sensors

MDPI

Systematic Review

Eyes in the Wild and Minds in the Loop: A Systematic Review of Artificial Intelligence for Camera-Trap Wildlife Monitoring

Areej Alsini

Department of Computer Science and Artificial Intelligence, College of Computing, Umm Al-Qura University, Makkah 24353, Saudi Arabia; aosini@uqu.edu.sa

Abstract

Camera traps generate imagery faster than people can review it, and artificial intelligence (AI) is increasingly used to analyse it. This systematic review asks to what extent AI applied to camera-trap imagery has moved from benchmark performance to ecologically usable monitoring. Following PRISMA 2020, Scopus was searched on 6 August 2026 for 2020–2026. Peer-reviewed English-language journal articles applying automated inference to imagery from unattended cameras observing free-ranging animals were included; captive or web-sourced imagery, non-image methods, reviews and conference papers were excluded. Of 637 records, 231 were read in full and 201 were included. Studies were appraised against six quality criteria, findings were synthesised as descriptive counts across five research questions, and performance was tabulated on each study's own metric, without meta-analysis. Output grew by about 29% per year. Detection and species classification each appear in 62% of studies, behaviour recognition in 5%, and study regions invert global biodiversity gradients. Only 62% of studies evaluate beyond their training data, a share unchanged since 2020; the measured generalisation penalty ranges from six to nearly seventy points. Only 35% report field deployment; costs appear only in deployment-oriented studies. The gap between benchmark performance and ecological practice remains the field's central challenge.

Keywords: camera traps; wildlife monitoring; deep learning; computer vision; object detection; species classification; domain generalisation; edge deployment; systematic review; PRISMA 2020

1. Introduction

1.1. Camera Traps and Camera-Trap Wildlife Monitoring

Camera traps—motion-, heat- or time-triggered cameras deployed unattended to observe free-ranging animals—are a standard instrument of terrestrial wildlife monitoring. They operate continuously without an observer, suit species that are rare, nocturnal, cryptic or wary, and produce a verifiable, archivable record; camera trapping was adopted in large part to correct the detection bias of direct observation [1,2]. Throughout this review, *camera trap* denotes the instrument, and *camera-trap wildlife monitoring* denotes the practice of deploying such instruments, alone or in networks, to estimate occupancy, abundance, density, activity or community composition.

The practice has matured over three decades. Photographic capture–recapture made camera traps a density-estimation tool for individually marked species such as tigers [3]; the random encounter model extended density estimation to species without individual markings [4]; and the methods were consolidated into a common analytical framework [1,5].

Check for updates

Academic Editor: Jochen Lang

Received: 28 September 2026
Revised: 22 October 2026
Accepted: 24 October 2026
Published: 27 October 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Comparative studies have shown that camera traps are an effective sampling tool relative to alternative survey methods [6], and coordinated networks now operate at regional to global scales [7]. Two field studies illustrate the range of effort involved: 291 trap-nights for four photographs of a Data Deficient armadillo [8] against 3741 camera-trap nights in a single steppe inventory [9].

1.2. The Image-Processing Bottleneck and the Rise of Automated Analysis

The same properties make camera traps prolific. Imagery accumulates at a rate set by wildlife, weather and vegetation rather than by the research team, and false triggers inflate it further. The conventional response—every image viewed, identified and coded by a person—does not scale, is slow relative to management decision cycles, and is hard to keep consistent across observers; even large citizen-science efforts such as Snapshot Serengeti [10] distribute this burden rather than remove it. The corpus of this review carries the same signature: of the 637 records its search identified, 82 were excluded precisely because they used camera traps with no automated inference step—ecological studies whose every image was reviewed by hand (Section 2.4).

Automated analysis of camera-trap imagery predates the period reviewed here. Early systems combined hand-crafted image features with classical classifiers [11]; convolutional neural networks were first applied to camera-trap species recognition in the mid-2010s [12,13]; and between 2018 and 2019, deep learning reached accuracies comparable to human volunteers on large benchmarks [14–16], object detectors were adapted to camera-trap data [17], and shared detection pipelines such as MegaDetector [18] began to move automated filtering into routine practice. The same period produced the first warning about transfer: models trained at one set of locations performed markedly worse at new ones [19]. These pre-2020 studies established the approaches—feature engineering, deep classification, detection-then-classification and citizen-science integration—on which the literature reviewed here builds; they are therefore summarised here as background rather than synthesised.

1.3. Previous Reviews and the Knowledge Gap

Several reviews have mapped parts of this territory. Weinstein [20] and Christin et al. [21] surveyed computer vision and deep learning across animal ecology; Schneider et al. [22] reviewed computer vision approaches to individual re-identification from camera-trap data; Tuia et al. [23] set out the opportunities and obstacles of machine learning for wildlife conservation; and Vélez et al. [24] compared the platforms that deliver camera-trap AI to practitioners. Three points of consensus emerge from this body of work: deep learning performs well on the data distribution it is trained on, automation can reduce the human labour of image review substantially, and progress depends on collaboration between ecologists and computer scientists. One finding remains in open conflict: some studies report steep losses when models meet new locations [19], whereas classifiers trained on broad, multi-site data have been reported to transfer acceptably to new locations (Section 3.4).

What none of these reviews provides is a systematic, quantified account of whether the systems now being built hold up outside the conditions that trained them, and whether they reach ecological practice. Earlier reviews were narrative rather than PRISMA-based, predate the rapid growth in output since 2020 (Section 3.1), and precede the arrival of transformer and vision–language models. In particular, no previous review has measured how often generalisation is tested, how large the penalty is when it is tested, what recovers it, or what field deployment costs in accuracy, hardware, labour and money. This is the knowledge gap that justifies the present review.

1.4. Core Question, Research Questions and Contributions

This review addresses a single core question: to what extent has artificial intelligence applied to camera-trap imagery moved from benchmark performance to ecologically usable monitoring? The title expresses its two halves. The eyes in the wild are the camera traps and the imagery they produce; the minds in the loop are the models that interpret that imagery and the ecologists who verify, adopt or reject their outputs. The literature sits at the junction of two communities—ecology and computer vision—that publish in different venues and mean different things by a result: a benchmark improvement in one, a change in what an ecologist can infer from a season of imagery in the other.

The primary objective is to map and critically appraise, in accordance with PRISMA 2020 [25], the peer-reviewed literature that applies automated inference to camera-trap imagery from 2020 onwards. Five research questions decompose the core question: RQ1 establishes the context (growth and venues), RQ2 characterises what is built (architectures and tasks), RQ3 where and on what it is built (taxonomic, geographic and ecosystem coverage), RQ4 whether it holds beyond its training data (evaluation rigour), and RQ5 whether it reaches practice (field deployment and ecological use). RQ1 is answered from all 637 identified records, because growth and venue structure are properties of the whole literature; RQ2–RQ5 are answered from the 201 included journal articles. The scope is camera-trap imagery only, as defined in Section 2.3; acoustic, telemetry, uncrewed aerial, thermal-only, footprint and seismic monitoring are excluded so that modality-specific questions about evaluation and deployment remain answerable. No taxonomic or document-type restriction is applied at search, so absences under RQ3 are properties of the literature, not of the query.

Beyond updating the descriptive map of the field, this review makes four contributions not available in earlier reviews:

1. a quantified, cross-study account of the generalisation penalty, reporting within- and beyond-training performance for 28 systems on each study's own metric, together with the measured cost of each recovery (Section 3.4);
2. a systematic ledger of field-deployment costs—accuracy lost on edge hardware, degradation in the field, energy, human hours and money—compiled from 24 deployed or deployment-oriented systems (Section 3.5);
3. evidence that the share of studies testing beyond their training data has not risen since 2020 (Section 3.4);
4. evidence that the publication base is shifting toward conference proceedings, from 30% of annual output in 2020 to 41% in 2025, with direct consequences for how future reviews must be designed (Section 3.1).

Section 2 reports the methods, Section 3 the results, Section 4 the discussion, Section 5 the limitations, and Section 6 the conclusions.

2. Materials and Methods

This review was conducted and reported in accordance with the PRISMA 2020 statement [25]. The PRISMA principles are not restated here; this section reports only how they were implemented in the present review. The workflow ran in seven stages: formulation of the research questions (Section 2.1); a structured database search (Section 2.2); application of pre-specified eligibility criteria (Section 2.3); title-and-abstract screening and full-text selection (Section 2.4); data extraction (Section 2.5); quality appraisal of the included studies (Section 2.6); and synthesis (Section 2.7). This review was not registered and no protocol was published; all definitions, reason categories and operational checks were fixed before full screening began.

2.1. Research Questions

The five research questions below decompose the core question stated in Section 1.

- RQ1. Has research on AI-assisted camera-trap monitoring grown between 2020 and 2026, and how is output distributed across venues and research communities?
- RQ2. Which model architectures and monitoring tasks characterise the field, and how has that composition shifted?
- RQ3. Which taxa, regions and ecosystems are represented, and where are the blind spots?
- RQ4. How rigorously are these systems evaluated—do studies test generalisation across sites, seasons and camera configurations, or report only within-dataset performance?
- RQ5. Do these systems reach ecological use and field deployment, and what constrains the gap between benchmark and practice?

2.2. Information Sources and Search Strategy

The electronic database Scopus was the information source. Scopus was chosen because it indexes both of the field's parent literature—the ecological and conservation journals and the computer science journals and major computer vision conference proceedings—within a single interface, and because it exports, for every record, the structured metadata (document type, source title, citation count and open-access status) on which the growth and venue analysis of RQ1 depends. The consequences of searching a single database are discussed in Section 5.

The search was built from two concept blocks (Table 1). The first block captures the instrument and is a required concept, so that every retrieved record concerns camera traps; it includes spelling variants and the term *trail camera*, which is common in North American and hunting-derived literature. The second block captures automated analysis at three levels of generality: field-level terms (deep learning, machine learning, computer vision, artificial intelligence), method-family terms (neural network, convolutional, YOLO, transformer), and generic terms (automated, automatic) that retrieve pipelines not described in machine learning vocabulary. The two blocks were combined with the Boolean operator AND, terms within each block with OR, and the truncation symbol (*) captured plural and derived forms.

The search proceeded in five steps. (1) The query was entered in the Scopus Advanced Search interface, restricted to the title, abstract and author-keyword fields (TITLE-ABS-KEY). (2) The publication-year limit (2020–2026) and the English-language limit were applied; no taxon, subject-area or document-type limits were applied, and conference proceedings were retained. (3) The search was executed on 6 August 2026 and returned 637 records. (4) All 637 records were exported with full bibliographic metadata and abstracts. (5) The export was checked for duplicate records (none were found, as expected from a single-database search) and passed to screening (Section 2.4). The complete search string, as executed, was:

```
TITLE-ABS-KEY ( ``camera trap*'' OR ``camera-trap*'' OR ``trail
camera*'' OR ``camera trapping'' ) AND TITLE-ABS-KEY ( ``deep learning''
OR ``machine learning'' OR ``computer vision'' OR ``artificial
intelligence'' OR ``neural network*'' OR ``convolutional'' OR ``YOLO''
OR ``transformer'' OR ``automated'' OR ``automatic'' ) AND PUBYEAR >
2019 AND PUBYEAR < 2027 AND ( LIMIT-TO ( LANGUAGE , ``English'' ) )
```

Table 1. Search strategy: concept blocks, keywords and limits applied in Scopus (TITLE-ABS-KEY fields).

Component	Keywords/Setting	Role
Concept 1: camera trapping (instrument)	"camera trap*" OR "camera-trap*" OR "trail camera*" OR "camera trapping"	Required concept; spelling variants and the trail-camera synonym
Concept 2a: field-level terms	"deep learning" OR "machine learning" OR "computer vision" OR "artificial intelligence"	Studies that name the discipline of the method
Concept 2b: method-family terms	"neural network*" OR "convolutional" OR "YOLO" OR "transformer"	Studies that name only the model family
Concept 2c: generic terms	"automated" OR "automatic"	Automated pipelines not described in machine learning vocabulary
Combination	Concept 1 AND (2a OR 2b OR 2c)	Every record must concern camera traps and automated analysis
Limits	PUBYEAR 2020–2026; language English	No taxon, subject-area or document-type limits
Database and date	Scopus, 6 August 2026	637 records retrieved

2.3. Eligibility Criteria

Studies were eligible where they applied an automated or computational inference step to camera-trap or trail-camera imagery. The inclusion criteria (IC1–IC6) and exclusion criteria (EC1–EC7) are set out in Table 2; each exclusion recorded in Figure 1 corresponds to one of these codes. Three operational rules govern the borderline cases. A camera trap is an unattended, fixed camera—motion-, heat-, time- or image-content-triggered – observing free-ranging animals. An automated inference step is an algorithm that decides what an image contains; image management, independence filtering and triggering hardware do not qualify, and neither does machine learning operating downstream of manually identified image content, such as species-distribution models or density estimation from manually classified records. Where a record's imagery provenance cannot be established (web-collected, museum or curated photographs presented under camera-trap vocabulary), the indeterminate-provenance rule excludes it, and full-text provenance checking is mandatory because the indexed fields cannot reveal the problem.

The start year of 2020 was chosen so that the review captures the period in which deep learning became the dominant approach and in which output grew most rapidly (Section 3.1); the foundational pre-2020 studies are summarised in Section 1 as background. Synthesis (RQ2–RQ5) is restricted to peer-reviewed journal articles, so that every synthesised study reports its methods at full length; the growth analysis of RQ1 uses all identified records. Journal indexing status and quartile were not used as eligibility criteria: every record is Scopus-indexed by construction, eligibility was decided on study content, and a quartile describes the publication venue rather than the quality of an individual study, which is appraised separately (Section 2.6).

2.4. Selection Process

Records were screened on title and abstract and coded Include, Maybe or Exclude, with a stated reason. Screening was performed by a single reviewer (the author), assisted by an AI pre-classification of every record. The reviewer verified every Include and Maybe code, and seeded 10% audits of both sides—16 eligibility exclusions and 25 verified Includes—

found no decision disagreement, against a pre-set rule that three or more errors would trigger a full re-screen. All definitions, reason categories and operational checks, including the indeterminate-provenance rule and a requirement that every retrieved document be checked against its record on title, authors and venue, were fixed before full screening and applied uniformly.

Table 2. Inclusion (IC) and exclusion (EC) criteria. The definitions in the text govern borderline cases.

Dimension	Inclusion Criteria	Exclusion Criteria
Imagery	**IC1.** Motion-, heat- or time-triggered cameras, or continuously recording cameras with automated image-content triggering, deployed unattended to observe free-ranging animals	**EC1.** Captive surveillance (laboratories, zoos, farms); web-, museum- or repository-sourced collections; provenance indeterminate after full-text reading (not camera-trap or provenance indeterminate)
Automated inference	**IC2.** Algorithmic detection, classification, localisation, counting or measurement of image content	**EC2.** Image management, independence filtering, spectrogram generation, triggering hardware; learning steps that consume coordinates or manual labels rather than images, e.g., Maxent (no automated method)
Tasks	**IC3.** Detection; species classification; individual re-identification; segmentation; tracking; counting; behaviour recognition	**EC3.** Analyses of the audio channel (acoustic monitoring) unless coupled to automated image analysis; generic computer vision with no animal application
Contribution types	**IC4.** Model development and evaluation; imagery datasets and benchmarks; augmentation and annotation methods; calibration and preprocessing; software tools and pipelines; deployment and edge computing	**EC4.** Bibliographic inventories, occurrence databases and geodata; editorials, notes and letters; reviews (review or editorial)
Subjects	**IC5.** Wild, free-ranging animals, including wild agricultural pests	**EC5.** Livestock and companion animals
Publication	**IC6.** Peer-reviewed journal article; published 2020–2026; English	**EC6.** Conference papers, book chapters, data papers and short surveys (recorded for RQ1, not synthesised; not a journal article). **EC7.** Full text unobtainable after retrieval attempts

Note: Criteria categories are shown in **bold**.

Every provisional decision from the AI-assisted pass was verified by the author against the sharpened definitions. Forty-one records initially coded Maybe were resolved (16 Include, 25 Exclude); ten provisional Includes were reversed through four targeted checks (non-image modalities, drifting imagery provenance, downstream-only machine learning and generic computer vision); and one record initially excluded as provenance-indeterminate was reinstated once its full text was obtained. No record was excluded by the automation tool alone. The verified title-and-abstract outcome, reported in Figure 1, is 255 records sought for retrieval and 382 excluded: not a journal article (EC6), 185; no automated method (EC2), 82; review or editorial (EC4), 70; not camera-trap or provenance indeterminate (EC1), 38; generic computer vision with no animal application (EC3), 4; other, 3. Of the 255 reports sought, 24 could not be obtained (EC7); all 231 retrieved reports were read in full, and 30 were excluded at that stage, leaving 201 included studies. Ambiguity was resolved conservatively throughout.

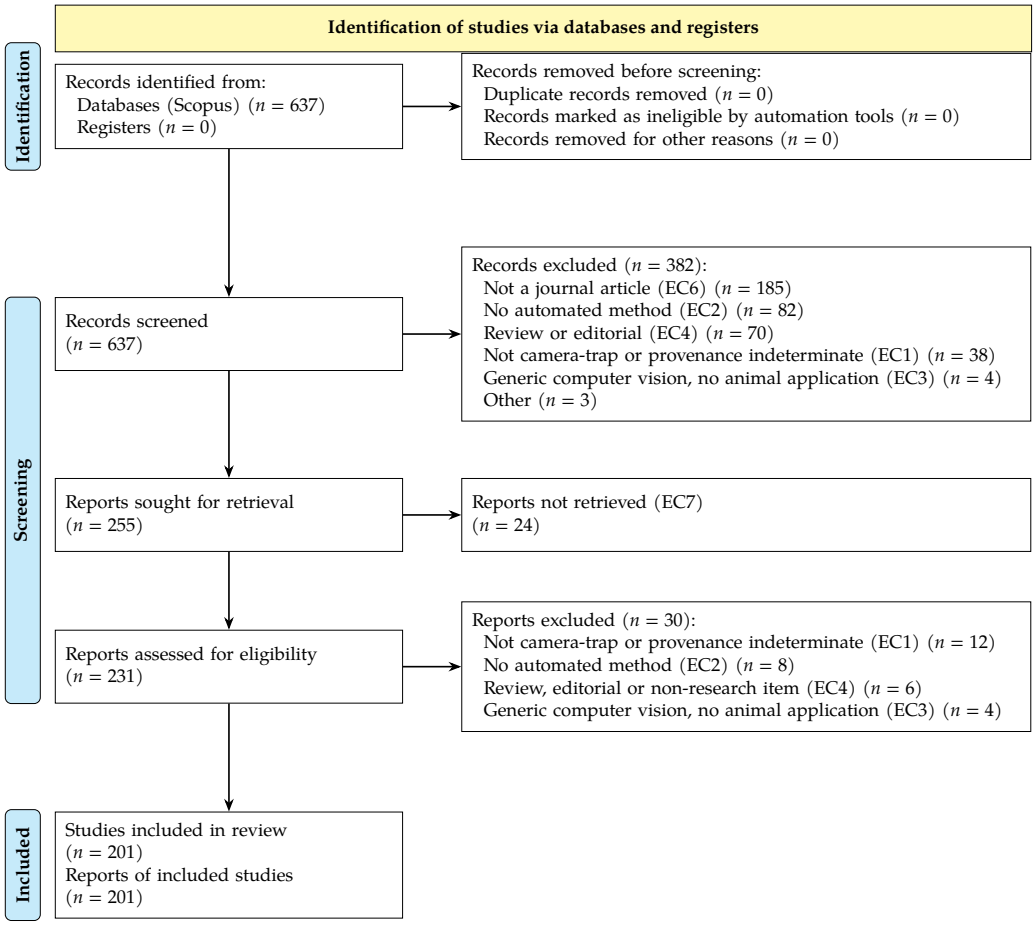


Figure 1. PRISMA 2020 flow diagram, following the official template for new systematic reviews that searched databases and registers only [25]. Exclusion reasons are mapped to the criteria codes of Table 2; AI pre-classification excluded no record without human verification (Section 2.4). The 201 included studies are the denominator for Sections 3.2–3.6.

2.5. Data Extraction

From every included study, the same fourteen items were recorded by the author (Table 3), chosen so that each research question has exactly the evidence it needs. Categorical items are recorded in a fixed vocabulary; metrics, locations and notes are recorded verbatim as free text; and *not reported* is recorded explicitly, because how often a study is silent on a field is itself a result. The generalisation test is coded as the strongest design reported, on an ordered scale from a same-site random split through temporal holdout, cross-camera and cross-site evaluation to cross-dataset transfer.

Table 3. The fourteen data items extracted from every included study.

#	Item	Content	Serves
1	Eligibility	Confirmed or reversed at full text	Selection
2	Reversal reason	Reason, where eligibility is reversed	Selection
3	Task(s)	Monitoring task(s) addressed	RQ2
4	Architecture family	CNN, YOLO, R-CNN, transformer, etc.	RQ2
5	Architecture named	Specific model(s), verbatim	RQ2
6	Taxa focus	Taxonomic group(s) studied	RQ3
7	Study location	Region of image collection, as stated	RQ3
8	Data source	Own deployment, public benchmark, or both	RQ3
9	Generalisation test	Strongest evaluation design beyond training data	RQ4
10	Headline metric	Principal performance figure, as the authors report it	RQ4
11	Deployment status	Archived imagery, released tool, prototype/field trial, operational	RQ5
12	Compute target	Workstation/server, cloud, edge/embedded, not stated	RQ5
13	Ecological use	None, feeds ecological analysis, informs management	RQ5
14	Note	Anything anomalous, free text	—

2.6. Quality Appraisal

Every included study was appraised against six quality criteria (QA1–QA6, Section 3.6, each answered Yes or No from the extraction items of Table 3. The criteria were framed around the research questions: two concern whether the study is reported fully enough to be reproduced and situated (QA1, QA2), one whether its computing requirements are stated (QA3), one whether its performance claims are tested beyond the training data (QA4), and two whether it connects to ecological use and field practice (QA5, QA6). Figure 2 summarises the procedure.

Quality appraisal was used to qualify the synthesis, not to exclude studies. Excluding studies that fail QA4 or QA6 would remove precisely the evidence that RQ4 and RQ5 are designed to measure and would make the field appear more rigorous and more deployed than it is. Instead, the appraisal determines which studies contribute to which conclusions: generalisation penalties (Section 3.4) are drawn only from studies meeting QA4, and the deployment costs of Section 3.5 are reported with an explicit statement of whether each system was field-deployed or bench-measured. Formal risk-of-bias tools designed for intervention or diagnostic-accuracy studies were not applied, because the included studies do not estimate treatment effects and the review does not pool effect estimates; the criterion-level results are reported in Section 3.6.

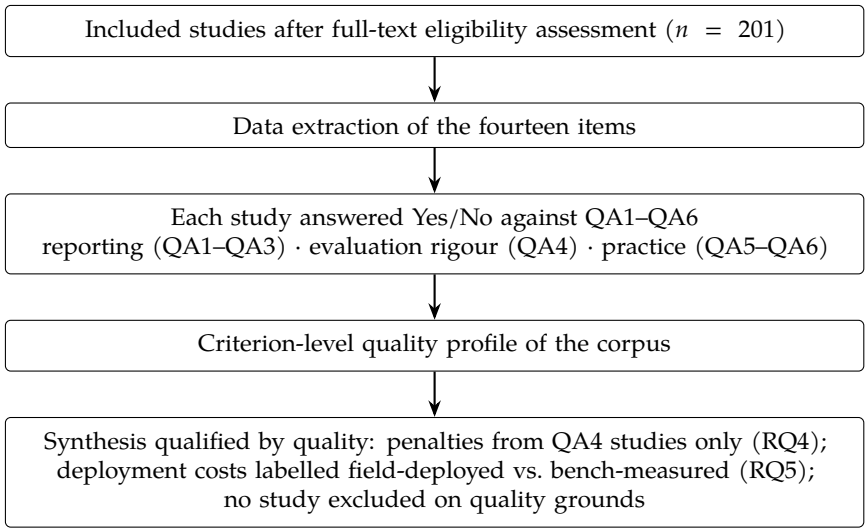


Figure 2. Quality-appraisal procedure applied to the included studies.

2.7. Synthesis Methods

Because the included studies address different tasks, on different datasets, with different performance metrics, a meta-analysis was neither feasible nor meaningful. The synthesis is therefore descriptive and narrative, organised by research question. Each research question is answered with counts and percentages over the relevant denominator (637 records for RQ1, 201 studies for RQ2–RQ5); task, architecture, taxon and region categories may overlap within a study, whereas evaluation design, deployment status, compute target, data source and ecological use are coded as single exclusive categories. Growth is summarised as the compound annual growth rate over the six complete years (2020–2025); 2026 is reported as a partial year. Where studies report performance within and beyond their training data, both values are tabulated on the study's own metric rather than converted to a common scale, and deployment evidence is tabulated in the units each study reports. The narrative synthesis then identifies the patterns, exceptions and explanations that recur across these tabulations.

3. Results

3.1. RQ1: Research Growth and Publication Landscape

RQ1 concerns the trajectory of the field and is answered from all 637 identified records, before eligibility screening; all quantities were recomputed from the corpus export. Output rose from 47 records in 2020 to 166 in 2025 (Table 4; Figure 3), a 3.5-fold increase equivalent to a compound annual growth rate of about 29%, monotonic across every complete year; the partial 2026 count of 105 records over roughly seven months annualises to 175–180, consistent with continued growth. Conference proceedings account for 200 records and their share of annual output rose from 30% in 2020 to 41% in 2025—a substantive finding: the field's centre of gravity is shifting toward the computer vision publication culture, so reviews restricted to journals will under-represent methodological work, increasingly over time. The corpus is dispersed across 340 venues with no dominant outlet: *Ecological Informatics* leads with 49 records (7.7%), ahead of *Methods in Ecology and Evolution* (26), *Ecology and Evolution* (21), *Remote Sensing in Ecology and Conservation* (18) and *Animals* (17), alongside computing venues. This is the structural signature of a field with two parent communities. Of the 637 records, 356 (56%) are open access; citations total 7973 (median 3, maximum 940), and 187 records are uncited, which is unsurprising given that 271 were published in 2025–2026.

Table 4. Records identified per year, by Scopus publication type.

Publication Type	2020	2021	2022	2023	2024	2025	2026 [1]	Total
Journal articles	29	33	49	44	71	83	70	379
Conference proceedings	14	12	11	31	43	68	21	200
Other document types	4	5	6	8	6	15	14	58
Total	47	50	66	83	120	166	105	637

[1] 1 January to 6 August 2026 only. Journal articles are Scopus document type *Article*; conference proceedings are *Conference paper* and *Conference review*; other document types comprise book chapters, reviews, data papers, editorials, books, notes, letters, short surveys and one retraction notice.

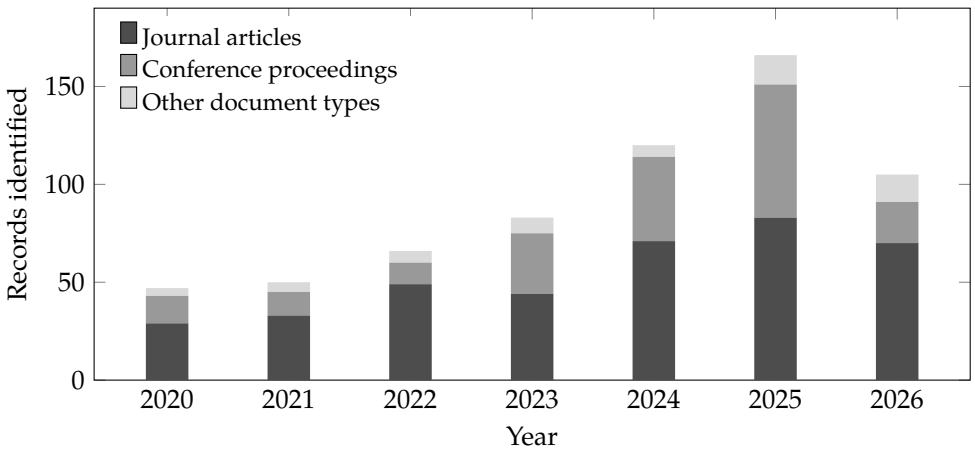


Figure 3. Annual publication output on AI-assisted camera-trap monitoring, 2020–2026. The 2026 column is a partial year, covering approximately seven months.

3.2. RQ2: Architectures and Monitoring Tasks

Every count in Sections 3.2–3.6 refers to the 201 studies included in the review [26–226], the final box of Figure 1. All counts derive from the full-text extraction fields of Section 2.5. Task, architecture, taxon and region categories overlap, because a study may address several tasks or name several regions, so those counts do not sum to 201; evaluation design, deployment status, compute target, data source and ecological use

are coded as single exclusive categories per study. RQ2 is answered along three separate dimensions: the monitoring tasks addressed (Section 3.2.1), the model architectures used (Section 3.2.2), and the tools and datasets contributed (Section 3.2.3).

3.2.1. Task Distribution Characteristics

Each study was coded to the monitoring tasks shown in Figure 4. *Detection*: locating animals in the frame, as bounding boxes or image-level presence. *Species classification*: assigning a taxon label to an image or detection. *Counting/density*: enumerating individuals or estimating abundance from imagery. *Tracking*: linking the same animal across video frames or image sequences. *Individual re-identification*: distinguishing individuals of one species. *Segmentation*: pixel-level delineation of the animal. *Empty/false-trigger filtering*: discarding blank or falsely triggered images before analysis, coded as a task in its own right, beyond the eligibility vocabulary of Section 2.3, because a substantial body of work addresses it directly. *Behaviour recognition*: labelling what the animal is doing.

Detection and species classification dominate: 125 of the 201 studies (62%) implement detection and 124 (62%) species classification (e.g., [42,93,135,178]). Empty or false-trigger filtering is the third most common task, at 75 studies (37%), usually as a preprocessing stage of a detection or classification pipeline (examples spanning filtering and segmentation: [64,132,177,201]). The secondary tasks follow: counting or density estimation appears in 32 studies (16%; e.g., [61,118,148,176,202]), individual re-identification in 24 (12%; e.g., [129,145,167,179]), segmentation in 21 (10%) and tracking in 19 (9%; e.g., [95,125]). Behaviour recognition is the rarest task, at 11 studies (5%; e.g., [38,134,207,221]), an order of magnitude below the two dominant tasks. The task profile is therefore steeply graded: the tasks that answer *what is present* are mature, whereas the tasks that answer *how many individuals* and *what they are doing*—the tasks closest to population and behavioural ecology—remain a small minority.

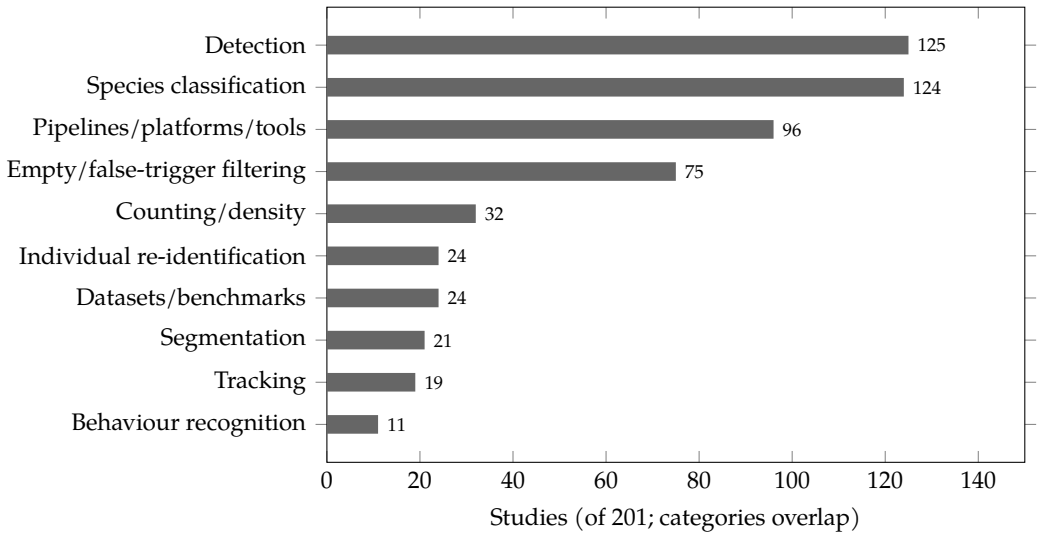


Figure 4. Monitoring tasks and contribution types documented at full text in the 201 included studies (RQ2). Categories overlap, so counts do not sum to 201. Behaviour recognition, the task closest to ecological mechanism, is the rarest task, an order of magnitude below detection and species classification.

3.2.2. Architecture Families and Reported Performance

The architecture is almost always identifiable: 192 of 201 studies (96%) name at least one model, and only 9 (4%) describe automated inference without identifying it (Table 5). Generic convolutional backbones (130 studies) lead, followed by the R-CNN family (82),

the YOLO family (64; [69,110]), transformers (34), classical machine learning or computer vision methods (32) and vision–language or foundation models (7; [115]). MegaDetector is named in 52 studies, about one in four; because it is itself a Faster R-CNN, it accounts for most of the R-CNN family's count. The families form a recognisable development pathway, from classical pipelines and generic convolutional classifiers, through two-stage (R-CNN) and one-stage (YOLO) detectors that localise animals before classifying them, to transformer backbones and, most recently, vision–language models that can be queried without task-specific training.

Table 5. Architecture families documented at full text in the 201 included studies (RQ2).

Architecture Family	*n*	Representative Members
Generic CNN backbones	130	ResNet, VGG, EfficientNet, Inception, MobileNet, DenseNet, ConvNeXt, Xception
R-CNN family	82	Faster/Mask R-CNN, including MegaDetector
YOLO family	64	YOLOv2–v10 and derivatives
MegaDetector (named)	52	De facto shared detection pipeline
Transformers	34	ViT-, Swin- and DETR-based detectors and classifiers
Classical ML/CV	32	Background subtraction, frame differencing, SIFT matching, SVM, random forest
Vision–language/foundation	7	CLIP-style, SigLIP and adapted multimodal models
No architecture named	9	4% of included studies

192 of 201 studies (96%) name at least one architecture. A study may use more than one family (for example, MegaDetector for detection and a CNN classifier for species), so rows overlap and do not sum to 192. Families were assigned by a documented keyword parse of the extraction fields *Architecture family* and *Architecture named*.

Full-text reading sharpens three of these families. The vision–language records agree that general-purpose models do not transfer to camera-trap recognition unaided: zero-shot CLIP retrieves Snapshot Serengeti attributes at 0.29 mean average precision, lifted to 0.60 by domain fine-tuning but still below a vision-only classifier's 0.68 [115], and an instruction-tuned vision–language model applied directly achieves near-perfect precision at recall as low as 0.07–0.20, so the working system delegates recognition to a YOLOv10 detector [114]. The verdict reverses for behaviour: a multimodal large language model applied zero-shot to three-way behaviour classification reaches 86.39% macro F1, at the price of prompt fragility (bare labels collapse SigLIP from 85.67% to 10.57%) [164]. Metadata fusion pays only where context is class-informative: temperature–date–location vectors lift a ResNet50 from 98.4% to 98.9%, while the same study's attention-fusion variant falls below its image-only baseline [142]; temporal fusion concentrates its gain in species with learnable activity rhythms [123]; and camera-identity co-supervision lifts cross-location macro F1 by 3.1 points on iWildCam, the gain tracking measured species–location mutual information and concentrating in rare classes [187].

Table 6 compares the performance that representative modern methods report in the included studies. Because each study uses its own data and metric, the table supports comparison of patterns rather than a ranking of models. Three patterns are consistent. First, within the data distribution on which they were trained, all modern families report high performance, between about 80% and 99% on accuracy-type metrics. Second, the difference between the within- and beyond-training performance of the same system is generally larger than the difference between families; in the one study that compared six architectures under the same cross-dataset test, the hybrid detector–transformer lost the fewest points [210]. Third, compact convolutional and YOLO models retain most of their accuracy when quantised for edge hardware [126,198], whereas vision–language models are not yet competitive for zero-shot species recognition but are already competitive for behaviour [115,164].

Table 6. Reported performance of representative modern methods in the included studies (RQ2). Values are on each study's own metric and dataset and are not directly comparable across rows.

Family	System and Task	Reported Performance
Generic CNN	Six architectures, species classification, 36 locations [34]	95.6% top-1 at trained locations; 68.7% at unseen locations
Generic CNN	ResNet50 with temperature–date–location fusion [142]	Accuracy 98.4% (image only) to 98.9% (with metadata)
Generic CNN	ConvNeXt, 88-class European mammal and bird classifier [135]	97.91% mAP (validation); 90.39% and 82.77% mAP at two held-out sites
Generic CNN (edge)	Custom 66 KB INT8 CNN, smart bait dispenser [198]	Recall 96.2%, F1 95.5%; <0.5 s inference on all four devices
Generic CNN (edge)	MobileNet v1 on Raspberry Pi, bait station [82]	Field accuracy 90%, specificity 96%; image processed in <2 s
R-CNN	MegaDetector (Faster R-CNN) zero-shot, three Bavarian regions [127]	Accuracy 96.0% (animal), 93.8% (person); daily detections vs. manual $r = 0.996$
R-CNN	Faster R-CNN, Chernobyl Exclusion Zone [92]	81.31% image-level accuracy; 50.88% event-level success at held-out cameras
YOLO	YOLOv6-N variant on Jetson Xavier NX [126]	$mAP_{0.5}$ 88.3% at 8.9 MB; 85.3% after INT8 quantisation at 6.15 ms/image
YOLO	YOLOv10x, real-time curlew alert network [159]	Live F1 95.05% and 96.03% (curlew and chick classes)
YOLO	YOLOv5 detector–classifier moved to a new habitat [181]	Accuracy 0.96 at the source location; 0.47 (MCC −0.01) at the new location
Transformer	Hybrid detector–ViT classifier, 45 species [210]	85.27% in distribution; 58.24% zero-shot cross-dataset, the smallest drop of six architectures
Vision–language	CLIP on Snapshot Serengeti attributes [115]	0.29 mAP zero-shot; 0.60 after fine-tuning; vision-only baseline 0.68
Vision–language	Multimodal LLM, zero-shot behaviour classification [164]	86.39% macro F1; SigLIP falls from 85.67% to 10.57% with bare-label prompts
Vision–language	Cross-continental transfer [215]	84.77% on African imagery; 16.17% on American imagery
Metric learning	Real-time individual re-identification [226]	Top-1 80–100% across six datasets; >10 images/s on a $200 edge device

3.2.3. Tool and Dataset Development

Two further categories code the contribution rather than the task: *datasets/benchmarks*, where the contribution is imagery data or an evaluation benchmark, and *pipelines/platforms/tools*, where it is a software system that wraps detection or classification capabilities for practical use. Almost half the corpus (96 studies, 48%) delivers a pipeline, platform or tool, whereas 24 (12%) contribute a dataset or benchmark (Figure 4). MegaDetector, named in 52 studies, is the field's de facto shared detection pipeline [116], and much tool development consists of placing a species classifier, a user interface or an ecological analysis downstream of it. A further 15 studies address the annotation bottleneck directly, through self-supervised, semi-supervised, few-shot, active or continual learning [43,83,189]. The volume of tool development contrasts with its uptake: of the 111 studies whose contribution is a dataset, benchmark, pipeline or tool, 55 never leave archived imagery (Section 3.5).

3.3. RQ3: Taxonomic, Geographic and Ecosystem Coverage

The taxonomic centre of gravity is mammalian: 111 studies (55%) work on mammal taxa and 99 (49%) on multi-species or community-level monitoring (the two overlap heavily, since most multi-species studies are mammal-dominated), against 38 (19%) on birds [141], 17 (8%) on insects or invertebrates [76], 10 (5%) on reptiles or amphibians [75,165] and 3 on fish or other aquatic taxa (Figure 5a). Full-text reading locates part of the cause in the trigger hardware: passive-infrared sensors fire on surface temperature contrast, so ectotherms can pass unrecorded, and the study that closed the gap did so by replacing the instrument—27 sites of always-on video cameras, whose million-plus frame-differenced videos, screened with YOLOv5, produced an 11,458-video dataset dominated by frogs (28.7%), spiders, lizards and scorpions [165].

Geography is stated in almost every study: 189 studies (94%) state where their imagery was collected, and only 12 (6%) present themselves as geographically unanchored methods work evaluated on unlocated or synthetic material. Where a place is stated, the distribution inverts biodiversity gradients (Figure 5b): Europe alone (75 studies) outnumbers Asia (35), Australia–Oceania (23) and Latin America (10) combined; North America adds 41 more; and the Neotropics, the most species-rich region on Earth, appear in one study in twenty. Africa (50) owes nearly half of its count (23 studies) to the Snapshot Serengeti benchmark and its descendants rather than to a broad continental literature. Independent project-level inventories report the same pattern [227].

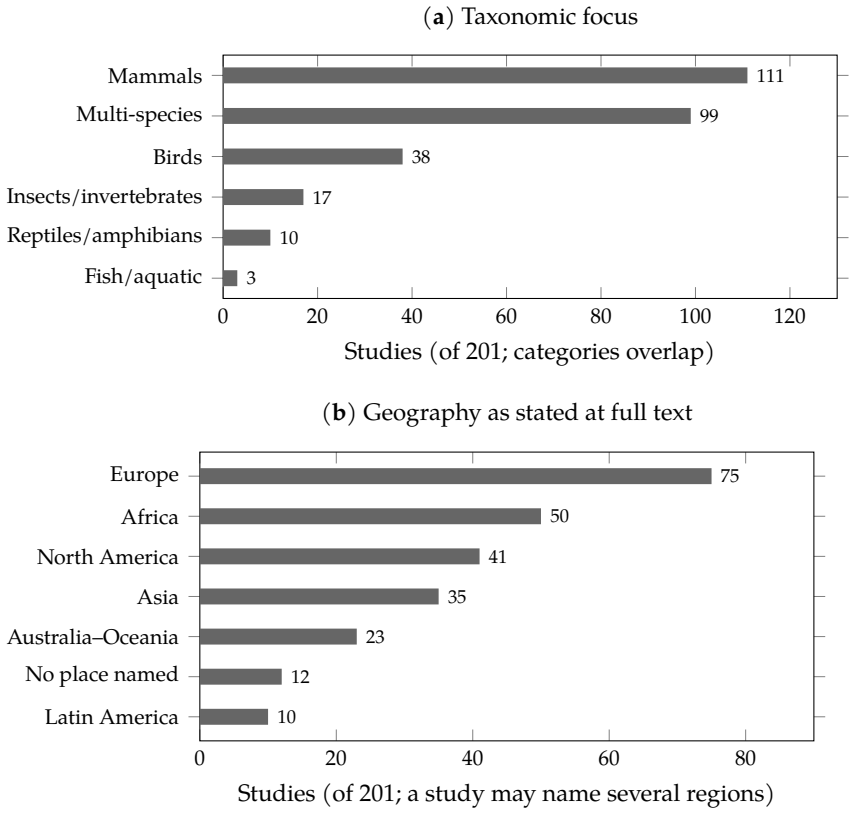


Figure 5. Coverage of the 201 included studies (RQ3). (**a**) Taxonomic focus. (**b**) Region of image collection as stated at full text; the distribution inverts global biodiversity gradients.

3.4. RQ4: Evaluation Rigour and Generalisation

For an ecologist, the rigour question is concrete: a model that scores well on the images it was trained around may still fail at the next valley, the next season, or the next camera model, and only a test on new material can tell. The corpus is far from uniform on this point.

124 of 201 studies (62%) report at least one evaluation beyond the training data: 61 hold out whole sites or locations [34,49,96,135], 39 test on a different dataset [36,215], 15 hold out a later period [42] and 9 hold out cameras within a site [92,94]. The remaining 61 (30%) evaluate only on a random split of the same material, and 16 (8%) report no quantitative evaluation of generalisation at all. The share has not moved: it lies between 51% and 80% in every year from 2020 to 2026 with no monotonic trend (Figure 6), and it is the same among field-deployed systems (42 of 70) as among archived-imagery studies (69 of 114). What follows is what the studies that did test it found; Table 7 holds the full evidence and per-study numbers.

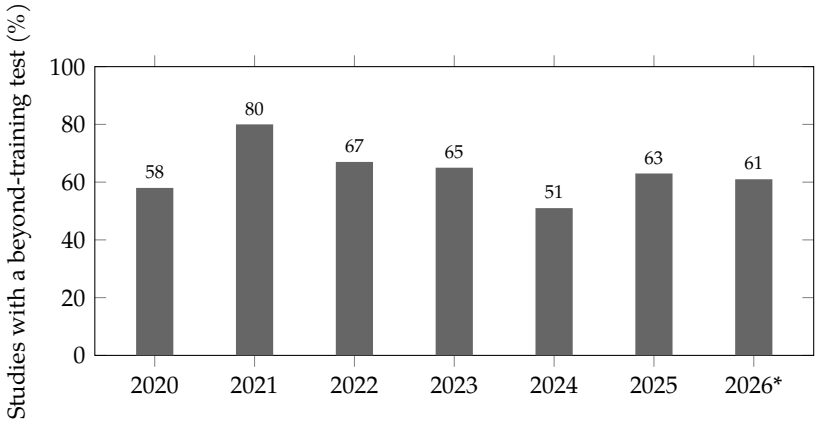


Figure 6. Share of included studies per year that report any evaluation beyond the training data (cross-site, cross-dataset, temporal or cross-camera holdout).

3.4.1. The Drop, Its Exceptions and Its Cause

Where studies test on new material, the drop is consistent and often severe. Figure 7 shows, for six comparable cases, how many points each system loses beyond its training data: unseen cameras cost a six-model comparison about a quarter of its accuracy [34], a cross-continental transfer almost everything [215], with other spatial cases in between [40,94]; time alone erodes fixed-site classifiers as the vegetation behind the animals changes [42,169]; and per-image scores flatter: one detector that was respectable per image got the whole capture event right only half the time at held-out cameras [92]. The exceptions share one property, training breadth: classifiers built on imagery from many places have passed demanding field checks at hundreds of unseen stations [49,96,135], and off-the-shelf MegaDetector matches manual detection indices in regions it was never tuned for [57,127], while at the same new site, a single-location-trained detector performs no better than chance [181]. The remaining limits lie in how images are made, not where: time-lapse animals too small for the model [59,62], failed triggers [182], taxa absent from pretraining [175]. Much of the drop has one intuitive cause: the camera never moves, so the model learns the scenery behind the animal, and recognising the trees works until it meets a new forest. One identifier jumped from 57% to 95% the moment the animal was cut out of the frame [170], others were caught matching backgrounds directly [77,99], and training with animals pasted onto varied backgrounds measurably shifts attention onto the animal [104,138,174,184].

3.4.2. Fixes, Honest Measurement and the Ecology Downstream

The recoveries are cheap in labels and reported with their price tags: adapting a broad existing model with a few hundred local images per species beats both the untouched global model and a locally built one [212]; scheduled or drift-triggered retraining restores decayed classifiers and cuts manual review severalfold [42,80]; and very small doses of targeted effort—five hundred local boxes, fifty annotations, an untested confidence

threshold finally tuned, an on-camera filter that re-calibrates itself in minutes—buy large recoveries [79,90,143,177]. Each fix carries a stated cost, from multiplied false alarms to an overfitting boundary [180], and specialised levers exist for the data floor, hard-to-see animals and newly arriving species [70,72,103]. Reported accuracy, in turn, means little until one knows what was held out: honest designs keep whole cameras or locations out of training [85,108,121], whereas leaked splits let a detector score 96% on paper and 26% in the field [208]; some cross-dataset claims test the recipe rather than the model, or internet photographs rather than deployments [52,59,83,146,162]; training at the survey's own empty-image ratio beats the field's default rebalancing [51]; confidence scores need calibrating [163]; and headline accuracy hides the rare classes ecologists most care about [85,108]. Whether any of this matters ecologically depends on the estimator, not on automation: occupancy and density models absorb label error rates up to about 20%, matching expert-label results across millions of images [36,96,205,218]—until errors arrive faster than the pooling window can soak up, at which point the same machinery biases occupancy by up to +166% while a semi-automated design preserves both the inference and 97.7% of the labour saving [124,148,183,199,202]. The spread in Table 7—from a six-point drop to a near seventy-point collapse, with recoveries priced in each currency—is why a single headline accuracy is not an adequate summary of any system in this corpus.

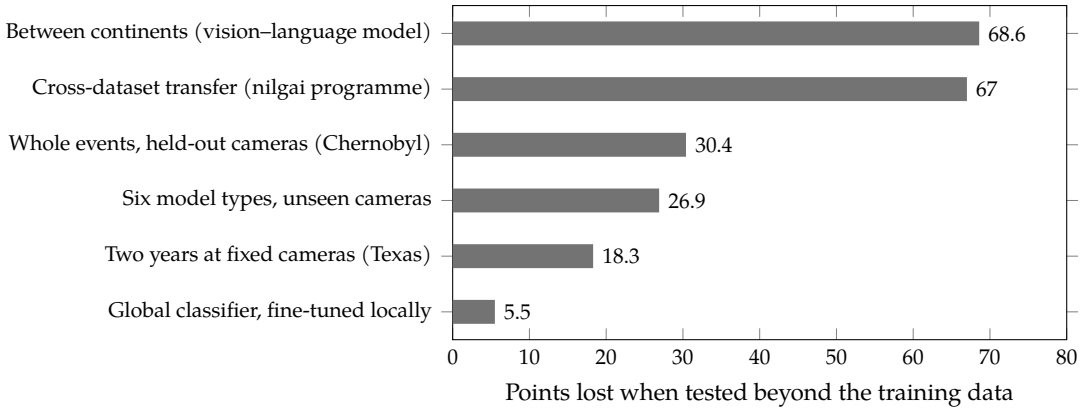


Figure 7. The generalisation penalty at a glance: how many points each of six systems loses when tested beyond its training data, on that system's own 0–100 metric [34,40,42,92,212,215]. The loss ranges from about six points (a global classifier fine-tuned on local data) to nearly seventy (cross-continental transfer of a vision–language model); the within- and beyond-training values for every case are given in Table 7. The values demonstrate the magnitude of the performance drop only: because the systems are evaluated with different metrics (accuracy, F1, mAP, MCC), absolute performance cannot be compared across systems.

Table 7. The generalisation penalty as reported by included studies that measure it (RQ4), with within- and beyond-training performance on each study's own metric.

Study Design	Within-Training Performance	Beyond-Training Performance
Six architectures, 47,279 images, 36 locations [34]	95.6% top-1 accuracy (trained locations)	68.7% (unseen locations)
Classification across habitat disturbance within one survey [94]	Overall F1 0.77 with detector-cropped images (0.47 whole images)	Mean F1 0.32 (new cameras, same disturbance); 0.27 (unseen disturbance); cropping recovers unseen-level F1 to 0.73
Faster R-CNN in the Chernobyl Exclusion Zone, twelve held-out cameras from an earlier deployment period [92]	81.31% image-level accuracy (train–test cameras, confidence 0.9)	50.88% event-level success under the strictest criterion (correct species and count)

Table 7. *Cont.*

Study Design	Within-Training Performance	Beyond-Training Performance
Vision–language model, cross-continental transfer [215]	84.77% (African imagery)	16.17% (American imagery); text-description fusion recovers 50.06%/39.96% (Cis/Trans)
Twenty-group classifier in a nilgai-management programme, cross-dataset test [40]	Average MCC 0.89 (south Texas deployment)	Average MCC 0.22 (CalTech dataset, southwestern USA)
Binary animal classifier at fixed Texas watering-fountain sites, 2017 model re-tested on 2019 imagery [42]	Day-model 87% sensitivity/96% specificity (2017)	81.4%/77.7% on 2019 imagery from the same sites; drift-triggered retraining restores 100%/100%
26-species Central African classifier [49]	—	Machine-label richness, activity, and occupancy (3 of 4 focal species; chimpanzee uncorrelated) matched expert labels at 227 independent stations
Fully automated global classifier (MegaDetector + SpeciesNet, no human review) vs. expert labels, three regions, 1227 cameras, ~3.8 M images [218]	Label true-positive rates >80% for most species	Mean occupancy-rate difference 0.03–0.07 and covariate direction agreement 83–94% across 16–19 shared species per region; models resilient to false-positive rates up to ~20%, divergence concentrated in rare and morphologically similar species
MLWIC2 base and regionally trained classifiers feeding occupancy models, three desert species, 388,108 images [124]	— (base model pretrained out of source; regional further training worth ~9–11 points of classification success)	Image accuracy 0.821–0.956 but false-positive rates 10.2–87.9% when fully automated; occupancy bias −33 to +166%, most-supported model structures and covariate directions not recovered; semi-automated empty-culling (97.7% labour saved) matches double review
Global classifier fine-tuned on 15 Australian Wet Tropics species, held-out camera-deployment areas [212]	Macro-F1 0.964 (same-deployment test)	0.909 out-of-distribution, against 0.839 (±0.171 across seeds) for a from-scratch local model and 0.133 for the unmodified global service under its full, non-comparable label space
Bat-hibernacula pipeline (detector, segmentation, classifier ensemble), six untrained sites [90]	99.3% classification accuracy at trained sites (F1 0.97–1.00, 13 species)	96.7–98.2% where camera placement matches training (<3 m, <45°); 17.8–90.8% at atypical placements (detector F1 0.38 at the worst site); retraining the detector with 500 site-specific annotations recovers 95.5–99.9% at all six
Hybrid detector–ViT classifier trained on a 45-species Ohio dataset, zero-shot on Caltech Camera Traps [210]	85.27% (45-class in-distribution test)	58.24% zero-shot cross-dataset—the smallest drop of six architectures tested (others fall 28.8–33.4 points)
YOLOv5 detector–classifier trained at one temperate-forest location, applied to the same smart-camera design in semi-arid heathland [181]	0.96 accuracy (source-location study, same model and hardware)	0.47 accuracy, MCC −0.01 (no better than random); 0% bird precision and recall; the globally trained MegaDetector at the same sites: 0.99 precision/0.98 recall
Pretrained MegaDetector applied zero-shot to an unseen park (159,272 images, 36 cameras) [57]	— (no training at the site)	Human images 99% precision/95% recall; site-week human-event index $r = 0.99$, mean difference 0.45% against manual classification
Pretrained MegaDetector (v4.1) applied zero-shot across three Bavarian regions (352,426 images, 159 cameras) [127]	— (no training in any region)	Accuracy 96.0% animal/93.8% person/99.3% vehicle at 95% confidence; daily detections vs. manual $r = 0.996$; diurnal activity overlap 0.41 vs. 0.42 manual; per-site person accuracy 52–97.9% in one region

Table 7. *Cont.*

Study Design	Within-Training Performance	Beyond-Training Performance
Pretrained MegaDetector applied zero-shot in Arctic tundra, split by trigger type (6224 images, 20 sites) [62]	— (no training at the site)	Motion-triggered images 94.6–95.7% correctly labelled; time-lapse images 55.7–61.6%; detection floor 60 px and maximum distance 222 m for the model, against 4 px and 2551 m for human review
Pretrained MegaDetector applied zero-shot to freshwater-lake time-lapse imagery of birds (142,801 images, two lakes) [182]	— (no training at the site; motion triggering unusable on lakes)	Recall 54.6% at 2.5% precision (0.1 confidence); recall 8.9% at 0.8; authors reject the automated workflow and screen all images manually
Pretrained MegaDetector applied zero-shot to time-lapse snake imagery, Everglades (~1.47M images, 42 cameras) [175]	— (no snake-specific training; a few thousand internet snake crops in pretraining)	94.5% of blanks removed; snake image-level recall 17.1% at precision 0.006 (manual recall 96.2%); large (>66 cm) species matched manual review at the detection level, smaller snakes largely missed; fine-tuning on study imagery raised false positives an order of magnitude
26-class multi-partner European classifier, out-of-sample sites [96]	97.3% balanced validation accuracy	93.6% image-level/96.7% sequence-level on 213,059 images from unseen sites; but squirrel recall 0.67, ibex precision 0.45, wolf precision 0.69
88-class European mammal+bird classifier (ConvNeXt), two held-out forest test sites in Germany and Poland [135]	97.91% mAP/93.85 mean F1 (validation split)	90.39% mAP at the German site but 82.77% at the Polish site (mean F1 82.43 vs. 62.03), the larger drop attributed to poorer image quality and lower camera angles; taxonomic output heads recover 94.23%/79.92% mAP across all levels as the stated fallback for uncertain species calls
Embedded blank filter, Snapshot-Serengeti-trained, one unseen station [143]	F1 0.78 across multiple park locations	F1 0.23 (precision 0.13) at the unseen station; on-device continual learning over location-specific synthetic data restores 0.48 overall, 0.81 post-calibration
Movement-mask ensemble sequence filter, active-learning-trained on two cameras, tested at unseen sites on three continents [177]	83–86% sequence accuracy at the two training cameras (12 rounds, ~50 annotations/class/round)	77–96% at unseen cameras (species-blind by design); a wind-exposed unseen camera collapses to 42% and one site-specific active-learning round restores 95%
Arctic small-mammal box-camera classifier (ResNet50, trained from scratch), transferred to two new study areas and a modified trap type [80]	98.2% validation accuracy (original network)	82.1% on the out-of-sample areas and 90.7% on the new trap type; retraining on 3276 of >400,000 new images restores 94.8% and 97.5%, and lifts the share of labels above 0.9 confidence from 73.7 to 94.9%
Temperate-trained video classifier applied zero-shot to tropical island ungulates (Guam) [153]	— (no training on the region; roe deer/boar training species morphologically similar)	Accuracy 78.2–84.2%, F1 $\geq$ 0.90 for deer and pig; the tropical-background African model manages only 39.3–67.4% on the same videos, so target similarity outweighs background similarity
Insect-trap detector fine-tuned on a published DIY-trap dataset, re-evaluated on the authors' own 27-plot grassland deployment [208]	96.1% mAP@0.5, 96.3% precision, 93% recall (source test set)	25.8% mAP@0.5, 41.6% precision, 29.1% recall on leakage-controlled field imagery; attributed to community shift, vegetation ingrowth and the removal of near-duplicate frames between splits
Ultra-lightweight MCU insect triggers (94k–411k parameters), field datasets unseen in training [201]	AUROC 0.918–0.964 on the withheld test set	AUROC 0.818–0.872 on one unseen field dataset but 0.588–0.691 on the other, cluttered-background one—out-of-domain success limited to simple scenes

Table 7. *Cont.*

Study Design	Within-Training Performance	Beyond-Training Performance
Bear pose estimation + MLP bipedalism classifier, test set exclusively from seven new 2023 camera stations [197]	— (no station shared with training)	96.1% bipedalism-vs-quadrupedy accuracy on predicted keypoints (97.4% on ground-truth keypoints); residual errors confined to posture-transition and sitting edge cases
GBIF-trained hierarchical insect classifier (1.12 million web-scraped images, 1127 taxa; missing-label-aware loss), applied without further training to the authors' own light-trap camera deployment in Costa Rica and Ecuador [203]	Balanced accuracy 98.6%/94.6%/85.7%/84.3% order through species (GBIF validation split)	Field balanced accuracy 98.3%/96.2%/93.3%/79.2% with recall 97.0%/92.7%/86.8%/58.5% order through species; species precision 29.2% over all predictions vs. 84.4% on confirmed taxa; genus precision approaches 100% only above 500 × 500 px crops, and ~21% of field crops fall below the consistently identifiable size—the web-to-field drop concentrates at fine taxonomic levels and small subjects

3.5. RQ5: Field Deployment and the Benchmark–Practice Gap

The last question is whether all of this leaves the laboratory. For the most part, it does not: 114 of 201 studies (57%) evaluate on archived imagery only, 17 (8%) release a software tool without a field trial, and 70 (35%) report a prototype or field trial (33) or an operational deployment (37). Computing on or near the camera is described in 41 studies (20%; [67,73]), cloud inference in 19, a workstation or server in 112, and 29 do not state their compute target at all. Of the 111 studies whose contribution is a dataset, benchmark, pipeline or tool (the union of the two RQ2 contribution categories, overlapping in nine), 55 never leave archived imagery—the benchmark–practice gap in one number. Connections to ecological end-use are stronger than deployment: 117 studies (58%) link their outputs to an ecological analysis (91; occupancy, density, abundance or activity) or to a management or conservation decision (26; [90,159,214,218]), while 84 (42%) stop at the classification metric. Most studies work on their own imagery (116, 58%) or combine it with a public benchmark (47); only 37 (18%) rely on public benchmarks alone. Human-in-the-loop and citizen-science workflows [84,133,165,204] are the shape in which automation most often reaches practice. What the deployed minority reports, assembled in Table 8, differs in kind from the benchmark literature: degradation, resource cost and human hours appear as first-class results—while even a study whose stated purpose is edge deployment can end up evaluating only on archived imagery [63].

3.5.1. What Running in the Field Actually Costs

Three costs recur in the deployed studies, and none appear in benchmark papers. First, *squeezing a model onto camera-side hardware costs accuracy, and the loss lands unevenly*: one classifier fell from 96% accuracy to a 0.70 F1 average after platform optimisation, with rare classes hit hardest [73], shrinking a detector cost precisely its smallest species [126], and a sub-300-milliwatt insect trigger held up on simple scenes but not complex ones [201]. Second, *the field itself degrades systems*: underwater housings fouled within a month [53], a 98-day insect tracker had to discard 87% of its raw detections as stationary false alarms [54], and the honest hardware ledgers record seven of thirty-six custom traps never working [181] and a charging fault silently consuming 154 of 814 satellite alerts [101]. Third, *in operation, error rates stop being metrics and become minutes, money and misdirected baits*: an Iridium-linked camera puts elephant alerts in Gabonese ecoguards' hands at a median of 7.35 min for about 1200 euros plus a monthly fee [101]; a curlew network raises real-time alerts whose stated purpose is intervention speed [159]; a reserve trial couples every classification to a

per-detection payment, turning false positives into a financial quantity [86]; an automated bait station decides in under two seconds whether to offer bait to an invasive canid, its threshold deliberately set low because the two kinds of error cost differently [82]; and filtering can move onto the camera itself, cutting detector runs by 84% using the sensor's own motion signal [64].

3.5.2. When Automation Pays, and Where People Stay in the Loop

Where systems run as infrastructure, the ledger is dominated by connectivity rather than by the model: a 65-camera cellular network costs more to build but 43% less over ten years than traditional trapping [122]; a feral-cat programme found the classifier alone saving 10% of costs but classifier-plus-connectivity saving 80% [140]; and an island biosecurity mesh tolerates poor precision on one species to guarantee a rodent alarm, repaying its start-up cost in twenty months [155]. At workflow scale the savings are just as concrete—five-to-eightfold processing speedups [57], twenty-three work-days of distance measurement done in under two [118], annotation of night video collapsing from over an hour per clip to under a minute [46], four-fifths of a reserve's labelling bill automated under a deliberately cautious discard-only-on-unanimity rule [93]—but with a consistent caveat: what decides whether assistance helps is the *structure* of the errors, not headline accuracy, because a missed animal costs a human four times what a correct label saves [84], and supervised first-pass workflows cut error fourfold while halving time [120]. Camera-plus-model combinations are even beginning to replace the survey instrument itself: baited cameras out-detected live traps for all six focal rodents at a third of the effort—with the live-trap estimates the biased ones [156]—and density pipelines now run from borrowed models [195] with the distance step handled training-free to twice the previous accuracy [217].

The recurring human quantity is the hours a model's confidence structure buys back, and the deployed studies treat the confidence threshold as an operational dial rather than a technical constant: the Arctic workflow auto-accepts labels above 0.9 confidence and retrains partly to push more images over that line [80]; the European classifier ships its threshold as a user-facing setting because partners disagree about which error matters [96]; a regional tool states feasibility as a laptop clearing a hundred videos in an afternoon [171]. Assistance does not erase expertise—experts recover abundance almost exactly where assisted novices still under-estimate [102,186]—and the tooling spectrum runs from a simple R pipeline at a third of manual cost [133] to a training-free filter whose false alarms are almost entirely livestock [132]. Privacy is handled inside the same workflows, from blurring before review [57] to removing human sequences at the honestly stated price of a third of true sequences missed [177] to transmitting labels rather than pictures [101]. The operational literature also shows the mirror image of the benchmark–practice gap—management programmes consumed as evidence that report no model accuracy at all [117,157,160]—and, where hardware and models both disappoint, the fallback is priced too: fifty hours of manual classification in one network [181], and 200,000 lake-bird images screened by hand [182].

Table 8. Reported deployment evidence in the included studies that provide it (RQ5), including deployment-oriented systems that were bench-measured rather than field-deployed. These papers report degradation and resource cost as first-class results; benchmark papers report neither.

System	Reported Evidence
Edge classification, three platforms [73]	96.1% accuracy before optimisation; macro F1 falling to 0.7 after, minority classes to 0.18; latency and current draw measured
Real-time re-identification [226]	40–60 cropped images/s on a laptop; >10/s on a $200 Cortex-A72 edge device, CPU-only; Top-1 80–100% across six datasets; 50% occlusion costs 9 points of Top-1 on average

Table 8. *Cont.*

System	Reported Evidence
Raspberry Pi Zero classification [128]	On-device classification and explainability at weighted F1 ≈ 0.95 (macro 0.87; one minority bird class at 0.16); 6 s per inference, 23.26 s per Grad-CAM visualisation; 50 mA (0.25 W) after HDMI shutdown on a solar power bank, 72-h prototype run
Cellular real-time bird classification [81]	Two years, eight feeder sites, UK and Netherlands; 42,973 images over 3/4G to server-side Faster-RCNN; 96.71% accuracy, 98.28% blank removal, sensitivity 88.79%; precision 75.32% average, spanning 35.0% (jay) to 100% (small, distant species; the stated mechanism: validation mAP 0.79 large vs. 0.46 small objects); untrained species mapped to the nearest class
Edge-optimised detector [162]	Parameter, FLOP and model-size reductions for deployment—yet evaluated only on archived imagery; the intelligent camera trap is listed as future work
INT8 wildlife detector on Jetson Xavier NX [126]	Reparameterised YOLOv6-N variant, $mAP_{0.5}$ 88.3% at 8.9 MB; 6.15 ms/image at INT8 keeping $mAP_{0.5}$ 85.3% under reparameterisation-aware training vs. 84.2% post-training quantisation; head pruning costs the smallest species (chevrotain AP 0.978 → 0.945); evaluated on archived imagery, the in-field system stated as the goal
24-h ectotherm video network, citizen-science-hosted [165]	27 always-on video traps over 7 months; $>10^6$ videos; YOLOv5 screening saves an estimated 1300 h of labelling; <5 cameras/site for 95% detection confidence of most modelled taxa
Real-time insect tracking on a 5-W edge computer [54]	98-day unattended season at 0.33 fps; 87% of raw detections removed as stationary false positives; 89% track precision, one class at 0.39
27-trap solar insect network, three grassland regions [208]	2654 trap-days over one season; on-device YOLOv5n gated by optical-flow motion detection (3.15 → 2.70 W); battery thresholds throttle recording windows; failures, overcast and vandalism gaps absorbed by two-week aggregation, each interval retaining ≥21 of 30 mowing events
Underwater camera-trap network, $125/unit [53]	5-month unattended marine deployments; biofouling degrades image quality and model accuracy within 1–2 months; analysis restricted to first-month predictions
MegaDetector-assisted recreation-ecology workflow [57]	506% overall/840% manual-component speedup on 159,272 images; 40% of volume fully automatic; privacy blurring before human review
Raspberry Pi video camera trap, in-capture-loop filtering [64]	PIR trigger replaced by hardware H.264 motion vectors (<0.4 ms/frame vs. 1.2 s YOLOv4-tiny inference); 84.1% fewer model runs; 2.20–6.25 W; multi-day field deployments
Edge classification, three platforms [73]	96.1% accuracy before optimisation; macro F1 falling to 0.7 after, minority classes to 0.18; latency and current draw measured
Real-time 3/4G reserve-wide trial with per-detection payments [86]	27 cameras, 400 km^2, 10 months; 96.38% sensitivity/87.14% precision over 12 taxa; GBP 185.20 auto-transferred over 18,520 detections; lion and cheetah precision ≈54/59% as the scheme's weak point
Edge bait station actuating pest control [82]	MobileNet v1 on Raspberry Pi; image processed <2 s, bait offered in 9.81 s; field accuracy 90%, specificity 96% at a deliberately low 5% threshold; 19 baits offered, 7 taken by target canids
Autonomous 4G camera network, 65 cameras, Natura 2000 site [122]	Solar-powered, no battery/SD visits (pilots autonomous 746 days); end-to-end pipeline to AI species ID; establishment +131% but annual costs −51%, 10-year total −43% (>500,000 euros) vs. traditional trapping; focal-species recall 0.42–0.50 keeps a human in the loop
Solar-autonomous insect camera trap, ~700 euros [139]	Five traps, 1919 h over a full season at ~4.4 W; on-device detection and tracking stores only cropped detections (5–10 KB vs. 0.5–1 MB per frame); curated-test top-1 0.972 falls to 0.2–0.49 for heterogeneous classes on 97,671 real-world images

Table 8. *Cont.*

System	Reported Evidence
Self-calibrating blank-filter camera trap, ~$100 [143]	On-device transfer + continual learning in 8–15 min at <0.6% of battery; F1 0.48 at an unseen station vs. 0.23 without calibration and 0.43 for zero-shot MegaDetector
Operational island-biosecurity mesh network, AI triage [155]	10 radio-linked cameras, 17 months; rodent alarm channel at 0.998 precision/0.934 recall reviewed within 24 h; downtime 4% of camera-days vs. 10% for the SD-card fleet it replaced, whose cards waited 56.5 days on average for review; annual maintenance $896 vs $7842, break-even at 20 months
Edge-evaluated insect pipeline, two full seasons [154]	12 UV traps, >5.7 million images, >3000 nights; Jetson Nano processes at 26.3 J/image (RPi4 148.9 J), near-real-time at ~0.5 fps; 2-min time-lapse reproduces tracked counts, enabling on-trap processing; three edge units uploaded abundance statistics for three months
Edge-evaluated metric-learning false-trigger filter [190]	0.076 s/image on a Jetson Nano at 80.17% accuracy/19.48% FPR—36.7% faster than the best YOLO comparator at comparable recall, trained from class labels only (no bounding boxes); archived imagery, communication-equipped traps the stated target
MCU insect trigger on a ~£20 ESP32-S3 board [201]	Int8-quantised models with no significant accuracy loss; <300 mW at maximum throughput, ~185 mW at half; lightest models exceed 8 frames/s, the most accurate 2; zero trigger-to-capture latency by retrospective design—but bench-measured only, and idle draw cannot match a PIR sensor's ≪1 mW
Smart bait-dispenser edge benchmark, four devices × six quantised CNNs [198]	66 KB INT8 custom CNN at recall 96.2%/F1 95.5%, beating four fine-tuned backbones; all devices <0.5 s mean inference; Portenta H7 selected at 0.185–0.19 W idle/inference, with an ATtiny85 motion-wake switch (6–30 μW sleep) proposed for multiday vaccination campaigns—bench-measured, field trials future work
Custom smart video traps for small fauna, 36 units, semi-arid heathland [181]	Downward-facing continuous video with on-board foreground-detection capture, solar-powered; 7 of 36 units never worked and 22 stopped after 66 video segments—failures first-class; only these traps detected nocturnal ectotherms and invertebrates (frogs, small reptiles, spiders, scorpions: zero on the paired commercial cameras)
Real-time 3/4G curlew alert network, 11 nesting sites [159]	Six-week live deployment (May–June 2024), 1072 images; solar cellular traps → SMTP → GPU-server YOLOv10x inference; alerts above 0.30 confidence for curlew and chick classes; live F1 95.05%/96.03% with 100% curlew precision; sub-threshold images discarded as blanks (98.28% claimed); residual errors concentrated in curlew–pheasant confusion, camera placement flagged as critical for chicks

3.6. Quality Appraisal of the Included Studies

Table 9 reports the criterion-level results of the quality appraisal described in Section 2.6. The six criteria (QA1–QA6) were designed for this review, because no validated quality-appraisal instrument exists for studies that develop and evaluate AI systems for ecological imagery; each is derived from an extraction item (Table 3), and together they are intended to characterise the reporting and validation profile of the corpus rather than to score or rank individual studies. The corpus is well reported but weakly validated. The reporting criteria are met almost universally: 96% of studies name their model (QA1) and 94% state where their imagery was collected (QA2), although one study in seven does not state its compute target (QA3). The criteria that concern evidence beyond the benchmark are met far less often: 62% of studies test performance beyond their training data (QA4), 58% connect their outputs to an ecological analysis or management decision (QA5), and only 35% validate their system in the field or in operation (QA6). The appraisal therefore locates the field's weakness in validation rather than in reporting, which is the same conclusion that RQ4 and RQ5 reach from the substantive evidence.

Table 9. Quality-appraisal criteria and results for the 201 included studies. Each criterion is answered Yes or No from the extraction item shown.

Code	Criterion (Extraction Item, Table 3)	Yes, *n*	Yes, %
QA1	Is at least one model architecture named? (item 5)	192	96
QA2	Is the region of image collection stated? (item 7)	189	94
QA3	Is the compute target stated? (item 12)	172	86
QA4	Is performance evaluated beyond the training data (cross-site, cross-dataset, temporal or cross-camera)? (item 9)	124	62
QA5	Are outputs linked to an ecological analysis or management decision? (item 13)	117	58
QA6	Is the system validated in a field trial or operational deployment? (item 11)	70	35

4. Discussion

4.1. Principal Findings by Research Question

The review set out to establish to what extent AI applied to camera-trap imagery has moved from benchmark performance to ecologically usable monitoring. The answers to the five research questions are as follows.

- RQ1 (growth and venues). The field has grown 3.5-fold between 2020 and 2025, about 29% per year, and the growth continued into 2026. Output is dispersed across 340 venues with no dominant outlet, and the share published as conference proceedings rose from 30% to 41%.
- RQ2 (architectures and tasks). Detection and species classification each appear in 62% of studies, whereas behaviour recognition appears in 5%. Generic convolutional backbones and two-stage detectors dominate, MegaDetector serves as the shared detection pipeline in one study in four, and transformers (34 studies) and vision–language models (7) are now present at the margins. Almost half of the studies deliver a pipeline or tool.
- RQ3 (coverage). The literature is centred on mammals (55%), with reptiles, amphibians and aquatic taxa rarely studied, in part because passive-infrared triggers miss ectotherms. Study regions invert global biodiversity gradients: Europe (75 studies) outnumbers Asia, Australia–Oceania and Latin America combined, and Latin America appears in 10 studies.
- RQ4 (evaluation rigour). 62% of studies test performance beyond their training data, a share that has stayed between 51% and 80% every year without a rising trend. Where the penalty is measured, it ranges from about six to nearly seventy points, is driven largely by models learning the static background, and can usually be recovered with modest local adaptation.
- RQ5 (deployment and ecological use). 57% of studies never leave archived imagery and only 35% report a field trial or operational system, yet 58% connect their outputs to an ecological analysis or management decision. Only deployed or deployment-oriented studies report accuracy loss on edge hardware, degradation in the field, and the energy, labour and money a system costs; human-in-the-loop workflows are the form in which automation most often reaches practice.

Read together, these answers reveal four asymmetries:

- Task concentration. Detection and species classification each appear in about three-fifths of studies, while behaviour recognition, the task closest to ecological mechanism, appears in one in twenty (11 studies; Figure 4).
- Geographic inversion. Almost every full text names its study region (94%), and the map that results is the inverse of the biodiversity map: Europe alone outnumbers

Asia, Australia–Oceania and Latin America combined, and Latin America appears in 10 studies (Figure 5).

- Uneven rigour. 62% of studies run some evaluation beyond their training data, but 30% evaluate only on a random split of the same material; the share has not improved since 2020 (Figure 6), and the penalty, when it is measured, ranges from six to nearly seventy points (Table 7).
- Minority deployment. 57% of studies never leave archived imagery and only 35% report a field trial or operational system, yet 58% connect their outputs to an ecological analysis or management decision (Table 8).

4.2. Comparison with Previous Reviews

The four asymmetries are not all new, and it is important to separate what this review confirms from what it adds. The predominance of species identification and counting over behavioural tasks is already visible in earlier surveys of computer vision in ecology [20]; geographic imbalance in camera-trap research is documented independently by project-level inventories [227]; and the difficulty of transferring models to new locations was demonstrated before the review period [19] and identified as a central obstacle for conservation applications [23]. On these points, the present review confirms earlier observations on a larger, systematically assembled corpus of 201 studies and turns qualitative statements into measured proportions.

The four contributions set out in Section 1.4 go beyond those observations, and the evidence qualifies the first three of them. The generalisation penalty documented in Table 7 is large but conditional—on the image-generation regime, the taxon, the camera geometry and the evaluation unit—and its recoveries are priced alongside it, so that a few hundred local images, scheduled retraining or a re-tuned threshold typically restore most of the lost performance (Section 3.4). That the share of studies testing generalisation has not risen since 2020 shows that awareness of the transfer problem has not yet changed evaluation practice. The deployment ledger (Table 8) shows that connectivity and maintenance, rather than the model, dominate the cost of a deployed system, and that edge optimisation costs accuracy unevenly, at the expense of rare classes.

4.3. Implications for Practice and Research

For practitioners, the evidence supports four recommendations: start from a model trained on imagery from many locations rather than from a model trained at a single site; budget for a few hundred locally annotated images to adapt it; judge it on the ecological estimate the project needs, not on a headline accuracy; and keep a human in the loop for rare and look-alike species, where errors concentrate. Where systems are to run as infrastructure, connectivity and maintenance, rather than the model, should dominate the budget.

For researchers, five practices would most change what the next review can conclude: state the architecture and the study location in the abstract as well as in the Methods, so that they are recoverable from indexed fields; make one honest generalisation test standard—a held-out camera reported beside the within-training figure—which is now cheap to pass, because a few hundred local images typically repair a failure; evaluate at the ecological unit, per event or estimate rather than per image; adopt the deployed minority's habit of reporting degradation, hardware, hours and money as first-class results; and aim new effort at the blind spots—behaviour recognition, ectothermic taxa and the biodiverse regions that the coverage map inverts—where a competent system is worth most.

5. Limitations

This review searched a single database (Scopus; Web of Science and other databases were not searched), in English only. This was a scope decision as much as a limitation, since the search already yields 637 records, full-text reading of all 231 retrievable reports and a 201-study synthesis, at the upper end of what a review that reads every retrievable full text can carry. Its cost is that Scopus's coverage of computer vision proceedings is incomplete, and the rising conference share (Section 3.1) means that this gap grows exactly where the field is moving. The search string has two blind spots—imagery described only in modality terms, and methodological papers phrased in learning-paradigm language or bare architecture names—and any such loss is selective, removing precisely the recent methodological work that RQ2 and RQ4 characterise. The string errs safely on the other side: out-of-scope pipelines that match both blocks inflate the screening burden and the *no automated method* count in Figure 1 but do not bias what is included.

Screening was single-reviewer with AI pre-classification; the seeded 10% audits of both sides found no disagreement, but audits sample rather than replicate, and dual independent screening would be the stronger design. Data extraction and quality appraisal were likewise performed by a single reviewer. One error class is invisible at abstract level: camera-trap vocabulary over imagery of other provenance, of which the 637 records contain a web-collected example [228]. Twenty-four reports could not be obtained and are absent from the synthesis.

The RQ2–RQ5 distributions rest on a documented keyword parse of free-text extraction fields into categories, with the small categories hand-checked. The parse is reproducible but not infallible: task and architecture categories overlap by design, region counts credit a study to every region it names, and the single-category fields (evaluation design, deployment status, compute target, data source and ecological use) required hand rulings for about one record in nine. The quality appraisal uses criteria derived from the extraction items rather than a validated risk-of-bias instrument because no such instrument exists for studies that develop and evaluate AI systems for ecological imagery without estimating an effect.

6. Conclusions

This systematic review of 201 studies published between 2020 and 2026 asked to what extent AI applied to camera-trap imagery has moved from benchmark performance to ecologically usable monitoring. Its answer is that the field has largely solved recognition within the data it is trained on but has only partly reached ecologically usable monitoring. Seven conclusions follow.

1. The field is growing quickly and changing venue. Output grew by about 29% per year, and the rising share of conference proceedings means that future reviews must search computer science sources as well as ecological ones.
2. Within-dataset performance is no longer the bottleneck. Modern convolutional, detector, transformer and edge-optimised models report roughly 80–99% on their own test data (Table 6).
3. Generalisation is the bottleneck, and its testing has not improved. Performance falls by six to nearly seventy points beyond the training data, yet only 62% of studies test for it, a share unchanged since 2020. A held-out-camera test should become a reporting standard.
4. The penalty is largely recoverable at low cost. Broad pretraining followed by adaptation with a few hundred local images, or scheduled retraining, restores most of the lost performance.

5. Ecological adequacy is decided at the estimator. Occupancy and density estimates tolerate moderate label error, but high false-positive rates can bias them severely; semi-automated workflows retain most of the labour saving while protecting the inference.
6. Deployment remains a minority, and its costs are real. Only 35% of studies reach the field; for these, connectivity and maintenance dominate cost, and edge optimisation degrades rare classes first.
7. Coverage gaps define the research agenda. Behaviour recognition, ectothermic taxa and the biodiverse tropics, Latin America in particular, are where new systems would add most ecological value.

Funding: This research work was funded by Umm Al-Qura University, Saudi Arabia, under grant number: 26UQU4300160GSSR01.

Institutional Review Board Statement: Not applicable. This study is a systematic review of published literature and did not involve humans or animals.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors extend their appreciation to Umm Al-Qura University, Saudi Arabia for funding this research work through grant number: 26UQU4300160GSSR01. A generative AI assistant (Claude, Anthropic) was used to produce provisional title-and-abstract screening decisions against the pre-specified eligibility criteria and to assist with grammar, spelling, vocabulary, and editorial polishing of this manuscript. All screening decisions were verified by the author, and all analyses, interpretations, and conclusions were produced by the author, who takes full responsibility for the content of the publication.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Burton, A.C.; Neilson, E.; Moreira, D.; Ladle, A.; Steenweg, R.; Fisher, J.T.; Bayne, E.; Boutin, S. Wildlife camera trapping: A review and recommendations for linking surveys to ecological processes. *J. Appl. Ecol.* **2015**, *52*, 675–685. [CrossRef]
2. Linkie, M.; Guillera-Arroita, G.; Smith, J.; Ario, A.; Bertagnolio, G.; Cheong, F.; Clements, G.R.; Dinata, Y.; Duangchantrasiri, S.; Fredriksson, G.; et al. Cryptic mammals caught on camera: Assessing the utility of range-wide camera trap data for conserving the endangered Asian tapir. *Biol. Conserv.* **2013**, *162*, 107–115. [CrossRef]
3. Karanth, K.U.; Nichols, J.D. Estimation of tiger densities in India using photographic captures and recaptures. *Ecology* **1998**, *79*, 2852–2862. [CrossRef]
4. Rowcliffe, J.M.; Field, J.; Turvey, S.T.; Carbone, C. Estimating animal density using camera traps without the need for individual recognition. *J. Appl. Ecol.* **2008**, *45*, 1228–1236. [CrossRef]
5. O'Connell, A.F.; Nichols, J.D.; Karanth, K.U. (Eds.) *Camera Traps in Animal Ecology: Methods and Analyses*; Springer: Tokyo, Japan, 2011. [CrossRef]
6. Wearn, O.R.; Glover-Kapfer, P. Snap happy: Camera traps are an effective sampling tool when compared with alternative methods. *R. Soc. Open Sci.* **2019**, *6*, 181748. [CrossRef] [PubMed]
7. Steenweg, R.; Hebblewhite, M.; Kays, R.; Ahumada, J.; Fisher, J.T.; Burton, C.; Townsend, S.E.; Carbone, C.; Rowcliffe, J.M.; Whittington, J.; et al. Scaling-up camera traps: Monitoring the planet's biodiversity with networks of remote sensors. *Front. Ecol. Environ.* **2017**, *15*, 26–34. [CrossRef]
8. Marín, P.; Mora, J.M.; López, L.I.; Pérez Arrieta, J.A.; Rodríguez, M.A.; Vega Cambronero, A.; Arias, I. New records on the distribution and habitat of the northern naked-tailed armadillo, *Cabassous centralis* (Mammalia, Cingulata, Chlamyphoridae) in Costa Rica. *Neotrop. Biol. Conserv.* **2021**, *16*, 451–460. [CrossRef]
9. Augugliaro, C.; Havmøller, R.W.; Monti, I.E.; Havmøller, L.W.; Janchivlamdan, C.; Lkhagvasuren, B. Non-volant mammal inventory of western Mongolian-Manchurian Grassland Ecoregion: A biogeographic crossroad worth preserving. *Check List* **2020**, *16*, 287–301. [CrossRef]
10. 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. *Sci. Data* **2015**, *2*, 150026. [CrossRef] [PubMed]

11. Yu, X.; Wang, J.; Kays, R.; Jansen, P.A.; Wang, T.; Huang, T. Automated identification of animal species in camera trap images. *EURASIP J. Image Video Process.* **2013**, *2013*, 52. [CrossRef]
12. Chen, G.; Han, T.X.; He, Z.; Kays, R.; Forrester, T.D. Deep convolutional neural network based species recognition for wild animal monitoring. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 858–862. [CrossRef]
13. Gomez Villa, A.; Salazar, A.; Vargas, F. Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks. *Ecol. Inform.* **2017**, *41*, 24–32. [CrossRef]
14. Norouzzadeh, M.S.; Nguyen, A.; Kosmala, M.; Swanson, A.; Palmer, M.S.; Packer, C.; Clune, J. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E5716–E5725. [CrossRef] [PubMed]
15. Willi, M.; Pitman, R.T.; Cardoso, A.W.; Locke, C.; Swanson, A.; Boyer, A.; Veldthuis, M.; Fortson, L. Identifying animal species in camera trap images using deep learning and citizen science. *Methods Ecol. Evol.* **2019**, *10*, 80–91. [CrossRef]
16. 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.; et al. Machine learning to classify animal species in camera trap images: Applications in ecology. *Methods Ecol. Evol.* **2019**, *10*, 585–590. [CrossRef]
17. Schneider, S.; Taylor, G.W.; Kremer, S.C. Deep learning object detection methods for ecological camera trap data. In Proceedings of the 2018 15th Conference on Computer and Robot Vision (CRV), Toronto, ON, Canada, 9–11 May 2018; pp. 321–328. [CrossRef]
18. Beery, S.; Morris, D.; Yang, S. Efficient pipeline for camera trap image review. *arXiv* **2019**, arXiv:1907.06772.
19. Beery, S.; Van Horn, G.; Perona, P. Recognition in Terra Incognita. In *Proceedings of the Computer Vision—ECCV 2018*; Lecture Notes in Computer Science; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer: Cham, Switzerland, 2018; Volume 11220, pp. 472–489. [CrossRef]
20. Weinstein, B.G. A computer vision for animal ecology. *J. Anim. Ecol.* **2018**, *87*, 533–545. [CrossRef] [PubMed]
21. Christin, S.; Hervet, É.; Lecomte, N. Applications for deep learning in ecology. *Methods Ecol. Evol.* **2019**, *10*, 1632–1644. [CrossRef]
22. Schneider, S.; Taylor, G.W.; Linquist, S.; Kremer, S.C. Past, present and future approaches using computer vision for animal re-identification from camera trap data. *Methods Ecol. Evol.* **2019**, *10*, 461–470. [CrossRef]
23. Tuia, D.; Kellenberger, B.; Beery, S.; Costelloe, B.R.; Zuffi, S.; Risse, B.; Mathis, A.; Mathis, M.W.; van Langevelde, F.; Burghardt, T.; et al. Perspectives in machine learning for wildlife conservation. *Nat. Commun.* **2022**, *13*, 792. [CrossRef] [PubMed]
24. Vélez, J.; McShea, W.J.; Shamon, H.; Castiblanco-Camacho, P.J.; Tabak, M.A.; Chalmers, C.; Fergus, P.; Fieberg, J. An evaluation of platforms for processing camera-trap data using artificial intelligence. *Methods Ecol. Evol.* **2023**, *14*, 459–477. [CrossRef]
25. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *PLoS Med.* **2021**, *18*, e1003583. [PubMed]
26. Ahmed, A.; Yousif, H.; Kays, R.; He, Z. Animal species classification using deep neural networks with noise labels. *Ecol. Inform.* **2020**, *57*, 101063. [CrossRef]
27. Ahumada, J.A.; Fegraus, E.; Birch, T.; Flores, N.; Kays, R.; O'Brien, T.G.; Palmer, J.; Schuttler, S.; Zhao, J.Y.; Jetz, W.; et al. Wildlife Insights: A Platform to Maximize the Potential of Camera Trap and Other Passive Sensor Wildlife Data for the Planet. *Environ. Conserv.* **2020**, *47*, 1–6. [CrossRef]
28. Brust, C.A.; Käding, C.; Denzler, J. Active and Incremental Learning with Weak Supervision. *KI—Kunstl. Intell.* **2020**, *34*, 165–180. [CrossRef]
29. Carl, C.; Schönfeld, F.; Profft, I.; Klamm, A.; Landgraf, D. Automated detection of European wild mammal species in camera trap images with an existing and pre-trained computer vision model. *Eur. J. Wildl. Res.* **2020**, *66*, 62. [CrossRef]
30. Falzon, G.; Lawson, C.; Cheung, K.W.; Vernes, K.; Ballard, G.A.; Fleming, P.J.S.; Glen, A.S.; Milne, H.; Mather-Zardain, A.; Meek, P.D. ClassifyMe: A field-scouting software for the identification of wildlife in camera trap images. *Animals* **2020**, *10*, 58. [CrossRef] [PubMed]
31. Guo, Y.; Rothfus, T.A.; Ashour, A.S.; Si, L.; Du, C.; Ting, T.F. Varied channels region proposal and classification network for wildlife image classification under complex environment. *IET Image Process.* **2020**, *14*, 585–591. [CrossRef]
32. Nipko, R.B.; Holcombe, B.E.; Kelly, M.J. Identifying Individual Jaguars and Ocelots via Pattern-Recognition Software: Comparing HotSpotter and Wild-ID. *Wildl. Soc. Bull.* **2020**, *44*, 424–433. [CrossRef]
33. Saito, H.; Otake, T.; Kato, H.; Tokutake, M.; Semba, S.; Tomioka, Y.; Kohira, Y. Battery-Powered Wild Animal Detection Nodes with Deep Learning. *IEICE Trans. Commun.* **2020**, *E103-B*, 1394–1402. [CrossRef]
34. Schneider, S.; Greenberg, S.; Taylor, G.W.; Kremer, S.C. Three critical factors affecting automated image species recognition performance for camera traps. *Ecol. Evol.* **2020**, *10*, 3503–3517. [CrossRef] [PubMed]
35. Shahinfar, S.; Meek, P.; Falzon, G. "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. *Ecol. Inform.* **2020**, *57*, 101085. [CrossRef]

36. Tabak, M.A.; Norouzzadeh, M.S.; Wolfson, D.W.; Newton, E.J.; Boughton, R.K.; Ivan, J.S.; Odell, E.A.; Newkirk, E.S.; Conrey, R.Y.; Stenglein, J.; et al. Improving the accessibility and transferability of machine learning algorithms for identification of animals in camera trap images: MLWIC2. *Ecol. Evol.* **2020**, *10*, 10374–10383. [CrossRef] [PubMed]
37. Wei, W.; Luo, G.; Ran, J.; Li, J. Zilong: A tool to identify empty images in camera-trap data. *Ecol. Inform.* **2020**, *55*, 101021. [CrossRef]
38. Bain, M.; Nagrani, A.; Schofield, D.; Berdugo, S.; Bessa, J.; Owen, J.; Hockings, K.J.; Matsuzawa, T.; Hayashi, M.; Biro, D.; et al. Automated audiovisual behavior recognition in wild primates. *Sci. Adv.* **2021**, *7*, abi4883. [CrossRef] [PubMed]
39. Duggan, M.T.; Groleau, M.F.; Shealy, E.P.; Self, L.S.; Utter, T.E.; Waller, M.M.; Hall, B.C.; Stone, C.G.; Anderson, L.L.; Mousseau, T.A. An approach to rapid processing of camera trap images with minimal human input. *Ecol. Evol.* **2021**, *11*, 12051–12063. [CrossRef] [PubMed]
40. Kutugata, M.; Baumgardt, J.; Goolsby, J.A.; Racelis, A.E. Automatic Camera-Trap Classification Using Wildlife-Specific Deep Learning in Nilgai Management. *J. Fish Wildl. Manag.* **2021**, *12*, 412–421. [CrossRef]
41. Miao, Z.; Liu, Z.; Gaynor, K.M.; Palmer, M.S.; Yu, S.X.; Getz, W.M. Iterative human and automated identification of wildlife images. *Nat. Mach. Intell.* **2021**, *3*, 885–895. [CrossRef]
42. Moallem, G.; Pathirage, D.D.; Reznick, J.; Gallagher, J.; Sari-Sarraf, H. An explainable deep vision system for animal classification and detection in trail-camera images with automatic post-deployment retraining. *Knowl.-Based Syst.* **2021**, *216*, 106815. [CrossRef]
43. Norouzzadeh, M.S.; Morris, D.; Beery, S.; Joshi, N.; Jojic, N.; Clune, J. A deep active learning system for species identification and counting in camera trap images. *Methods Ecol. Evol.* **2020**, *12*, 150–161. [CrossRef]
44. Radig, B.; Bodesheim, P.; Korsch, D.; Denzler, J.; Haucke, T.; Klasen, M.; Steinhage, V. Automated Visual Large Scale Monitoring of Faunal Biodiversity. *Pattern Recognit. Image Anal.* **2021**, *31*, 477–488. [CrossRef]
45. Schindler, F.; Steinhage, V. Identification of animals and recognition of their actions in wildlife videos using deep learning techniques. *Ecol. Inform.* **2021**, *61*, 101215. [CrossRef]
46. Schindler, F.; Steinhage, V. Saving costs for video data annotation in wildlife monitoring. *Ecol. Inform.* **2021**, *65*, 101418. [CrossRef]
47. Shepley, A.; Falzon, G.; Lawson, C.; Meek, P.; Kwan, P. U-infuse: Democratization of customizable deep learning for object detection. *Sensors* **2021**, *21*, 2611. [CrossRef] [PubMed]
48. Vargas-Felipe, M.; Pellegrin, L.; Guevara-Carrizales, A.A.; López-Monroy, A.P.; Escalante, H.J.; Gonzalez-Fraga, J.A. Desert bighorn sheep (Ovis canadensis) recognition from camera traps based on learned features. *Ecol. Inform.* **2021**, *64*, 101328. [CrossRef]
49. Whytock, R.C.; Świeżewski, J.; Zwerts, J.A.; Bara-Słupski, T.; Koumba Pambo, A.F.; Rogala, M.; Bahaa-el din, L.; Boekee, K.; Brittain, S.; Cardoso, A.W.; et al. Robust ecological analysis of camera trap data labelled by a machine learning model. *Methods Ecol. Evol.* **2021**, *12*, 1080–1092. [CrossRef]
50. Xi, T.; Wang, J.; Qiao, H.; Lin, C.; Ji, L. Image Filtering and Labelling Assistant (IFLA): Expediting the analysis of data obtained from camera traps. *Ecol. Inform.* **2021**, *64*, 101355. [CrossRef]
51. Yang, D.Q.; Li, T.; Liu, M.T.; Li, X.W.; Chen, B.H. A systematic study of the class imbalance problem: Automatically identifying empty camera trap images using convolutional neural networks. *Ecol. Inform.* **2021**, *64*, 101350. [CrossRef]
52. Yang, D.Q.; Tan, K.; Huang, Z.P.; Li, X.W.; Chen, B.H.; Ren, G.P.; Xiao, W. An automatic method for removing empty camera trap images using ensemble learning. *Ecol. Evol.* **2021**, *11*, 7591–7601. [CrossRef] [PubMed]
53. Bilodeau, S.M.; Schwartz, A.W.; Xu, B.; Pauca, V.P.; Silman, M.R. A low-cost, long-term underwater camera trap network coupled with deep residual learning image analysis. *PLoS ONE* **2022**, *17*, e0263377. [CrossRef] [PubMed]
54. Bjerge, K.; Mann, H.M.R.; Høye, T.T. Real-time insect tracking and monitoring with computer vision and deep learning. *Remote Sens. Ecol. Conserv.* **2021**, *8*, 315–327. [CrossRef]
55. Bodesheim, P.; Blunk, J.; Körschens, M.; Brust, C.A.; Käding, C.; Denzler, J. Pre-trained models are not enough: Active and lifelong learning is important for long-term visual monitoring of mammals in biodiversity research—Individual identification and attribute prediction with image features from deep neural networks and decoupled decision models applied to elephants and great apes. *Mamm. Biol.* **2022**, *102*, 875–897. [CrossRef]
56. Corva, D.M.; Semianiw, N.I.; Eichholtzer, A.C.; Adams, S.D.; Mahmud, M.A.P.; Gaur, K.; Pestell, A.J.L.; Driscoll, D.A.; Kouzani, A.Z. A Smart Camera Trap for Detection of Endotherms and Ectotherms. *Sensors* **2022**, *22*, 4094. [CrossRef] [PubMed]
57. Fennell, M.; Beirne, C.; Burton, A.C. Use of object detection in camera trap image identification: Assessing a method to rapidly and accurately classify human and animal detections for research and application in recreation ecology. *Glob. Ecol. Conserv.* **2022**, *35*, e02104. [CrossRef]
58. Haucke, T.; Kühl, H.S.; Hoyer, J.; Steinhage, V. Overcoming the distance estimation bottleneck in estimating animal abundance with camera traps. *Ecol. Inform.* **2022**, *68*, 101536. [CrossRef]
59. Hilton, M.L.; Goessling, J.M.; Knezevich, L.M.; Downer, J.M. Utility of machine learning for segmenting camera trap time-lapse recordings. *Wildl. Soc. Bull.* **2022**, *46*, e1342. [CrossRef]

60. Jia, L.; Tian, Y.; Zhang, J. Domain-Aware Neural Architecture Search for Classifying Animals in Camera Trap Images. *Animals* **2022**, *12*, 437. [CrossRef] [PubMed]
61. Johanns, P.; Haucke, T.; Steinhage, V. Automated distance estimation for wildlife camera trapping. *Ecol. Inform.* **2022**, *70*, 101734. [CrossRef]
62. Leorna, S.; Brinkman, T. Human vs. machine: Detecting wildlife in camera trap images. *Ecol. Inform.* **2022**, *72*, 101876. [CrossRef]
63. Nepovinnykh, E.; Eerola, T.; Biard, V.; Mutka, P.; Niemi, M.; Kunnasranta, M.; Kälviäinen, H. SealID: Saimaa Ringed Seal Re-Identification Dataset. *Sensors* **2022**, *22*, 7602. [CrossRef] [PubMed]
64. Riechmann, M.; Gardiner, R.; Waddington, K.; Rueger, R.; Leymarie, F.F.; Rueger, S. Motion vectors and deep neural networks for video camera traps. *Ecol. Inform.* **2022**, *69*, 101657. [CrossRef]
65. Ronoh, E.K.; Mirau, S.; Dida, M.A. Human-Wildlife Conflict Early Warning System Using the Internet of Things and Short Message Service. *Eng. Technol. Appl. Sci. Res.* **2022**, *12*, 8273–8277. [CrossRef]
66. Santangeli, A.; Chen, Y.; Boorman, M.; Sales Ligero, S.; Albert García, G. Semi-automated detection of tagged animals from camera trap images using artificial intelligence. *Ibis* **2022**, *164*, 1123–1131. [CrossRef]
67. Sato, R.; Saito, H.; Tomioka, Y.; Kohira, Y. Energy Reduction Methods for Wild Animal Detection Devices. *IEEE Access* **2022**, *10*, 24149–24161. [CrossRef]
68. Stančić, A.; Vyroubal, V.; Slijepčević, V. Classification Efficiency of Pre-Trained Deep CNN Models on Camera Trap Images. *J. Imaging* **2022**, *8*, 20. [CrossRef] [PubMed]
69. Tan, M.; Chao, W.; Cheng, J.K.; Zhou, M.; Ma, Y.; Jiang, X.; Ge, J.; Yu, L.; Feng, L. Animal Detection and Classification from Camera Trap Images Using Different Mainstream Object Detection Architectures. *Animals* **2022**, *12*, 1976. [CrossRef] [PubMed]
70. Wang, H.; Zhong, J.; Xu, Y.; Luo, G.; Jiang, B.; Hu, Q.; Lin, Y.; Ran, J. Automatically detecting the wild giant panda using deep learning with context and species distribution model. *Ecol. Inform.* **2022**, *72*, 101868. [CrossRef]
71. Westworth, S.O.A.; Chalmers, C.; Fergus, P.; Longmore, S.N.; Piel, A.K.; Wich, S.A. Understanding External Influences on Target Detection and Classification Using Camera Trap Images and Machine Learning. *Sensors* **2022**, *22*, 5386. [CrossRef] [PubMed]
72. Zhu, H.; Tian, Y.; Zhang, J. Class incremental learning for wildlife biodiversity monitoring in camera trap images. *Ecol. Inform.* **2022**, *71*, 101760. [CrossRef]
73. Zualkernan, I.; Dhou, S.; Judas, J.; Sajun, A.R.; Gomez, B.R.; Hussain, L.A. An IoT System Using Deep Learning to Classify Camera Trap Images on the Edge. *Computers* **2022**, *11*, 13. [CrossRef]
74. Battu, T.; Reddy Lakshmi, D.S. Animal image identification and classification using deep neural networks techniques. *Meas. Sens.* **2023**, *25*, 100611. [CrossRef]
75. Binta Islam, S.; Valles, D.; Hibbitts, T.J.; Ryberg, W.A.; Walkup, D.K.; Forstner, M.R.J. Animal Species Recognition with Deep Convolutional Neural Networks from Ecological Camera Trap Images. *Animals* **2023**, *13*, 1526. [CrossRef] [PubMed]
76. Bjerge, K.; Alison, J.; Dyrmann, M.; Frigaard, C.E.; Mann, H.M.; Høye, T.T. Accurate detection and identification of insects from camera trap images with deep learning. *PLoS Sustain. Transform.* **2023**, *2*, e0000051. [CrossRef]
77. Bohnett, E.; Holmberg, J.; Faryabi, S.P.; An, L.; Ahmad, B.; Rashid, W.; Ostrowski, S. Comparison of two individual identification algorithms for snow leopards (Panthera uncia) after automated detection. *Ecol. Inform.* **2023**, *77*, 102214. [CrossRef]
78. Bohnett, E.; Poya Faryabi, S.; Lewison, R.; An, L.; Bian, X.; Rajabi, A.M.; Jahed, N.; Rooyesh, H.; Mills, E.; Ramos, S.; et al. Human expertise combined with artificial intelligence improves performance of snow leopard camera trap studies. *Glob. Ecol. Conserv.* **2023**, *41*, e02350. [CrossRef]
79. Bothmann, L.; Wimmer, L.; Charrakh, O.; Weber, T.; Edelhoff, H.; Peters, W.; Nguyen, H.; Benjamin, C.; Menzel, A. Automated wildlife image classification: An active learning tool for ecological applications. *Ecol. Inform.* **2023**, *77*, 102231. [CrossRef]
80. Böhner, H.; Kleiven, E.F.; Ims, R.A.; Soininen, E.M. A semi-automatic workflow to process images from small mammal camera traps. *Ecol. Inform.* **2023**, *76*, 102150. [CrossRef]
81. Chalmers, C.; Fergus, P.; Wich, S.; Longmore, S.N.; Walsh, N.D.; Stephens, P.A.; Sutherland, C.; Matthews, N.; Mudde, J.; Nuseibeh, A. Removing Human Bottlenecks in Bird Classification Using Camera Trap Images and Deep Learning. *Remote Sens.* **2023**, *15*, 2638. [CrossRef]
82. Charlton, G.; Falzon, G.; Shepley, A.; Fleming, P.; Ballard, G.; Meek, P. The Sentinel Bait Station: An automated, intelligent design pest animal baiting system. *Wildl. Res.* **2024**, *51*, WR22183. [CrossRef]
83. Chen, H.; Lindshield, S.; Ndiaye, P.I.; Ndiaye, Y.H.; Pruetz, J.D.; Reibman, A.R. Applying Few-Shot Learning for In-the-Wild Camera-Trap Species Classification. *AI* **2023**, *4*, 574–597. [CrossRef]
84. Clarfeld, L.A.; Sirén, A.P.; Mulhall, B.M.; Wilson, T.L.; Bernier, E.; Farrell, J.; Lunde, G.; Hardy, N.; Gieder, K.D.; Abrams, R.; et al. Evaluating a tandem human-machine approach to labelling of wildlife in remote camera monitoring. *Ecol. Inform.* **2023**, *77*, 102257. [CrossRef]

85. Cunha, F.; dos Santos, E.M.; Colonna, J.G. Bag of tricks for long-tail visual recognition of animal species in camera-trap images. *Ecol. Inform.* **2023**, *76*, 102060. [CrossRef]
86. Fergus, P.; Chalmers, C.; Longmore, S.; Wich, S.; Warmenhove, C.; Swart, J.; Ngongwane, T.; Burger, A.; Ledgard, J.; Meijaard, E. Empowering Wildlife Guardians: An Equitable Digital Stewardship and Reward System for Biodiversity Conservation Using Deep Learning and 3/4G Camera Traps. *Remote Sens.* **2023**, *15*, 2730. [CrossRef]
87. He, W.; Luo, Z.; Tong, X.; Hu, X.; Chen, C.; Shu, Z. Long-Tailed Metrics and Object Detection in Camera Trap Datasets. *Appl. Sci.* **2023**, *13*, 6029. [CrossRef]
88. Jia, L.; Tian, Y.; Zhang, J. Neural architecture search based on packed samples for identifying animals in camera trap images. *Neural Comput. Appl.* **2023**, *35*, 10511–10533. [CrossRef]
89. Kartowisastro, I.H.; Latupapua, J. A Comparison of Adaptive Moment Estimation (Adam) and RMSProp Optimisation Techniques for Wildlife Animal Classification Using Convolutional Neural Networks. *Rev. D'Intelligence Artif.* **2023**, *37*, 1023–1030. [CrossRef]
90. Krivek, G.; Gillert, A.; Harder, M.; Fritze, M.; Frankowski, K.; Timm, L.; Meyer-Olbersleben, L.; von Lukas, U.F.; Kerth, G.; van Schaik, J. BatNet: A deep learning-based tool for automated bat species identification from camera trap images. *Remote Sens. Ecol. Conserv.* **2023**, *9*, 759–774. [CrossRef]
91. Lim, N.; Bifet, A.; Bull, D.; Frank, E.; Jia, Y.; Montiel, J.; Pfahringer, B. Showcasing the TAIAO project: Providing resources for machine learning from images of New Zealand's natural environment. *J. R. Soc. N. Zeal.* **2023**, *53*, 69–81. [CrossRef] [PubMed]
92. Maile, R.E.; Duggan, M.T.; Mousseau, T.A. The successes and pitfalls: Deep-learning effectiveness in a Chernobyl field camera trap application. *Ecol. Evol.* **2023**, *13*, e10454. [CrossRef] [PubMed]
93. Meng, D.Y.; Li, T.; Li, H.X.; Zhang, M.; Tan, K.; Huang, Z.P.; Li, N.; Wu, R.H.; Li, X.W.; Chen, B.H.; et al. A method for automatic identification and separation of wildlife images using ensemble learning. *Ecol. Inform.* **2023**, *77*, 102262. [CrossRef]
94. Norman, D.L.; Bischoff, P.H.; Wearn, O.R.; Ewers, R.M.; Rowcliffe, J.M.; Evans, B.; Sethi, S.; Chapman, P.M.; Freeman, R. Can CNN-based species classification generalise across variation in habitat within a camera trap survey? *Methods Ecol. Evol.* **2023**, *14*, 242–251. [CrossRef]
95. Ratnayake, M.N.; Amarathunga, D.C.; Zaman, A.; Dyer, A.G.; Dorin, A. Spatial Monitoring and Insect Behavioural Analysis Using Computer Vision for Precision Pollination. *Int. J. Comput. Vis.* **2023**, *131*, 591–606. [CrossRef]
96. Rigoudy, N.; Dussert, G.; Benyoub, A.; Besnard, A.; Birck, C.; Boyer, J.; Bollet, Y.; Bunz, Y.; Caussimont, G.; Chetouane, E.; et al. The DeepFaune initiative: A collaborative effort towards the automatic identification of European fauna in camera trap images. *Eur. J. Wildl. Res.* **2023**, *69*, 113. [CrossRef]
97. Sajithra Varun, S.; Nagarajan, G. DeepAID: A design of smart animal intrusion detection and classification using deep hybrid neural networks. *Soft Comput.* **2026**, *30*, 5303–5314. [CrossRef]
98. Simões, F.; Bouveyron, C.; Precioso, F. DeepWILD: Wildlife Identification, Localisation and estimation on camera trap videos using Deep learning. *Ecol. Inform.* **2023**, *75*, 102095. [CrossRef]
99. Suessle, V.; Arandjelovic, M.; Kalan, A.K.; Agbor, A.; Boesch, C.; Brazzola, G.; Deschner, T.; Dieguez, P.; Granjon, A.C.; Kuehl, H.; et al. Automatic Individual Identification of Patterned Solitary Species Based on Unlabeled Video Data. *J. WSCG* **2023**, *31*, 1–10. [CrossRef]
100. Walkup, D.K.; Ryberg, W.A.; Pierce, J.B.; Smith, E.; Childress, J.; East, F.; Pierce, B.L.; Brown, P.; Fielder, C.M.; Hibbitts, T.J. Testing the detection of large, secretive snakes using camera traps. *Wildl. Soc. Bull.* **2023**, *47*, e1408. [CrossRef]
101. Whytock, R.C.; Suijten, T.; van Deursen, T.; Świeżewski, J.; Mermiaghe, H.; Madamba, N.; Mouckoumou, N.; Zwerts, J.A.; Pambo, A.F.K.; Bahaa-el din, L.; et al. Real-time alerts from AI-enabled camera traps using the Iridium satellite network: A case-study in Gabon, Central Africa. *Methods Ecol. Evol.* **2023**, *14*, 867–874. [CrossRef]
102. Wiig, Ø.; Da Silva Teixeira, K.; Sena, L.; Santos De Oliveira, H.C.; Mendes-Oliveira, A.C. Identifying individual jaguars from camera-trap images using the HotSpotter program. *Mammalia* **2023**, *87*, 602–605. [CrossRef]
103. Zhang, Q.; Yi, X.; Guo, J.; Tang, Y.; Feng, T.; Liu, R. A few-shot rare wildlife image classification method based on style migration data augmentation. *Ecol. Inform.* **2023**, *77*, 102237. [CrossRef]
104. Zhong, Y.; Li, X.; Xie, J.; Zhang, J. A Lightweight Automatic Wildlife Recognition Model Design Method Mitigating Shortcut Learning. *Animals* **2023**, *13*, 838. [CrossRef] [PubMed]
105. Biswas, D.; Tiwari, A. Utilizing Computer Vision and Deep Learning to Detect and Monitor Insects in Real Time by Analyzing Camera Trap Images. *Nat. Eng. Sci.* **2024**, *9*, 280–292. [CrossRef]
106. Bjerge, K.; Karstoft, H.; Mann, H.M.; Høye, T.T. A deep learning pipeline for time-lapse camera monitoring of insects and their floral environments. *Ecol. Inform.* **2024**, *84*, 102861. [CrossRef]
107. Borstelmann, A.; Haucke, T.; Steinhage, V. The Potential of Diffusion-Based Near-Infrared Image Colorization. *Sensors* **2024**, *24*, 1565. [CrossRef] [PubMed]

108. Brookes, O.; Mirmehdi, M.; Stephens, C.; Angedakin, S.; Corogenes, K.; Dowd, D.; Dieguez, P.; Hic. PanAf20K: A Large Video Dataset for Wild Ape Detection and Behaviour Recognition. *Int. J. Comput. Vis.* **2024**, *132*, 3086–3102. [CrossRef]
109. Celis, G.; Ungar, P.; Sokolov, A.; Sokolova, N.; Böhner, H.; Liu, D.; Gilg, O.; Fufachev, I.; Pokrovskaya, O.; Ims, R.A.; et al. A versatile, semi-automated image analysis workflow for time-lapse camera trap image classification. *Ecol. Inform.* **2024**, *81*, 102578. [CrossRef]
110. Chen, L.; Li, G.; Zhang, S.; Mao, W.; Zhang, M. YOLO-SAG: An improved wildlife object detection algorithm based on YOLOv8n. *Ecol. Inform.* **2024**, *83*, 102791. [CrossRef]
111. Darras, K.F.A.; Balle, M.; Xu, W.; Yan, Y.; Zakka, V.G.; Toledo-Hernández, M.; Sheng, D.; Lin, W.; Zhang, B.; Lan, Z.; et al. Eyes on nature: Embedded vision cameras for terrestrial biodiversity monitoring. *Methods Ecol. Evol.* **2024**, *15*, 2262–2275. [CrossRef]
112. Doan, T.N.; Le-Thi, D.N. Wildlife Species Classification from Camera Trap Images Using Fine-tuning EfficientNetV2. *Int. J. Intell. Eng. Syst.* **2024**, *17*, 624–638. [CrossRef]
113. Efremov, V.; Leus, A.; Gavrilov, D.; Mangazeev, D.; Kholodnyak, I.; Radysh, A.; Zuev, V.; Vodichev, N. Multistage Algorithm Based on Neural Networks for Preprocessing and Processing of Photo and Video Data from Camera Traps. *Sci. Tech. Inf. Process.* **2024**, *51*, 506–513. [CrossRef]
114. Fergus, P.; Chalmers, C.; Matthews, N.; Nixon, S.; Burger, A.; Hartley, O.; Sutherland, C.; Lambin, X.; Longmore, S.; Wich, S. Towards Context-Rich Automated Biodiversity Assessments: Deriving AI-Powered Insights from Camera Trap Data. *Sensors* **2024**, *24*, 8122. [CrossRef] [PubMed]
115. Gabeff, V.; Rußwurm, M.; Tuia, D.; Mathis, A. WildCLIP: Scene and Animal Attribute Retrieval from Camera Trap Data with Domain-Adapted Vision-Language Models. *Int. J. Comput. Vis.* **2024**, *132*, 3770–3786. [CrossRef]
116. Gadot, T.; Istrate, Ş.; Kim, H.; Morris, D.; Beery, S.; Birch, T.; Ahumada, J. To crop or not to crop: Comparing whole-image and cropped classification on a large dataset of camera trap images. *IET Comput. Vis.* **2024**, *18*, 1193–1208. [CrossRef]
117. Gutiérrez-Zapata, S.; Santoro, S.; Gegundez-Arias, M.E.; Selva, N.; Calzada, J. Dog invasions in protected areas: A case study using camera trapping, citizen science and artificial intelligence. *Glob. Ecol. Conserv.* **2024**, *54*, e03109. [CrossRef]
118. Henrich, M.; Burgueño, M.; Hoyer, J.; Haucke, T.; Steinhage, V.; Kühl, H.S.; Heurich, M. A semi-automated camera trap distance sampling approach for population density estimation. *Remote Sens. Ecol. Conserv.* **2023**, *10*, 156–171. [CrossRef]
119. Hopkins, J.; Santos-Elizondo, G.M.; Villablanca, F. Detecting and monitoring rodents using camera traps and machine learning versus live trapping for occupancy modeling. *Front. Ecol. Evol.* **2024**, *12*, 1359201. [CrossRef]
120. Huebner, S.E.; Palmer, M.S.; Packer, C. Human Supervision is Key to Achieving Accurate AI-assisted Wildlife Identifications in Camera Trap Images. *Citiz. Sci. Theory Pract.* **2024**, *9*, 38. [CrossRef]
121. Jenkins, M.; Franklin, K.A.; Nicoll, M.A.C.; Cole, N.C.; Ruhomaun, K.; Tatayah, V.; Mackiewicz, M. Improving Object Detection for Time-Lapse Imagery Using Temporal Features in Wildlife Monitoring. *Sensors* **2024**, *24*, 8002. [CrossRef] [PubMed]
122. Kissling, W.D.; Evans, J.C.; Zilber, R.; Breeze, T.D.; Shinneman, S.; Schneider, L.C.; Chalmers, C.; Fergus, P.; Wich, S.; Geelen, L.H. Development of a cost-efficient automated wildlife camera network in a European Natura 2000 site. *Basic Appl. Ecol.* **2024**, *79*, 141–152. [CrossRef]
123. Liu, L.; Mou, C.; Xu, F. Improved Wildlife Recognition through Fusing Camera Trap Images and Temporal Metadata. *Diversity* **2024**, *16*, 139. [CrossRef]
124. Lonsinger, R.C.; Dart, M.M.; Larsen, R.T.; Knight, R.N. Efficacy of machine learning image classification for automated occupancy-based monitoring. *Remote Sens. Ecol. Conserv.* **2023**, *10*, 56–71. [CrossRef]
125. Malmström, M.; Kullberg, A.; Skog, I.; Axehill, D.; Gustafsson, F. Extended Target Tracking Utilizing Machine-Learning Software—With Applications to Animal Classification. *IEEE Signal Process. Lett.* **2024**, *31*, 354–358. [CrossRef]
126. Mao, W.; Li, G.; Li, X. Improved Re-Parameterized Convolution for Wildlife Detection in Neighboring Regions of Southwest China. *Animals* **2024**, *14*, 1152. [CrossRef] [PubMed]
127. Mitterwallner, V.; Peters, A.; Edelhoff, H.; Mathes, G.; Nguyen, H.; Peters, W.; Heurich, M.; Steinbauer, M.J. Automated visitor and wildlife monitoring with camera traps and machine learning. *Remote Sens. Ecol. Conserv.* **2023**, *10*, 236–247. [CrossRef]
128. Nazir, S.; Kaleem, M. Object classification and visualization with edge artificial intelligence for a customized camera trap platform. *Ecol. Inform.* **2024**, *79*, 102453. [CrossRef]
129. Nepovinnykh, E.; Chelak, I.; Eerola, T.; Immonen, V.; Kälviäinen, H.; Kholiavchenko, M.; Stewart, C.V. Species-Agnostic Patterned Animal Re-identification by Aggregating Deep Local Features. *Int. J. Comput. Vis.* **2024**, *132*, 4003–4018. [CrossRef]
130. Nguyen, T.T.T.; Eichholtzer, A.C.; Driscoll, D.A.; Semianiw, N.I.; Corva, D.M.; Kouzani, A.Z.; Nguyen, T.T.; Nguyen, D.T. SAWIT: A small-sized animal wild image dataset with annotations. *Multimed. Tools Appl.* **2024**, *83*, 34083–34108. [CrossRef]
131. Park, S.; Cho, M.; Kim, S.; Choi, J.; Song, W.; Kim, W.; Song, Y.; Park, H.; Yoo, J.; Seo, S.B.; et al. Exploring the potential application of a custom deep learning model for camera trap analysis of local urban species. *Landsc. Ecol. Eng.* **2024**, *20*, 595–604. [CrossRef]
132. Penn, M.J.; Miles, V.; Astley, K.L.; Ham, C.; Woodroffe, R.; Rowcliffe, M.; Donnelly, C.A. Sherlock—A flexible, low-resource tool for processing camera-trapping images. *Methods Ecol. Evol.* **2023**, *15*, 91–102. [CrossRef]

133. Petroni, L.; Natucci, L.; Massolo, A. An Ecologist-Friendly R Workflow for Expediting Species-Level Classification of Camera Trap Images. *Ecol. Evol.* **2024**, *14*, e70544. [CrossRef] [PubMed]
134. Schindler, F.; Steinhage, V.; van Beeck Calkoen, S.T.S.; Heurich, M. Action Detection for Wildlife Monitoring with Camera Traps Based on Segmentation with Filtering of Tracklets (SWIFT) and Mask-Guided Action Recognition (MAROON). *Appl. Sci.* **2024**, *14*, 514. [CrossRef]
135. Schneider, D.; Lindner, K.; Vogelbacher, M.; Bellafkir, H.; Farwig, N.; Freisleben, B. Recognition of European mammals and birds in camera trap images using deep neural networks. *IET Comput. Vis.* **2024**, *18*, 1162–1192. [CrossRef]
136. Schütz, A.K.; Louton, H.; Fischer, M.; Probst, C.; Gethmann, J.M.; Conraths, F.J.; Homeier-Bachmann, T. Automated Detection and Counting of Wild Boar in Camera Trap Images. *Animals* **2024**, *14*, 1408. [CrossRef] [PubMed]
137. Sharma, S.; Timilsina, S.; Gautam, B.P.; Watanabe, S.; Kondo, S.; Sato, K. Enhancing Sika Deer Identification: Integrating CNN-Based Siamese Networks with SVM Classification. *Electronics* **2024**, *13*, 2067. [CrossRef]
138. Shermeister, B.; Mor, D.; Levy, O. Leveraging camera traps and artificial intelligence to explore thermoregulation behaviour. *J. Anim. Ecol.* **2024**, *93*, 1246–1261. [CrossRef] [PubMed]
139. Sittinger, M.; Uhler, J.; Pink, M.; Herz, A. Insect detect: An open-source DIY camera trap for automated insect monitoring. *PLoS ONE* **2024**, *19*, e0295474. [CrossRef] [PubMed]
140. Smith, J.; Wycherley, A.; Mulvaney, J.; Lennane, N.; Reynolds, E.; Monks, C.A.; Evans, T.; Mooney, T.; Fancourt, B. Man versus machine: Cost and carbon emission savings of 4G-connected Artificial Intelligence technology for classifying species in camera trap images. *Sci. Rep.* **2024**, *14*, 14530. [CrossRef] [PubMed]
141. Song, Q.; Guan, Y.; Guo, X.; Guo, X.; Chen, Y.; Wang, H.; Ge, J.; Wang, T.; Bao, L. Benchmarking wild bird detection in complex forest scenes. *Ecol. Inform.* **2024**, *80*, 102466. [CrossRef]
142. Tøn, A.; Ahmed, A.; Imran, A.S.; Ullah, M.; Azad, R.M.A. Metadata augmented deep neural networks for wild animal classification. *Ecol. Inform.* **2024**, *83*, 102805. [CrossRef]
143. Velasco-Montero, D.; Fernández-Berni, J.; Carmona-Galán, R.; Sanglas, A.; Palomares, F. Reliable and efficient integration of AI into camera traps for smart wildlife monitoring based on continual learning. *Ecol. Inform.* **2024**, *83*, 102815. [CrossRef]
144. Wittmann, K.; Gamal Ibrahim, M.; Straw, A.D.; Klein, A.M.; Staab, M. Monitoring fast-moving animals—Building a customized camera system and evaluation toolset. *Methods Ecol. Evol.* **2024**, *15*, 836–842. [CrossRef]
145. Xu, N.; Ma, Z.; Xia, Y.; Dong, Y.; Zi, J.; Xu, D.; Xu, F.; Su, X.; Zhang, H.; Chen, F. A Serial Multi-Scale Feature Fusion and Enhancement Network for Amur Tiger Re-Identification. *Animals* **2024**, *14*, 1106. [CrossRef] [PubMed]
146. Yang, D.Q.; Meng, D.Y.; Li, H.X.; Li, M.T.; Jiang, H.L.; Tan, K.; Huang, Z.P.; Li, N.; Wu, R.H.; Li, X.W.; et al. A systematic study on transfer learning: Automatically identifying empty camera trap images using deep convolutional neural networks. *Ecol. Inform.* **2024**, *80*, 102527. [CrossRef]
147. Yang, Z.; Tian, Y.; Zhang, J. Adaptive image processing embedding to make the ecological tasks of deep learning more robust on camera traps images. *Ecol. Inform.* **2024**, *82*, 102705. [CrossRef]
148. Zampetti, A.; Mirante, D.; Palencia, P.; Santini, L. Towards an automated protocol for wildlife density estimation using camera-traps. *Methods Ecol. Evol.* **2024**, *15*, 2276–2288. [CrossRef]
149. Zeuss, D.; Bald, L.; Gottwald, J.; Becker, M.; Bellafkir, H.; Bendix, J.; Bengel, P.; Beumer, L.T.; Brandl, R.; Brändle, M.; et al. Nature 4.0: A networked sensor system for integrated biodiversity monitoring. *Glob. Change Biol.* **2023**, *30*, e17056. [CrossRef] [PubMed]
150. Aamir, M.; Wijers, M.; Loveridge, A.; Markham, A. A robust metric distance and height estimation pipeline for wildlife camera trap imagery. *Ecol. Inform.* **2025**, *92*, 103520. [CrossRef]
151. Akbar, R.R.S.; Rees, M.W.; Fleming, P.A.; Sohel, F. Body-part-based individual feral cat identification from camera trap images using deep learning. *Ecol. Inform.* **2025**, *90*, 103258. [CrossRef]
152. Appel, C.L.; Subramanian, A.; Koning, J.S.; Ngosi, M.; Sullivan, C.M.; Levi, T.; Lesmeister, D.B. Developing custom computer vision models with Njobvu-AI: A collaborative, user-friendly platform for ecological research. *Ecol. Appl.* **2025**, *35*, e70096. [CrossRef] [PubMed]
153. Bak, T.; Camp, R.J.; Burt, M.D.; Vogt, S. Automated methods for processing camera trap video data for distance sampling. *Pac. Conserv. Biol.* **2025**, *31*, PC25008. [CrossRef]
154. Bjerge, K.; Karstoft, H.; Høye, T.T. Towards edge processing of images from insect camera traps. *Remote Sens. Ecol. Conserv.* **2025**, *11*, 573–589. [CrossRef]
155. Brenner, L.J.; Rindlaub, N.; Matos, J.; Meyler, S.; Pollock, S.; Schuetzenmeister, F.; Holmes, N.D. Real-Time Island Biosecurity Surveillance: Evaluating a Wireless Camera Network for Ai-Assisted Early Detection of Invasive Mammals on Santa Cruz Island, CA. *West. North Am. Nat.* **2025**, *85*, 365–375. [CrossRef]
156. Brook, B.W.; Buettel, J.C.; Lunteren, P.v.; Rajmohan, P.P.; Aandahl, R.Z. MEWC: A user-friendly AI workflow for customised wildlife-image classification. *Peer Community J.* **2025**, *5*, e57. [CrossRef]

157. Burshakbayeva, L.M.; Narbayev, S.; Arynova, R.A.; Assylbekova, A.S.; Kazikhanova, S.R.; Nurgozhayeva, N.M.; Akhmetzhanova, N.A.; Berber, A.A. Using innovative technological methods for monitoring and accounting of rare and endangered species of wild animals in Kazakhstan. *Casp. J. Environ. Sci.* **2025**, *23*, 327–333. [CrossRef]
158. Cai, Y.; Tian, K.; Ji, L.; Xiao, Y.; Pang, D.; Hou, P.; Ji, Y.; Wang, L.; Li, X.; Lu, J.; et al. A novel target-oriented enhanced infrared camera trap data screening method. *Sci. Rep.* **2025**, *15*, 16323. [CrossRef] [PubMed]
159. Chalmers, C.; Fergus, P.; Wich, S.; Longmore, S.N.; Walsh, N.D.; Oliver, L.; Warrington, J.; Quinlan, J.; Appleby, K. AI-Driven Real-Time Monitoring of Ground-Nesting Birds: A Case Study on Curlew Detection Using YOLOv10. *Remote Sens.* **2025**, *17*, 769. [CrossRef]
160. Clarfeld, L.; Tang, C.; Huber, K.; Balantic, C.; Donovan, T. AMMonitor 2: Remote monitoring of biodiversity in an adaptive framework in R. *Methods Ecol. Evol.* **2025**, *16*, 477–485. [CrossRef]
161. Clarfeld, L.A.; Gieder, K.D.; Fuller, A.; Miao, Z.; Sirén, A.P.K.; Webb, S.M.; Morelli, T.L.; Wilson, T.L.; Kilborn, J.; Callahan, C.B.; et al. DeepFaune New England: A Species Classification Model for Trail Camera Images in Northeastern North America. *Ecol. Evol.* **2025**, *15*, e72174. [CrossRef] [PubMed]
162. Du, Z.; Wu, D.; Wen, Q.; Xu, F.; Liu, Z.; Li, C.; Luo, R. An Improved Lightweight Model for Protected Wildlife Detection in Camera Trap Images. *Sensors* **2025**, *25*, 7331. [CrossRef] [PubMed]
163. Dussert, G.; Chamaillé-Jammes, S.; Dray, S.; Miele, V. Being confident in confidence scores: Calibration in deep learning models for camera trap image sequences. *Remote Sens. Ecol. Conserv.* **2024**, *11*, 88–99. [CrossRef]
164. Dussert, G.; Miele, V.; Van Reeth, C.; Delestrade, A.; Dray, S.; Chamaillé-Jammes, S. Zero-shot animal behaviour classification with vision-language foundation models. *Methods Ecol. Evol.* **2025**, *16*, 1460–1472. [CrossRef]
165. Eichholtzer, A.C.; Nguyen, T.T.; Galletta, L.; Nguyen, H.S.; Corva, D.M.; Nguyen, D.T.; Driscoll, D.A. Integrating AI technologies and citizen science to fast-track small and ectothermic animal monitoring. *Ecol. Inform.* **2025**, *92*, 103457. [CrossRef]
166. Fargetta, G.; Anile, S.; Priano, L.; Spata, M.O.; Ortis, A.; Battiato, S. Bridging AI and wildlife conservation: Classifying wildcats from genetically validated images. *Ecol. Inform.* **2025**, *92*, 103467. [CrossRef]
167. Guo, C.; Miguel, A.; Maciejewski, A.A. Automatic Identification of Individual African Leopards in Unlabeled Camera Trap Images. *IEEE Trans. Autom. Sci. Eng.* **2024**, *22*, 2460–2471. [CrossRef]
168. Lech, P.; Marciniak, B.; Okarma, K. A Low-Cost Energy-Efficient IoT Camera Trap Network for Remote Forest Surveillance. *Electronics* **2025**, *14*, 4266. [CrossRef]
169. Li, H.X.; Zhang, M.; Meng, D.Y.; Geng, B.; Li, Z.K.; Huang, C.F.; Li, W.K.; Jiang, H.L.; Wu, R.H.; Li, X.W.; et al. An automatic identification method of common species based on ensemble learning. *Ecol. Inform.* **2025**, *86*, 103046. [CrossRef]
170. Ma, Y.; Tan, M.; Liu, X.; Zhang, Y.; Xu, Z.; Sun, W.; Ge, J.; Feng, L. Deep learning for Amur tiger re-identification in camera traps: A tool assisting population monitoring and spatio-temporal analysis. *Ecol. Indic.* **2025**, *171*, 113227. [CrossRef]
171. Magaldi, H.; Cornette, R.; Tibesigwa, J.J.; Katumba, R.; Rugonge, H.; Amarasekaran, B.; Anderson, N.; Cappelle, N.; Cardoso, A.W.; Cornélis, D.; et al. DeepForestVision: Automated wildlife identification for camera traps of African tropical forests. *Ecol. Solut. Evid.* **2025**, *6*, e70167. [CrossRef]
172. Mamapule, S.; Esiefarienrhe, M.; Obagbuwa, I.C. Automatic Identification and Counting of South African Animal Species in Camera Traps Using Deep Learning. *Int. J. Intell. Syst.* **2025**, *2025*, 1561380. [CrossRef]
173. Maslen, B.; Popovic, G.; Wang, D.; Jansen, A.; Warton, D. The Motion Picture: Leveraging Movement to Enhance AI Object Detection in Ecology. *Ecol. Evol.* **2025**, *15*, e71996. [CrossRef] [PubMed]
174. Mesnage, C.S.; Corbett, A.; Curry, J.; Rowlands, S.; Dutta, A.; Everson, R.; Simmons, B.I. Copy-Paste Augmentation Improves Automatic Species Identification in Camera Trap Images. *Ecol. Evol.* **2025**, *15*, e72357. [CrossRef] [PubMed]
175. Miller, S.; Kirkland, M.; Hart, K.M.; McCleery, R.A. Object detection-assisted workflow facilitates cryptic snake monitoring. *Remote Sens. Ecol. Conserv.* **2025**, *11*, 606–617. [CrossRef]
176. Mirka, B.; Lippitt, C.; Harris, G.; Converse, R.; Gurule, M.; Sesnie, S.; Butler, M.; Stewart, D.; Rossman, Z. A photogrammetric approach to the estimation of distance to animals in camera trap images. *Ecol. Inform.* **2025**, *89*, 103120. [CrossRef]
177. Mononen, T.; Hardwick, B.; Alcobia, S.; Barrett, A.; Blagoev, G.A.; Boyer, S.; Gonçalves, P.; Gottsberger, B.; Groner, E.; Ho, C.C.Y.; et al. An active ensemble classifier for detecting animal sequences from global camera trap data. *Methods Ecol. Evol.* **2025**, *16*, 2500–2516. [CrossRef]
178. Mulero-Pázmány, M.; Hurtado, S.; Barba-González, C.; Antequera-Gómez, M.L.; Díaz-Ruiz, F.; Real, R.; Navas-Delgado, I.; Aldana-Montes, J.F. Addressing significant challenges for animal detection in camera trap images: A novel deep learning-based approach. *Sci. Rep.* **2025**, *15*, 16191. [CrossRef] [PubMed]
179. Nepovinnykh, E.; Immonen, V.; Eerola, T.; Stewart, C.V.; Kälviäinen, H. Re-identification of patterned animals by multi-image feature aggregation and geometric similarity. *IET Comput. Vis.* **2025**, *19*, e12337. [CrossRef]

180. Okuley, O.S.; Aiello, C.M.; Glad, W.; Perkins, K.; Ianniello, R.; Darby, N.; Epps, C.W. Improving AI performance in wildlife monitoring through species and environment-specific training: A case study on desert Bighorn sheep. *Ecol. Inform.* **2025**, *89*, 103179. [CrossRef]
181. Pestell, A.J.L.; Rendall, A.R.; Sinclair, R.D.; Ritchie, E.G.; Nguyen, D.T.; Corva, D.M.; Eichholtzer, A.C.; Kouzani, A.Z.; Driscoll, D.A. Smart camera traps and computer vision improve detections of small fauna. *Ecosphere* **2025**, *16*, e70220. [CrossRef]
182. Sanderson, S.; Haines, G.E.; Reimchen, T.E.; Burton, A.C.; Beirne, C.; Hendry, A.P. Inferring bird communities on remote freshwater lakes through time-lapse imagery. *Can. J. Zool.* **2025**, *103*, 1–15. [CrossRef]
183. Santoro, S.; Gutiérrez-Zapata, S.; Calzada, J.; Selva, N.; Marín-Santos, D.; Beery, S.; Brandis, K.; Fernández de Viana, I.; Meek, P.; Mortelliti, A.; et al. Essential tools but overlooked bias: Artificial intelligence and citizen science classification affect camera trap data. *Methods Ecol. Evol.* **2025**, *16*, 2638–2652. [CrossRef]
184. Saradopoulos, I.; Potamitis, I.; Rigakis, I.; Konstantaras, A.; Barbounakis, I.S. Image Augmentation Using Both Background Extraction and the SAHI Approach in the Context of Vision-Based Insect Localization and Counting. *Information* **2025**, *16*, 10. [CrossRef]
185. Suessle, V.; Heurich, M.; Downs, C.T.; Weinmann, A.; Hergenroether, E. CNN-Based Flank Predictor for Quadruped Animal Species. *IET Comput. Vis.* **2025**, *19*, e70024. [CrossRef]
186. Sundaresan, A.; Parham, J.; Crall, J.; Warungu, R.; Muthami, T.; Miliko, J.; Mwangi, M.; Holmberg, J.; Berger-Wolf, T.; Rubenstein, D.; et al. Adapting the Re-ID Challenge for Static Sensors. *IET Comput. Vis.* **2025**, *19*, e70027. [CrossRef]
187. Wang, L.; Wang, S.; Deng, C.; Zhu, H.; Tian, Y.; Zhang, J. DeLoCo: Decoupled location context-guided framework for wildlife species classification using camera trap images. *Ecol. Inform.* **2025**, *85*, 102949. [CrossRef]
188. Weber, M.; Strijbis, J.; Osner, N.; Périquet-Pearce, S.; Crowther, T.W.; Werden, L.K. An open-source method for spatially and temporally explicit herbivory monitoring in semi-arid savannas. *J. Environ. Manag.* **2025**, *377*, 124690. [CrossRef] [PubMed]
189. Xie, P.; Gao, R.; Lu, W.; Shen, L. Weakly supervised bounding-box generation for camera-trap image based animal detection. *IET Comput. Vis.* **2024**, *19*, e12332. [CrossRef]
190. Zhu, R.; Zhao, E.; Hu, C.; Xie, J.; Zhang, J.; Hu, H. Metric learning unveiling disparities: A novel approach to recognize false trigger images in wildlife monitoring. *Ecol. Inform.* **2025**, *87*, 103091. [CrossRef]
191. Algasov, A.; Nepovinnykh, E.; Zolotarev, F.; Eerola, T.; Kälviäinen, H.; Stewart, C.V.; Otarashvili, L.; Holmberg, J.A. On Combining Animal Re-Identification Models to Address Small Datasets. *Int. J. Comput. Vis.* **2026**, *134*, 96. [CrossRef]
192. Ali, S.S.; Ali, S.D.; Javed, S.; Mudigansalage, S.; Hussain, I.; Lucas, C.; Dieuleveut, T.; Sorato, E.; Werghi, N. HBID24K: A New Benchmark Dataset for Vulnerable Houbara Bustard and Intruder Detection in Wildlife Monitoring. *Sci. Data* **2026**, *13*, 223. [CrossRef] [PubMed]
193. Amarathunga, D.C.; Duric, Z.; Hulthen, A.; Wang, A.; Chalise, M.; Ul Haque, M.; Parry, H. Image-based fine-scale analysis of insect movement patterns and environmental triggers using pitfall traps. *Smart Agric. Technol.* **2026**, *13*, 101712. [CrossRef]
194. Barnhill, A.; Towers, J.R.; Sutton, G.J.; Shaw, T.J.H.; Doniol-Valcroze, T.; Maier, A.; Nöth, E.; Bergler, C. Leveraging Co-Occurrence to Improve Deep Learning Photo-Identification in Social Animals. *Ecol. Evol.* **2026**, *16*, e73552. [CrossRef] [PubMed]
195. Bevan, P.A.; Pantazis, O.; Pringle, H.A.; Ferreira, G.B.; Ingram, D.J.; Madsen, E.K.; Thomas, L.; Thanet, D.R.; Silwal, T.; Rayamajhi, S.; et al. Deep learning-based ecological analysis of camera trap images is impacted by training data quality and quantity. *Remote Sens. Ecol. Conserv.* **2026**, *12*, 494–508. [CrossRef]
196. Bolduc, D.; Fauteux, D.; Genest, M.A.; Legagneux, P. Speeding up image annotation and AI-model training for wildlife images. *Ecol. Inform.* **2026**, *97*, 103926. [CrossRef]
197. Campanera-Moliné, A.; Sazatornil, V.; Brotons, L.; Lumbreras Ruiz, F.; Benavente, R.; Ramilo-Henry, M.; Ramos, C.; Palazón, S.; Ponsa, D. Inferring Brown Bear Hair Snare Interactions by Automatically Detecting Bipedalism on Camera Trap Images Using Pose Estimation and a Multilayer Perceptron. *Remote Sens. Ecol. Conserv.* **2026**. [CrossRef]
198. Chaturvedi, P.; Flies, A.S.; Han, S.C.; Connelly, W.M. A study on edge devices for image classification of the Tasmanian devil (Sarcophilus harrisii) for vaccine delivery. *Ecol. Solut. Evid.* **2026**, *7*, e70216. [CrossRef]
199. Cowans, A.; Lambin, X.; Hare, D.; Sutherland, C. Improving the integration of artificial intelligence into existing ecological inference workflows. *Methods Ecol. Evol.* **2024**, *17*, 228–237. [CrossRef]
200. Yang, F.; Zhao, C.; Wang, S.; Han, B.; Yang, L.; Xu, F. Adaptive mask resolution with collaborative edge optimization for wildlife instance segmentation. *Beijing Linye Daxue Xuebao/J. Beijing For. Univ.* **2026**, *48*, 26–42. [CrossRef]
201. Gardiner, R.J.; Rowlands, S.; Simmons, B.I. Towards scalable insect monitoring: Ultra-lightweight CNNs as on-device triggers for insect camera traps. *Methods Ecol. Evol.* **2025**, *17*, 357–370. [CrossRef]
202. Garthen, P.; Mudragada, H.S.C.; Schwarz, S.; Edelhoff, H.; Steinbauer, M.J.; Peters, W.; Kuecuekaytekin, K.; Bothmann, L. Distance-based population density estimates from camera traps: Strong impact of technical choices in automated pipelines. *Ecol. Inform.* **2026**, *95*, 103682. [CrossRef]
203. Grele, A.; Richards, L.A. BugNet: A rapid and scalable pipeline for automated insect monitoring using hierarchical data. *Front. Ecol. Evol.* **2026**, *14*, 1750931. [CrossRef]

204. Guo, C.; Miguel, A.; Maciejewski, A.A. A Human-in-the-Loop Solution for Individual Leopard Identification in Unlabeled Camera Trap Images. *IEEE Trans. Autom. Sci. Eng.* **2026**, *23*, 11020–11033. [CrossRef]
205. Henrich, M.; Fiderer, C.; Klamm, A.; Schneider, A.; Ballmann, A.; Stein, J.; Kratzer, R.; Reiner, R.; Greiner, S.; Twietmeyer, S.; et al. Camera traps and deep learning enable efficient large-scale density estimation of wildlife in temperate forest ecosystems. *Remote Sens. Ecol. Conserv.* **2025**, *12*, 148–163. [CrossRef]
206. Hitchcock, K.; Tollington, S.; Yarnell, R.W.; Williams, L.J.; Hamill, K.; Fergus, P. Optimizing the Accuracy and Efficiency of Camera Trap Image Analysis: Evaluating AI Model Performance and a Semi-Automated Workflow. *Remote Sens.* **2026**, *18*, 502. [CrossRef]
207. Hou, J.; Liu, C.; Liu, D.; Hull, V.; Wang, Y.; Zhao, X.; Tan, Y.; Shi, X.; Cheng, Y.; Tang, Z.; et al. Enhancing Wildlife Monitoring: An Advanced AI Approach for Accurate Giant Panda Behavior Detection and Conservation Insights. *Animals* **2026**, *16*, 943. [CrossRef] [PubMed]
208. Künast, R.; Jeschke, S.; Mäder, P.; Meyer, S. Prolonged reduction in grassland arthropod activity after mowing revealed by AI camera trap monitoring. *Basic Appl. Ecol.* **2026**, *94*, 1–13. [CrossRef]
209. Larsen, L.L.N.; Christensen, N.; Kristensen, T.; Faddersbøll, T.L.; Lassen, A.R.W.; Rasmussen, B.; Pagh, S.; Pertoldi, C. Automated Activity Tracking and Space Use Monitoring of Captive Jaguars with Machine Learning. *Animals* **2026**, *16*, 1504. [CrossRef] [PubMed]
210. Paul, A.; Raj, R.; Gourisaria, M.K.; Jha, A.V.; Bizon, N. YOLO-HARVEST: A hybrid ViT architecture with locality-enhanced attention for automated wildlife species classification. *Ecol. Inform.* **2026**, *94*, 103605. [CrossRef]
211. Picek, L.; Straka, J.; Jirik, M.; Belotti, E.; Duľa, M.; Krausová, J.; Bojda, M.; Cermak, V.; Bufka, L.; Dvořák, R.; et al. CzechLynx: A Dataset for Individual Identification and Pose Estimation of the Eurasian Lynx. *Sci. Data* **2026**, *13*, 511. [CrossRef] [PubMed]
212. Rajmohan, P.P.; Sharma, R.; Amir, Z.; Bruce, T.; Brook, B.W.; Morris, D.; Luskin, M.S. AI and computer vision for wildlife identification in camera trap images: Fine-tuning SpeciesNet outperforms local models for species classification. *Sci. Total Environ.* **2026**, *1042*, 181926. [CrossRef] [PubMed]
213. Ratnayake, M.N.; Gallon, L.; Toosi, A.N.; Dorin, A. A Motion-Based Compression and Tracking System for Video Camera Trap-Based Insect Behaviour Studies. *Int. J. Comput. Vis.* **2026**, *134*, 220. [CrossRef]
214. Ringwaldt, E.M.; Buettel, J.C.; Carver, S.; Brook, B.W. Epidemiological Dynamics of a Visually Apparent Disease: Camera Trapping and Machine-Learning Applied to Rumpwear in the Common Brushtail Possum. *Integr. Zool.* **2025**, *21*, 116–128. [CrossRef] [PubMed]
215. Santamaria, J.D.; Isaza, C.; Giraldo, J.H. WildIng: A Wildlife Image Invariant Representation Model for Geographical Domain Shift. *Int. J. Comput. Vis.* **2026**, *134*, 183. [CrossRef]
216. Sha, L.; Tian, Y.; Zhang, J. GDP-CT: Grouped data pruning for camera trap. *Expert Syst. Appl.* **2026**, *299*, 130234. [CrossRef]
217. Shahabaz, A.; Toczydłowska, J.; Gula, R.; Theuerkauf, J.; Sarkar, S. From relative to metric: Calibrating AI-based monocular depth learning models for distance sampling in wildlife monitoring applications. *Ecol. Inform.* **2026**, *96*, 103846. [CrossRef]
218. Thornton, D.; Morris, D.; King, T.; Perera-Romero, L.; Anderson, A.; Garcia-Anleu, R.; Fitkin, S.; Vynne, C. Identification of camera trap images by artificial intelligence and human experts produces similar multi-species occupancy models. *J. Appl. Ecol.* **2026**, *63*, e70370. [CrossRef]
219. Tleimat, J.M.; Krell, E.; McCracken, S.F. Needle in a haystack: Towards the use of machine learning to detect Texas tortoises in large datasets. *Amphib. Reptile Conserv.* **2026**, *20*, 67–81.
220. Ursem, L.; Bakker, E.S.; Jin, H.; van Leeuwen, C.H.A. Shorebird responses to fine-scale water level fluctuations and macrofauna biomass in a newly constructed freshwater wetland. *Restor. Ecol.* **2026**, *34*, e70329. [CrossRef]
221. Veazey, L.; Latty, C.; Chapman, Z.; Hollmen, T.E. Applying computer vision to accelerate monitoring and analysis of bird incubation behaviors: A case study using common eider nest camera footage. *Remote Sens. Ecol. Conserv.* **2025**, *12*, 21–37. [CrossRef]
222. Villalva, P.; Jordano, P. A Machine Learning Application to Camera-Traps: Robust Species Interactions Datasets for Analysis of Mutualistic Networks. *Ecol. Evol.* **2026**, *16*, e72584. [CrossRef] [PubMed]
223. Xu, N.; Dong, Y.; Zi, J.; Xia, Y.; Ma, Z.; Chen, F. Automatic re-identification of terrestrial mammals using deep learning and camera trap images. *Glob. Ecol. Conserv.* **2026**, *68*, e04223. [CrossRef]
224. Yousif, H.; Al-Milaji, Z. Image classification and object detection complexity optimization: Exploring deep learning models on camera trap and surveillance clips. *Results Control Optim.* **2026**, *22*, 100654. [CrossRef]
225. Zampetti, A.; Santini, L.; Ferreiro-Arias, I.; Paltrinieri, L.; Ortiz, I.; Cedeño-Panchez, B.A.; Baltzinger, C.; Beirne, C.; Bowler, M.T.; Forget, P.M.; et al. Introducing TropiCam-AI: A taxonomically flexible automated classifier of Neotropical arboreal mammals and birds from camera-trap data. *Methods Ecol. Evol.* **2026**, *17*, 1235–1247. [CrossRef]
226. Zábó, A.; Chaquinga, R.; Palacios Pérez, J.; Rubenstein, D.; Nagy, M.; Ahmad, A. RAPID: Real-time animal pattern re-identification on edge devices, an open-source tool for field deployment. *Methods Ecol. Evol.* **2026**, *17*, 2379–2394. [CrossRef]

227. Liang, L.W.; Fatimah, A.; Qian, Q.; Zhang, T.; Wang, X.; Zhang, F.Y.; Xiao, Q.; Guo, Z.; Jiang, A.L.; Zhao, X.; et al. Global Camera Trapping Inventory (GCTI) v.1.0: A worldwide inventory of camera-trapping studies for spatial biodiversity analysis. *Zool. Res.* **2026**, *474*, 1097–1105.
228. Gogula, S.; Khanna, M.R.; Neelappa, N.; Sundaram, A.; Kumar, E.R.; Kumar, S.R. Modernized wildlife surveillance and behaviour detection using a novel machine learning algorithm. *Int. J. Recent Innov. Trends Comput. Commun.* **2022**, *10*, 50–62. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.