Deep Learning for Combating Wildlife and Environmental Crime: A Survey

Falih Gozi Febrinanto[a,*], Sean P. McGowan[a], Jacob Maher[a], Phillip Cassey[a]

[a] *Wildlife Crime Research Hub, School of Biological Sciences, Adelaide University, Adelaide, 5001, SA, Australia*

Abstract

Wildlife and Environmental Crime (WEC) describes illegal activities that threaten biodiversity and harm the environment. In recent years, deep learning (DL) has shown strong potential in combating WEC by learning complex patterns in data through the extraction of high-dimensional features. DL also offers powerful capabilities for prediction, classification, and object tracking, and has become an increasingly valuable tool in conservation efforts. This paper provides a comprehensive review of recent advances in DL implementations to combat WEC. We introduce a new taxonomy that identifies key intervention points across four contexts: field (observations in natural ecosystems); records (administrative and legal data); physical (specimens in laboratories, markets, and enforcement facilities); and online platforms. Moreover, we review existing DL-based applications for WEC prevention and discuss key opportunities, as well as the remaining challenges. Finally, we outline future research directions that may enable more robust DL-based approaches to combat WEC.

Keywords: deep learning, wildlife crime, environmental crime, foundation models, wildlife monitoring, illegal wildlife trade

1. Introduction

Wildlife and Environmental Crime (WEC) describes the set of illicit activities that exploit or damage natural resources, ranging from animal trafficking to resource theft and habitat destruction [5]. This poses a huge risk to environmental security, biodiversity, and human wellbeing, including detrimental effects on ecosystem services and the introduction of invasive species and disease [20]. The main motivations of criminal actors vary, but they most commonly include financial gain and the maintenance of certain lifestyles, status, and identity [57]. According to a recent report, WEC is among the most lucrative illicit markets, generating approximately USD 110 to 281 billion in criminal gains annually[1]. Authorities work hard to prosecute offenders and implement preventive measures to safeguard future biodiversity and environmental assets. Several traditional approaches have been used to tackle these issues, from conducting border inspections to combat wildlife trafficking to manually monitoring wildlife habitats or digital spaces to detect illegal activities [84, 129]. However, the growing volume of items to be inspected and data to be analyzed makes traditional and manual labor challenging and difficult to sustain in combating WEC.

As challenges for combating WEC grow in scale and complexity, artificial intelligence (AI) has emerged as a powerful potential tool for understanding and addressing them. Traditional machine learning (ML) techniques, such as Support Vector Machines (SVMs) [124] and Bayesian models [82], have long been used to address conservation problems [62, 17]. However, recent advances in deep learning (DL), a specialized branch of ML that uses deep multi-layered neural networks [111, 99], have changed how complex tasks are approached, including the task of combating WEC [113, 81]. When highlighting the difference between ML and DL, one important concept is how they handle features in data [68]. In classical ML, experts often select or design features, such as size, color, shape, or structured columns, and the model then learns patterns based on these inputs. This can make the process easier to understand and control. In DL, the model can work directly with raw data, such as images, text, or audio, and automatically learn useful representations by building patterns from simple to complex ones. This makes DL particularly effective for high-dimensional and complex unstructured data, where it has often achieved strong performance compared with classical ML methods. In addition, foundation models (FMs) [106, 119], a subset of DL pre-trained on massive, diverse datasets using self-supervised or unsupervised learning, have been particularly impactful. These modern DL techniques have demonstrated strong performance across a range of applications, from species classification [97] and animal tracking [9] to hotspot prediction [132], making them a promising tool for strengthening current and future efforts to tackle WEC.

*Corresponding author

Email addresses: falih.febrinanto@adelaide.edu.au (Falih Gozi Febrinanto), sean.mcgowan@adelaide.edu.au (Sean P. McGowan), jacob.maher@adelaide.edu.au (Jacob Maher), phill.cassey@adelaide.edu.au (Phillip Cassey)

[1] https://www.fatf-gafi.org/en/topics/environmental-crime.html

Although modern DL techniques are increasingly applied to combat various types of crime, a limited number of surveys have systematically summarized recent advances and open challenges from a crime-focused perspective. Several existing survey papers have addressed related aspects of this domain, but they do not fully examine how DL can support WEC detection, investigation, and enforcement across different intervention points. For instance, Tuia et al. [126] provided a comprehensive review of machine learning and DL models across diverse data sources, ranging from mobile to stationary sensors. However, their work lacks a deeper analysis of the role of DL outside natural habitats, such as in controlled laboratory environments or across online and digital spaces. Lamba et al. [66] presented an extensive overview of DL applications and discussed key challenges in real-world deployment. Although the survey explained the end-to-end DL workflow in detail, it was published before the emergence of powerful FMs that leverage large-scale pre-training across multiple modalities. A recent review by Obisa et al. [95] discussed the use of AI, ML, and DL for detecting wildlife. Although it covered newer deep-learning models, its scope was mainly limited to wildlife-conservation tasks, such as animal monitoring. It did not specifically address WEC or examine the human actors involved in different crime types.

Other related surveys fall outside WEC's specific scope, as they primarily focus on wildlife management, human–wildlife interactions, or general environmental monitoring rather than crime-related activities. For example, Nandutu et al. [90] reviewed sensor technologies and learning-based methods for mitigating wildlife–vehicle collisions, while Ojija et al. [96] focused on human–wildlife conflict and recent AI-based solutions. More broadly, Alotaibi and Nassif [4] analyzed the role of AI in environmental monitoring, including water quality, air quality, and weather analysis. While these surveys are useful for understanding AI or DL applications in conservation and environmental domains, they do not analyze criminal actors or the broader enforcement context of WEC.

Here, we focus on tasks to combat WEC, and we aim to provide a clearer, more structured perspective by categorizing DL roles by intervention point: field contexts within natural ecosystems, record contexts involving historical or administrative data, physical contexts involving specimens outside natural habitats, and online contexts involving digital environments and online spaces. This perspective is useful because each intervention point produces different types of data and requires different modeling, deployment, and evaluation strategies. For example, field-based tasks may involve real-time surveillance, record-based tasks may assist with future prediction or risk analysis, physical-specimen tasks may support on-the-spot species identification given small pieces of species body parts, and online tasks may help detect illegal trade on digital platforms. By distinguishing these contexts, this categorization provides a structured overview of the field and offers practical guidance for selecting appropriate DL approaches to combat WEC.

The core contributions of this survey include:

- **A Novel Categorization.** We propose a new systematic categorization to classify current DL models for combating WEC into four intervention points: field, record, physical, and online contexts. This categorization provides a clear framework for understanding where DL models can be deployed.
- **A Comprehensive Review.** We conduct an extensive review to curate several state-of-the-art DL techniques for combating WEC, including recent foundation models, and assess practical applications for each model.
- **Challenges and Future Directions.** We identify the latest key challenges, including in-the-wild implementation challenges, and discuss future directions relevant to this topic, including the rise of agentic AI.

2. Preliminaries

In this section, we describe several related terms, definitions, and contexts used throughout this paper to provide a high-level understanding of the covered topics. First, we discuss the general definition of DL. Second, we provide a description of FMs. Third, we explain several related DL terms, including different learning paradigms and problem settings.

2.1. Deep Learning

DL is a subset of machine learning that uses multilayered artificial neural networks (ANN) to mimic how the human brain processes and filters information. It can be used to find complex patterns in large datasets across various modalities, including images, video, audio, text, time series, and graph data. DL enables image and object recognition, speech recognition, classification, clustering, forecasting, language translation, diagnostic support, and generative modeling. These systems operate under various paradigms, including supervised, unsupervised, self-supervised, and reinforcement learning, enabling neural networks to learn effectively from labeled data, unlabelled data, or interactive rewards.

DL models require training on available data before deployment for a specific task. Given an input feature vector x and its corresponding target y, we select a neural network model $f_\Theta(\cdot)$, parameterized by a set of parameters Θ, which maps the input to a prediction,

$$\hat{y} = f_\Theta(x). \tag{1}$$

To quantify the discrepancy, or to determine how accurate the prediction $\hat{y}$ is compared to the true target y, we define a loss function $\ell(\cdot)$:

$$\ell(\hat{y}, y), \tag{2}$$

which assigns a scalar error value to each prediction. Training consists of adjusting the parameters Θ to minimize the average loss over the training data.

DL is highly effective for combating WEC because it can process a diverse range of data, including images, audio, video, time series, and text. Additionally, DL scales efficiently with massive datasets, enabling the analysis of "big data" that traditional methods might struggle to process. Moreover, it can support real-time or near-real-time detection and monitoring, enabling wildlife protection efforts to move toward more active, immediate enforcement.

2.2. Foundation Models

FMs, a term coined by Bommasani et al. [14], are a branch of DL comprising large models pre-trained on large-scale datasets that serve as an adaptable basis for diverse downstream tasks [106]. Built on self-supervised learning, FMs include subsets such as Large Language Models (LLMs) for text generation and comprehension. Moreover, FMs' broader scope also enables the processing of audio, images, and video. Because these models are pre-trained on large datasets, they can be fine-tuned for specific applications, providing a unique opportunity to deploy them against WEC.

2.3. Related DL Terms

In this subsection, we explain several other related DL terms that cover learning paradigms, the learning objective, training procedure, problem settings, or the amount of labeled data available, as follows:

- **Supervised learning.** A learning paradigm where a model is trained on labeled examples, each input paired with a known correct answer, enabling the prediction of outputs for unseen inputs [68]. It covers both classification tasks and regression tasks.
- **Unsupervised learning.** A learning paradigm where a model discovers hidden structure or patterns in unlabeled data, typically applied to tasks such as clustering and anomaly detection [68].
- **Self-supervised learning.** A learning paradigm where a model learns representations from unlabeled data, such as by ensuring consistent representations of the same image under different transformations, allowing the model to learn useful features without human-annotated labels [21]. These learned representations can then be used for downstream tasks.
- **Semi-supervised learning.** A learning paradigm that trains a model using both a small amount of labeled data and a large amount of unlabeled data, reducing manual annotation while achieving strong performance [143].
- **Reinforcement learning.** A learning paradigm where a model learns a decision-making strategy by interacting with an environment, receiving numerical reward signals after each action, and updating its behavior to maximize cumulative reward over time [133].
- **Transfer learning.** A strategy that reuses knowledge learned by a model on one task or domain and applies it to a different but related task, thus reducing the need for large amounts of labeled data in the target domain [93].
- **Fine-tuning.** A specific type of transfer learning in which a pre-trained model's weights serve as a starting point and are updated using new data to adapt the model to a target task, requiring less data and training time [8].
- **Zero-shot learning.** A problem setting in which a model recognizes new categories without seeing labeled examples during training, instead relying on descriptions, such as attribute labels [101] or natural language text [103].
- **Few-shot learning.** A problem setting where a model must classify new, previously unseen categories from a very small number of labeled examples per category, typically ranging from one to five [136].
- **Active learning.** An iterative strategy in which the model selects the most informative or most uncertain unlabeled samples and requests feedback labels from a human expert to reduce annotation cost [131].
- **Prompt-based learning.** A paradigm in which a pre-trained model adapts to new tasks by conditioning on task instructions, called prompts, which may be fixed, hand-crafted text templates (hard prompts) or learnable vectors optimized during training (soft prompts), all without retraining the entire model [74].

2.4. Wildlife and Environmental Crime

WEC refers to a range of illegal activities that exploit or degrade natural resources, including crimes against animals, plants (and other taxa, e.g., fungi), and the habitats they live in. These crimes are often driven by financial gain and lifestyle choices, including consumption habits, hobbies, status-related preferences, and specific uses such as pets, drugs, and ornaments.

The negative impacts of WEC extend beyond financial losses to broader environmental, health, and social consequences. These impacts can create biosecurity threats, especially when interactions between wildlife, livestock, and people increase the risk of disease transmission, including zoonotic diseases that can spread from animals to humans. Moreover, WEC contributes to the introduction or spread of invasive species, which may disrupt local ecosystems

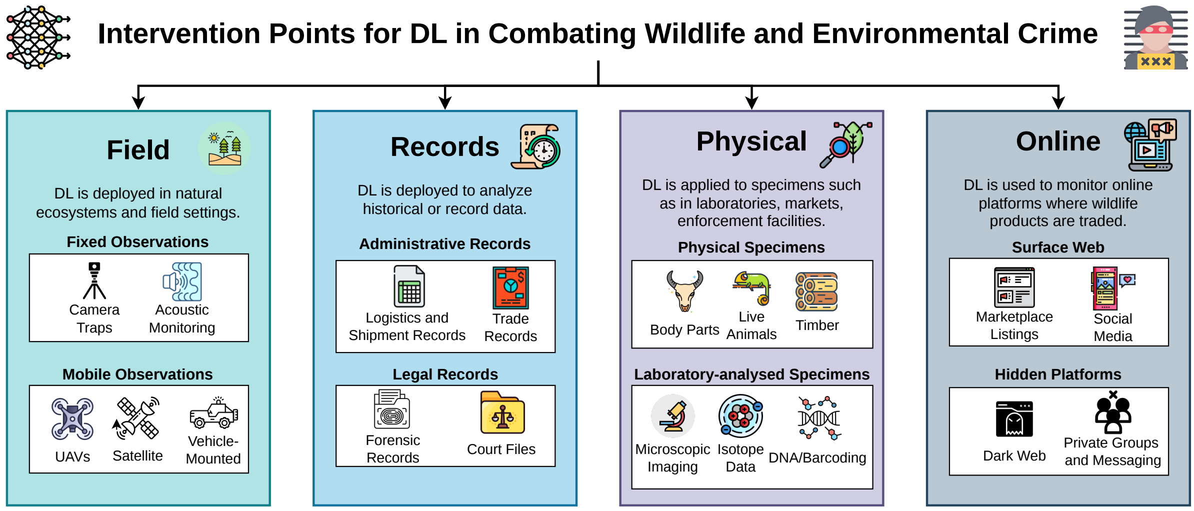

Figure 1: Categorization of DL implementation across different intervention points.

and further disturb native plants and animals. In addition, it can lead to animal welfare concerns when wildlife are injured, stressed, or killed as a result of human-animal conflict situations. Biodiversity loss is another important consequence, as WEC can damage habitats, reduce species populations, and disturb ecological balance. Several WEC types are described as follows:

- **Illegal trafficking:** unlawful trade and movement of wildlife or wildlife products, whether domestic or international, online or offline, including cross-border smuggling through concealment, false documentation, or other illicit methods.
- **Habitat destruction and degradation:** unlawful clearing, burning, draining, logging, and land conversion that can damage ecosystems and degrade wildlife habitats.
- **Poaching and illegal harvesting:** unauthorized taking, killing, capturing, or collecting of protected wildlife or related resources beyond legal limits or without permission.
- **Overexploitation:** unsustainable use of wildlife resources leading to population declines over time.
- **Biosecurity offenses:** breaches of laws intended to prevent the spread of invasive species or pathogens through wildlife, animals, plants, or related products.
- **Illegal dumping and pollution:** unlawful disposal of waste or release of pollutants that harm wildlife or habitats.
- **Natural resource theft:** illegal extraction of natural resources linked to wildlife or habitats, such as timber, fish, or protected biota, without lawful authority.

Historically, efforts to address these crimes have relied on traditional methods, such as border inspections, manual monitoring, and the analysis of descriptive statistics to inform decision-making. However, DL offers a huge opportunity to support and scale these efforts. It can help authorities generate data-driven recommendations and predictive insights, enabling earlier, more targeted interventions before environmental harm escalates.

3. Intervention Points

In this paper, we propose a novel categorization of deployment scenarios for DL to combat WEC. As illustrated in Figure 1, we group the literature by four distinct intervention points:

- **DL in field context:** DL is deployed in natural ecosystems and field settings.
- **DL in records context:** DL is deployed to analyze historical or record data.
- **DL in physical context:** DL is applied to specimens, such as in laboratories, markets, and enforcement facilities.
- **DL in online context:** DL is used to monitor online platforms where wildlife products are traded.

Categorizing DL deployment by intervention points clarifies where each DL method fits, thereby defining the data, constraints, and application targets for combating WEC. Moreover, it enables cleaner comparisons across studies, helping identify practical gaps and opportunities for real-world deployment. For each context, we describe how DL contributes and review DL models within that context.

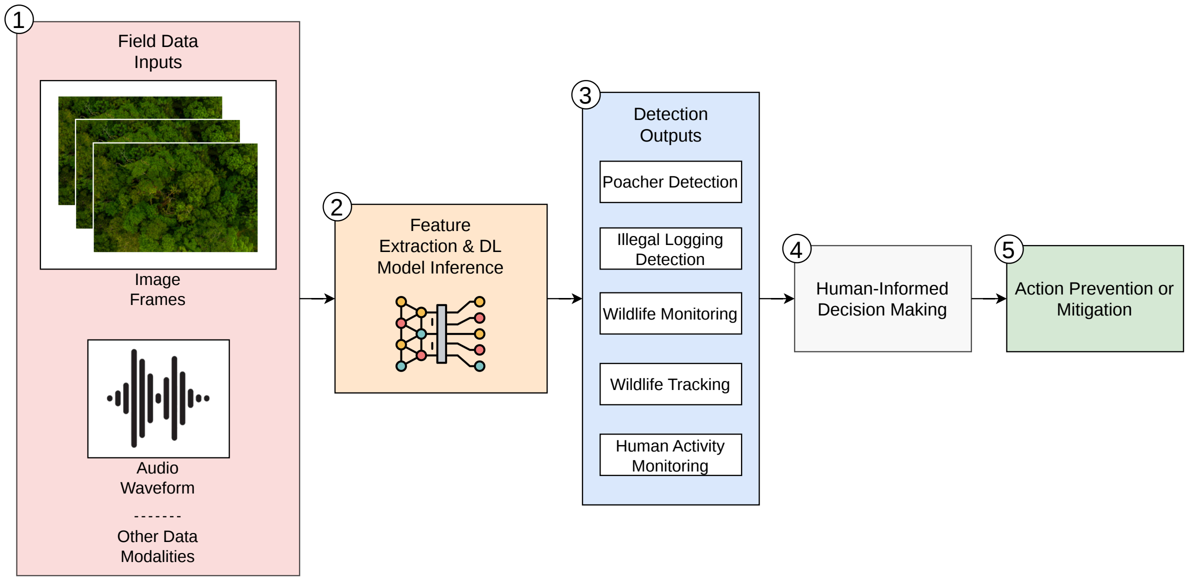

Figure 2: General deployment pipeline of DL in the field context for combating WEC. Field data inputs are collected through sensing devices and processed into numerical representations suitable for model input. A previously trained DL model is deployed to perform inference for prediction and detection tasks, such as identification, tracking, and monitoring. The resulting outputs are provided to humans, who can use this information to assess the situation, support mitigation, and take actions that may help prevent WEC.

Scope and selection. In this paper, we discussed 41 works related to DL for combating WEC. The candidate papers were selected through searches in Google Scholar, IEEE Xplore, ACM Digital Library, ScienceDirect, Springer Nature Link, and arXiv. We used search terms combining DL methods such as deep learning, Convolutional Neural Network (CNN), transformer, foundation model, LLM, vision-language model and other related DL terms in Section 2.3, with WEC terms such as seizure, wildlife crime, environmental crime and other related WEC terms in Section 2.4 for review, and organized into the four presented categorizations. We prioritized reputable sources, such as Q1 journals or CORE-ranked conferences, while also considering lower-ranked sources or high-quality preprints when they were directly relevant to the topic of DL for combating WEC. We excluded studies focused only on general conservation, ecology, or species monitoring when they did not explicitly or implicitly connect to illegal activity, enforcement, or crime actors. We also excluded some WEC-related works where no DL components were implemented.

3.1. DL in Field Context

In a field context, DL deployments are implemented inside natural ecosystems, including terrestrial and aquatic environments. DL supports data collection and monitoring for various decision-making needs. For example, DL models can process high-volume, heterogeneous field data, such as images, videos, audio, and time-series signals, to perform prediction tasks that support species monitoring in real-time or near-real-time settings. The general DL pipeline for combating WEC in the field context is shown in Figure 2.

This section reviews existing DL approaches for the field context, including those based on fixed sensing and mobile sensing, as summarized in Table 1.

3.1.1. Fixed Observations

Fixed observation refers to the collection of field data at known, stationary locations, either via deployed sensors or digitized observations from fixed monitoring points. Examples include camera traps, passive acoustic recorders, fixed surveillance cameras, and ranger observation posts. These observation points can capture or generate large volumes of data, making labeling, reviewing, and interpreting the outputs major bottlenecks for humans. DL can help automate this process for tasks such as object detection, classification, counting, and event recognition.

Camera Traps. Camera traps are now essential for ecological monitoring because they enable cost-effective and non-invasive observation of wildlife in their natural habitats. They are widely used for wildlife surveillance, and earlier work has focused on supporting species conservation rather than identifying the actors involved in illegal wildlife crime [128, 94, 18].

However, recent studies show how DL can process camera-trap data to help combat illegal activities that threaten biodiversity. For example, Fergus et al. [40] proposed a framework using camera traps to support a management and reward system in which each animal is assigned a digital twin. When an animal was detected, funds were released to local wildlife guardians, creating an interspecies payment structure that incentivizes protection against illegal activities. Detection was performed with a Faster

Table 1: DL Techniques for Combating Wildlife Crime in Field Context

Models	DL Techniques	Learning Settings	Applications	Data
Fergus et al. [40]	Faster R-CNN (ResNet-101)	Supervised	Species detection	Images via camera traps
Mitterwallner et al. [83]	MegaDetector v4.1	Supervised	Human & wildlife detection	Images via camera traps
Madhasu and Pande [81]	CCGAN + YOLO-CNAS	Supervised	Poacher & wildlife detection	Night-time camera-trap images
Katsis et al. [59]	CNN (ResNet-18)	Supervised	Gunshot detection	Audio via PAM
Vu et al. [130]	CNN (AlexNet, VGG-16, ResNet-18)	Supervised	Gunshot detection	Audio via PAM
Karthikeyan et al. [58]	CNN	Supervised	Chainsaw detection	Audio via PAM
Andreadis et al. [6]	CNN	Supervised	Chainsaw detection	Audio via IoT sensors
Stefanakis et al. [115]	LSTM-based DNN	Supervised	Chainsaw detection	Audio via PAM
SPOT [15]	Faster R-CNN (VGG-16)	Supervised	Poacher detection	Thermal infrared UAV images
ALSS-YOLO [49]	Modified YOLOv8-n	Supervised	Poacher & wildlife detection	Thermal infrared UAV images
Ramadan et al. [104]	CNN (TinyML)	Supervised	Illegal logging detection	Audio via IoT sensors
RegCD-Net [142]	CNN (ResNet-18)	Supervised	Illegal dumping detection	SUAV bi-temporal images
MultiSense DumpSpotter [146]	CNN (EfficientNet) + MLP + LLM (GPT-4o)	Supervised + prompt-based	Illegal dumpsite detection	Street-view images + geospatial/demographic metadata
Chen et al. [25]	YOLOv4 + NAS-FPN	Supervised	Illegal dumping detection	SUAV low-altitude images
Balaniuk et al. [10]	FCN + CNN	Supervised	Illegal mining detection	Satellite imagery (Sentinel)
Lv et al. [80]	Autoencoder + Isolation Forest	Unsupervised	Anomalous vessel detection	AIS vessel trajectory data
Paolo et al. [98]	Faster R-CNN	Supervised	Dark / IUU fishing vessel detection	SAR imagery (Sentinel)
Acharya et al. [1]	Faster R-CNN (ResNet-50)	Supervised	Bird-scaring line detection	EM video from fishing vessels
DeDOL [134]	DQN (Double DQN, Dueling)	Reinforcement	Patrol optimisation	Simulated grid-world data (real-time in-field information)

R-CNN model. They deployed 27 3G/4G camera traps across 400 km^2 in a South African reserve for 10 months, achieving an F1 score of 90.33% when identifying 12 animal species. Mitterwallner et al. [83] assessed the MegaDetector object-detection model for automated monitoring of people and wildlife in recreation areas. The model identified 19 wildlife species, human outdoor activities, and vehicles, and exhibited strong temporal accuracy for both during daytime and at night. Although that case study does not directly address illegal actors, the evaluation framework can still support authorities in analyzing patterns in both wildlife and human activity. More recently, Madhasu and Pande [81] introduced a nighttime wildlife surveillance approach that uses DL to detect animals. Although the system is designed primarily for detection, the authors claimed it could also help authorities identify smugglers or human intruders in forested regions. Their system first colorized night images with a customized Conditional GAN (CCGAN), then applied YOLO-CNAS for detection and multi-class classification. The approach delivered robust performance on nocturnal image-recognition tasks, achieving an F1 score of 95.17%.

Acoustic Monitoring. An acoustic monitoring system, usually called Passive acoustic monitoring (PAM), uses autonomous recording units to continuously capture sounds, producing vast amounts of audio data that require automated processing for practical analysis. DL approaches can process audio directly or through visual representations such as spectrograms to perform diverse bioacoustics classification tasks [16, 109]. DL is widely used to model PAM data and has supported conservation efforts such as monitoring species occurrences [54, 144]. Beyond conservation, recent work has also applied DL to PAM to combat wildlife crime, including illegal hunting and logging [130, 58].

Katsis et al. [59] used a type of CNN [71], specifically ResNet-18 [50], to classify gunshots from audio spectrograms and return candidate clips for human review. Their system targeted high recall to avoid missing rare gunshot events, achieving 0.95 recall. Another work by Vu et al. [130] developed a framework that pairs smartphone-based PAM with DL to map hunting activities in Chu Mom Ray National Park, Vietnam. They tested several CNN architectures, including AlexNet [64], VGG16 [114], and ResNet-18 [50]. They found that models perform well within a single site but transfer poorly across regions, highlighting the need for methods that better handle cross-region variability.

PAM data have also been applied to detecting illegal logging by identifying machinery sounds. CNN-based systems are common in this area, including work by Karthikeyan et al. [58] and Andreadis et al. [6]. A different approach by Stefanakis et al. [115] focused on chainsaw detection using a two-stage pipeline. First, a voice-activity detector extracted candidate segments from long recordings so the classifier did not need to process continuous audio, and second, a long short-term memory (LSTM) [145] determined whether each segment had chainsaw activity or not. Although their study does not explicitly specify the types of illegal actors, the method directly supports monitoring protected forests for unauthorized use of tools, such as chainsaws.

3.1.2. Mobile Observations

Mobile observation refers to situations in which data are collected or observed from moving platforms or agents that travel throughout natural ecosystems rather than remaining fixed in a single location. These include unmanned aerial vehicles (UAVs), satellite observation, vehicle-mounted systems, and ranger patrols.

UAV-Based Surveillance. UAVs, such as drones equipped with optical and thermal infrared cameras, can access remote or hazardous habitats and enable real-time monitoring over large areas. DL using UAV data is useful for conservation, for example, monitoring wildlife and vegetation over time without needing humans to manually inspect images for counting, change detection, or classification tasks [95].

DL with UAV data is also used to combat wildlife crime, including poaching, illegal logging, and illegal dumping. For example, SPOT (Systematic POacher deTector) [15] fine-tuned Faster R-CNN [107] (VGG-16 backbone [114]) on labeled thermal infrared frames from night UAV flights in African national parks to detect poachers and animals in near real time. Doull et al. [36] compared manual analysis with Faster R-CNN detection using drone-mounted thermal infrared (TIR) and RGB cameras. It also reported that automated analysis processed 1,400 images in about 2 hours, compared with 35 hours for manual processing. Another work, called ALSS-YOLO [49], was a lightweight detector for TIR UAV imagery designed to handle blur, overlap, low resolution, and noise. This case study can be used for poacher and wildlife detection and reported a precision of 0.887 on the BIRDSAI dataset [16].

For illegal logging, Ramadan et al. [104] presented a UAV-based Internet of Things (IoT) system with a CNN model for real-time detection of illegal logging by checking for human presence in the sites and reported better performance than classical ML baselines. For illegal dumping and waste monitoring, RegCD-Net [142] combined image registration to ensure that the source and target images shared the same perspective with change detection to detect where garbage patches had appeared in nature reserves using multiple small UAV settings, and it reported 96.74% precision with 20.66M parameters. Chen et al. [25] used SUAV imagery collected by DJI drones in two nature reserves in China and applied YOLOv4 [13] with feature fusion variants, reporting a precision of 91.34%.

Satellite Remote Sensing. Satellite imagery is widely used to monitor environmental and forest-cover changes across specific regions. It plays a crucial role in detecting environmental degradation and supporting conservation efforts in wildlife habitats [33]. Remote sensing also enables data-driven assessments of activities linked to WEC, such as illegal logging, habitat destruction, and poaching [45, 89]. Recent advances in DL have further improved accuracy and enabled near–real-time change detection compared to traditional algorithms.

Using satellite imagery, Balaniuk et al. [10] developed a low-cost and scalable DL pipeline to detect mines associated with illegal mining activities. Their approach adapts fully convolutional networks based on CNN architectures [71] for environmental impact classification, helping to identify hazardous or unauthorized mining operations around 263 mines that lack official concessions. Furthermore, Hojnik et al. [53] proposed a predictive modeling framework to generate illegal dumping risk maps in Maribor (Slovenia), demonstrating how remote sensing can support active environmental management. Paolo et al. [98] proposed a dataset and benchmark for detecting "dark" (non-broadcasted) fishing vessels, which can be associated with illegal fishing activities. It comprises nearly 1,000 analysis-ready satellite SAR (Synthetic Aperture Radar) images from the Sentinel-1 mission, each averaging 29,400-by-24,400 pixels. The paper provides a Faster R-CNN reference model to demonstrate the benchmark tasks on the dataset. Lv et al. [80] aimed to detect ships with anomalous Automatic Identification System (AIS) identities, where a vessel's trajectory behavior is inconsistent with its reported identity, which may indicate illegal fishing or smuggling activities. The method used an unsupervised autoencoder to learn implicit trajectory features from AIS data. Its proposed encoder, KACAN, combines convolutional components [71], attention [127], and a Kolmogorov-Arnold network (KAN) [78] components for representation learning. These learned features were concatenated with explicit rule-based trajectory features, such as speed statistics, acceleration, and heading changes, and a multidimensional isolation forest was then applied to detect anomalous ship identities.

Vehicle-mounted. Vehicle-mounted sensing refers to sensors deployed on moving vehicles, including land-based vehicles and sea-based platforms such as fishing vessels. This sensing source can provide useful data for DL systems to automate specific tasks for combating WEC.

Several works have been proposed around this intervention point. For example, Acharya et al. [1] proposed an Electronic Monitoring (EM) AI/ML workflow to automatically detect bird-scaring line (BSL) deployment on tuna longline vessels. Their method used a Faster R-CNN [107] object detector with a ResNet-50 [50] backbone, and they reported a detection precision of 0.87. This is important for compliance auditing because it can reduce manual EM video review and help authorities monitor whether bycatch mitigation measures are being used properly, which is critical for reducing the risk of seabird bycatch.

This vehicle-mounted sensing area has strong potential, but work directly targeting WEC is still limited. Another relevant study, although not directly implemented in natural habitats or designed specifically for WEC, is by Wu et al. [138], which introduced RoLID-11K, a large dashcam-based roadside-litter detection dataset. Roadside litter detection is closely aligned with illegal dumping detection from a waste-monitoring standpoint. In that study, the authors evaluated a wide range of deep-learning detectors, including transformer-based models (CO-DETR [151], DINO [21], RT-DETR [147], DiffusionDet [23], DEIMv2 [55]) and CNN models (YOLOv8 through YOLOv12) [125]. Another work, MultiSense DumpSpotter [146], tried to detect illegal dumping by integrating a multimodal DL architecture with LLMs to identify, classify, and analyze improper dumpsites. It fused a vision branch based on EfficientNet [123] with a custom MLP spatial branch to predict whether a sampled street-view location contained a potential dumpsite; sites flagged as positive were then passed to GPT-4o [56] to judge whether the dumpsite was improper or uncontrolled. The work also introduced UrbanDumpSight, a new dataset of over 4,000 street-view samples with geospatial and demographic metadata. Their results make the paper a useful technical reference for designing street-view or vehicle-mounted sensing pipelines for WEC-related applications.

Other Mobile Observation. In the case of field observation, it can also include other *real-time evidence* as data for DL to perform prediction or optimization to combat WEC in real ecosystems. Other real-time information can include footprints, a real-time indicator of human activity. Such information can support conservation enforcement in directly taking measurable actions in the field to prevent crime, assisted by DL output.

An example of other mobile observations is DeDOL [134]. Wang et al. (2019) modeled anti-poaching patrols as a Green Security Game with Real-Time Information (GSG-I), where rangers and poachers move sequentially and may observe real-time environmental traces, such as footprints, during patrols. They propose DeDOL, which combines game-theoretic strategy generation with a deep reinforcement learning model [133] to compute patrol strategies. In their work, the approach was evaluated in simulated grid-world anti-poaching environments and aimed to be transferable to real-time observations in real-world deployment.

3.2. DL in Records Context

The second intervention point is the records context, where DL is deployed to analyze historical or record data. Some useful analyses, such as understanding historical data patterns, help drive predictions of future values. The data can be sourced from administrative records such as logistics, shipments, or trade records, and from legal records such as past movement records, forensic, and case files. In this section, we review existing DL approaches for the records context as summarised in Table 2.

3.2.1. Administrative Records

WEC traces can be derived from historical data, such as logistics, shipment records, and trade records, where previous patterns of trade activities can be learned and later used to train DL models to identify rare or anomalous observations that may correspond to WEC activities. However, this is not limited to trade-related data, as administrative records can also cover broader areas, such as revealing poaching risk activities.

Several works cover this context, such as Chen et al. [26], who proposed a data-driven predict-and-plan framework for anti-poaching in a national nature reserve area using five years of SMART patrol records (2014–2018). They trained prediction-based ML and DL models, including a neural network with 5 hidden fully connected layers, bagging decision trees [12], and XGBoost [24], and an averaging ensemble. Then, it optimized multimodal patrol routes, such as walking and driving, to prioritize areas with a higher predicted poaching threat. Sanghera et al. [110] extended the trace-history approach to the maritime domain using AIS trajectory logs and a semi-supervised pipeline that combines ML and DL, where HMM (Hidden Markov Models)-based [86] segmentation and K-means clustering generate labels automatically, and random forest [108] and LSTM [145] models then classify vessel activities without pre-labeled data.

Datta et al. [32] presented a framework for analyzing US import trade data related to potentially suspicious wood and forest-product shipments, using their unsupervised anomaly-detection models, MEAD [30] and CHAD [31] to flag anomalous records for further investigation. The system combined a domain-knowledge module (incorporating CITES appendices, the IUCN red list[2], species-specific risk information, and expert-curated lists) with unsupervised anomaly-detection models using autoencoders with negative-sample generation for training, and NLP-based keyword extraction within the domain-knowledge module.

DIFFORACLE [63] was proposed as a patrol-planning method for green-security settings such as poaching prevention. It trained a conditional diffusion model to generate possible attacker behavior, such as poaching activity across connected park regions, given contextual information like historical patrol effort and spatial features. A Graph Convolutional Network (GCN) [61] is used as the diffusion model's backbone to model relationships between neighboring regions. The learned attacker model is then embedded in a game-theoretic optimization framework [141] to compute patrol strategies.

3.2.2. Legal Records

Similar to administrative records, legal records can also be useful as input for DL models to understand patterns from past cases. This aspect can include forensic records,

[2]https://www.iucnredlist.org/

Table 2: DL Techniques for Combating Wildlife Crime in Records Context

Models	DL Techniques	Learning Settings	Applications	Data
Chen et al. [26]	MLP + bagged trees + XGBoost	Supervised	Poaching threat prediction	SMART patrol records
Sanghera et al. [110]	HMM + LSTM	Semi-supervised	Fishing activity detection	AIS trajectory data
MEAD/ CHAD [32]	Autoencoder	Unsupervised	Illegal timber shipment detection	Trade shipping records
Coughlin et al. [29]	RoBERTa	Fine-tuning	Trafficking event extraction	EAGLE Network reports
Fu et al. [41]	LLM (DeepSeek-R1)	Prompt-based	Wildlife crime data extraction	Court verdicts (China)
Xiao and Ye [140]	DBN + CRF	Supervised	Wildlife crime pattern mining	Court judgments (China)
DIFFO-RACLE [63]	Conditional diffusion model + GCN	Generative modelling	Anti-poaching patrol planning	historical + synthetic poaching data

court files, and other legal documentation, where DL can assist decision-support systems or help analyze historical record patterns to support the WEC mitigation of highlighting similar future events.

For example, Coughlin et al. [29] developed a natural language processing (NLP) pipeline to automatically extract structured wildlife trafficking event data from monthly briefings published by the EAGLE (Eco Activists for Governance and Law Enforcement) Network. The pipeline used a RoBERTa-base transformer [77] to identify animals, products, countries, arrest counts, and quantities from unstructured report text. Fu et al. [41] proposed a framework to extract structured data from 247 court verdicts about wildlife crimes involving sea turtles in China (2012-2023) using the LLM, DeepSeek-R1 [44]. The model extracted 25 data items from each verdict, including seizure locations, species, quantities, values, and trafficking routes. The extracted data revealed that 1,236 individual sea turtles and 29,323 products were seized, with most seizures concentrated in southeastern coastal cities. This application specifically targeted illegal smuggling, trade, trafficking, and hunting of sea turtles. Xiao and Ye [140] proposed a characteristic mining method for wildlife crime based on judicial big data from China Judgments Online, covering 6,943 criminal judgment documents from 2013 to 2018. Using NLP-based information extraction techniques, the study employed a Deep Belief Network (DBN) [52] for spatial information extraction, specifically geographic entity recognition from text, and a Conditional Random Field (CRF) combined with rule-based methods for temporal information extraction. These methods were used to extract time-related case information and support analysis of the temporal patterns of wildlife crime in China. The result showed an F1 score of 87.4% for person information extraction.

3.3. DL in Physical Context

Physical intervention takes place outside a natural ecosystem. These include ports, airports, border crossings, markets, laboratories, and forensic facilities. In this context, DL supports and enhances analysis within established inspection or forensic workflows, especially in situations where manual processing would overwhelm human capacity.

In this section, we group physical context into two data observations. The first involves physical specimens, where DL is used for item-level identification. The second involves laboratory-analyzed specimens, where DL helps analyze specimen-derived data such as microscopic imaging, isotope data, and DNA/barcoding data for performing predictions. Several DL models, which are categorized as physical context interventions, are summarized in Table 3.

3.3.1. Physical Specimens

Physical data sources can be used to train DL for automatic tasks relevant to combating WEC. Using physical specimens, such as seized samples, helps authorities find useful evidence in physical contexts.

For example, Fein et al. [39] presented a pipeline for extracting and analyzing handwritten markings on seized elephant tusks to uncover connections among ivory trafficking networks. From 6,085 photographs spanning eight large seizures (2014-2019), the pipeline used a fine-tuned vision-language model (VLM) to detect handwriting, followed by semi-automated labeling using CLIP [103] embeddings and GPT-4o [56] for character recognition and description. The fine-tuned MM-GROUNDING-DINO [75] object detection model achieved 96% recall and 94% precision for handwriting detection. Another relevant recent work, Pirotta et al. [100], proposed a DL-based detector to identify trafficked marine wildlife, specifically shark fins, seahorses, and sea cucumbers, in airport and mail/traveler screening contexts using 3D X-ray CT scans. They used 68 physical marine samples to produce 298 original CT (Computed Tomography) scans, then expanded the training data using 3D Threat Image Projection (TIP) to train a detection model based on a 3D CNN with seven convolutional layers, each followed by ReLU activation and max-pooling. The model achieved high detection rates for shark fins, seahorses, and sea cucumbers, 95%, 96%, and

Table 3: DL Techniques for Combating Wildlife Crime in Physical Context

Models	DL Techniques	Learning Settings	Applications	Data
Fein et al. [39]	MM-DINO, CLIP, GPT-4o	Fine-tuning + prompt-based	Ivory trafficking network analysis	Seized ivory photographs
Pirotta et al. [100]	CNN	Supervised	Marine wildlife trafficking detection	3D X-ray CT scans of physical samples + TIP-augmented data
Chen et al. [27]	CNN	Supervised	Wood species identification	Vis/NIR hyperspectral data
Ergun [37]	CNN (ResNet-18, GoogLeNet, VGG-19)	Transfer learning	Wood species identification	Macroscopic wood images
Aghasanli et al. [2]	1D CNN	Transfer learning	Ivory source classification	Raman spectroscopy data
Zheng et al. [148]	CNN (ResNet-152)	Supervised	Wood species classification	Wood microscopic images
Zielinski et al. [150]	CNN (feature extraction)	Transfer learning	Wood species identification	Wood microscopic images
Prasetyo et al. [102]	multi-layer feedforward neural network	supervised	Shark and ray trade species identification	DNA melt-curve signatures
TabDDPM [135]	Diffusion Model + Transformer	Generative + supervised	Illegal ivory provenance tracing	Stable isotope data of ivory

86%, respectively, with false alarm rates of 2%, 9%, and 1%.

3.3.2. *Laboratory-analysed Specimens*

DL can assist with identification or prediction not only from raw data but also from laboratory-derived data, such as microscopic data, isotope data, and DNA data.

For example, Chen et al. [27] used CNN alongside several machine learning classifiers to deliver non-destructive species-level identification for trade enforcement, using Visible/ near-infrared hyperspectral data (383-2386 nm) from 800 Dalbergia wood samples. Similar work on wood identification from macroscopic images was presented by Ergun [37], which used transfer learning on pre-trained CNNs, including ResNet18 [50], GoogLeNet [122], VGG19 [114], and others. This paper evaluated multiple transfer learning architectures for wood species recognition from macro images and demonstrated strong performance across DL methods. It did not explicitly name a wildlife crime actor, but this case study is relevant to controlling illegal timber trade through improved species identification. Aghasanli et al. [2] proposed cross-domain transfer learning from inorganic mineral spectra to biological ivory identification using *Raman spectroscopy* data, and the best fine-tuned model reached about 99.7% accuracy. This work helps combat illegal ivory trade by distinguishing mammoth ivory, which may be legal in some jurisdictions, from illegal elephant ivory. Moreover, to reduce the *black-box* challenges of DL, Zheng et al. [148] proposed explainable DL for classifying wood microscopic images of endangered tree species. The paper used computer vision and DL to identify 15 CITES[3]-listed and look-alike endangered tree species from wood microscopic images of cross and tangential sections. A ResNet152 [50] model was trained on 850 images to discriminate species from genera, including Dalbergia, Pterocarpus, Swietenia, Carapa, Cedrela, and Swartzia, and achieved 99.3% classification accuracy on the 15-species dataset. The paper explicitly linked this application to illegal logging and illegal timber trade. Another relevant work, similar to the previous one, is Zielinski et al. [150], which used transfer learning with CNNs (pretrained on ImageNet) for wood microscopic image data of 77 Congolese timber species.

Prasetyo et al. [102] used a multi-layer feedforward neural network to identify shark and ray species, including CITES-listed species, from *DNA melt-curve* signatures. The data came from shark and ray trade samples collected across Java Island, Indonesia, including fishing ports, traditional markets, processing plants, export hubs, and inspection stations. The work aimed to support portable species identification for monitoring the shark and ray trade. The model achieved 79.41% test accuracy on independent samples. More recent work used TabDDPM [135] as a generative DL approach to trace the geographic provenance of illegal ivory using stable isotope analysis. The study used ivory isotope records from five regions: Asia, West Africa, Central Africa, East Africa, and Southern Africa. The pipeline first trained a diffusion model to generate synthetic isotopic signatures in tabular format, then trained a Transformer-based provenance classifier on the augmented data. It improved provenance classification on real isotope data, with accuracy rising from 87% to 94% and macro F1 from 80% to 89%.

3.4. *DL in Online Context*

Online interventions target WEC that is advertised, coordinated, or facilitated across various online platforms,

[3] https://cites.org/

including marketplaces, social media, messaging and closed-group platforms, and the dark web. DL can support these efforts by automatically identifying relevant wildlife-related content from both images and text. For example, foundation models such as Vision-Language Models (VLMs) and Large Language Models (LLMs) can help generate captions or descriptions from listing images and extract relevant information from unstructured titles and descriptions [11, 112]. This process is otherwise slow and expensive to perform manually, especially given the high volume and speed of online data. The use of foundation models in the online context is illustrated in Figure 3. Moreover, we summarize DL methods for this category in Table 4.

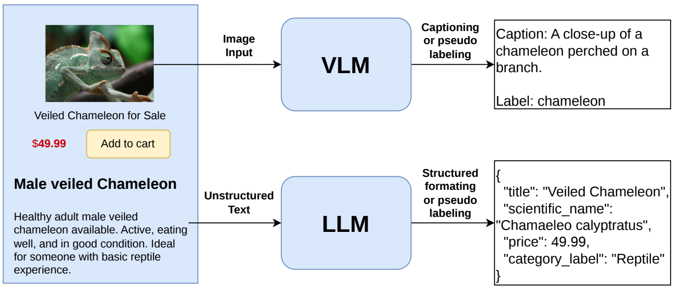

Figure 3: Foundation models assist with data extraction of illicit online marketplace activity.

3.4.1. Surface Web

The surface web consists of publicly accessible content indexed by search engines, including e-commerce sites, classified advertising platforms, and social media. DL pipelines in this space may suggest candidate content for information gathering [11], perform classification [88], and extract information for further analysis [112].

Marketplace listings. Marketplace listings enable many people to advertise their goods and services online, and the public can easily access these listings. This also creates space for illegal actors, such as wildlife traffickers, who might sell wildlife or wildlife products to targeted buyers [112]. These marketplace listings contain substantial information that can help authorities tackle illegal trade across platforms.

Monitoring marketplace listings requires a sophisticated pipeline to gather information from text, images, or videos to identify potential instances of illegal trafficking. This opens opportunities for DL to automate the extraction of useful indicators and detect suspicious listings. Several works have explored this.

LTS (Learning to Sample) [11] was a cost-aware labeling strategy that uses LLM inference to generate pseudo-labels, then trains cheaper specialized classifiers to spot illegal trafficking. It used foundation LLMs to assist with data extraction from text, especially GPT [56] or LLaMA [43] variants, to produce pseudo-labels indicating whether an ad is relevant. They also developed a CLIP [103]-based classifier to process multimodal ad images and text descriptions, passing them through a multi-layer perceptron to generate a final classification of animal versus non-animal products. The classifier reported an F1 score of up to 95%. Cardoso et al. [19] evaluated whether a transfer learning technique on deep vision models can identify online images associated with pangolin trafficking, including whole animals and traded parts. Their results showed strong performance across classification and detection settings, identifying over 90% of potential pangolin-trade instances. Furthermore, Sharma et al. [112] proposed a modular foundation-model-based vision–language pipeline using BLIP-2 [69] plus LLMs such as LLaMA [43] and FLAN-T5 [28] to analyze online advertisements for wildlife products and generate descriptions that include product type, species, and their status on the IUCN red list and CITES appendices. It targeted online sellers advertising wildlife-derived products and was designed as descriptive triage support for large-scale monitoring efforts. Finally, Kulkarni and Di Minin [65] incorporated vision models to detect online wildlife-trade content not only on online marketplaces but also on social media. It aimed to classify whether an image showed a live exotic animal in a captive or sale setting versus the same species in a wild or natural setting. Positive images were scraped from an exotic-animal trading website, while negative images of the same species in natural settings were collected from iNaturalist [4]. The developed models were tested using out-of-distribution images from Flickr, a community platform. They trained 24 CNN-based models across five architectures, including AlexNet [64], ResNet [50], and VGG [114].

Social media posts. Social media posts are also a good source for identifying illegal activities related to WEC. Social media platforms are accessible to the general public, and illicit activities, including wildlife trade, can also occur on these sites. The nature of social data, which continuously grows over time, makes it very difficult for humans to monitor manually. A smarter system is needed to detect and locate specific anomalous posts on social media that may be linked to WEC.

Several works have proposed using social media data to support mitigation efforts against WEC. For example, Momeny et al. [85] introduced a multimodal detection workflow that combined NLP models and deep CNNs such as DarkNet53 [105], InceptionResNetV2 [121], and EfficientNetV2 [123], followed by fusion to classify wildlife-trade-related YouTube posts. Their pipeline reportedly identified wildlife-trade-related content in 3,715 posts. Mou et al. [88] developed multimodal fusion approaches, such as BERT [34] plus ResNet [50] and BERT [34] plus ViT [35], to detect and recognize promotional behaviors related to the sale of wildlife products on online social networks, with a focus on ivory-product traders.

[4] https://www.inaturalist.org/

Table 4: DL Techniques for Combating Wildlife Crime in Online Context

Models	DL Techniques	Learning Settings	Applications	Data
LTS [11]	BERT, CLIP, GPT-4	Supervised (active learning)	Wildlife trafficking ad detection	Marketplace ads (eBay)
Cardoso et al. [19]	CNN (VGG-16, EfficientNet, DenseNet) + Faster R-CNN	Transfer learning	Pangolin trafficking detection	Online images (iNaturalist, Flickr)
Sharma et al. [112]	VLM (BLIP-2) + LLM (LLaMA 3.1, FLAN-T5)	Zero-shot / few-shot + fine-tuning	Wildlife product classification	Marketplace ads (text + images)
Kulkarni and Di Minin [65]	CNN (24 CNN-based models, including AlexNet, ResNet, and VGG)	Supervised (transfer learning + from scratch)	Online exotic-pet trade detection	Exotic-pet marketplace ads, iNaturalist, and Flickr images
Momeny et al. [85]	CNN (DarkNet-53, InceptionResNetV2, EfficientNetV2) + MPNet	Supervised	Wildlife trade detection	YouTube video frames + text
Mou et al. [88]	BERT, RoBERTa, ResNet-50, ViT	Supervised	Ivory trade post detection	Twitter/X posts (text + images)

Recent works follow the current trend of extracting structured data from unstructured content on online platforms using LLMs [149]. A common workflow involves three steps: (1) define the unstructured input text x and a textual prompt p for the LLM, (2) generate output tokens $\hat{t}$, and (3) parse these tokens into readable structured content output $\hat{y}$, often in a standardized format such as JSON, as illustrated below:

$$\text{Prompted Input:} \quad u = (p, x), \tag{3}$$

$$\text{Token Generation:} \quad \hat{t}_{1:T} = \phi_{\text{LLM}}(u), \tag{4}$$

$$\text{Structured Output:} \quad \hat{y} = \text{Parse}(\hat{t}_{1:T}), \tag{5}$$

where u is the fusion of the prompt p and unstructured text input x, ϕ_{LLM} is a pretrained LLM to generate token $\hat{t}$, T represents the maximum length of generated tokens and $\text{Parse}(\cdot)$ is a parsing function converting tokens to readable text.

3.4.2. Hidden Platforms

Research on hidden platforms generally focuses on detecting illegal traffickers or sellers who endanger biodiversity or endangered species. Hidden platforms include the deep web, where content is behind permission barriers (e.g., private messaging and closed groups), and the dark web, where hidden services are accessible via anonymous networks. Only a small number of studies have explored AI techniques to combat wildlife crime in these contexts.

Evidence from online wildlife-trade monitoring also shows that activity can shift from public posts to direct messaging channels, making detection harder once first contact is established [92]. Access in private or restricted environments typically requires authentication and permission to enter groups or conversations before useful data can be collected for DL-based prediction or analysis. It also raises ethical concerns because it requires hiding our true identity by bypassing traditional frameworks for informed consent and individual privacy [79], but the justification for doing so may outweigh the risk of illegal trade.

For the dark web, access is also difficult and usually begins with information gathering, such as scraping and search-based crawling of hidden structures. Importantly, current studies on dark-web wildlife monitoring have primarily relied on keyword-based analysis rather than on DL. Harrison et al. [47] found only one potential wildlife-related case in archived dark-web market data, while Stringham et al. [117] identified 3,332 wildlife-related ads from about 2 million ads, with most linked to recreational drug use.

Thus, DL contributions in this context are likely similar to those on open platforms, except for the process of gathering the raw data. After data collection, useful steps include extracting relevant content and converting unstructured posts into structured datasets, such as by labeling each post [11] for statistical analysis or proactive DL-based detection.

4. Applications

Section 3 reviewed DL approaches organized by *where* they are deployed: field, record, physical, and online contexts. In this section, we shift perspective and analyze *what computational tasks* DL can perform to combat WEC across these contexts. For the majority of use cases, the typical implementation of DL requires the following life cycle:

- **Problem definition:** Clearly describe the scope of the problem, including the data type, preferred output type, performance metrics, and any technical specifications required for effective implementation.
- **Data preparation:** Gather available datasets and appropriately augment and pre-process the data (including labeling for supervised tasks).

- **Model selection:** Choose the most appropriate model (pre-trained foundation model, or purpose-built model) for the task by considering data modality, speed, and available computational resources.

- **Training:** Train and fine-tune the model parameters, iteratively refining the model by evaluating its performance.

- **Deployment:** Integrate the model into the intended environment and monitor its performance to assess its effectiveness and failure modes.

Further details can be provided by considering four general applications for combating WEC:

- **Perception:** Extract structural information, such as classification ID or detected object, from data.

- **Pattern Analysis:** Extract patterns and relationships from data.

- **Forecasting:** Predict and anticipate future states based on past and current data.

- **Decision Support:** Prioritize and optimize real-world operational actions.

The following subsections describe these application families in more detail.

Figure 4: A conceptual illustration of DL for automatic species identification, where input wildlife images via camera-traps are processed by a DL model to predict species labels. Image courtesy of Jacob Maher, co-author of this paper, licensed under CC BY 4.0.

4.1. Perception

Perception tasks involve extracting structured information, including identification, detection, and tracking of objects across a wide range of data types. Identification tasks typically seek to assign a label to an input, i.e., assigning a species name to an animal in a photo, as illustrated in Figure 4. Therefore, labeled datasets are often required to train the model to accurately distinguish among different classes, making many identification problems a *supervised learning* paradigm. Supervised classification tasks in DL often follow the general idea of representation learning and use a classification head to map the learned latent representation to categorical labels [106], as follows:

$$\text{Representation Learning:} \quad z_i = f_{\mathrm{DL}}(x_i), \tag{6}$$
$$\text{Classification:} \quad \hat{y}_i = \tau(z_i), \tag{7}$$

where a DL encoder $f_{\mathrm{DL}}(\cdot)$ maps the input x_i to its latent representation z_i, which is then mapped to the categorical output label $\hat{y}_i$ using a classification head $\tau(\cdot)$.

Many existing foundation models achieve high accuracy in general-purpose identification tasks, which can later be adapted to more specific tasks [14] such as wildlife or environmental conservation. Detection and tracking problems, by contrast, focus on either one-time or continuous detection of objects within a dataset. As with identification tasks, various foundation models are designed to perform object detection and tracking [55, 8], which can be useful for combating WEC.

In the context of WEC, many problems related to identification and detection can be assisted by DL. Species classification is an application used to assign a taxonomic label or a higher-level category, such as genus or family, to an observation. This is commonly used to accelerate processing of acoustic data, camera-trap imagery, and other field-collected media, where manual labeling is often the primary bottleneck [94]. Recent methods using DL architectures have shown promise in discriminating look-alike species involved in wildlife trade. Liu et al. [76] used a CNN architecture, specifically ResNet with an attention mechanism, to differentiate CITES-listed tree species from their look-alikes. Su et al. [118] similarly demonstrated the effectiveness of using CNNs to discriminate the bird calls of 15 morphologically similar white-eye (*Zosterops*) species traded in wildlife markets. Techniques developed for field-based species identification can also assist in recognizing threatened species in online contexts, such as digital marketplaces. This allows the automatic detection of listings for particular species to alert authorities or experts to the illicit wildlife trade. The use of DL for detecting threatened species, such as pangolins, from images has shown strong promise for the automatic species identification on online platforms [19].

Object detection aims to localize instances of interest, such as animals or plants, and is often the first step in performing useful analytics. In wildlife-crime protection, detection is especially useful for filtering data, such as removing the large fraction of irrelevant parts in camera-trap footage and prioritizing frames containing objects of interest, such as animals, humans, or vehicles in protected areas, and supporting the ability to automate species counting [113]. Moreover, it is also useful for tracking, which refers to the continuous observation of animal and human activity over time to detect useful patterns, and often involves event detection as they occur via visual sensor streams such as cameras [9].

Despite the broad range of applications for perception tasks, these systems still rely heavily on labeled datasets. Foundation models offer useful opportunities as pretrained starting points, including YOLO variants [125], vision-language models such as CLIP [103] and BLIP-2 [69], and language models such as LLaMA [43] and BERT [34]. These models can then be adapted to support more specific perception tasks through transfer learning.

4.2. Pattern Analysis

Combating WEC often relies not just on identifying or tracking objects in data, but also on inferring informative patterns from it. Pattern analysis spans a wide range of problems, including anomaly detection, data mining, and causal discovery. Depending on the specific problem, these tasks may facilitate identifying structures that are not visible in any single sample by aggregating evidence across multiple observations.

Given the large volumes of data in wildlife and environmental contexts, there are many practical applications of pattern analysis to assist in understanding WEC. Anomaly detection aims to find rare or suspicious events that deviate from normal patterns by analyzing observations. In a WEC setting, anomalies might include unexpected human or vehicle presence in restricted zones, uncommon patterns of animal counts in object detection, or atypical changes in imagery. DL-based anomaly detection is often formulated using representation learning and then flagging outliers [106]. Recently, DL has been successfully used to detect anomalies in shipment data, alerting enforcement agencies to shipments that potentially carry harvested wood illegally [32].

Data mining refers to the process of analyzing very large datasets to identify patterns, relationships, or underlying structures that may not be immediately visible. In the context of wildlife crime, it can help reveal large-scale features within complex data and support a better understanding of wildlife offending and trade. By uncovering meaningful "portraits" from large data sources, data mining can provide decision-makers with useful information about wildlife and environmental crime (WEC). For example, in illegal wildlife trade cases, it may help identify links between sellers and buyers, recurring locations, commonly traded species, and pricing patterns. These insights can support efforts to prevent and deter future crime [140].

Inferring relationships in datasets by detecting significant causal relationships or correlations can help in understanding WEC activities and, therefore, their prevention. The correlation of search queries with illegal trade dynamics presents methods for WEC monitoring using search trends as a real-time proxy of demand [91]. Similarly, using Bayesian logistic regression, the smuggling of reptiles into Australia was shown to be predicted by the legal pet trade in the United States [116]. With the increased capabilities of data extraction and data mining, DL may offer more structured representations of WEC datasets and additional techniques for extracting causal and correlational relations beyond traditional statistical approaches by optimizing objective functions to capture larger structures or patterns in data collections.

4.3. Forecasting

Forecasting tasks predict future states based on historical and current data, producing anticipatory intelligence that enables proactive strategies to anticipate WEC activities. Particular practices that could benefit from DL-assisted forecasting include hotspot prediction, trade trend forecasting, and spatiotemporal projections of species and habitat distributions.

Hotspot prediction is the process of projecting specific locations where illegal activity is more likely to occur, producing spatial risk information to guide surveillance and enforcement teams [46]. These models typically combine historical incident data to predict future events, which may help the interception of WEC. For example, the prediction of poaching hotspots using neural networks improved when domain experts provided additional qualitative information on poaching activity patterns [46]. Working with domain experts enables the development of methods that leverage domain knowledge and historical data, yielding interpretable, data-efficient approaches.

The high predictive power of DL methods has been demonstrated by their ability to forecast financial trends accurately [7]. Using search trend data as predictor variables, DL approaches to short-term forecasting of lumber prices showed improved accuracy over traditional methods [51].

DL methods for spatiotemporal forecasting have been applied to broader ecological and conservation research to forecast species distributions and habitats. The utility of a range of machine learning methods, including CNNs and Random Forests, demonstrates that they often outperform traditional methods in predicting species distribution when sufficient data is available [60]. CNNs and Random Forests have also proved their efficacy in spatiotemporal problems by successfully predicting habitat suitability for seabird breeding [42]. These approaches, while critical for conservation efforts, may also help combat WEC by anticipating the dynamics of at-risk endangered species distributions.

Forecasting models follow a general idea across several processes: first, (1) defining the size of the series through window construction, (2) the DL model then performs representation learning to map the input series to a latent representation, and finally (3) a prediction is made for a defined step ahead, as follows:

$$\text{Window Construction:} \quad w_t = [x_{t-L+1}, \ldots, x_t], \tag{8}$$

$$\text{Representation Learning:} \quad h_t = f_{\mathrm{DL}}(w_t), \tag{9}$$

$$\text{Forecasting:} \quad \hat{x}_{t+H} = \tau(h_t), \tag{10}$$

where x_t represents an individual value (or vector of values) in time tick t, L is the window length, w_t is a time series input, $f_{\mathrm{DL}}(\cdot)$ maps the time series input w_t into its latent representation h_t, then the prediction head $\tau(\cdot)$ maps the latent representation into predicted value $\hat{x}_{t+H}$, with H representing how many steps ahead to predict.

Unlike perception tasks that involve species identification, forecasting tasks remain underexplored for combating WEC. This creates a promising direction for future development because forecasting has been extensively studied in other domains, such as demand forecasting in e-commerce [7] to ecological forecasting [67]. Adapting these methods for WEC data, such as patrol logs, seizure databases, and historical export-import databases such as CITES, could offer improvement to traditional approaches.

4.4. Decision Support

Decision support tasks focus on translating analytical outputs into real-world actions under uncertainty, such as allocating limited patrol resources or prioritizing enforcement targets. Unlike other applications, decision support assists in deciding *what to do next* given operational constraints, competing priorities, and limited resources. While incorporating DL methods in decision-making tasks may offer improvements to efficiency and outcomes, it is critical that interpretability and understanding are prioritized and, therefore, that the DL component acts only in a supporting role. For high-stakes decisions involving criminal justice or animal welfare, it is imperative that DL does not substitute human-based decision-making and only provides supporting evidence or guidance for further follow-up.

DL-based decision support for WEC is most commonly implemented in *field patrol optimization* [134, 26]. The DeDOL algorithm [134] combines game-theoretic reasoning with deep reinforcement learning (DRL) to optimize patrol strategies that respond to real-world observations, such as footprints. Furthermore, Chen et al. [26] implemented a predict-and-plan framework that uses DL-based threat predictions to optimize multimodal patrol routes.

In the online context, DL helps narrow an enormous search space to prioritize information for expert verification and follow-up. These techniques can expand the breadth and robustness of data scraping and may significantly speed up data analysis in the context of wildlife crime by enabling analysts to focus on high-risk trades under constraints of limited capacity to process them manually [11, 112].

5. Challenges

In this section, we discuss several challenges and open issues in improving the quality of DL models as an attempt to combat WEC, as summarized in the following points:

Domain Shift and Evolving Threats. A DL model trained under one set of conditions often fails when deployed in a different environment, a phenomenon known as domain shift or distribution shift [72]. For example, an acoustic classifier may perform well during controlled tests but degrade in real forests due to differences in background noise, sensor placement, and other contextual factors. Similarly, a camera-trap model trained in one habitat may struggle in another with different lighting, vegetation, and camera angles. The gap between training and deployment conditions is often identified only after a model is deployed in the field. Criminal tactics also evolve over time. Poachers adjust routes when patrol patterns shift, and online traffickers employ diverse strategies, platform accounts, and languages across social media, marketplaces, and other online platforms, all of which are considered evolving threats. A DL model trained on last year's patterns may fail to detect this year's threats.

Edge Deployment Under Field Constraints. WEC monitoring is sometimes deployed in settings with limited infrastructure, such as solar-powered camera traps and acoustic sensors in remote national parks, or portable screening devices at under-resourced border checkpoints. These settings impose limits on compute, memory, power, and connectivity that most DL architectures are not designed for. Running heavy models on low-power computing devices with unstable internet or satellite uplinks presents fundamentally different challenges from running them on high-performance computing resources. This condition, due to resource constraints, hinders the performance of DL models [73]. Model compression, effective inference scheduling, and mitigating failures such as overheating and battery drain can improve the quality of DL-based automatic monitoring.

Explainability for Enforcement. When a DL system flags suspicious activities such as an illegal shipment or identifies a poaching hotspot, the output needs to be actionable within enforcement and legal frameworks. A prosecutor cannot solely present a black-box confidence score or prediction results without explainable reasoning. This black-box nature of DL generally challenges the explainability [48]. For example, when a model detects chainsaw noise, it is not sufficient to output a confidence score alone. The system should also explain why the input was classified as a chainsaw event and provide an auditable trace that can serve as court evidence. Building systems that satisfy both technical performance and evidentiary standards remains an unsolved integration challenge.

Fusing Heterogeneous Data. A single wildlife trafficking case can leave traces, including seizure records, social media posts (including videos), marketplace listings in both text and image formats, satellite imagery, and network data showing import and export patterns. These signals can provide a richer picture than relying on a single modality [120]. Aligning and jointly reasoning over these heterogeneous data sources using DL remains a difficult open problem, especially when data fusion requires combining field-sensor and online data.

Rare-Event Problems. Illegal activities such as poaching incidents, illegal logging events, and trafficking seizures are rare compared to the volume of normal observations that monitoring systems collect regularly. This situation is, in technical terms, referred to as data imbalance [22]. Camera traps may record thousands of animals, acoustic monitors may process hours of environmental sound, satellite revisits may yield mostly unchanged scenes, and online scrapers may collect millions of listings. However, the number of illegal activities is very small. If the DL model uses supervised learning techniques, it may struggle with class imbalance. In addition, manually labeled data often requires a trained ecologist or someone with strong domain knowledge, making large-scale labeled datasets expensive and slow to build.

6. Future Directions

As we discussed the open issues in the earlier section, in this section, we explain broad future research directions to suggest a project area for researchers and practitioners for improving current DL models across different intervention points, as discussed in the following points:

Continual Learning. New data accumulate continuously, and their distribution may shift as criminal behaviors and environmental conditions change. Similarly, new tasks may emerge that the old trained model cannot handle. For example, in species identification, suppose one uses supervised learning with a multi-class objective; in the future, it may need to incorporate new types of species. Therefore, current classification models may not detect them, leading to misclassification. When we fine-tune the model on new data for new species, it may overwrite prior knowledge, leading to *catastrophic forgetting* [3]. Retraining from scratch may not be the best solution, as it requires access to all historical data and substantial computational resources. Another interesting direction is to implement a continual learning strategy that incrementally learns new examples without forgetting old patterns [38, 137]. This will help the detection system adapt over time and maintain good scalability.

Foundation Models for WEC-Specific Monitoring. While foundation models [8], such as large language models, vision-language models, and audio models, are used to combat WEC, they remain primarily general-purpose AI systems. It would be valuable to explore purpose-built foundation models trained specifically on diverse WEC-related tasks. Existing models, which were largely trained on data from well-studied ecosystems and species, could also be expanded to include a broader range of environments. This includes marine ecosystems, borderland and transboundary regions, and online environments in which WEC issues may arise.

Human-in-the-Loop. Future work could explore uncertainty-aware models that are able to recommend manual checks when decisions are uncertain. Instead of giving a hard prediction, the model could provide confidence scores or request human feedback before finalizing a result. This approach improves output quality and enables model refinement when corrected labels are used for fine-tuning. This technique is quite similar to the concept of active learning [131], in which models are updated over time and incorporate human input to ensure the quality of the labels used for updates. The illustration of active learning for the case of species identification is shown in Figure 5. Such a workflow strengthens AI–human teaming or human-in-the-loop [87], ensuring that humans remain in control of high-risk decisions and that predictions are reviewed appropriately.

Multi-Agent AI for Coordinated Surveillance. With the rise of agentic AI, there are many opportunities to automate DL processes, making them more proactive by implementing multiple agents simultaneously, thereby automating coordination [70]. For example, rather than treating each sensor as an independent detection pipeline, an acoustic monitor can trigger a nearby drone to fly over and obtain visual confirmation when it detects anomalous machine sounds. If there is a high chance of illegal activity, another AI agent could alert a ground patrol and help re-route and optimize their travel to the hotspot to intercept the actors involved. This requires a combination of DL algorithms, including vision models, audio models, deep reinforcement learning, and others. This will be an interesting future direction for providing more proactive mitigation of WEC.

Network Analysis. Graph or network data types can be used to depict the relationships of entities related to WEC. For example, wildlife trafficking operates through complex networks spanning many regions worldwide. It leaves traceable network patterns, such as importer–exporter relationships, where specialized DL models designed for network or graph data, such as graph learning methods [139], can be applied. Network or graph modeling can help detect anomalies or illegal activities at the node, edge, or whole-graph level. This opens new research directions that move beyond independently structured data, such as images,

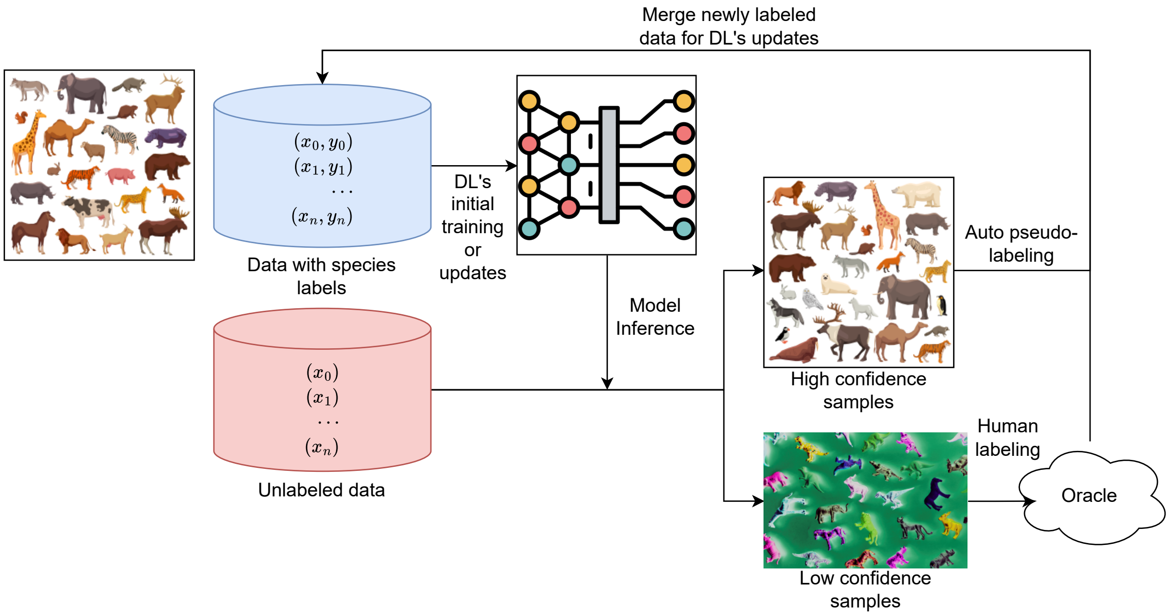

Figure 5: Illustration of active learning pipeline. Labeled data are first used to train an initial DL model. The trained model then predicts labels for a large pool of unlabeled data. High-confidence predictions can be used for pseudo-labeling, while low-confidence samples are selected for manual annotation. These newly labeled data are then combined to update the model further.

tabular data, or video, and instead incorporate network data, where graph learning architectures can propagate information between nodes via a *message passing* mechanism where it enables nodes in the netork to update their feature representations by iteratively exchanging and aggregating information with their neighboring nodes to potentially improve performance of related WEC detection or prediction tasks.

Test-Time Adaptation. As discussed in the open challenge above regarding domain shift and evolving threats, when a model deployed in one environment is later used in another, its performance often degrades. The model continues to produce outputs, but their reliability decreases as the visual or acoustic conditions change. DL-based test-time adaptation methods [72], which adjust model parameters using only unlabelled data from the new site, offer a promising solution. This future work aims to develop more robust DL-based models capable of supporting WEC detection and adapting effectively under domain shift or out-of-distribution scenarios.

7. Conclusion

Combating WEC is a complex and challenging task because it can occur both within and outside natural habitats and involves diverse sources and types of data. Despite these challenges, growing progress has been made through the use of AI, particularly DL, leading to the development of various frameworks that can support human efforts to protect wildlife and the environment from criminal actors. This paper provides a broad review of recent advances in DL for combating WEC. We propose a new categorisation that classifies existing DL models according to four intervention points: field, records, physical, and online contexts. We also discuss broad application tasks that illustrate DL's capabilities and highlight several remaining challenges and future research directions. We hope this survey helps shape researchers' and practitioners' understanding as they work together to develop impactful approaches for combating WEC.

CRediT authorship contribution statement

Falih Gozi Febrinanto: Conceptualization, Visualization, Writing – original draft, Writing – review & editing. **Sean P. McGowan:** Conceptualization, Writing – review & editing. **Jacob Maher:** Conceptualization, Writing – review & editing. **Phillip Cassey:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition.

References

[1] Acharya, D., Saqib, M., Devine, C., Untiedt, C., Little, L.R., Wang, D., Tuck, G.N., 2024. Using deep learning to automate the detection of bird scaring lines on fishing vessels. Biological Conservation 296, 110713.

[2] Aghasanli, A., Angelov, P., Kangin, D., Kerns, J., Shepherd, R., 2025. Transfer learning from inorganic materials to ivory detection. Scientific Reports 15, 15536.

[3] Aleixo, E.L., Colonna, J.G., Cristo, M., Fernandes, E., 2023. Catastrophic forgetting in deep learning: A comprehensive taxonomy. arXiv preprint arXiv:2312.10549 .

[4] Alotaibi, E., Nassif, N., 2024. Artificial intelligence in environmental monitoring: in-depth analysis. Discover Artificial Intelligence 4, 84.

[5] Anagnostou, M., Doberstein, B., 2022. Illegal wildlife trade and other organised crime: A scoping review. Ambio 51, 1615–1631.

[6] Andreadis, A., Giambene, G., Zambon, R., 2021. Monitoring illegal tree cutting through ultra-low-power smart iot devices. Sensors 21, 7593.

[7] Arsenault, P.D., Wang, S., Patenaude, J.M., 2025. A survey of explainable artificial intelligence (xai) in financial time series forecasting. ACM Computing Surveys 57, 1–37.

[8] Awais, M., Naseer, M., Khan, S., Anwer, R.M., Cholakkal, H., Shah, M., Yang, M.H., Khan, F.S., 2025. Foundation models defining a new era in vision: a survey and outlook. IEEE Transactions on Pattern Analysis and Machine Intelligence 47, 2245–2264.

[9] Balakrishnan, P., Leema, A.A., Kiruba, G.G., Gupta, A., Aryan, R., 2025. Deep-track: A real-time animal detection and monitoring system for mitigating human-wildlife conflict in fringe areas. Journal for Nature Conservation , 127063.

[10] Balaniuk, R., Isupova, O., Reece, S., 2020. Mining and tailings dam detection in satellite imagery using deep learning. Sensors 20, 6936.

[11] Barbosa, J.S., Gondhali, U., Petrossian, G., Sharma, K., Chakraborty, S., Jacquet, J., Freire, J., 2025. A cost-effective llm-based approach to identify wildlife trafficking in online marketplaces. Proceedings of the ACM on Management of Data 3, 1–23.

[12] Bhaduri, D., Toth, D., Holan, S.H., 2025. A review of tree-based methods for analyzing survey data. Wiley Interdisciplinary Reviews: Computational Statistics 17, e70010.

[13] Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M., 2020. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934 .

[14] Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M.S., Bohg, J., Bosselut, A., Brunskill, E., et al., 2021. On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258 .

[15] Bondi, E., Fang, F., Hamilton, M., Kar, D., Dmello, D., Choi, J., Hannaford, R., Iyer, A., Joppa, L., Tambe, M., et al., 2018. Spot poachers in action: Augmenting conservation drones with automatic detection in near real time, in: Proceedings of the AAAI Conference on Artificial Intelligence.

[16] Bondi, E., Jain, R., Aggrawal, P., Anand, S., Hannaford, R., Kapoor, A., Piavis, J., Shah, S., Joppa, L., Dilkina, B., et al., 2020. Birdsai: A dataset for detection and tracking in aerial thermal infrared videos, in: Proceedings of the IEEE/CVF Winter conference on applications of computer vision, pp. 1747–1756.

[17] Branco, V.V., Correia, L., Cardoso, P., 2023. The use of machine learning in species threats and conservation analysis. Biological Conservation 283, 110091.

[18] Bruce, T., Amir, Z., Allen, B.L., Alting, B.F., Amos, M., Augusteyn, J., Ballard, G.A., Behrendorff, L.M., Bell, K., Bengsen, A.J., et al., 2025. Large-scale and long-term wildlife research and monitoring using camera traps: a continental synthesis. Biological Reviews 100, 530–555.

[19] Cardoso, A.S., Bryukhova, S., Renna, F., Reino, L., Xu, C., Xiao, Z., Correia, R., Di Minin, E., Ribeiro, J., Vaz, A.S., 2023. Detecting wildlife trafficking in images from online platforms: A test case using deep learning with pangolin images. Biological Conservation 279, 109905.

[20] Cardoso, P., Amponsah-Mensah, K., Barreiros, J.P., Bouhuys, J., Cheung, H., Davies, A., Kumschick, S., Longhorn, S.J., Martínez-Muñoz, C.A., Morcatty, T.Q., et al., 2021. Scientists' warning to humanity on illegal or unsustainable wildlife trade. Biological Conservation 263, 109341.

[21] Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A., 2021. Emerging properties in self-supervised vision transformers, in: Proceedings of the IEEE/CVF international conference on computer vision, pp. 9650–9660.

[22] Carvalho, M., Pinho, A.J., Brás, S., 2025. Resampling approaches to handle class imbalance: a review from a data perspective. Journal of Big Data 12, 71.

[23] Chen, S., Sun, P., Song, Y., Luo, P., 2023. Diffusiondet: Diffusion model for object detection, in: Proceedings of the IEEE/CVF international conference on computer vision, pp. 19830–19843.

[24] Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785–794.

[25] Chen, W., Wang, H., Li, H., Li, Q., Yang, Y., Yang, K., 2021a. Real-time garbage object detection with data augmentation and feature fusion using suav low-altitude remote sensing images. IEEE Geoscience and Remote Sensing Letters 19, 1–5.

[26] Chen, W., Zhang, W., Liu, D., Li, W., Shi, X., Fang, F., 2021b. Data-driven multimodal patrol planning for anti-poaching, in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 15270–15277.

[27] Chen, Z., Xue, X., Wu, H., Gao, H., Wang, G., Ni, G., Cao, T., 2024. Visible/near-infrared hyperspectral imaging combined with machine learning for identification of ten dalbergia species. Frontiers in Plant Science 15, 1413215.

[28] Chung, H.W., Hou, L., Longpre, S., Zoph, B., Tay, Y., Fedus, W., Li, Y., Wang, X., Dehghani, M., Brahma, S., et al., 2024. Scaling instruction-finetuned language models. Journal of Machine Learning Research 25, 1–53.

[29] Coughlin, D., Gagnon, M., Grasso, V., Mou, G., Lee, K., Konrad, R., Raxter, P., Gore, M., 2022. Extracting and visualizing wildlife trafficking events from wildlife trafficking reports, in: 2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE. pp. 575–578.

[30] Datta, D., Islam, M.R., Self, N., Meadows, A., Simeone, J., Outhwaite, W., Keong, C.H., Smith, A., Walker, L., Ramakrishnan, N., 2020. Detecting suspicious timber trades, in: Proceedings of the AAAI conference on artificial intelligence, pp. 13248–13254.

[31] Datta, D., Muthiah, S., Simeone, J., Meadows, A., Ramakrishnan, N., 2021. Scrutinizing shipment records to thwart illegal timber trade, in: Proceedings of Outlier Detection Workshop, ACM SIGKDD, p. 9.

[32] Datta, D., Simeone, J.C., Meadows, A., Outhwaite, W., Keong Chen, H., Self, N., Walker, L., Ramakrishnan, N., 2025. Combating trade in illegal wood and forest products with machine learning. Plos one 20, e0311982.

[33] Delplanque, A., Théau, J., Foucher, S., Serati, G., Durand, S., Lejeune, P., 2024. Wildlife detection, counting and survey using satellite imagery: are we there yet? GIScience & Remote Sensing 61, 2348863.

[34] Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2019. Bert: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers), pp. 4171–4186.

[35] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2021. An image is worth 16x16 words: Transformers for image recognition at scale. ICLR .

[36] Doull, K.E., Chalmers, C., Fergus, P., Longmore, S., Piel, A.K., Wich, S.A., 2021. An evaluation of the factors affecting 'poacher'detection with drones and the efficacy of machine-learning for detection. Sensors 21, 4074.

[37] Ergun, H., 2024. Wood identification based on macroscopic images using deep and transfer learning approaches. PeerJ 12, e17021.

[38] Febrinanto, F.G., Xia, F., Moore, K., Thapa, C., Aggarwal, C., 2023. Graph lifelong learning: A survey. IEEE Computational Intelligence Magazine 18, 32–51.

[39] Fein, W., Horwitz, R.J., Brown III, J.E., Misra, A., Oviedo, F., White, K., Ferres, J.M.L., Wasser, S.K., 2025. Ai-driven detection and analysis of handwriting on seized ivory: A tool to uncover criminal networks in the illicit wildlife trade. arXiv preprint arXiv:2508.10219 .

[40] Fergus, P., Chalmers, C., Longmore, S., Wich, S., Warmenhove, C., Swart, J., Ngongwane, T., Burger, A., Ledgard, J., Meijaard, E., 2023. Empowering wildlife guardians: An equitable digital stewardship and reward system for biodiversity conservation using deep learning and 3/4g camera traps. Remote Sensing 15, 2730.

[41] Fu, C., Ren, W., Pagani-Núñez, E., Han, Y., Yang, J., Qiao, H., Li, Z., 2025. Ai-assisted data extraction helps uncover spatiotemporal patterns and socioeconomic drivers of wildlife crime involving sea turtles. Biological Conservation 312, 111511.

[42] Garcia-Quintas, A., Roy, A., Barbraud, C., Demarcq, H., Denis, D., Lanco Bertrand, S., 2023. Machine and deep learning approaches to understand and predict habitat suitability for seabird breeding. Ecology and Evolution 13, e10549.

[43] Grattafiori, A., Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Vaughan, A., et al., 2024. The llama 3 herd of models. arXiv preprint arXiv:2407.21783 .

[44] Guo, D., Yang, D., Zhang, H., Song, J., Wang, P., Zhu, Q., Xu, R., Zhang, R., Ma, S., Bi, X., et al., 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. arXiv preprint arXiv:2501.12948 .

[45] Guo, R., 2021. Using remote sensing imagery and machine learning to predict poaching in wildlife conservation parks, in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 15962–15963.

[46] Gurumurthy, S., Yu, L., Zhang, C., Jin, Y., Li, W., Zhang, X., Fang, F., 2018. Exploiting data and human knowledge for predicting wildlife poaching, in: Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies, pp. 1–8.

[47] Harrison, J.R., Roberts, D.L., Hernandez-Castro, J., 2016. Assessing the extent and nature of wildlife trade on the dark web. Conservation Biology 30, 900–904.

[48] Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M., Hussain, A., 2024. Interpreting black-box models: a review on explainable artificial intelligence. Cognitive Computation 16, 45–74.

[49] He, A., Li, X., Wu, X., Su, C., Chen, J., Xu, S., Guo, X., 2024. Alss-yolo: An adaptive lightweight channel split and shuffling network for tir wildlife detection in uav imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 17, 17308–17326.

[50] He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

[51] He, M., Li, W., Via, B.K., Zhang, Y., 2022. Nowcasting of lumber futures price with google trends index using machine learning and deep learning models. Forest Products Journal 72, 11–20.

[52] Hinton, G.E., 2009. Deep belief networks. Scholarpedia 4, 5947.

[53] Hojnik, B., Horvat, G., Mongus, D., Brumen, M., Kamnik, R., 2025. Towards sustainable waste management: Predictive modelling of illegal dumping risk zones using circular data loops and remote sensing. Sustainability 17, 8280.

[54] Hong, Z., Xie, Y., He, K., Zhou, Q., Miu, Y., Peng, H., Cao, Z., Fan, P., 2025. Deep audio detection networks from passive acoustic monitoring of a critically endangered primate. Integrative Zoology .

[55] Huang, S., Hou, Y., Liu, L., Yu, X., Shen, X., 2025. Real-time object detection meets dinov3. arXiv .

[56] Hurst, A., Lerer, A., Goucher, A.P., Perelman, A., Ramesh, A., Clark, A., Ostrow, A., Welihinda, A., Hayes, A., Radford, A., et al., 2024. Gpt-4o system card. arXiv preprint arXiv:2410.21276 .

[57] Janssen, J., Lemieux, A., Nivette, A., Ruiter, S., 2024. A scoping review on what motivates individuals to illegally harvest wildlife. Global Crime 25, 97–121.

[58] Karthikeyan, J., Sreehari, S., Koshy, J.R., Kavitha, K., 2021. Live acoustic monitoring of forests to detect illegal logging and animal activity, in: Advances in Computing and Network Communications: Proceedings of CoCoNet 2020, Volume 2. Springer, pp. 89–101.

[59] Katsis, L.K., Hill, A.P., Pina-Covarrubias, E., Prince, P., Rogers, A., Doncaster, C.P., Snaddon, J.L., 2022. Automated detection of gunshots in tropical forests using convolutional neural networks. Ecological Indicators 141, 109128.

[60] Kellenberger, B., Winner, K., Jetz, W., 2026. The performance and potential of deep learning for predicting species distributions. Global Ecology and Biogeography 35, e70184.

[61] Kipf, T.N., Welling, M., 2017. Semi-supervised classification with graph convolutional networks, in: International Conference on Learning Representations.

[62] de Knegt, H.J., Eikelboom, J.A., van Langevelde, F., Spruyt, W.F., Prins, H.H., 2021. Timely poacher detection and localization using sentinel animal movement. Scientific reports 11, 4596.

[63] Kong, L., Wang, H., Pan, Y., Kim, C.W., Song, M., Nguyen, A., Wang, T., Xu, H., Tambe, M., 2025. Robust optimization with diffusion models for green security, in: Conference on Uncertainty in Artificial Intelligence, PMLR. pp. 2325–2344.

[64] Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems 25.

[65] Kulkarni, R., Di Minin, E., 2023. Towards automatic detection of wildlife trade using machine vision models. Biological Conservation 279, 109924.

[66] Lamba, A., Cassey, P., Segaran, R.R., Koh, L.P., 2019. Deep learning for environmental conservation. Current Biology 29, R977–R982.

[67] Lapeyrolerie, M., Boettiger, C., 2023. Limits to ecological forecasting: Estimating uncertainty for critical transitions with deep learning. Methods in Ecology and Evolution 14, 785–798.

[68] LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. nature 521, 436–444.

[69] Li, J., Li, D., Savarese, S., Hoi, S., 2023. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models, in: International conference on machine learning, PMLR. pp. 19730–19742.

[70] Li, T., Zhu, K., Luong, N.C., Niyato, D., Wu, Q., Zhang, Y., Chen, B., 2022. Applications of multi-agent reinforcement learning in future internet: A comprehensive survey. IEEE Communications Surveys & Tutorials 24, 1240–1279.

[71] Li, Z., Liu, F., Yang, W., Peng, S., Zhou, J., 2021. A survey of convolutional neural networks: analysis, applications, and prospects. IEEE transactions on neural networks and learning systems 33, 6999–7019.

[72] Liang, J., He, R., Tan, T., 2025. A comprehensive survey on test-time adaptation under distribution shifts. International Journal of Computer Vision 133, 31–64.

[73] Liu, H.I., Galindo, M., Xie, H., Wong, L.K., Shuai, H.H., Li, Y.H., Cheng, W.H., 2024a. Lightweight deep learning for resource-constrained environments: A survey. ACM Computing Surveys 56, 1–42.

[74] Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., Neubig, G., 2023. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. ACM computing surveys 55, 1–35.

[75] Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Jiang, Q., Li, C., Yang, J., Su, H., et al., 2024b. Grounding dino: Marrying dino with grounded pre-training for open-set object detection, in: European conference on computer vision, Springer. pp. 38–55.

[76] Liu, S., Zheng, C., Wang, J., Lu, Y., Yao, J., Zou, Z., Yin, Y., He, T., 2024c. How to discriminate wood of cites-listed tree species from their look-alikes: using an attention mechanism with the resnet model on an enhanced macroscopic image dataset. Frontiers in Plant Science 15. doi:10.3389/fpls.2024.1368885.

[77] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V., 2019. Roberta: A robustly optimized bert pre-training approach. arXiv preprint arXiv:1907.11692 .

[78] Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljacic, M., Hou, T., Tegmark, M., 2025. Kan: Kolmogorov–arnold networks, in: International conference on learning representations, pp. 70367–70413.

[79] Logos, K., Brewer, R., Langos, C., Westlake, B., 2023. Establishing a framework for the ethical and legal use of web scrapers by cybercrime and cybersecurity researchers: learnings from a systematic review of australian research. International Journal of Law and Information Technology 31, 186–212.

[80] Lv, X., Shu, N., Jiang, R., Chang, C., Wu, T., 2025. An unsupervised deep learning approach for detecting anomalous identities of ships. Ocean Engineering 339, 122144.

[81] Madhasu, N., Pande, S.D., 2025. Revolutionizing wildlife protection: A novel approach combining deep learning and night-time surveillance. Multimedia Tools and Applications 84, 23465–23499.

[82] Miller, D.L., 2025. Bayesian views of generalized additive modelling. Methods in Ecology and Evolution 16, 446–455.

[83] Mitterwallner, V., Peters, A., Edelhoff, H., Mathes, G., Nguyen, H., Peters, W., Heurich, M., Steinbauer, M.J., 2024. Automated visitor and wildlife monitoring with camera traps and machine learning. Remote Sensing in Ecology and Conservation 10, 236–247.

[84] Moloney, G.K., Chaber, A.L., 2024. Where are you hiding the pangolins? screening tools to detect illicit contraband at international borders and their adaptability for illegal wildlife trafficking. Plos one 19, e0299152.

[85] Momeny, M., Kulkarni, R., Rinne, J., Soriano-Redondo, A., Di Minin, E., 2025. A multimodal learning approach for automated detection of wildlife trade on social media. bioRxiv , 2025–09.

[86] Mor, B., Garhwal, S., Kumar, A., 2021. A systematic review of hidden markov models and their applications. Archives of computational methods in engineering 28, 1429–1448.

[87] Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J., Fernández-Leal, Á., 2023. Human-in-the-loop machine learning: a state of the art. Artificial Intelligence Review 56, 3005–3054.

[88] Mou, G., Yue, Y., Lee, K., Zhang, Z., 2024. Wildlife product trading in online social networks: A case study on ivory-related product sales promotion posts, in: Proceedings of the International AAAI Conference on Web and Social Media, pp. 1096–1109.

[89] Mujetahid, A., Nursaputra, M., Soma, A.S., 2023. Monitoring illegal logging using google earth engine in sulawesi selatan tropical forest, indonesia. Forests 14, 652.

[90] Nandutu, I., Atemkeng, M., Okouma, P., 2022. Intelligent systems using sensors and/or machine learning to mitigate wildlife–vehicle collisions: A review, challenges, and new perspectives. Sensors 22, 2478.

[91] Ni, Q., Huang, Y., Xie, L., Zhang, M., Yao, Y., Xu, H., Zeng, C., Nijman, V., Xie, M., 2025. Digital footprints of wildlife crime: linking public search behavior to illegal trade dynamics for six keystone mammal taxa in china. Biological Conservation 308, 111273.

[92] Nijman, V., Smith, J.H., Foreman, G., Campera, M., Feddema, K., Nekaris, K., 2021. Monitoring the trade of legally protected wildlife on facebook and instagram illustrated by the advertising and sale of apes in indonesia. Diversity 13, 236.

[93] Niu, S., Liu, Y., Wang, J., Song, H., 2021. A decade survey of transfer learning (2010–2020). IEEE Transactions on Artificial Intelligence 1, 151–166.

[94] Norouzzadeh, M.S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M.S., Packer, C., Clune, J., 2018. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proceedings of the National Academy of Sciences 115, E5716–E5725.

[95] Obisa, M., Too, E., Osero, B., 2025. A review of artificial intelligence, machine learning, and deep learning and their applications in detecting wildlife animals. Measurement, Automation, and Control in Artificial Intelligence and Machine Learning , 1–32.

[96] Ojija, F., Ogwu, M.C., Ally, J., John, J.P., Stephano, A., Felix, N., Tekka, R., 2025. Artificial intelligence-driven solutions for mitigating human–wildlife conflict in biodiversity hotspots. Science Progress 108, 00368504251394584.

[97] Oliveira, M.B., Bernardino, H.S., Vieira, A.B., Augusto, D.A., 2025. Classification of animal species via deep neural networks and species distribution modeling: a systematic review. Artificial Intelligence Review 58, 230.

[98] Paolo, F., Lin, T.t.T., Gupta, R., Goodman, B., Patel, N., Kuster, D., Kroodsma, D., Dunnmon, J., 2022. xview3-sar: Detecting dark fishing activity using synthetic aperture radar imagery. Advances in Neural Information Processing Systems 35, 37604–37616.

[99] Parraga, O., More, M.D., Oliveira, C.M., Gavenski, N.S., Kupssinskü, L.S., Medronha, A., Moura, L.V., Simões, G.S., Barros, R.C., 2025. Fairness in deep learning: A survey on vision and language research. ACM Computing Surveys 57, 1–40.

[100] Pirotta, V., Meagher, P., Phan, H.T.H., Shen, K., Thompson, W., Dolman, B., Bending, Z., Smoothey, A.F., Hay, A., Rowe, C.E., et al., 2026. Marine wildlife trafficking: Use of ai algorithms for the autodetection of shark fins, seahorses and sea cucumbers. Frontiers in Ocean Sustainability 4, 1776978.

[101] Pourpanah, F., Abdar, M., Luo, Y., Zhou, X., Wang, R., Lim, C.P., Wang, X.Z., Wu, Q.J., 2022. A review of generalized zero-shot learning methods. IEEE transactions on pattern analysis and machine intelligence 45, 4051–4070.

[102] Prasetyo, A.P., Cusa, M., Murray, J.M., Agung, F., Muttaqin, E., Mariani, S., McDevitt, A.D., 2023. Universal closed-tube barcoding for monitoring the shark and ray trade in megadiverse conservation hotspots. Iscience 26.

[103] Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al., 2021. Learning transferable visual models from natural language supervision, in: International conference on machine learning, PmLR. pp. 8748–8763.

[104] Ramadan, M.N., Ali, M.A., Khoo, S.Y., Alkhedher, M., 2024. Ai-powered iot and uav systems for real-time detection and prevention of illegal logging. Results in Engineering 24, 103277.

[105] Redmon, J., Farhadi, A., 2018. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767 .

[106] Ren, J., Tang, T., Jia, H., Xu, Z., Fayek, H., Li, X., Ma, S., Xu, X., Xia, F., 2025. Foundation models for anomaly detection: Vision and challenges. AI Magazine 46, e70045.

[107] Ren, S., He, K., Girshick, R., Sun, J., 2016. Faster r-cnn: Towards real-time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence 39, 1137–1149.

[108] Resende, P.A.A., Drummond, A.C., 2018. A survey of random forest based methods for intrusion detection systems. ACM Computing Surveys (CSUR) 51, 1–36.

[109] Robinson, D., Miron, M., Hagiwara, M., Pietquin, O., 2025. NatureLM-audio: an audio-language foundation model for bioacoustics, in: The Thirteenth International Conference on Learning Representations (ICLR).

[110] Sanghera, A., Singh, P., Chu, E., Leem, S., Li, R., Ghattan, S., Tai, A.M.Y., 2026. Semi-supervised vessel trajectory analysis for unregulated fishing activity detection. Engineering Applications of Artificial Intelligence 169, 113933.

[111] Santos, C.F.G.d., Arrais, R.R., Silva, J.V.S.d., Silva, M.H.M.d., Neto, W.B.G.d.A., Lopes, L.T., Bileki, G.A., Lima, I.O., Rondon, L.B., Souza, B.M.d., et al., 2025. Isp meets deep learning: a survey on deep learning methods for image signal processing. ACM Computing Surveys 57, 1–44.

[112] Sharma, K., Barbosa, J.S., Roberts, S., Gondhali, U., Petrossian, G., Jacquet, J., Freire, J., Chakraborty, S., 2025. Descriptive analysis of online wildlife products using vision language models, in: Proceedings of the ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies, pp. 461–472.

[113] Simões, F., Bouveyron, C., Precioso, F., 2023. Deepwild: Wildlife identification, localisation and estimation on camera trap videos using deep learning. Ecological Informatics 75, 102095.

[114] Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition, in: ICLR.

[115] Stefanakis, N., Psaroulakis, K., Simou, N., Astaras, C., 2022. An open-access system for long-range chainsaw sound detection, in: 2022 30th European Signal Processing Conference (EUSIPCO), IEEE. pp. 264–268.

[116] Stringham, O.C., García-Díaz, P., Toomes, A., Mitchell, L., Ross, J.V., Cassey, P., 2021. Live reptile smuggling is predicted by trends in the legal exotic pet trade. Conservation Letters 14, e12833.

[117] Stringham, O.C., Maher, J., Lassaline, C.R., Wood, L., Moncayo, S., Toomes, A., Heinrich, S., Watters, F., Drake, C., Chekunov, S., et al., 2023. The dark web trades wildlife, but mostly for use as drugs. People and Nature 5, 999–1009.

[118] Su, S., Gu, D., Lai, J.Y., Arcilla, N., Su, T.Y., 2025. A novel deep learning-based bioacoustic approach for identification of look-alike white-eye (zosterops) species traded in wildlife markets. Ibis 167, 41–55. doi:10.1111/ibi.13309.

[119] Sun, J., Zheng, C., Xie, E., Liu, Z., Chu, R., Qiu, J., Xu, J., Ding, M., Li, H., Geng, M., et al., 2025. A survey of reasoning with foundation models: Concepts, methodologies, and outlook. ACM Computing Surveys 57, 1–43.

[120] Sun, R., Ren, Y., 2024. A multi-source heterogeneous data fusion method for intelligent systems in the internet of things. Intelligent systems with applications 23, 200424.

[121] Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A., 2017. Inception-v4, inception-resnet and the impact of residual connections on learning, in: Proceedings of the AAAI conference on artificial intelligence.

[122] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9.

[123] Tan, M., Le, Q., 2021. Efficientnetv2: Smaller models and faster training, in: International conference on machine learning, PMLR. pp. 10096–10106.

[124] Tanveer, M., Rajani, T., Rastogi, R., Shao, Y.H., Ganaie, M., 2024. Comprehensive review on twin support vector machines. Annals of Operations Research 339, 1223–1268.

[125] Tian, Y., Ye, Q., Doermann, D., 2025. Yolov12: Attention-centric real-time object detectors. arXiv preprint arXiv:2502.12524 .

[126] 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., 2022. Perspectives in machine learning for wildlife conservation. Nature communications 13, 792.

[127] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Advances in neural information processing systems 30.

[128] Villa, A.G., Salazar, A., Vargas, F., 2017. Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks. Ecological informatics 41, 24–32.

[129] Von Essen, E., Turnbull, J., Searle, A., Jørgensen, F.A., Hofmeester, T.R., van der Wal, R., 2023. Wildlife in the digital anthropocene: Examining human-animal relations through surveillance technologies. Environment and Planning E: Nature and Space 6, 679–699.

[130] Vu, T.T., Le, T.S., Clink, D.J., et al., 2024. Investigating hunting in a protected area in southeast asia using passive acoustic monitoring with mobile smartphones and deep learning. Ecological Indicators 167, 112501.

[131] Wang, K., Zhang, D., Li, Y., Zhang, R., Lin, L., 2016. Cost-effective active learning for deep image classification. IEEE Transactions on Circuits and Systems for Video Technology 27, 2591–2600.

[132] Wang, T., Zuo, Y., Manda, T., Hwarari, D., Yang, L., 2025. Harnessing artificial intelligence, machine learning and deep learning for sustainable forestry management and conservation: Transformative potential and future perspectives. Plants 14, 998.

[133] Wang, X., Wang, S., Liang, X., Zhao, D., Huang, J., Xu, X., Dai, B., Miao, Q., 2022. Deep reinforcement learning: A survey. IEEE Transactions on Neural Networks and Learning Systems 35, 5064–5078.

[134] Wang, Y., Shi, Z.R., Yu, L., Wu, Y., Singh, R., Joppa, L., Fang, F., 2019. Deep reinforcement learning for green security games with real-time information, in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 1401–1408.

[135] Wang, Y., Wei, C., Zhang, C., 2026. A generative deep learning framework for illegal ivory provenance via diffusion-based isotopic data augmentation. Microchemical Journal , 118344.

[136] Wang, Y., Yao, Q., Kwok, J.T., Ni, L.M., 2020. Generalizing from a few examples: A survey on few-shot learning. ACM computing surveys (csur) 53, 1–34.

[137] Wu, B., Ding, Z., Huang, J., 2026a. A review of continual learning in edge ai. IEEE Transactions on Network Science and Engineering .

[138] Wu, T., Xu, Q., He, X., Weekes, O., Brown, J., Duan, W., 2026b. Rolid-11k: A dashcam dataset for small-object roadside litter detection. arXiv preprint arXiv:2601.00398 .

[139] Xia, F., Peng, C., Ren, J., Febrinanto, F.G., Luo, R., Saikrishna, V., Yu, S., Kong, X., 2026. Graph learning. Foundations and Trends® in Signal Processing 19, 362–519.

[140] Xiao, W., Ye, P., 2025. Unveiling wildlife crime characteristics in china based on judicial big data. Discover Sustainability 6, 855.

[141] Xu, H., Ford, B., Fang, F., Dilkina, B., Plumptre, A., Tambe, M., Driciru, M., Wanyama, F., Rwetsiba, A., Nsubaga, M., et al., 2017. Optimal patrol planning for green security games with blackbox attackers, in: International Conference on Decision and Game Theory for Security, Springer. pp. 458–477.

[142] Yan, K., Dong, Y., Yang, Y., Xing, L., 2022. Multisuav collaboration and low-altitude remote sensing technology-based image registration and change detection network of garbage scattered areas in nature reserves. Remote Sensing 14, 6352.

[143] Yang, X., Song, Z., King, I., Xu, Z., 2022. A survey on deep semi-supervised learning. IEEE transactions on knowledge and data engineering 35, 8934–8954.

[144] You, L., Coyotl, E.P., Gunturu, S., Van Segbroeck, M., 2023. Transformer-based bioacoustic sound event detection on few-shot learning tasks, in: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE. pp. 1–5.

[145] Yu, Y., Si, X., Hu, C., Zhang, J., 2019. A review of recurrent neural networks: Lstm cells and network architectures. Neural computation 31, 1235–1270.

[146] Zhang, S., Ma, J., Jiang, F., 2025. Monitoring street-level improper dumpsites via a multi-modal and llm-based framework. Resources, Conservation and Recycling 218, 108227.

[147] Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y., Chen, J., 2024. Detrs beat yolos on real-time object detection, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 16965–16974.

[148] Zheng, C., Liu, S., Wang, J., Lu, Y., Ma, L., Jiao, L., Guo, J., Yin, Y., He, T., 2024. Opening the black box: explainable deep-learning classification of wood microscopic image of endangered tree species. Plant Methods 20, 56.

[149] Zhou, W., Zhou, J., Wang, H., Li, Z., He, Q., Han, S., Li, G., Zhou, X., He, Y., Liu, C., et al., 2026. Can llms clean up your mess? a survey of application-ready data preparation with llms. arXiv preprint arXiv:2601.17058 .

[150] Zielinski, K.M., Scabini, L., Ribas, L.C., da Silva, N.R., Beeckman, H., Verwaeren, J., Bruno, O.M., De Baets, B., 2025. Advanced wood species identification based on multiple anatomical sections and using deep feature transfer and fusion. Computers and Electronics in Agriculture 231, 109867.

[151] Zong, Z., Song, G., Liu, Y., 2023. Detrs with collaborative hybrid assignments training, in: Proceedings of the IEEE/CVF international conference on computer vision, pp. 6748–6758.