Vantage point photography and deep learning methods save time in monitoring seabird nesting colonies

Rosalin Wilkin,[1,*] Jillian Anderson,[1] Ishan Sahay,[2] Macus Ong,[1] Samantha Broadley,[1] Marine Gonse,[3] Greg McClelland,4 and Ruth Joy[5]

[1] Ecological Restoration Graduate School, Simon Fraser University, Burnaby, British Columbia, Canada

[2] Research Computing Group, Simon Fraser University, Burnaby, British Columbia, Canada

[3] Ecosystem Dynamics and Sustainability, IFREMER, Institut Agro, INRAE, Plouzané, France

[4] Canadian Wildlife Service, Environment and Climate Change Canada, Nanaimo, British Columbia, Canada

5 School of Environmental Science, Simon Fraser University, Burnaby, British Columbia, Canada

* Corresponding author: rosimissi@gmail.com

ABSTRACT
Monitoring seabird colonies is essential for assessing population health and sustainability amid increasing marine industry and climate change. Advancements in photography have led to high-resolution photogrammetry techniques for monitoring seabird colonies. However, manual counting of birds and nests in images, potentially over multiple dates and seasons is time-consuming and has limited a wider adoption of photogrammetry in colony monitoring. We addressed the task of automatically counting cormorants and their nests using an SLT camera, a Gigapan robotic camera mount, and image-stitching software. We applied this system to Vancouver's Ironworkers Memorial Second Narrows Bridge, home to British Columbia's largest *Nannopterum auritum* (Double-crested Cormorant) colony. The system takes overlapping images of the colony from a vantage point to create a panoramic image. We took 23 images of the bridge between April and September 2021. A subset of these images was used to train a deep-learning model that became the foundation of an automated pipeline to detect cormorants of different sizes, positions (standing, incubating, or sunning with wings outstretched), and nests (including only partial glimpses amongst the bridge girders). Our pipeline demonstrated potential for monitoring cormorant populations, by lowering manual effort while achieving high agreement with manual counts. Specifically, our pipeline reduced the manual time required to process images by 96%, while achieving an average agreement of 93.6% between manual and automated counts for both cormorants and nests. We found reduced performance from an application of our model to images of a novel colony; however, we suggest that with additional model training and fine-tuning our pipeline should provide an efficient and accurate alternative to manual counts for other colonial bird monitoring contexts. Our study showcases that high-resolution photogrammetry combined with deep learning methods enables the automatic identification and counting of birds and nests, significantly reducing the time and effort of long-term monitoring of colonially nesting birds.

Keywords: bridge colony, convolutional neural networks, cormorants, large-scale colony monitoring, panoramic images, vantage point survey

How to Cite

Wilkin, R., J. Anderson, I. Sahay, M. Ong, S. Broadley, M. Gonse, G. McClelland, and R. Joy (2025). Vantage point photography and deep learning methods save time in monitoring seabird nesting colonies. Ornithological Applications 127:duaf000.

LAY SUMMARY

- Monitoring seabirds is essential due to their status as indicators for ecosystem health.
- High-resolution photographs are helpful for seabird monitoring, but the time required to manually count the birds and nests in these images limits the number of colonies that can be monitored.
- We counted cormorants and nests using an automated pipeline consisting of a deep-learning object detection model and automated post-processing.
- We applied this system to a bridge that is home to the largest *Nannopterum auritum* (Double-crested Cormorant) colony in BC.
- The automated pipeline reduced human labor time from 225 to 25 min.
- Performance was poorer at a novel bridge site, suggesting that the model would require additional training for application to other contexts.
- Our study showcases that high-resolution photography combined with deep learning methods can efficiently count seabirds and their nests, providing an opportunity for long-term monitoring on nesting colonies.

La photographie à partir de points d'observation et les méthodes d'apprentissage profond permettent de gagner du temps dans le suivi des colonies de nidification d'oiseaux marins

RÉSUMÉ
Le suivi des colonies d'oiseaux marins est essentiel pour évaluer la santé et la durabilité des populations dans un contexte de développement de l'industrie maritime et de changements climatiques. Les progrès en photographie ont permis de mettre au point des techniques de photogrammétrie à haute résolution pour le suivi des colonies d'oiseaux marins. Toutefois, le comptage manuel des oiseaux et des nids sur les images, potentiellement étalées sur plusieurs dates et saisons, prend beaucoup de temps et a limité l'adoption de la photogrammétrie dans le suivi des colonies. Nous nous sommes intéressés au comptage automatique de cormorans et de leurs nids à l'aide d'une caméra à miroir semi-transparent, d'un support de caméra robotique Gigapan et d'un logiciel d'assemblage d'images. Nous avons mis en œuvre ce système à l'*Ironworkers Memorial Second Narrows Bridge* de Vancouver, un pont qui abrite la plus grande colonie de *Nannopterum auritum* en Colombie-Britannique. Ce système prend des images qui se chevauchent de la colonie à partir d'un point d'observation pour créer une image panoramique. Nous avons pris 23 images du pont entre avril et septembre 2021. Un sous-ensemble de ces images a été utilisé pour entraîner un modèle d'apprentissage profond qui est devenu la base d'un pipeline automatisé pour détecter les cormorans de différentes tailles et en différentes positions (debout, en incubation ou au soleil avec les ailes déployées) ainsi que les nids (y compris des aperçus partiels parmi les poutres du pont). Notre pipeline a démontré son potentiel pour le suivi des populations de cormorans, en réduisant l'effort manuel tout en obtenant une grande concordance avec les comptages manuels. Plus précisément, notre pipeline a permis de réduire de 96 % le temps de traitement manuel des images, tout en obtenant une concordance moyenne de 93,6 % entre les comptages manuels et automatisés des cormorans et des nids. Nous avons constaté une baisse des performances lors de l'application de notre modèle à une nouvelle colonie. Cependant, nous suggérons qu'avec un entraînement supplémentaire du modèle et un réglage de précision, notre pipeline devrait fournir une alternative efficace et précise aux comptages manuels dans d'autres contextes de suivi d'oiseaux coloniaux. Notre étude montre que la photogrammétrie à haute résolution combinée à des méthodes d'apprentissage profond permet l'identification et le comptage automatiques des oiseaux et des nids, ce qui réduit significativement le temps et l'effort de suivi à long terme d'oiseaux nichant en colonie.

Mots-clés : colonie sous un pont, réseaux neuronaux convolutionnels, cormorans, suivi de colonie à grande échelle, images panoramiques, inventaire par points d'observation

INTRODUCTION

Seabirds are one of the world's most threatened groups of vertebrates, with 1 in 3 species threatened with extinction (United Nations 2017). They are also effective marine sentinels as top predators, sensitive to marine ecosystem change and anthropogenic stressors, such as increasing pollution, commercial vessel traffic, depleting fish stocks, and climate change (Mallory et al. 2010). While seabirds are a relatively well-studied group, monitoring seabird populations remains challenging. Many species breed in remote or inaccessible areas, limiting our ability to collect demographic data essential to understanding drivers of population change and enabling effective management and conservation decisions.

Photography has long been used to monitor difficult-to-access seabird colonies (Weller and Derksen 1972, Piatt et al. 1990, Heubeck at al.2014), but manually counting nests and birds in these images remains labor intensive, and can introduce a bottleneck in post-processing time, ultimately reducing the number of colonies that can be monitored long term (Spampinato et al. 2015). As a result, the use of photography for monitoring purposes, especially tracking individual nests throughout the year for demographics, remains limited despite significant advances in photography technology.

Recent advances in computer vision models have taken advantage of the increasing complexity and scale of high-resolution images and, when automated, have provided a solution for expanding volumes of image data. Supervised deep-learning algorithms, such as convolutional neural networks (CNNs), use labelled image data to learn to detect features such as birds and their nests. New opportunities for monitoring seabird demographics have been realized through the rapid uptake of remotely piloted aircraft systems (RPAS), or drones, when included in a pipeline tied to computer vision and deep-learning models. In recent years, drone imagery coupled with deep-learning models have made important contributions to seabird colony monitoring (Hodgson et al. 2016, Hayes et al 2021). However, the use of drones is not always logistically and financially feasible (e.g., cost to hire qualified pilots), and for certain regions drone use is restricted or forbidden (e.g., within 5.6 km of Canadian airports). The GigaPan image system is a robotic camera technology that takes hundreds of images that are then stitched into a global high-resolution panorama, also called orthomosaic, using a dedicated software (GigaPan 2013). Initially developed by the National Aeronautics and Space Administration (NASA), GigaPan was adapted for taking high-resolution photos of remote objects, making it particularly well-suited for monitoring remote seabird colonies.

Nannopterum auritum (Double-crested Cormorants) are near-shore fish-eating top predators and as such, important indicators of local marine ecosystem health (Gress et al. 1973; Weseloh et al. 1995). In British Columbia, the number of breeding pairs has declined by ~66% since the 1980s (Adkins et al. 2014, Chastant et al. 2014). In this study, we adapted existing deep-learning methods to analyze images taken using a land-based vantage-point GigaPan image system to monitor a population of urban bridge-nesting cormorants. The GigaPan image system was coupled to a supervised deep-learning model for counting cormorants and their nests on the Ironworkers Memorial Second Narrows (IMSN) bridge, British Columbia, Canada. The IMSN colony has been recommended for annual monitoring due to its importance to the declining population of *N. auritum* in British Columbia (Pacific Flyway Council 2013). The proximity of the colony to the Port of Vancouver creates challenges, including drone restrictions, that have prevented the establishment of a long-term monitoring program. To address this, we developed (1) a non-invasive vantage point method of collecting nest and bird data on an urban bridge without disturbing the breeding colony; (2) built an automated deep-learning pipeline to process these data; and (3) assessed the

feasibility of moving beyond a single image occupancy estimate to a standardized methodology for evaluating nest success, breeding phenology, and colony dynamics within the cormorant breeding season.

METHODS

Study Location

There has been a recent shift of cormorants from natural rookeries to the Ironworkers Memorial Second Narrows (IMSN) bridge in Vancouver, British Columbia. *Nannopterum auritum* were first observed nesting at this urban breeding site in 2009, and their presence has increased gradually since (Carter and Drever 2016, Halpin and Drever 2017, Hemmera 2018). In 2019, the bridge became the largest *N. auritum* breeding colony in the province (Carter and Drever 2016; Halpin and Drever 2017). During the breeding season, the IMSN bridge is home to nesting pairs of cormorants, but also hosts nonbreeding cormorants. Our goal was to count both nests and cormorants (both in and out of nests) present on the bridge.

This study captures the breeding cycle of cormorants in 2021 on the IMSN bridge. Cormorants and their nests on the bridge were counted from April 4th to September 24th, 2021, to align with the cormorant nesting season (Chatwin et al. 2002). We consistently visited the same permanent sampling location (North Shore vantage point) at a 100–150 m distance from the bridge (49.298°, –123.025°) to capture one panoramic image of the cormorant breeding colony on the 2nd span (span 2) of the IMSN bridge (Figure 1).

Data Collection

[LEVEL HEADING 3] **High-resolution panoramic image of the Ironworkers Memorial Second Narrows bridge**

We captured high-resolution panoramas of the IMSN bridge from the vantage point using a Sony α7R IV Digital SLT (DSLT) camera with a Sony FE 200-600 mm lens. The aim was to capture a panorama of the IMSN bridge taken from the same North Shore vantage point, repeated across each week of the breeding season. The panorama was created by using automated control points across adjacent images and stitching together the overlapping high-resolution images. To simplify the stitching of multiple overlapping images, we used a robotic camera mount to create the panorama, a system developed by GigaPan Systems Panoramic Photography Equipment and Software Company (GigaPan 2013, Lynch et al. 2015). Each panorama of the IMSN bridge captured by the GigaPan system contained 3 rows and 11 columns of images and took 12–23 min to capture. The overlapping images were stitched together into a single image using the stitching software program PTGui Pro v11.31 software (https://www.ptgui.com). Once stitched, we exported each panorama as a TIFF file, downsizing most panoramas to 60–90% of their original resolution (depending on the rendered size) to ensure that the size of the file was <4 GB. Each of the 23 stitched panoramas had an approximate dimension of 60,000 × 30,000 pixels.

[LEVEL HEADING 3] **High-resolution panoramic images of Astoria-Megler Bridge**

Since 2015, large numbers of *N. auritum* moved from East Sand Island to begin nesting on the Astoria-Megler bridge that traverses the Columbia River between Astoria, Oregon and

Megler, Washington. The aim was to test if a model trained on one bridge had internally learned features that would transfer to detections at another colony, even when the bridge was significantly different in its architectural structure from the IMSN bridge on which it was trained. To assess the transferability of our panorama vantage point methods from the IMSN bridge to this novel structure, we took one panoramic image of each of 2 different bridge sections. One section with metal bridge trusses was similar to the IMSN bridge but green in color (the IMSN bridge is yellow), while another section was very different from the IMSN bridge with solid steel plate girders that the model was naive to. We created the panoramic images by taking multiple adjacent overlapping images from a boat using the Sony DSLT camera. The GigaPan robotic mount was not used as it relies on a stationary vantage point. Instead, we stitched the multiple overlapping images of a bridge section by matching control points across adjacent images. As there are limited naive locations with bridges that also house cormorant colonies to conduct this test, the image capture methods were modified to allow us to achieve the same quality of resulting panoramic images as at the IMSN bridge. We used the PtGui Pro software to render the panoramas such that the final image was the same dimension as the one for the IMSN bridge.

[LEVEL HEADING 3] **Automated pipeline construction**

To reduce the time required to manually count cormorants and their nests in the panoramic images, we built an automated counting pipeline using a deep-learning object detection model with a convolutional neural network (CNN) architecture training pipeline. Object detection is the computer-driven task of detecting instances of items in digital images and videos. A CNN is a supervised machine learning method that automatically learns to recognize increasingly complex visual features in images (e.g., edges, textures, and patterns) and uses these features to classify an image or locate objects in an image. CNNs can be trained on unstructured data, such as panoramic images, to detect objects such as cormorants and nests, thereby reducing dependence on biologists at the computer screen. Using a CNN as our foundation, we built an automated pipeline (Figure 2) to count cormorants and nests on the IMSN bridge. To build this pipeline, we followed 4 steps: (1) pre-processing, (2) annotation, (3) training, and (4) prediction & post-processing.

[LEVEL HEADING 3] **Pre-processing of panoramas**

Each panoramic image was split into 1,000 × 1,000 pixel tiles rendering 2,500–3,000 non-overlapping tiles per panoramic image. Previous studies applying CNN object detection algorithms to the field of wildlife monitoring used overlapping tiles to ensure each object is fully contained within at least one tile (Eikelboom et al. 2019, Hayes et al. 2021, Kabra et al. 2022, Lawrence et al. 2023). However, since bridge structures can obscure birds and nests from view, our model needed to learn to detect partial birds and nests. Thus, we used non-overlapping tiles to train the object detection model.

[LEVEL HEADING 3] **Annotation pipeline**

The object detection model was trained on a set of manually created annotations. Each annotation consists of a box and label, which together provide the location of a cormorant or a nest. The training set was incrementally built over 5 rounds of annotation, at which point we found training on additional annotations provided little improvement to model performance. In each training round, a set of tiles were uploaded to a self-hosted instance of

the *LOST* (Label Objects and Save Time; Jäger et al. 2019) annotation tool. Using this tool, human annotators drew labelled bounding boxes around all cormorants and nests (partial or whole) visible in the tile. In all but the initial annotation round, annotators were shown candidate annotations proposed by a CNN fine-tuned on the annotation data from previous rounds. These candidate annotations could then be removed, edited, or re-labelled by the annotator.

Once an annotation round was complete, the newly annotated tiles were divided into training, validation, and testing sets. Each round of annotation saw these sets grow, with new annotations added to the datasets of the previous rounds. For our final round of annotation (round 5), all newly annotated tiles were placed into the training set. The final annotated dataset was composed of a training set (85%), validation set (7.5%), and testing set (7.5%) produced from 1186 annotated tiles. The ratio between training, validation, and testing sets can vary depending on the task and amount of data available, with training data commonly comprising 70–90% of the total dataset (Eikelboom et al., 2019; Hayes et al., 2021; Lawrence et al., 2023; Weinstein et al., 2022). Ultimately, the goal in selecting a split ratio is to strike a balance between having enough data to train a model while also having enough data to assess its performance on new data.

[LEVEL HEADING 3] Training pipeline

The TensorFlow Object Detection API (Yu et al. 2020) provides access to deep-learning models pre-trained and evaluated on the Microsoft COCO dataset (Lin et al. 2014). This widely used benchmark dataset contains over 200,000 labelled images across 80 object categories. Based on the benchmarked metrics reported in the TensorFlow Object Detection API, we selected the CNN model architecture named CenterNet Hourglass104 512x512 (from now on referred to as CenterNet) due to its balance of precision and speed for fine-tuning with our dataset of cormorants and nests.

We adjusted several hyperparameters to optimize the model for our particular task. These adjustments included removing data augmentation transformations inconsistent with our ecological context (e.g., vertically flipping tiles) and limiting the number of objects the model could detect in a single tile (see Supplementary Material Table S1 for detailed information on all adjustments).

[LEVEL HEADING 3] Prediction pipeline and post-processing

Once training was complete, novel panoramas were tiled (as in the annotation pipeline) and fed through the trained CenterNet Hourglass model to generate a set of predicted detections. Each predicted detection consisted of a class label (cormorant or nest), bounding-box position within the image, and a detection score indicating the model's relative confidence in that detection. Finally, these detections underwent two post-processing steps, masking and nest deduplication, to allow comparison of automated pipeline results with manual counts.

First, we removed model predictions outside the IMSN span 2 region (Figure 1), using the polygon masking tool in *LOST* (Jäger et al. 2019). For each panorama, a polygon mask was drawn around the focal Span 2 region. For each detection, an "inner rectangle" was created by reducing the bounding box dimensions by 50% while keeping the center point. Only those detections whose inner rectangles intersected with the Span 2 polygon mask were kept.

Second, we merged duplicate detections of individual nests. A consequence of

training the model to identify partial objects was that a single object may be detected multiple times in adjacent tiles. While automatic deduplication provides a solution by identifying and merging duplicate detections, it also risks merging distinct detections. Since nests were both larger than birds (~1.4x larger) and much less likely to be near one another (typically built outside of "pecking-distance", or at least 70 cm away from the next nest), we used automatic deduplication for nests alone. A buffer of 10 pixels was added to the bounding box of each nest detection. If two buffers overlapped across a tile boundary, the nests were considered duplicates and merged into a single detection.

This prediction and post-processing procedure was also used to detect cormorants and nests in two panoramas of the Astoria-Megler bridge. The application of our automated pipeline to a location which it had not been tuned for provides an opportunity to evaluate the generalizability of the automated pipeline.

[LEVEL HEADING 3] **Evaluating automated pipeline performance**

We evaluated the performance of the pipeline by comparing automated pipeline results with human counts, using a set of 8 panoramas collected in 2020 that were excluded from model training and pipeline development. We manually annotated these 8 panoramas using the *LOST* annotation tool (Jäger et al. 2019) and then conducted 2 rounds of manual review. After manual review, we corrected annotation errors and treated the resulting set of annotations as our ground-truth. We used this ground-truth data to evaluate the entire automated pipeline performance and the performance impact of each post-processing step using mean average precision (mAP) and average recall (AR), variations of precision and recall.

Precision is the ratio of correct detections to total detections made by a model (i.e., True Positives / (True Positives + False Positives)). Recall is the ratio of correctly detected objects to all objects (i.e., True Positives / (True Positives + False Negatives)). In object detection models there are 2 thresholds that determine whether a model prediction contributes to the true positive (TP) or false positive (FP) count: detection score, and intersection-over-union (IoU). Each prediction made by the model includes a detection score between 0 and 1 that represents the relative confidence of the model in that prediction. Predictions with detection scores below a specified threshold are not considered detections and therefore contribute to neither the TP nor FP counts. Predictions above the detection score threshold are either a TP or FP prediction, depending on whether the prediction's bounding box corresponds to a ground-truth annotation. This is determined by calculating the IoU (area of intersection/area of union) between the predicted and annotated bounding-boxes, labelling those which exceed the IoU threshold as TPs.

mAP is a way to measure how well a model performed by averaging precision across object classes (cormorants and nests) and recall values (obtained by adjusting the detection score threshold). mAP can be calculated either across a range of IoU thresholds or for a fixed IoU threshold. Because our primary goal was to obtain accurate census counts, we prioritized object detection over localization and used a fixed IoU threshold of 0.1 to avoid penalizing localization errors. Localization errors occur when the model detects but incorrectly places an object, yielding an IoU between 0.1 and 0.5 (Hoiem, et al., 2012). While we consider these detections successful, the standard $mAP^{IoU=0.5}$ metric classifies these detections as false positive errors and their corresponding annotations as false negatives, thereby artificially lowering precision and recall. To avoid conflation of detection and localization performance, we used $mAP^{IoU=0.1}$ to assess the impact of each post-processing step in addition to the overall

automated pipeline's performance.

Average recall, AR, summarizes how well the model identifies objects by averaging across classes and a range of IoU thresholds from 0.5 to 0.95. However, rather than adjusting the detection score threshold, AR specifies a maximum number of detections allowed in an image. Given panoramas may contain hundreds of objects (cormorants and nests), standard variations of AR ($AR^{max=1, 10, 100}$) would not suffice. Instead, using $AR^{max=400}$ limited per-image detections to 400 and allowed theoretical detection of all objects while continuing to remove low-scoring detections from consideration.

In addition to calculating individual $mAP^{IoU=0.1}$ and $AR^{max=400}$ metrics, we also created a Precision-Recall (PR) curve for each of our two object classes (cormorants and nests) to visualize the automated pipeline's tradeoff in precision and recall. Each PR curve shows how precision changes as we adjust the detection score threshold to detect more or fewer objects, which affects recall, while keeping the IoU threshold fixed to 0.1. To construct the curve, we calculate the corresponding precision for 101 recall values evenly spaced between 0 and 1. By plotting precision against recall, we can visualize the impact of changes to one metric on the other. We also use the PR curves to find the detection score threshold which provides the best balance between precision and recall. This is done by calculating the F1 score, the harmonic mean of precision and recall, for every point on the PR curves. The point with the highest F1 score represents the best trade-off in precision and recall, and corresponds to an optimal detection score threshold for each class. Using these detection score thresholds, we counted the number of nests and cormorants detected by our pipeline using the 23 panoramic images taken of IMSN Span 2 during the 2021 breeding season in addition to 2 panoramic images taken of the Astoria-Megler bridge. The generated counts from the Astoria-Megler bridge were compared with corresponding manual counts to help us evaluate the generalizability of the pipeline to new sites.

The time and effort required to generate counts was also used to evaluate the potential of the automated pipeline as an alternative to previous counting methods. Prior to the development of the automated pipeline, research assistants used the default image software on their personal computers (e.g., Preview on macOS) to count cormorants and nests manually within panoramic images. The time required to generate these manual counts is contrasted with the time required to generate counts using the automated pipeline. These time comparisons do not include the upfront time needed to train research assistants (to conduct manual or automated counts), train deep-learning models, or deploy the automated pipeline.

RESULTS

Model Performance

After 5 rounds of annotation, human annotators had reviewed 1,186 tiles, generating 2,055 cormorant annotations and 1,292 nest annotations. The fine-tuned CNN model achieved an $mAP^{IoU=0.1}$ of 0.56 for cormorants, and 0.66 for nests (Table 1). Each post-processing step provided incremental improvements to the performance of the automated pipeline (Table 1). The final automated pipeline obtained a $mAP^{IoU=0.1}$ of 0.82 on cormorants and 0.91 on nests. Further, it achieved an $AR^{max=400}$ and 0.94 for nests. These results were consistent across all sections of the bridge, regardless of the distance of the section from the vantage point. Figure 4 provides additional detail on the performance of the pipeline by presenting precision-recall curves for both cormorants and nests, noting the highest F1 score for cormorants was 0.80,

and for nests was 0.89. These results are consistent with other applications of object detection to ecological contexts, which have obtained F1 scores between 0.72 and 0.92 (Hayes et al. 2021, Fennell et al. 2022, Pillay 2022, Green et al. 2023).

We compared counts generated by the automated object-detection pipeline with manual counts in 5 novel panoramas (i.e., those not used in model training). There was an average difference of 6.4 % between the nest counts generated from the 2 methods, with automated counts consistently lower than manual counts (Figure 5). There was a similar 6.4% average difference between model-based and manual cormorant counts but no consistent bias in the direction of the difference (i.e., the automated pipeline was neither consistently higher nor lower than the manual counts) (Figure 6). Three of the 5 cormorant count comparisons were accurate to within 3%. Similar magnitude differences occurred when comparing the lead annotator (i.e., lead author vs co-author counts) to other counters of the same image and found the counts between human annotators differed on average by 8.5% for cormorants and 2% for nests.

We repeated this automated CenterNet Hourglass object-detection pipeline for 2 very different structural sections of the Astoria-Megler bridge on which the model was not trained to assess its transferability to another context. The aim was to compare the automated pipeline counts and the manual counts to assess the transferability of the automated pipeline for cormorant and nest detection at other breeding colonies (Yosinski et al. 2014). For the panorama with a similar structure to the IMSN bridge but with green instead of yellow metal trusses, the cormorant count difference was 7.2% (<1% off the IMSN bridge performance), while the difference for nests was 60.4% (~10 times the IMSN bridge performance). For the panorama of the bridge section with steel plate girders, the difference in cormorant and nest counts was 28.4% and 77.2%, respectively. These results suggest that our automated pipeline could be used to detect cormorants on other bridges with comparable bridge architectures. However, to apply the pipeline to sites that are quite different from the IMSN bridge, including natural colonies, additional CNN model training is needed.

Time and Effort Requirements

The comparison between the manual and automated-pipeline methods revealed significant differences in the time and effort requirements of each method (Table 1). Overall, generating counts manually took 255 min for each image: 15 min to collect the image and 240 min to count the cormorants and nests. Alternatively, generating counts using the automated pipeline only took 40 min for each image: 15 min to collect the image, 10 min to configure the automated pipeline to run on a high-performance computing cluster, and 15 min for the automated counting pipeline to execute. Because automated counting was an unsupervised task, the automated pipeline approach required only 25 min of human labor time for each image. In particular, the automated pipeline resulted in a 96% reduction in the time required for active computer work compared with manual counting.

Common Errors in the Automated Pipeline

In general, automated pipeline errors can be grouped into 3 categories: (1) incorrect detections (false positives), (2) missing detections (false negatives), and (3) duplicate detections. Most often, incorrect detections occurred when *Columba livia* (Rock dove) were counted as cormorants (Figure 6A) or bridge features were detected as nests (Figure 6B). Almost all duplicate detections occurred when an object (nest or cormorant) was split across tiles. Since post-processing attempts to merge nests spanning tiles, the most common source

of duplicate detections was cormorants bifurcated across vertically adjacent tiles (Figure 6C). Finally, the automated pipeline most often missed detecting cormorants when only a small portion of the bird was visible (Figure 6D) and missed detecting nests when they were obscured by bridge beams (Figure 6E).

Application of Automated Cormorant and Nest Counting Pipeline

The panoramic images of the IMSN bridge from the North Shore vantage point covered 23 days of the 2021 breeding season (Figure 7). We ran the automated pipeline on all 23 panoramic images. The results indicate the number of nests increased from a low of 26 nests on April 4th, to a maximum of 155 nests detected on June 9th, before declining throughout the remainder of the breeding season. A similar pattern was seen with cormorant counts. The automated pipeline counted 121 cormorants on April 4th, our first day of monitoring. These counts increased until they reached a peak of 284 cormorants on May 12th, a date which precedes nestlings being large enough to be seen in panoramic images. Cormorant counts were variable from May 12th to July 5th, reflecting the life history of shared incubation and feeding of nestlings. After July 5th, counts consistently declined until reaching a low of 19 cormorants on September 24th, the final day of monitoring. This survey period encompassed the nest building, egg laying and chick-rearing season of 2021 for cormorants on the IMSN bridge.

DISCUSSION

Our study built a semi-automated deep-learning pipeline to process images across the breeding season of a cormorant nesting colony. We found that combining CNN object-detection with the convenience of a GigaPan image-capture system from a ground-based vantage point enabled us to create a high-resolution time series of cormorant and nest numbers across the breeding season on a Vancouver highway bridge. Our automated pipeline overcame the human labor bottleneck of manually counting hundreds of birds and nests for a season's worth of breeding colony images while producing counts within 10% of manual counts. We found the automated object detection pipeline was a time-saving alternative to manual counting by processing a single image in one-tenth the time it would take to process the same image manually.

The CNN model performance was improved by two post-processing steps. Masking out non-bridge regions of the images from the model resulted in a 0.26 increase in mAP for cormorants (0.56 to 0.82) and a 0.19 increase for nests (0.66 to 0.85). This is not surprising as this focused the model on the foreground bridge, and away from the background image that included grain elevators, boats, pedestrians and their pets, and other animals. The second post-processing step of identifying and merging duplicate nest detections in adjacent tiles improved the mAP of nests by 0.06. However, this step was not applied to cormorant detections, as it would have resulted in merging many distinct cormorants and led to a reduced mAP. More training on partial objects may improve the performance of the model. Nevertheless, our model achieved good performance for the vantage point of span 2 of the IMSN bridge with an overall F1 score of 0.80 for cormorants and 0.89 for nests.

The experiment of applying the object detection pipeline to a novel environment at the Astoria-Megler bridge revealed the limitations and opportunities of deploying the same model on additional colonies. We demonstrated transferability of the object detection algorithm for detecting cormorants on the beams of a similar (i.e., green instead of yellow)

bridge. However, the transferability to nest detection was not convincing and the automated pipeline did not perform well on novel bridge components. We recommend researchers interested in using our object detection model in novel locations to use location-specific human-labelled data to fine-tune our model. Previous work demonstrates that fewer than 1,000 annotations are needed to do such fine-tuning (Weinstein et al. 2022). In general, the object-detection pipeline proposed here has broad uses for other applications; however, fine-tuning is imperative before relying on the model for object detection tasks at new locations, or for new species.

Drone imagery has been combined with CNN-based object detection models in recent studies of seabirds in panoramic images (e.g., Hayes et al. 2021, Kellenberger et al. 2021, Cusick et al. 2024). Model performance metrics achieved by drones flying directly overhead of horizontal seabird colonies are typically higher than the F1 scores we report in this study (e.g., F1 scores > 0.90 in the studies listed above). This is a consequence of the simplicity of these colonies and the object detection task compared to the complexity of the bridge structure where cormorants build nests on crossbeams partially hidden from the vantage point view. As such, this should not be viewed as a weakness of our CNN model pipeline. The present study shows that it is possible to take advantage of the benefits of deep-learning models without the expense of a drone and can be done similarly with a more affordable land-based camera and tripod setup. This study suggests there is broad application for the GigaPan image system to be used in any setting where a suitable vantage point is available.

As many seabirds are vulnerable to anthropogenic disturbances that lead to the stress response of nest flushing and nest abandonment (Götmark 1992), there are additional motivations for using the GigaPan image-capture system beyond saving time. Conventional monitoring surveys of cormorants conducted using in-person observation on foot or by boat can cause disturbance and stress (Hipfner and Greenwood 2009, Chatwin et al. 2013). A suitably located observation blind is one way to reduce disturbance, however, this is not often logistically possible including at the bridges that are the focus of this study. As an alternative, imagery from drones is growing in popularity for surveying seabird colonies and may decrease colony disturbance (Lalach et al. 2023). However, some species such as *Pygoscelis adeliae* (Adelie Penguin) showed significant negative response to drones (Rümmler et al. 2016). The convenience of a GigaPan image system from a ground-based vantage point is that images on a high-quality camera can be easily captured from a distance of 250 m or more (250 m in this study, or 800 m in Lato et al. 2012) reducing or eliminating the risk of disturbing a nesting bird.

One further goal of our study was to establish the feasibility of moving beyond a single image of cormorant occupancy to create a standardized methodology for evaluating colony success within and across years. A benefit of the GigaPan and object detection pipeline is that it standardizes the image collection process. It eliminates the problem of different researchers reporting slightly different counts for cormorants and nests for the same image. Similarly, drones flown at different heights or by different pilots on different flight routes introduces variability that will affect CNN model performance. The proposed GigaPan image capture from a standardized vantage point eliminates this variability within an image or between flight lines and enables a stronger comparison between panoramas and across years.

While this study successfully developed machine learning methods for monitoring bridge-nesting cormorants, not all seabird colony monitoring projects will necessarily benefit from automated counting. For example, a project with limited scope may only collect images

on a single date, such as the peak of the breeding season (Lalach et al. 2023). In these cases, the upfront effort required to manually annotate data, train a model, and validate results may exceed the benefits of transitioning from manual to automated counting. However, as advised by Hentati-Sundberg (2023) and demonstrated in this study, active learning can be used to reduce this initial time investment by selecting an optimal training set.

Conclusion

Automated detection of seabirds and their nests using supervised deep-learning algorithms, such as convolutional neural networks, are at the cutting edge of methodology in seabird biology. The rising popularity of these models in seabird monitoring projects promises to open up the time and energy to ask more complex questions about these colonies and extract more data than would be possible with traditional surveys. This study was motivated by applying a new generation of camera and computing technology and associated big data tools to a classic ecological question, "How many nests are in this colony?". The large volume of image data from the IMSN bridge that is georeferenced to the beam in this study, creates opportunities to explore more ecological questions, with greater scientific complexity. The use of machine learning for seabird monitoring is growing, and this study demonstrates that projects of varying scope, time frame, species, and locations are challenges that can be addressed with similar pipelines to those presented here.

Acknowledgements
We thank Hoky Hsu and Rachel Stapleton for field support and Laurence Cojocaru for his help in the early development of the deep learning model. We are grateful for graduate student support from the Habitat Conservation Trust Fund's Conservation Economic Stimulus Initiative. We are indebted to Jenna Cragg and Ariel Lenske for their support of this project, and to Trudy Chatwin for planting the idea 'seed'.

Funding statement

This research was enabled in part by support provided by the Canadian Wildlife Service of Environment and Climate Change Canada, the Digital Research Alliance of Canada (alliance can.ca) and an NSERC Discovery Grant to R. Joy.

Ethics statement

We followed Simon Fraser University's Animal Care Protocol: 1337ES-22.

Conflict of interest statement

The authors declare no conflict of interest.

Author contributions

R.W., M.O., and R.J. conceived the idea. R.W., J.A., and I.S. developed and designed methods. R.W., M.O., and S.B. collected data in the field; R.W., J.A., and S.B. extracted data from images and model metrics. R.W, J.A., and I.S. analysed the data. R.W., J.A., I.S., S.B., M.G., G.M., and R.J. wrote or substantially edited the paper. G.M. and R.J. contributed substantial materials, resources, and funding.

Data availability
Analyses reported in this article can be reproduced using the data provided by Wilkins et al. (2025).

LITERATURE CITED

Adkins, J. Y., D. D. Roby, D. E. Lyons, K. N. Courtot, K. Collis, H. R. Carter, W. D. Shuford, P. J. Capitolo (2014). Recent population size, trends, and limiting factors for the Double- crested Cormorant in the western North America: Double- crested Cormorant population trends. *The Journal of Wildlife Management* 78:1131–1142.

Carter, H. R., and M. C. Drever (2016). Double-crested Cormorants and other seabirds nesting at the Second Narrows Bridge and Power Tower in Vancouver, British Columbia, in 2016. Unpubl. Report, Carter Biological Consulting, Victoria, British Columbia, Canada.

Chastant, J. E., D. T. King, D. V. C. Weseloh, and D. J. Moore (2014). Population dynamics of Double- crested Cormorants in two interior breeding areas. *The Journal of Wildlife Management* 78:3–11.

Chatwin, T. A., R. Joy, and A. E. Burger (2013). Set-back distances to protect nesting and roosting seabirds off Vancouver Island from boat disturbance. *Waterbirds* 36:43–52.

Chatwin, T. A., M. H. Mather, and T. D. Giesbrecht (2002). Changes in pelagic and double-crested cormorant nesting populations in the Strait of Georgia, British Columbia. *Northwestern Naturalist* 83:109–117.

Cusick, A., K. Fudala, P. P. Storozenko, J. Swiezewski, J. Kaleta, W. C. Oosthuizen, C. Pfeifer, and R. J. Bialik (2024). Using machine learning to count Antarctic Shag (*Leucocarbo bransfieldensis*) nests on images captured by Remotely Piloted Aircraft Systems. *Ecological Informatics* 82:102707.

Eikelboom, J. A. J., J. Wind, E. van de Ven, L. M. Kenana, B. Schroder, H. J. de Knegt, F. van Langevelde, and H. H. T. Prins (2019). Improving the precision and accuracy of animal population estimates with aerial image object detection. *Methods in Ecology and Evolution* 10:1875–1887.

Fennell, M., C. Beirne, and A. C. Burton (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. *Global Ecology and Conservation*, 35:e02104.

GigaPan (2013) About Gigapan. http://www.gigapan.com/cms/about-us

Götmark, F. (1992). The effects of investigator disturbance on nesting birds. *Current Ornithology* 9:63–104.

Green, K. M., M. K. Virdee, H. C. Cubaynes, A. I. Aviles-Rivero, P. T. Fretwell, P. C. Gray, D. W. Johnston, C.-B. Schönlieb, L. G. Torres, and J. A. Jackson (2023). Gray whale detection in satellite imagery using deep learning. *Remote Sensing in Ecology and Conservation* 9:829–840.

Gress, F., R. W. Risebrough, D. W. Anderson, L. F. Kiff and J. R. Jehl, Jr. (1973). Reproductive failures of Double-crested Cormorants in southern California and Baja California. *The Wilson Bulletin* 85:197–208.

Halpin, L. R., and M. C. Drever (2017). Double-crested Cormorants, Pelagic Cormorants and other seabirds nesting in Vancouver and the Strait of Georgia, British Columbia, in

2017. Unpublished report, Halpin Wildlife Research, Vancouver, British Columbia, Canada.

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

Hentati- Sundberg, J., A. B. Olin, S. Reddy, P. A. Berglund, E. Svensson, M. Reddy, S. Kasarareni, A. A. Carlsen, M. Hanes, S. Kad, and O. Olsson (2023). Seabird surveillance: combining CCTV and artificial intelligence for monitoring and research. *Remote Sensing in Ecology and Conservation* 9:568–581.

He, K., X. Zhang, S. Ren, and J. Sun (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, June 27–30*. pp. 770–778.

Hemmera (2018) Ironworkers Memorial Second Narrows Crossing bird management plan – phase 1: avian use study. Prepared for BC Ministry of Transportation and Infrastructure. Hemmera Envirochem Inc, Burnaby, British Columbia, Canada.

Heubeck, M., S. Gear, and M. P. Harris (2014). A photographic resurvey of seabird colonies on Foula, Shetland. *Scottish Birds* 34:291–302.

Hipfner, J., and J. Greenwood (2009). Timing and success of breeding in Pelagic Cormorants at Triangle Island, British Columbia, 2003–2008. *Northwestern Naturalist* 90:238–243.

Hodgson, J. C., S. M. Baylis, R. Mott, A. Herrod, and R. H. Clarke (2016). Precision wildlife monitoring using unmanned aerial vehicles. *Scientific Reports* 6:22574.

Jäger, J., G. Reus, J. Denzler, V. Wolff, and K. Fricke-Neuderth (2019). November 4. LOST: A flexible framework for semi-automatic image annotation. *arXiv* 1910.07486

Kabra, K., A. Xiong, W. Li, M. Luo, W. Lu, R. Garcia, D. Vijay, J. Yu, M. Tang, T. Yu, H. Arnold, A. Vallery, R. Gibbons, and A. Barman (2022). Deep object detection for waterbird monitoring using aerial imagery. In *2022 IEEE International Conference on Machine Learning and Applications (ICMLA), Nassau, Bahamas, December 12–15*. pp. 455–460.

Kellenberger, B., T. Veen, E. Folmer, and D. Tuia (2021). 21,000 birds in 4.5 h: Efficient large- scale seabird detection with machine learning. *Remote Sensing in Ecology and Conservation* 7:445–460.

Lalach, L. A. R., D. W. Bradley, D. F. Bertram, and L. K. Blight (2023). Using drone imagery to obtain population data of colony-nesting seabirds to support Canada's transition to the global Key Biodiversity Areas program. *Nature Conservation* 51:155–166.

Lato, M. J., G. Bevan, and M. Fergusson (2012). Gigapixel imaging and photogrammetry: Development of a new long range remote imaging technique. *Remote Sensing* 4:3006–3021.

Lawrence, B., E. de Lemmus, and H. Cho (2023). UAS-based real-time detection of Red-

cockaded Woodpecker cavities in heterogeneous landscapes using YOLO object detection algorithms. *Remote Sensing* 15:883.

Lin, T.-Y., M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. Lawrence Zitnick (2014). Microsoft COCO: Common objects in context. In *Computer Vision – ECCV (13th European Conference on Computer Vision) 2014, Zurich, Switzerland, September 6–12*. pp. 740–755.

Lynch, T. P., R. Alderman, and A. J. Hobday (2015). A high- resolution panorama camera system for monitoring colony- wide seabird nesting behaviour. *Methods in Ecology and Evolution* 6:491–499.

Mallory, M. L., S. A. Robinson, C. E. Hebert, and M. R. Forbes (2010). Seabirds as indicators of aquatic ecosystem conditions: A case for gathering multiple proxies of seabird health. *Marine Pollution Bulletin* 60:7–12

Pacific Flyway Council (2013). A Monitoring Strategy for the Western Population of Double-crested Cormorants within the Pacific Flyway. Pacific Flyway Council, U.S. Fish and Wildlife Service, Portland, Oregon, USA.

Piatt, J. F., B. D. Roberts, and S. A. Hatch (1990). Colony attendance and population monitoring of Least and Crested Auklets on St. Lawrence Island, Alaska. *The Condor* 92:97–106.

Pillay, N. (2022). *Counting Animals in Ecological Images*. Faculty of Science, Department of Statistical Sciences, University of Cape Town, South Africa. http://hdl.handle.net/11427/36534

Rümmler, M.-C., O. Mustafa, J. Maercker, H.-U. Peter, and J. Esefeld (2016). Measuring the influence of unmanned aerial vehicles on Adélie Penguins. *Polar Biology* 39:1329–1334.

Spampinato, C., G. M. Farinella, B. Boom, V. Mezaris, M. Betke, and R. B. Fisher (2015). Special issue on animal and insect behaviour understanding in image sequences. *EURASIP Journal on Image and Video Processing* 2015:1–4.

United Nations (Editors) (2017). Seabirds. In *The First Global Integrated Marine Assessment: World Ocean Assessment I*. Cambridge University Press, Cambridge, UK. 763–72.

Weinstein, B. G., L. Garner, V. R. Saccomanno, A. Steinkraus, A. Ortega, K. Brush, G. Yenni, A. E. McKellar, R. Converse, C. D. Lippitt, and A. Wegmann (2022). A general deep learning model for bird detection in high- resolution airborne imagery. *Ecological Applications* 32:e2694.

Weller, M. W., and D. V. Derksen (1972). Use of time-lapse photography to study nesting activities of birds. *The Auk* 89:196–200.

Weseloh, D. V., P. J. Ewins, J. Struger, P. Mineau, C. A. Bishop, S. Postupalsky and J. P. Ludwig (1995). Double-crested Cormorants of the Great Lakes: Changes in population size, breeding distribution and reproductive output between 1913 and 1991. *Colonial Waterbirds* 18:48–59.

Wilkin, R., J. Anderson, I. Sahay, M. Ong, S. Broadley, M. Gonse, G. McClelland, and R.

Joy (2025). Data from: Vantage point photography and deep learning methods save time in monitoring seabird nesting colonies. *Ornithological Applications* 127:duaf000. 10.20383/103.01188 [Dataset].

Yosinski, J., J. Clune, Y. Bengio, and H. Lipson (2014). How transferable are features in deep neural networks? In *Twenty-seventh International Conference on Neural Information Processing Systems, Montreal, QC, Canada, December 8–13*. pp. 3320–3328.

Yu, H., C. Chen, X. Du, Y. Li, A. Rashwan, L. Hou, P. Jin, F. Yang, F. Liu, J. Kim, and J. Li (2020). TensorFlow Model Garden. https://github.com/tensorflow/models

Zaidi, S. S. A., M. S. Ansari, A. Aslam, N. Kanwal, M. Asghar, and B. Lee (2022). A survey of modern deep learning-based object detection models. *Digital Signal Processing* 126:103514.

FIGURE 1. Span 2 of the Ironworkers Memorial Second Narrows (IMSN) Bridge from the North Shore vantage point, Vancouver, British Columbia. We included detections within the area marked by the yellow line, while detections outside this area were "masked out" (excluded) from the post-processed counts.

FIGURE 2. Depiction of the process used to build the automated pipeline for cormorant and nest detection on the Ironworkers Memorial Second Narrows (IMSN) Bridge. First, we split input panoramas into smaller 1,000 × 1,000 pixel tiles (Pre-processing). We then draw bounding boxes around cormorants and nests as annotations (Annotation Pipeline). The CNN model learned from the provided annotation examples (Training Pipeline) to draw bounding boxes on new unseen panoramic images (Prediction Pipeline). Then, two post-processing steps corrected for double counting the same nest captured on adjacent tiles and for identifying objects outside our area of interest (Post-processing). The remaining bounding boxes of cormorants and nests were summed to produce counts (Final Output) of the IMSN bridge.

FIGURE 3. Detections made by the final pipeline on a panorama of Ironworkers' Memorial Second Narrows bridge to generate counts of the second span of the bridge (Span 2). The inset provides a zoomed-out view of span 2 of the bridge, while the red rectangle notes the location corresponding to the zoomed-in view. The white grid shows the boundaries of the 1,000 × 1,000 pixel tiles, green boxes denote cormorant detections, and pink boxes denote nest detections.

FIGURE 4. Precision-recall curves comparing the cormorant and nest detections made by the complete automated pipeline. Circles indicate the highest F1 score for the Nest (F1 score = 0.89) and Cormorant (F1 score = 0.80) pipelines. The maximum F1 scores occurred at detection score thresholds of 0.21 for cormorants and 0.20 for nests. The F1 score is the harmonic average of precision and recall and is a single metric widely used to evaluate model performance.

FIGURE 5. Comparison of manual counts with the automated pipeline's counts on 5 of the 23 panoramic images of the IMSN bridge taken during the 2021 breeding season. Percent values along the top axis represent the percent difference between the manual and post-processed counts. Negative values correspond to lower post-processed counts relative to manual counts, while positive values correspond to higher post-processed counts compared to manual counts.

FIGURE 6. A series of images highlighting the complexity of the image task and common errors made by the automated pipeline. Teal boxes show correct detections, solid gold boxes show incorrect detections, and dashed gold boxes indicate objects missed by the automated pipeline. In panel (**A**), the model did not identify the gull but mislabeled the *Columba livia* (Rock dove). In panel (**B**), the model confused bridge rivets for a nest. In panel (**C**), the model detected the cormorant in two adjacent tiles causing double counting. In panel (**D**), the

model has missed the partial cormorant outlined in the shadows. In panel (**E**), the nest is obscured by bridge beams and was overlooked by the model.

FIGURE 7**.** Comparison of the number of nests (left) and cormorants (right) counted on the bridge bracings of Span 2 of the IMSN bridge across the 2021 breeding cycle. Each of the 23 bars corresponds to the post-processed CenterNet Hourglass model counts of a single panoramic image from the same North Shore vantage point. The width of the bar corresponds to the number of days between the vantage point surveys.

TABLE 1. Performance of the automated pipeline on 2020 data improves with each post-processing step (i.e., increasing rows have increased mean average precision). Cormorant and Nest $mAP^{IoU=0.1}$ compare model detections to expert annotations (i.e., what proportion of the predicted cormorants/nests were annotated by experts). The "Overall" column averages the $mAP^{IoU=0.1}$ values.

	$mAP^{IoU=0.1}$		
	Cormorants	**Nests**	**Overall**
Model	0.56	0.66	0.61
Model + masking	0.82	0.85	0.83
Model + masking + merging duplicate nests	0.82	0.91	0.87

TABLE 2. Time and effort requirements to count cormorants and nests in a single panoramic image of Span 2 at the peak of the nesting season on the IMSN bridge.

	Time (per image)				
	Time in the field	**Active computer time**	**Automated processing time**	**Total time**	**Human labor time**
Manual counts	15 min	240 min	0 min	255 min	255 min
Automated-pipeline counts	15 min	10 min	15 min	40 min	25 min

Figure 1

Figure 2

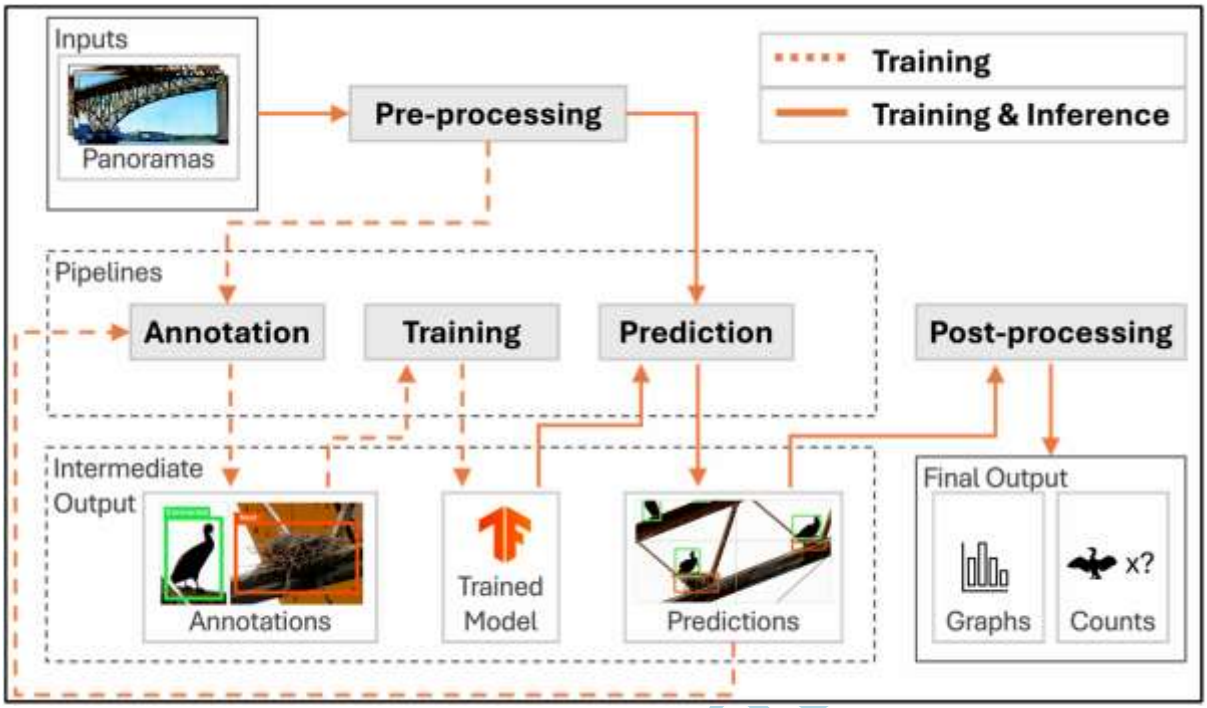

Figure 3

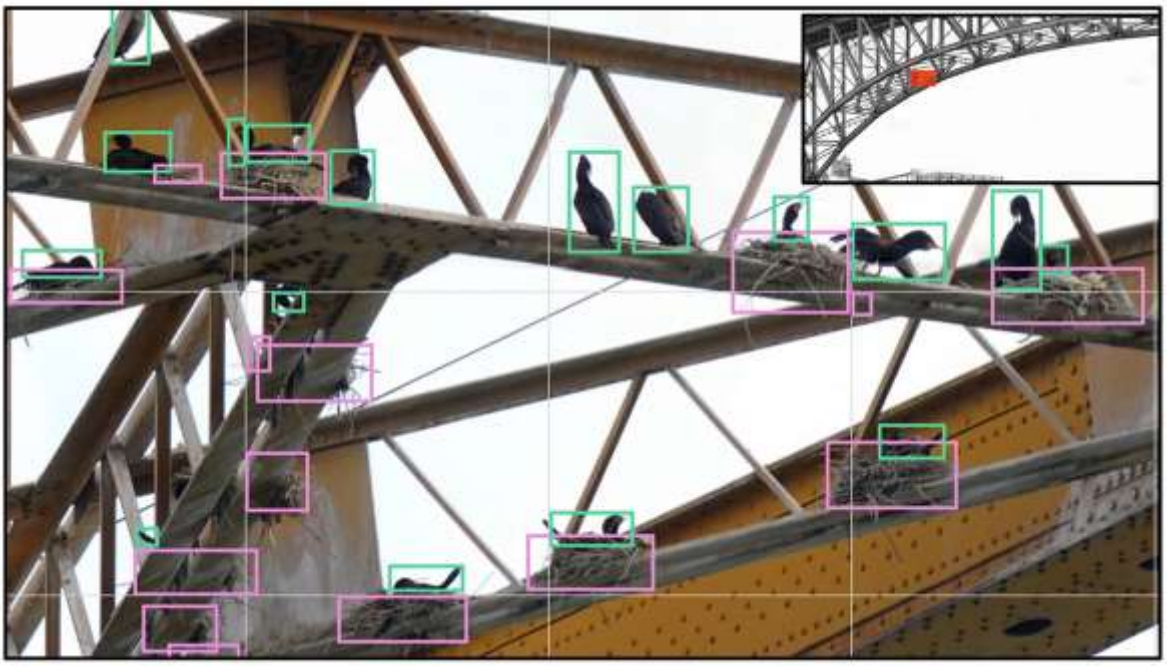

Figure 4

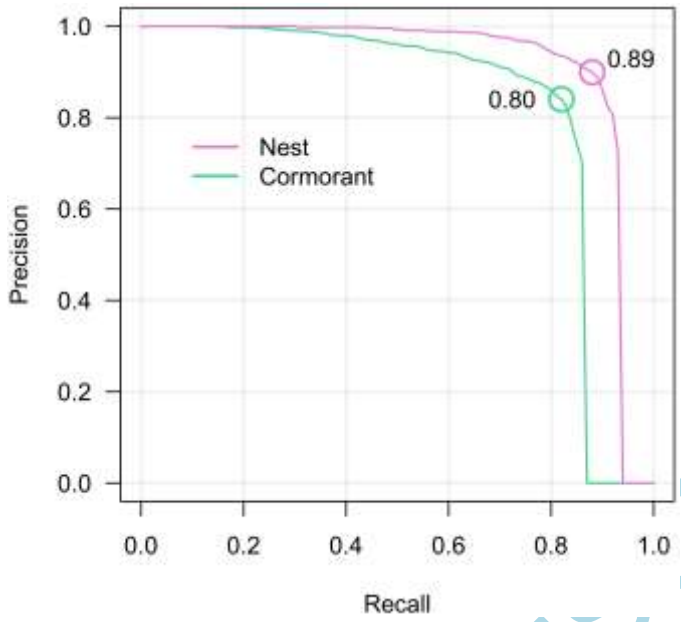

Figure 5

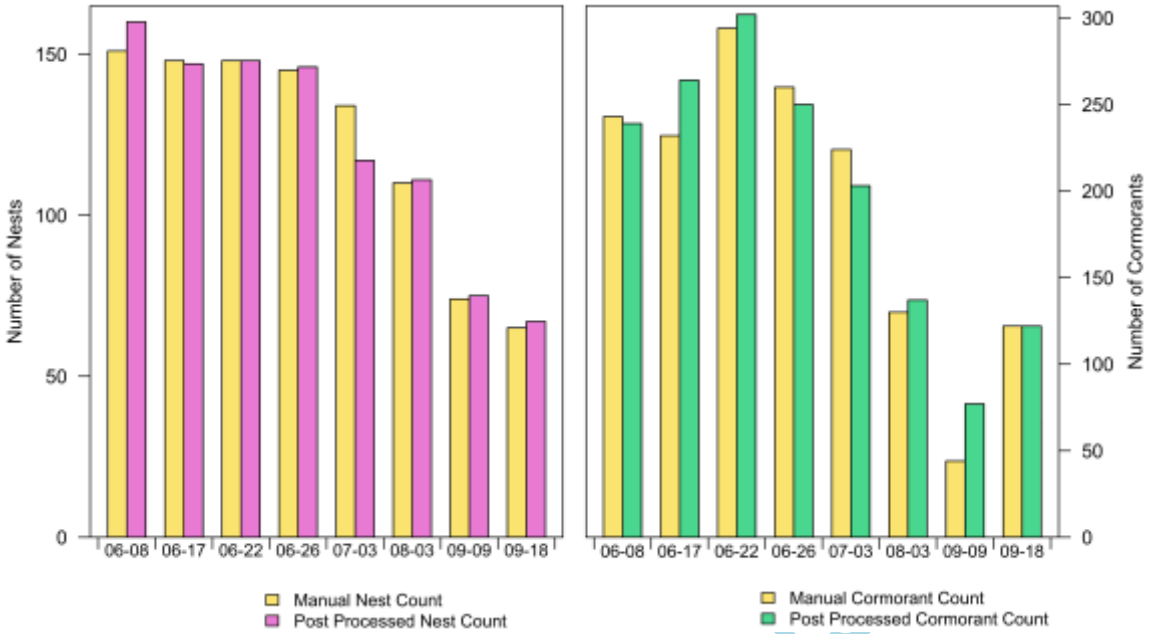

Figure 6

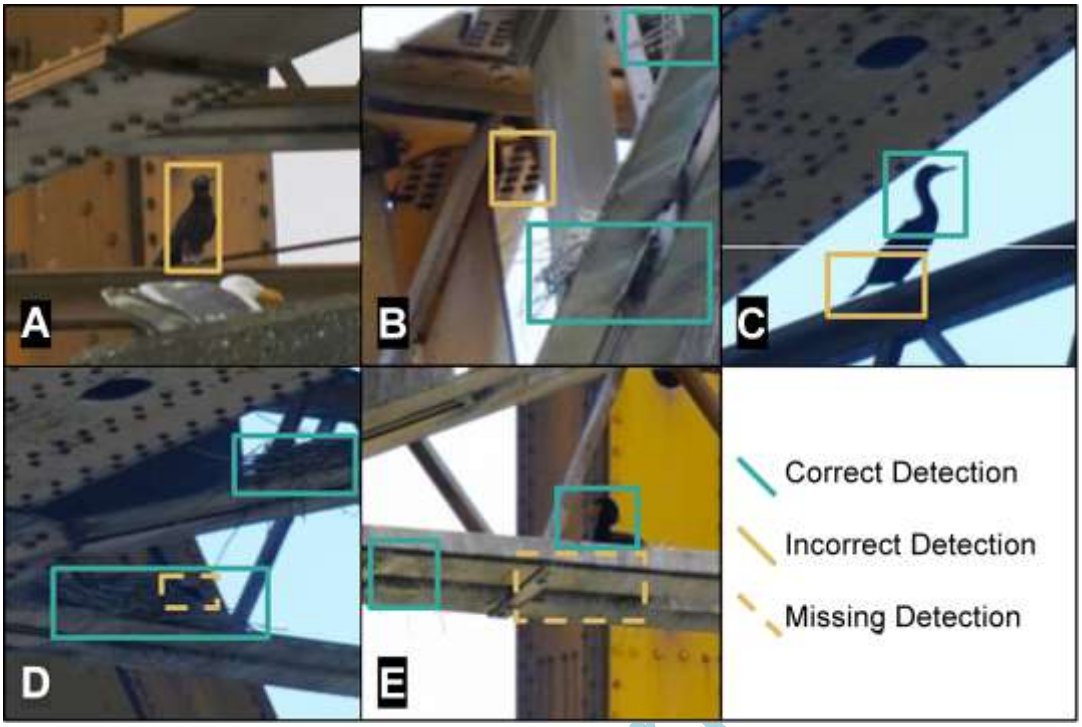

Figure 7