Ecological Informatics 91 (2025) 103380

Contents lists available at ScienceDirect

Ecological Informatics

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

Integrating AI into ecology for fully automated monitoring of endangered seabird breeding colonies

Inbal Schekler[a,b,c,*,1], Eyal Halabi[a,1], Nir Sapir[b], Ilan Shimshoni[a], Ohad Hatzofe[f], Yoav Perlman[c], Yosef Kiat[c,d,e]

[a] Department of Information Systems, University of Haifa, Haifa, Israel
[b] Department of Evolutionary and Environmental Biology and Institute of Evolution, University of Haifa, Haifa, Israel
[c] BirdLife Israel, Society for the Protection of Nature in Israel (SPNI), Tel-Aviv, Israel
[d] School of Zoology, Faculty of Life Sciences, Tel-Aviv University, Tel-Aviv, Israel
[e] Steinhardt Museum of Natural History, Tel-Aviv University, Tel-Aviv, Israel
[f] Science Division, Israel Nature and Parks Authority (NPA), Jerusalem, Israel

ARTICLE INFO

Keywords:
Bird conservation
Remote sensing and monitoring
You only look once (YOLOv8) algorithm
Deep learning (DL)
Computer vision
Colonial seabirds
Species ID

ABSTRACT

The rapid pace of human-related development and environmental instability has placed significant pressure on global ecosystems, threatening the existence of numerous threatened wildlife species, including birds. Accurate and detailed monitoring can play a crucial role in safeguarding biodiversity, supporting conservation efforts and enhancing ecosystem management. Traditional monitoring methods, such as direct observation and manual image analysis, are labor-intensive, prone to bias, and often inadequate for large and densely populated breeding colonies. In this study, we developed a fully automated deep-learning-based algorithm to identify, count, and map breeding seabirds in a large and densely populated mixed breeding colony containing breeding pairs of two visually similar species, the regionally-threatened Common Tern (*Sterna hirundo*) and the Little Tern (*Sternula albifrons*). Using YOLOv8 for initial object detection using remote-controlled cameras, we enhanced classification performance by integrating ecological and behavioral features, including spatial fidelity, movement patterns and size, through camera calibration techniques. Our algorithm successfully identified, counted, and mapped breeding individuals from both species with an average discrepancy of only 2 % compared to manual counts, and achieved over 90 % accuracy in correctly identifying the species of nesting individuals from both species. By providing high-resolution spatial mapping of nesting individuals, the system also offers valuable insights into habitat use and intra-colony dynamics. Additionally, incorporating behavioral tracking via video analysis allows for a more accurate differentiation between nesting and non-nesting individuals. Our methodology represents a significant advancement in automated wildlife monitoring by integrating artificial intelligence for automatic counting, mapping and classifying birds to enhance our understanding of ecological processes and to aid conservation. This study presents a robust, automated framework for seabird colony monitoring that minimizes human disturbance while maximizing accuracy and efficiency. By leveraging deep learning and ecological knowledge, this approach can revolutionize conservation monitoring, offering a scalable and cost-effective solution for tracking wildlife populations in an era of rapid environmental changes.

1. Introduction

The rapid pace of human-related development has placed significant pressure on ecosystems across the world, threatening numerous species (IUCN, 2022; Phillips et al., 2023). Addressing this challenge has become main focus of conservation efforts worldwide, with many countries implementing measures to protect animal populations and ensure their long-term survival. Seabirds are one of the most threatened groups of birds, with almost half of seabird species experiencing population declines (Croxall et al., 2012; Phillips et al., 2023). Understanding seabird population trends is vital for preserving both these species and the marine ecosystems they inhabit (Croxall et al., 2012). These

* Corresponding author.
E-mail address: goldinbal@gmail.com (I. Schekler).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.ecoinf.2025.103380
Received 4 March 2025; Received in revised form 5 August 2025; Accepted 5 August 2025
Available online 7 August 2025

seabirds, many of them extreme long-distance migrants, face an array of threats, including coastal and marine contamination, depletion of fish stocks, and climate changes, which adversely affect their nesting sites, foraging areas, reproductive outcomes, and survival rates (Brierley and Kingsford, 2009; Fay et al., 2015; Oro et al., 2004).

Effective monitoring strategies play a pivotal role in species conservation. When successfully implemented, they enable the detection of population changes and provide timely opportunities to address threats before a species faces extinction. Ideally, monitoring serves as a tool for evaluating the outcomes of various conservation measures within an adaptive management framework (Walsh et al., 1995) and supports the implementation of policy reforms. In the absence of robust monitoring data, it becomes challenging to prioritize research efforts, assess the success of management interventions, and provide decision-making guidelines (Scheele et al., 2018; Verdon et al., 2024). However, current monitoring practices predominantly rely on traditional methods such as on-site observations using binoculars or telescopes and manual camera monitoring (Pugesek and Stehn, 2016; Verdon et al., 2024). These approaches are time-intensive, laborious, and prone to limitations, including restricted spatial coverage. Moreover, the accuracy and completeness of the collected data can be compromised by observer bias, variations in expertise, and temporal inconsistencies of the monitoring (Legg and Nagy, 2006). Furthermore, approaching and undertaking the monitoring actions at close proximity to nesting colonies may disturb the birds and challenge their standardized monitoring, further complicating conservation efforts (Blackmer et al., 2004). The inefficiencies and inconsistencies of these widely used methods underscore the need for consistent, automated and more effective techniques to substantially improve wild animal monitoring, thereby enhancing conservation initiatives.

In recent years, advancements and developments in technological tools have transformed the way animals and biodiversity can be monitored and studied (August et al., 2015; Bridge et al., 2011; Delisle et al., 2021; Ghosh et al., 2023a; Ghosh et al., 2023b; Ghosh et al., 2024; Jiang and Wu, 2024; Pan et al., 2025). Various tools enable collection of large datasets with minimal time and resource investment or to process the data collected more efficiently. Methods such as aerial photography, unmanned aerial vehicle (UAV) surveys (Chabot et al., 2015; Noguchi et al., 2025; Weissensteiner et al., 2015), and the use of fixed position cameras for images and videos (Bolton et al., 2007; Kline et al., 2025; Schindler and Steinhage, 2021; Simões et al., 2023; Williams and DeLeon, 2020), all non-invasive methods, can improve the efficiency and accuracy of bird monitoring. These methods also help reduce the disturbance to the natural behavior of the birds in their breeding colonies, which are often particularly sensitive to human presence and other anthropogenic disturbances (Edney and Wood, 2021; Richter et al., 2018). However, animal monitoring through images and videos is however inefficient as it still heavily relies on manual image analysis, requiring the meticulous review of photographs and the counting of each species present. Some approaches employ semi-automated classification, which involves identifying a unique spectral signature for the target object to detect its occurrences within the image (Hodgson et al., 2018; Schwaller et al., 1989). However, this method proves challenging for high-density breeding sites or species that lack strong visual contrast with their environment, such as cormorants on dark rocks. Fully automated machine-learning algorithms have also been developed to identify birds in images (Jiang et al., 2025; Jones et al., 2020; Williams and DeLeon, 2020), but these usually require vast datasets of pre-annotated images (Jones et al., 2020). Moreover, many of these systems rely on single-frame analyses, which can lead to inaccuracies in breeding counts, as non-breeding birds may be miscounted, or breeding individuals may be overlooked if absent from the frame at the time of capture. Animal behavior research has also benefited from recent advances in deep learning (Saad Saoud et al., 2024), particularly through the development of models capable of estimating animal pose for behavioral analysis (Mathis et al., 2018; Zhang et al., 2023). These models enable the extraction of behavioral metrics such as movement rates and posture patterns, which can serve as indicators of an animal's overall well-being (Yang et al., 2024). However, many of these methods have been developed for domestic animals (An et al., 2023; Bhuiyan and Wree, 2023; Fang et al., 2021) or under controlled laboratory conditions (Forys et al., 2020; Segalin et al., 2021; Xiao et al., 2023). In the context of wildlife behavior monitoring, several studies have focused on detecting and classifying animals using drones (Cusick et al., 2024; Noguchi et al., 2025; Rančić et al., 2023; Schad and Fischer, 2023), which enable large-scale population surveys but do not support continuous behavioral monitoring. Camera traps, on the other hand, which are also widely used and increasingly analyzed using deep learning methods (Fennell et al., 2022; Johanns et al., 2022; Leorna et al., 2022; Song et al., 2024), do allow for continuous monitoring but are typically limited to a small number of individuals within restricted areas. Consequently, there is an urgent need for fully automated methods that can improve both the precision and efficiency of bird monitoring, enabling the study of fundamental behavioral traits through continuous, large-scale monitoring, such as at breeding colonies, for conservation purposes.

The terns studied in this project include the vulnerable Common Tern (*Sterna hirundo*) and the endangered Little Tern (*Sternula albifrons*) in Israel (Mayrose et al., 2017). Over recent decades, tern populations in Israel experienced significant changes, most notably the loss of suitable breeding habitats. This has led to the concentration of most of the breeding population into a single, large and dense colony that now consists approximately 90 % of the local breeding population for these species. Such aggregation heightens the population's vulnerability to catastrophic events, including predation and disease outbreaks, while potentially intensifying both interspecific and intraspecific competition (Boulinier and Lemel, 1996; Lewison et al., 2012). The breeding colony is located on an island within a large salt production facility that includes large pools, adjacent to the town of Atlit on the Mediterranean Sea coast of Northern Israel. The goals of this study are to develop an automated algorithm capable of (1) distinguishing between two visually similar species, (2) identifying and differentiating the behaviors of breeding individuals, (3) counting only breeding individuals, and (4) mapping the precise locations of breeders from both species. By employing a deep learning approach (YOLOv8), we monitored the breeding site by scanning it four times daily using two remotely controlled cameras, providing detailed reports on the number of breeding birds of each species and their spatial distribution. While cameras have been employed for seabird colony monitoring for several decades, to our knowledge, this study represents a novel application of fully automated methods for monitoring seabird colonies.

2. Materials and methods

2.1. Manual breeding populations counts

Manual annual counts of breeding Common and Little Terns from 2011 to 2024 were compiled from reports generated by the Atlit tern conservation project (Hatzofe and Mayrose, 2009; Kiat and Schekler, 2021; Ribak et al., 2017; Schekler et al., 2015; Schekler et al., 2016). These counts adhered to the methodologies described in the respective reports, which included observations conducted from a vehicle, a tower adjacent to the colony, or one or two remotely controlled cameras. The counts were carried out by professional observers, who changed over the years. Breeding counts were conducted approximately once a week throughout the breeding season, from April to August. The highest count recorded during the season was used as the final estimate of the size of the breeding population in each year. We performed an ANOVA analysis to evaluate whether the counting methods influenced the breeding population size estimates for each species using R software (version 4.0.3, R Core Team).

2.2. Data collection, processing and tagging

Data collection from the tern colony was carried out in 2023 and 2024 using two off-the-shelf remotely controlled cameras positioned on the island. The camera locations were selected to ensure full coverage and clear visibility of the colony, with particular emphasis on monitoring the Little Tern breeding area near the southern camera (Fig. 1S). This setup allowed for detailed observation and recording of the colony's dynamics, extent and additional properties (see below). The cameras (DH-SD5A232XA-HNR 2MP × 32 Starlight IR WizSense Network PTZ) were equipped with features such as remote horizontal and vertical movement, variable speed control, × 32 zoom, and online accessibility. The system allowed for setting up to 300 preset positions with specific zoom, yaw and pitch settings, enabling automated scanning sessions at predefined intervals and times between the present positions location, defined as "flags". Additionally, the cameras enabled the identification of individual birds through coded leg rings, nest counts, observations of adults, chicks, deceased individuals, identification of threats such as predators (e.g., Red Fox *Vulpes vulpes* or Hooded Crow *Corvus cornix*) and rising water levels that may harm nests found near the edges of the island. Automated scans were conducted four times daily at 10:00, 10:17, 15:00, and 15:17. Each camera focused on its designated area, with combined coverage of the entire island. The northern camera scanned 63 preset flags, while the southern camera covered 47 (Fig. 1S), pausing for 15 s at each flag. Scan durations were 15.75 min for the northern camera and 11.75 min for the southern. At the beginning and end of each scan, the cameras captured wide-angle views to assess potential threats. The cameras used in this study have not been observed to cause any disturbance to the birds. All scans are conducted silently during daylight hours, thereby eliminating potential disturbance from both infrared illumination (which is inactive during the day) and acoustic noise. All data were recorded and stored for analysis.

To train and refine a model for distinguishing between Common and Little Terns, tagged images from the colony were utilized. Tagging was performed using the RectLabel annotation tool (https://rectlabel.com/), creating a dataset of 1659 images derived from both cameras' flags. These images were split into training (70 %), validation (20 %), and testing (10 %) datasets. Seven categories were defined for tagging: Common Tern standing (1533 examples), sitting (3218), flying (61); Little Tern standing (82), sitting (466); chicks (538); and "other" (16 examples, including non-target species in the colony).

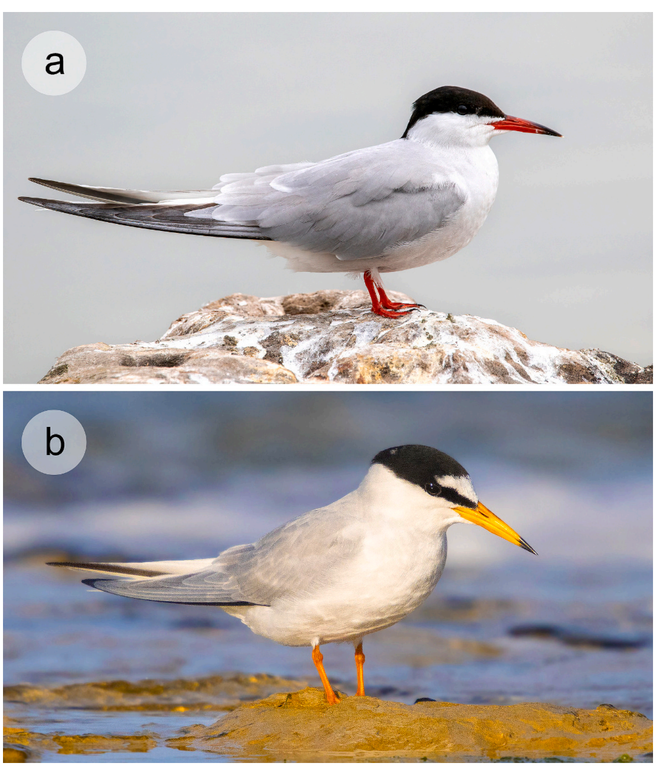

Fig. 1. The Common Tern (*Sterna hirundo*; a) and the Little Tern (*Sternula albifrons*; b). Their distinct physical characteristics includes the colour of the beak (red in the Common Tern, yellow in the Little Tern), the shape and colour of the forehead (entirely black in the Common Tern, black and white in the Little Tern), and relative size (the Common Tern is approximately twice as large). Common Tern image by Alexis Lours; Little Tern image by Julio Jesús Añel Perez (Cornell Lab of Ornithology, Macaulay Library). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.3. YOLOv8 network architecture

YOLOv8 (You Only Look Once version 8) is one of the latest generations of the YOLO series, designed for tasks such as object detection, image classification, and instance segmentation. Released by Ultralytics in January 2023, YOLOv8 builds on the architecture of YOLOv5 with substantial improvements in performance and usability (Varghese and Sambath, 2024). Known for its impressive accuracy and compact model size, YOLO models are accessible and efficient, as they can be trained on a single GPU, making them suitable for a broad spectrum of developers. Additionally, YOLOv8 can be fine-tuned for custom datasets, enabling enhanced performance tailored to specific applications and datasets, like in the case of our study where we fine-tuned it for tagged terns images. YOLOv8 introduces enhanced architectural features and optimizations, and evaluations on the COCO dataset demonstrate that YOLOv8 outperforms its predecessors (Varghese and Sambath, 2024).

The architecture of YOLOv8 is structured around four key components: the input, backbone, neck, and head, each optimized for efficiency and accuracy. The backbone has been enhanced with modifications such as replacing the 6 × 6 convolutional layer with a 3 × 3 layer and substituting the C3 module with the more efficient C2f module, improving feature extraction. The neck, responsible for aggregating and refining features across scales, eliminates the 1 × 1 convolutional sampling layer and also integrates the C2f module. In the head, a decoupled architecture separates classification and regression tasks, enabling more precise predictions. This design outputs bounding boxes for object localization, class predictions to identify object types, and objectness scores to estimate detection confidence. However, it is important to note that YOLOv8 does not inherently identify the size of the detected objects, nor does it utilize size information for classification purposes. This limitation means that while YOLOv8 excels in detecting and classifying objects based on learned features, it does not provide explicit measurements of object dimensions. This can be a critical consideration for applications requiring size-specific analyses, such as in our case, where one of the primary differences between the two similar tern species is their size. To improve model generalization and robustness, data augmentation was applied automatically during training using the YOLOv8 framework's default augmentation pipeline. The augmentation techniques included random horizontal flipping (50 % probability), random adjustments to image hue, saturation, and brightness (HSV augmentation), scaling (random zoom in/out), translation (random shifting in the x and y directions), minor rotation, and shear transformations. In addition, YOLOv8 applies advanced augmentation strategies such as Mosaic augmentation, which combines four images into one to enrich context diversity, and MixUp, which blends pairs of images and their labels with a weighted ratio. Several trials were conducted using different manual adjustments to the augmentation parameters, including varying the strength and probability of each transformation. However, these modifications did not result in improved model performance. Consequently, the default automatic augmentation provided by YOLOv8 was adopted for the final model training. The augmentation was disable in the final training cycles, enhancing the model's focus and improving accuracy in detecting real-world scenarios. YOLOv8 provides five model variants of increasing width and depth (YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x). Trials were conducted using several of these models (YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x). Given that the targets in this study were of moderate size, with no need to detect tiny objects, the lightweight YOLOv8s was found to sufficiently meet accuracy requirements while offering superior computational efficiency. Larger models did not yield noticeable improvements in detection performance and came at the cost of significantly higher computational demands.

2.4. Algorithm key components

To accurately detect and count breeding pairs of two closely related seabird species in a dense colony, we developed an algorithm by enhancing YOLOv8. Our approach addresses the challenge of distinguishing between visually similar species (Fig. 1) that nest in proximity, as well as differentiating breeding birds from non-breeding individuals for precise monitoring. This differentiation is achieved by

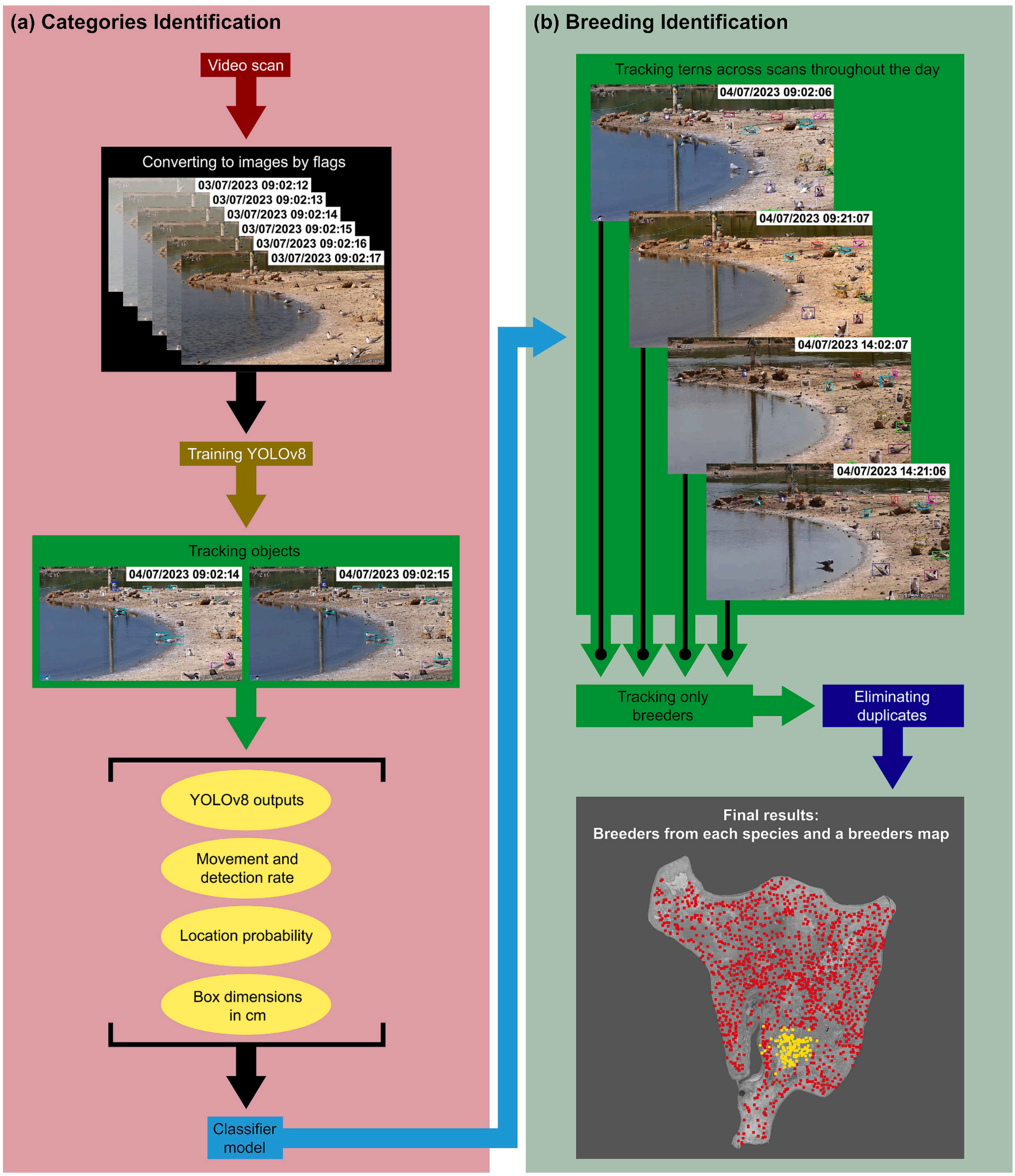

Fig. 2. Schematic representation of the algorithm for seabird breeder identification, counting and mapping. (a) Training phase for the identification of key categories, including Common Tern (standing, sitting, and flying), Little Tern (standing and sitting), chicks, and background classes. (b) Post training phase for identifying breeding individuals of Common Tern and Little Tern, followed by automated counting and spatial mapping of their locations.

tracking the exact locations of breeders throughout the day and leveraging behavioral traits characteristic of breeding seabirds like terns, which typically remain on their nests for long periods of time during the incubation period (Skutch, 1957). Furthermore, the algorithm generates georeferenced outputs, mapping the location of breeding pairs daily, thereby enabling spatial visualization of colony dynamics alongside accurate identification and counting of breeding terns. Our algorithm includes the following workflow and key components (Fig. 2):

2.4.1. Converting the video scans to images

In the first step, we converted the scan videos into images corresponding to the specific flags along the video timeline. Each flag was recorded for 15 s, during which images were captured at one-second intervals. We excluded the initial and final frames of the 15-s sequence, as they involved camera movement and focusing, and used 11 frames per flag for analysis.

2.4.2. Training YOLOv8 for species detection

The model was trained on tagged images to differentiate between Common Terns and Little Terns based on their distinct morphological characteristics. These include the colour of the beak which is red in the case of the Common Tern and yellow in the case of the Little Tern, and the shape and colour of the forehead which is all black for the Common Tern, and a combination of white and black for the Little Tern (Fig. 1). Another important distinction between the two species is their size, with the Common Tern being approximately twice the size of the Little Tern. While this size difference is fundamental for identification in the field, it cannot be utilized by YOLO for classification. The model was also trained on various behaviors, such as sitting, standing, and flying, to enable the identification of breeders, which spend most of the incubation period sitting. Additionally, it was trained to recognize chicks and other bird species occasionally present on the island, such as the Black-winged Stilt (*Himantopus himantopus*) and Spur-winged Lapwing (*Vanellus spinosus*). This comprehensive training forms the foundation for accurately distinguishing and localizing individual terns within the breeding colony.

2.4.3. Tracking objects

To classify objects accurately, we implemented a tracker to follow detected objects within each flagged area of the colony. Object tracking was achieved using the Intersection over Union (IoU) metric to correlate bounding boxes of detected objects between consecutive frames. IoU, a standard metric in object tracking, measures the degree of overlap between two bounding boxes. For our application, IoU scores represented the spatial consistency of an object's position across frames. A bounding box was associated with a track when its IoU score exceeded a threshold of 0.4. This threshold was determined empirically by testing a range of values to balance two opposing considerations: setting the threshold too low increased the risk of mistakenly linking two nearby individuals into a single track, while setting it too high increased the risk of generating multiple tracks for the same individual due to small frame-to-frame movements. Tracks were discarded if they could not be maintained for at least four frames. Valid tracks were integrated into the classification process. For each track, the mean IoU score of bounding boxes was calculated to define a movement rate. Additionally, a detection rate was derived as the proportion of frames in which the object appeared within its track.

2.4.4. Calculating box size in cm

To enhance species differentiation accuracy, we incorporated the physical size of the bounding box (in cm) into the classifier. Since Common Terns are approximately twice the size of Little Terns (Billerman et al., 2025), size can serve as a very important feature for classification. To determine real-world dimensions, the cameras were calibrated using a drone image of the colony anchored to geographic coordinates. By utilizing the drone image along with recorded zoom levels, pitch, and yaw for each flag, we calculated the precise location of each image pixel on the island. This calibration allowed us to convert pixel dimensions into centimeters, enabling the actual size of each bounding box to be calculated. This allowed incorporating it as a feature in the classifier.

2.4.5. Location probability

Seabirds, including terns, exhibit strong fidelity to their nesting locations over time (Blums et al., 2002; Doherty et al., 2002; Patrick and Weimerskirch, 2017). In this colony, Little Terns have consistently nested in specific areas for over 15 years, with only one recorded shift (Ribak et al., 2017). Location probabilities for each species were derived from tagged images and, where tagging was unavailable, YOLOv8 model predictions. These probabilities, specific to each flag, were added as a feature to improve classification.

2.5. Classifier model

The training dataset consisted of 111 tagged images, comprising 1793 annotated bounding boxes across eight categories: Common Tern sit (1431), Common Tern stand (221), Little Tern sit (84), Little Tern stand (3), chicks (12), background (38), and other (2). A background category was specifically introduced to address the YOLOv8 model's initial misclassification of stones. Each tracked bounding box was characterized by a range of features (Fig. 2a), including the final classifier integrated outputs from the YOLOv8 model, movement and detection rates calculated during the tracking process, location probabilities for each species, and box dimensions in centimeters, including width, height, and area. YOLOv8 output included confidence scores for each category and an identification frequency value, quantifying the consistency of YOLOv8 detections across the track. For example, if a track spanned 11 frames and YOLO identified the same bounding box as "Common Tern sit" in 9 frames and "Common Tern stand" in 2 frames, the track was assigned a score of 0.81 (9/11) for "Common Tern sit" and 0.19 (2/11) for "Common Tern stand", and with scores of 0 for all other categories.

2.5.1. Choosing classifier model and model evaluation

We evaluated three classification models: decision tree, random forest, and MLP (Multi-Layer Perceptron, a type of artificial neural network) (Taud and Mas, 2018), each tested with and without class weighting to address class imbalance, resulting in a total of 6 different models. Our dataset exhibited a highly imbalanced sample, with more Common Tern images than Little Tern images, reflecting the actual abundances in the colony. Class imbalance can bias models toward the majority class, reducing performance on minority classes. The class-weighting option assigns weights inversely proportional to class frequencies, emphasizing minority classes during training. Alternative methods to handle class imbalance, such as SMOTE (Synthetic Minority Over-sampling Technique) and undersampling, were considered but deemed less appropriate in this context. In particular, SMOTE generates synthetic samples of the minority class to artificially balance the dataset. However, since the Little Tern is genuinely less abundant in reality, we intentionally chose not to artificially inflate its representation in the dataset, as this would not reflect the true ecological distribution. Moreover, since SMOTE creates new examples by interpolating between existing ones, it can generate biologically implausible samples that do not represent realistic variation, potentially hindering rather than helping the learning process. Undersampling, which reduces the number of majority class (Common Tern) samples, was also avoided because it would require discarding a significant amount of valuable data from the majority class. Therefore, class weighting was preferred as it maintains the integrity of the real-world data distribution while effectively addressing class imbalance during model training. For each model, hyperparameters were optimized using grid search based on the F1-

score (Table 1S). Models were then evaluated using accuracy, recall, precision, and F1-score metrics. Accuracy measures the ratio of correct predictions to total predictions. Precision evaluates the proportion of true positives (TP) among all positive predictions (TP + FP), while recall (synonymous with True Positive Rate, TPR) assesses the proportion of true positives relative to actual positives (TP / [TP + FN]). The F1-score combines precision and recall into a single metric as their harmonic mean, providing a balanced evaluation particularly suited for imbalanced datasets. To evaluate the contribution of individual features to classification performance, we trained four variants of the same classifier. Beginning with the baseline model using YOLOv8 outputs (confidence and frequency), we incrementally added box dimensions (height, width, and area), location probability, and tracking rates (movement and detection). Each model was evaluated using precision, recall, F1-score, and accuracy metrics. These metrics together offer comprehensive insights into model performance.

2.6. *Breeding identification and counting*

To determine the number of breeders for both Common and Little Terns, from any given date, two additional post-classification procedures were implemented (Fig. 2b).

2.6.1. *Tracking breeders across scans along the day*

The classifier was applied at four distinct times each day (10:00, 10:17, 15:00, and 15:17), and individual terns were tracked across these scans using IoU. Only sitting terns with an IoU above 0.1 across scans were included in the count of breeding birds. The threshold was again determined empirically, as described in Section 2.4.3, but set lower (0.1 instead of 0.4) to account for the longer time interval between scans. At this temporal resolution, greater differences, such as changes in the birds' orientation, had to be accommodated. Terns with an IoU below this threshold were excluded, as their presence was deemed too transient to indicate active breeding which is expect to be consistent over time at each breeding site.

2.6.2. *Eliminating duplicate counts of breeders*

Overlapping regions between adjacent camera flags could lead to double-counting of the same individual. To address this, all detected terns' locations were projected onto a drone image of the island. In overlapping areas covered by two flags, only the flag closest to its respective camera was counted.

2.7. *Comparing manual and algorithmic counts*

We compared the results of the counts produced by the algorithm with the numbers reported by field observers of breeding Common and Little Terns in 2023 and 2024. Manual counting was done twice a week during the breeding season in these years using the same two remote-controlled cameras employed for algorithm development. The observer scanned the island once in every count, utilizing the camera adjacent to the relevant area, and completing a full scan of the island within 30 min to 2 h, depending on the number of terns. Technical issues with one or both cameras, along with communication problems, occasionally resulted in missing recordings due to hardware malfunctions. In 2024, one of the cameras was non-functional for part of the season. Consequently, comparable data between manual counts and camera observations were available for a total of 10 counting days: 8 in 2023 and 2 in 2024. We performed a Pearson correlation test between manual and algorithmic counts across all sampled days. In addition, we used a Bland-Altman analysis, a method to assess the agreement between two measurement techniques by plotting the difference between methods against their mean, allowing the identification of systematic biases and the range of acceptable differences (Martin Bland and Altman, 1986). Both analyses were performed using R software (version 4.0.3, R Core Team).

3. Results

3.1. *Annual manual counts*

The annual manual counts of the breeding population sizes for both

Table 1
The performance of the tested models across different categories: background, chick, Common Tern flying, Common Tern sitting, Common Tern standing, Little Tern sitting, Little Tern standing and other species. For each category, Precision, Recall, and F1-score are reported. In addition, overall model performance is summarized using micro, macro, and weighted averages for each metric. The support is for all models except YOLOv8 initial model that was trained with data support of: chick – 357, Common Tern Flying – 30, Common Tern Sitting – 2799, Common Tern Standing – 770, Little Tern Sitting – 464, Little Tern Standing – 80, Little Tern Flying – 6, Other – 16.

Model	Performance Metrics	Category								Weighted Avg	Macro Avg	Micro Avg
		Background	Chick	Common Tern			Little Tern		Other			
				Fly	Sit	Stand	Sit	Stand				
YOLOv8	Precision		0.45	0.37	0.94	0.73	0.56	0.18	0	0.6	0.29	0.58
	Recall		0.63	1	0.72	0.73	0.61	0.56	0	0.69	0.27	0.58
	F1-Score		0.52	0.55	0.81	0.73	0.27	0.27	0	0.73	0.27	0.58
Decision Tree	Precision	0.42	0.4	0	0.93	0.81	0.81	0	0	0.9	0.42	0.9
	Recall	0.34	0.33	0	1	0.71	0.76	0	0	0.9	0.39	0.9
	F1-Score	0.38	0.36	0	1	0.76	0.79	0	0	0.9	0.4	0.9
Decision Tree weights	Precision	0.24	0.33	1	0.93	0.72	0.82	0	0	0.88	0.5	0.88
	Recall	0.29	0.33	1	0.94	0.66	0.85	0	0	0.88	0.51	0.88
	F1-Score	0.26	0.33	1	0.94	0.69	0.83	0	0	0.88	0.5	0.88
Random Forest	Precision	0.6	0.33	1	0.94	0.84	0.88	0	0	0.91	0.53	0.92
	Recall	0.32	0.08	0.5	0.98	0.76	0.77	0	0	0.92	0.49	0.92
	F1-Score	0.41	0.13	0.67	0.96	0.8	0.82	0	0	0.91	0.5	0.92
Random Forest weights	**Precision**	**0.49**	**0.36**	**1**	**0.96**	**0.73**	**0.82**	**0**	**0**	**0.91**	**0.53**	**0.89**
	Recall	**0.5**	**0.33**	**1**	**0.94**	**0.83**	**0.9**	**0**	**0**	**0.91**	**0.59**	**0.89**
	F1-Score	**0.49**	**0.35**	**1**	**0.95**	**0.78**	**0.86**	**0**	**0**	**0.91**	**0.55**	**0.89**
Neural Network	Precision	0	0.11	0	0.84	0.82	0.56	0	0	0.8	0.29	0.83
	Recall	0	0.08	0	0.96	0.14	0.52	0	0	0.81	0.25	0.83
	F1-Score	0	0.1	0	0.9	0.25	0.54	0	0	0.77	0.25	0.83
Neural Network weights	Precision	0.06	0.04	0.02	0.96	0.48	0.42	0	0	0.85	0.28	0.66
	Recall	0.32	0.5	0.5	0.56	0.55	0.85	0	0	0.56	0.47	0.66
	F1-Score	0.1	0.07	0.07	0.71	0.51	0.56	0	0	0.66	0.29	0.66
	Support*	38	12	2	1431	221	84	3	2	1793	1793	1793

the Common and the Little Terns were significantly influenced by the counting method (Common Tern: $p < 0.001$, $F_2 = 25.18$; Little Tern: $p < 0.05$, $F_2 = 6.542$; Fig. 2S). the counts using the two remote-controlled cameras located on the breeding island yielded the highest estimates of breeders for the Common Tern, while counts conducted from a car resulted in the lowest numbers for both species.

3.2. Model classification performance

The classification performance of YOLOv8 alone, after training on tagged images, was significantly lower than most of the classifiers trained with the addition of improved features (Table 1, Fig. 3). YOLOv8 achieved overall categories weighted precision of 0.72 while the best-performing model, random forest with weights, achieved 0.91. While YOLOv8 alone classified Common Tern sitting with high accuracy (recall = 0.94, Table 1, Fig. 3), it identified Little Tern sitting correctly only about half of the time (56 %, Fig. 4, Table 1). Moreover, 36 % of Little Tern sitting instances were misclassified as Common Tern sitting (Fig. 3), highlighting a notable confusion between the two categories. Among the six classifiers examined after incorporating the algorithmic improvements (with optimal parameters for each model provided in Table 2S), the highest scores were observed for the random forest

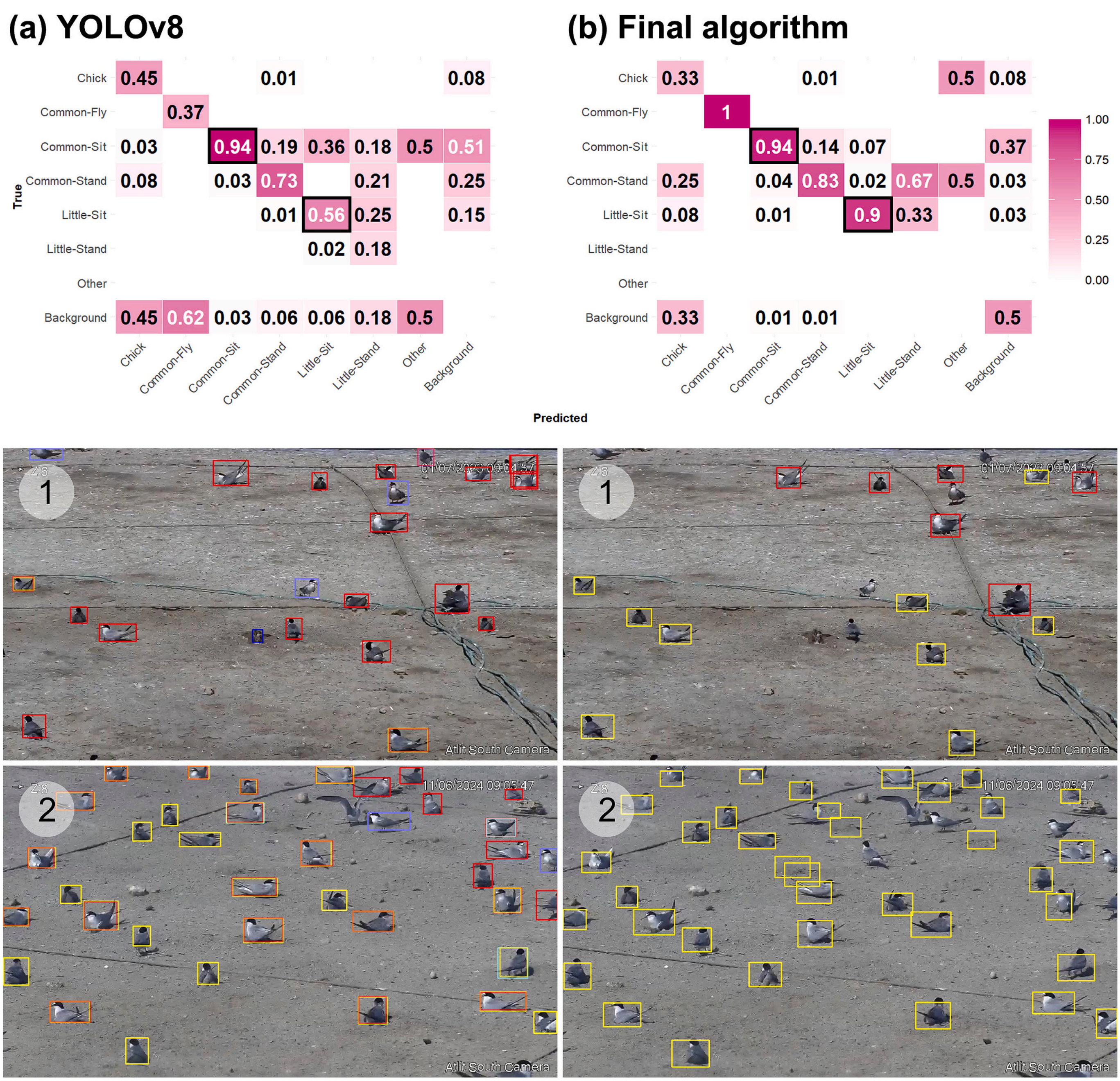

Fig. 3. Algorithm development and improvement. (a) Results of the algorithm after the initial YOLOv8 processing stage. (b) Results of the algorithm following algorithm improvements using the selected model (random forest with weights). Confusion matrices (upper panels) highlighting results for the "Common Tern sitting" and "Little Tern sitting" categories with black squares. Examples of YOLOv8 and final algorithm outputs from (1) July 1, 2023, at 09:04:57 and (2) June 11, 2024, at 09:05:47. "Common Tern sitting" detections are marked with red squares and "Little Tern sitting" detections with yellow squares. In instances of low detection confidence, YOLO may assign multiple category labels to a single bounding box. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

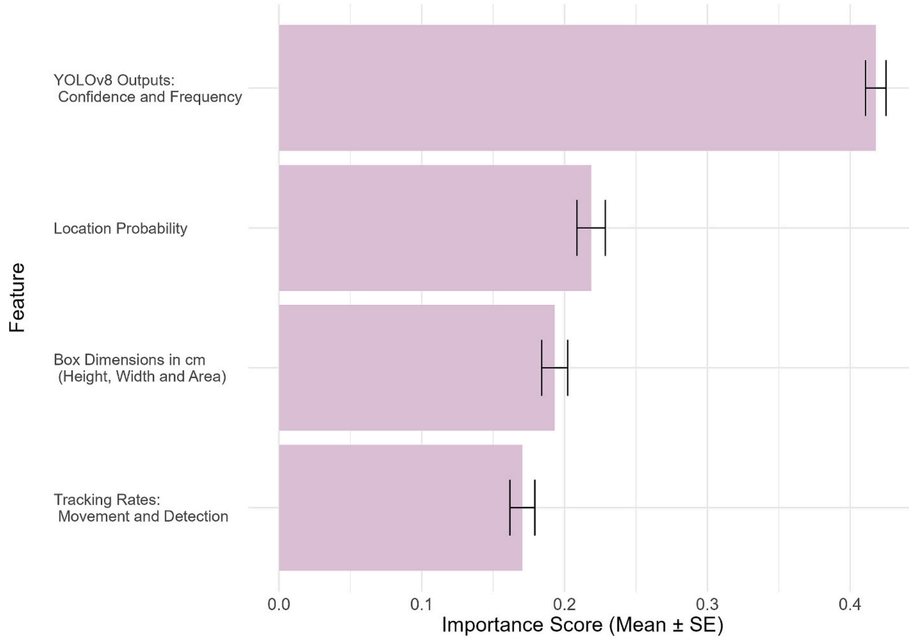

Fig. 4. The Relative importance of the features used to improve YOLOv8 model in the final algorithm (random forest with weights model). Values represent the mean ± standard error (SE) across five cross-validation folds. The relative importance of each feature is based on the number of times a variable is selected for splitting, weighted by the squared improvement of the model, averaged over all trees and scaled so that the sum adds to 1. The features were summarized to four main components across all the categories (detailed in Fig. 2S).

classifiers (both weighted and unweighted), which performed similarly. The weighted random forest classifier was selected as the final classifier due to its superior performance in identifying the Little Tern sitting category, which was one of the project's key focal categories, achieving a recall of 0.94. In this case, it misclassified Little Tern sitting as Common Tern sitting only in 7 % of the cases, while maintaining high

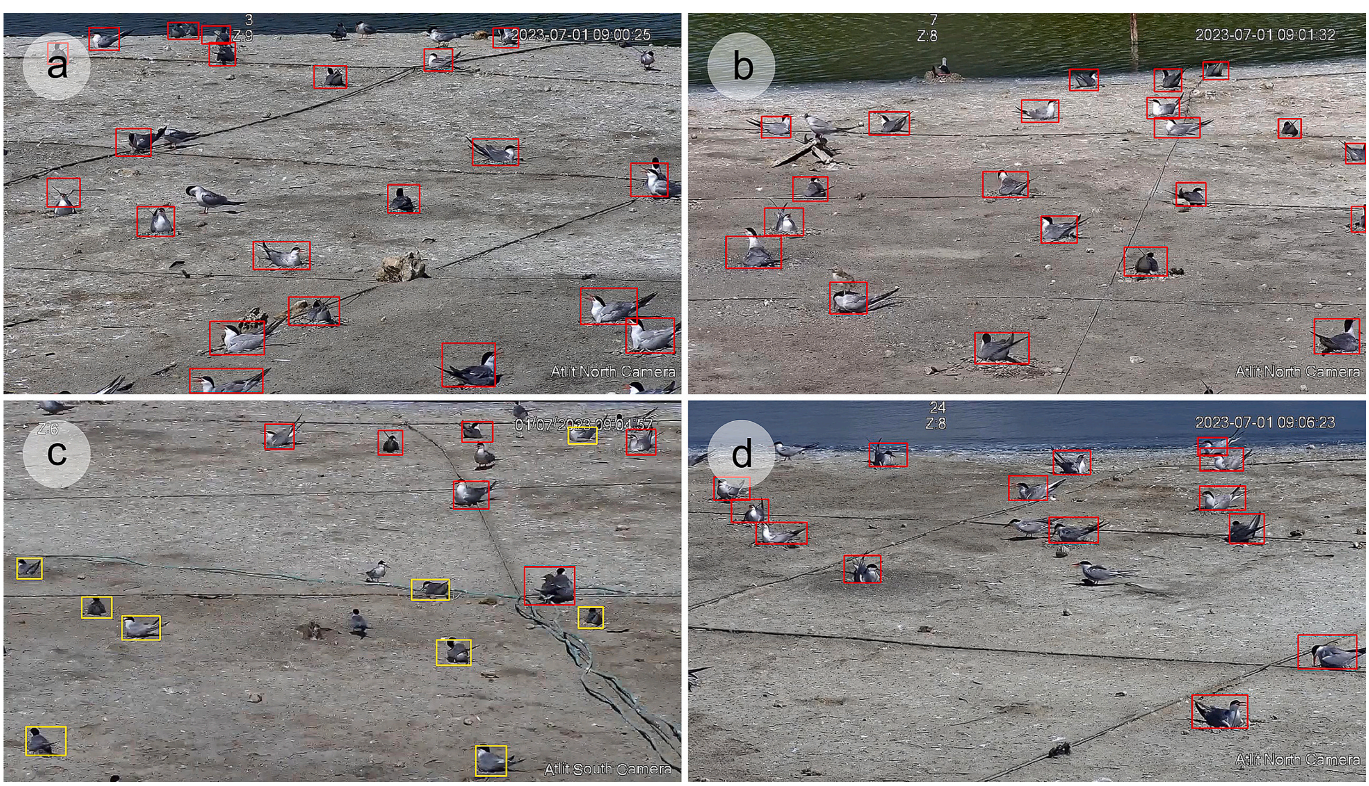

Fig. 5. Examples of the final algorithm performance using images from July 1, 2023 across different locations throughout the scanning of the island. Common Tern breeders in red squares and Little Tern breeders in yellow. (a) Identification of only Common Tern breeders, excluding standing individuals. (b) Identification of only Common Tern breeders, excluding other species breeders such as Black-winged Stilt (visible in the upper left corner). (c) Identification of both Little and Common Tern in close proximity, distinguishing between breeders and standing Little Terns. (d) Additional example of identification limited to only breeders terns despite close proximity to standing terns. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

accuracy for Common Tern sitting with a recall of 0.94 (Fig. 3, Table 1).

When assessing the contribution of individual features to the algorithm's performance improvement (Fig. 4, Fig. 3S, Table 3S), we found that all features substantially contributed. The location probability of Common and Little Terns at various island locations emerged as the most important feature (after YOLOv8 outputs), particularly for correctly classifying the Little Tern - Sit category (Table 3S), followed by box dimensions in cm and the tracking rates. However, the differences in the importance scores between these features were overall minor, indicating that all features contributed meaningfully to the algorithm's enhancement. Evaluating models with progressively added features (Table 3S) showed a consistent increase in weighted F1-score, rising from 0.85 for the baseline model to 0.87, 0.89, and 0.90 after the addition of box dimensions, location probability, and tracking rates, respectively. Specifically, box dimensions improved the classification of Common Tern - Sit (from F1: 0.91 to 0.93) and Little Tern - Sit (from F1: 0.50 to 0.58). The inclusion of location probability led to a substantial increase in the F1-score for Little Tern - Sit (from F1 0.58 to 0.83). Interestingly, the addition of tracking rates had the strongest impact on identifying the background category, nearly doubling its precision from

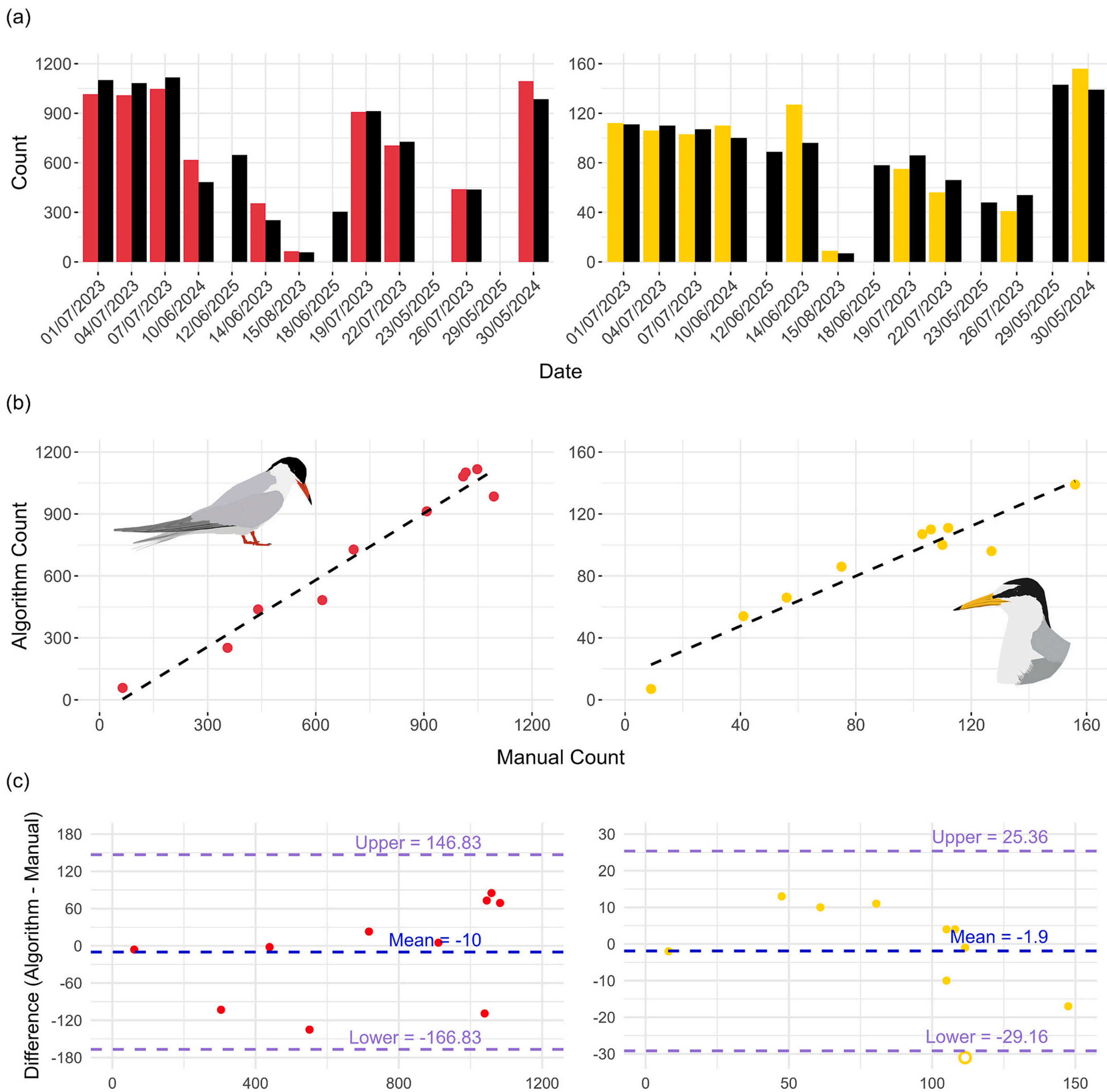

Fig. 6. (a) Manual breeding counts (red for Common Tern and yellow for Little Tern) and the final algorithm counts (in black) thought the 2023 and 2024 breeding seasons. (b) Correlation between manual and algorithm counts (red for Common Tern with $r = 0.98$ and $p < 0.001$, and yellow for Little Tern with $r = 0.96$ and $p < 0.001$). (c) Bland-Altman plot comparing manual counts with the final algorithm counts (red for Common Tern and yellow for Little Tern). The mean difference between the two methods is indicated by the blue dashed line, and the limits of agreement (range covering 95 % of the differences) are represented by purple dotted lines. Solid points represent observations within the limits of agreement, and hollow points indicate observations outside the limits. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

0.23 to 0.42.

Different categories exhibited distinct box sizes, with Little Tern having the smallest box area and width and the Common Tern flying category having the largest box area and height (Fig. 4S).

3.3. Final algorithm performance

The final breeding identification and counting algorithm successfully distinguished between standing and sitting terns for both Common (Fig. 5) and Little Terns (Fig. 5c). It also differentiated between Common and Little Terns regardless of their behavior (Fig. 5c) and between terns and other species breeding on the same island, such as Black-winged Stilts (Fig. 5b).

When comparing the breeding identification and counting algorithm with manual counts obtained via the two remote cameras in 2023 and 2024 (Fig. 6), the mean percentage difference was 2.01 % with a standard error of 3.42 %. By species, the mean percentage difference between the manual and algorithm counts was 4.47 ± 4.06 % for Common Terns and 0.43 ± 5.64 % for Little Terns. The Pearson correlation coefficient between the algorithm and manual counts is 0.98 for the Common Tern ($t_8 = 14.17$, $p < 0.001$) and 0.96 for the Little Tern ($t_8 = 9.1$, $p < 0.001$; Fig. 6). The Bland-Altman analysis showed wide agreement between the two methods for both species. For Common Terns, the mean difference was −10 with wider limits of agreement, reflecting greater variability at higher counts. For Little Terns, the mean difference was very small (−1.9) with narrow limits, indicating tight agreement. No proportional bias was observed for either species, with only one outlier detected for the Little Tern (Fig. 6c).In addition to reporting the number of breeding terns for both species, the final algorithm provided the location of the breeders on the island (Fig. 7).

4. Discussion

The ability to accurately monitor birds is critical for following and conserving their declining populations, which are increasingly threatened by human activities, climate changes, and habitat degradation (Brierley and Kingsford, 2009; Fay et al., 2015). In this study, we introduce a novel, efficient, and fully automated algorithm framework for the identification, counting, and spatial mapping of seabird breeding colonies. The algorithm represents a significant advancement in wildlife monitoring, offering a robust alternative to existing methods. While previous approaches have leveraged deep learning and camera systems to monitor birds and wild animal in general, they typically rely either on fixed camera traps (Chalmers et al., 2025; Fennell et al., 2022; Johanns et al., 2022; Noguchi et al., 2025) that provide continuous monitoring of a limited number of individuals or small areas, rendering them unsuitable for large water bird colonies, or on drones (Hayes et al., 2021; Rančić et al., 2023; Schad and Fischer, 2023),which can survey broad areas and capture behavior but are constrained by flight time and operational limitations. Our approach employs remote-controlled surveillance cameras, similar to those used in security systems, enabling continuous, long-term monitoring of large breeding colonies, specifically, over 1300 breeding pairs of Common and Little Terns, across extensive spatial and temporal scales. This persistent observation allows the detection of dynamic breeding behaviors, and by integrating ecological traits specific to water birds species into the algorithm, we present, for the first time, a method capable of monitoring an entire water bird colony throughout a complete breeding season. Despite the challenges of high colony density and species similarity, the framework successfully identified, counted, and mapped breeders with an average discrepancy of only 2 % compared to manual counts and achieved over 90 % accuracy in correctly identifying breeding individuals versus non-breeding individuals of both species.

Significant investment of time and resources was made to manually

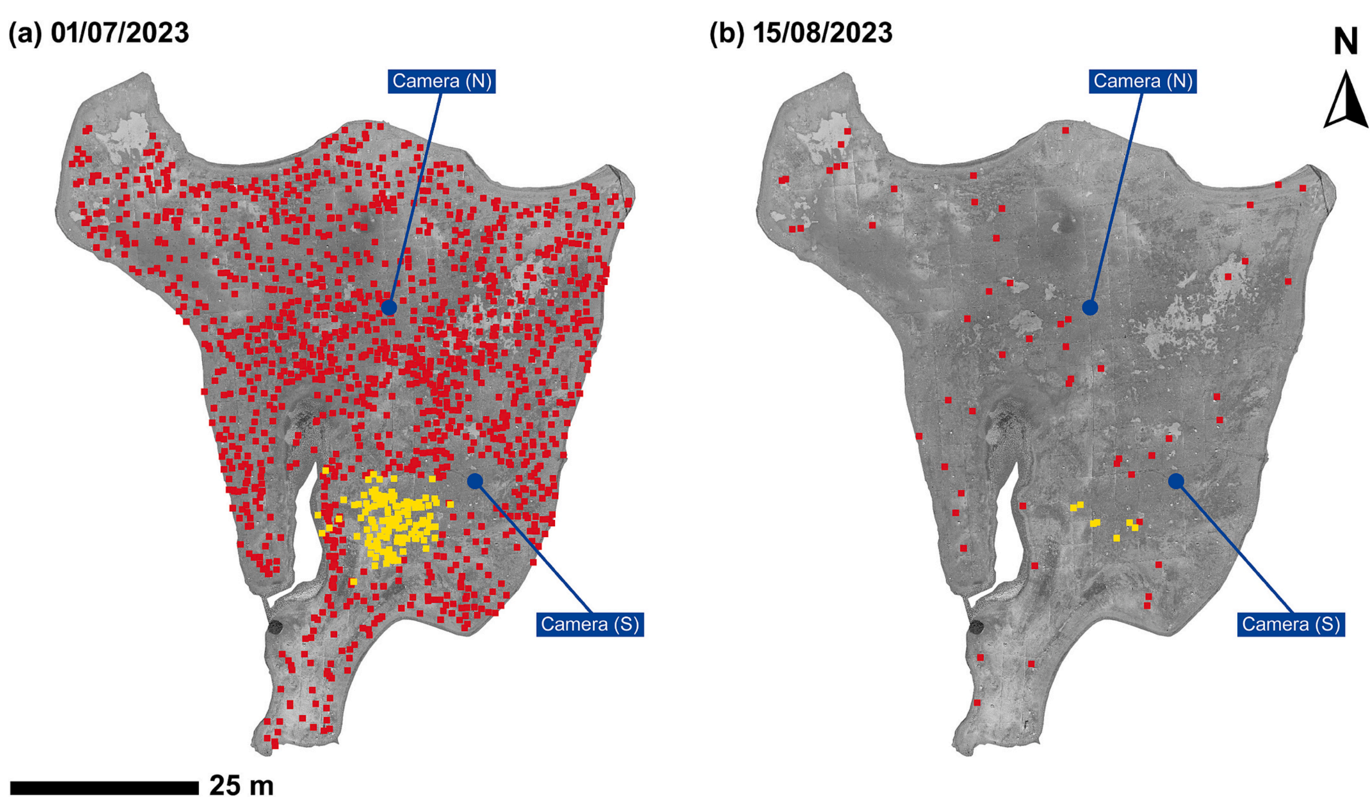

Fig. 7. Drone image of the terns' breeding island with the cameras locations (in blue circles), Common Tern breeders (red squares) and Little Tern breeders (yellow squares) in two dates along the breeding season: (a) 01/07/2023 and (b) 15/08/2023. The island area is 3200 m^2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

monitor these two species at the Atlit colony over a fourteen-year period, before the start of the present project. During this time, monitoring techniques evolved in an effort to improve accuracy. However, despite these efforts, the collected data cannot provide answers to one of the most fundamental questions of monitoring, whether the number of breeding individuals has changed over time. This issue is not unique to this project but reflects a broader challenge in conservation efforts. As technology improves and becomes more accessible, new, but usually still manual methods (Edney and Wood, 2021) are introduced into monitoring protocols, creating inconsistencies that hinder comparisons over time. In cases in which only traditional methods such as observation are used, variation between individual surveyors can lead to discrepancies of up to 20 %, particularly in large, densely populated colonies (Hutchinson, 1980). This situation highlights the critical need for a valid automated method to monitor bird colonies that remains reliable over time and does not interfere with breeding activities.

YOLOv8 is one of the latest advancements in deep learning classification, offering one of the highest performance available today (Varghese and Sambath, 2024). However, when applying it to distinguish between similar species or to identify specific behaviors, such as breeding, additional enhancements must be considered. On its own, after training on 1659 tagged images that yielded 5920 tagged classifications, YOLOv8 achievements were overall low, with a general weighted F1 score of 0.73. By integrating animal behavior and ecological knowledge into the model, we significantly improved its performance and achieved a general weighted F1 score of 0.91. In the Little Tern sitting category, for example, recall improved from 0.56 to 0.90.

This project was developed collaboratively by computer science experts and biologists with extensive experience in monitoring seabird colonies. It incorporated features that aid manual field monitoring in the algorithm architecture, a critical step toward leveraging AI in ecology. Embedding these features into the algorithm significantly improved its performance. Birds, and particularly seabirds, exhibit high fidelity to their breeding sites (Skutch, 1957). For example, the breeding location of Little Terns on the island has changed only once in the past fourteen years. Incorporating this site fidelity data into location probabilities further enhanced the model's accuracy. The drawback of this feature is that in the event of a sudden shift in breeding locations, the model would require updates and retraining. However, as such shifts are normally rare, we believe incorporating this feature is both appropriate and necessary for achieving good results, especially when monitoring similar species that breed in close proximity to one another. In addition, an essential behavior of breeding birds is that they spend most of their time incubating, remaining seated in the exact same place (Skutch, 1957). Traditional methods for monitoring breeding colonies, whether through direct observation, single-frame camera images, or drone snapshots (Davison et al., 2025; Neupane et al., 2022), provide only a single momentary view of the colony. This approach risks missing some breeders or incorrectly counting non-breeding birds present in the area (Hutchinson, 1980). In the present study, we utilized video footage for the first time to incorporate this behavior into the identification of breeders, and to our knowledge, this represents the first successful automatic differentiation between breeders and non-breeders within a colony.

Our method involved tracking individual birds for 11 s (excluding the start and end of each sequence) and integrating the frequency of YOLOv8 detections for each box as a feature in the final algorithm. Additional features, such as detection and movement scores, were also included. For example, a bird that stood in a place for a few seconds before flying away would receive a lower frequency score. Similarly, a bird that moved extensively but remained within the same small area and got high frequency score, would be assigned a low movement score. Beyond this, post-classification procedures evaluated the location of each box at different times of the day, increasing the certainty of whether a bird was a breeder. The use of video, rather than single images, for wildlife monitoring, holds immense potential for advancing research on wild animals (Kline et al., 2025; Simões et al., 2023). This approach can provide detailed and near real-time insights into behavioral traits such as nesting, food provisioning (Schindler and Steinhage, 2021), intra- and interspecific aggression, and the effects of weather on animals (Hentati-Sundberg et al., 2023; Kline et al., 2025; Schindler and Steinhage, 2021; Simões et al., 2023).

Another significant improvement we incorporated into our algorithm that, to the best of our knowledge, is being applied for the first time in automatic animal monitoring, is the camera calibration process. This process allows the algorithm to include bird size in centimeters as a feature, which is one of the most critical distinguishing physical characteristics between species in our case, as well as in other cases and species (Perktaş and Gosler, 2010). The calibration process enables us to extract valuable information from images that is often overlooked in current applications, specifically the size and precise location of the birds. Automatically monitoring bird locations can contribute significantly to understanding a wide range of biological traits, providing insights that can later be used as conservation tools. For example, tracking the spatial distribution of birds throughout the breeding season can reveal why certain areas are preferred as well as why this preference may change over time. Furthermore, the location map can be overlaid with additional environmental layers, such as sand moisture, soil temperature, elevation, distance to water or the presence of other species, providing deeper ecological context for understanding habitat selection and use. When combined with additional tools, such as leg rings or wing tags, this approach could facilitate a variety of research endeavors, including tracking individual traits over multiple timescales, studying the impacts of environmental changes on different individuals, and analyzing relationships within and between species. These advancements have the potential to deepen our understanding of animal behavior and ecology, ultimately supporting more effective conservation strategies.

We were unable to achieve high accuracy in identifying chicks, and we recommend collecting more targeted data to address this challenge. Targeted data collection of chicks must be conducted during a specific period of the breeding season, after the chicks have hatched. Moreover, to enable accurate identification, data should be collected across multiple developmental stages, as the appearance of a three-day-old chick differs considerably from that of a two-week-old. During the chick-rearing period, the optimal time for image collection is in the morning, when chicks are most active. As the day progresses, chicks tend to seek shelter from the sun, often hiding in vegetation or other shaded areas, making detection more challenging. An additional technique that can enhance chick detection, particularly given their small size and camouflaged appearance against the ground, is the use of thermal infrared cameras. These cameras can more effectively distinguish chicks from the background and even detect individuals concealed within vegetation (Ghosh et al., 2024). Identifying chicks in a large and densely populated colony presents significant difficulties, but it is both an important and urgent task. Avian species, especially water birds, including terns, serve as natural reservoirs for various diseases, such as avian influenza and avian malaria (Pantin-Jackwood and Swayne, 2009). Since 2022, multiple outbreaks of highly pathogenic avian influenza (HPAI) occurred worldwide, causing unprecedented mortality rates (Camphuysen and Gear, 2022). Over the past decade, two mass mortality events among tern chicks at this site, attributed to disease outbreaks, have led to an 80 % reduction in the chick population. These events underscore the urgent need for automated monitoring systems to track chicks and enable early detection of unusual mortality events.

Camera-based ecological monitoring and research are rapidly expanding in both scope and application (Cusick et al., 2024; Johanns et al., 2022; Lahoz-Monfort and Magrath, 2021; Rafiq et al., 2025; Reynolds et al., 2024; Song et al., 2024). Although the algorithm in this study was developed using data from a single tern colony in Israel, the methodology and technology employed are highly adaptable and scalable. The relevance of this technology extends far beyond terns and seabirds, offering potential applications across a wide range of avian species, other animal taxa, and ecological settings. For example, similar systems could be deployed at bat and bird communal roosting sites, other seabird colonies, or wetland breeding or wintering grounds. The approach is suitable for diverse landscapes as well, from open beach colonies, forested areas, urban roosts, or rugged cliffs, provided that camera placement offers sufficient visibility and stability. The increasing global availability of surveillance cameras, often already installed at key ecological sites (e.g., breeding colonies, migratory stopovers, wintering grounds), further facilitates the widespread adoption of this approach. In our study, we utilized standard surveillance equipment, a cost-effective and widely available technology. The total cost of the two cameras used was less than $3500, making this an affordable solution for long-term monitoring and research. In addition, model training was conducted using a GPU provided for free through Google Colab, eliminating the need for dedicated high-performance computing resources. By combining relatively low-cost equipment, minimal infrastructure requirements, and flexible model training pipelines, this approach is well-positioned for adaptation across taxa and ecosystems, thus enhancing the generalizability and utility of automated ecological monitoring. In the face of accelerating biodiversity loss, the development of accurate and reliable monitoring tools, like the algorithm proposed in this study, is critical for providing high-resolution data to support wildlife management, research, and conservation efforts. By combining computer vision with deep learning and ecological features, our method offers a scalable solution for identifying, counting, and mapping wildlife populations while minimizing human disturbance. This innovative approach has significant potential to improve the efficiency and accuracy of monitoring and research efforts, addressing global environmental challenges and informing effective conservation strategies.

Authors' contribution

Eyal Halabi, Inbal Schekler, Ilan Shimshoni and Nir Sapir conceived the study. Eyal Halabi and Inbal Schekler collected the data. Eyal Halabi, Inbal Schekler and Ilan Shimshoni designed the methodology. Eyal Halabi and Inbal Schekler wrote the algorithm code. Inbal Schekler and Eyal Halabi wrote the manuscript with input from all co-authors.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT 4 in order to improve the code, the text language and the readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

We thank the Data Science Research Centre at the University of Haifa for supporting this study. Additionally, we thank the Innovation and Sustainability Unit at the University of Haifa for the Impact project grant. We are grateful to the salt company team, a crucial partner in the tern study and conservation project in Atlit, the regional Israel Nature and Parks Authority team and the Society for the Protection of Nature in Israel. The camera system was sponsored equally by the salt company, INPA and SPNI. The cameras were deployed at the colony under permit 2495/2023/3.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2025.103380.

Data availability

The model source code is available at https://github.com/eyalh15/terns-project.

References

An, L., Ren, J., Yu, T., Hai, T., Jia, Y., Liu, Y., 2023. Three-dimensional surface motion capture of multiple freely moving pigs using MAMMAL. Nat. Commun. 14, 1–14. https://doi.org/10.1038/s41467-023-43483-w.

August, T., Harvey, M., Lightfoot, P., Kilbey, D., Papadopoulos, T., Jepson, P., 2015. Emerging technologies for biological recording. Biol. J. Linn. Soc. 115, 731–749. https://doi.org/10.1111/BIJ.12534.

Bhuiyan, M.R., Wree, P., 2023. Animal behavior for chicken identification and monitoring the health condition using computer vision: a systematic review. IEEE Access 11, 126601–126610. https://doi.org/10.1109/ACCESS.2023.3331092.

Billerman, S., Keeney, B.K., Schulenberg, T.S., 2025. Birds of the world. Cornell Laboratory of Ornithology, Ithaca, NY, USA.

Blackmer, A.L., Ackerman, J.T., Nevitt, G.A., 2004. Effects of investigator disturbance on hatching success and nest-site fidelity in a long-lived seabird, Leach's storm-petrel. Biol. Conserv. 116, 141–148. https://doi.org/10.1016/S0006-3207(03)00185-X.

Blums, P., Nichols, J.D., Hines, J.E., Mednis, A., 2002. Sources of variation in survival and breeding site fidelity in three species of European ducks. J. Anim. Ecol. 71, 438–450. https://doi.org/10.1046/J.1365-2656.2002.00613.X.

Bolton, M., Butcher, N., Sharpe, F., Stevens, D., Fisher, G., 2007. Remote monitoring of nests using digital camera technology. J. Field Ornithol. 78, 213–220. https://doi.org/10.1111/J.1557-9263.2007.00104.X.

Boulinier, T., Lemel, J.Y., 1996. Spatial and temporal variations of factors affecting breeding habitat quality in colonial birds : some consequences for dispersal and habitat selection. Acta Oecol. 17, 531–552.

Bridge, E.S., Thorup, K., Bowlin, M.S., Chilson, P.B., Diehl, R.H., Fléron, R.W., Hartl, P., Roland, K., Kelly, J.F., Robinson, W.D., Wikelski, M., 2011. Technology on the move: recent and forthcoming innovations for tracking migratory birds. Bioscience 61, 689–698. https://doi.org/10.1525/BIO.2011.61.9.7.

Brierley, A.S., Kingsford, M.J., 2009. Impacts of climate change on marine organisms and ecosystems. Curr. Biol. 19. https://doi.org/10.1016/j.cub.2009.05.046.

Camphuysen, C., Gear, S., 2022. Avian influenza leads to mass mortality of adult Great Skuas in Foula in summer 2022. Scott. Birds 312–323.

Chabot, D., Craik, S.R., Bird, D.M., 2015. Population census of a large common tern colony with a small unmanned aircraft. PLoS One 10, e0122588. https://doi.org/10.1371/JOURNAL.PONE.0122588.

Chalmers, C., Fergus, P., Wich, S., Longmore, S.N., Walsh, N.D., Oliver, L., Warrington, J., Quinlan, J., Appleby, K., 2025. AI-driven real-time monitoring of ground-nesting Birds: a case study on curlew detection using YOLOv10. Remote Sens 17, 769. https://doi.org/10.3390/RS17050769.

Croxall, J.P., Butchart, S.H.M., Lascelles, B., Stattersfield, A.J., Sullivan, B., Symes, A., Taylor, P., 2012. Seabird conservation status, threats and priority actions: a global assessment. Bird Conserv. Int. 22, 1–34. https://doi.org/10.1017/S0959270912000020.

Cusick, A., Fudala, K., Storożenko, P.P., Świeżewski, J., Kaleta, J., Oosthuizen, W.C., Pfeifer, C., Bialik, R.J., 2024. Using machine learning to count Antarctic shag (*Leucocarbo bransfieldensis*) nests on images captured by remotely piloted aircraft systems. Ecol. Inform. 82, 102707. https://doi.org/10.1016/J.ECOINF.2024.102707.

Davison, A.M., de Koning, K., Taubert, F., Schakel, J.K., 2025. Automated near real-time monitoring in ecology: status quo and ways forward. Ecol. Inform. 89, 103157. https://doi.org/10.1016/J.ECOINF.2025.103157.

Delisle, Z.J., Flaherty, E.A., Nobbe, M.R., Wzientek, C.M., Swihart, R.K., 2021. Next-generation camera trapping: systematic review of historic trends suggests keys to expanded research applications in ecology and conservation. Front. Ecol. Evol. 9, 617996. https://doi.org/10.3389/FEVO.2021.617996. /BIBTEX.

Doherty, P.F., Nichols, J.D., Tautin, J., Voelzer, J.F., Smith, G.W., Benning, D.S., Ray Bentley, V., Bidwell, J.K., Bollinger, K.S., Brazda, A.R., Buelna, E.K., Goldsberry, J. R., King, R.J., Roetker, F.H., Solberg, J.W., Thorpe, P.P., Wortham, J.S., 2002. Sources of variation in breeding-ground fidelity of mallards (*Anas platyrhynchos*). Behav. Ecol. 13, 543–550. https://doi.org/10.1093/BEHECO/13.4.543.

Edney, A.J., Wood, M.J., 2021. Applications of digital imaging and analysis in seabird monitoring and research. Ibis 163, 317–337. https://doi.org/10.1111/ibi.12871.

Fang, C., Zhang, T., Zheng, H., Huang, J., Cuan, K., 2021. Pose estimation and behavior classification of broiler chickens based on deep neural networks. Comput. Electron. Agric. 180, 105863. https://doi.org/10.1016/J.COMPAG.2020.105863.

Fay, R., Weimerskirch, H., Delord, K., Barbraud, C., 2015. Population density and climate shape early-life survival and recruitment in a long-lived pelagic seabird. J. Anim. Ecol. 84, 1423–1433. https://doi.org/10.1111/1365-2656.12390.

Fennell, M., Beirne, C., Burton, A.C., 2022. Use of object detection in camera trap image identification: assessing a method to rapidly and accurately classify human and

animal detections for research and application in recreation ecology. Glob. Ecol. Conserv. 35, e02104. https://doi.org/10.1016/J.GECCO.2022.E02104.
Forys, B.J., Xiao, D., Gupta, P., Murphy, T.H., 2020. Real-time selective markerless tracking of forepaws of head fixed mice using deep neural networks. eNeuro 7, 1–14. https://doi.org/10.1523/ENEURO.0096-20.2020.
Ghosh, S., Singh, A., Kumar, S., 2023a. Identification of medicinal plant using hybrid transfer learning technique. Indones. J. Electr. Eng. Comput. Sci. 31, 1605–1615. https://doi.org/10.11591/ijeecs.v31.i3.pp1605-1615.
Ghosh, S., Singh, A., Kumar, S., 2023b. BBBC-U-Net: optimizing U-Net for automated plant phenotyping using big bang big crunch global optimization algorithm. Int. J. Inf. Technol. 15, 4375–4387. https://doi.org/10.1007/S41870-023-01472-8. /TABLES/7.
Ghosh, S., Singh, A., Kumar, S., 2024. HPB3C-3PG algorithm: a new hybrid global optimization algorithm and its application to plant classification. Ecol. Inform. 81, 102581. https://doi.org/10.1016/J.ECOINF.2024.102581.
Hatzofe, O., Mayrose, A., 2009. Tern's Breeding in Isaral.
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. Ornithol. Appl. 123, 1–16. https://doi.org/10.1093/ORNITHAPP/DUAB022.
Hentati-Sundberg, J., Olin, A.B., Reddy, S., Berglund, P.A., Svensson, E., Reddy, M., Kasarareni, S., Carlsen, A.A., Hanes, M., Kad, S., Olsson, O., 2023. Seabird surveillance: combining CCTV and artificial intelligence for monitoring and research. Remote Sens. Ecol. Conserv. 9, 568–581. https://doi.org/10.1002/rse2.329.
Hodgson, J.C., Mott, R., Baylis, S.M., Pham, T.T., Wotherspoon, S., Kilpatrick, A.D., Raja Segaran, R., Reid, I., Terauds, A., Koh, L.P., 2018. Drones count wildlife more accurately and precisely than humans. Methods Ecol. Evol. 9, 1160–1167. https://doi.org/10.1111/2041-210X.12974.
Hutchinson, A.E., 1980. Estimating numbers of colonial nesting seabirds: a comparison of techniques. In: Proc. Colon. Waterbird Gr, 3, pp. 235–244.
IUCN, 2022. The IUCN red list of threatened species. Version 2022–2.
Jiang, L., Wu, L., 2024. Enhanced Yolov8 network with extended Kalman filter for wildlife detection and tracking in complex environments. Ecol. Inform. 84, 102856. https://doi.org/10.1016/J.ECOINF.2024.102856.
Jiang, D., Wang, H., Li, T., Gouda, M.A., Zhou, B., 2025. Real-time tracker of chicken for poultry based on attention mechanism-enhanced YOLO-Chicken algorithm. Comput. Electron. Agric. 237, 110640. https://doi.org/10.1016/J.COMPAG.2025.110640.
Johanns, P., Haucke, T., Steinhage, V., 2022. Automated distance estimation for wildlife camera trapping. Ecol. Inform. 70, 101734. https://doi.org/10.1016/J.ECOINF.2022.101734.
Jones, F.M., Arteta, C., Zisserman, A., Lempitsky, V., Lintott, C.J., Hart, T., 2020. Processing citizen science- and machine-annotated time-lapse imagery for biologically meaningful metrics. Sci Data 71 (7), 1–15. https://doi.org/10.1038/s41597-020-0442-6.
Kiat, Y., Schekler, I., 2021. Tern project 2021.
Kline, J., Zhong, A., Irizarry, K., Stewart, Charles V., Stewart, Christopher, Rubenstein, D. I., Berger-Wolf, T., 2025. WildWing: an open-source, autonomous and affordable UAS for animal behaviour video monitoring. Methods Ecol. Evol. 00, 1–10. https://doi.org/10.1111/2041-210X.70018. ;WGROUP:STRING:PUBLICATION.
Lahoz-Monfort, J.J., Magrath, M.J.L., 2021. A comprehensive overview of technologies for species and habitat monitoring and conservation. Bioscience 71, 1038–1062. https://doi.org/10.1093/BIOSCI/BIAB073.
Legg, C.J., Nagy, L., 2006. Why most conservation monitoring is, but need not be, a waste of time. J. Environ. Manag. 78, 194–199. https://doi.org/10.1016/J.JENVMAN.2005.04.016.
Leorna, S., Brinkman, T., Fullman, T., 2022. Estimating animal size or distance in camera trap images: photogrammetry using the pinhole camera model. Methods Ecol. Evol. 13, 1707–1718. https://doi.org/10.1111/2041-210X.13880.
Lewison, R., Oro, D., Godley, B.J., Underhill, L., Bearhop, S., Wilson, R.P., Ainley, D., Arcos, J.M., Boersma, P.D., Borboroglu, P.G., Boulinier, T., Frederiksen, M., Genovart, M., González-Solís, J., Green, J.A., Grémillet, D., Hamer, K.C., Hilton, G. M., Hyrenbach, K.D., Martínez-Abraín, A., Montevecchi, W.A., Phillips, R.A., Ryan, P.G., Sagar, P., Sydeman, W.J., Wanless, S., Watanuki, Y., Weimerskirch, H., Yorio, P., 2012. Research priorities for seabirds: improving conservation and management in the 21st century. Endanger. Species Res. 17, 93–121. https://doi.org/10.3354/ESR00419.
Martin Bland, J., Altman, D.G., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 327, 307–310. https://doi.org/10.1016/S0140-6736(86)90837-8.
Mathis, A., Mamidanna, P., Cury, K.M., Abe, T., Murthy, V.N., Mathis, M.W., Bethge, M., 2018. Deeplabcut: markerless pose estimation of user-defined body parts with deep learning. Nat. Neurosci. 21, 1281–1289. https://doi.org/10.1038/s41593-018-0209-y.
Mayrose, A., Lebinger, Z., Steinitz, O., Hatzofe, O., Haviv, E., Perelman, Y., Alon, D., Lider, N., 2017. The red book of birds in Israel. In: Society for the Protection of Nature in Israel, Israel Nature and Park Authority.
Neupane, S.B., Sato, K., Gautam, B.P., 2022. A literature review of computer vision techniques in wildlife monitoring. Int. J. Sci. Res. Publ. 16, 282–295.
Noguchi, N., Nishizawa, H., Shimizu, T., Okuyama, J., Kobayashi, S., Tokuda, K., Tanaka, H., Kondo, S., 2025. Efficient wildlife monitoring: deep learning-based detection and counting of green turtles in coastal areas. Ecol. Inform. 86, 103009. https://doi.org/10.1016/J.ECOINF.2025.103009.
Oro, D., Cam, E., Pradel, R., Martínez-Abraín, A., 2004. Influence of food availability on demography and local population dynamics in a long-lived seabird. Proc. R. Soc. Lond. B Biol. Sci. 271, 387–396. https://doi.org/10.1098/rspb.2003.2609.
Pan, W., Chen, J., Lv, B., Peng, L., 2025. Lightweight marine biodetection model based on improved YOLOv10. Alexandria Eng. J. 119, 379–390. https://doi.org/10.1016/J.AEJ.2025.01.077.
Pantin-Jackwood, M.J., Swayne, D.E., 2009. Pathogenesis and pathobiology of avian influenza virus infection in birds. Rev. Sci. Tech. 28, 113–136. https://doi.org/10.20506/rst.28.1.1869.
Patrick, S.C., Weimerskirch, H., 2017. Reproductive success is driven by local site fidelity despite stronger specialisation by individuals for large-scale habitat preference. J. Anim. Ecol. 86, 674–682. https://doi.org/10.1111/1365-2656.12636.
Perktaş, U., Gosler, A.G., 2010. Measurement error revisited: Its importance for the analysis of size and shape of birds, 45, 161–172. https://doi.org/10.3161/000164510X551309.
Phillips, R.A., Fort, J., Dias, M.P., 2023. Conservation status and overview of threats to seabirds. In: Conserv. Mar. Birds, pp. 33–56. https://doi.org/10.1016/B978-0-323-88539-3.00015-7.
Pugesek, B.H., Stehn, T.V., 2016. The Utility of Census or Survey for Monitoring Whooping Cranes in Winter.
Rafiq, K., Beery, S., Palmer, M.S., Harchaoui, Z., Abrahms, B., 2025. Generative AI as a tool to accelerate the field of ecology. Nat. Ecol. Evol. 2025, 1–8. https://doi.org/10.1038/s41559-024-02623-1.
Rančić, K., Blagojević, B., Bezdan, A., Ivošević, B., Tubić, B., Vranešević, M., Pejak, B., Crnojević, V., Marko, O., 2023. Animal detection and counting from UAV images using convolutional neural networks. Drones 2023 (7), 179. https://doi.org/10.3390/DRONES7030179.
Reynolds, S.A., Beery, S., Burgess, N., Burgman, M., Butchart, S.H.M., Cooke, S.J., Coomes, D., Danielsen, F., Di Minin, E., Durán, A.P., Gassert, F., Hinsley, A., Jaffer, S., Jones, J.P.G., Li, B.V., Mac Aodha, O., Madhavapeddy, A., O'Donnell, S.A. L., Oxbury, W.M., Peck, L., Pettorelli, N., Rodríguez, J.P., Shuckburgh, E., Strassburg, B., Yamashita, H., Miao, Z., Sutherland, W.J., 2024. The potential for AI to revolutionize conservation: a horizon scan. Trends Ecol. Evol. 40, 191–207. https://doi.org/10.1016/j.tree.2024.11.013.
Ribak, O., Kiat, Y., Schekler, I., Haztofe, O., 2017. Tern Project 2017.
Richter, S., Gerum, R.C., Schneider, W., Fabry, B., Le Bohec, C., Zitterbart, D.P., 2018. A remote-controlled observatory for behavioural and ecological research: a case study on emperor penguins. Methods Ecol. Evol. 9, 1168–1178. https://doi.org/10.1111/2041-210X.12971.
Saad Saoud, L., Sultan, A., Elmezain, M., Heshmat, M., Seneviratne, L., Hussain, I., 2024. Beyond observation: deep learning for animal behavior and ecological conservation. Ecol. Inform. 84, 102893. https://doi.org/10.1016/J.ECOINF.2024.102893.
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. ;PAGE:STRING:ARTICLE/CHAPTER.
Scheele, B.C., Legge, S., Armstrong, D.P., Copley, P., Robinson, N., Southwell, D., Westgate, M.J., Lindenmayer, D.B., 2018. How to improve threatened species management: an Australian perspective. J. Environ. Manag. 223, 668–675. https://doi.org/10.1016/J.JENVMAN.2018.06.084.
Schekler, I., Kiat, Y., Mayrose, A., Haztofe, O., 2015. Tern Project 2015.
Schekler, I., Kiat, Y., Haztofe, O., 2016. Tern Project 2016.
Schindler, F., Steinhage, V., 2021. Identification of animals and recognition of their actions in wildlife videos using deep learning techniques. Ecol. Inform. 61, 101215. https://doi.org/10.1016/J.ECOINF.2021.101215.
Schwaller, M.R., Olson, C.E., Ma, Z., Zhu, Z., Dahmer, P., 1989. A remote sensing analysis of Adélie penguin rookeries. Remote Sens. Environ. 28, 199–206. https://doi.org/10.1016/0034-4257(89)90113-2.
Segalin, C., Williams, J., Karigo, T., Hui, M., Zelikowsky, M., Sun, J.J., Perona, P., Anderson, D.J., Kennedy, A., 2021. The mouse action recognition system (MARS) software pipeline for automated analysis of social behaviors in mice. Elife 10. https://doi.org/10.7554/ELIFE.63720.
Simões, F., Bouveyron, C., Precioso, F., 2023. DeepWILD: wildlife identification, localisation and estimation on camera trap videos using deep learning. Ecol. Inform. 75, 102095. https://doi.org/10.1016/J.ECOINF.2023.102095.
Skutch, A.F., 1957. The incubation patterns of birds. Ibis 99 (1), 69–93. https://doi.org/10.1111/J.1474-919X.1957.TB01934.X.
Song, Q., Guan, Y., Guo, Xi, Guo, Xinhui, Chen, Y., Wang, H., Ge, J., Wang, T., Bao, L., 2024. Benchmarking wild bird detection in complex forest scenes. Ecol. Inform. 80, 102466. https://doi.org/10.1016/j.ecoinf.2024.102466.
Taud, H., Mas, J.F., 2018. Geomatic Approaches for Modeling Land Change Scenarios, Multilayer Perceptron (MLP). Springer, Cham. https://doi.org/10.1007/978-3-319-60801-3_27.
Varghese, R., Sambath, M., 2024. YOLOv8: A Novel object detection algorithm with enhanced performance and robustness. In: 2024 Int. Conf. Adv. Data Eng. Intell. Comput. Syst. ADICS, 2024. https://doi.org/10.1109/ADICS58448.2024.10533619.
Verdon, S.J., Davis, R.A., Tulloch, A., Legge, S.M., Watson, D.M., Woinarski, J.C.Z., Baker, G.B., Driessen, J., Geyle, H.M., Possingham, H., Garnett, S.T., 2024. Trends in monitoring of Australia's threatened birds (1990–2020): much improved but still inadequate. Emu 124, 21–36. https://doi.org/10.1080/01584197.2023.2275121.
Walsh, P., Halley, D., Harris, M., del Nevo, A., Sim, L., Tasker, M., 1995. Seabird Monitoring Handbook.
Weissensteiner, M.H., Poelstra, J.W., Wolf, J.B.W., 2015. Low-budget ready-to-fly unmanned aerial vehicles: an effective tool for evaluating the nesting status of canopy-breeding bird species. J. Avian Biol. 46, 425–430. https://doi.org/10.1111/JAV.00619.
Williams, H.M., DeLeon, R.L., 2020. Deep learning analysis of nest camera video recordings reveals temperature-sensitive incubation behavior in the purple martin (*Progne subis*). Behav. Ecol. Sociobiol. 74, 1–12. https://doi.org/10.1007/S00265-019-2789-2. /FIGURES/4.

Xiao, S., Wang, Y., Perkes, A., Pfrommer, B., Schmidt, M., Daniilidis, K., Badger, M., 2023. Multi-view tracking, re-ID, and social network analysis of a flock of visually similar birds in an outdoor aviary. Int. J. Comput. Vis. 131, 1532–1549. https://doi.org/10.1007/S11263-023-01768-Z. /FIGURES/1.
Yang, X., Dai, H., Wu, Z., Bahadur Bist, R., Subedi, S., Sun, J., Lu, G., Li, C., Liu, T., Chai, L., 2024. An innovative segment anything model for precision poultry monitoring. Comput. Electron. Agric. 222, 109045. https://doi.org/10.1016/J.COMPAG.2024.109045.
Zhang, W., Xu, Y., Bai, R., Li, L., 2023. Animal pose estimation algorithm based on the lightweight stacked hourglass network. IEEE Access 11, 5314–5327. https://doi.org/10.1109/ACCESS.2022.3231750.