RESEARCH ARTICLE

Seabird surveillance: combining CCTV and artificial intelligence for monitoring and research

Jonas Hentati-Sundberg[1] (iD), Agnes B. Olin[1,2] (iD), Sheetal Reddy[3,4], Per-Arvid Berglund[5] (iD), Erik Svensson[6] (iD), Mareddy Reddy[7] (iD), Siddharta Kasarareni[8], Astrid A. Carlsen[1] (iD), Matilda Hanes[9], Shreyash Kad[9] & Olof Olsson[10] (iD)

[1]Department of Aquatic Resources, Swedish University of Agricultural Sciences, Box 7018, 75007, Uppsala, Sweden
[2]Department of Ecology, Environment and Plant Sciences (DEEP), Stockholm University, 106 91, Stockholm, Sweden
[3]AI Sweden, Lindholmspiren 11, 41756, Gothenburg, Sweden
[4]Volvo Cars Corporation, Torslanda, 41878, Gothenburg, Sweden
[5]Baltic Seabird Project, Sanda Västerby 568, 62379, Klintehamn, Sweden
[6]Beräkning och Framsteg AB, Göteborg, Sweden
[7]Energy and Materials, Chalmers University of Technology, Kemivägen 10, 41296, Gothenburg, Sweden
[8]Chalmers Industriteknik, Gothenburg, Sweden
[9]Department of Electric Engineering, Chalmers University of Technology, Gothenburg, Sweden
[10]Stockholm Resilience Centre, Stockholm University, Stockholm, Sweden

Abstract

Ecological research and monitoring need to be able to rapidly convey information that can form the basis of scientifically sound management. Automated sensor systems, especially if combined with artificial intelligence, can contribute to such rapid high-resolution data retrieval. Here, we explore the prospects of automated methods to generate insights for seabirds, which are often monitored for their high conservation value and for being sentinels for marine ecosystem changes. We have developed a system of video surveillance combined with automated image processing, which we apply to common murres *Uria aalge*. The system uses a deep learning algorithm for object detection (YOLOv5) that has been trained on annotated images of adult birds, chicks and eggs, and outputs time, location, size and confidence level of all detections, frame-by-frame, in the supplied video material. A total of 144 million bird detections were generated from a breeding cliff over three complete breeding seasons (2019–2021). We demonstrate how object detection can be used to accurately monitor breeding phenology and chick growth. Our automated monitoring approach can also identify and quantify rare events that are easily missed in traditional monitoring, such as disturbances from predators. Further, combining automated video analysis with continuous measurements from a temperature logger allows us to study impacts of heat waves on nest attendance in high detail. Our automated system thus produces comparable, and in several cases significantly more detailed, data than those generated from observational field studies. By running in real time on the camera streams, it has the potential to supply researchers and managers with high-resolution up-to-date information on seabird population status. We describe how the system can be modified to fit various types of ecological research and monitoring goals and thereby provide up-to-date support for conservation and ecosystem management.

Introduction

The current rate of biodiversity loss requires urgent actions to avoid devastating consequences for the world's ecosystems (IPBES, 2019). Such actions must be supported by robust knowledge of how ecosystems work and respond to change, which in turn needs to be underpinned by high-quality data of sufficient coverage

(Mouquet et al., 2015; Sun & Scanlon, 2019). Furthermore, given the often abrupt and nonlinear nature of ecosystem changes, it is important that signals can be detected early on – something that requires high data resolution and signal-to-noise ratios (Clements et al., 2015; Hillebrand et al., 2020; Scheffer et al., 2009; Schmeller et al., 2018). Finally, in order to be useful in a practical setting, these data need to be up-to-date and cheap to obtain (Hampton et al., 2013; Purves, 2013).

The need for large amounts of high-resolution data is increasingly being met by automated systems, moving ecology into a 'big data' field with datasets of a size and resolution previously unimaginable (Hampton et al., 2013; Keitt & Abelson, 2021; Kelling et al., 2015; Michener & Jones, 2012). A few examples include online databases with billions of biotelemetry records (Davidson et al., 2021), several online data products from global-scale remote sensing (Wulder et al., 2012), regional networks of hundreds of audio recording units for species occurrence monitoring (Roe et al., 2021) and global networks of aquatic data collection buoys (Marcé et al., 2016; Roemmich et al., 2019).

Maximizing the utility of the vast amounts of data generated by automated sensor systems necessitates minimal latency between data acquisition, analysis and inference, in other words, high ecological velocity (Farley et al., 2018). One way to increase ecological velocity is to use artificial intelligence (AI), with machine learning algorithms processing continuous streams of data to provide inference and insights (Christin et al., 2019; Goodwin et al., 2022; Makiola et al., 2020). Early applications in this realm include real-time event identification in bio-loggers for tracking animal movement (Korpela et al., 2020), automatic identification and counting of animals from camera trap images (Norouzzadeh et al., 2021) and automated analysis of animal behaviour from video data (Williams & DeLeon, 2020).

Here, we created an automated data collection system based on video surveillance and deep learning (object detection) and applied it to colonially breeding seabirds (common murres *Uria aalge*; hereafter: 'murres') in the Baltic Sea. In recent years, seabird monitoring and research have become increasingly reliant on digital imaging techniques, such as time-lapse cameras and drones (Edney & Wood, 2021; Hinke et al., 2018; Jones et al., 2018; Rush et al., 2018; Schwaller et al., 2013). However, due to the large amounts of data produced, time spent manually processing images often constitutes a bottleneck (Edney & Wood, 2021). While there have been some recent applications of machine learning approaches to tackle this problem, this has so far primarily focused on counting birds from still frames (e.g. Hayes et al., 2021; Weinstein et al., 2022). In general, the

application of machine-learning algorithms to extract more sophisticated information on behaviour, demography and phenology from video data collected in the field has been very limited (but see Ditria et al., 2021; Schofield et al., 2019; Williams & DeLeon, 2020). However, the potential contribution to a better understanding of ecosystem dynamics is vast, especially if combined with high-resolution data on environmental conditions. As such, we envision our work as part of a move towards data-driven environmental management and research, making full use of the recent large leaps in sensor technology and automated analysis.

Materials and Methods

Technical setup for video recording

We used an off-the-shelf digital video surveillance (CCTV) system on the island of Stora Karlsö (57°17′ N, 17°58′ E), Baltic Sea, Sweden, to film murres at their nesting sites (Figs. 1 and 2). The installation was made in a previously constructed artificial breeding cliff, the Karlsö Auk Lab (Hentati-Sundberg et al., 2012). The CCTV system, which is easily scalable, consisted in 2019–2021 of five 2-megapixel resolution IP-cameras (Avtech AVM543P) The five cameras were mounted at five different ledges and covered 23 pairs on average over the three seasons. Details on the camera installation setup are provided in Supplementary Material S1. The cameras were powered by a 24-channel PoE network switch (AETEK C11-242-30-380) and connected to a Network Video Recorder (NVR, Avtech AVH8516) *via* a 4G-router (Teltonika RUT950) (Fig. 1). The router sets local IP addresses to all devices on the local network (LAN) and allows, via a static external IP address and port forwards, remote real-time access to the NVR and all cameras through a web interface (Fig. 1). The power consumption of the system is approximately 4.5 W for each camera, 65 W for the NVR and 2–7 W for the router. Built-in IR light (859 nm) in the cameras gives clear visibility of the birds even at night, without affecting them (Fig. 2B). Recording is continuous at 25 frames-per-second (FPS), generating approximately 24 Gb data camera^{-1} day^{-1} with H.264 video compression. Backups, generating *.avi* files, are made daily, although the storage capacity of the internal hard drives in the NVR (80 Tb) fits a whole three-month breeding season with 15+ cameras and thus easily fit our setup with five cameras.

Development of approach

The video material generated by the system is considerable – one breeding ledge with seven breeding pairs of

murres during three breeding seasons (May–July 2019–2021), generated 5600 h of footage. Manual analysis of the material to record the presence of adults, eggs and chicks is thus a daunting undertaking. To combat this problem, the first author of this paper arranged a 'hackathon' in collaboration with AI Sweden, the Swedish national centre for applied AI, in November 2019, where several teams competed to come up with an automated procedure for the task. As supervised machine learning algorithms need annotated data for training, we annotated 2668 randomly selected images with bounding boxes for adults, chicks and eggs prior to the hackathon (see Fig. 2 for examples of adults and chicks), using an online annotation platform provided by Annotell (https://www.annotell.com). Annotation took on average 1 min per image after some training, meaning that approximately 44 h were spent on annotation. All annotated images are freely available for training new generations of models (see "Data Availability Statement" section).

The teams participating in the hackathon got 2 weeks to develop a solution for automated seabird monitoring. The teams were provided with the full video material, the annotated images and access to a powerful computing infrastructure for model training and inference (Nvidia DGX, a datacentre class computer which has eight A100 graphical processing units (GPUs) suitable for AI computations). The winning team in the hackathon (authors S. R., M. R. and S. K. in this paper) developed an analysis pipeline written in Python 3, using the deep learning-based YOLO object detection algorithm, which detects pre-specified classes of objects and delineates their extent with bounding boxes (Redmon et al., 2016) (Figs. 1 and 2). The same pipeline is described and employed in this paper, with some minor modifications to the original setup. The final data product is bird detections with class (in this case: adult birds, chicks, eggs), timestamp, position, bounding box size and confidence level of all objects identified in each frame of the video.

Object detection using YOLO

The work presented in this paper builds on YOLO version 5 (YOLOv5) (https://github.com/ultralytics/yolov5). Since the hackathon, additional images have been annotated, and the current number of annotated images is 4061, containing 22 658 adult birds, 1300 chicks and 155
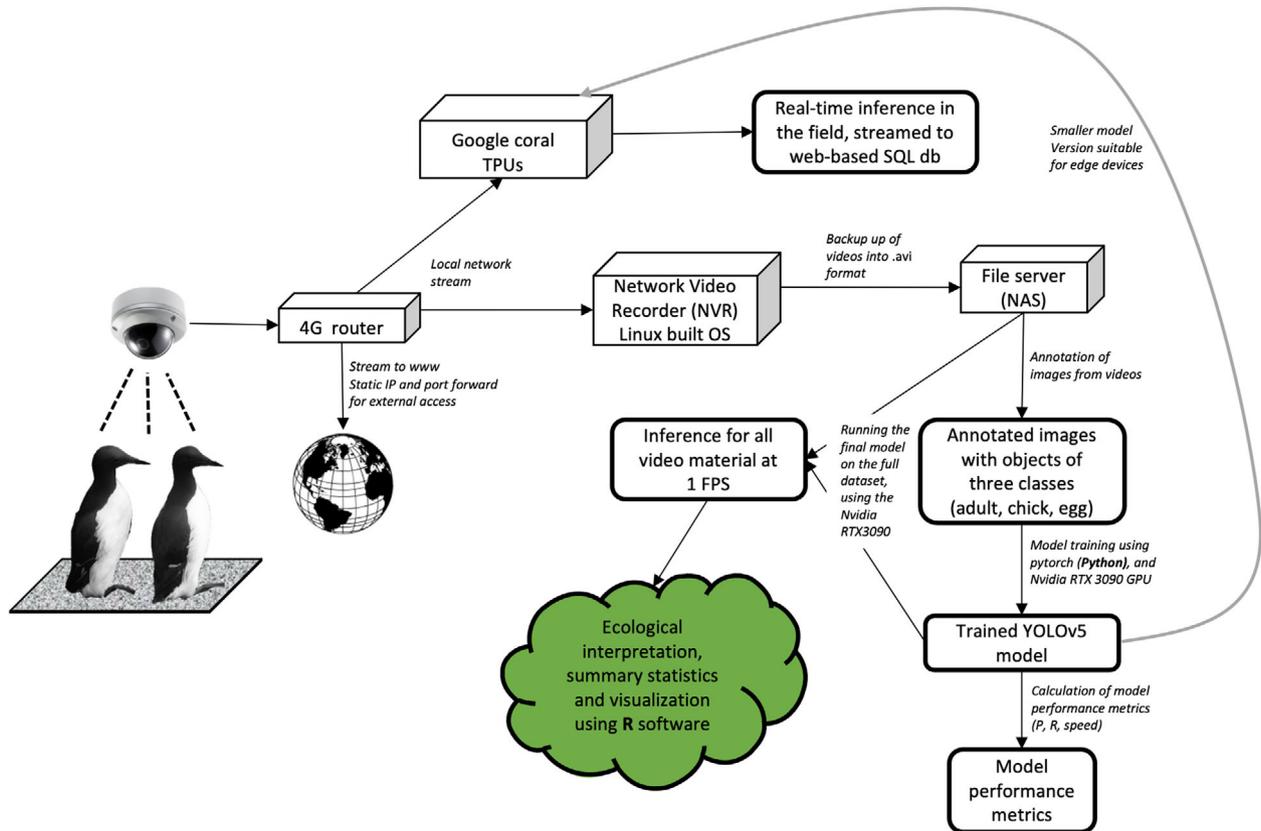

Figure 1. Illustration of the data collection and analysis pipeline used in this paper.
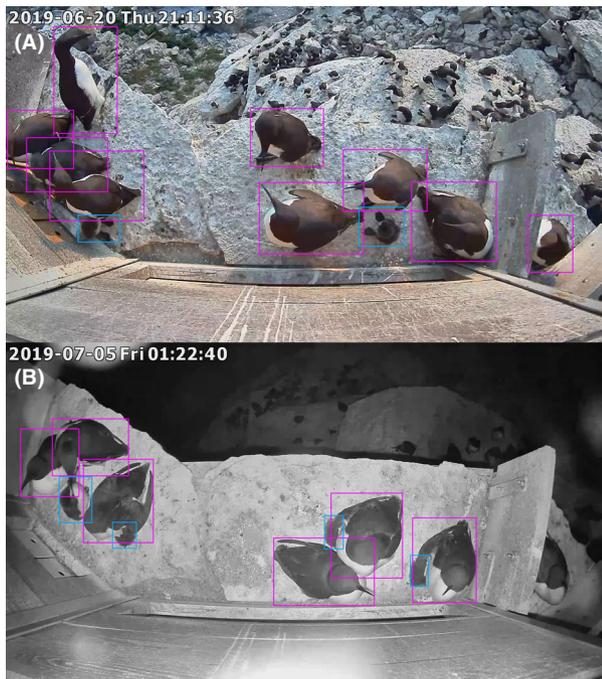
Figure 2. Example (A) day and (B) night-time images showing detected adult birds (magenta squares) and chicks (blue squares). Note that chicks partly covered by guarding parents are still detected.

eggs. Another 800 random background images (without birds, and thus no annotations) were added to the dataset for increasing robustness of the model based on unpublished YOLO recommendations. For training and testing of model performance, the annotated images were split into three sets: a train set, a validation set and a test set, each containing 80, 15 and 5% of the images, respectively (Table S1). The model was trained using the functions provided in the ultralytics YOLOv5 GitHub repository, using a Nvidia GEOFORCE RTX 3090 GPU.

In object detection, there is a trade-off between speed and accuracy (Redmon et al., 2016). Models with fewer parameters and running on smaller images are faster, whereas models with more parameters running on larger images are more accurate. YOLOv5 has pre-specified model sizes (nano, small, medium, large, x-large), and we evaluated models based on their Precision [P] (true positives/[true positives + false positives]) and Recall [R] (true positives/[true positives + false negatives]) and F1 score (the harmonic mean of P and R). In each model run, two parameters are set: Intersection-over-Union (IoU, overlap between the annotated and the predicted bounding box) and detection confidence (threshold level for a detected object to be retained). We ran a systematic parameter study across a range of IoU and threshold values for four models: Yolov5-small-640, Yolov5-small-960, Yolov5-medium-640 and Yolov5-medium-960 where

the number is the width of the images in pixels. Larger models were disregarded as they would lead to an unreasonable computation time given our current infrastructure, whereas smaller models (Yolov5-nano, and image sizes below 640 pixels width) were considered too inaccurate given initial experiments. Plots of F1 across models and model parameters are provided in Supplementary Material S2 (Figure S3). Based on these analyses, we used Yolov5-medium-960 for inference with an IoU threshold of 0.5 and a confidence threshold of 0.7 for the analyses reported in the paper. For all models, the training was stopped after 20 epochs of no improvements, using a stopping criterion provided by Yolov5. The Yolov5-medium-960 was trained for 130 epochs, with similar values for the other models.

When we ran inference with the final model on the complete dataset, we first used *ffmpeg* to resize the footage to videos at one image per second (every 25th frame) and resolution 960 × 544. This reduced the dataset size from 4 to 0.57 Tb. The model used for the inference was optimized and exported to TensorRT format. All image preparation, model training and model inferences were coded in Python 3 (van Rossum & Drake, 2009). Code for image processing and training and running the YOLOv5 model is available at https://github.com/BalticSeabird/SeabirdDetection. Examples of raw video data, all annotations and object detection outputs (SQLite database) are available at Data Dryad Repository (Hentati-Sundberg & Olin, 2022).

From detected objects to ecological inference

Using the object detection data, we showcase how AI can automate the study of several behavioural and demographic parameters in seabirds. We here present data from one camera, filming one of the artificial ledges ('Farallon 3') in the Karlsö Auk Lab with seven breeding pairs each year in 2019–2021. Several analytical processing steps from the raw detections to summary statistics were applied and are summarized under each section in the "Results". Post-processing and visualization of the object detection data were done in R 4.1.0 (R Core Team, 2021). Complete R code from detections in the SQLite database to ecological interpretation is available at https://github.com/BalticSeabird/ObjectDetectionInferences.

Real-time object detection in the field

In 2022, we implemented a pipeline for real-time object detection in the field. Rather than analysing pre-recorded video, we used the data streamed on the local network from the cameras (Fig. 1). For real-time applications,

inference speed becomes more important than for pre-recorded material, as a pre-requisite is that the analysis can run in (at a minimum) real time. Similarly, field sites often have limitations in electrical power which necessitates small processing units with limited power use. For the real-time experiments, we used some of the pre-trained models described above, but also smaller models (Yolov5-nano and smaller image sizes). We ran the models on the following three hardwares: a standard laptop CPU (Intel Core i5-8265U) with 8 Gb RAM, a Google Coral Dev Board2 with 4 GB RAM and a Nvidia GEFORCE RTX 2080 graphics card. We report values on price and power consumption of the three hardwares (Supplementary Material S2, Table S3) and speed obtained when running different model sizes and image sizes ("Results"). The inference results obtained from the devices were stored on a local SQLite database and then regularly pushed to a cloud-based SQL database, gaining global real-time access to data on presence of birds, chicks and eggs in the colony.

Results

Here, we first evaluate the performance of YOLOv5 in terms of accuracy and speed. Second, we provide some examples of how the object detection data can be used for ecological inference. Third, we explore the potential for real-time monitoring.

YOLO performance

The object detection algorithm run over three consecutive breeding seasons at one breeding ledge generated 139 342 009 detections of adults, 4 234 960 detections of chicks and 139 167 detections of eggs. The number of adults detected in each frame varied between 0 and 18.

The Precision and Recall of the YOLOv5-medium-960 model were 0.91 and 0.79, with an F1 score of 0.85, over all classes. The model performed better for adults than for chicks and eggs, with a P of 0.98, 0.84 and 0.92 and an R of 0.98, 0.74 and 0.64 for the three classes, respectively. High P means that there are few false positives, whereas the lower R means there are some false negatives (missing detections); in other words, at our chosen confidence threshold 0.7, the model is conservative. Eggs, which are almost always completely hidden under the incubating parent and not visible from above, were detected very rarely ($\leq$0.1% of detections) and therefore not analysed further. Full model performance scores are reported in Supplementary Material S2 (Table S2).

Video combined with AI dramatically reduced the time spent on analysis: we achieved an inference speed of 75 FPS for the Yolov5-medium-960 model. Images loaded on the GPU takes about 4.3 ms to process, which would ideally give a throughput of 230 FPS. However, loading and preprocessing the footage is a bottleneck in the setup. This means that we could process the whole 5600 h dataset in 75 h, running in the background with no human attention required. As manually annotating with bounding boxes takes about 1 min per image, achieving the same dataset would have taken 336 000 h (38 years) for a human. More realistically, analysing a 1 h video for recording number of adults and chicks present takes c. 10 min for an experienced analyst, which is 10 times slower than the computer and generates a dataset with much lower resolution.

We did not specifically monitor power consumption of the training and inference process, but the Nvidia RTX 3090 GPU has a power consumption of up to 350 W, meaning that the power use of the inference calculations may have been up to 9 kWh with the GPU running on 30%.

Ecological inference from object detection

Nest attendance, breeding activity and phenology

The object detection data can be studied at several different time scales to generate ecologically relevant information. If aggregated at short time scales (e.g. median number of adults detected per minute), it is possible to examine fine-scale attendance behaviour and, for example, identify disturbances (see Fig. 3A and further results below). At slightly longer time scales (e.g. median number of adults detected per hour), it is possible to examine diel cycles and breeding activity. Based on the object detections, there is clear variation in attendance patterns between days, as well as clear diel patterns (Fig. 3B).

It is also possible to track breeding activity and phenology across the full breeding season, providing key metrics for population monitoring. In our study system, this has traditionally been done through daily visits to determine the breeding status of each individual. Field data collected in this way matched inferences from the object detection data well, with birds detected in the middle of the night (median number of adults detected 00:00–01:00) providing a good predictor of active breeding (egg or chick) (Fig. 3C); the predicted and observed numbers were identical in 60% of 192 days in 2020 and 2021, and only off by 1 bird in over 20% in the remaining cases. Both breeding onset (first eggs laid) and chick departure (fledging) matched the onset and disappearance of night-time detections (Fig. 3C).

Chick growth

Growth is a key biological process that can provide information on, for example, food availability and
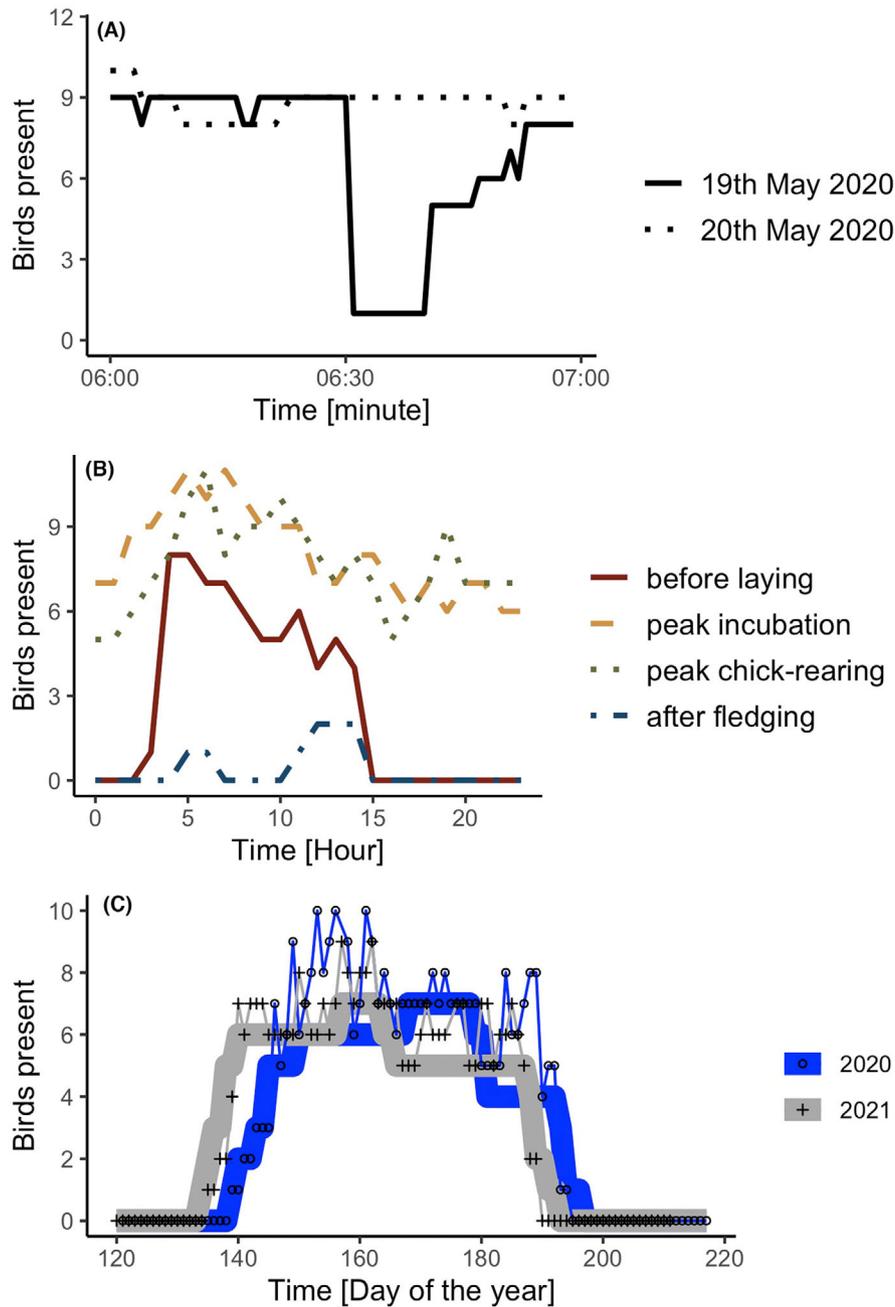
Figure 3. Attendance patterns for different time scales based on object detection. (A) Median number of adults per minute over the same hour (06:00–07:00) during two consecutive days in 2020, one with an eagle disturbance (solid) and one without (dotted) (see also Fig. 5). (B) Median number of adults per hour over 24 h for four different days in 2021 (before laying = 5th of May, peak incubation = 3rd of June, peak chick-rearing = 28th of June, after fledging = 29th of July). (C) Median number of adults detected during night (00:00–01:00) over the whole breeding season for 2020 and 2021 (thin lines and symbols), along with the actual number of active breeding pairs with egg or chick (thick lines) as based on daily, manual nest checks. Note that only one parent is usually present in the middle of the night and that the phenology variation between years are correctly tracked by the AI.

competition. Normally, one would have to catch the birds to take measurements, but it is possible that bounding boxes from object detection could be used to

monitor growth. To find out, we first manually assigned chick detections to pairs with known hatch dates, based on their location on the ledge, then for each chick

calculated the hourly average bounding box size. We found that bounding box sizes of chicks increased steadily within seasons, reflecting growth (Fig. 4A). In contrast, bounding box sizes of adults showed some variation within and between seasons but not a clear trend (Figure S4). There was a strong correlation between bounding box size and weight-at-age of murre chicks caught and weighed in the Auk Lab (linear regression, $P < 0.0001$, $R^2_{adj} = 0.87$) (Fig. 4B and C), suggesting that our method can potentially be used as a non-invasive method for chick growth estimations. Further work is necessary to determine whether this approach can capture variation between individuals, years and locations, especially as behaviour may also influence bounding box size.

Detection of disturbance events

Another important goal of population monitoring is to identify disturbance events, which may have effects on productivity. A recent study from Stora Karlsö colony revealed that the COVID-19 lockdown in 2020 led to an unusually high rate of disturbances of murres from white-tailed eagles, which in normal years are deterred by the large number of tourists on the island (Hentati-Sundberg, Berglund, et al., 2021). To investigate whether our automated system could be used to identify such events, we analysed the rate of disturbances in 2019, 2020 (lockdown year) and 2021 using the object detection data. We first calculated the maximum number of adult birds per minute and then analysed the change in number of birds from 1 min to another. We defined disturbances as occasions where four or more adult birds left from 1 min to the next (corresponding to >50% of the seven breeding pairs on the ledge in these years). The automated method identified a total of 277 such disturbance events involving 1708 birds. In line with the previous study, disturbances occurred much more frequently during the lockdown year 2020 than in the year before and after (Fig. 5). We validated this finding by randomly selecting 30 AI-identified disturbance events and manually checking the videos, from which we concluded that 29 included a disturbance at the indicated time, whereas one video did show a disturbance event but the time stamp in the video frame was wrong due to an unknown video system error. The automated method also correctly picked up two out of four disturbance events previously reported as supplementary videos in Hentati-Sundberg, Berglund, et al. (2021), whereas the other two videos included disturbances with fewer birds than the 4-bird limit described above, and were thus not picked up by the AI.

Attendance in relation to temperature

Finally, the detections can also be paired with environmental data to study responses to environmental change, such as increasing temperatures. As seabirds often breed on exposed cliffs and beaches, they may be particularly vulnerable to extreme temperatures. Maintaining internal temperatures on warm cliffs may drain both energy and water reserves, and long-term heat exposure could potentially affect physiological processes (Oswald & Arnold, 2012). As such, we expect the murres to spend more time away from sun-exposed breeding ledges at high temperatures.

To explore this relationship between temperature and nest attendance, we combined object detection data from the hottest part of the day (15:00–21:00) with temperature readings at 2-min intervals (using a COMET U0541 datalogger with a PT1000 temperature probe) for 2020, a season with unusually hot weather. For each 2-min interval, we extracted the median number of adults detected on the ledge and paired this with a temperature reading. As we were interested in responses during the active breeding phase, we only looked at the period during which at least 50% of breeding attempts were active.

At all temperatures, there were generally more birds than active breeding attempts, meaning that partners and/or non-breeding birds were present on the ledge (Fig. 6). However, at temperatures above 25°C, we observed a rise in the number of occasions where AI detected exactly the same number of birds as active breeding attempts (i.e. meaning no extra birds present), and at temperatures above 40°C, partners temporarily abandoned eggs and chicks (Fig. 6). These results are in line with results from detailed behavioural (manual) observations of the birds (Olin et al., 2023), where effects of day of year and time of day were also accounted for. We manually scrutinized the videos from the 33 occasions where the AI predicted egg/chick abandonment and concluded that 29 cases actually showed abandonment whereas 4 were due to missing detections, likely due to dirt on the camera lens. The results presented here show how high-resolution object detection data can help to provide fine-scale understanding of animal responses to climate change.

Potential for real-time AI

The field experiments running Yolov5 in the field revealed a clear potential for real-time AI. Despite being significantly slower than a high-end GPU, real-time inference at a 25 FPS frame rate was possible for Yolov5-nano-320px run on a standard laptop CPU, for Yolov5-small-320px
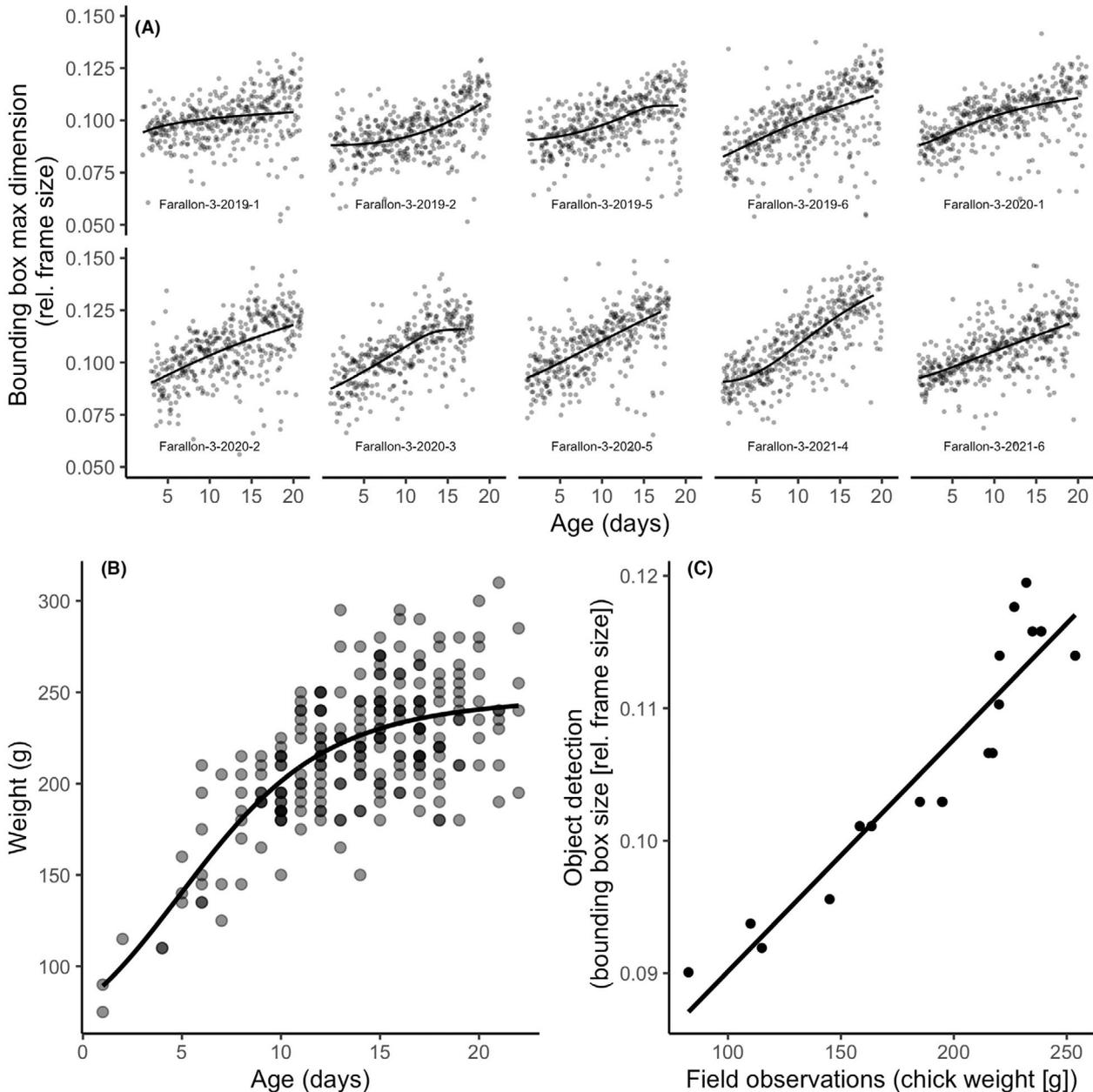
Figure 4. Chick growth as estimated from bounding box sizes in the object detection. (A) Growth curves for 10 chicks in 2019–2021, with hourly mean bounding box size (points) and fitted log-logistic model predictions. Chick IDs indicated in the panels, after the year. (B) Growth curve established in the same colony for murre chicks weighed at known age ($n = 246$) with fitted log-logistic model, and (C) correlation between bounding box size from object detection and weight-at-age, with linear regression line.

run at a Google coral, and for Yolov5-medium-320 as well as Yolov5-small-640 run on a medium-advanced consumer gaming GPU (Nvidia RTX 2080) (Table S4). However, none of these devices were able to run the Yolov5-medium-960 used for ecological inference in this paper in real time. This illustrates the trade-off between speed, accuracy, cost and energy consumption. Although

a high-end computing infrastructure is preferable for accuracy, it may not be realistic in a field setting, suggesting that 'edge' devices such as the Google corals (which could operate near the data collection edge) with low cost, low energy consumption and reasonable computing speed can be an attractive alternative for field and real-time applications.
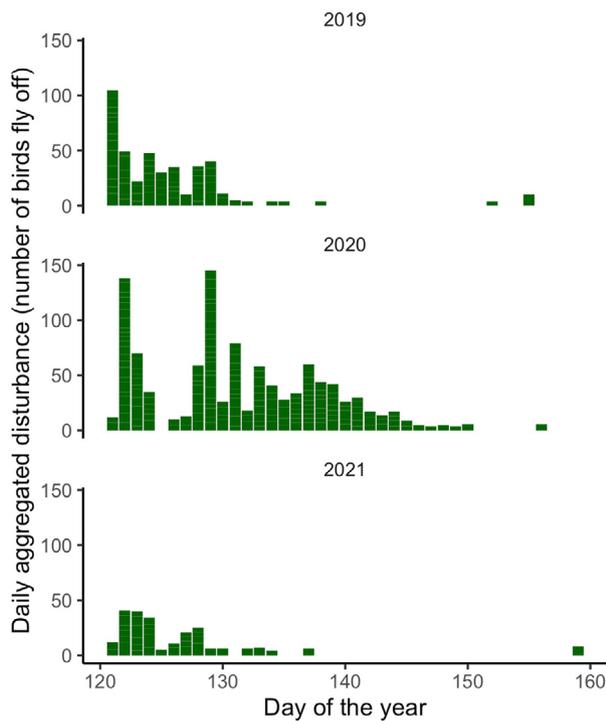
Figure 5. Number and timing of disturbance events (defined as >3 birds departing from the cliff from 1 min to another) in 2019–2021.

Discussion

In this paper, we present a new application of video surveillance combined with deep learning for ecological monitoring and research. In the context of seabirds, we show how our approach can, for example, be used to automatically provide standard monitoring metrics (phenology, nest attendance), detect rare events (eagle disturbances) and explore climate change impacts by linking object detections to automated temperature readings. Our automated method generates data of high resolution and complete temporal coverage, rendering it a useful complement to traditional field-based monitoring, which has greater spatial coverage and flexibility, but only tends to generate snapshots of data. It also reduces the disturbance to the animals and minimizes costs (5600 h of continuous data represents 140 weeks of full-time work if collected manually, and would require multiple field assistants throughout the field season). The approach is easily scalable to include multiple cameras and thereby achieve larger sample sizes.

By monitoring animal behaviour and demography, we can gain insights into broader ecosystem processes. For example, as changes in marine prey stocks are reflected in seabird phenology, attendance patterns and chick growth (Cairns, 1988; Piatt et al., 2007), seabirds essentially

function as ecosystem sensors. Our system uses technological sensors (cameras) to pick up information conveyed through the seabirds, which can ultimately provide up-to-date information on prey stocks, with the potential to feed into local fisheries management (Hentati-Sundberg, Olin, et al., 2021; ICES, 2021). In addition to quickly providing information useful for short-term management decisions, our system also provides data of incredible size and resolution, which can be useful for developing a deeper understanding of ecosystem dynamics (Keitt & Abelson, 2021; Michener & Jones, 2012; Mouquet et al., 2015; Purves, 2013). For example, combining temperature data with nest attendance on a minute-by-minute basis provides part of a much more detailed picture of how seabirds respond to climate change as compared to, for example, combining annual mean temperature with annual demographic metrics.

While machine learning approaches drastically reduce processing time, the set-up can require substantial time and effort (e.g. Ditria et al., 2021). Especially generating a sufficient number of annotations can be time-consuming. However, this time can be cut down by using 'active learning' (selecting an optimal training set) and by enrolling the public in classifying new data and in validating predictions from the model (Jones et al., 2018). We also echo previous calls (Christin et al., 2019; Weinstein, 2018) for increased collaboration between ecologists and computer scientists. In our case, the collaboration with AI Sweden, the Swedish national centre for applied AI, has not only benefited our work but also given them access to new datasets that has been highly useful for developing more widely applicable algorithms and workflows (e.g. a proof of concept for federated learning on video data with potential applications in the autonomous car industry, https://github.com/aidotse/fedbird).

Camera-based ecological monitoring and research is growing in scope and coverage (Bicknell et al., 2016; Norouzzadeh et al., 2021; Steenweg et al., 2017), and we believe there is potential for our setup beyond seabirds. Similar systems could, for example, be deployed at pinniped haul-out sites, watering holes, bat roosting sites or other locations regularly visited by animals. We relied on standard surveillance equipment, a cheap, reliable, and widespread technology, where the number of deployed CCTV cameras has now likely passed 1 billion (CNBC, 2019). In places with limited electricity and/or network connection (which is a common situation in ecological and environmental monitoring), our experiments with edge computing, *that is* AI models that run locally to send summarized information, could be further developed and applied (e.g. Rausch & Dustdar, 2019; Zhou et al., 2019). Possible real-time management
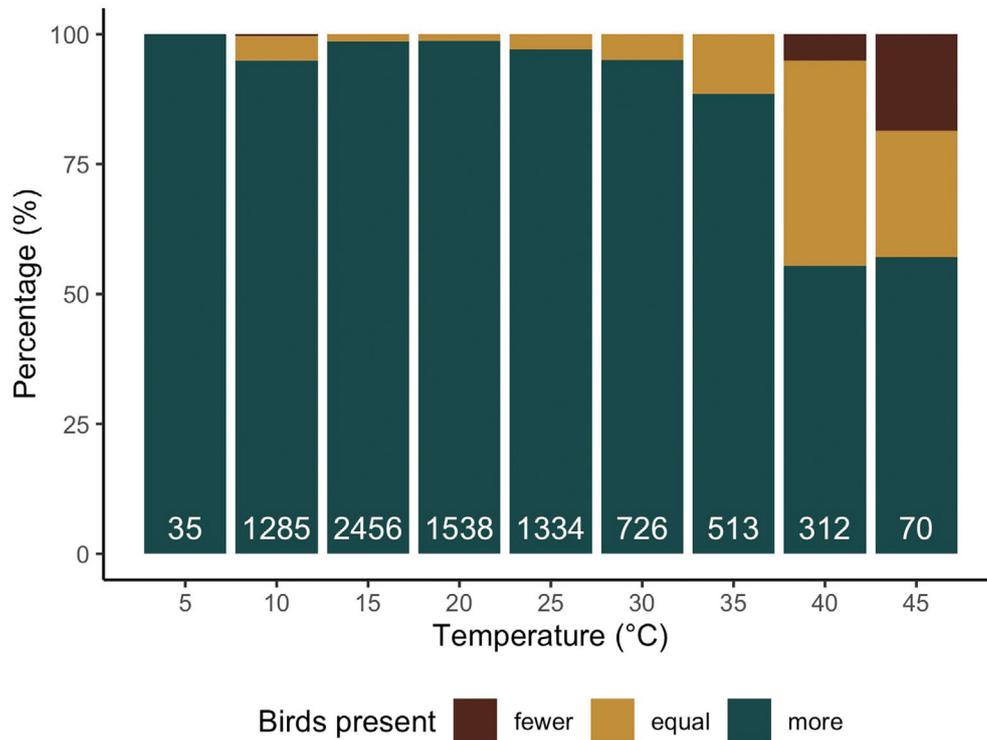
Figure 6. Bird presence in relation to temperature. Bars show temperature 5° intervals with the lower limit of each interval given on the axis. Number of observations per temperature interval is given on top of the bars. At temperatures above 40°, parents temporarily leave the ledge and thereby abandon chicks and eggs, to cool down in the water.

applications could include adjustment of fishing activities near the colony due to changed attendance patterns and/or chick growth, and changes in human activities (restricted access areas, scarecrows) due to eagle disturbances. As for the trained YOLOv5 model, this should be more or less directly applicable in colonies of common murre and the closely related Thick-billed murre (*Uria lomvia*), although local training data with other angles and distances will improve performance and result in a more general model (Weinstein et al., 2022). Priority for further annotations should be given to chicks and eggs, which were detected with lower confidence than adults in the current model version. If the approach is to be extended to other species, the trained model can still be used as a starting point, but should then be re-trained using new, annotated images (socalled transfer learning) (e.g. Weinstein et al., 2022).

While we have shown that useful information can be gained from simply detecting adult birds and their chicks, the described approach could be taken further by also developing a method for individual identification. Seabirds are long-lived animals, and interpreting behavioural changes benefits from knowledge of past experience (e.g. breeding success in previous year) and individual

characteristics (e.g. sex, age) (Clutton-Brock & Sheldon, 2010; McNamara & Houston, 1996). We believe that there is a lot of potential in identifying individuals based on features that are easily overlooked by human observers (see e.g. Chelak et al., 2021; Cheng et al., 2020; Ferreira et al., 2020) and in taking advantage of behaviours and movement patterns to identify individuals (Spiegel et al., 2017). The latter is likely to be particularly successful for seabirds, which often show very high site fidelity, down to the scale of centimetres (Harris et al., 1996). Combining individual identification with target tracking (Athar et al., 2020; Pereira et al., 2020; Walter & Couzin, 2021) would allow for continuous recording behaviour of individuals throughout extended monitoring periods. As such, while we here present results mainly pertaining to the behaviour at the ledge level (e.g. nest attendance, disturbance), this could in the future be specific to the pair (e.g. phenology, attendance and success at the nest level).

In this paper, we have used supervised learning, where algorithms are trained for specific tasks, such as classifying and counting birds, using annotated data. A major and growing field within AI is unsupervised learning, *that is* the identification of patterns without a training dataset

or pre-specified categories (Berry et al., 2019). With constantly improving sensors and processing technologies, we foresee the formation of 'data lakes' in which unsupervised learning algorithms trawl for new and unanticipated dynamics – opening up for research questions beyond our current imagination.

In this time of rapid biodiversity loss, we join the choir of ecologists arguing that we need to make better use of high-speed and high-resolution data collection systems (Hampton et al., 2013; LaDeau et al., 2017; Purves, 2013; Sun & Scanlon, 2019). In addition to supplying the data necessary for sustainable long-term management, real-time information systems offer the possibility for near-immediate responses by decision-makers (Grasso et al., 2019; Sun & Scanlon, 2019). AI has even been suggested as a means to not only automate data collection and analysis but even decision-making and the subsequent actions taken (Cantrell et al., 2017). Consequently, we foresee an increasing role of intelligent surveillance in ecological research and management, aided by the rapid parallel developments of sensors, communication technologies, and AI.

Acknowledgements

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

JHS and OO conceived the original ideas and designed methodology; JHS, PAB, AO and AAC collected data in the field. SR, MR, SiK, ShK, MH and ES trained models and developed analysis pipelines in the lab and in the field; JHS and AO analysed the post-processed object detection data and wrote the manuscript with contributions from all co-authors. All authors gave final approval for publication.

Data Availability Statement

Examples of raw video data, annotations, the final (trained) YOLOv5 model and the SQLite database with all object detections are available for download at Data Dryad Repository (Hentati-Sundberg & Olin, 2022). Python code for training and running the YOLOv5 model is available at https://github.com/BalticSeabird/ SeabirdDetections. R-code for ecological interpretations and figure generations are available at https://github.com/ BalticSeabird/ObjectDetectionInferences.

References

Athar, A., Mahadevan, S., Osep, A., Leal-Taixé, L. & Leibe, B. (2020) STEm-Seg: spatio-temporal embeddings for instance segmentation in videos. *Lecture Notes in Computer Science* (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 12356 LNCS, pp. 158–177.

Berry, M., Azilnah, M. & Yap, B.W. (2019) *Supervised and unsupervised learning for data science*. New York: Springer Nature.

Bicknell, A.W.J., Godley, B.J., Sheehan, E.V., Votier, S.C. & Witt, M.J. (2016) Camera technology for monitoring marine biodiversity and human impact. *Frontiers in Ecology and the Environment*, **14**, 424–432.

Brondizio, E.S., Settele, J., Diaz, S. & Ngo, H.T. (Eds.) (2019) *Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. Bonn: IPBES secretariat, pp. 1148.

Cairns, D.K. (1988) Seabirds as indicators of marine food supplies. *Biological Oceanography*, **5**, 261–271.

Cantrell, B., Martin, L.J. & Ellis, E.C. (2017) Designing autonomy: opportunities for new wildness in the Anthropocene. *Trends in Ecology & Evolution*, **32**, 156–166.

Chelak, I., Nepovinnykh, E., Eerola, T., Kalviainen, H. & Belykh, I. (2021) EDEN: deep feature distribution pooling for Saimaa ringed seals pattern matching. *Arxiv*. [Preprints].

Cheng, X., Zhu, J., Zhang, N., Wang, Q. & Zhao, Q. (2020) Detection features as attention (defat): a keypoint-free approach to Amur Tiger re-identification. *Proceedings - International Conference on Image Processing*. ICIP 2020, Abu Dhabi, United Arab Emirates. October, pp. 2231–2235.

Christin, S., Hervet, É. & Lecomte, N. (2019) Applications for deep learning in ecology. *Methods in Ecology and Evolution*, **10**, 1632–1644.

Clements, C.F., Drake, J.M., Griffiths, J.I. & Ozgul, A. (2015) Factors influencing the detectability of early warning signals of population collapse. *American Naturalist*, **186**, 50–58.

Clutton-Brock, T. & Sheldon, B.C. (2010) Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends in Ecology and Evolution*, **25**, 562–573.

CNBC. (2019) One billion surveillance cameras will be watching around the world in 2021, a new study says. *CNBC*.

Davidson, S., Bohrer, G., Kölzsch, A., Vinciguerra, C. & Kays, R. (2021) Mobilizing animal movement data: API use and the Movebank platform. *Biodiversity Information Science and Standards*, **5**: e74312.

Ditria, E.M., Jinks, E.L. & Connolly, R.M. (2021) Automating the analysis of fish grazing behaviour from videos using image classification and optical flow. *Animal Behaviour*, **177**, 31–37.

Edney, A.J. & Wood, M.J. (2021) Applications of digital imaging and analysis in seabird monitoring and research. *Ibis*, **163**, 317–337.

Farley, S.S., Dawson, A., Goring, S.J. & Williams, J.W. (2018) Situating ecology as a big-data science: current advances, challenges, and solutions. *Bioscience*, **68**, 563–576.

Ferreira, A.C., Silva, L.R., Renna, F., Brandl, H.B., Renoult, J.P., Farine, D.R. et al. (2020) Deep learning-based methods for individual recognition in small birds. *Methods in Ecology and Evolution*, **11**, 1072–1085.

Goodwin, M., Halvorsen, K.T., Jiao, L., Knausgård, K.M., Martin, A.H., Moyano, M. et al. (2022) Unlocking the potential of deep learning for marine ecology: overview, applications, and outlook. *ICES Journal of Marine Science*, **79**, 319–336.

Grasso, I., Archer, S.D., Burnell, C., Tupper, B., Rauschenberg, C., Kanwit, K. et al. (2019) The hunt for red tides: deep learning algorithm forecasts shellfish toxicity at site scales in coastal Maine. *Ecosphere*, **10**, e02960.

Hampton, S.E., Strasser, C.A., Tewksbury, J.J., Gram, W.K., Budden, A.E., Batcheller, A.L. et al. (2013) Big data and the future of ecology. *Frontiers in Ecology and the Environment*, **11**, 156–162.

Harris, M.P., Wanless, S. & Barton, T.R. (1996) Site use and fidelity in the common guillemot *Uria aalge*. *Ibis*, **138**, 399–404.

Hayes, M.C., Gray, P.C., Harris, G., Sedgwick, W.C., Crawford, V.D., Chazal, N. et al. (2021) Drones and deep learning produce accurate and efficient monitoring of large-scale seabird colonies. *Condor*, **123**, 1–16.

Hentati-Sundberg, J., Berglund, P.A., Hejdström, A. & Olsson, O. (2021) COVID-19 lockdown reveals tourists as seabird guardians. *Biological Conservation*, **254**, 108950.

Hentati-Sundberg, J., Olin, A., Evans, T.J., Isaksson, N., Berglund, P.A. & Olsson, O. (2021) A mechanistic framework to inform the spatial management of conflicting fisheries and top predators. *Journal of Applied Ecology*, **58**, 125–134.

Hentati-Sundberg, J. & Olin, A.B. (2022) *Common guillemots in the Baltic Sea studied with video surveillance and object detection: raw data, annotations, model, and model outputs*. Dryad, Dataset. https://doi.org/10.5061/dryad.xsj3tx9hx.

Hentati-Sundberg, J., Österblom, H., Kadin, M., Jansson, Å. & Olsson, O. (2012) The Karlsö Murre lab methodology can stimulate innovative seabird research. *Marine Ornithology*, **40**, 11–16.

Hillebrand, H., Donohue, I., Harpole, W.S., Hodapp, D., Kucera, M., Lewandowska, A.M. et al. (2020) Thresholds for ecological responses to global change do not emerge from empirical data. *Nature Ecology and Evolution*, **4**, 1502–1509.

Hinke, J., Barbosa, A., Emmerson, L.M., Hart, T., Juáres, M.A., Korczak-Abshire, M. et al. (2018) Estimating nest-level phenology and reproductive success of colonial seabirds using time-lapse cameras. *Methods in Ecology and Evolution*, **9**, 1853–1863.

ICES. (2021) *Baltic Sea ecoregion – fisheries overview*. Published 30 Nov 2021. https://doi.org/10.17895/ices.advice.9139

Jones, F., Allen, C., Arteta, C., Arthur, J., Black, C., Emmerson, L.M. et al. (2018) Data descriptor: time-lapse imagery and volunteer classifications from the Zooniverse penguin watch project. *Scientific Data*, **5**, 1–13.

Keitt, T.H. & Abelson, E.S. (2021) Ecology in the age of automation. *Science*, **373**, 858–859.

Kelling, S., Fink, D., La Sorte, F.A., Johnston, A., Bruns, N.E. & Hochachka, W.M. (2015) Taking a 'big data' approach to data quality in a citizen science project. *Ambio*, **44**, 601–611.

Korpela, J., Suzuki, H., Matsumoto, S., Mizutani, Y., Samejima, M., Maekawa, T. et al. (2020) Machine learning enables improved runtime and precision for bio-loggers on seabirds. *Communications Biology*, **3**, 1–9.

LaDeau, S.L., Han, B.A., Rosi-Marshall, E.J. & Weathers, K.C. (2017) The next decade of big data in ecosystem science. *Ecosystems*, **20**, 274–283.

Makiola, A., Compson, Z.G., Baird, D.J., Barnes, M.A., Boerlijst, S.P., Bouchez, A. et al. (2020) Key questions for next-generation biomonitoring. *Frontiers in Environmental Science*, **7**, 1–14.

Marcé, R., George, G., Buscarinu, P., Deidda, M., Dunalska, J., De Eyto, E. et al. (2016) Automatic high frequency monitoring for improved lake and reservoir management. *Environmental Science and Technology*, **50**, 10780–10794.

McNamara, J.M. & Houston, A.I. (1996) State-dependent life histories. *Nature*, **380**, 215–221.

Michener, W.K. & Jones, M.B. (2012) Ecoinformatics: supporting ecology as a data-intensive science. *Trends in Ecology and Evolution*, **27**, 85–93.

Mouquet, N., Lagadeuc, Y., Devictor, V., Doyen, L., Duputié, A., Eveillard, D. et al. (2015) Predictive ecology in a changing world. *Journal of Applied Ecology*, **52**, 1293–1310.

Norouzzadeh, M.S., Morris, D., Beery, S., Joshi, N., Jojic, N. & Clune, J. (2021) A deep active learning system for species identification and counting in camera trap images. *Methods in Ecology and Evolution*, **12**, 150–161.

Olin, A.B., Dück, L., Berglund, P.-A., Karlsson, E., Bohm, M., Olsson, O. et al. (2023) *Breeding failures and reduced nest attendance in response to heat stress in a high-latitude seabird*. Marine Ecology Progress Series.

Oswald, S.A. & Arnold, J.M. (2012) Direct impacts of climatic warming on heat stress in endothermic species: seabirds as bioindicators of changing thermoregulatory constraints. *Integrative Zoology*, 7, 121–136.

Pereira, T. D., Tabris, N., Li, J., Ravindranath, S., Papadoyannis, E.S., Wang, Z.Y. et al. (2020) SLEAP: multi-animal pose tracking. *bioRxiv*. [Preprint]. https://doi.org/10.1101/2020.08.31.276246

Piatt, J.F., Harding, A.M.A., Shultz, M., Speckman, S.G., Van Pelt, T.I., Drew, G.S. et al. (2007) Seabirds as indicators of marine food supplies: Cairns revisited. *Marine Ecology Progress Series*, 352, 221–234.

Purves, D. (2013) Time to model all life on earth. *Nature*, 493, 295–297.

R Core Team. (2021) *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rausch, T. & Dustdar, S. (2019) Edge intelligence: the convergence of humans, things, and AI. *Proceedings - 2019 IEEE International Conference on Cloud Engineering*, IC2E 2019, pp. 86–96.

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

Roe, P., Eichinski, P., Fuller, R.A., McDonald, P.G., Schwarzkopf, L., Towsey, M. et al. (2021) The Australian acoustic observatory. *Methods in Ecology and Evolution*, 12, 1802–1808.

Roemmich, D., Alford, M.H., Claustre, H., Johnson, K.S., King, B., Moum, J. et al. (2019) On the future of Argo: a global, full-depth, multi-disciplinary array. *Frontiers in Marine Science*, 6, 1–28.

Rush, G.P., Clarke, L.E., Stone, M. & Wood, M.J. (2018) Can drones count gulls? Minimal disturbance and semiautomated image processing with an unmanned aerial vehicle for colony-nesting seabirds. *Ecology and Evolution*, 8, 12322–12334.

Scheffer, M., Bascompte, J., Brock, W.A., Brovkin, V., Carpenter, S.R., Dakos, V. et al. (2009) Early-warning signals for critical transitions. *Nature*, 461, 53–59.

Schmeller, D.S., Weatherdon, L.V., Loyau, A., Bondeau, A., Brotons, L., Brummitt, N. et al. (2018) A suite of essential biodiversity variables for detecting critical biodiversity change. *Biological Reviews*, 93, 55–71.

Schofield, D., Nagrani, A., Zisserman, A., Hayashi, M., Matsuzawa, T., Biro, D. et al. (2019) Chimpanzee face recognition from videos in the wild using deep learning. *Science Advances*, 5, 1–10.

Schwaller, M.R., Southwell, C.J. & Emmerson, L.M. (2013) Continental-scale mapping of Adélie penguin colonies from Landsat imagery. *Remote Sensing of Environment*, 139, 353–364.

Spiegel, O., Leu, S.T., Bull, C.M. & Sih, A. (2017) What's your move? Movement as a link between personality and spatial dynamics in animal populations. *Ecology Letters*, 20, 3–18.

Steenweg, R., Hebblewhite, M., Kays, R., Ahumada, J., Fisher, J.T., Burton, C. et al. (2017) Scaling-up camera traps: monitoring the planet's biodiversity with networks of remote sensors. *Frontiers in Ecology and the Environment*, 15, 26–34.

Sun, A.Y. & Scanlon, B.R. (2019) How can big data and machine learning benefit environment and water management: a survey of methods, applications, and future directions. *Environmental Research Letters*, 14, 073001.

van Rossum, G. & Drake, F. (2009) *Python 3 reference manual*. Scotts valley: CreateSpace. Python Software Foundation.

Walter, T. & Couzin, I.D. (2021) Trex, a fast multi-animal tracking system with markerless identification, and 2D estimation of posture and visual elds. *eLife*, 10, 1–73.

Weinstein, B.G. (2018) A computer vision for animal ecology. *Journal of Animal Ecology*, 87, 533–545.

Weinstein, B.G., Garner, L., Saccomanno, V.R., Steinkraus, A., Ortega, A., Brush, K. et al. (2022) A general deep learning model for bird detection in high resolution airborne imagery. *Ecological Applications*, 32, 1–12.

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*). *Behavioral Ecology and Sociobiology*, 74, 7.

Wulder, M.A., Masek, J.G., Cohen, W.B., Loveland, T.R. & Woodcock, C.E. (2012) Opening the archive: how free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment*, 122, 2–10.

Zhou, Z., Chen, X., Li, E., Zeng, L., Luo, K. & Zhang, J. (2019) Edge intelligence: paving the last mile of artificial intelligence with edge computing. *Proceedings of the IEEE*, 107, 1762.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. Outside view of the Karlsö Auk Lab. JHS, PAB and AO from the author team are inside the construction where the electronics (network switch, temperature logger etc.) are located.

Figure S2. Inside view of the Karlsö Auk Lab during camera installation, with several visible network cables and wall-mounted rack cabinet in the background. Optic fibre

cable connects the network switch with the recording unit at the field office, at *c*. 35 m distance.

Figure S3. Performance of Yolov5 for different image and model sizes. (A) Yolov5-small, image width 640 pixels, (B) Yolov5-medium, image width 640 pixels, (C) Yolov5-small, image width 960 pixels, (D) Yolov5-medium, image width 960 pixels. Note how F1 is always higher for adults at a given combination of confidence threshold and IoU threshold and that F1 is highest for all classes in using image size 960 pixels and the medium model.

Figure S4. Size estimations of (A) adult birds and (B) chicks. Only days with more than 10 000 detections were included in the analysis.

Table S1. Number of images and objects per class in the training dataset used for all Yolo models.

Table S2. Performance metrics calculated for the Yolov5-medium-960 model for the validation dataset.

Table S3. Hardwares used for field experiments on real-time object detection. A100 was not used in the field but is included for comparison.