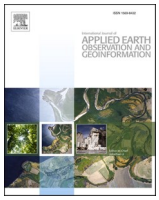
International Journal of Applied Earth Observation and Geoinformation 128 (2024) 103732

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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



A review of deep learning techniques for detecting animals in aerial and satellite images

Zeyu Xu *, Tiejun Wang *, Andrew K. Skidmore, Richard Lamprey

Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, P.O. Box 217, 7500 AE Enschede, the Netherlands

ARTICLE INFO

Keywords:
Biodiversity
Wildlife
Livestock
Remote sensing
Artificial intelligence
Object detection

ABSTRACT

Deep learning is an effective machine learning method that in recent years has been successfully applied to detect and monitor species population in remotely sensed data. This study aims to provide a systematic literature review of current applications of deep learning methods for animal detection in aerial and satellite images. We categorized methods in collated publications into image level, point level, bounding-box level, instance segmentation level, and specific information level. The statistical results show that YOLO, Faster R-CNN, U-Net and ResNet are the most used neural network structures. The main challenges associated with the use of these deep learning methods are imbalanced datasets, small samples, small objects, image annotation methods, image background, animal counting, model accuracy assessment, and uncertainty estimation. We explored possible solutions include the selection of sample annotation methods, optimizing positive or negative samples, using weakly and self-supervised learning methods, selecting or developing more suitable network structures. Future research trends we identified are video-based detection, very high-resolution satellite image-based detection, multiple species detection, new annotation methods, and the development of specialized network structures and large foundation models. We discussed existing research attempts as well as personal perspectives on these possible solutions and future trends.

1. Introduction

Biodiversity is declining worldwide at an accelerating pace, with multiple negative impacts on a good quality of life (Pörtner et al., 2021). The latest World Wildlife Fund (WWF) Living Planet Report showed that the populations of mammals, birds, amphibians, reptiles, and fish declined by an average of 68 % between 1970 and 2016 (Almond et al., 2020). However, monitoring global changes in biodiversity remains a major challenge because of the general lack of agreed international data standards and evaluation criteria (Turak et al., 2017).

A recent important step in biodiversity monitoring is the Essential Biodiversity Variables (EBV) framework, developed by the Group on Earth Biodiversity Observation Network (GEO BON) with the aim of distilling the complexity of biodiversity into a manageable list of priority measurements (Pereira et al., 2013). The species population is one of the most fundamental and priority variables to be monitored, as recognized by the EBV framework and supported by subsequent studies (Skidmore et al., 2015; Brummitt et al., 2017; McRae et al., 2017; Jetz et al., 2019; Skidmore et al., 2021).

Remote sensing has proven to be an important approach for monitoring species populations (Leyequien et al., 2007; Hollings et al., 2018; Skidmore et al., 2021). It has been demonstrated that some remote sensing platforms like unmanned aerial vehicles (UAVs, also called drones) can provide higher quality and more precise data than traditional ground counts (Hodgson et al., 2016). The remote sensing platforms currently used for animal detection are aircraft and satellites. Aerial enumeration is a classic animal detection method, which has been routinely conducted in many parts of the world to estimate the abundance of species and the rate of population growth (Jolly, 1969; Norton-Griffiths, 1978; Firchow et al., 1990; Mbugua, 1996; Chabot, 2009; Lamprey et al., 2020b; Delplanque et al., 2023b). In recent years, with the improvement of spatial resolution, satellite imagery has been increasingly used in animal detection (Yang et al., 2014; Xue et al., 2017; Cubaynes et al., 2019; Goncalves et al., 2020).

In animal detection based on remote sensing imagery, manual or visual detection (sometimes called “visual interpretation” or “ocular” detection) is still a commonly used method (Stapleton et al., 2014; Hodgson et al., 2016; Linchant et al., 2018; Bowler et al., 2020; Lamprey

* Corresponding authors.
E-mail addresses: z.xu-1@utwente.nl (Z. Xu), t.wang@utwente.nl (T. Wang).

https://doi.org/10.1016/j.jag.2024.103732
Received 11 October 2023; Received in revised form 16 February 2024; Accepted 20 February 2024
Available online 2 March 2024

et al., 2020b). These methods are costly and susceptible to subjectivity (Bowler et al., 2020). Automatic methods such as threshold methods, image differencing (LaRue et al., 2015), edge extraction algorithms (Lee et al., 2021), and traditional machine learning methods like Support Vector Machines (SVM, Cortes and Vapnik, 1995; van Gemert et al., 2015) and wavelet transforms (Farge, 1992; Bentley and McDonnell, 1994) have also been used to detect animals in remotely sensed images. However, their accuracy is unsatisfactory and still needs improvement (Goncalves et al., 2020).

Deep learning, a branch of Artificial Neural Networks (ANNs), has developed rapidly in recent years. Deep learning networks have multiple levels of data representation organized by abstractions, features, or concepts, while traditional ANN models typically have only one or two hidden layers (Oliveira et al., 2016). The key aspect of deep learning is that the layers of features are learned from data via a general-purpose learning procedure, not designed by human engineers, so it can easily take advantage of increases in the amount of available computation and data (LeCun et al., 2015). The aforementioned authors note that deep learning has significant advantages over traditional machine learning methods thanks to its efficient use of Graphics Processing Units (GPUs), activation functions like Rectified Linear Unit (RELU, Glorot et al., 2011), and regularization techniques like dropout (Srivastava et al., 2014); it has since been further improved with data augmentation methods like mixup (Zhang et al., 2018) and mosaic (Bochkovskiy et al., 2020).

Deep learning has achieved good results in image processing fields such as object detection (Ren et al., 2017), semantic segmentation (Ronneberger et al., 2015) and instance segmentation (He et al., 2017), and its extraction accuracy is superior to that of other machine learning algorithms (Osco et al., 2021). It enables the detection of animals occupying three pixels in remote sensing images (Wu et al., 2023) and can also yield specific information such as the posture and body length of animals in higher spatial resolution images (Mücher et al., 2022). Deep learning has been able to surpass manual detection results of citizen scientists in certain animal detection tasks (Torney et al., 2019). Compared to manual detection, one of the advantages of deep learning is that it is insensitive to interference factors such as shooting distance, shooting angle, and changes in animal body size (Eikelboom et al., 2019).

To our knowledge there has so far been no comprehensive review of animal detection based on deep learning and remote sensing. Wang et al., (2019a) reviewed studies on wildlife detection, with a particular emphasis on different remote sensing platforms. Tuia et al. (2022) wrote a perspective article on machine learning for wildlife conservation, aiming to link ecology and machine learning to showcase how relevant technological advances can be leveraged to address urgent animal conservation challenges. But their paper did not focus on deep learning methods. Yousefi et al. (2022) studied the application of deep learning and UAV platforms in livestock detection, but included some very low altitude side-view photos of animals and did not consider other remote sensing platforms like manned aircraft and satellites. In this review paper, we aim to provide a systematic literature review of current applications of deep learning methods for animal detection in aerial and satellite images. Specifically, we seek to answer the following questions: 1) which deep learning algorithms are most commonly used for detecting animals in aerial and satellite images? 2) what are the main challenges associated with the use of these algorithms? 3) What are the potential solutions to these challenges? and 4) what is the future research direction in this field?

2. Methodology

The main method we used to search for papers was to search by keywords, using the search string: TS = ((animal* OR wild animal* OR wildlife* OR livestock* OR mammals* OR bird* OR amphibians* OR reptile* OR fish* OR insect* OR animal) AND (satellite* OR drone* OR UAV* OR MAV* OR aircraft* OR aerial*) AND (deep learning* OR machine learning* OR artificial intelligence* OR CNN* OR convolutional neural network*)). We mined four databases: Web of Science (https://www.webofscience.com/wos), SCOPUS (https://www.scopus.com), Google Scholar (https://scholar.google.com), and the preprint platform ArXiv (https://arxiv.org). We included ArXiv because some influential research related to deep learning like YOLOv3 (Redmon and Farhadi, 2018) is published exclusively on this non peer-reviewed platform.

We searched the references of the publications we found and screened for relevant papers. As studies used public datasets, we also searched for papers based on public data from Zenodo (https://zenodo.org/) and Kaggle (https://www.kaggle.com/datasets) and included papers that we deemed to be relevant.

3. The collated papers

The cut-off date for our search was December 31, 2023. We found a total of 98 papers (excluding reviews). Fig. 1 shows the number of publications on animal detection based on deep learning and remote sensing; their number has increased greatly since 2017.

The worldwide distribution of the species studied is shown in Fig. 2. The research area covers seven continents; most studies were in Africa. The African animals studied are mainly wild animals on the savannahs, while the Oceania animal research focusses primarily on farm livestock and marine animals. Europe and South America are also major research areas for livestock detection.

4. Remote sensing platform

The primary remote sensing platforms utilized for animal detection are aircraft and satellites. The use of different platforms in each year's publications is shown in Fig. 3. The figure shows that studies using images acquired by UAVs were steadily increased over the years.

4.1. Aircraft

4.1.1. Unmanned aerial vehicles (UAVs)

UAVs have the advantages of miniaturization, high maneuverability, and the ability to perform more flexible tasks than manned aircraft and satellites; their disadvantages, particularly for large area wildlife enumeration, include low endurance and range, slow flight speeds and security issues. The flight altitude of UAVs can be changed in time to avoid clouds and obstacles, and the flight plan can also be adjusted to deal with the actual situation of animal activities (Naudé and Joubert, 2019). Currently, most images captured by UAVs in the collated papers are colour images composed of red, green, and blue bands (72 out of 77).

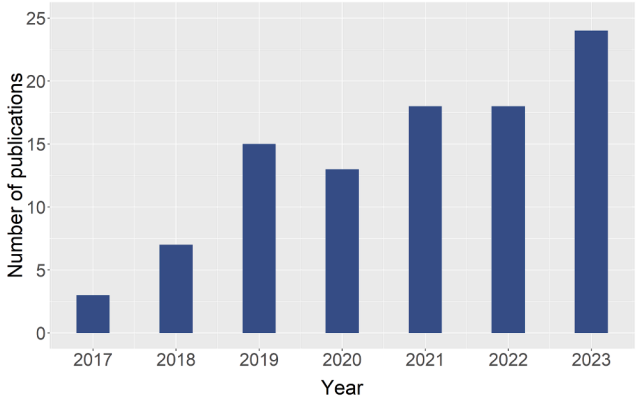


Fig. 1. Number of publications on animal detection based on deep learning and remote sensing from 2017 to 2023.

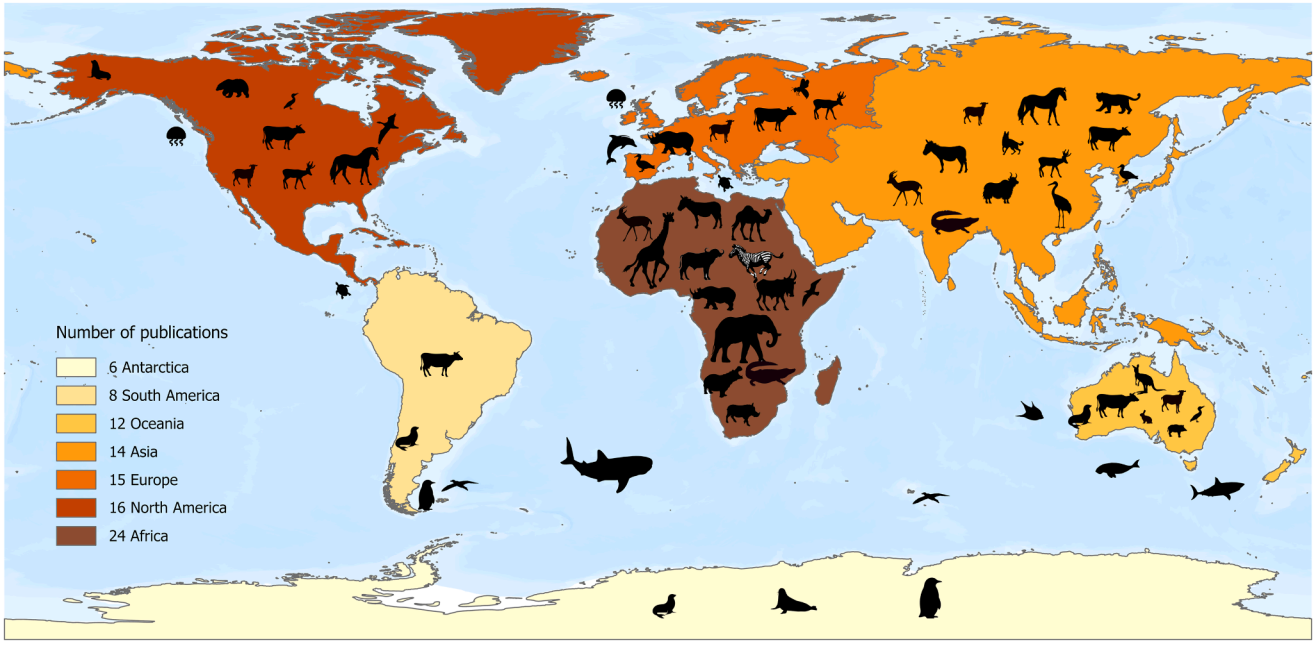


Fig. 2. Species studied in publications on animal detection based on deep learning and remote sensing from 2017 to 2023, indicating the global geographic distribution of the animals in studies. The position of species within the same continent is not distinguished. Details on the species and on the precise location of the studies can be found in Appendix 2.

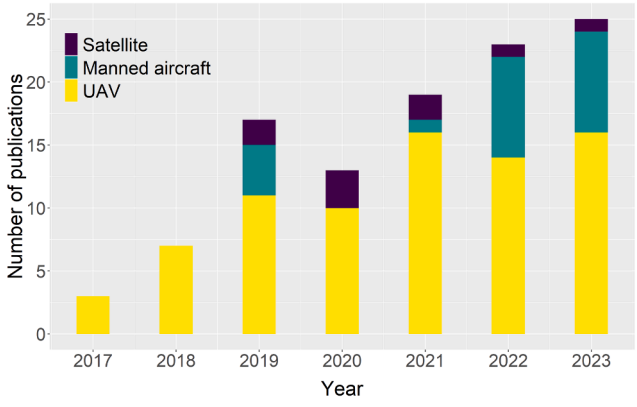


Fig. 3. Number of publications using different remote sensing platforms from 2017 to 2023.

Eight papers used thermal infrared imagery, accounting for 10.4 % of all studies using UAVs (Bondi et al., 2020; Ulhaq et al., 2021; Hinke et al., 2022; Chen et al., 2023a; Barrios et al., 2023; Krishnan et al., 2023; Xie et al., 2023; Zhang and Cai, 2023). In the papers we collated, other remote sensing platforms have not yet utilized thermal imagery. In terms of flight modes, most UAVs currently used in animal detection are multi-rotor UAVs and fixed-wing UAVs.

Recently, with the rapid development of consumer-grade UAVs, animal detection based on multi-rotor UAVs has been extensively studied. Commonly used multi-rotor UAV platforms include the DJI Phantom and Mavic series (Andrew et al., 2020; Barbedo et al., 2020a; Kellenberger et al., 2021; Petso et al., 2021; Sarwar et al., 2021; Shao et al., 2020). In the studies using multi-rotor UAVs, the reported spatial resolution could attain 1 cm. Some studies did not record spatial resolution, especially those with low flight altitudes with presumed higher spatial resolution (Andrew et al., 2021; Sundarama and Loganathan, 2020).

The ability of multi-rotor UAVs to hover allows the influence of flying height to be studied. Sarwar et al. (2021) set two heights (80 m and 120 m) to study the detection of sheep. Petso et al. (2021) set eight heights of 15 m, 20 m, 30 m, 40 m, 50 m, 70 m, 90 m, 110 m, and 130 m for livestock detection and found that an increase in altitude to minimize animal disturbance reduces detection ability. In addition, the optimal deep learning input resolution can also be determined by analyzing the structure and parameters of the neural network as the multi-rotor UAVs can ensure a stable flight height (Shao et al., 2020).

Fixed-wing UAVs are another commonly used UAV platform. They are often capable of covering a larger research area than multi-rotor UAVs. For instance, in a wildlife survey, over 600 acres were covered using the SenseFly eBee fixed-wing UAV (Shao et al., 2020). In terms of image quality control, fixed-wing UAVs are more prone to problems like image blur and sensor artifacts (Naudé and Joubert, 2019). SAVMAP is a widely used fixed-wing UAV wildlife dataset (Ofli et al., 2016) based on five flight missions over Kuzikus Park in eastern Namibia using a light monoplane UAV (SenseFly eBee).

4.1.2. *Manned aircraft*

Manned aircraft are also commonly used to detect animals by acquiring images vertically (Couturier et al., 1996; Norton-Griffiths, 1973) or obliquely (Lamprey et al., 2020b). In some cases, the flight altitude and the quality of images obtained by manned aircraft are similar to those of fixed-wing UAVs. However, manned aircraft have a higher altitude limit than UAVs: the highest flying-height mentioned in the 98 papers was over 500 m for manned aircraft (Naudé and Joubert, 2019), and 350 m for UAV (Ma et al., 2022; Peng et al., 2020). Currently, all research on manned aircraft images used Red-Green-Blue (RGB) three-band images according to the papers we collated.

The datasets for manned aircraft often come from wildlife surveys. The most used public dataset is the Aerial Elephant Dataset (AED, Naudé and Joubert, 2019). It is a collection of 2,101 aerial images of elephants in south central Africa. The images contain a total of 15,511 African elephants in different environmental backgrounds with spatial resolution from 2.4 cm to 13 cm. Table 1 lists the currently publicly available animal detection datasets and describes the relevant properties of the datasets.

Table 1
Description of publicly available datasets.

Dataset names (ref)	Species	Locations	Remote sensing platforms	Number of objects	URL
SAVMAP (Ofli et al., 2016)	multiple species (e.g., black rhino, zebra)	Namibia	UAV	1,183	https://zenodo.org/records/1204408
BIRDSAI (Bondi et al., 2020)	animal (e.g., elephant, lion, giraffe, crocodile)	South Africa, Malawi, Zimbabwe	UAV	220,000	https://sites.google.com/view/elizabethbondi/dataset
AnimalDrone (Zhu et al., 2021)	horse, sheep, zebra, giraffe, wolf, cow, yak, dog, antelope, boar	A: China B: web	UAV	4,049,168 (video)	https://github.com/VisDrone/AnimalDrone
WAID (Mou et al., 2023)	sheep, cattle, seal, camel, kiang, zebra	web	UAV	14,275	https://github.com/xiaohuicui/WAID
(Shao et al., 2020)	cattle	Japan	UAV	1,948	http://bird.nae-lab.org/cattle/
(Han et al., 2019)	livestock		UAV	4,996	https://github.com/hanl2010/Aerial-livestock-dataset/releases
(Gray et al., 2019b)	sea turtle	Costa Rica	UAV	2,161	https://doi.org/10.5061/dryad.5h06vv2
(Desai et al., 2022)	crocodile	India	UAV	480	https://doi.org/10.5061/dryad.s4mw6m98n
Aerial Seabirds West Africa (Kellenberger et al., 2021)	seabird (African royal tern, Caspian tern, slender-billed gull, grey-headed gull)	West African coast	UAV	21,516	https://lila.science/datasets/aerial-seabirds-west-africa/
(Hinke et al., 2022)	penguin, fur seal	Antarctica	UAV	6,314	https://zenodo.org/records/6714100
(Hayes et al., 2021)	penguin, albatross	Falkland (Malvinas) Islands, England	UAV	44,970	https://research.repository.duke.edu/concern/datasets/kp78gh20s?locale
BEE4EXP (Stojnić et al., 2021)	honeybee	Croatia	UAV	576 (video)	https://zenodo.org/records/7253878
(Weinstein et al., 2022)	bird	worldwide	UAV	> 250,000	https://doi.org/10.5281/zenodo.5033174
SheepCounter (Doll and Loos, 2023)	sheep		UAV	209,943	https://universe.roboflow.com/riisprivate/sheepcounter
ISOD, (Zhang and Cai, 2023)	elephant	South Africa, Malawi, Zimbabwe	UAV	22,837	https://zenodo.org/records/10020732
(Krishnan et al., 2023)	deer, cattle, horse	United States	UAV	478	https://projectportal.gri.msstate.edu/awir/
AED (Naudé and Joubert, 2019)	elephant	South Africa, Botswana, Zambezi	Manned aircraft	15,581	https://zenodo.org/records/3234780
(Eikelboom et al., 2019)	zebra, elephant, giraffe	Kenya	Manned aircraft	10,824	https://data.4tu.nl/articles/_/12713903/1
(Chabot et al., 2022)	polar bear	Baffin Bay, Canada	Manned aircraft	21	https://www.sciencedirect.com/science/article/pii/S1574954121003381?via%3Dihub
(Qian et al., 2023)	penguin	Antarctica	Manned aircraft	137,365	https://doi.org/10.5061/dryad.8931zcrv8
(DataCanary et al., 2017)	sea lion	Western Aleutian Islands	Manned aircraft and UAV	83,677	https://www.kaggle.com/competitions/noaa-fisheries-steller-sea-lion-population-count/

4.2. Satellites

The application of satellite data for animal detection primarily uses high-resolution images, with spatial resolution of 30 – 50 cm (Goncalves et al., 2020; Laradji et al., 2020; Bowler et al., 2020; Duporge et al., 2021; Robinson et al., 2021; Mücher et al., 2022). Brown et al. (2022) obtained images of different spatial resolutions through simulation and showed that 50 cm is the resolution threshold required for deep learning to be applied to cattle sized animals. Animals that have been successfully detected by high-resolution satellite imagery include elephants (Duporge et al., 2021), wildebeests (Yang et al., 2014; Wu et al., 2023), albatrosses (Bowler et al., 2020), seals (Goncalves et al., 2020), cattle (Laradji et al., 2020), and whales (Borowicz et al., 2019). However, at present satellite images are only used to identify single species (or classify different species into one category), and no research has yet shown that the images can be used to simultaneously identify different species when using deep learning methods. In the collated papers, most satellite-based studies (6 out of 9) used pansharpened Red-Green-Blue (RGB) bands. Only two studies directly used a single panchromatic band (Goncalves et al., 2020; Robinson et al., 2021), and one study used four bands containing near-infrared (RGB-NIR, Wu et al., 2023). In addition, there are currently no publicly available very high-resolution commercial satellite data in animal detection based on the collated papers.

Compared to aerial imagery, satellite imagery has greater spatial coverage and is ideal for large and remote areas (Hollings et al., 2018; Goncalves et al., 2020). However, high-resolution commercial satellite data is expensive (Bhardwaj et al., 2016), and its spatial resolution is still lower than that of many images obtained from aircraft. In terms of sampling strategies, satellites have fixed flight orbits and image parameters. For instance, WorldView-3 has a swath width of 13.1 km. Its revisit frequency is approximately one day, which ensures consistency in the data parameters and quality but makes it difficult to obtain repeated observation data on specific animal targets because of cloud cover and animal movement. The arrangement for acquiring aerial images from aircraft can be more flexible. The repeated observation area can be set according to each mission, and adjustments can be made, such as altering the flight altitude to take account of cloud cover (Hollings et al., 2018). Timely repeated area detection can provide richer data sources for post-processing. Although the flexible flight mode may result in inconsistent data quality (e.g., different spatial resolutions), corrections can be made based on flight parameters (Naudé and Joubert, 2019).

5. Deep learning methods for animal detection

5.1. Overview of methods

In general, research on animal detection covers several basic fields of image processing based on deep learning: scene recognition (He et al., 2016; Gao et al., 2021), object detection (Redmon et al., 2016), semantic segmentation (Ronneberger et al., 2015) and instance segmentation (He et al., 2017). Image annotation is the process of labelling images for computer vision tasks, which is the key prerequisite for successful deep learning applications. We therefore classified the current deep learning-

based animal detection studies into five categories: image level, point level, bounding box level, instance segmentation level, and specific information level based on the image annotation method and the degree of richness of the animal information obtained. These are shown diagrammatically in Fig. 4. Image level refers to judging whether an image contains the target animals. The images here can also be sub-images in a larger image. Point level refers to using points to mark the location of animals. The point here is usually the geometric center of the animal (Padubidri et al., 2021; Sarwar et al., 2021). In addition, point level also includes research using density maps generated from points. Bounding box level refers to studies that use rectangular boxes to mark animals. Instance segmentation refers to identifying and differentiating individual objects within an image, usually obtaining the precise boundaries while using bounding boxes. Specific information level refers to studies that can extract more information about animals, such as their posture and body length.

We counted the number of publications using methods at different detection levels from 2017 to 2023. From Fig. 5, we observe that the bounding box is the most commonly used detection level, which has generally increased over the years compared to the use of other detection levels.

We have detailed the dataset parameters, deep learning models, specific tasks and other information used in each study at each detection level in Appendix 1. To provide a more detailed visualization of the current use of deep learning algorithms in the collated papers, we synthesized spatial resolution and species information, and aggregated them together with deep learning networks in Fig. 6. Backbone CNN in this figure indicates studies that used or improved the base models such as ResNet and VGG. The information presented on spatial resolution is incomplete because it was not always explicitly stated in the collated papers.

Fig. 6 shows that the most used models are YOLO, Faster R-CNN, U-Net, and Backbone CNN (55.6 % used ResNet) and that current studies are mainly based on point level and bounding box level. The same neural network model can be applied at different detection levels, but at each level there are some dominant neural networks. At the bounding box level, over half of the studies used YOLO and Faster R-CNN networks. At the point level, U-Net and Backbone CNN dominate, and some other networks (e.g., fully convolutional network, FCN) have similar functions to U-Net. In the next section, the specific methods used at each level will be introduced and analyzed.

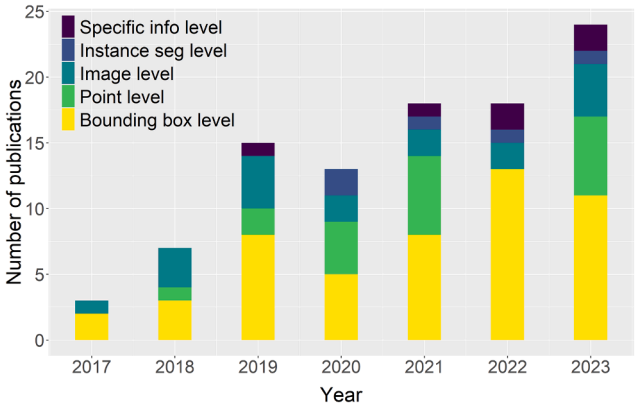


Fig. 5. Number of publications using deep learning methods at different detection levels from 2017 to 2023.

Research on animal detection can directly use existing deep learning methods, or it can improve or create new methods. Based on this, we divided the existing research into three categories: standard models, modified models and new models. The results are shown in Fig. 7. Most of the current research uses off-the-shelf computer vision models. The popular deep learning frameworks TensorFlow and PyTorch have both released their own object detection APIs and have been used in animal detection research (Duporge et al., 2021; Kabra et al., 2022). However,

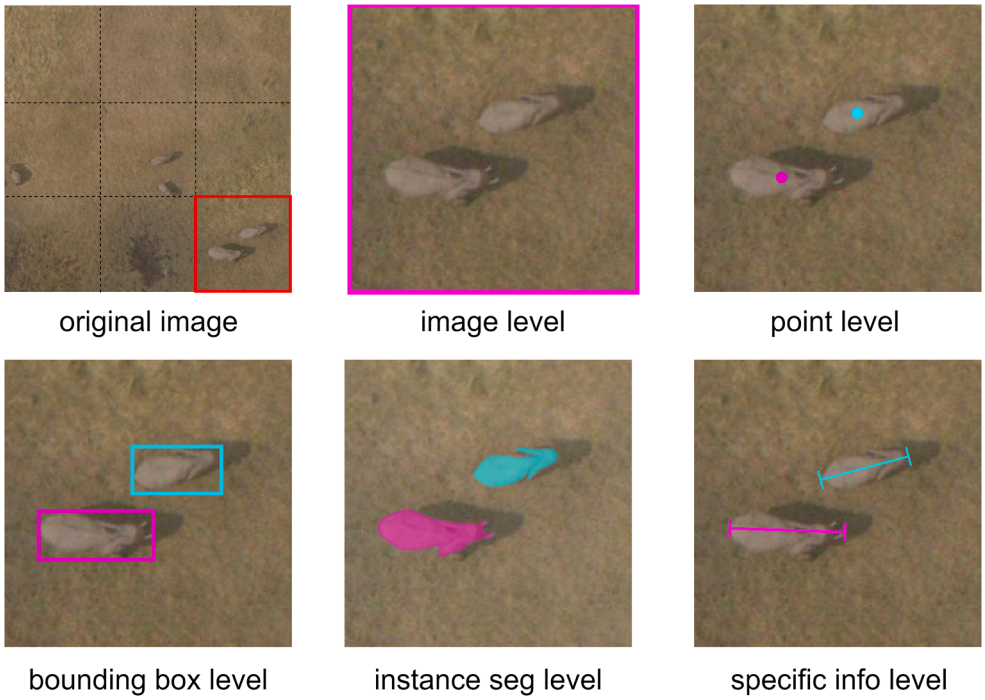


Fig. 4. The categorization of animal detection at different levels according to the image annotation method and the degree of richness of the animal information obtained. Because of limitations (the memory limit of the graphics card) the original image must be split into several sub-images. The image level directly identifies whether an animal is contained in a sub-image, but it does not yield the precise location. The point level, bounding box level, and instance seg[mentation] level mark the animal's location by different methods. The image above illustrating the specific info[rmation] level depicts extracting the length of the elephants, showing that more detailed information can be detected at this level.

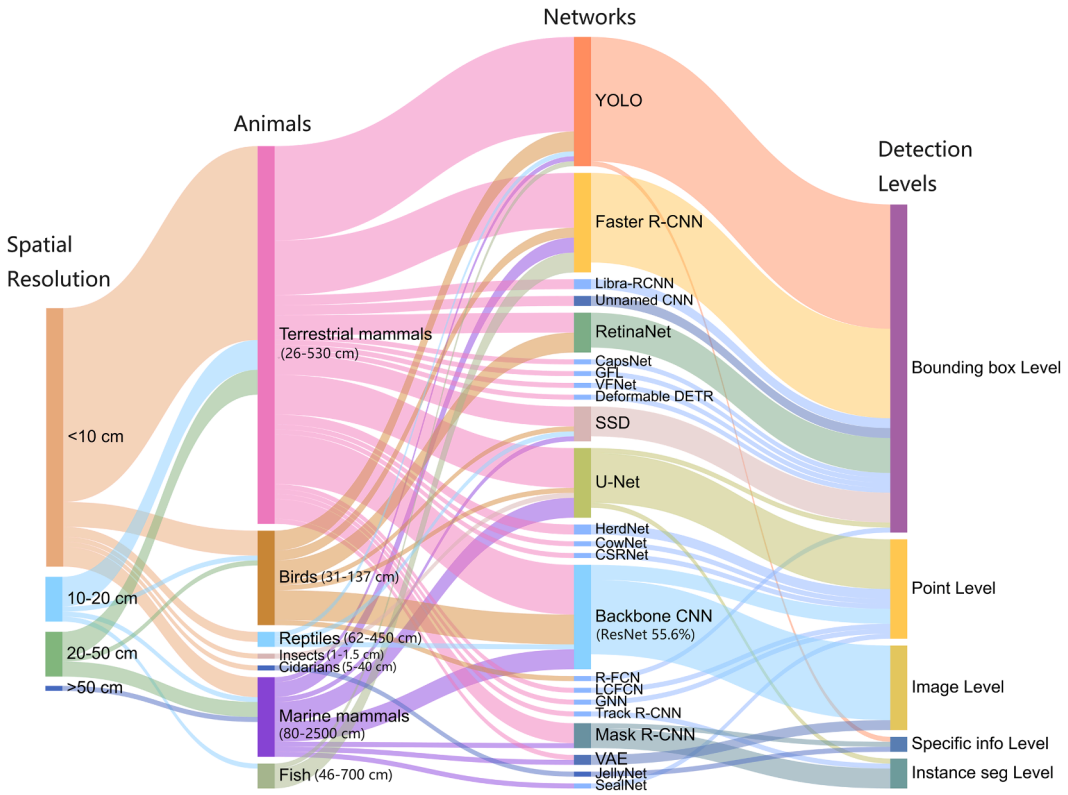


Fig. 6. An overview of the deep learning methods used in animal detection from 2017 to 2023. The first column is the spatial resolution of the image. The second column indicates the types of animals studied. The third column is the neural network structure used in the study. Backbone CNN refers to the use or adjustment of basic networks like VGG and ResNet. These networks can be used independently for image recognition, or they can serve as the backbone of other networks. The last column shows the different detection levels.

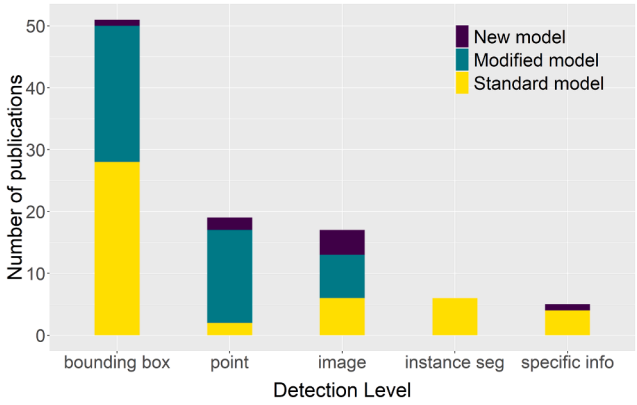


Fig. 7. Number of ways to use deep learning algorithms at different detection levels. 'Standard model' means using standard computer vision model, 'Modified model' means adapting the models to the animal detection tasks, 'New model' means designing new model specifically tailored to the task.

at the point level, more research uses new or modified deep learning models. Table 2 shows the current studies that have shared their code, which refers to the code created by the authors themselves. Research that directly uses open-source code or software is not listed in the table.

5.2. Deep learning methods used at different detection levels

5.2.1. Image level

Image level refers to directly identifying whether an image contains the animal objects without obtaining the location of the animals. In practical applications, it is common to divide the original image into sub-images (also called blocks) and then determine whether each sub-image contains the animal objects, e.g., "cattle" or "non-cattle" (Barbedo et al., 2019). Fig. 8 uses elephant detection as an example to illustrate the principle of the method. One of the most common ways to obtain sub-images is through a sliding window. By identifying the sub-image information acquired by the sliding window, the position of the animal objects in the image can also be roughly located (Dujon et al., 2021; Rivas et al., 2018). Although the image level can only roughly locate animal objects, the methods at this level are very fast in sample generation and model execution, so they are also widely used. The image level method is also usually used to pre-select the area containing animals and then accurately locate the animal positions using subsequent methods, e.g., traditional image processing methods like quadrant distinction and threshold methods (Barbedo et al., 2020b) and deep learning methods (Rahnemoonfar et al., 2019).

5.2.2. Point level

Point level represents the locations of the animals by points, usually the centroids. Since a single point is usually not directly used for network training, most studies expand points into density maps or other forms before training or within the deep learning algorithms.

There are currently three methods for point expansion. (1) Directly expanding circular or square areas centered on the points, and the area are assigned the same pixel values (Bowler et al., 2020; Wu et al., 2023). (2) Using algorithms like Gaussian kernel to obtain the density maps (Goncalves et al., 2020). A density map is a way to show where points or lines may be concentrated in each area and is also called intensity heatmap (Goncalves et al., 2020), intensity map (Laradji et al., 2020; Padubidri et al., 2021) or confidence map. (3) Generating the border class around the points (Kellenberger et al., 2018).

There are also different ways to regress the results to points. (1) After

Table 2
A list of the publicly accessible source code that was created and shared by authors.

Species	Deep learning models	Detection level	URL	Reference
whale	ResNet, DenseNet	image	https://zenodo.org/records/3356970	(Borowicz et al., 2019)
sea turtle	backbone CNN	image	https://zenodo.org/records/1973808	(Gray et al., 2019b)
livestock	VAE-GRF	image	https://github.com/HGangloff/vae_grf	(Gangloff et al., 2023)
snow goose	backbone CNN	point	https://github.com/Connor-Bowley/neuralNetwork	(Bowley et al., 2018)
cattle	CSRNet, LCFCN	point	https://github.com/IssamLaradji/cownter_strike	(Laradji et al., 2020)
seal	SealNet (derived from U-Net)	point	https://github.com/iceberg-project/Seals/tree/paper/SealNet_code	(Goncalves et al., 2020)
honeybee	U-Net like	point	https://github.com/vladan-stojnic/Detection-of-Small-Flying-Objects-in-UAV-Videos	(Stojnić et al., 2021)
donkey, camel, sheep, goat	HerdNet (derived from CenterNet)	point	https://github.com/Alexandre-Delplanque/HerdNet	(Delplanque et al., 2023a)
penguin	VGG-19	point	https://doi.org/10.5061/dryad.8931zcrv8	(Qian et al., 2023)
wildebeest	U-Net	point	https://github.com/zijing-w/Wildebeest-UNet	(Wu et al., 2023)
bird	Retinanet (Resnet 50 as backbone)	bounding box	https://zenodo.org/records/5156926	(Barbedo et al., 2019)
wildebeest	modified YOLOv3	bounding box	https://zenodo.org/records/2562058	(Torney et al., 2019)
elephant, animal form SAVMAP	YOLO v3 with super-resolution and altitude-augmented module	bounding box	https://github.com/Mowen111/SALT	(Xue et al., 2022)
crocodile	Inception v3, YOLO v5	bounding box	https://doi.org/10.5061/dryad.s4mw6m98n	(Desai et al., 2022)

thresholding the final feature maps, the centers of the resulting connected regions can be considered as the target points (Naudé and Joubert, 2019). (2) Obtaining the target points through clustering algorithms such as K-Means (Wu et al., 2023), local maximum algorithms (Delplanque et al., 2023a), Mean Shift and Gaussian Mixture Model (Sarwar et al., 2021; Wu et al., 2023). The regression can be included in the deep learning training process and the network then directly outputs the point coordinates of the animals, enabling the loss function to be calculated from the point information (Sarwar et al., 2021). It is also possible to let the neural network generate density maps or classification maps, and then further regress the maps to point coordinates (Bowler et al., 2020, an example is shown in Fig. 9).

Methods at point level are commonly used because in many cases the animal detection task only requires the animal's location to be ascertained and does not need more complex information (e.g., the animal's contours). Methods at this level are mainly based on backbone networks or fully convolutional networks for semantic segmentation (the pixel-level classification, Ronneberger et al., 2015). The most network structure used most in point level detection is U-Net. It is a classical semantic segmentation model that was developed in medical science but yields good results for remote sensing images (Naudé and Joubert, 2019; Padubidri et al., 2021; Sarwar et al., 2021).

Because density maps are commonly used, some researchers refer to the methods they used at point level as density map based. However, density maps are widely used in various convolutional neural networks. For example, CenterNet is also based on density mapping and has been successfully used for bounding box level detection (Zhou et al., 2019). Therefore, in this review we do not consider the use of density maps as a level distinguishing criterion.

5.2.3. Bounding box level

Methods at this level are currently the most widely used. As shown in Fig. 4, the bounding box is a rectangular box that encloses the animal object. Methods at this level can directly correspond to object detection methods in computer vision. The networks can be divided into anchor-based and anchor-free approaches depending on whether an anchor box is predetermined.

Anchor boxes are the predefined rectangular boxes before network training; they can be defined empirically or by clustering the training samples (Redmon and Farhadi, 2018). The most commonly-used anchor-based network structures are YOLO series or Faster R-CNN. YOLO is a one-stage detector that uses a single neural network to directly locate and classify objects. YOLO v1-v3 were developed by Joseph Redmon (Redmon et al., 2016; Redmon and Farhadi, 2017; Redmon and Farhadi, 2018), and v4-v9 (Bochkovskiy et al., 2020; Jocher et al., 2020; Li et al., 2022a; Wang et al., 2023, 2024; Ultralytics., 2023) were developed by other researchers (v6 and subsequent models are anchor-free or offer anchor-free variations). Fig. 10 shows an example of a flowchart for detecting animals using this type of method; the network structure in the figure is a simplified YOLOv3. Faster R-CNN is a widely used two-stage detector, developed from R-CNN (Girshick et al., 2014) and Fast R-CNN (Girshick, 2015). The first stage in this network is to find the possible anchor boxes of the objects, and the second stage is to classify these anchor boxes.

These anchor based network structures have been used to detect animals such as cattle (Brown et al., 2022; Mücher et al., 2022), birds (Hong et al., 2019), elephants (Eikelboom et al., 2019; Duporge et al., 2021), zebras (Eikelboom et al., 2019; Petso et al., 2021), giraffes (Eikelboom et al., 2019; Petso et al., 2021), wildebeest (Petso et al., 2021), and eagle rays (Desgarnier et al., 2022). Fig. 10 shows the process of detecting elephants. There are also studies that refined these networks to make them more suitable for animal detection (Torney et al., 2019; Peng et al., 2020).

The methods at the bounding box level can also be implemented without anchors. For instance, as mentioned in 5.2.2, CenterNet is based on density map. In recent years, the anchor-free method has increasingly attracted attention. It does not require preset anchors, has a high degree of automation, and maintains high detection accuracy. Currently, only a small number of studies on animal detection have utilized anchor-free networks, such as FSSCaps-DetCountNet (Sundarama and Loganathan, 2020) and Deformable DETR (Moreni et al., 2023).

5.2.4. Instance segmentation level

Instance segmentation can be considered as a special kind of image segmentation that involves identifying and separating each individual object (instance) within an image (Fig. 11). Generally, the images must have high spatial resolution to enable the extraction of useful features. The networks are usually modified bounding box level networks. The main method is Mask R-CNN (He et al., 2017); it has been used to detect cattle (Xu et al., 2020a; Xu et al., 2020b), sheep (Xu et al., 2020b; Luo et al., 2022), horses (Luo et al., 2022), kiangs (Luo et al., 2022), and whales (Gray et al., 2019).

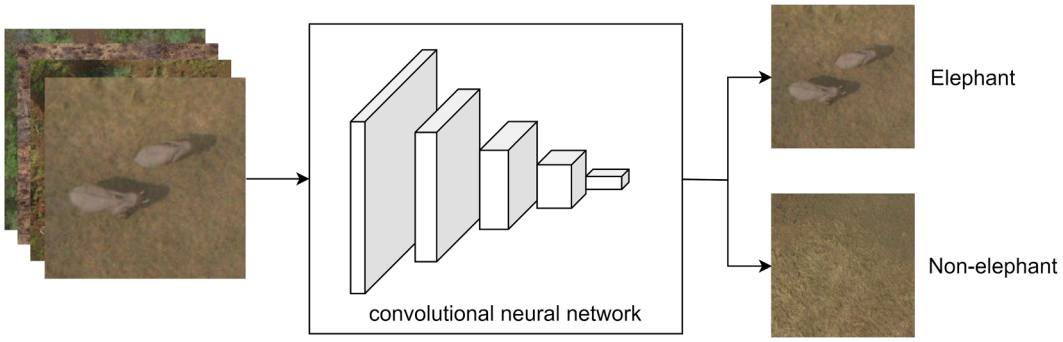


Fig. 8. An example of image level animal detection. In this case, sub-images are identified as “Elephant” or “Non-elephant”.

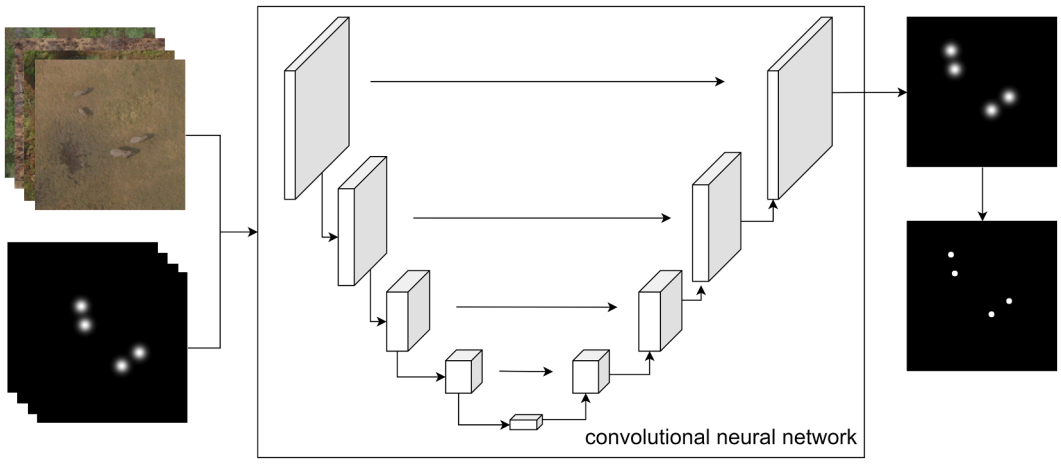


Fig. 9. An example of point level animal detection using the U-Net network (Ronneberger et al., 2015) and density map. The original point annotations are expanded into density maps through the Gaussian kernel function, and then the point results are obtained by regressing the output density maps.

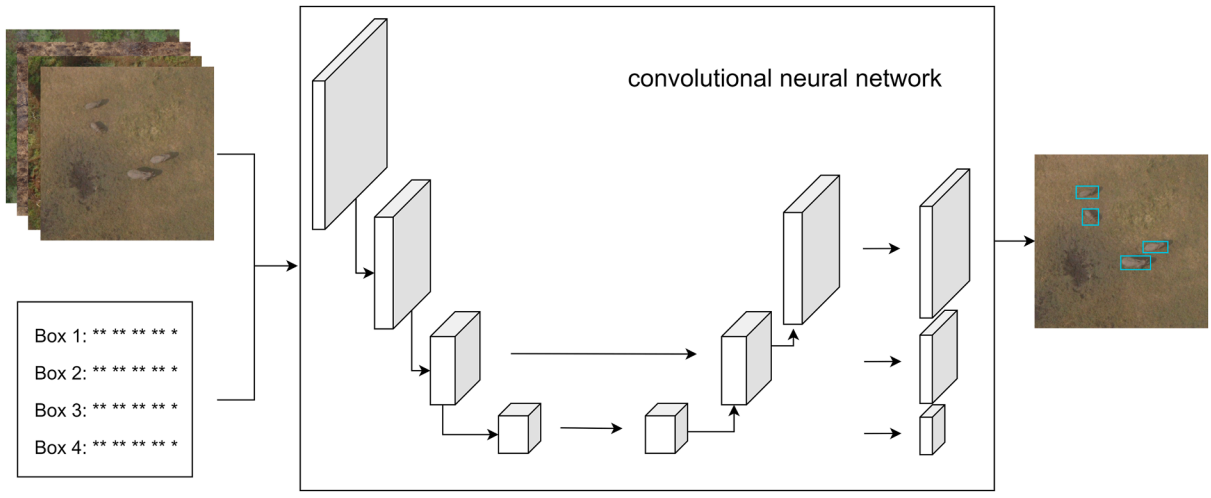


Fig. 10. An example of bounding box level animal detection process. The network structure is a simplified YOLOv3 network (Redmon and Farhadi, 2018). The input label is the coordinate position and object type (species of animal) of the bounding box marked in text form. Finally, the neural network generates predicted boxes of animal objects through multiple scales.

5.2.5. *Specific information level*

After identifying the species, location and numbers of the target animals, further analysis can obtain more information about the target animals. Using YOLOv3, Mücher et al. (2022) distinguished three poses of cattle (standing, grazing, and lying). Lenzi et al. (2023) differentiated between calves and adults in caribou. Mcilwaine et al. (2021) used deep convolutional network for jellyfish bloom detection and split images into two classes: ‘Bloom present’ and ‘No bloom present’. Animal body length can also be detected: Gray et al. (2019) used the instance segmentation results on whale objects to derive the principal direction by principal component analysis, and then obtained more accurate whale length information. Studies currently typically rely on images with very

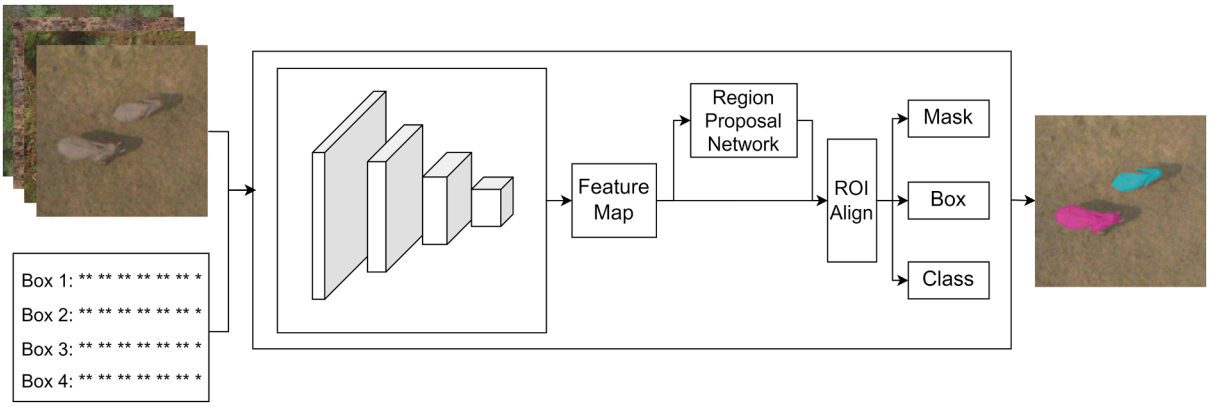


Fig. 11. An example of instance segmentation level animal detection. ROI = region of interest. The network structure is a schematic diagram of Mask R-CNN (He et al., 2017). The sample label is in text format and contains the coordinate information on the object contour point and the class (species) information. The mask output by this network represents the segmentation result of the animal target, the box represents the bounding box because Mask R-CNN has been developed from Faster R-CNN (a network for bounding box-based object detection), and classes represent the species. Finally, the output information is summarized to obtain the result of instance segmentation.

high spatial resolutions, or the animals studied are large in size. Like instance segmentation level, there are currently few studies of this kind, but there is great potential for future development.

5.3. Model evaluation techniques

5.3.1. Basic metrics

The most used basic evaluation metrics are precision, recall, and comprehensive accuracy score (F-measure, Chinchor, 1992). The metrics take the sample label as the ground truth value and the output result of deep learning method as the predicted value. The precision and recall are obtained by:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

where TP indicates the number of true positives and FP indicates the number of false positives (detection errors). FN indicates false negatives, i.e., the number of correct values that have not been detected. On this basis, the F-measure can be obtained by:

$$F = \frac{\left(1 + \beta^2\right) \cdot Precision \cdot Recall}{\beta^2 \cdot Precision + Recall} \tag{3}$$

Precision represents the correct proportion of detected animal objects, recall represents the detected proportion of actual animal objects, and the F-measure can more comprehensively represent the accuracy of detection. The value of β can influence the weighting of precision and recall in the overall precision evaluation. When$\beta = 1$, it is the currently most used F1-measure (also called F1-score).

Another commonly used precision evaluation index is the mean Average Precision (mAP). This method first draws the P-R curve according to the Precision and Recall under different thresholds, and then uses the area under the curve as the average precision.

5.3.2. Specific accuracy evaluation metrics

(1) Accuracy based on bounding box.

Intersection Over Union (IOU) ranges from 0 to 1, representing the ratio of the intersection area to the union area between the predicted bounding box and the true bounding box (see Fig. 12). When IOU = 0, the predicted box and the real box do not intersect; when IOU = 1, the predicted box and the real box overlap completely.

(2) Accuracy based on point.

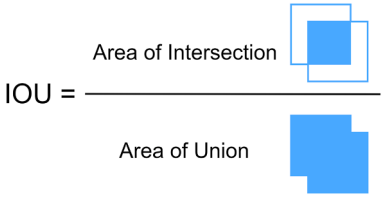


Fig. 12. Schematic diagram of the calculation of Intersection Over Union (IOU).

A predicted point that is within a specified distance from a point or points in the ground truth samples is considered as a correct detection. The distance can be the distance between pixels (Naudé and Joubert, 2019; Goncalves et al., 2020) or the real-world distance (Kellenberger et al., 2021).

(3) Accuracy based on the density map.

In some studies, the density map predicted by the deep network is used directly for calculating accuracy, and the ground truth of the density map is generally generated from points (Rahnemoonfar et al., 2019; Padubidri et al., 2021; Zhu et al., 2021). The main evaluation metrics are Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Grid Average Mean absolute Error (GAME).

6. Challenges in animal detection

6.1. Data imbalance

Images for total counts, sample counts, or straightforward animal detection are usually acquired over an area that is much larger than the target animals. In some images, animals will be densely clustered in a small area, whereas many other images may be negative (contain no animals). Data imbalance refers to this pixel imbalance between foreground information (animal objects) and background information. Some studies have proposed solutions to this problem; the four main solutions are summarized below.

(1) Optimization of sample datasets.

The most direct way to deal with data imbalance is to adjust the training samples. In most cases, the number of negative samples (samples that do not contain animals) exceeds the number of positive samples. Therefore, one solution is to increase the number of positive samples; another is to reduce the number of negative samples.

Positive samples are frequently augmented in animal detection. For instance, when detecting stingrays, Chou et al. (2018) opted for a data

augmentation method that used a generator called conditional Generative Latent Optimization (GLO) to increase the positive training samples. The generator is a kind of Generative Adversarial Network (GAN, Goodfellow et al., 2014) that creates fake data and has a discriminator that tries to differentiate real from fake, with both improving until the fake looks real. In this way, it can produce more samples with animals. This method is currently being extensively applied in many remote sensing tasks (Wang et al., 2019b; Ahmad et al., 2021), yet its use in animal detection remains limited. Currently, with the development of the Artificial Intelligence Generated Content (AIGC), in addition to GANs, there are more generative models available like stable diffusion (Rombach et al., 2022). Using these methods to increase the number of positive samples is a worthwhile research direction.

Imbalanced data can also be addressed using negative samples. A commonly used method is hard negative mining (Girshick et al., 2014). The hard negative samples refer to the objects that will likely confuse: for example, training not only on animals like elephants but also on big tree trunks, logs and rocks. In specific applications, Bowley et al. (2018) used feedback loops to optimize the selection of hard negative samples to replace manual selection. Peng et al. (2020) treated the hard negative samples as an independent class for training in kiang detection. These methods aim to minimize false positives in their studies and have yielded significant improvements.

Augmenting positive or negative samples has markedly improved detection accuracy. However, it is less effective when the data is very imbalanced (Naudé and Joubert, 2019). Furthermore, such methods may overly focus on extreme samples, or affect the overall distribution of the data, leading to risks such as overfitting (overfitting is when a model learns the training data's noise, which reduces its ability to generalize). Therefore, when using such methods, it is necessary to conduct a stricter assessment of the quality of the samples, especially in the evaluation of images generated by generative networks.

(2) Adjusting the deep learning network.

Refining the network is often an effective approach. Focal loss (Lin et al., 2020) was first introduced by Facebook AI Research (FAIR) as a method designed to address the issue of data imbalance and has since been widely adopted in various object detection tasks with abundant backgrounds and sparse foreground objects. In animal detection, Bowler et al. (2020) successfully used focal loss (Lin et al., 2020) in U-Net for albatross detection. Tversky loss (Salehi et al., 2017) is also an effective method for dealing with imbalanced data and has been successfully applied to wildebeest detection (Wu et al., 2023). The FSSCaps-DetCountNet developed by Sundarama and Loganathan (2020) also works well on challenging imbalance datasets. These studies showed that improvements in deep learning network can alleviate the data imbalance issue in animal detection. However, most of the improvement strategies come from outside the animal detection field. If there are more targeted methods for animal objects, the accuracy may be further improved.

(3) Two-step detection.

The first step in such methods is to find the small area containing the target animal(s) from a large area; this step is usually achieved by the method at image level. In the second step, in the block containing the animal objects, more accurate information on the location and quantity of the animal object is obtained within these sub-areas or tiles. Rahnemoonfar et al. (2019) developed a two-stage network called DisCountNet that uses theories from detection and from heat-map networks. The first stage (DiscNet) is used to select regions, and the second stage (CountNet) is used to count the objects inside the selected regions. They successfully applied this network to cattle detection. In whale detection, Guirado et al. (2019) also used the two-step approach. The first step, called "whale presence detection phase", involved the Inception v3 network; the second step, the "whale counting phase", involved the Faster R-CNN network. The main problem with this type of two-step approach is that errors may accumulate, and obvious deviations in the first step cannot be compensated by the second step. So, the first step must ensure detection is very accurate.

(4) Curriculum learning.

In curriculum learning (Bengio et al., 2009), the model can be trained adaptively on different parts of the dataset, usually on progressively more complex datasets. In animal detection, the model can first be trained on a sub-dataset (a dataset with a balanced ratio of animal objects to background), and then be gradually expanded in imbalanced data to obtain more comprehensive information. This method was successfully applied by Kellenberger et al. (2018) for mammal detection based on the SAVMAP dataset and by Chabot et al. (2022) for the detection of polar bears. While this method addresses the data imbalance issue efficiently, designing a "good" curriculum itself is not easy and may require more knowledge and experimentation. When using such methods, it is necessary to evaluate the time cost required for designing and validating the curriculum.

6.2. Small samples

"Small sample" refers to the number of training samples being limited. The problem of small samples in animal detection is mainly reflected in two aspects. One is that for some small tasks, such as livestock detection in some farms, the observation data is relatively small and not enough samples can be obtained. This may also apply to large imbalanced datasets, where a rare animal species occurs in few images. Secondly, and conversely, in larger tasks the animal objects to be annotated are also very large, and annotation is time-consuming. Therefore, such tasks require methods that make fewer demands on samples. To address this, current studies primarily involve methods based on data augmentation and learning strategies.

(1) Data augmentation methods.

Data augmentation methods can effectively expand the number of samples. Mirroring and rotation are commonly used methods to expand datasets. Han et al. (2019) expanded positive samples by flipping vertically, horizontally, and across both axes in livestock detection. Kellenberger et al. (2018) studied the impact of using rotation augmentation at different stages of training on detection accuracy and concluded that rotation is a useful augmentation method when combined with smaller learning rates and applied at later training stages. In recent years, new data augmentation methods have considerably improved computer vision accuracy. Some of these algorithms have been incorporated into new network architectures and applied in animal object detection research (e.g., the mosaic algorithm included in YOLOv4, Bochkovskiy et al., 2020). However, these new data augmentation methods have not been specifically evaluated in animal detection. The sample dataset can also be expanded through generating neural networks such as reported by Chou et al. (2018) for stingray detection. While these image enhancement methods offer some benefits, they might also add noise that affects training. Additionally, since the enhanced samples also come from the original data, the potential for accuracy improvement is limited. Therefore, it is necessary to consider the efficiency of the algorithm and use data augmentation algorithms in a limited way, such as by setting enhancement ratios and random variables to control the frequency of their use.

(2) Weakly supervised learning and self-supervised learning.

These methods differ from fully supervised learning in that they do not require a large number of fully labelled samples. Weakly supervised learning uses incomplete, inaccurate, or inconsistent labelled data for training models. Self-supervised learning generates its own supervision from input data without explicit labels. Research has indicated that improvements in these methods can reduce the need for annotated data in animal detection. In weakly supervised learning, algorithms such as VQ-VAE or VAE-GRF can be utilized to train on datasets that do not include animals, without the need for additional annotations (Pham et al., 2023; Gangloff et al., 2023). Subsequently, animals can be identified through anomaly detection methods (to identify unusual or anomalous patterns in a dataset). Moreover, research has shown that

with the weakly supervised learning approach, using a small number of accurate samples can achieve the detection accuracy of almost complete samples (Kellenberger et al., 2019b). By combining Transfer Sampling (TS) and a new window cropping strategy, Kellenberger et al. (2019a) found 80 % of animals in a challenging datasets using only half of the labelled data. Self-supervised methods can also play a role in reducing sample requirements. Zheng (2021) studied the application of self-supervised pre-training in wild animal target detection and found that his method can effectively reduce the number of training samples needed. These weakly supervised and self-supervised methods significantly reduce the cost of sample labelling, but currently, there is no method proposed to our knowledge that can completely eliminate the need for annotations. Notably, some of these methods also increase the complexity of the algorithms and the instability of the models. We have also observed that most research requiring lower supervision effort is conducted at the image or point level and can hardly achieve the performance of fully supervised learning. Nonetheless, as the volume of remote sensing data increases, the demand for sample volume will also grow, the importance of these methods is expected to rise significantly.

6.3. Small objects

The small objects here refer not to the actual size of the animals but to the small number of pixels the animal occupies in the images. Different datasets for object detection have different definitions of small objects. In the COCO dataset (a widely used dataset in common computer vision, Lin et al., 2014), objects smaller than 32 × 32 pixels are defined as small. The aerial image dataset DOTA assigns objects smaller than 50 pixels (side length) to small categories (Xia et al., 2018), but there are no animal objects in this dataset. In remote sensing- based animal detection there is no clear definition of small objects, even though most animal detection tasks encounter this issue. For example, in a study on wildebeest using Worldview data, each individual wildebeest occupies a square of about 3 × 3 pixels (Wu et al., 2023). Several studies have suggested methods to address the issue of small objects.

(1) Methods based on modifying network structures.

Some studies modified the network for detecting the small targets. Ma et al. (2022) used the HRNet, which is more suitable for small target detection, as the backbone of the Fast R-CNN network for large herbivore detection. Razaak et al. (2019) used a multi-scale approach of low-level feature combinations with deconvolutional modules on the SSD network to improve small object detection. Another commonly used method involves improving the receptive field. In deep learning, the receptive field refers to the area of the input data that a neuron can “see” or respond to. Ulhaq et al. (2021) integrated dilated convolution to YOLO to increase the receptive field for animal detection in low-resolution airborne thermal imagery. Li et al., (2022b) added a 3 × 3 maximum pooling layer to the SPP module to improve the receptive field in YOLOv5 and thereby to improve the detection accuracy of small and medium objects. Though effective, most of the methods employed are generic techniques from computer vision, with few customized for animal detection. In response to the characteristics of small targets, adjusting the convolutional structure or appropriately introducing the transformer model (Vaswani et al., 2017) to enhance the extraction and expression of features may improve the detection accuracy of small targets in dense areas. It should be noted that directly using models such as transformers may not be suitable for detection of small targets, and therefore these modules need to be adjusted to accommodate remote sensing image and animal detection.

(2) Methods based on adjusting samples.

Compared with methods based on modifying the network structure, the current method of adjusting samples is more specific to animal objects. Super-resolution reconstruction can be used to address the small object problem. Xue et al. (2022) used HAN resolution enhancement methods and improved the object sizes in samples on the AED and SAVMAP datasets. However, even under ideal conditions, the reconstructed high-resolution image may not fully achieve the quality of the original image. As animal objects in images are usually small, reconstruction errors easily lead to detection errors. Therefore, the role of super-resolution reconstruction methods needs to be confirmed in more types of remote sensing data. When the sample objects are too small, another effective method is to use point-based samples instead of bounding box-based samples, especially for animal objects of only a few pixels (Bowler et al., 2020; Goncalves et al., 2020; Wu et al., 2023).

It is worth noting that in some small object detection, the scale difference of the object is not obvious. However, some mainstream detectors such as YOLO v3-v5 and Faster R-CNN are multi-scale networks. In this case, the network structure and data characteristics are not completely matched. A few studies have addressed this issue. Torney et al. (2019) removed all but the final scale boxes in YOLOv3, as in their application for wildebeest detection the objects were only present at a single scale. Shao et al. (2020) chose a single scale network YOLOv2 and calculated the most suitable image input scale using the UAV flight altitude and the network parameters. These studies show that multi-scale networks are not always suitable for animal detection based on remote sensing images. Therefore, when applying existing deep learning models to animal detection, appropriately adding, deleting or adjusting the scale-related structures in the networks may bring about better detection results.

6.4. Image annotation methods

As demonstrated by Fig. 4, the two most used annotation methods are the bounding box based and point based methods. Both can be used to count and locate animals. There is a phenomenon that different researchers use different annotation methods on the same public dataset. For instance, when using the Aerial Elephant Dataset (Naudé and Joubert, 2019), Naudé and Joubert (2019) and Padubidri et al., (2021) chose point-based methods, while Delplanque et al. (2021) chose bounding box-based methods. Studies that discuss the issue of annotation methods favor the point-based methods (Rahnemoonfar et al., 2019; Padubidri et al., 2021; Sarwar et al., 2021). They suggest that bounding box based methods are less accurate when the animal distribution is dense, and the process of sample preparation is time-consuming. However, there are two problems in the current related research. The first is that some articles are theoretical, not empirical (Rahnemoonfar et al., 2019; Padubidri et al., 2021). The second is that although some papers report empirical research, it is difficult to judge whether the difference in accuracy is due to the difference in annotation methods or to the difference in neural network structures. Therefore, this question cannot yet be definitively answered, and more rigorous experimental evaluation is needed to clarify the effect of annotation methods. However, some recent studies have shown that if animals occupy only a few pixels in an image, the point based method is the most appropriate (Wu et al., 2023).

6.5. Image background

Currently, several studies address the complexity of the environment in which animals live. In some tasks, such as sheep detection, the ground objects are relatively simple: white sheep and green grass. In this case, the animal objects are easily identifiable. However, there are also background problems in sheep detection. Sarwar et al. (2021) found that in sunny weather, the border of a sheep may blur with the background due to sunlight being reflected by the grass; the U-Net-MS model they proposed using outperforms other networks in this case. In the detection of other animals, Han et al. (2019) divided the livestock detection images acquired by UAV into three categories according to the detection difficulty. The most difficult category had different colors of livestock and more distractions like snow, houses, and landforms in the background. In the detection of sea turtles, the individuals were distributed across a range of seabed depths from shore to 5 m. Experiments showed

that detection decreased significantly with each additional meter in depth and was close to zero at a depth of 5 m (Dujon et al., 2021). These studies show the influence of image backgrounds on the detection accuracy of deep learning. However, the background of animals in different studies varies widely, and no research has been done on a standardized index of background complexity. Therefore, the level of background complexity is difficult to directly compare across studies. For example, a background that is complex in some studies may be simple in others. Therefore, quantitative calculation of parameters related to environmental background (e.g., spatial heterogeneity) is needed to evaluate the application effects of deep learning methods more accurately in different backgrounds.

6.6. Animal counting

For most animal monitoring tasks, object counting is the ultimate goal. However, direct detection results may not meet the needs of counting. Therefore, further optimization is required to obtain more accurate counting results.

(1) Improvements in network architecture. Most current detection networks focus more on identifying the location of objects rather than counting them. To enhance their effectiveness for counting tasks, it's feasible to modify the network structure. For instance, Gonçalves et al. (2020) added a branch for counting to U-Net based on the WideResnet architecture. In addition, there is also research on making the network structure only output density map to specifically serve the task of counting (Padubidri et al., 2021; Meena et al., 2023).

(2) Post-detection processing. This involves the steps taken after objects are detected to refine or interpret the results. For example, after obtaining the results at the bounding box level, Eikelboom et al. (2019) used a correction factor to improve the accuracy of sample-based aerial counting.

(3) Accuracy assessment metrics. By establishing accuracy evaluation metrics that are more closely related to counting, the counting performance can be better reflected. This approach can thus further encourage the refinement of related methods. Currently, the commonly used metrics for evaluating counting accuracy include Mean Absolute Error, Root Mean Square Error and Mean Absolute Percentage Error (MAE, RMSE, MAPE, Sarwar et al., 2021; Delplanque et al., 2023a).

6.7. Accuracy assessment

The accuracy of each study is included in the Appendix 1. It is evident that the accuracy evaluation metrics in current studies are highly inconsistent. At the bounding box level, some studies directly use common indicators in computer vision, such as mAP, but these indicators are not always the most suitable for animal detection. Moreni et al. (2023) studied the difference between F1-score and mAP in animal detection and found that F1-score is more suitable than mAP because the animal detection task pays more attention to the location of the animals rather than the spatial range. At the point level, there are more detailed problems. Some studies use distance to determine whether objects have been detected correctly, but two questions arise. Should the pixel distance or the actual distance on the ground be used? How should the distance be determined? At present, different studies have different answers to these two questions (Naudé and Joubert, 2019; Kellenberger et al., 2021). For example, there are different methods for setting the threshold distance, as shown in Fig. 13. Even if the distance determination method is established, an issue remains: multiple predicted points near a ground truth point may satisfy the distance matching criteria (Fig. 13), and treating them all as true positives might significantly impact the accuracy. Different studies have opted for different methods and some studies do not even mention this problem that must be faced (Naudé and Joubert, 2019; Wu et al., 2023). There is therefore a need to develop a unified, scientific, and suitable accuracy assessment scheme for animal detection. For example, as illustrated in Fig. 13, the first method tends to classify a larger number of predicted points as True Positives (TP), resulting in an artificially inflated count. In contrast, the second method offers greater accuracy, which is also the standard approach recommended by the publishers of the COCO dataset (https://cocodataset.org/). Therefore, the formulation of a unified accuracy evaluation scheme should take into account these key details.

6.8. Uncertainty estimation

The ability to estimate uncertainty is essential for animal detection tasks and related ecological research. According to the collated papers, the sources of uncertainty include animals' camouflage, low spatial resolution images, background interference, and observer annotation errors (Guirado et al., 2019; Bowler et al., 2020; Kellenberger et al., 2021; Hinke et al., 2022; Lenzi et al., 2023). To quantify the uncertainty of model performance, Chalmers et al. (2021) employed 95 % confidence interval (CI), which represents an estimated range that is likely to include the true parameter value. In addressing the impact of uncertainty on the spatial accuracy of objects, Robinson et al., (2021) introduced two algorithms: an optimistic matching algorithm for dense scenes and a conservative matching algorithm for precise, one-to-one pairings. However, other studies lack quantitative research on uncertainty estimation. With increased emphasis on uncertainties in animal detection, there will be more methods employed to estimate them: for example, Bayesian methods (e.g. Bayesian Neural Networks, Monte Carlo dropout) and ensemble methods for uncertainty estimation, which have been used in remote sensing and deep learning (Cockx et al., 2014; Le et al., 2018; Loquercio et al., 2020; Abdar et al., 2021).

7. Future trends

7.1. Video-based detection

Many of the animal detection tasks in the papers we reviewed recorded data in video form. Instead of using the original video footage, in most research still images are extracted and analyzed (Sarwar et al., 2021). Information that helps improve detection accuracy, such as multi-frame timing information in the video, was not used in most studies. However, some recent studies have started focusing on directly processing video data and extracting information through techniques such as multi-frame processing and time series data processing. Zhu

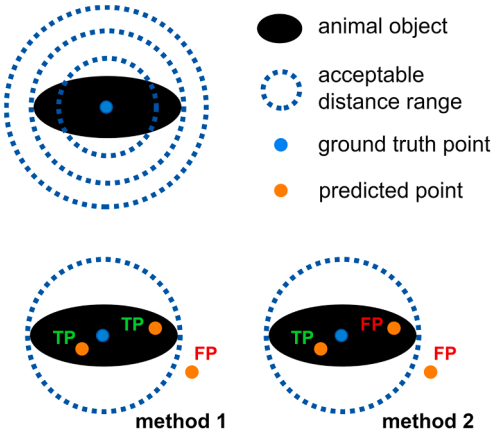


Fig. 13. Schematic diagram of point accuracy judgment. The upper illustration shows that there may be different distance threshold settings: less than the animal body length, equal to the animal body length, greater than the animal body length. The illustrations below show two TP (true positive) determination methods. Method 1 means that all prediction points that meet the distance range are considered TP, and method 2 means that within the acceptable distance range, only the prediction points closest to the real target are considered TP. FP indicates false positive.

et al. (2021) used a graph regularized flow attention network containing a temporal consistency module for video-based detection. Stojnić et al. (2021) used video stabilization and frame segmentation to detect small moving animals. In general, current optimization methods for video data in animal detection mainly come from the field of computer vision. Nevertheless, animals do not always exhibit rapid movement relative to the speed of movement of the background (e.g., video from an aircraft). Hence, in many instances within the video, the movement of animal targets is essentially imperceptible, unless shooting with an aircraft that can hover. This situation is quite different from video data in ordinary computer vision tasks. Consequently, it is crucial to explore whether current video processing approaches, such as optical flow and temporal fusion algorithms, are suitable for detecting animals in videos. Figuring out how to adjust these techniques for more efficient use remains a key issue in the realm of video-based animal detection.

7.2. Very high resolution satellite imagery

High resolution satellite imagery has traditionally been very costly, which is a major constraint on satellite-based animal detection. However, prices are now falling as more companies are offering < 1 m imaging capabilities from multiple constellations. In addition to Maxar's launched and developing WorldView constellation, Airbus and Planet company are similarly advancing their Pléiades Neo and Pelican constellations, respectively, all with spatial resolutions in the vicinity of 30 cm. These developments suggest that due to increased competition the price of commercial high resolution satellite imagery will likely fall. Notably, the Albedo constellation, set to launch next year, aims for a groundbreaking spatial resolution of 10 cm, potentially matching or even surpassing current aerial imagery. These developments indicate that animal detection based on very high-resolution satellite imagery has increasing potential for animal detection.

In addition, facilitating data sharing represents another avenue for advancing research utilizing satellite data. As we have enumerated in Table 1, there are shared datasets based on aircraft acquisitions. However, the sharing of very high-resolution commercial satellite data in the research community is currently not as prevalent, due to reasons like data policy restrictions. Addressing this issue may lead to meaningful progress in animal detection.

7.3. Multiple species detection

Although deep learning has significantly improved the accuracy of animal detection, most current studies concern single species detection. Some studies use datasets on more than one species but do not specifically distinguish them: for example, lumping different species together into the general categories like "animals", "mammals", or "livestock" (Kellenberger et al., 2018; Han et al., 2019). However, multi-species detection faces practical needs. For instance, it is critically important for detection in multi-species wildlife counts in African savannahs, where accurate numbers are needed for management (Lamprey et al., 2020a). With the improvement in image spatial resolution and the advancement of deep learning techniques, a few studies have begun to discuss multi-species detection (Delplanque et al., 2021) — a trend that is likely to continue. Most deep learning methods, such as YOLO, Faster R-CNN, etc., have efficient multi-category detection capabilities. But the main problem is that the animals are so small that there are little differences in features between species in the images. Therefore, optimizing the feature extraction part of the neural network, or using a stronger feature extractor, may significantly help solve this problem.

7.4. New annotation methods

New annotation methods can offer possibilities for improving accuracy and extracting more information in animal detection. The Oriented Bounding Box (OBB) adds a rotation angle relative to the vertical or horizontal direction based on the Horizontal Bounding Box (HBB). OBB can not only provide direction information of objects, but also effectively prevent boxes from overlapping in situations where objects like animals are densely packed. Currently, OBB has been widely used in remote sensing (Xia et al., 2018), but it has not yet been studied for animal detection. Another new annotation method is called 'keypoint annotation', which refers to labelling the key parts of objects, such as the head and limbs of animals. By using this method, the animal's pose, movement status and other information can be obtained. It should be noted that the keypoint here is not the keypoint features used in algorithms such as Scale-Invariant Feature Transform (SIFT) for image matching in some studies (Rahnemoonfar et al., 2019). Among the collected papers, some mentioned keypoint annotations without in-depth research, yet some included it in their future research plans (Hayes et al., 2021; Doll and Loos, 2023).

7.5. Specialized network structures

The mainstream applications in animal detection are still applied directly or simply adjusted to the off-the-shelf networks in computer vision. As there are already numerous network structures in the fields of object detection that can be used for animal detection, applying them directly can result in significant time savings and prompt updates to the latest network structures. However, some typical neural networks are not ideal for animal detection. For example, as mentioned in 6.3, the popular multi-scale structures may not be necessary in some animal detections. Some studies have begun to explore deep learning methods specifically for animal detection. Xue et al. (2022) proposed using an altitude-augmented module which can be usefully applied to aerial datasets with different altitudes. Additionally, although networks like U-Net were originally designed for semantic segmentation, some studies now use it to generate density maps to identify animal locations (Naudé and Joubert, 2019), which can also be viewed as a specialized improvement. Similar improvements can not only improve the accuracy of animal detection, but will also avoid deploying unnecessary computing resources. However, these studies are still few and do not reflect the characteristics of animal detection. We believe that, starting from the characteristics of animal targets (such as animal morphological characteristics and their aggregation, etc.) and the environment in which they live, rethinking the way scale issues are handled in the network are potential ideas that can make the models closer to animal detection and enhance their performance. It should be noted that overly detailed networks may not perform well in generalization, making them only suitable for very specific situations and requiring careful consideration in practical research.

7.6. Large foundation models

A foundation model is any model that is trained on broad data that can be adapted (e.g., fine-tuned) to a wide range of downstream tasks (Bommasani et al., 2022). In recent years, large foundation models like GPT-4 (OpenAI, 2023) have shown good application results. In image processing, there has also been significant advancement. Segment Anything Model (SAM) is a large model that can be applied to many scenarios such as interactive segmentation, boundary detection, semantic segmentation, instance segmentation, panoramic segmentation, etc. (the ViT-H version in SAM has 636 million parameters, Kirillov et al., 2023). Studies have shown that SAM can play a role in remote sensing tasks like image segmentation (Chen et al., 2023b). There are also models specially proposed for remote sensing data. Sun et al. (2023) developed a foundation model called RingMo using a large dataset by collecting two million remote sensing images and showed high accuracy in object detection tasks. However, this study did not explicitly include data related to animal detection. The aforementioned recent research underscores the potential for significant progress in large-scale models for animal monitoring. Nevertheless, large models typically require

extensive datasets (even when employing unsupervised methods) along with adequate hardware resources. Addressing these challenges is crucial for the development of such models, and collaborative sharing of data and computational resources could be a pivotal factor in their advancement.

8. Conclusion

This article presents a review of the application of deep learning for animal object detection in remote sensing. We first introduced different remote sensing platforms for animal detection. Then we compiled and analyzed the deep learning methods used in existing research and categorized them into five levels based on criteria such as annotation methods. Our analysis revealed that the most used neural network structures are YOLO, Faster R-CNN, U-Net, and ResNet, and that research at the bounding box level and point level predominates, but that with advances in deep learning technology and the improvement of image spatial resolution, future developments will probably be at the instance segmentation level and specific information level. After summarizing the deep learning methods, we identified the primary challenges in animal detection: data imbalance, small samples, small objects, image annotation methods, image background, animal counting, accuracy assessment and uncertainty estimation. We summarized the methods used in existing research to address these challenges and found that potential solutions for these challenges include handling positive or negative samples, adjusting network structures or annotation methods, and introducing new approaches like weakly supervised and self-supervised learning. At the same time, we also pointed out the shortcomings of the above solutions. Finally, we explored the trends in this field: considering video-based detection, very high-resolution satellite imagery, multiple species detection, new annotation methods, specialized network structures, and large foundation models.

CRediT authorship contribution statement

Zeyu Xu: Writing – original draft, Visualization, Validation, Resources, Project administration, Methodology, Investigation, Data curation. **Tiejun Wang:** Writing – review & editing, Validation, Supervision, Methodology. **Andrew K. Skidmore:** Writing – review & editing, Validation, Supervision. **Richard Lamprey:** Writing – review & editing, Validation, Supervision.

Declaration of competing interest

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

Data availability

We provided two supplementary files: one with tables on publications, datasets, methodologies, accuracies, and additional details at different detection levels, and the other with information about the species.

Acknowledgments

This work was supported by the China Scholarship Council, China (grant number 202104910129).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jag.2024.103732.

References

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U.R., Makarenkov, V., Nahavandi, S., 2021. A review of uncertainty quantification in deep learning: techniques, applications and challenges. Inf. Fusion 76, 243–297. https://doi.org/10.1016/j.inffus.2021.05.008.

Ahmad, T., Chen, X., Saqlain, A.S., Ma, Y., 2021. FPN-GAN: multi-class small object detection in remote sensing images. In: IEEE International Conference on Cloud Computing and Big Data Analytics, ICCCBDA, pp. 478–482. https://doi.org/10.1109/ICCCBDA51879.2021.9442506.

Almond, R., Grooten, M., Peterson, T., 2020. Living planet report 2020-bending the curve of biodiversity loss. World Wildlife Fund.

Andrew, W., Gao, J., Mullan, S., Campbell, N., Dowsey, A.W., Burghardt, T., 2021. Visual identification of individual Holstein-friesian cattle via deep metric learning. Comput. Electron. Agric. 185, 106133 https://doi.org/10.1016/j.compag.2021.106133.

Andrew, W., Greatwood, C., Burghardt, T., 2020. Fusing Animal Biometrics with Autonomous Robotics: Drone-based Search and Individual ID of Friesian Cattle (Extended Abstract). In: Presented at the 2020 IEEE Winter Applications of Computer Vision Workshops (WACVW). IEEE Computer Society, pp. 38–43. https://doi.org/10.1109/WACVW50321.2020.9096949.

Barbedo, J.G.A., Koenigkan, L.V., Santos, T.T., Santos, P.M., 2019. A study on the detection of cattle in UAV images using deep learning. Sensors 19, 5436. https://doi.org/10.3390/s19245436.

Barbedo, J.G.A., Koenigkan, L.V., Santos, P.M., 2020a. Cattle detection using oblique UAV images. Drones 4, 75. https://doi.org/10.3390/drones4040075.

Barbedo, J.G.A., Koenigkan, L.V., Santos, P.M., Ribeiro, A.R.B., 2020b. Counting cattle in UAV images-dealing with clustered animals and animal/background contrast changes. Sensors 20, 2126. https://doi.org/10.3390/s20072126.

Barrios, D.B., Valente, J., van Langevelde, F., 2023. Monitoring mammalian herbivores via convolutional neural networks implemented on thermal uav imagery. Available at SSRN. https://doi.org/10.2139/ssrn.4442721.

Bengio, Y., Louradour, J., Collobert, R., Weston, J., 2009. Curriculum learning. In: Proc. Annual International Conference on Machine Learning, ICML, pp. 41–48. https://doi.org/10.1145/1553374.1553380.

Bentley, P.M., McDonnell, J.T.E., 1994. Wavelet transforms: an introduction. Electron. Commun. Eng. J. 6, 175–186. https://doi.org/10.1049/ecej:19940401.

Bhardwaj, A., Sam, L., Martín-Torres, F.J., Kumar, R., 2016. UAVs as remote sensing platform in glaciology: present applications and future prospects. Remote Sens. Environ. 175, 196–204. https://doi.org/10.1016/j.rse.2015.12.029.

Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M., 2020. Yolov4: Optimal speed and accuracy of object detection. arXiv Prepr. arXiv:2004.10934. https://doi.org/10.48550/arXiv.2004.10934.

Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J.Q., Demszky, D., Donahue, C., Doumbouya, M., Durmus, E., Ermon, S., Etchemendy, J., Ethayarajh, K., Fei-Fei, L., Finn, C., Gale, T., Gillespie, L., Goel, K., Goodman, N., Grossman, S., Guha, N., Hashimoto, T., Henderson, P., Hewitt, J., Ho, D.E., Hong, J., Hsu, K., Huang, J., Icard, T., Jain, S., Jurafsky, D., Kalluri, P., Karamcheti, S., Keeling, G., Khani, F., Khattab, O., Koh, P.W., Krass, M., Krishna, R., Kuditipudi, R., Kumar, A., Ladhak, F., Lee, M., Lee, T., Leskovec, J., Levent, I., Li, X.L., Li, X., Ma, T., Malik, A., Manning, C. D., Mirchandani, S., Mitchell, E., Munyikwa, Z., Nair, S., Narayan, A., Narayanan, D., Newman, B., Nie, A., Niebles, J.C., Nilforoshan, H., Nyarko, J., Ogut, G., Orr, L., Papadimitriou, I., Park, J.S., Piech, C., Portelance, E., Potts, C., Raghunathan, A., Reich, R., Ren, H., Rong, F., Roohani, Y., Ruiz, C., Ryan, J., Ré, C., Sadigh, D., Sagawa, S., Santhanam, K., Shih, A., Srinivasan, K., Tamkin, A., Taori, R., Thomas, A. W., Tramèr, F., Wang, R.E., Wang, W., Wu, B., Wu, J., Wu, Y., Xie, S.M., Yasunaga, M., You, J., Zaharia, M., Zhang, M., Zhang, T., Zhang, X., Zhang, Y., Zheng, L., Zhou, K., Liang, P., 2022. On the opportunities and risks of foundation models. arXiv Prepr. arXiv: 2108.07258. https://doi.org/10.48550/arXiv.2108.07258.

Bondi, E., Jain, R., Aggrawal, P., Anand, S., Hannaford, R., Kapoor, A., Piavis, J., Shah, S., Joppa, L., Dilkina, B., Tambe, M., 2020. BIRDSAI: A dataset for detection and tracking in aerial thermal infrared videos. In: IEEE Winter Conference on Applications of Computer Vision, WACV, pp. 1747–1756. https://doi.org/10.1109/WACV45572.2020.9093284.

Borowicz, A., Le, H., Humphries, G., Nehls, G., Hoschle, C., Kosarev, V., Lynch, H.J., 2019. Aerial-trained deep learning networks for surveying cetaceans from satellite imagery. PLoS ONE 14, e0212532. https://doi.org/10.1371/journal.pone.0212532.

Bowler, E., Fretwell, P.T., French, G., Mackiewicz, M., 2020. Using deep learning to count albatrosses from space: assessing results in light of ground truth uncertainty. Remote Sens. 12, 2026. https://doi.org/10.3390/rs12122026.

Bowley, C., Mattingly, M., Barnas, A., Ellis-Felege, S., Desell, T., 2018. Detecting Wildlife in Unmanned Aerial Systems Imagery Using Convolutional Neural Networks Trained with an Automated Feedback Loop. In: Computational Science, ICCS. Lecture Notes in Computer Science, LNCS, 10860. Springer, Cham, pp. 69–82. https://doi.org/10.1007/978-3-319-93698-7_6.

Brown, J., Qiao, Y., Clark, C., Lomax, S., Rafique, K., Sukkarieh, S., 2022. Automated aerial animal detection when spatial resolution conditions are varied. Comput. Electron. Agric. 193, 106689 https://doi.org/10.1016/j.compag.2022.106689.

Brummitt, N., Regan, E.C., Weatherdon, L.V., Martin, C.S., Geijzendorffer, I.R., Rocchini, D., Gavish, Y., Haase, P., Marsh, C.J., Schmeller, D.S., 2017. Taking stock of nature: essential biodiversity variables explained. Biol. Conserv. 213, 252–255. https://doi.org/10.1016/j.biocon.2016.09.006.

Chabot, D., 2009. Systematic evaluation of a stock unmanned aerial vehicle (UAV) system for small-scale wildlife survey applications. McGill University.

Chabot, D., Stapleton, S., Francis, C.M., 2022. Using web images to train a deep neural network to detect sparsely distributed wildlife in large volumes of remotely sensed imagery: a case study of polar bears on sea ice. Ecol. Inform. 68, 101547 https://doi.org/10.1016/j.ecoinf.2021.101547.

Chalmers, C., Fergus, P., Curbelo Montanez, C.A., Longmore, S.N., Wich, S.A., 2021. Video analysis for the detection of animals using convolutional neural networks and consumer-grade drones. J. Unmanned Veh. Syst. 9, 112–127. https://doi.org/10.1139/juvs-2020-0018.

Chen, K., Liu, C., Chen, H., Zhang, H., Li, W., Zou, Z., Shi, Z., 2023b. RSPrompter: Learning to Prompt for Remote Sensing Instance Segmentation based on Visual Foundation Model. arXiv Prepr. arXiv:2306.16269. https://doi.org/10.48550/arXiv.2306.16269.

Chen, A., Jacob, M., Shoshani, G., Charter, M., 2023a. Using computer vision, image analysis and UAVs for the automatic recognition and counting of common cranes (Grus grus). J. Environ. Manage. 328, 116948 https://doi.org/10.1016/j.jenvman.2022.116948.

Chinchor, N., 1992. MUC-4 evaluation metrics. In: Proc. 4th conference on Message understanding, pp. 22–29. https://doi.org/10.3115/1072064.1072067.

Chou, Y.M., Chen, C.H., Liu, K.H., Chen, C.S., 2018. Stingray detection of aerial images using augmented training images generated by a conditional generative model. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPRW, pp. 1403–1409. https://doi.org/10.1109/CVPRW.2018.00189.

Cockx, K., Van de Voorde, T., Canters, F., 2014. Quantifying uncertainty in remote sensing-based urban land-use mapping. Int. J. Appl. Earth Obs. Geoinformation 31, 154–166. https://doi.org/10.1016/j.jag.2014.03.016.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297. https://doi.org/10.1007/BF00994018.

Couturier, S., Courtois, R., Crépeau, H., Rivest, L.-P., Luttich, S., 1996. Calving photocensus of the rivière George caribou herd and comparison with an independent census. Rangif. 16, 283–296. https://doi.org/10.7557/2.16.4.1268.

Cubaynes, H.C., Fretwell, P.T., Bamford, C., Gerrish, L., Jackson, J.A., 2019. Whales from space: four mysticete species described using new VHR satellite imagery. Mar. Mam. Sci. 35, 466–491. https://doi.org/10.1111/mms.12544.

DataCanary, Katie., Risdal, M., 2017. NOAA Fisheries Steller Sea Lion Population Count. Kaggle. https://kaggle.com/competitions/noaa-fisheries-steller-sea-lion-population-count.

Delplanque, A., Foucher, S., Lejeune, P., Linchant, J., Theau, J., 2021. Multispecies detection and identification of african mammals in aerial imagery using convolutional neural networks. Remote Sens. Ecol. Conserv. 8, 166–179. https://doi.org/10.1002/rse2.234.

Delplanque, A., Foucher, S., Théau, J., Bussière, E., Vermeulen, C., Lejeune, P., 2023a. From crowd to herd counting: how to precisely detect and count african mammals using aerial imagery and deep learning? ISPRS J. Photogramm. Remote Sens. 197, 167–180. https://doi.org/10.1016/j.isprsjprs.2023.01.025.

Delplanque, A., Lamprey, R., Foucher, S., Théau, J., Lejeune, P., 2023b. Surveying wildlife and livestock in Uganda with aerial cameras: deep learning reduces the workload of human interpretation by over 70%. Front. Ecol. Evol. 11 https://doi.org/10.3389/fevo.2023.1270857.

Desai, B., Patel, A., Patel, V., Shah, S., Raval, M.S., Ghosal, R., 2022. Identification of free-ranging mugger crocodiles by applying deep learning methods on UAV imagery. Ecological Informatics 72, 101874. https://doi.org/10.1016/j.ecoinf.2022.101874.

Desgarnier, L., Mouillot, D., Vigliola, L., Chaumont, M., Mannocci, L., 2022. Putting eagle rays on the map by coupling aerial video-surveys and deep learning. Biol. Conserv. 267, 109494 https://doi.org/10.1016/j.biocon.2022.109494.

Doll, O., Loos, A., 2023. Comparison of Object Detection Algorithms for Livestock Monitoring of Sheep in UAV images. In: Int. Workshop Camera traps, AI, and Ecology. https://doi.org/10.24406/publica-2164.

Dujon, A.M., Ierodiaconou, D., Geeson, J.J., Arnould, J.P.Y., Allan, B.M., Katselidis, K.A., Schofield, G., 2021. Machine learning to detect marine animals in UAV imagery: effect of morphology, spacing, behaviour and habitat. Remote Sens. Ecol. Conserv. 7, 341–354. https://doi.org/10.1002/rse2.205.

Duporge, I., Isupova, O., Reece, S., Macdonald, D.W., Wang, T., 2021. Using very-high-resolution satellite imagery and deep learning to detect and count african elephants in heterogeneous landscapes. Remote Sens. Ecol. Conserv. 7, 369–381. https://doi.org/10.1002/rse2.195.

Eikelboom, J.A.J., Wind, J., van de Ven, E., Kenana, L.M., Schroder, B., de Knegt, H.J., van Langevelde, F., Prins, H.H.T., 2019. Improving the precision and accuracy of animal population estimates with aerial image object detection. Methods Ecol. Evol. 10, 1875–1887. https://doi.org/10.1111/2041-210X.13277.

Farge, M., 1992. Wavelet transforms and their applications to turbulence. Annu. Rev. Fluid. Mech. 24, 395–458. https://doi.org/10.1146/annurev.fl.24.010192.002143.

Firchow, M., Vaughan, R., Mytton, R., 1990. Comparison of aerial survey techniques for pronghorns. Wildl. Soc. Bull. 1973–2006 (18), 18–23.

Gangloff, H., Pham, M.-T., Courtrai, L., Lefèvre, S., 2023. Unsupervised Anomaly Detection Using Variational Autoencoder with Gaussian Random Field Prior. In: 2023 IEEE International Conference on Image Processing (ICIP), pp. 1620–1624. https://doi.org/10.1109/ICIP49359.2023.10222900.

Gao, S.H., Cheng, M.M., Zhao, K., Zhang, X.Y., Yang, M.H., Torr, P., 2021. Res2Net: a new multi-scale backbone architecture. IEEE Trans. Pattern Anal. Mach. Intell. 43, 652–662. https://doi.org/10.1109/TPAMI.2019.2938758.

Girshick, R., 2015. Fast R-CNN. In: IEEE International Conference on Computer Vision, ICCV, pp. 1440–1448. https://doi.org/10.1109/ICCV.2015.169.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 580–587. https://doi.org/10.1109/CVPR.2014.81.

Glorot, X., Bordes, A., Bengio, Y., 2011. Deep Sparse Rectifier Neural Networks. In: Proc. the 14th International Conference on Artificial Intelligence and Statistics, PMLR, pp. 315–323.

Goncalves, B.C., Spitzbart, B., Lynch, H.J., 2020. SealNet: a fully-automated pack-ice seal detection pipeline for sub-meter satellite imagery. Remote Sens. Environ. 239, 111617 https://doi.org/10.1016/j.rse.2019.111617.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative Adversarial Nets. In: Advances in Neural Information Processing Systems. Curran Associates, Inc.

Gray, P.C., Bierlich, K.C., Mantell, S.A., Friedlaender, A.S., Goldbogen, J.A., Johnston, D. W., 2019a. Drones and convolutional neural networks facilitate automated and accurate cetacean species identification and photogrammetry. Methods Ecol. Evol. 10, 1490–1500. https://doi.org/10.1111/2041-210X.13246.

Gray, P.C., Fleishman, A.B., Klein, D.J., McKown, M.W., Bézy, V.S., Lohmann, K.J., Johnston, D.W., 2019b. A convolutional neural network for detecting sea turtles in drone imagery. Methods Ecol. Evol. 10, 345–355. https://doi.org/10.1111/2041-210X.13132.

Guirado, E., Tabik, S., Rivas, M.L., Alcaraz-Segura, D., Herrera, F., 2019. Whale counting in satellite and aerial images with deep learning. Sci. Rep. 9, 14259. https://doi.org/10.1038/s41598-019-50795-9.

Han, L., Tao, P., Martin, R.R., 2019. Livestock detection in aerial images using a fully convolutional network. Comput. vis. Media. 5, 221–228. https://doi.org/10.1007/s41095-019-0132-5.

Hayes, M.C., Gray, P.C., Harris, G., Sedgwick, W.C., Crawford, V.D., Chazal, N., Crofts, S., Johnston, D.W., 2021. Drones and deep learning produce accurate and efficient monitoring of large-scale seabird colonies. Ornithological Applications 123, duab022. https://doi.org/10.1093/ornithapp/duab022.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. In: IEEE International Conference on Computer Vision, ICCV, pp. 2961–2969. https://doi.org/10.1109/ICCV.2017.322.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.

Hinke, J.T., Giuseffi, L.M., Hermanson, V.R., Woodman, S.M., Krause, D.J., 2022. Evaluating thermal and color sensors for automating detection of penguins and pinnipeds in images collected with an unoccupied aerial system. Drones 6, 255. https://doi.org/10.3390/drones6090255.

Hodgson, J.C., Baylis, S.M., Mott, R., Herrod, A., Clarke, R.H., 2016. Precision wildlife monitoring using unmanned aerial vehicles. Sci. Rep. 6, 22574. https://doi.org/10.1038/srep22574.

Hollings, T., Burgman, M., van Andel, M., Gilbert, M., Robinson, T., Robinson, A., 2018. How do you find the green sheep? A critical review of the use of remotely sensed imagery to detect and count animals. Methods Ecol. Evol. 9, 881–892. https://doi.org/10.1111/2041-210X.12973.

Hong, S.J., Han, Y., Kim, S.Y., Lee, A.Y., Kim, G., 2019. Application of deep-learning methods to bird detection using unmanned aerial vehicle imagery. Sensors 19, 1651. https://doi.org/10.3390/s19071651.

Jetz, W., McGeoch, M.A., Guralnick, R., Ferrier, S., Beck, J., Costello, M.J., Fernandez, M., Geller, G.N., Keil, P., Merow, C., 2019. Essential biodiversity variables for mapping and monitoring species populations. Nat. Ecol. Evol. 3, 539–551. https://doi.org/10.1038/s41559-019-0826-1.

Jocher, G., Liu, C., Hogan, A., Yu, L., changyu98., Rai, P., Sullivan, T., 2020. ultralytics/yolov5, https://zenodo.org/record/7347926 (accessed June 2020). https://doi.org/10.5281/zenodo.7347926.

Jolly, G.M., 1969. Sampling methods for aerial censuses of wildlife populations. East Afr. Agric. for. J. 34, 46–49. https://doi.org/10.1080/00128325.1969.11662347.

Kabra, K., Xiong, A., Li, W., Luo, M., Lu, W., Yu, T., Yu, J., Singh, D., Garcia, R., Tang, M., Arnold, H., Vallery, A., Gibbons, R., Barman, A., 2022. Deep object detection for waterbird monitoring using aerial imagery. In: 21st IEEE International Conference on Machine Learning and Applications, pp. 455–460. https://doi.org/10.1109/ICMLA55696.2022.00073.

Kellenberger, B., Marcos, D., Tuia, D., 2018. Detecting mammals in UAV images: best practices to address a substantially imbalanced dataset with deep learning. Remote Sens. Environ. 216, 139–153. https://doi.org/10.1016/j.rse.2018.06.028.c.

Kellenberger, B., Marcos, D., Lobry, S., Tuia, D., 2019a. Half a percent of labels is enough: efficient animal detection in UAV imagery using deep CNNs and active learning. IEEE Trans. Geosci. Remote Sens. 57, 9524–9533. https://doi.org/10.1109/TGRS.2019.2927393.

Kellenberger, B., Marcos, D., Tuia, D., 2019b. When a few clicks make all the difference: improving weakly-supervised wildlife detection in UAV images. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 1414–1422. https://doi.org/10.1109/CVPRW.2019.00182.

Kellenberger, B., Veen, T., Folmer, E., Tuia, D., 2021. 21 000 birds in 4.5 h: efficient large-scale seabird detection with machine learning. Remote Sens. Ecol. Conserv. 7, 445–460. https://doi.org/10.1002/rse2.200.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., 2023. Segment anything. arXiv Prepr. arXiv:2304.02643. https://doi.org/10.48550/arXiv.2304.02643.

Krishnan, B.S., Jones, L.R., Elmore, J.A., Samiappan, S., Evans, K.O., Pfeiffer, M.B., Blackwell, B.F., Iglay, R.B., 2023. Fusion of visible and thermal images improves automated detection and classification of animals for drone surveys. Sci. Rep. 13, 10385. https://doi.org/10.1038/s41598-023-37295-7.

Lamprey, R., Ochanda, D., Brett, R., Tumwesigye, C., Douglas-Hamilton, I., 2020a. Cameras replace human observers in multi-species aerial counts in Murchison falls. Uganda. Remote Sens. Ecol. Conserv. 6, 529–545. https://doi.org/10.1002/rse2.154.

Lamprey, R., Pope, F., Ngene, S., Norton-Griffiths, M., Frederick, H., Okita-Ouma, B., Douglas-Hamilton, I., 2020b. Comparing an automated high-definition oblique camera system to rear-seat-observers in a wildlife survey in tsavo, Kenya: taking multi-species aerial counts to the next level. Biol. Conserv. 241, 108243 https://doi.org/10.1016/j.biocon.2019.108243.

Laradji, I., Rodriguez, P., Kalaitzis, F., Vázquez, D., Young, R., Davey, E., Lacoste, A., 2020. Counting Cows: Tracking Illegal Cattle Ranching From High-Resolution Satellite Imagery. Adv. Neural Inf. Process. Syst. https://doi.org/10.48550/arXiv.2011.07369.

LaRue, M.A., Stapleton, S., Porter, C., Atkinson, S., Atwood, T., Dyck, M., Lecomte, N., 2015. Testing methods for using high-resolution satellite imagery to monitor polar bear abundance and distribution. Wildl. Soc. Bull. 39, 772–779. https://doi.org/10.1002/wsb.596.

Le, M.T., Diehl, F., Brunner, T., Knoll, A., 2018. Uncertainty estimation for deep neural object detectors in safety-critical applications. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), pp. 3873–3878. https://doi.org/10.1109/ITSC.2018.8569637.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.

Lee, S., Song, Y., Kil, S.H., 2021. Feasibility analyses of real-time detection of wildlife using UAV-derived thermal and RGB images. Remote Sens. 13, 2169. https://doi.org/10.3390/rs13112169.

Lenzi, J., Barnas, A.F., ElSaid, A.A., Desell, T., Rockwell, R.F., Ellis-Felege, S.N., 2023. Artificial intelligence for automated detection of large mammals creates path to upscale drone surveys. Sci. Rep. 13, 947. https://doi.org/10.1038/s41598-023-28240-9.

Leyequien, E., Verrelst, J., Slot, M., Schaepman-Strub, G., Heitkönig, I.M.A., Skidmore, A., 2007. Capturing the fugitive: applying remote sensing to terrestrial animal distribution and diversity. Int. J. Appl. Earth Obs. Geoinf. 9, 1–20. https://doi.org/10.1016/j.jag.2006.08.002.

Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., Ke, Z., Li, Q., Cheng, M., Nie, W., 2022a. YOLOv6: A single-stage object detection framework for industrial applications. arXiv Prepr. arXiv:2209.02976. https://doi.org/10.48550/arXiv.2209.02976.

Li, Z., Namiki, A., Suzuki, S., Wang, Q., Zhang, T., Wang, W., 2022b. Application of low-altitude UAV remote sensing image object detection based on improved YOLOv5. Appl. Sci. 12, 8314. https://doi.org/10.3390/app12168314.

Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C. L., 2014. Microsoft coco: Common objects in context. In: Comput. Vis. ECCV. Lect. Notes Comput. Sci. (LNCS), 8693, pp. 740–755. https://doi.org/10.1109/ICCV.2015.181.

Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollar, P., 2020. Focal loss for dense object detection. IEEE Trans. Pattern Anal. Mach. Intell. 42, 318–327. https://doi.org/10.1109/ICCV.2017.324.

Linchant, J., Lhoest, S., Quevauvillers, S., Lejeune, P., Vermeulen, C., Ngabinzeke, J.S., Belanganayi, B.L., Delvingt, W., Bouche, P., 2018. UAS imagery reveals new survey opportunities for counting hippos. PLoS ONE 13, e0206413. https://doi.org/10.1371/journal.pone.0206413.

Loquercio, A., Segu, M., Scaramuzza, D., 2020. A general framework for uncertainty estimation in deep learning. IEEE Robot. Autom. Lett. 5, 3153–3160. https://doi.org/10.1109/LRA.2020.2974682.

Luo, W., Jin, Y., Li, X., Liu, K., 2022. Application of deep learning in remote sensing monitoring of large herbivores-a case study in Qinghai Tibet plateau. Pak. J. Zool. 54, 413. https://doi.org/10.17582/journal.pjz/20191205021259.

Ma, J., Hu, Z., Shao, Q., Wang, Y., Zhou, Y., Liu, J., Liu, S., 2022. Detection of large herbivores in UAV images: a new method for small target recognition in large-scale images. Diversity 14, 624. https://doi.org/10.3390/d14080624.

Mbugua, S., 1996. Counting elephants from the air–sample counts. Studying Elephants. Technical Handbook 21–27.

Mcilwaine, B., Rivas Casado, M., 2021. JellyNet: the convolutional neural network jellyfish bloom detector. Int. J. Appl. Earth Obs. Geoinf. 97, 102279 https://doi.org/10.1016/j.jag.2020.102279.

McRae, L., Deinet, S., Freeman, R., 2017. The diversity-weighted living planet index: controlling for taxonomic bias in a global biodiversity indicator. PLoS ONE 12, e0169156. https://doi.org/10.1371/journal.pone.0169156.

Meena, S.D., Manichandana, K.B.V., Potlur, R.S., Dhanyasri, M., Harshith, P., Sheela, J., 2023. Aerial imaging based sea lion count using modified U-net architecture. AIP Conf. Proc. 2869, 050024 https://doi.org/10.1063/5.0168211.

Moreni, M., Theau, J., Foucher, S., 2023. Do you get what you see? Insights of using mAP to select architectures of pretrained neural networks for automated aerial animal detection. PLoS ONE 18, e0284449. https://doi.org/10.1371/journal.pone.0284449.

Mou, C., Liu, T., Zhu, C., Cui, X., 2023. WAID: a large-scale dataset for wildlife detection with drones. Applied Sciences 13, 10397. https://doi.org/10.3390/app131810397.

Mücher, C.A., Los, S., Franke, G.J., Kamphuis, C., 2022. Detection, identification and posture recognition of cattle with satellites, aerial photography and UAVs using deep learning techniques. Int. J. Remote Sens. 43, 2377–2392. https://doi.org/10.1080/01431161.2022.2051634.

Naudé, J.J., Joubert, D., 2019. The Aerial Elephant Dataset: A New Public Benchmark for Aerial Object Detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPRW, pp. 48–55. https://openaccess.thecvf.com/content_CVPRW_2019/html/DOAI/Naude_The_Aerial_Elephant_Dataset_A_New_Public_Benchmark_for_Aerial_CVPRW_2019_paper.html.

Norton-Griffiths, M., 1973. Counting the Serengeti migratory wildebeest using two-stage sampling. Afr. J. Ecol. 11, 135–149. https://doi.org/10.1111/j.1365-2028.1973.tb00079.x.

Norton-Griffiths, M., 1978. Counting animals. Aeronautics in wildlife management, African Wildlife Leadership Foundation, p. 139.

Ofli, F., Meier, P., Imran, M., Castillo, C., Tuia, D., Rey, N., Briant, J., Millet, P., Reinhard, F., Parkan, M., 2016. Combining human computing and machine learning to make sense of big (aerial) data for disaster response. Big Data 4, 47–59. https://doi.org/10.1089/big.2014.0064.

Oliveira, T., Barbar, J., Soares, A., 2016. Computer network traffic prediction: a comparison between traditional and deep learning neural networks. Int. J. Big Data Intell. 3, 28. https://doi.org/10.1504/IJBDI.2016.073903.

OpenAI, 2023. GPT-4 Technical Report. arXiv Prepr. arXiv:2303.08774. https://doi.org/10.48550/arXiv.2303.08774.

Osco, L.P., Junior, J.M., Ramos, A.P.M., de Castro Jorge, L.A., Fatholahi, S.N., de Andrade Silva, J., Matsubara, E.T., Pistori, H., Gonçalves, W.N., Li, J., 2021. A review on deep learning in UAV remote sensing. Int. J. Appl. Earth Obs. Geoinf. 102, 102456 https://doi.org/10.1016/j.jag.2021.102456.

Padubidri, C., Kamilaris, A., Karatsiolis, S., Kamminga, J., 2021. Counting sea lions and elephants from aerial photography using deep learning with density maps. Anim. Biotelem. 9, 27. https://doi.org/10.1186/s40317-021-00247-x.

Peng, J.B., Wang, D.L., Liao, X.H., Shao, Q.Q., Sun, Z.G., Yue, H.Y., Ye, H.P., 2020. Wild animal survey using UAS imagery and deep learning: modified faster R-CNN for kiang detection in tibetan plateau. ISPRS J. Photogramm. Remote Sens. 169, 364–376. https://doi.org/10.1016/j.isprsjprs.2020.08.026.

Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., 2013. Essential biodiversity variables. Science 339, 277–278. https://doi.org/10.1126/science.1229931.

Petso, T., Jamisola, R.S., Mpoeleng, D., Mmereki, W., 2021. Individual Animal and Herd Identification Using Custom YOLO v3 and v4 with Images Taken from a UAV Camera at Different Altitudes. In: IEEE 6th International Conference on Signal and Image Processing, ICSIP. IEEE, pp. 33–39. https://doi.org/10.1109/ICSIP52628.2021.9688827.

Pham, M.-T., Gangloff, H., Lefèvre, S., 2023. Weakly Supervised Marine Animal Detection from Remote Sensing Images Using Vector-Quantized Variational Autoencoder. In: 2023 IEEE International Geoscience and Remote Sensing Symposium, IGARSS, pp. 5559–5562. https://doi.org/10.1109/IGARSS.2019.8898915.

Pörtner, H.O., Scholes, R.J., Agard, J., Archer, E., Arneth, A., Bai, X., Barnes, D., Burrows, M., Chan, L., Cheung, W.L., 2021. IPBES-IPCC co-sponsored workshop report on biodiversity and climate change. Zenodo. https://doi.org/10.5281/zenodo.5101133.

Qian, Y., Humphries, G.R.W., Trathan, P.N., Lowther, A., Donovan, C.R., 2023. Counting animals in aerial images with a density map estimation model. Ecol. Evol. 13, e9903. https://doi.org/10.1002/ece3.9903.

Rahnemoonfar, M., Dobbs, D., Yari, M., Starek, M.J., 2019. DisCountNet: discriminating and counting network for real-time counting and localization of sparse objects in high-resolution UAV imagery. Remote Sens. 11, 1128. https://doi.org/10.3390/rs11091128.

Razaak, M., Kerdegari, H., Argyriou, V., Remagnino, P., 2019. Multi-scale Feature Fused Single Shot Detector for Small Object Detection in UAV Images. In: Tzovaras, D., Giakoumis, D., Vincze, M., Argyros, A. (Eds.), Computer Vision Systems. ICVS 2019, Lecture Notes in Computer Science, LNCS, 11754. Springer, Cham, pp. 778–786. https://doi.org/10.1007/978-3-030-34995-0_71.

Redmon, J., Farhadi, A., 2018. Yolov3: An incremental improvement. arXiv Prepr. arXiv:1804.02767. https://doi.org/10.48550/arXiv.1804.02767.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You Only Look Once: Unified, Real-Time Object Detection. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 779–788. https://doi.org/10.1109/CVPR.2016.91.

Redmon, J., Farhadi, A., 2017. YOLO9000: Better, Faster, Stronger. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 6517–6525. https://doi.org/10.1109/CVPR.2017.690.

Ren, S.Q., He, K.M., Girshick, R., Sun, J., 2017. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39, 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.

Rivas, A., Chamoso, P., Gonzalez-Briones, A., Corchado, J.M., 2018. Detection of cattle using drones and convolutional neural networks. Sensors 18, 2048. https://doi.org/10.3390/s18072048.

Robinson, C., Ortiz, A., Hughey, L., Stabach, J., Ferres, J.M.L., 2021. Detecting cattle and elk in the wild from space. arXiv Prepr. arXiv:2106.15448. https://doi.org/10.48550/arXiv.2106.15448.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B., 2022. High-Resolution Image Synthesis With Latent Diffusion Models. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR. https://doi.org/10.2139/ssrn.4442721/CVPR52688.2022.01042.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (Eds.), Medical Image Computing and Computer-Assisted Intervention, MICCAI, Lecture Notes in Computer Science, LNCS. Springer, Cham, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.

Salehi, S.S.M., Erdogmus, D., Gholipour, A., 2017. In: Wang, Q., Shi, Y., Suk, HI., Suzuki, K. (Eds.), Machine Learning in Medical Imaging. MLMI 2017, Lecture Notes in Computer Science, LNCS, 10541. Springer, Cham, pp. 379–387. https://doi.org/10.1007/978-3-319-67389-9_44.

Sarwar, F., Griffin, A., Rehman, S.U., Pasang, T., 2021. Detecting sheep in UAV images. Comput. Electron. Agric. 187, 106219 https://doi.org/10.1016/j.compag.2021.106219.

Shao, W., Kawakami, R., Yoshihashi, R., You, S., Kawase, H., Naemura, T., 2020. Cattle detection and counting in UAV images based on convolutional neural networks. Int. J. Remote Sens. 41, 31–52. https://doi.org/10.1080/01431161.2019.1624858.

Skidmore, A.K., Pettorelli, N., Coops, N.C., Geller, G.N., Hansen, M., Lucas, R., Mücher, C.A., O'Connor, B., Paganini, M., Pereira, H.M., 2015. Environmental science: agree on biodiversity metrics to track from space. Nature 523, 403–405. https://doi.org/10.1038/523403a.

Skidmore, A.K., Coops, N.C., Neinavaz, E., Ali, A., Schaepman, M.E., Paganini, M., Kissling, W.D., Vihervaara, P., Darvishzadeh, R., Feilhauer, H., Fernandez, M., Fernández, N., Gorelick, N., Geijzendorffer, I., Heiden, U., Heurich, M., Hobern, D., Holzwarth, S., Muller-Karger, F.E., Van De Kerchove, R., Lausch, A., Leitão, P.J., Lock, M.C., Mücher, C.A., O'Connor, B., Rocchini, D., Roeoesli, C., Turner, W., Vis, J. K., Wang, T., Wegmann, M., Wingate, V., 2021. Priority list of biodiversity metrics to observe from space. Nat. Ecol. Evol. 5, 896–906. https://doi.org/10.1038/s41559-021-01451-x.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958. https://doi.org/10.5555/2627435.2670313.

Stapleton, S., LaRue, M., Lecomte, N., Atkinson, S., Garshelis, D., Porter, C., Atwood, T., 2014. Polar Bears from Space: Assessing Satellite Imagery as a Tool to Track Arctic Wildlife. PLoS ONE 9, e101513. https://doi.org/10.1371/journal.pone.0101513.

Stojnić, V., Risojević, V., Muštra, M., Jovanović, V., Filipi, J., Kezić, N., Babić, Z., 2021. A method for detection of small moving objects in UAV videos. Remote Sens. 13, 653. https://doi.org/10.3390/rs13040653.

Sun, X., Wang, P., Lu, W., Zhu, Z., Lu, X., He, Qibin, Li, J., Rong, X., Yang, Z., Chang, H., He, Qinglin, Yang, G., Wang, R., Lu, J., Fu, K., 2023. RingMo: A Remote Sensing Foundation Model With Masked Image Modeling. IEEE Trans. Geosci. Remote Sens. 61, 1–22. https://doi.org/10.1109/TGRS.2022.3194732.

Sundarama, D.M., Loganathan, A., 2020. FSSCaps-DetCountNet: fuzzy soft sets and CapsNet-based detection and counting network for monitoring animals from aerial images. J. Appl. Rem. Sens. 14, 026521 https://doi.org/10.1117/1.JRS.14.026521.

Torney, C.J., Lloyd-Jones, D.J., Chevallier, M., Moyer, D.C., Maliti, H.T., Mwita, M., Kohi, E.M., Hopcraft, G.C., 2019. A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images. Methods Ecol. Evol. 10, 779–787. https://doi.org/10.1111/2041-210X.13165.

Tuia, D., Kellenberger, B., Beery, S., Costelloe, B.R., Zuffi, S., Risse, B., Mathis, A., Mathis, M.W., van Langevelde, F., Burghardt, T., Kays, R., Klinck, H., Wikelski, M., Couzin, I.D., van Horn, G., Crofoot, M.C., Stewart, C.V., Berger-Wolf, T., 2022. Perspectives in machine learning for wildlife conservation. Nat. Commun. 13, 792. https://doi.org/10.1038/s41467-022-27980-y.

Turak, E., Brazill-Boast, J., Cooney, T., Drielsma, M., DelaCruz, J., Dunkerley, G., Fernandez, M., Ferrier, S., Gill, M., Jones, H., 2017. Using the essential biodiversity variables framework to measure biodiversity change at national scale. Biol. Conserv. 213, 264–271. https://doi.org/10.1016/j.biocon.2016.08.019.

Ulhaq, A., Adams, P., Cox, T.E., Khan, A., Low, T., Paul, M., 2021. Automated detection of animals in low-resolution airborne thermal imagery. Remote Sens. 13, 3276. https://doi.org/10.3390/rs13163276.

Ultralytics., 2023. ultralytics/ultralytics, https://github.com/ultralytics/ultralytics (accessed 11 January 2023).

van Gemert, J.C., Verschoor, C.R., Mettes, P., Epema, K., Koh, L.P., Wich, S., 2015. Nature conservation drones for automatic localization and counting of animals. Comput. vis. (LNCS) 8925, 255–270. https://doi.org/10.1007/978-3-319-16178-5_17.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. attention is all you need. arXiv Prepr. arXiv:1706.03762v7. https://arxiv.org/abs/1706.03762v7.

Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M., 2023. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7464–7475.

Wang, C.Y., Yeh, I.H., Liao, H.Y.M., 2024. YOLOv9: Learning what you want to learn using programmable gradient information. arXiv Prepr. arXiv: 2402.13616. https://doi.org/10.48550/arXiv.2402.13616.

Wang, G., Dong, G., Li, H., Han, L., Tao, X., Ren, P., 2019b. Remote sensing image synthesis via graphical generative adversarial networks. In: IEEE International Geoscience and Remote Sensing Symposium, pp. 10027–10030. https://doi.org/10.1109/IGARSS.2019.8898915.

Wang, D., Shao, Q., Yue, H., 2019a. Surveying wild animals from satellites, manned aircraft and unmanned aerial systems (UASs): a review. Remote Sens. 11, 1308. https://doi.org/10.3390/rs11111308.

Weinstein, B.G., Garner, L., Saccomanno, V.R., Steinkraus, A., Ortega, A., Brush, K., Yenni, G., McKellar, A.E., Converse, R., Lippitt, C.D., Wegmann, A., Holmes, N.D., Edney, A.J., Hart, T., Jessopp, M.J., Clarke, R.H., Marchowski, D., Senyondo, H., Dotson, R., White, E.P., Frederick, P., Ernest, S.K.M., 2022. A general deep learning model for bird detection in high-resolution airborne imagery. Ecol. Appl. 32, e2694. https://doi.org/10.1002/eap.2694.

Wu, Z., Zhang, C., Gu, X., Duporge, I., Hughey, L.F., Stabach, J.A., Skidmore, A.K., Hopcraft, J.G.C., Lee, S.J., Atkinson, P.M., 2023. Deep learning enables satellite-based monitoring of large populations of terrestrial mammals across heterogeneous landscape. Nat. Commun. 14, 3072. https://doi.org/10.1038/s41467-023-38901-y.

Xia, G.S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., Zhang, L., 2018. DOTA: a large-scale dataset for object detection in aerial images. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3974–3983. https://doi.org/10.1109/CVPR.2018.00418.

Xie, Y., Jiang, J., Bao, H., Zhai, P., Zhao, Y., Zhou, X., Jiang, G., 2023. Recognition of big mammal species in airborne thermal imaging based on YOLO V5 algorithm. Integr. Zool. 18, 333–352. https://doi.org/10.1111/1749-4877.12667.

Xu, B., Wang, W., Falzon, G., Kwan, P., Guo, L., Chen, G., Tait, A., Schneider, D., 2020a. Automated cattle counting using mask R-CNN in quadcopter vision system. Comput. Electron. Agric. 171, 105300 https://doi.org/10.1016/j.compag.2020.105300.

Xu, B., Wang, W., Falzon, G., Kwan, P., Guo, L., Sun, Z., Li, C., 2020b. Livestock classification and counting in quadcopter aerial images using mask R-CNN. Int. J. Remote Sens. 41, 8121–8142. https://doi.org/10.1080/01431161.2020.1734245.

Xue, M., Greenslade, T., Mirmehdi, M., Burghardt, T., 2022. Small or Far Away? Exploiting Deep Super-Resolution and Altitude Data for Aerial Animal Surveillance. In: IEEE/CVF Winter Conference on Applications of Computer Vision Workshops, WACVW, pp. 509–519. https://doi.org/10.1109/WACVW54805.2022.00057.

Xue, Y., Wang, T., Skidmore, A.K., 2017. Automatic counting of large mammals from very high resolution panchromatic satellite imagery. Remote Sens. 9, 878. https://doi.org/10.3390/rs9090878.

Yang, Z., Wang, T., Skidmore, A.K., de Leeuw, J., Said, M.Y., Freer, J., 2014. Spotting East African Mammals in Open Savannah from Space. PLoS ONE 9, e115989. https://doi.org/10.1371/journal.pone.0115989.

Yousefi, D.B.M., Rafie, A.S.M., Al-Haddad, S.A.R., Azrad, S., 2022. A Systematic Literature Review on the Use of Deep Learning in Precision Livestock Detection and Localization Using Unmanned Aerial Vehicles. IEEE Access 10, 80071–80091. https://doi.org/10.1109/ACCESS.2022.3194507.

Zhang, Y., Cai, Z., 2023. CE-RetinaNet: A Channel Enhancement Method for Infrared Wildlife Detection in UAV Images. IEEE Trans. Geosci. Remote Sens. 61, 1–12. https://doi.org/10.1109/TGRS.2023.3299651.

Zhang, H., Cissé, M., Dauphin, Y.N., Lopez-Paz, D., 2018. mixup: Beyond Empirical Risk Minimization. In: Proc. Int. Conf. Learn. Represent. (ICLR). https://doi.org/10.48550/arXiv.1710.09412.

Zheng, X., 2021. Self-supervised pretraining and controlled augmentation improve rare wildlife recognition in UAV images. In: IEEE International Conference on Computer Vision Workshops, pp. 732–741. https://doi.org/10.1109/ICCVW54120.2021.00087.

Zhou, X., Wang, D., Krähenbühl, P., 2019. Objects as points. arXiv Prepr. arXiv: 1904.07850. https://doi.org/10.48550/arXiv.1904.07850.

Zhu, P., Peng, T., Du, D., Yu, H., Zhang, L., Hu, Q., 2021. Graph regularized flow attention network for video animal counting from drones. IEEE Trans. Image Process. 30, 5339–5351. https://doi.org/10.1109/TIP.2021.3082297.