Ecological Informatics 86 (2025) 103009

Contents lists available at ScienceDirect

Ecological Informatics

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

Efficient wildlife monitoring: Deep learning-based detection and counting of green turtles in coastal areas

Naoya Noguchi [a], Hideaki Nishizawa [a,*], Taro Shimizu [a], Junichi Okuyama [b], Shohei Kobayashi [c,d], Kazuyuki Tokuda [e], Hideyuki Tanaka [e], Satomi Kondo [e]

[a] Graduate School of Informatics, Kyoto University, Japan
[b] Subtropical Coastal Research Group, Fisheries Technology Institute, Japan Fisheries Research and Education Agency, Japan
[c] Institute of Global Innovation Research, Tokyo University of Agriculture and Technology, Japan
[d] Japan Wildlife Research Center, Japan
[e] Everlasting Nature of Asia (ELNA), Ogasawara Marine Center, Japan

ARTICLE INFO

Keywords:
Abundance estimation
Coastal area
Deep learning
Sea turtle
Unmanned aerial vehicle
YOLOv7

ABSTRACT

Drones have recently been used to assess wildlife populations and their abundance. The automatic detection of target animals in drone footage enables efficient abundance estimation. However, accurately detecting animals remains challenging, especially in complex field environments. Moreover, automating the tracking of individuals across consecutive images and counting them along transect lines is necessary to apply drones to line-transect surveys. In this study, deep-learning-based You Look Only Once, Version 7 (YOLOv7) models were developed to automatically detect green turtles (*Chelonia mydas*) in Japanese coastal areas featuring coral reefs and seagrass beds. Drone footage yielded 103,296 annotated images of green turtles. The model was trained and validated using 78 % and 22 % of the images. The best model performances were 0.848, 0.853, and 0.922 for precision, recall, and mean average precision at the threshold of the intersection over union = 0.5, respectively. Then, the BoT-SORT object-tracking algorithm was implemented to track green turtles detected using the YOLOv7 model, and the counting of individuals was automated. When this automatic counting model was tested using eight drone footage clips, green turtles at the sea surface were successfully tracked and counted ($n = 3/3$); however, the performance in counting underwater green turtles was relatively poor ($n = 27/59$). The reduced performance might be attributable to accumulated errors in detecting green turtles while processing numerous images in the footage (approximately 60 fps). Nonetheless, relatively high precision was achieved by reducing false positives in complex coastal areas. The methods in this study should enhance the efficiency of long-term wildlife monitoring programs.

1. Introduction

Accurate and efficient monitoring of wild animal abundance is crucial for their conservation and management (Marques et al., 2013; Yoccoz et al., 2001). Estimating abundance using traditional methods, such as line transect surveys and capture-mark-recapture methods, is laborious and costly when conducting surveys at high frequency. Recently, technological and statistical advancement has enabled us to monitor wildlife efficiently using camera traps (Haucke et al., 2022; Johanns et al., 2022) and acoustics (Marques et al., 2013; Stevenson et al., 2015). Drones have also been commonly used to collect data on the population and abundance of both terrestrial (Corcoran et al., 2021; Ma et al., 2024; Vermeulen et al., 2013) and aquatic animals (Dunstan et al., 2020; Hodgson et al., 2013; Seymour et al., 2017). While drone-based methods have facilitated assessments with less cost and impact on animals, one of the major bottlenecks is the expertise and time required to manually analyze the imagery (Rees et al., 2018). Manually reviewing numerous images after a survey to determine their distribution and abundance is time-consuming. To overcome the problems related to the analysis of drone images, it is effective to automate

Abbreviations: CNN, Convolutional neural networks; YOLOv7, You Look Only Once, Version 7; png, Portable Network Graphics; AP, average precision; mAP, mean average precision; IoU, intersection over Union.

* Corresponding author.
E-mail address: nishizawa.hideaki.6s@kyoto-u.ac.jp (H. Nishizawa).

https://doi.org/10.1016/j.ecoinf.2025.103009
Received 1 August 2024; Received in revised form 8 January 2025; Accepted 9 January 2025
Available online 13 January 2025

methods for detecting and localizing target animals using deep-learning algorithms (Axford et al., 2024; Schad and Fischer, 2023).

Deep-learning animal detection from images shows promise in wildlife monitoring (Bakana et al., 2024; Roy et al., 2023). Recently, the technique has succeeded for drone imagery in detecting terrestrial mammals (Moreni et al., 2023; Peng et al., 2020), birds (Mpouziotas et al., 2023), and reptiles (Desai et al., 2022; Nath Tripathi et al., 2024), achieving 0.85 or more in an average precision or F1 score, indicating a balance between precision and recall. However, accurately detecting animals in complex environments (Moreni et al., 2023) or submerged marine animals (Colefax et al., 2019, 2021; Dujon et al., 2021) remains challenging. Among various deep-learning models applied to drone imagery (Axford et al., 2024), the You Look Only Once (YOLO) algorithm is characterized by its rapid object detection capabilities, making it suitable for analyzing extensive image datasets (Diwan et al., 2023). Additionally, the recent version of YOLO demonstrates satisfactory performance in detecting target objects within complex field environments (Chappidi and Sundaram, 2024; Rayner, 2024; Wu et al., 2022), such as the detection of gharials from drone images within a heterogenous riparian area (Nath Tripathi et al., 2024). Therefore, developing a YOLO-based model for processing drone-acquired images to detect marine animals in complex coastal areas should be explored further.

Automatic detection of animals has generally been applied to aerial images captured by drones, such as point-count surveys. However, for efficient line transection surveys using drones, automating counting individuals on lines is necessary. The tracking of individuals in consecutive images is important for this purpose. Although it is relatively straightforward to track an individual in a simple and unchanged background (e.g., laboratory conditions and images obtained by a stationary camera) (Dell et al., 2014), tracking individuals in natural environments using a moving camera remains a difficult task. Nonetheless, object tracking using a moving camera has been developed in the field of automotive safety (i.e., pedestrian tracking from a moving vehicle; Gavrila and Munder, 2007; Ragesh and Rajesh, 2019; Aharon et al., 2022) and has recently been applied to wildlife tracking (Jiang and Wu, 2024; Zhang et al., 2024). Therefore, applying such an object-tracking technique to moving drone images enables us to automate line transect surveys.

Monitoring the abundance of sea turtles is important because most sea turtle species are endangered, but some populations are recovering (Hays et al., 2024). Monitoring sea turtles at foraging grounds but not nesting beaches provides insights into whether foraging aggregations increase with nesting population recovery (Benson et al., 2007; Meylan et al., 2022; Seminoff et al., 2014); this is particularly important for green turtles (*Chelonia mydas*), which forage in coastal seagrass beds, where seagrass overgrazing has recently been reported on some foraging grounds (Christianen et al., 2014; Fourqurean et al., 2019). Automatic detection models of sea turtles from drone footage have been developed using deep learning tools, convolutional neural networks (CNN), olive ridley turtles (*Lepidochelys olivacea*) (Gray et al., 2019), and loggerhead turtles (*Caretta caretta*) (Dujon et al., 2021). However, the performance still needs to be improved [precision and recall: 0.163 and 0.765 (Gray et al., 2019); precision and recall: 0.20 and 0.75 (Dujon et al., 2021)]. Improvement in automatic detection and counting is necessary to detect green turtles that inhabit complex environments, including seagrass beds and coral reefs (Goatley et al., 2012).

In this study, a deep learning model was developed based on YOLOv7 (Wang et al., 2023) to automatically detect green turtles foraging in coastal areas using drone footage. The effects of the number of classes and transfer learning on the model's performance were investigated. Moreover, this study also developed and evaluated the automatic tracking and counting of detected green turtles using the BoT-SORT algorithm, which was developed for pedestrian tracking from a moving vehicle (Aharon et al., 2022).

2. Materials and methods

2.1. Drone image dataset collection

Drone image data were collected at a seagrass bed in Fukido River-mouth, Ishigakijima Island, Yaeyama Islands, Japan (24^{o}20' N, 124^{o}10' E) on August 4, October 25, and November 14, 2022. DJI Phantom 4 Pro v2, controlled by the Litchi application (https://flylitchi.com/), was used to travel along predefined autonomous routes with the angle of the gimbal camera of -90^{o} (nadir imagery) at a continuous speed of 3 m/s and a height of 30 m above sea level. The horizontal field of view was 45 m, and videos were recorded at 3840 × 2160 pixels with 59.94 fps. Surveys were completed between 8 and 9 a.m. and 4–5 p.m., when green turtles most frequently forage in this area (Okuyama et al., 2013) and avoid intense midday sun glare in the summer (Odzer et al., 2022).

Additionally, nadir imagery data of done were collected at Futami Bay, Chichijima Island, Ogasawara Islands, Japan (27^{o}5'N, 142^{o}12'E) at approximately 11 a.m. and 2 p.m. from January to July 2020, February to July 2021, and February to July 2022 using DJI Mavic 2 Zoom. The Ogasawara Islands are major rookeries for green turtles (Kondo et al., 2017) and have no major foraging grounds; however, both mature female and male green turtles aggregate in the bay mainly for mating and reproduction. The drone generally flowed along predefined routes at a continuous speed of 3 m/s and a height of 50 m above sea level. When green turtles were observed along the flight path, the drone hovered and zoomed temporarily to capture observational footage of the behaviors.

Since flights close to animals should also be avoided to reduce stress (Schad and Fischer, 2023), the altitude of the drone flight in this study was 20 m or higher; this follows previous sea turtle surveys using drones at altitudes of at least 20 m (Staines et al., 2022), generally 30–60 m (Dickson et al., 2022; Dujon et al., 2021; Dunstan et al., 2020).

2.2. Image processing and labeling

The overall framework for developing an automatic detection model for green turtles is illustrated in Fig. 1. The video footage obtained by the drones was manually checked for green turtles by NN and confirmed by HN. Because sea turtle species other than green turtles are extremely rare in the sites, underwater sea turtles that were unidentified clearly as green turtles were assumed to be green turtles. The drone video footage was converted to 4–5 images per second using the Portable Network Graphics (.png) format without reducing the resolution using QuickTime Player (Apple Inc.). Consequently, 8608 images of green turtles were obtained. Each image's outer rectangle of the target green turtle(s) was identified and annotated using LabelImg (https://github.com/HumanSignal/labelImg). Here, the following two patterns of labeling were applied: (1) all of the green turtles were labeled as one class "turtle" (13,566 turtles in 8608 images), and (2) green turtles at the sea surface (i.e., parts or the entire body are floating above the water) and underwater were labeled differently, "surface" and "underwater" classes (7232 and 6334 turtles), respectively. The 8608 images were augmented using horizontal mirroring, vertical mirroring, brightness enhancement, brightness reduction, and smoothing. The final augmented dataset consisted of 103,296 images, including 8608 original images.

2.3. Training and validation of the detection model

YOLOv7 (Wang et al., 2023) was adopted as the object detection model. The augmented datasets were split into a training dataset (78 %, n = 80,700/103,296) that was used to fit the network and a validation dataset (22 %, n = 22,596/103,296) that was used to estimate the network performance during the training process. The network was trained over 200 epochs with a batch size of eight. One epoch was input to the entire training dataset through the model during training.

This study developed two models: (1) without implementing transfer learning and (2) with transfer learning that used a CNN architecture pre-

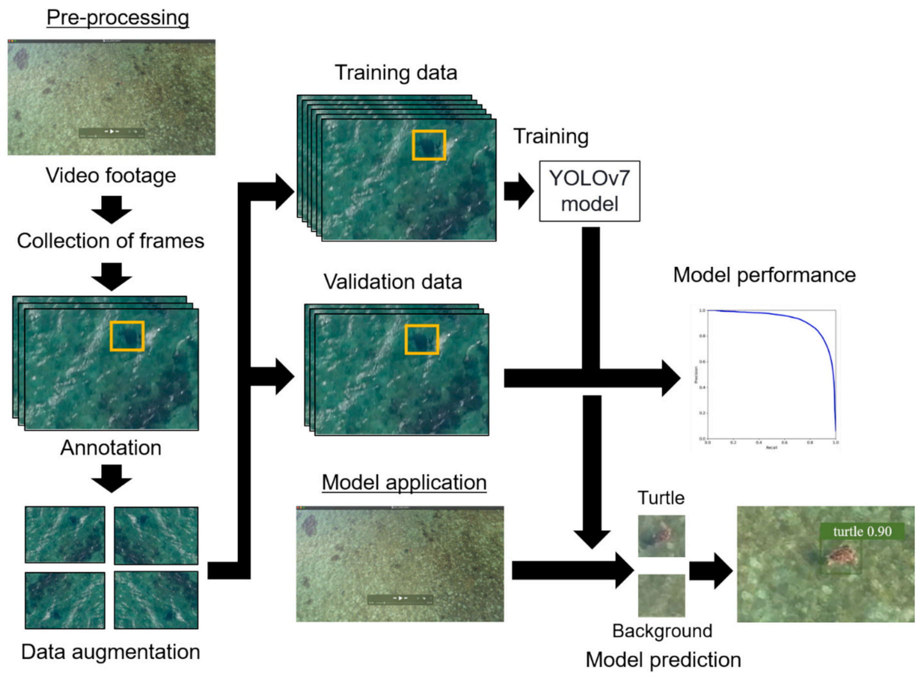

Fig. 1. Flow of developing and applying automatic detection of green turtles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

trained by the ImageNet dataset. Transfer learning is used to automatically extract features with the help of pre-trained models and potentially improve the performance of the model (Raza and Hong, 2020). The performances of these models were compared to evaluate the effectiveness of transfer learning.

In the validation of the YOLOv7-based model for detecting green turtles, the Intersection over Union (IoU) was defined as follows:

$$IoU = \frac{B \cap C}{B \cup C}$$

where B is the ground-truth bounding box of an object, and C is the predicted bounding box. In general, detections with IoU ≥ 0.50 were defined as true positives, and those with IoU < 0.50 as false positives. The model was evaluated using the precision, recall, and F1 scores that are calculated as

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

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

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

where TP, FP, and FN represent true positive (correctly identified objects), false positive (incorrectly identified objects), and false negative (objects that were not detected) detections, respectively.

The average precision (AP) is the average value of the highest precision (p) under different recall (r) conditions and is calculated as

$$AP = \int_0^1 \mathrm{p}(r)\mathrm{d}r$$

where the AP is calculated separately for each class. The mean average precision (mAP) is the AP value averaged across classes and is calculated as

$$mAP = \frac{1}{n}\sum_{i=1}^{n} APi$$

where n denotes the number of classes. In the case of one-class object detection, AP is equal to mAP. In addition to standard mAP when the threshold of IoU was 0.5 (hereafter, mAP@0.5), averaged mAP was calculated at varying thresholds of IoU from 0.5 to 0.95 (every 0.05; hereafter, mAP@0.5:0.95).

2.4. Object tracking and counting

By following a tracking-by-detection pipeline, an automatic counting model was developed for green turtles using drone footage based on the YOLOv7 and BoT-SORT algorithms (Fig. 2). First, the YOLOv7-based green turtle detector (developed in Section 2.3) was applied to each frame of the drone footage. The BoT-SORT algorithm was then used to track the target, which was detected as a green turtle. In consecutive frames, green turtles were not always successfully detected; however, the same track IDs were assigned when the turtle was detected again within 30 frames. The automatic carapace detection model (see Supplementary Material 1) was applied to the regions identified as green turtles. Detections for which the bounding boxes had an unrealistic aspect ratio for the green turtle carapace were eliminated using a ratio filter. The ratio of straight carapace length (SCL) to straight carapace width (SCW) of green turtles generally ranges from 1:1 to 1:1.5 (Eguchi et al., 2012; Salmon et al., 2018; Staines et al., 2022); the threshold of the ratio of length to width of the bounding box was set at 0.5 and 2.0, eliminating the detection with a ratio of <0.5 or > 2.0.

The total number of track IDs was calculated as the number of green turtles in each video. In consecutive frames, YOLOv7 may erroneously identify non-turtle objects as green turtles (i.e., false positives); however, false positives were assumed to be composed of fewer frames than green turtles. Therefore, a frame number filter was applied to eliminate false positives. By setting a threshold for the minimum number of frames in a track, tracks comprising smaller frames were eliminated. The number of false positives and negatives was evaluated by changing the threshold of the frame number filter from 0 to 200.

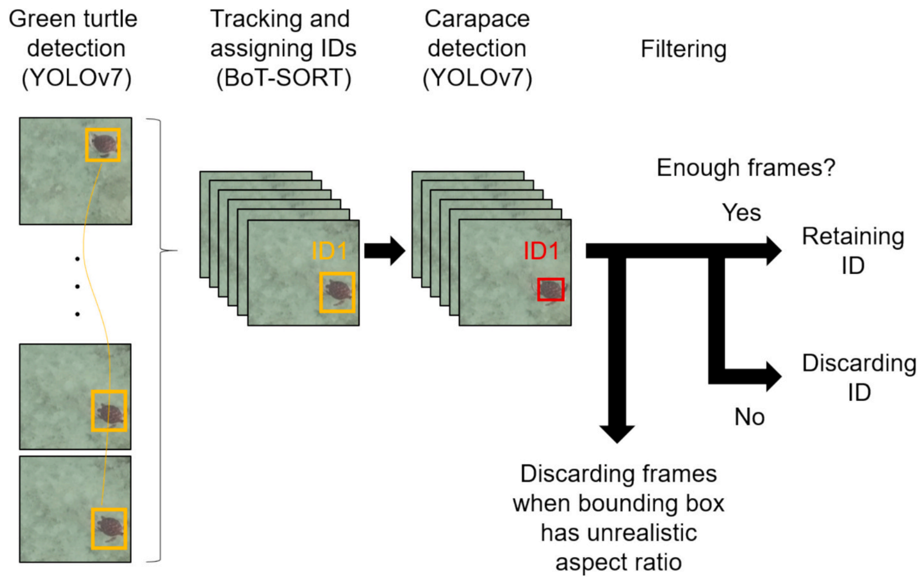

Fig. 2. Flow of automatic green turtle counting from drone footage based on the YOLOv7 and BoT-SORT algorithms. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.5. Testing of the model

The developed model was applied to newly obtained drone footage at the Fukido River mouth from July 5 to July 14, 2023, and five seagrass bed sites in Kumejima Island, Ryukyu Islands, Japan, from October 30 to November 4, 2023. A DJI Mavic 3 Pro drone controlled by the DJI Fly App was used to collect nadir imagery. The videos were recorded by a drone that flowed along predefined autonomous routes at a continuous speed of 3 m/s at heights of 20, 30, or 40 m above sea level from 8 to 9 a.m. or 4–5 p.m. when green turtles actively foraged (Okuyama et al., 2013) and the sun glare of the sea surface was limited (Odzer et al., 2022). Eight video footage clips were randomly selected, and the number of green turtles was manually counted. These footage clips were used to test the model. Precision and recall were calculated to evaluate the performance.

3. Results

3.1. Green turtle detection model

In the YOLOv7 models for detecting green turtles, the training and validation loss curves converged after 200 epochs, and no overfitting occurred during training (Supplementary Material 2). The simplest model that detects a single class "turtle" without implementing transfer learning showed the best performance with precision: 0.848, recall: 0.853, mAP@0.5: 0.922, and mAP@0.5:0.95: 0.620 (Table 1, Fig. 3). Precision and recall were balanced, and the F1 score reached 0.85 when the confidence was 0.538 (Fig. 3). The single-class detector that implemented transfer learning yielded slightly lower values of these metrics (Table 1). When the models were developed to detect two classes ("surface" and "underwater"), the performance decreased (Table 1).

3.2. Tracking and counting green turtles

The object detection part of the tracking and counting model implemented the simplest and best model in 3.1, which detects a single class "turtle" without implementing transfer learning. Across testing with the eight footage clips, manual observation by eye detected 62 green turtles (4–19 green turtles per footage), 3 and 59 of which were at the surface and underwater, respectively. When the frame number filter threshold was zero (i.e., no filter in which all detections were treated as green turtles once detected), the number of false negatives was minimized to eight, resulting in a recall of 0.871. However, many false positives ($n = 3619$) resulted in a precision of 0.015 (F1 score: 0.029). As the threshold of the frame number filter increased, the number of false positives decreased ($n = 12$ and precision = 0.625 when the threshold = 200); however, the number of false negatives increased (n = 42 and recall = 0.323 when the threshold = 200), resulting in F1 score = 0.426 (Fig. 4). False positives and negatives were balanced at the threshold 80–110 (precision: 0.462–0.492, recall: 0.452–484, F1 score: 0.471–0.488); therefore, the following results were obtained when the threshold was set to 90 that achieved the highest F1 score.

The model successfully tracked and counted green turtles at the sea surface (n = 3/3); however, tracking and counting underwater green turtles resulted in lower performance (n = 27/59) (Table 2). Lower performance was observed at sites with relatively large ripples or high turbidity (Fig. 5A, B). No true positives were detected in one of the eight footage clips (No. 4, n = 0/4). When the contrast between objects and background was clear under low turbidity conditions in sandy areas or when ripples were less pronounced due to cloudy conditions (Fig. 5C, D), performances were improved (No. 5–8, precision/recall: 0.429/0.750, 0.615/0.471, 0.480/0.632, and 0.500/0.600; F1 score: 0.533–0.546). Nonetheless, large number of false positives were detected when many coral bommies existed (No.7, Fig. 5D).

Table 1
Performance of models for one-class ("turtle") and two-class ("surface" and "underwater") detection with and without transfer learning (Y and N, respectively).

Model no.	Class	Transfer learning	Precision	Recall	F1 score	mAP@0.5	mAP@0.5:0.95
1	One	N	0.848	0.853	0.850	0.922	0.620
2	One	Y	0.847	0.844	0.845	0.921	0.615
3	Two	N	0.615	0.669	0.641	0.675	0.445
4	Two	Y	0.639	0.657	0.648	0.681	0.450

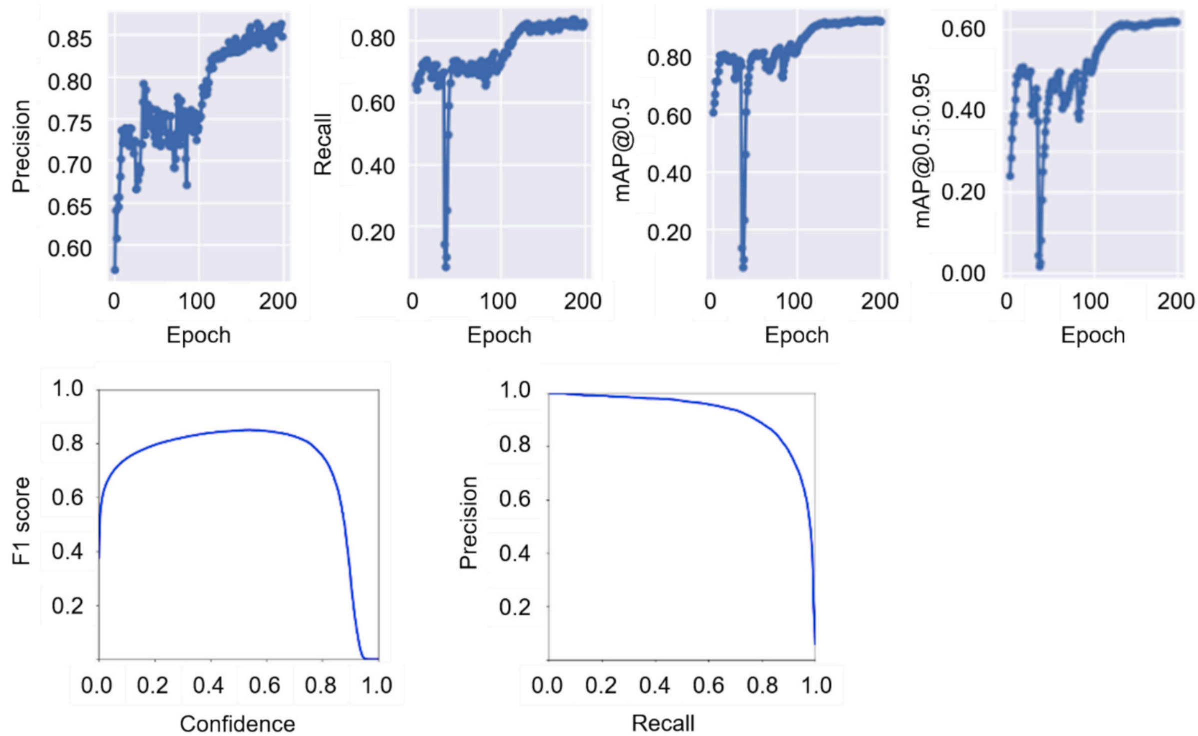

Fig. 3. Performance of the automatic green turtle detection model applied to the validation dataset (upper panels) (precision, recall, mAP@0.5, and mAP@0.5:0.95). The change in F1 score in relation to confidence, and the precision-recall curve in relation to the threshold of intersection over union (lower panels). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

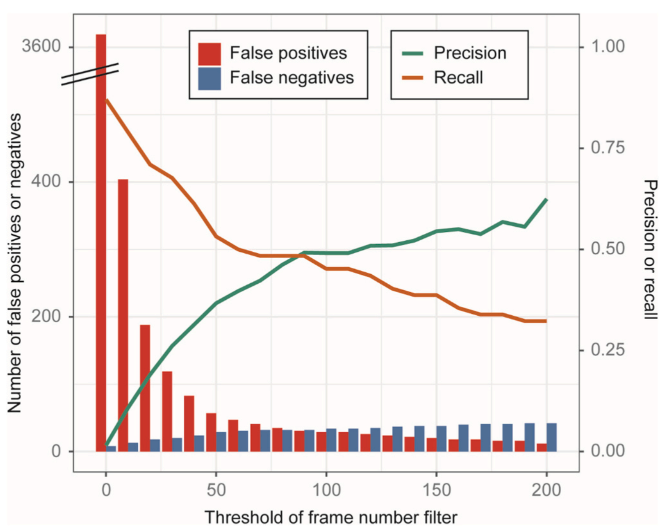

Fig. 4. False positives, false negatives, precision, and recall in relation to the threshold of the frame number filter in the automatic counting of green turtles applied to eight drone footage clips. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

4.1. Automatic detection of green turtles

Automatic detection of animals in complex field environments such as coastal areas from drone imagery is a key technique for efficient wildlife monitoring (Axford et al., 2024; Schad and Fischer, 2023). In detecting green turtles, the YOLOv7-based automatic detection model developed in this study achieved a precision of 0.848 when detecting green turtles as one class. The performance of our detection and localization model is also comparable to that of previous studies using drones (reviewed by Axford et al., 2024). Compared to previously developed CNN-based automatic detection models for sea turtles, which showed a precision of 0.163 (confidence = 0.93) (Gray et al., 2019) and 0.20 (confidence = 0.55) (Dujon et al., 2021), this model lowered the frequency of false positives. One possible factor that could improve precision is the amount and variation of the training datasets. In this study, the number of training images with data augmentation was 16.1 times and 9.1 times larger than 5010 images in Dujon et al. (2021) and 8826 images in Gray et al. (2019), respectively. In addition, training data were collected from various sites, including images of juveniles (Ishigakijima Island) and adults (Chichijima Island). We confirmed that training images of marine animals of various sizes on complex backgrounds improve the detection model (Dujon et al., 2021).

Transfer learning is generally considered an effective method for improving the model (Raza and Hong, 2020); however, transfer learning using ImageNet did not improve the model in this study. This may be because of the low feature similarity between the training datasets in the source and target domains. Detecting green turtles in drone imagery necessitates identifying local variations within the image. However, ImageNet data often contain a distinct global subject within the image (see Supplementary Material 3). The differences in resolution, shape, size, and location of objects in ImageNet compared to the image data obtained in this study may not improve model performance, as reported in medical imaging tasks (Raghu et al., 2019). Additionally, although transfer learning can accelerate convergence on a task, it does not consistently improve the performance of models trained on large datasets (He et al., 2019). Transfer learning may not yield observable improvements because of the utilization of relatively large training data. In transfer learning applications, it is essential to consider the characteristics and quantity of training data.

An increase in classification classes generally decreases the accuracy

Table 2
Performance of automatic counting of surface and underwater green turtles (model count vs. manual count) in eight drone footage clips from Fukido Rivermouth in Ishigakijima Island (Site A) and Kumejima Island (Sites B–E2) with a minimum frame threshold of 90.

No.	Site	Time	Altitude (m)	Surface	Underwater	FN	FP	Precision	Recall	F1 score
1	A	Jul. 13, 2023	30	1/1 (100 %)	0/3 (0 %)	3	2	0.333	0.250	0.286
2	A	Oct. 24, 2023	30	1/1 (100 %)	1/3 (33 %)	2	2	0.500	0.500	0.500
3	A	Oct. 24, 2023	20	0/0 (–)	1/5 (20 %)	4	2	0.333	0.200	0.250
4	B	Nov. 1, 2023	30	0/0 (–)	0/4 (0 %)	4	0	–	0.000	–
5	C	Oct. 31, 2023	30	0/0 (–)	3/4 (75 %)	1	4	0.429	0.750	0.546
6	D	Oct. 30, 2023	30	0/0 (–)	8/17 (47 %)	9	5	0.615	0.471	0.533
7	E1	Nov. 2, 2023	40	1/1 (100 %)	11/18 (61 %)	7	13	0.480	0.632	0.546
8	E2	Nov. 1, 2023	30	0/0 (–)	3/5 (60 %)	2	3	0.500	0.600	0.545
Total				3/3 (100 %)	27/59 (46 %)	32	31	0.492	0.484	0.488

FN: false negative, FP: false positive.

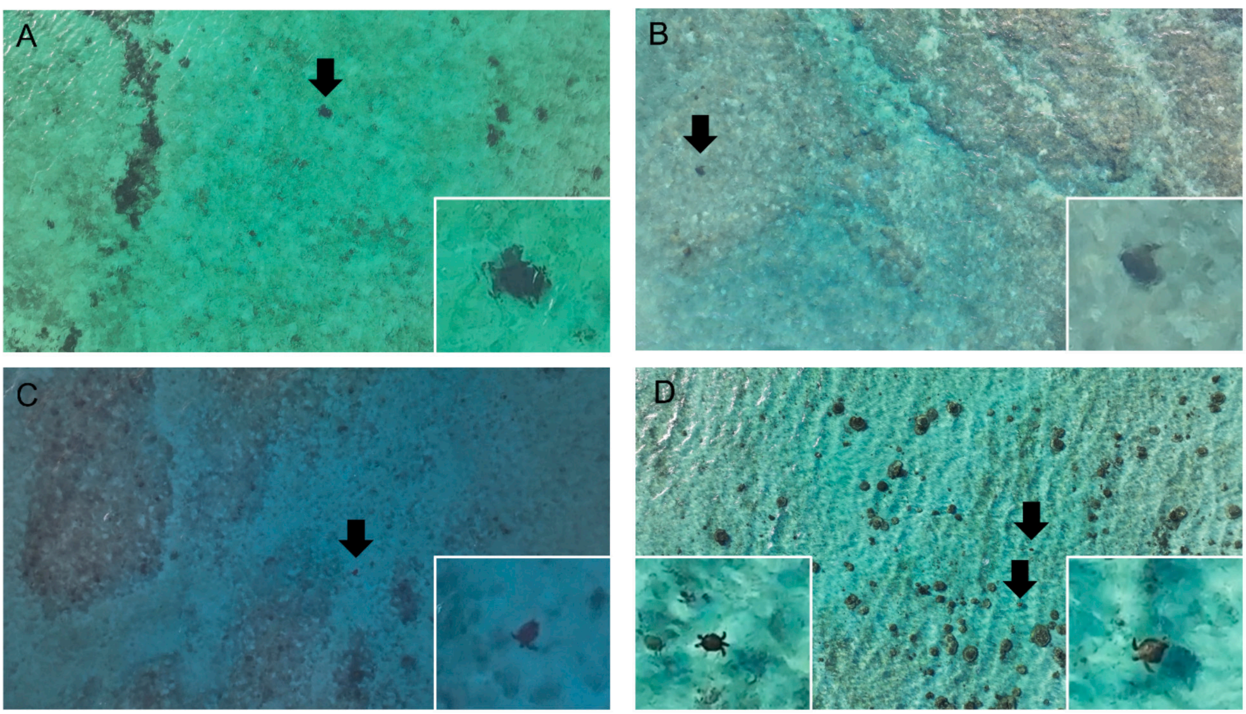

Fig. 5. Examples of drone imagery exhibiting unclear contrast between turtles and backgrounds because of ripples and turbidity (A and B from footage clips No. 2 and 4, respectively) and with clear contrast under cloudy conditions (C from footage clip No. 5) and within a sandy area (D from footage clip No. 7). Black arrows indicate turtles in the original drone images, with enlarged insets at the corners.

of the class probability P (class | object). Even if the object detection probability, P (object), is relatively high, the classification accuracy, which is defined as P (class | object) multiplied by P (object), may be low (e.g., Bjerge et al., 2023). Compared with the one-class detection model, the model classifying green turtle status as "surface" or "underwater" had lower performance in this study. The dataset exhibited minimal bias among classes; therefore, we did not implement class weighting, oversampling, or specific data augmentation to mitigate class imbalance. However, the comparatively lower number of "underwater" turtles in the training dataset may hinder the detection of underwater turtles. In addition, the contrast between objects and background decreases with increasing water depth (Colefax et al., 2019, 2021), complicating the detection of underwater animals (Dujon et al., 2021). The clarity of the turtles' edge and carapaces' brightness are critical characteristics for distinguishing between "surface" and "underwater." However, clarity and brightness are influenced by wave activity, turbidity, and weather conditions (Colefax et al., 2019, 2021). The data augmentation by brightness modification and smoothing might enhance object detection probability but reduce classification between "surface" and "underwater" turtles.

4.2. Automatic counting and tracking of green turtles

Accurate wildlife tracking in complex field environments using drone footage remains a major challenge in ecological informatics (Jiang and Wu, 2024; Zhang et al., 2024). Although the object detection model showed relatively high performance, the automatic counting model based on BoT-SORT resulted in relatively low performance. Since the detection of carapaces in green turtles was almost perfect (see Supplementary Material 1), the performance of the counting model was reduced by the accumulated errors in detecting green turtles in consecutive frames. Object detection is image-based; however, the counting model processes numerous images obtained by drones (approximately 60 fps), resulting in relatively high false positives and false negatives. Additionally, surface individuals are assumed to be detected more easily and continuously than underground individuals, which may be detected intermittently. Even if green turtles were successfully detected, detecting the same turtles intermittently resulted in the reassignment of track IDs and increased false positives (i.e., double counting). Therefore, selecting a threshold for the frame number filter is important based on the objectives of surveys and analyses.

The performance of the model varied across sites. High detection rates can be achieved when animals exist in uniform or homogeneous backgrounds; however, more false positives occur in heterogeneous habitats (Dujon and Schofield, 2019). False positives may increase when coral reefs or seagrass patches are similar in shape, colour, and size to green turtles, as observed in No. 7 (site E1). In addition, some green turtles on coral reefs were not successfully detected when their heads and limbs were blended with corals (site B). However, objects tend to be successfully detected when contrasted sharply with the background (Chabot and Francis, 2016; Patterson et al., 2015) when the drone footage was taken in sandy areas and/or cloudy conditions (No. 5 and site E). The ripples changed the shapes of the underwater turtles in consecutive images, likely complicating the tracking process. Low-altitude drone flight captures large objects in images; however, shape distortion caused by ripples may also be amplified. Therefore, lower altitudes may not improve the automatic tracking of objects (No. 3). The successful detection of surface green turtles can be attributed to the clarity of the edges in the background.

The BoT-SORT algorithm predicts the location of objects in consecutive frames using a Kalman filter module confined to linear motions, possibly reducing the robustness of object tracking in filed environments (Jiang and Wu, 2024; Zhang et al., 2024). Extending a Kalman filter to nonlinearity (Jiang and Wu, 2024) and/or introducing a ground motion compensation that corrects target position through feature matching of consecutive frames (Zhang et al., 2024) may improve tracking performance. Preprocessing videos, such as dehazing and enhancement based on spatio-temporal inter-frame relationships (Du et al., 2024; Qing et al., 2016) may also improve the performance. In addition, improving the YOLO model will enhance tracking performance. Adding attention modules is one possible solution to clarify the contrast between turtles and backgrounds (Feng and Jin, 2024). Nonetheless, it is difficult to detect objects that are not included in the training dataset. Three green turtles attached to remoras in the testing drone footage collected on Ishigakijima Island in October were not successfully counted because they were either detected in a limited number of frames or not detected. Therefore, adding these images and various images of underwater targets to the training datasets is required to further generalize the model.

4.3. Surveys of green turtles using drones

This study developed a deep learning model based on YOLOv7 that automatically detects green turtles in drone images. The model had relatively high precision for detecting green turtles in coastal areas with complex backgrounds. In addition, a model was developed to automatically count green turtles from drone footage using BoT-SORT. The counting model successfully counted green turtles at the sea surface; however, counting underwater green turtles showed relatively low performance. When applying the counting model, selecting an appropriate threshold depending on the objectives and sites is important. Currently, fully automated counting of green turtles from transecting drone videos remains difficult. However, implementing a semi-automated model that assists human visual inspection and provides stochastic abundance estimations based on the output of the model is recommended (Kellner et al., 2023). Infrared imaging can improve the detection accuracy of terrestrial animals (Ma et al., 2024) but does not work well for submerged animals (Colefax et al., 2021). Using a hyperspectral sensor mounted on a drone, which accentuates green colors, enhances the contrast between underwater animals and their background (Colefax et al., 2021); therefore, incorporating these techniques in future studies would increase the detectability of underwater turtles.

At present, "surface" and "underwater" turtles are classified based on whether parts or the entire body are floating above the water because of the difficulty in detecting the actual depth of target sea turtles. However, since the detectability is considered to decrease increasing water depth, the change in performance with the depth of target animals should be further evaluated to explore the applicability. Collecting drone images of sea turtles deployed depth recorders (Agabiti et al., 2024) or animal decoys set to know depths will help reveal the relationship. The use of animal decoys is a promising technique not only for understanding the detectability of targets in various substrate types and backgrounds (Benavides et al., 2020; Odzer et al., 2022), but also for augmentation of training data (Hong et al., 2019) that may improve detecting underwater turtles.

Occlusion (i.e., overlapping of individuals) is a problem in automatically detecting aggregating and schooling animals (Brack et al., 2018; Chabot and Francis, 2016; Moreni et al., 2023). This problem was not considered in this study because green turtles in the images were not heavily aggregated but could potentially become apparent when applying the model to mating turtles and large aggregations (Schofield et al., 2017). As the YOLO model struggles with small objects under high-density conditions (Redmon et al., 2016), detection may become difficult in the presence of occlusions. In such cases, it is necessary to adjust the CNN to allow the detection of multiple nearby turtles. For instance, the detection performance of dense, small target objects can be improved using the Soft NMS method (Bodla et al., 2017) to combine multiple weighted frames according to the confidence level of the IoU.

The altitude and time of the drone flight are important factors that influence detection. Flight at low altitudes provides images with clear objects that may improve detection probability but reduce the coverage of survey areas in one flight (Patterson et al., 2015) and possibly induce animal stress (Schad and Fischer, 2023). The altitude of drone flights should be selected to minimize the impact on animals. In this study, the selected altitude, based on previous studies (20 m or more, Dunstan et al., 2020, Dujon et al., 2021, Dickson et al., 2022, Staines et al., 2022), did not induce obvious escape responses from green turtles. In this study, the model was trained using images captured by drones at 30 m and 50 m, thereby enhancing its applicability to footage featuring animals of varying sizes from different altitudes. To reduce false positives associated with glare on the sea surface (Gray et al., 2019; Odzer et al., 2022), the drone was operated in the morning (8–9 a.m.) and evening (4–5 p.m.) during summer and autumn in this study. Avoiding sun glare considerably limits the available flying time. Therefore, selecting the appropriate altitude and timing remains a key issue in applying drone surveys to large-scale monitoring.

The model developed in this study focuses on green turtles. Other species of sea turtles were extremely rare in the study sites, but multiple species occasionally share the habits (e.g., Staines et al., 2022; Stokes et al., 2023). The model's applicability to the other species and species classification from images warrants further investigation. Visual identification of sea turtle species is possible through drone imagery via head (Staines et al., 2022) or carapace (Stokes et al., 2023) dimensions; therefore, classifying different sea turtle species, and other coastal animals, can be achieved by using additional training datasets. Upgrading the model in this study will facilitate the detection of various species in different backgrounds.

5. Conclusion

This study developed a detection and counting model for green turtles that will aid conventional manual monitoring, enabling a rapid understanding of the abundance of green turtles, even in complex coastal habitats, and achieving relatively high precision by minimizing false positives in complex coastal areas. It is also a pioneering work involving marine animal tracking and counting. The workflow developed in this study, which combines automatic detection, tracking, and counting, will apply to transect monitoring of other wildlife using drones in coastal and terrestrial habitats. Object-tracking is important for monitoring research areas with moving drones and recording moving animals with fixed-point cameras. Future reconstructions of animal trajectories using object tracking will provide insights into animal behavior. The methods developed in this study are anticipated to contribute to efficient long-term monitoring programs that facilitate understanding of population dynamics and conservation management.

CRediT authorship contribution statement

Naoya Noguchi: Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Conceptualization. **Hideaki Nishizawa:** Writing – review & editing, Writing – original draft, Investigation, Conceptualization. **Taro Shimizu:** Writing – review & editing, Resources, Formal analysis. **Junichi Okuyama:** Writing – review & editing, Resources, Investigation, Conceptualization. **Shohei Kobayashi:** Writing – review & editing, Resources, Investigation. **Kazuyuki Tokuda:** Writing – review & editing, Resources, Investigation. **Hideyuki Tanaka:** Writing – review & editing, Resources, Investigation. **Satomi Kondo:** Writing – review & editing, Resources, Investigation.

Declaration of competing interest

The authors declare that they have no conflicts of interest.

Acknowledgments

We thank the staff and volunteers of the Ogasawara Marine Center, Subtropical Coastal Research Group of Fisheries Technology Institute, Fisheries Cooperative Association on Kume Island, and Kumejima Town for their assistance with fieldwork. This study was partially supported by the Asahi Glass Foundation and JSPS KAKENHI (Grant Number 20K20007). We acknowledge anonymous reviewers for their constructive comments on the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2025.103009.

Data availability

The programs and code used in this study are available on GitHub (https://github.com/NishizawaHideaki/TurtleCount). The images and drone footage data used in this study are available from doi:https://doi.org/10.5061/dryad.1g1jwsv6z.

References

Agabiti, C., Tolve, L., Baldi, G., Zucchini, M., Tuccio, S., Restelli, F., Freggi, D., Luschi, P., Casale, P., 2024. Combining UAVs and multi-sensor dataloggers to estimate fine-scale sea turtle density at foraging areas: a case study in the Central Mediterranean. Endanger. Species Res. 54, 395–408. https://doi.org/10.3354/esr01350.

Aharon, N., Orfaig, R., Bobrovsky, B.-Z., 2022. BoT-SORT: robust associations multi-pedestrian tracking. arXiv:2206.14651. https://doi.org/10.48550/arXiv.2206.14651.

Axford, D., Sohel, F., Vanderklift, M.A., Hodgson, A.J., 2024. Collectively advancing deep learning for animal detection in drone imagery: successes, challenges, and research gaps. Eco. Inform. 83, 102842. https://doi.org/10.1016/j.ecoinf.2024.102842.

Bakana, S.R., Zhang, Y., Twala, B., 2024. WildARe-YOLO: a lightweight and efficient wild animal recognition model. Eco. Inform. 80, 102541. https://doi.org/10.1016/j.ecoinf.2024.102541.

Benavides, M.T., Fodrie, F.J., Johnston, D.W., 2020. Shark detection probability from aerial drone surveys within a temperate estuary. J. Unmann. Veh. Syst. 8, 44–56. https://doi.org/10.1139/juvs-2019-0002.

Benson, S.R., Forney, K.A., Carretta, J.V., Dutton, P.H., 2007. Abundance, distribution, and habitat of leatherback turtles (*Dermochelys coriacea*) off California, 1990−2003. Fish. Bull. 105, 337–347.

Bjerge, K., Geissmann, Q., Alison, J., Mann, H.M.R., Høye, T.T., Dyrmann, M., Karstoft, H., 2023. Hierarchical classification of insects with multitask learning and anomaly detection. Eco. Inform. 77, 102278. https://doi.org/10.1016/j.ecoinf.2023.102278.

Bodla, N., Singh, B., Chellappa, R., Davis, L.S., 2017. Soft-NMS—Improving object detection with one line of code. In: In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017, pp. 5561–5569.

Brack, I.V., Kindel, A., Oliveira, L.F.B., 2018. Detection errors in wildlife abundance estimates from unmanned aerial systems (UAS) surveys: synthesis, solutions, and challenges. Methods Ecol. Evol. 9, 1864–1873.

Chabot, D., Francis, C.M., 2016. Computer-automated bird detection and counts in high-resolution aerial images: a review. J. Field Ornithol. 87, 343–359.

Chappidi, J., Sundaram, D.M., 2024. Enhanced animal detection in complex outdoor environments using modified Yolo V7. Int. J. Intell. Syst. Appl. Eng. 12, 375–382. https://ijisae.org/index.php/IJISAE/article/view/5076.

Christianen, M.J.A., Herman, P.M.J., Bouma, T.J., Lamers, L.P.M., van Katwijk, M.M., van der Heide, T., Mumby, P.J., Silliman, B.R., Engelhard, S.L., van de Kerk, M., Kiswara, W., van de Koppel, J., 2014. Habitat collapse due to overgrazing threatens turtle conservation in marine protected areas. Proc. Roy. Soc. B 281, 20132890.

Colefax, A.P., Butcher, P.A., Pagendam, D.E., Kelaher, B.P., 2019. Reliability of marine faunal detections in drone-based monitoring. Ocean Coast. Manag. 174, 108–115. https://doi.org/10.1016/j.ocecoaman.2019.03.008.

Colefax, A.P., Kelaher, B.P., Walsh, A.J., Purcell, C.R., Pagendam, D.E., Cagnazzi, D., Butcher, P.A., 2021. Identifying optimal wavelengths to maximise the detection rates of marine fauna from aerial surveys. Biol. Conserv. 257, 109102. https://doi.org/10.1016/j.biocon.2021.109102.

Corcoran, E., Denman, S., Hamilton, G., 2021. Evaluating new technology for biodiversity monitoring: are drone surveys biased? Ecol. Evol. 11, 6649–6656. https://doi.org/10.1002/ece3.7518.

Dell, A.I., Bender, J.A., Branson, K., Couzin, I.D., de Polavieja, G.G., Noldus, L.P.J.J., Pérez-Escudero, A., Perona, P., Straw, A.D., Wikelski, M., Brose, U., 2014. Automated image-based tracking and its application in ecology. Trends Ecol. Evol. 29, 417–428. https://doi.org/10.1016/j.tree.2014.05.004.

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. Eco. Inform. 72, 101874. https://doi.org/10.1016/j.ecoinf.2022.101874.

Dickson, L.C.D., Tugwell, H., Katselidis, K.A., Schofield, G., 2022. Aerial drones reveal the dynamic structuring of sea turtle breeding aggregations and minimum survey effort required to capture climatic and sex-specific effects. Front. Mar. Sci. 9, 864694. https://doi.org/10.3389/fmars.2022.864694.

Diwan, T., Anirudh, G., Tembhume, J.V., 2023. Object detection using YOLO: challenges, architectural successors, datasets and applications. Multimed. Tools Appl. 82, 9243–9275. https://doi.org/10.1007/s11042-022-13644-y.

Du, D., Li, E., Si, L., Xu, F., Niu, J., 2024. End-to-End Underwater Video Enhancement: Dataset and Model. arXiv:2403.11506. https://doi.org/10.48550/arXiv.2403.11506.

Dujon, A.M., Schofield, G., 2019. Importance of machine learning for enhancing ecological studies using information-rich imagery. Endanger. Species Res. 39, 91–104. https://doi.org/10.3354/esr00958.

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.

Dunstan, A., Robertson, K., Fitzpatrick, R., Pickford, J., Meager, J., 2020. Use of unmanned aerial vehicles (UAVs) for mark-resight nesting population estimation of adult female green sea turtles at Raine Island. PLoS One 15 (6), e0228524. https://doi.org/10.1371/journal.pone.0228524.

Eguchi, T., Seminoff, J.A., LeRoux, R.A., Prosperi, D., Dutton, D.L., Dutton, P.H., 2012. Morphology and growth rates of the green sea turtle (*Chelonia mydas*) in a northern-most temperate foraging ground. Herpetologica 68, 76–87. http://www.bioone.org/doi/full/10.1655/HERPETOLOGICAD-11-00050.1.

Feng, J., Jin, T., 2024. CEH-YOLO: a composite enhanced YOLO-based model for underwater object detection. Eco. Inform. 82, 102758. https://doi.org/10.1016/j.ecoinf.2024.102758.

Fourqurean, J.W., Manuel, S.A., Coates, K.A., Massey, S.C., Kenworthy, W.J., 2019. Decadal monitoring in Bermuda shows a widespread loss of seagrasses attributable to overgrazing by the green sea turtle *Chelonia mydas*. Estuar. Coasts 42, 1524–1540.

Gavrila, D.M., Munder, S., 2007. Multi-cue pedestrian detection and tracking from a moving vehicle. Int. J. Comput. Vis. 7, 41–59. https://doi.org/10.1007/s11263-006-9038-7.

Goatley, C.H.R., Hoey, A.S., Bellwood, D.R., 2012. The role of turtles as coral reef macroherbivores. PLoS One 7 (6), e39979. https://doi.org/10.1371/journal.pone.0039979.

Gray, P.C., Fleishman, A.B., Klein, D.J., McKown, M.W., Bézy, V.S., Lohmann, K.J., Johnston, D.W., 2019. 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.

Haucke, T., Kühl, H.S., Hoyer, J., Steinhage, V., 2022. Overcoming the distance estimation bottleneck in estimating animal abundance with camera traps. Eco. Inform. 68, 101536. https://doi.org/10.1016/j.ecoinf.2021.101536.

Hays, G.C., Schofield, G., Papazekou, M., Chatzimentor, A., Katsanevakis, S., Mazaris, A. D., 2024. A pulse check for trends in sea turtle numbers across the globe. iScience 27, 109071. https://doi.org/10.1016/j.isci.2024.109071.

He, K., Girshick, R., Dollár, P., 2019. Rethinking ImageNet pre-training. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 4917–4926. https://doi.org/10.1109/ICCV.2019.00502.

Hodgson, A., Kelly, N., Peel, D., 2013. Unmanned aerial vehicles (UAVs) for surveying marine fauna: a dugong case study. PLoS One 8 (11), e79556. https://doi.org/10.1371/journal.pone.0079556.

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 (7), 1651. https://doi.org/10.3390/s19071651.

Jiang, L., Wu, L., 2024. Enhanced Yolov8 network with extended Kalman filter for wildlife detection and tracking in complex environments. Eco. Inform. 84, 102856. https://doi.org/10.1016/j.ecoinf.2024.102856.

Johanns, P., Haucke, T., Steinhage, V., 2022. Automated distance estimation for wildlife camera trapping. Eco. Inform. 70, 101734. https://doi.org/10.1016/j.ecoinf.2022.101734.

Kellner, K.F., Smith, A.D., Royle, J.A., Kéry, M., Belant, J.L., Chandler, R.B., 2023. The unmarked R package: twelve years of advances in occurrence and abundance modelling in ecology. Methods Ecol. Evol. 14, 1408–1415. https://doi.org/10.1111/2041-210X.14123.

Kondo, S., Morimoto, Y., Sato, T., Suganuma, H., 2017. Factors affecting the long-term population dynamics of green turtles (*Chelonia mydas*) in Ogasawara, Japan: influence of natural and artificial production of hatchlings and harvest pressure. Chelonian Conserv. Biol. 16, 83–92.

Ma, G., Li, W., Bao, H., Roberts, N.J., Li, Y., Zhang, W., Yang, K., Jiang, G., 2024. UAV equipped with infrared imaging for Cervidae monitoring: improving detection accuracy by eliminating background information interference. Eco. Inform. 81, 102651. https://doi.org/10.1016/j.ecoinf.2024.102651.

Marques, T.A., Thomas, L., Martin, S.W., Mellinger, D.K., Ward, J.A., Moretti, D.J., Harris, D., Tyack, P.L., 2013. Estimating animal population density using passive acoustics. Biol. Rev. 88, 287–309. https://doi.org/10.1111/brv.12001.

Meylan, P.A., Hardy, R.F., Gray, J.A., Meylan, A.B., 2022. A half-century of demographic changes in a green turtle (*Chelonia mydas*) foraging aggregation during an era of seagrass decline. Mar. Biol. 169, 74. https://doi.org/10.1007/s00227-022-04056-5.

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 (4), e0284449. https://doi.org/10.1371/journal.pone.0284449.

Mpouziotas, D., Karvelis, P., Tsoulos, I., Stylios, C., 2023. Automated wildlife bird detection from drone footage using computer vision techniques. Appl. Sci. 13, 7787. https://doi.org/10.3390/app13137787.

Nath Tripathi, R., Ramachandran, A., Agarwal, K., Tripathi, V., Badola, R., Ainul Hussain, S., 2024. UAV and deep learning: detection of selected riparian species along the Ganga River. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. XLVIII-1-2024. https://doi.org/10.5194/isprs-archives-XLVIII-1-2024-637-2024.

Odzer, M.N., Brooks, A.M.L., Heithaus, M.R., Whitman, E.R., 2022. Effects of environmental factors on the detection of subsurface green turtles in aerial drone surveys. Wildl. Res. 49 (1), 79–88. https://doi.org/10.1071/WR20207.

Okuyama, J., Nakajima, K., Noda, T., Kimura, S., Kamihata, H., Kobayashi, M., Arai, N., Kagawa, S., Kawabata, Y., Yamada, H., 2013. Ethogram of immature green turtles: behavioral strategies for somatic growth in large marine herbivores. PLoS One 8 (6), e65783. https://doi.org/10.1371/journal.pone.0065783.

Patterson, C., Koski, W., Pace, P., McLuckie, B., Bird, D.M., 2015. Evaluation of an unmanned aircraft system for detecting surrogate caribou targets in Labrador. J. Unmann. Veh. Syst. 4, 53–69.

Peng, J., Wang, D., Liao, X., Shao, Q., Sun, Z., Yue, H., Ye, H., 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.

Qing, C., Yu, F., Xu, X., Huang, W., Jin, J., 2016. Underwater video dehazing based on spatial–temporal information fusion. Multidim. Syst. Sign. Process. 27, 909–924. https://doi.org/10.1007/s11045-016-0407-2.

Ragesh, N.K., Rajesh, R., 2019. Pedestrian detection in automotive safety: understanding state-of-the-art. IEEE Access 7, 47864–47890. https://doi.org/10.1109/ACCESS.2019.2909992.

Raghu, M., Zhang, C., Kleinberg, J., Bengio, S., 2019. Transfusion: understanding transfer learning for medical imaging. Adv. Neural Inf. Proces. Syst. 32. https://doi.org/10.48550/arXiv.1902.07208.

Rayner, J.S., 2024. Machine Learning of Large Scale Imagery for Wildlife Conservation. Master's Thesis. Queensland University of Technology, Brisbane, Australia.

Raza, K., Hong, S., 2020. Fast and accurate fish detection design with improved YOLO-v3 model and transfer learning. Int. J. Adv. Comput. Sci. Appl. 11 (2), 7–16. https://doi.org/10.14569/IJACSA.2020.0110202.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2016, pp. 779–788.

Rees, A.F., Avens, L., Ballorain, K., Bevan, E., Broderick, A.C., Carthy, R.R., Christianen, M.J.A., Duclos, G., Heithaus, M.R., Johnston, D.W., Mangel, J.C., Paladino, F., Pendoley, K., Reina, R.D., Robinson, N.J., Ryan, R., Sykora-Bodie, S.T., Tilley, D., Varela, M.R., Whitman, E.R., Whittock, P.A., Wibbels, T., Godley, B.J., 2018. The potential of unmanned aerial systems for sea turtle research and conservation: a review and future directions. Endanger. Species Res. 35, 81–100. https://doi.org/10.3354/esr00877.

Roy, A.M., Bhaduri, J., Kumar, T., Raj, K., 2023. WilDect-YOLO: an efficient and robust computer vision-based accurate object localization model for automated endangered wildlife detection. Eco. Inform. 75, 101919. https://doi.org/10.1016/j.ecoinf.2022.101919.

Salmon, M., Mott, C.R., Bresette, M.J., 2018. Biphasic allometric growth in juvenile green turtles *Chelonia mydas*. Endanger. Species Res. 37, 301–308. https://doi.org/10.3354/esr00930.

Schad, L., Fischer, J., 2023. Opportunities and risks in the use of drones for studying animal behaviour. Methods Ecol. Evol. 14, 1864–1872. https://doi.org/10.1111/2041-210X.13922.

Schofield, G., Katselidis, K.A., Lilley, M.K.S., Reina, R.D., Hays, G.C., 2017. Detecting elusive aspects of wildlife ecology using drones: new insights on the mating dynamics and operational sex ratios of sea turtles. Funct. Ecol. 31, 2310–2319. https://doi.org/10.1111/1365-2435.12930.

Seminoff, J.A., Eguchi, T., Carretta, J., Allen, C.D., Prosperi, D., Rangel, R., Gilpatrick Jr., J.W., Forney, K., Peckham, S.H., 2014. Loggerhead Sea turtle abundance at a foraging hotspot in the eastern Pacific Ocean: implications for at-sea conservation. Endanger. Species Res. 24, 207–220. https://doi.org/10.3354/esr00601.

Seymour, A.C., Dale, J., Hammill, M., Halpin, P.N., Johnston, D.W., 2017. Automated detection and enumeration of marine wildlife using unmanned aircraft systems (UAS) and thermal imagery. Sci. Rep. 7, 45127. https://doi.org/10.1038/srep45127.

Staines, M.N., Smith, C.E., Madden Hof, C.A., Booth, D.T., Tibbetts, I.R., Hays, G.C., 2022. Operational sex ratio estimated from drone surveys for a species threatened by climate warming. Mar. Biol. 169, 152. https://doi.org/10.1007/s00227-022-04141-9.

Stevenson, B.C., Borchers, D.L., Altwegg, R., Swift, R.J., Gillespie, D.M., Measey, G.J., 2015. A general framework for animal density estimation from acoustic detections across a fixed microphone array. Methods Ecol. Evol. 6, 38–48. https://doi.org/10.1111/2041-210X.12291.

Stokes, H.J., Mortimer, J.A., Laloë, J.-O., Hays, G.C., Esteban, N., 2023. Synergistic use of UAV surveys, satellite tracking data, and mark-recapture to estimate abundance of elusive species. Ecosphere 14, e4444. https://doi.org/10.1002/ecs2.4444.

Vermeulen, C., Lejeune, P., Lisein, J., Sawadogo, P., Bouché, P., 2013. Unmanned aerial survey of elephants. PLoS One 8 (2), e54700. https://doi.org/10.1371/journal.pone.0054700.

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: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2023, pp. 7464–7475.

Wu, D., Jiang, S., Zhao, E., Liu, Y., Zhu, H., Wang, W., Wang, R., 2022. Detection of *Camellia oleifera* fruit in complex scenes by using YOLOv7 and data augmentation. Appl. Sci. 12 (22), 11318. https://doi.org/10.3390/app122211318.

Yoccoz, N.G., Nichols, J.D., Boulinier, T., 2001. Monitoring of biological diversity in space and time. Trends Ecol. Evol. 16, 446–453. https://doi.org/10.1016/S0169-5347(01)02205-4.

Zhang, Z., Zhao, Y., Fu, P., Luo, W., Shao, Q., Zhang, T., Yu, Z., 2024. A reliable unmanned aerial vehicle multi-target tracking system with global motion compensation for monitoring *Procapra przewalskii*. Eco. Inform. 81, 102556. https://doi.org/10.1016/j.ecoinf.2024.102556.