detection, whereas Wilsus (black banner) captured the animal four times resulting in a correct classification.