Science of the Total Environment 1042 (2026) 181926

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

AI and computer vision for wildlife identification in camera trap images: Fine-tuning SpeciesNet outperforms local models for species classification

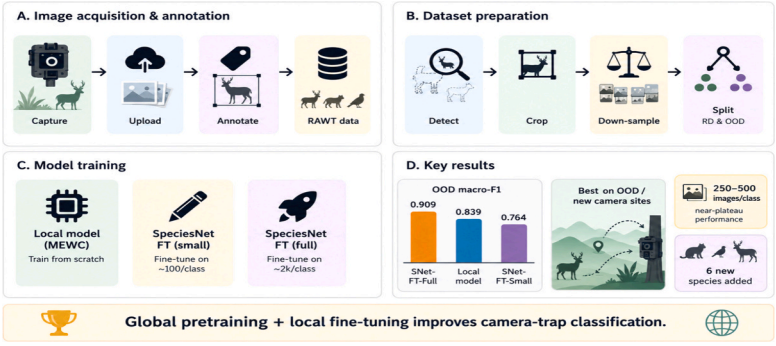
Prakash Palanivelu Rajmohan [a,b], Renuka Sharma [a], Zachary Amir [a,c,d], Tom Bruce [a], Barry W. Brook [e], Dan Morris [f], Matthew Scott Luskin [a,d,g,*]

[a] Wildlife Observatory of Australia (WildObs), Queensland Cyber Infrastructure Foundation (QCIF), Brisbane, Queensland, Australia
[b] School of Electrical Engineering and Computer Science, University of Queensland, Brisbane, Queensland, Australia
[c] Terrestrial Ecosystem Research Network, University of Queensland, Brisbane, Queensland, 4072, Australia
[d] Centre for Biodiversity and Conservation Science, University of Queensland, Brisbane, Queensland, Australia
[e] School of Natural Sciences, University of Tasmania, Hobart, 7005, Tasmania, Australia
[f] Google, Mountain View, CA, USA
[g] School of the Environment, University of Queensland, Brisbane, Queensland, Australia

HIGHLIGHTS

- Fine-tuning global classifier outperforms local models on camera-trap images.
- Just 250–500 images per class yields near-maximal F1-scores (>95%).
- Advantage is largest for out-of-distribution data (new camera sites).
- Adds 6 Wet Tropics species missing from the original model.
- Global pretraining + local fine-tuning unites scalability and precision.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Computer vision
Camera trap
SpeciesNet
Fine-tuning
Transfer learning
Wildlife classification

ABSTRACT

Wildlife camera traps generate millions of images that exceed the capacity of manual processing. Computer vision (CV), a branch of artificial intelligence (AI) and machine learning (ML), helps ecologists process images efficiently. The CV workflow generally starts with animal detection (e.g., with MegaDetector) and then, for those images with animals, the cropped image containing the animal (i.e., snip) is passed to a classifier to identify species. SpeciesNet is an open-source AI/ML classifier that recognises 2498 classes (mostly species-level) globally, and is therefore a 'global model'. However, SpeciesNet has substantial geographic and taxonomic gaps. Ecologists working in areas or with species beyond its scope may therefore build local classifiers for their particular sites. We hypothesised that a blended approach, fine-tuning SpeciesNet, could harness global feature representations and local taxonomic specialisation (i.e., classes limited to the study region). Within this context, we address three questions: (i) How do global, local, and fine-tuned classifiers compare? (ii) How many training images are required? (iii) How does performance vary between random distribution and out-of-distribution

* Corresponding author at: Wildlife Observatory of Australia (WildObs), Queensland Cyber Infrastructure Foundation (QCIF), Brisbane, Queensland, Australia.
E-mail address: m.luskin@uq.edu.au (M.S. Luskin).

https://doi.org/10.1016/j.scitotenv.2026.181926
Received 12 February 2026; Received in revised form 28 May 2026; Accepted 29 May 2026
0048-9697/© 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

testing? We used the Wildlife Observatory of Australia's tagged image repository for the 'Wet Tropics' rainforests (n = 454 camera deployments, 2,184,664 images, 121 species), and refined this to a balanced dataset of the 15 most common species for CV modelling. We found that (i) fine-tuning SpeciesNet delivered the highest performance, often exceeding 95% F1-score, (ii) performance plateaued after 250–500 local training images per class (species) for all three approaches, and (iii) these advantages were pronounced in out-of-distribution testing (i.e., for new cameras withheld from any model training). We conclude that fine-tuning SpeciesNet reconciles the longstanding tension between broad applicability and site-specific precision, accelerating image-to-inference workflows to achieve results within management-relevant timelines. Such advances move cameras further towards being an automated, easy, affordable, and efficient solution for wildlife monitoring, research, and conservation.

1. Introduction

There has been exponential growth in automated environmental monitoring technologies, transforming environmental monitoring and ecological research by generating unprecedented volumes of data with greater efficiency (Ahumada et al., 2020). Wildlife cameras are among the most widely used technologies for detecting terrestrial vertebrates, including for cryptic mammals previously lacking significant observations (Bruce et al., 2025). Beyond species presence, camera-trap datasets can also reveal individual characteristics, behaviour, health and even species interactions (Amir et al., 2022; Murray et al., 2021; Rowcliffe et al., 2014; Sollmann, 2018; Willi et al., 2019). However, the sheer volume of images generated by camera networks creates a significant bottleneck: manually reviewing millions of images is time- and resource-intensive, diverting effort from ecological interpretation and decision-making (Celis et al., 2024). This is a growing problem as networks of wildlife cameras are being deployed globally (Bruce et al., 2025). Efficiently processing camera trap images to produce accurate species identifications is a crucial step in extracting meaningful ecological insights, especially within timelines relevant to management, which we call the images-to-inferences pipeline. A surge in computer-vision (CV) methods has sought to address this by automating object detection and species classification, often lumped in under the artificial intelligence (AI) umbrella (Beery et al., 2019; Besson et al., 2022; Brook et al., 2025). However, these applications of CV are missing for locations and species lacking widely available training datasets and dedicated attention.

There are often two complementary CV steps used in ecological applications: object detectors and species classifiers. The workflow generally starts with a raw image from a camera-trap trigger being sent to an object-detection model, such as MegaDetector (Beery et al., 2019). Object detection models assess whether an image is empty, and if not apply a bounding box to denote the area of the image containing interesting information (i.e., an animal). Such detectors are designed to efficiently locate and delineate coarse-level object categories of interest (<5 classes; often blank, humans, vehicles, and animals). In the second stage, either the entire image containing the animal and its background, or more commonly, just the cropped bounding box area(s) containing the animal snips, is passed to a classifier to identify the species. The complementarity arises from MegaDetector performing efficient object detection across a broad range of image contexts (megapixels, lighting, habitats, and taxonomic scope), whereas a classifier typically accepts its snips but can assign dozens or even thousands of classes. The results from the classifier are then exported into standardized data formats (e.g., spreadsheets, JSON) for use in downstream ecological analysis and decisions for wildlife conservation and monitoring.

1.1. Global CV classifiers

Global species classifiers are trained on geographically and taxonomically diverse datasets spanning multiple continents. These models benefit from large sample sizes (>50M images, >10k per species) and recognize thousands of species, providing utility across ecosystems, and

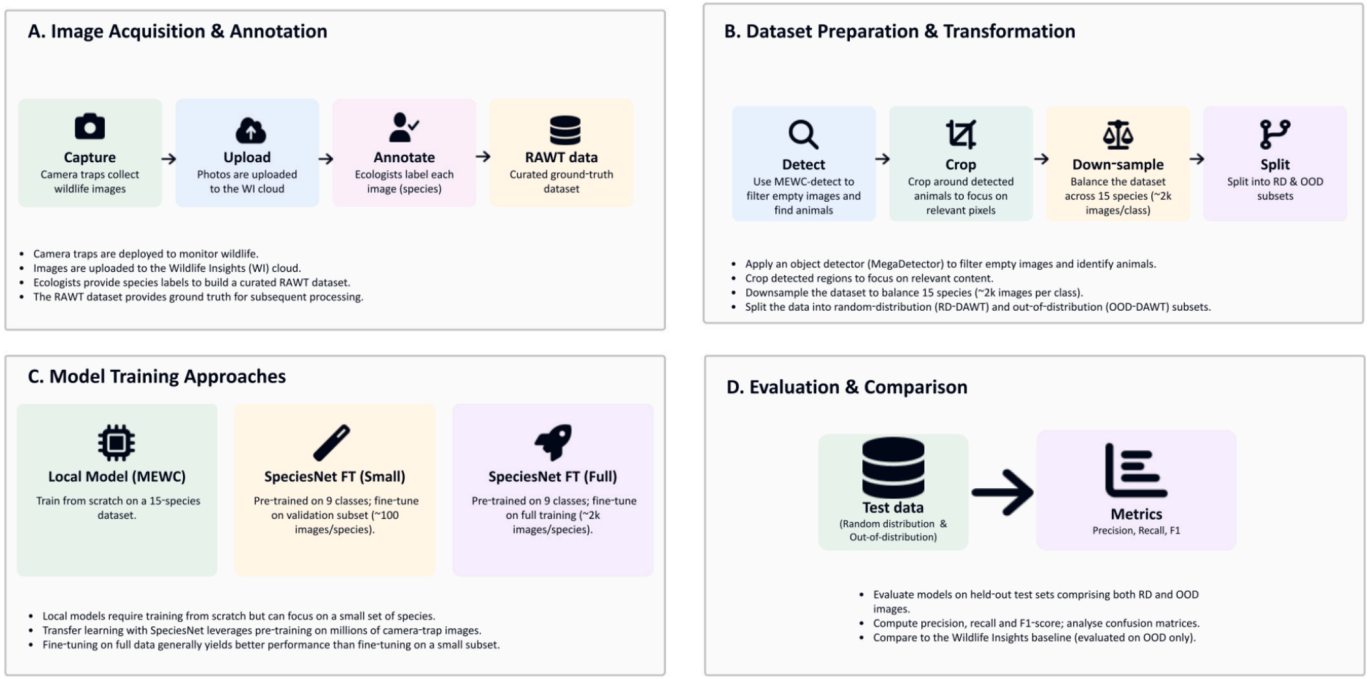


Fig. 1. Computer vision workflow. Computer vision (CV) workflow for the wildlife camera case study presented here.

even worldwide. They utilise sophisticated CV architectures designed for broad predictive generalisation, meaning the models generate reliable predictions on new, unseen data. Their key advantage lies in scalability and broad usability; however, they may exhibit suboptimal performance in regions or taxa that are underrepresented in their training datasets (Gadot et al., 2024).

SpeciesNet (Gadot et al., 2024) is an image classifier that was trained on an extensive and diverse image dataset provided by Wildlife Insights (WI), as well as a variety of publicly-available datasets. SpeciesNet's training data includes images from dozens of countries and has extensive taxonomic coverage but is not exhaustive (2498 taxonomic classes, mostly 'species' but also empty and vehicles). SpeciesNet's training data is biased towards the northern hemisphere's developed nations, and there are gaps where there is slower adoption of cameras or WI (commonly in developing areas with limited internet), regions with smaller physical areas like islands (often where endemism is also high) (Neave et al., 2024; Ngugi et al., 2022; Veron et al., 2019). Australia's Wet Tropics (AWT) rainforests embody the limitations of SpeciesNet because the region has many endemics and WI has also faced limited adoption due to data-sharing hesitancy (see Methods; Bruce et al., 2025). Thus, while global models offer impressive utility and scalability, their efficacy may be limited in some localised ecological and technical conditions. Finally, when deployed through online platforms such as WI, global models can function as black boxes with limited customisation (e.g., inability to add species or restrict predictions to a local species list), which can pose challenges for transparency and reproducibility in research.

1.2. Local CV classifiers

Local classifiers, in contrast, are often trained exclusively on datasets specific to a particular region, or other defined grouping. By learning from locally collected images, these models capture unique visual and ecological characteristics of their focal communities, but their performance deteriorates markedly outside the original training environment (Beery et al., 2018). Local classifiers also offer higher transparency and customisation to specific ecosystems. Local models require domain expertise and computational resources, as well as suitable amounts of labelled images and thus annotation effort (human time). There is also debate about the number of training images required for local models (Shahinfar et al., 2020) and the criteria for acceptable performance, and these models typically perform poorly when generalised beyond conditions similar to their training. Local classifiers can be used on cloud-based camera image management platforms such as TrapTagger and Agouti, and can be trained and used offline with the Mega Efficient Wildlife Classifier (MEWC) workflow (Brook et al., 2025).

1.3. Fine-tuning global classifiers

Fine-tuning is a transfer learning method that adapts a pre-trained global classifier by refining its weights to better recognize local features, such as species traits and habitat backgrounds (Yosinski et al., 2014). This hybrid approach aims to leverage the general visual representations from broad-scale global models and refine them with local data, aiming to combine their strengths. It is expected to enhance accuracy, data efficiency, and robustness to ecological variability, thereby reconciling the long-standing tension between global applicability and local precision. Fine-tuning, in this ecological CV context, often involves two approaches: (i) pruning classes from the global model to match the focal community, and (ii) retraining the global model with supplementary local images, which can include adding entirely new classes absent from the global training set. We clearly delineate these methods to enhance clarity, noting that the MEWC system requires custom coding for class pruning.

1.4. Performance of CV approaches

Despite the increasing adoption of both global and local classifiers in ecology, and recent emergence of fine-tuning, there are few systematic comparisons. Existing research has typically benchmarked individual models within specific contexts, such as MegaDetector for object detection across global datasets (Beery et al., 2019) or local classifiers within targeted ecological projects (Norouzzadeh et al., 2021), without comparisons of their trade-offs (Tabak et al., 2019). Given the rapid and widespread uptake of CV methods, comparative evaluations are urgently needed to clarify the trade-offs between scalability, accuracy, and robustness under both random-distribution and out-of-distribution scenarios. This is especially true for environmental managers and conservation practitioners that increasingly rely on CV-based workflows for high-stakes decisions (Christin et al., 2019).

Here, we address this gap using a systematic and empirical comparison of global, local, and fine-tuned species classifiers. We conduct experiments to (i) evaluate the performance of these three approaches (precision, recall, F1-score), (ii) determine how many local training images are required to meet performance thresholds, including the point of diminishing returns, and (iii) assess differences in performance for random distribution versus out-of-distribution applications. We train and test these models using images from AWT rainforest biodiversity hotspot, where a massive wildlife camera trap dataset has been made publicly available (WildObs, 2025). Our goal is to provide ecologists with practical guidance on how to process camera trap images efficiently.

2. Methods (Fig. 1)

2.1. Assumptions and limitations

We make the following assumptions in this work:

Limited class scope: The study is constrained to a 15-class classification problem, representing a subset of species commonly found in the AWT dataset. This limitation is intended to create a manageable and controlled experimental setting rather than to cover full species diversity.

Single-species classification task: This study focuses on a standard image classification task, where each image is assumed to contain a single identifiable species. In the rare cases where multiple animals were detected in an image by MegaDetector using MEWC, the detection with the highest confidence was used to generate the animal snip for classification. Formal multi-label classification (assigning multiple species labels to one image) is therefore beyond the scope of this study.

Methodological emphasis: Our objective is to evaluate the general effectiveness of three approaches: global SpeciesNet with class restriction, locally trained models, and SpeciesNet fine-tuned on local images. We prioritize methodological comparison over maximizing absolute metrics for any specific dataset, thereby avoiding overfitting and ad hoc tuning that could produce artificially inflated results.

2.2. Wildlife image acquisition and annotation

2.2.1. Data acquisition: camera deployment

The Wildlife Observatory of Australia (WildObs) established an extensive Australian Wet Tropics (AWT) camera-trap dataset by deploying 454 unbaited cameras between 2020 and 2024 (Anderson et al., 2025). These cameras generated 2,184,664 wildlife images of >120 vertebrate species including mammals, birds, and reptiles (Bruce et al., 2022; WildObs, 2025).

2.2.2. Data storage, management, and labelling

All raw camera-trap images were uploaded to WI's Google Cloud Platform buckets. Importantly, none of the AWT images were used to train SpeciesNet.

Images stored in WI were manually annotated by a team of trained ecologists to establish ground-truth species labels, requiring >1000 person hours. These labels served as the authoritative source for all the downstream tasks. Images containing multiple species, blanks, or vehicles were excluded from further analysis. This curated collection of annotated images constitutes the Raw Australian Wet Tropics (RAWT) dataset with a total of 412,822 images, of which 29,817 have been made publicly available in a curated format suitable for replication of our study, Wildlife Observatory of Australia (WildObs, 2025). Following annotation, all labelled images were automatically retrieved from WI using the University of Queensland's Bunya high-performance computer (Bunya, 2024) and stored in Australian Research Data Commons (ARDC) Nectar system for processing.

2.3. Preparation of image dataset

2.3.1. Object detection and double filtering

We first processed the RAWT dataset using MEWC. This workflow employs MegaDetector to remove irrelevant images (blanks, vehicles) and generate bounding boxes for animal localization, producing cropped animal snips. This secondary filtering improved dataset quality by ensuring classification focused exclusively on valid animal detections.

2.3.2. Cropping to animal detections

Using the bounding box outputs, we cropped each image to the detected animal snip, producing the Australian Wet Tropics Snipped (AWT-Snip) dataset. This deliberate design choice minimizes background content to prevent models from overfitting to habitat cues and camera placement, which degrades generalisation to new sites (Beery et al., 2020). Minimising background content forces models to learn morphological and textural features of the species themselves, a strategy shown to improve accuracy and robustness in ecological tasks (Norouzzadeh et al., 2021; Gadot et al., 2024).

The full AWT-Snip dataset exhibited a strongly long-tailed class distribution, with most species represented by fewer than 10 images (Fig. S1). We created a balanced dataset by down-sampling to the 15 most frequent species (Fig. 2), each with at least 2200 images, resulting in the balanced AWT dataset (~391,948 images). Six endemic species in this balanced dataset were absent from SpeciesNet's original training data: *Casuarius casuarius*, *Heteromyias cinereifrons*, *Hypsiprymnodon moschatus*, *Megapodius reinwardt*, *Orthonyx spaldingii*, and *Uromys caudimaculatus*.

2.3.3. Partitioning for robust evaluation

To evaluate model performance under different deployment scenarios, downsampled and balanced was partitioned into two complementary subsets:

Random Distribution (*RD*): RD data refers to images that likely share similar statistical properties and environmental conditions between the training and evaluating phases (same or similar site locations, ensuring consistency in species, habitat, imaging conditions, and camera settings). RD training and test images may be from the same cameras or locations with minimal domain shift. RD is useful when the application is expected to be for new images collected from the same sites it was trained on. For instance, a model trained on data from Sites 1, 2, and 3 collected between 2020 and 2022, would be used to classify new images from the same sites in 2024–2025. The model is expected to generalize well because it has learned relevant features specific to these sites and their species.

Out-of-distribution (*OOD*): OOD data refers to images that exhibit significant differences in statistical properties and environmental conditions between the training and evaluating phases. Specifically, OOD data consists of images collected from different site locations, species, habitats, or imaging conditions, introducing a domain shift that the classifier has not been explicitly trained to handle. For example, if a model is trained using images from Sites 1 and 2 but tested on data from Site 3. This tests if there is overfitting to background features or site-specific cues and instead of species-specific visual characteristics.

We created two evaluation datasets from the balanced AWT dataset by partitioning images into training (90%), validation (5%), and testing (5%) sets with equal representation per class (Table 1). We used a 90/5/5 split to maximise training data while still providing >100 test images per class for stable metric estimation. The fixed test set ensures consistent comparisons across the full training-size gradient (1–100%). Validation was kept small because SpeciesNet's hyper-parameters are pre-optimised; the 5% partition was repurposed for the low-data fine-tuning experiment (SNet-FT-Small). For the out-of-distribution split, this ratio additionally accommodated the requirement of disjoint camera deployments, assigning most cameras to training while preserving wholly independent validation and test deployments to prevent spatial leakage.

For the RD dataset, images were assigned to splits randomly,

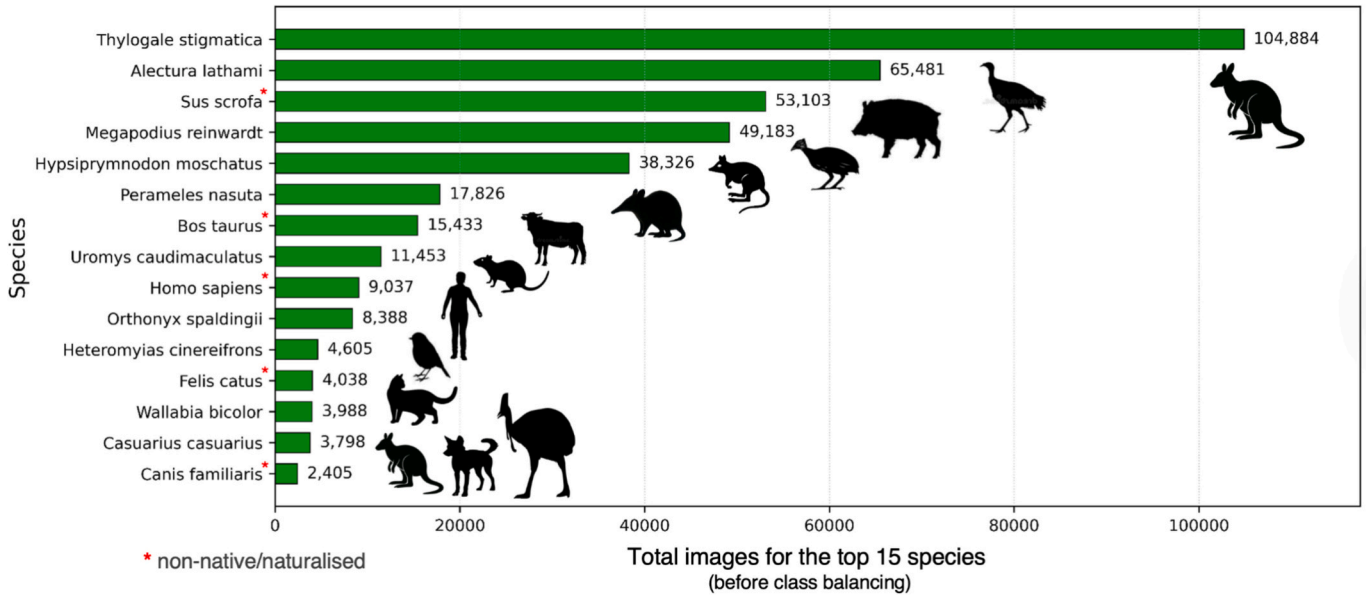


Fig. 2. Species image counts. Total images collected for the 15 species included in the CV training and testing (prior to class balancing).

Table 1
Dataset preprocessing summary. Statistics of each dataset after each pre-processing step. Downsampling was used to balance images per species by removing species with less than 2200 total images.

Datasets	Number of observations	Classes (species)	Used for downstream CV training?
All images	2,184,664	121	No
Species images, produce AWT-Snip	412,822	121	No
Restrict to 15 classes	391,948	15	No
Downsampled and balanced-RD	36,075	15	Yes
Downsampled and balanced-OOD	30,600	15	Yes

meaning images from the same camera could appear in multiple splits. For the OOD dataset, we partitioned by camera deployment, ensuring no camera's images appeared in multiple splits (Fig. 3). Because we use the official pretrained SpeciesNet model and perform no additional hyper-parameter tuning on our data, we repurposed the validation set images for downstream fine-tuning experiments.

2.4. Model training approaches

Types of CV approaches

2.4.1. Local model

We trained a site-specific local model using the MEWC framework, tailored to the 15-species community without external data. The model we chose for this experiment used an EfficientNetV2-M backbone (Tan and Le, 2021) to match the model used in SpeciesNet, but with ImageNet initialization and a new 15-class output layer. Training used Adam (Kingma and Ba, 2014) optimizer with strong augmentation and a two-phase curriculum. In the first phase, the entire EfficientNetV2-M backbone remained frozen for 10 epochs while the new classifier head was trained. In the second phase, the top block of the backbone was unfrozen and fine-tuned over four consecutive 7-epoch stages, with the severity of data augmentation and dropout rate progressively increased in each stage. To evaluate data requirements, we trained models across a gradient of training set sizes (1–100% of available data) with multiple random seeds, conducting this analysis under both random distribution and out-of-distribution splits. Complete training specifications, computational environment, and detailed curriculum are provided in Appendix S2.

2.4.2. SpeciesNet fine-tuned with limited local data (SNet-FT-Small)

We created SNet-FT-Small by fine-tuning SpeciesNet with minimal local data to combine global knowledge with local specialisation. Starting with pre-trained SpeciesNet weights, we pruned the output layer to our 15 target species, reusing weights for the 9 overlapping species and randomly initializing the 6 novel species. Fine-tuning used only the smaller training set (5% of images) with Adam optimizer, data augmentation (Shorten and Khoshgoftaar, 2019), and a 20-epoch budget, updating only the classifier and high-level layers while preserving lower-level features. This approach demonstrates minimal adaptation of a global classifier using limited local examples. Complete implementation details are provided in Appendix S3.

2.4.3. SpeciesNet fine-tuned with all local data (SNet-FT-Full)

We developed SNet-FT-Full by comprehensively fine-tuning SpeciesNet on the complete local training set (90% of images, ~2000 per species). Starting from the same pruned 15-class architecture used for SNet-FT-Small, we fine-tuned higher layers and the classifier head while preserving lower-level features through freezing. We evaluated this fine-tuned model across the same gradient of training set sizes used for the local model (Section 2.4.1) to assess data efficiency. This approach maintains global feature representations while specializing for local conditions through extensive local examples. Complete implementation details are provided in Appendix S4.

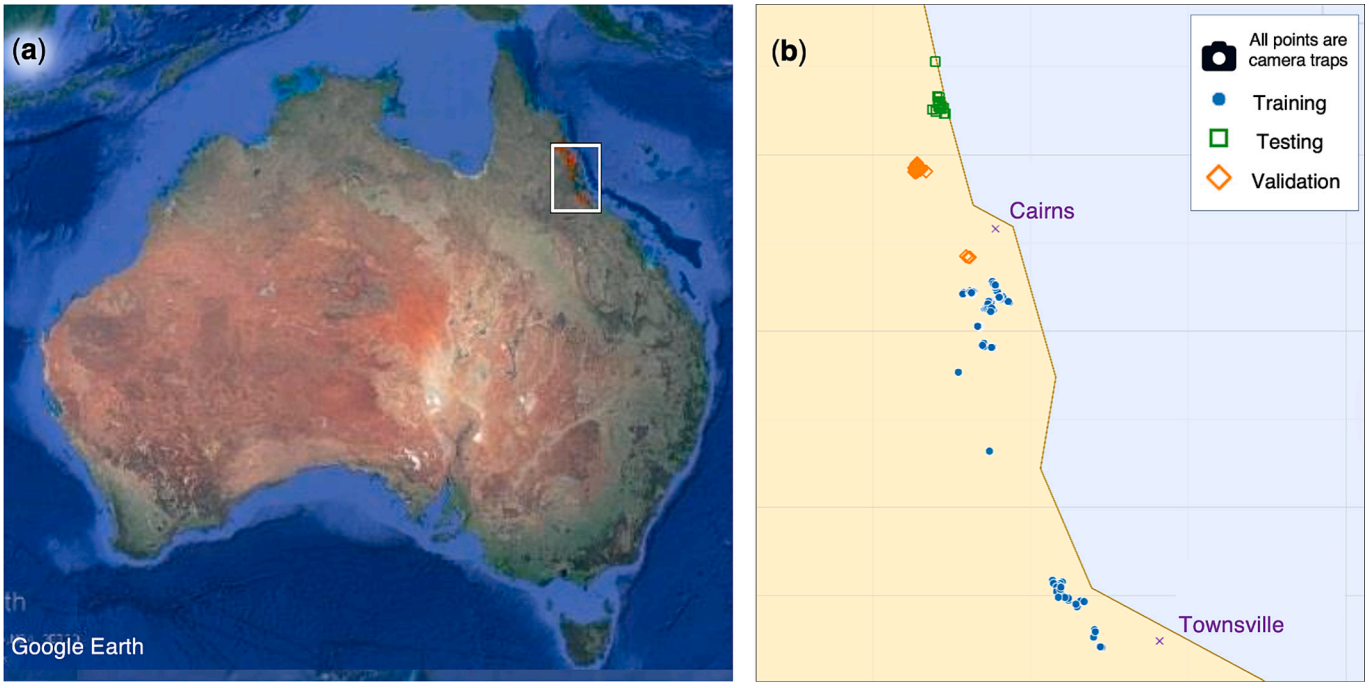


Fig. 3. Camera deployment map. Study site in the Australian wet tropics rainforests (a) and specific locations of 454 cameras that produced the wildlife image dataset (b). For random distribution (RD) training and testing (not shown), images from any camera could be used for either training or testing. For the out-of-distribution (OOD) training and testing (b), camera deployments were grouped into three areas and exclusively assigned to training, validation, or testing sets. Major cities are labelled for geographic reference (purple).

2.4.4. SpeciesNet used directly on the Wildlife Insights platform

We evaluated the unmodified global SpeciesNet model via the WI platform as a baseline out-of-the-box classifier. This approach represents the standard, most accessible workflow for researchers and practitioners in Australia. Using the standard WI pipeline (MegaDetector detection followed by SpeciesNet classification), we submitted only the held-out OOD test images with default settings. This approach provides a reproducible baseline but offers no customization for local species, as the model uses its full taxonomy of thousands of species (Table 2).

2.5. Model evaluation and comparison

2.5.1. Evaluation scenarios

We evaluated all models under both RD and OOD scenarios to assess performance in familiar versus novel conditions. RD tests model performance when deployment conditions match training data, while OOD tests generalisation to new locations with different environmental contexts. Both scenarios used the 90/5/5 train/validation/test splits described in Section 2.3 and Table S1a, enabling comparison of model robustness under ideal (RD) and challenging (OOD) conditions.

2.5.2. Performance metrics

We evaluated classification performance using precision, recall, and F1-score. The F1-score, as the harmonic mean of precision and recall, provides a balanced metric that accounts for both false positives (Type I errors) and false negatives (Type II errors). This balance is crucial for ecological applications where both misidentifications and missed detections carry significant costs. We report macro-averaged F1-scores in the main text, while complete per-class precision and recall values are provided in Supplementary Materials Tables. Accuracy was not evaluated for offline models, as the MEWC framework outputs precision, recall, and F1-score by design (Table 3).

2.5.3. Global model evaluation via Wildlife Insights

The global SpeciesNet model was evaluated using WI with default settings on the OOD test set only, since SpeciesNet was not trained on AWT imagery, making all images effectively from new regions. This provided an out-of-the-box performance baseline. All predictions were compared against ground-truth labels to compute precision, recall, and F1-scores.

2.5.4. Handling model variability and uncertainty

We accounted for variability in model training outcomes to ensure fair comparisons. The global SpeciesNet baseline has fixed pre-trained weights and produces deterministic predictions, while the fine-tuned SpeciesNet model showed minimal run-to-run variation due to stable initialization. In contrast, the locally-trained models exhibited greater variability. This occurred because they were initialized with general ImageNet weights, which provide features not optimised for wildlife recognition, leading to inconsistent convergence during training. To quantify this, we trained multiple local models with different random seeds and reported mean performance with standard deviations, providing variance-aware estimates of local model performance.

Table 2
Computer vision workflows evaluated in this study. All models used MegaDetector for animal detection and share the EfficientNetV2-M classification backbone. Base Model indicates the source of pre-trained weights (SpeciesNet or ImageNet). Deployment specifies whether the model was run through the Wildlife Insights online platform (WI SNet) or offline on local hardware. The Adaptation column describes the training strategy; numbers in parentheses denote (images per class in the random-distribution split, images per class in the out-of-distribution split) used for training or fine-tuning.

Model	Deployment	Base model	Adaptation
WI SNet	Online	SpeciesNet	None (unmodified)
SNet-FT-Full	Offline	SpeciesNet	Fine-tuned (2030, 1836 images/class)
SNet-FT-Small	Offline	SpeciesNet	Fine-tuned (120, 102 images/class)
Local models	Offline	ImageNet	Trained from scratch (2030, 1836 images/class)

Table 3
Evaluation metric definitions. Performance metrics and their ecological relevance for wildlife classification. We report the F1-score in the main text.

Metric	Focus	Penalizes	Best use case
Precision	Correctness of positives	False positives (Type 1 errors)	High cost for false alarms
Recall	Coverage of true positives	False negatives (Type 2 errors)	High cost for missing detections
F1-score	Balance of precision & recall	Both false positives and negatives	Imbalanced datasets or when both matter
Accuracy	Overall correctness	None specifically	Balanced datasets

3. Results

3.1. Random distribution (RD) performance

Under RD conditions with overlapping training and testing camera deployments, the fully fine-tuned global model (SNet-FT-Full) achieved superior performance with a macro-averaged F1-score of 0.964. This exceeded both the local model (0.937 ± 0.038) and the minimally adapted fine-tuned model (SNet-FT-Small, 0.934) (Table 4). SNet-FT-Full demonstrated balanced precision and recall (both 0.964), indicating consistent species identification with minimal false positives or missed detections. This approach exemplifies how publicly available global models can be effectively customized for regional applications using standard transfer learning techniques.

At the species level (Table S1), fine-tuning yielded large gains for several taxa, particularly endemic species like *Orthonyx spaldingii*, *Uromys caudimaculatus*, and *Wallabia bicolor*, where the fine-tuned global model markedly outperformed the local model. Conversely, the local model maintained comparable precision for highly distinctive species such as *Bos taurus* and *Casuarius casuarius*, suggesting that visually distinctive taxa can be effectively learned even by locally trained models with limited data.

Table 4
Classifier performance comparison. CV model performance for random (top) and out-of-distribution (bottom), showing mean values among the 15 species of Australian rainforest wildlife (15 classes). SNet = SpeciesNet, FT = Fine-tuning. Values in column headings indicate the number of images per class used under each split. Additional statistics are provided in Tables S3–S5.

Random distribution			
Metric	SNet-FT-Small (120 images per class)	SNet-FT-Full (2030 images per class)	Local model (2030 images per class)
Recall	0.933	0.964	0.937 ± 0.036
Precision	0.934	0.964	0.938 ± 0.047
F1-score	0.934	0.964	0.937 ± 0.038

Out-of-distribution			
Metric	SNet-FT-Small (102 images per class)	SNet-FT-Full (1836 images per class)	Local model (1836 images per class)
Recall	0.790	0.915	0.850 ± 0.199
Precision	0.784	0.925	0.849 ± 0.130
F1-score	0.764	0.909	0.839 ± 0.171

3.2. Out-of-distribution performance

In the OOD evaluation, where test images came from entirely unseen camera deployments, all models exhibited performance declines, but the magnitude of these drops differed markedly between approaches (Fig. S3). SNet-FT-Full again outperformed all other models, achieving a macro-averaged F1-score of 0.909, compared to 0.839 ± 0.171 for the local model and 0.764 for SNet-FT-Small. This represents a 7-point F1-score advantage over the local model and a 14-point advantage over the minimally fine-tuned model, underscoring the value of comprehensive fine-tuning for robust generalisation.

As an out-of-the-box global baseline, we evaluated the unmodified SpeciesNet model through the Wildlife Insights platform on the same balanced OOD test set (1530 images). This model was not pruned to the 15 target species or fine-tuned on local data, and it achieved 25.82% accuracy and a macro-F1 of 0.133 (Table 5; Fig. S4). The low scores largely reflect a label-space mismatch because predictions were assigned across the full global taxonomy, including genus-level labels, non-target taxa, and "No CV Result", rather than being restricted to the 15 focal species. We therefore present this result in a separate table (Table 5) as it represents the standard Wildlife Insights workflow and is not directly comparable to the 15-class classifiers in Table 4.

We next examined how classification performance scaled with the number of training images per species. Under both RD and OOD conditions, all models improved rapidly with the first 20 images per class and approached saturation after 250–500 images (Fig. 4, Figs. S2, S3). For instance, the fully fine-tuned SpeciesNet (SNet-FT-Full) attained >95% of its maximum macro-F1 with approximately 250 images per species, and additional training data beyond 500 images contributed only marginal gains. The local model displayed a similar pattern but converged to a lower asymptotic performance. These trends were consistent across both evaluation scenarios, indicating that a moderate number of locally labelled images is sufficient to achieve near-peak classification accuracy when fine-tuning a pre-trained global model.

The per-species precision and recall breakdowns were consistent with this scaling behaviour. SNet-FT-Full maintained high precision (0.925) and recall (0.915), while the local model exhibited greater variability (precision = 0.849 ± 0.130; recall = 0.850 ± 0.199), reflecting inconsistent performance across species. Notably, several rare or morphologically similar taxa (e.g., *Hypsiprymnodon moschatus*, *Thylogale stigmatica*) saw the greatest gains from fine-tuning, suggesting that global features enhanced recognition of difficult species in novel contexts; these species-level patterns are illustrated for representative taxa in (Fig. 5).

Table 5
Wildlife Insights baseline. Performance of the unmodified Wildlife Insights SpeciesNet baseline on the out-of-distribution test set of 1530 images (102 per species). All metrics are macro-averaged across the full unrestricted SpeciesNet or Wildlife Insights label space, not restricted to the 15 target species. The off-the-shelf model was evaluated without fine-tuning or class pruning. These values represent the standard Wildlife Insights workflow and are not directly comparable to the 15-class classifiers in Table 4, which were evaluated under a common pruned label space. The corresponding confusion matrix is provided in Fig. S4. Accuracy is reported as a percentage, while macro-precision, macro-recall, and macro-F1 are proportions on a 0–1 scale.

Metric	Value
Model	WI SpeciesNet (unmodified)
Evaluation setting	Off-the-shelf; no fine-tuning; no class pruning
Label space	Unrestricted global taxonomy
Accuracy	25.82%
Macro-precision	0.2284
Macro-recall	0.1106
Macro-F1	0.1329

4. Discussion

In this experiment, a fine-tuned global model with >2000 local training images (SNet-FT-Full) achieved the highest classification performance in both random-distribution (RD) and out-of-distribution (OOD) scenarios. The fine-tuned model's advantage was most pronounced under challenging OOD conditions: SNet-FT-Full attained an F1 of 0.909 on images from unseen camera deployments, versus 0.839 ± 0.171 for the local model. In the RD test, SNet-FT-Full reached a macro F1-score of 0.964, compared to 0.937 ± 0.038 for the local model and 0.934 for the minimally adapted global model (SNet-FT-Small). Even this 3% gain in the RD results is significant because it approaches or crosses the accuracy thresholds suggested as needing human verification (e.g., some rules-of-thumb are 95%). Notably, SNet-FT-Full excelled on species that proved difficult for the local model, underscoring its robust generalisation across taxa. The SNet-FT-Small model, trained on just 102–120 images per species, serves as a minimal-adaptation lower bound; as such, its performance represents the baseline achievable with very limited local data, whereas SNet-FT-Full captures the ceiling of the fine-tuning strategy. Further, fine-tuning SpeciesNet performance reaches very high levels with as little as 250 images. These results align with prior findings that deep learning classifiers can achieve reasonable accuracy with as few as several hundred labelled images per class, provided they represent the range of local variation in appearance, context, and camera configurations (Raghu et al., 2019; Torralba and Efros, 2011). Taken together, fine-tuning SpeciesNet consistently yielded better outcomes than training a new model.

The fine-tuned global model likely achieves its edge by effectively integrating broad learned features (e.g., key places to search for unique differences in body shapes) and local ecological features (e.g., the body shape of southern cassowaries, *Casuarius casuarius*). It also removes thousands of irrelevant classes present in the global model species list, while introducing new local species. The result was a single, streamlined classifier that generalised well to novel conditions (as seen in the OOD test) while remaining focused on the region's species. This synergy suggests that global classifier models like SpeciesNet can serve as powerful baselines that, with minimal retraining, adapt to new ecological contexts far more effectively than either approach alone. The benefits were pronounced for hard-to-classify species such as small birds (*Heteromyias cinereifrons*) and similar shaped species such as the two macroods (*Wallabia bicolor* and *Thylogale stigmatica*). In contrast, the local model remained competitive on highly distinctive species (e.g. cattle, *Bos taurus*), suggesting that visually obvious taxa can be learned adequately even with limited data. However, the local models exhibited higher variance and degraded generalisation compared to the fine-tuned model. Meanwhile, the off-the-shelf global classifier (SpeciesNet via Wildlife Insights) often misidentified species outside its original training set, resulting in a poor OOD baseline (macro-F1 = 0.1329).

A consistent source of error was the poor performance due to training images with multiple species, such Humans (*Homo sapiens*) walking domestic dogs (*Canis familiaris*). This co-occurrence produces ambiguous snips and overlapping detections that inflate off-diagonal entries in the confusion matrix (Fig. S4), with many *Homo sapiens* frames assigned as dogs and vice versa. This was especially problematic with snips that were dominated by the dog rather than the person and vice versa (Fig. 6). Our training was single-label by design, so these multi-object scenes were scored as errors, even when the model correctly recognises one of the co-occurring taxon. This pattern is consistent with context-induced confusions reported in camera classification (Beery et al., 2020), and suggests practical mitigations: (i) sequence-level aggregation, (ii) multi-label scoring for human–animal co-occurrence, and (iii) detector settings that favour separate, non-overlapping crops. The presence of dingoes (*Canis lupus dingo*; also called wild dogs *Canis familiaris dingo*) likely also added to misidentifications.

A key practical question is how many images are needed to train a reliable local model and for fine-tuning SpeciesNet. We found that

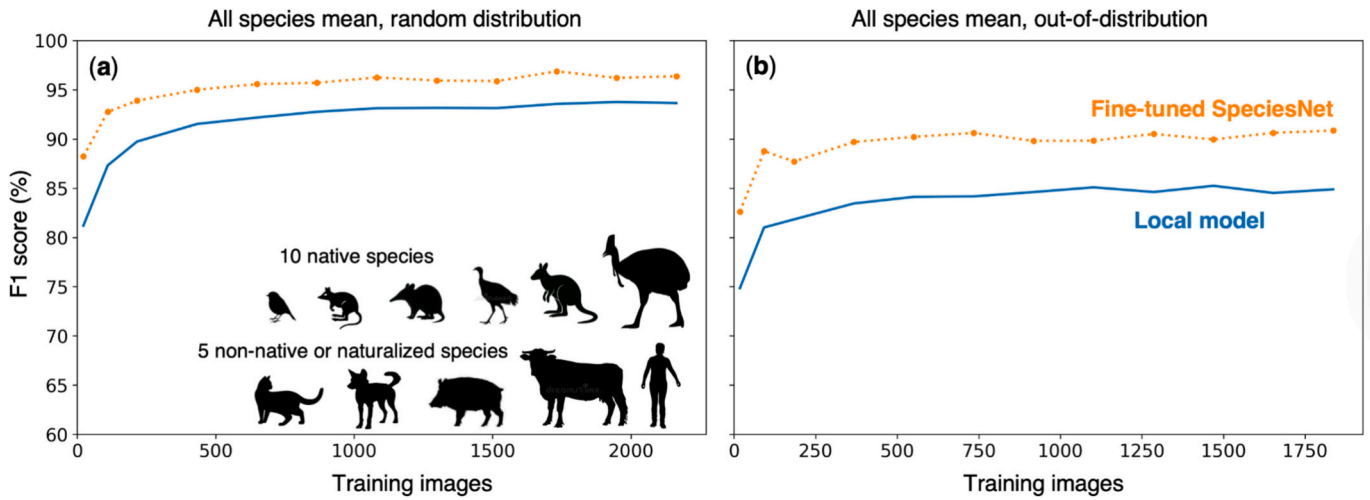


Fig. 4. Performance of local models and fine-tuned SpeciesNet classifiers as a function of the number of training images per species. (a) Random distribution (RD): test images were drawn from the same camera deployments as training. (b) Out-of-distribution (OOD): test images came from camera deployments entirely withheld from training. Values are macro-averaged F1-scores (mean across all 15 species; for local models, also across all random seeds). Fine-tuned SpeciesNet achieved higher F1-scores than local models at all training-set sizes, and the performance gap was larger under OOD conditions.

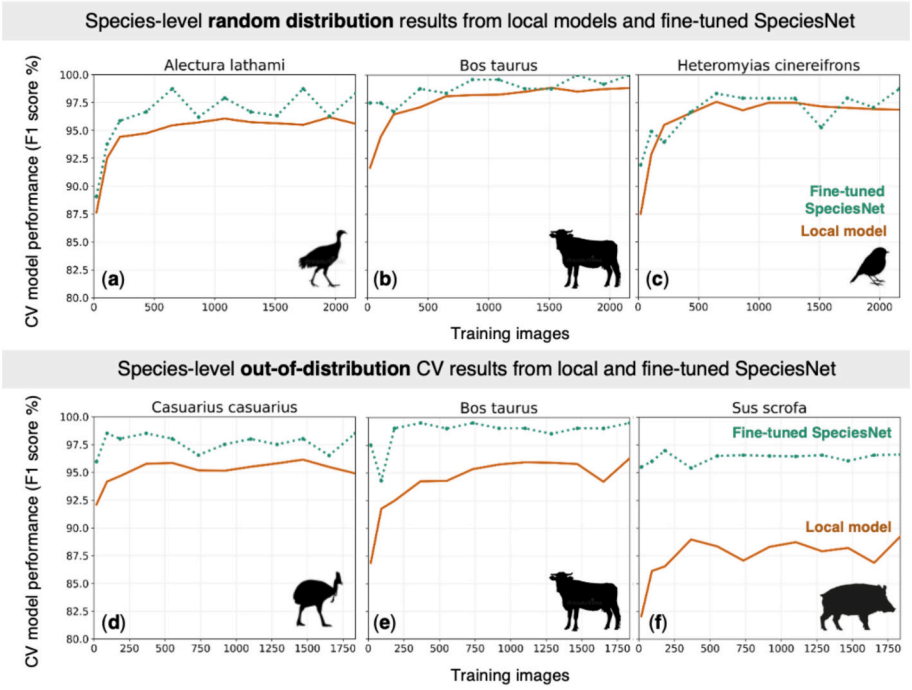


Fig. 5. Species-level performance. Per-species performance of CV models tested with random distribution (top) and out-of-distribution (bottom). Species chosen here show native and non-native species, birds and mammals, as well as a variety of body sizes (all 15 species are reported in the appendix). *Bos taurus* and *Sus scrofa* were already present in the SpeciesNet classes, while endemic *Casuarius casuarius* and *Heteromyias cinereifrons* are novel. The baseline SpeciesNet output for new species is where the green line crosses the y-axis (min value was 20 training images).

model accuracy improves rapidly with the first 20 images per class but plateaus beyond a certain threshold, often ~250 images, or even diminishing. This suggests that several hundred diverse, high-quality images per species may be sufficient (or even ideal) to approach the upper bound of local model or fine-tuning performance in this use case. Such inferences align with other deep learning models reports that achieve reasonable accuracy with as few as a few hundred labelled examples per category (Raghu et al., 2019; Torralba and Efros, 2011). This is an encouraging result – it implies that users do not necessarily need thousands of images for each species to build an effective classifier. Instead, there should be a focus on obtaining a few hundred representative images (capturing the range of individual variation (e.g. coat patterns), backgrounds and camera angles) for each target species. We note, however, that this threshold will vary with the number of species in the classifier and other forms of complexity; more images might be required for cryptic species that appear similar.

There are a number of practical considerations for using local and global models without fine-tuning. A practical downside to developing a local model trained only on ImageNet weights is that it requires higher effort in data collection and annotation, as well as AI/ML expertise, and computational resources. Local models are most suitable when new cameras are placed in the same locations as the training dataset (similar

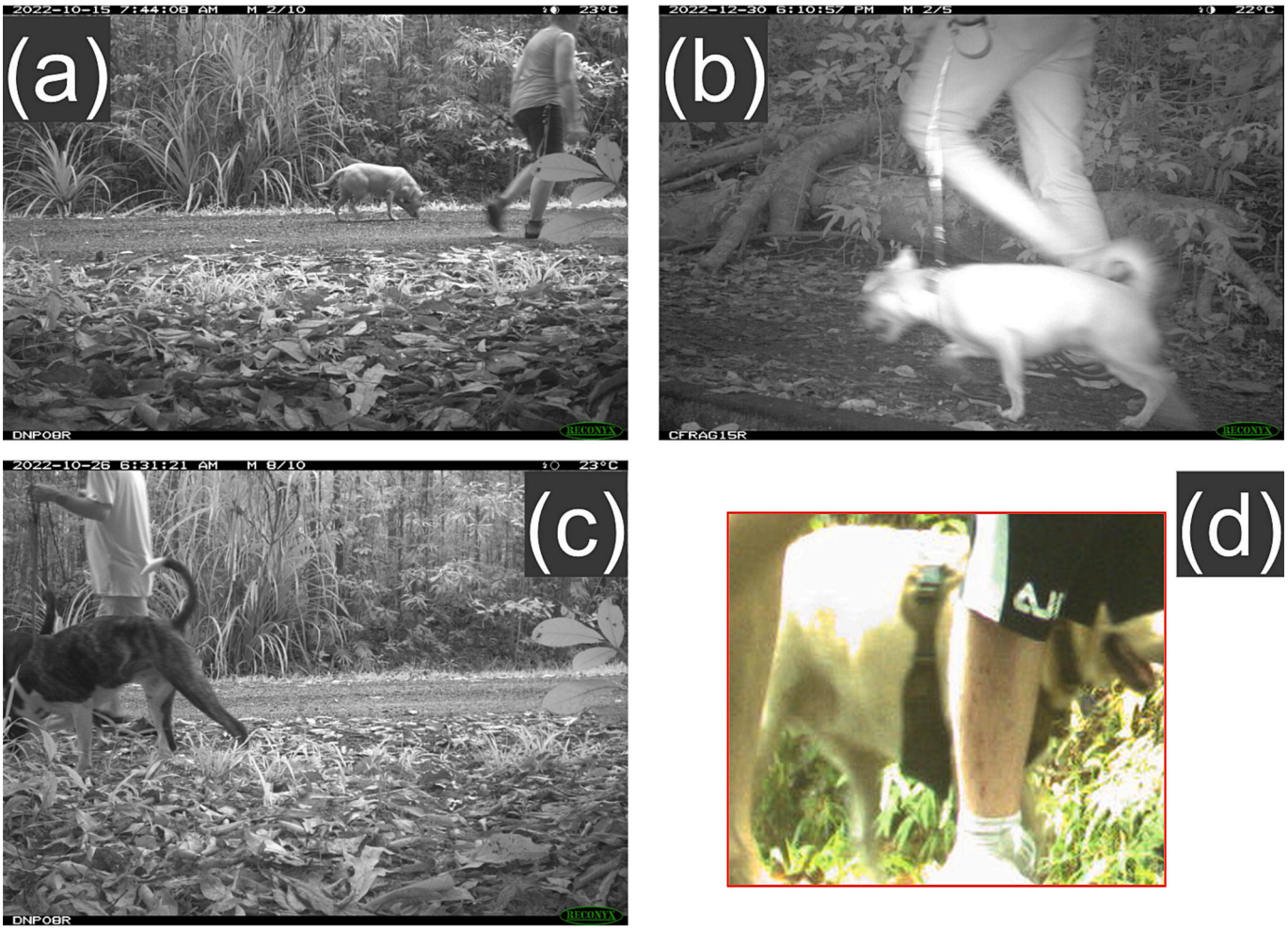


Fig. 6. Errors often resulted from training images with multiple species. This illustrates a key limitation of the single-label evaluation framework when multi-taxa occurrences are present. Panels (a–c) provide full images containing people and dogs (while obscuring human identities). Panel (d) shows a representative snip in which a human–dog misclassification occurred, where the handler's leg is visible in front of the dog.

to our random distribution results). On the contrary, their opposite practical advantages apply to using pre-trained global models like SpeciesNet, especially on online platforms with intuitive GUI like Wildlife Insights. If the focal site and study system are captured in SpeciesNet training, this is likely the best option. One downside to global models is when they function as plug-in-play black boxes with limited flexibility, such as if their use was limited to WI. For example, if they yield unexpected outputs (e.g. predicting a species that is biogeographically implausible at the site), this can undermine interpretability and confidence.

When using the downloadable SpeciesNet model offline, an alternative customisation method is to keep the full architecture but re-normalise its output probabilities (setting probabilities for excluded species to 0 and rescaling the remainder to sum to 1) across a user-defined subset of local species (van Lunteren, n.d.). This approach, however, only works effectively if all local species are already present in the global model's taxonomy. If new species are missing, users must either find a closely related proxy species from the global list or employ class pruning and weight updating as done in our fine-tuning method.

There are also practical considerations for fine-tuning a global model. An advantage is that it is computationally efficient (since the pre-trained backbone is reused) and data-efficient (requiring fewer images to reach high performance than training from scratch). Our approach of pruning the classifier head and updating weights directly integrates new species and focuses the model. Further refining the classes based on the focal species geographical distribution ensures its outputs were both accurate and context-appropriate, improving interpretability and trust. However, this approach does require some technical capability, though we attempt to mitigate this by providing open-source code and guidelines (see Appendix). Moreover, fine-tuning still relies on a moderate amount of local training data and might need periodic updates if environmental conditions or the species community change over time. In our case, fine-tuning SpeciesNet achieved excellent species recognition for an under-represented ecosystem.

5. Conclusion

Our approach accurately classified 15 commonly occurring species across the AWT, including the addition of 6 species previously missing from the global model, representing 4 endemic species. This study provides empirical guidance for ecologists using computer vision models for wildlife image classification. Fine-tuning a pre-trained global model (SpeciesNet) emerged as a highly effective and efficient strategy, essentially a hybrid approach that reconciles the trade-off between local and global approaches. The fine-tuned model not only outperformed a local model in both familiar and novel contexts, but did so with relatively fewer training images, illustrating the power of transfer learning for ecological applications. For other practitioners, this suggests that fine-tuning global models may be the optimal path forward when faced with limited training data for a new region or species of interest. Looking ahead, we envision a growing role for global and regionally pre-trained base classifier models in ecology and conservation. We have

demonstrated how a global model can be re-purposed to suit many different locales and taxa; similarly, base models trained on common taxa from a broader region (e.g., Australia) could be made available and efficiently fine-tuned for specific projects. This light-weight fine-tuning accelerates the development of accurate classifiers for biodiversity monitoring and delivering on management relevant timelines. This means that ecologists worldwide can harness state-of-the-art classifiers (trained on vast global data) and tweak them for their own projects without needing prohibitive data volumes or computing power, and we provide the code for users to do this.

Ethics and privacy statement

All wildlife camera-trap deployments, image handling and data processing were conducted in accordance with applicable institutional, site-access, data-governance and privacy requirements. The study used non-invasive camera-trap imagery and did not involve animal capture. Images containing people were processed only for the stated research purposes, no individuals were identified, and any examples shown in the manuscript were cropped, obscured or selected to protect privacy.

CRediT authorship contribution statement

Prakash Palanivelu Rajmohan: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Renuka Sharma:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation. **Zachary Amir:** Writing – review & editing, Validation, Resources, Data curation. **Tom Bruce:** Writing – review & editing, Validation, Data curation. **Barry W. Brook:** Writing – review & editing, Validation. **Dan Morris:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization. **Matthew Scott Luskin:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

Funding

This research was funded by a Wildlife Observatory of Australia (WildObs). WildObs was started with seed money from University of Queensland (UQ) Centre for Conservation Science and the School of the Environment. The project has been implemented through co-investment between UQ, the Australian Research Data Commons (ARDC) – Planet program, QCIF, and the Terrestrial Ecosystem Research Network (TERN). WildObs image platform was a collaborative project with Agouti, Wageningen University, and INBO in Europe, and we acknowledge their providing foundational support. WildObs is powered by storage and compute from the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS). WildObs has been shaped by scientists at universities in all states and territories, national and state governments, and NGOs such as Bush Heritage.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2026.181926.

Data availability

Data, code and materials availability: doi:10.15468/kjqw3f (for 391,948 curated images), https://github.com/WildObs/SpeciesNet-FineTuning (for code), wildobs.org.au (for accessing project resources and trained model information) and https://huggingface.co/WildObs/WildObs_QLD_WetTropics/tree/main (for downloading the trained model). The MEWC local model training repo is available at http://github.com/zaandahl/mewc. The full image set is at model_training_dataset_RD_QLD_Wet_Tropics.zip for Random Distribution (RD) setting and model_training_dataset_OOD_QLD_Wet_Tropics.zip for Out Of Distribution (OOD) setting.

References

Ahumada, J.A., Fegraus, E., Birch, T., Flores, N., Kays, R., O'Brien, T.G., Palmer, J., Schuttler, S., Zhao, J.Y., Jetz, W., Kinnaird, M., 2020. Wildlife insights: a platform to maximize the potential of camera trap and other passive sensor wildlife data for the planet. Environ. Conserv. 47 (1), 1–6. https://doi.org/10.1017/s0376892919000298.

Amir, Z., Sovie, A., Luskin, M.S., 2022. Inferring predator-prey interactions from camera traps: a Bayesian co-abundance modeling approach. Ecol. Evol. 12 (12), e9627. https://doi.org/10.1002/ece3.9627.

Anderson, S.E., Amir, Z., Bruce, T., Luskin, M.S., 2025. Range-wide camera trapping for the Australian cassowary reveals habitat associations with rainfall and forest quality. Ecol. Evol. 15 (6), e71464. https://doi.org/10.1002/ece3.71464.

Beery, S., van Horn, G., Perona, P., 2018. Recognition in Terra Incognita. arXiv [cs.CV]. arXiv. http://arxiv.org/abs/1807.04975.

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

Beery, S., Cole, E., Gjoka, A., 2020. The iWildCam 2020 Competition Dataset. arXiv [cs.CV]. arXiv. http://arxiv.org/abs/2004.10340.

Besson, M., Alison, J., Bjerge, K., Gorochowski, T.E., Høye, T.T., Jucker, T., Mann, H.M.R., Clements, C.F., 2022. Towards the fully automated monitoring of ecological communities. Ecol. Lett. 25 (12), 2753–2775. https://doi.org/10.1111/ele.14123.

Brook, B.W., Buettel, J.C., van Lunteren, P., Rajmohan, P.P., Aandahl, R.Z., 2025. MEWC: a user-friendly AI workflow for customised wildlife-image classification. Peer Community J. 5 (e57). https://doi.org/10.24072/pcjournal.565.

Bruce, T., Williams, S.E., Amin, R., L'Hotellier, F., Hirsch, B.T., 2022. Laying low: rugged lowland rainforest preferred by feral cats in the Australian Wet Tropics. Ecol. Evol. 12 (7), e9105. https://doi.org/10.1002/ece3.9105.

Bruce, T., Amir, Z., Allen, B.L., Alting, B.F., Amos, M., Augusteyn, J., Ballard, G.-A., Behrendorff, L.M., Bell, K., Bengsen, A.J., Bennett, A., Benshemesh, J.S., Bentley, J., Blackmore, C.J., Boscarino-Gaetano, R., Bourke, L.A., Brewster, R., Brook, B.W., Broughton, C., Buettel, J.C., Carter, A., Chiu-Werner, A., Claridge, A.W., Comer, S., Comte, S., Connolly, R.M., Cowan, M.A., Cross, S.L., Cunningham, C.X., Dalziell, A.H., Davies, H.F., Davis, J., Dawson, S.J., Di Stefano, J., Dickman, C.R., Dillon, M.L., Doherty, T.S., Driessen, M.M., Driscoll, D.A., Dundas, S.J., Eichholtzer, A.C., Elliott, T.F., Elsworth, P., Fancourt, B.A., Fardell, L.L., Faris, J., Fawcett, A., Fisher, D.O., Fleming, P.J.S., Forsyth, D.M., Garza-Garcia, A.D., Geary, W.L., Gillespie, G., Giumelli, P.J., Gracanin, A., Grantham, H.S., Greenville, A.C., Griffiths, S.R., Groffen, H., Hamilton, D.G., Harriott, L., Hayward, M.W., Heard, G., Heiniger, J., Helgen, K.M., Henderson, T.J., Hernandez-Santin, L., Herrera, C., Hirsch, B.T., Hohnen, R., Hollings, T.A., Hoskin, C.J., Hradsky, B.A., Humphrey, J.E., Jennings, P.R., Jones, M.E., Jordan, N.R., Kelly, C.L., Kennedy, M.S., Knipler, M.L., Kreplins, T.L., L'Herpiniere, K.L., Laurance, W.F., Lavery, T.H., Le Pla, M., Leahy, L., Leedman, A., Legge, S., Leitão, A.V., Letnic, M., Liddell, M.J., Lieb, Z.E., Linley, G.D., Lisle, A.T., Lohr, C.A., Maitz, N., Marshall, K.D., Mason, R.T., Matheus-Holland, D.F., McComb, L.B., McDonald, P.J., McGregor, H., McKnight, D.T., Meek, P.D., Menon, V., Michael, D.R., Mills, C.H., Miritis, V., Moore, H.A., Morgan, H.R., Murphy, B.P., Murray, A.J., Natusch, D.J.D., Neilly, H., Nevill, P., Newman, P., Newsome, T.M., Nimmo, D.G., Nordberg, E.J., O'Dwyer, T.W., O'Neill, S., Old, J.M., Oxenham, K., Pauza, M.D., Pestell, A.J.L., Pitcher, B.J., Pocknee, C.A., Possingham, H.P., Raiter, K.G., Rand, J.S., Rees, M.W., Rendall, A.R., Renwick, J., Reside, A., Rew-Duffy, M., Ritchie, E.G., Roach, C.P., Robley, A., Rog, S.M., Rout, T.M., Schlacher, T.A., Scomparin, C.R., Sitters, H., Smith, D.A., Somaweera, R., Spencer, E.E., Spindler, R.E., Stobo-Wilson, A.M., Stokeld, D., Streeting, L.M., Sutherland, D.R., Taggart, P.L., Teixeira, D., Thompson, G.G., Thompson, S.A., Thorpe, M.O., Todd, S.J., Towerton, A.L., Vernes, K., Waller, G., Wardle, G.M., Watchorn, D.J., Watson, A.W.T., Welbergen, J.A., Weston, M.A., Wijas, B.J., Williams, S.E., Woodford, L.P., Wooster, E.I.F., Znidersic, E., Luskin, M.S., 2025. Large-scale and long-term wildlife research and monitoring using camera traps: a continental synthesis. Biol. Rev. Camb. Philos. Soc. 100, 530–555. https://doi.org/10.1111/brv.13152.

Bunya, 2024. May 17. https://rcc.uq.edu.au/systems/high-performance-computing/bunya.

Celis, G., Ungar, P., Sokolov, A., Sokolova, N., Böhner, H., Liu, D., Gilg, O., Fufachev, I., Pokrovskaya, O., Ims, R.A., Zhou, W., Morris, D., Ehrich, D., 2024. A versatile, semi-automated image analysis workflow for time-lapse camera trap image classification. Eco. Inform. 81 (102578), 102578. https://doi.org/10.1016/j.ecoinf.2024.102578.

Christin, S., Hervet, É., Lecomte, N., 2019. Applications for deep learning in ecology. Methods Ecol. Evol. 10 (10), 1632–1644. https://doi.org/10.1111/2041-210x.13256.

Gadot, T., Istrate, Ştefan, Kim, H., Morris, D., Beery, S., Birch, T., Ahumada, J., 2024. To crop or not to crop: comparing whole-image and cropped classification on a large dataset of camera trap images. IET Comput. Vis. https://doi.org/10.1049/cvi2.12318.

Kingma, D.P., Ba, J., 2014. Adam: A Method for Stochastic Optimization. arXiv [cs.LG]. arXiv. http://arxiv.org/abs/1412.6980.

Murray, M.H., Fidino, M., Lehrer, E.W., Simonis, J.L., Magle, S.B., 2021. A multi-state occupancy model to non-invasively monitor visible signs of wildlife health with camera traps that accounts for image quality. J. Anim. Ecol. 90 (8), 1973–1984. https://doi.org/10.1111/1365-2656.13515.

Neave, G., Murphy, B.P., Rangers, T., Andersen, A.N., Davies, H.F., 2024. The intact and the imperilled: contrasting mammal population trajectories between two large adjacent islands. Wildl. Res. (East Melbourne, Melbourne, Vic.) 51 (8), WR24039. https://doi.org/10.1071/wr24039.

Ngugi, M.R., Neldner, V.J., Dodt, W.G., 2022. Assessing Change in Ecological Communities in K'gari (Fraser Island) Using Time Series Monitoring Datasets. Queensland Herbarium and Biodiversity Science, Department of Environment and Science. https://www.qld.gov.au/_data/assets/pdf_file/0025/286261/asssess-change-ecological-communities-kgari-fraser.pdf.

Norouzzadeh, M.S., Morris, D., Beery, S., Joshi, N., Jojic, N., Clune, J., 2021. A deep active learning system for species identification and counting in camera trap images. Methods Ecol. Evol. 12 (1), 150–161. https://doi.org/10.1111/2041-210x.13504.

Raghu, M., Zhang, C., Kleinberg, J., Bengio, S., 2019. Transfusion: understanding transfer learning for medical imaging. Adv. Neural Inf. Proces. Syst. 2019. February 13. https://proceedings.neurips.cc/paper_files/paper/2019/file/eb1e78328c46506b46a4ac4a1e378b91-Paper.pdf.

Rowcliffe, J.M., Kays, R., Kranstauber, B., Carbone, C., Jansen, P.A., 2014. Quantifying levels of animal activity using camera trap data. Methods Ecol. Evol. 5 (11), 1170–1179. https://doi.org/10.1111/2041-210x.12278.

Shahinfar, S., Meek, P., Falzon, G., 2020. How Many Images do I Need? Understanding How Sample Size Per Class Affects Deep Learning Model Performance Metrics for Balanced Designs in Autonomous Wildlife Monitoring. arXiv [cs.CV]. arXiv. https://doi.org/10.48550/arXiv.2010.08186.

Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. J. Big Data 6 (1), 60. https://doi.org/10.1186/s40537-019-0197-0.

Sollmann, R., 2018. A gentle introduction to camera-trap data analysis. Afr. J. Ecol. 56 (4), 740–749. https://doi.org/10.1111/aje.12557.

Tabak, M.A., Norouzzadeh, M.S., Wolfson, D.W., Sweeney, S.J., Vercauteren, K.C., Snow, N.P., Halseth, J.M., Di Salvo, P.A., Lewis, J.S., White, M.D., Teton, B., Beasley, J.C., Schlichting, P.E., Boughton, R.K., Wight, B., Newkirk, E.S., Ivan, J.S., Odell, E.A., Brook, R.K., Lukacs, P.M., Moeller, A.K., Mandeville, E.G., Clune, J., Miller, R.S., 2019. Machine learning to classify animal species in camera trap images: Applications in ecology. Methods Ecol. Evol. 10, 585–590. https://doi.org/10.1111/2041-210x.13120.

Tan, M., Le, Q., 2021. EfficientNetV2: smaller models and faster training. In: International Conference on Machine Learning, pp. 10096–10106. In: https://proceedings.mlr.press/v139/tan21a.html.

Torralba, A., Efros, A.A., 2011. Unbiased look at dataset bias. CVPR 2011, 1521–1528. https://doi.org/10.1109/cvpr.2011.5995347.

van Lunteren, P. (n.d.). markdown/finetune-speciesnet-simple.md at main · PetervanLunteren/AddaxAI. Github. Retrieved December 12, 2025, from https://github.com/PetervanLunteren/AddaxAI/blob/main/markdown/finetune-speciesnet-simple.md.

Veron, S., Haevermans, T., Govaerts, R., Mouchet, M., Pellens, R., 2019. Distribution and relative age of endemism across islands worldwide. Sci. Rep. 9 (1), 11693. https://doi.org/10.1038/s41598-019-47951-6.

WildObs, 2025. Wildlife Observatory of Australia. https://doi.org/10.15468/kjqw3f.

Willi, M., Pitman, R.T., Cardoso, A.W., Locke, C., Swanson, A., Boyer, A., Veldthuis, M., Fortson, L., 2019. Identifying animal species in camera trap images using deep learning and citizen science. Methods Ecol. Evol. 10 (1), 80–91. https://doi.org/10.1111/2041-210x.13099.

Yosinski, Clune, J., Bengio, Y., Lipson, A.H., 2014. How Transferable are Features in Deep Neural Networks? arXiv [cs.LG] arXiv. http://arxiv.org/abs/1411.1792.